MDP Exercise 11.4. (*Uniform*$[\theta,\, \theta+1]$) In Example 11.14, take $\theta_1 = \theta$ and $\theta_2 = \theta + 1$, where $\theta \in (-\infty, \infty)$ is a real-valued location parameter. For simplicity set $b(x) \equiv 1$, so each $X_i \sim$Uniform$[\theta,\, \theta+1]$. Show that $(X_{(1)},\, X_{(n)})$ remains a 2-dimensional minimal sufficient statistic for the 1-dimensional parameter θ.

MDP Exercise 11.5. For example, in Exercise 11.4, the minimal sufficient statistic $(X_{(1)},\, X_{(n)})$ is equivalent to the pair $(X_{(1)},\, R_n)$ where $R_n \equiv X_{(n)} - X_{(1)}$ is the sample range. Note that $0 \le X_{(1)} - \theta \le 1 - R_n$, so R_n, which is clearly ancillary hence provides no information about θ by itself, nonetheless governs the accuracy of $X_{(1)}$ as an estimator of θ. In fact, because R_n is ancillary, we can base inference about θ on the conditional distribution of $X_{(1)} \mid R_n$.

(i) Find this conditional distribution. (See CB Example 5.4.7 for a related discussion.) Use this conditional distribution to find an estimator $\tilde{\theta}_n$ that is conditionally unbiased for θ, thus unconditionally unbiased.

(ii) Let $\check{\theta}_n = X_{(1)} - \frac{1}{n+1}$. Show that $\check{\theta}_n$ is unbiased for θ, that $\mathrm{Var}(\tilde{\theta}_n) < \mathrm{Var}(\check{\theta}_n)$ for all n, and that $\lim_{n\to\infty} \mathrm{Var}(\tilde{\theta}_n)/\mathrm{Var}(\check{\theta}_n) = \frac{1}{2}$.

(iii) Find a confidence interval for θ, centered at $\tilde{\theta}$, whose conditional and unconditional confidence coefficient is $(1 - \alpha)$. □

Solution: (i) Let $U = X_{(1)} - \theta$, $V = X_{(n)} - \theta$, so $0 < U < V < 1$. Then

$$\Pr[u < U < V < v] = (v - u)^n,$$

so

$$f(u, v) = n(n-1)(v - u)^{n-2}.$$

Now let $R = V - U$, so

$$f(u, r) = n(n-1)r^{n-2}, \qquad 0 < u, r < 1,\ 0 < u + r < 1.$$

Thus

$$U \mid R \sim \text{Uniform}(0,\, 1 - R), \tag{11.a}$$

so

$$\mathrm{E}(U \mid R) = \tfrac{1}{2}(1 - R),$$
$$\mathrm{E}(X_{(1)} \mid R) = \theta + \tfrac{1}{2}(1 - R).$$

Therefore

$$\tilde{\theta}_n := X_{(1)} - \tfrac{1}{2}(1-R) = \tfrac{1}{2}(X_{(1)} + X_{(n)}) - \tfrac{1}{2}$$

is conditionally and unconditionally unbiased for θ.

(ii) From (9.9), $\mathrm{E}(X_{(1)}) = \theta + \frac{1}{n+1}$, hence $\check{\theta}_n$ is unbiased. From (9.10) and MDP Exercise 9.3,

$$\begin{aligned}
\mathrm{Var}(\check{\theta}_n) &= \frac{n}{(n+1)^2(n+2)},\\
\mathrm{Var}(\tilde{\theta}_n) &= \tfrac{1}{4}[\mathrm{Var}(X_{(1)}) + \mathrm{Var}(X_{(n)}) + 2\mathrm{Cov}(X_{(1)}, X_{(n)})]\\
&= \frac{1}{2}\Big[\frac{n}{(n+1)^2(n+2)} + \frac{1}{(n+1)^2(n+2)}\Big]\\
&= \frac{1}{2(n+1)(n+2)}.
\end{aligned}$$

Thus for $n > 1$,

$$1 > \frac{\mathrm{Var}(\tilde{\theta}_n)}{\mathrm{Var}(\check{\theta}_n)} = \frac{n+1}{2n} \to \frac{1}{2} \text{ as } n \to \infty.$$

(iii) From (11.a),

$$\text{(11.b)} \qquad \frac{X_{(1)} - \theta}{1-R} \,\Big|\, R \sim \text{Uniform}(0,\,1),$$

so

$$\begin{aligned}
(1-\alpha) &= \Pr\left[(1-R)\tfrac{\alpha}{2} < X_{(1)} - \theta < (1-R)(1-\tfrac{\alpha}{2})\right]\\
&= \Pr\left[\tilde{\theta}_n - \tfrac{1}{2}(1-R)(1-\alpha) < \theta < \tilde{\theta}_n + \tfrac{1}{2}(1-R)(1-\alpha)\right].
\end{aligned}$$

Note that the ancillary statistic $1-R$ governs the width of this confidence interval. $\square$

• **Completeness and minimal sufficiency.**

Theorem. If $\mathcal{S}^*$ is sufficient for $\mathcal{P}$ and $\mathcal{P}$ is complete on $\mathcal{S}^*$, then $\mathcal{S}^*$ is minimal sufficient for $\mathcal{P}$. Briefly: "a complete and sufficient subfield/statistic is minimal sufficient."

Proof. We must show that $\mathcal{S}^* \subseteq \mathcal{S}_0\,[\mathcal{P}]$ for any other sufficient subfield $\mathcal{S}_0$. Because $\mathcal{S}_0$ is sufficient, for each $S^* \in \mathcal{S}^*$ there exists an $\mathcal{S}_0$-measurable version g_{S^*} of $P[S^*|\mathcal{S}_0]\ \forall P \in \mathcal{P}$. Because $\mathcal{S}^*$ is sufficient, there exists an $\mathcal{S}^*$-measurable version h_{S^*} of $\mathrm{E}_P[g_{S^*}|\mathcal{S}^*]\ \forall P \in \mathcal{P}$. Thus for any $P \in \mathcal{P}$,

$$\mathrm{E}_P(h_{S^*}) = \mathrm{E}_P(g_{S^*}) = \mathrm{E}_P(I_{S^*}) \equiv P(S^*),$$

so $\mathrm{E}_P[h_{S^*} - I_{S^*}] = 0\ \forall P \in \mathcal{P}$. Because $h_{S^*} - I_{S^*}$ is $\mathcal{S}^*$-measurable and $\mathcal{P}$ is complete on $\mathcal{S}^*$, $h_{S^*} = I_{S^*}$ a.e. $[\mathcal{P}]$.

Set $S_0 = \{g_{S^*} = 1\} \in \mathcal{S}^*$; we shall show that $S_0 \triangle S^*$ is $\mathcal{P}$-null. First,

$$P(S^*) = \int I_{S^*}\, dP = \int_{S^*} h_{S^*}\, dP = \int_{S^*} g_{S^*}\, dP \le P(S^*) \quad \forall P \in \mathcal{P},$$

since $0 \le g_S^* \le 1$ a.e. $[\mathcal{P}]$. Therefore $S^* \cap \{g_{S^*} < 1\}$ is $\mathcal{P}$-null, hence $\{g_{S^*} \ne 1\} \cap S^*$ is $\mathcal{P}$-null. Next,

$$0 = \int_{(S^*)^c} I_{S^*}\, dP = \int_{(S^*)^c} h_{S^*}\, dP = \int_{(S^*)^c} g_{S^*}\, dP \ge 0 \quad \forall P \in \mathcal{P},$$

so $\{g_{S^*} = 1\} \cap (S^*)^c$ is $\mathcal{P}$-null. Thus $S_0 \triangle S^*$ is $\mathcal{P}$-null, as claimed.

Example. Let $X \sim N(\theta, 1)$ with $\theta = \pm 1$, i.e., $\mathcal{P} = \{N(1,1), N(-1,1)\}$. Show that X is minimal sufficient [verify!] but not complete, since X contains the nontrivial ancillary statistic $|X|$ [verify!].

Exercise 14.37.** Let $X_1, \ldots, X_n$ be an i.i.d. sample from the univariate normal location-scale family $N_1(\mu, \sigma^2)$. The MLEs of μ when σ^2 is known and when σ^2 is unknown are both $\bar{X}_n$, so trivially have the same asymptotic efficiency. This also follows from the fact that the $N_1(0,1)$ pdf is symmetric about 0, so μ and σ are orthogonal parameters. For a *finite* sample size $n \geq 2$, however, different confidence intervals are used for the two cases:

$$\sigma^2 \text{ unknown: } \bar{X}_n \pm \tfrac{s_n}{\sqrt{n}}\, t_{n-1;\alpha/2}, \qquad \sigma^2 \text{ known: } \bar{X}_n \pm \tfrac{\sigma}{\sqrt{n}}\, z_{\alpha/2}.$$

Show that the expected width of the first confidence interval is greater than the (non-random) width of the second, so that knowing σ^2 actually improves the accuracy of inference about μ (on average).

A. Unsuccessful analytic approach. We need to show that

$$\mathrm{E}\left(\frac{s_n}{\sqrt{n}}\, t_{n-1;\alpha/2}\right) > \frac{\sigma}{\sqrt{n}}\, z_{\alpha/2}. \tag{1}$$

It is tempting to try to establish (1) in two steps (we set $n-1=k$ for notational convenience):

(a) Show that $t_{k;\alpha/2} > z_{\alpha/2}$. This is equivalent to showing that $F_{1,k}$ is stochastically greater than χ_1^2, which in turn follows from the facts that $F_{1,k}$ is stochastically decreasing[43] in k and $F_{1,k} \xrightarrow{d} \chi_1^2$ as $k \to \infty$.

(b) Try to show that $\mathrm{E}(s_n) > \sigma$. Because $\frac{s_n^2}{\sigma^2} \sim \frac{\chi_k^2}{k}$,

$$\mathrm{E}\left(\tfrac{s_n}{\sigma}\right) = \sqrt{\tfrac{2}{k}}\,\frac{\Gamma(\frac{k+1}{2})}{\Gamma(\frac{k}{2})}, \tag{2}$$

so the hoped-for inequality in (b) can be re-written as

$$\sqrt{\tfrac{k}{2}} < \frac{\Gamma(\frac{k+1}{2})}{\Gamma(\frac{k}{2})}\ ? \tag{3}$$

[43] Attributed to P. L. Hsu (1938), J. Hajek (1962); see M. L. Eaton (1987), "Group induced orderings with some applications in statistics", CWI Newsletter No. 16, 3-31. The result can be obtained from the representation $F_{1,k} \sim U/(V/k)$ with $U \sim \chi_1^2$, $V \sim \chi_k^2$, $U \perp\!\!\!\perp V$. Condition on V and use the fact that the cdf of U is concave (since the pdf is decreasing). [Note to myself: can use majorization.]

Unfortunately the opposite is true. Because $\sqrt{x}$ is strictly concave, Jensen's Inequality yields

$$\mathrm{E}\left(\tfrac{s_n}{\sigma}\right) = \mathrm{E}\left(\sqrt{\frac{\chi_k^2}{k}}\right) < \sqrt{\mathrm{E}\left(\frac{\chi_k^2}{k}\right)} = 1, \tag{4}$$

so equivalently,

$$\frac{\Gamma(\frac{k+1}{2})}{\Gamma(\frac{k}{2})} < \sqrt{\frac{k}{2}}\ . \tag{5}$$

Comparison of (3) and (5): Set $x = \frac{k-1}{2}$ so (3) can be re-written as

$$\sqrt{x+\tfrac{1}{2}} < \frac{\Gamma(x+1)}{\Gamma(x+\frac{1}{2})}\ ? \tag{3$'$}$$

It follows from inequalities of Gautschi (1959, also see below) and Kershaw (1983), respectively, that for $x > 0$,

$$\sqrt{x} < \frac{\Gamma(x+1)}{\Gamma(x+\frac{1}{2})}, \tag{6}$$

$$\sqrt{x+\tfrac{1}{4}} < \frac{\Gamma(x+1)}{\Gamma(x+\frac{1}{2})}, \tag{7}$$

whereas (5) can be re-written as

$$\frac{\Gamma(x+1)}{\Gamma(x+\frac{1}{2})} < \sqrt{x+\tfrac{1}{2}}\ . \tag{5$'$}$$

Thus, although (3′) fails, it doesn't fail by much.

Gautschi's Inequality: Because the gamma function $\Gamma(x)$ is strictly log convex (verify by differentiating twice under the integral sign),

$$\Gamma(x+\tfrac{1}{2}) < \sqrt{\Gamma(x)\Gamma(x+1)},$$

so

$$\frac{\Gamma(x+\frac{1}{2})}{\Gamma(x+1)} < \sqrt{\frac{\Gamma(x)}{\Gamma(x+1)}} = \sqrt{\frac{1}{x}}.$$

B. Successful statistical approach. In 1961, John Pratt and P. K. Ghosh independently established the following identity,[44] which relates the expected length of a confidence interval $(l(X), u(X))$ for a real-valued parameter θ based on data $X \sim f_\theta(x)$ to the power function of the corresponding test: for any fixed θ_0,

$$\mathrm{E}_{\theta_0}[u(X) - l(X)] = \int_{\theta \neq \theta_0} \mathrm{Pr}_{\theta_0}[\theta \in (l(X), u(X))] d\theta. \tag{8}$$

If we let $\pi(\theta; \theta_0)$ denote the power function of the corresponding 2-sided test for testing $\theta = \theta_0$ vs. $\theta \neq \theta_0$, then

$$1 - \pi(\theta; \theta_0) = \mathrm{Pr}_\theta[l(X) < \theta_0 < u(X)],$$

so (8) can be written as

$$\mathrm{E}_{\theta_0}[u(X) - l(X)] = \int_{\theta \neq \theta_0} [1 - \pi(\theta_0; \theta)] d\theta. \tag{8}$$

(Note the interchange of θ and θ_0.) Therefore (1) will follow if we can show that the power of the 2-sided size-α t-test for testing $\mu = \mu_0$ with σ^2 unknown (but fixed at $\sigma^2 = 1$) is *less than* that of the 2-sided size-α Z-test for $\mu = \mu_0$ (with σ^2 known to $= 1$). $[Z \sim N(0, 1).]$

Because the 2-sided t_n-test is equivalent to the 1-sided $F_{1,n}$-test, where $F_{1,n} = t_n^2$, it therefore suffices to show that the power of the (non-central) $F_{1,n}$-test is greater than that of the 1-sided (non-central) χ_1^2-test, where $\chi_1^2 = Z^2$. In fact a stronger result was proved by Das Gupta and MDP (JASA 1974): For fixed value of the non-centrality parameter, the power of the non-central $F_{m,n}$-test is increasing in n (and decreasing in m). Therefore *the expected length of the t-based confidence interval on the left side of (1) is decreasing in n.* Now (1) follows from the fact that $t_n \to N(0, 1)$ as $n \to \infty$.

[44] This is proved via Fubini's Theorem, as in CB Exercise 2.14; see Lehmann and Romano *Testing Statistical Hypotheses*, 3rd edition, p.200.

MDP Exercise 16.5. Show that the Bayes estimator $\hat{\theta} \equiv \mathrm{E}[\Theta \mid X]$ is unbiased for θ iff it is perfect, i.e., $\hat{\theta} = \Theta$.

Solution. (The statement and proof in CB §7.5.2 are incorrect.) First,

$$\begin{aligned}
\mathrm{E}[(\hat{\theta} - \Theta)^2] &= \mathrm{E}(\hat{\theta}^2) - 2\mathrm{E}[\hat{\theta}\,\Theta] + \mathrm{E}(\Theta^2) \\
&= \mathrm{E}(\hat{\theta}^2) - 2\mathrm{E}[\mathrm{E}(\hat{\theta}\,\Theta \mid X)] + \mathrm{E}(\Theta^2) \\
&= \mathrm{E}(\hat{\theta}^2) - 2\mathrm{E}[\hat{\theta}\,\mathrm{E}(\Theta \mid X)] + \mathrm{E}(\Theta^2) \\
&\equiv \mathrm{E}(\hat{\theta}^2) - 2\mathrm{E}(\hat{\theta}^2) + \mathrm{E}(\Theta^2) \\
&= \mathrm{E}(\Theta^2) - \mathrm{E}(\hat{\theta}^2).
\end{aligned}$$

But by the unbiasedness of $\hat{\theta}$,

$$\begin{aligned}
\mathrm{E}[(\hat{\theta} - \Theta)^2] &= \mathrm{E}(\hat{\theta}^2) - 2\mathrm{E}[\hat{\theta}\,\Theta] + \mathrm{E}(\Theta^2) \\
&= \mathrm{E}(\hat{\theta}^2) - 2\mathrm{E}[\mathrm{E}(\hat{\theta}\,\Theta \mid \Theta)] + \mathrm{E}(\Theta^2) \\
&= \mathrm{E}(\hat{\theta}^2) - 2\mathrm{E}[\Theta\,\mathrm{E}(\hat{\theta} \mid \Theta)] + \mathrm{E}(\Theta^2) \\
&\equiv \mathrm{E}(\hat{\theta}^2) - 2\mathrm{E}(\Theta^2) + \mathrm{E}(\Theta^2) \\
&= \mathrm{E}(\hat{\theta}^2) - \mathrm{E}(\Theta^2),
\end{aligned}$$

so $\mathrm{E}[(\hat{\theta} - \Theta)^2] = 0$, i.e., $\hat{\theta} \equiv \mathrm{E}[\Theta \mid X] = \Theta$. Therefore $\hat{\theta}$ is perfect. The converse is trivial. $\square$

Exercise 17.3. Let $R = \frac{\psi_0 c_{10}}{\psi_1 c_{01}}$, so $c^* = \log R + \frac{1}{2}\left(\mu_1'\Sigma^{-1}\mu_1 - \mu_0'\Sigma^{-1}\mu_0\right)$. Let $d = \Sigma^{-1}(\mu_1 - \mu_0)$ and $\Delta^2 = (\mu_1 - \mu_0)'\Sigma^{-1}(\mu_1 - \mu_0)$. Then

$$\begin{aligned}
&P[\phi_\psi \text{ chooses } \mu_0 \mid \mu_1] = P[d'X < c^* \mid \mu_1] = P[N_1(d'\mu_1,\, d'\Sigma d) < c^*] \\
=&P[N_1((\mu_1 - \mu_0)'\Sigma^{-1}\mu_1,\, \Delta^2) < \log R + \tfrac{1}{2}\left(\mu_1'\Sigma^{-1}\mu_1 - \mu_0'\Sigma^{-1}\mu_0\right)] \\
=&P[N_1(0,\, \Delta^2) < \log R - \tfrac{1}{2}\Delta^2] \\
=&1 - \Phi\left(\tfrac{\frac{1}{2}\Delta^2 - \log R}{\Delta}\right) \qquad\qquad [\Phi = \text{cdf of } N_1(0,1).]
\end{aligned}$$

Similarly, $P[\phi_\psi \text{ chooses } \mu_1 \mid \mu_0] = 1 - \Phi\left(\frac{\frac{1}{2}\Delta^2 + \log R}{\Delta}\right)$.

Note: When $R = 1$ (e.g., if $\psi_0 = \psi_1$ and $c_{10} = c_{01}$), then

$$P[\phi_\psi \text{ chooses } \mu_0 \mid \mu_1] = P[\phi_\psi \text{ chooses } \mu_1 \mid \mu_0] = 1 - \Phi\left(\tfrac{1}{2}\Delta\right).$$

Δ is the *Mahalanobis distance* between $N_p(\mu_0, \Sigma)$ and $N_p(\mu_1, \Sigma)$. It is easy to verify that the Kullback-Leibler divergence between $N_p(\mu_0, \Sigma)$ and $N_p(\mu_1, \Sigma)$ is equal to $\frac{1}{2}\Delta^2$. (Compare to MDP eqn. (14.11).)

MDP Exercise 17.6. The posterior distribution is still given by (16.15):

$$\Theta \mid X \sim \text{Beta}(X + \alpha,\, n - X + \beta).$$

However, the expected posterior loss (EPL) for $\tilde{L} \equiv \frac{(a-\theta)^2}{\theta(1-\theta)}$ now becomes

$$\text{E}\left[\tfrac{(a-\Theta)^2}{\Theta(1-\Theta)} \,\middle|\, X\right] = \text{const} \cdot \int_0^1 (a-\theta)^2 \theta^{(X+\alpha-1)-1}(1-\theta)^{(n-X+\beta-1)-1} d\theta.$$

If $(X \neq 0, n)$ or $(X = 0,\, \alpha > 1)$ or $(X = n,\, \beta > 1)$ this is minimized when

$$(1) \qquad a = \text{E}[\text{Beta}(X + \alpha - 1,\, n - X + \beta - 1)] = \tfrac{X+\alpha-1}{n+\alpha+\beta-2},$$

which agrees with (17.21). If $(X = 0,\, \alpha = 1)$, the EPL $= \infty$ unless $a = 0$, so is trivially minimized at $a = 0$. If $(X = n,\, \beta = 1)$, the EPL $= \infty$ unless $a = 1$, so is trivially minimized at $a = 1$. These last two cases also agree with (17.21). Thus for $\alpha = \beta = 1$ the unique Bayes estimator is $\frac{X}{n}$, so this is admissible w.r.to $\tilde{L} \equiv \frac{(a-\theta)^2}{\theta(1-\theta)}$. Clearly admissibility w.r.to $L \equiv (a - \theta)^2$ is equivalent to admissibility w.r.to $\tilde{L}$ because, for any estimator d, the corresponding risk functions ar related by $\tilde{R}_d(\theta) = \frac{R_d(\theta)}{\theta(1-\theta)}$.

MDP Exercise 18.14. (i) Let $f_\theta(x) = \theta^{-1} I_{[0,\theta]}(x)$. By MDP Propositions 18.3 and 18.4, every admissible test (i.e., one with risk vector on the SW boundary) for testing f_1 vs. f_2 is a likelihood ratio test of the form (18.5). The LR is

$$\frac{f_2(x)}{f_1(x)} = \begin{cases} \frac{1}{2} & \text{if } 0 \le x \le 1; \\ \infty & \text{if } 1 < x \le 2. \end{cases}$$

Thus the LR test (18.5) has one of the following four forms:

A. $\quad c < \frac{1}{2}: \quad \phi_c(x) = 1 \;\text{ for all } 0 \le x \le 2;$

B. $\quad c = \frac{1}{2}: \quad \phi_c(x) = \begin{cases} \gamma_1(x) & \text{if } 0 \le x \le 1; \\ 1 & \text{if } 1 < x \le 2; \end{cases}$

C. $\quad \frac{1}{2} < c < \infty: \quad \phi_c(x) = \begin{cases} 0 & \text{if } 0 \le x \le 1; \\ 1 & \text{if } 1 < x \le 2; \end{cases}$

D. $\quad c = \infty: \quad \phi_c(x) = \begin{cases} 0 & \text{if } 0 \le x \le 1; \\ \gamma_2(x) & \text{if } 1 < x \le 2. \end{cases}$

Thus the SW boundary of the risk set is determined by the risk vectors (R_0, R_1) of these four classes of tests, as follows [verify from (18.2)]:

A. $\quad (R_0, R_1) = (1, 0);$

B. $\quad (R_0, R_1) = \left(\bar{\gamma}_1, \frac{1}{2}(1 - \bar{\gamma}_1)\right);$

C. $\quad (R_0, R_1) = (0, \frac{1}{2});$

D. $\quad (R_0, R_1) = \left(0, 1 - \frac{1}{2}\bar{\gamma}_2\right).$

where $\bar{\gamma}_1 = \int_0^1 \gamma_1(x)dx$, $\bar{\gamma}_2 = \int_1^2 \gamma_2(x)dx$, so $0 \le \bar{\gamma}_1 \le 1$, $0 \le \bar{\gamma}_2 \le 1$.

(ii) We may base all tests on the minimal sufficient statistic $X \equiv X_{\max}$, whose pdf is $f_\theta(x) = \theta^{-n} n x^{n-1} I_{[0,\theta]}(x)$. Now the LR is

$$\frac{f_2(x)}{f_1(x)} = \begin{cases} \frac{1}{2^n} & \text{if } 0 \le x \le 1; \\ \infty & \text{if } 1 < x \le 2. \end{cases}$$

The LR test (18.5) has the form A, B, C, or D with $\frac{1}{2}$ replaced by $\frac{1}{2^n}$. Note that these clases of tests are exactly the same as above. The SW boundary of the risk set is determined by the risk vectors $(R_0,\, R_1)$ of these four classes of tests, as follows [verify from (18.2)]:

A. $(R_0,\, R_1) = (1,\, 0)$;

B. $(R_0,\, R_1) = \left(\bar{\gamma}_1,\, \frac{1}{2^n}(1-\bar{\gamma}_1)\right)$;

C. $(R_0,\, R_1) = (0,\, \frac{1}{2^n})$;

D. $(R_0,\, R_1) = \left(0,\, 1 - \frac{1}{2^n}\bar{\gamma}_2\right)$.

where now $\bar{\gamma}_1 = n\int_0^1 \gamma_1(x)x^{n-1}dx$, $\bar{\gamma}_2 = n\int_1^2 \gamma_2(x)x^{n-1}dx$, so $0 \le \bar{\gamma}_1 \le 1$, $0 \le \bar{\gamma}_2 \le 2^n - 1$. Finally, because $\frac{1}{2^n} \to 0$ at an exponential rate, the entire risk set fills the unit square at an exponential rate.

MDP Exercise 18.22. MLR preserves monotonicity. *If $f(x|\lambda)$ has MLR in $T \equiv T(X)$ and $g(t)$ is nondecreasing in t, then*

$$\mathrm{E}_\lambda[g(T(X))] \equiv \int_{\mathcal{X}} g(t(x))\, f(x|\lambda)d\nu(x)$$

is nondecreasing in λ (ν is either counting measure or Lebesgue measure).

Proof. Set $h(\lambda) = \mathrm{E}_\lambda[g(T(X))]$. Then for any $\lambda_1 \le \lambda_2$ in Λ,

$$\begin{aligned}
&h(\lambda_2) - h(\lambda_1)\\
&= \int g(t(x))[f(x|\lambda_2) - f(x|\lambda_1)]d\nu(x)\\
&= \tfrac{1}{2}\iint [g(t(x)) - g(t(y))]\,[f(x|\lambda_2)f(y|\lambda_1) - f(y|\lambda_2)f(x|\lambda_1)]d\nu(x)d\nu(y)\\
&\ge 0,
\end{aligned}$$

since the 2 terms in brackets $[\cdot]$ are both ≥ 0 if $x \ge y$ or both ≤ 0 if $x \le y$.

MDP Exercise 18.27. Follow the outline of MDP Example 18.25:

$$\begin{aligned}
\sup_{\xi\in L,\,\sigma^2>0} f_{\xi,\sigma^2}(x) &= \sup_{\xi\in L,\,\sigma^2>0} \tfrac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\|x-\xi\|^2/2\sigma^2}\\
= \sup_{\sigma^2>0} \tfrac{1}{(2\pi\sigma^2)^{\frac{p}{2}}} e^{-\|x-P_L x\|^2/2\sigma^2} &= \tfrac{1}{(2\pi\hat{\sigma}_L^2)^{\frac{p}{2}}} e^{-p/2},
\end{aligned}$$

where P_L is the orthogonal projection matrix onto L and $\hat{\sigma}_L^2 = \frac{1}{p}\|x - P_L x\|^2$. Similarly define P_{L_0} and $\hat{\sigma}_{L_0}^2 = \frac{1}{p}\|x - P_{L_0}x\|^2$. Then the LRT statistic

$$\lambda \equiv \frac{\sup_{\xi\in L_0,\, \sigma^2>0} f_{\xi,\sigma^2}(x)}{\sup_{\xi\in L,\, \sigma^2>0} f_{\xi,\sigma^2}(x)}$$

satisfies

$$\lambda^{\frac{2}{p}} = \frac{\hat{\sigma}_L^2}{\hat{\sigma}_{L_0}^2} = \frac{\|x - P_L x\|^2}{\|x - P_{L_0}x\|^2} = \frac{\|(I - P_L)x\|^2}{\|(I - P_L)x\|^2 + \|(P_L - P_{L_0})x\|^2}$$

since $(I - P_L)(P_L - P_{L_0}) = 0$ [verify]. The LRT rejects H_0 if λ is too small, or equivalently if $F \equiv \frac{\|(P_L - P_{L_0})X\|^2}{\|(I - P_L)X\|^2}$ is too large. The numerator and denominator are independent by MDP Exercise 8.5 and are distributed as $\sigma^2\chi^2_{d-d_0}\left(\frac{\|(P_L - P_{L_0})\xi\|^2}{\sigma^2}\right)$ and $\sigma^2\chi^2_{p-d}$, resp., by MDP (8.111). (Note that $P_L - P_{L_0}$ and $I - P$ are projection matrices, and $(I - P)\xi = 0$ since $\xi \in L$.) Thus F has a noncentral F distribution, central when $H_0 : \xi \in L_0$ holds (since $(P_L - P_{L_0})\xi = 0$ in this case).

Comments on CB Exercise 10.31a assigned in 513 HW 18:

In my solution, a key step is to show that as $\min(n_1, n_2) \to \infty$,

$$\begin{aligned} S &\equiv \sqrt{\tfrac{n_1 n_2}{n_1+n_2}}\,(\hat{p}_1 - \hat{p}_2) \\ &\equiv \sqrt{\tfrac{n_2}{n_1+n_2}}\left[\sqrt{n_1}\,(\hat{p}_1 - p)\right] - \sqrt{\tfrac{n_1}{n_1+n_2}}\left[\sqrt{n_2}\,(\hat{p}_2 - p)\right] \xrightarrow{d} N(0,1). \end{aligned} \tag{1}$$

(In (1) it is assumed that n_1 and n_2 depend on a single index n s.t. $\min(n_1(n), n_2(n)) \to \infty$ as $n \to \infty$.) Several students asked if it is necessary to assume that $n_1 = n_2$, or more generally, that the ratio $\frac{n_1}{n_2} \to 1$.

The answer to this is "no". Consider the more general assumption that

$$\frac{n_1}{n_2} \to \rho \quad \text{for some } \rho \in [0, \infty]. \tag{2}$$

Note that the cases $\rho = 0$ and $\rho = \infty$ are included. Because $\hat{p}_1 \perp\!\!\!\perp \hat{p}_2$ and $\sqrt{n_i}\,(\hat{p}_i - p) \xrightarrow{d} N(0,1)$ as $n_i \to \infty$, $i = 1, 2$, under assumption (2) the result (1) will follow from Proposition 1 below with

$$\begin{aligned} Y_n &= (\sqrt{n_1}\,(\hat{p}_1 - p)\,,\ \sqrt{n_2}\,(\hat{p}_2 - p)), \\ Y &= N_2(0, I_2), \\ X_n = c_n &= \left(\sqrt{\tfrac{n_2}{n_1+n_2}}, -\sqrt{\tfrac{n_1}{n_1+n_2}}\right), \quad \text{(constant vectors)}, \\ c &= \left(\sqrt{\tfrac{1}{\rho+1}},\ -\sqrt{\tfrac{\rho}{\rho+1}}\right), \qquad \text{(a constant vector)}. \end{aligned}$$

Note that $\|c_n\|^2 = \|c\|^2 = 1$, so $c^tY \sim N(0,1)$.

Proposition 1. *Suppose that* $X_n \xrightarrow{p} c$ *and* $Y_n \xrightarrow{d} Y$ *as* $n \to \infty$, *where* X_n, Y_n, *and* Y *are random vectors in* $\mathbf{R}^p$ *and* $c \in \mathbf{R}^p$. *Then* $X_n^tY_n \xrightarrow{d} c^tY$.

Proof. Because $h(x, y) \equiv x^ty \le \|x\|\|y\|$, h is continuous in (x, y). Thus the result follows from MDP Theorem 10.6 p.142, re-stated just below. $\square$

Theorem 10.6. *Suppose that* $X_n \xrightarrow{p} c$ *in* $\mathbf{R}^k$ *and* $Y_n \xrightarrow{d} Y$ *in* $\mathbf{R}^l$. *If* $h(x, y)$ *is continuous then* $h(X_n, Y_n) \xrightarrow{d} h(c, Y)$. $\square$

However, *the result (1) does not require (2), i.e., does not require that* $\frac{n_1}{n_2}$ *converge to anything!* This follows from the next result:

Proposition 2. *If* $\|c_n\| \to 1$ *and* $Y_n \xrightarrow{d} N_p(0, I_p)$, *then* $c_n^tY_n \xrightarrow{d} N(0,1)$.

Proof. Assume to the contrary that $c_n^tY_n \overset{d}{\not\to} N(0,1)$. Then $\exists\, A \subseteq \mathbf{R}^1$ s.t. $\Pr[N(0,1) \in \partial A] = 0$ but

$$\Pr[c_n^tY_n \in A] \not\to \Pr[N(0,1) \in A]. \tag{4}$$

Therefore $\exists\, \epsilon > 0$ and a subsequence $\{n'\} \subset \{n\}$ s.t.

$$|\Pr[c_{n'}^tY_{n'} \in A] - \Pr[N(0,1) \in A]| \ge \epsilon \quad \forall n'. \tag{5}$$

However, $\{c_{n'}\}$ is bounded so $\{c_{n'}\}$ contains a convergent subsequence $\{c_{n''}\}$, i.e. $c_{n''} \to c$ for some c s.t. $\|c\| = 1$. Thus by Proposition 1 with n replaced by n'', X_n by $c_{n''}$, Y_n by $Y_{n''}$, and Y by $N_p(0, I_p)$,

$$c_{n''}^tY_{n''} \xrightarrow{d} c^tY \sim N(0,1),$$

which contradicts (5). $\square$

Discussion: A main property of convergence in distribution (aka. weak convergence of probability measures), is the Continuous Mapping Property: if $Y_n \xrightarrow{d} Y$ and h is continuous then $h(Y_n) \xrightarrow{d} h(Y)$ (cf. MDP Corollary 10.1, p.139). The Extended Continuous Mapping Property states conditions on a *sequence* of functions h_n s.t. $Y_n \xrightarrow{d} Y$ and $h_n \to h \Rightarrow h_n(Y_n) \xrightarrow{d} h(Y)$; cf. Billingsley (1968), Thm.5.5, p.34. However, as noted by F. Topsoe (Ann. Math. Statist. (1967), pp.1661-1665), *it is not necessary that the sequence of functions* $\{h_n\}$ *be convergent,* only that (a) $\{h_n\}$ be precompact in an appropriate sense[45] so that subsequences $\{h'_n\}$ contain convergent subsubsequences $\{h''_n\}$, and (b) $h_n(Y) \xrightarrow{d} h(Y)$.

Proposition 2 above is an example of this, with $h_n(y) = c_n^t y$. The condition $\|c_n\| \to 1$ does not require that $c_n \to$ anything. It does imply that (a) $\{c_n\}$ is bounded hence precompact, and (b) $\|c\| = 1$ for any limit point c of $\{c_n\}$, so $c^t Y \sim N(0,1)$, which facts are used in the above proof.

[Thanks to Jon Wellner for a helpful discussion.]

---

[45] e.g. precompact w.r.to the topology of uniform convergence on compact sets.

Solution to Exercise 18.43

For any random variable T, its cdf $F(t) \equiv P[T \le t]$ is nondecreasing, right continuous, and satisfies $F(t-) = P[T < t]$. Similarly, $G(t) \equiv P[T \ge t]$ is nonincreasing, left continuous, and satisfies $G(t+) = P[T > t]$.

Lemma 18.42. (i) *Both* $F(T)$ and $G(T)$ $\succeq_{\text{stoch}}$ $U \equiv \text{Uniform}[0,1]$.
(ii) *If T is a continuous rv, both* $F(T)$ and $G(T)$ $\sim U \equiv \text{Uniform}[0,1]$.

Proof. (i) Since F is nondecreasing, $A(u) \equiv \{t|F(t) \le u\}$ is a semi-infinite interval: either (a) $A(u) = (-\infty, t(u)]$ or (b) $A(u) = (-\infty, t(u))$. If (a),

$$P[F(T) \le u] = P[T \le t(u)] = F(t(u)) \le u$$

because $t(u) \in A(u)$ in this case. If (b), then

$$P[F(T) \le u] = P[T < t(u)] = F(t(u)-) \le u$$

because $t(u) - \epsilon \in A(u)$ for all $\epsilon > 0$ in this case. Thus $F(T) \succeq_{\text{stoch}} U$.

Now define $B(u) \equiv \{t|G(t) \le u\}$; either (a) $B(u) = [t(u), \infty)$ or (b) $B(u) = (t(u), \infty)$. If (a), then

$$P[G(T) \le u] = P[T \ge t(u)] = G(t(u)) \le u$$

because $t(u) \in B(u)$ in this case, while if (b) then

$$P[G(T) \le u] = P[T > t(u)] = G(t(u)+) \le u$$

because $t(u) + \epsilon \in B(u)$ for all $\epsilon > 0$ in this case. Thus $G(T) \succeq_{\text{stoch}} U$.
(ii) By continuity, $F(T) = 1 - G(T) \preceq_{\text{stoch}} 1 - U \sim U$, so $F(T) \sim U$, hence also $G(T) \sim U$. □

Now let T be a test statistic for testing $H_0 : \theta \in \Omega_0$ vs. $H_1 : \theta \in \Omega_1$, where H_0 is rejected for *large* values of T. Set $G_\theta(t) = P_\theta[T \ge t]$. The *p-value* $\equiv$ *attained significance level* associated with T is

$$p \equiv p(T_{\text{obs}}) \equiv \sup\nolimits_{\theta \in \Omega_0} G_\theta(T_{\text{obs}}), \tag{18.112}$$

where $T_{\text{obs}} \equiv T_{\text{observed}} \sim T$. Clearly $p \equiv p(T_{\text{obs}})$ is nonincreasing in T_{obs}. It follows from Lemma 18.42(i) that when $\theta \in \Omega_0$,

$$P_\theta[\, p \le u\,] \le P_\theta[G_\theta(T_{\text{obs}}) \le u] \le u.$$

Thus under H_0, *the p-value is stochastically larger than* $U \equiv$ Uniform[0,1].

Finally, consider the problem of testing $H_0 : \theta \le \theta_0$ vs. $H_1 : \theta > \theta_0$ (θ real) based on a test statistic $T \sim f_\theta$, where $f_\theta(t)$ has monotone likelihood ratio in θ (cf. §18.3). Then by Lemma 18.21, $G_\theta(t)$ is increasing in θ, so $p \equiv p(T_{\text{obs}}) = G_{\theta_0}(T_{\text{obs}})$ and p is stochastically *decreasing* in θ. If T is continuous then $p \sim \text{Uniform}[0,1]$ when $\theta = \theta_0$.

Definition 19.25. Let P, Q be probability measures on a measurable space $(\mathcal{X}, \mathcal{S})$ and let p, q be their corresponding pdfs w.r.to some measure ν. The *Hellinger distance* $H(P, Q)$ is defined via

$$H^2(P,Q) = \tfrac{1}{2}\int(\sqrt{p} - \sqrt{q})^2 d\nu = 1 - \int \sqrt{pq}\, d\nu. \tag{19.61}$$

Clearly $0 \le H(P,Q) \le 1$, $H(P,Q) = 0$ iff $P = Q$, and $H(P,Q) = 1$ iff the supports of P and Q are disjoint. Note that $0 \le \int \sqrt{pq}\, d\nu \le 1$. $\bar{\square}$

MDP Exercise 19.26. *Relations among Hellinger, total variation, and Kullback-Leibler distances.* Show that

(i) $$H^2(P,Q) \le D(P,Q) \le \sqrt{2} H(P,Q),$$

(ii) $$H^2(P,Q) \le \tfrac{1}{2} K(P,Q).$$

These imply that TV-separation is equivalent to H-separation, and that both imply KL-separation.

(iii)*** Does KL-separation imply TV≡H-separation?

Solution. (i) From (19.45)-(19.47) and (19.61),

$$H^2(P,Q) = 1 - \int \sqrt{pq} \le 1 - \int p \wedge q = D(P,Q).$$

Next, by (19.45)-(19.47) and the Cauchy-Schwartz inequality,

$$\begin{aligned} D(P,Q) &= \tfrac{1}{2}\int |p - q| = \tfrac{1}{2}\int |\sqrt{p} + \sqrt{q}|\,|\sqrt{p} - \sqrt{q}| \\ &\le \tfrac{1}{2}[\int |\sqrt{p} + \sqrt{q}|^2]^{1/2}[\int |\sqrt{p} - \sqrt{q}|^2]^{1/2} \\ &= \tfrac{1}{2}\left[\int p + \int q + 2\int \sqrt{pq}\right]^{1/2} \sqrt{2} H(P,Q) \\ &= \tfrac{1}{2}\left[2 + 2\int \sqrt{pq}\right]^{1/2} \sqrt{2} H(P,Q) \\ &\le \sqrt{2} H(P,Q), \end{aligned}$$

since $\int \sqrt{pq} \le 1$.

(ii) $$\begin{aligned} -K(P,Q) &= \int p \log \tfrac{q}{p} = 2\int p \log \sqrt{\tfrac{q}{p}} \\ &= 2\int p \log\left(1 + \sqrt{\tfrac{q}{p}} - 1\right) \\ &\le 2\int p\left(\sqrt{\tfrac{q}{p}} - 1\right) \\ &= 2(\int \sqrt{pq} - 1) = -2H^2(P,Q). \quad \text{QED} \end{aligned}$$

MDP Exercise 19.27. Finite families $\mathcal{P}_0$ and $\mathcal{P}_1$ are finitely distinguishable with an exponential error rate. (This strengthens Proposition 19.14).
Hint: Use Hellinger distance – recall (18.30). $\bar{\square}$

Solution. For convenience, suppose that $\mathcal{P}_0$ and $\mathcal{P}_1$ are parametrized by θ. As in the proof of Proposition 19.14, consider the sequence $\{\phi_n^*\}$ of finite sample size LRTs given in (18.45), that is,

$$\phi_n^*(x_1, \ldots, x_n) = \begin{cases} 0 & \text{if } \lambda_n \equiv \frac{\prod f_{\hat{\theta}_1}(x_i)}{\prod f_{\hat{\theta}_0}(x_i)} \leq 1, \\ 1 & \text{if } \lambda_n \equiv \frac{\prod f_{\hat{\theta}_1}(x)}{\prod f_{\hat{\theta}_0}(x)} > 1, \end{cases} \tag{18.45}$$

where $\hat{\theta}_i$ is the MLE under $\mathcal{P}_i$. Because $\mathcal{P}_0$ and $\mathcal{P}_1$ are finite, uniform consistency of $\{\phi_n^*\}$ again will follow from pointwise consistency; we must demonstrate the exponential rate. As in (18.30), for any $P_0 (\equiv \theta_0) \in \mathcal{P}_0$,

$$\begin{aligned}
P_0[\lambda_n > 1] &= P_0\left[\frac{\prod f_{\hat{\theta}_1}(x_i)}{\prod f_{\hat{\theta}_0}(x_i)} > 1\right] \leq P_0\left[\frac{\prod f_{\hat{\theta}_1}(x_i)}{\prod f_{\theta_0}(x_i)} > 1\right] \\
&\leq \mathrm{E}_0\left\{\left[\frac{\prod f_{\hat{\theta}_1}(x_i)}{\prod f_{\theta_0}(x_i)}\right]^{1/2}\right\} = \int \left[\prod f_{\hat{\theta}_1}(x_i)\right]^{1/2} \left[\prod f_{\theta_0}(x_i)\right]^{1/2} \\
&= \int \max_{\theta_1 \in \mathcal{P}_1} \left[\prod f_{\theta_1}(x_i) \prod f_{\theta_0}(x_i)\right]^{1/2} \\
&\leq \int \sum_{\theta_1 \in \mathcal{P}_1} \left[\prod f_{\theta_1}(x_i) \prod f_{\theta_0}(x_i)\right]^{1/2} \\
&= \sum_{\theta_1 \in \mathcal{P}_1} \rho_{\theta_1}^n \to 0 \quad \text{at an exponential rate,}
\end{aligned}$$

since $\mathcal{P}_1$ is finite and

$$\rho_{\theta_1} = \int [f_{\theta_1}(x_i) f_{\theta_0}(x_i)]^{1/2} = 1 - H^2(f_{\theta_1}, f_{\theta_0}) < 1.$$

Similarly, $P_1[\lambda_n \leq 1] \to 0$ at an exponential rate $\forall P_1 \in \mathcal{P}_1$. $\bar{\square}$