UC Berkeley
Department of Statistics
Fall 2016

STAT 210A: Introduction to Mathematical Statistics

Problem Set 4

Student: Ivana Malenica **Due:** Thursday, Oct. 6

Note: All measure-theoretic niceties about conditioning on measure-zero sets, almost-sure equality vs. actual equality, "all functions" vs. "all measurable functions," etc. are disregarded for the time being.

1. Bayesian law of large numbers

1. Let $p(x)$ and $q(x)$ denote two strictly positive densities with respect to a common dominating measure μ. The *Kullback–Leibler divergence* between p and q is defined as

$$D(p\|q) = \int_{\mathcal{X}} p(x) \log \frac{p(x)}{q(x)}\, d\mu(x).$$

Show that $D(p\|q) \geq 0$, with equality only in the case that $p(X) = q(X)$ almost surely (Hint: use Jensen's inequality; you may need a more general statement of Jensen's inequality than what is in Keener).

2. Consider a dominated likelihood model $\mathcal{P} = \{p_\theta(x) : \theta \in \Omega\}$, where the parameter space Ω is a finite set, and the densities are strictly positive on $\mathcal{X}$. Let λ denote a prior density w.r.t. the counting measure on Ω, and consider the Bayes posterior after observing a sample $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} p_{\theta_0}(x)$ for some *fixed* value θ_0 (that is, we are studying *frequentist* properties of the *Bayesian* posterior distribution).
If the prior λ puts positive mass on all values in Ω, show that as $n \to \infty$ with the true value θ_0 fixed, the posterior density eventually concentrates nearly all its mass on the true value θ_0. That is,

$$\mathbb{P}_{\theta_0}\left[\lambda(\theta_0 \mid X_1, \ldots, X_n) \geq 1 - \varepsilon\right] \to 1, \quad \text{for all } \varepsilon > 0.$$

(Hint: use the law of large numbers).

(Note: the requirement that the prior density should be nonzero everywhere is sometimes called Cromwell's Rule, after Oliver Cromwell's quotable plea to the Church of Scotland: "I beseech you, in the bowels of Christ, think it possible that you may be mistaken.")

2. Bayesian prediction

Consider a Bayesian model in which the prior distribution for Θ is uniform on $(0, 1)$ and given $\Theta = \theta$, $X_i, i \geq 1$, are i.i.d. Bernoulli with success probability θ. Find

$$\mathbb{P}(X_{n+1} = 1 \mid X_1, \ldots, X_n).$$

(Note: here the probability should be read as being taken over the joint distribution of $\Theta, X_1, X_2, \ldots$)

3. Ridge regression

Consider the i.i.d. linear observation model

$$Y_i \;=\; x_i'\beta + W_i, \qquad i = 1, \ldots n \tag{1}$$

where $\beta \in \mathbb{R}^d$, the design vectors $x_i \in \mathbb{R}^d$ are fixed and known, and $W_i \sim N(0, \sigma^2)$ is observation noise. Assume that $d < n$, and the design matrix $\mathbf{X}$ (the $n \times d$ matrix whose ith row is x_i') has full column rank.

(a) Assume that $\sigma^2 > 0$ is known, and that β is modeled as a fixed but unknown vector. Find the UMVU estimate of β based on Y.

(b) Now consider Bayesian estimation with the prior $\beta \sim N(\mu, \sigma^2 Q)$, where $Q \in \mathbb{R}^{d \times d}$ is a known, positive definite symmetric matrix. Find the posterior mean of β.

4. Absolute error loss

For a Bayesian model with a single real parameter Θ, assume that the posterior distribution of Θ given $X = x$ is absolutely continuous for all x. What is the Bayes estimator for the loss $L(\theta, d) = |\theta - d|$?

5. Exponential-exponential model

Consider a Bayesian model in which the prior distribution for Θ is $\lambda(\theta) = e^{-\theta} 1\{\theta > 0\}$ (the standard exponential distribution) and the density for X given $\Theta = \theta$ is

$$p_\theta(x) = e^{\theta - x} 1\{x > \theta\}.$$

(a) Find the marginal density for X, and the marginal expectation $\mathbb{E}[X]$.

(b) Find the Bayes estimator for θ under squared error loss. (Assume $X > 0$.)

6. Exponential families

This problem addresses the issue of implementing Bayes estimators for an s-parameter exponential family model in canonical form:

$$p_\theta(x) = e^{\theta' T(x) - A(\theta)} h(x)$$

where $x = (x_1, \ldots, x_n)$ and the random vector Θ has prior density $\lambda(\cdot)$.

(a) If $q(x)$ denotes the marginal density of X, show that for $i = 1, \ldots, n$, we have

$$\mathbb{E}\left[\sum_{j=1}^{s} \Theta_j \frac{\partial T_j(x)}{\partial x_i} \mid X = x\right] = \frac{\partial}{\partial x_i} \log q(x) - \frac{\partial}{\partial x_i} \log h(x).$$

(Assume here that all relevant quantities are suitably differentiable.)

(b) If $T(x) = x$, use part (a) to conclude that the posterior mean of Θ is given by

$$\nabla \log q(x) - \nabla \log h(x).$$

7. Jeffreys prior

For each distribution and each parameter, find the Jeffreys prior (possibly improper).

(a) Poisson distribution $X \sim \text{Pois}(\theta) = \frac{\theta^x e^{-\theta}}{x!}$, parameterized by θ and $\eta = \log \theta$.

(b) Normal distribution $X \sim N(\mu, \sigma^2)$, parameterized by μ (with σ^2 known) and by σ^2 (with μ known).