

An Example of an Improvable Rao–Blackwell Improvement, Inefficient Maximum Likelihood Estimator, and Unbiased Generalized Bayes Estimator

Tal Galili & Isaac Meilijson

To cite this article: Tal Galili & Isaac Meilijson (2016) An Example of an Improvable Rao–Blackwell Improvement, Inefficient Maximum Likelihood Estimator, and Unbiased Generalized Bayes Estimator, The American Statistician, 70:1, 108-113, DOI: [10.1080/00031305.2015.1100683](https://doi.org/10.1080/00031305.2015.1100683)

To link to this article: <https://doi.org/10.1080/00031305.2015.1100683>



© 2016 The Author(s). Published with license by Taylor & Francis© Tal Galili and Isaac Meilijson



Published online: 31 Mar 2016.



Submit your article to this journal [↗](#)



Article views: 4231



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 2 View citing articles [↗](#)

An Example of an Improvable Rao–Blackwell Improvement, Inefficient Maximum Likelihood Estimator, and Unbiased Generalized Bayes Estimator

Tal GALILI and Isaac MEILIJSON

The Rao–Blackwell theorem offers a procedure for converting a crude unbiased estimator of a parameter θ into a “better” one, in fact unique and optimal if the improvement is based on a minimal sufficient statistic that is complete. In contrast, behind every minimal sufficient statistic that is not complete, there is an improvable Rao–Blackwell improvement. This is illustrated via a simple example based on the uniform distribution, in which a rather natural Rao–Blackwell improvement is uniformly improvable. Furthermore, in this example the maximum likelihood estimator is inefficient, and an unbiased generalized Bayes estimator performs exceptionally well. Counterexamples of this sort can be useful didactic tools for explaining the true nature of a methodology and possible consequences when some of the assumptions are violated.

KEY WORDS: Improper prior; Minimal sufficiency; Uniform distribution; Uniformly minimum-variance unbiased estimator.

1. INTRODUCTION

Statistical theory courses usually start with basic notions for describing how much information on some unknown parameter θ can be obtained from a set of data Y . These are likelihood, sufficiency, minimal sufficiency, completeness, the Rao–Blackwell (Rao 1945; Blackwell 1947), and Lehmann–Scheffé theorems (Lehmann and Scheffé 1950, 1955). Familiarity with these notions is assumed.

The Rao–Blackwell theorem (RBT) offers a procedure (coined “Rao-Blackwellization” seemingly by Berkson 1955) for improving a crude unbiased estimator $g(Y)$ of the parameter θ into a better one (in mean-squared-error or any other convex loss function), by taking the conditional expectation of $g(Y)$ given some sufficient statistic $T = T(Y)$,

that is, $\hat{\theta}_{\text{RB}} = E_{\theta}[g(Y)|T]$ (this is a statistic because T is sufficient). The Lehmann–Scheffé theorem and RBT taken together state that the (unique) unbiased estimator based on a *complete* minimal sufficient statistic T achieves uniformly smaller expected loss under any convex loss function (the common term UMVUE, uniformly minimum-variance unbiased estimator, stresses only squared loss). Furthermore, if a parameter can at all be unbiasedly estimated, then it can also be unbiasedly estimated by a function of T , and the Rao–Blackwell improvement of the former automatically leads to the latter. If an unbiased estimator of a parameter is *not* a function of the sufficient statistic T , the Rao–Blackwell improvement based on T is a strict improvement over the original unbiased estimator.

Classical examples for these ideas are often based on unconstrained exponential families of distributions in which a complete sufficient statistic always exists, and the Rao–Blackwell improvement yields optimal unbiased results (see Abramovich and Ritov 2013). However, applying Rao–Blackwell improvements with *noncomplete* minimal sufficient statistic T will always yield some estimator that fails to have minimal possible variance, at least somewhere in the parameter space: Any non-degenerate function of T with mean identically zero (the existence of such is the definition of noncompleteness) is an unbiased estimator of the zero function that is left unchanged by Rao-Blackwellization because it is already a function of T , but is dominated (with strict variance inequality somewhere) by the zero statistic.

The didactically motivated family of examples to be introduced does not depend on delicate measure-theoretical pathologies of noncountably generated families of distributions. It deals instead with commonplace positive uniformly distributed random variables parameterized by their mean, and yields a *uniformly* dominated Rao–Blackwell improvement. Consider the uniform distributions $U((1-k)\theta, (1+k)\theta)$ with unknown mean $\theta > 0$ and known design parameter $k \in (0, 1)$, henceforth the *scale-uniform* family of distributions. This family does not satisfy the usual differentiability assumptions leading to Fisher Information, the Crámer-Rao bound and efficiency of maximum likelihood estimators (MLEs). For this family, MLE is inefficient.

While proper Bayes estimators cannot be unbiased (Bickel and Blackwell 1967), unbiased examples have been built under improper priors, such as the sample mean in the normal case. An explicit Bayes estimator built under a specific improper prior will be shown to be unbiased, simultaneously for all sample sizes and all k .

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The moral rights of the named author(s) have been asserted.

Tal Galili (E-mail: tal.galili@gmail.com) and Isaac Meilijson (E-mail: isaco@math.tau.ac.il), Department of Statistics and Operations Research, Tel-Aviv University, Tel Aviv-Yafo, Israel. The authors gratefully acknowledge David Gilat, Amit Meir, the referees, and the associate editor for their helpful comments. This work was supported in part by the European Research Council under ECEP7 European Research Council grant PSARPS-297519.

2. AN IMPROVABLE RAO-BLACKWELL IMPROVEMENT

Let X_1, X_2, \dots, X_n be a random sample from a scale-uniform distribution $X \sim U((1-k)\theta, (1+k)\theta)$, with unknown mean $E[X] = \theta$ and known design parameter $k \in (0, 1)$. In the search for “best” possible unbiased estimators for θ , it is natural to consider X_1 as an initial (crude) unbiased estimator for θ and then try to improve it. Since X_1 is not a function of $T = (X_{(1)}, X_{(n)})$, the minimal sufficient statistic for θ (where $X_{(1)} = \min(X_i)$ and $X_{(n)} = \max(X_i)$), it may be improved using the Rao-Blackwell theorem as follows:

$$\hat{\theta}_{RB} = E_{\theta}[X_1 | X_{(1)}, X_{(n)}] = \frac{X_{(1)} + X_{(n)}}{2} \quad (1)$$

since $X_1 | T \sim U(X_{(1)}, X_{(n)})$. Although this estimator has lower variance than X_1 (and \bar{X}_n), it is not even best among the linear unbiased estimators for θ . For this family of uniform distributions, there exists an unbiased estimator $\hat{\theta}_{LV}$ for θ with uniformly lower variance than the Rao-Blackwell improvement of X_1 . Both $\hat{\theta}_{LV}$ and $\hat{\theta}_{RB}$ are linear in $X_{(1)}$ and $X_{(n)}$, but with different coefficients. This will be shown in two steps: first, two unbiased estimators $\hat{\theta}_m$ and $\hat{\theta}_M$ of θ will be defined, as constant multiples of (the minimum) $X_{(1)}$ and (the maximum) $X_{(n)}$, respectively. Second, among the affine combinations of $\hat{\theta}_m$ and $\hat{\theta}_M$, a minimum-variance solution $\hat{\theta}_{LV}$ will be found, with strictly lower variance than $\hat{\theta}_{RB}$, uniformly in $\theta > 0$.

This idea of taking an affine combination of two unbiased estimators to define a new unbiased estimator appeared at least 200 years ago in a work by Laplace (see Stigler 1973), and is still found in modern use (e.g., see recent work by Damilano and Puig 2004).

As a preliminary, let Y_1, Y_2, \dots, Y_n be iid, with common $U(a, b)$ distribution, for some $a < b$, with $Y_{(1)} = \min(Y_i)$ and $Y_{(n)} = \max(Y_i)$. Then

$$\begin{aligned} E[Y_{(1)}] &= a + \frac{1}{n+1}(b-a) \\ E[Y_{(n)}] &= b - \frac{1}{n+1}(b-a) \end{aligned} \quad (2)$$

$$\text{var}[Y_{(1)}] = \text{var}[Y_{(n)}] = \frac{n}{(n+1)^2(n+2)}(b-a)^2 \quad (3)$$

$$\begin{aligned} \text{cov}[Y_{(1)}, Y_{(n)}] &= \frac{1}{(n+1)^2(n+2)}(b-a)^2 \\ \rho[Y_{(1)}, Y_{(n)}] &= \frac{1}{n}. \end{aligned} \quad (4)$$

Step 1: Substituting $a = (1-k)\theta$ and $b = (1+k)\theta$, the last formulas take the form

$$\begin{aligned} E_{\theta}[X_{(1)}] &= \left(1 - \frac{n-1}{n+1}k\right)\theta \\ E_{\theta}[X_{(n)}] &= \left(1 + \frac{n-1}{n+1}k\right)\theta \end{aligned} \quad (5)$$

$$\text{var}_{\theta}[X_{(1)}] = \text{var}_{\theta}[X_{(n)}] = \frac{4nk^2\theta^2}{(n+1)^2(n+2)} \quad (6)$$

$$\text{cov}_{\theta}[X_{(1)}, X_{(n)}] = \frac{4k^2\theta^2}{(n+1)^2(n+2)}$$

$$\rho_{\theta}[X_{(1)}, X_{(n)}] = \frac{1}{n} \quad (7)$$

giving rise to the “basic” unbiased estimators of θ

$$\begin{aligned} \hat{\theta}_m &= \frac{1}{1 - \frac{n-1}{n+1}k} X_{(1)} \\ \hat{\theta}_M &= \frac{1}{1 + \frac{n-1}{n+1}k} X_{(n)} \end{aligned} \quad (8)$$

with variances

$$\begin{aligned} \text{var}_{\theta}[\hat{\theta}_m] &= \frac{4nk^2\theta^2}{((1-k)(n-1) + 2)^2(n+2)} \\ \text{var}_{\theta}[\hat{\theta}_M] &= \frac{4nk^2\theta^2}{((1+k)(n-1) + 2)^2(n+2)}. \end{aligned} \quad (9)$$

These variances reveal that $X_{(n)}$ is strictly more informative about θ than $X_{(1)}$ throughout the range of n and k . Indeed, the variance of $\hat{\theta}_M$ is uniformly smaller than the variance of $\hat{\theta}_m$, even asymptotically:

$$\begin{aligned} \lim_{n \rightarrow \infty} n^2 \text{var}_{\theta}[\hat{\theta}_m] &= \frac{4k^2}{(1-k)^2} \theta^2 > \frac{4k^2}{(1+k)^2} \theta^2 \\ &= \lim_{n \rightarrow \infty} n^2 \text{var}_{\theta}[\hat{\theta}_M]. \end{aligned} \quad (10)$$

Step 2: When $k < 1$, the basic estimators $\hat{\theta}_m$ and $\hat{\theta}_M$ generate the more general family of unbiased estimators $\hat{\theta}_{mM}^{(\alpha)}$ of θ obtained as their affine combinations

$$\begin{aligned} \hat{\theta}_{mM}^{(\alpha)} &= (1-\alpha)\hat{\theta}_m + \alpha\hat{\theta}_M \\ &= \frac{1-\alpha}{1 - \frac{n-1}{n+1}k} X_{(1)} + \frac{\alpha}{1 + \frac{n-1}{n+1}k} X_{(n)} \end{aligned} \quad (11)$$

from which it is easy to identify the Rao-Blackwell improvement $\hat{\theta}_{RB}$ of X_1 as the case with $\alpha = \alpha_{RB} = \frac{1}{2}(1 + \frac{n-1}{n+1}k) > \frac{1}{2}$, and $\hat{\theta}_{LV}$ as the standard analytically derived case with

$$\alpha = \alpha_{LV} = \frac{\text{var}[\hat{\theta}_m] - \text{cov}[\hat{\theta}_m, \hat{\theta}_M]}{\text{var}[\hat{\theta}_m] + \text{var}[\hat{\theta}_M] - 2\text{cov}[\hat{\theta}_m, \hat{\theta}_M]} \quad (12)$$

and minimal variance

$$\begin{aligned} \text{var}[\hat{\theta}_{mM}^{(\alpha_{LV})}] &= \text{var}[(1-\alpha_{LV})\hat{\theta}_m + \alpha_{LV}\hat{\theta}_M] \\ &= \frac{\text{var}[\hat{\theta}_m]\text{var}[\hat{\theta}_M] - \text{cov}^2[\hat{\theta}_m, \hat{\theta}_M]}{\text{var}[\hat{\theta}_m] + \text{var}[\hat{\theta}_M] - 2\text{cov}[\hat{\theta}_m, \hat{\theta}_M]} \end{aligned} \quad (13)$$

that in the present case becomes

$$\alpha_{LV} = \frac{1}{2} \left(1 + \frac{2n}{(n-1)k + (n+1)\frac{1}{k}}\right), \quad (14)$$

which yields the following two representations for $\hat{\theta}_{LV}$

$$\begin{aligned} \hat{\theta}_{LV} &= \left(\frac{1}{2} - \frac{n}{(n-1)k + (n+1)\frac{1}{k}}\right)\hat{\theta}_m \\ &\quad + \left(\frac{1}{2} + \frac{n}{(n-1)k + (n+1)\frac{1}{k}}\right)\hat{\theta}_M \\ &= \frac{1}{2(k^2\frac{(n-1)}{(n+1)} + 1)} [(1-k)X_{(1)} + (1+k)X_{(n)}] \end{aligned} \quad (15)$$

with

$$\text{var}[\hat{\theta}_{LV}] = \frac{2k^2\theta^2}{((1+k^2)(n-1)+2)(n+2)}. \quad (16)$$

The asymptotic standard deviation $\sigma[\hat{\theta}_{LV}]$, proportional to $k\sqrt{\frac{2}{1+k^2}}$, is smaller than $\sigma[\hat{\theta}_M]$, proportional to $\frac{2k}{1+k}$. However, their asymptotic ratio ranges from only $\sqrt{2}$ (as $k \downarrow 0$) to 1 (as $k \uparrow 1$), which means that the two estimators $\hat{\theta}_M$ and $\hat{\theta}_{LV}$ are of comparable accuracy throughout the range, while $\hat{\theta}_m$ becomes infinitely worse as $k \rightarrow 1$.

We proceed now to analyze the Rao–Blackwell improvement $\hat{\theta}_{RB}$. From a probability point of view (i.e., when θ is a known constant), $E_\theta[X_1|X_{(1)}, X_{(n)}]$ treats $X_{(1)}$ and $X_{(n)}$ symmetrically. Indeed, the two are homoscedastic and stochastically equidistant from θ . From a statistics point of view (i.e., when *learning about* θ), $X_{(n)}$ is more informative than $X_{(1)}$, in the sense that θ is unbiasedly estimated more accurately by contracting $X_{(n)}$ than by expanding $X_{(1)}$. Thus, $\hat{\theta}_M$ should be intuitively expected to have a more decisive weight.

To see that indeed $\alpha_{RB} < \alpha_{LV}$, it is needed to check that $\frac{n-1}{n+1}k < \frac{2n}{(n-1)k+(n+1)\frac{1}{k}}$ or equivalently that $1 < \frac{2n}{(n-1)(1+k^2)\frac{n-1}{n+1}}$. But this last expression is monotone decreasing in k , with value $\frac{n+1}{n-1} > 1$ at $k = 1$.

The variance of $\hat{\theta}_{RB}$ can be expressed as

$$\text{var}[\hat{\theta}_{RB}] = \frac{2k^2\theta^2}{(n+2)(n+1)}. \quad (17)$$

The ratio $\frac{\text{var}[\hat{\theta}_{RB}]}{\text{var}[\hat{\theta}_{LV}]} = 1 + k^2\frac{(n-1)}{(n+1)}$ increases with n and, like $\frac{\text{var}[\hat{\theta}_M]}{\text{var}[\hat{\theta}_{LV}]}$, is bounded from above by 2.

2.1 A Remark on Minimal But NonComplete Sufficient Statistics

The scale-uniform example given illustrates that using the Rao–Blackwell theorem with a noncomplete minimal sufficient statistic on a crude initial (unbiased) estimator does not always yield an estimator with the lowest possible variance. And in fact, it never will. Any unbiased estimator that is a function of the minimal sufficient statistic is its own Rao–Blackwell improvement. If one such estimator has larger variance than another (such as $\hat{\theta}_m$ vs. $\hat{\theta}_M$), then the first has been improved, and if the two have equal variances, their average improves both. This is in essence the message in Torgersen’s converse to the Rao–Blackwell theorem (Torgersen 1988).

3. INEFFICIENCY OF THE MAXIMUM LIKELIHOOD ESTIMATOR

Under the usual differentiability assumptions, MLEs converge to the true parameter value at rate $n^{-\frac{1}{2}}$ *efficiently*, that is, with asymptotic variance that achieves the Crámer–Rao lower bound for the variance of unbiased estimators. In the model at hand these assumptions are violated and the parameter θ can be estimated, as seen above, with a faster rate of consistency n^{-1} .

For the scale-uniform family, the likelihood function $(\frac{1}{2k\theta})^n I_{\{\frac{X_{(n)}}{1+k} < \theta < \frac{X_{(1)}}{1-k}\}}$ is maximized at the minimal feasible θ value $\hat{\theta}_{MLE} = \frac{1}{1+k}X_{(n)}$, a slight departure from its unbiased correction $\hat{\theta}_M = \frac{1}{1+\frac{n-1}{n+1}k}X_{(n)}$ developed above. Thus, the asymptotic variance of $n(\hat{\theta}_{MLE} - \theta)$ is $\frac{4k^2}{(1+k)^2}\theta^2$, strictly bigger than the asymptotic variance $\frac{2k^2}{1+k^2}\theta^2$ achieved by $n(\hat{\theta}_{LV} - \theta)$, showing that MLE is inefficient in this case. The likelihood function discloses valuable deterministic information about θ , its feasible Low endpoint $L = \frac{X_{(n)}}{1+k}$ and High endpoint $H = \frac{X_{(1)}}{1-k}$. The MLE, $\hat{\theta}_{MLE} = L$, ignores the information provided by $X_{(1)}$ and pays a price for it in terms of efficiency.

The likelihood function exhibits a nice feature of the Rao–Blackwell improvement $\hat{\theta}_{RB} = \frac{1+k}{2}L + \frac{1-k}{2}H$: it is the *only* unbiased estimator of θ in the class $\hat{\theta}_{mM}^{(\alpha)}$ that is guaranteed to obtain values in the feasibility interval $[L, H]$, since it is the only one with coefficients adding up to 1 when viewed as a linear combination of L and H . Of course, feasibility is guaranteed for MLE ($= L$) and Bayesian estimators, the subject of the next section.

4. AN UNBIASED IMPROPER BAYES ESTIMATOR

In this section, we intend to complement the previous section by showing that, while MLE leads to the inefficient $\hat{\theta}_M$, Bayes estimation leads to a slight improvement over $\hat{\theta}_{LV}$, so the latter is a reasonably good estimator and the requirement of unbiasedness does not take a heavy toll. However, as will be seen, $\hat{\theta}_{LV}$ is uniformly dominated by another unbiased estimator, one that is not a linear combination of $X_{(1)}$ and $X_{(n)}$.

Letting ϑ stand for the Bayes-oriented random scale parameter and θ stand for its possible values, assume a smooth (improper) prior density $\lambda(\theta)$ on ϑ , proportional to θ^{-a} . As far as the posterior density of ϑ is concerned, it is only needed for its support, the feasibility interval $[L, H] = [\frac{X_{(n)}}{1+k}, \frac{X_{(1)}}{1-k}]$, where it becomes a proper posterior density from the first observation onward. The Bayes estimator of the scale parameter θ , the posterior expectation of ϑ , is

$$\begin{aligned} \hat{\theta}_{BAYES}^{(a)} &= \frac{\int_L^H \frac{\theta\lambda(\theta)}{(2k\theta)^n} d\theta}{\int_L^H \frac{\lambda(\theta)}{(2k\theta)^n} d\theta} \\ &= \frac{\int_L^H \frac{d\theta}{\theta^{n-1+a}}}{\int_L^H \frac{d\theta}{\theta^{n+a}}} = \frac{n-1+a}{n-2+a} \frac{L^{-(n-2+a)} - H^{-(n-2+a)}}{L^{-(n-1+a)} - H^{-(n-1+a)}} \\ &= \frac{n-1+a}{n-2+a} \left[1 - \frac{\frac{H}{L} - 1}{(\frac{H}{L})^{n-1+a} - 1} \right] L. \end{aligned} \quad (18)$$

The Bayes estimators under this conjugate bounded Pareto family are homogenous of order 1, that is, share with X_1 , \bar{X}_n , $\hat{\theta}_{RB}$, and $\hat{\theta}_{LV}$ their “respect” for the scale-parameter nature of θ , the property that the distribution of $\frac{\hat{\theta}}{\theta}$ is a pivotal quantity, that may depend on n and k but is independent of θ . Accordingly, their bias and standard deviation are constant multiples of θ . We show in the sequel that for the case $a = 2$, the Bayes estimator (henceforth $\hat{\theta}_{BAYES}^{(2)} = \hat{\theta}_{BAYES}$ for short) is unbiased, simultaneously on $n \geq 2$ and $k \in (0, 1)$.

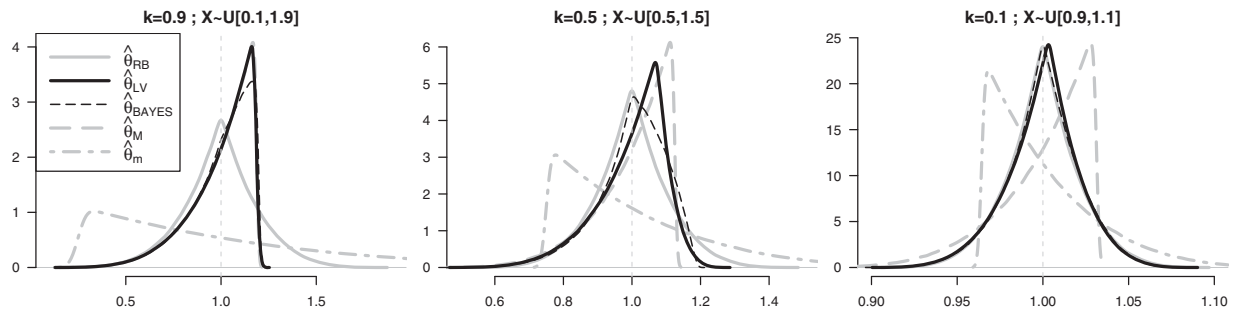


Figure 1. The simulated distribution (10^6 runs, sample size $n = 5$) of various unbiased estimators for $\theta = 1$ with design parameter $k = 0.9; 0.5; 0.1$.

The joint density of $X_{(1)}$ and $X_{(n)}$ is $f_{\theta}(x, y) = \frac{n(n-1)}{(2k\theta)^n} (y-x)^{n-2}$ on $[(1-k)\theta, (1+k)\theta]$. The expectation of $\hat{\theta}_{\text{BAYES}}$ can be evaluated for $\theta = 1$ as follows:

$$E[\hat{\theta}_{\text{BAYES}}] = \frac{n+1}{n} \left[E \left[\frac{X_{(n)}}{1+k} - \frac{n(n-1)}{(2k)^n} \int_{1-k}^{1+k} \int_x^{1+k} \frac{y}{1+k} \frac{\frac{1+k}{1-k} \frac{x}{y} - 1}{\left(\frac{1+k}{1-k} \frac{x}{y}\right)^{n+1} - 1} \times (y-x)^{n-2} dy dx \right] \right]. \quad (19)$$

The proof that $E[\hat{\theta}_{\text{BAYES}}] \equiv 1$ appears in the Appendix.

Figure 1 displays the empirical distribution of the various unbiased estimators of θ for sample size $n = 5$ from the uniform-scale distribution with $\theta = 1$ and $k = 0.9; 0.5; 0.1$.

The three panels display that $\hat{\theta}_{\text{RB}}$ (solid gray line) is symmetrically distributed around 1, with $\hat{\theta}_m$ (two-dash gray line) skewed to the right and $\hat{\theta}_M$ (long-dash gray line) skewed to the left. Both $\hat{\theta}_{\text{LV}}$ (solid line) and $\hat{\theta}_{\text{BAYES}}$ (thin dashed line) overlap $\hat{\theta}_M$ as $k \uparrow 1$ (leftmost panel), and approach $\hat{\theta}_{\text{RB}}$ as $k \downarrow 0$ (rightmost panel).

Since generalized Bayes rules need not be admissible, we cannot state from general principles that $\hat{\theta}_{\text{BAYES}}$ has lower variance than $\hat{\theta}_{\text{LV}}$. However, Figure 2 provides evidence that it does. Figure 2 displays for $\theta = 1$, all k that are multiples of $\frac{1}{50}$ and the choices $n = 3, 10, 25, 100$ of the sample size, the ratio of the empirical variance of $\hat{\theta}_{\text{BAYES}}$ (on 10^8 runs each, using the R software (R Core Team 2014)) to the theoretical variance of $\hat{\theta}_{\text{LV}}$ (16). These graphs are U-shaped, with value 1 toward the endpoints. The larger n is, the stronger the improvement of $\hat{\theta}_{\text{BAYES}}$ over $\hat{\theta}_{\text{LV}}$, but seemingly never by as much as 10%.

5. DISCUSSION

The concepts of minimal sufficiency and completeness of parametric families, at the basis of the theory of statistics, allow for the development of “optimal” statistical methods under proper sufficient conditions. The above example serves to point out the difficulty of exhibiting a “uniformly best estimator” in some settings. The purpose in presenting it is primarily pedagogical.

This article introduced the *scale-uniform* family of distributions $U((1-k)\theta, (1+k)\theta)$ (with unknown mean $\theta > 0$ and a known design parameter $k \in (0, 1)$). This family helps to illustrate the limitations of the Rao–Blackwell improvement when using a sufficient statistic that is minimal but not complete. It also serves to show that the maximum likelihood estimator may be inefficient for finite samples as well as asymptotically. Cases with inefficient MLE for finite samples are easily available: when estimating λ for the exponential distribution, the unbiased estimator $(\frac{n-1}{n} \frac{1}{\bar{X}})$ has lower variance than the MLE $(\frac{1}{\bar{X}})$. However, an *asymptotically* inefficient MLE is harder to come by.

The impressive progress in statistics in the last few decades has dimmed the role played by minimal sufficiency and completeness of parametric families in the development of statistical tools. Various nonparametric methods ranging from bootstrap and permutation tests to random forests and deep learning have emerged, playing now prominent roles in statistical practice. Nevertheless, the need for parametric methods continues to surface in this new era. For example, new online machine learning

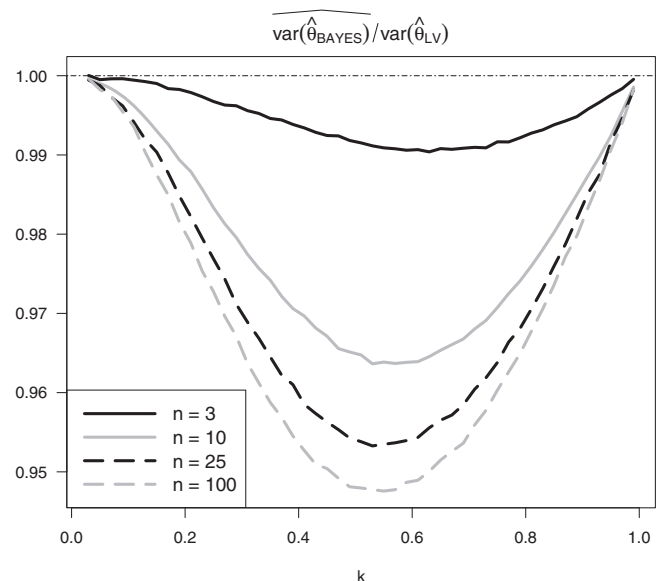


Figure 2. The ratio of the empirical variance of $\hat{\theta}_{\text{BAYES}}$ (on 10^8 runs each) to the theoretical variance of $\hat{\theta}_{\text{LV}}$ for various sample sizes over a range of k values.

methods are used in cases where the data become available in a sequential fashion. Since these methods need to be scalable for petabytes of information, data compression is essential (such as used by Google, Facebook, etc.). Another example is the use of map-reduce techniques that rely on summary statistics, either for efficiency in distributed computing, or for privacy-preserving properties (e.g., in hospital data, as is currently done in the medical informatics platform of the European Human Brain Project). A last example is of modern health care where the hope is to discover personalized medicines that can best treat the conditions of a specific individual. Such specificity is usually determined through inference, based on finding and analyzing a relatively small homogenous subsample of patients. In all three examples, the methods used often rely on minimal sufficient statistics, and whether this statistic is complete or not has consequences on the “optimal” properties of such methods. Hence, we would argue that while traditional statistical methods may have less of a need for old-school parametric assumptions, this need continues to resurface as we face new challenges in the modern era of massive, online, distributed, private, and personalized data.

A. APPENDIX: PROOFS

A.1 Proof of $\rho[Y_{(1)}, Y_{(n)}] = \frac{1}{n}$

For n iid $Y_i \sim U(a, b)$, the $n - 1$ observations except for $Y_{(1)}$ are conditionally iid given $Y_{(1)}$, distributed $U(Y_{(1)}, b)$. Therefore $E[Y_{(n)}|Y_{(1)}] = \frac{n-1}{n}b + \frac{1}{n}Y_{(1)}$. Since this conditional expectation is linear, its slope $\frac{1}{n}$ coincides with the slope of the linear regression that, for dependent and independent variables with equal variances is equal to their correlation coefficient.

A.2 Proof that $\hat{\theta}_{\text{BAYES}}$ is Unbiased

Apply $E[\frac{X_{(n)}}{1+k}] = 1 - \frac{2k}{(1+k)(n+1)}$, transform variables

$$\begin{aligned} u &= y - x ; v = \frac{x}{y} \\ x &= \frac{uv}{1-v} ; y = \frac{u}{1-v} \end{aligned} \quad (\text{A.1})$$

and evaluate the Jacobian as $|\det \frac{\partial(x,y)}{\partial(u,v)}| = \frac{u}{(1-v)^2}$. The integrand of the double integral in (19) is transformed into a function in which the new variable u appears only multiplicatively as u^n , and the inner integration over u can be easily performed once the endpoints of u are identified, for fixed v . For this purpose, u is represented in terms of v and x as $u = \frac{1-v}{v}x$. Then its lower endpoint $u_* = (1-k)\frac{1-v}{v}$ is obtained by the substitution $x = 1-k$ while its upper endpoint $u^* = (1+k)(1-v)$ is obtained from the equation resulting from the substitution $x = 1+k-v$. The inner integral is

$$\begin{aligned} \int_{u_*}^{u^*} u^n du &= \frac{1}{n+1}((u^*)^{n+1} - (u_*)^{n+1}) \\ &= \frac{1}{n+1} \left(\frac{1-v}{v}\right)^{n+1} ((1+k)^{n+1} v^{n+1} \\ &\quad - (1-k)^{n+1}). \end{aligned} \quad (\text{A.2})$$

Putting these elements together, (19) becomes

$$\begin{aligned} E[\hat{\theta}_{\text{BAYES}}] &= 1 + \frac{1-k}{1+k} \frac{1}{n} - \frac{(n-1)(1-k)^{n+1}}{(2k)^n(1+k)} \\ &\quad \times \int_{\frac{1-k}{1+k}}^1 \frac{\left(\frac{1+k}{1-k}v - 1\right)(1-v)^{n-2}}{v^{n+1}} dv, \end{aligned} \quad (\text{A.3})$$

which upon the rescaling $w = \frac{(1+k)v - (1-k)}{2k}$ becomes

$$\begin{aligned} E[\hat{\theta}_{\text{BAYES}}] &= 1 + \frac{1-k}{1+k} \frac{1}{n} - \frac{(n-1)(1-k)^n(1+k)}{(2k)^{n+1}} \\ &\quad \times \int_0^1 \frac{w(1-w)^{n-2}}{\left(w + \frac{1-k}{2k}\right)^{n+1}} dw. \end{aligned} \quad (\text{A.4})$$

For $E[\hat{\theta}_{\text{BAYES}}] = 1$ to hold (meaning that $\hat{\theta}_{\text{BAYES}}$ is unbiased), the identity

$$\int_0^1 \frac{w(1-w)^{n-2}}{\left(w + \frac{1-k}{2k}\right)^{n+1}} dw = \frac{(2k)^{n+1}}{n(n-1)(1-k)^{n-1}(1+k)^2} \quad (\text{A.5})$$

must hold. This formula appears explicitly in Gradshteyn & Ryzhik (Jeffrey and Zwillinger 2007, formula 3.197 (4)) as

$$\begin{aligned} \int_0^1 x^{\mu-1}(1-x)^{\nu-1}(1+ax)^{-\mu-\nu} dx \\ = (1+a)^{-\mu} B(\mu, \nu), \end{aligned} \quad (\text{A.6})$$

where B is the Beta function $B(\mu, \nu) = \int_0^1 t^{\mu-1}(1-t)^{\nu-1} dt$. Substituting $x = w$, $\mu = 2$, $\nu = n - 1$, $a = \frac{2k}{1-k}$ (after its inverse is taken out of the integral), identity (A.5) follows.

The above formula can also be accessed through *WolframAlpha.com* (LLC 2014).

[Received December 2014. Revised September 2015.]

REFERENCES

- Abramovich, F., and Ritov, Y. (2013), *Statistical Theory: A Concise Introduction*, Boca Raton, FL: CRC Press, pp. 37–40. [108]
- Berkson, J. (1955), “Maximum Likelihood and Minimum x 2 Estimates of the Logistic Function,” *Journal of the American Statistical Association*, 50, 130–162. [108]
- Bickel, P. J., and Blackwell, D. (1967), “A Note on Bayes Estimates,” *The Annals of Mathematical Statistics*, 38, 1907–1911. [108]
- Blackwell, D. (1947), “Conditional Expectation and Unbiased Sequential Estimation,” *The Annals of Mathematical Statistics*, 18, 105–110. [108]
- Damilano, G., and Puig, P. (2004), “Efficiency of a Linear Combination of the Median and the Sample Mean: The Double Truncated Normal Distribution,” *Scandinavian Journal of Statistics*, 31, 629–637. [109]
- Jeffrey, A., and Zwillinger, D. (2007), *Table of Integrals, Series, and Products*, New York: Academic Press, p. 317. [112]
- Lehmann, E. L., and Scheffé, H. (1950), “Completeness, Similar Regions, and Unbiased Estimation: Part i,” *Sankhyā: The Indian Journal of Statistics*, 10, 305–340. [108]
- (1955), “Completeness, Similar Regions, and Unbiased Estimation: Part ii,” *Sankhyā: The Indian Journal of Statistics (1933–1960)*, 15, 219–236. [108]
- C, W.A. (2014), ‘‘integrate (w*(1-w)^(n-2))/((w + (1-k)/(2k))^(n+1)) dw-wolfram|alpha,’’ available at <http://www.wolframalpha.com>. [112]

- Rao, C. R. (1945), "Information and Accuracy Attainable in the Estimation of Statistical Parameters," *Bulletin of the Calcutta Mathematical Society*, 37, 81–91. [108]
- R Core Team (2014), *R: A Language and Environment for Statistical Computing*, R Package Version 3.1.2, Vienna, Austria: R Foundation for Statistical Computing. [111]
- Stigler, S. M. (1973), "Studies in the History of Probability and Statistics. xxxii Laplace, Fisher, and the Discovery of the Concept of Sufficiency," *Biometrika*, 60, 439–445. [109]
- Torgersen, E. (1988), "On Bahadur's Converse of the Rao-Blackwell Theorem. Extension to Majorized Experiments," *Scandinavian Journal of Statistics*, 15, 273–280. [110]