Idea of equally likely outcomes : Flip a coin 3 times.

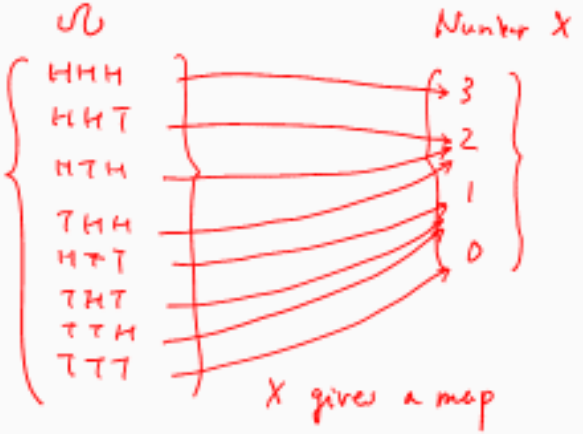
Ω = { HHH, HHT, HTH, THH, HTT, THT, TTH, TTT }

Random Variables

In probabilistic models, a random variable is a variable whose possible values are numerical outcomes of a random phenomenon. As a function or a map, it maps from an element (or an outcome) of a sample space to real values. The underlying probability over sample space makes it possible to define its cumulative distribution, which allows us to calculate quantities such as its expected value and variance, and the moments of its distribution.

X = # of heads

X = 3, 2, 1, 0



{X = 3} = { HHH }

{X = 2} = { HHT, HTH, THH }

Random variables.

A numerically valued map X of an outcome ω from a sample space Ω to the real line $\mathbb{R}$

$$X : \Omega \to \mathbb{R} : \omega \to X(\omega)$$

is called a **random variable (r.v.)**, and usually determined by an experiment. We conventionally denote random variables by uppercase letters X, Y, Z, U, V, etc., from the end-of-alphabet letters. In particular, a **discrete random variable** is a random variable that can take values on a finite set $\{a_1, a_2, \ldots, a_n\}$ of real numbers (usually integers), or on a countably infinite set $\{a_1, a_2, a_3, \ldots\}$. The statement such as "$X = a_i$" is an event since

$$\{\omega : X(\omega) = a_i\}$$

is a subset of a sample space Ω.

Frequency function.

$X: \Omega \to \mathbb{R}$: random variable $\Rightarrow$ $p: \{a_1, a_2, \vdots, a_n\} \to [0,1]$: frequency function : $a_i \to p(a_i)$

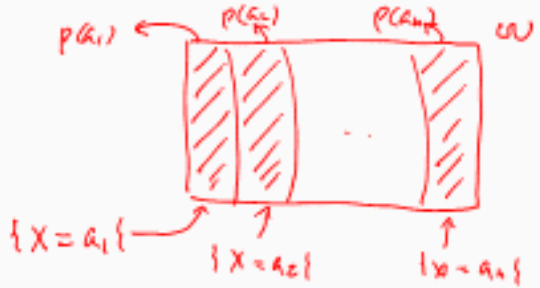
We can consider the probability of the event $\{X = a_i\}$, denoted by $P(X = a_i)$. The function

$$p(a_i) := P(X = a_i) = P(\{\omega : X(\omega) = a_i\})$$

over the possible values of X, say $a_1, a_2, \ldots$, is called a **frequency function**, or a **probability mass function**. The frequency function p must satisfy

$$\sum_i p(a_i) = 1,$$

where the sum is over the possible values of X. The frequency function will completely describe the probabilistic nature of random variable.



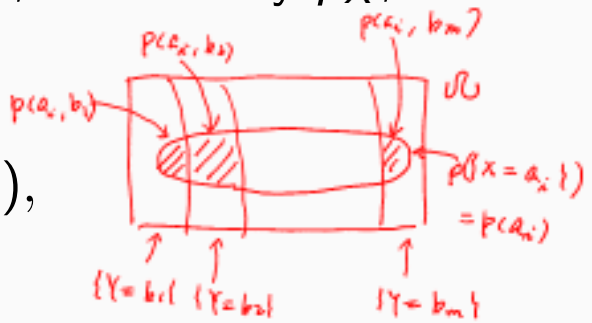
Joint distributions of discrete random variables.

When two discrete random variables X and Y are obtained in the same experiment, we can define their **joint frequency function** by

$$p(a_i, b_j) = P(X = a_i, Y = b_j)$$

where $a_1, a_2, \ldots$ and $b_1, b_2, \ldots$ are the possible values of X and Y, respectively. The **marginal frequency function** of X, denoted by p_X, can be calculated by

$$p_X(a_i) = P(X = a_i) = \sum_j p(a_i, b_j),$$



where the sum is over the possible values $b_1, b_2, \ldots$ of Y. Similarly, the marginal frequency function $p_Y(b_j) = \sum_i p(a_i, b_j)$ of Y is given by summing over the possible values $a_1, a_2, \ldots$ of X.

Example

An experiment consists of throwing a fair coin three times. Let X be the number of heads, and let Y be the number of heads before the first tail.

1. List the sample space Ω. $= \{HHH, HHT, \cdots, TTT\}$
2. Describe the events $\{X = 0\}$, $\{Y = 0\}$, and $\{X = 0, Y = 0\}$.
3. Find the frequency function p for X and Y. And compute the joint frequency $p(0, 0)$. $= P(\{TTT\}) = \frac{1}{8}$

2. $\{X=0\} = \{TTT\}$, $\{Y=0\} = \{TTT, THH, THT, TTH\}$, $\{X=0, Y=0\} = \{TTT\}$

$\{Y=1\} = \{HTT, HTH\}$

$\{Y=2\} = \{HHT\}$

$\{Y=3\} = \{HHH\}$

3. $P_X(0) = \frac{1}{8}$, $P_X(1) = \frac{3}{8}$, $P_X(2) = \frac{3}{8}$, $P_X(3) = \frac{1}{8}$

$P_Y(0) = \frac{1}{2}$, $P_Y(1) = \frac{1}{4}$, $P_Y(2) = \frac{1}{8}$, $P_Y(3) \frac{1}{8}$

Example

An experiment consists of throwing a fair coin three times. Let X be the number of heads, and let Y be the number of heads before the first tail.

1. List the sample space Ω.
2. Describe the events $\{X = 0\}$, $\{Y = 0\}$, and $\{X = 0, Y = 0\}$.
3. Find the frequency function p for X and Y. And compute the joint frequency $p(0, 0)$.

1. $\Omega = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$
2. $\{X = 0\} = \{TTT\}$, $\{Y = 0\} = \{TTT, THH, THT, TTH\}$, and $\{X = 0, Y = 0\} = \{TTT\}$,
3. $p_X(0) = \frac{1}{8}$; $p_X(1) = \frac{3}{8}$; $p_X(2) = \frac{3}{8}$; $p_X(3) = \frac{1}{8}$.
   $p_y(0) = \frac{1}{2}$; $p_y(1) = \frac{1}{4}$; $p_y(2) = \frac{1}{8}$; $p_y(3) = \frac{1}{8}$.
   $p(0, 0) = P(X = 0, Y = 0) = \frac{1}{8}$.

Cumulative distribution function.

Frequency function $p(a_i) = P(X = a_i) \longrightarrow X: \Omega \to \{a_1, a_2, \ldots, a_n\} \to p: \{a_1, \ldots, a_n\} \to [0,1]$

Example: X = # of heads

$p(0) = \frac{1}{8}$, $p(1) = \frac{3}{8}$, $p(2) = \frac{3}{8}$, $p(3) = \frac{1}{8}$

$\Rightarrow F(1.5) = P(X \le 1.5)$

$= P(\{TTT, TTH, THT, HTT\}) = \frac{1}{2}$

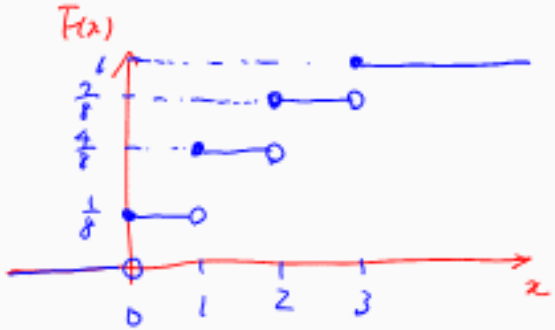
Another useful function is the **cumulative distribution function (cdf)**, and it is defined by

$F: (-\infty, \infty) \to [0,1]$

$$F(x) = P(X \le x), \qquad -\infty < x < \infty.$$

$= P(\{\omega : X(\omega) \le x\})$

The cdf of a discrete r.v. is a nondecreasing step function. It jumps wherever $p(x) > 0$, and the jump at a_i is $p(a_i)$. cdf's are usually denoted by uppercase letters, while frequency functions are usually denoted by lowercase letters.



Independent random variables.

Recall: Events A and B are independent if $P(A \cap B) = P(A)\,P(B)$

Random variables X and Y are independent if $\{X = a_i\}$ and $\{Y = b_j\}$ are independent for outcomes a_i and b_j for X and Y

$$\Updownarrow$$

$$p(a_i, b_j) = P(\{X = a_i\} \cap \{Y = b_j\}) = P(\{X = a_i\})\,P(\{Y = b_j\}) = p_X(a_i)\,p_Y(b_j)$$

Let X and Y be discrete random variables with joint frequency function $p(x, y)$. Then X and Y are said to be **independent**, if they satisfy

$$p(x, y) = p_X(x) p_Y(y)$$

for all possible values of (x, y).

Example

Continue the same experiment of throwing a fair coin three times. Let X be the number of heads, and let Y be the number of heads before the first tail.

$P(\{X \le x\})$ // — random variable; variable

1. Find the cdf $F(x)$ for X at $x = -1, 0, 1, 2, 2.5, 3, 4.5$.
2. Are X and Y independent?

1. $F(-1) = P(\{X \le -1\}) = P(\emptyset) = 0$ — Empty set

   $F(2.5) = P(\{X \le 2.5\}) = P(\{TTT, TTH, THT, HTT, THH, HTH, HHT\}) = \frac{7}{8}$

2. $P(0,0) = \frac{1}{8} \neq P_X(0)\,P_Y(0) = \left(\frac{1}{8}\right)\left(\frac{1}{2}\right) = \frac{1}{16} \Rightarrow X$ and Y are not independent.

Example

Continue the same experiment of throwing a fair coin three times. Let X be the number of heads, and let Y be the number of heads before the first tail.

1. Find the cdf $F(x)$ for X at $x = -1, 0, 1, 2, 2.5, 3, 4.5$.
2. Are X and Y independent?

1. $F(-1) = 0$, $F(0) = \frac{1}{8}$, $F(1) = \frac{1}{2}$, $F(2) = F(2.5) = \frac{7}{8}$, and $F(3) = F(4.5) = 1$.
2. Since $p_X(0) = \frac{1}{8}$, $p_Y(0) = \frac{1}{2}$, and $p(0,0) = \frac{1}{8}$, we find that $p(0,0) \neq p_X(0)p_Y(0)$. Thus, X and Y are not independent.

Continuous random variables.

Question: $P(X = 5.62) = 0 \Rightarrow$ We cannot use frequency function to determine a distribution when X is continuous.

No frequency function !!

A **continuous random variable** is a random variable whose possible values are real values such as 78.6, 5.7, 10.24, and so on. Examples of continuous random variables include temperature, height, diameter of metal cylinder, etc. In what follows, a random variable means a "continuous" random variable, unless it is specifically said to be discrete. The **probability distribution** of a random variable X specifies how its values are distributed over the real numbers. This is completely characterized by the **cumulative distribution function (cdf)**. The cdf

$P(X = t) = 0$ does not carry any information

$$F(t) := P(X \leq t).$$

represents the probability that the random variable X is less than or equal to t. Then we say that "the random variable X is distributed as $F(t)$."

Properties of cdf.

The cdf $F(t)$ must satisfy the following properties:

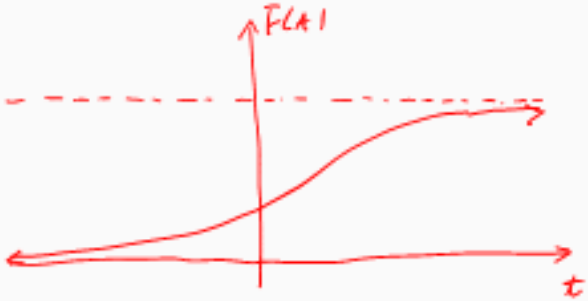
1. The cdf $F(t)$ is a positive function.

$$F(t) \geq 0 \text{ for all } t.$$

2. The cdf $F(t)$ increases monotonically.

$$F(s) \leq F(t) \text{ whenever } s < t.$$

3. The cdf $F(t)$ must tend to zero as t goes to negative infinity "$-\infty$", and must tend to one as t goes to positive infinity "$+\infty$."



$$\lim_{t \to -\infty} F(t) = 0$$

$$\lim_{t \to +\infty} F(t) = 1$$

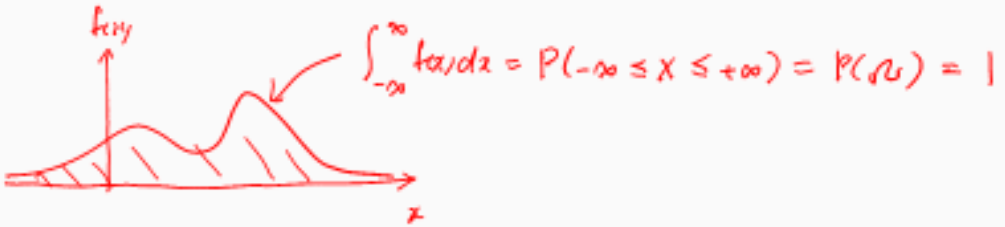
Probability density function (pdf).

It is often the case that the probability that the random variable X takes a value in a particular range is given by the area under a curve over that range of values. This curve is called the **probability density function (pdf)** of the random variable X, denoted by $f(x)$. Thus, the probability that "$a \leq X \leq b$" can be expressed by

$$P(a \leq X \leq b) = \int_a^b f(x)dx.$$

The pdf $f(x)$ must satisfy the following properties:

1. The pdf $f(x)$ is a positive function [that is, $f(x) \geq 0$].
2. The area under the curve of $f(x)$ (and above the x-axis) is one [that is, $\int_{-\infty}^{\infty} f(x)dx = 1$].



Relation between cdf and pdf.

The pdf $f(x)$ is related to the cdf $F(t)$ via

$$F(t) = P(X \leq t) = \int_{-\infty}^{t} f(x)dx.$$

This implies that such cdf $F(t)$ is a continuous function, and that the pdf $f(x)$ is the derivative of the cdf $F(x)$ if f is continuous at x, that is,

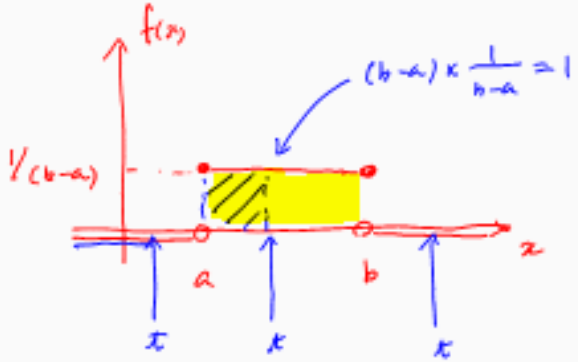
$$f(x) = \frac{dF(x)}{dx}.$$

Example

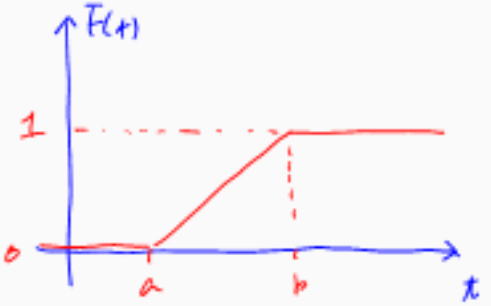
A random variable X is called a **uniform** random variable on $[a, b]$, when X takes any real number in the interval $[a, b]$ equally likely. Then the pdf of X is given by

$$f(x) = \begin{cases} 1/(b-a) & \text{if } a \le x \le b; \\ 0 & \text{otherwise [that is, if } x < a \text{ or } b < x]. \end{cases}$$

Find the cdf $F(t)$.



$$F(t) = \int_{-\infty}^{t} f(x)\,dx = \begin{cases} 0 & \text{if } x < a \\ \dfrac{t-a}{b-a} & \text{if } a \le x \le b \\ 1 & \text{if } x > b \end{cases}$$



Example

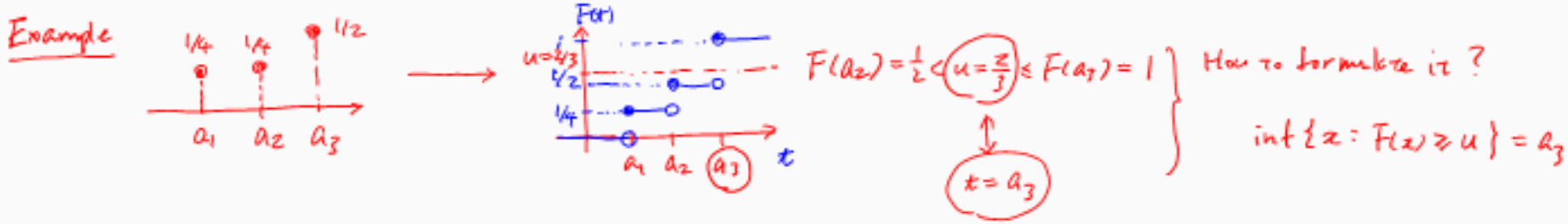
A random variable X is called a **uniform** random variable on $[a, b]$, when X takes any real number in the interval $[a, b]$ equally likely. Then the pdf of X is given by

$$f(x) = \begin{cases} 1/(b-a) & \text{if } a \leq x \leq b; \\ 0 & \text{otherwise [that is, if } x < a \text{ or } b < x]. \end{cases}$$

Find the cdf $F(t)$.

$$F(t) = \int_{-\infty}^{t} f(x)dx = \begin{cases} 0 & \text{if } t < a; \\ \frac{t-a}{b-a} & \text{if } a \leq t \leq b; \\ 1 & \text{if } t > b. \end{cases}$$

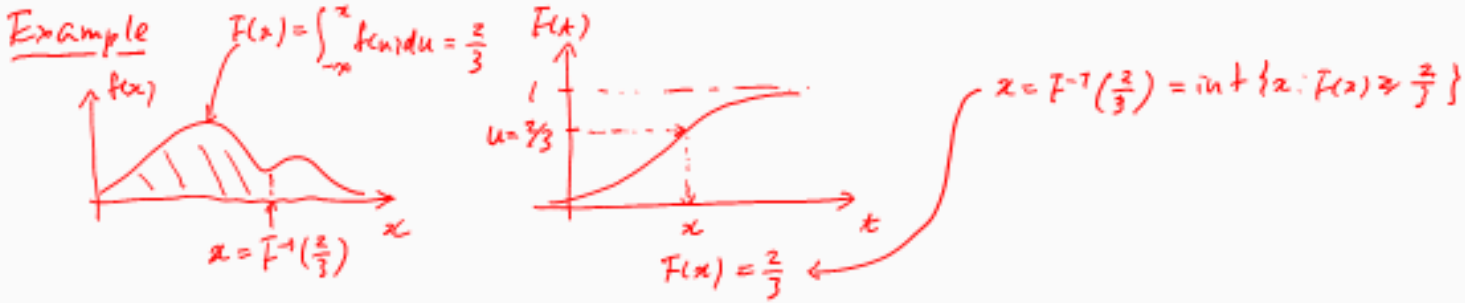
Quantile function.



Given a cdf $F(x)$, we can define the *quantile function* $Q(u) = \inf\{x : F(x) \geq u\}$ for $u \in (0, 1)$. In particular, if F is continuous and strictly increasing, then we have $Q(u) = F^{-1}(u)$. When $F(x)$ is the cdf of a random variable X, and $F(x)$ is continuous, we can find

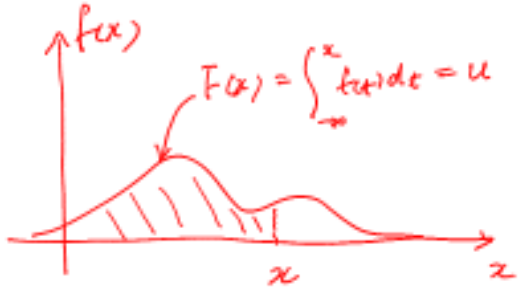
$$x_p = Q(p) \quad \Leftrightarrow \quad P(X \leq x_p) = p$$

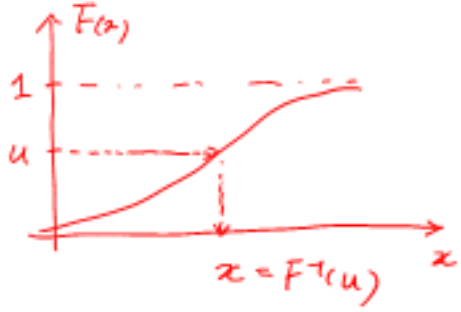
for every $p \in (0, 1)$. The value $x_{1/2} = Q(\frac{1}{2})$ is called the *median* of F, and the values $x_{1/4} = Q(\frac{1}{4})$ and $x_{3/4} = Q(\frac{3}{4})$ are called the *lower quartile* and the *upper quartile*, respectively.

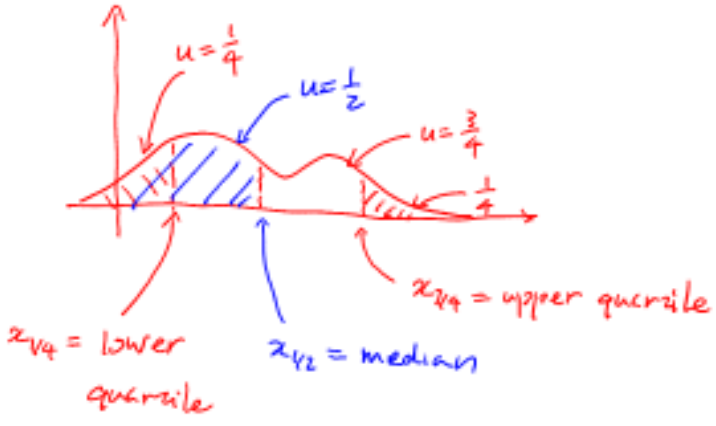


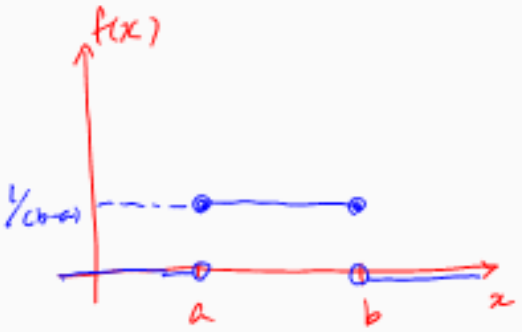
Quantile function for continuous random variable

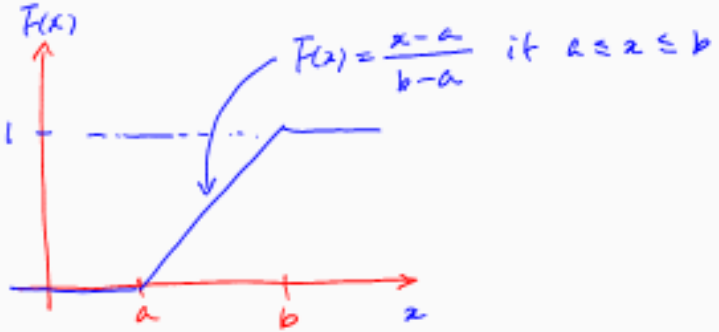
$$Q(u) = \inf\{x : F(x) \geq u\} = F^{-1}(u) \quad \text{if } F(x) \text{ is strictly increasing}$$

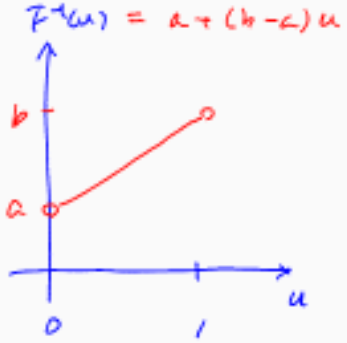












Example

Compute the quantile function $Q(u)$ for a uniform random variable X on $[a, b]$. Find the median of distribution.

Between a and b

$$Q(u) = F^{-1}(u) \quad \text{for } 0 < u < 1$$

How to find the inverse $x = F^{-1}(u)$?

$$u = F(x) = \frac{x-a}{b-a}$$

$$\Downarrow$$

$$u = \frac{x-a}{b-a}$$

$$(b-a)u = x - a$$

$$\Downarrow$$

$$x = a + (b-a)u = F^{-1}(u)$$

Example

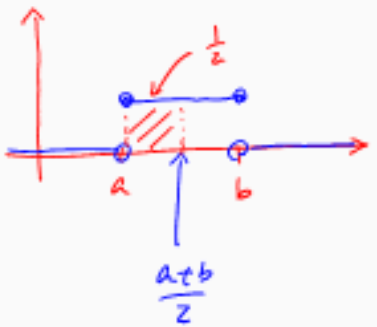
Compute the quantile function $Q(u)$ for a uniform random variable X on $[a, b]$. Find the median of distribution.

Restricting the domain $[a, b]$ for F, we find that $F(x) = \frac{x-a}{b-a}$ is strictly increasing function from (a, b) to $(0, 1)$, and therefore, that $F(x)$ has the inverse function $Q(u)$. By solving $u = \frac{x-a}{b-a}$ in terms of x, we obtain $Q(u) = a + (b-a)u$ for $0 < u < 1$. The median is given by $Q\left(\frac{1}{2}\right) = \frac{a+b}{2}$.

$\Downarrow$

$Q(\frac{1}{2}) = a + (b-a)(\frac{1}{2})$

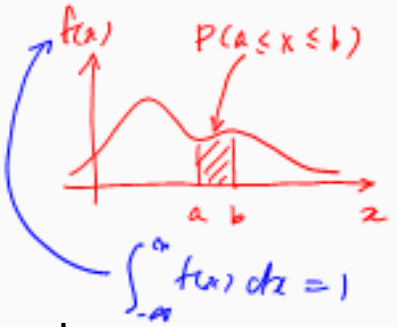
$= \frac{1}{2}a + \frac{1}{2}b = \frac{a+b}{2}$



Joint density function.

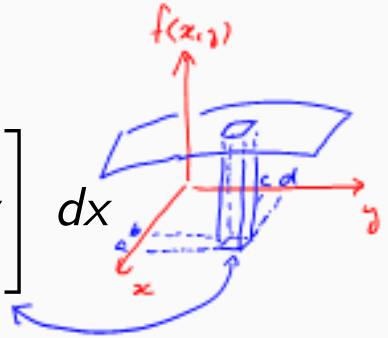
Consider a pair (X, Y) of random variables. A **joint density function** $f(x, y)$ is a nonnegative function satisfying

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



and is used to compute probabilities constructed from the random variables X and Y simultaneously by

$$P(a \le X \le b, c \le Y \le d) = \int_a^b \left[\int_c^d f(x,y)\,dy\right] dx$$



In particular,

$$F(u,v) = P(X \le u,\ Y \le v) = \int_{-\infty}^{u}\int_{-\infty}^{v} f(x,y)\,dy\,dx.$$

is called the **joint distribution function**.

Marginal densities and independence.

Recall marginal frequency function: $p(x,y) \longrightarrow p_X(x) = \sum_y p(x,y)$

Discrete case: X and Y are independent if $p(x,y) = p_X(x)\, p_Y(y)$

Given the joint density function $f(x,y)$, the distribution for each of X and Y is called the **marginal distribution**, and the **marginal density functions** of X and Y, denoted by $f_X(x)$ and $f_Y(y)$, are given respectively by

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy \quad \text{and} \quad f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx.$$

If the joint density function $f(x,y)$ for continuous random variables X and Y is expressed in the form

$$f(x,y) = f_X(x) f_Y(y) \quad \text{for all } x, y,$$

then X and Y are said to be independent.

$f_X(x)$

It must be a density function

$$P(a \le X \le b) = P(a \le X \le b, -\infty \le Y \le \infty) = \int_a^b \left[\int_{-\infty}^{\infty} f(x,y)\,dy \right] dx = \int_a^b f_X(x)\,dx$$

Continuous case

X and Y are independent

$\Updownarrow$

$\{X \le s\}$ and $\{Y \le t\}$ are independent for all s and t

$\Updownarrow$

$$\underbrace{P(X \le s, Y \le t)}_{\int_{-\infty}^{s}\left[\int_{-\infty}^{t} f(x,y)\,dy\right]dx} = \underbrace{P(X \le s)\,P(Y \le t)}_{\left[\int_{-\infty}^{s} f_X(x)\,dx\right]\left[\int_{-\infty}^{t} f_Y(y)\,dy\right] = \int_{-\infty}^{s}\left[\int_{-\infty}^{t} f_X(x) f_Y(y)\,dy\right]dx}$$

$\Updownarrow$

$$\int_{-\infty}^{s}\int_{-\infty}^{t} f(x,y)\,dy\,dx \overset{?}{=} \int_{-\infty}^{s}\int_{-\infty}^{t} f_X(x) f_Y(y)\,dy\,dx$$

$\Updownarrow$

$$f(x,y) = f_X(x) f_Y(y)$$

Example : Assume that $\lambda > 0$

Consider the following joint density function for random variables X and Y:

$$f(x,y) = \begin{cases} \lambda^2 e^{-\lambda y} & \text{if } 0 \le x \le y; \\ 0 & \text{otherwise.} \end{cases} \tag{2.1}$$

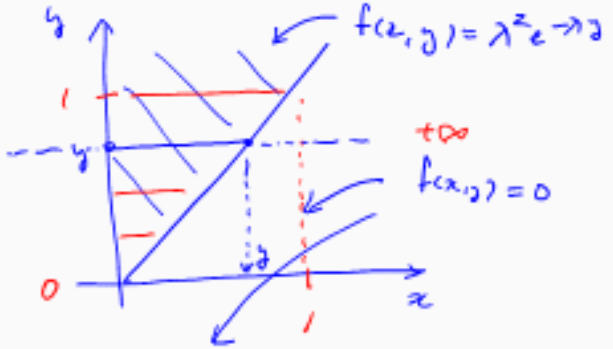
(a) Find $P(X \le 1, Y \le 1)$. (b) Find the marginal density functions $f_X(x)$ and $f_Y(y)$. (c) Are X and Y independent?

(a) $P(X \le 1, Y \le 1) = \int_{-\infty}^{1} \left[\int_{-\infty}^{1} f(x,y)\,dx \right] dy$

$= \int_0^1 \left[\int_0^y \lambda^2 e^{-\lambda y}\,dx \right] dy$

$= \int_0^1 \left[\lambda^2 e^{-\lambda y} \int_0^y dx \right] dy$

$= \int_0^1 (\lambda^2 e^{-\lambda y})(y)\,dy$



Example

Consider the following joint density function for random variables X and Y:

$$f(x,y) = \begin{cases} \lambda^2 e^{-\lambda y} & \text{if } 0 \le x \le y; \\ 0 & \text{otherwise.} \end{cases} \tag{2.1}$$

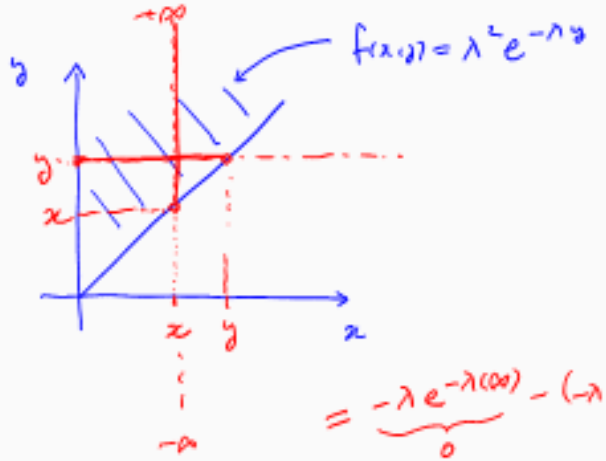
(a) Find $P(X \le 1, Y \le 1)$. (b) Find the marginal density functions $f_X(x)$ and $f_Y(y)$. (c) Are X and Y independent?

For (a) we can calculate it as follows.

$$P(X \le 1, Y \le 1) = \int_0^1 \int_0^y \lambda^2 e^{-\lambda y} dx dy = \int_0^1 \lambda^2 e^{-\lambda y} y dy$$

$$= \int_0^1 \lambda e^{-\lambda y} dy - \left[\lambda y e^{-\lambda y}\right]_0^1 = \left[-e^{-\lambda y}\right]_0^1 - \lambda e^{-\lambda} = 1 - e^{-\lambda} - \lambda e^{-\lambda}$$

Continue the solution.



(b) We obtain

$$f_X(x) = \int_x^{\infty} \lambda^2 e^{-\lambda y}\, dy = \left[-\lambda e^{-\lambda y}\right]_{y=x}^{y=\infty} = \lambda e^{-\lambda x}$$

$$f_Y(y) = \int_0^{y} \lambda^2 e^{-\lambda y}\, dx = \left[\lambda^2 e^{-\lambda y} x\right]_{x=0}^{x=y} = \lambda^2 y e^{-\lambda y}$$

for $0 \le x \le \infty$ and $0 \le y \le \infty$.

(c) Since $f(x,y) \neq f_X(x) f_Y(y)$, they are not independent.

Conditional density functions.

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

Conditional frequency function: $P(Y=y|X=x) = \frac{P(Y=y, X=x)}{P(X=x)} = \frac{p(x,y)}{p_X(x)}$

Suppose that two random variables X and Y has a joint density function $f(x,y)$. If $f_X(x) > 0$, then we can define the *conditional density function* $f_{Y|X}(y|x)$ given $X = x$ by

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

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

Similarly we can define the *conditional density function* $f_{X|Y}(x|y)$ given $Y = y$ by

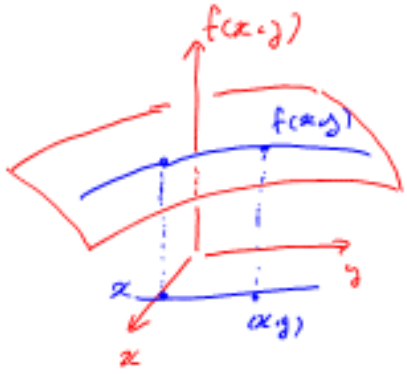
$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}$$

if $f_Y(y) > 0$. Then, clearly we have the following relation

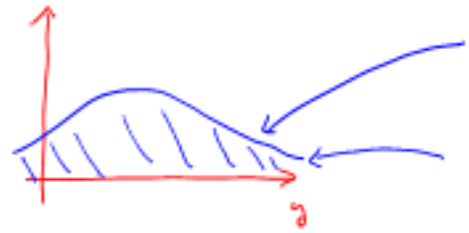
$$f(x,y) = f_{Y|X}(y|x)f_X(x) = f_{X|Y}(x|y)f_Y(y).$$

How does it make sense?

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



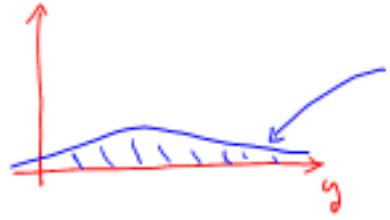
$z = f(x,y)$ for a fixed x



It looks like a density, but it is not?

$$\int_{-\infty}^{\infty} f(x,y)\,dy = f_X(x) \neq 1 \quad \text{in general}$$

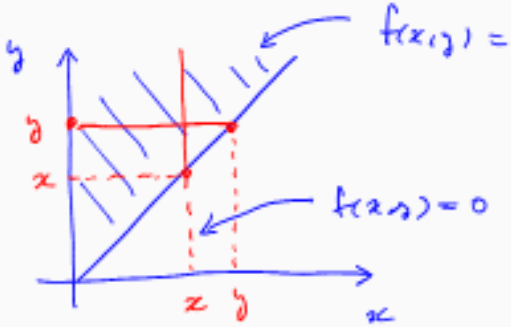
$$z = \frac{f(x,y)}{f_X(x)} = f_{Y|X}(y|z)$$



$$\int_{-\infty}^{\infty} \frac{f(x,y)}{f_X(x)}\,dy = \frac{1}{f_X(x)} \underbrace{\int_{-\infty}^{\infty} f(x,y)\,dy}_{f_X(x)} = 1$$

Example

Find the conditional density functions $f_{X|Y}(x|y)$ and $f_{Y|X}(y|x)$ for the joint density function (2.1) in the previous example.

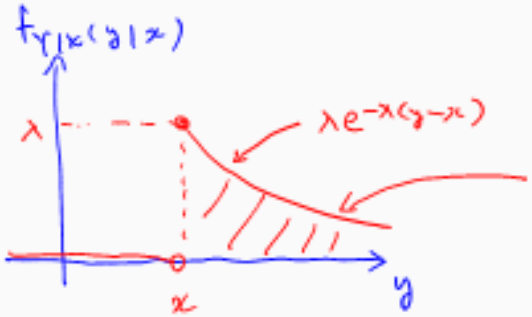


$f(x,y) = \lambda^2 e^{-\lambda y} \longrightarrow f_X(x) = \lambda e^{-\lambda x}$

$f_Y(y) = \lambda^2 y e^{-\lambda y}$

$f(x,y) = 0$

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} = \begin{cases} \dfrac{\lambda^2 e^{-\lambda y}}{\lambda e^{-\lambda x}} & \text{if } y \geq x \\ 0 & \text{if } y < x \end{cases} = \begin{cases} \lambda e^{-\lambda(y-x)} & \text{if } y \geq x \\ 0 & \text{if } y < x \end{cases}$$



$$\int_x^\infty \lambda e^{-\lambda(y-x)} = 1$$

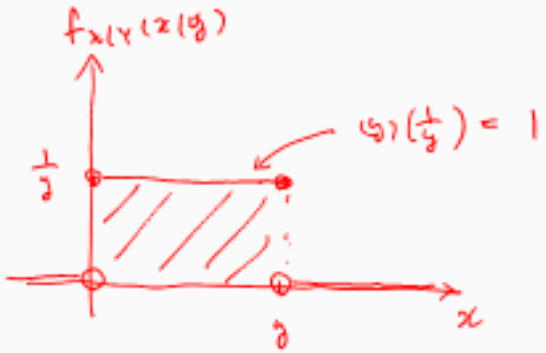
Example

Find the conditional density functions $f_{X|Y}(x|y)$ and $f_{Y|X}(y|x)$ for the joint density function (2.1) in the previous example.

$\lambda^2 e^{-\lambda y}$ if $0 \le x \le y$

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = \begin{cases} \frac{\lambda^2 e^{-\lambda y}}{\lambda^2 y e^{-\lambda y}} = \frac{1}{y} & \text{if } 0 \le x \le y; \\ 0 & \text{if } x > y \text{ or } x < 0. \end{cases}$$

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} = \begin{cases} 0 & \text{if } y < x; \\ \frac{\lambda^2 e^{-\lambda y}}{\lambda e^{-\lambda x}} = \lambda e^{-\lambda(y-x)} & \text{if } y \ge x. \end{cases}$$



Independence of random variables.

Recall independence: X and Y are independent if $p(x,y) = p_X(x)\,p_Y(y)$ $\leftarrow$ $P(A \cap B) = P(A)\,P(B)$

In theory we can consider a joint frequency function $p(x, y, z)$ for three discrete random variables X, Y, and Z. If it satisfies

$$P(X=x, Y=y, Z=z) = p(x, y, z) = p_X(x)p_Y(y)p_Z(z)$$

then they are said to be independent. In the same manner, if a joint density function $f(x, y, z)$ for continuous random variables$X$, Y and Z is expressed in the form

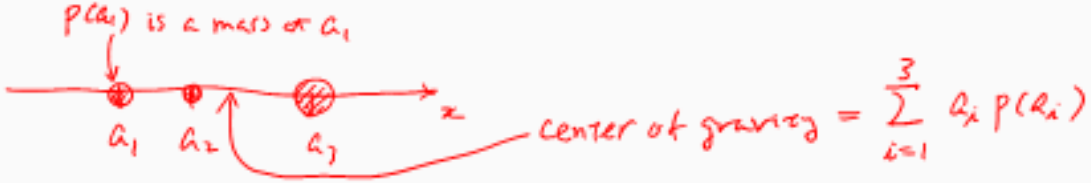
$P(a \le X \le b,\ c \le Y \le d,\ e \le Z \le f) = \int_a^b \int_c^d \int_e^f f(x,y,z)\,dz\,dy\,dx$

$$f(x, y, z) = f_X(x)f_Y(y)f_Z(z)$$

then they are independent. In general if $X_1, X_2, \ldots, X_n$ are **independent** random variables then we can construct the joint density function by

$$f(x_1, x_2, \ldots, x_n) = f_{X_1}(x_1) \times f_{X_2}(x_2) \times \cdots \times f_{X_n}(x_n)$$

Expectation.



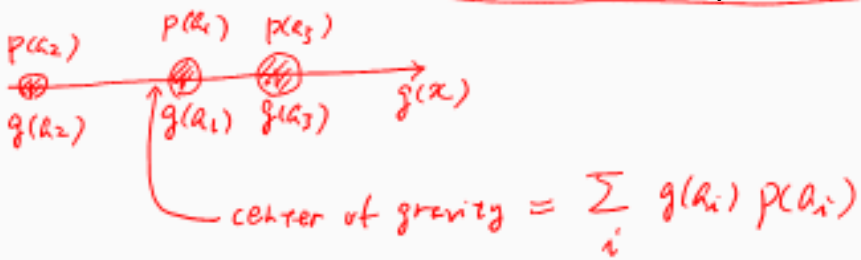
Let X be a discrete random variable whose possible values are $a_1, a_2, \ldots$, and let $p(x)$ is the frequency function of X. Then the **expectation** (**expected value** or **mean**) of the random variable X is given by

$$E[X] = \sum_i a_i p(a_i).$$

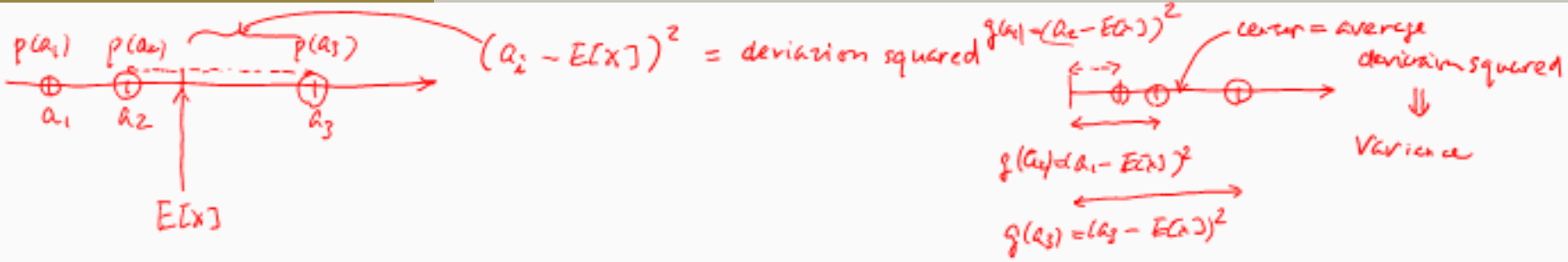
We often denote the expected value $E(X)$ of X by μ or μ_X. For a function g, we can define the expectation of function of random variable by

$$E[g(X)] = \sum_i g(a_i) p(a_i).$$

← Why do we need it ?



Variance.



The **variance** of a random variable X, denoted by $\mathrm{Var}(X)$ or σ^2, is the expected value of "the squared difference between the random variable and its expected value $E(X)$," and can be defined as

We need to prove it.

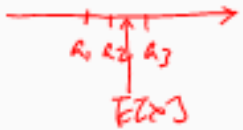
$$\mathrm{Var}(X) := E[(X - E(X))^2] = E[X^2] - (E[X])^2.$$

$g(X)$?

The square-root $\sqrt{\mathrm{Var}(X)}$ of the variance $\mathrm{Var}(X)$ is called the **standard error (SE)** (or **standard deviation (SD)**) of the random variable X.

Variance is large

Variance is small



Example

A random variable X takes on values 0, 1, 2 with the respective probabilities $P(X=0)=0.2$, $P(X=1)=0.5$, $P(X=2)=0.3$. Compute

1. $E[X] = \sum_{i=0}^{2} i\,p(i) = (0)(0.2) + (1)(0.5) + (2)(0.3) = 1.1$
2. $E[X^2] = \sum_{i=0}^{2} i^2 p(i) = (0)^2(0.2) + (1)^2(0.5) + (2)^2(0.3) = 1.7$
3. $\mathrm{Var}(X)$ and SD of X

$$\mathrm{Var}(x) = E[(x - \underbrace{E[x]}_{1.1})^2] = \sum_{i=0}^{2} (i-1.1)^2 p(i)$$

$$= (-1.1)^2(0.2) + (-0.1)^2(0.5) + (0.9)^2(0.3) = 0.49$$

$$\mathrm{Var}(x) = E[x^2] - (E[x])^2 = (1.7) - (1.1)^2 = 0.49$$

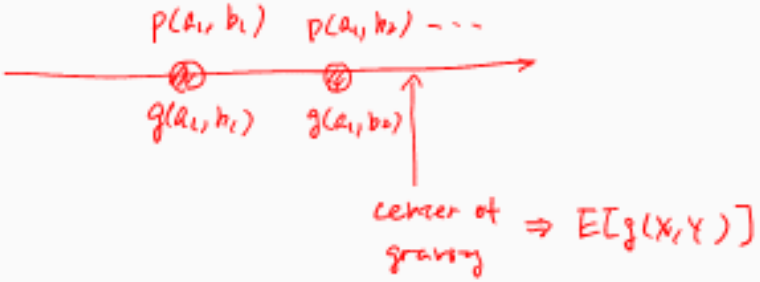
Example

A random variable X takes on values 0, 1, 2 with the respective probabilities $P(X = 0) = 0.2$, $P(X = 1) = 0.5$, $P(X = 2) = 0.3$. Compute

1. $E[X]$
2. $E[X^2]$
3. $\mathrm{Var}(X)$ and SD of X

1. $E[X] = (0)(0.2) + (1)(0.5) + (2)(0.3) = 1.1$
2. $E[X^2] = (0)^2(0.2) + (1)^2(0.5) + (2)^2(0.3) = 1.7$
3. $\mathrm{Var}(X) = E[(X - 1.1)^2] =$
   $(-1.1)^2(0.2) + (-0.1)^2(0.5) + (0.9)^2(0.3) = 0.49$. Also, using $\mathrm{Var}(X) = E[X^2] - (E[X])^2$ we can calculate
   $\mathrm{Var}(X) = (1.7) - (1.1)^2 = 0.49$.
   Then we obtain the SD of $\sqrt{0.49} = 0.7$.

Expectation for two variables.



→ P(X=x, Y=y)

Suppose that we have two random variables X and Y, and that $p(x, y)$ is their joint frequency function. Then the expectation of function $g(X, Y)$ of the two random variables X and Y is defined by

$$E[g(X, Y)] = \sum_{i,j} g(a_i, b_j)p(a_i, b_j),$$

where the sum is over all the possible values of (X, Y).

Properties of expectation.

One can think of the expectation $E(X)$ as "an operation on a random variable X" which returns the average value for X.

1. Let a be a constant, and let X be a random variable having the frequency function $p(x)$. Then we can show that

$$E[a + X] = \sum_x (a + x)p(x) = a + \sum_x x\,p(x) = a + E[X].$$

2. Let a and b be scalars, and let X and Y be random variables having the joint frequency function $p(x, y)$ and the respective marginal density functions $p_X(x)$ and $p_Y(y)$.

$$E[aX + bY] = \sum_{x,y} (ax + by)p(x, y) = a\sum_x x\left(\sum_y p(x,y)\right) + b\sum_y y\left(\sum_x p(x,y)\right)$$

$$= a\sum_x x\,p_X(x) + b\sum_y y\,p_Y(y) = aE[X] + bE[Y].$$

Claim: $Var(X) = E[(X - \underbrace{E[X]}_{\text{constant}})^2] = E[X^2] - (E[X])^2$

Proof: $E[(X - E[X])^2] = E[X^2 - \underbrace{2X\,E[X]}_{(2E[X])X} + \underbrace{(E[X])^2}_{\text{constant}}]$

$= E[X^2] - (2E[X])\,E[X] + (E[X])^2$

$= E[X^2] - 2(E[X])^2 + (E[X])^2$

$= E[X^2] - (E[X])^2$ QED.

Linearity property of expectation.

Let a and $b_1, \ldots, b_n$ be scalars, and let $X_1, \ldots, X_n$ be random variables. By applying the properties of expectation repeatedly, we can obtain

$$E\left[a + \sum_{i=1}^{n} b_i X_i\right] = a + E\left[\sum_{i=1}^{n} b_i X_i\right] = a + b_1 E[X_1] + E\left[\sum_{i=2}^{n} b_i X_i\right]$$
$$= \cdots = a + \sum_{i=1}^{n} b_i E[X_i].$$

It is also useful to observe the above property as that of "linear operator."

Expectation for independent random variables.

Suppose that X and Y are independent random variables. Then the joint frequency function $p(x, y)$ of X and Y can be expressed as

$$p(x, y) = p_X(x)p_Y(y).$$

By definition of independence

And the expectation of the function of the form $g_1(X) \times g_2(Y)$ is given by

$g(X,Y)$

Double summation $\sum_x \sum_y$

$$E[g_1(X)g_2(Y)] = \sum_{x,y} g_1(x)g_2(y)\, p(x, y) = \sum_x g_1(x) p_X(x) \left(\sum_y g_2(y) p_Y(y) \right)$$

$p_X(x)\, p_Y(y)$

$$= \left[\sum_x g_1(x)p_X(x) \right] \times \left[\sum_y g_2(y)p_Y(y) \right] = E[g_1(X)] \times E[g_2(Y)].$$

$E[g_1(X)]$ $\qquad$ $E[g_2(Y)]$

Covariance.

$$Var(X) = E[(X - E[X])^2]$$ (deviation)

$$\begin{aligned} Cov(X,Y) &= E[(X-\mu_x)(Y-\mu_y)] \\ &= E[XY - \mu_x Y - \mu_y X + \mu_x \mu_y] \\ &= E[XY] - \mu_x E[Y] - \mu_y E[X] + \mu_x \mu_y = E[XY] - \mu_x \mu_y \end{aligned}$$

Suppose that we have two random variables X and Y. Then the **covariance** of two random variables X and Y can be defined as

$$\mathrm{Cov}(X, Y) := E((X - \mu_x)(Y - \mu_y)) = E(XY) - E(X) \times E(Y),$$

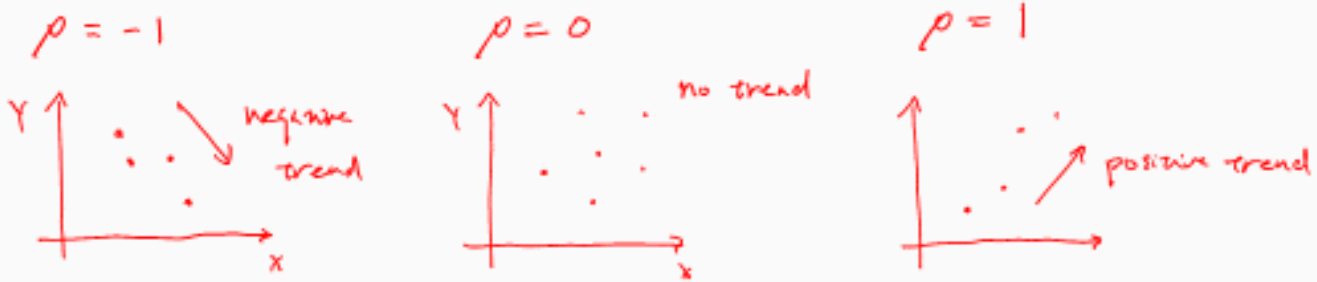
(relationship between X and Y)

where $\mu_x = E(X)$ and $\mu_y = E(Y)$. Then the **correlation coefficient**

(constant) (constant)

$$-1 \leq \rho = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} \leq 1$$

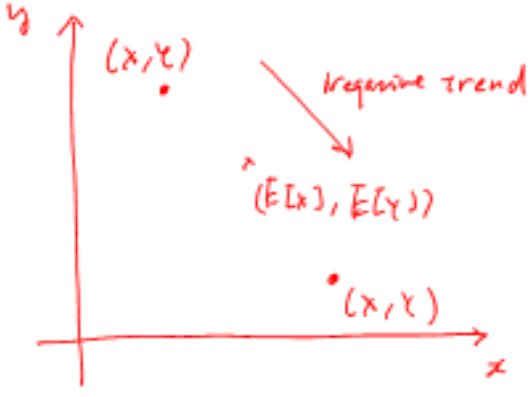
measures the strength of the dependence of the two random variables.



What is a relationship?

$Cov(X,Y) = E[(x - E[x])(Y - E[Y])] \longrightarrow$ positive

positive	positive
negative	negative



$E[(X - E[x])(Y - E[Y])] \longrightarrow$ negative

positive	negative
negative	positive

Properties of variance and covariance.

1. If X and Y are independent, then $\mathrm{Cov}(X, Y) = 0$ by observing that $E[XY] = E[X] \cdot E[Y]$.
2. In contrast to the expectation, the variance is *not* a linear operator. For two random variables X and Y, we have

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X, Y). \tag{2.2}$$

Moreover, if X and Y are independent, by observing that $\mathrm{Cov}(X, Y) = 0$ in (2.2), we obtain $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$. In general, we have

$$\mathrm{Var}(X_1 + \cdots + X_n) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n).$$

if $X_1, \ldots, X_n$ are independent random variables.

Variance and covariance under linear transformation.

Let a and b be scalars (that is, real-valued constants), and let X be a random variable. Then the variance of $aX + b$ is given by

$$\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X).$$

Exercise

Now let a_1, a_2, b_1 and b_2 be scalars, and let X and Y be random variables. Then similarly the covariance of $a_1X + b_1$ and $a_2Y + b_2$ can be given by

$$\mathrm{Cov}(a_1X + b_1, a_2Y + b_2) = a_1a_2\mathrm{Cov}(X, Y)$$

$\mathrm{Cov}(a_1X+b_1, a_2Y+b_2) = E[(X'-E[X'])(Y'-E[Y'])] = a_1a_2\,E[(X-E[X])(Y-E[Y])]$

$X' = a_1X+b_1 - E[a_1X+b_1] = (a_1E[X]+b_1) = a_1(X-E[X])$

$Y' = a_2(Y-E[Y])$

$E[(X-E[X])(Y-E[Y])] = \mathrm{Cov}(X,Y)$

Counterexample : $\mathrm{Cov}(X,Y)=0$ does not imply independence for X and Y.

Example

The joint frequency function $p(x, y)$ of two discrete random variables, X and Y, is given by

Y \ X	$x=-1$	0	1	$P_Y(y)$
$y=-1$	0	$1/2$	0	$1/2$
$y=1$	$1/4$	0	$1/4$	$1/2$
$P_X(x)$	$1/4$	$1/2$	$1/4$	

$E[(X-E(X))(Y-E(Y))]$
$= E[XY]$ since $E(X)=E(Y)=0$

1. Find the marginal frequency function for X and Y. $E[Y]=(-1)P_Y(-1)+(1)P_Y(1)=0$
2. Find $E[X]$ and $E[Y]$. $E[X]=(-1)P_X(-1)+(0)P_X(0)+(1)P_X(1)=0$
3. Find $\mathrm{Cov}(X, Y)$. $=E[XY]=(-1)(0)p(-1,1)+(0)(-1)p(0,-1)+(1)(1)p(1,1)=0$
4. Are X and Y independent? No.

$p(-1,-1)=0 \neq P_X(-1)P_Y(-1)=\left(\frac{1}{4}\right)\left(\frac{1}{2}\right)=\frac{1}{8}$

Solution.

1. $p_X(-1) = \frac{1}{4}$; $p_X(0) = \frac{1}{2}$; $p_X(1) = \frac{1}{4}$.
   $p_Y(-1) = \frac{1}{2}$; $p_X(1) = \frac{1}{2}$.
2. $E[X] = (-1)\left(\frac{1}{4}\right) + (0)\left(\frac{1}{2}\right) + (1)\left(\frac{1}{4}\right) = 0$
   $E[Y] = (-1)\left(\frac{1}{2}\right) + (1)\left(\frac{1}{2}\right) = 0$
3. $E[XY] = (-1)(1)\left(\frac{1}{4}\right) + (0)(-1)\left(\frac{1}{2}\right) + (1)(1)\left(\frac{1}{4}\right) = 0$
   Thus, we obtain $\mathrm{Cov}(X, Y) = E[XY] - E[X]Y[Y] = 0$.
4. No, X and Y are not independent, because
   $p(-1, -1) = 0 \neq p_X(-1)p_Y(-1) = \frac{1}{8}$.

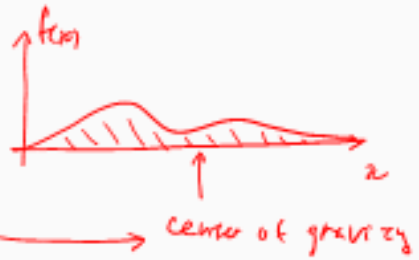
Expectation of continuous random variables.

$$P(a \le X \le b) = \int_a^b f(x)\,dx$$

Let $f(x)$ be the pdf of a continuous random variable X. Then we define the expectation of the random variable X by

improper integral

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



Furthermore, we can define the expectation $E[g(X)]$ of the function $g(X)$ of random variable by

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

if $g(x)$ is integrable with respect to $f(x)\,dx$, that is, $\int_{-\infty}^{\infty} |g(x)|\, f(x)\, dx < \infty$.

Recall: If X is a discrete random variable then

$$E[g(X)] = \sum_i g(a_i)\, p(a_i)$$

Properties of expectation.

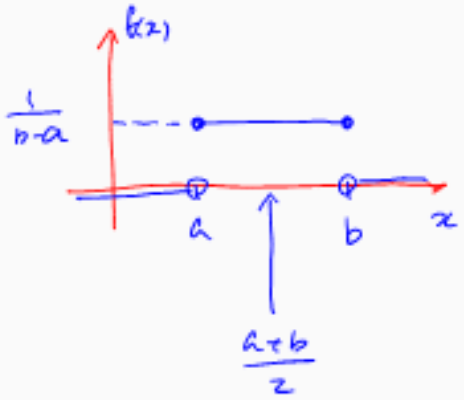
When X is a discrete random variable, $E[X]$ is considered as a "linear operator." This remains true even if X is a continuous random variable. Thus, we have the following properties (without proof):

1. $E[a + X] = a + E[X]$.
2. $E[aX + bY] = aE[X] + bE[Y]$.
3. $E\left[a + \sum_{i=1}^{n} b_i X_i\right] = a + \sum_{i=1}^{n} b_i E[X_i]$.
4. If X and Y are independent then

$$E[g_1(X)g_2(Y)] = E[g_1(X)] \times E[g_2(Y)].$$

Example

Let X be a uniform random variable on $[a, b]$. Compute $E[X]$ and $E[X^2]$, and find $\mathrm{Var}(X)$.



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

$$= \int_a^b x \left(\frac{1}{b-a}\right) dx = \left[\left(\frac{1}{b-a}\right)\frac{x^2}{2}\right]_a^b$$

$$= \frac{b^2-a^2}{2(b-a)} = \frac{(b-a)(b+a)}{2(b-a)} = \frac{a+b}{2}$$

$$E[X^2] = \int_a^b x^2 \frac{1}{b-a}\, dx = \left[\left(\frac{1}{b-a}\right)\frac{x^3}{3}\right]_a^b = \frac{b^3-a^3}{3(b-a)} = \frac{(b-a)(b^2+ab+a^2)}{3(b-a)} = \frac{b^2+ab+a^2}{3}$$

$$\mathrm{Var}(X) = E[X^2] - (E[X])^2 = \frac{a^2+ab+b^2}{3} - \left(\frac{a+b}{2}\right)^2 = \frac{a^2+ab+b^2}{3} - \frac{a^2+2ab+b^2}{4}$$

$$= \frac{4(a^2+ab+b^2) - 3(a^2+2ab+b^2)}{12} = \frac{a^2-2ab+b^2}{12} = \frac{(b-a)^2}{12}$$

Example

Let X be a uniform random variable on $[a, b]$. Compute $E[X]$ and $E[X^2]$, and find $\mathrm{Var}(X)$.

$$
\begin{aligned}
E[X] &= \int_a^b x \left(\frac{1}{b-a}\right) dx = \left(\frac{1}{b-a}\right) \left[\frac{x^2}{2}\right]_a^b \\
&= \frac{a+b}{2} \\
E[X^2] &= \int_a^b x^2 \left(\frac{1}{b-a}\right) dx = \left(\frac{1}{b-a}\right) \left[\frac{x^3}{3}\right]_a^b \\
&= \frac{a^2+ab+b^2}{3} \\
\mathrm{Var}(X) &= \frac{a^2+ab+b^2}{3} - \left(\frac{a+b}{2}\right)^2 = \frac{(b-a)^2}{12}
\end{aligned}
$$

Moment generating function.

Laplace transform

$f(x) \xleftrightarrow{\mathcal{L}f} [\mathcal{L}f](s) = \int_0^\infty e^{-sx} f(x)\,dx$ exists on (c_1, c_2)

Let X be a random variable. Then the *moment generating function* (*mgf*) of X is given by

$$M(t) = E\left[e^{tX}\right] \overset{\text{If } X \text{ is continuous}}{=} \int_{-\infty}^{\infty} e^{tx} f(x)\,dx$$

The mgf $M(t)$ is a function of t defined on some open interval (c_0, c_1) with $c_0 < 0 < c_1$. In particular when X is a continuous random variable having the pdf $f(x)$, the mgf $M(t)$ can be expressed as

$$M(t) = \int_{-\infty}^{\infty} e^{tx} f(x)\,dx.$$

When X is discrete the mgf is given by

$$M(t) = \sum_i e^{t a_i} p(a_i)$$

Properties of moment generating function.

(a) The most significant property of moment generating function is that "the moment generating function uniquely determines the distribution."

(b) Let a and b be constants, and let $M_X(t)$ be the mgf of a random variable X. Then the mgf of the random variable $Y = a + bX$ can be given as follows.

$$M_Y(t) = E\left[e^{tY}\right] = E\left[e^{t(a+bX)}\right] = e^{at}E\left[e^{(bt)X}\right] = e^{at}M_X(bt).$$

(c) Let X and Y be independent random variables having the respective mgf's $M_X(t)$ and $M_Y(t)$. We can obtain the mgf $M_Z(t)$ of the sum $Z = X + Y$ of random variables as follows.

$$\begin{aligned} M_Z(t) = E\left[e^{tZ}\right] &= E\left[e^{t(X+Y)}\right] = E\left[e^{tX}e^{tY}\right] \\ &= E\left[e^{tX}\right] \cdot E\left[e^{tY}\right] = M_X(t) \cdot M_Y(t). \end{aligned}$$

Properties of moment generating function, continued.

$$F(x) = \int_a^b f(x,t)\,dx$$

$$\frac{d}{dt}F(x) = \int_a^b \frac{\partial}{\partial t} f(x,t)\,dx$$

$$\frac{d}{dt}e^{tx} = x e^{tx}$$

$$\frac{d^2}{dt^2}e^{tx} = \frac{d}{dt}[x e^{tx}] = x^2 e^{tx}$$

$$\frac{d^r}{dt^r}e^{tx} = x^r e^{tx}$$

(d) When $t = 0$, it clearly follows that $M(0) = 1$. Now by differentiating $M(t)$ r times, we obtain

$$M^{(r)}(t) = \frac{d^r}{dt^r}E\left[e^{tX}\right] = E\left[\frac{d^r}{dt^r}e^{tX}\right] = E\left[X^r e^{tX}\right].$$

In particular when $t = 0$, $M^{(r)}(0)$ *generates* the r-th moment of X as follows.

$$M^{(r)}(0) = E[X^r], \quad r = 1, 2, 3, \ldots$$

$= E[X^r e^{0X}]$, $e^{0X} = 1$; $E[X^r]$: r-th moment of X

Example

Find the mgf $M(t)$ for a uniform random variable X on $[a, b]$. Then show that (a) $\lim_{t\to 0} M(t) = 1$ and (b) $\lim_{t\to 0} M'(t) = \frac{a+b}{2}$.

$$M_X(t) = E[e^{tx}] = \int_a^b e^{tx}\left(\frac{1}{b-a}\right)dx = \left[\frac{1}{b-a}\frac{e^{tx}}{t}\right]_a^b = \frac{e^{tb}-e^{ta}}{t(b-a)}$$

$$M_X(0) = \lim_{t\to 0} M_X(t) = \lim_{t\to 0}\left(\frac{1}{b-a}\right)\frac{e^{tb}-e^{ta}}{t} = \lim_{t\to 0}\left(\frac{1}{b-a}\right)\frac{be^{tb}-ae^{ta}}{1} = 1$$

$M_X(0) = E[e^{(0)x}] = E[1] = 1$

$$M_X'(t) = \left(\frac{1}{b-a}\right)\frac{d}{dt}\left(\frac{e^{tb}-e^{ta}}{t}\right) = \left(\frac{1}{b-a}\right)\left[\frac{(be^{tb}-ae^{ta})t-(e^{tb}-e^{ta})}{t^2}\right]$$

$M_X'(t) = E[xe^{tx}]$

$$M_X'(0) = \lim_{t\to 0} M_X'(t) = \cdots = \frac{a+b}{2}$$

$M_X'(0) = E[xe^{(0)x}] = E[x] = \frac{a+b}{2}$

Example

Find the mgf $M(t)$ for a uniform random variable X on $[a, b]$. Then show that (a) $\lim_{t\to 0} M(t) = 1$ and (b) $\lim_{t\to 0} M'(t) = \frac{a+b}{2}$.

$$M(t) = \int_a^b e^{tx}\left(\frac{1}{b-a}\right)dx = \begin{cases} \int_a^b \left(\frac{1}{b-a}\right)dx = 1 & \text{if } t = 0; \\ \left(\frac{1}{b-a}\right)\left[\frac{e^{tx}}{t}\right]_{x=a}^{x=b} = \frac{e^{tb}-e^{ta}}{t(b-a)} & \text{if } t \neq 0. \end{cases}$$

$$\lim_{t\to 0} M(t) = \lim_{t\to 0} \frac{e^{tb}-e^{ta}}{t(b-a)} = \lim_{t\to 0} \frac{be^{tb}-ae^{ta}}{b-a} = 1 \text{ by l'Hospital's rule.}$$

$$\lim_{t\to 0} M'(t) = \lim_{t\to 0} \frac{(be^{tb}-ae^{ta})t-(e^{tb}-e^{ta})}{t^2(b-a)}$$

$$= \lim_{t\to 0} \frac{b^2e^{tb}-a^2e^{ta}}{2(b-a)} = \frac{a+b}{2} \text{ again, by l'Hospital's rule.}$$

$= E[X] = M_X'(0)$

Exercises

Problem

Let $p(k)$, $k = -1, 0, 1$, be the frequency function for random variable X. Suppose that $p(0) = \frac{1}{4}$, and that $p(-1)$ and $p(1)$ are unknown.

1. *Show that $E[X^2]$ does not depend on the unknown values $p(-1)$ and $p(1)$.*
2. *If $E[X] = \frac{1}{4}$, then find the values $p(-1)$ and $p(1)$.*

Problem 2

The joint frequency function $p(x, y)$ of two discrete random variables, X and Y, is given by

		X		
		1	2	3
Y	1	c	$3c$	$2c$
	2	c	c	$2c$

1. *Find the constant c.*
2. *Find $E[X]$ and $E[XY]$.*
3. *Are X and Y independent?*

Problem 3

A pair (X, Y) of discrete random variables has the joint frequency function

$$p(x, y) = \frac{xy}{18}, \ x = 1, 2, 3 \text{ and } y = 1, 2.$$

1. *Find $P(X + Y = 3)$.*
2. *Find the marginal frequency function for X and Y.*
3. *Find $E[Y]$ and $\mathrm{Var}(Y)$.*
4. *Are X and Y independent? Justify your answer.*

Problem 4

1. *Determine the constant c so that $p(x)$ is a frequency function if $p(x) = cx$, $x = 1, 2, 3, 4, 5, 6$.*
2. *Similarly find c if $p(x) = c\left(\frac{2}{3}\right)^x$, $x = 1, 2, 3, 4, \ldots$.*

Geometric series: $\sum_{k=1}^{\infty} a^k = \frac{a}{1-a}$ if $|a| < 1$

$\sum_{x=1}^{\infty} p(x) = 1 \Rightarrow c\sum_{x=1}^{\infty}\left(\frac{2}{3}\right)^x = c\,\frac{\left(\frac{2}{3}\right)}{1-\left(\frac{2}{3}\right)} = 1$

Problem 5

Let X and Y be independent random variables.

1. *Show that* $\mathrm{Var}(aX) = a^2\mathrm{Var}(X)$.
2. *If* $E[X] = 1, E[Y] = 2$ *and* $\mathrm{Var}(X) = 4, \mathrm{Var}(Y) = 9$ *then find the mean and the variance of* $Z = 3X - 2Y$.

Problem

Suppose that X has the density function $f(x) = cx^2$ for $0 \leq x \leq 1$ and $f(x) = 0$ otherwise.

1. *Find c.*
2. *Find the cdf.*
3. *What is $P(.1 \leq X < .5)$?*

Problem 7

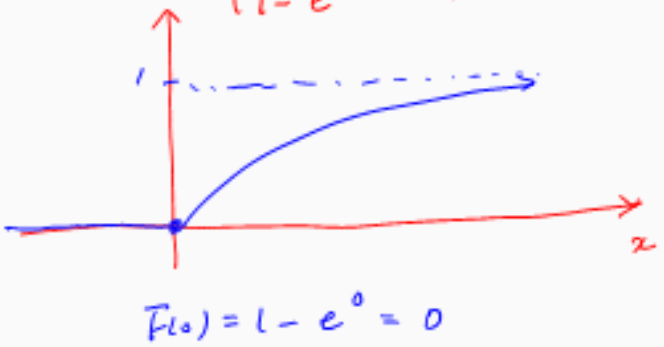
Let X be a random variable with the pdf

$$f(x) = \begin{cases} 2x & \text{if } 0 \leq x \leq 1; \\ 0 & \text{otherwise.} \end{cases}$$

1. *Find* $E[X]$.
2. *Find* $E[X^2]$ *and* $\mathrm{Var}(X)$.

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - e^{-\alpha x^\beta} & \text{if } x \geq 0 \end{cases} = \int_{-\infty}^{x} f(x)\,dx \quad \Leftarrow \quad f(x) = \frac{d}{dx} F(x) \text{ if } F \text{ is differentiable at } x$$

density function



$F(0) = 1 - e^0 = 0$

Answer I gave

$$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ \alpha\beta x^{\beta-1} e^{-\alpha x^\beta} & \text{if } x \geq 0 \end{cases}$$

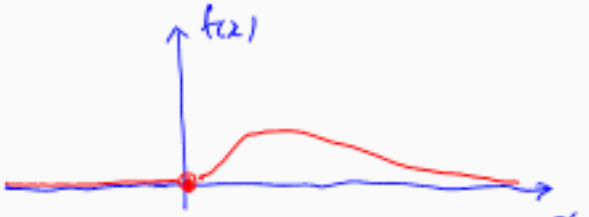
Problem

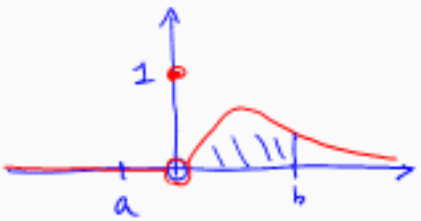
Let $F(x) = 1 - \exp(-\alpha x^\beta)$ for $x \geq 0$, $\alpha > 0$ and $\beta > 0$, and $F(x) = 0$ for $x < 0$. Show that F is a cdf, and find the corresponding density.

$$f(x) = \frac{d}{dx}\left[1 - e^{-\alpha x^\beta}\right] = -\left[-\alpha\beta x^{\beta-1} e^{-\alpha x^\beta}\right] = \alpha\beta x^{\beta-1} e^{-\alpha x^\beta} \quad \text{if } x > 0$$

$$f(x) = \frac{d}{dx} F(x) = \frac{d}{dx}[0] = 0 \quad \text{if } x < 0$$

You can set $f(0) = 0$ or anything $f(0) = 1$ for example





$$\int_a^b f(x)\,dx = F(b) - F(a)$$

$$\int_0^0 f(x)\,dx = 0$$

Problem 9

Suppose that random variables X and Y have the joint density

$$f(x,y) = \begin{cases} \frac{6}{7}(x+y)^2 & \text{if } 0 \le x \le 1 \text{ and } 0 \le y \le 1; \\ 0 & \text{otherwise.} \end{cases}$$

1. *Find the marginal density functions of X and Y.*
2. *Find the conditional density functions of X given $Y = y$ and of Y given $X = x$.*
3. *Find $E[X]$ and $E[Y]$.*

$E[aX+b] = a\,E[X] + b$

Problem 10

Discuss the following properties.

1. *Prove the property of variance:* $\mathrm{Var}(aX+b) = a^2\mathrm{Var}(X)$.
2. *Suppose that* $E[X] = \mu$ *and* $\mathrm{Var}(X) = \sigma^2$. *Then* $Z = (X-\mu)/\sigma$ *is said to be standardized, in which* σ *is often referred as "standard error." Show that* $E[Z] = 0$ *and* $\mathrm{Var}(Z) = 1$.

Problem 11

Recall that the expectation is an "linear operator." By using the definitions of variance and covariance in terms of the expectation, show that

1. $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X,Y)$, *and*
2. $\mathrm{Var}(X-Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) - 2\mathrm{Cov}(X,Y)$.

Problem 12

If X and Y are independent random variables with equal variances, find $\mathrm{Cov}(X+Y, X-Y)$.

Problem 13

Suppose that X and Y are independent random variables with their variances $\sigma_x^2 = \mathrm{Var}(X)$ and $\sigma_y^2 = \mathrm{Var}(Y)$. Let $Z = Y - X$.

1. *Find expressions for the covariance of X and Z in terms of σ_x and σ_y.*
2. *Find the correlation coefficient ρ of X and Z in terms of σ_x and σ_y.*

Problem 14

Let X and Y be independent random variables, and let α and β be scalars. Find an expression for the mgf $M_Z(t)$ of $Z = \alpha X + \beta Y$ in terms of the mgf's $M_X(t)$ and $M_Y(t)$ of X and Y.

$$M_Z(t) = E[e^{tZ}] = E[e^{t(\alpha X + \beta Y)}] = E[e^{t\alpha X} e^{t\beta Y}]$$

$$= E[e^{(t\alpha)X}] \times E[e^{(t\beta)Y}] = M_X(t\alpha)\, M_Y(t\beta)$$

↑ By independence

Optional problems

Problem

The Cauchy distribution function is defined by

$$F(x) = \frac{1}{2} + \frac{1}{\pi}\tan^{-1}(x), \quad -\infty < x < \infty.$$

1. *Show that this is a cdf.*
2. *Find the density function.*
3. *Suppose that a random variable X has the Cauchy distribution. Can you find $E[X]$? Why or why not?*

Problem

Let $f(x) = (1 + \alpha x)/2$ *for* $-1 \leq x \leq 1$ *and* $f(x) = 0$ *otherwise, where* $-1 \leq \alpha \leq 1$.

1. *Show that* f *is a density.*
2. *Find the cdf* $F(x)$*, and sketch the graph of* $F(x)$ *for* $-1 \leq \alpha < 0$, $\alpha = 0$, *and* $0 < \alpha \leq 1$.
3. *Find the formulas for the quantile function* $Q(u)$.
   Hint: Consider the following three cases separately: $-1 \leq \alpha < 0$, $\alpha = 0$, *and* $0 < \alpha \leq 1$.

Answers to exercises

Problem 11.

1. $E[X^2] = p(-1) + p(1) = 1 - p(0) = \frac{3}{4}$
2. $E[X] = -p(-1) + p(1) = \frac{1}{4}$. Together with $p(-1) + p(1) = \frac{3}{4}$, we obtain $p(-1) = \frac{1}{4}$ and $p(1) = \frac{1}{2}$.

Problem 12.

1. $\sum_{x=1}^{3}\sum_{y=1}^{2} p(x,y) = 10c = 1$ implies that $c = \frac{1}{10}$.
2. $E[X] = (1)\left(\frac{1}{5}\right) + (2)\left(\frac{2}{5}\right) + (3)\left(\frac{2}{5}\right) = \frac{11}{5}$
   $E[Y] = (1)\left(\frac{3}{5}\right) + (2)\left(\frac{2}{5}\right) = \frac{7}{5}$
   $E[XY] = (1)(1)\left(\frac{1}{10}\right) + (2)(1)\left(\frac{3}{10}\right) + (3)(1)\left(\frac{2}{10}\right) + (1)(2)\left(\frac{1}{10}\right) +$
   $(2)(2)\left(\frac{1}{10}\right) + (3)(2)\left(\frac{2}{10}\right) = \frac{31}{10}$
3. X and Y are not independent because
   $p(1,1) = \frac{1}{10} \neq p_X(1)p_Y(1) = \frac{3}{25}$. Or, you can find it by calculating

$$\mathrm{Cov}(X,Y) = E[XY] - E[X]E[Y] = \frac{1}{50}$$

Problem 13.

1. $P(X+Y=3) = p(1,2) + p(2,1) = \frac{2}{9}$
2. $p_X(y) = \frac{x}{6}$ for $x = 1, 2, 3$.
   $p_Y(y) = \frac{y}{3}$ for $y = 1, 2$.
3. $E[Y] = (1)\left(\frac{1}{3}\right) + (2)\left(\frac{2}{3}\right) = \frac{5}{3}$; $E[Y^2] = (1)^2\left(\frac{1}{3}\right) + (2)^2\left(\frac{2}{3}\right) = 3$
   $\mathrm{Var}(Y) = E[Y^2] - (E[Y])^2 = \frac{2}{9}$
4. Yes, since the joint frequency function satisfies
   $p(x,y) = p_X(x)p_Y(y)$ for all $x = 1, 2, 3$ and $y = 1, 2$.

Problem 14.

1. $\sum_{x=1}^{6} p(x) = c \times \frac{6(6+1)}{2} = 21c = 1$ Thus, we obtain $c = \frac{1}{21}$
2. $\sum_{x=1}^{\infty} p(x) = c \times \frac{(2/3)}{1-(2/3)} = 2c = 1$ Thus, we obtain $c = \frac{1}{2}$

Problem 15.

1. $\mathrm{Var}(aX) = E[(aX - E[aX])^2] = E[(aX - aE[X])^2] = E[a^2(X - E[X])^2] = a^2 E[(X - E[X])^2] = a^2 \mathrm{Var}(X)$.
2. $E[Z] = E[3X - 2Y] = (3)(1) - (2)(2) = -1$
   $\mathrm{Var}(Z) = \mathrm{Var}(3X + (-2)Y) = \mathrm{Var}(3X) + \mathrm{Var}((-2)Y) = (3)^2(4) + (-2)^2(9) = 72$

Problem 16.

1. $\int_0^1 cx^2\,dx = \frac{c}{3} = 1$. Thus, we have $c = 3$.
2. $F(t) = \begin{cases} 0 & \text{if } x < 0; \\ t^3 & \text{if } 0 \le x \le 1; \\ 1 & \text{if } x > 1. \end{cases}$
3. $P(.1 \le X < .5) = \int_{0.1}^{0.5} 3x^2\,dx = \left[x^3\right]_{0.1}^{0.5} = 0.124$

Problem 17.

1. $E[X] = \int_0^1 x(2x)\,dx = \left[\frac{2x^3}{3}\right]_0^1 = \frac{2}{3}$
2. $E[X^2] = \int_0^1 x^2(2x)\,dx = \left[\frac{x^4}{2}\right]_0^1 = \frac{1}{2}$ Then we obtain $\mathrm{Var}(X) = E[X^2] - (E[X])^2 = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{18}$.

Problem 18.

$F(x)$ is nonnegative and increasing for $x \geq 0$, and satisfies $F(0) = 0$, and $\lim_{x\to\infty} F(x) = 1$. Thus, $F(x)$ is cdf. The pdf is given by

$$f(x) = \frac{d}{dx}F(x) = \alpha\beta x^{\beta-1}\exp(-\alpha x^{\beta})$$

for $x \geq 0$.

Problem 19.

1. $f_X(x) = \int_0^1 \frac{6}{7}(x+y)^2 dy = \frac{2}{7}(3x^2+3x+1)$, $0 \le x \le 1$;
   $f_Y(y) = \int_0^1 \frac{6}{7}(x+y)^2 dx = \frac{2}{7}(3y^2+3y+1)$, $0 \le y \le 1$.
2. $f_{X|Y}(x|y) = \frac{3(x+y)^2}{3y^2+3y+1}$, $0 \le x \le 1$;
   $f_{Y|X}(y|x) = \frac{3(x+y)^2}{3x^2+3x+1}$, $0 \le y \le 1$;
3. $E[X] = \int_0^1 (x)\left(\frac{2}{7}\right)(3x^2+3x+1)dx = \frac{9}{14} = E[Y]$.

Problem 20.

1. Since $E[aX + b] = aE[X] + b$, by applying the definition of variance we obtain

$$\begin{aligned}\mathrm{Var}(aX + b) &= E[(aX + b - E[aX + b])^2] = E[a^2(X - E[X])^2] \\ &= a^2 E[(X - E[X])^2] = a^2 \mathrm{Var}(X)\end{aligned}$$

2. Since $Z = \left(\frac{1}{\sigma}\right) X - \left(\frac{\mu}{\sigma}\right)$ is a linear transformation, we obtain

$$E[Z] = \left(\frac{1}{\sigma}\right) E[X] - \left(\frac{\mu}{\sigma}\right) = 0$$

$$\mathrm{Var}(Z) = \left(\frac{1}{\sigma}\right)^2 \mathrm{Var}(X) = 1$$

Problem 21.

Let $\mu_x = E[X]$ and $\mu_y = E[Y]$. Then we show (a) as follows:

$$\begin{aligned}
\mathrm{Var}(X+Y) &= E[(X+Y-(\mu_x+\mu_y))^2] = E[((X-\mu_x)+(Y-\mu_y))^2] \\
&= E[(X-\mu_x)^2 + 2(X-\mu_x)(Y-\mu_y) + (Y-\mu_y)^2] \\
&= E[(X-\mu_x)^2] + 2E[(X-\mu_x)(Y-\mu_y)] + E[(Y-\mu_y)^2] \\
&= \mathrm{Var}(X) + 2\mathrm{Cov}(X,Y) + \mathrm{Var}(Y)
\end{aligned}$$

We can similarly verify (b).

Problem 22.

Let $\mu_x = E[X]$ and $\mu_y = E[Y]$. By applying the definition of variance and covariance, we obtain

$$\begin{aligned}
\mathrm{Cov}(X+Y, X-Y) &= E[(X+Y-(\mu_X+\mu_y))(X-Y-(\mu_x-\mu_y))] \\
&= E[(X-\mu_X)^2 - (Y-\mu_y)^2] \\
&= E[(X-\mu_X)^2] - E[(Y-\mu_y)^2] \\
&= \mathrm{Var}(X) - \mathrm{Var}(Y) = 0
\end{aligned}$$

where the last equality comes from the assumption of equal variances.

Problem 23.13

(a) Let $\mu_x = E[X]$ and $\mu_y = E[Y]$. Since $\mathrm{Cov}(X, Y) = 0$, we obtain

$$\begin{aligned}\mathrm{Cov}(X, Z) &= E[(X - \mu_x)((Y - X) - (\mu_y - \mu_x))] \\ &= E[(X - \mu_x)(Y - \mu_y) - (X - \mu_x)^2] \\ &= \mathrm{Cov}(X, Y) - \mathrm{Var}(X) = -\sigma_x^2\end{aligned}$$

(b) We can calculate

$$\mathrm{Var}(Z) = \mathrm{Var}(Y - X) = \mathrm{Var}(Y) + \mathrm{Var}(-X) = \sigma_y^2 + \sigma_x^2$$

Then we obtain

$$\rho = \frac{\mathrm{Cov}(X, Z)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Z)}} = -\frac{\sigma_x}{\sqrt{\sigma_x^2 + \sigma_y^2}}$$

Problem 24.

By the property of expectation for two independent random variables, we obtain

$$\begin{aligned} M_Z(t) &= E[e^{t(\alpha X+\beta Y)}] \\ &= E[e^{(t\alpha)X}] \times E[e^{(t\beta)Y}] = M_X(\alpha t) \times M_Y(\beta t) \end{aligned}$$