Chapter 6
Testing

Chapter 6

Testing

1 Neyman Pearson Tests

Basic Notation. Consider the hypothesis testing problem as in Examples 5.1.4 and 5.5.4, but with $\Theta_0 = \{0\}$, $\Theta_1 = \{1\}$ (simple hypotheses). Let ϕ be a critical function (or decision rule); let $\alpha =$ size or level $\equiv E_0\phi(X)$; and let $\beta =$ power $= E_1\phi(X)$.

Theorem 1.1 (Neyman - Pearson lemma). Let P_0 and P_1 have densities p_0 and p_1 with respect to some dominating measure μ (recall that $\mu = P_0 + P_1$ always works). Let $0 \le \alpha \le 1$. Then:
(i) There exists a constant k and a critical function ϕ of the form

$$(1) \qquad \phi(x) = \begin{cases} 1 & \text{if } p_1(x) > kp_0(x) \\ 0 & \text{if } p_1(x) < kp_0(x) \end{cases}$$

such that

$$(2) \qquad E_0\phi(X) = \alpha.$$

(ii) The test of (1) and (2) is a *most powerful* α level test of P_0 versus P_1.
(iii) If ϕ is a most powerful level α test of P_0 versus P_1, then it must be of the form (1) a.e. μ. It also satisfies (2) unless there is a test of size $< \alpha$ with power $= 1$.

Corollary 1 If $0 < \alpha < 1$ and β is the power of the most powerful level α test, then $\alpha < \beta$ unless $P_0 = P_1$.

Proof. Let $0 < \alpha < 1$.
(i) Now

$$P_0(p_1(X) > cp_0(X)) = P_0(Y \equiv p_1(X)/p_0(X) > c) = 1 - F_Y(c).$$

Let $k \equiv \inf\{c : 1 - F_Y(c) < \alpha\}$, and if $P_0(Y = k) > 0$, let $\gamma \equiv (\alpha - P_0(Y > k))/P_0(Y = k)$. Thus with

$$\phi(x) = \begin{cases} 1 & \text{if } \ p_1(x) > kp_0(x) \\ \gamma & \text{if } \ p_1(x) = kp_0(x) \\ 0 & \text{if } \ p_1(x) < kp_0(x), \end{cases}$$

we have

$$E_0\phi(X) = P_0(Y > k) + \gamma P_0(Y = k) = \alpha.$$

(ii) Let ϕ^* be another test with $E_0\phi^* \le \alpha$. Now

$$\int_{\mathcal{X}} (\phi - \phi^*)(p_1 - kp_0) d\mu = \int_{[\phi-\phi^*>0]\cup[\phi-\phi^*<0]} (\phi - \phi^*)(p_1 - kp_0) d\mu \ge 0,$$

and this implies that

$$\begin{aligned} \beta_\phi - \beta_{\phi^*} &= \int_{\mathcal{X}} (\phi - \phi^*) p_1 d\mu \\ &\ge k \int_{\mathcal{X}} (\phi - \phi^*) p_0 d\mu = k(\alpha - E_0\phi^*) \ge 0. \end{aligned}$$

Thus ϕ is most powerful.
(iii) Let ϕ^* be most powerful of level α. Define ϕ as in (i). Then

$$\begin{aligned} \int_{\mathcal{X}} (\phi - \phi^*)(p_1 - kp_0) d\mu &= \int_{[\phi\ne\phi^*]\cap[p_1-kp_0\ne 0]} (\phi - \phi^*)(p_1 - kp_0) d\mu \\ & \quad \begin{cases} \ge 0 & \text{as in } (ii) \\ > 0 & \text{if } \mu([\phi \ne \phi^*] \cap [p_1 \ne kp_0]) > 0 \end{cases} \\ &= 0 \end{aligned}$$

since > 0 contradicts ϕ^* being most powerful. Thus $\mu([\phi \ne \phi^*] \cap [p_1 \ne kp_0]) = 0$. Thus $\phi^* = \phi$ on the set where $p_1 \ne kp_0$. If ϕ^* were of size $< \alpha$ and power < 1, then it would be possible to include more points (or parts of points) in the rejection region, and thereby increase the power, until either the power is 1 or the size is α. Thus either $E_0\phi^*(X) = \alpha$ or $E_1\phi^*(X) = 1$.

Corollary proof. $\phi^{\#}(x) \equiv \alpha$ has power α, so $\beta \ge \alpha$. If $\beta = \alpha$, then $\phi^{\#} \equiv \alpha$ is in fact most powerful; and hence (iii) shows that $\phi(x) = \alpha$ satisfies (i); that is, $p_1(x) = kp_0(x)$ a.e. μ. Thus $k = 1$ and $P_1 = P_0$. □

- If $\alpha = 0$, let $k = \infty$ and $\phi(x) = 1$ whenever $p_1(x)/p_0(x) = \infty$; this is $\gamma = 1$.
- If $\alpha = 1$, let $k = 0$ and $\gamma = 1$, so that we reject for all x with $p_1(x) > 0$ or $p_0(x) > 0$.

Definition 1.1 If the family of densities $\{p_\theta : \theta \in [\theta_0, \theta_1] \subset \mathbb{R}\}$ is such that $p_{\theta'}(x)/p_\theta(x)$ is nondecreasing in $T(x)$ for each $\theta < \theta'$, then the family is said to have *monotone likelihood ratio* (MLR).

Definition 1.2 A test is of *size* α if

$$\sup_{\theta\in\Theta_0} E_\theta \phi(X) = \alpha.$$

Let $\mathcal{C}_\alpha \equiv \{\phi :\ \phi \text{ is of size } \alpha\}$. A test ϕ_0 is *uniformly most powerful of size* α (UMP of size α) if it has size α and

$$E_\theta \phi_0(X) \ge E_\theta \phi(X) \qquad \text{for all } \theta \in \Theta_1 \text{ and all } \phi \in \mathcal{C}_\alpha.$$

Theorem 1.2 (Karlin - Rubin). Suppose that X has density p_θ with MLR in $T(x)$.
(i) Then there exists a UMP level α test of $H :\ \theta \le \theta_0$ versus $K :\ \theta > \theta_0$ which is of the form

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

with $E_{\theta_0}\phi(X) = \alpha$.
(ii) $\beta(\theta) = E_\theta \phi(X)$ is increasing in θ for $\beta < 1$.
(iii) For all θ' this same test is the UMP level $\alpha' \equiv \beta(\theta')$ test of $H' : \theta \le \theta'$ versus $K' : \theta > \theta'$.
(iv) For all $\theta < \theta_0$, the test of (i) minimizes $\beta(\theta)$ among all tests satisfying $\alpha = E_{\theta_0}\phi$.

Proof. (i) and (ii): The most powerful level α test of θ_0 versus $\theta_1 > \theta_0$ is the ϕ above, by the Neyman - Pearson lemma, which guarantees the existence of c and γ. Thus ϕ is UMP of θ_0 versus $\theta > \theta_0$. According to the NP lemma (ii), this same test is most powerful of θ' versus θ''; thus (ii) follows from the NP corollary. Thus ϕ is also level α in the smaller class of tests of H versus K; and hence is UMP there also: note that with $\mathcal{C}_\alpha \equiv \{\phi : \sup_{\theta \le \theta_0} E_\theta \phi = \alpha\}$ and $\mathcal{C}_\alpha^{\theta_0} \equiv \{\phi : E_{\theta_0}\phi \le \alpha\}$, $\mathcal{C}_\alpha \subset \mathcal{C}_\alpha^{\theta_0}$.
(iii) The same argument works.
(iv) To minimize power, just apply the NP lemma with inequalities reversed. □

Example 1.1 (Hypergeometric). Suppose that we sample without replacement n items from a population of N items of which $\theta = D$ are defective. Let $X \equiv$ number of defective items in the sample. Then

$$P_D(X = x) \equiv p_D(x) = \frac{\binom{D}{x}\binom{N-D}{n-x}}{\binom{N}{n}}, \qquad \text{for } \; x = 0 \vee (n - N + D), \ldots, D \wedge n.$$

Since

$$\frac{p_{D+1}(x)}{p_D(x)} = \frac{D+1}{N-D} \frac{N-D-n+x}{D+1-x}$$

is increasing in x, this family of distributions has MLR in $T(X) = X$. Thus the UMP test of $H : D \le D_0$ versus $K : D > D_0$ rejects H if X is "too big": $\phi(X) = 1\{X > c\} + \gamma 1\{X = c\}$ where

$$P_{D_0}(X > c) + \gamma P_{D_0}(X = c) = \alpha.$$

Reminder: $E(X) = nD/N$ and $Var(X) = n(D/N)(1 - D/N)(1 - (n-1)/(N-1))$.

Example 1.2 (One-parameter exponential families). Suppose that

$$p_\theta(x) = c(\theta) \exp(Q(\theta)T(x))h(x)$$

with respect to the dominating measure μ where $Q(\theta)$ is increasing in θ. Then

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

with $E_{\theta_0}\phi(X) = \alpha$ is UMP level α for testing $H : \theta \le \theta_0$ versus $K : \theta > \theta_0$. [See pages 70 - 71 in TSH for binomial, negative binomial, Poisson, and exponential examples].

Example 1.3 (Log-concave location family). Suppose that g is a density with respect to Lebesgue measure on $\mathbb{R}$, and let $p_\theta(x) = g(x-\theta)$, $\theta \in \mathbb{R}$. Then p_θ has MLR in x if and only if g is log-concave. To see that p_θ has MLR in x if g is log-concave, note that the MLR holds if and only if

$$\frac{g(x-\theta')}{g(x-\theta)} \le \frac{g(x'-\theta')}{g(x'-\theta)} \quad \text{for all } \; x < x', \quad \theta < \theta';$$

this holds if and only if

$$\log g(x-\theta') - \log g(x-\theta) \le \log g(x'-\theta') - \log g(x'-\theta),$$

or equivalently

$$\log g(x'-\theta) + \log g(x-\theta') \le \log g(x-\theta) + \log g(x'-\theta'). \tag{3}$$

Now let $t \equiv (x'-x)/(x'-x+\theta'-\theta)$ and note that

$$\begin{aligned} x-\theta &= t(x-\theta') + (1-t)(x'-\theta),\\ x'-\theta' &= (1-t)(x-\theta') + t(x'-\theta). \end{aligned}$$

Thus log-concavity of g implies that

$$\begin{aligned} &\log g(x-\theta) \ge t\log g(x-\theta') + (1-t)\log g(x'-\theta), \quad \text{and}\\ &\log g(x'-\theta') \ge (1-t)\log g(x-\theta') + t\log g(x'-\theta). \end{aligned}$$

Adding these yields (3), so MLR holds. To prove that the MLR property of p_θ implies that g is log-concave, let $a < b$ be any real numbers, and set $x-\theta' = a$, $x'-\theta = b$, $x-\theta = x'-\theta'$. But then $x-\theta = x'-\theta' = (a+b)/2$ and (3) becomes

$$\log g(a) + \log g(b) \le 2\log g((a+b)/2).$$

Since this holds for all $a, b \in \mathbb{R}$ and g is measurable, this implies that g is concave by a theorem of Sierpinski (1920).

Example 1.4 (Noncentral t, χ^2, and F distributions). The noncentral t, χ^2, and F distributions have MLR in their noncentrality parameters. See Lehmann and Romano, page 224 for the t distribution; see Lehmann and Romano problem 7.4, page 307 for the χ^2 and F distributions.

Example 1.5 (Counterexample: Cauchy location family). The Cauchy location family $p_\theta(x) = \pi^{-1}(1+(x-\theta)^2)^{-1}$ does *not* have MLR.

Theorem 1.3 (Generalized Neyman-Pearson lemma). Let $f_0, f_1, \ldots, f_m$ be real-valued, $\mu-$integrable functions defined on a Euclidean space $\mathcal{X}$. Let ϕ_0 be any function of the form

$$\phi_0(x) = \begin{cases} 1 & \text{if } f_0(x) > k_1 f_1(x) + \cdots + k_m f_m(x)\\ \gamma(x) & \text{if } f_0(x) = k_1 f_1(x) + \cdots + k_m f_m(x)\\ 0 & \text{if } f_0(x) < k_1 f_1(x) + \cdots + k_m f_m(x) \end{cases}$$

where $0 \le \gamma(x) \le 1$. Then ϕ_0 maximizes

$$\int \phi f_0 d\mu$$

over all ϕ, $0 \le \phi \le 1$ such that

$$\int \phi f_i d\mu = \int \phi_0 f_i d\mu, \qquad i = 1, \ldots, m.$$

If $k_j \ge 0$ for $j = 1, \ldots, m$, then ϕ_0 maximizes $\int \phi f_0 d\mu$ over all functions ϕ, $0 \le \phi \le 1$ such that

$$\int \phi f_i d\mu \le \int \phi_0 f_i d\mu, \qquad i = 1, \ldots, m.$$

Proof. Note that

$$\int (\phi_0 - \phi)(f_0 - \sum_{j=1}^{k} k_j f_j) d\mu \geq 0$$

since the integrand is ≥ 0 by the definition of ϕ_0. Hence

$$\int (\phi_0 - \phi) f_0 d\mu \geq \sum_{j=1}^{m} k_j \int (\phi_0 - \phi) f_j d\mu \geq 0$$

in either of the above cases, and hence

$$\int \phi_0 f_0 d\mu \geq \int \phi f_0 d\mu.$$

This is a "short form" of the generalized Neyman-Pearson lemma; for a "long form" with more details and existence results, see Lehmann and Romano, TSH, page 77. □

Example 1.6 Suppose that $X_1, \ldots, X_n$ are i.i.d. from the Cauchy location family

$$p(x;\theta) = \frac{1}{\pi} \frac{1}{1 + (x-\theta)^2},$$

and consider testing $H: \theta = \theta_0$ versus $\theta > \theta_0$. Can we find a test ϕ of size α such that ϕ maximizes

$$\frac{d}{d\theta}\beta_\phi(\theta_0) = \frac{d}{d\theta} E_\theta \phi(X)\Big|_{\theta=\theta_0}?$$

For any test ϕ the power is given by

$$\beta_\phi(\theta) = E_\theta \phi(\underline{X}) = \int \phi(\underline{x}) p(\underline{x};\theta) d\underline{x},$$

so, if the interchange of $d/d\theta$ and $\int$ is justifiable, then

$$\beta'_\phi(\theta) = \int \phi(\underline{x}) \frac{\partial}{\partial\theta} p(\underline{x};\theta) d\underline{x}.$$

Thus, by the generalized N-P lemma with $f_0(\underline{x}) = (\partial/\partial) p(\underline{x};\theta_0)$ and $f_1 = p(\underline{x};\theta_0)$, any test of the form

$$\phi(\underline{x}) = \begin{cases} 1 & \text{if } \frac{\partial}{\partial\theta} p(\underline{x};\theta_0) > k p(\underline{x};\theta_0) \\ \gamma(x) & \text{if } \frac{\partial}{\partial\theta} p(\underline{x};\theta_0) = k p(\underline{x};\theta_0) \\ 0 & \text{if } \frac{\partial}{\partial\theta} p(\underline{x};\theta_0) < k p(\underline{x};\theta_0) \end{cases}$$

maximizes $\beta'_\phi(\theta_0)$ among all ϕ with $E_{\theta_0}\phi(\underline{X}) \leq \alpha$. This test is said to be *locally most powerful of size* α; cf. Ferguson, section 5.5, page 235. But

$$\frac{\partial}{\partial\theta} p(\underline{x};\theta_0) > k p(\underline{x};\theta_0)$$

is equivalent to

$$\frac{\frac{\partial}{\partial\theta} p(\underline{x};\theta_0)}{p(\underline{x};\theta_0)} > k,$$

or

$$\frac{\partial}{\partial\theta}\log p(\underline{x};\theta_0) > k,$$

or

$$S_n(\theta_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \dot{\mathbf{l}}_\theta(X_i;\theta_0) > k'.$$

Hence for the Cauchy family (with $\theta_0 \equiv 0$ without loss of generality), since

$$\frac{\partial}{\partial\theta}\log p(x;\theta) = \frac{2(x-\theta)}{1+(x-\theta)^2},$$

the locally most powerful test is given by

$$\phi(\underline{X}) = \begin{cases} 1 & \text{if } \ n^{-1/2}\sum_{i=1}^{n}\frac{2X_i}{1+X_i^2} > k' \\ 0 & \text{if } \ n^{-1/2}\sum_{i=1}^{n}\frac{2X_i}{1+X_i^2} < k' \end{cases}$$

where k' is such that $E_0\phi(\underline{X}) = \alpha$. Under $\theta = \theta_0 \equiv 0$, with $Y_i \equiv 2X_i/(1+X_i^2)$,

$$E_0 Y_i = 0, \qquad \mathrm{Var}_0(Y_i) = 1/2 = I_{\mathrm{loc}}(\mathrm{Cauchy}).$$

Hence, by the CLT, k' may be approximated by $2^{-1/2}z_\alpha$ where $P(Z > z_\alpha) = \alpha$. (It would be possible refine this first order approximation to k' by way of an Edgeworth expansion; see e.g. Shorack (2000), page 392.)

Note that $x/(1+x^2) \to 0$ as $x \to \infty$. Thus, if $\alpha < 1/2$ so that $k' > 0$, the rejection set of ϕ is a bounded set in $\mathbb{R}^n$; and since the probability that $\underline{X} = (X_1,\ldots,X_n)$ is in any given bounded set tends to 0 as $\theta \to \infty$, $\beta_\phi(\theta) \to 0$ as $\theta \to \infty$.

Example 1.7 Now consider the calculations of the previous example, but for $p_\theta(x) = g(x-\theta)$ where g is log-concave with finite Fisher information for location $I_g \equiv \int (g'(x))^2/g(x)dx < \infty$. Suppose that $X_1,\ldots,X_n$ are i.i.d. p_θ for some $\theta \in \mathbb{R}$. It is easily seen that the locally most powerful test of $H : \theta = \theta_0$ versus $K : \theta > \theta_0$ is of the form

$$\phi(\underline{X}) = \begin{cases} 1, & \text{if } \ S_n(\theta_0) > k \\ 0, & \text{if } \ S_n(\theta_0) \le k \end{cases}$$

where

$$S_n(\theta_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{-\frac{g'}{g}(X_i-\theta_0)\}.$$

Since $S_n(\theta_0) \to_d N(0, I_g)$ under θ_0, taking $k = \sqrt{I_g}z_\alpha$ yields a test of approximate size α for n large. We are interested in monotonicity of the power function $\beta_\phi(\theta)$ of this test. Now $X_i \stackrel{d}{=} Y_i + \theta$ where Y_i are i.i.d. g, and by log-concavity of g we know that g'/g is decreasing and hence that $-g'/g$ is increasing. Thus it follows that under P_θ

$$S_n(\theta_0) \stackrel{d}{=} \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{-\frac{g'}{g}(Y_i+\theta-\theta_0)\right\}$$

increases as θ increases, and hence

$$\begin{aligned}\beta_\phi(\theta) &= P_\theta(S_n(\theta_0) > \sqrt{I_g} z_\alpha) \\ &= P_0\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n -\frac{g'}{g}(Y_i + \theta - \theta_0) > \sqrt{I_g} z_\alpha\right)\end{aligned}$$

is non-decreasing as a function of θ. This example includes the cases when g is Normal, Laplace, logistic, Gamma with shape parameter larger than 2, and many more.

Consistency of Neyman - Pearson tests

Let P and Q be probability measures, and suppose that p and q are their densities with respect to a common $\sigma-$ finite measure μ on $(\mathcal{X}, \mathcal{A})$. Recall that the Hellinger distance $H(P, Q)$ between P and Q is given by

$$H^2(P, Q) = \frac{1}{2}\int(\sqrt{p} - \sqrt{q})^2 d\mu = 1 - \int \sqrt{pq}\, d\mu = 1 - \rho(P, Q)$$

where $\rho(P, Q) \equiv \int \sqrt{pq}\, d\mu$ is the *Hellinger affinity* between P and Q.

Proposition 1.1 $H(P, Q) = 0$ if and only if $p = q$ a.e. μ if and only if $\rho(P, Q) = 1$. Furthermore $\rho(P, Q) = 0$ if and only if $\sqrt{p} \perp \sqrt{q}$ in the Hilbert space $L_2(\mu)$.

Recall that if $X_1, \ldots, X_n$ are i.i.d. P or Q with joint densities

$$p_n(\underline{x}) = p(\underline{x}) = \prod_{i=1}^n p(x_i), \qquad \text{or} \qquad q_n(\underline{x}) = q(\underline{x}) = \prod_{i=1}^n q(x_i),$$

then $\rho(P^n, Q^n) = \rho(P, Q)^n \to 0$ unless $p = q$ a.e. μ (which implies $\rho(P, Q) = 1$).

Theorem 1.4 (Size and power consistency of Neyman-Pearson type tests). For testing p versus q the test

$$\phi_n(\underline{x}) = \begin{cases} 1 & \text{if } q_n(\underline{x}) > k_n p_n(\underline{x}) \\ 0 & \text{if } q_n(\underline{x}) < k_n p_n(\underline{x}) \end{cases}$$

with $0 < a_1 \le k_n \le a_2 < \infty$ for all $n \ge 1$ is size and power consistent if $P \ne Q$: both probabilities of error converge to zero as $n \to \infty$. In fact,

$$\begin{aligned}E_P\phi_n(\underline{X}) &\le k_n^{-1/2}\rho(P, Q)^n \le a_1^{-1/2}\rho(P, Q)^n, \\ E_Q(1 - \phi_n(\underline{X})) &\le k_n^{1/2}\rho(P, Q)^n \le a_2^{1/2}\rho(P, Q)^n.\end{aligned}$$

Proof. For the type I error probability we have

$$\begin{aligned}E_P\phi_n(\underline{X}) &= \int \phi_n(\underline{x})p_n(\underline{x})d\mu(\underline{x}) = \int \phi_n(\underline{x})p_n^{1/2}(\underline{x})p_n^{1/2}(\underline{x})d\mu(\underline{x}) \\ &\le k_n^{-1/2}\int \phi_n(\underline{x})p_n^{1/2}(\underline{x})q_n^{1/2}(\underline{x})d\mu(\underline{x}) \\ &\le k_n^{-1/2}\int p_n^{1/2}(\underline{x})q_n^{1/2}(\underline{x})d\mu(\underline{x}) = k_n^{-1/2}\rho(P^n, Q^n) = k_n^{-1/2}\rho(P, Q)^n.\end{aligned}$$

The argument for type II errors is similar:

$$\begin{aligned} E_Q(1-\phi_n(\underline{X})) &= \int (1-\phi_n(\underline{x}))q_n(\underline{x})d\mu(\underline{x}) = \int (1-\phi_n(\underline{x}))q_n^{1/2}(\underline{x})q_n^{1/2}(\underline{x})d\mu(\underline{x}) \\ &\le k_n^{1/2}\int (1-\phi_n(\underline{x}))p_n^{1/2}(\underline{x})q_n^{1/2}(\underline{x})d\mu(\underline{x}) \\ &\le k_n^{1/2}\int p_n^{1/2}(\underline{x})q_n^{1/2}(\underline{x})d\mu(\underline{x}) = k_n^{1/2}\rho(P^n,Q^n) = k_n^{+1/2}\rho(P,Q)^n. \end{aligned}$$

Now suppose that $P = P_{\theta_0}$ and $Q = P_{\theta_n}$ where $P_\theta \in \mathcal{P} \equiv \{P_\theta : \theta \in \Theta\}$ is Hellinger differentiable at θ_0 and $\theta_n = \theta_0 + n^{-1/2}h$. Thus

$$\begin{aligned} nH^2(P_{\theta_0}, P_{\theta_n}) &= \frac{1}{2}n\int \{\sqrt{p_{\theta_n}} - \sqrt{p_{\theta_0}}\}^2 d\mu \\ &\to \frac{1}{2}\frac{1}{4}h^T I(\theta_0)h, \end{aligned}$$

and consequently

$$\begin{aligned} \rho(P_{\theta_0}, P_{\theta_n})^n &= \left(1 - \frac{nH^2(P_{\theta_0}, P_{\theta_n})}{n}\right)^n \\ &\to \exp\left(-\frac{1}{8}h^T I(\theta_0)h\right). \end{aligned}$$

Hence from the same argument used to prove Theorem 1.4,

(a) $$\limsup_{n\to\infty} E_{\theta_0}\phi_n(\underline{X}) \le a_1^{-1/2}\exp\left(-\frac{1}{8}h^T I(\theta_0)h\right),$$

while

(b) $$\limsup_{n\to\infty} E_{\theta_n}(1-\phi_n(\underline{X})) \le a_2^{+1/2}\exp\left(-\frac{1}{8}h^T I(\theta_0)h\right).$$

If we choose $k_n = a$ for all n and fix h and a so that $a^{-1/2}\exp(-h^T I(\theta_0)h/8) = \alpha$, then $\sqrt{a} = \alpha^{-1}\exp(-h^T I(\theta_0)h/8)$, and hence the RHS of (b) is given by $\alpha^{-1}\exp(-h^T I(\theta_0)h/4)$.

2 Unbiased Tests; Conditional Tests; Permutation Tests

2.1 Unbiased Tests

Notation: Consider testing

$$H: \ \theta \in \Theta_0 \qquad \text{versus} \qquad K: \ \theta \in \Theta_1$$

where $X \sim P_\theta$, for some $\theta \in \Theta = \Theta_0 + \Theta_1$, is observed. Let ϕ denote a critical (or test) function.

Definition 2.1 ϕ is *unbiased* if $\beta_\phi(\theta) \geq \alpha$ for all $\theta \in \Theta_1$ and $\beta_\phi(\theta) \leq \alpha$ for all $\theta \in \Theta_0$. ϕ is *similar on the boundary* (SOB) if

$$\beta_\phi(\theta) = \alpha \qquad \text{for all} \ \ \theta \in \overline{\Theta}_0 \cap \overline{\Theta}_1 \equiv \Theta_B.$$

Remark 2.1 If ϕ is a UMP level α test, then ϕ is unbiased. Proof: compare ϕ with the trivial test function $\phi_0 \equiv \alpha$.

Remark 2.2 If ϕ is unbiased and $\beta_\phi(\theta)$ is continuous for $\theta \in \Theta$, then ϕ is SOB. Proof: Let θ_n's in Θ_0 converge to $\theta_0 \in \Theta_B$. Then $\beta_\phi(\theta_0) = \lim_n \beta_\phi(\theta_n) \leq \alpha$. Similarly $\beta_\phi(\theta_0) \geq \alpha$ by considering θ_n's in Θ_1 converging to θ_0. Hence $\beta_\phi(\theta_0) = \alpha$.

Definition 2.2 A *uniformly most powerful unbiased level α test* is a test ϕ_0 for which

$$E_\theta \phi_0 \geq E_\theta \phi \qquad \text{for all} \ \ \theta \in \Theta_1$$

and for all unbiased level α tests ϕ.

Lemma 2.1 If $\mathcal{P} = \{P_\theta : \ \theta \in \Theta\}$ is such that $\beta_\phi(\theta)$ is continuous for all test functions ϕ, then if ϕ_0 is UMP SOB for H versus K and if ϕ_0 is level α for H versus K, then ϕ_0 is UMP unbiased (UMPU) for H versus K.

Proof. The unbiased tests are a subset of the SOB tests by remark 2.2. Since ϕ_0 is UMP SOB, it is thus at least as powerful as any unbiased test. But ϕ_0 is unbiased since its power is greater than or equal to that of the SOB test $\phi \equiv \alpha$, and since it is level α. Thus ϕ_0 is UMPU. □

Remark 2.3 For a multiparameter exponential family with densities

$$\frac{dP_\theta}{d\mu}(x) = c(\theta) \exp\left(\sum \theta_j T_j(x)\right),$$

with $\theta = (\theta_1, \ldots, \theta_d) \in \mathbb{R}^d$, the power function $\beta_\phi(\theta)$ is continuous in θ for all ϕ.

Proof. Apply theorem 2.7.1 of chapter 2 of Lehmann and Romano (2005) with $\phi \equiv 1$ to find that $c(\theta)$ is continuous; then apply it again with ϕ denoting an arbitrary critical function. □

2.2 Application to one-parameter exponential families

Suppose that

$$p_\theta(\underline{x}) = c(\theta)\exp(\theta T(\underline{x}))h(\underline{x})$$

for $\theta \in \Theta \subset \mathbb{R}$ with respect to a $\sigma-$finite measure μ on some subset of $\mathbb{R}^n$.

Problems: Test

(1)	$H_1: \ \theta \le \theta_0$	versus	$K_1: \ \theta > \theta_0$;
(2)	$H_2: \ \theta \le \theta_1$ or $\theta \ge \theta_2$	versus	$K_2: \ \theta_1 < \theta < \theta_2$;
(3)	$H_3: \ \theta_1 \le \theta \le \theta_2$	versus	$K_3: \ \theta < \theta_1$ or $\theta_2 < \theta$;
(4)	$H_4: \ \theta = \theta_0$	versus	$K_4: \ \theta \ne \theta_0$.

Theorem 2.1 (1) The test ϕ_1 with $E_{\theta_0}\phi_1(T) = \alpha$ given by

$$\phi_1(T(\underline{X})) = \begin{cases} 1 & \text{if } \ T(\underline{X}) > c \\ \gamma & \text{if } \ T(\underline{X}) = c \\ 0 & \text{if } \ T(\underline{X}) < c \end{cases}$$

is UMP for H_1 versus K_1.
(2) The test ϕ_2 with $E_{\theta_i}\phi_2(T) = \alpha$, $i = 1, 2$ given by

$$\phi_2(T(\underline{X})) = \begin{cases} 1 & \text{if } \ c_1 < T(\underline{X}) < c_2 \\ \gamma_i & \text{if } \ T(\underline{X}) = c_i \\ 0 & \text{if } \ \text{otherwise} \end{cases}$$

is UMP for H_2 versus K_2.
(3) The test ϕ_3 with $E_{\theta_i}\phi_3(T) = \alpha$, $i = 1, 2$ given by

$$\phi_3(T(\underline{X})) = \begin{cases} 1 & \text{if } \ T(\underline{X}) < c_1 \ \text{ or } \ T(\underline{X}) > c_2 \\ \gamma_i & \text{if } \ T(\underline{X}) = c_i \\ 0 & \text{if } \ \text{otherwise} \end{cases}$$

is UMPU for H_3 versus K_3.
(4) The test ϕ_4 with $E_{\theta_0}\phi_4(T) = \alpha$ and $E_{\theta_0}T\phi_4(T) = \alpha E_{\theta_0}T$ given by

$$\phi_4(T(\underline{X})) = \begin{cases} 1 & \text{if } \ T(\underline{X}) < c_1 \ \text{ or } \ T(\underline{X}) > c_2 \\ \gamma_i & \text{if } \ T(\underline{X}) = c_i \\ 0 & \text{if } \ \text{otherwise} \end{cases}$$

is UMPU for H_4 versus K_4. Furthermore, if T is symmetrically distributed about a under θ_0, then $E_{\theta_0}\phi_4(T) = \alpha$, $c_2 = 2a - c_1$ and $\gamma_1 = \gamma_2$ determine the constants. The characteristic behavior of the power of these four tests is as follows:

Proof. (1) and (2) were proved earlier using the NP lemma (via MLR) and its generalized version respectively. For (2), see pages 81-82 in Lehmann and Romano (2005). For (3), see Lehmann and Romano (2005), page 121.

(4) We need only consider tests $\phi(\underline{x}) = \psi(T(\underline{x}))$ based on the sufficient statistic T, whose distribution is of the form $p_\theta(t) = c(\theta)e^{\theta t}$ with respect to some $\sigma-$finite measure ν. Since all power functions are continuous in the case of an exponential family, it follows that any unbiased test ψ satisfies $\alpha = \beta_\psi(\theta_0) = E_{\theta_0}\psi(T)$ and has a minimum at θ_0.

But by theorem 2.7.1, chapter 2, TSH, β_ψ is differentiable, and can be differentiated under the integral sign; hence

$$\begin{aligned}\beta'_\psi(\theta) &= \frac{d}{d\theta}\int \psi(t)c(\theta)\exp(\theta t)d\nu(t)\\ &= \frac{c'(\theta)}{c(\theta)}E_\theta\psi(T) + E_\theta(T\psi(T))\\ &= (-E_\theta T)E_\theta\psi(T) + E_\theta(T\psi(T))\end{aligned}$$

since, with $\psi_0 \equiv \alpha$, $0 = \beta'_{\psi_0}(\theta) = c'(\theta)/c(\theta) + E_\theta(T)$. Thus

$$0 = \beta'_\psi(\theta_0) = E_{\theta_0}(T\psi(T)) - \alpha E_{\theta_0}T.$$

Thus any unbiased test $\psi(T)$ satisfies the two conditions of the statement of our theorem. We will apply the generalized NP lemma to show that ϕ as given is UMPU.

Let

$$M \equiv \{(E_{\theta_0}\psi(T), E_{\theta_0}T\psi(T)) : \quad \psi(T) \ \text{ is a critical function}\}.$$

Then M is convex and contains $\{(u, uE_{\theta_0}T) : \quad 0 < u < 1\}$. Also M contains points (α, v) with $v > \alpha E_{\theta_0}T$; since, by problem 18 of chapter 3, Lehmann TSH, there exist tests (UMP one-sided ones) having $\beta'(\theta_0) > 0$. Likewise M contains points (α, v) with $v < \alpha E_{\theta_0}T$. Hence $(\alpha, \alpha E_{\theta_0}T)$ is an interior point of M.

Thus, by the generalized NP lemma (iv), there exist k_1, k_2 such that

$$\begin{aligned}\psi(t) &= \begin{cases}1 & \text{when } c(\theta_0)(k_1 + k_2 t)e^{\theta_0 t} < c(\theta')e^{\theta' t}\\ 0 & \text{when } c(\theta_0)(k_1 + k_2 t)e^{\theta_0 t} > c(\theta')e^{\theta' t}\end{cases}\\ \text{(a)} \qquad &= \begin{cases}1 & \text{when } a_1 + a_2 t < e^{bt}\\ 0 & \text{when } a_1 + a_2 t > e^{bt}\end{cases}\end{aligned}$$

having the property that it maximizes $E_{\theta'}\psi(T)$. But the region described in (a) is either one-sided or else the complement of an interval. By the Karlin-Rubin theorem 6.1.2 it cannot be one-sided (since one-sided tests have strictly monotone power functions violating $\beta'(\theta_0) = 0$). Thus

$$\text{(b)} \qquad \psi(T) = \begin{cases}1 & \text{if } T < c_1 \text{ or } T > c_2\\ 0 & \text{if } c_1 < T < c_2.\end{cases}$$

Since this test does not depend on $\theta' \neq \theta_0$, it is the UMP (within the class of level α tests having $\beta'(\theta_0) = 0$) test of H_4 versus K_4. Since $\psi_0 \equiv \alpha$ is in this class, ψ is unbiased. And this class of test includes the unbiased tests. Hence ψ is UMPU.

If T is distributed symmetrically about some point a under θ_0, then any test ψ symmetric about a that satisfies $E_{\theta_0}\psi(T) = \alpha$ will also satisfy

$$E_{\theta_0}T\psi(T) = E_{\theta_0}(T-a)\psi(T) + aE_{\theta_0}\psi(T) = 0 + a\alpha = \alpha E_{\theta_0}T$$

automatically. $\square$

2.3 UMPU tests for families with nuisance parameter via conditioning

Definition 2.3 Let T be sufficient for $\mathcal{P}_B \equiv \{P_\theta : \theta \in \Theta_B\}$, and let $\mathcal{P}^T \equiv \{P_\theta^T : \theta \in \Theta_B\}$. A test function ϕ is said to have *Neyman structure with respect to* T if

$$E(\phi(X)|T) = \alpha \qquad \text{a.s. } \mathcal{P}^T.$$

Remark 2.4 If ϕ has Neyman structure with respect to T, then ϕ is SOB.

Proof. $E_\theta \phi(X) = E_\theta E(\phi(X)|T) = E_\theta \alpha = \alpha$ for all $\theta \in \Theta_B$. $\Box$

Theorem 2.2 Let X be a random variable with distribution $P_\theta \in \mathcal{P} = \{P_\theta : \theta \in \Theta\}$, and let T be sufficient for $\mathcal{P}_B = \{P_\theta : \theta \in \Theta_B\}$. Then all SOB tests have Neyman structure with respect T if and only if the family of distributions $\mathcal{P}^T \equiv \{P_\theta^T : \theta \in \Theta_B\}$ is boundedly complete: i.e. if $E_P h(T) = 0$ for all $P \in \mathcal{P}^T$ with h bounded, then $h = 0$ a.e. $\mathcal{P}^T$.

Proof. Suppose that $\mathcal{P}^T$ is boundedly complete. Let ϕ be a SOB level α test; and define $\psi(T) \equiv E(\phi(X)|T)$. Now

$$\begin{aligned} E_\theta(\psi(T) - \alpha) &= E_\theta(E(\phi(X)|T)) - \alpha \\ &= E_\theta \phi(X) - \alpha = 0 \end{aligned}$$

for all $\theta \in \Theta_B$, and since $\psi(T) - \alpha$ is bounded, the bounded completeness of $\mathcal{P}^T$ implies $\psi(T) = \alpha$ a.e. $\mathcal{P}^T$. Hence $\alpha = \psi(T) = E(\phi(X)|T)$ a.e. $\mathcal{P}^T$, and ϕ has Neyman structure with respect to T.

Now suppose that all SOB tests have Neyman structure. Assume $\mathcal{P}^T$ is *not* boundedly complete. Then there exists h such that $|h| \le$some M with $E_\theta h(T) = 0$ for all $\theta \in \Theta_B$ and $h(T) \ne 0$ with probability > 0 for some $\theta_0 \in \Theta_B$. Define $\phi(T) \equiv ch(T) + \alpha$ where $c \equiv \{\alpha \wedge (1 - \alpha)\}/M$. Then $0 \le \phi(T) \le 1$ so ϕ is a critical function, and $E_\theta \phi(T) = \alpha$ for all $\theta \in \Theta_B$, so that ϕ is SOB. But $E(\phi(T)|T) = \phi(T) \ne \alpha$ with probability > 0 for the above θ_0, so ϕ does not have Neyman structure. This is a contradiction, and hence it follows that indeed $\mathcal{P}^T$ is boundedly complete. $\Box$

Remark 2.5 Suppose that:
(i) All critical functions ϕ have continuous power functions β_ϕ.
(ii) T is sufficient for $\mathcal{P}_B = \{P_\theta : \theta \in \Theta_B\}$ and $\mathcal{P}^T \equiv \{P_\theta^T : \theta \in \Theta_B\}$ is boundedly complete.
(Remark 2.3 says that (i) is always true for exponential families $p_\theta(x) = c(\theta)\exp(\sum \theta_j T_j(x))$; and theorem 4.3.1, TSH, page 116, allows us to check (ii) for these same families.) Then all unbiased tests are SOB and all SOB tests have Neyman structure. Thus if we can find a UMP Neyman structure test ϕ_0 and we can show that ϕ_0 is unbiased, then ϕ_0 is UMPU. Why is it easier to find UMP Neyman structure tests? Neyman structure tests are characterized by having conditional probability of rejection equal to α on each surface $T = t$. But the distribution on each such surface is independent of $\theta \in \Theta_B$ because T is sufficient for $\mathcal{P}^T$. Thus the problem has been reduced to testing a one parameter hypothesis for each fixed value of t; and in many problems we can easily find the most powerful test of this simple hypothesis.

2.4 Application to general exponential families; $k-$parameter

Consider the exponential family $\mathcal{P} = \{P_{\theta,\xi}\}$ given by

$$p_{\theta,\xi}(\underline{x}) = c(\theta,\xi)\exp\left(\theta U(\underline{x}) + \sum_{i=1}^{k}\xi_i T_i(\underline{x})\right)$$

with respect to a $\sigma-$finite dominating measure μ on some subset of $\mathbb{R}^n$ where Θ is convex, has dimension $k+1$, and contains interior points θ_i, $i = 1, 2$.

Problems: Test

(1)	$H_1: \ \theta \le \theta_0$	versus	$K_1: \ \theta > \theta_0$;
(2)	$H_2: \ \theta \le \theta_1$ or $\theta \ge \theta_2$	versus	$K_2: \ \theta_1 < \theta < \theta_2$;
(3)	$H_3: \ \theta_1 \le \theta \le \theta_2$	versus	$K_3: \ \theta < \theta_1$ or $\theta_2 < \theta$;
(4)	$H_4: \ \theta = \theta_0$	versus	$K_4: \ \theta \ne \theta_0$.

Theorem 2.3 The following are UMPU tests for the hypothesis testing problems 1-4 respectively:
(1) The test ϕ_1 given by

$$\phi_1(\underline{x}) = \begin{cases} 1 & \text{if } \ U > c(t) \\ \gamma(t) & \text{if } \ U = c(t) \\ 0 & \text{if if } \ U < c(t) \end{cases}$$

where $E_{\theta_0}(\phi_1(U)|T = t) = \alpha$ is UMPU for H_1 versus K_1.
(2) The test ϕ_2 given by

$$\phi_2(\underline{x}) = \begin{cases} 1 & \text{if } \ c_1(t) < U < c_2(t) \\ \gamma_i(t) & \text{if } \ U = c_i(t) \\ 0 & \text{if if else} \end{cases}$$

where $E_{\theta_i}(\phi_2(U)|T = t) = \alpha$, $i = 1, 2$, is UMPU for H_2 versus K_2.
(3) The test ϕ_3 given by

$$\phi_3(\underline{x}) = \begin{cases} 1 & \text{if } \ U < c_1(t) \ \text{ or } \ U > c_2(t) \\ \gamma_i(t) & \text{if } \ U = c_i(t) \\ 0 & \text{if if else} \end{cases}$$

where $E_{\theta_i}(\phi_3(U)|T = t) = \alpha$, $i = 1, 2$ is UMPU for H_3 versus K_3.
(4) The test ϕ_4 given by

$$\phi_4(\underline{x}) = \begin{cases} 1 & \text{if } \ U < c_1(t) \ \text{ or } \ U > c_2(t) \\ \gamma_i(t) & \text{if } \ U = c_i(t) \\ 0 & \text{if if else} \end{cases}$$

where $E_{\theta_0}(\phi_4(U)|T = t) = \alpha$ and $E_{\theta_0}\{U\phi_4(U)|T = t\} = \alpha E_{\theta_0}\{U|T = t\}$ is UMPU for H_4 versus K_4.

Remark 2.6 If $V = h(U, T)$ is increasing in U for each fixed t and is independent of T on Θ_B, then

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

is UMPU in (1).

Remark 2.7 If $h \equiv h(U,T) = a(t)U + b(t)$ with $a(t) > 0$, then the second constraint in (4) becomes

$$E_{\theta_0}\left\{\frac{V-b(t)}{a(t)}\phi|T=t\right\} = \alpha E_{\theta_0}\left\{\frac{V-b(t)}{a(t)}|T=t\right\}$$

or $E_{\theta_0}(V\phi|T=t) = \alpha E_{\theta_0}(V|T=t)$, and if this V is independent of T on the boundary, then the test is unconditional.

Example 2.1 (Comparing two Poisson distributions). Suppose that $X \sim$ Poisson(μ) and $Y \sim$ Poisson(ν) are independent. Consider testing $H : \nu \le \mu$ versus $K : \nu > \mu$. Then $\Theta_B = \{(\mu,\mu) : \mu \ge 0\}$. In this case we have

$$\begin{aligned} p_{\mu,\nu}(x,y) &= p(x,y;\mu,\nu) = \exp(-\mu)\frac{\mu^x}{x!}\cdot\exp(-\nu)\frac{\nu^y}{y!} \quad \text{for } x,y \in \{0,1,2,\ldots\} \\ &= c(\mu,\nu)\exp(x\log\mu + y\log\nu)\cdot\frac{1}{x!y!} \\ &= c(\mu,\nu)\exp(y\log(\nu/\mu) + (x+y)\log\mu)h(x,y). \end{aligned}$$

Thus our general theory applies with $\theta \equiv \log(\nu/\mu)$, $\xi = \log\mu$, $U(x,y) = y$, $T(x,y) = x+y$, where $T = X+Y$ is sufficient for $\xi = \log\mu$ for $(\mu,\nu) \in \Theta_B$. Note that testing H versus K is equivalent to testing $\theta \le 0$ versus $\theta > 0$. Therefore we carry out our test conditionally on T: noting that $(Y|T=t) \sim$ Binomial($t, \nu/(\mu+\nu)$) which reduces to Binomial $(t, 1/2)$ on the boundary $\mu = \nu$. Thus the UMPU test of H versus K of size α is given by $\phi(X,Y) = 1\{Y > c_t\} + \gamma_t 1\{Y = c_t\}$ where c_t, γ_t satisfy $P(\text{Bin}(t,1/2) > c_t) + \gamma_t P(\text{Bin}(t,1/2) = c_t) = \alpha$. [See Lehmann and Romano (2005), Section 4.5, pages 124-127.]

Example 2.2 (Comparing two Binomial distributions). Now suppose that $X \sim$ Binomial(m,p) and $Y \sim$ Binomial(n,q) are independent and consider testing $H : q \le p$ versus $K :\ q > p$. Then $\Theta_B = \{(p,p) :\ p \in [0,1]\}$. In this case we have

$$\begin{aligned} p_{p,q}(x,y) &= p(x,y;p,q) = \binom{m}{x}p^x(1-p)^{m-x}\cdot\binom{n}{y}q^y(1-q)^{n-y} \\ &= c(p,q)\exp\left(x\log\frac{p}{1-p} + y\log\frac{q}{1-q}\right)h(x,y) \\ &= c(p,q)\exp\left(y\log\frac{q/(1-q)}{p/(1-p)} + (x+y)\log\frac{p}{1-p}\right)h(x,y) \\ &\equiv c(p,q)\exp\left(y\theta + (x+y)\xi\right)h(x,y) = c(p,q)\exp\left(U(x,y)\theta + T(x,y)\xi\right)h(x,y) \end{aligned}$$

where

$$\theta \equiv \log\frac{q/1-q)}{p/(1-p)}$$

is the log of the *odds ratio*. Thus testing H versus K is equivalent to testing $\theta \le 0$ versus $\theta > 0$. In this case we carry out the test conditionally on $T = X+Y$. Let $\rho \equiv e^\theta$. Upon noting that

$$P(Y=y|X+Y=t) = C_t(\rho)\binom{n}{y}\binom{m}{t-y}\rho^y, \quad y \in \{0,\ldots,t\wedge y\}$$

which reduces to the hypergeometric distribution $(n,t,m+n)$ when $\theta = 0$ (and hence $\rho = 1$):

$$P_0(Y=y|X+Y=t) = \frac{\binom{n}{y}\binom{m}{t-y}}{\binom{m+n}{t}}, \quad y \in \{0,\ldots,t\wedge y\}.$$

Thus the UMPU test of H versus K of size α is given by $\phi(X,Y) = 1\{Y > c_t\} + \gamma_t 1\{Y = c_t\}$ satisfy $E_0\{\Phi(X,Y)|T = t\} = \alpha$.

Example 2.3 Suppose that $X_1, \ldots, X_n$ are i.i.d. $N(\mu, \sigma^2)$. Consider testing $H : \mu \le \mu_0$ versus $K : \mu > \mu_0$. Then

$$\begin{aligned} p(\underline{x}; \mu, \sigma^2) &= (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \right) \\ &= c(\mu, \sigma^2) \exp\left(\frac{n\mu}{\sigma^2}\overline{x} + \left(-\frac{1}{2\sigma^2}\right) \sum_{i=1}^n x_i^2 \right) \\ &= c(\mu, \sigma^2) \exp\left(\theta U(\underline{x}) + \xi T(\underline{x})\right) \end{aligned}$$

where $\theta \equiv n\mu/\sigma^2$, $\xi \equiv -1/(2\sigma^2)$, $U(\underline{x}) = \overline{x}$, and $T(\underline{x}) = \sum_{i=1}^n x_i^2$. Without loss of generality we can assume that $\mu_0 = 0$. (If not, replace the X_i's by $X_i - \mu_0$.) Now testing H versus K is equivalent to testing $H : \theta \le 0$ versus $K : \theta > 0$. Our general theory says that we should proceed by conditioning on the sufficient statistic T for the nuisance parameter ξ and carrying out the test conditionally on $T = t$: the UMPU conditional test is of the form $\phi(\underline{X})$ is of the form $\phi(\underline{X}) = 1\{U(\underline{X}) > c_t\}$ where c_t is chosen so that

$$E_\theta(\phi(\underline{X})|T = t\} = \alpha \quad \text{for each fixed} \quad t.$$

Since T is sufficient for ξ, the expectation in the last display does not depend on ξ for (θ, ξ) in Θ_B. But

$$V \equiv \frac{U}{\sqrt{T - nU^2}} = \frac{\overline{X}_n}{\sqrt{\sum_1^n X_i^2 - n\overline{X}_n^2}} \equiv h(U,T)$$

is monotone increasing in $U = \overline{X}_n$ for each fixed value of T. Moreover for $(\theta, \sigma^2) \in \Theta_B = \{0\} \times [0, \infty)$, the distribution of V does not depend on σ^2; i.e. V is ancillary. Since T is sufficient and complete for Θ_B, it follows from Basu's theorem that V and T are independent. Thus the test can be performed unconditionally: with

$$t_n \equiv \frac{\sqrt{n}\overline{X}_n}{\sqrt{\frac{1}{n-1}\sum_1^n (X_i - \overline{X}_n)^2}} = \sqrt{n(n-1)}V,$$

the test becomes $\phi(\underline{X}) = 1\{t_n \ge t_{n-1,\alpha}\}$ where $t_n \sim$ Student's t-distribution with $n-1$ degrees of freedom under Θ_B.

The following material is from Efron (1969). Another more geometric perspective is as follows: Note that under $\theta = 0$ the random vector

$$\frac{\underline{X}}{\|\underline{X}\|} = \frac{\underline{X}}{\sqrt{\sum_1^n X_i^2}}$$

takes values in the unit sphere S^{n-1} in $\mathbb{R}^n$: $S^{n-1} \equiv \{\underline{x} \in \mathbb{R}^n : \ \|\underline{x}\| = 1\}$. Moreover $\underline{X}/\|\underline{X}\|$ has a Uniform distribution on S^{n-1}. Also note that

$$\sqrt{n}V = \frac{\sqrt{n}\overline{X}_n/(\sum_1^n X_i^2)^{1/2}}{\left(1 - \overline{X}_n^2/(\sum_1^n X_i^2)\right)}$$

is a monotone increasing function of

$$S_n \equiv \frac{\sqrt{n}\overline{X}_n}{\|\underline{X}\|} = \langle \underline{X}/\|\underline{X}\|, \underline{1}n^{-1/2}\rangle = \cos(\Gamma_n)$$

where Γ_n is the angle between $\underline{X}/\|\underline{X}\| \in S^{n-1}$ and $\underline{1}n^{-1/2} \in S^{n-1}$. Since $\cos(w)$ is a decreasing function of $w \in [0, \pi/2]$, rejecting for large values of S_n corresponds to rejecting for small values of Γ_n, and our test becomes $\phi(\underline{X}) = 1\{\Gamma_n \le \gamma_{n,\alpha}\}$ where $\gamma_{n,\alpha}$ is the angle chosen so that $P_0(\underline{X}/\|\underline{X}\| \in \text{Cap}_{n,\alpha}) = \alpha$ where $\text{Cap}_{n,\alpha}$ denotes a spherical cap centered at $n^{-1/2}\underline{1}$ with area $\alpha\cdot$ Area of S^{n-1}. As it turns out, the area of such a cap is given by

$$\frac{1}{2}A_n I_{\sin^2(\gamma)}((n-1)/2, 1/2)$$

where $A_n = 2\pi^{n/2}\Gamma(n/2)$ and $I_x(a,b)$ is the Beta(a,b) distribution function

$$F_{a,b}(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\int_0^x t^{a-1}(1-t)^{b-1}dt.$$

Example 2.4 (Comparing two normal means when variances are equal – or not). Now suppose that

$$\begin{aligned} &X_1, \ldots, X_m \ \text{ are i.i.d. } N(\mu, \sigma^2), \quad \text{and} \\ &Y_1, \ldots, Y_n \ \text{ are i.i.d. } N(\nu, \tau^2), \end{aligned}$$

and the X's and Y's are independent. Consider testing $H: \ \mu \ge \nu$ versus $K: \ \mu < \nu$. The joint density of the X's and Y's is given by

$$p(\underline{x}, \underline{y}; \mu, \nu, \sigma^2, \tau^2) \quad = \quad c(\mu, \nu, \sigma^2, \tau^2)\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^m x_i^2 - \frac{1}{2\tau^2}\sum_{j=1}^n y_j^2 + \frac{\mu}{\sigma^2}\sum_{i=1}^m x_i + \frac{\nu}{\tau^2}\sum_{j=1}^n y_j\right).$$

Case 1: (Equal variances). Suppose that $\tau^2 = \sigma^2$. In this case we note that

$$m\mu\overline{x} + n\nu\overline{y} = \frac{mn}{N}(\overline{y} - \overline{x})\cdot(\nu - \mu) + (m\overline{x} + n\overline{y})\frac{m\mu + n\nu}{N}.$$

Thus the joint density can be rewritten in this case as

$$\begin{aligned} &c(\mu, \nu, \sigma^2, \tau^2)\exp\left(\frac{mn}{N}(\overline{y} - \overline{x})\frac{\nu - \mu}{\sigma^2} + (m\overline{x} + n\overline{y})\frac{m\mu + n\nu}{N\sigma^2} - \frac{1}{2\sigma^2}(\sum_{i=1}^m x_i^2 + \sum_{j=1}^n y_j^2)\right) \\ &= \quad c(\theta, \underline{\xi})\exp\left(\theta U(\underline{x}, \underline{y}) + \xi_1 T_1(\underline{x}, \underline{y}) + +\xi_2 T_2(\underline{x}, \underline{y})\right) \end{aligned}$$

where

$$\begin{aligned} &\theta \equiv \frac{mn}{N}\frac{\nu - \mu}{\sigma^2}, \quad \xi_1 \equiv \frac{m\mu + n\nu}{N\sigma^2}, \quad \xi_2 \equiv -\frac{1}{2\sigma^2}, \\ &U(\underline{X}, \underline{Y}) = \overline{Y} - \overline{X}, \ \ T_1 \equiv m\overline{X} + n\overline{Y}, \ \ T_2 \equiv \sum_{i=1}^m X_i^2 + \sum_{j=1}^n Y_j^2. \end{aligned}$$

Thus the UMPU tests of level α of $H_1 : \theta \le 0$ versus $K_1 : \theta > 0$, or equivalently H versus K, are of the form given in Theorem 2.3. Now note that

$$\begin{aligned} V &\equiv \frac{U}{\sqrt{T_2 - N^{-1}T_1^2 - \frac{mn}{N}U^2}} \equiv h(U, T_1, T_2) \\ &= \frac{\overline{Y} - \overline{X}}{\sqrt{\sum_1^m (X_i - \overline{X}_m)^2 + \sum_1^n (Y_j - \overline{Y})^2}} \end{aligned}$$

is a monotone increasing function of U for fixed values of T_1, T_2. Furthermore, under the null hypothesis $\mu = \nu$ the distribution of V does not depend on the common mean μ or on σ^2. Thus V is ancillary and by Basu's theorem it is independent of (T_1, T_2). Thus the rejection region of the UMPU test can be written as $t(\underline{X}, \underline{Y}) \ge c$ where

$$t(\underline{X}, \underline{Y}) \equiv \frac{\sqrt{\frac{mn}{N}}(\overline{Y} - \overline{X})}{\sqrt{\frac{\sum_1^m (X_i - \overline{X}_m)^2 + \sum_1^n (Y_j - \overline{Y})^2}{m+n-2}}} = \sqrt{\frac{mn(N-2)}{N}} V$$

has a $t_{N-2}-$distribution under $\mu = \nu$ and a non-central $t_{N-2}(\delta)$ distribution with $\delta \equiv \sqrt{mn/N}(\nu - \mu)/\sigma$ under the normality assumption with equal variances. [See Lehmann and Romano (2005), pages 159 - 161.]

Case 2: (Different variances). Suppose that $\tau^2 \ne \sigma^2$. In this case the joint density of all the X's and Y's on the boundary $\mu = \nu$ is given by

$$p(\underline{x}, \underline{y}; \mu, \mu, \sigma^2, \tau^2) = c(\mu, \mu, \sigma^2, \tau^2) \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^m x_i^2 - \frac{1}{\tau^2} \sum_{j=1}^n y_j^2 + \frac{\mu}{\sigma^2} \sum_{i=1}^m x_i + \frac{\mu}{\tau^2} \sum_{j=1}^n y_j \right).$$

The statistic

$$T = \left(\sum_1^m X_i, \sum_1^m X_i^2, \sum_1^n Y_j, \sum_1^n Y_j^2 \right)$$

is sufficient, but not complete (and also not boundedly complete) since $E_{\mu,\mu}(\overline{Y}_n - \overline{X}_m) = 0$ identically for all $(\mu, \mu, \sigma^2, \tau^2)$ (and similarly for a large class of bounded functions $h(\overline{Y_n} - \overline{X}_m)$). Thus the program of reducing by sufficiency fails. This is known as the *Behrens-Fisher* problem: see Lehmann and Romano (2005), Sections 6.6, 11.3.1, and 15.2; and Example 13.5.4. Reasonable alternative tests including the Welch approximate $t-$ test, are discussed in Lehmann and Romano (2005), section 11.3, pages 447-448.

Example 2.5 (Paired normals with nuisance shifts).

2.5 Permutation Tests

Consider testing

$$H_c : \; X_1, \ldots, X_m, Y_1, \ldots, Y_n \qquad \text{are i.i.d. with df } \; F \in \mathcal{F}_c$$

where $\mathcal{F}_c$ is the collection of all continuous distribution functions on $\mathbb{R}$, versus

$$K_1 : \; X_1, \ldots, X_m, Y_1, \ldots, Y_n \qquad \text{have joint density function } \; h.$$

We seek a most powerful similar test: ϕ is *similar* if

$$E_{(F,F)}\phi(\underline{X},\underline{Y}) = \alpha \qquad \text{for all } \; F \in \mathcal{F}_c. \tag{1}$$

But if $\underline{Z} \equiv (Z_1, \ldots, Z_N)$ with $N \equiv m+n$ denotes the ordered values of the combined sample $X_1, \ldots, X_m, Y_1, \ldots, Y_n$, then when H_c is true, $\underline{Z}$ is sufficient and complete; see e.g. Lehmann and Romano, TSH, page 118. Hence (1) holds if and only if (by theorem 2.2)

$$\begin{aligned} &E(\phi(\underline{X},\underline{Y})|\underline{Z} = \underline{z}) = \alpha \qquad \text{for a.e. } \; \underline{z} = (z_1, \ldots, z_N) \\ &\quad = \sum_{\pi \in \Pi} \phi(\pi \underline{z}) \frac{1}{N!} = \sum_{\underline{z}'} \phi(\underline{z}') \frac{1}{N!} \end{aligned} \tag{2}$$

where the sum is over all $N!$ permutations $\underline{z}'$ of $\underline{z}$. Thus if $\alpha = I/N!$, then any test which is performed conditionally on $\underline{Z} = \underline{z}$ and rejects for exactly I of the $N!$ permutations $\underline{z}'$ of $\underline{z}$ is a level α similar test; moreover (2) says that any level α similar test is of this form.

Definition 2.4 Tests satisfying (2) are called *permutation tests*. (Thus a test of H_c versus K_1 is similar if and only if it is a permutation test.)

We now need to find a most powerful permutation test by maximizing the conditional power. But

$$E_h(\phi(\underline{X},\underline{Y})|\underline{Z} = \underline{z}) = \sum_{\underline{z}'} \phi(\underline{z}') \frac{h(\underline{z}')}{\sum h(\underline{z}'')}.$$

Since the conditional densities under the composite null hypothesis and under the simple alternative h are

$$p_0(\underline{z}'|\underline{z}) = \frac{1}{N!} \qquad \text{and} \qquad p_1(\underline{z}'|\underline{z}) = \frac{h(\underline{z}')}{\sum h(\underline{z}'')}, \quad \underline{z}' \in \{\pi \underline{z} : \; \pi \in \Pi\},$$

the conditional power is maximized by rejecting for large values of

$$\frac{p_1(\underline{z}'|\underline{z})}{p_0(\underline{z}'|\underline{z})} = K_{\underline{z}} h(\underline{z}') \qquad \text{with } \; K_{\underline{z}} = \frac{N!}{\sum h(\underline{z}'')}.$$

Thus, at level $\alpha = I/N!$ we reject if

$$h(\underline{z}') > c(\underline{z})$$

where $c(\underline{z})$ is chosen so that we reject for exactly I of the $N!$ permutations $\underline{z}'$ of $\underline{z}$; or else we use a randomized version of such a test.

Example 2.6 Suppose now that we specify a particular alternative:

$$K_1 : \begin{array}{ll} X_1, \ldots, X_m & \text{are i.i.d. } \; N(\theta_1, \sigma^2) \\ Y_1, \ldots, Y_n & \text{are i.i.d. } \; N(\theta_2, \sigma^2) \end{array}$$

where $\theta_1 < \theta_2$ and σ^2 are fixed constants. Then the similar test of H_c that is most powerful again this simple K_1 rejects H for those permutations $\underline{z}'$ of $\underline{z}$ which lead to large values of

$$(2\pi\sigma^2)^{-N/2} \exp\left\{ -\frac{1}{2\sigma^2} \left(\sum_1^m (X_i - \theta_1)^2 + \sum_1^n (Y_j - \theta_2)^2 \right) \right\},$$

or small values of

$$\sum_1^m (X_i - \theta_1)^2 + \sum_1^n (Y_j - \theta_2)^2$$
$$= \sum_1^m X_i^2 + \sum_1^n Y_j^2 + m\theta_1^2 + n\theta_2^2 - 2\theta_1 \sum_{i=1}^m X_i - 2\theta_2 \sum_{j=1}^n Y_j,$$

or large values of

$$\theta_1 \sum_1^m X_i + \theta_2 \sum_1^n Y_j - \frac{m\theta_1 + n\theta_2}{N}(\sum_1^m X_i + \sum_1^n Y_j)$$
$$= \frac{mn}{N}(\theta_2 - \theta_1)(\overline{Y} - \overline{X}),$$

or large values of

$$\overline{Y} - \overline{X},$$

or large values of

$$\theta_1 \sum_1^m X_i + \theta_2 \sum_1^n Y_j - \theta_1(\sum_1^m X_i + \sum_1^n Y_j) = (\theta_2 - \theta_1)\sum_{j=1}^n Y_j,$$

or large values of

$$\frac{\sqrt{\frac{mn}{N}}(\overline{Y} - \overline{X})}{\sqrt{\frac{1}{N-2}\left\{\sum Z_i^2 - \frac{(\sum Z_i)^2}{N} - \frac{mn}{N}(\overline{Y} - \overline{X})^2\right\}}}$$
$$= \frac{\sqrt{\frac{mn}{N}}(\overline{Y} - \overline{X})}{\sqrt{\frac{1}{N-2}\left\{\sum(X_i - \overline{X})^2 + \sum(Y_j - \overline{Y})^2\right\}}} \equiv \tau(\underline{X}, \underline{Y}).$$

Thus the most powerful similar test of H_c versus K_1 is

$$\phi(\underline{X}, \underline{Y}) = \begin{cases} 1 & \text{if } \tau(\underline{X}, \underline{Y}) > c_\alpha(\underline{z}) \\ 0 & \text{if } \tau(\underline{X}, \underline{Y}) < c_\alpha(\underline{z}) \end{cases}$$

where $c_\alpha(\underline{z})$ is chosen so that exactly $\alpha N!$ of the permutations $\underline{z}'$ lead to rejection (if this is possible; if not we can use a randomized test). But we know that τ takes on at most $\binom{N}{m}$ distinct values according to each of the $\binom{N}{m}$ assignments $\underline{z}_c$ of m of the z_i's to be X_i's. Thus

$$\text{(3)} \qquad \phi(\underline{z}_c) = \begin{cases} 1 & \text{if } \tau(\underline{z}_c) > c_\alpha(\underline{z}) \\ 0 & \text{if } \tau(\underline{z}_c) < c_\alpha(\underline{z}) \end{cases}$$

where $c_\alpha(\underline{z})$ is chosen so that exactly $\alpha\binom{N}{m}$ of the assignments $\underline{z}_c$ of m of the z_i's to be X_i's leads to rejection.

Since the test (3) does not depend on which $\theta_1 < \theta_2$ or σ^2 we started with, the test is actually a UMP similar test of H_c versus $K \equiv \cup_{\theta_1 < \theta_2, \sigma^2} K_1$; i.e. different normal distributions with $\theta_1 < \theta_2$, σ^2 unknown.

Table 6.1: $\binom{5}{2}$ Possible Values of τ, $N = 5$, $m = 2$

combination	47	56	68	72	86	$\overline{Y} - \overline{X}$	$\sum Y_j$	τ
1	Y	Y	Y	X	X	−22.0	171	−1.436
2	Y	Y	X	Y	X	−18.7	175	−1.219
3	Y	X	Y	Y	X	−8.7	187	−0.566
4	Y	Y	X	X	Y	−7.0	189	−0.457
5	X	Y	Y	Y	X	−1.2	196	−0.076
6	Y	X	Y	X	Y	3.0	201	0.196
7	Y	X	X	Y	Y	6.3	205	0.414
8	X	Y	Y	X	Y	10.5	210	0.686
9	X	Y	X	Y	Y	13.8	214	0.903
10	X	X	Y	Y	Y	23.8	226	1.556

Example 2.7 Suppose that $(X_1, X_2) = (56, 72)$, $(Y_1, Y_2, Y_3) = (68, 47, 86)$. Thus $\overline{X} = 64$, $\overline{Y} = 67$, $\overline{Y} - \overline{X} = 3$. Here $\underline{Z} = (47, 56, 68, 72, 86)$, and $\binom{5}{2} = 5!/(2!3!) = 10$. (Note that $5! = 120$.) Note that $\binom{20}{10} = 184,756$, and, by Stirling's formula ($m! \sim \sqrt{2\pi m}(m/e)^m$) that

$$\binom{2m}{m} \sim \frac{1}{\sqrt{\pi m}} 2^{2m} \qquad \text{as } \; m \to \infty,$$

so the exact permutation test is difficult computationally for all but small sample sizes. But *sampling* from the permutation distribution is always possible.

Remark 2.8 We will call the present test "reject if $\tau > c_\alpha(\underline{z})$" the *permutation t - test*; it is the UMP similar test of H_c versus K specified above. If we consider the smaller null hypothesis

$$H_G : \; X_1, \ldots, X_m, Y_1, \ldots, Y_n \; \text{ i.i.d. } N(\theta, \sigma^2) \qquad \text{with } \theta, \; \sigma^2 \text{ unknown},$$

then we recall that the *classical t -test* "reject if $\tau > t_{m+n-2,\alpha}$" is the UMPU test of H_G versus K.

The classical $t-$test has greater power than the permutation $t-$test for H_G; but it is not a similar test of H_c. If we could show that for a.e. $\underline{z}$ the numbers

$$c_\alpha(\underline{z}) \qquad \text{and} \qquad t_{m+n-2,\alpha}$$

where just about equal, then the classical $t-$test and the permutation $t-$test would be almost identical.

Theorem 2.4 If $F \in \mathcal{F}_c$ has $E_F|X|^2 < \infty$ and if $0 < \liminf(m/N) \leq \limsup(m/N) < 1$, then

$$c_\alpha(\underline{z}) \to z_\alpha$$

where $P(N(0,1) > z_\alpha) = \alpha$. Since we also know that $t_{m+n-2,\alpha} \to z_\alpha$, it follows that $c_\alpha(\underline{z}) - t_{m+n-2,\alpha} \to 0$.

Proof. Let an urn contain balls numbered $z_1, \ldots, z_N$. Let $Y_1, \ldots, Y_n$ denote the numbers on n balls drawn without replacement. let $\overline{z} = N^{-1} \sum_{i=1}^N z_i$, $\sigma_z^2 = N^{-1} \sum_1^N (z_i - \overline{z})^2$, $m = N - n$. Then

$$E\overline{Y} = \overline{z}, \qquad \text{and} \qquad \sigma_N^2 \equiv Var(\overline{Y}) = \left(1 - \frac{n-1}{N-1}\right) \frac{\sigma_z^2}{n}.$$

Moreover, by the Wald - Wolfowitz - Noether - Hájek finite sampling CLT

$$\frac{\overline{Y} - \overline{z}}{\sigma_N} \to_d N(0,1)$$

as long as the Noether condition

$$\text{(a)} \qquad \eta_N \equiv \frac{\max_{1 \le i \le N} |z_i - \overline{z}|^2}{\sum_{i=1}^N |z_i - \overline{z}|^2} \to 0$$

holds.

Now rewrite the permutation $t-$ statistic τ: note that

$$\begin{aligned} \overline{Y} - \overline{X} &= \overline{Y} - \frac{1}{m}\sum_1^m X_i - \frac{1}{m}\sum_1^n Y_i + \frac{n}{m}\overline{Y} \\ &= \frac{N}{m}(\overline{Y} - \overline{z}), \end{aligned}$$

and hence

$$\begin{aligned} \tau &= \frac{\sqrt{\frac{mn}{N}}(\overline{Y} - \overline{X})}{\sqrt{\frac{1}{N-2}\left\{\sum Z_i^2 - \frac{(\sum Z_i)^2}{N} - \frac{mn}{N}(\overline{Y} - \overline{X})^2\right\}}} \\ &= \frac{\sqrt{\frac{N}{N-1}} \frac{\overline{Y} - \overline{z}}{\sqrt{\frac{1}{n}(1 - \frac{n-1}{N-1})}}}{\sqrt{\frac{N}{N-2}\sigma_z^2 - \frac{1}{N-2}\frac{N}{N-1}\frac{(\overline{Y}-\overline{z})^2}{\frac{1}{n}(1-\frac{n-1}{N-1})}}} \\ &= \sqrt{\frac{N-2}{N-1}} \frac{(\overline{Y} - \overline{z})/\sigma_N}{\sqrt{1 - \frac{1}{N-1}\frac{(\overline{Y}-\overline{z})^2}{\sigma_N^2}}} \\ &\to_d 1 \cdot \frac{Z}{\sqrt{1 - 0 \cdot Z^2}} = Z \sim N(0,1) \end{aligned}$$

if

$$\frac{\overline{Y} - \overline{z}}{\sigma_N} \to_d Z \sim N(0,1)$$

in probability or, better yet, almost surely; i.e. if

$$P\left(\frac{\overline{Y} - \overline{z}}{\sigma_N} \le t \Big| \underline{Z} = \underline{z}\right) \to \Phi(t)$$

in probability or almost surely. But this holds under the present hypotheses in view of the finite - sampling CLT 2.5 which follows, if we can show that

$$\text{(b)} \qquad \eta_N \to_{a.s.} 0$$

where η_N is key quantity in the Noether condition (a). To accomplish this note that even under the alternative hypothesis $F \neq G$ and $E_F|X|^2 < \infty$, $E_G|Y|^2 < \infty$,

$$\begin{aligned} \frac{1}{N}\sum_{i=1}^{N}(Z_i - \overline{Z})^2 &= \frac{1}{N}\left\{\sum_{i=1}^{N} Z_i^2 - N\overline{Z}^2\right\} \\ &= \frac{(m-1)}{N}S_X^2 + \frac{(n-1)}{N}S_Y^2 \\ &\to_{a.s.} \lambda\sigma_X^2 + (1-\lambda)\sigma_Y^2 \ge \min\{\sigma_X^2, \sigma_Y^2\} > 0 \end{aligned}$$

for any subsequence $N \to \infty$ for which $\lambda_N \equiv m/N \to \lambda$, and hence the denominator of η_N (divided by N) has a positive limit inferior almost surely. To see that the numerator converges almost surely to zero, first recall that $\max_{1\le i\le n}|X_i|/n \to_{a.s.} 0$ if and only if $E_F|X_1| < \infty$. Hence $\max_{1\le i\le n}|X_i|^2/n \to_{a.s.} 0$ if and only if $E_F|X_1|^2 < \infty$. Thus we rewrite the numerator divided by N as

$$\begin{aligned} \frac{1}{N}\max_{i\le N}|Z_i - \overline{Z}|^2 &\le \frac{2}{N}\left\{\max_{i\le N}|Z_i|^2 + \overline{Z}^2\right\} \\ &\le \frac{2}{N}\left\{\max\{\max_{i\le m}|X_i|^2,\ \max_{j\le n}|Y_j|^2\} + \left(\frac{m}{N}\overline{X} + \frac{n}{N}\overline{Y}\right)^2\right\} \\ &\le 2\max\{\frac{1}{m}\max_{i\le m}|X_i|^2,\ \frac{1}{n}\max_{j\le n}|Y_j|^2\} + \frac{2}{N}\left(\frac{m}{N}\overline{X} + \frac{n}{N}\overline{Y}\right)^2 \\ &\to_{a.s.} 0 + 0. \end{aligned}$$

Hence (b) holds (even under the alternative if $E_F X^2 < \infty$ and $E_G Y^2 < \infty$). □

Theorem 2.5 (Wald - Wolfowitz - Noether - Hájek finite - sampling central limit theorem). If $0 < \liminf(m/N) \le \limsup(m/N) < 1$, then

$$\frac{\overline{Y} - \overline{z}}{\sigma_N} \to_d Z \sim N(0,1) \qquad \text{as} \quad N \to \infty$$

if and only if

$$\text{(4)} \qquad \eta_N \equiv \frac{\max_{1\le i\le N}|z_i - \overline{z}|^2}{\sum_{i=1}^{N}|z_i - \overline{z}|^2} \to 0 \qquad \text{as} \quad N \to \infty.$$

Moreover,

$$\sup_t \left|P\left(\frac{\overline{Y} - \overline{z}}{\sigma_N} \le t\right) - \Phi(t)\right| \le 5\left(\frac{N}{m \wedge n}\eta_N\right)^{1/4} \qquad \text{for all} \quad N \ge 1.$$

Proof. See Hájek, *Ann. Math. Statist.* **32**, 506 - 523. For still better rates under stronger conditions, see Bolthausen (1984). □

3 Invariance in Testing; Rank Methods

3.1 Notation and Basic Results

Let $(\mathcal{X}, \mathcal{A}, P_\theta)$ be a probability space for all $\theta \in \Theta$, and suppose $\theta \neq \theta'$ implies $P_\theta \neq P_{\theta'}$. We observe $X \sim P_\theta$.

Suppose that $g : \mathcal{X} \to \mathcal{X}$ is one-to-one, onto $\mathcal{X}$, and measurable, and suppose that the distribution of gX when $X \sim P_\theta$ is some $P_{\theta'} = P_{\bar{g}\theta}$; that is

(1) $$P_\theta(gX \in A) = P_{\bar{g}\theta}(X \in A) \qquad \text{for all} \quad A \in \mathcal{A},$$

or equivalently

$$P_\theta(g^{-1}A) = P_{\bar{g}\theta}(A) \qquad \text{for all} \quad A \in \mathcal{A};$$

or, equivalently,

$$P_\theta(A) = P_{\bar{g}\theta}(gA) \qquad \text{for all} \quad A \in \mathcal{A}.$$

Hence

(2) $$E_\theta h(g(X)) = E_{\bar{g}\theta} h(X).$$

Suppose that $\bar{g}\Theta = \Theta$.

Let G denote a *group* of such tranformations g. We want to test $H : \theta \in \Theta_H$ versus $K : \theta \in \Theta_K$.

Proposition 3.1 $\overline{G}$ is a group of one-to-one transformations of Θ onto Θ and is homomorphic to G.

Proof. Suppose that $\bar{g}\theta_1 = \bar{g}\theta_2$. Then $P_{\theta_1} = P_{\theta_2}$ by (1). Thus $\theta_1 = \theta_2$ by assumption. Thus $\bar{g} \in \overline{G}$ is one-to-one.

Closure, associativity, and identity are easy.

If $X \sim P_\theta$, then $g_1 X \sim P_{\bar{g}_1\theta}$, and $(g_2 \circ g_1)X = g_2 \circ (g_1 X) \sim P_{\bar{g}_2 \circ \bar{g}_1 \theta}$, while $(g_2 \circ g_1)X \sim P_{\overline{g_2 \circ g_1}}$, so $\overline{g_2 \circ g_1} = \bar{g}_2 \circ \bar{g}_1$. If $X \sim P_\theta$, then $g^{-1}X \sim P_{\overline{g^{-1}}\theta}$, so $g \circ g^{-1} X \sim P_{\bar{g} \circ \overline{g^{-1}}\theta}$, while $g \circ g^{-1}X = X \sim P_\theta$, so $\bar{g} \circ \overline{g^{-1}} = \bar{e}$; thus $\bar{g}^{-1} = \overline{g^{-1}}$, and $\overline{G}$ is a group. $\square$

Definition 3.1 A group of one-to-one transformations of $\mathcal{X}$ onto $\mathcal{X}$ is said to leave the testing problem H versus K *invariant* provided $\bar{g}\Theta = \Theta$ and $\bar{g}\Theta_H = \Theta_H$ for all $g \in G$.

3.2 Orbits and maximal invariants

Definition 3.2 $x_1 \sim x_2 \text{mod}(G)$ if $x_2 = g(x_1)$ for some $g \in G$.

Proposition 3.2 $\sim$ is an equivalence relation.

Proof. Reflexive: $x_1 \sim x_1$ since $x_1 = e(x_1)$.
Symmetric: $g(x_1) = x_2$ implies $g^{-1}(x_2) = x_1$.
Transitive: $x_1 \sim x_2$ and $x_2 \sim x_3$ implies $x_1 \sim x_3$ since $g_1(x_1) = x_2$ and $g_2(x_2) = x_3$ implies $(g_2 \circ g_1)(x_1) = x_3$. $\square$

Definition 3.3 The equivalence classes of $\sim$ are called the *orbits* of G. Thus orbit$(x) = \{g(x) : g \in G\}$. A function ϕ defined on the sample space $\mathcal{X}$ is *invariant* if $\phi(g(x)) = \phi(x)$ for all $x \in \mathcal{X}$ and all $g \in G$.

Proposition 3.3 A test function ϕ is invariant if and only if ϕ is constant on each orbit of G.

Proof. This follows immediately from the definitions. $\square$

Definition 3.4 A measurable function $T : \mathcal{X} \to \mathbb{R}^k$ for some k is a *maximal invariant for G* (or GMI), if T is invariant and $T(x_1) = T(x_2)$ implies $x_1 \sim x_2$. That is, T is constant on the orbits of G and takes on distinct values on distinct orbits.

Theorem 3.1 Let T be a GMI. Then ϕ is invariant if and only if there exists a function h such that $\phi(x) = h(T(x))$ for all $x \in \mathcal{X}$.

Proof. Suppose that $\phi(x) = h(T(x))$. Then

$$\phi(gx) = h(T(gx)) = h(T(x)) = \phi(x),$$

so ϕ is invariant.

On the other hand, suppose that ϕ is $G-$invariant. Then $T(x_1) = T(x_2)$ implies $x_1 \sim x_2$ implies $g(x_1) = x_2$ for some $g \in G$. Thus $\phi(x_2) = \phi(gx_1) = \phi(x_1)$; that is, ϕ is constant on the orbit. It follows that ϕ is a function of T. $\square$

3.3 Examples

Example 3.1 (Translation group). Suppose that $\mathcal{X} = \mathbb{R}^n$ and $G = \{g : \ g\underline{x} = \underline{x} + c\underline{1}, \ c \in \mathbb{R}\}$. Then $T(\underline{x}) = (x_1 - x_n, \ldots, x_{n-1} - x_n)$ is a GMI.
Proof: Clearly T is invariant. Suppose that $T(\underline{x}) = T(\underline{x}^*)$. Then $x_i = x_i^* - (x_n^* - x_n)$ for $i = 1, \ldots, n-1$, and this holds trivially for $i = n$. Thus $\underline{x}^* = g(\underline{x}) = \underline{x} + c\underline{1}$ with $c = (x_n^* - x_n)$.

Example 3.2 (Scale group). Suppose that $\mathcal{X} = \{x \in \mathbb{R}^n : \ x_n \neq 0\}$, and $G = \{g : \ g\underline{x} = c\underline{x}, \ c \in \mathbb{R} \setminus \{0\}\}$. Then $T(\underline{x}) = (x_1/x_n, \ldots, x_{n-1}/x_n)$ is a GMI.
Proof: Clearly T is invariant. Suppose that $T(\underline{x}) = T(\underline{x}^*)$. Then $x_i^* = (x_n^*/x_n)x_i$ for $i = 1, \ldots, n-1$, this holds trivially for $i = n$. Thus $\underline{x}^* = g(\underline{x}) = c\underline{x}$ with $c = (x_n^*/x_n)$.

Example 3.3 (Orthogonal group). Suppose that $\mathcal{X} = \mathbb{R}^n$ and $G = \{g : \ g\underline{x} = \Gamma x, \ \Gamma$ an $n \times n$ orthogonal matrix$\}$. Then $T(\underline{x}) = \underline{x}^T\underline{x} = \sum_{i=1}^n x_i^2$ is a GMI.
Proof: $T(g\underline{x}) = \underline{x}^T\Gamma^T\Gamma\underline{x} = \underline{x}^T\underline{x}$, so T is invariant. Suppose that $T(\underline{x}) = T(\underline{x}^*)$. Then there exists $\Gamma = \Gamma_{x,x^*}$ such that $\underline{x}^* = \Gamma\underline{x}$.

Example 3.4 (Permutation group). Suppose that $\mathcal{X} = \mathbb{R}^n \setminus \{\text{ties}\}$, and $G = \{g : \ g(\underline{x}) = \pi\underline{x} = (x_{\pi(1)}, \ldots, x_{\pi(n)})$ for some permutation $\pi = (\pi(1), \ldots, \pi(n))$ of $\{1, \ldots, n\}$. Note that $\#(G) = n!$. Then $T(\underline{x}) = (x_{(1)}, \ldots, x_{(n)}) \equiv \underline{x}_{(\cdot)}$, the vector of ordered x's is a GMI.
Proof. $T(g\underline{x}) = T(\pi\underline{x}) = \underline{x}_{(\cdot)} = T(\underline{x})$, so T is invariant. Moreover, if $T(\underline{x}^*) = T(\underline{x})$, then $\underline{x}^* = \pi\underline{x}$ for some $\pi \in \Pi$, so T is maximal.

Example 3.5 (Rank transformation group). Suppose that $\mathcal{X} = \{\underline{x} \in \mathbb{R}^n : \ x_i \neq x_j \ \text{ for all } i \neq j\} = \mathbb{R}^n \setminus \{\text{ties}\}$, and $G = \{g : \ g(\underline{x}) = (f(x_1), \ldots, f(x_n)), \ f \ \text{ continuous and strictly increasing}\}$. Then $T(\underline{x}) \equiv \underline{r} = (r_1, \ldots, r_n)$ where $r_i \equiv \#\{j \leq n : \ x_j \leq x_i\}$ denotes the rank of x_i (among $x_1, \ldots, x_n$).
Proof: T is clearly invariant. If $T(\underline{x}^*) = T(\underline{x})$, then, relabeling if necessary, we have a picture as follows:

Example 3.6 (Sign group). Suppose that $\mathcal{X} = \mathbb{R}^n$ and that $G = \{g, e\}^n$ where $g(x) = -x$ and $e(x) = x$. Then $T(\underline{x}) = (|x_1|, \ldots, |x_n|)$ is a GMI.

Example 3.7 (Affine group). Suppose that $\mathcal{X} = \{x \in \mathbb{R}^n : \ x_{n-1} \neq x_n\}$ and that $G = \{g : \ g(\underline{x}) = a\underline{x} + b\underline{1}$ with $a \neq 0, \ b \in \mathbb{R}\}$. Then

$$T(\underline{x}) = \left(\frac{x_1 - x_n}{x_{n-1} - x_n}, \ldots, \frac{x_{n-2} - x_n}{x_{n-1} - x_n} \right)$$

is a GMI. Note that

$$T(\underline{x}) = \left(\frac{x_1 - \bar{x}}{s}, \ldots, \frac{x_n - \bar{x}}{s} \right)$$

is also a GMI (on $\mathcal{X} \equiv \{x \in \mathbb{R}^n : \ s > 0\}$ where $s^2 \equiv n^{-1} \sum_{i=1}^n (x_i - \bar{x})^2$).

Remark 3.1 In the previous example $G = G_2 \oplus G_1 = \text{scale} \oplus \text{translation} = \{g_2 \circ g_1 : \ g_1 \in G_1, \ g_2 \in G_2\}$. Then $Y = (x_1 - x_n, x_2 - x_n, \ldots, x_{n-1} - x_n)$ is a G_1- MI. In the space of the G_1-MI we have $Z = (y_1/y_{n-1}, \ldots, y_{n-2}/y_{n-1})$ is a G_2-MI. Thus Z is the GMI. If this works, it is OK; see theorem 6.2.2 on page 217 of TSH. But it doesn't always work. When $G = G_2 \oplus G_1$, it does work if G_1 is a normal subgroup of G. [Recall that G_1 is a normal subgroup of G if and only if $gG_1g^{-1} = G_1$ for all $g \in G$.]

Example 3.8 (Signed rank transformation group). Suppose that $\mathcal{X} = \mathbb{R}^N \setminus \{\text{ties}\}$ as in example 3.5 (but with N instead of n), but now let and

$$G = \{g : \ g(\underline{x}) = (f(x_1), \ldots, f(x_N)), \ f \text{ is odd, continuous, and strictly increasing}\}.$$

Then $T(\underline{x}) = (\underline{r}, \underline{s}) = (r_1, \ldots, r_m, s_1, \ldots, s_n)$ where $r_1, \ldots, r_m$ denote the ranks of $|x_{i_1}|, \ldots, |x_{i_m}|$ among $|x_1|, \ldots, |x_N|$ and $s_1, \ldots, s_n$ denote the rank of $|x_{j_1}|, \ldots, |x_{j_n}|$ among $|x_1|, \ldots, |x_N|$ and where $x_{i_1}, \ldots, x_{i_m} < 0 < x_{j_1}, \ldots, x_{j_n}$.
Proof: T is clearly invariant. To show maximal invariance, the picture is much as in example 3.5, but with the function f being odd; see Lehmann and Romano TSH pages 241-242.

Example 3.9 Suppose that $\mathcal{X} = \{(x_1, x_2) : \ x_2 > 0\}$ and that $G = \{g : \ g(x) = (x_1+b, x_2), \ b \in \mathbb{R}\}$. Then $T(x) = x_2$ is a GMI.

Example 3.10 Suppose that $\mathcal{X} = \{(x_1, x_2) : \ x_2 > 0\}$ as in example 3.9, but now suppose that the group $G = \{g : \ g(x) = (cx_1, cx_2), \ c > 0\}$ or $G = \{g : \ g(x) = (cx_1, |c|x_2), \ |c| \neq 0\}$. Then $T(x) = x_1/x_2$ is a GMI in the first case ($c > 0$), and $T(x) = |x_1|/x_2$ is a GMI in the second case ($c \neq 0$).

Example 3.11 Suppose that $\mathcal{X} = \{(x_1, x_2, x_3, x_4) : \ x_3, x_4 > 0\}$ and that $G = \{g : \ g(\underline{x}) = (cx_1 + a, cx_2 + b, cx_3, cx_4), \ a, b \in \mathbb{R}, \ c > 0\}$. Then $T(x) = x_3/x_4$ is a GMI.

3.4 UMP G-invariant tests

Theorem 3.2 If $T(X)$ is any $G-$invariant function and if $\nu(\theta)$ is a $\overline{G}MI$, then the distribution of $T(X)$ depends on θ only through $\nu(\theta)$.

Proof. Suppose that $\nu(\theta_1) = \nu(\theta_2)$. Then there exists $\bar{g} \in \overline{G}$ such that $\bar{g}\theta_1 = \theta_2$. Let g be the element of G corresponding to $\bar{g} \in \overline{G}$. Then by (1)

$$P_{\theta_1}(T(X) \in A) = P_{\theta_1}(T(gX) \in A) = P_{\bar{g}\theta_1}(T(X) \in A) = P_{\theta_2}(T(X) \in A)$$

for all $A \in \mathcal{A}$. Thus the distribution of T is a function of $\nu(\theta)$. $\square$

Theorem 3.3 Suppose that H versus K is invariant under G. Let $T(X)$ and $\delta \equiv \nu(\theta)$ denote the GMI and the $\overline{G}MI$; and suppose both are real-valued. Suppose that the densities $p_\delta(t) = (dP_\delta^T/d\mu)(t)$ with respect to some $\sigma-$finite measure μ have MLR in T; and suppose that H versus K is equivalent to $H_1 : \delta \leq \delta_0$ versus $K_1 : \ \delta > \delta_0$. Then there exists a UMP $G-$ invariant level α test of H versus K given by

$$\psi(T) = \begin{cases} 1 & \text{if } \ T > c \\ \gamma & \text{if } \ T = c \\ 0 & \text{if } \ T < c \end{cases}$$

with $E_{\delta_0}\psi(T) = \alpha$.

Proof. By theorem 3.1 any $G-$invariant test ϕ is of the form $\phi = \psi(T)$. By theorem 3.2, the distribution of T depends only on δ. Thus our theorem for UMP tests when there is MLR completes the proof. $\square$

Example 3.12 Tests of σ^2 for $N(\mu, \sigma^2)$. Let $X_1, \ldots, X_n$ be i.i.d. $N(\mu, \sigma^2)$. Consider testing $H : \ \sigma \leq \sigma_0$ versus $K : \ \sigma > \sigma_0$. Then $G = \{g : \ g(\underline{x}) = \underline{x} + c\underline{1}, \ c \in \mathbb{R}\}$ leaves H versus K invariant. By sufficiency we can restrict attention to tests based on $\overline{X}$ and $S \equiv \sum_1^n (X_i - \overline{X})^2$. let G^* denote the induced group $g^*(\overline{X}, S) = (\overline{X} + c, S)$. Thus S is a G^*MI by example 3.9. Now $S \sim \sigma^2 \chi^2_{n-1}$, which has MLR in S. Thus by theorem 3.3, the UMP G^*-invariant test of H versus K rejects H if $S > \sigma_0^2 \chi^2_{n-1,\alpha}$. By theorem 6.5.3, Lehmann and Romano, TSH (2005), page 229, it is also the UMP G-invariant test; also see Ferguson, page 157. (Recall from chapter 2 that this optimal normal theory test has undesirable robustness of level problems when the data fail to be normally distributed.)

Example 3.13 Two-sample t - test. Let $X_1, \ldots, X_m$ be i.i.d. $N(\mu, \sigma^2)$ and $Y_1, \ldots, Y_n$ be i.i.d. $N(\nu, \sigma^2)$, and consider testing $H: \ \nu \le \mu$ versus $K: \nu > \mu$. By sufficiency we can restrict attention to tests based on $(\overline{X}, \overline{Y}, S)$ with $S = \sum(X_i - \overline{X})^2 + \sum(Y_i - \overline{Y})^2$. Then the group $G = \{g: \ g(\underline{x}) = a\underline{x} + b\underline{1}, \ a > 0, \ b \in \mathbb{R}\}$ leaves H versus K invariant and if G^* denotes the induced group

$$g^*(\overline{X}, \overline{Y}, S) = (a\overline{X} + b, a\overline{Y} + b, a^2 S),$$

then $T(\overline{X}, \overline{Y}, S) = (\overline{Y} - \overline{X})/\sqrt{S}$ is a G^*-MI. Note that

$$t \equiv \frac{\sqrt{\frac{mn}{N}}(\overline{Y} - \overline{X})}{\sqrt{\frac{S}{N-2}}} = \sqrt{\frac{mn}{N}(N-2)}T \sim t_{m+n-2}(\delta)$$

with $\delta \equiv \sqrt{mn/N}(\nu - \mu)/\sigma$, and that H versus K is equivalent to $H' : \delta \le 0$ versus $K' : \delta > 0$. Since the non-central $t-$distributions have MLR, the UMP G^*-invariant test of H versus K is the two-sample $t-$test, "reject H if $t > t_{m+n-2,\alpha}$".

Example 3.14 (Sampling inspection by variables). Let $Y, Y_1, \ldots, Y_n$ be i.i.d. $N(\mu, \sigma^2)$. Let $p \equiv P(Y \le y_0) \equiv P(\text{good})$ for some fixed number y_0. Consider testing $H: p \ge p_0$ versus $K: \ p < p_0$. Now

$$\begin{aligned} p &= P(Y \le y_0) = P\left(\frac{Y - y_0 - (\mu - y_0)}{\sigma} \le \frac{y_0 - \mu}{\sigma}\right) \\ &= P\left(\frac{X - \theta}{\sigma} \le -\frac{\theta}{\sigma}\right) \\ &\qquad \text{where} \quad X_i \equiv Y_i - y_0 \sim N(\theta \equiv \mu - y_0, \sigma^2) \\ &= \Phi(-\theta/\sigma) = 1 - \Phi(\theta/\sigma), \end{aligned}$$

or $\theta/\sigma = \Phi^{-1}(1-p)$. Thus, on the basis of $X_1, \ldots, X_n$ we wish to test $H: \theta/\sigma \le c_0 \equiv \Phi^{-1}(1-p_0)$ versus $K: \theta/\sigma > c_0$. Now $\overline{X}$, $S = \sqrt{S^2}$ are sufficient. Also, H versus K is invariant under the group of example 3.10 with $c > 0$; and a GMI in the space of the sufficient statistic is $T = \sqrt{n}\overline{X}/S$. Now $T \sim t_{n-1}(\delta)$ where $\delta \equiv \sqrt{n}\theta/\sigma$, and the family of distributions has MLR in T. Also H versus K is equivalent to $H' : \delta \le \delta_0 \equiv \sqrt{n}\Phi^{-1}(1-p_0)$ versus $K' : \delta > \delta_0$. Thus the UMP $G-$invariant level α test of H versus K rejects H if $T > t_{n-1,\alpha}(\delta_0)$. [Note the use of the *non-central* t-distribution as a null distribution here!]

Example 3.15 (ANOVA - General Linear Model). The canonical form of ANOVA is as follows:

$\Theta:$ $\quad Z \sim N_n(\eta, \sigma^2 I), \qquad \eta \in V_k \subset \mathbb{R}^n$
where V_k is a subspace of $\mathbb{R}^n$ with dimension $k < n$,
and $\eta_i = 0, \ i = k+1, \ldots, n,$

$\Theta_0:$ $\quad Z \sim N_n(\eta, \sigma^2 I), \qquad \eta \in V_{k-r} \subset \mathbb{R}^n$
where V_{k-r} is a subspace of V_k with dimension $k - r < k$, and
and $\eta_i = 0, \ i = 1, \ldots, r, k+1, \ldots, n.$

We let

$$
\begin{aligned}
G_1 &\equiv \{g_1 :\ g_1 z = (z_1, \ldots, z_r, z_{r+1} + \Delta_{r+1}, \ldots, z_k + \Delta_k, z_{k+1}, \ldots, z_n),\ \text{with } \Delta_i \in \mathbb{R}\},\\
G_2 &= \left\{ \begin{array}{l} g_2 :\ g_2 z = (z_1^*, \ldots, z_r^*, z_{r+1}, \ldots, z_k, z_{k+1}, \ldots, z_n), \\ (z_1^*, \ldots, z_r^*) \text{ an orthogonal transformation of } (z_1, \ldots, z_r) \end{array} \right\},\\
G_3 &= \left\{ \begin{array}{l} g_3 :\ g_3 z = (z_1, \ldots, z_r, z_{r+1}, \ldots, z_k, z_{k+1}^*, \ldots, z_n^*), \\ (z_{k+1}^*, \ldots, z_n^*) \text{ an orthogonal transformation of } (z_{k+1}, \ldots, z_n) \end{array} \right\},\\
G_4 &= \{g_4 :\ g_4 z = cz, \quad \text{where } c \neq 0\};
\end{aligned}
$$

and, finally

$$G \equiv G_4 \oplus G_3 \oplus G_2 \oplus G_1 \equiv \{g_4 \circ g_3 \circ g_2 \circ g_1 :\ g_i \in G_i,\ i = 1, \ldots, 4\}.$$

Then H versus K is invariant under G.

Now $T_1(z) = (z_1, \ldots, z_r, z_{k+1}, \ldots, z_n)$ is a G_1MI.
In the space of the G_1MI, a G_2MI is $T_2(\underline{z}) = (\sum_{i=1}^r z_i^2, z_{k+1}, \ldots, z_n)$.
In the space of the $G_2 \oplus G_1$MI, a G_3MI is $T_3(\underline{z}) = (\sum_{i=1}^r z_i^2, \sum_{i=k+1}^n z_i^2)$.
In the space of the $G_3 \oplus G_2 \oplus G_1$MI, a G_4MI is $T(\underline{z}) = ((n-k)/r) \left(\sum_1^r z_i^2 / \sum_{k+1}^n z_i^2\right)$.

Now $T(z)$ is a GMI; thus any $G-$invariant test function for H versus K is a function of $T(z)$ by theorem 3.1. Similarly,

$$
\begin{aligned}
&(\sigma^2, \eta_1, \ldots, \eta_r) \ \text{ is a } \ \overline{G}_1\text{MI};\\
&(\sigma^2, \sum_{i=1}^r \eta_i^2) \ \text{ is a } \ \overline{G}_2 \oplus \overline{G}_1\text{MI}; \qquad \text{and a } \ \overline{G}_3 \oplus \overline{G}_2 \oplus \overline{G}_1\text{MI};
\end{aligned}
$$

and

$$\delta^2 \equiv \lambda^2 = \frac{\sum_{i=1}^r \eta_i^2}{\sigma^2} \qquad \text{is a } \ GMI.$$

Thus the distribution of any invariant test depends only on δ^2.

Now $T \sim F_{r,n-k}(\delta^2)$, which has MLR in T. Also, H versus K is equivalent to $H' :\ \delta = 0$ versus $K' : \delta > 0$. Thus the UMP $G-$invariant test of H versus K rejects H when $T > F_{r,n-k,\alpha}$.

Reduction to canonical form

The above analysis has been developed for the linear model in canonical form. Now the question is: how do we reduce a model stated in a more usual way to the canonical form? Suppose that

$$\underline{X} \sim N_n(\underline{\xi}, \sigma^2 I)$$

where $\underline{\xi} \equiv E\underline{X} = A\underline{\theta} \in \mathbf{L}$, where A is a (known) $n \times k$ matrix of rank k, $\underline{\theta}$ is a $k \times 1$ vector of (unknown) parameters, and $\mathbf{L}$ is the $k-$dimensional subspace of $\mathbb{R}^n$ spanned by the columns of the matrix A. Let B be a given $r \times k$ matrix, and consider testing

$$H :\ B\underline{\theta} = \underline{0} \qquad \text{or} \qquad \xi \in \mathbf{L}_1$$

where $\mathbf{L}_1$ is a $(k-r)-$dimensional subspace of $\mathbb{R}^n$.

To transform this form of the testing problem to canonical form, let T be an $n \times n$ orthogonal matrix with:

(i) the last $n-k$ rows of T are orthogonal to $\mathbf{L}$; i.e. orthogonal to the columns of A.

(ii) the rows $r+1,\ldots,k$ of T span $\mathbf{L}_1$.

Then set $\underline{Z} = T\underline{X}$. We compute

$$\underline{\eta} \equiv E\underline{Z} = TA\underline{\theta} = T\underline{\xi},$$

and note that:

(a) $\eta_{k+1} = \cdots = \eta_n = 0$ always by (i).

(b) $\eta_1 = \cdots = \eta_r = 0$ under H by (ii) since the first r rows of T are orthogonal to $\mathbf{L}_1$.

Now we will re-express the $F-$statistic we have derived in terms of the X's:

$$\begin{aligned} S^2(\underline{\eta}) &\equiv \sum_{i=1}^{n}(Z_i-\eta_i)^2 = \sum_{i=1}^{k}(Z_i-\eta_i)^2 + \sum_{i=k+1}^{n} Z_i^2 \\ &\geq \sum_{i=k+1}^{n} Z_i^2 \end{aligned}$$

by taking $\eta_i = \hat{\eta}_i = Z_i$, $i=1,\ldots,k$. But since T is orthogonal, $\underline{\eta} = T\underline{\xi}$, and $\underline{Z} = T\underline{X}$,

$$(3) \qquad S^2(\underline{\eta}) = \|\underline{Z}-\underline{\eta}\|^2 = (\underline{Z}-\underline{\eta})^T(\underline{Z}-\underline{\eta})$$

$$(4) \qquad = (\underline{X}-\underline{\xi})^T(\underline{X}-\underline{\xi}) = \sum_{i=1}^{n}(X_i-\xi_i)^2$$

so that

$$(5) \qquad \min_{\underline{\xi}\in\mathbf{L}} \sum_{i=1}^{n}(X_i-\xi_i)^2 = \sum_{i=1}^{n}(X_i-\hat{\xi}_i)^2 = \sum_{i=k+1}^{n} Z_i^2$$

where $\hat{\underline{\xi}}$ is the Least Squares (LS) estimator of $\underline{\xi}$ under $\underline{\xi} = A\underline{\theta} \in \mathbf{L}$. Similarly, under $H:\ \underline{\xi}\in\mathbf{L}_1$ (or $\eta_1 = \cdots = \eta_r = 0$ in the canonical form),

$$\begin{aligned} S^2(\eta) &= \sum_{i=1}^{r} Z_i^2 + \sum_{i=r+1}^{k}(Z_i-\eta_i)^2 + \sum_{i=k+1}^{n} Z_i^2 \\ &\geq \sum_{i=1}^{r} Z_i^2 + \sum_{k+1}^{n} Z_i^2 \end{aligned}$$

by taking $\eta_i = \hat{\hat{\eta}}_i = Z_i$ for $i = r+1,\ldots,k$, and hence by (4)

$$(6) \qquad \min_{\underline{\xi}\in\mathbf{L}_1} \sum_{i=1}^{n}(X_i-\xi_i)^2 \equiv \sum_{i=1}^{n}(X_i-\hat{\hat{\xi}}_i)^2 = \sum_{i=1}^{r} Z_i^2 + \sum_{i=k+1}^{n} Z_i^2$$

where $\hat{\hat{\underline{\xi}}}$ is the least squares estimate of $\underline{\xi}$ under the hypothesis $\underline{\xi}\in\mathbf{L}_1$. Combining (5) and (6) yields

$$\sum_{i=1}^{r} Z_i^2 = \sum_{i=1}^{n}(X_i-\hat{\hat{\xi}}_i)^2 - \sum_{i=1}^{n}(X_i-\hat{\xi}_i)^2;$$

here $\mathbf{L}_1$ is a subspace of dimension $k-r$ contained in $\mathbf{L}$, which is a subspace of dimension k contained in $\mathbb{R}^n$. Now since $X-\hat{\xi}\perp\mathbf{L}$

(7) $\quad \underline{X}-\underline{\hat{\xi}} \ \perp\ \mathbf{L};\quad \text{in particular, } \underline{X}-\underline{\hat{\xi}}\perp\underline{\hat{\xi}}-\underline{\hat{\hat{\xi}}}\in\mathbf{L}.$

Hence

$$\|\underline{X}-\underline{\hat{\hat{\xi}}}\|^2=\|\underline{X}-\underline{\hat{\xi}}\|^2+\|\underline{\hat{\xi}}-\underline{\hat{\hat{\xi}}}\|^2$$

by (7), and we have

$$\sum_{i=1}^r Z_i^2=\|\underline{\hat{\xi}}-\underline{\hat{\hat{\xi}}}\|^2=\sum_{i=1}^n(\hat{\xi}_i-\hat{\hat{\xi}}_i)^2,$$

and the $F-$statistic which yields the UMP $G-$invariant test of $H:\ \underline{\xi}\in\mathbf{L}_1$ versus $K:\ \underline{\xi}\notin\mathbf{L}_1$ may be written as

$$F=\frac{\sum_{i=1}^n(\hat{\xi}_i-\hat{\hat{\xi}}_i)^2/r}{\sum_{i=1}^n(X_i-\hat{\xi}_i)^2/(n-k)}=\frac{\left\{\sum_{i=1}^n(X_i-\hat{\hat{\xi}}_i)^2-\sum_{i=1}^n(X_i-\hat{\xi}_i)^2\right\}/r}{\sum_{i=1}^n(X_i-\hat{\xi}_i)^2/(n-k)}.$$

To re-express the noncentrality parameter of the distribution of F under the alternative hypothesis in terms of $\underline{\xi}$ (instead of $\underline{\eta}$), let $\underline{\xi}\in\mathbf{L}$, and let $\underline{\xi}^0$ denote the projection of $\underline{\xi}$ onto $\mathbf{L}_1$: thus $\underline{\xi}=\underline{\xi}^0+(\underline{\xi}-\underline{\xi}^0)$ where $\underline{\xi}^0\in\mathbf{L}_1$ and $\underline{\xi}-\underline{\xi}^0\perp\mathbf{L}_1$. Then

$$\delta^2=\sum_{i=1}^r\eta_i^2/\sigma^2=\sum_{i=1}^n\{\hat{\xi}_i(\xi)-\hat{\hat{\xi}}_i(\xi)\}^2/\sigma^2=\sum_{i=1}^n\{\xi_i-\xi_i^0\}^2/\sigma^2.$$

Here are three standard examples:

Example 3.16 (One-way layout).
Suppose that X_{ij}, $j=1,\ldots,n_i$ and $i=1,\ldots,I$ are independent with $X_{ij}\sim N(\xi_{ij},\sigma^2)$ where we suppose that $\xi_{ij}=\mu_i$ for $j=1,\ldots,n_i$ for each $i\in\{1,\ldots,I\}$. Now $n=n_1+\cdots+n_I$ and $k=I$. A traditional null hypothesis is $\mu_1=\cdots=\mu_I$, and hence $r=I-1$. Thus we can write $\underline{X}\sim N_n(A\mu,\sigma^2 I)$ where the $n\times I$ matrix $A=(\underline{a}_i)$ and the column vectors $\underline{a}_i$ are given by $a_{i,j}=1\{n_1+\cdots+n_{j-1}<i\le n_1+\cdots+n_j\}$. The subspace $\mathbf{L}$ of $\mathbb{R}^n$ is spanned by the columns of A, and the subspace $\mathbf{L}_1$ is the subspace spanned by the $n-$vector $\underline{1}$. To find the F statistic giving the UMP G-invariant test in this case we consider the sum of squares

$$\begin{aligned}S^2 &= \sum_{i=1}^I\sum_{j=1}^{n_i}(X_{ij}-\xi_{ij})^2=\sum_{i=1}^I\sum_{j=1}^{n_i}(X_{ij}-\mu_i)^2\\ &= \sum_{i=1}^I\sum_{j=1}^{n_i}(X_{ij}-\overline{X}_{i\cdot})^2+\sum_{i=1}^I\sum_{j=1}^{n_i}(\overline{X}_{i,\cdot}-\mu_i)^2=\sum_{i=1}^I\sum_{j=1}^{n_i}(X_{ij}-\overline{X}_{i\cdot})^2+\sum_{i=1}^I n_i(\overline{X}_{i,\cdot}-\mu_i)^2.\end{aligned}$$

Thus $\widehat{\xi}_{ij}=\widehat{\mu}_i=\overline{X}_{i,\cdot}$ for $j=1,\ldots,n_i$ and $i=1,\ldots,I$. Furthermore,

$$\sum_{i=1}^I\sum_{j=1}^{n_i}(X_{ij}-\xi_{ij})^2=\sum_{i=1}^I\sum_{j=1}^{n_i}(X_{ij}-\overline{X}_{\cdot,\cdot})^2+n(\overline{X}_{\cdot,\cdot}-\mu)^2,$$

where $\overline{X}_{\cdot\cdot} = \sum_i \sum_j X_{ij}/n$. Thus the test of $H : \mu_1 = \cdots = \mu_I$ versus $K : \mu_i \neq \mu_{i'}$ for some $i \neq i'$ is: reject if

$$F \equiv \frac{\sum_{i=1}^I n_i(\overline{X}_{i,\cdot} - \overline{X}_{\cdot\cdot})^2/(I-1)}{\sum_{i=1}^I \sum_{j=1}^{n_i}(X_{ij} - \overline{X}_{i\cdot})^2/(n-I)} > F_{I-1,n-I,\alpha}$$

where $P(F_{I-1,n-I} > F_{I-1,n-I,\alpha}) = \alpha$. The non-centrality parameter is

$$\delta^2 = \frac{1}{\sigma^2}\sum_{i=1}^I n_i(\mu_i - \mu_\cdot)^2$$

where $\mu_\cdot = n^{-1}\sum_{i=1}^I n_i\mu_i$, and hence the power function of this test is given by

$$\beta_\phi(\delta) = P_\delta(F_{I-1,n-I}(\delta^2) > F_{I-1,n-I,\alpha}).$$

Example 3.17 (Two-way layout with one observation per cell). Suppose that $X_{i,j}$, $i = 1, \ldots, I$, $j = 1, \ldots, J$ are independent with $X_{i,j} \sim N(\xi_{i,j}, \sigma^2)$. Suppose, moreover, that $\xi_{i,j} = \xi + \mu_i + \nu_j$ with $\sum_1^I \mu_i = 0$ and $\sum_1^J \nu_j = 0$; the μ_i's are called the *row effects* and the ν_j's are called the *column effects*. Thus we have $I + J + 1$ parameters and 2 restrictions, and hence the $\xi_{i,j}$'s are in a $k = I + J - 1$ dimensional subspace $\mathbf{L}$ of $\mathbb{R}^n$ where $n = I \cdot J$. Consider testing the null hypothesis of "no column effect": that is,

$$H : \ \nu_1 = \nu_2 = \cdots = \nu_J = 0.$$

This imposes $r = J - 1$ further restrictions on $\nu_1, \ldots, \nu_J$, and hence the $\xi_{i,j}$'s are in a $k - r = I$ dimensional subspace $\mathbf{L}_1$ of $\mathbb{R}^{I \cdot J}$. To identify the $F-$ statistic in this case, we first consider the sum of squares for the larger $k-$dimensional model: it is easily shown that

$$\begin{aligned} S^2 &= \sum_{i,j}(X_{i,j} - \xi_{i,j})^2 = \sum_{i,j}(X_{i,j} - \xi - \mu_i - \nu_j)^2 \\ &= \sum_{i,j}(X_{i,j} - \overline{X}_{i,\cdot} - \overline{X}_{\cdot,j} + \overline{X}_{\cdot,\cdot})^2 + \sum_{i,j}(\overline{X}_{i,\cdot} - \overline{X}_{\cdot,\cdot} - \mu_i)^2 \\ &\quad + \sum_{i,j}(\overline{X}_{\cdot,j} - \overline{X}_{\cdot,\cdot} - \nu_j)^2 + \sum_{i,j}(\overline{X}_{\cdot,\cdot} - \xi)^2 \end{aligned} \tag{8}$$

where

$$\overline{X}_{i,\cdot} \equiv \frac{1}{J}\sum_{j=1}^J X_{i,j}, \qquad \overline{X}_{\cdot,j} \equiv \frac{1}{I}\sum_{i=1}^J X_{i,j}, \qquad \overline{X}_{\cdot,\cdot} \equiv \frac{1}{IJ}\sum_{i,j} X_{i,j}.$$

It follows that

$$\widehat{\xi} = \overline{X}_{\cdot,\cdot}, \quad \widehat{\mu}_i = \overline{X}_{i,\cdot} - \overline{X}_{\cdot,\cdot}, \quad \widehat{\nu}_j = \overline{X}_{\cdot,j} - \overline{X}_{\cdot,\cdot},$$

and also that

$$S^2_{resid} \equiv \sum_{i,j}(X_{i,j} - \widehat{\xi}_{i,j})^2 = \sum_{i,j}(X_{i,j} - \overline{X}_{i,\cdot} - \overline{X}_{\cdot,j} + \overline{X}_{\cdot,\cdot})^2.$$

For testing the null hypothesis of no column effect, it follows from (8) that $\widehat{\widehat{\xi}} = \overline{X}_{\cdot,\cdot}$, $\widehat{\widehat{\mu}}_i = \overline{X}_{i,\cdot} - \overline{X}_{\cdot,\cdot}$, and hence

$$\sum_{i,j}(\widehat{\xi}_{i,j} - \widehat{\widehat{\xi}}_{i,j})^2 = \sum_{i,j}(\overline{X}_{\cdot,j} - \overline{X}_{\cdot,\cdot})^2 = I\sum_{j=1}^J(\overline{X}_{\cdot,j} - \overline{X}_{\cdot,\cdot})^2.$$

Thus the F-statistic for testing $H : \nu_1 = \cdots = \nu_J = 0$ is given by

$$F = \frac{I \sum_{j=1}^J (\overline{X}_{\cdot,j} - \overline{X}_{\cdot,\cdot})^2}{S^2_{resid}}.$$

This has a non-central $F_{J-1,IJ-I-J+1}(\delta^2)$ distribution under the general linear hypothesis where the non-centrality parameter is given by

$$\delta^2 = \frac{I \sum_{j=1}^J \nu_j^2}{\sigma^2}.$$

Example 3.18 (Linear regression). Suppose that $X_1, \ldots, X_n$ are independent with $\xi_i = E(X_i) = \beta_0 + \beta_1 z_i$ where $z_1, \ldots, z_n$ are known numbers. We will assume that $\sum_1^n z_i = 0$ and $\sum_1^n z_i^2$. For this model the matrix $A = (\underline{1}, \underline{z})$ where $\underline{z} = (z_1, \ldots, z_n)^T$. Thus $\mathbf{L}$ is the $k = 2$ dimensional subspace of $\mathbb{R}^n$ spanned by $\underline{1}$ and $\underline{z}$. Now consider testing $H : \beta_1 = 0$ versus $K : \beta_1 \neq 0$. Thus the subspace $\mathbf{L}_1$ is the one-dimensional subspace of $\mathbb{R}^n$ spanned by $\underline{1}$. The least squares estimators of β_0 and β_1 are $\widehat{\beta}_0 = \overline{X}_n$ and $\widehat{\beta}_1 = \sum_{i=1}^n z_i X_i / \sum_{i=1}^n z_i^2$, and the least squares estimator $\widehat{\widehat{\beta}}_0$ under the null hypothesis is just $\overline{X}_n$. Thus the F statistic giving the UMP G-invariant test of H versus K is given by

$$F = \frac{\sum_{i=1}^n (\widehat{\beta}_1 z_i)^2}{\sum_{i=1}^n (X_i - \widehat{\beta}_0 - \widehat{\beta}_1 z_i)^2/(n-2)} = \frac{\widehat{\beta}_1^2 \sum_1^n z_i^2}{\sum_{i=1}^n (X_i - \widehat{\beta}_0 - \widehat{\beta}_1 z_i)^2/(n-2)}$$

which has an $F_{1,n-2}(\delta^2)$ distribution where

$$\delta^2 = \frac{\beta_1^2 \sum_1^n z_i^2}{\sigma^2}.$$

3.5 Rank tests

First we need to be able to compute probabilities for rank vectors. Our first job here is to develop a fundamental formula due to Hoeffding which allows us to do this.

Let $Z_1, \ldots, Z_N$ be independent real-valued random variables with densities $f_1, \ldots, f_N$ respectively. Let

$$R_i \equiv \text{rank of } \ Z_i \text{ in } Z_1, \ldots, Z_N = \#\{j \leq N : \ Z_j \leq Z_i\} = N\mathbb{F}_N(Z_i)$$

for $i = 1, \ldots, N$ where $\mathbb{F}_N$ is the empirical distribution of the Z_i's. Thus

$$P(\underline{R} = \underline{r}) = \int \cdots \int_S f_1(z_1) \cdots f_N(z_N) dz_1 \cdots dz_N$$

where

$$S \equiv \{\underline{z} : \ R_i(\underline{z}) = r_i, \ i = 1, \ldots, N\} = \{\underline{z} : \ z_{d_1} < \cdots < z_{d_N}\}$$

where $d = r^{-1}$, the inverse permutation, $r \circ d = r \circ r^{-1} = e$. (Example: $N = 3$; $z = (10, 5, 8)$. Then $r = (3, 1, 2)$ and $d = (2, 3, 1)$.) Hence, letting $z_{d_i} \equiv v_i$,

$$S = \{V_1 < \cdots < V_N\},$$

and

$$\begin{aligned} P(\underline{R}=\underline{r}) &= \int\cdots\int_{v_1<\cdots<v_N} \frac{f_1(v_{r_1})\cdots f_N(v_{r_N})}{N!h(v_{r_1})\cdots h(v_{r_N})} N!h(v_1)\cdots h(v_N)d\underline{v} \\ &= \frac{1}{N!} E\frac{f_1(V_{(r_1)})\cdots f_N(V_{(r_N)})}{h(V_{(r_1)})\cdots h(V_{(r_N)})} \end{aligned}$$

where $V_{(1)} < \cdots < V_{(N)}$ are the order statistics of a sample of size N from h. This formula is one version of *Hoeffding's formula*.

Of course, sometimes direct calculation succeeds immediately. Here are two simple, but important, examples:

Example 3.19 Suppose that $F_i = F^{\Delta_i}$ with $\Delta_i > 0$, $i = 1, \ldots, N$ and F continuous. Then

$$\begin{aligned} P(\underline{R}=\underline{e}) &= P(X_1 < \cdots < X_N) = \int\cdots\int_{x_1<\cdots<x_N} \prod_{i=1}^N dF^{\Delta_i}(x_i) \\ &= \int\cdots\int_{0\le u_1\le\cdots\le u_N\le 1} \prod_{i=1}^N \Delta_i u_i^{\Delta_i-1} du_i \\ &= \int\cdots\int_{0\le u_2\le\cdots\le u_N\le 1} \prod_{i=3}^N \Delta_i u_i^{\Delta_i-1} \Delta_2 u_2^{\Delta_1+\Delta_2-1} du_2\cdots du_N \\ = \cdots\ &= \prod_{i=1}^N \frac{\Delta_i}{\sum_{j=1}^i \Delta_j}. \end{aligned}$$

This yields any probability $P(\underline{R}=\underline{r})$, $\underline{r}\in\Pi$, by relabeling:

$$P(\underline{R}=\underline{r}) = \prod_{i=1}^N \frac{\Delta_{d_i}}{\sum_{j=1}^i \Delta_{d_j}}.$$

Example 3.20 (Proportional hazards alternative). Similarly, suppose that $(1-F_i) = (1-F)^{\Delta_i}$ with $\Delta_i > 0$, $i = 1, \ldots, N$ and F continuous; this is equivalent to $\Lambda_i \equiv -\log(1-F_i) = \Delta_i\{-\log(1-F)\} = \Delta_i\Lambda$, the proportional hazards model. Then

$$\begin{aligned} P(\underline{R}=\underline{e}) &= P(X_1 < \cdots < X_N) = \int\cdots\int_{x_1<\cdots<x_N} \prod_{i=1}^N d\{1-(1-F(x_i))^{\Delta_i}\} \\ &= \int\cdots\int_{0\le u_1\le\cdots\le u_N\le 1} \prod_{i=1}^N \Delta_i(1-u_i)^{\Delta_i-1} du_i \\ = \cdots\ &= \prod_{i=1}^N \frac{\Delta_i}{\sum_{j=i}^N \Delta_j}. \end{aligned}$$

This yields any probability $P(\underline{R}=\underline{r})$, $\underline{r}\in\Pi$, by relabeling:

$$P(\underline{R}=\underline{r}) = \prod_{i=1}^N \frac{\Delta_{d_i}}{\sum_{j=i}^N \Delta_{d_j}}.$$

Now suppose that $X_1, \ldots, X_m$ are i.i.d. F and $Y_1, \ldots, Y_n$ are i.i.d. G, $F, G \in \mathcal{F}_c$; and let $\mathcal{G}$ denote the group of all strictly increasing continuous transformations of the real line onto itself, example 3.5.

Proposition 3.4
A. The two-sample problem of testing $H: \ F = G$ versus $K: \ F <_s G$, $F, G \in \mathcal{F}_c$, is invariant under $\mathcal{G}$.
B. The rank vector $\underline{R}$ is a $G-$MI.
C. $\psi(u) = G \circ F^{-1}(u)$ is a $\overline{G}-$MI.
D. The ordered Y ranks $Q_1 < \cdots < Q_n$ are sufficient for $\underline{R}$; $Q_i \equiv N\mathbb{H}_N(\mathbb{G}_n^{-1}(i/n)), \ i = 1, \ldots, n$.
E. Hoeffding's formula: suppose that F and G have densities f and g respectively, and that $f(x) = 0$ implies $g(x) = 0$. Then

$$P(\underline{Q} = \underline{q}) = \frac{1}{\binom{N}{n}} E_f \left\{ \prod_{j=1}^{n} \frac{g(V_{(q_j)})}{f(V_{(q_j)})} \right\}$$

where $V_{(1)} < \cdots < V_{(N)}$ are the order statistics of a sample of size N from F. Furthermore, this probability may be rewritten as

$$P(\underline{Q} = \underline{q}) = \frac{1}{\binom{N}{n}} E \left\{ \prod_{j=1}^{n} \psi'(U_{(q_j)}) \right\}$$

where $U_{(1)} < \cdots < U_{(N)}$ are the order statistics of a sample of N Uniform$(0,1)$ random variables.

Proof. Statements A - C follow easily from the preceding development. To prove E, we specialize Hoeffding's formula by taking $f_i = f$ for $i = 1, \ldots, m$, $f_i = g$ for $i = m+1, \ldots, N$, and $h = f$. Then

$$P(\underline{R} = \underline{r}) = \frac{1}{N!} E_F \left(\prod_{j=1}^{n} \frac{g}{f}(V_{(r_{m+j})}) \right) = \frac{1}{N!} E_F \left(\prod_{j=1}^{n} \frac{g}{f}(V_{(q_j)}) \right).$$

Hence

$$\begin{aligned} P(\underline{Q} = \underline{q}) &= \sum_{r: \ q(r) = q} P(\underline{R} = \underline{r}) = \frac{1}{N!} E_F \left(\prod_{j=1}^{n} \frac{g}{f}(V_{(q_j)}) \right) \sum_{r: \ q(r) = q} 1 \\ &= \frac{m!n!}{N!} E_F \left(\prod_{j=1}^{n} \frac{g}{f}(V_{(q_j)}) \right). \end{aligned}$$

Note that the claimed sufficiency of $\underline{Q}$ for $\underline{R}$ in D follows from these computations.

To see the second formula, note that $\psi'(u) = (g/f)(F^{-1}(u))$, and that

$$(F^{-1}(U_{(1)}, \ldots, F^{-1}(U_{(N)})) \stackrel{d}{=} (V_{(1)}, \ldots, V_{(N)}).$$

□

Here are several applications of Hoeffding's formula: we use the preceding results to find locally most powerful rank tests in several different two-sample testing problems: location, scale, and Lehmann alternatives of the proportional hazards type.

Proposition 3.5 (Locally most powerful rank test for location). Suppose that F has an absolutely continuous density f for which $\int |f'(x)|dx < \infty$. Then the locally most powerful rank test of $H : F = G$ versus $K : G = F(\cdot - \theta)$ with $\theta > 0$ is of the form

$$\phi(\underline{q}) = \begin{cases} 1 & \text{if } S_N \equiv \sum_{j=1}^n E_F\left(-\frac{f'}{f}(V_{(q_j)})\right) > k_\alpha \\ \gamma & \text{if } S_N = k_\alpha \\ 0 & \text{if } S_N < k_\alpha \end{cases}$$

where $V_{(1)} < \cdots < V_{(N)}$ are the order statistics in a sample of size N from F.

Proof. For a rank test, $\phi = \phi(\underline{Q})$, we want to maximize the slope of the power function at $\theta = 0$: i.e. to maximize the slope at $\theta = 0$ of

$$\beta_\phi(\theta) = E_\theta \phi(\underline{Q}) = \sum_{\underline{q}} \phi(\underline{q}) P_\theta(\underline{Q} = \underline{q}).$$

To do this we clearly want to find those $\underline{q}$ for which

$$\frac{d}{d\theta} P_\theta(\underline{Q} = \underline{q})\Big|_{\theta=0}$$

is large. But, by using proposition 3.4 and differentiation under the expectation (which can be justified by the assumption that $\int |f'(x)|dx < \infty$),

$$\begin{aligned} \frac{d}{d\theta} P_\theta(\underline{Q} = \underline{q})\Big|_{\theta=0} &= \frac{d}{d\theta} \frac{1}{\binom{N}{n}} E_F \left\{ \prod_{i=1}^n \frac{f(V_{(q_j)} - \theta)}{f(V_{(q_j)})} \right\} \Big|_{\theta=0} \\ &= \frac{1}{\binom{N}{n}} E_F \left\{ \frac{d}{d\theta} \prod_{i=1}^n \frac{f(V_{(q_j)} - \theta)}{f(V_{(q_j)})} \Big|_{\theta=0} \right\} \\ &= \frac{1}{\binom{N}{n}} E_F \left\{ - \sum_{j=1}^n \frac{f'}{f}(V_{(q_j)}) \right\} \end{aligned}$$

since

$$\begin{aligned} \frac{d}{d\theta} \prod_{j=1}^n \frac{f(x_j - \theta)}{f(x_j)} &= \sum_{k=1}^n \prod_{j=1, j\neq k}^n \frac{f(x_j - \theta)}{f(x_j)}\Big|_{\theta=0} \left\{ - \frac{f'(x_k - \theta)}{f(x_k)}\Big|_{\theta=0} \right\} \\ &= - \sum_{k=1}^n \frac{f'}{f}(x_k). \end{aligned}$$

□

Example 3.21 If F is $N(\mu, \sigma^2)$, then without loss (by the monotone transformation $g(X) = (X-\mu)/\sigma$, we may take $F = \Phi$, the standard $N(0, 1)$ distribution function. Then $-(f'/f)(x) = x$, so $E\{(-f'/f)(V_{(i)})\} = E(Z_{(i)})$ where $Z_{(1)} < \cdots < Z_{(N)}$ are the order statistics of N standard normal ($N(0, 1)$) random variables, and $S_N = \sum_{j=1}^n E(Z_{(q_j)})$. Note that $E(Z_{(i)})$ may be approximated by $\Phi^{-1}(i/(N + 1))$, or by $\Phi^{-1}((3i - 1)/(3N + 1))$.

Example 3.22 If F is logistic, $f(x) = e^{-x}/(1+e^{-x})^2$, then $f = F(1-F)$, and $-f'/f = 2F - 1$. Since $F(V_{(i)}) \overset{d}{=} U_{(i)}$ where $U_1, \ldots, U_N$ are uniform$(0,1)$ random variables with $EU_{(i)} = i/(N+1)$, the LMPRT of F versus $G = F(\cdot - \theta)$ rejects H for large values of $S_N = \sum_{j=1}^n Q_j$; this is the *Wilcoxon* statistic.

Proposition 3.6 (Locally most powerful rank test for scale). Suppose that F has an absolutely continuous density f for which $\int |xf'(x)|dx < \infty$. Then the locally most powerful rank test of $H : F = G$ versus $K : G = F(\cdot/\theta)$ with $\theta > 1$ is of the form

$$\phi(\underline{q}) = \begin{cases} 1 & \text{if } \; S_N \equiv \sum_{j=1}^n a_N(q_j) > k_\alpha \\ \gamma & \text{if } \; S_N = k_\alpha \\ 0 & \text{if } \; S_N < k_\alpha \end{cases}$$

where

$$a_N(i) \equiv E_F\{-1 - V_{(i)}\frac{f'}{f}(V_{(i)})\}$$

and $V_{(1)} < \cdots < V_{(N)}$ are the order statistics in a sample of size N from F.

Example 3.23 If $f(x) = e^{-x}1_{[0,\infty)}(x)$, then $(f'/f)(x) = -1$, and hence $a_N(i) = E_F\{-1 + V_{(i)}\} = E_F\{V_{(i)} - 1\}$ where $V_{(i)}$ are the order statistics of a sample of size N from F. But

$$V_{(i)} \overset{d}{=} \sum_{j=1}^{i} \frac{Z_j}{N-j+1}$$

where Z_j are i.i.d. exponential(1), and hence

$$E(V_{(i)}) \;=\; \sum_{j=1}^{i} \frac{1}{N-j+1} = \sum_{k=N-i+1}^{N} \frac{1}{k} = E\{-\log(1-U_{(i)})\}$$

since $F^{-1}(t) = -\log(1-t)$. These are the Savage scores for testing exponential scale change; the *approximate* scores are

$$a_N(i) = -\log\left(1 - \frac{i}{N+1}\right) - 1\ , \qquad i = 1, \ldots, N,$$

and the resulting test is sometimes called the "log-rank" test. Its modern derivation in survival analysis is via different considerations which allow for the introduction of censoring, and rewritten in a martingale framework. [Recall that $\sum_{k=1}^N k^{-1} - \log N \to \gamma = .5772\cdots$, Euler's constant, as $N \to \infty$, so

$$\begin{aligned} \sum_{k=N-i+1}^{N} \frac{1}{k} &= \sum_{k=1}^{N} \frac{1}{k} - \sum_{k=1}^{N-i} \frac{1}{k} \\ &= \sum_{k=1}^{N} \frac{1}{k} - \log N - \left(\sum_{k=1}^{N-i} \frac{1}{k} - \log(N-i)\right) - \log(1 - \frac{i}{N}) \\ &\doteq -\log(1 - \frac{i}{N}) \end{aligned}$$

for large N.]

Note that when F is exponential(1), then

$$(1 - G(x)) = 1 - F(x/\theta) = \exp(-x/\theta) = (1 - F(x))^{1/\theta},$$

or, $\Lambda_G = (1/\theta)\Lambda_F = \Delta\Lambda_F$ with $\Delta = 1/\theta$. Hence

$$\psi(u) = 1 - (1-u)^{1/\theta} = 1 - (1-u)^{\Delta},$$

and $\psi'(u) = \Delta(1-u)^{\Delta-1}$. Since the distribution of the ranks is the same for all (F, G) pairs with the same ψ, it follows that in fact the Savage test is the locally most powerful rank test of $H: \ F = G$ versus the Lehmann alternative $K: \ (1-G) = (1-F)^{\Delta}$, $\Delta < 1$.

Example 3.24 If F is $N(0, \sigma^2)$, the LMPRT of $F = G$ versus $K : G = F(\cdot/\theta)$, $\theta > 1$, rejects for large values of $S_N \equiv \sum_{j=1}^{n} a_N(Q_j)$ where $a_N(i) \equiv E(Z_{(i)}^2)$ and $Z_{(1)} < \cdots < Z_{(N)}$ is an ordered sample from $N(0,1)$. The approximate scores are $(\Phi^{-1}(i/(N+1))^2$.

Remark 3.2 Note that any rank statistic of the form S_N can be rewritten in terms of empirical distributions as follows:

$$S_N = \sum_{j=1}^{n} a_N(Q_j) = \sum_{j=1}^{n} a_N(R_{m+j}) = \sum_{i=1}^{N} a_N(i) Z_{Ni}$$

where $Z_{Ni} = 0$ or 1 according as the ith largest of the combined sample is an X or Y. Let $\mathbb{H}_N(x)$ = empirical df of the combined sample. Then $\mathbb{H}_N^{-1}(i/N) = i$th largest of the combined sample, $n\mathbb{G}_n(\mathbb{H}_N^{-1}(i/N))$ = the number of Y_i's $\le \mathbb{H}_N^{-1}(i/N)$, and $Z_{Ni} = \Delta\{n\mathbb{G}_n(\mathbb{H}_N^{-1})\}(i/N)$ where $\Delta h(y) \equiv h(y) - h(y-)$. Therefore we can write

$$\begin{aligned} S_N &= \sum_{i=1}^{N} a_N(i) Z_{Ni} = \sum_{i=1}^{n} a_N(i) \Delta\{n\mathbb{G}_n(\mathbb{H}_N^{-1})\}(i/N) \\ &= n \int_0^1 \phi_N(u) d\mathbb{G}_n(\mathbb{H}_N^{-1}(u)) \end{aligned}$$

where $\phi_N(u) \equiv \sum_{i=1}^{N} a_N(i) 1\{(i-1)/N < u \le i/N\}$ for $0 < u < 1$. If $\phi_N \to \phi$ and $\lambda_N \to \lambda$, then it is often true that under alternatives $F \ne G$,

$$\frac{1}{n} S_N = \int_0^1 \phi_N(u) d\mathbb{G}_n \circ \mathbb{H}_N^{-1}(u) \to_{a.s.} \int_0^1 \phi(u) dG \circ H^{-1}(u)$$

where $H = \lambda F + (1-\lambda) G$.

4 Efficiency of Tests

4.1 The Power of two tests

Example 4.1 (Power of the one-sample $t-$test:) Let $X_1, \ldots, X_n$ be i.i.d. (θ, σ^2). We wish to test $H : \theta \le \theta_0$ versus $K : \theta > \theta_0$. The classical test of H versus K rejects H when $t_n \equiv \sqrt{n}(\overline{X} - \theta_0)/S > t_{n-1,\alpha}$.
(i) This test has asymptotically correct level of significance (assuming $E(X^2) < \infty$ as we have by hypothesis) since, with $Z \sim N(0,1)$,

$$P_{\theta_0}(t_n > t_{n-1,\alpha}) \to P(Z > z_\alpha) = \alpha.$$

(ii) This test is consistent since, when a fixed $\theta > \theta_0$ is true

$$\begin{aligned} t_n &= \frac{\sqrt{n}(\overline{X} - \theta)}{S} + \frac{\sqrt{n}(\theta - \theta_0)}{S} \\ &\to_d Z + \infty = \infty \end{aligned}$$

and $t_{n-1,\alpha} \to z_\alpha$ so that $P_\theta(t_n > t_{n-1,\alpha}) \to 1$.
(iii) If $X_1, \ldots, X_n$ are i.i.d $(\theta_n, \sigma^2 \equiv (\theta_0 + n^{-1/2} c_n, \sigma^2)$ where $c_n \to c$, then

$$\begin{aligned} t_n &= \frac{\sqrt{n}(\overline{X} - \theta_n)}{S} + \frac{c_n}{S} \\ &\to_d Z + \frac{c}{\sigma} \sim N(c/\sigma, 1). \end{aligned}$$

Let $\beta_n^t(\theta)$ denote the power of the $t-$test based on $X_1, \ldots, X_n$ against the alternative θ. Then

$$\begin{aligned} &(1) \qquad \beta_n^t(\theta_n) = \beta_n^t(\theta_0 + n^{-1/2} c_n) \\ &(2) \qquad\qquad\;\; = P_{\theta_0 + c_n/\sqrt{n}}(t_n > t_{n-1,\alpha}) \to P(N(c/\sigma, 1) > z_\alpha). \end{aligned}$$

Example 4.2 Let $X_1, \ldots, X_n$ be i.i.d. with d.f. $F = F_0(\cdot - \theta)$ where F_0 has unique median 0 (so that $F_0(0) = 1/2$). We wish to test $H : \theta \le \theta_0$ versus $K : \ \theta > \theta_0$. Let $Y_i \equiv 1\{X_i \ge \theta_0\} = 1_{[\theta_0, \infty)}(X_i)$ for $i = 1, \ldots, n$. The *sign test* of H versus K rejects H when $S_n \equiv \sqrt{n}(\overline{Y}_n - 1/2)$ exceeds the upper α percentage $s_{n,\alpha}$ of its distribution when θ_0 is true.
(i) When θ_0 is true, Y_i is Bernoulli(1/2) so that $S_n \to_d Z/2 \sim N(0, 1/4)$. Since the exact distribution of $n\overline{Y}_n = \sum_1^n Y_i$ is Binomial$(n, 1/2)$ for all d.f.'s F as above, the test has exact level of significance α for all such F.
(ii) This test is consistent, since when a θ exceeding θ_0 is true

$$\begin{aligned} S_n &= \sqrt{n}(\overline{Y} - P_\theta(X \ge \theta_0)) + \sqrt{n}\{P_\theta(X > \theta_0) - 1/2)\} \\ &\to_d N(0, p(1-p)) + \infty = \infty \end{aligned}$$

with $p \equiv 1 - F(\theta_0 - \theta) > 1/2$ so that $P_\theta(S_n > s_{n,\alpha}) \to 1$.
(iii) If $X_1, \ldots, X_n$ are i.i.d. $F_0(\cdot - (\theta_0 + n^{-1/2} d_n))$ where $d_n \to d$ as $n \to \infty$ and where we now assume that F_0 has a strictly positive derivative f_0 at 0. Then, using $F_0(0) = 1/2$, we have

$$\begin{aligned} S_n &= \sqrt{n}(\overline{Y} - P_{\theta_0 + d_n/\sqrt{n}}(X \ge \theta_0)) + \sqrt{n}\{P_{\theta_0 + d_n/\sqrt{n}}(X > \theta_0) - 1/2)\} \\ &= n^{-1/2}\left\{\text{Binomial}(n, 1 - F_0(-d_n/\sqrt{n})) - n(1 - F_0(-d_n/\sqrt{n}))\right\} \\ &\qquad + \sqrt{n}(F_0(0) - F_0(-d_n/\sqrt{n})) \\ &\to_d Z/2 + df_0(0) \sim N(df_0(0), 1/4). \end{aligned}$$

Thus the power of the sign test $\beta_n^2(\theta)$ satisfies

$$\begin{aligned} \beta_n^s(\theta_0 + n^{-1/2}d_n) &\to P(Z/2 + df_0(0) > z_\alpha/2) = P(Z > z_\alpha - 2df_0(0) \\ &= P(N(2df_0(0), 1) > z_\alpha). \end{aligned} \tag{3}$$

4.2 Pitman Efficiency

Definition 4.1 *Pitman efficiency* is defined to be the limiting ratio of the sample sizes that produce equal asymptotic power against the same sequence of alternatives.

Now equal asymptotic power β in (2) and (3) requires that

$$\frac{c}{\sigma} = 2df_0(0). \tag{4}$$

If the $t-$test is based on N_t observations and the sign test is based on N_s observations, then equal alternatives in example 4.1 and example 4.2 requires that

$$c_{N_t}/\sqrt{N_t} = d_{N_s}/\sqrt{N_s}. \tag{5}$$

Thus the Pitman efficiency $e_{s,t}$ of the sign test with respect to the t test is just the limiting value of N_t/N_s subject to (4) and (5). Thus

$$\frac{N_t}{N_s} = \frac{c_{N_t}^2}{d_{N_s}^2} \to \frac{c^2}{d^2} = 4\sigma^2 f_0^2(0) = e_{s,t}.$$

Exercise 4.1 Evaluate $e_{s,t} = 4\sigma^2 f_0^2(0)$ in case:
(i) f_0 is Uniform$(-a, a)$;
(ii) f_0 is Normal$(0, a^2)$;
(iii) f_0 is Logistic$(0, a)$: (i.e. $f_0(x) = a^{-1}e^{-x/a}/[1 + \exp(-x/a)]^2$.
(iv) f_0 is t with k degrees of freedom;
(v) f_0 is double - exponential(a); $f_0(x) = (2a)^{-1}\exp(-a|x|)$.

A General calculation

We now consider the problem more generally. Suppose that $X_1, \dots, X_N$ have a joint distribution P_θ where θ is a real-valued parameter. We wish to test $H : \theta \le \theta_0$ versus $K : \theta > \theta_0$. Suppose that the T_1 test and the T_2 test are both consistent tests of H versus K; and that the T_i test rejects H if the statistic $T_{N,i}$ exceeds the upper α percent point of its distribution when $\theta = \theta_0$. Since both tests are consistent, it is useless to compare their limiting power under fixed alternatives; hence we will compare their power on a sequence of alternatives that approach θ_0 from above at the rate $1/\sqrt{N}$.

Suppose that for each $c > 0$ the statistics $T_{N,i}$ satisfy

$$P_{\theta_0 + c_N/\sqrt{N}}(T_{N,i} \le x) \to P(N(c\mu_i, \sigma_i^x) = P(N(c\mu_i/\sigma_i, 1) \le x)$$

for all x as $N \to \infty$ for any sequence of c_N's converging to c. Let the T_1- test (the T_2-test) use N_1 (use N_2) observations against the sequence of alternatives $c_{N_1}/\sqrt{N_1}$ (the sequence of alternatives $c_{N_2}/\sqrt{N_2}$) where $c_{N_1} \to c_1$ (where $c_{N_2} \to c_2$). Equal asymptotic power requires

$$\frac{c_1\mu_1}{\sigma_1} = \frac{c_2\mu_2}{\sigma_2},$$

and equal alternatives requires

$$\frac{c_{N_1}}{\sqrt{N_1}} = \frac{c_{N_2}}{\sqrt{N_2}};$$

solving these simultaneously leads to

$$\text{(6)} \qquad \frac{N_2}{N_1} = \frac{c_{N_2}^2}{c_{N_1}^2} \to \frac{(\mu_1/\sigma_1)^2}{(\mu_2/\sigma_2)^2} = e_{1,2}.$$

Note that the efficiency $e_{1,2}$ is independent of the common level of significance α of the tests, of the particular value of the asymptotic power β, and of the particular sequences that converge to the values of c_1 and c_2 that are specified by the choice of β. Since so much is summarized in a single number, the procedure is bound to have some shortcomings; however it can be extremely useful and informative.

The quantity $\epsilon_i \equiv (\mu_i/\sigma_i)^2$ is called the *efficacy* of the T_i-test, and hence the efficiency $e_{1,2}$ is the ratio of the efficacies.

Exercise 4.2 Define your idea of what the exact small sample efficiency $e_{s,t}(\alpha, \beta, n)$ of the sign test with respect to the $t-$test should be. Compute some values of it in case $X_1, \ldots, X_n$ are normal, and compare these values with the asymptotic value $e_{s,t} = 2/\pi \doteq .6366\ldots$ that was obtained in exercise 4.1.

Exercise 4.3 Now redefine Pitman efficiency to be the ratio of the squared distances from the alternative to the hypothesized valued θ_0 that produce equal asymptotic power as equal sample sized approach infinity. Show that you get the same answer as before.

Note that if T_1 and T_2 are estimating the same thing (that is, if $\mu_1 = \mu_2$), then $e_{1,2}$ is just the ratio of the limiting variances.

Also note that the typical test of $H : \theta \le \theta_0$ versus $K : \theta > \theta_0$ is of the form: reject H if

$$\frac{\sqrt{n}(T - E_{\theta_0}(T))}{\sqrt{Var_{\theta_0}(\sqrt{n}(T))}} > k_{n,\alpha} \to z_\alpha.$$

Thus when $\theta_0 + c/\sqrt{n}$ is true, intuitively we have (letting $m(\theta) \equiv E_\theta(T)$, and $\sigma_0^2 \equiv Var_{\theta_0}(\sqrt{n}T)$),

$$\begin{aligned}
\frac{\sqrt{n}(T - E_{\theta_0}(T))}{\sqrt{Var_{\theta_0}(\sqrt{n}(T))}} \quad &= \quad \frac{\sqrt{Var_{\theta_0+c/\sqrt{n}}(\sqrt{n}(T))}}{\sqrt{Var_{\theta_0}(\sqrt{n}(T))}} \frac{\sqrt{n}(T - E_{\theta_0+c/\sqrt{n}}(T))}{\sqrt{Var_{\theta_0+c/\sqrt{n}}(\sqrt{n}(T))}} \\
&\quad + \frac{\sqrt{n}(m(\theta_0 + c/\sqrt{n}) - m(\theta_0))}{\sqrt{Var_{\theta_0}(\sqrt{n}(T))}} \\
&\to_d \quad 1 \cdot Z + \frac{cm'(\theta_0)}{\sigma_0} \sim N(cm'(\theta_0)/\sigma_0, 1).
\end{aligned}$$

Thus we expect $(m'(\theta_0)/\sigma_0)^2$ to be the efficacy.

Exercise 4.4 Now consider testing $H : \theta = \theta_0$ versus $K : \theta \neq \theta_0$ on the basis of a two-sided test based on either the T_1 or the T_2 statistics consider previously. Show that the same formula for Pitman efficiency is appropriate for the two-side test also.

Exercise 4.5 Again consider testing $H : \theta = \theta_0$ versus $K : \theta \neq \theta_0$; but suppose now that

$$T_{n,i} \to_d \chi_k^2(c^2\delta_i^2) \qquad \text{as} \quad n \to \infty$$

under any sequence of alternatives $\theta_0 + c_n/\sqrt{n}$ having $c_n \to c > 0$ as $n \to \infty$. Here k is a fixed interger, and the limiting random variable has a noncentral chi-square distribution. Show that the Pitman efficiency criterion leads to $e_{1,2} = \delta_1^2/\delta_2^2$.

4.2 Some two-sample tests

Example 4.3 (The two-sample $t-$test). Let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ be independent samples from the distribution functions F and $G = F(\cdot - \theta)$ respectively. The classical test of $H: \ \theta \le 0$ versus $K: \ \theta > 0$ rejects H if

$$t_{m,n} \equiv \frac{\sqrt{\frac{mn}{N}}(\overline{Y} - \overline{X})}{\sqrt{\frac{m-1}{N-2}S_X^2 + \frac{n-1}{N-2}S_Y^2}} > t_{m+n-2,\alpha}.$$

(As noted in section 6.2,this test has certain optimality properties when F is a normal distribution.) If F is any d.f. having finite variance, then:

(i) When $\theta = 0$ we have $t_{m,n} \to_d N(0,1)$ provided $m \wedge n \to \infty$.

(ii) When $\theta > 0$ is true, then the test is consistent as $m \wedge n \to \infty$.

(iii) If $\lambda_N \equiv m/N \to \lambda \in (0,1)$ as $m \wedge n \to \infty$, then

$$P_{\theta=c/\sqrt{N}}(t_{m,n} > t_{m+n-2,\alpha}) \to P(c\sqrt{\lambda(1-\lambda)}/\sigma, 1) > z_\alpha).$$

Thus the efficacy of the two-sample $t-$test is

$$(7) \qquad \epsilon_t = \lambda(1-\lambda)/\sigma^2.$$

Example 4.4 (The Mann-Whitney and Wilcoxon tests). Let $X_1, \ldots, X_m$ be i.i.d. F and let $Y_1, \ldots, Y_n$ be i.i.d. G where F and G are continuous d.f.'s, and consider testing $H: \ F = G$ versus $K: \ F <_s G$ (i.e. $G(x) \le F(x)$ for all x and $G(x) < F(x)$ for some x).

The *Wilcoxon test* is "reject H if $W_{m,n} \equiv \sum_{j=1}^n Q_j = \sum_{j=1}^n R_{m+j}$ is too big". Tables of the exact null distribution of $W_{m,n}$ for small m, n are available, so the level is exactly α. Moreover, if H is true,

$$\frac{W_{m,n} - E_H(W_{m,n})}{\sqrt{Var_H(W_{m,n})}} = \frac{W_{m,n} - n(N+1)/2}{\sqrt{mn(N+1)/12}} \to_d N(0,1)$$

provided $m \wedge n \to \infty$; this follows from the Wald-Wolfowitz-Noether-Hájek permutational CLT since

$$\frac{N}{m \wedge n}\eta_N \equiv \frac{N}{m \wedge n}\frac{\max_i |a_i - \overline{a}|^2}{\sum_1^N (a_i - \overline{a})^2} = \frac{1}{m \wedge n}\frac{(N-1)^2/4}{(N^2-1)/12} \to 0$$

provided $m \wedge n \to \infty$.

The *Mann-Whitney test* is described as follows: let

$$U_{m,n} \equiv \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} 1\{X_i \le Y_j\}.$$

Mann and Whitney proposed to reject H if $U_{m,n}$ is "too big". Since

$$mnU_{m,n} + n(n+1)/2 = W_{m,n},$$

when H is true we have

$$\frac{U_{m,n} - 1/2}{\sqrt{(N+1)/12mn}} \to_d N(0,1) \qquad \text{as } m \wedge n \to \infty.$$

For arbitrary F and G

$$EU_{m,n} = E1\{X \le Y\} = P(X \le Y) = \int F dG,$$

while, for arbitrary continuous F and G

$$\begin{aligned} Var(\sqrt{mn}U_{m,n}) &= (n-1)\int (1-G)^2 dF \\ &\quad + (m-1)\int F^2 dG - (N-1)\left(\int F dG\right)^2 + \int F dG \\ &= (n-1)Var(1-G(X)) + (m-1)Var(F(Y)) + \int F dG(1 - \int F dG). \end{aligned}$$

We now consider the local alternatives $Y \stackrel{d}{=} X + c/\sqrt{N}$, or $G = F(\cdot - c/\sqrt{N})$. We also suppose that $\lambda_N = m/N \to \lambda$. Then

$$\begin{aligned} \frac{U_{m,n} - 1/2}{\sqrt{(N+1)/12mn}} &= \frac{U_{m,n} - \int F dG}{\sqrt{(N+1)/12mn}} + \frac{\int F dG - 1/2}{\sqrt{(N+1)/12mn}} \\ &\equiv Z_{m,n} + a_{m,n} \end{aligned}$$

where it seems intuitive that $Z_{m,n} \to_d Z \sim N(0,1)$ as $N \to \infty$ and

$$\begin{aligned} a_{m,n} &= \sqrt{\frac{12mn}{N(N+1)}}\sqrt{N}\left\{\int F dG - \frac{1}{2}\right\} \\ &= \sqrt{\frac{12mn}{N(N+1)}}\sqrt{N}\left\{\int F(x) dF(x - c/\sqrt{N}) - \int F dF\right\} \\ &= \sqrt{\frac{12mn}{N(N+1)}}\int \sqrt{N}(F(x - c/\sqrt{N}) - F(x)) dF(x) \\ &\to \sqrt{12\lambda(1-\lambda)}c\int f^2(x)dx \end{aligned}$$

assuming that F has density f with $\int f^2(x)dx < \infty$. Thus under $(F, G) = (F, F(\cdot - c/\sqrt{N}))$

$$\frac{U_{m,n} - 1/2}{\sqrt{(N+1)/12mn}} \to_d Z + c\sqrt{12\lambda(1-\lambda)}\int f^2 \sim N(c\sqrt{12\lambda(1-\lambda)}\int f^2, 1).$$

Thus the efficacy of the $U-$test is

$$\epsilon_U = 12\lambda(1-\lambda)\left(\int f^2(x)dx\right)^2. \tag{8}$$

Combining the efficacies in (7) and (8) for the $t-$test and the $U-$test respectively gives the Pitman efficiency of the $U-$test with respect to the $t-$test:

$$e_{U,t}(F) = \frac{12\lambda(1-\lambda)\{\int f^2\}^2}{\lambda(1-\lambda)/\sigma^2} = 12\sigma^2\left(\int f^2\right)^2.$$

Proposition 4.1 $e_{U,t}(F) \geq 108/125 = .864\ldots$.

Proof. first note that $e_{U,t}(F) = e_{U,t}(F(\cdot - a)/c)$. Thus it suffices to minimize $\int f^2(x)dx$ subject to the restrictions

$$\int x^2 f(x)dx = 1, \qquad \int xf(x)dx = 0, \qquad f(x) \geq 0, \qquad \int f(x)dx = 1.$$

Consider minimizing

$$B(f) \equiv \int_{-\infty}^{\infty}\{f^2(x) + f(x)2b(x^2-a^2)\}dx$$

with $b > 0$ subject to $f \geq 0$ and $\int f(x)dx = 0$. Now

$$\begin{aligned} f^2(x) + 2bf(x)(x^2-a^2) &= f(x)\{f(x) + 2b(x^2-a^2)\} \\ &= A\{A + 2b(x^2-a^2) \geq 0 \qquad \text{for } |x| \geq a. \end{aligned}$$

Thus take $f(x) = 0$ for $|x| \geq a$ and minimize the integrand pointwise for $|x| \leq a$. This yields $A \equiv f(x) = b(a^2-x^2)$. Thus the minimizer $f_{a,b}(x) \equiv f = b(a^2-x^2)1_{[-a,a]}(x)$. Choosing a and b so that $\int x^2 f(x)dx = 1$ and $\int f(x) = 1$ yields $a = \sqrt{5}$, $b = 3\sqrt{5}/100$, and hence $\int f^2(x)dx = 3\sqrt{5}/25$. Hence

$$e_{U,t} \geq 12\{\int f_{a,b}^2(x)dx\}^2 = 12(9\cdot 5)/625 = 108/125.$$

□

Proof of asymptotic normality of $U_{m,n}$ under local alternatives Suppose that

$$\begin{aligned} &X_{N,1},\ldots,X_{N,m} \qquad \text{are i.i.d. } F_N \\ &Y_{N,1}\ldots,Y_{N,n} \qquad \text{are i.i.d. } G_N \end{aligned}$$

where F_N, G_N, and H are continuous df's satisfying $\|F_N - H\|_\infty \to 0$ and $\|G_N - H\|_\infty \to 0$ as $N \to \infty$. Let $\mathbb{F}_m$ and $\mathbb{G}_n$ denote the empirical df's of $X_{N,1}\ldots,X_{N,m}$ and $Y_{N,1},\ldots,Y_{N,m}$ respectively. Now

$$U_{m,n} = \int \mathbb{F}_m d\mathbb{G}_n.$$

Consider

$$\begin{aligned}
\sqrt{\frac{mn}{N}}\left(U_{m,n}-\frac{1}{2}\right) &= \sqrt{\frac{mn}{N}}\left(\int \mathbb{F}_m d\mathbb{G}_n - \int F_N dG_N + \int F_N dG_N - \frac{1}{2}\right) \\
&= \sqrt{\frac{mn}{N}}\left\{\int (\mathbb{F}_m - F_N) d\mathbb{G}_n + \int F_N d(\mathbb{G}_n - G_N)\right\} \\
&\quad + \sqrt{\frac{mn}{N}}\left(\int F_N dG_N - \int G_N dG_N\right) \\
&\stackrel{d}{=} \sqrt{\frac{n}{N}}\int \mathbb{U}_m(F_N) d\mathbb{G}_n - \sqrt{\frac{m}{N}}\int \mathbb{V}_n(G_N) dF_N + \left(\frac{N+1}{12N}\right)^{1/2} a_{m,n};
\end{aligned}$$

here $\mathbb{U}_m$ is the empirical process of m i.i.d. Uniform$(0,1)$ rv's and $\mathbb{V}_n$ is the empirical process of n i.i.d. Uniform$(0,1)$ rv's independent of the random variables used to define $\mathbb{U}_m$. Thus for special constructions of $\mathbb{U}_m$ and $\mathbb{V}_n$ and independent Brownian bridge processes $\mathbb{U}$ and $\mathbb{V}$,

$$\begin{aligned}
\sqrt{\frac{mn}{N}}\left(U_{m,n}-\frac{1}{2}\right) &\to_d \sqrt{1-\lambda}\int \mathbb{U}(H)dH - \sqrt{\lambda}\int \mathbb{V}(H)dH + \frac{1}{\sqrt{12}}a \\
&= \int_0^1 \{\sqrt{1-\lambda}\,\mathbb{U}(t) - \sqrt{\lambda}\,\mathbb{V}(t)\}dt + \frac{a}{\sqrt{12}} \\
&\stackrel{d}{=} \int_0^1 \mathbb{U}(t)dt + \frac{a}{\sqrt{12}} \stackrel{d}{=} \frac{1}{\sqrt{12}}(Z+a) \sim \frac{1}{\sqrt{12}}N(a,1)
\end{aligned}$$

since

$$\sigma^2 = \int_0^1\int_0^1 (s\wedge t - st)dsdt = \frac{1}{12}.$$

Convergence of the first term above is justified by:

$$\begin{aligned}
&\left|\int \mathbb{U}_m(F_N)d\mathbb{G}_n - \int \mathbb{U}(H)dH\right| \\
&\quad \le \left|\int (\mathbb{U}_m(F_N) - \mathbb{U}(F_N))d\mathbb{G}_n\right| + \left|\int (\mathbb{U}(F_N) - \mathbb{U}(H))d\mathbb{G}_n\right| \\
&\qquad + \left|\int \mathbb{U}(H)d(\mathbb{G}_n - H)\right| \\
&\le \|\mathbb{U}_m - \mathbb{U}\|\int d\mathbb{G}_n + \|\mathbb{U}(F_N) - \mathbb{U}(H)\|\int d\mathbb{G}_n \\
&\qquad + \left|\int \mathbb{U}(H)d(\mathbb{G}_n - H)\right| \\
&\to_{a.s.} 0+0+0 = 0
\end{aligned}$$

where the convergence of the first term follows by the special (Skorokhod) construction of $\{\mathbb{U}_m, \mathbb{U}\}$ and $\int d\mathbb{G}_n = 1$; the convergence of the second term follows from $\|F_n - H\| \to 0$ and uniform continuity of $\mathbb{U}$ for a.e. fixed ω; and converence of the third term follows from Helly-Bray since $\mathbb{U}(H)$ is a bounded continuous function a.s. and $\mathbb{G}_n \to_d H$ almost surely. To see this last claim, note that

$$\begin{aligned}
\|\mathbb{G}_n - H\| &= \|\mathbb{G}_n - G_N + G_N - H\| \\
&\le n^{-1/2}\|\mathbb{V}_n(G_N)\| + \|G_N - H\| \\
&\le n^{-1/2}\|\mathbb{V}_n(G_N) - \mathbb{V}(G_N)\| + n^{-1/2}\|\mathbb{V}\| + \|G_N - H\| \\
&\to_{a.s.} 0+0+0 = 0.
\end{aligned}$$

Exercise 4.6 Evaluate $e_{U,t}(F) = 12\sigma^2 (\int f^2(x)dx)^2$ in case:

(i) f is Uniform$[-a, a]$.

(ii) f is Normal.

(iii) f is Logistic.

(iv) f is t_k.

(v) f is double-exponential.

Exercise 4.7 (General behavior of the centering constants for $U_{m,n}$). Suppose that

$$\|\sqrt{N}(f_N^{1/2} - h^{1/2}) - \frac{1}{2}\alpha h^{1/2}\|_2 \to 0, \qquad \text{and} \qquad \|\sqrt{N}(g_N^{1/2} - h^{1/2}) - \frac{1}{2}\beta h^{1/2}\|_2 \to 0.$$

Then

$$\|\sqrt{N}(F_N - H) - \int_{-\infty}^{\cdot} \alpha dH\|_\infty \to 0, \qquad \text{and} \qquad \|\sqrt{N}(G_N - H) - \int_{-\infty}^{\cdot} \beta dH\|_\infty \to 0.$$

Show that this implies (using $\int \alpha dH = 0 = \int \beta dH$) that

$$a_{m,n} \to \sqrt{12\lambda(1-\lambda)} \int (1-H)(\alpha - \beta) dH.$$

Check that the result for shift alternatives $H \equiv F$ and $G = F(\cdot - c/\sqrt{N})$ follows with $\alpha = 0$ and $\beta = -f'/f$.

4.4 Pitman efficiency via Le Cam's third lemma.

Often the limiting power and efficacy of a test can be easily derived via Le Cam's third lemma, lemma 3.3.14. Recall tht the essence of that lemma is that the joint limiting distribution of a statistic and the local log-likelihood ratio under the null hypothesis determines the joint limiting distribution of the statistic and the local log-likelhood ratio under the sequence of local alternatives. Here we simply illustrate this approach with the examples considered in section 6.4.1.

Example 4.5 The one-sample $t-$test again. Let $X_1, \ldots, X_n$ be i.i.d. $F = F_0(\cdot - \theta)$ with $\theta = E_F(X)$. Consider testing $H : \theta \le \theta_0$ versus $K : \theta > \theta_0$ using $t_n \equiv \sqrt{n}(\overline{X} - \theta_0)/S$. Suppose that F_0 has an absolutely continuous density f_0 and that $I_0 \equiv \int (f_0'/f_0)^2 f_0 dx < \infty$. Let $L_n \equiv \prod_{i=1}^n (f_n/f)(X_i)$ where $f_n(x) = f_0(x - \theta_n)$, $\theta_n = \theta_0 + c/\sqrt{n}$, and $f(x) = f_0(x - \theta_0)$. Thus with $\dot{\mathbf{l}}(x) \equiv -(f'/f)(x)$, under $P_n = P_f^n$,

$$\log L_n = \frac{c}{\sqrt{n}} \sum_{i=1}^n \dot{\mathbf{l}}(X_i) - \frac{c^2}{2} I(f_0) + o_p(1),$$

and, hence under P_n with $p_n(\underline{x}) = \prod_{i=1}^n f(x_i)$,

$$\begin{pmatrix} t_n \\ \log L_n \end{pmatrix} \to_d N_2 \left(\begin{pmatrix} 0 \\ -(c^2/2) I(f_0) \end{pmatrix}, \begin{pmatrix} 1 & \sigma_{tL} \\ \sigma_{tL} & c^2 I(f_0) \end{pmatrix} \right)$$

where

$$\sigma_{tL} = eE_f \frac{X-\theta_0}{\sigma} \dot{\mathbf{l}}(X) = cE_{f_0}\left\{ -\frac{f_0'(X)}{f_0(X)} \right\}.$$

But

$$1 = \frac{\theta_n - \theta_0}{c/\sqrt{n}} = \frac{1}{c/\sqrt{n}}\{E_{f_n}(X) - E_f(X)\} = E_f\left\{ X\left(\frac{(f_n/f)(X) - 1}{c/\sqrt{n}} \right) \right\},$$

where the right side converges to $E_f\{X(-f'/f)(X)\}$. Thus $1 = E_f\{X(-f'/f)(X)\}$ and $\sigma_{tL} = c/\sigma$. Hence it follows from Le Cam's third lemma with $q_n(\underline{x}) = \prod_{i=1}^n f_n(x_i)$ that, under Q_n,

$$\begin{pmatrix} t_n \\ \log L_n \end{pmatrix} \to_d N_2\left(\begin{pmatrix} c/\sigma \\ +(c^2/2)I(f_0) \end{pmatrix}, \begin{pmatrix} 1 & \sigma_{tL} \\ \sigma_{tL} & c^2 I(f_0) \end{pmatrix} \right).$$

Hence the efficacy of the $t-$test is (again) $\epsilon_t = 1/\sigma^2$.

Example 4.6 The one-sample sign test again. Now consider the sign statistic $S_n = \sqrt{n}(\overline{Y} - 1/2)$ where $Y_i \equiv 1_{(\theta_0,\infty)}(X_i)$ and L_n is as above. Then under P_n with $p_n(\underline{x}) = \prod_{i=1}^n f(x_i)$,

$$\begin{pmatrix} S_n \\ \log L_n \end{pmatrix} \to_d N_2\left(\begin{pmatrix} 0 \\ +(c^2/2)I(f_0) \end{pmatrix}, \begin{pmatrix} 1 & \sigma_{SL} \\ \sigma_{SL} & c^2 I(f_0) \end{pmatrix} \right)$$

where

$$\sigma_{SL} \equiv cE_f 1_{(\theta_0,\infty)}(X)\dot{\mathbf{l}}(X) = -c\int_{\theta_0}^{\infty} \frac{f'}{f}(x) f(x)dx = cf(\theta_0).$$

Hence, by Le Cam's third lemma

$$\begin{pmatrix} S_n \\ \log L_n \end{pmatrix} \to_d N_2\left(\begin{pmatrix} cf(\theta_0) \\ +(c^2/2)I(f_0) \end{pmatrix}, \begin{pmatrix} 1 & cf(\theta_0) \\ cf(\theta_0) & c^2 I(f_0) \end{pmatrix} \right)$$

and it follows that the efficacy of the sign test is $\epsilon_S = 4f^2(\theta_0) = 4f_0^2(0)$. Combining the two efficacies ϵ_t and ϵ_S yields the Pitman efficiency of the sign test relative to the $t-$test, $e_{S,t} = 4\sigma^2 f_0^2(0)$.

5 Confidence Sets and $p-$values

The theory of testing that has been developed in the previous sections in this chapter connects with with estimation theory via the construction of confidence sets. The material outlined in this section is drawn in large part from Sections 3.5 (pages 72-77) and 5.4 (pages 161-162) of Lehmann and Romano (2005), and Section 5.8 (pages 257 - 264) of Ferguson (1967).

First, a definition:

Definition 5.1 Let $\{S(x)\} \equiv \{S(x): \ x \in \mathcal{X}\}$ be a family of subsets of the parameter space Θ for a given sample space $\mathcal{X}$. Then $\{S(x)\}$ is said to be a family of *confidence sets of confidence level* $1-\alpha$ if

$$P_\theta(S(X) \text{ contains } \theta) = P_\theta(\theta \in S(X)) = 1-\alpha.$$

Construction of confidence sets from tests: Let $A(\theta_0)$ denotes the acceptance region of a size α nonrandomized test φ of the hypothesis $H_0: \ \theta = \theta_0$ against any alternative. That is

$$\varphi(x) = \begin{cases} 1 & \text{if } x \notin A(\theta_0) \\ 0 & \text{if } x \in A(\theta_0) \end{cases}$$

where $P_{\theta_0}(X \in A(\theta_0)) = 1-\alpha$. If we consider the sets $A(\theta)$, $\theta \in \Theta$, we have a family of acceptance regions, each a subset of $\mathcal{X}$ such that

$$P_\theta(X \in A(\theta)) = 1-\alpha.$$

Define $S(x) = \{\theta: \ x \in A(\theta)\}$, so that $\{\theta \in S(X)\} = \{X \in A(\theta)\}$. Then it follows that

$$P_\theta(\theta \in S(X)) = P_\theta(X \in A(\theta)) = 1-\alpha,$$

so the resulting family $\{S(x): \ x \in \mathcal{X}\}$ is a family of confidence sets of level $1-\alpha$.

Here are three examples:

Example 5.1 Suppose that $X \sim N(\mu, 1)$, and consider testing $H: \ \mu = \mu_0$ versus $K: \ \mu > \mu_0$. By the Karlin - Rubin theorem, the UMP test is $\varphi(X) = 1\{X > \mu_0 + z_\alpha\}$ with acceptance region $A(\mu_0) = \{x: \ x \le \mu_0 + z_\alpha\}$. Thus

$$\begin{aligned} A(\mu) &= \{x: \ x \le \mu + z_\alpha\}, \\ S(x) &= \{\mu: \ \mu \ge x - z_\alpha\}. \end{aligned}$$

Thus it follows that

$$\begin{aligned} 1-\alpha &= P_\mu(X \le \mu + z_\alpha) = P_\mu(\mu \ge X - z_\alpha) \\ &= P_\mu(\mu \ge \mu_\alpha(X)) \qquad \text{with } \mu_\alpha(X) \equiv X - z_\alpha \\ &= P_\mu(\mu \in [\mu_\alpha(X), \infty)) = P_\mu(\mu \in S(X)) \end{aligned}$$

where $S(X) \equiv [\mu_\alpha(X), \infty) = [X - z_\alpha, \infty)$, and hence the family $S(X)$ is a family of $1-\alpha$ level confidence sets for μ.

Example 5.2 Suppose that $X_1, \ldots, X_n$ are i.i.d. Weibull (α, β) as in Example 3.x.y. Consider testing $H: \ (\alpha, \beta) = (\alpha_0, \beta_0)$ versus $K: \ (\alpha, \beta) \neq (\alpha_0, \beta_0)$. From the theory developed in Chapter 4 we know that the likelihood ratio test based on the statistic

$$\lambda_n \equiv \lambda_n(\alpha_0, \beta_0) \equiv \frac{\sup_{\alpha>0,\beta>0} L_n(\alpha, \beta)}{L_n(\alpha_0, \beta_0)}$$

satisfies

$$2 \log \lambda_n(\alpha_0, \beta_0) \to_d \chi_2^2$$

when α_0, β_0 are true. Therefore the acceptance sets

$$A_n(\alpha_0, \beta_0) = \{\underline{X}: \ 2 \log \lambda_n(\alpha_0, \beta_0) \le \chi^2_{2,\delta}\}$$

where $P(\chi_2^2 > \chi^2_{2,\delta}) = \delta$ satisfy

$$P_{\alpha_0,\beta_0}(A_n(\alpha_0, \beta_0)) \to 1 - \delta, \qquad \text{as} \ \ n \to \infty,$$

and, similarly,

$$P_{\alpha,\beta}(A_n(\alpha, \beta)) \to 1 - \delta, \qquad \text{as} \ \ n \to \infty.$$

But then

$$P_{\alpha,\beta}(A_n(\alpha, \beta)) = P_{\alpha,\beta}((\alpha, \beta) \in S_n(\underline{X})) \to 1 - \delta$$

for the associated confidence sets

$$S_n(\underline{X}) = \{(\alpha, \beta) \in \mathbb{R}^{+2}: \ 2 \log \lambda_n(\alpha, \beta) \le \chi^2_{2,\delta}\}.$$

Example 5.3 (Lower confidence bounds for a distribution function) Suppose that $X_1, \ldots, X_n$ are i.i.d. with continuous distribution function F on $\mathbb{R}$. Consider testing $H: \ F = F_0$ (continuous) versus $K: \ F <_s F_0$ (i.e. $F(x) \ge F_0(x)$ with strict inequality for some x). One natural test statistic is

$$R_n \equiv \sup_{x \in \mathbb{R}} \frac{\mathbb{F}_n(x)}{F_0(x)}.$$

Now by Theorem 2.3.1 (and the discussion in Section 2.4) $\mathbb{F}_n \stackrel{d}{=} \mathbb{G}_n(F_0)$ where $\mathbb{G}_n$ is the empirical distribution function of i.i.d. Uniform$(0,1)$ random variables $\xi_1, \ldots, \xi_n$. Therefore

$$\begin{aligned} P_{F_0}(R_n > r) &= P_{F_0}\left(\sup_{x \in \mathbb{R}} \frac{\mathbb{G}_n(F_0(x))}{F_0(x)} > r\right) \\ &= P\left(\sup_{0<u<1} \frac{\mathbb{G}_n(u)}{u} > r\right) \\ &= 1/r, \qquad \text{for} \ \ r \ge 1; \end{aligned}$$

this is a result due to Daniels (1945); see e.g. Shorack and Wellner (1986), Theorem 9.1.2, page 345 . Thus the test

$$\varphi(\underline{X}) = \begin{cases} 1, & \text{if } R_n > 1/\alpha \\ 0, & \text{if } R_n \le 1/\alpha, \end{cases}$$

is a size α test with acceptance region

$$\begin{aligned} A(F_0) &= \left\{\underline{X} : \sup_{x\in\mathbb{R}} \frac{\mathbb{F}_n(x)}{F_0(x)} \le 1/\alpha\right\} = \{\underline{X} : \ \alpha\mathbb{F}_n(x) \le F_0(x) \ \text{ for all } x \in \mathbb{R}\}, \\ A(F) &= \{\underline{X} : \ \alpha\mathbb{F}_n(x) \le F(x) \ \text{ for all } x \in \mathbb{R}\}, \\ S(\underline{X}) &= \{F : \ F(x) \ge \alpha\mathbb{F}_n(x) \text{ for all } \mathbb{R}\}. \end{aligned}$$

Thus

$$\begin{aligned} P_F\{F \in S(\underline{X})\} &= P_F(F : \ F(x) \ge \alpha\mathbb{F}_n(x) \ \text{ for all } x \in \mathbb{R}) \\ &= P_F\left(\sup_{x\in\mathbb{R}} \frac{\mathbb{F}_n(x)}{F(x)} \le 1/\alpha\right) = 1-\alpha. \end{aligned}$$

This band is clearly most informative about F for small values of $F(x)$.

Example 5.4 (The Kolmogorov confidence band for a distribution function) Suppose that $X_1, \ldots, X_n$ are i.i.d. with continuous distribution function F. Consider testing $H: \ F = F_0$ (continuous) versus $K: \ F \ne F_0$ with strict inequality for some x). The classical Kolmogorov test statistic is $D_n \equiv D_n(F_0) = \sqrt{n}\sup_x |\mathbb{F}_n(x) - F_0(x)|$. Under the null hypothesis we have

$$D_n \stackrel{d}{=} \|\mathbb{U}_n(F_0)\|_\infty = \|\mathbb{U}_n\|_\infty \equiv \sup_{0\le t\le 1} |\mathbb{U}_n(t)|$$

where $\mathbb{U}_n = \sqrt{n}(\mathbb{G}_n - I)$ and $\mathbb{G}_n$ is the empirical distribution function of $\xi_1, \ldots, \xi_n$ i.i.d. $U(0,1)$. Thus for F_0 continuous

$$P_{F_0}(D_n(F_0) \le k_{n,\alpha}) = 1-\alpha$$

where $k_{n,\alpha}$ satisfies $P(\|\mathbb{U}_n\|_\infty > k_{n,\alpha}) = \alpha$. Note that $k_{n,\alpha} \to k_{\infty,\alpha}$ which satisfies $P(\|\mathbb{U}\|_\infty > k_{\infty,\alpha}) = \alpha$.

It follows that for all F continuous we have

$$P_F(\mathbb{F}_n(x) - k_{n,\alpha}n^{-1/2} \le F(x) \le \mathbb{F}_n(x) + k_{n,\alpha}n^{-1/2} \ \text{ for all } \ x \in \mathbb{R}) = 1-\alpha.$$

Since $0 \le F(x) \le 1$ we can replace the lower and upper bounds by

$$(\mathbb{F}_n(x) - k_{n,\alpha}n^{-1/2}) \vee 0 \quad \text{and} \quad (\mathbb{F}_n(x) + k_{n,\alpha}n^{-1/2}) \wedge 1$$

respectively. Note that the Kolmogorov bands are of constant width $2k_{n,\alpha}n^{-1/2}$ over most of the range and do not take advantage of the small variance $Var(\mathbb{F}_n(x)) = n^{-1}F(x)(1-F(x))$ when $F(x)$ is near 0 or 1. Since the Dvoretzky-Kiefer-Wolfowitz (1956) and Massart (1990) inequality says that

$$P(\|\mathbb{U}_n\|_\infty > \lambda) \le 2\exp(-2\lambda^2) \quad \text{for all } \lambda > 0,$$

and for all $n \ge 1$, replacing $k_{n,\alpha}$ by $((1/2)\log(2/\alpha))^{1/2}$ yields a slightly conservative band.

Example 5.5 (The Berk-Jones-Owen confidence band for a distribution function) Now consider the same setting as in Example 5.4, but consider the families of pointwise null and alternative hypotheses given by $H_t:\ F(t) = F_0(t)$ versus $K_t:\ F(t) \neq F_0(t)$ for $t \in \mathbb{R}$. Note that $H = \cap_t H_t$ and $K = \cup_t K_t$. We proceed by first developing tests for the pointwise null and alternative hypotheses based on likelihood ratio statistics: since $n\mathbb{F}_n(t) \sim \text{Bin}(n, F(t))$, we see that the likelihood ratio statistic for testing H_t versus K_t is given by

$$\begin{aligned} 2\log LR_t &= 2n\left(\mathbb{F}_n(t)\log\frac{\mathbb{F}_n(t)}{F_0(t)} + (1-\mathbb{F}_n(t))\log\frac{1-\mathbb{F}_n(t)}{1-F_0(t)}\right) \\ &\to_d \chi_1^2 \quad \text{under} \quad H_t. \end{aligned}$$

To test H versus K we use Roy's union-intersection principle: reject H in favor of $K = \cup_t K_t$ if

$$\begin{aligned} R_n &\equiv R_n(F_0) \equiv \sup_{t\in\mathbb{R}} 2\log LR_t \\ &= 2n\sup_{t\in\mathbb{R}}\left\{\mathbb{F}_n(t)\log\frac{\mathbb{F}_n(t)}{F_0(t)} + (1-\mathbb{F}_n(t))\log\frac{1-\mathbb{F}_n(t)}{1-F_0(t)}\right\} \end{aligned}$$

is "too large". R_n was proposed by Berk and Jones (1979) and is called the Berk-Jones statistic. It turns out that under the null hypothesis

$$R_n - r_n \to_d Y_4 \sim Ev^4$$

where $Ev(x) \equiv \exp(-\exp(-x))$ and $r_n \equiv \log_2(n) + (1/2)\log_3(n) - (1/2)\log(4\pi)$; see Berk and Jones (1979), Wellner and Koltchinskii (2003), Jager and Wellner (2007). Thus for any F continuous and $0 < \alpha < 1$ we can find $r_{n,\alpha} = r_n + c_{n,\alpha}$ so that

$$P_F(R_n(F) \le r_{n,\alpha}) \to 1-\alpha \quad \text{as} \quad n \to \infty.$$

Inversion of the resulting acceptance regions given by this family of tests yields the (approximate) $1-\alpha$ confidence bands for F proposed by Owen (1995):

$$P_F\left(\underline{BJ}_{n,\alpha}(x) \le F(x) \le \overline{BJ}_{n,\alpha}(x) \ \text{ for all } \ x \in \mathbb{R}\right) \to 1-\alpha$$

where $\underline{BJ}_{n,\alpha}$ and $\overline{BJ}_{n,\alpha}$ are obtained by inverting the acceptance region $\{R_n(F) \le r_{n,\alpha}\}$. For further details see Owen (1995) and Jager and Wellner (2004).

How should we choose a family of confidence sets? One natural criterion is to minimize the probability of covering false values. That is, we should try to make

$$P_{\theta'}(\theta \in S(X))$$

small if $\theta \neq \theta'$. It turns out that optimality properties of tests carry over or translate into optimality properties of confidence sets. One version of this is given in the following theorem.

Theorem 5.1 Let $A(\theta_0)$ be the acceptance region of a UMP test of size α of the hypothesis $H_0:\ \theta = \theta_0$ versus $H_1:\ \theta \in \Theta_1$. Then $\{S(x)\}$ defined by

$$\{\theta \in S(x)\} = \{x \in A(\theta)\}$$

minimizes $P_{\theta'}(\theta \in S(X))$ for all $\theta' \in \Theta_1$ among all level $1-\alpha$ families of confidence sets.

See e.g. Lehmann and Romano (2005), pages 72 - 77 and 164 - 168.

Suppose that $\theta = (\nu, \xi) \in \Theta$, $\nu \in \mathbb{R}$, $\xi \in \mathbb{R}^k$ for some k. A lower confidence bound for ν is a function $\underline{\nu}(x)$ such that

$$P_{\nu,\xi}(\underline{\nu}(X) \le \nu) \ge 1-\alpha \qquad \text{for all } \ \nu, \xi.$$

Similarly, a confidence interval for ν at confidence level $1-\alpha$ is given by $\underline{\nu}(x)$, $\overline{\nu}(x)$ satisfying

$$P_{\nu,\xi}(\underline{\nu}(X) \le \nu \le \overline{\nu}X)) \ge 1-\alpha \qquad \text{for all } \ \nu, \xi.$$

Connection with estimation: If $\underline{\nu}(X)$ satisfies

$$P_{\nu,\xi}(\underline{\nu}(X) \le \nu) = P_{\nu,\xi}(\underline{\nu}(X) \ge \nu) = 1/2,$$

then $\underline{\nu}(X)$ is a *median unbiased* estimator of ν. For the use of this in developing "R-estimators of location", see e.g. Hodges and Lehmann (1963).

p-values:

See Lehmann and Romano, pages 57, 63-65, 97-98, 108-109, and 139.

Consider a family of tests $\varphi_\alpha(X) = 1_{A_\alpha^c}(X)$ of H versus K with rejection regions A_α^c satisfying

$$(1) \qquad A_\alpha^c \subset A_{\alpha'}^c \qquad \text{if } \ \alpha < \alpha'.$$

Let

$$\begin{aligned} \hat{p} &= \hat{p}(X) \equiv \inf\{\alpha : \ X \in A_\alpha^c\} \\ &= \text{smallest significance value for which } \ H \\ &\qquad \text{would be rejected for the observed data } X \\ &\equiv \text{p-value of the test(s) } \ \varphi_\alpha. \end{aligned}$$

Example 5.6 Suppose that $X \sim N(\mu, \sigma^2)$, σ^2 known. For testing $H: \ \mu = 0$ versus $K: \ \mu > 0$, the UMP test is given by

$$\begin{aligned} A_\alpha^c &= \{x : \ x > \sigma z_\alpha\}, \qquad z_\alpha \equiv \Phi^{-1}(1-\alpha) \\ &= \{x : \ \Phi(x/\sigma) > \Phi(z_\alpha) = 1-\alpha\} \\ &= \{x : \ 1 - \Phi(x/\sigma) < \alpha\}, \end{aligned}$$

so $\hat{p}(X) = 1 - \Phi(X/\sigma)$. Alternatively,

$$\hat{p}(X(\omega)) = P_0(X \ge x)|_{x=X(\omega)} = \left(1 - \Phi\left(\frac{x}{\sigma}\right)\right)\Big|_{x=X(\omega)}.$$

Note that

$$P_0(\hat{p} \le u) \ = \ P_0(1 - \Phi(X/\sigma) \le u) = P_0(\Phi(X/\sigma) \ge 1-u) = u$$

for $0 < u < 1$ since $\Phi(X/\sigma) \sim \text{Uniform}(0,1)$ under P_0.

Lemma 5.1 Suppose that $X \sim P_\theta$ for some $\theta \in \Theta$ and $H: \theta \in \Theta_0$. Suppose that the test ϕ of H versus K has rejection regions A_α^c satisfying the nesting property (1).

(i) If

$$\sup_{\theta \in \Theta_0} P_\theta(X \in A_\alpha^c) \le \alpha \qquad \text{for all } \ 0 < \alpha < 1,$$

then the distribution of $\hat{p}$ under $\theta \in \Theta_0$ satisfies

$$P_\theta(\hat{p} \le u) \le u \qquad \text{for all } \ 0 \le u \le 1.$$

(ii) If for $\theta \in \Theta_0$

$$P_\theta(X \in A_\alpha^c) = \alpha \qquad \text{for all } \ 0 < \alpha < 1,$$

then

$$P_\theta(\hat{p} \le u) = u \qquad \text{for all } \ 0 \le u \le 1, \ \text{ and } \ \theta \in \Theta_0.$$

Proof. (i) If $\theta \in \Theta_0$, then $\{\hat{p} \le u\} \subset \{X \in A_v^c\}$ for all $u < v$. Thus $P_\theta(\hat{p} \le u) \le P_\theta(X \in A_v^c) \le v$ for all $v > u$, and hence, letting $v \searrow u$, $P_\theta(\hat{p} \le u) \le u$.
(ii) Since $\{X \in A_u^c\} \subset \{\hat{p} \le u\}$,

$$P_\theta(\hat{p} \le u) \ge P_\theta(X \in A_u^c) = u.$$

But since also $P_\theta(\hat{p} \le u) \le u$ from (i), the claimed equality follows. □

Example 5.7 (Test for μ when $X \sim N(\mu, \sigma^2)$, continued.) What is the distribution of $\hat{p}$ under $\mu > 0$?

$$\begin{aligned} P_\mu(\hat{p} \le u) &= P_\mu(1 - \Phi(X/\sigma) \le u) = P_\mu(X \ge \sigma\Phi^{-1}(1-u)) \\ &= P_\mu\left(\frac{X-\mu}{\sigma} \ge \Phi^{-1}(1-u) - \frac{\mu}{\sigma}\right) \\ &= 1 - \Phi(\Phi^{-1}(1-u) - \mu/\sigma). \end{aligned}$$

Note that since

$$\Phi^{-1}(1-u) - \mu/\sigma \le \Phi^{-1}(1-u),$$

it follows that

$$\Phi(\Phi^{-1}(1-u) - \mu/\sigma) \le 1 - u,$$

and hence

$$P_\mu(\hat{p} \le u) = 1 - \Phi(\Phi^{-1}(1-u) - \mu/\sigma) \ge 1 - (1-u) = u.$$

Moreover, if $X_1, \ldots, X_n$ are i.i.d. $N(\mu, 1)$ with $\mu > 0$ then the p-value is

$$\hat{p}_n(\omega) = P_0(\sqrt{n}\overline{X}_n > z)|_{z=\sqrt{n}\overline{X}_n(\omega)} = 1 - \Phi(\sqrt{n}\overline{X}_n(\omega)),$$

and since $1 - \Phi(x) \sim x^{-1}\phi(x)$ and $\overline{X}_n(\omega) \to \mu$ for all ω in a set with probability 1,

$$\hat{p}_n(\omega) \sim \frac{1}{\sqrt{2\pi}} \exp(-n\mu^2/2) \frac{1}{\sqrt{n}\mu} \to 0$$

exponentially quickly as $n \to \infty$ on a set of ω's with probability 1.