50.034 - Introduction to Probability and Statistics

Week 4 – Lecture 8

January–May Term, 2019

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

Lecture 8 Assigned Readings: 3.6, 3.7

Outline of Lecture

- Conditional distributions
- Conditional pmf/pdf
- Bayes' theorem for R.V.'s
- Random vectors
- Marginal pmf/pdf/cdf for multiple R.V.'s
- Independence of multiple R.V.'s

Introduction to Conditional Distributions

You have recently launched a mobile app with in-app purchases.

You decide to model the following R.V.'s:

- $X =$ amount spent on in-app purchases per user each week
- $Y =$ time (in hours) spent on the app per user each week

Clearly X and Y depend on each other. The distribution of X for a given user would change if we know the value of Y.

In general, we want to know how to adjust the distribution of one R.V. X, given observed values of another R.V. Y, e.g. $Y = 10$.

Such adjusted distributions are called **conditional distributions**.

Conditional Distributions

Recall: If A and B are events, such that $\Pr(B) > 0$, then the conditional probability of A given B is defined to be

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}.$$

Let X and Y be *any* R.V.'s, and let C' be a set of real numbers, such that $\Pr(Y \in C') > 0$. (Note: $\{Y \in C'\}$ is an event.)

The conditional distribution of X given $Y \in C'$ is defined to be the collection of all **conditional probabilities** of the form

$$\Pr(X \in C | Y \in C')$$

for all sets $C \subseteq \mathbb{R}$.

Conditional pmf

Let X and Y be **discrete** R.V.'s with joint pmf $p(x, y)$.
Let $p_Y(y)$ be the marginal pmf of Y.
[**Recall:** $p(x, y) = \Pr(X = x \text{ and } Y = y)$, $p_Y(y) = \Pr(Y = y)$.]

For any $y \in \mathbb{R}$ such that $p_Y(y) > 0$, the conditional probability mass function (conditional pmf) of X given $Y = y$, is a function denoted by $p_{X|Y}(x|y)$, and given by

$$p_{X|Y}(x|y) = \frac{p(x, y)}{p_Y(y)},$$

which is defined on all possible values x for X.

- **Note:** A conditional pmf is a pmf!
- If the context is clear, "$p_{X|Y}(x|y)$" is written as "$p(x|y)$".

Example 1

Doctors decide whether a patient is sick or not by performing a blood test. The test has two outcomes: positive, and negative. Let X and Y be R.V.'s defined on a group of patients, as follows:

- $X(\text{patient}) = \begin{cases} 0, & \text{if patient's blood test is negative;} \\ 1, & \text{if patient's blood test is positive.} \end{cases}$
- $Y(\text{patient}) = \begin{cases} 0, & \text{if patient is healthy;} \\ 1, & \text{if patient is sick.} \end{cases}$

The joint pmf $p(x, y)$ of X and Y is given as follows:

	Y=0	Y=1
X=0	0.72	0.005
X=1	0.18	0.095

(1): What does $p_{X|Y}(0|1)$ mean, and what is its value?
(2): What does $p_{X|Y}(1|0)$ mean, and what is its value?

Example 1

- $X(\text{patient}) = \begin{cases} 0, & \text{if patient's blood test is negative;} \\ 1, & \text{if patient's blood test is positive.} \end{cases}$
- $Y(\text{patient}) = \begin{cases} 0, & \text{if patient is healthy;} \\ 1, & \text{if patient is sick.} \end{cases}$

	Y=0	Y=1
X=0	0.72	0.005
X=1	0.18	0.095

(1): $p_{X|Y}(0|1)$ is the conditional probability that a patient's blood test is negative, given that the patient is sick.
Note: $p_Y(1) = \sum_x p(x,1) = 0.005 + 0.095 = 0.1$. Thus,

$$p_{X|Y}(0|1) = \frac{p(0,1)}{p_Y(1)} = \frac{0.005}{0.1} = 0.05.$$

(2): $p_{X|Y}(1|0)$ is the conditional probability that a patient's blood test is positive, given that the patient is healthy.
Note: $p_Y(0) = \sum_x p(x,0) = 0.72 + 0.18 = 0.9$. Thus,

$$p_{X|Y}(1|0) = \frac{p(1,0)}{p_Y(0)} = \frac{0.18}{0.9} = 0.02.$$

Conditional pdf

Let X and Y be **continuous** R.V.'s with joint pdf $f(x, y)$.
Let $f_Y(y)$ be the marginal pdf of Y.
For any $y \in \mathbb{R}$ such that $f_Y(y) > 0$, the conditional probability density function (conditional pdf) of X given $Y = y$, is a function denoted by $f_{X|Y}(x|y)$, and given by

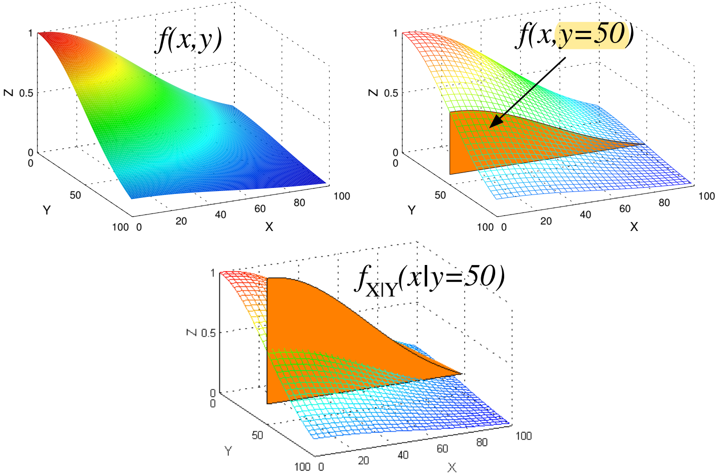
$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)},$$

which is defined on $-\infty < x < \infty$.

- If the context is clear, "$f_{X|Y}(x|y)$" is written as "$f(x|y)$".

Technicality: Since $\Pr(Y = y) = 0$ for a continuous R.V. Y, it seems like we are conditioning over an event with probability 0, which isn't allowed. There is a technical reason (beyond the scope of this course) on why conditional pdf still make sense; see discussion in Chap. 3.6 of textbook. Roughly speaking, $\lim_{\varepsilon \to 0} \Pr(|Y - y| < \varepsilon) > 0$.

Visualization of a conditional pdf



Note: $f_{X|Y}(x|y=50)$ is a normalization of $f(x,y=50)$.

- $f_{X|Y}(x|y=50) = \frac{f(x,y=50)}{f_Y(50)}$, i.e. divide by constant $f_Y(50)$.
- After normalization, $f_{X|Y}(x|y=50)$ is a legitimate pdf.

Properties of conditional pdf

A conditional pdf $f(x|y)$ is a pdf, so for a fixed value y,

- $f_{X|Y}(x|y)$ is a function in terms of x (since y is fixed).
- $f_{X|Y}(x|y) \geq 0$ for all $x \in \mathbb{R}$ (i.e. non-negative function).
- For any set $A \subseteq \mathbb{R}$, the conditional probability of the event $\{X \in A\}$, given that the event $\{Y = y\}$ has occurred, is

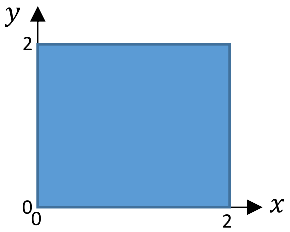
$$\Pr(X \in A|Y = y) = \int_A f_{X|Y}(x|y)\,dx.$$

Note: Since $\Pr(X \in \mathbb{R}|Y = y) = 1$, the conditional pdf $f_{X|Y}(x|y)$ must satisfy

$$\int_{-\infty}^{\infty} f_{X|Y}(x|y)\,dx = 1.$$

Example 2

An iron dart is thrown randomly onto a square magnetic board shown on the right. Let X and Y denote the x-coordinate and y-coordinate respectively of the point the dart lands on.



The joint pdf of X and Y is

$$f(x,y) = \begin{cases} \frac{1}{12}(x+2y), & \text{if } 0 \le x \le 2, 0 \le y \le 2; \\ 0, & \text{otherwise.} \end{cases}$$

Given that the Y-coordinate of the dart is 1, what is the probability that the X-coordinate is ≤ 1?

Example 2

Solution:
To rephrase the question, we want to find the value of the conditional probability $\Pr(X \leq 1|Y = 1)$.

First, we find the marginal pdf of Y:

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \int_0^2 \tfrac{1}{12}(x+2y)\,dx$$
$$= \left[\tfrac{1}{24}x^2 + \tfrac{1}{6}xy\right]_{x=0}^{x=2} = \tfrac{1}{3}y + \tfrac{1}{6}.$$

Hence, the conditional pdf of X given $Y = y$ is:

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{\frac{1}{12}(x+2y)}{\frac{1}{3}y+\frac{1}{6}} = \frac{x+2y}{2(2y+1)}.$$

Example 2

Substituting $y = 1$ into $f_{X|Y}(x|y)$, we get

$$f_{X|Y}(x|1) = \tfrac{x}{6} + \tfrac{1}{3}.$$

Therefore, the conditional probability $\Pr(X \leq 1|Y = 1)$ is:

$$\begin{aligned}
\int_{-\infty}^{1} f_{X|Y}(x|1)\,dx &= \int_{0}^{1} \left(\tfrac{x}{6} + \tfrac{1}{3}\right) dx \\
&= \left[\tfrac{x^2}{12} + \tfrac{1}{3}x\right]_{x=0}^{x=1} \\
&= \frac{5}{12} \\
&\approx 0.4167
\end{aligned}$$

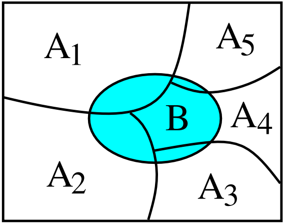
Recall: Law of total probability (Lecture 3)

Let $A_1, \ldots, A_k$ be **mutually exclusive** and **exhaustive** events in some sample space Ω.

- $A_1, \ldots, A_k$ are exhaustive if $A_1 \cup A_2 \cup \cdots \cup A_k = \Omega$.
- $A_1, \ldots, A_k$ are mutually exclusive if $A_i \cap A_j = \varnothing$ for all $i \neq j$.

Then for any event B, the **law of total probability** states that

$$\Pr(B) = \sum_{i=1}^{k} \Pr(B|A_i) P(A_i)$$



Question: Can we extend the law of total probability from conditional probabilities to conditional distributions?

Law of total probability for two R.V.'s

The **law of total probability for discrete R.V.'s** states that for two **discrete** R.V.'s X and Y,

$$p_X(x) = \sum_{y \in D_Y} p_{X|Y}(x|y) p_Y(y),$$

where D_Y is the set of possible values for Y.

- If we know the marginal pmf of Y, and the conditional pmf of X given $Y = y$, then we can find the marginal pmf of X.

The **law of total probability for continuous R.V.'s** states that for two **continuous** R.V.'s X and Y,

$$f_X(x) = \int_{-\infty}^{\infty} f_{X|Y}(x|y) f_Y(y)\, dy.$$

- If we know the marginal pdf of Y, and the conditional pdf of X given $Y = y$, then we can find the marginal pdf of X.

Bayes' theorem for two R.V.'s

The **Bayes' theorem for discrete R.V.'s** states that for two **discrete** R.V.'s X and Y,

$$p_{Y|X}(y|x) = \frac{p_{X|Y}(x|y)p_Y(y)}{p_X(x)}.$$

- Theorem relates the two conditional pmf's ($Y|X$ and $X|Y$).

The **Bayes' theorem for continuous R.V.'s** states that for two **continuous** R.V.'s X and Y,

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

- Theorem relates the two conditional pdf's ($Y|X$ and $X|Y$).

Example 3

A restaurant tracks its sales of set meals and the dining times of its customers. Suppose X and Y are continuous R.V.'s:

- X = Average dining time (in hrs) per customer each day.
- Y = Proportion of customers ordering set meals each day.

Based on historical data, the restaurant finds that the pdf of Y is

$$f_Y(y) = \begin{cases} \frac{6}{5}y + \frac{2}{5}, & \text{if } 0 \leq y \leq 1; \\ 0, & \text{otherwise}; \end{cases}$$

and that the conditional pdf of X given $Y = y$ (for $0 \leq y \leq 1$) is

$$f_{X|Y}(x|y) = \begin{cases} \frac{x+3y}{6y+2}, & \text{if } 0 \leq x \leq 2; \\ 0, & \text{otherwise.} \end{cases}$$

If the average dining time on a given day is 1 hour, what is the probability that at most 50% of the customers order set meals?

Example 3

Solution:
By the law of total probability for continuous R.V.'s,

$$f_X(x) = \int_{-\infty}^{\infty} f_{X|Y}(x|y) f_Y(y)\, dy = \int_0^1 \tfrac{x+3y}{6y+2} \cdot \left(\tfrac{6}{5}y + \tfrac{2}{5}\right) dy$$
$$= \left[\tfrac{1}{5}xy + \tfrac{3}{10}y^2\right]_{y=0}^{y=1} = \tfrac{1}{5}x + \tfrac{3}{10}$$

if $0 \leq x \leq 2$, and $f_X(x) = 0$ otherwise.

Thus, by the Bayes' theorem for continuous R.V.'s,

$$f_{Y|X}(y|x) = \frac{f_{X|Y}(x|y) f_Y(y)}{f_X(x)} = \frac{\frac{x+3y}{6y+2} \cdot \left(\frac{6}{5}y + \frac{2}{5}\right)}{\frac{1}{5}x + \frac{3}{10}} = \frac{2x+6y}{2x+3}$$

for each $0 \leq y \leq 1$, and $f_{Y|X}(y|x) = 0$ otherwise.

Example 3

Substituting $x = 1$ into the conditional pdf $f_{Y|X}(y|x)$, we get

$$f_{Y|X}(y|1) = \tfrac{6}{5}y + \tfrac{2}{5}.$$

Therefore, the probability that $Y \leq 0.5$ given $X = 1$ is

$$\begin{aligned}\Pr(Y \leq 0.5|X = 1) &= \int_{-\infty}^{0.5} f_{Y|X}(y|1)\,dy = \int_{0}^{0.5} (\tfrac{6}{5}y + \tfrac{2}{5})\,dy \\ &= \left[\tfrac{3}{5}y^2 + \tfrac{2}{5}\right]_{y=0}^{y=0.5} = 0.55.\end{aligned}$$

Random vectors

When dealing with many R.V.'s, it is useful to use vector notation.

- A random vector is a vector of (arbitrary) random variables.
  - A discrete random vector is a vector of discrete R.V.'s.
  - A continuous random vector is a vector of continuous R.V.'s.

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

Thus, we can extend this definition to multiple R.V.'s as follows:

- The joint distribution of a random vector $\mathbf{X} = (X_1, \ldots, X_n)$ is the collection of all probabilities of the form $\Pr(\mathbf{X} \in C)$, for all sets $C \subseteq \mathbb{R}^n$.
- If the context is clear, we can simply say "distribution of $\mathbf{X}$".

Joint pmf/pdf/cdf of random vectors

Analogously, we extend "joint pmf", "joint pdf" and "joint cdf" as follows:

- The joint pmf of a **discrete** random vector $\mathbf{X} = (X_1, \ldots, X_n)$ is the function $p(\mathbf{x}) = \Pr(\mathbf{X} = \mathbf{x})$, defined for all vectors $\mathbf{x} \in \mathbb{R}^n$.
- The joint pdf of a **continuous** random vector $\mathbf{X} = (X_1, \ldots, X_n)$ is a function $f(\mathbf{x})$ satisfying $f(\mathbf{x}) \geq 0$ and

$$\Pr(\mathbf{X} \in A) = \int_A f(\mathbf{x})\, d\mathbf{x}$$

  for all $A \subseteq \mathbb{R}^n$ and all $\mathbf{x} \in \mathbb{R}^n$.
- The joint cdf of **any** random vector $\mathbf{X} = (X_1, \ldots, X_n)$ is the function $F(\mathbf{x}) = \Pr(\mathbf{X} \leq \mathbf{x})$, defined for all vectors $\mathbf{x} \in \mathbb{R}^n$.
  - Given any vector $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$, we write "$\mathbf{X} \leq \mathbf{x}$" to mean that $X_1 \leq x_1$ and $X_2 \leq x_2$ and ... and $X_n \leq x_n$.

Marginal pmf/pdf

Next, we look at "marginal pmf" and "marginal pdf":

- Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a **discrete** random vector with joint pmf $p(\mathbf{x})$. Suppose D_i is the set of possible values for each R.V. X_i in $\mathbf{X}$. Then the marginal pmf of each X_i is

$$p_{X_i}(x_i) = \sum_{\substack{1 \leq j \leq n \\ j \neq i}} \sum_{x_j \in D_j} p(x_1, \ldots, x_n),$$

  defined for each possible value $x_i \in D_i$.

- Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a **continuous** random vector with joint pdf $f(\mathbf{x})$. Then the marginal pdf of each X_i is

$$f_{X_i}(x_i) = \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{(n-1) \text{ times}} f(x_1, \ldots, x_n) \underbrace{dx_1 \cdots dx_n}_{dx_i \text{ omitted}}.$$

Marginal cdf

Similarly, we can define "marginal cdf".

Definition: Let $\mathbf{X} = (X_1, \ldots, X_n)$ be an **arbitrary** random vector with joint cdf $F(\mathbf{x})$. Then the marginal cdf of each X_i is

$$F_{X_i}(x_i) = \lim_{\substack{x_1, \ldots, x_n \to \infty \\ x_i \text{ omitted}}} F(x_1, \ldots, x_n).$$

Remember: In general, the word "marginal" is used to indicate that there is some joint probability distribution for multiple R.V.'s.

- The marginal pmf is obtained from the joint pmf.
- The marginal pdf is obtained from the joint pdf.
- The marginal cdf is obtained from the joint cdf.

Independence of R.V.'s revisited

Recall: (Theorem in previous lecture)

- Two discrete R.V.'s X and Y are independent if and only if their joint pmf is the **product of the marginal pmf's**.
- Two continuous R.V.'s X and Y are independent if their joint pdf is the **product of the marginal pdf's**. if and only if

More generally, we can extend the theorem to multiple R.V.'s:

Theorem:

- A collection of discrete R.V.'s is independent if and only if the joint pmf is the **product of the marginal pmf's**.
- A collection of continuous R.V.'s is independent if the joint pdf is the **product of the marginal pdf's**.

Example 4

Let (X, Y, Z) be a continuous random vector with joint pdf

$$f(x, y, z) = \begin{cases} xy + z, & \text{if } 0 \leq x, y, z \leq 1; \\ 0, & \text{otherwise.} \end{cases}$$

What are the marginal pdf's for each of X, Y, Z? Are these random variables independent?

Example 4

Solution:

The marginal pdf of X is:

$$\begin{aligned} f_X(x) &= \int_0^1 \int_0^1 (xy+z)\,dy\,dz = \int_0^1 \left[\tfrac{1}{2}xy^2 + zy\right]_{y=0}^{y=1} dz \\ &= \int_0^1 \left(\tfrac{1}{2}x + z\right) dz = \left[\tfrac{1}{2}xz + \tfrac{1}{2}z^2\right]_{z=0}^{z=1} \\ &= \tfrac{x+1}{2}. \end{aligned}$$

Similarly, the marginal pdf of Y is:

$$\begin{aligned} f_Y(y) &= \int_0^1 \int_0^1 (xy+z)\,dx\,dz = \int_0^1 \left[\tfrac{1}{2}yx^2 + zx\right]_{x=0}^{x=1} dz \\ &= \int_0^1 \left(\tfrac{1}{2}y + z\right) dz = \left[\tfrac{1}{2}yz + \tfrac{1}{2}z^2\right]_{z=0}^{z=1} \\ &= \tfrac{y+1}{2}. \end{aligned}$$

Example 4

The marginal pdf of Z is:

$$\begin{aligned} f_Z(z) &= \int_0^1 \int_0^1 (xy+z)\,dx\,dy = \int_0^1 \left[\tfrac{1}{2}x^2y + zx\right]_{x=0}^{x=1} dy \\ &= \int_0^1 \left(\tfrac{1}{2}y + z\right) dy = \left[\tfrac{1}{4}y^2 + zy\right]_{y=0}^{y=1} \\ &= \tfrac{4z+1}{4}. \end{aligned}$$

Since

$$f_X(x)f_Y(y)f_Z(z) = \begin{cases} \frac{(x+1)(y+1)(4z+1)}{16}, & \text{if } 0 \le x, y, z \le 1; \\ 0, & \text{otherwise;} \end{cases}$$

we have that $f_X(x)f_Y(y)f_Z(z) \neq f(x,y,z)$.

Therefore, we conclude that X, Y, Z are not independent.

Independence of multiple general R.V.'s

Criteria for independence so far:

- joint pmf = product of marginal pmf's (discrete R.V.'s)
- joint pdf = product of marginal pdf's (continuous R.V.'s)

There is a more general condition for the independence of multiple **arbitrary** R.V.'s, in terms of "joint cdf" and "marginal cdf":
if some are discrete, some are continuous

Theorem: Let $\mathbf{X} = (X_1, \ldots, X_n)$ be an **arbitrary** random vector with joint cdf $F(\mathbf{x})$. Let $F_{X_i}(x_i)$ be the marginal cdf of each X_i. Then $X_1, \ldots, X_n$ are independent if and only if

$$F(\mathbf{x}) = F_{X_1}(x_1) \cdots F_{X_n}(x_n),$$

i.e. the joint cdf is the **product of the marginal cdf's**.

iid R.V.'s

Let $X_1, \ldots, X_n$ be R.V.'s.

If these n R.V.'s are **independent**, and each X_i has the same marginal cdf, then we say that $X_1, \ldots, X_n$ are independent and identically distributed.

This is a very common condition when dealing with multiple R.V.'s, so the term is usually abbreviated as i.i.d. or iid.

Any collection of iid R.V.'s is called a random sample, and the number n is called the sample size.

We will frequently be dealing with random samples later in this course, especially random samples with large sample sizes.

Summary

- Conditional distributions
- Conditional pmf/pdf
- Bayes' theorem for R.V.'s
- Random vectors
- Marginal pmf/pdf/cdf for multiple R.V.'s
- Independence of multiple R.V.'s