Stats 116 Problem Set 7

Due: Wednesday, May 18 5:00 p.m. on Gradescope

Please show your work for each exercise. If you collaborate with someone else—this is fine—be sure to note that in your homework submission. You must each write up separate answer sets. Questions are either from Ross's *A First Course in Probability* or our are home-cooked special sauce.

Question 7.1 (Blitzstein and Hwang Exercise 7.16)**:** Let X and Y have joint PDF

$$f_{X,Y}(x, y) = x + y, \quad \text{for } 0 \le x \le 1,\ 0 \le y \le 1.$$

(a) (2 pts) Check that this is a valid joint PDF.

(b) (2 pts) Are X and Y independent?

(c) (2 pts) Find the marginal PDFs of X and Y.

(d) (2 pts) Find the conditional PDF of Y given $X = x$.

Question 7.2 (Blitzstein and Hwang Exercise 7.23)**:** The *volume* of a region in n-dimensional Euclidean space $\mathbb{R}^n$ is the integral of 1 over that region. The *unit ball* in $\mathbb{R}^n$ is

$$\{x \in \mathbb{R}^n : \|x\|_2^2 \le 1\} = \{(x_1, \ldots, x_n) : x_1^2 + \cdots + x_n^2 \le 1\},$$

the ball of radius 1 centered at 0. The volume of the unit ball in n dimensions is

$$v_n = \frac{\pi^{n/2}}{\Gamma(n/2 + 1)},$$

where $\Gamma(a) = \int_0^\infty x^{a-1} e^{-x} dx$ is the gamma function. You may assume that $\Gamma(1) = 1$ (this is obvious) and $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ (this is not so obvious, but follows from, e.g., the Box-Muller transform in class), and of course you'll want to use that $\Gamma(a + 1) = a\Gamma(a)$ for any $a > 0$. For practice, verify that $v_2 = \pi$ and $v_3 = \frac{4}{3}\pi$. Let $U_1, U_2, \ldots, U_n \overset{\text{iid}}{\sim} \mathsf{Uni}[-1, 1]$.

(a) (2 pts) Find the probability that the vector $U = (U_1, U_2, \ldots, U_n)$ is in the unit ball in $\mathbb{R}^n$

(b) (2 pts) Evaluate the result in part (a) numerically for $n = 1, 2, \ldots, 10$, and plot the results (using a computer or your hands if you're quite good at making hand-drawn graphs). You should be able to do this without doing any integrals.

(c) (2 pts) Let c be a constant with $0 < c < 1$, and let X_n count how many of the U_j satisfy $|U_j| > c$. What is the distribution of X_n?

(d) (2 pts) For $c = 1/\sqrt{2}$, use the result of part (c) to give a simple, short derivation of what happens to the probability in part (a) as $n \to \infty$.

Question 7.3 (Blitzstein and Hwang Exercise 8.2 (2 pts))**:** Find the PDF of X^7 for $X \sim \mathsf{Exp}(\lambda)$.

Question 7.4 (Blitzstein and Hwang Exercise 8.11 (4pts))**:** Let T be a continuous random variable and $V = 1/T$. Show that their CDFs are related as follows:

$$F_v(v) = \begin{cases} F_T(0) + 1 - F_T(1/v) & \text{for } v > 0 \\ F_T(0) & \text{for } v = 0 \\ F_T(0) - F_T(1/v) & \text{for } v < 0. \end{cases}$$

Question 7.5 (Blitzstein and Hwang Exercise 8.22 (4pts))**:** Use a convolution integral to show that if $X \sim \mathsf{N}(\mu_1, \sigma^2)$ and $Y \sim \mathsf{N}(\mu_2, \sigma^2)$ are independent, then their sum $T = X + Y \sim \mathsf{N}(\mu_1 + \mu_2 + 2\sigma^2)$. You can use a standardization idea to reduce to the standard Normal case before setting up the integral. *Hint:* Complete the square.

Question 7.6 (Block matrix inversion)**:** In this question, you will develop a few ideas on performing inversion of a matrix by blocks, an important concept in statistics, probability, and numerical linear algebra. Let

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

be an $N \times N$ full rank matrix (so invertible), where $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times m}$, $C \in \mathbb{R}^{m \times n}$, and $D \in \mathbb{R}^{m \times m}$, where $m + n = N$. (We call A, B, C, D *sub-matrices* of M.) We will write M^{-1} in terms of inverses of its sub-matrices. To find the inverse of M, we write in terms of the block matrix multiplication

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} W & X \\ Y & Z \end{bmatrix} = I_N = \begin{bmatrix} I_n & 0 \\ 0 & I_m \end{bmatrix}.$$

Here $W \in \mathbb{R}^{n \times n}$, $X \in \mathbb{R}^{n \times m}$, $Y \in \mathbb{R}^{m \times n}$, and $Z \in \mathbb{R}^{m \times m}$. So working out the matrix multiplications, we must have

$$AW + BY = I_n, \quad CW + DY = 0, \quad AX + BZ = 0, \quad CX + DZ = I_m. \tag{7.1}$$

(a) $\star$ **(Extra credit: 2pts)** Assume that M is positive definite, meaning that $M = M^\top$ and $x^\top M x > 0$ whenever $x \neq 0$. Why are A and D invertible?

For the remainder of the question, you may assume all the given inverses exist.

(b) (2pts) Using the identities (7.1), show that

$$Y = -D^{-1}CW, \quad (A - BD^{-1}C)W = I, \quad \text{and so} \quad W = (A - BD^{-1}C)^{-1}.$$

(c) (2pts) Show similarly that

$$Z = (D - CA^{-1}B)^{-1}$$

and $X = -A^{-1}BZ$.

(d) (2pts) Conclude that

$$M^{-1} = \begin{bmatrix} (A - BD^{-1}C)^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -D^{-1}C(A - BD^{-1}C)^{-1} & (D - CA^{-1}B)^{-1} \end{bmatrix}.$$

Question 7.7 (Multivariate normal distributions)**:** Recall that a vector $Z \in \mathbb{R}^n$ has multivariate normal distribution with mean μ and covariance Σ if Z has density

$$f(z) = \frac{1}{(2\pi)^{n/2}\sqrt{\det(\Sigma)}} \exp\left(-\frac{1}{2}(z-\mu)^T\Sigma^{-1}(z-\mu)\right).$$

In this question, you will investigate conditioning on parts of a Gaussian, then use Question 7.6 to provide precise distributional answers. The specific identities you derive here are part of what makes using Gaussian distributions so important to (all) of multivariate statistics and applications of probability ideas in the sciences.

Let $X \in \mathbb{R}^n$ and $Y \in \mathbb{R}^m$ be jointly normally distributed (with mean 0), so that the vector

$$Z = \begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathsf{N}(0, \Sigma), \quad \text{where} \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

and $\Sigma_{11} \in \mathbb{R}^{n \times n}$ while $\Sigma_{22} \in \mathbb{R}^{m \times m}$, and $\Sigma_{12} = \Sigma_{21}^T$, so that Σ is symmetric. In this question, it will be convenient to work with what is often called the *precision* matrix, which is the inverse of the covariance Σ, i.e., $K = \Sigma^{-1}$. We also write K in the block matrix form

$$K = \begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{bmatrix},$$

where the formulae for K_{ij} follow from Question 7.6, so $K_{11} \in \mathbb{R}^{n \times n}$, $K_{12} \in \mathbb{R}^{n \times m}$, $K_{12} = K_{21}^T$, and $K_{22} \in \mathbb{R}^{m \times m}$, and K_{11} and K_{22} are symmetric.

Let $f(x, y)$ denote the density of the vector (X, Y).

(a) (2pts) Show that (ignoring the normalization term for the exponential) we have[1]

$$f(x, y) \propto \exp\left(-\frac{1}{2}\left(x^T K_{11} x + 2x^T K_{12} y + y^T K_{22} y\right)\right).$$

Hint: Because Σ is symmetric, so is K, and $K_{12} = K_{21}^T$.

(b) (2pts) Let $x \in \mathbb{R}^n, y \in \mathbb{R}^m$, $A \in \mathbb{R}^{n \times n}$ be a symmetric and invertible matrix, and $B \in \mathbb{R}^{n \times m}$ and $C \in \mathbb{R}^{m \times m}$. Show that

$$x^T A x + 2x^T B y + y^T C y = (x + A^{-1}By)^T A (x + A^{-1}By) + y^T(C - B^T A^{-1} B)y.$$

(c) (3pts) Let $f(x \mid y)$ denote the conditional density of x given $Y = y$. Show that

$$f(x \mid y) \propto \exp\left(-\frac{1}{2}(x + K_{11}^{-1}K_{12}y)^T K_{11}(x + K_{11}^{-1}K_{12}y)\right).$$

Hint: You can ignore any terms involving *only* y in the exponential. Why is this?

(d) (3pts) Show that conditional on $Y = y$, X has normal distribution with mean $-K_{11}^{-1}K_{12}y$ and covariance K_{11}^{-1}, that is,

$$X \mid Y = y \sim \mathsf{N}\left(-K_{11}^{-1}K_{12}y, K_{11}^{-1}\right).$$

[1] The symbol $\propto$ means "proportional to," and writing $f(x) \propto g(x)$ means that there is a fixed positive constant $c > 0$ such that $f(x) = cg(x)$.

(e) (2pts) Show that

$$\int_{\mathbb{R}^m} \exp\left(-\frac{1}{2}(x^T K_{11} x + 2x^T K_{12} y + y^T K_{22} y)\right) dy$$
$$= \exp\left(-\frac{1}{2}x^T(K_{11} - K_{12}K_{22}^{-1}K_{21})x\right)\sqrt{\det(K_{22}^{-1})(2\pi)^m}.$$

Hint: You may use that $\int_{\mathbb{R}^m} \exp(-\frac{1}{2}y^T Ay)dy = \sqrt{(2\pi)^m \det(A^{-1})}$, which we know from the normalization constants in the multivariate normal.

(f) (2pts) Conclude from part (e), which shows how to marginalize over y, that

$$X \sim \mathsf{N}(0, \Sigma_{11}).$$

This shows that whenever (X, Y) is jointly normal, then the marginals over X and Y are both normal as well, and moreover, we can directly obtain their means and covariances! *Hint:* Recall Question 7.6. What must $(K_{11} - K_{12}K_{22}^{-1}K_{21})^{-1}$ equal?

(g) (2pts) Assume that $n = m = 1$, so we have a bivariate normal distribution, where

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathsf{N}\left(0, \begin{bmatrix} \sigma_{11}^2 & \rho \\ \rho & \sigma_{22}^2 \end{bmatrix}\right).$$

Show that the variance of X *conditional* on knowing $Y = y$ is always smaller than the variance of X if $\rho \neq 0$. So knowledge about Y always decreases uncertainty about X!

Question 7.8* (Extra credit (4pts))**:** Let X_i, $i = 1, \ldots, n$ be a sequence of independent normal random variables with $X_i \sim \mathsf{N}(\mu_i, \sigma_i^2)$. Show that for any sequence of scalars $a_1, a_2, \ldots, a_n$,

$$\sum_{i=1}^n a_i X_i \sim \mathsf{N}\left(\sum_{i=1}^n a_i \mu_i, \sum_{i=1}^n a_i^2 \sigma_i^2\right).$$