MONOTONE CONVERGENCE OF BINOMIAL PROBABILITIES WITH AN APPLICATION TO MAXIMUM LIKELIHOOD ESTIMATION[1]

By Kumar Jogdeo

Courant Institute of Mathematical Sciences, New York University

0. Introduction and summary. It has been pointed out in the literature that the maximum likelihood (ML) estimator may be misleading in the presence of prior information. Many of these examples assume extreme sizes: one or infinity. In the present paper an example is considered where sample size may be any odd positive integer. This example is an amplification of the one given by Lehmann (1949) where he considers estimation of the probability of "success" p, based on a single observation of a Bernoulli random variable X. He states that with the prior information $\frac{1}{3} \leqq p \leqq \frac{2}{3}$ the ML estimator has uniformly larger expected squared error than any estimator $\delta(X)$ which is symmetric about $\frac{1}{2}$ and is such that $\frac{1}{3} \leqq \delta(0) \leqq \frac{1}{2} \leqq \delta(1) \leqq \frac{2}{3}$. In particular, the ML estimator is uniformly worse than the trivial estimator $\delta(X) \equiv \frac{1}{2}$. A natural question arises: does the same phenomenon occur for larger samples? In the following it has been shown that with $(2n + 1)$ observations if p is known to be in a small interval around $\frac{1}{2}$ then the trivial estimator is uniformly better than the ML estimator [now] based on $(2n + 1)$ observations. The interval having this property shrinks as n becomes large. The proof is based on a monotone convergence of certain binomial probabilities which itself may be of some interest.

1. Results. Throughout this section, S_{2n+1} will denote the sum of $(2n + 1)$ random variables and the probability of a success p will appear through either of the notations $P[\cdots \mid p]$ or $P_p[\cdots]$.

The problem of interest is to study the behavior of the ML estimator when p is restricted to the interval

$$\frac{1}{2} - \theta \leqq p \leqq \frac{1}{2} + \theta. \tag{1.1}$$

Due to the monotone character of the likelihood function on either side of its maximum, it is easy to verify that if

$$0 < \theta \leqq \frac{1}{2}(2n + 1), \tag{1.2}$$

then the ML estimator takes only two values

$$\begin{aligned} \hat{p}(S_{2n+1}) &= \tfrac{1}{2} + \theta \quad \text{if} \quad S_{2n+1} \geqq n + 1, \\ &= \tfrac{1}{2} - \theta \quad \text{if} \quad S_{2n+1} \leqq n. \end{aligned} \tag{1.3}$$

Suppose that the true value of p is $\frac{1}{2} + \lambda$ where $0 < \lambda \leqq \theta$. Then the expected

Received 6 March 1967.

[1] Research supported by Ford Foundation Grant for probability and statistics.

squared error of the ML estimator is

$$(1.4)\quad \begin{aligned} E[\hat{p}(S_{2n+1}) - \tfrac{1}{2} - \lambda]^2 &= (\theta - \lambda)^2 P[S_{2n+1} \geqq n + 1 \mid \tfrac{1}{2} + \lambda] \\ &\quad + (\theta + \lambda)^2 P[S_{2n+1} \leqq n \mid \tfrac{1}{2} + \lambda] \\ &= \theta^2 + \lambda^2 + 2\lambda\theta\{1 - 2P[S_{2n+1} \mid \tfrac{1}{2} + \lambda]\}. \end{aligned}$$

Note that the error given by (1.4) is unaltered if the true value were $(\frac{1}{2} - \lambda)$.

The expected squared error given above will be greater than λ^2, the error for the trivial estimator $\delta(S_{2n+1}) \equiv \frac{1}{2}$, throughout the interval $[\frac{1}{2} - \theta, \frac{1}{2} + \theta]$, provided

$$(1.5)\qquad \theta \geqq 2\lambda\{2P[S_{2n+1} \geqq n + 1 \mid \tfrac{1}{2} + \lambda] - 1\}, \qquad \lambda \leqq \theta,$$

which will certainly hold if

$$(1.6)\qquad P[S_{2n+1} \geqq n + 1 \mid \tfrac{1}{2} + \lambda] \leqq \tfrac{3}{4}, \qquad \lambda \leqq \theta.$$

The left side of (1.6) reaches its maximum at the highest possible value of λ viz. θ. On one hand the restriction $\theta \leqq \frac{1}{2}(2n + 1)$ is needed to keep ML estimator two valued and hence simple, however at the same time it is desirable to make the interval as wide as possible. This suggests $\theta = \frac{1}{2}(2n + 1)$.

THEOREM. *If the probability p of observing "1" in Bernoulli trials is restricted to the interval*

$$(1.7)\qquad \tfrac{1}{2} - 1/2(2n + 1) \leqq p \leqq \tfrac{1}{2} + 1/2(2n + 1),$$

where $n \geqq 1$ is a positive integer, then the ML *estimator based on $(2m + 1)$ observations, with $m \leqq n$, has uniformly larger expected squared error than the trivial estimator $\delta \equiv \frac{1}{2}$.*

It is clear from the preceding discussion that the theorem is true if the inequality (1.6) holds for $\theta = \frac{1}{2}(2n + 1)$. The following lemma is a much more precise result than the needed inequality.

LEMMA. *If S_{2n+1} denotes the sum of $(2n + 1)$ Bernoulli random variables then*

$$(1.8)\qquad P[S_{2n+1} \geqq n + 1 \mid (n + 1)/(2n + 1)] \downarrow \tfrac{1}{2} \text{ as } n \to \infty.$$

Here $\downarrow$ indicates monotone decreasing convergence.

PROOF. Throughout the proof n is an arbitrary but fixed integer larger than unity and

$$(1.9)\qquad \xi = (n + 1)/(2n + 1), \qquad \eta = n/(2n - 1).$$

The notation $P_\xi[\cdots]$ will be used instead of $P[\cdots \mid \xi]$.

The convergence to $\frac{1}{2}$ is well known and it remains to prove that

$$(1.10)\qquad P_\eta[S_{2n-1} \geqq n] - P_\xi[S_{2n+1} \geqq n + 1] > 0.$$

Considering $(2n + 1)$ trials as composed of two independent sets of $(2n - 1)$ trials and 2 trials it can be seen that

$$(1.11)\quad \begin{aligned} P_\xi[S_{2n+1} \geqq n + 1] &= P_\xi[S_{2n-1} \geqq n] \\ &\quad + \xi^2 P_\xi[S_{2n-1} = n - 1] - (1 - \xi)^2 P_\xi[S_{2n-1} = n]. \end{aligned}$$

Observing that

$$(1.12)\qquad \begin{aligned} \xi^2 P_\xi[S_{2n-1} = n-1] &- (1-\xi)^2 P_\xi[S_{2n-1} = n] \\ &= \binom{2n-1}{n}[\xi^{n+1}(1-\xi)^n - \xi^n(1-\xi)^{n+1}] \\ &= (1/(2n+1))\binom{2n-1}{n}\xi^n(1-\xi)^n, \end{aligned}$$

it suffices to show that

$$(1.13)\qquad P_\eta[S_{2n-1} \geqq n] - P_\xi[S_{2n-1} \geqq n] > (1/(2n+1))\binom{2n-1}{n}\xi^n(1-\xi)^n.$$

Since

$$(1.14)\qquad (d/dp)P_p[S_n \geqq k] = k\binom{n}{k}p^{k-1}(1-p)^{n-k}$$

it follows from the mean value theorem that for some ζ such that $\xi < \zeta < \eta$,

$$(1.15)\qquad \begin{aligned} P_\eta[S_{2n-1} \geqq n] &- P_\epsilon[S_{2n-1} \geqq n] \\ &= (\eta - \xi)\binom{2n-1}{n}n\zeta^{n-1}(1-\zeta)^{n-1} \\ &= (n/(2n+1)(2n-1)\binom{2n-1}{n}[\zeta(1-\zeta)]^{n-1}. \end{aligned}$$

Recalling (1.13), it remains to prove that

$$(1.16)\qquad ((2n-1)/n)[\xi(1-\xi)/\zeta(1-\zeta)]^{n-1}\xi(1-\xi) < 1.$$

Since $\frac{1}{2} < \xi < \zeta < \eta$ it is clear that $\xi(1-\xi) > \zeta(1-\zeta) > \eta(1-\eta)$ and

$$(1.17)\qquad \xi(1-\xi)/\zeta(1-\zeta) < \xi(1-\xi)/\eta(1-\eta) = 1 + 2/(2n-1)(2n+1)^2.$$

Further for $n > 1$,

$$(1.18)\qquad \begin{aligned} (n-1)\log[1 &+ 2/(n-1)(2n+1)^2] \\ &< (n-1)2/(n-1)(2n+1)^2 < \log 2, \end{aligned}$$

so that

$$(1.19)\qquad [\xi(1-\xi)/\zeta(1-\zeta)]^{n-1} < [\xi(1-\xi)/\eta(1-\eta)]^{n-1} < 2.$$

Using (1.19) together with

$$(1.20)\qquad ((2n-1)/n)\xi(1-\xi) = (2n-1)(n+1)/(2n+1)^2 < \tfrac{1}{2},$$

the inequality (1.16) follows. This completes the proof of the lemma.

2. Remarks. I. From (1.6) it follows that with a single observation the ML estimator is uniformly worse than the trivial estimator for the interval $[\frac{1}{4}, \frac{3}{4}]$.

II. When the sample size is an even positive integer, the ML estimator will take value $\frac{1}{2}$ with positive probability. However, this probability will be very small as n increases, enabling one to find an interval where the same behaviour of the ML estimator persists. For example, for the sample size of two,

$$(2.1)\qquad E[\hat{p}(S_2) - \tfrac{1}{2} - \lambda]^2 = \lambda^2 + \tfrac{1}{2}\theta^2 + 2\lambda^2\theta^2 - 4\theta\lambda^2, \qquad \lambda \geq \theta,$$

where p is restricted to $[(\frac{1}{2}) - \theta, (\frac{1}{2}) + \theta]$. It can be seen that the maximal θ for which the right side of (2.1) is larger than λ^2 is given by $\theta = 1 - \frac{3^{\frac{1}{2}}}{2}$. Thus the required interval is (.37, .63) which is slightly smaller than the one corresponding to the sample size of three.

I want to thank the referee for suggesting the maximal interval in the Remark II.

REFERENCE

LEHMANN, E. L. (1949). Notes on the theory of estimation. *Associated Students Store*. Univ. of California, Berkeley.