The Annals of Statistics
1982, Vol. 10, No. 1, 65–80

NATURAL EXPONENTIAL FAMILIES WITH QUADRATIC VARIANCE FUNCTIONS[1]

BY CARL N. MORRIS

University of Texas, Austin

The normal, Poisson, gamma, binomial, and negative binomial distributions are univariate natural exponential families with quadratic variance functions (the variance is at most a quadratic function of the mean). Only one other such family exists. Much theory is unified for these six natural exponential families by appeal to their quadratic variance property, including infinite divisibility, cumulants, orthogonal polynomials, large deviations, and limits in distribution.

1. Introduction. The normal, Poisson, gamma, binomial, and negative binomial distributions enjoy wide application and many useful mathematical properties. What makes them so special? This paper says two things: (i) they are *natural exponential families* (*NEFs*); and (ii) they have *quadratic variance functions* (*QVF*), i.e., the variance $V(\mu)$ is, at most, a quadratic function of the mean μ for each of these distributions.

Section 2 provides background on general exponential families, making two points. First, because of some confusion about the definition of exponential families, the terms "natural exponential families" and "natural observations" are introduced here to specify those exponential families and random variables whose convolutions comprise one exponential family. Second, the "variance function" $V(\mu)$ is introduced as a quantity that characterizes the NEF.

Only six univariate, one-parameter families (and linear functions of them) are natural exponential families having a QVF. The five famous ones are listed in the initial paragraph. The sixth is derived in Section 3 as the NEF generated by the hyperbolic secant distribution. Section 4 shows this sixth family contains infinitely divisible, generally skewed, continuous distributions, with support $(-\infty, \infty)$.

In Sections 6 through 10, natural exponential families with quadratic variance functions (NEF-QVF) are examined in a unified way with respect to infinite divisibility, cumulants, orthogonal polynomials, large deviations, and limits in distribution. Other insights are obtained concerning the possible limit laws (Section 10), and the self-generating nature of infinite divisibility in NEF-QVF distributions.

This paper concentrates on general NEF-QVF development, emphasizing the importance of the variance function $V(\mu)$, the new distributions, and the five unified results. Additional theory for NEF-QVF distributions, e.g., concerning classical estimation theory, Bayesian estimation theory, and regression structure, will be treated in a sequel to this paper. Authors who have established certain statistical results for NEF-QVF distributions

Received June 1981.

[1] Support for this work provided by NSF Grant No. MCS-8003416 at the University of Texas, and by a grant from the Department of Health and Human Services to The Rand Corporation, where the author is currently a consultant, is gratefully acknowledged.

AMS 1970 *subject classifications.* Primary, 60E05, 60E07, 60F10, 62E10, 62E15, 62E30.

Key words and phrases. Exponential families, natural exponential families, quadratic variance function, normal distribution, Poisson distribution, gamma distribution, exponential distribution, binomial distribution, negative binomial distribution, geometric distribution, hyperbolic secant distribution, orthogonal polynomials, moments, cumulants, large deviations, infinite divisibility, limits in distribution, variance function.

include Shanbag (1972) and Blight and Rao (1974) for Bhattacharya bounds; Laha and Lukacs (1960) and Bolger and Harkness (1965) for quadratic regression; and Duan (1979) for conjugate priors.

2. Natural exponential families and variance functions. A parametric family of distributions with natural parameter set $\Theta \subset R$ (the real line) is a univariate exponential family if random variables Y governed by these distributions satisfy

$$P_\theta(Y \in A) = \int_A \exp\{\theta T(y) - \psi(\theta)\}\xi(dy) \tag{2.1}$$

for some measure ξ not depending on $\theta \in \Theta$, $A \subset R$ a measurable set, and T a real valued measurable function (Lehmann, 1959, Barndorff-Nielsen, 1978). Any factor of the density of y not depending on θ is absorbed into ξ. In (2.1), θ is called the *natural parameter* and Θ, which is the largest set (an interval, possibly infinite) for which (2.1) is finite when $A = R$, is called the *natural parameter space.* Often θ is a nonlinear function of some more familiar parameter. The function $\psi(\theta)$ is determined by ξ so that (2.1) has unit probability if $A = R$.

The natural observation in (2.1) is $X = T(Y)$. Its distribution belongs to a *natural exponential family* (NEF),

$$P_\theta(X \in A) = \int_A \exp\{\theta x - \psi(\theta)\}\, dF(x) \tag{2.2}$$

with F a Stieltjes measure on R. If $0 \in \Theta$ and $\psi(0) = 0$ then F is a cumulative distribution function (CDF). Otherwise, choose any $\theta_0 \in \Theta$ and let $dF_0(x) \equiv \exp\{\theta_0 x - \psi(\theta_0)\}\, dF(x)$. Then F_0 is a CDF, and via (2.2) generates the same exponential family as F. Hence F in (2.2) is taken to be a CDF without loss of generality.

The modifier "natural" is needed to distinguish NEFs from exponential families not satisfying (2.2). For example, $Y \sim \text{Beta}(m\mu, m(1-\mu))$ and $Y \sim \text{Lognormal}(\alpha, \sigma^2)$ satisfy (2.1) with $T(y) = m\log(y/(1-y))$, $\theta = \mu$ and $T(y) = \log(y)$, $\theta = \alpha/\sigma^2$, respectively, but they are not NEFs. Convolutions of NEFs, $\Sigma X_i = \Sigma T(Y_i)$, are exponential families with natural parameter θ (and NEFs), but ΣY_i is not, unless T is linear. Cumulants of $X = T(Y)$ are derivatives of $\psi(\theta)$, but no simple expression yields cumulants of Y. Thus the results being developed here pertain only to NEFs.

Every univariate CDF F_0 possessing a moment generating function (MGF) in a neighborhood of zero *generates* a NEF, as follows. Define the cumulant generating function (CGF) $\psi(\theta)$ on Θ, the largest interval for which (2.3) exists, by

$$\psi(\theta) \equiv \log \int \exp(\theta x)\, dF_0(x), \qquad \theta \in \Theta. \tag{2.3}$$

Then the CDFs F_θ, $\theta \in \Theta$, defined by

$$dF_\theta(x) \equiv \exp(\theta x - \psi(\theta))\, dF_0(x) \tag{2.4}$$

form a NEF, with F_θ a CDF. The NEF so generated was called a *conjugate family,* (Khinchin, 1949) predating its use by Bayesians for quite a different purpose, and plays an important role in expansion theory (Barndorff-Nielsen and Cox, 1979). Given any $\theta^* \in \Theta$, F_{θ}^* generates the same NEF (2.3) and (2.4). Thus the NEF is closed in that it can be generated by any of its members. The mean, variance, and CGF are

$$\mu = E_\theta X = \int x\, dF_\theta(x) = \psi'(\theta) \tag{2.5}$$

$$V(\mu) = \text{Var}_\theta(X) = \int (x-\mu)^2\, dF_\theta(x) = \psi''(\theta) \tag{2.6}$$

$$\psi_\theta(t) = \log \int \exp(tx)\, dF_\theta(x) = \psi(t+\theta) - \psi(\theta). \tag{2.7}$$

In (2.3), $\psi(0) = 0$ since F_0 is a distribution function, so $F_\theta = F_0$ when $\theta = 0$ in (2.4). The results in (2.5) and (2.6) follow for $r = 1, 2$ from (2.7) and the property of the CGF that the rth cumulant, C_r of F_θ is given by

$$C_r \equiv \left.\frac{d^r\psi_\theta(t)}{dt^r}\right|_{t=0} = \psi^{(r)}(\theta). \tag{2.8}$$

The range of $\psi'(\theta)$ over $\theta \in \Theta$ will be denoted Ω, i.e., $\Omega = \psi'(\Theta)$. Ω is an interval (possibly infinite) of R. In applications, the mean $\mu \in \Omega$ of the distribution is a more standard parameterization than $\theta \in \Theta$ for most NEFs.

The variance $\psi''(\theta)$ depends on $\mu = \psi'(\theta)$ as does every function of θ because μ is a 1 $-$ 1 function of θ since $\psi''(\theta) > 0$. Formula (2.6) expresses this fact and we name $V(\mu)$, together with its domain Ω, the *variance function* (VF) of the NEF.

The variance function (Ω, V) characterizes the NEF, but no particular member of the NEF, because it determines the CGF, and hence the characteristic function, as follows. Given $V(\cdot)$ and μ_0 with $0 < V(\mu_0) < \infty$, define Ω as the largest interval containing μ_0 such that $0 < V(m) < \infty$ for all $m \in \Omega$. (Any other $\mu_0 \in \Omega$ regenerates Ω in the same way.) Now define $\psi(\cdot)$ by

$$\psi\left(\int_{\mu_0}^{\mu} \frac{dm}{V(m)}\right) = \int_{\mu_0}^{\mu} \frac{m\, dm}{V(m)} \tag{2.9}$$

for all $\mu \in \Omega$. Note that $\theta = \int_{\mu_0}^{\mu} dm/V(m)$ and the range of θ as μ varies over Ω is Θ, the natural parameter space. Validity of (2.9) follows from differentiation with respect to μ. In (2.9), $\psi(0) = 0$, $\psi'(0) = \mu_0$, and $\psi''(0) = V(\mu_0)$.

Observe that V without Ω may not characterize the NEF. For example, $V(\mu) = \mu^2$ characterizes two different NEFs, one with $\Omega_1 = (-\infty, 0)$ and the other with $\Omega_2 = (0, \infty)$. These correspond to the usual exponential distribution (Ω_2) and the negative of the usual exponential distribution (Ω_1), the sets being separated by points (in this case one point) with $V(\mu) \le 0$. If F_0 is not the CDF of a degenerate distribution in (2.3) and (2.4), $V(\mu)$ is strictly positive throughout Ω.

The mean-space Ω of a NEF is the interior of the smallest interval containing the support $\mathscr{X}$ of F in (2.2). Thus $\Omega = \mathscr{X}$ if $\mathscr{X}$ is an open interval, and the closure of Ω always contains $\mathscr{X}$ (Efron, 1978, Section 7).

The variance function characterizes the particular NEF within all NEFs, but not within a wider family of distributions. For example, the beta and lognormal distributions mentioned after (2.2) have QVF with

$$V_1(\mu) = \mu(1-\mu)/(m+1)$$

and

$$V_2(\mu) = \mu^2\{\exp(\sigma^2) - 1\}, \qquad \mu \equiv \exp(\alpha + \sigma^2/2),$$

respectively. These VFs match those of the binomial and gamma, respectively, and so fail to characterize the family within all exponential families.

Formula (2.8) is easily modified to generate cumulants of any NEF in terms of the mean $\mu = \psi'(\theta)$. If $C_r(\mu)$ is the rth cumulant expressed in terms of μ, then $C_1(\mu) = \mu$, $C_2(\mu) = V(\mu)$, and

$$C_{r+1}(\mu) = V(\mu)C_r'(\mu), \qquad r \ge 2, \tag{2.10}$$

primes denoting derivatives wrt μ. We have

$$C_{r+1}(\mu) = \psi^{(r+1)}(\theta) = d\psi^{(r)}(\theta)/d\theta$$
$$= \left\{\frac{d\psi^{(r)}(\theta)}{d\mu}\right\}\left(\frac{d\mu}{d\theta}\right) = C_r(\mu)C_2(\mu),$$

which is (2.10). Thus, suppressing dependence on μ,

$$C_3(\mu) = V'V, \qquad C_4(\mu) = V(V')^2 + V^2V'' \tag{2.11}$$

are expressed in terms of derivatives of the variance function, as can be higher cumulants, using higher derivatives of $V(\mu)$.

3. NEFs with quadratic variance functions. The development in Section 2 shows that many NEFs exist. A few have quadratic variance function (QVF)

$$V(\mu) = v_0 + v_1\mu + v_2\mu^2. \tag{3.1}$$

Those with QVF include the normal $N(\mu, \sigma^2)$ with $V(\mu) = \sigma^2$ (constant variance function); the Poisson, Poiss(λ) with $\mu = \lambda$, $V(\mu) = \mu$ (linear); the gamma Gam(r, λ), $\mu = r\lambda$, $V(\mu) = r\lambda^2 = \mu^2/r$; the binomial, Bin($r$, p), $\mu = rp$, $V(\mu) = rpq = -\mu^2/r + \mu(q = 1 - p)$; and the negative binomial NB(r, p), the number of successes before the rth failure, p = probability of success, $\mu = rp/q$, $V(\mu) = rp/q^2 = \mu^2/r + \mu(q \equiv 1 - p)$. Table 1 lists properties of these distributions.

The four operations (i) linear transformation $X \to (X - b)/c$, (ii) convolution, (iii) division, producing $\mathscr{L}(X_1)$ from $\mathscr{L}(X)$ with $X = X_1 + \cdots + X_n$, X_i iid, as discussed below, $\mathscr{L}(Z)$ signifying the law of Z, and (iv) generation of the NEF as in (2.3), (2.4), all produce NEFs, usually different from that of X, carrying each member of the original NEF to the same new NEF. Thus NEFs are equivalance classes under these operations. These operations also preserve the QVF property, as shown next, and so preserve NEF-QVF structure.

If X has a NEF distribution (2.4) and $V(\mu)$ is quadratic, as in (3.1), we shall say X has a NEF-QVF distribution. Let $c \neq 0$ and b be constants, and let $X^* = (X - b)/c$ be a linear transformation of X with mean, $\mu^* = (\mu - b)/c$. Then $V^*(\mu^*) \equiv \mathrm{Var}(X^*) = V(\mu)/c^2 = V(c\mu^* + b)/c^2$, so

$$V^*(\mu^*) = V(b)/c^2 + V'(b)\mu^*/c + v_2(\mu^*)^2. \tag{3.2}$$

Thus $v_2 \to v_2$, $v_1 \to V'(b)/c$, and $v_0 \to V(b)/c^2$ if $X \to (X - b)/c$.

Define $d \equiv v_1^2 - 4v_0v_2$ to be the *discriminant of* $V(\mu)$. Since

$$\{V'(\mu)\}^2 = (2v_2\mu + v_1)^2 = 4v_2V(\mu) + d, \tag{3.3}$$

the discriminant d^* of (3.2) is $[V'(b)/c]^2 - 4v_2V(b)/c^2 = d/c^2$. Thus d is unaltered by translations, and $d \to d/c^2$ if $X \to (X - b)/c$.

Now let X_i, $i = 1, \cdots, n$, be independent identically distributed (iid) as a NEF-QVF, with $V(\mu)$ as in (3.1). Define $X^* = (X_1 + \cdots + X_n - nb)/c$. Then X^* has a NEF-QVF distribution, $\mu^* \equiv EX^* = n(\mu - b)/c$, $\Omega^* = (\Omega - b)n/c$, and

$$\begin{aligned} V^*(\mu^*) &= nV(b + c\mu^*/n)/c^2 \\ &= nV(b)/c^2 + V'(b)\mu^*/c + v_2(\mu^*)^2/n. \end{aligned} \tag{3.4}$$

Thus, if $X \to (X_1 + \cdots + X_n - nb)/c$, then

$$v_0 \to nV(b)/c^2, \qquad v_1 \to V'(b)/c, \qquad v_2 \to v_2/n, \tag{3.5}$$

and the discriminant

$$d \to d/c^2. \tag{3.6}$$

Formulas (3.4) and (3.5) show that the QVF property is preserved and how the VF changes under convolution and linear transformation.

The convolution operation sometimes can be reversed. If X follows a NEF-QVF distribution, suppose n can be given so that $X = X_1^* + \cdots + X_n^*$ for iid variables $\{X_1^*\}$. This is possible for any n with infinitely divisible distributions, but otherwise only in certain cases. Distributions permitting this for some $n \geq 2$ are termed *divisible* here, paralleling the infinitely divisible terminology. The operation producing $\mathscr{L}(X_1^*)$ from $\mathscr{L}(X)$ is termed *division*, and X_1^* or $\mathscr{L}(X_1^*)$ is called a *divisor* of X or of $\mathscr{L}(x)$. The VF of X_1^* is obtained from that of X by reversing (3.5), taking $b = 0$, $c = 1$. Then $v_0 \to v_0/n$, $v_1 \to v_1$, and $v_2 \to nv_2$ under division.

If more generally for some n, $X = c(X_1^* + \cdots + X_n^*) + b$, i.e., X_1^* is the n-divisor of $(X - b)/c$, then $\mu^* = EX_1^* = (\mu - b)/nc$ and $V^*(\mu^*) = \mathrm{Var}(X_1^*) = V(b + cn\mu^*)/nc^2$. This is (3.4), replacing n by $1/n$. Thus (3.5) unifies n-convolutions and n-divisions of X, using $1/n$ to represent division. By combining these two operations, n can take fractional values, and for infinitely divisible distributions n can be any positive real number. Neither the convolution nor the division operation affects the discriminant d, as shown by (3.6).

Finally, in NEF-QVF distributions, $\psi(\theta)$ is related to θ and $V(\mu)$ by

$$\log(V(\mu)/V(\mu_0)) = 2v_2[\psi(\theta) - \psi(\theta_0)] + v_1[\theta - \theta_0]. \tag{3.7}$$

with $\mu = \psi'(\theta)$, $\mu_0 = \psi'(\theta_0)$. Formula (3.7) is non-trivial except in the normal case ($v_1 = v_2 = 0$), and is proved by differentiating wrt θ.

4. Finding all NEFs with QVF: the sixth family. Because the normal distribution $N(\mu, \sigma^2)$ has $V(\mu) = \sigma^2$, and the VF characterizes the NEF, it follows that the normal distribution is the unique NEF with constant variance function ($v_2 = v_1 = 0$). Similarly, the Poisson distributions, including linear transformations of the usual Poisson, are the only NEFs with strictly linear variance function, $V(\mu) = v_1\mu + v_0$, $v_1 \neq 0$.

Let us characterize all *strictly* quadratic NEF-QVF distributions. Suppose X has a NEF-QVF distribution with VF (3.1), and $v_2 \neq 0$. Define $X^* = aV'(X)$, taking $a = 1$ if $d = 0$, $a = |dv_2|^{-1/2}$ otherwise. Then X^* is a linear function of X with variance function given by (3.2),

$$V^*(\mu^*) = s + v_2(\mu^*)^2, \qquad s = -\mathrm{sgn}(dv_2). \tag{4.1}$$

We may regard (4.1) as the *canonical* member under linear transformations of the NEF-QVF with $v_2 \neq 0$ specified ($v_1 = 0$, $s = 0, \pm 1$). All other quadratic VFs, i.e., all other v_0, v_1, can be obtained from the canonical VF (4.1) by the linear transformation $X = (X^*/a - v_1)/2v_2$, the inverse of the transformation $X^* = aV'(X)$.

Six cases with $v_2 \neq 0$ in (4.1) correspond to combinations of $v_2 < 0$, $v_2 > 0$ and $s = -1$, 0, 1. The two having $v_2 < 0$ with $s = -1$ and $s = 0$ make $V(\mu) < 0$, so are impossible. Three others correspond to (linear transformations of) the gamma ($v_2 > 0$, $s = 0$), the binomial ($v_2 < 0$, $s = 1$), and the negative binomial ($v_2 > 0$, $s = -1$), cf. Table 1.

The case $v_2 > 0$, $s = 1$ remains. For $v_2 = 1$, the missing distribution is the natural observation of a beta exponential family. Let $y \sim \mathrm{Beta}(0.5 + \theta/\pi, 0.5 - \theta/\pi)$, $|\theta| < \pi/2$. The natural observation for this exponential family is $x \equiv \log\{y/(1-y)\}/\pi$, having density

$$f_{1,\theta}(x) \equiv \frac{\exp[\theta x + \log\{\cos(\theta)\}]}{2\cosh(\pi x/2)} \tag{4.2}$$

with respect to Lebesgue measure and support $-\infty < x < \infty$. The proof that X has this density follows almost immediately from the reflection formula (Abramowitz and Stegun, 1965, page 256), which implies that $\beta(0.5 + t, 0.5 - t) = \pi/\cos(\pi t)$.

The mean and variance of (4.2) are derived by differentiating the CGF $\psi(\theta) = -\log(\cos(\theta))$ to get $EX = \mu = \psi'(\theta) = \tan(\theta)$, $\mathrm{Var}\,X = V(\mu) = \psi''(\theta) = \csc^2(\theta) = 1 + \mu^2$, so X has a NEF-QVF distribution. Convolutions (r times) and infinite divisibility of (4.2), to be discussed in the next section, yield all distributions with $V(\mu) = r + \mu^2/r$, per the discussion surrounding (3.5), for any $r > 0$.

TABLE 1
Univariate natural exponential families with quadratic variance functions

	Normal $N(\lambda, \sigma^2)$	**Poisson** Poiss(λ)	**Gamma** Gam(r, λ)
Density	$\frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\lambda)^2/2\sigma^2}$	$\frac{\lambda^x e^{-\lambda}}{x!}$	$\left(\frac{x}{\lambda}\right)^{r-1} \frac{e^{-x/\lambda}}{\lambda\Gamma(r)}$
	$-\infty < x < \infty$ $-\infty < \lambda < \infty$ $\sigma^2 > 0$	$x = 0, 1, 2, \cdots$. $0 < \lambda < \infty$	$0 < x < \infty$ $0 < \lambda < \infty$ $r > 0$
Inf. Divis. (Sec. 6)	Yes	Yes	Yes
Elem. Distn.	$N(\lambda, 1)$	Poiss(λ)	Exponential (λ)
θ	λ/σ^2	$\log \lambda$	$-1/\lambda$
Θ	$(-\infty, \infty)$	$(-\infty, \infty)$	$(-\infty, 0)$
$\psi(\theta)$	$\frac{\sigma^2\theta^2}{2} = \frac{\lambda^2}{2\sigma^2}$	$e^\theta = \lambda$	$-r \log(-\theta)$ $= r \log(\lambda)$
mean $= \mu = \psi'(\theta)$	$\lambda = \theta\sigma^2$	$\lambda = e^\theta$	$r\lambda = -r/\theta$
Ω	$(-\infty, \infty)$	$(0, \infty)$	$(0, \infty)$
$V(\mu) = \psi''(\theta)$		$\lambda = e^\theta$	$r\lambda^2 = r/\theta^2$
$= v_2\mu^2 + v_1\mu + v_0$	$v_0 = \sigma^2$	$= \mu$	$= \mu^2/r$
$d = v_1^2 - 4v_0v_2$	0	1	0
$\log E_\theta e^{tX}$ $= \psi(t + \theta) - \psi(\theta)$	$t\lambda + t^2\sigma^2/2$	$\lambda(e^t - 1)$	$-r \log(1 - t\lambda)$
Orthog. Polynomials (Sec. 8)	Hermite	Poisson-Charlier	Generalized Laguerre
Exponential Large Deviations (Sec. 9), $\epsilon > 0$	No Conditions	$\lambda \geq \epsilon$	$r \geq \epsilon$
Limit Laws (Sec. 10)	Normal	Normal	Normal

TABLE 1—*Continued.*
Univariate natural exponential families with quadratic variance functions

	Binomial Bin(r, p)	**Negative Binomial** NB(r, p)	**NEF-GHS** NEF-GHS(r, λ)
Density	$\binom{r}{x} p^x q^{r-x}$	$\frac{\Gamma(x+r)}{\Gamma(r)x!} p^x q^r$	Sec. 5
	$x = 0, 1, \cdots, r$ $0 < p < 1$ $r = 1, 2, \cdots$.	$x = 0, 1, 2, \cdots$. $0 < p < 1$ $r > 0$	$-\infty < x < \infty$ $-\infty < \lambda < \infty$ $r > 0$
Inf. Divis. (Sec. 6)	No	Yes	Yes
Elem. Distn.	Bernoulli (p)	Geometric (p)	NEF-HS(λ)
θ	$\log(p/q)$	$\log(p)$	$\tan^{-1}(\lambda)$
Θ	$(-\infty, \infty)$	$(-\infty, 0)$	$\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$
$\psi(\theta)$	$-r \log(q)$ $= r \log(1 + e^\theta)$	$-r \log(q)$ $= -r \log(1 - e^\theta)$	$-r \log(\cos(\theta))$ $= (r/2) \log(1 + \lambda^2)$
mean $= \mu = \psi'(\theta)$	$rp = r/(1 + e^{-\theta})$	$\frac{rp}{q} = \frac{r}{e^{-\theta} - 1}$	$r\lambda = r \tan(\theta)$
Ω	$(0, r)$	$(0, \infty)$	$(-\infty, \infty)$
$V(\mu) = \psi''(\theta)$	$rpq = \frac{re^\theta}{(1 + e^\theta)^2}$	rp/q^2	$r(1 + \lambda^2)$
$= v_2\mu^2 + v_1\mu + v_0$	$= -\mu^2/r + \mu$	$= \mu^2/r + \mu$	$= \mu^2/r + r$
$d = v_1^2 - 4v_0v_2$	1	1	-4
$\log E_\theta e^{tX}$ $= \psi(t + \theta) - \psi(\theta)$	$r \log(pe^t + q)$	$r \log\left(\frac{q}{1 - pe^t}\right)$	$-r \log(\cos(t) - \lambda \sin(t))$
Orthog. Polynomials (Sec. 8)	Krawtchouk	Meixner	Pollaczek
Exponential Large Deviations (Sec. 9), $\epsilon > 0$	$\epsilon \le p \le 1 - \epsilon$	$r, p \ge \epsilon$	$r \ge \varepsilon$
Limit Laws (Sec. 10)	Normal Poisson	Normal Poisson Gamma	Normal Gamma

Linear transformations yield other distributions for a specified $v_2 = 1/r > 0$, producing arbitrary v_0, v_1, subject to $d = v_1^2 - 4v_0v_2 < 0$; from (3.3), linear transformations must preserve the sign of the discriminant, and $d = -4 < 0$ from $V(\mu)$ above. Then if X^* has $V^*(\mu^*) = r + (\mu^*)^2/r$, $X = X^*(v_0/r - v_1^2/4)^{1/2} - rv_1/2$ has VF

$$V(\mu) = v_0 + v_1\mu + \mu^2/r, \qquad r > 0, \qquad v_1^2 < 4v_0/r. \tag{4.3}$$

When $\theta = 0$, (4.2) is called the "hyperbolic secant" (HS) distribution (Johnson and Kotz, 1970). Convolution and infinite division when $\theta = 0$ produces the "generalized hyperbolic secant" (GHS) distributions (Baten, 1934; Harkness and Harkness, 1968), all symmetric distributions.

The NEFs generated by the HS distribution (and generated by their linear transformations) appear to be new. They have VFs given by (4.3), and are skewed when $\theta \neq 0$. The natural exponential family generated by the generalized hyperbolic secant distributions will be referred to as the *NEF-GHS* distributions, reserving *NEF-HS* for (4.2). Properties of these new distributions are developed in the following sections.

We have just found that all univariate natural exponential families with quadratic variance functions are the normal (constant), Poisson (linear), gamma, binomial, negative binomial, and NEF-GHS distributions and linear transformations of these. These six distributions are summarized in Table 1.

Each of these distributions contains a subfamily, which we term the family of *elementary distributions*, with leading coefficient in $V(\mu)$ having unit magnitude $\pm$ 1. These are the normal distribution with unit variance, ($v_0 = 1$), the usual Poisson ($v_1 = 1$), the exponential ($v_2 = 1$ in the gamma), the Bernoulli ($v_2 = -1$, $r = 1$ in the binomial), the geometric ($v_2 = 1$, $r = 1$ in the negative binomial), and the NEF-HS (4.2).

Remarkably, just six simple distributions, one elementary distribution from each class, including the $N(0, 1)$, the Poisson ($\lambda = 1$), the exponential (mean = 1), the Bernoulli ($p = \frac{1}{2}$), the geometric ($p = \frac{1}{2}$), and the hyperbolic secant ($\theta = 0$), generate six of the main families of distributions in statistics via the (commutative) processes of:

1. linear transformation: $X \to (X - b)/c$
2. generation of the NEF, per (2.3)–(2.4)
3. convolution
4. infinite division (division in the binomial case).

A fifth process, not commuting with the above, produces many other named univariate exponential families including the lognormal, the beta, the extreme value, the Pareto, and the Weibull distributions, which are not NEFs. These are derived as

5. nonlinear transformations of NEF-QVF distributions, i.e., $X \to Y = T^{-1}(X)$, T as in (2.1).

5. The NEF-GHS Distributions. The hyperbolic secant density $f_{1,0}(x)$ in (4.2) with $\theta = 0$ is symmetric about 0 with unit variance and CDF

$$F_{1,0}(x) = \frac{1}{\pi} \arctan\left(\sinh \frac{\pi x}{2}\right) + \frac{1}{2} \tag{5.1}$$

on $-\infty < x < \infty$ (Johnson and Kotz, 1970). Convolutions and divisions of this distribution, which is infinitely divisible (Feller, 1971, also proved in Section 6) are said to have the *generalized hyperbolic secant distribution* (GHS) (Johnson and Kotz, 1970). These densities, all symmetric about 0, take fairly simple form for integral convolutions, $f_{r,0}(x) = f_{1,0}(x) * \cdots * f_{1,0}(x)$ r times. Then

$$f_{2,0}(x) = \frac{x}{2\sinh(\pi x/2)}, \tag{5.2}$$

and for $r \geq 1$, integers only,

$$f_{r+2,0}(x) = \frac{x^2 + r^2}{r(r+1)} f_{r,0}(x). \tag{5.3}$$

Thus $f_{r,0}(x)$ is a polynomial of degree r divided by $\cosh(\pi x/2)$ if r is an odd integer, and divided by $\sinh(\pi x/2)$ if r is even. The $f_{r,0}(x)$ densities when r is not an integer are expressible as infinite products

$$f_{r,0}(x) = \frac{2^{r-2}}{\Gamma(r)} \prod_{j=0}^{\infty} \{1 + x^2/(r+2j)^2\}^{-1}$$

(Johnson and Kotz, 1970).

The MGF of $f_{r,0}(x)$ is $1/\cos^r(t)$ and the characteristic function $1/\cosh^r(t)$. For $r = 1$, then, the characteristic function is $1/\cosh(t)$, the same as the density, making the hyperbolic secant density $f_{1,0}(x)$ "self-conjugate," like the normal density.

We note that $f_{1,0}(x)$ is the distribution of $(2/\pi)\log|C_1|$ and $f_{r,0}(x)$ is of $(2/\pi)\log|C_1C_2 \cdots C_r|$ with $C_1, C_2, \cdots, C_r$ independent Cauchy random variables. This follows easily for $r = 1$ from the fact that C_1 is the ratio of two independent normals, so $Y = C_1^2/(1 + C_1^2) \sim$ Beta(0.5, 0.5), and then the argument preceding (4.2) shows $X = \log\{Y/(1-Y)\}/\pi = 2\log|C_1|/\pi$ has density $f_{1,0}(x)$. Convolutions give the result for general r. Thus, the product of independent Cauchy distributed variables is simply related to (5.3).

Now let us generate the NEF $f_{r,\theta}(x)$ from $f_{r,0}(x)$ by using (2.3) and (2.4) to get

$$f_{r,\theta}(x) \equiv \exp\{\theta x + r \log\cos(\theta)\} f_{r,0}(x), \qquad |\theta| < \pi/2, \tag{5.4}$$

with the CGF $\psi(\theta) = -r\log\{\cos(\theta)\}$ obtained from the MGF of $f_{r,0}(x)$. It has MGF $\{\cos(\theta)/\cos(\theta+t)\}^r$, mean $\mu = \psi'(\theta) = r\tan(\theta)$ and variance $V(\mu) = \psi''(\theta) = \mu^2/r + r$ with discriminant $d = -4$. It follows from (5.3) and (5.4) that

$$f_{r+2,\theta}(x) = \cos^r(\theta)\frac{x^2 + r^2}{r(r+1)} f_{r,\theta}(x). \tag{5.5}$$

Let $\lambda \equiv \tan(\theta)$ be the mean of the elementary density $f_{1,0}(x)$ (cf. Section 4). Then for any $r > 0$,

$$f^*_{r,\lambda}(x) = (1+\lambda^2)^{-r/2}\exp\{x\tan^{-1}(\lambda)\} f_{r,0}(x) \tag{5.6}$$

may be a more useful parameterization, $-\infty < \lambda < \infty$. The GHS distributions ($\lambda = 0$) are bell-shaped with exponentially decaying tails. For $\lambda \neq 0$, $f^*_{r,\lambda}(x)$ has skewness coefficient $2\lambda/(r + r\lambda^2)^{1/2}$, bounded above and below (as $\lambda \to \infty$) by $\pm 2/r^{1/2}$, the skewness of Gam(r, α). In fact, as $\lambda \to \pm\infty$, $f^*_{r,\lambda}(x)$ is approximated by the density of $\pm G$ with $G \sim$ Gam(r, $|\lambda|$), a fact proved in Section 10. Thus λ, or θ, is a shape parameter measuring the nonsymmetry of $f^*_{r,\lambda}(x)$, and compromising between a bell-shaped distribution when $\lambda = 0$ and a gamma distribution as $|\lambda| \to \infty$. Of course, as $r \to \infty$, this bell-shaped distribution approaches the normal. With $r = 1$, the compromise is between the HS distribution and an exponential (or minus an exponential), so as $\lambda \to \infty$, the probability that $X < 0$ vanishes rapidly while the modal value is positive but very close to zero. For any fixed r and $\theta \geq 0$, the right tail of $f_{r,\theta}(x)$ behaves like the tail of a Gam(r, $1/(\pi/2 - \theta)$) distribution.

Other properties of the NEF-GHS distributions are developed in the next sections as part of the general NEF-QVF theory.

6. Infinite divisibility of NEF-QVF distributions. The binomial distribution is not infinitely divisible, for no bounded distribution can be, but the other five NEF-QVF distributions are (Feller, 1971). A new proof is given here for all five distributions simultaneously, revealing these five to be "self-generating," a term defined below.

Khinchin's characterization (Gnedenko and Kolmogorov, 1954) that a distribution possessing a MGF be infinitely divisible amounts to requiring its CGF $\psi(t)$ satisfy

$$\psi''(t)/\psi''(0) = E\exp(tY) \tag{6.1}$$

for some random variable Y. We show for NEF-QVF distributions that Y may be taken as $Y = V'(\bar{X})$, $\bar{X} = (X_1 + X_2)/2$ being the average of two iid elementary distributions (cf. Section 4) in the family. In this case, $V'(\bar{X})$ is linear (or constant), V being quadratic; and $X_1 + X_2$, being a convolution, belongs to a NEF.

A family infinitely divisible distributions will be called *self-generating* if the random variable Y in (6.1) is a convolution of members of the family.

THEOREM 1. NEF-QVF *distributions are infinitely divisible, provided* $v_2 \geq 0$. *The elementary* NEF-QVF *distributions are self-generating.*

PROOF. First consider the elementary distributions (cf. Section 4), so $v_2 = 0$ or $v_2 = 1$. Let F_0 be the CDF in (2.4) of the distribution with $\theta = 0$, and $\psi(t)$ its CGF. With $\mu = \psi'(\theta)$ we have for F_θ in (2.4),

$$d\{\log \psi''(\theta)\}/d\theta = V(\mu)\, d\{\log V(\mu)\}/d\mu = V'(\mu) = 2v_2\mu + v_1 = 2v_2\psi'(\theta) + v_1.$$

Thus, integrating wrt θ, using $\psi(0) = 0$ to determine the constant of integration, and exponentiating:

$$\psi''(\theta) = \psi''(0) \exp\{2v_2\psi(\theta) + v_1\theta\}. \tag{6.2}$$

Let X_1 and X_2 be iid as F_0. Then

$$Y = V'(\bar{X}) = v_2(X_1 + X_2) + v_1 \text{ has MGF } E\exp(tY) = \exp\{2\psi(v_2t) + v_1t\}.$$

This is (6.2) at $\theta = t$, because $v_2\psi(\theta) = \psi(v_2\theta)$ when $v_2 = 0$ or $v_2 = 1$. Hence F_0, an elementary CDF, is infinitely divisible. Then nonelementary distributions, with $v_2 \neq 1$, are also infinitely divisible, being convolutions of divisors of the elementary distribution. This completes the proof.

7. Cumulants and moments.

The following theorem provides a simple relation between moments and cumulants that replaces complicated formulae relating moments to cumulants (e.g., expressions given in Kendall and Stuart, 1963). It holds in general, not only for exponential families.

THEOREM 2. *Suppose X has m moments, $EX^r \equiv M'_r$, and cumulants C_r. Then for* $1 \leq r \leq m$,

$$M'_r = \sum_0^{r-1} \binom{r-1}{i} M'_i C_{r-i}. \tag{7.1}$$

Central moments $M_r = E(X - M'_1)^r$ satisfy

$$M_1 = 0, \qquad M_2 = C_2, \qquad M_3 = C_3$$

and for $r \geq 4$,

$$M_r = C_r + \sum_2^{r-2} \binom{r-1}{i} M_i C_{r-i}. \tag{7.2}$$

PROOF. First show, by induction, that the rth derivative is

$$\phi^{(r)}(t) = \sum_0^{r-1} \binom{r-1}{j} \phi^{(j)}(t)\psi^{(r-j)}(t)$$

if $\phi(t)$ is the chf of X and $\psi(t) = \log \phi(t)$. Evaluating this at $t = 0$ yields (7.1). Replacing X by $X - \mu$ before computing ϕ and ψ, in which case $\phi'(0) = \psi'(0) = 0$, and substituting $t = 0$ yields (7.2). □

The cumulants $C_r = C_r(\mu)$ of NEFs can be computed from $V = V(\mu)$ and its derivatives, as observed in (2.11). If the NEF has QVF then $(V')^2 = 4v_2V + d$ by (3.3) and $V'' = 2v_2$ is constant. It follows from (2.11) and (7.2) that, with QVF, the fourth cumulant and moment satisfies

$$C_4 = 6v_2V^2 + dv \quad \text{and} \quad M_4 = (3 + 6v_2)V^2 + dV. \tag{7.3}$$

TABLE 2
Values of c_{mi} in (7.4)

$m \backslash i$	0	1	2	3	4	5	6	7
0	1	0	0	0	0	0	0	0
1	1	1	0	0	0	0	0	0
2	1	5	1	0	0	0	0	0
3	1	21	14	1	0	0	0	0
4	1	85	147	30	1	0	0	0
5	1	341	1408	627	55	1	0	0
6	1	1365	13013	11440	2002	91	1	0
7	1	5461	118482	196053	61490	5278	140	1

Of course, $V^{(r)}(\mu) = 0$ for QVF for all $r \geq 3$, leading to polynomial dependence of all cumulants (and moments) on $V(\mu)$, as Theorem 3 will show.

DEFINITION 1. The NEF-QVF cumulant coefficients $c_{m,i}$ are defined for integers $m = 0, 1, 2, \cdots$ and $0 \leq i \leq$ m by $c_{m,0} = c_{m,m} = 1$ and for any $1 \leq i \leq m - 1$,

$$c_{m,i} = c_{m-1,i-1} + (i+1)^2 c_{m-1,i}. \tag{7.4}$$

Note that Pascal's triangle for the binomial coefficients is generated by similar recursive formulae, suppressing $(i+1)^2$ in (7.4). Table 2 displays enough $c_{m,i}$ values to generate the first 17 cumulants, using Theorem 3, and hence the first 17 moments, using Theorem 2.

THEOREM 3. *For* NEF-QVF *distributions, with* $m = 1, 2, \cdots$

$$\psi^{(2m)}(\theta) = V \textstyle\sum_0^{m-1} c_{m-1,i}(2i+1)! v_2^i d^{m-1-i} V^i \tag{7.5}$$

$$\psi^{(2m+1)}(\theta) = \frac{1}{2} VV' \textstyle\sum_0^{m-1} c_{m-1,i}(2i+2)! v_2^i d^{m-1-i} V^i. \tag{7.6}$$

PROOF. For any function $T = T(V)$ having derivative T' wrt V, we have

$$\frac{d}{d\theta} T(V) = \frac{d\mu}{d\theta} \frac{dV}{d\mu} \frac{dT(V)}{dV} = VV'T'(V), \tag{7.7}$$

$$\frac{d}{d\theta} \{V'T(V)\} = V\{V''T(V) + (V')^2 T'(V)\} = 2v_2 VT(V) + V(4v_2 V + d)T'(V). \tag{7.8}$$

Now (7.5) clearly holds for $m = 1$, $\psi^{(2)}(\theta) = V$. Assume it holds for m and let $T(V)$ be the r.h.s. of (7.5), a polynomial of degree m in V. Then application of (7.7) to (7.5) term-by-term gives (7.6) immediately. Next, differentiate (7.6) term-by-term using (7.8) with $T(V) = V^{i+1}$ to get

$$\frac{1}{2} \textstyle\sum_0^{m-1} c_{m-1,i}(2i+2)! v_2^i d^{m-1-i}\{2v_2 V^{i+2} + (4v_2 V + d)(i+1)V^{i+1}\}.$$

Rearranging terms to get the coefficients of V^{i+1}, and using (7.4), yields (7.5). The proof follows by induction on m. □

Theorem 3 shows that for NEF-QVF distributions, the rth cumulant, and also the rth moment (using Theorems 2 and 3), is a polynomial of degree at most r in μ. It also shows that when $r = 2m$, the cumulants and moments are polynomials of degree m in $V = V(\mu)$. For r odd, an extra (linear) factor $V'(\mu)$ is required. Moreover, C_r^2 is a polynomial of degree r in V even when r is odd, because the square C_r^2 defined by (7.6) involves the product of $(V')^2 = 4v_2 V + d$ and a polynomial of degree $r - 1$.

The first eight NEF-QVF cumulants written out from Theorem 3 are

$$C_2 = V, \quad C_3 = V'V, \quad C_4 = 6v_2 V^2 + dV,$$
$$(7.9) \qquad C_5 = V'V(12v_2 V + d), \quad C_6 = 120v_2^2 V^3 + 30v_2\, dV^2 + d^2 V,$$
$$C_7 = V'V(360v_2^2 V^2 + 60v_2\, dV + d^2),$$
$$C_8 = 5040v_2^3 V^4 + 1680v_2^2\, dV^3 + 126v_2\, d^2V^2 + d^3V.$$

Using (7.1) and (7.2) with this information yields the central moments

$$M_2 = V, \quad M_3 = V'V, \quad M_4 = (3 + 6v_2)V^2 + dV,$$
$$(7.10) \qquad M_5 = C_5 + 10C_2C_3 = V'V((10 + 12v_2)V + d),$$
$$M_6 = (15 + 130v_2 + 120v_2^2)V^3 + (25 + 30v_2)\, dV^2 + d^2V.$$

The coefficients of skewness $\gamma_3 \equiv C_3/V^{1.5}$ and kurtosis $\gamma_4 \equiv C_4/V^2$ are

$$(7.11) \qquad \gamma_3 = V'V^{-1/2}, \qquad \gamma_3^2 = 4v_2 + d/V, \qquad \gamma_4 = 6v_2 + d/V = 2v_2 + \gamma_3^2$$

for NEF-QVF distributions.

The normal is the only symmetric NEF, with or without QVF, because the skewness, $V'(\mu)V^{-1/2}(\mu)$, vanishes for all μ only if $V(\mu)$ is constant. Thus, no symmetric distribution other than the normal (including the GHS, logistic, etc.) can generate a symmetric NEF-QVF family. This follows easily from (2.11).

In cases with $v_2 d = 0$, i.e., the normal ($v_2 = d = 0$), the Poisson ($v_2 = 0$), and the gamma ($d = 0$), (7.5) and (7.6) are well-known cumulant expressions. However, these expressions do unify results for the six distributions, and in the binomial ($v_2 = -1/r$, $d = 1$), negative binomial ($v_2 = 1/r$, $d = 1$), and NEF-GHS ($v_2 = 1/r$, $d = -4$) cases, Theorem 3 provides useful new expressions for the cumulants.

8. Orthogonal polynomials for NEF-QVF distributions. Let $f(x, \theta)$ be a NEF-QVF density proportional to $\exp(x\theta - \psi(\theta))$ relative to some measure as in (2.2). Define

$$(8.1) \qquad P_m(x, \mu) = V^m(\mu)\left\{\frac{d^m}{d\mu^m} f(x, \theta)\right\}/f(x, \theta)$$

for $m = 1, 2, \cdots$. Derivatives in (8.1) are taken with respect to the mean μ, not θ. We have

$$P_0(x, \mu) = 1, \quad P_1(x, \mu) = x - \mu, \quad P_2(x, \mu) = (x - \mu)^2 - V'(\mu)(x - \mu) - V(\mu),$$

and will show that $P_m(x, \mu)$ is a polynomial of degree m in both x and μ with leading term x^m, and that $\{P_m\}$ is a family of orthogonal polynomials.

From (8.1) it follows immediately that (with arguments suppressed),

$$(8.2) \qquad P_{m+1} = V^{m+1}f^{-1}\, d(P_m f V^{-m})/du = (P_1 - mV')P_m + VP'_m, \quad m \geq 1$$

with $P'_m \equiv \partial P_m(x, \mu)/\partial\mu$, $P^{(r)} \equiv \partial^r P_m(x, \mu)/\partial\mu^r$.

THEOREM 4. *The set $\{P_m(x, \mu) : m = 0, 1, \cdots\}$ is, for* NEF-QVF *families, an orthogonal system of polynomials with respect to $f(x, \theta) = \exp\{x\theta - \psi(\theta)\}$. $P_m(x, \mu)$ has exact degree m in both μ and x with leading term x^m. It is generated by*

$$(8.3) \qquad P_{m+1} = (P_1 - mV')P_m - m\{1 + (m - 1)v_2\}VP_{m-1}$$

for $m \geq 1$ with $P_0 = 1$, $P_1 = x - \mu$. Define $a_0 = 1$ and for $m \geq 1$,

$$(8.4) \qquad a_m \equiv m! \prod_{i=0}^{m-1} (1 + iv_2).$$

Then for $m \geq 1$, $r = 0, 1, \cdots, m$, the derivatives wrt μ *are*

$$(8.5) \qquad P_m^{(r)} = (-1)^r(a_m/a_{m-r})P_{m-r}.$$

Finally, $E_\theta P_m = 0$ for $m \geq 1$ and

$$(8.6) \qquad E_\theta P_m P_n = \delta_{mn} a_m V^m, \quad m, n \geq 0.$$

PROOF. Define $b_m \equiv (m+1)(1+mv_2)$. We start by proving (8.5) for $r=1$, $a_m/a_{m-1} = b_{m-1}$. Now $P'_m = -b_{m-1}P_{m-1}$ holds for $m=1$, so assume it holds for $m \geq 1$ and use it in (8.2) to write

$$P_{m+1} = (P_1 - mV')P_m - b_{m-1}VP_{m-1}.$$

Differentiating this wrt μ,

$$\begin{aligned} P'_{m+1} &= (-1 - mV'')P_m + (P_1 - mV')P'_m - b_{m-1}V'P_{m-1} - b_{m-1}VP'_{m-1} \\ &= -(1+2mv_2)P_m - b_{m-1}(P_1 - mV')P_{m-1} - b_{m-1}V'P_{m-1} - b_{m-1}VP'_{m-1} \\ &= -(1+2mv_2)P_m - b_{m-1}[(P_1 - (m-1)V')P_{m-1} + VP'_{m-1}] \\ &= -(1+2mv_2 + b_{m-1})P_m = -b_mP_m, \quad \text{(from (8.2))}. \end{aligned}$$

Induction on m now proves $P_m^{(1)} = -b_{m-1}P_{m-1}$. By iterating this $r-1$ more times,

$$P_m^{(r)} = (-1)^r(b_{m-1} \cdots b_{m-r})P_{m-r} = (-1)^r(a_m/a_{m-r})P_{m-r},$$

proving (8.5).

Equation (8.3) follows from (8.2) and (8.5) with $r=1$. That $P_m(x,\mu)$ is a polynomial of exact degree m in both x and μ with leading term x^m follows inductively from (8.3). Now for any $n < m$, using (8.1)

$$E_\theta X^n P_m = V^m \int x^n f^{(m)}(x,\theta)\, dF(x) = V^m \frac{\partial^m}{\partial \mu^m} E_\theta X^n = 0,$$

because Section 7 revealed $E_\theta X^n$ to be a polynomial of degree at most n in $\mu = \psi'(\theta)$. It follows that P_m is orthogonal to every polynomial of lower degree, and so (8.6) holds for $m \neq n$.

To prove (8.6) for $m = n \geq 1$, multiply (8.3) by P_{m-1} and take expectations to show $EP_1P_mP_{m-1} = b_{m-1}VEP_{m-1}$. Repeat this procedure with P_{m+1} to get $EP^2_{m+1} = EP_1P_mP_{m+1}$. Putting these together, $EP^2_m = b_{m-1}VEP^2_{m-1}$. Iterating,

$$EP^2_m = b_{m-1} \cdots b_0 V^m EP_0 = a_m V^m. \qquad \square$$

The polynomials of this section are known individually as the Hermite (normal distribution), Poisson-Charlier (Poisson distribution), Generalized Laguerre (gamma), Krawtchouk (binomial), Meixner (negative binomial), and Pollaczek (GHS) polynomials (symmetric GHS subfamilies only, not the NEF-GHS), (Szego, 1975). The main new results here lie in unifying these six polynomial systems and in the new polynomials and results provided by (8.3) through (8.6) for nonsymmetric members of the NEF-GHS family.

The system (8.1) forms a set of polynomials of the indicated degree for any natural exponential family. However $EP_2P_3 = V'''(\mu)V^3(\mu)$, for example, which is not zero unless $V(\mu)$ is quadratic. Thus, these polynomials form an orthogonal system only if QVF holds.

Finally, two useful facts follow from Theorem 4. The first (8.7) relates the orthogonal polynomials of different members in the exponential family. The second (8.8) provides the expectation of a polynomial defined with the "wrong" parameter.

COROLLARY 1. *Let $\mu, \mu_0 \in \Omega$ be given for two distributions in the same* NEF-QVF. *Then for any $m = 0, 1, \cdots$,*

$$P_m(x, \mu_0) = a_m \textstyle\sum_0^m \dfrac{(\mu - \mu_0)^{m-r}}{(m-r)!} \dfrac{1}{a_r} P_r(x, \mu) \tag{8.7}$$

with a_m defined by (8.4). *Hence*

$$E_\mu P_m(X, \mu_0) = \frac{a_m}{m!}(\mu - \mu_0)^m. \tag{8.8}$$

PROOF. Expand $P_m(x, \mu_0)$ in a Taylor series (of order m only) around μ and use (8.5) to get

$$P_m(x, \mu_0) = \sum_0^m \frac{(\mu_0 - \mu)^r}{r!} P_m^{(r)}(x, \mu) = \sum_0^m \frac{(\mu - \mu_0)^r}{r!} \frac{a_m}{a_{m-r}} P_{m-r}(x, \mu).$$

Interchanging r and $m - r$ yields (8.7). Then (8.8) follows because $E_\mu P_r(X, \mu) = \delta_{0r}$. □

9. A large deviation theorem. This section proves the sharpest possible large deviation theorem for the entire NEF-QVF class. First we have a lemma applying to any NEF, not just those with QVF.

LEMMA. *Let X have a* NEF *distribution. Then for all $t \geq 0$, writing $\sigma^2 = V(\mu)$,*

$$P\left(\frac{X - \mu}{\sigma} \geq t\right) \leq \exp(-B(t)), \qquad B(t) \equiv \sigma^2 \int_0^t \frac{(t - w)\, dw}{V(\mu + w\sigma)}. \tag{9.1}$$

PROOF. For simplicity, assume $\theta = 0$ so $\psi(0) = 0$, $\psi'(0) = \mu$, $\psi''(0) = \sigma^2$. For any $w \geq 0$, $P(X - \mu - t\sigma \geq 0) \leq E \exp\{w[X - \mu - t\sigma]\} = \exp\{\psi(w) - w(\mu + t\sigma)\}$. This is minimized at $w = w(t)$ satisfying $\psi'(w(t)) = \mu + t\sigma$. Thus $w(0) = 0$, since $\psi'(0) = \mu$. Define $B(t) = w(t)(\mu + t\sigma) - \psi(w(t))$ as the negative of the minimum value. Note $B(0) = 0$. Now $B'(t) = w(t)\sigma$, after simplification, so $B'(0) = 0$, and $B''(t) = w'(t)\sigma$. Differentiating $\psi'(w(t)) = \mu + t\sigma$ with respect to t, $\psi''(w(t))w'(t) = \sigma$. But $\psi''(w(t)) = V(\psi'(w(t)) = V(\mu + t\sigma)$. Since $\sigma^2 = V(\mu)$, $B''(t) = \sigma^2/V(\mu + t\sigma)$. Using Taylor's theorem with integral remainder for expansion about $t = 0$, and $B(0) = B'(0) = 0$, gives $B(t) = \int_0^t (t - w)B''(w)\, dw$. This is (9.1). □

With QVF distributions we can go further, for then in (9.1),

$$V(\mu + t\sigma)/\sigma^2 = \{V(\mu) + t\sigma V'(\mu) + v_2 t^2\sigma^2\}/\sigma^2 = 1 + \gamma t + v_2 t^2,$$

with $\gamma = V'(\mu)/\sigma$ the skewness coefficient of X. Suppose γ, $v_2 \leq C < \infty$, a constant, taking $C \geq 1$ for convenience. Then for all $t \geq 0$, $C \geq 1$, $0 < 1 + \gamma t + v_2 t^2 \leq 1.5C(1 + t^2)$ (consider cases $t \leq 1$ and $t \geq 1$ separately). Formula (9.1) for $B(t)$ gives

$$\begin{aligned} B(t) &= \int_0^t \frac{(t - w)\, dw}{1 + \gamma w + v_2 w^2} \geq \frac{1}{1.5C} \int_0^t \frac{(t - w)\, dw}{1 + w^2} \\ &= [t \cdot \arctan(t) - 0.5 \log(1 + t^2)]/1.5C \leq \pi t/3C. \end{aligned} \tag{9.2}$$

This behaves like $\pi t/3C$ for t large.

THEOREM 5. *Suppose γ, the skewness, and v_2 are bounded above by absolute constants as $t \to \infty$. If X has a* NEF-QVF *distribution, there exists an absolute constant $b > 0$ such that*

$$P\left(\frac{X - \mu}{\sigma} \geq t\right) \leq \exp(-bt) \tag{9.3}$$

for all $t \geq 1$. If $|\gamma|$ and v_2 are both bounded above, then there exists $b_0 > 0$ such that

$$P\left(\left|\frac{X - \mu}{\sigma}\right| \geq t\right) \leq \exp(-b_0 t) \tag{9.4}$$

for all $|t| \geq 1$.

PROOF. (9.3) follows from the discussion preceding the theorem, taking $b = \pi/3C$ *if* γ, $v_2 \leq C$. Then (9.4) follows from (9.3) by $X \to -X$ which sends $\gamma \to -\gamma$ and leaves v_2 unchanged. □

The skewness $\gamma^2 = 4v_2 + d/V(\mu)$ from (7.11), and is invariant under linear transformations $X \to (X - b)/c$. Thus Theorem 5 requires v_2 bounded above, and if $d > 0$, $V(\mu)$ bounded away from zero. This happens, for example if there exists $\varepsilon > 0$ fixed such that (in the notation of Table 1): $\lambda \geq \varepsilon$ (Poisson); $r \geq \varepsilon$ (gamma and NEF-GHS); $p(1-p) \geq \varepsilon$ (binomial); $p \geq \varepsilon$ and $r \geq \varepsilon$ (negative binomial); and no conditions (normal). These conditions are noted in one of the last lines of Table 1.

If $X_1, X_2, \cdots$ all have NEF-QVF distributions, $EX_i = \mu_i$, $\operatorname{Var} X_i = \sigma_i^2$, then

$$\lim_{k\to\infty} P(\max_{1\leq i\leq k} | X_i - \mu_i |/\sigma_i \geq t_k) = 0 \tag{9.5}$$

as $k \to \infty$ if $t_k \to \infty$ faster than $\log k$, and if v_{2i} and $| \gamma_i |$ are uniformly bounded. This follows from (9.4) because the l.h.s. of (9.5) is bounded by $\sum_{i=1}^{k} P(| X_i - \mu_i | \geq t_k\sigma_i)$. This fact is used in (Morris, 1967) to prove that certain tests in NEF-QVF distributions are ε-Bayes for large k.

The exponential bounds (9.3), (9.4) are the best possible for NEF-QVF distributions because the three elementary distributions with $v_2 = 1$ (the exponential, geometric, and hyperbolic secant) whose CDF's can be evaluated, have exponential tails exactly. Improvements are possible in the normal and Poisson cases, however. For example, in the normal case $B(t) = (1/2)t^2$ from (9.2) and insertion of this in (9.1) gives a sharper than exponential bound for large t.

10. Limits in distribution. If $V(\mu)$ takes an approximate form, $V^*(\mu)$, say, it is reasonable that the NEF corresponding to $V(\mu)$ should have approximately the distribution of the NEF corresponding to $V^*(\mu)$. That is, because $V(\mu)$ characterizes the NEF, nearness of V to V^* forces nearness of the characteristic functions. This also implies convergence of all moments and cumulants because they are functions of $V(\mu)$, via (7.5), (7.6) and (7.2), or through (2.9), the MGF.

For example, suppose $X \sim \operatorname{Bin}(r, p)$, $\mu = rp$, $V(\mu) = -\mu^2/r + \mu$. Then $V(\mu) \to V^*(\mu) \equiv \mu$ as $r \to \infty$ if $p \to 0$ and μ is fixed. Similarly $X \sim \operatorname{NB}(r, p)$, $\mu = rp/(1-p)$, and $V(\mu) = \mu^2/r + \mu \to V^*(\mu) = \mu$ as $r \to \infty$ and $p \to 0$, holding μ constant. Because $V^*(\mu) = \mu$ is the Poisson variance function, we thereby obtain the familiar Poisson limits for the binomial and negative binomial distributions as $r \to \infty$ with rp fixed.

Other distributional approximations hold. For $\operatorname{NB}(r, p)$ let $\mu = rp/(1-p)$ be large and r be fixed (i.e., $p \to 1$). Then $V(\mu) = \mu^2/r + \mu = (\mu^2/r)(1 + r/\mu) \sim V^*(\mu) \equiv \mu^2/r$ in ratio. Since V^* is the VF of $\operatorname{Gam}(r, \mu/r)$, we have $\operatorname{NB}(r, p) \doteq \operatorname{Gam}(r, 1/(1-p))$ for p near 1. For $r = 1$ this means Geometric $(p) \doteq$ Exponential $(1/(1-p))$ as $p \to 1$.

For similar reasons with r fixed, the NEF-GHS family with $\mu = r\lambda$, $V(\mu) = r + \mu^2/r$ has approximately a $\operatorname{Gam}(r, \mu/r) = \operatorname{Gam}(r, \lambda)$ distribution for μ (and λ) large (i.e. if $r/\mu = 1/\lambda \to 0$), for then $V(\mu) \sim V^*(\mu) = \mu^2/r$, the VF of $\operatorname{Gam}(r, \mu/r)$. This assertion appears at the end of Section 5.

The central limit theorem (CLT) and weak law of large numbers (WLLN) also follow for NEFs by showing the VF of the appropriately scaled $\sum_1^n X_i$ is approximately constant (CLT) or vanishes (WLLN) as $n \to \infty$. In other words, the expectation μ^* of $X^* = \sum (X_i - \alpha)/n^{1/2}$ is bounded if α satisfies $| \alpha - \mu | < C/n^{1/2}$, C a constant. Then $V^*(\mu^*) = \operatorname{Var}(X^*) = V(\alpha + \mu^*/n^{1/2}) \doteq V(\alpha)$ is nearly a constant function of μ^* for n large. Thus, X^* is asymptotically normal. Alternatively, if $X^* = \sum X_i/n$, then $\mu^* = EX^* = \mu$, and $V^*(\mu^*) = V(\mu)/n \to 0$ as $n \to \infty$. Because this limit is the variance of a constant, the WLLN is proved.

Convolutions and scalings of X map $X \to \sum_1^n X_i/c$ and $V \to V^*(\mu) = nV(c\mu/n)/c^2$ and so $v_i^* = v_i c^{i-2}/n^{1-i}$; cf. (3.4) and (3.5) with $b = 0$. Limit theorems for such transformations can never increase the complexity of V (because $v_i = 0 \Rightarrow v_i^* = 0$), or the order (degree) of V, if V is a polynomial. Thus the normal and Poisson distributions can be limits of the *strictly* quadratic NEF-QVF distributions, but not conversely. Interesting limit theorems make V^* *less* complex than V; i.e., some of the v_i^* vanish in the limit. The three limit distributions just considered (the Poisson, gamma, and normal) do this with $V^*(\mu)$ a monomial. Because such limits never increase the order of a polynomial V, the normal, with constant variance function, is the most widely reached nontrivial limit law. Of course the WLLN, with $V(\mu)$ vanishing, is even more widely achieved.

Within NEF-QVF families, the discriminant d changes to $d^* = d/c^2$ under convolution and linear transformations $X \to \sum (X_i - b)/c$. If $c \to \infty$, limits of these distributions have $d^* = 0$, i.e., must be normal or gamma limits. Otherwise limit distributions must preserve the sign of d. Hence the Poisson, binomial, and negative binomial distributions, all having $d > 0$, can be limits of one another. The NEF-GHS distributions, being the only NEF-QVF distributions with $d < 0$, cannot be limits of any other NEF-QVF distributions.

REFERENCES

ABRAMOWITZ, M., and STEGUN, I. A. (ed.) (1965). *Handbook of Mathematical Functions.* Dover, New York.

BARNDORFF-NIELSEN, O. (1978). *Information and Exponential Families in Statistical Theory.* Wiley, New York.

BARNDORFF-NIELSEN, O. and COX, D. R. (1979). Edgeworth and saddle-point approximations with statistical applications. *J. Roy. Statist. Soc. B* **41** No. 3, 279–312.

BATEN, W. D. (1934). The probability law for the sum of n independent variables, each subject to the law (1/2h) sech (πx/2h). *Bull. Amer. Math. Soc.* **40** 284–290.

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** No. 2, 703–705.

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. (1978). The geometry of exponential families. *Ann. Statist.* **6** 362–376.

FELLER, W. (1971). *An Introduction to Probability Theory and its Applications,* Vol. II., 2nd ed. Wiley, New York.

GNEDENKO, B. V., and KOLMOGOROV, A. N. (1954). *Limit Distributions for Sums of Independent Random Variables.* Addison-Wesley, Cambridge, Mass.

HARKNESS, W. L., and HARKNESS, M. L. (1968). Generalized hyperbolic secant distributions. *J. Amer. Statist. Assoc.* **63** 329–337.

JOHNSON, N. L., and KOTZ, S. (1970). *Continuous Univariate Distributions–2.* Houghton-Mifflin, Boston.

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

KHINCHIN, A. I. (1949). *Mathematical Foundations of Statistical Mechanics.* Dover, New York.

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

LEHMANN, E. (1959). *Testing Statistical Hypotheses.* Wiley, New York.

MORRIS, C. (1967). A class of epsilon Bayes tests of a simple null hypothesis on many parameters in exponential families. Technical Report No. 57, Department of Statistics, Stanford University.

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. American Mathematical Society, Colloquium Publications, Providence, R.I.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF TEXAS
AUSTIN, TX 78712