Conjugate Priors for Exponential Families

Author(s): Persi Diaconis and Donald Ylvisaker

Source: *The Annals of Statistics*, Mar., 1979, Vol. 7, No. 2 (Mar., 1979), pp. 269-281

Published by: Institute of Mathematical Statistics

Stable URL: https://www.jstor.org/stable/2958808

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at https://about.jstor.org/terms

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve and extend access to *The Annals of Statistics*

This content downloaded from
128.148.194.11 on Tue, 06 Feb 2024 17:13:44 +00:00
All use subject to https://about.jstor.org/terms

The Annals of Statistics
1979, Vol. 7, No. 2, 269–281

CONJUGATE PRIORS FOR EXPONENTIAL FAMILIES

BY PERSI DIACONIS[1] AND DONALD YLVISAKER[2]

Stanford University and The University of California, Los Angeles

Let X be a random vector distributed according to an exponential family with natural parameter $\theta \in \Theta$. We characterize conjugate prior measures on Θ through the property of linear posterior expectation of the mean parameter of X : $E\{E(X|\theta)|X = x\} = ax + b$. We also delineate which hyperparameters permit such conjugate priors to be proper.

1. Introduction. Modern Bayesian statistics is dominated by the notion of conjugate priors. The usual definition is that a family of priors is conjugate if it is closed under sampling (Lindley [1972], pages 22–23 or Raiffa and Schlaifer [1961], pages 43–57). Consider the following example: let S_n be the number of heads in n independent tosses of a coin with unknown parameter p. The accepted family of conjugate priors for p is the beta family with densities

$$f(p; \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}, \qquad \alpha > 0, \beta > 0.$$

Let h be any positive bounded measurable function on the unit interval and observe that a prior density proportional to $h(p)f(p; \alpha, \beta)$ leads to a posterior density of p, given $S_n = x$, proportional to $h(p)f(p; \alpha + x, \beta + h - x)$. Thus, the family $\{h(\cdot)f(\cdot; \alpha, \beta)|\alpha > 0, \beta > 0, h \text{ positive, bounded, measurable}\}$ with each member normalized to be a prior density, is closed under sampling. Now beta priors have the additional property that the posterior expectation of the mean parameter p is a linear function of S_n. That is, there are numbers a_n, b_n such that

$$E[p|S_n = k] = \frac{\int_0^1 p^{k+1}(1-p)^{n-k} f(p; \alpha, \beta)dp}{\int_0^1 p^k(1-p)^{n-k} f(p; \alpha, \beta)dp} = a_n k + b_n$$

holds for $k = 0, 1, 2, \ldots, n$. A principal result of this paper is that, subject to regularity conditions, the conjugate priors typically used satisfy, and are characterized by, a similar relation of posterior linearity:

$$E\{E(X|\theta)|X = x\} = ax + b. \tag{1.1}$$

The regularity conditions assumed below allow such standard examples as the normal prior for normal location, the gamma prior for the Poisson, the inverse Wishart prior for normal covariance, and the beta prior for the negative binomial.

Received September 1977; revised April 1978.

[1]Research supported in part by NSF Grant MPS74-21416 and by the Energy Research and Development Administration under contract EY-76-C-03-0515.

[2]Research supported in part by NSF Grants MPS 72-4591, MPS 75-7120 and MCS 77-02121.

AMS 1970 *subject classifications.* 62E10, 62A15.

Key words and phrases. Conjugate priors, linearity of regression, Bayesian analysis, characterization theorems, exponential families, credibility theory, admissibility.

This content downloaded from
128.148.194.11 on Tue, 06 Feb 2024 17:13:44 +00:00
All use subject to https://about.jstor.org/terms

The Dirichlet prior for the multinomial is also suitably characterized but requires separate treatment. Our results are also of interest in the following two contexts:

Credibility theory and linear Bayesian analysis. Linear Bayes prediction has been used since 1920 by the actuarial professional under the heading of credibility theory (Kahn [1975]). Our Theorem 2, that (1.1) holds for exponential families with their customary conjugate priors, is a rigorous treatment of some recent results of Jewell [1974b] on what is there termed exact credibility. In work unconnected with credibility theory per se, Ericson [1969], [1970] noted that when (1.1) holds, a and b can be given expression in terms of the means and variances of the underlying distributions. Independently, Hartigan [1969] made essentially the same observation and went on to use the a and b so determined to form a linear Bayes predictor. Efron and Morris [1973] have an extension of the empirical Bayes approach they developed for normal location problems to situations where the Bayes estimate is linear Bayes. Now, in fact, when Theorems 3, 4 and 5 below are in force, they imply that the assumption of (1.1) for fixed a and b is exactly the assumption of a specific prior distribution.

Admissibility and Karlin's theorem. When is the estimate $aX + b$ an admissible estimate of $E(X|\theta)$ with squared error as loss? A sufficient condition for one-dimensional exponential families was given by Karlin [1958] and discussed further by Ping [1964], Cohen [1966] and Stone [1967]. Our Theorems 1 and 2 give a simple interpretation of Karlin's condition and n dimensional extensions when $0 < a < 1$, and the parameter space Θ is open. Then, if $\pi_{b/(1-a),\,(1-a)/a}$ denotes the conjugate prior defined in (2-3), Karlin's sufficient condition becomes

$$\int_u^v \pi^{-1}_{\frac{b}{1-a},\,\frac{1-a}{a}}(\theta)\,d\theta \to \infty$$

as u and v approach the boundary of Θ. It is easy to show that this is equivalent to asking that $\pi_{b/(1-a),\,(1-a)/a}$ be a proper prior. Theorems 1 and 2 show that if $0 < a < 1$ and $b/(1-a)$ is in the set $\mathcal{X}$ defined in Section 2, then $aX + b$ is a proper Bayes estimate with respect to $\pi_{b/(1-a),\,(1-a)/a}$. Since the Bayes risk can be seen to be finite, admissibility follows. Of course, Karlin's result leads to admissibility when $a = 1$. Stein's result on the inadmissibility of the mean of a multivariate normal shows that $a = 1$ need not be admissible in three or more dimensions.

The characterization theorems here have been given previously in special cases. Johnson [1956], [1967] characterized the gamma prior for the Poisson mean and related results are given in Goldstein [1975] and in Chapters 5 and 6 of Kagan, Linnik and Rao [1973]. The results given here are considerably more general and, in some cases, more precise than those previously found.

Section 2 of this paper studies the usual notion of conjugate prior and establishes (1.1) under mild regularity conditions. Moreover, Theorem 1 gives precise conditions on the "hyperparameters" of the conjugate prior to guarantee integrability (see Novick and Hall [1965] in this connection). Section 3 is devoted to a proof that (1.1) characterizes conjugate priors when the observation space is sufficiently rich.

This content downloaded from
128.148.194.11 on Tue, 06 Feb 2024 17:13:44 +00:00
All use subject to https://about.jstor.org/terms

Section 3 also comments on the possibility of (1.1) holding when a is a matrix. In Section 4 we consider problems particular to the case of a discrete observation space.

2. Conjugate priors in exponential families. This section contains requisite notation and terminology associated with a d-parameter exponential family of distributions. Depending on the setting, Theorem 1 gives sufficient or necessary and sufficient conditions on the "hyperparameters" of a conjugate prior distribution for it to be proper. Theorem 2 then establishes linear posterior expectation under regularity conditions. Theorem 2 has been in the folklore of the subject and a full proof for the 1-dimensional case has recently appeared in Jewell [1974a], [1975]. The section closes with a brief Bayesian interpretation of Theorems 1 and 2.

Start with a fixed σ-finite measure μ on the Borel sets of R^d. Consider the convex hull of the support set of the measure μ, and then let $\mathcal{X}$ be the interior of this convex set. It will always be assumed that $\mathcal{X}$ is a nonempty open set in R^d, so that the observation set is genuinely d-dimensional. For $\theta \in R^d$, define $M(\theta) = \ln \int e^{\theta \cdot x} d\mu(x)$ and let $\Theta = \{\theta | M(\theta) < \infty\}$. The standard use of Hölder's inequality in this context shows that Θ is a convex set—it is called the *natural parameter space*. It is further assumed that Θ is a nonempty open set in R^d—in the terminology of Barndorff-Neilson [1970], we restrict attention to regular exponential families. The openness of Θ is indeed a regularity condition on the measure μ—one which is employed crucially in Theorem 2.

The *exponential family* $\{P_\theta\}$ of probability measures *through* μ is determined by

$$dP_\theta(x) = e^{x \cdot \theta - M(\theta)} d\mu(x), \qquad \Theta \in \theta. \tag{2.1}$$

Expectation under P_θ will be denoted by E_θ or $E\{\ |\theta\}$. Now suppose X is a random vector with distribution P_θ. Then if one differentiates the identity $\int_{\mathcal{X}} dP_\theta(x) = 1$ in θ and makes admissible interchanges of differentiation and integration, one finds

$$\text{(i)}\ E(X|\theta) = E_\theta(X) = \nabla M(\theta) = \left(\frac{\partial M(\theta)}{\partial \theta_1}, \ldots, \frac{\partial M(\theta)}{\partial \theta_d} \right)' = (M_1(\theta), \ldots, M_d(\theta))'$$

(2.2)

$$\text{(ii)}\ E_\theta(X - \nabla M(\theta))(X - \nabla M(\theta))' = M''(\theta) = \left(\frac{\partial^2 M(\theta)}{\partial \theta_i \partial \theta_j} \right)_{i,j=1}^{d} = (M_{ij}(\theta))_{i,j=1}^{d}.$$

Because $\mathcal{X}$ is assumed open in R^d, the Hessian $M''(\theta)$ must be positive definite at each θ—for otherwise there is a θ_0 and a vector $c \neq 0$ so that $c'(X - \nabla M(\theta_0)) = 0$ a.s. P_{θ_0} and then a.e. μ. Furthermore, from (2.2i), $\nabla M(\theta)$ must be in the convex hull of the support of μ. It is then easy to see that $\nabla M(\theta)$ cannot be a boundary point of this convex set, so $\nabla M(\theta) \in \mathcal{X}$ for each $\theta \in \Theta$.

Recall that Θ is to be a nonempty convex open set in R^d and let $d\theta$ denote the Lebesgue measure on the Borel sets of Θ. Define a family $\{\tilde{\pi}_{n_0, x_0}\}$ of measures on the same space according to

$$d\tilde{\pi}_{n_0, x_0}(\theta) = e^{n_0 x_0 \cdot \theta - n_0 M(\theta)} d\theta, \qquad n_0 \in R^1, x_0 \in R^d. \tag{2.3}$$

This content downloaded from
128.148.194.11 on Tue, 06 Feb 2024 17:13:44 +00:00
All use subject to https://about.jstor.org/terms

If $\tilde{\pi}_{n_0, x_0}$ can be normalized to a probability measure π_{n_0, x_0} on Θ, it will be termed a *distribution conjugate to the exponential family* $\{P_\theta\}$ of (2.1). The province of the parameter (n_0, x_0) which allows such a normalization is the subject matter of the first theorem. A Bayesian interpretation of the theorem appears at the end of this section.

THEOREM 1. *If* $n_0 > 0$ *and* $x_0 \in \mathcal{X}$, *then* $\tilde{\pi}_{n_0, x_0}(\Theta) < \infty$. *Conversely, if* $\tilde{\pi}_{n_0, x_0}(\Theta) < \infty$ *and* $\Theta = R^d$ *then* $n_0 > 0$; *while if* $\tilde{\pi}_{n_0, x_0}(\Theta) < \infty$ *with* $n_0 > 0$, *then* $x_0 \in \mathcal{X}$.

PROOF. For the forward portion of the theorem, we first bound $e^{-M(\theta)}$ as follows. Let A be a compact convex subset of R^d. For $\theta \in \Theta$, $\int_A e^{z\cdot\theta} d\mu(z) \leqslant \int e^{z\cdot\theta} d\mu(z) < \infty$ and therefore $\mu(A) < \infty$. Moreover if $\mu(A) > 0$, write $\mu_A(B) = \mu(A \cap B)/\mu(A)$, $x_A = \int z d\mu_A(z)$, and use Jensen's inequality to get

$$e^{-M(\theta)} = \left(\int e^{z\cdot\theta} d\mu(z)\right)^{-1} \leqslant \left[\mu(A)\right]^{-1}\left(\int e^{z\cdot\theta} d\mu_A(z)\right)^{-1} \leqslant \left[\mu(A)\right]^{-1} e^{-\theta\cdot x_A}. \tag{2.4}$$

The full force of (2.4) comes from the observation that means of the form x_A are dense in $\operatorname{supp}(\mu)$.

If $x_0 \in \mathcal{X}$, one can write $x_0 = \sum_{j=1}^{d+1} \lambda_j x_j$ where the λ_j are nonnegative and sum to 1, the x_j are in $\operatorname{supp}(\mu)$ and do not lie in a $(d-1)$-dimensional hyperplane. A dense set D of $\mathcal{X}$ can be so represented with the added requirement that the λ_j are positive. For $x_0 \in D$ we can then require positive λ_j together with compact convex sets A_j so that $x_j = x_{A_j}$, $j = 1, \ldots, d+1$, since these means are dense in $\operatorname{supp}(\mu)$. If $x_0 \in D$, write $\Theta_k = \Theta \cap \{\theta \mid \theta \cdot x_k = \max_j \theta \cdot x_j\}$ and invoke (2.4) to obtain

$$\begin{aligned} \int_\Theta e^{n_0 x_0\cdot\theta - n_0 M(\theta)} d\theta &= \sum_{k=1}^{d+1} \int_{\Theta_k} e^{n_0 x_0\cdot\theta - n_0 M(\theta)} d\theta \\ &\leqslant \sum_{k=1}^{d+1} \left[\mu(A_k)\right]^{-n_0} \int_{\Theta_k} \exp\left(n_0 \theta \cdot \left(\sum_{j=1}^{d+1} \lambda_j (x_j - x_k)\right)\right) d\theta. \end{aligned} \tag{2.5}$$

In the kth integral on the right side of (2.5), make the change of variable $\sigma_j = \theta \cdot (x_j - x_k)$, $j \neq k$, with $|J_k| = |\partial\sigma/\partial\theta|$, say. Since the x_j do not fall in a $(d-1)$-dimensional hyperplane, $|J_k| \neq 0$ and the right side of (2.5) becomes

$$\sum_{k=1}^{d+1} \left[\mu(A_k)\right]^{-n_0} |J_k|^{-1} \int_{\sigma(\Theta) \cap \{\sigma_j \leqslant 0, j \neq k\}} \exp\left(n_0 \sum_{j \neq k} \lambda_j \sigma_j\right) d\sigma < \infty.$$

Thus, $\tilde{\pi}_{n_0, x_0}(\Theta) < \infty$ on a dense set D of $\mathcal{X}$. For fixed $n_0 > 0$, $\tilde{\pi}_{n_0, x_0}(\Theta) < \infty$ on a convex set of x_0 according to Hölder's inequality—hence, it must be finite on all of $\mathcal{X}$.

For the converse direction, observe first that the integrand $f_{n_0, x_0}(\theta) = e^{n_0 x_0\cdot\theta - n_0 M(\theta)}$ satisfies

$$\begin{aligned} &\text{(i)} \quad \nabla f_{n_0, x_0}(\theta) = n_0 (x_0 - \nabla M(\theta)) f_{n_0, x_0}(\theta) \\ &\text{(ii)} \quad f''_{n_0, x_0}(\theta) = \left[n_0^2 (x_0 - \nabla M(\theta))(x_0 - \nabla M(\theta))' - n_0 M''(\theta)\right] f_{n_0, x_0}(\theta). \end{aligned} \tag{2.6}$$

This content downloaded from
128.148.194.11 on Tue, 06 Feb 2024 17:13:44 +00:00
All use subject to https://about.jstor.org/terms

If f_{n_0, x_0} were integrable with $n_0 < 0$ and $\Theta = R^d$, we would have by (2.6ii) the integrability of a positive convex function over all of R^d, a contradiction. So suppose $n_0 > 0$ and $x_0 \notin \mathcal{X}$. Choose a unit vector ξ_1 so that, by translating μ if necessary, $\xi_1 \cdot x_0 \geqslant 0$ and $\xi_1 \cdot x \leqslant 0$ for $x \in \mathcal{X}$. Let $\xi_1, \ldots, \xi_d$ be orthonormal in R^d and make the change of variable $\sigma_i = \xi_i \cdot \theta$, $i = 1, \ldots, d$, in $\int f_{n_0, x_0}(\theta) d\theta$. Since $\nabla M(\theta) \in \mathcal{X}$, $\xi_1 \cdot \nabla f_{n_0, x_0}(\theta) \geqslant 0$ for all θ by (2.6i). Thus, integration with respect to σ_1 is integration of a positive nondecreasing function. For f_{n_0, x_0} to be integrable it is necessary that $\xi_1 \cdot \theta$ be bounded above on Θ for almost every $(d\theta)$ choice of $\sigma_2, \ldots, \sigma_d$. Let $\theta_0 \in \Theta$ so that

$$\int e^{z \cdot \theta_0} d\mu(z) = \int \exp\left(\sum_{j=1}^d (z \cdot \xi_j)(\theta_0 \cdot \xi_j)\right) d\mu(z) < \infty.$$

Since $\xi_1 \cdot z \leqslant 0$ on $\mathcal{X}$, $\theta_0 + u\xi_1$ must also be in Θ for any positive u because $e^{z \cdot (\theta_0 + u\xi_1)} = e^{z \cdot \theta_0 + uz \cdot \xi_1} \leqslant e^{z \cdot \theta_0}$ on $\mathcal{X}$. However, $\xi_1 \cdot (\theta_0 + u\xi_1) = \xi_1 \cdot \theta_0 + u$ and, therefore, $\xi_1 \cdot \theta$ is not bounded above on Θ. This contradiction means that x_0 must be in $\mathcal{X}$ and the proof of the theorem is complete.

The following result unifies many standard Bayesian calculations.

THEOREM 2. *Suppose Θ is open in R^d. If θ has the distribution π_{n_0, x_0}, $n_0 > 0$ and $x_0 \in \mathcal{X}$, then $E(\nabla M(\theta)) = x_0$.*

PROOF. The required result translates through (2.6) to

$$\int_\Theta \nabla f_{n_0, x_0}(\theta) d\theta = 0 \tag{2.7}$$

Consider the first component of (2.7) and assume for now that Fubini's theorem applies in order to write

$$\int_\Theta \frac{\partial}{\partial \theta_1} f_{n_0, x_0}(\theta) d\theta = \int \cdots \int \left[\lim_{\theta_1 \to \bar{\theta}_1} f_{n_0, x_0}(\theta) - \liminf_{\theta_1 \to \underline{\theta}_1} f_{n_0, x_0}(\theta)\right] d\theta_2, \ldots, d\theta_d \tag{2.8}$$

where $\underline{\theta}_1 = \underline{\theta}_1(\theta_2, \ldots, \theta_d) < \theta_1 < \bar{\theta}_1 = \bar{\theta}_1(\theta_2, \ldots, \theta_d)$ for $\theta \in \Theta$. The last two limits will be shown to be zero. Consider the first of them when $\bar{\theta}_1 = +\infty$. Use (2.4) in conjunction with a set A so that $(x_0)_1 < (x_A)_1$, hold $\theta_2, \ldots, \theta_d$ fixed and let $\theta_1 \to \infty$ to see that the limit is zero in such a case. If $\bar{\theta}_1(\theta_2, \ldots, \theta_d) < \infty$, take θ_1^* so that $(\theta_1^*, \theta_2, \ldots, \theta_d)' \in \Theta$. Then

$$\int_{x_1 \leqslant 0} e^{\bar{\theta}_1 x_1 + \cdots + \theta_d x_d} d\mu(x) \leqslant \int_{x_1 \leqslant 0} e^{\theta_1^* x_1 + \cdots + \theta_d x_d} d\mu(x) < \infty.$$

Now since $M(\bar{\theta}_1, \theta_2, \ldots, \theta_d) = +\infty$,

$$\int_{x_1 > 0} e^{\theta_1 x_1 + \cdots + \theta_d x_d} d\mu(x) \to \int_{x_1 > 0} e^{\bar{\theta}_1 x_1 + \cdots + \theta_d x_d} d\mu(x) = +\infty,$$

as $\theta_1^* \leqslant \theta_1 \to \bar{\theta}_1 = \bar{\theta}_1(\theta_2, \ldots, \theta_d)$, by monotone convergence. Thus $M(\theta_1, \ldots, \theta_d) \to \infty$ as $\theta_1 \to \bar{\theta}_1$, and the first limit on the right side of (2.8) is zero. A similar argument applies to the second limit.

It remains to be seen that Fubini's theorem has been correctly applied at (2.8). From (2.2ii) with $\theta_2, \ldots, \theta_d$ fixed, $(\partial/\partial\theta_1) M(\theta)$ is a strictly increasing function of

This content downloaded from
128.148.194.11 on Tue, 06 Feb 2024 17:13:44 +00:00
All use subject to https://about.jstor.org/terms

θ_1. Thus $(\partial/\partial\theta_1)f_{n_0, x_0}(\theta)$ changes sign at most once from positive to negative as θ_1 varies over $(\underline{\theta}_1, \bar{\theta}_1)$, (from (2.6i)). In particular, one deduces from the argument of the previous paragraph that there is a unique point $\theta_1^* = \theta_1^*(\theta_2, \ldots, \theta_d)$ at which $(\partial/\partial\theta_1)f_{n_0, x_0}(\theta) = 0$ and $f_{n_0, x_0}(\theta)$ has a maximum. Hence with $\theta^* = (\theta_1^*, \theta_2, \ldots, \theta_d)'$,

$$\int \left| \frac{\partial}{\partial \theta_1} f_{n_0, x_0}(\theta) \right| d\theta_1 = 2f_{n_0, x_0}(\theta^*) = 2\max_{\theta_1} f_{n_0, x_0}(\theta).$$

Absolute integrability in the left side of (2.8) evidently requires the integrability of $f_{n_0, x_0}(\theta^*)$ in $\theta_2, \ldots, \theta_d$. To proceed on this, let $x_0 = \sum_{j=1}^{d+1} \lambda_j x_j$ where the λ_j are nonnegative and sum to 1, the x_j do not lie in a $(d-1)$-dimensional hyperplane and $x_j = x_{A_j}$ for compact convex A_j, $j = 1, \ldots, d+1$. Moreover, by translating μ if necessary, choose $x_0 = 0$. Then from (2.4),

$$\begin{aligned} (2.9)\quad & \int \cdots \int \max_{\theta_1} f_{n_0, x_0}(\theta) d\theta_2, \ldots, d\theta_d \\ & \leqslant \int \cdots \int \max_{\theta_1} \min_k [\mu(A_n)]^{-n_0} e^{-n_0 \theta \cdot x_k} d\theta_2, \ldots, d\theta_d \\ & \leqslant c \int \cdots \int \exp(-n_0 \min_{\theta_1} \max_k \theta \cdot x_k) d\theta_2, \ldots, d\theta_d. \end{aligned}$$

For $0 \neq \theta^{(2)} = (\theta_2, \ldots, \theta_d)' = \|\theta^{(2)}\|\eta$ one finds

$$\begin{aligned} \min_{\theta_1} \max_k \left[\theta_1 x_{k1} + \theta_2 x_{k2} + \cdots + \theta_d x_{kd} \right] &= \|\theta^{(2)}\| \min_{\theta_1} \max_k \left[\frac{\theta_1}{\|\theta^{(2)}\|} x_{k1} + \eta \cdot x_k^{(2)} \right] \\ &= \|\theta^{(2)}\| \min_{\theta_1} \max_k \left[\theta_1 x_{k1} + \eta \cdot x_k^{(2)} \right]. \end{aligned}$$

If

$$\inf_{\|\eta\|=1} \min_{\theta_1} \max_k \left[\theta_1 x_{k1} + \eta \cdot x_k^{(2)} \right] = \delta$$

is positive, the right side of (2.9) is bounded by

$$c \int_{R^{d-1}} \cdots \int e^{-n_0 \|\theta^{(2)}\| \delta} d\theta_2, \ldots, d\theta_d < \infty.$$

To see that δ is positive, observe that $0 = \theta \cdot \sum_{j=1}^{d+1} \lambda_j x_j = \sum_{j=1}^{d+1} \lambda_j \theta \cdot x_j$ so $\max_k \theta \cdot x_k \geqslant 0$ for any θ. Then $\min_{\theta_1} \max_k \theta \cdot x_k \geqslant 0$ for any $\theta^{(2)}$. But $\min_{\theta_1} \max_k \theta \cdot x_k$ is a continuous function in $\theta^{(2)}$ so $\inf_{\|\eta\|=1} \min_{\theta_1} \max_k [\theta_1 x_{k1} + \eta \cdot x_k^{(2)}] = \delta > 0$ unless there is an η^* with $\|\eta^*\| = 1$ so that $\min_{\theta_1} \max_k [\theta_1 x_{k1} + \eta^* \cdot x_k^{(2)}] = 0$. Now if this were the case, there would be a vector $(\theta_1, \eta^{*\prime})' = \xi$ with $\max_k \xi \cdot x_k = 0$ and also, since $\sum \lambda_j \xi \cdot x_k = 0$, $\min_k \xi \cdot x_k = 0$. This contradicts the fact that the x_k are not contained in a $d-1$ dimensional hyperplane. The proof of Theorem 2 is completed by applying the same arguments to the other coordinates in (2.7).

REMARKS. To apply Theorem 2 to a sample $X_1, \ldots, X_n$ of size n from P_θ note that if π_{n_0, x_0} is the prior distribution of θ, the posterior distribution is $\pi_{n_0+n, (n_0 x_0 + n\bar{X})/(n_0+n)}$ with $\bar{X}$ the mean of the sample. Theorem 2 yields

$$(2.10)\qquad E\{\nabla M(\theta) | X_1, \ldots, X_n\} = \frac{n_0 x_0 + n\bar{X}}{n_0 + n},$$

i.e., the conditional expectation of the mean parameter is a linear combination of

This content downloaded from
128.148.194.11 on Tue, 06 Feb 2024 17:13:44 +00:00
All use subject to https://about.jstor.org/terms

the prior expectation of the mean parameter x_0 and $\bar{X}$. The weights in the linear combination are proportional to n_0 and the sample size—in this sense n_0 might be thought of as a prior sample size. The restrictions of Theorem 1 on n_0, x_0 to guarantee proper conjugate priors are generally consistent with this interpretation.

The openness of Θ is not used in the proof of Theorem 1. The situation there with $n_0 < 0$ is inconclusive if $\Theta \neq R^d$. In fact, one can have finiteness with $n_0 < 0$. For example, for the exponential family corresponding to the geometric distribution on the nonnegative integers in dimension 1, $\tilde{\pi}_{n_0, x_0}(\Theta) < \infty$ with n_0 in $(-1, 0)$ as long as $x_0 < 0$. It can be shown that when $n_0 < 0$ the Bayes rule (for squared error loss) has infinite Bayes risk.

3. Characterization of conjugate priors—continuous case. This section is concerned with the converse to Theorem 2: can one conclude from the linearity at (2.10) that θ had a conjugate prior? The answer is yes, if it is assumed that the support of μ is sufficiently rich. The latter restriction will be clarified somewhat in Section 4.

In the statement of Theorem 3 and that of Theorem 4, below, there is an assumed form for a conditional expectation. Each univariate expectation can be interpreted in the following way (cf. Strauch ([1965]):

$$E(Y|Z) = g(Z) \quad \text{if and only if} \quad E(Y^+|Z) - E(Y^-|Z) = g(Z) \text{ a.s.} \tag{3.1}$$

With such an interpretation in mind, we are able to avoid the explicit assumption that means are finite when we postulate the form of a conditional expectation.

THEOREM 3. *Suppose Θ is open in R^d. Let X be a sample of size one from P_θ of* (2.1) *and suppose the support of μ contains an open interval I_0 in R^d. If θ has a prior distribution which does not concentrate at a single point, and if*

$$E(\nabla M(\theta)|X) = aX + b \tag{3.2}$$

for some constant a and constant vector b, then $a \neq 0$, τ is absolutely continuous $(d\theta)$ with $d\tau(\theta) = ce^{a^{-1}b\cdot\theta - a^{-1}(1-a)M(\theta)}d\theta$.

PROOF. From (3.1) and (3.2), $E(M_i^+|X) - E(M_i^-|X)$ is finite a.s. so $E(M_i^+|X)$ $E(M_i^-|X) < \infty$ for $i = 1, \ldots, d$, a.s. Since X has the positive density $f(x) = \int e^{x\theta - M(\theta)}d\tau(\theta)$ with respect to μ, $E(M_i^+|X) \cdot E(M_i^-|X) < \infty$ a.e. μ for each i. Observe that

$$E(M_i^{\pm}|x) = \int M_i^{\pm} e^{x\theta - M(\theta)}d\tau(\theta)/f(x) \tag{3.3}$$

with probability 1 in x for each i, and then that (3.3) holds a.e. μ for each i. Therefore, all integrals on the right side of (3.3) are finite a.e. μ and may be freely manipulated. From (3.2) one finds

$$\int \nabla M(\theta) e^{x\cdot\theta - M(\theta)}d\tau(\theta) = (ax + b)f(x) \quad \text{a.e.} \quad \mu. \tag{3.4}$$

Now if $a = 0$ in (3.4), $\int(\nabla M(\theta) - b)e^{x\cdot\theta - M(\theta)}d\tau(\theta)$ vanishes on an open interval I_0 of R^d. But then $\nabla M(\theta) - b$ must vanish on the support of τ and so is zero on at

This content downloaded from
128.148.194.11 on Tue, 06 Feb 2024 17:13:44 +00:00
All use subject to https://about.jstor.org/terms

least two points. Such a conclusion violates the strict convexity of M (from (2.2) and below).

In (3.4) replace x by $z = x + iy$ and observe that

$$Q(z) = \int(\nabla M(\theta) - az - b)e^{z\cdot\theta - M(\theta)}d\tau(\theta)$$

vanishes at least on Re $z \in I_0$. Then for a choice of $x_0 = \text{Re } z_0 \in I_0$, $Q(x_0 + iy)$ vanishes for all y and

$$\int\left(\frac{\nabla M(\theta) - ax_0 - b}{a}\right)e^{x_0\cdot\theta + iy\cdot\theta - M(\theta)}d\tau(\theta) = iy\int e^{x_0\cdot\theta + iy\cdot\theta - M(\theta)}d\tau(\theta). \tag{3.5}$$

In (3.5), write $m(\theta)' = (m_1(\theta), \ldots, m_d(\theta))' = (\nabla M(\theta) - ax_0 - b/a)$ and let $dF(\theta) = e^{x_0\cdot\theta - M(\theta)}d\tau(\theta)$. Then one has

$$\int e^{iy\cdot\theta}m(\theta)dF(\theta) = iy\int e^{iy\cdot\theta}dF(\theta). \tag{3.6}$$

The argument now proceeds from (3.6) along the lines of Lemma 6.1.1 of Kagan, Linnik and Rao [1973].

Begin with the first equation from (3.6). Multiply both sides of the equation by the factor

$$\left(\frac{1}{2\pi}\right)^d \prod_{k=1}^d \frac{1 - e^{-ih_ky_k}}{iy_k}\cdot\frac{e^{-i\alpha_ky_k} - e^{-i\beta_ky_k}}{iy_k},$$

with $\alpha_k < \beta_k$, $k = 1, \ldots, d$, and then integrate over $-T \leqslant y_k \leqslant T$, $k = 1, \ldots, d$. On the right hand side one finds

$$\begin{aligned}
&\int_{-T}^{T}\cdots\int_{-T}^{T}\left(\frac{1}{2\pi}\right)^d\prod_{k=1}^d\left(\frac{1 - e^{-ih_ky_k}}{iy_k}\right)\\
&\quad\cdot(e^{-i\alpha_1y_1} - e^{-i\beta_1y_1})\prod_{k=2}^d\int_{\alpha_k}^{\beta_k}e^{-iy_ku_k}du_k\int e^{iy\cdot\theta}dF(\theta)\\
&\quad= \int_{\alpha_2}^{\beta_2}\cdots\int_{\alpha_d}^{\beta_d}du_1, \ldots, du_d\int dF(\theta)\int_{-T}^{T}\cdots\int_{-T}^{T}\left(\frac{1}{2\pi}\right)^d\prod_{k=2}^d
\end{aligned}$$

$$\begin{aligned}
&\left(\frac{1 - e^{-ih_ky_k}}{iy_k}e^{i-y_k(u_k - \theta_k)}\right)\\
&\times\left(\frac{e^{-i\alpha_1y_1} - e^{-i(\alpha_1 + h_1)y_1}}{iy_1} - \frac{e^{-i\beta_1y_1} - e^{-i(\beta_1 + h_1)y_1}}{iy_1}\right)dy_1, \ldots, dy_d\\
&\to\int_{\alpha_2}^{\beta_2}\cdots\int_{\alpha_d}^{\beta_d}\{F[(\alpha_1, u_2, \ldots, u_d), (\alpha_1 + h_1, u_2 + h_2, \ldots, u_d + h_d)]\\
&- F[(\beta_1, u_2, \ldots, u_d), (\beta_1 + h_1, u_2 + h_2, \ldots, u_d + h_d)]\}du_2, \ldots, du_d
\end{aligned} \tag{3.7}$$

as $T \to \infty$, where $F[(\boldsymbol{\alpha}, \boldsymbol{\beta})]$ denotes the F measure of the d-dimensional interval $(\boldsymbol{\alpha}, \boldsymbol{\beta})$. Proceeding in the same way with the left hand side of (3.6), one has

$$\begin{aligned}
&\int_{-T}^{T}\cdots\int_{-T}^{T}\left(\frac{1}{2\pi}\right)^d\prod_{k=1}^d\left(\frac{1 - e^{-ih_ky_k}}{iy_k}\right)\prod_{h=1}^d\int_{\alpha_k}^{\beta_k}e^{-iu_ky_k}du_k\int e^{-y\cdot\theta}m_1(\theta)dF(\theta)\\
&\to\int_{\alpha_1}^{\beta_1}\cdots\int_{\alpha_d}^{\beta_d}F^{(1)}[(u_1, \ldots, u_d), (u_1 + h_1, \ldots, u_d + h_d)]du_1, \ldots, du_d
\end{aligned} \tag{3.8}$$

This content downloaded from
128.148.194.11 on Tue, 06 Feb 2024 17:13:44 +00:00
All use subject to https://about.jstor.org/terms

where $dF^{(1)}(\theta) = m_1(\theta)dF(\theta)$. In (3.7) and (3.8) let $h_k \to -\infty$, $k = 1, \ldots, d$, to obtain

$$\int_{\alpha_2}^{\beta_2} \cdots \int_{\alpha_d}^{\beta_d} [F(\alpha_1, u_2, \ldots, u_d) - F(\beta_1, u_2, \ldots, u_d)]du_2, \ldots, du_d = \int_{\alpha_1}^{\beta_1} \cdots \int_{\alpha_d}^{\beta_d} F^{(1)}(u_1, \ldots, u_d)du_1, \ldots, du_d \tag{3.9}$$

for all $\boldsymbol{\alpha}, \boldsymbol{\beta}$. Now for fixed α_1, β_1 it follows from (3.9) that

$$F(\alpha_1, u_2, \ldots, u_d) - F(\beta_1, u_2, \ldots, u_d) = \int_{\alpha_1}^{\beta_1} F^{(1)}(u_1, \ldots, u_d)du_1 \tag{3.10}$$

for almost all $u_2, \ldots, u_d$. Hence (3.10) holds simultaneously for all rational (α_1, β_1) except possibly for a set N of $(u_2, \ldots, u_d)$ of Lebesgue measure 0. Then aside from N, $F(u_1, \ldots, u_d)$ must be absolutely continuous in u_1. It will be argued that this conclusion holds for all $u_2, \ldots, u_d$, i.e., that $N = \varnothing$.

Note first that since F is nondecreasing, continuous in u_1 for almost all $u_2, \ldots, u_d$, it is, in fact, a continuous function of all d variables. Hence for each fixed α_1, β_1, the left side of (3.10) is continuous in $u_2, \ldots, u_d$. But the right side of (3.10) is

$$\int_{-\infty}^{u_2} \cdots \int_{-\infty}^{u_d} \int_{\alpha_1}^{\beta_1} \int_{-\infty}^{u_1} m_1(\theta)dF(\theta)$$

and so must also be continuous in $u_2, \ldots, u_d$ since F is continuous. Finally then, (3.10) holds for all $u_2, \ldots, u_d$ and, again from continuity, simultaneously for all α_1, β_1.

Given that $F(u_1, \ldots, u_d)$ is absolutely continuous in u_1 for every $u_2, \ldots, u_d$, Fubini's theorem insures that F is absolutely continuous with respect to d-dimensional Lebesgue measure and, from (3.10), $\partial F/\partial u_1 = -F^{(1)}(u_1, \ldots, u_d)$. From the remaining equations at (3.6) get the full relation:

$$\frac{\partial F}{\partial u_r} = -F^{(r)}(u_1, \ldots, u_d) = -\int_{-\infty}^{u_1} \cdots \int_{-\infty}^{u_d} m_r(\theta)dF(\theta), \qquad r = 1, \ldots, d. \tag{3.11}$$

Write $dF(\theta) = f(\theta)d\theta = (\partial^d/\Pi\partial\theta_i)F(\theta_1, \ldots, \theta_d)d\theta$ and use (3.11) to see that $\partial F/\partial u_r$ is also absolutely continuous with

$$\frac{\partial^d}{\Pi\partial u_i}\frac{\partial F}{\partial u_r} = \frac{\partial f}{\partial u_r} = -m_r f, \qquad r = 1, \ldots, d.$$

Now $m = a^{-1}(\nabla m - ax_0 - b)$ so

$$\nabla f = \left(ba^{-1} + x_0 - a^{-1}\nabla M\right)f \tag{3.12}$$

from which it follows that $f = c\exp(ba^{-1}\cdot\theta + x_0\cdot\theta - a^{-1}M(\theta))$. Recall that $dF(\theta) = e^{x_0\cdot\theta - M(\theta)}d\tau(\theta)$ and so find $d\tau(\theta) = c\exp(a^{-1}b\cdot\theta - ((1-a)/a)M(\theta))$, the desired conclusion.

REMARK. It is not generally possible to ask that (1.1) hold with $a = A$, a matrix, A not proportional to I. Here is a brief look at the situation for the present exponential family setting. Begin with the statement of Theorem 3 but with a

This content downloaded from 128.148.194.11 on Tue, 06 Feb 2024 17:13:44 +00:00
All use subject to https://about.jstor.org/terms

matrix A and a prior measure τ not supported on a $(d-1)$-dimensional hyperplane. Proceed through the proof of Theorem 3 to (3.4). At this point one can argue that $|A| \neq 0$ since, if it were, there would be some vector ξ with $\xi \cdot (\nabla M(\theta) - b) = 0$ on the support of τ, a contradiction. Continue through the proof to the conclusion

$$\nabla f = -A^{-1}(\nabla M - Ax_0 - b)f \tag{3.13}$$

where f is the density of the measure $e^{x_0 \cdot \theta - M(\theta)} d\tau(\theta)$. Now it is not generally true that (3.13) has a solution f, but then (3.2) could not have applied. For an example, take A to be a diagonal matrix with entries along the diagonal all distinct. It can be easily argued that (3.13) can be solved for such an A only if $M(\theta)$ has the special form $\sum_{i=1}^{d} \mu_i(\theta_i)$. Other types of matrices lead to similar, though less agreeable, conditions on M. When (3.13) does allow a solution, the prior measures τ which result are what Jewell [1974b] (in the context of the multivariate normal distribution) refers to as enriched priors.

4. Characterization of conjugate priors—discrete case. The converse of Theorem 2 is less complete when the support of μ does not contain an interval. Consider for instance the problem of estimating a binomial parameter p from a sample of size n. If π is a prior measure of the Borel sets of [0, 1], the conditions of posterior linearity become:

$$\int_0^1 p^{k+1}(1-p)^{n-k} d\pi(p) = (ak + b) \int_0^1 p^k (1-p)^{n-k} d\pi(p) \tag{4.1}$$

$$\text{for} \quad k = 0, 1, 2, \ldots, n.$$

These are merely restrictions on the first $n + 1$ moments of the measure π and any π which has the same first $n + 1$ moments as a beta measure will satisfy (4.1). In this section we give theorems characterizing the conjugate priors of all commonly occurring exponential families on the nonnegative integers. Theorem 4 specializes to a characterization of the beta distribution as the unique family allowing linear posterior expectation for negative binomial random variables for example. The case of Poisson variables is not covered by Theorem 4, but this has already been treated by Johnson [1957, 1967]. Theorem 5 then deals with the assumptions needed to characterize the binomial distribution.

Suppose X is a sample of size 1 from P_θ of (2.1) and let the support of μ be the nonnegative integers. For this setting, Θ is always an interval which is unbounded to the left. Our regularity assumption would have Θ be an open interval, and for Theorem 4 it will be assumed that $\Theta = (-\infty, \theta_0)$ with $\theta_0 < \infty$, i.e., that μ does not have a moment generating function on all of R. Under this setup, we have

THEOREM 4. *Suppose θ has a prior distribution τ on $\Theta = (-\infty, \theta_0)$ with $\theta_0 < \infty$, and assume τ is not concentrated on a single point. If*

$$E(M'(\theta) | X = x) = ax + b \qquad \text{for} \quad x = 0, 1, \cdots, \tag{4.2}$$

then $a > 0$, τ is absolutely continuous with respect to Lebesgue measure, and $d\tau(\theta) = ce^{a^{-1}b\theta - a^{-1}(1-a)M(\theta)} d\theta$.

This content downloaded from
128.148.194.11 on Tue, 06 Feb 2024 17:13:44 +00:00
All use subject to https://about.jstor.org/terms

PROOF. One may proceed as in the proof of Theorem 3 to the equation (3.4). If a were zero, (3.4) would give

$$\int_{-\infty}^{\theta_0}(M'(\theta) - b)e^{x\theta - M(\theta)}d\tau(\theta) = 0, \qquad x = 0, 1, \cdots. \tag{4.3}$$

Make the change of variables $t = e^{\theta}$ in (4.3) and so produce a signed measure on $[0, e^{\theta_0}]$ having all moments zero. This signed measure must in fact be the zero measure, since the moment problem is determined on a compact interval. This implies $M'(\theta) - b$ is zero on the support of τ, a contradiction. So $a \neq 0$ and (4.2) can be written as

$$\int_{-\infty}^{\theta_0}e^{x\theta}(M'(\theta) - b)e^{-M(\theta)}d\tau(\theta) = ax\int_{-\infty}^{\theta_0}e^{x\theta}e^{-M(\theta)}d\tau(\theta), \qquad x = 0, 1, \cdots. \tag{4.4}$$

Transform the left side of (4.4) as follows:

$$\begin{aligned}&\int_{-\infty}^{\theta_0}\Big[\int_{-\infty}^{\theta}xe^{xy}dy\Big](M'(\theta) - b)e^{-M(\theta)}d\tau(\theta)\\ &= \int_{-\infty}^{\theta_0}xe^{xy}\Big[\int_{y}^{\theta_0}(M'(\theta) - b)e^{-M(\theta)}d\tau(\theta)\Big]dy\\ &= -\int_{-\infty}^{\theta_0}xe^{x\theta}\Big[\int_{-\infty}^{\theta}(M'(y) - b)e^{-M(y)}d\tau(y)\Big]d\theta.\end{aligned} \tag{4.5}$$

In (4.5) the interchange of integrations can be easily justified and (4.4) has been invoked with $x = 0$ to produce the final equality. Replacing the left side of (4.4) by the right side of (4.5) one has

$$\int_{-\infty}^{\theta_0}e^{x\theta}\Big[-\int_{-\infty}^{\theta}(M'(y) - b)e^{-M(y)}d\tau(y)\Big]d\theta = a\int_{-\infty}^{\theta_0}e^{x\theta}e^{-M(\theta)}d\tau(\theta) \quad \text{for} \quad x = 1, 2, \cdots. \tag{4.6}$$

Make again the change of variable $t = e^{\theta}$ in (4.6) to produce a signed measure on $[0, e^{\theta_0}]$ all of whose moments are zero except possibly the zeroth. Such a signed measure must concentrate on the origin and, in the present circumstance, puts no weight there. But then

$$ae^{-M(\theta)}d\tau(\theta) = -\Big[\int_{-\infty}^{\theta}(M'(y) - b)e^{-M(y)}d\tau(y)\Big]d\theta. \tag{4.7}$$

From (4.7), τ is absolutely continuous with a density f which satisfies the differential equation

$$af'(\theta) - aM'(\theta)f(\theta) = -(M'(\theta) - b)f(\theta). \tag{4.8}$$

The conclusion follows easily from (4.8).

As the discussion at the beginning of the section indicates, a different formulation is required if the beta distribution is to be characterized for binomial observations. This is accomplished in Theorem 5. For simplicity, in the statement and proof of this theorem we use the notation of the mean parameter as opposed to the natural parameter.

THEOREM 5. *Let τ be a prior distribution for p on the Borel sets of* $[0, 1]$ *and assume τ does not concentrate on a single point. If for each $n = 1, 2, \ldots,$ there are*

This content downloaded from
128.148.194.11 on Tue, 06 Feb 2024 17:13:44 +00:00
All use subject to https://about.jstor.org/terms

numbers a_n and b_n for which

$$\int_0^1 p^{k+1}(1-p)^{n-k}d\tau(p) = (a_n k + b_n)\int_0^1 p^k(1-p)^{n-k}d\tau(p) \tag{4.9}$$

for $k = 0, 1, \ldots, n$, then

$$a_n = \frac{a}{1 + a(n-1)}, \qquad b_n = \frac{b}{1 + a(n-1)} \tag{4.10}$$

with $a > 0$, $b > 0$, $a + b < 1$ and τ is a beta distribution.

PROOF. Ericson's [1969] result, in conjunction with the linearity of (4.9), implies that

$$E(p|S_n = k) = \frac{k\,\mathrm{Var}(p) + E(p)E(p(1-p))}{n\,\mathrm{Var}(p) + E(p(1-p))}.$$

This yields (4.10) with $a = (\mathrm{Var}(p)/E(pE(1-p)))$ and $b = (E(p(1-p))/E(1-p))$. Take $n = 1$ in (4.9) and sum over $k = 0, 1$ to obtain $\int p d\tau(p) = b + a\int p d\tau(p)$, so $\int p d\tau(p) = (b/(1-a))$. Now take $k = n$ in (4.9) to write

$$\begin{aligned}\int_0^1 p^{n+1}d\tau(p) &= \frac{an + b}{1 + a(n-1)}\int_0^1 p^n d\tau(p) \\ &= \cdots = \left\{\prod_{j=1}^n \frac{aj + b}{1 + a(j-1)}\right\}\frac{b}{1-a}.\end{aligned} \tag{4.11}$$

In this way it is clear that all moments of τ are determined by a and b, hence τ is also. Moreover, (4.9) can be achieved by using the beta density with $\alpha = b/a$ and $\beta = (1 - (a + b)/a)$.

The proof of Theorem 5 generalizes in a straightforward way to yield a characterization of the Dirichlet family as the unique family allowing linear posterior expectation for multinomial observations.

Acknowledgment. We thank Bradley Efron, Joseph Keller, Carl Morris and David Siegmund for helpful remarks as this work progressed.

REFERENCES

BARNDORFF-NIELSEN, O. (1970). *Exponential Families-Exact Theory*. Various Publication Series No. 19. Aarhus Universitet, Aarhus.

COHEN, A. (1966). All admissible linear estimates of the mean vector. *Ann. Math. Statist.* **37** 458–463.

EFRON, B. and MORRIS, C. (1973). Stein's estimation rule and its competetitors—an empirical Bayes approach. *J. Amer. Statist. Assoc.* **68** 117–130.

ERICSON, W. A. (1969). A note on the posterior mean of a population mean. *J. Roy. Statist. Soc. Series B* **31** 332–334.

ERICSON, W. A. (1970). On the posterior mean and variance of a population mean. *J. Amer. Statist. Assoc.* **65** 649–652.

GOLDSTEIN, M. (1975). Uniqueness relations for linear posterior expectations. *J. Roy. Statist. Soc. Ser B* **37** 402–405.

HARTIGAN, J. A. (1969) Linear Bayesian methods. *J. Roy. Statist. Soc. Series B* **31** 446–454.

JEWELL, W. S. (1974a). Credible means are exact Bayesian for exponential families. *Astin Bull.* **8** 77–90.

JEWELL, W. S. (1974b). Exact multidimensional credibility. *Mitt. der Verein. Schweiz. Versich.-Math.* **74** 193–214.

This content downloaded from
128.148.194.11 on Tue, 06 Feb 2024 17:13:44 +00:00
All use subject to https://about.jstor.org/terms

JEWELL, W. S. (1975). Regularity conditions for exact credibility. *Astin Bull.* **8** 336–341.
JOHNSON, W. L. (1957). Uniqueness of a result in the theory of accident proneness. *Biometrika* **44** 430–531.
JOHNSON, W. L. (1967). Note on a uniqueness relation in certain accident proneness models. *J. Amer. Statist. Assoc.* **62** 288–289.
KAGAN, A. M., LINNICK, YU. V. and RAO, C. R. (1973). *Characterization Problems in Mathematical Statistics*. Wiley, New York.
KAHN, P. M. (1975). *Credibility Theory and Application*. Academic Press, New York.
KARLIN, S. (1958). Admissibility for estimation with quadratic loss. *Ann. Math. Statist.* **29** 406–436.
LINDLEY, D. V. (1972). *Bayesian Statistics, A Review*. SIAM, Philadelphia.
NOVICK, M. R. and HALL, W. J. (1965). A Bayesian indifference procedure. *J. Amer. Statist. Assoc.* **60** 1104–1117.
PING, C. (1964). Minimax estimates of parameters of distributions belonging to the exponential family. *Chinese Math.—Acta* **5** 277–299.
RAIFFA, H. and SCHLAIFER, R. (1961). *Applied Statistical Decision Theory*. Graduate School of Business Administration, Harvard Univ., Boston.
STONE, M. (1967). Generalized Bayes decision functions, admissibility, and the exponential family. *Ann. Math. Statist.* **38** 818–822.
STRAUCH, R. E. (1965). Conditional expectations of random variables without expectations. *Ann. Math. Statist.* **36** 1556–1559.

BELL LABORATORIES
600 MOUNTAIN AVENUE
MURRAY HILL, NEW JERSEY 07974

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CALIFORNIA
LOS ANGELES, CALIFORNIA 90024

This content downloaded from
128.148.194.11 on Tue, 06 Feb 2024 17:13:44 +00:00
All use subject to https://about.jstor.org/terms