Homework 1 Due: Oct 10th, 2021

1. For $x_1, \dots, x_n \in \mathbb{R}$, let $\bar{x}$ and $\tilde{x}$ be the sample mean and the sample median, respectively. Show that $\sum_{i=1}^n (x_i - \bar{x})^2 \le \sum_{i=1}^n (x_i - c)^2$ and $\sum_{i=1}^n |x_i - \tilde{x}| \le \sum_{i=1}^n |x_i - c|$ for $c \in \mathbb{R}$.

2. Given $(x_1, y_1), \dots, (x_n, y_n) \in \mathbb{R}^2$, find $(\hat{\alpha}, \hat{\beta}) = \arg\min_{(\alpha,\beta)\in\mathbb{R}^2} \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2$ provided that x_i's are not all the same. Then show that $\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$, where $\bar{y} = \frac{1}{n}\sum_{i=1}^n y_i$ is the sample mean of y_i's and $\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i$ are fitted values.

3. In survival analysis, we are interested in the probability distribution of time spans. The hazard function h associated with a random variable T is defined by $h(t) = \lim_{\delta \searrow 0} \frac{1}{\delta} \mathbb{P}(t < T \le t + \delta \mid T > t)$. Assume that T has a continuous distribution with probability density function f.
   (a) Express h in terms of f and F, where $F(t) = \int_{-\infty}^t f(u)\,\mathrm{d}u$ is the cumulative distribution function.
   (b) Express f in terms of h and H, where $H(t) = \int_{-\infty}^t h(u)\,\mathrm{d}u$ is the cumulative hazard function.

4. Consider a sample of i.i.d. random variables $X_1, \dots, X_n$, each from a continuous distribution with cumulative distribution function F and probability density function f. Let $X_{(k)}$ be the k^{th} order statistic. By expressing the probability density functions, determine
   (a) the joint distribution of $X_{(1)}, \dots, X_{(n)}$,
   (b) the joint distribution of $X_{(i)}$ and $X_{(j)}$, where $1 \le i < j \le n$, and
   (c) the distribution of $X_{(k)}$.

   Then relate the uniform distribution on $[0, 1]$ to the beta distribution with integer parameters.

5. Show that the logarithm of the gamma function is convex as follows. Consider $X \sim \text{Gamma}(\alpha, 1)$, whose probability density function is given by $x \mapsto \frac{1}{\Gamma(\alpha)} x^{\alpha-1} \mathrm{e}^{-x} \mathbb{1}_{[x>0]}$. Justify that $\frac{\mathrm{d}^2}{\mathrm{d}\alpha^2} \log \Gamma(\alpha) = \text{Var}(\log X)$.

6. Let $X_n = \frac{1}{n}\sum_{i=1}^n Z_i^2$ where $Z_1, \dots, Z_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$. Show that $\mathbb{P}\{|X_n - 1| \ge u\} \le 2\,\mathrm{e}^{-nu^2/8}$ for $u \in (0, 1)$.

7. Let $X \sim \mathcal{N}(0, 1)$ and $Q \sim \chi_1^2$ be independent. Show that $X/\sqrt{Q}$ has the standard Cauchy distribution, whose probability density function is given by $t \mapsto \frac{1}{\pi}\frac{1}{1+t^2}$.

8. Let X be a vector consisting of i.i.d. random variables $X_1, \dots, X_n$, where $n \ge 2$. Show that X is normally distributed with zero mean if and only if $AX \overset{d}{=} X$ for any constant $n \times n$ orthogonal matrix A.

9. Let $\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}_{k+\ell}\left(\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{pmatrix} \right)$, where $X \sim \mathcal{N}_k(\mu_X, \Sigma_{XX})$.
   (a) Find the joint distribution of X and $Y - BX$, where B is a constant $\ell \times k$ matrix.
   (b) Derive the conditional distribution of Y given $X = x$.

10. Let $X_{ij} = \mu + \alpha_i + \varepsilon_{ij}$, $1 \le i \le m$, $1 \le j \le n$, where α_i's and ε_{ij}'s are independent random variables, $\alpha_i \sim \mathcal{N}(0, \sigma_\alpha^2)$ and $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$. Denote $\bar{X}_i = \frac{1}{n}\sum_{j=1}^n X_{ij}$ and $\bar{X} = \frac{1}{m}\sum_{i=1}^m \bar{X}_i$. Find the joint distribution of $\text{SSA} = n\sum_{i=1}^m (\bar{X}_i - \bar{X})^2$ and $\text{SSE} = \sum_{i=1}^m \sum_{j=1}^n (X_{ij} - \bar{X}_i)^2$.

11. Let $X \sim \mathcal{N}_n(0_n, \Sigma)$ and $f \in C^1(\mathbb{R}^n; \mathbb{R})$. Assume that f is polynomially bounded in the sense that there exist constants $k_1, \dots, k_n, a > 0$ such that $|f(x)| \le (1 + |x_1|)^{k_1} \cdots (1 + |x_n|)^{k_n}$ for all $x = (x_1, \dots, x_n)^\top$ with $\|x\| \ge a$. Show that $\mathbb{E}[Xf(X)] = \Sigma\,\mathbb{E}[\nabla f(X)]$, where ∇f denotes the gradient of f.

12. (a) Show that $\mathbb{E}[\max_{1\le i\le n} X_i] \le \sqrt{2\log n}$ if X_i's are random variables with $\mathbb{E}[\mathrm{e}^{tX_i}] \le \mathrm{e}^{t^2/2}$ for $t > 0$.
    (b) Show that $\frac{x}{x^2+1}\varphi(x) \le 1 - \Phi(x) \le \frac{1}{x}\varphi(x)$ for $x > 0$, where Φ and φ are the cumulative distribution function and the probability density function of $\mathcal{N}(0, 1)$, respectively. Then deduce that $\lim_{t\to\infty} \Phi^{-1}(1 - \frac{1}{t})/\sqrt{2\log t} = 1$.
    (c) Show that $\sqrt{2\log(2n)} \ge \mathbb{E}[\max_{1\le i\le n} |X_i|] \ge \Phi^{-1}(1 - \frac{1}{2(n+1)})$ for $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$.

Homework 2 Due: Oct 31st, 2021

1. Let X be a random element having probability density function f_θ, where $\theta \in \Theta \subset \mathbb{R}^k$. If θ has a prior density q, then the posterior is given by $q(\theta|x) = f_\theta(x)q(\theta)/f(x)$, where $f(x) = \int_\Theta f_\theta(x)q(\theta)\,\mathrm{d}\theta$. Show that a statistic $T(X)$ is sufficient for θ if and only if $q(\theta|x)$ depends on x through $T(x)$ for any q.

2. (a) Let X be a random variable following the uniform distribution on $[\theta, \theta+1]$, where $\theta \in \mathbb{R}$ is unknown. Show that, as a sample of size one, X is minimal sufficient but is not boundedly complete.

   (b) Let X be a random variable distributed as $\mathbb{P}\{X = -1\} = p$ and $\mathbb{P}\{X = k\} = (1-p)^2 p^k$ for $k = 0, 1, 2, \cdots$, where $p \in (0, 1)$ is unknown. Show that, as a sample of size one, X is minimal sufficient and boundedly complete but is not complete.

3. Let $X_1, \ldots, X_n$ be i.i.d. random variables, each having probability density function $x \mapsto \mathrm{e}^{-(x-\theta)} \mathbb{1}_{[x>\theta]}$, where $\theta \in \mathbb{R}$ is unknown. Show that the sample minimum $X_{(1)}$ is a complete sufficient statistic, and that the sample mean $\bar{X}$ is a complete statistic but is not sufficient.

4. Let $\begin{pmatrix} X_1 \\ Y_1 \end{pmatrix}, \ldots, \begin{pmatrix} X_n \\ Y_n \end{pmatrix} \overset{\text{i.i.d.}}{\sim} \mathcal{N}_2\left(0_2, \begin{pmatrix} 1 & \theta \\ \theta & 1 \end{pmatrix}\right)$, where $\theta \in (-1, 1)$.

   (a) Show that $\big(\sum_{i=1}^n (X_i^2 + Y_i^2), \sum_{i=1}^n X_i Y_i\big)$ is minimal sufficient but is not complete.

   (b) Show that $T_X = \sum_{i=1}^n X_i^2$ and $T_Y = \sum_{i=1}^n Y_i^2$ are both ancillary, while (T_X, T_Y) is not ancillary.

5. Let $X_1, \ldots, X_n$ be independent random variables with $X_i \sim \mathrm{Gamma}(\alpha_i, \theta)$, whose probability density function is $x \mapsto \frac{1}{\Gamma(\alpha_i)\theta^{\alpha_i}} x^{\alpha_i - 1} \mathrm{e}^{-x/\theta} \mathbb{1}_{[x>0]}$. Show that $T = \sum_{i=1}^n X_i$ is independent of $(X_1, \ldots, X_n)/T$.

6. Let X be a random variable such that $\mathbb{E}[X] = \mu \in \mathbb{R}$ and $\mathrm{Var}(X) = \sigma^2 > 0$, where μ and σ^2 are unknown. Consider estimating μ under the squared error loss. Show that the linear estimator $\hat{\mu}_{a,b} = a + bX$ is inadmissible if $b < 0$; $b > 1$; or $b = 1$ and $a \neq 0$.

7. Let $X \sim \mathcal{N}_n(\mu, I_n)$ and $f \in C^1(\mathbb{R}^n; \mathbb{R}^n)$.

   (a) Show that $\mathbb{E}_\mu\big[\|X - f(X) - \mu\|^2\big] = \mathbb{E}_\mu\big[n + \|f(X)\|^2 - 2\,\mathrm{tr}(\nabla f(X))\big]$ under appropriate conditions.

   (b) Show that $\hat{\mu} = (1 - \frac{n-2}{\|X\|^2})X$ satisfies that $\mathbb{E}_\mu\big[\|\hat{\mu} - \mu\|^2\big] < \mathbb{E}_\mu\big[\|X - \mu\|^2\big]$ if $n \geq 3$.
   Hence, the estimator X for μ is inadmissible under the squared error loss.

8. Note that a parameter is said to be estimable if it has an unbiased estimator.

   (a) Suppose that S is a sufficient statistic taking values in $\{s_1, \ldots, s_d\}$. Show that estimable parameters form a linear space of dimension $\leq d$, with equality holding if and only if S is complete.

   (b) Let $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathrm{Bernoulli}(p)$, where $p \in (0, 1)$. Show that the odds $\frac{p}{1-p}$ is not estimable.

9. Let $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathrm{Poisson}(\lambda)$, where $\lambda \in (0, \infty)$. Find the uniformly minimum variance unbiased estimator for $\mathrm{e}^{a\lambda}$, given $a \in \mathbb{R}$. Explain why unbiased estimators can be ridiculous when $a < -n$.

10. (a) Let X be a random element having probability density function f_θ, where $\theta \in \Theta \subset \mathbb{R}^k$. Suppose that $T = T(X)$ is an unbiased estimator for $\gamma(\theta) \in \mathbb{R}$. Show that $\mathrm{Var}_\theta(T) \geq \sup_\varepsilon |\gamma(\theta+\varepsilon) - \gamma(\theta)|^2 / D(\varepsilon; \theta)$, where $D(\varepsilon; \theta) = D_{\chi^2}(f_{\theta+\varepsilon} \| f_\theta) = \int \frac{(f_{\theta+\varepsilon} - f_\theta)^2}{f_\theta}$.

    (b) Let $X_1, \ldots, X_n$ be i.i.d. random variables, each having probability density function $x \mapsto \mathrm{e}^{-(x-\theta)} \mathbb{1}_{[x>\theta]}$, where $\theta \in \mathbb{R}$. Show that $\sup_{\varepsilon \in (0,\infty)} \varepsilon^2 / D(\varepsilon; \theta) = \frac{a^2}{(\mathrm{e}^a - 1)n^2}$ for some $a \in (1, 2)$, and that the uniformly minimum variance unbiased estimator for θ has variance $\frac{1}{n^2}$.

11. Let X be a random variable having probability density function f_θ, where $\theta \in \mathbb{R}^k$. Recall that the Fisher information matrix is given by $\mathcal{I}(\theta) = \mathbb{E}_\theta\big[(\frac{\partial}{\partial \theta} \log f_\theta(X))(\frac{\partial}{\partial \theta} \log f_\theta(X))^\top\big]$. Let $h_\theta(x) = f_\theta(x)/S_\theta(x)$ be the hazard function, where $S_\theta(x) = \int_x^\infty f_\theta(u)\,\mathrm{d}u$ is the survival function. Show that, under appropriate regularity conditions, $\mathcal{I}(\theta) = \mathbb{E}_\theta\big[(\frac{\partial}{\partial \theta} \log h_\theta(X))(\frac{\partial}{\partial \theta} \log h_\theta(X))^\top\big]$.

Homework 3 Due: Nov 16th, 2021

1. Let $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, 1)$, where $\mu \in \mathbb{R}$. Find the uniformly minimum variance unbiased estimator for μ^2, and show that it has a larger variance than the Cramér–Rao lower bound.

2. Let X be a random element having probability density function f_θ, where $\theta \in \Theta \subset \mathbb{R}^k$. Suppose that $T = T(X)$ is an unbiased estimator for $\gamma(\theta) \in \mathbb{R}$ and q is a probability density function on Θ. Show that, under appropriate regularity conditions, $\int_\Theta \mathrm{Var}_\theta(T) q(\theta)\,\mathrm{d}\theta \geq \big(\int_\Theta \dot\gamma(\theta) q(\theta)\,\mathrm{d}\theta\big)^\top \big(\int_\Theta \mathcal{I}(\theta) q(\theta)\,\mathrm{d}\theta\big)^{-1} \big(\int_\Theta \dot\gamma(\theta) q(\theta)\,\mathrm{d}\theta\big)$, where $\mathcal{I}(\theta)$ is the Fisher information matrix and $\dot\gamma = \nabla\gamma$ is the gradient of γ.

3. Let X be a random element having probability density function f_θ, where $\theta \in \Theta \subset \mathbb{R}$. Suppose that $T = T(X)$ is an unbiased estimator for $\gamma(\theta) \in \mathbb{R}$. Show that, under appropriate regularity conditions, $\mathrm{Var}_\theta(T) \geq \gamma_k(\theta)^\top \mathcal{J}_k(\theta)^{-1} \gamma_k(\theta)$, where k is a positive integer, $\gamma_k(\theta) = (\gamma^{(i)}(\theta))_{1 \leq i \leq k}$, and $\mathcal{J}_k(\theta)$ is a $k \times k$ matrix with entries $\mathcal{J}_k(\theta)[i,j] = \mathbb{E}_\theta\big[\frac{1}{f_\theta(X)^2} \frac{\partial^i f_\theta(X)}{\partial\theta^i} \frac{\partial^j f_\theta(X)}{\partial\theta^j}\big]$ for $1 \leq i, j \leq k$.

4. Let $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} \mathrm{Uniform}([0,\theta])$, where $\theta \in (0,\infty)$. Find the uniformly minimum variance unbiased estimator and the maximum likelihood estimator for θ, and show that they are inadmissible by considering estimators of the form $cX_{(n)}$, where $X_{(n)}$ is the sample maximum and $c \in (0,\infty)$.

5. Let $(X_1, \dots, X_m) \sim \mathrm{Multinomial}(N; p_1, \dots, p_m)$, where N is a given positive integer, and the probabilities p_i's are unknown parameters. Show that the maximum likelihood estimators for $\mathrm{Var}(X_i)$ and $\mathrm{Cov}(X_i, X_j)$ are $X_i(N - X_i)/N$ and $-X_iX_j/N$, respectively, where $i \neq j$.

6. (a) Let $X_1, \dots, X_n$ be i.i.d. random variables from a mixture of Gaussian distributions, each having probability density function $x \mapsto \frac{1}{2\sqrt{2\pi}} \exp(-\frac{(x-\mu)^2}{2}) + \frac{1}{2\sqrt{2\pi\sigma^2}} \exp(-\frac{(x-\mu)^2}{2\sigma^2})$, where $\mu \in \mathbb{R}$ and $\sigma > 0$ are unknown. Show that the likelihood function is unbounded, and thus the maximum likelihood estimates do not exist.

   (b) Let X be a random variable having probability density function $x \mapsto f(x-\theta)$, where $f(x) = \frac{1}{4\sqrt{-\pi x}} \mathrm{e}^x \mathbb{1}_{[x<0]} + \frac{1}{2\pi\sqrt{x(1-x)}} \mathbb{1}_{[0<x<1]} + \frac{1}{4\sqrt{\pi(x-1)}} \mathrm{e}^{1-x} \mathbb{1}_{[x>1]}$ and $\theta \in \mathbb{R}$ is unknown. Show that the likelihood equation has a unique root, but no maximum likelihood estimator exists.

   (c) Let $X_1, \dots, X_n$ be i.i.d. random variables, each having probability density function $x \mapsto \frac{1}{\pi} \frac{1}{1+(x-\mu)^2}$, where $\mu \in \mathbb{R}$ is unknown. Show that multiple solutions of the likelihood equation can indeed occur in this Cauchy location case.

7. (a) Let $X_1, \dots, X_n | \tau \overset{\text{i.i.d.}}{\sim} \mathrm{Exponential}(\tau)$ with rate parameter $1/\tau \sim \mathrm{Gamma}(\alpha_0, 1)$. Find the Bayes estimator for the scale parameter τ under the squared error loss.

   (b) Let $X_1, \dots, X_n | \theta \overset{\text{i.i.d.}}{\sim} \mathrm{Uniform}([0,\theta])$ with $\log\theta \sim \mathcal{N}(\theta_0, \sigma_0^2)$. Find the Bayes estimator for θ under the zero-one loss $(\theta, a) \mapsto \mathbb{1}_{[\theta \notin [a, a+\epsilon)]}$ with $\epsilon \searrow 0$.

8. Let $X|\mu \sim \mathcal{N}_n(\mu, I_n)$, where μ has a prior distribution.

   (a) Show that $\mathbb{E}[\mu|X = x] = x + \nabla \log f(x)$, where f is the marginal probability density function of X.

   (b) Show that $\mathbb{E}[\mu|X] = (1 - \frac{1}{1+\tau^2})X$ if $\mu \sim \mathcal{N}_n(0_n, \tau^2 I_n)$.

   (c) In an empirical Bayes approach, hyperparameters should be estimated using the sample. Restricting our attention to the marginal model in 8b, show that $\frac{n-2}{\|X\|^2}$ is an unbiased estimator for $\frac{1}{1+\tau^2}$ if $n \geq 3$.

9. (a) Given a target distribution π and a proposal kernel q, consider the acceptance probabilities $a_{\mathrm{B}}(\xi, \eta) = \frac{\pi(\eta)q(\eta,\xi)}{\pi(\xi)q(\xi,\eta) + \pi(\eta)q(\eta,\xi)}$. Verify the detailed balance condition for the corresponding Markov chain. Compare $a_{\mathrm{B}}(\xi, \eta)$ with $a_{\mathrm{MH}}(\xi, \eta) = \min\big(1, \frac{\pi(\eta)q(\eta,\xi)}{\pi(\xi)q(\xi,\eta)}\big)$.

   (b) The Bernoulli factory problem asks one to simulate an event of probability $f(p)$ for a known function f, given $p \in [0,1]$ which cannot be evaluated but where events of probability p can be simulated. Show that the following two-coin algorithm is a solution for $f(p) = \frac{pq_1}{(1-p)q_0 + pq_1}$:

   Repeat sampling $C \sim \mathrm{Bernoulli}(p)$ and $C' \sim \mathrm{Bernoulli}(q_C)$ until $C' = 1$, then output C.

Homework 4 Due: Nov 30th, 2021

1. Let $X \sim P_\theta$ be a random element having probability density function f_θ, where $\theta \in \{0,1\}$. Show that $\mathbb{P}_0\{\phi(X)=1\}+\mathbb{P}_1\{\phi(X)=0\} \ge \int \min(f_0,f_1)$ for any $\{0,1\}$-valued test function ϕ, with equality holding if $\phi = \mathbb{1}_{[f_1>f_0]}$. Also show that $\sup_A |P_0(A)-P_1(A)| = 1-\inf_\phi \left[\mathbb{P}_0\{\phi(X)=1\}+\mathbb{P}_1\{\phi(X)=0\}\right]$.

2. Let $X_1,\dots,X_n \overset{\text{i.i.d.}}{\sim}$ Bernoulli(p), where $p \in (0,1)$.
   (a) Find the most powerful test for $H_0: p = 1/3$ against $H_1: p = 2/3$ at level 0.05.
   (b) Find the uniformly most powerful test for $H_0: p = 1/2$ against $H_1: p > 1/2$ at level 0.05.

3. Let $X \sim P$. Denote by ϕ_{α_0} the most powerful test for $H_0: P \in \mathcal{P}_0$ against $H_1: P = P_1$ at level α_0. Typically $\phi_{\alpha_0} \le \phi_{\tilde{\alpha}_0}$ if $\alpha_0 < \tilde{\alpha}_0$. This may be false, even if the most powerful tests are nonrandomized.
   (a) Find the most powerful test for $H_0: P = P_0$ against $H_1: P = P_1$ at level α_0 for $\alpha_0 \in \{0.05, 0.1\}$, where P is a discrete distribution given by

P	1	2	3
P_0	0.85	0.1	0.05
P_1	0.70	0.2	0.10

   (b) Find the most powerful test for $H_0: P \in \{P_0, Q_0\}$ against $H_1: P = P_1$ at level α_0 for $\alpha_0 \in \{\frac{5}{13}, \frac{6}{13}\}$, where P is a discrete distribution given by

P	1	2	3	4
P_0	$\frac{2}{13}$	$\frac{4}{13}$	$\frac{3}{13}$	$\frac{4}{13}$
Q_0	$\frac{4}{13}$	$\frac{2}{13}$	$\frac{1}{13}$	$\frac{6}{13}$
P_1	$\frac{4}{13}$	$\frac{3}{13}$	$\frac{2}{13}$	$\frac{4}{13}$

4. Let F and G be continuous increasing distribution functions, and define $k(u) = G(F^{-1}(u))$ for $u \in (0,1)$.
   (a) Show that F and G are stochastically ordered, say $G \le F$, if and only if $k(u) \le u$ for all u.
   (b) Suppose that F and G have probability density functions f and g, respectively. Show that they are monotone likelihood ratio ordered, say g/f nondecreasing, if and only if k is convex.
   (c) Deduce that monotone likelihood ratio implies stochastic ordering.

5. Let X be a random variable having probability density function f_θ, where $\theta \in \Theta \subset \mathbb{R}$. Show that, for a sample of size one, $\{f_\theta\}_{\theta\in\Theta}$ has monotone likelihood ratio in X if and only if $\frac{\partial^2}{\partial\theta\partial x}\log f_\theta(x) \ge 0$ for all θ and x, provided that this derivative always exists.

6. Let X be a random element having probability density function f_θ, where $\theta \in \Theta \subset \mathbb{R}$. Suppose that $\{f_\theta\}_{\theta\in\Theta}$ has monotone likelihood ratio in $T = T(X)$, whose distribution is continuous. Consider the problem of testing $H_0: \theta \le \theta_0$ against $H_1: \theta > \theta_0$. The p-value of the uniformly most powerful test is given by $\widehat{p} = S(T)$, where $S(t) = \mathbb{P}_{\theta_0}\{T > t\}$. Show that if $\theta \le \theta_0$, then $\mathbb{P}_\theta\{\widehat{p} \le u\} \le u$ for $u \in [0,1]$.

7. Let $X_1,\dots,X_n \overset{\text{i.i.d.}}{\sim}$ Pareto(θ, c), each having probability density function $x \mapsto \theta c^\theta x^{-\theta-1}\mathbb{1}_{[x>c]}$, where $\theta > 1$ is unknown while $c > 0$ is known. Denote by μ the mean of X_1. Find the uniformly most powerful test for $H_0: \mu \le \mu_0$ against $H_1: \mu > \mu_0$ at level α_0.

8. Let $X_1,\dots,X_n \overset{\text{i.i.d.}}{\sim}$ Uniform($[0,\theta]$), where $\theta > 0$.
   (a) Find the uniformly most powerful test for $H_0: \theta \le \theta_0$ against $H_1: \theta > \theta_0$ at level α_0.
   (b) Find the uniformly most powerful test for $H_0: \theta = \theta_0$ against $H_1: \theta \ne \theta_0$ at level α_0.

9. Let $X \sim \mathcal{N}(\mu, 1)$. Show that there exists no uniformly most powerful test for $H_0: \mu = \mu_0$ against $H_1: \mu \in \{\mu_-, \mu_+\}$ at level α_0, where $\mu_- < \mu_0 < \mu_+$.

10. For each of the following models, construct a $(1-\alpha_0)$ confidence region for the parameter θ, using the uniformly most powerful unbiased test for $H_0: \theta = \theta_0$ against $H_1: \theta \ne \theta_0$ at level α_0.
   (a) Let $X_1,\dots,X_n \overset{\text{i.i.d.}}{\sim}$ Gamma(ν, θ), each having probability density function $x \mapsto \frac{1}{\Gamma(\nu)\theta^\nu}x^{\nu-1}\mathrm{e}^{-x/\theta}\mathbb{1}_{[x>0]}$, where $\theta > 0$ is unknown while $\nu > 0$ is known.
   (b) Let $X_i \sim \mathcal{N}(t_i\theta, \sigma^2)$, $1 \le i \le n$, be independent random variables, where only t_i's are known.

Homework 5 Due: Dec 14th, 2021

1. Let $X_1, X_2, \ldots$ be random variables with zero mean. Show that if $\operatorname{Var}(X_i) = O(1)$ and $\operatorname{Corr}(X_i, X_j) \to 0$ as $|i-j| \to \infty$, then $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ converges to zero in probability.

2. (a) Show that if $\mathbb{E}[(X_n - Y_n)^2]/\operatorname{Var}(X_n) \to 0$, then $\operatorname{Corr}(X_n, Y_n) \to 1$.

   (b) Suppose that X_n and Y_n are normalized with mean 0 and variance 1. Show that if $X_n \xrightarrow{d} X$ and $\operatorname{Corr}(X_n, Y_n) \to 1$, then $Y_n \xrightarrow{d} X$.

   (c) Suppose that X_n and Y_n have zero means and equal variances. Disprove the statement in 2b.

3. Applying the Lindeberg–Feller central limit theorem, prove the following propositions.

   (a) Let $X_1, X_2, \ldots$ be independent random variables with zero mean and finite variance. If there is a constant $\delta > 0$ such that $\frac{1}{s_n^{2+\delta}}\sum_{i=1}^n \mathbb{E}\big[|X_i|^{2+\delta}\big] \xrightarrow{(n\to\infty)} 0$, where $s_n = \sqrt{\sum_{i=1}^n \operatorname{Var}(X_i)}$, then $\frac{1}{s_n}\sum_{i=1}^n X_i \xrightarrow{d} \mathcal{N}(0,1)$.

   (b) Let $X_1, X_2, \ldots$ be i.i.d. random variables with zero mean and variance $\sigma^2 < \infty$. If c_{ni}'s are constants such that $\max_{1\le i\le n} c_{ni}^2 \big/ \sum_{j=1}^n c_{nj}^2 \xrightarrow{(n\to\infty)} 0$, then $\sum_{i=1}^n c_{ni}X_i \Big/ \sqrt{\sum_{j=1}^n c_{nj}^2} \xrightarrow{d} \mathcal{N}(0,\sigma^2)$.

4. Let $\varepsilon_1, \varepsilon_2, \ldots$ be i.i.d. random variables with mean μ and variance σ^2. Given a constant $\beta \in [-1,1)$, define $X_1 = \varepsilon_1$ and $X_n = \beta X_{n-1} + \varepsilon_n$ for $n \ge 2$. Find the asymptotic distribution of $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$.

5. Let $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \operatorname{Beta}(\theta, 2)$, where $\theta \in (0,\infty)$. Check that $\hat{\theta}_n = 2\bar{X}_n/(1-\bar{X}_n)$ is a method of moments estimator for θ, where $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. Find the asymptotic distribution of $\hat{\theta}_n$. Then show that $\hat{\theta}_n$ is not efficient by calculating the Cramér–Rao lower bound.

6. Let $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \operatorname{Bernoulli}(p)$, where $p \in (0,1)$. Find the asymptotic distribution of the maximum likelihood estimator for $\operatorname{Var}_p(X_1) = p(1-p)$. Be careful about the case where $p = \frac{1}{2}$.

7. Let $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \operatorname{Uniform}([0,\theta])$, where $\theta \in (0,\infty)$. Show that the maximum likelihood estimator for θ has an asymptotically exponential distribution of convergence rate n, rather than an asymptotically normal distribution of convergence rate $\sqrt{n}$.

8. (a) Let $\{X_{nj} : 1 \le j \le k_n\}_{n=1}^\infty$ be a triangular array of independent random variables. Show that if $\sum_{j=1}^{k_n} \operatorname{Var}(X_{nj}) \xrightarrow{(n\to\infty)} 1$ and $\max_{1\le j\le k_n}|X_{nj}| \le C_n$ for some constants $C_n = o(1)$, then $\sum_{j=1}^{k_n}(X_{nj} - \mathbb{E}X_{nj}) \xrightarrow{d} \mathcal{N}(0,1)$.

   (b) Let $Y_n \sim \operatorname{Binomial}(n, p_n)$ with $p_n = 1/\sqrt{n}$. Show that $n^{3/4}\,(\hat{p}_n - p_n) \xrightarrow{d} \mathcal{N}(0,1)$ where $\hat{p}_n = Y_n/n$.

9. Let $X_{ij} \sim \mathcal{N}(\mu_i, \sigma^2)$, $1 \le i \le n$, $1 \le j \le k$, be independent random variables, where μ_i's and σ^2 are unknown. Suppose that k is fixed but n tends to infinity. Show that the maximum likelihood estimator for σ^2 is not consistent, which exemplifies that maximum likelihood estimators may not behave well when the number of incidental parameters is of the order of the total sample size.

10. Test the Hardy–Weinberg model for probabilities of the three genotypes of a single gene with two alleles: $p_1 = \theta^2$, $p_2 = 2\theta(1-\theta)$, $p_3 = (1-\theta)^2$, $0 < \theta < 1$, and observed frequencies y_1, y_2, y_3, using

    (a) Pearson's frequency chi-square test; (b) Wald's test; (c) the likelihood ratio test; (d) Rao's score test.

11. For each of the following models, find an approximation to the posterior density of θ that is good for sufficiently large samples.

    (a) Let $X_1, \ldots, X_n|\theta \overset{\text{i.i.d.}}{\sim} \operatorname{Poisson}(\theta)$ with prior density $\theta \mapsto (\theta+1)^{-2}\mathbb{1}_{(0,\infty)}(\theta)$.

    (b) Let $X_1, \ldots, X_n|\theta \overset{\text{i.i.d.}}{\sim} \operatorname{Geometric}(\theta)$ with prior density $\theta \mapsto \frac{\pi}{2}\sin(\pi\theta)\mathbb{1}_{(0,1)}(\theta)$.

Homework 6 Due: Dec 28th, 2021

1. Let $(R_1,\dots,R_n)$ be uniformly distributed over the $n!$ permutations of $\{1,2,\cdots,n\}$, based on which T is a statistic. The Hájek projection of T is $\hat{T}=\arg\min_S \mathbb{E}[(T-S)^2]$ among linear statistics $S=\sum_{i=1}^n a(i,R_i)$.

   (a) Show that $\hat{T}=\frac{n-1}{n}\sum_{i=1}^n \hat{a}(i,R_i)-(n-2)\mathbb{E}[T]$, where $\hat{a}(i,j)=\mathbb{E}[T|R_i=j]$.

   (b) Relate $\rho=\frac{1}{n}\sum_{i=1}^n(i-\frac{n+1}{2})(R_i-\frac{n+1}{2})/\operatorname{Var}(R_i)$ to $\hat{\tau}$, where $\tau=\frac{1}{n(n-1)}\sum_{i\neq j}\operatorname{sgn}(i-j)\operatorname{sgn}(R_i-R_j)$.

   (c) Show that $\operatorname{Var}(\tau)=\frac{2(2n+5)}{9n(n-1)}$ and $\operatorname{Var}(\hat{\tau})=\frac{4(n+1)^2}{9n^2(n-1)}$.

2. Let $X_1,X_2,\dots$ be i.i.d. random variables, and let h be a symmetric m-variate function such that $\mathbb{E}[|h(X_1,\dots,X_m)|]<\infty$. Define $U_{k,n}=\binom{n}{m}^{-1}\sum_c h(X_{(k-1)n+i_1},\dots,X_{(k-1)n+i_m})$, where $\{i_1,\dots,i_m\}$ runs over all m-combinations of $\{1,\cdots,n\}$. Find the asymptotic distribution of $\overline{U}_n=K_n^{-1}\sum_{k=1}^{K_n}U_{k,n}$ as $n\to\infty$, assuming that $0<\operatorname{Var}(h_1(X_1))<\infty$ where $h_1(x_1)=\mathbb{E}[h(X_1,X_2,\dots,X_m)|X_1=x_1]$.

3. Let $X_1,X_2,\dots$ be i.i.d. random variables, and let h be a symmetric bivariate function such that $\mathbb{E}[h(X_1,X_1)^2]$ and $\mathbb{E}[h(X_1,X_2)^2]$ are finite. Define $U_n=\frac{2}{n(n-1)}\sum_{1\le i<j\le n}h(X_i,X_j)$ and $V_n=\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n h(X_i,X_j)$.

   (a) Show that $\sqrt{n}U_n$ and $\sqrt{n}V_n$ are asymptotically equivalent in the sense that $\sqrt{n}\,(V_n-U_n)=o_{\mathbb{P}}(1)$. Then deduce the limiting distribution of $\sqrt{n}\,(V_n-\mathbb{E}[h(X_1,X_2)])$.

   (b) Consider $h(x,y)=xy$. Assume that $\mathbb{E}[X_1]=0$. Find the asymptotic relative efficiency of V_n with respect to U_n in terms of asymptotic mean squared error.

4. Let $X_1,X_2,\dots$ be i.i.d. random variables with an unknown probability density function f. Consider the kernel density estimator $\hat{f}_n(x)=\frac{1}{nh_n}\sum_{i=1}^n K\big(\frac{x-X_i}{h_n}\big)$ with bandwidth $h_n>0$, where K is a known probability density function on $\mathbb{R}$. Assume that the second order derivative f'' is continuous on a neighborhood of x, and that K is normalized so that $\int_{-\infty}^{\infty}uK(u)\,\mathrm{d}u=0$ and $\int_{-\infty}^{\infty}u^2K(u)\,\mathrm{d}u=1$.

   (a) Show that $\mathbb{E}[\hat{f}_n(x)]=f(x)+\frac{1}{2}f''(x)h_n^2+o(h_n^2)$, provided $h_n=o(1)$.

   (b) Show that $\sqrt{nh_n}\big(\hat{f}_n(x)-\mathbb{E}[\hat{f}_n(x)]\big)\xrightarrow{d}\mathcal{N}(0,f(x)R_K)$, provided $h_n=o(1)$ and $nh_n\to\infty$, where $R_K=\int_{-\infty}^{\infty}K(u)^2\,\mathrm{d}u$ represents the roughness of the kernel K.

   (c) Show that the optimal bandwidth h_n^{opt} is of order $n^{-1/5}$ by minimizing approximately the pointwise mean squared error $\operatorname{MSE}(\hat{f}_n(x))=\mathbb{E}\big[|\hat{f}_n(x)-f(x)|^2\big]$.

5. It is common in insurance to assume that the times of arrival of claims follow a homogeneous Poisson process (with unit intensity). Given n successive arrivals of claims $0<t_1<t_2<\cdots<t_n$, derive the following tests for this model: (a) Kolmogorov–Smirnov; (b) Carmér–von Mises; (c) Anderson–Darling.

6. Let $X_1,\dots,X_n$ be i.i.d. random variables, each having cumulative distribution function F. The empirical distribution function is given by $F_n(x)=\frac{1}{n}\sum_{i=1}^n \mathbb{1}_{[X_i\le x]}$. Given $p\in(0,1)$, define $\xi_p=\inf\{x:F(x)\ge p\}$ and $\hat{\xi}_{p,n}=\inf\{x:F_n(x)\ge p\}$. Show that if F is differentiable at ξ_p with $F'(\xi_p)=f(\xi_p)>0$, then
$$\sqrt{n}\begin{pmatrix}\bar{X}_n-\mathbb{E}[X_1]\\ \hat{\xi}_{p,n}-\xi_p\end{pmatrix}\xrightarrow{d}\mathcal{N}_2\left(0_2,\begin{pmatrix}\operatorname{Var}(X_1) & \frac{1}{f(\xi_p)}\big(p\,\mathbb{E}[X_1]-\int_{x\le\xi_p}x\,\mathrm{d}F(x)\big)\\ \frac{1}{f(\xi_p)}\big(p\,\mathbb{E}[X_1]-\int_{x\le\xi_p}x\,\mathrm{d}F(x)\big) & p(1-p)/f(\xi_p)^2\end{pmatrix}\right).$$

7. Let $X_1,\dots,X_n$ be i.i.d. random variables, each having cumulative distribution function F. There are absolute constants $A,B\in(0,\infty)$ such that $\mathbb{P}\{\sup_{x\in\mathbb{R}}|F_n(x)-F(x)|>u\}\le A\mathrm{e}^{-nu^2/B}$ for all $u>0$, referred to as the DKW inequality.

   (a) Consider $\hat{\xi}_p=\inf\{x:F_n(x)\ge p\}$ for estimating the p-quantile $\xi_p=\inf\{x:F(x)\ge p\}$. For $\epsilon>0$, let $\delta_\epsilon=\min\{F(\xi_p+\epsilon)-p,\,p-F(\xi_p-\epsilon)\}$. Show that $\mathbb{P}\{|\hat{\xi}_p-\xi_p|>\epsilon\}\le 2A\mathrm{e}^{-n\delta_\epsilon^2/B}$.

   (b) Given $\alpha\in(0,1)$, let $L(x)=\max\{F_n(x)-\epsilon_n,0\}$ and $U(x)=\min\{F_n(x)+\epsilon_n,1\}$, where $\epsilon_n=\sqrt{\frac{B}{n}\log\frac{A}{\alpha}}$. Show that $\{[L(x),U(x)]\}_{x\in\mathbb{R}}$ are simultaneous confidence intervals for $\{F(x)\}_{x\in\mathbb{R}}$ of confidence level $1-\alpha$.

   (c) Prove the DKW inequality with $A\le 4$ and $B\le 8$.

Bonus

1. Given a loss function L, we say that a decision rule T is unbiased if $\mathbb{E}_P[L(P,T)] \le \mathbb{E}_P[L(P',T)]$ for all populations P and P'. Demonstrate the meaning of unbiasedness in the context of point estimation and hypothesis testing, respectively.

2. (a) Let X be a random element having probability density function f_θ, where $\theta \in \Theta \subset \mathbb{R}^k$. Suppose that $T = T(X)$ is an estimator for $\gamma(\theta) \in \mathbb{R}$. Show that, under appropriate regularity conditions, $\int_\Theta \mathbb{E}_\theta[(T-\gamma(\theta))^2]q(\theta)\,\mathrm{d}\theta \ge \big(\int_\Theta \dot\gamma(\theta)q(\theta)\,\mathrm{d}\theta\big)^\top \big(\mathcal{I}_q + \int_\Theta \mathcal{I}(\theta)q(\theta)\,\mathrm{d}\theta\big)^{-1}\big(\int_\Theta \dot\gamma(\theta)q(\theta)\,\mathrm{d}\theta\big)$, where q is a probability density function on Θ and $\mathcal{I}_q = \int_\Theta (\frac{\partial}{\partial\theta}\log q(\theta))(\frac{\partial}{\partial\theta}\log q(\theta))^\top q(\theta)\,\mathrm{d}\theta$.

   (b) For each constant $M > 1$ define $\mathcal{F}_M$ to be the set of probability density functions f on $[-\frac{1}{2},\frac{1}{2}]$ such that $M^{-1} \le f \le M$ and $\max(|f'|,|f''|) \le M$. Let $X_1, X_2, \dots$ be i.i.d. random variables, each having probability density function $f \in \mathcal{F}_M$. Show that there exists a constant $C = C_M > 0$ such that $\sup_{f\in\mathcal{F}_M} \mathbb{E}_f[(T_n - f(0))^2] \ge Cn^{-4/5}$ for every estimator $T_n = T_n(X_1,\dots,X_n)$. Proceed as follows.

      i. Define $f_t(x) = 1 + t^2\psi(x/t)$, where $\psi : \mathbb{R} \to \mathbb{R}$ is a compactly supported smooth function such that $|\psi''| \le M$, $\int_\mathbb{R} \psi(x)\,\mathrm{d}x = 0$, and $\psi(0) = 1$. Then $\{f_t\}_{|t|<\delta} \subset \mathcal{F}_M$ if $\delta > 0$ is small enough.

      ii. Determine an upper bound for $\mathcal{I}_1(t) = \int_\mathbb{R} (\frac{\partial}{\partial t}\log f_t(x))^2 f_t(x)\,\mathrm{d}x$.

      iii. Let q_1 be a smooth probability density function with $q_1(t) = 0$ for $|t| \ge \delta$. If $q_h(\cdot) = q_1(\cdot/h)/h$, then $\int_\mathbb{R} \mathbb{E}_t[(T_n - f_t(0))^2]\,q_{h_n}(t)\,\mathrm{d}t \ge c/(h_n^{-4} + nh_n)$ for some constant $c > 0$.

   Conclude that the kernel density estimator is optimal in a minimax sense.

3. Let $X_1,\dots,X_n \overset{\text{i.i.d.}}{\sim} \text{Uniform}([(1-\kappa)\theta,(1+\kappa)\theta])$, where $\theta \in (0,\infty)$ is unknown while $\kappa \in (0,1)$ is known. Study the estimator for θ of the form $aX_{(1)} + bX_{(n)}$, where $X_{(i)}$ is the i^{th} order statistic.

4. Let $X_1,\dots,X_{30}$ be independent random variables with $X_i \sim \text{Bernoulli}(0.4)$ for $i \le \theta$ and $X_i \sim \text{Bernoulli}(0.7)$ for $i > \theta$. Estimate θ from the following dataset:

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
X_i	0	1	0	0	1	0	0	1	1	1	0	1	0	1	1
i	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
X_i	1	1	0	1	0	1	1	1	1	1	1	0	0	0	0

   Formulate this as a general problem and derive some optimality results.

5. Given $X \sim P_\theta$, consider the region estimation of θ under the loss $(\theta, C) \mapsto S(\text{vol}(C)) - \mathbb{1}_C(\theta)$, where $S : [0,\infty) \to [0,1]$ is an increasing function and $\text{vol}(C)$ is the volume of C.

   (a) Find the Bayes estimator (credible set) C_Π corresponding to a prespecified prior Π of θ.

   (b) Determine the smallest radius of C_Π as σ varies, if $P_\theta = \mathcal{N}_n(\theta, I_n)$, $\Pi = \mathcal{N}_n(\mu, \sigma^2 I_n)$, and $S(t) = \frac{t}{1+t}$.

6. Find the asymptotic distribution of the correlation coefficient of a sample of i.i.d. bivariate random vectors with finite fourth moments. What if the underlying population is normal and the Fisher transformation $r \mapsto \text{artanh}(r) = \frac{1}{2}\log(\frac{1+r}{1-r})$ is used?

7. Let $X_i \sim \mathcal{N}(f(t_i), \sigma^2)$ be independent random variables, where t_i's are known. Consider the estimator $\hat f_n = f_{\tau_n,\hat\beta_n}$ for f, where $f_{\tau,\beta}(t) = \beta_1\mathbb{1}_{[t<\tau]} + \beta_2\mathbb{1}_{[t\ge\tau]}$ and $\hat\beta_n = \arg\min_{\beta\in\mathbb{R}^2}\sum_{i=1}^n (X_i - f_{\tau_n,\beta}(t_i))^2$.

   (a) Evaluate the integrated mean squared error of $\hat f_n$, assuming appropriate conditions.

   (b) Suppose that $t_i = i/n$. Find asymptotic expressions for $\int \text{Bias}(\hat f_n)^2$ and $\int \text{Var}(\hat f_n)$.