Stat 643 Spring 2010
Assignments

1 Characteristic Functions

1. (Some simple stuff about complex analysis.)

a) Find the series expansions of $\exp(x)$, $\cos(x)$, and $\sin(x)$ for real numbers x. Substitute the imaginary number $\iota\theta$ (for θ real) into the first of these and then say why the definition

$$\exp(\iota\theta) \equiv \cos(\theta) + \iota\sin(\theta)$$

is formally a sensible definition.

b) Show that for real a, b

$$\exp(\iota(a+b)) = \exp(\iota a)\exp(\iota b)$$

c) For real a, b define

$$\exp(a+\iota b) = \exp(a)\exp(\iota b)$$

and argue that for complex numbers u and v,

$$\exp(u+v) = \exp(u)\exp(v)$$

d) Suppose that f and g are complex-valued functions of a real variable, and that X and Y are independent random variables. Show (assuming that $\mathrm{E}f(X)$ and $\mathrm{E}g(Y)$ exist, i.e. $\mathrm{E}|\mathrm{Re}(f(X))| < \infty$,$\mathrm{E}|\mathrm{Im}(f(X))| < \infty$,$\mathrm{E}|\mathrm{Re}(g(Y))| < \infty$, and $\mathrm{E}|\mathrm{Im}(g(Y))| < \infty$) that

$$\mathrm{E}f(X)g(Y) = \mathrm{E}f(X)\,\mathrm{E}g(Y)$$

e) Show that for f a complex-valued function of a real variable

$$|\mathrm{E}f(X)| \le \mathrm{E}\,|f(X)|$$

(Hint: Since $\mathrm{E}f(X)$ is complex, $\exists$ a complex number α with $|\alpha| = 1$ so that $\alpha\mathrm{E}f(X) = |\mathrm{E}f(X)|$. Then $\mathrm{Re}\,\alpha f \le |\alpha f| = |f|$ and $\mathrm{E}\alpha f(X)$ must be real.)

2. Show (from the definition of $\exp(\iota\theta)$) that for $a < b$

$$\left|\frac{\exp(-\iota ta) - \exp(-\iota tb)}{\iota t}\right| = \left|\int_a^b \exp(-\iota tx)\,dx\right|$$

3. (Rosenthal 11.5.7) Let μ_n be the discrete uniform distribution on $\left\{0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}\right\}$ and λ be the U$(0,1)$ distribution. Find the characteristic functions ϕ_n for μ_n and ϕ for λ. Show that

$$\phi_n(t) \to \phi(t) \quad \forall t \in \mathbb{R}$$

4. Problem 10.9 of Athreya and Lahiri.

5. Problem 10.10 of Athreya and Lahiri.

6. Problem 10.14 of Athreya and Lahiri.

7. Problem 10:15 of Athreya and Lahiri.

8. Problem 10.19 of Athrey and Lahiri.

9. Problem 10.20 of Athreya and Lahiri.

10. Prove the following simple consequence of Corollary 10.4.2 of Athreya and Lahiri. k-dimensional random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$ have the same distribution if and only if for every k-vector $\boldsymbol{a}$, $\boldsymbol{a}'\boldsymbol{X}$ and $\boldsymbol{a}'\boldsymbol{Y}$ have the same distribution.

2 Central Limit Theorems

11. (Ferguson, page 34) Suppose that $\{X_j\}$ are independent random variables with

$$P\left[X_j = -\sqrt{j}\right] = P\left[X_j = \sqrt{j}\right] = \frac{1}{2}$$

For $S_i = \sum_{j=1}^{i} X_j$ and $B_i^2 = \text{Var}S_i$, does S_i/B_i have a normal limit in distribution? Ague carefully one way or the other.

12. (Ferguson, page 34) Suppose that $\{X_j\}$ are iid with mean μ and variance σ^2. Let $T_i = \sum_{j=1}^{i} z_{ij} X_j$ for a double array of numbers $\{z_{ij}\}$. For $\mu_i = \text{E}T_i$ and $\sigma_i^2 = \text{Var}T_i$, show that $(T_i - \mu_i)/\sigma_i \xrightarrow{d} \text{N}(0,1)$ provided that

$$\max_{j \le i} \frac{z_{ij}^2}{\sum_{j=1}^{i} z_{ij}^2} \underset{i\to\infty}{\longrightarrow} 0$$

13. (Ferguson, page 34) (Records) Suppose that $\{Z_j\}$ are iid continuous random variables. We will say that a record occurs at i if $Z_i > Z_j \ \forall\ j < i$. Let $R_i = 1$ if a record occurs at i and $R_i = 0$ otherwise. It is then possible to argue that $\{R_i\}$ are independent Bernoulli random variables, with $P[R_i = 1] = 1/i$. (You need not argue this if you don't see how to do so.) Let $S_n = \sum_{i=1}^{n} R_i$ be the number of records in the first n observations. Show that

$$\frac{S_n - \text{E}S_n}{\sqrt{\text{Var}S_n}} \xrightarrow{d} \text{N}(0,1)$$

14. (Chung, page 223) Find an example where Lindeberg's condition is satisfied, but Liapounov's is not for any $\delta > 0$.

15. Problem 11.2 of Athreya and Lahiri.

16. Problem 11.4 of Athreya and Lahiri.

17. Problem 11.10 of Athreya and Lahiri.

18. Problem 11.14 of Athreya and Lahiri.

19. Suppose that $\{Y_j\}_{j=0,1,\ldots}$ is a sequence of iid Bern(p) random variables. For $j = 1, 2, \ldots$ let

$$X_j = Y_j + Y_{j-1}$$

Does $\{X_j\}$ satisfy the hypotheses of Theorem 23 for some m? With $S_n = \sum_{j=1}^{n} X_j$ are there a_n and b_n such that $(S_n - a_n)/b_n \xrightarrow{d} \text{N}(0,1)$?

20. Find an example of a 2-dependent sequence of uniformly bounded random variables for which $\sigma_n^2 \to \infty$, but $\sigma_n / n^{1/3} \nrightarrow \infty$. For your example, try to determine whether or not

$$\frac{S_n - \text{E}S_n}{\sigma_n} \xrightarrow{d} \text{N}(0,1)$$

3 Conditional Expectation

21. Prove Theorem 32 from lecture. That is, prove that
a) If $c \in \mathbb{R}$, $\mathrm{E}[c|\mathcal{G}] = c$ a.s.
b) If X is $\mathcal{G}$-measurable, then $\mathrm{E}[X|\mathcal{G}] = X$ a.s.
c) If $X \geq 0$ a.s., then $\mathrm{E}[X|\mathcal{G}] \geq 0$ a.s.
d) If $c_1, c_2, \ldots, c_k$ are real numbers and $X_1, X_2, \ldots, X_k$ have expectations, $\mathrm{E}\left[\sum_{i=1}^k c_i X_i|\mathcal{G}\right] = \sum_{i=1}^k c_i \mathrm{E}[X_i|\mathcal{G}]$ a.s.
e) If $|\mathrm{E}\,[X|\mathcal{G}]| \leq \mathrm{E}[|X|\,|\mathcal{G}]$ a.s.

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

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

24. (Rosenthal 13.4.1) Let A and B be events, with $0 < P(B) < 1$. Let $\mathcal{G} = \sigma(B)$ be the σ-algebra generated by B.
a) Describe$\mathcal{G}$ explicitly.
b) Compute $P(A|\mathcal{G})$ explicitly.
c) Relate $P(A|\mathcal{G})$ to the usual notation $P(A|B) = P(A \cap B)/P(B)$.

25. (Rosenthal 13.2.4) Let A and B be disjoint events. Show that
a) $0 \leq P(A|\mathcal{G}) \leq 1$ a.s.
b) $P(A \cup B|\mathcal{G}) = P(A|\mathcal{G}) + P(B|\mathcal{G})$ a.s.

26. Problem 12.2 of Athreya and Lahiri.

27. Problem 12.12 of Athreya and Lahiri.

28. Problem 12.14 of Athreya and Lahiri.

29. Problem 12.20 of Athreya and Lahiri.

30. Problem 12.30 of Athreya and Lahiri.

4 Sufficiency and Related Notions

31. Prove the $\mathbb{R}^1$ version of Shao's Lemma 1.2, page 37 (Lehmann's Theorem/Lemma 51 from class). (Hint: Indicators, simple functions, non-negative functions, general functions.)

32. Suppose that $X = (X_1, X_2, ..., X_n)$ has independent components, where each X_i is generated as follows. For independent random variables W_i ~normal$(\mu, 1)$ and Z_i ~Poisson (μ), $X_i = W_i$ 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 $[0,1] \times [0, \infty)$.)

33. (Ferguson) Consider the probability distributions on $(\mathcal{R}^1, \mathcal{B}_1)$ defined as follows. For $\boldsymbol{\theta} = (\theta, p) \in \boldsymbol{\Theta} = \mathcal{R} \times (0,1)$ and $X \sim P_{\boldsymbol{\theta}}$, suppose

$$P_{\boldsymbol{\theta}}[X = x] = \begin{cases} (1-p)\,p^{x-\theta} & \text{for } x = \theta, \theta+1, \theta+2, \ldots \\ 0 & \text{otherwise} \end{cases}$$

Let $X_1, X_2, ..., X_n$ be iid according to $P_{\boldsymbol{\theta}}$.

a) Argue that the family $\{P^n_{\boldsymbol{\theta}}\}_{\boldsymbol{\theta} \in \boldsymbol{\Theta}}$ is not dominated by a σ-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 $\boldsymbol{\theta}$. (Argue that the conditional distribution of X given $T(X)$ doesn't depend upon θ.)

c) Argue that the factorization theorem can be applied if θ is known (the first factor of the parameter space $\boldsymbol{\Theta}$ 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.

34. Argue carefully for the sufficiency of the order statistic in symmetric problems. That is, for $(\mathcal{X}, \mathcal{B}) = \left(\mathbb{R}^k, \mathcal{B}^k\right)$ (Euclidean k-space with the Borel σ-algebra) and $\pi = (\pi_1, \pi_2, \ldots, \pi_k)$ a permutation of the first k positive integers, for$x \in \mathbb{R}^k$, let $\pi(x) \equiv (x_{\pi_1}, x_{\pi_2}, \ldots, x_{\pi_k})$ and for $B \in \mathcal{B}^k$ let $\pi(B) \equiv \left\{y \in \mathbb{R}^k | y = \pi(x) \text{ for some } x \in B\right\}$. Suppose that for each $P \in \mathcal{P}$, it is the case that $P(\pi(B)) = P(B)$ for all permutations π. Show by direct argument (not trying to appeal to the factorization theorem) that in this circumstance the order statistic $T(X) \equiv \left(X_{(1)}, X_{(2)}, \ldots, X_{(k)}\right)$ is sufficient for $\mathcal{P}$. (Hint: For $B \in \mathcal{B}^k$, try $Y(x)$ equal to the fraction of of all permuations that place x in B as a version of $P(B|\sigma(T))$.)

35. Suppose that X' is exponential with mean λ^{-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 μ on $\mathcal{X} = \{0\} \cup (1, \infty)$ consisting of a point mass of 1 at 0 plus Lebesgue measure on (1,∞). Give a formula for the R-N derivative of P^X_λ wrt μ on $\mathcal{X}$.

b) Suppose that $X_1, X_2, ..., X_n$ are iid with the marginal distribution P^X_λ. 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.

36. As a complement to Problem 35, consider this "random censoring" model. Suppose that X' is as in Problem 35. 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' and an indicator of whether it was X' or the random censoring time that is seen). 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 μ).

a) Find an R-N derivative of P^X_λ wrt μ on $\mathcal{X}$.

b) Suppose that $X_1, X_2, ..., X_n$ are iid with the marginal distribution P^X_λ. 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.)

37. (Bickel and Doksum (2001), page 48) Suppose that the distributions $\mathcal{P} = \{P_\theta\}_{\theta \in \Theta}$ are dominated by a σ-finite measure μ and that $\theta_0 \in \Theta$ is such that $f_{\theta_0} = \frac{dP_{\theta_0}}{d\mu} > 0$ a.e. μ. Consider the function of x and θ

$$\Lambda(\theta, x) = \frac{f_\theta(x)}{f_{\theta_0(x)}}$$

The random function of θ, $\Lambda(\theta, X)$ can be thought of as a "statistic." Argue that it is minimal sufficient.

38. (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 μ, where $f_\theta(x) > 0\ \forall \theta$ a.e. μ. Let F_θ be the cdf of P_θ. 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 θ in a model where $X_1, X_2, ..., X_n$ are iid P_θ.

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

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

39. (Shao Exercise 2.25) Show that if T is a sufficient statistic and $T = \psi(S)$, where ψ is another statistic and S is measurable, then S is sufficient.

40. (Shao Exercise 2.41) Suppose that $X_1, X_2, ..., X_n$ are iid $P_{\boldsymbol{\theta}}$ for $\boldsymbol{\theta} = (\theta_1, \theta_2) \in (0,1) \times \{1,2\}$, where $P_{(\gamma,1)}$ is the Poisson(γ) distribution and $P_{(\gamma,2)}$ is the Bernoulli(γ) distribution. Find a two-dimensional minimal sufficient statistic for $\boldsymbol{\theta}$ (and argue carefully for minimal sufficiency).

41. (Shao Exercise 2.45) Suppose that $X_1, X_2, ..., X_n$ are iid P_θ for $\theta \in \mathbb{R}$, where if $\theta \neq 0$ the distribution P_θ is $\mathrm{N}(\theta, 1)$, while P_0 is $\mathrm{N}(0, 2)$. Show that $\overline{X}$ is complete but not sufficient for θ.

42. (Shao Exercise 2.56) Suppose that $\boldsymbol{X}_1, \boldsymbol{X}_2, ..., \boldsymbol{X}_n$ are iid $\mathrm{MVN}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ for $\boldsymbol{\mu} = (\mu, \mu)'$ and $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}$. Argue that $\left(\overline{X}_1, \overline{X}_2, S_1^2, S_2^2\right)$ is minimal sufficient but not boundedly complete. What is a first order ancillary statistic here?

43. (Shao Exercise 2.57) Suppose that $X_1, X_2, ..., X_n$ are iid $\mathrm{N}(\gamma, \gamma^2)$ for $\gamma \in \mathbb{R}$. Find a minimal sufficient statistic and show that it is not complete.

5 Models and Measures of Statistical Information

44. (Shao Exercise 2.14) (Truncated Exponential Families) Suppose that distributions $P_{\boldsymbol{\eta}}$ have R-N derivatives wrt to some σ-finite measure μ of the form $f_{\boldsymbol{\eta}}(x) = K(\boldsymbol{\eta}) \exp\left(\sum_{i=1}^k \eta_i T_i(x)\right) h(x)$. For a measurable set A, consider the family of distributions $Q_{\boldsymbol{\eta}}^A$ on A with R-N derivatives

$$g_{\boldsymbol{\eta}}(x) \propto f_{\boldsymbol{\eta}}(x) I\left[x \in A\right]$$

Argue that this is an exponential family and say what you can about the natural parameter space for this family in comparison to that of the $P_{\boldsymbol{\eta}}$ family.

45. 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 σ-finite measure μ. Consider the parametric family of distributions with parameter $\theta \in [0,1]$ and densities wrt μ of the form

$$f_\theta(x) = (1-\theta) f_0(x) + \theta f_1(x) \ ,$$

and suppose that X has density f_θ.

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 Θ 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 Θ_0 if the posterior probability assigned to Θ_0 is at least .5. Describe as explicitly as you can the subset of $\mathcal{X}$ where this is the case.

46. Prove the following Theorem:

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

$$g_\theta(t) = \int_{\mathcal{S}} f_\theta(t,s) d\gamma(s) \ \text{ and } \ f_\theta(s|t) = \frac{f_\theta(t,s)}{g_\theta(t)}$$

Suppose that $\mathcal{P}$ is FI regular at θ_0 and that $g'_\theta(t) = \int_{\mathcal{S}} f'_\theta(t,s)d\gamma(s)\ \forall t$. Then

$$I_X(\theta_0) \geq I_T(\theta_0)$$

47. Prove the discrete X version of the following Theorem:
$I_X(P;Q) \geq I_{T(X)}\left(P^T;Q^T\right)$ and there is equality if and only if $T(X)$ is sufficient for $\{P,Q\}$

48. Prove that $I_X(P;Q) \geq I_{T(X)}\left(P^T;Q^T\right)$ in the context of Problem 46 (assume P and Q are elements of $\mathcal{P}$).

49. Suppose that Y has pdf f on $(0,\infty)$ and conditional on $Y=y$, B is Bernoulli $p(y)$, for $p(\cdot)$ an increasing function $(0,\infty) \to [0,1]$. Let

$$X = B \cdot Y$$

($X=Y$ when $B=1$ and $X=0$ otherwise. X is a censored version of Y, where the censoring probability depends on the realized value of Y.) P^X, the distribution of X, is then absolutely continuous wrt $\delta_0+\lambda$, for δ_0 a unit point mass at 0 and λ Lebesgue measure on $(0,\infty)$. An R-N derivative is

$$g(x) = \begin{cases} \int_0^\infty (1-p(y))\,f(y)\,dy & \text{if } x=0 \\ p(x)\,f(x) & \text{if } x>0 \end{cases}$$

In this context, suppose that f is the exponential density with mean β

$$f(y|\beta) = I[y>0]\frac{1}{\beta}\exp(-\frac{y}{\beta})$$

and that

$$p(y) = \begin{cases} y & \text{if } y<1 \\ 1 & \text{if } y \geq 1 \end{cases}$$

a) Show how to calculate the Fisher information in X about β at the value β_0, $I_X(\beta_0)$.
b) Under what conditions would you expect this to be essentially the same as the Fisher information in Y?

50. (Schervish Problem 47 of Chapter 2). Suppose that for $\theta = (p_Y, p_Z)$ we assume that $Y \sim$Bi(n,p_Y) and conditioned on $Y=y$, $Z \sim$Bi(y,p_Z). Let $X=(Y,Z)$ and find $\boldsymbol{I}_X(\theta_0)$.

6 Decision Theory

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

52. Consider the two state decision problem with $\Theta=\{1,2\}$, P_1 the Bernoulli $\left(\frac{1}{4}\right)$ distribution and $P_2$the Bernoulli $\left(\frac{1}{2}\right)$distribution, $\mathcal{A}=\Theta$ 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 $\mathcal{S}$, 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 $\mathcal{D}^*$ has a corresponding element of $\mathcal{D}_*$ 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 $\boldsymbol{g}=(p,1-p)$. 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 $\boldsymbol{g}=(\frac{1}{2},\frac{1}{2})$.

53. Consider a two state decision problem with $\Theta = \{1, 2\}$, where the observable $X = (X_1, X_2)$ has iid Bernoulli $\left(\frac{1}{4}\right)$coordinates if $\theta = 1$ and iid Bernoulli $\left(\frac{1}{2}\right)$coordinates if $\theta = 2$. Suppose that $\mathcal{A} = \Theta$ and $L(\theta, a) = I[\theta \neq a]$. Consider the behavioral decision rule $\phi_{\boldsymbol{x}}$ defined by

$$\begin{array}{rl} \phi_x(\{1\}) = 1 & \text{if } x_1 = 0 \\ \phi_x(\{1\}) = \frac{1}{2} = \phi_x(\{2\}) & \text{if } x_1 = 1 \end{array}$$

a) Show that $\phi_{\boldsymbol{x}}$ 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 52.)

b) Find a behavioral decision rule that is a function of the sufficient statistic $X_1 + X_2$and is risk equivalent to $\phi_{\boldsymbol{x}}$. (Note that this rule is inadmissible.)

54. Consider the squared error loss estimation of $p \in (0, 1)$, based on X ~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 $\delta_1$with probability $\frac{1}{2}$ and δ_2 with probability $\frac{1}{2}$.

a) Write out expressions for the risk functions of δ_1, δ_2 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.

55. Suppose that $\Theta = \Theta_1 \times \Theta_2$and that a decision rule ϕ is such that for each θ_2, ϕ is admissible when the parameter space is $\Theta_1 \times \{\theta_2\}$. Show that ϕ is then admissible when the parameter space is Θ.

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

57. Shao Exercise 2.64

58. Shao Exercise 2.78

59. (Ferguson) Prove or give a counterexample: If $\mathcal{C}_1$ and $\mathcal{C}_2$ are complete classes of decision rules, then $\mathcal{C}_1 \cap \mathcal{C}_2$ is essentially complete.

60. (Ferguson and others) Suppose that X ~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 56 and prove that $\frac{X}{n}$ is admissible under this loss function as well.

61. 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 σ-finite measure μ, 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?

62. Suppose that X ~Bernoulli (p) and that one wishes to estimate p with loss $L(p, a) = |p - a|$. Consider the estimator δ with $\delta(0) = \frac{1}{4}$ and $\delta(1) = \frac{3}{4}$.

a) Write out the risk function for δ 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 δ is Bayes.

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

63. Consider a decision problem where P_θ 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.

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

a) Show that δ is an equalizer rule.

b) Show that δ is generalized Bayes versus Lebesgue measure on Λ.

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

d) Prove that δ is minimax for this problem.

7 Asymptotics of Likelihood-Based Inference

65. Consider the situation of problem 35 and maximum likelihood estimation of λ.

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

b) Give a simple estimator of λ based on M alone. Prove that this estimator is consistent for λ. 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 (Wald) confidence intervals for λ based on the MLE $\hat{\lambda}_n$ and two different approximations to $I_1(\lambda)$.

66. Consider the situation of Problems 35 and 65. 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 λ based on the asymptotic χ^2 distribution for the LRT statistic for testing $\mathrm{H}_0 : \lambda = \lambda_0$ vs $\mathrm{H}_1 : \lambda \neq \lambda_0$ 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 λ 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 (Wald) 90% confidence intervals for λ based on the asymptotic distribution of the MLE here. Then give an approximate 90% interval based on inverting the LRTs of $\mathrm{H}_0{:}\lambda = \lambda_0$ vs $\mathrm{H}_1{:}\lambda \neq \lambda_0$.

67. Suppose that X_1, X_2, X_3 and X_4 are independent binomial random variables, $X_i \sim$binomial (n, p_i). Consider the problem of testing $\mathrm{H}_0{:}p_1 = p_2$ 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 $\mathrm{H}_0{:}p_1 = p_2$ and $\mathrm{H}_0{:}p_3 = p_4$ (a fact that might be useful in directly showing the χ^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 χ^2 under the null hypothesis.

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

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

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

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

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

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

d) Show that provided $\bar{x} \in (0, 2)$ the loglikelihood has a maximizer

$$\widehat{\hat{\theta}}_n = \ln\left(\frac{\bar{x} - 1 + \sqrt{-3\bar{x}^2 + 6\bar{x} + 1}}{2(2 - \bar{x})}\right) \quad .$$

Prove that an estimator defined to be $\widehat{\hat{\theta}}_n$ 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 θ. What do these look like based on, say, $\widehat{\hat{\theta}}_n$?

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

69. Consider a one-parameter exponential family of the form in Definition 65. ($\eta \in \mathcal{R}^1$.) 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 μ). Consider first a single observation from this family of distributions, X.

a) Argue that the model $\mathcal{P}$ is FI regular at every η_0 in the interior of Γ.

b) Give an expression for the FI in X about η at η_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_η.

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

e) Verify that the hypotheses of Theorem 170 hold at every η_0 in the interior of Γ.

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

g) Verify that the hypotheses of Theorem 173 hold at every η_0 in the interior of Γ.

70. Under the hypotheses of Theorem 173, consider testing the one-sided hypothesis H$_0$:$\theta \leq \theta_0$ vs H$_1$:not H$_0$ using a likelihood ratio test. Add to the hypotheses of Theorem 173 the assumption that the θ_0 probability that $L_n(\theta)$ is unimodal (i.e. $L_n'(\theta) > 0$ for $\theta < \widehat{\theta}_n$ and $L_n'(\theta) < 0$ for $\theta > \widehat{\theta}_n$) converges to 1. Consider the LR type statistic

$$2\left(L_n(\widehat{\theta}_n) - \max_{\theta < \theta_0} L_n(\theta)\right)$$

that is asymptotically equal to

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

and find the limiting distribution of $2\ln\lambda(x)$ under θ_0.

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

71. Prove Corollary 175.

72. Prove Theorem 176.

8 Invariance

73. Consider the estimation problem with n iid P_θ observations, where P_θ is the exponential distribution with mean θ. 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 θ 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).

74. 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 θ under squared error loss. (It may well help you to visualize this problem to "sketch" the joint density here for $\theta = 17$.)

75. (Problem 11, page 389 Schervish) Let $X_1, \ldots X_n$ be iid $\mathrm{N}(\mu, \sigma^2)$ and for $\mathcal{A} = \{0, 1\}$ let

$$L(\theta, a) = RI[\mu \geq \mu_0] I[a = 1] + I[\mu < \mu_0] I[a = 0]$$

Let $\mathcal{G}$ be a group that for $c > 0$ acts on $\mathcal{X}$ as

$$g_c(x_1, \ldots x_n) = (c(x_1 - \mu_0) + \mu_0, \ldots, c(x_n - \mu_0) + \mu_0)$$

Find $\overline{\mathcal{G}}$ and $\widetilde{\mathcal{G}}$ so that the problem is invariant, and show that the t-test is equivariant.

9 Information Inequalities for Estimation

76. 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 σ-finite measure μ and satisfies

i) $f_\theta(x) > 0$ for all θ 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 δ with $\mathrm{E}_\theta|\delta(X)| < \infty$ for all θ, $\mathrm{E}_\theta \delta(X)$ can be differentiated under the integral sign at all points of Θ.

Let h be a function from Θ to $\mathcal{R}^1$ such that h' is continuous and nonvanishing on Θ. 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 $\mathcal{P}$.

77. Show that for $X \sim \mathrm{U}(0, \theta)$, the hypotheses of Theorem 197 are not met. What does Theorem 198 provide for a lower bound on the variance of an unbiased estimator of θ? Identify an unbiased estimator of θ based on X and compare its variance to the Chapman-Robbins lower bound.

78. 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 $\boldsymbol{p}$.

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

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

(See Shao's Theorem 3.3 page 169 for the multiparameter C-R inequality.)

10 Testing

79. Consider again the scenario of problem 52. There you sketched the risk set $\mathcal{S}$. Here sketch the corresponding set $\mathcal{V}$.

80. Consider a composite versus composite testing problem in a family of distributions $\mathcal{P} = \{P_\theta\}$ dominated by the σ-finite measure μ, 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 $\phi(x) \doteq 1$ is Bayes, while if $G(\Theta_1) = 0$ then $\phi(x) \doteq 0$ 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

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

c) Suppose that X is Normal $(\theta, 1)$. Find Bayes tests of $\text{H}_0 : \theta = 0$ vs $\text{H}_1 : \theta \neq 0$

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 Δ a point mass distribution at 0.

81. Consider the two distributions P_0 and P_1 on $\mathcal{X} = (0, 1)$ with densities wrt Lebesgue measure

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

a) Find a most powerful level $\alpha = .2$ test of $\text{H}_0 : X \sim P_0$ vs $\text{H}_1 : X \sim P_1$.

b) Plot both the 0-1 loss risk set $\mathcal{S}$ and the set $\mathcal{V} = \{\beta_\phi(\theta_0), \beta(\theta_1)\}$ 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?

82. 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.

x	-2	-1	0	1	2
$\theta = -1$	$\frac{\alpha}{2}$	$\frac{1}{2} - \alpha$	α	$\frac{1}{2} - \alpha$	$\frac{\alpha}{2}$
$\theta \neq -1$	θc	$\left(\frac{1-c}{1-\alpha}\right)\left(\frac{1}{2} - \alpha\right)$	$\left(\frac{1-c}{1-\alpha}\right)\alpha$	$\left(\frac{1-c}{1-\alpha}\right)\left(\frac{1}{2} - \alpha\right)$	$(1-\theta)\,c$

Find the size α likelihood ratio test of $\text{H}_0 : \theta = -1$ vs $\text{H}_1 : \theta \neq -1$. Show that the test $\phi(x) = I[x = 0]$ is of size α and is strictly more powerful that the LRT whatever be θ. (This simple example shows that LRTs need not necessarily be in any sense optimal.)