CHAPTER 2. ANALYTIC PROPERTIES

DIFFERENTIABILITY AND MOMENTS

The cumulant generating function has several nice properties. Among these are the fact that its defining expression may be differentiated under the integral sign. In this manner one obtains the moments of X from the derivatives of ψ.

One needs first to establish a simple bound.

2.1 Lemma

Let $B = \text{conhull}\,\{b_i : i=1,\ldots,I\} \subset R^k$. Let $C \subset B^\circ$ be compact and let $b_0 \in C$. Then there are constants K_ℓ (depending on C,B) $\ell=0,1,\ldots$ such th

$$(1) \qquad ||x||^\ell e^{b\cdot x} \le K_\ell \sum_{i=1}^{I} e^{b_i\cdot x} \qquad \forall\ b \in C,\quad x \in R^k \quad .$$

Also,

$$(2) \qquad \left| \frac{e^{b\cdot x} - e^{b_0\cdot x}}{||b - b_0||} \right| \le K_1 \sum_{i=1}^{I} e^{b_i\cdot x} \quad , \quad b \in C,\quad x \in R^k \quad .$$

Proof. Let $\varepsilon > 0$. Note that there exists a $K_{\ell,\varepsilon} < \infty$ such that

$$|r|^\ell \le K_{\ell,\varepsilon}\, e^{\varepsilon|r|} \qquad \forall \quad r \in R$$

since

$$\lim_{|r|\to\infty} |r|^\ell/e^{\varepsilon|r|} = 0 \quad .$$

Let $\{e_i : i=1,\ldots,k\}$ denote the elementary (orthogonal) unit vectors in R^k. Then

$$||x||^\ell \le k^{(\ell-2)/2} \sum_{i=1}^{k} |x_i|^\ell \le K'_{\ell,\varepsilon} \sum_{i=1}^{k} e^{\varepsilon|x_i|} < K'_{\ell,\varepsilon} \sum_{i=1}^{k} (e^{\varepsilon e_i\cdot x} + e^{-\varepsilon e_i\cdot x}),$$

where $K'_{\ell,\varepsilon} = k^{(\ell-2)/2} K_{\ell,\varepsilon}$. Choose $\varepsilon > 0$ such that $(b \pm \epsilon e_i) \in B$, $i=1,\ldots,k$, for all $b \in C$. See Figure 2.1(1). By convexity

$$e^{(b\pm\varepsilon e_i)\cdot x} \leq \max(e^{b_i\cdot x}) \quad ,$$

since $e^{a\cdot x}$ is convex in $a \in R^k$ and $(b\pm\varepsilon e_i) \in B = \text{conhull}\ \{b_i\}$. Then

$$||x||^\ell e^{b\cdot x} \leq K'_{\ell,\varepsilon}\ e^{b\cdot x} \sum_{i=1}^{k} (e^{\varepsilon e_i\cdot x} + e^{-\varepsilon e_i\cdot x}) \leq K'_{\ell,\varepsilon} \sum_{i=1}^{k} (e^{(b+\varepsilon e_i)\cdot x}$$

$$+ e^{(b-\varepsilon e_i)\cdot x}) \leq 2k\ K'_{\ell,\varepsilon} \max(e^{b_i\cdot x}) \leq 2k\ K'_{\ell,\varepsilon} \sum_{i=1}^{I} e^{b_i\cdot x} \quad .$$

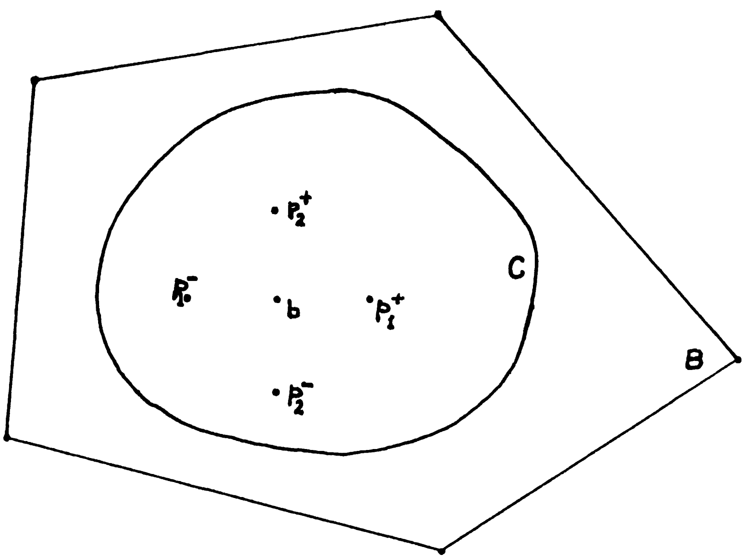

Figure 2.1(1): B, C, and $p_{i\pm} = b \pm \varepsilon e_i$ for the proof of Lemma 2.1.

This proves (1), with $K_\ell = 2k\, K'_{\ell,\varepsilon}$.

Note that (2) may also be written

$$\left| \frac{e^{(b-b_0)\cdot x} - 1}{b - b_0} \right| \le K_1 \sum_{i=1}^{I} e^{(b_i - b_0)\cdot x} \quad .$$

Hence it suffices to prove (2) in the case where $b_0 = 0$, (so that $e^{b_0\cdot x} = 1$) and we make this assumption below. Note that $re^r - e^r + 1 \ge 0$ and also that for $r \le 0$ $1 - e^r < |r|$. Using the first inequality when $b\cdot x > 0$ and the second when $b\cdot x < 0$ yields

$$\left| \frac{e^{b\cdot x} - 1}{||b||\ ||x||} \right| \le \left| \frac{e^{b\cdot x} - 1}{b\cdot x} \right|$$

$$\le \frac{\max(b\cdot x e^{b\cdot x},\ |b\cdot x|)}{|b\cdot x|} .$$

Hence

$$\left| \frac{e^{b\cdot x} - 1}{||b||} \right| \le ||x|| e^{b\cdot x} + ||x|| \le 2K_1 \sum_{i=1}^{p} e^{b_i\cdot x}$$

by (1) since $b \in C$ and $0 \in C$. ||

FORMULAS FOR MOMENTS

Let $\ell_i \ge 0$ be non-negative integers with $\sum_{i=1}^{k} \ell_i = \ell$. Formal calculation yields

(1) $$\frac{\partial^\ell}{\prod_{i=1}^{k} \partial\theta_i^{\ell_i}} \lambda(\theta) = \int \Big(\prod_{i=1}^{k} x_i^{\ell_i}\Big) e^{\theta\cdot x} \nu(dx) \quad .$$

In particular

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

These calculations are justified by the following theorem.

2.2 Theorem

Suppose $\theta_0 \in N^\circ$. Then all derivatives of λ and of ψ exist at

θ_0. They are given by the above expressions (1), (2) derived by formally differentiating under the integral sign.

Proof. We prove only (2). (The proof of the general formula (1) is similar and proceeds by induction on ℓ. See Exercise 2.2.1.) Let $\theta_0 \in N^\circ$. Then there is a $B = \text{conhull}\{\theta_i : i=1,\dots,I\} \subset N^\circ$ and $C \subset B^\circ$, C compact, with $\theta_0 \in C^\circ$.

Let

(3) $$d(\theta, x) = \frac{e^{\theta\cdot x} - e^{\theta_0\cdot x} - (\theta-\theta_0)\cdot xe^{\theta_0\cdot x}}{||\theta - \theta_0||} .$$

By Lemma 2.1

(4) $$\sup_{\theta\in C} |d(\theta, x)| \le 2K_1 \Sigma e^{\theta_i\cdot x} .$$

Also

(5) $$|d(\theta, x)| \to 0 \quad \text{as} \quad \theta \to \theta_0$$

since $\nabla e^{\theta\cdot x}\big|_{\theta=\theta_0} = xe^{\theta_0\cdot x}$. Hence

$$\int d(\theta, x)\nu(dx) \to 0 \quad \text{as} \quad \theta \to \theta_0$$

by the dominated convergence theorem, so that

(6) $$\frac{\lambda(\theta) - \lambda(\theta_0) - (\theta - \theta_0)\cdot \int xe^{\theta_0\cdot x}\nu(dx)}{||\theta - \theta_0||} \underset{\theta\to\theta_0}{\to} 0 ,$$

which proves (1). ||

Theorem 2.2 immediately yields the following fundamental formulae. For $f : R^k \to R$ introduce the notation $D_2 f$ for the k×k matrix $(\frac{\partial^2 f}{\partial x_i \partial x_j})$. An alternate expression is $\nabla'\nabla f$ since ∇' converts each element of the (column) vector $\frac{\partial f}{\partial x_i}$ into the row vector $(\partial(\frac{\partial f}{\partial x_i})/\partial x_j : j=1,\dots,k)$, and hence $D_2 f = \nabla'\nabla f$.

2.3 Corollary

Consider a standard exponential family. Let $\theta \in N^\circ$. Then

(1) $$E_\theta(X) = \nabla\psi(\theta)$$

(2) $$\operatorname{cov}_\theta X = D_2\psi(\theta) = \nabla'\nabla\psi(\theta) \quad .$$

Notation. In the sequel we frequently use the notation

(1') $$\xi(\theta) = \nabla\psi(\theta) = E_\theta(X)$$

and

(2') $$\Sigma(\theta) = D_2\psi(\theta) = \Sigma_\theta(X) \ .$$

Proof. Calculating formally,

$$\nabla\psi(\theta) = \int x e^{\theta\cdot x}\nu(dx)/\int e^{\theta\cdot x}\nu(dx)$$

$$= E_\theta(X) \ .$$

The calculation is justified by Theorem 2.2. This proves (2). The proof of (1) is similar. ||

2.4 Examples

The reader is invited to use Corollary 2.3 to calculate the familiar formulae for mean and variance in the classic exponential families such as (univariate) normal, multinomial, Poisson, gamma, negative binomial, etc..

For the multivariate normal distribution Corollary 2.3 provides a benefit in the reverse direction. Let Y be m-variate normal (μ, Σ), as in Example 1.14. Fix $\mu = 0$. Direct calculation (not using Corollary 2.3) yields the familiar result

(1) $$E(Y_iY_j) = \sigma_{ij} = (-(\mathcal{Q})^{-1})_{ij} = -\theta^{ij}$$

when $\mu = 0$, where $\mathcal{Q}^{-1} = (\theta^{ij})$. Calculation using Corollary 2.3 and the formula 1.14(3) for the cumulant generating function thus yields for $i \neq j$

(2) $$\frac{\partial}{\partial\theta_{ij}} (-\tfrac{1}{2})\log|-\mathcal{Q}| = -\theta^{ij}/(1 + \delta_{ij})$$

since the corresponding canonical statistics are $Y_iY_j/(1+\delta_{ij})$. Let $B = -\mathcal{Q}$. Then (2) shows that for any positive definite symmetric matrix, B,

$$\text{(3)} \qquad \frac{\partial}{\partial b_{ij}} \log|B| = 2b^{ij}/(1+\delta_{ij}) \quad \text{where} \quad B^{-1} = (b^{ij}) \ .$$

Hence, also,

$$\text{(4)} \qquad \frac{\partial}{\partial b_{ij}} |B| = 2b^{ij}|B|/(1+\delta_{ij}) \ . \qquad ||$$

The convexity of ψ together with Theorem 2.2 yields the following useful result.

2.5 Corollary

Let $\theta_1, \theta_2 \in N^{\circ}$. Then

$$\text{(1)} \qquad (\theta_1 - \theta_2) \cdot (\xi(\theta_1) - \xi(\theta_2)) \geq 0 \ .$$

Equality holds in (1) if and only if $P_{\theta_1} = P_{\theta_2}$. Consequently $\xi(\theta_1) = \xi(\theta_2)$ if and only if $P_{\theta_1} = P_{\theta_2}$. (If $\{p_\theta\}$ is minimal this happens only when $\theta_1 = \theta_2$.)

Proof. ψ is convex. Hence the directional derivative of ψ in direction $\theta_1 - \theta_2$ is non-decreasing as one moves along the line from θ_2 to θ_1. That is,

$$\text{(2)} \qquad (\theta_1 - \theta_2) \cdot \nabla\psi(\theta_2 + \rho(\theta_1 - \theta_2)) = (\theta_1 - \theta_2) \cdot \xi(\theta_2 + \rho(\theta_1 - \theta_2))$$

is non-decreasing in ρ. This yields (1).

If $P_{\theta_1} \neq P_{\theta_2}$ then ψ is strictly convex on the line joining θ_2 and θ_1. Hence (2) is strictly increasing for $\rho \in (0,1)$. This yields the remaining assertions of the corollary. (The parenthetical assertion is contained in Theorem 1.13.) ||

The final corollary to Theorem 2.2 establishes the possibility of differentiating inside the integral sign for expectations involving exponential families. The result is stated only for real valued statistics, but obviously generalizes to higher dimensional statistics.

2.6 Corollary

Let $T : R^k \to R$. Let

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

Then $N(T)$ is convex. Define

(2) $$h(\theta) = \int T(x) e^{\theta \cdot x} \nu(dx) = e^{\psi(\theta)} E_\theta(T(X))$$

for $\theta \in N(T)$. Then all derivatives of h exist at every $\theta \in N^\circ(T)$, and they may be computed under the integral sign. In particular

(3) $$\nabla E_\theta(T(X)) = \int (x - \xi(\theta)) T(x) \exp(\theta \cdot x - \psi(\theta)) \nu(dx) .$$

Proof. Suppose $T(x) \geq 0$. Applying Theorem 2.2 to the measure $\omega(dx) = T(x)\nu(dx)$ yields the desired results. For general T the corollary follows upon using the above to separately treat T^+ and T^-. ||

Note that if T and $|T|^{-1}$ are bounded then $N(T) \supset N$.

ANALYTICITY

The moment generating function is analytic. This fact is implicit in the proof of Theorem 2.2. As a preliminary we extend the definition of λ and ψ to the complex domain.

Let

$$\lambda : \mathbb{C}^k \to \mathbb{C}$$

be defined by the same expression as previously, i.e.

(1) $$\lambda(\theta) = \int \exp(\theta \cdot x) \nu(dx) .$$

For $\theta \in \mathbb{C}^k$ let Re θ denote the vector with coordinates $(\text{Re } \theta_1, \ldots, \text{Re } \theta_k)$. Note that for $x \in R^k$

(2) $$|e^{\theta \cdot x}| = e^{(\text{Re } \theta) \cdot x} .$$

Hence $\lambda(\theta)$ exists for Re $\theta \in N$.

2.7 Theorem

$\lambda(\theta)$ is analytic on $\{\theta \in \mathbb{C}^k : \text{Re } \theta \in N^\circ\}$.

Proof. Lemma 2.1 (and its proof) apply for $b \in \mathbb{C}^k$, $x \in R^k$. Similarly the proof of Theorem 2.2(2) is valid verbatim for $\theta \in \mathbb{C}^k$. Thus $\nabla\lambda(\theta)$ exists for Re $\theta \in N^\circ$ (and has the expression 2.2(2)). This implies that λ is analytic on this domain. ||

Two important properties of analytic functions are: (i) they can be expanded in a Taylor series; and (ii) they are analytic in each variable separately. Thus, for a fixed value of $(\theta_2,\ldots,\theta_k)$, $\lambda((\cdot,\theta_2,\ldots,\theta_k))$ is analytic. $\lambda((\cdot,\theta_2,\ldots,\theta_k))$ is determined by its values on any subset having an accumulation point. This is the basis for the following result.

2.8 Lemma

Let $T : R^k \to R$, and let

(1) $$h(\theta) = \int T(x)e^{\theta\cdot x}\nu(dx), \quad \text{for} \quad \text{Re } \theta \in N(T), \quad \text{as defined in 2.6(1).}$$

Then h is analytic on $\{\theta \in \mathbb{C}^k : \text{Re } \theta \in N^\circ(T)\}$.

Let L be a line in R^k, and let $B \subset L \cap N(T)$ be any subset of $L \cap N(T)$ having an accumulation point in $N^\circ(T)$. Then

(2) $$h(\theta) = 0 \quad \forall\, \theta \in B$$

implies $h(\theta) = 0$ for all $\theta \in R^k$ such that $\theta \in L \cap N^\circ(T)$.

Proof. The first assertion follows upon applying Theorem 2.7 to $T^+(x)\nu(dx)$, and $T^-(x)\nu(dx)$.

Next, one may apply linked affine transformations as in Proposition 1.6. Because of this it suffices to consider the case where $L = \{\theta \in R^k : \theta_2 = \ldots = \theta_k = 0\}$. $h((\theta_1,0,\ldots,0))$ is an analytic function of $\theta_1 \in \mathbb{C}$, as already noted. Hence (2) implies $h(\theta,0,..,0) \equiv 0$ on its domain of analyticity, which is $\{(\theta,0,\ldots,0) : \text{Re } \theta \in L \cap N^\circ(T)\}$. This proves the analyticity, which is $\{(\theta,0,\ldots,0)\colon \operatorname{Re}\theta \in L \cap N^{o}(T)\}$. This proves the second assertion. ||

Note that, more generally, if B is as above then the values of h on B uniquely determine by analytic continuation its value on all of $L \cap N^{o}(T)$.

(Straight lines play a special role in the above lemma. However we note that there is a valid generalization of the above lemma in which L can be replaced by a suitable one dimensional curve determined as the locus of points satisfying $(n - 1)$ simultaneous analytic equations (C. Earle (1980), personal communication). For example L may be taken to be the curve $x_1^2 + x_2^2 = 1$, $x_3 = \ldots = x_k = 0$.)

2.9 Example

A question which arises, in statistical estimation theory, is whether the positive part James-Stein estimator for an unknown normal mean,

$$\delta(x) = (1 - (k-2)||x||^{-2})^{+}x, \quad x \in R^k,$$

can possibly be generalized Bayes for squared error loss. This is equivalent to asking whether $\delta(\cdot)$ can be the gradient of a cumulant generating function for some measure $\nu(d\theta)$ having $N = R^k$. (Note interchange of roles of θ and x.) See Theorem 4.16. The answer is, "No." To see this note that $\delta(x) \equiv 0$ for $||x|| \le 1$. Hence if $\delta(x) = \nabla\psi(x) = \nabla\lambda(x)/\lambda(x)$ for $||x|| < 1$ it follows by analyticity that $\psi(x) \equiv 0$ on its domain of analyticity, which in this case is R^k. This implies $\delta(x) \equiv 0$, a contradiction. ||

2.10 Example

The question arises in the theory of hypothesis tests as to whether the unit square,

$$S = \{x \in R^k : |x_i| \le 1\}, \qquad k \ge 2,$$

can be a Bayes acceptance region for testing the mean of a normal distribution. Placed in a general context, the question is whether there exist two distinct non-zero finite measures G_0 and G_1 (concentrated on disjoint sets Θ_0 and $\Theta_1 \subset R^k$) such that

(1) $$d(x) = \int e^{\theta\cdot x-\theta^2/2}(G_0(d\theta) - G_1(d\theta)) \geq 0 \quad \text{if } x \in S,$$

and $d(x) \leq 0$ if $x \notin S$. The answer is, "No."

Proof. Let $\mu_i(d\theta) = e^{-\theta^2/2}G_i(d\theta)$, $i = 0,1$. Then $d(x) = \lambda_0(x) - \lambda_1(x)$ where λ_i is the moment generating function of μ_i. Note that $N_{\mu_i} = R^k$, $i = 0,1$. Hence $d(\cdot)$ is analytic on R^k.

For convenience consider only the case $k = 2$. Expand d in a Taylor series about (1,1) as

$$d((1, 1) + (y_1, y_2)) = \sum_{i=0}^{\infty} \sum_{j=0}^{i} a_{j,i-j}\, y_1^j y_2^{i-2}$$

($a_{00} = 0$ since $d((1, 1)) = 0$.) Let i' be the smallest index for which $\sum_{j=0}^{i'} |a_{j,i'-j}| > 0$. i' exists since $d \not\equiv 0$ if (1) is valid.

Suppose i' is even. Then for $y_i \geq 0$, $i = 1,2$,

(2) $$\sum_{j=0}^{i'} a_{j,i'-j}\, y_1^j y_2^{i'-j} = \sum_{j=0}^{i'} a_{j,i'-j}(-y_1)^j(-y_2)^{i'-j}.$$

There are values (y_1, y_2) in the first quadrant for which (2) $\neq 0$, since (2) is a non-zero homogeneous polynomial. Suppose (y_1^0, y_2^0) is such a value. Then

$$|\rho|^{-i'}d((1, 1)) + (\rho y_1^0, \rho y_2^0)) = \sum_{j=0}^{i'} a_{i',i'-j}(y_1^0)^j(y_2^0)^{i'-j} + o(\rho)$$

$$= c + o(\rho) \quad \text{as } |\rho| \to 0$$

with $c \neq 0$. If $c > 0$ it follows that $d((1, 1)) + (\rho y_1^0, \rho y_2^0)) > 0$ for $\rho > 0$ sufficiently small; and this would contradict (1). If $c < 0$ it follows that $d((1, 1) + (\rho y_1^0, \rho y_2^0)) < 0$ for $\rho < 0$ sufficiently small; and this would also contradict (1).

If i' is odd analogous reasoning yields

$$||y||^{-i'}\, d((1, 1) + (y_1, -y_2)) = ||y||^{-i'}d((1, 1) + (-y_1, y_2)) + o(1)$$

as $||y|| \to 0$, and that there are values of $(y_1^0, -y_2^0) > 0$ for which $\lim_{\rho\downarrow 0} ||y||^{-i'}d((1, 1) + \rho(y_1^0, -y_2^0)) \neq 0$. It follows that there are values of

y in either the fourth quadrant or the second quadrant for which $d((1, 1) + y) > 0$. This again contradicts (1).

Hence (1) is impossible. ||

COMPLETENESS

2.11 Remarks

A family $\{F_\theta : \theta \in \Theta\}$ of probability distributions (or their associated densities, if these exist) is called statistically *complete* if $T : R^k \to R$ with

(1) $$\int T(x)F_\theta(dx) = 0 \qquad \forall\ \theta \in \Theta$$

implies

(2) $$T(x) = 0 \quad \text{a.e.} \quad (F_\theta) \qquad \forall\ \theta \in \Theta \quad .$$

(Implicit in (1) is the condition that $\int |T(x)|F_\theta(dx) < \infty \ \ \forall\ \theta \in \Theta$.)

Standard exponential families are complete if the parameter space is large enough. This result, which is equivalent to the uniqueness theorem for Laplace transforms, is proved in Theorem 2.12. (The uniqueness theorem for Laplace transforms states that if $N_\mu^\circ \cap N_\nu^\circ \neq \phi$ then $\lambda_\mu = \lambda_\nu$ if and only if $\mu = \nu$.) The most convenient way to prove this theorem seems to be to invoke the uniqueness theorem for Fourier-Stieltjes transforms (equals characteristic functions) which is described in the next paragraph.

Let $\mathrm{Im} = \{bi \in \mathbb{C} : b \in R\}$ denote the pure imaginary numbers. Let F be a finite (non-negative) measure on R^k. The function $\kappa : R^k \to \mathbb{C}$ defined by

$$\kappa_F(b) = \lambda_F(bi) \qquad b \in R^k$$

is the *Fourier-Stieltjes transform* (or, Fourier transform, or, *characteristic function*) of F. Hence λ_F restricted to the domain $(\mathrm{Im})^k$ is equivalent to κ_F. Note that κ_F always exists (i.e. $\mathrm{Re}((\mathrm{Im})^k) = 0 \subset N$). The uniqueness theorem for Fourier transforms is as follows.

Theorem. Let F and G be two finite non-negative measures on R^k. Then $F = G$

if and only if $\kappa_F \equiv \kappa_G$ (i.e. $\lambda_F(bi) = \lambda_G(bi)\ \forall\ b \in R^k$).

Proof. This is a standard result in the theory of characteristic functions. Proofs abound. A quick proof may be found in Feller (1966, XV,3). (This proof is explicitly for R, but generalizes immediately to R^k.) ||

Here is the classic result on completeness of exponential families.

2.12 Theorem

Let $\{p_\theta\}: \theta \in \Theta\}$ be a standard exponential family. Suppose $\Theta^\circ \neq \phi$. Then $\{p_\theta\}$ is complete.

Proof. Let $\theta_0 \in \Theta^\circ$. One may translate coordinates using Proposition 1.6 so that $\theta_0 = 0$. There is thus no loss of generality in assuming $\theta_0 = 0$.

Suppose $\int T(x)p_\theta(x)\nu(dx) = 0\ \forall\ \theta \in \Theta$. Then, letting $T = T^+ - T^-$,

$$(1) \qquad \int T^+(x)e^{\theta\cdot x}\nu(dx) = \int T^-(x)e^{\theta\cdot x}\nu(dx) \qquad \forall\ \theta \in \Theta\ .$$

Let $F(dx) = T^+(x)\nu(dx)$, $G(dx) = T^-(x)\nu(dx)$. Then (1) becomes

$$(2) \qquad \lambda_F(\theta) = \int e^{\theta\cdot x}F(dx) = \int e^{\theta\cdot x}G(dx) = \lambda_G(\theta) \qquad \forall\ \theta \in \Theta$$

Both $\lambda_F(\cdot)$ and $\lambda_G(\cdot)$ are analytic on the domain $\Theta^\circ \times (\mathrm{Im})^k$. (2) states that they agree on $\Theta \times 0 \subset \Theta \times (\mathrm{Im})^k$. Hence $\lambda_F(x) = \lambda_G(z)$ for all z such that re $z \in \Theta^\circ$. (This follows directly from analyticity. Alternately one may apply the second half of Lemma 2.8 to all lines which intersect Θ .) In particular

$$(3) \qquad \lambda_F(0 + bi) = \lambda_G(0 + bi) \qquad \forall\ b \in R^k$$

since $0 \in \Theta^\circ$. Thus, F = G by Theorem 2.11. This says that $T^+(x)\nu(dx) = T^-(x)\nu(dx)$, which implies $T^+ = T^-$ a.e.(ν), which implies $T = 0$ a.e.(ν). Hence $\{p_\theta\}$ is complete. ||

Note from the above that any canonical family is complete. From this we derive:

2.13 Corollary

A standard family with $N^\circ \neq \phi$ is uniquely determined by its Laplace transform (or by its cumulant generating function).

Note that the corollary applies to all minimal families since they always have $N^\circ \neq \phi$.

Proof. Consider the standard families in R^k generated by the measures μ and ν. Suppose $N^\circ_\mu \neq \phi$ and $\psi_\mu = \psi_\nu$. Then $N_\mu = N_\nu = N$.

Let $\omega = (\mu + \nu)/2$. Then, ω generates an exponential family with $\lambda_\omega = (\lambda_\mu + \lambda_\nu)/2$. Hence $N_\omega = N$ and $\psi_\omega = \psi_\mu = \psi_\nu$.

Let $T = \frac{d\mu}{d\omega} - \frac{d\nu}{d\omega}$. Then

$$\int T(x)e^{\theta\cdot x-\psi_\omega(\theta)}\omega(dx) = (\tfrac{1}{2})(\int e^{\theta\cdot x-\psi_\mu(\theta)}\mu(dx) - \int e^{\theta\cdot x-\psi_\nu(\theta)}\nu(dx))$$

$$= \frac{1}{2} - \frac{1}{2} = 0 \qquad \forall\ \theta \in N$$

Hence $T = 0$ a.e.(ω) by Theorem 2.12; which implies $\mu = \nu$. ||

Theorem 2.12 has many other important applications in statistics. It plays an important role, for example, in the theory of unbiased estimates and in the construction of unbiased tests. Some aspects of this role are described in the exercises and in succeeding chapters.

MUTUAL INDEPENDENCE

Lehamnn (1959, p. 162-163) describes a nice proof of the independence of $\bar{X}$, S^2 in a normal sample. A different but related proof is a special instance of an argument which applies in several important exponential families. (See Example 2.15.) The basic parts of the argument are due to Neyman (1938) and Basu (1955), but the full result in Theorem 2.14, below, was only recently proved by Bar-Lev (1983) and by Barndorff-Nielsen and Blaesild (1983). The proof below follows that in the second of these papers. See the exercises for an additional related result of Bar-Lev and for several applications of this theorem.

Through most of this subsection we consider the situation where

θ and x are, respectively, partitioned as $\theta' = (\theta'_{(1)}, \theta'_{(2)})$, $x' = (x'_{(1)}, x'_{(2)})$. As in Sections 1.7 and 1.15, problems not in this form can sometimes be reduced to this form through use of linked linear transformations on θ and x. Where convenient, we write $\psi(\theta) = \psi(\theta_{(1)}, \theta_{(2)})$. We use the notation $Y \sim \text{Expf}\,(\theta)$ to mean that the distributions of Y form a standard exponential family with natural parameter θ. We also use the notation $X \perp Y$ to mean that X and Y are independent.

2.14 Theorem

Let $X \sim \text{Expf}\,(\theta)$ with $\theta^\circ \in \Theta^\circ$. Let $X' = (X'_{(1)}, X'_{(2)})$ where $X_{(i)}$ is k_i dimensional, and let $h(X_{(1)})$ be a k_2 dimensional statistic. Let

$$\rho_1(\theta_{(1)}, \theta_{(2)}) = \log E_\theta\,(\exp((\theta_{(1)} - \theta^\circ_{(1)}) \cdot X_{(1)} + (\theta_{(2)} - \theta^\circ_{(2)}) \cdot h(X_{(1)}))) \tag{1}$$

$$\rho_2(\theta_{(2)}) = \log E_\theta\,(\exp((\theta_{(2)} - \theta^\circ_{(2)}) \cdot (X_{(2)} - h(X_{(1)}))) .$$

Then the following conditions are equivalent:

$$X_{(1)} \perp (X_{(2)} - h(X_{(1)})) \quad \text{under} \quad \theta^\circ \tag{2}$$

$$X_{(1)} \perp (X_{(2)} - h(X_{(1)})) \quad \text{for all} \quad \theta \in \Theta \tag{2'}$$

$$\psi(\theta_{(1)}, \theta_{(2)}) = \rho_1(\theta_{(1)}, \theta_{(2)}) + \rho_2(\theta_{(2)}) \quad \forall \quad \theta \in \Theta \tag{3}$$

$$(X_{(2)} - h(X_{(1)})) \sim \text{Expf}\,(\theta_{(2)}) \tag{4}$$

$$(X_{(1)}, h(X_{(1)})) \sim \text{Expf}\,(\theta_{(1)}, \theta_{(2)}) \quad . \tag{5}$$

Proof. For convenience, assume without loss of generality that $\theta^\circ = 0$. (See Proposition 1.6.) Let ω denote the joint distribution under 0 of $V = (X_{(1)}, h(X_{(1)}), X_{(2)} - h(X_{(1)}))$. Consider the standard exponential family generated by ω, with natural parameter space N_V. Note that, in general,

$\{X_{(1)} \perp (X_{(2)} - h(X_{(1)}))\} \Leftrightarrow \{(X_{(1)}, h(X_{(1)})) \perp (X_{(2)} - h(X_{(1)}))\}$. The equivalence of (2) and (2') is seen in this fashion to be a special case of Exercise 1.7.1.

(2) $\Rightarrow$ (3) follows from a direct calculation.

(3) $\Rightarrow$ (2): Let ω_1 denote the distribution under $\theta^\circ = 0$ of $(V_{(1)}, V_{(2)}) = (X_{(1)}, h(X_{(1)}))$ and ω_2 that of $V_{(3)} = X_{(1)} - h(X_{(1)})$. Let $\omega^* = \omega_1 \times \omega_2$. Then the cumulant generating function ψ^* of ω^* satisfies

$$\psi^*(\theta_{(1)}, \theta_{(2)}, \theta_{(2)}) = \rho_1(\theta_{(1)}, \theta_{(2)}) + \rho_2(\theta_{(2)}), \quad (\theta_{(1)}, \theta_{(2)}) \in \Theta \quad .$$

Furthermore, the cumulant generating function of the linear function $(V_{(1)}, V_{(2)} + V_{(3)})$ is ψ^{**} given by

$$\begin{aligned}\psi^{**}(\theta_{(1)}, \theta_{(2)}) &= \psi^*(\theta_{(1)}, \theta_{(2)}, \theta_{(2)}) = \rho_1(\theta_{(1)}, \theta_{(2)}) + \rho_2(\theta_{(2)}) \\ &= \psi(\theta_{(1)}, \theta_{(2)}) , \qquad \theta \in \Theta \quad .\end{aligned}$$

It follows from Corollary 2.13, since $\Theta^\circ \neq \phi$, that $(V_{(1)}, V_{(2)} + V_{(3)})$ has the same distribution under θ° as $(X_{(1)}, X_{(2)})$. Thus $(X_{(1)}, X_{(2)} - h(X_{(1)}))$ has the same joint distribution under θ° as $(V_{(1)}, V_{(2)} + V_{(3)} - h(V_{(1)}))$. But, $V_{(2)} + V_{(3)} - h(V_{(1)}) = V_{(3)}$. Hence $X_{(1)} \perp (X_{(2)} - h(X_{(1)}))$ under θ° since $V_{(1)} \perp V_{(3)}$.

(2) $\Rightarrow$ (4) and (5), as can be seen by direct calculation of the marginal distributions involved via the standard formulae (6) and (8), below.

(4) $\Rightarrow$ (2): The marginal density of $V_{(3)} = X_{(2)} - h(X_{(1)})$ relative to the marginal distribution ω_2 is

$$(6) \qquad q_\theta(v_{(3)}) = \int \exp(\theta_{(1)} \cdot v_{(1)} + \theta_{(2)} \cdot h(v_{(1)}) + \theta_{(2)} \cdot v_{(3)} - \psi(\theta))\, \omega(dv_{(1)} \mid v_{(3)}) \qquad \text{(a.e.)}$$

where $\omega(\cdot|\cdot)$ denotes the indicated conditional distribution. By (4) $q_\theta(v_{(3)}) = \exp(\theta_{(2)} v_{(3)} - \rho_2(\theta_{(2)}))$ (a.e.). Setting $\theta_{(2)} = 0$ yields

(7) $\exp(\psi(\theta_{(1)}, 0) - \rho_2(0))$

$$= \int \exp(\theta_{(1)} \cdot v_{(1)})\, \omega(dv_{(1)}|v_{(3)}) , \quad (\theta_{(1)}, 0) \in \Theta , \quad \text{(a.e.)} .$$

Here the Laplace transform of $\omega(\cdot|v_{(3)})$ exists on an open set and is independent of $v_{(3)}$ (a.e.). It follows from another application of Corollary 2.13 that $\omega(\cdot|v_{(3)})$ is independent of $v_{(3)}$ (a.e.). So, $V_{(1)}$ is independent of $V_{(3)}$. This verifies (2).

The proof that (5) ⇒ (2) is similar. The marginal joint density of $V_{(1)}$, $V_{(2)}$ is

(8) $$q'_\theta(v_{(1)}, v_{(2)}) = \int \exp(\theta_{(1)} \cdot v_{(1)} + \theta_{(2)} \cdot h(v_{(1)}) + \theta_{(2)} \cdot v_{(3)}$$

$$- \psi(\theta))\, \omega'(dv_{(3)}|v_{(1)}) \quad \text{(a.e.)} .$$

Setting $\theta_{(1)} = 0$ and cancelling terms in (5) implies

$$\exp(\psi(0, \theta_{(2)}) - \rho(0, \theta_{(2)})) = \int \exp(\theta_{(2)} \cdot v_{(3)})\, \omega'(dv_{(3)}|v_{(1)}) \quad \text{(a.e.)}.$$

Hence, as before $\omega'(\cdot|v_{(1)})$ is independent of $v_{(1)}$ (a.e.), which yields (2). ||

2.15 Examples

(i) Let $Y_1,\ldots,Y_n$ be independent $N(\mu,\sigma^2)$ variables. Then (Example 1.12) $(Y_i, Y_i^2) \sim \text{Expf}(\mu/\sigma^2, -1/2\sigma^2)$. Hence $(\Sigma Y_i, \Sigma Y_i^2) \sim \text{Expf}(\mu/\sigma^2, -1/2\sigma^2)$. Also $(\Sigma Y_i, (\Sigma Y_i)^2/n) \sim \text{Expf}(\mu/\sigma^2, -1/2\sigma^2)$. This verifies 2.14(5). Hence $\Sigma Y_i - \Sigma Y_i^2/n = \Sigma(Y_i - \bar{Y})^2 \sim \text{Expf}(-1/2\sigma^2)$ and is independent of T_n by 2.14(4) and 2.14(2').

(ii) Similarly, let $X_1,\ldots,X_n$ be independent $\Gamma(\alpha, \sigma)$. Then (Example 1.12) $\Gamma(\Sigma X_i, \Sigma \ln X_i) \sim \text{Expf}(-1/\sigma, n\alpha)$. The marginal distribution of X_i is also $(n\alpha, \sigma)$; hence $(\Sigma X_i, \ln \Sigma X_i) \sim \text{Expf}(-1/\sigma, n\alpha)$. Again, Theorem 2.14 yields that $(\Sigma \ln X_i - \ln \Sigma X_i) \perp \Sigma X_i$. This is often re-expressed in the form $\tilde{X}/\bar{X} \perp \bar{X}$ where here $\tilde{X} = (\prod_{i=1}^{n} X_i)^{1/n}$ denotes the geometric mean of the observations. Also, $\ln(\tilde{X}/\bar{X}) \sim \text{Expf}(n\alpha)$. See the Exercises for a double extension of this conclusion.

There are further applications of this theorem. For some of these see the exercises and the references cited above. In particular there are several applications to problems involving the inverse Gaussian distribution. See Chapter 3.

CONTINUITY THEOREM

The continuity theorem for Laplace transforms refers to the limiting behavior of a sequence of measures and the associated Laplace transforms. We first need a standard definition and some related remarks.

2.16 Definition

Consider R^k. Let C denote the space of continuous (real-valued) functions on R^k. Let $C_0 \subset C$ denote the subspace of continuous functions with compact support -- i.e.

$$c(x) = 0 \quad \text{for } ||x|| > r, \quad \text{some } r < \infty .$$

A (non-negative) measure ν is called *locally finite* if $\nu(\{x : ||x|| \le r\}) < \infty \ \forall \ r \in R$. Except where specifically noted, all measures are assumed to be locally finite, σ-finite, and non-negative. Let $\{\nu_n\}$ be a sequence of measures. We say

$$\nu_n \to \nu \ \text{(weak*)} \quad \text{if}$$

(1) $$\int c(x)\nu_n(dx) \to \int c(x)\nu(dx) \quad \forall \ c \in C_0 .$$

Here are several important facts concerning weak* convergence.

For ν finite let V_ν denote the *cumulative distribution function*: $V_\nu(t) = \nu(\{x : x_i \le t_i, \ i=1,\ldots,k\})$.

(i) Then $\nu_n \to \nu$ if and only if

(2) $$V_{\nu_n}(t) \to V_\nu(t) \quad \forall \ t \in R^k \quad \text{at which } V_\nu(\cdot) \text{ is continuous.}$$

(ii) Suppose $\nu_n \to \nu$. Then $\liminf_{n\to\infty} \nu_n(R^k) \ge \nu(R^k)$. Suppose there is a $c \in C$, $c \geq 0$, with

(3) $$\lim_{||x||\to\infty} c(x) = \infty$$

such that

(3') $$\limsup_{n\to\infty} \int c(x)\nu_n(dx) < \infty \quad .$$

Then

(4) $$\lim_{n\to\infty} \nu_n(R^k) = \nu(R^k) < \infty \quad .$$

(iii) Furthermore, (4) implies

(5) $$\int c(x)\nu_n(dx) \to \int c(x)\nu(dx)$$

for all bounded $c \in C$. (Condition (3), (3') is sometimes referred to by saying the sequence is *tight*.)

(iv) If $\nu_n \geq 0$ is any bounded sequence (i.e. $\limsup_{n\to\infty} \nu_n(R^k) < \infty$) then there is a subsequence $\{\nu_{n_i}\}$ and a finite measure ν such that $\nu_{n_i} \to \nu$.

For a proof of these facts see Neveu (1965).

2.17 Theorem

Let $S \subset R^k$ and let $B = \text{conhull } S$. Suppose $B^\circ \neq \phi$. Let ν_n be a sequence of measures on R^k such that

(1) $$\liminf_{n\to\infty} \sup_{b\in S} \lambda_{\nu_n}(b) < \infty \qquad \forall\ b \in S\ .$$

Let $b_0 \in B^\circ$. Then there exists a subsequence $\{n_i\}$ and a locally finite measure ν such that

(2) $$e^{b_0\cdot x}\ \nu_{n_i}(dx) \to e^{b_0\cdot x}\nu(dx)$$

and

(3) $$\lambda_{\nu_{n_i}}(b) \to \lambda_\nu(b) \qquad \forall\ b \in B^\circ \quad .$$

The convergence in (3) is uniform on compact subsets of B°, (Condition (3) is of course equivalent to

$$\psi_{\nu_{n_i}}(b) \to \psi_\nu(b) \quad \forall \ b \in B^\circ \qquad .$$

Condition (1) implies the measures ν_n are locally finite.)

Remark. Lemma 2.1 together with (3) shows that

$$\int xe^{b\cdot x}\,\nu_{n_i}(dx) \to \int xe^{b\cdot x}\nu(dx), \qquad b \in B^\circ \quad , \tag{4}$$

and similarly for higher moments of x. Hence

$$\nabla\lambda_{\nu_{n_i}}(b) \to \nabla\lambda_\nu(b), \qquad b \in B^\circ \tag{5}$$

and similarly for higher order partial derivatives of λ. See Exercise 2.17.1.

Similar reasoning also shows that

$$e^{\theta\cdot b}\nu_{n_i}(d\theta) \to e^{\theta\cdot b}\nu(d\theta) \quad \text{weak*} \quad \forall \ b \in B_0 \quad . \tag{6}$$

Hence the measure ν in (2) does not depend on the choice of $b_0 \in B_0$.

Proof. We exploit Proposition 1.6 and assume without loss of generality that $b_0 = 0 \in B^\circ$. It also suffices to assume that B is a convex polytope (i.e. B = conhull $\{b_i : i=1,\ldots,m\}$) since the interior of any convex set is a countable union of such polytopes, and a compact subset of the interior will be contained in one of them.

Now,

$$\lim_{||x||\to\infty} \sum_{i=1}^m e^{b_i\cdot x} = \infty$$

by Lemma 2.1. Thus, for some subsequence $\{n_j'\}$

$$\limsup_{n\to\infty} \int\left(\sum_{i=1}^m e^{b_i\cdot x}\right)\nu_{n_j'}(dx) < \infty \qquad , \tag{7}$$

by (1). Hence, the sequence $\{\nu_{n_j'}\}$ is tight, and there exists a further subsequence $\{\nu_{n_i}\}$ and a limiting measure ν such that $\nu_{n_i} \to \nu$. This immediately implies that also $e^{b\cdot x}\nu_{n_i}(dx) \to e^{b\cdot x}\nu(dx)$ for any $b \in R^k$.

Let $b \in B^\circ$. Then $\lim_{||x||\to\infty} (\Sigma e^{b_i \cdot x}/e^{b\cdot x}) = \infty$, again by Lemma 2.1. As in (7)

$$\limsup_{n\to\infty} \int \frac{\Sigma e^{b_i \cdot x}}{e^{b\cdot x}} e^{b\cdot x} \nu_{n_i}(dx) < \infty .$$

Hence the sequence $e^{b\cdot x}\nu_{n_i}(dx)$ is also tight. This implies $\int e^{b\cdot x}\nu_{n_i}(dx) \to \int e^{b\cdot x}\nu(dx)$, which yields (3).

Let $C \subset B^\circ$ be compact. Then

$$||x|| e^{b\cdot x} \leq K \sum_{i=1}^{m} e^{b_i \cdot x}$$

by Lemma 2.1. This yields

$$\limsup_{i\to\infty} \sup_{b\in C} ||\nabla\lambda_{\nu_{n_i}}(b)|| \leq \limsup_{i\to\infty} K \int \sum_{i=1}^{m} e^{b_i \cdot x} \nu_{n_i}(dx)$$

$$\leq \limsup_{i\to\infty} K \sum_{i=1}^{m} \lambda(b_i) < \infty .$$

The functions $\lambda_{\nu_{n_i}}(\cdot)$ are thus uniformly (in $\{n_i\}$) uniformly continuous on C. The convergence in (3) is therefore uniform on C. ||

2.18 Uniform Convergence

Theorem 2.17 shows that if

$$\psi_{\lambda_i}(b) \to \psi_\lambda(b) \quad \text{for all} \quad b \in B^\circ \neq \phi \quad \text{then} \quad \nu_i \to \nu$$

There is a useful uniform version of this statement. Let

(1) $$\{\nu_{\alpha n} : n=1,\ldots,\alpha \in A\}$$

be a family of sequences of measures and $\{\nu_\alpha : \alpha \in A\}$ be a family of measures. All of these are assumed locally finite. We say

$$\nu_{\alpha n} \to \nu_\alpha \text{ (weak*) uniformly in } \alpha$$

when for each $c \in C_0$

(2) $$\int c(x)\nu_{\alpha n}(dx) \underset{n\to\infty}{\to} \int c(x)\nu_\alpha(dx)$$

uniformly over $\alpha \in A$. For notational convenience in the following, let $V_\alpha = V_{\nu_\alpha}$, etc.

Proposition. Suppose the family of cumulative distribution functions $\{V_\alpha : \alpha \in A\}$ is equicontinuous at every $x \in R^k$. Then $\nu_{\alpha n} \to \nu_\alpha$ uniformly in α if and only if

(3) $$V_{\alpha n} \to V_\alpha \quad \text{uniformly for} \quad \alpha \in A$$

Proof. The necessity of (3) is proved by applying (2) to continuous functions c satisfying

$$c(x) = \begin{cases} 1 & x_i \le x_{0i} - \delta \quad \text{for all } i=1,\dots,k \\ 0 & x_i > x_{0i} + \delta \quad \text{for some } i=1,\dots,k \end{cases}$$

and then choosing δ sufficiently small.

Conversely, (3) implies $\int g(x)d(V_{\alpha n}(x) - V_\alpha(x)) = \int (V_{\alpha n}(x) - V_\alpha(x))dg(x) \to 0$ uniformly in α for each differentiable $g \in C_0$. If $c \in C_0$ and $\varepsilon > 0$ there is a differentiable $g \in C_0$ with $|g - c| < \varepsilon$. Then $|\int(c(x) - g(x))\, d(V_{\alpha n}(x) - V_\alpha(x))| < 2\varepsilon$ uniformly for all $\alpha \in A$ and all n. Combining these facts yields the uniform convergence of $\nu_{\alpha n}$ to ν_α. ||

Extra care in the proof of the above proposition will show that if the $\{V_\alpha : \alpha \in A\}$ are equicontinuous uniformly over $x \in S$ and $\nu_{\alpha n} \to \nu_\alpha$ uniformly in α then (3) holds uniformly for $\alpha \in A$, $x \in S$.

2.19 Theorem

Let $\{\nu_{\alpha n}\}$ and $\{\nu_\alpha\}$ be as in 2.18(1). Suppose B = conhull S, and $B^\circ \ne \phi$. Let $\lambda_\alpha = \lambda_{\nu_\alpha}$, etc. Suppose

(1) $$\lambda_{\alpha n}(b) \underset{n\to\infty}{\to} \lambda_\alpha(b) \qquad \forall\ b \in S$$

uniformly over $\alpha \in A$, and suppose

(2) $$\sup_{b\in S} \sup_{\alpha} \lambda_\alpha(b) < \infty \quad .$$

Then $\nu_{\alpha n} \to \nu_\alpha$ uniformly over $\alpha \in A$.

Proof. If $\nu_{\alpha n} \not\to \nu_\alpha$ uniformly over $\alpha \in A$, there is a $c \in C_0$ and a sequence α_n such that

(3) $$\lim_{n\to\infty} |\int c(\theta)(\nu_{\alpha_n n}(d\theta) - \nu_{\alpha_n}(d\theta))| > 0 \quad .$$

In view of (3) there exists a subsequence n_i and limiting measures $\nu_1^* \neq \nu_2^*$ such that if we write $\nu_{\alpha_{n_i} n_i} = \omega_i$ and $\nu_{\alpha_{n_i}} = \bar{\omega}_i$ then

(4) $$\omega_i \to \nu_1^* , \qquad \lambda_{\omega_i}(b) \to \lambda_{\nu_1^*}(b) , \qquad b \in B ;$$

and

(5) $$\bar{\omega}_i \to \nu_2^* , \qquad \lambda_{\bar{\omega}_i}(b) \to \lambda_{\nu_2^*}(b) \qquad b \in B \quad ,$$

by Theorem 2.17. (To establish (4) we exploit (2) to guarantee condition 2.17(1) for the sequence $\{\omega_{n_i}\}$.)

Assumption (1) implies $\lambda_{\nu_1^*}(b) = \lambda_{\nu_2^*}(b)$, $b \in B$, which implies $\nu_1^* = \nu_2^*$. This is a contradiction. It follows that $\nu_{\alpha n} \to \nu_\alpha$ uniformly over $\alpha \in A$. ||

TOTAL POSITIVITY

2.20 Definitions

Let $S \subset R$ and $h : S \to R$. Let $\{x_0 < \ldots < x_n\} \subset S$. The sequence $\{x_i \in S : i=0,1,\ldots,n\}$ is called a strictly changing sequence for h having order n if

(1) $$(\operatorname{sgn} h(x_{i-1}))(\operatorname{sgn} h(x_i)) = -1 \qquad i=1,\ldots,n \quad .$$

The number $S^-(h)$ -- *the number of strict sign changes of* h -- is the maximal order of a sequence of strict sign changes of h. Clearly $0 \le S^-(h) \le \infty$. Let $S^-(h) = n < \infty$ and let $\{x_i \in S : i=0,\ldots,n\}$ be a strictly changing sequence for h having order n. Then *the (strict) initial sign of* h is

$$(2) \qquad IS^-(h) = \operatorname{sgn} h(x_0) \quad .$$

(It is easy to check that this definition is well-formulated -- i.e. does not depend on the chosen strictly changing sequence for h.)

Similarly a sequence $\{x_i \in S : i=0,\ldots,n\}$ is called a weakly changing sequence for h having order n if

$$(3) \qquad (\operatorname{sgn} h(x_{2i}))(\operatorname{sgn} h(x_{2j+1})) \le 0$$

$$\text{for} \quad i=0,\ldots,[n/2], \quad j=0,\ldots,[(n-1)/2] \quad .$$

This means that zeros of the sequence $\{\operatorname{sgn} h(x_i) : i=0,1,\ldots,n\}$ can be reassigned as either a (+1) or a (-1) in a manner so that the resulting sequence of ±1's alternates in sign. The number $S^+(h)$ is the maximal order of such a sequence. Clearly, $0 \le S^+(h) \le \infty$, and

$$(4) \qquad S^+(h) \ge S^-(h) \quad .$$

Let $S^+(h) = n < \infty$ and let $\{x_i \in S : i=0,\ldots,n\}$ be a weakly changing sequence for h of order n. Then

$$(5) \qquad IS^-(h) = \begin{cases} +1 & \text{if} \quad h(x_{2i}) > 0 \quad \text{for some} \quad i=0,\ldots,[n/2] \\ 0 & \text{if} \quad h(x_i) \equiv 0 \quad i=0,\ldots,n \\ -1 & \text{if} \quad h(x_{2i}) < 0 \quad \text{for some} \quad i=0,\ldots,[n/2] \quad . \end{cases}$$

It can be checked that this definition is well formulated.

2.21 Theorem

Let $\{p_\theta\}$ be a standard one parameter exponential family. Let $g : R \to R$ such that $\nu\{x : g(x) \ne 0\} > 0$. Let

(1) $$h(\theta) = E_\theta(g(x)) , \quad \theta \in N°(g) .$$

Then

(2) $$S^+(h) \leq S^-(g) .$$

If equality holds in (2) then

(3) $$IS^+(h) = IS^-(g) .$$

Remark. The domain of h in (1) is restricted to $N°(g)$. The theorem remains true if the domain of h is all of $N(g)$. We leave this generalization as an exercise.

The sign-change-preserving properties (2), (3) are equivalent to "Total Positivity of $\{p_\theta\}$ of order ∞." Karlin (1968) is a very useful, standard reference on this topic. See also Brown, Johnstone, and MacGibbon (1981).

Proof. Let
$$\hat{g}(\theta) = \int e^{\theta x} g(x)dx = e^{\psi(\theta)}h(\theta) .$$

It suffices to prove $\hat{g}$ has the properties of h in (2), (3). The proof is by induction on $n = S^-(g)$. Assume without loss of generality that $IS^-(g) = +1$.

When $n = 0$ the result is trivial since then $g \geq 0$ and $\nu(\{x : g(x) > 0\}) > 0$ so that $\hat{g}(\theta) > 0$ for all $\theta \in N(h)$, as claimed in (2).

Assume the theorem is true for $n \leq N$. Suppose $n = N + 1$. Let $\xi_1 = \inf\{x : g(x) < 0\}$. $\xi_1 > -\infty$ since $IS^-(g) = +1$. Let

$$u(\theta) = \frac{d}{d\theta}(e^{-\theta\xi_1}\hat{g}(\theta)) = \int (x - \xi_1)g(x)e^{\theta x}\nu(dx) .$$

Now, $S^-((x - \xi_1)g(x)) \leq N = n - 1$, as can easily be checked from the definition of ξ_1. Hence $S^+(u) \leq N$ by the induction hypothesis. Integration yields that $S^+(U) \leq N + 1$ where

(4) $$U(\theta) = \int^\theta u(t)dt = e^{-\xi_1\theta}\,\hat{g}(\theta) .$$

(2) follows from (4). (3) may be verified by concentrating the above argument on the case where $S^+(u) = N$ and $S^+(U) = N + 1$, and using the induction hypothesis

to keep track of $IS^+(u)$ and consequently of $IS^+(U)$. ||

The above property for $n = 1$ is equivalent to the strict monotone likelihood ratio property. The following is an important consequence of this.

<u>2.22 Corollary</u>

Let $\{p_\theta\}$ be a standard one parameter exponential family. Suppose $g : R \to R$ is non-decreasing and not essentially a constant (ν). Then $E_\theta(g)$ is strictly increasing on $N°(g)$.

(*Remark.* Again, the result is true on the full domain, $N(g)$, but we leave verification of this as an exercise.)

Proof. Let ess inf $g(\cdot) < c <$ ess sup $g(\cdot)$; then $g(\cdot) - c$ satisfies the hypotheses of Theorem 2.21 with $S^-(g-c) = 1$. Hence $E_\theta(g) - c > 0$ (or < 0) for $\theta \in N°(g)$ whenever $\theta > \theta_1(c)$ (whenever $\theta < \theta_1(c)$). It follows that g is strictly increasing on $N°(g)$. ||

It is possible to derive from the above some results concerning sign changes for multidimensional families. In general, these results appear very weak by comparison with their univariate cousins. Here is an example of such a result which will be useful later.

<u>2.23 Corollary</u>

Let $\{p_\theta\}$ be a standard k parameter exponential family. Let $\theta_0 \in N$ and $v \in R^k$. Let $\theta_\rho = \theta_0 + \rho v$. Suppose $g : R^k \to R$ satisfies

$$(1) \qquad g(x) \begin{cases} \leq 0 & v \cdot x \leq \alpha \\ \geq 0 & v \cdot x \geq \alpha \end{cases}$$

for some $\alpha \in R$. Let

$$h(\rho) = E_{\theta_\rho}(g(X)) \quad .$$

Then $S^+(h) \leq 1$. If $S^+(h) = 1$ then $IS^+(h) = -1$.

Proof. Apply Theorem 2.22 to the one parameter exponential family $\{p_{\theta_\rho}\}$ of densities of $v \cdot X$. Observe that

$$E_{\theta_\rho}(g(x) \mid v \cdot x = t) = g^*(t)$$

is independent of ρ by Theorem 1.7, and (1) guarantees that $S^-(g^*) \leq 1$. These observations enable the desired application of the theorem. ||

PARTIAL ORDER PROPERTIES

The preceding multidimensional result is not very satisfactory; the hypotheses on h are too restrictive. Better results may be obtained by considering partial orderings and imposing suitable restrictions on the exponential family. We give one simple result as an appetizer for what may be obtained.

For this result define the partial ordering, $\propto$, on R^k by $x \propto y$ if $x_i \leq y_i$, $i=1,\ldots,k$. A function $h : R^k \to R$ is non-decreasing relative to this ordering if $x \propto y$ implies $h(x) \leq h(y)$. The following preparatory lemma is also of independent interest.

2.24 Lemma

Let X have coordinates $X_1,\ldots,X_k$ which are independent random variables with distributions $F_1,\ldots,F_k$, respectively. Suppose h_1, h_2 are non-decreasing relative to the partial ordering $\propto$. Then

$$E(h_1(X)h_2(X)) \geq E(h_1(X))E(h_2(X)) \quad . \tag{1}$$

Proof. The proof is by induction on k. Note that for k = 1 the result is well known. This observation enables one to rewrite and reduce the left side of (1) as

$$\int\cdots\int h_1(x)h_2(x) \prod_{i=1}^{k} F_i(dx_i)$$

$$= \int\cdots\int \left(\int h_1(x)h_2(x)F_k(dx_k)\right) \prod_{i=1}^{k-1} F_i(dx_i)$$

$$\geq \int\cdots\int [\int h_1(x)F_k(dx_k)][\int h_2(x)F_k(dx_k)] \prod_{i=1}^{k-1} F_i(dx_i) \quad .$$

Each function in square brackets is clearly non-decreasing in $(x_1,\ldots,x_{k-1})$. Hence, by induction, (1) is valid. ||

Here is the application to exponential families.

2.25 Theorem

Consider a minimal standard exponential family for which the canonical coordinate variables $X_1,\ldots,X_k$ are independent. Let h be non-decreasing relative to the partial ordering $\propto$. Then $E_\theta(h)$ is a non-decreasing function of θ on $N^\circ(h)$. (This result may be extended to all of $N(h)$.)

Proof. Write

$$\frac{\partial}{\partial\theta_j} E_\theta(h) = \int(x_j - \xi_j(\theta))h(x)p_\theta(x)\nu(dx) \quad .$$

Note that both $x_j - \xi_j(\theta)$ and $h(x)$ are non-decreasing functions of x. Hence

$$\frac{\partial}{\partial\theta_j} E_\theta(h) \geq E(X_j - \xi_j(\theta))E_\theta(h(X)) = 0$$

by Lemma 2.24. It follows that $E_\theta(h)$ is non-decreasing in each coordinate of θ and hence (equivalently) is non-decreasing relative to $\propto$. ||

The preceding theorem is merely a sample of the available results. Other assumptions may replace the independence assumption, above. Notably, the conclusion of Lemma 2.24 remains valid if the joint distribution, F, of X has a density f with respect to Lebesgue measure which is monotone likelihood ratio in each pair of coordinates when the others are held fixed. (Exercise.) (There is also a lattice variable version of this fact.) Such densities are called multivariate totally positive of order 2 (= MTP_2). Suppose $\{p_\theta\}$ is a minimal standard exponential family whose dominating measure, ν, is MTP_2. It follows by the proof of the theorem above that then h non-decreasing implies $E_\theta(h)$ non-decreasing in θ.

Under suitable conditions it is also possible to derive analogous "order preserving" results for other partial orderings. For example, one may consider the partial ordering induced by a convex cone $C \subset R^k$, under which $x \propto_C y$ if $y - x \in C$.

A rather different but very fruitful partial ordering is that leading the notion of Schur convexity. Define $x \propto_S y$ if $\sum_{i=1}^{k} x_i = \sum_{i=1}^{k} y_i$ and if $\sum_{i=1}^{k'} x_{[i]} \leq \sum_{i=1}^{k'} y_{[i]}$, $1 \leq k' < k$, where $x_{[i]}, i=1,\ldots,k$, denote the coordinates of x written in decreasing order, etc. Then h is called Schur convex if it is non-decreasing relative to the ordering $\propto_S$. (Obviously any such function must be a symmetric function of $x_1,\ldots,x_k$.)

For further information about these and other partial orderings, consult Marshall and Olkin (1979), Karlin and Rinott (1981), Eaton (1982), and references cited in these works.

EXERCISES

2.2.1 Generalize 2.1(2) to

(1) $$||x||^{\ell} \frac{e^{\theta \cdot x} - e^{\theta_0 \cdot x}}{||\theta - \theta_0||} \le K_{\ell+1} \sum_{i=1}^{I} e^{b_i \cdot x} \quad .$$

Thus

(2) $$\frac{(\Pi x_i^{\ell_i})(e^{\theta \cdot x} - e^{\theta_0 \cdot x} - (\theta - \theta_0) \quad x e^{\theta_0 \cdot x})}{||\theta - \theta_0||} \le 2K_{\ell+1} \Sigma e^{b_i \cdot x} \quad .$$

Use this to prove 2.2(1) by induction on ℓ.

2.3.1 Consider a one-dimensional standard exponential family with $K \subset [0, \infty)$. Show that

(1) $$(E_0[(1 - a)^X])^2 \ge E_0[(1 - 2a)^X], \quad 0 \le a < \tfrac{1}{2} \quad ,$$

and $\mathrm{Var}_0 X < \infty$ imply

(2) $$E_0(X) \ge \mathrm{Var}_0 X \quad .$$

[Let $e^{\theta} = (1 - a)$ and show by differentiating at $\theta = 0^-$ that (1) implies $\psi'(0^-) \ge \psi''(0^-)$. The finiteness of $\mathrm{Var}_0 X$ guarantees that $\psi''(0^-) = \mathrm{Var}_0 X < \infty$, etc., S. Zamir (personal communication).] (It is not known if (1) implies (2) without the assumption that $\mathrm{Var}_0 X < \infty$.)

2.4.1 Canonical one-parameter exponential families for which $\mathrm{Var}_{\theta}(X)$ is a quadratic function of $E_{\theta}(X)$ are called *quadratic variance function* families (= QVF). See Morris (1982, 1983). Verify that the following six families have the QVF property:

(1) $N(\mu, \sigma^2)$ μ known

(2) $P(\lambda)$

(3) $\Gamma(\alpha, \sigma)$ α known

(4) Bin (r, p) r known

(5) Neg. Bin. (r, p) r known

(6) ν has density $f(x) = (2 \cosh(\frac{\pi x}{2}))^{-1}$, $-\infty < x < \infty$,

relative to Lebesgue measure. ($X = \pi^{-1} \log(Y/(1 - Y))$ where $Y \sim$ Beta $(\tfrac{1}{2}, \tfrac{1}{2})$.)

[In (6) $\psi(\theta) = -\log(\cos\theta)$. This is called the hyperbolic secant distribution. The generalized hyperbolic secant distributions are produced from these by infinite divisibility and convolution. These families are the only QVF families (Morris, 1982). See also Bar-Lev and Enis (1985).]

<u>2.5.1</u> Let $\{p_\theta\}$ be a canonical one-dimensional exponential family. Then $N^\circ = (\theta_1, \theta_2)$, $\xi(N^\circ) = (\xi_1, \xi_2)$ for some $-\infty \le \theta_1 < \theta_2 \le \infty$ and $-\infty \le \xi_1 < \xi_2 \le \infty$. If $K = [x_1, \infty)$ then $\xi_1 = x_1$. (Theorem 3.6 is a multivariate generalization of this result.)

<u>2.10.1</u> Let $\{p_\theta\}$ be a two-dimensional canonical exponential family. Find a convex subset of N such that h bounded and $E_\theta(h) = 0$ for all $\theta \in \partial N$ implies $h = 0$ a.e.(ν). (Hence, the family $\{p_\theta : \theta \in \partial N\}$ is "boundedly complete".) Conclude that every test of Θ_0 versus $\Theta_1 = N - \Theta_0$ is "admissible". (i.e. Let $\pi_\phi(\theta) = E_\theta(\phi)$. Then $\pi_{\phi_1}(\theta) \leqq \pi_{\phi_2}(\theta)$, $\theta \in \Theta_0$, and $\pi_{\phi_1}(\theta) \geqq \pi_{\phi_2}(\theta)$, $\theta \in \Theta_1$, implies $\pi_{\phi_1}(\theta) \equiv \pi_{\phi_2}(\theta)$.) [∂N contains an infinite number of line segments. See Farrell (1968).]

<u>Similar Tests and Unbiased Tests</u>

<u>2.12.0</u> Let $\Theta_i \subset \Theta$, i=0,1. A critical test function ϕ, $0 \le \phi \le 1$, is called level α *unbiased* if $E_\theta(\phi) \le \alpha$, $\theta \in \Theta_0$, and $E_\theta(\phi) \ge \alpha$, $\theta \in \Theta_1$. It is called *similar* (level α) if $E_\theta(\phi) \equiv \alpha$, $\theta \in \bar{\Theta}_0 \cap \bar{\Theta}_1 \cap N$. The following problems consider the common case where $\Theta_0 \cup \Theta_1 = N$ so that $\partial\Theta_0 \cap N = \bar{\Theta}_0 \cap \bar{\Theta}_1 \cap N$. Exercises 2.21.3, 2.21.4 and 2.21.5 contain further applications of these concepts. See also 7.12.1.

<u>2.12.1</u> Let $\{p_\theta\}$ be a regular canonical family and let $\theta' = (\theta'_{(1)}, \theta'_{(2)})$, $X' = (X'_{(1)}, X'_{(2)})$ be partitioned vectors. (Regularity is convenient but not essential here.) Let $L = \{\theta : \theta_{(1)} = 0\}$. Assume $L \cap N^{\circ} \ne \phi$.

(i) Show that a critical function ϕ is similar on L if and only if

$$(1) \qquad \alpha = \int \phi(x)\nu(dx_{(1)} | x_{(2)}) \qquad \text{a.e.}(\nu) \quad .$$

(Tests with property (1) are said to have *Neyman* structure. Note that the right side of (1) is $E_\theta(\phi|x_{(2)})$ for $\theta \in L$.)

(ii) Show that ϕ is similar on L and satisfies

$$(2) \qquad v \cdot \nabla E_\theta(\phi) = 0 \quad \forall \ v \in L^\perp = \{v : v \cdot \theta = 0 \quad \forall \ \theta \in L\}$$

if and only if ϕ satisfies (1) and

$$(3) \qquad \int x_{(1)}\phi(x)\nu(dx_{(1)}|x_{(2)}) = 0 \quad \text{a.e. } (\nu) .$$

(Note that (2) is a necessary condition for a test of H_0: $\theta \in L$ versus H_1: $\theta \notin L$ to be unbiased. (3) expresses the fact that $v \cdot \nabla E_\theta(\phi|x_{(2)}) = 0$ for all $\theta \in L$, $v \in L^\perp$, $x_{(2)}$. See Lehmann (1959) for many applications of (1) and (3) to the construction of U.M.P.U. tests.)

2.12.2 (i) Let $X \sim N(\theta, I)$ in R^k, $k \geq 2$. Show there does not exist a non-constant level α similar test of $\Theta_0 = \{\theta : \theta_i \leq 0$ for some i$\}$. [Use Example 2.10.]

(ii) Show there exists a non-constant similar test of $\Theta_0 = \{\theta : \theta_i = 0$ for some i$\}$, but there does not exist a non-constant unbiased test of this hypothesis.

2.12.3 Let $X \in R^k$, $X_i \sim P(\lambda_i)$, independent. Show there exists a non-trivial similar test of $\{\lambda : \lambda_i \leq 1 \ \forall \ i\}$ but there does not exist a non-trivial unbiased test of this hypothesis.

2.13.1

Let $X = (X_{ij})$ be a matrix $\Gamma(\alpha, I)$ variable. (See Exercise 1.14.4.) Observe that $\log |X|$ has the same Laplace transform as $\sum_{i=1}^m \log Y_i$ where Y_i are independent $\Gamma(\alpha - (i-1)/2, 1)$ variables. Hence $|X|$ has the same distribution as $\prod_{i=1}^m Y_i$. Reinterpret this result to show equality of the distribution of the determinant of a Wishart (n, I) matrix and a product of independent χ^2 -variables.

2.13.2

Let F, G be two distributions on R^k. Let $\mu^F_{i_1,\ldots,i_k} = E(\prod_{j=1}^k X_j^{i_j})$ and similarly for μ^G. Suppose

(1) $$\mu^F_{i_1,\ldots,i_k} = \mu^G_{i_1,\ldots,i_k} \qquad i_j = 0,1,\ldots \qquad j = 1,\ldots,k ,$$

and

(2) $$\limsup_{n\to\infty} \frac{1}{n} m_{j,2n}^{\frac{1}{2}n} < \infty , \qquad j = 1,\ldots,k$$

where $m_{j,2n} = E(|X_j|^{2n})$. (Note $m_{j,2n} = \mu_{0,\ldots,2n,\ldots,0}$.) Then $F = G$. (Condition (2) is slightly weaker than the necessary and sufficient condition,

(3) $$\sum_{n=1}^{\infty} m_{j,2n}^{-(\frac{1}{2}n)} = \infty , \qquad j = 1,\ldots,k \quad ,$$

for (1) to imply equality of F and G. See Feller (1966, Sections XV4 and VII3) and references cited therein.)

[Use Stirling's formula to show that $\Sigma m_{j,n}\theta^n/n!$ converges absolutely for $|\theta| < \varepsilon$, $j=1,\ldots,k$, and hence that $\lambda_F = \lambda_G$ on an open set in R^k.]

<u>2.14.1</u> (Bar-Lev (1983).)

Let $X \sim \text{Expf}\ (\theta)$ with $\Theta^\circ \neq \phi$. Let $\not{\Sigma}\,(X_{(2)}|x_{(1)})$ denote the indicated conditional covariance matrix. Show that $\not{\Sigma}_\theta(X_{(2)}|x_{(1)})$ depends only on θ if and only if $X_{(1)} \perp (X_{(2)} - h(X_{(1)}))$ for some (measurable) function h.

[Integrate $\not{\Sigma}_\theta(X_{(2)}|x_{(1)})$ on θ starting at $0 \in \Theta^\circ$ to find that the conditional cumulant generating function of $X_{(2)}$ under P_0 is

(2) $$\psi(\theta|x_{(1)}) = \rho(\theta_{(2)}) + \theta_{(2)} \cdot h(x_{(1)}) \qquad \text{for some functions } \rho ,\ h \quad .$$

Show that (2) implies $X_{(2)} - h(X_{(1)}) \perp X_{(1)}$ under P_0.]

<u>2.14.2</u>

Suppose $X \sim \text{Expf}\ (\theta)$ with $\Theta^\circ \neq \phi$. Then the following are equivalent:

(1) $X_{(1)} \perp X_{(2)}$ for some $\theta^\circ \in \Theta$, or for all $\theta \in \Theta$,

(2) $\psi(\theta_{(1)}, \theta_{(2)}) = \psi_1(\theta_{(1)}) + \psi_2(\theta_{(2)})$ for some functions ψ_1 and ψ_2 ,

(3) $X_{(i)} \sim \text{Expf}\ (\theta_{(i)})$ for $i = 1$ and 2 ,

(4) $\text{cov}_\theta(X_{(1)}, X_{(2)}) = 0 \quad \forall\, \theta \in \Theta$.

[For (1) - (3) apply Theorem 2.14 with $h \equiv 0$ and check $\psi_i = \rho_i$, $i=1,2$. For (4) $\Rightarrow$ (2) use 2.3(2) and integrate.]

2.14.3 (Patil (1965), Barndorff-Nielsen and Blaesild (1983).)

Let $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ be a family of distributions on $\mathcal{Y}, \mathcal{B}$. Let $X : \mathcal{Y} \to R^k$ (measurable), $\Theta \subset R^k$ with $\Theta^\circ \neq \phi$. Suppose

$$\ln E_\theta \exp((\beta - \theta) \cdot X(Y)) = \rho(\beta) - \rho(\theta), \qquad \beta, \theta \in \Theta$$

for some function $\rho(\cdot)$. Then $X \sim \text{Expf}(\theta)$. [Use Corollary 2.13.]

2.14.4

Let X have a k-dimensional multinomial (N, π) distribution. Write $X'_{(1)} = (X_1, \ldots, X_{k_1})'$, $X'_{(2)} = (X_{k_1+1}, \ldots, X_k)'$. Show that the marginal distributions of both $X_{(1)}$ and $X_{(2)}$ form an exponential family, but $X_{(1)}$ is not independent of $X_{(2)}$ as one might expect from Theorem 2.14(2). Why not? [The fact that X is not a minimal family is irrelevant; for $k \geq 3$, $k_1 \leq k-2$ the same phenomenon occurs in the minimal model defined as in 1.2(7).]

2.15.1

Let the independent symmetric $m \times m$ matrices, X_i, $i=1,\ldots,n$, have matrix $\Gamma(\alpha_i, \not{\Sigma})$ distributions. (See Exercise 1.14.4). Show that $Z = Z_1, \ldots, Z_n$ with $Z_j = |X_j| / |\sum_{i=1}^n x_i|$ is independent of $\sum_{i=1}^n x_i$. Show that the distributions of $\ln Z = \{\ln Z_j : j=1,\ldots,n\}$ form an exponential family, and identify the canonical statistic and parameter for this distribution. (This generalizes Example 2.15(ii). The distributions of Z form the so-called multivariate beta distribution. See, e.g., Muirhead (1982). When $m = 1$ the X_i have ordinary Γ distributions and the distribution of Z is a Dirichlet distribution. See Exercise 5.6.2.

2.16.1

Suppose $\nu_n \to \nu$ with $\nu(R^k) < \infty$. Then

$$\limsup \nu_n(R^k) \leq \nu(R^k) \tag{1}$$

if and only if the sequence $\{\nu_n\}$ is tight. [Let $c(x) = i$ if $r_i \leq ||x|| \leq r_{i+1}$

and choose $r_i \ni \sup_n \nu_n(\{||x|| \geq r_i\}) \leq 1/i^2$, $i=1,2,\ldots$.] Hence a convergent sequence of probability measures has a probability measure as its limit if and only if it is tight.

2.17.1

Verify 2.17(4),(5). [From Lemma 2.1(1)

$$(1) \qquad ||\int xe^{b\cdot x}\,\nu_i(dx)|| \leq \sum_{j=1}^{J} ||\int \left\{ \frac{xe^{b\cdot x}}{\Sigma e^{b_j\cdot x}} \right\} e^{b_j\cdot x}\,\nu_i(dx)||$$

and the quantity in braces in (1) is $O(1/(1 + ||x||))$. Now use 2.16(1).]

2.17.2

Let $S \subset R^k$ and $B = \text{conhull } S$. Let ν_n be a bounded sequence of measures on R^k $(\nu_n(R^k) \leq K_1 < \infty)$ with $\lambda_{\nu_n}(b) < \infty$, $b \in S$, $n=1,\ldots$. Suppose $0 \in B^\circ$. Define $P_{n,b}$ by

$$(1) \qquad \frac{dP_{n,b}}{d\nu_n} = \exp(b\cdot x - \psi_{\nu_n}(b)) \quad .$$

Suppose for each $b \in S$ there is a $K = K(b)$ such that

$$(2) \qquad \limsup_{n\to\infty} P_{n,b}(\{||x|| < K\}) > 0 \quad .$$

Then there is a subsequence $\{n'\} \subset \{n\}$ and a non-zero limiting measure ν such that for all $b \in B^\circ$

$$(3) \qquad e^{b\cdot x}\,\nu_{n'}(dx) \to e^{b\cdot x}\,\nu(dx)\ , \qquad \lambda_{\nu_{n'}}(b) \to \lambda_\nu(b)$$

[As in the proof of Theorem 2.17 it suffices to consider the case where S is finite. Then $K = \max\{K(b) : b \in S\} < \infty$. If $b \in S$, $||b|| \leq K_0$ then $0 < \varepsilon \leq \int_{||x||\leq K} e^{b\cdot x}\nu_{n'}(dx)/\lambda_{\nu_{n'}}(b) \leq K_1 e^{K_0 K}/\lambda_{\nu_{n'}}(b)$. Hence 2.17(1) is satisfied on $S \cap \{b : ||b|| \leq K_0\}$. $\nu \neq 0$ since $0 \in B^\circ$]

2.18.1

Let $\{\nu_{\alpha n} : \alpha \in A\}$, $n=1,2,\ldots$ be a family of sequences of measures on $X = \{0,1,\ldots\}$. Show that $\nu_{\alpha n} \to \nu_\alpha$ uniformly in α if and only if

$\nu_{\alpha n}(\{x\}) \to \nu_\alpha(\{x\})$ uniformly in α for each $x \in X$.

2.19.1

Let $\{p_\theta\}$ be an exponential family with supp $\nu \subset \{0,1,\ldots\}$ and $\nu(0) > 0$, $\nu(1) > 0$. Let $X_1,\ldots,X_n$ be a random sample and, as usual, let $S_n = \sum_{i=1}^{n} X_i$. Define $\theta_n(\lambda)$ by

$$\xi(\theta_n(\lambda)) = \lambda/n \quad . \tag{1}$$

Let $F_{\lambda,n}$ denote the distribution of S_n under the parameter $\theta_n(\lambda)$. Show that $F_{\lambda,n} \to P(\lambda)$ and the convergence is uniform in λ over $\lambda \in [a,b]$ for $0 < a < b < \infty$. (A slight elaboration of the argument yields uniformity over $[0, b]$.) Generalize this result to the case where p_θ is a k-dimensional exponential family. [Show $\psi''(\theta_n(\lambda)) \to 0$ as $n \to \infty$ since $\theta_n(\lambda) \to -\infty$, uniformly for $\lambda \in [a, b]$. Hence $\log E_{\theta_n(\lambda)} e^{\beta S_n} = \lambda(e^\beta - 1) + o(1)$ as $n \to \infty$ uniformly for $\lambda \in [a, b]$. Then apply Theorem 2.19. In the non-degenerate k-dimensional case the limit distribution is the product of independent Poisson variables.] (A special case of the above is the well known result Bin $(n, \lambda/n) \to \mathcal{P}(\lambda)$. The general form of the above statement was pointed out to me by I. Johnstone.)

2.21.1

Let X be non-central χ^2 with m degrees of freedom and non-centrality parameter θ . Show that the distributions of X have the sign-change preserving properties 2.21(2), (3). [Use Exercise 1.12.1(1). Write $E_\theta(h(X)) = E_\theta(E(h(X)|K))$.]

2.21.2

Let X be a one-dimensional exponential family and $\theta_0 \in N^\circ$.

(i) Show that the (essentially unique) level α test of the form

$$\phi(x) = \begin{array}{ll} 1 & x > x_0 \\ \gamma & x = x_0 \\ 0 & x < x_0 \end{array} \tag{1}$$

is the U.M.P. level α test of H_0: $\theta \leq \theta_0$ versus H_1: $\theta > \theta_0$.

(ii) Similarly, show that the (essentially unique) level α test of the form

$$(2) \qquad \phi(x) = \begin{cases} 1 & x > x_2 \text{ or } x < x_1 \\ \gamma_i & x = x_i \\ 0 & x_1 < x < x_2 \end{cases}$$

satisfying

$$(3) \qquad E_{\theta_0}(x\phi(x)) = 0$$

is the U.M.P.U. level test of H_0: $\theta = \theta_0$ versus H_1: $\theta \neq \theta_0$.

[(i) Let ϕ' be any different level α test. Then $S^-(\phi - \phi') = 1$. $E_{\theta_0}(\phi - \phi') = 0$ by definition. Now use Theorem 2.18. (ii) Condition (3) is the one-dimensional version of 2.12.1(3). Again use Theorem 2.18.] (It is also possible to show by a continuity argument that level α tests of the form (1) and (2), (3) always exist.)

2.21.3

Consider a 2×2 contingency table. (See Exercise 1.8.1.) Describe the general form of the U.M.P.U. level α tests of the following null hypotheses. In each case the alternative is the complement of H_0.

(i) H_0: $p_{11}p_{22}/p_{12}p_{21} \leq 1$

(ii) H_0: $p_{11}p_{22}/p_{12}p_{21} = 1$

(iii) H_0: $p_{11} \leq p_{12}$

(iv) H_0: $p_{12} = p_{21}$.

(This corresponds to the exact form of McNemar's test. See, e.g. Fleiss (1981).) [Use Exercise 2.21.2 and, for (i), (ii), Exercise 1.15.1. See Lehmann (1959).]

2.21.4

Consider a 2×2 contingency table. Let $c > 0$, $c \neq 1$. Show there exist non-trivial similar tests of the null hypothesis

H_0: $p_{11}/(p_{11} + p_{12}) = cp_{21}/(p_{21} + p_{22})$ of conditional probabilities in a given proportion, even though this is not a log-linear hypothesis. [Use randomized tests. Consider the conditional distribution given Y_{i+}, $i=1,2$ under which Y_{11} and Y_{21} are independent binomials. (This case is of interest on its own merits.) Consider the special case $Y_{1+} = 1 = Y_{2+}$ for which the condition for similarity reduces to four linear equations in the four variables $\phi(y)$ for the four conditionally possible outcomes, y. This test is unbiased for the one-sided version of H_0, but not for H_0 as defined above. Is there, in general, an unbiased test of H_0? Is there, in general, a U.M.P.U. test of either the one- or two-sided hypothesis in either the original model or the conditional (independent binomial) model? The somewhat analogous question of the existence of similar and of unbiased tests for the Behrens-Fisher problem of equality of means for two normal samples with unknown variances is solved in Wijsman (1958) and in Linnik (1968).]

2.21.5

Let $X_1,\ldots,X_n$ be a sequence of independent failure times, assumed to have a $\Gamma(\alpha, \sigma)$ distribution. Describe the U.M.P.U. tests of H_0: $\alpha = 1$ versus H_1: $\alpha > 1$ and H_1': $\alpha \neq 1$. [Use Exercise 2.21.2 and Example 2.15.]

2.25.1

Suppose ν has density f with respect to Lebesgue measure on R^k and f is MTP_2 (i.e. has monotone likelihood ratio) in each pair of coordinates. Prove the conclusions of Lemma 2.24 and Theorem 2.25. Prove these also for the case where f, as above, is a density with respect to counting measure on the lattice of points with integer coordinates. [If $h(x_1,\ldots,x_k)$ is non-decreasing then, under ν, $E(h(X_1,\ldots,X_{k-1}, X_k) \mid X_k = x_k)$ is also non-decreasing.]

2.25.2

Let $\{p_\theta\}$ be a canonical k-parameter exponential family with

$\theta_0 \in N^\circ$. Let H_0: $\theta \le \theta_0$ and H_1: $\theta > \theta_0$. (i) Show that any Bayes or generalized Bayes test, α, of H_0 versus H_1 has the strong monotonocity property

$$(1) \qquad \begin{aligned} \phi(x) > 0 \quad & y > x \Rightarrow \phi(y) = 1 \\ \phi(x) < 1 \quad & y < x \Rightarrow \phi(y) = 0 \quad . \end{aligned}$$

Assume $\theta_0 = 0$ and consider $\nabla \int p_\theta(x)[G_1(d\theta) - G_0(d\theta)]$ where G_i denotes the (generalized) prior measure restricted to H_i.] (ii) Suppose the dominating measure ν is MTP_2. Show that any (generalized) Bayes test is unbiased. [Use the above and Exercise 2.25.1.]

<u>2.25.3</u> (Slepian's Inequality)

Let X, Y be k-dimensional normal variables with mean 0 and non-singular covariance matrices A, B, respectively. Suppose

$$a_{ii} = b_{ii}, \quad a_{ij} \ge b_{ij} \qquad 1 \le i,j \le k \quad .$$

Then, for any $C \in R^k$,

$$(1) \qquad \Pr\{X \le C\} \ge \Pr\{Y \le C\} \quad .$$

[If $Z^{(\rho)} \sim N(0, A + \rho(B - A))$ then

$$(2) \qquad \frac{\partial}{\partial\rho} P(Z^{(\rho)} \le C) = \sum_{i \ne j} \alpha_{ij} \frac{\partial}{\partial\theta_{ij}} P(Z^{(\rho)} \le C)$$

where each $\alpha_{ij} \ge 0$. Note that for $i \ne j$

$$(3) \qquad \frac{\partial\lambda}{\partial\theta_{ij}} = \theta_{ij} \exp(-\ln|\Sigma|/2) = \theta_{ij}\lambda$$

by 2.4(2). Hence

$$(4) \qquad \frac{\partial p_\theta(Z)}{\partial\theta_{ij}} = \theta_{ij}\, p_\theta(Z) = \frac{\partial^2 p_\theta(Z)}{\partial Z_i\, \partial Z_j}$$

from Corollary 2.13. Combine (2) and (4) to yield (1).] (For an alternate proof of Slepian's inequality see Saw (1977). For generalizations see Joag-Dev, Perlman, and Pitt (1983) and Brown and Rinott (1986).)