Tutorial Worksheet 1 - Reviews of Probability Theory (with Solutions)

Problem 1. (Transformation of Random Variables)

(a) (a) Let X_1, X_2, X_3 be independent $N(0,1)$ random variables. Find the probability density function of $U = X_1^2 + X_2^2 + X_3^2$.

Solution: We know from class that X_i^2 has the χ_1^2 distribution, i.e., the Gamma$(1/2, 1/2)$ distribution. Since X_1, X_2, X_3 are independent, so are X_1^2, X_2^2, X_3^2, so $U = X_1^2 + X_2^2 + X_3^2$ is a sum of 3 independent Gamma$(1/2, 1/2)$ random variables, and as such has the Gamma$(3/2, 1/2)$ distribution (or χ_3^2 distribution). Thus

$$f_U(u) = \begin{cases} \frac{(1/2)^{3/2}}{\Gamma(3/2)} u^{1/2} e^{-(1/2)u} & \text{if } x > 0 \\ 0 & \text{otherwise.} \end{cases}$$

The expression can be simplified by noting that $\Gamma(3/2) = (1/2)\Gamma(1/2) = \frac{\sqrt{\pi}}{2}$.

(b) Suppose that the random vector $\boldsymbol{Y} = (Y_1, Y_2, Y_3)$ is uniformly distributed on the sphere of radius 1 centred at the origin; that is, $\boldsymbol{Y}$ has joint probability density function (pdf)

$$f_{\boldsymbol{Y}}(y_1, y_2, y_3) = \begin{cases} \frac{3}{4\pi} & \text{if } (y_1, y_2, y_3) \in S \\ 0 & \text{otherwise,} \end{cases}$$

where $S = \left\{(y_1, y_2, y_3) \in \mathbb{R}^3 : y_1^2 + y_2^2 + y_3^2 \leq 1\right\}$ is the sphere of radius 1 centred at $(0,0,0)$. Let $V = Y_1^2 + Y_2^2 + Y_3^2$. Find the probability density function of V and find $E[V]$.

Solution: We can find the pdf of V by first computing the CDF of V. For $v \in [0,1]$ we have

$$F_V(v) = P(V \leq v) = P\left(Y_1^2 + Y_2^2 + Y_3^2 \leq (\sqrt{v})^2\right) = P\left((Y_1, Y_2, Y_3) \in S_{\sqrt{v}}\right),$$

where $S_{\sqrt{v}}$ is the sphere of radius $\sqrt{v}$. Thus, for $v \in [0,1]$

$$\begin{aligned} F_V(v) &= \iiint_{S_{\sqrt{v}}} f_Y(y_1, y_2, y_3)\, dy_1 dy_2 dy_3 = \iiint_{S_{\sqrt{v}}} \frac{3}{4\pi} dy_1 dy_2 dy_3 \\ &= \frac{3}{4\pi} \times \text{ Volume of } S_{\sqrt{v}} \\ &= \frac{3}{4\pi} \times \frac{4\pi(\sqrt{v})^3}{3} = v^{3/2}. \end{aligned}$$

We also have $F_V(v) = 0$ for $v < 0$ and $F_V(v) = 1$ for $v > 1$. Differentiating, we obtain the pdf of V as

$$f_V(v) = F_V'(v) = \begin{cases} \frac{3\sqrt{v}}{2} & \text{for } 0 \leq v \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$E[V] = \frac{3}{2} \int_0^1 v^{3/2} dv = \frac{3}{2}\left[\frac{2}{5} v^{5/2}\right]_0^1 = \frac{3}{5}.$$

(c) Let X and Y be independent and identically distributed uniform random variables on the interval $(0,1)$. Define $U = \frac{1}{2}(X+Y)$ to be the average and define $V = X$.

(i) Find the joint probability density function of (U, V) and draw the sample space of (U, V).

(ii) Find the marginal probability density function of U.

Solution: (i) The inverse transformation is $X = V$ and $Y = 2U - V$. The matrix of partial derivatives of the inverse transformation is

$$\begin{bmatrix} 0 & 1 \\ 2 & -1 \end{bmatrix}$$

with determinant -2. The joint pdf of (X, Y) is

$$f_{XY}(x, y) = \begin{cases} 1 & \text{for } 0 \le x \le 1, 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

So the joint pdf of (U, V) is

$$f_{UV}(u, v) = \begin{cases} 2 & \text{for } (u, v) \in S_{UV} \\ 0 & \text{otherwise} \end{cases}$$

where S_{UV} is the sample space of (U, V), determined by the constraints $0 \le v \le 1$ and $0 \le 2u - v \le 1$, or $0 \le v \le 1$ and $v/2 \le u \le (v + 1)/2$. The sample space of (U, V) is plotted in Fig. 1.

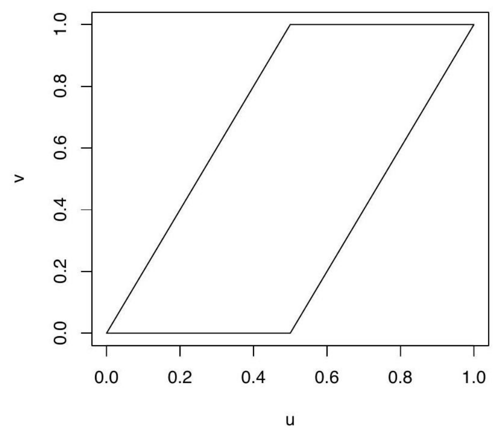

Figure 1: Sample space of (U, V).

(ii) To get the marginal pdf of U we integrate the joint pdf $f_{UV}(u, v)$ over the variable v from $-\infty$ to ∞. From Figure 1 we see that the limits of integration are different depending on whether $u \in [0, 0.5]$ or $u \in (0.5, 1]$. For $u \in [0, 0.5]$ we have

$$f_U(u) = \int_0^{2u} (2)dv = 4u.$$

For $u \in (0.5, 1]$ we have

$$f_U(u) = \int_{2u-1}^{1} (2)dv = 2(1 - (2u - 1)) = 4(1 - u)$$

Clearly, for $u \notin [0, 1]$ we have $f_U(u) = 0$. To summarize,

$$f_U(u) = \begin{cases} 4u & \text{for } u \in [0, 0.5] \\ 4(1 - u) & \text{for } u \in (0.5, 1] \\ 0 & \text{otherwise.} \end{cases}$$

(d) Let $X = (X_1, X_2)^T$ be uniformly distributed on the positive quadrant intersected with the disk of radius 1 centred at the origin; i.e., X has joint pdf $f_X(x_1, x_2) = \frac{4}{\pi} I_{S_X}(x_1, x_2)$, where

$$S_X = \left\{(x_1, x_2) \in \mathbb{R}^2 : x_1 \ge 0, x_2 \ge 0, x_1^2 + x_2^2 \le 1\right\}.$$

Let $Y_1 = X_1^2$ and $Y_2 = X_2^2$. Find the joint pdf of $(Y_1, Y_2)^T$ and the marginal pdf of Y_1.

Solution: We use the bivariate change of variable formula to get the joint pdf of (Y_1, Y_2). The inverse transformation is $X_1 = Y_1^{1/2}$ and $X_2 = Y_2^{1/2}$. The Jacobian is

$$\mathbf{J} = \det \begin{bmatrix} \frac{1}{2\sqrt{y_1}} & 0 \\ 0 & \frac{1}{2\sqrt{y_2}} \end{bmatrix} = \frac{1}{4\sqrt{y_1 y_2}}$$

The support of (Y_1, Y_2) is

$$S_Y = \left\{(y_1, y_2) \in \mathbb{R}^2 : y_1 \geq 0, y_2 \geq 0, y_1 + y_2 \leq 1\right\}$$

By the change of variable formula, the joint pdf of (Y_1, Y_2) is given by

$$f_Y(y_1, y_2) = f_X\left(\frac{1}{2\sqrt{y_1}}, \frac{1}{2\sqrt{y_2}}\right)|\mathbf{J}| = \frac{4}{\pi}\frac{1}{4\sqrt{y_1 y_2}} I_{S_Y}(y_1, y_2) = \frac{1}{\pi\sqrt{y_1 y_2}} I_{S_Y}(y_1, y_2).$$

The marginal pdf of Y_1 is obtained by integration. For $y_1 \in [0, 1]$,

$$\begin{aligned} f_{Y_1}(y_1) &= \int_{-\infty}^{\infty} f_Y(y_1, y_2)\, dy_2 = \int_0^{1-y_1} \frac{1}{\pi\sqrt{y_1 y_2}} dy_2 \\ &= \frac{1}{\pi\sqrt{y_1}}\left[2\sqrt{y_2}\right]_0^{1-y_1} \\ &= \frac{2\sqrt{1-y_1}}{\pi\sqrt{y_1}}, \end{aligned}$$

and $f_{Y_1}(y_1) = 0$ otherwise.

(e) Let X_1, X_2, X_3 be independent and identically distributed Exponential (λ) random variables (the Exponential (λ) distribution has pdf $f_X(x) = \lambda e^{-\lambda x}$ for $x > 0$ and $f_X(x) = 0$ for $x \leq 0$, and df $F_X(x) = 1 - e^{-\lambda x}$ for $x > 0$ and $F_X(x) = 0$ for $x \leq 0$). Find $P\left(X_1 + X_2 + X_3 \leq \frac{3}{2}\right)$. (Write down the appropriate 3-dimensional integral and evaluate it).

Solution: The appropriate 3-dimensional integral is (by symmetry every order in which the integration is done is equally easy)

$$\begin{aligned} &P\left(X_1 + X_2 + X_3 \leq \frac{3}{2}\right) \\ =& \int_0^{3/2}\int_0^{3/2-x_3}\int_0^{3/2-x_2-x_3} f_X(x_1) f_X(x_2) f_X(x_3)\, dx_1 dx_2 dx_3 \\ =& \int_0^{3/2}\int_0^{3/2-x_3} f_X(x_2) f_X(x_3)\left(1 - e^{-\lambda(3/2-x_2-x_3)}\right) dx_2 dx_3 \\ =& \int_0^{3/2} \lambda e^{-\lambda x_3}\int_0^{3/2-x_3}\left[\lambda e^{-\lambda x_2} - \lambda e^{-\lambda(3/2-x_3)}\right] dx_2 dx_3 \\ =& \int_0^{3/2} \lambda e^{-\lambda x_3}\left[1 - e^{-\lambda(3/2-x_3)} - \lambda\left(\frac{3}{2} - x_3\right) e^{-\lambda(3/2-x_3)}\right] dx_3 \\ =& 1 - e^{-3\lambda/2} - \frac{3\lambda}{2} e^{-3\lambda/2} - \left(\frac{3\lambda}{2}\right)^2 e^{-3\lambda/2} + \lambda^2 e^{-3\lambda/2}\frac{(3/2)^2}{2} \\ =& 1 - e^{-3\lambda/2}\left[1 + \frac{3\lambda}{2} + \frac{1}{2}\left(\frac{3\lambda}{2}\right)^2\right]. \end{aligned}$$

Problem 2. (Order Statistics)

(a) Let $X_1, \ldots, X_n$ be a sequence of independent random variables uniformly distributed on the interval $(0,1)$, and let $X_{(1)}, \ldots, X_{(n)}$ denote their order statistics. For fixed k, let $g_n(x)$ denote the probability density function of $nX_{(k)}$. Find $g_n(x)$ and show that

$$\lim_{n\to\infty} g_n(x) = \begin{cases} \frac{x^{k-1}}{(k-1)!}e^{-x} & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases}$$

which is the Gamma$(k,1)$ density.

Solution: The pdf of $X_{(k)}$ is

$$f_k(x_k) = \begin{cases} \frac{n!}{(k-1)!(n-k)!} x_k^{k-1}(1-x_k)^{n-k} & \text{for } 0 < x_k < 1 \\ 0 & \text{otherwise.} \end{cases}$$

Let $X = nX_{(k)}$. Then the set of possible values of X (the sample space of X) is $\{x : 0 < x < n\}$. So for $x \in (0,n)$, by the (1-dimensional) change of variable formula the pdf of X is given by

$$g_n(x) = f_k\left(\frac{x}{n}\right)\frac{1}{n} = \frac{(n-1)!}{(k-1)!(n-k)!}\left(\frac{x}{n}\right)^{k-1}\left(1-\frac{x}{n}\right)^{n-k}$$

and $g_n(x) = 0$ for $x \notin (0,n)$. Now fix $x > 0$ and let $n > x$. Then

$$\begin{aligned} g_n(x) &= \frac{(n-1)!}{(k-1)!(n-k)!}\left(\frac{x}{n}\right)^{k-1}\left(1-\frac{x}{n}\right)^{n-k} \\ &= \frac{(n-1)\times\ldots\times(n-k+1)}{n^{k-1}}\left(1-\frac{x}{n}\right)^{-k}\frac{1}{(k-1)!}x^{k-1}\left(1-\frac{x}{n}\right)^{n} \\ &\to \frac{1}{(k-1)!}x^{k-1}e^{-x}, \end{aligned}$$

as desired, since

$$\begin{aligned} \frac{(n-1)\times\ldots\times(n-k+1)}{n^{k-1}} &= \left(\frac{n-1}{n}\right)\times\ldots\times\left(\frac{n-k+1}{n}\right) \\ &\to 1\times\ldots\times 1 = 1, \end{aligned}$$

$(1-x/n)^{-k}$ clearly converges to 1 as $n\to\infty$ (for fixed k), and $(1-x/n)^n \to e^{-x}$ as $n\to\infty$ (from calculus). Since $g_n(x) = 0$ for $x < 0$ it is obvious that $g_n(x) \to 0$ as $n\to\infty$ if $x < 0$.

(b) (i) Let $X_1, \ldots, X_6$ be independent random variables uniformly distributed on the interval $(0,1)$. Find the pdf of $U = \min\{\max(X_1,X_2), \max(X_3,X_4), \max(X_5,X_6)\}$.

Solution: Let $Y_1 = \max(X_1,X_2), Y_2 = \max(X_3,X_4)$, and $Y_3 = \max(X_5,X_6)$. Then Y_1, Y_2, Y_3 are independent and identically distributed random variables, each with cdf

$$F_Y(y) = P(X_1 \leq y, X_2 \leq y) = [P(X_1 \leq y)]^2 = \begin{cases} y^2 & \text{for } 0 < y < 1 \\ 0 & \text{for } y \leq 0 \\ 1 & \text{for } y \geq 1 \end{cases}$$

and pdf

$$f_Y(y) = \begin{cases} 2y & \text{for } 0 < y < 1 \\ 0 & \text{otherwise.} \end{cases}$$

Then $U = \min(Y_1,Y_2,Y_3)$ has cdf

$$\begin{aligned} F_U(u) &= 1 - P(\min(Y_1,Y_2,Y_3) > u) \\ &= 1 - P(Y_1 > u, Y_2 > u, Y_3 > u) \\ &= 1 - (1-F_Y(u))^3 \\ &= \begin{cases} 1-(1-u^2)^3 & \text{for } 0 < u < 1 \\ 0 & \text{for } u \leq 0 \\ 1 & \text{for } u \geq 1. \end{cases} \end{aligned}$$

By differentiation, then, U has pdf given by

$$f_U(u) = \begin{cases} 6u\left(1-u^2\right)^2 & \text{for } 0<u<1 \\ 0 & \text{otherwise.} \end{cases}$$

(ii) Let $X_1, \ldots, X_n$ be independent and identically distributed random variables, each with a Uniform distribution on the interval $(0,1)$. Let $X = \min(X_1, \ldots, X_n)$ and $Y = \max(X_1, \ldots, X_n)$. Find $P\left(X < \frac{1}{2} < Y\right)$ and $E\left[X^3\right]$.

Solution: The joint pdf of (X, Y) is

$$f(x,y) = \begin{cases} n(n-1)(y-x)^{n-2} & \text{for } 0<x<y<1 \\ 0 & \text{otherwise.} \end{cases}$$

Therefore,

$$\begin{aligned} P\left(X<\frac{1}{2}<Y\right) &= \int_{1/2}^{1}\int_{0}^{1/2} n(n-1)(y-x)^{n-2}dxdy \\ &= \int_{1/2}^{1} n\left[-\left.(y-x)^{n-1}\right|_0^{1/2}\right]dy \\ &= \int_{1/2}^{1} ny^{n-1}dy - \int_{1/2}^{1} n\left(y-\frac{1}{2}\right)^{n-1}dy \\ &= \left.y^n\right|_{1/2}^{1} - \left.\left(y-\frac{1}{2}\right)^n\right|_{1/2}^{1} = 1-\frac{1}{2^n}-\frac{1}{2^n} = 1-\frac{1}{2^{n-1}}. \end{aligned}$$

The marginal pdf of X is

$$f_X(x) = \begin{cases} n(1-x)^{n-1} & \text{for } 0<x<1 \\ 0 & \text{otherwise.} \end{cases}$$

Therefore,

$$E\left[X^3\right] = \int_0^1 x^3 n(1-x)^{n-1}dx = n\int_0^1 x^3(1-x)^{n-1}dx$$

The integral on the right is $B(4,n)$, where $B(\cdot,\cdot)$ is the beta function. So

$$E\left[X^3\right] = nB(4,n) = n\frac{\Gamma(4)\Gamma(n)}{\Gamma(n+4)} = \frac{6n}{(n+3)(n+2)(n+1)n} = \frac{6}{(n+3)(n+2)(n+1)}.$$

(c) Let $X_1, \ldots, X_n, Y_1, \ldots, Y_n$ be independent Uniform $(0,1)$ random variables. We form n rectangles, where the i-th rectangle has adjacent sides of length X_i and Y_i, for $i = 1, \ldots, n$. Let A_i be the area of the i th rectangle, $i = 1, \ldots, n$, and define $A_{\max} = \max(A_1, \ldots, A_n)$. Find the pdf of $A_{\max}$.

Solution: First note that $A_i = X_iY_i$ for $i = 1, \ldots, n$, and $A_1, \ldots, A_n$ are independent since the vectors $(X_1, Y_1), \ldots, (X_n, Y_n)$ are independent. Also, the A_i are identically distributed, each with support $(0,1)$. We first find $P(A_i \le a)$ for any $a \in (0,1)$, which is $P(X_iY_i \le a)$. If $Y_i \le a$ then the limits for X_i are from 0 to 1. If $Y_i > a$ then the limits for X_i are from 0 to a/Y_i. The joint pdf of (X_i, Y_i) is $f(x,y) = 1$ for $0 < x, y < 1$. Thus we have

$$\begin{aligned} P(A_i \le a) = P(X_iY_i \le a) &= \int_0^a\int_0^1 dxdy + \int_a^1\int_0^{a/y} dxdy \\ &= a + a\int_a^1 \frac{1}{y}dy = a + a(\ln 1 - \ln a) = a(1-\ln a). \end{aligned}$$

Next, we get the cdf of $A_{\max}$ as

$$F(a) = P(A_{\max} \le a) = P(A_1 \le a, \ldots, A_n \le a) = P(A_1 \le a)^n = (a(1-\ln a))^n,$$

for $0 < a < 1$. Finally, we obtain the pdf f of $A_{\max}$ by differentiation:

$$f(a) = F'(a) = n(a(1 - \ln a))^{n-1}[1 - \ln a - 1] = -n \ln a (a(1 - \ln a))^{n-1},$$

which is valid for $a \in (0,1)$, and $f(a) = 0$ for $a \le 0$ or $a \ge 0$.

(d) Let $X_1, \ldots, X_n$ be independent and identically distributed Exponential (λ) random variables. Compute $E\left[X_{(1)} e^{-\lambda X_{(2)}}\right]$, where $X_{(1)}$ and $X_{(2)}$ are the first and second order statistics of $X_1, \ldots, X_n$.

Solution: The joint pdf of $X_{(1)}$ and $X_{(2)}$ is

$$f_{12}(x_1, x_2) = \begin{cases} n(n-1)\lambda^2 e^{-\lambda x_1} e^{-\lambda x_2} \left(e^{-\lambda x_2}\right)^{n-2} & \text{for } 0 < x_1 < x_2 < \infty \\ 0 & \text{otherwise} \end{cases}$$

Then

$$\begin{aligned} E\left[X_{(1)} e^{-\lambda X_{(2)}}\right] &= n(n-1)\lambda^2 \int_0^\infty \int_{x_1}^\infty x_1 e^{-\lambda x_2} e^{-\lambda x_1} e^{-\lambda(n-1)x_2} dx_2 dx_1 \\ &= (n-1)\lambda \int_0^\infty x_1 e^{-\lambda x_1} e^{-\lambda n x_1} dx_1 \\ &= \frac{n-1}{n+1} \int_0^\infty x_1 \lambda(n+1) e^{-\lambda(n+1)x_1} dx_1 \\ &= \frac{n-1}{\lambda(n+1)^2} \end{aligned}$$

since the final integral is the mean of an Exponential$(\lambda(n+1))$ distribution.

(e) Let X_1, X_2, X_3 be independent Uniform $(0,1)$ random variables, and let $X_{(1)}, X_{(2)}, X_{(3)}$ denote their order statistics. Let X be the area of the square with side length $X_{(2)}$ and let Y be the area of the rectangle with side lengths $X_{(1)}$ and $X_{(3)}$. Find $P(X > Y)$, $E[X]$, and $E[Y]$.

Solution: We wish to compute $P\left(X_{(2)}^2 > X_{(1)} X_{(3)}\right)$. The joint pdf of $\left(X_{(1)}, X_{(2)}, X_{(3)}\right)$ is $f_{123}(x_1, x_2, x_3) = 6$ for $0 < x_1 < x_2 < x_3 < 1$ and equals 0 otherwise. The solution is computed by integrating $f_{123}(x_1, x_2, x_3)$ over the region $A = \left\{(x_1, x_2, x_2) : x_2^2 - x_1 x_3 > 0\right\}$. If we set the inner integral over x_2 and note that for given values of $0 < x_1 < x_3 < 1$, the possible values of x_2 are $\sqrt{x_1 x_3} < x_2 < x_3$. The integral can be written as

$$\begin{aligned} \iiint_A f_{123}(x_1, x_2, x_3)\, dx_1 dx_2 dx_3 &= 6 \int_0^1 \int_0^{x_3} \int_{\sqrt{x_1 x_3}}^{x_3} dx_2 dx_1 dx_3 \\ &= 6 \int_0^1 \int_0^{x_3} \left(x_3 - x_1^{1/2} x_3^{1/2}\right) dx_1 dx_3 \\ &= 6 \int_0^1 \left(x_3^2 - x_3^{1/2} \frac{2}{3} x_3^{3/2}\right) dx_3 \\ &= 2 \int_0^1 x_3^2 dx_3 = \frac{2}{3}. \end{aligned}$$

The marginal pdf of $X_{(2)}$ and the joint pdf of $\left(X_{(1)}, X_{(3)}\right)$ are given by

$$f_2(x_2) = \begin{cases} 6x_2(1 - x_2) & 0 < x_2 < 1 \\ 0 & \text{otherwise} \end{cases}$$

and

$$f_{13}(x_1, x_3) = \begin{cases} 6(x_3 - x_1) & 0 < x_1 < x_3 < 1 \\ 0 & \text{otherwise} \end{cases}$$

respectively. Then

$$\begin{aligned} E[X] &= E\left[X_{(2)}^2\right] = \int_0^1 x_2^2 6x_2(1 - x_2)\, dx_2 = \frac{6}{4} - \frac{6}{5} = \frac{3}{10} \\ E[Y] &= E\left[X_{(1)} X_{(3)}\right] = \int_0^1 \int_0^{x_3} x_1 x_3 6(x_3 - x_1)\, dx_1 dx_3 = 6 \int_0^1 \left(\frac{x_3^4}{2} - \frac{x_3^4}{3}\right) dx_3 = \frac{1}{5}. \end{aligned}$$

Problem 3. (Convergence of Random Variables and Limit Theorems)

(a) Let X and $X_1, X_2, \dots$ be random variables each having a $N(0,1)$ distribution. Suppose (X_n, X) has a bivariate normal distribution for each n and the correlation between X_n and X is $\rho(X_n, X) = \rho_n$, for $n \geq 1$.

(i) Show that X_n converges to X in distribution as $n \to \infty$ (for arbitrary correlations ρ_n).

Solution: If Φ is the standard normal cdf and F_n is the cdf of X_n for $n \geq 1$, then $F_n(x) = \Phi(x)$ for all x (this is given). Therefore, $F_n(x) \to \Phi(x)$ as $n \to \infty$ (trivially). It is also given that $\Phi(x)$ is the cdf of X. Therefore, X_n converges to X in distribution.

(ii) If $\rho_n \to 1$ as $n \to \infty$, show that X_n converges to X in probability as $n \to \infty$.

Solution: Let $\epsilon > 0$ be given. Both X_n and X are zero mean so $X_n - X$ also has zero mean. So by Chebyshev'e inequality,

$$P(|X_n - X| > \epsilon) \leq \frac{\operatorname{Var}(X_n - X)}{\epsilon^2}. \tag{1}$$

But

$$\operatorname{Var}(X_n - X) = \operatorname{Var}(X_n) + \operatorname{Var}(X) - 2\operatorname{Cov}(X_n, X) = 1 + 1 - 2\rho_n = 2(1 - \rho_n). \tag{2}$$

So if $\rho_n \to 1$ as $n \to \infty$ then $P(|X_n - X| > \epsilon) \to 0$ as $n \to \infty$ for any $\epsilon > 0$. Thus, X_n converges to X in probability.

(iii) Show that if $\rho_n = 1 - a^n$ for some constant $a \in (0,1)$, then X_n converges to X with probability 1 as $n \to \infty$. Do you get convergence with probability 1 if $a = 0$? If $a = 1$? Prove your answers.

Solution: Plugging $\rho_n = 1 - a^n$ into (2) we have $\operatorname{Var}(X_n - X) = 2a^n$. Plugging this into Chebyshev's inequality in (1) we have

$$P(|X_n - X| > \epsilon) \leq \frac{2a^n}{\epsilon^2}.$$

If $a \in (0,1)$ the sum on the right converges and so X_n converges to X with probability 1 by the sufficient condition from class. If $a = 0$ the sum is again convergent (it is equal to 0) and so X_n converges to X with probability 1 . If $a = 1$ then $\rho_n = 0$ so that X_n is independent of X for any n (since (X_n, X) is bivariate normal). In this case $X_n - X$ has a $N(0,2)$ distribution and so $P(|X_n - X| > \epsilon) = 2(1 - \Phi(\epsilon/\sqrt{2}))$, where Φ is the standard normal cdf. This is some positive constant for every n and so $P(|X_n - X| > \epsilon)$ does not converge to 0 as $n \to \infty$. Therefore, X_n does not converge to X in probability. This then implies that X_n does not converge to X with probability 1.

(b) Suppose 80 points are uniformly distributed in the ball in $\mathbb{R}^3$ centred at the origin with radius 1. Let D_i be the Euclidean distance from the origin of the i th point, $i = 1, \dots, 80$, and let $\bar{D} = \frac{1}{80}\sum_{i=1}^{80} D_i$. Use the central limit theorem to find a value c satisfying $P(|\bar{D} - E[\bar{D}]| \leq c) = 0.95$.

Solution: We first find the distribution of D_i. For $x \in (0,1), D_i \leq x$ if and only if the i-th point in the ball is in the ball in $\mathbb{R}^3$ centred at the origin of radius x. That is, $P(D_i \leq x) = \frac{4\pi x^3/3}{4\pi/3} = x^3$. From this we have that the pdf of D_i is

$$f_D(x) = \begin{cases} 3x^2 & \text{for } x \in [0,1] \\ 0 & \text{otherwise.} \end{cases}$$

The mean, second moment, and variance are then easily computed to be

$$E[D_i] = \frac{3}{4}, \quad E\left[D_i^2\right] = \frac{3}{5}, \quad \operatorname{Var}(D_i) = \frac{3}{80}.$$

By the central limit theorem

$$\begin{aligned}P(|\bar{D}-3/4|\le c) &= P\left(\frac{\sqrt{80}|\bar{D}-3/4|}{\sqrt{3/80}}\le\frac{c\sqrt{80}}{\sqrt{3/80}}\right)\\ &\approx P\left(|Z|\le\frac{c\sqrt{80}}{\sqrt{3/80}}\right),\end{aligned}$$

where Z has a $N(0,1)$ distribution. In order for this probability to be equal to 0.95 we need $\frac{c\sqrt{80}}{\sqrt{3/80}}$ equal to 1.96. Solving for c gives

$$c=\frac{1.96\sqrt{3}}{80}\approx 0.0424.$$

(c) (i) Suppose that $\{X_n\}$ is a sequence of zero-mean random variables and X is a zero mean random variable, and suppose that $E\left[(X_n-X)^2\right]\le C/n^p$ for every n, for some constants C and $p>1$. Show that $X_n\to X$ almost surely.

Solution: Let $\epsilon>0$ be given. Since $E[X_n-X]=0$ we have by Chebyshev's inequality that

$$P(|X_n-X|>\epsilon)\le\frac{\mathrm{Var}(X_n-X)}{\epsilon^2}=\frac{E\left[(X_n-X)^2\right]}{\epsilon^2}\le\frac{C}{\epsilon^2 n^p}.$$

Therefore,

$$\sum_{n=1}^{\infty}P(|X_n-X|>\epsilon)\le\sum_{n=1}^{\infty}\frac{C}{\epsilon^2 n^p}<\infty$$

if $p>1$. By a sufficient condition from class this implies that $X_n\to X$ almost surely.

(ii) Suppose that $\{X_n\}$ is a sequence of non-negative random variables. Show that $E[X_n]\to 0$ as $n\to\infty$ implies that $X_n\to 0$ in probability, but that the converse is false in general.

Solution: Suppose that $E[X_n]\to 0$. Let $\epsilon>0$ be given. Then by Markov's inequality

$$P(|X_n-0|>\epsilon)\le\frac{E[X_n]}{\epsilon}\to 0$$

as $n\to\infty$. Hence, $X_n\to 0$ in probability. To prove that the reverse implication is false in general we give a counterexample. Let the distribution of X_n be given by

$$X_n=\begin{cases}0 & \text{with probability } 1-\frac{1}{n}\\ n & \text{with probability } \frac{1}{n}.\end{cases}$$

Then if $\epsilon>0$ is given $P(|X_n-0|>\epsilon)=P(X_n=n)=\frac{1}{n}\to 0$ as $n\to\infty$, so $X_n\to 0$ in probability. However, $E[X_n]=n\left(\frac{1}{n}\right)=1$ for all n, which does not converge to 0.

(d) Suppose that $\{X_n\}$ and $\{Y_n\}$ are sequences of random variables and X and Y are random variables such that $X_n\to X$ in distribution and $Y_n\to Y$ in distribution. Give an example where it is not true that X_n+Y_n converges to $X+Y$ in distribution. [Hint: Consider $Y=-X$]

Solution: Let W and Z be independent $N(0,1)$ random variables. Let $X_n=W$ for all n and $Y_n=Z$ for all n. Let $X=W$ and $Y=-W$. It is easy to see that X and Y are both $N(0,1)$, and it is clear that $X_n\to X$ in distribution and $Y_n\to Y$ in distribution. But $X+Y=0$ whereas X_n+Y_n is distributed as $N(0,2)$ for all n. Therefore, it is not true that X_n+Y_n converges to $X+Y$ in distribution.

(e) Consider the probability

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} F\left(X_i\right)-\frac{1}{2}\right| \geq \frac{1}{\sqrt{3n}}\right).$$

Use Chebyshev's inequality to bound this probability. Use the central limit theorem to approximate this probability for large n.

Solution: The random variable $\frac{1}{n}\sum_{i=1}^{n} F\left(X_i\right)$ has mean $\frac{1}{2}$ and variance $\frac{1}{12n}$. Applying Chebyshev's inequality, we have

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} F\left(X_i\right)-\frac{1}{2}\right| \geq \frac{1}{\sqrt{3n}}\right) \leq \frac{\operatorname{Var}\left(\frac{1}{n}\sum_{i=1}^{n} F\left(X_i\right)\right)}{(1/\sqrt{3n})^2} = \frac{1/12n}{1/3n} = \frac{1}{4}.$$

By the central limit theorem, $Z_n = \frac{\frac{1}{n}\sum_{i=1}^{n} F(X_i)-1/2}{\sqrt{1/12n}}$ has approximately a $N(0,1)$ distribution for large n. Therefore,

$$\begin{aligned} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} F\left(X_i\right)-\frac{1}{2}\right| \geq \frac{1}{\sqrt{3n}}\right) &= P\left(\left|\frac{1}{n}\sum_{i=1}^{n} F\left(X_i\right)-\frac{1}{2}\right|\sqrt{12n} \geq \frac{\sqrt{12n}}{\sqrt{3n}}\right) \\ &= P\left(\left|Z_n\right| \geq 2\right) \approx 0.0456. \end{aligned}$$

Problem 4. (Moment-generating function)

(a) Let $X \sim \text{Binomial}(n,p)$. Find the moment-generating function of X in terms of n and p. (Hint: X is the sum of n independent Bernoulli random variables.)

Solution: X is the sum of n independent Bernoulli random variables $X_1, \ldots, X_n$, each with moment generating function

$$M_{X_i}(t) = \mathbb{E}\exp\left(tX_i\right) = pe^t + (1-p).$$

Combining these and applying independence yields

$$M_X(t) = \mathbb{E}\exp(tX) = \mathbb{E}\exp\left(t\left(X_1+\ldots+X_n\right)\right) = \prod_{i=1}^{n}\mathbb{E}\exp\left(tX_i\right) = \left(1-p+pe^t\right)^n.$$

(b) Let $X_1, X_2, \ldots$ be independent and identically distributed random variables, each with a Poisson distribution with mean 1. Let $S_n = X_1+\ldots+X_n$ for $n \geq 1$ and let $M_n(t)$ be the moment generating function of S_n.

(i) Find the smallest n such that $P\left(M_n\left(S_n\right) > 1\right) \geq 0.99$ using the exact probability.

Solution: First, the common moment generating function of the X_i is

$$M(t) = E\left[e^{tX_i}\right] = \sum_{k=0}^{\infty} e^{tk}\frac{1}{k!}e^{-1} = e^{-1}\sum_{k=0}^{\infty}\frac{\left(e^t\right)^k}{k!} = e^{-1}e^{e^t} = e^{e^t-1}.$$

Then $M_n(t) = M(t)^n = e^{n\left(e^t-1\right)}$ and $M_n\left(S_n\right) = e^{n\left(e^{S_n}-1\right)}$. So,

$$P\left(M_n\left(S_n\right)>1\right) = P\left(e^{n\left(e^{S_n}-1\right)}>1\right) = P\left(n\left(e^{S_n}-1\right)>0\right) = P\left(e^{S_n}>1\right) = P\left(S_n>0\right).$$

The exact distribution of S_n is Poisson(n), so $P\left(S_n>0\right) = 1-P\left(S_n=0\right) = 1-e^{-n}$. Setting this to .99 gives $e^{-n} = .01$, or $n = \ln 100 = 4.605$. So $n = 5$ is the smallest n.

(ii) Find the smallest n such that $P\left(M_n\left(S_n\right) > 1\right) \geq 0.99$ using the central limit theorem.

Solution: By the central limit theorem, the distribution of $\left(S_n-n\right)/\sqrt{n}$ is approximated by a $N(0,1)$ distribution, so that (from part i),

$$P\left(M_n\left(S_n\right)>1\right) = P\left(S_n>0\right) = P\left(\left(S_n-n\right)/\sqrt{n} > -\sqrt{n}\right) \approx 1-\Phi(-\sqrt{n}),$$

where $\Phi(\cdot)$ is the standard normal cdf. Setting this to 0.99 gives $\Phi(-\sqrt{n}) = 0.01$, or (from the table) $-\sqrt{n} = -2.33$, or $n = 5.429$. So $n = 6$ is the smallest n according to this approximation.

(c) Let X have a Gamma $(3,3)$ distribution. Conditional on $X = x$ let Z have a normal distribution with mean x and variance 2. Finally, let $Y = e^Z$. Find $E[Y]$ and $\operatorname{Var}(Y)$.

Solution: Conditioned on $X = x, E[Y]$ is the moment generating function of a $N(x,2)$ distribution evaluated at 1 . Letting $M_x(\cdot)$ denote this mgf, we have $M_x(t) = e^{xt+t^2}$, and so $E[Y \mid X = x] = M_x(1) = e^{x+1}$. Similarly, $E\left[Y^2 \mid X = x\right] = E\left[e^{2Z} \mid X = x\right] = M_x(2) = e^{2x+4}$. Then by the law of total expectation, $E[Y] = E\left[e^{X+1}\right] = eM_X(1)$ and $E\left[Y^2\right] = E\left[e^{2X+4}\right] = e^4 M_X(2)$, where $M_X(\cdot)$ is the moment generating function of X. Since X has a Gamma $(3,3)$ distribution we have $M_X(t) = \left(\frac{3}{3-t}\right)^3$, and so

$$E[Y] = e\left(\frac{3}{2}\right)^3 = \frac{27e}{8} \approx 9.174 \quad \text{and} \quad E\left[Y^2\right] = 27e^4.$$

Then

$$\operatorname{Var}(Y) = E\left[Y^2\right] - E[Y]^2 = 27e^4 - \left(\frac{27}{8}\right)^2 e^2 \approx 1390.$$

(d) Let X be a random variable with Exponential(λ) distribution. Recall that the moment generating function of X is $M_X(t) = \frac{\lambda}{\lambda - t}$ for $t < \lambda$.

(i) Find $E\left[X^n\right]$, where n is any positive integer. You may use the mgf or compute $E\left[X^n\right]$ directly.

Solution: Using the mgf, we write the mgf as

$$M_X(t) = \frac{1}{1 - t/\lambda} = \sum_{n=0}^{\infty}\left(\frac{t}{\lambda}\right)^n$$

which is valid for $|t| < \lambda$. Comparing to $M_X(t) = \sum_{n=0}^{\infty} \frac{E[X^n]}{n!} t^n$ (which is the MacLaurin series expansion of $M_X(t)$ when it is valid (it is valid for all $t < \lambda$)), we get that $E\left[X^n\right] = \frac{n!}{\lambda^n}$. Computing $E\left[X^n\right]$ directly, we have

$$E\left[X^n\right] = \int_0^{\infty} x^n \lambda e^{-\lambda x} dx = \lambda \frac{\Gamma(n+1)}{\lambda^{n+1}} \int_0^{\infty} \frac{\lambda^{n+1}}{\Gamma(n+1)} x^{(n+1)-1} e^{-\lambda x} dx = \frac{n!}{\lambda^n},$$

since the integrand in the last integral is a Gamma$(n+1, \lambda)$ density and so the integral is 1, and $\Gamma(n+1) = n!$.

(ii) Find $M_Y(t)$, the mgf of $Y = \ln X$, for $t > -1$.

Solution: Write the mgf of Y as $M_Y(t) = E\left[e^{tY}\right] = E\left[e^{t\ln X}\right] = E\left[X^t\right]$. Computing $E\left[X^t\right]$ directly we get an integral similar to the one in part(a) with n replaced by t. For $t > -1$ we get a proper Gamma$(t+1, \lambda)$ density in the final integral and so we get

$$M_Y(t) = E\left[X^t\right] = \frac{\Gamma(t+1)}{\lambda^t}.$$

(e) Let $X_1, \ldots, X_n$ be independent Poisson(1) random variables and let $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$. Let $k > 1$ be given. Show that

$$P(\bar{X} \geq k) \leq \left(\frac{e^{k-1}}{k^k}\right)^n.$$

Solution: We find the mgf of $\bar{X}$ and then apply Chernoff's bound. The mgf of each X_i is

$$M_{X_i}(t) = E\left[e^{tX_i}\right] = \sum_{j=0}^{\infty} e^{tj} \frac{1}{j!} e^{-1} = e^{-1} e^{e^t} = e^{e^t - 1}.$$

Then the mgf of $\bar{X}$ is given by

$$M_X(t) = E\left[e^{t\bar{X}}\right] = M_{X_i}\left(\frac{t}{n}\right)^n = e^{n\left(e^{t/n}-1\right)}.$$

Chernoff's bound gives

$$P(\bar{X} > k) \leq \min_{t>0} \frac{e^{n\left(e^{t/n}-1\right)}}{e^{tk}} = e^{n\left(e^{t/n}-1\right)-tk}. \tag{3}$$

Differentiating the exponent with respect to t gives $e^{t/n} - k$. Setting this to 0 and solving for t gives $t = n \ln k$. Plugging this into the RHS of Eqn. (3) then gives

$$P(\bar{X} > k) \leq e^{n(k-1)-\ln k^{kn}} = \frac{e^{n(k-1)}}{k^{kn}} = \left(\frac{e^{k-1}}{k^k}\right)^n,$$

as desired.

Problem 5. (Bivariate and Multivariate Distributions)

(a) **Accounting for voter turnout.** Let N be the number of people in the state of Iowa. Suppose pN of these people support Hillary Clinton, and $(1-p)N$ of them support Donald Trump, for some $p \in (0,1)$. N is known (say $N = 3{,}000{,}000$) and p is unknown.

(i) Suppose that each person in Iowa randomly and independently decides, on election day, whether or not to vote, with probability 1/2 of voting and probability 1/2 of not voting. Let V_{Hillary} be the number of people who vote for Hillary and V_{Donald} be the number of people who vote for Donald. Show that

$$\mathbb{E}\left[V_{\text{Hillary}}\right] = \frac{1}{2}pN, \quad \mathbb{E}\left[V_{\text{Donald}}\right] = \frac{1}{2}(1-p)N.$$

What are the standard deviations of V_{Hillary} and V_{Donald}, in terms of p and N? Explain why, when N is large, we expect the fraction of voters who vote for Hillary to be very close to p.

Solution: Recall that a Binomial(n,p) random variable has mean np and variance $np(1-p)$. A total of pN people support Hillary, each voting independently with probability $\frac{1}{2}$, so $V_{\text{Hillary}} \sim \text{Binomial}\left(pN, \frac{1}{2}\right)$. Then

$$\mathbb{E}\left[V_{\text{Hillary}}\right] = \frac{1}{2}pN, \quad \text{Var}\left[V_{\text{Hillary}}\right] = \frac{1}{4}pN,$$

and the standard deviation of V_{Hillary} is $\sqrt{\frac{1}{4}pN}$. Similarly, as $(1-p)N$ people support Donald, $V_{\text{Donald}} \sim \text{Binomial}\left((1-p)N, \frac{1}{2}\right)$, so

$$\mathbb{E}\left[V_{\text{Donald}}\right] = \frac{1}{2}(1-p)N, \quad \text{Var}\left[V_{\text{Donald}}\right] = \frac{1}{4}(1-p)N,$$

and the standard deviation of V_{Donald} is $\sqrt{\frac{1}{4}(1-p)N}$. The fraction of voters who vote for Hillary is

$$\frac{V_{\text{Hillary}}}{V_{\text{Hillary}} + V_{\text{Donald}}} = \frac{V_{\text{Hillary}}/N}{V_{\text{Hillary}}/N + V_{\text{Donald}}/N}.$$

As $\mathbb{E}\left[V_{\text{Hillary}}/N\right] = \frac{1}{2}p$ (a constant) and $\text{Var}\left[V_{\text{Hillary}}/N\right] = \frac{1}{4}p/N \to 0$ as $N \to \infty$, V_{Hillary}/N should be close to $\frac{1}{2}p$ with high probability when N is large. Similarly, as $\mathbb{E}\left[V_{\text{Donald}}/N\right] = \frac{1}{2}(1-p)$ and $\text{Var}\left[V_{\text{Donald}}/N\right] = \frac{1}{4}(1-p)/N \to 0$ as $N \to \infty$, V_{Donald}/N should be close to $\frac{1}{2}(1-p)$ with high probability when N is large. Then the fraction of voters for Hillary should, with high probability, be close to

$$\frac{\frac{1}{2}p}{\frac{1}{2}p + \frac{1}{2}(1-p)} = p.$$

(The above statements "close to with high probability" may be formalized using Chebyshev's inequality, which states that a random variable is, with high probability, not too many standard deviations away from its mean.)

(ii) Now suppose there are two types of voters - "passive" and "active". Each passive voter votes on election day with probability 1/4 and doesn't vote with probability 3/4, while each active voter votes with probability 3/4 and doesn't vote with probability 1/4. Suppose that a fraction q_H of the people who support Hillary are passive and $1-q_H$ are active, and a fraction q_D of the people who support Donald are passive and $1-q_D$ are active. Show that

$$\mathbb{E}\left[V_{\text{Hillary}}\right]=\frac{1}{4} q_H p N+\frac{3}{4}\left(1-q_H\right) p N, \quad \mathbb{E}\left[V_{\text{Donald}}\right]=\frac{1}{4} q_D(1-p) N+\frac{3}{4}\left(1-q_D\right)(1-p) N .$$

What are the standard deviations of V_{Hillary} and V_{Donald} in terms of p, N, q_H, and q_D? If we estimate p by $\hat{p}$ using a simple random sample of $n=1000$ people from Iowa. Explain why $\hat{p}$ might not be a good estimate of the fraction of voters who will vote for Hillary.

Solution: Let $V_{\mathrm{H,p}}$ and $V_{\mathrm{H,a}}$ be the number of passive and active voters who vote for Hillary, and similarly define $V_{\mathrm{D,p}}$ and $V_{\mathrm{D,a}}$ for Donald. There are $q_H p N$ passive Hillary supporters, each of whom vote independently with probability $\frac{1}{4}$, so

$$V_{\mathrm{H,p}} \sim \operatorname{Binomial}\left(q_H p N, \frac{1}{4}\right) .$$

Similarly,

$$\begin{aligned}
& V_{\mathrm{H,a}} \sim \operatorname{Binomial}\left(\left(1-q_H\right) p N, \frac{3}{4}\right), \\
& V_{\mathrm{D,p}} \sim \operatorname{Binomial}\left(q_D(1-p) N, \frac{1}{4}\right), \\
& V_{\mathrm{D,a}} \sim \operatorname{Binomial}\left(\left(1-q_D\right)(1-p) N, \frac{3}{4}\right),
\end{aligned}$$

and these four random variables are independent. Since $V_{\text{Hillary}}=V_{\mathrm{H,p}}+V_{\mathrm{H,a}}$,

$$\begin{gathered}
\mathbb{E}\left[V_{\text{Hillary}}\right]=\mathbb{E}\left[V_{\mathrm{H,p}}\right]+\mathbb{E}\left[V_{\mathrm{H,a}}\right]=\frac{1}{4} q_H p N+\frac{3}{4}\left(1-q_H\right) p N, \\
\operatorname{Var}\left[V_{\text{Hillary}}\right]=\operatorname{Var}\left[V_{\mathrm{H,p}}\right]+\operatorname{Var}\left[V_{\mathrm{H,a}}\right]=\frac{3}{16} q_H p N+\frac{3}{16}\left(1-q_H\right) p N=\frac{3}{16} p N,
\end{gathered}$$

and the standard deviation of V_{Hillary} is $\sqrt{\frac{3}{16} p N}$. Similarly,

$$\begin{aligned}
\mathbb{E}\left[V_{\text{Donald}}\right] & =\mathbb{E}\left[V_{\mathrm{D,p}}\right]+\mathbb{E}\left[V_{\mathrm{D,a}}\right]=\frac{1}{4} q_D(1-p) N+\frac{3}{4}\left(1-q_D\right)(1-p) N, \\
\operatorname{Var}\left[V_{\text{Donald}}\right] & =\operatorname{Var}\left[V_{\mathrm{D,p}}\right]+\operatorname{Var}\left[V_{\mathrm{D,a}}\right]=\frac{3}{16} q_D(1-p) N+\frac{3}{16}\left(1-q_D\right)(1-p) N \\
& =\frac{3}{16}(1-p) N,
\end{aligned}$$

and the standard deviation of V_{Donald} is $\sqrt{\frac{3}{16}(1-p) N}$.

The quantity $\hat{p}$ estimates p, but in this case p may not be the fraction of voters who vote for Hillary: By the same argument as in part (a), the fraction of voters who vote for Hillary is given by

$$\begin{aligned}
\frac{V_{\text{Hillary}}}{V_{\text{Hillary}}+V_{\text{Donald}}} & =\frac{V_{\text{Hillary}} / N}{V_{\text{Hillary}} / N+V_{\text{Donald}} / N} \\
& \approx \frac{\frac{1}{4} q_H p+\frac{3}{4}\left(1-q_H\right) p}{\frac{1}{4} q_H p+\frac{3}{4}\left(1-q_H\right) p+\frac{1}{4} q_D(1-p)+\frac{3}{4}\left(1-q_D\right)(1-p)},
\end{aligned}$$

where the approximation is accurate with high probability when N is large. When $q_H \neq q_D$, this is different from p : For example, if $q_H=0$ and $q_D=1$, this is equal to $\frac{p}{p+(1-p)/3}$ which is greater than p, reflecting the fact that Hillary supporters are more likely to vote than are Donald supporters.

(iii) We do not know q_H and q_D. However, suppose that in our simple random sample, we can observe whether each person is passive or active, in addition to asking them whether they support Hillary or Donald. Suggest estimators $\hat{V}_{\text{Hillary}}$ and $\hat{V}_{\text{Donald}}$ for $\mathbb{E}[V_{\text{Hillary}}]$ and $\mathbb{E}[V_{\text{Donald}}]$ using this additional information. Show, for your estimators, that

$$\mathbb{E}\left[\hat{V}_{\text{Hillary}}\right] = \frac{1}{4}q_H pN + \frac{3}{4}(1-q_H)\,pN, \quad \mathbb{E}\left[\hat{V}_{\text{Donald}}\right] = \frac{1}{4}q_D(1-p)N + \frac{3}{4}(1-q_D)(1-p)N.$$

Solution: Let $\hat{p}$ be the proportion of the 1000 surveyed people who support Hillary. Among the surveyed people supporting Hillary, let $\hat{q}_H$ be the proportion who are passive. Similarly, among the surveyed people supporting Donald, let $\hat{q}_D$ be the proportion who are passive. (Note that these are observed quantities, computed from our sample of 1000 people.) Then we may estimate the number of voters for Hillary and Donald by

$$\begin{aligned}
\hat{V}_{\text{Hillary}} &= \frac{1}{4}\hat{q}_H\hat{p}N + \frac{3}{4}(1-\hat{q}_H)\,\hat{p}N \\
\hat{V}_{\text{Donald}} &= \frac{1}{4}\hat{q}_D(1-\hat{p})N + \frac{3}{4}(1-\hat{q}_D)(1-\hat{p})N.
\end{aligned}$$

$\hat{q}_H\hat{p}$ is simply the proportion of the 1000 surveyed people who both support Hillary and are passive. Hence, letting $X_1, \ldots, X_{1000}$ indicate whether each surveyed person both supports Hillary and is passive, we have

$$\hat{q}_H\hat{p} = \frac{1}{n}(X_1 + \ldots + X_n).$$

Each $X_i \sim \text{Bernoulli}(q_H p)$, so linearity of expectation implies $\mathbb{E}[\hat{q}_H\hat{p}] = q_H p$. Similarly, $(1-\hat{q}_H)\,\hat{p}, \hat{q}_D(1-\hat{p})$, and $(1-\hat{q}_D)(1-\hat{p})$ are the proportions of the 1000 surveyed people who support Hillary and are active, support Donald and are passive, and support Donald and are active, so the same argument shows $\mathbb{E}[(1-\hat{q}_H)\,\hat{p}] = (1-q_H)\,p, \mathbb{E}[\hat{q}_D(1-\hat{p})] = q_D(1-p)$, and $\mathbb{E}[(1-\hat{q}_D)(1-\hat{p})] = (1-q_D)(1-p)$. Then applying linearity of expectation again yields

$$\mathbb{E}\left[\hat{V}_{\text{Hillary}}\right] = \mathbb{E}[V_{\text{Hillary}}], \quad \mathbb{E}\left[\hat{V}_{\text{Donald}}\right] = \mathbb{E}[V_{\text{Donald}}].$$

(b) Let X and Y be continuous random variables with joint density function

$$f(x,y) = \begin{cases} \frac{y^3}{2}e^{-y(x+1)} & \text{for } x > 0, y > 0 \\ 0 & \text{otherwise.} \end{cases}$$

(i) Find the marginal pdf of X.

Solution: Integrating the joint pdf over y gives the marginal pdf of X. For $x > 0$, we have

$$f_X(x) = \int_0^\infty \frac{y^3}{2}e^{-y(x+1)}dy = \frac{\Gamma(4)}{2(x+1)^4}\int_0^\infty \frac{(x+1)^4}{\Gamma(4)}y^3 e^{-y(x+1)}dy = \frac{3}{(x+1)^4},$$

as the integrand in the last integral over y is a Gamma$(4, x+1)$ pdf, and so integrates (over y) to 1. For $x \le 0, f_X(x) = 0$.

(ii) Find the marginal pdf of Y and $E[Y]$.

Solution: Integrating the joint pdf over x gives the marginal pdf of Y. For $y > 0$, we have

$$f_Y(y) = \int_0^\infty \frac{y^3}{2}e^{-y(x+1)}dx = \frac{y^2}{2}e^{-y}\int_0^\infty ye^{-yx}dx = \frac{y^2}{2}e^{-y}$$

as the integrand in the last integral over x is an Exponential (y) pdf, and so integrates (over x) to 1 . For $y \le 0, f_Y(y) = 0$. We can recognize the marginal distribution of Y to be Gamma$(3,1)$, and so $E[Y] = 3/1 = 3$.

(c) **Existence of multivariate normal.**

(i) Suppose $(X_1, \ldots, X_k) \sim \mathcal{N}(0, \Sigma)$ for a covariance matrix $\Sigma \in \mathbb{R}^{k \times k}$. Let $Y_1, \ldots, Y_m$ be linear combinations of $X_1, \ldots, X_k$, given by

$$Y_j = a_{j1}X_1 + \ldots + a_{jk}X_k$$

for each $j = 1, \ldots, m$ and some constants $a_{j1}, \ldots, a_{jk} \in \mathbb{R}$. Consider the matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mk} \end{pmatrix}.$$

By computing the means, variances, and covariances of $Y_1, \ldots, Y_m$, show that

$$(Y_1, \ldots, Y_m) \sim \mathcal{N}\left(0, A\Sigma A^T\right).$$

Solution: Any linear combination of $Y_1, \ldots, Y_m$ is a linear combination of $X_1, \ldots, X_k$, so $(Y_1, \ldots, Y_m)$ is multivariate normal. Using linearity of expectation and bilinearity of covariance, we compute

$$\begin{aligned} \mathbb{E}[Y_i] &= a_{i1}\mathbb{E}[X_1] + \ldots + a_{ik}\mathbb{E}[X_k] = 0, \\ \operatorname{Cov}[Y_i, Y_j] &= \operatorname{Cov}[a_{i1}X_1 + \ldots + a_{ik}X_k, a_{j1}X_1 + \ldots + a_{jk}X_k] \\ &= \sum_{r=1}^{k}\sum_{s=1}^{k} a_{ir}a_{js} \operatorname{Cov}[X_r, X_s] = \sum_{r=1}^{k}\sum_{s=1}^{k} a_{ir}a_{js}\Sigma_{rs} = a_i \Sigma a_j^T, \end{aligned}$$

where a_i and a_j denote rows i and j of the matrix A. (The computation of covariance is valid for both $i \neq j$ and $i = j$; in the latter case this yields $\operatorname{Var}[Y_i] = \operatorname{Cov}[Y_i, Y_i]$.) As $a_i \Sigma a_j^T = \left(A\Sigma A^T\right)_{ij}$, this implies by definition $Y \sim \mathcal{N}\left(0, A\Sigma A^T\right)$.

(ii) Let $A \in \mathbb{R}^{k \times k}$ be any matrix, let $\Sigma = AA^T$, and let $\mu \in \mathbb{R}^k$ be any vector. Show that there exist random variables $Y_1, \ldots, Y_k$ such that $(Y_1, \ldots, Y_k) \sim \mathcal{N}(\mu, \Sigma)$. (Hint: Let $X_1, \ldots, X_k \overset{IID}{\sim} \mathcal{N}(0, 1)$, and let each Y_j be a certain linear combination of $X_1, \ldots, X_k$ plus a certain constant.)

Solution: Take $X_1, \ldots, X_k \overset{iid}{\sim} \mathcal{N}(0, 1)$, define $Z_j = a_{j1}X_1 + \ldots + a_{jk}X_k$ for each $j = 1, \ldots, k$, and let $Y_j = Z_j + \mu_j$. As $(X_1, \ldots, X_k)$ have the multivariate normal distribution $\mathcal{N}(0, I)$ where I is the $k \times k$ identity matrix, and as $AIA^T = AA^T = \Sigma$, part (a) implies $(Z_1, \ldots, Z_k) \sim \mathcal{N}(0, \Sigma)$. Then $(Y_1, \ldots, Y_k) \sim \mathcal{N}(\mu, \Sigma)$ (since adding the vector $\mu = (\mu_1, \ldots, \mu_k)$ does not change the variances and covariances of $Y_1, \ldots, Y_k$ but shifts their means by $\mu_1, \ldots, \mu_k$).

(d) (i) Let (X, Y) be a random point uniformly distributed on the unit disk $\left\{(x, y) : x^2 + y^2 \leq 1\right\}$. Show that $\operatorname{Cov}[X, Y] = 0$, but that X and Y are not independent.

Solution: X has the same distribution as $-X$, so $\mathbb{E}[X] = \mathbb{E}[-X] = -\mathbb{E}[X]$, hence $\mathbb{E}[X] = 0$. Similarly $\mathbb{E}[Y] = 0$. Also (X, Y) has the same joint distribution as $(-X, Y)$, so $\mathbb{E}[XY] = \mathbb{E}[-XY] = -\mathbb{E}[XY]$, hence $\mathbb{E}[XY] = 0$. Then $\operatorname{Cov}[X, Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 0$. On the other hand, conditional on $X = x, Y$ is uniformly distributed on the interval $\left[-\sqrt{1 - x^2}, \sqrt{1 - x^2}\right]$. As this depends on x, X and Y are not independent.

(ii) Let $X, Y \sim \mathcal{N}(0, 1)$ be independent. Compute $\mathbb{P}[X + Y > 0 \mid X > 0]$. (Hint: Use the rotational symmetry of the bivariate normal PDF.)

Solution: We have

$$\mathbb{P}[X + Y > 0 \mid X > 0] = \frac{\mathbb{P}[X + Y > 0, X > 0]}{\mathbb{P}[X > 0]}.$$

Since X has the same distribution as $-X, \mathbb{P}[X > 0] = \mathbb{P}[-X > 0] = \mathbb{P}[X < 0]$, so $\mathbb{P}[X > 0] = \frac{1}{2}$. To compute $\mathbb{P}[X + Y > 0, X > 0]$, note that this is the integral of the bivariate normal PDF $f_{X,Y}(x, y)$ in the region to the right of the y-axis and above the line $y=-x$. The integral of $f_{X,Y}(x,y)$ over all of $\mathbb{R}^2$ must equal 1 ; hence by rotational symmetry of $f_{X,Y}(x,y)$ around the origin, the integral over any wedge formed by two rays extending from the origin is $\theta/(2\pi)$ where θ is the angle formed by these rays. For the above region, this angle is $3\pi/4$, so $\mathbb{P}[X+Y>0, X>0]=3/8$. Then $\mathbb{P}[X+Y>0 \mid X>0]=3/4$.

(e) (i) A family of Yellow-Faced (YF) gophers consisting of 2 parents and 3 children are kept in a laboratory. In addition to these a family of YF gophers with 2 parents and 4 children, a family of Big Pocket (BP) gophers with 2 parents and 5 children, and a family of BP gophers with 1 mother and 4 children are also kept in the laboratory. A sample of 4 gophers is selected at random from among all the gophers in the laboratory. What is the probability that the sample consists of one adult female, one adult male, and 2 children, with both adults of the same genus (either both YF or both BP).

Solution: Let X_1 be the number of male YF gophers, X_2 the number of female YF gophers, X_3 the number of male BP gophers, X_4 the number of female BP gophers, and X_5 the number of child gophers in the sample. The sample size is $n=4$ and the total number of gophers in the laboratory is 23 . The total numbers of YF male, YF female, BP male, BP female, and children gophers are $2,2,1,2$ and 16 respectively. The joint distribution of $(X_1, X_2, X_3, X_4, X_5)^T$ is Multivariate Hypergeometric and the desired probability is

$$
\begin{aligned}
&P\left(X_1=1, X_2=1, X_3=0, X_4=0, X_5=2\right)+P\left(X_1=0, X_2=0, X_3=1, X_4=1, X_5=2\right) \\
&=\frac{\binom{2}{1}\binom{2}{1}\binom{1}{0}\binom{2}{0}\binom{16}{2}}{\binom{23}{4}}+\frac{\binom{2}{0}\binom{2}{0}\binom{1}{1}\binom{2}{1}\binom{16}{2}}{\binom{23}{4}} \\
&=\frac{2 \times 2 \times 1 \times 1 \times 120}{8855}+\frac{1 \times 1 \times 1 \times 2 \times 120}{8855}=\frac{720}{8855} \approx 0.0813 .
\end{aligned}
$$

(ii) Let X and Y be arbitrary random variables, and let $g(\cdot)$ and $h(\cdot)$ be arbitrary real valued functions defined on $\mathbb{R}$. For each of the following statements say whether it is TRUE or FALSE. If TRUE prove it and if FALSE give a counterexample.

(I) If X and Y are uncorrelated then so are $g(X)$ and $h(Y)$ for any g and h.

(II) If $g(X)$ and $h(Y)$ are uncorrelated for all g and h then X and Y are uncorrelated.

Solution:

(I) FALSE. Take X to have a $N(0,1)$ distribution. Then $E[X]=0$ and $E\left[X^3\right]=0$, and take $Y=X^2$. Then $\operatorname{Cov}(X, Y)=\operatorname{Cov}\left(X, X^2\right)=E\left[X^3\right]-E[X] E\left[X^2\right]=0$. Take $g(X)=X^2$ and $h(Y)=Y$. Then $\operatorname{Cov}(g(X), h(Y))=\operatorname{Cov}\left(X^2, X^2\right)=\operatorname{Var}\left(X^2\right)=2>0$, where $\operatorname{Var}\left(X^2\right)$ is the variance of a χ_1^2 distribution.

(II) TRUE. Let A and B be arbitrary subsets of $\mathbb{R}$, and take $g(X)=I_A(X)$ and $h(Y)=I_B(Y)$. Since $\operatorname{Cov}(g(X), h(Y))=0$ we have $E\left[I_A(X) I_B(Y)\right]=E\left[I_A(X)\right] E\left[I_B(Y)\right]$, which is equivalent to $P(X \in A, Y \in B)=P(X \in A) P(Y \in B)$. Since A and B were arbitrary this implies that X and Y are independent. But then X and Y are uncorrelated.