PhD Prelim Exam
THEORY
(Majors and Co-majors)

Summer 2010
(Given on 7/8/10)

Part I: A Distance Function on Probability Distributions

Notation: For a random variable Z on a probability space (psp) $(\Omega, \mathcal{F}, P)$, recall the definition of the norm

$$\|Z\|_2 \equiv \left[\mathrm{E}|Z|^2\right]^{1/2} \equiv \left[\int_\Omega |Z(\omega)|^2 dP(\omega)\right]^{1/2}$$

so that $\|Z\|_2^2 \equiv \mathrm{E}|Z|^2$. When necessary for clarity, a subscript "$(\Omega, \mathcal{F}, P)$" will be used in the norm $\|Z\|_2 \equiv \|Z\|_{2,(\Omega,\mathcal{F},P)}$ to indicate the underlying psp on which a random variable is defined.

1. Let X and Y be random variables on a psp $(\Omega, \mathcal{F}, P)$. Prove Minkowski's inequality:

$$\|X - Y\|_2 \leq \|X\|_2 + \|Y\|_2,$$

noting that the expectations may not be finite.

Hint: Write $|X - Y||X - Y|$ and apply the Cauchy-Schwarz inequality to $|X| \cdot |X - Y|$ and $|Y| \cdot |X - Y|$.

Definition: Let $\mathcal{C}_2$ denote the set of all probability distributions on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ with finite 2nd moment. That is, if $\mu \in \mathcal{C}_2$, then μ is a probability measure on the Borel σ-algebra $\mathcal{B}(\mathbb{R})$ of $\mathbb{R}$ and $\int_{\mathbb{R}} |x|^2 d\mu(x) < \infty$.

Notation: For a random variable X defined on a psp $(\Omega, \mathcal{F}, P)$, the distribution of X, denoted by PX^{-1}, is a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ given as $PX^{-1}(A) \equiv P(X \in A), A \in \mathcal{B}(\mathbb{R})$.

Definition: For $\mu, \nu \in \mathcal{C}_2$, define a corresponding nonnegative distance function $d_2(\mu, \nu)$ between distributions μ and ν as

$$\left[d_2(\mu,\nu)\right]^2 \equiv \inf \left\{ \int_\Omega |X(\omega) - Y(\omega)|^2 dP(\omega) : \begin{array}{l} (\Omega, \mathcal{F}, P) \text{ is a psp with random variables } X, Y \\ \text{defined on this psp so that } PX^{-1} = \mu,\ PY^{-1} = \nu \\ \text{(meaning } X \text{ has distribution } \mu,\ Y \text{ has distribution } \nu) \end{array} \right\}.$$

In short, we may write

$$\left[d_2(\mu,\nu)\right]^2 \equiv \inf \|X - Y\|^2_{2,(\Omega,\mathcal{F},P)}$$

where the infimum is over all probability spaces $(\Omega, \mathcal{F}, P)$ on which some random variables X and Y are defined so that X has distribution μ and Y has distribution ν.

2. For $\mu, \nu \in \mathcal{C}_2$, give an example of a psp $(\Omega, \mathcal{F}, P)$ having random variables X, Y where $PX^{-1} = \mu$, $PY^{-1} = \nu$ and, using this example, show that $d_2(\mu, \nu)$ is finite.

3. d_2 is a *metric* on $\mathcal{C}_2$ because the following three properties hold for any $\mu, \nu, \gamma \in \mathcal{C}_2$:

(i) $d_2(\mu, \mu) = 0$, (ii) $d_2(\mu, \nu) = d_2(\nu, \mu)$, (iii) $d_2(\mu, \gamma) \leq d_2(\mu, \nu) + d_2(\nu, \gamma)$.

Verify that property (i) above holds. (You need **not** argue that (ii) or (iii) hold.)

Note: The development in Questions **4,5,6** below serves to show that: for $\mu, \nu \in \mathcal{C}_2$, there exists a psp $(\Omega_0, \mathcal{F}_0, P_0)$ and random variables X_0, Y_0 on this psp such that $P_0 X_0^{-1} = \mu$, $P_0 Y_0^{-1} = \nu$ and

$$\left[d_2(\mu,\nu)\right]^2 = \int_{\Omega_0} |X_0(\omega) - Y_0(\omega)|^2 dP_0(\omega) \equiv \|X_0 - Y_0\|^2_{2,(\Omega_0,\mathcal{F}_0,P_0)}.$$

That is, **the infimum in the definition of** $\left[d_2(\mu,\nu)\right]^2$ **is always achieved.**

4. For each integer $n \geq 1$, there exists a psp $(\Omega_n, \mathcal{F}_n, P_n)$ and random variables X_n, Y_n on this psp such that $P_n X_n^{-1} = \mu$, $P_n Y_n^{-1} = \nu$ and

$$[d_2(\mu,\nu)]^2 \leq \mathrm{E}_n|X_n - Y_n|^2 < [d_2(\mu,\nu)]^2 + 1/n,$$

where $\mathrm{E}_n|X_n - Y_n|^2 \equiv \|X_n - Y_n\|^2_{2,(\Omega_n,\mathcal{F}_n,P_n)}$. Using this fact, show that the sequence of random variables $\{X_n\}_{n=1}^{\infty}$ is tight **and** that $\{X_n - Y_n\}_{n=1}^{\infty}$ is also tight.

5. The conclusion of Question 4 implies that the sequence of pairs $\{(X_n, Y_n)\}_{n=1}^{\infty}$ defined there is tight. By Helly's selection theorem, there then exists a subsequence (X_{n_k}, Y_{n_k}) of (X_n, Y_n) and a random vector (X_0, Y_0), with (X_0, Y_0) defined on some psp $(\Omega_0, \mathcal{F}_0, P_0)$, such that $(X_{n_k}, Y_{n_k}) \xrightarrow{d} (X_0, Y_0)$ as $k \to \infty$ (where $\xrightarrow{d}$ denotes convergence in distribution). Using this, show that the distribution of X_0 is μ **and** that the distribution of Y_0 is ν, stating any standard results that you use.

6. By Skorohod's embedding theorem, there exists a psp $(\Omega, \mathcal{F}, P)$ and random vectors (U_k, V_k), $k \geq 0$, on this psp such that $(U_k, V_k) \to (U_0, V_0)$ almost surely(P) as $k \to \infty$ and (U_k, V_k) has the same distribution as (X_{n_k}, Y_{n_k}), $k \geq 1$, while (U_0, V_0) has the same distribution as (X_0, Y_0). Using this along with results in Questions **4-5**, show that

$$\left[d_2(\mu,\nu)\right]^2 = \mathrm{E}_0|X_0 - Y_0|^2 \equiv \|X_0 - Y_0\|^2_{2,(\Omega_0,\mathcal{F}_0,P_0)}.$$

Part II. Inequalities Involving the Distance Function

Abbreviated Notation: Recall the function d_2 provides a distance between two *distributions* in $\mathcal{C}_2$. If X and Y are arbitrary random variables (possibly defined on different probability spaces) having finite 2nd moments, then for convenience may write $d_2(X, Y)$ to denote the d_2-distance between the *distribution* of X and the *distribution* of Y. That is,

$$d_2(X, Y) \equiv d_2(\mu_X, \nu_Y),$$

where above $\mu_X, \nu_Y \in \mathcal{C}_2$ represent probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ corresponding to the distributions of X and Y.

Beginning of Questions 7,8,9. Let $U_1, \ldots, U_n$ be *independent* random variables on some common psp and $V_1, \ldots, V_n$ be *independent* random variables on a second (possibly different) psp, where the distributions of $U_1, \ldots, U_n, V_1, \ldots, V_n$ belong to $\mathcal{C}_2$.

7. Construct a common psp $(\Omega, \mathcal{F}, P)$ on which there exist pairs of random variables $(X_1, Y_1), \ldots, (X_n, Y_n)$ having the four properties:

   (i) the n pairs of the random variables are independent (though within any pair, X_i and Y_i need not be independent).

   (ii) X_i has the same distribution as U_i for $i = 1, \ldots, n$.

   (iii) Y_i has the same distribution as V_i for $i = 1, \ldots, n$.

   (iv) $[d_2(U_i, V_i)]^2 = \mathrm{E}|X_i - Y_i|^2$ for $i = 1, \ldots, n$.

   Hint: Use the result developed on Page 2.

8. Suppose $a_1, \ldots, a_n$ are real-valued constants. If U_i and V_i have the same mean for each $i = 1, \ldots, n$, show that

$$\left[d_2\left(\sum_{i=1}^{n} a_i U_i, \sum_{i=1}^{n} a_i V_i\right)\right]^2 \leq \sum_{i=1}^{n} a_i^2 [d_2(U_i, V_i)]^2.$$

9. If U_1 has real-valued mean a and V_1 has real-valued mean b, show that

$$[d_2(U_1 - a, V_1 - b)]^2 \leq [d_2(U_1, V_1)]^2.$$

Part III. The Distance Function and Convergence in Distribution

10. Suppose μ_n, $n \geq 0$, is a sequence of probability distributions on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that μ_n converges in distribution to μ_0 as $n \to \infty$ ($\mu_n \xrightarrow{d} \mu_0$). Provide three equivalent characterizations (alternative formulations) of $\mu_n \xrightarrow{d} \mu_0$.

11. Suppose μ_n, $n \geq 0$, is a sequence of probability distributions in $\mathcal{C}_2$. Prove that, as $n \to \infty$,

$$d_2(\mu_n, \mu_0) \to 0 \textbf{ if and only if } \mu_n \xrightarrow{d} \mu_0 \text{ and } \int_{\mathbb{R}} |x|^2 d\mu_n(x) \to \int_{\mathbb{R}} |x|^2 d\mu_0(x).$$

    Hint: To show the "$\Rightarrow$" implication, apply the result from Page 2 and the inequality $\big|\|X\|_2 - \|Y\|_2\big| \leq \|X - Y\|_2$ based on Question **1**; to show the "$\Leftarrow$" implication, use Skorohod's embedding theorem.

Part IV. Establishing the Bootstrap

12. On a common psp $(\Omega, \mathcal{F}, P)$, suppose X_n, $n \geq 1$, is a sequence of independent, identically distributed (iid) random variables with $\mathrm{E}X_1^2 < \infty$. For $M_n = \max_{1 \leq i \leq n} |X_i|$, show that $M_n/\sqrt{n} \to 0$ as $n \to \infty$ almost surely(P).

Hint: Recall that for a generic random variable Y, $\sum_{n=1}^{\infty} P(|Y| > \sqrt{n}) \leq \mathrm{E}|Y|^2$.

Beginning of Questions 13-15. On a psp $(\Omega, \mathcal{F}, P)$, suppose X_n, $n \geq 1$, is a sequence of iid random variables with common cumulative distribution function (cdf) F and variance $0 < \sigma^2 < \infty$. Let $F_n(x) = \sum_{i=1}^n \mathbb{I}(X_i \leq x)/n$, $x \in \mathbb{R}$, represent the empirical cdf based on $X_1, \ldots, X_n$, $n \geq 1$, where $\mathbb{I}(\cdot)$ denotes the indicator function. Write μ_{F_n} and μ_F to denote the probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ prescribed by cdfs F_n and F, respectively.

Recall that $X_1, \ldots, X_n$, F_n and μ_{F_n} are all measurable functions of $\omega \in \Omega$ where, for fixed $x \in \mathbb{R}$ and $A \in \mathcal{B}(\mathbb{R})$, $F_n(x, \omega) = \sum_{i=1}^n \mathbb{I}(X_i(\omega) \leq x)/n$ and $\mu_{F_n}(A, \omega) = \sum_{i=1}^n \mathbb{I}(X_i(\omega) \in A)/n$.

13. Show that $d_2(\mu_{F_n}, \mu_F) \to 0$ as $n \to \infty$ almost surely(P), stating any standard results that you use.

Hint: See Question **11**.

The bootstrap can be used to estimate the sampling distribution of the sample mean $\bar{X}_n = \sum_{i=1}^n X_i/n$. Given observations $X_1, \ldots, X_n$, create iid random variables $X_1^*, \ldots, X_n^*$, defining each X_i^* by a random and uniform selection from $\{X_1, \ldots, X_n\}$ (X_i^* assumes the value X_j with probability $1/n$, $j = 1, \ldots, n$). That is, a "bootstrap" sample $X_1^*, \ldots, X_n^*$ is a random sample from the probability distribution μ_{F_n} (given $X_1, \ldots, X_n$) and the expected value of a bootstrap observation is $\mathrm{E}_* X_1^* = \bar{X}_n$ (where E_* denotes expectation with respect to the distribution of X_1^*). The distribution of $Z_n^* \equiv \sqrt{n}(\bar{X}_n^* - \mathrm{E}_* X_1^*)$, defined by the sample mean $\bar{X}_n^* = \sum_{i=1}^n X_i^*/n$ of a bootstrap random sample from μ_{F_n}, approximates the distribution of $Z_n \equiv \sqrt{n}(\bar{X}_n - \mathrm{E}X_1)$.

14. Considering the distribution of Z_n^* given $X_1, \ldots, X_n$, show $d_2(Z_n^*, Z) \to 0$ as $n \to \infty$ almost surely(P), where Z denotes a random variable having a normal $\mathcal{N}(0, \sigma^2)$ distribution with mean 0 and variance σ^2.

Hint: See Questions **8**, **9**, **11**, **13** and note $d_2(Z_n^*, Z) \leq d_2(Z_n^*, Z_n) + d_2(Z_n, Z)$ by Question **3**.

15. Another way to investigate the bootstrap distribution is through the Berry-Esseen lemma, which states

$$\sup_{x \in \mathbb{R}} \left| P(Z_n^* \leq x | X_1, \ldots, X_n) - \Phi\left(\frac{x}{\sqrt{\sigma_{*n}^2}}\right) \right| \leq \frac{5}{\sqrt{n}} \frac{\mathrm{E}_*|X_1^*|^3}{(\sigma_{*n}^2)^{3/2}}, \quad \sigma_{*n}^2 \equiv \mathrm{Var}_*(X_1^*),$$

where $\Phi(\cdot)$ denotes the standard normal cdf and $P(Z_n^* \leq x | X_1, \ldots, X_n)$, $x \in \mathbb{R}$, denotes the cdf of the bootstrap quantity Z_n^* while $\sigma_{*n}^2 \equiv \mathrm{Var}_*(X_1^*)$ and $\mathrm{E}_*|X_1^*|^3$ denote the variance and absolute 3rd moment of a bootstrap random variable X_1^* (given $X_1, \ldots, X_n$). Using this and the result of Question 12, prove that the distribution of Z_n^* (given $X_1, \ldots, X_n$) converges to a normal $\mathcal{N}(0, \sigma^2)$ distribution as $n \to \infty$, almost surely(P).

1. If $\|X\|_2 = \infty$ or $\|Y\|_2 = \infty$ or $\|X - Y\|_2 = 0$, then the inequality holds trivially. So suppose $\|X\|_2 + \|Y\|_2 < \infty$ and $\|X - Y\|_2 > 0$. Then,

$$\begin{aligned} \|X - Y\|_2^2 = \mathrm{E}|X - Y|^2 &\leq \mathrm{E}|X| \cdot |X - Y| + \mathrm{E}|Y| \cdot |X - Y| \\ &\leq \|X\|_2 \cdot \|X - Y\|_2 + \|Y\|_2 \cdot \|X - Y\|_2, \end{aligned}$$

by the Cauchy-Schwarz (Holder's) inequality. Divide both sides by $\|X - Y\|_2 > 0$.

2. Let $(\Omega, \mathcal{F}, P)$ be any psp with random variables X and Y such that $PX^{-1} = \mu$ and $PY^{-1} = \nu$. One possibility is $\Omega = \mathbb{R}^2$, $\mathcal{F} = \mathcal{B}(\mathbb{R}^2)$, $P = \mu \times \nu$ and $X(\omega) = \omega_1, Y(\omega) = \omega_2$ for $\omega = (\omega_1, \omega_2) \in \Omega$. Another possibility is $\Omega = (0,1)$, $\mathcal{F} = \mathcal{B}((0,1))$, $P \equiv$Lebesgue measure, $X(\omega) = \phi_\mu(\omega)$, $Y(\omega) = \phi_\nu(\omega)$, $\omega \in (0,1)$, where $\phi_\gamma(\omega) = \inf\{x \in \mathbb{R} : \gamma((-\infty, x]) \geq \omega\}$ is the quantile function of a probability distribution/measure γ.

Then, by Minkowski's inequality in Question 1,

$$d_2(\mu, \nu) \leq \|X - Y\|_{2,(\Omega,\mathcal{F},P)} \leq \|X\|_{2,(\Omega,\mathcal{F},P)} + \|Y\|_{2,(\Omega,\mathcal{F},P)} < \infty$$

because, as $\mu, \nu \in \mathcal{C}_2$,

$$\|X\|_{2,(\Omega,\mathcal{F},P)}^2 = \int_{\mathbb{R}} |x|^2 d\mu(x) < \infty, \quad \|Y\|_{2,(\Omega,\mathcal{F},P)}^2 = \int_{\mathbb{R}} |y|^2 d\nu(y) < \infty.$$

3. Let $(\Omega, \mathcal{F}, P)$ be any psp on which a random variable X exists such that $PX^{-1} = \mu$. (One possibility is $\Omega = \mathbb{R}$, $\mathcal{F} = \mathcal{B}(\mathbb{R})$, $P = \mu$, $X(\omega) = \omega$ for $\omega \in \Omega$.) Now set $Y(\omega) = X(\omega)$ for $\omega \in \Omega$ so that $PY^{-1} = \mu$. By definition,

$$0 \leq [d_2(\mu, \mu)]^2 \leq \int_\Omega |X - Y|^2 dP = \int_\Omega |X - X|^2 dP = 0.$$

4. By definition, $\{X_n\}_{n\geq 1}$ is tight if, for any $\epsilon > 0$, there exists some $M \equiv M_\epsilon > 0$ such that $\sup_n P_n(|X_n| > M) < \epsilon$ holds. But for all n, the distribution of X_n is the same: $P_n X_n^{-1} = \mu$. Hence, there exists some $M > 0$ where

$$\sup_n P_n(|X_n| > M) = P_1(|X_1| > M) = \mu([-M, M]^c) < \epsilon$$

holds (i.e., $\lim_{M\to\infty} P_1(|X_1| > M) = 0$.)

By Markov's inequality and the fact that $\sup_n \mathrm{E}_n|X_n - Y_n|^2 \leq 1 + [d_2(\mu, \nu)]^2 < \infty$ (cf. Question 2), we find that

$$\sup_n P_n(|X_n - Y_n| > M) \leq \sup_n \frac{\mathrm{E}_n|X_n - Y_n|^2}{M^2} \leq \frac{1 + [d_2(\mu, \nu)]^2}{M^2}$$

can be made arbitrarily small for large M, implying the tightness of $\{|X_n - Y_n|\}_{n\geq 1}$.

5. $h(x, y) = x$ and $g(x, y) = y$ are continuous functions of $(x, y) \in \mathbb{R}^2$. Because $(X_{n_k}, Y_{n_k}) \xrightarrow{d} (X_0, Y_0)$, the continuous mapping theorem gives $h(X_{n_k}, Y_{n_k}) = X_{n_k} \xrightarrow{d} X_0 = h(X_0, Y_0)$ and $g(X_{n_k}, Y_{n_k}) = Y_{n_k} \xrightarrow{d} Y_0 = g(X_0, Y_0)$. But for all $k \geq 1$, the distribution of X_{n_k} is μ while the distribution of Y_{n_k} is ν. Therefore, X_0 and Y_0 must have distributions μ and ν.

6. By Question 5, X_0 and Y_0 have distributions μ and ν so that, by definition,
$$[d_2(\mu,\nu)]^2 \le \mathrm{E}_0|X_0 - Y_0|^2$$
holds. Also, for all $k \ge 1$, note that due to the equivalence of distributions
$$\int_\Omega |U_k - V_k|^2 dP \equiv \mathrm{E}|U_k - V_k|^2 = \mathrm{E}_{n_k}|X_{n_k} - Y_{n_k}|^2 \le [d_2(\mu,\nu)]^2 + 1/n_k.$$
$(U_k, V_k) \to (U_0, V_0)$ a.s.(P) implies $|U_k - V_k|^2 \to |U_0 - V_0|^2$ a.s.(P) so that, by Fatou's lemma,
$$\mathrm{E}_0|X_0 - Y_0|^2 = \mathrm{E}|U_0 - V_0|^2 \le \liminf_{k\to\infty} \mathrm{E}|U_k - V_k|^2 \le [d_2(\mu,\nu)]^2 + \liminf_{k\to\infty} 1/n_k = [d_2(\mu,\nu)]^2.$$
Now $[d_2(\mu,\nu)]^2 = \mathrm{E}_0|X_0 - Y_0|^2$ follows.

7. For $i = 1, \ldots, n$, there exists a psp $(\Omega_i, \mathcal{F}_i, P_i)$ and random variables $\tilde{X}_i$ and $\tilde{Y}_i$ such that $\tilde{X}_i$ has the same distribution as U_i, $\tilde{Y}_i$ has the same distribution as V_i, and $\|\tilde{X}_i - \tilde{Y}_i\|^2_{2,(\Omega_i,\mathcal{F}_i,P_i)} = \mathrm{E}_i|\tilde{X}_i - \tilde{Y}_i|^2 = [d_2(U_i, V_i)]^2$ (cf. Part II or Questions 4-6).
To "build in" independence, use a product space $\Omega = \Omega_1 \times \cdots \times \Omega_n$, $\mathcal{F} = \mathcal{F}_1 \times \cdots \times \mathcal{F}_n$ and a product measure $P = P_1 \times \cdots \times P_n$. For $\omega = (\omega_1, \ldots, \omega_n) \in \Omega$, let $Z_i(\omega) = \omega_i$ be the ith coordinate projection and define $X_i(\omega) = \tilde{X}_i(Z_i(\omega))$ and $Y_i(\omega) = \tilde{Y}_i(Z_i(\omega))$ for $i = 1, \ldots, n$. Now $(X_1, Y_1), \ldots, (X_n, Y_n)$ are independent pairs and each (X_i, Y_i) has the same distribution as $(\tilde{X}_i, \tilde{Y}_i)$ (i.e., X_i, U_i (or Y_i, V_i) have the same distribution and $\mathrm{E}|X_i - Y_i|^2 = [d_2(U_i, V_i)]^2$).

8. Using the construction in Question 7, $(X_1, \ldots, X_n)$ has the same distribution as $(U_1, \ldots, U_n)$ (the variables in each vector are independent) and $(Y_1, \ldots, Y_n)$ has the same distribution as $(V_1, \ldots, V_n)$. So, $\sum_{i=1}^n a_i X_i$ has the same distribution as $\sum_{i=1}^n a_i U_i$ and $\sum_{i=1}^n a_i Y_i$ has the same distribution as $\sum_{i=1}^n a_i V_i$. Also, each $\mathrm{E}(X_i - Y_i) = 0$ because X_i has the same distribution as U_i, Y_i has the same distribution as V_i, and V_i and U_i have the same mean. By definition,
$$\left[d_2\left(\sum_{i=1}^n a_i U_i, \sum_{i=1}^n a_i V_i\right)\right]^2 \le \mathrm{E}\left(\sum_{i=1}^n a_i(X_i - Y_i)\right)^2 = \sum_{i=1}^n a_i^2 \mathrm{E}\,(X_i - Y_i)^2$$
using that the variance of an independent sum is the sum of variances. Finally, $\mathrm{E}\,(X_i - Y_i)^2 = [d_2(U_i, V_i)]^2$, $i = 1, \ldots, n$ by the construction.

9. There exists a psp $(\Omega_1, \mathcal{F}_1, P_1)$ and random variables X_1 and Y_1 such that X_1 has the same distribution as U_1, Y_1 has the same distribution as V_1, and $\|X_1 - Y_1\|^2_{2,(\Omega_1,\mathcal{F}_1,P_1)} = \mathrm{E}_1|X_1 - Y_1|^2 = [d_2(U_1, V_1)]^2$ (cf. Part II or Questions 4-6). $X_1 - a$ has the same distribution as $U_1 - a$ and $Y_1 - b$ has the same distribution as $V_1 - b$. So by definition,
$$[d_2(U_1 - a, V_1 - b)]^2 \le \mathrm{E}|(X - a) - (Y - b)|^2 = \mathrm{E}|X_1 - Y_1|^2 - (a - b)^2 \le [d_2(U_1, V_1)]^2.$$
Alternatively, let $(\Omega, \mathcal{F}, P)$ be any psp on which random variables X and Y exist with X having the same distribution as U_1 and Y having the same distribution as V_1 (see Solution for Question 1). Then, by definition
$$[d_2(U_1 - a, V_1 - b)]^2 \le \mathrm{E}|(X_1 - a) - (Y_1 - b)|^2 = \mathrm{E}|X - Y|^2 - (a - b)^2 \le \mathrm{E}|X - Y|^2.$$
Since the psp $(\Omega, \mathcal{F}, P)$ is arbitrary and $[d_2(U_1 - a, V_1 - b)]^2$ is a lower bound on $\mathrm{E}|X - Y|^2$, $[d_2(U_1 - a, V_1 - b)]^2 \le [d_2(U_1, V_1)]^2$ follows by the infimum definition.

10. Some equivalent characterizations of $\mu_n \xrightarrow{d} \mu_0$ are

(a) $\mu_n(A) \to \mu_0(A)$ for any $A \in \mathcal{B}(\mathbb{R})$ where $\mu_0(\partial A) = 0$.

(b) $\int_{\mathbb{R}} f(x)d\mu_n(x) \to \int_{\mathbb{R}} f(x)d\mu_0(x)$ for all bounded, continuous functions f on $\mathbb{R}$.

(c) $\{\mu_n\}_{n\geq 1}$ is tight and all convergent subsequences of μ_n converge in distribution to μ_0.

(d) characteristic functions converge $\int_{\mathbb{R}} e^{ixt}d\mu_n(x) \to \int_{\mathbb{R}} e^{ixt}d\mu_0(x)$, all $t \in \mathbb{R}$

11. $\Leftarrow$ implication. Suppose $\mu_n \xrightarrow{d} \mu_0$ and $\int_{\mathbb{R}} |x|^2 d\mu_n(x) \to \int_{\mathbb{R}} |x|^2 d\mu_0(x)$. By Skorohod's embedding theorem, there exists a psp $(\Omega, \mathcal{F}, P)$ and random variables Y_n, $n \geq 0$, such that Y_n has distribution $PY_n^{-1} = \mu_n$ and $Y_n \to Y_0$ a.s.(P). By definition,

$$[d_2(\mu_n, \mu_0)]^2 \leq \mathrm{E}|Y_n - Y_0|^2, \quad n \geq 1.$$

Note that $|Y_n - Y_0|^2 \leq [2\max\{|Y_n|, |Y_0|\}]^2 \leq 4(|Y_0|^2 + |Y_n|^2) \equiv g_n$, for all $\omega \in \Omega$ and $n \geq 1$ and that $g_n \to 8|Y_0|$ a.s.(P) and

$$\begin{aligned}\int_\Omega g_n dP = 4\int_\Omega |Y_n| + |Y_0| dP &= 4\int_{\mathbb{R}} |x|^2 d\mu_n(x) + 4\int_{\mathbb{R}} |x|^2 d\mu_0(x) \\ &\to 8\int_{\mathbb{R}} |x|^2 d\mu_0(x) = 8\int_\Omega |Y_0| dP.\end{aligned}$$

As $|Y_n - Y_0|^2 \to 0$ a.s.(P), the extended Dominated Convergence Theorem then yields

$$[d_2(\mu_n, \mu_0)]^2 \leq \mathrm{E}|Y_n - Y_0|^2 \to 0.$$

$\Rightarrow$ implication. Suppose $d_2(\mu_n, \mu_0) \to 0$. For each $n \geq 1$, there exists a psp $(\Omega_n, \mathcal{F}, P_n)$ and random variables X_n and Y_n such that $P_n X_n^{-1} = \mu_n$, $P_n Y_n^{-1} = \mu_0$ and $[d_2(\mu_n, \mu_0)]^2 = \mathrm{E}_n|X_n - Y_n|^2$. Suppose Y_0 is a random variable with distribution μ_0. Because Y_n has distribution μ_0 for all $n \geq 0$, $Y_n \xrightarrow{d} Y_0$ trivially. Also, $X_n - Y_n \xrightarrow{p} 0$ because, for any $\epsilon > 0$,

$$P_n(|X_n - Y_n| > \epsilon) \leq \frac{\mathrm{E}_n|X_n - Y_n|^2}{\epsilon^2} = \frac{[d_2(\mu_n, \mu_0)]^2}{\epsilon^2} \to 0.$$

By Slutsky's theorem, $X_n = (X_n - Y_n) + Y_n \xrightarrow{d} Y_0$; in other words, $\mu_n \xrightarrow{d} \mu_0$. By Minkowski's inequality,

$$\left|[\mathrm{E}_n|X_n|^2]^{1/2} - [\mathrm{E}_n|Y_n|^2]^{1/2}\right| \leq [\mathrm{E}_n|X_n - Y_n|^2]^{1/2} = d_2[\mu_n, \mu_0] \to 0$$

and note $\mathrm{E}_n|X_n|^2 = \int_{\mathbb{R}} |x|^2 d\mu_n(x)$, $\mathrm{E}_n|Y_n|^2 = \int_{\mathbb{R}} |x|^2 d\mu_0(x)$.

12. Pick an arbitrary $C > 0$. Then

$$\sum_{n=1}^{\infty} P(|X_n| > C\sqrt{n}) = \sum_{n=1}^{\infty} P(|X_1/C| > \sqrt{n}) \leq \mathrm{E}|X_1/C|^2 < \infty.$$

By the Borel-Cantelli lemma, $P(|X_n| > C\sqrt{n}$ infinitely often$) = 0$ so that $P(|X_n|/\sqrt{n} \leq C$ eventually$) = 1$. Let $B_m \equiv \{\omega \in \Omega : |X_n(\omega)|/\sqrt{n} \leq 1/m \text{ eventually}\}$, $m \geq 1$, and $A = \cap_{m=1}^{\infty} B_m$. Then $B_{m+1} \subset B_m$ so that

$$P(A) = \lim_{m\to\infty} P(B_m) = \lim_{m\to\infty} 1 = 1.$$

For $\omega \in A$, $|X_n(\omega)|/\sqrt{n} \to 0$ so that, for an arbitrary $\epsilon > 0$, there exists $N \equiv N_{\omega,\epsilon}$ such that $|X_n(\omega)|/\sqrt{n} < \epsilon$ for $n \geq N$, implying that

$$\frac{M_n(\omega)}{\sqrt{n}} \le \max\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{N}|X_i(\omega)|, \epsilon\right\}$$

for $n \ge N$. For large n, $M_n(\omega)/\sqrt{n} \le \epsilon$ holds implying $M_n(\omega)/\sqrt{n} \to 0$ for $\omega \in A$.

13. By the Glivenko-Cantelli lemma, $\sup_{x\in\mathbb{R}}|F_n(x)-F(x)| \to 0$ a.s.(P), $F_n(x) \equiv F_n(x,\omega) \equiv \sum_{i=1}^n \mathbb{I}(X_i(\omega) \le x)/n$, $\omega \in \Omega$. By the SLLN, $\int_{\mathbb{R}} x^2 d\mu_{F_n}(x) = \sum_{i=1}^n X_i^2/n \to \mathrm{E}X_1^2 = \int_{\mathbb{R}} x^2 d\mu_F(x)$ since $\mathrm{E}X_1^2 < \infty$. Let $A = \{\omega \in \Omega : \sup_{x\in\mathbb{R}}|F_n(x,\omega)-F(x)| \to 0, \sum_{i=1}^n X_i^2(\omega)/n \to \mathrm{E}X_1^2\}$ so that $P(A) = 1$. For a fixed $\omega \in A$, $F_n(x) = F_n(x,\omega) \to F(x)$ for all $x \in \mathbb{R}$ (implying $\mu_{F_n} = \mu_{F_n}(\omega) \xrightarrow{d} \mu_F$) and $\int_{\mathbb{R}} x^2 d\mu_{F_n}(x) \to \int_{\mathbb{R}} x^2 d\mu_F(x)$. Therefore, by Question 11, $d_2(\mu_{F_n}, \mu_F) \to 0$ for $\omega \in A$.

14. By the CLT, $Z_n = \sqrt{n}(\bar{X}_n - \mathrm{E}X_1) \xrightarrow{d} Z$, where Z denotes a $\mathcal{N}(0,\sigma^2)$ variable, and $\mathrm{E}Z_n^2 = \mathrm{Var}(X_1) = \sigma^2 \to \mathrm{E}Z^2$. By Question 11, $d_2(Z_n, Z) \to 0$.

By Questions 8 and 9,

$$\begin{aligned}[d_2(Z_n^*, Z_n)]^2 \le \sum_{i=1}^n \frac{1}{n}[d_2(X_i^* - \mathrm{E}_*X_1^*, X_i - \mathrm{E}X_1)]^2 &= [d_2(X_1^* - \mathrm{E}_*X_1^*, X_1 - \mathrm{E}X_1)]^2 \\ &\le [d_2(X_1^*, X_1)]^2 = [d_2(\mu_{F_n}, \mu_F)]^2\end{aligned}$$

so that $d_2(Z_n^*, Z_n) \to 0$ a.s.(P) by Question 13.

Hence, $d_2(Z_n^*, Z) \le d_2(Z_n^*, Z_n) + d_2(Z_n, Z) \to 0$ a.s.(P).

15. Note $\sigma_{*n}^2 = \mathrm{Var}_*(X_1^*) = \mathrm{E}_*(X_1^*)^2 - (\mathrm{E}_*X_1^*)^2 = \sum_{i=1}^n X_i^2/n - (\bar{X}_n)^2 \to \mathrm{E}X_1^2 - [\mathrm{E}X_1]^2 = \sigma^2$ a.s.(P) by the SLLN. Also,

$$\frac{\mathrm{E}_*|X_1^*|^3}{\sqrt{n}} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{|X_i|^3}{n} \le \frac{M_n}{\sqrt{n}}\sum_{i=1}^n X_i^2/n \to 0 \text{ a.s.}(P)$$

by $M_n/\sqrt{n} \to 0$ a.s.(P) from Question 14 and $\sum_{i=1}^n X_i^2/n \to \mathrm{E}X_1^2$ a.s.(P). Hence,

$$\sup_{x\in\mathbb{R}}\left|P(Z_n^* \le x|X_1,\ldots,X_n) - \Phi\left(\frac{x}{\sqrt{\sigma_{*n}}}\right)\right| \le \frac{5}{\sqrt{n}}\frac{\mathrm{E}_*|X_1^*|^3}{(\sigma_{*n}^2)^{3/2}} \to 0 \text{ a.s.}(P).$$

When $\sigma_{*n}^2 = \mathrm{Var}_*(X_1^*) \to \sigma^2$ (which happens a.s.(P)),

$$\sup_{x\in\mathbb{R}}\left|\Phi\left(\frac{x}{\sqrt{\sigma_{*n}^2}}\right) - \Phi\left(\frac{x}{\sigma}\right)\right| \to 0.$$

(i.e., if $\sigma_{*n}^2 \to \sigma^2$ and Y is a standard normal variable, $Y\sigma_{*n} \xrightarrow{d} Y\sigma$). Hence,

$$\sup_{x\in\mathbb{R}}\left|P(Z_n^* \le x|X_1,\ldots,X_n) - \Phi\left(\frac{x}{\sigma}\right)\right| \to 0 \text{ a.s.}(P)$$

and the distribution of Z_n^*, given $X_1 \ldots, X_n$, converges to a normal $\mathcal{N}(0,\sigma^2)$, a.s.(P).

Please begin every answer to a numbered sub-part of this question on a new sheet of paper!!

Part 1

Consider a decision problem with parameters $\theta \in \Theta$, decision rules $\delta \in D$, and risk function $R(\theta,\delta)$.

1) Say carefully and completely what it means for a decision rule $\delta^* \in D$

a) to be admissible,

b) to be a Bayes rule, and

c) to be a minimax rule.

2) State and carefully prove a theorem giving sufficient conditions on Θ and a prior distribution under which a corresponding Bayes rule is necessarily admissible.

Part 2

Let $\mathcal{P}$ be a partition of the finite index set $\mathcal{I} = \{1,2,\ldots,N\}$, i.e., $\mathcal{P} = \{S_j^{\mathcal{P}}\}_{j=1}^{k^{\mathcal{P}}}$ is a collection of ($k^{\mathcal{P}}$) subsets of $\mathcal{I}$ with the property that every element of $\mathcal{I}$ belongs to exactly one $S_j^{\mathcal{P}}$.

Define a loss function for pairs of partitions measuring their dissimilarity as follows. For

$$I(i,i',\mathcal{P}) = I\left[\text{there is an } S_j^{\mathcal{P}} \in \mathcal{P} \text{ such that } i \in S_j^{\mathcal{P}} \text{ and } i' \in S_j^{\mathcal{P}}\right]$$

take

$$L(\mathcal{P}_1,\mathcal{P}_2) = \sum_{i<i'}\left[I(i,i',\mathcal{P}_1)\left(1-I(i,i',\mathcal{P}_2)\right) + I(i,i',\mathcal{P}_2)\left(1-I(i,i',\mathcal{P}_1)\right)\right] \quad . \qquad (*)$$

Then let Δ be a distribution over (the finite set of) partitions. For a random partition $\mathcal{Q} \sim \Delta$ define the probability

$$\Gamma_{i,i'}^{\Delta} = \Delta\left[\text{there is an } S_j^{\mathcal{Q}} \in \mathcal{Q} \text{ such that } i \in S_j^{\mathcal{Q}} \text{ and } i' \in S_j^{\mathcal{Q}}\right] \quad .$$

We consider representing the distribution Δ by some (representative) single "small-expected-loss-and-simple" partition. For a fixed partition $\mathcal{P}$ and random partition $\mathcal{Q} \sim \Delta$, an expected loss incurred representing Δ by $\mathcal{P}$ is

$$\mathrm{E}_{\Delta} L(\mathcal{P},\mathcal{Q}) = \sum_{q} \Delta[\mathcal{Q} = q] L(\mathcal{P},q) \quad .$$

Further, one might measure the complexity of a partition $\mathcal{P}$ by a (positive) "entropy" measure

$$C(\mathcal{P}) = -\sum_{j=1}^{k^{\mathcal{P}}} \frac{N_j^{\mathcal{P}}}{N} \ln\left(\frac{N_j^{\mathcal{P}}}{N}\right)$$

for

$$N_j^{\mathcal{P}} = \left|S_j^{\mathcal{P}}\right| = \text{the number of indices in } S_j^{\mathcal{P}} \quad .$$

Then associated with the partition $\mathcal{P}$ as representing a distribution Δ are the two non-negative measures of undesirability

$$\mathrm{E}_\Delta L(\mathcal{P},\mathcal{Q}) \text{ and } C(\mathcal{P}) \quad .$$

(A big value of either of these is bad.)

3) For a fixed partition $\mathcal{P}$ and random partition $\mathcal{Q} \sim \Delta$, find a lower bound on the expected loss $\mathrm{E}_\Delta L(\mathcal{P},\mathcal{Q})$ in terms of the probabilities $\Gamma_{ii'}^{\Delta}$.

4) Propose an admissibility criterion for this context that any partition chosen to represent Δ should satisfy, and describe the criterion's relationship to minimization of

$$\mathrm{E}_\Delta L(\mathcal{P},\mathcal{Q}) + \lambda C(\mathcal{P}) \text{ for some } \lambda > 0 \quad .$$

In a statistical context, we might think of elements of a partition as "clusters" of indices. Suppose, for example, that $X_1, X_2, \ldots, X_N$ are independent, $X_i \sim \text{Binomial}(10,\theta_i)$, and that for some partition $\mathcal{P}$, $i \in S_j^{\mathcal{P}}$ and $i' \in S_j^{\mathcal{P}}$ implies that $\theta_i = \theta_{i'}$. Given Δ a (prior) distribution over partitions and fixed probability density function (pdf) h on $(0,1)$, we might then phrase a Bayes clustering model by assuming that *a priori* $\mathcal{P} \sim \Delta$ and that given $\mathcal{P} = \left\{S_j^{\mathcal{P}}\right\}_{j=1}^{k^{\mathcal{P}}}$ there are parameters $\pi_1^{\mathcal{P}}, \pi_2^{\mathcal{P}}, \ldots, \pi_{k^{\mathcal{P}}}^{\mathcal{P}}$ that are *a priori* independent identically distributed (iid) with marginal pdf h, and that $i \in S_j^{\mathcal{P}}$ implies that $\theta_i = \pi_j^{\mathcal{P}}$.

Consider a very small case with $N = 3$, Δ uniform on partitions, and the uniform density $h(\theta) \equiv 1$ on $(0,1)$.

5) Suppose that what is observed is $X_1 = 2, X_2 = 8,$ and $X_3 = 4$. What is the posterior distribution over partitions, say $\Delta(\cdot\,|\,\mathbf{X})$? (If you find it useful, you may abbreviate the Beta function $\int_0^1 u^{\alpha-1}(1-u)^{\beta-1}\,du$ as $\mathrm{B}(\alpha,\beta)$, and you need not evaluate Beta functions.)

6) Based on an answer to **5)**, one might compute a posterior expected loss for any particular partition/clustering $\mathcal{P}$. Here, show how to compute the posterior expected loss

$$\mathrm{E}_{\Delta(\cdot|\mathbf{X})}\, L(\mathcal{P},\mathcal{Q}) \text{ for } \mathcal{Q} \sim \Delta(\cdot|\mathbf{X})$$

of the clustering/partition $\mathcal{P}$ that consists of the single set $\{1,2,3\}$.

The following fact may be helpful at some point in **Parts 3-4** of this question. These parts have nothing to do with earlier parts of the question. They concern entirely different topics.

For X with probability density function (pdf) with respect to Lebesgue measure on $\Re = (-\infty,\infty)$

$$f(x) = I[x \ge 0]\exp(-x) \ ,$$

it follows that (for $i = \sqrt{-1}$)

$$\mathrm{E}\exp(itX) = (1-it)^{-1} = \frac{1}{1+t^2} + i\frac{t}{1+t^2} \text{ for } t \in \Re \ .$$

Part 3

Suppose that g is a pdf that is positive and continuous on $[0,\infty)$. For $X_1, X_2, \ldots$ independent identically distributed (iid) with marginal pdf g, let

$$M_n \equiv n \cdot \min\{X_1, X_2, \ldots, X_n\}$$

7) Show directly that M_n converges in distribution (to an Exponential limit).

For the case where $g(x) = I[x \ge 0]\exp(-x)$, it is possible to show (don't try to do so here) that there is convergence in distribution as $n \to \infty$

$$\begin{pmatrix} M_n \\ \sqrt{n}(\bar{X}_n - 1) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} W_1 \\ W_2 \end{pmatrix}$$

where W_1 and W_2 are **independent** random variables with appropriate marginal distributions.

8) For this case of g, and for t_1 and t_2 real numbers (and $i = \sqrt{-1}$), evaluate

$$\lim_{n\to\infty} \mathrm{E}\exp\left(i\left(t_1 M_n + t_2\sqrt{n}\left(\bar{X}_n - 1\right)\right)\right) \ .$$

Part 4

Consider the parametric family of densities (with respect to Lesbegue measure on $[0,\infty)$)

$$g(x \mid \eta, \lambda) = I[x \geq \eta]\frac{1}{\lambda}\exp\left(-\frac{x-\eta}{\lambda}\right)$$

and two different statistical models for n iid observations built using it.

Model A: $Y_1, Y_2, \ldots, Y_n$ are iid with density $g(y \mid 0, \lambda)$, but what is observed are values

$$\mathbf{Z}_j = \left(Z_{1j}, Z_{2j}\right) \equiv \left(I\left[Y_j \geq \eta\right],\ Y_j \cdot I\left[Y_j \geq \eta\right]\right)$$

and $\eta > 0$ and $\lambda > 0$ are unknown parameters. (This two-parameter model for the $\mathbf{Z}_j$ has an unknown *censoring* point as one parameter.)

Model B: $X_1, X_2, \ldots, X_n$ iid with density $g(x \mid \eta, \lambda)$ are observed and $\eta > 0$ and $\lambda > 0$ are unknown parameters. (This two-parameter model for the X_j has an unknown *truncation* point as one parameter.)

9) Identify low-dimensional sufficient statistics for (η, λ) in both of Models A and B. Argue carefully that they are sufficient.

10) Say very carefully and exactly what prevents both of Models A and B from satisfying the standard Fisher Information regularity conditions.

11) Consider maximum likelihood estimators of the parameter vector (η, λ) (say $\widehat{(\eta, \lambda)}_n = \left(\hat{\eta}_n, \hat{\lambda}_n\right)$) for both of Models A and B. For Model A identify $\hat{\eta}_n$ and a single equation (involving $\hat{\eta}_n$) that must be solved to find $\hat{\lambda}_n$. For Model B identify an explicit form for $\widehat{(\eta, \lambda)}_n$.

12) For the case of Model B, what is an approximate joint distribution for your maximum likelihood estimator $\widehat{(\eta, \lambda)}_n$ in **11)**? Explain carefully. **Hint:** Consider the limiting distribution of

$$\begin{pmatrix} n(\hat{\eta}_n - \eta) \\ \sqrt{n}(\hat{\lambda}_n - \lambda) \end{pmatrix}$$

and recall the preamble before **8)**.

13) Consider testing of the hypothesis $H_0 : (\eta, \lambda) = (\eta_0, \lambda_0)$ in Model B, and in particular likelihood ratio testing of the hypothesis. For $l_n(\eta, \lambda)$ the log-likelihood for this problem and $\widehat{(\eta, \lambda)}_n$ the maximum likelihood estimator of (η, λ), the statistic

$$2\left(l_n\left(\widehat{(\eta,\lambda)}_n\right) - l_n\left((\eta_0, \lambda_0)\right)\right)$$

converges in distribution under the null hypothesis (despite the fact that standard regularity conditions are not satisfied). Derive the limiting distribution. **Hint:** Notice that a second-order Taylor expansion of $l_n(\eta, \lambda)$ about $\widehat{(\eta, \lambda)}_n$ is valid on $(0, \hat{\eta}_n] \times (0, \infty)$ despite the fact that $l_n(\eta, \lambda)$ is discontinuous along the line $\eta = \hat{\eta}_n$. You don't need to show the remainder term is negligible. (It is.)

Note Title

Theory II Key 2010 Statistics Prelim

11/28/2009

1) Book Work

2) Book Work — The simplest set of conditions is that Θ is finite and the prior puts positive mass on all elements of Θ

3)
$$E_\Delta L(P,Q) = \sum_{i<i'} I[i,i',P]\, \Delta\left[\text{there is no } S_j^Q \in Q \text{ such that } i \in S_j^Q \text{ and } i' \in S_j^Q\right]$$
$$+ \sum_{i<i'} \left(1 - I[i,i',P]\right) \Delta\left[\text{there is an } S_j^Q \in Q \text{ such that } i \in S_j^Q \text{ and } i' \in S_j^Q\right]$$
$$= \sum_{i<i'} I[i,i',P]\left(1 - \Gamma_{ii'}^{\Delta}\right) + \sum_{i<i'} \left(1 - I[i,i',P]\right) \Gamma_{ii'}^{\Delta}$$

Then for each i, i' pair, the smallest possible value of the sum of the two corresponding terms is $\min(T^{\Delta}_{ii'}, 1-T^{\Delta}_{ii'})$ so

$$E_{\Delta} L(P,Q) \geq \sum_{i<i'} \min\left(T^{\Delta}_{ii'}, 1-T^{\Delta}_{ii'}\right)$$

4) A partition P might be termed admissible if there is no other partition P^* with both

$$E_{\Delta} L(P^*, Q) \leq E_{\Delta} L(P, Q)$$

and
$$C(P^*) \leq C(P)$$

where one of the inequalities is strict. The same logic used to prove the finite 4 version of 2) will show that for $\lambda \in (0,\infty)$ a minimizer of

$$E_{\Delta} L(P,Q) + \lambda C(P)$$

will be admissible.

5) There are 5 possible partitions, namely

$$\underbrace{\{\{1,2,3\}\}}_{A}, \quad \underbrace{\{\{1\},\{2\},\{3\}\}}_{B}, \quad \underbrace{\{\{1\},\{2,3\}\}}_{C}, \quad \underbrace{\{\{2\},\{1,3\}\}}_{D}, \quad \underbrace{\{\{3\},\{1,2\}\}}_{E}$$

For these partitions, if $f(x|\theta)$ is the Binomial $(10,\theta)$ pmf, the corresponding joint pmf's are

$$A \quad f(x_1|\theta_1^A)\, f(x_2|\theta_1^A)\, f(x_3|\theta_1^A) = f(\underset{\sim}{x}|A,\underset{\sim}{\theta}^A)$$

$$B \quad f(x_1|\theta_1^B)\, f(x_2|\theta_2^B)\, f(x_3|\theta_3^B) = f(\underset{\sim}{x}|B,\underset{\sim}{\theta}^B)$$

$$C \quad f(x_1|\theta_1^C)\, f(x_2|\theta_2^C)\, f(x_3|\theta_2^C) = f(\underset{\sim}{x}|C,\underset{\sim}{\theta}^C)$$

$$D \quad f(x_1|\theta_1^D)\, f(x_2|\theta_2^D)\, f(x_3|\theta_1^D) = f(\underset{\sim}{x}|D,\underset{\sim}{\theta}^D)$$

$$E \quad f(x_1|\theta_1^E)\, f(x_2|\theta_1^E)\, f(x_3|\theta_2^E) = f(\underset{\sim}{x}|E,\underset{\sim}{\theta}^E)$$

Then with $\Delta(A) = \Delta(B) = \Delta(C) = \Delta(E) = \Delta(F) = \frac{1}{5}$ a joint density for

The partition, $\underset{\sim}{\theta}$, and $\underset{\sim}{X}$ is

$$\frac{1}{5} I\left[\text{partition is } \mathcal{P}\right] f(\underset{\sim}{x} \mid \mathcal{P}, \underset{\sim}{\theta}^{\mathcal{P}}) \cdot 1$$

for $\mathcal{P}$ a partition, $\underset{\sim}{\theta}^{\mathcal{P}}$ in $(0,1)^{k^{\mathcal{P}}}$. A joint density for the partition and $\underset{\sim}{X}$ is obtained by integrating out $\underset{\sim}{\theta}^{\mathcal{P}}$. This produces for $X_1 = 2$, $X_3 = 8$ and $X_3 = 4$

A $$\frac{1}{5}\int_0^1 \binom{10}{2} p^2(1-p)^8 \binom{10}{8} p^8(1-p)^2 \binom{10}{4} p^4(1-p)^6 \, dp = \frac{1}{5}\binom{10}{2}\binom{10}{8}\binom{10}{4} B(15,17)$$

B $$\frac{1}{5}\binom{10}{2}\binom{10}{8}\binom{10}{4} B(3,9)\, B(9,3)\, B(5,7)$$

C $$\frac{1}{5}\binom{10}{2}\binom{10}{8}\binom{10}{4} B(3,9)\, B(13,9)$$

D $$\frac{1}{5}\binom{10}{2}\binom{10}{8}\binom{10}{4} B(7,15)\, B(9,3)$$

E $$\frac{1}{5}\binom{10}{2}\binom{10}{8}\binom{10}{4} B(11,11)\, B(5,7)$$

So the posterior here is

$$\Delta(A|X) = B(15,17)/\text{sum} \qquad \Delta(B|X) = B(3,9)\,B(9,3)\,B(5,7)\,/\text{sum}$$

$$\Delta(C|X) = B(3,9)\,B(13,9)/\text{sum} \qquad \Delta(D|X) = B(7,15)\,B(9,3)/\text{sum}$$

$$\Delta(E|X) = B(11,11)\,B(5,7)/\text{sum}$$

For $\text{sum} = B(15,17) + B(3,9)B(9,3)\,B(5,7) + B(3,9)\,B(13,9) + B(7,15)B(9,3)$
$+ B(11,11)\,B(5,7)$

6) Since the candidate partition puts all indices in the same equivalence class, one needs to evaluate the sum over all pairs of indices, the posterior probabilities that the pair is in different equivalence classes — These are

$$\Delta(\text{1 and 2 are in different classes}\,|X) = \Delta(B|X) + \Delta(C|X) + \Delta(D|X)$$

$$\Delta(\text{1 and 3 are in different classes}\,|X) = \Delta(B|X) + \Delta(C|X) + \Delta(E|X)$$

$$\Delta(\text{2 and 3 are in different clases}\,|X) = \Delta(B|X) + \Delta(D|X) + \Delta(E|X)$$

So the sum of these is $3\Delta(B|X) + 2\Delta(C|X) + 2\Delta(D|X) + 2\Delta(E|X)$
and this is the posterior risk of the partition with the single class $\{1,2,3\}$

7) For $t > 0$ consider

$$P\left[M_n > t\right] = P\left[\min X_i > \frac{t}{n}\right] = P\left[\text{all } X_i > \frac{t}{n}\right]^n$$

$$= \left[1 - \int_0^{t/n} g(x)\,dx\right]^n$$

for some $0 \le a_n \le \frac{t}{n}$

$$= \left[1 - g(a_n)\frac{t}{n}\right]^n \longrightarrow e^{-g(0)t}$$

That is, the limiting distribution of M_n is Exponential with mean $1/g(0)$

8) Here $g(0)=1$ and $M_n \xrightarrow{d} \text{Exp}(1)$ while the CLT says that $\sqrt{n}\,(\bar{X}_n - 1) \xrightarrow{d} N(0,1)$. The expected value in question is the characteristic function of the random vector which must then converge to the characteristic function of the (joint) limit dsn. This is (by the facts before Part 3 and before 8) and the fact that the std normal chf is $e^{-t^2/2}$)

$$(1-it_1)^{-1}\, e^{-t_2^2/2}$$

9) <u>Model A</u>: The joint dsn of a $\underset{\sim}{Z}_j$ is concentrated on $\{(0,0)\} \cup \{(1,x) \mid x \geq 0\}$. A dominating σ-finite measure for the joint dsn is

counting measure on $\{(0,0)\}$ + Lebesgue measure on Borel sets of x's corresponding to sets of $(1,x)$ pairs

A R-N derivative of the (joint) dsn of a $\underset{\sim}{Z}$ is

$$I\left[\underset{\sim}{Z}=(0,0)\right]\left(1-\exp\left(-\frac{\eta}{\lambda}\right)\right)+I\left[Z_1=1\right]\frac{1}{\lambda}\exp\left(-\frac{Z_2}{\lambda}\right)I\left[Z_2\geq\eta\right]$$

So a joint density for all n $\underset{\sim}{Z}_j$ is

$$\left(1-\exp\left(-\frac{\eta}{\lambda}\right)\right)^{\#\left[\underset{\sim}{Z}_j=(0,0)\right]}\cdot\frac{1}{\lambda^{\#\left[Z_{1j}=1\right]}}\exp\left(-\frac{\sum Z_{2j}I\left[Z_{1j}=1\right]}{\lambda}\right)I\left[\eta\leq\min\left\{Z_{2j}\text{ with }Z_{1j}=1\right\}\right]$$

$$\left(\#\left[Z_{1j}=0\right],\ \sum_j Z_{2j}I\left[Z_{1j}=1\right],\ \min\left\{Z_{2j}\text{ with }Z_{1j}=1\right\}\right)$$

is sufficient for (η,λ)

Model B: The pdf for an X_j here is the two-parameter exponential pdf

$$\frac{1}{\lambda} \exp\left(-\frac{x-\eta}{\lambda}\right) I[x \geq \eta]$$

so that the joint pdf for the n X_j's is

$$\frac{1}{\lambda^n} \exp\left(-\sum\left(\frac{x_i-\eta}{\lambda}\right)\right) I[\eta \leq \min x_i]$$

The factorization theorem then shows that

$$\left(\min X_j, \sum X_j\right)$$

is sufficient for (η, λ)

10) Model A: The dsn for a single $\underset{\sim}{Z}$ has density

$$I[\underset{\sim}{Z}=(0,0)]\left(1-\exp\left(-\frac{\eta}{\lambda}\right)\right) + I[Z_1=1] \frac{1}{\lambda} \exp\left(-\frac{Z_2}{\lambda}\right) I[Z_2 \geq \eta]$$

Pick a parameter (η_0, λ_0) and any open ball containing it. There is an interval of η's for which the points (η, λ_0) are in that open ball. For (z_1, z_2) of the form $(1, z_2)$ with z_2 in that interval of η's, this density is not even cont^{s} as a function of (η, λ) on the open ball around (η_0, λ_0), let alone differentiable. This set of possible observations has positive probability for those (η, λ) in the ball and thus can not be simply eliminated from consideration. So regularity conditions don't hold.

<u>Model B</u>: This is the same argument, applied to the function

$$\frac{1}{\lambda} \exp\left(-\frac{x-\eta}{\lambda}\right) I[x \geq \eta]$$

and observations $x \in [0, \infty)$.

II) <u>Model A</u>: The log-likelihood here is

$$\#[z_{1j}=0] \log\left(1-\exp\left(-\frac{\eta}{\lambda}\right)\right) - \#[z_{1j}=1] \log \lambda - \frac{\sum z_{2j} I[z_{1j}=1]}{\lambda}$$

for $\lambda > 0$ and $\eta \leq \min\{z_{2j}$ with $z_{1j} = 1\}$

(and is $-\infty$ otherwise)

For fixed λ this is maximized as a function of η by

$$\hat{\eta}_n = \min\{ z_{2j} \mid z_{1j} = 1 \}$$

Then

$$\frac{\partial}{\partial\lambda} \ell(\hat{\eta}_n, \lambda) = -\#[z_{1j}=0] \frac{\exp\left(-\frac{\hat{\eta}}{\lambda}\right)\left(\frac{\hat{\eta}}{\lambda^2}\right)}{1-\exp\left(-\frac{\hat{\eta}}{\lambda}\right)} - \#[z_{1j}=1]\frac{1}{\lambda}$$

$$+ \frac{1}{\lambda^2} \sum z_{2j} I[z_{1j}=1]$$

and setting $\frac{\partial}{\partial\lambda} \ell(\hat{\eta}_n, \lambda) = 0$ and simplifying produces

$$\lambda \#[z_{1j}=1] = (-1)\#[z_{1j}=0]\, \hat{\eta}_n \frac{\exp\left(-\frac{\hat{\eta}_n}{\lambda}\right)}{1-\exp\left(-\frac{\hat{\eta}_n}{\lambda}\right)} + \sum z_{2j} I[z_{1j}=1]$$

$$\lambda = (-1)\frac{\#[z_j = 0]}{\#[z_j = 1]}\,\hat{\eta}\,\frac{\exp\left(-\frac{\hat{\eta}_n}{\lambda}\right)}{1-\exp\left(-\frac{\hat{\eta}_n}{\lambda}\right)} + \frac{\sum z_{2j} I[z_j = 1]}{\#[z_j = 1]}$$

which must be solved for λ to produce $\hat{\lambda}_n$

<u>Model B:</u> The joint pdf for the n X_j's is

$$\frac{1}{\lambda^n} \exp\left(-\sum\left(\frac{x_i - \eta}{\lambda}\right)\right) I[\eta \le \min x_i]$$

is clearly maximized as a function of η for fixed λ by $\hat{\eta}_n = \min X_i$.

Then for $l(\eta, \lambda)$ the loglikelihood

$$l(\hat{\eta}_n, \lambda) = n \log \lambda - \frac{1}{\lambda}\sum(X_i - \hat{\eta}_n)$$

which is maximized as a function of λ by $\hat{\lambda}_n = \frac{1}{n}\sum(X_i - \hat{\eta}_n)$

$$= \bar{X}_n - \hat{\eta}_n$$

12) From the preamble before 8), for large n

$$n \min(X_i - \eta) \quad \text{and} \quad \sqrt{n}\left(\overline{X}_n - (\eta+\lambda)\right)$$

are approximately independent, the first exponential with mean λ and the second normal with mean 0 and std dev λ

Now

$$\hat{\eta}_n = \min X_i = (\min X_i - \eta) + \eta = \min(X_i - \eta) + \eta$$

$$= \frac{1}{n}\left[n \min(X_i - \eta)\right] + \eta$$

and $$\overline{X}_n - \hat{\eta}_n = \frac{1}{\sqrt{n}}\left[\sqrt{n}\left(\overline{X}_n - (\eta+\lambda)\right) + \sqrt{n}(\eta+\lambda)\right] - \hat{\eta}_n$$

So $$\begin{pmatrix} \hat{\eta}_n \\ \hat{\lambda}_n \end{pmatrix} = \begin{pmatrix} \frac{1}{n}\left[n \min(X_i - \eta)\right] + \eta \\ \frac{1}{\sqrt{n}}\left[\sqrt{n}\left(\overline{X}_n - (\eta+\lambda)\right)\right] + \lambda + \frac{1}{n}\left[n \min(X_i - \eta)\right] \end{pmatrix}.$$

and $$\begin{pmatrix} n(\hat{\eta}_n-\eta) \\ \sqrt{n}(\hat{\lambda}_n-\lambda) \end{pmatrix} = \begin{pmatrix} n\min(X_i-\eta) \\ \sqrt{n}(\overline{X}_n-(\eta+\lambda)) + \frac{1}{\sqrt{n}}[n\min(X_i-\eta)] \end{pmatrix}$$

Now since $n\min(X_i-\eta)$ converges in dsn, $\frac{1}{\sqrt{n}}[n\min(X_i-\eta)]$ converges to 0 in probability. So we get convergence in dsn to a joint dsn of $\text{Exp}(\lambda) \times N(0,\lambda^2)$.

Then in approximate terms $\hat{\eta}_n$ and $\hat{\lambda}_n$ might be thought of as

1) independent

2) $\hat{\eta}_n$ with the marginal of $\eta+W$ for W exponential with mean $= \frac{\lambda}{n}$

and 3) $\hat{\lambda}_n$ with a normal dsn with mean $=\lambda$ and variance $= \frac{\lambda^2}{n}$

13) For $0<\eta\le\min X_i$ and $\lambda>0$ the loglikelihood is

$$\ell_n(\eta,\lambda) = -n\log\lambda - \frac{1}{\lambda}\sum(X_i-\eta)$$

Partials are

$$\frac{\partial}{\partial\lambda}\ell_n = -\frac{n}{\lambda} + \frac{1}{\lambda^2}\sum(X_i-\eta) \qquad \frac{\partial}{\partial\eta}\ell_n = \frac{n}{\lambda}$$

$$\frac{\partial^2\ell_n}{\partial\lambda^2} = \frac{n}{\lambda^2} - \frac{2}{\lambda^3}\sum(X_i-\eta) \qquad \frac{\partial^2}{\partial\lambda\partial\eta}\ell_n = \frac{\partial^2}{\partial\eta\partial\lambda}\ell_n = -\frac{n}{\lambda^2} \qquad \frac{\partial^2}{\partial\eta^2} = 0$$

Then a second-order Taylor expansion of ℓ_n at $\widehat{(\eta,\lambda)}_n$ gives

$$2\left(\ell_n\left(\widehat{(\eta,\lambda)}_n\right) - \ell_n\left((\eta_0,\lambda_0)\right)\right) \approx -2\left(\eta_0-\hat\eta_n,\ \lambda_0-\hat\lambda_n\right)\begin{pmatrix} n/\hat\lambda_n \\ 0 \end{pmatrix}$$

$$- \left(\eta_0 - \hat{\eta}_n, \lambda_0 - \hat{\lambda}_n\right) \begin{pmatrix} 0 & -\frac{n}{\hat{\lambda}_n^2} \\ -\frac{n}{\hat{\lambda}_n^2} & -\frac{n}{\hat{\lambda}_n^2} \end{pmatrix} \begin{pmatrix} \eta_0 - \hat{\eta}_n \\ \lambda_0 - \hat{\lambda}_n \end{pmatrix}$$

converges to std normal in dsn by slutzky

$$= 2\left(\hat{\eta}_n - \eta_0\right)\frac{n}{\hat{\lambda}_n} + \frac{n}{\hat{\lambda}_n^2}\left(\hat{\lambda}_n - \lambda_0\right)^2 + 2\left(\hat{\lambda}_n - \lambda_0\right)\left(\hat{\eta}_n - \eta_0\right)\frac{n}{\hat{\lambda}_n^2}$$

converges in dsn under (η_0, λ_0)

converges to Exp(1) by slutzky independent of

converges to 0 in probability under (η_0, λ_0)

converges to λ_0 in probability under (η_0, λ_0)

Since twice an Exp(1) r.v. is χ^2_2 we have that the likelihood ratio statistic is approximately the sum of independent χ^2_2 and χ^2_1 variables and the limiting distribution is χ^2_3.

Part I. Assume that (X, Y) has the bivariate probability density function (pdf)

$$f(x,y) = \begin{cases} C_\theta x^{\theta-1}\exp(-y) & \text{if } 0 < x < y < \infty \\ 0 & \text{otherwise} \end{cases}$$

for some C_θ with $\theta > 0$.

1. Determine the constant C_θ in terms of the gamma function $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1}e^{-t}dt$, $\alpha > 0$.
2. Find the range of θ for which $E(X^{-2})$ exists.
3. Find the joint moment generating function of (X, Y).
4. Show that X and $Y - X$ are independent.
5. Find the pdf for the marginal distribution of Y.

Part II. Let $X_1, \cdots, X_n$ be independently and identically distributed random variables with the probability density function (pdf)

$$f_X(x;\theta) = \begin{cases} \theta e^{-\theta x} & \text{if } x > 0 \\ 0 & \text{otherwise,} \end{cases}$$

and $\theta > 0$. Instead of observing $X_1, \cdots, X_n$, suppose that we observe $Y_1, \cdots, Y_n$ where each Y_i is independently degenerated from X_i by the following steps:

[Step 1] Generate X_i from $f_X(x;\theta)$ above.

[Step 2] Generate U_i from the Exp(1) distribution, independently from X_i, where Exp(1) is the exponential distribution with mean 1.

[Step 3] If $X_i \le U_i$, then set $Y_i = X_i$. Otherwise, go to Step 1.

6. Show that Y_1 has pdf given by

$$f_Y(y;\theta) = (\theta+1)e^{-(1+\theta)y}, \quad y > 0.$$

[Hint: use Bayes formula to find the cumulative distribution function of Y_1.]

7. Based on the observations, $Y_1, \cdots, Y_n$, find the maximum likelihood estimator $\hat{\theta}$ of θ.

8. Find the limiting distribution of $\sqrt{n}(\hat{\theta} - \theta)$ as $n \to \infty$.

9. Determine a variance stabilizing transformation $g : (0, \infty) \to \mathbb{R}$ such that $\sqrt{n}[g(\hat{\theta}) - g(\theta)]$ is asymptotically pivotal as $n \to \infty$.

Part III. Let X be a random variable. We wish to test

$$H_0 : X \sim f_0 \quad \text{against} \quad H_1 : X \sim f_1$$

where f_0 and f_1 are two probability density functions.

10. Provide a decision rule $d(X)$, where $d(X) = 1$ if H_0 is rejected and $d(X) = 0$ otherwise, such that $d(X)$ minimizes the quantity

$$a\alpha + b\beta,$$

where α and β are the probabilities of type I and type II errors, respectively, and a and b are two given positive constants.

1. $C_\theta = 1/\Gamma(\theta)$

2. The marginal distribution of X is $Gamma\,(\theta, 1)$ distribution. Thus, $E\left(X^k\right) = \Gamma(\theta + k)/\Gamma(\theta)$ if $\theta + k > 0$. Thus, $E\left(X^{-2}\right)$ exists if $\theta > 2$.

3. Joint MGF:

$$\begin{aligned}
M_{X,Y}(t_1, t_2) &= E\{\exp(t_1 X + t_2 Y)\} \\
&= \int_0^\infty \int_x^\infty C_\theta x^{\theta-1} \exp(t_1 x + t_2 y)\exp(-y)dydx \\
&= (1-t_2)^{-1}\int_0^\infty C_\theta x^{\theta-1}\exp(t_1 x)\exp\{-(1-t_2)x\}dx, \quad 1-t_2 > 0 \\
&= (1-t_2)^{-1}(1-t_1-t_2)^{-\theta}, \quad 1-t_1-t_2>0,\ 1-t_2>0
\end{aligned}$$

4. The proof follows by showing that the joint MGF of X and $Y - X$ is decomposed as a product of the marginal MGF's:

$$\begin{aligned}
E\{\exp[t_1 X + t_2(Y-X)]\} &= E\{\exp[(t_1 - t_2)X + t_2 Y]\} \\
&= (1-t_1)^{-\theta} I_{(-\infty,1)}(t_1) \times (1-t_2)^{-1} I_{(-\infty,1)}(t_2) \\
&= E\{\exp(t_1 X)\}\, E\{\exp[t_2(Y-X)]\}.
\end{aligned}$$

5. Since $X \sim Gamma\,(\theta, 1)$, $Y - X \sim Gamma\,(1,1)$, and X and $Y-X$ are independent, the marginal distribution of $Y = X + (Y - X)$ is $Gamma\,(\theta+1, 1)$. Thus, the pdf for the marginal distribution of Y is

$$f_Y(y;\theta) = \frac{1}{\Gamma(\theta+1)} y^\theta \exp(-y)$$

for $y > 0$.

6. Note that, by Bayes theorem, for $y > 0$,

$$\begin{aligned}
Pr\{Y \le y\} &= Pr\{X \le y \mid U \le X\} \\
&= \frac{Pr(X \le y, U \le X)}{Pr(U \le X)} \\
&= \frac{\int_0^y \int_u^\infty e^{-u} f_X(x;\theta)\,dxdu}{\int_0^\infty \int_u^\infty e^{-u} f_X(x;\theta)\,dxdu} \\
&= \frac{\int_0^y e^{-(\theta+1)u}du}{\int_0^\infty e^{-(\theta+1)u}du}.
\end{aligned}$$

Thus, the pdf of Y is

$$f_Y(y;\theta) = (\theta+1)\,e^{-(\theta+1)y}, \quad y>0.$$

7. The likelihood function is

$$L(\theta) = (\theta+1)^n \exp\left\{-(\theta+1)\sum_{i=1}^n y_i\right\}$$

The maximum value of $L(\theta)$ can be obtained by solving

$$\frac{\partial}{\partial\theta}\ln L(\theta) = 0$$

which leads to

$$\hat{\theta} = 1/\bar{y} - 1$$

where $\bar{y} = \sum_{i=1}^n y_i/n$. The MLE $\hat{\theta}$ of θ is

$$\hat{\theta} = -1 + \left(\sum_{i=1}^n y_i/n\right)^{-1}.$$

8. The limiting distribution is

$$\sqrt{n}\left(\hat{\theta}-\theta\right) \to N(0, 1/I(\theta))$$

where $I(\theta) = E\{-\partial^2 \ln f(X;\theta)/\partial\theta^2\} = (1+\theta)^{-2}$. Thus, we have

$$\sqrt{n}\left(\hat{\theta}-\theta\right) \to N\left(0, (1+\theta)^2\right).$$

9. Choose a transformation g so that the derivative $g'(\theta)$ at $\theta>0$ satisfies $[g'(\theta)(1+\theta)]^2 = 1$ or $g'(\theta) = 1/(1+\theta)$, so that $g(\theta) = \log(\theta+1)$ will do. Then,

$$\sqrt{n}\left[g(\hat{\theta}) - g(\theta)\right] \to N\left(0, [g'(\theta)(1+\theta)]^2 = 1\right)$$

has a limiting standard normal for all $\theta>0$.

10. Consider a decision function $d(x)$ such that

$$d(x) = \begin{cases} 1 & \text{if we reject } H_0 \\ 0 & \text{otherwise.} \end{cases}$$

Then, $\alpha = Pr\{d(X) = 1 \mid \theta = 0\}$ and $\beta = Pr\{d(X) = 0 \mid \theta = 1\}$ where

$$\theta = \begin{cases} 0 & \text{if } X \sim f_0 \\ 1 & \text{if } X \sim f_1, \end{cases}$$

and thus $D = aPr\{d(X) = 1 \mid \theta = 0\} + bPr\{d(X) = 0 \mid \theta = 1\}$. Now, consider a loss function

$$L[\theta, d(x)] = \begin{cases} a & \text{if } \theta = 0, d(x) = 1 \\ b & \text{if } \theta = 1, d(x) = 0 \\ 0 & \text{otherwise.} \end{cases}$$

If we consider a uniform prior $\pi_i = P(\theta = i) = 1/2$, $i = 0, 1$, we get the risk function

$$R(\theta, d) = \begin{cases} aP_\theta\{d(X) = 1\} & \text{if } \theta = 0 \\ bP_\theta\{d(X) = 0\} & \text{if } \theta = 1. \end{cases}$$

Since $E\{R(\theta, d)\} = 0.5aP_0\{d(X) = 1\} + 0.5bP_1\{d(X) = 0\} = 0.5D$, minimizing D is equivalent to finding the Bayes solution. Note that the posterior distribution is

$$h(\theta \mid x) = \frac{f_\theta(x)}{f_0(\theta) + f_1(\theta)}, \quad \theta = 0, 1.$$

Now,

$$E\{L[\theta, d(X)] \mid X = x\} = \begin{cases} b\frac{f_1(x)}{f_0(x)+f_1(x)} & \text{if } d(x) = 0 \\ a\frac{f_0(x)}{f_0(x)+f_1(x)} & \text{if } d(x) = 1 \end{cases}$$

is minimized by choosing $d(x)$ such that

$$d(x) = \begin{cases} 0 & \text{if } b\frac{f_1(x)}{f_0(x)+f_1(x)} \leq a\frac{f_0(x)}{f_0(x)+f_1(x)} \\ 1 & \text{if } b\frac{f_1(x)}{f_0(x)+f_1(x)} > a\frac{f_0(x)}{f_0(x)+f_1(x)}. \end{cases}$$

Therefore, the optimal decision rule is to reject H_0 if

$$\frac{f_1(x)}{f_0(x)} > \frac{a}{b}.$$

Suppose discrete random variable $X \geq 0$, taking values in the non-negative integers, has probability mass function $p(x)$. The probability generating function (pgf) for X is $g(s) = E\left[s^X\right] = \sum_{x=0}^{\infty} p(x)s^x$ whenever this sum converges.

1. Write down the pgf of a Bernoulli(p) random variable, $0 \leq p \leq 1$.
2. Verify that $\frac{p}{1-(1-p)s}$, $s < \frac{1}{1-p}$ is the pgf of a geometric random variable with probability mass function $p(x) = p(1-p)^x, x = 0, 1, \ldots$, and $0 < p < 1$.

Another distribution defined on the nonnegative integers has pgf

$$g(s) = \frac{(1-p) + (2p-1)s}{1-(1-p)s}. \tag{1}$$

Assume $0 < p < 1$ for the rest of this problem.

3. What is $p(0) = P(X = 0)$ for X with pgf given by (1)?
4. What is $E(X)$ for X with pgf given by (1)?
5. What is $E[X(X-1)]$ for X with pgf given by (1)?

Consider a population of reproducing particles, where all particles die synchronously at discrete time steps. When a particle dies, it gives birth to a random number of new particles. Let Z_n count the number of particles in the population after n time steps, and suppose the population initiates with $Z_0 = 1$ particle. At the end of the first time step, the original particle dies and gives birth to $Z_1 = X_{11}$ particles. At the end of the second time step, the Z_1 particles die and give birth to a total of $Z_2 = X_{21} + X_{22} + \cdots + X_{2Z_1}$ particles, where X_{2j} is the number of particles contributed by dying particle $j \in \{1, \ldots, Z_1\}$. In general, Z_{n-1} particles die and give birth to a total of $Z_n = X_{n1} + \cdots + X_{nZ_{n-1}}$ particles at time step n.

Assume the family sizes $\{X_{ij}\}$ are independent and identically distributed random variables with a common pgf $g(s)$.

6. Show that the probability generating function of Z_n satisfies

$$g_n(s) = g_{n-1}\left[g(s)\right], \tag{2}$$

where $g_{n-1}(s)$ is the pgf for Z_{n-1} and $g(s)$ is the common pgf for all X_{ij}.

In the next seven questions, through question 12, assume the pgf $g(s)$ is given by Eq. (1).

7. Use Eq. (2) to show that $E(Z_n) = 1$ for all n.

8. On the other hand, show that $\mathrm{Var}(Z_n) \to \infty$ as $n \to \infty$.

By induction, the functional recursion (2) delivers $g_n(s)$ as the n composition of $g(s)$, *i.e.*,

$$g_n(s) = \underbrace{g(g(\cdots g(s)))}_{n \text{ times}}.$$

Furthermore, the pgf of Eq. (1) is closed under functional composition, yielding

$$g_n(s) = \frac{n(1-p) - [n(1-p) - p]s}{p + n(1-p) - n(1-p)s}.$$

Now, suppose data are collected in order to estimate p of Eq. (1). In a single experiment, a population is initiated with one particle, $Z_0 = 1$, and the number of particles, Z_2, at the second time point is recorded. The experiment was independently repeated m times, and m_0 populations of size 0, m_1 populations of size 1, and no other population sizes were observed at time point 2 (*i.e.*, $m_0 + m_1 = m$).

9. Obtain a maximum likelihood estimate of p.

The population is said to be extinct by generation n if $Z_n = 0$.

10. What is the probability, $P(Z_n = 0)$, that the process is extinct by generation n?
11. What would be a maximum likelihood estimate of this probability given the data (m_0, m_1)?
12. Show that $Z_n \to 0$ in distribution as $n \to \infty$.

Henceforth, consider the general setup described after question 5 with the arbitrary probability generating function $g(s)$.

13. Let G be the number of time steps until extinction, so $G = \min\{i \geq 1 : Z_i \geq 1, Z_{i+1} = 0\}$, and define $f_k = P(G \leq k)$. Show that

$$f_{k+1} = g(f_k)$$

for any pgf $g(s)$.

14. How can f_k be used to compute $E[G]$?

Suppose discrete random variable $X \geq 0$, taking values in the non-negative integers, has probability mass function $p(x)$. The probability generating function (pgf) for X is $g(s) = E\left[s^X\right] = \sum_{x=0}^{\infty} p(x)s^x$ whenever this sum converges.

1. Write down the pgf of a Bernoulli(p) random variable, $0 \leq p \leq 1$.

Solution:

$$h(s) = (1-p) + ps$$

2. Verify that $\frac{p}{1-(1-p)s}$, $s < \frac{1}{1-p}$ is the pgf of a geometric random variable with probability mass function $p(x) = p(1-p)^x, x = 0, 1, \ldots,$ and $0 < p < 1$.

Solution:

$$\begin{aligned} h(s) &= \sum_{x=0}^{\infty} p(1-p)^x s^x \\ &= p \sum_{x=0}^{\infty} [s(1-p)]^x \\ &= \frac{p}{1-(1-p)s}, \text{ assuming } s(1-p) < 1 \end{aligned}$$

Or one could work backwards, proving $\frac{d^x h(s)}{ds^n}(0) = p(x)$ by an induction argument.

Another distribution defined on the nonnegative integers has pgf

$$g(s) = \frac{(1-p) + (2p-1)s}{1-(1-p)s}. \tag{1}$$

Assume $0 < p < 1$ for the rest of this problem.

3. What is $p(0) = P(X = 0)$ for X with pgf given by (1)?

Solution: $p(0) = g(0) = 1 - p$

4. What is $E(X)$ for X with pgf given by (1)?

Solution: $E(X) = g'(1)$, so first find the derivative

$$\begin{aligned} g'(s) &= \frac{[1-(1-p)s](2p-1)+(1-p)[(1-p)+(2p-1)s]}{[1-(1-p)s]^2} \\ &= \frac{p^2}{[1-(1-p)s]^2} \\ g'(1) &= \frac{p^2}{p^2} = 1 \end{aligned}$$

5. What is $E[X(X-1)]$ for X with pgf given by (1)?

Solution: $E[X(X-1)] = g''(1)$, so

$$\begin{aligned} g''(s) &= \frac{(1-p)p^2}{[1-(1-p)s]^4} \\ g''(1) &= \frac{1-p}{p^2} \end{aligned}$$

Consider a population of reproducing particles, where all particles die synchronously at discrete time steps. When a particle dies, it gives birth to a random number of new particles. Let Z_n count the number of particles in the population after n time steps, and suppose the population initiates with $Z_0 = 1$ particle. At the end of the first time step, the original particle dies and gives birth to $Z_1 = X_{11}$ particles. At the end of the second time step, the Z_1 particles die and give birth to a total of $Z_2 = X_{21} + X_{22} + \cdots + X_{2Z_1}$ particles, where X_{2j} is the number of particles contributed by dying particle $j \in \{1, \ldots, Z_1\}$. In general, Z_{n-1} particles die and give birth to a total of $Z_n = X_{n1} + \cdots + X_{nZ_{n-1}}$ particles at time step n.

Assume the family sizes $\{X_{ij}\}$ are independent and identically distributed random variables with a common pgf $g(s)$.

6. Show that the probability generating function of Z_n satisfies

$$g_n(s) = g_{n-1}[g(s)], \tag{2}$$

where $g_{n-1}(s)$ is the pgf for Z_{n-1} and $g(s)$ is the common pgf for all X_{ij}.

Solution:

$$\begin{aligned} E\left[s^{Z_n}\right] &= E\left[E\left[s^{X_{n1}+\cdots+X_{nt}} \mid Z_{n-1}=t\right]\right] \\ &= E\left[E\left[s^{X_{n1}} \mid Z_{n-1}=t\right] \cdots E\left[s^{X_{nt}} \mid Z_{n-1}=t\right]\right] \quad \text{by independence of } X_{ij} \\ &= E\left\{[g(s)]^{Z_{n-1}}\right\} = g_{n-1}\left[g(s)\right] \end{aligned}$$

In the next seven questions, through question 12, assume the pgf $g(s)$ is given by Eq. (1).

7. Use Eq. (2) to show that $E(Z_n) = 1$ for all n.

Solution: By chain rule for derivatives, we have

$$g_n'(s) = g_{n-1}'\left[g(s)\right] g'(s)$$

In addition $g(1) = g'(1) = 1$. Thus,

$$g_n'(1) = g_{n-1}'(1) = g_{n-2}'(1) = \cdots = g_1'(1)$$

and $g_1'(1) = g(1) = 1$, so $E[Z_n] = g_n'(1) = 1$.

8. On the other hand, show that $\mathrm{Var}(Z_n) \to \infty$ as $n \to \infty$.

Solution: Note $g_n''(1) = E[Z_n(Z_n - 1)] = \mathrm{Var}(Z_n)$ because $E(Z_n) = 1$. Thus,

$$\begin{aligned} g_n''(s) &= g_{n-1}''\left[g(s)\right]\left[g'(s)\right]^2 + g_{n-1}'\left[g(s)\right] g''(s) \\ g_n''(1) &= g_{n-1}''(1) + \frac{1-p}{p^2} \\ g_n''(1) &= \frac{n(1-p)}{p^2} \end{aligned}$$

which approaches ∞ as $n \to \infty$.

By induction, the functional recursion (2) delivers $g_n(s)$ as the n composition of $g(s)$, *i.e.*,

$$g_n(s) = \underbrace{g(g(\cdots g(s)))}_{n \text{ times}}.$$

Furthermore, the pgf of Eq. (1) is closed under functional composition, yielding

$$g_n(s) = \frac{n(1-p) - [n(1-p) - p]s}{p + n(1-p) - n(1-p)s}.$$

Now, suppose data are collected in order to estimate p of Eq. (1). In a single experiment, a population is initiated with one particle, $Z_0 = 1$, and the number of particles, Z_2, at the second time point is recorded. The experiment was independently repeated m times, and m_0 populations of size 0, m_1 populations of size 1, and no other population sizes were observed at time point 2 (*i.e.*, $m_0 + m_1 = m$).

9. Obtain a maximum likelihood estimate of p.

Solution:

$$g_2(s) = \frac{2(1-p) - (2-3p)s}{2 - p - 2(1-p)s}$$

so

$$g_2(0) = P(Z_2 = 0) = \frac{2(1-p)}{2-p}$$

Also,

$$\frac{dg_2(s)}{ds} = \frac{2(1-p)[2(1-p) - (2-3p)s] - (2-3p)[2 - p - 2(1-ps)]}{[2-p-2(1-p)s]^2}$$

so

$$\frac{dg_2(0)}{ds} = P(Z_2 = 1) = \frac{p^2}{(2-p)^2}$$

The likelihood is

$$L(p) = \left(\frac{2(1-p)}{2-p}\right)^{m_0} \left(\frac{p^2}{(2-p)^2}\right)^{m_1} \propto (1-p)^{m_0} p^{2m_1} (2-p)^{-(m_0+2m_1)}$$

with log likelihood

$$l(p) = m_0 \log(1-p) + 2m_1 \log(p) - (m_0 + 2m_1)\log(2-p).$$

Maximizing this analytically yields solution

$$\hat{p} = \frac{4m_1}{4m_1 + m_0}.$$

The population is said to be extinct by generation n if $Z_n = 0$.

10. What is the probability, $P(Z_n = 0)$, that the process is extinct by generation n?

Solution:
$$g_n(0) = \frac{n(1-p)}{p + n(1-p)}$$

11. What would be a maximum likelihood estimate of this probability given the data (m_0, m_1)?

Solution: By functional invariance,
$$\widehat{g}_n(0) = \frac{n(1-\hat{p})}{\hat{p} + n(1-\hat{p})} = \frac{nm_0}{4m_1 + nm_0}.$$

12. Show that $Z_n \to 0$ in distribution as $n \to \infty$.

Solution: $g_n(0)$ is the distribution function evaluated at 0:
$$g_n(0) = P(Z_n = 0) = P(Z_n \leq 0),$$
and
$$\lim_{n\to\infty} g_n(0) = \lim_{n\to\infty} \frac{n(1-p)}{p + n(1-p)} = 1$$
shows the distribution function $P(Z_n \leq z)$ converges to the degenerate distribution located at 0.

Henceforth, consider the general setup described after question 5 with the arbitrary probability generating function $g(s)$.

13. Let G be the number of time steps until extinction, so $G = \min\{i \geq 1 : Z_i \geq 1, Z_{i+1} = 0\}$, and define $f_k = P(G \leq k)$. Show that
$$f_{k+1} = g(f_k)$$
for any pgf $g(s)$.

Solution: Use the law of total probability and condition on the number of particles Z_1 at time step 1. For exposition, let G_i be the number of generations it takes for the population initiated by the ith particle in generation 1 to go extinct, $i \in \{1, \dots, Z_1\}$.

$$\begin{aligned} f_{k+1} = P(G \leq k+1) &= \sum_{i=0}^{\infty} P(G_1 \leq k, \dots, G_i \leq k \mid Z_1 = i) P(Z_1 = i) \\ &= \sum_{i=0}^{\infty} P(G_1 \leq k) \cdots P(G_i \leq k) P(X_{11} = i) \\ &= \sum_{i=0}^{\infty} (f_k)^i p(i) \\ &= g(f_k) \end{aligned}$$

14. How can f_k be used to compute $E[G]$?

Solution:

$$E[G] = \sum_{k=0}^{\infty} (1 - f_k)$$

Part I. Let X_1 and X_2 be independent and identically distributed (iid) Bernoulli observations with $P_p[X_1 = 1] = p = 1 - P_p[X_1 = 0]$ where $0 \le p \le 1$ is unknown. It is desired to estimate $\theta = P_p[X_1 = X_2]$.

1. What is the range of θ?
2. Find an unbiased estimator of θ.
3. Find the uniformly minimum-variance unbiased estimator (UMVUE) of θ. Justify your answer. Explain why this estimator may or may not be reasonable.
4. Find the maximum likelihood estimator (MLE) of θ.

Part II. Let $X_1, ..., X_n$ be a random sample with probability density function (pdf) given by

$$f(x|\theta) = \frac{2}{\sqrt{\pi\theta}} \exp\left(-\frac{x^2}{\theta}\right), \quad x > 0,$$

where $\theta > 0$.

Note: If needed, you may use the following facts: $\frac{2X_1^2}{\theta} \sim \chi_1^2$, a chi-square variable with 1 degree of freedom. The mean and variance of a chi-square variable χ_1^2 are 1 and 2, respectively.

5. Find the CRLB (Cramèr-Rao Lower Bound) for estimating θ.
6. Using the CRLB, find the UMVUE of θ.
7. Given $\theta_0 > 0$ and test size $\alpha \in (0, 1)$, show that the likelihood ratio test for $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \ne \theta_0$ is given by

   $$\text{Reject } H_0 \text{ if } W_n < \theta_0 c_1 \text{ or } W_n > \theta_0 c_2$$

   where $0 < c_1 < c_2 < \infty$ are appropriate constants and $W_n = 2\sum_{i=1}^n X_i^2$. Use chi-square quantiles to describe the required form of c_1, c_2.

   Note: An explicit numerical determination of c_1, c_2 is not possible for a given $\alpha \in (0, 1)$.
8. Use the likelihood ratio test in 7. to find a confidence interval for θ with confidence coefficient $1 - \alpha$. Justify your solution.

Part III. Let (X_i, Y_i), $i = 1, ..., n$ be a random sample from a (bivariate) uniform distribution defined on a circle centered at (0,0) and with radius $\theta > 0$. That is, the joint probability density function of (X_1, Y_1) is given by

$$f(x, y) = \begin{cases} \frac{1}{\pi\theta^2} & \text{if } \sqrt{x^2 + y^2} < \theta \\ 0 & \text{otherwise.} \end{cases}$$

Define $R_n = \max_{i=1,...,n} \sqrt{X_i^2 + Y_i^2}$

Note: If needed, you may use the following: $P_\theta[X_1^2 + Y_1^2 \leq s^2] = (s/\theta)^2$, $s \in [0, \theta]$.

9. For $0 \leq r \leq \theta$, show that $P_\theta(R_n \leq r) = (r/\theta)^{2n}$.

10. Show that R_n is a mean squared error consistent estimator of θ.

11. Suppose a confidence interval for θ of the form $I_n = (0, R_n 5^{1/(2n+\sqrt{n})}]$. For a given $\theta > 0$, determine the coverage probability of I_n as $n \to \infty$.

12. Consider a prior for θ which has a density given by

$$\pi(\theta) = \begin{cases} \theta^{n-1} & \theta > 0 \\ 0 & \text{otherwise.} \end{cases}$$

    Using this prior and random sample $(X_1, Y_1), \ldots, (X_n, Y_n)$, $n \geq 2$, find the Bayes estimator of θ under the loss function $L(t, \theta) = (t - \theta)^2/\theta$, where $t \in \mathbb{R}$, $\theta > 0$.

13. Consider a statistic $T_n = 1$ if $R_n < 2^{-1/n}$ and, otherwise, $T_n = 0$. Using the prior in 12., show that T_n is the simple Bayes test of $H_0 : \theta \geq 1$ vs $H_1 : \theta < 1$.

14. For testing $H_0 : \theta \geq 1$ vs $H_1 : \theta < 1$, give the power function of the test rule T_n in 13. as well as the size of this test rule.

1. $\theta = P_p(X_1 = X_2) = P_p(X_1 = 0, X_2 = 0) + P_p(X_1 = 1, X_2 = 1) = p^2 + (1-p)^2 \in [1/2, 1]$ for $p \in [0, 1]$. The parameter θ has a minimum of $1/2$ when $p = 1/2$.

2. Using the indicator function $\mathbb{I}(\cdot)$, the estimator $Y = \mathbb{I}[X_1 = X_2]$ (1 if $X_1 = X_2$ and 0 if $X_1 \neq X_2$) is an unbiased estimator of θ as $\theta = P_p(X_1 = X_2) = \mathrm{E}_p Y$.

3. The Bernoulli(p) model is a regular exponential family and, from this, it can be shown that $S = X_1 + X_2$ is a complete and sufficient statistic. Note that $Y = \mathbb{I}[X_1 = X_2] = \mathbb{I}[S \neq 1]$ is unbiased for θ and a function of the complete and sufficient statistic S. Therefore, Y is the UMVUE of θ by the Lehmann-Scheffe theorem.

   Alternatively, because Y is unbiased for θ, the UMVUE could be derived as

   $$\mathrm{E}[Y|S = s] = P(X_1 = X_2|S = s) = P_p(X_1 = X_2, S = s)/P_p(S = s)$$

   which equals 1 if $s = 0, 2$ and equals 0 if $s = 1$. That is, $\mathrm{E}[Y|S] = \mathbb{I}[S \neq 1] = Y$. (The above conditional probability as the ratio of probabilities is well-defined for any $s = 0, 1, 2$ when $0 < p < 1$; when $p = 0$, $0 = X_1 = X_2 = S$ and $E[Y|S] = 1$ deterministically and when $p = 1$, $X_1 = X_2 = 1, S = 2, E[Y|S] = 1$.)

   Y is not a reasonable estimator because it can lie outsidfe the range of θ.

4. The MLE of p is $\hat{p} = (X_1 + X_2)/2$ in the Bernoulli model, so that $\hat{\theta} = \hat{p}^2 + (1 - \hat{p})^2$ is the MLE of θ.

5. As this is a regular exponential family, the CRLB for θ is $[d\theta/d\theta]^2/I_n(\theta) = 1/(nI_1(\theta))$, where $I_n(\theta) = nI_1(\theta)$ is the Fisher information number for θ based on $X_1, \ldots, X_n$ and $I_1(\theta)$ is the Fisher information number based on X_1. We have

   $$I_1(\theta) = -\mathrm{E}_\theta \frac{d^2 \log f(X_1|\theta)}{d\theta^2} = \mathrm{E}_\theta \left(\frac{2X_1^2}{\theta^3} - \frac{1}{2\theta^2} \right) = \frac{1}{2\theta^2}$$

   using $\mathrm{E}_\theta(2X_1^2/\theta) = \mathrm{E}\chi_1^2 = 1$. So the CRLB for θ is $2\theta^2/n$.

6. Let $T = 2\sum_{i=1}^n X_i^2/n$. Then $\mathrm{E}_\theta T = \theta \mathrm{E}_\theta(2X_1^2/\theta) = \theta$ so that T is an unbiased estimator of θ and $\mathrm{Var}_\theta(T) = \theta^2 \mathrm{Var}_\theta(2X_1^2/\theta)/n = \theta^2 \mathrm{Var}(\chi_1^2)/n = 2\theta^2/n$ =CRLB. Therefore, T is the UMVUE of θ.

7. Given positive data $X_1, \dots, X_n$, the likelihood function for θ is

$$L(\theta) = (2/\sqrt{\pi})^n \theta^{-n/2} \exp[-\sum_{i=1}^{n} X_i^2/\theta]$$

and the MLE is $\hat{\theta} = W/n$, where $W = 2\sum_{i=1}^{n} X_i^2$. The likelihood ratio test statistic is then

$$\lambda_n(\theta_0) = \frac{L(\theta_0)}{L(\hat{\theta})} = n^{-n/2}(W_{0n})^{n/2} \exp[(n - W_{0n})/2] = g(W_{0n})$$

where $W_{0n} = W/\theta_0$ and $g(x) = (x/n)^{n/2} \exp[(n-x)/2]$, $x > 0$. The size $\alpha \in (0,1)$ likelihood ratio test is given by rejecting H_0 if $\lambda_n(\theta_0) < \lambda$ where $\lambda \in (0,1)$ is chosen so that $P_{\theta_0}(\lambda_n(\theta_0) < \lambda) = \alpha$.

Note $\lim_{x\to 0^+} g(x) = 0 = \lim_{x\to\infty} g(x)$ and $g(\cdot)$ is concave down with a maximum of $g(n) = 1$ at $x = n$; that is, $g(\cdot)$ is increasing on $(0, n]$ and decreasing on $[n, \infty)$ which can be checked from $\log g(x) = (n/2)\log x - (n-x)/2 - n(n/2)$, $x > 0$. So for $\lambda \in (0,1)$, $g(W_{0n}) = \lambda_n(\theta_0) < \lambda$ if and only if $W_{0n} < c_1$ or $W_{0n} > c_2$ where $0 < c_1 < n < c_2$ are constants for which $g(c_1) = g(c_2) = \lambda$. Hence, the likelihood ratio test is equivalent to rejecting $H_0 : \theta = \theta_0$ if $W_{0n} < c_1$ or $W_{0n} > c_2$ where constants $0 < c_1 < n < c_2$ are chosen so that $g(c_1) = g(c_2)$ and

$$\alpha = P_{\theta_0}(W_{0n} < c_1 \text{ or } W_{0n} > c_2) = 1 - P_{\theta_0}(c_1 \le W_{0n} \le c_2).$$

Under $H_0 : \theta = \theta_0$, W_{0n} has a χ^2_n distribution. If $\chi^2_{n,\gamma}$ denotes the γ-quantile of a χ^2_n variable (i.e., $P(\chi^2_n \le \chi^2_{n,\gamma}) = \gamma \in (0,1)$), pick $c_1 = \chi^2_{n,\gamma_1}$ and $c_2 = \chi^2_{n,\gamma_2}$ so that $\gamma_2 - \gamma_1 = 1 - \alpha$. That is, we need to choose $\gamma_1 \in (0, \alpha)$ so that $g(c_1) = g(c_2)$ for $c_1 = \chi^2_{n,\gamma_1}$ and $c_1 = \chi^2_{n,1-\alpha+\gamma_1}$.

8. The acceptance region for the size α likelihood ratio test of $H_0 : \theta = \theta_0$ is $A(\theta_0) = \{(Y_1, \dots, Y_n) : c_1 \le 2\sum_{i=1}^{n} Y_i^2/\theta_0 \le c_2\}$. Given positive data $X_1, \dots, X_n$ and inverting the test, a confidence region for θ is

$$\{\theta > 0 : (X_1, \dots, X_n) \in A(\theta)\} = \left\{\theta > 0 : c_1 \le \frac{2\sum_{i=1}^{n} X_i^2}{\theta} \le c_2\right\} = \left[\frac{2\sum_{i=1}^{n} X_i^2}{c_2}, \frac{2\sum_{i=1}^{n} X_i^2}{c_1}\right].$$

9. For $0 \le r \le \theta$, $P_\theta(R_n \le r) = \prod_{i=1}^n P(X_i^2 + Y_i^2 \le r^2) = (r/\theta)^{2n}$.

10. The variable R_n has a probability density function given by $f(r) = 2n\theta^{-2n} r^{2n-1}$ supported on $0 \le r \le \theta$. The mean and second moment of R_n are then

$$\mathrm{E}_\theta R_n = 2n\theta^{-2n} \int_0^\theta r^{2n} dr = \frac{2n\theta}{2n+1}, \quad \mathrm{E}_\theta R_n^2 = 2n\theta^{-2n} \int_0^\theta r^{2n+1} dr = \frac{2n\theta^2}{2n+2}.$$

So the bias of R_n converges to zero, $\mathrm{E}_\theta R_n - \theta = -\theta/(2n+1) \to 0$, as $n \to \infty$. Similarly, $\mathrm{E}_\theta R_n^2 \to \theta^2$ as $n \to \infty$ so that the variance of R_n converges to zero:

$$\mathrm{E}_\theta R_n^2 - [\mathrm{E}_\theta R_n]^2 \to \theta^2 - \theta^2 = 0.$$

Therefore, R_n is mean squared error consistent.

11. For any $\theta > 0$, $P_\theta(\theta \in I_n) = P_\theta(\theta \le R_n 5^{1/(2n+\sqrt{n})}) = P_\theta(\theta 5^{-1/(2n+\sqrt{n})} \le R_n) = 1 - P_\theta(R_n < \theta 5^{-1/(2n+\sqrt{n})}) = 1 - 5^{-2n/(2n+\sqrt{n})} \to 1 - 5^{-1} = 4/5$ as $n \to \infty$.

12. Assuming $R_n > 0$, the likelihood function is $L(\theta) = \pi^{-n}\theta^{-2n}\mathbb{I}(0 < R_n \le \theta)$ so that the posterior density $f(\theta)$ of θ is proportional to $L(\theta)\pi(\theta)$ or $f(\theta) = C\theta^{-n-1}\mathbb{I}(R_n \le \theta)$ for a normalizing constant C; it turns out that $C = nR_n^n$, but this is not needed here. The posterior risk function is

$$\begin{aligned}\int_{R_n}^\infty L(t,\theta) f(\theta) d\theta &= t^2 C \int_{R_n}^\infty \theta^{-n-2} d\theta - 2tC \int_{R_n}^\infty \theta^{-n-1} d\theta + C \int_{R_n}^\infty \theta^{-n} d\theta \\ &= t^2 C \frac{R_n^{-n-1}}{n+1} - 2tC \frac{R_n^{-n}}{n} + \int_{R_n}^\infty C\theta^{-n} d\theta\end{aligned}$$

which, taking derivatives in t, is minimized at $t = R_n(n+1)/n$. The Bayes estimator is $R_n(n+1)/n$.

13. The simple Bayes test is to reject H_0 if the posterior probability over the H_1-parameter space exceeds $1/2$; that is, reject if

$$\int_0^1 f(\theta) d\theta = nR_n \int_{R_n}^1 \theta^{-n-1} d\theta = nR_n^n(-n^{-1} + R_n^{-n}/n) = 1 - R_n^n > 1/2$$

which is equivalent to rejecting if $2^{-1/n} > R_n$.

14. The power function of the test is $P_\theta(T=1) = P_\theta(R_n < 2^{-1/n})$ so that

$$P_\theta(T=1) = \begin{cases} 1 & \text{if } \theta \le 2^{-1/n} \\ (2^{-1/n}/\theta)^{2n} = \theta^{-2n}4^{-1} & \text{if } \theta > 2^{-1/n} \end{cases}$$

The size of the test is $\sup_{\theta \ge 1} P_\theta(T=1) = 4^{-1}$.