A Class of Parameter Functions for Which the Unbiased Estimator Does Not Exist

By Shande Chen

University of North Texas Health Science Center at Fort Worth

Unbiased estimation is a popular criterion in small sample point estimation. However, there is limited knowledge about conditions under which the unbiased estimate does not exist. In this paper, in analogy to the binomial estimation, we give a class of parameter functions for which no unbiased estimator exists. The minimax bias estimators for these functions are obtained. The relationship between our results and the sample size and the interim review is commented.

1. Introduction Point estimation is a very important area in statistical inference. In small-sample estimation, a lot of attention is paid to unbiased estimation and a complete theory, the Blackwell-Rao and Lehmann-Scheffé theorems, about the uniformly minimum variance unbiased estimator (UMVU) has been developed. This topic appears in intermediate to advanced statistical inference textbooks, such as [3] and [12], as well as many lecture notes, such as [6]. According to the classic works by Lehmann and Scheffé [13], Halmos [7] and Bahadur [1], the condition of unbiasedness is generally a strong one. However, there is very limited knowledge about under what conditions the unbiased estimate does not exist. One well-known example is in binomial samples. From a sample of n Bernoulli trials with success probability π, only polynomials in π of degree no more than n can be estimated unbiasedly. There are few other examples for the non-existence of the unbiased estimator in statistical inference textbooks. Some approaches, primarily focusing on binomial estimation problems, are suggested by Bhattacharyya [2], Siraždinov [14], and Hall [5], when the unbiased estimator does not exist. We believe more examples for non-existence of the unbiased estimator will enrich both the theory of point estimation and statistical education.

A special problem comes from the practice of clinical trials. To determine the sample size, several parameters need to be assumed. For simplicity, we consider the one-sided one-sample z-test based on a sample from $N(\mu, 1)$, and testing H_0: $\mu = 0$ vs H_1: $\mu > 0$. Usually a target value μ_1 of the alternative is assumed, and the sample size is determined by the significance level and the power at this target value. Based on such a design, the whole power function can be obtained, so investigators can know what the power of the study is if the true difference is some value of μ. It is noted that the power of the study depends on the unknown parameter (treatment effect), and hence it is also unknown. Sometimes, this target value is given according to a well-established clinical significance, e.g., a test is required to have 80% power to detect H_1: $\mu = 1$, where people think $\mu \geq 1$ is clinically significant. However, it is not rare that people choose an estimated μ from a previous small study as the target. There could be two interpretations for using such an estimated μ in power calculation. One is treating this estimated value as a reasonable non-random target value, so the power at the target is not related to the power of the study (at the unknown treatment effect). The other interpretation is that the estimated μ is a random variable from a sample of a smaller size and is used to estimate the power at the unknown parameter μ for a larger sample. Usually clinicians want to know how likely a clinical trial will conclude a significant result. Therefore such an estimator for the study power at the unknown parameter provides an educated guess at best. Of course, the question of whether this kind of estimator for the power can be unbiased arises naturally. This provides a motivation for our study.

In the next section, we give a class of parameter functions for which no unbiased estimator exists, and construct some minimax bias estimators for them. Finally, more discussion about our results and the relationship with the sample size and the interim review is presented.

2. Main Results Let $X_1, \cdots, X_n$ be *i.i.d.* from a parametric distribution $F(\cdot;\mu)$. Let $\{\mathcal{X}, \Omega, \mathcal{M}\}$ be a reference measure on the sample space $\mathcal{X}$ for X. Typically, when $\mathcal{X}$ is all the real numbers, the reference measure is the Lebesgue measure, and when $\mathcal{X}$ is discrete (e.g., all the natural numbers or $\{0,1\}$), it is the counting measure. Let $Z_n = Z(X_1, \cdots, X_n)$ be a statistic. The distribution (probability measure) for Z_n induced from the product measure of F, the reference measure on the sample space of Z_n will depend on the nature of the sample space, e.g., the Lebesgue measure if the sample space is real. Therefore the densities or conditional densities for $X_1, \cdots, X_n$ and their functions are well defined by the corresponding probability measures and the reference measure. We consider estimating $\theta(\mu) = P(Z_n \in \mathcal{R}|\mu)$, a function of μ. When Z_n is the test statistic and $\mathcal{R}$ is the rejection region of a test, $\theta(\mu)$ is the power of the test at the unknown parameter μ. Let $0 < m < n$ and let $Z_m = Z(X_1, \cdots, X_m)$. Let $g(\cdot; Z_n)$ denote the conditional density of Z_m given Z_n.

THEOREM 1. *For fixed $0 < m < n$, assume that Z_m is complete and sufficient with respect to $X_1, \cdots, X_m$ and that Z_n is complete and sufficient with respect to $X_1, \cdots, X_n$.*

(A) Assume that there exit a subset $\mathcal{A}$ of the sample space of Z_m and two non-empty sets $\mathcal{B}_1 \subset \mathcal{R}$ and $\mathcal{B}_0 \subset \mathcal{R}^c$, where $\mathcal{R}^c$ is the complement of $\mathcal{R}$, such that for each $y \in \mathcal{B}_0 \cup \mathcal{B}_1$, $\mathcal{A}$ belongs to the support of $g(\cdot; y)$ and has a nonzero probability measure with respect to $g(\cdot; y)$. Then, there is no unbiased estimator $T_m = T(X_1, \cdots, X_m)$ for $\theta(\mu)$, such that $0 \le T_m \le 1$.

(B) Assume $g(\cdot; y)$ is continuous on the sample space of Z_n in the sense of the weak convergence of the distribution function $G(\cdot; y)$. And assume the intersection of the closures of $\mathcal{R}$ and $\mathcal{R}^c$ is not empty. Then, there is no unbiased estimator $T_m = h(Z_m)$, such that $h(\cdot)$ is a continuous and bounded function, for $\theta(\mu)$.

PROOF. For simplicity, we use $m = 1$ and $n = 2$ in the proof. Suppose that T_1 is an unbiased estimator for $\theta(\mu)$. Since Z_1 is sufficient with respect to X_1, $h(Z_1) \equiv \mathbf{E}(T_1|Z_1)$ is also an unbiased estimator and bounded by 0 and 1. Thus, without loss of generality, we can assume $T_1 = h(Z_1)$. Based on *i.i.d.* assumption, $g(\cdot; Z_n)$ is not degenerate. Since Z_2 is sufficient with respect to X_1, X_2, $\mathbf{E}[h(Z_1)|Z_2] = \int h(u) g(u; Z_2) du$ is also an unbiased estimator. It is obvious that

$1_{\mathcal{R}}(Z_2)$, where $1_{\mathcal{R}}(\cdot)$ is the indicator function for $\mathcal{R}$, is unbiased for $\theta(\mu)$. It follows from completeness of Z_2 that

$$1_{\mathcal{R}}(y) = \int h(u)g(u;y)du, \quad a.s. \tag{2.1}$$

for y in the sample space of Z_2. Since $0 \leq h(u) \leq 1$ and $\mathcal{A}$ has a nonzero probability measure, for $y \in \mathcal{B}_1$, (2.1) implies that $h(u) = 1, u \in \mathcal{A}$. However, if $y \in \mathcal{B}_0$, (2.1) implies that $h(u) = 0, u \in \mathcal{A}$. This leads to a contradiction and completes the proof for (A).

To prove (B), consider two sequences of y in (2.1), one in $\mathcal{R}$ and the other in $\mathcal{R}^c$, both tend to a point y_0 in the intersection of the closures. Applying the Helly-Bray theorem, the right hand side of (2.1) converges to a common value $\int h(u)g(u;y_0)du$ for both sequences. But from left hand side of (2.1), the integral converges to 1 for the sequence in $\mathcal{R}$ and to 0 for that in $\mathcal{R}^c$, which will lead to a contradiction. □

Remark. Theorem 1 allows a vector parameter, in which both Z_m and Z_n are vectors. Thus, the theorem applies to power for hypothesis testing with nuisance parameters, such as the one-sample t-test.

Remark. The completeness for Z_m is not needed to prove Theorem 1. In general, Z_m and Z_n should have the same properties, so including completeness for Z_m does not narrow the application of the theorem in practice.

Remark. Based on the assumption that $X_1, \cdots, X_n$ are *i.i.d* from $F(\cdot;\mu)$, for any m such that $0 < m < n$, neither Z_m nor $X_1, \cdots, X_m$ is sufficient with respect to $X_1, \cdots, X_n$. This illustrates Theorem 1. It also shows the importance of sufficiency.

Remark. In (B) of Theorem 1, the condition on $h(\cdot)$ can be replaced by other conditions, such as uniform integrability, to imply the convergence in mean.

From Theorem 1, the power at the unknown parameter of a test of sample size n cannot be estimated unbiasedly from a sample of a smaller size. One-sample and two-sample tests for proportion are examples, in which (A) is easily verified. In the simpler one-sample case, the power is a polynomial in the population proportion of degree n, therefore non-existence of an unbiased estimate for it from a sample of fewer Bernoulli trials also follows from a well-known fact. It is noted that $\theta(\mu)$ is not necessarily a power function. Many functions of $\theta(\mu)$ can be obtained by selecting different $\mathcal{R}$.

Now we consider data from a normal population with unknown mean μ and known variance σ^2, without loss of generality we may assume $\sigma = 1$. We are interested in testing $H_0: \mu = 0$ vs. $H_1: \mu > 0$, and hence a z-test will be performed. Suppose a level α and a sample size n are specified. The null hypothesis is rejected if $Z_n = \sum_{i=1}^n X_i/\sqrt{n} \geq z_\alpha$, where $X_1, \cdots, X_n$ are *i.i.d.* from $N(\mu, 1)$. The power at the unknown μ is $\bar{\Phi}(z_\alpha - \delta)$, where $\bar{\Phi}$ and z_α are the upper tail probability function and $100(1-\alpha)$ percentile for the standard normal distribution, respectively, and $\delta = \sqrt{n}\mu$. Again for fixed $0 < m < n$, the sufficiency and completeness of Z_m and Z_n are well-known. It is not difficult to show that conditional on Z_n, Z_m has a normal distribution with mean $\sqrt{t}Z_n$ and variance $1-t$, where $t = m/n$. Therefore, both (A) and (B) of Theorem 1 apply. These conclusions are easily extended to the case of unknown σ^2 (a nuisance parameter).

The MLE (maximum likelihood estimator) based on $X_1, \cdots, X_m$ for the above power is $\bar{\Phi}(z_\alpha - Z_m/\sqrt{t})$. Since Z_m is distributed as $N(\sqrt{m}\mu, 1)$ and the MLE is in the form of $P(Y > \sqrt{t}z_\alpha - z_m)$, where Y is distributed as $N(0,t)$ and independent of Z_m, the expectation of the MLE can be easily obtained by the convolution for the sum of two independent normal variates. It follows that

$$\mathbf{E}[\bar{\Phi}(z_\alpha - Z_m/\sqrt{t})] = \bar{\Phi}\{[z_\alpha - \delta]/\sqrt{1/t+1}\}. \tag{2.2}$$

Therefore the bias of the MLE is determined by the power $\bar{\Phi}(z_\alpha - \delta)$ and the ratio $t = m/n$.

Since there is no unbiased estimator for the unknown power based on $X_1, \cdots, X_m$, minimum-bias estimation becomes a reasonable approach. Some theory and examples for minimum-bias estimation can be found in [5]. The following theorem provides another approach to construct such estimators.

THEOREM 2. *For fixed $0 < m < n$ and $X_1, \cdots, X_m$ being i.i.d. from $N(\mu, 1)$, the estimator $1_{\{z \geq \sqrt{t}z_\alpha\}}(Z_m)$ is the minimax bias estimator for $\bar{\Phi}(z_\alpha - \delta)$ (where $\delta = \sqrt{n}\mu$) among estimators $h(Z_m)$ with monotone increasing h and $0 \leq h(z) \leq 1$, that is, it minimizes $\sup_\delta |b(\delta; h)|$, where $b(\delta; h) = \mathbf{E}[h(Z_m)|\delta] - \bar{\Phi}(z_\alpha - \delta)$ is the bias of the estimator at δ.*

PROOF. Let $\gamma = z_\alpha - \delta$ and let $k(\cdot) = h(\cdot + \sqrt{t}z_\alpha)$. Then

$$\mathbf{E}[h(Z_m)|\delta] = \int k(u)\phi(u + \sqrt{t}\gamma)du = k(\infty) - \int \Phi(u + \sqrt{t}\gamma)dk(u),$$

where Φ and ϕ are the distribution and density functions of $N(0,1)$, respectively. For $h = h^* = 1_{\{z \geq \sqrt{t}z_\alpha\}}$, $k = k^* = 1_{\{u \geq 0\}}$, the bias is $\Phi(\gamma) - \Phi(\sqrt{t}\gamma)$. By differentiating this bias with respect to γ, it follows that the absolute bias achieves its maximum at γ^* which satisfies $\phi(\gamma) = \sqrt{t}\phi(\sqrt{t}\gamma)$, yielding $\gamma_1^* = \sqrt{-\log(t)/(1-t)}$ and $\gamma_2^* = -\gamma_1^*$. For any $\rho > 0$, $\Phi(u+\rho) - \Phi(u-\rho) \leq \Phi(\rho) - \Phi(-\rho)$. This implies that

$$\begin{aligned} &\int k(u)\phi(u + \sqrt{t}\gamma_2^*)du - \int k(u)\phi(u + \sqrt{t}\gamma_1^*)du \\ &= \int [\Phi(u + \sqrt{t}\gamma_1^*) - \Phi(u - \sqrt{t}\gamma_1^*)]dk(u) \leq \Phi(\sqrt{t}\gamma_1^*) - \Phi(\sqrt{t}\gamma_2^*). \end{aligned} \tag{2.3}$$

At γ_1^*, the bias of h^* is positive. If an estimator h has less absolute bias at γ_1^*, then we must have $\int k(u)\phi(u + \sqrt{t}\gamma_1^*)du < 1 - \Phi(\sqrt{t}\gamma_1^*)$. But from (2.3), $\int k(u)\phi(u + \sqrt{t}\gamma_2^*)du < 1 - \Phi(\sqrt{t}\gamma_2^*)$, which implies that the absolute bias of h is greater than h^* at γ_2^*. This proves the minimax property of h^*.

If another h is also a minimax estimator, then (2.3) must hold as an equality. But this is true if and only if $h = h^*$ *a.s.*, and hence the uniqueness follows. □

Remark. Let $F(\cdot)$ be a distribution function with density f which is symmetric around 0 and unimodal. Let Z_m have distribution function $F(\cdot - \sqrt{t}\delta)$. To estimate $1 - F(c - \delta)$, where c is a given constant, the bias of $1_{\{z > \sqrt{t}c\}}(Z_m)$ is $F(\gamma) - F(\sqrt{t}\gamma)$

(where $\gamma = c - \delta$), which is positive for $\gamma > 0$ and skew symmetric at 0, and is equal to zero at 0 and ∞. Therefore its maximum absolute bias is achieved at γ_1^* and $\gamma_2^* = -\gamma_1^*$ for some $\gamma_1^* > 0$. Similarly to the proof of Theorem 2, $1_{\{z>\sqrt{t}c\}}(Z_m)$ is the minimax bias estimator for $1 - F(c - \delta)$. Non-existence of unbiased estimators follows as a direct consequence.

Remark. The approach in the proof of Theorem 2 is different from that in [5]. A simple exact form of the minimax bias estimator may be rare in general if no unbiased estimator exists.

3. Discussion We have found a class of parameter functions for which the unbiased estimator does not exist. Intuitively, the probability of an event from a sample of a bigger size, in general, cannot be estimated unbiasedly from a sample of a smaller size. When the sample is binomial, these functions are polynomials in the success probability of degree greater than the sample size. Therefore, our conclusion is an extension of the binomial case, and may be used as another example to illustrate a case where the unbiased estimator does not exist. The conditions in Theorem 1 or Theorem 2 are satisfied for many commonly–used distributions. Hence, non-existence of an unbiased estimator is not rare. The present paper may help us to better understand about unbiasedness and other estimation criteria.

There are two issues related to the sample size calculation or power determination. The first is in the practice of the clinical trial when the target value is an estimate from a pilot study. In this case, the target power is similar to the MLE for the power at the unknown true parameter. Since usually the target power is quite high (80% or more), from (2.2), this estimate is biased to be lower. Therefore, the traditional methods are conservative. The second issue is in statistical education. Many introductory statistics textbooks contain sample size formulas, especially for the one-sample z-test (known variance), e.g., [9]. Thus, it is important to explain the meaning of the target value well and, if necessary, to comment on possible consequences when an estimate is used as the target value.

In addition, in interim reviews, there is an analogous situation. Here, let us briefly discuss some related issues. One tool used in interim reviews is the conditional power [8, 10, 11]. The conditional power is a conditional probability of rejecting the null hypothesis at the end of a trial given part of the data (accumulated at the interim review). The conditional power evaluated at the unknown parameters certainly describes the tendency best, but is unknown. The expectation of this unknown conditional power is exactly the power of the test at the unknown parameters. Therefore, there does not exist any "unbiased" estimator for this conditional power (i.e., the expectation of the difference between an estimator and this conditional power vanishes). This may explain why stochastic curtailment based on conditional powers will change the type I and II errors. On the other hand, once we take an interim look, the parameter of interest becomes conditional power, the concept of unconditional power is no longer relevant. In other words, the estimate of conditional power is far more important than the estimate of power. Another similar situation is in the sample size adjustment techniques pioneered by Stein [15]. Some more recent developments are in [16] and [4]. The typical idea in sample size adjustment is to adjust the sample size by estimating the unknown variance. From the results in Section 2, any adjustment of the sample size based on an estimated improvement at an interim review is biased, and such adjustments may need more studies before they can be applied.

Though we start from a problem in point estimation theory, as discussed, our results are also related to a few applications. The results should be useful in statistical theory, application and education.

Acknowledgments The author wishes to thank Professor W. J. Hall for the reference about minimum-bias estimation and beneficial discussion, and Wendy Conner, M.S., for helpful remarks. The author also thanks two reviewers for their helpful comments and Drs. D. Oakes and J. E. Kolassa for their editorial work.

REFERENCES

[1] R.R. Bahadur. On unbiased estimates of uniformly minimum variance. *Sankhya*, 18:211–224, 1957.

[2] A. Bhattacharyya. Notes on the use of unbiased and biased statistics in the binomial population. *Calcutta Statistical Association Bulletin*, 5:149–164, 1954.

[3] P. Bickel and K. Doksum. *Mathematical Statistics: Basic Ideas and Selected Topics*. Holden-Day, Oakland, CA, 1977.

[4] A.L. Gould. Interim analyses for monitoring clinical trials that do not materially affect the type I error rate. *Statistics in Medicine*, 11:55–66, 1992.

[5] W.J. Hall. Estimators with minimum bias. In R. Bellman, editor, *Mathematical Optimization Techniques*, Berkeley, 1963. University of California Press.

[6] W.J. Hall. A course in statistical inference. Technical Report 91/01, Departments of Statistics and Biostatistics, University of Rochester, Rochester, NY, 1991.

[7] P.R. Halmos. The theory of unbiased estimators. *Annals of Mathematical Statistics*, 17:34–43, 1946.

[8] M. Halperin, K.K.G. Lan, J.H. Ware, N.J. Johnson, and D.L. DeMets. An aid to data monitoring in long-term clinical trials. *Controlled Clinical Trials*, 3:311–323, 1982.

[9] D.V. Huntsberger and P.P. Billingsley. *Elements of Statistical Inference*. Wm. C. Brown, Dubuque, IA, 6th edition, 1987.

[10] K.K.G. Lan, R. Simon, and M. Halperin. Stochastically curtailed tests in long-term clinical trials. *Sequential Analysis*, 1:207–219, 1982.

[11] K.K.G. Lan and J. Wittes. The b-value: A tool for monitoring data. *Biometrics*, 44:579–585, 1988.

[12] E.L. Lehmann. *Theory of Point Estimation*. Wiley, 1983.

[13] E.L. Lehmann and H. Scheffé. Completeness, similar regions, and unbiased estimation–part I. *Sankhya*, 10:305–340, 1950.

[14] S.H. Siraždinov. Concerning estimations with minimum bias for a binomial distribution. *Theory of Probability and Its Applications*, 1:174–176, 1956. In Russian with English summary.

[15] C. Stein. A two-sample test for a linear hypothesis whose power is independent of the variance. *Annals of Mathematical Statistics*, 16:243–258, 1945.

[16] J. Wittes and E. Brittain. The role of internal pilot studies in increasing the efficiency of clinical trials. *Statistics in Medicine*, 9:65–72, 1990.

University of North Texas Health Science Center
School of Public Health, Department of Biostatistics
3500 Camp Bowie Blvd.
Fort Worth, TX 76116 USA
schen@hsc.unt.edu