Unifying the Named Natural Exponential Families and their Relatives

Carl N. Morris and Kari F. Lock
Harvard University*

March 9, 2009

Abstract

Five of the six univariate natural exponential families (NEF) with quadratic variance functions (QVF), meaning their variances are at most quadratic functions of their means, are the Normal, Poisson, Gamma, Binomial, and Negative Binomial distributions. The sixth is the NEF-CHS, the NEF generated from convolved Hyperbolic Secant distributions. These six NEF-QVFs and their relatives are unified in this paper and in the main diagram, Figure 1, via arrows that connect NEFs with many other named distributions. Relatives include all of Pearson's families of conjugate distributions (e.g. Inverted Gamma, Beta, F, and Skewed-t), conjugate mixtures (including two Polya urn schemes), and conditional distributions (including Hypergeometrics and Negative Hypergeometrics). Limit laws that also relate these distributions are indicated by solid arrows in Figure 1.

Keywords: Normal, Poisson, Gamma, Binomial, Pearson families, quadratic variance functions

1 INTRODUCTION

Statisticians appreciate that the Normal, Poisson, Gamma, Binomial, and Negative Binomial distributions reach powerfully into every realm of theoretical and applied statistics. These distributions are five of the six natural exponential families (NEFs) that have quadratic variance functions (QVF), i.e. the variance is at most a quadratic function, $V(\mu)$, of the mean μ.

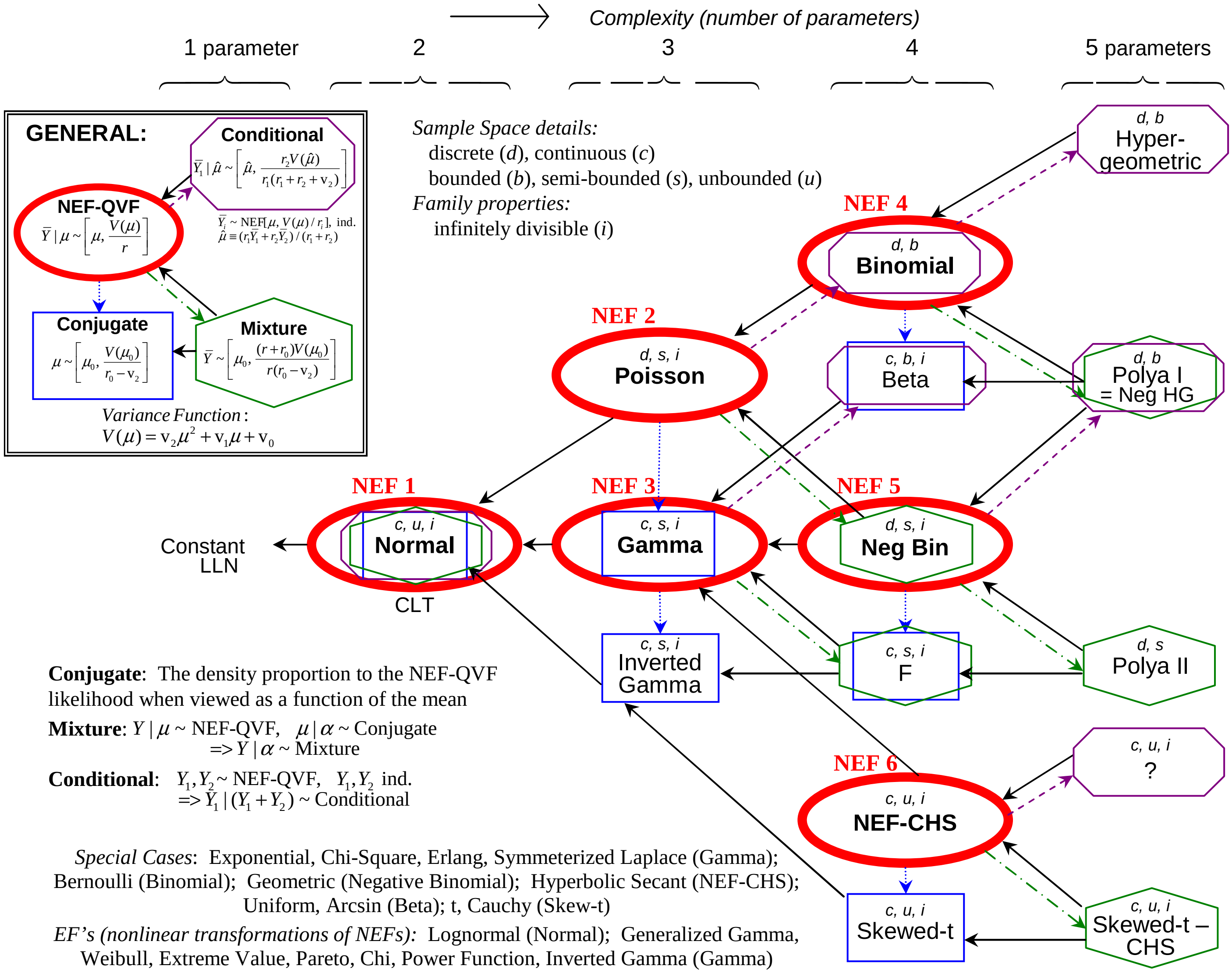
Figure 1 shows the six NEF-QVFs in red ellipses, with arrows that connect them to distributions via 3 different relationships. Section 2 introduces general

*Carl. N. Morris (morris@stat.harvard.edu) is Professor of Statistics, and Kari F. Lock (klock@fas.harvard.edu) is a doctoral student, both of the Department of Statistics, Harvard University, Cambridge, MA 02138. We are grateful for support from Harvard's Clark Fund. We thank Cindy Christiansen for her contributions to a 1988 version of Figure 1, and Cindy and Joe Blitzstein for their valuable discussions.

Natural Exponential Families with Quadratic Variance Functions (NEF-QVF), and their Relatives

Carl Morris &
Kari Lock, 2008

Figure 1: The six NEF-QVF distributions (in ellipses), each with its conjugate (rectangles), conjugate mixture (hexagons), and conditional (octagons). Limit laws are portrayed with solid arrows.



Conjugate: The density proportion to the NEF-QVF likelihood when viewed as a function of the mean

Mixture: $Y \mid \mu \sim$ NEF-QVF, $\mu \mid \alpha \sim$ Conjugate
$\Rightarrow Y \mid \alpha \sim$ Mixture

Conditional: $Y_1, Y_2 \sim$ NEF-QVF, Y_1, Y_2 ind.
$\Rightarrow Y_1 \mid (Y_1 + Y_2) \sim$ Conditional

Special Cases: Exponential, Chi-Square, Erlang, Symmeterized Laplace (Gamma); Bernoulli (Binomial); Geometric (Negative Binomial); Hyperbolic Secant (NEF-CHS); Uniform, Arcsin (Beta); t, Cauchy (Skew-t)

EF's (nonlinear transformations of NEFs): Lognormal (Normal); Generalized Gamma, Weibull, Extreme Value, Pareto, Chi, Power Function, Inverted Gamma (Gamma)

PARAMETERS: Location (L), Scale (S), Exponential Family Generation (G), Convolution (C), and Population Size (N)

Table 1: Key facts for the six NEF-QVF distributions. The mean, μ, variance function, $V(\mu)$, natural parameter, η, and cumulant function, ψ(η) are given for the elementary distributions ($r = 1$), Y_1. Convolutions (bottom half of the table) are sums of r iid elementary distributions.

	NORMAL	POISSON	GAMMA	BINOMIAL	NEG BIN	NEF-CHS
Generator	N(0,1)	Pois(1)	Expo(1)	Bern(1/2)	Geom(1/2)	HSec
Elementary, Y_1	**N(μ,1)**	**Pois(μ)**	**μExpo(1)**	**Bern(p)**	**Geom(p)**	**NEF-HS(μ)**
$\mu = EY_1 = EY$	μ	μ	μ	p	$p/(1-p)$	μ
$V(\mu) = Var(Y_1)$	1	μ	μ^2	$-\mu^2 + \mu$	$\mu^2 + \mu$	$\mu^2 + 1$
η	μ	$\log(\mu)$	$1 - (1/\mu)$	$\log(p/(1-p))$	$\log(2p)$	$\tan^{-1}(\mu)$
$\psi(\eta)$	$\eta^2/2$	$e^\eta - 1$	$-\log(1-\eta)$	$\log((1+e^\eta)/2)$	$-\log(2-e^\eta)$	$-\log(\cos(\eta))$
$\mathbf{Y \equiv \sum_{i=1}^{r} Y_i}$	**N($r\mu$, r)**	**Pois($r\mu$)**	**μ Gam(r)**	**Bin(r, p)**	**NBin(r, p)**	**NEF-CHS(r, μ)**
Support	$(-\infty, \infty)$	$\{0, 1, 2, \ldots\}$	$[0, \infty)$	$\{0, 1, \ldots, r\}$	$\{0, 1, 2, \ldots\}$	$(-\infty, \infty)$
Parameters	L,G: $\mu \in \mathbb{R}$ S,C: $r > 0$	C, G: $r\mu > 0$	S,G: $\mu > 0$ C: $r > 0$	G: $p \in [0, 1]$ C: $r > 0$	G: $p \in [0, 1]$ C: $r > 0$	G: $\mu \in \mathbb{R}$ C: $r > 0$

NEFs as a special subset of exponential families (EFs), and covers the NEF parameters. Section 3 introduces the variance function (VF), focusing on NEF-QVF distributions. Besides these famous five distributions, the sixth and only other NEF-QVF is the NEF generated by convolved (including infinite division) of Hyperbolic Secant (HS) distributions, labeled the "NEF-CHS" family, with "C" indicating convolution and division.

Four types of arrows, shapes, and colors in Fig. 1 summarize relationships among the six univariate NEF-QVFs and other related univariate distributions. Pearson's families of distributions, in blue rectangles in Fig. 1, placed directly below each NEF-QVF and connected by a dotted arrow, arise as conjugate (or prior) distributions for the six NEF-QVFs (Section 4). Conjugate mixtures (marginal distributions of NEFs) stemming from these Pearson conjugates are shown in green hexagons at the lower right of each NEF-QVF and connected by a dotted and dashed arrow (Section 5). The purple octagons to the upper right of each NEF-QVF and connected by a dashed arrow are the conditional distributions of one NEF, given the sum of two independent members of the same NEF (Section 6). Limits in distribution are displayed with leftward-pointing solid arrows ($\leftarrow$), achieved via simplified variance functions (Section 7).

This paper provides an overview of key ideas most of which are from Morris (1982, 1983), although there are some new realizations and results, plus Fig. 1 itself. Most proofs and many other results are left to those papers and to the references. Further details, technicalities, and proofs for the more probabilistic results presented (sections 2, 3, 7) can be found in Morris 1982, and further details for the more statistical results (sections 4, 5, 6) can be found in Morris 1983. These papers are available online along with a longer (unpublished) version of this paper at *http://www.stat.harvard.edu/People/Faculty/Carl_N._Morris/*.

A related diagram of Leemis and McQueston (2008) shows 76 named univariate distributions, and some of their relationships, including many of those in Fig. 1, plus others. While Figure 1 here includes many of their distributions (including by special cases or by transformations), our purpose is to reveal the unified structure that connects all six NEF-QVFs and their relatives.

The power of Figure 1 is its parallelism for the six NEF-QVFs. Figure 1 shows each NEF's distributional relatives via conjugacy (rectangles), mixtures (hexagons), and conditioning (octagons). Relatives are illustrated by different shapes, as summarized in the "General" box on Figure 1. This "General" pattern is repeated six times, once for each NEF-QVF. This paper's goal is to explain each diagrammatic shape and the connecting arrows so readers can use Figure 1 as a quick reference when working with these distributions, and to provide insight into their intertwined relationships. We hope readers will appreciate this powerful glimpse into the beautiful unification of these distributions we all encounter regularly and love to work with.

2 NATURAL EXPONENTIAL FAMILIES

2.1 Defining an NEF

Natural exponential families are a subclass of all exponential families. The distributions of a random variable X form a univariate *exponential family* (EF) if their densities or probability mass functions (PMF) have the form

$$\exp\{\ A(x)B(\theta) + C(x) + D(\theta)\}, \tag{2.1}$$

with θ as the parameter of interest. Special subclasses of these are the *natural exponential families* (NEF), where $A(\cdot)$ is linear. If $X \sim EF$, then $Y = A(X)$ is termed the *natural observation* and $Y \sim NEF$. Some authors have used the terminologies *linear* (Patil 1985), and *canonical* or *standard* (Brown 1986), as synonyms for *natural*.

To segue into NEF language, define η as the *natural parameter*, and $\psi(\cdot)$ as the *cumulant function*, where $\eta \equiv B(\theta)$ and $\psi(\eta) \equiv -D(\theta)$. Define $dF_0(x) \equiv d(e^{C(x)})$. Then the distribution of the *natural observation* $Y = A(X)$, (2.1) can be written in the general form for all NEFs:

$$P(Y \in B) = \int_B \exp\{\ \eta y - \psi(\eta)\ \}\ dF_0(y) = \int_B dF_\eta(y). \tag{2.2}$$

A univariate NEF is a parametric family of distributions with random variables Y satisfying (2.2). The natural parameter, η, lies in the *natural parameter space*, $\mathbf{H}$ (the Greek capital eta), a nondegenerate interval that contains 0 iff $F_0(\cdot)$ is a CDF. Note that if $F_0(\cdot)$ is a CDF, then $\psi(0) = 0$. Taking $B = (-\infty, y]$, (2.2) yields the family of CDFs, $F_\eta(y)$. If 0 is in $\mathbf{H}$, the CDF F_0 has moment generating function (MGF) $M_0(t) = \exp(\psi(t))$, and $\mathbf{H}$ is the interval on which $M_0(t)$ is finite. The MGF of the NEF for each η is $M_\eta(t) = \exp\{\ [\psi(\eta+t) - \psi(\eta)]\}$

In NEF terms, $\psi(\eta)$ is the *cumulant function* for the NEF (not to be confused with the *cumulant generating function*, $\log(M_\eta(t))$) because for NEFs the k^{th} cumulant of Y, C_k for $k = 1, 2, \ldots$, is

$$C_k = \psi^{(k)}(\eta) = \frac{d^k \psi(\eta)}{d\eta^k}. \tag{2.3}$$

The k^{th} derivative of the cumulant generating function evaluated at $t = 0$ equals the k^{th} derivative of the cumulant function, $\psi(\eta)$, justifying (2.3). The first two cumulants, the mean and variance, are

$$\mu \equiv EY = \psi'(\eta) \equiv C_1(\mu) \equiv C_1, \text{ and} \tag{2.4}$$

$$V(\mu) \equiv \mathrm{Var}(Y) = \psi''(\eta) \equiv C_2(\mu) \equiv C_2. \tag{2.5}$$

The mean, μ, lies in the mean space $\Omega \equiv \psi'(\mathbf{H})$, whose closure is the smallest interval that contains the sample space. The variance function $V(\mu)$ is central to NEF-QVF theory, see Section 3.

NEFs have several advantages over EFs. Derivatives of $\psi(\eta)$ yield cumulants of $Y = A(X)$, the natural observation, but not the moments of X (unless $A(X)$ is linear). Convolutions of NEFs extend the NEF family by this 2nd (convolution) parameter, and preserve the cumulant function up to a constant, but convolutions of EFs that are not also NEFs usually are complicated. Sufficient statistics for independent NEFs are sums of the natural observations $Y_i = A(X_i)$, and not of X_i.

To illustrate, consider $Y \sim N(\mu, \sigma^2)$ with σ known. In (2.2), $\eta = \mu/\sigma^2$, $\psi(\eta) = \sigma^2\eta^2/2$ and $dF_0(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/(2\sigma^2)} dy$, so Y belongs to an NEF. Differentiating the cumulant function, $\psi(\eta)$, yields the mean $\psi'(\eta) = \eta\sigma^2 = \mu$, and variance $\psi''(\eta) = \sigma^2$. Convolutions of Ys will still be Normal. Alternatively, when X is LogNormal, $X = \exp(Y)$, then X follows (2.1) with $A(x) = \log x$, so X lies in an EF but not in an NEF. X does not have a corresponding cumulant function (in fact it doesn't even have a moment generating function), and the density function of a convolution of even two LogNormals is intractable.

2.2 NEF Parameters

The definition of a one-dimensional NEF given in (2.2) may be extended to include three other (sometimes coincident) parameters in addition to the natural parameter, so that distributions also can involve a convolution parameter, a location parameter, and a scale parameter. However we still have a one parameter NEF, as these other three usually are considered to be fixed and known. We choose one "simplest" family by choosing the three additional parameters to take the simplest values for each NEF-QVF, naming this choice the *elementary family*: Normal$(\mu, 1)$, Poisson(μ), μ Exponential(1), Bernoulli(p), Geometric(p), and NEF-HS(μ). Starting with any of these elementary families, their convolutions (and divisions, whenever possible) add in the *convolution parameter*, $r > 0$. If r is an integer, r is the number of convolved elementary distributions. For example, NBin$(r,\, p)$ is the convolution (sum) of r i.i.d. Geom(p) distributions and Bin(r, p) is the convolution of r i.i.d Bern(p) distributions. When infinite division holds for the family, i.e. all NEF-QVFs other than Binomials, r can be any positive real number. Convolutions of members within an NEF then stay within the 1-parameter NEF, with a different value of r, provided the convolved distributions have the same natural parameter η.

The convolution parameter r is considered known throughout this paper, except for the Poisson case, when it is a function of the natural parameter, and thus it is convenient to separate r from η, the natural parameter, or equivalently from μ. Let $Y_i \overset{\text{iid}}{\sim}$ Elementary Family, $Y \equiv \sum_{i=1}^{r} Y_i$, and $\overline{Y} = Y/r$. In general, we choose to let μ and $V(\mu)$ pertain to Y_i. To keep the expectation as μ we choose to write densities in terms of $\overline{Y}$, in preference to Y, and then

$$dF_{\eta,r}(\overline{y}) = \exp\{r[\eta\overline{y} - \psi(\eta)]\} dF_{0,r}(\overline{y}). \tag{2.6}$$

When $r = 1$, this simplifies to (2.2). However, if (2.2) is more convenient, one may use (2.2) even if $r \neq 1$ by absorbing the known constant r into η, with $\psi(\eta)$

altered accordingly. Table 1 gives μ, $V(\mu)$, η, and $\psi(\eta)$ for each elementary distribution ($r = 1$).

NEF location and scale parameters arise via linear transformations. If $Y \sim NEF$, $Y^* \equiv a_0 \pm a_1 Y$ also belong to another NEF, as determined by the known location parameter $a_0 \in \mathbb{R}$ and known scale parameter $a_1 > 0$. This allows, for example, for Poisson or Binomial distributions with non-integer support.

Sometimes parameters play dual roles, as displayed in the last column of Table 1. For the Normal, the mean is both the natural parameter and the location parameter, while the standard deviation is the scale parameter while its square, the variance, is the inverted convolution parameter. Each column in Figure 1 is headed by the number of unique parameters, with successively larger values to the right.

3 NEF-QVFs

3.1 Variance Functions

The *variance function* (VF), $V(\mu)$, expresses the variance of a distribution in terms of its mean, μ. By (2.4) and (2.5),

$$\frac{d\mu}{d\eta} = \frac{d\psi'(\eta)}{d\eta} = \psi''(\eta) = Var(Y) > 0, \tag{3.1}$$

so μ is 1-1 in η, increasing monotonically making $Var(Y) = \psi''(\eta)$ be a function of μ.

Cumulants are computable recursively as functions of μ,

$$C_{k+1} = \frac{dC_k}{d\eta} = \frac{dC_k}{d\mu} \cdot \frac{d\mu}{d\eta} = C_k'(\mu)V(\mu). \tag{3.2}$$

Denoting V' as $dV(\mu)/d\mu$, cumulants are expressible in terms of V as $C_2 = V$, $C_3 = V' \cdot V$, $C_4 = V \cdot (V')^2 + V^2 \cdot V''$, and higher cumulants derived via (3.2). Cumulants are convertible to central moments, M_k for $k \geq 2$, by $M_2 = C_2$, $M_3 = C_3$, $M_4 = C_4 + 3C_2^2$, and for $k \geq 4$, $M_k = C_k + \sum_{i=2}^{k-2} \binom{k-1}{i} M_i C_{k-i}$. Therefore, the variance function and it's domain, Ω (the mean space, which is the convex closure of the sample space), completely determine the NEF via it's MGF (of course, no specific μ is determined). The VF characterizes an NEF uniquely among NEFs (it does not do so among EFs), which makes meaningful the notation $Y \sim NEF[\mu, V(\mu)]$, where square brackets always denote [mean, variance]. Together with its domain Ω, $V(\mu)$ uniquely specifies an NEF family.

3.2 The Six NEF-QVFs

Quadratic variance functions (QVFs), satisfy

$$V(\mu) = v_2\mu^2 + v_1\mu + v_0. \tag{3.3}$$

Precisely six NEF-QVFs exist, displayed as red ellipses in Figure 1. The VF for each elementary NEF-QVF is shown in Table 1, and is multiplied by r for convolutions and divided by r when we divide the convolution by r. Figure 1 indicates whether the support of each distribution is continuous or discrete, is bounded, semi-bounded (bounded only above or below), or unbounded, and whether the distribution is infinitely divisible or not. Table 1 provides further details.

The first five NEF-QVFs serve at the core of statistics, as sampling distributions, although the 6th is largely unknown. Each of the first five has a "story" that can identify its use. Normal distributions arise from the central limit theorem. Poisson distributions count the occurrence of rare events. Gammas arise as convolutions and divisions of Exponentials, which are memoryless. Binomials, Bin(r, p), count the number of successes in r i.i.d. Bernoulli(p) trials. Negative Binomials, NBin(r, p), being convolutions of r i.i.d. Geometrics, each count the successes before r failures in successive Bernoulli(p) trials.

The 6^{th} NEF-QVF, the NEF-CHS, arises from the Hyperbolic Secant (HS) distribution. The HS is a continuous, infinitely divisible distribution, $Y \in \mathbb{R}$, with a symmetric, bell-shaped density function, the hyperbolic secant, as its PDF $f(y) = 0.5 \cdot sech(\pi y/2) = 0.5/cosh(\pi y/2)$. It has a finite MGF, exponentially decaying tails, mean 0 and unit variance, and a distributional representation of $(2/\pi)log(|Cauchy|)$ (see Johnson and Kotz (1970); Morris (1982), and Manoukian and Nadeau (1988)). Convolutions and divisions of this distribution yield all the *Convolved Hyperbolic Secant* (CHS) distributions. The CHS historically has been called the Generalized Hyperbolic Secant (GHS), Harkness and Harkness (1968), however, "convolved" is more descriptive than "generalized", as the latter can take many meanings. Exponential family generation of the CHS produces the NEF-CHS (Morris 1982, Sec. 5). Statisticians lack convenient skewed sampling distributions with support on all reals, and the NEF-CHS provides one.

Numberings 1-5 of the columns of Fig. 1 index parameter complexity. These count the number of parameters (including location and scale) needed to identify a single distribution within each type. The six NEF-QVFs have two parameters (Normal), three parameters (Poisson and Gamma), or four parameters. Fig. 1 also recognizes sample space complexities, with discrete (d) distributions placed higher up, and continuous (c) distributions placed lower down. Boundedness complexities are: (b) bounded both above and below, e.g. Binomial; (s) semi-bounded (e.g. Gamma); and (u) unbounded in both directions (e.g. Normal). Bounded distributions lie closer to the upper right of Fig. 1, and unbounded distributions are toward the left or bottom. These three complexities help suggest a useful ordering for these six NEFs: NEF-1 Normal (constant VF); NEF-2 Poisson (linear VF); NEF-3 Gamma; NEF-4 Binomial; NEF-5 Negative Binomial, and NEF-6 NEF-CHS.

4 PEARSON CONJUGATES

4.1 Conjugate Families

Distributions conjugate to an NEF, "conjugate priors" in Bayesian terminology, take the form of the NEF's likelihood function, with parameters $r_0 > 0$ and $\mu_0 \in \Omega$ replacing the convolution parameter r and the sufficient statistic $\overline{y}$ in (2.6). If $r_0 > 0$ and μ_0 are such that the integral is finite with respect to Lebesgue measure on $\eta \in \mathbf{H}$, this is a density function on η

$$g_1(\eta)d\eta = K \exp\left(\, r_0[\eta\mu_0 - \psi(\eta)] \, \right) d\eta, \text{ so} \tag{4.1}$$

$$g(\mu)d\mu = K \exp\left(-r_0 \left[\int (\mu - \mu_0)/V(\mu) \; d\mu \right] \right) d\mu/V(\mu) \tag{4.2}$$

with $K = K(r_0, \mu_0)$ as the normalizing constant. The main interest here concerns these conjugate densities $g(\mu)$ as functions of the NEF mean μ.

For example, starting with $\overline{Y} \sim (1/r)Bin(r,p)$, the conjugate distribution replaces $\overline{y}$ with μ_0 (or Y with $r_0\mu_0$) and r with r_0, and uses the measure $dp/V(p)$, leading to the density $Kp^{r_0\mu_0}(1-p)^{r_0(1-\mu_0)}dp/p(1-p)$. This is a Beta density, so that $\mu = p \sim Beta(r_0\mu_0, r_0(1-\mu_0))$ arises as conjugate to Binomial NEFs.

Figure 1 exhibits the six conjugate families by blue rectangles directly below each NEF-QVF (follow the dotted arrows). No arrow is shown for the Normal, which is its own conjugate, so the Normal conjugate is earmarked by the rectangle inside the Normal's ellipse. Other conjugate families on μ are: Gammas for Poisson NEFs, Inverted Gammas (reciprocals of Gamma random variables) for Gamma NEFs, Betas for Binomials, F-distributions for Negative Binomials, and Skewed-t distributions (Skates 1993; Esch 2003) for the NEF-CHS. Symmetric Skewed-t conjugates are Student's t_n distributions, which include the Cauchy, t_1. We have renamed as "Skewed-t" here what was originally labeled Skew-t (e.g. Skates 1993) because the term "Skew-t" more recently has been taken to refer to certain other distributions, and not Pearson's original Type IV, Skew-t distributions.

Karl Pearson a century ago characterized all continuous distributions for which the derivative of the log density is a linear function divided by a quadratic function. These distributions became known as Pearson families (Johnson and Kotz, 1970; Kendall, Stuart, and Ord, 1987). Remarkably, the six NEF-QVF conjugates derived above correspond precisely to all of Pearson's families. Thus, to honor Pearson, we refer to these NEF-QVF conjugates as "Pearson Conjugates" (PC). Besides the Normal, Gammas are Pearson's type III, Inverted Gammas are Pearson's type V, Betas are Pearson's type I, Fs are Pearson's type VI, and Skewed-ts are Pearson's type IV. All other Pearson distributions are special cases of these six.

Pearson derived moment expressions for all his distributions. The means and variances of these Pearson conjugates, in our notation, are such that

$$\mu \sim PC\left[\mu_0, \frac{V(\mu_0)}{r_0 - v_2}\right]. \tag{4.3}$$

Remarkably, the conjugate distribution's VF in (4.3), with $V(\mu_0) = v_2\mu_0^2 + v_1\mu_0 + v_0$, is the *same* VF as that of the NEF. Of course μ may not exist, and $r_0 > v_2$ is required for the variance to exist. All finite higher Pearson moments also are expressible in terms of the VF.

4.2 Pearson Conjugates as Prior Distributions

The term "conjugate distribution" without "prior" is adopted here since the Pearson distributions are widely used in statistics, and because they arise as conjugate priors for the means of NEF-QVF distributions. Conjugate prior distributions give rise to posterior distributions in the same family, so with $\overline{y}$ distributed as (2.6),

$$\mu|\overline{y}, \mu_0, r_0 \sim PC\left[\mu_{\overline{y}}^* \equiv \frac{r_0\mu_0 + r\overline{y}}{r_0 + r}, \ \frac{V(\mu_{\overline{y}}^*)}{r + r_0 - v_2}\right]. \tag{4.4}$$

The PC here is in the same family as that in (4.3), so the posterior VF again agrees with that of the NEF. The conjugate prior also is convenient because the posterior mean, $\mu_{\overline{y}}^*$, is linear in $\overline{y}$, a result that characterizes conjugate distributions of all NEFs (Diaconis and Ylvisaker, 1979). A new and useful fact, easily proved, is that Jeffrey's prior for an NEF mean μ has posterior mean as a linear function of $\overline{y}$ if and only if the VF is quadratic. Thus the Jeffrey's posterior is easiest to work with (only) for NEF-QVF distributions.

Conjugate priors allow a range of choices of the mean and variance, via choosing μ_0 and r_0 in (4.3). If the data have NEF-QVF distributions, then among all distributions with a pre-specified prior mean and variance, Pearson conjugates achieve the minimax risk for estimating μ with squared error loss. Thus Pearson conjugates are the simple, safe, and robust choices among prior distributions with the same prior mean and variance. (Jackson et al, 1970; Morris, 1983, Theorem 5.5). Walter and Hamedani (1991), Consonni and Veronese (1992), and Diaconis, Khare, and Saloff-Coste (2008) have studied other aspects of these PC distributions from a unified perspective, as have Gutierrez-Pena and Smith (1997) in a multivariate setting.

5 CONJUGATE MIXTURES

A *Pearson conjugate mixture* refers to the marginal distribution of $\overline{Y}$, when $\overline{Y}|\mu \sim NEF[\mu, V(\mu)/r]$ and $\mu|\mu_0, r_0 \sim PC[\mu_0, V(\mu_0)/(r_0 - v_2)]$, as in (4.3) with $r_0 > v_2$. Conjugate mixtures are shown as green hexagons in Figure 1, each at the lower right of its associated NEF-QVF; follow the green dotted and dashed arrows.

The six NEF-QVFs, mixed with their Pearson conjugates, have "Pearson mixtures" (PM) as marginal distributions, with the VF determining the family. Then

$$\overline{Y} \sim PM\left[\mu_0, \frac{r + r_0}{r(r_0 - v_2)}V(\mu_0)\right], \tag{5.1}$$

where means and variances of $\overline{Y}$ follow directly from Adam's Law ($E(Y) = E[E(Y|X)]$) and from Eve's Law ($Var(Y) = E[Var(Y|X)] + Var[E(X|Y)]$). Once again, the PM inherits the VF of the originating NEF.

In NEF-1 through NEF-5 order, NEF-QVF conjugate mixtures are well-known distributions: Normals; Negative Binomials; F distributions; Polya I (or Beta-Binomial); and Polya II distributions (explained next). NEF-CHS mixtures, with μ having a Skewed-t, form a continuous unbounded five parameter family of distributions, labeled "Skewed-t - CHS" in Fig. 1.

Polya's urn schemes involve sampling from a binary urn, initially with B blue and W white balls, under double replacement sampling (Feller 1950). Polya I sampling, continues until a prescribed number r balls are drawn. Polya II sampling only continues until r blue balls have been drawn. The random variable Y in each case is the number of white balls drawn. $Bin(r,p)$ and $NBin(r,p)$ distributions mixed with $p \sim Beta(W,B)$ (so for the Negative Binomial, $\mu = p/(1-p) \sim \frac{B}{W}F_{2W,2B}$) produce these Polya I and Polya II distributions, respectively. More generally, W and B need not be integers. The marginal distribution includes a new *population* parameter $N = W + B$ that emerges to produce a 5^{th} parameter. As $N \to \infty$, these revert back to the four parameter Binomial and Negative Binomial.

6 CONDITIONAL DISTRIBUTIONS

Let $\overline{Y_i} \overset{\text{ind}}{\sim} NEF-QVF[\mu, V(\mu)/r_i]$, $i = 1,2$. Denote the uniformly minimum variance unbiased estimator (UMVUE) of μ as $\hat{\mu} \equiv (r_1\overline{Y}_1 + r_2\overline{Y}_2)/(r_1+r_2) \sim NEF-QVF[\mu, V(\mu)/(r_1+r_2)]$. The *conditional distribution* then has first two moments

$$\overline{Y}_1 |\ \hat{\mu} \ \sim \ \left[\ \hat{\mu},\ \ \left(\frac{r_2}{v_2+r_1+r_2}\right)\frac{V(\hat{\mu})}{r_1}\ \right]. \tag{6.1}$$

This can be proved simply and simultaneously for all NEF-QVF distributions by using the completeness and sufficiency of $\hat{\mu}$, (Morris 1983, Sec. 4). Again, conditional distributions inherited their VF $V(\cdot)$ from the NEF-QVF.

Figure 1 uses dashed arrows to locate these conditional distributions, in the purple octagons at the upper right of each NEF-QVF. The conditional distributions for Normals are Normals, conditional Poissons are Binomials, conditional Gammas are (multiples of) Betas, conditional Binomials are Hypergeometrics (e.g. for Fisher's Exact Test), and conditional Negative Binomials are Negative Hypergeometrics. Formula (6.1) provides the first two moments for the NEF-CHS, a distribution not yet named or studied.

In Figure 1, the Polya I and Negative Hypergeometric distributions coincide, under appropriate re-parameterizations. To see this, consider Polya's urn scheme. Let X be the number of white balls drawn from an urn starting with W white and B blue balls, drawing until r blue balls appear, so $X \sim \text{NegHG}(r, \frac{W}{W+B})$. Alternatively, consider an urn with r white and $B - r + 1$ blue balls, draw W balls from the urn with double replacement, and let Y be the number of white balls drawn, so $Y \sim$ Polya I. Then $Y \sim X$, i.e. $\Pr(X = a) = \Pr(Y = a)$ for all a.

7 LIMITS IN DISTRIBUTION

Limits of distributions (as parameter values converge toward a boundary of the parameter space) are indicated by solid arrows in Figure 1. Limits always reduce parameter complexity, so all arrows are left-directed. Many more limits exist, but Figure 1 reduces clutter by not showing transitive limits. E.g. Negative Binomials have Poisson limits, and Poissons have Normal limits, so Negative Binomials must have Normal limits, but that is not shown with a separate arrow.

7.1 NEF Convergence via Variance Functions

Because a VF characterizes an NEF family, convergence of NEF VFs implies convergence of distributions. For example, for $Bin(r, p)$, when $r \to \infty$ and $p \to 0$ with mean $\lambda \equiv rp$ held constant, the variance $rp(1-p) = \lambda(1-\lambda/r)$ asymptotes to λ. Since this limiting variance λ equals the mean for this NEF, the asymptotic distribution must be the NEF with that VF, i.e. Poisson. Likewise, a $NBin(r, p)$ with mean $\lambda = rp/(1-p)$ held constant as $r \to \infty$ and $p \to 0$, has variance $rp/(1-p)^2 = \lambda(1+\lambda/r) \to \lambda$. Hence, Binomials and Negative Binomials also have Poisson limits in distribution.

Negative Binomials and NEF-CHS distributions have Gamma limits when r stays fixed and $\mu \to \infty$. The mean of $\mathrm{NBin}(r, p)/r$ is $\mu = p/(1-p)$, and its variance is $(\mu^2+\mu)/r$, while the NEF-CHS has mean μ and variance $(\mu^2+1)/r$. For large μ, both of these VFs approximate (in ratio) μ^2/r, the VF of the Gamma family.

7.2 Convergence of NEF-QVF Relatives

Each conditional distribution (octagons in Figure 1) has its NEF-QVF (Figure 1 ellipses) as a limit. Let Y_1 and $Y_2 \sim$ NEF-QVF, with convolutions parameters r_1 and r_2 respectively. As $r_2 \to \infty$ with r_1 fixed, Y_2 dominates $Y_1 + Y_2$, so $Y_1|Y_1 + Y_2 \to Y_1|Y_2 \sim Y_1 \sim$ NEF-QVF.

As the convolution parameter for the conjugate (r_0) goes to infinity, the conjugate mixture distribution (hexagons on Figure 1) limits back to the NEF-QVF from which it came. If $r_0 \to \infty$, we essentially know μ exactly, so $\overline{Y} \to \overline{Y}|\mu \sim$ NEF-QVF. Alternatively, as the convolution parameter for the NEF-QVF (r) goes to infinity, the conjugate mixture distribution limits to the mean parameter's Pearson conjugate distribution. As $r \to \infty$, then $\bar{Y} \to \mu$ by the LLN, so $\bar{Y} \to \mu \sim PC$.

The remaining limit arrows in Figure 1 follow the "Four Color Rule" for NEF-QVFs. If one red ellipse (NEF-QVF) limits to another red ellipse, then each distributional relative (blue square, green hexagon, and purple octagon) also limits (leftward) to the corresponding color relative of the limiting NEF-QVF.

8 CONCLUSION

This paper has reviewed unifications of the six univariate one-parameter NEF-QVF families. Additional probabilistic results and proofs that unify NEF-QVFs are in Morris (1982, 1983, 1985). Many of their properties can be proved for all six of these families by using the quadratic nature of their VFs. These include proofs of probabilistic results for infinite divisibility, cumulant and moment formulae, orthogonal polynomials, and large deviation bounds (Morris 1982). Statistical results, also provable in a unified way include unbiased estimation, Bhattacharyya bounds (via orthogonal polynomials), and additional results for Pearson conjugate distributions (Morris 1983).

Infinitely many univariate NEFs exist, but almost none are named, besides those with QVFs. The most beautiful NEF with a non-quadratic VF has a cubic monomial VF, $V(\mu) \propto \mu^3$, this being the Tweedie's and Wald's Inverse Gaussian distribution (Shesadri 1993, Tweedie 1957). Letac and Mora (1990) identified all possible NEFs with cubic VFs, showing that there are exactly six that are not QVFs. See also Letac (1992).

The Multivariate Normal is the only multivariate NEF with a fully-parameterized covariance matrix. Patil (1985) and Brown (1986) adopted the term "Linear Exponential Family" (LEF) to refer to multivariate NEFs. The Multinomial distribution is a multivariate NEF (LEF) with a quadratic covariance matrix, but with a very restrictive covariance parameterization. Also see Bar-Lev et. al. (1994) on multivariate NEFs.

NEFs lead naturally to quasi-likelihood methods, as pioneered by Wedderburn (1974), who originated the term "variance function". The VF is central to quasi-likelihood methods and to generalized linear models (McCullagh and Nelder 1989).

The distributions and relationships in Figure 1 form the core for probability and statistics courses because five of the NEF-QVFs arise widely as sampling distributions for real data. One beautiful aspect of Figure 1 is that once one realizes that the relationship between a particular NEF-QVF and its relatives is understood (for example, Binomial, the Beta conjugate, the Polya I conjugate mixture (Beta-Binomial), and the Hypergeometric conditional distribution), one realizes that the other five NEF-QVFs enjoy parallel relationships. A student that learns how to do Bayesian inference with a Normal likelihood with Normal priors can use Figure 1 and the results summarized here to see how to make parallel inferences for Gamma likelihoods with their Inverted Gamma priors, or for any other NEF-QVF likelihoods with their conjugate distributions. Once one understands the conditional distribution for Binomials is Hypergeometric, then one also realizes from Figure 1 that the conditional distribution of Negative Binomials is Negative Hypergeometric. We offer Figure 1 here, believing that it will aid and deepen insights and understanding for students, for faculty, and for practitioners of probability and statistics.

References

[1] Bar-Lev, S.; Bshouty, D.; Enis, P.; Letac, G.; Lu, I-L.; Richards, D. (1994) The diagonal multivariate natural exponential families and their classification. J. Theoret. Probab. 7, 883−929.

[2] Barndorff-Nielson, O. (1978), *Information and Exponential Families in Statistical Theory,* New York:Wiley.

[3] Brown, L. D. (1986), *Fundamentals of Statistical Exponential Families,*, Hayward, California: Institute of Mathematical Statistics, Lecture Notes - Monograph Series, **9**.

[4] Consonni, G. and Veronese, P. (1992). "Conjugate priors for exponential families having quadratic variance functions", *Journal of the American Statistical Association*, 87, 11231127.

[5] Diaconis, P., and Ylvisaker, D. (1979), "Conjugate Priors for Exponential Families", *The Annals of Statistics*, 7, 269-281.

[6] Diaconis, P., Khare, K., Saloff-Coste, L. (2008). "Gibbs Sampling, Exponential Families and Orthogonal Polynomials," To appear in *Statistical Science*.

[7] Esch, D. (2003). "Applications and Extensions of Three Statistical Models," Ph.D. dissertation, Harvard University, May 22, 2003.

[8] Feller, W. (1950) *In Introduction to Probability Theory and its Applications,* (Vol 1), New York:Wiley.

[9] Gutierrez-Pena, E. and Smith, A. (1997). Exponential and Bayes in conjugate families: Review and extensions, Test 6, 190.

[10] Harkness, W. L., and Harkness, M. L. (1968), "Generalized Hyperbolic Secant Distributions," *Journal of the American Statistical Association*, 63, 329-337.

[11] 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.

[12] Johnson, N. L., and Kotz, S. (1970), *Continuous Univariate Distributions - 2*, Boston:Houghton-Mifflin.

[13] Kendall, M., Stuart, A., and Ord, K. (1987). *Kendall's Advanced Theory of Statistics : Volume 1 Distribution Theory* (5th ed.), New York:Oxford.

[14] Leemis, L.M. and McQueston, J. T. (2008). "Univariate Distribution Relationships," *The American Statistician*, 62, 1, 45−53.

[15] Letac, G. (1992). "Lectures on Natural Exponential Families and Their Variance Functions." *Monografias de matematica*, 50, I.M.P.A., Rio de Janeiro.

[16] Letac, G. and Mora, M. (1990). "Natural Real Exponential Families with Cubic Variance Functions," *The Annals of Statistics,* 18, 1, 1−37.

[17] Manoukian, E. B. and Nadeau, P. (1988). "A Note on the Hyperbolic-Secant Distribution," *The American Statistician*, 42, 1, 77−79.

[18] McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models* (2nd ed), Boca Raton:Chapman & Hall.

[19] Morris, C. N. (1982), "Natural Exponential Families with Quadratic Variance Functions," *The Annals of Statistics*, 10, 65−80.

[20] Morris, C. N. (1983), "Natural Exponential Families with Quadratic Variance Functions: Statistical Theory," *The Annals of Statistics*, 11, 515−529.

[21] Morris, C. N. (1985), "Natural Exponential Families," *Encyclopedia of Statistical Sciences* (Vol 6), ed. Kotz and Johnson, New York:Wiley, 157−159.

[22] Patil, G.P. (1985). "Linear Exponential Family", *Encyclopedia of Statistical Sciences* (Vol 5), ed. Kotz and Johnson, New York:Wiley, 22-24.

[23] Seshadri, V. (1993). *The Inverse Gaussian Distribution: A Case Study in Exponential Families"*, Oxford University Press.

[24] Skates, S. (1993). "On Secant Approximations to Cumulative Distribution Functions," *Biometrika*, 80, 1, 223−235.

[25] Tweedie, M. C. K. (1957). "Statistical Properties of Inverse Gaussian Distribution I, II" *The Annals of Mathematical Statistics,* 28, 2 362-377, 3 396−705.

[26] Walter, G. G. and Hamedani, G. G. (1991). "Bayes Empirical Bayes Estimation for Natural Exponential Families with Quadratic Variance Functions," *The Annals of Statistics,* 19, 3, 1191-1224.

[27] Wedderburn, R. M. W. (1974), "Quasi-likelihood Functions, Generalized Linear Models, and the Gauss-Newton Method," *Biometrika*, 61, 439-447.