The Annals of Statistics
1983, Vol. 11, No. 2, 515–529

NATURAL EXPONENTIAL FAMILIES WITH QUADRATIC VARIANCE FUNCTIONS: STATISTICAL THEORY[1]

BY CARL N. MORRIS

University of Texas

The normal, Poisson, gamma, binomial, negative binomial, and NEF-GHS distributions are the six univariate natural exponential families (NEF) with quadratic variance functions (QVF). This sequel to Morris (1982) treats certain statistical topics that can be handled within this unified NEF-QVF formulation, including unbiased estimation, Bhattacharyya and Cramer-Rao lower bounds, conditional distributions and moments, quadratic regression, conjugate prior distributions, moments of conjugate priors and posterior distributions, empirical Bayes and G_2 minimax, marginal distributions and their moments, parametric empirical Bayes, and characterizations.

1. Introduction. Certain probabilistic properties of univariate *natural exponential families* (NEF) having *quadratic variance functions* (QVF) were developed in Morris (1982). There are six basic NEF-QVF distributions: normal, Poisson, gamma, binomial, negative binomial, and the NEF generated by the generalized hyperbolic secant (GHS) distributions. Affine transformations of these basic distributions generate all other NEF-QVFs. In addition to introducing and characterizing all NEF-QVFs, these families are studied in a unified way using the quadratic nature of their *variance functions* (VF) in Morris (1982) with respect to their infinite divisibility, moment and cumulant properties, large deviation behavior, limits in distribution, and their systems of orthogonal polynomials.

This sequel is concerned with NEF-QVF results of a more statistical nature. Section 2 starts by summarizing some basic NEF-QVF properties, with others introduced as needed later. Section 3 treats unbiased estimation of arbitrary analytic functions, including moments and cumulants. Cramer-Rao and Bhattacharyya lower bounds for unbiased estimators arise naturally and easily in this theory, which is based on the NEF-QVF orthogonal polynomials.

Conditional distributions and quadratic regression, i.e. distributions and moments of X_1 given $Y = X_1 + X_2$ with X_1, X_2 independent NEF-QVF distributions are the subject of Section 4. The rth conditional moment is shown to be a polynomial of degree r in Y if and only if the NEF has QVF.

Bayesian analysis with conjugate prior distributions is studied in Section 5. The conjugate prior distributions are those whose posterior means are linear in the natural observation. The conjugate prior distribution on the mean of a NEF is a Pearson family if the NEF has a QVF. The moments of the conjugate prior distribution and of the posterior distribution of μ are easily expressed in terms of the variance function for X. Finally, the conjugate prior distribution is the minimax choice for NEF-QVF distributions among prior distributions with specified mean and variance. That is, the statistician who knows only the first two prior moments can safely use the conjugate prior distribution.

Received September 1982; revised November 1982.

[1] Support for this work provided by NSF Grant No. MCS-8104250, Department of Mathematics, University of Texas, Austin, Texas 78712.

AMS 1970 *subject classifications.* Primary 60E05, 60E07, 60F10, 62E15, 62E30

Key words and phrases. Exponential families, natural exponential families, quadratic variance function, normal distribution, Poisson distribution, gamma distribution, binomial distribution, negative binomial distribution, NEF-GHS distribution, unbiased estimation, Bhattacharyya bounds, quadratic regression, conjugate priors, Bayesian analysis, posterior moments, G_2 minimax, parametric empirical Bayes, and characterizations.

Section 6 briefly reviews the marginal distributions of X and their moments when X has a NEF-QVF distribution, given μ, and μ has a conjugate prior distribution. Formulas for estimating the moments of the prior distribution from X are developed based on the NEF-QVF orthogonal polynomials, a useful result in random effects models and parametric empirical Bayes inference. Distributions arising in Sections 4–6, as conditional distributions, as conjugate prior distributions, and as marginal distributions, include the normal, gamma, beta, F, reciprocal gamma, t (as conjugate prior for the NEF-GHS), binomial, hypergeometric, negative hypergeometric, and negative binomial. Other named exponential families arise as nonlinear transformations of natural exponential families with quadratic variance functions, including lognormal, Weibull, extreme value, Pareto, and Cauchy distributions. The exponential, chi squared, Rayleigh, Bernoulli, and geometric distributions are special cases of the NEF-QVF family. Thus, most well-known univariate distributions are related to the NEF-QVF distributions, as summarized in Table 1.

A form for the Bayes rule is provided in Section 7 for Bayesian situations. This immediately suggests parametric empirical Bayes estimators, generalizing the James-Stein (1961) estimator to all NEF-QVF distributions.

In writing this paper, only results that can be proved in general for NEF-QVF distributions are presented. Many old results are given with new proofs, and some new ones are included. This paper also assembles known results from scattered sources and incorporates them here in the NEF-QVF framework.

2. A review of natural exponential families with quadratic variance functions. A parametric family of distributions with *natural parameter space* $\Theta \subset R$ (the real line) is a *natural exponential family* (NEF) if random variables X governed by these distributions satisfy

$$P_\theta(X \in A) = \int_A \exp\{x\theta - \psi(\theta)\}\, dF(x), \tag{2.1}$$

with F a Stieltjes measure on R not depending on $\theta \in \Theta$, the *natural parameter*, and sets $A \subset R$. The cumulant generating function $\psi(\theta)$ gives (2.1) unit probability. The random variable X is the *natural observation*. Exponential families that are not NEFs are nonlinear transformations $Y = K(X)$ of NEFs.

The natural observation X has mean and variance

$$\psi'(\theta) = \mu = E_\theta X = \int x\, dF_\theta(x) \tag{2.2}$$

$$\psi''(\theta) = V(\mu) = \mathrm{Var}_\theta(X) = \int (x-\mu)^2\, dF_\theta(x) \tag{2.3}$$

and cumulants $C_r(\mu) = \psi^{(r)}(\theta)$, $r = 1, 2, \cdots$. The function $V(\mu)$ in (2.3) on its domain $\Omega \equiv \psi'(\Theta)$ is called the *variance function* (VF) of the NEF and characterizes the NEF (but no particular member of the NEF).

In Morris (1982) it is shown that exactly six basic types of NEFs have *quadratic variance functions* (QVF)

$$V(\mu) = v_0 + v_1\mu + v_2\mu^2. \tag{2.4}$$

These natural exponential families with quadratic variance functions (NEF-QVF) are summarized in Table 1 here, and Table 1 of Morris (1982), as (a) the normal, $N(\mu, \sigma^2)$ with $V(\mu) = \sigma^2$ (constant variance function); (b) the Poisson Poiss(λ) with $\mu = \lambda$, $V(\mu) = \mu$ (linear variance function); (c) the gamma, Gam(r, λ), $\mu = r\lambda$, $V(\mu) = r\lambda^2 = \mu^2/r$; (d) the binomial, Bin$(r, p)$, $\mu = rp$, $V(\mu) = rpq = -\mu^2/r + \mu$, $(q \equiv 1 - p)$; (e) the negative binomial, NB(r, p), $0 \le p \le 1$, $\mu = rp/q$, $V(\mu) = rp/q^2 = \mu^2/r + \mu$, $(q \equiv 1 - p)$; and (f) the NEF generated by the generalized hyperbolic secant (GHS) distribution, NEF-GHS, with $V(\mu) = \mu^2/r + r$, $r > 0$. The NEF-GHS is a family of continuous distributions with support $-\infty < x < \infty$ and $\psi(\theta) = -\log\cos(\theta)$.

The six basic types of distributions can be extended by convolutions and location and scale changes all of which preserve both the NEF and the QVF properties, as follows. Let $X_1, \cdots, X_n$ be iid (independent, identically distributed) as a NEF-QVF distribution. Then $X^* \equiv \sum(X_i - b)/c$ has a NEF-QVF distribution with mean $\mu^* = n(\mu - b)/c$ and variance function $\mathrm{Var}(X^*) = V^*(\mu^*) \equiv v_0^* + v_1^*\mu^* + v_2^*(\mu^*)^2$ with

$$v_0^* = nV(b)/c^2, \qquad v_1^* = V'(b)/c, \qquad v_2^* = v_2/n. \tag{2.5}$$

The discriminant of $V(\mu)$ is

$$d \equiv v_1^2 - 4v_0v_2. \tag{2.6}$$

Then d^*, the discriminant of V^*, *is* $d^* = d/c^2$ which is unchanged by convolution and translation. Formula (2.5) also holds for all real $n > 0$ when the NEF-QVF distribution is infinitely divisible, i.e. for all but the binomial cases.

Each of the six NEF-QVF families has up to four parameters, being the location (b), scale (c), convolution (n) (including division), and exponential (μ) parameters. The normal family has but two parameters because the exponential and convolution parameters also serve as the location and scale parameters. The Poisson family has three parameters because μ also is the convolution parameter. The gamma family has three parameters because μ also is the scale parameter. Affine transformations of the usual binomial, negative binomial, and NEF-GHS families of distributions are properly four parameter families.

3. Unbiased estimation theory. Every analytic function $g(\mu)$ has a unique uniform minimum variance unbiased estimator (UMVUE) $\hat{g}(X)$ if $X \sim$ NEF-QVF and if (3.7) is finite, Seth (1949), and Abbey and David (1970), except in the binomial (n, p) case when this holds only if $g(p)$ is a polynomial of degree not exceeding n. Let $f(x, \theta) = \exp(x\theta - \psi(\theta))$ be the NEF-QVF density and define

$$P_m(x, \mu) = V^m(\mu)\left\{\frac{d^m f(x, \theta)}{d\mu^m}\right\} \Big/ f(x, \theta) \tag{3.1}$$

for $m = 0, 1, 2, \cdots$ (in the binomial (n, p) case, $P_m = 0$ for $m > n$). *Define* $a_0 = b_0 = 1$ and a_m, b_m for $m \geq 1$ by

$$a_m = m! \prod_0^{m-1} (1 + iv_2) \equiv m!\, b_m. \tag{3.2}$$

Then from Morris (1982a),

$$\{P_0, (x, \mu) = 1, P_1(x, \mu) = x - \mu, \cdots\}$$

is a complete set of orthogonal polynomials, P_m of degree m in both x and μ. We have

$$E_\mu P_m(X, \mu)P_n(X, \mu) = \delta_{mn} a_m V^m(\mu) \tag{3.3}$$

with δ the Kronecker delta and

$$E_\mu P_m(X, \mu_0) = b_m(\mu - \mu_0)^m. \tag{3.4}$$

Formula (3.3) shows the orthogonality of the polynomials, that all but P_0 have mean 0, and their variances are $a_m V^m(\mu)$. Formula (3.4) yields the expectation of P_m under the "wrong" distribution and is the basis for constructing unbiased estimators, as follows. More about these polynomials is available in Morris (1982, Section 8).

THEOREM 3.1. *Let $g(\mu)$ be an analytic function of $\mu \in \Omega$ and choose μ_0 in the interior* *of Ω so that*

$$g(\mu) = \textstyle\sum_0^\infty c_i(\mu - \mu_0)^i/i! \tag{3.5}$$

with $c_i \equiv g^{(i)}(\mu_0)$ the ith derivative at μ_0. Assume (3.7) *is finite. Then the unique unbiased estimator of $g(\mu)$ is*

$$\hat{g}(X) = \textstyle\sum_0^\infty c_i P_i(X, \mu_0)/a_i. \tag{3.6}$$

(Different choices of $\mu_0 \in \Omega$ lead to different representations in (3.6) *but always to the same $\hat{g}$). Also,*

$$\operatorname{Var}_\mu\{\hat{g}(X)\} = \{g'(\mu)\}^2 V(\mu) + \{g''(\mu)\}^2 V^2(\mu)/\{2(1+v_2)\} + \textstyle\sum_3^\infty \{g^{(i)}(\mu)\}^2 V^i(\mu)/a_i. \tag{3.7}$$

PROOF. (3.6) is unbiased for (3.5) because of (3.4) and the definition (3.2) of b_i. Uniqueness follows because the exponential families are complete. The variance of (3.6) when $\mu = \mu_0$ is, from (3.3),

$$\operatorname{Var}_{\mu_0}\{\hat{g}(X)\} = \textstyle\sum_1^\infty c_i^2 V^i(\mu_0)/a_i$$

and since this holds for all μ_0 we have (3.7). □

The first term on the rhs (right hand side) of (3.7) is the Cramer-Rao lower bound for the variance of an unbiased estimator. The sum of the first k terms of the rhs of (3.7) is the kth Bhattacharyya bound, also a lower bound for the variance of an unbiased estimator. Clearly, the kth bound is attained in NEF-QVF distributions if and only if $g(\mu)$ is a polynomial of degree at most k, if and only if the UMVUE $\hat{g}(x)$ is a polynomial of degree k. Fend (1959) and Rao (1952) showed this last result. Formulas (3.1), (3.6), (3.7) are not new. Seth (1949) used this system of polynomials to get $\hat{g}(X)$ in (3.6) from $g(\mu)$. Abbey and David (1970) and Blight and Rao (1974) developed the expression (3.7) for the variance and the Bhattacharyya bounds for these families, and Shanbag (1972) also characterized the Bhattacharyya bounds in NEF-QVF cases. Patil and Shorrock (1965) prove that for general exponential families the first two Bhattacharyya bounds are identical if and only if $g(\mu)$ is linear. For NEF-QVF families this follows easily from (3.7) because $g''(\mu) = 0$ is equivalent to $g(\mu)$ being linear. Guoying and Ping (1981) have made recent contributions to existence of the UMVUE and the Cramer-Rao and Bhattacharyya bounds for NEF-QVF distributions using the polynomial system (3.1).

Certain unbiased estimators are worth noting especially. Let $X \sim$ NEF-QVF$(\mu, V(\mu))$. Theorem 3.2 deals with unbiased estimation of the variance function $V(\mu)$.

THEOREM 3.2. *Define $V^*(X) \equiv V(X)/b_2$, and let $c = 2v_2(2+3v_2)/b_2$, and note $b_2^2/b_4 = 1/(1+c)$. Recall $d = v_1^2 - 4v_0v_2$. Then $V^*(X)$ is the* UMVUE *of $V(\mu)$,*

$$E_\mu V^*(X) = V(\mu) \quad \textit{for all } \mu. \tag{3.8}$$

$$\operatorname{Var}_\mu(V^*(X)) = cV^2(\mu) + dV(\mu). \tag{3.9}$$

An unbiased estimate of (3.9) *is given inside the brackets of* (3.10),

$$E_\mu\left\{\frac{c}{1+c} V^{*2}(X) + \frac{1}{1+c} dV^*(X)\right\} = \operatorname{Var}_\mu\{V^*(X)\}. \tag{3.10}$$

PROOF. We have $EV(X) = V(\mu) + v_2\operatorname{Var}(X) = (1+v_2)V(\mu)$, proving (3.8). The variance of the quadratic function $V^*(X)$ is then given by the first two terms of (3.7) with $g(\mu) = V(\mu)$, and so equals $(V')^2 V + (2v_2)^2 V^2/2b_2$, suppressing the argument μ in V and V'. Since $(V')^2 = 4v_2 V + d$, (3.9) follows. It follows easily from (3.8) and (3.9) that

$$E_\mu\{V^{*2}(X) - dV^*(X)\}/(1+c) = V^2(\mu), \tag{3.11}$$

and then (3.10) follows by using (3.8), (3.11) to determine the unique unbiased estimator of (3.9). □

The third and fourth cumulants, polynomials of degree 3 and 4, $C_3(\mu) = V'(\mu)V(\mu)$ and $C_4(\mu) = 6v_2V^2(\mu) + dV(\mu)$ in NEF-QVF distributions also have UMVUEs, easily determined by the relations

$$EC_3(X) = b_3C_3(\mu) \tag{3.12}$$

$$EC_4(X) = b_4C_4(\mu) + dv_2b_2V(\mu). \tag{3.13}$$

These can be proven directly by expanding $C_r(X)$ around $X = \mu$ in a Taylor series of order r. A quicker proof for (3.13) uses (3.11) and (3.8) to construct the UMVUE for the rhs of (3.13), with $C_4(\mu) = 6v_2V^2(\mu) + dV(\mu)$, and then shows this UMVUE simplifies to $C_4(X)$.

Central moments $M_r(\mu) = E(X - \mu)^r$ also have UMVUEs. That already is established in (3.8) and (3.12) for $M_2(\mu) = V(\mu)$, $M_3(\mu) = C_3(\mu)$, but we also have

$$E\left[\left(\frac{1 + 2v_2}{b_4}\right)\{3V^2(X) - 2dV(X)\}\right] = M_4(\mu). \tag{3.14}$$

This expression follows from writing $M_4(\mu) = (3 + 6v_2)V^2(\mu) + dV(\mu)$, Morris (1982, Section 7) and using (3.11) and (3.8).

As an application of these ideas, let $\hat{p} = X \sim (1/n)\mathrm{Bin}(n, p)$ with $n \geq 4$, i.e. $\hat{p} \sim \mathrm{NEF}(p, V(p) = pq/n)$, $q = 1 - p$. We have $\mu = p$, $v_2 = -1/n$, $d = 1/n^2$, $c = -(4 - 6/n)/(n - 1)$ and $1/(1 + c) = b_2^2/b_4 = n(n - 1)/(n - 2)(n - 3)$. Then the UMVUE of pq/n is $V^*(\hat{p}) = \hat{p}\hat{q}/(n - 1)$, with variance $cp^2q^2/n^2 + pq/n^3$ from (3.9) which has UMVUE given by (3.10),

$$E_\mu\left[\frac{\hat{p}\hat{q}}{n(n - 2)(n - 3)}\left\{1 - 4\hat{p}\hat{q} - \frac{2n - 4}{(n - 1)^2}\hat{p}\hat{q}\right\}\right] = \mathrm{Var}_\mu\{V^*(\hat{p})\}. \tag{3.15}$$

Many other examples of unbiased estimators are possible of course. For example $g(\mu) = (\mu - \mu_0)^m$ has UMVUE $\hat{g}(X) = P_m(X, \mu_0)/b_m$ from (3.6). Then using (3.7) or Morris (1982, formula (8.7)), this estimator has variance

$$\mathrm{Var}_\mu\hat{g}(X) = m!\ V^m(\mu) \textstyle\sum_{i=0}^{m-1} \binom{m}{i} \dfrac{\delta^i}{i!\, b_{m-i}} \tag{3.16}$$

with $\delta \equiv (\mu - \mu_0)^2/V(\mu)$.

4. Conditional distributions. Let X_1 and X_2 have independent NEF-QVF distributions, X_i having density

$$\exp\{\theta x_i - \nu_i\psi(\theta)\} \tag{4.1}$$

so that $X_i \sim \mathrm{NEF\text{-}QVF}(\nu_i\mu, \nu_iV(\mu))$ where $\mu = \psi'(\theta)$, $V(\mu) = \psi''(\theta)$, and $\nu_i > 0$ is known. This is only a notational generalization of the earlier model when $\nu_i \neq 1$. We call ν_i the "convolution parameter" because when ν_i is an integer in (4.1) X_i has the density of the convolution of ν_i NEF-QVF(μ, $V(\mu)$) variates. More generally (4.1) makes sense for any positive ν_i if the NEF is infinitely divisible (among QVF distributions, only the binomial is not infinitely divisible). We must remember that $\psi_i(\theta) = \nu_i\psi(\theta)$, $\mu_i = \nu_i\mu$ and $V_i(\mu_i) = \nu_iV(\mu_i/\nu_i)$ are the CGF, mean, and VF in the previous notation, requiring slight adjustments when applying earlier NEF results to (4.1).

Let X_1, X_2 be independent, X_i having density (4.1), and define in this section

$$Y \equiv X_1 + X_2, \qquad \nu \equiv \nu_1 + \nu_2, \qquad \pi_i \equiv \nu_i/\nu, \tag{4.2}$$

with Y a complete sufficient statistic for μ, $Y \sim \mathrm{NEF}(\nu\mu, \nu V(\mu))$ also having density of the form (4.1). For all NEFs, the conditional distribution of X_1 given Y is independent of the parameter μ, and

$$EX_1 \mid Y = \pi_1 Y. \tag{4.3}$$

This is most easily proved by appeal to the following lemma, which will be used repeatedly in this section.

LEMMA 4.1. *Let S be a complete sufficient statistic* (css) *for a parameter μ. Let T be a statistic with mean $ET = g(\mu)$. Then $ET \mid S = \hat{g}(S)$ is that unique function of S which is unbiased for $g(\mu)$, $E\hat{g}(S) = g(\mu)$ (given by* (3.6) *for NEF-QVF distributors).*

PROOF OF LEMMA. $ET \mid S$ is an unbiased estimate of $g(\mu)$, depending only on S. By completeness, it must equal $\hat{g}(S)$. □

Formula (4.3) now follows from Lemma 4.1 because both sides of the equation have the same expectation $\nu_1 V(\mu)$, and Y is a css for μ.

If the NEF also has QVF, $\operatorname{Var}(X_i) = \nu_i V(\mu)$, we have

$$\operatorname{Var}(X_1 \mid Y) = \frac{\nu_1 \nu_2}{\nu + v_2} V\left(\frac{Y}{\nu}\right). \tag{4.4}$$

This follows by showing when V is quadratic that $EX_1^2 \mid Y = (\pi_1 Y)^2 + (\nu_1\nu_2/(\nu + v_2)) \cdot V(Y/\nu)$, which is done by checking that both sides have the same expectation $EX_1^2 = \nu_1 V(\mu) + \nu_1^2 \mu^2$, and then applying Lemma 4.1.

Moreover, if we have $\operatorname{Var}(X_1 \mid Y) = Q(Y)$ with quadratic $Q(y) = q_2 y^2 + q_1 y + q_0$, then $\operatorname{Var}(X_1) = \nu_1 V(\mu)$ also must equal $EQ(Y) + \operatorname{Var}(\pi_1 Y) = Q(\nu\mu) + q_2 \nu V(\mu) + \pi_1^2 \nu V(\mu)$, making $V(\mu)$ quadratic. Thus, a necessary and sufficient condition within NEF distributions that the conditional variance (4.4) be quadratic in Y is that X_i have one of the six NEF-QVF distributions. Bolger and Harkness (1965) proved this result without the NEF assumption, in which case a seventh distribution, the Cauchy, also was shown to have a similar property.

Tweedie (1946) characterized those distributions for which $ES^2 \mid \bar{X}$ is quadratic in $\bar{X}$, S^2 the sample variance $\sum_1^n (X_i - \bar{X})^2/(n-1)$ of n iid observations $X_1, \cdots, X_n$, as being the six NEF-QVF families or the Cauchy family.

In Tweedie's case, iid NEF-QVF(μ, $V(\mu)$) distributions for X_i yield

$$ES^2 \mid \bar{X} = nV(\bar{X})/(n + v_2), \tag{4.5}$$

which is proved by checking that the rhs of (4.5) has mean $V(\mu) = ES^2$ and using Lemma 4.1. On the other hand, if $ES^2 \mid \bar{X}$ is quadratic in $\bar{X}$ then it follows that $\operatorname{Var}(X_i \mid \bar{X})$ also is quadratic in $\bar{X}$ and, by the remarks of the paragraph following (4.4), that the NEF is one of the six QVF families. This proves Tweedie's result.

Note that $ES^2 \mid \bar{X}$ always is a non-constant function of $\bar{X}$ for NEFs unless the $\{X_i\}$ are normal. Since every distribution with a MGF belongs to a NEF, we have another proof that S^2 is statistically independent of $\bar{X}$ only for the normal distribution.

Laha and Lukacs (1960) showed Tweedie's result holds when S^2 is replaced by a more general quadratic function. Their result is extended to monomials of degree $r \geq 2$ as follows.

THEOREM 4.1. *Let X_1, X_2 have independent* NEF *densities* (4.1) *with $\nu_1, \nu_2 > 0$. Then $V(\mu)$ is quadratic if and only if $E(X_1^r \mid Y)$ is a polynomial of degree r in $Y = X_1 + X_2$ for all $r = 1, 2, 3, \cdots$.*

PROOF. If $V(\mu)$ is a QVF then for any $r = 1, 2, \cdots$ EX_1^r is a polynomial of degree r in μ, c.f. Morris (1982, Section 7). The unbiased estimator $H_r(Y)$ of EX_1^r is a polynomial of degree r in Y, given by Theorem 3.1. Now, completeness of Y guarantees that $H_r(Y) = EX_1^r \mid Y$.

To prove the converse, assume $Q(Y) = EX_1^2 \mid Y$ is quadratic with leading coefficient q_2. Then $EX_1^2 = \nu_1^2\mu^2 + \nu_1 V(\mu)$ also equals $EQ(Y) = Q(\nu\mu) + q_2\nu V(\mu)$, or equivalently:

$$(\nu_1 - q_2\nu)V(\mu) = Q(\nu\mu) - \nu_1^2\mu^2.$$

Thus, either $V(\mu)$ is quadratic or $\nu_1 = q_2\nu$. If the latter, $Q(\nu\mu) = \nu_1^2\mu^2$, so $Q(Y) = \pi_1^2 Y^2 = (EX_1 \mid Y)^2$. This means $\mathrm{Var}(X_1 \mid Y) = 0$, or equivalently $\nu_2 = 0$, a contradiction. □

Lemma 4.1 can be used to prove many other conditional expectation formulas. For example, let X_1 and X_2 in (4.1) have NEF-QVF distributions and $r \geq 0$ be an integer. Let $P_r(X_1, \nu_1\mu_0)$ be the rth orthogonal polynomial with $EX_1 = \nu_1\mu_0$ and let $P_r^*(Y, \nu\mu_0)$ be the corresponding polynomial for the distribution $Y = X_1 + X_2$ in (4.2). Then

$$EP_r(X_1, \nu_1\mu_0) \mid Y = \pi_1^r \frac{b_r}{b_r^*} P_r(Y, \nu\mu_0). \tag{4.6}$$

Here we must redefine $b_r \equiv \prod_0^{r-1} (1 + i\nu_2/\nu_1)$ and $b_r^* = \prod_0^{r-1} (1 + i\nu_2/\nu)$ in adjusting (3.2) to account for $\nu_1 \neq 1$, $\nu \neq 1$. Formula (4.6) follows easily from Lemma 4.1 and (3.4), which states that both sides of (4.6) have expectation $b_r(\nu_1\mu - \nu_1\mu_0)^r$.

Formula (4.6) gives an explicit method for applying the Rao-Blackwell theorem to $g(X_1)$, ie. for computing the UMVUE $Eg(X_1) \mid Y$, if $g(X_1)$ can be expressed as $\sum_0^\infty c_r P_r(X_1, \nu_1\mu_0)$ for some μ_0.

Finally, let X_1, X_2 be independent with $X_i \sim$ NEF-QVF $(\nu_i\mu, \nu_i V(\mu))$ as in (4.1) and $Y = X_1 + X_2$. The conditional distributions are: (a) normal case, $X_1 \sim N(\nu_i\mu, \nu_i V)$, $X_1 \mid Y \sim N(\pi_1 Y, \pi_1\pi_2\nu V)$; (b) Poisson case, $X_i \sim \mathrm{Poiss}(\nu_i\mu)$, $X_1 \mid Y \sim \mathrm{Bin}(Y, \pi_1)$; (c) gamma case, $X_i \sim \mathrm{Gam}(\nu_i, \mu)$, $X_1 \mid Y \sim Y \cdot \mathrm{Beta}(\nu_1, \nu_2)$; (d) binomial case, $X_i \sim \mathrm{Bin}(\nu_i, p)$, $\mu = p$, $X_1 \mid Y \sim \mathrm{HG}(Y, \pi_1; \nu)$, the hypergeometric distribution with density $\binom{\nu_1}{x_1}\binom{\nu_2}{y - x_1}/\binom{\nu}{y}$ on $x_1 = 0, 1, \cdots, y$; and (e)negative binomial case, $X_i \sim \mathrm{NB}(\nu_i, p)$, $\mu = p/(1-p)$, $X_1 \mid Y \sim \mathrm{NHG}(Y, \pi_1; \nu)$, the negative hypergeometric (NHG) distribution on $x_1 = 0, 1, \cdots, Y$ with density when $Y = y$:

$$\binom{y}{x_1} \frac{\Gamma(\nu)\Gamma(x_1 + \nu_1)\Gamma(y - x_1 + \nu_2)}{\Gamma(y + \nu)\Gamma(\nu_1)\Gamma(\nu_2)} \tag{4.7}$$

This is also the marginal density of X_1 if $X_1|p \sim \mathrm{Bin}(y, p)$ and $p \sim \mathrm{Beta}(\nu_1, \nu_2)$, and so (4.7) also has been called the "beta-binomial distribution". The conditional distribution of X_1 given $X_1 + X_2$ when X_i has a NEF-GHS$(\nu_i\mu, \nu_i(1 + \mu^2))$ density is unnamed and apparently has not been considered.

These distributional results are summarized in Table 1. The means and variances of these distributions are given by (4.3) and (4.4).

5. Conjugate prior and posterior distributions. The sample mean X of n iid NEF$(\mu, V(\mu))$ distributions has a NEF$(\mu, V(\mu)/n)$ density

$$\exp\{nx\theta - n\psi(\theta)\}. \tag{5.1}$$

Let $\mu_0 \in \Omega$, the mean space, and $m > 0$. The conjugate prior distribution on θ mimics (5.1), being

$$g^*(\theta) = K \exp\{m\mu_0\theta - m\psi(\theta)\} \tag{5.2}$$

with $K = K(m, \mu_0)$ chosen to make $\int_\Theta g^*(\theta)\, d\theta = 1$. Here $g^*(\theta)$ is a two parameter family of densities for θ having a NEF with natural parameter μ_0, convolution parameter $m > 0$, m not necessarily an integer, even in the binomial case, and CGF $= -\log K(m, \mu_0)$. We think of (5.2) as a distribution on $\mu = \psi'(\theta)$, and not on θ. This usually is a non-linear transformation of the NEF for θ and therefore is an exponential family that is not a NEF, except in the case of the normal distribution. The density of μ on Ω with respect to $d\mu$ is

$$g(\mu) = K \exp\{m\mu_0\theta(\mu) - m\psi(\theta(\mu))\} V^{-1}(\mu) \tag{5.3}$$

with $\theta(\mu)$ denoting the inverse function of $\mu = \psi'(\theta)$. Jackson et al. (1970) note that (5.3) is a Pearson family in μ when X, has a NEF-QVF distribution. This is shown in Theorem 5.1. The converse also may be true, but is unproven. Then an expectation identity for these prior distributions is proven in Theorem 5.2.

THEOREM 5.1. *The densities* (5.3) *form a Pearson family on μ if $V(\mu)$ is quadratic.*

PROOF. $-\log g(\mu)$ has derivative $\{m(\mu - \mu_0) + V'(\mu)\}/V(\mu)$. This is the ratio of a linear to a quadratic function of μ, Pearson's condition, Kendall and Stuart (1963), if $V(\mu)$ is quadratic. □

THEOREM 5.2. *Let $h(\mu)$ have continuous derivative $h'(\mu)$ on Ω and be such that $Eh(\mu)(\mu - \mu_0)$ exists when μ has density* (5.3). *Assume at the end points of $\Omega = (a, b)$, a and/or b possibly infinite, that* $\lim h(\mu)V(\mu)g(\mu) = 0$ *as $\mu \to a$ or b. Then*

$$E(\mu - \mu_0)h(\mu) = \frac{1}{m} Eh'(\mu)V(\mu). \tag{5.4}$$

PROOF. We use integration by parts to write

$$\begin{aligned}\int_a^b h'(\mu)V(\mu)g(\mu)\,d\mu &= \int_a^b V(\mu)g(\mu)\,dh(\mu) = h(b)V(b)g(b) - h(a)V(a)g(a)\\ &\quad - \int_a^b h(\mu)[V'(\mu)g(\mu) - \{m(\mu - \mu_0) + V'(\mu)\}g(\mu)]\,d\mu\\ &= mE(\mu - \mu_0)h(\mu). \square\end{aligned}$$

The endpoint condition for Theorem (5.2) holds for all NEF-QVF distributions whenever $Eh(\mu)(\mu - \mu_0)$ exists. Define $M_r \equiv E(\mu - \mu_0)^r$ for $r = 0, 1, 2, \cdots$.

THEOREM 5.3. *For $r \geq 1$ and $V(\mu)$ quadratic, $M_0 = 1$, $M_1 = 0$ and*

$$M_{r+1} = \frac{r}{m - rv_2}\{V'(\mu_0)M_r + V(\mu_0)M_{r-1}\}. \tag{5.5}$$

PROOF. Let $h(\mu) = (\mu - \mu_0)^r$ in (5.4) and write $V(\mu) = V(\mu_0) + (\mu - \mu_0)V'(\mu_0) + v_2(\mu - \mu_0)^2$. Then (5.4) gives $M_{r+1} = (1/m)V(\mu_0)rM_{r-1} + (1/m)rV'(\mu_0)M_r + (1/m)v_2rM_{r+1}$. We show $M_1 = 0$ by choosing $h(\mu) = 1$ in (5.4). □

The expectations in Theorem 5.3 may not always exist. We have $K > 0$ in (5.3) whenever $m > 0$ and μ_0 is an interior point of Ω and $E\mu$ also always exists under these conditions. For $r \geq 2$, M_r exists if and only if r satisfies $(r - 1)v_2 < m$. This holds for all $r \geq 2$ when $v_2 \leq 0$, but not always for the reciprocal gamma, F, and t priors of the three distributions with $v_2 > 0$.

The central moments of the conjugate prior can be characterized in terms of the quadratic variance function $V(\mu)$ of the sampling distribution X by using (5.5). The first four are given below, defining $c_r \equiv \prod_1^{r-1}(m - iv_2)$ for $r \geq 2$, e.g. $c_2 = m - v_2$, $c_3 = (m - v_2)(m - 2v_2)$.

$$E\mu = \mu_0, \quad M_2 = \mathrm{Var}(\mu) = V(\mu_0)/c_2 \tag{5.6}$$

$$M_3 = E(\mu - \mu_0)^3 = 2V'(\mu_0)V(\mu_0)/c_3 \tag{5.7}$$

$$M_4 = E(\mu - \mu_0)^4 = \{3(m + 6v_2)V^2(\mu_0) + 6dV(\mu_0)\}/c_4. \tag{5.8}$$

The first three cumulants of μ are simple multiples of the first three cumulants of X at $\mu = \mu_0$, but that doesn't hold up for the fourth cumulant, $M_4 - 3M_2^2$, as can be checked by calculating it when $m = 1$ from (5.8) and (5.6) and comparing with the fourth cumulant $C_4(\mu) = 6v_2 V^2(\mu) + dV(\mu)$ of a NEF-QVF.

Formulas (5.5) – (5.8) can be used to determine central moments of the Beta($m\mu_0$, $m(1 - \mu_0)$) distribution because it is the conjugate prior to the Bernoulli distribution with $V(p) = p(1-p)$, $d = 1$, $v_2 = -1$ and $c_r = (1/m)(m + r - 1)^{(r-1)}$.

The conjugate prior distributions (5.3) on μ for the six NEF-QVF cases have two parameters, equivalent to the mean μ_0 and variance $V(\mu_0)/c_2$. These prior distributions on μ are: (a) normal if X is normal; (b) gamma if X is Poisson; (c) reciprocal gamma (the distribution of 1 divided by a gamma) if X is gamma; (d) beta if X is binomial; (e) F (the ratio of independent gammas) if X is negative binomial (because if p has a beta distribution then $\mu = p/q$ is the ratio of gammas). The NEF-GHS distribution has conjugate prior density

$$g(\mu) = K\exp\{m\mu_0\tan^{-1}(\mu)\}(1 + \mu^2)^{-(m+2)/2}. \tag{5.9}$$

This is a scaled Student's t_{m+1} density when $\mu_0 = 0$. Thus the t-distribution, and the Cauchy as $m \to 0$, arise as prior distributions conjugate to the NEF-GHS family. Table 1 summarizes these statements.

Now let X have a NEF-QVF density (5.1) and μ have the conjugate prior density (5.3). Then the posterior density has the same form as (5.3),

$$g_0(\mu) = K_0\exp\{Nx_0\theta(\mu) - N\psi(\theta(\mu))\}V^{-1}(\mu), \tag{5.10}$$

with $N \equiv n + m$, and $x_0 \equiv (nx + m\mu_0)/(n + m)$ the weighted average of the sample and *a priori* means. Formulas (5.4)–(5.8) also hold for the posterior distribution (5.10) by substituting N and x_0 for m and μ_0. Thus x_0 is the posterior mean

$$E(\mu \mid X = x) = x_0 = (1 - B)\,x + B\mu_0 \tag{5.11}$$

with B a shrinking factor,

$$B = \frac{EV(\mu)/n}{\mathrm{Var}(X)} = \frac{m}{m + n} = \frac{V(\mu_0) + v_2\tau_0^2}{V(\mu_0) + (n + v_2)\tau_0^2}, \tag{5.12}$$

and $\tau_0^2 \equiv \mathrm{Var}(\mu) = V(\mu_0)/c_2$ from (5.6). All but the last equality in (5.12) hold for any NEF, but that depends on QVF. It follows because for quadratic functions $V(\circ)$, $EV(\mu) = V(E\mu) + v_2\mathrm{Var}(\mu) = V(\mu_0) + v_2\tau_0^2$ and so $\mathrm{Var}(X) = E\,\mathrm{Var}(X \mid \mu) + \mathrm{Var}(EX \mid \mu) = EV(\mu)/n + \mathrm{Var}(\mu) = V(\mu_0)/n + (1 + v_2/n)\tau_0^2$. □

Ericson (1969) notes that (5.11) holds in NEFs with B given by the first equality in (5.12), where $\mathrm{Var}(X) = EV(\mu)/n + \mathrm{Var}(\mu)$. Diaconis and Ylvisaker (1979) characterize conjugate priors in NEFs as those having the linear property in (5.11).

THEOREM 5.4. *Let $M_r^* = E(\mu - x_0)^r \mid X$ be the* r*th central moment of μ given X, assuming X given μ has the* NEF-QVF *distribution* (5.1) *and μ has the conjugate prior distribution* (5.3). *Let $N = m + n$ and x_0 be given by* (5.11). *Then $M_0^* = 1$, $M_1^* = 0$, and for $r = 1, 2, \cdots$*

$$M_{r+1}^* = \frac{r}{N - rv_2}\{V'(x_0)M_r^* + V(x_0)M_{r-1}^*\}. \tag{5.13}$$

so that in special cases

$$M_2^* = \mathrm{Var}(\mu \mid X = x) = (1 - B)E\left\{\frac{V(\mu)}{n} \mid X = x\right\} = \frac{V(x_0)}{N - v_2} \tag{5.14}$$

TABLE 1

Natural Exponential Families with Quadratic Variance Functions, and related distributions. $X \mid \mu \sim$ NEF-QVF $(\mu, V(\mu)/n)$.

$X \mid \mu$: NEF-QVF	(a) Normal $N(\mu, \frac{V}{n})$	(b) Poisson $\frac{1}{n}$ Poiss $(n\mu)$	(c) Gamma Gam $(n, \mu/n)$	(d) Binomial Bin (n, p) $(\mu = p)$	(e) Negative Binomial NB (n, p) $(\mu = p/(1-p))$	(f) NEF-GHS $(\mu, (1+\mu^2)/n)$
$V(\mu)$	V	μ	μ^2	$\mu(1-\mu)$	$\mu + \mu^2$	$1 + \mu^2$
Special cases	—	—	exponential $(n = 1)$ Rayleigh Chi squared	Bernoulli $(n = 1)$	geometric $(n = 1)$	hyperbolic secant $(n = 1, \mu = 0)$
Exponential families: via monotone non-linear transformations	lognormal (e^x)	—	extreme value Weibull uniform $(n = 1)$	—	—	beta
Conditional distributions n_1X_1 given $Y = n_1X_1 + n_2X_2$ (Sec. 4)	normal	binomial	beta (scaled)	hyper-geometric	negative hyper-geometric or beta-binomial	—
Conjugate prior distribution on μ and posterior distribution of $\mu \mid X$ (Sec. 5)	normal	gamma	reciprocal gamma	beta	F	t, Cauchy and others
Marginal distribution of X (Sec. 6)	normal	negative binomial	F	negative hyper-geometric or beta-binomial	beta-Pascal	—

and

$$M_3^* = E\{(\mu - x_0)^3 \mid X = x\} = \frac{2V'(x_0)V(x_0)}{(N - v_2)(N - 2v_2)}. \tag{5.15}$$

PROOF. Because the posterior density (5.10) has the same form as the prior density (5.3), this result follows immediately from Theorem 5.3. □

Note that the squared posterior skewness of μ given x is

$$\gamma_3^2(x_0) = \frac{4(N - v_2)}{(N - 2v_2)^2}\{4v_2 + d/V(x_0)\}, \tag{5.16}$$

being proportional to the squared skewness of X given μ computed at $\mu = x_0$, $\{4v_2 + d/V(\mu_0)\}/n$. When n/N and $(N - v_2)/N$ are nearly unity, as with large samples, the posterior skewness of μ is approximately double the skewness of the sample mean.

Duan (1979) showed that M_{r^*} is a polynomial of degree r in X when X has a NEF-QVF distribution given μ. Formula (5.13) also proves that result. Duan does this to show for NEF-QVF distributions that the monomials $\{\mu^k, k = 0, 1, 2, \cdots\}$ form a "fan sequence", i.e. a basis of functions with subspaces invariant under the operator $E_\mu E^x$ with respect to conjugate prior distributions. He uses this and data from repeated problems to test that a specified prior distribution is appropriate.

We close this section by considering a little known but important robustness property of conjugate prior distributions. For a NEF-QVF distribution, let Π_0 be the class of all prior distributions on μ with specified mean μ_0 and variance $\tau_0^2 = V(\mu_0)/(m - v_2)$. The parameters μ_0, m for the conjugate prior distribution are equivalent to μ_0, τ_0^2. Let $\pi_0 \in \Pi_0$ be the conjugate prior on μ with these two moments.

DEFINITION. An estimator $t(x)$ of μ is said to be *empirical Bayes minimax* for squared error loss with respect to Π_0 (Morris, 1983), or *G_2 minimax* (Jackson, et al. 1970), if it is minimax with respect to the risk function $r(\pi, t) \equiv E_\pi(t(X) - \mu)^2$. This is a double expectation, X given μ random and μ distributed according to $\pi \in \Pi_0$.

Denote the Bayes estimator with respect to $\pi \in \Pi_0$ by $t_\pi(x) = E_\pi \mu \mid X = x$, the posterior mean, $t_{\pi_0}(x)$ being the linear estimator (5.11). Jackson, et al. (1970) prove the following theorem.

THEOREM 5.5. *Let X have a* NEF-QVF$(\mu, V(\mu)/n)$ *distribution. Then the conjugate prior $\pi_0 \in \Pi_0$ is G_2 minimax; equivalently t_{π_0} is empirical Bayes minimax with respect to Π_0 and squared error loss. That is*

$$r(\pi, t_{\pi_0}) = r_0 \equiv r(\pi_0, t_{\pi_0}) \le r(\pi_0, t) \tag{5.17}$$

for every estimator t and for all $\pi \in \Pi_0$.

PROOF. First compute $r(\pi, t_{\pi_0})$

$$\begin{aligned}
&= E_\pi\{(1 - B)X + B\mu_0 - \mu\}^2 = (1 - B)^2 E_\pi V(\mu)/n + B^2 E_\pi(\mu - \mu_0)^2 \\
&= (1 - B)^2\{V(\mu_0) + v_2\tau_0^2\}/n + B^2\tau_0^2 \\
&= (1 - B)\{V(\mu_0) + v_2\tau_0^2\}/n \equiv r_0
\end{aligned}$$

with B given by (5.12) and $\tau_0^2 = V(\mu_0)/(m - v_2)$. Observe that r_0 is independent of π. The inequality in (5.17) holds because t_{π_0} is the Bayes rule for π_0, and is strict if $t \ne t_{\pi_0}$. □

Theorem 5.5 justifies using the conjugate prior in Bayes and empirical Bayes practice when one has little knowledge of the distribution of μ beyond its first two moments. In that case choosing $\pi \ne \pi_0$ can be risky because the statistician thinks his risk is $r(\pi, t_\pi) <$

r_0 but it may actually be $r(\pi^*, t_\pi) > r_0$ if some other $\pi^* \in \Pi_0$ obtains. Only the conjugate prior avoids this hazard.

The concept of empirical Bayes minimax is mentioned here with G_2 minimax because it is part of a general approach to empirical Bayes inference (Morris, 1983).

6. Marginal distributions arising from conjugate prior distributions. We consider here the marginal moments and distributions of X when $X \mid \mu \sim$ NEF-QVF$(\mu, V(\mu)/n)$ with density (5.1) given μ and $\mu \sim CP(\mu_0, V(\mu_0)/(m - v_2))$, the conjugate prior density with mean μ_0, variance $\tau_0^2 \equiv V(\mu_0)/(m - v_2)$, and density (5.3).

The NEF-QVF orthogonal polynomials are $P_r(X, \mu_0)$ at μ_0. In this context, one must be careful to use $V(\mu)/n$, not $V(\mu)$ for the variance function in the polynomials, Morris (1982, Section 8), and we continue to define $b_r = \pi_0^{r-1}(1 + iv_2/n)$ when $n \neq 1$, and $a_r \equiv r!b_r$. Let M_r denote the rth central moment of μ, determined by Theorem 5.3.

THEOREM 6.1. *With E denoting expectation over the marginal distribution of X,*

$$EP_r(X, \mu_0) = b_r M_r. \tag{6.1}$$

Thus $EX = \mu_0$ and

$$\operatorname{Var}(X) = E(X - \mu_0)^2 = \frac{V(\mu_0)}{n} + b_2\tau_0^2 = \frac{V(\mu_0)}{n}\frac{m+n}{m - v_2}. \tag{6.2}$$

PROOF. $E\{EP_r(X, \mu_0) \mid \mu\} = b_r E(\mu - \mu_0)^r$ via (3.4), proving (6.1). Then (6.2) follows because $b_2 M_2 = b_2 \tau_0^2 = EP_2(X, \mu_0)$

$$\begin{aligned} &\equiv E\{(X - \mu_0)^2 - V'(\mu_0)(X - \mu_0)/n - V(\mu_0)/n\} \\ &= \operatorname{Var}(X) - V(\mu_0)/n. \ \square \end{aligned}$$

Higher central moments of X can be developed with some difficulty by expressing $(X - \mu_0)^r$ in terms of the orthogonal polynomials, and applying (6.1). Eg., it can be checked that

$$\begin{aligned} (X - \mu_0)^3 = {} & P_3(X, \mu_0) + 3V'(\mu_0)P_2(X, \mu_0)/n \\ & + \{(3 + 6v_2)V(\mu_0)/n + d\}(X - \mu_0)/n + V'(\mu_0)V(\mu_0)/n^2 \end{aligned} \tag{6.3}$$

so the third marginal moment is

$$\begin{aligned} E(X - \mu_0)^3 &= b_3 M_3 + 3V'(\mu_0)b_2 M_2/n + V'(\mu_0)V(\mu_0)/n^2 \\ &= V'(\mu_0)V(\mu_0)\left(\frac{2b_3}{c_3} + \frac{3b_2}{nc_2} + \frac{1}{n^2}\right). \end{aligned} \tag{6.4}$$

The marginal distributions of X are: (a) $N(\mu_0, V/n + \tau_0^2)$ for $X \mid \mu \sim N(\mu, V/n)$ and $\mu \sim N(\mu_0, \tau_0^2)$; (b) $(1/n)NB(m, n\mu_0/(m + n\mu_0))$ for $X \mid \mu \sim (1/n)$ Poiss$(n\mu)$ and $\mu \sim$ Gam$(m, \mu_0/m)$; (c) $(\mu_0 m/(m + 1))F_{2n,2m+2}$ for $X \mid \mu \sim$ Gam$(n, \mu/n)$ and $\mu \sim \mu_0 m/$Gam$(m + 1, 1)$; (d)$(1/n)$NHG$(n, \mu_0; m)$ if $X \mid p \sim 1/n$ Bin(n, p) and $p \sim$ Beta$(m\mu_0, m(1 - \mu_0))$; and (e) the beta-Pascal distribution $1/n$ BPasc$(n, p_0 = m\mu_0/(m\mu_0 + m + 1); m)$ with mean μ_0 and variance $((\mu_0 + \mu_0^2)/n)(m + n)/(m - 1)$ for $Y \equiv nX$ on the integers $Y = 0, 1, 2, 3, \cdots$ if $X \mid \mu \sim (1/n)$ $NB(n, p = \mu/(1 + \mu))$ and $\mu \sim \mu_0(m/(m + 1))F_{2m\mu_0, 2m+2}$ ($p \sim$ Beta$(m\mu_0, m + 1)$). The NEF-GHS marginal distribution, based on its conjugate prior, seems not to have been considered. These results are contained in Table 1.

7. Parametric empirical Bayes estimation for NEF-QVF distributions. Suppose the estimation problem of Section 5 is repeated k times, the statistician observing k independent sufficient statistics X_i, each a sample mean computed from n observations, with NEF-QVF distributions

$$X_i | \mu_i \overset{\text{ind}}{\sim} \text{NEF}(\mu_i, V(\mu_i)/n), \quad i = 1, \cdots, k. \tag{7.1}$$

The k parameters $\mu_1, \cdots, \mu_k$ may differ, but independently follow the same conjugate prior distribution

$$\mu_i \overset{\text{ind}}{\sim} CP(\mu_0, \tau_0^2 = V(\mu_0)/(m - v_2)), \quad i = 1, \cdots, k \tag{7.2}$$

with first two moments (μ_0, τ_0^2) and density (5.3). Thus the k random pairs (X_i, μ_i) are independent and exchangeable. Were (μ_0, τ_0^2) known, the posterior mean (5.11) would provide good point estimates for the μ_i, at least for squared error loss function. It they are unknown, we may proceed as follows.

Let $\bar{X} = \sum X_i/k$ and $S = \sum_1^k (X_i - \bar{X})^2$ be the mean and the sum of squares between the k groups.

THEOREM 7.1. *With (X_i, μ_i) distributed as in (7.1), (7.2) then*

$$E\mu_i | X_i = (1 - B)X_i + BE\bar{X} \tag{7.3}$$

with

$$B = \frac{v_2}{n + v_2} \frac{k - 1}{k} + \frac{n}{n + v_2} E\left\{\frac{V(\bar{X})}{n}\right\} \frac{k - 1}{ES}, \tag{7.4}$$

where B is defined in (5.12) and E in (7.3), (7.4) involves marginal distributions.

PROOF. Note that

$$EV(X_i) = E\{EV(X_i) \mid \mu_i\} = E\{v_2 \text{Var}(X_i|\mu_i) + V(\mu_i)\} = \left(\frac{v_2}{n} + 1\right)EV(\mu_i),$$

and hence we can find an unbiased estimate of $EV(\mu_i)$ by averaging the $V(X_i)$. However $(1/k) \sum V(X_i) = v_2S/k + V(\bar{X})$. Thus $E\{v_2S/k + V(\bar{X})\} = ((v_2/n) + 1)EV(\mu_i)$ and it follows that

$$B \equiv \frac{EV(\mu_i)/n}{\text{Var}(X_i)} = \frac{E\{v_2S/k + V(\bar{X})\}/(n + v_2)}{ES/(k - 1)},$$

the shrinking factor defined in (5.12), simplifies to (7.4). □

Formulas (7.3) and (7.4) strongly suggest parametric empirical Bayes (PEB) estimators when (μ_0, τ^2) are unknown and $k \geq 4$. If

$$EV(\bar{X})(k - 1)/ES \doteq EV(\bar{X})(k - 3)/S, \tag{7.5}$$

one can remove expectations from (7.3), (7.4) to approximate the posterior mean by $\hat{\mu}_i$ defined by

$$E\mu_i | X_i = \hat{\mu}_i \equiv (1 - \hat{B})X_i + \hat{B}\bar{X}. \tag{7.6}$$

Here we take

$$\hat{B} = \frac{v_2}{n + v_2} \frac{k - 1}{k} + \frac{n}{n + v_2} \hat{B}_{JS} \tag{7.7}$$

with $\hat{B}_{JS} \equiv (k - 3)V(\bar{X})/nS$, the naive extension of the James-Stein shrinking factor.

DISCUSSION.

1. Formula (7.6) is the James-Stein estimator (1961) if X_i is normal.
2. Non-normal empirical linear Bayes rules were discussed in general by Efron-Morris (1973, Section 9) and by Robbins (1982) for distributions with quadratic variance functions.

3. The difference $\hat{B} - \hat{B}_{JS}$ ordinarily has the sign of v_2 and diminishes as $1/n$. Thus, relative to $\hat{B}$, $\hat{B}_{JS}$ overshrinks in the binomial case ($v_2 = -1$), undershrinks (is conservative) by about $1/n$ in the gamma, negative binomial and NEF-GHS cases ($v_2 = 1$), and is correct for the Poisson case ($v_2 = 0$). For Poisson estimation,

$$\hat{B} = (k - 3)\bar{X}/nS.$$

4. Note we use $\hat{B}_{JS}$ by taking σ^2 to be $V(\bar{X})$ in the James-Stein coefficient, which is independent of i, even though $\mathrm{Var}(X_i \mid \mu_i) = V(\mu_i)/n$ actually depends on i. However, most applications have unequal sample sizes n_i for each component. Those more difficult problems can be treated in a manner analogous to Theorem 7.1 by using the results of Theorem 6.1 to estimate (5.12),

$$B_i \equiv \{V(\mu_0) + v_2\tau_0^2\}/\{V(\mu_0) + (n_i + v_2)\tau_0^2\}.$$

5. The approximation (7.5) will improve as either n or k increases, either justifying the normal theory independence of $\bar{X}$, S or making $\bar{X}$ degenerate at μ_0.

6. The approximation $(k-1)/ES \doteq E(k-3)/S$ used in (7.5) is exact in the normal case when S has a Chi squared distribution. Of course it is terrible for discrete distributions, for then $P(S = 0) > 0$, but since $B \leq 1$, one should force $\hat{B} \leq 1$, and with this modification (7.6), (7.7) will be much better behaved.

8. Concluding remarks. The references to this paper indicate a scattered, somewhat redundant and disconnected literature concerning the NEF-QVF distributions. Because of this scatteredness, some authors understandably were unaware of related work. To help cure this deficiency, and possibly similar deficiencies in this paper, I encourage readers to forward any further missing references to me.

Several results in Morris (1982) appeared earlier without citations. The results (7.2), (7.3) page 74 there about the relation of moments to lower order moments and cumulants appear in Kendall and Stuart (1963, Exercise 3.9). A statement page 77 about the Pollaczek polynomials should have noted that they were defined for the full NEF-GHS family (Szego, 1975, page 395). Seth (1949, Section 5) discussed the family of NEF-QVF orthogonal polynomials.

An important paper extending the theory here is that of Nelder and Wedderburn (1972) who introduce the variance function and use it in exponential families to find algorithms for maximum likelihood estimation of NEF linear regression parameters. Also see Wedderburn's (1974) further development of that theory.

REFERENCES

ABBEY, J. L., and DAVID, H. T. (1970). The construction of uniformly minimum variance unbiased estimators for exponential distributions. *Ann. Math. Statist.* **41** 1217–1222.

BLIGHT, B. J. N., and RAO, P. V. (1974). The convergence of Bhattacharya bounds. *Biometrika* **61** 137–142.

BOLGER, E. M., and HARKNESS, W. L. (1965). Characterizations of some distributions by conditional moments. *Ann. Math. Statist.* **36** 703–705.

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

DUAN, N. (1979). Significance test for prior distributions: the modified efficient score test and its asymptotic theory. Technical Report No. 135, Department of Statistics, Stanford University.

EFRON, B. and MORRIS, C. (1973). Stein's estimation rule and its competitors—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. B* **31** 332–334.

FEND, A. V. (1959). On the attainment of the Cramer-Rao and Bhattacharya bounds for the variance of an estimate. *Ann. Math. Statist.* **30** 381–388.

GUOYING, L., and PING, C. (1981). Some results about the type of C-R lower bound. *Communications in Science.* Address: Research Institute of System Sciences, Chinese Academy of Sciences, Beijing 100080, People's Republic of China.

JACKSON, D. A., O'DONOVAN, T. M., ZIMMER, W. J., and DEELY, J. J. (1970). Minimax estimators in the exponential family. *Biometrika* **57** 439–443.

JAMES, W. and STEIN, C. (1961). Estimation with quadratic loss. *Proceedings of the Fourth Berkeley Symposium.* University of California Press **1** 361–379.

KENDALL, M., and STUART, A. (1963). *The Advanced Theory of Statistics*, Vol. I, *Distribution Theory*, 2nd ed. Griffin, London.

LAHA, R. G., and LUKACS, E. (1960). On a problem connected with quadratic regression. *Biometrika* **47** 335–343.

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

MORRIS, C. N. (1983). Parametric empirical Bayes inference: Theory and Applications. *J. Amer. Statist. Assoc.* **78** 47–65.

NELDER, J. A. and WEDDERBURN, R. W. M. (1972). Generalized linear models. *J. Roy. Statist. Soc. Ser. A* **135** 370–384.

PATIL, G. P. and SHORROCK, R. (1965). On certain properties of the exponential-type families. *J. Roy. Statist. Soc.* **27** 94–99.

RAO, C. R. (1952). Some theorems on minimum variance estimation. *Sankhyā* **12** 27–42.

ROBBINS, H. (1982). Recent thoughts on empirical Bayes. Neyman Memorial Lecture, IMS Annual Meeting, Cincinnati.

SETH, G. R. (1949). On the variance of estimates. *Ann. Math. Statist.* **20** 1–27.

SHANBAG, D. N. (1972). Some characterizations based on the Bhattacharya matrix. *J. Appl. Probab.* **9** 580–587.

SZEGO, G. (1975). *Orthogonal Polynomials*, 4th ed. Vol. 23. Amer. Math. Soc. Colloquiums Publications, Providence, R. I.

TWEEDIE, M. C. K. (1946). The regression of the sample variance on the sample mean. *J. Lond. Math. Soc.* **21** 22–28.

WEDDERBURN, R. M. W. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika* **61** 439–447.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF TEXAS
AUSTIN, TEXAS 78712