CHAPTER 1. BASIC PROPERTIES

TANDARD EXPONENTIAL FAMILIES

.1 Definitions (Standard Exponential Family): Let ν be a σ-finite measure n the Borel subsets of R^k. Let

1) $$N = N_\nu = \{\theta : \int e^{\theta \cdot x} \nu(dx) < \infty\} \quad .$$

et

2) $$\lambda(\theta) = \int e^{\theta \cdot x} \nu(dx)$$

Define $\lambda(\theta) = \infty$ if the integral in (2) is infinite.) Let

$$\psi(\theta) = \log \lambda(\theta) \quad ,$$

nd define

3) $$p_\theta(x) = \exp(\theta \cdot x - \psi(\theta)) \, , \qquad \theta \in N$$

et $\Theta \subset N$. The family of probability densities

$$\{p_\theta : \theta \in \Theta\}$$

s called a k-dimensional *standard exponential family* (of probability ensities). The associated distributions

$$P_\theta(A) = \int_A p_\theta(x) \nu(dx) \quad , \qquad \theta \in \Theta$$

re also referred to as a standard exponential family (of probability istributions).

N is called the *natural parameter space*. ψ has many names. We will call it the *log Laplace transform* (of ν) or the *cumulant generating function*. $\theta \in \Theta$ is sometimes referred to as a *canonical parameter*, and

$x \in X$ is sometimes called a *canonical observation*, or value of a *canonical statistic*.

The family is called *full* if $\Theta = N$. It is called *regular* if N is open, i.e. if

$$N = N^\circ$$

where N° denotes the interior of N, defined as int. $N = \{\cup Q\colon Q \subset N, \ Q \text{ is open}\}$.

As customary, let the support of ν (supp ν) denote the minimal closed set $S \subset R^k$ for which $\nu(S^{\text{comp}}) = 0$. Let

$$H = \text{convex hull (supp } \nu) = \text{conhull (supp } \nu) \ . \tag{4}$$

and let $K = K_\nu = \bar{H}$. K is called the *convex support* of ν. (The convex hull of a set $S \in R^k$ is the set $\{y\colon \exists \ \{x_i\} \subset S, \ \{\alpha_i\}, \ 0 < \alpha_i, \ \Sigma\alpha_i = 1 \ni y = \Sigma\alpha_i x_i\}$.)

For $S \subset R^k$ the dimension of S, dim S, is the dimension of the linear space spanned by the set of vectors $\{(x_1 - x_2)\colon x_1, x_2 \in S\}$. A k-dimensional standard family is called *minimal* if

$$\dim N = \dim K = k \ . \tag{5}$$

Note that if K is compact then $N = R^k$, so that the family is regular.

(The exponential families described above can be called finite dimensional exponential families. Various writers have recently begun to investigate infinite dimensional generalizations. See Soler (1977), Mandelbaum (1983), and Lauritzen (1984) for some results and references.)

Standard exponential families abound in statistical applications. Often a reduction by sufficiency and reparametrization is, however, needed in order to recognize the standard exponential family hidden in specific settings. Here are two of the most fruitful examples.

1.2 Example (Normal samples): Let $Y_1,\ldots,Y_n$ be independent identically distributed normal variables with mean μ and variance σ^2. Thus, each Y_i has density (relative to Lebesgue measure)

(1) $$\phi_{\mu,\sigma^2}(y) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp(-(y-\mu)^2/2\sigma^2)$$

and cumulative distribution function Φ_{μ,σ^2} . Consider the statistics

$$\bar{Y} = n^{-1} \sum_{i=1}^{n} Y_i$$

$$S^2 = n^{-1} \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

$$X_1 = \bar{Y} , \quad X_2 = n^{-1} \sum_{i=1}^{n} Y_i^2 = S^2 + \bar{Y}^2 .$$

The joint density of $Y = Y_1,\ldots,Y_n$ can be written in two distinct revealing ways, as

(2) $$f_{\mu,\sigma^2}(y) = (2\pi\sigma^2)^{-n/2}\exp(-ns^2/2\sigma^2 - n(\bar{y}-\mu)^2/2\sigma^2) ,$$

or as

(3) $$f_{\mu,\sigma^2}(y) = (2\pi\sigma^2)^{-n/2}\exp((n\mu/\sigma^2)x_1 + (-n/2\sigma^2)x_2)\exp(-n\mu^2/2\sigma^2) .$$

From the first of these one sees that $\bar{Y}$ and S^2 are sufficient statistics. (One can also derive from this expression that $\bar{Y}$ and S^2 are independent (see sections 2.14 - 2.15) with $\bar{Y}$ being normal mean μ, variance σ^2/n and $V = S^2$ being $(\sigma^2/n)\cdot\chi^2_{n-1}$ -- i.e. having density

(4) $$f(v) = (n/2\sigma^2)^{m/2}(\Gamma(m/2))^{-1} v^{(m/2 -1)} \exp(-nv/2\sigma^2)\chi_{(0,\infty)}(v)$$

with $m = n-1$.)

$X = (X_1,X_2)$ is also sufficient. This can be seen from the factorization (3), or from the fact that X is a 1-1 function of $(\bar{Y},S^2)$. Let ν denote the marginal measure on R^2 corresponding to X -- i.e.

$$\nu(A) = \int_{(x_1,x_2)\in A} dy_1 \ldots dy_n .$$

(It can be checked that when $n \geq 2$, $\nu(dx) = (\frac{n}{2})^{n/2} (\pi^{\frac{1}{2}}\Gamma((n-1)/2))^{-1}(x_2-x_1^2)^{\frac{n-3}{2}} dx$ over the region $K = \{(x_1,x_2): x_1^2 \leq x_2\}$. When $n = 1$ ν is supported on the curve $\{(x_1,x_2): x_1^2 = x_2\}$.) Then the density of X relative to ν is

(5) $$p_{\theta_1,\theta_2}(x) = \exp(\theta_1 x_1 + \theta_2 x_2 - \psi(\theta))$$

with

$$\theta_1 = n\mu/\sigma^2 , \qquad \theta_2 = -n/2\sigma^2$$

and

$$\psi(\theta) = -\theta_1^2/4\theta_2 - (n/2)\log(-2\theta_2/n) .$$

Thus the distributions of the sufficient statistic form a 2 dimensional exponential family with canonical parameters (θ_1,θ_2) related to the original parameters as above.

This family is minimal. The natural parameter space is

$$N = \{(\theta_1, \theta_2) : \theta_1 \in R , \quad \theta_2 < 0\} .$$

The above can of course be generalized to multivariate normal distributions. See Example 1.14.

1.3 Example (Multinomial distribution):

Let $X = (X_1,\ldots,X_k)$ be multinomial (N,π) -- that is

$$\Pr\{X = x\} = \binom{N}{x_1,\ldots,x_k}\Pi\, \pi_i^{x_i} .$$

Let ν be the measure concentrated on the set $\{x : x_i$ integers, $x_i \geq 0$, $i=1,\ldots,k$, $\sum_{i=1}^{k} x_i = N\}$, and given by

(1) $$\nu(\{x\}) = \binom{N}{x_1,\ldots,x_k} = \frac{N!}{x_1!\ldots x_k!} .$$

Then the density of X relative to ν is

(2) $$p_\theta(x) = \exp(\sum_{i=1}^{k} \theta_i x_i - \psi(\theta))$$

where

(3) $$\theta_i = \log \pi_i \qquad i=1,\ldots,k$$

and

$$(4) \qquad \psi(\theta) \;=\; N \log\left(\sum_{i=1}^{k} e^{\theta_i}\right) \quad .$$

This is a k dimensional exponential family with canonical statistic X . Its canonical parameter is related to π by (3). It has parameter space

$$(5) \qquad \Theta \;=\; \{(\log \pi_i) : \; 0 < \pi_i, \; \Sigma\pi_i = 1\} \quad .$$

Note that this exponential family is not full. The full family has densities $\{p_\theta\}$ as above with $\Theta = N = R^k$. (For Θ as in (5) $\psi(\theta) \equiv 0$, however ψ as defined in (4), rather than $\psi \equiv 0$, is the appropriate cumulant generating function, as defined in 1.1(3) for the full family.) However

$$(6) \qquad p_\theta \;=\; p_{\theta+a1}$$

for all $a \in R$ where $1' = (1,\ldots,1)$. Hence expanding this family to be a full family does not introduce any new distributions.

The above phenomenon is related to the fact that the above family is not minimal since $\dim K = k-1 < k$. To reduce to a minimal family let $X^* \in R^{k-1}$ be given by $(X_1,\ldots,X_{k-1})$. Then X^* is sufficient. (In fact, it is essentially equivalent to X since $X_k = N - \sum_{i=1}^{k-1} X_i^*$ a.e.(ν) .) Let $\theta^* \in R^{k-1}$ be given by $\theta_i^* = \theta_i - \theta_k$, and let $\nu^*(\{x^*\}) = \binom{N}{x_1^*,\ldots,x_{k-1}^*,\,N-\sum_{i=1}^{k-1} x_i^*}$. Then the density of X^* relative to ν^* is

$$(7) \qquad p^*_{\theta^*}(x^*) \;=\; \exp\left(\sum_{i=1}^{k-1} \theta^*_1 x^*_1 - \psi^*(\theta^*)\right)$$

where

$$(8) \qquad \psi^*(\theta^*) \;=\; N \log\left(1 + \sum_{i=1}^{k-1} e^{\theta_i^*}\right) \quad .$$

This is a full minimal standard exponential family with $N = R^{k-1}$.

Note that

$$
(9) \qquad \begin{aligned}
\pi_i &= \exp(\theta_i^*)/(1 + \Sigma\exp(\theta_i^*)) \qquad i=1,\ldots,k-1 \ , \\
\pi_k &= 1/(1 + \Sigma\exp(\theta_i^*)) \ .
\end{aligned}
$$

Here, each different $\theta^* \in R^{k-1} = N$ corresponds to a different distribution.

Reductions by reparametrization and sufficiency like those in the above examples are frequent in statistical applications. Together with proper choice of the dominating measure, ν, they lead to the representation of problems involving exponential families in terms of problems involving standard exponential families. This is formally explained in the next few paragraphs.

1.4 Definition:

Let $\{F_\omega : \omega \in \Omega\}$ be a family of distributions on a probability space $\mathcal{Y},\mathcal{B}$. Suppose $F_\omega \ll \mu$, $\omega \in \Omega$. Suppose there exist functions

$$
\begin{aligned}
&C : \Omega \to (0,\infty) \\
&R : \Omega \to R^k \\
&T : \mathcal{Y} \to R^k \qquad \text{(Borel measurable)} \\
&h : \mathcal{Y} \to [0,\infty) \quad \text{(Borel measurable)}
\end{aligned}
$$

such that

$$
(1) \qquad f_\omega(y) = \frac{dF_\omega}{d\mu} = C(\omega)h(y)\exp(R(\omega) \cdot T(y)) \ .
$$

Then $\{F_\omega\}$ (or, $\{f_\omega\}$) is called a *k dimensional exponential family* of distributions (or, of densities).

1.5 Proposition:

Any k dimensional exponential family (1.4(1)) can be reduced by sufficiency, reparametrization, and proper choice of ν to a k dimensional standard exponential family (1.1(3)). The sufficient statistic is $X = T(Y)$, and its distributions form an exponential family with canonical parameter $\theta = R(\omega)$.

Proof: $X = T(Y)$ is sufficient by virtue of 1.4(1) and the Neyman factorization theorem. (See e.g. Lehmann (1959) Chapter 2 Theorem 8.) Let $\mu^*(dy) = h(y)dy$ and let $\nu(A) = \mu^*(T^{-1}(A))$ for Borel measurable sets $A \subset R^k$. Then the induced densities of X with respect to ν exist and have the desired form 1.1(3) with $\theta = R(\omega)$ and $\psi(\theta) = -\log C(R^{-1}(\theta))$. (Note that if $R(\omega_1) = R(\omega_2)$, then $f_{\omega_1} = f_{\omega_2}$ and hence $C(\omega_1) = C(\omega_2)$.) ||

In spite of appearances the above reduction process is not really unique. Any standard exponential family can be transformed to a different, but equivalent, form by linearly transforming X and Θ with linked nonsingular affine transformations. This is described in the following proposition.

1.6 Proposition:

Let $\{p_\theta\}$ be a k-dimensional standard exponential family. Let M be a non-singular $k \times k$ matrix and let

$$Z = MX + z_0$$

(1)

$$\phi = (M')^{-1}\theta + \phi_0 .$$

Then the distributions of Z also form a k-dimensional standard exponential family which is equivalent to the original family.

Proof: The equivalency assertion is immediate since the transformations (1) are 1-1. Furthermore, the density of Z relative to the measure ν_2 defined by $\nu_2(A) = \nu(M^{-1}(A - z_0))$ is

$$\exp(\theta' x(z) - \psi(\theta))$$

(2)

$$= \exp((\phi - \phi_0)' MM^{-1}(z - z_0) - \psi(M'(\phi - \phi_0)))$$

$$= \exp(\phi' z - \psi(M'(\phi - \phi_0)) + \phi' z_0 - \phi_0' z + \phi_0 \cdot z_0).$$

(By definition $A - z_0 = \{x : \exists\, z \in A,\ \ x = z - z_0\}$.)

Let $\nu_1(dz) = \exp(-\phi_0' z)\nu_2(dz)$ and $\psi_1(\phi) = \psi(M'(\phi - \phi_0)) - \phi' z_0 + \phi_0 \cdot z_0$. The densities of Z relative to ν_1 are

(3) $$\exp\{\phi' z - \psi_1(\phi)\} ,$$

which, as claimed, form a k parameter exponential family. The natural parameter space for this family is $M'^{-1}\Theta + \phi_0$ and the cumulant generating function is ψ_1 . ||

Proposition 1.6 shows that one may apply an arbitrary affine transformation either to Θ or to X. In this way one may assume without loss of generality that Θ (or X) lies in a convenient position in R^k . One application of this process will be discussed at some length in Section 3.11, and such transformations will be used wherever convenient.

MARGINAL DISTRIBUTIONS

The proof of Proposition 1.6 yields a statement about marginal distributions generated under linear projections by standard exponential families. The result is important in its own right, and useful in the proof of Theorem 1.8, as well.

Some preliminary remarks will be helpful. Let $\mathcal{M}_1 : R^k \underset{\text{onto}}{\rightarrow} R^m$ be a linear map. $\mathcal{M}_1$ is represented by an (m×k) matrix, M_1, of rank m. There is then a linear map $\mathcal{M}_2 : R^k \underset{\text{onto}}{\rightarrow} R^{k-m}$ which is orthogonally complementary to $\mathcal{M}_1$ -- that is, the rows of the corresponding ((k-m)×k) matrix, M_2, of rank (k-m) are orthogonal to those of M_1 . (The rows of M_2 can be chosen to be orthonormal, but that is not necessary here.) Let M denote the (k×k) nonsingular matrix $M = \binom{M_1}{M_2}$. If $x \in R^k$ then $Z = Mx$ can be written as $\binom{Z_1}{Z_2}$ with $Z_1 \in R^m$, $Z_2 \in R^{m-k}$.

Let $M = \binom{M_1}{M_2}$ as defined above. Then M^{-1} exists and can be written as

(1) $$M^{-1} = (M_1^-, M_2^-)$$

where M_1^- is (k×m), M_2^- is (k×(k-m)) and

(2) $$(M_1^-)' M_2^- = 0$$

since M_1 and M_2 are orthogonally complementary.

Let $\theta \in R^k$ and $\phi = (M^{-1})'\theta = \begin{pmatrix} (M_1^-)' \\ (M_2^-)' \end{pmatrix}\theta = \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}$. Then

(3) $$\theta' x = \theta' M^{-1} M x$$

$$= \phi_1' z_1 + \phi_2' z_2$$

by (2). For typographical reasons let $M_i'^- = (M_i^-)'$.

The special case where $M_1(x_1,\ldots,x_k) = (x_1,\ldots,x_m)$ is worth noting. Here

(4) $$M_1 = (I_{m\times m},\ 0_{m\times(k-m)}) = M_1'^-$$

$$M_2 = (0_{(k-m)\times m},\ I_{(k-m)\times(k-m)}) = M_2'^-$$

Somewhat more generally, if the rows of M_1 and M_2 are orthonormal then

(4') $$M_1 = M_1'^-$$
$$M_2 = M_2'^-$$

<u>1.7 Theorem</u>:

Consider a standard exponential family. Let $M_1 : R^k \underset{\text{onto}}{\to} R^m$ and $\theta = M'\begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}$ as described above. Fix $\phi_2^0 \in M_2'^-(N) \subset R^{k-m}$. Consider the family of distributions of $Z_1 = M_1 X$ over the parameter space $\Phi_{\phi_2^0} = M_1'^-(\{\theta \in \Theta :\ M_2'^-\theta = \phi_2^0\})$. These form an m dimensional standard exponential family generated by the marginal measure defined by

(5) $$\nu_{\phi_2^0}(A) = \int_{M_1^{-1}(A)} \exp(\phi_2^{0\prime} M_2 x)\nu(dx) .$$

The natural parameter space for this family is $N_{\phi_2^0} = M_1'^-(\{\theta \in N :\ M_2'^-\theta = \phi_2^0\})$.

The statistic Z_1 is sufficient for the family of densities $\{p_\theta(x) :\ M_2'^-\theta = \phi_2^0\}$.

Proof: A direct proof is as easy as an appeal to Proposition 1.6. The density of Z relative to the appropriate dominating measure $\nu(M^{-1}\,\cdot)$ is

(6) $$\exp(\theta' x - \psi(\theta)) \;=\; \exp(\phi_1 \cdot z_1 + \phi_2 \cdot z_2 - \psi(M'\phi))\ .$$

When $\phi_2 = \phi_2^0$ the factor $\exp(\phi_2^0 \cdot z_2)$ can be absorbed into the dominating measure, yielding $\nu_{\phi_2^0}(\cdot)$ as defined in (5). The resulting family of densities is the standard exponential family claimed in the statement of the theorem. (Note that (6) also provides a formula for the cumulant generating function of this family.)

The assertions concerning $N_{\phi_2^0}$ and sufficiency follow from (6), with $\phi_2 = \phi_2^0$, and the Neyman factorization theorem. ||

For the special case where M_1 is as described in (4), one sees that for fixed $\theta_{k+1},\ldots,\theta_m$ the distributions of $Z_1 = (X_1,\ldots,X_k)$ form an exponential family.

Note that the theorem does not say that the family of distributions of $Z_1 = M_1X$ form a standard exponential family with natural parameter ϕ_1 if the parameter θ ranges over *all* of Θ . In fact such a claim is generally false unless Θ is of dimension $\leq$m and satisfies

(7) $$\Theta \subset \{\theta : M_2^{\prime -}\theta = \phi_2^0\} \qquad \text{for some } \phi_2^0 \in R^{k-m}\ ,$$

as will be the case in Theorem 1.9; or

(8) $$Z_1 \text{ and } Z_2 \text{ are independent for some } \theta \in \Theta\ .$$

(It will be seen in the next chapter that (8) implies independence of Z_1 and Z_2 for all $\theta \in \Theta$.)

(8) Remark. The preceding theorem may be given an alternative interpretation. Let L be a linear variety in R^k -- that is $L = x_0 + V$ for some m dimensional linear subspace $V \subset R^k$. Let $P : R^k \to L$ be any affine projection onto L -- that is, P is affine, $P^2 = P$, and P is the identity on L. Let Q denote the orthogonal projection onto $V^{\perp} = \{w \in R^k :\ v'w = 0\ \ \forall\ \ v \in V\}$. Let

$\theta_{(2)} \in V^{\perp}$. Then the family of distributions of P(X) as θ ranges over $\{\theta \in N : \; Q\theta = \theta_{(2)}\}$ forms an exponential family.

To verify the above, note that there are linear isometries

$$S_1 : \; R^m \underset{\text{onto}}{\overset{\rightarrow}{1\text{-}1}} L \qquad\qquad S_2 : \; R^{k-m} \underset{\text{onto}}{\overset{\rightarrow}{1\text{-}1}} V^{\perp} \; .$$

The theorem applies to the maps $M_1 = S_1^{-1} \circ P$, $M_2 = S_2^{-1} \circ Q$, and yields a statement concerning the distributions of $M_1(X)$. This converts directly to the above statement about the distributions of $P(X) = S_1(M_1(X))$ over the appropriate parameter space since S_1 is a linear isometry, and S_1 and S_2 are orthogonal, etc.

1.8 EXAMPLE (Log-linear models): Consider a multinomial (N, π) variable as described in Example 1.3. Consider the family of distributions for which the natural parameter 1.3(3) satisfies

(1) $$\theta = B\beta + \theta_0 , \qquad \beta \in R^m$$

where B is a specified k×m matrix of rank m. Assume, in addition, that

(2) $$B = (1_k, B_{(2)})$$

where $1_k' = (1,\ldots,1)$ and $B_{(2)}$ is $k \times (m-1)$ of rank $(m-1)$. This is a *log-linear multinomial model*. The name derives from the fact that the linear constraint (1) can also be written in the form $\log \pi = B\beta$ where $(\log \pi)_i = \log \pi_i$, $i=1,\ldots,k$. Condition (2) is imposed because $P_\theta = P_{\theta+a1}$, as noted in 1.3(6). Because of (2) for every $\beta_{(2)}' = (\beta_2,\ldots,\beta_m)$ there is a unique $\beta_1 = \beta_1(\beta_{(2)})$ such that

(3) $$\sum_{i=1}^{k} \pi_i = \sum_{i=1}^{k} e^{\theta_i} = 1 .$$

Let $M_1 = B'$ and let $M = \binom{M_1}{M_2}$ as in 1.7. Theorem 1.7 yields that $Z_{(1)} = M_1X = B'X$ is a sufficient statistic. The distributions of $Z_{(1)}$ form an m-dimensional exponential family with corresponding natural parameter

$M_1'^{-}\theta = \beta + B^{-}\theta_0$. This family is not minimal since $(Z_{(1)})_1 = N$ w.p. 1 . As in Example 1.3 one may reduce to an equivalent minimal family having dimension $(m-1)$ and canonical statistic $Z^*_{(1)} = B'_{(2)}X = (Z_{(1),2},\ldots,Z_{(1),m})'$.

Here is a famous log-linear model arising in genetics. Suppose a parent population contains alleles G,g at a certain locus, with frequency $p,q = 1-p$, respectively. Under the assumptions of random mating and no selection a generation of N offspring will have genotypes GG, Gg, gg according to a multinomial distribution with π given by

$$\pi_1 = p^2 , \qquad \pi_2 = 2pq , \qquad \pi_3 = q^2 . \tag{4}$$

Such a multinomial distribution is called a *Hardy-Weinberg distribution.* This corresponds to a log-linear model with

$$B = \begin{pmatrix} 1 & 2 \\ 1 & 1 \\ 1 & 0 \end{pmatrix} \qquad \theta_0 = \begin{pmatrix} 0 \\ \log 2 \\ 0 \end{pmatrix} . \tag{5}$$

Thus, $z_{(1)} = \binom{N}{2x_1 + x_2}$ is a sufficient statistic for the distributions of this log-linear family, and $z^*_{(1)} = 2x_1 + x_2$ is a minimal sufficient statistic.

(This log-linear family can be imbedded in a useful way in the original multinomial family as follows:

Let

$$M = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 1 & 2 \\ -1/3 & 2/3 & -1/3 \end{pmatrix} = \binom{M_1}{M_2} . \tag{6}$$

Then

$$M^{-1} = \begin{pmatrix} 5/12 & -1/12 & -1/2 \\ 1/6 & 1/6 & 1 \\ -1/12 & 5/12 & -1/2 \end{pmatrix} = (M_1^{-}, M_2^{-}) .$$

Let $\phi_0' = (0,0,-\ln 2)$ and $z_0' = (0,0,\frac{N}{3})$. According to Proposition 1.6 $Z = MX + z_0$ is the canonical statistic for an exponential family with corresponding canonical parameter $\phi = (M^{-1})'\theta + \phi_0$. In terms of the original variables $z_1 = 2x_1 + x_2$, $z_2 = 2x_3 + x_2$, $z_3 = x_2$, and $\phi_3 = (\frac{1}{2})\log(\pi_2^2/4\pi_1\pi_3)$, etc. The log-linear family described above is therefore the family of marginal distributions of (z_1, z_2) under the restriction $\phi_3 = 0$. The family of distributions corresponding to the restriction $\phi_3 = \phi_3^0 \neq 0$ also has a natural genetic interpretation as the distribution of a population after variable selection of genotypes. See Barndorff-Nielsen (1978, p.123); the generalization of this model to a multiallelic locus is also described there.)

REDUCTION TO A MINIMAL FAMILY

Any exponential family which is not minimal can be reduced to a minimal standard family through sufficiency, reparametrization, and proper choice of ν. This involves only a minor extension of the process used above in Proposition 1.5 and Theorem 1.7. This reduction is unique up to the appearance of linked affine transformations as in Proposition 1.6. Here are the details.

1.9 Theorem

Any k dimensional exponential family can be reduced by sufficiency, reparametrization, and proper choice of ν to an m dimensional minimal standard exponential family, for some $m \leq k$. Let X,θ and Z,ϕ denote the canonical statistic and parameter for two such reductions to an m_1 and an m_2 dimensional minimal family, respectively. Then $m_1 = m_2$ and (X,θ), (Z,ϕ) are related as in 1.6(1).

Proof. The reduction to a minimal standard family will be performed in three steps. First, one may apply Proposition 1.5 to reduce to a standard k dimensional family.

Suppose for this family that $\dim \Theta = m' < k$. Thus $\Theta \subset \theta_0 + V$ where V is an m'-dimensional linear subspace. One may let P be the orthogonal projection on V and M_1, M_2 the corresponding orthonormal matrices described above in Theorem 1.7. Then $M_2\Theta = \phi_2^0$, a constant vector. By Theorem 1.7, $Z_1 = M_1X$ is sufficient, and its distributions form a standard exponential family, whose parameter space has dimension m'.

Thus it now suffices to consider the case of a standard m' dimensional exponential family whose parameter space also has dimension m' . Suppose for this family that dim $K = m < m'$. Then $K \subset x_0 + V$, similar to the previous situation. Let P be the orthogonal projection on V, and M_1, M_2 as above. Observe that

$$\text{(1)} \qquad \theta \cdot x = \theta' M_1' M_1 x + \theta' M_2' M_2 x$$

$$= \theta' M_1' M_1 x + \theta' M_2' M_2 x_0 \qquad \text{a.e.}\nu \quad .$$

It follows that $Z_1 = M_1 X$ is a sufficient statistic whose distributions form a standard exponential family with natural parameter $M_1\theta$. (Actually Z is not merely sufficient, but is actually equivalent to X under ν.) Since dim $(M_1 K)$ = dim $(M_1 \Theta) = m$ this family is the desired minimal family formed from the original family through reduction by sufficiency and reparametrization.

Suppose $\{p_\omega : \omega \in \Omega\}$ is a standard k dimensional exponential family relative to ν, and (X,θ), (Z,ϕ) denote the canonical statistics and parameters for two reductions of $\{p_\omega\}$ to a minimal standard exponential family. For the next step let $P_\theta^{(1)}$, $P_\phi^{(2)}$ denote their respective probability distributions with dimensions m_1 and m_2 respectively, etc.. Let $\omega_0 \in \Omega$. Since X and Z are each sufficient

$$\text{(4)} \qquad \frac{dP_\omega}{dP_{\omega_0}} = \frac{dP^{(1)}_{\theta(\omega)}}{dP^{(1)}_{\theta(\omega_0)}}(X(y)) = \frac{dP^{(2)}_{\phi(\omega)}}{dP^{(2)}_{\phi(\omega_0)}}(Z(y)) \qquad \text{a.e.}(\nu) \quad .$$

Now,

$$\frac{dP^{(1)}_{\theta(\omega)}}{dP^{(1)}_{\theta(\omega_0)}}(x) = \frac{p^{(1)}_{\theta(\omega)}(x)}{p^{(1)}_{\theta(\omega_0)}(x)}$$

$$= \exp(((\theta(\omega) - \theta(\omega_0)) \cdot x - (\psi^{(1)}(\theta(\omega)) - \psi^{(1)}(\theta(\omega_0))) \quad ;$$

and similarly for $P^{(2)}$. Hence (4) yields

(5) $$(\theta(\omega) - \theta(\omega_0)) \cdot x(y) - U^{(1)}(\theta(\omega))$$

$$= (\phi(\omega) - \phi(\omega_0)) \cdot z(y) - U^{(2)}(\phi(\omega)) \qquad \text{a.e. } (\nu)$$

for all $\omega \in \Omega$.

Suppose $m = m_1 < m_2$. Since $\dim \{\phi(\omega) : \omega \in \Omega\} = m_2 > m$ there exist values $\alpha_i \in R$, $\omega_i \in \Omega$, $i=1,\ldots,m+1$, such that $0 = \sum_{i=1}^{m+1} \alpha_i(\theta(\omega_i) - \theta(\omega_0))$ and $\phi^* = \sum_{i=1}^{m+1} \alpha_i(\phi(\omega_i) - \phi(\omega_0)) \neq 0$. It follows from (5) that

(6) $$\phi^* \cdot z(y) = \text{const} \qquad \text{a.e. } (\nu) \quad .$$

But, (6) implies $K_2 \subset \{z : \phi^* \cdot z = \text{const}\}$ so that $\dim K_2 < m_2$. This contradicts the fact that the distributions of Z form a minimal standard family of dimension m_2. Hence $m_1 = m_2 = m$.

Now choose $\omega_1,\ldots,\omega_m$ so that $\{\theta(\omega_i) - \theta(\omega_0) : i=1,\ldots,m\}$ span R^m. The preceding argument shows that $\{\phi(\omega_i) - \phi(\omega_0) : i=1,\ldots,m\}$ must also span R^m. Let M, non-singular, be chosen so that

$$\phi(\omega_i) - \phi(\omega_0) = (M')^{-1}(\theta(\omega_i) - \theta(\omega_0)) \qquad i=1,\ldots,m \quad .$$

Then, as in 1.6(3),

(7) $$(\theta(\omega_i) - \theta(\omega_0)) \cdot x(y) - U(\theta(\omega_i))$$

$$= (\phi(\omega_i) - \phi(\omega_0)) \cdot Mx(y) - U(\phi(\omega_i))$$

$$= (\phi(\omega_i) - \phi(\omega_0)) \cdot z(y) - U(\phi(\omega_i)) \qquad \text{a.e. } (\nu) \quad .$$

Let $y_0 \in K$ be a value for which (7) is valid for $i=1,\ldots,m$. Then (7) yields

$$(8)\qquad (\phi(\omega_i) - \phi(\omega_0)) \cdot M(x(y) - x(y_0))$$

$$= (\phi(\omega_i) - \phi(\omega_0)) \cdot (z(y) - z(y_0)) \quad \text{a.e. } (\nu) \quad i=1,\ldots,m \;.$$

This implies $M(x(y) - x(y_0)) = z(y) - z(y_0)$; which verifies 1.6(1) with $z_0 = z(y_0)$. ||

1.10 Definition

Let $\{p_\theta\}$ be a k-dimensional exponential family. Theorem 1.9 shows that there is a unique value, m, such that $\{p_\theta\}$ can be reduced to a minimal exponential family of dimension m. This value is called the *order* of the family p.

If $\{p_\theta\}$ is a standard family it is clear that its order m satisfies

$$(1)\qquad m \le \min(\dim \Theta, \dim K) \quad .$$

In most cases equality holds in (1); however, it is possible to have inequality, even when $\{p_\theta\}$ is full.

In view of Theorem 1.9 there is no loss of generality in confining oneself to the study of minimal standard exponential families. A full minimal standard exponential family is also called a *canonical exponential family*.

RANDOM SAMPLES

A nearly trivial but very important application of the first part of Theorem 1.9 involves independent identically distributed (i.i.d.) observations from an exponential family.

1.11 Theorem

Let $X_1,\ldots,X_n$ be i.i.d. observations from some k-dimensional standard exponential family with natural parameter space N and convex support K. Then $S = \sum_{i=1}^{n} X_i$ is a sufficient statistic. The distributions of S form a standard k-dimensional family with natural parameter space N and convex support $nK = \{s : \exists\, x \in K,\ s = nx\}$. The order of the families corresponding to S and to X_i are equal.

Proof: The joint density of $X_1,\ldots,X_n$ with respect to $\nu\times \ldots \times\nu$ is

$$p_\theta(x_1,\ldots,x_n) = \exp\Big(\sum_{i=1}^{n} (\theta \cdot x_i - \psi(\theta))\Big)$$

$$= \exp\Big(\sum_{i=1}^{n} (\theta_i \cdot x_i - \psi(\theta_i))\Big) \quad \text{with} \quad \theta_i \equiv \theta \ .$$

Hence $X_1,\ldots,X_n$ are canonical statistics from an nk-dimensional exponential family whose parameter space satisfies $\Theta = \{(\theta_1,\ldots,\theta_n) \in R^{nk} : \theta_i \equiv \theta \in R^k\}$. Applying Theorem 1.7 yields that S is sufficient and comes from a standard k-dimensional family with natural parameter space N and convex support nK. (All this is also obvious from the fact that

$$p_\theta(x_1,\ldots,x_n) = \exp\Big(\theta \sum_{i=1}^{n} x_i - n\psi(\theta)\Big) \quad .)$$

It is easily checked that any linear map which transforms the distributions of X_i to a minimal family also transforms those of S to one, and conversely. This yields the assertion concerning the order of the families corresponding to S and X_i . ||

Note that the cumulant generating function for the exponential family generated by S is

$$n\psi(\theta) \quad . \tag{1}$$

The sufficient statistic $\bar{X} = n^{-1}S$ also has distributions from an exponential family. (Apply Theorem 1.6.) Here, the natural parameter space is $n\mathcal{N}$ and the convex support is K. The cumulant generating function for $\bar{X}$ corresponding to the point $\phi = n\theta$ in its natural parameter space is

(2) $$n\psi(\phi/n) \quad .$$

(Under appropriate additional conditions a family of distributions for which there is a nontrivial sufficient statistic based on a sample of size n must be an exponential family. See Dynkin (1951) and Hipp (1974).)

1.12 Examples

Example 1.2 displays an instance of this theorem. If Y is normal with mean μ and variance σ^2 then $X = (Y, Y^2)$ is the canonical statistic of a minimal standard exponential family having canonical parameter $\theta = (\mu/\sigma^2, -1/2\sigma^2)$. Thus if one has i.i.d. observations $Y_1,\ldots,Y_n$ then $S = \sum_{i=1}^{n} X_i = (\sum_{i=1}^{n} Y_i, \sum_{i=1}^{n} Y_i^2)$ is a sufficient statistic; and its distributions form a minimal standard exponential family.

As another example, suppose Y is a member of the *gamma family* with unknown index, α, and scale, σ. The density of Y relative to Lebesque measure on $(0, \infty)$ is

(1) $$f(y) = (\sigma^\alpha \Gamma(\alpha))^{-1} y^{(\alpha-1)} e^{-y/\sigma} , \qquad y > 0 \quad .$$

We will use the notation $Y \sim \Gamma(\alpha, \sigma)$. Note that $\Gamma(m/2, 2) = \chi^2_m$. These distributions form a two-dimensional exponential family with canonical statistic (Y, ln Y) and canonical parameters $(-1/\sigma, \alpha)$. If $Y_1,\ldots,Y_n$ are i.i.d. with density (1) then $S_1 = \sum_{i=1}^{n} Y_i$ and $S_2 = \sum_{i=1}^{n} \ln Y_i$ form a two-dimensional exponential family. It is interesting to note that the marginal distribution of S_1/n also has a density of the form (1) with index $n\alpha$ and scale $n\sigma$. (Here, as well as in the preceding normal example, S_1 is strongly reproductive in the terminology of Barndorff-Nielsen and Blaesild (1983b). For more details see Theorem 2.14 and Example 2.15.)

Another example of interest is provided by the Poisson distribution; where Y has probability function

(2) $$\Pr\{Y = y\} = \lambda^y e^{-\lambda}/y! \qquad y=0,1,\ldots \quad .$$

We will use the notation $Y \sim P(\lambda)$. Then $X = Y$ comes from a one-dimensional exponential family with canonical parameter $\theta = \ln \lambda$. The distribution of $S = \sum_{i=1}^{n} Y_i$ is itself Poisson with natural parameter $\theta + \ln n = \ln n\lambda$.

CONVEXITY PROPERTY

Here is an important fundamental fact about exponential families.

1.13 Theorem

(i) N is a convex set and ψ is convex on N.

(ii) ψ is lower semi-continuous on R^k and is continuous on $N°$.

(iii) $P_{\theta_1} = P_{\theta_2}$ if and only if

(1) $$\psi(\alpha\theta_1 + (1 - \alpha)\theta_2) = \alpha\psi(\theta_1) + (1 - \alpha)\psi(\theta_2)$$

for some $0 < \alpha < 1$. In this case (1) is then valid for all $0 \leq \alpha \leq 1$.

(iv) If $\dim K = k$ (in particular, if $\{p_\theta\}$ is minimal) then ψ is strictly convex on N, and $P_{\theta_1} \neq P_{\theta_2}$ for any $\theta_1 \neq \theta_2 \in N$.

Proof: Let $\theta_1, \theta_2 \in N$, $0 < \alpha < 1$. Then by Hölder's inequality

(2) $$\begin{aligned}\exp(\psi(\alpha\theta_1 + (1 - \alpha)\theta_2)) &= \int\exp((\alpha\theta_1 + (1 - \alpha)\theta_2) \cdot x)\nu(dx) \\ &= \int(\exp \theta_i \cdot x)^\alpha \cdot (\exp \theta_2 \cdot x)^{(1-\alpha)} \nu(dx) \\ &\leq (\int\exp(\theta_1 \cdot x)\nu(dx))^\alpha (\int\exp(\theta_2 \cdot x)\nu(dx))^{(1-\alpha)} \\ &= \exp(\alpha\psi(\theta_1) + (1 - \alpha)\psi(\theta_2)) \quad .\end{aligned}$$

This proves the convexity of ψ, and the convexity of N follows easily.

There is strict inequality in (1) unless

(3) $$\theta_1 \cdot x = \theta_2 \cdot x + K \qquad (\text{a.e. } (\nu))$$

for some constant K; in which case there is equality. (3) is equivalent to

$e^{\theta_1 \cdot x} = e^K e^{\theta_2 \cdot x}$ a.e.(ν) which is equivalent to the assertion $P_{\theta_1} = P_{\theta_2}$.

If (3) holds for some $\theta_1 \neq \theta_2$ then $\dim K \leq k - 1$. Hence $\dim K = k$ implies $P_{\theta_1} \neq P_{\theta_2}$ for any $\theta_1 \neq \theta_2 \in N$.

Finally, for the continuity assertions, note first that $\lambda(\theta) = \int \exp(\theta \cdot x)\nu(dx)$ is lower semi-continuous by Fatou's lemma. Hence ψ is lower semi-continuous. Any convex function defined and finite on a convex set N of R^k must be continuous on $N°$. (We leave this as an exercise on convex sets.) ||

Be careful about the above result -- the fact that ψ is strictly convex on N does not imply that N is strictly convex; for a simple example, see Example 1.2 which involves a minimal family for which

$$N = \{(\theta_1, \theta_2) : \theta_1 \in R, \theta_2 < 0\} .$$

Usually ψ is continuous on all of N. However examples can be constructed when $k \geq 2$ where this is not the case.

This simple theorem has an interesting direct application.

1.14 Example

Let Y be m-variate normal with mean μ and covariance matrix Σ. We will use the notation $Y \sim N(\mu, \Sigma)$. Also, $\delta_{ij} = 1$ if i=j and = 0 if i≠j . The density of Y with respect to Lebesgue measure is

$$(1) \qquad \phi_{\mu,\Sigma}(y) = (2\pi)^{-m/2}|\Sigma|^{-\frac{1}{2}}\exp(\mathrm{tr}(-\Sigma^{-1}(y - \mu)(y - \mu)'/2))$$

$$= (2\pi)^{-m/2}|\Sigma|^{-\frac{1}{2}}\exp((\Sigma^{-1}\mu) \cdot y + \mathrm{tr}((-\Sigma^{-1}/2)(yy')) - \mu'\Sigma^{-1}\mu/2) .$$

It follows that the distributions of Y form an (m + m(m+1)/2) dimensional exponential family with canonical statistics $Y_1,\ldots,Y_m$, $\{Y_iY_j/(1 + \delta_{ij}): i \leq j\}$ and corresponding canonical parameters $(\Sigma^{-1}\mu)_1,\ldots,(\Sigma^{-1}\mu)_m$, $\{(-\Sigma^{-1})_{ij} : i \leq j\}$. For the following it is convenient to label these statistics $X_1,\ldots,X_m$, $\{X_{ij} : i \leq j\}$ and the corresponding parameters as $(\theta_1,\ldots,\theta_m, \{\theta_{ij} : i \leq j\})$. Write $\theta = (\theta_1,\ldots,\theta_m)$, $\mathcal{Q} = (\theta_{ij})$. Ignoring the factor $(2\pi)^{-m/2}$, which can be absorbed into the measure ν, the cumulant generating function is

$$(2)\qquad \psi(\cdot) = (-\tfrac{1}{2})\log|\Sigma^{-1}| + (\mu'\Sigma^{-1}\mu)/2 = (-\tfrac{1}{2})\log(|-\mathcal{Q}|) - \theta'\mathcal{Q}\theta/2 \quad .$$

Note that $N = \{(\theta, \{\theta_{ij} : i \le j\}) : -\mathcal{Q}$ is positive definite$\}$. It is easy to check that N is open, so that this family is regular. By Theorem 1.12

$$(3)\qquad \psi(0, \{\theta_{ij} : i \le j\}) = (-\tfrac{1}{2})\log(|-\mathcal{Q}|)$$

is strictly convex in the variables $\{\theta_{ij} : i \le j\}$ over the set where $\mathcal{Q}$ is positive definite. To reinterpret this result slightly, let $B = -\mathcal{Q}$; then (3) yields that

$$(4)\qquad \log|B| \quad \text{is strictly concave}$$

as a function of the variables $\{b_{ij} : i \le j\}$ over the region where B is positive definite. (4) yields

$$(5)\qquad |B^{-1}| = |B|^{-1} \quad \text{is strictly convex} \quad .$$

((4) can also be proven by directly calculating $\frac{\partial^2}{\partial b_{ij}\,\partial b_{k\ell}} \log|B|$, and showing the resulting $\binom{k+1}{2} \times \binom{k+1}{2}$ matrix is positive definite. The above proof is much simpler !)

CONDITIONAL DISTRIBUTIONS

Let ν be a given σ-finite measure on the Borel subsets of R^k, and $P \ll \nu$ a probability measure with density p. Assume (without loss of generality) that $0 \in N$ so that ν is finite. Let $M_1 : R^k \to R^m$ be linear, $M_1(x) = M_1 x$. Then the conditional measure of ν given $z_1 = M_1(X)$ exists. It will be denoted by $\nu(\cdot|M_1X = z_1)$ or $\nu(\cdot|z_1)$. The conditional distribution of P given $M_1(X)$ exists and has density proportional to $p(\cdot)$ relative to $\nu(\cdot|z_1)$ over the set $\{x : M_1(X) = z_1\}$. (More generally these facts are true if M_1 is any Borel measurable function. See, for example, Neveu (1965).)

The above situation resembles that described in 1.7. Let $M_2 : R^k \to R^{k-m}$ be an orthogonal complement of M_1. Then

$$M_2 : \{x : M_1(x) = z_1\} \to R^{k-m}$$

is 1 - 1.

We will also use the symbol $\nu(\cdot|z_1)$ for the equivalent conditional distribution of $M_2(X)$ given $M_1(X) = z_1$. As before,

$$\phi = M'^{-1}\theta = \begin{pmatrix} M_1'^{-} \\ M_2'^{-} \end{pmatrix}\theta = \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} .$$

It is always possible to choose M_2 to be "orthonormal" so that

$$M_2^{-} = M_2', \quad \text{and so} \quad M_2'^{-} = M_2 .$$

To do so simplifies somewhat the resulting formulae.

1.15 Theorem

The distribution of $Z_2 = M_2X$ given $Z_1 = M_1X$ depends only on $\phi_{(2)} = M_2'^{-}\theta$. For fixed $Z_1 = z_1$ these distributions form the (k-m) dimensional exponential family generated by the measure defined by $\nu(\cdot|z_1)$.

Let N_{z_1} denote the natural parameter space of this conditional family. Then $\phi_2 \in M_2'^{-}N$ implies

(1) $$\phi_2 \in N_{M_1X} \quad \text{a.e.}(\nu) .$$

Furthermore, if $\{p_\theta\}$ is regular then

(2) $$M_2'^{-}N \subset N_{M_1X} \quad \text{a.e.}(\nu) .$$

Proof: The conditional density of Z_2 given $Z_1 = z_1$ is proportional to

$$p_\theta((z_1, z_2)) = \exp(\phi_1 \cdot z_1 + \phi_2 \cdot z_2 - \psi(\theta)) .$$

Hence the density of Z_2 given $Z_1 = z_1$ relative to $\nu(\cdot|z_1)$ can be written as

(3) $$p_\phi(z_2) = \exp(\phi_2 \cdot z_2 - \psi_z(\phi_2))$$

where

(4) $$\psi_{z_1}(\phi_2) = \ln(\int \exp(\phi_2 \cdot z_2)\nu(dz_2|z_1)) .$$

The natural parameter space N_{z_1} is the set$\{\phi_2\}$, for which the integral on the right of (4) is finite. Let $\phi_2 \in M_2'^{-}N$. There is thus a $\theta \in N$

for which $\phi_2 = M_2'^{-}\theta$. Let ν^* denote the marginal measure on R^m defined by $\nu^*(A) = \nu(M_1^{-1}(A))$. Then

$$\infty > \int \exp(\theta \cdot x)\nu(dx) = \int\{\int \exp(\phi_1 \cdot z_1 + \phi_2 \cdot z_2)\nu(dz_2|z_1)\}\nu^*(dz_1) \quad .$$

Hence

$$\infty > \int \exp(\phi_2 \cdot z_2)\nu(dz_2|z_1)$$

for almost every $z_1(\nu^*)$. This verifies (1).

Suppose $\{p_\theta\}$ is regular. Let $\{\theta_i: i=1,\ldots,\} \subset N$ be a countable dense subset of N. $\{M_2'^{-}\ \theta_i : i=1,\ldots\}$ is dense in $M_2'^{-}N$. $M_2'^{-}$ is a linear map. Hence $M_2'^{-}N$ is convex and open since N is convex (by Theorem 1.13) and open (by assumption). It follows that

$$\text{(5)} \qquad \text{conhull } \{M_2'^{-}\ \theta_i : i=1,\ldots\} = M_2'^{-}N \quad .$$

(We leave (5) as an exercise on convex sets.)

Since $\{\theta_i\}$ is countable it follows from (1) that

$$M_2'^{-}\ \theta_i \subset N_{M_1X} \quad \text{for all} \quad i=1,\ldots, \quad \text{a.e.}(\nu) \quad .$$

Thus

$$M_2'^{-}N = \text{conhull } \{M_2'^{-}\ \theta_i : i=1,\ldots\} \subset N_{M_1X} \quad \text{a.e.}(\nu) \ ,$$

since N_{M_1X} is convex; which proves (2). ||

The above result can be given an alternate interpretation under which the conditional distributions of X given X ∈ L form an exponential family, for L a given linear variety in R^k. See 1.7(8). We omit the details.

Here are two important simple applications of the above ideas.

1.16 Example

Let $X_1,\ldots,X_k$ be independent Poisson variables with expectations λ_i . See 1.12(2). Then $X = (X_1,\ldots,X_k)$ is the canonical statistic of a standard exponential family with natural parameter θ: $\theta_i = \ln \lambda_i \quad i=1,\ldots,k$. The dominating measure has $\nu(\{x\}) = 1/\prod_{i=1}^{k} x_i!$. Let $N > 0$ be an integer.

Then the distributions of X given $\sum_{i=1}^{k} X_i = N$ form a standard exponential family with dominating measure

$$(1) \qquad \nu(\{x\} \mid \sum_{i=1}^{k} x_i = N) = 1/\prod_{i=1}^{n} x_i!, \qquad \text{for} \qquad \sum_{i=1}^{k} x_i = N .$$

This measure is proportional to the measure 1.3(1) which generates the multinomial distribution. Hence the conditional distribution is multinomial (N,π).

The value of π can be easily computed as follows: orthogonally project onto $\{\theta: \Sigma\theta_i = 0\}$ which is the linear subspace parallel to $\{x : \Sigma x_i = N\}$. This yields $(\theta - \bar{\theta}1)$ (where $\bar{\theta} = k^{-1}\sum_{i=1}^{k}\theta_i$) as the natural parameter of the conditional multinomial distribution. Thus

$$\pi_i = ce^{\theta_i - \bar{\theta}}$$

with $c = (\sum_{i=1}^{k} e^{\theta_i - \bar{\theta}})^{-1}$. Substituting $\theta_i = \ln \lambda_i$ yields

$$(2) \qquad \pi_i = \lambda_i / \sum_{i=1}^{k} \lambda_i .$$

1.17 Example

Let X be k-variate normal with mean μ and covariance Σ. For Σ given the distributions of X form a standard exponential family with natural parameter $\theta = \Sigma^{-1}\mu$. (This can easily be checked directly or derived from Example 1.14 by using Theorem 1.7.) The dominating measure for this family is proportional to $\nu(dx) = \exp(-x'\Sigma^{-1}x/2)dx$.

Let $z_1 = (x_1,\ldots,x_m)$, $z_2 = (x_{m+1},\ldots,x_k)$. The conditional distributions of Z_2 given $Z_1 = z_1$ form an exponential family. The natural parameter for this family is just $\phi_2 = (\theta_{m+1},\ldots,\theta_k)'$.

Partition Σ as

$$(1) \qquad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \quad \text{with} \quad \Sigma_{11}(m \times m) , \quad \text{etc.}$$

Then

(2) $$\Sigma^{-1} = \begin{pmatrix} (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1} & -\Sigma_{11}^{-1}\Sigma_{12}(\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1} \\ -(\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1}\Sigma_{21}\Sigma_{11}^{-1} & (\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1} \end{pmatrix}$$

$$= \begin{pmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{pmatrix}, \text{ say.}$$

((2) is a general formula for block symmetric positive definite matrices. Note that $\Sigma^{12} = -\Sigma_{11}^{-1}\Sigma_{12}(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1} = -\Sigma_{22}^{-1}\Sigma_{21}(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1}$.) Note that the natural parameter can be written as

$$\phi_2 = \begin{pmatrix} \theta_{m+1} \\ \vdots \\ \theta_k \end{pmatrix} = (\Sigma^{-1}\mu)_2 = \Sigma^{21}\mu_{(1)} + \Sigma^{22}\mu_{(2)}$$

where

$$\mu = \begin{pmatrix} \mu_{(1)} \\ \mu_{(2)} \end{pmatrix} .$$

Consider the case where $z_1 = 0$. The conditional dominating measure is

$$\nu(dz_2|0) = c \exp(-z_2'\Sigma^{22} z_2/2)$$

and is thus a normal density with mean 0, variance-covariance $(\Sigma^{22})^{-1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} = \Sigma^*$, say. It follows that the conditional density of Z_2 given $Z_1 = 0$ is normal with this covariance matrix and with mean μ^* given by

$$\Sigma^{*-1}\mu^* = \phi_2 ,$$

since ϕ_2 must be the value of the natural parameter for both the unconditional and conditional family. Hence

(3) $$\mu^* = \Sigma^*\phi_2 = \Sigma^*(\Sigma^{21}\mu_{(1)} + \Sigma^{22}\mu_{(2)}) = -\Sigma_{21}\Sigma_{11}^{-1}\mu_{(1)} + \mu_{(2)} .$$

For $z_1 \neq 0$ it is convenient to use the location invariance of the normal family. The conditional distribution under (μ,Σ) of $Z_{(2)}$ given $Z_{(1)} = z_{(1)}$ is the same as the conditional distribution under $(\begin{pmatrix} \mu_{(1)} - z_{(1)} \\ \mu_{(2)} \end{pmatrix}, \Sigma)$ of $Z_{(2)}$ given $Z_{(1)} = 0$. By the preceding this is normal with covariance matrix $\Sigma^* = (\Sigma^{22})^{-1}$ and mean $\mu_{(2)} - \Sigma_{21}\Sigma_{11}^{-1}(\mu_{(1)} - z_{(1)})$.

EXERCISES

1.1.1 (a) Let C be any closed convex set in R^k. Show that there exists a standard exponential family with $N = C$. [$C = \bigcap_{i=1}^{\infty} \{\theta: v_i \cdot \theta \le c_i\}$ with $||v_i|| = 1$. Let ν_i denote Lebesgue measure on the ray $\{x: x = \alpha v_i, \alpha > 0\}$ and let $\nu = \sum_{i=1}^{\infty} 2^{-i} \exp(c_i v_i \cdot x)\nu_i/(1+||x||^2)$. The result is also true, but harder to prove, if C is an open convex set.]

(b) Let $C = \{(\theta_1, \theta_2): ||\theta||^2 < 1\} \cup \{(0, 1)\}$ and show there exists an exponential family with $N = C$.

1.2.1 Verify 1.2(5) (including the formula for ν which precedes it). Note that when n = 1 the measure ν can be described by the relations $x_2 = x_1^2$ and $\nu(dx_1) = dx_1/\sqrt{2\pi}$.

1.7.1 (i) Let Z = MX as in Theorem 1.7. Show that Z_1 is independent of Z_2 for some $\theta \in \Theta$ if and only if Z_1 is independent of Z_2 for all $\theta \in \Theta$.

(ii) Give an example to show that the assertion is false if Z_1, Z_2 are non-linear transformations of X. [(i) Assume independence at $\theta = 0$.
(ii) Let X be bivariate normal with mean μ and covariance I, and $Z_1 = ||x||$, $Z_2 = \tan^{-1}(x_2/x_1)$.]

1.7.2 Consider the situation of Theorem 1.7. Suppose the original family $\{p_\theta: \theta \in N\}$ is full and minimal. Then the family of distributions of Z_1 for $\phi_1 \in \Phi_{\phi_2^0}$ is full. It is minimal if and only if there is a $\theta \in \text{int}\, N$ with $M_2'\phi_2^0 = \theta$. [For a situation where the family of distributions of Z_1 is not minimal use Exercise 1.1.1(b), let M be as in (4), and let $\phi_2^0 = 1$.]

1.7.3 (a) Show that if 1.7(7) or (8) are satisfied then the distributions of $Z_1 = M_1X$ form a standard exponential family with natural parameter ϕ_1.

(b) Give an example to show that the distributions of $Z_1 = M_1X$ may form a standard exponential family with natural parameter different from ϕ_1 even when 1.7(7) and (8) fail. [Consider the distribution of X_1 when X is multinomial of dimension $k \geq 3$, or equivalently, of X_1^* with X^* as in Example 1.3. There are also some other interesting instances of this phenomenon.]

1.8.1 (Contingency table under independence). Consider a 2×2 contingency table in which the observations are Y_{ij}, $1 \leq i, j \leq 2$, and have a multinomial (N, p) distribution with $p = \{p_{ij}, 1 \leq i, j \leq 2\}$. Under the model of independence $p_{ij} = p_{i+} \cdot p_{+j}$ where $p_{i+} = \sum_j p_{ij}$, etc.. Write this independence model as a log-linear model in a fashion so that the coordinates of the natural (minimal) sufficient statistic are independent binomial variables. Generalize to the model of independence in an r×c contingency table. (For further log-linear models in contingency tables, see Haberman (1974), (1979).)

1.10.1 Show that in any standard exponential family of dimension k and order m, $m + k \geq \dim K + \dim \Theta$. Give an example in which $m < \min(\dim \Theta, \dim K)$. [The simplest example has $m = 0$, $\dim \Theta = \dim K = 1$, $k = 2$.]

1.12.1 From many points of view the negative binomial distributions are the discrete analog to the gamma distributions. The *negative binomial*, $NB(\alpha, p)$, distribution has probability function

$$P(x) = \frac{\Gamma(x + \alpha)}{\Gamma(x+1)\Gamma(\alpha)} (1 - p)^{\alpha} p^x , \quad x=0,1,\ldots \quad .$$

Show that for fixed α the family $NB(\alpha, \cdot)$ is a one parameter exponential family, but that -- unlike the $\Gamma(\alpha, \sigma)$ situation -- the family $NB(\alpha, p)$ $\alpha > 0$, $0 < p < 1$ is not an exponential family.

1.12.2 Let ν denote counting measure on $\{(0,0), (1,1), (2,0), (3,1), (4,0), \ldots\} \subset R^k$. Show that the exponential family generated by ν has the following properties: X_1 has a *geometric distribution*, $Ge(p_1) = NB(1,p)$; X_2 has a binomial distribution, $B(p_2)$; $(X_1 - X_2)/2$ has a geometric distribution $Ge(p_3)$ and $(X_1 - X_2)/2$ is independent of X_2. Write p_1, p_2, p_3 in terms of the natural parameters θ_1, θ_2.

1.12.3 Let $Z_1,\ldots,Z_m$ be i.i.d. $N(\mu,\sigma^2)$. Let $X = \sum_{i=1}^{m} Z_i^2$. Then X has a scaled non-central χ^2 distribution with m degrees of freedom, non-centrality parameter $\delta = m\mu^2/\sigma^2$, and scale parameter σ^2. Denote this distribution by $\chi_m^2(\delta, \sigma^2)$. (i) This distribution has density

$$(1) \qquad g(x) = \sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} \frac{(\frac{x}{\sigma^2})^{(m/2)+k-1} e^{-x/(2\sigma^2)}}{\sigma^2\Gamma(k + \frac{m}{2})\, 2^{k + m/2}} , \qquad x > 0 ,$$

where $\lambda = \delta/2$. (From the form of (1) it is evident that $K = k \sim P(\lambda)$ and $X|K \sim \Gamma(k + m/2, \sigma^2)$; thus (X/σ^2) K is central χ^2 with $k + \frac{m}{2}$ degrees of freedom.) (ii) The distributions of X can also be represented as the marginal distribution of X_1 from a canonical two parameter exponential family generated by a measure ν supported on $\{(x_1, x_2)\colon x_1 > 0,\ x_2 = 0,1,\ldots\}$. [(i) By change of variables and series expansion prove (1) for the case m=1. (1) for general m then follows from facts about sums of Poisson and gamma variables. (ii) Let ν be the measure generated by (1) with $\sigma^2 = 1$, $\lambda = 1$.]

1.13.1 (i) Show that when k = 1 then ψ must be continuous on N. [Use 1.13 and convexity of N.]

(ii) More generally, let $\theta_0,\theta_1 \in N$ and $\theta_\rho = (1 - \rho)\theta_0 + \rho\theta_1$ and show $\psi(\theta_\rho)$ is continuous in ρ for $0 \le \rho \le 1$. [Reason as in (i), or use Theorem 1.7 and (i).]

(iii) Give an example of an exponential family in which ψ is not continuous on N. [Exercise 1.1.1(b) provides an example.]

1.13.2 Generalize 1.13.1(ii) as follows: let $\theta \in N$ and $\theta_j \in N$, j=1,...,J. Let $\tau_i \in \operatorname{conhull}\{\theta, \theta_1,\ldots,\theta_J\}$, i=1,... and $\tau_i \to \theta$. Then $\psi(\tau_i) \to \psi(\theta)$. [Write $\tau_i = \sum_j \alpha_{ij}[(1 - \rho_i)\theta_j + \rho_i\theta]$ with $\alpha_{ij} \ge 0$, $\sum_j \alpha_{ij} = 1$, and $\rho_i \uparrow 1$.

Use 1.13.1(ii) and the fact that $\psi(\tau_i) \le \sum_j \alpha_{ij}\, \psi((1 - \rho_i)\theta_j + \rho_i\theta)$.]

1.13.3 Let $Y = (Y_0 = 1, Y_1,\ldots Y_n)'$ be the initial state and n further observations from an S-state Markov chain with transition matrix P (i.e., $P(Y_\ell = j|Y_{\ell-1} = i) = p_{ij}$, $1 \le i,j \le S$, $\ell=1,\ldots,n$). Let N denote the sample transition matrix, $N = \{n_{ij}\}$,

$$(1) \qquad n_{ij} = \sum_{\ell=1}^{n} \chi_{\{i,j\}}(Y_{\ell-1}, Y_\ell) .$$

Suppose $p_{ij} > 0$, $1 \le i,j \le S$. Show that the distributions of Y form an S^2 dimensional exponential family with canonical statistic $N = \{n_{ij}\}$ and canonical parameters $\{\log p_{ij}\}$. Show that if $n \ge 3$ the family has order $S^2 - 1$. [Let E_{ij} denote the matrix with i,j-th entry 1 and all other entries 0. Show that for given $1 \le i,j \le K$ there exist sample points N_1, N_2 having positive probability and that $N_1 + (E_{ii} - E_{ji}) = N_2$ and (other) points N_1, N_2 such that $N_1 + (E_{ii} - E_{11}) = N_2$.]

1.14.1 Univariate General Linear Model (G.L.M.). Let Y be m-variate normal, $Y \sim N(\mu, \sigma^2 I)$, $\mu \in R^m$, $\sigma^2 > 0$. (a) Show that this is an m+1 dimensional exponential family. (b) In the G.L.M. μ is restricted by

$$\mu = B\beta , \qquad \beta \in R^r$$

with B a known $m\times r$ matrix. Assume (for convenience) B has rank r. Show that this is a full (r+1) dimensional exponential family. [Use Example 1.14 and Theorem 1.7.]

1.14.2 Matrix normal distribution. Let $\mu = \{\mu_{ij}\}$ be an $m\times q$ matrix and let $\Gamma = \{\gamma_{ij}\}$ and $\Sigma = \{\sigma_{ij}\}$ be $m\times m$ and $q\times q$ positive definite matrices, respectively. Let $Y = \{Y_{ij}\}$ be an $m\times q$ random matrix whose entries have a multivariate normal distribution with

$$E\, Y_{ij} = \mu_{ij} \qquad \mathrm{Cov}\, Y_{ij} Y_{i'j'} = \gamma_{ii'}\sigma_{jj'} \quad .$$

This is the matrix normal distribution, denoted by $Y \sim N(\mu, \Gamma, \Sigma)$.

(a) Show that Y has density (relative to Lebesgue measure on R^{mq})

$$f(y) = (2\pi)^{-mq/2}\, |\Gamma|^{-m/2}\, |\Sigma|^{-q/2} \exp \mathrm{tr}\, (-\Gamma^{-1}(Y - \mu)\Sigma^{-1}(Y - \mu)'/2) \quad .$$

[See Arnold (1981, Theorem 17.4).]

(b) Reduce this to an $mq + \frac{m(m+1)q(q+1)}{4}$ dimensional minimal exponential family with canonical parameters $\theta_{ij} = \Gamma^{-1}\mu\Sigma^{-1}$, $1 \le i \le m$, $1 \le j \le q$, and $\theta_{ii'jj'} = \gamma^{ii'}\sigma^{jj'}$, $1 \le i \le i' \le m$, $1 \le j \le j' \le q$, where $\Gamma^{-1} = \{\gamma^{ij}\}$,

$\Sigma^{-1} = \{\sigma^{ij}\}$.

(c) Show that if $m \geq 2$ and $q \geq 2$ this is not a full exponential family. Rather, Θ is an $mq + m(m+1)/2 + q(q+1)/2 - 1$ dimensional differentiable manifold inside of N.

(An alternate notation involves writing $Y = (Y_{(1)},\ldots,Y_{(q)})$ and defining $(\text{vec } Y)' = (Y'_{(1)},\ldots,Y'_{(q)})$. Then $Y \sim N(\mu, \Gamma, \Sigma)$ is the same as $\text{vec } Y \sim N(\text{vec } \mu, \Sigma \otimes \Gamma)$ where $\otimes$ denotes the Kronecker product.

1.14.3 Multivariate Linear Model (M.L.M.). Here $Y \sim N(\mu, I, \Sigma)$ with Σ positive definite and

$$\mu = B\beta$$

with B a known $m\times r$ matrix and β an $(r\times q)$ matrix of parameters. Assume (for convenience) B has rank r. Show that this can be reduced to a full minimal regular exponential family of dimension $rq + q(q+1)/2$.

1.14.4 Wishart distribution. Let $X = (x_{ij})$ and $\Sigma = (\sigma_{ij})$ be symmetric $m\times m$ positive definite matrices. The matrix $\Gamma(\alpha, \Sigma)$ distribution has density

$$p_{\alpha,\Sigma}(X) = \frac{|X|^{\alpha-(m+1)/2} \exp\operatorname{tr}(-\Sigma^{-1}X)}{\Gamma_m(\alpha)\,|\Sigma|^{\alpha}} \tag{1}$$

where

$$\Gamma_m(\alpha) = \pi^{m(m-1)/4} \prod_{i=1}^{m} \Gamma(\alpha - (i-1)/2)\ , \qquad \alpha > (m-1)/2\ .$$

Show this is an exponential family, and describe the natural observations, natural parameters, and cumulant generating function.

(If Y_i, $i=1,\ldots,n$, are independent $N(0, \Sigma)$ vectors then $\sum_{i=1}^{n} Y_iY_i' = X$ has the $\Gamma(\frac{n}{2}, 2\Sigma)$ distribution. This is also called the *Wishart* (n, Σ) distribution and denoted by $W(n, \Sigma)$. See e.g. Arnold (1981). Also $\sum_{i=1}^{n} (Y_i - \bar{Y})(Y_i - \bar{Y})' \sim W(n-1, \Sigma)$.)

1.15.1 Consider a 2×2 contingency table (see Exercise 1.8.1). Find the conditional distribution of Y_{ij} given $Y_{i+} = \sum_{j=1}^{2} Y_{ij}$ and $Y_{+j} = \sum_{i=1}^{2} Y_{ij}$. Show that these conditional distributions depend only on the given values Y_{i+}, Y_{+j} and on the *odds ratio* $p_{11}p_{22}/p_{12}p_{21}$ and form a one-parameter exponential family. [Under the independence model the distribution is hypergeometric and independent of p.]