Stat 643 Problems

1. Let $(\Omega, \mathcal{A})$ be a measurable space. Suppose that $\mu$, $P$ and $Q$ are $\sigma$-finite positive measures on this space with $P \ll Q$ and $Q \ll \mu$.
   a) Show that $P \ll \mu$ and $\frac{dP}{d\mu} = \frac{dP}{dQ} \cdot \frac{dQ}{d\mu}$ a.s. $\mu$.
   b) Show that if $\mu \ll Q$, then $\frac{d\mu}{dQ} = \left(\frac{dQ}{d\mu}\right)^{-1}$ a.s. $\mu$.

2. (Cressie) Let $\Omega = [-\lambda, \lambda]$ for some $\lambda > 0$, $\mathcal{A}$ be the set of Borel sets and $P$ be Lebesgue measure divided by $2\lambda$. For a subset of $\Omega$, define the symmetric set for $A$ as $-A = \{ -\omega | \omega \in A \}$ and let $\mathcal{C} = \{ A \in \mathcal{A} | A = -A \}$.
   a) Show that $\mathcal{C}$ is a sub $\sigma$-algebra of $\mathcal{A}$.
   b) Let $X$ be an integrable random variable. Find $E(X|\mathcal{C})$.

3. Let $\Omega = [0, 1]^2$, $\mathcal{A}$ be the set of Borel sets and $P = \frac{1}{2}\Delta + \frac{1}{2}\mu$ for $\Delta$ a distribution placing a unit point mass at the point $(\frac{1}{2}, 1)$ and $\mu$ 2-dimensional Lebesgue measure. Consider the variable $X(\omega) = \omega_1$ and the sub $\sigma$-algebra of $\mathcal{A}$ generated by $X, C$.
   a) For $A \in \mathcal{A}$, find $E(I_A|\mathcal{C}) = P(A|\mathcal{C})$.
   b) For $Y(\omega) = \omega_2$, find $E(Y|\mathcal{C})$.


5. Prove the fact stated in the paragraph above Result 5 in the typed notes.

6. Show directly that the following families of distributions have countable equivalent subsets:
   a) $\mathcal{P}$, for $P_\theta$ uniform on the integers $\theta - 1$ through $\theta + 1$,
      $\Theta = \{\ldots, -2, -1, 0, 1, 2, 3, \ldots\}$.
   b) $\mathcal{P}$, for $P_\theta$ uniform on the 2-dimensional disk of radius $\theta$ centered at the origin,
      $\Theta = (0, \infty)$.
   c) The Poisson $(\theta)$ family, $\Theta = (0, \infty)$.

7. Suppose that $X = (X_1, X_2, \ldots, X_n)$ has independent components, where each $X_i$ is generated as follows. For independent random variables $W_i \sim \text{normal}(\mu, 1)$ and $Z_i \sim \text{Poisson}(\mu)$, $X_i = W_i$ with probability $p$ and $X_i = Z_i$ with probability $1 - p$.
   Suppose that $\mu \in [0, \infty)$. Use the factorization theorem and find low dimensional sufficient statistics in the cases that:
   a) $p$ is known to be $\frac{1}{2}$, and
   b) $p \in [0, 1]$ is unknown.

   (In the first case the parameter space is $\Theta = \{\frac{1}{2}\} \times [0, \infty)$, while in the second it is $[0, 1] \times [0, \infty)$.)

8. (Ferguson) Consider the probability distributions on $(\mathcal{R}^1, \mathcal{B}_1)$ defined as follows. For $\theta = (\theta, p) \in \Theta = \mathcal{R} \times (0, 1)$ and $X \sim P_\theta$, suppose
\[ P_\theta[X = x] = \begin{cases} (1 - p)p^{x-\theta} & \text{for } x = \theta, \theta + 1, \theta + 2, \ldots \\ 0 & \text{otherwise} \end{cases} \]

Let \( X_1, X_2, \ldots, X_n \) be iid according to \( P_\theta \).

a) Argue that the family \( \{P_\theta^\alpha\}_{\alpha \in \Theta} \) is not dominated by a \( \sigma \)-finite measure, so that the factorization theorem cannot be applied to identify a sufficient statistic here.

b) Argue from first principles that the statistic \( T(X) = (\min X_i, \sum X_i) \) is sufficient for the parameter \( \theta \).

c) Argue that the factorization theorem can be applied if \( \theta \) is known (the first factor of the parameter space \( \Theta \) is replaced by a single point) and identify a sufficient statistic for this case.

d) Argue that if \( \rho \) is known, \( \min X_i \) is a sufficient statistic.

9. Suppose that \( X' \) is exponential with mean \( \lambda^{-1} \) (i.e. has density \( f_\lambda(x) = \lambda \exp(-\lambda x)I[x \geq 0] \)) with respect to Lebesgue measure on \( \mathbb{R}^1 \), but that one only observes \( X = X'I[X' > 1] \). (There is interval censoring below \( x = 1 \).)

a) Consider the measure \( \mu \) on \( \mathcal{X} = \{0\} \cup (1,\infty) \) consisting of a point mass of \( 1 \) at \( 0 \) plus Lebesgue measure on \( (1,\infty) \). Give a formula for the R-N derivative of \( P_\lambda^X \) wrt \( \mu \) on \( \mathcal{X} \).

b) Suppose that \( X_1, X_2, \ldots, X_n \) are iid with the marginal distribution \( P_\lambda^X \). Find a 2-dimensional sufficient statistic for this problem and argue that it is indeed sufficient.

c) Argue carefully that your statistic from b) is minimal sufficient.

10. Suppose that \( X_1, X_2, \ldots, X_n \) are iid with marginal density with respect to Lebesgue measure on \( (0,2) \) of the form

\[ f_\theta(x) = \begin{cases} \theta x & x \in (0,1) \\ \frac{1}{2}(2 - \theta) & x \in [1,2] \end{cases} \]

Find a low dimensional sufficient statistic and prove that it is minimal.

11. (Lehmann TPE page 65) Let \( X_1, X_2, \ldots, X_m \) and \( Y_1, Y_2, \ldots, Y_n \) be independent normal random variables with \( \text{EX} = \xi \), \( \text{VarX} = \sigma^2 \), \( \text{EY} = \eta \) and \( \text{VarY} = \tau^2 \). Find minimal sufficient statistics for the following situations:

a) no restrictions on the parameters,

b) \( \sigma = \tau \), and

c) \( \xi = \eta \).

12. (Lehmann TPE page 65) Let \( f \) be a positive integrable function over \( (0,\infty) \) and let \( p_\theta(x) = c(\theta) f(x)I[0 < x < \theta] \). If \( X_1, X_2, \ldots, X_n \) are iid with the marginal density \( p_\theta(x) \), show that \( \max X_i \) is sufficient for \( \theta \).

13. Schervish Exercise 2.8. And show that \( Z \) is ancillary.
14. Schervish Exercise 2.13. (I think you need to assume that \( \{x_i\} \) spans \( R^k \).)

15. Schervish Exercise 2.16.


17. Schervish Exercise 2.22. What is a first order ancillary statistic here?

18. Schervish Exercise 2.27.

19. Suppose that \( S \) is a convex subset of \([0, \infty)^k\). Argue that there exists a finite \( \Theta \) decision problem that has \( S \) as its set of risk vectors. (Consider problems where \( X \) is degenerate and carries no information.)

20. Consider the two state decision problem with \( \Theta = \{1, 2\} \), \( P_1 \) the Bernoulli \((\frac{1}{2})\) distribution and \( P_2 \) the Bernoulli \((\frac{1}{3})\) distribution, \( A = \Theta \) and \( L(\theta, a) = I[\theta \neq a] \).

   a) Find the set of risk vectors for the four nonrandomized decision rules. Plot these in the plane. Sketch \( S \), the risk set for this problem.

   b) For this problem, show explicitly that any element of \( D^* \) has a corresponding element of \( D_s \) with identical risk vector and vice versa.

   c) Identify the the set of all admissible risk vectors for this problem. Is there a minimal complete class for this decision problem? If there is one, what is it?

   d) For each \( p \in [0, 1] \), identify those risk vectors that are Bayes versus the prior \( g = (p, 1-p) \). For which priors are there more than one Bayes rule?

   e) Verify directly that the prescription "choose an action that minimizes the posterior expected loss" produces a Bayes rule versus the prior \( g = (\frac{1}{2}, \frac{1}{2}) \).

21. Consider a two state decision problem with \( \Theta = \{1, 2\} \), where the observable \( X = (X_1, X_2) \) has iid Bernoulli \((\frac{1}{3})\) coordinates if \( \theta = 1 \) and iid Bernoulli \((\frac{1}{2})\) coordinates if \( \theta = 2 \). Suppose that \( A = \Theta \) and \( L(\theta, a) = I[\theta \neq a] \). Consider the behavioral decision rule \( \phi_x \) defined by

   \[
   \begin{align*}
   \phi_x(\{1\}) &= 1 & \text{if } x_1 = 0, \\
   \phi_x(\{1\}) &= 0 & \text{if } x_1 = 1.
   \end{align*}
   \]

   a) Show that \( \phi_x \) is inadmissible by finding a rule with a better risk function. (It may be helpful to figure out what the risk set is for this problem, in a manner similar to what you did in problem 20.)

   b) Use the construction from Result 21 and find a behavioral decision rule that is a function of the sufficient statistic \( X_1 + X_2 \) and is risk equivalent to \( \phi_x \). (Note that this rule is inadmissible.)

22. Consider the squared error loss estimation of \( p \in (0, 1) \), based on \( X \sim \text{binomial} (n, p) \), and the two nonrandomized decision rules \( \delta_1(x) = \frac{x}{n} \) and \( \delta_2(x) = \frac{1}{2} \left( \frac{x}{n} + \frac{1}{2} \right) \). Let
ψ be a randomized decision function that chooses δ₁ with probability 1/2 and δ₂ with probability 1/2.

a) Write out expressions for the risk functions of δ₁, δ₂ and ψ.
b) Find a behavioral rule that is risk equivalent to ψ.
c) Identify a nonrandomized estimator that is strictly better than ψ or φ.

23. Suppose that Θ = Θ₁ × Θ₂ and that a decision rule φ is such that for each θ₂, φ is admissible when the parameter space is Θ₁ × {θ₂}. Show that φ is then admissible when the parameter space is Θ.

24. Suppose that w(θ) > 0 ∀θ. Show that φ is admissible with loss function L(θ, a) iff it is admissible with loss function w(θ)L(θ, a).

25. Let P₀ and P₁ be two distributions on X and f₀ and f₁ be densities of these with respect to some dominating σ-finite measure μ. Consider the parametric family of distributions with parameter p ∈ [0, 1] and densities wrt μ of the form

$$f_p(x) = (1 - p)f_0(x) + pf_1(x)$$

and suppose that X has density f_p.

a) If G({0}) = G({1}) = 1/2, find the squared error loss formal Bayes estimator of p versus the prior G.
b) Suppose now that G is the uniform distribution on [0, 1]. Find the squared error loss Bayes estimator of p versus the prior G.
c) Argue that your estimator from b) is admissible.

26. For the model of problem 25, suppose that P₀ is the binomial (n, 1/4) distribution and P₁ is the binomial (n, 3/4) distribution. Find an ancillary statistic. (Hint: consider the probabilities that X takes the values 0 and n.) Is the family of problem 25 then necessarily boundedly complete?

27. (Ferguson) Prove or give a counterexample: If C₁ and C₂ are complete classes of decision rules, then C₁ ∩ C₂ is essentially complete.


29. (Ferguson and others) Suppose that X ~ binomial (n, p) and one wishes to estimate p ∈ (0, 1). Suppose first that L(p, a) = p⁻¹(1 - p)⁻¹(p - a)².

a) Show that $\frac{X}{n}$ is Bayes versus the uniform prior on (0, 1).
b) Argue that $\frac{X}{n}$ is admissible in this decision problem.

Now consider ordinary squared error loss, L(p, a) = (p - a)².

c) Apply the result of problem 24 and prove that $\frac{X}{n}$ is admissible under this loss function as well.
30. Consider a two state decision problem where $\Theta = \mathcal{A} = \{0, 1\}$, $P_0$ and $P_1$ have respective densities with respect to a dominating $\sigma$-finite measure $\mu$, $f_0$ and $f_1$ and the loss function is $L(\theta, a)$.
   a) For $G$ an arbitrary prior distribution, find a formal Bayes rule versus $G$.
   b) Specialize your result from a) to the case where $L(\theta, a) = I[\theta \neq a]$. What connection does the form of these rules have to the theory of simple versus simple hypothesis testing?

31. (Ferguson and others) Suppose that one observes independent binomial random variables, $X \sim \text{binomial}(n, p_1)$ and $Y \sim \text{binomial}(n, p_2)$ and needs to estimate $p_1 - p_2$ with squared error loss, $L(p, a) = (p_1 - p_2 - a)^2$. (We will assume that the parameter space is $(0, 1)^2$ and that the action space is $(-1, 1)$.) Find the form of a rule Bayes against a prior uniform over the parameter space.

32. Consider a "linear loss two action" decision problem, where $\Theta = \mathbb{R}^1$, $\mathcal{A} = \{0, 1\}$ and $L(\theta, a) = (\theta - \theta_0)I[a = 0]I[\theta \geq \theta_0] + (\theta_0 - \theta)I[\theta = 1]I[\theta < \theta_0]$ for some $\theta_0$ of interest. Find the form of a Bayes rule for this problem. (Hint: Consider the difference $L(\theta, 0) - L(\theta, 1)$.)

33. Problem 24, page 211 Schervish.

34. Problem 34, page 142 Schervish.

35. Problem 36, page 142 Schervish.


41. Consider the family of bivariate observations $X = (Y, Z)$ on $\mathcal{X} = (0, 1)^2$ with densities wrt two-dimensional Lebesgue measure

$$f_\eta(x) \propto \exp(\eta_1 y + \eta_2 z)$$

for a bivariate parameter vector $\eta = (\eta_1, \eta_2)$.
   a) Find the natural parameter space for this family of distributions.
   b) What is the normalizing constant $K(\eta)$ for this family? Use it to find $E_\eta Y$.
   c) Find a UMVUE of the quantity $E_\eta Y_1 - E_\eta Z_1$ based on iid vectors $X_1, X_2, \ldots, X_n$ with marginal (bivariate) distribution specified by$f_\eta(x)$. Argue carefully that your estimator is UMVUE.
42. Suppose that $X \sim \text{Bernoulli} (p)$ and that one wishes to estimate $p$ with loss $L(p, a) = |p - a|$. Consider the estimator $\delta$ with $\delta(0) = \frac{1}{4}$ and $\delta(1) = \frac{3}{4}$.

a) Write out the risk function for $\delta$ and show that $R(p, \delta) \leq \frac{1}{4}$.

b) Show that there is a prior distribution placing all its mass on $\{0, \frac{1}{2}, 1\}$ against which $\delta$ is Bayes.

c) Prove that $\delta$ is minimax in this problem and identify a least favorable prior.

43. Consider a decision problem where $P_\theta$ is the Normal $(\theta, 1)$ distribution on $\mathcal{X} = \mathbb{R}^1$, $\mathcal{A} = \{0, 1\}$ and $L(\theta, a) = I[\alpha = 0]I[\theta > 5] + I[\alpha = 1]I[\theta \leq 5]$.

a) If $\Theta = (- \infty, 5] \cup [6, \infty)$ guess what prior distribution is least favorable, find the corresponding Bayes decision rule and prove that it is minimax.

b) If $\Theta = \mathbb{R}^1$, guess what decision rule might be minimax, find its risk function and prove that it is minimax.

44. Consider the scenario of problem 29. Show that the estimator there is minimax for the weighted squared error loss and identify a least favorable prior.

45. (Ferguson and others) Consider (inverse mean) weighted squared error loss estimation of $\lambda$, the mean of a Poisson distribution. That is, let $\Lambda = (0, \infty)$, $\mathcal{A} = \Lambda$, $P_\lambda$ be the Poisson distribution on $\mathcal{X} = \{0, 1, 2, 3, \ldots\}$ and $L(\lambda, a) = \lambda^{-1}(\lambda - a)^2$. Let $\delta(X) = X$.

a) Show that $\delta$ is an equalizer rule.

b) Show that $\delta$ is generalized Bayes versus Lebesgue measure on $\Lambda$.

c) Find the Bayes estimators wrt the $\Gamma(\alpha, \beta)$ priors on $\Lambda$.

d) Prove that $\delta$ is minimax for this problem.

46. (Berger page 385) Suppose that $X$ is Poisson with mean $\lambda$ and that one wishes to test $H_0 : \lambda = 1$ vs $H_\alpha : \lambda = 2$ with "$0 - 1$ loss," i.e. $\Lambda = \mathcal{A} = \{1, 2\}$ and $L(\lambda, a) = I[\lambda \neq a]$.

a) Find the form of the nonrandomized Bayes tests for this problem.

b) Using your answer to part a), find enough risk points for nonrandomized Bayes rules to sketch the part of the lower boundary of the risk set relevant to finding a minimax rule.

c) On the basis of your sketch from part b), find a minimax rule for this problem and identify a least favorable distribution.

47. Consider the left-truncated Poisson distribution with pmf $f_\theta(x) = \exp(-\theta)\theta^x/x!(1-\exp(-\theta))$ for $x = 1, 2, 3, \ldots$ Suppose that $X_1, X_2, \ldots, X_n$ are iid variables with this distribution.

a) Show that $T = \sum_{i=1}^n X_i$ is a complete sufficient statistic for $\theta \in (0, \infty)$.

b) It turns out that the pmf of $T$ is given by $g_\theta(t) = S_{\theta t}^n n!\theta^t/t!(\exp(\theta) - 1)^n$ for $t = n, n+1, n+2, \ldots$ where $S_{\theta}^n$ is the so called "Stirling number of the second
kind" (whose exact definition is not important here). Find the UMVUE of \( \theta \). (It
will involve the \( S^0_i \)'s.)

48. Consider an observable \( X \) taking values in \( X \) and with distribution \( P_\theta \) for \( \theta \in \Theta \). For a statistic \( T \), let \( B_0 = B(T) \) and \( P^0_\theta \) be the restriction of \( P_\theta \) to \( B_0 \). Prove that if \( T \) is complete and sufficient and \( \phi(T) \) is unbiased for \( \gamma(\theta) \), then another random variable \( U \) is an unbiased estimator of \( \gamma(\theta) \) iff \( E(U|T) = \phi(T) \) a.s. \( P^0_\theta \forall \theta \).

49. Suppose that under the usual set-up \( T \) and \( S \) are two real-valued statistics.
   a) Suppose that \( E_\theta(T(X))^2 < \infty \) and \( E_\theta(S(X))^2 < \infty \forall \theta \). Show that if \( T(X) \)
and \( S(X) \) are both unbiased estimators of \( \theta \) and \( T(X) \) and \( S(X) - T(X) \) are
uncorrelated \( \forall \theta \), then \( \text{Var}_\theta T(X) \leq \text{Var}_\theta S(X) \).
   b) Suppose that \( S(X) \) is an unbiased estimator of \( \theta \), \( E_\theta(S(X))^2 < \infty \forall \theta \) and that
\( R = E_\theta[S|T] \) doesn’t depend upon \( \theta \). Argue that \( R \) is preferable to \( S \) under
squared error loss.

50. Suppose that \( X \) is a real-valued random variable with density (wrt the \( \sigma \)-finite
measure \( \mu \)) defined by \( f_\eta(x) = \exp(a(\eta) + \eta x)h(x) \) for a positive function \( h(\cdot) \)
\( \forall x \in \mathbb{R}^1 \) and \( \eta \in (c, d) = \Gamma \subset \mathbb{R}^1 \), the natural parameter space.
   a) Find \( \gamma(\eta) = E_\eta X \).
   b) For a fixed \( \eta \in \Gamma \), let \( M_\eta(s) = E_\eta \exp(sX) \). Find \( M_\eta(s) \), determine for which
values of \( s \) it is finite, and show that it depends on \( \eta \) only through \( \gamma(\eta) \).


52. Problem 19, page 341 Schervish. Also, find the MRE estimator of \( \theta \) under squared
error loss for this model.

53. Show that the C-R inequality is not changed by a smooth reparameterization. That is,
suppose that \( \mathcal{P} = \{P_\theta\} \) is dominated by a \( \sigma \)-finite measure \( \mu \) and satisfies
i) \( f_\theta(x) > 0 \) for all \( \theta \) and \( x \),
ii) for all \( x \), \( \frac{d}{d\theta} f_\theta(x) \) exists and is finite everywhere on \( \Theta \subset \mathbb{R}^1 \) and
iii) for any statistic \( \delta \) with \( E_\theta(\delta(X)|< \infty \) for all \( \theta \), \( E_\theta \delta(X) \) can be differentiated
under the integral sign at all points of \( \Theta \).

Let \( h \) be a function from \( \Theta \) to \( \mathbb{R}^1 \) such that \( h' \) is continuous and nonvanishing on \( \Theta \). Let
\( \eta = h(\theta) \) and define \( Q_\eta = P_\theta \). Show that the information inequality bound obtained
from \( \{Q_\eta\} \) evaluated at \( \eta = h(\theta) \) is the same as the bound obtained from \( \mathcal{P} \).

54. Let \( N = (N_1, N_2, \ldots, N_k) \) be multinomial \( (n, p_1, p_2, \ldots, p_k) \) where \( \sum p_i = 1 \).
   a) Find the form of the Fisher information matrix based on the parameter \( p \).
   b) Suppose that \( p_i(\theta) \) for \( i = 1, 2, \ldots, k \) are differentiable functions of a real
parameter \( \theta \in \Theta \), and open interval, where each \( p_i(\theta) \geq 0 \) and \( \sum p_i(\theta) = 1 \).

Suppose that \( h(y_1, y_2, \ldots, y_k) \) is a continuous real-valued function with continuous
first partial derivatives and define $q(\theta) = h(p_1(\theta), p_2(\theta), ..., p_k(\theta))$. Show that the information bound for unbiased estimators of $q(\theta)$ in this context is 

$$
\frac{(q'(\theta))^2}{nI_1(\theta)} \quad \text{where} \quad I_1(\theta) = \sum_{i=1}^{k} p_i(\theta) \left( \frac{d}{d\theta} \log p_i(\theta) \right)^2.
$$

55. As an application of Lemma 61 (or 61') consider the following. Suppose that $X \sim U(\theta_1, \theta_2)$ and consider unbiased estimators of $\gamma(\theta_1, \theta_2) = \frac{\theta_1 + \theta_2}{2}$. For $\theta_1 < \theta_1' < \theta_2$ and $\theta_1 < \theta_2' < \theta_2$, apply Lemma 61 with $g_1(x) = 1 - \frac{f_{\theta_1, \theta_2}(x)}{f_{\theta_1', \theta_2}(x)}$ and $g_2(x) = 1 - \frac{f_{\theta_1, \theta_2}(x)}{f_{\theta_1, \theta_2'}(x)}$. What does Lemma 61 then say about the variance of an unbiased estimator of $\gamma(\theta_1, \theta_2)$? (If you can do it, sup over values of $\theta_1'$ and $\theta_2'$. I've not tried this, so am not sure how it comes out.)

56. Problem 11, page 389 Schervish.

57. Take the simplest possible $n$-dimensional statistical estimation problem, $X_1, X_2, ..., X_n$ iid Bernoulli $(p)$, and make your way through the results on maximum likelihood, carefully seeing and saying what they assume and what they yield. (Find $z_{\theta_0}(p)$, $l_n(x^n, p)$, verify hypotheses of the theorems and see what they say about estimation of $p$. For sake of argument, let $T_n = \frac{1}{n-5} \sum_{i=1}^{n-5} X_i$ and have a look at the corresponding $\hat{p}_n$. What do the theorems say about the behavior of $X$ when $p = 0$ or 1? What are approximate confidence intervals for $p$? Etc.)

58. Suppose that $X_1, X_2, ...$ are iid, each taking values in $\mathcal{X} = \{0, 1, 2\}$ with R-N derivative of $P_\theta$ wrt to counting measure on $\mathcal{X}$

$$
f_{\theta}(x) = \frac{\exp(x\theta)}{1 + \exp(2\theta)}.
$$

a) Find an estimator of $\theta$ based on $n_0 = \sum_{i=1}^{n} I[X_i = 0]$ that is $\sqrt{n}$ consistent (i.e. for which $\sqrt{n}(\hat{\theta} - \theta)$ converges in distribution).

b) Find in more or less explicit form a "one step Newton modification" of your estimator from a).

c) Prove directly that your estimator from b) is asymptotically normal with variance $1/I_1(\theta)$. (With $\hat{\theta}_n$ the estimator from a) and $\tilde{\theta}_n$ the estimator from b),

$$
\tilde{\theta}_n = \hat{\theta}_n - \frac{L_n'(\hat{\theta}_n)}{L_n''(\hat{\theta}_n)},
$$

and write $L_n'(\hat{\theta}_n) = L_n'(\theta) + (\hat{\theta}_n - \theta)L_n''(\theta) + \frac{1}{2}(\hat{\theta}_n - \theta)^2L_n''(\theta^*)$ for some $\theta^*$ between $\hat{\theta}_n$ and $\theta$.)

d) Show that provided $\bar{x} \in (0, 2)$ the loglikelihood has a maximizer
\[
\hat{\theta}_n = \log\left( \frac{\bar{x} - 1 + \sqrt{-3\bar{x}^2 + 6\bar{x} + 1}}{2(2 - \bar{x})} \right).
\]

Prove that an estimator defined to be \(\hat{\theta}_n\) when \(\bar{x} \in (0, 2)\) will be asymptotically normal with variance \(1/I_1(\theta)\).

e) Show that the "observed information" and "expected information" approximations lead to the same large sample confidence intervals for \(\theta\). What do these look like based on, say, \(\hat{\theta}_n\)?

By the way, a version of nearly everything in this problem works in any one parameter exponential family. See the "Theory II" question on last year's prelim.

59. Prove the following "filling-in" lemma:

**Lemma** Suppose that \(g_0\) and \(g_1\) are two distinct, positive probability densities defined on an interval in \(\mathcal{R}\). If the ratio \(g_1/g_0\) is nondecreasing in a real-valued function \(T(x)\), then the family of densities \(\{g_\alpha | \alpha \in [0, 1]\}\) for \(g_\alpha = \alpha g_1 + (1 - \alpha)g_2\) has the MLR property in \(T(x)\).

60. Consider the situation of problem 9 and maximum likelihood estimation of \(\lambda\).

a) Show that with \(M = \# [X_i's equal to 0]\), in the event that \(M = n\) there is no MLE of \(\lambda\), but that in all other cases there is a maximizer of the likelihood. Then argue that for any \(\lambda > 0\), with \(P_\lambda\) probability tending to 1, the MLE of \(\lambda\), say \(\hat{\lambda}_n\), exists.

b) Give a simple estimator of \(\lambda\) based on \(M\) alone. Prove that this estimator is consistent for \(\lambda\). Then write down an explicit one-step Newton modification of your estimator from a).

c) Discuss what numerical methods you could use to find the MLE from a) in the event that it exists.

d) Give two forms of large sample confidence intervals for \(\lambda\) based on the MLE \(\hat{\lambda}_n\) and two different approximations to \(I_1(\lambda)\).

61. Consider the two distributions \(P_0\) and \(P_1\) on \(X = (0, 1)\) with densities wrt Lebesgue measure

\[
f_0(x) = 1 \text{ and } f_1(x) = 3x^2.
\]

a) Find a most powerful level \(\alpha = .2\) test of \(H_0 : X \sim P_0\) vs \(H_1 : X \sim P_1\).

b) Plot both the 0-1 loss risk set \(S\) and the set \(V = \{\beta_0(\theta_0), \beta(\theta_1)\}\) for the simple versus testing problem involving \(f_0\) and \(f_1\). What test is Bayes versus a uniform prior? This test is best of what size?

c) Take the result of Problem 59 as given. Consider the family of mixture distributions \(\mathcal{P} = \{P_\theta\}\) where for \(\theta \in [0, 1]\), \(P_\theta = (1 - \theta)P_0 + \theta P_1\). In this family find a UMP level \(\alpha = .2\) test of the hypothesis \(H_0 : \theta \leq .5\) vs \(H_1 : \theta > .5\) based on a single observation \(X\). Argue carefully that your test is really UMP.
62. Consider a composite versus composite testing problem in a family of distributions $\mathcal{P} = \{P_\theta\}$ dominated by the $\sigma$-finite measure $\mu$, and specifically a Bayesian decision-theoretic approach to this problem with prior distribution $G$ under 0-1 loss.

a) Show that if $G(\Theta_0) = 0$ then $\phi(x) \equiv 1$ is Bayes, while if $G(\Theta_1) = 0$ then $\phi(x) \equiv 0$ is Bayes.

b) Show that if $G(\Theta_0) > 0$ and $G(\Theta_1) > 0$ then the Bayes test against $G$ has Neyman-Pearson form for densities $g_0$ and $g_1$ on $\mathcal{X}$ defined by

$$g_0(x) = \frac{\int_{\Theta_0} f_\theta(x) dG(\theta)}{G(\Theta_0)} \quad \text{and} \quad g_1(x) = \frac{\int_{\Theta_1} f_\theta(x) dG(\theta)}{G(\Theta_1)}.$$ 

c) Suppose that $X$ is Normal $(\mu, \sigma^2)$. Find Bayes tests of $H_0$ vs $H_1$.

i) supposing that $G$ is the Normal $(0, \sigma^2)$ distribution, and

ii) supposing that $G$ is $\frac{1}{2}(N + \Delta)$, for $N$ the Normal $(0, \sigma^2)$ distribution, and $\Delta$ a point mass distribution at 0.

63. Suppose that $X$ is Poisson with mean $\theta$. Find the UMP test of size $\alpha = .05$ of $H_0 : \theta \leq 4$ or $\theta \geq 10$ vs $H_1 : 4 < \theta < 10$. The following table of Poisson probabilities $P_\theta[X = x]$ will be useful in this enterprise.

<table>
<thead>
<tr>
<th>$x$</th>
<th>$\theta = 4$</th>
<th>$\theta = 10$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>.0183</td>
<td>.0000</td>
</tr>
<tr>
<td>1</td>
<td>.0733</td>
<td>.0005</td>
</tr>
<tr>
<td>2</td>
<td>.1465</td>
<td>.0023</td>
</tr>
<tr>
<td>3</td>
<td>.1954</td>
<td>.0076</td>
</tr>
<tr>
<td>4</td>
<td>.1954</td>
<td>.0189</td>
</tr>
<tr>
<td>5</td>
<td>.1563</td>
<td>.0378</td>
</tr>
<tr>
<td>6</td>
<td>.1042</td>
<td>.0631</td>
</tr>
<tr>
<td>7</td>
<td>.0595</td>
<td>.0901</td>
</tr>
<tr>
<td>8</td>
<td>.0298</td>
<td>.1126</td>
</tr>
<tr>
<td>9</td>
<td>.0132</td>
<td>.1251</td>
</tr>
<tr>
<td>10</td>
<td>.0053</td>
<td>.1251</td>
</tr>
<tr>
<td>11</td>
<td>.0019</td>
<td>.1137</td>
</tr>
<tr>
<td>12</td>
<td>.0006</td>
<td>.0948</td>
</tr>
</tbody>
</table>

64. Suppose that $X$ is exponential with mean $\lambda^{-1}$. Set up the two equations that will have to be solved simultaneously in order to find UMP size $\alpha$ test of $H_0 : \lambda \leq .5$ or $\lambda \geq 2$ vs $H_1 : \lambda \in (.5, 2)$.

65. Suppose that for $i = 1, 2, ...$ the random vectors $\mathbf{X}_i = (Y_i, Z_i)$ are iid with distribution $P_\theta$ described as follows. $Y_i$ and $Z_i$ are independent, $Y_i \sim \text{binomial}(n, p)$ and $Z_i \sim \text{geometric}(p)$ (i.e. with pmf $g_p(z) = p(1 - p)^{z-1}$ for $z = 1, 2, ...$).

a) Give the C-R lower bound on the variance of an unbiased estimator of $p$ based on $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_n$.

b) Find the maximum likelihood estimator of $p$ based on $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_n$ and discuss its asymptotic properties.

c) Give both the "expected information" and "observed information" large sample confidence intervals for $p$ based on your estimator from b).

66. Consider again the scenario of problem 20. There you sketched the risk set $\mathcal{S}$. Here sketch the corresponding set $\mathcal{V}$. 

-10-
67. Consider the estimation problem with \( n \) iid \( P_\theta \) observations, where \( P_\theta \) is the exponential distribution with mean \( \theta \). Let \( G \) be the group of scale transformations on \( X = (0, \infty)^n \), \( G = \{ g_\epsilon | \epsilon > 0 \} \) where \( g_\epsilon(x) = \epsilon x \).

a) Show that the estimation problem with loss \( L(\theta, \alpha) = (\log(\alpha/\theta))^2 \) is invariant under \( G \) and say what relationship any equivariant nonrandomized decision rule must satisfy.

b) Show that the estimation problem with loss \( L(\theta, \alpha) = (\theta - \alpha)^2/\theta^2 \) is invariant under \( G \), and say what relationship any equivariant nonrandomized estimator must satisfy.

c) Find the generalized Bayes estimator of \( \theta \) in situation b) if the "prior" has density wrt Lebesgue measure \( g(\theta) = \theta^{-1}I[\theta > 0] \). Argue that this estimator is the best equivariant estimator in the situation of b).

68. (Problem 37, page 122 TSH.) Let \( X_1, ..., X_n \) and \( Y_1, ..., Y_m \) be independent samples from \( N(\xi, 1) \) and \( N(\eta, 1) \), and consider the hypotheses \( H_0 : \eta \leq \xi \) and \( H_1 : \eta > \xi \). Show that there exists a UMP test of size \( \alpha \), and it rejects \( H_0 \) when \( \bar{Y} - \bar{X} \) is too large. (If \( \xi_1 < \eta_1 \) is a particular alternative, the distribution assigning probability 1 to the point with \( \eta = \xi = (m\xi_1 + n\eta_1)/(m + n) \) is least favorable at size \( \alpha \).)

69. Problem 57 Schervish, page 293.

70. Problem 12, parts a)-c) Schervish, page 534.

71. Fix \( \alpha \in (0, 5) \) and \( c \in (\frac{\alpha}{2-2\alpha}, \alpha) \). Let \( \Theta = \{ -1 \} \cup [0, 1] \) and consider the discrete distributions with probability mass functions as below.

<table>
<thead>
<tr>
<th>( x )</th>
<th>(-2)</th>
<th>(-1)</th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \theta = -1 )</td>
<td>( \frac{\alpha}{2} )</td>
<td>( \frac{1}{2} - \alpha )</td>
<td>( \alpha )</td>
<td>( \frac{1}{2} - \alpha )</td>
<td>( \frac{\alpha}{2} )</td>
</tr>
<tr>
<td>( \theta \neq -1 )</td>
<td>( \theta c )</td>
<td>( \frac{\alpha - \frac{\alpha}{2}}{1 - \alpha} )</td>
<td>( \frac{1 - \alpha}{1 - \alpha} )</td>
<td>( \frac{\alpha}{1 - \alpha} )</td>
<td>( \frac{1 - \alpha}{2 - \alpha} )</td>
</tr>
</tbody>
</table>

Find the size \( \alpha \) likelihood ratio test of \( H_0 : \theta = -1 \) vs \( H_1 : \theta \neq -1 \). Show that the test \( \phi(x) = I[x = 0] \) is of size \( \alpha \) and is strictly more powerful that the LRT whatever be \( \theta \).

(This simple example shows that LRTs need not necessarily be in any sense optimal.)

72. Consider the situation of problem 58.

a) Find the form of UMP tests of \( H_0 : \theta \leq \theta_0 \) vs \( H_1 : \theta > \theta_0 \). Discuss how you would go about choosing an appropriate constant to produce a test of approximately size \( \alpha \). Discuss how you would go about finding an optimal (approximately) 90% confidence set for \( \theta \).

b) Beginning with a third order Taylor expansion of the loglikelihood at \( \theta_0 \) about the MLE, prove directly that the asymptotic \( \chi^2 \) distribution holds under \( H_0 \) for the likelihood ratio statistic for testing \( H_0 : \theta = \theta_0 \) vs \( H_1 : \theta \neq \theta_0 \).

c) Suppose that \( n = 20 \) and \( n_0 = 7 \), \( n_1 = 6 \) and \( n_2 = 7 \). Plot the corresponding loglikelihood and give the value of the MLE. Based on the MLE and its normal limiting distribution, then give an approximate 90% two-sided confidence interval
for $\theta$. Finally, based on the asymptotic $\chi^2$ distribution of the likelihood ratio statistic, give a different approximate 90% confidence interval for $\theta$.

73. Consider the situation of Problems 9 and 60. Below are some data artificially generated from an exponential distribution.

$0.24, 3.20, 0.14, 1.86, 0.58, 1.15, 0.32, 3.82, 0.58, 1.60, 0.15, 0.23, 0.58, 0.11, 2.32, 2.53, 0.88$

a) Plot the loglikelihood function for the uncensored data (the $x'$ values given above). Give approximate 90% two-sided confidence intervals for $\lambda$ based on the asymptotic $\chi^2$ distribution for the LRT statistic for testing $H_0 : \lambda = \lambda_0$ vs $H_1 : \lambda \neq \lambda_0$ and based on the asymptotic normal distribution of the MLE.

Now consider the censored data problem where any value less than 1 is reported as 0. Modify the above data accordingly and do the following.

b) Plot the loglikelihood for the censored data (the derived $x$ values). How does this function of $\lambda$ compare to the one from part a)? It might be informative to plot these on the same set of axes.

c) It turns out (you might derive this fact) that

$$I_1(\lambda) = \frac{1}{\lambda^2} \left( \frac{\exp(-\lambda)}{1 - \exp(-\lambda)} \right) \left( 1 + \lambda^2 - \exp(-\lambda) \right).$$

Give two different approximate 90% confidence intervals for $\lambda$ based on the asymptotic distribution of the MLE here. Then give an approximate 90% interval based on inverting the LRTs of $H_0 : \lambda = \lambda_0$ vs $H_1 : \lambda \neq \lambda_0$.

74. Suppose that $X_1, X_2, X_3$ and $X_4$ are independent binomial random variables, $X_i \sim \text{binomial} (n_i, p_i)$. Consider the problem of testing $H_0 : p_1 = p_2$ and $p_3 = p_4$ against the alternative that $H_0$ does not hold.

a) Find the form of the LRT of these hypotheses and show that the log of the LRT statistic is the sum of the logs of independent LRT statistics for $H_0 : p_1 = p_2$ and $H_0 : p_3 = p_4$ (a fact that might be useful in directly showing the $\chi^2$ limit of the LRT statistic under the null hypothesis).

b) Find the form of the Wald tests and show directly that the test statistic is asymptotically $\chi^2$ under the null hypothesis.

c) Find the form of the $\chi^2$ tests of these hypotheses and again show directly that the test statistic is asymptotically $\chi^2$ under the null hypothesis.

75. If $X_1, X_2, \ldots, X_n$ are iid normal $(\mu, \sigma^2)$, find the form of the LRT of $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$. Show directly that the LRT statistic is asymptotically $\chi^2$ with 1 degree of freedom. Invert the LRTs to produce a large sample 90% confidence procedure for $\mu$. 

-12-