Appendix

We begin this appendix with a statement and proof of a result due to Basu (1955). Consider a measurable space $(\mathcal{X}, \mathcal{B})$ and a probability model $\{P_\theta | \theta \in \Theta\}$ defined on $(\mathcal{X}, \mathcal{B})$. Consider a statistic T defined on $(\mathcal{X}, \mathcal{B})$ to $(\mathcal{Y}, \mathcal{B}_1)$, and let $\mathcal{B}_T = \{T^{-1}(B) | B \in \mathcal{B}_1\}$. Thus $\mathcal{B}_T$ is a σ-algebra and $\mathcal{B}_T \subseteq \mathcal{B}$. Conditional expectation given $\mathcal{B}_T$ is denoted by $\mathcal{E}(\cdot | \mathcal{B}_T)$.

Definition A.1. The statistic T is a *sufficient statistic* for the family $\{P_\theta | \theta \in \Theta\}$ if, for each bounded $\mathcal{B}$ measurable function f, there exists a $\mathcal{B}_T$ measurable function $\hat{f}$ such that $\mathcal{E}_\theta(f | \mathcal{B}_T) = \hat{f}$ a.e. P_θ for all $\theta \in \Theta$.

Note that the null set where the above equality does not hold is allowed to depend on both θ and f. The usual intuitive description of sufficiency is that the conditional distribution of $X \in \mathcal{X}$ ($\mathcal{L}(X) = P_\theta$ for some $\theta \in \Theta$) given $T(X) = t$ does not depend on θ. Indeed, if $P(\cdot | t)$ is such a version of the conditional distribution of X given $T(X) = t$, then $\hat{f}$ defined by $\hat{f}(x) = h(T(x))$ where

$$h(t) = \int f(x) P(dx | t)$$

serves as a version of $\mathcal{E}_\theta(f | \mathcal{B}_T)$ for each $\theta \in \Theta$.

Now, consider a statistic U defined on $(\mathcal{X}, \mathcal{B})$ to $(\mathcal{Z}, \mathcal{B}_2)$.

Definition A.2. The statistic U is called an *ancillary statistic* for the family $\{P_\theta | \theta \in \Theta\}$ if the distribution of U on $(\mathcal{Z}, \mathcal{B}_2)$ does not depend on $\theta \in \Theta$—that is, if for all $B \in \mathcal{B}_2$,

$$P_\theta\{U^{-1}(B)\} = P_\eta\{U^{-1}(B)\}$$

for all $\theta, \eta \in \Theta$.

In many instances, ancillary statistics are functions of maximal invariant statistics in a situation where a group acts transitively on the family of probabilities in question—see Section 7.5 for a discussion.

Finally, given a statistic T on $(\mathfrak{X}, \mathfrak{B})$ to $(\mathfrak{Y}, \mathfrak{B}_1)$ and the parametric family $\{P_\theta | \theta \in \Theta\}$, let $\{Q_\theta | \theta \in \Theta\}$ be the induced family of distributions of T on $(\mathfrak{Y}, \mathfrak{B}_1)$—that is,

$$Q_\theta(B) = P_\theta(T^{-1}(B)), \qquad B \in \mathfrak{B}_1.$$

Definition A.3. The family $\{Q_\theta | \theta \in \Theta\}$ is called *boundedly complete* if the only bounded solution to the equation

$$\int h(y) Q_\theta(dy) = 0, \qquad \theta \in \Theta$$

is the function $h = 0$ a.e. Q_θ for all $\theta \in \Theta$.

At times, a statistic T is called boundedly complete—this means that the induced family of distributions of T is boundedly complete according to the above definition. If the family $\{Q_\theta | \theta \in \Theta\}$ is an exponential family on a Euclidean space and if Θ contains a nonempty open set, then $\{Q_\theta | \theta \in \Theta\}$ is boundedly complete—see Lehmann (1959, page 132).

Theorem (Basu, 1955). If T is a boundedly complete sufficient statistic and if U is an ancillary statistic, then, for each θ, $T(X)$ and $U(X)$ are independent.

Proof. It suffices to show that, for bounded measurable functions h and k on $\mathfrak{Y}$ and $\mathfrak{Z}$, we have

$$\text{(A.1)} \qquad \mathcal{E}_\theta h(T(X)) k(U(X)) = \mathcal{E}_\theta h(T(X)) \mathcal{E}_\theta k(U(X)).$$

Since U is ancillary, $a = \mathcal{E}_\theta k(U(X))$ does not depend on θ, so $\mathcal{E}_\theta(k(U) - a) = 0$ for all θ. Hence

$$\mathcal{E}_\theta\left[\mathcal{E}_\theta((k(U) - a) | \mathfrak{B}_T)\right] = 0 \quad \text{for all } \theta.$$

Since T is sufficient, there is a $\mathfrak{B}_T$ measurable function, say $\hat{f}$, such that $\mathcal{E}_\theta((k(U) - a) | \mathfrak{B}_T) = \hat{f}$ a.e. P_θ. But since $\hat{f}$ is $\mathfrak{B}_T$ measurable, we can write $\hat{f}(x) = \psi(T(x))$ (see Lehmann, 1959, Lemma 1, page 37). Also, since k is bounded, $\hat{f}$ can be taken to be bounded. Hence

$$\mathcal{E}_\theta \psi(T) = 0 \qquad \text{for all } \theta$$

and ψ is bounded. The bounded completeness of T implies that ψ is 0 a.e. Q_θ where $Q_\theta(B) = P_\theta(T^{-1}(B))$, $B \in \mathcal{B}_1$. Thus $h(T)\psi(T) = 0$ a.e. Q_θ, so

$$\begin{aligned} 0 = \mathcal{E}_\theta h(T)\psi(T) &= \mathcal{E}_\theta\left[h(T)\mathcal{E}_\theta((k(U) - a)|\mathcal{B}_T)\right] \\ &= \mathcal{E}_\theta\left[\mathcal{E}_\theta(h(T)(k(U) - a)|\mathcal{B}_T)\right] \\ &= \mathcal{E}_\theta h(T)(k(U) - a). \end{aligned}$$

Thus (A.1) holds. □

This Theorem can be used in many of the examples in the text where we have used Proposition 7.19.

The second topic in this Appendix concerns monotone likelihood ratio and its implications. Let $\mathcal{X}$ and $\mathcal{Y}$ be subsets of the real line.

Definition A.4. A nonnegative function k defined on $\mathcal{X} \times \mathcal{Y}$ is *totally positive of order 2* (TP-2) if, for $x_1 < x_2$ and $y_1 < y_2$, we have

$$k(x_1, y_1)k(x_2, y_2) \geqslant k(x_1, y_2)k(x_2, y_1). \tag{A.2}$$

In the case that $\mathcal{Y}$ is a parameter space and $k(\cdot, y)$ is a density with respect to some fixed measure, it is customary to say that k has a monotone likelihood ratio when k is TP-2. This nomenclature arises from the observation that, when k is TP-2 and $y_1 < y_2$, then the ratio $k(\cdot, y_2)/k(\cdot, y_1)$ is nondecreasing in x—assuming that $k(\cdot, y_1)$ does not vanish. Some obvious examples of TP-2 functions are: $\exp[xy]$, x^y for $x > 0$, y^x for $y > 0$. If $x = g(s)$ and $y = h(t)$ where g and h are both increasing or decreasing, then $k_1(s, t) \equiv k(g(s), h(t))$ is TP-2 whenever k is TP-2. Further, if $\psi_1(x) \geqslant 0$, $\psi_2(y) \geqslant 0$, and k is TP-2, then $k_1(x, y) = \psi_1(x)\psi_2(y)k(x, y)$ is also TP-2.

The following result due to Karlin (1956) is of use in verifying that some of the more complicated densities that arise in statistics are TP-2. Here is the setting. Let $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ be Borel subsets of R^1 and let μ be a σ-finite measure on the Borel subsets of $\mathcal{Y}$.

Lemma (Karlin, 1956). Suppose g is TP-2 on $\mathcal{X} \times \mathcal{Y}$ and h is TP-2 on $\mathcal{Y} \times \mathcal{Z}$. If

$$k(x, z) = \int g(x, y) h(y, z) \mu(dy)$$

is finite for all $x \in \mathcal{X}$ and $z \in \mathcal{Z}$, then k is TP-2.

Proof. For $x_1 < x_2$ and $z_1 < z_2$, the difference

$$\Delta = k(x_1, z_1) k(x_2, z_2) - k(x_1, z_2) k(x_2, z_1)$$

can be written

$$\Delta = \int\int g(x_1, y_1) g(x_2, y_2) [h(y_1, z_1) h(y_2, z_2) - h(y_1, z_2) h(y_2, z_1)]$$
$$\times \mu(dy_1) \mu(dy_2).$$

Now, write Δ as the double integral over the set $\{y_1 < y_2\}$ plus the double integral over the set $\{y_1 > y_2\}$. In the integral over the set $\{y_1 > y_2\}$, interchange y_1 and y_2 and then combine with the integral over $\{y_1 < y_2\}$. This yields

$$\Delta = \int\int_{\{y_1 < y_2\}} [g(x_1, y_1) g(x_2, y_2) - g(x_1, y_2) g(x_2, y_1)]$$
$$\cdot [h(y_1, z_1) h(y_2, z_2) - h(y_1, z_2)(h(y_2, z_1)] \mu(dy_1) \mu(dy_2).$$

On the set $\{y_1 < y_2\}$, both of the bracketed expressions are non-negative as g and h are TP-2. Hence $\Delta \geqslant 0$ so k is TP-2. □

Here are some examples.

◆ **Example A.1.** With $\mathcal{X} = (0, \infty)$, let

$$f(x, m) = \frac{x^{(m/2)-1} \exp[-\frac{1}{2}x]}{2^{m/2} \Gamma(m/2)}$$

be the density of a chi-squared distribution with m degrees of freedom. Since $x^{m/2}$, $x \in \mathcal{X}$ and $m > 0$, is TP-2, $f(x, m)$ is TP-2. Recall that the density of a noncentral chi-squared distribution with

p degrees of freedom and noncentrality parameter $\lambda \geqslant 0$ is given by

$$h(x, \lambda) = \sum_{j=0}^{\infty} \frac{(\lambda/2)^j \exp[-\frac{1}{2}\lambda]}{j!} f(x, p + 2j).$$

Observe that $f(x, p + 2j)$ is TP-2 in x and j and $(\lambda/2)^j \exp[-\frac{1}{2}\lambda]/j!$ is TP-2 in j and λ. With $\mathcal{Y} = \{0, 1, \dots\}$ and μ as counting measure, Karlin's Lemma implies that $h(x, \lambda)$ is TP-2. ◆

◆ **Example A.2.** Recall that, if χ_m^2 and χ_n^2 are independent random variables, then $Y = \chi_m^2/\chi_n^2$ has a density given by

$$f(y|m, n) = \frac{\Gamma((m+n)/2)}{\Gamma(m/2)\Gamma(n/2)} \frac{y^{(m/2)-1}}{(1+y)^{(m+n)/2}}, \qquad y > 0.$$

If the random variable χ_m^2 is noncentral chi-squared, say $\chi_p^2(\lambda)$, rather than central chi-squared, then $Y = \chi_p^2(\lambda)/\chi_m^2$ has a density

$$h(y|\lambda) = \sum_{j=0}^{\infty} \frac{(\lambda/2)^j \exp[-\frac{1}{2}\lambda]}{j!} f(y|p + 2j, n).$$

Of course, Y has an unnormalized $F(p, m; \lambda)$ distribution according to our usage in Section 8.4. Since $f(y|p + 2j, n)$ is TP-2 in y and j, it follows as in Example A.1 that h is TP-2. ◆

The next result yields the useful fact that the noncentral Student's t distribution is TP-2.

Proposition A.1. Suppose $f \geqslant 0$ defined on $(0, \infty)$ satisfies

(i) $\int_0^\infty e^{ux} f(x)\, dx < +\infty \quad$ for $u \in R^1$
(ii) $f(x/\eta)$ is TP-2 on $(0, \infty) \times (0, \infty)$.

For $\theta \in R^1$ and $t \in R^1$, define k by $k(t, \theta) = \int_0^\infty e^{\theta t x} f(x)\, dx$. Then k is TP-2.

Proof. First consider $t \in R^1$ and $\theta > 0$. Set $v = \theta x$ in the integral defining k to obtain

$$k(t, \theta) = \frac{1}{\theta} \int_0^\infty e^{tv} f\left(\frac{v}{\theta}\right) dv.$$

Now apply Karlin's Lemma to conclude that k is TP-2 on $R^1 \times (0, \infty)$. A similar argument shows that k is TP-2 on $R^1 \times (-\infty, 0)$. Since $k(t, 0)$ is a constant, it is now easy to show that k is TP-2 on $R^1 \times R^1$. □

◆ **Example A.3.** Suppose X is $N(\mu, 1)$ and Y is χ^2_n. The random variable $T = X/\sqrt{Y}$, which is, up to a factor of $\sqrt{n}$, a noncentral Student's t random variable, has a density that depends on μ—the noncentrality parameter. The density of T_1 (derived by writing down the joint density of X and Y, changing variables to T and $W = \sqrt{Y}$, and integrating out W) can be written

$$h(t, \mu) = \frac{2\exp\left[-\frac{1}{2}\mu^2\right]}{\Gamma(n/2)(1 + t^2)^{(n+1)/2}}$$

$$\times \int_0^\infty \exp[\psi(t)\mu x] \exp[-x^2] x^{-n}\, dx$$

where

$$\psi(t) = \sqrt{2}\, t(1 + t^2)^{-1/2}$$

is an increasing function of t. Consider the function

$$k(v, \mu) = \int_0^\infty \exp[v\mu x] \exp[-x^2] x^{-n}\, dx.$$

With $f(x) = \exp[-x^2]x^{-n}$, Proposition A.1 shows k, and hence h, is TP-2. ◆

We conclude this appendix with a brief description of the role of TP-2 in one sided testing problems. Consider a TP-2 density $p(x|\theta)$ for $x \in \mathcal{X} \subseteq R^1$ and $\theta \in \Theta \subseteq R^1$. Suppose we want to test the null hypothesis $H_0: \theta \in (-\infty, \theta_0] \cap \Theta$ versus $H_1: \theta \in (\theta_0, \infty) \cap \Theta$. The following basic result is due to Karlin and Rubin (1956).

Proposition A.2. Given any test ϕ_0 of H_0 versus H_1, there exists a test ϕ of the form

$$\phi(x) = \begin{cases} 1 & \text{if } x > x_0 \\ \gamma & \text{if } x = x_0 \\ 0 & \text{if } x < x_0 \end{cases}$$

with $0 \leqslant \gamma \leqslant 1$ such that $\mathcal{E}_\theta \phi \leqslant \mathcal{E}_\theta \phi_0$ for $\theta \leqslant \theta_0$ and $\mathcal{E}_\theta \phi \geqslant \mathcal{E}_\theta \phi_0$ for $\theta > \theta_0$. For any such test ϕ, $\mathcal{E}_\theta \phi$ is nondecreasing in θ.