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

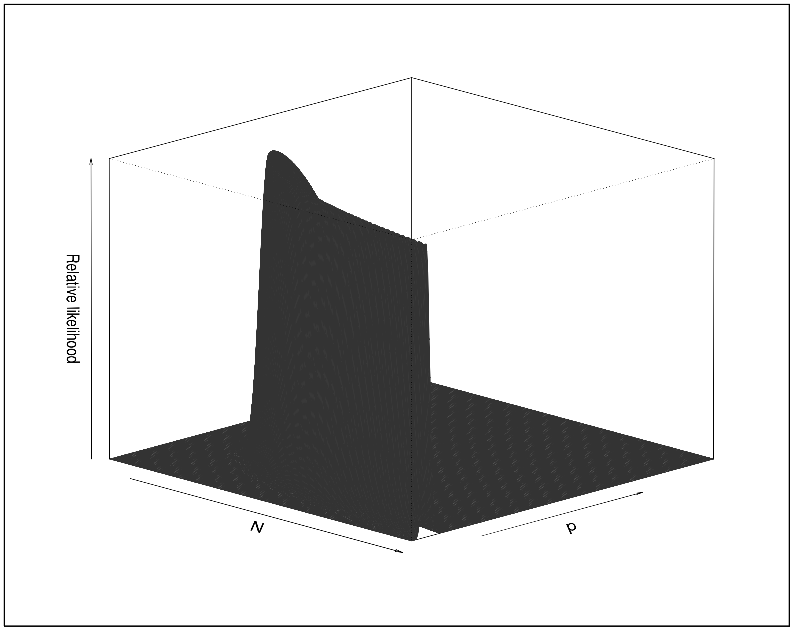

Figure 1: Binomial likelihood surface for the sample $\boldsymbol{x} = (16, 18, 22, 25, 27)$. Source: Elaborated by the author.

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

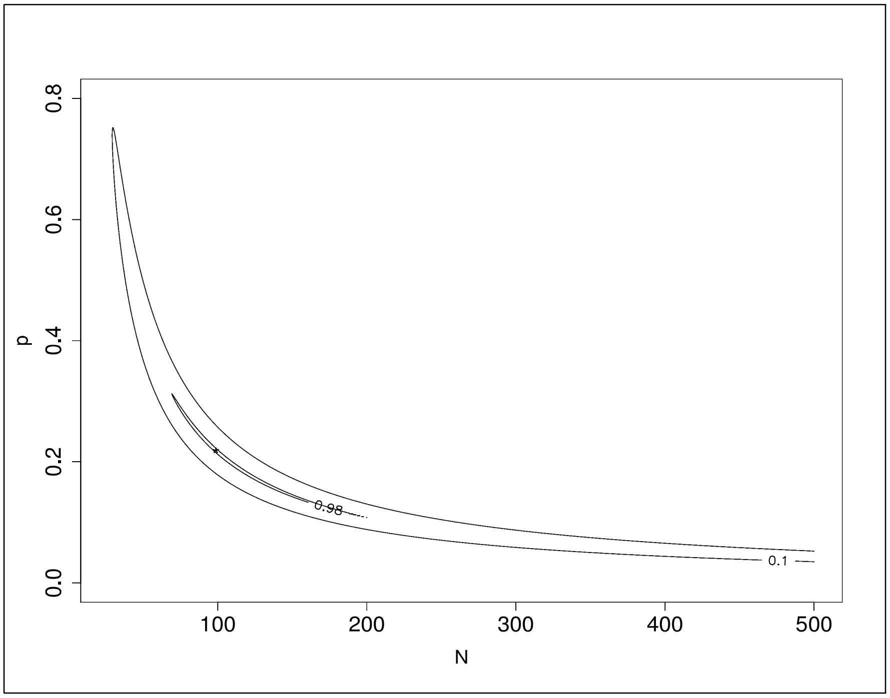

Figure 2: Contour plot of the relative likelihood surface given in Figure 1. The MLE is marked with an asterisk. Source: Elaborated by the author.

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.

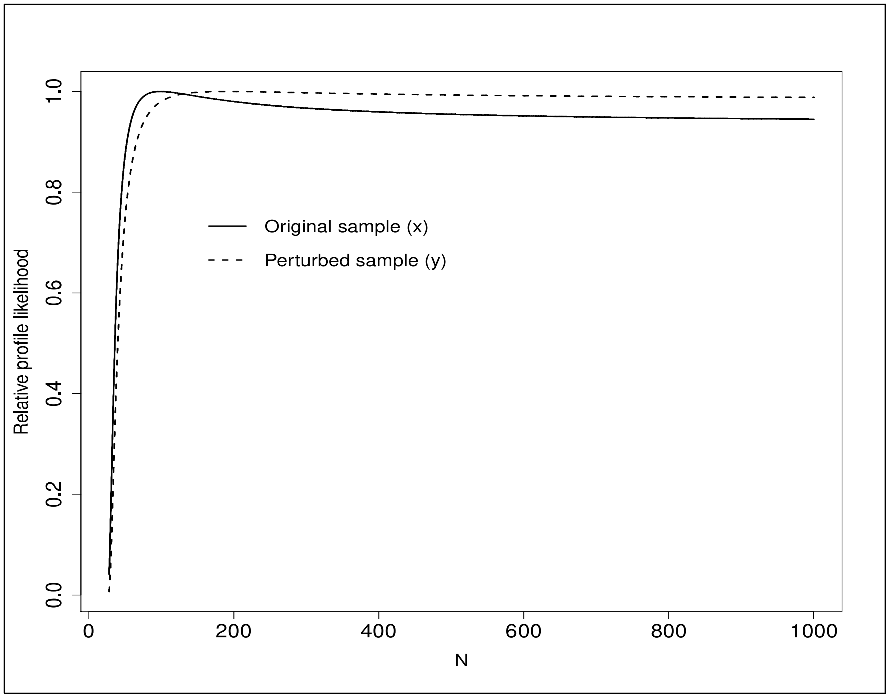

Figure 3: The relative profile likelihood function of N for the original and perturbed sample, $\boldsymbol{x} = (16, 18, 22, 25, 27)$ and $\boldsymbol{y} = (16, 18, 22, 25, 28)$ respectively. Source: Elaborated by the author.

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:

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

where

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

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).

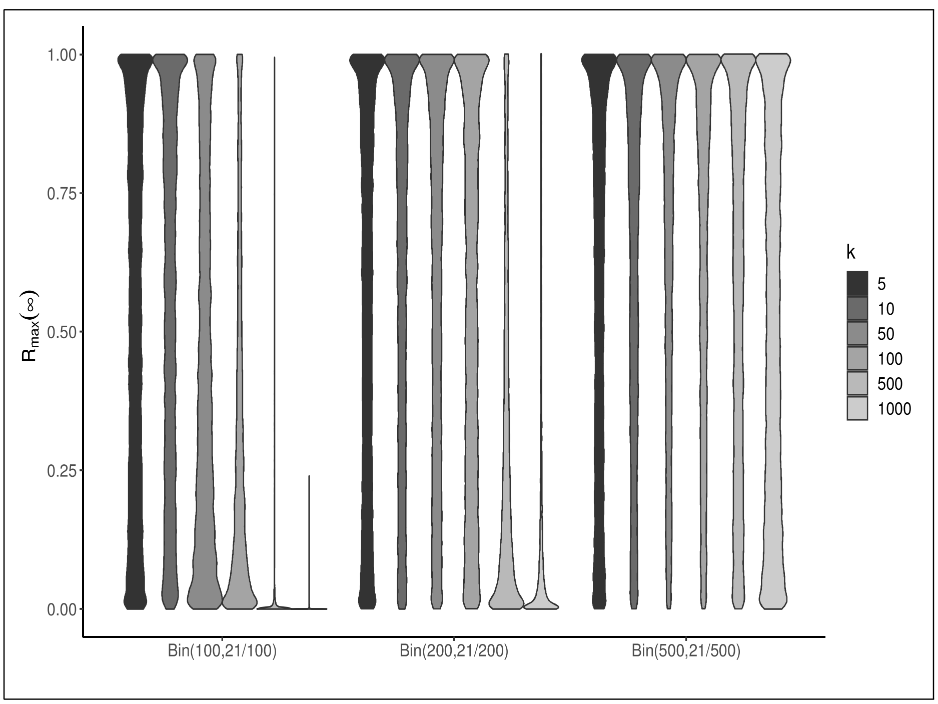

Figure 4: Violin plots for binomial simulation scenarios where $\widehat{\mu} > \widehat{\sigma}^2$. Source: Elaborated by the author.

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.

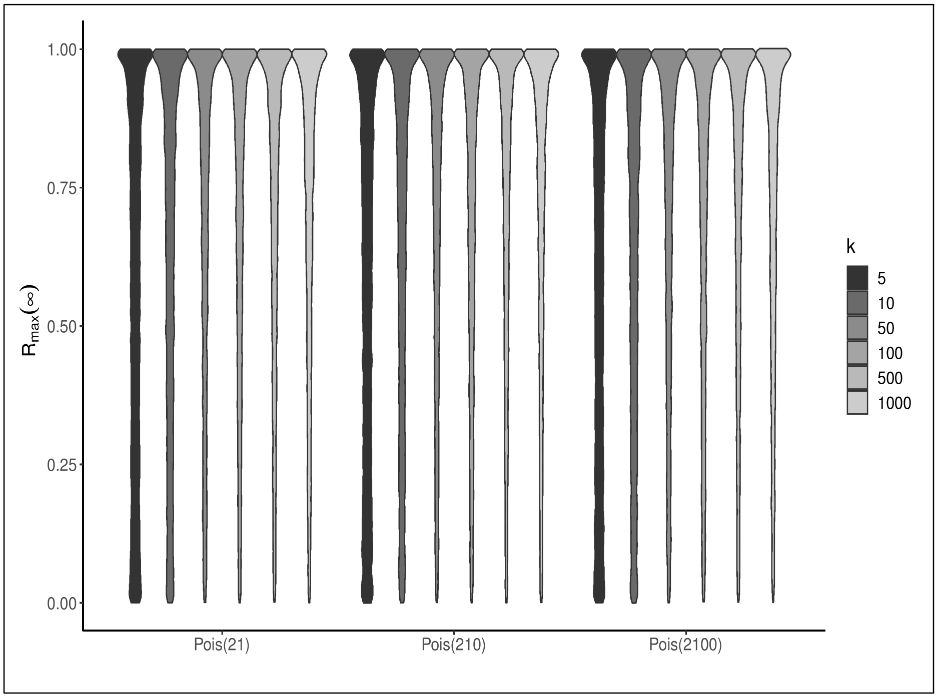

Figure 5: Violin plots for Poisson simulation scenarios where $\widehat{\mu} > \widehat{\sigma}^2$. Source: Elaborated by the author.

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 $0 \leq p \leq 1$ and $N \geq r_k$, 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.

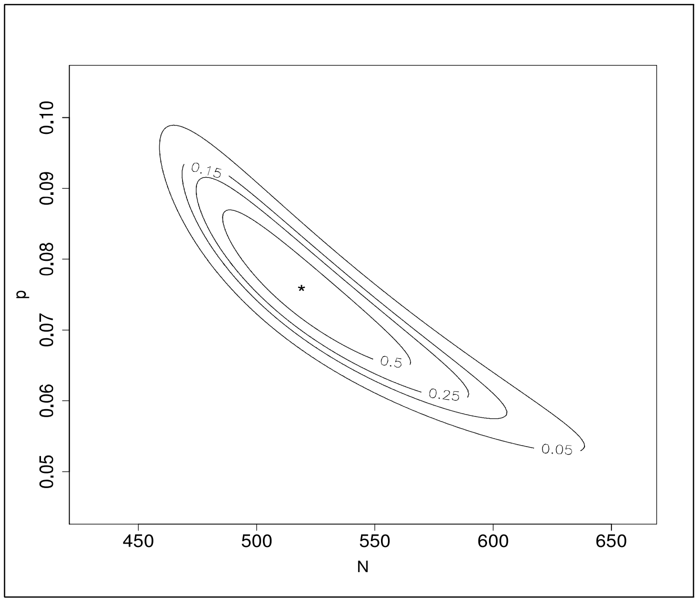

Figure 6: Likelihood contours as a function of N and p for $k = 18$, $r_k = 394$ and $S = 4544$. The MLE is marked with an asterisk. Source: Elaborated by the author.

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.

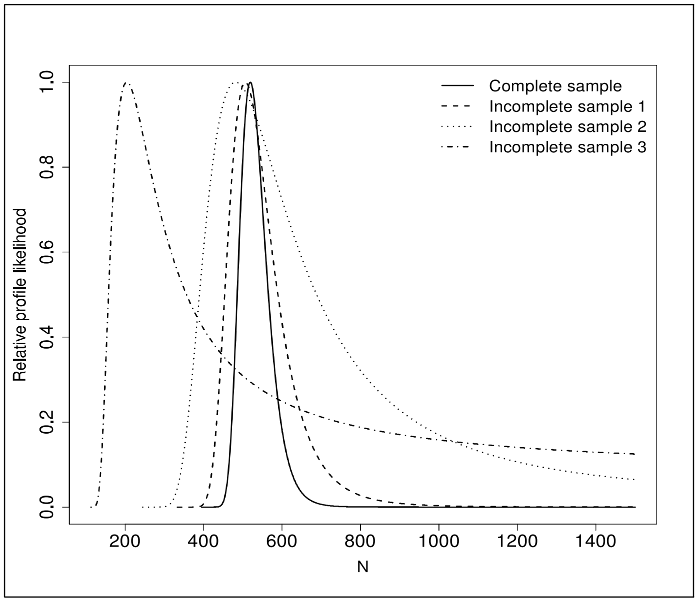

Figure 7: Relative profile likelihood functions of N for the complete sample ($k = 18$, $r_k = 394$, $S = 4544$) and the incomplete samples presented in Table 3. Source: Elaborated by the author.

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.