Exercise (1.2)**.** Let $\mathcal{C}$ be a collection of subsets of Ω and let $\Gamma = \{\mathcal{F} : \mathcal{F}$ is a σ-field on Ω and $\mathcal{C} \subset \mathcal{F}\}$. Show that $\Gamma \neq \emptyset$ and $\sigma(\mathcal{C}) = \bigcap_{\mathcal{F}\in\Gamma} \mathcal{F}$.

Proof. Note that the power set 2^Ω consists of all subsets of Ω, and thus lies in Γ. This entails that $\Gamma \neq \emptyset$. By definition, $\sigma(\mathcal{C})$ is the smallest σ-field on Ω containing $\mathcal{C}$, and clearly $\sigma(\mathcal{C}) \in \Gamma$, so $\sigma(\mathcal{C}) \subset \bigcap_{\mathcal{F}\in\Gamma} \mathcal{F}$. On the other hand, it can be seen that $\mathcal{F} \cap \sigma(\mathcal{C})$ is still a member of Γ for every $\mathcal{F} \in \Gamma$, and this allows us to conclude $\sigma(\mathcal{C}) \supset \bigcap_{\mathcal{F}\in\Gamma} \mathcal{F}$. Combining the two inclusion relations completes the proof. □

Exercise (1.4)**.** Let $\mathcal{C}$ be the collection of intervals of the form $(a, b]$, where $-\infty < a < b < \infty$, and let $\mathcal{D}$ be the collection of closed sets on $\mathbb{R}$. Show that $\mathcal{B} = \sigma(\mathcal{C}) = \sigma(\mathcal{D})$, where $\mathcal{B}$ is the Borel σ-field on $\mathbb{R}$.

Proof. Note that for collections $\mathcal{C}_1$ and $\mathcal{C}_2$ of subsets, if $\mathcal{C}_1 \subset \sigma(\mathcal{C}_2)$ and $\mathcal{C}_2 \subset \sigma(\mathcal{C}_1)$, then $\sigma(\mathcal{C}_1) = \sigma(\mathcal{C}_2)$. Since every open set on $\mathbb{R}$ can be expressed as a union of a sequence of finite open intervals, the Borel σ-field $\mathcal{B}$ generated by the collection of open sets can also be generated by the collection of finite open intervals. To see $\mathcal{B} = \sigma(\mathcal{C})$, it suffices that $(a, b) = \bigcup_{n=1}^\infty (a, b - \frac{1}{n}]$ and $(a, b] = \bigcap_{n=1}^\infty (a, b + \frac{1}{n})$, where $-\infty < a < b < \infty$. To see $\mathcal{B} = \sigma(\mathcal{D})$, any open set is the complement of a closed set, and vice versa. □

Exercise (1.8)**.** Let $\{A_n\}$ be a sequence of events on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Define

$$\limsup_n A_n = \bigcap_{n=1}^\infty \bigcup_{i=n}^\infty A_i \quad \text{and} \quad \liminf_n A_n = \bigcup_{n=1}^\infty \bigcap_{i=n}^\infty A_i.$$

Show that $\mathbb{P}(\liminf_n A_n) \leq \liminf_n \mathbb{P}(A_n)$ and $\limsup_n \mathbb{P}(A_n) \leq \mathbb{P}(\limsup_n A_n)$.

Proof. We have $\mathbb{P}(\liminf_n A_n) = \lim_n \mathbb{P}(\bigcap_{i=n}^\infty A_i)$ and $\mathbb{P}(\limsup_n A_n) = \lim_n \mathbb{P}(\bigcup_{i=n}^\infty A_i)$ by the continuity of probability. Then it remains to use $\mathbb{P}(\bigcap_{i=n}^\infty A_i) \leq \mathbb{P}(A_n)$ and $\mathbb{P}(\bigcup_{i=n}^\infty A_i) \geq \mathbb{P}(A_n)$ in these limits. □

Exercise (1.10)**.** Let $F(x_1, \ldots, x_k)$ be a c.d.f. on $\mathbb{R}^k$. Show that

(a) $F(x_1, \ldots, x_{k-1}, x_k) \leq F(x_1, \ldots, x_{k-1}, x'_k)$ if $x_k \leq x'_k$.

(b) $\lim_{x_i\to-\infty} F(x_1, \ldots, x_k) = 0$ for any $1 \leq i \leq k$.

(c) $F(x_1, \ldots, x_{k-1}, \infty) = \lim_{x_k\to\infty} F(x_1, \ldots, x_{k-1}, x_k)$ is a c.d.f. on $\mathbb{R}^{k-1}$.

Proof. (a) $\{X_1 \leq x_1, \ldots, X_{k-1} \leq x_{k-1}, X_k \leq x_k\} \subset \{X_1 \leq x_1, \ldots, X_{k-1} \leq x_{k-1}, X_k \leq x'_k\}$.
(b) $\{X_1 \leq x_1, \ldots, X_k \leq x_k\} \searrow \emptyset$ as $x_i \to -\infty$.
(c) $\{X_1 \leq x_1, \ldots, X_{k-1} \leq x_{k-1}, X_k \leq x_k\} \nearrow \{X_1 \leq x_1, \ldots, X_{k-1} \leq x_{k-1}\}$ as $x_k \to \infty$. □

Exercise (1.24)**.** Let f be an integrable Borel function on $(\Omega, \mathcal{F}, \nu)$. Show that for each $\varepsilon > 0$, there is a δ_ε such that $A \in \mathcal{F}$ and $\nu(A) < \delta_\varepsilon$ imply $\int_A |f| \,\mathrm{d}\nu < \varepsilon$.

Proof. Pick M_ε so that $\int_{\{|f|>M_\varepsilon\}} |f| \,\mathrm{d}\nu < \frac{\varepsilon}{2}$ by the dominated convergence theorem. Then put $\delta_\varepsilon = \frac{\varepsilon}{2M_\varepsilon}$. Once $A \in \mathcal{F}$ satisfies $\nu(A) < \delta_\varepsilon$, we have $\int_A |f| \,\mathrm{d}\nu \leq \int_{\{|f|>M_\varepsilon\}} |f| \,\mathrm{d}\nu + \int_A M_\varepsilon \,\mathrm{d}\nu < \frac{\varepsilon}{2} + M_\varepsilon\delta_\varepsilon = \varepsilon$. □

Exercise (1.28)**.** Show that $\sum_{i=0}^\infty \sum_{j=0}^\infty f(i, j) = \sum_{j=0}^\infty \sum_{i=0}^\infty f(i, j)$ does not hold for

$$f(i, j) = \mathbb{1}_{[i=j]} - \mathbb{1}_{[i=j-1]}.$$

Does this contradict Fubini's theorem?

Proof. Summing $\sum_j f(i, j) = 0$ over i yields 0, while summing $\sum_i f(i, j) = \mathbb{1}_{[j=0]}$ over j yields 1. They are not equal because f is not summable and has varying signs, and hence Fubini's theorem cannot work. □

Exercise (1.40)**.** Let X be a random variable having a continuous c.d.f. F. Show that $Y = F(X)$ has the uniform distribution on $(0, 1)$.

Proof. For $p \in [0,1]$, let $F^{-1}(p) = \inf\{x \in \mathbb{R} : F(x) \geq p\}$, then $\mathbb{P}\{F(X) \leq p\} = \mathbb{P}\{X \leq F^{-1}(p)\} = p$. □

Exercise (1.43)**.** Let $X \sim \mathcal{N}_k(\mu, \Sigma)$ with a positive definite Σ.

(a) Let $Y = AX + c$, where A is an $l \times k$ matrix of rank $l \leq k$ and $c \in \mathbb{R}^l$. Show that $Y \sim \mathcal{N}_l(A\mu + c, A\Sigma A^\top)$.

(b) Show that the components of X are independent if and only if Σ is a diagonal matrix.

(c) Let Λ be positive definite and $Y \sim \mathcal{N}_m(\eta, \Lambda)$ be independent of X. Show that $(X, Y) \sim \mathcal{N}_{k+m}((\mu, \eta), D)$, where D is a block diagonal matrix whose two diagonal blocks are Σ and Λ.

Proof. (a) It's straightforward that for $u \in \mathbb{R}^l$,

$$\begin{aligned}\mathbb{E}\exp(u^\top Y) &= \mathbb{E}\exp((A^\top u)^\top X)\exp(u^\top c)\\ &= \exp((A^\top u)^\top \mu + \tfrac{1}{2}(A^\top u)^\top \Sigma (A^\top u))\exp(u^\top c)\\ &= \exp(u^\top (A\mu + c) + \tfrac{1}{2}u^\top (A\Sigma A^\top)u).\end{aligned}$$

(b) If the components of X are independent, then the off-diagonal elements of Σ representing pairwise covariances of the components of X must vanish. As for the converse, writing $X = (X_1, \ldots, X_k)$ and $\mu = (\mu_1, \ldots, \mu_k)$, we have

$$\mathbb{E}\exp(t\cdot X) = \exp\left\{\sum_{i=1}^k \left(t_i\mu_i + \frac{1}{2}\sigma_i^2 t_i^2\right)\right\} = \prod_{i=1}^k \exp\left(t_i\mu_i + \frac{1}{2}\sigma_i^2 t_i^2\right) = \prod_{i=1}^k \mathbb{E}\exp(t_i X_i), \quad \forall t = (t_1, \ldots, t_k),$$

provided that $\Sigma = \operatorname{diag}(\sigma_1^2, \ldots, \sigma_k^2)$, and thus the independence of X_i's is guaranteed by Kac's theorem (see, e.g., https://math.stackexchange.com/a/287321).

(c) By the independence of X and Y, one can see that

$$\begin{aligned}\mathbb{E}\exp((t,s)\cdot(X,Y)) &= \mathbb{E}\exp(t\cdot X)\,\mathbb{E}\exp(s\cdot Y)\\ &= \exp(t\cdot\mu + \tfrac{1}{2}t\cdot\Sigma t)\exp(s\cdot\eta + \tfrac{1}{2}s\cdot\Lambda s)\\ &= \exp((t,s)\cdot(\mu,\eta) + \tfrac{1}{2}(t,s)\cdot D(t,s))\end{aligned}$$

for $t \in \mathbb{R}^k$ and $s \in \mathbb{R}^m$. □

Exercise (1.46)**.** Let X_1 and X_2 be independent random variables having the standard normal distribution. Obtain the joint Lebesgue p.d.f. of (Y_1, Y_2), where $Y_1 = \sqrt{X_1^2 + X_2^2}$ and $Y_2 = X_1/X_2$. Are Y_i's independent?

Solution. The Jacobian of the mapping $(x_1, x_2) \mapsto (y_1, y_2) = (\sqrt{x_1^2 + x_2^2}, x_1/x_2)$ is

$$\det\left(\frac{\partial y_i}{\partial x_j}\right)_{1\leq i,j\leq 2} = \det\begin{pmatrix} x_1/\sqrt{x_1^2+x_2^2} & x_2/\sqrt{x_1^2+x_2^2} \\ 1/x_2 & -x_1/x_2^2\end{pmatrix} = -\frac{\sqrt{x_1^2+x_2^2}}{x_2^2}.$$

Noticing that $(x_1, x_2) \mapsto (y_1, y_2)$ is two-to-one a.e., substituting $x_1^2 = y_1^2 y_2^2/(1+y_2^2)$ and $x_2^2 = y_1^2/(1+y_2^2)$ gives the p.d.f.

$$f_{Y_1,Y_2}(y_1,y_2) = 2\left|\det\left(\frac{\partial x_j}{\partial y_i}\right)\right| f_{X_1,X_2}(x_1,x_2) = 2\frac{x_2^2}{\sqrt{x_1^2+x_2^2}}\frac{1}{2\pi}\exp\left(-\frac{x_1^2+x_2^2}{2}\right) = \underbrace{\frac{1}{\pi}\frac{1}{1+y_2^2}}_{=f_{Y_2}(y_2)}\cdot\underbrace{y_1\exp\left(-\frac{y_1^2}{2}\right)}_{=f_{Y_1}(y_1)}$$

for $(y_1, y_2) \in (0,\infty)\times\mathbb{R}$, while it's clear that $f_{Y_1,Y_2}(y_1,y_2) = 0$ otherwise. Thus, Y_i's are independent. ////

Exercise (1.52)**.** Let $X_1, \dots, X_n$ be independent and $X_i \sim \mathcal{N}(0, \sigma_i^2)$, $i = 1, \dots, n$. Let

$$\tilde{X} = \frac{\sum_{i=1}^n \sigma_i^{-2} X_i}{\sum_{i=1}^n \sigma_i^{-2}} \quad \text{and} \quad \tilde{S}^2 = \sum_{i=1}^n \sigma_i^{-2}(X_i - \tilde{X})^2.$$

Apply Cochran's theorem to show that $\tilde{X}^2$ and $\tilde{S}^2$ are independent and that $\tilde{S}^2 \sim \chi^2_{n-1}$.

Proof. Let $Z = (\sigma_1^{-1}X_1, \dots, \sigma_n^{-1}X_n) \sim \mathcal{N}_n(0, I_n)$ and write the weight vector as $w = (\sigma_1^{-1}, \dots, \sigma_n^{-1})$. Then, $\tilde{X} = w^\top Z / w^\top w = Z^\top w / w^\top w$, and $\tilde{X}^2 = Z^\top w w^\top Z / (w^\top w)^2 = \frac{1}{w^\top w} Z^\top \frac{ww^\top}{w^\top w} Z$. Also,

$$\tilde{S}^2 = \sum_{i=1}^n \sigma_i^{-2}(X_i^2 + \tilde{X}^2 - 2X_i\tilde{X}) = Z^\top Z + w^\top w \frac{Z^\top w w^\top Z}{(w^\top w)^2} - 2Z^\top w \frac{w^\top Z}{w^\top w} = Z^\top \left(I_n - \frac{ww^\top}{w^\top w} \right) Z.$$

In view of Cochran's theorem, we just point out that $\frac{ww^\top}{w^\top w}$ and $I_n - \frac{ww^\top}{w^\top w}$ are orthogonal projections onto $\operatorname{span}(w)$ and $\operatorname{span}(w)^\perp$, respectively, and $\operatorname{tr}(I_n - \frac{ww^\top}{w^\top w}) = n - 1$. This completes the proof. □

Exercise (1.53)**.** Let $X \sim \mathcal{N}_n(\mu, I_n)$ and A_i be an $n \times n$ symmetric matrix satisfying $A_i^2 = A_i$, $i = 1, 2$. Show that a necessary and sufficient condition that $X^\top A_1 X$ and $X^\top A_2 X$ are independent is $A_1 A_2 = 0$.

Remark. The idempotence of A_i's is a redundant condition. See, e.g., `https://zhuanlan.zhihu.com/p/85314322`.

Proof. For sufficiency, we only need $I_n - A_1 - A_2$ to be idempotent by Cochran's theorem, which does hold since

$$(I_n - A_1 - A_2)^2 - (I_n - A_1 - A_2) = -(A_1 - A_1^2 - A_1A_2) - (A_2 - A_2A_1 - A_2^2) = A_1A_2 + A_2A_1$$

and $A_2A_1 = A_2^\top A_1^\top = (A_1A_2)^\top$. For necessity, summation of independent $X^\top A_1 X \sim \chi^2_{\operatorname{tr}(A_1)}(\mu^\top A_1 \mu)$ and $X^\top A_2 X \sim \chi^2_{\operatorname{tr}(A_2)}(\mu^\top A_2 \mu)$ gives $X^\top (A_1 + A_2) X \sim \chi^2_{\operatorname{tr}(A_1+A_2)}(\mu^\top (A_1 + A_2)\mu)$, and thus $A_1 + A_2$ is idempotent (see, e.g., `https://zhuanlan.zhihu.com/p/85314314`). It follows that

$$0 = (A_1 + A_2)^2 - (A_1 + A_2) = A_1A_2 + A_2A_1 + A_1^2 + A_2^2 - A_1 - A_2 = A_1A_2 + A_2A_1.$$

But $A_1A_2 = A_1^2 A_2 = -A_1A_2A_1$ is symmetric, i.e., $A_1A_2 = A_2^\top A_1^\top = A_2A_1$. Hence, we obtain $A_1A_2 = 0$. □

Exercise (1.58)**.** Let $X \sim \mathcal{N}_k(\mu, \Sigma)$ with a positive definite Σ.

(a) Show that $\mathbb{E}X = \mu$ and $\operatorname{Var}(X) = \Sigma$.

(b) Let A be an $l \times k$ matrix and B be an $m \times k$ matrix. Show that AX and BX are independent if and only if $A\Sigma B^\top = 0$.

(c) Suppose that $k = 2$, $X = (X_1, X_2)$, $\mu = 0$, $\operatorname{Var}(X_1) = \operatorname{Var}(X_2) = 1$, and $\operatorname{Cov}(X_1, X_2) = \rho$. Show that $\mathbb{E}\max\{X_1, X_2\} = \sqrt{(1-\rho)/\pi}$.

Proof. (a) Since $\mathbb{E}\exp(t^\top X) = \exp(t^\top \mu + \frac{1}{2} t^\top \Sigma t) = 1 + t^\top \mu + \frac{1}{2} t^\top (\Sigma + \mu\mu^\top) t + o(\|t\|^2)$ for $t \in \mathbb{R}^k$, we have $\mathbb{E}X = \frac{\partial}{\partial t}\big|_{t=0} \mathbb{E}\exp(t^\top X) = \mu$ and $\operatorname{Var}(X) = \mathbb{E}XX^\top - \mathbb{E}X\mathbb{E}X^\top = \frac{\partial^2}{\partial t \partial t^\top}\big|_{t=0} \mathbb{E}\exp(t^\top X) - \mu\mu^\top = \Sigma$.

(b) Since the linear transforms AX and BX of X are jointly normally distributed, they are independent if and only if $\operatorname{Cov}(AX, BX) = A\operatorname{Var}(X)B^\top = A\Sigma B^\top$ vanishes.

(c) Note that $X_1 - X_2 \sim \mathcal{N}(0, 2 - 2\rho)$ and $Z \sim \mathcal{N}(0, 1)$ satisfies that

$$\mathbb{E}|Z| = 2\int_0^\infty z \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \mathrm{d}z = \sqrt{\frac{2}{\pi}}.$$

Therefore, $\mathbb{E}\max\{X_1, X_2\} = \mathbb{E}(X_1 + X_2 + |X_1 - X_2|)/2 = \mathbb{E}\sqrt{2 - 2\rho}|Z|/2 = \sqrt{(1-\rho)/\pi}$. □

Exercise (1.71)**.** Let ϕ be a ch.f. on $\mathbb{R}^k$. Show that $|\phi| \le 1$ and ϕ is uniformly continuous on $\mathbb{R}^k$.

Proof. Suppose that $\phi(t) = \mathbb{E}\exp(\sqrt{-1}t^\top X)$ for a random vector X. Clearly $|\phi(t)| \le \mathbb{E}|\exp(\sqrt{-1}t^\top X)|$, where $|\exp(\sqrt{-1}t^\top X)| = 1$, and thus $|\phi(t)| \le 1$. As $t - s \to 0$, we have

$$|\phi(t) - \phi(s)| \le \mathbb{E}|\exp(\sqrt{-1}(t-s)^\top X) - 1||\exp(\sqrt{-1}s^\top X)| = \mathbb{E}|\exp(\sqrt{-1}(t-s)^\top X) - 1| \to 0$$

by the bounded convergence theorem. □

Exercise (1.79)**.** Find an example of two random variables X and Y such that X and Y are not independent but their ch.f.'s satisfy $\phi_X(t)\phi_Y(t) = \phi_{X+Y}(t)$ for all $t \in \mathbb{R}$.

Solution. Let $X = Y$ be a Cauchy random variable with scale $\sigma > 0$, whose p.d.f. is $f(x) = \frac{1}{\pi\sigma}\frac{1}{1+(x/\sigma)^2}$ w.r.t. the Lebesgue measure on $\mathbb{R}$. Then their ch.f.'s are both $\phi(t) = \exp(-\sigma|t|)$. One can also see that $X + Y$ is a Cauchy random variable with scale 2σ, and its ch.f. is $\phi_{X+Y}(t) = \exp(-2\sigma|t|) = \phi(t)^2$. ////

Exercise (1.82)**.** Let (X_1, X_2) be $\mathcal{N}_k(\mu, \Sigma)$ with a $k \times k$ positive definite $\Sigma = \begin{pmatrix}\Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}$, where X_1 is a random l-vector and Σ_{11} is an $l \times l$ matrix. Show that the conditional distribution of X_2 given $X_1 = x_1$ is

$$\mathcal{N}_{k-l}(\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(x_1 - \mu_1), \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}),$$

where $\mu_i = \mathbb{E}X_i$, $i = 1, 2$. (Hint: consider $X_2 - \mu_2 - \Sigma_{21}\Sigma_{11}^{-1}(X_1 - \mu_1)$ and $X_1 - \mu_1$.)

Proof. The key trick is to find some $(k-l) \times l$ matrix B such that $X_2 - BX_1$ is independent of X_1. Since $(X_1, X_2 - BX_1)$ is normally distributed, it suffices that $\mathrm{Cov}(X_2 - BX_1, X_1) = \Sigma_{21} - B\Sigma_{11}$ vanishes, i.e., $B = \Sigma_{21}\Sigma_{11}^{-1}$. Then $\mathrm{Var}(X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1) = \mathrm{Cov}(X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1, X_2) = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$, which is the Schur complement of Σ_{11} in Σ. Thus, the conditional distribution of $X_2 = (X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1) + \Sigma_{21}\Sigma_{11}^{-1}X_1$ given $X_1 = x_1$ is $\mathcal{N}_{k-l}(\mu_2 - \Sigma_{21}\Sigma_{11}^{-1}\mu_1, \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}) + \Sigma_{21}\Sigma_{11}^{-1}x_1$, as desired. □

Exercise (1.86)**.** Let X and Y be integrable random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{A} \subset \mathcal{F}$ be a σ-field. Show that $\mathbb{E}[Y\mathbb{E}(X|\mathcal{A})] = \mathbb{E}[X\mathbb{E}(Y|\mathcal{A})]$, assuming that both integrals exist.

Proof. Since both $\mathbb{E}[\bullet]$ and $\mathbb{E}[\bullet|\mathcal{A}]$ are linear operators, and we may write $X = X^+ - X^-$ and $Y = Y^+ - Y^-$, it suffices to consider the case that both X and Y are non-negative. Then, the monotone convergence theorem applies to $X = \lim_{m\to\infty}\min(X, m)$ and $Y = \lim_{n\to\infty}\min(Y, n)$, so we can reduce our attention to bounded random variables. What remains to do is plugging $\mathbb{E}[Y\mathbb{E}(X|\mathcal{A})|\mathcal{A}] = \mathbb{E}(X|\mathcal{A})\mathbb{E}(Y|\mathcal{A}) = \mathbb{E}[X\mathbb{E}(Y|\mathcal{A})|\mathcal{A}]$ into the right-hand side of $\mathbb{E}[\bullet] = \mathbb{E}[\mathbb{E}(\bullet|\mathcal{A})]$. □

Exercise (1.99)**.** Let $X_1, X_2, \ldots$ be i.i.d. random variables and Y be a discrete random variable taking positive integer values. Assume that Y and X_i's are independent. Let $Z = \sum_{i=1}^Y X_i$.
(a) Obtain the ch.f. of Z.
(b) Show that $\mathbb{E}Z = \mathbb{E}Y\mathbb{E}X_1$.
(c) Show that $\mathrm{Var}(Z) = \mathbb{E}Y\,\mathrm{Var}(X_1) + \mathrm{Var}(Y)(\mathbb{E}X_1)^2$.

Proof. (a) $\phi_Z(t) = \mathbb{E}\exp(\sqrt{-1}tZ) = \mathbb{E}\mathbb{E}[\exp(\sqrt{-1}tZ)|Y] = \mathbb{E}[\prod_{i=1}^Y \mathbb{E}\exp(\sqrt{-1}tX_i)] = \mathbb{E}[\phi_{X_1}(t)^Y]$.
(b) $\mathbb{E}Z = \mathbb{E}\mathbb{E}[Z|Y] = \mathbb{E}[Y\mathbb{E}X_1] = \mathbb{E}Y\mathbb{E}X_1$.
(c) $\mathrm{Var}(Z) = \mathbb{E}\,\mathrm{Var}(Z|Y) + \mathrm{Var}(\mathbb{E}[Z|Y]) = \mathbb{E}[Y\,\mathrm{Var}(X_1)] + \mathrm{Var}(Y\mathbb{E}X_1) = \mathbb{E}Y\,\mathrm{Var}(X_1) + \mathrm{Var}(Y)(\mathbb{E}X_1)^2$. □

Exercise (1.127)**.** Let $X, X_1, X_2, \ldots, Y_1, Y_2, \ldots, Z_1, Z_2, \ldots$ be random variables. Prove:
(a) If $X_n \xrightarrow{d} X$, then $X_n = O_{\mathbb{P}}(1)$.
(b) If $X_n = O_{\mathbb{P}}(Z_n)$ and $\mathbb{P}(Y_n = 0) = 0$, then $X_nY_n = O_{\mathbb{P}}(Y_nZ_n)$.
(c) If $X_n = O_{\mathbb{P}}(Z_n)$ and $Y_n = O_{\mathbb{P}}(Z_n)$, then $X_n + Y_n = O_{\mathbb{P}}(Z_n)$.

(d) If $\mathbb{E}|X_n| = O(a_n)$, then $X_n = O_{\mathbb{P}}(a_n)$, where $a_n \in (0, \infty)$.
(e) If $X_n \xrightarrow{\text{a.s.}} X$, then $\sup_n |X_n| = O_{\mathbb{P}}(1)$.

Proof. In what follows, pick $\varepsilon > 0$ at first.

(a) Let $M > 0$ be a continuous point of the c.d.f. of $|X|$ such that $\mathbb{P}(|X| > M) < \varepsilon/2$. By the continuous mapping theorem, $|X_n| \xrightarrow{d} |X|$ and thus $\mathbb{P}(|X_n| > M) \to \mathbb{P}(|X| > M)$. There exists some N such that $\sup_{n>N} \mathbb{P}(|X_n| > M) < \mathbb{P}(|X| > M) + \varepsilon/2 < \varepsilon$. Also, we can choose $M' > M$ satisfying that $\max_{n \le N} \mathbb{P}(|X_n| > M') < \varepsilon$, which implies that $\sup_n \mathbb{P}(|X_n| > M') < \varepsilon$.

(b) Choose $M > 0$ such that $\sup_n \mathbb{P}(|X_n| > M|Z_n|) < \varepsilon$. Then we have $\sup_n \mathbb{P}(|X_nY_n| > M|Y_nZ_n|) < \varepsilon$ since $\{|X_nY_n| > M|Y_nZ_n|\} \subset \{|X_n| > M|Z_n|\} \setminus \{Y_n = 0\} \subset \{|X_n| > M|Z_n|\}$.

(c) Choose $M_1, M_2 > 0$ such that $\sup_n \mathbb{P}(|X_n| > M_1|Z_n|) < \varepsilon/2$ and $\sup_n \mathbb{P}(|Y_n| > M_2|Z_n|) < \varepsilon/2$. Let $M = M_1 + M_2$. Taking complements, $\{|X_n + Y_n| > M|Z_n|\} \subset \{|X_n| > M_1|Z_n|\} \cup \{|Y_n| > M_2|Z_n|\}$. Therefore, $\sup_n \mathbb{P}(|X_n + Y_n| > M|Z_n|) \le \sup_n[\mathbb{P}(|X_n| > M_1|Z_n|) + \mathbb{P}(|Y_n| > M_2|Z_n|)] < \varepsilon$.

(d) Suppose that $\mathbb{E}|X_n| < Ca_n$ for a constant $C > 0$. Let $M = C/\varepsilon$. By Markov's inequality, $\mathbb{P}(|X_n| > Ma_n) \le \mathbb{E}|X_n|/(Ma_n) < C/M = \varepsilon$ for all n.

(e) Choose $M > 0$ such that $\mathbb{P}(|X| > M) < \varepsilon/2$. Since $\mathbb{P}(\bigcup_{k>n}\{|X_k - X| > M\}) \to 0$, there exists some N so that $\mathbb{P}(\bigcup_{k>N}\{|X_k - X| > M\}) < \varepsilon/2$. This leads to $\mathbb{P}(\sup_{k>N} |X_k| > 2M) < \varepsilon/2 + \varepsilon/2 = \varepsilon$, using $\{\sup_{k>N} |X_k| > 2M\} \subset \{|X| > M\} \cup \bigcup_{k>N}\{|X_k - X| > M\}$. To complete the proof, note that $\{X_k\}_{k \le N}$ is also bounded in probability, and $\sup_n |X_n| = \max(|X_1|, \ldots, |X_N|, \sup_{k>N} |X_k|)$. □

Exercise (1.129)**.** Let $\{F_n\}$ be a sequence of c.d.f.'s on $\mathbb{R}$, $G_n(x) = F_n(a_nx+c_n)$, and $H_n(x) = F_n(b_nx+d_n)$, where $\{a_n\}$ and $\{b_n\}$ are sequences of positive numbers and $\{c_n\}$ and $\{d_n\}$ are sequences of real numbers. Suppose that $G_n \to_w G$ and $H_n \to_w H$, where G and H are nondegenerate c.d.f.'s. Show that $a_n/b_n \to a > 0$, $(c_n - d_n)/b_n \to b \in \mathbb{R}$, and $H(ax + b) = G(x)$ for all $x \in \mathbb{R}$.

Proof. Write $\alpha_n = a_n/b_n$ and $\beta_n = (c_n - d_n)/b_n$, then $G_n(x) = H_n(\alpha_nx + \beta_n)$. Pick a subsequence n_k such that $\alpha_{n_k} \xrightarrow[(k\to\infty)]{} a \in [0, \infty]$ and a further subsequence n_{k_j} such that $\beta_{n_{k_j}} \xrightarrow[(j\to\infty)]{} b \in [-\infty, \infty]$.

- Assume $a = 0$ and $|b| < \infty$ for contradiction.

  Pick $\delta, \varepsilon > 0$ such that H is continuous at $b - \delta$ and $b + \varepsilon$. For all $x \in \mathbb{R}$ at which G is continuous,

  $$H_{n_{k_j}}(b - \delta) \le G_{n_{k_j}}(x) \le H_{n_{k_j}}(b + \varepsilon)$$

  when $j \gg 1$. Letting $j \to \infty$ gives $H(b - \delta) \le G(x) \le H(b + \varepsilon)$. Since the continuous points of any monotone function are dense,

  $$H(b-) \le G(-\infty) \le G(\infty) \le H(b),$$

  so $H(b-) = 0$ and $H(b) = 1$. This means H is degenerate.

- Assume $a < \infty$ and $|b| = \infty$ for contradiction.

  - Consider the case when $b = \infty$.
    If G is continuous at $x \in \mathbb{R}$ and H is continuous at $M > 0$, then $G_{n_{k_j}}(x) \ge H_{n_{k_j}}(M)$ for $j \gg 1$. Letting $j \to \infty$ gives $G(x) \ge H(M)$. Since the continuous points of any monotone function are dense, we obtain $G(-\infty) \ge H(\infty) = 1$, which is absurd.
  - Consider the case when $b = -\infty$.
    Similar arguments yield $G(\infty) \le H(-\infty) = 0$.

- Assume $a = \infty$ for contradiction.

  Now that $H_n(x) = G_n(\alpha_n^{-1}x + (d_n - c_n)/a_n)$, the above results prevent $a^{-1} = \lim_{k\to\infty} \alpha_{n_k}^{-1}$ to vanish.

We have shown $a \in (0, \infty)$ and $b \in \mathbb{R}$. Let $\mathcal{C}$ consist of those x such that G is continuous at x and H is continuous at $ax + b$. Clearly $\mathbb{R} \setminus \mathcal{C}$ is at most countable. For any $x \in \mathcal{C}$, we have

$$G(x) = \lim_{j\to\infty} G_{n_{k_j}}(x) \begin{cases} \leq \inf\limits_{y\in\mathcal{C}\cap(x,\infty)} \lim\limits_{j\to\infty} H_{n_{k_j}}(ay+b) = \inf\limits_{y\in\mathcal{C}\cap(x,\infty)} H(ay+b) = H(ax+b) \\ \geq \sup\limits_{y\in\mathcal{C}\cap(-\infty,x)} \lim\limits_{j\to\infty} H_{n_{k_j}}(ay+b) = \sup\limits_{y\in\mathcal{C}\cap(-\infty,x)} H(ay+b) = H(ax+b) \end{cases}$$

and thus $G(x) = H(ax + b)$ for all $x \in \mathbb{R}$ by monotonicity and right-continuity. Henceforth, a and b are uniquely determined. Taking n_{k_j} to approach lim sup and lim inf completes the proof. □

Exercise (1.136)**.** Let $X_1, X_2, \ldots$ be independent random variables with $\mathbb{P}(X_n = \pm 2^{-n}) = 1/2$, $n = 1, 2, \cdots$. Show that $\sum_{i=1}^n X_i \xrightarrow{d} U \sim \text{Uniform}(-1, 1)$.

Proof. In view of Lévy's continuity theorem, we just compute ch.f.'s.
• $\phi_{X_n}(t) = (\mathrm{e}^{\sqrt{-1}t\cdot 2^{-n}} + \mathrm{e}^{-\sqrt{-1}t\cdot 2^{-n}})/2 = \cos(2^{-n}t) \implies \phi_{\sum_{i=1}^n X_i}(t) = \prod_{i=1}^n \phi_{X_i}(t) = \sin(t)/(2^n \sin(2^{-n}t))$.
• $\phi_U(t) = \frac{1}{2}\int_{-1}^1 \mathrm{e}^{\sqrt{-1}tu}\,\mathrm{d}u = (\mathrm{e}^{\sqrt{-1}t} - \mathrm{e}^{-\sqrt{-1}t})/(2\sqrt{-1}t) = \sin(t)/t$.
It's clear that $\phi_{\sum_{i=1}^n X_i} \to \phi_U$ since $2^n \sin(2^{-n}t) \to t$. □

Exercise (1.150)**.** Let $X_1, X_2, \ldots$ be i.i.d. random variables satisfying $\mathbb{P}(X_1 = 2^j) = 2^{-j}$, $j = 1, 2, \cdots$. Show that the WLLN does not hold for $\{X_n\}$, i.e., $\frac{1}{n}\sum_{i=1}^n X_i - a_n \xrightarrow{\mathbb{P}} 0$ does not hold for any sequence of real numbers $\{a_n\}$.

Proof. By Theorem 1.13 in Shao's book, it's equivalent to disprove $n\mathbb{P}(|X_1| > n) \to 0$. To see this, it holds for $n = 2^k$ that $2^k\mathbb{P}(|X_1| > 2^k) = 2^k \sum_{j>k} 2^{-j} = 1$. □

Exercise (1.155)**.** Let $\{X_n\}$ be a sequence of random variables and let $\bar{X} = \sum_{i=1}^n X_i/n$.
(a) Show that if $X_n \xrightarrow{\text{a.s.}} 0$, then $\bar{X} \xrightarrow{\text{a.s.}} 0$.
(b) Show that if $X_n \xrightarrow{L^r} 0$, then $\bar{X} \xrightarrow{L^r} 0$, where $r \geq 1$ is a constant.
(c) Show that the result in (b) may not be true for $r \in (0, 1)$.
(d) Show that $X_n \xrightarrow{\mathbb{P}} 0$ may not imply $\bar{X} \xrightarrow{\mathbb{P}} 0$.

Proof. (a) By the Stolz–Cesáro theorem, $\{X_n \to 0\} \subset \{\bar{X} \to 0\}$.

(b) Note that $\|\bar{X}\|_{L^r} \leq \sum_{i=1}^n \|X_i\|_{L^r}/n$.

(c&d) By Markov's inequality, the convergence to zero in L^r implies the convergence to zero in probability. Hence, it suffices to find an example that $\mathbb{E}|X_n|^r \to 0$ whereas the limit distribution of $\bar{X}$ exists and is nontrivial. Consider independent $X_1, X_2, \ldots$ such that $\mathbb{P}(X_n = n^a) = \mathbb{P}(X_n = -n^a) = \frac{1}{2}n^{-b}$ and $\mathbb{P}(X_n = 0) = 1 - n^{-b}$, where $a, b > 0$ are constants to be determined. Since $\mathbb{E}|X_n|^r = n^{ra-b}$, we just require that $ra < b$. Then, we want to show that $\bar{X}$ is asymptotically normal using Lindeberg's CLT. Note that $\sum_{i=1}^n \text{Var}(X_i) = \sum_{i=1}^n i^{2a-b} \sim \frac{1}{2a-b+1}n^{2a-b+1}$, so we simply choose $b = 2a - 1$. Now Lindeberg's condition becomes $\sum_{i=1}^n \mathbb{E}[(X_i/n)^2 \mathbb{1}_{\{|X_i|/n > \varepsilon\}}] \to 0$ for any $\varepsilon > 0$, which certainly holds if $\max_{i\leq n}|X_i|/n < \varepsilon$ for n large enough, and we can require $a < 1$ to fulfill it. In summary, what we need is that $ra < 2a - 1$ for some $a \in (\frac{1}{2}, 1)$. This is true when $r \in (0, 1)$. □

Exercise (1.161)**.** Suppose that $X_n \sim \text{Binomial}(\theta, n)$, where $0 < \theta < 1$, $n = 1, 2, \cdots$. Define $Y_n = \log(X_n/n)$ when $X_n \geq 1$ and $Y_n = 1$ when $X_n = 0$. Show that $Y_n \xrightarrow{\text{a.s.}} \log\theta$ and $\sqrt{n}(Y_n - \log\theta) \xrightarrow{d} \mathcal{N}(0, \frac{1-\theta}{\theta})$. Establish similar results when $X_n \sim \text{Poisson}(n\theta)$.

Proof. First, we show the asymptotics of $\{X_n\}$. Note that $X_n \stackrel{d}{=} \sum_{i=1}^n \xi_i$, where $\xi_1, \xi_2, \ldots \stackrel{\text{i.i.d.}}{\sim} \text{Bernoulli}(\theta)$. Thus, $\mathbb{E}|X_n - n\theta|^4 = \sum_i \mathbb{E}(\xi_i - \theta)^4 + \binom{4}{2}\sum_{j>k}\mathbb{E}[(\xi_j - \theta)^2(\xi_k - \theta)^2] \leq C_\theta n^2$ for a constant $C_\theta > 0$. It follows that $\mathbb{P}(|X_n/n - \theta| \geq \varepsilon) \leq (n\varepsilon)^{-4}\mathbb{E}|X_n - n\theta|^4 \leq C_\theta\varepsilon^{-4}n^{-2}$ are summable for any fixed $\varepsilon > 0$. By the Borel–Cantelli lemma, $\mathbb{P}(\limsup_n\{|X_n/n - \theta| \geq \varepsilon\}) = 0$, which implies that $X_n/n \xrightarrow{\text{a.s.}} \theta$. Clearly

$\{X_n/n \to \theta\} \subset \{Y_n \to \log\theta\}$, so we have $Y_n \xrightarrow{\text{a.s.}} \log\theta$. By the CLT, $(\sum_{i=1}^n \xi_i - n\theta)/\sqrt{n} \xrightarrow{d} \mathcal{N}(0, \theta(1-\theta))$, and so is $\sqrt{n}(X_n/n - \theta)$. Then the delta method together with $\log x = \log\theta + \frac{1}{\theta}(x-\theta) + \cdots$ leads to $\sqrt{n}(Y_n - \log\theta) \xrightarrow{d} \frac{1}{\theta}\mathcal{N}(0, \theta(1-\theta)) = \mathcal{N}(0, \frac{1-\theta}{\theta})$. If the distribution of X_n is replaced with Poisson$(n\theta)$, then the above argument still holds as long as the distribution of ξ_i is replaced with Poisson(θ), in which case the SLLN remains the same while the CLT becomes $\sqrt{n}(X_n/n - \theta) \xrightarrow{d} \mathcal{N}(0, \theta)$ and $\sqrt{n}(Y_n - \log\theta) \xrightarrow{d} \mathcal{N}(0, \frac{1}{\theta})$. □

Exercise (1.162)**.** Let $X_1, X_2, \ldots$ be independent random variables such that $X_j \sim \text{Uniform}(-j, j)$. Show that Lindeberg's condition is satisfied and state the resulting CLT.

Proof. Note that $\mathbb{E}X_j = 0$ and $\text{Var}(X_j) = \mathbb{E}X_j^2 = \frac{1}{2j}\int_{-j}^{j} x^2 \, \mathrm{d}x = \frac{j^2}{3}$. It follows that $\sum_{j=1}^n \text{Var}(X_j) = \frac{n(n+1)(2n+1)}{18} \sim \frac{n^3}{9}$. Let $\sigma_n = \frac{n^{3/2}}{3}$, then Lindeberg's condition reads $\sum_{i=1}^n \mathbb{E}[(X_i/\sigma_n)^2 \mathbb{1}_{\{|X_i|/\sigma_n > \varepsilon\}}] \to 0$ for any fixed $\varepsilon > 0$. It does hold, because $\max_{i \le n} |X_i|/\sigma_n \le \frac{3}{n^{1/2}} < \varepsilon$ for n large enough. As a result, we have the CLT that $\frac{X_1 + \cdots + X_n}{n^{3/2}/3} \xrightarrow{d} \mathcal{N}(0,1)$. □

Exercise (2.3)**.** Show that $\{P_\theta : \theta \in \Theta\}$ is an exponential family and find its canonical form and natural parameter space, when

(a) P_θ is Poisson(θ), $\theta \in \Theta = (0, \infty)$;

(b) P_θ is NegativeBinomial(θ, r) with a fixed r, $\theta \in \Theta = (0,1)$;

(c) P_θ is Exponential(a, θ) with a fixed a, $\theta \in \Theta = (0, \infty)$;

(d) P_θ is Gamma(α, γ), $\theta = (\alpha, \gamma) \in \Theta = (0,\infty) \times (0, \infty)$;

(e) P_θ is Beta(α, β), $\theta = (\alpha, \beta) \in \Theta = (0,1) \times (0,1)$;

(f) P_θ is Weibull(α, θ) with a fixed $\alpha > 0$, $\theta \in \Theta = (0, \infty)$.

Proof. (a) Let ν be the counting measure on $\{0, 1, 2, \cdots\}$, then

$$\frac{\mathrm{d}P_\theta}{\mathrm{d}\nu}(x) = \mathrm{e}^{-\theta}\frac{\theta^x}{x!} = f_\xi(x) = \exp(x\xi - \mathrm{e}^\xi)\frac{1}{x!},$$

where $\xi = \log\theta \in \Xi = \mathbb{R}$.

(b) Let ν be the counting measure on $\{r, r+1, \cdots\}$, then

$$\frac{\mathrm{d}P_\theta}{\mathrm{d}\nu}(x) = \binom{x-1}{r-1}\theta^r(1-\theta)^{x-r} = f_\xi(x) = \exp\{(x-r)\xi + r\log(1-\mathrm{e}^\xi)\}\binom{x-1}{r-1},$$

where $\xi = \log(1-\theta) \in \Xi = (-\infty, 0)$.

(c) Let ν be the Lebesgue measure on (a, ∞), then

$$\frac{\mathrm{d}P_\theta}{\mathrm{d}\nu}(x) = \theta^{-1}\mathrm{e}^{-(x-a)/\theta} = f_\xi(x) = \exp\{(a-x)\xi + \log\xi\},$$

where $\xi = \theta^{-1} \in \Xi = (0, \infty)$.

(d) Let ν be the Lebesgue measure on $(0, \infty)$, then

$$\frac{\mathrm{d}P_\theta}{\mathrm{d}\nu}(x) = \frac{1}{\Gamma(\alpha)\gamma^\alpha}x^{\alpha-1}\mathrm{e}^{-x/\gamma} = f_\xi(x) = \exp\{\xi_1\log x - \xi_2 x - \log\Gamma(\xi_1+1) + (\xi_1+1)\log\xi_2\},$$

where $\xi = (\xi_1, \xi_2) = (\alpha - 1, \gamma^{-1}) \in \Xi = (-1, \infty) \times (0, \infty)$.

(e) Let ν be the Lebesgue measure on $(0,1)$, then

$$\frac{\mathrm{d}P_\theta}{\mathrm{d}\nu}(x) = \frac{1}{\mathrm{B}(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1} = f_\xi(x) = \exp\{\xi_1\log x + \xi_2\log(1-x) - \log\mathrm{B}(\xi_1+1, \xi_2+1)\},$$

where $\xi = (\xi_1, \xi_2) = (\alpha-1, \beta-1) \in \Xi = (-1, 0) \times (-1, 0)$.

(f) Let ν be the Lebesgue measure on $(0,\infty)$, then

$$\frac{\mathrm{d}P_\theta}{\mathrm{d}\nu}(x)=\frac{\alpha}{\theta}x^{\alpha-1}\mathrm{e}^{-x^\alpha/\theta}=f_\xi(x)=\exp\{-x^\alpha\xi+\log(\alpha\xi)\}x^{\alpha-1},$$

where $\xi=\theta^{-1}\in\Xi=(0,\infty)$. □

Exercise (2.4)**.** Show that the family of Exponential(a,θ) with two unknown parameters a and θ is not an exponential family.

Proof. Note that distributions in an exponential family have a common support, whereas the support (a,∞) of Exponential(a,θ) varies with a. □

Exercise (2.14)**.** Let X be a random variable with a p.d.f. f_θ in an exponential family $\{P_\theta:\theta\in\Theta\}$ and let A be a Borel set. Show that the distribution of X truncated on A (i.e., the conditional distribution of X given $X\in A$) has a p.d.f. $f_\theta\mathbb{1}_A/P_\theta(A)$ that is in an exponential family.

Proof. Note that $\mathbb{P}_\theta(X\in B|X\in A)=\mathbb{P}_\theta(X\in B\cap A)/\mathbb{P}_\theta(X\in A)=\int_B f_\theta\mathbb{1}_A\,\mathrm{d}P_\theta/P_\theta(A)$ for any Borel set B. Write $f_\theta(x)=\exp\{\xi(\theta)^\top T(x)-K(\theta)\}h(x)$, then $f_\theta(x)\mathbb{1}_A(x)/P_\theta(A)=\exp\{\xi(\theta)^\top T(x)-\tilde K(\theta)\}\tilde h(x)$ with $\tilde K(\theta)=K(\theta)+\log P_\theta(A)$ and $\tilde h(x)=h(x)\mathbb{1}_A(x)$. □

Exercise (2.20)**.** Let $X_1,\dots,X_n$ be i.i.d. random variables having the exponential distribution Exponential(a,θ), $a\in\mathbb{R}$, and $\theta>0$. Show that the smallest order statistic $X_{(1)}$ has the exponential distribution Exponential$(a,\theta/n)$ and that $2\sum_{i=1}^n(X_i-X_{(1)})/\theta$ has the chi-square distribution χ^2_{2n-2}.

Proof. The distribution of $X_{(1)}$ follows from $\mathbb{P}\{X_{(1)}>x\}=\prod_{i=1}^n\mathbb{P}\{X_i>x\}=\mathrm{e}^{-n(x-a)/\theta}$, $x>a$. Using Exercise 2.24, the joint Lebesgue p.d.f. of $X_{(1)},X_{(2)}-X_{(1)},\dots,X_{(n)}-X_{(1)}$ is

$$\begin{aligned}(x_1,y_2,\dots,y_n)\mapsto\quad & n!\theta^{-1}\mathrm{e}^{-(x_1-a)/\theta}\mathbb{1}_{[x_1>a]}\theta^{-1}\mathrm{e}^{-(x_1+y_2-a)/\theta}\cdots\theta^{-1}\mathrm{e}^{-(x_1+y_n-a)/\theta}\mathbb{1}_{[0<y_2<\cdots<y_n]}\\ &=n\theta^{-1}\mathrm{e}^{-n(x_1-a)/\theta}\mathbb{1}_{[x_1>a]}(n-1)!\theta^{-1}\mathrm{e}^{-y_2/\theta}\cdots\theta^{-1}\mathrm{e}^{-y_n/\theta}\mathbb{1}_{[0<y_2<\cdots<y_n]},\end{aligned}$$

and thus $(X_{(i+1)}-X_{(1)})_{1\le i\le n-1}$ has the same distribution as the order statistics $(Y_{(i)})_{1\le i\le n-1}$ of an i.i.d. sample $Y_1,\dots,Y_{n-1}$ from Exponential$(0,\theta)$. Then $2\sum_{i=1}^n(X_i-X_{(1)})/\theta=2\sum_{i=1}^{n-1}(X_{(i+1)}-X_{(1)})/\theta$ has the same distribution as $2\sum_{i=1}^{n-1}Y_{(i)}/\theta=2\sum_{i=1}^{n-1}Y_i/\theta$. Note that $2Y_i/\theta\sim\chi^2_2$. □

Exercise (2.24)**.** Prove the claims in Example 2.9 for the distributions related to order statistics. Suppose that X_i has a c.d.f. F having a Lebesgue p.d.f. f. Then the joint Lebesgue p.d.f. of $X_{(1)},\dots,X_{(n)}$ is

$$g(x_1,x_2,\dots,x_n)=n!f(x_1)f(x_2)\cdots f(x_n)\mathbb{1}_{[x_1<x_2<\cdots<x_n]}.$$

The joint Lebesgue p.d.f. of $X_{(i)}$ and $X_{(j)}$, $1\le i<j\le n$, is

$$g_{i,j}(x,y)=\frac{n![F(x)]^{i-1}[F(y)-F(x)]^{j-i-1}[1-F(y)]^{n-j}f(x)f(y)}{(i-1)!(j-i-1)!(n-j)!}\mathbb{1}_{[x<y]}$$

and the Lebesgue p.d.f. of $X_{(i)}$ is

$$g_i(x)=\frac{n!}{(i-1)!(n-i)!}[F(x)]^{i-1}[1-F(x)]^{n-i}f(x).$$

Proof. The first equality makes sense because $(x_1,\dots,x_n)\mapsto(x_{(1)},\dots,x_{(n)})$ is $n!$-to-1. Then the next two equalities are consequences of integral, where we may use $\mathbb{P}(\bigcap_{i=1}^n\{a<X_{(i)}<b\})=\mathbb{P}(\bigcap_{i=1}^n\{a<X_i<b\})$ to deduce that $\int_a^b\int_a^{x_n}\cdots\int_a^{x_2}f(x_1)f(x_2)\cdots f(x_n)\,\mathrm{d}x_1\cdots\mathrm{d}x_{n-1}\mathrm{d}x_n=\frac{1}{n!}(\int_a^b f(x)\,\mathrm{d}x)^n$, $\forall-\infty\le a<b\le\infty$, and replace n with $i-1,j-i-1,n-j,n-i$. □

Exercise (2.27)**.** Let $X_1, \dots, X_n$ be i.i.d. random variables from $P_\theta \in \{P_\theta : \theta \in \Theta\}$. In the following cases, find a sufficient statistic for $\theta \in \Theta$ that has the same dimension as θ.
(a) P_θ is the Poisson distribution Poisson(θ), $\theta \in (0, \infty)$.
(b) P_θ is the negative binomial distribution NegativeBinomial(θ, r) with a known r, $\theta \in (0, 1)$.
(f) P_θ is the log-normal distribution LogNormal(μ, σ^2), $\theta = (\mu, \sigma^2) \in \mathbb{R} \times (0, \infty)$.

Solution. By the Fisher–Neyman factorization theorem, $T(X)$ is sufficient if the corresponding p.d.f. has the form $f_\theta(x) = g_\theta(T(x))h(x)$.
(a) $f_\theta(x) = \prod_{i=1}^n (\mathrm{e}^{-\theta} \frac{\theta^{x_i}}{x_i!}) = \mathrm{e}^{-n\theta} \theta^{\sum_{i=1}^n x_i} \prod_{i=1}^n \frac{1}{x_i!} \implies T(X) = \sum_{i=1}^n X_i$.
(b) $f_\theta(x) = \prod_{i=1}^n [\binom{x_i - 1}{r-1} \theta^r (1-\theta)^{x_i - r}] = \theta^{nr}(1-\theta)^{\sum_{i=1}^n x_i - nr} \prod_{i=1}^n \binom{x_i-1}{r-1} \implies T(X) = \sum_{i=1}^n X_i$.
(f) $$\begin{aligned} f_\theta(x) &= \prod_{i=1}^n \left[\tfrac{1}{\sqrt{2\pi\sigma^2}} x_i^{-1} \mathrm{e}^{-(\log x_i - \mu)^2/(2\sigma^2)} \mathbb{1}_{(0,\infty)}(x_i)\right] \\ &= (2\pi\sigma^2)^{-n/2} \mathrm{e}^{-[\sum_{i=1}^n (\log x_i)^2 - 2\mu \sum_{i=1}^n \log x_i + n\mu^2]/(2\sigma^2)} \prod_{i=1}^n [x_i^{-1} \mathbb{1}_{(0,\infty)}(x_i)] \end{aligned} \implies T(X) = \textstyle\sum_{i=1}^n (\log X_i, (\log X_i)^2).$$

////

Exercise (2.29)**.** Let $\phi(\cdot)$ be a positive Borel function on $(\mathbb{R}, \mathcal{B})$ such that $\int_a^b \phi(x)\,\mathrm{d}x < \infty$ for any a and b, $-\infty < a < b < \infty$. Let $\theta = (a, b)$, $\Theta = \{(a, b) \in \mathbb{R}^2 : a < b\}$, and

$$f_\theta(x) = c(\theta)\phi(x)\mathbb{1}_{(a,b)}(x),$$

where $c(\theta) = [\int_a^b \phi(x)\,\mathrm{d}x]^{-1}$. Then $\{f_\theta : \theta \in \Theta\}$, called a truncation family, is a parametric family dominated by the Lebesgue measure on $\mathbb{R}$. Let $X_1, \dots, X_n$ be i.i.d. random variables having the p.d.f. f_θ. Then show that $X_{(1)}$ (or $X_{(n)}$) is sufficient for a (or b) if we consider a subfamily $\{f_{(a,b)} : a < b\}$ with a fixed b (or a).

Proof. Since the p.d.f. of $(X_1, \dots, X_n)$ is $\boldsymbol{x} \mapsto \prod_{i=1}^n f_\theta(x_i) = [c(\theta)]^n \mathbb{1}_{(a,\infty)}(x_{(1)}) \mathbb{1}_{(-\infty,b)}(x_{(n)}) \prod_{i=1}^n \phi(x_i)$, the desired result is easily seen using the Fisher–Neyman factorization theorem. □

Exercise (2.31)**.** Let $X_1, \dots, X_n$ be i.i.d. random variables having a distribution $P \in \mathcal{P}$, where $\mathcal{P}$ is the family of distributions on $\mathbb{R}$ having continuous c.d.f.'s. Let $T = (X_{(1)}, \dots, X_{(n)})$ be the vector of order statistics. Show that, given T, the conditional distribution of $X = (X_1, \dots, X_n)$ is a discrete distribution putting probability $1/n!$ on each of the $n!$ points $(X_{i_1}, \dots, X_{i_n}) \in \mathbb{R}^n$, where $\{i_1, \dots, i_n\}$ is a permutation of $\{1, \cdots, n\}$; hence, T is sufficient for $P \in \mathcal{P}$.

Proof. Since X_j's are i.i.d., $(X_{i_j})_{1 \le j \le n}$ and $(X_j)_{1 \le j \le n}$ have the same distribution. Note that $T(X_{i_1}, \dots, X_{i_n}) = T(X_1, \dots, X_n)$. Thus, the conditional distributions of $(X_{i_j})_{1 \le j \le n}$ and $(X_j)_{1 \le j \le n}$ given T are the same. To make it explicit, $\mathbb{P}(X_{i_j} = x_{(j)}, \forall j \mid T = (x_{(j)})_{1 \le j \le n}) = \mathbb{P}(X_j = x_{(j)}, \forall j \mid T = (x_{(j)})_{1 \le j \le n})$, which characterizes the distribution of $(X_1, \dots, X_n)$ given T. □

Exercise (2.41)**.** Let $X_1, \dots, X_n$ be i.i.d. random variables having a population P in a parametric family indexed by (θ, j), where $\theta \in (0, 1)$, $j = 1, 2$, and $n \ge 2$. When $j = 1$, P is the Poisson distribution Poisson(θ). When $j = 2$, P is the binomial distribution Binomial$(\theta, 1)$.
(a) Show that $T = \sum_{i=1}^n X_i$ is not sufficient for (θ, j).
(b) Find a two-dimensional minimal sufficient statistic for (θ, j).

Solution. Let $W = \mathbb{1}_{[X_{(n)} \le 1]}$. We will show that (T, W) is minimal sufficient for (θ, j). Since W is not a function of T, we then conclude that T is not sufficient. Put $0^0 = 1$. The p.d.f. of $(X_1, \dots, X_n)$ is

$$\begin{aligned} f_{\theta,j} : x = (x_1, \dots, x_n) \mapsto & \prod_{i=1}^n \left[(\mathrm{e}^\theta \theta^{x_i})^{\mathbb{1}_{[j=1]}} \frac{1}{x_i!} \left(\theta^{x_i}(1-\theta)^{1-x_i} \mathbb{1}_{[x_i \le 1]}\right)^{\mathbb{1}_{[j=2]}} \right] \\ &= \mathrm{e}^{n\theta \mathbb{1}_{[j=1]}} \theta^{T(x)} \left((1-\theta)^{n-T(x)} W(x)\right)^{\mathbb{1}_{[j=2]}} \prod_{i=1}^n \frac{1}{x_i!}, \end{aligned}$$

and thus (T, W) is sufficient by the factorization theorem. In order for (T, W) to be minimal, we claim that $(T(x), W(x)) = (T(y), W(y))$ for all x and y such that $f_{\theta,j}(y)/f_{\theta,j}(x)$ does not depend on (θ, j). Note that

$$\frac{f_{\theta,j}(y)}{f_{\theta,j}(x)} = \theta^{T(y)-T(x)}\big((1-\theta)^{T(x)-T(y)}W(y)/W(x)\big)^{\mathbb{1}_{[j=2]}}\prod_{i=1}^{n}\frac{x_i!}{y_i!},$$

where we set $0/0 = 0$ and $\infty^0 = 1$. With $j = 1$ and θ varying, one can see that $T(x) = T(y)$. Then alternating j leads to $W(x) = W(y)$. ////

Exercise (2.49)**.** Let T and S be two statistics such that $S = \psi(T)$ for a measurable ψ. Show that
(a) if T is complete, then S is complete;
(b) if T is complete and sufficient and ψ is one-to-one, then S is complete and sufficient;
(c) the results in (a) and (b) still hold if the completeness is replaced by the bounded completeness.

Proof. (a) Let f be a measurable function such that $\mathbb{E}f(S) = 0$. Then $\mathbb{E}\tilde{f}(T) = 0$, where $\tilde{f} = f \circ \psi$. It follows that $f(S) = \tilde{f}(T) \overset{\text{a.s.}}{=} 0$.

(b) Since ψ is one-to-one, we have $\sigma(S) = \sigma(T)$, and hence $X|S$ and $X|T$ have the same conditional distribution. This makes the sufficiency clear, while the completeness follows from (a).

(c) The argument proceeds verbatim once f is further required to be bounded. □

Exercise (2.59)**.** Let $X_1, \dots, X_n$ be i.i.d. random variables having the exponential distribution Exponential(a, θ).
(a) Show that $\sum_{i=1}^n (X_i - X_{(1)})$ and $X_{(1)}$ are independent for any (a, θ).
(b) Show that $(X_{(n)} - X_{(i)})/(X_{(n)} - X_{(n-1)})$, $i = 1, \cdots, n-2$, are independent of $(X_{(1)}, \sum_{i=1}^n (X_i - X_{(1)}))$.

Proof. Although we may derive joint p.d.f.'s directly as in Exercise 2.20 to prove the independence, it is good practice to utilize Basu's theorem. Let $Y_i = X_i - a \sim$ Exponential$(0, \theta)$ and $Z_i = Y_i/\theta \sim$ Exponential$(0, 1)$.

(a) Treat $\theta \in (0, \infty)$ as fixed and $a \in \mathbb{R}$ as unknown. On one hand, $\sum_{i=1}^n (X_i - X_{(1)}) = \sum_{i=1}^n (Y_i - Y_{(1)})$ is ancillary. On the other hand, $X_{(1)}$ is sufficient and complete. To see this, the joint p.d.f. of X_i's is

$$\boldsymbol{x} = (x_1, \dots, x_n) \mapsto \prod_{i=1}^{n}\big(\theta^{-1}\mathrm{e}^{-(x_i-a)/\theta}\mathbb{1}_{[x_i>a]}\big) = \theta^{-n}\mathrm{e}^{-\sum_{i=1}^n x_i/\theta}\cdot \mathrm{e}^{na/\theta}\mathbb{1}_{[x_{(1)}>a]},$$

so the factorization theorem gives the sufficiency; and for any Borel function φ such that $\mathbb{E}\varphi(X_{(1)}) = 0$, we can use $X_{(1)} \sim$ Exponential$(a, \theta/n)$ shown in Exercise 2.20 to obtain

$$\int_a^\infty \varphi(x)\frac{n}{\theta}\mathrm{e}^{-n(x-a)/\theta}\,\mathrm{d}x = 0 \quad\text{or}\quad \int_a^\infty \varphi(x)\mathrm{e}^{-nx/\theta}\,\mathrm{d}x = 0,$$

whose derivative w.r.t. a should be $\varphi(a)\mathrm{e}^{-na/\theta} = 0$ for almost all a, and therefore φ vanishes a.e. on $\mathbb{R}$ containing the range (a, ∞) of $X_{(1)}$.

(b) Treat $(a, \theta) \in \mathbb{R} \times (0, \infty)$ as unknown. Clearly

$$(X_{(n)} - X_{(i)})/(X_{(n)} - X_{(n-1)}) = (Z_{(n)} - Z_{(i)})/(Z_{(n)} - Z_{(n-1)}), \quad i = 1, \cdots, n-2,$$

are ancillary. Since the joint p.d.f. of X_i's is

$$\boldsymbol{x} = (x_1, \dots, x_n) \mapsto \prod_{i=1}^{n}\big(\theta^{-1}\mathrm{e}^{-(x_i-a)/\theta}\mathbb{1}_{[x_i>a]}\big) = \theta^{-n}\mathrm{e}^{-[\sum_{i=1}^n (x_i - x_{(1)}) + n(x_{(1)}-a)]/\theta}\mathbb{1}_{[x_{(1)}>a]},$$

$(X_{(1)}, \sum_{i=1}^n (X_i - X_{(1)}))$ is sufficient by the factorization theorem. Using (a) and Exercise 2.20, the p.d.f. of $(X_{(1)}, \sum_{i=1}^n (X_i - X_{(1)}))$ is

$$(x, y) \mapsto \frac{n}{\theta}\mathrm{e}^{-n(x-a)/\theta}\mathbb{1}_{(a,\infty)}(x)\cdot\frac{1}{\Gamma(n-1)\theta^{n-1}}y^{n-2}\mathrm{e}^{-y/\theta}\mathbb{1}_{(0,\infty)}(y).$$

For any bivariate Borel function ψ such that $\mathbb{E}\psi(X_{(1)}, \sum_{i=1}^n (X_i - X_{(1)})) = 0$, we have

$$\iint_{(a,\infty)\times(0,\infty)} \psi(x, y)y^{n-2}\mathrm{e}^{-(nx+y)/\theta}\,\mathrm{d}x\mathrm{d}y = 0.$$

Then differentiating w.r.t. a gives

$$\int_0^\infty \psi(a, y)y^{n-2}\mathrm{e}^{-y/\theta}\,\mathrm{d}y = 0 \quad \text{for almost all } a \in \mathbb{R}.$$

By the uniqueness of the Laplace transform, $\psi(a, y)y^{n-2} = 0$ for almost all $y > 0$, provided the above integral equality for all $\theta > 0$. Therefore, $\psi = 0$ a.e. on $\mathbb{R} \times (0, \infty)$, which justifies the completeness of $(X_{(1)}, \sum_{i=1}^n (X_i - X_{(1)}))$. □

Exercise (2.65)**.** Let $X_1, \ldots, X_n$ be i.i.d. random variables having the exponential distribution Exponential$(0, \theta)$, $\theta \in (0, \infty)$. Consider estimating θ with the squared error loss. Calculate the risks of the sample mean $\bar{X}$ and $cX_{(1)}$, where c is a positive constant. Is $\bar{X}$ better than $cX_{(1)}$ for some c?

Solution. Note that $\mathrm{MSE}(\bar{X}) = \mathrm{Var}(\bar{X}) = \frac{1}{n}\mathrm{Var}(X_1) = \frac{1}{n}\theta^2$. Recall that $X_{(1)} \sim$ Exponential$(0, \theta/n)$ by Exercise 2.20. Hence, substituting $y = nx/\theta$,

$$\mathrm{MSE}(cX_{(1)}) = \int_0^\infty (cx - \theta)^2 \frac{n}{\theta}\mathrm{e}^{-nx/\theta}\,\mathrm{d}x = \int_0^\infty \theta^2 \left(\frac{c^2}{n^2}y^2 - 2\frac{c}{n}y + 1\right)\mathrm{e}^{-y}\,\mathrm{d}y = \left(2\frac{c^2}{n^2} - 2\frac{c}{n} + 1\right)\theta^2,$$

where $2\frac{c^2}{n^2} - 2\frac{c}{n} + 1 = 2(\frac{c}{n} - \frac{1}{2})^2 + \frac{1}{2}$. When $n = 1$, $\mathrm{MSE}(\bar{X}) < \mathrm{MSE}(cX_{(1)})$ if and only if $c > 1$. When $n \geq 2$, $\mathrm{MSE}(\bar{X}) \leq \mathrm{MSE}(cX_{(1)})$ for all $c > 0$, with equality holding if and only if $n = 2$ and $c = 1$. ////

Exercise (2.67)**.** Let $X_1, \ldots, X_n$ be i.i.d. random variables having the exponential distribution Exponential$(0, \theta)$, $\theta \in (0, \infty)$. Consider the hypotheses

$$H_0 : \theta \leq \theta_0 \quad \text{versus} \quad H_1 : \theta > \theta_0,$$

where $\theta_0 > 0$ is a fixed constant. Obtain the risk function (in terms of θ) of the test rule $T_c(X) = \mathbb{1}_{(c,\infty)}(\bar{X})$, under the 0-1 loss.

Solution. Note that $n\bar{X} \sim \mathrm{Gamma}(n, \theta)$. Hence, the risk of T_c is given by

$$R(T_c) = \mathbb{P}\{\bar{X} > c\}\mathbb{1}_{[\theta\leq\theta_0]} + \mathbb{P}\{\bar{X} \leq c\}\mathbb{1}_{[\theta>\theta_0]},$$

where

$$\mathbb{P}\{\bar{X} > c\} = \frac{1}{(n-1)!\theta^n}\int_{nc}^\infty x^{n-1}\mathrm{e}^{-x/\theta}\,\mathrm{d}x$$

and $\mathbb{P}\{\bar{X} \leq c\} = 1 - \mathbb{P}\{\bar{X} > c\}$. ////

Exercise (2.104)**.** Let $X_1, \ldots, X_n$ be i.i.d. from the uniform distribution on $(\theta - \frac{1}{2}, \theta + \frac{1}{2})$, where $\theta \in \mathbb{R}$ is unknown. Show that $(X_{(1)} + X_{(n)})/2$ is strongly consistent for θ and also consistent in MSE.

Proof. Since these two kinds of consistency are preserved under addition, it suffices to prove the assertion with $(X_{(1)} + X_{(n)})/2$ replaced by $X_{(1)} + 1/2$ and $X_{(n)} - 1/2$, respectively. Note that the distribution is symmetric about θ, so applying the mapping $x \mapsto 2\theta - x$ will complete the proof once we figure out the consistency of $X_{(n)} - 1/2$. It's clear that $\mathbb{P}\{X_{(n)} - 1/2 < \theta - \varepsilon\} = \prod_{i=1}^n \mathbb{P}\{X_i < \theta + 1/2 - \varepsilon\} = (1 - \varepsilon)^n$ for any $\varepsilon \in (0, 1)$. By the Borel–Cantelli lemma, the convergence of $\sum_{n=1}^\infty \mathbb{P}\{X_{(n)} - 1/2 < \theta - \varepsilon\} = (1 - \varepsilon)/\varepsilon$ leads to the strong consistency of $X_{(n)} - 1/2$ for θ. Also, the p.d.f. of $\theta - (X_{(n)} - 1/2)$ is given by $\varepsilon \mapsto n(1 - \varepsilon)^{n-1}\mathbb{1}_{(0,1)}(\varepsilon)$. Therefore, $\mathrm{MSE}(X_{(n)} - 1/2) = \int_0^1 \varepsilon^2 n(1 - \varepsilon)^{n-1}\,\mathrm{d}\varepsilon = 2/[(n + 2)(n + 1)] \to 0$. □

Exercise (2.110)**.** Let $g_1, g_2, \ldots$ be continuous functions on $(a,b) \subset \mathbb{R}$ such that $g_n \to g$ uniformly in any closed subinterval of (a,b). Let T_n be a consistent estimator of $\theta \in (a,b)$. Show that $g_n(T_n)$ is consistent for $\vartheta = g(\theta)$.

Proof. Note that $|g_n(T_n) - g(\theta)| \le |g_n(T_n) - g(T_n)| + |g(T_n) - g(\theta)|$. By the continuous mapping theorem, $g(T_n) \xrightarrow{\mathbb{P}_\theta} g(\theta)$. For any $\varepsilon > 0$, $\{|g_n(T_n) - g(T_n)| > \varepsilon\} \subset \{|T_n - \theta| > \delta\} \cup \{\|g_n - g\|_{L^\infty([\theta-\delta,\theta+\delta])} > \varepsilon\}$ for a $\delta = \delta_\theta$ such that $[\theta-\delta, \theta+\delta] \subset (a,b)$, and thus $g_n(T_n) - g(T_n) \xrightarrow{\mathbb{P}_\theta} 0$. □

Exercise (2.111)**.** Let $X_1, \ldots, X_n$ be i.i.d. from P with unknown mean $\mu \in \mathbb{R}$ and variance $\sigma^2 > 0$, and let $g(\mu) = 0$ if $\mu \neq 0$ and $g(0) = 1$. Find a consistent estimator of $\vartheta = g(\mu)$.

Solution. Consider the estimator $T = \mathbb{1}_{\{|\bar{X}| < n^{-1/4}\}}$ for $g(\mu) = \mathbb{1}_{[\mu=0]}$, where $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$ is the sample mean. Since both g and T take values in $\{0,1\}$, it suffices to show that $\lim_n \mathbb{P}\{n^{1/4}|\bar{X}| < 1\} = g(\mu)$. If $\mu \neq 0$, we have $\bar{X} \xrightarrow{\mathbb{P}} \mu$ by the weak law of large numbers, and therefore $n^{1/4}|\bar{X}| \xrightarrow{\mathbb{P}} \infty$. If $\mu = 0$, we have $\sqrt{n}\bar{X} \xrightarrow{d} Z \sim \mathcal{N}(0, \sigma^2)$ by the central limit theorem, and therefore $n^{1/4}|\bar{X}| \xrightarrow{d} \lim_n |Z|/n^{1/4} = 0$. ////

Exercise (3.1)**.** Let $X_1, \ldots, X_n$ be i.i.d. binary random variables with $\mathbb{P}(X_i = 1) = p \in (0,1)$.
(a) Find the UMVUE of p^m, $m \le n$.
(b) Find the UMVUE of $\mathbb{P}(X_1 + \cdots + X_m = k)$, where m and k are positive integers $\le n$.
(c) Find the UMVUE of $\mathbb{P}(X_1 + \cdots + X_{n-1} > X_n)$.

Solution. Let $T(x) = \sum_{i=1}^n x_i$, then $T(X)$ is complete and sufficient for p, since the joint p.m.f. of X_i's is

$$x = (x_1, \ldots, x_n) \mapsto p^{T(x)}(1-p)^{n-T(x)} = \exp\left\{T(x)\ln(\tfrac{p}{1-p}) + n\ln(1-p)\right\},$$

which belongs to an exponential family. By the Lehmann–Scheffé theorem, the UMVUE of $g(p)$, if exists, has the form of $h(T(X))$ for a known $h(\cdot)$. One can see that $T(X) \sim \text{Binomial}(p, n)$, and thus

$$\mathbb{E}[h(T(X))] = \sum_{t=0}^n h(t)\binom{n}{t}p^t(1-p)^{n-t} = \sum_{t=0}^n h(t)\binom{n}{t}\theta^t \Big/ (\theta+1)^n,$$

where $\theta = \frac{p}{1-p} \in (0, \infty)$.

(a) It's a special case of (b) with $k = m$, so the UMVUE of p^m is $h_m(T(X))$ such that

$$h_m(t) = \tbinom{n-m}{t-m}/\tbinom{n}{t} = \tfrac{(n-m)!t!}{n!(t-m)!} \text{ if } t \in \{m, m+1, \cdots, n\} \text{ and } h_m(t) = 0 \text{ otherwise.}$$

(b) Note that $\mathbb{P}(X_1 + \cdots + X_m = k) = \binom{m}{k}p^k(1-p)^{m-k} = \binom{m}{k}\theta^k/(\theta+1)^m$, so its UMVUE is $h_{m,k}(T(X))$ satisfying

$$\sum_{t=0}^n h_{m,k}(t)\binom{n}{t}\theta^t = \binom{m}{k}\theta^k(\theta+1)^{n-m} = \sum_{t=k}^{n-m+k}\binom{m}{k}\binom{n-m}{t-k}\theta^t,$$

i.e.,

$$h_{m,k}(t) = \tbinom{m}{k}\tbinom{n-m}{t-k}/\tbinom{n}{t} \text{ if } t \in \{k, k+1, \cdots, n-m+k\} \text{ and } h_{m,k}(t) = 0 \text{ otherwise.}$$

(c) The UMVUE of

$$\begin{aligned}\mathbb{P}(X_1 + \cdots + X_{n-1} > X_n) &= \mathbb{P}(X_1 + \cdots + X_{n-1} > 0)\mathbb{P}(X_n = 0) + \mathbb{P}(X_1 + \cdots + X_{n-1} > 1)\mathbb{P}(X_n = 1) \\ &= [1 - (1-p)^{n-1}](1-p) + [1 - (1-p)^{n-1} - (n-1)p(1-p)^{n-2}]p \\ &= 1 - (1-p)^{n-1} - (n-1)p^2(1-p)^{n-2}\end{aligned}$$

is $h(T(X))$, where $h(t) = 1 - h_{n-1,0}(t) - \frac{2}{n}h_{n,2}(t)$. ////

Exercise (3.6)**.** Let $X_1, \ldots, X_n$ be i.i.d. having the exponential distribution Exponential(a, θ) with parameters $\theta > 0$ and $a \in \mathbb{R}$.
(a) Find the UMVUE of a when θ is known.
(b) Find the UMVUE of θ when a is known.
(c) Find the UMVUE's of θ and a.
(d) Assume that θ is known. Find the UMVUE's of $\mathbb{P}(X_1 \geq t)$ and $\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{P}(X_1 \geq t)$ for a fixed $t > a$.
(e) Find the UMVUE of $\mathbb{P}(X_1 \geq t)$ for a fixed $t > a$.
Remark. One may refer to Exercise 2.59 and its proof.

Solution. The joint p.d.f. of X_i's is $x = (x_1, \ldots, x_n) \mapsto \theta^{-n}\mathrm{e}^{-n(\bar{x}-a)/\theta}\mathbb{1}_{\{x_{(1)}>a\}}$, where $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$ and $x_{(1)} = \min_{1\leq i\leq n} x_i$. Note that $X_{(1)} \sim$ Exponential$(a, \theta/n)$.

(a) Since $X_{(1)}$ is sufficient and complete and $\mathbb{E}X_{(1)} = a + \theta/n$, the UMVUE of a is $X_{(1)} - \theta/n$.

(b) Since $\bar{X}$ is a sufficient and complete statistic such that $\mathbb{E}\bar{X} = a + \theta$, the UMVUE of θ is $\bar{X} - a$.

(c) Since $(X_{(1)}, \bar{X})$ is sufficient and complete, the UMVUE of (θ, a) is $(\frac{n}{n-1}(\bar{X} - X_{(1)}), \frac{nX_{(1)}-\bar{X}}{n-1})$.

(d) Note that $\mathbb{P}(X_1 \geq t) = \mathrm{e}^{-(t-a)_+/\theta}$ and $\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{P}(X_1 \geq t) = -\theta^{-1}\mathrm{e}^{-(t-a)/\theta}\mathbb{1}_{[t>a]}$. If $h_\theta(X_{(1)})$ is unbiased for $\mathrm{e}^{-(t-a)_+/\theta}$, then the UMVUE's of $\mathbb{P}(X_1 \geq t)$ and $\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{P}(X_1 \geq t)$ are given by $h_\theta(X_{(1)})$ and $-h_\theta(X_{(1)})/\theta$, respectively, provided $t > a$. Straightforward calculations yield

$$\mathrm{e}^{-(t-a)_+/\theta} = \mathbb{E}[h_\theta(X_{(1)})] = \int_a^\infty h_\theta(x)\frac{n}{\theta}\mathrm{e}^{-n(x-a)/\theta}\,\mathrm{d}x = \frac{n}{\theta}\mathrm{e}^{na/\theta}\int_a^\infty h_\theta(x)\mathrm{e}^{-nx/\theta}\,\mathrm{d}x.$$

Dividing both sides by $\mathrm{e}^{na/\theta}$ and then differentiating them w.r.t. a, we obtain that

$$h_\theta(x) = \tfrac{n-1}{n}\mathrm{e}^{-(t-x)/\theta} \text{ if } x < t \text{ and } h_\theta(x) = 1 \text{ if } x > t.$$

(e) Recall that the sufficient and complete statistic $(X_{(1)}, n(\bar{X} - X_{(1)}))$ has the p.d.f.

$$(x, y) \mapsto \frac{n}{\theta}\mathrm{e}^{-n(x-a)/\theta}\mathbb{1}_{(a,\infty)}(x) \cdot \frac{1}{\Gamma(n-1)\theta^{n-1}}y^{n-2}\mathrm{e}^{-y/\theta}\mathbb{1}_{(0,\infty)}(y).$$

The UMVUE of $\mathbb{P}(X_1 \geq t)$ is $\phi(X_{(1)}, n(\bar{X} - X_{(1)}))$ satisfying $\mathbb{E}[\phi(X_{(1)}, n(\bar{X} - X_{(1)}))|X_{(1)}] = h_\theta(X_{(1)})$, i.e.,

$$(\star) \qquad \int_0^\infty \phi(x, y)y^{n-2}\mathrm{e}^{-y/\theta}\,\mathrm{d}y = h_\theta(x)\Gamma(n-1)\theta^{n-1} = \int_0^\infty h_\theta(x)y^{n-2}\mathrm{e}^{-y/\theta}\,\mathrm{d}y.$$

When $x > t$, the uniqueness of the Laplace transform gives $\phi(x, y) = 1$. When $x < t$, it follows from

$$(\star) = \int_{t-x}^\infty \frac{n-1}{n}[z - (t-x)]^{n-2}\mathrm{e}^{-z/\theta}\,\mathrm{d}z = \int_0^\infty \frac{n-1}{n}[y - (t-x)]^{n-2}\mathbb{1}_{(t-x,\infty)}(y)\mathrm{e}^{-y/\theta}\,\mathrm{d}y$$

that

$$\phi(x, y) = \frac{n-1}{n}\frac{[y - (t-x)]^{n-2}}{y^{n-2}}\mathbb{1}_{(t-x,\infty)}(y) = \frac{n-1}{n}\left(1 - \frac{t-x}{y}\right)^{n-2}\mathbb{1}_{[x+y>t]}.$$

Thus, $\phi(x, y)$ has been determined almost everywhere. ////

Exercise (3.37)**.** Let $X_1, \ldots, X_n$ be i.i.d. from the uniform distribution Uniform$(0, \theta)$ with $\theta > 0$.
(a) Show that

$$\frac{\partial}{\partial\theta}\int h(x)f_\theta(x)\,\mathrm{d}x = \int h(x)\frac{\partial}{\partial\theta}f_\theta(x)\,\mathrm{d}x$$

does not hold for $h(X) = X_{(n)}$.
(b) Show that the inequality

$$\mathrm{Var}(T(X)) \geq \frac{1}{\mathbb{E}\left[\frac{\partial}{\partial\theta}\log f_\theta(X)\right]^2}$$

does not hold for the UMVUE $T(X)$ of θ.

Proof. (a) Note that $f_\theta(x) = \theta^{-n}\mathbb{1}_{[x_{(n)}<\theta]}$ and $\mathbb{P}(X_{(n)} > t) = 1-(t/\theta)^n$ for $t \in (0,\theta)$. One can see that

$$\int h(x)f_\theta(x)\,\mathrm{d}x = \mathbb{E}X_{(n)} = \int_0^\infty \mathbb{P}(X_{(n)} > t)\,\mathrm{d}t = \int_0^\theta \left[1-\left(\frac{t}{\theta}\right)^n\right]\mathrm{d}t = \frac{n}{n+1}\theta$$

$$\implies \int h(x)\frac{\partial}{\partial\theta}f_\theta(x)\,\mathrm{d}x = \int h(x)\frac{-n}{\theta}f_\theta(x)\,\mathrm{d}x = \frac{-n^2}{n+1} \neq \frac{n}{n+1} = \frac{\partial}{\partial\theta}\int h(x)f_\theta(x)\,\mathrm{d}x.$$

(b) Clearly $T(X) = \frac{n+1}{n}X_{(n)}$. On one hand,

$$\mathbb{E}X_{(n)}^2 = \int_0^\infty 2t\mathbb{P}(X_{(n)} > t)\,\mathrm{d}t = \int_0^\theta 2t\left[1-\left(\frac{t}{\theta}\right)^n\right]\mathrm{d}t = \frac{n}{n+2}\theta^2$$

$$\implies \mathrm{Var}(T(X)) = \left(\frac{n+1}{n}\right)^2\left[\mathbb{E}X_{(n)}^2 - (\mathbb{E}X_{(n)})^2\right] = \frac{\theta^2}{n(n+2)}.$$

On the other hand,

$$\frac{\partial}{\partial\theta}\log f_\theta(x) = \frac{\frac{\partial}{\partial\theta}f_\theta(x)}{f_\theta(x)} = \frac{-n}{\theta} \implies \mathbb{E}\left[\frac{\partial}{\partial\theta}\log f_\theta(X)\right]^2 = \frac{n^2}{\theta^2}.$$

Therefore,

$$\mathrm{Var}(T(X))\cdot\mathbb{E}\left[\tfrac{\partial}{\partial\theta}\log f_\theta(X)\right]^2 = n/(n+2) < 1.$$

This completes the proof. □

Exercise (3.39)**.** Let X be a single sample from the double exponential distribution DoubleExponential(μ,θ) with $\mu = 0$ and $\theta > 0$. Find the UMVUE's of the following parameters and, in each case, determine whether the variance of the UMVUE attains the Cramér–Rao lower bound.
(a) $\vartheta = \theta$; (b) $\vartheta = \theta^r$, where $r > 1$; (c) $\vartheta = (1+\theta)^{-1}$.

Solution. Note that $|X| \sim$ Exponential$(0,\theta)$ is a sufficient and complete statistic for θ, since X has the p.d.f. $f_\theta(x) = \exp(-|x|/\theta)/(2\theta)$, $x \in \mathbb{R}$. Also, using $\frac{\partial}{\partial\theta}\log f_\theta(x) = |x|/\theta^2 - 1/\theta$, we obtain the Fisher information $\mathcal{I}(\theta) = \mathrm{Var}(\frac{\partial}{\partial\theta}\log f_\theta(X)) = \mathrm{Var}(|X|/\theta^2) = 1/\theta^2$.

(a&b) The UMVUE $|X|^r/\Gamma(r+1)$ of θ^r has variance $[\Gamma(2r+1)/\Gamma(r+1)^2 - 1]\theta^{2r}$, and the Cramér–Rao lower bound is $(r\theta^{r-1})^2/\mathcal{I}(\theta) = r^2\theta^{2r}$, which is attained if and only if $r = 1$.

(c) The UMVUE of $(1+\theta)^{-1}$ is $\exp(-|X|)$ with variance $(1+2\theta)^{-1} - (1+\theta)^{-2} = \theta^2/[(1+2\theta)(1+\theta)^2]$, which doesn't attain the Cramér–Rao lower bound $[-(1+\theta)^{-2}]^2/\mathcal{I}(\theta) = \theta^2/(1+\theta)^4$. ////

Exercise (3.53)**.** Let $h(x_1,x_2,x_3) = \mathbb{1}_{(-\infty,0)}(x_1+x_2+x_3)$. Define the U-statistic with this kernel and find h_k and ζ_k, $k = 1,2,3$.

Solution. Suppose that $X_1,\ldots,X_n$ are i.i.d. random variables with a left-continuous c.d.f. F, then $\sum_{i=1}^n X_i$ has the c.d.f. F^{*n}. The U-statistic with kernel h is $U_n = \binom{n}{3}^{-1}\sum_c \mathbb{1}_{\{X_i+X_j+X_k<0\}}$, where $\sum_c$ denotes the summation over the $\binom{n}{3}$ combinations $\{i,j,k\}$ of 3 distinct elements from $\{1,\cdots,n\}$. Besides, we have

- $h_1(x_1) = \mathbb{E}[h(x_1,X_2,X_3)] = \mathbb{P}(X_2+X_3 < -x_1) = F^{*2}(-x_1)$ with $\zeta_1 = \mathrm{Var}(h_1(X_1)) = \mathrm{Var}(F^{*2}(-X_1))$,
- $h_2(x_1,x_2) = \mathbb{E}[h(x_1,x_2,X_3)] = F(-x_1-x_2)$ with $\zeta_2 = \mathrm{Var}(h_2(X_1,X_2)) = \mathrm{Var}(F(-X_1-X_2))$, and
- $h_3 = h$ with $\zeta_3 = \mathrm{Var}(\mathbb{1}_{\{X_1+X_2+X_3<0\}}) = p(1-p)$, where $p = \mathbb{P}(X_1+X_2+X_3 < 0) = F^{*3}(0)$. ////

Exercise (3.54)**.** Let $X_1, \dots, X_n$ be i.i.d. random variables having finite $\mu = \mathbb{E}X_1$ and $\bar{\mu} = \mathbb{E}X_1^{-1}$. Find a U-statistic that is an unbiased estimator of $\mu\bar{\mu}$ and derive its variance and asymptotic distribution.

Solution. Consider $h(x_1, x_2) = \frac{1}{2}(\frac{x_1}{x_2} + \frac{x_2}{x_1})$. Then the U-statistic with kernel h is

$$U_n = \frac{1}{n(n-1)} \sum_{1 \le i < j \le n} \left(\frac{X_i}{X_j} + \frac{X_j}{X_i} \right),$$

which is unbiased for $\mathbb{E}[h(X_1, X_2)] = \mu\bar{\mu}$. Let $h_1(x_1) = \mathbb{E}[h(x_1, X_2)] = \frac{1}{2}(\bar{\mu}x_1 + \mu/x_1)$ and $h_2 = h$. It can be seen that

$$\zeta_1 = \operatorname{Var}(h_1(X_1)) = \tfrac{1}{4}[\bar{\mu}^2 \operatorname{Var}(X_1) + \mu^2 \operatorname{Var}(X_1^{-1}) + 2\mu\bar{\mu}(1 - \mu\bar{\mu})],$$
$$\zeta_2 = \operatorname{Var}(h_2(X_1, X_2)) = \tfrac{1}{2}(\mathbb{E}[X_1^2]\mathbb{E}[X_1^{-2}] + 1 - 2\mu^2\bar{\mu}^2),$$

and then

$$\operatorname{Var}(U_n) = \binom{n}{2}^{-1} \sum_{k=1}^{2} \binom{2}{k}\binom{n-2}{2-k}\zeta_k = \frac{4(n-2)\zeta_1 + 2\zeta_2}{n(n-1)} \sim \frac{4\zeta_1}{n}$$

is determined. Moreover, $\sqrt{n}(U_n - \mu\bar{\mu}) \xrightarrow{d} \mathcal{N}(0, 4\zeta_1)$ by Theorem 3.5. ////

Exercise (3.63)**.** Under the model $X = Z\beta + \varepsilon$ and the assumption that ε is distributed as $\mathcal{N}_n(0_n, \sigma^2 I_n)$ with an unknown $\sigma^2 > 0$, find the UMVUE's of $(l^\top\beta)^2$, $l^\top\beta/\sigma$, and $(l^\top\beta/\sigma)^2$ for an estimable $l^\top\beta$.

Solution. Denote $\mu = Z\beta$, then $X \sim \mathcal{N}_n(\mu, \sigma^2 I_n)$ and $l^\top\beta = c^\top\mu$ for some c such that $Z^\top c = l$. Let

$$\hat{X} = Z(Z^\top Z)^- Z^\top X = \operatorname{proj}_{\operatorname{Col}(Z)} X \sim \mathcal{N}_n(\mu, \sigma^2 \operatorname{proj}_{\operatorname{Col}(Z)})$$

and $\hat{\sigma} = \sqrt{S/(n-r)}$, where $r = \operatorname{rank}(Z)$ and

$$S = \|X - \hat{X}\|^2 = X^\top(I_n - \operatorname{proj}_{\operatorname{Col}(Z)})X = \varepsilon^\top(I_n - \operatorname{proj}_{\operatorname{Col}(Z)})\varepsilon \sim \sigma^2\chi^2_{n-r}.$$

One can see that $\|X - \mu\|^2 - S = \|\hat{X} - \mu\|^2 = \hat{X}^\top\hat{X} - 2\mu^\top\hat{X} + \mu^\top\mu$. Hence, the p.d.f. of X is given by

$$x \mapsto (2\pi\sigma^2)^{-n/2} \exp\{-(S + \hat{X}^\top\hat{X} - 2\mu^\top\hat{X} + \mu^\top\mu)/(2\sigma^2)\},$$

from which $(\hat{X}, S)$ is complete and sufficient for (μ, σ^2). Also, $\hat{X}$ and S are independent, since $\operatorname{proj}_{\operatorname{Col}(Z)} X$ and $(I_n - \operatorname{proj}_{\operatorname{Col}(Z)})X$ are jointly normal and uncorrelated. Note that $\mathbb{E}[\hat{X}\hat{X}^\top] = \mu\mu^\top + \sigma^2 \operatorname{proj}_{\operatorname{Col}(Z)}$ and $c^\top \operatorname{proj}_{\operatorname{Col}(Z)} c = l^\top(Z^\top Z)^- l$, so the UMVUE of $(l^\top\beta)^2 = c^\top\mu\mu^\top c$ is $(c^\top\hat{X})^2 - \hat{\sigma}^2 l^\top(Z^\top Z)^- l$. The UMVUE's of $l^\top\beta/\sigma$ and $(l^\top\beta/\sigma)^2$ are given by $c^\top\hat{X}/(\mu_{-1/2,n-r}\hat{\sigma})$ and $(c^\top\hat{X})^2/(\mu_{-1,n-r}\hat{\sigma}^2) - l^\top(Z^\top Z)^- l$, respectively, where $\mu_{\alpha,m} = \int_0^\infty s^\alpha \frac{s^{m/2-1}\mathrm{e}^{-s/2}}{2^{m/2}\Gamma(m/2)}\,\mathrm{d}s/m^\alpha = \frac{2^\alpha\Gamma(m/2+\alpha)}{\Gamma(m/2)m^\alpha}$ is the α^{th} moment of χ^2_m/m. ////

Exercise (3.107)**.** Let $X_1, \dots, X_n$ be i.i.d. random variables having the Lebesgue p.d.f.

$$f_{\alpha,\beta}(x) = \alpha\beta^{-\alpha}x^{\alpha-1}\mathbb{1}_{(0,\beta)}(x),$$

where $\alpha > 0$ and $\beta > 0$ are unknown.
(a) Obtain moment estimators of α and β.
(b) Obtain the asymptotic distribution of the moment estimators of α and β derived in (a).

Solution. (a) It's easily seen that $\mu_k(\alpha, \beta) = \int x^k f_{\alpha,\beta}(x)\,\mathrm{d}x = \alpha\beta^k/(\alpha + k)$. The solution $(\hat{\alpha}, \hat{\beta})$ to the estimating equation $\mu_k(\hat{\alpha}, \hat{\beta}) = \hat{\mu}_k = \frac{1}{n}\sum_{i=1}^n X_i^k$ for $k = 1, 2$ is then given by

$$\hat{\alpha} = -1 + \sqrt{\hat{\mu}_2/(\hat{\mu}_2 - \hat{\mu}_1^2)} \text{ and } \hat{\beta} = (1 + \sqrt{1 - \hat{\mu}_1^2/\hat{\mu}_2})\hat{\mu}_2/\hat{\mu}_1.$$

(b) By the central limit theorem, $\sqrt{n}(\hat{\mu}_1-\mu_1,\hat{\mu}_2-\mu_2)\xrightarrow{d}\mathcal{N}_2(0_2,\Sigma)$, where

$$\Sigma=\begin{pmatrix}\mathrm{Var}(X_1) & \mathrm{Cov}(X_1,X_1^2)\\ \mathrm{Cov}(X_1^2,X_1) & \mathrm{Var}(X_1^2)\end{pmatrix}=\begin{pmatrix}\mu_2-\mu_1^2 & \mu_3-\mu_1\mu_2\\ \mu_3-\mu_1\mu_2 & \mu_4-\mu_2^2\end{pmatrix}.$$

Note that $\mu_{k+\ell}-\mu_k\mu_\ell=k\ell\alpha\beta^{k+\ell}/[(\alpha+k)(\alpha+\ell)(\alpha+k+\ell)]$. Let

$$D=\frac{\partial(\mu_1,\mu_2)}{\partial(\alpha,\beta)}=\begin{pmatrix}\beta/(\alpha+1)^2 & \alpha/(\alpha+1)\\ 2\beta^2/(\alpha+2)^2 & 2\alpha\beta/(\alpha+2)\end{pmatrix}.$$

It follows from the delta method that $\sqrt{n}(\hat{\alpha}-\alpha,\hat{\beta}-\beta)\xrightarrow{d}\mathcal{N}_2(0_2,D^{-1}\Sigma(D^{-1})^\top)$. ////

Exercise (3.108)**.** Let $X_1,\ldots,X_n$ be i.i.d. from the following discrete distribution:

$$\mathbb{P}(X_1=1)=\frac{2(1-\theta)}{2-\theta},\quad \mathbb{P}(X_1=2)=\frac{\theta}{2-\theta},$$

where $\theta\in(0,1)$ is unknown.
(a) Obtain an estimator of θ using the method of moments.
(b) Obtain the AMSE of the moment estimator in (a).

Solution. (a) Note that $\mathbb{E}X_1=2/(2-\theta)$. Thus, solving $2/(2-\hat{\theta})=\bar{X}=\frac{1}{n}\sum_{i=1}^n X_i$ leads to $\hat{\theta}=2-2/\bar{X}$.

(b) Note that $\mathrm{Var}(X_1)=\frac{2+2\theta}{2-\theta}-(\frac{2}{2-\theta})^2=\frac{2\theta-2\theta^2}{(2-\theta)^2}$. We have $\sqrt{n}(\bar{X}-\frac{2}{2-\theta})\xrightarrow{d}\xi\sim\mathcal{N}(0,\frac{2\theta-2\theta^2}{(2-\theta)^2})$ by the central limit theorem. Then the delta method gives $\sqrt{n}(\hat{\theta}-\theta)\xrightarrow{d}(\frac{2}{(2-\theta)^2})^{-1}\xi\sim\mathcal{N}(0,\frac{1}{2}\theta(1-\theta)(2-\theta)^2)$, from which one can see that $\mathrm{AMSE}_{\hat{\theta}}(\theta)=\theta(1-\theta)(2-\theta)^2/(2n)$. ////

Exercise (4.1)**.** Show that the priors in the following cases are conjugate priors:
(a) $X_1,\ldots,X_n$ are i.i.d. from $\mathcal{N}_k(\theta,I_k)$, $\theta\in\mathbb{R}^k$, and $\Pi=\mathcal{N}_k(\mu_0,\Sigma_0)$;
(b) $X_1,\ldots,X_n$ are i.i.d. from Binomial(θ,k), $\theta\in(0,1)$, and $\Pi=\mathrm{Beta}(\alpha,\beta)$;
(d) $X_1,\ldots,X_n$ are i.i.d. from Exponential$(0,\theta)$, $\theta>0$, and $\Pi=\mathrm{InverseGamma}(\alpha,\gamma)$.

Proof. Denote the sample mean by $\bar{X}=\frac{1}{n}\sum_{i=1}^n X_i$.

(a) Note that

$$\begin{aligned}\pi(\theta|x)&\propto\pi(\theta)f(x|\theta)\\ &\propto\exp\Big\{-\frac{1}{2}(\theta-\mu_0)^\top\Sigma_0^{-1}(\theta-\mu_0)\Big\}\prod_{i=1}^n\exp\Big\{-\frac{1}{2}(x_i-\theta)^\top(x_i-\theta)\Big\}\\ &\propto\exp\Big\{-\frac{1}{2}\Big[\theta^\top\Sigma_0^{-1}\theta-2\theta^\top\Sigma_0^{-1}\mu_0+\sum_{i=1}^n(\theta^\top\theta-2\theta^\top x_i)\Big]\Big\}\\ &=\exp\Big\{-\frac{1}{2}\big[\theta^\top(\Sigma_0^{-1}+nI_k)\theta-2\theta^\top(\Sigma_0^{-1}\mu_0+n\bar{x})\big]\Big\}\\ &=\exp\Big\{-\frac{1}{2}\big[\theta^\top\Sigma^{-1}\theta-2\theta^\top\Sigma^{-1}\mu(x)\big]\Big\}\\ &\propto\exp\Big\{-\frac{1}{2}(\theta-\mu(x))^\top\Sigma^{-1}(\theta-\mu(x))\Big\},\end{aligned}$$

where

$$\Sigma=(\Sigma_0^{-1}+nI_k)^{-1}\quad\text{and}\quad\mu(x)=\Sigma(\Sigma_0^{-1}\mu_0+n\bar{x}),$$

so we have $\theta|X\sim\mathcal{N}_k(\mu(X),\Sigma)$.

(b) Note that

$$\begin{aligned}\pi(\theta|x) &\propto \pi(\theta)f(x|\theta)\\ &\propto \theta^{\alpha-1}(1-\theta)^{\beta-1}\prod_{i=1}^{n}\left[\theta^{x_i}(1-\theta)^{k-x_i}\right]\\ &= \theta^{\alpha+n\bar{x}-1}(1-\theta)^{\beta+nk-n\bar{x}-1},\end{aligned}$$

so we have $\theta|X \sim \text{Beta}(\alpha+n\bar{X}, \beta+nk-n\bar{X})$.

(d) Note that

$$\begin{aligned}\pi(\theta|x) &\propto \pi(\theta)f(x|\theta)\\ &\propto \theta^{-\alpha-1}\exp(-\gamma/\theta)\prod_{i=1}^{n}\left[\theta^{-1}\exp(-x_i/\theta)\right]\\ &= \theta^{-(\alpha+n)-1}\exp\{-(\gamma+n\bar{x})/\theta\},\end{aligned}$$

so we have $\theta|X \sim \text{InverseGamma}(\alpha+n, \gamma+n\bar{X})$. □

Exercise (4.4)**.** Let $X_1,\ldots,X_n$ be i.i.d. from Uniform$(0,\theta)$, where $\theta>0$ is unknown. Let the prior of θ be LogNormal(μ_0,σ_0^2), where $\mu_0\in\mathbb{R}$ and $\sigma_0>0$ are known constants.

(a) Find the posterior p.d.f. of $\vartheta=\log\theta$.

(b) Find the r^{th} posterior moment of θ.

(c) Find a value that maximizes the posterior p.d.f. of θ.

Solution. Denote the largest order statistic by $X_{(n)}=\max_{1\le i\le n}X_i$.

(a) Note that $\vartheta=\log\theta\sim\mathcal{N}(\mu_0,\sigma_0^2)$, and thus

$$\begin{aligned}\pi(\vartheta|x) &\propto \pi(\vartheta)f(x|\vartheta)\\ &\propto \exp\left\{-\frac{(\vartheta-\mu_0)^2}{2\sigma_0^2}\right\}\prod_{i=1}^{n}\frac{\mathbb{1}\{x_i<\exp(\vartheta)\}}{\exp(\vartheta)}\\ &\propto \exp\left\{-\frac{\vartheta^2-2\mu_0\vartheta}{2\sigma_0^2}-n\vartheta\right\}\mathbb{1}\{x_{(n)}<\exp(\vartheta)\}\\ &= \exp\left\{-\frac{\vartheta^2-2(\mu_0-n\sigma_0^2)\vartheta}{2\sigma_0^2}\right\}\mathbb{1}\{\log x_{(n)}<\vartheta\}\\ &\propto \exp\left\{-\frac{[\vartheta-(\mu_0-n\sigma_0^2)]^2}{2\sigma_0^2}\right\}\mathbb{1}\{\vartheta>\log x_{(n)}\},\end{aligned}$$

so $\vartheta|X$ is distributed as $\mathcal{N}(\mu_0-n\sigma_0^2,\sigma_0^2)$ truncated on $(\log X_{(n)},\infty)$. Hence,

$$\pi(\vartheta|X)=\frac{C_X^{-1}}{\sqrt{2\pi}\sigma_0}\exp\left\{-\frac{[\vartheta-(\mu_0-n\sigma_0^2)]^2}{2\sigma_0^2}\right\}\mathbb{1}\{\vartheta>\log X_{(n)}\}\quad\text{where}\quad C_X=\Phi\left(\frac{\mu_0-n\sigma_0^2-\log X_{(n)}}{\sigma_0}\right).$$

(b) Direct calculations lead to

$$\begin{aligned}&\mathbb{E}[\theta^r|X]=\mathbb{E}[\exp(r\vartheta)|X]\\ &=\int_{\log X_{(n)}}^{\infty}\frac{C_X^{-1}}{\sqrt{2\pi}\sigma_0}\exp\left\{r\vartheta-\frac{[\vartheta-(\mu_0-n\sigma_0^2)]^2}{2\sigma_0^2}\right\}\mathrm{d}\vartheta\\ &=C_X^{-1}\int_{\log X_{(n)}}^{\infty}\frac{1}{\sqrt{2\pi}\sigma_0}\exp\left\{-\frac{[\vartheta-(\mu_0-n\sigma_0^2+r\sigma_0^2)]^2}{2\sigma_0^2}\right\}\exp\left\{\frac{(2\mu_0-2n\sigma_0^2+r\sigma_0^2)r\sigma_0^2}{2\sigma_0^2}\right\}\mathrm{d}\vartheta\\ &=C_X^{-1}\Phi\left(\frac{\mu_0-(n-r)\sigma_0^2-\log X_{(n)}}{\sigma_0}\right)\exp\left\{\left[\mu_0-\left(n-\frac{r}{2}\right)\sigma_0^2\right]r\right\}.\end{aligned}$$

(c) Since $\frac{\mathrm{d}\vartheta}{\mathrm{d}\theta} = 1/\theta = \exp(-\vartheta)$, we have

$$\pi(\theta|X) = \pi(\vartheta|X)\frac{\mathrm{d}\vartheta}{\mathrm{d}\theta} \propto \exp\Big\{-\frac{[\vartheta-(\mu_0-n\sigma_0^2)]^2}{2\sigma_0^2} - \vartheta\Big\}\mathbb{1}\{\vartheta > \log X_{(n)}\}.$$

Thus,

$$\max \pi(\theta|X) \iff \min\big\{[\vartheta-(\mu_0-n\sigma_0^2)]^2 + 2\sigma_0^2\vartheta\big\} \text{ s.t. } \vartheta > \log X_{(n)}.$$

It's clear that the optimal value is given by $\theta^* = \max\{\exp(\vartheta^*), X_{(n)}\}$, where $\vartheta^* = \mu_0 - (n+1)\sigma_0^2$. ////

Exercise (4.7)**.** Let $X_1, \dots, X_n$ be i.i.d. binary with $\mathbb{P}(X_1 = 1) = p \in (0, 1)$. Find the Bayes action w.r.t. the uniform prior on $[0, 1]$ in the problem of estimating p under the loss $L(p, a) = (p - a)^2/[p(1 - p)]$.

Solution. Denote the sample mean by $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$. We have $p|X \sim \text{Beta}(n\bar{X}+1, n-n\bar{X}+1)$, since $\pi(p|x) \propto \pi(p)f(x|p) = 1\cdot\prod_{i=1}^n[p^{x_i}(1-p)^{1-x_i}] = p^{n\bar{x}}(1-p)^{n-n\bar{x}}$. It follows that

$$\begin{aligned}\mathbb{E}[L(p,a)|X] &= \int_0^1 \frac{p^2-2ap+a^2}{p(1-p)}\cdot\frac{p^{n\bar{X}}(1-p)^{n-n\bar{X}}}{\mathrm{B}(n\bar{X}+1, n-n\bar{X}+1)}\,\mathrm{d}p \\ &= \frac{\mathrm{B}(n\bar{X}+2, n-n\bar{X}) - 2a\,\mathrm{B}(n\bar{X}+1, n-n\bar{X}) + a^2\,\mathrm{B}(n\bar{X}, n-n\bar{X})}{\mathrm{B}(n\bar{X}+1, n-n\bar{X}+1)} \\ &= \frac{n\bar{X}+1}{n-n\bar{X}} - 2a\frac{n+1}{n-n\bar{X}} + a^2\frac{n(n+1)}{n\bar{X}(n-n\bar{X})}.\end{aligned}$$

The Bayes action is then given by $\arg\min_a \mathbb{E}[L(p,a)|X] = \bar{X}$. ////

Exercise (4.23)**.** Let $X_1, \dots, X_n$ be i.i.d. with the Lebesgue p.d.f. $f_\theta(x) = \sqrt{2\theta/\pi}\exp(-\theta x^2/2)\mathbb{1}_{[0,\infty)}(x)$, where $\theta > 0$ is unknown. Let the prior of θ be Gamma(α, γ) with known α and γ. Find the Bayes estimator of θ and its Bayes risk under the loss function $L(\theta, a) = (a - \theta)^2/\theta$.

Solution. Denote $T = T(X) = \sum_{i=1}^n X_i^2/2$. We have $\theta|X \sim \text{Gamma}(\alpha + n/2, \gamma/(1+\gamma T))$, since $\pi(\theta|x) \propto \pi(\theta)f(x|\theta) \propto \theta^{\alpha-1}\exp(-\theta/\gamma)\mathbb{1}_{(0,\infty)}(\theta)\cdot\prod_{i=1}^n[\theta^{1/2}\exp(-\theta x_i^2/2)] = \theta^{\alpha+n/2-1}\exp\{-\theta[1/\gamma+T(x)]\}\mathbb{1}_{(0,\infty)}(\theta)$. It follows that

$$\begin{aligned}\mathbb{E}[L(\theta,a)|X] &= \int_0^\infty \frac{\theta^2-2a\theta+a^2}{\theta}\cdot\frac{\theta^{\alpha+n/2-1}\exp\{-\theta/[\gamma/(1+\gamma T)]\}}{\Gamma(\alpha+n/2)[\gamma/(1+\gamma T)]^{\alpha+n/2}}\,\mathrm{d}\theta \\ &= (\alpha+n/2)\gamma/(1+\gamma T) - 2a + a^2/[(\alpha+n/2-1)\gamma/(1+\gamma T)].\end{aligned}$$

The Bayes action is then given by $\hat{\theta}(X) = \arg\min_a \mathbb{E}[L(\theta,a)|X] = (\alpha+n/2-1)\gamma/(1+\gamma T)$, whose Bayes risk is $r(\pi,\hat{\theta}) = \mathbb{E}\{\mathbb{E}[L(\theta,\hat{\theta}(X))|\theta]\} = \mathbb{E}\{\mathbb{E}[L(\theta,\hat{\theta}(X))|X]\} = \mathbb{E}\{\gamma/(1+\gamma T)\}$. Note that $\sqrt{\theta}X_i|\theta \overset{\text{i.i.d.}}{\sim} |\mathcal{N}(0,1)|$ and thus $2\theta T|\theta \sim \chi_n^2 = \text{Gamma}(n/2, 2)$. One can see that $T = T(X)$ has the conditional Lebesgue p.d.f. $g(t|\theta) = [\Gamma(n/2)]^{-1}\theta^{n/2}t^{n/2-1}\exp(-\theta t)\mathbb{1}_{(0,\infty)}(t)$ and the marginal Lebesgue p.d.f.

$$g(t) = \int g(t|\theta)\pi(\theta)\,\mathrm{d}\theta = t^{n/2-1}\mathbb{1}_{(0,\infty)}(t)\int_0^\infty \frac{\theta^{n/2+\alpha-1}\exp\{-\theta(t+1/\gamma)\}}{\Gamma(n/2)\Gamma(\alpha)\gamma^\alpha}\,\mathrm{d}\theta = \frac{\gamma^{n/2}t^{n/2-1}\mathbb{1}_{(0,\infty)}(t)}{\mathrm{B}(n/2,\alpha)(1+\gamma t)^{\alpha+n/2}}.$$

Hence, the Bayes risk $r(\pi,\hat{\theta}) = \int \frac{\gamma}{1+\gamma t}g(t)\,\mathrm{d}t = \gamma\frac{\gamma^{n/2}}{\mathrm{B}(n/2,\alpha)}\frac{\mathrm{B}(n/2,\alpha+1)}{\gamma^{n/2}} = \gamma\alpha/(\alpha+n/2)$. ////

Exercise (4.34)**.** Let X be an observation from Gamma(α, θ) with a known α and an unknown $\theta > 0$. Show that $X/(\alpha + 1)$ is an admissible estimator of θ under the squared error loss, using Theorem 4.3.

Proof. Consider $\theta \sim \Pi = \text{InverseGamma}(k, \sigma)$ where $k > 2$, then $\theta|X \sim \text{InverseGamma}(\alpha+k, X+\sigma)$, since

$\pi(\theta|x) \propto f(x|\theta)\pi(\theta) \propto \theta^{-\alpha}\exp(-x/\theta)\cdot\theta^{-k-1}\exp(-\sigma/\theta) = \theta^{-(\alpha+k)-1}\exp\{-(x+\sigma)/\theta\}$. Thus, the Bayes estimator of θ w.r.t. the prior Π under the squared error loss is the posterior mean

$$\hat{\theta}(X) = \mathbb{E}[\theta|X] = (X+\sigma)/(\alpha+k-1),$$

whose risk function is given by

$$\mathrm{MSE}_{\hat{\theta}}(\theta) = (\mathbb{E}[\hat{\theta}(X)|\theta]-\theta)^2 + \mathrm{Var}(\hat{\theta}(X)|\theta) = \{[\sigma-(k-1)\theta]^2 + \alpha\theta^2\}/(\alpha+k-1)^2.$$

It follows from $\mathbb{E}[\theta] = \sigma/(k-1)$ and $\mathrm{Var}(\theta) = \sigma^2/[(k-1)^2(k-2)]$ that

$$r^*(\Pi) = \mathbb{E}[\mathrm{MSE}_{\hat{\theta}}(\theta)] = \{(k-1)^2\,\mathrm{Var}(\theta) + \alpha\mathbb{E}[\theta^2]\}/(\alpha+k-1)^2 = \sigma^2/[(k-1)(k-2)(\alpha+k-1)].$$

For $T(X) = X/(\alpha+1)$, we have

$$\mathrm{MSE}_T(\theta) = (\mathbb{E}[T(X)|\theta]-\theta)^2 + \mathrm{Var}(T(X)|\theta) = [(-\theta)^2+\alpha\theta^2]/(\alpha+1)^2 = \theta^2/(\alpha+1),$$

and $r_T(\Pi) = \mathbb{E}[\mathrm{MSE}_T(\theta)] = \sigma^2/[(k-1)(k-2)(\alpha+1)]$. One can see that

$$r_T(\Pi) - r^*(\Pi) = \sigma^2/[(k-1)(\alpha+1)(\alpha+k-1)].$$

Let $O = (\theta_-, \theta_+)$ with $0 < \theta_- < \theta_+ < \infty$. By the mean value theorem, there exists some $\theta_0 \in (\theta_-, \theta_+)$ such that

$$\Pi(O) = [\Gamma(k)]^{-1}\sigma^k\theta_0^{-k-1}\exp(-\sigma/\theta_0)(\theta_+-\theta_-).$$

Hence,

$$\begin{aligned} 0 \le [r_T(\Pi)-r^*(\Pi)]/\Pi(O) &= \sigma^{2-k}\Gamma(k)\theta_0^{k+1}\exp(\sigma/\theta_0)/[(k-1)(\alpha+k-1)(\alpha+1)(\theta_+-\theta_-)] \\ &\le \sigma^{2-k}\Gamma(k)\theta_+^{k+1}\exp(\sigma/\theta_-)/[(k-1)(\alpha+k-1)(\alpha+1)(\theta_+-\theta_-)]. \end{aligned}$$

Consider $\Pi_j = \mathrm{InverseGamma}(k_j, \sigma_j)$, $j = 1, 2, \cdots$, where $k_j = 2 + o(1) > 2$ and $\sigma_j = o(1)$ as $j \to \infty$. Clearly $\hat{\theta}_j(X) = (X+\sigma_j)/(\alpha+k_j-1) \to X/(\alpha+1) = T(X)$. In order to show the admissibility of T, Blyth's method (Theorem 4.3 in Shao's book) requires $[r_T(\Pi_j) - r^*(\Pi_j)]/\Pi_j(O) \to 0$, which is equivalent to $\sigma_j^{2-k_j} \to 0$. However, this is actually impossible! ☹ □

Exercise (4.48)**.** Let $X_1, \ldots, X_n$ be i.i.d. from Exponential(μ, θ), where θ is known and $\mu \in \mathbb{R}$ is unknown.
(a) show that $X_{(1)} - \theta\log 2/n$ is an MRIE of μ under the absolute error loss $L(\mu, a) = |a-\mu|$;
(b) show that $X_{(1)} - t$ is an MRIE under the loss function $L(\mu, a) = \mathbb{1}_{(t,\infty)}(|a-\mu|)$.

Proof. Recall from Exercise 2.20 and Exercise 2.59 that $X_{(1)} \sim \mathrm{Exponential}(\mu, \theta/n)$ is sufficient and complete. Since $D = (X_i - X_n)_{1\le i\le n-1}$ is ancillary, it is independent of $X_{(1)}$ by Basu's theorem. Clearly $X_{(1)}$ is a location invariant estimator of μ, so $X_{(1)} - u_*$ is an MRIE if $u_* \in \arg\min_{u\in\mathbb{R}}\mathbb{E}_0[L(0, X_{(1)}-u)]$.

(a) To minimize $\mathbb{E}_0|X_{(1)} - u|$, we should take u as the median of $X_{(1)} \sim \mathrm{Exponential}(0, \theta/n)$. Thus, solving $1/2 = \mathbb{P}_0(X_{(1)} > u_*) = \exp(-nu_*/\theta)$ leads to $u_* = \theta\log 2/n$.

(b) It's straightforward that $\mathbb{E}_0\mathbb{1}\{|X_{(1)}-u| > t\} = \mathbb{P}_0(|X_{(1)}-u| > t) = 1 - \mathbb{P}_0(u-t \le X_{(1)} \le u+t)$ attains its minimum if and only if $[u-t, u+t]$ is taken as $[0, 2t]$. That is, $u_* = t$. □

Exercise (4.58)**.** Let $X_1, \ldots, X_n$ be i.i.d. with a Lebesgue p.d.f. $f_\sigma(x) = (2/\sigma)[1-(x/\sigma)]\mathbb{1}_{(0,\sigma)}(x)$, where $\sigma > 0$ is an unknown scale parameter. Find Pitman's estimator of σ^h for $n = 2, 3, 4$.

Solution. The Pitman estimator of σ^h is given by

$$T_n = \frac{\int_0^\infty t^{n+h-1}\prod_{i=1}^n f_1(tX_i)\,\mathrm{d}t}{\int_0^\infty t^{n+2h-1}\prod_{i=1}^n f_1(tX_i)\,\mathrm{d}t} = \frac{\int_0^{1/X_{(n)}} t^{n+h-1}\prod_{i=1}^n (1-tX_i)\,\mathrm{d}t}{\int_0^{1/X_{(n)}} t^{n+2h-1}\prod_{i=1}^n (1-tX_i)\,\mathrm{d}t} =: \frac{U_n(n+h)}{U_n(n+2h)}.$$

Let $V_{0,n} = 1$ and $V_{k,n} = \sum_{1\le i_1<\cdots<i_k\le n}\prod_{j=1}^k X_{i_j}$ for $1 \le k \le n$. We have $\prod_{i=1}^n(1-tX_i) = \sum_{k=0}^n(-1)^k V_{k,n}t^k$, and $U_n(r) = \int_0^{1/X_{(n)}} t^{r-1}\prod_{i=1}^n(1-tX_i)\,\mathrm{d}t = \sum_{k=0}^n(-1)^k V_{k,n}/[(r+k)X_{(n)}^{r+k}]$. Hence, T_n is determined. ////

Exercise (4.69)**.** Let $X_1, \dots, X_n$ be i.i.d. from $\mathcal{N}(\mu, \sigma^2)$ with unknown $\theta = (\mu, \sigma^2)$. Show that $\bar{X}$ is a minimax estimator of μ under the loss function $L(\theta, a) = (a - \mu)^2/\sigma^2$ when $\Theta = \mathbb{R} \times (0, \infty)$.

Proof. Note that $\bar{X}$ has constant risks, since $R_{\bar{X}}(\theta) = \mathbb{E}[L(\theta, \bar{X})|\theta] = \mathbb{E}[(\bar{X} - \mu)^2|\theta]/\sigma^2 = 1/n$. Assume for contradiction that there exists an estimator T of μ such that $\sup_{\theta\in\Theta} R_T(\theta) < R_{\bar{X}}$, then we have $\sup_{\mu\in\mathbb{R}} \mathbb{E}[(T - \mu)^2|(\mu, \sigma_0^2)] < \sigma_0^2/n$ for, say, $\sigma_0^2 = 1$. Consider $\mu \sim \mathcal{N}(0, \frac{\sigma_0^2}{\kappa})$, whence $\mu|X \sim \mathcal{N}(\frac{n\bar{X}}{n+\kappa}, \frac{\sigma_0^2}{n+\kappa})$. Therefore, $\sup_{\mu\in\mathbb{R}} \mathbb{E}[(T - \mu)^2|\mu] \geq \mathbb{E}\{\mathbb{E}[(T - \mu)^2|\mu]\} = \mathbb{E}\{\mathbb{E}[(\mu - T)^2|X]\} \geq \mathbb{E}\{\mathrm{Var}(\mu|X)\} = \frac{\sigma_0^2}{n+\kappa}$. Letting $\kappa \searrow 0$ gives $\sup_{\mu\in\mathbb{R}} \mathbb{E}[(T - \mu)^2|\mu] \geq \sigma_0^2/n$, which is absurd. □

Exercise (4.77)**.** Let $X_1, \dots, X_n$ be i.i.d. from Exponential(a, θ) with a known θ and an unknown $a \in \mathbb{R}$. Under the squared error loss, show that $X_{(1)} - \theta/n$ is the unique minimax estimator of a.

Proof. From Example 4.12 in Shao's book, $T = X_{(1)} - \theta/n$ is Pitman's estimator of a. Clearly $X_{(1)}$ is location invariant with $\mathbb{E}_a|X_{(1)} - a|^3 < \infty$, so T is admissible according to Stein[i)]. By Theorem 4.7 and Theorem 4.13 in Shao's book, T is the unique minimax estimator. □

Exercise (4.127)**.** Let $X_1, \dots, X_n$ be i.i.d. such that $\log X_i \sim \mathcal{N}(\theta, \theta)$ with an unknown $\theta > 0$.

(a) Obtain the likelihood equation and show that one of the solutions of the likelihood equation is the unique MLE of θ.

(b) Using Theorem 4.17, obtain the asymptotic distribution of the MLE of θ.

Solution. (a) It is straightforward that the likelihood function is

$$L_n(\theta) = \prod_{i=1}^n \left[\frac{X_i^{-1}}{\sqrt{2\pi\theta}} \exp\left\{-\frac{(\log X_i - \theta)^2}{2\theta}\right\}\right] = \frac{\cancel{\prod_{i=1}^n X_i^{-1}}}{(2\pi\theta)^{n/2}} \exp\left\{-\frac{\sum_{i=1}^n (\log X_i)^2 + n\theta^2 - \cancel{2\theta\sum_{i=1}^n \log X_i}}{2\theta}\right\},$$

whose logarithm is

$$\ell_n(\theta) = -\frac{n}{2}\log(2\pi\theta) - \frac{\sum_{i=1}^n (\log X_i)^2}{2\theta} - \frac{n}{2}\theta.$$

Then the score function is given by

$$s_n(\theta) = \frac{\partial \ell_n(\theta)}{\partial\theta} = -\frac{n}{2\theta^2}\left[\theta - \frac{\sum_{i=1}^n (\log X_i)^2}{n} + \theta^2\right].$$

Solving the likelihood equation $s_n(\theta) = 0$ so that $\theta > 0$, one can derive the MLE

$$\hat{\theta}_n = \left[\sqrt{1 + 4\textstyle\sum_{i=1}^n (\log X_i)^2/n} - 1\right]/2.$$

Note that $s_n > 0$ on $(0, \hat{\theta}_n)$ and $s_n < 0$ on $(\hat{\theta}_n, \infty)$, so ℓ_n achieves its maximum exactly at $\hat{\theta}_n$.

(b) Since

$$\nabla s_n(\theta) = \frac{n}{2\theta^2} - \frac{\sum_{i=1}^n (\log X_i)^2}{\theta^3},$$

the Fisher information matrix can be calculated as

$$\mathcal{I}_n(\theta) = \mathbb{E}_\theta[-\nabla s_n(\theta)] = \frac{\sum_{i=1}^n \mathbb{E}_\theta[(\log X_i)^2]}{\theta^3} - \frac{n}{2\theta^2} = n\left(\frac{\theta^2 + \theta}{\theta^3} - \frac{1}{2\theta^2}\right) = n \cdot \frac{\theta + 1/2}{\theta^2}.$$

We have $[\mathcal{I}_n(\theta)]^{1/2}(\hat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}(0, 1)$, or $\hat{\theta}_n \overset{\circ}{\sim} \mathcal{N}(\theta, [\mathcal{I}_n(\theta)]^{-1})$. ////

[i)] Stein, Charles. The Admissibility of Pitman's Estimator of a Single Location Parameter. Ann. Math. Statist. 30 (1959), no. 4, 970–979. doi:10.1214/aoms/1177706080. `https://projecteuclid.org/euclid.aoms/1177706080`

Exercise (4.134)**.** Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be i.i.d. random 2-vectors. Suppose that both X_1 and Y_1 are binary, $\mathbb{P}(X_1 = 1) = \frac{1}{2}$, $\mathbb{P}(Y_1 = 1|X_1 = 0) = \exp(-a\theta)$, and $\mathbb{P}(Y_1 = 1|X_1 = 1) = \exp(-b\theta)$, where $\theta > 0$ is unknown and $a > 0$ and $b > 0$ are known constants.

(a) Suppose that (X_i, Y_i), $i = 1, \cdots, n$, are observed. Find the MLE of θ and its nondegenerate asymptotic distribution.

(b) Suppose that only $Y_1, \ldots, Y_n$ are observed. Find the MLE of θ and its nondegenerate asymptotic distribution.

(c) Calculate the asymptotic relative efficiency of the MLE in (a) w.r.t. the MLE in (b). How much efficiency is lost in the special case of $a = b$?

Solution. (a) Let $N_{xy} = \sum_{i=1}^n \mathbb{1}\{X_i = x, Y_i = y\}$. Then the log-likelihood function is

$$\ell_n(\theta) = N_{00}\log(1 - \mathrm{e}^{-a\theta}) - N_{01}a\theta + N_{10}\log(1 - \mathrm{e}^{-b\theta}) - N_{11}b\theta - n\log 2,$$

and the score function is

$$s_n(\theta) = \nabla\ell_n(\theta) = a[N_{00}/(\mathrm{e}^{a\theta} - 1) - N_{01}] + b[N_{10}/(\mathrm{e}^{b\theta} - 1) - N_{11}].$$

Note that

$$\nabla s_n(\theta) = -a^2\mathrm{e}^{a\theta}N_{00}/(\mathrm{e}^{a\theta} - 1)^2 - b^2\mathrm{e}^{b\theta}N_{10}/(\mathrm{e}^{b\theta} - 1)^2 \le 0.$$

The MLE $\hat{\theta}_n$ of θ is the root of s_n, which may be found numerically. We have $[\mathcal{I}_n(\theta)]^{1/2}(\hat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}(0, 1)$, where

$$\mathcal{I}_n(\theta) = \mathbb{E}_\theta[-\nabla s_n(\theta)] = n[a^2/(\mathrm{e}^{a\theta} - 1) + b^2/(\mathrm{e}^{b\theta} - 1)]/2.$$

(b) Let $M_y = \sum_{i=1}^n \mathbb{1}\{Y_i = y\}$. Then the log-likelihood function is

$$\ell_n(\theta) = M_0\log(2 - \mathrm{e}^{-a\theta} - \mathrm{e}^{-b\theta}) + M_1\log(\mathrm{e}^{-a\theta} + \mathrm{e}^{-b\theta}) - n\log 2,$$

and the score function is

$$s_n(\theta) = \nabla\ell_n(\theta) = [M_0/(2 - \mathrm{e}^{-a\theta} - \mathrm{e}^{-b\theta}) - M_1/(\mathrm{e}^{-a\theta} + \mathrm{e}^{-b\theta})](a\mathrm{e}^{-a\theta} + b\mathrm{e}^{-b\theta}).$$

The MLE $\hat{\theta}_n$ of θ is the root of s_n, which may be found numerically. We have $[\mathcal{I}_n(\theta)]^{1/2}(\hat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}(0, 1)$, where

$$\mathcal{I}_n(\theta) = \mathrm{Var}_\theta(s_n(\theta)) = n(a\mathrm{e}^{-a\theta} + b\mathrm{e}^{-b\theta})^2/[(2 - \mathrm{e}^{-a\theta} - \mathrm{e}^{-b\theta})(\mathrm{e}^{-a\theta} + \mathrm{e}^{-b\theta})].$$

(c) Clearly the asymptotic relative efficiency of the MLE in (a) w.r.t. the MLE in (b) is given by

$$\begin{aligned}\mathrm{ARE} &= \frac{[a^2/(\mathrm{e}^{a\theta} - 1) + b^2/(\mathrm{e}^{b\theta} - 1)]/2}{(a\mathrm{e}^{-a\theta} + b\mathrm{e}^{-b\theta})^2/[(2 - \mathrm{e}^{-a\theta} - \mathrm{e}^{-b\theta})(\mathrm{e}^{-a\theta} + \mathrm{e}^{-b\theta})]} \\ &\overset{(a=b)}{=} \frac{a^2/(\mathrm{e}^{a\theta} - 1)}{(a\mathrm{e}^{-a\theta})^2/[(1 - \mathrm{e}^{-a\theta})\mathrm{e}^{-a\theta}]} = 1,\end{aligned}$$

and thus no efficiency is lost in the case where $a = b$. ////

Exercise (4.140)**.** Let $X_1, \ldots, X_n$ be i.i.d. binary random variables with $\mathbb{P}(X_1 = 1) = p$, where $p \in (0, 1)$ is unknown. Let $\hat{\vartheta}_n$ be the MLE of $\vartheta = p(1 - p)$.

(a) Show that $\hat{\vartheta}_n$ is asymptotically normal when $p \ne \frac{1}{2}$.

(b) When $p = \frac{1}{2}$, derive a nondegenerate asymptotic distribution of $\hat{\vartheta}_n$.

Solution. From Example 4.29 in Shao's book, $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ is the unique MLE of p, and thus we have $\hat{\vartheta}_n = \bar{X}_n(1 - \bar{X}_n)$. By the CLT, $\sqrt{n}(\bar{X}_n - p) \xrightarrow{d} \sqrt{p(1-p)}Z$, where $Z \sim \mathcal{N}(0, 1)$. It remains to apply the delta method. Note that $g(p) = p(1 - p)$ has derivatives $g'(p) = 1 - 2p$ and $g''(p) = -2$.

(a) It's immediate that

$$\sqrt{n}(\hat{\vartheta}_n - \vartheta) = \sqrt{n}(g(\bar{X}_n) - g(p)) \xrightarrow{d} g'(p)\sqrt{p(1-p)}Z \sim \mathcal{N}(0, p(1-p)(1-2p)^2).$$

(b) Since $g'(p) = 0$, we turn to the second order. Clearly $\vartheta = 1/4$ and

$$n(\hat{\vartheta}_n - \vartheta) = n(g(\bar{X}_n) - g(p)) \xrightarrow{d} \tfrac{1}{2}g''(p)p(1-p)Z^2 = -\tfrac{1}{4}Z^2,$$

where $Z^2 \sim \chi^2_1$. ////

Exercise (6.2)**.** Consider $H_0 : P = P_0$ versus $H_1 : P = P_1$. Suppose that P_j has the p.d.f. f_j w.r.t. a σ-finite measure ν, where $j = 0, 1$. Let $\beta(P)$ be the power function of a UMP test T of size $\alpha \in (0, 1)$. Show that $\alpha < \beta(P_1)$ unless $P_0 = P_1$.

Proof. Note that $T(X)$ is at least as good as $T_0(X) \equiv \alpha$, and thus $\beta(P_1) \geq \mathbb{E}_1[T_0(X)] = \alpha$. If the equality holds, then T_0 is also UMP. By the uniqueness part of Theorem 6.1 in Shao's book, we have $f_1 \xlongequal{\nu\text{-a.e.}} cf_0$ for a constant c, since $0 < \alpha < 1$. But f_j's are p.d.f.'s, which entails $c = 1$. This leads to $P_0 = P_1$. □

Exercise (6.4)**.** Let H_0 and H_1 be simple hypotheses and let $\alpha \in (0, 1)$. Suppose that T_* is a UMP test of size α for testing H_0 versus H_1 and that $\beta < 1$, where β is the power of T_* when H_1 is true. Show that $1 - T_*$ is a UMP test of size $1 - \beta$ for testing H_1 versus H_0.

Proof. Suppose that the sample X has the p.d.f. f_j w.r.t. a σ-finite measure ν when H_j is true, $j = 0, 1$. By the uniqueness of the UMP test,

$$T_*(X) \xlongequal{\text{a.s.}} \mathbb{1}\{f_1(X) > cf_0(X)\} + \delta(X)\mathbb{1}\{f_1(X) = cf_0(X)\}$$

for some constant $c \geq 0$ and some statistic δ. Using $0 < \alpha \leq \beta < 1$, we have $c \in (0, \infty)$. It follows that

$$1 - T_*(X) \xlongequal{\text{a.s.}} \mathbb{1}\{f_0(X) > c^{-1}f_1(X)\} + [1 - \delta(X)]\mathbb{1}\{f_0(X) = c^{-1}f_1(X)\}$$

is a UMP test for testing H_1 versus H_0 by the Neyman–Pearson lemma (Theorem 6.1 in Shao's book). □

Exercise (6.5)**.** Let X be a sample of size 1 from a Lebesgue p.d.f. f_θ. Find a UMP test of size $\alpha \in (0, \frac{1}{2})$ for $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$ when
(a) $f_\theta(x) = 2\theta^{-2}(\theta - x)\mathbb{1}_{(0,\theta)}(x)$, $\theta_0 < \theta_1$;
(c) f_{θ_0} is the p.d.f. of $\mathcal{N}(0, 1)$ and f_{θ_1} is the p.d.f. of Cauchy$(0, 1)$;

Solution. By the Neyman–Pearson lemma, it suffices to determine a test T of the form

$$T(X) = \mathbb{1}\{f_{\theta_1}(X) > cf_{\theta_0}(X)\}, \text{ where } c \geq 0 \text{ is a constant such that } \mathbb{E}_{\theta_0}[T(X)] = \alpha.$$

(a) Since $\frac{f_{\theta_1}(x)}{f_{\theta_0}(x)} = \frac{\theta_0^2(\theta_1 - x)}{\theta_1^2(\theta_0 - x)}\mathbb{1}_{(0,\theta_0)}(x) + \infty\mathbb{1}_{[\theta_0,\theta_1)}(x)$ increases with x, we have $T(X) = \mathbb{1}\{X > x_\alpha\}$ for some constant $x_\alpha \in (0, \theta_0)$. Note that $\alpha = \mathbb{E}_{\theta_0}[T(X)] = \mathbb{P}_{\theta_0}\{X > x_\alpha\} = (1 - \frac{x_\alpha}{\theta_0})^2$, so $x_\alpha = (1 - \sqrt{\alpha})\theta_0$.

(c) Since $\frac{f_{\theta_1}(x)}{f_{\theta_0}(x)} = \sqrt{\frac{2}{\pi}} \cdot \frac{\exp(x^2/2)}{1+x^2}$ increases with $|x| \in (1, \infty)$ and decreases with $|x| \in (0, 1)$, we have

$$T(X) = \mathbb{1}\{|X| > z_\alpha^+\} + \mathbb{1}\{|X| < z_\alpha^-\}$$

for some constants $z_\alpha^+ \in (1, \zeta)$ and $z_\alpha^- \in (0, 1)$ such that $\frac{f_{\theta_1}(z_\alpha^+)}{f_{\theta_0}(z_\alpha^+)} = \frac{f_{\theta_1}(z_\alpha^-)}{f_{\theta_0}(z_\alpha^-)}$, or

$$T(X) = \mathbb{1}\{|X| > z_\alpha\}$$

for some constant $z_\alpha \in [\zeta, \infty)$, where $\zeta \approx 1.5852$ satisfies $\exp(\zeta^2/2) = 1 + \zeta^2$. Then $z_\alpha^\pm$ can be obtained by solving $\alpha = \mathbb{P}_{\theta_0}\{|X| > z_\alpha^+\} + \mathbb{P}_{\theta_0}\{|X| < z_\alpha^-\} = 2(1 - \Phi(z_\alpha^+)) + (2\Phi(z_\alpha^-) - 1) = 1 - 2\Phi(z_\alpha^+) + 2\Phi(z_\alpha^-)$, and $z_\alpha = \Phi^{-1}(1 - \frac{\alpha}{2})$ solves $\alpha = \mathbb{P}_{\theta_0}\{|X| > z_\alpha\} = 2(1 - \Phi(z_\alpha)) \leq 2(1 - \Phi(\zeta)) \approx 0.1129$. ////

Exercise (6.14)**.** Let $X_1, \ldots, X_n$ be i.i.d. from a Lebesgue p.d.f. f_θ, where $\theta \in \Theta \subset \mathbb{R}$. Find a UMP test of size α for testing $H_0 : \theta \le \theta_0$ versus $H_1 : \theta > \theta_0$ when
(a) $f_\theta(x) = \theta^{-1}\mathrm{e}^{-x/\theta}\mathbb{1}_{(0,\infty)}(x)$, $\theta > 0$;
(c) f_θ is the p.d.f. of $\mathcal{N}(1, \theta)$;

Solution. We will show that $\{f_\theta\}_{\theta\in\Theta}$ has monotone likelihood ratio in some real-valued statistic $Y(X)$, and thus the Karlin–Rubin theorem (Theorem 6.2 in Shao's book) can come into play. Then it suffices to determine a test T of the form $T(X) = \mathbb{1}\{Y(X) > c\}$, where c is a constant such that $\mathbb{E}_{\theta_0}[T(X)] = \alpha$.

(a) Note that if $\theta_1 < \theta_2$, then $\frac{f_{\theta_2}(x)}{f_{\theta_1}(x)} = \frac{\theta_1^n}{\theta_2^n}\exp\{(\frac{1}{\theta_1} - \frac{1}{\theta_2})\sum_{i=1}^n x_i\}$ increases with $Y(x) := \sum_{i=1}^n x_i$. Since $Y(X) \sim \mathrm{Gamma}(n, \theta) = \frac{\theta}{2}\chi^2_{2n}$, we may take $c = \frac{\theta_0}{2}\chi^2_{2n,\alpha}$, where $\chi^2_{r,\alpha}$ is the $(1-\alpha)$th quantile of χ^2_r.

(c) Note that if $\theta_1 < \theta_2$, then $\frac{f_{\theta_2}(x)}{f_{\theta_1}(x)} = (\frac{\theta_1}{\theta_2})^{n/2}\exp\{\frac{1}{2}(\frac{1}{\theta_1} - \frac{1}{\theta_2})\sum_{i=1}^n (x_i-1)^2\}$ is an increasing function of $Y(x) := \sum_{i=1}^n (x_i - 1)^2$. Since $Y(X) \sim \theta\chi^2_n$, we may take $c = \theta_0\chi^2_{n,\alpha}$. ////

Exercise (6.28)**.** Suppose that X has the p.d.f.

$$f_\theta(x) = \exp\{\eta(\theta)Y(x) - \xi(\theta)\}h(x),$$

where $\eta(\theta)$ is a nondecreasing function of $\theta \in \Theta \subset \mathbb{R}$. Consider hypotheses

$$H_0 : \theta_1 \le \theta \le \theta_2 \quad \text{versus} \quad H_1 : \theta < \theta_1 \text{ or } \theta > \theta_2,$$

or

$$H_0 : \theta = \theta_0 \quad \text{versus} \quad H_1 : \theta \ne \theta_0.$$

Show that a UMP test does not exist.

Proof. We may consider only the former testing problem, since the latter one is the special case where $\theta_1 = \theta_2 = \theta_0$. Assume for contradiction that T_* is a UMP test for testing H_0 versus H_1 with size α. Note that $\{f_\theta\}_{\theta\in\Theta}$ has monotone likelihood ratio in $Y(X)$, so there exist some UMP tests T_1 and T_2 for testing $\theta \ge \theta_1$ versus $\theta < \theta_1$ and $\theta \le \theta_2$ versus $\theta > \theta_2$, respectively, with size α. Since $\mathbb{E}_\theta[T_1(X)] \le \alpha$ for $\theta \in [\theta_1, \infty) \supset [\theta_1, \theta_2]$, we have $\mathbb{E}_\theta[T_*(X)] \ge \mathbb{E}_\theta[T_1(X)]$ for $\theta \in (-\infty, \theta_1) \cup (\theta_2, \infty) \supset (-\infty, \theta_1)$. It follows from the uniqueness of the UMP test (Theorem 6.2(iv) in Shao's book) that $T_* \overset{\text{a.s.}}{=\!=} T_1$, which has the power function $\beta(\theta) = \mathbb{E}_\theta[T_*(X)]$ that is strictly *decreasing* for θ such that $\beta(\theta) \in (0, 1)$ by Theorem 6.2(ii) in Shao's book. By analogy, $T_* \overset{\text{a.s.}}{=\!=} T_2$ and $\beta(\theta)$ is strictly *increasing* for θ such that $\beta(\theta) \in (0, 1)$. This is absurd, unless $\beta(\theta) \equiv 0$ or $\beta(\theta) \equiv 1$, which implies uninteresting $\alpha \in \{0, 1\}$. □

Exercise (6.41)**.** Let (X_0, X_1, X_2) be a random vector following Multinomial$(n; p_0, p_1, p_2)$ with unknown $p_1, p_2 \in (0, 1)$. Derive a UMPU test of size α for testing $H_0 : p_0 = p^2,\ p_1 = 2p(1-p),\ p_2 = (1-p)^2$ versus $H_1 : H_0$ is not true, where $p \in (0, 1)$ is unknown.

Solution. Note that X has the p.m.f.

$$x \mapsto \binom{n}{x_0, x_1, x_2} p_0^{x_0}p_1^{x_1}p_2^{x_2} = \binom{n}{x_0, x_1, x_2} p_1^n \exp\left\{x_0 \log\left(\frac{p_0p_2}{p_1^2}\right) + (x_2 - x_0)\log\left(\frac{p_2}{p_1}\right)\right\}.$$

Since H_0 is equivalent to the hypothesis that

$$\log(p_0p_2/p_1^2) = -\log(4),$$

it follows from Theorem 6.4(iv) in Shao's book that a UMPU test of size α is given by

$$\begin{aligned} T_*(X) &= \mathbb{1}\{X_0 < c_1(X_2 - X_0)\} + \mathbb{1}\{X_0 > c_2(X_2 - X_0)\} \\ &\quad + \gamma_1(X_2 - X_0)\mathbb{1}\{X_0 = c_1(X_2 - X_0)\} + \gamma_2(X_2 - X_0)\mathbb{1}\{X_0 = c_2(X_2 - X_0)\}, \end{aligned}$$

where c_j's and γ_j's are Borel functions determined by

$$\mathbb{E}_0[T_*(X)|X_2-X_0]=\alpha \quad\text{and}\quad \mathbb{E}_0[T_*(X)X_0|X_2-X_0]=\alpha\mathbb{E}_0[X_0|X_2-X_0],$$

where $\mathbb{E}_0$ is the expectation under H_0. One can see that

$$\mathbb{P}(X_0=x_0|X_2-X_0=u)=\frac{\mathbb{1}_A(x_0)}{Z_u}\frac{(p_0p_2/p_1^2)^{x_0}}{x_0!(n-2x_0-u)!(x_0+u)!},$$

where $A=\{0,1,\cdots,n\}\cap[-u,\frac{n-u}{2}]$, and Z_u is the constant such that $\sum_{x_0}\mathbb{P}(X_0=x_0|X_2-X_0=u)=1$, which depends on (p_0,p_1,p_2) through $\log(p_0p_2/p_1^2)$. Then everything is ready. ////

Exercise (6.44)**.** Let X_j, $j=1,2,3$, be independent from Poisson(λ_j), $j=1,2,3$, respectively. Show that there exists a UMPU test of size α for testing $H_0:\lambda_1\lambda_2\le\lambda_3^2$ versus $H_1:\lambda_1\lambda_2>\lambda_3^2$.

Proof. Note that X has the p.m.f.

$$x\mapsto \mathrm{e}^{-\lambda_1-\lambda_2-\lambda_3}\frac{\lambda_1^{x_1}\lambda_2^{x_2}\lambda_3^{x_3}}{x_1!x_2!x_3!}=\frac{\mathrm{e}^{-\lambda_1-\lambda_2-\lambda_3}}{x_1!x_2!x_3!}\exp\left\{x_1\log\left(\frac{\lambda_1\lambda_2}{\lambda_3^2}\right)+(x_2-x_1)\log(\lambda_2)+(x_3+2x_1)\log(\lambda_3)\right\}.$$

Since H_0 is equivalent to $\log(\lambda_1\lambda_2/\lambda_3^2)\le 0$, it follows from Theorem 6.4(i) in Shao's book that a UMPU test of size α is given by $T_*(X)=\mathbb{1}\{X_1>c(U)\}+\gamma(U)\mathbb{1}\{X_1=c(U)\}$, where $U=(X_2-X_1,X_3+2X_1)$, and c and γ are Borel functions determined by $\mathbb{E}_0[T_*(X)|U]=\alpha$, where $\mathbb{E}_0$ is the expectation under H_0. □

Exercise (6.50)**.** Suppose that we have two independent samples, $X_{i1},X_{i2},\dots,X_{in}\overset{\text{i.i.d.}}{\sim}\mathcal{N}(\mu_i,\sigma_i^2)$, $i=1,2$, where $n\ge 2$. Let $S_i^2=\frac{1}{n-1}\sum_{j=1}^n(X_{ij}-\bar{X}_i)^2$ where $\bar{X}_i=\frac{1}{n}\sum_{j=1}^n X_{ij}$. Show that a UMPU test of size α for testing $H_0:\sigma_2^2=\Delta_0\sigma_1^2$ versus $H_1:\sigma_2^2\ne\Delta_0\sigma_1^2$

$$\text{rejects } H_0 \text{ when } \max\left(\frac{S_2^2}{\Delta_0S_1^2},\frac{\Delta_0S_1^2}{S_2^2}\right)>\frac{1-c}{c},$$

where c satisfies $\int_0^c f_{(n-1)/2,(n-1)/2}(v)\,\mathrm{d}v=\alpha/2$ and $f_{a,b}$ is the p.d.f. of Beta(a,b).

Proof. Note that the joint p.d.f. of X_{ij}'s is

$$x\mapsto\frac{1}{(2\pi)^n(\sigma_1^2\sigma_2^2)^{n/2}}\exp\left(-\sum_{i=1}^2\frac{1}{2\sigma_i^2}\sum_{j=1}^n x_{ij}^2+\sum_{i=1}^2\frac{n\mu_i}{\sigma_i^2}\bar{x}_i-\sum_{i=1}^2\frac{n\mu_i^2}{2\sigma_i^2}\right)=\exp\{\theta Y(x)+\varphi\cdot U(x)-\zeta(\theta,\varphi)\},$$

where

$$\theta=\frac{1}{2\Delta_0\sigma_1^2}-\frac{1}{2\sigma_2^2},\ \varphi=\left(-\frac{1}{2\sigma_1^2},\frac{n\mu_1}{\sigma_1^2},\frac{n\mu_2}{\sigma_2^2}\right),\ Y(x)=\sum_{j=1}^n x_{2j}^2,\ U(x)=\left(\sum_{j=1}^n x_{1j}^2+\frac{1}{\Delta_0}\sum_{j=1}^n x_{2j}^2,\bar{x}_1,\bar{x}_2\right),$$

and $\zeta(\theta,\varphi)=n\log(2\pi)+\frac{n}{2}\sum_{i=1}^2[\frac{\mu_i^2}{\sigma_i^2}+\log(\sigma_i^2)]$. To apply Lemma 6.7(ii) in Shao's book, consider

$$V=\frac{F}{1+F}=\frac{S_2^2}{\Delta_0S_1^2+S_2^2}=\frac{Y-nU_3^2}{\Delta_0U_1-n\Delta_0U_2^2-nU_3^2},$$

where $F=S_2^2/(\Delta_0S_1^2)$, and U_k denotes the k^{th} component of U for $k=1,2,3$. Under $H_0:\theta=0$, one can see that U is complete and sufficient (for φ), $F\sim F_{n-1,n-1}$ and $V\sim$ Beta$(\frac{n-1}{2},\frac{n-1}{2})$. By Basu's theorem, V and U are independent when H_0 is true. Therefore, a UMPU test is given by

$$T_*=\mathbb{1}\{V<c_1\}+\mathbb{1}\{V>c_2\}=\mathbb{1}\{F<c_1/(1-c_1)\}+\mathbb{1}\{F>c_2/(1-c_2)\},$$

where c_i's are constants determined by $\mathbb{E}_0[T_*]=\alpha$ and $\mathbb{E}_0[T_*V]=\alpha\mathbb{E}_0[V]$, where $\mathbb{E}_0$ is the expectation under H_0. This leads to $c_1=c$ and $c_2=1-c$, using $f_{a,b}(v)=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}v^{a-1}(1-v)^{b-1}\mathbb{1}_{(0,1)}(v)$. It follows that $T_*=\mathbb{1}\{\max(1/F,F)>(1-c)/c\}$. □

Exercise (6.85)**.** Let $X_1, \ldots, X_n$ be i.i.d. from the discrete uniform distribution on $\{1, \cdots, \theta\}$, where θ is an integer ≥ 2. Find a level α LR test for
(a) $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$, where θ_0 is a known integer ≥ 2;
(b) $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$.

Solution. Note that the likelihood function is given by $L(\theta) = \prod_{i=1}^n (\theta^{-1}\mathbb{1}\{X_i \leq \theta\}) = \theta^{-n}\mathbb{1}\{X_{(n)} \leq \theta\}$, where $X_{(n)} = \max_{1\leq i\leq n} X_i$. Clearly the unrestricted MLE of θ is $X_{(n)}$.

(a) Under H_0 the restricted maximum likelihood is $L(X_{(n)})\mathbb{1}\{X_{(n)} \leq \theta_0\}$. Thus, the LR test has the rejection region $\{\mathbb{1}\{X_{(n)} \leq \theta_0\} \leq c\} = \{X_{(n)} > \theta_0\}$, where $c \in (0, 1)$ is constant. Its significance size is 0.

(b) Under H_0 the restricted maximum likelihood is $L(\theta_0)$. Thus, the LR test has the rejection region of the form $\{\theta_0^{-n}\mathbb{1}\{X_{(n)} \leq \theta_0\}/X_{(n)}^{-n} \leq c\} = \{X_{(n)} \leq c^{1/n}\theta_0\} \cup \{X_{(n)} > \theta_0\}$, where $c \in (0, 1)$ is constant. Its significance size is $(\lfloor c^{1/n}\theta_0\rfloor/\theta_0)^n \leq \alpha$, and here we require c to satisfy that $\lfloor c^{1/n}\theta_0\rfloor \leq \alpha^{1/n}\theta_0$. ////

Exercise (6.92)**.** Let X_1 and X_2 be independently distributed as Poisson(λ_1) and Poisson(λ_2), respectively. Find an LR test of significance level α for testing
(a) $H_0 : \lambda_1 = \lambda_2$ versus $H_1 : \lambda_1 \neq \lambda_2$;
(b) $H_0 : \lambda_1 \geq \lambda_2$ versus $H_1 : \lambda_1 < \lambda_2$. (Is this test a UMPU test?)

Solution. Note that the likelihood function is given by $L(\lambda_1, \lambda_2) = \mathrm{e}^{-\lambda_1-\lambda_2}\lambda_1^{X_1}\lambda_2^{X_2}/(X_1!X_2!)$. Clearly the unrestricted MLE of (λ_1, λ_2) is (X_1, X_2).

(a) Under H_0 the restricted MLE of each of λ_1 and λ_2 is $\bar{X} = (X_1 + X_2)/2$. Thus, the LR test has the rejection region of the form $\{L(\bar{X}, \bar{X})/L(X_1, X_2) \leq c\} = \{\bar{X}^{X_1+X_2}/(X_1^{X_1}X_2^{X_2}) \leq c\}$, where $c \in (0, 1)$ is constant and such that the probability of a type I error doesn't exceed α for any $\lambda_1 = \lambda_2$.

(b) Under H_0 the restricted MLE of (λ_1, λ_2) is (X_1, X_2) if $X_1 \geq X_2$ and $(\bar{X}, \bar{X})$ otherwise. Thus, the LR test has the rejection region of the form

$$\{L(X_1, X_2)\mathbb{1}\{X_1 \geq X_2\}+L(\bar{X}, \bar{X})\mathbb{1}\{X_1 < X_2\} \leq c_0L(X_1, X_2)\} = \{\bar{X}^{X_1+X_2}/(X_1^{X_1}X_2^{X_2}) \leq c_0\}\cap\{X_1 < X_2\},$$

where $c_0 \in (0, 1)$ is constant and such that the probability of a type I error doesn't exceed α for any $\lambda_1 \geq \lambda_2$. Since $L(\lambda_1, \lambda_2) = \exp\{X_1 \log(\frac{\lambda_1}{\lambda_2}) + (X_1 + X_2) \log(\lambda_2) - (\lambda_1 + \lambda_2)\}/(X_1!X_2!)$, a nonrandomized UMPU test is $T_* = \mathbb{1}\{X_1 \leq c(X_1 + X_2)\}$ for some Borel function c, by Theorem 6.4 in Shao's book. In order for the LR test to be a UMPU test, it suffices that $\psi(y) = u \log(\frac{u}{2}) - y \log(y) - (u - y) \log(u - y)$ is increasing on $[0, \frac{u}{2}]$ for any positive integer u. Indeed, $\psi'(y) = \log(\frac{u-y}{y}) > 0$ for $y \in [0, \frac{u}{2})$. This leads to an affirmative answer. Moreover, c can be obtained using $X_1|X_1 + X_2 = u \sim \text{Binomial}(\frac{1}{2}, u)$ when $\lambda_1 = \lambda_2$. ////

Exercise (6.96)**.** Let $X_{i1}, \ldots, X_{in_i} \overset{\text{i.i.d.}}{\sim} \text{Beta}(\theta_i, 1)$, $i = 1, 2$, be independent. For testing $H_0 : \theta_1 = \theta_2$ versus $H_1 : \theta_1 \neq \theta_2$, find the forms of the LR test, Wald's test, and Rao's score test.

Solution. Let $T_i = -\sum_{j=1}^{n_i} \log(X_{ij}) \sim \chi^2_{2n_i}/(2\theta_i)$ for $i = 1, 2$. Then the log-likelihood function is given by $\ell(\theta_1, \theta_2) = \sum_{i=1}^2\sum_{j=1}^{n_i} \log(\theta_i X_{ij}^{\theta_i-1}) = \sum_{i=1}^2[n_i \log(\theta_i) - (\theta_i - 1)T_i]$, the score function is $s(\theta_1, \theta_2) = \nabla\ell(\theta_1, \theta_2) = (n_i/\theta_i-T_i)_{i=1}^2$, and the Fisher information matrix is $\mathcal{I}(\theta_1, \theta_2) = -\mathbb{E}\,\nabla s(\theta_1, \theta_2) = \text{diag}(\{n_i/\theta_i^2\}_{i=1}^2)$. Clearly the unrestricted MLE of θ_i is $\hat{\theta}_i = n_i/T_i$ for $i = 1, 2$, and under H_0 the restricted MLE of $\theta_1 = \theta_2$ is $\tilde{\theta} = (n_1 + n_2)/(T_1 + T_2)$. They satisfy that $s(\hat{\theta}_1, \hat{\theta}_2) = 0_2$ and $s(\tilde{\theta}, \tilde{\theta}) = 0_2$. Denote $d = (1, -1)^\top$. We have

- the LR test: $\{2[\ell(\hat{\theta}_1, \hat{\theta}_2) - \ell(\tilde{\theta}, \tilde{\theta})] > c\} = \{n_1 \log(\frac{n_1}{T_1}) + n_2 \log(\frac{n_2}{T_2}) - (n_1 + n_2) \log(\frac{n_1+n_2}{T_1+T_2}) > \frac{c}{2}\}$
- Wald's test: $\{(\hat{\theta}_1 - \hat{\theta}_2)^2/(d^\top\mathcal{I}(\hat{\theta}_1, \hat{\theta}_2)^{-1}d) > c\} = \{(\frac{n_1}{T_1} - \frac{n_2}{T_2})^2 > c(\frac{n_1}{T_1^2} + \frac{n_2}{T_2^2})\}$
- Rao's score test: $\{s(\tilde{\theta}, \tilde{\theta})^\top\mathcal{I}(\tilde{\theta}, \tilde{\theta})^{-1}s(\tilde{\theta}, \tilde{\theta}) > c\} = \{(n_1 + n_2)(n_1T_2 - n_2T_1)^2 > cn_1n_2(T_1 + T_2)^2\}$

where the critical values $c \in [0, \infty)$ may be different line by line. ////

Exercise (7.1)**.** Let $X_{i1}, \ldots, X_{in_i} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu_i, \sigma_i^2)$, $i = 1, 2$, be two independent samples, where all parameters are unknown. Let $\bar{X}_i$ and S_i^2 be the sample mean and the sample variance of the i^{th} sample, $i = 1, 2$.
(a) Let $\theta = \mu_1 - \mu_2$. Assume that $\sigma_1 = \sigma_2$. Show that

$$t(X, \theta) = \frac{(\bar{X}_1 - \bar{X}_2 - \theta)/\sqrt{1/n_1 + 1/n_2}}{\sqrt{[(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2]/(n_1 + n_2 - 2)}}$$

is a pivotal quantity and construct a confidence interval for θ with confidence coefficient $1 - \alpha$, using $t(X, \theta)$.
(b) Let $\theta = \sigma_2^2/\sigma_1^2$. Show that $\mathrm{Re}(X, \theta) = S_2^2/(\theta S_1^2)$ is a pivotal quantity and construct a confidence interval for θ with confidence coefficient $1 - \alpha$, using $\mathrm{Re}(X, \theta)$.

Solution. Note that $\bar{X}_i \sim \mathcal{N}(\mu_i, \sigma_i^2/n_i)$ and $S_i^2/\sigma_i^2 \sim \chi_{n_i - 1}^2/(n_i - 1)$ are independent.

(a) Since $\bar{X}_1 - \bar{X}_2 \sim \mathcal{N}(\mu_1 - \mu_2, \sigma_1^2/n_1 + \sigma_2^2/n_2)$ and $(n_1 - 1)S_1^2/\sigma_1^2 + (n_2 - 1)S_2^2/\sigma_2^2 \sim \chi_{n_1 + n_2 - 2}^2$ are independent, we have $t(X, \theta) \sim t_{n_1 + n_2 - 2}$, and thus a $(1 - \alpha)$ confidence interval for θ is given by

$$\bar{X}_1 - \bar{X}_2 \pm t_{n_1 + n_2 - 2, \frac{\alpha}{2}} \sqrt{(1/n_1 + 1/n_2)[(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2]/(n_1 + n_2 - 2)},$$

where $t_{r, \frac{\alpha}{2}}$ denotes the $(1 - \frac{\alpha}{2})^{\text{th}}$ quantile of t_r.

(b) We have $\mathrm{Re}(X, \theta) = (S_2^2/\sigma_2^2)/(S_1^2/\sigma_1^2) \sim F_{n_2 - 1, n_1 - 1}$, and thus a $(1 - \alpha)$ confidence interval for θ is given by $[S_2^2/(c_2 S_1^2), S_2^2/(c_1 S_1^2)]$, where c_1 and c_2 denote the $(\frac{\alpha}{2})^{\text{th}}$ and the $(1 - \frac{\alpha}{2})^{\text{th}}$ quantiles of $F_{n_2 - 1, n_1 - 1}$, respectively. ////

Exercise (7.2)**.** Let X_i, $i = 1, 2$, be independent with the p.d.f.'s $\lambda_i \mathrm{e}^{-\lambda_i x} \mathbb{1}_{(0, \infty)}(x)$, $i = 1, 2$, respectively.
(a) Let $\theta = \lambda_1/\lambda_2$. Show that $\theta X_1/X_2$ is a pivotal quantity and construct a confidence interval for θ with confidence coefficient $1 - \alpha$, using this pivotal quantity.
(b) Let $\theta = (\lambda_1, \lambda_2)$. Show that $\lambda_1 X_1 + \lambda_2 X_2$ is a pivotal quantity and construct a confidence set for θ with confidence coefficient $1 - \alpha$, using this pivotal quantity.

Solution. Note that $X_i \sim \chi_2^2/(2\lambda_i)$ for $i = 1, 2$.

(a) We have $\theta X_1/X_2 = (\lambda_1 X_1)/(\lambda_2 X_2) \sim F_{2,2}$, and thus a $(1 - \alpha)$ confidence interval for θ is given by $[c_1 X_2/X_1, c_2 X_2/X_1]$, where c_1 and c_2 denote the $(\frac{\alpha}{2})^{\text{th}}$ and the $(1 - \frac{\alpha}{2})^{\text{th}}$ quantiles of $F_{2,2}$, respectively.

(b) We have $\lambda_1 X_1 + \lambda_2 X_2 \sim \chi_4^2/2$, and thus a $(1 - \alpha)$ confidence set for θ is given by $\{(\lambda_1, \lambda_2) : \lambda_1 X_1 + \lambda_2 X_2 \in [c_1/2, c_2/2]\}$, where c_1 and c_2 denote the $(\frac{\alpha}{2})^{\text{th}}$ and the $(1 - \frac{\alpha}{2})^{\text{th}}$ quantiles of χ_4^2, respectively. ////