Stat 643 Problems, Fall 2002

Assignment 1

Starred problems are "recommended" but not to be turned in.

*1. Let \((\Omega, \mathcal{A})\) be a measurable space. Suppose that \(\mu, \nu, \rho\) are \(\sigma\)-finite positive measures on this space with \(\nu \ll \mu\) and \(\rho \ll \nu\).
   a) Show that \(\nu \ll \rho\) a.s. \(\mu\).
   b) Show that if \(\mu \ll \rho\), then \(\frac{d\mu}{d\rho} = \left(\frac{d\rho}{d\mu}\right)^{-1}\) 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 \(C = \{ A \in \mathcal{A} | A = -A\}\).
   a) Show that \(C\) is a sub \(\sigma\)-algebra of \(\mathcal{A}\).
   b) Let \(X\) be an integrable random variable. Find \(E(X|C)\).

*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 \(E(I_A|C) = P(A|C)\).
   b) For \(Y(\omega) = \omega_2\), find \(E(Y|C)\).

4. 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 \([0, 1] \times [0, \infty)\).)

5. (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 \(\{P_\theta\}_{\theta \in \Theta}\) is not dominated by a \(\sigma\)-finite measure, so that the factorization theorem can not 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 \( p \) is known, \( \min X_i \) is a sufficient statistic.

6. 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 \( \mathcal{R}^1 \), 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 \( P^X_\lambda \) wrt \( \mu \) on \( \mathcal{X} \).

   b) Suppose that \( X_1, X_2, \ldots, X_n \) are iid with the marginal distribution \( P^X_\lambda \). 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.

7. As a complement to Problem 6, consider this "random censoring" model. Suppose that \( X' \) is as in Problem 6. Suppose further that independent of \( X' \), \( Z \) is exponential with mean 1. We will suppose that with \( W = \min(X', Z) \) one observes
\[ X = (I[W = X'], W) \] (one sees either \( X' \) or a random censoring time less than \( X' \)). The family of distributions of \( X \) is absolutely continuous with respect to the product of counting measure and Lebesgue measure on \( \mathcal{X} = \{0,1\} \times \mathcal{R}^1 \) (call this dominating measure \( \mu \)).

   a) Find an R-N derivative of \( P^X_\lambda \) wrt \( \mu \) on \( \mathcal{X} \).

   b) Suppose that \( X_1, X_2, \ldots, X_n \) are iid with the marginal distribution \( P^X_\lambda \). Find a minimal sufficient statistic here. (Actually carefully proving minimal sufficiency looks like it would be hard. Just make a good guess and say what needs to be done to finish the proof.)

8. Schervish Exercise 2.8. And show that \( Z \) is ancillary.

9. Schervish Exercise 2.13. (I think you need to assume that \( \{x_i\} \) spans \( \mathcal{R}^k \).)

10. Schervish Exercise 2.18.

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

12. Schervish Exercise 2.27.

1A. (Bickel and Doksum (2001), page 48) Suppose that the distributions \( \mathcal{P} = \{P_\theta\}_{\theta \in \Theta} \) are dominated by a \( \sigma \)-finite measure \( \mu \) and that \( \theta_0 \in \Theta \) is such that \( f_{\theta_0} = \frac{dP_{\theta_0}}{d\mu} > 0 \) a.e. \( \mu \). Consider the function of \( x \) and \( \theta \)
The random function of $\theta, \Lambda(\theta, X)$ can be thought of as a "statistic." Argue that it is minimal sufficient.

1B. (Truncation) Consider a family of distributions $\mathcal{P} = \{P_{\theta}\}_{\theta \in \Theta}$ on $(\mathcal{R}^1, \mathcal{B}_1)$ absolutely continuous wrt Lebesgue measure $\mu$, where $f_{\theta}(x) > 0 \forall \theta$ a.e. $\mu$. Let $F_{\theta}$ be the cdf of $P_{\theta}$. For $d \in \mathcal{R}^1$ let $P_{\theta,d}$ have density

$$f_{\theta,d}(x) = I[x > d] \frac{f_{\theta}(x)}{1 - F_{\theta}(d)}$$

wrt to Lebesgue measure. Suppose that $T(X)$ is sufficient for $\theta$ in a model where $X_1, X_2, ..., X_n$ are iid $P_{\theta},$

a) Prove or give a counterexample that $(T(X), \min X_i)$ is sufficient for $(\theta, d)$ in a model where $X_1, X_2, ..., X_n$ are iid $P_{\theta,d}$.

b) If $T(X)$ is minimal sufficient for $\theta$, is $(T(X), \min X_i)$ guaranteed to be minimal sufficient for $(\theta, d)$ in a model where $X_1, X_2, ..., X_n$ are iid $P_{\theta,d}$?
Assignment 2

13. Schervish Exercise 2.34 (consider \( n \) iid observations).


15. Schervish Exercise 2.42.


17. Schervish Exercise 2.46.

18. Prove the following version of Theorem 24 (that Vardeman gave only a discrete case argument for):

Suppose that \( X = (T, S) \) and the family \( \mathcal{P} \) of distributions of \( X \) on \( \mathcal{X} = T \times S \) is dominated by the measure \( \mu = \nu \times \gamma \) (a product of \( \sigma \)-finite measures on \( T \) and \( S \) respectively). With \( \frac{dP_\theta}{d\mu}(t, s) = f_\theta(t, s) \), let

\[
g_\theta(t) = \int_S f_\theta(t, s) d\gamma(s) \quad \text{and} \quad f_\theta(s|t) = \frac{f_\theta(t, s)}{g_\theta(t)}
\]

Suppose that \( \mathcal{P} \) is FI regular at \( \theta_0 \) and that \( g'_\theta(t) = \int_S f'_\theta(t, s) d\gamma(s) \forall t \). Then

\[
I_X(\theta_0) \geq I_T(\theta_0)
\]

19. Let \( P_0 \) and \( P_1 \) be two distributions on \( \mathcal{X} \) and \( f_0 \) and \( f_1 \) be densities of these with respect to some dominating \( \sigma \)-finite measure \( \mu \). Consider the parametric family of distributions with parameter \( \theta \in [0, 1] \) and densities wrt \( \mu \) of the form

\[
f_\theta(x) = (1 - \theta)f_0(x) + \theta f_1(x)
\]

and suppose that \( X \) has density \( f_\theta \).

a) If \( G(\{0\}) = G(\{1\}) = \frac{1}{2} \), find the posterior distribution of \( \theta|X \).

b) Suppose now that \( G \) is the uniform distribution on \([0, 1]\). Find the posterior distribution of \( \theta|X \).

i) What is the mean of this posterior distribution (one Bayesian way of inventing a point estimator for \( p \))? 

ii) Consider the partition of \( \Theta \) into \( \Theta_0 = [0, .5] \) and \( \Theta_1 = (.5, 1.0] \). One Bayesian way of inventing a test for \( H_0: \theta \in \Theta_0 \) is to decide in favor of \( \Theta_0 \) if the posterior probability assigned to \( \Theta_0 \) is at least .5. Describe as explicitly as you can the subset of \( \mathcal{X} \) where this is the case.

24. Prove the discrete \( X \) version of Theorem 30.
2A. Prove the inequality in Theorem 30 in the context of Problem 18 (assume $P$ and $Q$ are elements of $\mathcal{P}$).
Assignment 3

20. Consider the situation of problem 6 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 $I_1(\lambda)$.

21. Consider the situation of Problems 6 and 20. Below are some data artificially
    generated from an exponential distribution.
    
    \[ .24, 3.20, .14, 1.86, .58, 1.15, .32, .66, 1.60, .34, 
    .61, .09, 1.18, 1.29, .23, .58, .11, 3.82, 2.53, .88 \]
    
    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 $\tilde{x}$ 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 $H_1 : \lambda \neq \lambda_0$.

22. Suppose that $X_1$, $X_2$, $X_3$ and $X_4$ are independent binomial random variables,
    $X_i \sim \text{binomial}(n, p_i)$. 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.

23. Suppose that \( X_1, X_2, \ldots \) are iid, each taking values in \( \mathcal{X} = \{0, 1, 2\} \) with RN derivative of \( P_\theta \) wrt to counting measure on \( \mathcal{X} \\
\frac{\exp(x\theta)}{1 + \exp\theta + \exp(2\theta)}
\)

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 \( L'_n(\hat{\theta}_n) = L'_n(\theta) + (\hat{\theta}_n - \theta)L''_n(\theta) + \frac{1}{2}(\hat{\theta}_n - \theta)^2L'''_n(\theta^n) \) for some \( \theta^n \) between \( \hat{\theta}_n \) and \( \theta \).

d) Show that provided \( \bar{x} \in (0, 2) \) the loglikelihood has a maximizer
\[ \hat{\theta}_n = \log \left( \frac{x - 1 + \sqrt{-3x^2 + 6x + 1}}{2(2 - \bar{x})} \right) \]
Prove that an estimator defined to be \( \tilde{\theta}_n \) 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, \( \tilde{\theta}_n \)?

By the way, a version of nearly everything in this problem works in any one parameter exponential family.

3A. Consider a one-parameter exponential family of the form in Fact 10 on the list of 643 Results. (\( \eta \in \mathcal{R}^1 \).) Without loss of generality, assume that \( h(x) > 0 \ \forall x \) and in fact, for part e) below, you may assume that \( h(x) = 1 \) (i.e. a varying \( h \) has been absorbed into the dominating measure \( \mu \)). Consider first a single observation from this family of distributions, \( X \).

a) Argue that the model \( \mathcal{P} \) is FI regular at every \( \eta_0 \) in the interior of \( \Gamma \).

b) Give an expression for the FI in \( X \) about \( \eta \) at \( \eta_0 \).
c) Evaluate the K-L information $I_X(\eta_0, \eta)$.

Now suppose that $X_1, X_2, \ldots, X_n$ are iid $P_\eta$.

d) What is the form of the likelihood equation for this problem?

e) Verify the hypotheses of Theorem 33 hold at every $\eta_0$ in the interior of $\Gamma$.

f) What is the general form of a "one-step Newton 'improvement" on a consistent estimator of $\eta$, say $\hat{\eta}$?

g) Verify that the hypotheses of Theorem 36 hold at every $\eta_0$ in the interior of $\Gamma$.

3B. Under the hypotheses of Theorem 36, consider testing the one-sided hypothesis $H_0: \theta \leq \theta_0$ vs $H_1: \theta > \theta_0$ using a likelihood ratio test. Add to the hypotheses of Theorem 36 the assumption that the $\theta_0$ probability that $L_n(\theta)$ is unimodal (i.e. $L'_n(\theta) > 0$ for $\theta < \hat{\theta}_n$ and $L'_n(\theta) < 0$ for $\theta > \hat{\theta}_n$) converges to 1. Consider the LR type statistic

$$2 \left( L_n(\hat{\theta}_n) - \max_{\theta \leq \theta_0} L_n(\theta) \right)$$

that is asymptotically equal to

$$2 I[\theta_0 < \hat{\theta}_n] \left( L_n(\hat{\theta}_n) - L_n(\theta_0) \right)$$

and find the limiting distribution of $2\log \lambda(x)$ under $\theta_0$.

It may help to think about the following (non-asymptotic) distribution problem: Consider $Z \sim N(0, 1)$. What is the distribution of $Y = I[Z \geq 0] \cdot Z^2$.
Assignment 4

*25. (Optional) Consider the problem of Bayesian inference for the binomial parameter $p$. In particular, for sake of convenience, consider the Uniform $(0, 1)$ (Beta($\alpha, \beta$) for $\alpha = \beta = 1$) prior distribution.

(a) It is possible to argue from reasonably elementary principles that in this binomial context, where $\Theta = (0, 1)$, the Beta posteriors have a consistency property. That is, simple arguments can be used to show that for any fixed $p_0$ and any $\epsilon > 0$, for $X_n \sim \text{binomial}(n, p_0)$, the random variable

$$Y_n = \int_{p_0 - \epsilon}^{p_0 + \epsilon} \frac{1}{B(\alpha + X_n, \beta + (n - X_n))} p^{\alpha + X_n - 1} (1 - p)^{\beta + (n - X_n) - 1} dp$$

(which is the posterior probability assigned to the interval $(p_0 - \epsilon, p_0 + \epsilon)$) converges in $p_0$ probability to 1 as $n \to \infty$. This problem is meant to lead you through this argument. Let $\epsilon > 0$ and $\delta > 0$.

i) Argue that there exists $m$ such that if $n \geq m$, $\left| \frac{x_n}{n} - \frac{\alpha + x_n}{\alpha + \beta + n} \right| < \frac{\epsilon}{3}$ for all $x_n = 0, 1, ..., n$.

ii) Note that the posterior variance is $\frac{(\alpha + x_n)(\beta + n - x_n)}{\alpha + \beta + n + 1}$. Argue there is an $m'$ such that if $n \geq m'$ the probability that the posterior assigns to

$$\left( \frac{\alpha + x_n}{\alpha + \beta + n} - \frac{\epsilon}{3}, \frac{\alpha + x_n}{\alpha + \beta + n} + \frac{\epsilon}{3} \right)$$

is at least $1 - \delta$ for all $x_n = 0, 1, ..., n$.

iii) Argue that there is an $m''$ such that if $n \geq m''$ the $p_0$ probability that

$$\left| \frac{x_n}{n} - p_0 \right| < \frac{\epsilon}{3}$$

is at least $1 - \delta$.

Then note that if $n \geq \max(m, m', m'')$ i) and ii) together imply that the posterior probability assigned to $\left( \frac{x_n}{n} - \frac{2\epsilon}{3}, \frac{x_n}{n} + \frac{2\epsilon}{3} \right)$ is at least $1 - \delta$ for any realization $x_n$. Then provided $\left| \frac{x_n}{n} - p_0 \right| < \frac{\epsilon}{3}$ the posterior probability assigned to $(p_0 - \epsilon, p_0 + \epsilon)$ is also at least $1 - \delta$. But iii) says this happens with $p_0$ probability at least $1 - \delta$. That is, for large $n$, with $p_0$ probability at least $1 - \delta$, $Y_n \geq 1 - \delta$. Since $\delta$ is arbitrary, (and $Y_n \leq 1$) we have the convergence of $Y_n$ to 1 in $p_0$ probability.

(b) Corollary 45 suggests that in an iid model, posterior densities for large $n$ tend to look normal (with means and variances related to the likelihood material). The posteriors in this binomial problem are Beta $\left( \alpha + x_n, \beta + (n - x_n) \right)$ (and we can think of $X_n \sim \text{binomial}(n, p_0)$ as derived as the sum of $n$ iid Bernoulli ($p_0$) variables). So we ought to expect Beta distributions for large parameter values to look roughly normal. To illustrate this (and thus essentially what Corollary 45 says in this case), do the following. For $\rho = .3$ (for example ... any other value would do as well), consider the Beta $\left( \alpha + n \rho, \beta + n (1 - \rho) \right)$ (posterior) distributions for $n = 10, 20, 40$ and 100. For $p_n \sim \text{Beta} \left( \alpha + n \rho, \beta + n (1 - \rho) \right)$ plot the probability densities for the variables
on a single set of axes along with the standard normal density. Note that if \( W \) has
pdf \( f(\cdot) \), then \( aW + b \) has pdf \( g(\cdot) = \frac{1}{a} f\left(\frac{x-b}{a}\right) \). (Your plots are translated and
rescaled posterior densities of \( \rho \) based on possible observed values \( x_n = \frac{3}{n} \).)

If this is any help in doing this plotting, I tried to calculate values of the Beta
function using MathCad and got the following:
\[
\begin{align*}
(B(4, 8))^{-1} &= 1.32 \times 10^3, \\
(B(7, 15))^{-1} &= 8.14 \times 10^5, \\
(B(13, 29))^{-1} &= 2.291 \times 10^{11} \text{ and } (B(31, 71))^{-1} = 2.967 \times 10^{27}.
\end{align*}
\]

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

27. Consider the following model. Given parameters \( \lambda_1, \ldots, \lambda_N \) variables \( X_1, \ldots, X_N \)
are independent Poisson variables, \( X_i \sim \text{Poisson}(\lambda_i) \). \( M \) is a parameter taking values in
\( \{1, 2, \ldots, N\} \) and if \( i \leq M \), \( \lambda_i = \mu_1 \), while if \( i > M \), \( \lambda_i = \mu_2 \). (\( M \) is the number
of Poisson means that are the same as that of the first observation.) With parameter vector
\( \theta = (M, \mu_1, \mu_2) \) belonging to \( \Theta = \{1, 2, \ldots, N\} \times (0, \infty) \times (0, \infty) \) we wish to make
inference about \( M \) based on \( X_1, \ldots, X_N \) in a Bayesian framework.

As matters of notational convenience, let \( S_m = \sum_{i=1}^{m} X_i \) and \( T = S_N \).

a) If, for example, a prior distribution \( G \) on \( \Theta \) is constructed by taking \( M \) uniform
on \( \{1, 2, \ldots, N\} \) independent of \( \mu_1 \) exponential with mean 1, independent of \( \mu_2 
\) exponential with mean 1, it is possible to explicitly find the (marginal) posterior
of \( M \) given that \( X_1 = x_1, \ldots, X_N = x_N \). Don't actually bother to finish the
somewhat messy calculations needed to do this, but show that this is possible
(indicate clearly why appropriate integrals can be evaluated explicitly).

b) Suppose now that \( G \) is constructed by taking \( M \) uniform on \( \{1, 2, \ldots, N\} \)
independent of \( (\mu_1, \mu_2) \) with joint density \( g(\cdot, \cdot) \) wrt Lebesgue measure on
\( (0, \infty) \times (0, \infty) \). Describe in as much detail as possible a SSS based method for
approximating the posterior of \( M \), given \( X_1 = x_1, \ldots, X_N = x_N \). (Give the
necessary conditionals up to multiplicative constants, say how you're going to use
them and what you'll do with any vectors you produce by simulation.)
28. Consider the simple two-dimensional discrete distribution for \( \theta = (\theta_1, \theta_2) \) given in the table below.

\[
\begin{array}{cccc}
\theta_2 & 1 & 2 & 3 & 4 \\
4 & 0 & 0 & .2 & .1 \\
3 & 0 & 0 & .05 & .05 \\
2 & .2 & .1 & 0 & 0 \\
1 & .2 & .1 & 0 & 0 \\
\end{array}
\]

a) This is obviously a very simple setup where anything one might wish to know or say about the distribution is easily derivable from the table. However, for sake of example, suppose that one wished to use SSS to approximate \( Q = P[\theta_1 \leq 2] = .6 \). Argue carefully that Gibbs Sampling (SSS) will fail here to produce a correct estimate of \( Q \). What is it about the transition matrix that describes SSS here that "causes" this failure?

b) Consider the very simple version of the Metropolis-Hastings algorithm where the matrix \( T = (t_{\theta', \theta}) \) is \( \frac{1}{8} \mathbf{I} \) (for \( \mathbf{I} \) an \( 8 \times 8 \) matrix of 1's) and \( \theta' \) and \( \theta \) are vectors for which the corresponding probabilities in the table above are positive. Find the transition matrix \( P \) and verify that it has the distribution above as it invariant distribution.

29. Suppose for sake of argument that I am interested in the properties of the discrete distribution with probability mass function

\[
f(x) = \begin{cases} 
k \frac{\sin x}{2x} & \text{for } x = 1, 2, \ldots \\ 0 & \text{otherwise} \end{cases}
\]

but, sadly, I don't know \( k \).

a) Describe a MCMC method I could use to approximate the mean of this distribution.

(b) One doesn't really have to resort to MCMC here. One could generate iid observations from \( f(x) \) (instead of a MC with stationary distribution \( f(x) \)) using the "rejection method." (And then rely on the LLN to produce an approximate mean for the distribution.) Prove that the following algorithm (based on iid Uniform (0, 1) random variables and samples from a geometric distribution) will produce \( X \) from this distribution.

**Step 1** Generate \( X^* \) from the geometric distribution with probability mass function

\[
h(x) = \begin{cases} 
\frac{1}{2x} & \text{for } x = 1, 2, \ldots \\ 0 & \text{otherwise} \end{cases}
\]

**Step 2** Generate \( U \) that is Uniform (0, 1)
Step 3  Note that \( h(x) \geq \frac{|\sin x|}{2x} \). If \( Uh(x) < \frac{|\sin x|}{2x} \) set \( X = X^* \). Otherwise return to Step 1.

30. 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.)

31. Consider the two state decision problem with \( \Theta = \{1, 2\} \), \( P_1 \) the Bernoulli \( \left( \frac{1}{2} \right) \) distribution and \( P_2 \) the Bernoulli \( \left( \frac{1}{3} \right) \) 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) (This is tedious. Do it if you like, it is not required.) For this problem, show explicitly that any element of \( D^* \) has a corresponding element of \( D \), with identical risk vector and vice versa.
   c) Identify 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 \( g = \left( \frac{1}{2}, \frac{1}{2} \right) \).

32. Consider a two state decision problem with \( \Theta = \{1, 2\} \), where the observable \( X = (X_1, X_2) \) has iid Bernoulli \( \left( \frac{1}{2} \right) \) coordinates if \( \theta = 1 \) and iid Bernoulli \( \left( \frac{1}{3} \right) \) 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
   \[
   \phi_x(\{1\}) = 1 \quad \text{if } x_1 = 0, \quad \phi_x(\{1\}) = \frac{1}{2} \quad \text{and } \phi_x(\{2\}) = \frac{1}{2} \quad \text{if } x_1 = 1.
   \]
   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 31.)
   b) Use the construction from Result 61 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.)

33. 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 \( \psi \) be a randomized decision function that chooses \( \delta_1 \) with probability \( \frac{1}{2} \) and \( \delta_2 \) with probability \( \frac{1}{2} \).
   a) Write out expressions for the risk functions of \( \delta_1 \), \( \delta_2 \), and \( \psi \).
   b) Find a behavioral rule that is risk equivalent to \( \psi \).
   c) Identify a nonrandomized estimator that is strictly better than \( \psi \) or \( \phi \).
34. Suppose that $\Theta = \Theta_1 \times \Theta_2$ and that a decision rule $\phi$ is such that for each $\theta_2$, $\phi$ is admissible when the parameter space is $\Theta_1 \times \{\theta_2\}$. Show that $\phi$ is then admissible when the parameter space is $\Theta$.

35. Suppose that $w(\theta) > 0 \forall \theta$. Show that $\phi$ is admissible with loss function $L(\theta, a)$ iff it is admissible with loss function $w(\theta)L(\theta, a)$.

36. Problem 9, page 209 of Schervish.

4A. Suppose that I am interested in the properties of a discrete joint distribution with (joint) pmf proportional to

$$g(x, y) = \begin{cases} \frac{\sin x}{x+y} & \text{for } x = 1, 2, \ldots \text{ and } y = 1, 2, \ldots \\ 0 & \text{otherwise} \end{cases}$$

I decide to use MCMC to generate a large number of $(x, y)$ pairs whose empirical distribution should approximate this discrete distribution.

a) In light of the logic in part b) of Problem 29, describe a straightforward SSS algorithm that uses values from a standard distribution for the "y" substitutions and a rejection algorithm to get values for "x" substitutions.

b) As a second MCMC algorithm here, describe a "Gibbs/Metropolis hybrid" algorithm that uses values from a standard distribution in a "Gibbs step" alternating with a "Metropolis/Hastings step" to create the $(x, y)$ pairs. Argue that indeed your algorithm has the correct invariant distribution.
Assignment 5

37. (Ferguson) Prove or give a counterexample: If $C_1$ and $C_2$ are complete classes of decision rules, then $C_1 \cap C_2$ is essentially complete.

38. (Ferguson and others) Suppose that $X \sim \text{binomial }(n, p)$ and one wishes to estimate $p \in (0, 1)$. 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.
   c) Show that $\frac{X}{n}$ is minimax and identify a least favorable prior.

Now consider ordinary squared error loss, $L(p, a) = (p - a)^2$.
   d) Apply the result of problem 35 and prove that $\frac{X}{n}$ is admissible under this loss function as well.

39. 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?

40. 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.

41. Consider a decision problem where $P_\theta$ is the Normal $(\theta, 1)$ distribution on $\mathcal{X} = \mathcal{R}^1$, $\mathcal{A} = \{0, 1\}$ and $L(\theta, a) = I[a = 0]I[\theta > 5] + I[a = 1]I[\theta \leq 5]$.
   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 $\Theta = \mathcal{R}^1$, guess what decision rule might be minimax, find its risk function and prove that it is minimax.

42. (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.
43. Problem 14, page 340 Schervish.

44. Let $N = (N_1, N_2, ..., N_k)$ be multinomial $(n, p_1, p_2, ..., p_k)$ 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 $i = 1, 2, ..., k$ 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 $h(y_1, y_2, ..., y_k)$ 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

   $$\frac{(q'(\theta))^2}{nI_1(\theta)}$$

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

45. Show that the C-R inequality is not changed by a smooth reparameterization. That is, suppose that $P = \{P_0\}$ 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 $\Theta \subset \mathcal{R}^1$ 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 $\mathcal{R}^1$ 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 $P$.

46. As an application of Lemma 88 (or 88') 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 88 with $g_1(\theta) = 1 - \frac{f_{\theta_1', \theta_2}(x)}{f_{\theta_1, \theta_2}(x)}$ and $g_2(\theta) = 1 - \frac{f_{\theta_1, \theta_2'}(x)}{f_{\theta_1, \theta_2}(x)}$.

   What does Lemma 88 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.)

47. Problem 11, page 389 Schervish.

48. Consider the estimation problem with $n$ iid $P_\theta$ observations, where $P_\theta$ is the exponential distribution with mean $\theta$. Let $\mathcal{G}$ be the group of scale transformations on $\mathcal{X} = (0, \infty)^n$, $\mathcal{G} = \{g_c|c > 0\}$ where $g_c(x) = cx$.

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

   b) Show that the estimation problem with loss $L(\theta, a) = (\theta - a)^2/\theta^2$ is invariant under $\mathcal{G}$, 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).

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

50. Let $f(u, v)$ be the bivariate probability density of the distribution uniform on the square in $(u, v)$-space with corners at $(\sqrt{2}, 0), (0, \sqrt{2}), (-\sqrt{2}, 0)$ and $(0, -\sqrt{2})$.

Suppose that $X = (X_1, X_2)$ has bivariate probability density $f(x|\theta) = f(x_1 - \theta, x_2 - \theta)$. Find explicitly the best location equivariant estimator of $\theta$ under squared error loss. (It may well help you to visualize this problem to "sketch" the joint density here for $\theta = 17$.)

51. Consider again the scenario of problem 31. There you sketched the risk set $\mathcal{S}$. Here sketch the corresponding set $\mathcal{V}$.

52. 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 $\mathcal{R}^1$. 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 $T(x)$.

53. Consider the two distributions $P_0$ and $P_1$ on $\mathcal{X} = (0, 1)$ with densities wrt Lebesgue measure $f_0(x) = 1$ 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 $\mathcal{S}$ and the set $\mathcal{V} = \{\beta_\alpha(\theta_0), \beta(\theta_1)\}$ 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 52 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.

54. 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
c) Suppose that \( X \) is Normal \((\theta, 1)\). Find Bayes tests of \( H_0 : \theta = 0 \) vs \( H_1 : \theta \neq 0 \)

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.

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

\[
\begin{array}{c|cccccc}
   x & -2 & -1 & 0 & 1 & 2 \\
\hline
\theta = -1 & \frac{\alpha}{2} & \frac{\alpha}{2} - \alpha & \alpha & \frac{\alpha}{2} - \alpha & \frac{\alpha}{2} \\
\theta \neq -1 & \theta c & \left(\frac{1-c}{1-\alpha}\right)(\frac{1}{2} - \alpha) & \theta c & \left(\frac{1-c}{1-\alpha}\right)(\frac{1}{2} - \alpha) & \frac{\alpha}{2} \\
\end{array}
\]

Find the size \( \alpha \) likelihood ratio test of \( H_0 : \theta = -1 \) vs \( H_1 : \theta \neq -1 \). Show that the test \( \hat{\phi}(x) = I[x = 0] \) 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.)

56. 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.

\[
\begin{array}{ccccccccccccccc}
   x & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\
\hline
   \theta = 10 & .0000 & .0005 & .0023 & .0076 & .0189 & .0378 & .0631 & .0901 & .1126 & .1251 & .1251 & .1137 & .0948 \\
   x & 13 & 14 & 15 & 16 & 17 & 18 & 19 & 20 & 21 & 22 & 23 & 24 & 25 \\
   \theta = 4 & .0002 & .0001 & .0000 \\
   \theta = 10 & .0729 & .0521 & .0347 & .0217 & .0128 & .0071 & .0037 & .0019 & .0009 & .0004 & .0002 & .0001 & .0000 \\
\end{array}
\]

57. 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 \( H_1 : \lambda \in (.5, 2) \).

58. Consider the situation of problem 23. 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 \( \theta \).

59. (Problem 37, page 122 TSH.) Let \( X_1, \ldots, X_n \) and \( Y_1, \ldots, Y_m \) 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 \( \alpha \).)
60. Problem 57 Schervish, page 293.