2022Fall Probability & Mathematical Statistics	2022/11/10

Homework 7

Professor: Ziyu Shao — *Due: 2022/11/20 10:59pm*

1. (a) Show the proof of general Bayes' Rule (four cases).

	Y **discrete**	Y **continuous**
X **discrete**	$P(Y=y\|X=x)=\frac{P(X=x\|Y=y)P(Y=y)}{P(X=x)}$	$f_Y(y\|X=x)=\frac{P(X=x\|Y=y)f_Y(y)}{P(X=x)}$
X **continuous**	$P(Y=y\|X=x)=\frac{f_X(x\|Y=y)P(Y=y)}{f_X(x)}$	$f_{Y\|X}(y\|x)=\frac{f_{X\|Y}(x\|y)f_Y(y)}{f_X(x)}$

   (b) Show the proof of general LOTP (four cases).

	Y **discrete**	Y **continuous**
X **discrete**	$P(X=x)=\sum_y P(X=x\|Y=y)P(Y=y)$	$P(X=x)=\int_{-\infty}^{\infty} P(X=x\|Y=y)f_Y(y)dy$
X **continuous**	$f_X(x)=\sum_y f_X(x\|Y=y)P(Y=y)$	$f_X(x)=\int_{-\infty}^{\infty} f_{X\|Y}(x\|y)f_Y(y)dy$

2. (a) The Cauchy distribution has PDF

$$f(x)=\frac{1}{\pi\left(1+x^2\right)}$$

   for all x. Find the CDF of a random variable with the Cauchy PDF. Hint: Recall that the derivative of the inverse tangent function $\tan^{-1}(x)$ is $\frac{1}{1+x^2}$.

   (b) The Pareto distribution with parameter $a>0$ has PDF

$$f(x)=\frac{a}{x^{a+1}}$$

   for $x\geq 1$ (and 0 otherwise). This distribution is often used in statistical modeling. Find the CDF of a Pareto r.v. with parameter a; check that it is a valid CDF.

   (c) Let $Z\sim\mathcal{N}(0,1)$, and c be a nonnegative constant. Find $E[\max(Z-c,0)]$, in terms of the standard Normal CDF Φ and PDF φ.

3. The Exponential is the analog of the Geometric in continuous time. This problem explores the connection between Exponential and Geometric in more detail, asking what happens to a Geometric in a limit where the Bernoulli trials are performed faster and faster but with smaller and smaller success probabilities.

   Suppose that Bernoulli trials are being performed in continuous time; rather than only thinking about first trial, second trial, etc., imagine that the trials take place at points on a timeline. Assume that the trials are at regularly spaced times 0, Δt, $2\Delta t$, $\ldots$, where Δt is a small positive number. Let the probability of success of each trial be $\lambda\Delta t$, where λ is a positive constant. Let G be the number of failures before the first success (in discrete time), and T be the time of the first success (in continuous time).

   (a) Find a simple equation relating G to T. Hint: Draw a timeline and try out a simple example.

   (b) Find the CDF of T. Hint: First find $P(T > t)$.

   (c) Show that as $\Delta t \to 0$, the CDF of T converges to the Expo(λ) CDF, evaluating all the CDFs at a fixed $t \geq 0$.

4. The Laplace distribution has PDF

$$f(x) = \frac{1}{2}e^{-|x|}$$

   for all real x. The Laplace distribution is also called a symmetrized Exponential distribution. Explain this in the following two ways.

   (a) Plot the PDFs and explain how they relate.

   (b) Let $X \sim$ Expo(1) and S be a random sign (1 or -1, with equal probabilities), with S and X independent. Find the PDF of $S \cdot X$ (by first finding the CDF), and compare the PDF of $S \cdot X$ and the Laplace PDF.

5. This problem explores a visual interpretation of covariance. Data are collected for $n \geq 2$ individuals, where for each individual two variables are measured (e.g., height and weight). Assume independence across individuals (e.g., person 1's variables give no information about the other people), but not within individuals (e.g., a person's height and weight may be correlated).

   Let $(x_1, y_1), \ldots, (x_n, y_n)$ be the n data points. The data are considered here as fixed, known numbers – they are the observed values after performing an experiment. Imagine plotting all the points (x_i, y_i) in the plane, and drawing the rectangle determined by each pair of points. For example, the points $(1, 3)$ and $(4, 6)$ determine the rectangle with vertices $(1, 3), (1, 6), (4, 6), (4, 3)$.

   The signed area contributed by (x_i, y_i) and (x_j, y_j) is the area of the rectangle they determine if the slope of the line between them is positive, and is the negative of the area of the rectangle they determine if the slope of the line between them is negative.

(Define the signed area to be 0 if $x_i = x_j$ or $y_i = y_j$, since then the rectangle is degenerate.) So the signed area is positive if a higher x value goes with a higher y value for the pair of points, and negative otherwise. Assume that the x_i are all distinct and the y_i are all distinct.

(a) The sample covariance of the data is defined to be

$$r = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

where

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \text{ and } \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$$

are the sample means. (There are differing conventions about whether to divide by $n-1$ or n in the definition of sample covariance, but that need not concern us for this problem.)

Let (X, Y) be one of the (x_i, y_i) pairs, chosen uniformly at random. Determine precisely how $\text{Cov}(X, Y)$ is related to the sample covariance.

(b) Let (X, Y) be as in (a), and $(\tilde{X}, \tilde{Y})$ be an independent draw from the same distribution. That is, (X, Y) and $(\tilde{X}, \tilde{Y})$ are randomly chosen from the n points, independently (so it is possible for the same point to be chosen twice).

Express the total signed area of the rectangles as a constant times $E((X - \tilde{X})(Y - \tilde{Y}))$. Then show that the sample covariance of the data is a constant times the total signed area of the rectangles.

Hint: Consider $E((X - \tilde{X})(Y - \tilde{Y}))$ in two ways: as the average signed area of the random rectangle formed by (X, Y) and $(\tilde{X}, \tilde{Y})$, and using properties of expectation to relate it to $\text{Cov}(X, Y)$. For the former, consider the n^2 possibilities for which point (X, Y) is and which point $(\tilde{X}, \tilde{Y})$; note that n such choices result in degenerate rectangles.

(c) Based on the interpretation from (b), give intuitive explanations of why for any r.v.s W_1, W_2, W_3 and constants a_1, a_2, covariance has the following properties:

(i) $\text{Cov}(W_1, W_2) = \text{Cov}(W_2, W_1)$;

(ii) $\text{Cov}(a_1 W_1, a_2 W_2) = a_1 a_2 \text{Cov}(W_1, W_2)$;

(iii) $\text{Cov}(W_1 + a_1, W_2 + a_2) = \text{Cov}(W_1, W_2)$;

(iv) $\text{Cov}(W_1, W_2 + W_3) = \text{Cov}(W_1, W_2) + \text{Cov}(W_1, W_3)$.