Review Session 3 – Solutions

1 Solutions

1.1 Previous Core Competency Problems

Problem 1 (2018 Summer Practice, # 12). Suppose that $U_1, U_2 \overset{i.i.d.}{\sim} U(0,1)$. Let $V_1 := \max(U_1, U_2)$, $V_2 := \min(U_1, U_2)$.

(a) Find $\mathbb{P}(V_1 \geq x, V_2 \leq y)$, where $x, y \in [0,1]$.

(b) Hence or otherwise find the joint density for (V_1, V_2).

(c) Hence or otherwise compute $\mathbb{E}(V_1^2 + V_2^2)$.

Solution

First, we compute the cdf

$$\mathbb{P}(V_1 \leq x, V_2 \leq y) = \mathbb{P}(V_1 \leq x) - \mathbb{P}(V_1 \leq x, V_2 > y) = \begin{cases} x^2 - (x-y)^2 & x \geq y \\ x^2 & x < y \end{cases}$$

Thus,

$$\frac{\partial^2}{\partial y \partial x}\mathbb{P}(V_1 \leq x, V_2 \leq y) = \begin{cases} 2 & x \geq y \\ 0 & x < y \end{cases}$$

is the pdf of (V_1, V_2). Then,

$$\mathbb{P}(V_1 \geq x_0, V_2 \leq y_0) = \int_0^{y_0} \int_{x_0}^1 2\, dx\, dy = 2y_0 - 2x_0 y_0$$

if $x_0 \geq y_0$ since all points in $(x_0, 1) \times (0, y_0)$ lie in $\{x \geq y\}$ if $x_0 \geq y_0$.

If $x_0 < y_0$, then not all points in $(x_0, 1) \times (0, y_0)$ lie in $\{x \geq y\}$. We must compute the area of the region where the density is 2 and not 0. One can draw this region and see it is a rectangle minus a right triangle and so its area is

$$(1 - x_0)y_0 - (y_0 - x_0)^2/2 = y_0 - y_0^2/2 - x_0^2/2 \implies \mathbb{P}(V_1 \geq x_0, V_2 \leq y_0) = 2y_0 - y_0^2 - x_0^2 \text{ if } y_0 > x_0.$$

Finally,

$$\mathbb{E}[V_1^2 + V_2^2] = \mathbb{E}[U_1^2 + U_2^2] = 2/3$$

Problem 2 (2018 Summer Practice, # 13). Suppose $X_1, X_2, X_3 \overset{i.i.d.}{\sim} N(0,1)$. Let (Y_1, Y_2, Y_3) be defined as follows:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}$$

(a) Find the joint distribution of (Y_1, Y_2, Y_3).

(b) Show that $Y_1^2 + Y_2^2 + Y_3^2 = X_1^2 + X_2^2 + X_3^2$.

(c) Hence or otherwise derive the distribution of $(X_1 - \overline{X})^2 + (X_2 - \overline{X})^2 + (X_3 - \overline{X})^2$, where $\overline{X} = \frac{X_1 + X_2 + X_3}{3}$.

Solution

(a) We have $Y_1, Y_2, Y_3 \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ since the given matrix is orthonormal.

(b) The matrix, being orthonormal, is also norm preserving meaning $Y_1^2 + Y_2^2 + Y_3^2 = X_1^2 + X_2^2 + X_3^2$.

(c) We claim that $S^2 = \sum_{i=1}^{3}(X_i - \overline{X})^2$ is a χ_2^2 distribution. First, we have

$$S^2 = \sum_{i=1}^{3} X_i^2 - 2X_i\overline{X} + \overline{X}^2 = \sum_{i=1}^{3} X_i^2 + \sum_{i=1}^{3} \overline{X}^2 - 2\overline{X}\sum_{i=1}^{3} X_i = \sum_{i=1}^{3} X_i^2 - 3\overline{X}^2.$$

Recall S^2 and $\overline{X}$ are independent since they are uncorrelated projections of (X_1, X_2, X_3) (which are also Gaussian) onto orthogonal subspaces. We also have that $\sqrt{3} \cdot \overline{X} \sim \mathcal{N}(0,1) \implies 3\overline{X}^2 \sim \chi_1^2$. Then, comparing the mgf's of $\sum_{i=1}^{3} X_i \sim \chi_3^2$ and $3\overline{X}^2 + S^2 \sim \chi_1^2 + S^2$ gives us that $S^2 \sim \chi_2^2$.

Problem 3 (2019 September, # 4)**.** Let $X_1, \ldots, X_n$ be a random sample (i.i.d.) from a density function f. The corresponding CDF is denoted by F. Denote by $X_{(1)} < \cdots < X_{(n)}$ the order statistics, i.e. a rearrangement of $X_1, \ldots, X_n$ according to their values.

(i) For $n = 2$, derive the density function of $X_{(1)}$ in terms of f and F. [**Hint:** You may want to find the distribution function of $X_{(1)}$ first.]

(ii) For $n = 3$, derive the density function of $X_{(2)}$ in terms of f and F.

(iii) For any n and k, derive the density function of $X_{(k)}$ in terms of f and F.

Solution

We derive the general formula for n and k. We claim

$$f_{X_{(k)}}(x) = \frac{n!}{(n-k)!(k-1)!} F(x)^{k-1}(1-F(x))^{n-k} f(x). \tag{1}$$

We show this by induction. First, we have the cdf of $X_{(1)}$ is given by $\mathbb{P}(X_{(1)} \le x) = 1 - (1-F(x))^n$. Taking the derivative of this gives us density:

$$f_{X_{(1)}}(x) = n(1-F(x))^{n-1} f(x).$$

Thus, (1) holds for $k = 1$. Next, suppose (1) holds for k. Then,

$$\mathbb{P}(X_{(k+1)} \le x) = \mathbb{P}(X_{(k)} \le x, X_{(k+1)} \le x) = \mathbb{P}(X_{(k)} \le x) - \mathbb{P}(X_{(k+1)} > x, X_{(k)} \le x).$$

The first term on the RHS above is the cdf of $X_{(k)}$ evaluated at x and the second term is

$$\mathbb{P}(X_{(k+1)} > x, X_{(k)} \le x) = \binom{n}{k} F(x)^k (1-F(x))^{n-k}.$$

Then, taking the derivative of the difference gives us, using the induction hypothesis:

$$\frac{n!}{(n-k)!(k-1)!} F(x)^{k-1}(1-F(x))^{n-k} f(x) - \binom{n}{k}\left(kf(x)F(x)^{k-1}(1-F(x))^{n-k} - F(x)^k(n-k)f(x)(1-F(x))^{n-k-1}\right).$$

This simplifies to

$$\frac{n!}{(n-k-1)!k!} F(x)^k (1-F(x))^{n-k-1} f(x),$$

as desired.

Problem 4 (2020 May, # 1)**.** Suppose we have a random variable $\xi \sim$ Uniform$(0,1)$. Suppose that conditioning on ξ, we have i.i.d. Bernoulli(ξ) random variables $X_1, X_2, \ldots, X_n, X_{n+1}$, i.e. $P(X_i = 1|\xi) = 1 - P(X = 0|\xi) = \xi$. Calculate

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

Solution

We have

$$\begin{aligned}\mathbb{P}(X_{n+1}=1|X_1,\dots,X_n) &= \frac{\mathbb{P}(X_{n+1}=1,X_1,\dots,X_n)}{\mathbb{P}(X_1,\dots,X_n)} \\ &= \frac{\int_0^1 \xi^{1+\sum_{i=1}^n X_i}(1-\xi)^{n-\sum_{i=1}^n X_i}\,d\xi}{\int_0^1 \xi^{\sum_{i=1}^n X_i}(1-\xi)^{n-\sum_{i=1}^n X_i}\,d\xi} \\ &= \frac{\frac{\Gamma(2+\sum_{i=1}^n X_i)\Gamma(n+1-\sum_{i=1}^n X_i)}{\Gamma(3+n)}}{\frac{\Gamma(\sum_{i=1}^n X_i+1)\Gamma(n-\sum_{i=1}^n X_i+1)}{\Gamma(n+2)}} \\ &= \frac{1+\sum_{i=1}^n X_i}{n+2}.\end{aligned}$$

We can also proceed via

$$\mathbb{P}(X_{n+1}=1|X_1,\dots,X_n) = \mathbb{E}_{\xi|X_1,\dots,X_n}[\mathbb{P}(X_{n+1}=1|X_1,\dots,X_n,\xi)] = \mathbb{E}[\xi|X_1,\dots,X_n].$$

However, $\xi|X_1,\dots,X_n \sim$ Beta $(1+\sum_{i=1}^n X_i, 1+n-\sum_{i=1}^n X_i)$ using properties of the Beta-Bernoulli conjugate family. Thus, the above posterior mean is $\frac{1+\sum_{i=1}^n X_i}{2+n}$.

Problem 5 (2020 May, # 4)**.** Let Z_1, Z_2, Z_3 be i.i.d. $N(0,1)$ random variables. Let $R=\sqrt{Z_1^2+Z_2^2+Z_3^2}$.

(i) Find the distribution of R and write down its density function.

(ii) Suppose that we have two independent random variables $X \sim$ Gamma(α,λ) and $Y \sim$ Gamma(β,λ), where $\alpha,\beta,\lambda>0$. Let

$$U = X+Y \qquad \text{and} \qquad V=\frac{X}{X+Y}.$$

Find the joint density (p.d.f.) of (U,V) and identify the joint distribution (c.d.f.).
Hint: density function of Gamma$(\alpha,\lambda) = \lambda^\alpha x^{\alpha-1}e^{-\lambda x}/\Gamma(\lambda)$.

Solution

(i) $R^2 = Z_1^2+Z_2^2+Z_3^2 \sim \chi_3^2$. Recall that $\chi_k^2 \sim$ Gamma$(k/2, 1/2)$ (using the shape-rate parametrization given in the hint). Thus, by the hint, R^2 has density for $r>0$:

$$f_{R^2}(r) = \frac{1}{2^{3/2}\Gamma(3/2)} r^{k/2-1}e^{-r/2}.$$

Then, we use the pdf transformation law to find the distribution of $\sqrt{R^2}$ for $r>0$:

$$f_R(r) = f_{R^2}(r^2)\cdot|2r| = \frac{1}{2^{1/2}\Gamma(3/2)} r^2 e^{-r^2/2}.$$

(ii) The inverse transformation of $(X,Y)\mapsto(U,V)$ is $(U,V)\mapsto(U\cdot V, U-U\cdot V)$ which has Jacobian $-U$. Thus, by the pdf transformation law we have for $u>0$ and $v\in(0,1)$:

$$f_{U,V}(u,v) = \frac{\lambda^{\alpha+\beta}(uv)^{\alpha-1}(u-uv)^{\beta-1}e^{-\lambda u}}{\Gamma(\lambda)^2}\cdot u = \frac{\lambda^{\alpha+\beta}u^{\alpha+\beta-1}v^{\alpha-1}(1-v)^{\beta-1}e^{-\lambda u}}{\Gamma(\lambda)^2}.$$

Since the joint pdf factors into the marginal pdf's of U and V, we see that (U,V) is the joint distribution of an independent Gamma$(\alpha+\beta,\lambda)$ and a Beta(α,β).

Problem 6 (2020 September, # 4)**.** Suppose we generate $U\sim$ Unif$[0,3]$. Let V denote the value of the integer nearest to U (so V takes values in $\{0,1,2,3\}$). Let X denote the rounding error i.e. the absolute distance between U and V.

(a) What is the distribution of V?

(b) What is the distribution of the X.

(c) Are X and V independent?

(d) Are X and U independent?

Solution

(a)

$$\begin{aligned}\mathbb{P}(V=0) &= \mathbb{P}(U \le 1/2) = 1/6\\ \mathbb{P}(V=1) &= \mathbb{P}(1/2 < U < 3/2) = 1/3\\ \mathbb{P}(V=2) &= \mathbb{P}(3/2 < U < 5/2) = 1/3\\ \mathbb{P}(V=3) &= \mathbb{P}(U > 5/2) = 1/6.\end{aligned}$$

(b) It's straightforward to see

$$\mathbb{P}(X \le x | V = i) = \begin{cases} 2x & x \le 1/2 \\ 1 & \text{otherwise,} \end{cases}$$

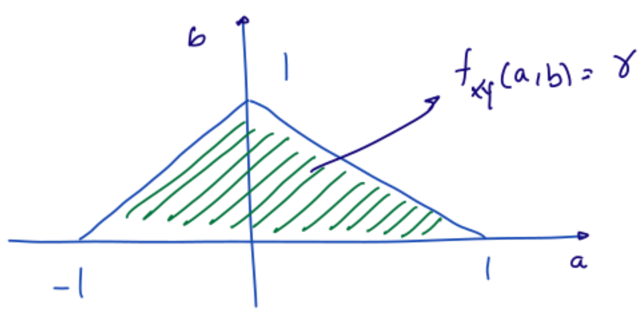
for any $i \in \{0, 1, 2, 3\}$.

(c) Using (b), X and V are independent.

(d) X and U are not independent. Note that if $U \in (1/4, 3/4)$, then $X \ge 1/4$. Thus,

$$\mathbb{P}\left(X < 1/4, U \in (1/4, 3/4)\right) = 0 \neq \mathbb{P}(X < 1/4)\mathbb{P}(U \in (1/4, 3/4)).$$

Problem 7 (2021 May, # 4)**.** The joint pdf $f_{X,Y}(a,b)$ of a random variable X and Y is zero outside the triangular region shown below and is equal to a fixed number γ on the triangular region.



Answer the following questions:

(i) Calculate the value of γ.

(ii) Are X and Y independent? Prove your answer and then give an intuitive explanation.

(iii) Calculate $\text{Cov}(X,Y)$. Does the result you obtain make sense?

(iv) Calculate the joint CDF of two random variables $Z = X + Y$ and $W = X - Y$.

Solution

(i) The area of the triangular region is 1. Thus, $\gamma = 1$.

(ii) No, X and Y are not independent. We have $\mathbb{P}(Y > 1/2 || X| > 0.99) = 0$, yet $\mathbb{P}(Y > 1/2) > 0$.

(iii) We have

$$\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \int_{-1}^{0}\int_{0}^{x+1} xy\,dy\,dx + \int_{0}^{1}\int_{0}^{1-x} xy\,dy\,dx - \mathbb{E}[X]\mathbb{E}[Y] = 0 - \mathbb{E}[X]\mathbb{E}[Y].$$

Next, the marginal density of X is

$$f_X(x) = \begin{cases} \int_0^{x+1} 1\,dy = x+1 & x \in [-1,0] \\ \int_0^{1-x} 1\,dy = 1-x & x \in (0,1]. \end{cases}.$$

Thus,

$$\mathbb{E}[X] = \int_{-1}^{0} x(x+1)\,dx + \int_0^1 x(2-2x)\,dx = 0 \implies \mathrm{Cov}(X,Y) = 0.$$

The result makes sense since X and Y are positively correlated on half of the support and negatively correlated on the other half of the support. So, we should expect their overall correlation to be zero.

(iv) We have that if $F(X,Y) = (Z,W)$, then $F^{-1}(Z,W) = \left(\frac{Z+W}{2}, \frac{Z-W}{2}\right)$. Then the Jacobian of F^{-1} has determinant $|\det(JF^{-1})| = 1/2$. Thus, the pdf of (Z,W) is $1/2$ and the cdf is $F_{Z,W}(z,w) = zw/2$.

1.2 Additional Practice

Problem 8 (Casella & Berger, Exercise 4.17). Let X be an $\exp(1)$ random variable, and define Y to be the integer part of $X+1$, i.e. $Y = \lfloor X+1 \rfloor$.

(a) Find the distribution of Y.

(a) Find the conditional distribution of $X-4$ given $Y \geq 5$.

Solution

(a) $\mathbb{P}(Y = n+1) = \int_n^{n+1} e^{-x}\,dx = e^{-n}(1-e^{-1})$ so that $Y \sim$ geometric$(1-e^{-1})$.

(a) $\mathbb{P}(X-4 \leq x | Y \geq 5) = \mathbb{P}(X-4 \leq x | X \geq 4) = \mathbb{P}(X \leq x) = e^{-x}$ by the memoryless property of the exponential distribution.

Problem 9 (Casella & Berger, Exercise 4.54). Find the pdf of $\prod_{i=1}^n X_i$ where the X_i's are independent $\mathrm{Unif}([0,1])$ random variables. Hint: try to calculate the cdf, and remember the relationship between uniforms and exponentials.

Solution

Recall from review session 2 that each $-\log(X_i) \sim \exp(1)$ which means that $\sum_{i=1}^n -\log(X_i) \sim$ gamma$(n,1)$. Then, the cdf of the product is

$$\mathbb{P}\left(\prod_{i=1}^n X_i \leq x\right) = \mathbb{P}\left(\sum_{i=1}^n -\log(X_i) \geq -\log(x)\right) = \int_{-\log(x)}^{\infty} \frac{1}{\Gamma(n)} z^{n-1} e^{-z}\,dz.$$

The pdf is the derivative of the above RHS which is (by the fundamental theorem of calculus):

$$\frac{d}{dx}\int_{-\log(x)}^{\infty} \frac{1}{\Gamma(n)} z^{n-1} e^{-z}\,dz = -\frac{1}{\Gamma(n)}(-\log(x))^{n-1} e^{\log(x)} \frac{d}{dx}(-\log(x)) = \frac{1}{\Gamma(n)}(-\log(x))^{n-1},$$

for $x \in (0,1)$.

Problem 10 (Casella & Berger, Exercise 4.59). For any three random variables X, Y, Z with finite variances, prove that

$$\mathrm{Cov}(X,Y) = \mathbb{E}[\mathrm{Cov}(X,Y|Z)] + \mathrm{Cov}(\mathbb{E}[X|Z], \mathbb{E}[Y|Z]),$$

where $\mathrm{Cov}(X,Y|Z)$ is the covariance of X and Y conditional on Z.

Solution

We first note that we may assume WLOG that $\mathbb{E}[X] = \mathbb{E}[Y] = \mathbb{E}[Z] = 0$ since, by the covariance being invariant under shifts, we can substitute X, Y, Z with $X - \mathbb{E}[X]$, $Y - \mathbb{E}[Y]$, and $Z - \mathbb{E}[Z]$ in the desired expression. Then, we can write by the tower law,

$$\mathrm{Cov}(X, Y) = \mathbb{E}[XY] = \mathbb{E}[\mathbb{E}[XY|Z]].$$

Adding and subtracting $\mathbb{E}[X|Z]\mathbb{E}[Y|Z]$, we get

$$\mathrm{Cov}(X, Y) = \mathbb{E}[\mathbb{E}[XY|Z] - \mathbb{E}[X|Z]\mathbb{E}[Y|Z]] + \mathbb{E}[\mathbb{E}[X|Z]\mathbb{E}[Y|Z]].$$

The first term above is $\mathbb{E}[\mathrm{Cov}(X, Y|Z)]$ and the second term is $\mathrm{Cov}(\mathbb{E}[X|Z], \mathbb{E}[Y|Z])$ since $\mathbb{E}[\mathbb{E}[X|Z] = \mathbb{E}[X] = 0$.