IMPROVING ON THE MLE OF p FOR A BINOMIAL(n,p) WHEN p IS AROUND $\frac{1}{2}$

FRANÇOIS PERRON
Université de Montréal

We consider the problem of estimating the parameter p of a binomial (n,p) distribution when p lies in a symmetric interval of length $m/\sqrt{n}$, $m < \sqrt{n}$. We establish sufficient conditions for the domination of the maximum likelihood estimator with quadratic loss. We suggest three other estimators for the estimation of p. The first two dominate the maximum likelihood estimator. The first one comes from Moors (1985) and corresponds to the bayesian estimator with respect to the symmetric prior concentrated on the end points if and only if $m \leq 1$ when n is odd or $m \leq \sqrt{n/(n-1)}$ when n is even. The second estimator comes from Charras and van Eeden (1991); it is in fact the maximum likelihood estimator for the problem where m is replaced by m_0, $0 < m_0 < m$. We give an algorithm for the selection of m_0. The third is the Bayes estimator with respect to the prior having a density proportional to $\left(p(1-p)\right)^{-1}$. This estimator dominates the maximum likelihood estimator for some values of (n,m) but not for all of them. We give simple sufficient conditions for the domination of the Bayes estimator over the maximum likelihood estimator. It is clear that the maximum likelihood estimator is inappropriate when either n or m is small. When n is large, all of the estimators have approximately the same behaviour except for the last. Numerical evaluations illustrate our comments.

1. Introduction

Assume that the statistician observes x, the realisation of X, a binomial(n,p) random variable. Consider the problem of estimating the proportion p with quadratic loss when p lies in a symmetric interval around $\frac{1}{2}$. In many situations, prior knowledge tells us that this symmetric interval has length less than 1 simply because successes and failures are not rare events. For example, in an effort to protect the privacy of the respondant, Warner (1965) has developed a method where one is interested in the estimation of π and $p = \pi P + (1-\pi)(1-P)$. In his setup, P is known, $\frac{1}{2} < P < 1$ and the distribution of X is a binomial(n,p), therefore, $1 - P \leq p \leq P$. In the following, m will be fixed, $0 < m < \sqrt{n}$, and we shall set $p = (1 + \theta/\sqrt{n})/2$, $\Theta(m) = \{\theta \in \mathbb{R} : \|\theta\| \leq m\}$.

In this problem, the maximum likelihood estimator δ_{mle} is the truncation of the empirical proportion (x/n) on the parameter space. It is given by $\delta_{\text{mle}}(x) = \{1 + [|2x/n - 1| \wedge (m/\sqrt{n})]\,\text{sgn}(2x/n - 1)\}/2$. This estimator is inadmissible because it takes values on the boundary of the parameter space, (see, Sacks, 1963, DasGupta 1985 or Charras and van Eeden, 1991). Actually, Charras and van Eeden (1991) specifically treat our problem in their Example 5.2. They propose that we modify the maximum likelihood

AMS subject classifications: 62F10, 62F15, 62F30.