50.034 - Introduction to Probability and Statistics

Week 4 – Cohort Class

January–May Term, 2019

SINGAPORE UNIVERSITY OF
TECHNOLOGY AND DESIGN

CC Week 4 Assigned Readings: (Revise: 3.4, 3.5, 3.6, 3.7)

Outline of Cohort Class

Exercises on the following topics:

- Joint distributions
- Marginal pmf/pdf/cdf
- Independence of R.V.'s
- Conditional distributions

Joint distributions

The joint distribution of any R.V.'s X and Y is the collection of all probabilities of the form $\Pr\left((X, Y) \in C\right)$, for all sets $C \subseteq \mathbb{R}^2$.

- C here contains pairs of real numbers.
- **Interpretation:** For any set $C \subseteq \mathbb{R}^2$, this distribution gives the probability $\Pr((X, Y) \in C)$ of how likely pairs of X-values and Y-values take on pairs of values in C.

There are other ways to represent the same information given by the joint distribution of two R.V.'s:

- joint pmf [only for **discrete** R.V.]
- joint pdf [only for **continuous** R.V.]
- joint cdf [for **any** R.V.]

Important Note on Notation:

- The comma in $\Pr(X = x, Y = y)$ represents "and".
- $\Pr(X = x, Y = y)$, $\Pr(X = x \text{ and } Y = y)$, $\Pr((X, Y) = (x, y))$ all mean exactly the same probability that $X = x$ **and** $Y = y$.

Joint pmf and joint pdf

Definition: If X and Y are **discrete** R.V.'s, then the joint pmf of X and Y is the function $p(x, y) = \Pr(X = x, Y = y)$.

- **Note:** $p(x, y) \geq 0$ and $\sum_x \sum_y p(x, y) = 1$.

Definition: If X and Y are **continuous** R.V.'s, then a joint pdf of X and Y is a function $f(x, y)$ satisfying the following:

- $f(x, y)$ is a non-negative function, i.e. $f(x, y) \geq 0$ for all x, y
- For any set $A \subseteq \mathbb{R}^2$, the probability of event $\{(X, Y) \in A\}$ is

$$\Pr\left((X, Y) \in A\right) = \iint_A f(x, y)\, dx\, dy.$$

Recall: $\iint_A f(x, y)\, dx\, dy$ is the volume of the region under the graph of $f(x, y)$ over A (on the xy-plane).

Note: Every joint pdf $f(x, y)$ must satisfy:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dx\, dy = 1.$$

Marginal pmf and joint pdf

If X and Y are **discrete** R.V.'s with joint pmf $p(x, y)$, then:

- The marginal pmf of X is $p_X(x) = \sum_{y \in D_Y} p(x, y)$;
- The marginal pmf of Y is $p_Y(x) = \sum_{x \in D_X} p(x, y)$;

where D_X and D_Y are the sets of possible values for X and Y.

If X and Y are **continuous** R.V.'s with joint pdf $f(x, y)$, then:

- The marginal pdf of X is $f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy$.
- The marginal pdf of X is $f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx$.

Important Remarks:

- The word "marginal" indicates that the pmf/pdf is obtained from a joint distribution.
- A marginal pmf/pdf is a legitimate pmf/pdf.
  - $p_X(x) \geq 0$, $p_Y(y) \geq 0$, $f_X(x) \geq 0$, $f_Y(y) \geq 0$ for all x, y.
  - $\sum_x p_X(x) = \sum_y p_Y(y) = 1$, $\int_{-\infty}^{\infty} f_X(x)\, dx = \int_{-\infty}^{\infty} f_Y(y)\, dy = 1$.

Joint cdf and marginal cdf

Definition: The joint cdf of any given R.V.'s X and Y is the function $F(x, y) = \Pr(X \leq x, Y \leq y)$, for $-\infty < x, y < \infty$.

Definition: Let X and Y be arbitrary R.V.'s with joint cdf $F(X, y)$.

- The marginal cdf of X is $F_X(x) = \lim_{y\to\infty} F(x, y)$.
- The marginal cdf of Y is $F_Y(y) = \lim_{x\to\infty} F(x, y)$.

Remarks:

- If X and Y are **discrete** R.V.'s with joint pmf $p(x, y)$, then:

$$F(a, b) = \Pr(X \leq a, Y \leq b) = \sum_{x\leq a} \sum_{y\leq b} p(x, y).$$

- If X and Y are **continuous** R.V.'s with joint pdf $f(x, y)$, then:

$$F(a, b) = \Pr(X \leq a, Y \leq b) = \int_{-\infty}^{b} \int_{-\infty}^{a} f(x, y)\, dx\, dy.$$

Exercise 1 (45 mins)

Let X and Y be continuous R.V.'s such that (X, Y) must belong to the rectangle in the xy-plane containing all points (x, y) that satisfy $0 \leq x \leq 3$ and $0 \leq y \leq 4$. Suppose that the joint cdf of X and Y at every point (x, y) in this rectangle is specified as follows:

$$F(x, y) = \frac{1}{156} xy(x^2 + y).$$

Determine the following:

1. The joint cdf of X and Y (i.e. at every point in $\mathbb{R}^2$, not just in the rectangle).
2. $\Pr(1 \leq X \leq 2, 1 \leq Y \leq 2)$.
3. $\Pr(2 \leq X \leq 4, 2 \leq Y \leq 4)$.
4. The cdf of Y.
5. The joint pdf of X and Y.
6. The pdf of X.
7. $\Pr(Y \leq X)$.

(You do not have to do these parts in order.)

Exercise 1 - Solution

1. [Determine the joint cdf of X and Y.]

First, we recall some basic properties of joint cdf's:

- For fixed y_0, $F(x, y_0)$ (as a function of x) is **non-decreasing**.
- For fixed x_0, $F(x_0, y)$ (as a function of y) is **non-decreasing**.
- The limits of $F(x, y)$ at $(\pm\infty, \pm\infty)$ are:

$$\lim_{\substack{x\to-\infty \\ y\to-\infty}} F(x, y) = 0 \quad \text{and} \quad \lim_{\substack{x\to\infty \\ y\to\infty}} F(x, y) = 1.$$

Case 1: If $(x_0, y_0) \in \mathbb{R}^2$ satisfies either $x_0 < 0$ **or** $y_0 < 0$, then the set $\{(x, y) \in \mathbb{R}^2 | x \leq x_0, y \leq y_0\}$ does not contain any point in the given rectangle, hence $F(x_0, y_0) = 0$ in this case.

Case 2: If $(x_0, y_0) \in \mathbb{R}^2$ satisfies both $x_0 > 3$ **and** $y_0 > 4$, then the set $\{(x, y) \in \mathbb{R}^2 | x \leq x_0, y \leq y_0\}$ contains all points in the given rectangle, hence $F(x_0, y_0) = 1$ in this case.

Exercise 1 - Solution

Case 3: If $(x_0, y_0) \in \mathbb{R}^2$ satisfies both $x_0 > 3$ **and** $0 \le y_0 \le 4$, then the set $\{(x, y) \in \mathbb{R}^2 | x \le x_0, y \le y_0\}$ when intersected with the rectangle gives the subset:

$$\{(x, y) \in \mathbb{R}^2 | 0 \le x \le 3, 0 \le y \le y_0\},$$

hence $F(x_0, y_0) = F(3, y_0) = \frac{1}{52} y_0(9 + y_0)$ in this case.

Case 4: If $(x_0, y_0) \in \mathbb{R}^2$ satisfies both $0 \le x_0 \le 3$ **and** $y_0 > 4$, then the set $\{(x, y) \in \mathbb{R}^2 | x \le x_0, y \le y_0\}$ when intersected with the rectangle gives the subset:

$$\{(x, y) \in \mathbb{R}^2 | 0 \le x \le x_0, 0 \le y \le 4\},$$

hence $F(x_0, y_0) = F(x_0, 4) = \frac{1}{39} x_0(x_0^2 + 4)$ in this case.

Note: We have covered all possible cases.

- Every point $(x_0, y_0) \in \mathbb{R}^2$ either has been considered in one of the four cases, or is contained in the given rectangle.

Exercise 1 - Solution

By combining these four cases, together with the given value of $F(x, y)$ on the rectangle, we get:

$$F(x,y) = \begin{cases} \frac{1}{156}xy(x^2+y), & \text{if } 0 \leq x \leq 3 \text{ and } 0 \leq y \leq 4; \\ \frac{1}{52}y(9+y), & \text{if } x > 3 \text{ and } 0 \leq y \leq 4; \\ \frac{1}{39}x(x^2+4), & \text{if } 0 \leq x \leq 3 \text{ and } y > 4; \\ 1, & \text{if } x > 3 \text{ and } y > 4; \\ 0, & \text{otherwise.} \end{cases}$$

Exercise 1 - Solution

2. [Determine $\Pr(1 \leq X \leq 2, 1 \leq Y \leq 2)$.]

Recall: $F(x, y) = \frac{1}{156}xy(x^2 + y)$ if $0 \leq x \leq 3$ and $0 \leq y \leq 4$.

Since X and Y are continuous R.V.'s, we have

$$
\begin{aligned}
& \Pr(1 \leq X \leq 2, 1 \leq Y \leq 2) \\
= & \Pr(1 < X \leq 2, 1 < Y \leq 2) \\
= & \Pr(X \leq 2, Y \leq 2) - \Pr(X \leq 1, Y \leq 2) \\
& - \Pr(X \leq 2, Y \leq 1) + \Pr(X \leq 1, Y \leq 1) \\
= & F(2,2) - F(1,2) - F(2,1) + F(1,1) \\
= & \tfrac{1}{156}\left[24 - 6 - 10 + 2\right] \\
= & \tfrac{5}{78}.
\end{aligned}
$$

Exercise 1 - Solution

3. [Determine $\Pr(2 \leq X \leq 4, 2 \leq Y \leq 4)$.]

Recall: $F(x, y) = \frac{1}{156}xy(x^2 + y)$ if $0 \leq x \leq 3$ and $0 \leq y \leq 4$.
Also, from part 1., $F(x, y) = \frac{1}{52}y(9 + y)$ if $x > 3$ and $0 \leq y \leq 4$.

Since X and Y are continuous R.V.'s, we have

$$\begin{aligned}
&\Pr(2 \leq X \leq 4, 2 \leq Y \leq 4) \\
&= \Pr(2 < X \leq 4, 2 < Y \leq 4) \\
&= \Pr(X \leq 4, Y \leq 4) - \Pr(X \leq 2, Y \leq 4) \\
&\quad - \Pr(X \leq 4, Y \leq 2) + \Pr(X \leq 2, Y \leq 2) \\
&= F(4, 4) - F(2, 4) - F(4, 2) + F(2, 2) \\
&= \frac{4(13)}{52} - \frac{8(8)}{156} - \frac{2(11)}{52} + \frac{4(6)}{156} \\
&= \frac{25}{78}.
\end{aligned}$$

Exercise 1 - Solution

4. [Determine the cdf of Y.]

From part 1., we already have:

$$F(x,y) = \begin{cases} \frac{1}{156}xy(x^2+y), & \text{if } 0 \le x \le 3 \text{ and } 0 \le y \le 4; \\ \frac{1}{52}y(9+y), & \text{if } x > 3 \text{ and } 0 \le y \le 4; \\ \frac{1}{39}x(x^2+4), & \text{if } 0 \le x \le 3 \text{ and } y > 4; \\ 1, & \text{if } x > 3 \text{ and } y > 4; \\ 0, & \text{otherwise.} \end{cases}$$

We know that the cdf of Y is $F_Y(y) = \lim_{x\to\infty} F(x,y)$. Thus:

$$F_Y(y) = \begin{cases} \frac{1}{52}y(9+y), & \text{if } 0 \le y \le 4; \\ 1, & \text{if } y > 4; \\ 0, & \text{if } y < 0. \end{cases}$$

Exercise 1 - Solution

5. [Determine the joint pdf of X and Y.]

Let $f(x, y)$ denote the joint pdf of X and Y. Since (X, Y) must belong to the given rectangle, we know that $f(x, y) = 0$ whenever (x, y) is not in the rectangle.
Given any $(x_0, y_0) \in \mathbb{R}^2$, we know

$$f(x_0, y_0) = \frac{\partial^2 F(x,y)}{\partial x \partial y}\Big|_{(x,y)=(x_0,y_0)},$$

provided this 2nd-order partial derivative exists at $(x, y) = (x_0, y_0)$.

We check that

$$\frac{\partial^2}{\partial x \partial y}\left(\frac{1}{156} xy(x^2 + y)\right) = \frac{1}{156} \cdot \frac{\partial}{\partial y}(3x^2 y + y^2) = \frac{1}{156}(3x^2 + 2y).$$

Therefore,

$$f(x, y) = \begin{cases} \frac{1}{156}(3x^2 + 2y), & \text{if } 0 \le x \le 3 \text{ and } 0 \le y \le 4; \\ 0, & \text{otherwise.} \end{cases}$$

Exercise 1 - Solution

6. [Determine the pdf of X.]

From part 5., the joint pdf of X and Y is

$$f(x, y) = \begin{cases} \frac{1}{156}(3x^2 + 2y), & \text{if } 0 \leq x \leq 3 \text{ and } 0 \leq y \leq 4; \\ 0, & \text{otherwise.} \end{cases}$$

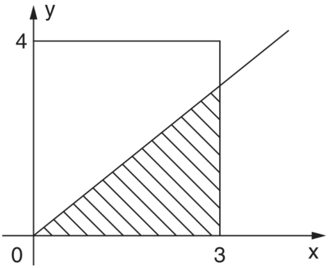
Thus, the pdf of X is the marginal pdf of X, which equals

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f(x, y)\, dy \\ &= \int_0^4 \tfrac{1}{156}(3x^2 + 2y)\, dy \\ &= \left[\tfrac{1}{156}(3x^2 y + y^2) \right]_{y=0}^{y=4} \\ &= \tfrac{1}{39}(3x^2 + 4). \end{aligned}$$

Exercise 1 - Solution

7. [Determine $\Pr(Y \leq X)$.]

The shaded region on the right shows the region containing all points (x, y) on the xy-plane satisfying $y \leq x$.



From part 5., the joint pdf of X and Y in this shaded region is

$$f(x,y) = \frac{1}{156}(3x^2 + 2y).$$

Thus,

$$\Pr(Y \leq X) = \int_0^3 \int_0^x \frac{1}{156}(3x^2 + 2y)\,dy\,dx = \int_0^3 \left[\frac{3x^2y + y^2}{156}\right]_{y=0}^{y=x} dx$$
$$= \int_0^3 \frac{3x^3 + x^2}{156}\,dx = \frac{1}{156} \cdot \left[\frac{3}{4}x^4 + \frac{1}{3}x^3\right]_{y=0}^{y=3} = \frac{93}{208}.$$

Conditional distribution/pmf/pdf

Let $C' \subseteq \mathbb{R}$, and let X and Y be **arbitrary** R.V.'s. The conditional distribution of X given $Y \in C'$ is the collection of all conditional probabilities of the form $\Pr(X \in C | Y \in C')$ for all sets $C \subseteq \mathbb{R}$.

If X and Y are **discrete** R.V.'s with joint pmf $p(x, y)$, and if $y \in \mathbb{R}$ such that $p_Y(y) > 0$, then the conditional pmf of X given $Y = y$, is the function $p_{X|Y}(x|y) = \frac{p(x,y)}{p_Y(y)}$.

If X and Y are **continuous** R.V.'s with joint pdf $f(x, y)$, and if $y \in \mathbb{R}$ such that $f_Y(y) > 0$, then the conditional pdf of X given $Y = y$, is the function $f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}$.

Exercise 2 (25 mins)

Let X and Y be continuous R.V.'s with joint pdf

$$f(x,y) = \begin{cases} c \cdot \sin x, & \text{if } 0 \leq x \leq \frac{\pi}{2} \text{ and } 0 \leq y \leq 3; \\ 0, & \text{otherwise;} \end{cases}$$

where c is an unspecified constant.

1. Determine the value of c.
2. Determine the marginal pdf of X.
3. Determine the marginal pdf of Y.
4. Are X and Y independent or dependent?
5. Determine the value of $\Pr(1 < Y < 2 | X = \frac{\pi}{6})$.

Exercise 2 - Solution

1. [Determine the value of c.]

Since $f(x, y)$ is a pdf, it must satisfy $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dx\, dy = 1$.

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dx\, dy = \int_0^3 \int_0^{\frac{\pi}{2}} c \cdot \sin x\, dx\, dy$$

$$= \int_0^3 \Big[- c \cdot \cos x \Big]_{x=0}^{x=\frac{\pi}{2}} dy = \int_0^3 c\, dy = 3c.$$

Thus, $c = \frac{1}{3}$.

Exercise 2 - Solution

2. [Determine the marginal pdf of X.]

Let $f_X(x)$ be the marginal pdf of X.

Using $c = \frac{1}{3}$ from part 1., we get that

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy = \int_0^3 \tfrac{1}{3} \sin x\, dy$$
$$= \left[\tfrac{1}{3} y \sin x\right]_{y=0}^{y=3} = \sin x$$

if $0 \leq x \leq \frac{\pi}{2}$, and $f_X(x) = 0$ otherwise.

Therefore:

$$f_X(x) = \begin{cases} \sin x, & \text{if } 0 \leq x \leq \frac{\pi}{2}; \\ 0, & \text{otherwise.} \end{cases}$$

Exercise 2 - Solution

3. [Determine the marginal pdf of Y.]

Let $f_Y(y)$ be the marginal pdf of Y.

Using $c = \frac{1}{3}$ from part 1., we get that

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \int_0^{\frac{\pi}{2}} \tfrac{1}{3}\sin x\,dx$$
$$= \Big[-\tfrac{1}{3}\cos x \Big]_{x=0}^{x=\frac{\pi}{2}} = \tfrac{1}{3}$$

if $0 \le y \le 3$, and $f_Y(y) = 0$ otherwise.

Therefore:

$$f_Y(y) = \begin{cases} \frac{1}{3}, & \text{if } 0 \le y \le 3; \\ 0, & \text{otherwise.} \end{cases}$$

Exercise 2 - Solution

4. [Are X and Y independent or dependent?]

So far, we have:

$$f(x,y) = \begin{cases} \frac{1}{3}\sin x, & \text{if } 0 \le x \le \frac{\pi}{2} \text{ and } 0 \le y \le 3; \\ 0, & \text{otherwise;} \end{cases}$$

$$f_X(x) = \begin{cases} \sin x, & \text{if } 0 \le x \le \frac{\pi}{2}; \\ 0, & \text{otherwise.} \end{cases}$$

$$f_Y(y) = \begin{cases} \frac{1}{3}, & \text{if } 0 \le y \le 3; \\ 0, & \text{otherwise.} \end{cases}$$

Since $f(x,y) = f_X(x)f_Y(y)$, i.e. the joint pdf is the product of the marginal pdf's, we conclude that X and Y are **independent**.

Exercise 2 - Solution

5. [Determine the value of $\Pr(1 < Y < 2 | X = \frac{\pi}{6})$.]

From part 4., we know that X and Y are independent.
Thus, $\Pr(1 < Y < 2 | X = \frac{\pi}{6}) = \Pr(1 < Y < 2)$.

- Knowing the value of Y does not give additional information about X. The probability that $1 < Y < 2$ remains the same, after we are given that $X = \frac{\pi}{6}$.

From part 3., the marginal pdf of Y is

$$f_Y(y) = \begin{cases} \frac{1}{3}, & \text{if } 0 \leq y \leq 3; \\ 0, & \text{otherwise.} \end{cases}$$

Therefore,

$$\Pr(1 < Y < 2 | X = \tfrac{\pi}{6}) = \Pr(1 < Y < 2) = \int_1^2 \tfrac{1}{3}\, dy = \tfrac{1}{3}.$$

Exercise 2 - Solution

Without using the fact that X and Y are independent, we could also calculate the conditional probability $\Pr(1 < Y < 2 | X = \frac{\pi}{6})$ directly:

Alternative Method for part 5.

The conditional probability of Y given $X = x$ is

$$f_{Y|X}(y|x) = \frac{f(x, y)}{f_X(x)}.$$

From part 2., $f_X(\frac{\pi}{6}) = \sin\frac{\pi}{6} = \frac{1}{2}$. Thus:

$$f_{Y|X}(y|\tfrac{\pi}{6}) = \frac{f(x, \frac{\pi}{6})}{f_X(\frac{\pi}{6})} = \frac{\frac{1}{3}\sin\frac{\pi}{6}}{\frac{1}{2}} = \frac{1}{3}.$$

Summary

Exercises on the following topics:

- Joint distributions
- Marginal pmf/pdf/cdf
- Independence of R.V.'s
- Conditional distributions

Reminder: Homework Set 4 is due next Cohort Class.