STAT 330 - Mathematical Statistics

Cameron Roopnarine

Last updated: September 21, 2020

Contents

Chapter 2

Univariate Random Variable

LECTURE 1 | 2020-09-09

Review of:

- Probability
- Random variables (discrete and continuous)
- Expectation and variance
- Moment generating function

2.1 Probability

DEFINITION 2.1.1: Probability model

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

DEFINITION 2.1.2: Sample space

A **sample space** S is a set of all the distinct outcomes for a random experiment, with the property that in a single trial, one and only one of these outcomes occurs.

EXAMPLE 2.1.3

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 2.1.4: † Sigma algebra

A collection of subsets of a set S is called **sigma algebra**, denoted by β, if it satisfies the following properties:

(I) $\varnothing \in \beta$

(II) If $A \in \beta$, then $\bar{A} \in \beta$

(III) If $A_1, A_2, \ldots \in \beta$, then $\bigcup_{i=1}^{\infty} A_i \in \beta$

DEFINITION 2.1.5: Probability set function

Let β be a sigma algebra associated with the sample space S. A **probability set function** is a function P with domain β that satisfies the following axioms:

(I) $P(A) \geqslant 0$ for all $A \in \beta$

(II) $P(S) = 1$

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

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

EXAMPLE 2.1.6

Toss a coin twice, given one event A,

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

since $|S| = 4$. P satisfies the three properties, therefore P is a probability function.

PROPOSITION 2.1.7: Additional Properties of the Probability Set Function

Let β be a sigma algebra associated with the sample space S and let P be a probability set function with domain β. If $A, B \in \beta$, then:

(1) $P(\varnothing) = 0$

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

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

(4) If $A \subset B$, then $P(A) \leqslant P(B)$

Note for (4), $A \subset B$ means $a \in A$ implies $a \in B$.

Proof of: 2.1.7

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

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

and by (II) we have

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

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 $P(\varnothing) = 0$ as required.

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

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

and since $P(\varnothing) = 0$ by the result of (1) it follows that

$$P(A \cup B) = P(A) + P(B)$$

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

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

as required.
Proof of (4): Since

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

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

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

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

EXERCISE 2.1.8

Let β be a sigma algebra associated with the sample space S and let P be a probability set function with domain β. If $A, B \in \beta$ then prove the following:

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

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

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

as required.

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

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

Therefore,

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

as required.

DEFINITION 2.1.9: Conditional probability

Let β be a sigma algebra associated with the sample space S and suppose $A, B \in \beta$ with $P(B) > 0$. Then the **conditional probability** of A given that B has occurred is

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

DEFINITION 2.1.10: Independent events

Let β be a sigma algebra associated with the sample space S and suppose $A, B \in \beta$. A and B are **independent events** if

$$P(A \cap B) = P(A)P(B)$$

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

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

EXAMPLE 2.1.11

Toss a coin twice.

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

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

also

$$P(B) = \frac{2}{4}$$

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

therefore A and B are independent.

2.2 Random Variables

DEFINITION 2.2.1: Random variable

A **random variable** 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 2.2.2

Toss a coin twice. X: # of H in two tosses
Possible values of X: $0, 1, 2$. Given $x \in \mathbb{R}$.

$$\{X \leqslant x\}$$

- $x < 0$ then $\{X \leqslant x\} = \varnothing$
- $0 \leqslant x < 1$ then

then

$$\{X \leqslant x\} = \{X = 0\} = \{(T, T)\} \subseteq S$$

therefore X is a random variable.

DEFINITION 2.2.3: Cumulative distribution function

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

$$F(x) = P(X \leqslant x)$$

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

DEFINITION 2.2.4: Properties of the 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) $\lim_{x\to\infty} F(x) = 1$ and $\lim_{x\to-\infty} F(x) = 0$.
By looking at:
- $x \to \infty$: $\{X \leqslant x\} \to S$
- $x \to -\infty$: $\{X \leqslant x\} \to \varnothing$

(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 $a < b$

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

(5) For all b

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

LECTURE 2 | 2020-09-09

2.3 Discrete Random Variables

DEFINITION 2.3.1: 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 2.3.2

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

DEFINITION 2.3.3: Probability function

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

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

DEFINITION 2.3.4: Support set

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

PROPOSITION 2.3.5: Properties of the Probability Function

(1) $f(x) \geqslant 0$ *for* $x \in \mathbb{R}$
(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,

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

  Therefore,

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

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

- **Binomial.** $X \sim \text{Binomial}(n, p)$. 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)

  $$P(X = 1) = p$$

  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.

  $$P(X_i = 1) = p$$

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

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

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

  $$f(x) = P(X = x) = (1 - p)^x p$$

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

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

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

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

- **Poisson.** $X \sim \text{Poisson}(\mu)$ where $X = 0, 1, \ldots$.

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

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

EXERCISE 2.3.6

Verify all that all the probability models above are indeed probability functions using Proposition 2.3.5.

Solution. TODO

2.4 Continuous Random Variables

DEFINITION 2.4.1: 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 2.4.2: 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: $f(x) \neq P(X = x)$

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

since

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

DEFINITION 2.4.3: Properties of the 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) $f(x) = F'(x) = \displaystyle\lim_{h \to 0} \frac{F(x+h) - F(x)}{h}$

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

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

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

EXAMPLE 2.4.4

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 2.4.5

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 $P(-2 < X < 3)$.

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) $F(x) = P(X \leqslant x)$.

Case 1: $x < 1$.

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

Case 2: $x \geqslant 1$.

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

(iii) $P(-2 < X < 3)$. 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 | 2020-09-14

We first introduce a function that will be used.

DEFINITION 2.4.6: Gamma function

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

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

PROPOSITION 2.4.7: 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) $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$

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: † 2.4.7

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 $\Gamma\left(\frac{1}{2}\right)$ explicitly now.

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

EXAMPLE 2.4.8

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,

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

EXAMPLE 2.4.9

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, $X \sim \text{Weibull}(\theta, \beta)$. 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 $y = \left(\frac{x}{\theta}\right)^\beta \implies x = \theta y^{1/\beta}$ and $dx = \frac{\theta}{\beta}y^{(1/\beta)-1}\,dy$. Therefore,

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

EXAMPLE 2.4.10: Normal

The p.d.f. is given by

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

$x \in \mathbb{R}$, $-\infty < \mu < \infty$, $\sigma^2 > 0$. 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 $dx = \frac{\sqrt{2}}{2} y^{-1/2}\, dy$. Therefore,

$$= \frac{2}{\sqrt{2\pi}} \int_{0}^{\infty} e^{-y} \frac{\sqrt{2}}{2} y^{-1/2}\, dy = \frac{1}{\sqrt{\pi}} \int_{0}^{\infty} y^{1/2-1} e^{-y}\, dy = \left(\frac{1}{\sqrt{\pi}}\right) \Gamma\left(\frac{1}{2}\right) = 1$$

<u>Case 2</u>: 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 $z = \frac{x-\mu}{\sigma} \implies x = \mu + \sigma z$ 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.

2.7 Expectation

DEFINITION 2.7.1: Expectation of discrete random variable

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

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

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

DEFINITION 2.7.2: Expectation of continuous random variable

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

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

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

EXAMPLE 2.7.3: 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 $\mathbf{E}[X]$.
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, $\mathbf{E}[X]$ does not exist!

EXAMPLE 2.7.4: 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 $\mathbf{E}[X]$.
Solution.

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

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

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

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

EXAMPLE 2.7.5: Bernoulli and Binomial Random Variable

Suppose $X \sim$ Bernoulli(p).

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

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

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

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,

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

EXAMPLE 2.7.6

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 $\mathbf{E}[X]$ and for what values of θ, does $\mathbf{E}[X]$ 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 $\mathbf{E}[X]$ exists. Also,

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

DEFINITION 2.7.7: 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

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

provided the sum converges absolutely; that is, provided

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

DEFINITION 2.7.8: 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

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

provided the integral converges absolutely; that is, provided

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

THEOREM 2.7.9: 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,

$$\mathbf{E}[aX + b] = a\mathbf{E}[x] + b$$

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

Proof of: 2.7.9

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

DEFINITION 2.7.10: Variance

The variance of a random variable is defined as

$$\sigma^2 = \mathbf{Var}[X] = \mathbf{E}\left[(X-\mu)^2\right]$$

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

DEFINITION 2.7.11: Special Expectations

(I) The mean of a random variable

$$\mathbf{E}[X] = \mu$$

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

$$\mathbf{E}\left[X^k\right]$$

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

$$\mathbf{E}\left[(X-\mu)^k\right]$$

(IV) † The kth factorial of a random variable

$$\mathbf{E}\left[X^{(k)}\right] = \mathbf{E}[X(X-1)\cdots(X-k+1)]$$

(V) The variance of a random variable

$$\mathbf{Var}[X] = \mathbf{E}\left[(X-\mu)^2\right] = \sigma^2$$

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

THEOREM 2.7.12: Properties of Variance

If X is a random variable, then

$$\mathbf{Var}[X] = \mathbf{E}\left[X^2\right] - \mu^2$$

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

Proof of: 2.7.12

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

EXAMPLE 2.7.13

Suppose $X \sim \text{Poisson}(\theta)$, the p.f. is defined as $f(x) = \frac{\theta^x}{x!}e^{-\theta}$ for $x = 0, 1, 2, \ldots$. Find $\mathbf{E}[X]$ and $\mathbf{Var}[X]$.
Solution.

$$\sum_{x=0}^{\infty} |x| f(x) < \infty$$

Therefore,

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

Let $y = x - 1$, then

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

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

$$\mathbf{Var}\left[X\right] = \mathbf{E}\left[X^2\right] - \mu^2$$

Let's find $\mathbf{E}\left[X^2\right]$:

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

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 $y = x - 2$:

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

Therefore,

$$\mathbf{Var}\left[X\right] = \mathbf{E}\left[X^2\right] - \mu^2 = (\theta^2 + \theta) - \theta^2 = \theta$$

Lecture 4 | 2020-09-14

EXAMPLE 2.7.14

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

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

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,

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

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 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.

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

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,

- $\mathbf{E}\left[X\right] = \dfrac{\beta\Gamma(\alpha+1)}{\Gamma(\alpha)} = \alpha\beta$
- $\mathbf{E}\left[X^2\right] = \dfrac{\beta^2\Gamma(\alpha+2)}{\Gamma(\alpha)} = \dfrac{\beta^2(\alpha+1)\Gamma(\alpha+1)}{\Gamma(\alpha)} = \alpha(\alpha+1)\beta^2$
- $\mathbf{Var}\left[X\right] = \mathbf{E}\left[X^2\right] - \mu^2 = \alpha(\alpha+1)\beta^2 - \alpha^2\beta^2 = \alpha\beta^2$

2.10 Moment Generating Functions

DEFINITION 2.10.1: Moment generating function

Suppose X is a random variable, then

$$M(t) = \mathbf{E}\left[e^{tX}\right]$$

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

REMARK 2.10.2

If we are able to find some $h > 0$ such that for any $t \in (-h, h)$,

$$\mathbf{E}\left[e^{tX}\right] < \infty$$

say $M(t)$ is the m.g.f. of X.

EXAMPLE 2.10.3

Suppose $X \sim$ Gamma(α, β). Find $M(t)$. Recall the p.d.f. is

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

Solution.

$$
\begin{aligned}
M(t) &= \mathbf{E}\left[e^{tX}\right] \\
&= \int_{-\infty}^{\infty} e^{tx} f(x)\, dx \\
&= \int_{0}^{\infty} e^{tx} \cdot \frac{x^{\alpha-1} e^{-x/\beta}}{\Gamma(\alpha)\beta^{\alpha}}\, dx \\
&= \int_{0}^{\infty} \frac{x^{\alpha-1} e^{-x\left(\frac{1}{\beta}-t\right)}}{\Gamma(\alpha)\beta^{\alpha}}\, dx \\
&= \int_{0}^{\infty} \frac{x^{\alpha-1} e^{-x/\tilde{\beta}}}{\Gamma(\alpha)\beta^{\alpha}}\, dx
\end{aligned}
$$

where

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

Continuing,

$$
\begin{aligned}
&= \int_{0}^{\infty} \frac{x^{\alpha-1} e^{-x/\tilde{\beta}}}{\Gamma(\alpha)\tilde{\beta}^{\alpha}} \cdot \frac{\tilde{\beta}^{\alpha}}{\beta^{\alpha}}\, dx \\
&= \frac{\tilde{\beta}^{\alpha}}{\beta^{\alpha}} \times 1 \\
&= (1-\beta t)^{-\alpha}
\end{aligned}
$$

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

EXAMPLE 2.10.4

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

Solution.

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

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

Three important properties of $M(t)$.

THEOREM 2.10.5: 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: 2.10.5

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

as required.

EXAMPLE 2.10.6

(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)

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

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

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

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

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

THEOREM 2.10.7: 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

$$M^{(k)}(0) = \mathbf{E}\left[X^k\right]$$

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

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

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

Proof of: 2.10.7

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

EXAMPLE 2.10.8

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

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

We know,

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

For $\mathbf{Var}[X]$ we find $M''(t)$.

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

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

$$\mathbf{Var}[X] = \mathbf{E}\left[X^2\right] - \mu^2 = \alpha\beta^2(\alpha + 1) - (\alpha\beta)^2 = \alpha\beta^2$$

EXAMPLE 2.10.9

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

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

Therefore,

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

Now,

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

Therefore,

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

So,

$$\mathbf{Var}[X] = \mathbf{E}\left[X^2\right] - \mu^2 = \theta^2 + \theta - (\theta)^2 = \theta$$

THEOREM 2.10.10: 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,

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

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

EXAMPLE 2.10.11

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

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

Solution.

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

(ii)

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

Therefore,

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

Also,

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

and

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

Therefore,

$$\mathbf{Var}[Y] = \mathbf{E}\left[Y^2\right] - \mu^2 = 5 - 1 = 4$$

(iii) $M_Y(t) = \exp\left\{2t^2 - t\right\}$ 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 | 2020-09-21

EXAMPLE 2.10.12: 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

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

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

$P(X \leqslant x, Y \leqslant y)$: "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 (we consider their intersection).

REMARK 3.1.3

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

$$F(x_1, \ldots, x_n) = P(X_1 \leqslant x_1, \ldots, X_n \leqslant x_n)$$

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) $\lim_{x \to -\infty} F(x, y) = 0$ and $\lim_{y \to -\infty} F(x, y) = 0$ by looking at

$$\underbrace{\{X \leqslant x\}}_{\to 0} \cap \underbrace{\{Y \leqslant y\}}_{\to 0}$$

when $(x, y) \to (-\infty, -\infty)$.

(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

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

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

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

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

$$f(x, y) = P(X = x, Y = y)$$

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 of 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 $R \subseteq R^2$

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

EXAMPLE 3.2.4

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

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

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

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

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

EXAMPLE 3.2.6

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

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:

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

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

$$f_2(y) = P(Y = y) = qp^y$$

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

(iii) Find $P(X \leqslant Y)$.

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

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

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

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 $\frac{\partial^2}{\partial x \partial y} [F(x, y)]$ 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 $R \subseteq \mathbb{R}^2$:

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

EXAMPLE 3.3.4

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

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

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

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

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

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

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

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

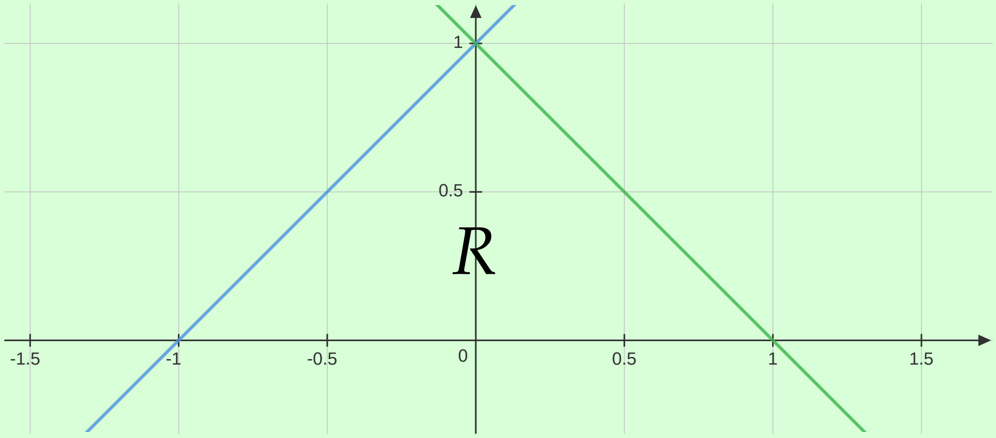
where $x_\ell(y)$ and $x_u(y)$ are continuous for $y_\ell \leqslant y \leqslant y_u$. If $f(x, y)$ is continuous on R, then

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

We use ℓ for “lower” amd u for “upper.”

EXAMPLE 3.3.7

Describe the region R below 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) $P(X \leqslant 1/3, Y \leqslant 1/2)$

(b) $P(X \leqslant Y)$

(c) $P(X + Y \leqslant 1/2)$

(d) $P(XY \leqslant 1/2)$

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

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

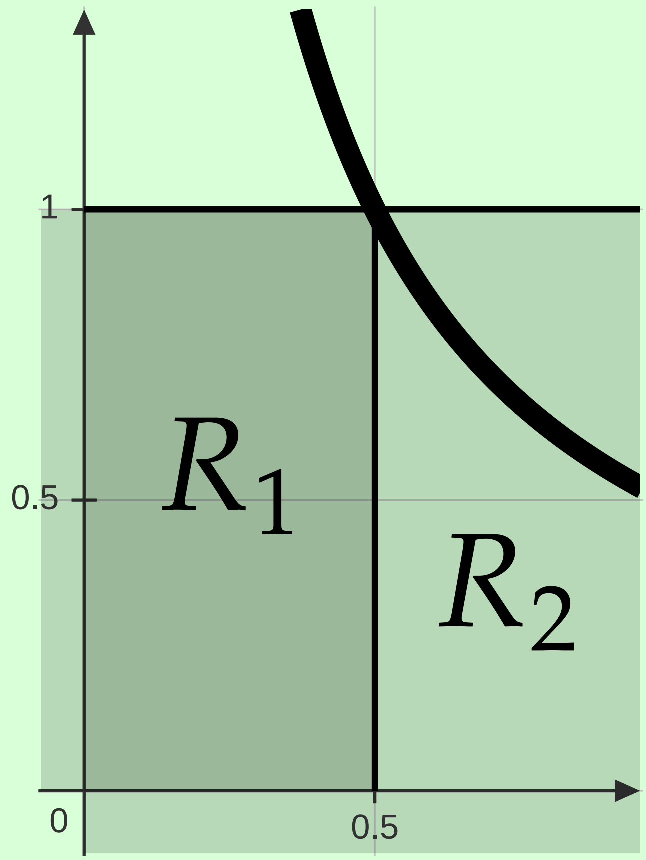
(b) $R = \{(x,y) : 0 \leqslant x \leqslant 1, x \leqslant y \leqslant 1\}$.

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

(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) $P(X \leqslant 1/3, Y \leqslant 1/2)$

  (b) $P(X \leqslant Y)$

  (c) $P(X + Y \geqslant 1)$

(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) $P(X \leqslant 1/3, Y \leqslant 1/2)$.

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

Therefore,

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

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

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

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

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

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

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

(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 $t \leqslant 0$, $F_T(t) = P(T \leqslant t) = 0$, so we only focus on $t > 0$, so $F_T(t) = P(T \leqslant t)$.

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

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, taking the derivative, the p.d.f. of T is

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

Author's note: The length of the last examples leave me to say one thing:

"This one does not spark joy."