STAT 330 - Mathematical Statistics

Cameron Roopnarine

Last updated: October 21, 2020

Contents

Chapter 2

Univariate Random Variable

LECTURE 1 | 2020-09-09

Review probability model, random variable (r.v.), expectation, and moment generating function.

2.1 Probability Model and Random Variable

DEFINITION 2.1.1: Probability model

A **probability model** is used for a random experiment and consists of three components:
(I) Sample space
(II) Event
(III) Probability (density) function

DEFINITION 2.1.2: Sample space

A **sample space** S is the collection of all possible outcomes of one single random experiment.

DEFINITION 2.1.3: Event

An **event** is a subset of S and is denoted by A.

EXAMPLE 2.1.4

Toss a coin twice. This is a random experiment because we do not know the outcome before we toss the coin twice.

- $S = \{(H,H),(H,T),(T,H),(T,T)\}$

Define A: First toss is an H.
Clearly, $A = \{(H,H),(H,T)\} \subseteq S$, so A is an event.

DEFINITION 2.1.5: Probability function

A **probability function** $\mathbb{P}(\,\cdot\,)$ is a function that satisfies the following axioms:

(I) $\mathbb{P}(A) \geqslant 0$ for any event A

(II) $\mathbb{P}(S) = 1$

(III) *Additivity property*: If $A_1, A_2, A_3, \ldots$ are pairwise mutually exclusive events; that is, $A_i \cap A_j = \emptyset$ for all $i \neq j$, then

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$

EXAMPLE 2.1.6

Toss a coin twice, given one event A,

$$\mathbb{P}(A) = \frac{\#\text{ of outcomes in } A}{4}$$

since $|S| = 4$. $\mathbb{P}(\,\cdot\,)$ satisfies the three properties, therefore $\mathbb{P}(\,\cdot\,)$ is a probability function.

PROPOSITION 2.1.7: Additional Properties of the Probability Set Function

Let A and B be events with sample space S and let $\mathbb{P}(\,\cdot\,)$ be a probability function, then

(1) $\mathbb{P}(\emptyset) = 0$

(2) If A and B are mutually exclusive events, then $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$

(3) $\mathbb{P}(\bar{A}) = 1 - \mathbb{P}(A)$

(4) If $A \subseteq B$, then $\mathbb{P}(A) \leqslant \mathbb{P}(B)$

Proof of: 2.1.7

Proof of (1): Let $A_1 = S$ and $A_i = \emptyset$ for $i = 2, 3, \ldots$. Since $\bigcup_{i=1}^{\infty} A_i = S$, then by (III) it follows that

$$\mathbb{P}(S) = \mathbb{P}(S) + \sum_{i=2}^{\infty} \mathbb{P}(\emptyset)$$

and by (II) we have

$$1 = 1 + \sum_{i=2}^{\infty} \mathbb{P}(\emptyset)$$

By (I) the right side is a series of non-negative numbers which must converge to the left side which is 1 which is finite which results in a contradiction unless $\mathbb{P}(\emptyset) = 0$ as required.

Proof of (2): Let $A_1 = A$, $A_2 = B$, and $A_i = \emptyset$ for $i = 3, 4, \ldots$. Since $\bigcup_{i=1}^{\infty} A_i = A \cup B$, then by (III)

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) + \sum_{i=3}^{\infty} \mathbb{P}(\emptyset)$$

and since $\mathbb{P}(\emptyset) = 0$ by the result of (1) it follows that

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$$

Proof of (3): Since $S = A \cup \bar{A}$ and $A \cap \bar{A} = \emptyset$ then by (II) and by (2) it follows that

$$1 = \mathbb{P}(S) = \mathbb{P}(A \cup \bar{A}) = \mathbb{P}(A) + \mathbb{P}(\bar{A})$$

as required.
Proof of (4): Since

$$B = (A \cap B) \cup (\bar{A} \cap B) = A \cup (\bar{A} \cap B)$$

and $A \cap (\bar{A} \cap B) = \emptyset$ then by (2)

$$\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(\bar{A} \cap B)$$

But by (I), $\mathbb{P}(\bar{A} \cap B) \geqslant 0$, so the result now follows.

EXERCISE 2.1.8

Let A and B be events with sample space S and let $\mathbb{P}(\,\cdot\,)$ be a probability function, then prove the following:

1. $0 \leqslant \mathbb{P}(A) \leqslant 1$
2. $\mathbb{P}(A \cap \bar{B}) = \mathbb{P}(A) - \mathbb{P}(A \cap B)$
3. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$

1. $\mathbb{P}(A) \geqslant 0$ follows from (I). From (3) we have $\mathbb{P}(\bar{A}) = 1 - \mathbb{P}(A)$. But from (I) $\mathbb{P}(\bar{A}) = \geqslant 0$ and therefore $\mathbb{P}(A) \leqslant 1$.
2. Since $A = (A \cap B) \cup (A \cap \bar{B})$ and $(A \cap B) \cap (A \cap \bar{B}) = \emptyset$, then by (2)

$$\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap \bar{B})$$

as required.

3. $\mathbb{P}(A \cup B) = (A \cap \bar{B}) + \mathbb{P}(A \cap B) + \mathbb{P}(\bar{A} \cap B)$. By the previous result,

$$\mathbb{P}(A \cap \bar{B}) = \mathbb{P}(A) - \mathbb{P}(A \cap B) \quad \text{and} \quad \mathbb{P}(\bar{A} \cap B) = \mathbb{P}(B) - \mathbb{P}(A \cap B)$$

Therefore,

$$\begin{aligned} \mathbb{P}(A \cup B) &= (\mathbb{P}(A) - \mathbb{P}(A \cap B)) + \mathbb{P}(A \cap B) + (\mathbb{P}(B) - \mathbb{P}(A \cap B)) \\ &= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) \end{aligned}$$

as required.

DEFINITION 2.1.9: Conditional probability

Suppose A and B are two events with $\mathbb{P}(B) > 0$. Then the **conditional probability** of A given that B has occurred is

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

DEFINITION 2.1.10: Independent events

Suppose A and B are two events. A and B are **independent events** if and only if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$

Clearly, $\mathbb{P}(A \mid B) = \mathbb{P}(A)$ if and only if A and B are independent since

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A)\mathbb{P}(B)}{\mathbb{P}(B)} = \mathbb{P}(A)$$

EXAMPLE 2.1.11

Toss a coin twice.

- A: First toss is H
- B: Second toss is T

$$\mathbb{P}(A) = \frac{\text{\# of outcomes in } A}{4} = \frac{2}{4} \quad \text{and} \quad \mathbb{P}(B) = \frac{2}{4}$$

$$\mathbb{P}(A \cap B) = \frac{1}{4} = \mathbb{P}(A)\mathbb{P}(B)$$

therefore A and B are independent.

DEFINITION 2.1.12: Random variable

A **random variable** (r.v.) X is a function from a sample space S to the real numbers $\mathbb{R}$; that is,

$$X : S \to \mathbb{R}$$

satisfies for any given $x \in \mathbb{R}$ $\{X \leqslant x\}$ is an event.

$$\{X \leqslant x\} = \{\omega \in S : X(\omega) \leqslant x\} \subseteq S$$

EXAMPLE 2.1.13

Toss a coin twice. Let X be the number of heads (H) in two tosses. Verify that X is a random variable.
Solution. Possible values of X: $0, 1, 2$. Given $x \in \mathbb{R}$, $\{X \leqslant x\}$.

- $x < 0 \implies \{X \leqslant x\} = \emptyset$
- $x = 0 \implies \{X \leqslant x\} = \{X = 0\} = \{(T, T)\} \subseteq S$
- $x = 1 \implies \{X \leqslant x\} = \{X = 1\} = \{(H, T), (T, H)\} \subseteq S$
- $x = 2 \implies \{X \leqslant x\} = \{X = 2\} = \{(H, H)\} \subseteq S$

Thus, X is a random variable.

DEFINITION 2.1.14: Cumulative distribution function

The **cumulative distribution function** (c.d.f.) of a random variable X is defined by

$$F(x) = \mathbb{P}(X \leqslant x)$$

for all $x \in \mathbb{R}$. Note that the c.d.f. is defined for all $\mathbb{R}$.

DEFINITION 2.1.15: Properties — Cumulative Distribution Function

(1) F is a non-decreasing function; that is, if $x_1 \leqslant x_2$, then $F(x_1) \leqslant F(x_2)$.
By looking at:
- $\{X \leqslant x_1\} \subseteq \{X \leqslant x_2\}$ if $x_1 \leqslant x_2$.

(2) $\lim\limits_{x\to\infty} F(x) = 1$ and $\lim\limits_{x\to-\infty} F(x) = 0$.
By looking at:
- $x \to \infty$: $\{X \leqslant x\} \to S$
- $x \to -\infty$: $\{X \leqslant x\} \to \emptyset$

(3) $F(x)$ is a right continuous function; that is, for any $a \in \mathbb{R}$,

$$\lim_{x\to a^+} F(a) = F(a)$$

(4) For all $a < b$

$$\mathbb{P}(a < X \leqslant b) = \mathbb{P}(X \leqslant b) - \mathbb{P}(X \leqslant a) = F(b) - F(a)$$

(5) For all b

$$\mathbb{P}(X = b) = \mathbb{P}(\text{jump at } b) = \lim_{t\to b^+} F(t) - \lim_{t\to b^-} F(t) = F(b) - \lim_{t\to b^-} F(t)$$

LECTURE 2 | 2020-09-09

2.2 Discrete Random Variables

DEFINITION 2.2.1: Discrete random variable

If a random variable X can only take finite or countable values, X is a **discrete random variable**.

In this case, $F(x)$ is a right-continuous step function.

REMARK 2.2.2

When we say **countable**, we mean something you can enumerate such as $\mathbb{Z}$ or $\mathbb{N}^+$.

DEFINITION 2.2.3: Probability function

If X is a discrete random variable, then the **probability function** (p.f.) of X is given by

$$f(x) = \begin{cases} \mathbb{P}(X = x) = F(x) - \lim\limits_{\varepsilon\to 0^+} F(x - \varepsilon) & \text{if } X \text{ can take value } x \\ 0 & \text{if } X \text{ cannot take value } x \end{cases}$$

DEFINITION 2.2.4: Support set

The set $A = \{x : f(x) > 0\}$ is called the **support set** of X. These are all the possible values X can take.

PROPOSITION 2.2.5: Properties of the Probability Function

(1) $f(x) \geqslant 0$ *for all* $x \in \mathbb{R}$
(2) $\sum\limits_{x\in A} f(x) = 1$

Review some commonly used discrete random variables:

- **Bernoulli.** $X \sim$ Bernoulli(p) where X can only take two possible values 0 (failure) or 1 (success). Let p be the probability of a success for a single trial. So,

$$\mathbb{P}(X=1)=p \quad \text{and} \quad \mathbb{P}(X=0)=1-p$$

  Therefore,

$$f(x)=\mathbb{P}(X=x)=p^x(1-p)^{1-x}$$

  Example: Toss a coin twice. Let X be the number of heads. Then $X \sim$ Bernoulli(p)

- **Binomial.** $X \sim$ Binomial(n,p). Suppose we have **Bernoulli Trials**:
  - We run n trials
  - Each trial is independent of each other
  - Each trial has two possible outcomes: 0 (failure), 1 (success)

$$\mathbb{P}(X=1)=p$$

  Let X be the number of success across these n trials and p be the success probability for a single trial.

$$X=\sum_{i=1}^{n} X_i$$

  X_i is the outcome of the ith trial.

$$\mathbb{P}(X_i=1)=p$$

  where $X_i \sim$ Bernoulli(p). Therefore,

$$f(x)=\mathbb{P}(X=x)=\binom{n}{x}p^x(1-p)^{n-x}$$

- **Geometric.** $X \sim$ Geometric(p). Let X be the number of failures before the first success. X can take values $0,1,2,\ldots$.

$$f(x)=\mathbb{P}(X=x)=(1-p)^x p$$

  Example. $X =$ number of tails before you get the first head.

- **Negative Binomial.** $X \sim$ Negative Binomial(r,p). Let X be the number of failures before you get r success. X can take values $0,1,2,\ldots$

$$f(x)=\mathbb{P}(X=x)=\binom{x+r-1}{x}(1-p)^x p^{r-1} p$$

  Example. $X =$ number of tails before you get the rth head.

- **Poisson.** $X \sim$ Poisson(μ) where $X=0,1,\ldots$.

$$f(x)=\mathbb{P}(X=x)=\frac{\mu^x}{x!}e^{-\mu}$$

  where $x=0,1,2,\ldots$.

2.3 Continuous Random Variables

DEFINITION 2.3.1: Continuous random variable

Suppose X is a random variable with c.d.f. F. If F is a continuous function for all $x \in \mathbb{R}$ and F is differentiable except possibly at countably many points, then X is called a **continuous random variable**.

Note that this is not a rigorous definition, but it will be used in this course.

DEFINITION 2.3.2: Probability density function, Support set

The **probability density function** (p.d.f.) of a continuous random variable is

$$f(x) = \begin{cases} F'(x) & \text{if } F(x) \text{ is differentiable at } x \\ 0 & \text{otherwise} \end{cases}$$

The set $A = \{x : f(x) > 0\}$ is called the **support set** of X.

Continuous case: $f(x) \neq \mathbb{P}(X = x)$

$$\mathbb{P}(x < X \leqslant x + \delta) \approx f(x)\delta$$

since

$$\lim_{\delta \to 0} \frac{F(x+\delta) - F(x)}{\delta} = F'(x) = f(x)$$

DEFINITION 2.3.3: Properties — Probability Density Function

(I) $f(x) \geqslant 0$ for all $x \in \mathbb{R}$

(II) $\displaystyle\int_{-\infty}^{\infty} f(x)\,dx = \lim_{x\to\infty} F(x) - \lim_{x\to-\infty} F(x) = 1$

(III) $\displaystyle f(x) = F'(x) = \lim_{h\to 0} \frac{F(x+h) - F(x)}{h}$

(IV) $\displaystyle F(x) = \int_{-\infty}^{x} f(t)\,dt$ since $F(-\infty) = 0$.

(V) $\displaystyle \mathbb{P}(a < X \leqslant b) = \mathbb{P}(X \leqslant b) - \mathbb{P}(X \leqslant a) = F(b) - F(a) = \int_a^b f(x)\,dx$

(VI) $\displaystyle \mathbb{P}(X = b) = F(b) - \lim_{a\to b^-} F(a) = F(b) - F(b) = 0 \neq f(b)$ since F is continuous.

EXAMPLE 2.3.4

Suppose the c.d.f. of X is

$$F(x) = \begin{cases} 0 & x \leqslant a \\ \dfrac{x-a}{b-a} & a < x < b \\ 1 & x \geqslant b \end{cases}$$

Find the p.d.f. of X.
Solution.

$$f(x) = \begin{cases} \dfrac{1}{b-a} & a < x < b \\ 0 & \text{otherwise} \end{cases}$$

We note that $X \sim \text{Uniform}[a, b]$

EXAMPLE 2.3.5

Let the p.d.f. be defined as follows.

$$f(x) = \begin{cases} \dfrac{\theta}{x^{\theta+1}} & x \geqslant 1 \\ 0 & x < 1 \end{cases}$$

(i) For what values of θ is f a p.d.f.

(ii) Find $F(x)$.
(iii) Find $\mathbb{P}(-2 < X < 3)$.

Solution.

(i) Note that $\dfrac{\theta}{x^{\theta+1}} \geqslant 0$ for all $\theta \geqslant 0$.

Case 1: $\theta = 0$. $f(x) \equiv 0$, then f cannot be a pdf since $\int_{-\infty}^{\infty} f(x)\,dx = 0 \neq 1$

Case 2: $\theta > 0$.

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{1} f(x)\,dx + \int_{1}^{\infty} f(x)\,dx = \int_{1}^{\infty} \frac{\theta}{x^{\theta+1}}\,dx = \left[-x^{-\theta}\right]_1^{\infty} = 1$$

Therefore, f is a p.d.f. when $\theta > 0$.

(ii) $F(x) = \mathbb{P}(X \leqslant x)$.

Case 1: $x < 1$.

$$\mathbb{P}(X \leqslant x) = \int_{-\infty}^{x} f(t)\,dt = 0$$

Case 2: $x \geqslant 1$.

$$\mathbb{P}(X \leqslant x) = \int_{-\infty}^{x} f(t)\,dt = \int_{-\infty}^{1} f(t)\,dt + \int_{1}^{x} f(t)\,dt = \int_{1}^{x} \frac{\theta}{t^{\theta+1}}\,dt = \left[-t^{-\theta}\right]_1^{x} = 1 - x^{-\theta}$$

(iii) $\mathbb{P}(-2 < X < 3)$. Either use the c.d.f. we found or the p.d.f.

Using the c.d.f. we have

$$F(3) - F(-2) = (1 - 3^{-\theta}) - 0$$

Using the p.d.f. we have

$$\int_{-2}^{3} f(x)\,dx = \int_{-2}^{1} f(x)\,dx + \int_{1}^{3} f(x)\,dx = \int_{1}^{3} f(x)\,dx = \text{exercise}$$

Lecture 3 | 2020-09-13

We first introduce a function that will be used.

DEFINITION 2.3.6: Gamma function

The **gamma function**, denoted $\Gamma(\alpha)$ for all $\alpha > 0$, is given by

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x}\,dx$$

PROPOSITION 2.3.7: Properties of the Gamma Function

(1) $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$ *for* $\alpha > 1$
(2) $\Gamma(n) = (n-1)!$ *when* $n \geqslant 1$ *is a positive integer*
(3) $\Gamma\left(\dfrac{1}{2}\right) = \sqrt{\pi}$

We don't need to know the following proof, but I checked it out for fun. Content not found in the syllabus is usually labelled with a dagger (†).

Proof of: † 2.3.7

Proof of (1). Suppose $\alpha > 1$.

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}\,dx$$

Let $u = x^{\alpha-1} \implies du = (\alpha-1)x^{\alpha-2}\,dx$ and $dv = e^{-x}\,dx \implies v = -e^{-x}$. Now, recall from MATH 138:

$$\int u\,dv = uv - \int v\,du$$

So,

$$\begin{aligned}
\Gamma(\alpha) &= \left[(\alpha-1)x^{\alpha-2}\left(-e^{-x}\right)\right]_0^\infty - \int_0^\infty \left(-e^{-x}\right)(\alpha-1)x^{\alpha-2}\,dx \\
&= 0 + (\alpha-1)\int_0^\infty e^{-x}x^{\alpha-2}\,dx \\
&= (\alpha-1)\Gamma(\alpha)
\end{aligned}$$

Proof of (2). Using (1):

$$\begin{aligned}
\Gamma(\alpha) &= (\alpha-1)\Gamma(\alpha-1) \\
&= (\alpha-1)(\alpha-2)\Gamma(\alpha-3) \\
&= (\alpha-1)(\alpha-2)\cdots(3)(2)(1)\Gamma(1)
\end{aligned}$$

We know that $\Gamma(1) = 1$ by using the definition (trivial), therefore the result now follows.
Proof of (3). Sketch:

- Let $u = x^2$, so $du = 2x\,dx$. Let $\alpha = \dfrac{1}{2}$, so the integral looks like:

$$\Gamma\left(\frac{1}{2}\right) = 2\int_0^\infty e^{-u^2}\,du$$

- Compute $\left[\Gamma\left(\frac{1}{2}\right)\right]^2$. Using polar coordinates, compute the following double integral.

$$4\int_0^\infty \int_0^\infty e^{-(u^2+v^2)}\,dv\,du$$

  One will have to compute the Jacobian Matrix.
- Solve for $\Gamma\left(\dfrac{1}{2}\right)$ explicitly now.

Author's note: This was covered in MATH 237 when I took it (F19).

EXAMPLE 2.3.8

The p.d.f. is given by

$$f(x) = \begin{cases} \dfrac{x^{\alpha-1}e^{-x/\beta}}{\Gamma(\alpha)\beta^\alpha} & x > 0 \\ 0 & x \leqslant 0 \end{cases}$$

when $\alpha > 0$ and $\beta > 0$. We say that $X \sim \text{Gamma}(\alpha, \beta)$.
We also say that α is the scale parameter and β is the shape parameter for this distribution.
Verify that $f(x)$ is a p.d.f.

Solution. Showing $f(x) \geqslant 0$ is trivial. Now,

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_0^{\infty} \frac{x^{\alpha-1}e^{-x/\beta}}{\Gamma(\alpha)\beta^{\alpha}}\,dx$$

Let $y = x/\beta \implies x = y\beta$ and $dx = \beta\,dy$. Therefore,

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_0^{\infty} \frac{y^{\alpha-1}\beta^{\alpha-1}e^{-y}}{\Gamma(\alpha)\beta^{\alpha}}(\beta)\,dy = \frac{1}{\Gamma(\alpha)}\int_0^{\infty} y^{\alpha-1}e^{-y}\,dy = 1$$

EXAMPLE 2.3.9

Suppose the p.d.f. is given by

$$f(x) = \begin{cases} \dfrac{\beta}{\theta^{\beta}}x^{\beta-1}e^{-(x/\theta)^{\beta}} & x > 0 \\ 0 & x \leqslant 0 \end{cases}$$

with $\alpha > 0$ and $\beta > 0$. Then, $X \sim$ Weibull(θ, β). Verify that $f(x)$ is a p.d.f.
Solution. $f(x) \geqslant 0$ for every $x \in \mathbb{R}$. Now,

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_0^{\infty} \frac{\beta}{\theta^{\beta}}x^{\beta-1}e^{-(x/\theta)^{\beta}}\,dx$$

Let $y = (x/\theta)^{\beta} \implies x = \theta y^{1/\beta}$ and $dx = (\theta/\beta)y^{(1/\beta)-1}\,dy$. Therefore,

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_0^{\infty} \frac{\beta}{\theta^{\beta}}\theta^{\beta-1}y^{(\beta-1)/\beta}e^{-y}\frac{\theta}{\beta}y^{(1/\beta)-1}\,dy = \int_0^{\infty} e^{-y}\,dy = \Gamma(1) = 1$$

EXAMPLE 2.3.10: Normal

The p.d.f. is given by

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

for $x \in \mathbb{R}$, $\mu \in \mathbb{R}$, $\sigma^2 > 0$. Verify that $f(x)$ is a p.d.f.
Solution.
$f(x) \geqslant 0$ obviously.
Case 1: $\mu = 0$ and $\sigma^2 = 1$, then we say X follows a **standard normal** distribution. We want to show that

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{x^2}{2}\right\}\,dx = 1$$

Since the function is symmetrical around 0, we have the following equivalent integral.

$$2\int_0^{\infty} \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{x^2}{2}\right\}\,dx$$

Let $y = x^2/2 \implies x = \sqrt{2y}$ and $dx = \dfrac{\sqrt{2}}{2}y^{-1/2}\,dy$. Therefore,

$$= \frac{2}{\sqrt{2\pi}}\int_0^{\infty} e^{-y}\frac{\sqrt{2}}{2}y^{-1/2}\,dy = \frac{1}{\sqrt{\pi}}\int_0^{\infty} y^{1/2-1}e^{-y}\,dy = \left(\frac{1}{\sqrt{\pi}}\right)\Gamma\left(\frac{1}{2}\right) = 1$$

Case 2: For general μ and σ^2,

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}\,dx$$

Let $z = \dfrac{x-\mu}{\sigma} \implies x = \mu + \sigma z$ and $dx = \sigma\, dz$. Therefore,

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{z^2}{2}\right\} \sigma\, dz = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{z^2}{2}\right\} dz = 1$$

using Case 1.

2.4 Expectation

DEFINITION 2.4.1: Expectation of discrete random variable

Suppose X is a discrete random variable with support A and p.f. $f(x)$. Then,

$$\mathbb{E}[X] = \sum_{x \in A} x f(x)$$

if $\sum_{x\in A} |x| f(x) < \infty$ (finite). If $\sum_{x\in A} |x| f(x) = \infty$ (infinite), then $\mathbb{E}[X]$ does not exist.

DEFINITION 2.4.2: Expectation of continuous random variable

Suppose X is a continuous random variable with support A and p.d.f. $f(x)$. Then,

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x)\, dx$$

if $\int_{-\infty}^{\infty} |x| f(x)\, dx < \infty$ (finite). Similarly, if $\int_{-\infty}^{\infty} |x| f(x)\, dx = \infty$ (infinite), then $\mathbb{E}[X]$ does not exist.

EXAMPLE 2.4.3: Discrete

Suppose

$$f(x) = \frac{1}{x(x+1)} = \frac{1}{x} - \frac{1}{x+1}$$

for $x = 1, 2, \ldots$. The support set is $A = \{1, 2, \ldots\}$. We note that $f(x)$ is a p.f. since $f(x) \geqslant 0$ and

$$\sum_{x \in A} f(x) = \sum_{x=1}^{\infty} \left(\frac{1}{x} - \frac{1}{x+1}\right) = 1 - \frac{1}{2} + \frac{1}{2} - \frac{1}{3} + \cdots = 1$$

Find $\mathbb{E}[X]$.
Solution.

$$\sum_{x\in A} |x| f(x) = \sum_{x=1}^{\infty} x \left(\frac{1}{x} - \frac{1}{x+1}\right) = \sum_{x=1}^{\infty} \frac{1}{x+1} = \infty$$

Therefore, $\mathbb{E}[X]$ does not exist!

EXAMPLE 2.4.4: Continuous

Let the p.d.f. be defined as $f(x) = \dfrac{1}{x^2+1}$ for $x \in \mathbb{R}$. This is known as the Cauchy distribution (or Student's T-distribution with 1 degree of freedom). Find $\mathbb{E}[X]$.

Solution.

$$\int_{-\infty}^{\infty} |x| f(x)\,dx = \int_{-\infty}^{\infty} |x| \frac{1}{x^2+1}\,dx = 2\int_{0}^{\infty} \frac{x}{x^2+1}\,dx = \left[\ln|x^2+1|\right]_0^{\infty} = \infty$$

$\mathbb{E}[X]$ does not exist! The following is **wrong**:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x)\,dx = \int_{-\infty}^{\infty} \frac{x}{1+x^2}\,dx = 0$$

since the integral above with $|x|$ is infinite. You must always remember to check that the $\mathbb{E}[X]$ is finite (using $|X|$) for both the discrete and continuous case.

EXAMPLE 2.4.5: Bernoulli and Binomial Random Variable

Suppose $X \sim \text{Bernoulli}(p)$.

$$\mathbb{P}(X=1) = p \quad \text{and} \quad \mathbb{P}(X=0) = 1-p$$

We know $\mathbb{E}[X] = (1)\mathbb{P}(X=1) + (0)\mathbb{P}(X=0) = p$
Now suppose $X \sim \text{Binomial}(n,p)$. Find $\mathbb{E}[X]$.
Solution.

$$\mathbb{E}[X] = \sum_{x \in A} x f(x) = \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x}$$

This is hard to do. But, we know we can use the relationship between the Binomial and Bernoulli random variable so,

$$X = \sum_{i=1}^{n} X_i$$

Therefore,

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbb{E}[X_i] = np$$

EXAMPLE 2.4.6

Suppose for a random variable X the p.d.f. is given by $f(x) = \dfrac{\theta}{x^{\theta+1}}$ for $x \geqslant 1$ and 0 when $x < 1$. Assume $\theta > 0$. Find $\mathbb{E}[X]$ and for what values of θ, does $\mathbb{E}[X]$ exist.
Solution.

$$\int_{-\infty}^{\infty} |x| f(x)\,dx = \int_{1}^{\infty} (x) \frac{\theta}{x^{\theta+1}}\,dx = \theta \int_{1}^{\infty} \frac{1}{x^{\theta}}\,dx < \infty \iff \theta > 1$$

from MATH 138. So, if $\theta > 1$ then $\mathbb{E}[X]$ exists. Also,

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x)\,dx = \int_{1}^{\infty} \frac{\theta x}{x^{\theta+1}}\,dx = \theta \int_{1}^{\infty} \frac{1}{x^{\theta}}\,dx = \frac{\theta}{\theta - 1}$$

DEFINITION 2.4.7: Expectation (Discrete)

If X is a discrete random variable with probability function $f(x)$ and support set A, then the **expectation** of the random variable $g(X)$ is defined by

$$\mathbb{E}[g(X)] = \sum_{x \in A} g(x) f(x)$$

provided the sum converges absolutely; that is, provided

$$\sum_{x \in A} |g(x)| f(x) < \infty$$

DEFINITION 2.4.8: Expectation (Continuous)

If X is a continuous random variable with p.d.f. $f(x)$ and support set A, then the **expectation** of the random variable $g(X)$ is defined by

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f(x)\, dx$$

provided the integral converges absolutely; that is, provided

$$\int_{-\infty}^{\infty} |g(x)| f(x)\, dx < \infty$$

THEOREM 2.4.9: Expectation is a Linear Operator

Suppose X is a random variable with probability (density) function $f(x)$, and a and b are real constants, and $g(x)$ and $h(x)$ are real-valued functions. Then,

$$\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$$

$$\mathbb{E}[ag(X) + bh(X)] = a\mathbb{E}[g(X)] + b\mathbb{E}[h(X)]$$

Proof of: 2.4.9

Omitted from the lecture and hence these notes. See Course Notes for most of the proof.

DEFINITION 2.4.10: Variance

The variance of a random variable is defined as

$$\sigma^2 = \mathbb{V}(X) = \mathbb{E}[(X - \mu)^2]$$

where $\mu = \mathbb{E}[X]$.

DEFINITION 2.4.11: Special Expectations

(I) The kth moment (about the origin) of a random variable

$$\mathbb{E}[X^k]$$

(II) The kth moment about the mean of a random variable

$$\mathbb{E}[(X-\mu)^k]$$

THEOREM 2.4.12: Properties of Variance

If X is a random variable, then

$$\mathbb{V}(X) = \mathbb{E}[X^2] - \mu^2$$

where $\mu = \mathbb{E}[X]$. Note that the variance of X exists if $\mathbb{E}[X^2] < \infty$.

Proof of: 2.4.12

Omitted from the lecture and hence these notes. See Course Notes for most of the proof.

EXAMPLE 2.4.13

Suppose $X \sim \text{Poisson}(\theta)$, the p.f. is defined as $f(x) = \dfrac{\theta^x}{x!}e^{-\theta}$ for $x = 0, 1, 2, \ldots$. Find $\mathbb{E}[X]$ and $\mathbb{V}(X)$.
Solution. The support is non-negative, so $|x| = x$. Therefore,

$$\mathbb{E}[X] = \sum_{x=0}^{\infty} x\frac{\theta^x}{x!}e^{-\theta} = \sum_{x=1}^{\infty} \frac{x}{x!}\theta^x e^{-\theta} = \theta \sum_{x=1}^{\infty} \frac{\theta^{x-1}}{(x-1)!}e^{-\theta}$$

Let $y = x - 1$, then

$$\mathbb{E}[X] = \theta \sum_{y=0}^{\infty} \frac{\theta^y}{y!}e^{-\theta}$$

We know $e^\theta = \displaystyle\sum_{y=0}^{\infty} \frac{\theta^y}{y!}$, so $\mathbb{E}[X] = \theta$.

$$\mathbb{V}(X) = \mathbb{E}[X^2] - \mu^2$$

Let's find $\mathbb{E}[X^2]$:

$$\begin{aligned}
\mathbb{E}[X^2] &= \sum_{x=0}^{\infty} x^2 \frac{\theta^x}{x!}e^{-\theta} \\
&= \sum_{x=1}^{\infty} \frac{x}{(x-1)!}\theta^x e^{-\theta} \\
&= \sum_{x=1}^{\infty} \frac{(x-1)+1}{(x-1)!}\theta^x e^{-\theta} \\
&= \sum_{x=1}^{\infty} \frac{x-1}{(x-1)!}\theta^x e^{-\theta} + \sum_{x=1}^{\infty} \frac{1}{(x-1)!}\theta^x e^{-\theta}
\end{aligned}$$

Looking at the first sum (since the second sum was computed before):

$$\sum_{x=2}^{\infty} \frac{\theta^2}{(x-2)!}\theta^{x-2}e^{-\theta} + \theta$$

Let $y = x - 2$:

$$\mathbb{E}[X^2] = \sum_{y=0}^{\infty} \frac{\theta^2 \theta^y}{y!} e^{-\theta} + \theta = \theta^2 + \theta$$

Therefore,

$$\mathbb{V}(X) = \mathbb{E}[X^2] - \mu^2 = (\theta^2 + \theta) - \theta^2 = \theta$$

Lecture 4 | 2020-09-13

EXAMPLE 2.4.14

If $X \sim \text{Gamma}(\alpha, \beta)$, prove that

$$\mathbb{E}[X^p] = \frac{\beta^p \Gamma(\alpha + p)}{\Gamma(\alpha)}$$

for $p > -\alpha$.

Solution. Recall that

$$f(x) = \begin{cases} \dfrac{x^{\alpha-1} e^{-x/\beta}}{\Gamma(\alpha)\beta^\alpha} & x > 0 \\ 0 & x \leqslant 0 \end{cases}$$

So,

$$\mathbb{E}[X^p] = \int_{-\infty}^{\infty} x^p f(x)\, dx = \int_0^{\infty} x^p \frac{x^{\alpha-1} e^{-x/\beta}}{\Gamma(\alpha)\beta^\alpha}\, dx$$

There are two methods to solve this integral:

Method 1: Rewrite the function as the p.d.f. of a gamma distribution.

$$= \int_0^{\infty} \frac{x^{p+\alpha-1} e^{-x/\beta}}{\Gamma(\alpha)\beta^\alpha}\, dx$$

which is close to the p.d.f. of $\text{Gamma}(p + \alpha, \beta)$.

$$= \int_0^{\infty} \frac{x^{p+\alpha-1} e^{-x/\beta}}{\Gamma(\alpha + p)\beta^{\alpha+p}} \times \underbrace{\frac{\Gamma(\alpha + p)\beta^{\alpha+p}}{\Gamma(\alpha)\beta^\alpha}}_{\text{constant}}\, dx = \frac{\Gamma(\alpha + p)\beta^p}{\Gamma(\alpha)} \times 1$$

Method 2: Rewrite the function as a gamma function.

$$\mathbb{E}[X^p] = \int_0^{\infty} \frac{x^{(p+\alpha)-1} e^{-x/\beta}}{\Gamma(\alpha)\beta^\alpha}\, dx$$

Let $y = x/\beta \implies x = \beta y$ and $dx = \beta\, dy$. Therefore,

$$= \int_0^{\infty} \frac{\beta^{p+\alpha-1} y^{(p+\alpha)-1} e^{-y}}{\Gamma(\alpha)\beta^\alpha} (\beta)\, dy = \frac{\beta^p}{\Gamma(\alpha)} \int_0^{\infty} y^{(p+\alpha)-1} e^{-y}\, dy = \frac{\Gamma(p + \alpha)}{\Gamma(\alpha)} \beta^p$$

Additionally,

- $\mathbb{E}[X] = \dfrac{\beta \Gamma(\alpha + 1)}{\Gamma(\alpha)} = \alpha\beta$
- $\mathbb{E}[X^2] = \dfrac{\beta^2 \Gamma(\alpha + 2)}{\Gamma(\alpha)} = \dfrac{\beta^2 (\alpha + 1)\Gamma(\alpha + 1)}{\Gamma(\alpha)} = \alpha(\alpha + 1)\beta^2$
- $\mathbb{V}(X) = \mathbb{E}[X^2] - \mu^2 = \alpha(\alpha + 1)\beta^2 - \alpha^2\beta^2 = \alpha\beta^2$

2.5 Moment Generating Functions

DEFINITION 2.5.1: Moment generating function

Suppose X is a random variable, then

$$M(t) = \mathbb{E}[e^{tX}]$$

is called the **moment generating function** (m.g.f.) of X if $M(t)$ exists for $t \in (-h, h)$ with some $h > 0$.

REMARK 2.5.2

If we are able to find some $h > 0$ such that for any $t \in (-h, h)$, $\mathbb{E}[e^{tX}] < \infty$, then we say $M(t)$ is the m.g.f. of X.

EXAMPLE 2.5.3

Suppose $X \sim \text{Gamma}(\alpha, \beta)$. Find $M(t)$. Recall the p.d.f. is

$$f(x) = \begin{cases} \dfrac{x^{\alpha-1}e^{-x/\beta}}{\mathit{\Gamma}(\alpha)\beta^\alpha} & x > 0 \\ 0 & x \leqslant 0 \end{cases}$$

Solution.

$$\begin{aligned} M(t) &= \mathbb{E}[e^{tX}] \\ &= \int_{-\infty}^{\infty} e^{tx} f(x)\, dx \\ &= \int_0^{\infty} e^{tx} \frac{x^{\alpha-1}e^{-x/\beta}}{\mathit{\Gamma}(\alpha)\beta^\alpha}\, dx \\ &= \int_0^{\infty} \frac{x^{\alpha-1}e^{-x\left(\frac{1}{\beta}-t\right)}}{\mathit{\Gamma}(\alpha)\beta^\alpha}\, dx \\ &= \int_0^{\infty} \frac{x^{\alpha-1}e^{-x/\tilde{\beta}}}{\mathit{\Gamma}(\alpha)\beta^\alpha}\, dx \end{aligned}$$

where

$$\tilde{\beta} = \frac{1}{\left(\dfrac{1}{\beta} - t\right)}$$

Continuing,

$$\begin{aligned} &= \int_0^{\infty} \frac{x^{\alpha-1}e^{-x/\tilde{\beta}}}{\mathit{\Gamma}(\alpha)\tilde{\beta}^\alpha} \left(\frac{\tilde{\beta}^\alpha}{\beta^\alpha}\right) dx \\ &= \frac{\tilde{\beta}^\alpha}{\beta^\alpha}(1) \\ &= (1 - \beta t)^{-\alpha} \end{aligned}$$

The moment generating function must be non-negative since $1 - \beta t > 0$ and therefore, $t < 1/\beta$. Take $h = 1/\beta$.

EXAMPLE 2.5.4

If $X \sim$ Poisson(θ), the p.f. is given by $f(x) = \dfrac{\theta^x e^{-\theta}}{x!}$ for $x = 0, 1, 2, \ldots$. Find $M(t)$.
Solution.

$$
\begin{aligned}
M(t) &= \mathbb{E}[e^{tX}] \\
&= \sum_{x=0}^{\infty} e^{tx} \frac{\theta^x e^{-\theta}}{x!} \\
&= \sum_{x=0}^{\infty} \frac{\left(e^t\theta\right)^x e^{-\theta}}{x!} \\
&= e^{-\theta} \sum_{x=0}^{\infty} \frac{\left(e^t\theta\right)^x}{x!} \\
&= e^{-\theta} \exp\left\{e^t\theta\right\} \\
&= \exp\left\{\theta\left(e^t - 1\right)\right\}
\end{aligned}
$$

for all $t \in \mathbb{R}$.

Three important properties of $M(t)$.

THEOREM 2.5.5: Moment Generating Function of a Linear Function

Suppose the random variable X has moment generating function $M_X(t)$ defined for $t \in (-h, h)$ for some $h > 0$. Let $Y = aX + b$ where $a, b \in \mathbb{R}$ and $a \neq 0$. Then, the moment generating function of Y is

$$M_Y(t) = e^{bt} M_X(at)$$

for $|t| < \dfrac{h}{|a|}$.

Proof of: 2.5.5

$$
\begin{aligned}
M_Y(t) &= \mathbb{E}[e^{tY}] \\
&= \mathbb{E}[e^{t(aX+b)}] \\
&= e^{bt}\mathbb{E}[e^{atX}] && \text{exists for } |at| < h \\
&= e^{bt} M_X(at) && \text{for } |t| < \frac{h}{|a|}
\end{aligned}
$$

as required.

EXAMPLE 2.5.6

(i) If $Z \sim N(0, 1)$, find $M_Z(t)$.
(ii) If $X \sim N(\mu, \sigma^2)$, find $M_X(t)$.
Solution.

(i)

$$\begin{aligned}
M_Z(t) &= \mathbb{E}[e^{tZ}] \\
&= \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\} dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2 - 2tx}{2}\right\} dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(x-t)^2 - t^2}{2}\right\} dx && \text{complete the square} \\
&= \exp\left\{\frac{t^2}{2}\right\} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(x-t)^2}{2}\right\} dx
\end{aligned}$$

where the integral is the p.d.f. of $N(\mu = t, \sigma^2 = 1)$. Therefore,

$$\mathbb{E}[e^{tZ}] = \exp\left\{\frac{t^2}{2}\right\}$$

(ii) $X = \sigma Z + \mu$ where $Z \sim N(0, 1)$.

$$\begin{aligned}
M_X(t) &= e^{\mu t} M_Z(\sigma t) \\
&= e^{\mu t} \exp\left\{\frac{(\sigma t)^2}{2}\right\} \\
&= \exp\left\{\frac{(\sigma t)^2}{2} + \mu t\right\}
\end{aligned}$$

THEOREM 2.5.7: Moments from Moment Generating Function

Suppose the random variable X has moment generating function $M(t)$ defined for $t \in (-h, h)$ for some $h > 0$. Then, $M(0) = 1$ and

$$M^{(k)}(0) = \mathbb{E}[X^k]$$

for $k = 1, 2, \ldots$ where

$$M^{(k)}(t) = \frac{d^k}{dt^k}[M(t)]$$

is the kth derivative of $M(t)$.

Proof of: 2.5.7

Omitted from the lecture and hence these notes. See Course Notes for most of the proof.

EXAMPLE 2.5.8

Gamma(α, β) has m.g.f. $M(t) = (1 - \beta t)^{-\alpha}$ for $t < 1/\beta$. What is $\mathbb{E}[X]$ and $\mathbb{V}(X)$?
Solution. For $\mathbb{E}[X]$ we find $M'(t)$.

$$M'(t) = (-\alpha)(1 - \beta t)^{-\alpha-1}(-\beta) = (\alpha\beta)(1 - \beta t)^{-\alpha-1}$$

We know,

$$\mathbb{E}[X] = M'(0) = \alpha\beta$$

For $\mathbb{V}(X)$ we find $M''(t)$.

$$M''(t) = (\alpha\beta)(-\alpha - 1)(-\beta)(1 - \beta t)^{-\alpha-2}$$

Now, $M''(0) = \alpha\beta^2(\alpha+1) = \mathbb{E}[X^2]$. Therefore,

$$\mathbb{V}(X) = \mathbb{E}[X^2] - \mu^2 = \alpha\beta^2(\alpha+1) - (\alpha\beta)^2 = \alpha\beta^2$$

EXAMPLE 2.5.9

The m.g.f. of Poisson(θ) is $M(t) = \exp\{\theta(e^t - 1)\}$. Find $\mathbb{E}[X]$ and $\mathbb{V}(X)$.
Solution.

$$M'(t) = \exp\{\theta(e^t - 1)\}\,\theta e^t$$

Therefore,

$$\mathbb{E}[X] = M'(0) = \theta$$

Now,

$$M''(t) = \exp\{\theta(e^t - 1)\}\,\theta^2 e^{2t} + \theta e^t \exp\{e^\theta(e^t - 1)\}$$

Therefore,

$$M''(0) = \mathbb{E}[X^2] = \theta^2 + \theta$$

So,

$$\mathbb{V}(X) = \mathbb{E}[X^2] - \mu^2 = \theta^2 + \theta - (\theta)^2 = \theta$$

THEOREM 2.5.10: Uniqueness Theorem for Moment Generating Functions

Suppose the random variable X has moment generating function $M_X(t)$ and the random variable Y has moment generating function $M_Y(t)$. $M_X(t) = M_Y(t)$ for all $t \in (-h, h)$ for some $h > 0$ if and only if X and Y have the same distribution; that is,

$$\mathbb{P}(X \leqslant s) = F_X(s) = F_Y(s) = \mathbb{P}(Y \leqslant s)$$

for all $s \in \mathbb{R}$.

EXAMPLE 2.5.11

Suppose X has m.g.f. $M_X(t) = \exp\left\{\dfrac{t^2}{2}\right\}$.

(i) Find m.g.f. of $Y = 2X - 1$
(ii) Find $\mathbb{E}[Y]$ and $\mathbb{V}(Y)$
(iii) What is the distribution of Y.

Solution.

(i) $M_Y(t) = e^{-t}\exp\left\{\dfrac{(2t)^2}{2}\right\} = \exp\{2t^2 - t\}$.

(ii)

$$M_Y'(t) = \exp\{2t^2 - t\}\,(4t - 1)$$

Therefore,

$$\mathbb{E}[Y] = M_Y'(0) = -1$$

Also,

$$M_Y''(t) = \exp\{2t^2 - t\}\,(4t-1)^2 + 4\exp\{2t^2 - t\}$$

and

$$\mathbb{E}[Y^2] = M_Y''(0) = 1 + 4 = 5$$

Therefore,

$$\mathbb{V}(Y) = \mathbb{E}[Y^2] - \mu^2 = 5 - 1 = 4$$

(iii) $M_Y(t) = \exp\{2t^2 - t\}$ is the m.g.f. of $N(-1, 4)$ since if $X \sim N(\mu, \sigma^2)$, then (by previous example)

$$M_X(t) = e^{\mu t} \exp\left\{\frac{\sigma^2 t^2}{2}\right\}$$

LECTURE 5 | 2020-09-20

EXAMPLE 2.5.12: Uniqueness Theorem

Suppose $M_X(t) = (1 - 2t)^{-1}$. What is the distribution of X?
Solution. $X \sim \text{Gamma}(\alpha = 1, \beta = 2)$.

Chapter 3

Multivariate Random Variables

3.1 Joint and Marginal Cumulative Distribution Functions

Purpose: to characterize a joint distribution of two random variables.

DEFINITION 3.1.1: Joint cumulative distribution function

Suppose X and Y are random variables defined on a sample space S. The **joint cumulative distribution function** of X and Y is given by

$$F(x, y) = \mathbb{P}(X \leqslant x, Y \leqslant y)$$

for $(x, y) \in \mathbb{R}^2$.

$\mathbb{P}(X \leqslant x, Y \leqslant y)$: "What is the probability these two events occur simultaneously"

REMARK 3.1.2

Since $\{X \leqslant x\}$ and $\{Y \leqslant y\}$ are both events, $F(x, y)$ is well-defined as we consider $\{X \leqslant x\} \cap \{Y \leqslant y\}$.

REMARK 3.1.3

If we have more than two random variables, say $X_1, X_2, \dots, X_n$ We can similarly define the cumulative distribution function as

$$F(x_1, \dots, x_n) = \mathbb{P}(X_1 \leqslant x_1, \dots, X_n \leqslant x_n)$$

However, in this course we will only focus on two events X and Y.

DEFINITION 3.1.4: Joint cumulative distribution function

(I) F is non-decreasing in x for fixed y

(II) F is non-decreasing in y for fixed x

(III) $\lim\limits_{x \to -\infty} F(x, y) = 0$ and $\lim\limits_{y \to -\infty} F(x, y) = 0$

By looking at

$$\underbrace{\{X \leqslant x\}}_{\substack{\to 0 \\ \text{as } x \to -\infty}} \cap \underbrace{\{Y \leqslant y\}}_{\substack{\to 0 \\ \text{as } y \to -\infty}}$$

(IV)

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

DEFINITION 3.1.5: Marginal distribution function

The **marginal distribution function** of X is given by

$$F_1(x) = \lim_{y\to\infty} F(x,y) = \mathbb{P}(X \leqslant x)$$

for $x \in \mathbb{R}$.
The **marginal distribution function** if Y is given by

$$F_2(y) = \lim_{x\to\infty} F(x,y) = \mathbb{P}(Y \leqslant y)$$

for $y \in \mathbb{R}$.

REMARK 3.1.6

The definition of marginal distribution function tells us that we can know all information about marginal c.d.f. from the joint c.d.f. but the marginal c.d.f. cannot give full information about joint c.d.f.

3.2 Bivariate Discrete Distributions

DEFINITION 3.2.1: Joint discrete random variables

Suppose X and Y are both discrete random variables, then X and Y are **joint discrete random variables** X and Y.

DEFINITION 3.2.2: Joint probability function, Support set

Suppose X and Y are discrete random variables. The **joint probability function** of X and Y is given by

$$f(x,y) = \mathbb{P}(X = x, Y = y)$$

for $(x,y) \in \mathbb{R}^2$.
The set $A = \{(x,y) : f(x,y) > 0\}$ is called the **support set** of (X,Y).

DEFINITION 3.2.3: Properties — Joint Probability Function

(I) $f(x,y) \geqslant 0$ for $(x,y) \in \mathbb{R}^2$

(II) $\displaystyle\sum_{(x,y)\in A} f(x,y) = 1$

(III) For any set $R \subseteq \mathbb{R}^2$

$$P\left[(X,Y) \in R\right] = \sum_{(x,y)\in R} f(x,y)$$

EXAMPLE 3.2.4

Suppose we want to find $\mathbb{P}(X \leqslant Y)$. What is the corresponding set R?
Solution. $R = \{(x,y) : x \leqslant y\}$
Suppose we want to find $\mathbb{P}(X + Y \leqslant 1)$. What is the corresponding set R?
Solution. $R = \{(x,y) : x + y \leqslant 1\}$

DEFINITION 3.2.5: Marginal probability function

Suppose X and Y are discrete random variables with joint probability function $f(x,y)$.
The **marginal probability function** of X is given by

$$f_1(x) = \mathbb{P}(X = x) = \mathbb{P}(X = x, Y < \infty) = \sum_y f(x,y)$$

for $x \in \mathbb{R}$.
The **marginal probability function** of Y is given by

$$f_2(y) = \mathbb{P}(Y = y) = \mathbb{P}(X < \infty, Y = y) = \sum_x f(x,y)$$

for $y \in \mathbb{R}$.

EXAMPLE 3.2.6

Suppose that X and Y are discrete random variables with joint p.f. $f(x,y) = kq^2p^{x+y}$ where

- $x = 0, 1, 2, \ldots$
- $y = 0, 1, 2, \ldots$
- $0 < p < 1$
- $q = 1 - p$

(i) Determine k.
(ii) Find marginal p.f. of X and find marginal p.f. of Y.
(iii) Find $\mathbb{P}(X \leqslant Y)$.

Solution.

(i) $k > 0$ since if $k = 0$ then the summation of the joint p.f. will be 0 (but needs to be 1).

$$\sum_{x=0}^{\infty}\sum_{y=0}^{\infty} f(x,y) = 1$$

Therefore,

$$k\left(\sum_{x=0}^{\infty}\sum_{y=0}^{\infty} p^{x+y}q^2\right) = kq^2\left(\sum_{x=0}^{\infty} p^x\right)\left(\sum_{y=0}^{\infty} p^y\right) = kq^2\left(\frac{1}{1-p}\right)\left(\frac{1}{1-p}\right) = k$$

Thus, $k = 1$.

(ii) Marginal p.f. of X:

$$f_1(x) = \mathbb{P}(X = x) = \sum_{y=0}^{\infty} q^2p^{x+y} = q^2p^x\left(\sum_{y=0}^{\infty} p^y\right) = q^2p^x\left(\frac{1}{1-p}\right) = p^x(1-p)$$

Support of X: $x = 0, 1, 2, \ldots$.
By symmetry,

$$f_2(y) = \mathbb{P}(Y = y) = qp^y$$

Support of Y: $y = 0, 1, 2, \ldots$.

(iii) Find $\mathbb{P}(X \leqslant Y)$.

$$\begin{aligned}
\mathbb{P}(X \leqslant Y) &= \sum_{x=0}^{\infty}\sum_{y=x}^{\infty}(q^2p^{x+y}) \\
&= \sum_{x=0}^{\infty} q^2p^x \sum_{y=x}^{\infty} p^y \\
&= \sum_{x=0}^{\infty} q^2p^x \left(\frac{p^x}{1-p}\right) \\
&= q\sum_{x=0}^{\infty} p^{2x} \\
&= q\left(\frac{1}{1-p^2}\right) \\
&= \frac{1}{1+p}
\end{aligned}$$

REMARK 3.2.7: Interesting Fact

If X and Y are *continuous* random variables and have the same distribution and ***independent***,

$$\mathbb{P}(X \leqslant Y) = \frac{1}{2}$$

3.3 Bivariate Continuous Distributions

DEFINITION 3.3.1: Joint probability density function, Support set

Suppose that $F(x, y)$ is a continuous function and that

$$f(x, y) = \frac{\partial^2}{\partial x \partial y}[F(x, y)]$$

exists and is a continuous function except possibly along a finite number of curves. Suppose also that

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

Then X and Y are said to be continuous random variables with **joint probability density function** f. The set $A = \{(x, y) : f(x, y) > 0\}$ is called the support set of (X, Y).

REMARK 3.3.2

We will arbitrarily define $f(x, y)$ to be equal to 0 when $\dfrac{\partial^2}{\partial x \partial y}[F(x, y)]$ does not exist, although we can define it to be any real number.

DEFINITION 3.3.3: Properties — Joint Probability Density Function

(I) $f(x, y) \geqslant 0$ for all $(x, y) \in \mathbb{R}^2$

(II) For any set $R \subseteq \mathbb{R}^2$:

$$P\left[(X, Y) \in R\right] = \iint\limits_{(x,y)\in R} f(x, y)dx\,dy$$

EXAMPLE 3.3.4

To find $\mathbb{P}(X \leqslant Y)$, the region is $R = \{(x, y) : x \leqslant y\}$. Therefore,

$$\mathbb{P}(X \leqslant y) = \iint\limits_{x\leqslant y} f(x, y)dx\,dy$$

DEFINITION 3.3.5: Marginal probability density function

Suppose X and Y are continuous random variables with p.d.f. $f(x, y)$. The **marginal probability density function** of X is given by

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y)\,dy$$

for $x \in \mathbb{R}$ and the **marginal probability density function** of Y is given by

$$f_2(y) = \int_{-\infty}^{\infty} f(x, y)\,dx$$

for $y \in \mathbb{R}$.

$$P\left[(X, Y) \in \mathbb{R}\right] = \iint\limits_{R} f(x, y)dx\,dy = \int_x \int_y f(x, y)\,dx\,dy$$

Helpful theorem from MATH 237 that some of you may have forgot:

THEOREM 3.3.6: †

y first, then x
Let $R \subset \mathbb{R}^2$ be defined by

$$y_\ell(x) \leqslant y \leqslant y_u(x) \quad \text{and} \quad x_\ell \leqslant x \leqslant x_u$$

where $y_\ell(x)$ and $y_u(x)$ are continuous for $x_\ell \leqslant x \leqslant x_u$. If $f(x,y)$ is continuous on R, then

$$\iint_R f(x,y)dA = \int_{x_\ell}^{x_u} \int_{y_\ell(x)}^{y_u(x)} f(x,y)\,dy\,dx$$

x first, then y
Let $R \subset \mathbb{R}^2$ be defined by

$$x_\ell(y) \leqslant x \leqslant x_u(y) \quad \text{and} \quad y_\ell \leqslant y \leqslant y_u$$

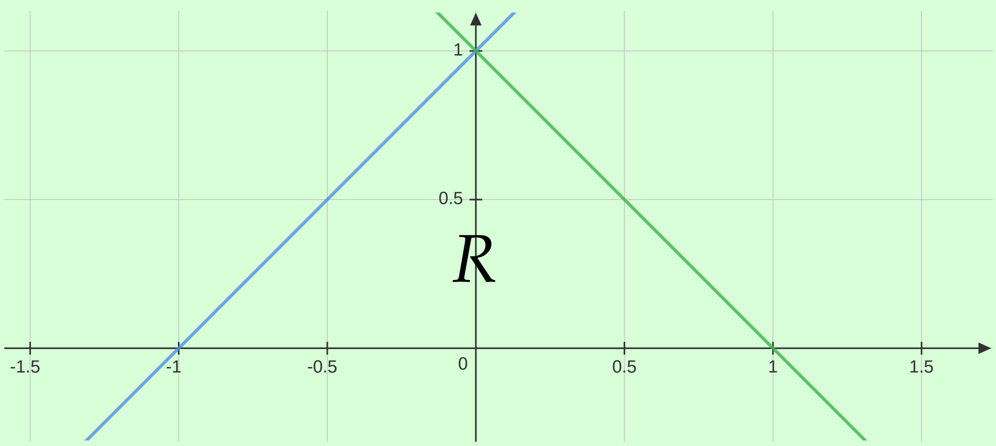
where $x_\ell(y)$ and $x_u(y)$ are continuous for $y_\ell \leqslant y \leqslant y_u$. If $f(x,y)$ is continuous on R, then

$$\iint_R f(x,y)dA = \int_{y_\ell}^{y_u} \int_{x_\ell(y)}^{x_u(y)} f(x,y)\,dx\,dy$$

We use ℓ for "lower" amd u for "upper."

EXAMPLE 3.3.7

Describe the region R above the x-axis.



Solution. R can be described by the set of two inequalities (you can actually verify this in Desmos if you *really* forgot how this works):

$$0 \leqslant y \leqslant 1$$

$$y - 1 \leqslant x \leqslant 1 - y$$

Using the theorem above,

$$\int_0^1 \int_{y-1}^{1-y} f(x,y)\,dx\,dy$$

LECTURE 6 | 2020-09-20

Author's note: Diagrams will be omitted for most of the text, unless the example is not trivial. Students are encouraged to draw the diagrams when following the examples.

EXAMPLE 3.3.8

Let X and Y be continuous random variables with joint p.d.f.

$$f(x,y) = \begin{cases} x+y & 0 \leqslant x \leqslant 1, \quad 0 \leqslant y \leqslant 1 \\ 0 & \text{otherwise} \end{cases}$$

(i) Show $f(x,y)$ is a joint p.d.f.
(ii) Find
  (a) $\mathbb{P}(X \leqslant 1/3, Y \leqslant 1/2)$
  (b) $\mathbb{P}(X \leqslant Y)$
  (c) $\mathbb{P}(X + Y \leqslant 1/2)$
  (d) $\mathbb{P}(XY \leqslant 1/2)$
(iii) Find marginal p.d.f. of X and Y.

Solution.

(i) Note that $f(x,y) \geqslant 0$. We need to show

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

$$\begin{aligned}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dy\,dx &= \int_0^1\int_0^1 (x+y)\,dy\,dx \\ &= \int_0^1 \left[x + \frac{y^2}{2}\right]_0^1 dx \\ &= \int_0^1 \left(x + \frac{1}{2}\right) dx \\ &= \left[\frac{x^2}{2} + \frac{x}{2}\right]_0^1 \\ &= 1\end{aligned}$$

(ii) (a) Take $R = \{(x,y) : 0 \leqslant x \leqslant 1/3, 0 \leqslant y \leqslant 1/2\}$.

$$\begin{aligned}\int_0^{1/3}\int_0^{1/2} (x+y)\,dy\,dx &= \int_0^{1/3} \left[xy + \frac{y^2}{2}\right]_0^{1/2} dx \\ &= \int_0^{1/3} \left(\frac{x}{2} + \frac{1}{8}\right) dx \\ &= \left[\frac{x^2}{4} + \frac{x}{8}\right]_0^{1/3} \\ &= \frac{1}{36} + \frac{1}{24} \\ &= \frac{5}{72}\end{aligned}$$

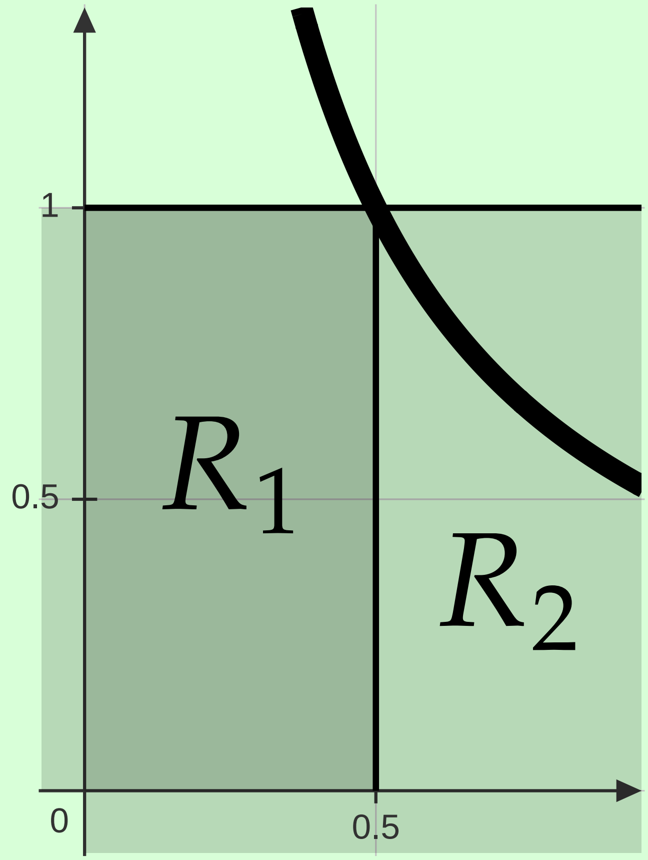
(b) $R = \{(x, y) : 0 \leqslant x \leqslant 1, x \leqslant y \leqslant 1\}$.

$$\begin{aligned}
\int_0^1 \int_x^1 (x+y)\,dy\,dx &= \int_0^1 \left[xy + \frac{y^2}{2}\right]_x^1 dx \\
&= \int_0^1 x + \frac{1}{2} - x^2 - \frac{x^2}{2}\,dx \\
&= \left[\frac{x^2}{2} + \frac{x}{2} - \frac{x^3}{3} - \frac{x^3}{2}\left(\frac{1}{3}\right)\right]_0^1 \\
&= \frac{1}{2}
\end{aligned}$$

(c) $R = \{(x, y) : 0 \leqslant x \leqslant 1/2, 0 \leqslant y \leqslant (1/2) - x\}$

$$\begin{aligned}
\int_0^{1/2} \int_0^{(1/2)-x} (x+y)\,dy\,dx &= \int_0^{1/2} \left[xy + \frac{y^2}{2}\right]_0^{(1/2)-x} dx \\
&= \int_0^{1/2} \left(\frac{x}{2} - x^2 + \frac{1}{8} - \frac{x}{2} + \frac{x^2}{2}\right) dx \\
&= \int_0^{1/2} \frac{1}{8} - \frac{x^2}{2}\,dx \\
&= \left[\frac{x}{8} - \frac{x^3}{2}\left(\frac{1}{3}\right)\right]_0^{1/2} \\
&= \frac{1}{24}
\end{aligned}$$

(d) This example is a bit complicated, so I included a figure.



Note the curve drawn is $xy = 1/2$. R_1 can be described with:

$$0 \leqslant x \leqslant \frac{1}{2}$$

$$0 \leqslant y \leqslant 1$$

R_2 (region below the curve) can be described with:

$$\frac{1}{2} \leqslant x \leqslant 1$$

$$0 \leqslant y \leqslant \left(\frac{1}{2}\right)/x$$

Therefore, we need to evaluate two double integrals.

$$\int_0^{1/2} \int_0^1 (x+y)\,dy\,dx + \int_{1/2}^1 \int_0^{(1/2)/x} (x+y)\,dy\,dx = \frac{3}{4}$$

(iii) The support of X is $[0, 1]$.

$$f_1(x) = 0 \iff x < 0 \text{ or } x > 1$$

Therefore, we focus on $0 \leqslant x \leqslant 1$.

$$f_1(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \int_0^1 (x+y)\,dy = \left[x + \frac{y^2}{2}\right]_0^1 = x + \frac{1}{2}$$

Thus,

$$f_1(x) = \begin{cases} x + \frac{1}{2} & 0 \leqslant x \leqslant 1 \\ 0 & \text{otherwise} \end{cases}$$

$f_2(y)$ is similar by symmetry.

EXAMPLE 3.3.9

Suppose

$$f(x,y) = \begin{cases} ke^{-x-y} & 0 < x < y < \infty \\ 0 & \text{otherwise} \end{cases}$$

is the joint p.d.f. of (X,Y).

(i) Find k.

(ii) Find

  (a) $\mathbb{P}(X \leqslant 1/3, Y \leqslant 1/2)$

  (b) $\mathbb{P}(X \leqslant Y)$

  (c) $\mathbb{P}(X + Y \geqslant 1)$

(iii) Marginal p.d.f. of X and Y.

(iv) Suppose $T = X + Y$, find the p.d.f. of T.

Solution.

(i) We know $f(x,y) \geqslant 0 \iff k \geqslant 0$. Actually, $k > 0$ since if $k = 0$, then $f(x,y) \equiv 0$. We solve k by solving the following:

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

Therefore,

$$\begin{aligned}
&= \int_0^\infty \int_x^\infty ke^{-x-y}\,dy\,dx \\
&= k\int_0^\infty e^{-x}\left[-e^{-y}\right]_x^\infty\,dx \\
&= k\int_0^\infty e^{-x}e^{-x}\,dx \\
&= k\int_0^\infty e^{-2x}\,dx \\
&= k\left[-\frac{1}{2}e^{-2x}\right]_0^\infty \\
&= \frac{k}{2}
\end{aligned}$$

Thus, $k/2 = 1 \implies k = 2$.

(ii) (a) $\mathbb{P}(X \leqslant 1/3, Y \leqslant 1/2)$.

$$R = \{(x,y) : 0 \leqslant x \leqslant 1/3, x \leqslant y \leqslant 1/2\}$$

Therefore,

$$
\begin{aligned}
\mathbb{P}(X \leqslant 1/3, Y \leqslant 1/2) &= \int_0^{1/3} \int_x^{1/2} 2e^{-x-y}\, dy\, dx \\
&= 2\int_0^{1/3} e^{-x}\left[-e^{-y}\right]_x^{1/2}\, dx \\
&= 2\int_0^{1/3} e^{-x}\left(-e^{-1/2} + e^{-x}\right)\, dx \\
&= 2\int_0^{1/3} -e^{-1/2}e^{-x} + e^{-2x}\, dx \\
&= 2\left(-e^{-1/2}\left[-e^{-x}\right]_0^{1/3} + \left[-\frac{1}{2}e^{-2x}\right]_0^{1/3}\right) \\
&= 2\left(-e^{-1/2}\left(-e^{-1/3}+1\right) + \left(-\frac{1}{2}\right)\left(e^{-2/3}-1\right)\right) \\
&= 2\left(1/2 + e^{-5/6} - e^{-1/2} - \frac{1}{2}e^{-2/3}\right) \\
&= 1 - e^{-2/3} + 2\left(e^{-5/6} - e^{-1/2}\right) \\
&\approx 0.1427
\end{aligned}
$$

(b) $\mathbb{P}(X \leqslant Y)$. Note that the region is the same as the support. Therefore,

$$\mathbb{P}(X \leqslant Y) = \iint_{x \leqslant y} f(x, y) dx\, dy = 1$$

(c) $\mathbb{P}(X + Y \geqslant 1)$. Note that this region is a bit complicated, so we will consider $1 - \mathbb{P}(X + Y < 1) = 1 - \mathbb{P}(X + Y \leqslant 1)$. The equal sign does not account for any area (it's continuous, but not required to know in this course).

$$R = \{(x, y) : 0 \leqslant x \leqslant 1/2, x \leqslant y \leqslant 1 - x\}$$

$$
\begin{aligned}
\mathbb{P}(X + Y \leqslant 1) &= \int_0^{1/2} \int_x^{1-x} 2e^{-x}e^{-y}\, dy\, dx \\
&= 2\int_0^{1/2} e^{-x}\left[e^{-y}\right]_x^{1-x}\, dx \\
&= 2\int_0^{1/2} e^{-x}\left(-e^{x-1} + e^{-x}\right)\, dx \\
&= 2\int_0^{1/2} -e^{-1} + e^{-2x}\, dx \\
&= 2\left[-xe^{-1} - \frac{1}{2}e^{-2x}\right]_0^{1/2} \\
&= 2\left(\left(-\frac{1}{2}e^{-1} - \frac{1}{2}e^{-2(1/2)}\right) - \left(0 - \frac{1}{2}\right)\right) \\
&= 2\left(-\frac{1}{2}e^{-1} - \frac{1}{2}e^{-1} + \frac{1}{2}\right) \\
&= 2\left(-e^{-1} + \frac{1}{2}\right) \\
&= 1 - 2e^{-1}
\end{aligned}
$$

Thus, $\mathbb{P}(X+Y \geqslant 1) = 1 - \mathbb{P}(X+Y \leqslant 1) = 1 - (1 - 2e^{-1}) = 2e^{-1}$.

(iii) Marginal p.d.f. of X. The support of X is $(0, \infty)$. We know $x > 0$, so

$$f_1(x) = \int_{-\infty}^{\infty} f(x,y)\, dy = \int_{x}^{\infty} 2e^{-x-y}\, dy = 2e^{-x}\left[-e^{-y}\right]_x^{\infty} = 2e^{-2x}$$

The marginal p.d.f. of Y. The support of Y is $(0, \infty)$. We know $y > 0$, so

$$f_2(y) = \int_0^y 2e^{-x-y}\, dx = 2e^{-y}\left[-e^{-x}\right]_0^y = 2e^{-y}\left(1 - e^{-y}\right) = 2e^{-y} - 2e^{-2y}$$

(iv) Suppose $T = X + Y$, find the p.d.f. of T. We first find the c.d.f. of T, then we take the derivative of T.
Support of T is $(0, \infty)$.
When $t \leqslant 0$, $F_T(t) = \mathbb{P}(T \leqslant t) = 0$, so we only focus on $t > 0$, so $F_T(t) = \mathbb{P}(T \leqslant t)$.

$$R = \{(x,y) : 0 \leqslant x \leqslant t/2, x \leqslant y \leqslant t - x\}$$

Therefore,

$$\begin{aligned}
F_T(t) &= \int_0^{t/2} \int_x^{t-x} 2e^{-x-y}\, dy\, dx \\
&= 2\int_0^{t/2} e^{-x}\left[-e^{-y}\right]_x^{t-x}\, dx \\
&= 2\int_0^{t/2} e^{-x}\left(-e^{x-t} + e^{-x}\right)\, dx \\
&= 2\int_0^{t/2} -e^{-t} + e^{-2x}\, dx \\
&= 2\left[-xe^{-t} - \frac{1}{2}e^{-2x}\right]_0^{t/2} \\
&= 2\left(\left(-\frac{t}{2}e^{-t} - \frac{1}{2}e^{-t}\right) - \left(0 - \frac{1}{2}\right)\right) \\
&= 2\left(-\frac{t}{2}e^{-t} - \frac{1}{2}e^{-t} + \frac{1}{2}\right) \\
&= 1 - e^{-t} - te^{-t}
\end{aligned}$$

So,

$$F_T(t) = \begin{cases} 1 - e^{-t} - te^{-t} & t > 0 \\ 0 & t \leqslant 0 \end{cases}$$

Therefore, by computing $\frac{d}{dt}[F_T(t)]$, the p.d.f. of T is

$$f_T(t) = \begin{cases} te^{-t} & t > 0 \\ 0 & t \leqslant 0 \end{cases}$$

3.4 Independence

DEFINITION 3.4.1: Independent

For any two random variables, we say X and Y are **independent** if and only if

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$$

for any two sets A and B.

THEOREM 3.4.2

For joint discrete random variables or joint continuous random variables. Suppose X and Y are have joint p.f. (discrete) or joint p.d.f. (continuous): $f(x, y)$. Then X and Y are independent if and only if

$$f(x, y) = f_1(x)f_2(y)$$

$$F(x, y) = F_1(x)F_2(y)$$

where $F(x, y)$ is the joint c.d.f. and $F_1(x)$ is the marginal c.d.f. of X and $F_2(y)$ is the marginal c.d.f. of Y.

THEOREM 3.4.3

If X and Y are independent, then $g(X)$ and $h(Y)$ are independent. Where g and h are two real-valued functions.

EXAMPLE 3.4.4

If X and Y are independent, then X^2 and Y^2 are independent. However, if X^2 and Y^2 are independent, then X and Y may not be independent. Can you find an example here? Choose X where

$$\mathbb{P}(X = 1) = \mathbb{P}(X = -1) = \frac{1}{2}$$

EXAMPLE 3.4.5

Consider the joint discrete random variable $f(x, y) = q^2p^{x+y}$, where $x = 0, 1, \ldots$ and $y = 0, 1, 2, \ldots$. Then $f_1(x) = qp^x$ and $f_2(y) = qp^y$. Therefore,

$$f(x, y) = f_1(x)f_2(y)$$

shows that X and Y are independent.

Consider $f(x, y) = \begin{cases} x + y & 0 \leqslant x \leqslant 1 \\ 0 & 0 \leqslant y \leqslant y \end{cases}$ We've shown that

$$f_1(x) = \begin{cases} x + \frac{1}{2} & 0 \leqslant x \leqslant 0 \\ 0 & \text{otherwise} \end{cases}$$

$$f_2(y) = \begin{cases} y + \frac{1}{2} & 0 \leqslant y \leqslant 0 \\ 0 & \text{otherwise} \end{cases}$$

Here we see that $f(x, y) \neq f_1(x)f_2(y)$ Therefore, X and Y are not independent.

THEOREM 3.4.6: Factorization Theorem for Independence

Suppose X and Y are random variables with joint p.(d).f. $f(x, y)$. Suppose also that A is the support set of (X, Y), A_1 is the support set of X, and A_2 is the support set of Y. Then, X and Y are independent random variables if and only if there exist non-negative functions $g(x)$ and $h(y)$ such that

$$f(x, y) = g(x)h(y) \quad \forall (x, y) \in A_1 \times A_2$$

where $A_1 \times A_2 = \{(x, y) : x \in A_1, y \in A_2\}$.

REMARK 3.4.7

Equivalently, we can check that
- The support of A is a rectangle.
- The range of X does not depend on the values of y and the range of Y does not depend on the values of x.

EXAMPLE 3.4.8

$f(x, y) = \dfrac{\theta^{x+y}e^{-2\theta}}{x!y!}$ where $x, y \in \mathbb{Z}_{\geqslant 0}$. Are X and Y independent or not?
Solution. Find the marginal p.f. of X and Y

$$f(x, y) = \frac{\theta^x}{x!}e^{-\theta} \cdot \frac{\theta^y}{y!}e^{-\theta}$$

The range of X does not depend on the value of y. Therefore, X and Y are independent.

$$f_1(x) = \sum_{y=0}^{\infty} f(x, y) = \frac{\theta^x e^{-\theta}}{x!} \quad x \in \mathbb{Z}_{\geqslant 0}$$

$$f_2(y) = \sum_{x=0}^{\infty} f(x, y) = \frac{\theta^y e^{-\theta}}{y!} \quad y \in \mathbb{Z}_{\geqslant 0}$$

If we've shown that X and Y are independent, then we can verify

$$f(x, y) = g(x)h(y)$$

With $f_1(x) = C_1 g(x)$ and $f_2(y) = C_2 h(y)$ where $C_1, C_2 \in \mathbb{R}$ is a constant. We know that $C_1 C_2 = 1$.

EXAMPLE 3.4.9

If X and Y have joint p.d.f.

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

where $-1 \leqslant x \leqslant 1$ and $0 \leqslant y \leqslant 1$. Are X and Y independent? Find $f_1(x)$ and $f_2(y)$.
Solution. $f(x, y) = \underbrace{(1 - x^2)}_{h(x)} \frac{3}{2} \underbrace{y}_{g(y)}$. and $A = \{(x, y) : -1 \leqslant x \leqslant 1, 0 \leqslant y \leqslant 1\}$ is a rectangle. Therefore X and Y are independent. So,

$$f_1(x) = C_1 h(x) = C(1 - x^2) \quad \text{for } -1 \leqslant x \leqslant 1$$

So, let's consider the integral:

$$\int_{-1}^{1} f_1(x)\,dx = C_1 \int_{-1}^{1} (1 - x^2)\,dx = 1 \implies C_1 = \frac{3}{4}$$

Using our previous result, we know that

$$f_2(y) = \frac{1}{C_1} h(y) = \frac{4}{3} \cdot \frac{3}{2} y = 2y \text{for } 0 \leqslant y \leqslant 1$$

EXAMPLE 3.4.10: Uniform Distribution on a Semicircle

$f(x,y) = \dfrac{2}{\pi}$ where $0 \leqslant x \leqslant \sqrt{1-y^2}$ and $-1 \leqslant y \leqslant 1$. The area of the semicircle is given by $\pi/2$. Are X and Y independent and find $f_1(x)$ and $f_2(y)$.

Solution. $f(x,y) = \dfrac{2}{\pi}$ Take $g(x) = 1$ and $h(y) = 2/\pi$. Also, this is not a rectangle, so X and Y are not a rectangle. Similarly, for a particular value of $x = 0$ then $-1 \leqslant y \leqslant 1$, so y depends on x.
The support of X is $[0,1]$.

$$f_1(x) = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{2}{\pi} \, dy = \frac{4}{\pi}\sqrt{1-x^2}$$

The support of Y is $[-1,1]$.

$$f_2(y) = \int_{0}^{\sqrt{1-y^2}} \frac{2}{\pi} \, dx = \frac{2}{\pi}\sqrt{1-y^2}$$

Neither of these marginal distributions are uniform.

LECTURE 8 | 2020-09-27

Independence:

(1) Definition

(2) Check independence: $f(x,y) = f_1(x)f_2(y)$

(3) If X and Y are independent, then $g(X)$ and $h(Y)$ are independent. The converse is not true.

(4) Factorization Theorem

  (i) $f(x,y) = g(x)h(y)$

  (ii) A is a rectangle, or equivalent statements.

  X and Y are independent if and only if (i) and (ii) are satisfied.

3.5 Joint Expectation

This section: extend the definition of expectation from univariate to bivariate cases.

DEFINITION 3.5.1: Joint exepectation

Suppose $h(x,y)$ is a real valued function, then

$$\mathbb{E}[h(X,Y)] = \begin{cases} \displaystyle\sum_x \sum_y h(x,y) f(x,y) & X, Y \text{ are joint discrete} \\ \displaystyle\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x,y) f(x,y) \, dx \, dy & X, Y \text{ are joint continuous} \end{cases}$$

EXAMPLE 3.5.2

$$\mathbb{E}[XY] = \begin{cases} \displaystyle\sum_x \sum_y xyf(x,y) & X, Y \text{ are joint discrete} \\ \displaystyle\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf(x,y)\,dx\,dy & X, Y \text{ are joint continuous} \end{cases}$$

$$\mathbb{E}[X] = \begin{cases} \displaystyle\sum_x \sum_y xf(x,y) & X, Y \text{ are joint discrete} \\ \displaystyle\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xf(x,y)\,dx\,dy & X, Y \text{ are joint continuous} \end{cases}$$

Alternatively,

$$\mathbb{E}[X] = \sum_x xf_1(x) = \sum_x x\left[\sum_y f(x,y)\right]$$

THEOREM 3.5.3: Properties

(1) Linearity: exchange the order of summation and expectation

$$\mathbb{E}[ag(X,Y) + bh(X,Y)] = a\mathbb{E}[g(X,Y)] + b\mathbb{E}[h(X,Y)]$$

$$\mathbb{E}\left[\sum_{i=1}^{n} a_i X_i\right] = \sum_{i=1}^{n} a_i \mathbb{E}[X_i]$$

(2) Independence: if X and Y are independent, then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$. More generally, we have $\mathbb{E}[g(X)h(Y)] = \mathbb{E}[g(X)]\mathbb{E}[h(Y)]$. Furthermore, if $X_1, \ldots, X_n$ are independent, then

$$\mathbb{E}\left[\prod_{i=1}^{n} h(X_i)\right] = \prod_{i=1}^{n} \mathbb{E}[h(X_i)]$$

DEFINITION 3.5.4: Covariance

The **covariance** of X and Y is defined as

$$\mathsf{Cov}(X,Y) = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)] = \mathbb{E}[XY] - \mu_X \mu_Y$$

where $\mu_X = \mathbb{E}[X]$ and $\mu_Y = \mathbb{E}[Y]$.

PROPOSITION 3.5.5

If X and Y are independent, then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y] \implies \mathsf{Cov}(X,Y) = 0$.

THEOREM 3.5.6: Variance Formula

(1) $\mathsf{Cov}(X, X) = \mathbb{V}(X)$
(2) $\mathbb{V}(aX + bY) = a^2\mathbb{V}(X) = b^2\mathbb{V}(Y) + 2ab\mathsf{Cov}(X, Y)$
(3) $\mathsf{Cov}(X + Y, Z) = \mathsf{Cov}(X, Z) + \mathsf{Cov}(Y, Z)$
(4)

$$\mathbb{V}\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 \mathbb{V}(X_i) + \underbrace{\sum_{i \neq j} a_i a_j \mathsf{Cov}(X_i, X_j)}_{\binom{n}{2} \text{ terms}} = \sum_{i=1}^{n} a_i^2 \mathbb{V}(X_i) + 2\underbrace{\sum_{i<j} a_i a_j \mathsf{Cov}(X_i, X_j)}_{\binom{n}{2} \text{ terms}}$$

(5) If $X_1, \dots, X_n$ are independent, then

$$\mathbb{V}\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 \mathbb{V}(X_i)$$

EXAMPLE 3.5.7

Suppose the joint p.f. of X and Y is $f(x, y) = \dfrac{\theta^{x+y} e^{-2\theta}}{x!y!}$, where $x \in \mathbb{Z}_{\geqslant 0}$ and $y \in \mathbb{Z}_{\geqslant 0}$.
Find $\mathbb{V}(2X + 3Y)$.
Solution.

$$f(x, y) = \underbrace{\left(\frac{\theta^x e^{-\theta}}{x!}\right)}_{g(x)} \underbrace{\left(\frac{\theta^y e^{-\theta}}{y!}\right)}_{h(y)}$$

Thus, the range of X does not depend on Y. Therefore, X and Y are independent. In other words, we can write

$$f_1(x) = C\frac{\theta^x e^{-\theta}}{x!} \quad x \in \mathbb{Z}_{\geqslant 0}$$

Since $\displaystyle\sum_{x=0}^{\infty} \frac{\theta^x e^{-\theta}}{x!} = 1$ as it is Poisson we get that $C = 1$. Also,

$$f_2(y) = \frac{\theta^y e^{-\theta}}{y!} \quad y \in \mathbb{Z}_{\geqslant 0}$$

Thus, $\mathbb{V}(X) = \theta$ and $\mathbb{V}(Y) = \theta$. Finally,

$$\mathbb{V}(2X + 3Y) = 4\mathbb{V}(X) + 9\mathbb{V}(Y) = 13\theta$$

EXAMPLE 3.5.8

The joint p.d.f. of X and Y is $f(x, y) = \begin{cases} x + y & x, y \in [0, 1] \\ 0 & \text{otherwise} \end{cases}$.
Find $\mathbb{V}(X + Y)$.
Solution. We know $\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y) + 2\mathsf{Cov}(X, Y)$. Recall that

$$f_1(x) = \begin{cases} x + \frac{1}{2} & x \in [0, 1] \\ 0 & \text{otherwise} \end{cases} \qquad f_2(y) = \begin{cases} y + \frac{1}{2} & y \in [0, 1] \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbb{E}[X] = \int_0^1 x\left(x + \frac{1}{2}\right) dx = \frac{7}{12}$$

$$\mathbb{E}[X^2] = \int_0^1 x^2 \left(x + \frac{1}{2}\right) dx = \frac{5}{12}$$

$$\implies \mathbb{V}(X) = \mathbb{E}[X^2] - \mu_X^2 = \frac{5}{12} - \left(\frac{7}{12}\right)^2 = \frac{11}{144}$$

We know that $\mathbb{E}[Y] = 7/12$, $\mathbb{V}(Y) = 11/144$. Now,

$$\mathbb{E}[XY] = \int_0^1 \int_0^1 xy(x+y)\, dy\, dx = \frac{1}{3}$$

$$\implies \mathsf{Cov}(X,Y) = \mathbb{E}[XY] - \mu_X \mu_Y = \frac{1}{3} - \left(\frac{7}{12}\right)\left(\frac{7}{12}\right) = -\frac{1}{144}$$

Hence,

$$\mathbb{V}(X+Y) = \mathbb{V}(X) + \mathbb{V}(Y) + 2\mathsf{Cov}(X,Y) = \frac{11}{144} + \frac{11}{144} - \frac{2}{144} = \frac{20}{144} = \frac{5}{36}$$

DEFINITION 3.5.9: Correlation coefficient

The **correlation coefficient** of X and Y is defined as

$$\rho(X,Y) = \frac{\mathsf{Cov}(X,Y)}{\sqrt{\mathbb{V}(X)}\sqrt{\mathbb{V}(Y)}}$$

REMARK 3.5.10

$\rho(X,Y)$ can only be used to characterize **linear** association between X and Y. For example, there might be exist some quadratic relationship between X and Y but $\rho(X,Y) \to 0$.

EXAMPLE 3.5.11

$Y = X^2$ and $X \sim N(0,1)$. Note that $\rho(X,Y) = 0$, but obviously there is some relationship between X and Y.

THEOREM 3.5.12

$-1 \leqslant \rho(X,Y) \leqslant 1$.
(1) $\rho(X,Y) = 1 \implies Y = aX + b$ *with* $a > 0$.
(2) $\rho(X,Y) = -1 \implies Y = aX + b$ *with* $a < 0$.

EXAMPLE 3.5.13

Let $f(x,y) = \begin{cases} x+y & x,y \in [0,1] \\ 0 & \text{otherwise} \end{cases}$. Find $\rho(X,Y)$.

Solution. Recall that $\mathbb{V}(X) = \mathbb{V}(Y) = 11/144$ and $\mathsf{Cov}(X,Y) = -1/144$. So,

$$\rho(X,Y) = \frac{-1/144}{\sqrt{11/144}\sqrt{11/144}} = -\frac{1}{11}$$

Lecture 9 | 2020-10-04

Last lecture we talked about joint expectation, more specifically:

- Definition

- Linearity property
- Expectation of product in independent case

$$\mathbb{E}[g(X)h(Y)] = \mathbb{E}[g(X)]\mathbb{E}[h(Y)]$$

  if X is independent of Y.
- Covariance $\text{Cov}(X, X) = \mathbb{V}(X)$. Also, $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y] \iff \text{Cov}(X, Y) = 0$.
- Correlation

3.6 Conditional Distributions

DEFINITION 3.6.1: Conditional probability function

Suppose that X and Y have joint p.f. $f(x, y)$ and marginal p.f. $f_1(x)$ and $f_2(y)$.
The **conditional probability function** of X given $Y = y$ is

$$f_1(x \mid y) = \frac{f(x, y)}{f_2(y)} \quad \text{provided } f_2(y) > 0$$

The **conditional probability function** of Y given $X = x$

$$f_2(y \mid x) = \frac{f(x, y)}{f_1(x)} \quad \text{provided } f_1(x) > 0$$

PROPOSITION 3.6.2: Properties of Conditional Probability Function

$f_1(x \mid y)$ and $f_2(y \mid x)$ are both probability functions; that is,

$$f_1(x \mid y) \geqslant 0 \quad \textit{and} \quad \sum_x f_1(x \mid y) = 1 \implies f_1(x \mid y) \textit{ is a p.f.}$$

$$f_2(y \mid x) \geqslant 0 \quad \textit{and} \quad \sum_y f_2(y \mid x) = 1 \implies f_2(y \mid x) \textit{ is a p.f.}$$

DEFINITION 3.6.3: Conditional probability density function

Suppose that X and Y have joint p.d.f. $f(x, y)$ and marginal p.d.f. $f_1(x)$ and $f_2(y)$.
The **conditional probability density function** of X given $Y = y$ is

$$f_1(x \mid y) = \frac{f(x, y)}{f_2(y)} \quad \text{provided } f_2(y) > 0$$

The **conditional probability density function** of Y given $X = x$

$$f_2(y \mid x) = \frac{f(x, y)}{f_1(x)} \quad \text{provided } f_1(x) > 0$$

PROPOSITION 3.6.4: Properties of Conditional Probability Function

$f_1(x \mid y)$ and $f_2(y \mid x)$ are both probability density functions; that is,

$$f_1(x \mid y) \geqslant 0 \quad \text{and} \quad \int_{-\infty}^{\infty} f_1(x \mid y)\,dx = 1 \implies f_1(x \mid y) \text{ is a p.d.f.}$$

$$f_2(y \mid x) \geqslant 0 \quad \text{and} \quad \int_{-\infty}^{\infty} f_2(y \mid x)\,dy = 1 \implies f_2(y \mid x) \text{ is a p.d.f.}$$

EXAMPLE 3.6.5

Let $f(x,y) = \begin{cases} 8xy & 0 < y < x < 1 \\ 0 & \text{otherwise} \end{cases}$

Find

(i) $f_1(x \mid y)$

(ii) $f_2(y \mid x)$

Solution.

(i) To find $f_1(x \mid y)$, we need to calculate $f_2(y)$.

$$f_2(y) = \int_y^1 8xy\,dx = -4y^3 + 4y \quad 0 < y < 1$$

By definition,

$$f_1(x \mid y) = \frac{f(x,y)}{f_2(y)} = \frac{8xy}{4y - 4y^3} = \frac{2x}{1-y^2} \quad 0 < y < 1$$

Given $0 < y < 1$, the support of X is $y < x < 1$.

(ii) To find $f_2(y \mid x)$, we need to calculate $f_1(x)$.

$$f_1(x) = \int_0^x 8xy\,dy = 4x^3 \quad 0 < x < 1$$

By definition,

$$f_2(y \mid x) = \frac{f(x,y)}{f_1(x)} = \frac{8xy}{4x^3} = \frac{2y}{x^2} \quad 0 < x < 1$$

Given $0 < x < 1$, the support of Y is $0 < y < x$.

EXAMPLE 3.6.6

$f(x,y) = \begin{cases} x+y & x,y \in [0,1] \\ 0 & \text{otherwise} \end{cases}$

Recall that $f_1(x) = x + 1/2$ for $0 \leqslant x \leqslant 1$ and $f_2(y) = y + 1/2$ for $0 \leqslant y \leqslant 1$. Therefore,

$$f_1(x \mid y) = \frac{f(x,y)}{f_2(y)} = \frac{x+y}{y+1/2}$$

Given $0 \leqslant y \leqslant 1$, the support of X is $0 \leqslant x \leqslant 1$.

$$f_2(y \mid x) = \frac{x+y}{x+1/2}$$

Given $0 \leqslant x \leqslant 1$, the support of Y is $0 \leqslant y \leqslant 1$.

EXAMPLE 3.6.7

$f(x,y) = q^2p^{x+y}$ where $x \in \mathbb{Z}_{\geqslant 0}$ and $y \in \mathbb{Z}_{\geqslant 0}$. Note we derived that $f_1(x) = qp^x$ and $f_2(y) = qp^y$. Therefore,

$$f_1(x \mid y) = \frac{f(x,y)}{f_2(y)} = qp^x = f_1(x)$$

$$f_2(y \mid x) = \frac{f(x,y)}{f_1(x)} = qp^y = f_2(y)$$

This is another way to show independence of X and Y.

REMARK 3.6.8: Applications of Conditional Distribution

- Check independence. X and Y are independent if and only if

$$f_1(x \mid y) = f_1(x) \quad \text{and} \quad f_2(y \mid x) = f_2(y)$$

- We can use conditional distribution to find joint distribution.
  Product rule: $f(x,y) = f_1(x \mid y)f_2(y) = f_2(y \mid x)f_1(x)$.

EXAMPLE 3.6.9: Product rule

Suppose $Y \sim \text{Poisson}(\theta)$ and $X \mid Y = y \sim \text{Binomial}(y,p)$. Find the marginal p.f. of X.
Before we get to the solution of this problem, let's consider a physical setup.

- Y: number of students who go to Tim Hortons in one day. Note that $Y \sim \text{Poisson}(\theta)$.
- $X \mid Y = y$: number of students among these y visitors

What is the distribution of X? We guess that $X \sim \text{Poisson}(\theta p)$.
Solution.

$$f_1(x \mid y) = \binom{y}{x} p^x (1-p)^{y-x} \quad x = 0, 1, \dots, y$$

$$f_2(y) = \frac{\theta^y}{y!}e^{-\theta} \quad y = 0, 1, 2 \dots$$

$$\begin{aligned} f(x,y) &= f_1(x \mid y)f_2(y) \\ &= \left(\frac{y!}{x!(y-x)!}p^x(1-p)^{y-x}\right)\frac{\theta^y}{y!}e^{-\theta} \\ &= \left(\frac{\theta^x p^x}{x!}\right)\frac{\theta^{y-x}(1-p)^{y-x}}{(y-x)!}e^{-\theta} \end{aligned}$$

(X,Y) support is $x = 0, 1, \dots, y$ and $y = 0, 1, \dots$. Therefore,

$$\begin{aligned} f_1(x) &= \sum_y f(x,y) \\ &= \sum_{y=x}^{\infty}\left[\left(\frac{(\theta p)^x}{x!}\right)\left(\frac{(\theta(1-p))^{y-x}}{(y-x)!}e^{-\theta}\right)\right] \\ &= \frac{e^{-\theta}(\theta p)^x}{x!}\sum_{h=0}^{\infty}\frac{[\theta(1-p)]^h}{h!} && h = y - x \\ &= \frac{e^{-\theta}(\theta p)^x}{x!}e^{\theta(1-p)} \\ &= \frac{(\theta p)^x}{x!}e^{-\theta p} \end{aligned}$$

Therefore, $x = 0, 1, \dots$ and so $X \sim \text{Poisson}(\theta p)$.

EXAMPLE 3.6.10

Suppose Y has p.d.f. $f_2(y) = \dfrac{y^{\alpha-1}}{\Gamma(\alpha)}e^{-y}$ for $y > 0$; that is, $Y \sim \text{Gamma}(\alpha, \beta = 1)$. The conditional p.d.f. of X given $Y = y$ is

$$f_1(x \mid y) = ye^{-xy} \quad \text{for } x > 0, y > 0$$

Find the marginal p.d.f. of X.
Solution. Firstly, find the joint p.d.f. of (X, Y) is

$$f(x, y) = f_1(x \mid y)f_2(y) = ye^{-xy}\frac{y^{\alpha-1}}{\Gamma(\alpha)}e^{-y} = \frac{y^{\alpha}}{\Gamma(\alpha)}e^{-(x+1)y}$$

The support of X is $(0, \infty)$. Recall that the gamma function is $\Gamma(\alpha) = \displaystyle\int_0^{\infty} x^{\alpha-1}e^{-x}\,dx$.
The marginal p.d.f. of X is

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y)\,dy = \int_0^{\infty} \frac{y^{\alpha}e^{-(x+1)y}}{\Gamma(\alpha)}\,dy$$

Let $t = (x + 1)y$, therefore $y = t/(x + 1)$ and $dy = dt/(x + 1)$.

$$\int_0^{\infty} \frac{t^{\alpha}}{(x+1)^{\alpha}\Gamma(\alpha)}e^{-t}\frac{1}{x+1}\,dt = \frac{1}{(x+1)^{\alpha+1}\Gamma(\alpha)}\int_0^{\infty} t^{\alpha}e^{-t}\,dt = \frac{1}{(x+1)^{\alpha+1}\Gamma(\alpha)}\Gamma(\alpha+1)$$

By 2.3.7, we know that $\Gamma(\alpha + 1) = (\alpha)\Gamma(\alpha)$. Therefore,

$$\frac{\Gamma(\alpha+1)}{(x+1)^{\alpha+1}\Gamma(\alpha)} = \frac{(\alpha)\Gamma(\alpha)}{(x+1)^{\alpha+1}\Gamma(\alpha)} = \frac{\alpha}{(x+1)^{\alpha+1}}$$

That is, $f_1(x) = \dfrac{\alpha}{(x+1)^{\alpha+1}}$ and the support of X is positive.

LECTURE 10 | 2020-10-04

Recall that $f_2(y \mid x) = \begin{cases} \text{p.f.} & X \text{ and } Y \text{ are joint discrete} \\ \text{p.f.} & X \text{ and } Y \text{ are joint continuous} \end{cases}$

Therefore, we can define expectation of expectation based on $f_2(y \mid x)$.

3.7 Conditional Expectation

DEFINITION 3.7.1: Conditional expectation

The **conditional expectation** of $g(Y)$ given $X = x$ is defined as

$$\mathbb{E}[g(Y) \mid X = x] = \begin{cases} \displaystyle\sum_y g(y)f_2(y \mid x) & Y \text{ is discrete} \\ \displaystyle\int_{-\infty}^{\infty} g(y)f_2(y \mid x)\,dy & Y \text{ is continuous} \end{cases}$$

REMARK 3.7.2

- Supplementary notes: $\mathbb{E}[g(Y) \mid X = x]$ is denoted by $\mathbb{E}[g(Y) \mid x]$.

We're interested in

1. $\mathbb{E}[Y \mid X = x]$, $g(Y) = Y$.
2. $\mathbb{V}(Y \mid X = x) = \mathbb{E}[Y^2 \mid X = x] - (\mathbb{E}[Y \mid X = x])^2$.
3. $\mathbb{E}[e^{tY} \mid X = x]$, that is, $g(Y) = e^{tY}$.

THEOREM 3.7.3: Independence

If X and Y are independent, then

$$\mathbb{E}[g(Y) \mid X = x] = \mathbb{E}[g(Y)] \quad \textit{and} \quad \mathbb{E}[h(X) \mid Y = y] = \mathbb{E}[h(X)]$$

In other words, the conditional expression becomes an unconditional one.

EXAMPLE 3.7.4

If X and Y are independent, then

$$\mathbb{E}[Y \mid X = x] = \mathbb{E}[Y] \quad \text{and} \quad \mathbb{V}(Y \mid X = x) = \mathbb{V}(Y)$$

Also, $\mathbb{V}(Y \mid X = x) = \mathbb{E}[Y^2 \mid X = x] - (\mathbb{E}[Y \mid X = x])^2 = \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2$

THEOREM 3.7.5: Substitution Rule

Let h be a bivariate function, then

$$\mathbb{E}[h(X, Y) \mid X = x] = \mathbb{E}[h(x, Y) \mid X = x]$$

EXAMPLE 3.7.6

- $\mathbb{E}[X + Y \mid X = x] = \mathbb{E}[x + Y \mid X = x] = x + \mathbb{E}[Y \mid X = x]$
- $\mathbb{E}[XY \mid X = x] = \mathbb{E}[xY \mid X = x] = x\mathbb{E}[Y \mid X = x]$

THEOREM 3.7.7

The conditional expectation has all properties of expectation like linearity.

EXAMPLE 3.7.8

$f(x, y) = \begin{cases} 8xy & 0 < y < x < 1 \\ 0 & \text{otherwise} \end{cases}$ We've found that $f_1(x \mid y) = (2x)/(1 - y)^2$ for $0 < y < 1$ and $y < x < 1$.

$$\mathbb{E}[X \mid Y = y] = \int_{-\infty}^{\infty} x f_1(x \mid y)\, dx = \int_y^1 (x) \frac{2x}{1 - y^2}\, dx = \left(\frac{2}{3}\right) \frac{1 - y^3}{1 - y^2} = \left(\frac{2}{3}\right) \frac{y^2 + y + 1}{y + 1}$$

$$\mathbb{E}[X^2 \mid Y = y] = \int_y^1 (x^2) \frac{2x}{1 - y^2}\, dy = \left(\frac{2}{4}\right) \frac{1 - y^4}{1 - y^2} = \left(\frac{1}{2}\right)(y^2 + 1) \quad 0 < y < 1$$

$$\mathbb{V}(X \mid Y = y) = \left(\frac{1}{2}\right)(1 + y^2) - \left(\frac{4}{9}\right) \frac{(1 + y + y^2)^2}{(1 + y)^2} \quad 0 < y < 1$$

EXAMPLE 3.7.9

Suppose $Y \sim \text{Poisson}(\theta)$ and $X \mid Y = y \sim \text{Binomial}(y, p)$. Then,

$$\mathbb{E}[X \mid Y = y] = yp \quad \text{and} \quad \mathbb{V}(X \mid Y = y) = yp(1 - p)$$

REMARK 3.7.10

Note that $\mathbb{E}[g(Y) \mid X] \neq \mathbb{E}[g(Y) \mid X = x]$. $\mathbb{E}[g(Y) \mid X]$ is a random variable because it's a function of X, denoted by $h(X)$. It's value is given by $h(x) = \mathbb{E}[g(Y) \mid X = x]$ for $X = x$.
How to get it? Two steps.

- Step 1: Find $\mathbb{E}[g(Y) \mid X = x] = h(x)$
- Step 2: Replace x by X to get the random variable $\mathbb{E}[g(Y) \mid X] = h(X)$.

EXAMPLE 3.7.11

Suppose $Y \sim \text{Poisson}(\theta)$ and $X \mid Y = y \sim \text{Binomial}(y, p)$. Then,

$$\mathbb{E}[X \mid Y = y] = yp \implies \mathbb{E}[X \mid Y] = Yp$$

These concepts lead to the Double Expectation Theorem.

THEOREM 3.7.12: Double Expectation

$\mathbb{E}[g(Y)] = \mathbb{E}[\mathbb{E}[g(Y) \mid X]]$. *In particular,* $\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y \mid X]]$.

EXAMPLE 3.7.13

Suppose $Y \sim \text{Poisson}(\theta)$ and $X \mid Y = y \sim \text{Binomial}(y, p)$. Find $\mathbb{E}[X]$.
Solution. By Theorem 3.7.12 we have

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X \mid Y]] = \mathbb{E}[Yp] = p\mathbb{E}[Y] = p\theta$$

Recall that we've shown that $X \sim \text{Poisson}(p\theta) \implies \mathbb{E}[X] = p\theta$.

Lecture 11 | 2020-10-18

Last lecture:

- Conditional expectation: $\mathbb{E}[g(Y) \mid X = x]$
- Properties:
  (1) Independence
  (2) Substitution rule
  (3) Linearity
- Definition of $\mathbb{E}[g(Y) \mid X]$
  - a random variable and function of X, denoted as $h(X)$
  - two step method to find $h(X)$
- Double expectation theorem: $\mathbb{E}[g(Y)] = \mathbb{E}[\mathbb{E}[g(Y) \mid X]]$

THEOREM 3.7.14

$$\mathbb{V}(Y) = \mathbb{E}[\mathbb{V}(Y \mid X)] + \mathbb{V}(\mathbb{E}[Y \mid X])$$

REMARK 3.7.15

$\mathbb{V}(Y \mid X)$ is a random variable and function of X.

Two steps:

1. $\mathbb{V}(Y \mid X = x) = \mathbb{E}[Y^2 \mid X = x] - (\mathbb{E}[Y \mid X = x])^2$
2. replace x with X to get $\mathbb{V}(Y \mid X)$

EXAMPLE 3.7.16

$Y \sim$ Poisson(θ), $X \mid Y = y \sim$ Binomial(y, p). Find $\mathbb{V}(X)$.
Solution. We know that $X \sim$ Poisson$(p\theta)$, then $\mathbb{V}(X) = p\theta$. But we can alternatively use the Double Expectation Theorem.

$$\mathbb{V}(X) = \mathbb{E}[\mathbb{V}(X \mid Y)] + \mathbb{V}(\mathbb{E}[X \mid Y])$$

To find $\mathbb{V}(X \mid Y)$,

$$\mathbb{V}(X \mid Y = y) = yp(1-p) \implies \mathbb{V}(X \mid Y) = Yp(1-p)$$

To find $\mathbb{E}[X \mid Y]$,

$$\mathbb{E}[X \mid Y = y] = yp \implies \mathbb{E}[X \mid Y] = Yp$$

Therefore,

$$\mathbb{V}(X) = \mathbb{E}[Yp(1-p)] + \mathbb{V}(pY) = p(1-p)\mathbb{E}[Y] + p^2\mathbb{V}(Y) = p(1-p)\theta + p^2\theta = p\theta$$

EXAMPLE 3.7.17

Suppose $X \sim$ Uniform$[0, 1]$ and $Y \mid X = x \sim$ Binomial$(10, x)$. Find $\mathbb{E}[Y]$ and $\mathbb{V}(Y)$.

$$\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y \mid X]]$$

Two steps to find $\mathbb{E}[Y \mid X]$.

$$\mathbb{E}[Y \mid X = x] = 10x \implies \mathbb{E}[Y \mid X] = 10X$$

$$\mathbb{E}[Y] = \mathbb{E}[10X] = 10\mathbb{E}[X] = 10\left(\frac{1+0}{2}\right) = 5$$

$$\mathbb{V}(Y) = \mathbb{E}[\mathbb{V}(Y \mid X)] = \mathbb{V}(\mathbb{E}[Y \mid X])$$

Two steps to find $\mathbb{V}(Y \mid X)$.

$$\mathbb{V}(Y \mid X = x) = 10x(1-p) \implies \mathbb{V}(Y \mid X) = 10X(1-X)$$

$$\begin{aligned}
\mathbb{V}(Y) &= \mathbb{E}[10X(1-X)] + \mathbb{V}(10X) \\
&= 10\mathbb{E}[X] - 10\mathbb{E}[X^2] + 100\mathbb{V}(X) \\
&= 10\left(\frac{1+0}{2}\right) - 10\left[\mathbb{V}(X) + (\mathbb{E}[X])^2\right] + 100\mathbb{V}(X) \\
&= 5 - 10\left[\frac{(0-1)^2}{12} + \left(\frac{1+0}{2}\right)^2\right] + 100\left[\frac{(0-1)^2}{12}\right] \\
&= 5 - 10\left(\frac{1}{12} + \frac{1}{4}\right) + 100\left(\frac{1}{12}\right) \\
&= 5 - 10\left(\frac{1}{3}\right) + \frac{100}{12} \\
&= 10
\end{aligned}$$

EXAMPLE 3.7.18

Suppose $Y \sim \text{Poisson}(\theta)$ and $X \mid Y = y \sim \text{Binomial}(y, p)$. Find the m.g.f. of X using the Double Expectation Theorem. [We *could* use the formula sheet to find $M_X(t)$ since we already know $X \sim \text{Poisson}(p\theta)$]

Solution. By definition, the m.g.f. of X is

$$M_X(t) = \mathbb{E}[e^{tX}] = \mathbb{E}[\mathbb{E}[e^{tX} \mid Y]]$$

Given $Y = y$,

$$\begin{aligned}
\mathbb{E}[e^{tX} \mid Y = y] &= \sum_{x=0}^{y} e^{tx}\binom{y}{x}p^x(1-p)^{y-x} \\
&= \sum_{x=0}^{y}\binom{y}{x}(pe^t)^x(1-p)^{y-x} \\
&= (1-p+pe^t)^y
\end{aligned}$$

Therefore, $\mathbb{E}[e^{tX} \mid Y] = (1-p+pe^t)^Y$. Therefore,

$$\begin{aligned}
M_X(t) &= \mathbb{E}\left[(1-p+pe^t)^Y\right] \\
&= \sum_{y=0}^{\infty}(1-p+pe^t)^y\frac{\theta^y e^{-\theta}}{y!} \\
&= e^{-\theta}\sum_{y=0}^{\infty}\frac{\left[\theta(1-p+pe^t)\right]^y}{y!} \\
&= e^{-\theta}\,\text{exp}\left\{\theta(1-p+pe^t)\right\} \\
&= \text{exp}\left\{\theta p(e^t-1)\right\}
\end{aligned}$$

Actually, this is the m.g.f. of Poisson(θp).

3.8 Joint Moment Generating Functions

DEFINITION 3.8.1: Joint moment generating function

If X and Y are two random variables, then

$$M(t_1, t_2) = \mathbb{E}[e^{t_1 X + t_2 Y}]$$

is called the **joint moment generating function** of X and Y if $M(t_1, t_2)$ exists for $|t_1| < h_1$ and $|t_2| < h_2$ for some $h_1 > 0$, $h_2 > 0$.

REMARK 3.8.2

In general, suppose $X_1, \dots, X_n$ are random variables, then

$$M(t_1, \dots, t_n) = \mathbb{E}\left[\exp\left\{\sum_{i=1}^{n} t_i X_i\right\}\right]$$

is the **joint moment generating function** if it exists for $|t_i| < h_i$ for $i = 1, \dots, n$.

REMARK 3.8.3: Applications of Joint Moment Generating Functions

(1) From joint m.g.f. to marginal m.g.f. Given $M(t_1, t_2)$ for $|t_1| < h_1$ and $|t_2| < h_2$,

$$M_X(t_1) = M(t_1, t_2 = 0) = \mathbb{E}[e^{t_1 X}] \quad |t_1| < h_1$$

$$M_Y(t_2) = M(0, t_2) = \mathbb{E}[e^{t_2 Y}] \quad |t_2| < h_2$$

(2) Independence Property. X and Y are independent if and only if

$$M(t_1, t_2) = M_X(t_1) M_Y(t_2)$$

More generally, if $X_1, \dots, X_n$ are independent, then

$$M(t_1, \dots, t_n) = \prod_{i=1}^{n} M_{X_i}(t_i)$$

EXAMPLE 3.8.4

Suppose $f(x, y) = e^{-y}$ for $0 < x < y$ is the joint p.d.f. of (X, Y). Find the joint m.g.f. of X and Y. Are they independent? Find the marginal p.d.f. of X and Y.

Solution.

$$\begin{aligned}
M(t_1, t_2) &= \mathbb{E}[e^{t_1 X + t_2 Y}] \\
&= \int_0^\infty \left[\int_0^y e^{t_1 x + t_2 y} e^{-y} \, dx \right] dy \\
&= \int_0^\infty e^{(t_2 - 1)y} \left[\frac{1}{t_1} e^{t_1 x} \right]_0^y dy \\
&= \frac{1}{t_1} \int_0^\infty e^{(t_2-1)y} (e^{t_1 y} - 1) \, dy \\
&= \frac{1}{t_1} \int_0^\infty e^{(t_1 + t_2 - 1)y} - e^{(t_2 - 1)y} \, dy \\
&= \frac{1}{t_1} \left(\frac{1}{1 - t_1 - t_2} - \frac{1}{1 - t_2} \right) \\
&= \frac{1}{(1 - t_1 - t_2)(1 - t_2)}
\end{aligned}$$

with $t_2 - 1 < 0$ and $t_1 + t_2 - 1 < 0$. Therefore, $t_2 < 1$ and $t_1 + t_2 < 1$.

$$M_X(t_1) = M(t_1, t_2 = 0) = \frac{1}{1 - t_1}$$

which is the m.g.f. of Exponential(1).

$$M_Y(t_2) = M(t_1 = 0, t_2) = \frac{1}{(1 - t_2)^2}$$

which is the m.g.f. of Gamma$(\alpha = 2, \beta = 1)$. Note that the joint support is a triangle (not a rectangle), so obviously $M(t_1, t_2) \neq M_X(t_1) M_Y(t_2)$. Thus, X and Y are not independent.

EXAMPLE 3.8.5: Additivity of Poisson Random Variables

Suppose $X \sim$ Poisson(θ_1) and $Y \sim$ Poisson(θ_2) with X and Y independent. Prove that $X + Y \sim$ Poisson$(\theta_1 + \theta_2)$.

Solution. We can try to find the p.d.f. of $X + Y$ (direct method). Alternatively, find $M_{X+Y}(t)$.

$$\begin{aligned}
M_{X+Y}(t) &= \mathbb{E}[e^{tX + tY}] \\
&= \mathbb{E}[e^{tX} e^{tY}] && X \text{ and } Y \text{ independent} \\
&= \mathbb{E}[e^{tX}] \mathbb{E}[e^{tY}] \\
&= \exp\{\theta_1 (e^t - 1)\} \exp\{\theta_2 (e^t - 1)\} \\
&= \exp\{(\theta_1 + \theta_2)(e^t - 1)\}
\end{aligned}$$

which is the m.g.f. of Poisson$(\theta_1 + \theta_2)$.