

Table of Contents

- 2. [Univariate Random Variables](#)
 - 2.1 [Introduction to probability model](#)
 - 2.2 [Discrete random variable](#)
 - 2.3 [Continuous random variable](#)
 - 2.4 [Expectation](#)
 - 2.5 [Moment generating function](#)
- 3. [Joint Distribution](#)

2 Univariate Random Variables

2.1 Introduction to probability model

- **Probability model** is used to describe a random experiment.

It consists of three important components:

- i. **Sample space** S : a collection of all possible outcomes of one random experiment.

e.g. Toss a coin: $S = \{H, T\}$

e.g. Toss a coin twice: $S = \{(H, H), (H, T), (T, H), (T, T)\}$

- ii. **Event**: denoted by A, B, C , etc. It is a subset of sample space.

e.g. Toss a coin twice:

Define A as 1st toss is tail, $A = \{(T, T), (T, H)\} \subseteq S$

- iii. **Probability function** P : It is a function of events.

It satisfies properties (axioms):

- a. $0 \leq P(A) \leq 1$ for any event A .

- b. $P(S) = 1$

- c. Countable additivity: If A_1, A_2, \dots are assumed to be pairwise mutually exclusive events (i.e. $A_i \cap A_j = \emptyset$ for $i \neq j$), $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$.

We can now prove the following properties:

- a. $P(\emptyset) = 0$.

Proof: Let $A_i = \emptyset$ for $i \geq 1$, $A_i \cap A_j = \emptyset$ for $i \neq j$, by axioms we have $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$, or in other words, $P(\emptyset) = \sum_{i=1}^{\infty} P(\emptyset)$. Additionally, $0 \leq P(\emptyset) \leq 1$, therefore, $P(\emptyset) = 0$.

- b. Let A denote an event. Let \bar{A} denote the complementary event of A , which means \bar{A} satisfies two conditions:

- a. $\bar{A} \cap A = \emptyset$, and

- b. $\bar{A} \cup A = S$.

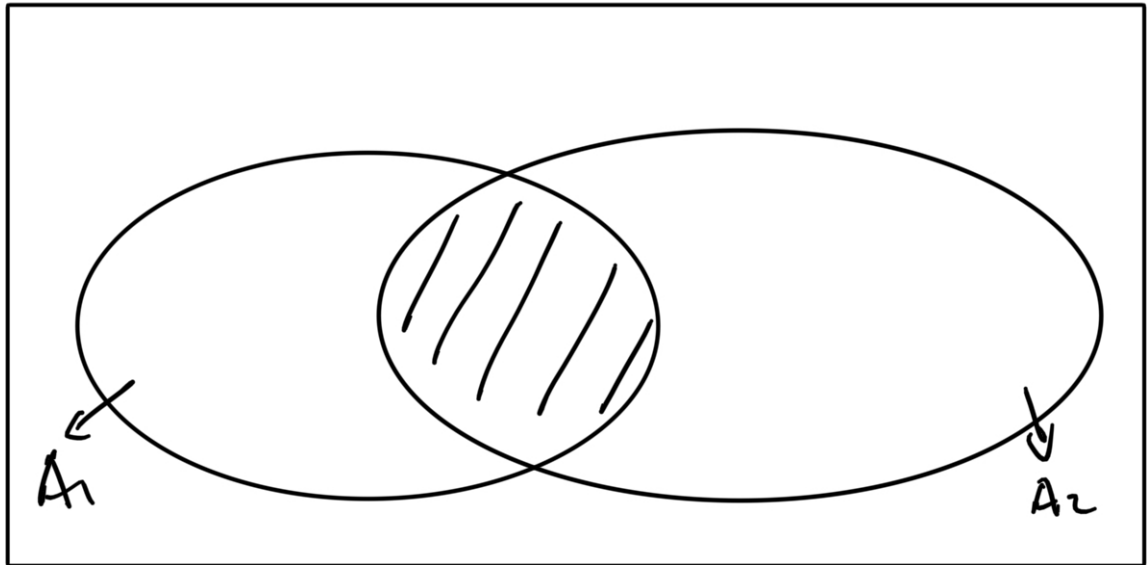
Prove $P(A) + P(\bar{A}) = 1$:

Proof: Define $A_1 = A$, $A_2 = \bar{A}$, $A_i = \emptyset$ for $i \geq 3$, so $A_i \cap A_j = \emptyset$ for $i \neq j$, by axioms we have $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$, in other words, $P(S) = P(A) + P(\bar{A}) + \sum_{i=3}^{\infty} 0$, therefore, $P(A) + P(\bar{A}) = 1$.

- c. If A_1 and A_2 are mutually exclusive, then $P(A_1 \cup A_2) = P(A_1) + P(A_2)$.

Proof: Define $A_i = \emptyset$ for $i \geq 3$, so $S = A_i \cap A_j = \emptyset$, for $i \neq j$. Then $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$, or in other words, $P(A_1 \cup A_2) = P(A_1) + P(A_2) + 0$.

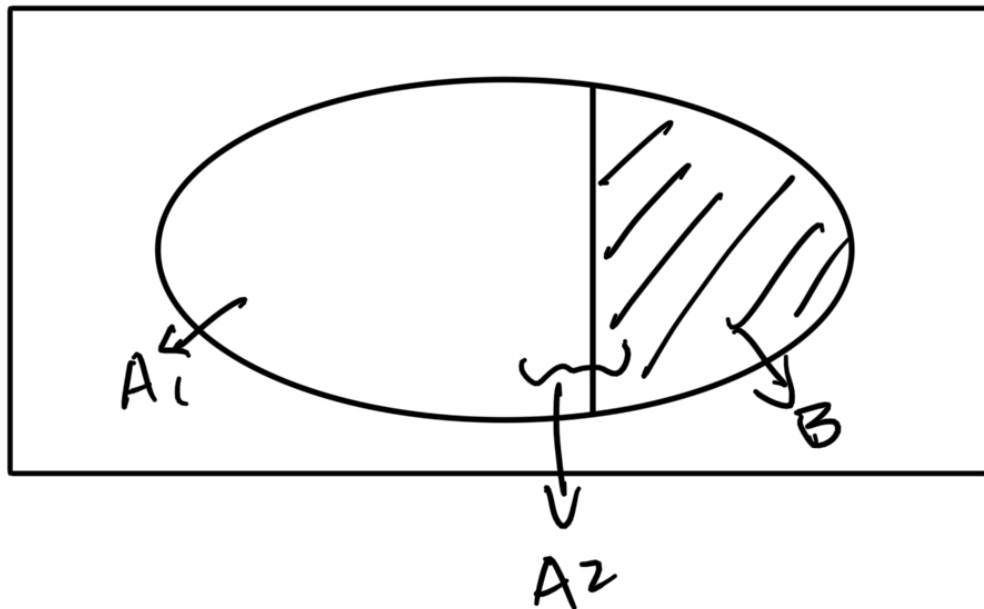
- d. In general, $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$.



Proof: Define $B = \{\omega | \omega \in A_1, \omega \notin A_2\}$, since $A_1 = B \cup (A_1 \cap A_2)$, we can get $B \cap (A_1 \cap A_2) = \emptyset$, $B \cup (A_1 \cap A_2) = A_1$, $B \cap (A_1 \cap A_2) = \emptyset$, $B \cap A_2 = \emptyset$, and therefore $B \cup A_2 = A_1 \cup A_2$.

Then $P(A_1 \cup A_2) = P(B \cup A_2) = P(B) + P(A_2)$. Note $P(A_1 \cup A_2) = P(A_2) + P(B)$ and $P(B) = P(A_1) - P(A_1 \cap A_2)$. Hence, $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$.

e. If $A_1 \subseteq A_2$, then $P(A_1) \leq P(A_2)$



Proof: $A_2 \setminus A_1 := B = \{\omega | \omega \in A_2, \omega \notin A_1\}$, we have $B \cap A_1 = \emptyset$, $B \cup A_1 = A_2$. Then $P(A_2) = P(A_1 \cup B) = P(A_1) + P(B) \geq P(A_1)$.

e.g. Toss a coin twice

Then $S = \{(H, H), (H, T), (T, H), (T, T)\}$ for any event A ,

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

Verify P is a probability function.

- **Conditional probability**

Suppose A and B denote two events. Provided $P(B) > 0$, then the conditional probability of A given B is

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

◦ Independence of two events

Suppose A and B denotes two events. We say A and B are independent if and only if

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

- Proposition: If A and B are independent, then $P(A|B) = P(A)$ (We assume $P(B) > 0$)

Proof: $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A)$

e.g. Toss a coin twice

$A :=$ 1st toss is a head $= \{(H, T), (H, H)\}$

$B :=$ 2nd toss is a head $= \{(T, H), (H, H)\}$

For any event C, $P(C) = \frac{\# \text{ of elements in } C}{4}$

Verify A and B are independent.

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

By definition, $A \cap B = \{(H, H)\} \implies P(A \cap B) = \frac{1}{4}$

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

Hence, $P(A \cap B) = P(A)P(B)$.

• Random variable (r.v.) X, Y, ζ, η

Random variable is a function from sample space to real line.

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

Specifically, given any $\omega \in S$, $X(\omega) \in \mathbb{R}$.

This function satisfies that for any $x \in \mathbb{R}$, $\{X \leq x\} = \{\omega | X(\omega) \leq x\}$ is an event.

e.g. Toss a coin twice

$X : \#$ of heads in two tosses.

$X : (H, H) \mapsto 2$.

We need to check for any x , $\{X \leq x\}$ is an event.

1. $x \geq 2$, $\{X \leq x\} = \{\omega | X(\omega) \leq x\} = S$
2. $x \in [1, 2)$, what is $\{X \leq x\}$?
3. $x \in [0, 1)$, what is $\{X \leq x\}$?
4. $x < 0$, what is $\{X \leq x\}$?

• Cumulative distribution of X (c.d.f.)

For any $x \in \mathbb{R}$, the c.d.f. of X is defined as $F(x) = P(X \leq x)$.

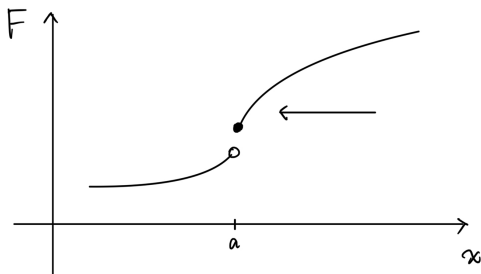
It satisfies the following property:

- $F(x)$ is a non-decreasing function, i.e., if $x_1 \leq x_2$, then $F(x_1) \leq F(x_2)$.

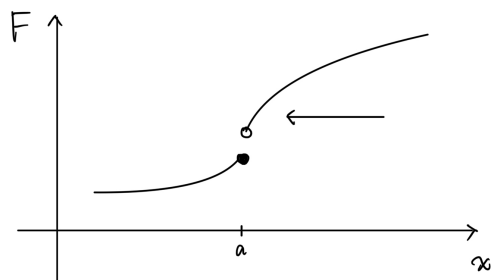
Proof: $\{X \leq x_1\}$ is an event. $\{X \leq x_1\} \subseteq \{X \leq x_2\}$ if $x_1 < x_2$, since $\{\omega | X(\omega) \leq x_1\} \subseteq \{\omega | X(\omega) \leq x_2\}$.

- $\lim_{x \rightarrow -\infty} F(x) = 0, \lim_{x \rightarrow \infty} F(x) = 1$.

- $F(x)$ is a right-continuous function, i.e., for any $a \in \mathbb{R}$, $\lim_{x \rightarrow a^+} F(x) = F(a)$.



right-continuous



not right-continuous

1, 2 and 3 are three basic properties of a c.d.f.

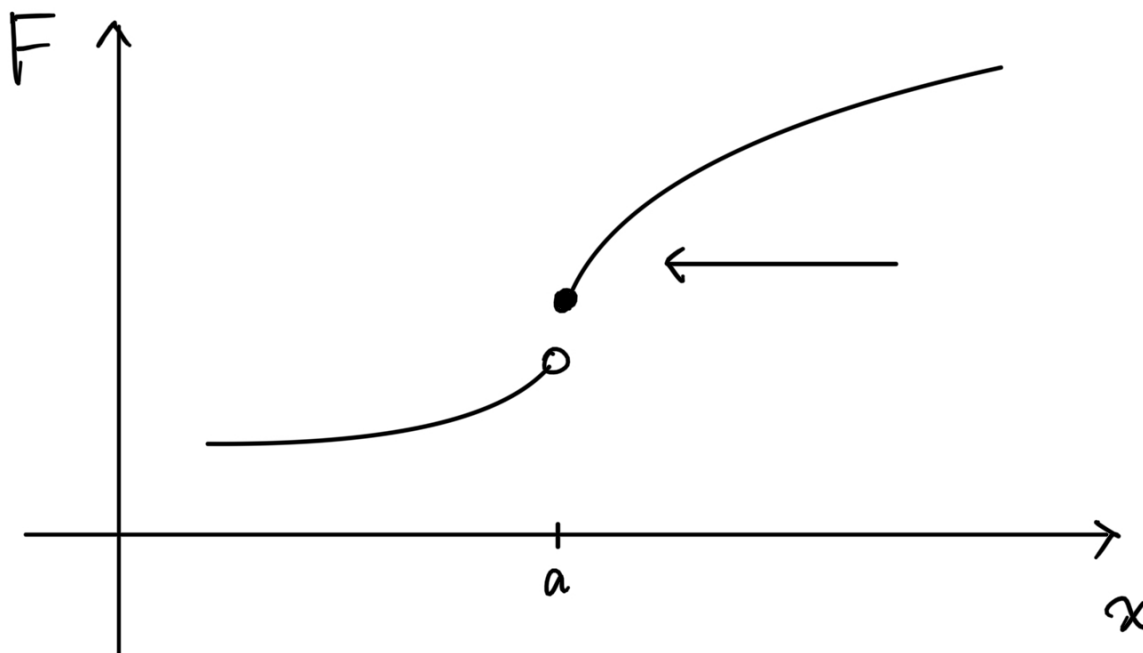
Some extra properties of a c.d.f.:

iv. $P(a < X \leq b) = F(b) - F(a)$.

Proof: Define $A = \{X \leq b\}$, $B := \{X \leq a\}$, $C = \{a < x \leq b\}$, we want to prove: $P(a < X \leq b) = P(X \leq b) - P(X \leq a) \iff P(C) = P(A) - P(B)$. Note $B \cap C = \emptyset$, $B \cup C = A$. Then $P(A) = P(B \cup C) = P(B) + P(C)$.

v. $P(X = a) = P(X \leq a) - P(X < a) = F(a) - F(a^-)$.

Proof: $P(X = a) = P(X \leq a) - P(X < a) = F(a) - \lim_{x \rightarrow a^-} F(x) = \lim_{x \rightarrow a^+} F(x) - \lim_{x \rightarrow a^-} F(x)$.



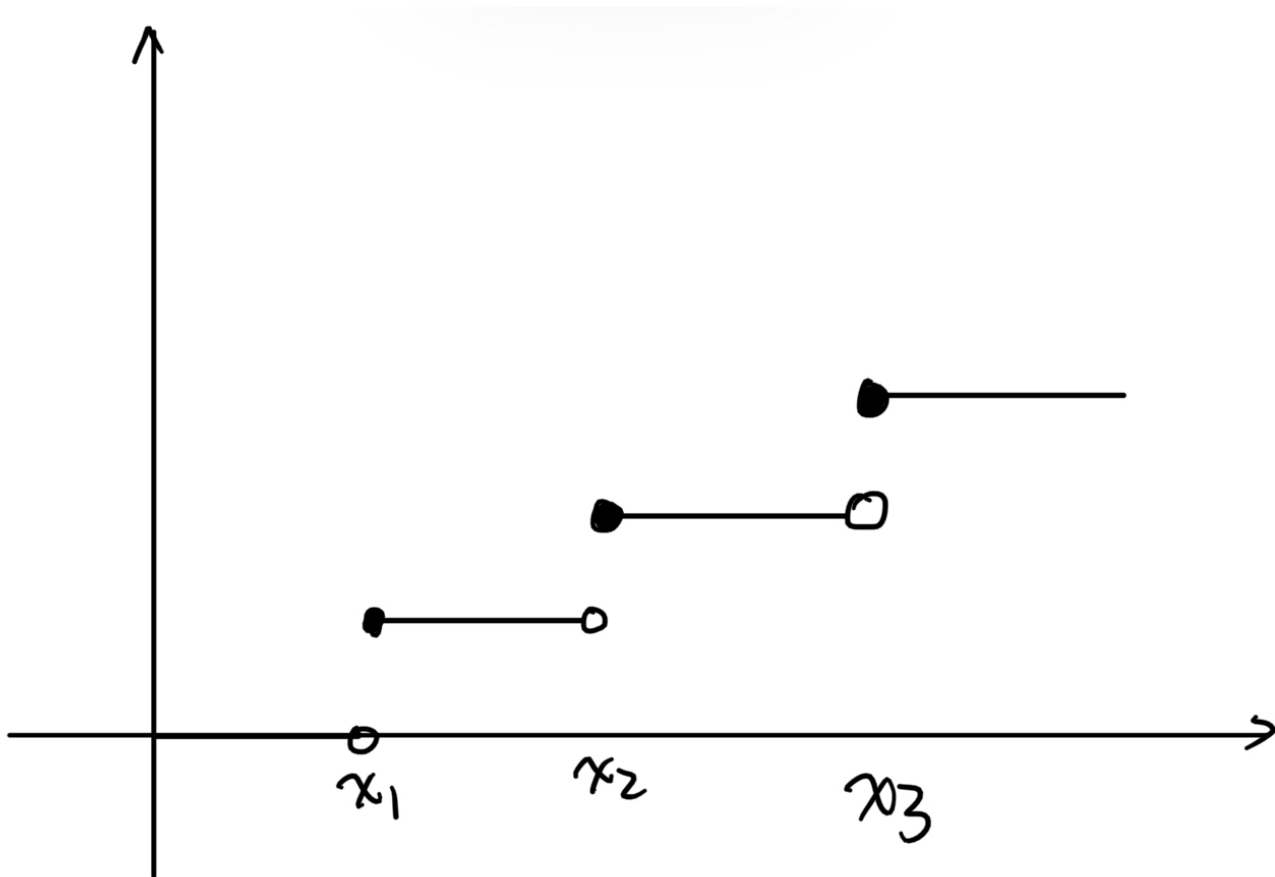
right-continuous

2.2 Discrete random variable

Definition:

If a random variable X can only take on a finite or countably infinite number of values, then X is called a discrete random variable.

- **cdf** of a discrete r.v. is a right continuous step function



- **Probability function (pf):** $f(x) = P(X = x)$.

For a discrete r.v., $f(x) \begin{cases} > 0 & \text{if } X \text{ can take value } x \\ = 0 & \text{if } X \text{ cannot take value } x \end{cases}$

- **Support:** The set $A = \{x : f(x) > 0\}$ is called the support of X . These are all the possible values that X can take.
- Properties of a p.f. f for a discrete r.v. X .

i. $f(x) \geq 0$ for any $x \in \mathbb{R}$.

ii. $\sum_{x \in A} f(x) = 1$.

Proof: The support of X is a countable set, $A = \{x_1, \dots, x_n\}$. Let $B_i = \{X = x_i\}$ is an event for $i = 1, \dots, n$. B_i are pairwise mutually exclusive events, i.e. $B_i \cap B_j = \emptyset$ for $i \neq j$. Then, $\bigcup_{i=1}^n B_i = S$. Then, $1 = P(S) = P\left(\bigcup_{i=1}^n B_i\right) = \sum_{i=1}^n P(B_i) = \sum_{i=1}^n P(X = x_i)$.

- Some commonly used discrete r.v.

i. Bernoulli r.v. $X \sim \text{Bern}(p)$.

X can only take two possible values, 0 and 1. $A = \{0, 1\}$.

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

ii. Binomial distribution

Toss a coin n times.

a. different tosses are independent

b. probability of getting a head is fixed, which is denoted by p .

X : # of heads across n tosses, then $X \sim \text{Bin}(n, p)$.

Hence the support of X , $A = \{0, 1, 2, \dots, n\}$.

The p.f. of X is $f(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$, $x \in A$.

$$\sum_{x=0}^n \binom{n}{x} p^x (1-p)^{n-x} = [p + (1-p)]^n = 1$$

iii. Geometric distribution

X : # of failures before the first success.

The support of X is $A = \{0, 1, \dots\}$.

$$f(x) = P(X = x) = (1-p)^x p, x \in A.$$

$$\sum_{x=0}^{\infty} (1-p)^x p = \frac{p}{1-(1-p)} = 1$$

iv. Negative binomial r.v. $X \sim \text{NegBin}(r, p)$

X : # of failures before the r th success.

v. Poisson r.v. $X \sim \text{Poisson}(\mu)$

The support of X , $A = \{0, 1, \dots\}$.

The probability function $f(x) = P(X = x) = \frac{\mu^x}{x!} e^{-\mu}$, $x \in A$.

$$\sum_{x \in A} f(x) = \sum_{x=0}^{\infty} \frac{\mu^x}{x!} e^{-\mu} = e^{-\mu} \sum_{x=0}^{\infty} \frac{\mu^x}{x!} = e^{-\mu} e^{\mu} = 1.$$

$$\text{Aside: } e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}.$$

2.3 Continuous random variable

Definition: If the collection of all possible values X can take is an interval or the real line, then X is called a continuous r.v.

- Remark: If X is continuous r.v., its cdf $F(x)$ is continuous everywhere. Moreover, F is differentiable almost everywhere. It is not differentiable at at most countable locations.

- Probability density function (pdf):

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

- Support of X : $A = \{x | f(x) > 0\}$.

- Basic property of f :

$$\text{i. } f(x) \geq 0 \text{ for any } x \in \mathbb{R}.$$

$$\text{ii. } \int_{-\infty}^{\infty} f(x) dx = 1.$$

- Extra properties of f :

$$\text{i. } F(x) = \int_{-\infty}^x f(t) dt = F(x) - F(-\infty) \text{ (find cdf from pdf).}$$

$$\text{ii. } f(x) = \begin{cases} F'(x) & \text{if } F \text{ is differentiable at } x \\ 0 & \text{otherwise} \end{cases} \text{ (find pdf from cdf).}$$

$$\text{iii. } P(X = x) = 0 \text{ and } f(x) \neq P(X = x) \text{ for any } x.$$

$$\text{If } F \text{ is differentiable at } x, \text{ then } f(x) = \lim_{h \rightarrow 0} \frac{F(x+h) - F(x)}{h}$$

$$\implies F(x+h) - F(x) \approx f(x) \cdot h$$

$$\implies P(x < X \leq x+h) \approx f(x) \cdot h.$$

$$\text{iv. } P(a < X \leq b) = F(b) - F(a) = P(a < X < b) = P(a \leq X \leq b)$$

Example (Uniform distribution):

Suppose the cdf is

$$F(x) = \begin{cases} 0 & x \leq a \\ \frac{x-a}{b-a} & a < x < b \\ 1 & x \geq b \end{cases}$$

Find pdf $f(x)$:

$$\text{The pdf is: } f(x) = \begin{cases} 0 & x \leq a \\ \frac{1}{b-a} & a < x < b \\ 0 & x \geq b \end{cases}$$

Example:

Define a function

$$f(x) = \begin{cases} \frac{\theta}{x^{\theta+1}} & x \geq 1 \\ 0 & \text{otherwise} \end{cases}$$

i. Find for what values of θ , f is a pdf?

Solution: $f(x) \geq 0$ for any $x \in \mathbb{R}$, therefore $\theta \geq 0$. $\int_{-\infty}^{\infty} f(x) dx = \int_1^{\infty} \frac{\theta}{x^{\theta+1}} dx$.

Case 1: $\theta = 0$, $\int_{-\infty}^{\infty} f(x) dx = 0 \neq 1$.

Case 2: $\theta > 0$, $\int_{-\infty}^{\infty} f(x) dx = \int_1^{\infty} \frac{\theta}{x^{\theta+1}} dx = -\frac{1}{x^{\theta}} \Big|_1^{\infty} = 1$.

ii. Find $F(x)$ if f is a pdf.

Solution: $F(x) = \int_{-\infty}^x f(t) dt$

Case 1: $x \leq 1$, $F(x) = \int_{-\infty}^x f(t) dt = 0$.

Case 2: $x > 1$, $F(x) = \int_{-\infty}^x f(t) dt = \int_1^x \frac{\theta}{t^{\theta+1}} dt = -\frac{1}{t^{\theta}} \Big|_1^x = 1 - \frac{1}{x^{\theta}}$.

iii. Find $P(2 < X < 3)$ and $P(-2 < X < 3)$.

Solution:

$$P(2 < X < 3) = F(3) - F(2) = \left(1 - \frac{1}{3^{\theta}}\right) - \left(1 - \frac{1}{2^{\theta}}\right) = \frac{1}{2^{\theta}} - \frac{1}{3^{\theta}}.$$

$$P(-2 < X < 3) = F(3) - F(-2) = \left(1 - \frac{1}{3^{\theta}}\right) - 0 = 1 - \frac{1}{3^{\theta}}.$$

$$P(-2 < X < 3) = \int_{-2}^3 f(x) dx = \int_{-2}^1 f(x) dx + \int_1^3 f(x) dx = \int_{-2}^1 0 dx + \int_1^3 \frac{\theta}{x^{\theta+1}} dx = -\frac{1}{x^{\theta}} \Big|_1^3 = 1 - \frac{1}{3^{\theta}}.$$

◦ Gamma function, $\Gamma(\alpha)$, $\alpha > 0$.

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

$$a. \Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1).$$

$$b. \Gamma(n) = (n - 1)! \text{ when } n \text{ is a positive integer, } \Gamma(1) = 1.$$

$$c. \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}.$$

Example (Gamma distribution):

The pdf is

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

if $\alpha > 0, \beta > 0$ are constants.

Verify f is a pdf.

Solution:

$$a. f(x) \geq 0 \text{ for any } x \in \mathbb{R}.$$

$$b. \int_{-\infty}^{\infty} f(x) dx = \int_{-\infty}^0 f(x) dx + \int_0^{\infty} f(x) dx = 0 + \int_0^{\infty} \frac{x^{\alpha-1} e^{-x/\beta}}{\beta^{\alpha} \Gamma(\alpha)} dx.$$

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

$$\text{Let } y = \frac{x}{\beta} \implies x = \beta y, dx = \beta dy.$$

$$\text{Then, } \int_0^{\infty} \frac{x^{\alpha-1} e^{-x/\beta}}{\beta^{\alpha} \Gamma(\alpha)} dx = \int_0^{\infty} \frac{(\beta y)^{\alpha-1} e^{-y}}{\beta^{\alpha} \Gamma(\alpha)} \beta dy = \frac{1}{\Gamma(\alpha)} \int_0^{\infty} y^{\alpha-1} e^{-y} dy = \frac{1}{\Gamma(\alpha)} \Gamma(\alpha) = 1.$$

Example (Weibull distribution):

The pdf is

$$f(x) = \begin{cases} \frac{\beta}{\theta^{\beta}} x^{\beta-1} \exp\left\{-\left(\frac{x}{\theta}\right)^{\beta}\right\} & x > 0 \\ 0 & x < 0 \end{cases}$$

where $\alpha > 0, \beta > 0$ are constants, $X \sim \text{Weibull}(\theta, \beta)$.

Verify f is a pdf.

Solution:

$$a. f(x) \geq 0 \text{ for any } x \in \mathbb{R}.$$

$$\begin{aligned} \text{b. } \int_{-\infty}^{\infty} f(x)dx &= \int_{-\infty}^0 f(x)dx + \int_0^{\infty} f(x)dx = 0 + \int_0^{\infty} \frac{\beta}{\theta^{\beta}} x^{\beta-1} \exp \left\{ -\left(\frac{x}{\theta}\right)^{\beta} \right\} dx. \\ \text{Let } y &= \left(\frac{x}{\theta}\right)^{\beta} \implies x = \theta y^{\frac{1}{\beta}}, dx = \theta^{\frac{1}{\beta}} y^{\frac{1}{\beta}-1} dy. \\ \text{Then, } \int_{-\infty}^{\infty} f(x)dx &= \int_0^{\infty} \frac{\beta}{\theta^{\beta}} (\theta y^{\frac{1}{\beta}})^{\beta-1} \exp \{-y\} \theta^{\frac{1}{\beta}} y^{\frac{1}{\beta}-1} dy = \Gamma(1) = 1. \end{aligned}$$

Exmample (Normal distribution/Gaussian distribution):

The pdf is

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

for $x \in \mathbb{R}$,

where $\mu \in \mathbb{R}$, $\sigma > 0$ are constants, $X \sim \text{Normal}(\mu, \sigma)$.

Verify f is a pdf.

Solution:

$$\text{a. } f(x) \geq 0 \text{ for any } x \in \mathbb{R}.$$

$$\text{b. } \int_{-\infty}^{\infty} f(x)dx = 1.$$

To verify 2, we start from a special case, where $\mu = 0$, $\sigma = 1$.

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \text{ i.e., } \int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 1.$$

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 2 \int_0^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx. \text{ Let } y = \frac{x^2}{2} \implies x = \sqrt{2y}, dx = \sqrt{2}dy.$$

$$\text{Then, } 2 \int_0^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{\pi}} \int_0^{\infty} e^{-y} y^{1-1/2} dy = \frac{1}{\sqrt{\pi}} \Gamma(1/2) = 1.$$

Prove $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ is a pdf for any $\mu \in \mathbb{R}$, $\sigma > 0$.

$$\text{a. } f(x) \geq 0 \text{ for any } x \in \mathbb{R}.$$

$$\text{b. } \int_{-\infty}^{\infty} f(x)dx = 1?$$

$$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx.$$

$$\text{Let } z = \frac{x-\mu}{\sigma} \implies x = \mu + \sigma z, dx = \sigma dz$$

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{z^2}{2}} dz = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = 1.$$

2.4 Expectation

- Definition of expectation for discrete r.v.

Suppose that X is a discrete r.v. with support A and p.f. $f(x)$.

Then, $E(X) = \sum_{x \in A} x f(x)$ provided $\sum_{x \in A} |x| f(x) < \infty$.

- Definition of expectation for continuous r.v.

Suppose that X is a continuous r.v. with support A and pdf $f(x)$.

Then $E(X) = \int_{-\infty}^{\infty} x f(x) dx$ provided $\int_{-\infty}^{\infty} |x| f(x) dx < \infty$.

Example (Cauchy distribution):

The pdf of X is $f(x) = \frac{1}{\pi(1+x^2)}$ for $x \in \mathbb{R}$.

Find $E(X)$.

Solution:

$$\int_{-\infty}^{\infty} |x| f(x) dx = \int_{-\infty}^{\infty} |x| \frac{1}{\pi(1+x^2)} dx = 2 \int_0^{\infty} \frac{x}{\pi(1+x^2)} dx = \frac{\ln(1+x^2)}{\pi} \Big|_0^{\infty} = \infty.$$

Therefore, $E(X)$ does not exist.

Example:

Suppose p.f. $f(x) = \frac{1}{x(x+1)}$ for $x = 1, 2, 3, \dots$, the support of X is $A = \{1, 2, 3, \dots\}$.

- Show f is a p.f.

Solution:

$$\text{i. } f(x) \geq 0 \text{ for any } x \in \mathbb{R}.$$

$$\text{ii. } \sum_{x \in A} f(x) = \sum_{x \in A} \frac{1}{x(x+1)} = \sum_{x=1}^{\infty} \left(\frac{1}{x} - \frac{1}{x+1} \right) = 1 - \frac{1}{2} + \frac{1}{2} - \frac{1}{3} + \frac{1}{3} - \frac{1}{4} + \dots = 1.$$

- Find $E(X)$.

Solution: $E(X) = \sum_{x \in A} x f(x) = \sum_{x \in A} x \frac{1}{x(x+1)} = \sum_{x \in A} \frac{1}{x+1} = \sum_{x=1}^{\infty} \frac{1}{x+1} = \infty$.
 $E(X)$ does not exist.

More examples of expectations:

i. Binomial Distribution, $X \sim \text{Bin}(n, p)$.

Solution 1: $E(X) = \sum_{x \in A} x f(x) = \sum_{x=0}^n x \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} = \sum_{x=1}^n x \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} = \sum_{x=1}^n \frac{n!}{(x-1)!(n-x)!} p^x (1-p)^{n-x}$

Let $y = x - 1$, then $\sum_{x=1}^n \frac{n!}{(x-1)!(n-x)!} p^x (1-p)^{n-x} = np \sum_{y=0}^{n-1} \frac{(n-1)!}{y!(n-1-y)!} p^y (1-p)^{n-1-y} = np$, since $\sum_{y=0}^{n-1} \frac{(n-1)!}{y!(n-1-y)!} p^y (1-p)^{n-1-y}$ is a pf of $\text{Bin}(n-1, p)$.

Solution 2: For the i th trial, $X_i = \begin{cases} 1 & \text{if the } i\text{th outcome is a success} \\ 0 & \text{otherwise} \end{cases}$.

Then, $P(X_i = 1) = p$. Let $X = \sum_{i=1}^n X_i$, then $X \sim \text{Bin}(n, p)$.

$E(X) = E(\sum_{i=1}^n X_i) = \sum_{i=1}^n E(X_i) = \sum_{i=1}^n 1 \cdot P(X_i = 1) = np$.

ii. Suppose X is a continuous r.v. with pdf $f(x) = \begin{cases} \frac{\theta}{x^{\theta+1}} & x \geq 1 \\ 0 & \text{otherwise} \end{cases}$, where $\theta > 0$ is a constant. Find $E(X)$, and determine the values of θ for which $E(X)$ exists.

Solution: $\int_{-\infty}^{\infty} |x| f(x) dx = \int_1^{\infty} x \frac{\theta}{x^{\theta+1}} dx = \int_1^{\infty} \frac{\theta}{x^{\theta}} dx < \infty$ iff $\theta > 1$.

When $\theta > 1$, $E(X) = \int_{-\infty}^{\infty} x f(x) dx = \int_1^{\infty} \frac{\theta x}{x^{\theta+1}} dx = \theta \int_1^{\infty} \frac{1}{x^{\theta}} dx = \left(\frac{\theta}{1-\theta} x^{1-\theta} \right) \Big|_1^{\infty} = \frac{\theta}{\theta-1}$.

When $\theta \leq 1$, $E(X)$ does not exist.

• Expectation of a function of X

Suppose that X is a r.v., what is $E(g(X))$, where g is a real function?

For example, $g(x) = x^2$.

Let $Y = g(X)$, find $E(Y)$.

• Case 1: If X is a discrete r.v. with support A and p.f. $f(x)$, then $E(g(X)) = \sum_{x \in A} g(x) f(x)$ provided $\sum_{x \in A} |g(x)| f(x) < \infty$.

• Case 2: If X is a continuous r.v. with support A and pdf $f(x)$, then $E(g(X)) = \int_{-\infty}^{\infty} g(x) f(x) dx$ provided $\int_{-\infty}^{\infty} |g(x)| f(x) dx < \infty$.

• Linearity Property: If a and b are two constants, then $E[ag(X) + bg(X)] = aE(g(X)) + bE(h(X))$.

• Variance: $\text{Var}(X) = E[(X - \mu)]^2 = E(X^2) - \mu^2 = E(X^2) - [E(X)]^2$ where $\mu = E(X)$.

• Moments:

• k th moment about 0: $E(X^k)$.

• k th moment about mean: $E[(X - \mu)^k]$, where $\mu = E(X)$.

Example (Poisson distribution):

Suppose $X \sim \text{Poisson}(\mu)$, where $\mu > 0$ is a constant.

Find $E(X)$ and $\text{Var}(X)$.

Solution: $E(X) = \sum_{x=0}^{\infty} x \frac{\mu^x}{x!} e^{-\mu} = \mu e^{-\mu} \sum_{x=1}^{\infty} \frac{\mu^{x-1}}{(x-1)!}$.

Let $y = x - 1$, then $E(X) = \mu e^{-\mu} \sum_{y=0}^{\infty} \frac{\mu^y}{y!} = \mu e^{-\mu} e^{\mu} = \mu$.

$E(X^2) = \sum_{x=0}^{\infty} x^2 \frac{\mu^x}{x!} e^{-\mu} = \sum_{x=1}^{\infty} \frac{x \mu^x}{(x-1)!} e^{-\mu} = \sum_{x=1}^{\infty} \frac{(x-1+1) \mu^x}{(x-1)!} e^{-\mu} = \sum_{x=1}^{\infty} \frac{(x-1)^2 \mu^x}{(x-1)!} e^{-\mu} + \sum_{x=1}^{\infty} \frac{\mu^x}{(x-1)!} e^{-\mu} = \sum_{x=2}^{\infty} \frac{(x-1) \mu^x}{(x-1)!} e^{-\mu} + \sum_{x=2}^{\infty} \frac{\mu^x}{(x-2)!} e^{-\mu}$.

Let $y = x - 2$, then $\sum_{y=0}^{\infty} \frac{\mu^{y+2}}{y!} e^{-\mu} = \mu^2 e^{-\mu} \sum_{y=0}^{\infty} \frac{\mu^y}{y!} = \mu^2$.

That means $E(X^2) = \mu^2 + \mu$, and $\text{Var}(X) = E(X^2) - [E(X)]^2 = \mu^2 + \mu - \mu^2 = \mu$.

Example (Gamma distribution):

Suppose $X \sim \text{Gamma}(\alpha, \beta)$. Find $E(X^k)$, $k > 0$.

pdf of X is $f(x) = \begin{cases} \frac{x^{\alpha-1} e^{-x/\beta}}{\beta^{\alpha} \Gamma(\alpha)} & x > 0 \\ 0 & \text{otherwise} \end{cases}$.

Solution: $E(X^k) = \int_{-\infty}^{\infty} x^k f(x) dx = \int_0^{\infty} x^k \frac{x^{\alpha-1} e^{-x/\beta}}{\beta^{\alpha} \Gamma(\alpha)} dx$. Let $y = \frac{x}{\beta} \implies x = \beta y, dx = \beta dy$.

Then, $E(X^k) = \int_0^{\infty} \frac{(\beta y)^k (\beta y)^{\alpha-1} e^{-y}}{\beta^{\alpha} \Gamma(\alpha)} \beta dy = \frac{\beta^k}{\Gamma(\alpha)} \int_0^{\infty} y^{k+\alpha-1} e^{-y} dy = \frac{\beta^k}{\Gamma(\alpha)} \Gamma(k+\alpha) = \frac{\beta^k \Gamma(k+\alpha)}{\Gamma(\alpha)}$.

In particular, if $k = 1$, $E(X) = \frac{\beta \Gamma(1+\alpha)}{\Gamma(\alpha)} = \frac{\beta \alpha \Gamma(\alpha)}{\Gamma(\alpha)} = \alpha \beta$.

$k = 2$, $E(X^2) = \frac{\beta^2 \Gamma(2+\alpha)}{\Gamma(\alpha)} = \frac{\beta^2 (\alpha+1) \alpha \Gamma(\alpha)}{\Gamma(\alpha)} = \alpha(\alpha+1) \beta^2$.

$\text{Var}(X) = E(X^2) - [E(X)]^2 = \alpha(\alpha+1) \beta^2 - (\alpha \beta)^2 = \alpha \beta^2$.

Alternatively:

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

Define $\alpha^* = k + \alpha$, then $E(X^k) = \int_0^\infty \frac{x^{\alpha^*-1} e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)} dx = \int_0^\infty \frac{x^{\alpha^*-1} e^{-x/\beta}}{\beta^{\alpha^*} \Gamma(\alpha^*)} \frac{\beta^{\alpha^*} \Gamma(\alpha^*)}{\beta^\alpha \Gamma(\alpha)} dx = \frac{\beta^{\alpha^*} \Gamma(\alpha^*)}{\beta^\alpha \Gamma(\alpha)} \int_0^\infty \frac{x^{\alpha^*-1} e^{-x/\beta}}{\beta^{\alpha^*} \Gamma(\alpha^*)} dx = \frac{\beta^{\alpha^*} \Gamma(\alpha^*)}{\beta^\alpha \Gamma(\alpha)} = \frac{\beta^{k+\alpha} \Gamma(k+\alpha)}{\beta^\alpha \Gamma(\alpha)} = \frac{\beta^k \Gamma(k+\alpha)}{\Gamma(\alpha)}.$

2.5 Moment generating function

- Definition: Suppose X is a random variable, then $M(t) = E(E^{tx})$ is called the moment generating function (mgf) of X if $M(t)$ exists for $t \in (-h, h)$ for some $h > 0$.

Example (Gamma distribution):

Suppose $X \sim \text{Gamma}(\alpha, \beta)$. Find the mgf of X .

Solution: $M(t) = E(e^{tX}) = \int_{-\infty}^\infty e^{tx} f(x) dx = \int_0^\infty e^{tx} \frac{x^{\alpha-1} e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)} dx = \int_0^\infty \frac{x^{\alpha-1} e^{-(1/\beta - t)x}}{\beta^\alpha \Gamma(\alpha)} dx$. (Note: $1/\beta > t$, otherwise the integral diverges.)

Let $y = (1/\beta - t)x$, then $x = \frac{y}{1/\beta - t} = \frac{\beta y}{1 - t\beta}$, $dx = \frac{\beta}{1 - t\beta} dy$.

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

Example (Poisson distribution):

Suppose $X \sim \text{Poisson}(\mu)$. Find the mgf of X .

Solution: $M(t) = E(e^{tX}) = \sum_{x=0}^\infty e^{tx} \frac{\mu^x}{x!} e^{-\mu} = e^{-\mu} \sum_{x=0}^\infty \frac{(\mu e^t)^x}{x!} = e^{-\mu} e^{\mu e^t} \sum_{x=0}^\infty \frac{(\mu e^t)^x}{x!} e^{-e^t \mu} = e^{\mu(e^t - 1)}.$

Example (Normal distribution):

Suppose $X \sim N(0, 1)$. Find the mgf of X .

Solution: $M(t) = E(e^{tX}) = \int_{-\infty}^\infty e^{tx} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} e^{tx - \frac{x^2}{2}} dx = \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x^2 - 2tx + t^2) + \frac{1}{2}t^2} dx = e^{\frac{1}{2}t^2} \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-t)^2} dx = e^{\frac{1}{2}t^2}.$

Question: How to find the mgf of $N(\mu, \sigma^2)$?

- Three important properties of mgf

- Suppose the mgf of X is $M(t)$. If $Y = aX + b$, where a and b are constants, then the mgf of Y is $M_Y(t) = e^{bt} M(at)$.

If $Y \sim N(\mu, \sigma^2)$, then $X = \frac{Y - \mu}{\sigma} \sim N(0, 1)$.

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

$M_Y(t) = e^{\mu t} M_X(\sigma t) = e^{\mu t} e^{\frac{1}{2}\sigma^2 t^2}.$

- Find the k th moment of X about 0 from $M(t)$:

$E(X^k) = M^{(k)}(0) = \left. \frac{d^k}{dt^k} M(t) \right|_{t=0}.$

$M(t) = E(e^{tX})$, $M'(t) = E(X e^{tX})$.

In particular, $E(X) = M'(0)$, $E(X^2) = M''(0)$. Then, $\text{Var}(X) = E(X^2) - [E(X)]^2 = M''(0) - [M'(0)]^2.$

Example (Gamma distribution):

If $X \sim \text{Gamma}(\alpha, \beta)$, $M(t) = \left(\frac{1}{1 - t\beta} \right)^\alpha$, where $t < \frac{1}{\beta}$.

Find $E(X)$ and $\text{Var}(X)$.

Solution: $M'(t) = \alpha\beta(1 - \beta t)^{-\alpha-1}$, $M''(t) = \alpha(\alpha + 1)\beta^2(1 - \beta t)^{-\alpha-2}.$

Then, $E(X) = M'(0) = \alpha\beta$, $E(X^2) = M''(0) = \alpha(\alpha + 1)\beta^2.$

- Uniqueness of mgf.

Namely, X and Y have the same distribution iff X and Y have the same mgf.

Example: X has mgf $M(t) = e^{t^2/2}$

- Find mgf of $Y = 2X - 1$.

Solution: $M_Y(t) = e^{-t} M_X(2t) = e^{-t} e^{2t^2}.$

- Find $E(Y)$ and $\text{Var}(Y)$.

Solution: $M'_Y(t) = (4t - 1)e^{2t^2 - t}$. $E(X) = M'_Y(0) = -1.$

$M''_Y(t) = 4e^{2t^2 - t} + (4t - 1)^2 e^{2t^2 - t}$. $E(Y^2) = M''_Y(0) = 1 + 4 = 5.$

$\text{Var}(Y) = E(Y^2) - [E(Y)]^2 = 5 - (-1)^2 = 4.$

- What is the distribution of Y ?

Solution: $Y \sim N(-1, 4)$, since $M_Y(t) = e^{-t} e^{2t^2}.$

3 Joint distribution

3.1 Joint and Marginal cdfs

- Definition of joint cdf

Suppose that X and Y are two r.v.s. The joint cdf of X and Y is defined by $F(x, y) = P(X \leq x, Y \leq y)$ for $x, y \in \mathbb{R}$.

Remark: This definition can be extended to n r.v.s. X_1, X_2, \dots, X_n .

Joint cdf is $F(x_1, x_2, \dots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n)$.

However, we will focus on the case of $n = 2$.

- Properties of joint cdf

i. Fix y , $F(x, y)$ is monotone increasing function of x , i.e., $F(x_1, y) \leq F(x_2, y)$ if $x_1 < x_2$.

Proof: $F(x_1, y) = P(X \leq x_1, Y \leq y)$, since $\{X \leq x_1, Y \leq y\} \subset \{X \leq x_2, Y \leq y\}$, $F(x_1, y) \leq F(x_2, y)$.

ii. Fix x , $F(x, y)$ is monotone increasing function of y , i.e., $F(x, y_1) \leq F(x, y_2)$ if $y_1 < y_2$.

iii. $\lim_{x \rightarrow -\infty} F(x, y) = 0 = \lim_{y \rightarrow -\infty} F(x, y)$.

Proof: $F(x, y) = P(X \leq x, Y \leq y) \leq P(X \leq x)$, and consider $\lim_{x \rightarrow -\infty} P(X \leq x) = 0$, additionally, by property of joint cdf, $F(x, y) \geq 0$, then by squeeze theorem, $\lim_{x \rightarrow -\infty} F(x, y) = 0$.

iv. $\lim_{x \rightarrow \infty, y \rightarrow \infty} F(x, y) = 1$.

Proof: Consider set $Axy = \{X \leq x\} \cup \{Y \leq y\}$, then as $x, y \rightarrow \infty$, $P(\overline{Axy}) \rightarrow 0$, then $F(x, y) = P(Axy) \rightarrow 1$.

v. How to find marginal cdf from the joint one?

$$F_1(x) = P(X \leq x) = \lim_{y \rightarrow \infty} F(x, y).$$

Define $Ax = \{X \leq x\}$, $By = \{Y \leq y\}$.

As $y \rightarrow \infty$, $Ax \cup By \rightarrow Ax$.

$$F_2(y) = P(Y \leq y) = \lim_{x \rightarrow \infty} F(x, y).$$

3.2 Joint Discrete r.v.s

- Definition: If both X and Y are discrete r.v.s, then as a pair, $X \& Y_{(X,Y)}$ are joint discrete r.v.s X and Y .

- Definition of joint p.f.:

The joint p.f. of X and Y is given by $f(x, y) = P(X = x, Y = y)$ for any $x, y \in \mathbb{R}$.

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

- Basic properties of joint p.f.:

i. $f(x, y) \geq 0$ for any $(x, y) \in \mathbb{R}^2$.

ii. $\sum_{(x,y) \in A} f(x, y) = 1$.

Question: How to find probability over a region $C \subseteq \mathbb{R}^2$?

iii. $P((X, Y) \in C) = \sum_{(x,y) \in C} f(x, y)$.

Question: How to find marginal p.f. from the joint one?

iv. $f_1(x) = P(X = x) = P(X = xY < \infty) = \sum_{y \in \mathbb{R}} f(x, y)$.

E.g. Suppose X and Y are independent discrete r.v.s with joint p.f. $f(x, y) = kq^2p^{x+y}$ for $x = 0, 1, \dots$ and $y = 0, 1, \dots$, and 0 elsewhere. Here $p \in (0, 1)$ is a constant, $q = 1 - p$.

a. Find k .

Solution: Since $f(x, y) \geq 0$ for any $(x, y) \in \mathbb{R}^2$, $k > 0$. Since $\sum_{x=0}^{\infty} \sum_{y=0}^{\infty} f(x, y) = 1$, Then,

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

Therefore, $k = 1$

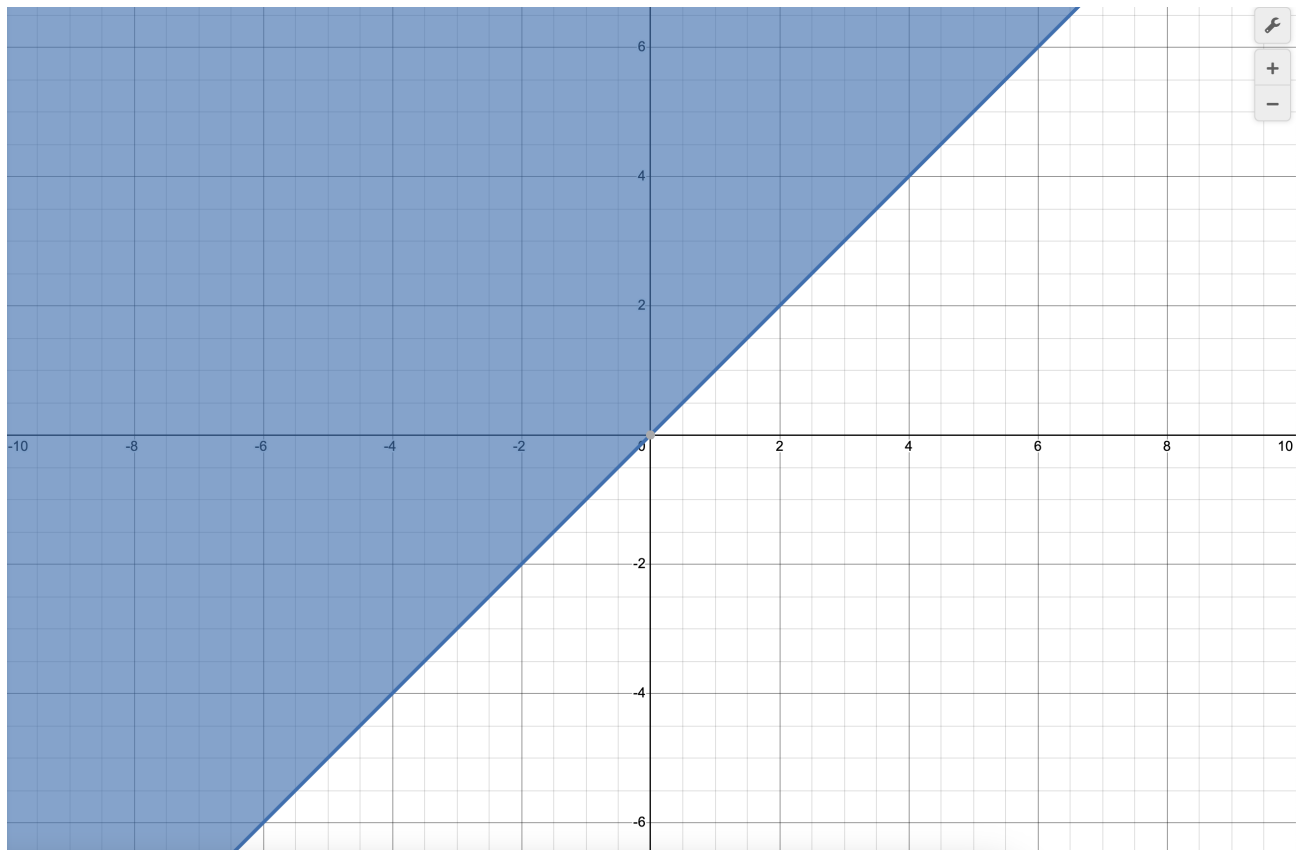
b. Find the marginal p.f. of X and find marginal p.f. of Y .

Solution: The support of X is $Ax = \{0, 1, 2, \dots\}$.

Here, $f_1(x) = \sum_{y \in \mathbb{R}} f(x, y) = 0$ if $x \notin Ax$

Given $X \in Ax$, then $f_1(x) = \sum_{y \in \mathbb{R}} f(x, y) = \sum_{y=0}^{\infty} f(x, y) = \sum_{y=0}^{\infty} p^{x+y} q^2 = q^2 p^x \sum_{y=0}^{\infty} p^y = q^2 p^x \frac{1}{1-p} = qp^x$.

c. $P(X \leq Y)$



Solution: $P(X \leq Y) = \sum_{(x,y) \in C} f(x,y)$ where $C = \{(x,y) \in \mathbb{R}^2 : x \leq y\}$, therefore, $P(X \leq Y) = \sum_{y=0}^{\infty} \sum_{x=0}^y p^{x+y} q^2 = \sum_{x=0}^{\infty} p^x q^2 \sum_{y=x}^{\infty} p^y = \sum_{x=0}^{\infty} p^x q^2 \frac{p^x}{1-p} = q \sum_{x=0}^{\infty} p^{2x} = q \frac{1}{1-p^2} = \frac{1}{1+p}$.

3.3 Joint Continuous r.v.s

- Definition: If joint cdf of (X, Y) can be written as $F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) du dv$ then X and Y are joint continuous r.v.s with joint pdf $f(x, y)$.

Namely, $f(x, y) = \begin{cases} \frac{\partial^2}{\partial x \partial y} F(x, y) & \text{if exists} \\ 0 & \text{o.w.} \end{cases}$.

- Definition of joint support: $A = \{(x, y) \in \mathbb{R}^2 : f(x, y) > 0\}$.
- Properties of joint pdf:

- $f(x, y) \geq 0$ for any $(x, y) \in \mathbb{R}^2$.
- $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1$.

Question: How to find probability over a region $C \subseteq \mathbb{R}^2$?

iii. $P((X, Y) \in C) = \iint_{(x,y) \in C} f(x, y) dx dy$.

Question: How to find marginal pdf from the joint one?

iv. $f_1(x) = \int_{-\infty}^{\infty} f(x, y) dy$ and $f_2(y) = \int_{-\infty}^{\infty} f(x, y) dx$.

E.g. X and Y are joint continuous r.v.s with joint pdf $f(x, y) = \begin{cases} x + y & \text{if } 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{o.w.} \end{cases}$.

a. Show f is a joint pdf.

Solution: $f(x, y) \geq 0$ for any $(x, y) \in \mathbb{R}^2$.

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = \int_0^1 \int_0^1 (x + y) dx dy = \int_0^1 \left(\frac{x^2}{2} + xy \right) \Big|_{x=0}^{x=1} dy = \int_0^1 \left(\frac{1}{2} + y \right) dy = \frac{1}{2} + \frac{1}{2} = 1.$$

b. Find

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

Solution: $P(X \leq 1/3, Y \leq 1/2) = \int_0^{1/3} \int_0^{1/2} (x + y) dy dx = \int_0^{1/3} \left(xy + \frac{y^2}{2} \right) \Big|_{y=0}^{y=1/2} dx = \int_0^{1/3} \left(\frac{x}{2} + \frac{1}{8} \right) dx = \frac{1}{36} + \frac{1}{24} = \frac{5}{72}$.

b. $P(X \leq Y)$

$$\text{Solution: } P(X \leq Y) = \iint_C f(x, y) dx dy = \int_0^1 dx \int_x^1 (x + y) dy = \int_0^1 dy \int_0^y (x + y) dx = \int_0^1 \left(\frac{x^2}{2} + xy \right) \Big|_{x=0}^{x=y} dy = \int_0^1 \left(\frac{y^2}{2} + y^2 \right) dy = \frac{1}{2}.$$

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

Solution: Let $C = \{(x, y) | x + y \leq \frac{1}{2}, 0 \leq x \leq 1, 0 \leq y \leq 1\}$.

$$\text{Then, } P(X + Y \leq 1/2) = \iint_C f(x, y) dx dy = \int_0^{1/2} \int_0^{1/2-x} (x + y) dy dx = \int_0^{1/2} \left(xy + \frac{y^2}{2} \right) \Big|_{y=0}^{y=1/2-x} dx = \int_0^{1/2} \left(\frac{x}{2} - \frac{x^2}{4} + \frac{1}{8} \right) dx = \int_0^{1/2} \left(-\frac{x^2}{2} + \frac{1}{8} \right) dx = \left(-\frac{x^3}{6} + \frac{x}{8} \right) \Big|_0^{1/2} = \frac{1}{24}.$$

d. $P(XY \leq 1/2)$

Solution: Find $P(XY > 1/2)$ first.

$$P(XY > 1/2) = \int_0^{1/2} \int_0^{1/2/x} (x + y) dy dx = \int_0^{1/2} \left(xy + \frac{y^2}{2} \right) \Big|_{y=0}^{y=1/2x} dx = \int_0^{1/2} \left(x - \frac{1}{8x} \right) dx = \left(\frac{x^2}{2} + \frac{1}{8x} \right) \Big|_0^{1/2} = \frac{1}{4}.$$

$$\text{Therefore, } P(XY \leq 1/2) = 1 - P(XY > 1/2) = 1 - \frac{1}{4} = \frac{3}{4}$$

c. Find marginal pdf of X and Y .

Solution: The support of X is $[0, 1]$.

$$\text{Given } x \in [0, 1], f_1(x) = \int_{-\infty}^{\infty} f(x, y) dy = \int_0^1 (x + y) dy = \left(xy + \frac{y^2}{2} \right) \Big|_{y=0}^{y=1} = x + \frac{1}{2}.$$

E.g. Suppose $f(x) = \begin{cases} ke^{-x-y} & 0 < x < y < \infty \\ 0 & \text{o.w.} \end{cases}$ is the joint pdf of (X, Y) .

a. Find k .

Solution: $f(x, y) \geq 0$ for any $(x, y) \in \mathbb{R}^2$, therefore, $k \geq 0$.

$$\text{Now, } \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = \int_0^{\infty} \int_x^{\infty} ke^{-x-y} dy dx = \int_0^{\infty} ke^{-x} (-e^{-y}) \Big|_x^{\infty} dx = \int_0^{\infty} ke^{-2x} dx = -\frac{k}{2} e^{-2x} \Big|_0^{\infty} = \frac{k}{2} = 1, \text{ therefore, } k = 2.$$

b. Find:

a. $P(X \leq \frac{1}{3}, Y \leq \frac{1}{2})$

$$\text{Solution: Let } C = \{(x, y) | x \leq 1/3, y \leq 1/2, 0 < x < y\}. \text{ Then, } P(X \leq \frac{1}{3}, Y \leq \frac{1}{2}) = \iint_C f(x, y) dx dy = \int_0^{1/3} \int_x^{1/2} 2e^{-x-y} dy dx = \int_0^{1/3} 2e^{-x} (-e^{-y}) \Big|_x^{1/2} dx = \int_0^{1/3} 2e^{-x} (-e^{-1/2} + e^{-x}) dx = \int_0^{1/3} 2e^{-x} (e^{-x} - e^{-1/2}) dx = \int_0^{1/3} 2e^{-2x} - 2e^{-1/2} e^{-x} dx = -e^{-2x} + 2e^{-1/2} e^{-x} \Big|_0^{1/3} = 1 - 2e^{-1/2} - e^{-2/3} - e^{-5/6}.$$

b. $P(X \leq Y)$

Solution: $P(X \leq Y) = 1$

c. $P(X + Y \geq 1)$

Solution: Let $C = \{(x, y) | x + y \geq 1, 0 < x < y\}$

Let's find $P(X + Y < 1)$ first.

$$P(X + Y < 1) = \iint_{x, y \in \mathbb{R}} 2e^{-x-y} dy dx = \int_0^{1/2} \int_x^{1-x} 2e^{-x-y} dy dx = \int_0^{1/2} 2e^{-x} (-e^{-y}) \Big|_x^{1-x} dx = \int_0^{1/2} 2e^{-x} (-e^{-x+1} + e^{-x}) dx = \int_0^{1/2} 2e^{-2x} - 2e^{-1} dx = -e^{-2x} - 2e^{-1} x \Big|_0^{1/2} = 1 - 2e^{-1}.$$

$$\text{Hence, } P(X + Y \geq 1) = 1 - P(X + Y < 1) = 2e^{-1}.$$

c. Find marginal pdf of X and Y .

Joint support is $A = \{(x, y) | 0 < x < y < \infty\}$. The support of X is $A_X = \{0 < x < \infty\}$.

$$\text{Given } x \in (0, \infty), f_1(x) = \int_{-\infty}^{\infty} f(x, y) dy = \int_x^{\infty} 2e^{-x-y} dy = 2e^{-x} (-e^{-y}) \Big|_x^{\infty} = 2e^{-2x}.$$

The support of Y is $A_Y = \{0 < y < \infty\}$.

$$\text{Given } y \in (0, \infty), f_2(y) = \int_{-\infty}^{\infty} f(x, y) dx = \int_0^y 2e^{-x-y} dx = 2e^{-y} (-e^{-x}) \Big|_0^y = 2e^{-y} - 2e^{-2y}.$$

d. Find the distribution of $T = X + Y$.

Solution: The support of T is $A_T = \{0 < t < \infty\}$.

a. If $t \leq 0$, $P(T \leq t) = 0$.

b. If $t > 0$, $F_T(t) = P(T \leq t) = P(X + Y \leq t) = \iint_{(x,y) \in C} 2e^{-x-y} dx dy = \int_0^{t/2} \int_x^{t-x} 2e^{-x-y} dy dx = \int_0^{t/2} (-2e^{-x}e^{-y})|_x^{t-x} = -e^{-2x} - 2e^{-t}x|_0^{t/2} = 1 - e^{-t} - te^{-t}$.
The pdf of T is $f_T(t) = \frac{d}{dt}F_T(t) = e^{-t} + te^{-t} = e^{-t} - e^{-t} + te^{-t} = te^{-t}$ for $t > 0$ and 0 otherwise.

3.4 Independent of random variables

- Definition: For any two r.v.s X and Y , we say X and Y are independent if and only if $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$ for any $A, B \subseteq \mathbb{R}$.

Here, $X \in A$ is an event, meaning $\{\omega \in \Omega : X(\omega) \in A\}$.

e.g. Let $A = (-\infty, x)$, $B = (-\infty, y)$, $x, y \in \mathbb{R}$.

Therefore, if X and Y are independent, $P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y) = F_1(x)F_2(y)$ for any $x, y \in \mathbb{R}$.

Conclusion: X and Y are independent if and only if $F(x, y) = F_1(x)F_2(y)$ for any $x, y \in \mathbb{R}$. (Above shows this is a necessary condition, proof of this is a sufficient condition is beyond the scope of this course.)

Suppose X and Y has joint p.f. or joint p.d.f, which is denoted by $f(x, y)$, and marginal p.f. or marginal p.d.f, denoted by $f_1(x)$ and $f_2(y)$, then X and Y are independent iff $f(x, y) = f_1(x)f_2(y)$ for every $x, y \in \mathbb{R}$.

Remark: If X and Y are independent, then $g(X)$ and $h(Y)$ must be independent for any real functions g and h .

e.g. If X is independent of Y , then X^2 is independent of Y^2 . But X^2 is independent of Y^2 , we cannot conclude X is independent of Y .

Suppose $P(X = 1) = P(X = -1) = \frac{1}{2}$. Let $Y = X$. $P(X = 1, Y = 1) = P(X = 1) = \frac{1}{2}$, but $P(X = 1)P(Y = 1) = \frac{1}{4}$.
 $P(Y^2 = 1) = P(X^2 = 1) = 1$.

Example: (Joint Discrete r.v.s)

Consider the joint p.f. of X and Y is $f(x, y) = q^2 p^{x+y}$ for $x = 0, 1, \dots$ and $y = 0, 1, \dots$, and 0 elsewhere. Here $p \in (0, 1)$ is a constant, $q = 1 - p$.

Marginal p.f. of X is $f_1(x) = qp^x$ for $x = 0, 1, \dots$ and 0 elsewhere.

Marginal p.f. of Y is $f_2(y) = qp^y$ for $y = 0, 1, \dots$ and 0 elsewhere.

Thus, $f(x, y) = f_1(x)f_2(y)$ for every $x, y \in \mathbb{R}$ therefore, X and Y are independent.

Example (Joint Continuous r.v.s)

Suppose the joint pdf of X and Y is $f(x, y) = \begin{cases} x + y & \text{if } 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{o.w.} \end{cases}$.

The marginal pdf of X is $f_1(x) = x + \frac{1}{2}$ for $x \in [0, 1]$ and 0 otherwise.

The marginal pdf of Y is $f_2(y) = y + \frac{1}{2}$ for $y \in [0, 1]$ and 0 otherwise.

Hence, $f(x, y) \neq f_1(x)f_2(y)$ for $x \in (0, 1)$ and $y \in (0, 1)$, therefore, X and Y are not independent.

- Factorization theorem for independence

Condition 1: $f(x, y) = g(x)h(y)$ for every $x, y \in \mathbb{R}$ for some function g and h where $f(x, y)$ denotes the joint p.f. or joint p.d.f. of X and Y .

Condition 2: Let A be the joint support of X and Y , and let A_1 be the marginal support of X and A_2 be the marginal support of Y . Then, $A = A_1 \times A_2 = \{(x, y) \in \mathbb{R}^2 : x \in A_1, y \in A_2\}$. (Interpretation: A is a rectangle or the range of X and Y are independent.)

Conditions 1 and 2 are satisfied if and only if X and Y are independent.

Example: If the joint p.f. of X and Y is $f(x, y) = \frac{\mu^{x+y} e^{-2\mu}}{x!y!}$ for $x = 0, 1, \dots$ and $y = 0, 1, \dots$ and 0 elsewhere.

i. Is X independent of Y ?

Solution: Condition 1: $f(x, y) = \frac{\mu^{x+y} e^{-2\mu}}{x!y!} = \frac{\mu^x e^{-\mu}}{x!} \frac{\mu^y e^{-\mu}}{y!}$. If we take $g(x) = \begin{cases} \frac{\mu^x e^{-\mu}}{x!} & \text{if } x = 0, 1, \dots \\ 0 & \text{o.w.} \end{cases}$ and $h(y) =$

$\begin{cases} \frac{\mu^y e^{-\mu}}{y!} & \text{if } y = 0, 1, \dots \\ 0 & \text{o.w.} \end{cases}$, then $f(x, y) = g(x)h(y)$ for every $x, y \in \mathbb{R}$.

Condition 2: $A = \{(x, y) \in \mathbb{R}^2 : x \in A_1, y \in A_2\}$, where $A_1 = \{0, 1, \dots\}$ and $A_2 = \{0, 1, \dots\}$.

Therefore, by factorization theorem, X and Y are independent.

ii. Find the marginal p.f. of X and Y .

Solution: A shortcut: $f_1(x) = C \cdot g(x)$ for some constant C .

Property 1: $f_1(x) \geq 0$ for any $x \in \mathbb{R}$. Here $g(x) = \begin{cases} \frac{\mu^x e^{-\mu}}{x!} & \text{if } x = 0, 1, \dots \\ 0 & \text{o.w.} \end{cases}$, therefore, $C \geq 0$.

Property 2: The support of X is $A_1 = \{0, 1, \dots\}$. Therefore, $\sum_0^\infty f_1(x) = \sum_0^\infty C \frac{\mu^x e^{-\mu}}{x!} = C \sum_0^\infty \frac{\mu^x e^{-\mu}}{x!}$, then $C = 1$.

Therefore, $f_1(x) = \begin{cases} \frac{\mu^x e^{-\mu}}{x!} & \text{if } x = 0, 1, \dots \\ 0 & \text{o.w.} \end{cases}$.

Similarly, $f_2(y) = \begin{cases} \frac{\mu^y e^{-\mu}}{y!} & \text{if } y = 0, 1, \dots \\ 0 & \text{o.w.} \end{cases}$.

Example (Joint Continuous r.v.s)

Suppose the joint pdf of X and Y is $f(x, y) = \begin{cases} \frac{3}{2}y(1-x^2) & -1 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{o.w.} \end{cases}$.

i. Is X independent of Y ?

Solution: Condition 1: $f(x, y) = (\frac{3}{2}y)(1-x^2)$, then $g(x) = \begin{cases} 1-x^2 & \text{if } -1 \leq x \leq 1 \\ 0 & \text{o.w.} \end{cases}$ and $h(y) = \begin{cases} \frac{3}{2}y & \text{if } 0 \leq y \leq 1 \\ 0 & \text{o.w.} \end{cases}$.

Then $f(x, y) = g(x)h(y)$ for every $x, y \in \mathbb{R}$.

Condition 2: $A = \{(x, y) \in \mathbb{R}^2 : x \in A_1, y \in A_2\}$, where $A_1 = [-1, 1]$ and $A_2 = [0, 1]$.

Therefore, by factorization theorem, X and Y are independent.

ii. Find the marginal pdf of X and Y .

Solution: A shortcut: $f_1(x) = C \cdot g(x)$ for some constant C , the support of X is $A_1 = [-1, 1]$.

Property 1: $f_1(x) \geq 0$ for any $x \in \mathbb{R}$. Here $g(x) = \begin{cases} 1-x^2 & \text{if } -1 \leq x \leq 1 \\ 0 & \text{o.w.} \end{cases}$, therefore, $C \geq 0$.

Property 2: $\int_{-\infty}^{\infty} f_1(x)dx = \int_{-1}^1 C(1-x^2)dx = C \left(x - \frac{x^3}{3} \right) \Big|_{-1}^1 = 2C \left(1 - \frac{1}{3} \right) = 1$, therefore, $C = \frac{3}{4}$.

Therefore, $f_1(x) = \begin{cases} \frac{3}{4}(1-x^2) & \text{if } -1 \leq x \leq 1 \\ 0 & \text{o.w.} \end{cases}$.

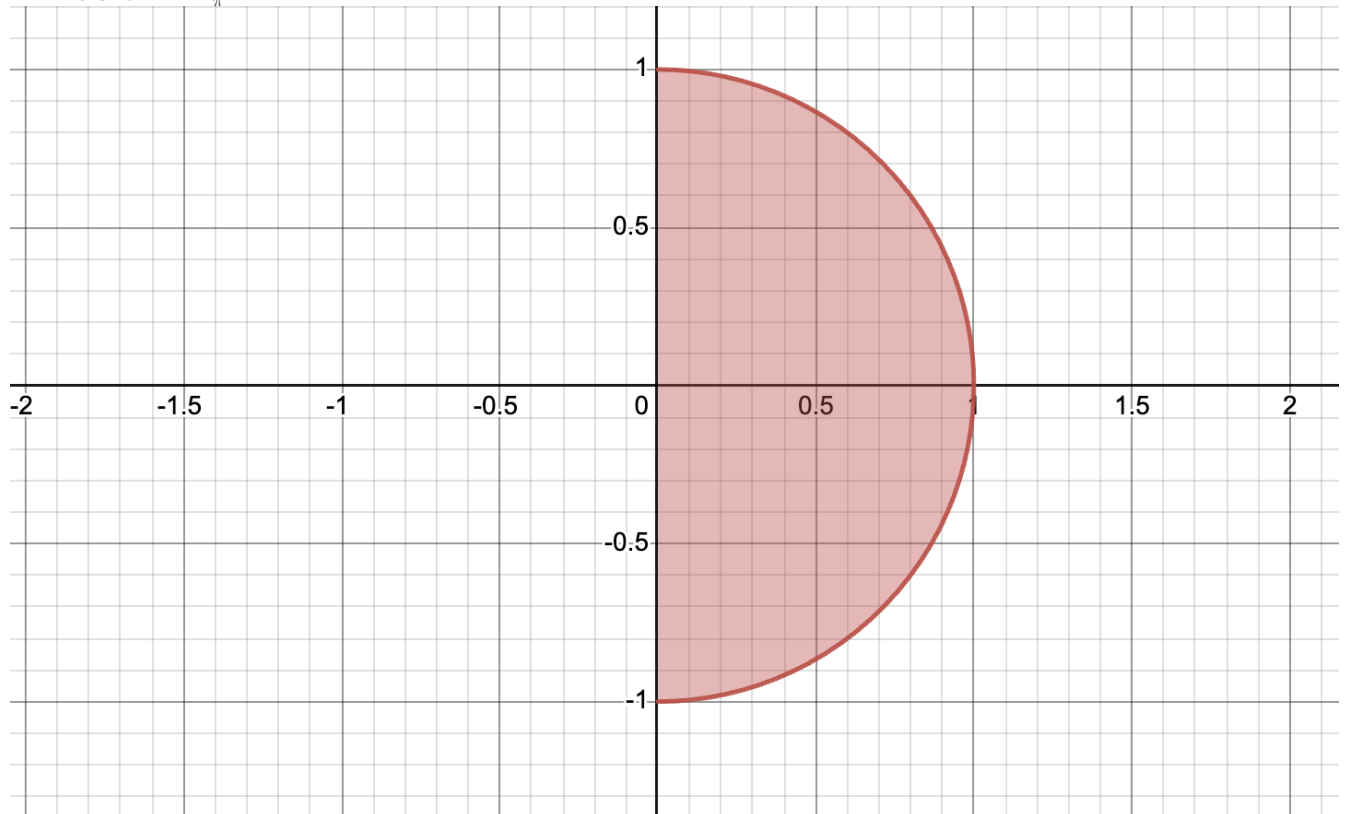
Support of Y is $A_2 = [0, 1]$, given $y \in [0, 1]$, $f_2(y) = \frac{f(x, y)}{f_1(x)} = \frac{\frac{3}{2}y(1-x^2)}{\frac{3}{4}(1-x^2)} = 2y$. Therefore, $f_2(y) = \begin{cases} 2y & \text{if } 0 \leq y \leq 1 \\ 0 & \text{o.w.} \end{cases}$.

Example (Uniform distribution over a region)

Suppose (X, Y) follows a uniform distribution over $C = \{(x, y) | x \geq 0, x^2 + y^2 \leq 1\}$.

Namely, $f(x, y) = \begin{cases} c & \text{if } (x, y) \in C \\ 0 & \text{o.w.} \end{cases}$.

Here, by graph, $c = \frac{2}{\pi}$.



i. Is X independent of Y ?

Solution: Given $x \in [0, 1]$, Y can take value in $[-\sqrt{1-x^2}, \sqrt{1-x^2}]$, therefore, X and Y are not independent.

ii. Find the marginal pdf of X and Y .

Solution: The support of X is $A_1 = [0, 1]$, given $x \in [0, 1]$, $f_1(x) = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{2}{\pi} dy = \frac{4}{\pi} \sqrt{1-x^2}$.

The support of Y is $A_2 = [-1, 1]$, given $y \in [-1, 1]$, $f_2(y) = \int_0^{\sqrt{1-y^2}} \frac{2}{\pi} dx = \frac{2}{\pi} \sqrt{1-y^2}$.

3.5 Joint expectation

- Definition: Suppose $h(x, y)$ is a bivariate function, then $E[h(x, y)] = \begin{cases} \sum_x \sum_y h(x, y) f(x, y) & \text{if } X \text{ and } Y \text{ are joint discrete} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y) f(x, y) dx dy & \text{if } X \text{ and } Y \text{ are joint continuous} \end{cases}$, provided $E[|h(x, y)|] < \infty$.

$$\text{e.g. } E[XY] = \begin{cases} \sum_x \sum_y (xy) f(x, y) & \text{if } X \text{ and } Y \text{ are joint discrete} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (xy) f(x, y) dx dy & \text{if } X \text{ and } Y \text{ are joint continuous} \end{cases}, \text{ provided } E[|XY|] < \infty.$$

e.g. $E[X]$ (i.e. $h(x, y) = x$)

- Method 1:

$$E(X) = \begin{cases} \sum_x \sum_y x f(x, y) & \text{joint discrete} \\ \iint_{\mathbb{R}^2} x f(x, y) dx dy & \text{joint continuous} \end{cases}$$

- Method 2: find the marginal distribution, i.e., the marginal p.f. or marginal p.d.f. of X first, denoted by $f_1(x)$, then

$$E(X) = \begin{cases} \sum_x x f_1(x) & \text{joint discrete} \\ \int_{\mathbb{R}^2} x f_1(x) dx & \text{joint continuous} \end{cases}$$

- Properties of joint expectation:

i. linearity: $E[ag(X, Y) + bh(X, Y)] = aE[g(X, Y)] + bE[h(X, Y)]$ where a, b are constants, g, h are bivariate functions.

ii. Under independence assumption (X is independent of Y), $E(XY) = E(X)E(Y)$ and $E[g(X)h(Y)] = E[g(X)]E[h(Y)]$. Further, if X_1, \dots, X_n are independent, then $E[\prod_{i=1}^n h_i(X_i)] = \prod_{i=1}^n E[h_i(X_i)]$.

- Covariance of X and Y

Definition: Covariance of X and Y is defined as $Cov(X, Y) = E[(X - E(X))(Y - E(Y))] = E(XY) - E(X)E(Y)$.

If X and Y are independent, then $Cov(X, Y) = 0$.

An example where X and Y are uncorrelated, but not independent.

Let $X \sim N(0, 1)$ and $Y = X^2$, then $E(X) = 0$, $E(XY) = E(X^3)$, $Cov(X, Y) = 0$.

Now, we find a pair of a and b such that $P(X \leq a, Y \leq b) \neq P(X \leq a)P(Y \leq b)$. Consider $a = -2, b = 1$, then $P(X \leq a) = P(X \leq -2) > 0$, $P(Y \leq b) = P(X^2 \leq 1) = P(-1 \leq X \leq 1) > 0$, but $P(X \leq a, Y \leq b) = P(X \leq -2, Y \leq 1) = 0$.

- Results for covariance

i. $Cov(X, X) = E[(X - \mu_X)(X - \mu_X)] = E[(X - \mu_X)^2] = Var(X)$.

ii. $Cov(X + Y, Z) = Cov(X, Z) + Cov(Y, Z)$.

- Variance formula

$$Var(aX + bY) = Cov(aX + bY, aX + bY)$$

$$\text{i. } Cov(aX, aX) + Cov(aX, bY) + Cov(bY, aX) + Cov(bY, bY) = Var(aX) + 2abCov(X, Y) + Var(bY) = a^2Var(X) + 2abCov$$

$$\text{ii. } Var\left(\sum_{i=1}^n\right) = \sum_{i=1}^n Var(X_i) + 2 \sum_{i < j} Cov(X_i, X_j)$$

iii. If X_1, \dots, X_n are independent,

$$Var\left(\sum_{i=1}^n\right) = \sum_{i=1}^n Var(X_i)$$

Example 1: Suppose the joint p.f. of X and Y is $f(x, y) = \begin{cases} \frac{\mu^{x+y} e^{-2\mu}}{x!y!} & \text{if } x = 0, 1, \dots \text{ and } y = 0, 1, \dots \\ 0 & \text{o.w.} \end{cases}$. Find $Var(2X + 3Y) =$

$$4Var(X) + 12Cov(X, Y) + 9Var(Y).$$

Solution: Since X and Y are independent, $Cov(X, Y) = 0$, therefore, $Var(2X + 3Y) = 4Var(X) + 9Var(Y)$.
 Previously, we find $X \sim Poisson(\mu)$, $Y \sim Poisson(\mu)$, therefore $Var(X) = \mu$, $Var(Y) = \mu$.
 Hence, $Var(2X + 3Y) = 4\mu + 9\mu = 13\mu$.

Example 2: Suppose the joint p.f. of X and Y is $f(x, y) = \begin{cases} x + y & \text{if } 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{o.w.} \end{cases}$. Find $Var(X + Y)$.

Solution:

$$\begin{aligned} Var(X + Y) &= Var(X) + 2Cov(X, Y) + Var(Y) \\ &= 2Var(X) + 2Cov(X, Y) \end{aligned}$$

the marginal pdf of X is $f_1(x) = \begin{cases} x + 1/2 & \text{if } 0 \leq x \leq 1 \\ 0 & \text{o.w.} \end{cases}$.

$$\text{then, } E(X) = \int_0^1 x \left(x + \frac{1}{2}\right) dx = \int_0^1 \left(x^2 + \frac{x}{2}\right) dx = \left(\frac{x^3}{3} + \frac{x^2}{4}\right) \Big|_0^1 = \frac{7}{12}.$$

$$E(X^2) = \int_0^1 x^2 \left(x + \frac{1}{2}\right) dx = \int_0^1 \left(x^3 + \frac{x^2}{2}\right) dx = \left(\frac{x^4}{4} + \frac{x^3}{6}\right) \Big|_0^1 = \frac{5}{12}.$$

$$Var(X) = E(X^2) - (E(X))^2 = \frac{5}{12} - \left(\frac{7}{12}\right)^2 = \frac{11}{144}.$$

$$Cov(X, Y) = E(XY) - E(X)E(Y), \text{ where } E(X)E(Y) = \left(\frac{7}{12}\right)^2 = \frac{49}{144}.$$

$$\begin{aligned} E(XY) &= \int_0^1 \int_0^1 (xy)(x + y) dx dy \\ &= \int_0^1 \int_0^1 (x^2y + xy^2) dx dy \\ &= \int_0^1 \left(\frac{x^3y}{3} + \frac{x^2y^2}{2}\right) \Big|_{x=0}^{x=1} dy \\ &= \int_0^1 \left(\frac{y}{3} + \frac{y^2}{2}\right) dy \\ &= \left(\frac{y^2}{6} + \frac{y^3}{6}\right) \Big|_{y=0}^{y=1} \\ &= \frac{1}{3} \end{aligned}$$

$$Cov(X, Y) = 1/3 - 49/144 = -1/144.$$

$$Var(X + Y) = 2Var(X) + 2Cov(X, Y) = 2\frac{11}{144} + 2\frac{-1}{144} = \frac{20}{144}.$$

- Correlation coefficient

Definition: Correlation coefficient of X and Y is defined as $\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}$.

i. Used to describe linear association between X and Y .

ii. Unit free

iii. $-1 \leq \rho(X, Y) \leq 1$.

(not required): Use the fact $|E(XY)| \leq \sqrt{E(X^2)}\sqrt{E(Y^2)}$ to prove $-1 \leq \rho(X, Y) \leq 1$.

- Properties of correlation coefficient:

i. $\rho(X, Y) = 1 \implies Y = aX + b$ for some constants $a > 0$ and b .

ii. $\rho(X, Y) = -1 \implies Y = aX + b$ for some constants $a < 0$ and b .