

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 μ on \mathcal{X} , for $\theta \in \Theta \subseteq \mathbb{R}^n$. Additionally, assume the parameter θ is itself random, following *prior density* $q(\theta)$ with respect to the Lebesgue measure on Θ .

Then, we can write the *posterior density* (distribution of θ 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 θ 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, θ is conditionally independent of X given $T(X)$ if and only if $T(X)$ is sufficient.

Problem 2 (Convexity of $A(\eta)$ and Ξ_1). Let $\mathcal{P} = \{p_\eta : \eta \in \Xi_1\}$ denote an s -parameter exponential family in canonical form

$$p_\eta(x) = e^{\eta' T(x) - A(\eta)} h(x), \quad A(\eta) = \log \int_{\mathcal{X}} e^{\eta' T(x)} h(x) d\mu(x),$$

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 (μ -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 Ξ_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 $q_1 = c^{-1}$, $f_1(x)^{1/c} = e^{\eta_1' T(x)} h(x)$.

(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 η that give normalizable densities) is a convex subset of \mathbb{R}^s .

Problem 3 (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 η ; that is, show $\mathbb{P}_\eta(X \leq c)$ is non-increasing in η , for every $c \in \mathbb{R}$.

Moral: This exercise confirms something that we should intuitively expect to be true: that increasing the natural parameter η , 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 4 (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 ν . 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 η . 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 η_1 and η_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 θ , 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 $T(X) = X/n$).

Problem 5 (Multinomial family). The multinomial family is a multi-category version of the binomial, it measures the number of times each category comes up if we sample a d -category random variable with distribution π on n independent trials. Throughout this problem assume $d \geq 3$.

If $X \sim \text{Multinom}(n, \pi)$, with all $\pi_j > 0$ and $\sum_j \pi_j = 1$, then X has density

$$p_\pi(x) = \pi_1^{x_1} \pi_2^{x_2} \cdots \pi_d^{x_d} \cdot \frac{n!}{x_1! x_2! \cdots x_d!}$$

Note: The coordinates of $X = (X_1, \dots, X_d)$ are *not* i.i.d. samples; each one corresponds to a different bin and X_1 is not independent of X_2 .

- (a) Rewrite the densities as a $(d - 1)$ -parameter exponential family, giving an explicit form for $T(x)$, $h(x)$, η , and $A(\eta)$. Is $X = (X_1, \dots, X_d)$ minimal sufficient?
- (b) Suppose a certain gene has two alleles **A** and **a**, and $\theta \in (0, 1)$ is the unknown prevalence of allele **a** in a well-mixed population. Then the proportion of individuals in the population with genotypes **aa**, **Aa**, and **AA** is θ^2 , $2\theta(1 - \theta)$, and $(1 - \theta)^2$, respectively.

We can estimate θ by sampling n independent individuals from the population and counting the number who have each genotype. These counts will have a joint multinomial distribution with probability parameter

$$\pi(\theta) = (\theta^2, 2\theta(1 - \theta), (1 - \theta)^2).$$

Hence, scientific considerations might lead us to use the multinomial subfamily indexed by θ :

$$\mathcal{P} = \{\text{Multinom}(n, \pi(\theta)) : \theta \in (0, 1)\}.$$

Can \mathcal{P} be written as a one-parameter exponential family? Find a minimal sufficient statistic for \mathcal{P} .

Moral: The “true” (minimal) dimension of an exponential family is not always the dimension in which it is initially given to us.