Diagramación en Latex: Universidad Nacional de Colombia, Medellín.
Editado y diagramado en Medellín.

Índice

ROPOSAL FOR THE MONITORING THE CLEANING OF THE BOGOTÁ RIVER TO APPEAR ON ITS ECOSYSTEM SERVICES

CAROLINA VILLEGAS VARGAS

COMITÉ EDITORIAL REVISTA DE LA FACULTAD DE CIENCIAS

COMITÉ CIENTÍFICO

Revista de la Facultad de Ciencias Universidad Nacional de Colombia, Sede Medellín
V 11 N°2 julio-diciembre 2022 • ISSN-e 2357-5749

EDITORIAL

VÍCTOR IGNACIO LÓPEZ RÍOS[a], JOSÉ A. MONTOYA[b]

Este número de la revista lo integran cinco artículos relacionados con el tópico de verosimilitudes planas, un tópico especial en el área de la Estadística, a continuación se presenta una pequeña introducción al respecto.

La función de verosimilitud es un ingrediente fundamental en varios enfoques de inferencia estadística, entre los que figuran el método bayesiano, la verosimilitud integrada, el método de la verosimilitud maximizada o perfil, entre otros. En particular, la función de verosimilitud es la piedra angular del método clásico de estimación por máxima verosimilitud y sus estimadores generalmente poseen propiedades asintóticas deseables para un marco teórico de inferencia estadística. Cuando se cuenta con una muestra observada, de tamaño finito, resultado de la realización de un experimento, la función de verosimilitud es una herramienta inferencial que posee información sobre valores del vector de parámetros que explican con mayor probabilidad lo que se ha observado. Cuando se estandariza la verosimilitud de manera que su máximo sea uno, se obtiene la verosimilitud relativa, la cual permite ordenar valores posibles del vector de parámetros con base en qué tan probable hacen a la muestra observada, ello con respecto a la probabilidad máxima de observar la muestra. Esta virtud de la función de verosimilitud respecto a jerarquizar los valores posibles del vector de parámetros, a la luz de la muestra observada, permite atender tanto de forma intuitiva como formal problemas estadísticos de inferencia tales como estimación puntual, estimación por intervalos, pruebas de hipótesis, pruebas de significancia, entre otros. Aún más, cuando el vector de parámetros es de dimensión mayor que uno, la función de verosimilitud contiene información sobre las relaciones paramétricas heredadas, tanto del modelo de probabilidad y su parametrización, como de la cantidad y calidad de la muestra observada.

En situaciones donde la forma de la función de verosimilitud de un parámetro escalar es aproximadamente acampanada, su análisis es sencillo y las inferencias a partir de ésta son esencialmente simples, pues la forma de esta clase de funciones de verosimilitud se puede caracterizar o aproximar bastante bien con solo dos cantidades estadísticas: el estimador de máxima verosimilitud y la información observada de Fisher. Ambas cantidades, ampliamente discutidas en la literatura estadística, son de gran utilidad pues el estimador de máxima verosimilitud permite identificar la localización de la función de verosimilitud en el espacio paramétrico (dominio de la función) y la información observada de Fisher describe la curvatura de la función a su alrededor. Así, con ambas estadísticas se puede reconstruir la forma de la función de verosimilitud y en consecuencia las inferencias que produce esta función. Algo similar ocurre con superficies de verosimilitud o en dimensiones mayores de esta función, donde formas cuadráticas que involucran al estimador de máxima verosimilitud (un vector) y a la matriz de información observada de Fisher, caracterizan la forma de función de verosimilitud.

Por otra parte, cuando la forma de función de verosimilitud de un parámetro escalar dista mucho de aproximarse a una normal o ser acampanada y además es difícil encontrar una reparametrización que lo logre, entonces el estimador de máxima verosimilitud y la información observada de Fisher resultan estadísticas que por sí solas son incapaces

[a]Editor en Jefe Revista Facultad de Ciencias, Ph. D. Profesor Titular Escuela de Estadística, Universidad Nacional de Colombia, Sede Medellín.

[b]Ph. D. en Probabilidad y Estadística. Profesor de Estadística, Departamento de Matemáticas. Universidad de Sonora, México.

de reconstruir la forma de esta clase de verosimilitudes y reproducir todas las inferencias que se obtienen a partir de ellas. En estas situaciones, antes de cualquier inferencia o decisión de buscar procedimientos alternativos, se requiere analizar la forma completa de la función de verosimilitud, con la finalidad de determinar la génesis de dicha forma.

En el caso de verosimilitudes planas, éstas son generalmente un síntoma de problemas inferenciales relacionados con la naturaleza del modelo (que incluye su especificación), reparametrizaciones, sobreparametrización, cantidad y calidad de los datos experimentales, porcentaje de censura en los datos, bondad y ajuste del modelo a los datos, entre otros. Aunque en la literatura estadística existen estudios que han arrojado luz sobre lo que en realidad está detrás del problema de verosimilitudes planas, es necesaria una mayor difusión de situaciones en las que pueden ocurrir este tipo de verosimililtudes y las múltiples causas que pueden originarlas. En general, es indispensable profundizar en el análisis de la forma de la función de verosimilitud, pues sólo así se tendrá una fundamentación razonable, en caso de criticar las inferencias obtenidas a partir de ésta.

Los primeros cinco artículos abordan diversas situaciones en las que pueden surgir verosimilitudes planas: insuficiente tamaño de muestra, problema de no-identificabilidad práctica, ocurrencia de modelos empotrados, modelos anidados, así como una relación subyacente entre los parámetros del modelo. Algunos de los ejemplos que se incluyen en estos manuscritos son ampliamente conocidos en la literatura estadística y fueron considerados con la finalidad de mostrar problemáticas específicas asociadas a verosimilitudes planas. En la mayoría de los manuscritos se incluye un amplio análisis de diversos escenarios de simulación.

El sexto artículo se enmarca en el área de las matemáticas. En este artículo los autores presentan los gráficos existenciales para el cálculo proposicional paraconsistente, $KT4P$. El sistema $KT4P$, se encuentra caracterizado por una semántica de mundos posibles, además, $KT4P$ es paraconsistente, es decir, no colapsa en la presencia de contradicciones. Los autores presentan todas las pruebas de manera completa, rigurosa y detallada.

El séptimo artículo se enmarca en el área de la Química, como una contribución a la enseñanza del equilibrio químico. Los autores presentan el formalismo que integra la cinética y la termodinámica y lo aplican al estudio del modelo de reacción $A \rightleftharpoons B$. Los resultados obtenidos permiten explicar con claridad los principales aspectos cinéticos y energéticos que caracterizan el equilibrio químico.

El octavo artículo se enmarca en el ára de las Ciencias Naturales. El autor propone un indicador para hacerle seguimiento a la recuperación de los servicios ecosistémicos del río Bogotá, y de esta manera darle contexto biológico, social, cuantitativo y cualitativo a la sentencia del Consejo de Estado del año 2014.

El comité Editorial agradece enormemente los valiosos comentarios emitidos por los árbitros nombrados para la evaluación de todos los artículos presentados en este número y que permitieron mejorar sustancialmente la calidad de cada uno de ellos.

Revista de la Facultad de Ciencias Universidad Nacional de Colombia, Sede Medellín
V 11 N°2 julio-diciembre de 2022 • ISSN-e 2357-5749 • Research paper • Pages 8 to 24
DOI: https://doi.org/10.15446/rev.fac.cienc.v11n2.97888

FLAT LIKELIHOODS: BINOMIAL CASE[a]

VEROSIMILITUDES PLANAS: CASO BINOMIAL

JOSÉ A. MONTOYA[b*]

Recibido 22-08-2021, aceptado 9-12-2021, versión final 16-12-2021
Research Paper

ABSTRACT: The shape of the likelihood function is often considered as one of the underlying causes of strange or counterintuitive estimation results. However, strange likelihood shapes may be a symptom of inferential issues related with the nature of the model and experimental data. In the cases discussed here, binomial flat likelihoods are related not only to sample size, but also to an embedded Poisson model problem. It is essential to understand the shapes of the likelihood function in order for being able to legitimately criticize likelihood inferences. This is particularly important since the likelihood function is a key ingredient in many inferential methods.
KEYWORDS: Flat likelihood; threshold parameter; embedded model; Poisson distribution; likelihood contours; profile likelihood function.

RESUMEN: La forma de la función de verosimilitud es frecuentemente considerada como una de las causas subyacentes de resultados de estimación extraños o contradictorios. Sin embargo, formas extrañas de la verosimilitud pueden ser un síntoma de problemas inferenciales relacionados con la naturaleza del modelo y los datos experimentales. En los casos discutidos aquí, verosimilitudes binomiales planas están relacionadas no solamente con el tamaño de muestra sino también con un problema de un modelo Poisson empotrado. Es fundamental entender las formas de la función de verosimilitud con el fin de poder criticar, legítimamente, inferencias por verosimilitud. Esto es de particular importancia puesto que la función de verosimilitud es un ingrediente fundamental en muchos métodos inferenciales.
PALABRAS CLAVE: Verosimilitud plana; parámetro umbral; modelo empotrado; distribución Poisson; contornos de verosimilitud; función de verosimilitud perfil.

1. INTRODUCTION

The likelihood function is the basis of the classical maximum likelihood estimation method. In fact, under regularity conditions on the model, maximum likelihood estimates are strongly consistent, asymptotically normal, and asymptotically efficient. In addition, the likelihood function plays a key role in Bayesian inference, integrated likelihood method, profile likelihood method, and many others inferential statistic approaches.

[a]Montoya, J. A. (2022). Flat Lihelihoods: Binomial Case. *Rev. Fac. Cienc.*, 11 (**2**), 8–24. DOI: https://doi.org/10.15446/rev.fac.cienc.v11n2.97888

[b]PhD in Probability and Statistics. Statistics Professor. Department of Mathematics. University of Sonora, México.
[*]Corresponding author: arturo.montoya@unison.mx

The shape of the likelihood function is often cited as an underlying cause of these strange or unintuitive results, and academic literature has consistently illustrated that statistical methods based on the likelihood function can lead to misleading, strange, unstable, preposterous or ridiculous estimates (Harter & Moore, 1966; Breusch *et al.*, 1997; Berger *et al.*, 1999; Martins & Stedinger, 2000; Pewsey, 2000; Martins & Stedinger, 2001; El Adlouni *et al.*, 2007; Tumlinson, 2015). The criticisms to maximum likelihood estimation, associated with strange likelihood functions that grow rapidly to infinity, and that arise from the use of density functions having singularities, are invalid. This issue is well known; see Montoya *et al.* (2009), Liu *et al.* (2015), and the references cited therein. However, there are some others criticisms to maximum likelihood estimation that occur quite often and that arise when the shape of the likelihood function is flat. In fact, flat likelihoods are used to promete integrated likelihoods, Bayesian posteriors, penalized methods and new optimization methods (Berger *et al.*, 1999; Tsionas, 2001; Frery *et al.*, 2004; Li & Sudjianto, 2005; Ghosh *et al.*, 2006; Cole *et al.*, 2013; Lima & Cribari-Neto, 2019).

Some authors have related the problem of flat likelihoods with an overparameterization of the model, inappropriate reparametrizations, embedded models, a limited amount and quality of experimental data, and also a poor model fit when sample size is large (Cheng & Iles, 1990; Catchpole & Morgan, 1997; Raue *et al.*, 2009; Sundberg, 2010; Farcomeni, & Tardella, 2012; Kreutz *et al.*, 2013). Although these studies are based on relevant statistical models (exponential family of distributions, three-parameters distributions, capture-recapture models and dynamical models), and they helped to shed some light on what lies behind the problem of flat likelihoods, more work is needed to expand our understanding.

In this article we focus our research on the binomial model; perhaps one of the simplest models in statistical literature, but one with *ad hoc* characteristics for a novel study of flat likelihoods. For instance, when parameters N and p are unknown, this model does not belong to the exponential family. In addition, the parameter N is the right extreme of the support of the binomial distribution; that is, the binomial probability model is a non-regular discrete model. On the other hand, the binomial model has an embedded Poisson distribution; that is, the limiting case of the two-parameter binomial distribution is the one-parameter Poisson distribution. Furthermore, the binomial model has a low dimension parametric space (2-dimensional), which in principle shows no signs of overparameterization. Historically the number of Bernoulli trials N has had a different logical level than the one considered on p, in the sense that quantity N has been usually considered fixed and known, while the probability of success p of each Bernoulli trial has been conceived unknown. Now, in the context of abundance estimation, where N represents the size of the population to be estimated and the parameter p is the probability of capture, the value of N could be determined completely through a census; but the determination of the exact value of p does not seem to be so simple because it represents a much more abstract concept.

On the other hand, it is quite attractive to address the problem of estimating parameters N and p, based on independent success counts $x_1, x_2, \cdots, x_k$ from a binomial distribution. This problem has been considered hard

for many scientists, since classic estimators of N tend to be unstable under slight perturbations of the sample, so many statistic literature focusses on providing stable estimators for N (Draper & Guttman, 1971; Olkin *et al.*, 1981; Carroll & Lombard, 1985; Casella, 1986; Kahn, 1987; Raftery, 1988; Hall, 1994; Gupta *et al.*, 1999; DasGupta & Rubin, 2005). It is noteworthy that only few statistical literature provides the graph of the binomial likelihood function. Berger *et al.* (1999) and Aitkin & Stasinopoulos (1989) had shown that the shape of the profile and conditional likelihood function of N can be essentially flat; that is, it can be nearly constant over a large range of N values.

In this paper we perform simulations to investigate the conditions under which the binomial model leads to flat likelihoods. In all cases, we analyze the limit of the relative profile likelihood of N when parameter N goes to infinity, and show that flat likelihoods occur frequently. In most cases, this limit value has an empirical distribution that assigns a negligible probability to values close to zero. Specifically, for a fixed sample size, the mass of the empirical distribution for this limit value is shifted to the right of zero, assigning a large probability to values close to one when N, the 100th percentile of the binomial distribution, moves away from Np, the expected value of the binomial random variable. We also show that flat likelihoods frequently occur when data coming from a Poisson model, embedded in the binomial distribution, are modeled by a binomial model; this occurred for all the sample sizes here analyzed. Therefore, when faced to the problem of estimating parameters of the binomial model, we suggest adding information into inference process and even changing the observation protocol. However, in any case, we emphasize the importance of plotting the likelihood function to assess the presence of flat likelihoods.

In general, when dealing with flat likelihoods we should not only focus on finding stable estimators or solving numerical problems during the optimization of the likelihood function, but also tackle the issues involved in the genesis of such behaviour. The major contribution of this paper is in honing our intuition and understanding of the multifactorial nature of the flat likelihoods problem, contributing in some way to the scant literature concerning this subject.

2. BINOMIAL LIKELIHOOD

Suppose $X_1, \cdots, X_k$ are i.i.d. binomial distributed random variables with unknown parameters (N, p), and assume that N is the parameter of interest and p a nuisance parameter. The probability mass function of a binomial random variable X_i is

$$P(X_i = x; N, p) = \binom{N}{x} p^x (1-p)^{N-x},$$

where $x = 0, 1, 2, \cdots, N$ and $0 \leq p \leq 1$. We here denote the binomial distribution as $Bin(N, p)$.
The likelihood function of (N, p) based on an observed sample $\boldsymbol{x} = (x_1, \cdots, x_k)$ is

$$\begin{aligned} L(N,p) &= \prod_{i=1}^{k} \binom{N}{x_i} p^{x_i} (1-p)^{N-x_i} \\ &= p^t (1-p)^{kN-t} \left[\prod_{i=1}^{k} \binom{N}{x_i} \right], \end{aligned} \tag{1}$$

for $0 \leq p \leq 1$ and $N \geq x_{\max}$, where $t = \sum_{i=1}^{k} x_i$ and $x_{\max} = \max\{x_1, \cdots, x_k\}$. The maximum likelihood estimate (MLE) of (N, p) is the value of (N, p) which maximizes (1). Unfortunately, this estimator does not have a closed-form expression, so it must be numerically approximated. The MLE of (N, p) will be denoted here as $\left(\widehat{N}, \widehat{p}\right)$.

The relative likelihood function of (N, p) is obtained by dividing the likelihood in (1) by the maximum value it takes, corresponding to its value at the MLE,

$$R(N,p) = \frac{L(N,p)}{\max\limits_{N,p} L(N,p)} = \frac{L(N,p)}{L\left(\widehat{N}, \widehat{p}\right)}. \tag{2}$$

The profile likelihood of N is easy to compute, since the restricted maximum likelihood estimate of p is $\widehat{p}(N) = \widehat{\mu}/N$, where $\widehat{\mu} = t/k$. Thus, the profile likelihood and its corresponding relative likelihood function of N, standardized to be one at the maximum of the likelihood function, are

$$L_{\max}(N) = \max_{p} L(N,p) = \left(\frac{t}{kN}\right)^t \left(1 - \frac{t}{kN}\right)^{kN-t} \left[\prod_{i=1}^{k} \binom{N}{x_i} \right], \tag{3}$$

$$R_{\max}(N) = \frac{L_{\max}(N)}{\max\limits_{N,p} L(N,p)} = \frac{L_{\max}(N)}{L_{\max}\left(\widehat{N}\right)}. \tag{4}$$

In general, the profile likelihood is a method devised to handle the estimation of parameters of interest, in the presence of nuisance parameters. To a considerable extent, the profile likelihood function can be thought of, and used, as if it were an authentic likelihood function. More details about the profile likelihood and most of its properties are described in Kalbfleisch (1985, Section 10.3), Pawitan (2001, Section 3.4), Sprott (2000, p. 66), Barndorff-Nielsen & Cox (1994, pp. 89-91), Serfling (2002, pp. 155-160), and Murphy & Van Der Vaart (2000), among others. A relevant strength of the profile likelihood function is that it can be used to study and visualize various aspects of a full likelihood function, such as for instance those associated with the flatness of the likelihood surface. The following example illustrates how the graph of the profile likelihood of N can reveal the flat nature of the likelihood function of (N, p).

3. EXAMPLE: BINOMIAL FLAT LIKELIHOOD

Olkin *et al.* (1981) considered the problem of estimating parameter N, based on independent observations $x_1, \cdots, x_k$ from a $Bin(N, p)$, with unknown parameters N and p. They showed that both maximum likelihood

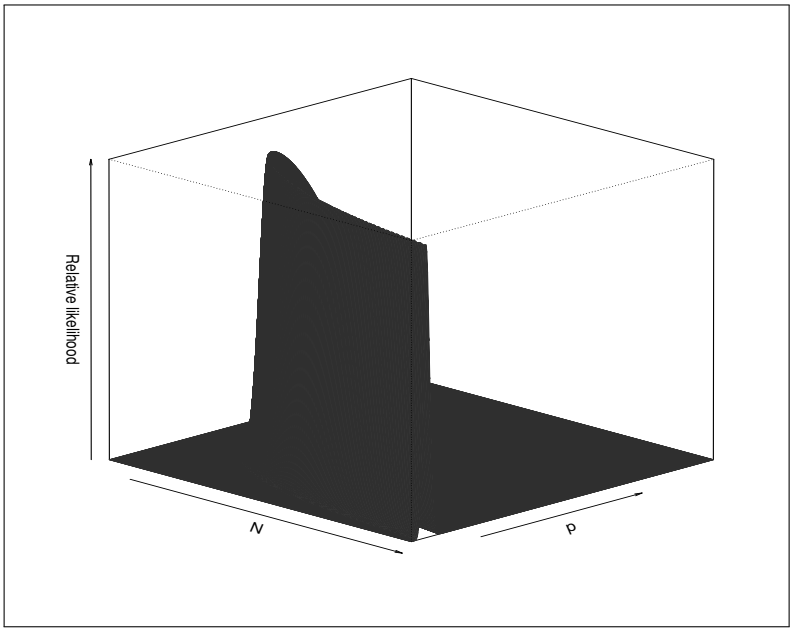


Figure 1: Binomial likelihood surface for the sample $\boldsymbol{x} = (16, 18, 22, 25, 27)$. Source: Elaborated by the author.

and moment estimators for parameter N can be extremely unstable, in the sense that switching an observed x_i to $x_i + 1$ can result in a massive change in the estimate of N. The authors also mentioned that the flatness of the likelihood causes the MLE (among others) to be extremely sensitive, even to slight perturbations on the data. This issue is illustrated in the following example.

Let us consider the sample $\boldsymbol{x} = (16, 18, 22, 25, 27)$ from Olkin *et al.* (1981, p. 638), which was simulated from a $Bin(75, 0.32)$. Figures 1 and 2 show the likelihood surface and the corresponding contour plot at two different levels: 0.98 and 0.1. A visual examination of these plots shows that the likelihood function has a ridge with a relatively flat top. In addition, we can see that parameters are inter-related, i.e. the contours are not parallel to the axes. Figure 3 shows the relative profile likelihood of N for both, the original sample and the perturbed one, where perturbation consist in modifying $x_{\max}$ by $x_{\max}$ +1 on the original sample. Both likelihoods have essentially the same planar form; however, the corresponding MLE's are different ($\hat{N} = 99$ and $\hat{N} = 190$). This is the MLE instability pointed out by Olkin *et al.* (1981) as a harder problem.

Sprott (2000, p. 11) assures that both the MLE and the observed information exhibit two features of the likelihood function, the former is a measure of its location relative to the parameter-axes while the latter is a measure of its curvature in a neighborhood of the MLE. Thus, when the likelihood function of a scalar parameter is smooth and symmetric with respect to the MLE, then the MLE and the observed information are usually the main features determining the shape of the likelihood function. It should be stressed that flat likelihoods can not be determined by these quantities without loss of information.
Intuitively, the sensitivity of the MLE depends on the shape of the likelihood function. If the likelihood

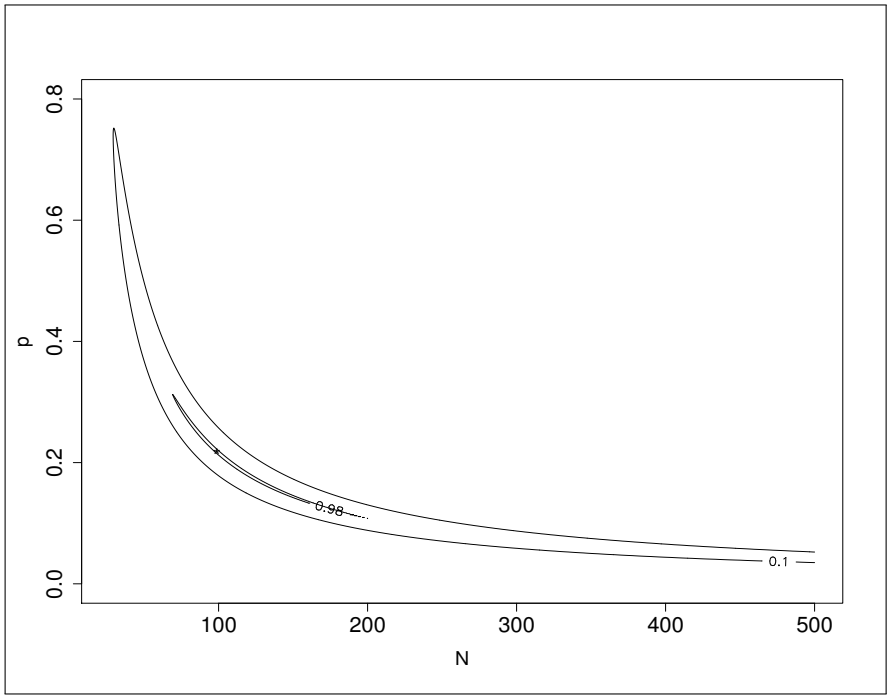


Figure 2: Contour plot of the relative likelihood surface given in Figure 1. The MLE is marked with an asterisk. Source: Elaborated by the author.

of N is highly curved around MLE, then the MLE is a stable estimate of N. On the other hand, if the likelihood is flat near MLE, then the MLE becomes highly unstable. But, perhaps more important than those facts is that a flat likelihood can produce likelihood-confidence intervals of N where the upper limit of can become infinite, and the question is: how often does a flat binomial likelihood occur? In the next section we investigate how frequently flat likelihoods of N are; this is done by exploring the limit of the relative profile likelihood of N, as N approaches infinity. Now, considering that the profile likelihood function for the parameter of interest is invariant under reparameterizations of the nuisance parameters, then the graph of the profile likelihood function of N do not change under reparameterizations of p. Here, we will analyze two cases, the one corresponding to binomial samples and also the case of a Poisson sample.

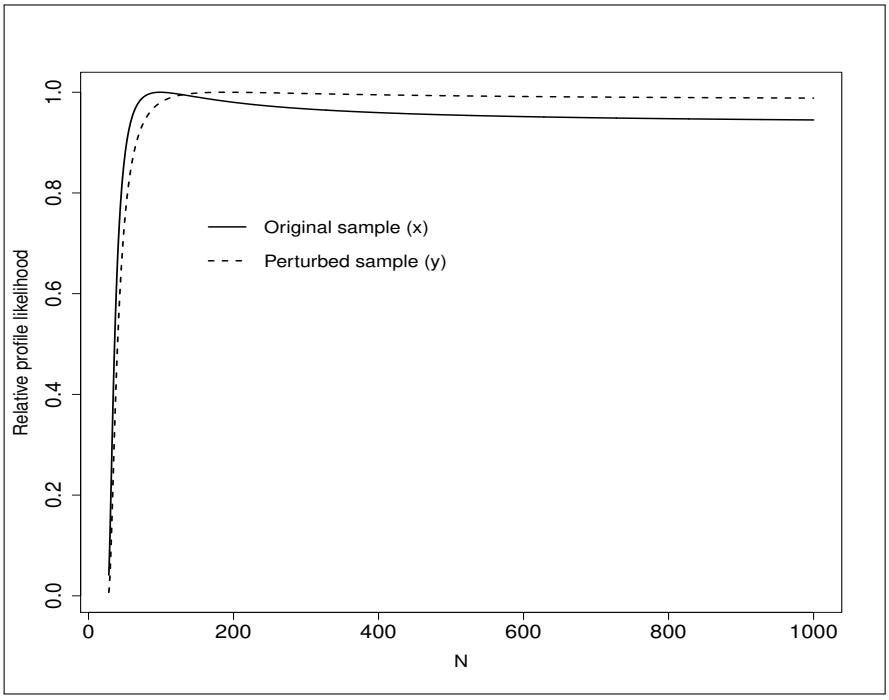


Figure 3: The relative profile likelihood function of N for the original and perturbed sample, $\boldsymbol{x} = (16, 18, 22, 25, 27)$ and $\boldsymbol{y} = (16, 18, 22, 25, 28)$ respectively. Source: Elaborated by the author.

4. FLATNESS OF THE RELATIVE PROFILE LIKELIHOOD OF N

Suppose that $R_{\max}(\infty)$ is the limit of the relative profile likelihood function of N, as N approaches infinity,

$$R_{\max}(\infty) = \lim_{N\to\infty} R_{\max}(N) = \lim_{N\to\infty} \frac{L_{\max}(N)}{L_{\max}\left(\widehat{N}\right)}. \tag{5}$$

If $R_{\max}(\infty)$ is smaller than one but is close to one, then for every real number $\varepsilon > 0$, there exists a natural number N_0 such that for all $N > N_0$, we have $R_{\max}(N) \in (R_{\max}(\infty) - \varepsilon, R_{\max}(\infty) + \varepsilon)$. Therefore, there exists a set of values of N that makes the observed sample as likely as the MLE of N does. That is, a flat likelihood is expected over a large range of N. Thus, it appears that the occurrence frequency of this type of behaviour can be used as an indicator for measuring the degree of flatness of the relative profile likelihood of N. In particular, when $R_{\max}(\infty) = 1$ the likelihood function is maximized at $\widehat{N} = \infty$ and $\widehat{p} = 0$. Thus, the binomial model would be mathematically rejected in favor of a Poisson model. Note that $R_{\max}(\infty) \approx 1$ can be interpreted as an embedded model problem. In fact, when you have a series of count data, defined in certain space-time limits, one could naturally think in a Poisson model, but it turns out that population size is a parameter of interest (N is assumed to be finite), in such a case the expected number of successes would be Np, but p is also an unknown parameter.

Olkin *et al.* (1981) showed that for $\widehat{\mu} \leq \widehat{\sigma}^2$, where $\widehat{\sigma}^2 = \sum_{i=1}^{k} (x_i - \bar{x})^2 / k$, the likelihood function is

maximized at $\widehat{N}=\infty$ and $\widehat{p}=0$. Thus, when $\widehat{\mu}\leq\widehat{\sigma}^2$, $R_{\max}(\infty)=1$. Otherwise, if $\widehat{\mu}>\widehat{\sigma}^2$, $L(N,p)$ is maximized at one or more positive (finite) values of N. Therefore we express $R_{\max}(\infty)$ as follows:

$$R_{\max}(\infty)=\begin{cases}\lim\limits_{N\to\infty}\dfrac{L_{\max}(N)}{\max\limits_{N}L_{\max}(N)} & ,\text{ if }\widehat{\mu}>\widehat{\sigma}^2,\\ 1 & ,\text{ if }\widehat{\mu}\leq\widehat{\sigma}^2.\end{cases}\tag{6}$$

For computational convenience, we can rewrite (12) as follows:

$$R_{\max}(\infty)=\begin{cases}\exp\left[l^{\infty}_{\max}-\max\limits_{N}l_{\max}(N)\right] & ,\text{ if }\widehat{\mu}>\widehat{\sigma}^2,\\ 1 & ,\text{ if }\widehat{\mu}\leq\widehat{\sigma}^2,\end{cases}\tag{7}$$

where

$$l^{\infty}_{\max}=t\ln(\bar{x})-k\bar{x}-\sum_{i=1}^{k}\ln\left[\Gamma(x_i+1)\right]\tag{8}$$

and

$$\begin{aligned}l_{\max}(N)=t\ln\left(\frac{\bar{x}}{N}\right)+(kN-t)\ln\left(1-\frac{\bar{x}}{N}\right)+k\ln\left[\Gamma(N+1)\right]\\ -\sum_{i=1}^{k}\left\{\ln\left[\Gamma(N-x_i+1)\right]+\ln\left[\Gamma(x_i+1)\right]\right\}.\end{aligned}\tag{9}$$

The R optimize function (from the built-in package stats) is used here to maximize (9) for N.

4.1. Simulation results: Binomial sample

In this section a simple experiment conducted to show the behavior of the relative profile likelihood function of N as N approaches infinity, $R_{\max}(\infty)$, is presented. Binomial samples of size 5, 10, 50, 100, 500, and 1000 were simulated 10000 times, considering parameter values $(N,p)=(100,21/100)$, $(200,21/200)$, $(500,21/500)$. Sample sizes for this simulation scenarios were selected considering small sample sizes, as well as very large ones, compared to the number of unknown parameters in the model. In the case of the selection of N and p values, we chose them in order to get $Np=21$. This selection was motivated by Example 1 presented in Carroll & Lombard (1985), that involves counting impala herds in Kruger National Park; there $\widehat{N}\widehat{p}=\bar{x}=21$. Now, to obtain the maximum of $l_{\max}(N)$ in (9), we set the optimize() parameters as lower=$x_{\max}$, upper=10000, and tol=0.0001.

Table 1 shows simulation results considering a sample classification according to whether $\widehat{\mu}>\widehat{\sigma}^2$ or $\widehat{\mu}\leq\widehat{\sigma}^2$. For these binomial scenarios, the fraction of cases where $\widehat{\mu}\leq\widehat{\sigma}^2$ exhibited an effect due to sample size. It can be observed that firstly this percentage increases and then decreases. In all these cases, $R_{\max}(\infty)=1$. On the other hand, violin plots displayed in Figure 4 show the shape of the frequency distribution of $R_{\max}(\infty)$ for each simulation scenario in which $\widehat{\mu}>\widehat{\sigma}^2$. The examination of Figure 4 reveals that the degree of flatness of

Table 1: Classification (in percentage) of 10000 simulated samples of a binomial random variable.

$Bin(N,p)$	k	$\widehat{\mu} > \widehat{\sigma}^2$	$\widehat{\mu} \leq \widehat{\sigma}^2$
$(100, 21/100)$	5	83.36	16.64
	10	81.94	18.06
	50	91.39	8.61
	100	96.74	3.26
	500	99.98	0.02
	1000	100	0
$(200, 21/200)$	5	76.89	23.11
	10	73.19	26.81
	50	76.89	23.11
	100	82.05	17.95
	500	96.81	3.19
	1000	99.65	0.35
$(500, 21/500)$	5	73.06	26.94
	10	67.73	32.27
	50	64.74	35.26
	100	66.34	33.66
	500	77.40	22.60
	1000	84.15	15.85

the relative profile likelihood of N can be high and that flat likelihoods occur very frequently. In particular, for a fixed sample size, flat likelihoods occur more frequently when Np, the expected value of the binomial random variable, is located far from N (the extreme right of the support of the binomial distribution or 100th percentile).

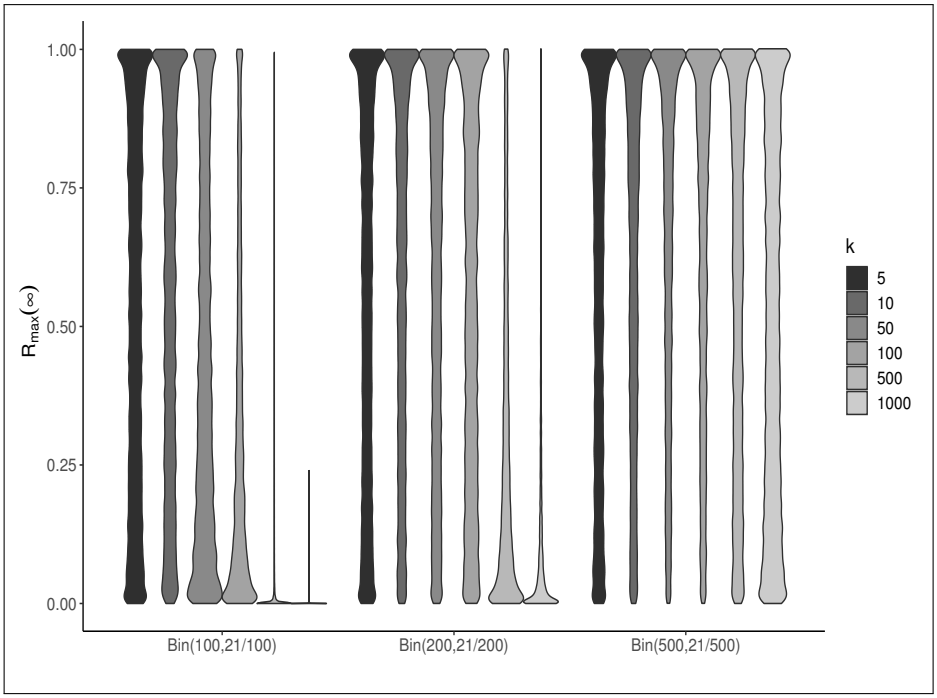


Figure 4: Violin plots for binomial simulation scenarios where $\widehat{\mu} > \widehat{\sigma}^2$. Source: Elaborated by the author.

Table 2: Classification (in percentage) of 10000 simulated samples of a Poisson random variable.

$Pois(\lambda)$	k	$\widehat{\mu} > \widehat{\sigma}^2$	$\widehat{\mu} \leq \widehat{\sigma}^2$
21	5	28.14	71.86
	10	35.45	64.55
	50	43.58	56.42
	100	45.55	54.45
	500	48.14	51.86
	1000	49.54	50.46
210	5	28.97	71.03
	10	34.65	65.35
	50	43.44	56.56
	100	45.05	54.95
	500	47.89	52.11
	1000	48.64	51.36
2100	5	28.62	71.38
	10	35.35	64.65
	50	43.63	56.37
	100	45.55	54.45
	500	48.57	51.43
	1000	48.51	51.49

4.2. Simulation results: Poisson sample

In this section we explore the flatness of the relative profile likelihood of N, when data from a Poisson model are modeled with a binomial model. Here, this case is not considered as a problem of incorrect specification of the underlying model (binomial distribution), because the Poisson model is embedded in the binomial distribution; that is, the Poisson model is obtained when $N \to \infty$, $p \to 0$, and the mean value $Np = \lambda$ remains constant.

We simulate data from a Poisson model and we again show the behaviour of $R_{\max}(\infty)$. Samples of size 5, 10, 50, 100, 500, and 1000 were simulated 10000 times from a Poisson distribution with a mean of $\lambda = 21, 210$, and 2100. We denote the Poisson distribution as $Pois(\lambda)$. To obtain the maximum of $l_{\max}(N)$ in (9), we set the optimize() parameters as lower=$x_{\max}$, upper=1000000, and tol=0.0001.

Table 2 shows simulation results considering a sample classification according to whether $\widehat{\mu} > \widehat{\sigma}^2$ or $\widehat{\mu} \leq \widehat{\sigma}^2$. For these Poisson scenarios, the percentage of cases where $\widehat{\mu} \leq \widehat{\sigma}^2$ exhibited a strong effect due to sample size. In all these cases, a decreasing trend is obtained when increasing sample size. In the case of a sample of size 1000, these proportions are close to 50 percent. On the other hand, violin plots displayed in Figure 5 show the shape of the frequency distribution of $R_{\max}(\infty)$ for each simulation scenario in which $\widehat{\mu} > \widehat{\sigma}^2$. This figure also shows that flat likelihoods frequently occur when data from a Poisson model, embedded in the binomial distribution, are modeled by a binomial model and this happened for all the sample sizes analyzed in this study. In fact, Figure 5 shows that in all Poisson scenarios considered here, $R_{\max}(\infty)$ values are concentrated close to one, when the sample is large.

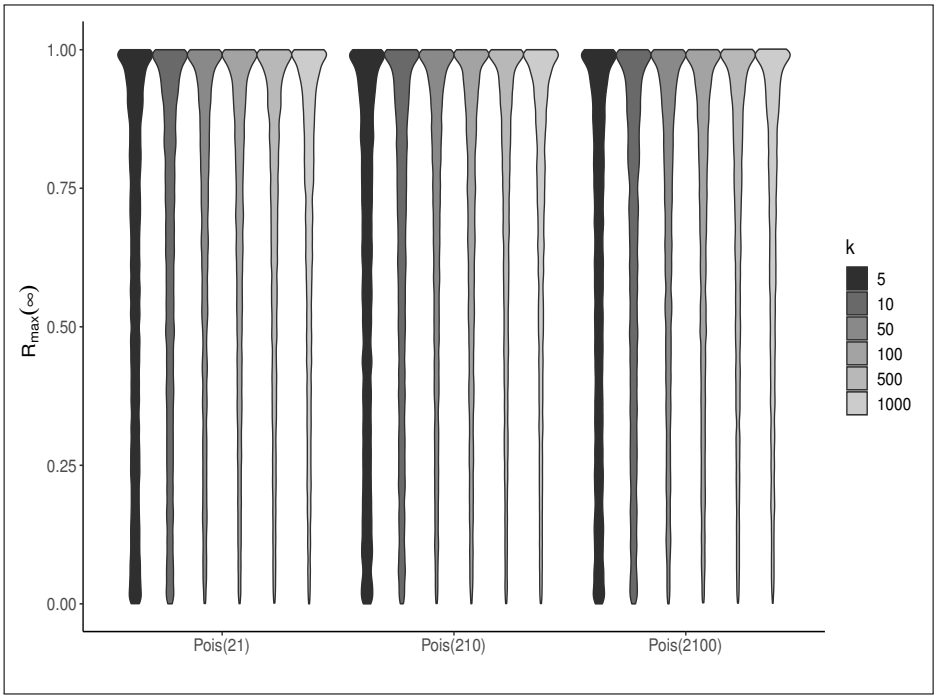


Figure 5: Violin plots for Poisson simulation scenarios where $\widehat{\mu} > \widehat{\sigma}^2$. Source: Elaborated by the author.

5. ESTIMATION OF MODEL PARAMETERS

When a binomial experimental design yields a sequence of independent success counts, with no further information and a binomial model is assumed, difficulties involved in estimating parameters N and p are not only related to the limited amount of information provided by the data, to make possible appropriate and useful inferences, but also to the approximate nature of the models under consideration (the family of binomial distributions and the Poisson embedded model in this family), used to assign probabilities. To properly estimate (N, p) parameters, not only a large sample size (relative to the number of parameters in the model) is required, but also a sample containing sufficient information about the threshold parameter N (the 100-quantile or the "right end" of the support of the binomial distribution) and the underlying relationship between p and N.

Fisher (1941) mentioned that "With a sufficiently large sample N is necessarily one less than the number of frequency classes, and is thus determined without reference to the actual frequencies". However, a considerable increase in sample size is a mathematical solution to the problem of estimating parameter N of the binomial model, but not a practical one. Therefore, in the presence of small samples, a feasible solution to the problem of estimating parameter N involves the incorporation of additional information into the inference process. Here, Bayesian methods and the integrated likelihood provide a formal framework for statistical inference about N. However, it is important to emphasize that in the presence of a flat binomial likelihood, seemingly reasonable inferences about N, obtained under these approaches, are strongly determined by prior

information (Kahn, 1987; Aitkin & Stasinopoulos, 1989).

Since the problem of flat binomial likelihoods is imminent, a strategy that might work to satisfactorily estimate (N,p) is modifying the observation protocol. However, this can surely lead to probability model changes. It is possible that parameters (N,p) will be still present in the new model, as will be shown in the following example; however, it is more likely for them to disappear or change status, as in the case of hierarchical models involving covariates. In any case, it is necessary to plot the likelihood function to assess the presence of flat likelihoods. This is also illustrated in the following example.

Example: Removal sampling

This example illustrates a model that can be very effective for estimating parameter N, but not necessarily precludes the possibility of getting flat likelihoods. Here, we consider a removal scheme used when estimating animal populations. In this type of removal sampling, population size estimation is based on the results of a series of trappings, and trapped animals are removed from the population. Population size N is the parameter of interest and the binomial probability of capture during a single trapping p is a nuisance parameter.

Suppose that the probability of capturing x_i animals during the ith trapping, given that y_i animals had previously been captured, is

$$P(x_i \mid y_i;N,p) = \binom{N-y_i}{x_i} p^{x_i} (1-p)^{N-y_i-x_i}.$$

Thus, the joint probability of sample $x_1,\cdots,x_k$ is

$$P(x_1,\cdots,x_k;N,p) = \left(\frac{r_k!}{x_1!\cdots x_k!}\right)\binom{N}{r_k} p^{r_k}(1-p)^{kN-S},$$

where $r_i = \sum_{j=1}^{i} x_j$ and $S = \sum_{i=1}^{k} r_i$. Therefore, the likelihood function of (N,p) based on the sample of catches actually observed in k trappings is

$$L(N,p) = C\binom{N}{r_k} p^{r_k}(1-p)^{kN-S}, \tag{10}$$

for $0 \le p \le 1$ and $N \ge r_k$, where $C = r_k!/(x_1!\cdots x_k!)$.
In this case, the restricted maximum likelihood estimate of p is $\widehat{p}(N) = r_k/(r_k+kN-S)$. Thus, the profile likelihood function of N is

$$L_{\max}(N) = C\binom{N}{r_k}\left(\frac{r_k}{r_k+kN-S}\right)^{r_k}\left(\frac{kN-S}{r_k+kN-S}\right)^{kN-S}. \tag{11}$$

An example of this kind of removal experiment, appertaining to rat populations, gave $k=18$, $r_k=394$ and $S=4544$; see Moran (1951). A contour plot of the relative likelihood of (N,p), corresponding to (10), is shown in Figure 6. The surface is well behaved (it is no flat) with a unique maximum defining the MLE. In this case, profile likelihood of N is asymmetric with respect to $\widehat{N}$, MLE of the parameter N, but it is not flat.

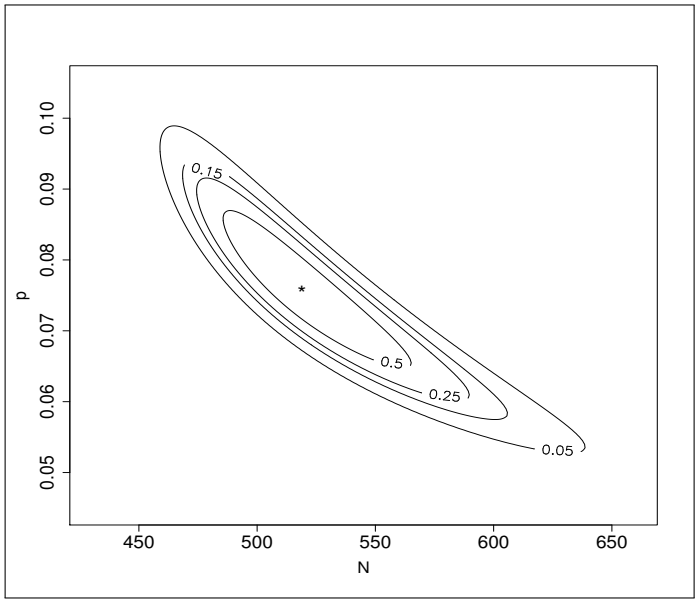


Figure 6: Likelihood contours as a function of N and p for $k = 18$, $r_k = 394$ and $S = 4544$. The MLE is marked with an asterisk.
Source: Elaborated by the author.

However, the shape of this function can dramatically change and become flat when the k successive trappings decrease. Figure 7 shows the relative profile likelihood function of N, corresponding to (11), for the complete and incomplete samples summarized in Table 3. From this figure it can be observed that the MLE decreases when k decreases. But, more important than this is perhaps the fact that small sample sizes (incomplete samples 2 and 3) yield asymmetric likelihoods. Furthermore, the likelihood function corresponding to incomplete sample 3 is fairly flat over a large range of N values. Table 4 shows the MLE's of N and the 5 % likelihoods intervals.

In summary, it is important to analyze the shape of the profile likelihood function. Given an observed sample, when profile likelihood function is steeply curved, the parameter estimate is better constrained, in comparison with a nearly flat curve, since this suggests that many hypotheses would be almost equally likely. On the other hand, when we already have a model under study, it is possible to perform sample simulations, in order to explore the shape of the profile likelihood of the parameter of interest, as it is shown in Section 4, where it was possible to identify and characterize the occurrence of an identifiability parameter problem; being also able to determine adequate sample sizes that could move us away from the possibility of dealing with samples that yield to flat likelihoods.

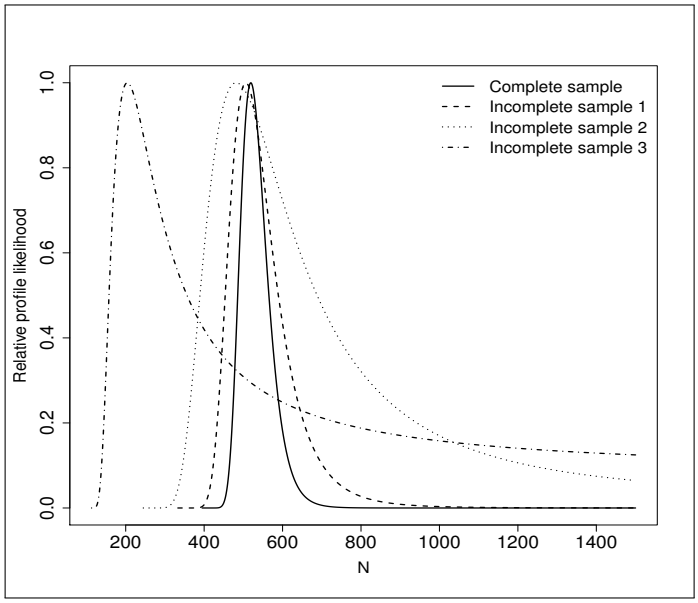


Figure 7: Relative profile likelihood functions of N for the complete sample ($k = 18$, $r_k = 394$, $S = 4544$) and the incomplete samples presented in Table 3. Source: Elaborated by the author.

Table 3: Statistical summary of incomplete samples, corresponding to the removal experiment performed in rat population example.

Sample	k	r_k	S
Incomplete 1	13	332	2691
Incomplete 2	8	244	1206
Incomplete 3	3	112	242

Table 4: MLE and 5% likelihoods intervals for the parameter N.

Sample	Lower limit	MLE	Upper limit
Complete	458	519	639
Incomplete 1	415	506	753
Incomplete 2	334	482	1713
Incomplete 3	134	204	> 10000

6. CONCLUSIONS

The cases discussed here, show that flat likelihoods warn about the presence of two possible problems that must be considered and analyzed before applying any inferential method. The first one is associated not only with the limited amount and the quality of experimental data in order to make appropriate and useful inferences about the threshold parameter, but also with the underlying relationship between the parameters of the model. This problem occurs if the expected value of the binomial random variable is located far away from the threshold parameter, which is the right end of the distribution support, that is, the 100th percentile. The second problem is related to embedded models. In our case, the Poisson model, which only depends on one

parameter, is embedded in the binomial distribution family. The problem arises when the best Poisson model and the best binomial model are indistinguishable; that occurs when both models assign similar probabilities to the observed sample. An interesting fact is that this behavior persists despite of increasing the sample size.

To summarize, the purpose of the preceding is to reemphasize about the importance of taking into account the shape of the likelihood in statistical inference. Since the likelihood function is widely used in many statistical inferential methods, it is important not only to recognize factors that can cause flat likelihoods but also to study the degree of severity of this flattening, all these in order to apply or develop *ad hoc* statistical and computational methods that allow us to get valid inferences.

References

Aitkin, M. & Stasinopoulos, M. (1989). Likelihood analysis of a binomial sample size problem. *Contributions to Probability and Statistics (pp. 399-411)*. Springer. New York.

Barndorff-Nielsen, O. E. & Cox, D. R. (1994). Inference and asymptotics. Chapman & Hall/CRC. Boca Raton.

Berger, J. O., Liseo, B.& Wolpert, R. L. (1999). Integrated likelihood methods for eliminating nuisance parameters. *Statistical Science*, 14(**1**), 1-28.

Breusch, T. S., Robertson, J. C. & Welsh, A. H. (1997). The emperor's new clothes: a critique of the multivariate t regression model. *Statistica Neerlandica*, 51(**3**), 269-286.

Carroll, R. J. & Lombard, F. (1985). A note on *N* estimators for the binomial distribution. *Journal of the American Statistical Association*, 80(**390**), 423-426.

Casella, G. (1986). Stabilizing binomial *n* estimators. *Journal of the American Statistical Association*, 81(**393**), 172-175.

Catchpole, E. A.& Morgan, B. J. (1997). Detecting parameter redundancy. *Biometrika*, 84(**1**), 187-196.

Cheng, R. C. H. & Iles, T. C. (1990). Embedded models in three-parameter distributions and their estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, 52(**1**), 135-149.

Cole, S. R., Chu, H. & Greenland, S. (2013). Maximum likelihood, profile likelihood, and penalized likelihood: a primer. *American Journal of Epidemiology*, 179(**2**), 252-260.

DasGupta, A. & Rubin, H. (2005). Estimation of binomial parameters when both *n*, *p* are unknown. *Journal of Statistical Planning and Inference*, 130(**1-2**), 391-404.

Draper, N.; Guttman, I. (1971), Bayesian estimation of the binomial parameter. *Technometrics*, **13**(3), 667-673.

El Adlouni, S., Ouarda, T. B., Zhang, X., Roy, R. & Bobée, B. (2007). Generalized maximum likelihood estimators for the nonstationary generalized extreme value model. *Water Resources Research*, 43(**3**), W03410.

Farcomeni, A. & Tardella, L. (2012). Identifiability and inferential issues in capture-recapture experiments with heterogeneous detection probabilities. *Electronic Journal of Statistics*, 6, 2602-2626.

Fisher, R. A. (1941). The negative binomial distribution. *Annals of Eugenics*, 11(**1**), 182-187.

Frery, A. C., Cribari-Neto, F.& De Souza, M. O. (2004). Analysis of minute features in speckled imagery with maximum likelihood estimation. *EURASIP Journal on Advances in Signal Processing*, 2004(**16**), 2476-2491.

Ghosh, M., Datta, G. S., Kim, D. & Sweeting, T. J. (2006). Likelihood-based inference for the ratios of regression coefficients in linear models. *Annals of the Institute of Statistical Mathematics*, 58(**3**), 457-473.

Gupta, A. K., Nguyen, T. T. & Wang, Y. (1999). On maximum likelihood estimation of the binomial parameter *n*. *Canadian Journal of Statistics*, 27(**3**), 599-606.

Hall, P. (1994). On the erratic behavior of estimators of *N* in the binomial *N*, *p* distribution. *Journal of the American Statistical Association*, 89(**425**), 344-352.

Harter, H. L. & Moore, A. H. (1966). Local-maximum-likelihood estimation of the parameters of three-parameter lognormal populations from complete and censored samples. *Journal of the American Statistical Association*, 61(**315**), 842-851.

Kahn, W. D. (1987). A cautionary note for Bayesian estimation of the binomial parameter *n*. *The American Statistician*, 41(**1**), 38-40.

Kalbfleisch, J. G. (1985). Probability and Statistical Inference, Vol. 2. Springer-Verlag. New York.

Kreutz, C., Raue, A., Kaschek, D. & Timmer, J. (2013). Profile likelihood in systems biology. *The FEBS Journal*, 280(**11**), 2564-2571.

Li, R. & Sudjianto, A. (2005). Analysis of computer experiments using penalized likelihood in Gaussian Kriging models. *Technometrics*, 47(**2**), 111-120.

Lima, V. M. & Cribari-Neto, F. (2019). Penalized maximum likelihood estimation in the modified extended Weibull distribution. *Communications in Statistics-Simulation and Computation*, 48(**2**), 334-349.

Lindsey, J. K. (1996). Parametric statistical inference. Oxford University Press. New York.

Liu, S., Wu, H. & Meeker, W. Q. (2015). Understanding and addressing the unbounded "likelihood" problem. *The American Statistician*, 69(**3**), 191-200.

Martins, E. S. & Stedinger, J. R. (2000). Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resources Research*, 36(**3**), 737-744.

Martins, E. S. & Stedinger, J. R. (2001). Generalized maximum likelihood Pareto-Poisson estimators for partial duration series. *Water Resources Research*, 37(**10**), 2551-2557.

Montoya, J. A., Díaz-Francés, E. & Sprott, D. A. (2009). On a criticism of the profile likelihood function. *Statistical Papers*, 50(**1**), 195-202.

Moran, P. A. P. (1951). A mathematical theory of animal trapping. *Biometrika*, 38(**3-4**), 307-311.

Murphy, S. A. & Van Der Vaart, A. W. (2000). On profile likelihood. *Journal of the American Statistical Association*, 95(**450**), 449-465.

Olkin, I., Petkau, A. J. & Zidek, J. V. (1981). A comparison of *n* estimators for the binomial distribution. *Journal of the American Statistical Association*, 76(**375**), 637-642.

Pawitan, Y. (2001). In All Likelihood: Statistical Modelling and Inference Using Likelihood. Oxford University Press. New York.

Pewsey, A. (2000). Problems of inference for Azzalini's skewnormal distribution. *Journal of Applied Statistics*, 27(**7**), 859-870.

Raftery, A. E. (1988). Inference for the binomial *N* parameter: A hierarchical Bayes approach. *Biometrika*, 75(**2**), 223-228.

Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U. & Timmer, J. (2009). Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, 25(**15**), 1923-1929.

Serfling, R. J. (2002). Approximation Theorems of Mathematical Statistics. John Wiley & Sons. New York.

Sprott, D. A. (2000). Statistical inference in science. Springer-Verlag. New York.

Sundberg, R. (2010). Flat and multimodal likelihoods and model lack of fit in curved exponential families. *Scandinavian Journal of Statistics*, 37(**4**), 632-643.

Tsionas, E. G. (2001). Likelihood and Posterior Shapes in Johnson's System. *Sankhyā: The Indian Journal of Statistics, Series B*, 63(**1**), 3-9.

Tumlinson, S. E. (2015). On the non-existence of maximum likelihood estimates for the extended exponential power distribution and its generalizations. *Statistics & Probability Letters*, 107, 111-114.

Revista de la Facultad de Ciencias Universidad Nacional de Colombia, Sede Medellín
V 11 N°2 julio-diciembre de 2022 • ISSN-e 2357-5749 • Research paper • Pages 25 to 38
DOI: https://doi.org/10.15446/rev.fac.cienc.v11n2.97782

UNDERSTANDING A LIKELIHOOD FLAT PROBLEM: INFERENCES ON THE RATIO OF REGRESSION COEFFICIENTS IN LINEAR MODELS[a]

ENTENDIENDO UN PROBLEMA DE VEROSIMILITUD PLANA: INFERENCIAS SOBRE EL COCIENTE DE COEFICIENTES DE REGRESIÓN EN MODELOS LINEALES

JORGE ESPÍNDOLA ZEPEDA[b], JOSÉ A. MONTOYA[c*]

Recibido 13-08-2021, aceptado 6-04-2022, versión final 13-06-2022
Research Paper

ABSTRACT: In this paper, we analyze a flat likelihood function shape that arises when performing inferences on the ratio of two regression coefficients in a linear regression model, parameter of interest in various applications. Due to this shape, infinite length likelihood-confidence intervals can be obtained. In the cases discussed here these likelihood-confidence intervals are related to the nested models problem, which is analyzed in detail through three illustrative simulated cases. It is essential to understand the shapes of the likelihood function in order to legitimately criticize likelihood inferences. This is of particular importance since the likelihood function is a key ingredient used in many inference methods.
KEYWORDS: Shape of the likelihood function; nested models; linear regression model; profile likelihood function.

RESUMEN: En este artículo se analiza una forma plana de la función de verosimilitud que surge cuando se realizan inferencias sobre la razón de dos coeficientes de regresión en un modelo lineal, parámetro de interés en diversas aplicaciones. Debido a esta forma pueden obtenerse intervalos de verosimilitud-confianza de longitud infinita. En los casos que se discuten aquí, estos intervalos de verosimilitud-confianza están relacionados con el problema de modelos anidados, que es analizado a detalle a través de la simulación de tres casos ilustrativos. Es fundamental comprender las formas de la función de verosimilitud para criticar de manera legítima las inferencias por verosimilitud. Esto es de particular importancia ya que la función de verosimilitud es un ingrediente clave utilizado en muchos métodos inferenciales.
PALABRAS CLAVE: Forma de la función de verosimilitud; modelos anidados; modelo de regresión lineal; función de verosimilitud perfil.

[a]Espíndola Zepeda, J. & Montoya, J. A. (2022). Understanding a Likelihood Flat Problem: Inferences On The Ratio of Regression Coefficients in Linear Models. *Rev. Fac. Cienc.*, 11 (**2**), 25–38. DOI: https://doi.org/10.15446/rev.fac.cienc.v11n2.97782
[b]M.Sc. in Mathematics. PhD student. Department of Mathematics. University of Sonora, México.
[c]PhD in Probability and Statistics. Statistics Professor. Department of Mathematics. University of Sonora, México.
[*]Corresponding author: arturo.montoya@unison.mx

1. INTRODUCTION

Academic literature has consistently illustrated that statistical methods based on the likelihood function can lead to misleading, strange, unstable, preposterous or ridiculous estimates, (Harter & Moore, 1966; Breusch *et al.*, 1997; Berger *et al.*, 1999; Martins & Stedinger, 2000; Pewsey, 2000; Martins & Stedinger, 2001; El Adlouni *et al.*, 2007; Tumlinson, 2015). The shape of the likelihood function is often cited as an underlying cause of these strange or unintuitive results. The criticisms of maximum likelihood estimates, associated with strange likelihood functions that grow rapidly to infinite and caused by the use of density functions that have singularities, are invalid. This issue is well known; see Montoya *et al.* (2009), Liu *et al.* (2015), and the references cited therein. However, there are also criticisms of the maximum likelihood estimation that occur quite often and that arise when the shape of the likelihood function becomes flat. In fact, flat likelihoods are used to promote integrated likelihoods (Berger *et al.*, 1999; Ghosh *et al.*, 2006), Bayesian posteriors (Tsionas, 2001), penalized methods (Li & Sudjianto, 2005; Cole *et al.*, 2013; Lima & Cribari-Neto, 2019) and also new optimization methods (Frery *et al.*, 2004).

Some authors have found that flat likelihoods can be related with an overparameterization of the model (Catchpole & Morgan, 1997), inappropriate reparametrizations (Farcomeni, & Tardella, 2012), embedded models (Cheng & Iles, 1990), limited amount and quality of experimental data (Raue *et al.*, 2009; Kreutz *et al.*, 2013), and even a poor model adjustment to the data, when the sample size is large (Sundberg, 2010). Although these studies are based on relevant statistical models (exponential family of distributions, three-parameters distributions, capture-recapture models and dynamical models), and all helped to clarify what is behind the problem of these flat likelihoods, additional work is needed to enhance our understanding.

This article is focused on the linear regression model. We consider the profile likelihood function to analyze the ratio of two regression coefficients, our parameter of interest, which naturally appears in many statistical applications that usually arise in biopharmaceutical and economic research, among others. This parameter can mean: a measure of the relative potency of a test drug versus a reference drug (Fay *et al.*, 2009); the location of a turning point in a quadratic specification where the marginal impact of a regressor changes sign (Rosenblad, 2020); the interpretation of the marginal effect of one regressor when interacted with another (Hirschberg & Lye, 2010), among many others meanings.

Profile likelihood function for the ratio of two regression coefficients was given by Ghosh *et al.* (2003) and reviewed again in Ghosh *et al.* (2006), where they showed that inferences about this parameter, based on profile likelihood-confidence intervals, may result the entire real line. This fact occurs because we are dealing with an essentially flat likelihood function that has a horizontal asymptote, located at some distance above the global minimum.

In this paper, we analyze conditions under which a confidence interval for the interest parameter, based on

the profile likelihood function, has an infinite length and even becomes the entire real line. These conditions are related to the nested models problem. In the cases discussed here, infinite length confidence intervals arise when a hypothesis test, under an F statistic, shows that full and reduced regression models fit the data equally well. When this situation is considered, the shape of the likelihood function is very informative and explains this model criticism problem.

The paper is structured as follows. Section 2 includes some results used to perform hypothesis testing about the coefficients of the classical linear regression model. In addition, some connections between these tests and a pivotal quantity (or pivot) for the ratio of two regression coefficients, our parameter of interest, are presented. Section 3 shows how the shape of the profile likelihood function of the parameter of interest is determined by the pivotal quantity. Additionally, connections between the shape of the likelihood and hypotheses tests about the coefficients of the linear regression model are shown. In Section 4 we include three illustrative simulated cases to exemplify theoretical results shown in previous sections. Finally, some concluding remarks are presented in Section 5.

2. F-TESTS AND NESTED REGRESSION MODEL

Cox (2006, p. 3) described that a statistical inference process can be basically divided into two stages: choice of the family of models and model criticism. The family that is chosen in the first stage is often fully specified, except for a limited number of unknown parameters. In our case, we address flat likelihoods and estimation issues, within the multiple linear regression framework. For simplicity, and without loss of generality, we have selected a linear regression model with two predictor variables. In the second stage, we must ask ourselves whether the selected model is consistent with the data, if some changes on it may be needed, or even if the whole model should be modified. A particular problem related to this stage arises when comparing two nested models. Two models are called nested if one contains all the predictors of the other one and also some additional predictors. In linear regression, this is known as reduced models (Searle & Gruber, 1971).

Consider the regression model

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + e_i, \tag{1}$$

for $i = 1,\ldots,n$, $n > 2$, where e_i's are independent random errors, normally distributed with mean 0 and common variance σ^2. Here, regression coefficients, β_1 and β_2, are considered real unknown parameters. Model presented in (1) can be written in matrix form

$$y = X\beta + e, \tag{2}$$

where $y^T = (y_1,\ldots,y_n)$, $X^T = (x_1,\ldots,x_n)$, $x_i = (x_{i1},x_{i2})^T$, $i = 1,\ldots,n$, $\beta^T = (\beta_1,\beta_2)$, $e^T = (e_1,\ldots,e_n)$, and rank of X equals 2.

The nested modelling problem associated with the aim of this article occurs when $y_i = e_i$ or $y_i = \beta_1 x_{i1} + e_i$ fits the data evenly well. The typical statistical hypotheses, related to these nested problems are $H_{01} : \beta_1 = \beta_2 = 0$ and $H_{02} : \beta_2 = 0$, respectively. For the nested model $y_i = e_i$, called the white noise problem, we test the hypothesis $H_{01} : \beta_1 = \beta_2 = 0$, using the standard test statistic

$$F_1 = \frac{n\hat{\beta}^T C\hat{\beta}}{2MSE}, \tag{3}$$

where $C = (c_{ij}) = n^{-1}X^T X$ is a positive matrix, $i, j = 1, 2$, $\hat{\beta}^T = (\hat{\beta}_1, \hat{\beta}_2) = (nC)^{-1}X^T y$, and $MSE = (y - X\hat{\beta})^T (y - X\hat{\beta})/(n-2)$. It is well known that F_1 has an F distribution with 2 and $n-2$ degrees of freedom. For $0 < \alpha < 1$, the hypothesis $H_{01} : \beta_1 = \beta_2 = 0$ is rejected at α significance level when F_1 is greater than the $(1-\alpha)$ quantile of the F distribution with 2 and $n-2$ degrees freedom, quantile denoted here as $F_{2,n-2,1-\alpha}$.

On the other hand, for the nested model $y_i = \beta_1 x_{i1} + e_i$, we can test the null hypothesis $H_{02} : \beta_2 = 0$ using the typical test statistic

$$F_2 = \frac{n|C|\hat{\beta}_2^2/c_{11}}{MSE}. \tag{4}$$

This statistic has an F distribution with 1 and $n-2$ degrees of freedom. Thus, when F_2 is greater than $F_{1,n-2,1-\alpha}$, the $(1-\alpha)$ quantile of the F distribution with 1 and $n-2$ degrees freedom, the hypothesis $H_{02} : \beta_2 = 0$ is rejected at α significance level.

Let us now consider the problem of making inferences on $\theta_1 = \beta_1/\beta_2$, the ratio of the regression coefficients β_1 and β_2; when $\beta_1 = \beta_2 = 0$ or $\beta_2 = 0$, it can cause indeterminacy and lead to some paradoxes. For such reason, inferences about $\beta^T = (\beta_1, \beta_2)$ play an important role in the determination of plausible values of θ_1, and this must be fully understood. Particularly, those inferences related to hypotheses H_{01} and H_{02}.
Ghosh *et al.* (2006) showed that

$$F_1(\theta_1) = \frac{n|C|(\hat{\beta}_2\theta_1 - \hat{\beta}_1)^2/Q(\theta_1)}{MSE} \tag{5}$$

has an F distribution with 1 and $n-2$ degrees of freedom, where $Q(\theta_1) = c_{11}\theta_1^2 + 2c_{12}\theta_1 + c_{22}$, being $F_1(\theta_1)$ a pivotal quantity (Sprott, 2008, p. 63) related to the hypotheses tests H_{01} and H_{02}. They also found the following relationship between $F_1(\theta_1)$ and the statistic F_1:

$$\sup_{\theta_1} F_1(\theta_1) = 2F_1. \tag{6}$$

Now, considering (3) and (6), the hypothesis $H_{01} : \beta_1 = \beta_2 = 0$ is rejected at α significance level if

$$\sup_{\theta_1} F_1(\theta_1) > 2F_{2,n-2,1-\alpha}. \tag{7}$$

On the other hand, there is also a relationship between $F_1(\theta_1)$ and F_2 that is as follows:

$$\lim_{|\theta_1|\to\infty} F_1(\theta_1) = F_2, \tag{8}$$

since $\lim_{|\theta_1|\to\infty}(\hat{\beta}_2\theta_1-\hat{\beta}_1)^2/Q(\theta_1)=\hat{\beta}_2^2/c_{11}$. Hence, observing (4) and (8), the hypothesis $H_{02}:\beta_2=0$ is rejected at α significance level if

$$\lim_{|\theta_1|\to\infty} F_1(\theta_1) > F_{1,n-2,1-\alpha}. \tag{9}$$

The profile likelihood function of θ_1 is presented in the next section, where it is used to get a closed mathematical expression for the likelihood ratio test statistic, associated to the pivotal quantity $F_1(\theta_1)$. In addition, we show some connections between relative profile likelihood function and confidence intervals for θ_1. But, even more important than that is to show how profile likelihood function shape is associated with the nature of the nested models problem. Thus, the characteristics of this problem are naturally inherited by $\theta_1=\beta_1/\beta_2$ and their confidence intervals.

3. INFERENCE ABOUT PARAMETER θ_1

The likelihood approach is one of the most popular techniques for deriving statistics, as explained in Casella & Berger (2002, p. 315), and this is performed via the likelihood function. The resulting statistics play an important role in statistical inference since, besides the fact that they can be used to obtain point estimates, it is possible to construct confidence intervals or perform hypothesis tests based on them. In our case, the likelihood function for model (2) is given by

$$L(\beta_1,\beta_2,\sigma)\propto\sigma^{-n}\exp\left[-\frac{1}{2\sigma^2}(y-X\beta)^T(y-X\beta)\right],$$

and since $(y-X\beta)^T(y-X\beta)=(y-X\hat{\beta}+X\hat{\beta}-X\beta)^T(y-X\hat{\beta}+X\hat{\beta}-X\beta)$ and $(X\hat{\beta}-X\beta)^T(y-X\hat{\beta})=0$, the likelihood function can be also written as

$$L(\beta_1,\beta_2,\sigma)\propto\sigma^{-n}\exp\left\{-\frac{1}{2\sigma^2}\left[SSE+(\beta-\hat{\beta})^TX^TX(\beta-\hat{\beta})\right]\right\}, \tag{10}$$

where $SSE=(y-X\hat{\beta})^T(y-X\hat{\beta})$.

To make inferences regarding a parameter of interest, in the presence of nuisance parameters, the likelihood approach relies on the profile likelihood function (Sprott, 2008, p. 66). This is obtained by fixing values for the parameter of interest and maximizing with respect to nuisance parameters. Here, $\theta_1=\beta_1/\beta_2$ is the parameter of interest and β_2 and σ are considered nuisance parameters. Hence, in order to obtain the profile likelihood function for θ_1, the likelihood function (10) is reparametrized in terms of $(\theta_1,\beta_2,\sigma)$.

$$L(\theta_1,\beta_2,\sigma)\propto\sigma^{-n}\exp\left\{-\frac{1}{2\sigma^2}\left[SSE+n(\beta_2\phi-\hat{\beta})^TC(\beta_2\phi-\hat{\beta})\right]\right\}, \tag{11}$$

where $\phi^T=(\theta_1,1)$. Then, maximizing with respect to σ and β_2, the resulting profile likelihood function of θ_1 is

$$L_P(\theta_1)\propto\left[SSE+n|C|(\hat{\beta}_2\theta_1-\hat{\beta}_1)^2/Q(\theta_1)\right]^{-n/2}, \tag{12}$$

where $Q(\theta_1) = \phi^T C \phi = c_{11}\theta_1^2 + 2c_{12}\theta_1 + c_{22} > 0$.

By the likelihood invariance property, the MLE of θ_1 is $\hat{\theta}_1 = \hat{\beta}_1/\hat{\beta}_2$. Now, the relative profile likelihood function of θ_1 is a standardized version that takes a value of one at the maximum of the profile likelihood function of θ given in (12),

$$R(\theta_1) = \frac{L_P(\theta_1)}{L_P(\hat{\theta}_1)} = \left[1 + \frac{n|C|(\hat{\beta}_2\theta_1 - \hat{\beta}_1)^2/Q(\theta_1)}{(n-2)MSE}\right]^{-n/2}. \tag{13}$$

A level c profile likelihood region for θ_1 is given by

$$\{\theta_1 : R(\theta_1) \geq c\},$$

where $0 \leq c \leq 1$. Since $-2\ln[R(\theta_1)]$ follows an asymptotically χ^2 distribution (Sprott, 2008; Kalbfleisch, 1985), approximate likelihood-confidence intervals can be easily obtained. $R(\theta_1)$ is a transformation of the pivotal quantity $F_1(\theta_1)$ defined in (5),

$$R(\theta_1) = \left[1 + \frac{F_1(\theta_1)}{n-2}\right]^{-n/2}, \tag{14}$$

then, a $100(1-\alpha)\%$ likelihood-confidence interval for θ_1 is

$$LCI(y) = \{\theta_1 : R(\theta_1) > c_{1,n-2,1-\alpha}\}, \tag{15}$$

with

$$c_{1,n-2,1-\alpha} = \left[1 + \frac{F_{1,n-2,1-\alpha}}{n-2}\right]^{-n/2},$$

where $F_{1,n-2,1-\alpha}$ is the $(1-\alpha)$ quantile of the F distribution with 1 and $n-2$ degrees freedom, and $0 < \alpha < 1$. Equation (15) implies that, in a θ_1 versus $R(\theta_1)$ plot, the likelihood-confidence interval will consist of all θ_1 values where relative profile likelihood function exceeds the horizontal cutoff $c_{1,n-2,1-\alpha}$.

On the other hand, since $F_1(\theta_1)$ properties are inherited by $R(\theta_1)$, then, in order not only to understand the shape of the profile likelihood function of θ_1, but also to interpret it and to carefully state the results from statistical inferences for $\theta_1 = \beta_1/\beta_2$ and their corresponding hypotheses tests for (β_1, β_2), it is necessary to make a link between the characteristics of the pivotal quantity $F_1(\theta_1)$ and $R(\theta_1)$.

In particular, since the quantity $F_1(\theta_1)$ attains its maximum and minimum and has a horizontal asymptote, then $R(\theta_1)$ will also reach its maximum and minimum and will have a horizontal asymptote. Relative profile likelihood function $R(\theta_1)$ crosses the horizontal asymptote, as illustrated in Figure 1. Note that, by (8) and (14), $F_1(\theta_1)$ inherits the asymptote to $R(\theta_1)$ function, when $|\theta_1|$ becomes larger, so that $R(\theta_1)$ approaches to the asymptote $(1 + F_2/(n-2))^{-n/2}$. Moreover, by varying $c_{1,n-2,1-\alpha}$ level, $(0 < \alpha < 1)$, the length of the likelihood-confidence intervals for θ_1 can be finite or infinite. Something far more interesting is the fact that

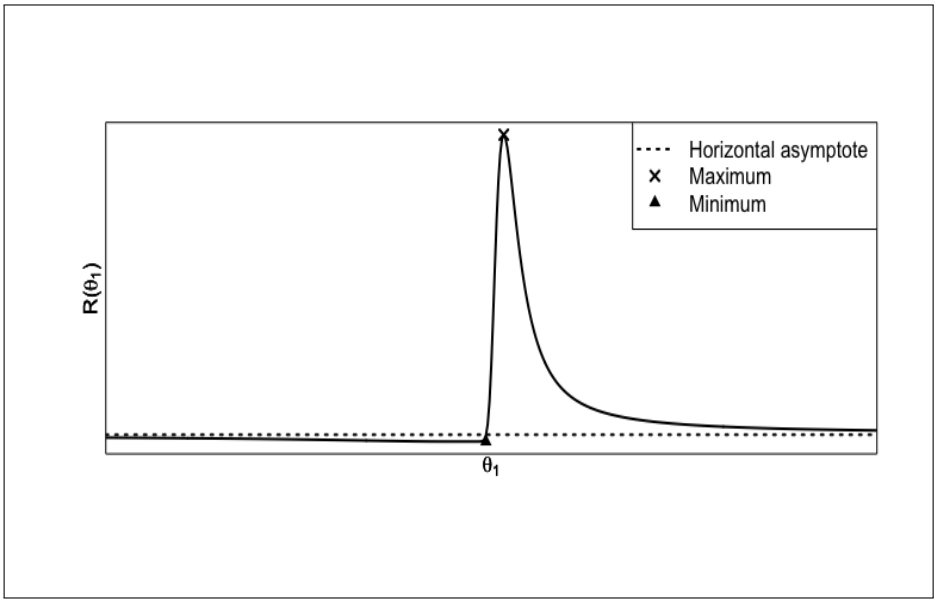


Figure 1: An illustration of the relative profile likelihood function of θ_1. Source: Elaborated by the authors.

likelihood inferences for θ_1 are related to the classical hypotheses tests on the regression coefficients β_1 and β_2, as is explained below.

A case of likelihood-confidence intervals of infinite length for θ_1 occurs when

$$\inf_{\theta_1 \in \mathbb{R}} R(\theta_1) > c_{1,n-2,1-\alpha}. \tag{16}$$

This fact leads to the typical criticism of absurd likelihood-confidence intervals. That is, there is always a confidence level $\alpha < 1$ for which the associated likelihood-confidence interval is the entire real line. This issue should not be taken lightly, because it conceals information about the plausibility of the hypothesis $H_{01} : \beta_1 = \beta_2 = 0$. Since (16) can equivalently be written as

$$2F_1 < F_{1,n-2,1-\alpha},$$

and $F_{1,n-2,1-\alpha} < 2F_{2,n-2,1-\alpha}$, see Casella & Berger (2002, p. 258) and Casella *et al.* (2001, p. 5-7), then the hypothesis $H_{01} : \beta_1 = \beta_2 = 0$ is not rejected at α significance level. Therefore, intervals of this type are related to a non-rejection of H_{01}. On the other hand, when H_{01} is rejected, it is usual to analyze if any of the regression coefficients is zero. The connection between $H_{02} : \beta_2 = 0$ and the infinite length likelihood-confidence intervals is shown below.

Another possibility of likelihood-confidence intervals of infinite length for θ_1 occurs when

$$\lim_{|\theta_1| \to \infty} R(\theta_1) \geq c_{1,n-2,1-\alpha}. \tag{17}$$

In this case, likelihood-confidence interval may result the entire real line or the union of two infinite intervals, and that occurs when $c_{1,n-2,1-\alpha}$ is less than the minimum value of $R(\theta_1)$ or when it lies between the

horizontal asymptote and the minimum value of $R(\theta_1)$, respectively. Note that an infinite length interval is also obtained when the value of $c_{1,n-2,1-\alpha}$ coincides exactly with the asymptote. However, in any case, the hypothesis $H_{02}:\beta_2=0$ is not rejected at α significance level, because (17) is equivalent to

$$F_2 < F_{1,n-2,1-\alpha}.$$

Overall, likelihood-confidence intervals of infinite length suggest moving back to the model criticism stage and testing for model adequacy, in terms of the significance of the regression model coefficients.

4. ILLUSTRATIVE EXAMPLES

This section exemplifies the relationship between parameter θ_1 inferences and hypotheses tests regarding nested models involving β_1 and β_2. In that sense, the three following cases are considered:

- Example 1: $H_{01}:\beta_1=\beta_2=0$ is not rejected,
- Example 2: $H_{02}:\beta_2=0$ is not rejected but $H_{01}:\beta_1=\beta_2=0$ is rejected,
- Example 3: $H_{01}:\beta_1=\beta_2=0$ and $H_{02}:\beta_2=0$ are both rejected.

In all these examples sample size is set at $n=25$ and significance level is fixed at $\alpha=0.05$, so $c_{1,n-2,1-\alpha}$ shown in (15) takes a value of 0.11, which will be used in all these examples. In the same way, $\sigma=5$ remains fixed, as well as the covariable (x_1,x_2) values obtained from Rawlings *et al.* (1998, p. 177). In view of all the above, the procedure results as follows:

- In all the examples, β_1 and β_2 values are provided.
- Given β_1 and β_2 values, y_i is then simulated from a normal distribution with mean $\beta_1 x_{i1}+\beta_2 x_{i2}$ and a standard deviation σ.
- Once F_1 statistic is computed, then $H_{01}:\beta_1=\beta_2=0$ is tested. The critical value corresponding to this hypothesis test is $F_{2,n-2,1-\alpha}=3.42$.
- In case of $H_{01}:\beta_1=\beta_2=0$ rejection, then $H_{02}:\beta_2=0$ is tested by computing F_2 statistic and comparing it with $F_{1,n-2,1-\alpha}=4.27$ critical value.
- Finally, a plot of the relative profile likelihood function for θ_1 is constructed and a $100(1-\alpha)\,\%$ likelihood-confidence interval for this parameter is computed.

4.1. Example 1: $H_{01}:\beta_1=\beta_2=0$ is not rejected

In this example $\beta_1=0.2$ and $\beta_2=0.02$ are considered, simulating then y_i values from a normal distribution with mean $0.2x_{i1}+0.02x_{i2}$ and standard deviation σ. Once the sample for this regression model is generated, the hypothesis $H_{01}:\beta_1=\beta_2=0$ is tested at α significance level. In this case, the test statistic value

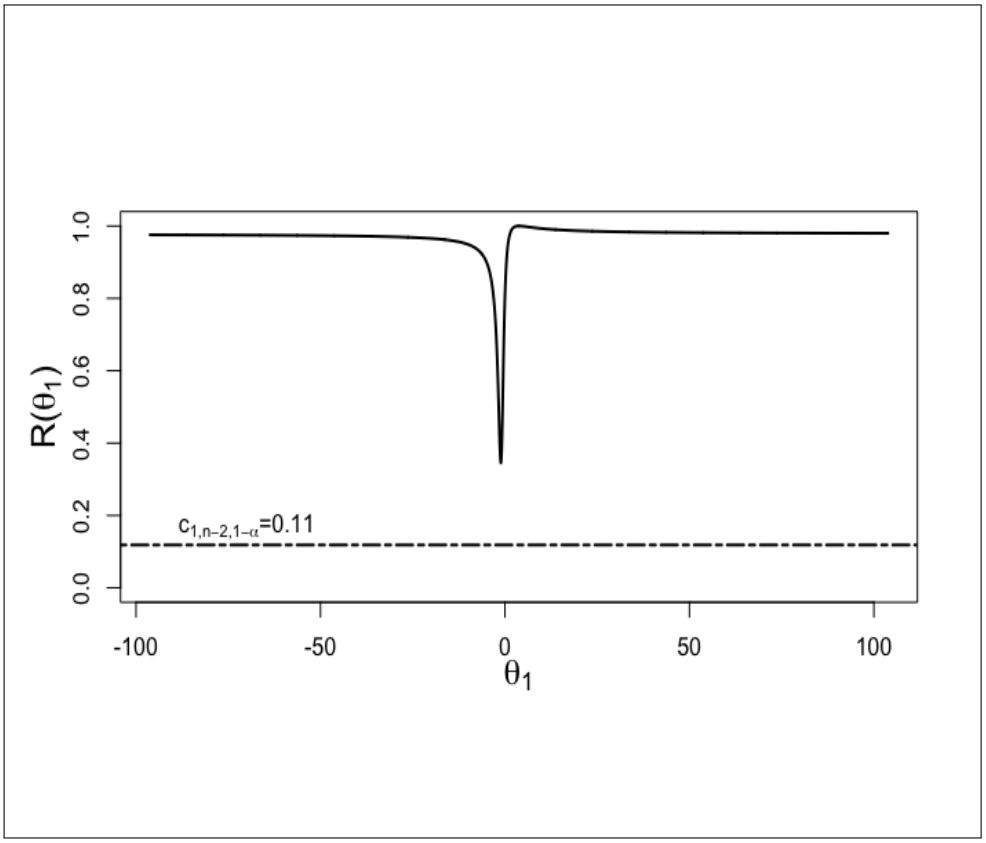


Figure 2: Relative profile likelihood function of θ_1 for Example 1: $H_{01}:\beta_1=\beta_2=0$ is not rejected at α significance level. Source: Elaborated by the authors.

results $F_1 = 1.02$. Now, since $F_1 < F_{2,n-2,1-\alpha}$, there is no statistical evidence to reject the hypothesis $H_{01}:\beta_1=\beta_2=0$; so that these covariables could be excluded from the model. However, if this information is not considered and the inference process regarding θ_1 continues, a relative profile likelihood function of θ_1, like the one shown in Figure 2, is obtained. This type of graph clearly shows a function that is always above the horizontal cutoff $c_{1,n-2,1-\alpha}$; that is, the whole $100(1-\alpha)\,\%$ likelihood-confidence interval for θ_1 results in the real line. Moreover, as explained in Equation (16), when the infimum of the relative profile likelihood function for θ_1 is greater than $c_{1,n-2,1-\alpha}$, then, there is no statistical evidence to reject $H_{01}:\beta_1=\beta_2=0$, at the proposed α significance level.

4.2. Example 2: $H_{02}:\beta_2=0$ is not rejected but $H_{01}:\beta_1=\beta_2=0$ is rejected

In this example $\beta_1 = 0.5$ and $\beta_2 = 0.02$, so y_i values are simulated from a normal distribution with mean $0.5x_{i1}+0.02x_{i2}$ and standard deviation σ. Under this scenario $F_1 = 20.6$, so it follows that $F_1 > F_{2,n-2,1-\alpha}$, and at α significance level there is statistical evidence to reject the hypothesis $H_{01}:\beta_1=\beta_2=0$. According to the procedure previously stated, hypothesis $H_{02}:\beta_2=0$ is now tested. In this case $F_2 = 1.63$, so $F_2 < F_{1,n-2,1-\alpha}$; that is, there is no statistical evidence to reject the hypothesis $H_{02}:\beta_2=0$, at α significance

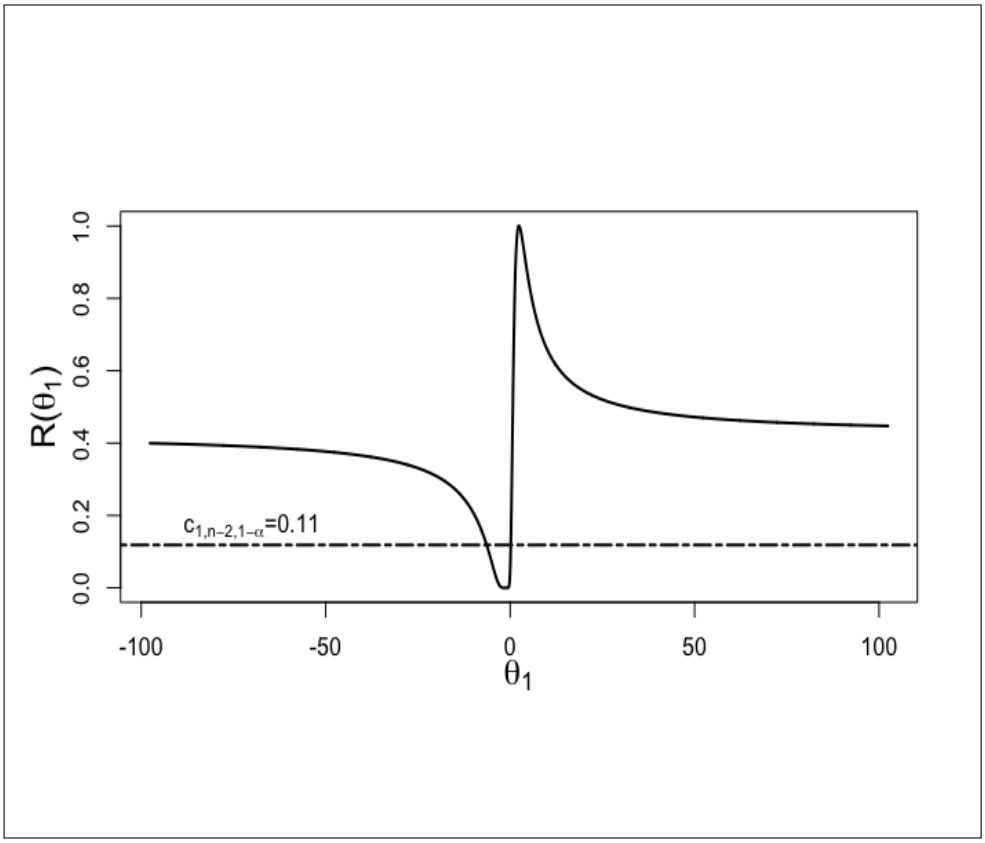


Figure 3: Relative profile likelihood function of θ_1 for Example 2: At α significance level $H_{02}:\beta_2=0$ is not rejected but $H_{01}:\beta_1=\beta_2=0$ is rejected. Source: Elaborated by the authors.

level, so x_2 variable could be excluded from the regression model. However, if this variable remains in the model and the inference process about parameter θ_1 continues, a $100(1-\alpha)\,\%$ likelihood-confidence interval of infinite length for θ_1 is also obtained. That occurs because the asymptote of the relative profile likelihood function for θ_1 is above the cutoff $c_{1,n-2,1-\alpha}$, as shown in Figure 3. Note that, in this case, the infimum of the relative profile likelihood is not above this cutoff, like in previous example where $H_{01}:\beta_1=\beta_2=0$ is not rejected, but the asymptote of the relative profile likelihood function for θ_1, as explained in Equation 17, is above the cutoff $c_{1,n-2,1-\alpha}$, so there is no statistical evidence to reject $H_{02}:\beta_2=0$, at α significance level.

4.3. Example 3: $H_{01}:\beta_1=\beta_2=0$ and $H_{02}:\beta_2=0$ are both rejected

To exemplify this scenario $\beta_1=0.5$ and $\beta_2=0.6$ are considered, so y_i values are generated based on a normal distribution with mean $0.5x_{i1}+0.6x_{i2}$ and standard deviation σ. Now, since $F_1=27.44$, then $F_1>F_{2,n-2,1-\alpha}$; therefore, the hypothesis $H_{01}:\beta_1=\beta_2=0$ is rejected at α significance level. Similarly, as $F_2=4.64$, then $F_2>F_{1,n-2,1-\alpha}$ and $H_{02}:\beta_2=0$ is also rejected at α significance level. When continuing with the inference process regarding θ_1, under this situation, a finite $100(1-\alpha)\,\%$ likelihood-confidence interval for θ_1 is

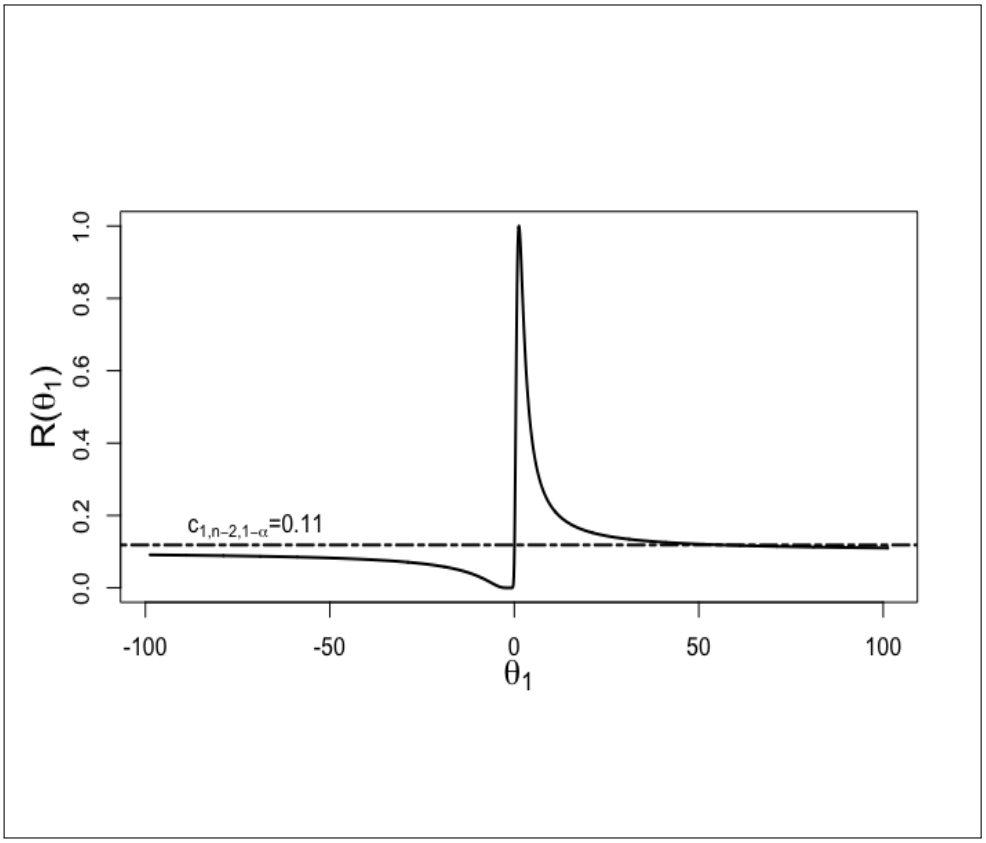


Figure 4: Relative profile likelihood function of θ_1 for Example 3: $H_{01} : \beta_1 = \beta_2 = 0$ y $H_{02} : \beta_2 = 0$ are both rejected at α significance level. Source: Elaborated by the authors.

obtained. In this case the asymptote of the relative profile likelihood function for θ_1 is below the cutoff $c_{1,n-2,1-\alpha}$, as shown in Figure 4.

As can be observed in the examples included here, when an infinite $100(1-\alpha)\,\%$ likelihood-confidence interval for θ_1 is obtained, it implies that there will be no statistical evidence to reject any of the hypotheses $H_{01} : \beta_1 = \beta_2 = 0$ or $H_{02} : \beta_2 = 0$, at α significance level. Note that, as mentioned in previous section, when hypothesis $H_{02} : \beta_2 = 0$ is not rejected, at α significance level, then a $100(1-\alpha)\,\%$ likelihood-confidence interval of infinite length, for parameter $\theta_1 = \beta_1/\beta_2$, is obtained.

5. CONCLUSIONS

The purpose of this article is to emphasize the importance of deeply understanding and analyzing the shape of the likelihood function and their corresponding inferences. In the cases discussed here, the infinite length of likelihood-confidence intervals are due to the fact of considering full models when reduced models are not rejected by the data. Occurrence of flat likelihood shapes, like the ones presented here, are very informative regarding model criticisms and should not be the reason for seeking or promoting alternative estimation

methods such as integrated likelihoods or posterior Bayesian distributions.

Acknowledgements

We thank Dr. Gudelia Figueroa Preciado for her help in the development of this paper and for her insightful discussions. We also thank the anonymous referees for their helpful suggestions.

Funding

This article was supported by Consejo Nacional de Ciencia y Tecnología. Doctoral fellowship of the first author.

References

Berger, J. O., Liseo, B. & Wolpert, R. L. (1999). Integrated likelihood methods for eliminating nuisance parameters. *Statistical Science*, 14(**1**), 1-22.

Breusch, T. S., Robertson, J. C. & Welsh, A. H. (1997). The emperor's new clothes: a critique of the multivariate t regression model. *Statistica Neerlandica*, 51(**3**), 269-286.

Casella, G. & Berger, R. L. (2002). Statistical inference. Duxbury. Pacific Grove, CA.

Casella, G., Berger, R. L. & Santana, D. (2001). Solutions Manual for Statistical Inference.

Catchpole, E. A. & Morgan, B. J. (1997). Detecting parameter redundancy. *Biometrika*, 84(**1**), 187-196.

Cheng, R. C. H. & Iles, T. C. (1990). Embedded models in three-parameter distributions and their estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, 52(**1**), 135-149.

Cole, S. R., Chu, H. & Greenland, S. (2013). Maximum likelihood, profile likelihood, and penalized likelihood: a primer. *American Journal of Epidemiology*, 179(**2**), 252-260.

Cox, D. R. (2006). Principles of statistical inference. Cambridge university Press. Cambridge.

El Adlouni, S., Ouarda, T. B., Zhang, X., Roy, R. & Bobée, B. (2007). Generalized maximum likelihood estimators for the nonstationary generalized extreme value model. *Water Resources Research*, 43 (**3**), W03410.

Farcomeni, A.& Tardella, L. (2012). Identifiability and inferential issues in capture-recapture experiments with heterogeneous detection probabilities. *Electronic Journal of Statistics*, 6, 2602-2626.

Fay, M. P., Noubary, F. & Saul, A. (2009). Robust Noninferiority Tests for Potency of a Test Drug Against a Reference Drug. *Statistics in Biopharmaceutical Research*, 1(**3**), 291-300.

Frery, A. C., Cribari-Neto, F. & De Souza, M. O. (2004). Analysis of minute features in speckled imagery with maximum likelihood estimation. *EURASIP Journal on Advances in Signal Processing*, 2004(**16**), 2476-2491.

Ghosh, M., Datta, G. S., Kim, D. & Sweeting, T. J. (2006). Likelihood-based inference for the ratios of regression coefficients in linear models. *Annals of the Institute of Statistical Mathematics*, 58(**3**), 457-473.

Ghosh, M., Yin, M. & Kim, Y. H. (2003). Objective Bayesian inference for ratios of regression coefficients in linear models. *Statistica Sinica*, 13(**2**), 409-422.

Harter, H. L. & Moore, A. H. (1966). Local-maximum-likelihood estimation of the parameters of three-parameter lognormal populations from complete and censored samples. *Journal of the American Statistical Association*, 61(**315**), 842-851.

Hirschberg, J. & Lye, J. (2010). A reinterpretation of interactions in regressions. *Applied Economics Letters*, 17(**5**), 427-430.

Kalbfleisch, J. G. (1985). Probability and Statistical Inference, Vol. 2. Springer-Verlag. New York.

Kreutz, C., Raue, A., Kaschek, D.& Timmer, J. (2013). Profile likelihood in systems biology. *The FEBS journal*, 280(**11**), 2564-2571.

Li, R. & Sudjianto, A. (2005). Analysis of computer experiments using penalized likelihood in Gaussian Kriging models. *Technometrics*, 47(**2**), 111-120.

Lima, V. M. & Cribari-Neto, F. (2019). Penalized maximum likelihood estimation in the modified extended Weibull distribution. *Communications in Statistics-Simulation and Computation*, 48(**2**), 334-349.

Liu, S., Wu, H. & Meeker, W. Q. (2015). Understanding and addressing the unbounded "likelihood" problem. *The American Statistician*, 69(**3**), 191-200.

Martins, E. S. & Stedinger, J. R. (2000). Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resources Research*, 36(**3**), 737-744.

Martins, E. S. & Stedinger, J. R. (2001). Generalized maximum likelihood Pareto-Poisson estimators for partial duration series. *Water Resources Research*, 37(**10**), 2551-2557.

Montoya, J. A., Díaz-Francés, E.& Sprott, D. A. (2009). On a criticism of the profile likelihood function. *Statistical Papers*, 50(**1**), 195-202.

Pewsey, A. (2000). Problems of inference for Azzalini's skewnormal distribution. *Journal of Applied Statistics*, **27**(7), 859-870.

Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U. & Timmer, J. (2009). Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, 25(**15**), 1923-1929.

Rawlings, J. O., Pantula, S. G. & Dickey, D. A. (1998). Applied regression analysis: a research tool. Springer-Verlag. New York.

Rosenblad, A. K. (2020). The mean, variance, and bias of the OLS based estimator of the extremum of a quadratic regression model for small samples. *Communications in Statistics-Theory and Methods*, 51(**9**), 2870-2886.

Searle, S. R. & Gruber, M. H. (1971). Linear models. John Wiley & Sons. New York.

Sprott, D. A. (2008). Statistical inference in science. Springer-Verlag. New York.

Sundberg, R. (2010). Flat and multimodal likelihoods and model lack of fit in curved exponential families. *Scandinavian Journal of Statistics*, 37(**4**), 632-643.

Tsionas, E. G. (2001). Likelihood and Posterior Shapes in Johnson's System. *Sankhyā: The Indian Journal of Statistics, Series B*, 63(**1**), 3-9.

Tumlinson, S. E. (2015). On the non-existence of maximum likelihood estimates for the extended exponential power distribution and its generalizations. *Statistics & Probability Letters*, 107, 111-114.

Revista de la Facultad de Ciencias Universidad Nacional de Colombia, Sede Medellín
V 11 N°2 julio-diciembre de 2022 • ISSN-e 2357-5749 • Research paper • Pages 39 a 53
DOI: https://doi.org/10.15446/rev.fac.cienc.v11n2.98450

FLAT LIKELIHOODS: THREE-PARAMETER WEIBULL MODEL CASE[a]

VEROSIMILITUDES PLANAS: CASO DEL MODELO WEIBULL DE TRES PARÁMETROS

JOSÉ A. MONTOYA[b*], GUDELIA FIGUEROA PRECIADO[c]

Recibido 15-09-2021, aceptado 27-01-2022, versión final 16-02-2022
Research paper

ABSTRACT: Criticisms of maximum likelihood estimation frequently occur when likelihood function shape becomes flat. Although some research have been done regarding the possible causes of a flat likelihood, more work is needed to expand our knowledge on this subject. In this paper we analyze the origin of Weibull flat likelihoods. In particular, we study the severity of the likelihood flatness by examining the limit behaviour of the relative profile likelihood for the three-parameter Weibull threshold parameter, when this parameter goes to infinity. In the cases discussed here, flat likelihoods are not only related to sample size but also to an embedded model problem. Due to the widespread use of the likelihood function in inferential statistical methods, it is important not only to identify factors that can cause flat likelihoods, but also to study the severity of this flattening, in order to develop or apply *ad hoc* statistical and computational methods for making inferences.
KEYWORDS: Flat likelihood function; threshold parameter; embedded model; GEV distribution; likelihood contours; profile likelihood function.

RESUMEN: Críticas a la estimación por máxima verosimilitud ocurren frecuentemente cuando la forma de la función de verosimilitud es plana. Aunque se ha realizado investigación respecto a las posibles causas de una verosimilitud plana, es necesario un mayor trabajo para expandir nuestro conocimiento sobre este tema. En este artículo se analiza el origen de verosimilitudes Weibull planas. En particular, se estudia la severidad de la planura de la verosimilitud a través de examinar el comportamiento del límite de la verosimilitud perfil relativa del parámetro umbral del modelo Weibull de tres parámetros, cuando este parámetro se va a infinito. En los casos que aquí se presentan, las verosimilitudes planas no están solamente relacionadas con el tamaño de la muestra sino también con un problema de modelos empotrados. Dado el amplio uso de la función de verosimilitud en métodos estadísticos inferenciales, es importante no solamente identificar los factores que pueden ocasionar verosimilitudes planas, sino también estudiar lo severo de este aplanamiento, a fin de aplicar métodos estadísticos y computacionales *ad hoc*, al realizar inferencias.
PALABRAS CLAVE: Función de verosimilitud plana; parámetro umbral; modelo empotrado; Distribución de VEG, contornos de verosimilitud; función de verosimilitud perfil.

[a]Montoya, J. A. & Figueroa-Preciado, G. (2022). Flat likelihoods: Three-parameter Weibull model case. *Rev. Fac. Cienc.*, 11 (**2**), 39–53. DOI: https://doi.org/10.15446/rev.fac.cienc.v11n2.98450

[b]PhD in Probability and Statistics. Statistics Professor. Department of Mathematics. University of Sonora, México.
[*]Corresponding author: arturo.montoya@unison.mx
[c]PhD in Probability and Statistics. Statistics Professor. Department of Mathematics. University of Sonora, México.

1. INTRODUCTION

Statistical methods based on likelihood have been subject to severe criticisms, generally focused on getting ridiculous estimation values (Harter & Moore, 1966; Breusch *et al.*, 1997; Berger *et al.*, 1999; Martins & Stedinger, 2000; Pewsey, 2000; Martins & Stedinger, 2001; El Adlouni *et al.*, 2007; Tumlinson, 2015). The shape of the likelihood function is usually associated with these strange results; however, abnormal likelihoods can be thought of as a symptom, not a cause; see Montoya (2008), Montoya *et al.* (2009), and Liu *et al.* (2015).

Maximum likelihood estimation criticisms occur quite often and many of them arise when the shape of the likelihood function is flat. Actually, the occurrence of flat likelihoods have been used to promote integrated likelihoods, Bayesian posteriors, penalized methods as well as new optimization methods (Berger *et al.*, 1999; Tsionas, 2001; Frery *et al.*, 2004; Li & Sudjianto, 2005; Ghosh *et al.*, 2006; Cole *et al.*, 2013; Lima & Cribari-Neto, 2019). However, the problem of flat likelihoods is complex and multifactorial. Some authors have related this problem with an overparameterization of the model, inappropriate reparameterizations, embedded models, a limited amount and quality of experimental data, and also a poor model fit when sample size is large (Cheng & Iles, 1990; Catchpole & Morgan, 1997; Raue *et al.*, 2009; Sundberg, 2010; Farcomeni, & Tardella, 2012; Kreutz *et al.*, 2013).

In this article we focus on the three-parameter Weibull model. This distribution is widely used in reliability studies to describe failure or waiting times in normal or accelerated life testing (Elmahdy & Aboutahoun, 2013; Elmahdy, 2015). It is also commonly applied in many other contexts such as in forestry, when studying tree diameters distribution (Green *et al.*, 1994); in medicine, for the study of the survival times of cancer patients (Khan *et al.*, 2011); rainfall studies in hydrology (Koutsoyiannis, 2004; Silva & Peiris, 2017; Bolívar-Cimé *et al.*, 2015), and also for the study of tides in climatology (De Haan, 1990). The three-parameter Weibull model has been also used by Deng *et al.* (2018) to study the fracture strength of brittle ceramics and other materials; for describing the voltage breakage of electric circuits (Hirose & Lai, 1997), and to analyze the time needed to complete a task, in cognitive psychology (Cousineau, *et al.*, 2002), among many other applications.

The three-parameter Weibull density function is given by

$$f(x;\alpha,\gamma,\beta) = \frac{\beta}{\gamma}\left(\frac{x-\alpha}{\gamma}\right)^{\beta-1} \exp\left[-\left(\frac{x-\alpha}{\gamma}\right)^{\beta}\right], \tag{1}$$

where $x \geq \alpha$, $\gamma \geq 0$, and $\beta \geq 0$. In this parameterization α is a threshold parameter, while γ and β are scale and shape parameters, respectively. When $\alpha = 0$, a two parameter Weibull distribution is obtained.

Many researchers have reported that maximum likelihood estimation of the threshold parameter, of a three-parameter Weibull distribution, can sometimes cause both computational and statistical difficulties. Several

papers in statistical literature have devoted considerable effort to address these difficulties, usually proposing other estimation procedures, as alternatives to the maximum likelihood approach. These difficulties have been mainly linked with the *non-regular threshold problem* and the *embedded model problem*, as can be found in Smith & Naylor (1987), Cheng & Iles (1990) and Hirose & Lai (1997). The first problem occurs when working with density functions that have singularities, since these are inherited by the likelihood function; hence, for certain parameter values, likelihood becomes unbounded. When density functions are replaced by the probability of the observed sample, then the likelihood function poses no problems of this sort, as stated by Barnard (1967), Barnard & Sprott (1983), Montoya *et al.* (2009) and Liu *et al.* (2015), among others. Now, the *embedded model problem*, is characterized by a relatively flat or constant likelihood function over a wide range of values of the threshold parameter. Cheng & Iles (1990) considered that this is a model selection problem. They mentioned that a three-parameter Weibull distribution has an embedded two-parameter model (the extreme value distribution) and that its presence is the reason behind the relatively flat likelihood, which gives rise to the reported computational difficulties. On the other hand, Hirose & Lai (1997) considered that the embedded model problem occurs when having one more parameter that can not be estimated given the sample size.

In this paper we perform simulations to investigate to what extent the three-parameter Weibull model lead to flat likelihoods. We analyze the limit of the relative profile likelihood of threshold parameter α, when this parameter goes to infinity, and show that flat likelihoods occur frequently. In most cases, this limit value has an empirical distribution that assigns a negligible probability to values close to zero. Moreover, the empirical distribution of this limit value estimates that at least 12 percent of all likelihood functions that result from three-parameter Weibull model under small sample size $n = 20$ are essentially flat or constant likelihoods over a huge range of values the threshold parameter. We also show that flat likelihoods occur very frequently when data of a Gumbel model, embedded in the three-parameter Weibull distribution, are modelled by a three-parameter Weibull model. Specifically, for all sample size under study, the empirical distribution of this limit value estimates that at least 45 percent of all likelihood functions that result from three-parameter Weibull model under data of a Gumbel model are essentially flat or constant likelihood functions over a huge range of values the threshold parameter.

In Section 2 we present the three parameter Weibull likelihood and relative likelihood functions of the vector parameters, as well as the profile likelihood function of threshold parameter. Section 3 focuses on analyzing an example presented by Hirose & Lai (1997), that allows us to show that a lack of parameter identifiability based on observed data makes impossible to distinguish between parameter candidates. On the other hand, Section 4 is devoted to analyze some simulation scenarios where the frequency of occurrence of certain behaviour of the relative profile likelihood, is taken as an indicator for the degree of non-identifiability of the parameters, based on the sample. Finally, some conclusions are presented in Section 6.

2. WEIBULL LIKELIHOOD

Suppose $X_1, X_2, \cdots, X_k$ are i.i.d. random variables, distributed as a three-parameter Weibull with unknown parameters (α, γ, β). Let $\boldsymbol{x} = (x_1, x_2, \cdots, x_k)$ be realization of this sample. The likelihood function of the vector parameters (α, γ, β), based on the observed sample $\boldsymbol{x}$ and in (1), is

$$L(\alpha, \gamma, \beta) = \left(\frac{\beta}{\gamma}\right)^k \left[\prod_{i=1}^{k}\left(\frac{x_i - \alpha}{\gamma}\right)\right]^{\beta-1} \exp\left[-\sum_{i=1}^{k}\left(\frac{x_i - \alpha}{\gamma}\right)^{\beta}\right], \tag{2}$$

for $\gamma \geq 0$, $\beta \geq 0$, and $\alpha \leq x_{(1)}$, where $x_{(1)}$ is the minimum value of the observed sample. Just for notational simplicity, sample dependence of L is not explicitly written.

The relative likelihood function of (α, γ, β) is a standardized version of (2); this takes the value of one at its maximum, the MLE of (α, γ, β),

$$R(\alpha, \gamma, \beta) = \frac{L(\alpha, \gamma, \beta)}{\max\limits_{\alpha,\gamma,\beta} L(\alpha, \gamma, \beta)} = \frac{L(\alpha, \gamma, \beta)}{L\left(\widehat{\alpha}, \widehat{\gamma}, \widehat{\beta}\right)}, \tag{3}$$

where, $\widehat{\alpha}$, $\widehat{\gamma}$, and $\widehat{\beta}$ are the MLE's of parameters α, γ, and β, respectively.

Now, assume that vector parameter (α, γ, β) can be partitioned as follows: an interest parameter ψ and a nuisance parameter λ. Then, the profile likelihood of ψ and its corresponding relative likelihood function are defined as

$$L_{\max}(\psi) = \max_{\lambda} L(\psi, \lambda), \tag{4}$$

$$R_{\max}(\psi) = \frac{L_{\max}(\psi)}{\max\limits_{\psi,\lambda} L(\psi, \lambda)} = \frac{L_{\max}(\psi)}{L_{\max}(\widehat{\psi})}. \tag{5}$$

For example, the profile likelihood and its corresponding relative likelihood function of $\psi = \alpha$ is

$$L_{\max}(\alpha) = \max_{\gamma,\beta} L(\alpha, \gamma, \beta), \tag{6}$$

$$R_{\max}(\alpha) = \frac{L_{\max}(\alpha)}{\max\limits_{\alpha,\gamma,\beta} L(\alpha, \gamma, \beta)} = \frac{L_{\max}(\alpha)}{L_{\max}(\widehat{\alpha})}. \tag{7}$$

In general, the profile or maximized likelihood is a powerful though simple inferential method for estimating a parameter of interest separately from the remaining unknown parameters of the statistical model. More details about the profile likelihood approach are described in Kalbfleisch (1985, Section 10.3), Pawitan (2001, Section 3.4), Sprott (2000, p. 66), Barndorff-Nielsen & Cox (1994, pp. 89-91), Serfling (2002, pp. 155-160), and Murphy & Van Der Vaart (2000). An important and mostly unknown aspect about the profile likelihood function is that it can be used to study and visualize various aspects of the full likelihood function, such as for instance those associated with the flatness of the likelihood surface. The following example illustrates how the graph of a profile likelihood, particularly the case of α, can reveal the flat nature of the likelihood function of (α, γ, β).

Table 1: Simulated sample from a three-parameter Weibull distribution with vector parameters $(0, 3.27, 10.16)$.

3.13	2.84	3.35	2.78	2.66	3.47	3.58	2.97	3.26	3.34
2.36	3.41	2.28	2.77	2.99	2.97	3.49	3.53	3.11	2.75

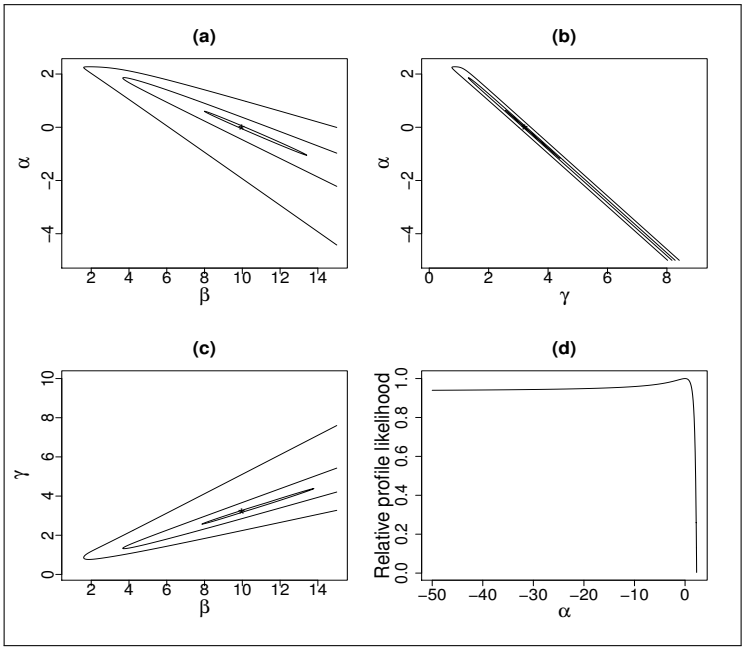


Figure 1: Relative profile likelihood functions corresponding to the data in Table 1. Source: Elaborated by the authors.

3. EXAMPLE: WEIBULL FLAT LIKELIHOOD

The dataset in Table 1 has been generated using a three-parameter Weibull distribution, with vector parameters $(\alpha, \gamma, \beta) = (0, 3.27, 10.16)$. The selection of these values was motivated by the example included in Hirose & Lai (1997), that involves $k = 20$ breakdown voltages. Figure 1(a-c) shows contour plots of the relative profile likelihood of (β, α), (γ, α), and (β, γ) for the simulated sample, displaying only three contour levels at 0.995, 0.75, and 0.05; in each one of these cases, the parameter MLE's are marked with an asterisk. It is clear from looking these elongated contours, that the three parameters are interrelated. In addition, the graph shows that the likelihood function has a ridge with a relatively flat top. In fact, Figure 1(d) shows that the relative profile likelihood of α is relatively flat over a huge range of α values.

When for a fixed data set it is observed that two different values of the parameters give rise to the same likelihood (probability of observed data), then it would be impossible to distinguish between these two parameter candidates, based on the data alone; thus, it would be not viable to identify the true parameter. In that sense, when the relative profile likelihood function of α, $L_{\max}(\alpha)$, is flat over a wide range of α values, including the MLE such as in Figure 1(d), then the probability of the observed sample is the same over all this α value set. Therefore, the lack of identifiability of these parameters occurs at least for the observed da-

ta. Since the profile likelihood function for the parameter of interest is invariant under reparameterizations of the nuisance parameters, then, the graph of the profile likelihood function of α will not change under reparameterizations of (γ, β).

Sprott (2000, p. 11) stated that the MLE and the observed information exhibit two features of the likelihood function, the MLE as a measure of its location relative to the parameter-axes, while the latter as a measure of its curvature in a neighborhood of the MLE. Thus, when the likelihood function of a scalar parameter is smooth and symmetric around the MLE, then the MLE and the observed information are usually the main features for determining the shape of the likelihood function. However, it should be stressed that flat likelihoods can not be determined by these quantities without loss of information.

Intuitively, the sensitivity of the MLE depends on the shape of the likelihood function. If the likelihood of α is very curved around MLE, this turns out to be a stable estimate of α. On the other hand, if the likelihood is flat near the MLE, this becomes highly unstable. Something important to mention, besides these facts, is that a flat likelihood can lead to likelihood-confidence intervals of α where its lower limit becomes minus infinity. In the next section we investigate the frequency of the occurrence of flat likelihoods of threshold parameter α by exploring the limit of the relative profile likelihood of α, as α approaches to minus infinity. Here we will analyze two situations: one corresponding to three-parameter Weibull samples and the other one considers samples of the embedded Gumbel model.

4. FLATNESS OF THE RELATIVE PROFILE LIKELIHOOD OF α

In this section we illustrate, to a certain degree, the non-identifiability of the parameters for the three-parameter Weibull model, considering a given sample size and focusing on the limit of the relative profile likelihood function of α, as α approaches to minus infinity,

$$R_{\max}(\infty) = \lim_{\alpha \to -\infty} R_{\max}(\alpha) = \lim_{\alpha \to -\infty} \frac{L_{\max}(\alpha)}{\max\limits_{\alpha,\gamma,\beta} L(\alpha, \gamma, \beta)}. \tag{8}$$

If $R_{\max}(\infty)$ is smaller than one, but close to it, then for every real number $\varepsilon > 0$, there is a real number α_0 such that for all $\alpha < \alpha_0$, we have $R_{\max}(\alpha) \in (R_{\max}(\infty) - \varepsilon, R_{\max}(\infty) + \varepsilon)$. Thus, there is a value set for α that makes the observed sample as likely as the MLE does. The frequency of occurrence of this type of behavior is like an indicator for the degree of non-identifiability of the parameters, based on the sample. In particular, $R_{\max}(\infty) = 1$ can be interpreted as an extreme case of lack of identifiability of the three-parameter Weibull model, based on the observed data. This situation seem to suggest that the dataset may be fitted with the embedded two-parameter model (the extreme value distribution), which corresponds to the limiting case of (1) as $\beta \to \infty$, $\gamma \to \infty$, and $\alpha \to -\infty$ such that $\gamma + \alpha \to \mu$, and $\gamma/\beta \to \sigma$ for some constants μ and σ.

Cheng & Iles (1990) showed that

$$log\left[L_{\max}(\alpha)\right] = l_0(\widehat{\mu},\widehat{\sigma}) + \frac{l_1(\widehat{\mu},\widehat{\sigma})}{\alpha} + O\left(\frac{1}{\alpha^2}\right), \tag{9}$$

where

$$l_0(\mu,\sigma) = -klog(\sigma) + \frac{1}{\sigma}\sum_{i=1}^{k}(x_i-\mu) - \sum_{i=1}^{k}\exp\left[\frac{(x_i-\mu)}{\sigma}\right] \tag{10}$$

is the log-likelihood function corresponding to the two-parameter extreme value model,

$$l_1(\mu,\sigma) = \sum_{i=1}^{k}(x_i-\mu) + \frac{1}{2\sigma}\sum_{i=1}^{k}(x_i-\mu)^2 - \frac{1}{2\sigma}\sum_{i=1}^{k}(x_i-\mu)^2\exp\left[\frac{(x_i-\mu)}{\sigma}\right], \tag{11}$$

and $(\widehat{\mu},\widehat{\sigma})$ is the value of (μ,σ) that maximizes $l_0(\mu,\sigma)$. Thus, for computational convenience, an approximate version of the relative profile likelihood will be used in order to calculate this limit. Our approximate version of $R_{\max}(\infty)$ is therefore

$$\widetilde{R}_{\max}(\infty) = \begin{cases} \dfrac{\exp\left[l_0(\widehat{\mu},\widehat{\sigma})\right]}{\max\limits_{\alpha,\gamma,\beta} L(\alpha,\gamma,\beta)} & \text{, if } l_1(\widehat{\mu},\widehat{\sigma}) < 0, \\ 1 & \text{, if } l_1(\widehat{\mu},\widehat{\sigma}) \geq 0. \end{cases} \tag{12}$$

4.1. Some scenarios for samples from the three-parameter Weibull distribution

An experiment was conducted to analyze the behavior of the relative profile likelihood function of α, as α approaches to minus infinity. Weibull samples of size 20, 50, 125, and 200 were simulated 10000 times, considering the following scenarios.

- Case 1, $(\alpha,\gamma,\beta) = (\alpha, 4.71, 21.26)$ and $\alpha = -3.46$, or -2.96 or -2.21. The selection of these values was motivated by Example 1 from Cheng & Iles (1990), concerning fibre strengths measurements, where it was found that $\widehat{\alpha} = -3.46$, $\widehat{\gamma} = 4.71$ and $\widehat{\beta} = 21.26$.
- Case 2, $(\alpha,\gamma,\beta) = (\alpha, 3, 10)$ and $\alpha = 0.1$, or 0.5 or 1: These parameter values were used by Hirose & Lai (1997), in order to show that the MLE of α take the value $-\infty$ with probability over 0.4, for the case of interval-grouped observations.
- Case 3, $(\alpha,\gamma,\beta) = (0.1, 5.7, 7), (0.1, 5.7, 5.5), (0.1, 8, 5)$. These values correspond to credible values of the posterior distribution obtained by Green *et al.* (1994) in three samples of tree diameters (samples 3-1).

Table 2 summarizes a classification of all the simulated samples for the above scenarios, according to whether $0 \leq \widetilde{R}_{\max}(\infty) < 1$ or $\widetilde{R}_{\max}(\infty) = 1$. Note that $\widetilde{R}_{\max}(\infty) = 1$ corresponds to samples where the MLE of α actually diverged to $-\infty$. At a first glance, it can be observed that as the sample size increases, there is a decrease in the proportion of samples where $\widetilde{R}_{\max}(\infty) = 1$. For all scenarios in Case 1, $\widetilde{R}_{\max}(\infty)$ assumes the value 1 with probability close to 0.36, 0.30, 0.20 and 0.15 when $n = 20$, 50, 125, and 200. For all scenarios

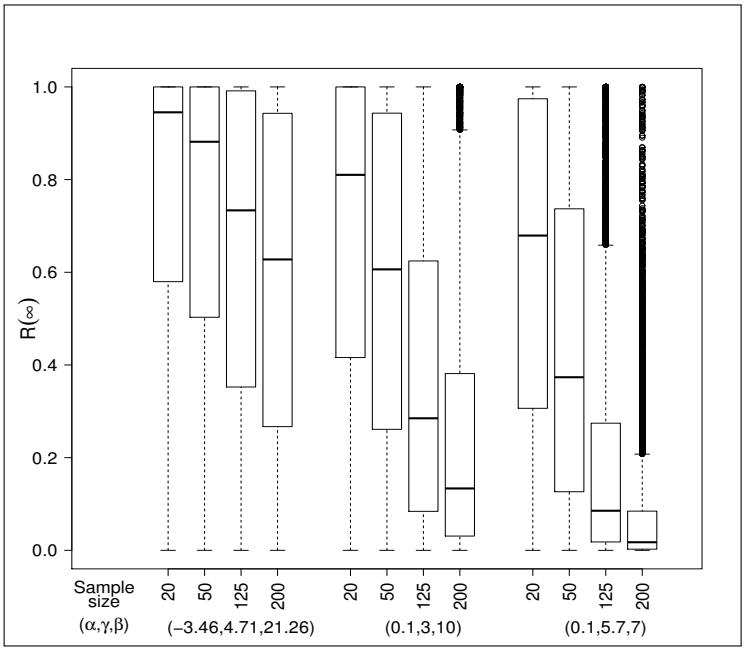


Figure 2: Box plots for three scenarios simulation Weibull of the Table 2. Source: Elaborated by the authors.

in Case 2, these probabilities are close to 0.26, 0.15, 0.04, and 0.01. In Case 3, the scenario with the highest proportion of samples where $\widetilde{R}_{\max}(\infty) = 1$ is the one where $(\alpha, \gamma, \beta) = (0.1, 5.7, 7)$, corresponding to sample 3 from Green *et al.* (1994). Here, $\widetilde{R}_{\max}(\infty)$ assumes the value 1 with probabilities close to 0.18, 0.06, 0.007 and 0, for $n = 20$, 50, 125, and 200, respectively. Thus, a general conclusion is that an embedded Gumbel model is no longer a reasonable model for the data, when sample size increases.

Box plots displayed in Figure 2 provide an idea of the shape of the frequency distribution of $\widetilde{R}_{\max}(\infty)$, for three of the simulation scenarios included in Table 2. Looking at these boxplots we can observe a large degree of non-identifiability of parameters for the three-parameter Weibull model, and notice that flat likelihoods can occur very frequently. Hence, in this kind of scenarios, it is difficult to identify the parameters of the three-parameter Weibull distribution; particularly parameter α, that represent the lower end of the support of this model.

4.2. Embedded Gumbel model in the three-parameter Weibull distribution

It is interesting to analyze scenarios such as the ones included in this section, where we simulate data from an embedded two-parameter model (the extreme value distribution) and the behavior of $\widetilde{R}_{\max}(\infty)$ is again studied. Samples of size 20, 50, 125, and 200 were simulated 10000 times from a Gumbel distribution with parameters $\mu = \alpha + \gamma$ and $\sigma = \gamma/\beta$, where the values assigned to (α, γ, β) correspond to those presented in Table 2. Once again, the samples are classified according to whether $0 \leq \widetilde{R}_{\max}(\infty) < 1$ or $\widetilde{R}_{\max}(\infty) = 1$, summarizing the results in Table 3. As we previously stated, $\widetilde{R}_{\max}(\infty) = 1$ is linked to those samples whe-

Table 2: Classification (in percentage) of 10000 simulated samples of a three-parameter Weibull random variable.

(α,γ,β)	k	$0 \leq \widetilde{R}_{\max}(\infty) < 1$	$\widetilde{R}_{\max}(\infty) = 1$
$(-3.46, 4.71, 21.26)$	20	62.24	37.76
	50	70.05	29.95
	125	79.30	20.70
	200	85.09	14.91
$(-2.96, 4.71, 21.26)$	20	63.15	36.85
	50	70.21	29.79
	125	79.06	20.94
	200	84.66	15.34
$(-2.21, 4.71, 21.26)$	20	63.98	36.02
	50	69.83	30.17
	125	79.35	20.65
	200	84.03	15.97
$(0.1, 3, 10)$	20	74.26	25.74
	50	84.66	15.34
	125	95.93	4.07
	200	98.71	1.29
$(0.5, 3, 10)$	20	73.69	26.31
	50	85.79	14.21
	125	95.66	4.34
	200	98.67	1.33
$(1, 3, 10)$	20	73.48	26.52
	50	85.67	14.33
	125	95.14	4.86
	200	98.52	1.48
$(0.1, 5.7, 7)$	20	81.59	18.41
	50	93.26	6.74
	125	99.26	0.74
	200	99.93	0.07
$(0.1, 5.7, 5.5)$	20	86.20	13.80
	50	96.92	3.08
	125	99.86	0.14
	200	100	0
$(0.1, 8, 5)$	20	88.40	11.60
	50	98.04	1.96
	125	99.92	0.08
	200	100	0

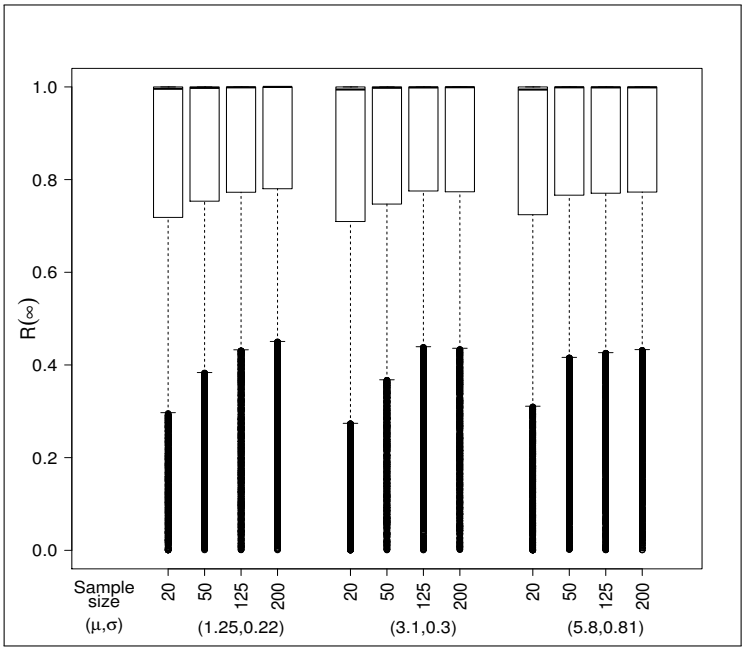


Figure 3: Box plots for three scenarios simulation Gumbel of the Table 3. Source: Elaborated by the authors.

re the MLE of α diverged to $-\infty$. These results show that in all scenarios, $\widetilde{R}_{\max}(\infty)$ assumes the value 1 with probabilities larger than 0.45. Thus, when data coming from Gumbel model, embedded in the three-parameter Weibull distribution, are modeled by a three-parameter Weibull model, flat likelihoods will occur very frequently. In order to analyze these samples in the graphical way previously presented, some box plots are displayed in Figure 3. These provide a clear idea of the shape of the frequency distribution of $\widetilde{R}_{\max}(\infty)$ for three of the simulation scenarios included in Table 3, showing that the flatness of the likelihood suggests that the embedded two-parameter model seems a reasonable model for the data.

It is important to note that once a sample size is set, for all the simulation scenarios shown in this section, the maximum computation time was two minutes, using a personal computer. On the other hand, all the simulations presented here were performed in the R software, version 4.0.3, and source code is available upon request.

Table 3: Classification (in percentage) of 10000 simulated samples of a Gumbel random variable.

(μ,σ)	k	$0 \leq \widetilde{R}_{\max}(\infty) < 1$	$\widetilde{R}_{\max}(\infty) = 1$
$(1.25, 0.22)$	20	53.40	46.60
	50	52.22	47.78
	125	51.72	48.28
	200	50	50
$(1.75, 0.22)$	20	54.38	45.62
	50	52.23	47.77
	125	51.70	48.30
	200	51.50	48.50
$(2.5, 0.22)$	20	53.02	46.98
	50	51.81	48.19
	125	51.88	48.12
	200	50.82	49.18
$(3.1, 0.3)$	20	53.83	46.17
	50	52.50	47.50
	125	51.75	48.25
	200	51.45	48.55
$(3.5, 0.3)$	20	53.81	46.19
	50	53.55	46.45
	125	51.54	48.46
	200	52.06	47.94
$(4, 0.3)$	20	53.14	46.86
	50	52.90	47.10
	125	52.22	47.78
	200	51.14	48.86
$(5.8, 0.81)$	20	53.91	46.09
	50	51.58	48.42
	125	51.85	48.15
	200	51.72	48.28
$(5.8, 1.04)$	20	53.44	46.56
	50	53.12	46.88
	125	52.23	47.77
	200	50.88	49.12
$(8.1, 1.6)$	20	53.15	46.85
	50	52.61	47.39
	125	51.71	48.29
	200	50.93	49.07

5. CONCLUSIONS

The cases discussed here were carefully selected to illustrate the potential information that flat likelihoods can provide, since their occurrence can be interpreted as a warning about an insufficient sample size for making appropriate and useful inferences about the threshold parameter, and the underlying relationship between the parameters of the model. Furthermore, flat likelihoods can also suggest that data may be coming from an embedded Gumbel model in the family of three-parameter Weibull distributions.

In general, dealing with flat likelihoods goes beyond finding stable estimators or solving numerical problems that arise during the optimization of the likelihood function, which can usually be addressed throughout an appropriate reparameterization. The flat likelihood problem embraces many issues and those should be carefully analyzed to understand the genesis of such a behaviour. The most valuable part of this paper relies on honing our intuition and understanding the multifactorial nature of this problem, contributing, in a certain way, to the scant literature concerning this subject.

In order to broaden the knowledge regarding the use of flat likelihoods in inference, future work is planned to analyze the advantages and disadvantages of using a Bayesian approach in these situations, as well as to examine the similarities and differences with the likelihood approach.

References

Barnard, G. A. (1967). The use of the likelihood function in statistical practice. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1, 27–40.

Barnard, G. A. & Sprott D. A. (1983). Likelihood. In: Kotz S, Johnson NL (eds) Encyclopedia of statistical science, Vol 4. Wiley, New York, pp 639–644.

Barndorff-Nielsen, O. E. & Cox, D. R. (1994). Inference and asymptotics. Chapman & Hall/CRC. Boca Raton.

Berger, J. O., Liseo, B. & Wolpert, R. L. (1999). Integrated likelihood methods for eliminating nuisance parameters. *Statistical Science*, 14(**1**), 1-28.

Bolívar-Cimé, A., Díaz-Francés, E. & Ortega, J. (2015). Optimality of profile likelihood intervals for quantiles of extreme value distributions: application to environmental disasters. *Hydrological Sciences Journal*, 60(**4**), 651-670.

Breusch, T. S., Robertson, J. C. & Welsh, A. H. (1997). The emperor's new clothes: a critique of the multivariate t regression model. *Statistica Neerlandica*, 51(**3**), 269-286.

Catchpole, E. A. & Morgan, B. J. (1997). Detecting parameter redundancy. *Biometrika*, 84(**1**), 187-196.

Cheng, R. C. H. & Iles, T. C. (1990). Embedded models in three-parameter distributions and their estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, 52(**1**), 135-149.

Cole, S.R., Chu, H. & Greenland, S. (2013). Maximum likelihood, profile likelihood, and penalized likelihood: a primer. *American Journal of Epidemiology*, 179(**2**), 252-260.

Cousineau, D., Goodman, V. W. & Shiffrin, R. M. (2002). Extending statistics of extremes to distributions varying in position and scale and the implications for race models. *Journal of Mathematical Psychology*, 46(**4**), 431-454.

De Haan, L. (1990). Fighting the arch-enemy with mathematics. *Statistica neerlandica*, 44(**2**), 45-68.

Deng, B., Jiang, D. & Gong, J. (2018). Is a three-parameter Weibull function really necessary for the characterization of the statistical variation of the strength of brittle ceramics?. *Journal of the European Ceramic Society*, 38(**4**), 2234-2242.

El Adlouni, S., Ouarda, T. B., Zhang, X., Roy, R. & Bobée, B. (2007). Generalized maximum likelihood estimators for the nonstationary generalized extreme value model. *Water Resources Research*, 43(**3**), W03410.

Elmahdy, E. E. & Aboutahoun, A. W.(2013). A new approach for parameter estimation of finite Weibull mixture distributions for reliability modeling. *Applied Mathematical Modelling*, 37(**4**), 1800-1810.

Elmahdy, E. E. (2015). A new approach for Weibull modeling for reliability life data analysis. *Applied Mathematics and computation*, 250, 708-720.

Farcomeni, A. & Tardella, L. (2012). Identifiability and inferential issues in capture-recapture experiments with heterogeneous detection probabilities. *Electronic Journal of Statistics*, 6, 2602-2626.

Frery, A. C., Cribari-Neto, F. & De Souza, M. O. (2004). Analysis of minute features in speckled imagery with maximum likelihood estimation. *EURASIP Journal on Advances in Signal Processing*, 2004(**16**), 2476-2491.

Ghosh, M., Datta, G. S., Kim, D. & Sweeting, T. J. (2006). Likelihood-based inference for the ratios of regression coefficients in linear models. *Annals of the Institute of Statistical Mathematics*, 58(**3**), 457-473.

Green, E. J., Roesch, F. A., Smith, A. F. & Strawderman, W. E. (1994). Bayesian Estimation for the Three-Parameter Weibull Distribution with Tree Diameter Data. *Biometrics*, 50(**1**), 254-269.

Harter, H. L. & Moore, A. H. (1966). Local-maximum-likelihood estimation of the parameters of three-parameter lognormal populations from complete and censored samples. *Journal of the American Statistical Association*, 61(**315**), 842-851.

Hirose, H. & Lai, T. L. (1997). Inference from grouped data in three-parameter Weibull models with applications to breakdown-voltage experiments. *Technometrics*, 39(**2**), 199-210.

Khan, H. M., Albatineh, A., Alshahrani, S., Jenkins, N. & Ahmed, N. U. (2011). Sensitivity analysis of predictive modeling for responses from the three-parameter Weibull model with a follow-up doubly censored sample of cancer patients. *Computational Statistics & Data Analysis*, 55(**12**), 3093-3103.

Kalbfleisch, J. G. (1985). Probability and Statistical Inference, Vol. 2. Springer-Verlag. New York.

Koutsoyiannis, D. (2004). Statistics of extremes and estimation of extreme rainfall: I. Theoretical investigation/Statistiques de valeurs extrêmes et estimation de précipitations extrêmes: I. Recherche théorique. *Hydrological Sciences Journal*, 49(**4**).

Kreutz, C., Raue, A., Kaschek, D. & Timmer, J. (2013). Profile likelihood in systems biology. *The FEBS journal*, 280(**11**), 2564-2571.

Li, R. & Sudjianto, A. (2005). Analysis of computer experiments using penalized likelihood in Gaussian Kriging models. *Technometrics*, 47(**2**), 111-120.

Lima, V. M. & Cribari-Neto, F. (2019). Penalized maximum likelihood estimation in the modified extended Weibull distribution. *Communications in Statistics-Simulation and Computation*, 48(**2**), 334-349.

Liu, S., Wu, H. & Meeker, W. Q. (2015). Understanding and addressing the unbounded "likelihood" problem. *The American Statistician*, 69(**3**), 191-200.

Martins, E. S. & Stedinger, J. R. (2000). Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resources Research*, 36(**3**), 737-744.

Martins, E. S. & Stedinger, J. R. (2001). Generalized maximum likelihood Pareto-Poisson estimators for partial duration series. *Water Resources Research*, 37(**10**), 2551-2557.

Montoya, J. A. (2008). La verosimilitud perfil en la Inferencia Estadística. Centro de Investigación en Matemáticas, A. C., Guanajuato, Gto., México.

Montoya, J. A., Díaz-Francés, E. & Sprott, D.A. (2009). On a criticism of the profile likelihood function. *Statistical Papers*, 50(**1**), 195-202.

Murphy, S. A. & Van Der Vaart, A. W. (2000). On profile likelihood. *Journal of the American Statistical Association*, 95(**450**), 449-465.

Pawitan, Y. (2001). In All Likelihood: Statistical Modelling and Inference Using Likelihood. Oxford University Press. New York.

Pewsey, A. (2000). Problems of inference for Azzalini's skewnormal distribution. *Journal of Applied Statistics*, 27(**7**), 859-870.

Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U. & Timmer, J. (2009). Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, 25(**15**), 1923-1929.

Serfling, R. J. (2002). Approximation Theorems of Mathematical Statistics. John Wiley & Sons. New York.

Silva, H. P. T. N. & Peiris, T. S. G.(2017). Statistical modeling of weekly rainfall: a case study in Colombo city in Sri Lanka. *Proceedings of the Engineering Research Conference (MERCon), Moratuwa, IEEE*, 241-246.

Smith, R. L. & Naylor, J. C. (1987). A comparison of maximum likelihood and Bayesian estimators for the three parameter Weibull distribution. *Journal of the Royal Statistical Society*, 36(**3**), 358–369.

Sprott, D. A. (2000). Statistical inference in science. Springer-Verlag. New York.

Sundberg, R. (2010). Flat and multimodal likelihoods and model lack of fit in curved exponential families. *Scandinavian Journal of Statistics*, 37(**4**), 632-643.

Tsionas, E. G. (2001). Likelihood and Posterior Shapes in Johnson's System. *Sankhyā: The Indian Journal of Statistics, Series B*, 63(**1**), 3-9.

Tumlinson, S. E. (2015). On the non-existence of maximum likelihood estimates for the extended exponential power distribution and its generalizations. *Statistics & Probability Letters*, 107, 111-114.

Revista de la Facultad de Ciencias Universidad Nacional de Colombia, Sede Medellín
V 11 N°2 julio-diciembre de 2022 • ISSN-e 2357-5749 • Research paper • Pages 54 a 73
DOI: https://doi.org/10.15446/rev.fac.cienc.v11n2.99967

FLAT LIKELIHOODS: SKEW NORMAL DISTRIBUTION CASE[a]

VEROSIMILITUDES PLANAS: CASO DE LA DISTRIBUCIÓN NORMAL SESGADA

JOSÉ A. MONTOYA[b*], GUDELIA FIGUEROA PRECIADO[c]

Recibido 06-12-2021, aceptado 10-03-2022, versión final 04-05-2022
Research paper

ABSTRACT: Several references argue in favor of alternative estimation methods, rather than the likelihood one, when the likelihood function exhibits flat regions. However, in the case of the skew normal distribution we present a discussion describing the interpretation of those flat likelihoods. This distribution is widely used in several interesting applications and contains the normal distribution as a nested model and the half-normal as an embedded model. Here, we show that flat likelihoods provide relevant information that should be carefully analyzed before discarding its use and proposing other estimation methods. Two well-known examples, that have been reported as troublesome, are analyzed here, including also an exhaustive computational study. The analysis of different scenarios allows to understand not only the reason of this likelihood function shape, but also to discover the information this behavior provides.
KEYWORDS: Flat likelihood function; parameter relationship; embedded models; likelihood contours; profile likelihood, simulation study.

RESUMEN: Diversas referencias argumentan a favor de métodos de estimación alternativos al de verosimilitud, cuando la función de verosimilitud exhibe regiones planas. Sin embargo, para el caso de la distribución normal sesgada se presenta una discusión donde se describe la interpretación de esas verosimilitudes planas. Esta distribución es ampliamente utilizada en diversas aplicaciones interesantes y tiene a la distribución normal como modelo anidado y a la distribución half-normal como modelo empotrado. Aquí se muestra que las verosimilitudes planas proporcionan información relevante que debe analizarse cuidadosamente, antes de abandonar su uso y proponer otros métodos de estimación. Se analizan dos ejemplos muy conocidos, que han sido reportados como problemáticos y se incluye también un estudio computacional exhaustivo. El análisis de diferentes escenarios permite comprender no sólo la razón de esta forma de la función de verosimilitud, sino también descubrir la información que este comportamiento proporciona.
PALABRAS CLAVE: Función de verosimilitud plana; relación entre parámetros, modelos empotrados, contornos de verosimilitud; verosimilitud perfil, estudio de simulación.

[a]Montoya, J. A. & Figueroa-Preciado, G. (2022). Flat likelihoods: Skew normal distribution case. *Rev. Fac. Cienc.*, 11 (**2**), 54–73. DOI: https://doi.org/10.15446/rev.fac.cienc.v11n2.99967

[b]PhD in Probability and Statistics. Statistics Professor. Department of Mathematics. University of Sonora, México.
[*]Corresponding author: arturo.montoya@unison.mx
[c]PhD in Probability and Statistics. Statistics Professor. Department of Mathematics. University of Sonora, México.

1. INTRODUCTION

Many data generated in real life processes often exhibit asymmetry, being then poorly described with symmetric distributions, like the popular and easy to use normal distribution. The need for study and analysis of skewed distributions is fundamental, and various proposals can be found in statistical literature, like those presented in Jones (2015). Here, we focus on a particular and appropriate distribution to analyze unimodal and skewed data, that is proposed by Azzalini (1985) in a well-known paper, where skew normal distribution is introduced; being this an extension of the normal distribution family, where a shape parameter is added in order to control skewness. In this way, this three parameter distribution includes the normal distribution, when considering that skewness parameter is equal to zero. Skew normal distribution allows to reflect different degrees of skewness, so it can be used in many interesting applications, like the one related to the analysis of microRNA data that can be found in Hossain & Beyene (2015), reliability studies applied to a strength-stress model, presented in Gupta & Brown (2001), and also addressing geostatistical problems, like the ones exposed in Allard & Naveau (2007), among many applications. On the other hand, some properties and characterizations of the skew normal distribution are discussed in Genton (2004), Gupta *et al.* (2004), Figueiredo & Gomes (2013), as well as in Azzalini (2005).

The skew normal density function is given by

$$f(x;\xi,\omega,\alpha) = \frac{2}{\omega\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\xi}{\omega}\right)^2\right) \Phi\left(\alpha\left(\frac{x-\xi}{\omega}\right)\right), \tag{1}$$

where $x \in \mathbb{R}$, $\alpha, \xi \in \mathbb{R}$, $\omega \in \mathbb{R}^+$, and

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz.$$

It is common to use the notation $X \sim SN(\xi,\omega,\alpha)$ when a random variable X has the density function presented in (1). In this parametrization α is the shape parameter, while ξ and ω are location and scale parameters, respectively. When $\alpha = 0$, a regular normal distribution $N\left(\xi,\omega^2\right)$ is obtained.

In some statistical literature, such as Azzalini & Arellano-Valle (2013) and Arellano-Valle & Azzalini (2008), two main inferential problems linked to skew normal distribution are exposed. One occurs when parameter $\alpha = 0$, since for any random sample $X_1, X_2, \cdots, X_n$ from (1), the profile log-likelihood function for α has an inflection point, and the expected information matrix becomes singular, even when the distribution is identifiable. Nevertheless, this singularity problem is due to parametrization of model (1). When the original parametrization (ξ,ω,α) is replaced for the centred one (μ,σ,γ_1), proposed by Azzalini (1985),

$$\mu = \xi + \omega\mu_\alpha, \quad \sigma^2 = \omega^2\left(1-\mu_\alpha^2\right), \quad \gamma_1 = \frac{4-\pi}{2}\left(\frac{\mu_\alpha}{\sqrt{1-\mu_\alpha^2}}\right)^3, \tag{2}$$

where $\mu_\alpha = \alpha\sqrt{2/\pi}/\sqrt{1+\alpha^2}$, then all these problems vanish (Azzalini & Arellano-Valle, 2013). It is important to emphasize that Rubio & Genton (2016) identified another problem related to the skew normal distribution, occurring when α is in an interval around zero, and although the skew normal is identifiable, it is hard to distinguish from a normal distribution. It is relevant to notice that parameter α is quite important, since it controls several features like the mode, the mean, the tail-weight and the spread of the distribution; so it is expected to require more data to be properly estimated.

On the other hand, Azzalini & Arellano-Valle (2013) and Arrué & Arellano-Valle & Gómez (2016) have pointed out another estimation problem that is also linked to the skew normal model. They found a non negligible probability that the maximum likelihood estimator (MLE) of the skewness parameter α goes to ∞ or $-\infty$, in samples of moderate sizes; that is, the MLE of α can become divergent. Azzalini & Arellano-Valle (2013) stated that if $X \sim SN(0,1,\alpha)$, then skew normal density (1), seen as a function of α, is proportional to the standard normal distribution $\Phi(\cdot)$ when this is evaluated in αx,

$$f(x;0,1,\alpha) \propto \Phi(\alpha x).$$

They also mentioned the fact that $\Phi(\cdot)$ is a monotonically increasing function, so then the maximum of the log-likelihood for an observed sample $x = x_i$ occurs if α goes to ∞ when $x_i > 0$, or if α goes to $-\infty$ when $x_i < 0$. They related this behavior with the probability of a divergent MLE, considering the fact that $P(X < 0) = 0.5 + \pi^{-1}\arctan\alpha$, when $\xi = 0$ and $\omega = 1$; then, the probability of a divergent MLE can be written as

$$p_{n,\alpha} = \left(\frac{1}{2} - \frac{\arctan\alpha}{\pi}\right)^n + \left(\frac{1}{2} + \frac{\arctan\alpha}{\pi}\right)^n. \tag{3}$$

Azzalini & Arellano-Valle (2013) reported that this probability goes rapidly to zero, when sample size n goes to ∞; but in the case of small or moderate sample sizes, there is a non negligible probability, ($p_{25,\alpha=5} \approx 0.197$ and $p_{50,\alpha=5} \approx 0.039$). These authors also mentioned that there is no characterization for those samples coming from a three parameter skew normal distribution $SN(\xi,\omega,\alpha)$, whose MLE diverges. It is important to note that the MLE is just one feature of the likelihood function shape, and what is truly fundamental is to understand the reason of that shape; only such a knowledge provides the basis for legitimate criticisms that could support new estimation methods, alternative computational optimizations, adequate reparametrizations, etcetera; such a knowledge could even provide practical clues for a model selection. Actually, as we previously stated, the skew normal singularity problem can be addressed with an adequate parametrization of the model presented in (1), as is explained in Pewsey (2000), and some references cited therein.

On the other hand, we are aware that there are some issues that deserve to be analyzed in depth, like the one related with a non negligible probability for the MLE of α of being divergent; conducting such a study may help to understand why the likelihood function for parameter α can result flat or even increasing over a wide range of parameter values. Now, although we consider important to quantify the frequency of cases where the MLE of α goes to ∞ or $-\infty$, in a skew normal with two unknown parameters, we consider even more important to understand the flatness of the likelihood function for α and its relationship with a half normal

embedded model problem.

In this manuscript we basically analyze the behavior of the shape of the likelihood function for parameter α, since it controls the shape in a skew normal model. In the analyses presented here we consider two cases: the first one concerns with a skew normal model, where only parameter α is considered unknown; this case is included just for the understanding of the simplest case. The second one is related to a skew normal model where α and ξ, shape and location parameters, are both unknown. We believe that these cases provide enough information for the understanding of the shape that takes the likelihood function in a skew normal model, since location parameter ξ is linked to the support of the distribution in the embedded half normal model.

In order to meet the objectives we have set out in this manuscript, we conducted some simulation studies. In the case of unknown parameter α, we analyze the limit of the relative likelihood function for α, when α goes to infinity. Now, when (ξ, α) are both unknown, we study the limit of the profile relative likelihood function for α, when α goes to infinity; in both cases we denote this limit as $R(\infty)$. When only parameter α is unknown, simulation results showed that skew normal samples, with no negative values lead to an increasing likelihood function for α, reaching its maximum when α goes to ∞, $R(\infty) = 1$, and there is a large probability that a hypothesis test for an embedded half normal model will not be rejected. On the other hand, when a skew normal sample has at least one negative observation, then a half normal model is not plausible, $R(\infty) = 0$, result that any reasonable inferential method should yield, since the probability of assuming a negative value for a half normal variable is zero. Furthermore, mathematically, the likelihood function for α turns out informative and provides reliable confidence intervals.

In the case of unknown α and ξ parameters, our simulation results indicated that those skew normal samples leading to profile likelihood functions for α whose maximum is reached at infinity, $R(\infty) = 1$, there is a large probability of not rejecting a hypothesis of an embedded shifted half normal model (shifted ξ units). Conversely, when $0 \leq R(\infty) < 1$, the relationship between parameter α value and sample size n, plays a fundamental role for determining the flatness of the likelihood function. For instance, when $\alpha \leq 3$ and $n \geq 200$, it is possible to construct 95 % likelihood-confidence intervals with a finite upper bound, since most of the times $R(\infty) \approx 0$. Again, for any of these skew normal samples, the shape of the likelihood function for parameter α is informative and provides reasonable inferences.

In Section 2 we introduce the likelihood function for a skew normal model where three parameters are unknown; presenting also the relative and profile likelihood functions for a specific parameter of interest in this model. Section 3 includes two illustrative examples, taken from statistical literature. The first one allows us to detect when centred parametrization can yield a well-behaved likelihood function, and the second one enables us to identify cases where the relative profile likelihood for α grows until reaching its maximum, decreasing later and staying essentially flat; features that are revealed, in a different way, when using centred

parametrization. A more in-depth exploration about the flatness of the likelihood function is performed in Section 4, where two simulation studies are included. Finally, some important conclusions are presented in Section 5.

2. SKEW NORMAL LIKELIHOOD

In this section we define the likelihood function and the profile and relative likelihood functions for the general $SN(\xi,\omega,\alpha)$ case, where all three parameters are unknown. The cases where only α is unknown or (ξ,α) are both unknown, discussed in this manuscript, can be easily obtained just fixing the real values of the remaining parameters. Definitions included in this manuscript are useful and valid for the purposes described here, but more details regarding the theoretical likelihood framework can be found in Barndorff-Nielsen (1988) and Pawitan (2001), just to mention some.

Suppose $X_1, X_2, \cdots, X_n$ are i.i.d. random variables, skew normal distributed with unknown parameters (ξ,ω,α). Let $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)$ be realization of this sample. The likelihood function of the parameters vector (ξ,ω,α), based on the observed sample $\boldsymbol{x}$ and in model (1), can be written as

$$L(\xi,\omega,\alpha) \propto \omega^{-n} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\xi}{\omega}\right)^2\right)\prod_{i=1}^{n}\Phi\left(\alpha\left(\frac{x_i-\xi}{\omega}\right)\right), \tag{4}$$

where $\alpha,\xi \in \mathbb{R}$, and $\omega \in \mathbb{R}^+$. For notation simplicity, the dependence of L on the sample is not made explicit.

When considering the centred reparametrization given in (2), then the likelihood function of the parameter vector (μ,σ,γ_1), can be written as:

$$L(\mu,\sigma,\gamma_1) \propto L\left(\xi = \xi_{\mu,\sigma,\gamma_1}, \omega = \omega_{\sigma,\gamma_1}, \alpha = \alpha_{\gamma_1}\right), \tag{5}$$

where

$$\xi_{\mu,\sigma,\gamma_1} = \mu - \sigma c_{\gamma_1}, \quad \omega_{\sigma,\gamma_1} = \sigma\sqrt{1+c_{\gamma_1}^2}, \quad \alpha_{\gamma_1} = \frac{c_{\gamma_1}}{\sqrt{b^2-(1-b^2)\,c_{\gamma_1}^2}},$$

$b = \sqrt{2/\pi}$, and $c_{\gamma_1} = [2\gamma_1/(4-\pi)]^{1/3}$. It is important to note that throughout this manuscript, we use both, direct parametrization (ξ,ω,α) of the skew normal model, as well as centred reparametrization (μ,σ,γ_1), in order to exemplify and analyze the likelihood shape. Nevertheless, it will be sufficient to define the relative and profile likelihood functions for the centred parametrization.

The relative likelihood function of (μ,σ,γ_1) is a standardized version of (5), that takes the value of one at its maximum, the MLE of (μ,σ,γ_1),

$$R(\mu,\sigma,\gamma_1) = \frac{L(\mu,\sigma,\gamma_1)}{\max\limits_{\mu,\sigma,\gamma_1} L(\mu,\sigma,\gamma_1)} = \frac{L(\mu,\sigma,\gamma_1)}{L(\widehat{\mu},\widehat{\sigma},\widehat{\gamma_1})}, \tag{6}$$

where $\widehat{\mu}$, $\widehat{\sigma}$, and $\widehat{\gamma_1}$ are the MLE's of parameters μ, σ, and γ_1. Note that the likelihood function for (ξ, ω, α) is obtained just replacing in (6) the function $L(\mu, \sigma, \gamma_1)$ by $L(\xi, \omega, \alpha)$ given in (4).

Assume that the full parameter (μ, σ, γ_1) can be partitioned into an interest parameter ψ and a nuisance parameter λ. Then, the profile likelihood and its corresponding relative likelihood function of ψ, standardized to be one at the maximum of the likelihood function, are defined as

$$L_{\max}(\psi) = \max_{\lambda} L(\psi, \lambda),$$
$$R_{\max}(\psi) = \frac{L_{\max}(\psi)}{\max_{\psi,\lambda} L(\psi, \lambda)} = \frac{L_{\max}(\psi)}{L_{\max}(\widehat{\psi})}.$$

For example, the profile likelihood and its corresponding relative likelihood function of $\psi = \gamma_1$ is

$$L_{\max}(\gamma_1) = \max_{\mu,\sigma} L(\mu, \sigma, \gamma_1), \tag{7}$$
$$R_{\max}(\gamma_1) = \frac{L_{\max}(\gamma_1)}{\max_{\mu,\sigma,\gamma_1} L(\alpha, \gamma, \beta)} = \frac{L_{\max}(\gamma_1)}{L_{\max}(\widehat{\gamma_1})}. \tag{8}$$

Now, it is easy to note that profile likelihood of $\psi = \alpha$ (and its corresponding relative likelihood) can be obtained just substituting in (7-8) the function $L(\mu, \sigma, \gamma_1)$ by the function $L(\xi, \omega, \alpha)$ given (4).

In general, the profile or maximized likelihood is a powerful though simple inferential method for estimating a parameter of interest separately from the remaining unknown parameters of the statistical model. More details about the profile likelihood approach are described in Kalbfleisch (1985, Section 10.3), Pawitan (2001, Section 3.4), Sprott (2000, p. 66), Barndorff-Nielsen & Cox (1994, pp. 89-91), Serfling (2002, pp. 155-160), and Murphy & Van Der Vaart (2000). Thus, the profile likelihood function may be used to study and visualize various aspects of a full likelihood function, such as for instance those associated with the flatness of the likelihood surface, as well as the intrinsic relationship between those parameters.

3. EXAMPLES: SKEW NORMAL LIKELIHOOD

In this section we analyze two data sets. In Example 1 we show that centred reparametrization (2) proposed by Azzalini (1985) allows to symmetrize the likelihood function, when data come from a skew normal model where α is closed to zero; so this parameter space transformation favors inferences about model parameters. However, as will be shown in Example 2, when data come from a skew normal model where α is far away from zero, this skew normal model reparametrization generates likelihood functions whose shapes reveal features linked to flat or monotone increasing likelihoods, but those features emerge in a different way.

3.1. Example 1

In this example we construct profile likelihood functions based on a simulated random sample $\boldsymbol{x_1} = (x_1, x_2, \cdots, x_n)$ of size $n = 202$, from a skew normal model $X \sim SN(64.07, 17.68, 0)$, using

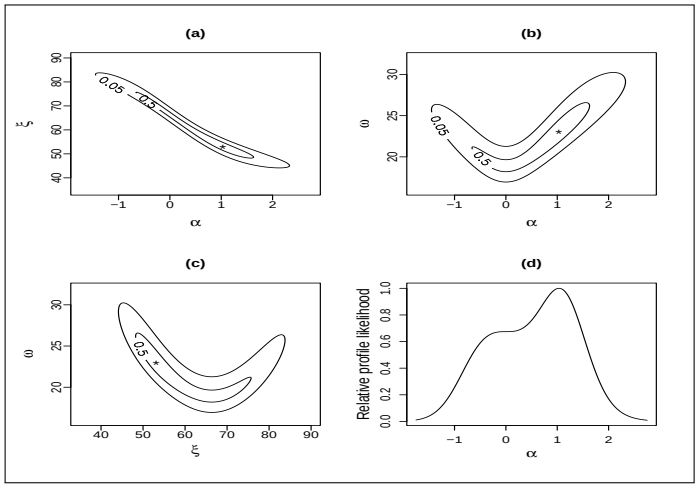


Figure 1: Unknown three parameters ξ, ω, and α: Relative profile likelihood functions for direct parameters corresponding to the simulated sample $\boldsymbol{x_1}$ with size $n = 202$ from a skew normal with parameters $\xi = 64.07$, $\omega = 17.68$, and $\alpha = 0$. Source: Elaborated by the authors.

R software version 4.0.4, and using a seed in order to make it replicable in case of interest: set.seed(as.numeric(as.Date("2021-09-03"))). Sample size selection and parameter values for ξ and ω were selected according to the sample size and the MLE's that were obtained with the AIS, (Australian Institute of Sport) weight data (in kg); see Arellano-Valle & Azzalini (2008). AIS data contain biomedical measurements on 202 Australian athletes and it is available in the library *sn* in R. Shape parameter was fixed in $\alpha = 0$, just for comparing the likelihood shape when direct (ξ, ω, α) or centred (μ, σ, γ_1) reparametrization are used.

Based on simulated $\boldsymbol{x_1}$ data, Figures 1-2 show the profile likelihood contours for parameters in the $SN(\xi, \omega, \alpha)$ and $SN(\mu, \sigma, \gamma_1)$ models, respectively, considering all parameters unknown. As can be observed in Figure 1, for this simulated sample, direct reparametrization (ξ, ω, α) generates elongated likelihood contours, sometimes like a boomerang shape. In this case, profile likelihood function for parameter α is not flat; it is bimodal, with the highest hump corresponding to the MLE location. Everything seems to indicate that this skew normal parametrization brings to light a model parameters relationship. On the other hand, Figure 2 leads to what could be called well-behaved and nearly symmetric likelihoods, favoring simple, efficient and reasonable inferences.

3.2. Example 2

This example is based on sample size and parameter values considered in Frontier data, where a set of $n = 50$ values were sampled from $SN(0, 1, 5)$; this dataset is presented by Azzalini & Capitanio (1999). It is well-known that the MLE of α, in this dataset, diverges when a skew normal model with unknown parameters (ξ, ω, α) is considered; see Azzalini & Arellano-Valle (2013).

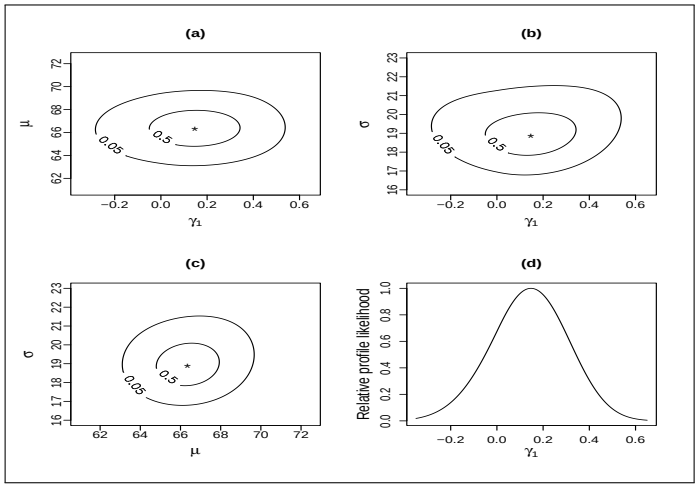


Figure 2: Unknown three parameters μ, σ, and γ_1: Relative profile likelihood functions for centred parameters corresponding to the simulated sample $\boldsymbol{x_1}$ with size $n = 202$ from a skew normal with parameters $\xi = 64.07$, $\omega = 17.68$, and $\alpha = 0$. Source: Elaborated by the authors.

In view of the foregoing, we construct some profile likelihood functions based on a sample $\boldsymbol{x_2} = (x_1, x_2, \cdots, x_n)$ of size $n = 50$, that was simulated from a skew normal random variable, $X \sim SN(0,1,5)$ using the R software. The sample can be easily replicated by setting a seed with the instruction set.seed(as.numeric(as.Date("2021-09-06"))). This sample allows us to show a relative profile likelihood for α, that grows until it reaches its maximum and then decreases, staying essentially flat.

In the case of a $SN(\xi, \omega, \alpha)$ model with all parameters unknown, Figure 3 shows the shape of different relative profile likelihood contours when direct parametrization (ξ, ω, α) is used. It can be seen that the contour plot, associated to the relative profile likelihood of (ξ, ω) based on the observed sample $\boldsymbol{x_2}$, estimates quite well some plausible regions for these parameters. However, contour plots associated to the relative profile likelihood functions of (ξ, α) and (ω, α), lengthen as α goes to infinity. Actually, relative profile likelihood of α seems flat for values greater than 40. It is important to highlight that a similar behavior is observed when considering a $SN(\xi, 1, \alpha)$ model, where just ξ and α, the location and shape parameters are unknown; see Figure 4.

Finally, Figure 5 shows the profile likelihood function of γ_1, based on our 50 simulated observations contained in $\boldsymbol{x_2}$, and also the one that is based on Frontier's data, a set of $n = 50$ values sampled from $SN(0,1,5)$. For these two samples we assume a skew normal model with ξ and α unknown, when using direct parametrization, and μ y γ_1 are treated as unknown, in case of a centred parametrization. In this figure we can observe that the MLE of γ_1, in Frontier data, is reached at the right end of the support of the parameter space for γ_1, given by $\sqrt{(2)}\,(4-\pi)\,/\,(\pi-2)^3\,/2$, and occurring when $\alpha \to \infty$. As we previously stated, this is also the case in a skew normal model with three unknown parameters. On the other hand, for our simulated data, the set of γ_1 values with a relative profile likelihood greater than 0.15, $\{\gamma_1 : R_{\max}(\gamma_1) \geq 0.15\}$, provides a

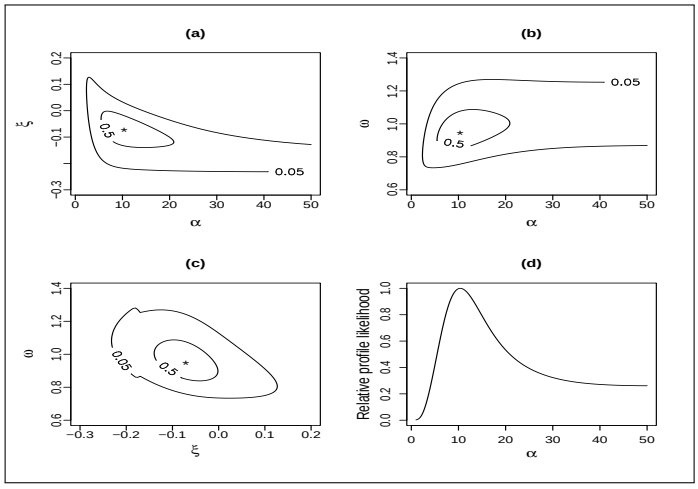


Figure 3: Unknown three parameters ξ, ω, and α: Relative profile likelihood functions for direct parameters corresponding to the simulated sample $\boldsymbol{x_2}$ with size $n = 50$ from a skew normal with $\xi = 0$, $\omega = 1$, and $\alpha = 5$. Source: Elaborated by the authors.

95 % likelihood-confidence interval for γ_1. This includes the right end of the support of the parameter space for γ_1, associated to $\alpha \to \infty$, that is linked to a well-defined model, as this simply corresponds to the half normal distribution. Thus, this does not represent a degenerate case, but a well-defined model.

It is interesting to deeply analyze this example, since it reveals that some features linked to flat or monotone increasing likelihoods for shape parameter α, in the skew normal model, do not vanish when using centred parametrization; actually, those features are inherited and revealed in a different way.

Next section is devoted to analyze, in greater detail, this flat likelihood phenomenon; taking into account the relationship between this likelihood flatness and the closeness between skew normal models and their embedded models.

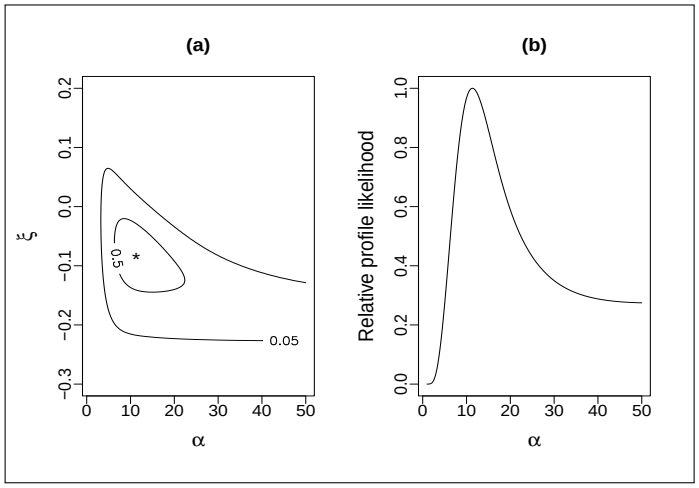


Figure 4: Unknown two parameters ξ and α: Relative profile likelihood functions for direct parameters corresponding to the simulated sample $\boldsymbol{x_2}$ with size $n = 50$ from a skew normal with $\xi = 0$, $\omega = 1$, and $\alpha = 5$. Source: Elaborated by the authors.

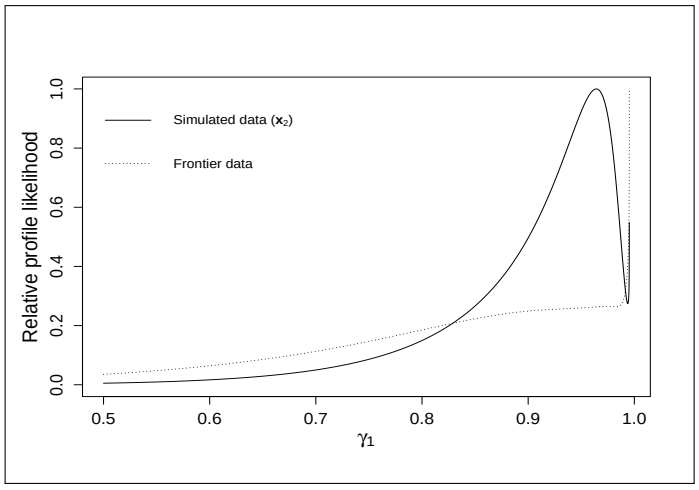


Figure 5: Two unknown parameters: μ and γ_1: Relative profile likelihood functions for γ_1 corresponding to the simulated sample $\boldsymbol{x_2}$, with size $n = 50$ from a skew normal with parameters $\xi = 0$, $\omega = 1$, and $\alpha = 5$, and Frontier data, a set of $n = 50$ values sampled from $SN(0,1,5)$. Source: Elaborated by the authors.

4. FLATNESS OF THE RELATIVE PROFILE LIKELIHOOD OF α

As mentioned by Sprott (2000, p. 11), the MLE and the observed information exhibit two features of the likelihood function; the former is a measure of its location relative to the parameter-axes, while the latter is a measure of its curvature in a neighborhood of the MLE. Thus, when the likelihood function of a scalar parameter is smooth and symmetric with respect to the MLE, then the MLE and the observed information are usually the main features determining the shape of the likelihood function. It should be stressed that flat likelihoods can not be determined by these quantities, without loss of information.

Intuitively, the sensitivity of the MLE depends on the shape of the likelihood function. If the likelihood of α is very curved around MLE, then the MLE results a stable estimate of α. On the other hand, if the likelihood is flat near the MLE, then this becomes highly unstable. However, more important than those facts is the one related with likelihood-confidence intervals of α arising from a flat likelihood, where upper limit can become infinity.

In this section we study, throughout computer simulations, the degree of practical parameter non-identifiability of parameter α, in the sense explained in Cole (2020, p. 38) that "*a parameter is practically non-identifiable if the log-likelihood has a unique maximum, but the parameters's likelihood-based confidence region tends to infinity in either or both directions*". Here, we study the frequency of the occurrence of flat likelihoods of shape parameter α by exploring the limit of the relative likelihood of α, as α approaches infinity. First of all, we analyze the simplest case when only parameter α is unknown, considering later a skew normal model with all parameters unknown. Always having in mind the existence of an embedded half normal model in this flat likelihood problem, our scenarios were designed considering the closeness between the skew normal and half normal models. Total variation distance (TV), as exposed in Huber (1981, p. 34), Jurečková (2006, p. 12) and Rieder (1994), was used to measure the closeness between these distributions and computed through the function TotalVarDist included in the R library *distrEx*. The probability of divergence of the MLE of α, provided in (3), was also measured.

For all simulated samples where an increasing likelihood function pattern ($\widehat{\alpha} \to \infty$) was detected, a p-value associated to a Kolmogorov-Smirnov distribution-free test was obtained; this procedure makes possible to test for general differences in two distributions, as it is explained in Hollander et al. (2014). Resulting p-values allow to quantify the extent to which these samples provide evidence against an embedded half normal model. Note that no p-value computation is needed when the divergence condition for the MLE of α is not met (al least one negative observation), since no negative values are allowed in a half normal model.

4.1. The unknown shape parameter case

To illustrate the degree of practical parameter non-identifiability of the shape parameter in the skew normal model, given a sample size, we will focus on the limit of the relative likelihood function of α as α approaches

infinity,

$$R(\infty) = \lim_{\alpha\to\infty} R(\alpha) = \lim_{\alpha\to\infty} \frac{L(\xi=0,\omega=1,\alpha)}{\max\limits_{\xi=0,\omega=1,\alpha} L(\xi,\omega,\alpha)}, \tag{9}$$

where $L(\xi,\omega,\alpha)$ is given in (4). Note that centred parametrization (2) yields

$$\mu = \mu_\alpha, \quad \sigma^2 = \left(1-\mu_\alpha^2\right), \quad \gamma_1 = \frac{4-\pi}{2}\left(\frac{\mu_\alpha}{\sqrt{1-\mu_\alpha^2}}\right)^3,$$

where $\mu_\alpha = \alpha\sqrt{2/\pi}/\sqrt{1+\alpha^2}$. In this way, as $\alpha\to\infty$, $\mu_\alpha \to \sqrt{2/\pi}$, and then

$$\mu = \sqrt{\frac{2}{\pi}}, \quad \sigma^2 = 1-\frac{2}{\pi}, \quad \gamma_1 = \frac{\sqrt{2}\,(4-\pi)}{(\pi-2)^{3/2}}. \tag{10}$$

Moreover, with this reparametrization $\gamma_1 \in [-\gamma_{10},\gamma_{10}]$, where $\gamma_{10} = \sqrt{(2)}\,(4-\pi)\,/\,(\pi-2)^3\,/2$; this can be interpreted as computationally favorable when searching the MLE of γ_1. For all the foregoing considerations, in our computations we use the following expression for $R(\infty)$, based on (9):

$$R(\infty) = \begin{cases} 1, & \text{if all } x_i > 0, \\ \dfrac{L\left(\mu=\sqrt{2/\pi},\sigma=1-2/\pi,\gamma_1=\gamma_{10}\right)}{\max\limits_{\gamma_1\in[-\gamma_{10},\gamma_{10}]} L\left(\mu=\mu_{\gamma_1},\sigma=\sigma_{\gamma_1},\gamma_1\right)}, & \text{in another case,} \end{cases} \tag{11}$$

where $\mu_{\gamma_1} = c_{\gamma_1}/\sqrt{1+c_{\gamma_1}^2}$ and $\sigma_{\gamma_1} = 1/\sqrt{1+c_{\gamma_1}^2}$.

If $R(\infty)$ is smaller than one, but close to it, then for every real number $\varepsilon > 0$ there exists a real number α_0 such that, for all $\alpha > \alpha_0$, we have $R(\alpha) \in (R(\infty)-\varepsilon, R(\infty)+\varepsilon)$. Thus, there is a set of values of α that makes the observed sample as likely as does the MLE. The frequency of occurrence of this type of behavior represents an index of the degree of practical parameter non-identifiability. In particular, when all parameters are unknown, $R(\infty) = 1$ can be interpreted as an extreme case of a practical parameter non-identifiability of the shape parameter in the skew normal model. This situation could suggest that dataset may be fitted with an embedded half normal model, whose parameters are given in (10).

In order to select adequate simulation scenarios that allow us to analyze the behavior of $R(\infty)$, and also its relationship with an embedded half normal model, obtained when α goes to infinity, we focus on the closeness between the skew normal and half normal models. As we previously stated we use the total variation distance, TV, as a measure of the closeness between these distributions. Figure 6a shows the plots for skew normal densities with shape parameter values $\alpha = 0, 5, 50$, and $\alpha\to\infty$; while in Figure 6b we plot TV as a function of α, in order to easily identify TV distances between the densities shown in Figure 6a. Particularly, in Figure 6b we can observe a $TV = 0.5$ between normal density ($\alpha = 0$) and half normal density, difference that is visually evident from Figure 6a, where we can observe that a skew normal with $\alpha = 50$ seems very close to half normal model and $TV = 0.006365677$, quite smaller than $TV = 0.5$. Now, TV distance between

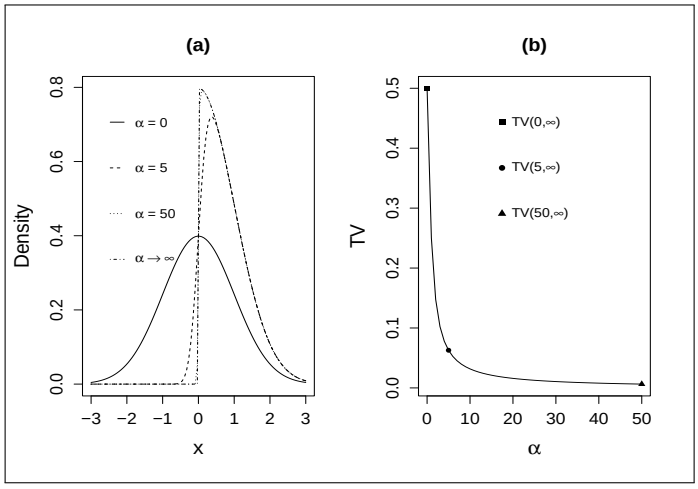


Figure 6: (a) Skew normal densities with shape parameter values $\alpha = 0, 5, 50$ and $\alpha \to \infty$. (b) TV as a function α, measuring closeness between skew normal and an embedded half-normal model. In this plot, TV distances are obtained considering $\alpha = 0, 5, 50$ versus $\alpha = \infty$, according to densities plotted in (a). Source: Elaborated by the authors.

Table 1: Percentages $100p_{n,\alpha}$ of divergence of the MLE of α, for simulation scenarios.

n	TV			
	0.10 ($\alpha = 3$)	0.06 ($\alpha = 5$)	0.03 ($\alpha = 11$)	0.006 ($\alpha = 50$)
25	6.71	19.74	48.09	85.24
50	0.45	3.90	23.13	72.67
100	0	0.15	5.35	52.80
200	0	0	0.29	27.88

a skew normal model where $\alpha = 5$ is closer from a half normal model ($\alpha \to \infty$), than the one obtained when $\alpha = 0$.

Considering the cases previously described, we now simulate independent random samples from a skew normal random variable, $X \sim SN(0, 1, \alpha)$, setting α values in $3, 5, 11$, and 50; these values yield TV distances of $0.1, 0.06, 0.03$, and 0.006, respectively. On the other hand, we select sample sizes of $n = 25$, 50, 100, and 200, in order to have a wide variety for the probability of divergence of the MLE of α. These values are computed throughout $p_{n,\alpha}$ given in (3), and are shown in Table 1.

Table 2 shows the classification of 5000 simulated samples, for each of the selected scenarios, according to whether $0 \leq \widetilde{R}_{\max}(\infty) < 1$ or $\widetilde{R}_{\max}(\infty) = 1$. Note that $\widetilde{R}_{\max}(\infty) = 1$ corresponds to those samples whose MLE of α diverged to ∞. Actually, empirical frequencies corresponding to this column in Table 2, are similar to the percentages shown in Table 1, obtained from the theoretical divergence probabilities. Moreover, all samples considered in this column have positive observations, and at least 95.55 % of the p-values yielded by a Kolmogorov-Smirnov test, under the assumption of an embedded half normal model, were greater than 0.05. Thus, at 0.05 significance level, a half normal model will not be rejected approximately 95 % of the

Table 2: Classification (in percentage) of 5000 simulated samples $X \sim SN(0,1,\alpha)$.

α	n	$0 \leq \widetilde{R}_{\max}(\infty) < 1$	$\widetilde{R}_{\max}(\infty) = 1$
3	25	93.38	6.62
	50	99.66	0.34
	100	100	0
	200	100	0
5	25	80.28	19.72
	50	96.20	3.80
	100	99.82	0.18
	200	100	0
11	25	51.24	48.76
	50	77.24	22.76
	100	94.74	5.26
	200	99.68	0.32
50	25	15.06	84.94
	50	25.94	74.06
	100	46.28	53.72
	200	72.24	27.76

times, regardless of the α and n values considered in Table 1.

On the other hand, when samples had at least one negative observation, and divergence condition for the MLE of α is not met, for any of the considered scenarios we found that $R(\infty) \in \left[0, 1.691853 \times 10^{-316}\right]$; that is, the right tail of the likelihood functions for α and γ_1, decreases to practically zero. For the particular case when $TV = 0.006$ and $n = 25$, Figure 7 shows the enveloping curves for the relative likelihood functions of γ_1, where simulated samples had at least one negative observation (753 samples, 15.06 %). In this figure we can observe that all relative likelihoods for γ_1 reached the value $R(\infty) = 0$ at $\gamma_1 = \gamma_{10}$. Now, Figure 8 shows the enveloping curves when considering a large TV and also a large sample size ($TV = 0.1$, $n = 200$), in samples with at least one negative observation (5000 samples, 100 %); so likelihood-confidence intervals for γ_1 or α, at typical confidence levels (90 %, 95 %, and 99 %), yield finite upper bounds, since $R(\infty) = 0$.

In summary, simulation results show that when skew normal samples have no negative observations, likelihood function grows until it reaches its maximum at $\alpha = \infty$, being highly probable that an embedded half normal model will not be rejected under a hypothesis test. Now, in case of at least one negative observation, an embedded half normal model results not plausible ($R(\infty) = 0$). Furthermore, mathematically, likelihood yields reliable confidence intervals, sometimes with endpoints $\left(A_{\gamma_1}, \gamma_{10}\right)$, or equivalently (A_α, ∞). Thus, for any of these skew normal samples, likelihood shape resulted informative, providing reasonable inferences.

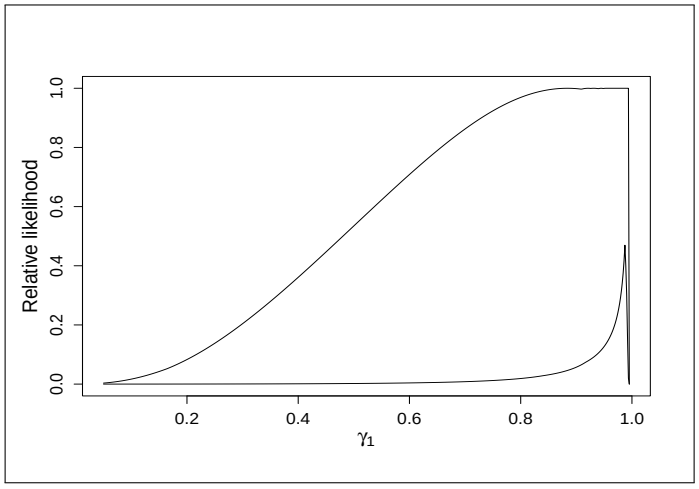


Figure 7: Enveloping curves for relative likelihood functions of γ_1 for simulated skew normal samples where: $TV = 0.006$, $\alpha = 50$ and $n = 25$. Source: Elaborated by the authors.

4.2. The case of unknown shape and location parameters

In the case that parameters (ξ, α) are unknown, we will focus on the limit of the relative profile likelihood function of α as α approaches infinity,

$$R(\infty) = \lim_{\alpha\to\infty} R_{\max}(\alpha) = \lim_{\alpha\to\infty} \frac{L_{\max}(\xi, \omega = 1, \alpha)}{\max\limits_{\xi,\omega=1,\alpha} L(\xi, \omega, \alpha)}, \tag{12}$$

where $L(\xi, \omega, \alpha)$ is given in (4). Note that, in this case, centred parametrization results

$$\mu = \xi + \mu_\alpha, \quad \sigma^2 = \left(1 - \mu_\alpha^2\right), \quad \gamma_1 = \frac{4-\pi}{2}\left(\frac{\mu_\alpha}{\sqrt{1-\mu_\alpha^2}}\right)^3.$$

Then, as $\alpha \to \infty$, $R(\infty)$ can be written as:

$$R(\infty) = \frac{\max\limits_{\xi\in\mathbb{R}} L\left(\mu = \xi + \sqrt{2/\pi}, \sigma = 1 - 2/\pi, \gamma_1 = \gamma_{10}\right)}{\max\limits_{\xi\in\mathbb{R}, \gamma_1\in[-\gamma_{10},\gamma_{10}]} L\left(\mu = \mu_{\xi,\gamma_1}, \sigma = \sigma_{\gamma_1}, \gamma_1\right)}, \tag{13}$$

where $\mu_{\xi,\gamma_1} = \xi + c_{\gamma_1}/\sqrt{1 + c_{\gamma_1}^2}$ and $\sigma_{\gamma_1} = 1/\sqrt{1 + c_{\gamma_1}^2}$.

In order to analyze the case where shape and location parameters are unknown, and also for being able of comparing new results with those obtained in previous section, same samples and scenarios will be considered in this section. This makes possible to computationally evaluate the profile likelihood of α when only parameter α is unknown, and also the case when parameters (ξ, α) are both unknown. We again compute the p-values that a Kolmogorov-Smirnov test yields, when an embedded shifted half normal model is assumed,

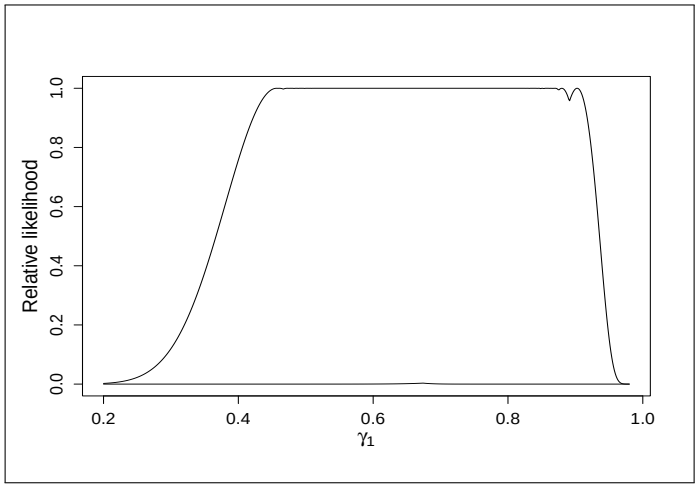


Figure 8: Enveloping curves for relative likelihood functions of γ_1, for simulated skew normal samples where: $TV = 0.1$, $\alpha = 3$ and $n = 200$. Source: Elaborated by the authors.

in order to quantify the occurrence of flat relative profile likelihoods for parameter α.

Table 3 shows a classification of the simulated samples, according to whether $0 \leq \widetilde{R}_{\max}(\infty) < 1$ or $\widetilde{R}_{\max}(\infty) = 1$. Note that $\widetilde{R}_{\max}(\infty) = 1$ corresponds to samples where the MLE of α diverged to ∞. It is important to mention that those scenarios where at least 10 % of their samples fell in category $\widetilde{R}_{\max}(\infty) = 1$, (five hundred or more of their samples), at least 95.14 % of them yielded a p-value greater than 0.05 in a Kolmogorov-Smirnov test, providing no evidence against a shifted half normal model. Thus, at 0.05 significance level, a shifted half normal model is not rejected in a Kolmogorov-Smirnov test, 95 % of the times, independently of sample size n and the α value considered in the simulation scenario; see Tabla 1. Those scenarios with few samples in category $\widetilde{R}_{\max}(\infty) = 1$ were excluded from this analysis, in order to obtain a reliable estimate for the frequency of p-values smaller than 0.05.

On the other hand, when divergence of the MLE of α is not met, the limit of the relative profile likelihood of α, denoted here as $R(\infty)$, has an interesting behavior. For small sample sizes it is highly probable to observe flat likelihoods, but when sample sizes are large, frequency distribution is located and concentrated close to zero. All these can be observed in Figure 9. Scenarios where $TV = 0.006$ ($\alpha = 50$) are not included in this figure, since the limited number of $R(\infty)$ values could yield to non-reliable comparisons. Nevertheless, similar behavior of $R(\infty)$ distribution has been observed in many performed simulations, but due to space limitation are not included here.

In summary, simulation results showed that in a skew normal sample whose relative profile likelihood of α reaches its maximum at infinity, there is a large probability that an embedded shifted half normal model will not be rejected in a hypothesis test. On the other hand, when the relative profile likelihood of α is smaller

Table 3: Classification (in percentage) of 5000 simulated samples $X \sim SN(0,1,\alpha)$.

α	n	$0 \leq \widetilde{R}_{\max}(\infty) < 1$	$\widetilde{R}_{\max}(\infty) = 1$
3	25	64.18	35.82
	50	91.66	8.34
	100	99.56	0.44
	200	100	0
5	25	37.76	66.24
	50	72.86	27.14
	100	95.04	4.96
	200	99.86	0.14
11	25	10.48	89.52
	50	32.30	67.70
	100	65.48	34.52
	200	92.28	7.72
50	25	2.72	97.28
	50	3.94	96.06
	100	9.24	90.76
	200	26.26	73.74

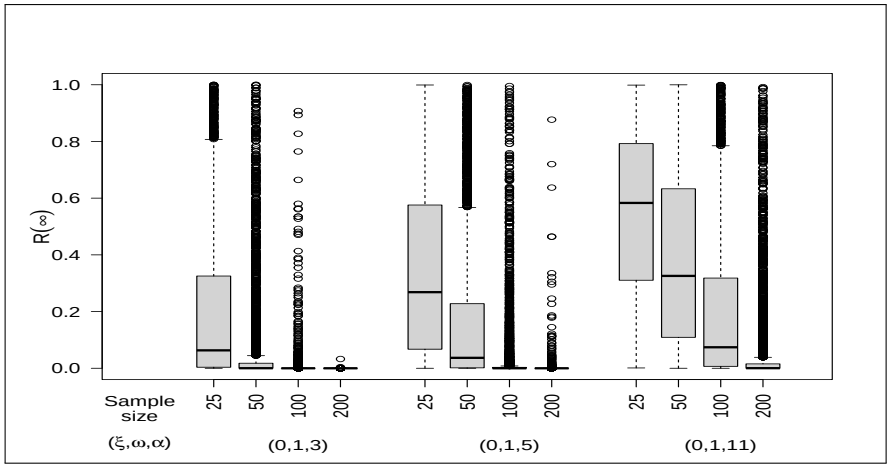


Figure 9: Box plots for three of the skew normal simulation scenarios presented in the Table 1. Source: Elaborated by the authors.

than 1, the relationship between α and n will determine the degree of flatness of the likelihood function. For instance, in Figure 9, for the case where $\alpha \leq 3$ y $n \geq 200$, it is possible to construct 95% likelihood-confidence intervals with a finite upper bound, since most of the times $R(\infty) \approx 0$; thus, for any of these skew normal samples, likelihood shape is informative and provides reasonable inferences.

5. CONCLUSIONS

Possible conditions of the divergence of the MLE for shape parameter in a skew normal model were studied here, throughout the analysis of its likelihood and profile likelihood function. The two particular cases included in this manuscript and the exhaustive simulation study that was carried out, allowed us to analyze the behavior of the shape of the likelihood function for this parameter. The occurrence of flat likelihoods for shape parameter, where $R(\infty) = 1$, suggested a practical parameter non-identifiability of a skew normal model, observing than an embedded half normal model, or a shifted one, are usually not rejected under a hypothesis test. Some other results presented here, show the fundamental role between shape parameter values and sample size, for determining the flatness of the likelihood function.

As we have shown in this work, flatness of a likelihood function is informative and should not be ignored, much less neglected. The cases discussed in this manuscript reveal, for certain kind of reparametrization, practical non-identifiability problems that can be caused by embedded models not rejected by the data. Flat likelihood regions can also occur when data is unable to provide enough information, either to distinguish nested models or to adequately estimate parameters involved in the distribution under study. The analysis of the shape of the likelihood function and particularly its flatness yields valuable information that usually uncover aspects that must be carefully analyzed. This issue is particularly important when inferential methods are carried out in family models involving nested or embedded models and containing many parameters and covariables.

References

Allard, D. & Naveau, P. (2008). A new spatial skew-normal random field model. *Communications in Statistics - Theory and Methods*, 36(**9**), 1821-1834.

Arellano-Valle, R. B. & Azzalini, A. (2008). The centred parametrization for the multivariate skew-normal distribution. *Journal of multivariate analysis*, 99(**7**), 1362-1382.

Arrué, J., Arellano-Valle, R. B. & Gómez, H. W. (2016). Bias reduction of maximum likelihood estimates for a modified skew-normal distribution, *Journal of Statistical Computation and Simulation*, 86(**15**), 2967-2984.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12, 171-178.

Azzalini, A. (2005). The skew–normal distribution and related multivariate families. *Scandinavian Journal of Statistics*, 32(**2**), 159-188.

Azzalini, A. & Arellano-Valle, R. B. (2013). Maximum penalized likelihood estimation for skew-normal and skew-t distributions. *Journal of Statistical Planning and Inference*, 143(**2**), 419-433.

Azzalini, A. & Capitanio, A. (1999). Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(**3**), 579-602.

Barndorff-Nielsen, O. E. & Cox, D. R. (1994). Inference and asymptotics. Chapman & Hall/CRC. Boca Raton.

Barndorff-Nielsen, O. E. (1988). Parametric Statistical Models and Likelihood. Lecture Notes in Statistics. Springer, New York.

Cole, D. J. (2020). Parameter Redundancy and Identifiability. Chapman and Hall/CRC. New York.

Figueiredo, F. & Gomes, M. I. (2013). The skew-normal distribution in SPC. *REVSTAT - Statistical Journal*, 11(**1**), 83-104.

Gupta, A. K., Nguyen, T. T. & Sanqui, J. A. (2004). Characterization of the skew-normal distribution. *Annals of the Institute of Statistical Mathematics*, 56, 351-360.

Genton, M. G. (2004). Skew-elliptical Distributions and Their Applications: a Journey Beyond Normality. Chapman & Hall/CRC. Boca Raton.

Gupta, R. C. & Brown, N. (2001). Reliability studies of the skew-normal distribution and its application to a strength-stress model. *Communications in Statistics-Theory and Methods*, 30, 2427-2445.

Hollander, M., Wolfe, Douglas A. & Chicken, E. (2014). Nonparametric Statistical Methods. Wiley. New Jersey.

Hossain, A. & Beyene, J. (2015). Application of skew-normal distribution for detecting differential expression to microRNA data, *Journal of Applied Statistics*, 42(**3**), 477-491.

Huber, P. J. (1981). Robust Statistics. Wiley. New York

Jones, M. C. (2015). On families of distributions with shape parameters. *International Statistical Review*, 83(**2**), 175-192.

Jurečková J. & Picek, J. (2006). Robust Statistical Methods with R. Chapman & Hall/CRC. New York.

Kalbfleisch, J. G. (1985). Probability and Statistical Inference, Vol. 2. Springer-Verlag. New York.

Murphy, S. A. & Van Der Vaart, A. W. (2000). On profile likelihood. *Journal of the American Statistical Association*, 95(**450**), 449-465.

Pawitan, Y. (2001). In All Likelihood: Statistical Modelling and Inference Using Likelihood. Oxford University Press. New York.

Pewsey, A. (2000). Problems of inference for Azzalini's skew normal distribution, *Journal of Applied Statistics*, 27(**2**), 859-870.

Rieder, H. (1994). Robust Asymptotic Statistics. Springer-Verlag. New York.

Rubio, F. J. & Genton, M. G. (2016). Bayesian linear regression with skew-symmetric error distributions with applications to survival analysis. *Statistics in Medicine*, 35(**14**), 2441-2454.

Serfling, R. J. (2002). Approximation Theorems of Mathematical Statistics. John Wiley & Sons. New York.

Sprott, D. A. (2000). Statistical inference in science. Springer-Verlag. New York.

Revista de la Facultad de Ciencias Universidad Nacional de Colombia, Sede Medellín
V 11 N°2 julio-diciembre de 2022 • ISSN-e 2357-5749 • Research paper • Pages 74 a 99
DOI: https://doi.org/10.15446/rev.fac.cienc.v11n2.100986

FLAT LIKELIHOODS: SIR-POISSON MODEL CASE[a]

VEROSIMILITUDES PLANAS: CASO DEL MODELO SIR-POISSON

JOSÉ A. MONTOYA[b*], GUDELIA FIGUEROA PRECIADO[c], MAYRA TOCTO ERAZO[d]

Recibido 11-02-2022, aceptado 29-06-2022, versión final 30-06-2022
Research paper

ABSTRACT: Systems of differential equations are used as the basis to define mathematical structures for moments, like the mean and variance of probability distributions of random variables. Nevertheless, the integration of a deterministic model and a probabilistic one, in order to describe a random phenomenon, and take advantage of the observed data for making inferences on certain population dynamic characteristics, can lead to parameter identifiability problems for the observed sample. Furthermore, approaches to deal with those problems are usually inappropriate. In this paper, the shape of the likelihood function of a SIR-Poisson model is used to describe the relationship between flat likelihoods and the practical parameter identifiability problem. In particular, we show how a flattened shape for the profile likelihood of the basic reproductive number R_0, arises as the observed sample (over time) becomes smaller, causing ambiguity regarding the shape of the average model behavior. We conducted some simulation studies to analyze the flatness severity of the R_0 likelihood, and the coverage frequency of the likelihood-confidence regions for the model parameters. Finally, we describe some approaches to deal the practical identifiability problem, showing the impact that those can have on inferences. We believe this work can help to raise awareness on the way statistical inferences can be affected by a priori parameter assumptions and the underlying relationship between them, as well as those arising by model reparameterizations and incorrect model assumptions.
KEYWORDS: Flat likelihood function; ordinary differential equations; SIR model, basic reproductive number; likelihood contours; profile likelihood function.

RESUMEN: Sistemas de ecuaciones diferenciales son usados como base para definir estructuras matemáticas para momentos, como la media y la varianza de distribuciones de probabilidad de variables aleatorias. Sin embargo, integrar un modelo determinista con uno de probabilidad para representar un fenómeno aleatorio y aprovechar los datos observados de dicho fenómeno para hacer inferencia sobre características de la dinámica poblacional, puede producir problemas de identificabilidad de parámetros para la muestra observada. Más aún, dichos problemas suelen ser tratados de forma inapropiada. En este artículo se utiliza la forma de la función de verosimilitud del modelo SIR-Poisson para describir la relación entre funciones de verosimilitud planas y el problema de identificabilidad práctica de parámetros. En particular, se muestra cómo la forma aplanada de la función de verosimilitud perfil del número reproductivo básico R0 aparece a medida que la muestra observada (en el tiempo) se reduce y produce ambigüedad en la forma del comportamiento medio del modelo. Se efectuan diversos estudios de simulación con el fin de analizar

[a]Montoya, J. A., Figueroa-Preciado, G. & Tocto Erazo, Mayra (2022). Flat likelihoods: SIR-Poisson model case. *Rev. Fac. Cienc.*, 11 (**2**), 74–99. DOI: https://doi.org/10.15446/rev.fac.cienc.v11n2.100986

[b]PhD in Probability and Statistics. Statistics Professor. Department of Mathematics. University of Sonora, México.
[*]Corresponding author: arturo.montoya@unison.mx
[c]PhD in Probability and Statistics. Statistics Professor. Department of Mathematics. University of Sonora.
[d]PhD in Mathematics. Mathematics Professor. Department of Mathematics. University of Sonora, México.

la severidad de la planura de la verosimilitud de R_0 y la frecuencia de cobertura de las regiones de verosimilitud-confianza de los parámetros del modelo. Finalmente, se describen algunos tratamientos que se han utilizado para abordar el problema de identificabilidad práctica y se muestra el impacto que éstos pueden tener en las inferencias. Se considera que este trabajo ayudará a tomar conciencia sobre cómo las inferencias estadísticas se ven afectadas por suposiciones a priori sobre los parámetros y la relación subyacente entre ellos, así como por reparametrizaciones del modelo y por suposiciones incorrectas sobre el mismo.
PALABRAS CLAVE: Función de verosimilitud plana; ecuaciones diferenciales ordinarias; modelo SIR; número reproductivo básico; contornos de verosimilitud; función de verosimilitud perfil.

1. INTRODUCTION

Systems of differential equations are mathematical tools broadly used for modeling physical, chemical, biological and epidemiological processes, among others. In particular, mathematical models based on ordinary differential equations have been widely proposed to study different infectious diseases such as dengue, Zika, Ébola, COVID-19, cholera, among others. For this modeling approach, the SIR model (Kermack & McKendrick, 1927) has been the basis of many models (Arino & Van den Driessche, 2003; Pandey *et al.*, 2013; Khan *et al.*, 2014; Xiao & Zou, 2014; Cosner, 2015; Lee & Castillo-Chavez,, 2015; Hendron & Bonsall, 2016; Phaijoo & Gurung, 2016; Kim *et al.*, 2017; Ghosh *et al.*, 2018; Mishra *et al.*, 2018; Mishra & Gakkhar, 2018; Nguyen *et al.*, 2018; Qi *et al.*, 2018; Sasmal *et al.*, 2018; Tuncer *et al.*, 2018; Vinh *et al.*, 2018; Tocto-Erazo *et al.*, 2020; Tocto-Erazo *et al.*, 2021). These models usually provide a threshold value, known as the basic reproductive number (R_0), that indicates the severity of the disease, as well as the existence of endemic states. Estimation of parameter R_0 is relevant in public health decision making, being very helpful for implementing adequate controls for infectious diseases.

In the problem regarding the estimation of parameter R_0, as well as some other parameters of systems of differential equations, it is common to use a Poisson model or a Negative Binomial distribution, to describe the frequency distribution of random variables associated to different observable characteristics of the dynamic under study, like the number of infected people, number of deaths, and some others (Chowell *et al.*, 2007; Chowell *et al.*, 2008; Lloyd-Smith, 2007; Chowell *et al.*, 2012; Capistrán *et al.*, 2012; Acuña-Zegarra *et al.*, 2020; Acuña-Zegarra *et al.*, 2021; Núñez-López *et al.*, 2021). In these cases it is common to use systems of differential equations as the basis for defining mathematical structures for the mean and variance of these probability distributions. In this paper we consider the Poisson model, whose mean and variance are fully determined by a SIR model. We selected both models (the SIR and the Poisson one) due to: convenience regarding the number of unknown parameters and computational effort; they are representative in the statistical and ordinary differential equations literature. Despite the mathematical simplicity of these models, they enable us to illustrate the essence of the parameter identifiability problem for an observed sample, the main objective of this manuscript; allowing also to show not only how this problem have been handled, but also the impact it can have on resulting inferences.

Regarding the negative binomial case, even though it is also a counting model used in the differential equations parameter estimation, this will not be addressed here. This model has two unknown parameters, a probability of exit and an overdispersion parameter; both involved in the mathematical expression of its mean or expected value. Therefore, it could happen that the occurrence of a lack of parameter identifiability, based on the observed sample, could not only be caused by the relationship between the mean of the negative binomial distribution and the SIR model parameters, and their own relationship, but also could be due (to some extent) by the negative binomial model itself, caused by the relationship between its parameters. In fact, we consider that the analysis of this distribution, in the context studied in this manuscript, is a research topic in its own right.

The problem of structural and practical identifiability, in models involving systems of differential equations, has already been reported in literature by several authors (Rosenbaum *et al.*, 1999; Raue *et al.*, 2009; Zhan *et al.*, 2015; Tuncer *et al.*, 2016; Gábor *et al.*, 2017; Kao & Eisenberg, 2018; Marquis *et al.*, 2018; Saccomani & Thomaseth, 2018). In this paper we are interested in the problem of parameter identifiability for an observed sample (practical identifiability), which mainly occurs when relatively flat likelihood functions are observed, regarding to their maximum, yielding likelihood-confidence regions that cover a large part of the parameter space, and possibly extending infinitely (Raue *et al.*, 2009; Cole, 2020). In particular, for the case of the SIR-Poisson model we are interested in describing the relationship between flat likelihood functions and the practical parameter identifiability problem, considering not only a standard reparameterization of the SIR model, but also another one involving the basic reproductive number R_0. Moreover, we are also interested about quantifying the frequency of occurrence of flat likelihoods, and to understand the observed sample configurations that cause these kind of likelihoods. In this paper we also show some approaches that have been used to address the problem of parameter identifiability, pointing out the impact that these can have on inferences. This is done in order to understand the effect that our assumptions and reparameterizations have on inferences, and to become aware about the validity of the uncertainty quantification in the statistical estimation of the parameters of systems of differential equations.

In this paper we perform some simulations in order to explore the extent to which the SIR-Poisson model leads to flat likelihoods, analyzing the coverage frequency of the likelihood-confidence regions of the model parameters. In particular, we focus on the behavior of the relative profile likelihood function of R_0, when R_0 takes large values, showing that flat likelihoods frequently occur when the observed sample (over time) is reduced, producing ambiguity in the shape of the average model behavior. Thus, in real applications, the upper limit of R_0 confidence intervals would be practically infinite, a result that is not very useful for practical purposes, in a real problem.

However, the coverage frequencies of these intervals are still close to the nominal probability coverage. Finally, in this paper we describe three approaches that have been used to address the problem of practical

parameter identifiability, showing the impact these can have on inferences: one of them involves setting values to the unknown parameters; another approach relies on estimating the growth rate of the SIR model, by assuming an exponential growth model for the data; and the third one involves performing Monte Carlo simulations to estimate R_0, based on a prior knowledge for the distributions of the parameters of the SIR model. In either case, we illustrate how quantitative claims about parameters and uncertainty estimation are both completely determined by these estimation procedures.

This paper is organized as follows. In Section 2, a SIR model is reparametrized in terms of an unknown parameter vector that involves the mean and variance of a Poisson model. The likelihood function, as well as the relative likelihood function are defined and used to construct likelihood regions for this parameter vector. Profile likelihood function is also defined and used to test hypothesis concerning a scalar parameter. All these functions allow to easy visualize plausible and implausible parameter values. Section 3 includes an example where a sample from a SIR-Poisson distribution is simulated, in order to show situations where different parameter values give rise to same likelihood, for the observed data; illustrating the practical parameter identifiability problem. On the other hand, in Section 4, a simulation study to analyze the shape of the profile likelihood function of parameter R_0, the basic reproductive number, is performed. The selected scenarios and experimental features allow to analyze how severe the flatness of this function can become. Section 5 includes three common approaches used to deal with flat likelihoods in the SIR-Poisson model case. Some situations that may lead to inappropriate or overestimated inferences are carefully discussed. Finally, some general conclusions are presented in Section 6.

2. SIR-POISSON LIKELIHOOD

Suppose that $X_1, X_2, \cdots, X_n$ are independent random variables, observed at times $t_1, \ldots, t_n$, respectively. Suppose furthermore that X_i is Poisson distributed with mean

$$\lambda_i = \int_{t_{i-1}}^{t_i} \beta \frac{I(t)}{N} S(t)\, dt, \tag{1}$$

where N is a constant population, $\beta > 0$, and $\gamma > 0$ are the transmission and recovery rates, $I(t)$ and $S(t)$ are solutions for the SIR model given by

$$\begin{aligned} \dot{S} &= -\beta \frac{I(t)}{N} S(t), \\ \dot{I} &= \beta \frac{I(t)}{N} S(t) - \gamma I(t), \\ \dot{R} &= \gamma I(t), \end{aligned} \tag{2}$$

$N = S + I + R$, and initial conditions at time $t_0 = 0$ are fixed and known: $S(t_0) = S_0$, $I(t_0) = I_0$, and $R(t_0) = N - S_0 - I_0$.

Note that λ_i in (1) relies on the SIR model reparameterization given in (2), so for a sake of clarity we will write it as $\lambda_i(\theta)$ instead of λ_i, in order to emphasize that SIR model is reparametrized in terms of the unknown parameters vector $\theta = (\beta, \gamma)$, and also that the unknown parameters of the Poisson model, governing its mean and variance, are β and γ. This kind of Poisson model will be called SIR-Poisson model. It is important here to note that $\theta = (\beta, R_0)$ indicates that the SIR-Poisson model is reparametrized in terms of the transmission rate β and the basic reproductive number (or basic reproductive ratio) $R_0 = \beta/\gamma$; that is, $(\beta, \gamma) \leftrightarrow (\beta, \beta/R_0)$. We consider this reparameterization since it involves R_0, a parameter of considerable interest in infectious disease control literature, for which some estimation techniques have been developed when there is a presence of parameter identifiability problems.

Let $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)$ be an observed sample from $X_1, X_2, \cdots, X_n$. The likelihood function of parameter vector $\theta = (\beta, \gamma)$, associated to the observed sample $\boldsymbol{x}$, can be written as

$$L(\theta; \boldsymbol{x}) = \prod_{i=1}^{n} \frac{1}{x_i!} [\lambda_i(\theta)]^{x_i} \exp[-\lambda_i(\theta)], \tag{3}$$

where $\theta = (\beta, \gamma) \in \mathbb{R}^+ \times \mathbb{R}^+$. Note that the likelihood function of (β, R_0) is obtained by replacing $\theta = (\beta, \beta/R_0)$ en (3).

The relative likelihood function of θ is a standardized version of (3). It takes the value of one at the maximum likelihood estimate of θ,

$$R(\theta; \boldsymbol{x}) = \frac{L(\theta; \boldsymbol{x})}{\max_{\theta} L(\theta; \boldsymbol{x})} = \frac{L(\theta; \boldsymbol{x})}{L\left(\widehat{\theta}; \boldsymbol{x}\right)}, \tag{4}$$

where $\widehat{\theta} = \left(\widehat{\beta}, \widehat{\gamma}\right)$ is the maximum likelihood estimator (MLE) of parameter $\theta = (\beta, \gamma)$. Using the invariance likelihood property, the MLE for R_0 results $\widehat{R}_0 = \widehat{\beta}/\widehat{\gamma}$. The relative likelihood in (4) measures the plausibility of any specified θ value, relative to that of $\widehat{\theta}$. This likelihood function ranks all possible θ values according their respective plausibilities, in light of the observed sample $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)$. Note that $\theta \in \mathbb{R}^+ \times \mathbb{R}^+$, thus, given a plot of $R(\theta; \boldsymbol{x})$ we can easily distinguish plausible and implausible values for $\theta = (\beta, \gamma)$, or equivalently for $\theta = (\beta, R_0)$. Furthermore, a plot of $R(\theta; \boldsymbol{x})$ can be used also to visualize the flatness of the likelihood surface.

Another practical way to make inferences about θ, which also allows us to visualize and describe the likelihood surface, is through likelihood regions. The likelihood region for θ, at likelihood level c, where $c \in [0, 1]$, is given by

$$\mathcal{C}_\theta(c; \boldsymbol{x}) = \{\theta : R(\theta; \boldsymbol{x}) \geq c\}. \tag{5}$$

Any θ value within this region has a relative likelihood $R(\theta; \boldsymbol{x}) \geq c$, while $R(\theta; \boldsymbol{x}) < c$ for those outside this region; so just looking at the contour plot we can easily identify plausible and implausible θ values, at likelihood level c. Now, varying c from 0 to 1, we obtain a complete set of nested likelihoods regions

converging to the MLE $\widehat{\theta}$, as $c \to 1$; see Sprott (2000, pp. 14-15).

On the other hand, a $100\,(1-\alpha)$ % confidence interval can be assigned to $\mathcal{C}_\theta\,(c;\boldsymbol{x})$ in (5), considering the asymptotical distribution of the likelihood ratio $-2\ln R\,(\theta;\boldsymbol{X})$, under $H_0 : \theta = \theta_0$. This likelihood ratio has a Chi-squared distribution with two degrees of freedom, χ^2_2, (Kalbfleisch, 1985; Serfling, 2002). When $-2\ln c = q_{2,1-\alpha}$, where $q_{2,1-\alpha}$ is the quantile of probability $1-\alpha$ of a Chi-squared distribution with two degrees of freedom, the confidence level for $\mathcal{C}_\theta\,(c;\boldsymbol{x})$ results $100\,(1-\alpha)$ %. For example, when $c = 0.05$, the confidence level is approximately 95 %. Moreover,

$$P\text{-value} = P\left\{\chi^2_2 \geq -2\ln R\,(\theta_0;\boldsymbol{x})\right\} = R\,(\theta_0;\boldsymbol{x})\,,$$

and every specified θ value has an associated P-value greater or equal than $c = 0.05$ when lying within this relative likelihood contour, and less than $c = 0.05$ otherwise.

Now, assume that vector parameter θ can be partitioned as follows: an interest parameter ψ and a nuisance parameter λ. Then, the profile likelihood of ψ and its corresponding relative likelihood function are defined as

$$\begin{aligned} L_{\max}\,(\psi;\boldsymbol{x}) &= \max_{\lambda} L\,(\psi,\lambda;\boldsymbol{x})\,, \\ R_{\max}\,(\psi;\boldsymbol{x}) &= \frac{L_{\max}\,(\psi;\boldsymbol{x})}{\max\limits_{\psi,\lambda} L\,(\psi,\lambda;\boldsymbol{x})} = \frac{L_{\max}\,(\psi;\boldsymbol{x})}{L_{\max}\,(\widehat{\psi};\boldsymbol{x})} \end{aligned} \tag{6}$$

For example, we might be interested in making inferences about R_0, so the profile likelihood of $\psi = R_0$, and its corresponding relative likelihood function can be calculated as

$$\begin{aligned} L_{\max}\,(R_0;\boldsymbol{x}) &= \max_{\beta} L\,(\theta = (\beta,\beta/R_0)\,;\boldsymbol{x})\,, \\ R_{\max}\,(R_0;\boldsymbol{x}) &= \frac{L_{\max}\,(R_0;\boldsymbol{x})}{\max\limits_{\theta} L\,(\theta;\boldsymbol{x})} = \frac{L_{\max}\,(R_0;\boldsymbol{x})}{L_{\max}\left(\widehat{R}_0;\boldsymbol{x}\right)}. \end{aligned} \tag{7}$$

The relative profile likelihood varies between 0 and 1, and ranks all possible R_0 values based on the observed sample $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)$ on times $t_1, \ldots, t_n$. Thus, a plot of $R_{\max}\,(R_0;\boldsymbol{x})$ allows to distinguish plausible and implausible R_0 values. Note that we might be interested in making inferences about β, so we can use $\theta = (\beta,\gamma)$ and consider $(\psi,\lambda) = (\beta,\gamma)$.

A level c profile likelihood region (commonly an interval) for the scalar parameter of interest ψ is given by

$$\mathcal{C}_\psi\,(c;\boldsymbol{x}) = \{\psi : R_{\max}\,(\psi;\boldsymbol{x}) \geq c\}\,, \tag{8}$$

where $0 \leq c \leq 1$. Considering that the profile likelihood ratio statistic $-2\ln R_{\max}\,(\psi;\boldsymbol{X})$ follows an asymptotically χ^2_1 distribution, if the null hypothesis $H_0 : \psi = \psi_0$ is true, then (8) is a confidence interval for

the scalar parameter ψ, with approximate confidence level determined from $-2\ln R_{\max}(\psi_0;\boldsymbol{X}) \sim \chi^2_1$. When $-2\ln c = q_{1,1-\alpha}$, where $q_{1,1-\alpha}$ is the quantile of probability $1-\alpha$ of a Chi-squared distribution with one degree of freedom, the confidence level for $\mathcal{C}_\psi(c;\boldsymbol{x})$ is $100(1-\alpha)$ %. For example, the confidence level will be approximately 90 %, 95 %, and 99 % if $c = 0.25$, 0.146, and 0.036, respectively; see Kalbfleisch (1985, pp. 115-116).

In general, the profile or maximized likelihood is a powerful but simple inferential method for estimating a parameter of interest, separately from the remaining unknown parameters of the statistical model. More details about the profile likelihood approach are described in Kalbfleisch (1985, Section 10.3), Pawitan (2001, Section 3.4), Sprott (2000, p. 66), Barndorff-Nielsen & Cox (1994, pp. 89-91), Serfling (2002, pp. 155-160), and Murphy & Van Der Vaart (2000). An important and mostly unknown aspect about the profile likelihood function is that it can be used to study and visualize several features of the whole likelihood function, such as for instance those associated with the flatness of the likelihood surface. The example included in the following section illustrates how the graph of a profile likelihood, particularly the case of R_0, can reveal the flat nature of the likelihood function of a SIR-Poisson model.

3. EXAMPLE: SIR-POISSON LIKELIHOOD

The dataset in Table 1 has been generated using a SIR-Poisson distribution, with vector parameters $(\beta,\gamma) = (1.786946, 0.4761905)$, $N = 763$, and initial conditions $I_0 = 1$ and $S_0 = 762$. Figure 1 shows, in grayscale, the relative likelihood function for (β,γ) and its corresponding contour plot. In particular, the $c = 0.05$ contour level is indicated with a dashed line, and the MLE is marked with an asterisk ($\widehat{\beta} = 1.845084$ and $\widehat{\gamma} = 0.520862$). Just looking at the shape of the likelihood surface, and its elongated contours, there is evidence of the close relationship between parameters β and γ. Nevertheless, the region of plausible values for both parameters is clearly into the plots limits. In this case, profile likelihood functions for β and γ are well-behaved, in the sense that they grow, reach a maximum and then decrease; see Figure 2(a)-(b). The 95 % confidence intervals for β and γ are $\mathcal{C}_\beta(c = 0.146;\boldsymbol{x}) = [1.644828, 2.063218]$ and $\mathcal{C}_\gamma(c = 0.146;\boldsymbol{x}) = [0.3068966, 0.7609195]$, respectivamente.

It is important to note that in many applications, where parameters of differential equation models are estimated, only data with an increasing behavior are available or used, data that apparently have already reached a peak. In the following example we simulate this kind of scenario, assuming that only 40 % of the observations are available; that is, only the first four observations in Table 1, $\boldsymbol{x} = (2, 12, 59, 130)$ are known. Now, Figure 3 shows the relative likelihood function for (β,γ) and its corresponding contour plot; in this plot the $c = 0.05$ contour level is marked with a dashed line and the MLE is indicated with an asterisk ($\widehat{\beta} = 1.9047994$ and $\widehat{\gamma} = 0.5635199$). In this case, the resulting contours are even more elongated than those obtained when considering the 100 % of the simulated sample, emphasizing again the close relationship between both parameters, β and γ. Now, the region of plausible values for both parameters is truncated within

Table 1: Simulated sample from a SIR-Poisson distribution with $\beta = 1.786946$, $\gamma = 0.4761905$, $N = 763$, $I_0 = 1$, and $S_0 = 762$.

i	1	2	3	4	5	6	7	8	9	10
t_i	1	2	3	4	5	6	7	8	9	10
x_i	2	12	59	130	206	171	92	42	19	8

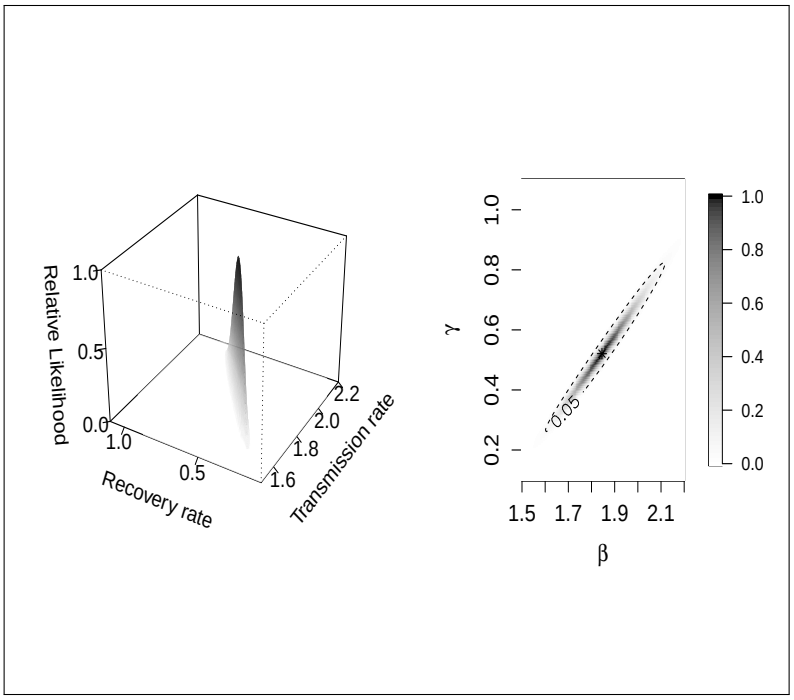


Figure 1: Likelihood surface and contour plot for the relative likelihood of (β, γ), corresponding to data presented in Table 1. The grayscale bar shows different values for the relative likelihood; the 95 % confidence region is plotted with a dashed line, at $c = 0.05$ contour level, and the MLE is marked with an asterisk. Source: Elaborated by the authors.

the limits of the plots. This can be better observed in the profile likelihood functions of both parameters, plotted in Figure 4(a)-(b); particularly, the γ profile likelihood function assigns a high plausibility to those γ values close to zero. Note that due to the intrinsic relationship exhibited by both parameters, the β profile likelihood function also presents a similar behavior. The upper limits of the 95 % confidence intervals for β and γ are 2.925301 and 1.656964, respectively.

Now, regarding the model reparameterization (β, R_0), the shape of the likelihood surface is much flatter than the observed in the other cases, with elongated and curved contours, as can be seen observed in Figure 5. This plot shows that the likelihood function has a ridge, with a relatively flat top.

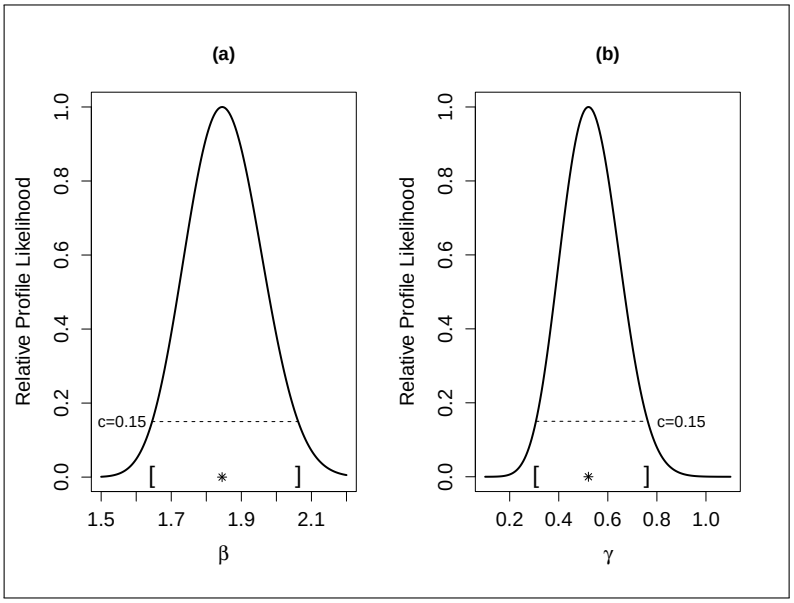


Figure 2: Relative profile likelihood functions for β and γ, based on the simulated data presented in Table 1. In both cases, (a) and (b), the 95 % confidence intervals are indicated in brackets, and the MLE is marked with an asterisk. Source: Elaborated by the authors.

The profile relative likelihood of R_0, that is shown in Figure 6, illustrates this kind of flatness, where the likelihood increases until reaching its maximum, decreasing then slowly; actually, plausibility of $R_0 = 6$ and $R_0 = 10$ are 0.85 and 0.72, respectively. The lower limit of the 95 % confidence interval for R_0 results 1.769076.

When for a fixed data set we observe two different parameter values giving rise to the same likelihood (probability of observed data), then it results practically impossible to distinguish between these two parameter candidates, based on the data alone; thus, it would be not feasible to identify the true parameter value. In that sense, when the relative profile likelihood function of R_0 is flat over a wide range of R_0 values, including the MLE such as in Figure 6, then the probability of the observed sample is the same over all this R_0 value set. Therefore, the lack of identifiability of this parameter occurs at least for the observed data. Now, since the profile likelihood function for the parameter of interest is invariant under reparameterizations of the nuisance parameters, then, the plot of the profile likelihood function of R_0 will not change under reparameterizations of (β, γ).

pagebreak

Sprott (2000, p. 11) stated that the MLE and the observed information exhibit two important features of the likelihood function, the MLE as a measure of its location relative to the parameter-axes, while the latter as a proxy of its curvature, in a neighborhood of the MLE. Thus, when the likelihood function of a scalar parameter is smooth and symmetric around the MLE, then the MLE and the observed information are usually

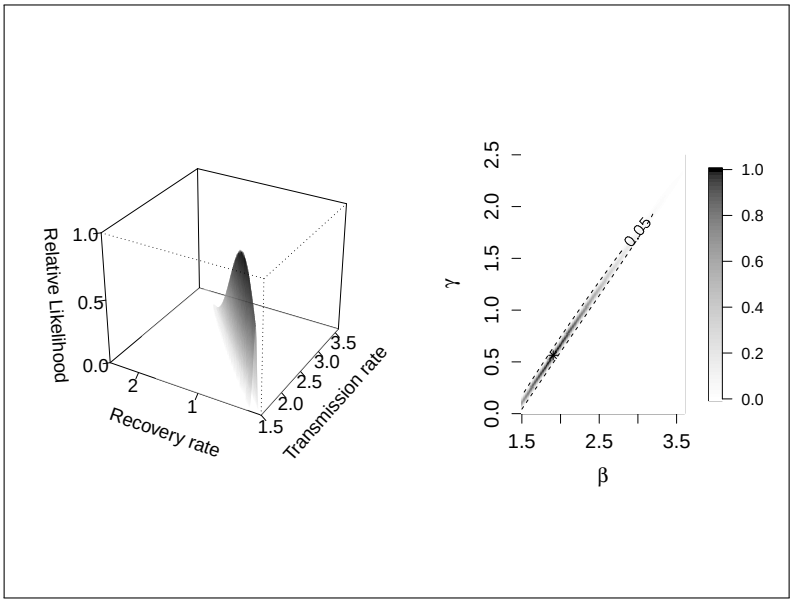


Figure 3: Likelihood surface and contour plots of the relative likelihood of (β,γ) corresponding to $\boldsymbol{x} = (2,12,59,130)$ data in Table 1. Grayscale bar shows different relative likelihood values; the 95 % confidence region is denoted with a dashed line at $c = 0.05$ contour level, whereas the MLE is marked with an asterisk. Source: Elaborated by the authors.

the main features that determine the shape of the likelihood function. However, it should be stressed that flat likelihoods can not be determined by these quantities, without loss of information.

Intuitively, the sensitivity of the MLE depends on the shape of the likelihood function. If the R_0 likelihood is very curved around MLE, then, a stable estimate of R_0 is obtained. On the other hand, if the likelihood is flat near the MLE, this estimate becomes highly unstable. Beside these facts, it is important to mention that a flat likelihood can lead to R_0 likelihood-confidence intervals whose upper limit becomes infinity.

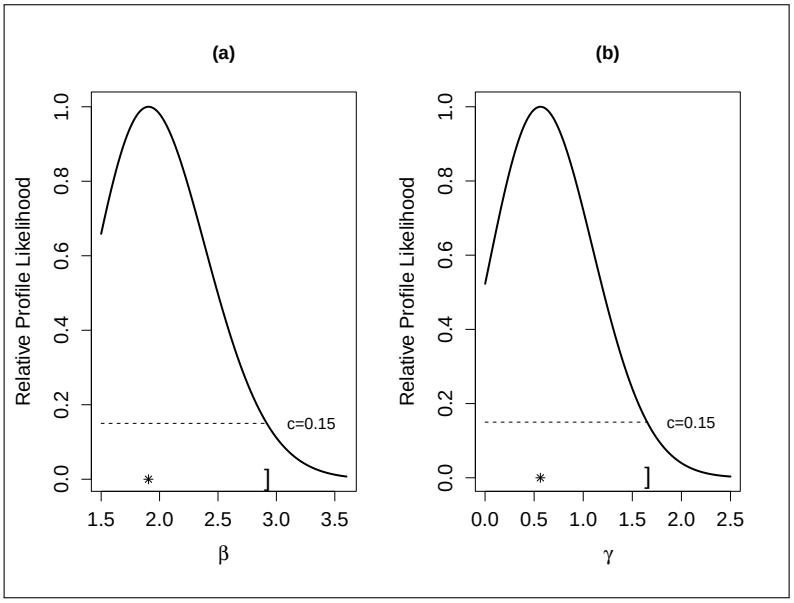


Figure 4: The relative profile likelihood function of β and γ for the simulated data $\boldsymbol{x} = (2, 12, 59, 130)$ in Table 1. In both cases, (a) and (b), the upper limit of the 95% confidence interval is denoted with a bracket, whereas the MLE is marked with an asterisk. Source: Elaborated by the authors.

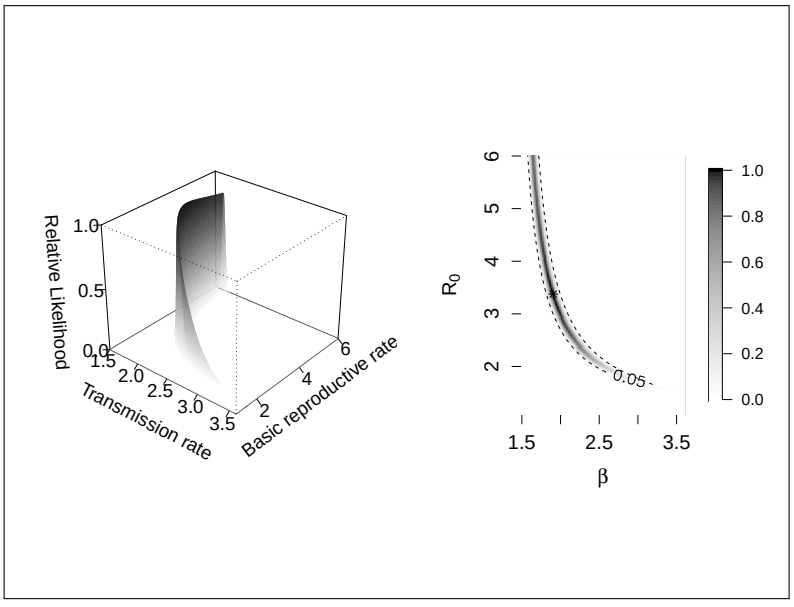


Figure 5: Likelihood surface and contour plots of the relative likelihood of (β, R_0), corresponding to $\boldsymbol{x} = (2, 12, 59, 130)$ data in Table 1. Grayscale bar shows different relative likelihood values; the 95% confidence region is denoted with a dashed line at $c = 0.05$ contour level, whereas the MLE is marked with an asterisk. Source: Elaborated by the authors.

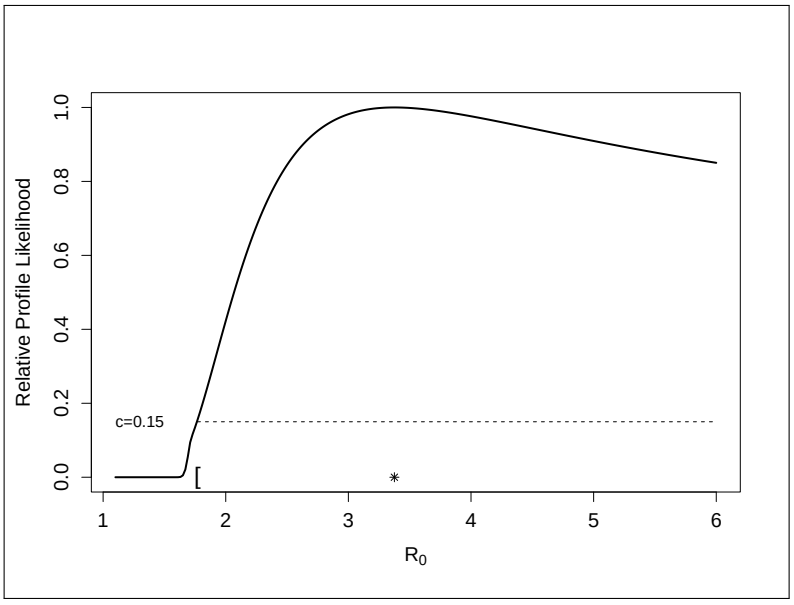


Figure 6: Relative profile likelihood function of R_0 for simulated $\boldsymbol{x} = (2, 12, 59, 130, 206)$ data in Table 1. The upper limit of the 95 % confidence interval is indicated with a bracket and the MLE is marked with an asterisk. Source: Elaborated by the authors.

4. FLATNESS OF THE RELATIVE PROFILE LIKELIHOOD OF R_0

In this section we will conduct some experiments, in order to analyze how the shape of the R_0 profile likelihood function changes, as the observed sample size varies over time. For that purpose, we will examine the distribution of the relative profile likelihood function of R_0, evaluated at different values, all of them larger than the nominal ones used in the simulation scenarios. In particular, we are interested in analyzing the severity of the flatness of the profile likelihood of R_0, when the observed sample (over time) produces ambiguity regarding the shape of the average behavior of the SIR-Poisson model. Furthermore, in all experiments considered here, we will compute the coverage frequency of the likelihood-confidence regions for (β, R_0), β and R_0, in order to study the effect that flat likelihoods could have on the coverage probabilities of these intervals.

In this simulation study we consider 5000 SIR-Poisson samples of sizes $n_1 = 40\,\%n$, $n_2 = 60\,\%n$, $n_3 = 80\,\%n$, and $n_4 = 100\,\%n$, simulated according the following scenarios for n and R_0:

- Case 1, $n = 23$ and $(\beta, \gamma) = (0.893473, 0.4761905)$, that yields $n_1 = 9$, $n_2 = 14$, $n_3 = 18$, and $n_4 = 23$, and $R_0 = 1.876293$.
- Case 2, $n = 10$ and $(\beta, \gamma) = (1.786946, 0.4761905)$ that yields $n_1 = 4$, $n_2 = 6$, $n_3 = 8$, and $n_4 = 10$, and $R_0 = 3.752587$.
- Case 3, $n = 7$ and $(\beta, \gamma) = (2.680419, 0.4761905)$ that yields $n_1 = 3$, $n_2 = 4$, $n_3 = 6$, and $n_4 = 7$, and $R_0 = 5.62888$.

In our simulation scenarios, samples of size n were selected in such a way that maximum is already reached and $I(n) / \max\{I(t)\} \approx 0.31$, where $I(t)$ is the solution of the SIR model. Figure 7 shows the SIR model solution $I(t)$, for each of the R_0 values considered in our simulation cases. It is noteworthy that the n selected values in our simulation experiment allow that λ_i, the mean of the SIR-Poisson model given in (1), grows until reaching its maximum, decreasing later to values close to zero. In this way, $n_1 = 40\,\%n$, $n_2 = 60\,\%n$, $n_3 = 80\,\%n$, and $n_4 = 100\,\%n$, represent observed sample sizes that produce a gradient of information associated to different ambiguity levels regarding the shape of the average behavior of the model.

Based on the simulated samples, we compute the coverage frequency of the likelihood confidence regions $\mathcal{C}_{(\beta,R_0)}(0.05;\boldsymbol{X})$, $\mathcal{C}_{(\beta)}(0.146;\boldsymbol{X})$ and $\mathcal{C}_{(R_0)}(0.146;\boldsymbol{X})$, whose nominal coverage probability is 95 %. Furthermore, for each simulation experiment we also computed the empirical distribution for $R_{\max}\left(R_0 = R_0^U;\boldsymbol{X}\right)$, where $R_0^U = 10$, $R_0^U = 15$, and $R_0^U = 20$. It is worth mentioning that MLEs were estimated using the constrOptim function of the R 4.0.3 version software, and that all results were obtained considering only those samples where the constrOptim function yielded a successful performance of the Nelder and Mead optimization method. For example, in Case 1 $(\beta, \gamma) = (0.893473, 0.4761905)$ with $n = 9$ and Case 2 $(\beta, \gamma) = (1.786946, 0.4761905)$ with $n = 4$ convergence of the method was obtained for the 95 % and 99 % of all 5000 simulated samples. In all remaining scenarios, 100 % of simulated samples successfully converged.

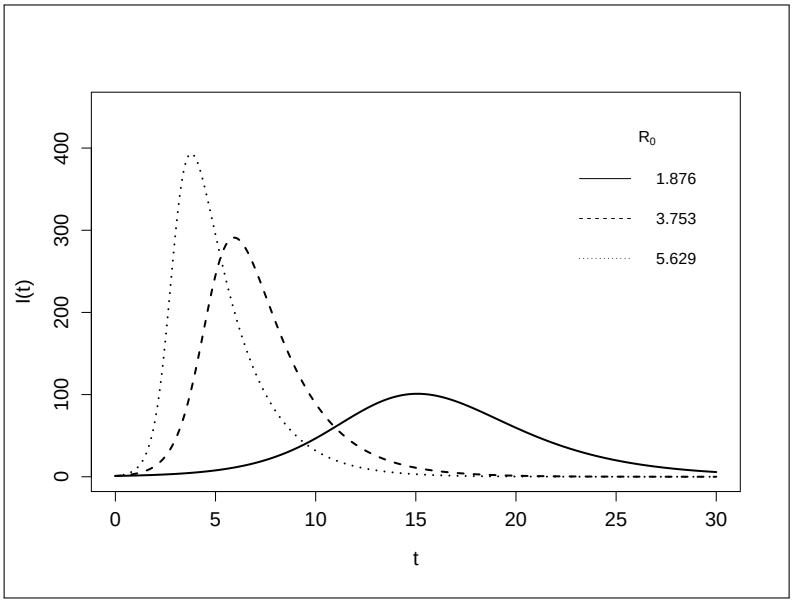


Figure 7: $I(t)$ solutions for the SIR model for different R_0 values considered in simulation scenarios. Source: Elaborated by the authors.

As can be observed in Table 2, in all cases the coverage frequency of the likelihood-confidence regions for the SIR-Poisson model parameters is close to the 95 % nominal probability coverage. On the other hand, it can be observed in Figures 8 to 10, that those scenarios where sample size yield ambiguity regarding the shape of the average behavior of the SIR-Poisson model, the median of the distribution of $R_{\max}\left(R_0 = R_0^U; \boldsymbol{X}\right)$ in $R_0^U = 10$, $R_0^U = 15$, and $R_0^U = 20$ takes positive values that are far from 0, with a downward trend thereafter, though slightly leaning, allowing to infer that the shape of the R_0 relative profile likelihood function becomes fundamentally flat for $R_0 \in [10, 20]$.

For instance, in all cases (Case 1-3) where sample size is $n_1 = 40\,\%n$, the 0.35 level quantile of the R_0^U empirical distributions was larger than the relative likelihood level at $c = 0.146$. That is, each one of the corresponding samples yielded 95 % confidence intervals for R_0, where the interval upper limit was greater or equal to $R_0^U = 20$. This result can be considered as uninformative, given the fact that real R_0 values were 1.876293, 3.752587, 5.62888, and almost absurd in the practical context of an epidemic. On the other hand, for all scenarios in cases with $100\,\%n$ and $R_0^U = 20$, is clear that at least 91 % of the times, the 95 % confidence intervals for R_0 have a finite upper limit, less than $R_0^U = 20$; thus, $R_{\max}\left(R_0 = 20; \boldsymbol{x}\right)$ was less or equal than the relative likelihood at $c = 0.146$ level.

Table 2: Coverage frequency of the likelihood-confidence regions for the SIR-Poisson model parameters.

$(\beta,\gamma) \to R_0 = \beta/\gamma$	n	$\mathcal{C}_{(\beta,R_0)}(0.05;\boldsymbol{X})$	$\mathcal{C}_{(\beta)}(0.146;\boldsymbol{X})$	$\mathcal{C}_{(R_0)}(0.146;\boldsymbol{X})$
$(0.893473, 0.4761905) \to R_0 = 1.876293$	9	96.58	97.36	97.30
	14	95.34	95.24	95.52
	18	95.58	95.40	95.40
	23	95.40	95.32	95.78
$(1.786946, 0.4761905) \to R_0 = 3.752587$	4	96.88	97.89	97.75
	6	94.78	94.74	95.32
	8	94.86	94.72	94.72
	10	94.92	94.82	94.60
$(2.680419, 0.4761905) \to R_0 = 5.62888$	3	95.74	96.62	96.38
	4	94.80	95.18	95.00
	6	94.78	94.88	95.14
	7	94.36	94.24	94.52

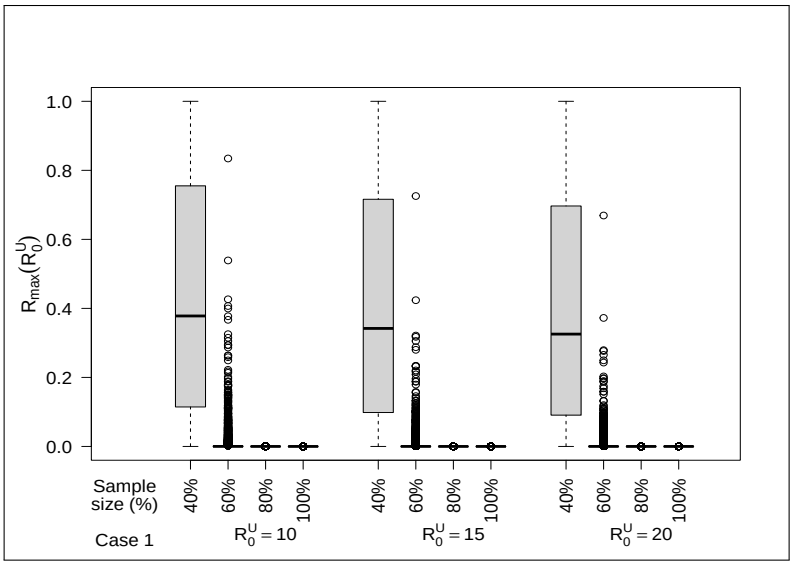


Figure 8: Box plots for $R_{\max}\left(R_0 = R_0^U\right)$ from simulated data under Case 1 parameter values. Source: Elaborated by the authors.

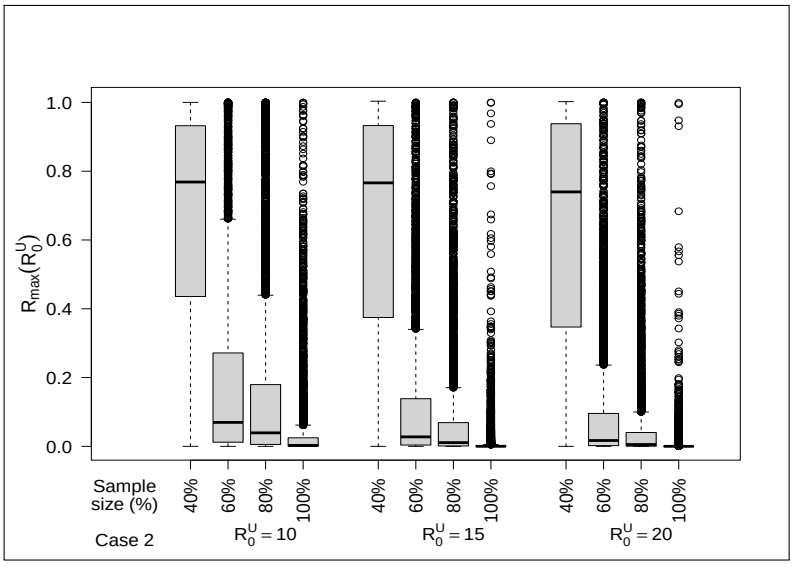


Figure 9: Box plots for $R_{\max}\left(R_0 = R_0^U\right)$ from simulated data under Case 2 parameter values. Source: Elaborated by the authors.

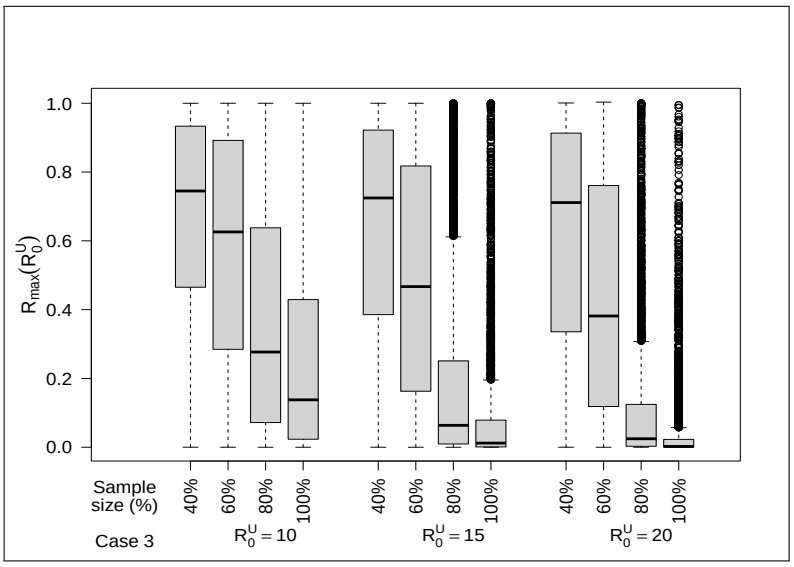


Figure 10: Box plots for $R_{\max}\left(R_0 = R_0^U\right)$ from simulated data under Case 3 parameter values. Source: Elaborated by the authors.

5. SOME ESTIMATION PROCEDURES IN THE PRESENCE OF FLAT LIKELIHOODS

In this section we describe three common parameter estimation approaches, basically applied to the SIR-Poisson model in the presence of flat likelihoods, an indication of a lack of parameter identifiability situation, based on the sample. One approach involves setting values to unknown parameters; another one concerns with estimating the growth rate $\delta = \beta - \gamma$ of the SIR model, assuming a data exponential growth. The third one uses Monte Carlo simulations, based on prior knowledge of the parameter distributions of the SIR model. In addition, we emphasize the impact that these practices can have on R_0 inferences.

5.1. Setting values to unknown parameters

In literature concerning parameter estimation in differential equations, the practice of setting values to an unknown parameter stands out (Khan *et al.*, 2014; Funk *et al.*, 2016; Gui-Quan *et al.*, 2017; Ghosh *et al.*, 2017; Vinh *et al.*, 2018; Guanghu *et al.*, 2019). In general, assigning values to an unknown parameter in a model, in order to estimate some other parameters of that model, causes statistical inferences (estimators, confidence intervals, precision, coverage probability, etc.) to depend on the selected parameter values, that are considered to be known. Our model under study, a SIR-Poisson distribution, is not an exception. To clarify this situation, we will just consider the observed SIR-Poisson model sample $\boldsymbol{x} = (2, 12, 59, 130)$, which represents 40 % of a simulated sample from a SIR-Poisson distribution with vector parameters $(\beta, R_0) = (1.786946, 3.752587)$, $N = 763$, and initial conditions $I_0 = 1$ and $S_0 = 762$. The likelihood functions linked to this sample were shown and described in Section 3. In particular, it was stated that the likelihood of (β, R_0), shown in Figure 3, displays a flat surface with elongated and curved contours, indicating the strong relationship between these parameters, based on the observed sample. Moreover, the relative profile likelihood of R_0 grows until achieving its maximum, decreasing slowly thereafter and assigning high plausibility values (close to 1) not only to true R_0 value, but also to some larger values like $R_0 = 6$ and $R_0 = 10$ (with plausibility 0.85 and 0.72, respectively).

In this particular case, setting parameter β in a fixed and known value β^* yields a likelihood $L(R_0;\boldsymbol{x}) = L(\beta = \beta^*, R_0;\boldsymbol{x})$ that only depends on R_0, the parameter of interest. Figure 11 shows the corresponding relative likelihood, standardized to one at its maximum, for different β values (included $\beta_0 = 1.786946$, the true value used in data simulation). Now, just for the sake of comparison, R_0 relative profile likelihood function and the true R_0 value are marked in this figure.

From this plot becomes clear that inferences are almost completely determined by the value that is specified for β, and only the likelihood corresponding to $\beta = \beta_0$ assigns a non-negligible plausibility to the true R_0 value. Moreover, the three β specified values give rise to likelihoods with different amplitudes, yielding different precision in the likelihood-confidence intervals for R_0, that are constructed considering the same relative likelihood level $0 < c < 1$. In fact, these amplitudes are determined by the relationship between β

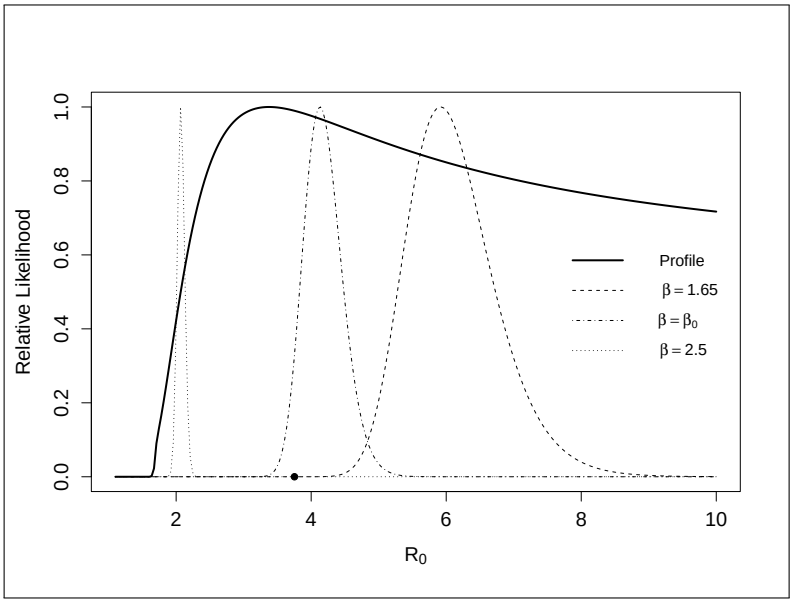


Figure 11: Relative likelihood function of R_0 for different β values ($\beta = 1.85$, $\beta = \beta_0$, and $\beta = 2.5$), and relative profile likelihood function of R_0. True R_0 value is marked with a black dot. Source: Elaborated by the authors.

and R_0, displayed in the contour plot of Figure 5. It is worth mentioning that simulated samples from previous section (Case 2, with $n = 4$) were used to estimate the coverage probability of the 95 % confidence intervals for R_0 that these likelihoods yield. Coverage frequencies associated to $\beta^* = 1.65$, 1.786946, and 2.5 were 2.8 %, 95.38 % and 0 %, respectively.

5.2. Estimation of $\delta = \beta - \gamma$ assuming an exponential growth model

Another approach for estimating parameters of differential equations is to estimate the parameters of a general model, but considering a simpler model for that data; one containing fewer parameters that are apparently involved in the general model. For example, many authors propose compartment models for the dynamics under study using an exponential growth model, where the growth rate δ is estimated independently from the general model (Chowell *et al.*, 2007; Chowell *et al.*, 2012; Towers *et al.*, 2016). Once δ estimates are obtained, they are used to set the parameter values of the general compartment model. However, a problem with this practice is that estimating δ, based on an exponential growth model, does not consider the intrinsic relationship between the parameters of the general differential equation model, originally proposed to model the dynamics under study. Consequently, inferences about these parameters could be very different from those that would be obtained if δ were estimated with the general compartment model.

We show this estimation approach considering the SIR model presented in (2), as a general model, and using a Poisson distribution with mean $f(t;\delta)$ to fit the observed data $\{(t_i;x_i)\}_{i=0}^{n}$, where $\ln(f(t;\delta)) = \ln(I_0) + \delta t$,

δ is the parameter that describes the growth rate in this exponential model, I_0 is the number of initially observed cases, and x_i is the number of new cases during the time interval $(t_{i-1}, t_i]$. In this case, the likelihood function of δ, that yields this exponential growth model is

$$L_{Exp}(\delta) \propto \exp\left(\delta \sum_{i=1}^{n} t_i x_i\right) \exp\left[-I_0 \sum_{i=1}^{n} \exp(\delta t_i)\right]. \tag{9}$$

Moreover, we will use the data set $\boldsymbol{x} = (2, 12, 59, 130)$, presented in previos subsection, in order to compare δ inferences that arise from $L_{Exp}(\delta)$, given in (9), against those inferences arising from the profile likelihood function of the exponential growth rate $\beta - \gamma$ (Ma, 2020), of the SIR model given in (2). In particular, and with certain notation abuse, this profile can be obtained by simply reparameterizing the SIR model with parameters β and γ, in terms of new parameters β and $\delta = \beta - \gamma$ (that is, $\gamma = \beta - \delta$), and then maximizing on β for each fixed δ value.

Figure 12 shows the relative version of (9), standardized to be one at its maximum. Moreover, and just for a comparison purpose, the δ profile likelihood function based on the SIR-Poisson model is also included in this figure, where the true δ value, obtained as the difference between β and γ true values, and used for data simulation, is marked with a black dot. In this figure we can observe that the relative likelihood of δ, yielded by the exponential model $R_{Exp}(\delta)$, is narrower than the profile likelihood of δ, and is located to its left. Moreover, $R_{Exp}(\delta)$ assigns negligible plausibility to many δ values that are plausible under the profile likelihood, included the true δ value. It is important to note that Case 2 simulated samples with $n = 4$ were also used here, in order to estimate the coverage probability of the 95 % confidence intervals of δ, yielded by $R_{Exp}(\delta)$ and δ profile likelihood functions (this last one computed under the basis of the SIR-Poisson general model). The resulting coverage frequencies associated to $R_{Exp}(\delta)$ and the profile likelihood were 0.04 % and 95.38 %, respectively.

5.3. Monte Carlo Simulation

Another approach to estimate the basic reproductive number R_0, of a system of differential equations with parameter θ, involves obtaining a mathematical expression for $R_0 = h(\theta)$, where $h(\cdot)$ is a function of θ, and assume a joint distribution $F(\cdot)$ for this parameter, that allows to generate $\theta_1, \theta_2, \cdots, \theta_M$ values, that are evaluated in $h(\cdot)$ in order to obtain R_0, $R_{01}, R_{02}, \cdots, R_{0M}$ values of the parameter of interest. These values are used to estimate the distribution of the basic reproductive number and to construct confidence intervals for R_0.

A difficulty that arises when trying to properly apply this practice, is obtaining a joint distribution $F(\theta)$ that rescues the intrinsic relationship between the parameters of the differential equation model, that is proposed to model the dynamics of the phenomenon under study. Typically, according to literature on differential equation parameter estimation, it is common to make literature reviews to collect information about an estimated range for the parameters of the differential equation model, and use this information to propose

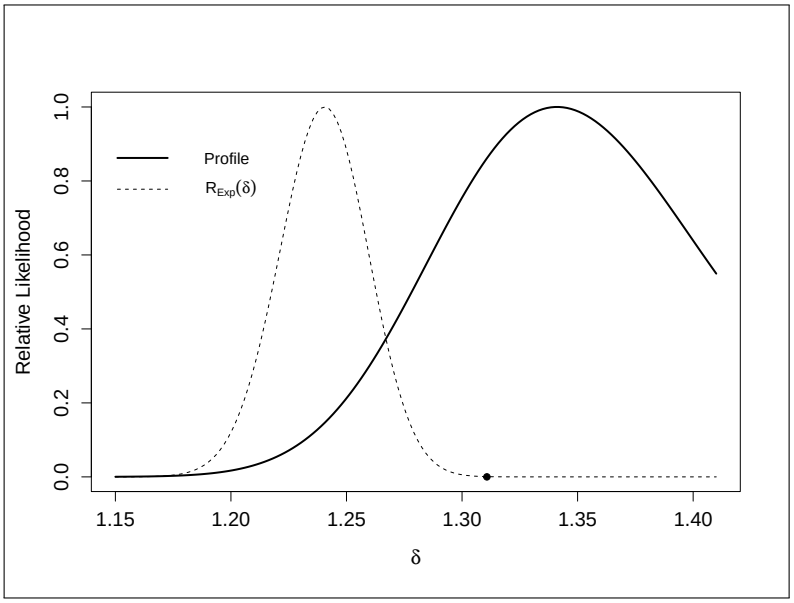


Figure 12: Relative likelihood and relative profile likelihood functions of δ, obtained from a model with an exponential growth. True δ value is marked with a black dot. Elaborated by the authors.

independent distributions for each of the model parameters (Pandey *et al.*, 2013; Camacho *et al.*, 2014; Funk *et al.*, 2016; Saldaña *et al.*, 2020).
It is also common to combine the approach described in the previous section and use the asymptotic distribution of the δ MLE, the parameter representing the growth rate of the exponential model, instead of a uniform distribution (Chowell *et al.*, 2007; Towers *et al.*, 2016).

Intuitively, setting aside the relationship between model parameters and assuming independence in order to construct $F(\theta)$, causes that implausible θ values become considered in the calculation of R_0, and that occurs given the intrinsic relationship between model parameters. In addition, misspecification of any of the parameter distributions may lead also to inaccurate inferences. To exemplify this practice and the kind of results it can yield, we will consider the SIR model in (2), but reparametrized in terms of $\theta = (\beta, \delta)$. To specify $\theta = (\beta, \delta)$ distribution, we assume independence and propose a uniform distribution $(1.2, 2)$ for β, in such a way that this interval includes true $\beta = 1.786946$ value. On the other hand, we use a normal distribution $N\left(\widehat{\delta}, \widehat{\sigma}^2\right)$ for δ, whose mean $\widehat{\delta} = 1.240564$ and variance $\widehat{\sigma}^2 = 1.240564$ are the MLE of δ and the Fisher information inverse, respectively; both obtained through the likelihood given in (9). Considering the proposed distributions, $M = 10000$ random samples for β and δ were simulated; in all cases it occurred that $\beta > \delta$, obtaining then 10000 computations of $R_0 = \beta/(\beta - \delta)$, which were used to build the histogram shown in Figure 13, where true R_0 value is marked with a black dot. It is interesting to observe that this frequency distribution is mainly concentrated into the interval $(2.604502, 3.480875)$, whose lower and upper limits are the 2.5 % and 97.5 % probability quantiles, respectively, and this interval does not include true $R_0 = 3.752587$ value.

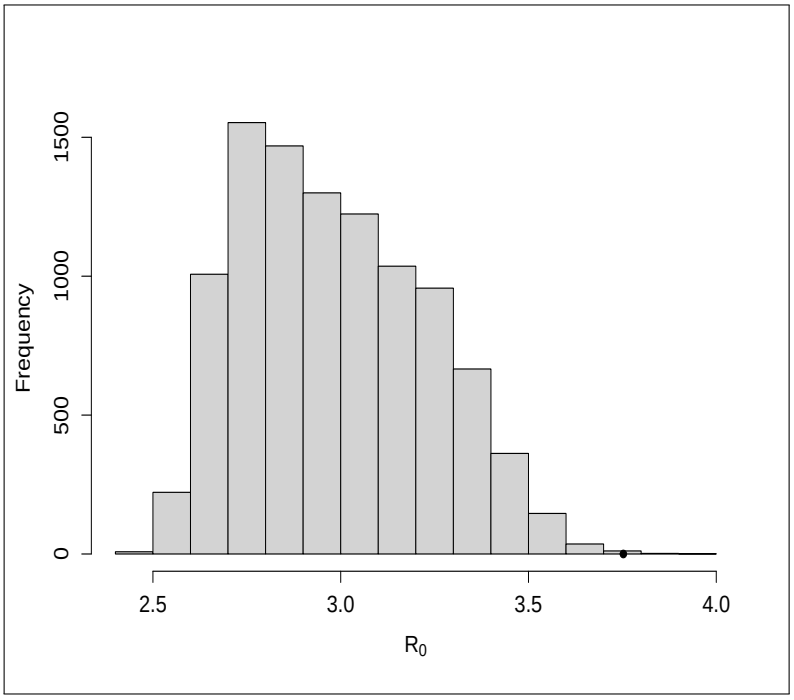


Figure 13: Histogram of R_0 values obtained from the simulation of 1000 random samples of β and δ. True R_0 value is marked with a black dot. Elaborated by the authors.

6. CONCLUSIONS

The shape of the likelihood function helps to detect practical parameter identifiability problems, revealing the intrinsic relationship of model parameters, in light of the observed sample. Moreover, the profile likelihood function allows to visualize and analyze the severity of the identifiability problem for a scalar parameter of interest, associated to the model under study, such as R_0, the basic reproductive number.
In the cases studied here, narrow likelihood surfaces with elongated or curved contours exhibit the problem of practical identifiability, whenever one of their tails does not stop decreasing and the surface plot appears to be truncated or cutoff within the parameter space. In general, much of the corresponding trajectory of the ridge or peak of the likelihood surface is nearly flat or constant, assigning to these points a plausibility as high as that of the MLE. Thus, fixing one parameter and maximizing the likelihood surface regarding some other one, is the way how the profile likelihood provides information about the ridge likelihood surface flatness.

In our simulation study, occurrence of the practical identifiability problem did not affect the coverage frequencies of the likelihood-confidence regions, that were always close to the nominal coverage probability.

On the other hand, the vast majority of practical parameter identifiability problems occurred when the observed sample of the SIR-Poisson model came from the initial and increasing part of the mean of the model, over time. In these cases, flat likelihoods can be interpreted in the sense that data do not provide sufficient information to discern about the shape of the average model behavior, over time. Thus, there are many parameter combinations yielding average behaviors that make the observed sample as likely as the average behavior associated with the MLE.

The methods discussed here for dealing with the practical identifiability problem should be cautiously used. Setting parameters can lead to false inference precision, also called overestimation or spurious accuracy. On the other hand, parameter estimation in systems of differential equations, with lack of consideration on the mathematical structure and the underlying parameters relationship involved in such a model, in light of the data, may lead to inappropriate inferences.

Perhaps, a strategy for obtaining prior information concerning the parameters of the model, in order to rescue their relationship, would be to consider general restrictions regarding the characteristics of the phenomenon under study, this could allow to generate feasible solutions of the system of differential equations, and from these to determine the values of the parameters that generated them, and use that information to propose a joint prior distribution. There is also the possibility of using this joint prior distribution in a Bayesian inference framework.

Extreme care should be taken when applying the estimation practices identified in this article, as they may lead to inappropriate conclusions about the basic reproductive number value, as well as the scenarios projections of an epidemic, among others. The latter becomes highly relevant when resulting estimates are used in public health decision making.
Perhaps the support and knowledge of experts in the field of the phenomenon being modeled can help to make better decisions to identify the relationship between the parameters of the model, and hence achieving the formulation of more appropriate hypotheses regarding the phenomenon under study.

References

Acuña-Zegarra, M. A., Díaz-Infante, S., Baca-Carrasco & D., Olmos-Liceaga, D. (2021). COVID-19 optimal vaccination policies: a modeling study on efficacy, natural and vaccine-induced immunity responses. *Mathematical Biosciences*, 6337, 108614.

Acuña-Zegarra, M. A., Santana-Cibrian, M. & Velasco-Hernández, J. X. (2020). Modeling behavioral change and COVID-19 containment in Mexico: A trade-off between lockdown and compliance. *Mathematical Biosciences*, 325, 108370.

Arino, J. & Van den Driessche, P. (2003). A multi-city epidemic model. *Mathematical Population Studies*, 10(**3**), 175-193.

Barndorff-Nielsen, O.E. & Cox, D.R. (1994). Inference and asymptotics. Chapman & Hall/CRC. Boca Raton.

Camacho, A., Kucharski, A. J., Funk, S., Breman, J., Piot, P. & Edmunds, W. J. (2014). Potential for large outbreaks of Ebola virus disease. *Epidemics*, 9, 70-78.

Capistrán, M.A., Christen, J. A. & Velasco-Hernández, J. X. (2012). Towards uncertainty quantification and inference in the stochastic SIR epidemic model. *Mathematical Biosciences*, 240(**2**), 250-259.

Chowell, G., Diaz-Dueñas, P., Miller, J. C., Alcazar-Velazco, A., Hyman, J. M., Fenimore, P. W.& Castillo-Chavez, C. (2007). Estimation of the reproduction number of dengue fever from spatial epidemic data. *Mathematical Biosciences*, 208(**2**), 571-589.

Chowell, G., Torre, C. A., Munayco-Escate, C., Suarez-Ognio, L., Lopez-Cruz, R., Hyman, J. M. & Castillo-Chavez, C. (2008). Spatial and temporal dynamics of dengue fever in Peru: 1994?2006. *Epidemiology & Infection*, 136(**12**), 1667–1677.

Chowell, G., Towers, S., Viboud, C., Fuentes, R., Sotomayor, V., Simonsen, L.,Miller, M. A., Lima, M., Villarroel, C., Chiu, M., Villarroel, J. E. & Olea, A. (2012). The influence of climatic conditions on the transmission dynamics of the 2009 A/H1N1 influenza pandemic in Chile. *BMC Infectious Diseases*, 12(**1**), 1-12.

Cole, D. J. (2020). Parameter redundancy and identifiability. Chapman & Hall/CRC. Boca Raton.

Cosner, C. (2015). Models for the effects of host movement in vector-borne disease systems. *Mathematical Biosciences*, 270, 192-197.

Funk, S., Kucharski, A. J., Camacho, A., Eggo, R. M., Yakob, L., Murray, L. M. & Edmunds, W. J. (2016). Comparative Analysis of Dengue and Zika Outbreaks Reveals Differences by Setting and Virus. *PLoS Neglected Tropical Diseases*, 10(**12**), e0005173.

Gábor, A., Villaverde, A. F.& Banga, J. R. (2017). Parameter identifiability analysis and visualization in large-scale kinetic models of biosystems. *BMC Systems Biology*, 11(**1**), 1-16.

Ghosh, I., Sardar, T.& Chattopadhyay, J. (2017). A Mathematical Study to Control Visceral Leishmaniasis: An Application to South Sudan. *Bulletin of Mathematical Biology*, 79(**5**), 1100-1134.

Ghosh, I., Tiwari, P. K., Samanta, S., Elmojtaba, I. M., Al-Salti, N. & Chattopadhyay, J. (2018). A simple SI-type model for HIV/AIDS with media and self-imposed psychological fear. *Mathematical Biosciences*, 306, 160-169.

Guanghu, Z., Tao, L., Jianpeng, X., Bing, Z., Tie, S., Yonghui, Z., Lifeng, L., Zhiqiang, P., Aiping, D., Wenjun, M. & Yuantao, H. (2019). Effects of human mobility, temperature and mosquito control on the spatiotemporal transmission of dengue. *Science of The Total Environment*, 651, 969-978.

Gui-Quan, S., Jun-Hui, X., Sheng-He, H., Zhen, J,; Ming-Tao, L. & Liqun, L. (2017). Transmission dynamics of cholera: Mathematical modeling and control strategies. *Communications in Nonlinear Science and Numerical Simulation*, 45, 235-244.

Hendron, R. W. S. & Bonsall, M. B. (2016). The interplay of vaccination and vector control on small dengue networks. *Journal of Theoretical Biology*, 407, 349-361.

Kalbfleisch, J. G. (1985). Probability and Statistical Inference, Vol. 2. Springer-Verlag. New York.

Kao, Y. H. & Eisenberg, M. C. (2018). Practical unidentifiability of a simple vector-borne disease model: Implications for parameter estimation and intervention assessment. *Epidemics*, 25, 89-100.

Kermack, W. O. & McKendrick, A. G. (1927). Contribution to the mathematical theory of epidemics. *Proccedings of the Royal Society A*, 115(**772**), 700–721.

Khan, A., Hassan, M. & Imran, M. (2014). Estimating the basic reproduction number for single-strain dengue fever epidemics. *Infectious Diseases of Poverty*, 3(**1**), 1-17.

Kim, J. E., Lee, H., Lee, C. H. & Lee, S. (2017). Assessment of optimal strategies in a two-patch dengue transmission model with seasonality. *PLoS ONE*, 12(**3**), e0173673.

Lee, S. & Castillo-Chavez, C. (2015). The role of residence times in two-patch dengue transmission dynamics and optimal strategies. *Journal of Theoretical Biology*, 374, 152-164.

Lloyd-Smith, J. O. (2007). Maximum likelihood estimation of the negative binomial dispersion parameter for highly overdispersed data, with applications to infectious diseases. *PLoS ONE*, 2(**2**), e180.

Ma, J. (2020). Estimating epidemic exponential growth rate and basic reproduction number. *Infectious Disease Modelling*, 5, 129-141.

Marquis, A.D., Arnold, A., Dean-Bernhoft, C., Carlson, B.E. & Olufsen, M. S. (2018). Practical identifiability and uncertainty quantification of a pulsatile cardiovascular model. *Mathematical Biosciences*, 304, 9-24.

Mishra, A., Ambrosio, B., Gakkhar, S. & Aziz-Alaoui, M. A. (2018). A network model for control of dengue epidemic using sterile insect technique. *Mathematical Biosciences & Engineering*, 15(**2**), 441-460.

Mishra, A. & Gakkhar, S. (2018). Non-linear dynamics of two-patch model incorporating secondary dengue infection. *International Journal of Applied and Computational Mathematics*, 4(**19**), 1-22.

Murphy, S. A. & Van Der Vaart, A. W. (2000). On profile likelihood. *Journal of the American Statistical Association*, 95(**450**), 449-465.

Nguyen, V. K., Parra-Rojas, C. & Hernandez-Vargas, E. A. (2018). The 2017 plague outbreak in Madagascar: Data descriptions and epidemic modelling. *Epidemics*, 25, 20-25.

Núñez-López, M., Ramos, L. A. & Velasco-Hernández, J. X. (2021). Migration rate estimation in an epidemic network. *Applied Mathematical Modelling*, 89, 1949-1964.

Pandey, A., Mubayi, A. & Medlock, J. (2013). Comparing vector-host and SIR models for dengue transmission. *Mathematical Biosciences*, 246(**2**), 252-259.

Pawitan, Y. (2001). In All Likelihood: Statistical Modelling and Inference Using Likelihood. Oxford University Press. New York.

Phaijoo, G. R. & Gurung, D. B. (2016). Mathematical study of dengue disease transmission in multi-patch environment. *Applied Mathematics*, 7(**14**), 1521-1533.

Qi, L., Xue, M., Cui, J.A., Wang, Q. & Wang, T. (2018). Schistosomiasis transmission model and its control in Anhui province. *Bulletin of Mathematical Biology*, 80(**9**), 2435-2451.

Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U. & Timmer, J. (2009). Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, 25(**15**), 1923-1929.

Rosenbaum, E. A., Pechen De D'angelo, A. M., Bergoc, R. M. & Venturino, A. (1999). Modelling acetylcholinesterase kinetics: The identifiability problem in parameter estimation. *Journal of Biological Systems*, 7(**01**), 95-111.

Saccomani, M. P. & Thomaseth, K. (2018). The union between structural and practical identifiability makes strength in reducing oncological model complexity: a case study. *Complexity*, 2018.

Saldaña, F., Flores-Arguedas, H., Camacho-Gutiérrez, J. A. & Barradas, I. (2020). Modeling the transmission dynamics and the impact of the control interventions for the COVID-19 epidemic outbreak. *Mathematical Biosciences and Engineering*, 17(**4**), 4165-4183.

Sasmal, S. K., Ghosh, I., Huppert, A. & Chattopadhyay, J. (2018). Modeling the spread of Zika virus in a stage-structured population: effect of sexual transmission. *Bulletin of Mathematical Biology*, 80(**11**), 3038-3067.

Serfling, R. J. (2002). Approximation Theorems of Mathematical Statistics. John Wiley & Sons. New York.

Sprott, D. A. (2000), Statistical inference in science. Springer-Verlag. New York.

Tocto-Erazo, M. R., Espíndola-Zepeda, J. A., Montoya-Laos, J. A., Acuña-Zegarra, M. A., Olmos-Liceaga, D., Reyes-Castro, P. A. & Figueroa-Preciado, G. (2020), Lockdown, relaxation, and acme period in COVID-19: A study of disease dynamics in Hermosillo, Sonora, Mexico. *PLoS ONE*, 15(**12**), e0242957.

Tocto-Erazo, M. R., Olmos-Liceaga, D. & Montoya, J. A. (2021). Effect of daily periodic human movement on dengue dynamics: The case of the 2010 outbreak in Hermosillo, Mexico. *Applied Mathematical Modelling*, 97, 559-567.

Towers, S., Brauer, F., Castillo-Chavez, C., Falconar, A.K., Mubayi, A. & Romero-Vivas, C. M. (2016). Estimate of the reproduction number of the 2015 Zika virus outbreak in Barranquilla, Colombia, and estimation of the relative role of sexual transmission. *Epidemics*, 17, 50-55.

Tuncer, N., Gulbudak, H., Cannataro, V. L. & Martcheva, M. (2016). Structural and practical identifiability issues of immuno-epidemiological vector-host models with application to rift valley fever. *Bulletin of Mathematical Biology*, 78(**9**), 1796-1827.

Tuncer, N., Mohanakumar, C., Swanson, S. & Martcheva, M. (2018). Efficacy of control measures in the control of Ebola, Liberia 2014-2015. *Journal of Biological Dynamics*, 12(**1**), 913-937.

Vinh, D. N., Ha, D.T.M., Hanh, N.T., Thwaites, G., Boni, M. F., Clapham, H. E. & Thuong, N. T. T. (2018). Modeling tuberculosis dynamics with the presence of hyper-susceptible individuals for Ho Chi Minh City from 1996 to 2015. *BMC Infectious Diseases*, 18(**1**), 1-13.

Xiao, Y. & Zou, X. (2014). Transmission dynamics for vector-borne diseases in a patchy environment. *Journal of Mathematical Biology*, 69(**1**), 113-146.

Zhan, C., Li, B.Y.S. & Yeung, L.F. (2015). Structural and practical identifiability analysis of S-system. *IET Systems Biology*, 9(**6**), 285-293.

Revista de la Facultad de Ciencias Universidad Nacional de Colombia, Sede Medellín
V 11 N°2 julio-diciembre de 2022 • ISSN-e 2357-5749 • Artículo Investigación • Páginas 100 a 147
DOI: https://doi.org/10.15446/rev.fac.cienc.v11n2.99198

GRÁFICOS EXISTENCIALES PARACONSISTENTES ALFAK[a]

ALFAK PARACONSISTENT EXISTENTIAL GRAPHS

MANUEL SIERRA-ARISTIZÁBAL[b] *

Recibido 25-10-2021, aceptado 06-04-2022, versión final 16-05-2022
Artículo Investigación

RESUMEN: En este trabajo, se presentan los gráficos existenciales para el cálculo proposicional paraconsistente, $KT4P$. Este sistema deductivo, es un fragmento de la lógica proposicional modal S4, y se construye a partir del cálculo proposicional clásico positivo, junto con un operador de negación débil. El sistema $KT4P$, se encuentra caracterizado por una semántica de mundos posibles, además, $KT4P$ es paraconsistente, es decir, no colapsa en la presencia de contradicciones. Los gráficos existenciales para este sistema paraconsistente, se presentan en el estilo de los gráficos alfa de Charles Sanders Peirce, junto con reglas para bucles y rizos similares a las presentadas recientemente por Arnold Oostra, para los gráficos existenciales intuicionistas. Todas las pruebas, son presentadas de manera completa, rigurosa y detallada.
PALABRAS CLAVE: Gráficos existenciales; lógica paraconsistente; semántica de mundos posibles; afirmación fuerte; negación débil.

ABSTRACT: In this paper, the existential graphs for the paraconsistent propositional calculus, $KT4P$, are presented. This deductive system is a fragment of the modal propositional logic S4, and is constructed from the positive classical propositional calculus, together with a weak negation operator. The $KT4P$ system is characterized by a semantics of possible worlds, in addition, $KT4P$ is paraconsistent, that is, it does not collapse in the presence of contradictions. The existential graphs for this paraconsistent system are presented in the style of the alpha graphics of Charles Sanders Peirce (late nineteenth century), along with rules for loops and scroll similar to those recently introduced (early twenty-first century) by Arnold Oostra, for intuitionistic existential graphics. All evidence is presented in a complete, rigorous and detailed manner.
KEYWORDS: Existential graphs; paraconsistent logic; semantics of possible worlds; strong affirmation; weak negation.

1. INTRODUCCIÓN

Los gráficos existenciales, alfa, beta y gama, fueron creados por Charles Sanders Peirce a finales del siglo XIX, ver Roberts (1992) y Peirce (1965). Los gráficos alfa corresponden al cálculo proposicional clásico, los gráficos beta corresponden a la lógica clásica de relaciones de primer orden. Los gráficos gama fueron presentados por Peirce y, posteriormente, extendidos por Jay Zeman, construyendo gráficos existenciales

[a]Sierra-Aristizábal, M. (2022). Gráficos existenciales paraconsistentes AlfaK. *Rev. Fac. Cienc.*, 11 (**2**), 100–147. DOI: https://doi.org/10.15446/rev.fac.cienc.v11n2.99198
[b]Universidad EAFIT, ORCID: 0000-0001-8157-2577
*Autor para la correspondencia: msierra@eafit.edu.co, mahesiar@outlook.com

para las lógicas modales S4, S4.2 y S5 en Zeman (1963). Por otro lado, Geraldine Brady y Todd Trimble, han propuesto modelos categóricos para los gráficos existenciales alfa en Brady & Trimble (2000). Recientemente, Arnold Oostra presentó gráficos existenciales para el cálculo proposicional intuicionista en Oostra (2010) y Oostra (2021), para el cálculo de relaciones intuicionista en Oostra (2011), y para las lógicas modales *S*4, *S*4.2 y *S*5, versiones intuicionistas, en Oostra (2012).

Los gráficos existenciales tienen una extraordinaria característica, citando a Fernando Zalamea:
"La axiomatización del cálculo proposicional clásico y de la lógica clásica de primer orden puramente relacional, con las mismas reglas, por medio de los sistemas Alfa y Beta, explicita raíces técnicas comunes desapercibidas en las presentaciones tradicionales de la lógica clásica. En efecto, las mismas reglas detectan, en el contexto del lenguaje Alfa, un manejo proposicional, y en el contexto extendido del lenguaje Beta, un manejo cuantificacional. Esta situación es toda una revelación en la historia de la lógica. Constituye, de manera precisa, la única presentación conocida de los cálculos clásicos que utiliza globalmente las mismas reglas axiomáticas para controlar el manejo local de los conectivos y de los cuantificadores" en Zalamea (2010).

En el presente trabajo, se presentan los gráficos existenciales para el cálculo proposicional paraconsistente $KT4P$. Este sistema deductivo, es un fragmento de la lógica proposicional modal *S*4, y se construye a partir del cálculo proposicional clásico positivo (es decir, el cálculo proposicional clásico sin los axiomas para la negación clásica), junto con un operador de negación débil (el cual, semánticamente se comporta como la posibilidad de la negación clásica). El sistema $KT4P$, se encuentra caracterizado por una semántica de mundos posibles, y además, es paraconsistente (es decir, no colapsa en la presencia de contradicciones). Los gráficos existenciales para este sistema paraconsistente, se presentan en el estilo de los gráficos alfa de Charles Sanders Peirce, junto con reglas para bucles y rizos, similares a las presentadas por Arnold Oostra, para los gráficos existenciales intuicionistas. Todas las pruebas son presentadas de manera completa, rigurosa y detallada.

La caracterización semántico deductiva de los gráficos existenciales paraconsistentes, se logra mediante la siguiente estrategia: En la sección 2, se presenta el sistema deductivo de lógica proposicional modal $KT4P$. En la sección 3, se presenta la semántica de mundos posibles para este sistema. En la sección 4, se presenta la caracterización de $KT4P$ con la semántica de la sección 3, esto se logra en la proposición 4.8 (los teoremas de $KT4P$ son las fórmulas válidas de la semántica y solo ellas). En la sección 5, se presenta la versión inicial de los gráficos existenciales paraconsistentes, el sistema AlfaK. En la sección 6, se presenta la equivalencia entre $KT4P$ y AlfaK, inicialmente, en la proposición 6.2, se prueba que los teoremas de $KT4P$ son teoremas gráficos de AlfaK, en la proposición 6.4, se prueba que los teoremas gráficos de $KT4P$ son válidos en la semántica de mundos posibles, finalmente, en la proposición 6.8, se prueba que los teoremas de $KT4P$ son exactamente los teoremas gráficos. En la sección 7, se presenta el sistema AlfaK_P, de gráficos existenciales paraconsistentes, en el formato original de Peirce (borrado en regiones pares, escritura en regiones impares,

iteración, desiteración), con reglas adicionales para los rizos y bucles, en la proposición 7.4, se prueba que los teoremas gráficos de AlfaK son teoremas gráficos de AlfaK_P, en la proposición 7.5, se prueba la recíproca, concluyéndose en la proposición 7.6 la equivalencia de los mismos. En la sección 8, se presentan algunas conexiones de AlfaK_P con los gráficos existenciales gama de Peirce y con el sistema Gamma-LD presentado en Sierra (2021), también se define un operador de incompatibilidad, como versión del operador de buen comportamiento en las lógicas paraconsistentes, finalmente, se utiliza AlfaK para dar solución a una versión de la paradoja del mentiroso.

2. SISTEMA DE LÓGICA MODAL KT4P

En esta sección, se presenta el sistema deductivo de lógica proposicional modal $KT4P$, sus conexiones con el cálculo proposicional clásico, y algunos de sus teoremas.

Definición 2.1. *(Fórmulas de $KT4P$, negación local, afirmación fuerte y equivalencia material). El conjunto FK, de fórmulas de $KT4P$, se construye a partir de un conjunto FA, de fórmulas atómicas, y de la constante λ, de la siguiente manera.*

1. *$P \in FA$ implica $P \in FK$.*
2. *$\lambda \in FK$*
3. *$X \in FK$ implica $-X \in FK$.*
4. *$X, Y \in FK$ implica $X \bullet Y, X \cup Y, X \supset Y \in FK$.*
5. *Solo $1, 2, 3$ y 4 determinan FK.*

NL (Negación local) $\sim X = X \supset -\lambda$. *AF (Afirmación fuerte)* $+X = \sim -X = -X \supset -\lambda$.
EQ (Equivalencia material) $X \equiv Y = (X \supset Y) \bullet (Y \supset X)$

Definición 2.2. *(Sistema deductivo $KT4P$). El sistema $KT4P$ consta de los axiomas (donde $X, Y, Z \in FK$):*

Axλ. λ

Ax1. $X \supset (Z \supset X)$

Ax2. $(X \supset (Y \supset Z)) \supset ((X \supset Y) \supset (X \supset Z))$.

Ax3. $X \supset (X \cup Y)$.

Ax4. $X \supset (Y \cup X)$.

Ax5. $(X \supset Y) \supset ((Z \supset Y) \supset ((X \cup Z) \supset Y))$.

Ax6. $(X \bullet Y) \supset X$.

Ax7. $(X \bullet Y) \supset Y$.

Ax8. $(X \supset Y) \supset ((X \supset Z) \supset (X \supset (Y \bullet Z)))$.

Ax9. $((X \supset Y) \supset X) \supset X$.

Ax10. $X \cup -X$.

Ax11. $(-X \supset -\lambda) \supset (-Y \supset -(X \supset Y))$.

Ax12. $-(-X \supset -\lambda) \supset -X$.

Ax13. $(-X \supset -\lambda) \supset (-X \supset Y)$.

Ax14. Si X es un axioma entonces $-X \supset -\lambda$ *es un axioma.*

Como única regla de inferencia se tiene el modus ponens Mp: de X y $X \supset Z$ *se infiere Z.*

Observación 2.1. *El axioma Ax*λ *puede omitirse, y simplemente se define* $\lambda = P \supset P$.

Definición 2.3. *(Teoremas de KT4P). Sean* $X, X_1, \ldots, X_n \in FK$. *X es un teorema de KT4P, denotado* $X \in TK$, *si existe una demostración de X a partir de los axiomas utilizando la regla Mp, es decir, X es la última fila de una secuencia finita de líneas, en la cual, cada una de las líneas es un axioma, o se infiere de dos filas anteriores, utilizando la regla de inferencia Mp. El número de líneas de la secuencia es referenciado como la longitud de la demostración de X. Y es un teorema (o consecuencia) de* $\{X_1, \ldots, X_n\}$, *denotado* $Y \in TK \cup \{X_1, \ldots, X_n\}$, *si existe una demostración de Y, a partir de los axiomas y de los supuestos* $\{X_1, \ldots, X_n\}$.

Hecho 2.1. *(Teorema de deducción en KT4P). Sean* $X, Y, X_1, \ldots, X_n \in FK$.

a. TD (Teorema de deducción): Si Y es consecuencia de $\{X, X_1, \ldots, X_n\}$ *en KT4P, entonces* $X \supset Y \in TK \cup \{X_1, \ldots, X_n\}$

b. Ax11 $\supset (+(X \supset Y) \supset (+X \supset +Y))$.

*c. Ax*12 $\equiv (-+X \supset -X)$.

*d. Ax*13 $\equiv (+X \supset (-X \supset Y))$.

*e. Ax*14 *equivale a, X es un axioma implica* $+X$ *es un axioma.*

Prueba. Parte *a*. Se prueba utilizando inducción matemática sobre la longitud, L, de la demostración de Y a partir de los supuestos $\{X, X_1, \ldots X_n\}$.
Paso base. $L = 1$. Se tienen tres casos, Y es un axioma, Y es X o $Y \in \{X_1, \ldots, X_n\}$, en el primer caso resulta que $Y \in TK$, utilizando $Ax1$ $Y \supset (X \supset Y)$ y Mp se concluye que $X \supset Y \in TK$. En el segundo caso, por $Ax2$ se tiene $(X \supset ((X \supset X) \supset X)) \supset ((X \supset (X \supset X)) \supset (X \supset X)) \in TK$, por $Ax1$ se tienen $X \supset ((X \supset X) \supset X) \in TK$ y $X \supset (X \supset X) \in TK$, aplicando Mp dos veces se deriva $X \supset X \in TK$, y como X es Y se concluye que $X \supset Y \in TK$. Lo anterior, junto con el tercer caso, implica que $X \supset Y \in TK \cup \{X_1, \ldots, X_n\}$.
Paso inductivo. Como hipótesis inductiva se tiene que, si la longitud de la demostración de W a partir de $\{X, X_1, \ldots, X_n\}$ es menor que L entonces $X \supset W \in TK \cup \{X_1, \ldots, X_n\}$.
Si la longitud de la demostración de Y a partir de $\{X, X_1, \ldots, X_n\}$ es L, se consideran cuatro casos, Y es un axioma, $Y \in \{X_1, \ldots, X_n\}$, Y es X o Y se infiere de fórmulas anteriores utilizando Mp. En los tres primeros casos se procede como en el paso base. En el cuarto caso, se tienen Z y $Z \supset Y$ (ambas conscuencias de

$\{X, X_1, \ldots, X_n\}$), por la hipótesis inductiva se infiere que $X \supset Z \in TK \cup \{X_1, \ldots, X_n\}$ y $X \supset (Z \supset Y) \in TK \cup \{X_1, \ldots, X_n\}$. Utilizando $Ax2$ $(X \supset (Z \supset Y)) \supset ((X \supset Z) \supset (X \supset Y)) \in TK \cup \{X_1, \ldots, X_n\}$, por Mp se sigue $(X \supset Z) \supset (X \supset Y) \in TK \cup \{X_1, \ldots, X_n\}$, de nuevo por Mp, se concluye $X \supset Y \in TK \cup \{X_1, \ldots, X_n\}$. Por el principio de inducción matemática, se ha probado que, si Y es consecuencia de $\{X, X_1, \ldots X_n\}$ en $KT4P$, entonces $X \supset Y \in TK \cup \{X_1, \ldots, X_n\}$.

Parte b. Supóngase que $+(X \supset Y)$ y $+X$. Por $Ax11$ y definición de $+$ se obtiene $+X \supset (-Y \supset -(X \supset Y))$, aplicando Mp se sigue $-Y \supset -(X \supset Y)$, como se tiene $+(X \supset Y)$, según definición de $+$ significa $-(X \supset Y) \supset -\lambda$ utilizando $Ax1$ y Mp se infiere $-Y \supset (-(X \supset Y) \supset -\lambda)$, por $Ax2$ y Mp se deriva $(-Y \supset -(X \supset Y)) \supset (-Y \supset -\lambda)$, como se tiene el antecedente, por Mp se deduce $-Y \supset -\lambda$, lo cual por la definición de $+$ es $+Y$. Aplicando la parte a (TD) dos veces, se concluye que $+(X \supset Y) \supset (+X \supset +Y) \in TK$.

Partes c, d y e. Consecuencia directa de aplicar la definición de $+$ en los axiomas 12, 13 y 14. □

Proposición 2.1. *(Afirmación fuerte de los teoremas). Sea $X \in FK$.*
$X \in TK$ implica $+X \in TK$.

Prueba. Supóngase que $X \in TK$, se probará que $+X \in TK$, por inducción sobre la longitud de la demostración de X.

Paso base. La longitud de la demostración de X es 1, es decir X es un axioma, pero si X es un axioma entonces, por $Ax14$, $+X$ es axioma, y, por lo tanto, $+X \in TK$.

Paso de inducción. Como hipótesis inductiva, se tiene que si la longitud de la demostración de Y es menor que L entonces $+Y$ es un teorema. Supóngase que la demostración de X tiene longitud $L > 1$. Se tiene entonces que X es un axioma o X es consecuencia de pasos anteriores utilizando la regla de inferencia Mp. En el primer caso se procede como en el paso base. En el segundo caso se tienen, para alguna fórmula Z, demostraciones de $Z \supset X$ y de Z, ambas de longitud menor que L. Aplicando la hipótesis inductiva resulta que $+(Z \to X)$,$+Z \in TK$. Por $Ax11$ y el Hecho 2.1b se tiene $+(Z \supset X) \supset (+Z \supset +X)$, aplicando dos veces la regla Mp resulta que $+X \in TK$. Por lo que, según el principio de inducción matemática, se ha probado $X \in TK$ implica $+X \in TK$. □

Hecho 2.2. *($KT4P$ incluye CPC). Sea $X \in FK$.*

a. $\sim X \supset (X \supset Y) \in TK$.

b. $(\sim X \supset X) \supset X \in TK$.

c. Los teoremas del cálculo proposicional clásico, CPC, son teoremas de $KT4P$.

Prueba. Parte a. Supóngase que $\sim X$, es decir $X \supset -\lambda$. Supuesto 2, X, por Mp se infiere $-\lambda$. Por $Ax\lambda$ se tiene que $\lambda \in TK$, por la proposición 2.1 se sigue que $+\lambda \in TK$. Por $Ax13$ y el Hecho 2.1d se tiene $+\lambda \supset (-\lambda \supset Y)$, aplicando Mp dos veces se infiere Y. Por TD (Hecho 2.1 a) dos veces se concluye $\sim X \supset (X \supset Y)$.

Parte b. Supóngase que $\sim X \supset X$, por definición de $\sim$, se sigue $(X \supset -\lambda) \supset X$, por $Ax9$ $((X \supset -\lambda) \supset X) \supset X$

y Mp se sigue X. Aplicando TD (Hecho 2.1 a), se concluye que $(\sim X \supset X) \supset X$.
Parte c. Los axiomas $Ax1,\ldots,Ax9$, junto con los Hechos 2.2 a y 2.2 b, y con la regla de inferencia Mp, constituyen una axiomatización del cálculo proposicional clásico, *CPC*, presentada por Alfred Tarski en 1952, ver Doop (1969). Por lo tanto, los teoremas del cálculo proposicional clásico, *CPC*, son teoremas de $KT4P$. □

Hecho 2.3. *(Resultados Clásicos en KT4P). En lo que sigue se utilizaran los siguientes resultados del cálculo proposicional clásico CPC, los cuales, por el Hecho 2.2c, valen en KT4P (para detalles de las pruebas en CPC ver Sierra (2010) y Hamilton (1978)).*

Sean $X,Y,Z \in FK$.

I•. Introducción de •: de X y Y se infiere $X \bullet Y$.	*E•. Eliminación de •: de* $X \bullet Y$ *se infieren X y Y.*
SH. Silogismo hipotético: de $X \supset Y$ *y* $Y \supset Z$ *se infiere* $X \supset Z$.	*Exp. Exportación:* $(X \supset (Y \supset Z)) \equiv ((X \bullet Y) \supset Z)$.
N∪. Negación de ∪: $\sim (X \cup Y) \equiv (\sim X \bullet \sim Y)$.	*N•. Negación de •:* $\sim (X \bullet Y) \equiv (\sim X \cup \sim Y)$.
N ⊃. Negación de ⊃: $\sim (X \supset Y) \equiv (X \bullet \sim Y)$.	*Tras.Transposición:* $(X \supset Y) \equiv (\sim Y \supset \sim X)$.
Imp. Implicación: $(X \supset Y) \equiv (\sim X \cup Y)$.	*Id. Principio de identidad :* $X \supset X$.
PR. Principio de retorsión: de $(\sim X \supset X) \equiv X$.	*DN. Doble negación:* $X \equiv \sim\sim X$.

DI. Demostración indirecta: Si X implica $Y \bullet \sim Y$*, entonces, se infiere* $\sim X$.

Definición 2.4. *(Posibilidad en KT4P). Para* $X \in FK$. $\otimes X = \sim + \sim X = (-(X \supset -\lambda) \supset -\lambda) \supset -\lambda$.

Hecho 2.4. *(Demostración indirecta en KT4P). Sea* $X,Y \in FK$.

a. DI (Demostración indirecta): $-X \supset (\sim Z \bullet Z)$ *implica X.*

b. AF• (Afirmación fuerte de la conjunción): $+(X \bullet Y) \equiv (+X \bullet +Y)$.

c. N− (Negación local de la negación): $\sim -X \equiv +X$, $-X \equiv \sim +X$, $+X \equiv (-X \supset -\lambda)$.

Versiones del Ax10 $d. +X \supset X$. $e. X \supset \otimes X$. $f. -X \supset - +X$. $g. \sim X \supset -X$.

Versiones del Ax12. $h. -X \equiv - +X$, $-X \equiv -(-X \supset -\lambda)$. $i. +X \supset + +X$. $j. +X \equiv + +X$.

Versiones de ⊗

$k. \otimes \sim X \equiv -X, \otimes X \equiv - \sim X$.	$l. \sim \otimes \sim X \equiv +X, \otimes \sim X \equiv \sim +X$.
$ll. \otimes \otimes X \equiv \otimes X$.	$m. \sim \otimes X \equiv + \sim X$.
$n. Ax11 \equiv (+(X \supset Y) \supset (+X \supset +Y))$.	$o. -\lambda \supset Y$.

Prueba. Parte a. Supóngase que $-X \supset (\sim Z \bullet Z)$. Supuesto 2, $-X$, por Mp se sigue $\sim Z \bullet Z$, por $E\bullet$ se derivan $\sim Z$ y Z, utilizando $Ax3$ $Z \supset (Z \cup X)$ y Mp resulta $Z \cup X$, aplicando SD se infiere X. Según TD se obtiene $-X \supset X$, por Id se tiene $X \supset X$, utilizando $Ax5$ $(-X \supset X) \supset ((X \supset X) \supset ((X \cup -X) \supset X))$ y Mp dos veces se deduce $(X \cup -X) \supset X$, utilizando $Ax10$ $X \cup -X$ y Mp se concluye que X.
Parte b. Por $Ax6$, y $Ax7$ se tienen $(X \bullet Y) \supset X$ y $(X \bullet Y) \supset Y$, y por $Ax14$ resultan $+((X \bullet Y) \supset X)$ y $+((X \bullet Y) \supset Y)$. Utilizando el $Ax11$, Hecho 2.1 b y Mp se infieren $+(X \bullet Y) \supset +X$ y $+(X \bullet Y) \supset +Y$.

Utilizando $Ax8$ y Mp dos veces se infiere $+(X \bullet Y) \supset (+X \bullet +Y)$.
Por otro lado, supuesto 1, X, y supuesto 2, Y, por $I\bullet$ resulta $X \bullet Y$, aplicando TD dos veces se obtiene $X \supset (Y \supset (X \bullet Y))$. Por la proposición 2.1 se infiere $+(X \supset (Y \supset (X \bullet Y)))$, y utilizando $Ax11$, Hecho 2.1 b y Mp dos veces resulta $+X \supset +(Y \supset (X \bullet Y))$, como además por $Ax11$ y Hecho 2.1 b se tiene $+(Y \supset (X \bullet Y)) \supset (+Y \supset +(X \bullet Y))$, entonces por SH se obtiene $+X \supset (+Y \supset +(X \bullet Y))$, utilizando Exp se infiere $(+X \bullet +Y) \supset +(X \bullet Y)$. Como ya se probó la recíproca, entonces por parte $I\bullet$ y EQ resulta que $+(X \bullet Y) \equiv (+X \bullet +Y)$.
Parte c. Por la definición de $+$ se tiene $\sim -X \equiv +X$, aplicando $Tras$ y DN también se tiene $-X \equiv \sim +X$. Y también, de $\sim -X \equiv +X$, por la definición de $\sim$, resulta que $+X \equiv (-X \supset -\lambda)$.
Parte d. Por $Ax10$ se tiene $X \cup -X$, por Imp y DN resulta $\sim -X \supset X$, por la parte c, se concluye $+X \supset X$.
Parte e. Por la parte d, se tiene $+ \sim X \supset \sim X$, por $Tras$ y DN se sigue $X \supset \sim + \sim X$, lo cual por definición de $\otimes$, significa $X \supset \otimes X$.
Parte f. Por la parte c se tiene $-X \equiv \sim +X$, por $Ax10$ se tiene $+X \cup - +X$, aplicando Imp y DN se deriva $\sim +X \supset - +X$, según EQ, $E\bullet$ y SH se concluye $-X \supset - +X$.
Parte g. Por $Ax10$ se tiene $X \cup -X$, usando Imp y DN, se deriva $\sim X \supset -X$.
Parte h. Por $Ax12$ y el Hecho 2.1 c se tiene $- +X \supset -X$, de la parte f se tiene $-X \supset - +X$, aplicando EQ, se concluye $-X \equiv - +X$. Utilizado la parte c, se concluye que, $-X \equiv -(-X \supset -\lambda)$.
Parte i. Por $Ax12$ y el Hecho 2.1 c se tiene $- +X \supset -X$. Supuesto 1, +X. Supuesto 2, $\sim ++X$, por la parte c se tiene $-(+X) \equiv \sim +(+X)$, aplicando EQ, $E\bullet$ y Mp se obtiene $- +X$, por la parte h, se sabe que $-X \equiv - +X$, por lo que se infiere $-X$, por la parte c se tiene $-X \equiv \sim +X$, y entonces por EQ y Mp resulta $\sim +X$, lo cual no es el caso, por DI se deriva, $++X$. Aplicando TD, se concluye que $+X \supset ++X$.
Parte j. Por la parte d, se tiene $+(+X) \supset (+X)$, y por la parte i se sabe que $+X \supset ++X$, utilizando $I \bullet$ y EQ resulta $+X \equiv ++X$.
Parte k. Por definición de $\otimes$, se tiene $\otimes(\sim X) \equiv \sim + \sim (\sim X)$, por DN se sigue $\otimes \sim X \equiv \sim +X$, por la parte c se tiene $-X \equiv \sim +X$. Aplicando SH se concluye $\otimes \sim X \equiv -X$. Tomando X como $\sim Z$, se obtiene $\otimes \sim\sim Z \equiv - \sim Z$, y por DN se deriva $\otimes Z \equiv - \sim Z$.
Parte l. Por la parte k se tiene $\otimes \sim X \equiv -X$, aplicando $Tras$ se sigue $\sim \otimes \sim X \equiv \sim -X$, por la parte c se tiene $\sim -X \equiv +X$, utilizando SH se concluye $\sim \otimes \sim X \equiv +X$. Aplicando $Tras$ y DN también se tiene $\otimes \sim X \equiv \sim +X$.
Parte ll. Por la parte j, se tiene $+(\sim X) \equiv ++(\sim X)$, por la parte l, se tiene $\sim \otimes \sim X \equiv +X$, resultando que $\sim \otimes \sim (\sim X) \equiv \sim \otimes \sim (\sim \otimes \sim (\sim X))$, por DN se deriva $\sim \otimes X \equiv \sim \otimes \otimes X$, finalmente por $Tras$ resulta $\otimes X \equiv \otimes \otimes X$.
Parte m. Por la definición de $\otimes$ se tiene $\otimes X \equiv \sim + \sim X$, aplicando $Tras$ y DN se obtiene $\sim \otimes X \equiv + \sim X$.
Parte n. Supóngase que $+(X \supset Y) \supset (+X \supset +Y)$, por $Tras$ se deriva $\sim (+X \supset +Y) \supset \sim +(X \supset Y)$, aplicando $N \supset$ resulta $(+X \bullet \sim +Y) \supset \sim +(X \supset Y)$, lo cual por la definición de $+$ y el Hecho 2.4 c significa $((-X \supset -\lambda) \bullet -Y) \supset -(X \supset Y)$, finalmente, por Exp de deriva $(-X \supset -\lambda) \supset (-Y \supset -(X \supset Y))$, y como por el Hecho 2.1 b se tiene la recíproca, entonces por $I\bullet$ y la definición de $\equiv$ se concluye que $Ax11 \equiv (+(X \supset Y) \supset (+X \supset +Y))$.

Parte o. Supóngase $-\lambda$, por $Ax\lambda$ se tiene λ, por $Ax14$ se deriva $+\lambda$, y como también se tiene $-\lambda$, utilizando $Ax13$ y Hecho 2.1d y Mp dos veces se infiere Y. □

Notación. Basado en el Hecho 2.1 partes c, d y e, y en el Hecho 2.4 parte n, las equivalencias de los axiomas $Ax11$ a $Ax14$, se utilizarán y referenciarán, indistintamente, como los mismos axiomas.

3. SEMÁNTICA PARA EL SISTEMA KT4P

En esta sección, se presenta la semántica de mundos posibles para el sistema $KT4P$, en la proposición 3.2, se prueba que los teoremas del sistema $KT4P$ son fórmulas válidas en la semántica propuesta.

Definición 3.1. *(Modelos para $KT4P$).*
$(S, Ma, <, V)$ es un modelo para $KT4P$, significa que, S es un conjunto no vacío de mundos posibles, Ma es un mundo posible, llamado mundo actual, $<$, es una relación binaria en S, V es una valuación de $S \times FK$ en $\{0,1\}$. La relación, $<$, satisface las siguientes restricciones:
Reflexividad de $<$. RR: $(\forall M \in S)\,(M < M)$.
Transitividad de $<$. RT: $(\forall K, M, N \in S)\,(K < M \,y\, M < N \,implica\, K < N)$.

Definición 3.2. *(Verdad en $KT4P$). En el modelo $(S, Ma, <, V)$, con $X, Y \in FK$.*
$V(M,X) = 1$ se abrevia como $M(X) = 1$, y significa que en el mundo posible M, la fórmula X es verdadera.
$V(M,X) = 0$ se abrevia como $M(X) = 0$, y significa que en el mundo posible M, la fórmula X es falsa.
La valuación V satisface las siguientes reglas:

1. $V\lambda$. $M(\lambda) = 1$.
2. $V \supset$. $M(X \supset Y) = 1$ *equivale a* $M(X) = 1$ *implica* $M(Y) = 1$.
3. $V\bullet$. $M(X \bullet Y) = 1$ *equivale a* $M(X) = M(Y) = 1$.
4. $V\cup$. $M(X \cup Y) = 1$ *equivale a* $M(X) = 1$ *o* $M(Y) = 1$.
5. $V-$. $M(-X) = 1$ *equivale a* $(\exists P \in S)\,(M < P \,y\, P(X) = 0)$.

Hecho 3.1. *(Verdad para la negación local, afirmación fuerte y posibilidad). Para $X \in FK$*

a. $V \sim$.$M(\sim X) = 1$ *equivale a* $M(X) = 0$.
b. $V +$.$M(+X) = 1$ *equivale a* $(\forall N \in S)\,(M < N \,implica\, N(X) = 1)$.
c. $M(+X) = 1$ *implica* $M(X) = 1$.
d. $M(-X) = 0$ *implica* $M(X) = 1$.
e. $V -+$.$M(-+X) = 1$ *equivale a* $M(+X) = 0$.

f. $V\otimes .M(\otimes X)=1$ *equivale a* $(\exists P\in S)\,(M<P\,y\,P(X)=1)$.

Prueba. Parte *a*. Por $V\lambda$ se tiene que para todo $M\in S$, en cualquier modelo, $M(\lambda)=1$, según $V-$ resulta que $M(-\lambda)=0$. Por definición de $\sim$, se tiene que, $M(\sim X)=1$ equivale a $M(X\supset -\lambda)=1$, lo cual equivale a $M(X)=1$ implica $M(-\lambda)=1$, como $M(-\lambda)=0$, entonces se infiere $M(X)=0$, por lo que, $M(\sim X)=1$ implica $M(X)=0$. Para la recíproca, si $M(X)=0$, por $V\supset$ se sigue que $M(X\supset -\lambda)=1$. Por lo tanto, $M(\sim X)=1$ equivale a $M(X)=0$.
Parte *b*. $M(+X)=1$ por la definición de $+$ equivale a $M(\sim -X)=1$, lo cual por la parte *a*, $V\sim$, equivale a $M(-X)=0$, y esto por $V-$ equivale a $(\forall N\in S)\,(M<N\,\text{implica}\,N(X)=1)$.
Parte *c*. Si $M(+X)=1$, entonces por la parte *b*, $V+$, se sigue $(\forall N\in S)\ (M<N\,\text{implica}\,N(X)=1)$, en particular, $M<M$ implica $M(X)=1$, por la restricción *RR* se tiene que $M<M$, por lo tanto, $M(X)=1$.
Parte *d*. Si $M(-X)=0$, entonces por $V-$ se tiene $(\forall N\in S)\,(M<N\,\text{implica}\,N(X)=1)$, en particular, $M<M$ implica $M(X)=1$, por la restricción *RR* se tiene que $M<M$, por lo tanto, $M(X)=1$.
Parte *e*. Si $M(-+X)=1$, por $V-$ equivale a $(\exists P\in S)\ (M<P\,\text{y}\,P(+X)=0)$, por la parte *b*, $V+$, $(\exists Q\in S)$ $(P<Q\,\text{y}\,Q(X)=0)$, como $M<P$ y $P<Q$ por la restricción *RT* se sigue que $M<Q$, aplicando $V+$ se deriva $M(+X)=0$. Para probar la recíproca, supóngase que $M(+X)=0$, por la restricción *RR* se tiene que $M<M$, luego, por $V-$ se deduce $M(-+X)=1$.
Parte *f*. $M(\otimes X)=1$, por definición de $\otimes$ equivale a $M(\sim +\sim X)=1$, por $V\sim$ equivale a $M(+\sim X)=0$, utilizando $V+$ equivale a $(\exists P\in S)\ (M<P\,\text{y}\,P(\sim X)=0)$, lo cual por $V\sim$ equivale a $(\exists P\in S)$ $(M<P\,\text{y}\,P(X)=1)$. □

Definición 3.3. *(Validez en KT4P). Para* $X,X_1,\ldots,X_n\in FK$*, se dice que una fórmula X es válida, denotado* $X\in VK$*, si y solo sí X es verdadera en todos los modelos para KT4P, es decir, X es verdadera en el mundo actual de todos los modelos para KT4P. Se dice que,* $\{X_1,\ldots,X_n\}$ *valida a Y si y solo sí* $(X_1\bullet X_2\bullet\ldots\bullet X_n)\supset Y\in VK$.

Proposición 3.1. *(Validez de los axiomas). Sea* $X\in FK$.
Si X es un axioma de KT4P entonces $X\in VK$.

Prueba. Parte 1. $Ax\lambda$. λ. Por $V\lambda$ se tiene para todo $M\in S$, $V(\lambda)=1$. Por lo tanto, $Ax\lambda\in VK$.
Parte 2. $Ax1,\ldots,Ax9$. Si X es uno de los axiomas $Ax1,\ldots,Ax9$, utilizando las reglas $V\bullet$, $V\cup$, $V\supset$ y procediendo como es habitual para la validez del cálculo proposicional clásico en Sierra (2010) y Hamilton (1978), se concluye que $X\in VK$, es decir, $Ax1,\ldots,Ax9\in VK$.
Parte 3. $Ax10$. Supóngase que $X\cup -X\notin VK$, por lo que existe un modelo, tal que en el mundo actual M, $M(X\cup -X)=0$, por $V\cup$ resulta $M(X)=0$ y $M(-X)=0$, utilizando $V-$ resulta que para cada $N\in S$, $M<N$ implica $N(X)=1$, por la restricción *RR*, $M<M$, y entonces $M(X)=1$, lo cual no es el caso. Por lo tanto, $Ax10\in VK$.
Parte 4. $Ax11$. Supóngase que $+(X\supset Y)\supset(+X\supset +Y)\notin VK$, por lo que existe un modelo, tal que en el mundo actual M, $M(+(X\supset Y)\supset(+X\supset +Y))=0$, por $V\supset$ se sigue $M(+(X\supset Y))=1$ y $M(+X\supset +Y)=0$, por $V\supset$ se deriva $M(+X)=1$ y $M(+Y)=0$, por $V+$, existe $P\in S$, $M<P$ y $P(Y)=0$, como $M(+(X\supset Y))=1$ por $V+$ se tiene que para todo $N\in S$, $M<N$ implica $M(X\supset Y)=1$, y como

$M < P$, resulta $P(X \supset Y) = 1$, como $P(Y) = 0$, por $V \supset$ se sigue $P(X) = 0$, lo cual por $V+$ implica $M(+X) = 0$, lo cual no es cierto. Por lo tanto, utilizando el Hecho 2.4n, $Ax11 \in VK$.
Parte 5. $Ax12$. Supóngase que $-+X \supset -X \notin VK$, por lo que existe un modelo, tal que en el mundo actual M, $M(-+X \supset -X) = 0$, por $V+$ se sigue $M(-+X) = 1$ y $M(-X) = 0$. Como $M(-+X) = 1$ existe $P \in S$, $M < P$ y $P(+X) = 0$, por $V+$ existe $Q \in S$, $P < Q$ y $Q(X) = 0$, se tiene $M < P$ y $P < Q$, por la restricción RT se afirma $M < Q$. Se tiene $M(-X) = 0$, aplicando $V-$, para cada $N \in S$, $M < N$ implica $N(X) = 1$, en particular, como $M < Q$, entonces $Q(X) = 1$, lo cual no es cierto. Por lo tanto, utilizando el Hecho 2.1c, $Ax12 \in VK$.
Parte 6. $Ax13$. Supóngase que $M(+X) = 1$, por $V+$ se tiene que, para todo $N \in S$, $M < N$ implica $N(X) = 1$. Supuesto 2, $M(-X) = 1$, por $V-$, existe $P \in S$, $M < P$ y $P(X) = 0$. Supuesto 3, $M(Y) = 0$, se tiene $M < P$, por lo que, del primer resultado se infiere $P(X) = 1$, lo cual no es el caso, por lo que, $M(Y) = 1$. Por TD y el Hecho 2.1 d, $Ax13 \in VK$.
Parte 7. $Ax14$. Supóngase que X es uno de los axiomas $Ax1$ a $Ax13$, por las partes 1 a 6, resulta que $X \in VK$. Supuesto 2, $+X \notin VK$, por lo que existe un modelo, tal que en el mundo actual M, $M(+X) = 0$, aplicando $V+$, existe $P \in S$, $M < P$ y $P(X) = 0$, lo cual significa que $X \notin VK$, lo cual no es el caso, por lo que $+X \in VK$. Por lo tanto, si X es axioma entonces $+X \in VK$, y utilizando el Hecho 2.1e, si X es axioma entonces $-X \supset -\lambda \in VK$. □

Proposición 3.2. *(Teorema de validez). Sean $X, Y \in FK$.*

a. Si $X \in TK$ entonces $X \in VK$.

b. Si Y es una consecuencia de $\{X_1, \ldots, X_n\}$ entonces $\{X_1, \ldots, X_n\}$ valida a Y.

Prueba. Parte a. Supóngase que $X \in TK$, se prueba que $X \in VK$ por inducción sobre la longitud, L, de la demostración de X.
Paso Base $L = 1$. Significa que X es un axioma, lo cual por la proposición 3.1 se concluye que $X \in VK$.
Paso de inducción. Como hipótesis inductiva, se tiene que para cada fórmula Y, si $Y \in TK$ y la longitud de la demostración de Y es menor que L (donde $L > 1$) entonces $Y \in VK$. Si $X \in TK$ y la longitud de la demostración de X es L entonces, X es un axioma o X es consecuencia de aplicar Mp en pasos anteriores de la demostración. En el primer caso se procede como en el caso base. En el segundo caso se tienen para alguna fórmula Y, demostraciones de Y y de $Y \supset X$, donde la longitud de ambas demostraciones es menor que L, utilizando la hipótesis inductiva se infiere que $Y \in VK$ y $Y \supset X \in VK$, por lo que en el mundo actual, M, de cualquier modelo se tienen $M(Y) = 1$ y $M(Y \supset X) = 1$, por $V \supset$ resulta que $M(X) = 1$, en consecuencia $X \in VK$. Utilizando el principio de inducción matemática, se ha probado que, para cada $X \in TK$, $X \in TK$ implica $X \in VK$.
Parte b. Supóngase que, Y es una *consecuencia* de $\{X_1, \ldots, X_n\}$, aplicando TD (Hecho 2.1) y exp, se tiene $(X_1 \bullet X_2 \bullet \ldots \bullet X_n) \supset Y \in TK$, por la parte a se infiere, $(X_1 \bullet X_2 \bullet \ldots \bullet X_n) \supset Y \in VK$, lo cual por definición significa que $\{X_1, \ldots, X_n\}$ valida a Y. □

4. CARACTERIZACIÓN SEMÁNTICO-DEDUCTIVA KT4P

En esta sección, se presenta la caracterización de $KT4P$ con la semántica de la sección 3. La completitud se prueba en la proposición 4.7 (las fórmulas válidas de la semántica son teoremas de $KT4P$), y la caracterización se logra en la proposición 4.8 (los teoremas de $KT4P$ son las fórmulas válidas de la semántica y solo ellas).

Definición 4.1. *(Extensión localmente consistente y completa)., (Una extensión de un conjunto de fórmulas C de $KT4P$, se obtiene alterando el conjunto de fórmulas de C de tal manera que, se preserven los teoremas de C, y que el lenguaje de la extensión coincida con el lenguaje de $KT4P$. Una extensión es localmente consistente si no existe ninguna $X \in FK$ tal que tanto X como $\sim X$ sean teoremas de la extensión. Un conjunto de fórmulas es localmente inconsistente si de ellas se deriva una contradicción local, es decir, se deriva $Z\bullet \sim Z$ para alguna $Z \in FK$. Una extensión es localmente completa si para toda $X \in FK$, o bien X es teorema de la extensión o bien $\sim X$ es teorema de la extensión. Para llegar a la demostración de completitud en la proposición 4.7, se sigue la estrategia del modelo canónico, presentada en Henkin (1949).*

Proposición 4.1. *(Extensión localmente consistente). Para $X \in FK$.*

a. $KT4P$ es localmente consistente.

b. Si $E \in EXT(K)$, $X \notin TK - E$ y $E_x \in EXT(K)$ se obtiene añadiendo $\sim X$ como nueva fórmula a E, entonces, E_x es consistente.

Prueba. Parte a. Supóngase que $KT4P$ no fuese localmente consistente, por lo que debe existir $X \in FK$ tal que $X, \sim X \in TK$, por la proposición 3.2, $X, \sim X \in VK$, pero esto es imposible, ya que si $\sim X \in VK$, entonces para todo modelo $(S, M_a, <, V)$, se tienen $V(M_a, \sim X) = 1$, es decir, según $V \sim$, $M_a(X) = 0$, por lo que $X \notin VK$, lo cual no es el caso. Por lo tanto, $KT4P$ es localmente consistente.
Parte b. Sea $X \notin TK - E$, y sea E_x la extensión obtenida añadiendo $\sim X$ como nueva fórmula a E. Supóngase que E_x es inconsistente, por lo que, para alguna $Z \in FK$, se tiene $Z, \sim Z \in TK - Ex$. Ahora bien, por el Hecho 2.2 se tiene que $\sim Z \supset (Z \supset X) \in TK$ y por lo tanto de $\sim Z \supset (Z \supset X) \in TK - E_x$, aplicando dos veces Mp se obtiene que $X \in TK - E_x$. Pero E_x tan sólo se diferencia de E en que tiene $\sim X$ como axioma adicional, así que 'X es un teorema de E_x' es equivalente a 'X es un teorema de E a partir del conjunto $\{\sim X\}$'. Por TD resulta que $\sim X \supset X \in TK - E$, y por PR se infiere que $X \in TK - E$, lo cual no es el caso, y por lo tanto, E_x es consistente.

Proposición 4.2. *(Extensión localmente consistente y completa)*
Si $E \in EXT(K)$ es localmente consistente entonces existe $E' \in EXT(E)$ que es localmente consistente y completa.

Prueba. Sea $X_0, X_1, X_2, \ldots$ una enumeración de todas las fórmulas de $KT4P$. Se construye una sucesión E_0', E_1', E_2', $\ldots$ de extensiones de E como sigue: Sea $E_0' = E$. Si $X_0 \in TK - E_0'$, sea $E_1' = E_0'$, en caso contrario añádase $\sim X_0$ como nueva fórmula para obtener E_1' a partir de E_0'. En general, dado $t \geq 1$, para construir

$E_t^{'}$ a partir de $E_{t-1}^{'}$, se procede así: si $X_{t-1} \in TK - E_{t-1}^{'}$, entonces $E_t^{'} = E_{t-1}^{'}$, en caso contrario, sea $E_t^{'}$ la extensión de $E_{t-1}^{'}$ obtenida añadiendo $\sim X_{t-1}$ como nueva fórmula. La prueba se realiza por inducción matemática sobre t.

Paso base. $t = 0$. Como E es consistente y $E_0^{'} = E$, por hipótesis, resulta que $E_0^{'}$ es consistente.

Paso inductivo. Hipótesis inductiva: $E_{t-1}^{'}$ es localmente consistente con $t \geq 1$. Por la proposición 4.1b, $E_t^{'}$ es localmente consistente.

Paso inductivo. Hipótesis inductiva: $E_{t-1}^{'}$ es localmente consistente con $t \geq 1$. Por la proposición 4.1b, $E_t^{'}$ es localmente consistente.

Por el principio de inducción matemática, se concluye que, todo $E_t^{'}$ es localmente consistente.

Se define $E^{'}$, como aquella extensión de E, la cual tiene como nuevas fórmulas a aquellas fórmulas que son nuevas fórmulas de al menos uno de los $E_t^{'}$. Si $E^{'}$ no es localmente consistente, entonces existe $X \in FK$ tal que, $X, \sim X \in TK - E^{'}$, pero las demostraciones de X y $\sim X$ en $E^{'}$ son sucesiones finitas de fórmulas, de modo que cada demostración solamente puede contener casos particulares de un número finito de axiomas o nuevas fórmulas de $E^{'}$, por lo que, debe existir un t, suficientemente grande, para que todas estas nuevas fórmulas utilizadas sean elementos de $E_t^{'}$, resultando que X,$\sim X \in TK - E_t^{'}$, lo cual es imposible ya que $E_t^{'}$ es localmente consistente. Por lo tanto, $E^{'}$ es localmente consistente.

Para probar que $E^{'}$ es completo, sea $X \in FK$. X debe aparecer en la lista X_0, X_1, X_2, …, supóngase que X es X_k. Si $X_k \in TK - E_k^{'}$, entonces $X_k \in TK - E^{'}$, puesto que $E^{'} \in EXT(E_k^{'})$, si $X_k \notin TK - E_k^{'}$, entonces de acuerdo con la construcción de $E_{k+1}^{'}$, $\sim X_k$ es una nueva fórmula de $E_{k+1}^{'}$, con lo que $\sim X_k \in TK - E_{k+1}^{'}$, y entonces $\sim X_k \in TK - E^{'}$. Así, en todo caso se tiene que $X_k \in TK - E^{'}$ o $\sim X_k \in TK - E^{'}$, por lo que $E^{'}$ es localmente completo. □

Proposición 4.3. *(Consistencia local subordinada). Sean* $Y, Z_1, \ldots, Z_k \in FK$.
Si $\{+Z_1, \ldots, +Z_k, \otimes Y\}$ *es localmente consistente entonces* $\{Z_1, \ldots, Z_k, Y\}$ *es localmente consistente.*

Prueba. Supóngase que $\{Z_1, \ldots, Z_k, Y\}$ es localmente inconsistente en $KT4P$, por lo que existe una fórmula $W \in FK$ tal que, a partir de $\{Z_1, \ldots, Z_k, Y\}$ se infiere $W\bullet \sim W$ en $KT4P$, utilizando TD y Exp resulta que $(Z_1 \bullet \ldots \bullet Z_k \bullet Y) \supset (W\bullet \sim W) \in TK$ y por DI, en $KT4P$ resulta $\sim (Z_1 \bullet \ldots \bullet Z_k \bullet Y) \in TK$, lo cual por $N\bullet$ significa, $(Z_1 \bullet \ldots \bullet Z_k) \supset \sim Y \in TK$. Utilizando la proposición 2.1 resulta que $+((Z_1 \bullet \ldots \bullet Z_k) \supset \sim Y) \in TK$, por $Ax11$, el Hecho 2.4 n y Mp se infiere $+(Z_1 \bullet \ldots \bullet Z_k) \supset + \sim Y \in TK$, por el Hecho 2.4$b$ se obtiene $(+Z_1 \bullet \ldots \bullet +Z_k) \supset + \sim Y \in TK$, lo cual, por DN, $N \supset$ y la definición de $\otimes$, equivale a $\sim (+Z_1 \bullet \ldots \bullet +Z_k \bullet \otimes Y) \in TK$, por lo que $\{+Z_1, \ldots, +Z_k, \otimes Y\}$ es localmente inconsistente en $KT4P$. Se ha probado que, $\{Z_1, \ldots, Z_k, Y\}$ localmente inconsistente implica que $\{+Z_1, \ldots + Z_k, \otimes Y\}$ localmente inconsistente, es decir, $\{+Z_1, \ldots, +Z_k, \otimes Y\}$ localmente consistente implica $\{Z_1, \ldots, Z_k, \}$ localmente consistente. □

Definición 4.2. *(Subordinado). Sean* $E, F \in EXT(K)$ *localmente consistentes y completas. Se dice que* F *es subordinada de* E *si y solamente si existe* $Y \in FK$, *tal que* $\otimes Y \in E$, *y además para cada* $Z \in FK$, *tal que* $+Z \in E$, *se tiene que* $Y, Z \in F$.

Proposición 4.4. *(Extensión subordinada localmente consistente y completa) Para $E \in EXT(K)$, $X \in FK$. Si E es localmente consistente y completa y $\otimes X \in E$, entonces existe $F \in EXT(K)$ localmente consistente y completa tal que, $X \in F$ y F es subordinada de E.*

Prueba. Supóngase que $\otimes X \in E$. Sea $E_X = \{X\} \cup \{Z : +Z \in E\}$, como E es localmente consistente, entonces por la proposición 4.3 , E_X también es localmente consistente. Al adicionar a E_X los axiomas de $KT4P$ y todas sus consecuencias, se obtiene una extensión de $KT4P$ que incluye a E_X, utilizando la proposición 4.2, se construye una extensión localmente consistente y completa F de $KT4P$ la cual incluye a E_X. Como $X \in E_X$, también $X \in F$. Si $+W \in E$, por definición de E_X, $W \in E_X$, por lo que $W \in F$. Por lo que, F es subordinado de E. □

Proposición 4.5. *(Propiedades de la subordinación). Para $E, F, G \in EXT(K)$ localmente consistentes y completas.*

a. (Transitividad). Si G es subordinado de F y F es subordinado de E entonces G es subordinado de E.

b. (Reflexividad). F es subordinado de F.

Prueba. Parte a, supóngase que G es subordinado de F y F es subordinado de E. Como G es subordinado de F entonces existe $\otimes Z \in F$ tal que $Z \in G$. Si $\otimes Z \notin E$, entonces al ser una extensión localmente completa, $\sim \otimes Z \in E$, por el Hecho 2.4m, resulta $+ \sim Z \in E$, por el Hecho 2.4i, se sigue $++ \sim Z \in E$, y al ser F subordinado de E resulta que $+ \sim Z \in F$, lo cual, por el Hecho 2.4m, significa que $\sim \otimes Z \in F$, pero esto es imposible ya que F es localmente consistente. Por lo que, $\otimes Z \in E$.
Sea $W \in FK$, tal que $+W \in E$, por el Hecho 2.4i, $+W \supset ++W \in E$, aplicando Mp, $++W \in E$, y como F es subordinado de E, se infiere que $+W \in F$, y como, G es subordinado de F, entonces $W \in G$. En resumen, existe $\otimes Z \in E$ tal que para cada fórmula $+W \in E$ se tiene que $Z \in G$ y $W \in G$, y por lo tanto, G es subordinado de E.
Parte b. Sea X el axioma $Ax1.1$, por lo que $X \in TK$, y como por el Hecho 2.4e se tiene $X \supset \otimes X$, resulta que $\otimes X \in TK$, entonces $X \in F$ y $\otimes X \in F$. Supóngase que $+W \in F$, por el Hecho 2.4d se tiene $+W \supset W$, resultando que $W \in F$. Por lo tanto, F subordinada de F. □

Proposición 4.6. *(Construcción de un modelo). Si $E' \in EXT(K)$ es localmente consistente, entonces existe un modelo en el cual todo $X \in TK - E'$ es verdadero.*

Prueba. Se define el modelo $(S, ME_a, <, V)$ de la siguiente manera: sean $E, F, G, \ldots$, extensiones localmente consistentes y completas de E' (E_a la inicial y las demás subordinadas), presentadas en las proposiciones 4.2 y 4.4. A cada extensión F, se le asocia un mundo posible MF, sean S el conjunto de tales mundos posibles y ME_a el mundo actual. La relación de accesibilidad, $<$, se construye así: $MF < MG$ si y solamente si G es subordinado de F.
Para cada $MF \in S$ y para cada $X \in F$, $V(MF, X) = 1$ si $X \in F$ y $V(MF, X) = 0$ si $X \notin F$, donde F es la extensión localmente consistente y completa asociada a MF. Nótese que V es funcional, por ser F localmente consistente y completa. Para afirmar que M es un modelo, se deben garantizar las reglas 1 a 5 de la definición 3.2.

1. Por $Ax\lambda$ se tiene $\lambda \in TK$, por lo que $\lambda \in F$, es decir, $V(MF,\lambda) = 1$. Por lo tanto, se satisface la definición $V\lambda$.

2. Para el caso del condicional $X \supset Y$. Utilizando $N\supset$, $I\bullet$ y $E\bullet$, se tiene la siguiente cadena de equivalencias: $V(MF,X \supset Y) = 0$, es decir $\sim (X \supset Y) \in F$, o sea que $X\bullet \sim Y \in F$, resultando que $X \in F$ y $\sim Y \in F$, lo cual significa que $V(MF,X) = 1$ y $V(MF,Y) = 0$, por lo que se satisface la definición $V\supset$.

3. Para el caso de la conjunción $X \bullet Y$. Utilizando $I\bullet$ y $E\bullet$, se tiene la siguiente cadena de equivalencias: $V(MF,X \bullet Y) = 1$, es decir $X \bullet Y \in F$, por lo que $X \in F$ y $Y \in F$, lo cual significa que $V(MF,X) = 1$ y $V(MF,Y) = 1$, por lo que se satisface la definición $V\bullet$.

4. Para el caso de la disyunción $X \cup Y$. Utilizando $N\cup$, $I\bullet$ y $E\bullet$, se tiene la siguiente cadena de equivalencias: $V(MF,X \cup Y) = 0$, es decir $\sim (X \cup Y) \in F$, o sea que $\sim X\bullet \sim Y \in F$, de donde $\sim X \in F$ y $\sim Y \in F$, es decir $V(MF,X) = 0$ y $V(MF,Y) = 0$, por lo que se satisface la definición $V\cup$.

5. Para el caso de la regla $V-$. Sean MF es un mundo asociado a F, MG es un mundo asociado a G y $Z \in FK$. Supóngase que $V(MF,-Z) = 1$, por lo que $-Z \in F$, y por Hecho 2.4k $\otimes \sim Z \in F$, por la proposición 4.4, existe G subordinada de F, tal que $\sim Z \in G$, resultando que, $(\exists MG \in S)(MF < MG$ y $MG(Z) = 0)$.

Para probar la recíproca, supóngase que $(\exists MG \in S)(MF < MG$ y $MG(Z) = 0)$. Si $V(MF,-Z) = 0$, entonces resulta que $\sim -Z \in F$, y por Hecho 2.4c, $+Z \in F$, y como $MF < MG$, es decir, G es subordinada de F, entonces $Z \in G$, es decir, $MG(Z) = 1$, resultando, por la hipótesis, que G es localmente inconsistente, lo cual no es el caso. Por lo tanto, $V(MF,-Z) = 1$. Como ya se probó la recíproca, entonces, se satisface la definición $V-$.
Con base en el análisis anterior, se infiere que V es una valuación, y como las restricciones RT y RR, se garantizan por la proposición 4.5, se concluye finalmente que, M es un modelo.
Para finalizar la prueba, sea X un teorema de E', por lo que X está en E'. Por lo tanto, utilizando la definición de V, resulta que $V(ME_a,X) = 1$, es decir, X es verdadera en el modelo $M = (S,ME_a,<,V)$. □

Proposición 4.7. *(Completitud de KT4P) Para* $X,X_1, \ldots, X_n \in FK$.

a. Si $X \in VK$ *entonces* $X \in TK$.

b. Si $\{X_1,\ldots,X_n\}$ *valida a Y entonces Y es una consecuencia de* $\{X_1,\ldots,X_n\}$.

Prueba. Parte *a*. Si $X \notin TK$, entonces, por la proposición 4.1b, la extensión E', obtenida añadiendo $\sim X$ como nueva fórmula, es consistente. Así pues, según la proposición 4.6, existe un modelo M tal que todo teorema de E' es verdadero en M, y como $\sim X \in TK - E'$, entonces $\sim X$ es verdadero en M, es decir, X es falso en M, y por lo tanto, $X \notin VK$. Se ha probado que, $X \notin TK$ implica $X \notin VK$, es decir, $X \in VK$ implica $X \in TK$.
Parte *b*. Supóngase que $\{X_1,\ldots,X_n\}$ valida a Y, es decir, $(X_1 \bullet X_2 \bullet \ldots \bullet X_n) \supset Y \in VK$, por la parte *a*, se

deriva que, $(X_1 \bullet X_2 \bullet \ldots \bullet X_n) \supset Y \in TK$. Si se asumen $\{X_1, \ldots, X_n\}$, por $Conj$ y Mp se infiere Y, por lo tanto,Y es una consecuencia de $\{X_1, \ldots, X_n\}$. □

Proposición 4.8. *(Caracterización semántico-deductiva de KT4P). Para X, Y,*X_1*, …,* $X_n \in FK$.

a. $X \in VK$ *si y solo si* $X \in TK$.

b. $\{X_1, \ldots, X_n\}$ *valida a Y si y solo si Y es una consecuencia de* $\{X_1, \ldots, X_n\}$.

Prueba. Consecuencia de las proposiciones 3.2 y 4.7. □

5. ALFAK EN EL ESTILO ORIGINAL DE CHARLES SANDERS PEIRCE

En esta sección, se presenta la versión inicial de los gráficos existenciales paraconsistentes, para el sistema AlfaK. Para la construcción de los gráficos existenciales, se utiliza una variante, de la notación propuesta por Peirce en el volumen 4, párrafo 378 de *Collected Papers* Peirce (1965).

Definición 5.1. *(Gráficos existenciales). El conjunto, GK, de gráficos existenciales para el sistema AlfaK, se construye a partir de un conjunto de gráficos atómicos, GA, y de la constante* λ *(gráfico vacío,* $\lambda =' \_ '$*), de la siguiente manera.*

1. $P \in GA$ *implica* $P \in GK$.
2. $\lambda \in GK$.
3. $X \in GK$ *implica* $\{X\} \in GK$.
4. $X, Y \in GK$ *implica* $XY, [X(Y)\_], [\_(X)(Y)\_] \in GK$.
5. *Solo* $1, 2, 3$ *y* 4 *determinan GK*.

Definición 5.2. *(Tipos de curvas). En el gráfico* $[X(Y)]$*, o de manera más precisa* $[X(Y)\_]$*, la estructura* $[\ldots(\ldots)\_]$*, se llama rizo. La parte* $[\ldots]$ *del rizo (los corchetes), se llama curva externa del rizo. La parte* $(\ldots)$ *del rizo (los paréntesis), se llama curva interna del rizo o lazo. En* $[X\,(Y)]$*, X se denomina antecedente del rizo y Y consecuente del rizo.*

La parte que se encuentra entre ambas curvas $[X\,(\ldots)]$ (aquella donde se encuentra el antecedente X) se llama *región externa del rizo* o *región del antecedente*. La parte que se encuentra rodeada por la curva interna o lazo $[\ldots(Y)]$ (aquella donde se encuentra el consecuente Y) se llama *región interna del rizo* o *región del lazo* o *región del consecuente*. La línea que conecta la curva interna con la externa $[\ldots(\ldots)\_]$, se llama *enlace del rizo* (usualmente el enlace, no se hace explícito, pero está presente).
En el gráfico $\{Z\}$, la *estructura* $\{\ldots\}$, se llama *corte*. La parte que se encuentra rodeada por el corte $\{Z\}$ (aquella donde se encuentra Z) se llama *región interna del corte* o simplemente *región del corte*.

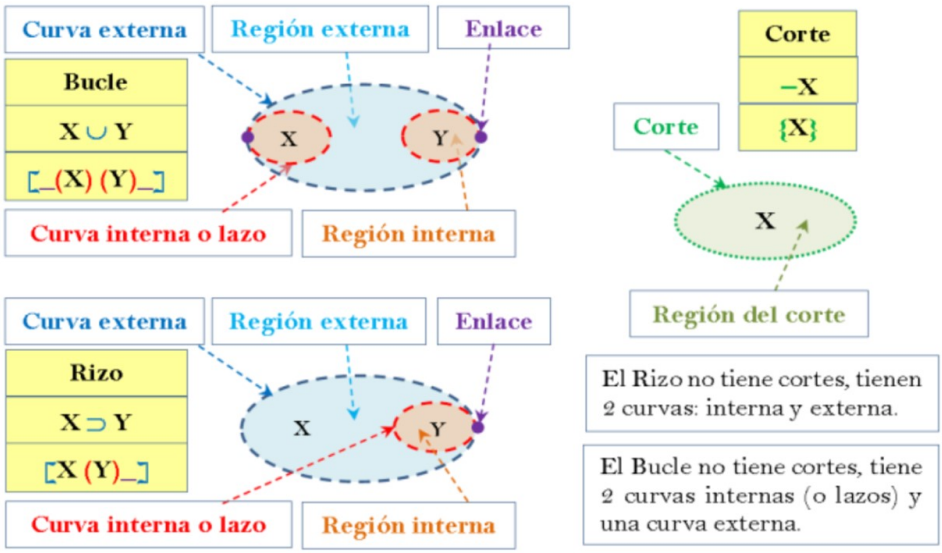


Figura 1: Curvas y regiones de los rizos, bucles y cortes. Oostra (2010).

Un rizo es el *ensamble* de dos *cortes encajados*, es decir, cuando se borra el enlace de un rizo $[X\,(Y)\,\_]$, se obtienen los *cortes encajados* $\{X\,\{Y\}\}$ (no hay cortes entre ellos). Cuando se *escribe el enlace* (es decir, se realiza el *ensamble*) entre los cortes encajados $\{X\,\{Y\}\}$, se dice que ha sido *ensamblado* el rizo $[X\,(Y)\,\_]$, o de manera más concisa $[X\,(Y)]$. Como se verá más adelante, el borrado y la escritura del enlace, no generan siempre gráficos lógicamente equivalentes.

En el gráfico $[(X)\,(Y)]$, o de manera más precisa $[\_\,(X)\,(Y)\,\_]$, la estructura $[\_\,(\ldots)\,(\ldots)\,\_]$, se llama *bucle*. Las partes $[(\ldots)\,(\ldots)]$ del *bucle* (los paréntesis), se llaman *curvas internas del bucle* o *lazos*. La parte $[\ldots]$ del *bucle* (los corchetes), se llama *curva externa del bucle*. En $[(X)\,(Y)]$, X y Y se denominan *disyuntos* del bucle.

La parte que se encuentra entre ambas curvas $[\_\,(\ldots)\,(\ldots)\,\_]$ se llama *región externa del bucle*. La parte que se encuentra rodeada por las curvas internas o lazos $[(X)\,(Y)]$ (aquella donde se encuentran los disyuntos X y Y) se llaman *regiones internas del bucle* o, cada una de ellas, *región interna del lazo* o *región del disyunto*. Las *líneas que conectan* las curvas internas con la externa $[\_\,(\ldots)\,(\ldots)\,\_]$, se llaman *enlaces del bucle* (usualmente esta línea no se hace explícita, pero está presente).

Un *bucle* es el *ensamble doble* de dos *cortes encajados en otro corte*, es decir, cuando se *borra un enlace del bucle* $[\_\,(X)\,(Y)\,\_]$, se obtiene el rizo $[\{X\}\,(Y)\,\_]$, cuando se *borra el enlace* de este rizo se obtienen los *cortes encajados* $\{\{X\}\,\{Y\}\}$. Cuando se escriben *los enlaces* (es decir, se realiza *el ensamble del bucle*) entre los cortes encajados $\{\{X\}\,\{Y\}\}$, se dice que ha sido *ensamblado* el bucle $[\_\,(X)\,(Y)\,\_]$, o de manera más concisa $[(X)\,(Y)]$. Un *bucle* es el *ensamble simple* de un *rizo*, es decir, cuando se *borra un enlace* de un bucle $[\_\,(X)\,(Y)\,\_]$, se obtiene el *rizo* $[\{X\}\,(Y)\,\_]$, o el *rizo* $[\_\,(Y)\,\{X\})]$.

La representación de los rizos y bucles, en el formato tradicional, se muestra en la Figuras 1 y 2.

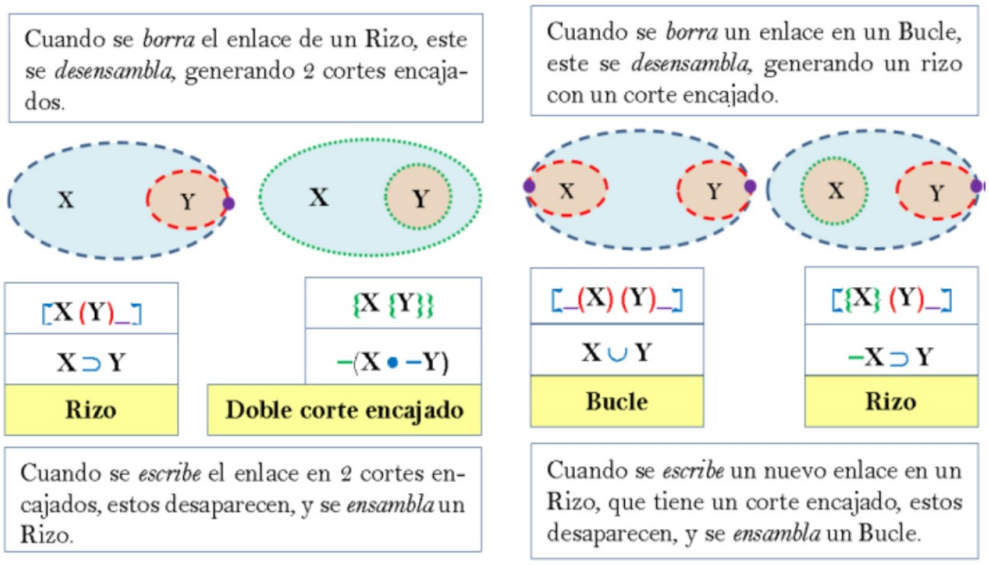


Figura 2: Transformación de bucle en rizo, y de rizo en doble corte encajado, Oostra (2010).

Definición 5.3. *(Tipos de regiones) Sean $X,Y,Z \in GK$. Se dice que un gráfico X se encuentra en una región par, denotado $X \in RP$ o X^{Rp}, si X se encuentra rodeado por un número par de curvas (internas, externas) y/o cortes. X se encuentra en una región impar, denotado $X \in RI$ o X^{Ri}, si X se encuentra rodeado por un número impar de curvas (internas, externas) y/o cortes.*
$X \in Rn$ o X^{R_n} significa que el gráfico X se encuentra en una región rodeada por n curvas y/o cortes ($n = 0,1,2,3,\ldots$), donde n puede ser par o impar.
$X \in RU$ o X^{Ru} significa que X se encuentra en una región de curvas, es decir, X está rodeado únicamente por curvas (no aparece ningún corte).

Notación. *(Rizo con región externa vacía y rizo para la afirmación alterna)* Para $X \in GK$.
$[(X)] = [\lambda(X)]$. $\quad *X = [\{X\}] = [\{X\}(\{\lambda\})]$, y se dice que $*X$ es un *gráfico fuerte*.

Definición 5.4. *(Sistema graficador AlfaK). El sistema deductivo AlfaK consta de:*

*a. Axioma. Ax*λ. λ.

b. Reglas primitivas de transformación de gráficos (y algunas derivadas), donde $X,Y,Z \in GK$.

Borrado de gráficos.
1. $XZ| \Rightarrow X$. 2. $[Y\{XZ\}(W)]| \Rightarrow [Y\{X\}(W)]$. 3. $[(Y)(XZ)]| \Rightarrow [(Y)(X)]$.
Escritura de gráfico.
4. $[X(Y)]| \Rightarrow [XZ(Y)]$. 5. $[Y(W[X(K)])]| \Rightarrow [Y(W[XZ(K)])]$.
Borrado y escritura del enlace. En un rizo o bucle, un *enlace* puede ser *borrado* (generando el doble corte encajado o el rizo), o, en un doble corte encajado, un *enlace* puede ser *escrito* (ensamblando el rizo), o, en un rizo, el segundo enlace puede ser escrito (ensamblando el bucle), según se cumpla alguna de las siguientes

condiciones.

Borrado del enlace.

6. $[X\,[Z\,(Y)\,\_]\,(Z)]\,|\Rightarrow[X\,\{Z\,\{Y\}\}\,(Z)]$. 7. $[\_\,(*X)\,(Y)]\Rightarrow[\{*X\}\,(Y)]$.

Escritura del enlace.

8. $[(Z)\,\{X\}]\,|\Rightarrow[(Z)\,(X)\,\_]$. 9. $[(Y)\,X]\,|\Rightarrow[(Y)\,(\{X\})\,\_]$.

Borrado y escritura del enlace.

10. Regla derivada. $[*X\,(Y)]\Leftrightarrow[\_\,(\{X\})\,(Y)]$.

Borrado y escritura de rizos.

11. $X\Leftrightarrow[(X)]$. 12. $\{X\}\Leftrightarrow\{[(X)]\}$.

13. $[Y\,(X)]\Leftrightarrow[Y\,([(X)])]$. 14. $[(Y)\,(X)]\Leftrightarrow[(Y)\,([(X)])]$.

Borrado y escritura de la afirmación fuerte. Una *afirmación fuerte* puede ser borrada (sin borrar el grafico que afirma), o, la *afirmación fuerte* puede ser *escrita* a la izquierda del gráfico dado, según se cumpla alguna de las siguientes condiciones.

15. $\{X\}\Leftrightarrow\{*X\}$. 16. $\{[*X\,(Y)]\}\,|\Rightarrow\{[X\,(Y)]\}$. Iteración y desiteración de gráficos. Un *gráfico* se puede *iterar* en su misma región, o en una región de un rizo, que *no* sea parte del *gráfico* que se itera, o en una región de un corte, que *no* sea parte del *gráfico* que se itera, e inversamente, un *gráfico* que pueda ser considerado como escrito por iteración, se puede *desiterar*, si se cumple alguna de las siguientes condiciones.

Iteración y desiteración de un gráfico en región de curvas.

17. $Z\,[X\,(Y)]\Leftrightarrow Z\,[XZ\,(Y)]$. 18. $Z\,[Y\,(X)]\Leftrightarrow Z\,[Y\,(XZ)]$.

18.1 $Z\,[(Y)\,(X)]\Leftrightarrow Z\,[(Y)\,(XZ)]$. 19. $[YZ\,(X)]\Leftrightarrow[YZ\,(XZ)]$.

20. $[(Z\,[X\,(Y)])\,(W)]\Leftrightarrow[(Z\,[ZX\,(Y)])\,(W)]$.

Iteración y desiteración de un lazo.

21. $[(X)\,(X)]\Leftrightarrow[(X)]$.

Iteración y desiteración de un gráfico fuerte ($*Z$).

22. $*Z\,\{*ZY\}\Rightarrow *Z\,\{Y\}$ 23. $*Z\,\{Y\,[X\,(W)]\}\Leftrightarrow *Z\,\{Y\,[X*Z\,(W)]\}$.

Preservación de las transformaciones.

24. Regla derivada. $X>>Y$ implica $[W\,(ZX)]\,|\Rightarrow[W\,(ZY)]$.

Generación de reglas de graficación mediante la afirmación fuerte. A partir de las reglas de graficación, primitivas o derivadas, se generan nuevas reglas de graficación con la *afirmación fuerte*.

25. Regla derivada. $X\Rightarrow Y$ implica $*X\Rightarrow *Y$.

Afirmación fuerte de teoremas gráficos.

26. $Y\in TG$ implica $*Y\in TG$.

Reglas implícitas

Concatenación. Dos gráficos que se encuentren en una misma región, pueden se concatenados. E inversamente, dos gráficos que se encuentren concatenados, pueden ser separados, quedando ambos en la misma región.

27. $X,Y\Leftrightarrow YX$.

Conmutatividad. Dos gráficos concatenados pueden ser reescritos cambiando el orden de la concatenación.

28. $XY \Leftrightarrow YX$.
Asociatividad. En tres gráficos que se encuentren concatenados, es *irrelevante el orden* en que fueron concatenados. Inicialmente el primero se concatena con el segundo y este resultado se concatena con el tercero, o el primero se concatena con el resultado de concatenar el segundo con el tercero.
29. $XY,Z \Leftrightarrow X,YZ \Leftrightarrow XYZ$.
Notación. $X \Rightarrow Z$ significa que, mediante una *regla de transformación aplicada* al gráfico X, se infiere el gráfico Z. Si también Z se transforma en X, entonces se escribe $X \Leftrightarrow Z$. $X| \Rightarrow Z$ significa que, $X \Rightarrow Z$, pero en general, no vale $Z \Rightarrow X$. $X >> Z$ significa que X se transforma en Z utilizando un número finito de reglas de transformación.

Observación 5.1. *Las reglas* 27, 28 *y* 29, *reciben el nombre de reglas implícitas, puesto que, dada su obviedad y naturalidad gráfica, pueden no ser referenciadas, pero realmente son aplicadas.*

Definición 5.5. *(Teoremas de AlfaK). Sea $X \in GK$. X es un teorema gráfico de AlfaK, denotado $X \in TG$, si existe una demostración de X a partir del axioma $Ax\lambda$, utilizando las reglas de transformación de gráficos, es decir, X es la última fila de una secuencia finita de líneas, en la cual, cada una de las líneas es $Ax\lambda$, o se infiere de filas anteriores, utilizando las reglas de transformación. O dicho brevemente, $X \in TG$ si y solo si $\lambda >> X$. El número de líneas, de la secuencia finita, es referenciado como la longitud de la demostración de X.*

Proposición 5.1. *(Teorema de deducción gráfico y demostración indirecta). Para X,Y, $Z \in GK$.*

a. $TDG.XY >> Z$ *implica* $X >> [Y(Z)]$.

b. $TDG.Y >> Z$ *implica* $[Y(Z)]$.

c. $\{\_\} >> Z$.

d. $DI.X >> \{\_\}$ *implica* $\{X\}$.

e. $DI.\ \{X\} >> \{\_\}$ *implica* X.

Prueba. Parte *a*. Supóngase que $XY >> Z$. Si además se supone X, entonces resulta $Y >> Z$, aplicando la regla 24 se deriva $[Y(Y)] >> [Y(Z)]$. Además, por $Ax\lambda$ se tiene λ y por la regla 11 se obtiene $[\_(\lambda)]$, es decir, $[\_(\_)]$, aplicando la regla 4 resulta $[Y(\_)]$, utilizando la regla 19 se infiere $[Y(Y)]$, como ya se probó $[Y(Y)] >> [Y(Z)]$, se deduce $[Y(Z)]$. Por lo tanto, $X >> [Y(Z)]$, es decir, $XY >> Z$ implica $X >> [Y(Z)]$.
Parte *b*. Caso particular de la parte *a* (tomando $X = \lambda$).
Parte *c*. Supóngase que $\{\_\}$, por la regla 15 se infiere $\{*\lambda\}$. Por $Ax\lambda$ se tiene $\lambda \in TG$, utilizando la regla 26 se tiene $*\lambda \in TG$, por la regla 11 se obtiene $[(*\lambda)]$, utilizando la regla 4 se sigue $[\{Z\}(*\lambda)]$, aplicando la regla 8 se concluye $[(Z)(*\lambda)]$, utilizando la regla 7 se deriva $[(Z)\{*\lambda\}]$, y como se tienen $\{*\lambda\}$, por la regla 17 se deduce $[(Z)]$, por la regla 11 se afirma Z. Por lo tanto, $\{\_\} >> Z$.
Parte *d*. Supóngase que $X >> \{\_\}$, por la parte *c*., se tiene $\{\_\} >> \{X\}$, por lo que, $X >> \{X\}$, aplicando la parte *b*, se infiere $[X(\{X\})]$, utilizando la regla 9 se sigue $[(\{X\})(\{X\})]$, por la regla 21 se deriva $[(\{X\})]$,

lo cual por regla 11 significa $\{X\}$.
Parte *e*. Supóngase que $\{X\} >>> \{\_\}$, por la parte *c*, se tiene $\{\_\} >> X$, por lo que, $\{X\} >> X$, aplicando la parte *b*, se infiere $[\{X\}(X)]$, utilizando la regla 8 se sigue $[(X)(X)]$, por la regla 21 se deriva $[(X)]$, lo cual por la regla 11 significa X. □

6. EQUIVALENCIA ENTRE KT4P Y ALFAK

En esta sección, se presenta la equivalencia entre $KT4P$ y AlfaK, inicialmente, en la proposición 6.2, se prueba que los teoremas de $KT4P$ son teoremas gráficos de AlfaK, en la proposición 6.4, se prueba que los teoremas gráficos de $KT4P$ son válidos en la semántica de mundos posibles, finalmente, en la proposición 6.8, se prueba que los teoremas de $KT4P$ son exactamente los teoremas gráficos.

Definición 6.1. *(Interpretación de $KT4P$ en términos de AlfaK)*

a. Por definición, $FA = GA$ (las fórmulas atómicas son los mismos gráficos atómicos).

b. Función de traducción, $(\_)'$ de FK en GK. Sean $X,Y \in FK$ y $P \in FA$.

1. $P' = P$.
2. $(X \supset Y) = \left[X'\left(Y'\right)\right]$.
3. $(X \cup Y)' = \left[(X)'(Y)'\right]$.
4. $(-X)' = \left\{X'\right\}$.
5. $(X \bullet Y)' = X'Y'$.
6. $(+X)' = \left[\left\{X'\right\}\right] = *X'$.
7. $\lambda' = \lambda$.

Proposición 6.1. *(Axiomas de $KT4P$ son teoremas de AlfaK) Sea $X \in FK$.*
Si X es axioma de $KT4P$ entonces $X' \in TG$.

Prueba. Utilizando $Ax\lambda$ y las reglas primitivas de la definición 5.4 se tienen:

1. $Ax\lambda.\lambda$. También es un axioma de AlfaK.

2. $Ax1$. $X \supset (Z \supset X)$. Por $Ax\lambda$ se tiene λ, por la regla 11 se infiere $[(\lambda)]$, es decir, $[()]$, utilizando la regla 4 se sigue $\left[X'()\right]$, aplicando la regla 19 resulta $\left[X'\left(X'\right)\right]$, por la regla 13 se infiere $\left[X'\left(\left[\left(X'\right)\right]\right)\right]$, finalmente por la regla 5 se concluye $\left[X\left(\left[Z'\left(X'\right)\right]\right)\right]$, es decir, $(Ax1)'$ es un teorema gráfico.

3. $Ax2$. $(X \supset (Y \supset Z)) \supset ((X \supset Y) \supset (X \supset Z))$. Supóngase que $\left[X'\left(\left[Y'\left(Z'\right)\right]\right)\right]$, $\left[X'\left(Y'\right)\right]$ y X'. Por la regla 17 se deducen $\left[\left(\left[Y'\left(Z'\right)\right]\right)\right]$ y $\left[\left(Y'\right)\right]$, aplicando la regla 11 se siguen $\left[Y'\left(Z'\right)\right]$ y Y',

utilizando la regla 17 se infiere $\left[\left(Z'\right)\right]$ finalmente por la regla 11 se concluye Z'. Por TDG dos veces, se ha probado que $(Ax2)'$ es un teorema gráfico.

4. $Ax3$. $X \supset (X \cup Y)$ y $Ax4$. $X \supset (Y \cup X)$. Supóngase que X', por $Ax\lambda$, se tiene λ, por la regla 11 se obtiene $[(\lambda)]$, es decir, $[(\_)]$, utilizando la regla 4 se sigue $\left[\left\{Y'\right\}()\right]$, aplicando la regla 18 resulta $\left[\left\{Y'\right\}\left(X'\right)\right]$, finalmente, aplicando la regla 8 se concluye $\left[\left(Y'\right)\left(X'\right)\right]$. Por TDG, se ha probado que $(Ax3)'$ y $(Ax4)'$ son teoremas gráficos.

5. $Ax5$. $(X \supset Y) \supset ((Z \supset Y) \supset ((X \cup Z) \supset Y))$. Supóngase que $\left[X'\left(Y'\right)\right]$, $\left[Z'\left(Y'\right)\right]$ y $\left[\left(X'\right)\left(Z'\right)\right]$. Por la regla 18.1 se deriva $\left[\left(\left[X'\left(Y'\right)\right]X'\right)\left(Z'\right)\right]$, por la regla 20 se sigue $\left[\left(\left[\left(Y'\right)\right]X'\right)\left(Z'\right)\right]$, utilizando la regla 14 se obtiene $\left[\left(Y'X'\right)\left(Z'\right)\right]$, por la regla 3 resulta $\left[\left(Y'\right)\left(Z'\right)\right]$, utilizando la regla 18.1 se infiere $\left[\left(Y'\right)\left(\left[Z'\left(Y'\right)\right]Z'\right)\right]$, según la regla 20 $\left[\left(Y'\right)\left(\left[\left(Y'\right)\right]Z'\right)\right]$, por la regla 14 resulta $\left[\left(Y'\right)\left(Y'Z'\right)\right]$, utilizando la regla 3 se deriva $\left[\left(Y'\right)\left(Y'\right)\right]$, aplicando la regla 21 se concluye $\left[\left(Y'\right)\right]$, finalmente, por la regla 11 se obtiene Y'. Por lo tanto, según TDG se ha probado que $(Ax5)'$ es un teorema gráfico.

6. $Ax6$. $(X \bullet Y) \supset X$ y $Ax7$. $(X \bullet Y) \supset Y$. Supóngase que $X'Y'$, por la regla 1 se concluyen X' y Y'. Por TDG, se puede afirmar que $(Ax6)'$ y $(Ax7)'$ son teoremas gráficos.

7. $Ax8$. $(X \supset Y) \supset ((X \supset Z) \supset (X \supset (Y \bullet Z)))$. Supóngase que $\left[X'\left(Y'\right)\right]$, $\left[X'\left(Z'\right)\right]$ y X'. Por la regla 17 se infieren $\left[\left(Y'\right)\right]$ y $\left[\left(Z'\right)\right]$, por la regla 11 se derivan Y' y Z', utilizando la regla 27, se concluye $Y'Z'$. Por lo tanto, utilizando TDG se ha probado que $(Ax8)'$ es un teorema gráfico.

8. $Ax9$. $((X \supset Y) \supset X) \supset X$. Supóngase que $\left[\left[X'\left(Y'\right)\right]\left(X'\right)\right]$, aplicando la regla 6 resulta $\left[\left\{X'\left\{Y'\right\}\right\}\left(X'\right)\right]$, según la regla 2 se sigue $\left[\left\{X'\right\}\left(X'\right)\right]$, utilizando la regla 8 se deriva $\left[\left(X'\right)\left(X'\right)\right]$, por la regla 21 se obtiene $\left[\left(X'\right)\right]$, aplicando la regla 11 se concluye X'. Por lo tanto, utilizando TDG, se ha probado que $(Ax9)'$ es un teorema gráfico.

9. $Ax10$. $X \cup -X$. Por $Ax\lambda$ y la regla 11 se tiene $[(\lambda)]$, es decir, $[(\_)]$, aplicando la regla 4 se infiere $\left[X'()\right]$, por la regla 19 se deriva $\left[X'\left(X'\right)\right]$, utilizando la regla 9 se concluye $\left[\left(\left\{X'\right\}\right)\left(X'\right)\right]$. Por lo tanto, $(Ax10)'$ es un teorema gráfico.

10. $Ax11$. $(-X \supset -\lambda) \supset (-Y \supset -(X \supset Y))$, lo cual por la definición de $+$ y Exp equivale a $(+X \bullet -Y) \supset -(X \supset Y)$. Supóngase que $*X'$ y $\left\{Y'\right\}$, por la regla 12 se obtiene $\left\{\left[\left(Y'\right)\right]\right\}$, según la regla 23 resulta $\left\{\left[*X'\left(Y'\right)\right]\right\}$, aplicando la regla 16 se concluye $\left\{\left[X'\left(Y'\right)\right]\right\}$. Finalmente, por TDG se afirma que $[(+X \bullet -Y) \supset -(X \supset Y)]'$ es un teorema gráfico, se concluye finalmente que, $(Ax11)'$ es un teorema gráfico.

11. $Ax12.\ -+X \supset -X$. Supóngase que $\left\{*X^{'}\right\}$, por la regla 15 se tiene $\left\{*X^{'}\right\} \Leftrightarrow \left\{X^{'}\right\}$, resultando $\left\{X^{'}\right\}$. Por lo tanto, utilizando TDG y el Hecho 2.1c, $(Ax12)^{'}$ es un teorema gráfico.

12. $Ax13.\ +X \supset (-X \supset Y)$. Supóngase que $*X^{'}$ y $\left\{X^{'}\right\}$, por la regla 15 se infiere $\left\{*X^{'}\right\}$, aplicando la regla 22 se sigue $\{\_\}$, es decir, $\{\lambda\}$, por la regla 15 se infiere $\{*\lambda\}$. Por $Ax\lambda$ se tiene $\lambda \in TG$, utilizando la regla 26 se tiene $*\lambda \in TG$, por la regla 11 se obtiene $[(*\lambda)]$, utilizando la regla 4 se sigue $\left[\left\{Y^{'}\right\}(*\lambda)\right]$, aplicando la regla 8 se concluye $\left[\left(Y^{'}\right)(*\lambda)\right]$, utilizando la regla 7 se deriva $\left[\left(Y^{'}\right)\{*\lambda\}\right]$, y como se tienen $\{*\lambda\}$, por la regla 17 se deduce $\left[\left(Y^{'}\right)\right]$, por la regla 11 se afirma $Y^{'}$. Finalmente, por TDG dos veces, se concluye que $(Ax13)^{'}$ es un teorema gráfico.

13. $Ax14$. Si X es axioma de AlfaK entonces $+X$ es axioma de AlfaK. Corresponde a la regla 26: Si $X^{'}$ es un teorema gráfico entonces $*X^{'}$ es un teorema gráfico. Por lo tanto, $(Ax14)^{'}$ es válido en AlfaK. □

Proposición 6.2. *(Los teoremas de $KT4P$ son teoremas gráficos). Para $X \in FK$.*

a. Si $X \in TK$ entonces $X^{'} \in TG$.

b. Si Y es consecuencia de X entonces $X^{'} >> Y^{'}$.

Prueba. Parte a. Inducción sobre la longitud de la demostración de X en $KT4P$.
Paso base. Si la longitud de la demostración es 1, entonces X es un axioma, por la proposición 6.1 $X^{'} \in TG$.
Paso inducción. Como hipótesis inductiva se tiene que: si $Y \in TK$ y la longitud de la demostración de Y es menor que L entonces $Y^{'}$ es un teorema gráfico. Supóngase $X \in TK$ y que la longitud de la demostración de X es L, por lo que X es un axioma o se obtiene de pasos anteriores utilizando Mp. En el primer caso se procede como en el paso base. En el segundo caso se tienen Y y $Y \supset Z$ en pasos previos de la demostración de X, es decir, las longitudes de las demostraciones de Y y $Y \supset X$ son menores que L, por la hipótesis inductiva resulta que $Y^{'} \in TG$ y $\left[Y^{'}\left(X^{'}\right)\right] \in TG$, por la regla 17 resulta $\left[\left(X^{'}\right)\right] \in TG$, finalmente, aplicando la regla 11, se concluye que $X^{'} \in TG$.
Por el principio de inducción matemática, se probado que los teoremas de AlfaK son teoremas gráficos.
Parte b. Si Y es *consecuencia* de X, entonces $X \supset Y \in TK$, por la parte a, $\left[X^{'}\left(Y^{'}\right)\right] \in TG$, es decir, $\lambda >> \left[X^{'}\left(Y^{'}\right)\right]$, si se supone $X^{'}$, por la regla 17 se sigue $\left[\left(Y^{'}\right)\right]$, aplicando la regla 11 resulta $Y^{'}$, por lo que $X^{'} >> Y^{'}$. □

Definición 6.2. *(Interpretación de AlfaK en términos de $KT4P$).*
Función de traducción, $(\_)^{''}$ de GK en FK. Sean $X, Y \in GK$, $P \in GA$

1. $P^{''} = P$.
2. $[X(Y)]^{''} = X^{''} \supset Y^{''}$.
3. $[(X)(Y)]^{''} = X^{''} \cup Y^{''}$.

4. $\{X\}'' = -X''$.

5. $(XY)'' = X'' \bullet Y''$.

6. $(*X)'' = \sim -X'' = +X''$.

7. $\lambda'' = \lambda$.

Proposición 6.3. *(Las reglas primitivas de AlfaK son reglas válidas en la semántica de $KT4P$). Para $X,Y,Z \in GK$.*
Si $X \Rightarrow Y$ entonces, de X'' se infiere válidamente Y''.

Prueba. Basta con validar las reglas primitivas de la definición 5.4.
Regla 1: $XZ| \Rightarrow X$. Por $Ax6$ y $Ax7$ se tienen $\left(X'' \bullet Z''\right) \supset X''$, y $\left(X'' \bullet Z''\right) \supset Z''$, es decir, $(XZ)'' \Rightarrow X''$, y $(XZ)'' \Rightarrow Z''$. Por lo tanto, $(\text{Regla } 1)''$ es una regla válida en $KT4P$.

Regla 2: $[Y\{XZ\}(W)]| \Rightarrow [Y\{X\}(W)]$. Supóngase que $M\left(\left(Y' \bullet -\left(X'' \bullet Z''\right)\right) \supset W''\right) = 1$. Supuesto 2, $M\left(Y'' \bullet -X''\right) = 1$, por $V\bullet$ se sigue $M\left(Y''\right) = 1$ y$M\left(-X''\right) = 1$, aplicando $V-$ existe $P \in S, M < P$ y $P\left(X''\right) = 0$, por $V\bullet$ se deriva que $P\left(X'' \bullet Z''\right) = 0$, de nuevo por $V-$ se infiere $M\left(-\left(X'' \bullet Z''\right)\right) = 1$, como se tiene $M\left(Y''\right) = 1$, por $V\bullet$ se deriva $M\left(Y'' \bullet -\left(X''Z''\right)\right) = 1$, utilizando el supuesto inicial, según $V \supset$ se afirma $M\left(W''\right) = 1$. Aplicando TDG, se concluye que $M\left(Y'' \bullet -X''\right) = 1$ implica $M\left(W''\right) = 1$, lo cual por $V \supset$ significa $M\left(\left(Y'' \bullet -X''\right) \supset W''\right) = 1$. Por lo tanto, $(Regla2)''$ es una regla válida en $KT4P$.

Regla 3: $[(Y)(XZ)]| \Rightarrow [(Y)(X)]$. Supuesto 1, $M\left(Y'' \cup \left(X'' \bullet Z''\right)\right) = 1$. Supuesto 2, $M\left(Y'' \cup X''\right) = 0$, por $V\cup$ resulta que $M\left(Y''\right) = 0$ y $M\left(X''\right) = 0$, utilizando el supuesto 1, y aplicando $V \supset$ se infiere $M\left(X'' \bullet Z''\right) = 1$, lo cual por $V\bullet$ implica que $M\left(X''\right) = 1$, lo cual no es el caso, por lo que $M\left(Y'' \cup X''\right) = 1$. Por lo tanto, $(Regla3)''$ es una regla válida en $KT4P$.

Regla 4: $[X(Y)]| \Rightarrow [XZ(Y)]$. Supóngase que $M\left(X'' \supset Y''\right) = 1$. Si $M\left(X'' \bullet Z''\right) = 1$, por $V\bullet$ se sigue que $M\left(X''\right) = 1$, y como $M\left(X'' \supset Y''\right) = 1$, utilizando $V \supset$ se infiere que $M\left(Y''\right) = 1$, por lo que, $M\left(X'' \bullet Z''\right) = 1$ implica $M\left(Y''\right) = 1$, y por lo tanto, según $V \supset$ se obtiene, $M\left(\left(X'' \bullet Z''\right) \supset Y''\right) = 1$. En consecuencia, $(Regla4)''$ es una regla válida en $KT4P$.

Regla 5: $[Y(W[X(K)])]| \Rightarrow [Y(W[XZ(K)])]$. Supuesto 1, $M\left(Y'' \supset \left(W'' \bullet \left(X'' \supset K''\right)\right)\right) = 1$. Supuesto 2, Y'', por $V \supset$ y supuesto 1 se deriva $M\left(W'' \bullet \left(X'' \supset K''\right)\right) = 1$, utilizando $V\bullet$ se obtiene $M\left(W''\right) = 1$ y $M\left(X'' \supset K''\right) = 1$. Supuesto 3, $M\left(X'' \bullet Z''\right) = 1$, por $V\bullet$ resulta $M\left(X''\right) = 1$, y como se tiene $M\left(X'' \supset K''\right) = 1$, por $V \supset$ se deriva $M\left(K''\right) = 1$, por lo que, $M\left(X'' \bullet Z''\right) = 1$ implica $M\left(K''\right) = 1$,

lo cual por $V\supset$ significa que $M\left(\left(X''\bullet Z''\right)\supset K\right)=1$, y como se tiene $M\left(W''\right)=1$, por $V\bullet$ se sigue $M\left(W''\bullet\left(\left(X''\bullet Z''\right)\supset K\right)\right)=1$. Descargando el supuesto 2 se deduce $M\left(Y''\supset\left(W''\bullet\left(\left(X''\bullet Z''\right)\supset K\right)\right)\right)=1$. Por lo tanto, $(\text{Regla } 5)''$ es una regla válida en $KT4P$.

Regla 6: $[X\,[Z\,(Y)\,\_]\,(Z)]\,|\Rightarrow[X\,\{Z\,\{Y\}\}\,(Z)]$. Supóngase que $M\left(\left(X''\bullet\left(Z''\supset Y''\right)\right)\supset Z''\right)=1$. Supuesto 2, $M\left(\left(X''\bullet-\left(Z''\bullet-Y''\right)\right)\supset Z''\right)=0$, por $V\supset$ se obtiene $M\left(X''\bullet-\left(Z''\bullet-Y''\right)\right)=1$ y $M\left(Z''\right)=0$, por $V\bullet$ se sigue $M\left(-\left(Z''\bullet-Y''\right)\right)=1$ y $M\left(X''\right)=1$, como $M\left(Z''\right)=0$, utilizando el supuesto inicial, por $V\supset$ resulta $M\left(X''\bullet\left(Z''\supset Y''\right)\right)=0$, y como $M\left(X''\right)=1$, por $V\bullet$ se deriva $M\left(Z''\supset Y''\right)=0$, lo cual por $V\supset$ significa $M\left(Y''\right)=0$ y $M\left(Z''\right)=1$, lo cual no es el caso, por lo que, $M\left(\left(X''\bullet-\left(Z''\bullet-Y''\right)\right)\supset Z''\right)=1$. Por lo tanto, $(\text{Regla } 6)''$ es una regla válida en $KT4P$.

Regla 7: $[(*X)\,(Y)]\Rightarrow[\{*X\}\,(Y)]$. Supóngase que $M\left(+X''\cup Y''\right)=1$. Supuesto 2, $M\left(-+X''\right)=1$, aplicando $V-+$ se sigue $M\left(+X''\right)=0$, aplicando $V\cup$, en el supuesto inicial se infiere $M\left(Y''\right)=1$, por lo que, $M\left(-+X''\right)=1$ implica $M\left(Y''\right)=1$, utilizando $V\supset$ se concluye que $M\left(-+X''\supset Y''\right)=1$. Por lo tanto, $(\text{Regla } 7)''$ es una regla válida en $KT4P$.

Regla 8: $[(Y)\,\{X\}]\,|\Rightarrow[(Y)\,(X)]$. Supuesto 1, $M\left(-X''\supset Y''\right)=1$. Supuesto 2, $M\left(X''\cup Y''\right)=0$, lo cual por $V\cup$ significa $M\left(X''\right)=0$ y $M\left(Y''\right)=0$, utilizando el supuesto 1 y $V\supset$ implica que $M\left(-X''\right)=0$, por $V-$ se sigue para todo modelo N, si $M<N$ entonces $N\left(X''\right)=1$, por la restricción RR se sabe que $M<M$, de donde $M\left(X''\right)=1$, lo cual no es el caso, en consecuencia, $M\left(X''\cup Y''\right)=1$. Por lo tanto, $(\text{Regla } 8)''$ es una regla válida en $KT4P$.

Regla 9: $[(Y)\,X]\,|\Rightarrow[(Y)\,(\{X\})]$. Supuesto 1, $M\left(X''\supset Y''\right)=1$. Supuesto 2, $M\left(Y''\cup-X''\right)=0$, por $V\cup$ se infiere $M\left(Y''\right)=0$ y $M\left(-X''\right)=0$, aplicando $V-$ y la restricción RR se sigue $M\left(X''\right)=1$, por el supuesto 1 y $V\supset$ se genera $M\left(Y''\right)=1$, lo cual es imposible, entonces $M\left(Y''\cup-X''\right)=1$. Por lo tanto, $(\text{Regla } 9)''$ es una regla válida en $KT4P$.

Regla 11: $X\Leftrightarrow[(X)]$. Supóngase que $M\left(X''\right)=1$. Por $V\supset$ resulta que $M\left(\lambda''\supset X''\right)=1$. Para la recíproca, supóngase que $M\left(\lambda''\supset X''\right)=1$, como por $V\lambda$ se tiene que $M\left(\lambda''\right)=1$, utilizando $V\supset$ se deriva $M\left(X''\right)=1$. Se ha probado que, $M\left(X''\right)=1$ es equivalente a $M\left(\lambda''\supset X''\right)=1$. Por lo tanto, $(\text{Regla } 11)''$ es una regla válida en $KT4P$.

Regla 12: $\{X\}\Leftrightarrow\{[(X)]\}$. $M\left(-X''\right)=1$ equivale a, existe $P\in S$, $P<M$ y $P\left(X''\right)=0$, por la regla

$(\text{Regla } 11)''$, probada en el párrafo anterior, es equivalente a existe $P \in S$, $P < M$ y $P\left(\lambda'' \supset X''\right) = 0$, lo cual por $V-$ equivale a $M\left(-\left(\lambda'' \supset X''\right)\right) = 1$. Por lo tanto, $(\text{Regla } 12)''$ es una regla válida en $KT4P$.

Regla 13: $[Y(X)] \Leftrightarrow [Y([(X)])]$. $M\left(Y'' \supset X''\right) = 1$ por $V \supset$ equivale a, $M\left(Y''\right) = 1$ implica $M\left(X''\right) = 1$, aplicando $(\text{Regla } 11)''$ (probado más arriba) equivale a, $M(Y) = 1$ implica $M\left(\lambda'' \supset X''\right) = 1$, de nuevo por $V \supset$ es equivalente a $M\left(Y \supset \left(\lambda'' \supset X''\right)\right) = 1$. Por lo tanto, $(\text{Regla } 13)''$ es una regla válida en $KT4P$.

Regla 14: $[(Y)(X)] \Leftrightarrow [(Y)([(X)])]$. $M\left(Y'' \cup X''\right) = 1$ por $V\cup$ equivale a, $M\left(Y''\right) = 1$ o $M\left(X''\right) = 1$, aplicando $(\text{Regla } 11)''$ (probado más arriba) equivale a, $M(Y) = 1$ o $M\left(\lambda'' \supset X''\right) = 1$, de nuevo por $V\cup$ es equivalente a $M\left(Y \cup \left(\lambda'' \supset X''\right)\right) = 1$. Por lo tanto, $(\text{Regla } 14)''$ es una regla válida en $KT4P$.

Regla 15: $\{X\} \Leftrightarrow \{*X\}$. Supóngase que, $M\left(-+X''\right) = 0$, aplicando $V-+$ se sigue que $\left(+X''\right) = 1$, por $V+$ se tiene que, para todo $N \in S$, $M < N$ implica $N\left(X''\right) = 1$, lo cual por $V-$ significa que $M\left(-X''\right) = 0$, por lo que, $M\left(-+X''\right) = 0$ implica $M\left(-X''\right) = 0$, es decir, $M\left(-X''\right) = 1$ implica $M\left(-+X''\right) = 1$. Para probar la recíproca, supóngase que, $M\left(-+X''\right) = 1$, por $V-+$ se deriva $M\left(+X''\right) = 0$, lo cual por $V+$ significa que, existe $P \in S$, $M < P$ y $P\left(X''\right) = 0$, utilizando $V-$ se infiere $M\left(-X''\right) = 1$, por lo que, $M\left(-+X''\right) = 1$ implica $M\left(-X''\right) = 1$. Como también se tiene la recíproca, entonces $M\left(-+X''\right) = 1$ equivale a $M\left(-X''\right) = 1$. Por lo tanto, $(\text{Regla } 15)''$ es una regla válida en $KT4P$.

Regla 16: $\{[*X(Y)]\} \mid \Rightarrow \{[X(Y)]\}$. Supóngase que $M\left(-\left(+X'' \supset Y''\right)\right) = 1$. Por $V-$ significa existe $P \in S$, $M < P$ y $P\left(+X'' \supset Y''\right) = 0$, lo cual por $V \supset$ significa $P\left(+X''\right) = 1$ y $P\left(Y''\right) = 0$, y por $V+$ se infiere, para todo $N \in S$, $P < N$ implica $N\left(X''\right) = 1$, como por la restricción RR, $P < P$ resulta que $P\left(X''\right) = 1$. Si se tuviese que $M\left(-\left(X'' \supset Y''\right)\right) = 0$, entonces por $V \supset$ se tendría para todo $N \in S$, $M < N$ implica $M\left(X'' \supset Y''\right) = 1$, en particular, como $M < P$, entonces $P\left(X'' \supset Y''\right) = 1$, y como ya se obtuvo $P\left(X''\right) = 1$, por $V \supset$ se deduce $P\left(Y''\right) = 1$, lo cual no es el caso, por lo que $M\left(-\left(X'' \supset Y''\right)\right) = 1$. Por lo tanto, $(\text{Regla } 16)''$ es una regla válida en $KT4P$.

Regla 17: $Z[X(Y)] \Leftrightarrow Z[XZ(Y)]$. Supóngase que $M\left(Z'' \bullet \left(X'' \supset Y''\right)\right) = 1$, por $V\bullet$ equivale a $M\left(Z''\right) = 1$ y $M\left(X'' \supset Y''\right) = 1$. Supóngase que $M\left(X'' \bullet Z''\right) = 1$, por $V\bullet$ se sigue $M\left(X''\right) = 1$, como se tiene $M\left(X'' \supset Y''\right) = 1$, por $V \supset$ se infiere $M\left(Y''\right) = 1$, por lo que $M\left(\left(X'' \bullet Z''\right) \supset Y''\right) = 1$, como además se tiene $M\left(Z''\right) = 1$, por $V\bullet$ se concluye $M\left(Z'' \bullet \left(\left(X'' \bullet Z''\right) \supset Y''\right)\right) = 1$. Por lo tanto, la implicación directa de $(\text{Regla } 17)''$ es una regla válida en $KT4P$.

Para la implicación recíproca, supóngase que $M\left(Z''\bullet\left(\left(X''\bullet Z''\right)\supset Y''\right)\right)=1$, por $V\bullet$ se sigue $M\left(Z''\right)=1$ y $M\left(X''\bullet Z''\supset Y''\right)=1$. Supuesto 2, $M\left(X''\right)=1$, y como $M\left(Z''\right)=1$, por $V\bullet$ se obtiene $M\left(X''\bullet Z''\right)=1$, pero $M\left(\left(X''\bullet Z''\right)\supset Y''\right)=1$, aplicando $V\supset$ se infiere $M\left(Y''\right)=1$, por lo que, $M\left(X''\right)=1$ implica $M\left(Y''\right)=1$, lo cual por $V\supset$ significa $M\left(X''\supset Y''\right)=1$, como $M\left(Z''\right)=1$, por $V\bullet$ se deriva $M\left(Z''\bullet(X\supset Y)\right)=1$. Por lo tanto, la implicación recíproca de $(\text{Regla 17})''$ es una regla válida en $KT4P$.

Regla 18: $Z[Y(X)]\Leftrightarrow Z[Y(XZ)]$. Supóngase que $M\left(Z''\bullet\left(Y''\supset X''\right)\right)=1$, lo cual por $V\bullet$ implica $M\left(Z''\right)=1$ y $M\left(Y''\supset X''\right)=1$. Supuesto 2, $M\left(Z''\bullet\left(Y''\supset\left(X''\bullet Z''\right)\right)\right)=0$, y como $M\left(Z''\right)=1$, por $V\bullet$ se sigue $M\left(Y''\supset\left(X''\bullet Z''\right)\right)=0$, aplicando $V\supset$ se deriva $M\left(Y''\right)=1$ y $M\left(X''\bullet Z''\right)=0$. Se tienen $M\left(Y''\supset X''\right)=1$ y $M\left(Y''\right)=1$, por $V\supset$ resulta $M\left(X''\right)=1$, pero $M\left(X''\bullet Z''\right)=0$, según $V\bullet$ se infiere $M\left(Z''\right)=0$, lo cual no es el caso, por lo que el supuesto 2 es falso, es decir $M\left(Y''\supset\left(X''\bullet Z''\right)\right)=1$, y como $M\left(Z''\right)=1$, por $V\bullet$ se sigue $M\left(Z''\bullet\left(Y''\supset\left(X''\bullet Z''\right)\right)\right)=1$. Por lo tanto, la implicación directa de $(\text{Regla 18})''$ es una regla válida en $KT4P$.
Para la implicación recíproca, supóngase que $M\left(Z''\bullet\left(Y''\supset\left(X''\bullet Z''\right)\right)\right)=1$, lo cual por $V\bullet$ significa, $M\left(Z''\right)=1$ y $M\left(Y''\supset\left(X''\bullet Z''\right)\right)=1$, por $V\supset$ resulta que, $M\left(Y''\right)=1$ implica $M\left(X''\bullet Z''\right)=1$, pero de $M\left(X''\bullet Z''\right)=1$ por $V\bullet$ se deriva $M\left(X''\right)=1$, luego, $M\left(Y''\right)=1$ implica $M\left(X''\right)=1$, es decir $M\left(Y''\supset X''\right)=1$, y como $M\left(Z''\right)=1$, por $V\bullet$ se sigue que $M\left(Z''\bullet\left(Y''\supset X''\right)\right)=1$. Por lo tanto, la implicación recíproca de $(\text{Regla 18})''$ es una regla válida en $KT4P$.

Regla 18.1: $Z[(Y)(X)]\Leftrightarrow Z[(Y)(XZ)]$. Supóngase que $M\left(Z''\bullet\left(Y''\cup X''\right)\right)=1$, lo cual por $V\bullet$ implica $M\left(Z''\right)=1$ y $M\left(Y''\cup X''\right)=1$. Supuesto 2, $M\left(Z''\bullet\left(Y''\cup\left(X''\bullet Z''\right)\right)\right)=0$, y como $M\left(Z''\right)=1$, por $V\bullet$ se sigue $M\left(Y''\cup\left(X''\bullet Z''\right)\right)=0$, aplicando $V\bullet$ se deriva $M\left(Y''\right)=0$ y $M\left(X''\bullet Z''\right)=0$, pero $M\left(Z''\right)=1$, aplicando $V\bullet$ se deriva $M\left(X''\right)=0$, y como $M\left(Y''\cup X''\right)=1$, por $V\cup$ se obtiene $M\left(Y''\right)=1$, lo cual no es el caso, por lo que, $M\left(Z''\bullet\left(Y''\cup\left(X''\bullet Z''\right)\right)\right)=1$. Por lo tanto, la implicación directa de $(\text{Regla 18.1})''$ es una regla válida en $KT4P$.
Para la implicación recíproca, supóngase que $M\left(Z''\bullet\left(Y''\cup\left(X''\bullet Z''\right)\right)\right)=1$, lo cual por $V\bullet$ significa, $M\left(Z''\right)=1$ y $M\left(Y''\cup\left(X''\bullet Z''\right)\right)=1$, por $V\cup$ resulta que, $M\left(Y''\right)=1$ o $M\left(X''\bullet Z''\right)=1$, pero de $M\left(X''\bullet Z''\right)=1$ por $V\bullet$ se deriva $M\left(X''\right)=1$, luego, $M\left(Y''\right)=1$ o $M\left(X''\right)=1$, es decir $M\left(Y''\cup X''\right)=1$, y como $M\left(Z''\right)=1$, aplicando $V\bullet$ se concluye que $M\left(Z''\bullet\left(Y''\cup X''\right)\right)=1$. Por lo tanto, la implicación recíproca de $(\text{Regla 18.1})''$ es una regla válida en $KT4P$.

Regla 19: $[YZ(X)]\Leftrightarrow[YZ(XZ)]$. Supóngase que $M\left(\left(Y''\bullet Z''\right)\supset X''\right)=1$. Supuesto 2, $M\left(Y''\bullet Z''\right)=1$,

por $V\supset$ resulta $M\left(X''\right)=1$, del supuesto 2 por $V\bullet$ se deriva $M\left(Z''\right)=1$, y al tener $M\left(X''\right)=1$, de nuevo por $V\bullet$ se deduce $M\left(X''\bullet Z''\right)=1$, por lo que, $M\left(Y''\bullet Z''\right)=1$ implica $M\left(X''\bullet Z''\right)=1$, lo cual por $V\supset$ significa $M\left(\left(Y''\bullet Z''\right)\supset\left(X''\bullet Z''\right)\right)=1$. Por lo tanto, la implicación directa de $(\text{Regla 19})''$ es una regla válida en $KT4P$.
Para la implicación recíproca, supóngase que $M\left(\left(Y''\bullet Z''\right)\supset\left(X''\bullet Z''\right)\right)=1$. Si $M\left(Y''\bullet Z''\right)=1$, por $V\supset$ se sigue $M\left(X''\bullet Z''\right)=1$, por $V\bullet$ se infiere $M\left(X''\right)=1$, de donde, $M\left(Y''\bullet Z''\right)=1$ implica $M\left(X''\right)=1$, lo cual por $V\supset$ significa $M\left(\left(Y''\bullet Z''\right)\supset X''\right)=1$. Por lo tanto, la implicación recíproca de $(\text{ID.G.3 Regla 19})''$ es una regla válida en $KT4P$.

Regla 20: $[(Z[X(Y)])(W)]\Leftrightarrow[(Z[ZX(Y)])(W)]$. Supóngase que $M\left(\left(Z''\bullet\left(X''\supset Y''\right)\right)\cup W''\right)=1$. Supuesto 2 $M\left(\left(Z''\bullet\left(\left(Z''\bullet X''\right)\supset Y''\right)\right)\cup W''\right)=0$, por $V\cup$ se obtienen $M\left(Z''\bullet\left(\left(Z''\bullet X''\right)\supset Y''\right)\right)=0$ y $M\left(W''\right)=0$, lo cual junto al supuesto inicial, por $V\cup$ implica $M\left(Z''\bullet\left(X''\supset Y''\right)\right)=1$, por $V\bullet$ implica $M\left(Z''\right)=1 y M\left(X''\supset Y''\right)=1$, y como se tiene $M\left(Z''\bullet\left(\left(Z''\bullet X''\right)\supset Y''\right)\right)=0$ por $V\bullet$ se deriva $M\left(\left(Z''\bullet X''\right)\supset Y''\right)=0$, por $V\supset$ resultan $M\left(Z''\bullet X''\right)=1$ y $M\left(Y''\right)=0$, y como $M\left(X''\supset Y''\right)=1$, por $V\supset$ se afirma $M\left(X''\right)=0$, pero de $M\left(Z''\bullet X''\right)=1$, se sigue por $V\bullet$ que $M\left(X''\right)=1$, generando contradicción, por lo que, el supuesto 2 es falso, es decir $M\left(\left(Z''\bullet\left(\left(Z''\bullet X''\right)\supset Y''\right)\right)\cup W''\right)=1$. Por lo tanto, la implicación directa de $(\text{Regla 20})''$ es una regla válida en $KT4P$.

Para la implicación recíproca, supóngase que $M\left(\left(Z''\bullet\left(\left(Z''\bullet X''\right)\supset Y''\right)\right)\cup W''\right)=1$. Supuesto 2, $M\left(\left(Z''\bullet\left(X''\supset Y''\right)\right)\cup W''\right)=0$, por $V\cup$ se sigue $M\left(Z''\bullet\left(X''\supset Y''\right)\right)=0$ y $M\left(W''\right)=0$, este resultado con el supuesto inicial y $V\cup$ implica $M\left(Z''\bullet\left(\left(Z''\bullet X''\right)\supset Y''\right)\right)=1$, es decir $M\left(\left(Z''\bullet X''\right)\supset Y''\right)=1$ y $M\left(Z''\right)=1$, como $M\left(Z''\bullet\left(X''\supset Y''\right)\right)=0$, por $V\bullet$ se infiere $M\left(X''\supset Y''\right)=0$, es decir $M\left(X''\right)=1$ y $M\left(Y''\right)=0$, al tener $M\left(X''\right)=M\left(Z''\right)=1$, por $V\bullet$ resulta $M\left(X''\bullet Z''\right)=1$, pero $M\left(\left(Z''\bullet X''\right)\supset Y''\right)=1$, entonces por $V\supset$ se deduce $M\left(Y''\right)=1$, lo cual no es el caso, por lo que $M\left(\left(Z''\bullet\left(X''\supset Y''\right)\right)\cup W''\right)=1$. Por lo tanto, la implicación recíproca de $(\text{Regla 20})''$ es una regla válida en $KT4P$.

Regla 21: $[(X)(X)]\Leftrightarrow[(X)]$. Supóngase que $M\left(X''\cup X''\right)=1$. Si se tuviese que $M\left(X''\right)=0$, por $V\cup$ se tendría $M\left(X''\right)=1$, contradicción, por lo que $M\left(X''\right)=1$. Para la recíproca, supóngase que $M\left(X''\right)=1$, pero $M\left(X''\cup X''\right)=0$, por $V\cup$ resulta $M\left(X''\right)=0$, contradicción, por lo que $M\left(X''\cup X''\right)=1$. Por lo tanto, $(\text{Regla 21})''$ es una regla válida en $KT4P$.

Regla 22: $*X\{*XY\}|\Rightarrow *X\{Y\}$. Supóngase que, $M\left(+X''\bullet-\left(+X''\bullet Y''\right)\right)=1$, por $V\bullet$ se sigue

$M\left(+X''\right)=1$ y $M\left(-\left(+X''\bullet Y''\right)\right)=1$, lo cual según $V-$ significa que existe $P\in S$, $M<P$ y $P\left(+X''\bullet Y''\right)=0$. Supuesto 2, $M\left(+X''\bullet -Y''\right)=0$, y como $M\left(+X''\right)=1$, por $V\bullet$ se sigue $M\left(-Y''\right)=0$, aplicando $V-$, resulta que para cada $N\in S$, $M<N$ implica $N\left(Y''\right)=1$, y como $M<P$, entonces $P\left(Y''\right)=1$, pero $P\left(+X''\bullet Y''\right)=0$, por $V\bullet$ se deriva $P\left(+X''\right)=0$, por $V+$ se infiere que existe $Q\in S$, $P<Q$ y $Q\left(X''\right)=0$, como $M<P$ y $P<Q$ por la restricción RT se sigue $M<Q$, por $V+$ se concluye que $M\left(+X''\right)=0$, lo cual no es el caso, por lo que, $M\left(+X''\bullet -Y''\right)=1$. En consecuencia, $M\left(+X''\bullet -\left(+X''\bullet Y''\right)\right)=1$ implica $M\left(+X''\bullet -Y''\right)=1$. Por lo tanto, $(\text{Regla } 22)''$ es una regla válida en $KT4P$.

Regla 23: $*Z\{Y[X(W)]\}\Leftrightarrow *Z\{Y[X*Z(W)]\}$. Supóngase que $M\left(+Z''\bullet -\left(Y''\bullet\left(X''\supset W''\right)\right)\right)=1$, por $V\bullet$ se siguen $M\left(+Z''\right)=1$ y $M\left(-\left(Y''\bullet\left(X''\supset W''\right)\right)\right)=1$, por $V-$ se sigue la existencia de $P\in S, M<P$ y $P\left(Y''\bullet\left(X''\supset W''\right)\right)=0$.

Supuesto 2, $M\left(+Z''\bullet -\left(Y''\bullet\left(\left(X''\bullet +Z''\right)\supset W''\right)\right)\right)=0$, como $M\left(+Z''\right)=1$, por $V\bullet$ resulta $M\left(-\left(Y''\bullet\left(\left(X''\bullet +Z''\right)\supset W''\right)\right)\right)=0$, aplicando $V-$ se infiere para cada $N\in S$, $M<N$ implica $N\left(Y''\bullet\left(\left(X''\bullet +Z''\right)\supset W''\right)\right)=1$, en particular, como $M<P$ entonces $P\left(Y''\bullet\left(\left(X''\bullet +Z''\right)\supset W''\right)\right)=1$, es decir, por $V\bullet P\left(\left(X''\bullet +Z''\right)\supset W''\right)=1$ y $P\left(Y''\right)=1$, pero $P\left(Y''\bullet\left(X''\supset W''\right)\right)=0$, por $V\bullet$ se afirma $P\left(X''\supset W''\right)=0$, lo cual por $V\supset$ significa $P\left(X''\right)=1$ y $P\left(W''\right)=0$, pero $P\left(\left(X''\bullet +Z''\right)\supset W''\right)=1$, por $V\supset$ se afirma $P\left(X''\bullet +Z''\right)=0$, pero $P\left(X''\right)=1$, por $V\bullet$ resulta $P\left(+Z''\right)=0$ aplicando $V+$ se tiene que existe $Q\in S$, $P<Q$ y $Q\left(Z''\right)=0$, pero como $M<P$ y $P<Q$, por la restricción RT se deriva que $M<Q$, además se sabe que $M\left(+Z''\right)=1$, por $V+$ significa que para todo $N\in S$, $M<N$ implica $N\left(Z''\right)=1$, en particular, como $M<Q$, entonces $Q\left(Z''\right)=1$, lo cual no es el caso, y entonces, $M\left(+Z''\bullet -\left(Y''\bullet\left(\left(X''\bullet +Z''\right)\supset W''\right)\right)\right)=1$. Por lo tanto, la implicación directa de $(\text{Regla } 23)''$ es una regla válida en $KT4P$.

Para probar la recíproca, supóngase que $M\left(+Z''\bullet -\left(Y''\bullet\left(\left(X''\bullet +Z''\right)\supset W''\right)\right)\right)=1$, según $V\bullet$ se tienen $M\left(-\left(Y''\bullet\left(\left(X''\bullet +Z''\right)\supset W''\right)\right)\right)=1$ y $M\left(+Z''\right)=1$, aplicando $V-$ se sigue que existe $P\in S$, tal que $M<P$ y $P\left(Y''\bullet\left(\left(X''\bullet +Z''\right)\supset W''\right)\right)=0$. Supuesto 2, $M\left(+Z''\bullet -\left(Y''\bullet\left(X''\supset W''\right)\right)\right)=0$, como $M\left(+Z''\right)=1$, por $V\bullet$ se infiere $M\left(-\left(Y''\bullet\left(X''\supset W''\right)\right)\right)=0$, utilizando $V-$ se obtiene para cada $N\in S$, $M<N$ implica $N\left(Y''\bullet\left(X''\supset W''\right)\right)=1$, en particular, como $M<P$ entonces $P\left(Y''\bullet\left(X''\supset W''\right)\right)=1$, aplicando $V\bullet$, resulta $P\left(X''\supset W''\right)=1$ y $P\left(Y''\right)=1$, pero $P\left(Y''\bullet\left(\left(X''\bullet +Z''\right)\supset W''\right)\right)=0$, utilizando $V\bullet$ se afirma que $P\left(\left(X''\bullet +Z''\right)\supset W''\right)=0$, por $V\supset$ resulta $P\left(X''\bullet +Z''\right)=1$ y $P\left(W''\right)=0$, y al tener $P\left(X''\supset W''\right)=1$ por $V\supset$ se sigue $P\left(X''\right)=0$, pero, como

$P\left(X'' \bullet +Z''\right) = 1$, por $V\bullet$ se infiere $P\left(X''\right) = 1$, lo cual es imposible, por lo que el supuesto 2 no es el caso, es decir, $M\left(+Z'' \bullet -\left(Y'' \bullet \left(X'' \supset W''\right)\right)\right) = 1$. Por lo tanto, la implicación recíproca de $(Regla23)''$ es una regla válida en $KT4P$.

Regla 26: $Y \in TG$ implica $*Y \in TG$. Si $+Y''$ es inválida, entonces existe un modelo, $(S, Ma, <, V)$, tal que $Ma(+Y'') = 0$, por $V+$ se infiere que existe $Q \in S$, $Ma < Q$ y $Q(Y'') = 0$, lo cual implica que Y'' es falsa en el modelo $(S - \{Ma\}, Q, <, V)$, es decir, Y'' es inválida. Por lo tanto, $+Y''$ es inválida implica que Y'' es inválida, o de forma equivalente, Y'' es válida implica que $+Y''$ es válida. En consecuencia,$(\text{Regla } 26)''$ se satisface en $KT4P$.

Regla 27:$X, Y \Leftrightarrow YX$. Por $V\bullet$, $M(X'') = 1$ y $M(Y'') = 1$ es equivalente a $M(X'' \bullet Y'') = 1$. Por lo tanto, $(Regla27)''$ es una regla válida en $KT4P$.

Regla 28: $YX \Leftrightarrow YX$. Por $V\bullet$, $M(X'' \bullet Y'') = 1$ es equivalente a $M(X'') = 1$ y $M(Y'') = 1$, lo cual es equivalente a (tienen el mismo significado) $M(Y'') = 1$ y $M(X'') = 1$, y esto por $V\bullet$, es equivalente a $M(Y'' \bullet X'') = 1$. Por lo tanto, $(Regla28)''$ es una regla válida en $KT4P$.

Regla 29: $XY, Z \Leftrightarrow X$, $YZ \Leftrightarrow XYZ$. Aplicando $V\bullet$, se tiene la siguiente cadena de equivalencias, $M((X'' \bullet Y'') \bullet Z'') = 1$, significa $M(X'' \bullet Y'') = 1$ y $M(Z'') = 1$, es decir, $M(X'') = 1$ y $M(Y'') = 1$, y además $M(Z'') = 1$, lo cual significa lo mismo que $M(X'') = 1$, y además, $M(Y'') = 1$ y $M(Z'') = 1$, o sea, $M(X'') = 1$, y $M(Y'' \bullet Z'') = 1$, lo cual es equivalente a $M(X'' \bullet (Y'' \bullet Z'')) = 1$. Por lo tanto, $(Regla29)''$ es una regla válida en $KT4P$. □

Proposición 6.4. *(Los teoremas gráficos son válidos en la semántica de KT4P.) Para $X, Y \in GK$.*

a. Si $X \in TG$ entonces $X'' \in VK$.

b. Si $X >> Y$ entonces X'' valida a Y''.

Prueba. Parte *a*. Por inducción sobre la longitud, L, de la demostración de X en AlfaK.
Paso base. $L = 1$, por lo que X es un axioma, es decir, $X = \lambda$, por lo que $X'' = \lambda'' = \lambda$ y por $V\lambda$, $\lambda \in VK$.
Paso inductivo. Como hipótesis inductiva se tiene que, si $Y \in TG$ y la longitud de la demostración de Y es menor que$L$, entonces $Y'' \in VK$. Supóngase que la longitud de la demostración de X es L. Si X es un axioma, se procede como en el paso base, en caso contrario, X se infiere de pasos anteriores utilizando las reglas de construcción de gráficos, se tiene Z y se aplica la regla $Z \Rightarrow X$, para inferir X. Como la longitud de la demostración de Z es menor que L, por la hipótesis inductiva se afirma que $Z'' \in VK$, como $Z \Rightarrow X$ es una de las reglas de AlfaK, por la proposición 6.3, se garantiza que de X'' se infiere válidamente Z'', por lo tanto, $X'' \in VK$.

Por el principio de inducción matemática, se ha probado que los teoremas gráficos son válidos.
Parte *b*. Si $X >> Y$, es decir, $[X(Y)] \in TG$, por la parte *a* se sigue $X'' \supset Y'' \in VK$, es decir, X'' valida a Y''. □

Proposición 6.5. *(Los teoremas gráficos son teoremas de KT4P). Para $X,Y \in GK$.*

a. Si $X \in TG$ entonces $X'' \in TK$.

b. Si $X >> Y$ entonces Y'' es consecuencia de X''.

Prueba. Parte *a*. Por la Proposición 4.8*a* se tiene que, $X'' \in VK$ si y solo si $X'' \in TK$, y por la Proposición 6.4*a* se tiene que, si $X \in TG$ entonces $X'' \in VK$. Por lo tanto, si $X \in TG$ entonces $X'' \in TK$.
Parte *b*. Consecuencia de las proposiciones 4.8*b* y 6.4*b*. □

Proposición 6.6. *(Los teoremas gráficos son teoremas de KT4P). Para $X,Y \in GK$*

a. Si $X \in TG$ entonces $X'' \in TK$.

b. Si $X >> Y$ entonces Y'' es consecuencia de X''.

Prueba. Parte *a*. Por la Proposición 4.8*a* se tiene que, $X'' \in VK$ si y solo si $X'' \in TK$, y por la Proposición 6.4*a* se tiene que, si $X \in TG$ entonces $X'' \in VK$. Por lo tanto, si $X \in TG$ entonces $X'' \in TK$.
Parte *b*. Consecuencia de las proposiciones 4.8*b* y 6.4*b*. □

Definición 6.3. *(Composición de las funciones de traducción) Sean $T1 = (\_)' : FK \to GK$ y $T2 = (\_)'' : GK \to FK$, las funciones de traducción presentadas en definiciones 6.1 y 6.2.*
Se definen: la función compuesta $T1oT2 : GK \to GK$ tal que $(T1oT2)(X) = T1(T2((X))$, la función compuesta, $T2oT1 : FK \to FK$ tal que $(T2oT1)(X) = T2(T1(X))$, la función identidad en FK, $Id_{FK} : FK \to FK$ tal que $(Id_{FK})(X) = X$, la función identidad en GK, $Id_{GK} : GK \to GK$ tal que $(Id_{GK})(X) = X$.

Definición 6.4. *(Complejidad de los gráficos) Sean $P \in GA$, $X,Y \in GK$.*
La función, C, complejidad de un gráfico, asigna a cada gráfico un entero no negativo, de la siguiente forma:

1. $C(P) = C(\lambda) = 0$
2. $C(\{X\}) = 1 + C(X)$.
3. $C(XY) = 1 + \max\{C(X), C(Y)\}$.
4. $C([(X)(Y)]) = 2 + \max\{C(X), C(Y)\}$.
5. $C([X(Y)]) = 1 + \max\{C(X), C(Y) + 1\}$.

Definición 6.5. *(Complejidad de las fórmulas) Sean $P \in FA$; $X,Y \in FK$.*
La función, K, complejidad de una fórmula, asigna a cada fórmula un entero no negativo, de la siguiente forma:

1. $K(P) = K(\lambda) = 0$.
2. $K(-X) = 1 + K(X)$.
3. $K(X \bullet Y) = K(X \cup Y) = K(X \supset Y) = 1 + \max\{K(X), K(Y)\}$.

Proposición 6.7. *(Funciones de traducción inversas). Sean $P \in GA$, $P \in FA, G, G1, G2 \in GK$ y $X, X1, X2 \in FK$.*

a. $T1oT2 = Id_{GK}$. *b. $T2oT1 = Id_{FK}$.*
c. $T1$ es la función inversa de $T2$. *d. $T2$ es la función inversa de $T1$.*

Prueba. Parte *a*. Inducción sobre la complejidad, *C*, del gráfico *G*.
Paso base. $C(G) = 0$, entonces hay dos casos.
Caso 1: $G = P$. $(T1oT2)(P) = T1(T2(P)) = T1(P) = P$.

Caso 2: $G = \lambda\,.\,(T1oT2)(\lambda) = T1(T2(\lambda)) = T1(\lambda) = \lambda$.
Paso inductivo. $C(G) \geq 1$. Como hipótesis inductiva se tiene que

$$(T1oT2)(G1) = G1,\ (T1oT2)(G2) = G2.$$

Hay cuatro casos. Caso 3:

$$G = \{G1\}.(T1oT2)(\{G1\}) = T1(T2(\{G1\})) = T1(-T2(G1)) = \{T1(T2(G1))\} = \{G1\}.$$

Caso 4: $G = G1G2$.

$$(T1oT2)(G1G2) = T1(T2(G1G2)) = T1(T2(G1) \bullet T2(G2)) = T1(T2(G1))T1(T2(G2)) = G1G2.$$

Caso 5:

$$\begin{aligned} G &= [G1(G2)]\,(T1oT2)([G1(G2)]) = T1(T2([G1(G2)])) = T1(T2(G1) \supset T2(G2)) \\ &= [T1(T2(G1))(T1(T2(G2)))] = [G1(G2)] \end{aligned} \tag{1}$$

Caso 6:

$$\begin{aligned} G &= [(G1)(G2)]\,.(T1oT2)([(G1)(G2)]) = T1(T2([(G1)(G2)])) = T1(T2(G1) \cup T2(G2)) \\ &= [(T1(T2(G1)))(T1(T2(G2)))] = [(G1)(G2)] \end{aligned} \tag{2}$$

Por el principio de inducción matemática se ha probado que $(T1oT2) = Id_{GK}$.
Parte b. Inducción sobre la complejidad, K, de la fórmula X.
Paso base. $K(X) = 0$, entonces hay dos casos.
Caso 1: $X = P.(T2oT1)(P) = T2(T1(P)) = T2(P) = P$.
Caso 2: $X = \lambda\,.\,(T2oT1)(\lambda) = T2(T1(\lambda)) = T2(\lambda) = \lambda$.
Paso inductivo. $K(X) \geq 1$. Como hipótesis inductiva se tiene que

$$(T2oT1)(X1) = X1,\ (T2oT1)(X2) = X2.$$

Hay cuatro casos. Caso 3.

$$X = -X1.\,(T2oT1)(-X1) = T2(T1(-X1)) = T2(\{T1(X1)\}) = -(T2oT1)(X1) = -X1.$$

Caso 4. $X = X1 \bullet X2$.

$$\begin{aligned} (T2oT1)(X1 \bullet X2) &= T2(T1(X1 \bullet X2)) = T2(T1(X1)T1(X2)) = T2(T1(X1)) \bullet T2(T1(X2)) \\ &= X1 \bullet X2 \end{aligned} \tag{3}$$

Caso 5. $X = X1 \supset X2$.

$$\begin{aligned} (T2oT1)(X1 \supset X2) &= T2(T1(X1 \supset X2)) = T2([T1(X1)(T1(X2))]) \\ &= T2(T1(X1)) \supset T2(T1(X2)) = X1 \supset X2 \end{aligned} \tag{4}$$

Caso 6. $X = X1 \cup X2$.

$$(T2oT1)(X1 \cup X2) = T2(T1(X1 \cup X2)) = T2([(T1(X1))(T1(X2))]) \\ = T2(T1(X1)) \cup T2(T1(X2)) = X1 \cup X2. \tag{5}$$

Por el principio de inducción matemática se ha probado que $(T2oT1) = Id_{FK}$.
Partes c y d. Consecuencia directa de las partes a y b. □

Proposición 6.8. *(KT4P y AlfaK son isomorfos)* $G,H \in GK$ *y* $X,Y \in FK$.
a. $G \in TG$ *si y solo si* $G'' \in TK$. b. $X' \in TG$ *si y solo si* $X \in TK$.
c. $G >> H$ *si y solo si* H'' *es consecuencia de* G''.
d. $X' >> Y'$ *si y solo si* Y *es consecuencia de* X.

Prueba. Parte a. Por la proposición 6.6a se tiene que, si $G \in TG$ entonces $G'' \in TK$, además, por la proposición 6.2a se tiene que, si $G'' \in TK$ entonces $\left(G''\right)' \in TG$, pero por la proposición 6.7a se sabe que, $\left(G''\right)' = G$, resultando que, si $G'' \in TK$ entonces $G \in TG$, y como se tiene la recíproca, se concluye que, $G \in TG$ si y solo si $G'' \in TK$.
Parte b. Por la proposición 6.2a se tiene que, si $X \in TK$ entonces $X' \in TG$, además, por la proposición 6.6a se tiene que, si $X' \in TG$ entonces $(X')'' \in TK$, pero por la proposición 6.7b se sabe que, $\left(X'\right)'' = X$, resultando que, si $X' \in TG$ entonces $X \in TK$, y como se tiene la recíproca, se concluye que, $X' \in TG$ si y solo si $X \in TK$.
Parte c. Por la proposición 5.1b., $G >> H$ equivale a $[G(H)] \in TG$, lo cual por la parte a, equivale a $G'' \supset H'' \in TK$, y por TD, Mp y la definición 2.3, equivale a H'' es consecuencia de G''.
Parte d. Por la proposición 5.1b., $X' >> Y'$ equivale a $\left[X'\left(Y'\right)\right] \in TG$, lo cual por la parte b, equivale a $X \supset Y \in TK$, y por TD, Mp y la definición 2.3, equivale a Y es consecuencia de X. □

Hecho 6.1. *(Reglas derivadas en AlfaK). Las reglas* 10*,* 24 *y* 25 *son derivadas.*

Prueba. Por la proposición 6.8, basta observar que, las traducciones de estas reglas son válidas en $KT4P$. □

7. ALFAK EN EL ESTILO ORIGINAL DE CHARLES SANDERS PEIRCE

En esta sección, se presenta el sistema AlfaK_P, de gráficos existenciales paraconsistentes, en el formato original de Peirce (borrado en regiones pares, escritura en regiones impares, iteración, desiteración), con reglas para los rizos y bucles. En la proposición 7.4, se prueba que los teoremas gráficos de AlfaK son teoremas gráficos de AlfaK_P, en la proposición 7.5, se prueba la recíproca, concluyéndose en la proposición 7.6 la equivalencia de éstos.

Proposición 7.1. *(Interiorizando las inferencias en KT4P). Para* $X,Y,Z,W,V \in FK$.

Si $X \supset Y \in VK$ *entonces:*

a. $(X \bullet Z) \supset (Y \bullet Z) \in VK$. *b.* $-(Y \bullet Z) \supset -(X \bullet Z) \in VK$.

c. $((Z \bullet Y) \supset W) \supset ((Z \bullet X) \supset W) \in VK$. *d.* $-(W \bullet -(X \bullet Z)) \supset -(W \bullet -(Y \bullet Z)) \in VK$.

e. $-((X \bullet Z) \supset W) \supset -((Y \bullet Z) \supset W) \in VK$. *f.* $((Z \bullet (X \supset W)) \supset V) \supset ((Z \bullet (Y \supset W)) \supset V) \in VK$.

g. $(W \supset (X \bullet Z)) \supset (W \supset (Y \bullet Z)) \in VK$. *h.* $((W \bullet -(X \bullet Z)) \supset V) \supset ((W \bullet -(Y \bullet Z)) \supset V) \in VK$.

Prueba. Supóngase que $X \supset Y \in VK$, por las proposiciones 2.1 y 4.8 se infiere que $+(X \supset Y) \in VK$, por lo que, se tiene el resultado inicial, para todo modelo, $(S,Ma,<,V)$, y para cada $M \in S$, $M(+(X \supset Y)) = 1$ y $M(X \supset Y) = 1$.

Parte *a*. Supóngase que $M(X \bullet Z) = 1$, por $V\bullet$ se tiene $M(X) = 1$, $M(Z) = 1$, y como $M(X \supset Y) = 1$, por $V \supset$ se sigue $M(Y) = 1$, utilizando $V\bullet$ se deriva $M(Y \bullet Z) = 1$. Se ha probado, $M(X \bullet Z) = 1$ implica $M(Y \bullet Z) = 1$, lo cual por $V \supset$ significa que $M((X \bullet Z) \supset (Y \bullet Z)) = 1$, para todo $M \in S$. Por lo tanto, $(X \bullet Z) \supset (Y \bullet Z) \in VK$.

Parte *b*. Supóngase que $M(-(Y \bullet Z)) = 1$, entonces por $V-$, existe $P \in S, M < P$ y $P(Y \bullet Z) = 0$. Supuesto 2, $M(-(X \bullet Z)) = 0$, por $V-$ se sigue que, para cada $N \in S$, $M < N$ implica $N(X \bullet Z) = 1$, y como $M < P$ entonces $P(X \bullet Z) = 1$, por $V\bullet$ se tiene $P(X) = 1$ y $P(Z) = 1$, como $P(Y \bullet Z) = 0$, por $V\bullet$ se deriva $P(Y) = 0$, pero como $P(X) = 1$, y por el resultado inicial, $P(X \supset Y) = 1$, entonces por $V \supset$, se obtiene $P(Y) = 1$, lo cual no es el caso, por lo tanto, $M(-(X \bullet Z)) = 1$. Se ha probado que, $M(-(Y \bullet Z)) = 1$ implica $M(-(X \bullet Z)) = 1$, lo cual por $V \supset$ significa que $M(-(Y \bullet Z) \supset -(X \bullet Z)) = 1$, para todo $M \in S$. Por lo tanto, $-(Y \bullet Z) \supset -(X \bullet Z) \in VK$.

Parte *c*. Supóngase que $M((Z \bullet Y) \supset W) = 1$. Supuesto 2, $M((Z \bullet X) \supset W) = 0$, por $V \supset$ se obtiene $M(Z \bullet X) = 1$ y $M(W) = 0$, por $V\bullet$ se sigue $M(Z) = 1$ y $M(X) = 1$, como $M((Z \bullet Y) \supset W) = 1$ y $M(W) = 0$, por $V \supset$ resulta $M(Z \bullet Y) = 0$, y como $M(Z) = 1$, por $V\bullet$ se obtiene $M(Y) = 0$, pero se tiene $M(X) = 1$, y por el resultado inicial, $M(X \supset Y) = 1$, entonces por $V \supset$, se obtiene $M(Y) = 1$, lo cual no es el caso, por lo tanto, $M(Z \bullet X \supset W) = 1$. Se ha probado que, $M(Z \bullet Y \supset W) = 1$ implica $M(Z \bullet X \supset W) = 1$, lo cual por $V \supset$ significa que $M((Z \bullet Y \supset W) \supset (Z \bullet X \supset W)) = 1$, para todo $M \in S$. Por lo tanto, $(Z \bullet Y \supset W) \supset (Z \bullet X \supset W) \in VK$.

Parte *d*. Supóngase que $M(-(W \bullet -(X \bullet Z))) = 1$, entonces por $V-$, existe $P \in S, M < P$ y $P(W \bullet -(X \bullet Z)) = 0$. Supuesto 2, $M(-(W \bullet -(Y \bullet Z))) = 0$, por $V-$ se sigue que, para cada $N \in S$, $M < N$ implica $N(W \bullet -(Y \bullet Z)) = 1$, y como $M < P$ entonces $P(W \bullet -(Y \bullet Z)) = 1$, por $V\bullet$ se tiene $P(W) = 1$ y $P(-(Y \bullet Z)) = 1$, por $V-$ se sigue que, existe $Q \in S$, $P < Q$ y $Q(Y \bullet Z) = 0$, como $P(W) = 1$ y $P(W \bullet -(X \bullet Z)) = 0$, por $V\bullet$ se deriva $P(-(X \bullet Z)) = 0$, lo cual por $V-$ se deriva que, para cada $N \in S$, $P < N$ implica que $N(X \bullet Z) = 1$, como $P < Q$, en particular, $Q(X \bullet Z) = 1$, por $V\bullet$ se sigue $Q(X) = 1$ y $Q(Z) = 1$, por el resultado inicial, se tiene $Q(X \supset Y) = 1$, por $V \supset$ se sigue $Q(Y) = 1$, y como $Q(Z) = 1$, por $V\bullet$ se deriva $Q(Y \bullet Z) = 1$, lo cual no es el caso, por lo tanto, $M(-(W \bullet -(Y \bullet Z))) = 1$. Se ha probado que, $M(-(W \bullet -(X \bullet Z))) = 1$ implica $M(-(W \bullet -(Y \bullet Z))) = 1$, lo cual por $V \supset$ significa que $M(-(W \bullet -(X \bullet Z)) \supset -(W \bullet -(Y \bullet Z))) = 1$, para todo $M \in S$. Por lo tanto, $-(W \bullet -(X \bullet Z)) \supset -(W \bullet -(Y \bullet Z)) \in VK$.

Parte *e*. Supóngase que $M(-((X \bullet Z) \supset W)) = 1$, entonces por $V-$, existe $P \in S, M < P$ y $P((X \bullet Z) \supset W) =$

0, lo cual por $V\supset$ significa $P(X\bullet Z)=1$ y $P(W)=0$, por $V\bullet$ se sigue $P(X)=1$ y $P(Z)=1$. Supuesto 2, $M(-((Y\bullet Z)\supset W))=0$, por $V-$ se sigue que, para cada $N\in S$, $M<N$ implica $N((Y\bullet Z)\supset W)=1$, y como $M<P$ entonces $P((Y\bullet Z)\supset W)=1$, por el resultado inicial, se tiene que $P(X\supset Y)=1$, y como $P(X)=1$, por $V\supset$ se infiere $P(Y)=1$, al tener $P(Z)=1$, por $V\bullet$ se tiene $P(Y\bullet Z)=1$, este resultado junto con el hecho $P((Y\bullet Z)\supset W)=1$, genera $P(W)=1$, lo cual no es el caso, por lo tanto, $M(-((Y\bullet Z)\supset W))=1$. Se ha probado que, $M(-((X\bullet Z)\supset W))=1$ implica $M(-((Y\bullet Z)\supset W))=1$, lo cual por $V\supset$ significa que $M(-((X\bullet Z)\supset W)\supset -((Y\bullet Z)\supset W))=1$, para todo $M\in S$. Por lo tanto, $-((X\bullet Z)\supset W)\supset -((Y\bullet Z)\supset W)\in VK$.

Parte f. Supóngase que $M((Z\bullet(X\supset W))\supset V)=1$. Supuesto 2, $M((Z\bullet(Y\supset W))\supset V)=0$, por $V\supset$ se obtiene $M(Z\bullet(Y\supset W))=1$ y $M(V)=0$, por $V\bullet$ se sigue $M(Z)=1$ y $M(Y\supset W)=1$. Al tener $M((Z\bullet(X\supset W))\supset V)=1$ y $M(V)=0$, por $V\supset$ se deriva $M(Z\bullet(X\supset W))=0$, como se tiene $M(Z)=1$, por $V\bullet$ se deduce $M(X\supset W)=0$, por $V\supset$ significa $M(X)=1$, $M(W)=0$, por el resultado inicial, se sabe que $M(X\supset Y)=1$, y como $M(X)=1$, por $V\supset$ se deriva $M(Y)=1$, lo cual junto con $M(Y\supset W)=1$, por $V\supset$ se obtiene $M(W)=1$, lo cual no es el caso, por lo tanto, $M((Z\bullet(Y\supset W))\supset V)=1$. Se ha probado que, $M((Z\bullet(X\supset W))\supset V)=1$ implica $M((Z\bullet(Y\supset W))\supset V)=1$, lo cual por $V\supset$ significa que $M(((Z\bullet(X\supset W))\supset V)\supset((Z\bullet(Y\supset W))\supset V))=1$, para todo $M\in S$. Por lo tanto, $((Z\bullet(X\supset W))\supset V)\supset((Z\bullet(Y\supset W))\supset V)\in VK$.

Parte g. Supóngase que $M(W\supset(X\bullet Z))=1$. Supuesto 2, $M(W)=1$, por $V\supset$ se tiene $M(X\bullet Z)=1$, por $V\bullet$ se sigue $M(X)=1$ y $M(Z)=1$, por el resultado inicial, se tiene que $M(X\supset Y)=1$, y como $M(X)=1$, por $V\supset$ resulta $M(Y)=1$, y por $V\bullet$ se obtiene $M(Y\bullet Z)=1$, resultando que, $M(W)=1$ implica $M(Y\bullet Z)=1$, lo cual por $V\supset$, significa $M(W\supset(Y\bullet Z))=1$. Se ha probado que, $M(W\supset(X\bullet Z))=1$ implica $M(W\supset(Y\bullet Z))=1$, lo cual por $V\supset-$ significa que $M((W\supset(X\bullet Z))\supset(W\supset(Y\bullet Z)))=1$, para todo $M\in S$. Por lo tanto, $(W\supset(X\bullet Z))\supset(W\supset(Y\bullet Z))\in VK$.

Parte h. Supóngase que $M((W\bullet-(X\bullet Z))\supset V)=1$. Supuesto 2, $M((W\bullet-(Y\bullet Z))\supset V)=0$, por $V\supset$ se infiere $M(W\bullet-(Y\bullet Z))=1$ y $M(V)=0$, por $V\bullet$ se sigue $M(W)=1$ y $M(-(Y\bullet Z))=1$, aplicando $V-$, existe $P\in S$, $M<P$ y $P(Y\bullet Z)=0$, como $M((W\bullet-(X\bullet Z))\supset V)=1$ y $M(V)=0$, por $V\supset$ se deriva $M(W\bullet-(X\bullet Z))=0$, y como $M(W)=1$, por $V\bullet$ se deduce $M(-(X\bullet Z))=0$, por $V-$, para cada $N\in S$, $M<N$ implica $N(X\bullet Z)=1$, y como $M<P$, entonces, en particular $P(X\bullet Z)=1$, por $V\bullet$ significa $P(X)=1$ y $P(Z)=1$, por el resultado inicial, se sabe que $P(X\supset Y)=1$, y como $P(X)=1$, por $V\supset$ resulta $P(Y)=1$, y como $P(Z)=1$, por $V\bullet$ se sigue $P(Y\bullet Z)=1$, lo cual no es el caso, por lo tanto, $M((W\bullet-(Y\bullet Z))\supset V)=1$.

Se ha probado que, $M((W\bullet-(X\bullet Z))\supset V)=1$ implica $M((W\bullet-(Y\bullet Z))\supset V)=1$, lo cual por $V\supset$ significa que $M(((W\bullet-(X\bullet Z))\supset V)\supset((W\bullet-(Y\bullet Z))\supset V))=1$, para todo $M\in S$. Por lo tanto, $h.$ $((W\bullet-(X\bullet Z))\supset V)\supset((W\bullet-(Y\bullet Z))\supset V)\in VK$. □

Proposición 7.2. *(Interiorizando las inferencias en AlfaK). Para $X,Y,Z,W,V\in GK$.*
Si $X>>Y$ entonces:

$a.\ XZ >> YZ.$ $b.\ \{YZ\} >> \{XZ\}.$
$c.\ [ZY(W)] >> [ZX(W)].$ $d.\ \{W\{XZ\}\} >> \{W\{YZ\}\}.$
$e.\ \{[XZ(W)]\} >> \{[YZ(W)]\}.$ $f.\ [Z[X(W)](V)] >> [Z[Y(W)](V)].$
$g.\ [W(XZ)] >> [W(YZ)].$ $h.\ [W\{XZ\}(V)] >> [W\{YZ\}(V)].$

Prueba. Si $X >> Y$, por la proposición 5.1b (TDG), se infiere que $[X(Y)] \in TG$. Considerando este hecho, las partes $a, b, \ldots, h$ de esta proposición (proposición 7.2), son consecuencia directa de las proposiciones 6.8 y 7.1. □

Proposición 7.3. *(Inferencias gráficas en regiones pares e impares). Para* $X, Y, Z, W, V \in GK$.
Si $X >> Y$ *entonces*
$a.\ X >> Y$ *en cualquier región par.* $b.\ Y >> X$ *en cualquier región impar.*

Prueba. Inducción en el número, n, de cortes que rodean a X.
Paso base. $n = 0$. $XZ >> YZ$ se satisface por la proposición 7.2 parte a.
$n = 1$. Hay dos posibilidades, $\{XZ\}$ y $[ZX(W)]$. Por la proposición 7.2 partes b y c, se tiene que $\{YZ\} >> \{XZ\}$ y $[ZY(W)] >> [ZX(W)]$, por lo que, la proposición se satisface cuando $n = 1$.
$n = 2$. Hay cinco posibilidades, $\{W\{XZ\}\}$, $\{[XZ(W)]\}$, $[Z[X(W)](V)]$, $[W(XZ)]$ y $[W\{XZ\}(V)]$. Por la proposición 7.2 partes d, e, f,..., h, se tiene que, $\{W\{XZ\}\} >> \{W\{YZ\}\}$, $\{[XZ(W)]\} >> \{[YZ(W)]\}$, $[Z[X(W)](V)] >> [Z[Y(W)](V)]$, $[W(XZ)] >> [W(YZ)]$ y $[W\{XZ\}(V)] >> [W\{YZ\}(V)]$, por lo que, la proposición se satisface cuando $n = 2$.
Paso inductivo. Parte a. Como hipótesis inductiva se tiene que, si X está rodeada por $2n$ cortes, entonces $X >> Y$. Por la proposición 7.2 partes $d, e, f, \ldots, h$, se tiene que, $\{W\{XZ\}\} >> \{W\{YZ\}\}$, $\{[XZ(W)]\} >> \{[YZ(W)]\}$, $[Z[X(W)](V)] >> [Z[Y(W)](V)]$, $[W(XZ)] >> [W(YZ)]$ y $[W\{XZ\}(V)] >> [W\{YZ\}(V)]$, en la región rodeada por $2n$ cortes, y estos son los únicos casos por los cuales a X, se le pueden agregar otros dos cortes. Por lo tanto, si X está rodeada por $2n+2$ cortes, es decir, por $2(n+1)$ cortes, entonces $X >> Y$.
Parte b. Como hipótesis inductiva se tiene que, si X está rodeada por $2n+1$ cortes, entonces $Y >> X$. Por la proposición 7.2 partes $d, e, f, \ldots, h$, se tiene que, $\{W\{YZ\}\} >> \{W\{XZ\}\}$, $\{[YZ(W)]\} >> \{[XZ(W)]\}$, $[Z[Y(W)](V)] >> [Z[X(W)](V)]$, $[W(YZ)] >> [W(XZ)]$ y $[W\{YZ\}(V)] >> [W\{XZ\}(V)]$, en la región rodeada por $2n+1$ cortes, y estos son los únicos casos por los cuales a X, se le pueden agregar otros dos cortes. Por lo tanto, si X está rodeada por $2n+1+2$ cortes, es decir, por $2(n+1)+1$ cortes, entonces $Y >> X$.
Por el principio de inducción matemática, se ha probado la proposición. □

Definición 7.1. *(Sistema graficador AlfaK_P al estilo Peirciano) El sistema graficador AlfaK_P, es el sistema AlfaK al estilo Peirciano. En este sistema, se presentan las reglas de inferencia, haciendo referencia la paridad de las regiones en las cuales se aplican (subíndice* 1 *indica que la región es impar, subíndice* 2 *indica que la región es par). En este formato fueron presentados originalmente, los sistemas de gráficos existenciales, por su creador Charles Sanders Peirce, ver Roberts (1992).*

El sistema deductivo AlfaK_P consta del *axioma*. $Ax\lambda.\lambda$. La hoja de aserción donde se dibujan los gráficos existenciales.
El sistema graficador AlfaK_P consta de las siguientes *reglas primitivas de transformación de gráficos (y algunas derivadas)*.

1. Escritura y borrado de gráficos
   Borrado de gráficos. Un gráfico puede ser *borrado* cuando se encuentra en región par.

$$1p.\ \text{En región par, } (XY)_2 \,|\, \Rightarrow X_2$$

   Escritura de gráficos. En una región *impar* puede ser escrito cualquier gráfico.

$$2p.\ \text{En región impar, } X_1 \,|\, \Rightarrow (XY)_1$$

2. Escritura y borrado de enlaces
   Escritura de segundo enlace en un rizo. En un rizo, el segundo *enlace* puede ser *escrito* (ensamblando un bucle), cuando el antecedente es un corte y la región externa es *impar*.

$$3p.\ [(Y)\{X\}_1] \,|\, \Rightarrow [(Y)(X)\_1], \text{ cuando } \{X\} \text{ está en región impar.}$$

   Borrado de segundo enlace de un bucle. En un bucle, el *segundo enlace* que se encuentre región *par* puede ser *borrado* (ensamblando un rizo).

$$4p.\ \text{Regla derivada. } [(Y)(X)\_2] \,|\, \Rightarrow [(Y)\{X\}_2], \text{ cuando el enlace está en región par.}$$

   Borrado y escritura de segundo enlace de bucle. En un bucle, el *segundo enlace* puede ser *borrado* (generando un rizo), cuando el disyunto es fuerte. Y recíprocamente, en un rizo, si el antecedente es un corte fuerte, entonces, el *segundo enlace* puede ser *escrito* (ensamblando un bucle).

$$5p.\ [(Y)(*X)\_] \Leftrightarrow [(Y)\{*X\}], \text{ en cualquier región.}$$

   Borrado y escritura de enlace de rizo. Si en un rizo, el enlace se encuentra en región *impar*, entonces puede ser *borrado*, cuando el consecuente es fuerte.

$$6p.\ \text{Regla derivada. } [Y(*X)\_1] \,|\, \Rightarrow \{Y\{*X\}_1\}, \text{ cuando el enlace está en región impar.}$$

   Escritura de enlace en doble corte encajado. Si en un doble corte encajado, la región externa es *par* y la región interna es fuerte, entonces, el enlace puede ser *escrito* (generando un rizo).

$7p$. Regla derivada. $\{Y\{*X\}_2\} \mid \Rightarrow [Y(*X)\_2]$, cuando $\{*X\}$ está en región par.

Borrado de enlace de rizo en región externa de un rizo. En cualquier región que incluya un rizo, cuyo antecedente es otro rizo, ocurre que, si el enlace del rizo encajado es *par*, entonces puede ser *borrado*.

$8p$. $[X[Z(Y)\_2](Z)] \mid \Rightarrow [X\{Z\{Y\}_2\}(Z)]$, cuando el enlace está en región par.

Borrado y escritura de enlace de rizo con antecedente fuerte. En cualquier región que incluya un rizo, cuyo antecedente es fuerte, puede *escribirse* el segundo enlace, ensamblando un bucle (el nuevo disyunto, es el corte del gráfico que estaba fuertemente afirmado). Y recíprocamente, si en un bucle, un disyunto es un corte, entonces este enlace puede ser *borrado*, generando un rizo (el antecedente es la afirmación fuerte del gráfico que estaba en la región del corte).

$9p$. $[*X(Y)] \Leftrightarrow [\_(\{X\})(Y)]$, en cualquier región.

Escritura de segundo enlace en un rizo. En cualquier región que incluya un rizo, con región de enlace impar, puede *escribirse* el segundo enlace, ensamblando un bucle (el nuevo disyunto es el corte del antecedente).

$10p$. $[X_1(Y)] \mid \Rightarrow [\_1(\{X\})(Y)]$, cuando el enlace está en región impar.

3. Escritura y borrado de rizos
   Borrado y escritura de rizos. En cualquier región, par o impar, puede ser escrito o borrado un rizo.

$11p$. $X \Leftrightarrow [(X)]$, en cualquier región.

4. Escritura y borrado de doble corte
   Borrado del doble corte en región par. En cualquier región *par*, al gráfico $\{\{\{X\}\}\}$, se le puede *borrar* el doble corte.

$12p$. Regla derivada. En región par, $\{\{\{X\}_2\}\} \mid \Rightarrow \{X\}_2$

Escritura del doble corte en región impar. En cualquier región *impar*, al gráfico $\{X\}$, se le puede *escribir* (rodear con) un doble corte.

$13p$. Regla derivada. En región impar, $\{X\}_1 \mid \Rightarrow \{\{\{X\}_1\}\}$

Escritura del doble corte de gráfico fuerte que esté en región par. En un *gráfico fuerte* que se encuentre en región *par*, puede ser *escrito* un *doble corte*.

$14p$. Regla derivada. En región par, $*X2| \Rightarrow \{\{*X_2\}\}$

Borrado y escritura de doble corte en cuádruple corte. En cualquier región, en un *cuádruple corte* puede ser *borrado* un doble corte, además, en un *doble corte* puede ser *escrito* otro doble corte.

$15p$. Regla derivada. $\{\{\{\{X\}\}\}\} \Leftrightarrow \{\{X\}\}$, en cualquier región.

5. Escritura y borrado de afirmación fuerte
   Borrado de la afirmación fuerte. La afirmación fuerte puede ser borrada en cualquier región par.

$16p$. En región par, $*X_2| \Rightarrow X_2$

Escritura de la afirmación fuerte. La afirmación fuerte puede ser escrita en región impar.

$17p$. Regla derivada. En región impar, $X_1| \Rightarrow *X_1$

Borrado y escritura de la afirmación fuerte duplicada. En cualquier región, a una *afirmación fuerte* se le puede *escribir* otra *afirmación fuerte*. En una afirmación *fuerte duplicada*, se puede *borrar* una de ellas.

$18p$. Regla derivada. $*X \Leftrightarrow **X$, en cualquier región.

Borrado y escritura de la afirmación fuerte en región de corte. En cualquier región de corte, una *afirmación fuerte* se puede *borrar*. En una región de corte, se puede *escribir* una *afirmación fuerte*.

19p. $\{*X\} \Leftrightarrow \{X\}$, en cualquier región.

6. Iteración y desiteración de gráficos
   Iteración y desiteración de gráficos en región de solo curvas. Un *gráfico* se puede *iterar* o *desiterar* en cualquier región, par o impar, si la región solo se encuentra rodeada por cero o más curvas (no se encuentra rodeada por algún corte).

$20p$. X en región de curvas Ru, $Y\,(X)^{Ru} \Leftrightarrow Y\,(XY)^{Ru}$

Iteración y desiteración de gráficos fuertes en región de corte.
Un *gráfico fuerte* se puede *iterar* o *desiterar* en cualquier región, par o impar, *aún* si la región se encuentra rodeada por algún corte.

$21p$. $*Y\,(X)^{Rn} \Leftrightarrow *Y\,(*YX)^{Rn}$, en cualquier región Rn.

Iteración y desiteración de gráficos en región de doble corte. Un *gráfico* se puede *iterar* o *desiterar* en cualquier región de doble corte.

$$22p.\ \text{Regla derivada. } \{Y\,\{X\}\} \Leftrightarrow \{Y\,\{YX\}\}, \text{ en cualquier región.}$$

Iteración y desiteración de lazo. En cualquier región, par o impar. Si en un bucle, las regiones de ambos lazos son iguales, se puede desiterar uno de los lazos (generando un rizo con región externa vacía). Y recíprocamente, en un rizo con región externa vacía, se puede iterar el lazo (ensamblándose un bucle con ambos lazos iguales).

$$23p.\ [(X)\,(X)] \Leftrightarrow [(X)], \text{ en cualquier región.}$$

7. Preservación de las transformaciones
   Preservación directa de las transformaciones. Supóngase que un *gráfico inicial* se transforma en uno *terminal*. Entonces, en una región par, el *gráfico inicial* puede ser reemplazado por el *gráfico terminal*.

$$24p.\ X >> Y \text{ implica que, para } X \text{ en región par, } X_2 >> Y$$

   Preservación inversa de las transformaciones. Supóngase que un *gráfico inicial* se transforma en uno*terminal*. Entonces, en una región impar, el *gráfico terminal* puede ser reemplazado por el *gráfico inicial*.

$$25p.\ \text{Regla derivada. } X >> Y \text{ implica que, para } Y \text{ en región impar, } Y_1 >> X$$

8. Afirmación fuerte de teoremas gráficos
   Los teoremas gráficos pueden ser fuertemente afirmados. En cualquier región, los teoremas gráficos, al ser fuertemente afirmados, generan un nuevo teorema.

$$26p.\ Y \in TG \text{ implica } *Y \in TG, \text{ en cualquier región.}$$

9. Reglas implícitas
   Concatenación. Dos gráficos que se encuentren en una misma región, pueden se concatenados. E inversamente, dos gráficos que se encuentren concatenados, pueden ser separados en la misma región.

$$27p.\ X, Y \Leftrightarrow YX, \text{ en cualquier región.}$$

   Conmutatividad. Dos gráficos concatenados pueden ser reescritos cambiando el orden.

$$28p.\ XY \Leftrightarrow YX, \text{ en cualquier región.}$$

Asociatividad. En tres gráficos que se encuentren concatenados, es *irrelevante* el orden en que fueron concatenados. Inicialmente el primero se concatena con el segundo y este resultado se concatena con el tercero, o el primero se concatena con el resultado de concatenar el segundo con el tercero.

$29p.\ XY,Z \Leftrightarrow X, YZ \Leftrightarrow XYZ$, en cualquier región.

Observación 7.1. *Respecto a la regla* $19p$*, en AlfaK,* $\{[X(Y)]\}$ *no implica* $\{[*X(Y)]\}$*, el corte debe ser más interno. Respecto a la regla* $22p$*, en AlfaK,* $Y\{YX\}$ *no implica* $Y\{X\}$*. Respecto a la regla* $24p$*, como caso particular se tiene:* $X >> Y$ *implica* $*X \Rightarrow *Y$.

Proposición 7.4. *(AlfaK_P valida a AlfaK). Para* $X,Y \in GK$

a. Los teoremas de AlfaK son teoremas de AlfaK_P.

b. $X >> Y$ *en AlfaK implica* $X >> Y$ *en AlfaK_P.*

Prueba. Parte a. Supóngase las reglas primitivas, del sistema AlfaK, presentadas en la definición 5.4. A partir de ellas se prueban las reglas primitivas de AlfaK_P, presentadas en la definición 7.1. En la prueba de todas las reglas, para hacer la generalización a regiones pares e impares, se utiliza la proposición 7.3.
Regla $1p$ se sigue de la regla 1. Regla $2p$ se sigue de la regla $1p$ y la proposición 7.1. Regla $3p$ se sigue de la regla 8. Regla $5p$ se sigue de la regla 7. Regla $8p$ se sigue de la regla 6. Regla $9p$ se sigue de la regla 10 (cuya validez está garantizada por Hecho 6.1).
Regla $10p$. Supóngase que, $M\left(X'' \supset Y''\right) = 1$. Supuesto 2, $M\left(Y'' \cup -X''\right) = 0$, por $V\cup$ se sigue que $M\left(Y''\right) = 0$ y $M\left(-X''\right) = 0$, por $V-$ se tiene que, para cada $N \in S$, $M < N$ implica $N\left(X''\right) = 1$, como $M < M$, entonces, $M\left(X''\right) = 1$, y como $M\left(X'' \supset Y''\right) = 1$, por $V \supset$ resulta $M\left(Y''\right) = 1$, lo cual no es el caso, por lo que, $M\left(Y'' \cup -X''\right) = 1$. Por lo tanto, $((X \supset Y) \supset (Y \cup -X))'' \in VK$, es decir, $[X1(Y)] >> [\_{}_1(\{X\})(Y)]$. La regla $10p$ se sigue de este resultado.
Regla $11p$. Se sigue de la regla 11.
Regla $16p$. Por el Hecho $2.4d$ se tiene $+X' \supset X' \in TK$, lo cual por la proposición 4.8, implica que $+X' \supset X' \in VK$, es decir, $+X'$ valida a X', por la proposición $6.8c$, resulta $*X >> X$. La regla $16p$ se sigue de este resultado.
Regla $19p$. Se sigue de la regla 15 y la proposición 7.2.
Regla $20p$. Se sigue de las reglas 17 a 20 y la proposición 7.2.
Regla $21p$. Se sigue de las reglas $20p$, 22 y 23 y la proposición 7.2.
Regla $23p$. Se sigue de la regla 21 y la proposición 7.2. Regla $24p$. Se sigue de la proposición $7.3a$.
Regla $26p$. Se sigue de la regla 26 y la proposición 7.2.
Regla $27p$. Se sigue de la regla 27 y la proposición 7.2. Regla $28p$. Se sigue de la regla 28 y la proposición 7.2. Regla $29p$. Se sigue de la regla 29 y la proposición 7.2.
Parte b. $X >> Y$ en AlfaK, por TDG significa que $[X(Y)]$ es teorema de AlfaK, por la parte a, se sigue que $[X(Y)]$ es teorema de AlfaK_P. Si se supone X, por la regla $20p$ se obtiene $[(Y)]$, aplicando la regla $11p$ se deriva Y, por lo tanto, $X >> Y$ en AlfaK_P. □

Proposición 7.5. *(AlfaK valida a AlfaK_P) Los teoremas de AlfaK_P son teoremas de AlfaK.*

Prueba. Supóngase las reglas primitivas, del sistema AlfaK_P, presentadas en la definición 7.1. A partir de ellas se prueban las reglas primitivas de AlfaK, presentadas en la definición 5.4.
Reglas 1, 2 y 3. Casos particulares de la regla $1p$.
Reglas 4 y 5. Casos particulares de la regla $2p$.
Regla 6. Caso particular de la regla $8p$.
Regla 7. Caso particular de la regla $5p$.
Regla 8. Caso particular de la regla $3p$.
Regla 9. Caso particular de la regla $10p$.
Reglas 11, 12, 13 y 14. Casos particulares de la regla $11p$.
Regla 15. Caso particular de la regla $19p$.
Regla 16. Caso particular de la regla $16p$.
Reglas 17, 18, 18.1, 19 y 20. Casos particulares de la regla $20p$.
Regla 21. Caso particular de la regla $23p$.
Reglas 22 y 23. Casos particulares de la regla $21p$.
Regla 26. Caso particular de la regla $26p$.
Regla 27. Caso particular de la regla $27p$.
Regla 28. Caso particular de la regla $28p$.
Regla 29. Caso particular de la regla $29p$. □

Proposición 7.6. *(AlfaK equivalente a AlfaK_P). Para* $X, Y \in GK$.

a. X es teorema de AlfaK_P si y solo si X es teorema de AlfaK.

b. $X >> Y$ *en AlfaK si y solo si* $X >> Y$ *en AlfaK_P.*

Prueba. Consecuencia directa de las proposiciones 7.4 y 7.5. □

Hecho 7.1. *(Reglas derivadas en AlfaK_P).*
Las reglas $4p$, $6p$, $7p$, $12p$, $13p$, $14p$, $15p$, $17p$, $18p$, $22p$ *y* $25p$ *de AlfaK_P son derivadas en AlfaK y en AlfaK_P.*

Prueba. Por las proposiciones 6.8 y 7.6, basta observar que, las traducciones de estas reglas son válidas en AlfaK. □

8. CONCLUSIONES

En esta sección, se presentan algunas conexiones de AlfaK_P con los gráficos existenciales gama de Peirce y con el sistema Gamma-LD, (Sierra, 2021), también se define un operador de incompatibilidad, como versión

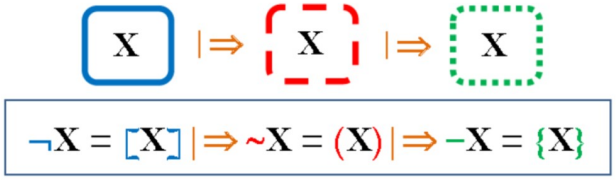


Figura 3: Negaciones paracompleta, clásica y paraconsistente. Fuente: Elaboración propia.

del operador de buen comportamiento en las lógicas paraconsistentes, finalmente, se utiliza AlfaK para dar solu-ción a una versión de la paradoja del mentiroso.

8.1. Conclusión 1

(Respecto al corte paraconsistente). Para $X \in FK$, el corte paraconsistente $\left\{X'\right\}$, es decir $-X$, el cual por el Hecho 2.4*k*, significa que $\otimes \sim X$, y esto semánticamente afirma que es posible la negación clásica de X, lo cual coincide con la interpretación del *corte quebrado* en los gráficos gama de Peirce en Zeman (1963). Además, por el Hecho 2.4*c* se tiene $\sim -(\sim X) \equiv +(\sim X)$, pero en el sistema Gamma-LD presentado en Sierra (2021), se tiene que $+ \sim X$ es la negación intuicionista generalizada, es decir, $+ \sim X \equiv \neg X$, resultando que, $\sim - \sim X \equiv \neg X$, o de manera gráfica, $(\left\{\left(X'\right)\right\}) \Leftrightarrow \left[X'\right]$. De la negación intuicionista también se tiene $+ \sim (\sim X) \equiv \neg(\sim X)$, y entonces por *DN*, $+X \equiv \neg \sim X$, utilizando *Tras*, resulta $\sim +X \equiv \sim \neg \sim X$, por el Hecho 2.4*c* se sigue, $-X \equiv \sim \neg \sim X$, o de manera gráfica, $\left(\left[\left(X'\right)\right]\right) \Leftrightarrow \left\{X'\right\}$, lo cual significa que, en gamma-LD, y por lo tanto en el sistema gama de Peirce, se encuentran las tres negaciones: clásica, paraconsistente y paracompleta (negación intuicionista generalizada), más aún, se tiene que $\left[X'\right] >> \left(X'\right) >> \left\{X'\right\}$, lo cual, con los gráficos tradicionales, puede ser representado como se muestra en la Figura 3.

Para pasar de un corte al otro, simplemente se definen reglas de borrado de cortes:

$$\textit{Borrado inicial o suave}: \left[X'\right] >> \left(X'\right). \quad \textit{Borrado secundario o fuerte}: \left(X'\right) >> \left\{X'\right\}.$$

En Sierra (2021), cuando se restringe el lenguaje de *LD* al lenguaje de *LI*, entonces la restricción asociada a los gráficos Gamma-*LD*, llamada Alfa-*LI*, coincide con los gráficos existenciales intuicionistas presentados en Oostra (2010).

De manera similar, cuando se restringe el lenguaje de *LD* al lenguaje de $KT4P$, entonces la restricción asociada a los gráficos Gamma-*LD*, debe coincidir con los gráficos existenciales paraconsistentes AlfaK.

Definición 8.1. *(Fórmulas fuertemente inconsistentes) Sea $X \in FK$. X es fuertemente inconsistente, significa que se tiene $X \supset -\lambda$, es decir, que se tiene $\sim X$. En caso contrario, X es fuertemente consistente.*

8.2. Conclusión 2

(Respecto a la afirmación fuerte). Para $X \in FK$, por el Hecho 2.4c, se tiene $+X \equiv (-X \supset -\lambda) \equiv\sim -X$, lo cual en AlfaK se representa como $\left[\left\{X'\right\}(\{\_\})\right]$, y significa que $\left\{X'\right\}$, es decir, $-X$ es fuertemente inconsistente. Si se supone $+X$, se sigue $-X \supset -\lambda$, por el Hecho 2.4o, se tiene $-\lambda \supset Z$, resultando por *SH* que $-X \supset Z$, es decir, $\left\{X'\right\} >> Z'$. En resumen, cuando se tiene $+X$, entonces, si se deriva $-X$, el sistema colapsa (se pueden inferir todas las fórmulas), es decir, $-X$ se comporta como si fuera la negación clásica en presencia de contradicciones.

En las lógicas paraconsistentes, por ejemplo, en las primeras lógicas paraconsistentes presentadas por Newton da Costa en da Costa (1993), se utilizan el llamado operador de *buen comportamiento*, con cual simplemente se permite que la negación paraconsistente de una fórmula dada se comporte como la negación clásica, es decir, la negación paraconsistente de la fórmula dada sea fuertemente inconsistente. En el caso de AlfaK, el análogo al buen comportamiento se define a continuación, en términos de la afirmación fuerte.

Definición 8.2. *(Operador incompatibilidad con la negación) Sea $X \in FK$.*
X es incompatible con su negación, denotado X^I o IX, significa que, $X \supset +X$, es decir, $X^I \equiv (X \supset +X)$.
En AlfaK, X'^I se representa como $\left[X'\left(*X'\right)\right]$, es decir, $\left[X'\left(\left[\left\{X'\right\}(\{\_\})\right]\right)\right]$.

Proposición 8.1. *(Caracterización semántica del operador de incompatibilidad con la negación). Sean $X \in FK, (S, Ma, <, V)$ un modelo de $KT4P$ y $M \in S$.*

a. $VI.\ M\left(X^I\right) = 0$ equivale a $M(X) = 1$ y $(\exists P \in S)(M < P\, y\, P(X) = 0)$.

b. $VI.\ M(X^I) = 1$ equivale a, si $M(X) = 1$ entonces $(\forall N \in S)(M < N\, implica\, N(X) = 1)$.

c. $VI.\ M(X^I) = 1$ equivale a, si $(\exists P \in S)\,(M < P\, y\, P(X) = 0)$ entonces $M(X) = 0$.

Prueba. Parte *a*. $M(X^I) = 0$, por definición equivale a $M(X \supset +X) = 0$, por $V \supset$ resulta equivalente a, $M(X) = 1$ y $M(+X) = 0$, por $V+$ equivale a $M(X) = 1$ y $(\exists P \in S)(M < P\, y\, P(X) = 0)$.
Las partes *b* y *c* son consecuencias directas de la parte *a*. □

Proposición 8.2. *(Iteración y desiteración de gráficos incompatibles con la negación). Sean $X, Y \in FK$.*
Si un gráfico es incompatible con su negación, se puede iterar o desiterar en cualquier región, par o impar.

$$30p.\ Si\ Y^I\ entonces\ Y\,(X)^{Rn} \Leftrightarrow Y\,(YX)^{Rn},\ en\ cualquier\ región\ Rn.$$

Prueba. Considerando la regla $20p$, las proposiciones 7.2 a 7.4, y que solo interesa la iteración en región par y la desiteración en impar (ya que iteración en impar y la desiteración en par, son realmente las reglas de escritura $2p$ y borrado $1p$), entonces para probar $30p$, es suficiente con garantizar que, las siguientes fórmulas son consecuencias de Y^I:

a. $(Y \bullet - - X) \supset (Y \bullet - - (X \bullet Y))$.

b. $(Y \bullet (-X \supset Z)) \supset (Y \bullet (-(X \bullet Y) \supset Z))$.

c. $(Y \bullet -(X \supset Z)) \supset (Y \bullet -((X \bullet Y) \supset Z))$.

Parte *a*. Supóngase que $M\left(Y^I\right) = 1$ y $M(Y \bullet --X) = 1$, por $V\bullet$ se tiene $M(Y) = 1$ y $M(--X) = 1$, por $V-$ resulta que existe $P \in S$, $M < P$ y $P(-X) = 0$. Supuesto 2, $M(--(X \bullet Y)) = 0$, por $V-$ y $M < P$, resulta que $P(-(X \bullet Y)) = 1$, y entonces existe $Q \in S$, $P < Q$ y $Q(X \bullet Y) = 0$, y como $M\left(Y^I\right) = M(Y) = 1$, por la proposición 8.1*b* se infiere que, para todo $N \in S$, $M < N$ implica $N(Y) = 1$, y como $M < P$ y $P < Q$, por la restricción RT, se sigue $M < P$, y así $Q(Y) = 1$, pero $Q(X \bullet Y) = 0$, aplicando $V\bullet$ se deriva $Q(X) = 0$, pero $P(-X) = 0$ y $P < Q$, aplicando $V-$ se infiere $Q(X) = 1$, lo cual no es el caso, por lo que, $M(--(X \bullet Y)) = 1$, y como $M(Y) = 1$, por $V\bullet$ se concluye que $M(Y \bullet --(X \bullet Y)) = 1$.
Parte *b*. Supóngase que $M\left(Y^I\right) = M(Y \bullet (-X \supset Z)) = 1$, por $V\bullet$ se tiene $M(Y) = 1$ y $M(-X \supset Z) = 1$. Supuesto 2, $M(-(X \bullet Y) \supset Z) = 0$, según $V \supset$ se obtiene $M(-(X \bullet Y)) = 1$ y $M(Z) = 0$, aplicando $V-$, existe $P \in S$, $M < P$ y $P(X \bullet Y) = 0$, como $M\left(Y^I\right) = M(Y) = 1$ y $M < P$, por la proposición 8.1*b* se infiere que, $P(Y) = 1$, pero $P(X \bullet Y) = 0$, por $V\bullet$ se deriva que $P(X) = 0$, como $M(-X \supset Z) = 1$ y $M(Z) = 0$, por $V \supset$ resulta $M(-X) = 0$, por $V-$ y $M < P$, se deriva $P(X) = 1$, lo cual no es el caso, por lo que, $M(-(X \bullet Y) \supset Z) = 1$, y como $M(Y) = 1$, por $V\bullet$ se concluye que $M(Y \bullet (-(X \bullet Y) \supset Z)) = 1$.
Parte *c*. Supóngase que $M\left(Y^I\right) = M(Y \bullet -(X \supset Z)) = 1$, por $V\bullet$ se tiene $M(Y) = 1$ y $M(-(X \supset Z)) = 1$, aplicando $V-$, existe $P \in S$, $M < P$ y $P(X \supset Z) = 0$, por $V \supset$ se obtienen $P(X) = 1$ y $P(Z) = 0$. Supuesto 2, $M(-((X \bullet Y) \supset Z)) = 0$, por $V-$ y $M < P$ se sigue que, $P((X \bullet Y) \supset Z) = 1$, y como $P(Z) = 0$, por $V \supset$ se infiere que $P(X \bullet Y) = 0$, pero $P(X) = 1$, utilizando $V\bullet$ se sigue $P(Y) = 0$, y como $M\left(Y^I\right) = 1$ y $M < P$, por la proposición 8.1*c* se infiere que, $M(Y) = 0$, lo cual no es el caso, por lo que, $M(-((X \bullet Y) \supset Z)) = 1$, y como $M(Y) = 1$, por $V\bullet$ se concluye que $M(Y \bullet -((X \bullet Y) \supset Z)) = 1$. □

Proposición 8.3. *(Relación entre afirmación fuerte e incompatibilidad con la negación). Sean* $X \in FK$. $+X| \Rightarrow X^I$.

Prueba. Supóngase que $+X$, por $Ax1$ y Mp se infiere $X \supset +X$, es decir, X^I. La recíproca, se refuta con el modelo $(\{Ma,P\},Ma,<,V)$, donde $Ma < Ma$, $Ma < P$, $P < P$, $Ma\left(X^I\right) = 1$ y $Ma(+X) = Ma(X) = P(X) = 0$. □

8.3. Conclusión 3

(Respecto a las aplicaciones). Muchas *paradojas lógicas* involucran los conceptos de verdad o falsedad, por ejemplo, la siguiente variante de la paradoja del mentiroso Bochenski (1976) y Smullyan (1997). Considérese la situación en la cual se tiene una oración que se muestra en la Figura 4

1. Cuando se identifican en la *lógica clásica*, ser el caso con verdadero (Z) y no ser el caso con falso $(\sim Z)$, entonces se tiene la paradoja: si Z (la oración es el caso), como se tiene que $Z \equiv \sim Z$ (la oración dice que es falsa), entonces resulta que $\sim Z$ (también es falsa), y si $\sim Z$ (la oración es falsa), como

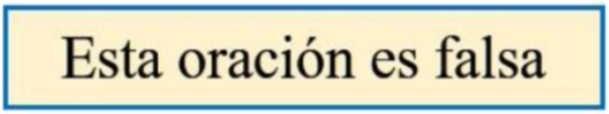


Figura 4: Variante de la paradoja del mentiroso, Smulllyan (1997).

se tiene que $Z \equiv \sim Z$ (la oración dice que es falsa), entonces resulta que Z (es el caso). Es decir, $(Z \supset \sim Z) \bullet (\sim Z \supset Z)$, por *Imp* y *DN* se deriva $(\sim Z \cup \sim Z) \bullet (Z \cup Z)$, resultando $Z \bullet \sim Z$. Como la lógica clásica no soporta las contradicciones, resulta inútil en este caso.

2. Cuando se identifican en la *lógica intuicionista*, ser el caso como verdadero (Z), y no ser el caso con falso $(\neg Z)$, entonces se tiene la paradoja: Si se supone que Z (la oración es el caso), como se tiene que $Z \leftrightarrow \neg Z$ (la oración dice que es falsa), entonces se deriva $\neg Z$ (la oración es falsa), por lo que, $Z \rightarrow \neg Z$. Si se supone que $\neg Z$ (la oración es falsa), y como se sabe que $Z \leftrightarrow \neg Z$ (la oración dice que es falsa), se infiere Z (Z es el caso), por lo que $\neg Z \rightarrow Z$. Se ha probado que, $(Z \rightarrow \neg Z) \wedge (\neg Z \rightarrow Z)$.

De van Dalen (2013), se sabe que la lógica intuicionista está caracterizada por una semántica de mundos posibles (similar a la semántica de $KT4P$), en la cual, se tiene la regla $V\neg : V(M, \neg X) = 1$ si y solo si para todo $N \in S$, $M < N$ implica que $V(N,X) = 0$, la regla $V \rightarrow$: $V(M, X \rightarrow Y) = 1$ si y solo si para todo $N \in S$, $M < N$ implica que $V(N, X \supset Y) = 1$, y además la relación de accesibilidad es reflexiva y transitiva.
Supóngase que $V(M, Z \rightarrow \neg Z) = 1$, $V(M, \neg Z \rightarrow Z) = 1$, pero $V(M,Z) = 0$, aplicando $V \rightarrow$ y reflexividad, resulta que $V(M, \neg Z) = 0$, por $V\neg$, se deriva que existe $P \in S$, $M < P$ y $V(P,Z) = 1$, y como se tiene $V(M, Z \rightarrow \neg Z) = 1$, utilizando $V \rightarrow$ y $M < P$ se obtiene que $V(P, \neg Z) = 1$, lo cual por $V\neg$ y reflexividad significa que $V(P,Z) = 0$, lo cual no es el caso, por lo tanto, $V(M, Z \rightarrow \neg Z) = V(M, \neg Z \rightarrow Z) = 1$ implica $V(M,Z) = 1$.
Supóngase que $V(M, Z \rightarrow \neg Z) = 1, V(M, \neg Z \rightarrow Z) = 1$, pero $V(M, \neg Z) = 0$, por $V\neg$ se sigue que existe $P \in S, M < P$ y $V(P,Z) = 1$, como $V(M, Z \rightarrow \neg Z) = 1$, por $V \rightarrow$ y $M < P$ se deduce $V(P, \neg Z) = 1$, lo cual por $V\neg$ y reflexividad significa $V(P,Z) = 0$, lo cual no es el caso, por lo tanto, $V(M, Z \rightarrow \neg Z) = V(M, \neg Z \rightarrow Z) = 1$ implica $V(M, \neg Z) = 1$.
Se tiene entonces que, de $(Z \rightarrow \neg Z) \wedge (\neg Z \rightarrow Z)$, se deriva la contradicción $Z \wedge \neg Z$, pero la lógica intuicionista no es paraconsistente (no soporta las contradicciones, ya que tiene como teorema $Z \rightarrow (\neg Z \rightarrow X)$) Heyting (1956). Por lo tanto, la lógica intuicionista, resulta inútil en el caso de esta paradoja. □

Hecho 8.1. *(Las fórmulas fuertemente inconsistentes no tienen modelos) Para $X \in FK$.*
Si existe un modelo $(S, M, <, V)$ de $KT4P$ y $V(M,X) = 1$ entonces X es fuertemente consistente en $KT4P$.

Prueba. Sea X fuertemente inconsistente en $KT4P$, por lo que $X \supset -\lambda \in FK$, por la proposición 4.8*a* se infiere $X \supset -\lambda \in VK$, es decir, para todo modelo M, $V(M, X \supset -\lambda) = 1$. Si hay un modelo tal que $V(M,X) = 1$, entonces por $V \supset$ se sigue que $V(M, -\lambda) = 1$, lo cual por $V-$ implica que existe $P \in S$, $M < P$ y $V(P, \lambda) = 0$, lo cual contradice $V\lambda$. Por lo tanto, no existe tal modelo. Se ha probado que, si X es fuertemente inconsistente en $KT4P$ entonces no existe un modelo $(S, M, <, V)$ de $KT4P$ tal que

$V(M,X) = 1$, es decir, si existe un modelo $(S,M,<,V)$ de $KT4P$, tal que $V(M,X) = 1$, entonces X es fuertemente consistente en $KT4P$. □

Proposición 8.4. *(Solución a la variante de la paradoja del mentiroso) En el sistema KT4P, representando lo que dice la oración es el caso como Z, la oración Z es verdadera como +Z, y la oración Z es falsa, es decir, Z no es verdadera (o Z puede ser refutada) como −Z, entonces la situación (se tiene una oración que dice: esta oración es falsa) queda representada por la fórmula $Z \equiv -Z$, la cual en KT4P no es fuertemente inconsistente y se pueden obtener conclusiones válidas.*

Prueba en AlfaK_P: Se tiene $Z \equiv -Z$ en $KT4P$, por lo que, en AlfaK_P se tienen $\left[Z'\left(\left\{Z'\right\}\right)\right]$ y $\left[\left\{Z'\right\}\left(Z'\right)\right]$. Supóngase que $*Z'$, por la regla de borrado $16p$ se obtiene Z', aplicando la regla de desiteración $20p$, en la primera premisa, se infiere $\left[\_\left(\left\{Z'\right\}\right)\right]$, utilizando la regla de borrado de rizo $11p$, se deduce $\left\{Z'\right\}$, por la regla de escritura en región de corte $19p$, se deriva $\left\{*Z'\right\}$, como se tiene $*Z'$, utilizando la regla de desiteración $21p$, se sigue $\{\_\}$, aplicando la proposición $5.1d$, demostración indirecta gráfica, se concluye, $\left\{*Z'\right\}$, por la regla de borrado en región de corte $19p$, se deriva $\left\{Z'\right\}$, aplicando la regla de desiteración $20p$, en la segunda premisa, se infiere $\left[\left(Z'\right)\right]$, utilizando la regla de borrado de rizo $11p$, se deduce Z', finalmente, por la regla de concatenación $27p$, se ha probado $Z'\left\{Z'\right\}$, es decir, $Z \bullet -Z$, por lo que, Z es el caso, pero es falsa y no es verdadera, ya que es refutable.

De acuerdo con el Hecho 8.1, para garantizar que no hay paradoja, es decir que $Z \equiv -Z$ no es fuertemente inconsistente, basta verificar que $Z \equiv -Z$ tiene un modelo. Considere el modelo $(\{MA,M1\},MA,<,V)$ tal que, $MA < MA$, $M1 < M1$, $MA < M1$, $V(MA,+Z) = V(M1,Z) = V(M1,+Z) = 0$ y $V(MA,Z) = V(MA,-Z) = V(M1,-Z) = 1$. Por lo tanto, $V(MA,Z \equiv -Z) = 1$. Es decir, $Z \equiv -Z$ tiene un modelo, y por lo tanto es fuertemente consistente en $KT4P$ y no hay paradoja. □

Comentario. En Sierra (2021), se presenta otra solución a la variante de la paradoja del mentiroso, presentada más arriba, utilizando el sistema de gráficos existenciales Gamma-LD. Los resultados obtenidos contrastan fuertemente: en Gamma-LD se infiere que Z no es el caso, Z no es falsa y Z no es verdadera, mientras que en AlfaK se infiere que Z es el caso, Z es falsa y Z no es verdadera. Esto se debe al contraste de los enfoques de las lógicas paracompletas versus las lógicas paraconsistentes, por ejemplo, en AlfaK se tiene $\left[\left(Y'\right)\left\{X'\right\}\right] | \Rightarrow \left[\left(Y'\right)\left(X'\right)\right]$, es decir, $-X \supset Y | \Rightarrow X \cup Y$, mientras que en Alfa intuicionista se tiene $\left[\left(Y'\right)\left(X'\right)\right] | \Rightarrow \left[\left(Y'\right)\left[X'\right]\right]$, es decir, $X \vee Y | \Rightarrow \neg X \rightarrow Y$.

Agradecimientos

El autor agradece a los profesores Arnold Oostra Van Noppen de la Universidad del Tolima, Fernando Zalamea Traba de la Universidad Nacional de Colombia y Yuri Poveda Quiñones de la Universidad Tecnológica de Pereira, por sus inspiradoras publicaciones sobre los gráficos existenciales de Peirce; a la profesora María Elena Álvarez Ospina de la Universidad EAFIT y a la Ingeniera Gloria Rúa Marín egresada

de la Universidad EAFIT, por sus inspiradoras motivaciones sobre gráficos existenciales paraconsistentes y paracompletos.

Referencias

Bochenski, I. (1976). Historia de la lógica formal. Ed. Gredos. Madrid.

Brady, G. & Trimble, T. (2000). A categorical interpretation of C.S. Peirce's propositional logic Alpha. *Journal of Pure and Applied Algebra*. 149, 213-239.

da Costa, N. (1993). Sistemas formais inconsistentes. Ed. UFPR. Universidade Federal Do Paraná. Curitiba. Brasil.

Doop, J. (1969). Nociones de lógica formal. Ed. Tecnos S.A. Madrid.

Hamilton, A. (1978). Logic for mathematicians. Cambridge University Press. Cambridge.

Henkin, L. (1949). The completeness of the first order functional calculus. *The Journal of Symbolic Logic*, 14(**3**), 159 –166.

Heyting, A. (1956). Intuitionism: An introduction (Studies in logic and the foundations of mathematics). North-holland publishing company.

Oostra, A. (2010). Los gráficos Alfa de Peirce aplicados a la lógica intuicionista. *Cuadernos de Sistemática Peirceana*, 2, 25-60.

Oostra, A. (2011). Gráficos existenciales beta intuicionistas. *Cuadernos de Sistemática Peirceana*, 3, 53-78.

Oostra, A. (2012). Los gráficos existenciales gama aplicados a algunas lógicas modales intuicionistas. *Cuadernos de Sistemática Peirceana*, 4, 27-50.

Oostra A. (2021). Equivalence Proof for Intuitionistic Existential Alpha Graphs. Diagrammatic Representation and Inference. Diagrams 2021. *Lecture Notes in Computer Science*, 12909. Springer, Cham.

Peirce, C. (1965). Charles S. Peirce, Collected Papers of Charles Sanders Peirce. Charles Hartshorne, Paul Weiss and Arthur W. Burks (Eds.). Cambridge (Massachusetts): Harvard University Press (1931–1958).

Roberts, D. (1992). The existential graphs in Computters Math. *Applie*, 23, 6-9, 639-663.

Sierra, M. (2010). Argumentación deductiva con diagramas y árboles de forzamiento. Fondo Editorial Universidad EAFIT. Medellín.

Sierra, M. (2021). Lógica doble LD y gráficos existenciales gamma LD. *Revista Facultad de Ciencias Básicas*, 17, 2.

Smullyan, R. (1997). Como se llama este libro. Ed. Cátedra. Madrid.

van Dalen, D. (2013). Logic and structure. Springer-Verlag. London.

Zalamea, F. (2010). Los gráficos existenciales Peirceanos. Universidad Nacional de Colombia. Bogotá.

Zeman, J. (1963). The Graphical Logic of C.S. Peirce, Ph.D. Thesis, University of Chicago.

Revista de la Facultad de Ciencias Universidad Nacional de Colombia, Sede Medellín
V 11 N°2 julio-diciembre de 2022 • ISSN-e 2357-5749 • Artículo Investigación • Páginas 148 to 161
DOI: https://doi.org/10.15446/rev.fac.cienc.v11n2.99977

UN MÉTODO FORMAL PARA LA ARMONIZACIÓN CONCEPTUAL DEL EQUILIBRIO QUÍMICO[a]

A FORMAL METHOD FOR THE CONCEPTUAL HARMONIZATION OF CHEMICAL EQUILIBRIUM

DIEGO FERNANDO PATIÑO-SIERRA[b], DANIEL BARRAGÁN[bc*]

Recibido 7-12-2021, aceptado 6-05-2022, versión final 27-05-2022
Artículo Investigación

RESUMEN: La conceptualización es el principal reto que se debe afrontar al enseñar el equilibrio químico. En la enseñanza de los cursos de química general suele predominar la aproximación cinética al concepto de equilibrio químico, mientras que la fundamentación energética de este concepto se deja para cursos posteriores. En este trabajo se presenta el formalismo que integra la cinética y la termodinámica y se aplica al estudio del modelo de reacción $A \rightleftharpoons B$. Los resultados obtenidos permiten explicar con claridad los principales aspectos cinéticos y energéticos que caracterizan el equilibrio químico. Este trabajo se desarrolló con el propósito de contribuir a fortalecer en el docente la apropiación conceptual del equilibrio químico, y de ofrecerle herramientas que le den autonomía para la elaboración de material didáctico novedoso para llevar al aula de clase.
PALABRAS CLAVE: Avance de reacción; equilibrio dinámico; equilibrio químico; función energía de Gibbs.

ABSTRACT: Conceptualization is the main challenge to be faced when teaching chemical equilibrium. In the teaching of general chemistry courses, the kinetic approach to chemical equilibrium usually predominates, while the energetic foundation of this concept is left for later courses. In this paper, we present the formalism that integrates kinetics and thermodynamics, applying it to the study of the $A \rightleftharpoons B$ reaction model. The results obtained clearly explain the main kinetic and energetic aspects that characterize chemical equilibrium. This work is developed to strengthen the teacher's conceptual appropriation of chemical equilibrium and offer tools that give autonomy for the elaboration of novel didactic material to take to the classroom.
PALABRAS CLAVE: Chemical equilibrium; dynamic equilibrium; extent of reaction; Gibbs energy function.

1. INTRODUCCIÓN

En el aula de clase se pueden confrontar el lenguaje cotidiano y los sucesos de la cotidianidad con los conceptos y fenómenos propios de la química. La acidez de un suelo, la corrosión de un metal, la oxidación de

[a]Patiño-Sierra, D. F. & Barragán, D. (2022). Un método formal para la armonización conceptual del equilibrio químico. *Rev. Fac. Cienc.*, 11 (**2**), 148–161. DOI: https://doi.org/10.15446/rev.fac.cienc.v11n2.99977

[b]Semillero de Investigación Filosofía y Educación de la Ciencia, Facultad de Ciencias, Universidad Nacional de Colombia Sede Medellín, Medellín, Colombia.

[c]Escuela de Química, Facultad de Ciencias, Universidad Nacional de Colombia Sede Medellín, Medellín, Colombia.

[*]Corresponding author: dalbarraganr@unal.edu.co

una fruta, la acidificación de lagos y mares, y hasta el origen y propiedades de las gemas, son sucesos, o fenómenos, de nuestro entorno para los cuales culturalmente hemos desarrollado un lenguaje social propio. Sin embargo, a estos sucesos de la cotidianidad está ligado un lenguaje, unos conceptos, y un formalismo propio de la ciencia. El ejercicio docente demanda establecer puentes de comunicación coherentes, para que esta trasposición entre los saberes sociales y el conocimiento escolar no termine en abstracciones simplistas que lleven a trivializar o a complicar el aprendizaje de las ciencias (Bradley & Steenberg, 2006; Hoppe, 1980).

El lenguaje de la química está codificado en una serie de complejas representaciones simbólicas que caracterizan la materia desde la escala atómica hasta la macroscópica. El fin del lenguaje simbólico de la química es expresar las conclusiones, los conceptos, que se obtienen como consecuencia de la interacción entre evidencias experimentales, leyes fundamentales de la naturaleza, teorías científicas y modelos. Es así como, mediante reglas y convenciones los símbolos representan las sustancias, la estructura y la geometría molecular, la reactividad y los mecanismos de reacción (Taber, 2009). La enseñanza de la química no se puede reducir a dar un tratamiento superficial al lenguaje simbólico de la química, cuando esto sucede el estudiante percibe la química como un saber abstracto, carente de lógica y de aplicabilidad. (Raviolo, 2006).

En el estudio del equilibrio químico, acompañan al lenguaje químico la reactividad, la cinética, la termodinámica, las analogías, la simulación, la contextualización y los modelos matemáticos. Todo lo anterior permite destacar al equilibrio químico como un concepto integrador y unificador de la química. Dado lo anterior no sorprende que el equilibrio químico sea considerado uno de los temas de mayor complejidad, tanto para la enseñanza, como para el aprendizaje. La enseñanza tradicional, la soportada únicamente en el lenguaje, hace que el nivel de abstracción sea alto al momento de estudiar el equilibrio químico. Sin embargo, el uso de animaciones y de simuladores, al igual que la experimentación bien planeada, son estrategias que el docente debe considerar al momento de enseñar el equilibrio químico (Rogers *et al.*, 2000).

Dado que hay una estrecha relación entre el dominio conceptual del docente y la argumentación científica del estudiante, (Kaya, 2013), es necesario que el docente disponga de material, más allá de lo disponible en libros de texto y la internet, que le permita identificar y clarificar las posibles concepciones erróneas que tenga sobre determinada temática. La complejidad asociada a la enseñanza y el aprendizaje del equilibrio químico ha sido estudiada, con el propósito de hacer propuestas para la enseñanza, desde las concepciones erróneas de los estudiantes al argumentar sobre problemas específicos (Driel & Gräber, 2002; Davenport *et al.*, 2014). Por ejemplo, la respuesta, o el cambio del equilibrio, ante una perturbación, ya sea por adición de un gas inerte, cambio en el volumen de reacción o adición de reactivos, demanda en el estudiante una concepción más cuantitativa que cualitativa del equilibrio químico. Los estudiantes fallan al analizar la respuesta del equilibrio químico a una perturbación cuando su análisis se basa solo en un principio de acción-reacción, en lugar de considerar la magnitud del cociente de reacción frente a la constante de equilibrio (Tyson *et al.*, 1999). Los docentes que adoptan el llamado principio de Le Châtelier como concepto central para la

enseñanza del equilibrio químico tienen concepciones erradas sobre la argumentación de los aspectos cinéticos y termodinámicos. Esas concepciones erradas tienen principalmente origen en la aproximación simple de analizar el carácter dinámico del equilibrio a partir de la respuesta que busca amortiguar una perturbación (Cheung, 2009).

Los esfuerzos realizados para contribuir a contrarrestar la complejidad asociada a la enseñanza-aprendizaje del equilibrio químico, y las concepciones erróneas, consideran el uso de analogías, escenarios de contexto, simulaciones computacionales, actividades lúdicas, experimentación y hasta elaboradas secuencias didácticas. Una propuesta de secuencia didáctica para el estudio del equilibrio químico considera las siguientes etapas de análisis: de una reacción química que no alcanza el equilibrio (incompleta), de reacciones químicas opuestas (cinéticamente reversibles), de reacciones en equilibrio químico dinámico (incremento del rendimiento de reacción), el avance de una reacción desde un estado estacionario al equilibrio, el avance de una reacción desde un estado de equilibrio a otro, y el significado de la magnitud de la constante de equilibrio (Ghirardi *et al.*, 2014).

En este trabajo se presenta el formalismo en el que el docente se puede apoyar para que, a través del desarrollo de un material propio, dé un manejo práctico, ilustrativo y armonizado al concepto de equilibrio químico. La innovación docente llevada al aula de clase no versa necesariamente sobre la creatividad para enseñar a partir de lo que se tiene disponible, sino para generar nuevo material académico que lleve la enseñanza, el aprendizaje, y por tanto la educación, hacia nuevos escenarios para crear, transmitir y apropiar conocimiento.

2. METODOLOGÍA

Para el modelo de reacción $A \rightleftharpoons B$, el cual consideramos tiene lugar en sistema cerrado a temperatura y presión constantes, describimos el cambio en la composición (de la concentración de las sustancias químicas) y en la energía química de la reacción en función de la variable independiente progreso o avance de reacción. El modelo matemático que se obtiene tiene solución analítica, lo cual facilita la simulación utilizando software libre de fácil acceso, tal como WolframAlpha, o una hoja de cálculo.

2.1. El modelo $A \rightleftharpoons B$

El modelo de ecuación química $A \rightleftharpoons B$ se utiliza para describir procesos unimoleculares, es decir procesos fisicoquímicos en los que participa solo una entidad molecular o sustancia química. Estos procesos fisicoquímicos pueden ser reacciones químicas, cambios de fase o transporte de sustancias.

Una reacción química unimolecular se puede entender a partir de la redistribución interna de electrones y/o átomos, o como consecuencia de cambios conformacionales. Son ejemplos de reacciones unimoleculares la transformación del ciclopropano en propeno y la conversión del cis-2-buteno en trans-2-buteno. En estas

reacciones unimoleculares se acepta que la velocidad de reacción es proporcional a la concentración, lo que permite describir la cinética mediante una ley de velocidad de primer orden (Gillespie *et al.*, 1990; Gillespie *et al.*, 1994).

Las reacciones unimoleculares, de la forma $A \rightleftharpoons B$, y con ley de velocidad de primer orden, son un sistema conveniente para el estudio de la estequiometría, la cinética y la termodinámica de las reacciones químicas en el aula de clase, ya que las ecuaciones matemáticas y sus soluciones facilitan el ejercicio pedagógico y didáctico por parte del docente. Sin embargo, hay que tener presente que el estudio experimental de las reacciones unimoleculares puede llevar a modelos cinéticos y termodinámicos más complejos que el correspondiente a una ley de velocidad de orden uno, sobre todo si estas se llevan a cabo en fase gaseosa (Martínez-Nuñez, 2002; Troe, 2012).

El modelo de ecuación $A \rightleftharpoons B$ también es útil para describir el equilibrio de fases de las sustancias químicas, por ejemplo, el equilibrio líquido-vapor. Una botella cerrada y con agua hasta la mitad de su volumen, y a temperatura y presión constantes, contiene agua en fase líquida y agua en fase vapor. La cantidad de agua en cada una de las fases depende del volumen del recipiente, de la temperatura y de la presión. Manteniendo el volumen del recipiente constante, si cambia la temperatura, cambia la cantidad de cada una de las fases. Ahora, si el volumen y la temperatura son constantes, la adición de agua líquida, o la extracción de vapor de agua de la botella, llevaran a cambios en las cantidades que alcanzan el equilibrio. El modelo $A \rightleftharpoons B$ permite analizar cualitativamente el equilibrio líquido-vapor, al igual que el equilibrio sólido-líquido (hielo-agua), sin embargo, el modelamiento matemático no es tan sencillo como asumir una ley de velocidad de primer orden (Atkins & De Paula, 2014).

Se considera finalmente, en esta sección, procesos de transferencia de masa, de cantidad de sustancia, entre dos fases. La extracción líquido-líquido de una sustancia, como puede ser la del ácido benzoico disuelto en benceno que se extrae con disolución acuosa de hidróxido de sodio; la distribución o reparto de yodo molecular entre aceite y agua; la distribución de un fármaco entre un tejido y el plasma sanguíneo. Si se analiza los ejemplos anteriores con el modelo de ecuación química $A \rightleftharpoons B$, es mejor expresarlo de la siguiente manera: $\alpha_A \rightleftharpoons \beta_A$. Con este modelo de ecuación se describe la transferencia de la sustancia A, entre las fases α y β, proceso para el que es válido asumir una ley de velocidad de primer orden (Turner, 1994; Kujawski *et al.*, 2012; Harris & Logan, 2014).

El objeto conceptual de estudio en este trabajo es el equilibrio químico y todos los análisis que se hagan al respecto tienen validez para cualquiera de los procesos fisicoquímicos, con o sin reacción química, que se han mencionado en esta sección.

2.2. Formalismo

Se considera una reacción unimolecular que tiene lugar en un volumen de 1 litro, a temperatura y presión constantes. El volumen de 1 litro es conveniente ya que la concentración molar de las especies químicas (C) y la cantidad de sustancia (N), de cada una de ellas, son cantidades iguales. Sea r la variable que describe el progreso, o avance de la reacción, en términos de N o de C (Petrucci *et al.*, 2017; Novak, 2019; Honig, 2020).

La reacción unimolecular que se considera se describe mediante la ecuación química (1):

$$A \rightleftharpoons B \tag{1}$$

A medida que la reacción ocurre, la disminución en la concentración de los reactivos o el aumento en la concentración de los productos, divido por el coeficiente estequiométrico respectivo de cada especie, determina el cambio en el avance de la reacción. Así, de la ecuación (1) se obtiene la ecuación (2):

$$-dC_A = dC_B = dr \tag{2}$$

Después de integrar la ecuación (2), $-\int_{C_{A,0}}^{C_A} dC_A = \int_{C_{B,0}}^{C_B} dC_B = \int_0^r dr$, se obtienen las ecuaciones (3) y (4), que expresan la composición del sistema en función del avance de reacción:

$$C_A = C_{A,0} - r \tag{3}$$

$$C_B = C_{B,0} + r \tag{4}$$

El cociente de reacción, Q_r, es una cantidad útil en el estudio del equilibrio químico ya que expresa la proporción relativa entre productos y reactivos en cualquier instante de la reacción. A partir de las ecuaciones (3) y (4), el cociente de reacción se escribe según la ecuación (5):

$$Q_r = \frac{C_{B,0} + r}{C_{A,0} - r} \tag{5}$$

En el estado de equilibrio químico el cociente de reacción es igual a la constante analítica de equilibrio (la que se expresa en función de concentraciones y no de actividades), $Q_r = K$.

Se considera ahora la energía libre por mol de cada sustancia (G), la cual se conoce como el potencial químico (μ). En disoluciones ideales o muy diluidas, el potencial químico de las sustancias químicas A y B, se expresa mediante las ecuaciones (6) y (7):

$$\mu_A = \bar{G}_A = \mu_A^0 + RT \ln C_A \tag{6}$$

$$\mu_B = \bar{G}_B = \mu_B^0 + RT \ln C_B \tag{7}$$

La energía total de la reacción, la expresada mediante la función de Gibbs, depende de la composición según la ecuación (8):

$$G_r = N_A \mu_A + N_B \mu_B \tag{8}$$

Después de reemplazar las ecuaciones (3), (4), (6) y (7) en (8) se obtiene la ecuación (9) que expresa la energía de la reacción en función del avance:

$$G_r = \left((C_{A,0}-r)\,\mu_A^0 + (C_{B,0}+r)\,\mu_B^0\right) + RT\left((C_{A,0}-r)\ln(C_{A,0}-r)\right) + \left((C_{B,0}+r)\ln(C_{B,0}+r)\right) \tag{9}$$

En la ecuación (9) se hace uso de las siguientes consideraciones: $V = 1L$; $C_A = N_A$; $C_B = N_B$.

En este momento ya se tiene la composición de la reacción, el cociente de reacción y la energía de reacción en función del avance de reacción. Ahora se necesita expresar el avance de reacción en función de las concentraciones iniciales y de parámetros cinéticos. Sea ν_1 la velocidad de reacción de $A \rightarrow B$, con constante cinética k_1, y ν_{-1} la velocidad de reacción de $B \rightarrow A$, con constante cinética k_{-1}, a partir de leyes de velocidad de primer orden, la velocidad total de reacción, que es igual a la variación temporal del avance de reacción, se escribe mediante la ecuación (10):

$$\frac{dr}{dt} = \nu_{total} = \nu_1 - \nu_{-1} = k_1\,(C_{A,0}-r) - k_{-1}\,(C_{B,0}+r) \tag{10}$$

Para integrar la ecuación (10), se agrupan los términos para obtener la ecuación :

$$\int_0^r \frac{dr}{k_1\,(C_{A,0}-r) - k_{-1}\,(C_{B,0}+r)} = \int_0^t dt \tag{11}$$

La ecuación (11) tiene solución analítica exacta, de modo que el avance de reacción r cambia en el tiempo según los valores de las concentraciones iniciales y de las constantes cinéticas, como se muestra en la ecuación (12):

$$r = \frac{(k_1 C_{A,0} - k_{-1} C_{B,0})\left(e^{-(k_1+k_{-1})t} - 1\right)}{-(k_1+k_{-1})} \tag{12}$$

Revisemos lo que se tiene con el formalismo anterior: la ecuación (12) se resuelve en un intervalo de tiempo, para unos valores dados de las concentraciones iniciales y de las constantes cinéticas, y como se está estudiando el equilibrio químico, se resuelve hasta obtener un valor constante del avance de reacción. Los valores obtenidos para r se reemplazan en todas las ecuaciones anteriores, y se obtiene así la variación en el tiempo de las concentraciones, el cociente de reacción, los potenciales químicos y la energía del sistema. Con todos estos resultados hay que implementar estrategias que contribuyan a armonizar la conceptualización del equilibrio químico.

2.3. El equilibrio químico

En los libros de texto de química general y termodinámica se encuentran diversas definiciones del equilibrio químico (Petrucci *et al.*, 2017; Brown *et al.*, 2004), por ahora solo se tomará de referencia la dada por la IUPAC (Chalk *et al.*, 2019), la cual es: "*procesos reversibles, procesos que pueden llevarse en la dirección de avance o de retroceso mediante el cambio (infinitesimal) de una variable, y que finalmente alcanzan un punto donde las velocidades en ambas direcciones son idénticas, de modo que el sistema da la apariencia de tener una composición estática en la que la energía de Gibbs, G, es mínima. En el equilibrio la suma de*

los potenciales químicos de los reactivos es igual a la de los productos, de modo que: $\Delta G_r^0 = -RT \ln K$. *La constante de equilibrio K, es dada por la Ley de Acción de Masas*".

Una definición debe tener la característica de ser una proposición que precisa el significado de una unidad léxica, en este caso un concepto, para mejorar su comprensión, a la vez que describe sus cualidades o características esenciales. En la definición de la IUPAC identificamos elementos proposicionales no conexos, unos que podrían ser propios de la cinética y otros de la termodinámica. También identificamos algunos términos ambiguos o subjetivos, como son el de "dirección", "punto" y "apariencia". Igualmente, sin justificación, se relacionan la composición estática con un mínimo de la función energía de Gibbs. El breve análisis realizado a la definición propuesta por la IUPAC es una muestra de lo difícil que resulta en química dar definiciones, lo confusas que pueden llegar a ser las que encontramos en los textos, y de ahí en adelante las desafortunadas consecuencias en la enseñanza y en el aprendizaje (Raviolo & Martínez-Aznar, 2003; Raviolo, 2006; Quílez-Pardo, 2009).

3. RESULTADOS

Para el modelo de reacción $A \rightleftharpoons B$ se dan valores a las cantidades iniciales de reactivos $C_{A,0}$ y $C_{B,0}$, y a las constantes cinéticas k_1 y k_{-1}. Con estos datos se resuelve la ecuación (12) en el tiempo hasta alcanzar un valor constante, como se muestra en la Figura 1. Se observa que la reacción avanza desde $r = 0$ hasta $r = 0.104$, en el intervalo de tiempo de 0 a 250. En el equilibrio químico la variable independiente r tiene un valor que depende de los parámetros de la reacción, así que este valor no es característico de la reacción en estudio. Lo importante es conocer la trayectoria temporal que sigue la reacción a medida que esta avanza desde unas condiciones iniciales hacia el equilibrio. En la Figura 1 también se observa la respuesta de la reacción en equilibrio a una perturbación. En el tiempo 250 la concentración de equilibrio A se incrementó en 0.020 provocando que la reacción vuelva a comenzar, pero esta vez desde unas concentraciones iniciales diferentes: $C_{A,0} = 0.0655; C_{B,0} = 0.1145$. Estas concentraciones ya no son las de equilibrio, así que la reacción avanzará hasta alcanzar un nuevo estado de equilibrio, desde $r = 0$ hasta $r = 0.014$. Tengamos presente que r es una variable independiente que muestra la trayectoria de avance de una reacción desde su inicio, por lo tantos sus valores no son acumulativos, son una coordenada sobre la que se describe un proceso. El cambio temporal en las concentraciones de las sustancias químicas, la cinética, depende de la trayectoria seguida por la reacción durante el avance hacia el equilibrio. Así, los resultados mostrados en la Figura 1 se reemplazan en las ecuaciones (3) y (4), para obtener el cambio temporal en la composición de la reacción, como se muestra en la Figura 2. Se observa en la Figura 2, que en el intervalo de tiempo de 0 a 250 la concentración de A disminuye, mientras que la de B aumenta, describiendo una cinética característica de una ley de velocidad de primer orden. En el tiempo 250 las concentraciones de equilibrio son 0.0455 y 0.1145, respectivamente para A y B. En el tiempo 250 el equilibrio químico se perturba incrementando en 0.02 la concentración de A, y la reacción se reactiva para alcanzar una nueva composición de equilibrio, con concentraciones de 0.0515 y 0.1285, para A y B respectivamente. Como respuesta a la perturbación,

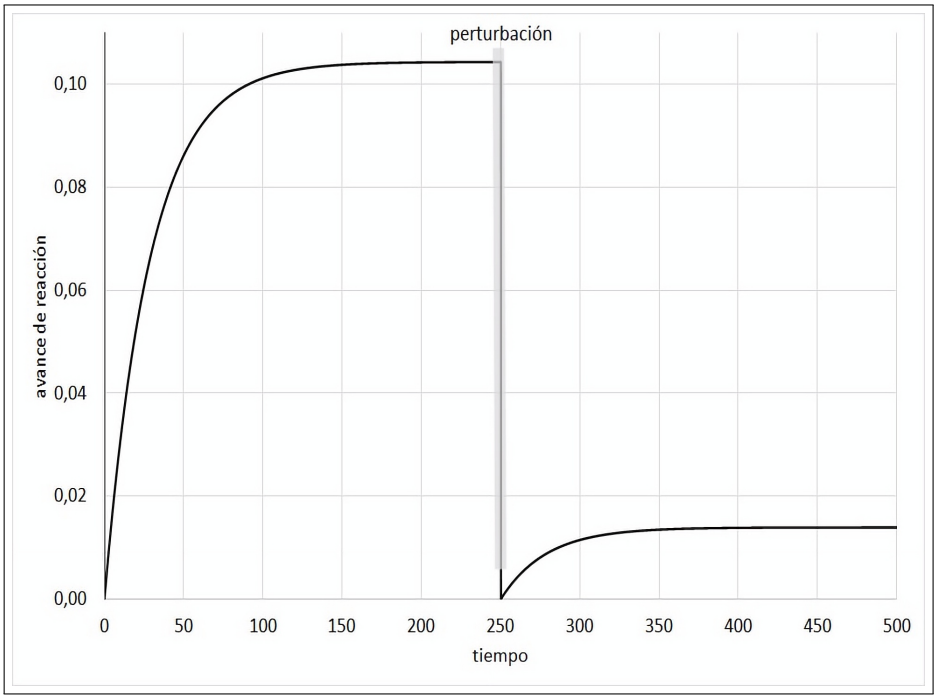


Figura 1: Avance de la reacción $A \rightleftharpoons B$ hacia el equilibrio químico. Los valores de los parámetros son: $C_{A,0} = 0.150; C_{B,0} = 0.010; k_1 = 0.025; k_{-1} = 0.01$. Fuente: Elaboración propia

la reacción $A \rightleftharpoons B$ restaura el estado de equilibrio químico alcanzando una nueva composición, que, por la naturaleza de la perturbación, corresponde a un incremento en las concentraciones de equilibrio de las sustancias A y B, dicho incremento se muestra en la figura como $\delta C_{A,eq}$ y $\delta C_{B,eq}$. Queda claro que, como respuesta a una perturbación en la concentración de equilibrio, una reacción química avanza hacia restaurar el estado de equilibrio con nuevas concentraciones de equilibrio de las sustancias químicas involucradas. Así, como el valor del avance de reacción en el equilibrio no caracteriza la reacción, tampoco lo hacen las concentraciones de equilibrio. Lo que se conserva como característica, como propiedad, de la reacción es la proporción relativa entre productos y reactivos en el equilibrio. De los datos de la Figura 2 se observa que haciendo uso de la ecuación (5), se obtiene que antes de la perturbación, el cociente de reacción en el equilibrio es $Q_r = 0.1145/0.0455 \approx 2.5$, y después de la perturbación es $Q_r = 0.1285/0.0515 \approx 2.5$. Este valor del cociente de reacción coincide con el de la constante de equilibrio $K = k_1/k_{-1} = 0.025/0,01 = 2.5$. A medida que la reacción avanza hacia el equilibrio, el cociente de reacción cambia hasta alcanzar un valor constante, y este valor constante es independiente de las concentraciones iniciales de las sustancias químicas. Entonces, la cantidad que caracteriza una reacción en estado de equilibrio, que es la propiedad característica, es el valor del cociente de reacción, que en el equilibrio químico es la misma constante analítica de equilibrio (Petrucci *et al.*, 2017).

A partir de los resultados de la ecuación (12) también se pueden evaluar las velocidades de reacción, reemplazando en la ecuación (10). Los resultados obtenidos se muestran en la Figura 3. Se observa que

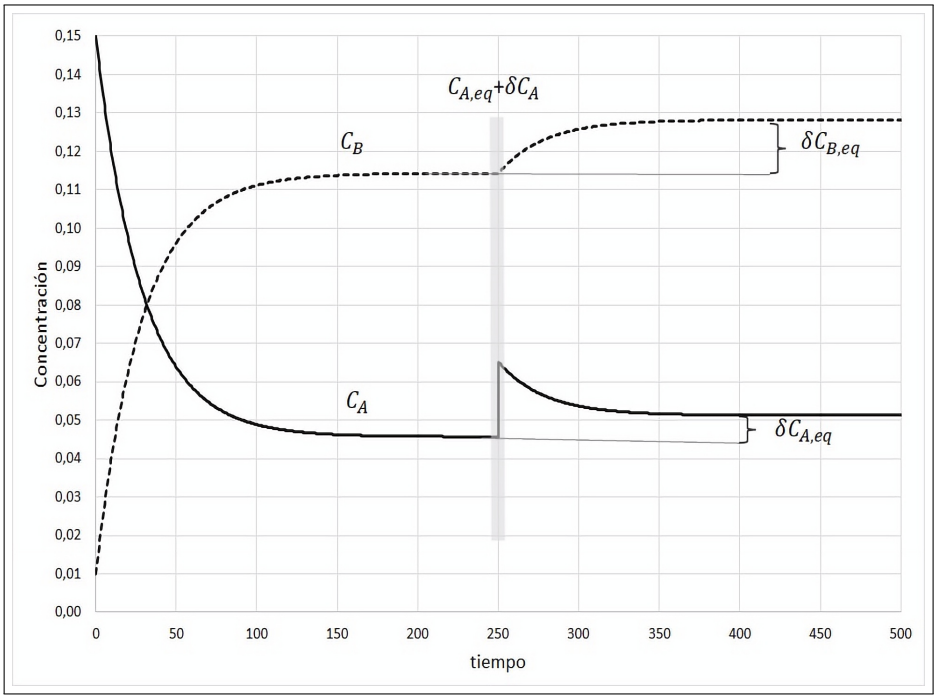


Figura 2: Avance de la composición de la reacción $A \rightleftharpoons B$ hacia el equilibrio químico. Los valores iniciales de los parámetros son: $C_{A,0} = 0.15; C_{B,0} = 0.01; k_1 = 0.025; k_{-1} = 0.01$. En el tiempo 250 la composición de equilibrio se perturba incrementado en 0.02 la concentración C_A. Fuente: Elaboración propia

a medida que la reacción avanza, curva (a) en la Figura 3, hacia el equilibrio, la velocidad total de reacción disminuye hasta hacerse 0 en el equilibrio, esto es consistente con afirmar que la reacción global terminó, y que las concentraciones de las sustancias ya no cambian. La curva (b) en la figura muestra la disminución en la velocidad directa de reacción, debido a que la concentración de reactivo va disminuyendo, mientras que la curva (c) muestra el aumento en la velocidad de reacción inversa, esto por el aumento en la concentración del producto. Cuando la reacción alcanza el equilibrio, las dos velocidades, directa e inversa, se hacen iguales y diferentes de cero. Es decir, el equilibrio químico es dinámico en el sentido que las velocidades de los procesos directos e inversos se mantienen: hay una continua transformación de reactivos en productos y de productos en reactivos. En el tiempo 250 se aplica la perturbación, aumento de la concentración de A. En la Figura 3 se observa que la reacción se reactiva con una velocidad total diferente de cero y que disminuye hasta hacerse 0 nuevamente. Igual sucede con las velocidades de reacción de los procesos directo e inverso, hasta que alcanzar un nuevo valor constante, igual, y diferente de cero.

Finalmente, los resultados obtenidos de la ecuación (12) se reemplazan en la ecuación (9) para analizar la energía de Gibbs de la reacción. La segunda ley de la termodinámica postula que todo sistema cerrado, a temperatura y presión constantes, evoluciona irreversiblemente hasta alcanzar el estado de equilibrio, y que ese estado de equilibrio es globalmente estable y se caracteriza por valores extremos en las funciones de estado, o potenciales termodinámicos. Para un sistema de reacción química, la función energía de Gibbs,

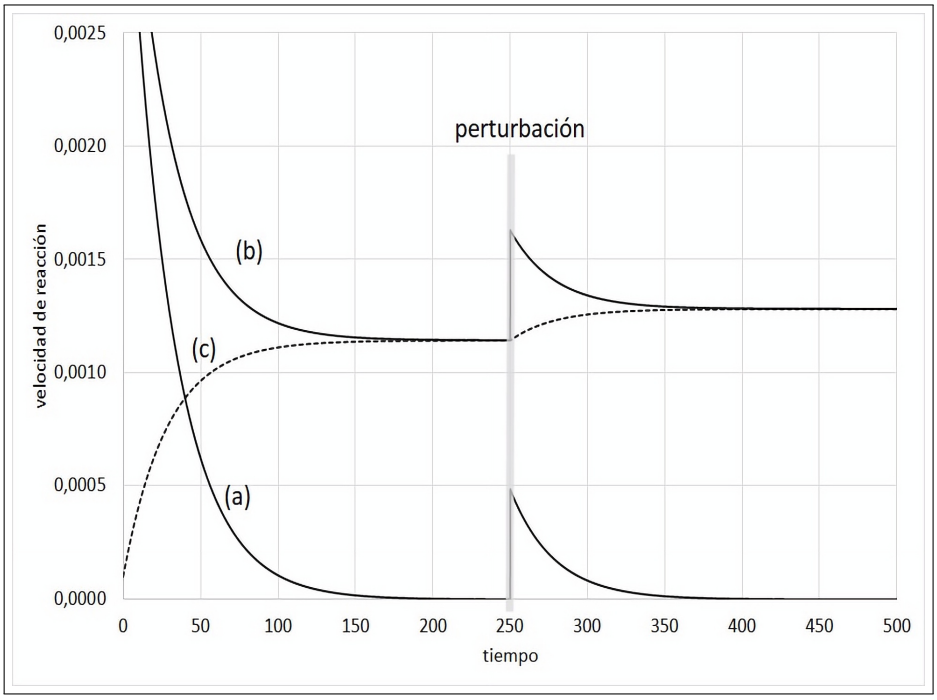


Figura 3: Velocidades de reacción de la reacción $A \rightleftharpoons B$ hacia el equilibrio químico. Los valores iniciales de los parámetros son: $C_{A,0} = 0.15; C_{B,0} = 0.01; k_1 = 0.025; k_{-1} = 0,01$. En el tiempo 250 la composición de equilibrio se perturba incrementado en 0.02 la concentración C_A. (a) velocidad total de reacción, $\nu_{total} = \nu_1 - \nu_{-1}$; (b) velocidad directa de reacción, $\nu_1 = k_1 \left(C_{A,0} - r\right)$); (c) velocidad inversa de reacción, $\nu_{-1} = k_{-1} \left(C_{B,0} + r\right)$. Fuente: Elaboración propia.

bajo las restricciones impuestas, tiene un mínimo global en el equilibrio químico. En los libros de texto el enunciado anterior se ilustra con una función convexa, donde la variable independiente suele denominarse "coordenada de reacción", "extensión de reacción", "grado de avance", "progreso de la reacción", pero no es claro cómo construir dicha curva (Dumié *et al.*, 1987; Burgess, 2003; Petrucci *et al.*, 2017; Atkins & De Paula, 2014).

En la Figura 1 se mostró que la variable independiente r, el avance de reacción alcanza un valor determinado en el equilibrio, dadas unas concentraciones iniciales de las sustancias químicas. La función energía de Gibbs evaluada con los valores de r, la denominamos G_r en la ecuación (9), cambiará a medida que avanza la reacción hasta alcanzar el valor de equilibrio. Si se modifican las concentraciones iniciales, la reacción avanza hasta un nuevo valor de r en el equilibrio, y por lo tanto pasa lo mismo con G_r. De los resultados de la Figura 2 se mostró que, para diferentes concentraciones iniciales, las concentraciones de equilibrio son diferentes, pero el cociente de reacción Q_r es el mismo. Así que el mínimo de la función de Gibbs en el equilibrio debe corresponder al valor característico de Q_r. Es necesario hacer otra consideración, y es que si la reacción tiene únicamente concentración inicial de A, y no ha avanzado, $r = 0$, el valor de G_r corresponde a la cantidad $N_A \mu_A$. De manera similar, si la reacción tiene únicamente concentración inicial de B, y no ha avanzado, el valor de G_r corresponde a la cantidad $N_B \mu_B$. El análisis anterior permite identificar los extremos de la

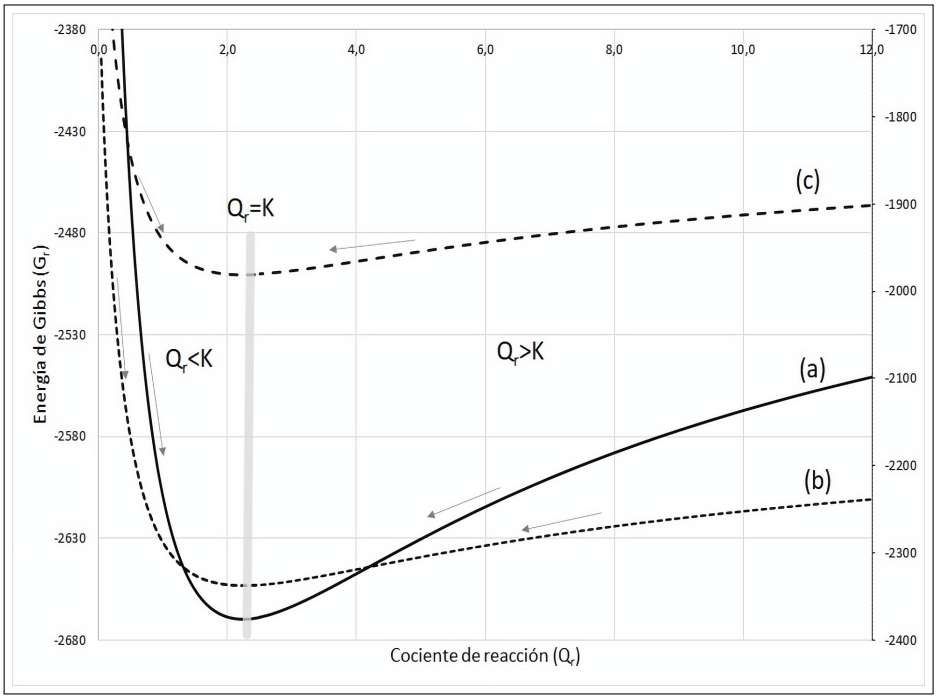


Figura 4: Función convexa energía de Gibbs. (a) $N_A + N_B = 0.30$; (b) $N_A + N_B = 0.25$; (c) $N_A + N_B = 0.20$;$k_1 = 0.025$; $k_{-1} = 0.01$; $\mu_A^0 = -3000$;$\mu_B^0 = -5000$. Fuente: Elaboración propia

función convexa de Gibbs, siempre y cuando a lo largo de la trayectoria de la curva se conserve la cantidad total de sustancia: $N_A + N_B = constante$. Es decir, la función convexa de Gibbs representa las trayectorias, el avance de reacción, para las que la cantidad total de sustancia se conserva: puede que inicialmente se tenga solo $N_A = x$, o sólo $N_B = x$, o una cierta cantidad de N_A y N_B, de modo que $N_A + N_B = x$. Para todas esas posibles trayectorias en la coordenada r, una vez se alcanza el estado de equilibrio, se tiene el mismo valor de Q_r.

En la Figura 4 se muestra la función convexa de la energía de Gibbs, donde la "coordenada" que describe el cambio de la función es el cociente de reacción. Las curvas (a), (b) y (c) corresponden a diferentes concentraciones iniciales de las sustancias químicas, y por lo tanto a una diferente cantidad total de sustancia conservada. Sin embargo, en todas ellas el mínimo coincide con el mismo valor del cociente de reacción, ya que la temperatura y la presión permanecen constantes. Los valores extremos de la curva dependen de los valores asignados a las energías libres molares estándar de las sustancias. En esta Figura 4 también se puede aclarar la respuesta del equilibrio químico ante una perturbación en las concentraciones. Se considera que las condiciones iniciales de reacción corresponden a las de la curva (c), y que la perturbación sea un incremento en las concentraciones iniciales, por ejemplo, las correspondientes a la curva (b), entonces como respuesta a la perturbación la función convexa de Gibbs describe una trayectoria diferente en r, pero con el mismo valor mínimo del cociente de reacción. Un error frecuente es interpretar la respuesta a la perturbación sobre la misma curva, calculando el cociente de reacción con la perturbación y leyendo sobre la curva si se encuentra a la derecha o la izquierda del mínimo. Las fechas sobre las curvas de la Figura 4 indican el

sentido en el que la curva se construye según las proporciones iniciales de las sustancias, y no corresponden a la respuesta de una perturbación. Es claro que, sí la reacción en equilibrio se perturba, se modifican las concentraciones y por lo tanto ya no se conserva la cantidad total de sustancia. El análisis de la respuesta a una perturbación, de un sistema de reacción en equilibrio químico, ha sido objeto de análisis con aportes muy interesantes (Quílez-Pardo, 1997; Quílez-Pardo, 2002; Torres, 2007).

4. CONCLUSIONES

Expresar todas las cantidades de una reacción química en función de la variable independiente avance de reacción, r, permite un análisis coherente de los aspectos cinéticos y energéticos que caracterizan el estado de equilibrio químico. El formalismo presentado en este trabajo se puede replicar para diferentes estequiometrías, por ejemplo, reacciones bimoleculares tales como $2AA \rightleftharpoons B$ o $A + CA \rightleftharpoons B$, permitiendo así al docente el desarrollo de un amplio y original material para llevar al aula de clase.

Los resultados obtenidos permiten expresar de manera clara y sustentada las siguientes características de una reacción en estado de equilibrio químico (en sistema cerrado y a temperatura y presión constantes):

- Un proceso cinéticamente reversible: es un proceso donde la velocidad de reacción directa e inversa son iguales, por lo tanto, la velocidad total de reacción es cero.
- Es un proceso donde la composición química es macroscópicamente invariable.
- Es un proceso termodinámicamente estable, caracterizado por un mínimo en la función energía de Gibbs, para un valor característico del cociente de reacción. De manera equivalente, la reacción avanza hacia el equilibrio químico a medida que el cociente de reacción cambia y la energía de Gibbs disminuye hasta alcanzar un mínimo.
- La propiedad fisicoquímica que caracteriza y define el estado de equilibrio químico es la constante de equilibrio. Como consecuencia de cualquier perturbación el cociente de reacción se aleja del valor de la constante de equilibrio, provocando que la reacción avance hasta que el cociente de reacción restaura su valor característico. Así, el estado de equilibrio queda definido unívocamente por la constante de equilibrio, bajo las restricciones impuestas.

Referencias

Atkins, P. & De Paula, J. (2014). Physical Chemistry. OUP. Oxford.

Bradley, J. D. & Steenberg, E. (2006). Symbolic language in chemistry - a new look at an old problem. *Proc 19th ICCE*, Seoul, 140.

Brown, T. L., LeMay Jr, H. E., Bursten, B. E. & Burdge, J. R. (2004). Química: la ciencia central. Pearson educación.

Burgess, A. E. (2003). The A to G of chemical equilibrium: Aspects depicted by helmholtz energy using its relationship to Gibbs energy. *J. Chem. Educ.* 80(**12**), 1476.

Chalk, S. J., McNaught, A. D. & Wilkinson, A. (2019). IUPAC. Compendium of Chemical Terminology.

Cheung, D. (2009). The adverse effects of Le Chatelier's principle on teacher understanding of chemical equilibrium. *J. Chem. Educ.* 86(**4**), 514.

Davenport, J. L., Leinhardt, G., Greeno, J., Koedinger, K., Klahr, D., Karabinos, M. & Yaron, D. J. (2014). Evidence-based approaches to improving chemical equilibrium instruction. *J. Chem. Educ.* 91(**10**), 1517-1525.

Dumié, M., Boulil, B. & Henri-Rosseau, O. (1987). On the mínimum of the Gibbs free energy involved in chemical equilibrium. *J. Chem. Educ.* 64(**3**), 201-204.

Driel, J. H. V. & Gräber, W. (2002). The teaching and learning of chemical equilibrium. In Chemical education: Towards research-based practice. Springer. Dordrecht. 271 p.

Ghirardi, M., Marchetti, F., Pettinari, C., Regis, A. & Roletto, E. (2014). A teaching sequence for learning the concept of chemical equilibrium in secondary school education. *J. Chem. Educ.* 91(**1**), 59-65.

Gillespie, R. J., Humphreys, D. A., Baird, C. & Robinson, E. A. (1990). Química, Tomo 2; Editorial Reverté. Barcelona. 819 p.

Gillespie, R. J., Eaton, D. R., Humphreys, D. A. & Robinson, E. A. (1994). Atoms, molecules, and reactions: An introduction to chemistry. Prentice Hall.

Harris, M. F. & Logan, J. L. (2014). Determination of log K ow Values for Four Drugs. *J. Chem. Educ.* 91(**6**), 915-918.

Honig, J. M. (2020). Thermodynamics: principles characterizing physical and chemical processes. Academic Press.

Hoppe, R. (1980). On the symbolic language of the chemist. *Angew. Chem.* 19(**2**), 110-125.

Kaya, E. (2013). Argumentation Practices in Classroom: Pre-service teachers' conceptual understanding of chemical equilibrium. *Int. J. Sci. Educ.* 35(**7**), 1139-1158.

Kujawski, J. B., Janusz, A. & Kuzma, W. (2012). Prediction of log P: ALOGPS application in medicinal chemistry education. *J. Chem. Educ.* 89(**1**), 64-67.

Martínez-Nuñez, E. (2002). Dinámica de reacciones unimoleculares en fase gas. Desviaciones del comportamiento estadístico. *Quim. Nova.* 25(**4**), 579-588.

Novak, I. (2019). Reversible Reactions: Extent of Reaction and Theoretical Yield. *J. Chem. Educ.* 97(**2**), 443-447.

Petrucci, R. H., Herring, F. G., Madura, J. D. & Bissonnette, C. (2017). General chemistry. Pearson. Canada.

Quílez-Pardo, J. & López, V. S. (1995). Errores conceptuales en el estudio del equilibrio químico: nuevas aportaciones relacionadas con la incorrecta aplicación del principio de Le Chatelier. *Enseñ. Cienc.* 72-80.

Quílez Pardo, J. (1997). Superación de errores conceptuales del equilibrio químico mediante una metodología basada en el empleo exclusivo de la constante de equilibrio químico. *Educ. Química.* 8(**1**), 46-54.

Quílez-Pardo, J. (2002). Una propuesta curricular para la enseñanza de la evolución de los sistemas en equilibrio químico que han sido perturbados. *Educ. Química*, 13(**3**), 170-187.

Quílez-Pardo, J. (2009). Análisis de los errores que presentan los libros de texto universitarios de química general al tratar la energía libre de Gibbs. *Enseñ. Cienc.* 317-330.

Raviolo, A. & Martínez-Aznar, M. (2003). Una revisión sobre las concepciones alternativas de los estudiantes en relación con el equilibrio químico. Clasificación y síntesis de sugerencias didácticas. *Educ. Química.* 14(**3**), 159-165.

Raviolo, A. (2006). Las imágenes en el aprendizaje y en la enseñanza del equilibrio químico. *Educ. Química.* 17(**4e**), 300-307.

Rogers, F., Huddle, P. A. & White, M. W. (2000). Simulations for teaching chemical equilibrium. *J. Chem. Educ.* 77(**7**), 920.

Taber, K. S. (2009). Learning at the symbolic level. In: Multiple representations in chemical education. Gilbert, J. K.; Treagust, D. eds. Springer, Dordrecht, 45-105.

Torres, E. M. (2007). Effect of a perturbation on the chemical equilibrium: Comparison with Le Châtelier's principle. *J. Chem. Educ.* 84(**3**), 516.

Troe, J. U. R. C. E. N. (2012). Unimolecular reactions: experiments and theories. In: Kinetics of Gas Reactions. Eyring H.; Henderson D. & Jost W. eds. Physical Chemistry: An Advanced Treatise, 6, 835-929.

Turner, D. E. (1994). An experiment to demonstrate the effect of pH on partition coefficients in liquid-liquid extraction. *J.Chem. Educ.* 71(**2**), 173.

Tyson, L., Treagust, D. F. & Bucat, R. B. (1999). The complexity of teaching and learning chemical equilibrium. *J.Chem. Educ.* 76(**4**), 554-558.

Revista de la Facultad de Ciencias Universidad Nacional de Colombia, Sede Medellín
V 11 N°2 julio-diciembre de 2022 • ISSN-e 2357-5749 • Artículo Investigación • Páginas 162 to 177
DOI: https://doi.org/10.15446/rev.fac.cienc.v11n2.100727

PROPUESTA DE SEGUIMIENTO DE LA LIMPIEZA DEL RÍO BOGOTÁ A PARTIR DE SUS SERVICIOS ECOSISTÉMICOS[a]

PROPOSAL FOR THE MONITORING THE CLEANING OF THE BOGOTÁ RIVER TO APPEAR ON ITS ECOSYSTEM SERVICES

CAROLINA VILLEGAS VARGAS[b*]

Recibido 27-01-2022, aceptado 28-06-2022, versión final 30-06-2022
Artículo Investigación

RESUMEN: A partir de la sentencia del río Bogotá del Consejo de Estado del año 2014, esta investigación propone un indicador para hacerle seguimiento a la recuperación de los servicios ecosistémicos del río, y de esta manera darle contexto biológico, social, cuantitativo y cualitativo a la sentencia. Para conocer el estado de los servicios ecosistémicos se diseñó una encuesta que evaluó 15 servicios ecosistémicos por medio de la escala de Likert del acuerdo, que fue aplicada a 266 personas adultas en dos lugares de la ribera del río: apenas éste llega a la ciudad conurbada y apenas sale de la ciudad. Entre los resultados más importantes se encuentra que la salud del ecosistema río Bogotá se deteriora al pasar por la urbe y recibir sus desechos, y, contrario a este resultado se encuentra que el beneficio de la naturaleza más valorado por la población encuestada fue cuidar de la naturaleza para las generaciones futuras. Un resultado optimista con respecto al futuro que más que certezas deja preguntas acerca de cómo en entornos deteriorados aparecen valores trascendentales que evocan un mundo en paz y en unidad con la naturaleza.
PALABRAS CLAVE: Beneficios; bienestar; contribuciones; naturaleza.

ABSTRACT: Based on the socio-environmental conflict and the Bogotá River ruling of the Council of State in 2014, this research proposes an indicator to monitor the recovery of the river's ecosystem services, and in this way give quantitative and qualitative social biological context to the river. sentence. To know the state of ecosystem services, a survey was designed that evaluated the valuation of 15 ecosystem services through the Likert scale of the agreement, which was applied to 266 adults in two places on the riverbank, as soon as the river reaches to the conurbated city and barely leaves the city. As the most important results, we found that the health of the Bogotá River ecosystem deteriorates as it passes through the city and receives its waste. Contrary to this expected result, we found that the benefit of nature most valued by the surveyed population was caring for nature for future generations. An optimistic result with respect to the future that, more than certainties, leaves questions about how transcendental values appear in deteriorated environments that evoke a world in peace and in unity with nature.
PALABRAS CLAVE: Benefits; contributions; nature; well-being.

[a]Villegas-Vargas, C. (2022). Propuesta de seguimiento de la limpieza del río Bogotá a partir de sus servicios ecosistémicos. *Rev. Fac. Cienc.*, 11 (**2**), 162–177. DOI: https://doi.org/10.15446/rev.fac.cienc.v11n2.100727
[b]Jardín Botánico de Bogotá José Celestino Mutis
[*]Autor correspondencia: carolinavillegasvargas@gmail.com

1. INTRODUCCIÓN

El río es el eje de la estructura ecológica principal de la Sabana de Bogotá, y es proveedor de servicios ecosistémicos y beneficios de la naturaleza para la ciudad (van der Hammen, 1998). Es en el río Bogotá donde el ambiente, la sociedad y la urbe confluyen (Osorio-Osorio, 2007), generando impactos de magnitudes aún no dimensionadas y que como sociedad urbana estamos en deuda de solucionar (Consejo de Estado, 2014).Es en este río en donde la diada sociedad-naturaleza evidencian lo problemático.

El río como eje ecológico regional presenta una alta biodiversidad, esta condición dual de corredor ecológico y de recolector de los desechos es el ejemplo perfecto de la compleja relación de la sociedad con la naturaleza. Este escenario ejemplifica la problemática socioambiental alrededor del agua en la que confluyen asuntos de saneamiento básico, salud pública, contaminación e infraestructura (Osorio-Osorio, 2008).

En consonancia con la importancia que tiene el río para la ciudad y para el país, esta investigación pretende evaluar cuáles son los servicios ecosistémicos más valorados por las personas que viven o pasan cerca de la ribera del cuerpo de agua, en el sector de las microcuencas Tintal y Torca.

Teniendo en cuenta que los servicios ecosistémicos son las contribuciones de la naturaleza al bienestar humano (Díaz *et al.*, 2019), se quiso indagar cómo la gente valora 15 servicios ecosistémicos en la ribera del río Bogotá, en dos lugares de la ciudad por donde pasa el río: la desembocadura del humedal La Conejera (localidad de Suba) que corresponde al lugar en donde el río entra en contacto con la ciudad densa, y la salida del río luego de recorrer la urbe y recibir sus vertimientos en el puente de la Avenida Longitudinal de Occidente -ALO-, en la localidad de Bosa, 500 metros antes de recibir las aguas del río Tunjuelo (ver Figura 1).

La valoración de los servicios ecosistémicos se inaugura con la publicación de Costanza *et al.* (1997), quienes afirmaron que la sostenibilidad de la vida humana en el planeta depende de los servicios ecológicos que, además, pueden tener un costo económico infinito. Luego de esta publicación hay un auge en la consolidación del concepto que continúa en el año 2000 con la declaración de los Objetivos del Milenio (Naciones Unidas ONU, 2000). Es así como en todo el planeta se inicia una evaluación de los ecosistemas que, después de un arduo trabajo de más de 1300 científicos dieron como resultado la publicación Evaluación de Ecosistemas del Milenio (MEA, 2005).Esta evaluación hizo recomendaciones y les dio énfasis a las evaluaciones científicas para que fueran consideradas por los tomadores de decisiones a nivel global, recomendaciones para que las decisiones sobre los ecosistemas sean pertinentes, no prescriptivas y con información de utilidad oportuna sobre si la acción u omisión es acorde con los escenarios estudiados (Corvalán *et al.*, 2005). Además, se consolida la estrecha relación espejo entre la salud humana y salud del ecosistema, que ha sido definida por la Organización Mundial de la Salud como “una salud” (Corvalán *et al.*, 2005).

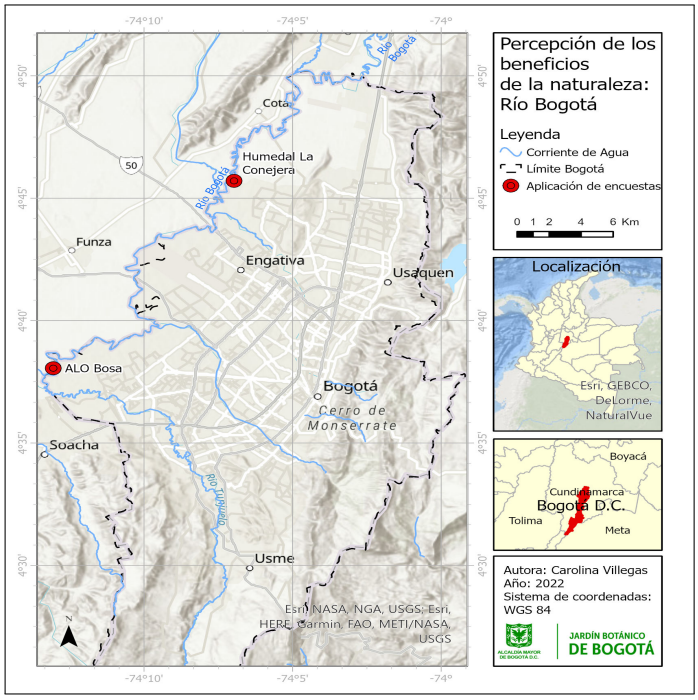


Figura 1: Mapa de la zona de estudio. Fuente: Elaboración propia

Como siguiente referente de dimensiones planetarias, en el año 2010, se inició la consolidación de la Plataforma Intergubernamental de Ciencia y Política sobre Biodiversidad y Servicios de los Ecosistemas, (en adelante IPBES, por sus siglas en inglés). Esta plataforma nació como un panel inspirado en el Panel Intergubernamental de Expertos en Cambio Climático (IPCC), que surge como un panel científico que contribuye al Convenio Macro de las Naciones Unidas Contra el Cambio Climático. En la evaluación de la IPBES del año 2019 los servicios ecosistémicos definidos como las contribuciones de la naturaleza al bienestar se clasificaron en 18 categorías materiales, no materiales y regulatorias. También se incluyó y legitimó el conocimiento de comunidades indígenas y comunidades locales con enfoques diversos en este tipo de estudios (Díaz *et al.*, 2019).

1.1. El río Bogotá y su descontaminación

En cuanto al río Bogotá, límite natural de la microcuenca y de la ciudad de Bogotá, en el año 2014, el Consejo de Estado lo declaró en situación de catástrofe ambiental, ecológica y económico-social por la acción y omisión de: habitantes, industrias, entidades nacionales, entidades distritales y municipios. Para atender la situación del río, la sentencia sobre la descontaminación del río Bogotá del 28 de marzo del 2014, ordenó a todas las entidades del Estado construir parámetros y acciones de coordinación, para construir políticas, planes y programas medioambientales para recuperar, sanear y conservar la cuenca del río Bogotá

(Orozco-Roa, 2016).

Esta sentencia, precedente de acción colectiva por promover políticas públicas y generar decisiones trascendentales en la protección del medio ambiente y los derechos colectivos (Guiza-Suárez *et al.*, 2015) ha promovido la construcción de espacios de coordinación e interacción entre entidades del ámbito nacional, departamental, distrital, municipal, privados, de entidades educativas y de organizaciones de la sociedad civil (Observatorio Regional Ambiental y de Desarrollo Sostenible del Río Bogotá ORARBO, 2022). Asimismo, ha ampliado las perspectivas del conflicto socioambiental, para concretarse en acuerdos de protección de derechos humanos fundamentales vulnerados, en particular el derecho a gozar de un ambiente sano (Guiza-Suárez *et al.*, 2015).

Esta investigación también pretende aunarse a esta iniciativa y hacerle seguimiento a esta sentencia, mediante la propuesta de un indicador cuantitativo y cualitativo de la valoración social de los servicios ecosistémicos alrededor del río Bogotá, que pueda servir como línea base para evaluar la recuperación del río en el mediano y el largo plazo.

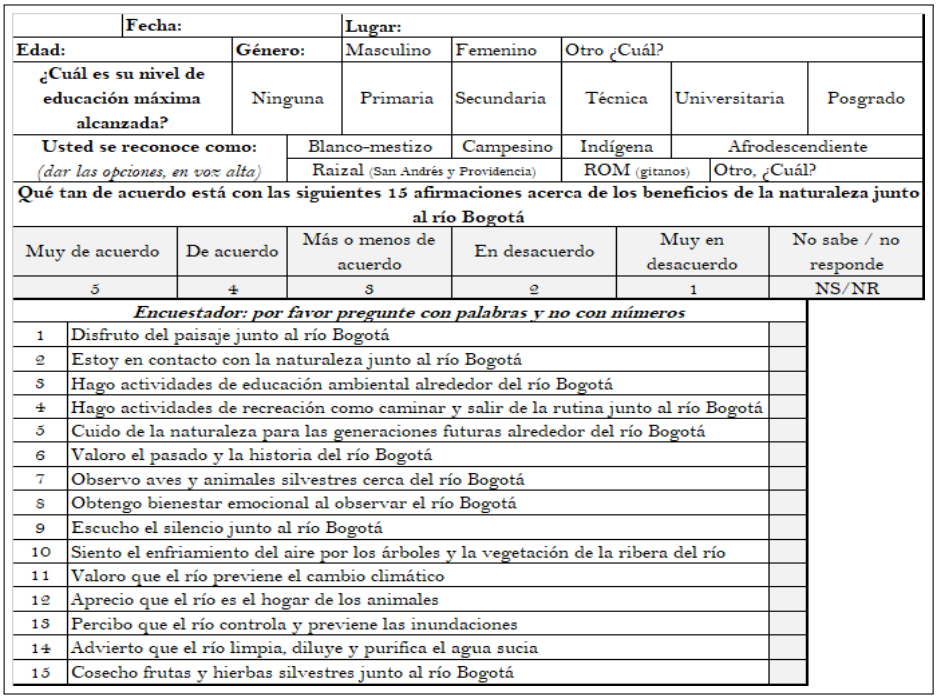
Para esta investigación, reconocer el valor que tiene la naturaleza en la sociedad influye en la manera como la sociedad puede regularse sin recurrir a la monetización o a la economía. Asimismo, evaluar los servicios que provee el ecosistema desde diferentes perspectivas, y no solo la económica, permite abordar el problema del crecimiento de las ciudades y las demandas humanas por bienestar, corrigiendo la clásica tendencia que favorece a la riqueza privada y al capital, frente a la riqueza pública y al capital natural (Sukhdev *et al.*, 2014). Abordar las áreas verdes urbanas, de las que hace parte la ribera del río Bogotá, como fuentes de bienestar desde una perspectiva multidisciplinaria más que economicista o biologicista, puede ser de utilidad para los tomadores de decisiones (Pereira-Prado, 2015).

2. MATERIALES Y MÉTODOS

2.1. Delimitación espacial

El río Bogotá nace a 3.300 m.s.n.m. en el municipio Villapinzón (Cundinamarca), recorre 380 km y desemboca a 280 m.s.n.m. en el río Magdalena, en el municipio de Girardot (Cundinamarca). Cubre un área total de 589.143 hectáreas, riega a 45 municipios del departamento, recibe las aguas de 19 subcuencas (Corporación Autónoma Regional de Cundinamarca CAR, 2006), y de acuerdo al último censo de 2018 recibe los vertimientos de 7′181.469 habitantes de Bogotá (Departamento Nacional de Estadística, 2018), además de otros municipios.

Este río recibe las externalidades del 26 % de la actividad económica del país a través de tres afluentes que lo contaminan hasta convertirlo en uno de los más contaminados del mundo: el río Juan Amarillo, que le aporta 123 toneladas de sólidos en suspensión de desechos al día (ton/día); el río Fucha con 590 ton/día y el



Fecha:		Lugar:				
Edad:	**Género:**	Masculino	Femenino	Otro ¿Cuál?		
¿Cuál es su nivel de educación máxima alcanzada?	Ninguna	Primaria	Secundaria	Técnica	Universitaria	Posgrado
Usted se reconoce como: *(dar las opciones, en voz alta)*	Blanco-mestizo	Campesino	Indígena	Afrodescendiente		
	Raizal (San Andrés y Providencia)		ROM (gitanos)	Otro, ¿Cuál?		

Qué tan de acuerdo está con las siguientes 15 afirmaciones acerca de los beneficios de la naturaleza junto al río Bogotá

Muy de acuerdo	De acuerdo	Más o menos de acuerdo	En desacuerdo	Muy en desacuerdo	No sabe / no responde
5	4	3	2	1	NS/NR

Encuestador: por favor pregunte con palabras y no con números

1	Disfruto del paisaje junto al río Bogotá	
2	Estoy en contacto con la naturaleza junto al río Bogotá	
3	Hago actividades de educación ambiental alrededor del río Bogotá	
4	Hago actividades de recreación como caminar y salir de la rutina junto al río Bogotá	
5	Cuido de la naturaleza para las generaciones futuras alrededor del río Bogotá	
6	Valoro el pasado y la historia del río Bogotá	
7	Observo aves y animales silvestres cerca del río Bogotá	
8	Obtengo bienestar emocional al observar el río Bogotá	
9	Escucho el silencio junto al río Bogotá	
10	Siento el enfriamiento del aire por los árboles y la vegetación de la ribera del río	
11	Valoro que el río previene el cambio climático	
12	Aprecio que el río es el hogar de los animales	
13	Percibo que el río controla y previene las inundaciones	
14	Advierto que el río limpia, diluye y purifica el agua sucia	
15	Cosecho frutas y hierbas silvestres junto al río Bogotá	

Figura 2: Encuesta diseñada para llevar a cabo esta investigación

río Tunjuelo con 616 ton/día (Corporación Autónoma Regional de Cundinamarca CAR, 2006).

2.2. Diseño de encuesta

Para definir la valoración social de los servicios ecosistémicos primero se realizó una caracterización preliminar de los servicios y del contexto (Rincón-Ruiz *et al.*, 2014). La metodología seleccionada fue la aplicación de encuestas (Martín-López *et al.*, 2012).

La encuesta es una técnica de investigación para explorar, describir y explicar una realidad social de una muestra de casos representativa de una población, que usa un conjunto expandido de instrumentos y procedimientos que se centra en la recolección precisa de datos y variables dependientes (Rojas-Tejada *et al.*, 1998). El cálculo de la muestra mínima estándar para las ciencias sociales con un nivel de confianza de 95 %, un porcentaje de error del 6 % y una población afectada de 100.000 personas, fue de 266 personas (Morales-Vallejo, 2008), número total de encuestas aplicadas. Los análisis estadísticos se llevaron a cabo mediante el programa Excel y RStudio (RStudio, 2019).

La población encuestada se clasificó por edad, género, nivel de educación y autorreconocimiento étnico (Martín-López *et al.*, 2012). La encuesta sin preguntas adicionales no analizadas en esta publicación está disponible en la Figura 2. En el diseño de la encuesta se tuvo especial cuidado con el lenguaje para incluir a personas con educación formal y a personas desescolarizadas: palabras fáciles de entender para una variedad

Figura 3: Escala de Likert para responder a la valoración de los servicios ecosistémicos en la ribera del río Bogotá. El valor máximo fue de 5 y correspondió a la valoración de "Muy de acuerdo", le sigue "De acuerdo" con valor de 4, "Más o menos de acuerdo" con valor de 3, "En desacuerdo" con valor de 2, "Muy en desacuerdo" con valor de 1, "No sabe" y "No responde" con valor de 0. Esta encuesta fue diseñada y aplicada en papel como de manera electrónica con el programa Survey 123 (Esri, 2017). Fuente: Elaboración propia.

de público adulto, con cuidado de no considerar a las personas como ignorantes en el área específica de los servicios ecosistémicos, sino como receptores innatos de los beneficios de la naturaleza esenciales para todo ser humano (Himes & Muraca, 2018). La encuesta no pretendió enseñarle a la gente nada nuevo ni evaluar el nivel de educación ambiental de la población encuestada; por el contrario, el propósito fue evaluar la manera como las personas han percibido la recuperación del río y la manera como se ha democratizado el concepto de naturaleza (Ainscough *et al.*, 2019; Taylor & Bogdan, 2018).

Se hizo énfasis en construir conceptos coherentes, lógicos y realistas, pensando en darle a la encuesta un contexto operativo de fácil entendimiento por la población encuestada (Nahlik *et al.*, 2012). Para desarrollar este lenguaje incluyente se tuvo especial atención en denominar cada servicio ecosistémico con adjetivos entendibles para la mayor cantidad de público posible. En este sentido, por ejemplo, no se preguntó por el servicio ecosistémico "hábitat de animales"; sino que se preguntó qué tan de acuerdo está con el beneficio de la naturaleza "aprecio que el río es el hogar de los animales", ver Figura 2.

La encuesta y las categorías de clasificación social, como instrumentos de valoración social fue elaborada teniendo en cuenta varias fuentes bibliográficas, entre ellas Martín-López *et al.* (2011), Martín-López *et al.* (2012),Pereira-Prado (2015) y Vilardy-Quiroga & González-Nóvoa (2018). Para darle mayor profundidad a cada uno de los servicios ecosistémicos evaluados se tuvo en cuenta la escala de Likert del acuerdo, que evalúa el nivel de afinidad acerca de 15 afirmaciones diseñadas. Esta escala de Likert del acuerdo fue escogida debido a que muchas veces las afirmaciones fueron incongruentes con la realidad de contaminación y desolación que prevalece en áreas afectadas por el río. La escala del acuerdo se explica en la Figura 3, que fue un infograma que se le entregó a cada persona que respondió la encuesta para que tuviera presente las opciones y no las olvidara en el momento de la respuesta.

Si bien el servicio ecosistémico "generaciones futuras" que se incluyó en esta investigación no está contenido en las referenciadas y clásicas clasificaciones de servicios ecosistémicos, como la Evaluación de los Ecosistemas del Milenio MEA (2005), la IPBES se ha abierto a incluir dentro de las investigaciones servicios más locales, plurales y diversos (Díaz *et al.*, 2015; Díaz *et al.*, 2018; Pascual *et al.*, 2017). En este contexto de ampliación del enfoque se incluyó este servicio ecosistémico investigado por Lin *et al.* (2017), atendiendo a nuevos beneficios del ecosistema percibidos por las personas. Además, se incluyó el servicio ecosistémico cultural "silencio", que corresponde a bajos niveles de ruido que permite escuchar los sonidos de la naturaleza silenciados por la presencia de la ciudad.

3. RESULTADOS

Se aplicaron en total 266 encuestas en la ribera del río Bogotá, a personas mayores de 18 años que pasaron por el lugar de estudio, 96 encuestas en cercanía al humedal La Conejera y 170 en cercanía del puente de la ALO. El trabajo de campo se llevó a cabo en días hábiles y en horas de la mañana del 14 de mayo al 3 de noviembre del año 2019.

La población encuestada mostró mayor representatividad del género masculino (60 %) que del femenino (40 %). La presencia baja del género femenino en el lugar de estudio, suponemos que fue por razones relacionadas con las actividades del cuidado que las confinan al hogar, de acuerdo con los tiempos en los cuales se aplicaron las encuestas, que fueron días y horas hábiles.

En cuanto al nivel de educación, el grupo mejor representado reportó tener educación secundaria (41.7 %), seguido de personas con formación primaria (21.4 %), formación técnica (17.3 %), formación universitaria (13.9 %), ninguna formación (3.8 %), y por último posgrado (1.9 %), ver Tabla 1.

En cuanto a la clasificación etaria, la encuesta no incluyó a jóvenes ni a niños. La mayoría de las encuestas (91 %) fueron aplicadas a adultos de entre 19 y 60 años, y muy pocos adultos mayores de 60 años (9 %). La composición de la población está representada en su mayoría por personas que se definen a sí mismas como blanco-mestizos (60.8 %), le siguen en orden numérico, los campesinos (28.5 %), los afrodescendientes (5.4 %), indígenas (3.5 %), personas del pueblo ROM -Gitano- (0.4 %) y por último otros (1.5 %).

Tabla 1: Características de la población encuestada. $(*)$ La clasificación de autorreconocimiento étnico de "campesino" se incluyó de acuerdo a la sentencia de la Corte Suprema de Justicia 2028/2018, que falló a favor de 1770 campesinas y campesinos para ser incluidos como sujetos políticos para el Estado colombiano, y que reconoce en el sujeto político campesino un grupo culturalmente diferenciado y que necesita especial protección con la premisa de "para que el campesinado cuente, tiene que ser contado" (Dejusticia, 2019).

Clasificación	Grupos	Valor Absoluto	Valor Relativo
Distribución	Femenino	107	40 %
géneros	Masculino	159	60 %
	Ninguna	9	3.8 %
	Primaria	57	21.4 %
Nivel de	Secundaria	111	41.7 %
Educación	Técnica	46	17.3 %
	Universitaria	37	13.9 %
	Posgrado	5	1.9 %
Clasificación	Adultos 19-60 años	231	91 %
etaria*	Adultos mayores	22	9 %
	Blanco mestizo	158	60.8 %
	Campesino*	74	28.5 %
Autorreconocimiento	Indígena	9	3.5 %
étnico	Afrodescendiente	14	5.4 %
	ROM(Pueblo gitano)	1	0.4 %
	Otro	4	1.5 %

Mediante el programa R (RStudio, 2019), el análisis estadístico descriptivo de las variables "servicios ecosistémicos" mostró los siguientes resultados:
En la muestra de el Humedal La Conejera los servicios ecosistémicos presentaron una curtosis alta, es decir, una agrupación de los datos en el centro de la curva, mayor de uno, haciendo evidente que los datos se concentran en torno a la media, y que en la mayoría de los casos muestra una deviación estándar de entre 0.3 y 1.3; menos hogar de animales y recolección de frutos, que mostraron curtosis cercanas a cero, es decir, con una curva normal menos pronunciada con datos distribuidos a lo largo del rango.

En la muestra del puente de la ALO los servicios ecosistémicos presentaron una curtosis bajísima cercana a cero, haciendo evidente que las variables presentan valores variados que se distribuyen a lo largo del rango y en la mayoría de los casos muestra una deviación estándar alta, en comparación con Humedal La Conejera, de entre 1.1 y 1.6.

En cuanto a la prueba de Shapiro, todos los resultados arrojaron que ninguna variable tuvo distribución normal. Todas las variables analizadas presentaron asimetrías negativas de altos niveles de acuerdo, menos recolección de frutos que viró hacia el desacuerdo.

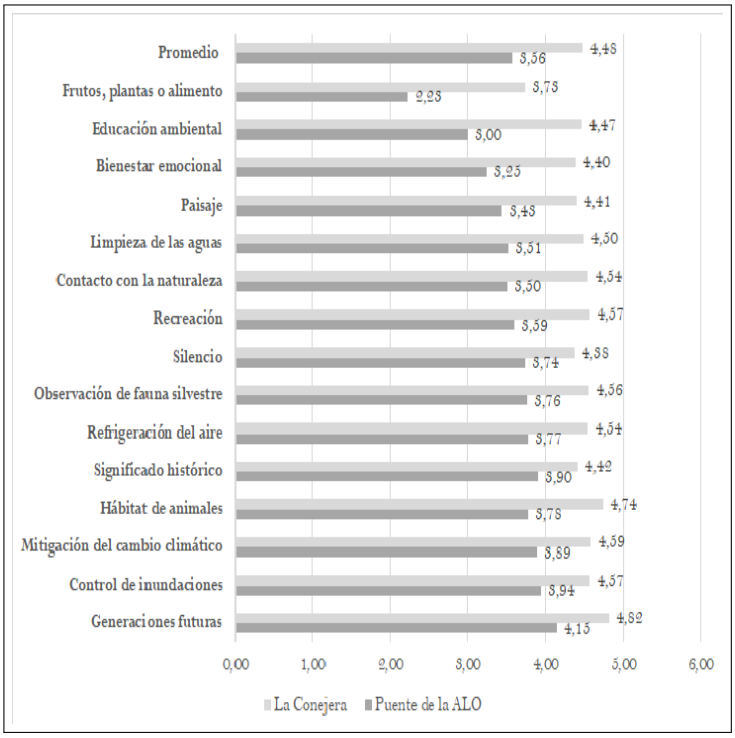


Figura 4: Promedio de la valoración social de servicios ecosistémicos evaluados. Fuente:Elaboración propia.

Acorde con el análisis de los promedios de la valoración de los beneficios de la naturaleza evaluados con la escala de Likert del acuerdo se encontró que el río Bogotá presenta un mejor estado de los servicios ecosistémicos en la desembocadura del humedal La Conejera de la localidad de Suba, que corresponde a la entrada del río a la ciudad densamente construida, en comparación con el puente de la Avenida Longitudinal de Occidente ALO de la Localidad de Bosa, en donde el río sale de la ciudad (ver Figura 4).

El valor promedio de los servicios ecosistémicos del río asignado por la población encuestada en la desembocadura del humedal La Conejera fue de 4.48 -unidades de valoración social en la escala de Likert del acuerdo-; mientras que el valor promedio en el puente de la ALO fue de 3.56. Este resultado hace evidente que la salud del ecosistema río Bogotá se deteriora al pasar por la urbe y recibir sus desechos (ver Figura 4), en este espacio el río pierde 0.92 -unidades de valoración social en la escala de Likert del acuerdo- (diferencia entre el promedio de la valoración de los servicios ecosistémicos del río desde que entra hasta que sale de la ciudad).

El servicio ecosistémico más importante para la población encuestada fue el valor trascendental de legado "generaciones futuras", un servicio del tipo cultural, que presentó un valor promedio en la escala de Likert del acuerdo de 4.82 en La Conejera y de 4.15 en el puente de la ALO, (Figura 4). Le siguen en orden de

importancia, “el control de inundaciones” y “mitigación del cambio climático”; estos dos servicios del tipo “regulación”, además del servicio “hábitat de animales”, servicio de apoyo (MEA, 2005). Estos resultados muestran que la población encuestada valora al río por sus diversas funciones ecosistémicas culturales, de regulación y de apoyo.

Además de los 15 servicios ecosistémicos evaluados, la población encontró como nuevos beneficios de la naturaleza: resiliencia de la naturaleza, agua para riego de cultivos, cuidado del agua, esperanza, transporte fluvial.

4. DISCUSIÓN

El beneficio de la naturaleza más valorado por la población encuestada fue “cuidar de la naturaleza para las generaciones futuras”. Este servicio ecosistémico emergente hace evidente que la población ha asimilado conceptos políticos relacionados con una ecologización del discurso (Lemkow, 2002). Este servicio también puede ser definido como un “valor trascendental” (Raymond & Kenter, 2016), que alude nociones éticas deseables de “un mundo en paz y unidad con la naturaleza”. Estos “valores trascendentales” afectan el comportamiento e incluyen maneras en que se interioriza el conocimiento y la evidencia en el comportamiento proambiental. Kenter *et al.* (2015) afirman que los valores trascendentales son valores incompletos y ambiguos sobre estados finales deseables, o comportamientos que trascienden situaciones específicas y guían la selección o evaluación de comportamientos y eventos (Schwartz & Bilsky, 1987).

El servicio ecosistémico emergente “generaciones futuras” y el nuevo servicio ecosistémico “resiliencia de la naturaleza” están en consonancia con lo propuesto por Díaz *et al.* (2018); quienes invitan a tener en cuenta en las investigaciones de valoración social de los servicios ecosistémicos de escala local y diversos, nuevos servicios que no hagan parte de los establecidos por las Evaluaciones del Milenio (Corvalán *et al.*, 2005) y que involucren otras culturas y otras maneras de relacionarse con la naturaleza.

En el momento actual, la percepción de deterioro ambiental ha generado en el común de las personas un sesgo pesimista derivado del abuso del mensaje del riesgo climático (Huertas & Corraliza, 2016; DStocknes, 2014), y/o una “ecofatiga”, sentimiento de indefensión que traslada la responsabilidad a otros y la consecuente sensación de que lo que cada persona haga es muy poco y no tiene consecuencias eficaces ante la gravedad del problema climático. Contrario a este sentimiento pesimista, los resultados de esta investigación muestran que las percepciones de las personas alrededor de la naturaleza, a escala local, en la ribera del río Bogotá son optimistas con respecto al futuro. Este extraño optimismo, así lo denomino por lo fétido y deteriorado que es del río, me contagió de un optimismo que a veces entiendo como irracional, aunque está relacionado también con un sentimiento de nostalgia por un río que hace muchos años habitó la infancia de algunas personas encuestadas, que los recuerdan limpio y que lo proyectan con una mirada optimista de un mundo mejor. Una mirada que muestra como aquellos que vivimos alejados del río Bogotá lo valoramos por

prejuicios y de maneras desvinculadas al lugar y a la experiencia. Pareciera que solo viviendo cerca de él o recorriéndolo con frecuencia se pudieran generar conexiones humanas con el entorno. Estas conexiones son la base de la topofilia o "apego al lugar" (Tuan, 1974), concepto que explica que los vínculos del ser humano con su entorno son un "vínculo afectivo" que se aprende en la práctica y en contacto con la naturaleza.

Esta mirada local de los beneficios de la naturaleza en la ribera del río Bogotá podrían incidir en potenciales miradas optimistas de futuro acordes con la topofilia, que además, pueden ser bidireccionales, es decir, pueden ir tanto de la naturaleza a los humanos como de los humanos a la naturaleza (Escalera, 2018), a manera de miradas que embellecen el entorno y lo rodean de un halo optimista.

Por último, Ainscough *et al.* (2019) afirman que el trabajo sobre servicios ecosistémicos, además de integrar el concepto de sostenibilidad, debe ser interdisciplinario. En concordancia a lo expuesto por este autor quiero invitar a interpretar estos mismos datos desde diferentes perspectivas del conocimiento para enriquecer la discusión acerca de lo natural en la urbe, y hacer aportes para mejorar el estado y la relación con nuestro gran río.

5. CONCLUSIONES

A pesar del deterioro evidente del río Bogotá y de haber perdido todos los servicios del tipo "sostenimiento" perceptibles por las personas, esta corriente de agua sique siendo valorada por sus capacidad para proveer servicios ecosistémicos del tipo cultural, de regulación y de apoyo.

El beneficio más valorado por la población encuestada fue "cuidar de la naturaleza para las generaciones futuras", valor de legado que alude a valores emergentes o a valores de lo natural. Un resultado algo ambiguo que va en contra del sesgo pesimista derivado del abuso del mensaje del riesgo climático.

En su paso por la ciudad el río Bogotá se va marchitando, va perdiendo servicios ecosistémicos y la capacidad de generar bienestar ambiental.

Agradecimientos

Este trabajo fue financiado en su totalidad por el Jardín Botánico de Bogotá José Celestino Mutis. Se realizó con el apoyo de los estudiantes de prácticas de la Universidad de Ciencias Ambientales UDCA: Samuel Mera Ferreira y Julián Stivens Gómez Melo, estudiantes de la carrera de la carrera de Ciencias Ambientales.

Referencias

Ainscough, J. de Vries Lentsch, A., Metzger, M., Rounsevell, M., Schröter, M., Delbaere, B. de Groot, R. & Staes J. (2019). Navigating pluralism: Understanding perceptions of the ecosystem services concept. *Ecosyst. Serv.*, 36. Doi:10.1016/j.ecoser.2019.01.004

Consejo de Estado (2014). Sentencia del río Bogotá. [En línea]. Consejo de Estado. [Consultada en marzo de 2022]. Disponible en: [En línea]. CAR. [Consultada en marzo de 2022]. Disponible en: https://www.consejodeestado.gov.co/documentos/boletines/141/AC/25000-23-27-000-2001-90479-01(AP).pdf

Corporación Autónoma Regional de Cundinamarca CAR (2006). Plan de Ordenación y Manejo de la Cuenca Hidrográfica del Río Bogotá. Elaboración del diagnóstico, prospectiva y formulación de la cuenca hidrográfica del río Bogotá. Subcuenca del Río Bogotá Sector Tibitoc-Soacha - 2120-10. [En línea]. CAR. [Consultada en marzo de 2022]. Disponible en: https://www.car.gov.co/uploads/files/5ac25b9d3b786.pdf

Corvalán, C., Hales, S. & McMichael, A.J. (Eds.) (2005). Ecosystems and human well-being: health synthesis, Millennium ecosystem assessment. World Health Organization, Genova, Swiza 53 p.

Costanza, R., d' Arge, R., de Groot, R., Farber, S., Grasso, M., Hannon, B., Limburg, K., Naeem, S., O'Neill, R. V., Paruelo, J., Raskin, R. G., Sutton, P. & van den Belt, M. (1997). The value of the world's ecosystem services and natural capital. *Nature*, 387(**6630**), 253–260. https://doi.org/10.1038/387253a0

Dejusticia (2019, junio 2). ¿En qué va la sentencia que pide medidas para contar al campesinado?. [En línea]. Centro de Estudios de Derecho, Justicia y Sociedad. [Consultada en junio de 2022]. Disponible en: https://www.dejusticia.org/asi-va-la-sentencia-que-pide-contar-al-campesinado/

Departamento Nacional de Estadística DANE(2018). ¿Dónde Estamos?. [En línea]. Gobierno de Colombia. [Consultada en junio de 2022]. Disponible en: `https://sitios.dane.gov.co/cnpv/#!/donde_estamos`

Díaz, S., Demissew, S., Carabias, J., Joly, C., Lonsdale, M., Ash, N., Larigauderie, A., Adhikari, J. R., Arico, S., Báldi, A., Bartuska, A., Baste, I. A., Bilgin, A., Brondizio, E., Chan, K. M., Figueroa, V. E., Duraiappah, A., Fischer, M., Hill, R. & Zlatanova, D. (2015). The IPBES Conceptual Framework-Connecting nature and people. *Current Opinion in Environmental Sustainability*, 14, 1–16. https://doi.org/10.1016/j.cosust.2014.11.002

Díaz, S., Pascual, U., Stenseke, M., Martín-López, B., Watson, R., Molnár, Z., Hill, R., Chan, K., Baste, I., Brauman, K., Polasky, S., Church, A., Lonsdale, M., Larigauderie, A., Leadley, P., van Oudenhoven, A., Plaat, F., Schröter, M., Lavorel, S. & Shirayama, Y. (2018). Assessing nature's contributions to people. *Science*, 359, 270–272. https://doi.org/10.1126/science.aap8826

Díaz, S., Setelle, J., Brondizio, E. S., Ngo, H. T., Gueze, M., Agard, J., Arneth, A., Balvanera, P., Brauman, K. A., Butchart, S. H. M., Chan, K. M. A., Garibaldi, L. A., Ichii, K., Liu, J., Subramanian, S. M., Midgley, G. F., Miloslavich, P., Molnar, Z., Obura, D. & Zayas, C. N. (2019). The global assessment report on biodiversity and ecosystem services. Summary for policy makers. IPBES secretariat. 56 p.

Dick, J., Turkelboom, F., Woods, H., Iniesta-Arandia, I., Primmer, E., Saarela, S.R., Bezák, P., Mederly, P., Leone, M., Verheyden, W., Kelemen, E., Hauck, J., Andrews, C., Antunes, P., Aszalós, R., Baró, F., Barton, D.N., Berry, P., Bugter, R., Carvalho, L., Czúcz, B., Dunford, R., Garcia Blanco, G., Geam*ănă*, N., Giuc*ă*, R., Grizzetti, B., Izakovičová, Z., Kertész, M., Kopperoinen, L., Langemeyer, J., Montenegro Lapola, D., Liquete, C., Luque, S., Martínez Pastur, G., Martin-Lopez, B., Mukhopadhyay, R., Niemela, J., Odee, D., Peri, P.L., Pinho, P., Patrício-Roberto, G.B., Preda, E., Priess, J., Röckmann, C., Santos, R., Silaghi, D., Smith, R., V*ă*dineanu, A., van der Wal, J.T., Arany, I., Badea, O., Bela, G., Boros, E., Bucur, M., Blumentrath, S., Calvache, M., Carmen, E., Clemente, P., Fernandes, J., Ferraz, D., Fongar, C., García-Llorente, M., Gómez-Baggethun, E., Gundersen, V., Haavardsholm, O., Kalóczkai, A., Khalalwe, T., Kiss, G., Köhler, B., Lazányi, O., Lellei-Kovács, E., Lichungu, R., Lindhjem, H., Magare, C., Mustajoki, J., Ndege, C., Nowell, M., Nuss Girona, S., Ochieng, J., Often, A., Palomo, I., Pataki, G., Reinvang, R., Rusch, G., Saarikoski, H., Smith, A., Soy Massoni, E., Stange, E., Vågnes Traaholt, N., Vári, A., Verweij, P., Vikström, S., Yli-Pelkonen, V. Zulian, G. (2018). Stakeholders' perspectives on the operationalisation of the ecosystem service concept: Results from 27 case studies. *Ecosyst. Serv.* 29, 552–565. https://doi.org/10.1016/j.ecoser.2017.09.015

Escalera-Reyes, J. (2018). ¿Servicios de los ecosistemas o en los socioecosistemas?: Una mirada crítica al marco de los servicios ecosistémicos desde la Antropología. En B. Santamarina Campos, A. Coca & O. Beltran (Eds.), Antropología ambiental: Conocimientos y prácticas locales a las puertas del Antropoceno (pp. 71–82). Icaria.

Esri (2017). Survey123. Estados Unidos de América.

Giesecke Sara Lafosse, M. P. (2020). Elaboración y pertinencia de la matriz de consistencia cualitativa para las investigaciones en ciencias sociales. *Desde el Sur*, 12(**2**), 397–417. https://doi.org/10.21142/DES-1202-2020-0023

Guiza-Suárez, L., Londoño-Toro, B. & Rodríguez-Barajas, C. D. (2015). La judicialización de los conflictos ambientales: un estudio del caso de la cuenca hidrográfica del Río Bogotá (CHRB), Colombia. *Rev. Int. Contam. Ambient.* 31(**2**), 195–209.

Himes, A. & Muraca, B. (2018). Relational values: the key to pluralistic valuation of ecosystem services. *Curr. Opin. Environ. Sustain.* 35, 1–7. https://doi.org/10.1016/j.cosust.2018.09.005

Huertas, C. & Corraliza, J. A. (2016). Resistencias psicológicas en la percepción del cambio climático. *Papeles de relaciones ecosociales y cambio global*, 136, 107–1019.

Kenter, J. O., O'Brien, L., Hockley, N., Ravenscroft, Fazey, L., Irvine, K. N., Reed, M. S., Christie, M., Brady, E., Bryce, R., Church, A., Cooper, N., Davies, A., Evely, A., Everard, M., Fish, R., Fisher, J. A., Jobstvogt, N., Molloy, C., Orchard-Webb, J., Ranger, S., Ryan, M., Watson, V. & Williams, S. (2015). What are shared and social values of ecosystems? *Ecol. Econ.*, 111, 86–99. https://doi.org/10.1016/j.ecolecon.2015.01.006

Lemkow, L. (2002). Sociología ambiental: pensamiento socioambiental y ecología social del riesgo. Icaria, Barcelona. 232 p.

Lin, Y.-P., Lin, W.-C., Li, H.-Y., Wang, Y.-C., Hsu, C.-C., Lien, W.-Y., Anthony, J. & Petway, J. R. (2017). Integrating Social Values and Ecosystem Services in Systematic Conservation Planning: A Case Study in Datuan Watershed. *Sustainability*, 9(**5**), 718. https://doi.org/10.3390/su9050718

Martín-López, B., García-Lorente, M., Palomo, I. & Montes, C. (2011). The conservation against development paradigm in protected areas: Valuation of ecosystem services in the Doñana social-ecological system (southwestern Spain). *Ecol. Econ.*, 70, 1481–1491. https://doi.org/10.1016/j.ecolecon.2011.03.009

Martín-López, B., Iniesta-Arandia, I., García-Llorente, M., Palomo, I., Casado-Arzuaga, I., García del Amo, D., Gómez-Baggethun, E., Oteros-Rozas, E., Palacios-Agundez, I., Willaarts, B., González, J.A., Santos-Martín, F., Onaindia, M., López-Santiago, C. & Montes, C.(2012). Uncovering Ecosystem Service Bundles through Social Preferences. *PLoS ONE* 7, e38970. https://doi.org/10.1371/journal.pone.0038970

Millennium Ecosystem Assessment MEA (2005). Ecosystems and human well-being: synthesis. Island Press, Washington, DC. 43 p.

Morales-Vallejo, P. (2008). Estadística aplicada a las ciencias sociales. Universidad Pontificia de Comillas, Madrid. 364 p.

Naciones Unidas ONU, (2000). 55/2 Declaración del milenio. [En línea]. Organización de las Naciones Unidas [Consultada en marzo de 2022]. Disponible en: https://www.un.org/spanish/milenio/ares552.pdf

Nahlik, A. M., Kentula, M. E., Fennessy, M. S. & Landers, D. H.(2012). Where is the consensus? A proposed foundation for moving ecosystem service concepts into practice. *Ecol. Econ.* 77, 27–35. https://doi.org/10.1016/j.ecolecon.2012.01.001

Observatorio Regional Ambiental y de Desarrollo Sostenible del Río Bogotá ORARBO. (2022). Directorio Actores Ambientales-Observatorio Regional Ambiental y de Desarrollo Sostenible del Río Bogotá. [En línea]. gov.co. [Consultada en junio de 2022]. Disponible en: http://www.orarbo.gov.co/es/directorio-actores-ambientales

Orozco-Roa, P. A. (2016). Alternativas para el manejo de aguas pluviales en medios urbanos. Estudio de caso: Implementación y manejo de los canales pluviales en las cuencas del Salitre y Tintal en el marco del proceso de recuperación Río Bogotá 2000- 2014 (Tesis de pregrado en Gestión y Desarrollo Urbanos). Bogotá D.C., Universidad Colegio Mayor de Nuestra Señora de Cundinamarca, Facultad de Ciencia Política y Gobierno, Bogotá, D.C. 70 p.

Osorio-Osorio, J.A., (2007). El Río Tunjuelo en la historia de Bogotá: 1900 - 1990, 1. ed. ed. Alcaldía Mayor de Bogotá, Bogotá. 112 p.

Osorio-Osorio, J. A. (2008). La historia del agua en Bogotá: una exploración bibliográfica sobre la cuenca del rio Tunjuelo, en el siglo XX. *Mem. Soc.*, 12(**25**), 107–116.

Pascual, U., Balvanera, P., Díaz, S., Pataki, G., Roth, E., Stenseke, M., Watson, R. T., Başak Dessane, E., Islar, M., Kelemen, P. H., Maris, V., Quaas, M., Subramanian, S., Wittmer, H., Adlan, A., Ahn, A., Al-Hafedh, Y. S., Amankwah, E., Asah, S. T. & Yagi, N. (2017). Valuing nature's contributions to people: The IPBES approach. *Current Opinion in Environmental Sustainability*, 26–27, 7–16. https://doi.org/10.1016/j.cosust.2016.12.006

Páramo Morales, D. (2015). Editorial: La teoría fundamentada (Grounded Theory), metodología cualitativa de investigación científica. *Revista científica Pensamiento y Gestión*, 39, i–xi. https://doi.org/10.14482/pege.39.8439

Pereira-Prado, M. M. (2015). Las áreas verdes urbanas como generadoras de ecoservicios para el bienestar humano. Propuesta de gestión de parques para la Localidad de Engativá (Tesis de Maestría). Bogotá D.C., Pontificia Universidad Javeriana, Facultad de Estudios Ambientales y Rurales, Bogotá, D.C. 189 p.

Raymond, C. M. & Kenter, J.O. (2016). Transcendental values and the valuation and management of ecosystem services. *Ecosyst. Serv.*, 21, 241-257. https://doi.org/10.1016/j.ecoser.2016.07.018

Rincón-Ruiz, A., Echeverry-Duque, M., Piñeros, A. M., Tapia, C.H., David, A., Arias-Arévalo, P. & Zuluaga, P. A.(2014). Valoración integral de la biodiversidad y los servicios ecosistémicos: aspectos conceptuales y metodológicos. Instituto de Investigación de Recursos Biológicos Alexander von Humboldt, Bogotá, D.C., Colombia. 150 p.

Rojas Tejada, A. J., Fernández Prados, J. S. & Pérez Meléndez, C. (1998). Investigar mediante encuestas: Fundamentos técnicos y aspectos prácticos. Síntesis. 224 p.

RStudio (2019). RStudio. Boston.

Schwartz, S.H. & Bilsky, W. (1987). Toward a universal psychologyval structure of human values. *J. Personal. Soc. Psycology*, 53(*3*), 550–562.

Stocknes, P. E. (2014). Rethinking climate communications and the psychological climate paradox. *Energy Research & Social Science*, 1, 161–170. https://doi.org/10.1016/j.erss.2014.03.009

Sukhdev, P., Wittmer, H. & Miller, D. (2014). The economics of ecosystems and biodiversity (TEEB) challenges and responser, in: Helm, D., Hepburn, C. (Eds.), Nature in the Balance: The Economics of Biodiversity. Oxford University, Oxford. 16 p.

Tuan, Y. (1974). Topofilia: Un estudio de las percepciones, actitudes y valores sobre el entorno. Prentice-Hall. 351 p.

Taylor, S. J. & Bogdan, R. (1996). Introducción a los métodos cualitativos de investigación: La búsqueda de significados. Paidós. 343 p.

van der Hammen, T. (1998). Plan ambiental de la Cuenca Alta del río Bogotá: análisis y orientaciones para el ordenamiento territorial. Corporación Autónoma Regional de Cundinamarca CAR, Bogotá D.C. 142 p.

Vilardy-Quiroga, S. P. & González-Nóvoa, P. A. (Eds.)(2011). Repensando la Ciénaga: nuevas miradas y estrategias para la sostenibilidad en la Ciénaga Grande de Santa Marta. Univ. del Magdalena [u.a.], Santa Marta, Col. 226 p.