Stat 581 Supplementary Exercises

These exercises are derived from a variety of sources.

1 Measure theory and probability

1. Let $\mathcal{A}$ be an algebra on $\mathcal{X}$, closed under complements and finite unions. Show that $\mathcal{A}$ is closed under finite intersections.

2. Let $\mathcal{A}$ be an algebra that contains the set $(-\infty, 1/n]$ for $n \in \{1, 2, \ldots\}$. Is $(-\infty, 0] \in \mathcal{A}$ necessarily? If so, prove, otherwise, explain. Repeat the question for the case that $\mathcal{A}$ is a σ-algebra.

3. Let $\mathcal{C}$ be the collection of open intervals on the real line, and let $\mathcal{G}$ be the collection of open sets on the real line. Show that $\sigma(\mathcal{G}) = \sigma(\mathcal{C})$ (Hint: recall that $\sigma(\mathcal{G})$ is the *smallest* σ-algebra that contains the open sets).

4. Let $\mathcal{C}, \mathcal{G}$ be the open intervals and open sets, and let $\mathcal{D}, \mathcal{F}$ be the closed sets and closed intervals. Show that $\sigma(C) = \sigma(G) = \sigma(D) = \sigma(F)$.

5. Let $(\mathcal{X}, \mathcal{A}, \mu)$ be a measure space. Show that if $\{A_n\} \subset \mathcal{A}$, $A_n \subset A_{n+1}$, then $\mu(A_n) \uparrow \mu(\cup A_n)$.

6. Let $(\Omega, \mathcal{A}, \mu)$ be a measure space and let $f \in (m\mathcal{A})^+$, i.e. $f(\omega) > 0$ a.e. μ. Prove that $\int f d\mu > 0$.

7. Let $\mathcal{A}$ be a σ-algebra of Ω and let $X : \Omega \to \mathbb{R}$.

   (a) Show that the collection $\mathcal{B}$ of sets B for which $\{\omega : X(\omega) \in B\} \in \mathcal{A}$ is a σ-algebra on $\mathbb{R}$.

   (b) Now suppose $\{\omega : X(\omega) < x\} \in \mathcal{A}$ for every $x \in \mathbb{R}$. Show that X is $\mathcal{A}$-measurable.

      i. First, show $\{\omega : X(\omega) \in (a, b)\} \in \mathcal{A}$ for all $a < b$.

      ii. Second, show $\{\omega : X(\omega) \in G\} \in \mathcal{A}$ for open sets $G \subset \mathbb{R}$.

      iii. Third, show that $\mathcal{B}$ contains the Borel sets of $\mathbb{R}$.

8. Let P be a probability measure on $(\mathbb{R}, \mathcal{B})$ with CDF $F(x) = P((-\infty, x])$. Using the basic limit theorems on measures, show that F has a left limit ($\exists c : \lim_{\epsilon \uparrow 0} F(x+\epsilon) = c$) and is continuous on the right ($\lim_{\epsilon \downarrow 0} F(x + \epsilon) = F(x)$).

9. Let $Z \sim$ normal(0,1), $Y \sim$ binary(1/2) independent of Z, and $X = YZ$. Let P be the probability measure for X on $(\mathbb{R}, \mathcal{B})$. Find a measure μ that dominates P, and find the density of P with respect to μ.

10. Let $X_1, \ldots, X_n$ be i.i.d. "failure times" with pdf $p(x|\theta) = e^{-x/\theta}/\theta$ on $0 < x < \infty$. Suppose we only see the censored variables $Y_1, \ldots, Y_n$ where $Y_i = \min(X_i, t)$ and t is a known censoring time.

    (a) Calculate $\Pr(Y = t|\theta)$ and the CDF $F(y)$ for a single observation $Y = \min(X, t)$.

    (b) Explain why the class of distributions $\mathcal{P}_Y = \{\Pr(Y \in \cdot|\theta) : 0 < \theta < \infty\}$ for the Y_i's is not dominated by Lebesgue measure μ. Define a new measure ν which is a simple modification of μ that dominates $\mathcal{P}_Y$.

    (c) Find the corresponding Radon-Nikodym derivatives, i.e. the probability densities $p(y|\theta)$ for Y, such that $\Pr(Y \in A|\theta) = \int_A p(y|\theta)\nu(d\theta)$.

11. Let $(\Omega, \mathcal{F}, \Pr)$ be a probability space, $\mathcal{G}$ be a sub σ-algebra of $\mathcal{F}$, $X \in \mathrm{m}\mathcal{F}$ and $Z \in \mathrm{m}\mathcal{G}^+$, with Z being bounded. Show that $Z\mathrm{E}[X|\mathcal{G}]$ is a version of $\mathrm{E}[ZX|\mathcal{G}]$.

    (a) Show the result holds when Z is an indicator function.

    (b) Show the result holds when Z is a simple function.

    (c) Show the result holds when $Z \in \mathrm{m}\mathcal{G}^+$ and bounded.

    (*) Show the result holds when $\mathrm{E}[|ZX|] < \infty$.

12. Let $f(x, y)$ be the joint pdf of two continuous random variables X and Y, and let $f(x|y)$ be the conditional pdf of X "given $Y = y$" defined in the usual way. Show formally that conditional probabilities can be obtained from $f(x|y)$, i.e., that

$$\int_A f(x|Y)dx \;\text{ is a version of }\; \Pr(X \in A|\sigma(Y)) \equiv \mathrm{E}[1_A(X(\omega))|Y(\omega)].$$

13. Suppose $X \in m\mathcal{F}^+$ is a positive integrable random variable that is independent of a sub-σ-algebra $\mathcal{H}$. Show that $\mathrm{E}[X|\mathcal{H}] = \mathrm{E}[X]$:

    (a) Let X^* be any $\sigma(X)$-measurable simple function. Show that $\mathrm{E}[X^*|\mathcal{H}] = \mathrm{E}[X^*]$ and so $\int_H X^* dP = \mathrm{E}[X^*]\Pr(H)$ for $H \in \mathcal{H}$.

    (b) Now let $\{X_n\}$ be a sequence of $\sigma(X)$-measurable simple functions such that $X_n(\omega) \uparrow X(\omega)\ \forall \omega \in \Omega$. Use the appropriate limit theorem to show that $\mathrm{E}[X|\mathcal{H}] = \mathrm{E}[X]$.

(note that this result holds for general integrable random variables X).

14. Let $X_1, \ldots, X_n$ be i.i.d. random variables with $\mathrm{E}[|X_1|] < \infty$. Obtain $\mathrm{E}[X_1|\bar{X}]$ and confirm that it satisfies Kolmogorov's conditions. One way to proceed is as follows:
    (a) Show that for $G \in \sigma(\bar{X})$, then $X = (X_1, \ldots, X_n) \in G \Rightarrow X_\pi = (X_{\pi_1}, \ldots, X_{\pi_n}) \in G$ for any permutation π (hint: first show this for sets of the form $G = \{X : \bar{X} \in (a, b)\}$.
    (b) Show that $\mathrm{E}[X_1|\bar{X}] = \cdots = \mathrm{E}[X_n|\bar{X}]$.
    (c) Obtain $\mathrm{E}[\mathrm{X}_1|\bar{\mathrm{X}}]$.

2 Exponential families

1. The entropy of a random variable measures our inability to predict it. For a continuous random variable with density $p(y)$, the entropy is given by

$$-\int \ln p(y) p(y) dy,$$

which is the negative of the "average height" of the density.

   (a) Using the fact that $\ln x \leq x - 1 \; \forall x \geq 0$ show that

$$-q \ln q \leq -q \ln p + (p - q) \; \forall p \geq 0, q \geq 0.$$

   (b) Let $p_\theta(y) = \exp\{t(y) \cdot \theta - a(\theta)\}$ be a member of a continuous K-parameter regular exponential family with $\theta \in \Theta$, the natural parameter space. Let $q(y)$ be any other probability density such that $\mathrm{E}_q[t(y)] = \mathrm{E}_\theta[t(y)] \equiv \int t(y) p_\theta(y) \; dy$. Show that the entropy of $p_\theta(y)$ is at least as great as that of $q(y)$, i.e.

$$-\int q(y) \ln q(y) dy \leq -\int p_\theta(y) \ln p_\theta(y) dy.$$

2. Let $q(y)$ be a probability density, $t(y) : \mathcal{Y} \rightarrow \mathbb{R}^K$ and ψ_0 be the q-expectation of $t(y)$, i.e.

$$\int t(y) q(y) \; dy = \psi_0.$$

Suppose we are to sample data from q but model it as having come from a member of the exponential family $\{p_\theta(y) = \exp\{t(y) \cdot \theta - a(\theta\} : \theta \in \Theta\}$, of which q is not necessarily a member.

(a) Let

$$\tilde{l}(\theta) = \int \ln p_\theta(y) q(y) \; dy$$

be the q-expectation of the log-likelihood. Obtain a set of K equations which define the maximizer θ_0 of $\tilde{l}(\theta)$ in terms of derivatives of $a(\theta)$ and ψ_0. Also obtain $\psi_{\theta_0} = \mathrm{E}_{\theta_0}[t(y)]$, the expectation of $t(y)$ under $p_{\theta_0}(y)$, and compare it to ψ_0.

(b) Let $Y_1, \ldots, Y_n \sim$ i.i.d. q and let $l(\theta) = \frac{1}{n} \sum \ln p_\theta(y_i)$.

i. For a given θ, what will $l(\theta)$ converge to as $n \to \infty$?

ii. Let $\hat{\theta}$ be the maximizer of $l(\theta)$ and $\hat{\psi} = \mathrm{E}_{\hat{\theta}}[t(Y)]$. What is $\hat{\psi}$ converging to?

iii. What is $\hat{\theta}$ converging to ?

3. Consider the Kullback-Leibler loss:

$$L(\theta, d) = \int \log \frac{p(x|\theta)}{p(x|d)} p(x|\theta) \; \mu(dx),$$

which measures the predictive accuracy of $p(x|d)$ against the truth $p(x|\theta)$. Note that $L(\theta, d) \geq 0$ unless $p(x|\theta) = p(x|d)$ a.e. μ.

(a) Obtain an expression for this loss function for estimating the natural parameter in a (multiparameter) exponential family model in the natural parameter space, and show that this loss function is convex.

(b) Obtain expressions for the loss function when the model is

i. Poisson with unknown mean;

ii. normal with unknown mean and variance;

iii. gamma with unknown shape and scale.

4. Show that if $t(x)$ satisfies a linear constraint, then the exponential family generated by $t(x)$ is not identifiable.

5. Consider an exponential family $\mathcal{P} = \{p(x|\eta) = \exp(t(x) \cdot \eta - A(\eta)) : \eta \in \mathcal{H}\}$. Derive expressions for $\mathrm{E}[t(x)|\eta]$ and $\mathrm{Var}[t(x)|\eta]$ as a function of the derivatives of $A(\eta)$. Apply this result to obtain $\mathrm{E}[t(x)|\eta]$ and $\mathrm{Var}[t(x)|\eta]$ for the following families:

(a) the normal(μ, σ^2) family;

(b) the beta(a, b) family;

(c) the gamma(a, b) family, having mean a/b and variance a/b^2.

3 Decision problems

1. Suppose we want to estimate a parameter θ under a strictly convex loss function. Let X be the data and let $T = T(X)$ be a sufficient statistic that is not a 1-1 function of X. Show that any estimator $\hat{\theta}(X)$ that is a function of X and not of T is inadmissible. (Hint: For any estimator $\delta(X)$, find an estimator based on T that dominates it.) In your construction, where did you use the fact that T is sufficient, and not just any function of X?

2. Based on $\bar{X} \sim N(\theta, 1/n)$, suppose you need to decide among three actions: 1) stating nothing, 2) stating $\theta < 0$ or 3) stating $\theta > 0$. Refer to these decisions numerically as $d = 0$, $d = -1$ and $d = 1$, respectively, and let the loss be $L(\theta, d) = 1 - d \times \text{sign}(\theta)$.

   (a) Consider a decision rule of the form $\delta(\bar{x}) = \text{sign}(\bar{x}) \times 1(|\bar{x}| > c)$. Compute the risk as a function of θ and n, and plot the risk function for several values of n.

   (b) Consider a decision rule based on the z-test: If the test of $\theta = 0$ is rejected at level $\alpha = 0.05$, then take d to be the sign of $\bar{X}$. If the test doesn't reject, then take $d = 0$. Compute the risk function of this procedure as a function of n for several values of n, compare to the risk function in (a) and comment.

4 Admissibility

1. Obtain an example of a model and loss function for which there is a θ_0 such that $\delta(X) = \theta_0$ is not admissible.

2. Suppose we want to estimate θ under a strictly convex loss function. Let X be the data and let T_1, T_2 be any two sufficient statistics such that $T_2 = g(T_1)$ for some known function g. Let $\mathcal{C}_1, \mathcal{C}_2$ be the classes of estimators that are functions of T_1 and T_2, respectively.

   (a) Show that both $\mathcal{C}_1$ and $\mathcal{C}_2$ are complete classes.

   (b) Show that $\mathcal{C}_2 \subset \mathcal{C}_1$.

   (c) Show that if $\delta \in \mathcal{C}_1 \setminus \mathcal{C}_2$, then δ is inadmissible.

   (d) Based on the result, what sort of sufficient statistic should be used to construct an estimator?

3. Let $X \sim \text{binary}(\theta)$, $\theta \in \{\theta_l, \theta_h\}$ where $0 < \theta_l < \theta_h < 1$. Consider estimation of θ with 0-1 loss.

   (a) Characterize all non-randomized estimators of θ.

   (b) Characterize all estimators of θ, in terms of the non-randomized estimators, and draw the risk set.

   (c) Characterize all admissible estimators.

   (d) Identify the priors for which the corresponding Bayes estimators are admissible.

4. Let $X_j \sim N(\theta_j, 1),\ j = 1, 2$ and let $L((\theta_1, \theta_2), d) = (\theta_1 - d)^2$. Show that $\delta((X_1, X_2)) = \text{sign}(X_2)$ is an admissible procedure, and explain this counter-intuitive result.

5. Let $X \sim N(\theta, 1)$ and $L(\theta, d) = (\theta - d)^2$.

   (a) Show formally that $\delta(X) = \theta_0$ is an admissible estimator.

   (b) Consider a randomized estimator of the form $\delta(X, U) = \theta_0 \times 1(U < c) + \theta_1 \times 1(U > c)$, where $U \sim$ uniform(0,1). Decide whether or not δ is admissible, and prove your result.

6. Recall we proved that if $X \sim \text{normal}(\theta, 1)$ then X is admissible for θ under squared error loss. Using this fact, show the following:

   (a) If $X \sim$ normal (θ, σ_0^2), σ_0^2 known, then X is admissible for θ.

   (b) If $X \sim$ normal (θ, σ^2), σ^2 unknown, then X is admissible for θ.

   (c) If $X_1, \ldots, X_n \sim$ i.i.d. normal (θ, σ_0^2), σ_0^2 known, then $\bar{X}$ is admissible for θ.

   (d) If $X_1, \ldots, X_n \sim$ i.i.d. normal (θ, σ^2), σ^2 unknown, then $\bar{X}$ is admissible for θ.

7. Let $\theta \in (0, \infty)$ be an unknown parameter and X be a random variable such that $\text{E}[X|\theta] = \theta$ and $\text{Var}[X|\theta] = v(\theta)$ where $v(\theta)$ is specified. Consider estimation of θ by linear functions of the form

$$\delta_a(X) = aX$$

   for $a \in (0, 1)$, with squared-error loss

$$L(\delta(X), \theta) = [\delta(X) - \theta]^2.$$

   Let $\mathcal{A}$ be the set of all such estimators, indexed by $a \in (0, 1)$.

(a) For $v(\theta) = \theta^2$, calculate the risk function of δ_a, and find a value of a that makes δ_a admissible within the class $\mathcal{A}$ (i.e., no member of $\mathcal{A}$ dominates it). Show that this estimator dominates the unbiased estimator $\delta_1(X) = X$.

(b) For $v(\theta) = \theta$, prove that every member of $\mathcal{A}$ is admissible among the class $\mathcal{A}$ (i.e., no member dominates another), and that no member of $\mathcal{A}$ dominates δ_1.

(c) Suppose $v(\theta) = \theta^k$ where k is a positive integer. Find a closed form expression for the Bayes estimator in the class $\mathcal{A}$ when θ has prior density $p(\theta) = e^{-\theta},\ \theta > 0$.

8. Consider testing $H_0 : X \sim P_0$ versus $H_1 : X \sim P_1$, where P_0 and P_1 have densities $p_0(x)$ and $p_1(x)$ with respect to a common dominating measure μ on $\mathcal{X}$.

(a) Obtain the Bayes risk of a decision rule δ under the priors $(\pi_0, \pi_1) = (1, 0)$ and $(\pi_0, \pi_1) = (0, 1)$.

(b) For each prior, obtain the form of all Bayes rules.

(c) Identify which Bayes rules under these two priors are admissible.

5 Bayesian estimation

1. Let $X_1, \ldots, X_n \sim$ i.i.d. normal(θ, σ^2) where σ^2 is known. Using the prior distribution $\theta \sim$ normal$(\theta_0, \sigma^2/\kappa_0)$,

(a) Obtain the Bayes estimator under squared error loss.

(b) Compute the risk function of the Bayes estimator.

(c) Suppose the true population mean is μ. For what values of θ_0 and κ_0 will the Bayes estimator beat the sample mean, in terms of risk? Write your answer in terms of $(\mu - \theta_0)^2, \kappa_0, \sigma^2$ and n.

2. Let $X \sim N(\theta, 1)$, where $\theta \in \mathbb{R}$. Use a limiting Bayes argument to show that X is admissible under squared error loss (Hint: Use LC thm 7.8.7 on page 415, and note from the proof in class that it is sufficient for the condition to hold for all open sets of the form $\Theta_0 = \{\theta : |\theta - \theta_0| < r|\}$ for some $\theta_0 \in \Theta_0$ and $r > 0$) .

3. Let $\boldsymbol{X} = X_1, \ldots, X_n \sim$ i.i.d. normal(θ, σ^2). Consider the prior distribution $\pi(\theta, \sigma^2)$ where $\pi(\theta|\sigma^2)$ is the normal$(\theta_0, \sigma^2/\kappa_0)$ density and $\pi(\sigma^2)$ is the inverse-gamma density, so that $1/\sigma^2 \sim$ gamma$(a = \nu_0/2, b = \nu_0\sigma_0^2/2)$ and $1/\sigma^2$ has expectation $a/b = 1/\sigma_0^2$.

For this model and prior, obtain $\pi(\sigma^2|\theta, \boldsymbol{X})$, $\pi(\theta|\sigma^2, \boldsymbol{X})$, $\pi(\sigma^2|\boldsymbol{X})$ and $\pi(\theta|\boldsymbol{X})$. If you prefer, you can write everything in terms of the precision $\tau^2 = 1/\sigma^2$ instead.

4. Let $X_1, \ldots, X_n \sim$ i.i.d. Poisson(θ), and let $\theta \sim$ gamma(a, b) with expectation a/b. Suppose you observe $X_1, \ldots, X_n$ and want to predict the value of a future observation from this population. Obtain the form of the *predictive distribution* $p(X^*|x_1, \ldots, x_n)$ based on the Poisson model $X_1, \ldots, X_n, X^* \sim$ i.i.d. Poisson(θ) and the gamma(a, b) prior distribution. Note that the predictive distribution does not depend on any unknown parameters, meaning that you can actually use it to make predictions. Letting μ be the true population mean, what does the predictive distribution converge to as $n \to \infty$? Explain why this limiting distribution makes sense.

5. Let $\boldsymbol{X} \sim$ multinomial($n, \boldsymbol{\theta}$), where $\boldsymbol{\theta} \in \mathcal{S}_K$, the K-dimensional simplex, and $\boldsymbol{X}$ is a K-vector of counts that sum to n.

   (a) Find a class of conjugate prior distributions for $\{p(\boldsymbol{x}|\boldsymbol{\theta}) : \boldsymbol{\theta} \in \mathcal{S}_K\}$, and obtain the posterior expectation of $\mathrm{E}[\boldsymbol{\theta}|\boldsymbol{X}]$.

   (b) Obtain Jeffrey's default prior distribution for $\boldsymbol{\theta}$. Is it a proper probability distribution?

6. Let $X_1, X_2, \ldots$ be i.i.d. exponential(1) and suppose we observe $T = X_1 + \cdots + X_M$, where $M \geq 1$ is unknown. We want to estimate M under the loss function $L(M, \delta) = (M - \delta)^2/M$, i.e., if M is very large you can be off by more than if M is small. Suppose our prior distribution for M has density $\pi(m) = p(1-p)^{m-1}$.

   (a) Is T UMRUE for M?

   (b) Obtain the posterior distribution of M given T.

   (c) Show that the posterior expected loss $\mathrm{E}[L(M, \delta)|T]$ for any estimator $\delta(t)$ can be expressed as $1 + \lambda - 2\delta + (1 - \exp^{-\lambda})\delta^2/\lambda$, where $\lambda = (1-p)T$. From this, obtain the Bayes estimator.

   (d) Compare the (non-Bayesian) risk functions of the unbiased estimator T and the Bayes estimator.

7. A positive random variable X has density $p(x|\theta)$ where $\theta \in \Theta = \{0, 1\}$ and

$$\begin{aligned} p(x|0) &= e^{-x} \text{ for } x > 0 \\ p(x|1) &= e^{-(x-1)/2}/2 \text{ for } x > 1 \ . \end{aligned}$$

Suppose you have a prior distribution π for θ such that $\pi(0) = \gamma = 1 - \pi(1)$, and consider estimating θ with zero-one loss. Recall under this loss, a posterior mode is a Bayes estimator.

(a) Find the posterior distribution $\pi(\theta|x)$ and the Bayes estimator when $0 < \gamma < 1$.

(b) Find the Bayes estimator if $\gamma = 1$.

(c) Describe all Bayes estimators if $\gamma = 0$. Thinking intuitively, find the Bayes estimator in this set that has minimum (non-Bayes) risk. You don't have to show that this estimator has minimum risk, but you do need to identify the right estimator. Is this estimator Bayes with respect to another prior on θ, i.e. another value of γ? Is it admissible?

8. Consider Bayesian inference using a posterior density $\pi(\theta|x)$:

   (a) Find the form of the Bayes estimator under absolute loss $L(\theta, d) = |\theta - d|$, and prove your result.

   (b) Find the form of the Bayes estimator under zero-one loss $L(\theta, d) = 1(\theta \neq d)$ for the case that $\pi(\theta|x)$ is discrete.

9. Let $\boldsymbol{X} \sim N_p(\boldsymbol{\theta}, \mathbf{I})$. Try using Blyth's method with normal priors to show admissibility of $\boldsymbol{X}$ as an estimator of $\boldsymbol{\theta}$ under squared error loss. What goes wrong?

10. Based on $\bar{X} \sim N(\theta, 1/n)$, suppose you need to decide among three actions: 1) stating nothing, 2) stating $\theta < 0$ or 3) stating $\theta > 0$. Refer to these decisions numerically as $d = 0$, $d = -1$ and $d = 1$, respectively, and let the loss be $L(\theta, d) = 1 - d \times \text{sign}(\theta)$. Find the Bayes rule under a $N(0, \tau^2)$ prior for θ. Plot the (non-Bayes) risk function of the Bayes procedure, and compare it to the risk function of the "z-test" procedure detailed in Exercise 3.2.

11. Let $\boldsymbol{X} \sim P_\eta$ for some unknown value of $\boldsymbol{\eta}$, where P_η has density $p(\boldsymbol{x}|\boldsymbol{\eta}) = \exp(\boldsymbol{\eta} \cdot \boldsymbol{x} - A(\boldsymbol{\eta}))h(\boldsymbol{x})$. Show that the Bayes estimator under prior $\pi(\boldsymbol{\eta})$ and sum of squared error loss is given by

$$\hat{\eta}_j = \frac{\partial}{\partial x_j} \log \frac{p_\pi(\boldsymbol{x})}{h(\boldsymbol{x})}.$$

6 Shrinkage estimators

1. Let $X \sim p(x|\theta)$ for some $\theta \in \Theta$, and define $\mu = \mu(\theta) = \mathrm{E}[X|\theta] \in \mathbb{R}^p$. Show that if $M = \mu(\Theta)$ is convex and $\mu_0 \notin \bar{M}$, then $w\mu_0 + (1-w)X$ is not admissible for estimating μ under squared error loss.

2. Derive the version of Stein's lemma for multiparameter exponential families given by LC lemma 1.5.15 using Fubini's theorem.

3. Note that the risk advantage of $\boldsymbol{\delta}_{JS}$ over $\boldsymbol{X}$ is largest when $||\boldsymbol{\theta}||^2$ is small, i.e. $\boldsymbol{\theta} \approx \mathbf{0}$. Suppose instead you thought that $\boldsymbol{\theta} \approx \boldsymbol{\theta}_0$. Derive an alternative estimator that performs well when $\boldsymbol{\theta} \approx \boldsymbol{\theta}_0$, in that it beats $\boldsymbol{\delta}_{JS}$ at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ and also dominates $\boldsymbol{X}$ everywhere. Show that your estimator meets these latter two criteria.

4. Let $\boldsymbol{X} \sim N_p(\boldsymbol{\theta}, \mathbf{I})$. Note that $\boldsymbol{X}$ can be written as $\boldsymbol{X} = S\boldsymbol{\theta} + \boldsymbol{R}$ where $\boldsymbol{R} \cdot \boldsymbol{\theta} = 0$. Obtain the joint distribution of S and $\boldsymbol{R}$.

5. Consider the hierarchical model where $\boldsymbol{X} \sim N_p(\boldsymbol{\theta}, \mathbf{I})$ and $\boldsymbol{\theta} \sim N_p(\mathbf{0}, \tau^2\mathbf{I})$, and only $\boldsymbol{X}$ is observed.

   (a) For the case that τ^2 is known, obtain an estimator δ of $\boldsymbol{\theta}$ that minimizes the expected squared error

   $$\mathrm{E}[(\boldsymbol{\theta} - \delta)^2] = \int (\boldsymbol{\theta} - \delta(\boldsymbol{x}))^2 p(\boldsymbol{x}|\boldsymbol{\theta}) p(\boldsymbol{\theta}|\tau^2)\ d\boldsymbol{x} d\boldsymbol{\theta}.$$

   (b) For the case that τ^2 is unknown, find an unbiased estimator of $\tau^2/(\tau^2+1)$ based on $\boldsymbol{X}$.

6. For estimating $\boldsymbol{\theta}$ based on $\boldsymbol{X} \sim N_p(\boldsymbol{\theta}, \mathbf{I})$, consider the class of adaptive shrinkage estimators of the form $\delta_c(\boldsymbol{x}) = (1 - \frac{c}{||\boldsymbol{x}||^2})\boldsymbol{x}$.

   (a) Under squared error loss, find the simplest expression you can for the risk function of δ_c.

   (b) Now suppose we are in a hierarchical situation where $\boldsymbol{\theta} \sim N_p(\mathbf{0}, \tau^2\mathbf{I})$. Obtain a closed-form expression for the risk $R(\tau^2, \delta_c)$ of δ_c, where now risk means the loss averaged over $\boldsymbol{X}$ and $\boldsymbol{\theta}$ (this would be the Bayes risk, if the distribution for $\boldsymbol{\theta}$ were thought of as a prior). Find the value of $\tilde{c}$ of c that minimizes $R(\tau^2, \delta_c)$.

(c) For small, medium and large values of p, plot $R(\tau^2, \delta_{\tilde{c}})$ as a function of τ^2, along with $R(\tau^2, \frac{\tau^2}{1+\tau^2}\boldsymbol{X})$. Describe and interpret what you see.

7. For $\boldsymbol{X} \sim N_p(\boldsymbol{\theta}, \mathbf{I})$ show that $\mathrm{E}[||\boldsymbol{X}||^{-2}|\boldsymbol{\theta}] \leq \mathrm{E}[||\boldsymbol{X}||^{-2}|\mathbf{0}]$.

8. Let $X \sim N_p(\boldsymbol{\theta}, \sigma^2\mathbf{I})$ and $S/\sigma^2 \sim \chi^2_n$ be independent. For the case that σ^2 is unknown, consider the class of estimators of the form $\delta_c(\boldsymbol{x}, s) = (1 - \frac{cs}{\boldsymbol{x}\cdot\boldsymbol{x}})\boldsymbol{x}$.

   (a) Compute the risk function of δ_c. Your final answer will depend on $\mathrm{E}[(\boldsymbol{X}\cdot\boldsymbol{X})^{-1}]$ but no other uncalculated expectations.

   (b) Find the value of $\tilde{c}$ of c that minimizes the risk function.

   (c) Compare the risk function of δ_{JS} in the case $\sigma^2 = 1$ is known to the risk of $\delta_{\tilde{c}}$. How much is lost by not knowing σ^2?

7 Minimax estimation

1. Let $X \sim \text{binomial}(n, \theta)$. Obtain the minimax estimator of θ under loss $L(\theta, d) = (\theta - d)^2/[\theta(1-\theta)]$. Compute the risk of this estimator, and compare it to a plot of the risk function of the minimax estimator under squared error loss. Conversely, plot the risk function of this estimator under squared error loss, and compare it to the minimax estimator in that case.

2. An estimator δ is said to be ϵ-Bayes w.r.t. a prior π if $R(\pi, \delta) \leq R(\pi, \delta_\pi) + \epsilon$. An estimator is said to be extended Bayes if for each $\epsilon > 0$, there is a prior π_ϵ for which it is ϵ-Bayes. Show that if δ is constant risk and is extended Bayes, then it is minimax (this result is similar to LC Theorem 5.1.12)

3. Find a minimax estimator of $\theta = \mathrm{E}[X]$ in each of the following scenarios, showing all work that leads to your result.

   (a) $X \sim N(\theta, \sigma^2)$, σ^2 unknown, with loss $L(\theta, d) = (\theta - d)^2/\sigma^2$.

   (b) $X \sim \text{Poisson}(\theta)$ under loss $L(\theta, d) = (\theta - d)^2/\theta$.

   (c) $X \sim \text{binomial}(n, \theta)$ under loss $L(\theta, d) = (\theta - d)^2/[\theta(1-\theta)]$.

   Now generalize these results to the case that X has density $p(x|\psi) = h(x)\exp(\psi x - A(\psi))$ for some unknown parameter $\psi \in \Psi$, and where $\theta = \mathrm{E}[X|\psi]$. Specify any conditions on $\{p(x|a) : \psi \in \Psi\}$ you need.

4. Let $X_1, \ldots, X_p$ be independent random variables with with $\mathrm{E}[X_i] = \mu_i$ and $\mathrm{Var}[X_i] = \sigma_i^2 < M$, for some known $M < \infty$. Determine whether or not $\bar{X}$ is minimax for estimating $\bar{\mu} = \sum \mu_i / p$ under squared error loss.

5. The following to scenarios indicate some limitations of the minimax criteria:

   (a) Let $\mathcal{P} = \{P_\theta : \theta \in [0, 1]\}$ be some model, and consider estimation of θ based on the loss $L(\theta, d) = (1-\theta)d + \theta(1-d)$ and a sample $X \sim P_\theta$. Show that $\delta(x) = 1/2$ is minimax.

   (b) Let $X \sim$ binomial(n, θ) and consider estimating θ under the loss $L(\theta, d) = \min((\theta - d)^2/\theta^2, 2)$. Show that $\delta(x) = 0$ is the unique minimax estimator.

6. Consider estimating $\theta = \mu_y - \mu_x$ under squared error loss based on $X_1, \ldots, X_m \sim$ i.i.d. $N(\mu_x, \sigma_x^2)$ and $Y_1, \ldots, Y_n \sim$ i.i.d. $N(\mu_y, \sigma_y^2)$.

   (a) Find a minimax estimator for θ in the case that σ_x^2 and σ_y^2 are known.

   (b) Find a minimax estimator for θ in the case that σ_x^2 and σ_y^2 are known to be less some number C.

   (c) Do you think the minimax estimators from parts (a) and (b) are unique minimax? Why or why not?

8 Equivariance

1. Show that if $\{P_\theta : \theta \in \Theta\}$ is invariant under a group $\mathcal{G}$, then the induced transformation classes $\bar{\mathcal{G}}$ and $\tilde{\mathcal{G}}$ are also groups of bijections.

2. Let $X \sim N(\theta, 1)$, $\theta \in \mathbb{R}$, and and consider estimation of the sign of θ under the loss

$$L(\theta, d) = \begin{cases} 0 & \text{if } d = s(\theta) \\ 1 & \text{if } d = 0 \\ c & \text{if } d = -s(\theta), \end{cases}$$

   where $s()$ is the sign function.

   (a) Show the estimation problem is invariant under the transformation $g : x \to -x$. Characterize the class of equivariant estimators in terms of $|x|$ and $s(x)$.

(b) Obtain the Bayes estimator under the prior $\theta \sim N(0, \tau^2)$ and determine whether or not it is equivariant. Describe the estimator as $\tau^2 \to \infty$.

3. Let $\mathcal{P} = \{p(x|\theta) = e^{-x/\theta}/\theta : x > 0, \theta \in \Theta = \mathbb{R}^+\}$. Consider estimation of θ under the loss $L(\theta, d) = (1 - d/\theta)^2$ based on one observation $X \sim P_\theta$, $\theta \in \Theta$.

   (a) Characterize the class of estimators equivariant under the group of functions $\mathcal{G} = \{g : x \to cx, c > 0\}$.

   (b) Calculate the risk function of each such estimator.

   (c) Identify the equivariant estimator that uniformly minimizes the risk.

4. Let $\mathcal{P}$ be the class of all probability distributions on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ with continuous CDFs. Let $X_1, \dots, X_n$ be an i.i.d. sample from $P \in \mathcal{P}$, and consider estimation of the median via the loss $L(P, d) = (P(-\infty, d] - 1/2))^2$. Show that the estimation problem is invariant under the group $\mathcal{G}$ of transformations of the form $g_h(x_1, \dots, x_n) = (h(x_1), \dots, h(x_n))$, where h is a continuous strictly increasing function on $\mathbb{R}$. Identify the induced transformations $\bar{\mathcal{G}}$ and $\tilde{\mathcal{G}}$.

5. Let $\{\Theta, L, D\}$ be an invariant estimation problem where the induced group acts transitively and commutatively on Θ, and let δ_0 be an equivariant estimator. Prove that δ is equivariant iff

$$\delta(x) = \tilde{g}_x \delta_0(x)$$

   for some invariant function $\tilde{g}_x : \mathcal{X} \to \tilde{\mathcal{G}}$, i.e. $\tilde{g}_{gx} = \tilde{g}_x \ \forall g \in \mathcal{G}, x \in \mathcal{X}$.

6. Obtain the UMRE estimator for the vector location problem $\boldsymbol{X} \sim p_0(\boldsymbol{x} - \boldsymbol{\theta})$ under the group $\mathcal{G} = \{g : \boldsymbol{x} \to \boldsymbol{x} + a, a \in \mathbb{R}^p\}$ under squared error loss.

7. Consider a scale model $\mathcal{P} = \{p_\theta(x) = p_1(x/\theta)/\theta : \theta > 0\}$, where p_1 is a known probability density on $\mathbb{R}^+$.

   (a) Show that the model is invariant under the group $\mathcal{G} = \{g : x \to cx, c > 0\}$, and identify the induced group $\bar{\mathcal{G}}$ on the parameter space.

   (b) Show that the problem of estimating θ is invariant under $\mathcal{G}$ for loss functions of the form $L(\theta, d) = f(d/\theta)$, where f attains its minimum value of zero at $d = \theta$. Identify the induced group $\tilde{\mathcal{G}}$ on the decision space.

   (c) Characterize the class of equivariant estimators and the UMREE in terms of f and p_1.

(d) In the case $f(r) = (1-r)^2$, find the UMREE in terms of p_1.

(e) For the case $f(r) = (1-r)^2$, show that the risk of any equivariant estimator can be expressed as a posterior risk under a (possibly improper) prior distribution. Show that the UMREE can be viewed as a Bayes estimator under this prior.

8. Obtain the UMREE for the univariate location problem $X \sim p_0(x-\theta),\ \theta \in \mathbb{R}$, for the group $\mathcal{G} = \{g : x \to x + c, c \in \mathbb{R}\}$ and absolute loss $L(\theta, d) = |\theta - d|$.