Estimation With Inadequate Information

E. L. LEHMANN*

As a possible explanation for the existence of absurd UMVU estimators, a class of estimation problems is defined: estimating the probability of an "unobservable" event, for example, estimating the probability of an event of interest occurring over a long period of time from observations over a much shorter period. It is pointed out that in typical cases of this kind, no reasonable unbiased estimator can exist. Consideration of the maximum likelihood estimator suggests the possibility that in fact there may then often exist no reasonable estimators.

KEY WORDS: Unbiased estimation; Unobservable events; Inadequate information.

1. INTRODUCTION

The possibility of an absurd unbiased estimator with uniformly minimum variance (UMVU) is usually illustrated by one or another variant (see, e.g., Kendall and Stuart 1979, Ex.17.26) of the following situation.

Example 1. Let X have the Poisson distribution $P(\lambda)$ with $E(X) = \lambda$, and let the estimand be $g(\lambda) = e^{-b\lambda}$, $0 < b < \infty$. The equations

$$\sum \delta(x)\, \lambda^x \left(\frac{e^{-\lambda}}{x!}\right) = e^{-b\lambda} \qquad \text{for all } \lambda > 0$$

have the unique solution $\delta(x) = (1 - b)^x$. For $0 < b < 1$ this decreases strictly from 1 at $x = 0$ to 0 at $x = \infty$ and is a reasonable estimator of $g(\lambda)$, which decreases strictly from 1 at $\lambda = 0$ to 0 at $\lambda = \infty$. However, for $b > 1$, $\delta(x)$ oscillates between positive and negative values as x is even or odd and no longer appears to bear much relation to the function it estimates.

The purpose of this note is to suggest that this is not an isolated example but is rather a special case of a class of situations in which, in a certain sense, the available information is inadequate for the existence of reasonable unbiased (or possibly any) estimators.

2. PROBLEMS WITH INADEQUATE INFORMATION

Suppose that the observations are represented by X with a distribution P_θ, $\theta \in \Omega$, and that Y with distribution Q_θ is independent of X and unobserved. For the model (X, Y) with distribution $P_\theta \times Q_\theta$, $\theta \in \Omega$, we assume the existence of a complete sufficient statistic T. The estimand to be estimated on the basis of X is assumed to be of the form

$$g(\theta) = P_\theta(T \in A) \tag{1}$$

for some set A. Whether T falls into A cannot be observed on the basis of X. For short, the event $T \in A$ is therefore called *unobservable*.

To see how Example 1 fits into this scheme, suppose that $b > 1$ and that Y is distributed independently of X according to a Poisson distribution with $E(Y) = (b - 1)\lambda$. For (X, Y), the statistic $T = X + Y$ is sufficient and complete, and the estimand of Example 1 is $e^{-b\lambda} = P(T = 0)$. The situation may be described somewhat more concretely in terms of a Poisson process with parameter λ. The quantity $e^{-b\lambda}$ is the probability that no event occurs in a time interval of length b. This probability is estimated by observing the process for a time interval of length 1. No difficulty arises when $b < 1$ but X is no longer adequate when $b > 1$, and the inadequacy increases as b increases. Correspondingly, the oscillations of the estimator $(1 - b)^x$, and hence its unreasonableness, increase with $b > 1$. The case $b = 1$, in which $e^{-b\lambda} = P(X = 0)$ is marginal; the UMVU estimator is the indicator of the set $\{X = 0\}$. (Note: The term UMVU estimator may seem inappropriate here since there exists only one unbiased estimator. However, in this and the other situations considered here, X itself is typically a sufficient statistic for some observations $X_1, \ldots, X_n$; in the present case, the X_i might be the numbers of events occurring in n nonoverlapping intervals of length $1/n$.)

When estimating (1) on the basis of X, an unbiased estimator frequently will not exist, as is illustrated by the following two examples.

Example 2. Consider n binomial trials with success probability p, of which the first $m < n$ are observed but the remaining n-m are not. If X_i is 1 or 0 when the ith trial is, respectively, a success or a failure, let $X = X_1 + \cdots + X_m$, $Y = X_{m+1} + \cdots + X_n$, and $T = X + Y$ and consider the estimand $g(p) = P(T = n) = p^n$. No unbiased estimator of p^n exists since only polynomials of degree less than m have unbiased estimators based on X.

Example 3. Let $X_1, \ldots, X_m$ $(m > 1)$ be iid according to the normal distribution $N(\xi, \sigma^2)$ and consider the problem of estimating $\Phi((u - \xi)/\sigma) = P(X_i \leq u)$. This problem was treated by Kolmogorov (1950), Lieberman and Resnikoff (1955), and others. For the sake of simplicity sup-

* E.L. Lehmann is Professor of Statistics, Department of Statistics, University of California, Berkeley, CA 94720. This article was prepared under National Science Foundation Grant MCS-79-03716. The author would like to thank the referee for a very careful reading of the original version, which resulted in substantial improvements.

September 1983, Volume 78, Number 383
Theory and Methods Section

pose that σ^2 is known. Then $\bar{X}$ is a complete sufficient statistic, distributed as $N(\xi, \tau^2)$ with $\tau^2 = \sigma^2/m$. For $\tau = 1$, the problem reduces to that of estimating $\Phi((u - \xi)/b)$ on the basis of a single random variable X distributed as $N(\xi, 1)$, where $b = \sqrt{m} > 1$.

The corresponding problem with $b < 1$ is of the type considered here. This is seen by letting Y be distributed independently of X with normal distribution $N(\xi, \beta^2)$, β^2 known. Then $T = \alpha X + (1 - \alpha)Y$ with $\alpha = \beta^2/(1 + \beta^2)$ is a complete sufficient statistic distributed as $N(\xi, \alpha)$, and

$$g(\xi) = P(T < u) = \Phi\left(\frac{u - \xi}{\sqrt{\alpha}}\right), \qquad (2)$$

which reduces to the earlier problem for $\alpha = b^2$. We shall see below that no unbiased estimator of (2) exists for $\alpha < 1$.

There is, however, a class of problems, illustrated by Example 1, in which an unbiased estimator of (1) does exist. Later in this section, we show that in such cases $\delta(X)$ will always be unsatisfactory.

Theorem 1. $\delta(X)$ is an unbiased estimator of (1) if and only if

$$\begin{aligned} E[\delta(X) \mid t] &= 1 \quad \text{if} \quad t \in A \\ &= 0 \quad \text{if} \quad t \notin A \quad (\text{a.e. } \mathcal{P}^T). \end{aligned} \qquad (3)$$

Proof. Let $\eta(t) = E[\delta(X) \mid t]$. If $I_A(t)$ denotes the indicator of A and $\delta(X)$ is unbiased, we have $E_\theta[\eta(T)] = E_\theta[I_A(T)] = g(\theta)$, and hence by completeness of T, $\eta(t) = I_A(t)$ (a.e. $\mathcal{P}^T$). The converse is obvious.

Before discussing the general implications of (3), let us use this result to show that no unbiased estimator $\delta(X)$ of $g(\theta)$ exists in Example 3. An easy calculation shows that the conditional distribution of X, given $T = t$, is normal with mean t and constant variance, say γ^2. Equation (3) thus becomes

$$\begin{aligned} \frac{1}{\sqrt{2\pi}\,\gamma} \int_{-\infty}^{\infty} \delta(x) \exp\left(-\frac{1}{2\gamma^2}(x - t)^2\right) dx \\ = 1 \quad \text{if} \quad t \leqslant u \\ = 0 \quad \text{if} \quad t > u. \end{aligned}$$

This can have no solution since the left side is an analytic function of t. An analogous proof shows that no unbiased estimator $\delta(X)$ of $P_\sigma(T \leqslant r)$ exists when $T = X + Y$, where X and Y are independently distributed as $\sigma^2\chi_m^2$ and $\sigma^2\chi_n^2$, respectively.

As we have seen, unbiased estimators of (1) often do not exist. However, there are important classes of problems for which the mathematics forces the existence of such estimators, which are then liable to share some of the unfortunate features of the estimator $\delta(X)$ of Example 1.

Suppose that there exist $t_0 \in A$ and $t_1 \notin A$ such that the conditional distributions of X given t_0 and given t_1, respectively, are not mutually singular. Then if $0 < g(\theta) < 1$ for all θ, any unbiased estimator must take on values either both ≥ 1 and < 0 or both > 1 and ≤ 0. (Under weak additional assumptions it has to take on values both > 1 and < 0.)

That $\delta(X)$ takes on values outside the range of the estimand $g(\theta)$ is an unpleasant property shared by some other UMVU estimators (for example, by those of some variance components). What is more unusual is the oscillatory character of the estimator when one expects it to be monotone.

The following result shows that in many situations (3) precludes $\delta(X)$ from being monotone.

Theorem 2. Suppose X and T are real valued and that the conditional distribution of X given t is strictly stochastically increasing in t in the sense that $t < t'$ implies

$$P(X \leq x \mid t) > P(X \leq x \mid t') \qquad (4)$$

for all x not satisfying $P(X \leq x \mid t) = P(X \leq x \mid t') = 0$ or 1. Suppose in addition that for any $t < t'$ the conditional distributions of X, given t and t', are not mutually singular. Then if $\delta(X)$ is unbiased for estimating

$$g(\theta = P_\theta(T \leq r), \qquad (5)$$

it cannot be monotone.

Proof. By Theorem 1, δ satisfies

$$\begin{aligned} E[\delta(X) \mid t] &= 1 \quad \text{if} \quad t \leq r \\ &= 0 \quad \text{if} \quad t > r. \end{aligned} \qquad (6)$$

Suppose $\delta(X)$ is nonincreasing. Then $t < t'$ implies $E[\delta(X) \mid t] > E[\delta(X) \mid t']$ unless there exists a constant c such that

$$P[\delta(X) = c \mid t] = P[\delta(X) = c \mid t'] = 1.$$

It thus follows from (6) that $\delta(X) = 1$ with probability 1 under the conditional distributions of X, given t with $t \leq r$, and that $\delta(X) = 0$ with probability 1 under the conditional distributions X, given t with $t > r$. This violates the assumed nonsingularity of these conditional distributions.

3. SOME FURTHER EXAMPLES

In generalization of Example 1, consider the following problem.

Example 4. Let X be distributed as in Example 1, but let the estimand be

$$g_r(\lambda) = P(T = r) \quad \text{or} \quad h_r(\lambda) = P(T \leqslant r), \qquad (7)$$

where T is distributed as $P(b\lambda)$. Note that $h_r(\lambda)$ decreases from 1 to 0 as λ increases from 0 to infinity, while $g_r(\lambda)$ for $r > 0$ increases from 0 to a maximum at $\lambda = r/b$ and then decreases.

The UMVU estimators of $g_r(\lambda)$ and $h_r(\lambda)$ are, respectively,

$$\begin{aligned} \delta_r(x) &= \binom{x}{r} b^r(1 - b)^{x-r} \quad \text{if } x \geqslant r \\ &= 0 \qquad\qquad\qquad\quad \text{if } 0 \leqslant x < r \end{aligned} \qquad (8)$$

and

$$\eta_r(x) = \sum_{j=0}^{r} \delta_j(x). \quad (9)$$

It is easy to check that if $b > 1$, then in both cases the estimator will take on both positive and negative values for arbitrarily large values of x. In fact, for any given $b > 1$,

$$\frac{\delta_r(x+1)}{\delta_r(x)} \to -(b-1) \quad \text{as} \quad x \to \infty$$

so that for sufficiently large x, $\delta_r(x)$ oscillates between successive positive and negative values, and the magnitude of the oscillations increases with b. That $\eta_r(x)$ exhibits the same behavior follows from the fact that $\eta_r(x)/\delta_r(x) \to 1$ as $x \to \infty$.

Example 5. Let $X + m$ be the number of binomial trials required to obtain m successes, so that X has the negative binomial distribution

$$P(X = x) = \binom{m+x-1}{m-1} p^m q^x, \quad x = 0, 1, \ldots$$

and let

$$g(q) = (1-q)^N = \mathrm{P}(T = 0),$$

where $T + N$ is the number of binomial trials required to obtain N successes (with the same success probability p). Then the UMVU estimator of $g(q)$ is given by

$$\sum \delta(x) \binom{m+x-1}{m-1} q^x \equiv (1-q)^{N-m}. \quad (10)$$

If $N < m$, say $m - N = k < m$, $\delta(x)$ reduces to

$$\delta(x) = \frac{k(k+1)\cdots(m-1)}{(x+k)(x+k+1)\cdots(x+m-1)},$$

which decreases from 1 at $x = 0$ to 0 as $x \to \infty$, as one would expect. On the other hand, consider the case $N > m$, $N - m = n > 0$, say. This corresponds to the situation in which $T = X + Y$, where $Y + n$ is the number of binomial trials required to obtain n successes. The assumptions of Theorem 1 thus hold, and it is easily checked that those of Theorem 2 do too. The estimator (10) now reduces to

$$\delta(x) = (-1)^x \frac{n(n-1)\cdots(n-x+1)}{m(m+1)\cdots(m+x-1)} \quad \text{for } x = 0, \ldots, n$$

and $\delta(x) = 0$ for $x > n$. As in the earlier examples, we find that the oscillations of $\delta(x)$ (for $x \leq n$) increase in magnitude with n, that is, as the unobserved information embodied in the estimand increases.

Example 6. Let $X_1, \ldots, X_n; X_{n+1}, \ldots, X_N$ be iid according to the uniform distribution on $(0, \theta)$, where it is assumed known that $\theta > 1$ but θ is otherwise unknown, and let

$$\begin{aligned} X &= 1 \quad \text{if } \max(X_1, \ldots, X_n) < 1 \\ &= \max(X_1, \ldots, X_n) \quad \text{otherwise} \\ Y &= 1 \quad \text{if } \max(X_{n+1}, \ldots, X_N) < 1 \\ &= \max(X_{n+1}, \ldots, X_N) \quad \text{otherwise} \end{aligned}$$

and $T = \max(X, Y)$. Then the factorization criterion shows T to be sufficient for θ, and it is easily checked that T is complete. Suppose only $X_1, \ldots, X_n$ are observed and that it is desired to estimate

$$g(\theta) = \frac{1}{\theta^N} = P_\theta(T < 1).$$

The estimator $\delta(X)$ given by

$$\delta(1) = 1; \delta(x) = \frac{n-N}{nx^N} \quad \text{for} \quad x > 1 \quad (11)$$

is UMVU. The example satisfies the conditions of Theorem 2 and although $g(\theta)$ is strictly decreasing, one can, in fact, easily see directly that $\delta(x)$ is not monotone since it is positive for $x = 1$ and negative but increasing for $x > 1$. The UMVU estimator is given by (11) also when $N < n$ but in that case is nonincreasing as one would expect.

4. CONCLUDING REMARKS

In the preceding sections we have indicated a class of estimation problems—estimating the probability of an unobservable event—that frequently lead to absurd UMVU estimators. The suggested explanation is that in these situations the data do not contain enough information to provide reasonable unbiased estimators. The question naturally arises whether in such cases any reasonable estimators exist. This requires another investigation.

As a first clue, one might consider the performance of the maximum likelihood estimator (MLE). It is the most widely used estimating procedure, cannot take on values outside the range of the estimand, and in all our examples is a monotone function of x when $g(\theta)$ is monotone in θ.

Example 1 (continued). In the context of Example 1, the MLE is $\delta(X) = e^{-bX}$, and

$$E\,\delta(X) = \exp[\lambda(e^{-b} - 1)] \equiv h(\lambda). \quad (12)$$

The bias of $\delta(X)$ is therefore

$$b(\lambda) = h(\lambda) - g(\lambda) = e^{-a\lambda} - e^{-b\lambda}, \quad (13)$$

where $a = 1 - e^{-b}$. One can easily see that $a < b$, so that on the average $\delta(X)$ overestimates $e^{-b\lambda}$ for all λ. The bias function $b(\lambda)$ tends to 0 as $\lambda \to \infty$ and has a unique maximum at

$$\lambda^* = \frac{\log b - \log a}{b - a}. \quad (14)$$

Table 1 shows the value of λ^* and the maximum bias

Table 1. Maximum Bias $b(\lambda^*)$ in Example 1

b	$\frac{1}{4}$	$\frac{1}{3}$	$\frac{1}{2}$	1	2	3	4	5
λ^*	4.250	3.250	2.249	1.247	.739	.561	.465	.403
$b(\lambda^*)$	.045	.060	.088	.167	.300	.401	.478	.537
$g(\lambda^*)$	.346	.339	.325	.287	.228	.186	.155	.133

$b(\lambda^*)$ for various values of b. It is striking to observe the deterioration of the situation for $b \geqslant 1$. Biases of .4 and .5 in estimating probabilities are hardly tolerable.

It is also noteworthy that for any fixed λ, the relative bias $b(\lambda)/e^{-\lambda} \sim e^{(b-1)\lambda}$ as $b \to \infty$ and hence tends to infinity exponentially as $b \to \infty$.

Example 6 (continued). The MLE of $1/\theta^N$ in the context of Example 6 is

$$\delta(X) = 1 \quad \text{if} \quad X < 1$$
$$= 1/X^N \quad \text{if} \quad X > 1$$

and its bias is

$$b(\theta) = \frac{N}{n - N}(\theta^{-N} - \theta^{-n}).$$

The bias is thus always positive. It is zero at $\theta = 1$ and tends to be zero as $\theta \to \infty$. It has a unique maximum at

$$\theta^* = \left(\frac{n}{N}\right)^{1/(n-N)}. \tag{15}$$

Table 2 shows for $n = 3$ and 5 the value of θ^* and the maximum bias for various values of N. As in the preceding example, the maximum bias increases steeply when the information becomes inadequate (i.e., for $N > n$).

Table 2. Maximum Bias $b(\lambda^*)$ in Example 6

N	1	2	4	6	8	10	12	15
				$n = 3$				
θ^*	1.732	1.500	1.333	1.260	1.217	1.188	1.167	—
$b(\theta^*)$	.192	.296	.422	.500	.555	.597	.630	—
$g(\theta^*)$	.577	.444	.316	.250	.208	.179	.157	—
				$n = 5$				
θ^*	1.357	1.291	1.250	1.200	1.170	1.149	1.133	1.116
$b(\theta^*)$	.217	.279	.328	.402	.457	.500	.535	.577
$g(\theta^*)$	.543	.465	.410	.335	.286	.250	.223	.192

The results of Sections 2 and 3 together with the rather limited numerical evidence just presented suggest that, in the situations considered here, reasonable estimators of the probabilities (1) may not exist. To prove that a reasonable estimator is incompatible with acceptable bias control, it would be necessary to determine the estimator that minimizes the maximum absolute bias subject to the appropriate monotonicity (and range) conditions. Beyond this, it would be important to determine the estimators that minimize the maximum risk without imposing such conditions.

Situations like those described here are not purely mathematical artifacts. In particular, the estimation of probabilities of minor or major catastrophes over long periods of time from observations over much shorter periods tend to be just of this type. However, when the probabilities being estimated are very small, the difficulty of the estimation problem may be overshadowed by the unreliability of the model in the extreme part of the distribution.

[*Received March 1981. Revised January 1983.*]

REFERENCES

KENDALL, M., and STUART, A. (1979), *The Advanced Theory of Statistics* (Vol. 2, 4th ed.), New York: Macmillan.

KOLMOGOROV, A.N. (1950), "Unbiased Estimates," *Izvestija Akademii Nauk Armjanskoĭ SSR, Serija Matematika*, 14, 303–326 (American Mathematical Society Translation 98).

LIEBERMAN, G.J., and RESNIKOFF, G.J. (1955), "Sampling Plans for Inspection by Variables," *Journal of the American Statistical Association*, 50, 457–516.