THE ESTIMATION OF $p \geq q$ FROM TWO INDEPENDENT BINOMIAL SAMPLES $b(n,p)$ AND $b(m,q)$

ELOÍSA DÍAZ-FRANCÉS AND DAVID A. SPROTT

ABSTRACT. The estimation of the probability of success p from independent binomial samples $b(p,n)$ and $b(q,m)$ where $p \geq q$ is considered using a likelihood approach.

1. INTRODUCTION

Consider the model $\mathcal{B}$ of two independent binomial samples $x \sim b(n,p)$, $y \sim b(m,q)$ under the restriction $p \geq q$. The problem is to make quantitative statements about p, based on the information contained in both samples.

The focus of previous approaches to this problem has been on point estimates, for instance the maximum likelihood estimate (mle), and their optimal properties, such as mean squared error (mse), bias, variance, admissible and minimax estimators, the use of certain loss functions, etc. See, for example, Robertson and Waltman (1968), Sackrowitz (1970), Johnson (1971), and Hengartner (1999). One conclusion of these approaches has been that in order to have higher precision for estimating p, depending on the value of q, y should be discarded.

Here the consequences of basing inferences about p on the whole observed likelihood function based on both samples are examined. The use of the whole likelihood function, and not just its mle, has been increasingly used in problems of estimation since Fisher (1956, p. 73), (e.g. Barnard, Jenkins and Winston (1963), Edwards (1992), Sprott (2000)). The purpose is to make quantitative statements of uncertainty about unknown parameters based on all of the sample information. "The likelihood supplies a natural order of preference among the possibilities under consideration", Fisher (1956, p. 73). For a single parameter the results can usually be given in the form of graph of the likelihood function supplemented by a set of nested likelihood intervals – see Sprott (2000, Section 2.8) – as in Figure 1.

2. The Profile Likelihood Function

Under the binomial model $\mathcal{B}$, the joint likelihood function based on the two independent samples is

$$L_{\mathcal{B}} \propto \underbrace{p^x(1-p)^{n-x}}_{\text{First Sample}} \underbrace{q^y(1-q)^{m-y}}_{\text{Second Sample}}, \text{ where } p \geq q.$$

For this model, the marginal distribution of x does not provide a marginal likelihood of p since the range of p depends on q. Therefore in the absence of knowledge of q the maximized or profile likelihood of p under the restricted model $\mathcal{B}$ will be considered, which is

$$L_M(p; x, y) \propto L_{\mathcal{B}}[p, \hat{q}(p); x, y],$$

where $\hat{q}(p)$ is the restricted mle of q for a specified value of p,

$$\hat{q}(p) = \min(\hat{q}, p) = \begin{cases} \hat{q}, & \text{if } p \geq \hat{q} \\ p, & \text{if } p < \hat{q} \end{cases},$$

so that $\hat{q}(p) \leq p$.

Therefore the profile likelihood of p is

$$L_M(p; x, y) \propto \begin{cases} L_A(p; x, y) = p^x(1-p)^{n-x}\, p^y(1-p)^{m-y} & \text{if } p < \hat{q} \\ L_B(p; x, y) = p^x(1-p)^{n-x}\, \hat{q}^y(1-\hat{q})^{m-y} & \text{if } p \geq \hat{q} \end{cases}. \quad (2.1)$$

The global maximum of $L_M(p; x, y)$ is at $\hat{p}^*$,

$$\hat{p}^* = \begin{cases} (x+y)/(n+m) & \text{if } \hat{p} < \hat{q} \\ x/n & \text{if } \hat{p} \geq \hat{q} \end{cases}. \quad (2.2)$$
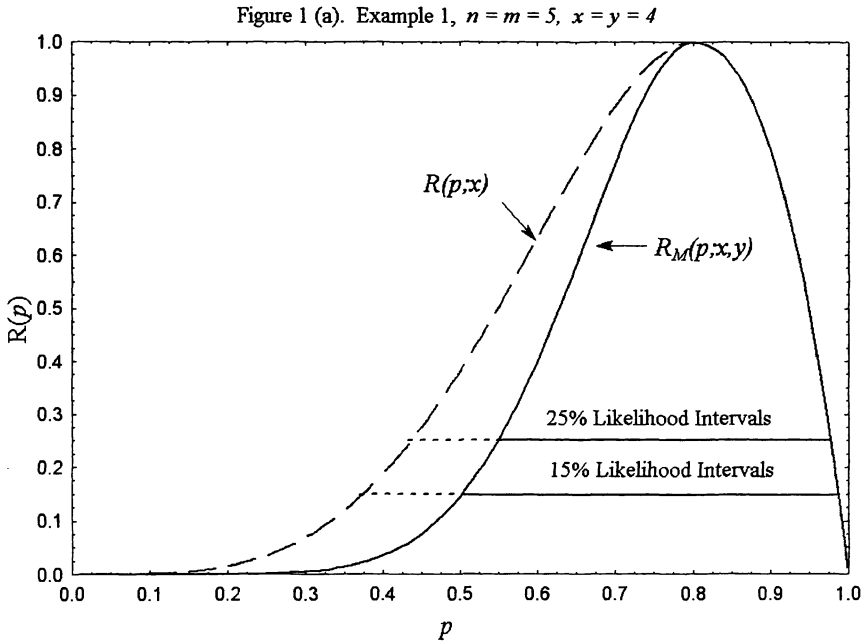
The relative profile likelihood function $R_M(p) = L_M(p)/L_M(\hat{p}^*)$ is $L_M(p)$ standardized to be one at its maximum.

The likelihood ratio LR,

$$LR = L_M(\hat{p}^*; x, y)/L_{\mathcal{G}}(\hat{p}, \hat{q}; x, y)$$

$$= \begin{cases} L_A\left[(x+y)/(n+m); x, y\right] / L_{\mathcal{G}}(\hat{p}, \hat{q}; x, y) & \text{if } \hat{p} < \hat{q} \\ L_B(x/n; x, y) / L_{\mathcal{G}}(\hat{p}, \hat{q}; x, y) = 1 & \text{if } \hat{p} \geq \hat{q} \end{cases}, \quad (2.3)$$

can be used to assess the restricted model $\mathcal{B}$ relative to the general model $\mathcal{G}$ of two independent binomial samples, where

$$L_{\mathcal{G}} = p^x(1-p)^{n-x}q^y(1-q)^{m-y}, \quad (2.4)$$

Figure 1 (a). Example 1, $n = m = 5$, $x = y = 4$

where no restrictions are imposed on either p or q.

Whenever $\hat{p} < \hat{q}$, the data can provide evidence against the restricted model as the distance between $\hat{p}$ and $\hat{q}$ increases. In contrast, if $\hat{p} \geq \hat{q}$, the data cannot give evidence against the restricted model, since $LR = 1$. The quantity LR can be interpreted in terms of how less probable are x, y under $\mathcal{B}$ than under $\mathcal{G}$. Another interpretation is in terms of $P(-2 \log LR \geq -2 \log LR_o)$, where LR_o is the observed value of (2.3). The likelihood ratio statistic $-2 \log LR$ will not have the χ^2 distribution. But its exact distribution is easily calculated with modern computing power by enumerating from (2.3) all samples $(x, y | n, m)$ for which $LR \leq LR_o$. The resulting test of the model $\mathcal{B}$ against $\mathcal{G}$ is

$$P = \max_{p \geq q} \left[\sum_{LR \leq LR_o} \binom{n}{x} p^x (1-p)^{n-x} \binom{m}{y} q^y (1-q)^{m-y} \right]. \qquad (2.5)$$

This will be illustrated in the following examples.

3. Examples

Example 1: If $\hat{p} \geq \hat{q}$, then from (2.2) the maximum of $L_M(p; x, y)$ occurs in $L_B(p; x, y)$ at $\hat{p} = x/n$, and is $\hat{p}^x(1-p)^{n-x}\hat{q}^y(1-\hat{q})^{m-y}$. This facilitates comparing $R_M(p; x, y)$ with $R(p; x)$, the binomial relative likelihood of p based on x alone. Under this condition, $R_M(p; x, y)$ is given by

$$\begin{cases} R_A(p; x, y) &= p^{x+y}(1-p)^{n+m-x-y}/\hat{p}^x(1-\hat{p})^{n-x}\hat{q}^y(1-\hat{q})^{m-y} & \text{if } p < \hat{q} \\ R_B(p; x, y) &= p^x(1-p)^{n-x}/\hat{p}^x(1-\hat{p})^{n-x} & \text{if } p \geq \hat{q} \end{cases}.$$

In the range $p \geq \hat{q}$ the component $R_B(p; x, y)$ of $R_M(p; x, y)$ is the same as the ordinary relative likelihood function $R(p; x)$ of p based on x alone obtained from $L_{\mathcal{G}}$, (2.4). In the range $p < \hat{q}$ the remaining component $R_A(p; x, y) \leq R(p; x)$, so that the function $R_M(p; x, y)$ lies entirely within the function $R(p; x)$ in this range, indicative of the higher precision of $R_M(p; x, y)$ in this range. These facts are exemplified in Figure 1(a) with $x = 4, y = 4$, $n = m = 5$, so that $\hat{p} = \hat{q} = .8$.

If $\hat{q}$ is sufficiently small then $R_M(p; x, y) = R_B(p; x)$ over practically the entire range of p, and hence equal to $R(p; x)$. For example if $x = 4$, $y = 1$, $n = m = 5$, the range of R_B is $p \geq .2$ and $R_M(p)$ is indistinguishable from $R(p)$ as shown in Figure 1(b).

From (2.3) $-2\log LR = 0$ so that there is no evidence against the model $\mathcal{B}$ in this situation.

Example 2: The remaining case is that of $\hat{p} < \hat{q}$ in (2.2). Here the maximum occurs in L_A, so that comparisons with $R(p; x)$ are complicated. Their maxima occur at different places, so that the likelihoods have more complicated differences. Also this is the case where the data can present evidence against the model $\mathcal{B}$. These facts are exemplified in Figure 2 with $x = 1$, $y = 4$, $m = n = 5$, so that $\hat{p} = 0.2 < \hat{q} = 0.8$, and $\hat{p}^* = 0.5$. Here $LR_o = 0.1455$. This means that the maximum probability of $x = 1, y = 4|m = n = 5$ under model $\mathcal{B}$ is only 0.1455 of the maximum probability under the general model $\mathcal{G}$. Thus $\mathcal{B}$ is somewhat implausible relative to $\mathcal{G}$. Using $-2\log LR_o = 3.8552$ in (2.5) gives $P = 0.055$, the exact significance level obtained with $p = q = 0.5$. Based on this the observations contradict model $\mathcal{B}$ at the 0.055 level of significance. Thus the inferences about p based on $\mathcal{B}$ would be suspect. This suggests reverting to the general model $L_{\mathcal{G}}$ in which y contains no information about

Figure 1 (b). Example 1, $n = m = 5$, $x = 4$, $y = 1$
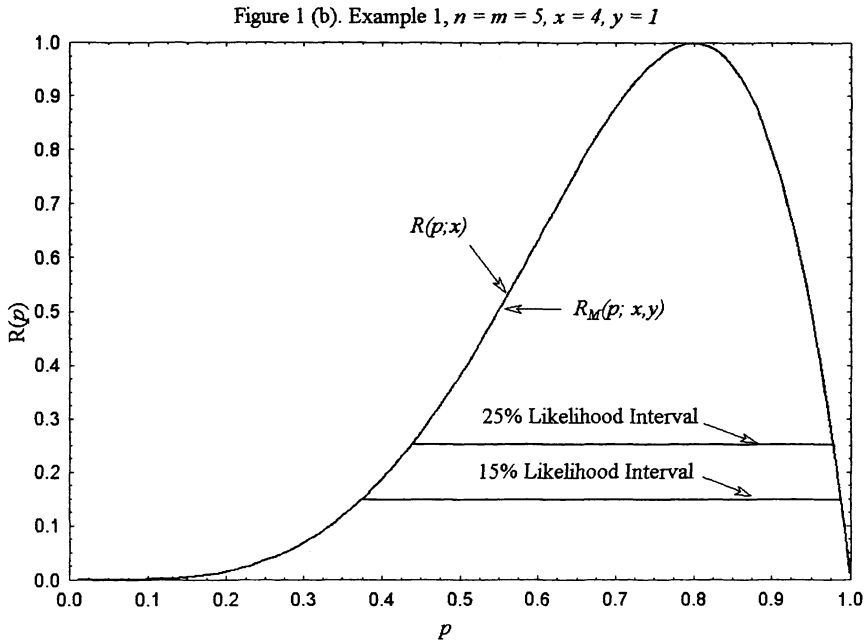
p. It is interesting that the $\chi^2_{(1)}$ approximation applied to the distribution of $-2 \log LR$ in this example gives $P = 0.0496$,. This is perhaps surprising since $-2 \log LR$ is not supposed to have the χ^2 distribution.

4. Some Conclusions

(1) The likelihood function to be used is determined by the model that is selected based on scientific considerations. In model $\mathcal{B}$, x and y both contribute to the estimation of p. The relative contribution of each is determined automatically by the profile likelihood function.

(2) More generally, as shown in the Figures, these likelihood functions are often too intricate in shape to be described by point estimates and their properties such as mse. The latter are not adequate to make informative quantitative estimation statements using all of the sample information. In contrast, the purpose of the likelihood function is to make such statements.

(3) The likelihood function also leads to an assessment of the legitimacy of the underlying assumption that $p > q$, for example by using the likelihood ratio test (2.5).
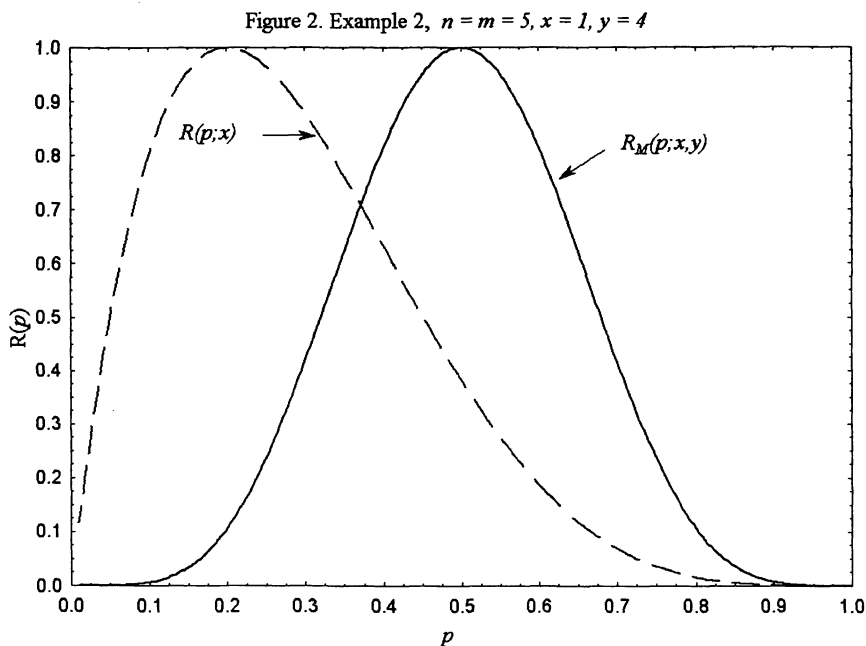
Figure 2. Example 2, $n = m = 5, x = 1, y = 4$

REFERENCES

[1] Barnard, G. A., Jenkins, G. M. and Winston, C. B. (1963). Likelihood inference and time series (with discussion). *Journal of the Royal Statistical Society* A 125, 321-372.

[2] Edwards, A. W. F. (1992). *Likelihood.* The Johns Hopkins University Press: Baltimore

[3] Fisher, R.A. (1956). *Statistical Methods and Scientific Inference.* Oliver and Boyd: Edinburgh.

[4] Hengartner, N. W. (1999). A Note on Maximum Likelihood Estimation. *American Statistician* 53, 123-125.

[5] Johnson, B. (1971). On the Admissible estimators for Certain Fixed Sample binomial Problems. *Annals of Mathematical Statistics* 42, 1579-1587.

[6] Robertson, T. and Waltman, P. (1968). On Estimating Monotone Parameters. *Annals of Mathematical Statistics* 39, 1030-1039.

[7] Sackrowitz, H. (1970). Estimation for Monotone Parameter Sequences: The Discrete Case. *Annals of Mathematical Statistics* 41, 609-620.

[8] Sprott, D. A. (2000). *Statistical Inference in Science.* New York: Springer-Verlag.

ELOÍSA DÍAZ-FRANCÉS
CENTRO DE INVESTIGACIÓN
EN MATEMÁTICAS
GUANAJUATO, MÉXICO.

diazfran@cimat.mx

DAVID A. SPROTT
CENTRO DE INVESTIGACIÓN
EN MATEMÁTICAS
GUANAJUATO, MÉXICO.
AND
UNIVERSITY OF WATERLOO
ONTARIO, CANADA.

sprott@cimat.mx