The Annals of Statistics
2002, Vol. 30, No. 5, 1345–1375

SINGLE OBSERVATION UNBIASED PRIORS[1]

BY XIAO-LI MENG[2] AND ALAN M. ZASLAVSKY[3]

University of Chicago and Harvard University

This paper studies a class of *default* priors, which we call single observation unbiased priors (SOUP). A prior for a parameter is a SOUP if the corresponding posterior mean of the parameter based on a single observation is an unbiased estimator of the parameter. We prove that, under mild regularity conditions, a default prior for a convolution parameter is "noninformative" in the sense of yielding a posterior inference invariant under amalgamation only if it is a SOUP. Therefore, when amalgamation invariance is desirable, as in our motivating example of performing imputation for census undercount, SOUP is the only possible coherent "noninformative" prior for Bayesian predictions that will be utilized under aggregation. The use of SOUP also mutually calibrates Bayesian and frequentist inferences for aggregates of convolution parameters across many small areas. We describe approaches that identify SOUPs in many common models, in particular a constructive duality method that identifies SOUPs in pairs of distribution families. We introduce O-completeness, a necessary and sufficient condition for a prior distribution to be uniquely characterized by the corresponding posterior mean. Uniqueness of the SOUP is determined by the O-completeness of the dual family. O-completeness of a natural exponential family is implied by its completeness. Hence, the Diaconis–Ylvisaker characterization of the conjugate prior for natural exponential families via linear posterior expectation is a direct consequence of the completeness of such families. For most of the examples we have examined, the inverse of the variance function is the SOUP for the mean parameter of the corresponding family, suggesting that Hartigan's results on asymptotic unbiasedness can be generalized to some families with discrete parameters. We also discuss a possible extension of Berger's result on the inadmissibility of unbiased estimators, as the nonexistence of SOUP can be a first step in establishing such inadmissibility.

Received November 1999; revised November 2001.

[1]Supported in part by the Census Bureau through a contract with the National Opinion Research Center at the University of Chicago.

[2]Supported in part by NSF Grants DMS-92-04504, DMS-95-05043, DMS-96-26691 and DMS-00-72510.

[3]Supported in part by Joint Statistical Agreements 88-2, 89-8, 90-23 and 91-31 between the Bureau of the Census and Harvard University.

AMS 2000 subject classifications. Primary 62F15; secondary 62C15.

Key words and phrases. Admissibility, affine duality, amalgamation invariance, Bayes linear prediction, census undercount, completeness, O-completeness, conjugate priors, duality, generalized Bayes estimators, multiple imputation, beta, binomial, F, gamma, negative binomial, Poisson, location families, scale families.

1. Motivations.

1.1. *Imputing census undercount.* A census was conducted in an area, but some households were omitted due to undercounting. Given the observed count $Y = y$ of households in a census block, a small geographical area roughly corresponding to a city block, we need to impute the correct total, N, for that block. (Block indices are suppressed here.) The imputation procedure will be repeated independently several times to create multiple imputations [Rubin (1987)]. To do so, we need a model to link Y to N. We are willing to assume that omissions of households are independent, and thus the binomial model $Y \mid N \sim \mathrm{B}(N, p)$ is reasonable. The probability of inclusion p is either estimated from a distinct survey [Hogan (1993)], or, as in many applications of multiple imputation, p itself is first drawn from some posterior distributions; see Zaslavsky [(1989), part II] for the details of the imputation procedure for census undercount.

As posterior prediction is the only general procedure currently available for creating proper multiple imputations [Rubin (1987, 1996), Meng (1994)], we need a prior for N for each block. In many applications of this sort, the analyst imputes each block independently (conditional on p) and adopts a *default* prior for all blocks because the imputer only has vague information about N, relative to the information in the data, for individual blocks. In fact, when tens of thousands of blocks are involved, as in census applications (e.g., for congressional districts), and block boundaries are drawn in a largely arbitrary manner, constructing a plausible subjective prior for each individual block is impractical. (On the other hand, dependencies among blocks can be modeled through appropriate models for p, which represents a census coverage rate that is determined by the characteristics of housing units and residents of the block, and is not sensitive to the way block boundaries are drawn.) Furthermore, the operating properties of procedures using a common default prior and independence of blocks are easier to characterize than when many poorly constructed "real" priors are used. Thus, in such applications, a simple general purpose prior with good properties can be more suitable than a complicated prior constructed using a full hierarchical Bayes approach.

This raises the question of what default priors are suitable for such problems. As Kass and Wasserman's (1996) review illustrates, formal rules for selecting priors are based on theoretical properties, typically invariance of some sort, that are considered desirable for the problem at hand. Most questions of importance in a census concern aggregates of block counts over several or many blocks. It is thus desirable and important to require that the imputed block totals are consistent as the number of blocks gets large. (An adequate number of multiple imputations typically would be chosen so that the mean of the imputations for the area of interest approximates the posterior mean.) The easiest way of guaranteeing this consistency is to require the imputed total N^* to be an unbiased estimator for the true total N (for each block),

$$\mathbf{E}[N^* \mid N] = \mathbf{E}\big[\mathbf{E}[N^* \mid Y] \,\big|\, N\big] = N, \tag{1.1}$$

where the leftmost expectation is over both the undercount model and the imputation model, $P(N \mid Y = y)$. The unbiasedness of N^* is also desirable from a policy perspective. When benefits are distributed on the basis of population, a foreseeable bias may be more objectionable, and certainly less defensible, on grounds of fairness than an unknown error with mean zero [e.g., Schirm and Preston (1992), Section 4].

We call a prior distribution satisfying (1.1) a single observation unbiased prior (SOUP), emphasizing the small-sample exact criterion which it must satisfy. By reduction to sufficient statistics, this definition also applies to many situations with more than one observation. Since a posterior mean under a proper prior cannot be unbiased for the corresponding parameter [e.g., Lehmann (1983), Chapter 4] except for some pathological cases [e.g., Bickel and Mallows (1988)], we know that a SOUP is almost always improper. For our motivating example, we will show that there is exactly one SOUP, $\pi(N) \propto N^{-1}$, corresponding to the unbiased estimator Y/p. The corresponding posterior is the negative binomial $\mathrm{NB}(Y, p)$.

1.2. *Coherent noninformative priors for convolution parameters.* In the application described above, we have argued that the block boundaries are nearly uninformative about population distribution. This suggests that, as long as p is the same in several blocks, the inference should be the same regardless of whether we independently impute for each block and then add the imputed populations, or add the observed populations and then impute for the aggregated area. Otherwise the prior incorporates information about the meaning of the original block definitions, which we have argued should be regarded as completely arbitrary. Under regularity conditions, we can show that SOUP is the only possible coherent prior satisfying this amalgamation invariance property for any convolution parameter.

Consider $Y_i \overset{\text{indep}}{\sim} \mathrm{B}(N_i, p)$, $i = 1, 2$, with p known, and suppose the estimand is $N_+ = N_1 + N_2$. The amalgamation invariance property for a *default* prior π means that we can compute $P_\pi(N_i \mid Y_i, p)$, $i = 1, 2$, separately, and then obtain $P_\pi(N_+ \mid Y_1, Y_2; p)$ via convolution, or we can directly obtain $P_\pi(N_+ \mid Y_1 + Y_2, p)$. In other words, the operator of aggregating data commutes with the operator of aggregating the convolution parameters under the corresponding posterior.

Suppose π is such an amalgamation invariant prior for $B(N, p)$, and (i) that the corresponding posterior mean $m(y) = \mathbf{E}_\pi(N \mid Y = y)$ is finite for all $y \in \mathcal{N}^+ = \{0, 1, \dots, \}$ and (ii) that $m(y)$ is an asymptotically consistent estimator of N, namely, as $N \to \infty$, $m(y)/N \to 1$. Obviously, both (i) and (ii) are desirable minimum restrictions on π. Now since the posterior mean operator is linear, the amalgamation invariance property of π implies

$$m(y_1) + m(y_2) = m(y_1 + y_2) \qquad \forall\, y_1, y_2 \in \mathcal{N}^+, \tag{1.2}$$

which implies $m(y) = cy$, $\forall\, y \in \mathcal{N}^+$. The consistency requirement (ii) then implies that $c = 1/p$ and thus $m(Y) = Y/p$, the unbiased estimator of N. This means that the only candidate for π is the SOUP $\pi(N) \propto N^{-1}$. Since the resulting posterior is $\mathrm{NB}(Y, p)$, we see $\pi(N)$ is indeed amalgamation invariant.

This result applies to any univariate family $\{f(y \mid \theta),\ y \in \Omega;\ \theta \in \Theta\}$ such that

$$Y_j \mid \theta_j \overset{\text{indep}}{\sim} f(y \mid \theta_j),\ j = 1, 2, \quad \Longrightarrow \quad Y_1 + Y_2 \mid \theta_1, \theta_2 \sim f(y \mid \theta_1 + \theta_2), \tag{1.3}$$

and in particular to convolution families. For this generalization, we need the regularity condition that the only continuous additive functions [i.e., $g(\xi_1 + \xi_2) = g(\xi_1) + g(\xi_2)$] on either Ω or Θ are constant multiples [i.e., $g(\xi) = c\xi$]. This assumption is satisfied by common statistical models where Ω and Θ are R, R^+ or $\mathcal{N}^+$. Assuming $\mathbf{E}_\theta(y)$ is continuous in Θ (which is always the case when Θ is discrete), (1.3) implies $\mathbf{E}_\theta(Y) = c_1\theta$. The amalgamation invariance property of π then implies (1.2) with $\mathcal{N}^+$ replaced by Ω. Assuming $m(y)$ is continuous in y, (1.2) implies $m(y) = c_2 y$, and thus $\mathbf{E}_\theta[m(Y)] = c_1 c_2 \theta$. Consequently, if $m(y)$ is consistent for θ, then $c_2 = c_1^{-1}$, and hence π must be a SOUP.

If we do not insist on using the same form for $\pi(N_1)$, $\pi(N_2)$ and $\pi(N_+)$, then it may be possible to obtain amalgamation invariance using priors other than SOUP. But these are no longer *default* priors, as their construction will generally depend on information about the blocks, which we are supposing is not available. Thus the only *coherent default* prior under aggregation is the SOUP. However, like any formal rule for choosing prior distributions, SOUP should not be used in contexts for which it is not designed. A bowl of warm chicken soup is good for some souls, but certainly not for all.

1.3. *Background and overview.* With this motivation, this paper provides some general theory and methods for finding SOUP and studies existence and uniqueness of SOUP for some common models. For more theoretical studies of unbiased Bayes estimators, see Bickel and Mallows (1988) and Consonni and Veronese (1993). Our results differ from theirs in that we are concerned with the prior distribution only as a means to obtain a posterior prediction, while they are interested in constructing realistic joint distributions with the "unbiased" property, and therefore are concerned with the propriety of the priors. Another related literature concerns finding proper (or limiting improper) priors to show that a given estimator is generalized Bayes in order to prove admissibility, as reviewed in Rukhin (1995). Although our objectives are different, the theoretical results we present may play a complementary role for that purpose because the nonexistence of the SOUP can be a first step in establishing inadmissibility. There is also a large literature on formal rules for selecting prior distributions that have desirable properties, as reviewed by Kass and Wasserman (1996). But this work typically focuses on properties relevant to inferences about individual parameters, not on properties of aggregated predictions as in our motivating example and in many other multiple imputation applications.

The rest of the paper develops theory and examples for SOUPs. Section 2 describes some general methods for finding a SOUP, particularly a duality method which identifies SOUPs in pairs. Section 3 introduces the notion of O-completeness and demonstrates its use for assessing the uniqueness of a SOUP and for other prior determination problems. Sections 4–6 study SOUP, respectively, for some convolution families, exponential families, and scale and location families. Section 7 investigates SOUPs for transformations of various mean parameters. Section 8 presents several informative counterexamples and discusses related open problems, including connections between the nonexistence of SOUP and inadmissibility.

2. Methods for finding SOUPs.

2.1. *Definition and nonexistence.* Let $\mathcal{P}_\Theta = \{f(y \mid \theta), y \in \Omega, \theta \in \Theta\}$ be a family of densities with respect to μ, where $\Omega = \bigcup_{\theta \in \Theta}\{y : f(y \mid \theta) > 0\}$, $\pi(\theta)$ is a prior density of θ on Θ with respect to ν, and $\theta_\pi(y) = \mathbf{E}_\pi[\theta \mid Y = y]$ is the corresponding posterior mean of θ. We call π a SOUP for θ (under f) if

$$\mathbf{E}_f[\theta_\pi(Y) \mid \theta] = \theta \qquad \forall \theta \in \Theta, \tag{2.1}$$

where $\mathbf{E}_f$ is with respect to $f(y \mid \theta)$. Mathematically, (2.1) defines an integral equation for π with the constraints that $\pi(\theta) \geq 0$ for $\theta \in \Theta$ and (implicitly) that $\mathbf{E}_\pi(\theta \mid y)$ exists almost surely with respect to $f(y \mid \theta)$ (but the zero-measure set can depend on θ). The quadratic loss underlying (2.1) can be extended to more general loss functions with corresponding Bayes estimators [Hartigan (1965)]; see Section 7.

For given f and Θ, there may be no solution to (2.1). Below are several obvious situations where SOUP does not exist; less obvious examples appear in Section 8.

(I) Unbiased estimators do not exist (e.g., for the reciprocal of the Poisson mean parameter).

(II) Any unbiased estimator must assume values outside the parameter space. This occurs, for example, when the parameter space is restricted (e.g., $\theta \geq 0$) and $f(y \mid \theta)$ is not degenerate at its boundary (e.g., when $\theta = 0$); see Berger (1990) for examples and general theory.

(III) No improper prior can yield a proper posterior. An example is the equal mixture of $N(0, 1)$ and $N(2\theta, 1)$, namely, $f(y \mid \theta) = (\phi(y) + \phi(y - 2\theta))/2$ with $\Theta = R$, where $\hat{\theta} = Y \in R$ is unbiased for θ.

Even when a SOUP exists, directly solving (2.1) for π is generally difficult. One way to simplify is first to choose an unbiased estimator, say $\hat{\theta}(Y)$, and then to find a SOUP by solving $\mathbf{E}(\theta \mid y) = \hat{\theta}(y)$, that is,

$$\frac{\int_\Theta \xi f(y \mid \xi)\pi(\xi)\nu(d\xi)}{\int_\Theta f(y \mid \xi)\pi(\xi)\nu(d\xi)} = \hat{\theta}(y) \qquad \forall y \in \Omega. \tag{2.2}$$

Such a SOUP, while still *for* θ, we call more specifically "a SOUP *corresponding to* $\hat{\theta}(y)$." For complete families (e.g., regular exponential families), the unbiased estimator is unique (a.e.), so these two phrases carry the same meaning. An advantage of (2.2) is that, by selecting a desirable unbiased estimator, we can exclude pathological SOUPs.

Since solving (2.2) for π is the same as showing $\hat{\theta}(y)$ is a generalized Bayes estimator, there is a close relationship between the nonexistence of SOUP and the inadmissibility of $\hat{\theta}(Y)$, for admissible estimators are typically also generalized Bayes estimator [e.g., Brown (1971, 1979, 1980), Berger and Srinivasan (1978), Rukhin (1995)]. However, the two concepts are not equivalent. An obvious counterexample is the sample mean of a multivariate normal with dimension at least three (i.e., the Stein paradox), which is inadmissible although the constant prior is a SOUP (Section 6.2). Nevertheless, if $\hat{\theta}(Y)$ is inadmissible, then we should suspect that the corresponding SOUP does not exist and vice versa; see Section 8.

2.2. *Three simple methods for finding SOUPs.* Formally solving (2.2) can still be quite difficult in general. Fortunately, there are a number of simpler methods for finding SOUPs for parameters of many common models, with or without transformations, as summarized in Table 1 (discussed in Sections 4–6) and Table 2 (discussed in Section 7). In particular, for a continuous mean parameter of the natural observation from an exponential family, under regularity conditions, a SOUP is given by the Fisher information; see Section 5 for details. This fact was observed by Hartigan (1965), whose general result on asymptotically unbiased priors is also quite useful in *suggesting* candidates for SOUP for other continuous parameters of an exponential family, as explored in Section 7.

In general, we have identified the following three methods. First, in some cases, we can directly use the form of the likelihood $L(\theta \mid y) = f(y \mid \theta)$, much as when we seek a conjugate prior, to identify a posterior distribution $P(\theta \mid y)$ whose mean parameter is $\hat{\theta}(y)$. The corresponding SOUP is then given by $\pi(\theta) \propto P(\theta \mid y)/L(\theta \mid y)$. This is the approach we use for our motivating example (see Section 4.1).

This seemingly "tautological" method also suggested to us a more "mechanical" approach. The essence of this second approach is first to identify a density, not necessarily the correct posterior (as in the applications of Section 6), such that we can express all quantities in (2.2) as expectations with respect to this density. For all the cases listed in Table 1 and Table 2, we can transform (2.2) into the identity

$$\mathbf{Cov}_y[\alpha(\xi, y), \beta_\pi(\xi, y)] = 0 \qquad \forall\, y \in \Omega, \tag{2.3}$$

where α is a known function, β_π is a function involving the unknown π and $\mathbf{Cov}_y$ is with respect to a known density $p_y(\xi)$, which is free of π. An obvious solution for π in (2.3) is to let $\beta_\pi(\xi, y)$ be free of ξ and then to solve it for nonnegative π. When $p_y(\xi)$ identified is already the posterior under a SOUP, this method is just the first method.

TABLE 1
The SOUP for some common families

	Model	Parameter	Variance Function	SOUP	Unbiased estimator
1a	$Y \sim \mathrm{B}(\theta, p)$	$\theta \in \{0, 1, 2, \ldots\}$	$\theta p(1-p)$	θ^{-1}	Y/p
1b	$\theta \sim \mathrm{NB}(Y, p)$	$Y \in \{0, 1, 2, \ldots\}$	$Y(1-p)p^{-2}$	Y^{-1}	θp
2a	$Y \sim \mathrm{B}(n, \theta)$	$\theta \in [0, 1]$	$n\theta(1-\theta)$	$\theta^{-1}(1-\theta)^{-1}$	Y/n
2b	$\theta \sim \mathrm{Beta}(Y, n-Y)$	$Y \in \{0, 1, \ldots, n\}$	$Y(n-Y)[n^2(n+1)]^{-1}$	$Y^{-1}(n-Y)^{-1}$	$n\theta$
3a	$Y \sim \mathrm{Poisson}(\theta)$	$\theta \in (0, +\infty)$	θ	θ^{-1}	Y
3b	$\theta \sim \mathrm{Gamma}(Y, 1)$	$Y \in \{0, 1, \ldots\}$	Y	Y^{-1}	θ
4a	$Y \sim \mathrm{NB}(n, (1+\theta)^{-1})$	$\theta \in (0, +\infty)$	$n\theta(1+\theta)$	$\theta^{-1}(1+\theta)^{-1}$	$(Y-n)/n$
4b	$\theta \sim \chi^2_{2(Y-n)}/\chi^2_{2(n+1)}$	$Y \in \{n, n+1, \ldots\}$	$Y(Y-n)[n^2(n-1)]^{-1}$	$Y^{-1}(Y-n)^{-1}$	$n(\theta+1)$
5	$Y \sim f(Y-\theta)$	$\theta \in (-\infty, +\infty)$	$V(Y \mid \theta=0)$	Constant	$Y - E(Y \mid \theta=0)$
6	$Y \sim f(Y/\theta)/\theta, Y \geq 0$	$\theta \in (0, +\infty)$	$\theta^2 V(Y \mid \theta=1)$	θ^{-2}	$Y/E(Y \mid \theta=1)$

A third method for finding SOUPs is via a duality. When we find a SOUP $\pi(\theta)$ corresponding to $\hat{\theta}(y)$ under $f(y \mid \theta)$ by solving (2.2), we have also found a SOUP, $\pi^*(y) \propto \int f(y \mid \theta)\pi(\theta)\,d\theta$, for the *parameter* $\hat{\theta}(y)$ under the *dual family* $f^*(\theta \mid y) = P(\theta \mid y)$. In other words, if we switch the roles of y and θ and regard the posterior $P(\theta \mid y)$ under SOUP $\pi(\theta)$ as a sampling family $f^*(\theta \mid y)$, then the marginal density of y, $\pi^*(y)$, is a SOUP *corresponding* to θ, the unbiased estimator of $\hat{\theta}(y)$ because $\mathbf{E}_{f^*}(\theta \mid y) = \hat{\theta}(y)$ and (2.1) can be reexpressed as the counterpart of (2.2) for the dual family

$$\frac{\int_\Omega \hat{\theta}(y) f^*(\theta \mid y)\pi^*(y)\mu(dy)}{\int_\Omega f^*(\theta \mid y)\pi^*(y)\mu(dy)} = \theta \qquad \forall \theta \in \Theta. \tag{2.4}$$

Because of this duality, our presentation is made in terms of "dual" pairs. The first eight examples in Table 1 constitute four dual pairs, while the last two are self-dual, and Table 2 consists of seven pairs. This duality method is very effective for finding SOUPs for models with continuous variables but discrete parameters, because such a model is a dual model of a discrete model with continuous parameters, for which a SOUP may readily be available, possibly by Hartigan's method (see Section 5).

2.3. *Affine duality.* In all examples in Table 1, our unbiased estimator $\hat{\theta}(Y)$ is linear in Y. Consequently, the duality described in Section 2.2 implies the following *affine duality*:

$$\begin{aligned} &\mathbf{E}_f[Y \mid \theta] = c_1\theta + c_2 \qquad \forall \theta \in \Theta \\ &\text{and} \\ &\mathbf{E}_\pi[\theta \mid Y] = (Y - c_2)/c_1 \qquad \forall Y \in \Omega, \end{aligned} \tag{2.5}$$

where c_1 and c_2 are constants. This affine duality is stronger than the SOUP property, and can be made false for any SOUP by applying an arbitrary nonlinear transformation to Y, although a one-to-one transformation of Y does not affect the SOUP property itself. Conversely, if π is a SOUP and the transformation $Y \to \theta_\pi(Y) \equiv \mathbf{E}_\pi[\theta \mid Y]$ is one-to-one, then an affine dual pair is obtained by replacing Y with $\theta_\pi(Y)$; this transformation defines a natural scale for the observation. (In most of our examples, however, we leave Y on the scale of the familiar distributional form, which is a linear transformation of the unbiased estimator.) This argument suggests that affine duality applies in most cases.

If $Y \to \theta_\pi(Y)$ is not required to be one-to-one, it is possible to construct examples of SOUP families in which $\theta_\pi(y_1) = \theta_\pi(y_2)$ but $P(\theta \mid y_1) \neq P(\theta \mid y_2)$, so imputations of θ based on two distinct observed values of Y agree in expectation but not in distribution. For example, define a family of distributions on the integers by $f(y \mid \theta) = 1/2$ for θ odd and $y = \theta$ or $y = \theta - 2$, $f(y \mid \theta) = 1$ for θ even and $y = \theta$, and $f(y \mid \theta) = 0$ otherwise. The constant prior on the integers is a SOUP. Then for y any even number, $\theta_\pi(y) = \theta_\pi(y-1) = y$, but the posterior distributions given y or $y-1$ are supported, respectively, on $\{y\}$ or $\{y-1, y+1\}$. In such cases it would be inferentially invalid to reduce Y to $\theta_\pi(Y)$.

Affine duality obviously implies $\mathbf{E}_\pi\{\mathbf{E}_f[Y \mid \theta] \mid Y\} = Y$, $\forall Y \in \Omega$, which is a special case of linearity of the posterior expectation of the mean parameter investigated by Diaconis and Ylvisaker (1979):

$$\mathbf{E}_\pi\big\{\mathbf{E}_f[Y \mid \theta] \big| Y\big\} = aY + b \qquad \forall Y \in \Omega. \tag{2.6}$$

Diaconis and Ylvisaker (1979) establish that if Y is from a continuous regular exponential family, then only conjugate prior densities on the natural parameter satisfy (2.6). They also establish similar results for discrete Y, still from a regular exponential family, when the (natural) parameter space is an open interval $(-\infty, \theta_0)$, where $\theta_0 < +\infty$. Their results, however, are not directly applicable to our setting because they were characterizing *proper* priors (and thus $a \neq 1$ as implied by their results). Our SOUPs for exponential families may be obtained, however, as limits of these (conjugate) proper priors as the number of prior observations approaches zero; that is, as $n_0 \to 0$ in the notation of Diaconis and Ylvisaker (1979).

3. The uniqueness of SOUP and *O*-completeness.

3.1. *A necessary and sufficient condition*: *O-completeness.* Once a SOUP corresponding to an unbiased estimator is found, its uniqueness is of both theoretical and practical interest; no other considerations are needed to select a prior if it is unique. Showing the uniqueness of the SOUP characterizes priors by posterior means, in the spirit of Diaconis and Ylvisaker (1979). Hartigan (1965) gives asymptotic uniqueness results, which can be used to deduce uniqueness in some finite-sample cases under additional assumptions (see Section 5.2). It is not surprising that whether a prior can be uniquely determined by posterior means depends on the richness of the posterior family, namely, a certain kind of "completeness." To be precise, we introduce the following notion.

DEFINITION 1. A density/probability family $\mathscr{P}_\Lambda = \{P(Z \mid \lambda),\ \lambda \in \Lambda\}$, where Z can be of any dimension and $\mathbf{E}_\lambda(Z)$ exists for all $\lambda \in \Lambda$, is called O-complete if for any nonnegative real-valued function $g(Z)$ satisfying $\mathbf{E}_\lambda(g(Z)) > 0$, $\forall \lambda \in \Lambda$, $\mathbf{Cov}_\lambda(g(Z), Z) = 0$ for all λ implies $g(Z) \equiv$ constant (a.e. $\mathscr{P}_\Lambda$); that is, there exists a constant C and a set A such that $P_\lambda(A) = 1$ for all $\lambda \in \Lambda$ and $g(z) = C$ for all $z \in A$.

In other words, O-completeness means that, except for the trivial cases of constant functions, there cannot be any fixed-sign function of Z that is uncorrelated (i.e., orthogonal, hence the term "O-completeness") with Z and has nonzero mean for all λ. The following general result, which goes beyond the uniqueness of SOUP, shows that the O-completeness is a necessary and sufficient condition for the posterior mean function to uniquely determine the prior density, proper or improper.

THEOREM 1. *Suppose for a sampling family $\mathscr{P}_\Theta = \{f(y \mid \theta),\ y \in \Omega, \theta \in \Theta\}$ with respect to μ, we have a posterior family $\mathscr{P}_\Omega^\pi = \{P_\pi(\theta \mid y) \propto f(y \mid \theta)\pi(\theta),\ y \in \Omega\}$ with respect to ν whose posterior mean $\mathbf{E}_\pi(\theta \mid y)$ exists for all $y \in \Omega$, where $\pi(\theta)$ is a proper or improper prior density with support $\Theta_\pi \subset \Theta$. Then $\pi(\theta)$ is uniquely determined by $\{\mathbf{E}_\pi(\theta \mid y),\ y \in \Omega\}$ among all $\tilde{\pi} \in \{\tilde{\pi}(\theta) : \Theta_{\tilde{\pi}} \subseteq \Theta_\pi\}$, up to a proportionality constant (a.e. $\mathscr{P}_\Omega^\pi$), if and only if $\mathscr{P}_\Omega^\pi$ is O-complete.*

PROOF. Suppose for a $\tilde{\pi} \in \{\tilde{\pi}(\theta) : \Theta_{\tilde{\pi}} \subseteq \Theta_\pi\}$, $\mathbf{E}_{\tilde{\pi}}(\theta \mid y) = \mathbf{E}_\pi(\theta \mid y)$ for all $y \in \Omega$. Then, by reexpressing $\mathbf{E}_{\tilde{\pi}}(\theta \mid y)$ in terms of $\mathbf{E}_\pi$ via importance sampling, we obtain

$$\mathbf{E}_\pi\left(\frac{\tilde{\pi}(\theta)}{\pi(\theta)}\theta \,\Big|\, y\right) = \mathbf{E}_\pi\left(\frac{\tilde{\pi}(\theta)}{\pi(\theta)} \,\Big|\, y\right)\mathbf{E}_\pi(\theta \mid y) \qquad \forall y \in \Omega. \tag{3.1}$$

Since $\mathbf{E}_\pi(\tilde{\pi}(\theta)/\pi(\theta) \mid y) > 0$ for all $y \in \Omega$, if $\mathscr{P}_\Omega^\pi$ is O-complete, then $\tilde{\pi}(\theta) \propto \pi(\theta)$ (a.e. $\mathscr{P}_\Omega^\pi$).

On the other hand, if $\mathscr{P}_\Omega^\pi$ is not O-complete, then there exists a nonnegative real-valued function $\eta(\theta)$ such that $\mathbf{E}_\pi(\eta(\theta) \mid y) > 0$ and

$$\frac{\mathbf{E}_\pi(\eta(\theta)\theta \mid y)}{\mathbf{E}_\pi(\eta(\theta) \mid y)} = \mathbf{E}_\pi(\theta \mid y) \qquad \forall y \in \Omega, \tag{3.2}$$

but for any constant C, $P_\pi(\eta(\theta) = C \mid y_C) < 1$ for some $y_C \in \Omega$. It follows then that the use of either $\pi(\theta)$ or of $\tilde{\pi}(\theta) \equiv \pi(\theta)\eta(\theta)$ leads to the same posterior mean $\mathbf{E}_\pi(\theta \mid y)$ for all $y \in \Omega$, although $\pi(\theta)/\tilde{\pi}(\theta)$ is not almost surely constant with respect to $P_\pi(\theta \mid y_C)$ and thus to $\mathscr{P}_\Omega^\pi$. Necessity thus follows. □

The restriction $\Theta_{\tilde{\pi}} \subseteq \Theta_\pi$ in Theorem 1 is a theoretical necessity, but is automatic for all the SOUPs found in this paper, where $\Theta_\pi = \Theta$. The need for this condition is illustrated by considering $\mathcal{P}_\Theta = \{\mathrm{B}(1, p_\theta),\ \theta = 0, 1, 2\}$, where p_0, p_1, $p_2 \in (0, 1)$ are three distinct numbers with $p_1 = 1/2$. Suppose $\pi_1(\theta)$ concentrates on $\{0, 1\}$ with prior odds $O_1 = \pi_1(0)/\pi_1(1)$. The corresponding posterior for θ is $\mathrm{B}(1, \alpha(y))$, where $\alpha(y) = (1 + 2O_1 p_0^y (1 - p_0)^{1-y})^{-1}$. Since for any $Z \mid p \sim \mathrm{B}(1, p)$, $\mathbf{Cov}_p(g(Z), Z) = p(1 - p)[g(1) - g(0)]$, $\mathrm{B}(1, \alpha(y))$ is O-complete for both $y = 0$ and $y = 1$ (more than we need). Consequently, by Theorem 1, $\pi_1(\theta)$ is the only prior with $\{\mathbf{E}_\pi(\theta \mid y) \equiv \alpha(y), y = 0, 1\}$ among all π's on Θ such that $\pi(2) = 0$. However, if we let $\pi_2(\theta)$ be a prior that concentrates on $\{0, 2\}$ with prior odds $O_2 = \pi_2(0)/\pi_2(2)$, then one can verify that $\mathbf{E}_{\pi_2}(\theta \mid y) = \mathbf{E}_{\pi_1}(\theta \mid y)$ for both $y = 0$ and $y = 1$ as long as

$$O_1 = \frac{p_0 - p_2}{4p_0(1 - p_0)(2p_2 - 1)} \quad \text{and} \quad O_2 = \frac{p_2(1 - p_2)(2p_0 - 1)}{p_0(1 - p_0)(2p_2 - 1)},$$

where we choose $p_0 > p_2 > 1/2$ or $p_0 < p_2 < 1/2$.

3.2. *Connecting O-completeness with completeness.* Since both notions are about richness of a probability family, not surprisingly, there is a close relationship between O-completeness and the standard notion of completeness of the same family [e.g., Lehmann (1983)]. Clearly, if $P(Z \mid \lambda)$ is such that for any absolutely continuous function $g(Z)$ [Hudson (1978)],

$$\mathbf{E}_\lambda\big[g(Z)\big(Z - \mathbf{E}_\lambda(Z)\big)\big] = \mathbf{E}_\lambda[\alpha(Z)g'(Z)] \tag{3.3}$$

holds, where $\alpha(z)$ is a positive (a.e.) real-valued function, then O-completeness is implied by completeness if we restrict g in Definition 1 to be absolutely continuous. In fact, when Z is a natural observation from an exponential family, we have the following general result.

THEOREM 2. *Let $\mathcal{P}_\Lambda = \{f(z \mid \lambda) = a(z)\exp(c^\top(\lambda)z - b(\lambda));\ z \in \Omega \subset R^k$, $\lambda \in \Lambda \subset R^d\}$ be a family of densities with respect to μ, where Λ is such that for any differentiable function $h(\lambda)$ on Λ, $\frac{\partial h(\lambda)}{\partial \lambda} = 0$, $\forall \lambda \in \Lambda$, implies $h(\lambda) =$* constant $\forall \lambda \in \Lambda$. *Suppose the score function $S(\lambda \mid z) = \frac{\partial c^\top(\lambda)}{\partial \lambda} z - \frac{\partial b(\lambda)}{\partial \lambda}$ is well defined for all $\lambda \in \Lambda$. Then the O-completeness of $\mathcal{P}_\Lambda$ is implied by its completeness.*

PROOF. Under the assumptions of the theorem, for any real-valued function $g(y)$ such that $\mathbf{Cov}_\lambda(g(Z), Z)$ exists for all $\lambda \in \Lambda$,

$$\frac{\partial \mathbf{E}_\lambda[g(Z)]}{\partial \lambda} = \mathbf{E}_\lambda[g(Z)S(\lambda \mid Z)] = \frac{\partial c^\top(\lambda)}{\partial \lambda}\mathbf{E}_\lambda[g(Z)Z] - \frac{\partial b(\lambda)}{\partial \lambda}\mathbf{E}_\lambda[g(Z)]. \tag{3.4}$$

Taking $g(z) \equiv 1$ in (3.4) yields $\frac{\partial b(\lambda)}{\partial \lambda} = \frac{\partial c^\top(\lambda)}{\partial \lambda}\mathbf{E}_\lambda(Z)$; combining this with (3.4), we obtain

$$\frac{\partial \mathbf{E}_\lambda[g(Z)]}{\partial \lambda} = \frac{\partial c^\top(\lambda)}{\partial \lambda}\mathbf{Cov}_\lambda\big(g(Z), Z\big). \tag{3.5}$$

Consequently, if $\mathbf{Cov}_\lambda(g(Z), Z) = 0$, $\forall \lambda \in \Lambda$, then $\mathbf{E}_\lambda[g(Z) - C] = 0$, $\forall \lambda \in \Lambda$, which implies $g(z) = C$ (a.s. $\mathcal{P}_\Lambda$) if $\mathcal{P}_\Lambda$ is complete. $\square$

This result is more generally applicable than using (3.3). By taking the derivative of $\mathbf{E}_\lambda[g(Z)]$ with respect to λ, we avoid requiring $g(z)$ to be absolutely continuous; indeed, Z can even be a discrete variable. Combined with Theorem 1, it also provides a simpler proof of Diaconis and Ylvisaker's (1979) well-known characterization of conjugate priors for the natural parameter θ of a continuous exponential family (their Theorem 3). Since (2.6) holds trivially when π is conjugate and the natural parameter is the natural observation of the posterior family of θ, Theorems 1 and 2 imply that conjugate priors are the only possible priors for (2.6) to hold when the posterior family is complete. In fact, a key part of Diaconis and Ylvisaker's (1979) proof largely parallels the proof of the completeness of a natural exponential family when the parameter space contains an open interval in R^d (i.e., their condition that μ contains an open interval in R^d, as the sample space corresponds to the parameter space for the posterior family).

However, Theorem 2 does not cover the nonexponential families, nor is it useful for establishing the O-completeness of $\mathcal{P}_\Lambda$ when it is not complete (see Sections 5.2–5.4). For nonexponential families, sometimes it is possible to establish useful identities resembling (3.5), for example, Lemmas 1 and 2 of Section 4.1. When a family is not complete, an effective method for determining O-completeness is to turn $\mathbf{Cov}_\lambda(g(Z), Z) = 0$ into a set of moment conditions for an unknown density depending on g, and then to invoke well-known results on determining a density via its moments; see Sections 5.2–5.4.

Although we are unable to prove or disprove the conjecture that completeness implies O-completeness in general, it is easy to construct a family that is O-complete but not complete. Besides the discrete parameter cases in Sections 5.2–5.4, the following simple example is particularly intriguing. Consider $\mathcal{P} = \{P_N : N \geq 1\}$, where P_N is the uniform distribution on $\{1, \ldots, N\}$. While $\mathcal{P}$ is complete, Stigler (1972) showed that $\mathcal{P} \setminus \{P_n\} \equiv \{P_N : N \geq 1,\ N \neq n\}$ is not complete for any $n \geq 1$. However, if $\mathbf{Cov}_N(g(Z), Z) = 0$, $\forall N \geq 2$, then since $N\mathbf{Cov}_N(g(Z), Z) - (N+1)\mathbf{Cov}_{N+1}(g(Z), Z) = 0$, we obtain

$$g(N+1) = \frac{1}{N}\sum_{k=1}^{N} g(k), \qquad N \geq 2. \tag{3.6}$$

Since $\mathbf{Cov}_N(g(Z), Z) = 0$ for $N = 2$, which implies $g(2) = g(1)$, we have from (3.6) that $g(Z) = g(1)$ for any positive integer Z. Consequently, $\mathcal{P} \setminus \{P_1\}$ is O-complete, although it is not complete. It is interesting to note, however, that $\mathcal{P} \setminus \{P_1, P_2\}$ is not O-complete because without the restriction induced by $N = 2$, any $g(Z)$ satisfying $g(k) = (2g(1) + g(2))/3$, $\forall k > 4$ and $g(3) = g(1)$ is uncorrelated with Z for any $N \geq 3$. Thus, like completeness, O-completeness is a property of a family, not a property of a statistic or of the parametric form, which was Stigler's (1972) key point and motivation in presenting his example.

4. Convolution families. This section provides a theoretical study of SOUPs for our motivating example. It also reveals some general desirable properties of SOUP for convolution parameters. The Poisson–Gamma pair is a regular exponential family as well as a convolution family, and thus its exposition is deferred to Section 5.3.

4.1. *The* $\mathrm{B}(\theta, p)$ *and* $\mathrm{NB}(Y, p)$ *pair.* Suppose $Y \mid \theta \sim \mathrm{B}(\theta, p)$ with known $p \in (0, 1)$, where $\theta = N$ in Section 1. The form of the likelihood function $L(\theta \mid Y) = \binom{\theta}{Y} p^Y (1-p)^{\theta - Y}$ suggested trying the negative binomial $\mathrm{NB}(Y, p)$ as a candidate posterior density, $P(\theta \mid Y) = \binom{\theta-1}{Y-1} p^Y (1-p)^{\theta-Y}$, $\theta \geq Y$. Since the mean of $\mathrm{NB}(Y, p)$, Y/p, is indeed an unbiased estimator of θ, we obtain $\pi(\theta) = P(\theta \mid Y)/L(\theta \mid Y) \propto \theta^{-1}$, $\theta \in \mathcal{N}^+ \equiv \{0, 1, 2, \ldots\}$ as our SOUP. Note that the "infinite" mass in the prior at $\theta = 0$ enters the calculation only when $Y = 0$, in which case it implies that $P(\theta = 0 \mid Y = 0) = 1$, meaning that $\mathrm{NB}(0, p)$ is defined to be a point mass at $\theta = 0$. In our census example, this implies that no households are imputed into areas which showed no population, a desirable property as any number of such areas could be arbitrarily defined anywhere. This is a direct consequence of the amalgamation invariance of Section 1.2. More fundamentally, it reflects our desire to be *noninformative*: if we have no prior information about the areas then it is not possible to impute anything other than zero for an area with no enumerated households.

By affine duality, $\pi^*(Y) \propto \sum_{\theta=Y}^{\infty} f(Y \mid \theta)\pi(\theta) \propto Y^{-1}$ is a SOUP for the Y parameter of $\mathrm{NB}(Y, p)$. To verify, respectively, the uniqueness of $\pi(\theta)$ for $\mathrm{B}(\theta, p)$ and of $\pi^*(Y)$ for $\mathrm{NB}(Y, p)$, we first note that both $\mathcal{P}_{\mathrm{B}}(p) = \{\mathrm{B}(\theta, p) : \theta \in \mathcal{N}^+\}$ and $\mathcal{P}_{\mathrm{NB}}(p) = \{\mathrm{NB}(Y, p) : Y \in \mathcal{N}^+\}$ are complete for fixed $p \in (0, 1)$. Second, both families are O-complete. (Note that Theorem 2 is not applicable here because neither of them is an exponential family when the parameter of interest is the convolution parameter, not p.) For $\mathcal{P}_{\mathrm{B}}(p)$, O-completeness follows immediately from its completeness and the following result.

LEMMA 1. *Let* $\mathbf{E}_{n,p}$ *denote expectation with respect to* $\mathrm{B}(n, p)$, *where* $n \geq 1$ *and* $0 \leq p \leq 1$. *Let* $\mathbf{Cov}_{n,p}$ *and* $\mathbf{V}_{n,p}$ *accordingly denote covariance and variance. Then for any real-valued function* g,

$$\mathbf{Cov}_{n,p}\big(g(X), X\big) = \mathbf{V}_{n,p}(X) \times \mathbf{E}_{n-1,p}[g(X+1) - g(X)]. \tag{4.1}$$

Similarly, for $\mathcal{P}_{NB}(p)$, the O-completeness is implied by its completeness and the following identity.

LEMMA 2. *Let* $\mathbf{E}_{r,p}$ *denote expectation with respect to* $\mathrm{NB}(r, p)$, *where* $r \geq 1$ *and* $0 < p \leq 1$. *Let* $\mathbf{Cov}_{r,p}$ *and* $\mathbf{V}_{r,p}$ *accordingly denote covariance and variance.*

Then for any real-valued function g,

$$\mathbf{Cov}_{r,p}\big(g(X), X\big) = \mathbf{V}_{r,p}(X) \times \mathbf{E}_{r+1,p}[g(X) - g(X-1)], \tag{4.2}$$

as long as $\mathbf{E}_{r,p}[g(X)X]$ *exists.*

These identities involve a shift in the convolution parameter, and thus are different from those previously found [e.g., Hudson (1978), Hwang (1982)]; see the Remark in Appendix A, where we prove these lemmas. Note for $\mathcal{P}_{NB}(p)$, $g(0)$ is not regulated by O-completeness because $X = 0$ enters (4.2) only when $r = 0$, in which case $\mathbf{V}_{r,p}(X) = r(1-p)/p^2 = 0$. But this does not affect the uniqueness of our SOUP because $\pi(0)$ must be infinite in order for $\mathbf{E}_\pi(\theta \mid Y = 0) = 0$, a desirable property as discussed earlier.

4.2. *Mutually calibrating Bayesian and frequentist inferences.* Consider an inference for $\theta \in \Theta$, the convolution parameter of $f(Y \mid \theta)$; in our census example, $\theta = N$. Given a prior π, multiple imputation with infinitely many imputations gives a posterior interval for θ of the form $\mathbf{E}_\pi(\theta \mid Y) \pm z_{\alpha/2}\sqrt{\mathbf{V}_\pi(\theta \mid Y)}$ [Rubin (1987)]. In order for this interval to have approximately the nominal frequency coverage $1 - \alpha$, we need $\mathbf{V}_\pi(\theta \mid Y)$ to be a reasonable estimator of $\mathbf{V}_f[\mathbf{E}_\pi(\theta \mid Y) \mid \theta]$. Standard large-sample theory guarantees this when $\theta \to \infty$ for most choices of π. However, this does not guarantee approximate equality for small θ, and more problematically, not even for sums of many small θ_b's each of which is based on a small sample, an important consideration for our census application.

However, when π is a SOUP corresponding to $\hat{\theta}(Y) = aY$, where a is a constant, and when the corresponding posterior family $P(\theta \mid Y)$ is also a convolution family, as in Section 4.1, $\mathbf{V}_\pi(\theta \mid Y)$ is an unbiased estimator of $\mathbf{V}_f[\mathbf{E}_\pi(\theta \mid Y) \mid \theta]$. This is because for any convolution family both the mean and variance parameters are proportional to the convolution parameter, and thus $\mathbf{V}_\pi(\theta \mid Y) \propto \mathbf{E}_\pi(\theta \mid Y)$ as a function of Y. Consequently, $\mathbf{E}_f[\mathbf{V}_\pi(\theta \mid Y) \mid \theta] \propto \theta$ when π is a SOUP. On the other hand, because of the same convolution property of $f(Y \mid \theta)$, $\mathbf{V}_f[\mathbf{E}_\pi(\theta \mid Y) \mid \theta] = \mathbf{V}_f(aY \mid \theta) \propto \theta$. This implies that

$$\mathbf{E}_f[\mathbf{V}_\pi(\theta \mid Y) \mid \theta] = \mathbf{V}_f[\mathbf{E}_\pi(\theta \mid Y) \mid \theta] \qquad \forall \theta \in \Theta, \tag{4.3}$$

because the ratio of the two sides approaches unity when $\theta \to \infty$, under the assumption that π leads to consistent inference. Similarly, we obtain the "dual" form of (4.3),

$$\mathbf{E}_\pi[\mathbf{V}_f(Y \mid \theta) \mid Y] = \mathbf{V}_\pi[\mathbf{E}_f(Y \mid \theta) \mid Y] \qquad \forall Y \in \Omega. \tag{4.4}$$

For the pair in Section 4.1, both sides of (4.3) are $\theta(1-p)/p$, and both sides of (4.4) are $(1-p)Y$.

Identity (4.3) implies that multiple imputation inferences for aggregates over many imputed values will be calibrated, even if each imputed value is small.

On the other hand, if a non-SOUP prior is used for the imputation and yields systematically smaller posterior variances, its coverage must be dependent on the correctness of the prior information, including (when small areas are being aggregated) the correctness of any assumptions of conditional independence; these may be hard to verify.

While (4.3) addresses the frequentist analyst's concerns about the calibration of Bayesian inferences, (4.4) can be related to the Bayesian's evaluation of the frequentist's typical procedures. The left-hand side of this equation is the Bayesian's posterior expectation of the sampling variance of Y, which the frequentist typically estimates by substituting his unbiased estimate of θ into the relationship of variance to mean. The right-hand side of (4.4) is the posterior variance of θ since the inner expectation is the expectation of the unbiased estimator (assuming without loss of generality that $a = 1$). The equation therefore says that the Bayesian who uses the SOUP can regard the frequentist's inference for the variance as an acceptable plug-in estimate of his posterior variance. In other words, under SOUP, the Bayesian and frequentist inferences for the convolution parameter are mutually calibrated when aggregating over small areas or samples.

While (4.3) also holds for any location family with flat SOUP (see Section 6.2), it does not necessarily hold for other nonconvolution families even when a SOUP is used. In fact, for scale families (Section 6.1) the posterior variance, $\mathbf{V}_\pi(\theta \mid Y)$, need not even be finite.

5. Exponential families. Several methods are available for exponential families, including the inverse variance rule described in Section 5.1, the duality method, and the method based on (2.3). We save the last method for Section 6, where the first two methods are not directly applicable. [Note also that Hartigan's (1998) maximum likelihood prior is a SOUP when the MLE is an unbiased estimator.] The normal distribution is omitted here as it is covered by the results in Section 6.

5.1. *The inverse variance rule for continuous exponential families.* For the mean parameter of the natural observation of a one-dimensional exponential family $\{f(y \mid \phi) = a(y)\exp(\phi y - b(\phi)) : \phi \in \Phi\}$, where Φ is an interval (open, closed, or half-open) on the real line, the construction of SOUP is particularly simple if ϕ is continuous and $f(y \mid \phi) \to 0$ for any fixed y when ϕ approaches a boundary point of Φ. The mean parameter is $\theta = \mathbf{E}_f(Y \mid \phi) = b'(\phi)$, and it is well known [e.g., Berger (1985)] that

$$\frac{\int_\Phi b'(\phi) f(y \mid \phi)\, d\phi}{\int_\Phi f(y \mid \phi)\, d\phi} = -\frac{\int_\Phi e^{\phi y}\, de^{-b(\phi)}}{\int_\Phi e^{\phi y - b(\phi)}\, d\phi} = y. \tag{5.1}$$

Consequently, $\pi(\phi) \propto 1$ is a SOUP for ϕ. The corresponding prior for θ is then given by

$$\pi(\theta) \propto \frac{1}{b''(\phi(\theta))} = I_f(\theta) = \frac{1}{V_f(Y \mid \phi(\theta))}, \tag{5.2}$$

where $\phi = \phi(\theta)$ is the inverse mapping of $\theta = b'(\phi)$, which is one-to-one because $b''(\phi) = V_f(Y \mid \phi) > 0$, and $I_f(\theta)$ is the Fisher information for $\theta = b'(\phi)$ under f.

This elegant result was first obtained by Hartigan (1965). It is also implied by Theorem 2 of Diaconis and Ylvisaker (1979), although their theorem is applicable only in the limit to any case in Table 1. It obviously applies if θ is a linear transformation of the mean parameter of Y, and it generalizes to multidimensional exponential families [Hartigan (1965), Section 8]. In some cases, to apply this rule we must appropriately define the corresponding posterior at the boundary points; typically these involve degenerate distributions (see Section 5.2). Furthermore, the condition that $f(y \mid \phi) \to 0$ for any fixed y when ϕ approaches any boundary point of Φ is crucial for deriving (5.1); see Section 8.1 for a counterexample.

Each of two exponential family pairs in Table 1, the Binomial–Beta pair and the Negative-Binomial–F pair, describes not a single family of distributions but a class of families indexed by a convolution parameter n. Because the variance of the convolution is n times the variance of the generating variable, by (5.2) the same prior must be a SOUP for every member of the convolution class. Creation of a class of families by convolution should not be confused, however, with inference in a single family where θ itself is a convolution parameter, as in Section 4.

5.2. *The* $\mathrm{B}(n,\theta)$ *and* $\mathrm{Beta}(Y, n-Y)$ *pair.* Consider the family $\mathrm{B}(n,\theta)$, where n is known. By (5.2) the SOUP is given by $\pi(\theta) \propto \theta^{-1}(1-\theta)^{-1}$. The posterior of θ is then $\mathrm{Beta}(Y, n-Y)$, which is improper when $Y=0$ or $Y=n$. In these two cases, the infinite mass at $\theta = 0$ (when $Y=0$) or at $\theta=1$ (when $Y=n$) can be taken to define proper (degenerate) posterior distributions: $P(\theta = 0 \mid Y = 0) = P(\theta = 1 \mid Y = n) = 1$. This SOUP can also be viewed as the limiting case of Theorem 5 of Diaconis and Ylvisaker (1979) with their $b \to 0$, although their proof does not apply when $b = 0$. The dual SOUP under $\mathcal{P}_{\mathrm{Beta}}(n) = \{\mathrm{Beta}(Y, n-Y) : Y = 0, 1, \dots, n\}$ is $\pi^*(Y) \propto Y^{-1}(n-Y)^{-1}$. The calculation of $P(Y \mid \theta)$, the dual (posterior) binomial distribution, goes through if the identity

$$\frac{n}{Y(n-Y)} \frac{\Gamma(n)}{\Gamma(Y)\Gamma(n-Y)} = \binom{n}{Y}$$

is extended formally to $Y = 0, n$, so $P(Y = 0 \mid \theta = 0) = P(Y = n \mid \theta = 1) = 1$.

For any fixed $n \geq 1$, $\mathcal{P}_{\mathrm{B}}(n) = \{\mathrm{B}(n,\theta) : \theta \in [0,1]\}$ is complete, and thus the unbiased estimator $\hat{\theta}(Y) = Y/n$ is unique. Furthermore, $\mathcal{P}_{\mathrm{B}}(n)$ is O-complete because of its completeness and Lemma 1 (or by using Theorem 2 with the modification at $\theta = 0$ and $\theta = 1$). Thus, by Theorem 1, $\pi^*(Y) \propto Y^{-1}(n-Y)^{-1}$ is the only SOUP corresponding to $\hat{Y}(\theta) = n\theta$ (a.e. $\mathcal{P}_{\mathrm{Beta}}(n)$).

However, for fixed $n \geq 1$, $\mathcal{P}_{\mathrm{Beta}}(n)$ is neither complete nor O-complete because its parameter space $Y \in \{0, 1, \dots, n\}$ is not rich enough, the same problem encountered in Diaconis and Ylvisaker [(1979), Section 4]. Thus, $\hat{Y}(\theta) = n\theta$ is not the only unbiased estimator of Y under $\mathscr{P}_{\text{Beta}}(n)$, nor is $\pi(\theta) \propto \theta^{-1}(1-\theta)^{-1}$ the only SOUP for θ under $\mathscr{P}_{\text{B}}(n)$. However, if we require the SOUP for θ to be free of n, then $\pi(\theta) \propto \theta^{-1}(1-\theta)^{-1}$ is the only SOUP almost surely with respect to $\mathscr{P}^*_{\text{Beta}} = \{\text{Beta}(Y, n-Y) : Y = 0, 1, \ldots, n;\ n = 1, 2, \ldots\} = \{\text{Beta}(\alpha, \beta) : \alpha \in \mathcal{N}^+,\ \beta \in \mathcal{N}^+,\ \alpha + \beta \geq 1\}$, because $\mathscr{P}^*_{\text{Beta}}$ is O-complete. Note here we define $\text{Beta}(0, \beta) = \mathbb{1}_{\{\theta=0\}}$ and $\text{Beta}(\alpha, 0) = \mathbb{1}_{\{\theta=1\}}$, where $\alpha, \beta \geq 1$.

To verify the O-completeness of $\mathscr{P}^*_{\text{Beta}}$, suppose $\mathbf{Cov}_{\alpha,\beta}(g(Z), z) = 0$ for a $g(Z) \geq 0$ satisfying $\mathbf{E}_{\alpha,\beta}[g(Z)] > 0$, for all α and β allowed by $\mathscr{P}^*_{\text{Beta}}$. Taking $\alpha = k \geq 1$ and $\beta = 1$ then yields

$$(5.3) \quad 0 < \int_0^1 Z^k g(Z)\, dZ = \frac{k}{k+1} \int_0^1 Z^{k-1} g(Z)\, dZ < \infty, \qquad k = 1, 2, \ldots.$$

Consequently, $0 < \int_0^1 g(Z)\, dZ < \infty$ and thus without loss of generality we assume $\int_0^1 g(Z)\, dZ = 1$. The recursion (5.3) then implies $\mathbf{E}_g(Z^k) = (k+1)^{-1}$, $\forall k \geq 1$. Therefore, $g(Z) \equiv 1$ because $\{\mu_k = (k+1)^{-1}, k = 1, 2, \ldots\}$ is also the set of moments for $U(0, 1)$ and the set uniquely (a.e. Lebesgue) determines the density since $\sum_{k=0}^{\infty} \frac{\mu_k t^k}{k!} = \frac{e^t - 1}{t} < \infty$ for some (actually all) $t > 0$ [e.g., Stuart and Ord (1987), Section 4.21]. We remark that the O-completeness of $\mathscr{P}^*_{\text{Beta}}$ can also can be deduced from Hartigan's (1965) asymptotic uniqueness result, because his result asserts that as $n \to \infty$, SOUP is unique, and thus under the additional requirement that SOUP is free of n, $\pi(\theta) \propto \theta^{-1}(1-\theta)^{-1}$ is the only possibility (a.e. $\mathscr{P}^*_{\text{Beta}}$).

The Binomial–Beta pair generalizes readily to a Multinomial–Dirichlet pair: $\mathbf{Y} \mid \theta \sim \text{Multinomial}(n, \theta)$ has SOUP $\pi(\theta) \propto \prod \theta_i^{-1}$, and the dual distribution $\theta \mid \mathbf{Y} \sim \text{Dirichlet}(\mathbf{Y})$ has SOUP $\pi^*(\mathbf{Y}) \propto \prod Y_i^{-1}$.

5.3. *The* Poisson(θ) *and* Gamma(Y, 1) *pair.* Having considered the binomial family, it is natural to consider its limit as $n \to \infty$, the Poisson distribution, $Y \mid \theta \sim \text{Poisson}(\theta)$. By (5.2), the SOUP is given by $\pi(\theta) \propto \theta^{-1}$, a continuous counterpart of the discrete SOUP for $\text{B}(\theta, p)$ of Section 4.1. The corresponding posterior distribution $P(\theta \mid Y)$ is Gamma(Y, 1). The posterior of θ is degenerate when $Y = 0$, with an "infinite" mass at $\theta = 0$; we define $P(\theta = 0 \mid Y = 0) = 1$, on the same argument as in the case of $\text{B}(\theta, p)$.

The dual SOUP under the Gamma family $\mathscr{P}_{\text{Gamma}} = \{\text{Gamma}(Y, 1) : Y \in \mathcal{N}^+\}$ is $\pi^*(Y) \propto Y^{-1}$. As noted in Section 4, the Poisson–Gamma pair may also be regarded as a convolution pair. Consequently, the intriguing variance relationships observed in Section 4.2 hold for this pair: both sides of (4.3) are θ and both sides of (4.4) are Y. These are also trivially true because for both Poisson(θ) and Gamma(Y, 1), the mean parameter is the same as the variance parameter. The Gamma family with unknown scale parameter, Gamma(c, θ), where c is known, is covered by the results in Section 6.1.

Since $\mathscr{P}_{\text{Poisson}} = \{\text{Poisson}(\theta),\ \theta \in [0, \infty)\}$ is complete and O-complete, there is only one unbiased estimator $\hat{\theta}(Y) = Y$ for θ under $\mathscr{P}_{\text{Poisson}}$ and one SOUP corresponding to $\hat{Y}(\theta) = \theta$ under $\mathcal{P}_{\text{Gamma}}$. The O-completeness of $\mathcal{P}_{\text{Poisson}}$ follows its completeness and either Theorem 2 (with modification at $\theta = 0$) or more directly the identity $\mathbf{E}_\lambda[g(Z)(Z-\lambda)] = \lambda \mathbf{E}_\lambda[g(Z+1) - g(Z)]$, a relative of (5.1) of Hudson (1978).

However, $\mathcal{P}_{\text{Gamma}}$ is not complete due to the discrete nature of Y, so there are unbiased estimators of Y other then $\hat{Y}(\theta) = \theta$. Nevertheless, $\mathcal{P}_{\text{Gamma}}$ is O-complete, and thus SOUP $\pi(\theta) \propto \theta^{-1}$ is unique for θ (a.e. $\mathcal{P}_{\text{Gamma}}$). The O-completeness is verified by taking $Z \sim \text{Gamma}(k, 1)$ in $\mathbf{Cov}_k(g(Z), Z) = 0$ for $k = 1, 2, \dots$, which leads to $\mathbf{E}_h(Z^k) = k!$, where $h(Z) \propto g(Z)e^{-Z}\mathbb{1}_{(0,\infty)}(Z)$. Consequently $g(Z)$ must be constant because the exponential distribution $p(Z) = e^{-Z}\mathbb{1}_{(0,\infty)}(Z)$ is the only (a.e. Lebesgue) possible distribution having moments $\mu_k = k!$, $k = 1, 2, \dots$. This uniqueness can also be viewed as a natural extension, to the improper prior cases, of Johnson's (1957) characterization of Gamma distributions as the only priors to yield linear posterior expectations for the Poisson distribution.

5.4. *The* $\text{NB}(n, (\theta+1)^{-1})$ *and F pair.* Consider $\text{NB}(n, p)$ with $p = (\theta+1)^{-1}$ so $\mathbf{E}(Y \mid \theta) = n(1+\theta)$ is linear in θ. Since $\mathbf{V}(Y \mid \theta) = n\theta(1+\theta)$, by (5.2), a SOUP corresponding to $\hat{\theta}(Y) = (Y-n)/n$ is $\pi(\theta) \propto \theta^{-1}(\theta+1)^{-1}$, with infinite mass at $\theta = 0$. The dual family is the (scaled) F family,

$$\mathcal{P}_F(n) = \left\{ \frac{Y-n}{n+1} F_{2(Y-n),2(n+1)} \equiv \frac{\chi^2_{2(Y-n)}}{\chi^2_{2(n+1)}} : Y = n, n+1, \dots \right\},$$

where $\chi_0^2 \equiv 0$. The dual SOUP is $\pi^*(Y) \propto Y^{-1}(Y-n)^{-1}$, which is improper.

Since $\mathcal{P}_{\text{NB}}(n) = \{\text{NB}(n, (\theta+1)^{-1}) : \theta \in [0, \infty)\}$ is complete and O-complete, the unbiased estimator for θ is unique under $\mathcal{P}_{\text{NB}}(n)$ and the dual SOUP $\pi^*(Y) \propto Y^{-1}(Y-n)^{-1}$ is unique under $\mathcal{P}_F(n)$. The O-completeness of $\mathcal{P}_{\text{NB}}(n)$ follows from its completeness and Lemma 2 (or Theorem 2 with modification at $\theta = 0$). However, $\mathcal{P}_F(n)$ is not complete because of the discrete nature of Y, and thus there are unbiased estimators of Y other than $\hat{Y}(\theta) = n(1+\theta)$. But $\mathcal{P}_F(n)$ is O-complete, because $\mathbf{Cov}_Y(g(\theta), \theta) = 0$ implies, after letting $\zeta = \theta/(1+\theta)$ and $m = Y - n$, $\mathbf{E}_h[\zeta^m] = \frac{m}{n}\mathbf{E}_h[\zeta^{m-1} - \zeta^m]$, $m = 1, 2, \dots$, where the expectation is taken over $h(\zeta) \propto (1-\zeta)^{n-1} g(\zeta/(1-\zeta))$. It follows that

$$\mu_m \equiv \mathbf{E}_h[\zeta^m] = \frac{m}{m+n}\mathbf{E}_h[\zeta^{m-1}] = \cdots = \frac{m!n!}{(m+n)!}, \qquad m \geq 1.$$

Since $\text{Beta}(1, n) = n(1-\zeta)^{n-1}$ has the same set of moments, and this set uniquely determines the density, we conclude $g(\theta) = \text{constant}$. Consequently, $\pi(\theta) \propto \theta^{-1}(\theta+1)^{-1}$ is the only SOUP for θ under $\mathcal{P}_{\text{NB}}(n)$.

It is informative to note that if we consider $Y \mid p \sim \text{NB}(n, p)$, then the SOUP corresponding to the unbiased estimator $\hat{p}(Y) = (n-1)/(Y-1)$ is $\pi(p) \propto p^{-2}(1-p)^{-1}$ with the dual family $\{\text{Beta}(n-1, Y-n),\ Y \geq n\}$. This example demonstrates that when $\mathbf{E}(Y \mid p)$ is not linear in p, (5.2) may not yield a SOUP: in this case $\mathbf{E}(Y \mid p) = n/p$ and $\mathbf{V}(Y \mid p) = n(1-p)/p^2$. We also note that $\theta \mid Y \sim \chi^2_{2(Y-n)}/\chi^2_{2(n+1)}$ is the same as $p = (1+\theta)^{-1} \mid Y \sim \mathrm{Beta}(n+1, Y-n)$, which is different from the dual/posterior distribution $\mathrm{Beta}(n-1, Y-n)$ corresponding to the SOUP for p. This is not surprising since unbiasedness and thus the SOUP property are not invariant to reparametrization; more on this in Section 7.

6. Location and scale families.

6.1. *Scale families.* Suppose that f is a continuous scale family generated by $f_0(\cdot)$, $f(y \mid \theta) = f_0(y/\theta)/\theta$, where $\theta, y \in (0, \infty)$. Then $\hat{\theta}(y) = y/\mu_0$ is an unbiased estimator of θ, where $\mu_0 = \int y f_0(y)\,dy$ is assumed to exist. Applying (2.2), finding a SOUP in this case amounts to solving

$$\frac{\int_0^\infty f_0(y/\zeta)\pi(\zeta)\,d\zeta}{\int_0^\infty f_0(y/\zeta)\pi(\zeta)/\zeta\,d\zeta} = \frac{y}{\int_0^\infty \zeta f_0(\zeta)\,d\zeta} \qquad \forall\, y \in (0, \infty) \tag{6.1}$$

for π. With a change of variable $\xi = y/\zeta$ (y fixed), we can reexpress (6.1) in the form of (2.3),

$$\mathbf{E}[\xi^{-1}\pi(y\xi^{-1})] = \mathbf{E}[\xi]\mathbf{E}[\xi^{-2}\pi(y\xi^{-1})] \qquad \forall\, y \in (0, \infty), \tag{6.2}$$

where $\xi \sim f_0$. Thus a solution is obtained by letting $\beta_\pi(\xi, y) \equiv \xi^{-2}\pi(y\xi^{-1})$ be free of ξ. This yields a SOUP, $\pi(\theta) \propto \theta^{-2}$. Affine duality here implies no new results, because $f^*(\theta \mid y) = f_0^*(\theta/y)/y$, where $f_0^*(x) = f_0(x^{-1})x^{-2}$, is just another scale family.

It is noteworthy that π is not the usual scale-invariant prior $\pi(\theta) \propto \theta^{-1}$ [e.g., Berger (1985), Section 3.3.2], but it is a *relatively scale invariant prior* (see below). Also, the condition that θ varies over $(0, \infty)$ is important; if we restrict, say, $\theta \geq \theta_0 > 0$, then there is no SOUP, a case of (II) in Section 2.1.

Discrete scale families may be defined for which the sample and parameter spaces are "exponential lattices," $\Omega = \Theta = \{c^j : j = 0, \pm 1, \pm 2, \ldots\}$. The SOUP for this case is $\pi(\theta) \propto \theta^{-1}$; the difference from the continuous case arises because the Jacobian is 1 for the change of variables from (6.1) to (6.2).

The uniqueness of this SOUP can be determined using O-completeness for specific forms of $f_0(\cdot)$ [e.g., for the family of exponential distributions, $f_0(z) = e^{-z}$, $z \geq 0$]. To assert uniqueness without knowing the form of $f_0(t)$, we must restrict the family of priors. As discussed in Berger [(1985), Section 3.3], when a noninformative prior is sought for a scale parameter, the first choice would be a *relatively scale invariant prior* [Hartigan (1964)], essentially one which looks the same after rescaling of the variable, satisfying

$$\pi\left(\frac{\theta_1}{\theta_2}\right) = \pi(1)\frac{\pi(\theta_1)}{\pi(\theta_2)} \qquad \text{for any positive } \theta_1 \text{ and } \theta_2. \tag{6.3}$$

Assuming $\pi(\theta)$ is continuous at one point at least, (6.3) implies that π must be a power function, that is, $\pi(\theta) \propto \theta^c$ for some $c \in R$. Within such a family, (6.2) becomes

$$\mathbf{E}[\xi^{-(c+1)}] = \mathbf{E}[\xi]\mathbf{E}[\xi^{-(c+2)}] \qquad \text{where } \xi \sim f_0. \tag{6.4}$$

Since $g_1(\xi) = \xi$ and $g_2(\xi) = \xi^{-(c+2)}$ are either concordant (i.e., both are increasing functions when $c < -2$) or discordant [i.e., $g_2(\xi)$ is decreasing when $c > -2$], by the second Chebyshev inequality, (6.4) can be true only when $g_2(\xi)$ is a constant function (a.e. f_0); that is, $c = -2$. Hence, $\pi(\theta) \propto \theta^{-2}$ is the unique SOUP among all continuous relatively scale invariant priors.

6.2. *Location families.* Suppose that f is a location family, $f(y \mid \theta) = f_0(y - \theta)$, where both y and θ vary on the whole real line. An unbiased estimator for θ is $\hat{\theta}(y) = y - \mu_0$, where $\mu_0 = \int_{-\infty}^{\infty} y f_0(y)\, dy$ is assumed to exist. In analogy to (6.2), finding a corresponding SOUP amounts to solving

$$\mathbf{E}[\xi\pi(y - \xi)] = \mathbf{E}[\xi]\mathbf{E}[\pi(y - \xi)] \qquad \forall\, y \in (-\infty, \infty), \tag{6.5}$$

where $\xi \sim f_0$. The standard location invariant prior, $\pi(\theta) \propto$ constant, is an obvious solution to (6.5), by taking β_π in (2.3) to be a constant. This result also applies for vector-valued and lattice-valued location families. Again, θ must be unbounded, and affine duality does not generate a new family.

Analogously to Section 6.1, we consider *relatively location invariant priors*, which satisfy

$$\pi(\theta_1 - \theta_2) = \pi(0)\frac{\pi(\theta_1)}{\pi(\theta_2)} \qquad \text{for any } \theta_1 \text{ and } \theta_2. \tag{6.6}$$

Under the same assumption of continuity at one point at least, (6.6) implies $\pi(\theta) \propto e^{c\theta}$ for some $c \in R$. Within this family, by applying the same second Chebyshev inequality as in Section 6.1, we can conclude $c = 0$ is the only solution. Again, if $f_0(\cdot)$ is of some specific form [e.g., $N(\theta, 1)$], then it may be possible to establish uniqueness without the restriction to relatively location invariant priors.

7. Nonlinear transformations of mean parameters.

7.1. *Transformed mean parameters and their unbiased estimators.* Since SOUP is not invariant to a nonlinear reparametrization, finding a SOUP for a nonlinear transformation is more complicated than the usual multiplication by the corresponding Jacobian factor. As theoretical illustrations, here we consider various nonlinear transformations $\theta = \vartheta(\gamma)$ of the original mean parameter (or its conventional linear transformation) $\gamma = n, p, \lambda$, or $(1 - p)/p$ as introduced in Sections 4 and 5. These transformed parameters are themselves mean parameters of nonlinear transformations $Y = y(Z)$ of the original observation Z, and by duality $Y = y(Z)$ is the mean parameter of the corresponding dual family. Note that determining the SOUP for a transformed parameter can also be regarded as determining an unbiased prior for the original parameter with the loss function defined by square error of the transformed parameter.

A number of examples are summarized in Table 2. The numbering of the distributions corresponds to that in Table 1, but for distributions for which more than one transformation is given, the alternatives are indicated by Roman numerals. The definitions of θ for the Negative Binomial cases are given, for clarity, in terms of p as well as γ. We omit the binomial case with $\theta = (1-p)^K$ because it is an obvious extension of the case with $\theta = p^K$. We also omit the lognormal case, $Z \mid \theta \sim \mathrm{N}(\gamma, 1)$ with $\theta = e^\gamma$, $Y = e^{Z-1/2}$ and $\pi(\theta) \propto 1/\theta^2$, which is a scale family as in Section 6.1. We note that for cases 1, 2, 3(i), 4(i) and 4(iii), the posterior distribution $\theta \mid Z$ degenerates to a point mass at 0 for all $Z < K$ so there is no loss of information by the many-to-one transformation $Z \to y(Z) = 0$ for $Z < K$.

While there is a large literature on finding unbiased estimators for functions of parameters [see Voinov and Nikulin (1993)], we find the following result particularly convenient in constructing Table 2. The unbiased estimators we consider involve shifts in the convolution parameter and/or the observation.

LEMMA 3. *Suppose that $f_n(z \mid \phi) = a_n(z)\exp(z\phi - nb(\phi))$ is a density with respect to μ, either Lebesgue or counting measure, where $n \in \mathcal{N}(\phi) = \{n : f_n(z \mid \phi)$ is a proper density$\}$. Let K_1 and K_2 be two constants such that $K_2 + n \in \mathcal{N}(\phi)$ and $\mathcal{Z}_{n+K_2} + K_1 \subseteq \mathcal{Z}_n$, where $\mathcal{Z}_m$ denotes the support of $a_m(z)$. Then*

$$\mathbf{E}_{f_n}\left[\frac{a_{n+K_2}(Z-K_1)}{a_n(Z)}\,\middle|\,\phi\right] = e^{K_1\phi + K_2 b(\phi)}. \tag{7.1}$$

PROOF.

$$\begin{aligned}
&\mathbf{E}_{f_n}\left[\frac{a_{n+K_2}(Z-K_1)}{a_n(Z)}\,\middle|\,\phi\right]\\
&\quad= \int_{\mathcal{Z}_n} \frac{a_{n+K_2}(z-K_1)}{a_n(z)} a_n(z) e^{z\phi - nb(\phi)}\mu(dz)\\
&\quad= e^{K_1\phi + K_2 b(\phi)} \int_{\mathcal{Z}_n} a_{n+K_2}(z-K_1) e^{(z-K_1)\phi - (n+K_2)b(\phi)}\mu(dz)\\
&\quad= e^{K_1\phi + K_2 b(\phi)} \int_{\mathcal{Z}_n - K_1} f_{n+K_2}(z;\phi)\mu(dz) = e^{K_1\phi + K_2 b(\phi)}.
\end{aligned}$$
□

Consequently, when $\mathcal{N}(\phi)$ is free of ϕ, which is true for all the exponential cases in Table 2, $y(Z) = a_{n+K_2}(z-K_1)/a_n(z)$ is an unbiased estimator of $\vartheta(\phi) = \exp(K_1\phi + K_2 b(\phi))$ for K_1 and K_2 satisfying the conditions of Lemma 3. The values of K_1, K_2 corresponding to each case are shown in Table 2. Note that

TABLE 2

SOUP for some transformed parameters and observations[1]

	Model $Z \mid \gamma$ and Posterior (dual) model $\gamma \mid Z$	Parameter $\theta = \theta(\gamma)$ and Observation $Y = y(Z)$	SOUP for θ, $\pi_\theta^{\mathcal{S}}(\theta)$ and SOUP for Y, $\pi^*(y)$	Corresponding prior for γ, $\pi_\theta^{\mathcal{S}}(\gamma)$ (continuous γ)	Values of K or a, K_1, K_2
1a	$Z \sim \mathrm{B}(\gamma, p)$	$F_K(\gamma)$	$1/\theta = 1/F_K(\gamma)$		$K \geq 1$
1b	$\begin{cases}(K-1) + \mathrm{NB}(Z-K+1, p), & Z \geq K \\ P(\theta = 0 \mid Z = z) = 1, & z < K\end{cases}$	$F_K(Z)/p^K$	$1/y$		$K_2 = K,\ K_1 = 0$
2a	$Z \sim \mathrm{B}(n, \gamma)$	γ^K	$\theta^{1/K-2}(1-\theta^{1/K})^{-1}$	$\gamma^{-K}(1-\gamma)^{-1}$	$0 < K \leq n$
2b	$\begin{cases}\gamma \sim \mathrm{Beta}(Z-K+1, n-Z), & Z \geq K \\ P(\theta = 0 \mid Z = z) = 1, & z < K\end{cases}$	$F_K(Z)/F_K(n)$	$1/(n-z)y$		$K_1 = -K_2 = K$
3(i)a	$Z \sim \mathrm{Poisson}(\gamma)$	γ^K	$\theta^{1/K-2}$	γ^{-K}	$K > 0$
3(i)b	$\begin{cases}\gamma \sim \mathrm{Gamma}(Z-K+1, 1), & Z \geq K \\ P(\theta = 0 \mid Z = z) = 1, & z < K\end{cases}$	$F_K(Z)$	$1/y$		$K_1 = K,\ K_2 = 0$
3(ii)a	$Z \sim \mathrm{Poisson}(\gamma)$	$e^{a\gamma}$	$\theta^{-2}/(\log\theta)$	$e^{-a\gamma}/\gamma$	$a > -1$
3(ii)b	$\gamma \sim \mathrm{Gamma}(Z, (a+1)^{-1})$	$(a+1)^Z$	$1/zy$		$K_2 = a,\ K_1 = 0$
4(i)a	$Z \sim \mathrm{NB}(n, (1+\gamma)^{-1}) - n$	$\gamma^K = [(1-p)/p]^K$	$\theta^{1/K-2}(1+\theta^{1/K})^{-1}$	$\gamma^{-K}(1+\gamma)^{-1}$	$K > 0$
4(i)b	$\begin{cases}(1+\gamma)^{-1} = p \sim \mathrm{Beta}(n+K, Z-K), & Z \geq K \\ P(\theta = 0 \mid Z = z) = 1, & z < K\end{cases}$	$F_K(Z)/F_K(n+K-1)$	$1/(z+n)y$		$K_1 = K_2 = K$
4(ii)a	$Z \sim \mathrm{NB}(n, (1+\gamma)^{-1}) - n$	$(1+\gamma)^{-K} = p^K$	$\theta^{-2}(1-\theta^{1/K})^{-1}$	$(1+\gamma)^K\gamma^{-1}$	$-\infty < K < n$
4(ii)b	$(1+\gamma)^{-1} = p \sim \mathrm{Beta}(n-K, Z)$	$F_K(n-1)/F_K(n+Z-1)$	$1/zy$		$K_2 = -K,\ K_1 = 0$
4(iii)a	$Z \sim \mathrm{NB}(n, (1+\gamma)^{-1}) - n$	$[\gamma/(1+\gamma)]^K = (1-p)^K$	$\theta^{-(2+1/K)}(\theta^{-1/K}-1)^{-2}$	$[\gamma/(1+\gamma)]^{-K}$	$K > 0$
4(iii)b	$\begin{cases}(1+\gamma)^{-1} = p \sim \mathrm{Beta}(n-1, Z-K+1), & Z \geq K \\ P(\theta = 0 \mid Z = z) = 1, & z < K\end{cases}$	$F_K(Z)/F_K(n+Z-1)$	$1/y$		$K_1 = K,\ K_2 = 0$

[1] K is an integer; a is a real; $F_K(t) \equiv \begin{cases} t!/(t-K)!, & t \geq K, \\ 0, & t < K. \end{cases}$ In 4(i–iii), Z is distributed on $\{0, 1, 2, \ldots\}$, that is, the distribution of the number of *failures* before n successes.

to apply this result to the Poisson case we need first to introduce a convolution family of the sum of n Poisson(λ) variables, for which the density is $\exp(z \log \lambda - n\lambda) n^z/z!$; hence $a_n(z) = n^z/z!$.

7.2. *The use of Hartigan's asymptotic construction.* All the SOUPs in Table 2 for transformed exponential families can be obtained by applying Hartigan's (1965) asymptotic construction. Under a host of regularity conditions, Hartigan (1965) has shown that with i.i.d. samples $\mathbf{X} = \{X_1, \ldots, X_n\}$ from $f(x \mid \theta)$ and using loss function $L(\delta, \theta)$, the Bayes decision $\delta_\pi(\mathbf{X})$ under prior π is asymptotically unbiased, namely,

$$\mathbf{E}\big[L\big(\delta_\pi(\mathbf{X}), \tilde{\theta}\big) \,\big|\, \theta\big] \geq \mathbf{E}\big[L\big(\delta_\pi(\mathbf{X}), \theta\big) \,\big|\, \theta\big] \qquad \forall \tilde{\theta}, \theta \in \Theta,$$

if and only if

$$\pi(\theta) \propto I_f(\theta) \left[\frac{\partial^2 L(\delta, \theta)}{\partial \delta^2}\right]_{\delta=\theta}^{-1/2}, \tag{7.2}$$

where $I_f(\theta)$ is the Fisher information of $f(x \mid \theta)$. For squared error loss $L(\delta, \theta) = (\delta - \theta)^2$, Hartigan's result (7.2) implies that the asymptotically unbiased prior is

$$\pi(\theta) \propto I_f(\theta). \tag{7.3}$$

For the exponential families discussed in Section 5, this gives the exact SOUP (5.2), as shown there.

In general, when $I_f(\theta)$ exists, (7.3) is a good candidate for SOUP even when the exact argument of Section 5 is not applicable. In particular, since for a one-to-one differentiable transformation $\theta = \vartheta(\gamma)$, $I_f(\gamma) = I_f(\vartheta(\gamma))(\vartheta'(\gamma))^2$, (7.3) suggests that when we already have a SOUP for γ, denoted by $\pi_\gamma^{\mathcal{S}}(\gamma)$, and want to seek a SOUP for $\theta = \vartheta(\gamma)$, a likely candidate is

$$\pi_\theta(\theta) \propto \pi_\gamma^{\mathcal{S}}\big(\vartheta^{-1}(\theta)\big) J^2(\theta), \tag{7.4}$$

where $J(\theta) = \frac{d\gamma}{d\theta}$ is the Jacobian from $\theta \to \gamma = \vartheta^{-1}(\theta)$. In other words, the usual multiplicative factor $|J|$ is replaced by J^2. Equivalently, in terms of γ, a good candidate of SOUP *for the new parameter* $\theta = \vartheta(\gamma)$ is

$$\pi_\theta(\gamma) \propto \pi_\gamma^{\mathcal{S}}(\gamma) |J(\vartheta(\gamma))|. \tag{7.5}$$

To avoid confusion, in (7.4) and (7.5) we used the superscript $\mathcal{S}$ to denote a SOUP, the subscript of π to index the SOUP parameter, that is, the parameter whose posterior mean is an unbiased estimator of itself, and the function argument to index the density parameter, that is, the parameter for which we choose to compute the density. For the exponential family given in Lemma 3, it is sometimes convenient to use the canonical parameter ϕ. Because $\gamma = nb'(\phi)$, $\pi_\gamma^{\mathcal{S}}(\gamma) \propto 1/b''(\phi(\gamma))$, the prior for ϕ corresponding to (7.5) is

$$\pi_\theta(\phi) \propto \pi_\theta\big(nb'(\phi)\big)|b''(\phi)| \propto \big|J\big(\vartheta\big(nb'(\phi)\big)\big)\big| \propto b''(\phi) \left|\frac{d\theta}{d\phi}\right|^{-1}. \tag{7.6}$$

For $\theta = \vartheta(\phi) = \exp(K_1\phi + K_2 b(\phi))$, (7.6) becomes

$$\pi_\theta(\phi) \propto \frac{b''(\phi)}{K_1 + K_2 b'(\phi)} e^{-K_1\phi - K_2 b(\phi)}. \tag{7.7}$$

For all the transformed exponential families in Table 2, we verified that under this $\pi_\theta(\phi)$,

$$\mathbf{E}_\pi\left[e^{K_1\phi + K_2 b(\phi)} \mid Z\right] = \frac{a_{n+K_2}(Z - K_1)}{a_n(Z)}, \tag{7.8}$$

and thus $\pi_\theta(\phi) = \pi_\theta^{\mathcal{S}}(\phi)$. This is somewhat remarkable, for θ is not the mean parameter for the natural observation and thus Hartigan's exact argument, given in Section 5.1, does not apply to these cases.

On the other hand, this asymptotic approach does not always produce a SOUP. In the examples we have checked, (7.7) yields a SOUP whenever

$$\frac{b''(\phi)}{K_1 + K_2 b'(\phi)} = \exp\bigl(C_1\phi + C_2 b(\phi)\bigr), \tag{7.9}$$

where C_1 and C_2 are constants, and not otherwise. For example, we have verified by numerical integration that (7.7) is not a SOUP for $Z \mid \lambda \sim \text{Poisson}(\lambda)$ with $\theta = \lambda^{K_1} e^{K_2\lambda}$ where $K_1 > 0$, $K_2 > -1$, $K_2 \neq 0$, which violates (7.9) with $b(\phi) = e^\phi = \lambda$. The unbiased estimator from Lemma 3 for $\theta = \lambda^{K_1} e^{K_2\lambda}$ is $Y = y(Z) = (a+1)^{Z-K}[z!/(z-K)!]\mathbb{1}_{\{z \geq K\}}$, but we are unable to find a SOUP corresponding to $y(Z)$. We suspect that there may be a limited set of families and values of K_1, K_2 for which (7.7) delivers $\pi_\theta^{\mathcal{S}}(\phi)$, that is, those corresponding to solutions of the differential equation (7.9).

8. Counterexamples and open problems. We conclude by presenting several less obvious examples where there is no SOUP, and discuss related open problems. We also report our unsuccessful attempts to find a unified principle underlying the inverse variance rule given in Section 5.1, hoping to stimulate an ultimately successful resolution, which would be valuable since the inverse variance rule is almost trivial to apply in practice.

8.1. *The inverse Gaussian family, inadmissibility and Berger's lemma.* The density of the inverse Gaussian distribution $\text{IG}(\mu, \lambda)$, where $\mu > 0$ and $\lambda > 0$, is

$$f(y \mid \mu, \lambda) = \left(\frac{\lambda}{2\pi y^3}\right)^{1/2} \exp\left\{-\frac{\lambda(y-\mu)^2}{2\mu^2 y}\right\}, \qquad y > 0. \tag{8.1}$$

This is a continuous exponential family with natural parameter $\phi = -\lambda/(2\mu^2)$, mean parameter μ and variance parameter μ^3/λ. Here we are interested in the mean parameter $\theta = \mu$, for which Y is the unique (a.e. Lebesgue) unbiased estimator. The λ parameter is assumed known, but it is not set to unity to allow reduction via sufficiency; namely, if $\{Y_1, \dots, Y_n\}$ are i.i.d. samples from (8.1), then $\bar{Y}n \mid \mu, \lambda \sim \mathrm{IG}(\mu, n\lambda)$.

An interesting feature of this family is that while $\lim_{\theta\to 0} f(y \mid \theta) = 0$ for any $y > 0$, when $\theta \to \infty$, $f(y \mid \theta)$ approaches the density of $\lambda\chi_1^{-2}$, which is positive for any positive y and whose mean is infinite. Consequently, the inverse variance rule (5.2), which gives $\pi(\theta) \propto \theta^{-3}$, does not lead to a SOUP for θ. Instead, it is easy to verify that under this prior, the posterior mean is always less than $\bar{Y}_n$:

$$\mathbf{E}_\pi(\theta \mid \bar{Y}_n, \lambda) = \frac{\Phi(n\lambda/\bar{Y}_n)}{\Phi(n\lambda/\bar{Y}_n) + (\bar{Y}_n/n\lambda\sqrt{2\pi})e^{-n\lambda/(2\bar{Y}_n)}}\bar{Y}_n, \tag{8.2}$$

where $\Phi(x)$ is the c.d.f. of $N(0, 1)$.

In fact, there is no SOUP, no matter how pathological, for θ for this family. A proof is given in Appendix B. This suggests a possible extension of Berger's (1990) lemma on the behavior of unbiased estimators when the underlying distribution family $\{f(y \mid \theta),\ \theta \in \Theta\}$ is not degenerate at the boundary of Θ. Berger's lemma states that in such cases any unbiased estimator of θ must assume values outside Θ. However, Berger's lemma does not apply to cases where the nondegenerate boundary value of θ is $\pm\infty$, as in the current case where the unbiased estimator $\bar{Y}_n$, which is also the maximum likelihood estimator and uniformly minimum variance unbiased estimator, is obviously always within the parameter space.

We conjecture a natural extension of Berger's lemma, that when $\{f(y \mid \theta),\ \theta \in \Theta\}$ is not degenerate at the boundary of Θ, whether that boundary is finite or not, there is no SOUP for θ. Of course, this weaker version still implies the inadmissibility of any unbiased estimator [under the regularity conditions of Brown (1980)], which was Berger's main emphasis. For $\mathrm{IG}(\mu, \lambda)$, $\bar{Y}_n$ is inadmissible because it is uniformly dominated by $Y_n^{(b)} = \bar{Y}_n \mathbb{1}_{\{\bar{Y}_n \le b\}} + w_n(b)\bar{Y}_n \mathbb{1}_{\{Y_n > b\}}$ for any $b > 0$ with an appropriate choice of $w_n(b) \in (0.5, 1)$. Other constructions are given in Hsieh, Korwar and Rukhin (1990).

8.2. *Nonexistence of SOUP for odd functions of location parameters.* We now show that there are infinitely many examples where SOUP does not exist even if $f(y \mid \theta)$ degenerates on any boundary point of the function of θ being estimated and none of (I)–(III) of Section 2.1 occurs. This further substantiates Berger's (1990) conclusion: "...inadmissibility of unbiased estimators is likely to be the rule, rather than exception." While more general results can be obtained with more complicated mathematical details, we present the following simple yet informative result concerning odd functions of a location parameter.

THEOREM 3. *Let $\{f(z - \gamma),\ z \in R,\ \gamma \in R\}$ be a unimodal symmetric location family on the real line, where $f(z)$ is strictly decreasing for $z > 0$. Let $\theta = \theta(\gamma)$ be an odd function of γ with $\theta(\gamma) > 0$ when $\gamma > 0$, and $Y(Z)$ be* *an unbiased estimator of* $\theta(\gamma)$. *If* $Y(z)$ *is an odd function of* z *and has* (*at least*) *a positive root* $z_0 > 0$, *then there is no SOUP corresponding to* $Y(Z)$.

PROOF. We assume that π is a SOUP and demonstrate a contradiction. By the definition of SOUP, $\int_{-\infty}^{\infty}\pi(\gamma)\theta(\gamma)f(z-\gamma)\,d\gamma = 0$ for $z = -z_0, z_0$, because $Y(-z_0) = -Y(z_0) = 0$. It follows that

$$\int_0^{\infty}\pi(\gamma)\theta(\gamma)f(\gamma \pm z_0)\,d\gamma = \int_0^{\infty}\pi(-\gamma)\theta(\gamma)f(\gamma \mp z_0)\,d\gamma. \tag{8.3}$$

Combining the two versions of (8.3) yields $\int_0^{\infty}\theta(\gamma)[\pi(\gamma)+\pi(-\gamma)][f(\gamma - z_0) - f(\gamma + z_0)]\,d\gamma = 0$. This implies $\pi(\gamma) = 0$ (a.e. Lebesgue) because $\theta(\gamma) > 0$ and $f(\gamma - z_0) - f(\gamma + z_0) > 0$ for all $\gamma > 0$ under our assumptions. But $\pi(\gamma) = 0$ (a.e. Lebesgue) is not a SOUP. □

If there is any unbiased estimator there must be an odd one, $Y^*(Z) = [Y(Z) - Y(-Z)]/2$, since $\mathbf{E}[Y^*(Z) \mid \gamma] = [\theta(\gamma) - \theta(-\gamma)]/2 = \theta(\gamma)$. An odd estimator is natural given the symmetry of this problem. In particular a unique unbiased estimator must be odd. The assumption that $Y(z)$ has at least one positive root is also satisfied in many examples, as demonstrated by the following lemma, the proof of which is given in Appendix C.

LEMMA 4. *Let* $\{f(z-\gamma),\ z \in R,\ \gamma \in R\}$ *be any location family, and assume* $\mathbf{E}(|Z|^K \mid \gamma = 0)$ *is finite, where* K *is a positive integer. Then there is a unique unbiased polynomial estimator of degree* K *for* γ^K, *given by* $Y(Z) = \sum_{i=0}^{K} a_i Z^i$, *where*

$$a_K = 1, \quad a_j = -\sum_{i=j+1}^{K}\binom{i}{j} a_i b_{i-j}, \qquad j = K-1, \dots, 1, 0, \tag{8.4}$$

with $b_k = \mathbf{E}(Z^k \mid \gamma = 0)$, $k = 0, 1, \dots, K$. *Furthermore, if* $K \geq 3$ *is an odd integer and* $f(z)$ *is a nondegenerate symmetric and unimodal density, then* $Y(z)$ *must have at least one positive root.*

An immediate corollary of Theorem 3 and Lemma 4 is that for the normal family $N(\gamma, 1)$ family and many other complete symmetric unimodal location families (e.g., Laplace), there is no SOUP for odd powers of γ. This implies that of all positive integer powers of the location parameter, only the location parameter itself has a SOUP, since there is obviously no SOUP for even powers because $\gamma = 0$ is the boundary of the transformed parameter (i.e., γ^{2k}) space and $f(z \mid \gamma = 0)$ is not degenerate.

It is interesting to observe that for the normal family $N(\gamma, 1)$, Hartigan's (1965) asymptotic construction for $\theta = \gamma^K$, $K \geq 3$ an odd integer, not only does not yield a SOUP but in fact gives an improper posterior for all n. This is because the

Fisher information for $N(\theta^{1/K}, n^{-1})$ is $I_n(\theta) \propto \theta^{-2(K-1)/K}$, so the corresponding posterior measure is $P(\theta \mid \mathbf{X}) \propto \theta^{-2(K-1)/K} \exp(-\frac{n(\bar{X}_n - \theta^{1/K})^2}{2})$, which is not a proper density on R for K an odd integer, $K \geq 3$. This further complicates the determination of when Hartigan's (1965) asymptotic method gives SOUP for parameters of continuous exponential families.

8.3. *SOUP and the inverse variance rule.* Outside continuous exponential families, Hartigan's (1965) asymptotic Fisher information prior (7.3) is obviously unlikely to yield a SOUP: Fisher information may not even be defined. However, in each of the examples of Table 1, we have

$$\text{SOUP} \propto \frac{1}{\text{variance function}}, \tag{8.5}$$

an obvious extension of Hartigan's information rule (5.2). In particularly, (8.5) holds for our examples from discrete exponential families and nonexponential families, for which neither Hartigan's (1965) exact result nor Diaconis and Ylvisaker's (1979) Theorem 2 is applicable.

Attempting to find a unifying principle that would explain (8.5), we have identified other common features. For example, in the first and third pairs, the parameter of interest is a convolution parameter. However, (8.5) does not yield a SOUP for the convolution family generated by a variable taking three values; in fact, there is no such SOUP. In all ten cases of Table 1, the variances are quadratic functions of the means (and of the parameters of interest), leading us to think along the lines of Morris (1982, 1983); on the other hand, the variance is cubic in the mean for the inverse Gaussian family, which has no SOUP (Section 8.1). But a counterexample to (8.5) also has this "quadratic" property: for a discrete scale family on an "exponential lattice," the SOUP for the scale parameter is the inverse of the standard deviation, not of the variance (Section 6.2). These unsuccessful attempts cause us to believe that there are some subtle reasons for (8.5), and that one key to understanding (8.5) is to define "natural observation" and "natural parameter" for nonexponential families. The works of Diaconis and Ylvisaker (1985), Consonni and Veronese (1992) and Gutiérrez-Peña and Smith (1995, 1997), among others, may be useful for generalizing Hartigan's results, but we have not been successful in developing the required extensions.

Finally we remark on another intriguing phenomenon. Firth (1993) demonstrated that the use of Jeffreys' prior, namely, $\pi(\theta) \propto I_f^{1/2}(\theta)$, removes the first-order bias (i.e., the n^{-1} term) in the *posterior mode* of the canonical parameter of an exponential family. However, as shown in Section 5.1, the use of the square of Jeffreys' prior, namely the Fisher information, removes all the bias in the *posterior mean* for the same parameter. It appears that the invariance of an estimator to one-to-one transformation, which holds for posterior mode but not for posterior expectation, is the mathematical reason for the discrepancy, which is exactly the

Jacobian factor, as made clear in Section 7.2. This observation is further substantiated by the fact that for posterior intervals, which are also invariant to one-to-one transformation, the use of Jeffreys' prior removes the first-order bias (i.e., the $n^{-1/2}$ term) in the frequentist coverage for univariate problems [e.g., Welch and Peers (1963), Stein (1986), Nicolaou (1993)]. More statistical insight into this discrepancy would be helpful in understanding the use of Fisher information as a default prior.

APPENDIX A

PROOF OF LEMMAS 1 AND 2. In both lemmas, $\mathbf{E}[|g(X)X|] < \infty$ implies $\mathbf{E}[|g(X)|] < \infty$ because $|g(x)| \le |g(x)x|$ for $x = 1, 2, \ldots$. The proof of Lemma 1 then follows from $\mathbf{E}_{n,p}[g(X)X] = np\mathbf{E}_{n-1,p}[g(X+1)]$, which is easy to verify, and $\mathbf{E}_{n,p}[g(X)] = p\mathbf{E}_{n-1,p}[g(X+1)] + (1-p)\mathbf{E}_{n-1,p}[g(X)]$, a consequence of the divisibility of $\mathrm{B}(n, p)$.

The proof of Lemma 2 is more involved. First, it is easy to verify directly that $\mathbf{E}_{r+1,p}[g(X)] = \frac{p}{r(1-p)}\mathbf{E}_{r,p}[g(X)(X-r)]$. Then, since $\mathbf{E}_{r,p}(X) = r/p$, the left-hand side of (4.2) is

$$
\begin{aligned}
&\mathbf{E}_{r,p}[g(X)(X-r)] - \frac{r(1-p)}{p}\mathbf{E}_{r,p}[g(X)]\\
&\quad = \frac{r(1-p)}{p}\{\mathbf{E}_{r+1,p}[g(X)] - \mathbf{E}_{r,p}[g(X)]\}\\
&\quad = \frac{r(1-p)}{p^2}\Biggl[\mathbf{E}_{r+1,p}[g(X)](1-(1-p))\\
&\qquad\qquad - \sum_{k=r}^{\infty} g(k)\binom{k-1}{r-1} p^{r+1}(1-p)^{k-r}\Biggr]\\
&\quad = \frac{r(1-p)}{p^2}\Biggl[\mathbf{E}_{r+1,p}[g(X)]\\
&\qquad\qquad - \sum_{k=r}^{\infty} g(k)\biggl[\binom{k-1}{r} + \binom{k-1}{r-1}\biggr] p^{r+1}(1-p)^{k-r}\Biggr]\\
&\quad = \frac{r(1-p)}{p^2}\Biggl[\mathbf{E}_{r+1,p}[g(X)] - \sum_{k=r+1}^{\infty} g(k-1)\binom{k-1}{r} p^{r+1}(1-p)^{k-(r+1)}\Biggr],
\end{aligned}
$$

which is the same as the right-hand side of (4.2). Note in the third equation $\binom{r-1}{r}$ is defined to be zero. $\square$

REMARK. This result is different from (5.2) of Hudson (1978),

$$(1-p)\mathbf{E}_{r,p}[g(X)] = \mathbf{E}_{r,p}\left[\frac{X-r}{X-1}g(X-1)\right],$$

which we have not been able to use for verifying O-completeness.

APPENDIX B

PROOF OF THE NONEXISTENCE OF SOUP FOR IG$(\theta, 1)$. We prove by contradiction. Suppose there is a nonnegative real measurable function $\pi(\theta)$, the value $+\infty$ allowed, on $(0, \infty)$ such that (2.2) is satisfied with $\hat{\theta}(y) = y$ for all $y > 0$, where $f(y \mid \theta)$ is the density of IG$(\theta, 1)$. This implies, by letting $t = (2\theta^2)^{-1}$,

$$\text{(B.1)} \qquad 0 < \int_0^\infty g(t)e^{-yt}\,dt = y\int_0^\infty \sqrt{2t}g(t)e^{-yt}\,dt < \infty \qquad \forall y > 0,$$

where $g(t) = (2t)^{-2}\pi(1/\sqrt{2t})e^{\sqrt{2t}}$. Because $e^{-yt} = y\int_t^\infty e^{-sy}\,ds$ for any $t, y > 0$, and because all integrands are nonnegative, by Fubini's theorem, (B.1) implies

$$\text{(B.2)} \qquad 0 < \int_0^\infty G(t)e^{-yt}\,dt = \int_0^\infty \sqrt{2t}g(t)e^{-yt}\,dt < \infty \qquad \forall y > 0,$$

where $0 \le G(t) = \int_0^t g(s)\,ds$, which is finite for any $t \in (0, \infty)$ because of its monotonicity and integrability with respect to $e^{-t}\,dt$. By the uniqueness of the inverse Laplace transform, (B.2) implies $g(t) = G(t)/\sqrt{2t}$ (a.e. Lebesgue). Define $\tilde{g}(t) = G(t)/\sqrt{2t}$ for all $t \in (0, \infty)$, and since the value of g on a set of measure 0 does not alter G, $G(t) = \int_0^t \tilde{g}(s)\,ds$ for all $t \in (0, \infty)$. Since $G(t)$ is absolutely continuous, $\tilde{g}(t)$ is also absolutely continuous, which in turn implies $G(t)$ is differentiable and $G'(t) = \tilde{g}(t)$ for all $t \in (\varepsilon, \infty)$, for any $\varepsilon > 0$. It follows that $G(t) = \sqrt{2t}G'(t)$ for all $t \in (0, \infty)$, implying $G(t) = ce^{\sqrt{2t}}$ for all $t > 0$, where c is some nonnegative constant. However, since $G(t) = \int_0^t \tilde{g}(s)\,ds \downarrow 0$ when $t \downarrow 0$, $C = 0$. This implies $\tilde{g}(t) \equiv 0$, or equivalently $g(t) = 0$ (a.e. Lebesgue), which is not possible if (B.1) holds. □

REMARK. The solution $G(t) = ce^{\sqrt{2t}}$ is the same as $\pi(\theta) = c\theta^{-3}$, the inverse variance rule. However, the nondegeneracy of $f(y \mid \theta = 1/\sqrt{2t})$ at $t = 0$ forces $\lim_{t\downarrow 0} G(t) = 0$, leading to the contradiction.

APPENDIX C

PROOF OF LEMMA 4. First we give a unique construction for an unbiased polynomial estimator of γ^K of degree K. If such an estimator $Y(Z) = \sum_{i=0}^K a_i Z^i$ exists, then

$$\text{(C.1)} \qquad \gamma^K = \mathbf{E}[Y(Z) \mid \gamma] = \sum_{j=0}^K \left(\sum_{i=j}^K a_i \binom{i}{j} b_{i-j}\right)\gamma^j \qquad \forall \gamma \in R.$$

This implies $a_K = 1$ and all the coefficients of γ^j are zero for all $j = 0, \ldots, K-1$, which leads to (8.4).

When $K = 2M + 1 \geq 3$ and f is symmetric and nondegenerate, $b_{2m} > 0$ and $b_{2m+1} = 0$ for $m = 0, 1, \ldots, M$. It follows then, by induction, $a_{2m} = 0$ for all $m = 0, 1, \ldots, M$ (or use the fact that $Y(z)$ has to be an odd function because of its uniqueness). Let $0 \leq r < M$ be the smallest integer such that $a_{2r+1} \neq 0$. Then $Y(z) = z^{2r+1} g(z^2)$, where $g(w) = \sum_{t=r}^{M} a_{2t+1} w^{t-r}$ and $g(0) = a_{2r+1} \neq 0$. It suffices to prove that $g(w)$ must have at least one positive root. If it does not, then since the leading coefficient in $g(w)$ is 1, $g(w)$ must be bounded below by some $g_{\min} > 0$. Analogously to the proof of Theorem 3, using the symmetry and unimodality of f, the unbiasedness condition can be reexpressed as

$$\begin{aligned}\gamma^{2M+1} &= \int_0^\infty z^{2r+1} g(z^2)\big(f(z-\gamma) - f(z+\gamma)\big)\,dz \\ &\geq g_{\min} \int_0^\infty z^{2r+1}\big(f(z-\gamma) - f(z+\gamma)\big)\,dz \\ &= g_{\min} \int_{-\infty}^\infty z^{2r+1} f(z-\gamma)\,dz,\end{aligned}$$

or equivalently,

$$\gamma^{2M+1} \geq g_{\min} \sum_{j=0}^{r} \binom{2r+1}{2j+1} b_{2(r-j)} \gamma^{2j+1} > 0 \qquad \forall \gamma > 0. \tag{C.2}$$

However, (C.2) is not possible when $\gamma \downarrow 0$, because the middle expression is a polynomial in γ of order less than $2M + 1$ and thus goes to zero slower than γ^{2M+1}. □

Acknowledgments. We thank the late R. Bahadur, P. Diaconis, J. Hartigan, P. McCullagh, M. Stein, S. Stigler and M. Wichura for providing helpful ingredients during our eleven years of cooking SOUP; we started when both of us were students at Harvard and thus the majority of work reported here was done when Meng was at the University of Chicago. We also thank the Editor, J. Berger, and several reviewers for constructive comments which led to improvement of the presentation.

REFERENCES

BERGER, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. Springer, New York.

BERGER, J. O. (1990). On the inadmissibility of unbiased estimators. *Statist. Probab. Lett.* **9** 381–384.

BERGER, J. O. and SRINIVASAN, C. (1978). Generalized Bayes estimators in multivariate problems. *Ann. Statist.* **6** 783–801.

BICKEL, P. J. and MALLOWS, C. L. (1988). A note on unbiased Bayes estimates. *Amer. Statist.* **42** 132–134.

BROWN, L. D. (1971). Admissible estimates, recurrent diffusions, and insoluble boundary value problems. *Ann. Math. Statist.* **42** 855–903.

BROWN, L. D. (1979). A heuristic method for determining admissibility of estimators, with applications. *Ann. Statist.* **7** 960–994.

BROWN, L. D. (1980). A necessary condition for admissibility. *Ann. Statist.* **8** 540–544.

CONSONNI, G. and VERONESE, P. (1992). Conjugate priors for exponential families having quadratic variance functions. *J. Amer. Statist. Assoc.* **87** 1123–1127.

CONSONNI, G. and VERONESE, P. (1993). Unbiased Bayes estimates and improper priors. *Ann. Inst. Statist. Math.* **45** 303–315.

DIACONIS, P. and YLVISAKER, D. (1979). Conjugate priors for exponential families. *Ann. Statist.* **7** 269–281.

DIACONIS, P. and YLVISAKER, D. (1985). Quantifying prior opinion. In *Bayesian Statistics 2* (J. M. Bernardo, D. V. Lindley and A. F. M. Smith, eds.) 133–156. North-Holland, Amsterdam.

FIRTH, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika* **80** 27–38. [Correction (1995) **82** 667.]

GUTIÉRREZ-PEÑA, E. and SMITH, A. F. M. (1995). Conjugate parameterizations for natural exponential families. *J. Amer. Statist. Assoc.* **90** 1347–1356. [Correction (1996) **91** 1757.]

GUTIÉRREZ-PEÑA, E. and SMITH, A. F. M. (1997). Exponential and Bayesian conjugate families: Review and extensions (with discussion). *Test* **6** 1–90.

HARTIGAN, J. (1964). Invariant prior distributions. *Ann. Math. Statist.* **35** 836–845.

HARTIGAN, J. (1965). The asymptotically unbiased prior distribution. *Ann. Math. Statist.* **36** 1137–1152.

HARTIGAN, J. (1998). The maximum likelihood prior. *Ann. Statist.* **26** 2083–2103.

HOGAN, H. (1993). The 1990 post-enumeration survey: Operations and results. *J. Amer. Statist. Assoc.* **88** 1047–1060.

HSIEH, H. K., KORWAR, R. M. and RUKHIN, A. L. (1990). Inadmissibility of the maximum likelihood estimator of the inverse Gaussian mean. *Statist. Probab. Lett.* **9** 83–90.

HUDSON, H. M. (1978). A natural identity for exponential families with applications in multiparameter estimation. *Ann. Statist.* **6** 473–484.

HWANG, J. T. (1982). Improving upon standard estimators in discrete exponential families with applications to Poisson and negative binomial cases. *Ann. Statist.* **10** 857–867.

JOHNSON, N. L. (1957). Uniqueness of a result in the theory of accident proneness. *Biometrika* **44** 530–531.

KASS, R. E. and WASSERMAN, L. (1996). The selection of prior distributions by formal rules. *J. Amer. Statist. Assoc.* **91** 1343–1370.

KENDALL, M., STUART, A. and ORD, J. K. (1987). *Kendall's Advanced Theory of Statistics* **1**. *Distribution Theory*, 5th ed. Oxford, New York.

LEHMANN, E. L. (1983). *Theory of Point Estimation*. Wiley, New York.

MENG, X. L. (1994). Multiple-imputation inferences with uncongenial sources of input (with discussion). *Statist. Sci.* **9** 538–573.

MORRIS, C. N. (1982). Natural exponential families with quadratic variance functions. *Ann. Statist.* **10** 65–80.

MORRIS, C. N. (1983). Natural exponential families with quadratic variance functions: Statistical theory. *Ann. Statist.* **11** 515–529.

NICOLAOU, A. (1993). Bayesian intervals with good frequentist behavior in the presence of nuisance parameters. *J. Roy. Statist. Soc. Ser. B* **55** 377–390.

RUBIN, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. Wiley, New York.

RUBIN, D. B. (1996). Multiple imputation after 18+ years (with discussion). *J. Amer. Statist. Assoc.* **91** 473–520.

RUKHIN, A. L. (1995). Admissibility: A survey of a concept in progress. *Inter. Statist. Rev.* **63** 95–115.

SCHIRM, A. L. and PRESTON, S. H. (1992). Comment on "Should we have adjusted the U.S. Census of 1980?" by D. A. Freedman and W. C. Navidi. *Survey Methodology* **18** 35–43.

STEIN, C. (1986). On the coverage probability of confidence sets based on prior distributions. In *Sequential Methods in Statistics*, 3rd ed. (G. B. Wetherbil and K.D. Glazebrook, eds.) 485–514. Chapman and Hall, London.

STIGLER, S. M. (1972). Completeness and unbiased estimation. *Amer. Statist.* **26** 28–29.

VOINOV, V. G. and NIKULIN, M. S. (1993). *Unbiased Estimators and Their Applications* **1**. *Univariate Case*. Kluwer, Dordrecht.

WELCH, B. L. and PEERS, H. W. (1963). On formulae for confidence points based on integrals of weighted likelihoods. *J. Roy. Statist. Soc. Ser. B* **25** 318–329.

ZASLAVSKY, A. M. (1989). Representing the census undercount: Reweighting and imputation methods. Ph.D. dissertation, Dept. Mathematics, MIT.

DEPARTMENT OF STATISTICS
HARVARD UNIVERSITY
1 OXFORD STREET
CAMBRIDGE, MASSACHUSETTS 02138
E-MAIL: meng@stat.harvard.edu

DEPARTMENT OF HEALTH CARE POLICY
HARVARD MEDICAL SCHOOL
180 LONGWOOD AVE
BOSTON, MASSACHUSETTS 02115-5899
E-MAIL: zaslavsky@hcp.med.harvard.edu