Section 7
CSE312 26su

Administrivia

- Pset 5 due yesterday(late deadline tomorrow at 11:59PM)
- Pset 6 due **Wednesday Aug 12**

Content Review

Review: Joint Distributions

	Discrete	Continuous
Joint PMF/PDF	$p_{X,Y}(x,y) = \mathbb{P}(X = x, Y = y)$	$f_{X,Y}(x,y) \neq \mathbb{P}(X = x, Y = y)$
Joint range/support $\Omega_{X,Y}$	$\{(x,y) \in \Omega_X \times \Omega_Y : p_{X,Y}(x,y) > 0\}$	$\{(x,y) \in \Omega_X \times \Omega_Y : f_{X,Y}(x,y) > 0\}$
Joint CDF	$F_{X,Y}(x,y) = \sum_{t \leqslant x, s \leqslant y} p_{X,Y}(t,s)$	$F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(t,s)\, ds dt$
Normalization	$\sum_{x,y} p_{X,Y}(x,y) = 1$	$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx dy = 1$
Marginal PMF/PDF	$p_X(x) = \sum_y p_{X,Y}(x,y)$	$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) dy$
Expectation	$\mathbb{E}[g(X,Y)] = \sum_{x,y} g(x,y) p_{X,Y}(x,y)$	$\mathbb{E}[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y) dx dy$
Independence must have	$\forall x, y, p_{X,Y}(x,y) = p_X(x) p_Y(y)$ $\Omega_{X,Y} = \Omega_X \times \Omega_Y$	$\forall x, y, f_{X,Y}(x,y) = f_X(x) f_Y(y)$ $\Omega_{X,Y} = \Omega_X \times \Omega_Y$

Review: Law of Total Probability

Law of Total Probability (r.v. version): If X is a discrete random variable, then

$$\mathbb{P}(A) = \sum_{x \in \Omega_X} \mathbb{P}(A|X = x) p_X(x) \qquad \text{discrete } X$$

Continuous Law of Total Probability:

$$\mathbb{P}(A) = \int_{x \in \Omega_X} \mathbb{P}(A|X = x) f_X(x) dx$$

Review: Conditional Expectation for Events

Conditional expectation: The expected value of random variable X given that event A has occurred, written $\mathbb{E}[X|A]$, is defined as

$$\mathbb{E}[X|A] = \sum_{x \in \Omega_X} x \cdot \mathbb{P}(X = x|A).$$

Review: Conditional Distributions and Expectation

	Discrete	Continuous
Conditional PMF/PDF	$p_{X\|Y}(x\|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$	$f_{X\|Y}(x\|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$
Conditional Expectation	$\mathbb{E}[X\|Y=y] = \sum_x x p_{X\|Y}(x\|y)$	$\mathbb{E}[X\|Y=y] = \int_{-\infty}^{\infty} x f_{X\|Y}(x\|y) dx$

Review: Law of Total Expectation

- **Law of Total Expectation (Event Version):** Let events $A_1 \dots, A_n$ partition the sample space. Then

$$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X|A_i]\,\mathbb{P}(A_i).$$

- **Law of Total Expectation (RV Version):** Suppose X and Y are random variables. Then

$$\mathbb{E}[X] = \sum_{y \in \Omega_Y} \mathbb{E}[X|Y=y] \cdot \mathbb{P}(Y=y). \qquad \text{discrete } Y$$

$$\mathbb{E}[X] = \int_{y \in \Omega_Y} \mathbb{E}[X|Y=y]\, f_Y(y)dy \qquad \text{continuous } Y$$

Review: Central Limit Problem Template

a) Setup the problem - write event you are interested in, in terms of sum of random variables. (what do we want to solve for/what is the probability we want to be true?)

- Write the random variable we're interested in as a sum of i.i.d., random variables.
- Apply CLT to $X = \sum_{i=1}^{n} X_i$ approximating it by $Y \sim N(\mu, \sigma^2)$.
- Write the probability we're interested in.

b) *If the RVs are discrete*, apply continuity correction.

c) Normalize RV to have mean 0 and standard deviation 1: $Z = \frac{Y-\mu}{\sigma}$

d) Replace RV in probability expression with $Z \sim N(0, 1)$

e) Write in terms of $\Phi(z) = P(Z \leq z)$

f) Look up in the Phi table (or do a reverse Phi table lookup if we're looking for a value of z that gives us a certain probability)

Problems

Task 1: Content Review Questions

a) Select one: Suppose we have n independent and identically distributed random variables $X_1, X_2, \ldots, X_n$, each with mean μ and variance σ^2. Let $X = \sum_{i=1}^{n} X_i$. Then as n grows large, the Central Limit Theorem tells us that X behaves similarly to which normal distribution?

- ○ $X \sim \mathcal{N}(n\mu, n\sigma^2)$
- ○ $X \sim \mathcal{N}(\mu, n\sigma^2)$
- ○ $X \sim \mathcal{N}(n\mu, \sigma^2)$
- ○ $X \sim \mathcal{N}(n\mu, n^2\sigma^2)$

b) **Marginal PDF.** Let X and Y be continuous random variables with joint PDF $f_{X,Y}(x,y)$. Which of the following correctly expresses the marginal PDF $f_X(x)$?

- ○ $\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx$
- ○ $\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy$
- ○ $\frac{f_{X,Y}(x,y)}{f_Y(y)}$
- ○ $\int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(t,s)\,ds\,dt$

c) **Independence and Support.** True or False: If the joint support $\Omega_{X,Y}$ of the random variables (X, Y) is a circle defined by $x^2 + y^2 \leq 1$, and $\Omega_X = \Omega_Y = [0, 1]$ then X and Y are independent.

- ○ True
- ○ False

d) **Continuous Law of Total Probability.** Let A be an event and X be a continuous random variable with PDF $f_X(x)$. Which of the following is the correct expression for the Continuous Law of Total Probability?

- ○ $\mathbb{P}(A) = \int_{-\infty}^{\infty} \mathbb{P}(A \mid X = x)\, dx$
- ○ $\mathbb{P}(A) = \int_{-\infty}^{\infty} \mathbb{P}(A \cap X = x) f_X(x)\, dx$
- ○ $\mathbb{P}(A) = \int_{-\infty}^{\infty} \mathbb{P}(X = x \mid A)\mathbb{P}(A)\, dx$
- ○ $\mathbb{P}(A) = \int_{-\infty}^{\infty} \mathbb{P}(A \mid X = x) f_X(x)\, dx$

e) Select all the true statements. Suppose that $X_1, X_2, \ldots, X_{30}$ are i.i.d. random variables, and let $X = \sum_{i=1}^{30} X_i$. Let Y denote the normal random variable used to approximate X via the CLT. Then:

- ○ If $X_i \sim \text{Ber}(p)$, then we use continuity correction to approximate

$$\mathbb{P}(12 \leq X \leq 20) \approx \mathbb{P}(11.5 \leq Y \leq 20.5)$$

- ○ If $X_i \sim \text{Exp}(\lambda)$, then we use continuity correction to approximate

$$\mathbb{P}(12 \leq X \leq 20) \approx \mathbb{P}(11.5 \leq Y \leq 20.5)$$

- ○ If $\mathbb{P}(X_i = 0) = \mathbb{P}(X_i = 2) = \frac{1}{2}$, then we use continuity correction to approximate

$$\mathbb{P}(12 \leq X \leq 20) \approx \mathbb{P}(11 \leq Y \leq 21)$$

- ○ If $X_i \sim \text{Poi}(\lambda)$, then we use continuity correction to approximate

$$\mathbb{P}(12 < X < 20) \approx \mathbb{P}(12.5 \leq Y \leq 19.5)$$

Task 2: Joint PMFs

Suppose X and Y have the following joint PMF:

X/Y	1	2	3
0	0	0.2	0.1
1	0.3	0	0.4

a) Identify the range of X (Ω_X), the range of Y (Ω_Y), and their joint range ($\Omega_{X,Y}$).

b) Find the marginal PMF for X, $p_X(x)$ for $x \in \Omega_X$.

c) Find the marginal PMF for Y, $p_Y(y)$ for $y \in \Omega_Y$.

d) Are X and Y independent? Why or why not?

e) Find $\mathbb{E}[X^3 Y]$.

Task 2 Solution: a)

a) Identify the range of X (Ω_X), the range of Y (Ω_Y), and their joint range $(\Omega_{X,Y})$.

Task 2 Solution: a)

a) Identify the range of X (Ω_X), the range of Y (Ω_Y), and their joint range $(\Omega_{X,Y})$.

$\Omega_X = \{0, 1\}$, $\Omega_Y = \{1, 2, 3\}$, and $\Omega_{X,Y} = \{(0, 2), (0, 3), (1, 1), (1, 3)\}$

Task 2 Solution: b)

b) Find the marginal PMF for X, $p_X(x)$ for $x \in \Omega_X$.

Task 2 Solution: b)

b) Find the marginal PMF for X, $p_X(x)$ for $x \in \Omega_X$.

Note that $\Omega_X = \{0, 1\}$.

$$p_X(0) = \sum_y p_{X,Y}(0, y) = 0 + 0.2 + 0.1 = 0.3$$

$$p_X(1) = 1 - p_X(0) = 0.7$$

Task 2 Solution: c)

c) Find the marginal PMF for Y, $p_Y(y)$ for $y \in \Omega_Y$.

Task 2 Solution: c)

c) Find the marginal PMF for Y, $p_Y(y)$ for $y \in \Omega_Y$.

Note that $\Omega_Y = \{1, 2, 3\}$.

$$p_Y(1) = \sum_x p_{X,Y}(x, 1) = 0 + 0.3 = 0.3$$

$$p_Y(2) = \sum_x p_{X,Y}(x, 2) = 0.2 + 0 = 0.2$$

$$p_Y(3) = \sum_x p_{X,Y}(x, 3) = 0.1 + 0.4 = 0.5$$

Task 2 Solution: d)

d) Are X and Y independent? Why or why not?

Task 2 Solution: d)

d) Are X and Y independent? Why or why not?

X and Y are not independent. Recall that a *necessary* condition for X and Y to be independent is that $\Omega_{X,Y} = \Omega_X \times \Omega_Y$. The joint range $\Omega_{X,Y}$ does not satisfy this criteria, so it cannot be independent.

Task 2 Solution: e)

e) Find $\mathbb{E}[X^3Y]$.

Task 2 Solution: e)

e) Find $\mathbb{E}[X^3Y]$.

Note that $X^3 = X$ since X takes values in $\{0,1\}$.

$$\mathbb{E}[X^3Y] = \mathbb{E}[XY] = \sum_{(x,y)\in\Omega_{X,Y}} xyp_{X,Y}(x,y) = 1\cdot 1\cdot 0.3 + 1\cdot 3\cdot 0.4 = 1.5$$

Task 6: Continuous Joint Density

The joint density of X and Y is given by

$$f_{X,Y}(x,y) = \begin{cases} xe^{-(x+y)} & x > 0, y > 0 \\ 0 & \text{otherwise.} \end{cases}$$

and the joint density of W and V is given by

$$f_{W,V}(w,v) = \begin{cases} 2 & 0 < w < v, 0 < v < 1 \\ 0 & \text{otherwise.} \end{cases}$$

Are X and Y independent? Are W and V independent?

Task 6 Solution

For two random variables X, Y to be independent, we must have $f_{X,Y}(x, y) = f_X(x) f_Y(y)$ for all $x \in \Omega_X$, $y \in \Omega_Y$. Let's start with X and Y by finding their marginal PDFs. By definition, and using the fact that the joint PDF is 0 outside of $y > 0$, we get:

$$f_X(x) = \int_0^\infty x e^{-(x+y)} dy = e^{-x} x$$

Task 6 Solution

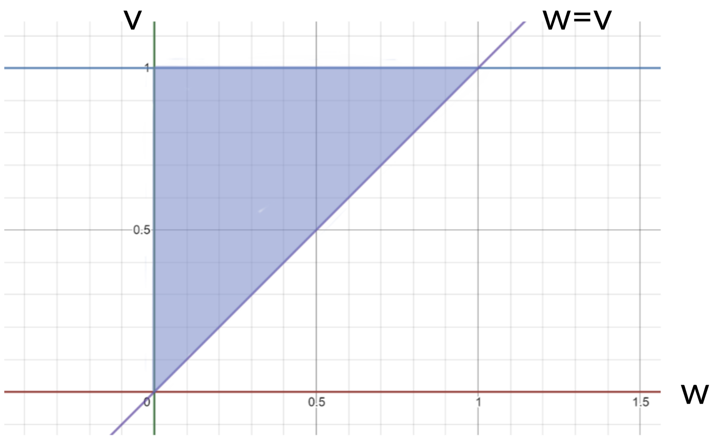
We do the same to get the PDF of Y, again over the range $x > 0$:

$$f_Y(y) = \int_0^\infty xe^{-(x+y)}dx = e^{-y}$$

Since $e^{-x}x \cdot e^{-y} = xe^{-x-y} = xe^{-(x+y)}$ for all $x, y > 0$, X and Y are independent.

Task 6 Solution

We can see that W and V are not independent simply by observing that $\Omega_W = (0,1)$ and $\Omega_V = (0,1)$, but $\Omega_{W,V}$ is not equal to their Cartesian product. Specifically, looking at their range of $f_{W,V}(w,v)$. Graphing it with w as the "x-axis" and v as the "y-axis", we see that :



Task 7: 3 points on a line

Three values X_1, X_2, X_3 are selected uniformly at random, each between 0 and 1 (continuous independent uniform distributions). What is the probability that X_2 is greater than X_1 but less than X_3?

Task 7 Solution

Method 1: Condition on X_2.

$$\begin{aligned}
\mathbb{P}(X_1 < X_2 < X_3) &= \int_{-\infty}^{\infty} \mathbb{P}(X_1 < X_2 < X_3 \mid X_2 = x)\, f_{X_2}(x)\, dx && \text{Continuous LoTP} \\
&= \int_{-\infty}^{\infty} \mathbb{P}(X_1 < x, X_3 > x | X_2 = x)\, f_{X_2}(x)\, dx \\
&= \int_{-\infty}^{\infty} \mathbb{P}(X_1 < x, X_3 > x)\, f_{X_2}(x)\, dx && \text{Indep of } X_1, X_3 \text{ from } X_2 \\
&= \int_{-\infty}^{\infty} \mathbb{P}(X_1 < x)\, \mathbb{P}(x < X_3)\, f_{X_2}(x)\, dx && \text{Indep of } X_1, X_3 \\
&= \int_{-\infty}^{\infty} F_{X_1}(x)\, (1 - F_{X_3}(x))\, f_{X_2}(x)\, dx \\
&= \int_0^1 x\, (1 - x)\, 1\, dx \\
&= \frac{x^2}{2} - \frac{x^3}{3}\Big|_0^1 = \frac{1}{6}
\end{aligned}$$

Task 7 Solution

Method 2: Integrate over each variable. First, find the joint PDF of the three variables using independence

$$f_{X_1,X_2,X_3}(x_1,x_2,x_3) = f_{X_1}(x_1)f_{X_2}(x_2)f_{X_3}(x_3) = \begin{cases} 1 & \text{if} \quad 0 \le x_1, x_2, x_3 \le 1 \\ 0 & \text{otherwise} \end{cases}$$

So, the probability of the event $X_1 < X_2 < X_3$ is exactly the volume of the region where $x_1 < x_2 < x_2$. In order to integrate over these region, start with x_2 (this has a range of 0 to 1), then x_1 (since $x_1 < x_2$, its range is 0 to x_2), and finally x_3 (since $x_2 < x_3 < 1$, its range is x_2 to 1).

$$\begin{aligned}
\mathbb{P}(X_1 < X_2 < X_3) &= \int_0^1 \int_0^{x_2} \int_{x_2}^1 f_{X_1,X_2,X_3}(X_1,X_2,X_3)\, dx_3 dx_1 dx_2 \\
&= \int_0^1 \int_0^{x_2} \int_{x_2}^1 1\, dx_3 dx_1 dx_2 \\
&= \int_0^1 \int_0^{x_2} x_3 \Big|_{x_2}^1 dx_1 dx_2 = \int_0^1 \int_0^{x_2} 1 - x_2\, dx_1 dx_2 \\
&= \int_0^1 (1-x_2)x_1 \Big|_0^{x_2} dx_2 = \int_0^1 (1-x_2)x_2\, dx_2 \\
&= \frac{x^2}{2} - \frac{x^3}{2} \Big|_0^1 = \frac{1}{6}
\end{aligned}$$

Task 10: Bad Computer (CLT) - part a

Each day, the probability your computer crashes is 10%, independent of every other day. Suppose we want to evaluate the computer's performance over the next 100 days.

a) Let X be the number of crash-free days in the next 100 days. What distribution does X have? Identify $\mathbb{E}[X]$ and $Var(X)$ as well. Write an exact expression for $\mathbb{P}(X \geq 87)$.

Task 10 part a Solution

Since X counts the number of crash-free days (successes) in 100 days (trials), where each trial is a success with probability 0.9, we can see that X is binomial with $n = 100$ and $p = 0.9$, or $X \sim \mathsf{Binomial}(100, 0.9)$. Hence, $\mathbb{E}[X] = np = 90$ and $Var(X) = np(1-p) = 9$. Finally,

$$\mathbb{P}(X \geq 87) = \sum_{k=87}^{100} \binom{100}{k} (0.9)^k (1-0.9)^{100-k}$$

Task 10: Bad Computer (CLT) - part b

Each day, the probability your computer crashes is 10%, independent of every other day. Suppose we want to evaluate the computer's performance over the next 100 days.

a) Let X be the number of crash-free days in the next 100 days. What distribution does X have? Identify $\mathbb{E}[X]$ and $Var(X)$ as well. Write an exact expression for $\mathbb{P}(X \geq 87)$.

b) Approximate the probability of at least 87 crash-free days out of the next 100 days using the Central Limit Theorem.

Task 10 part b Solution

From the previous part, we know that $X = 90$ and $\text{Var}(X) = 9$.

$$\begin{aligned}\mathbb{P}(X \geq 87) &= \mathbb{P}(86.5 < X < 100.5) = \mathbb{P}(\frac{86.5 - 90}{3} < \frac{X - 90}{3} < \frac{100.5 - 90}{3}) \\ &\approx \mathbb{P}(-1.17 < \frac{X - 90}{3} < 3.5) \approx \Phi(3.5) + \Phi(1.17) - 1 \approx 0.8788\end{aligned}$$

Notice that, if you had used $86.5 < X$ in place of $86.5 < X < 100.5$, your answer would have been nearly the same, because $\Phi(3.5)$ is so close to 1.

Task 8: Min and max of i.i.d. random variables (optional)

Let $X_1, X_2, \dots, X_n$ be i.i.d. random variables each with CDF $F_X(x)$ and pdf $f_X(x)$. Let $Y = \min(X_1, \dots, X_n)$ and let $Z = \max(X_1, \dots, X_n)$. Show how to write the CDF and pdf of Y and Z in terms of the functions $F_X(\cdot)$ and $f_X(\cdot)$.