STAT547C: Topics in Probability **Winter Term 1 (Fall) 2020-21**

Assignment 4

Due date: Dec. 3 by 11:59 pm. *Instructor: Benjamin Bloem-Reddy*

Instructions

Academic integrity policy: I encourage you to discuss verbally with other students about the assignment. However, you should write your answers by yourself. For example, copying (either manually or electronically) part of a function or a LaTeX equation is not permitted; and if you use online resources, you must cite them. If you discuss the assignment with anyone (a classmate or anyone else), you must say so at the top of your solutions. Also, refrain from looking at answer keys from other schools, previous years, or the Math Stack Exchange. For more information, see:
`http://learningcommons.ubc.ca/guide-to-academic-integrity/`

1. *Justify your answers formally.*
2. Submit your work before class on the due date to Gradescope (.pdf) and Canvas (.tex).

Throughout this assignment, let $(\Omega, \mathcal{H}, \mathbb{P})$ be a (background) probability space.

1 Conditional densities

A topic we did not cover in lecture is **conditional densities**. (You may find pp. 155-156 of Çinlar helpful.) Suppose that the joint distribution π of X and Y (random variables taking values in $(D, \mathcal{D})$ and $(E, \mathcal{E})$, respectively) has the form

$$\pi(dx, dy) = \mu_0(dx)\nu_0(dy)p(x, y) , \quad x \in D,\ y \in E ,$$

where μ_0 and ν_0 are σ-finite measures and p is a positive function that belongs to $\mathcal{D} \otimes \mathcal{E}$. (Often, $D = E = \mathbb{R}^D$ and $\mu_0 = \nu_0 =$Lebesgue.) This π can be put in the form (9.20) from the lecture notes:

$$\pi(dx, dy) = \mu(dx)K(x, dy) = [\mu_0(dx)m(x)][\nu_0(dy)k(x, y)] , \tag{1}$$

with

$$m(x) = \int_E \nu_0(dy)p(x, y) , \quad k(x, y) = \begin{cases} p(x, y)/m(x) , & m(x) > 0 , \\ \int_D \mu_0(dx')p(x', y) , & m(x) = 0 . \end{cases} \tag{2}$$

Then the function $y \mapsto k(x, y)$ is called the **conditional density** (with respect to ν_0) of Y given that $X = x$.

To illustrate, consider independent gamma random variables $Y \overset{d}{=} \gamma_{a,c}$ and $Z \overset{d}{=} \gamma_{b,c}$. Let $X = Y + Z$.

(a) What is the joint distribution of X and Y, $\pi(dx, dy)$?
Hint: To start, construct the joint distribution of X, Y, Z, making use of the Dirac measure.
[10 mark(s)]

(b) What is the distribution of X? What is its density with respect to the Lebesgue measure?

[10 mark(s)]

(c) Let μ denote the distribution of X. The conditional density of Y given $X = x$ is obtained from the kernel

$$K(x, dy) = \frac{\pi(dx, dy)}{\mu(dx)} \ . \tag{3}$$

What is K? What is k? In particular, what is K when $a = b = 1$ (corresponding to exponential random variables)?

[10 mark(s)]

Question total: [30 mark(s)]

2 Sufficiency & co.

For this question, we will work directly on $(\Omega, \mathcal{H})$, which we assume to be a standard measurable space. Working directly on $(\Omega, \mathcal{H})$ is purely for convenience, and everything can be transferred to some other (standard) measurable space. Let $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ be a family of probability measures on $(\Omega, \mathcal{H})$ indexed by Θ (i.e., it is a statistical model). The family $\mathcal{P}$ is **dominated** by a measure ν if $P_\theta \ll \nu$ for each $P_\theta \in \mathcal{P}$.

a) **Sufficient σ-algebras.** We talked briefly in class about sufficient statistics. The measure-theoretic definition of sufficiency deals with sufficient σ-algebras. Specifically, a sub-σ-algebra $\mathcal{F} \subset \mathcal{H}$ is sufficient for $\mathcal{P}$ if there is a transition probability kernel K from $(\Omega, \mathcal{F})$ into $(\Omega, \mathcal{H})$ such that $K(\cdot, A)$ is a (regular) version of $P_\theta[A \mid \mathcal{F}]$ for all $A \in \mathcal{H}$, $\theta \in \Theta$ (that is, $K(\omega, A) = P_\theta[A \mid \mathcal{F}](\omega)$ for each $\omega \in \Omega$).

Suppose that $\mathcal{P}$ is dominated by a probability measure P_* and that each P_θ has a density f_θ with respect to P_*. Furthermore, assume that f_θ belongs to $\mathcal{F}$. Show that $P_*[A \mid \mathcal{F}]$ is a version of $P_\theta[A \mid \mathcal{F}]$ for each $A \in \mathcal{H}$ and $\theta \in \Theta$, and therefore $\mathcal{F}$ is sufficient for $\mathcal{P}$.

[15 mark(s)]

b) Let $(D, \mathcal{D})$ be a measurable space, and $S : \Omega \to D$ be $\mathcal{H}/\mathcal{D}$-measurable. Define

$$K(x, B) = \delta_{S(x)}(B) \ , \quad x \in \Omega, \ B \in \mathcal{D} \ . \tag{4}$$

Show (by direct computation) that, using the notation of Theorem 9.7 in the lecture notes,

i) $Kf = f \circ S$ for $f : D \to \mathbb{R}_+$ belonging to $\mathcal{D}_+$.

ii) $\mu K = \mu \circ S^{-1}$ is a measure on $(D, \mathcal{D})$ for every measure μ on $(\Omega, \mathcal{H})$.

iii) $\mu K f = \mu(f \circ S)$ for every measure μ on $(\Omega, \mathcal{H})$ and f belonging to $\mathcal{D}_+$.

[15 mark(s)]

c) Returning to part a) above, since we're working directly on $(\Omega, \mathcal{H})$, our random variable of interest is $X(\omega) = \omega$. As defined in part b), $S : \Omega \to D$ is a statistic of X, and the properties in part b) hold for any statistic. We're interested in sufficient statistics. For simplicity, let $(D, \mathcal{D}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and assume that S is surjective. (We can always assume that D is the image of S so there is no loss of generality.)

Let $\mathcal{F} = \sigma S$, and assume that the model $\mathcal{P}$ is dominated by P_*, and that each P_θ has density with respect to P_*

$$p_\theta(x) = f_\theta(x)h(x) \ , \quad x \in \Omega, \quad \text{with } f_\theta \text{ belonging to } \sigma S. \tag{5}$$

For $s \in \mathbb{R}$, let $S^{-1}\{s\} = \{x \in E : S(x) = s\}$.

Since $(\Omega, \mathcal{H})$ is standard, we know that a regular version of the conditional distribution of X given σS exists for each P_θ and for P_*. Call these $K_\theta(s, dx)$ and $K_*(s, dx)$, respectively.

Show that for all $\theta \in \Theta$,

$$K_\theta(s, dx) = \frac{K_*(s, dx)h(x)}{\int_\Omega K_*(s, dx')h(x')} = \frac{K_*(s, dx)h(x)}{\int_{S^{-1}\{s\}} K_*(s, dx')h(x')} \ , \tag{6}$$

which shows that σS is sufficient.

[15 mark(s)]

Question total: [45 mark(s)]

3 The Merit of Information (Optional)

Let $\mathcal{G} \subset \mathcal{F}$ and $\mathbb{E}[X]^2 < \infty$.

1. Show that

$$\mathbb{E}(\{X - \mathbb{E}(X|\mathcal{F})\}^2) + \mathbb{E}(\{\mathbb{E}(X|\mathcal{F}) - \mathbb{E}(X|\mathcal{G})\}^2) = \mathbb{E}(\{X - \mathbb{E}(X|\mathcal{G})\}^2) \ .$$

Dropping the term $\mathbb{E}(\{\mathbb{E}(X|\mathcal{F}) - \mathbb{E}(X|\mathcal{G})\}^2)$ on the LHS, we can interpret the resulting inequality as saying that more information means a smaller mean square error.

[15 mark(s)]

2. Argue that $\mathbb{E}(\{X - \mathbb{E}(X|\mathcal{F})\}^2) = \mathbb{E}(\{X - \mathbb{E}(X|\mathcal{G})\}^2)$ if one of the following is true:

   - $X = a$ almost surely, for some $a \in \mathbb{R}$.
   - X is $\mathcal{G}$-measurable.
   - $\mathcal{F} = \mathcal{G}$.

[10 mark(s)]

Question total: [25 mark(s)]

Assignment total: [100 mark(s)]