

# Stats 210A, Fall 2024

## Homework 2

**Due on:** Wednesday, Sep. 18

**Instructions:** You may disregard measure-theoretic niceties about conditioning on measure-zero sets, almost-sure equality vs. actual equality, “all functions” vs. “all measurable functions,” etc. (unless the problem is explicitly asking about such issues).

**Problem 1** (Bayesian interpretation of sufficiency). Assume we have a family  $\mathcal{P}$  defined by densities  $p_\theta(x)$  with respect to a common measure  $\mu$  on  $\mathcal{X}$ , for  $\theta \in \Theta \subseteq \mathbb{R}^n$ . Additionally, assume the parameter  $\theta$  is itself random, following *prior density*  $q(\theta)$  with respect to the Lebesgue measure on  $\Theta$ .

Then, we can write the *posterior density* (distribution of  $\theta$  given  $X = x$ ) as

$$q_{\text{post}}(\theta | x) = \frac{p_\theta(x)q(\theta)}{\int_{\Theta} p_\zeta(x)q(\zeta) d\zeta}.$$

**Note:** this manipulation of the densities generally works even though we might worry about conditioning on a measure zero set. Feel free to make similar manipulations yourself in the problem.

- (a) Suppose a statistic  $T(X)$  has the property that, for any prior distribution  $q(\theta)$ , the posterior distribution  $q_{\text{post}}(\theta | x)$  depends on  $x$  only through  $T(x)$ . Show that  $T(X)$  is sufficient for  $\mathcal{P}$ .
- (b) Conversely, show that, if  $T(X)$  is sufficient for  $\mathcal{P}$  then, for any prior  $q$ , the posterior depends on  $x$  only through  $T(x)$ .

**Moral:** If we have a prior opinion about  $\theta$  in the form of a distribution, and then we update our opinion using Bayes’ rule after observing  $X$ , then we will naturally adhere to the sufficiency principle. This gives an alternative epistemological motivation for the principle. Another interpretation of the result of this exercise is that, in the Bayesian model,  $\theta$  is conditionally independent of  $X$  given  $T(X)$  if and only if  $T(X)$  is sufficient.

**Problem 2** (Uniform location-scale family). Let  $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \text{Unif}[\mu - \sigma, \mu + \sigma]$ , with  $\mu \in \mathbb{R}$  and  $\sigma > 0$  unknown.

- (a) Show that  $T(X) = (X_{(1)}, X_{(n)})$  is minimal sufficient.
- (b) Suppose that we wish to estimate  $\mu$  under the squared error loss. The sample mean  $\bar{X}$  may appear to be a reasonable estimator of  $\mu$ , but we might worry about the fact that it is not a function of  $T(X)$ .

Guided by the sufficiency principle, we could instead consider the estimator

$$\delta^*(X) = \frac{X_{(1)} + X_{(n)}}{2}.$$

Use Monte Carlo simulation to simulate the distribution of both estimators for  $\mu = 0, \sigma = 1, n = 1000$ . Report the root mean squared error (RMSE), defined as  $\sqrt{\text{MSE}}$  (as calculated by Monte Carlo).

For each estimator, plot a histogram of simulated estimates. You can either put both histograms on the same plot, or in side-by-side plots. If you do the latter, it might be more enlightening to use the same

range for the horizontal and vertical axes in both plots, so the differences between the histograms are more evident. Include your code.

- (c) In the previous part, we only compared the two estimators for one setting of the parameters  $\mu, \sigma$ . Show that, for both estimators, the MSE for generic  $(\mu, \sigma)$  does not depend on  $\mu$ , and is proportional to  $\sigma^2$ . That is, show that

$$\text{MSE}(\mu, \sigma; \delta) = \sigma^2 \text{MSE}(0, 1; \delta), \quad \text{for } \delta = \bar{X}, \delta^*.$$

Complete the argument to conclude that  $\bar{X}$  is inadmissible whenever  $\text{MSE}(0, 1; \delta^*) < \text{MSE}(0, 1; \bar{X})$ .

- (d) (**Optional:** Not graded, no extra points) In this and the next optional parts, you will calculate the MSE analytically using some nice manipulations.

If  $B \sim \text{Beta}(\alpha, \beta)$  then its density is proportional to  $x^{\alpha-1}(1-x)^{\beta-1}$  on  $x \in [0, 1]$ .

If  $U_1, \dots, U_n \stackrel{\text{i.i.d.}}{\sim} U[0, 1]$ , show that

$$U_{(n)} \sim \text{Beta}(n, 1), \quad \text{which has density } p(x) = nx^{n-1},$$

and

$$U_{(1)}/U_{(n)} \sim \text{Beta}(1, n-1) \quad \text{which has density } p(x) = (n-1)(1-x)^{n-2},$$

independently of  $U_{(n)}$ .

**Hint:** For the first part, start by writing down the CDF of  $U_{(n)}$ .

**Hint:** For the second part, you may use without proof the fact that, conditional on  $U_{(n)} = u$ , the remaining  $n-1$  values are i.i.d.  $\text{Unif}[0, u]$ , then proceed similarly to what you did for the first part.

- (e) (**Optional:** Not graded, no extra points) Compute the MSE of each estimator as a function of  $n, \mu$ , and  $\sigma$ , and show that  $\delta^*$  strictly dominates  $\bar{X}$  for  $n > 2$  (the estimators coincide for  $n = 2$ ). What happens to the ratio of their MSE's as  $n \rightarrow \infty$ ?

**Hint:** The results from the previous part should be useful. You may use without proof that  $\text{Beta}(\alpha, \beta)$  has mean  $\frac{\alpha}{\alpha+\beta}$  and variance  $\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$ .

**Moral:** Understanding and respecting the statistical structure of a model sometimes helps us to come up with estimators that perform dramatically better than the estimator we would have naïvely thought of. Here is a case where applying the sufficiency principle helped us get a much better estimator than the sample mean.

**Problem 3** (Convexity of  $A(\eta)$  and  $\Xi_1$ ). Let  $\mathcal{P} = \{p_\eta : \eta \in \Xi_1\}$  denote an  $s$ -parameter exponential family in canonical form

$$p_\eta(x) = e^{\eta^T T(x) - A(\eta)} h(x), \quad A(\eta) = \log \int_{\mathcal{X}} e^{\eta^T T(x)} h(x) d\mu(x),$$

where  $\Xi_1 = \{\eta : A(\eta) < \infty\}$  is the natural parameter space.

Recall Hölder's inequality: if  $q_1, q_2 \geq 1$  with  $q_1^{-1} + q_2^{-1} = 1$ , and  $f_1$  and  $f_2$  are ( $\mu$ -measurable) functions from  $\mathcal{X}$  to  $\mathbb{R}$ , then

$$\|f_1 f_2\|_{L^1(\mu)} \leq \|f_1\|_{L^{q_1}(\mu)} \|f_2\|_{L^{q_2}(\mu)}, \quad \text{where } \|f\|_{L^q(\mu)} = \left( \int_{\mathcal{X}} |f(x)|^q d\mu(x) \right)^{1/q}.$$

**Note:**  $q_1 = q_2 = 2$  reduces to Cauchy-Schwarz.

- (a) Show that  $A(\eta) : \mathbb{R}^s \rightarrow [0, \infty]$  is a convex function: that is, for any  $\eta_1, \eta_2 \in \mathbb{R}^s$  (not just in  $\Xi_1$ ), and  $c \in [0, 1]$  then

$$A(c\eta_1 + (1-c)\eta_2) \leq cA(\eta_1) + (1-c)A(\eta_2) \tag{1}$$

**Hint:** try  $q_1 = c^{-1}$ ,  $f_1(x)^{1/c} = e^{\eta_1^T T(x)} h(x)$ .

- (b) Use the previous part to show that  $\Xi_1 \subseteq \mathbb{R}^s$  is convex.

**Moral:** The natural parameter space for any exponential family (meaning the set of all parameters  $\eta$  that give normalizable densities) is a convex subset of  $\mathbb{R}^s$ .

**Problem 4** (Expectation of an increasing function). (a) Assume  $X \sim P$  is a real-valued random variable. Show that if  $f(x)$  and  $g(x)$  are non-decreasing functions of  $x$ , then

$$\text{Cov}(f(X), g(X)) \geq 0$$

**Hint:** first show  $\mathbb{E}[(f(X_1) - f(X_2))(g(X_1) - g(X_2))] = 2\text{Cov}(f(X_1), g(X_1))$ , where  $X_1, X_2 \stackrel{\text{i.i.d.}}{\sim} P$ .

(b) Let  $p_\eta(x)$  be a one-parameter canonical exponential family with non-decreasing sufficient statistic  $T(x)$ , where  $x \in \mathcal{X} \subseteq \mathbb{R}$ :

$$p_\eta(x) = e^{\eta T(x) - A(\eta)} h(x).$$

Let  $\psi(x)$  be any non-decreasing bounded function. Show that, for  $\eta \in \Xi_1^\circ$ , the interior of the natural parameter space,  $\frac{d}{d\eta} \mathbb{E}_\eta[\psi(X)] \geq 0$ .

**Hint:** find an expression for  $\frac{d}{d\eta} \mathbb{E}_\eta[\psi(X)]$  by using methods akin to the ones we used in class to derive the differential identities. You may assume it is justified to differentiate under the integral sign.

(c) Conclude that  $X$  is stochastically increasing in  $\eta$ ; that is, show  $\mathbb{P}_\eta(X \leq c)$  is non-increasing in  $\eta$ , for every  $c \in \mathbb{R}$ .

**Moral:** This exercise confirms something that we should intuitively expect to be true: that increasing the natural parameter  $\eta$ , which “tilts” the distribution toward larger values of  $T(X)$ , will also shift the distribution of  $X$  to the right if  $T$  is an increasing function. It also illustrates the usefulness of differential identities for understanding exponential families’ structure.

**Problem 5** (Mean parameterization of an exponential family). Consider the  $s$ -parameter exponential family  $\mathcal{P} = \{P_\eta : \eta \in \Xi\}$  on  $\mathcal{X}$  with densities  $p_\eta(x) = e^{\eta' T(x) - A(\eta)} h(x)$  with respect to a common dominating measure  $\nu$ . Assume  $\Xi = \Xi_1^\circ$ , the interior of the full natural parameter space, and that  $\text{Var}_\eta(a' T(X)) > 0$  for all  $a \neq 0$  and  $\eta \in \Xi$ .

Define the *mean parameter*

$$\mu(\eta) = \mathbb{E}_\eta[T(X)].$$

We will show that this is a one-to-one mapping, so  $\mathcal{P}$  can be alternatively be parameterized by  $\mu(\eta)$  instead of  $\eta$ . The Bernoulli, Poisson, and exponential distributions are exponential families that are most often parameterized by their means, and parameterizations of other distributions like the normal and binomial are closely related to the mean parameterization.

Throughout this problem, you may use without proof that if the variance of any statistic  $S(X)$  is positive under one  $P_\eta \in \mathcal{P}$  then it is positive under all  $P_\eta \in \mathcal{P}$  (as an optional exercise, try to prove this).

(a) For  $s = 1$ , show that  $\eta \mapsto \mathbb{E}_\eta[T(X)]$  is a one-to-one mapping; that is, show that if  $\eta_1 \neq \eta_2$  then  $\mathbb{E}_{\eta_1}[T(X)] \neq \mathbb{E}_{\eta_2}[T(X)]$ .

**Hint:** You can use the differential identities.

(b) For  $s > 1$  and  $\eta_1, \eta_2 \in \Xi$ , consider the subfamily whose parameter space is the line segment between  $\eta_1$  and  $\eta_2$ . For  $\theta \in [0, 1]$ , let

$$\eta(\theta) = (1 - \theta)\eta_1 + \theta\eta_2.$$

Show that this subfamily is a one-parameter exponential family on  $\mathcal{X}$  with natural parameter  $\theta$ , and write it in standard exponential family form.

(c) Combine (a) and (b) to show that  $\eta \mapsto \mathbb{E}_\eta[T(X)]$  is a one-to-one mapping for  $s \geq 1$ .

**Moral:** We can always parameterize an exponential family by the mean of  $T(X)$ . This parameterization is very often more intuitive than the natural parameterization, and it gives us the standard parameterization for families like the Poisson (with sufficient statistic  $T(X) = X$ ) and binomial (with sufficient statistic  $T(X) = X/n$ ).