Chapter 3　Random Variable

Jingrui SUN

Email: sunjr@sustech.edu.cn

Southern University of Science and Technology

September 8, 2024

Outline I

Outline II

Outline III

Outline

1. An example: (Again !!)

Experiment: A coin is thrown 3 times. Then the sample space is

$$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}.$$

Now, perform this experiment and the total number of heads (appeared) is observed and recorded.

Part 3.1.1: Motivation

If the outcome is $\omega_1 \triangleq HHH$, then the total number of heads is 3.

If the outcome is $\omega_2 \triangleq HHT$, then the total number of heads is 2.

If the outcome is $\omega_3 \triangleq HTH$, then the total number of heads is 2.

... ...

If the outcome is $\omega_7 \triangleq TTH$, then the total number of heads is 1.

If the outcome is $\omega_8 \triangleq TTT$, then the total number of heads is 0.

Part 3.1.1: Motivation

If we let X denote the total number of heads, then we can see that:

(1) X is a "variable" (the values of X are numbers and it can take different values).

(2) The value of X depends upon the outcome, i.e. depends upon the particular ω in the sample space Ω.

In short, X can be viewed as a function from Ω to the real number.

(3) Since the outcome is random, and so the value of X is also random.

So, essentially, $\underline{X \text{ is a random number}}$.

We usually call such random number X as random variable.

Part 3.1.2: Definition

1. **Definition:** A random variable is a real-valued function defined on the points of a sample space.

2. Notes:

(1) First a random experiment with all the possible outcomes which form the **sample space** Ω. Then for each $\omega \in \Omega$ (a point in Ω) we assign a real value. Thus a real-valued "function" on Ω.

(2) Since the random variable X is a function on Ω, we usually write it as $X(\omega)$.

Part 3.1.2: Definition

3. Notations

(1) Random Variable $\longrightarrow$ r.v.

(2) Use capital letters X, Y, Z etc (or $X(\omega)$, $Y(\omega)$, $Z(\omega)$) to denote the random variables.

(3) Return to our example in the above we see that "the total number of heads appeared", X, is a random variable.

Further question: For what outcomes, to which the total number of heads is 2 (exactly 2 !!) ?

Answer: Clearly $\{HHT, HTH, THH\}$.

This is an event (because it is a subset of Ω), denoted by

$$A = \{HHT, HTH, THH\}.$$

That is, the event A is formed by such outcomes for which the number of heads is exactly 2, i.e.

$$\begin{aligned} A &= \{\omega \in \Omega; X(\omega) = 2\} \\ &= \{HHT, HTH, THH\}. \end{aligned}$$

Since A is an event, then we can consider its probability, i.e.

$$P(A) = P(\{\omega \in \Omega; X(\omega) = 2\}).$$

Usually we simply write it as $P(X(\omega) = 2)$ or even, more simply $P(X = 2)$.

Part 3.1.2: Definition

Now what's the meaning of $P(X \leq 2)$ or $P(X < 2)$?

The former is $P(B)$, where

$$B = \{HHT, HTH, HTT, THH, THT, TTH, TTT\}.$$

Write down the latter yourself.

Outline

1. Examples

Recall the example in §3.1.1 (Always keep this example in mind. It will be helpful in understanding many concepts).

Experiment: A fair coin is thrown 3 times:

$$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\},$$

or simply

$$\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6, \omega_7, \omega_8\}.$$

The meaning of the latter Ω is self-explained.

Part 3.2.1: Concept

Let X denote the number of heads, then X is a random variable. Indeed,

$$X(\omega_1) = 3, \quad X(\omega_2) = 2, \quad X(\omega_3) = 2, \quad X(\omega_4) = 1,$$
$$X(\omega_5) = 2, \quad X(\omega_6) = 1, \quad X(\omega_7) = 1, \quad X(\omega_8) = 0.$$

We see that, the r.v. X takes the values of $\{0, 1, 2, 3\}$.

These values are called the <u>possible values</u> of the r.v. X.

For the above r.v. X, it can take only a <u>finitely many number</u> of values (The number of possible values is finite).

We call it a <u>discrete</u> random variable. (In our example, there are 4 possible values i.e. $\{0, 1, 2, 3\}$).

Part 3.2.1: Concept

2. **Definition:** A random variable is called a <u>discrete</u> r.v., if it can take only a finitely many or at most a countable infinitely many number of values.

3. Notes:

"Countably infinitely many number of values" means all the possible values can be written as a sequence.

For example, the non-negative integers $\{0, 1, 2, 3, 4, \cdots\}$.

Part 3.2.1: Concept

4. Notation:

For a r.v. X, we shall use the lowercase letter x to denote the possible values.

Hence, for a discrete r.v. X, all the possible values can be rewritten as $\{x_1, x_2, \cdots, x_n\}$ (finitely many) or $\{x_1, x_2, \cdots, x_n, \cdots\}$ (sequence).

Note that, here, x_1, say, is a particular real number, it is not a random variable (The random variable is X).

In the above example, the random variable X is the total number of heads, so, the possible values of X are $\{0, 1, 2, 3\}$.

Here we may write $x_1 = 0, x_2 = 1, x_3 = 2, x_4 = 3$.

Part 3.2.2: p.m.f. for discrete random variables

1. Example: Again, see the above example. Since the coin is fair, all the 8 different outcomes have the same probability $\frac{1}{8}$. Hence,

$$
\begin{aligned}
P(X=0) &= \frac{1}{8}, \quad \text{since } (X=0) = \{\mathrm{TTT}\} = \{\omega_8\}; \\
P(X=1) &= \frac{3}{8}, \quad \text{since } (X=1) = \{\omega_4, \omega_6, \omega_7\}; \\
P(X=2) &= \frac{3}{8}, \quad \text{since } (X=2) = \{\omega_2, \omega_3, \omega_5\}; \\
P(X=3) &= \frac{1}{8}, \quad \text{since } (X=3) = \{\mathrm{HHH}\} = \{\omega_1\}.
\end{aligned}
$$

Part 3.2.2: p.m.f. for discrete random variables

Now, we let ("$\triangleq$" means "defined as")

$$p(0) \triangleq P(X = 0) = \frac{1}{8}; \quad p(1) \triangleq P(X = 1) = \frac{3}{8};$$
$$p(2) \triangleq P(X = 2) = \frac{3}{8}; \quad p(3) \triangleq P(X = 3) = \frac{1}{8};$$

i.e, in general, let (lowercase letter $p(\cdot)$)

$$p(x_i) \triangleq P(X = x_i),$$

we then get a function p.

Part 3.2.2: p.m.f. for discrete random variables

This function p is called the Probability Mass Function of the random variable X.

Note that this function p is a real function in the ordinary meaning. i.e. $p : \mathbb{R} \to \mathbb{R}$.

Essentially, this function can be rewritten as

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

Note also that, this function p has the property that

$$p(x_i) \geq 0, \quad \forall x_i, \quad \text{and}$$

$$\sum_{i=1}^{4} p(x_i) = \frac{1}{8} + \frac{3}{8} + \frac{3}{8} + \frac{1}{8} = 1.$$

Part 3.2.2: p.m.f. for discrete random variables

2. **Definition:** Suppose X is a discrete random variable and all the possible values of X are $x_1, x_2, \cdots$ (finitely or countable infinitely many number of values). Then the function p defined by

$$p(x_i) = P(X = x_i). \tag{3.2.1}$$

is called the probability mass function of X.

"Probability Mass Function" will be denoted by "p.m.f."

Part 3.2.2: p.m.f. for discrete random variables

3. Meaning: The p.m.f. of a r.v. X tells us two things:

(1) all the possible values;

(2) the probability of taking each value. Hence the p.m.f. tells us everything about the r.v. X. Namely, "p.m.f." contains all the information about the r.v. X.

4. Properties

Let $p(x_i)(i = 1, 2, \cdots)$ be the p.m.f of a r.v. X, then

(i)

$$p(x_i) \geq 0 \quad \text{for each } x_i, \tag{3.2.2}$$

(ii)

$$\sum_i p(x_i) = 1. \tag{3.2.3}$$

Proof: (i) is easy. In fact for each $x_i, p(x_i)$ is a probability and thus non-negative.

Part 3.2.2: p.m.f. for discrete random variables

To prove (ii), let's assume all the possible values of the random variable X form a set E, where E is either finite or countable infinite. Then

$$\begin{aligned}
\sum_i p(x_i) = \sum_{x_i \in E} p(x_i) &= \sum_{x_i \in E} P(X = x_i) \\
&= P(\cup_{x_i \in E}\{X = x_i\}) \quad (\sigma\text{-additive property}) \\
&= P(\Omega) \quad (\text{All the possible values of } X) \\
&= 1.
\end{aligned}$$

Note that by (i) and (ii), we also have that, $p(x_i) \leq 1$ for each x_i.

1. Example (again)

A fair coin is tossed 3 times.

X : The total number of heads appeared.

Then as above the p.m.f. is given by

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

i.e. all the possible values: $\{0, 1, 2, 3\}$ and

$$P(X = 0) = \frac{1}{8}, \quad P(X = 1) = \frac{3}{8}, \quad P(X = 2) = \frac{3}{8}, \quad P(X = 3) = \frac{1}{8}.$$

Part 3.2.3: Cumulative Distribution Function

Now, for <u>any</u> real value x, we consider $P(X \le x)$. Recall:

$$P(X \le x) = P(\{\omega \in \Omega; X(\omega) \le x\}).$$

For example,

$$P(X \le 5.4) = P(\{\omega \in \Omega; X(\omega) \le 5.4\}).$$

Suppose that

(1) $x = -2$, say, then $\{\omega \in \Omega; X(\omega) \leq -2\} = \emptyset$ (Impossible event since all the possible values are $\{0, 1, 2, 3\}$!!)

$$\Rightarrow P(X \leq -2) = P(\emptyset) = 0.$$

(2) $x = -0.34$, then, again

$$P(X \leq -0.34) = 0.$$

Part 3.2.3: Cumulative Distribution Function

(3) $x = 0$, say, then

$$\{\omega \in \Omega; X(\omega) \leq 0\} = \{\omega \in \Omega; X(\omega) = 0\}.$$

(The above is due to the fact that all the possible values of X are: $0, 1, 2, 3$)

$$\Rightarrow P(X \leq 0) = P(X = 0) = \frac{1}{8}.$$

(4) $x = 0.503$, say, then $\{X \leq 0.503\} = \{X(\omega) = 0\}$

$$\Rightarrow P(X \leq 0.503) = P(X = 0) = \frac{1}{8}.$$

Part 3.2.3: Cumulative Distribution Function

(5) Similarly, $P(X \leq 0.9999) = \frac{1}{8}$.

(6) $x = 1$, say, then $\{X \leq 1\} = \{X = 0\} \cup \{X = 1\}$

$$\begin{aligned}
\Rightarrow P(X \leq 1) &= P(\{X = 0\} \cup \{X = 1\}) \\
&= P(X = 0) + P(X = 1) \quad \text{(Think why here!!)} \\
&= \frac{1}{8} + \frac{3}{8} = \frac{4}{8} = \frac{1}{2}.
\end{aligned}$$

Similarly,

(7) $P(X \leq 1.5) = \frac{1}{2}$.

(8)

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$$
$$= \frac{1}{8} + \frac{3}{8} + \frac{3}{8} = \frac{7}{8}.$$

(9) $P(X \leq 2.99) = P(X = 0) + P(X = 1) + P(X = 2) = \frac{7}{8}$.

(10)

$$
\begin{aligned}
P(X \leq 3) &= P(X=0) + P(X=1) + P(X=2) + P(X=3) \\
&= \frac{1}{8} + \frac{3}{8} + \frac{3}{8} + \frac{1}{8} \\
&= 1.
\end{aligned}
$$

(11)

$$
\begin{aligned}
P(X \leq 1093) &= P(X=0) + P(X=1) + P(X=2) + P(X=3) \\
&= 1.
\end{aligned}
$$

In fact, for any $x \geq 3$, we have $P(X \leq x) = 1$.

Thus for any real value x, we can get a corresponding value by $P(X \leq x)$. Hence we get another function. We denote it as

$$F(x) \triangleq P(X \leq x),$$

This function $F(x)$ (again, in the ordinary meaning) is called the <u>Cumulative Distribution Function</u> of the random variables.

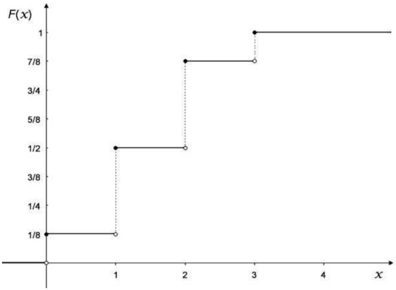
2. Graph

For the example above, $F(x)$ is actually

$$F(x) = \begin{cases} 0, & \text{if } x < 0, \\ \dfrac{1}{8}, & \text{if } 0 \leq x < 1, \\ \dfrac{1}{2}, & \text{if } 1 \leq x < 2, \\ \dfrac{7}{8}, & \text{if } 2 \leq x < 3, \\ 1, & \text{if } x \geq 3. \end{cases}$$

Part 3.2.3: Cumulative Distribution Function

So, the graph of $F(x)$ is



We can see that $F(x)$ is a "Step FunctioN".

Also, this function is non-negative and increasing, and

it jumps at 0 (with jump value $\frac{1}{8}$), 1 (with jump value $\frac{3}{8}$), 2 (with jump value $\frac{3}{8}$), and 3 (with jump value $\frac{1}{8}$),

i.e. it jumps at x whenever $p(x) > 0$ and that the jump value at x_i is $p(x_i)$.

Also

$$
\begin{aligned}
&\text{(i)} \quad 0 \leq F(x) \leq 1, \quad \forall x, \\
&\text{(ii)} \quad \lim_{x \to -\infty} F(x) = 0, \\
&\text{(iii)} \quad \lim_{x \to +\infty} F(x) = 1.
\end{aligned}
$$

Part 3.2.3: Cumulative Distribution Function

The similar properties hold true for other discrete random variables.

Also, the behaviour of the graph is also similar. (But, possibly for countable infinite number of jumps!)

We now give a general definition.

The cumulative distribution function is usually denoted by c.d.f..

Part 3.2.3: Cumulative Distribution Function

3. **Definition:** Suppose X is a random variable. The cumulative distribution function F of the random variable X is defined for all the real number x, where $-\infty < x < +\infty$, by

$$F(x) = P(X \leq x). \tag{3.2.4}$$

Note: In the above definition, different from p.m.f., we have ignored the adjective "discrete". The reason for this will be clear later.

4. Properties:

Suppose X is a random variable and $F(x)$ is the Cumulative Distribution Function (c.d.f.) of X defined by (2.3.4), i.e.,

$$F(x) = P\{X \leq x\}, \quad \forall x \in R.$$

Then the c.d.f. $F(x)$ possesses the following properties:

(i) $0 \leq F(x) \leq 1, \quad \forall x \in R.$

(ii) $F(x)$ is a nondecreasing function of X, that is,

$$F(x_1) \leq F(x_2), \quad \forall x_1 \leq x_2.$$

(iii) $\lim_{x \to +\infty} F(x) = 1$.

(iv) $\lim_{x \to -\infty} F(x) = 0$.

(v) $F(x)$ is right continuous, i.e., $\lim_{y \downarrow x} F(y) = F(x)$ for all x.

Part 3.2.3: Cumulative Distribution Function

Proof: (i) is clear since $F(x)$ is a probability.

(ii) If $x_1 < x_2$, then $\{\omega : X(\omega) \le x_1\} \subset \{\omega : X(\omega) \le x_2\}$, we have $P\{X \le x_1\} \le P\{X \le x_2\}$. That is, $F(x_1) \le F(x_2)$.

(iii) It is equivalent to showing that for any increasing sequence $\{x_n\}$ which tends to $+\infty$, $\lim_{n\to\infty} F(x_n) = 1$.

Now let $\{x_n\}$ be an increasing sequence $\{x_n\}$ which tends to $+\infty$, and let

$$A_n = \{X \le x_n\}.$$

Then $\{A_n\}$ is an increasing sequence of events. So

$$\lim_{n\to\infty} F(x_n) = \lim_{n\to\infty} P(A_n) = P(\lim_{n\to\infty} A_n) = P(\cup_{n=1}^{\infty} A_n).$$

But we can show that

$$\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} (X \leq x_n) = \{X < \infty\} = \Omega.$$

Hence,

$$\lim_{n\to\infty} F(x_n) = P(\cup_{n=1}^{\infty} A_n) = P(\Omega) = 1.$$

(iv) can be similarly proved. In short, $\lim_{x\to-\infty} F(x) = 0$ is equivalent to the statement that for any decreasing sequence $\{x_n\}$ such that $x_n \downarrow -\infty$, we have $\lim_{n\to\infty} F(x_n) = 0$.

Part 3.2.3: Cumulative Distribution Function

(v) Suppose $x_n \downarrow x$, then

$$B_n = \{X \leq x_n\}, \quad n \geq 1$$

are decreasing events whose intersection is just $\{X \leq x\}$ (Verify this yourself !!).

Hence the continuity property of the probability measure yields that

$$\lim_{n \to \infty} P\{X \leq x_n\} = P\{X \leq x\},$$

i.e. $x_n \downarrow x$ implies that

$$\lim_{n \to \infty} F(x_n) = F(x), \quad \forall x \in R.$$

That is that F(x) is a right-continuous function of x.

Part 3.2.4: Relation Between "p.m.f." and "c.d.f."

Recall : For a discrete random variable X,

$$\text{p.m.f. : } p(x_i) = P(X = x_i),$$

$$\text{c.d.f. : } F(x) = P(X \leq x).$$

There exists a close link between them.

For simplicity, let's just consider the case that all the possible values of the r.v. X are non-negative integers $\{0, 1, 2, 3, 4, 5, \cdots\}$.

1. If we know p.m.f., i.e. $p(0), p(1), \cdots$

$$\Rightarrow \text{ c.d.f. } F(x) = P(X \leq x) = \sum_{x_i \leq x} p(x_i).$$

For example,

$$F(17.1) = p(0) + p(1) + \cdots + p(17) = \sum_{k=0}^{17} p(k),$$

$$F(14) = p(0) + \cdots + p(14) = \sum_{k=0}^{14} p(k).$$

2. If we know c.d.f., i.e. $F(x)$ is known for each $x \in R$, then

$$\Rightarrow p(x_i) \text{ ??}$$

(This is usually the case in application !!)

Suppose we want to get $p(17) = P(X = 17)$, say, then

$$\begin{aligned} p(17) &= P(X = 17) \\ &= P(X \leq 17) - P(X \leq 16) \\ &= F(17) - F(16). \end{aligned}$$

Part 3.2.4: Relation Between "p.m.f." and "c.d.f."

Similarly, for example,

$$p(304) = F(304) - F(303).$$

The reason for the above key step is, for example,

$$(X \le 17) = (X \le 16) \cup (X = 17)$$
$$\Rightarrow P(X \le 17) = P(X \le 16) + P(X = 17).$$

(Think why here !!)

$$\Rightarrow P(X = 17) = P(X \le 17) - P(X \le 16),$$

i.e. $p(17) = F(17) - F(16)$.

Part 3.2.5: Worked Questions

1. **Exercise 1**: The p.m.f. of a discrete r.v. X is given by

$$p_i = c \cdot \frac{\lambda^i}{i!} \quad (i = 0, 1, 2, \cdots),$$

where $\lambda > 0$ is a known positive value and c is an unknown constant. Find $P\{X = 0\}$, and $P\{X > 2\}$.

Solution: $\sum_{i=0}^{\infty} p_i = 1 \Rightarrow c\sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = 1$.

But we know $e^x = \sum_{i=0}^{\infty} \frac{x^i}{i!} \Rightarrow \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = e^{\lambda}$.

Hence $ce^{\lambda} = 1, c = e^{-\lambda}$.

p.m.f.: $p_i = e^{-\lambda} \cdot \frac{\lambda^i}{i!}$.

So, $P\{X = 0\} = p_0 = e^{-\lambda} \times \frac{\lambda^0}{0!} = e^{-\lambda}$.

And

$$P\{X > 2\} = \sum_{i=3}^{\infty} p_i, \quad \text{(but this will cause some difficulty)}.$$

Part 3.2.5: Worked Questions

However,

$$\{X > 2\}^c = \{X \leq 2\} = \{X = 0\} \cup \{X = 1\} \cup \{X = 2\}.$$

Therefore

$$\begin{aligned} P\{X > 2\} &= 1 - P\{X \leq 2\} \\ &= 1 - P\{X = 0\} - P\{X = 1\} - P\{X = 2\} \\ &= 1 - e^{-\lambda} - e^{-\lambda} \cdot \frac{\lambda^1}{1!} - e^{-\lambda} \cdot \frac{\lambda^2}{2!} \\ &= 1 - e^{-\lambda} - \lambda e^{-\lambda} - \frac{\lambda^2}{2} e^{-\lambda} \\ &= 1 - e^{-\lambda}\left(1 + \lambda + \frac{\lambda^2}{2}\right). \end{aligned}$$

2. **Exercise 2**: Suppose X is a discrete r.v. whose p.m.f. is given by

$$p_1 = \frac{1}{4}, \quad p_2 = \frac{1}{2}, \quad p_3 = \frac{1}{8}, \quad p_4 = \frac{1}{8},$$

(i.e. All the possible values are $\{1, 2, 3, 4\}$)

(i) Find the c.d.f. of $X, F(x)$ say.

(ii) Plot the c.d.f..

Part 3.2.5: Worked Questions

Solution:

(1) If $x < 1$, then

$$F(x) = P\{X \leq x\} = P(\emptyset) = 0.$$

(2) If $1 \leq x < 2$, then

$$F(x) = P\{X \leq x\} = P\{X = 1\} = \frac{1}{4}.$$

(3) If $2 \leq x < 3$, then

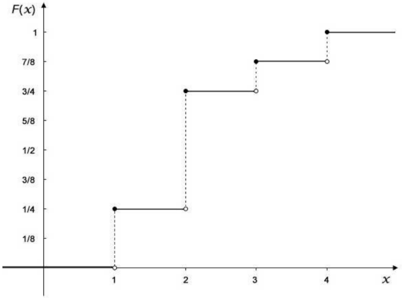
$$F(x) = P\{X \leq x\} = P\{X = 1\} + P\{X = 2\} = \frac{1}{4} + \frac{1}{2} = \frac{3}{4}.$$

(4) If $3 \le x < 4$, then

$$F(x) = P\{X \le x\} = P\{X = 1\} + P\{X = 2\} + P\{X = 3\}$$
$$= \frac{1}{4} + \frac{1}{2} + \frac{1}{8} = \frac{6}{8} + \frac{1}{8} = \frac{7}{8};$$

(5) If $x \ge 4$, then $F(x) = P\{X \le x\} = \sum_{i=1}^{4} P\{X = i\} = 1$.

The graph of F(x) is



Outline

Part 3.3.1: The Bernoulli Random Variable

1. **Definition:** A Bernoulli random variable takes on only two values: 0 and 1, with probabilities $1 - p$ and p, respectively, where

$$0 < p < 1. \tag{3.3.1}$$

2. p.m.f.: Bernoulli random variable X :

x_i	0	1
$p(x_i)$	$1-p$	p

i.e.

$$p(0) = P(X = 0) = 1 - p \tag{3.3.2}$$

3. Note: The p.m.f. of a Bernoulli random variable X satisfies:

$$p(0) \geq 0, \quad p(1) \geq 0, \quad p(0) + p(1) = 1.$$

Part 3.3.2: The Binomial Random Variables

1. **Definition:** A random variable is called a Binomial random variable, or, (more often) the r.v. X obeys the Binomial distribution, if all the possible values of X are

$$\{0, 1, 2, 3, \cdots, n\}$$

where n is a positive integer $(n \geq 1)$ and that

$$p(k) = P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad (3.3.4)$$

where $0 < p < 1$. (Trivial, if $p = 0$ or $p = 1$).

2. p.m.f.: Usually we let $q = 1 - p$ (and since $0 < p < 1$ we also have $0 < q < 1$),

x_i	0	1	2	$\cdots$	n
$p(x_i)$	q^n	$\binom{n}{1} pq^{n-1}$	$\binom{cn}{2} p^2 q^{n-2}$	$\cdots$	p^n

i.e.

$$p(k) = \binom{n}{k} p^k q^{n-k} = P(X = k), \tag{3.3.5}$$

where $k = 0, 1, 2, \cdots, n$ and $0 < p < 1, 0 < q < 1$ and $p + q = 1$

Part 3.3.2: The Binomial Random Variables

3. Notes:

(1) By (3.3.5) we again have $p(k) \geq 0$ for all $k = 0, 1, \cdots, n$, and

$$\sum_{k=0}^{n} p(k) = 1, \tag{3.3.6}$$

Indeed, $\sum_{k=0}^{n} p(k) = \sum_{k=0}^{n} \binom{k}{n} p^k q^{n-k} = (p+q)^n = 1.$

(2) If $n = 1$, we return to the Bernoulli r.v..

4. Notation:

For a Binomial r.v. X, the distribution depends upon two parameters: n and p, so, usually we write it as $B(n, p)$. Thus:

"X is a Binomial r.v. with parameters n and p"

$$\begin{aligned}
&\Longleftrightarrow \text{"}X \text{ obeys the Binomial distribution with parameters } n \text{ and } p\text{"}\\
&\Longleftrightarrow \text{"}X \sim B(n, p)\text{"}\\
&\Longleftrightarrow \text{"}p(k) = P(X = k) = \binom{n}{k} p^k q^{n-k}, k = 0, 1, 2, \cdots, n.\text{"}
\end{aligned}$$

Part 3.3.2: The Binomial Random Variables

5. Classic situation:

Suppose that the Bernoulli independent trials, each having a probability p $(0 < p < 1)$ of being a success, are performed n times.

Let X denote the number of successes recorded. Then $X \sim B(n, p)$.

Reason: Clear! Think yourself.

Part 3.3.2: The Binomial Random Variables

Hint: To get the values of $P\{X = k\}$, just note the following:

(1) First, "k successes in n trials" implies "k successes and $n-k$ failures in n trials".

(2) Secondly, assume that, for example, the first k trials are successes and the followed $n-k$ trials are failures, then by independence we know that the probability is

$$p^k(1-p)^{n-k} = p^k q^{n-k}.$$

Part 3.3.2: The Binomial Random Variables

(3) Thirdly, among n trials, the different ways of choosing k successes are $\binom{n}{k}$. But the different ways are "disjoint" events and hence by additive property, there are altogether $\binom{n}{k}$ terms with each term having the same probability of $p^k q^{n-k}$.

(4) Finally we get, for all $k = 0, 1, 2, \cdots, n$, we have

$$p(k) = P(X = k) = \binom{n}{k} p^k q^{n-k}.$$

Hence we get the Binomial distribution.

1. **Definition:**

We say that a r.v. X obeys the Poisson distribution if all the possible values of X are non-negative integers $\{0, 1, 2, 3, \cdots\}$ and the p.m.f. takes the form of

$$p(k) = P(X = k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!}, \tag{3.3.7}$$

where $k = 0, 1, 2, 3 \cdots$, and $\lambda > 0$ is a constant.

Note that, again, we have

$$p_k \geq 0, \quad \forall k, \quad \text{and} \quad \sum_{k=0}^{\infty} p_k = 1. \tag{3.3.8}$$

Indeed,

$$\sum_{k=0}^{\infty} p_k = \sum_{k=0}^{\infty} e^{-\lambda} \cdot \frac{\lambda^k}{k!} = e^{-\lambda} \cdot \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} \cdot e^{\lambda} = e^0 = 1.$$

2. Applications:

The most important discrete random variable:

(1) The number of telephone calls received in a time interval.

(2) The number of radioactive particles observed in a time interval, (or, the number of other kind of particles.)

(3) Queueing theory: the length of some kind of queues.

(4) Traffic studies: the number of cars, say, arrived at a junction.

(5) $\cdots$

3. Parameter:

The constant $\lambda > 0$ in (3.3.7) is called the parameter of the Poisson distribution (Also called: the mean; density; etc).

The meaning of λ will be clear later.

3. Notation:

Note that the Poisson distribution depends upon the parameter λ.

"X obeys the Poisson distribution with parameter λ"

$\Longleftrightarrow$ "the r.v. X has a Poisson distribution with parameter λ"

$\Longleftrightarrow$ "$X \sim \text{Poisson}(\lambda)$"

4. c.d.f. and the table:

If $X \sim \text{Poisson}(\lambda)$, then the c.d.f. of X :

$$F_X(x) = P(X \leq x) = \sum_{k \leq x} P(X = k) = \sum_{k \leq x} e^{-\lambda} \cdot \frac{\lambda^k}{k!}.$$

For example,

$$\begin{aligned} F_X(8) &= \sum_{k \leq 8} e^{-\lambda} \cdot \frac{\lambda^k}{k!} = e^{-\lambda} \cdot \sum_{k=0}^{8} \frac{\lambda^k}{k!} \\ &= e^{-\lambda} \left(1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \cdots + \frac{\lambda^8}{8!} \right). \end{aligned}$$

Part 3.3.3: The Possion Random Variable

Also, say

$$F_X(3.52) = \sum_{k \le 3.52} e^{-\lambda} \cdot \frac{\lambda^k}{k!} = \sum_{k=0}^{3} e^{-\lambda} \cdot \frac{\lambda^k}{k!}$$
$$= e^{-\lambda} \left(1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} \right).$$

The value of $F(x)$ can be obtained by checking the tables, for example, presented in our reference book. See Table III Pages 610-612.

Part 3.3.4: The Geometric Random Variable

The independent Bernoulli trials each having a probability p $(0 < p < 1)$ of being success are performed until a success occurs.

Let X be the number of trials required.

Then the p.m.f. of X is as follows: Possible values: $\{1, 2, 3, 4, \cdots\}$

$$\begin{aligned}
p(1) &= P\{X = 1\} = p \\
p(2) &= P\{X = 2\} \\
&= P\{\text{1st failure and 2nd success}\} \\
&= P\{\text{1st failure}\} \cdot P\{\text{2nd success}\} \\
&= (1 - p)p \\
&\cdots\cdots
\end{aligned}$$

The last second equality holds since trials are independent.

Part 3.3.4: The Geometric Random Variable

$$
\begin{aligned}
p(n) &= P\{X = n\} \\
&= P\{\text{the first } (n-1) \text{ times failure and the } n\text{th success}\} \\
&= P\{\text{the first } (n-1) \text{ times failure}\} \cdot P\{\text{the } n\text{th success}\} \\
&= P\{\text{the 1st failure}\} \cdots P\{\text{the } (n-1)\text{th failure}\} \\
&\quad \cdot P\{\text{the } n\text{th success}\} \\
&= \underbrace{(1-p)\cdots(1-p)}_{n-1} p \\
&= (1-p)^{n-1} p.
\end{aligned}
$$

Part 3.3.4: The Geometric Random Variable

Therefore, the p.m.f. of X is:

$$p(n) = P(X = n) = (1-p)^{n-1}p, \quad n = 1, 2, \ldots$$

Does the equality $\sum_{n=1}^{\infty} p_n = 1$ hold? Yes. Because

$$\begin{aligned}\sum_{n=1}^{\infty} p_n &= \sum_{n=1}^{\infty} q^{n-1}p = p \cdot \sum_{n=1}^{\infty} q^{n-1} = p \cdot \sum_{n=0}^{\infty} q^n \\ &= p \cdot \frac{1}{1-q} (\because 0 < q < 1) = p \cdot \frac{1}{p} = 1. \quad \checkmark\checkmark\end{aligned}$$

Part 3.3.5: The Negative Binomial Random Variable

Suppose that the independent Bernoulli trials, each having a probability p $(0 < p < 1)$, of being a success, are performed until a total of r successes is accumulated, where $r \geq 1$ is a positive integer.

Let X be the number of trials required.

The possible values of X are $\{r, r+1, r+2, \ldots\}$. Let

$$A_n = \{\text{the } n\text{th being a success}\},$$
$$B = \{r-1 \text{ successes in the first } n-1 \text{ times independent trials}\}.$$

Part 3.3.5: The Negative Binomial Random Variable

$$p(n) = P\{X = n\} = P(A_n \cap B) = P(B)P(A_n) = P(B)p$$

$$= \binom{n-1}{r-1} p^{r-1}(1-p)^{(n-1)-(r-1)}p$$

(Since Binomial distribution with parameters $n-1$ and p)

$$= \binom{n-1}{r-1} p^{r}(1-p)^{n-r}.$$

Part 3.3.5: The Negative Binomial Random Variable

Therefore, the p.m.f. of X is:

$$p(n) = P(X = n) = \binom{n-1}{r-1} p^r (1-p)^{n-r}, \quad n = r, r+1, \ldots$$

Note: By the property of p.m.f., we know that $\sum_{n=r}^{\infty} P\{X = n\} = 1$, i.e.

$$\sum_{n=r}^{\infty} \binom{n-1}{r-1} p^r (1-p)^{n-r} = 1.$$

Interestingly, we have given a probabilistic proof for the above equality.

Outline

Part 3.4.1: Concept

1. Example: Consider the "lifetime" of an electronic component. Then it is a r.v.. For this r.v., any positive real number is possible, i.e., all the possible values are $(0, +\infty)$ (or $[0, +\infty)$!!). Such a r.v. is not "discrete", we call it "continuous".

2. A random variable is called a <u>continuous</u> random variable if all the possible values are neither finite nor countable infinite, usually take an interval (finite or infinite) of real numbers, such as $[a, b]$, $[0, +\infty)$, or $(-\infty, +\infty)$.

Part 3.4.2: Cumulative Distribution Functions

1. **Definition** (similar to discrete r.v.): Let X be a continuous random variable, then the function F defined by

$$F(x) \triangleq P(X \leq x), \quad x \in \mathbb{R}$$

   is called the cumulative distribution function (c.d.f.) of X.

2. Recall: $(X \leq x) = \{\omega \in \Omega; X(\omega) \leq x\}$ is an event for any $x \in (-\infty, +\infty)$ and thus has the "probability".

Part 3.4.2: Cumulative Distribution Functions

3. Properties:

   (i) $0 \leq F(x) \leq 1, \ \forall x.$

   (ii) $F(x)$ is a non-decreasing (increasing) function of $x \in \mathbb{R}$.

   (iii) $\lim_{x \to \infty} F(x) = 1.$

   (iv) $\lim_{x \to -\infty} F(x) = 0.$

   The above properties are the same for both "discrete r.v.'s" and "continuous r.v.'s".

   The proofs are the same as provided before for discrete r.v.'s.

   In particular, properties (iii) and (iv) are just the "limit" properties of the probability measure.

Part 3.4.3: Probability Density Function

1. **Definition:** Let $F(x)$ be the c.d.f. of a continuous r.v. X. If there exists a function $f(x) \geq 0$ such that

$$F(x) = \int_{-\infty}^{x} f(y)dy, \quad \forall x \in \mathbb{R},$$

then we call $f(x)$ a probability density function (p.d.f.) of X.

In this course, we only consider continuous r.v.'s that have a probability density function.

Part 3.4.3: Probability Density Function

2. Important facts:

   (i) The p.d.f. of a r.v. X is not unique in the usual sense!

   (ii) If $f(x)$ is a p.d.f. of X, then $\int_{-\infty}^{+\infty} f(x)dx = 1$.

   (iii) Relation between c.d.f. and p.d.f.:

$$F'(x) = f(x), \quad F(x) = \int_{-\infty}^{x} f(y)dy.$$

Part 3.4.3: Probability Density Function

3. Remarks:

   (1) Although the properties regarding p.d.f. for a continuous random variable and those properties regarding p.m.f. for a discrete random variable are very similar, the reason is quite different.

   Indeed, for example, $p(x) \geq 0$ is due to the fact that the p.m.f. $p(x)$ is a probability, while the reason for $f(x) \geq 0$ for a p.d.f. is that it is a <u>derivative of an increasing function</u> of the c.d.f. $F(x)$.

(2) It is important to note that, although the c.d.f. of a random variable is unique, the p.d.f. is not unique.

Indeed, changing a density at a finite set (even countable!!) of points does not change the p.d.f.. Hence, strictly speaking, it is incorrect to speak of the density of a r.v.; rather, we should refer to a density.

But we shall not pay much attention to this point. In many cases, we still call the p.d.f. of a random variable.

We now use an example to illustrate this point.

Part 3.4.3: Probability Density Function

Example: Suppose the c.d.f. of a certain continuous r.v. is given by

$$F(x) = \begin{cases} 0, & \text{if } x < 0, \\ x^2, & \text{if } 0 \leq x < 1, \\ 1, & \text{if } x \geq 1. \end{cases}$$

What is a p.d.f. of this r.v.?

It is clear that $F(x)$ is differentiable everywhere except possibly at points $\{0, 1\}$ (In fact, it has a derivative at zero).

For this example, it is easy to see, (except at points $\{0, 1\}$) we have

$$F'(x) = \begin{cases} 0, & \text{if } x < 0, \\ 2x, & \text{if } 0 < x < 1, \\ 0, & \text{if } x > 1. \end{cases}$$

Part 3.4.3: Probability Density Function

Then how about the points $\{0,1\}$? We could assign any value!!! (But better to be non-negative!!)

Suppose we assign $f(0) = f(1) = 0$, then a $p.d.f.$ is given by

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

Another p.d.f. (for the same r.v. with the unique c.d.f. $F(x)$ as above) could be

$$g(x) = \begin{cases} 2x, & \text{if } 0 \le x \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

Note: At point 1, $f(1) = 0$, but $g(1) = 2$.

Part 3.4.3: Probability Density Function

We could even give another density by changing the values at $x = \frac{1}{7}$, 1, and 6, say, as follows

$$h(x) = \begin{cases} 2x, & \text{if } 0 < x < 1, \text{ but } x \neq \dfrac{1}{7}, \\ 3, & \text{if } x = \dfrac{1}{7}, \\ \pi, & \text{if } x = 1, \\ 17.6, & \text{if } x = -6, \\ 0, & \text{otherwise.} \end{cases}$$

The three densities are, of course, different, but they have the same c.d.f. $F(x)$ defined above. Remember: Any r.v. has just one c.d.f.

Part 3.4.3: Probability Density Function

4. Notes:

(1) By the relation $F(x) = \int_{-\infty}^{x} f(y)dy$, we have for any $a \leq b$,

$$F(b) - F(a) = \int_a^b f(y)dy = \int_a^b f(x)dx.$$

Indeed, by the simple property of the integral, if $a \leq b$, then

$$\begin{aligned} F(b) - F(a) &= \int_{-\infty}^{b} f(y)dy - \int_{-\infty}^{a} f(y)dy \\ &= \int_{-\infty}^{a} f(y)dy + \int_a^b f(y)dy - \int_{-\infty}^{a} f(y)dy \\ &= \int_a^b f(y)dy = \int_a^b f(x)dx. \end{aligned}$$

(2) The property and role of the p.d.f. in continuous r.v.'s are similar to the p.m.f. in discrete r.v.'s.

However, an important difference is that, for continuous r.v.'s,

$$f(x) \neq P(X = x)$$

In fact, later we shall prove that for a continuous random variable X, we have

$$P(X = x) = 0, \quad \forall x \in (-\infty, +\infty).$$

Part 3.4.4: Basic Formula

1. For any r.v. X (discrete or continuous), let $F(x)$ be the c.d.f. of X, then for any $a < b$,

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

Proof. First recall that here, for $a < b$

$$(a < X \leq b) = \{\omega \in \Omega; a < X(\omega) \leq b\}.$$

But it is easy to see (since $a < b$)

$$\begin{aligned} &\{\omega \in \Omega; X(\omega) \leq b\} \\ &\quad = \{\omega \in \Omega; X(\omega) \leq a\} \cup \{\omega \in \Omega; a < X(\omega) \leq b\}, \end{aligned}$$

i.e.

$$(X \leq b) = (X \leq a) \cup (a < X \leq b).$$

Part 3.4.4: Basic Formula

However, it is easy to see that

$$(X \leq a) \cap (a < X \leq b) = \emptyset.$$

Thus by the property of the probability measure,

$$\begin{aligned} P(X \leq b) &= P(X \leq a) + P(a < X \leq b) \\ \Rightarrow P(a < X \leq b) &= P(X \leq b) - P(X \leq a) \end{aligned}$$

Now, by the definition $P(X \leq b) = F(b)$ and $P(X \leq a) = F(a)$,

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

2. For a continuous r.v. X, let $F(x)$ and $f(x)$ be its c.d.f. and p.d.f., respectively, then for any $a < b$,

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

Part 3.4.4: Basic Formula

3. Let X be a continuous r.v.. For any real value b,

$$P(X = b) = 0.$$

Intuitively, it is clear that

$$P(X = b) = \lim_{n\to\infty} P\left(b - \frac{1}{n} < X \leq b\right),$$

so, $P(X = b) = \lim_{n\to\infty} \int_{b-\frac{1}{n}}^{b} f(x)dx = \int_{b}^{b} f(x)dx = 0.$

Recall, $P(X = b) = 0$ is not true for discrete random variables!!

4. Finally, we get the Basic Formula as follows: Suppose X is a continuous r.v. with probability density function $f(x)$, then for any $-\infty < a < b < +\infty$,

$$\begin{aligned} P(a < X \leq b) &= P(a \leq X \leq b) = P(a \leq X < b) \\ &= P(a < X < b) = \int_a^b f(x)dx. \end{aligned}$$

That is, the probability of the four intervals: $[a, b]$, $[a, b)$, $(a, b]$, (a, b) is the same.

This is only true for a continuous random variables.

From this formula, we see that $f(x) \neq P(X = x)$ in general.

5. How about $P(X > a)$, for example?

$$P(X > a) = P(a < X) = P(a < X < +\infty) = \int_a^{+\infty} f(x)dx.$$

Also, $P(X \geq a) = \int_a^{+\infty} f(x)dx$ (Why?)

Further, comparing with the definition of the c.d.f. $F(x)$, we get

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

Part 3.4.4: Basic Formula

that is,

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

Also,

$$P(X < a) = P(-\infty < X < a) = \int_{-\infty}^{a} f(x)dx.$$

Part 3.4.4: Basic Formula

6. In short, just remember that for any $a < b$, (a, b can be $\pm\infty$)

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

together with the fact that (if a, b are both finite)

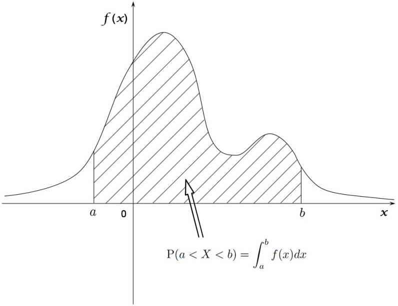
$$\begin{aligned} P(a < X < b) &= P(a \leq X < b) \\ &= P(a < X \leq b) = P(a \leq X \leq b). \end{aligned}$$

Part 3.4.5: Geometric Meaning

For a continuous r.v., a p.d.f. is more basic. The formula

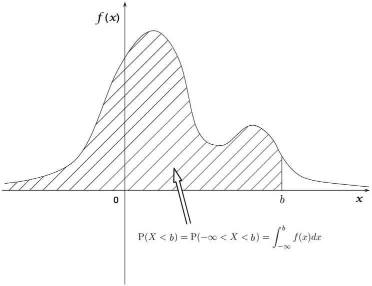
$$P(a < X < b) = \int_a^b f(x)dx$$

tells us that if X is a continuous r.v. with p.d.f. $f(x)$, then $P(a < X < b)$ is just the areas of the following shaded ones.



Part 3.4.5: Geometric Meaning

Similarly,



Now, the geometric meaning of $\int_{-\infty}^{+\infty} f(x)dx = 1$ should be clear.

Part 3.4.6: Worked Exercises

Example 1. A continuous r.v. X whose p.d.f. is given by

$$f(x) = \begin{cases} c(3-x), & \text{if } 0 < x < 3, \\ 0, & \text{otherwise.} \end{cases}$$

(1) Find the value of c.

(2) What is the c.d.f. of this r.v.?

Solution:

(1) $\int_{-\infty}^{+\infty} f(x)dx = 1$, i.e., $\int_0^3 c(3-x)dx = 1$,

$$\Rightarrow c\left[\int_0^3 (3-x)dx\right] = 1 \quad \Rightarrow \quad c = \frac{2}{9}.$$

(2) $F(x) = P\{X \leq x\} = \int_{-\infty}^{x} f(y)dy$.

If $x < 0$, then $\int_{-\infty}^{x} f(y)dy = \int_{-\infty}^{x} 0 \cdot dy = 0$;

If $0 \leq x < 3$, then

$$\int_{-\infty}^{x} f(y)dy = \int_{0}^{x} f(y)dy = \int_{0}^{x} \frac{2}{9}(3-y)dy = 1 - \frac{1}{9}(3-x)^2;$$

If $x \geq 3$, then $F(x) = 1$. Therefore,

$$F(x) = \begin{cases} 0, & \text{if } x < 0, \\ 1 - \dfrac{1}{9}(3-x)^2, & \text{if } 0 \leq x < 3, \\ 1, & \text{if } x \geq 3. \end{cases}$$

Example 2. Consider the r.v. X whose c.d.f. is

$$F(x) = \begin{cases} 0, & \text{if } x < 0, \\ 1 - e^{-2x} - 2xe^{-2x}, & \text{if } x \geq 0. \end{cases}$$

Find a p.d.f. of X.

Solution: Clearly, we may differentiate this function everywhere except perhaps at $x = 0$.

If $x < 0$, then $F'(x) = 0$.

If $x > 0$, then $F'(x) = 4xe^{-2x}$.

As to $x = 0$, we could let it to be 0. Hence, a density is given by

$$f(x) = \begin{cases} 4xe^{-2x}, & \text{if } x > 0, \\ 0, & \text{otherwise.} \end{cases}$$

Outline

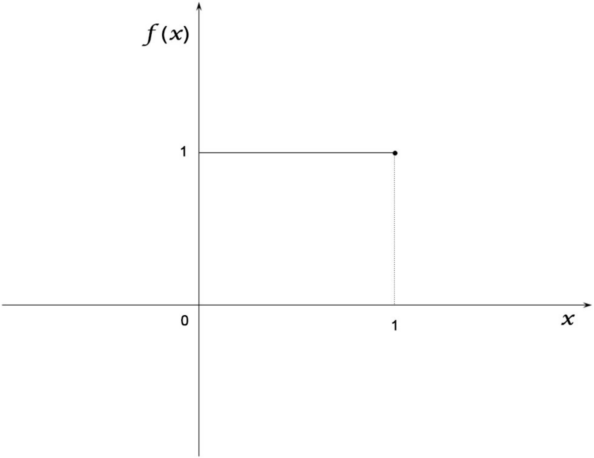
Part 3.5.1: Uniform Random Variable

1. **Definition:** A r.v. X is said to be uniformly distributed over the interval $[0,1]$, if its p.d.f. $f(x)$ is given by

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

2. Check: $f(x) \geq 0$ and $\int_{-\infty}^{+\infty} f(x)dx = 1$.

3. Graph of p.d.f. f(x):



4. c.d.f. $F(x)$: Recall: $F(x) = \int_{-\infty}^{x} f(y)dy$,

   If $x < 0$, then

$$F(x) = \int_{-\infty}^{x} f(y)dy = \int_{-\infty}^{x} 0 \cdot dy = 0;$$

   if $0 \leq x \leq 1$, then

$$F(x) = \int_{-\infty}^{x} f(y)dy = \int_{-\infty}^{0} f(y)dy + \int_{0}^{x} f(y)dy$$
$$= \int_{-\infty}^{0} 0dy + \int_{0}^{x} 1dy = x;$$

Part 3.5.1: Uniform Random Variable

If $x > 1$, then

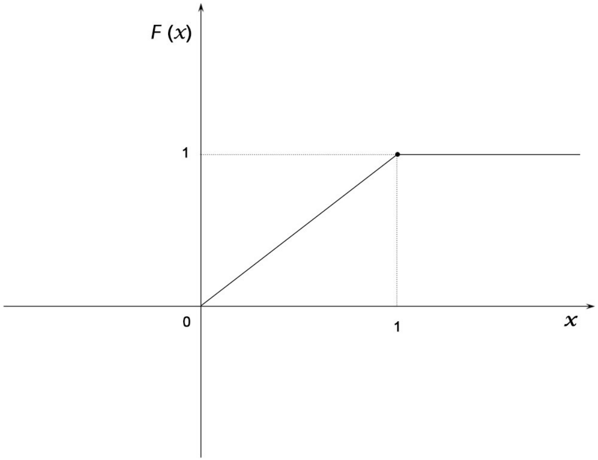
$$F(x) = \int_{-\infty}^{x} f(y)dy = \int_{-\infty}^{0} f(y)dy + \int_{0}^{1} f(y)dy + \int_{1}^{x} f(y)dy$$
$$= \int_{-\infty}^{0} 0dy + \int_{0}^{1} 1dy + \int_{1}^{x} 0dy = 1.$$

In short,

$$F(x) = \begin{cases} 0, & \text{if } x \leq 0, \\ x, & \text{if } 0 \leq x \leq 1, \\ 1, & \text{if } x > 1. \end{cases}$$

Part 3.5.1: Uniform Random Variable

Graph:



Part 3.5.1: Uniform Random Variable

5. **General case**: A r.v. X is said to be uniformly distributed over the interval $[a, b]$, where $-\infty < a < b < +\infty$, if its p.d.f. $f(x)$ is given by

$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{if } a \le x \le b, \\ 0, & \text{otherwise.} \end{cases}$$

Try yourself to obtain the form of the c.d.f. of this general case.

Part 3.5.2: The Exponential Distribution

1. **Definition:** A r.v. X is said to be exponentially distributed with parameter λ, if its p.d.f. $f(x)$ is given by

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

   where the constant λ satisfies $\lambda > 0$.

2. Check: $f(x) \geq 0$ and $\int_{-\infty}^{+\infty} f(x)dx = 1$.

3. Graph: Depending upon the parameter λ. See Figure 2-9 in the Book (Page 47, the 1st edition).

4. c.d.f. $F(x)$: Recall: $F(x) = \int_{-\infty}^{x} f(y)dy$.

   If $x < 0$, then

$$\int_{-\infty}^{x} f(y)dy = \int_{-\infty}^{x} 0 \cdot dy = 0.$$

   If $x \geq 0$, then

$$\begin{aligned} F(x) &= \int_{-\infty}^{x} f(y)dy = \int_{-\infty}^{0} f(y)dy + \int_{0}^{x} f(y)dy \\ &= \int_{-\infty}^{0} 0dy + \int_{0}^{x} \lambda e^{-\lambda y} dy = 1 - e^{-\lambda x}. \end{aligned}$$

   Hence,

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & \text{if } x \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

1. **Definition:** A r.v. X is said to follow a gamma distribution with parameter (α, λ) where $\alpha > 0, \lambda > 0$, if its p.d.f. is given by

$$f(x) = \begin{cases} \frac{\lambda e^{-\lambda x} \cdot (\lambda x)^{\alpha-1}}{\Gamma(\alpha)}, & x \geq 0, \\ 0, & x < 0, \end{cases}$$

where $\Gamma(\alpha)$, called the gamma function, is defined as

$$\Gamma(\alpha) = \int_0^\infty e^{-y} \cdot y^{\alpha-1} dy.$$

(Hence, X can only take non-negative values)

Part 3.5.3: The Gamma Distribution

Easy to see:

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

Indeed, letting $\lambda x = y$ (and noting $\lambda > 0$) yields

$$\int_0^\infty \lambda e^{-\lambda x}(\lambda x)^{\alpha-1}dx = \int_0^\infty e^{-y} \cdot y^{\alpha-1}dy = \Gamma(\alpha).$$

2. Some simple facts about the gamma function

$$\Gamma(\alpha) = \int_0^\infty e^{-y} y^{\alpha-1} dy, \quad (\alpha > 0).$$

The integration by parts of $\Gamma(\alpha)$ yields that for $\alpha > 1$

$$\begin{aligned}\Gamma(\alpha) &= -\int_0^\infty y^{\alpha-1} de^{-y} = -\, y^{\alpha-1} e^{-y}\Big|_0^\infty + \int_0^\infty e^{-y} dy^{\alpha-1} \\ &= \int_0^\infty e^{-y} dy^{\alpha-1} = \int_0^\infty e^{-y} \cdot (\alpha-1) y^{\alpha-1-1} dy \\ &= (\alpha-1) \int_0^\infty e^{-y} y^{\alpha-1-1} dy \\ &= (\alpha-1)\Gamma(\alpha-1).\end{aligned}$$

Part 3.5.3: The Gamma Distribution

If $\alpha > 2$ we could continue

$$\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1) = (\alpha - 1)(\alpha - 2)\Gamma(\alpha - 2).$$

If $\alpha > 3$, we could get

$$\Gamma(\alpha) = (\alpha - 1)(\alpha - 2)(\alpha - 3)\Gamma(\alpha - 3),$$

etc. By repeated applying of the recursion formula

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

one can reduce the form of $\Gamma(\alpha)$ until $\Gamma(x)$ for $0 < x \leq 1$.

Part 3.5.3: The Gamma Distribution

In particular, if $\alpha = n$ (a positive integer) then

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

Now

$$\begin{aligned}\Gamma(1) &= \int_0^\infty e^{-y} y^{1-1} dy \quad \left(\because \Gamma(\alpha) = \int_0^\infty e^{-y} y^{\alpha-1} dy\right) \\ &= \int_0^\infty e^{-y} dy = 1,\end{aligned}$$

$\Rightarrow \Gamma(n) = (n-1)!$.

Part 3.5.3: The Gamma Distribution

If $\alpha = n + \frac{1}{2}$ (n: positive integer), then

$$\begin{aligned}
\Gamma\left(n+\frac{1}{2}\right) &= \left(n+\frac{1}{2}-1\right)\Gamma\left(n+\frac{1}{2}-1\right) \\
&= \frac{2n-1}{2}\Gamma\left(\frac{2n-1}{2}\right) \\
&= \frac{2n-1}{2}\cdot\frac{2n-3}{2}\Gamma\left(\frac{2n-3}{2}\right) = \cdots \\
&= \frac{2n-1}{2}\cdot\frac{2n-3}{2}\cdots\frac{1}{2}\cdot\Gamma\left(\frac{1}{2}\right).
\end{aligned}$$

In short, for a positive integer n

$$\Gamma\left(n+\frac{1}{2}\right) = \frac{(2n-1)!!}{2^n}\Gamma\left(\frac{1}{2}\right).$$

Part 3.5.3: The Gamma Distribution

What is $\Gamma\left(\frac{1}{2}\right)$? By definition,

$$\Gamma\left(\frac{1}{2}\right) = \int_0^\infty x^{\frac{1}{2}-1} e^{-x} dx = \int_0^\infty x^{-\frac{1}{2}} e^{-x} dx = \int_0^\infty \frac{e^{-x}}{\sqrt{x}} dx.$$

Let $x = \frac{y^2}{2}$, then $dx = ydy$,

$$\Rightarrow \Gamma\left(\frac{1}{2}\right) = \int_0^\infty \frac{e^{-\frac{y^2}{2}}}{\frac{y}{\sqrt{2}}} ydy = \sqrt{2} \int_0^\infty e^{-\frac{y^2}{2}} dy.$$

But it is easy to see

$$\int_{-\infty}^0 e^{-\frac{y^2}{2}} dy = \int_0^\infty e^{-\frac{y^2}{2}} dy.$$

Part 3.5.3: The Gamma Distribution

Hence $\int_0^{\infty} e^{-\frac{y^2}{2}}\,dy = \frac{1}{2}\int_{-\infty}^{+\infty} e^{-\frac{y^2}{2}}\,dy$

$$\begin{aligned}\Rightarrow \Gamma\left(\frac{1}{2}\right) &= \frac{\sqrt{2}}{2}\int_{-\infty}^{+\infty} e^{-\frac{y^2}{2}}\,dy = \frac{\sqrt{\pi}}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-\frac{y^2}{2}}\,dy \\ &= \sqrt{\pi}\end{aligned}$$

Here we have used the fact that

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-\frac{y^2}{2}}\,dy = 1 \quad \text{(why?)}$$

Part 3.5.3: The Gamma Distribution

Hence for $n + \frac{1}{2}$ with n being a positive integer, we get

$$\begin{aligned}
\Gamma\left(\frac{1}{2}\right) &= \sqrt{\pi}, \\
\Gamma\left(1+\frac{1}{2}\right) &= \frac{1}{2}\Gamma\left(\frac{1}{2}\right) = \frac{\sqrt{\pi}}{2}, \\
\Gamma\left(2+\frac{1}{2}\right) &= \frac{3}{2}\cdot\frac{1}{2}\Gamma\left(\frac{1}{2}\right) = \frac{3!!}{2^2}\sqrt{\pi}.
\end{aligned}$$

In general

$$\Gamma\left(n+\frac{1}{2}\right) = \frac{(2n-1)!!}{2^n}\sqrt{\pi}.$$

Part 3.5.3: The Gamma Distribution

3. Now, return to the gamma distribution, if $\alpha = 1$, then

$$f(x) = \begin{cases} \frac{\lambda e^{-\lambda x}(\lambda x)^{1-1}}{\Gamma(1)}, & x \geq 0, \\ 0, & x < 0, \end{cases}$$

i.e.

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

That is just the p.d.f. of an exponentially distributed with parameter $\lambda > 0$.

If $\alpha = n$ (positive integer), then

$$f(x) = \begin{cases} \frac{\lambda e^{-\lambda x}(\lambda x)^{n-1}}{(n-1)!}, & x \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

Outline

1. **Definition:** We say that X is a standard normal random variable (or, X is standard normally distributed) if the p.d.f. is given by

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

In this case, we write $X \sim N(0, 1)$.

2. Check

$$f(x) \geq 0, \qquad \int_{-\infty}^{+\infty} f(x)dx = 1.$$

Let

$$I = \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx = \int_{-\infty}^{+\infty} e^{-\frac{y^2}{2}} dy$$

Then

$$I^2 = \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx \cdot \int_{-\infty}^{+\infty} e^{-\frac{y^2}{2}} dy = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-\frac{x^2+y^2}{2}} dxdy.$$

Changing of variables to polar coordinates yields

$$x = r\cos\theta, \quad y = r\sin\theta.$$

Part 3.6.1: Standard Normal Random Variables

Then

$$J = \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos\theta & r(-\sin\theta) \\ \sin\theta & r\cos\theta \end{vmatrix} = r\cos^2\theta + r\sin^2\theta = r.$$

Also

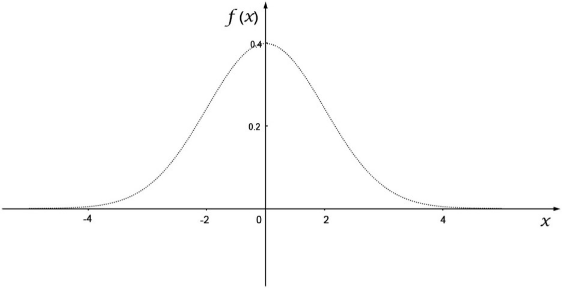
$$x^2 + y^2 = r^2\cos^2\theta + r^2\sin^2\theta = r^2.$$

Hence

$$I^2 = \int_0^\infty \int_0^{2\pi} e^{-\frac{r^2}{2}} \cdot r dr d\theta = \int_0^\infty r e^{-\frac{r^2}{2}} \left[\int_0^{2\pi} d\theta\right] \cdot dr$$
$$= 2\pi \int_0^\infty r e^{-\frac{r^2}{2}} dr = \left(-2\pi e^{-\frac{r^2}{2}}\right)\Big|_0^{+\infty} = 2\pi$$
$$\Rightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx = 1.$$

3. Graph of the p.d.f. $f(x)$:

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



Part 3.6.1: Standard Normal Random Variables

(1) Bell-shaped cure;

(2) Symmetric with $x = 0$. Indeed,

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

(3) Maximal value of $f(x)$: $\frac{1}{\sqrt{2\pi}} \approx 0.4$, i.e., $f(x)$ taking its maximal value at $x = 0$ and

$$f(0) = \frac{1}{\sqrt{2\pi}} e^{-\frac{0^2}{2}} = \frac{1}{\sqrt{2\pi}} \approx 0.4.$$

4. Cumulative Distribution Function (c.d.f.):

   Recall: $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(y)dy$. $X \sim N(0,1)$, so

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{y^2}{2}} dy.$$

   For this c.d.f. $F(x)$, we usually denote it as

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{y^2}{2}} dy.$$

   It can be proved that the closed form for $\Phi(x)$ is impossible, but the values of $\Phi(x)$ ($\forall x$) can be found by numerical method. (See the table!)

5. Calculation:

(1) For non-negative x, the value of $\Phi(x)$ can be obtained by checking the table directly. For example:

$$\Phi\left(\frac{2}{3}\right) \approx \Phi(0.67) \approx 0.7486.$$

(2) For negative x, the value of $\Phi(x)$ can be obtained by using the formula

$$\Phi(x) = 1 - \Phi(-x)$$

i.e., $\Phi(x) + \Phi(-x) = 1$. For example:

$$\Phi\left(-\frac{1}{3}\right) = 1-\Phi\left(\frac{1}{3}\right) \approx 1-\Phi(0.33) \approx 1-0.6294 = 0.3707.$$

(3) Proof of $\Phi(x) + \Phi(-x) = 1$:

Suppose $x \geq 0$, then

$$\Phi(x) + \Phi(-x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{y^2}{2}} dy + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-x} e^{-\frac{y^2}{2}} dy.$$

For the latter, let $z = -y$, then

$$\begin{aligned}\int_{-\infty}^{-x} e^{-\frac{y^2}{2}} dy &= \int_{+\infty}^{x} e^{-\frac{(-z)^2}{2}} d(-z) = -\int_{+\infty}^{x} e^{-\frac{z^2}{2}} dz \\ &= \int_{x}^{+\infty} e^{-\frac{z^2}{2}} dz = \int_{x}^{+\infty} e^{-\frac{y^2}{2}} dy.\end{aligned}$$

Part 3.6.1: Standard Normal Random Variables

Thus,

$$\Phi(x) + \Phi(-x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{y^2}{2}} dy + \frac{1}{\sqrt{2\pi}} \int_{x}^{+\infty} e^{-\frac{y^2}{2}} dy$$
$$= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = 1.$$

If $x < 0$, then let $a = -x$, then $a > 0$ and so

$$\Phi(x) + \Phi(-x) = \Phi(-a) + \Phi(a) = 1.$$

(4) **Summary:** Use the following formula to find the probabilities:

If $a < b$,

$$P(a \leq X \leq b) = \Phi(b) - \Phi(a).$$

For any a,

$$P(X \leq a) = \int_{-\infty}^{a} f(x)dx = \Phi(a),$$

$$P(X > a) = 1 - P(X \leq a) = 1 - \Phi(a),$$

where $X \sim N(0, 1)$.

(5) An Examples: If $X \sim N(0, 1)$, then

$$P(2 < X < 2.5) = \Phi(2.5) - \Phi(2) = 0.9938 - 0.9772 = 0.0166.$$

Also,

$$\begin{aligned} P(-1.63 \leq X < 0.85) &= \Phi(0.85) - \Phi(-1.63) \\ &= \Phi(0.85) - (1 - \Phi(1.63)) \\ &= \Phi(0.85) + \Phi(1.63) - 1 \\ &= 0.8023 + 0.9484 - 1 = 0.7507. \end{aligned}$$

Part 3.6.2: General Normal Random Variables

1. **Definition:** We say that X is a normal random variable with parameters μ and σ^2 (or, X is normally distributed with parameters μ and σ^2), denoted by $X \sim N(\mu, \sigma^2)$, if the p.d.f is given by

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < +\infty.$$

Note that, for a normal random variable, the p.d.f. has two parameters μ and σ^2 (we assume $\sigma > 0$). Note also that if $\mu = 0$ and $\sigma^2 = 1$, we return to the standard normal p.d.f..

2. Check

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

Let $y = \frac{x-\mu}{\sigma}$, then

$$dy = \frac{1}{\sigma} dx \quad \Rightarrow \quad dx = \sigma dy.$$

Also:

$$x \to -\infty \quad \Leftrightarrow \quad y \to -\infty \quad (\because \sigma > 0),$$
$$x \to +\infty \quad \Leftrightarrow \quad y \to +\infty \quad (\because \sigma > 0).$$

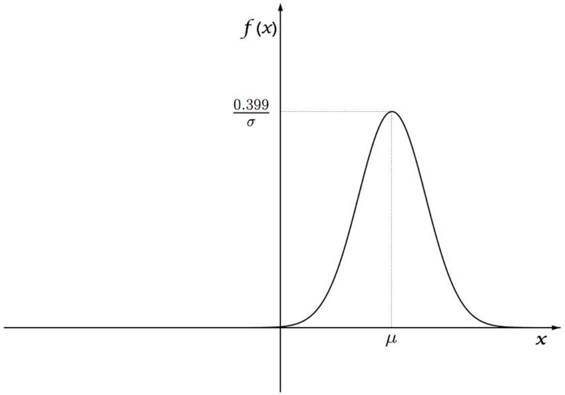
Hence

$$\begin{aligned}
\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx &= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\,dx \\
&= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty} e^{-\frac{1}{2}y^2}\cdot\sigma\,dy \\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-\frac{y^2}{2}}\,dy \\
&= 1.
\end{aligned}$$

Part 3.6.2: General Normal Random Variables

3. Graph of the p.d.f. $f(x)$: $X \sim N(\mu, \sigma^2)$,

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



Part 3.6.2: General Normal Random Variables

(1) Bell-shaped curve.

(2) Symmetric with $x = \mu$ (μ can be negative!).

(3) Maximal value: $\frac{1}{\sqrt{2\pi}\sigma} \approx \frac{0.399}{\sigma}$, i.e., $f(x)$ taking its maximal value at $x = \mu$ and thus

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

4. Notes:

   For general normal distribution, we have two parameters μ and σ^2.

   The probabilistic meaning of these two parameters will be clear later (see Chapter 5).

   Here, we just mention that the parameter is σ^2 not σ (σ is called the "standard deviation").

   For example, if we say, $X \sim N(3, 5)$ then $\mu = 3$, $\sigma^2 = 5$ and thus $\sigma = \sqrt{5}$.

Part 3.6.3: Linear Transformation of Normal Distributions

1. Problem: If $X \sim N(\mu, \sigma^2\Big)$, let $Y = a \cdot X + b$ where $a \neq 0$ and b are two constants. $Y \sim$?? (Normal ?? Parameters ??)

2. **Conclusion: (Important!!)** If $X \sim N(\mu, \sigma^2)$ and $Y = aX + b$ where a and b are two constants ($a \neq 0$), then Y is also normally distributed with parameters $a\mu + b$ and $a^2\sigma^2$, i.e.,

$$Y = aX + b \sim N(a\mu + b, a^2\sigma^2).$$

Note: The most important fact is that Y is still normally distributed.

Part 3.6.3: Linear Transformation of Normal Distributions

Proof. Let $F_X(\cdot)$ and $F_Y(\cdot)$ be the c.d.f's of X and Y. Then

$$F_Y(y) = P\{Y \leq y\} = P\{aX + b \leq y\} = P\{aX \leq y - b\}.$$

Consider two cases: $a > 0$ or $a < 0$.

(1) If $a > 0$, then $F_Y(y) = P\{aX \leq y - b\} = P\left\{X \leq \frac{y-b}{a}\right\}$,

$$\Rightarrow F_Y(y) = F_X\left(\frac{y-b}{a}\right) \Rightarrow \frac{dF_Y(y)}{dy} = \frac{d}{dy}F_X\left(\frac{y-b}{a}\right)$$

$$\Rightarrow F_Y'(y) = F_X'\left(\frac{y-b}{a}\right) \times \frac{1}{a}$$

Part 3.6.3: Linear Transformation of Normal Distributions

Let $f_Y(y)$ and $f_X(x)$ be the p.d.f.'s of Y and X. Then

$$f_Y(y) = f_X\left(\frac{y-b}{a}\right) \cdot \frac{1}{a}.$$

Now $X \sim N(\mu, \sigma^2)$ i.e., $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$,

$$\Rightarrow f_X\left(\frac{y-b}{a}\right) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\left(\frac{y-b}{a}-\mu\right)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-b-a\mu)^2}{2\sigma^2 a^2}}.$$

Therefore $f_Y(y) = f_X\left(\frac{y-b}{a}\right) \cdot \frac{1}{a} = \frac{1}{\sqrt{2\pi}a\sigma} e^{-\frac{(y-(a\mu+b))^2}{2\sigma^2 a^2}}$.

$$Y \sim N(a\mu + b, a^2\sigma^2).$$

Part 3.6.3: Linear Transformation of Normal Distributions

(2) If $a < 0$, then

$$
\begin{aligned}
F_Y(y) &= P\{aX \leq y - b\} = P\left\{X \geq \frac{y-b}{a}\right\} \quad (a < 0) \\
&= 1 - P\left\{X < \frac{y-b}{a}\right\} \\
&= 1 - P\left\{X \leq \frac{y-b}{a}\right\} \quad (X \text{ continuous}) \\
\Rightarrow \frac{dF_Y(y)}{dy} &= (-1) \cdot F_X'\left(\frac{y-b}{a}\right) \cdot \frac{1}{a} \\
\Rightarrow f_Y(y) &= (-1) \cdot \frac{1}{a} f_X\left(\frac{y-b}{a}\right)
\end{aligned}
$$

Part 3.6.3: Linear Transformation of Normal Distributions

$$\Rightarrow f_Y(y) = (-1) \cdot \frac{1}{a} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{[y-(a\mu+b)]^2}{2\sigma^2 a^2}}$$
$$= \frac{1}{\sqrt{2\pi}(-a\sigma)} e^{-\frac{[y-(a\mu+b)]^2}{2a^2\sigma^2}}.$$

Thus,

$$Y \sim N(a\mu + b, a^2\sigma^2).$$

Part 3.6.3: Linear Transformation of Normal Distributions

3. **Corollary:** If $X \sim N(\mu, \sigma^2)$, and let $Y = \frac{X-\mu}{\sigma}$, then

$$Y \sim N(0, 1).$$

Proof. Recall if $X \sim N(\mu, \sigma^2)$, $Y = aX + b$ $(a \neq 0)$, then

$$Y \sim N(a\mu + b, a^2\sigma^2).$$

Now, $Y = \frac{X-\mu}{\sigma}$, hence $a = \frac{1}{\sigma}$, $b = -\frac{\mu}{\sigma}$ and it follows that

$$a\mu + b = \frac{1}{\sigma} \cdot \mu - \frac{\mu}{\sigma} = 0; \qquad a^2\sigma^2 = \left(\frac{1}{\sigma}\right)^2 \cdot \sigma^2 = 1.$$

Therefore $Y \sim N(0, 1)$.

Part 3.6.4: Calculation of General Normal Distribution

1. Principle: Suppose $X \sim N(\mu, \sigma^2)$. Let $Y = \frac{X-\mu}{\sigma}$, then by the above corollary, $Y \sim N(0,1)$.

   Then we can use results about the standard normal distribution.

Part 3.6.4: Calculation of General Normal Distribution

2. Example: $X \sim N(3, 9)$, find $P(2 < X < 5)$.

   Solution: $\mu = 3$, $\sigma = 3$. Let $Y = \frac{X-\mu}{\sigma} = \frac{X-3}{3} \sim N(0, 1)$.

$$\begin{aligned}
P(2 < X < 5) &= P\left(-\frac{1}{3} < Y < \frac{2}{3}\right) \\
&= \Phi\left(\frac{2}{3}\right) - \Phi\left(-\frac{1}{3}\right) \quad (\because Y \sim N(0,1)) \\
&= \Phi\left(\frac{2}{3}\right) - \left(1 - \Phi\left(\frac{1}{3}\right)\right) \\
&= \Phi\left(\frac{2}{3}\right) + \Phi\left(\frac{1}{3}\right) - 1 \approx 0.3779.
\end{aligned}$$

Outline

Part 3.7.1: Concept

1. Recall if $X \sim N(\mu, \sigma^2)$, then

$$Y \triangleq aX + b \sim N(a\mu + b, a^2\sigma^2).$$

We usually call "Y is a function of the r.v. X".

More exactly, if $y = ax + b$ is an ordinary function (i.e. $\mathbb{R} \to \mathbb{R}$), where x and y are ordinary variables (nothing with randomness), then by replacing the argument x by the r.v. X, we get $aX + b$ (more clearly, $aX(\omega) + b$), which is a random variable, denoted by $Y(\omega)$. This is called a function of the r.v. X.

Similarly, if $y = x^2$ is another ordinary function and X is a r.v. then X^2 is a new r.v..

2. In general: If $y = g(x)$ is an ordinary function and X is a r.v., then replacing the argument x by the r.v. X, we get another r.v. $Y = g(X)$.

   This r.v. Y is called a function of the original r.v. X.

3. More examples: Let X be a random variable.

   (1) Function form: $y = e^x$ (i.e., $g(x) = e^x$), then $Y = e^X$ is a function of the r.v. X corresponding to the ordinary function $y = e^x$.

   (2) Function form: $z = e^{2x+3}$, then $Z(\omega) = e^{2X+3}$ is a r.v..

   (3) Function form: $z = x^k$ (k: positive integer), then $Z(\omega) = X^k$ is a r.v..

Part 3.7.2: Main Question

If X is a r.v. with c.d.f. $F_X(x)$ and $g(x)$ is a function, then $Y = g(X)$ is a r.v..

What is the distribution of Y?

No general solution! However, for some cases, we may get solutions.

Next we use examples to illustrate the general idea.

Part 3.7.3: General idea

1. Example 1. If $X \sim N(\mu, \sigma^2\Big)$, then $aX + b \sim ?$ (see before)

2. Example 2. If $X \sim N(0, 1)$, then what is the distribution of X^2?

   Solution: Let $Y = X^2$ and assume the c.d.f. of Y is $F_Y(y)$.

   $$F_Y(y) = P\{Y \leq y\} = P\{X^2 \leq y\}.$$

   If $y < 0$, then

   $$\{\omega; X^2(\omega) \leq y\} = \emptyset \;\Rightarrow\; F_Y(y) = 0$$

If $y \geq 0$, then

$$\{\omega; X^2(\omega) \leq y\} = (|X| \leq \sqrt{y}) = (-\sqrt{y} \leq X \leq \sqrt{y}),$$

and hence

$$\begin{aligned} F_Y(y) &= P\{X^2 \leq y\} = P\{-\sqrt{y} \leq X \leq \sqrt{y}\} \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}) \end{aligned}$$

By the "chain rule" of differentiation, for the case of $y > 0$ we obtain

$$\begin{aligned} F'_Y(y) = \frac{dF_Y(y)}{dy} &= F'_X(\sqrt{y})\frac{1}{2}y^{\frac{1}{2}-1} - F'_X(-\sqrt{y}) \cdot (-1)\frac{1}{2}y^{\frac{1}{2}-1} \\ &= \frac{1}{2}y^{-\frac{1}{2}}\left[F'_X(\sqrt{y}) + F'_X(-\sqrt{y})\right] \end{aligned}$$

Part 3.7.3: General idea

Let $f_Y(y)$ and $f_X(x)$ be the p.d.f.'s of Y and X respectively, then

$$f_Y(y) = \frac{1}{2\sqrt{y}}\Big[f_X(\sqrt{y}) + f_X(-\sqrt{y})\Big].$$

Note that until now, we haven't used the fact that $X \sim N(0,1)$ yet and hence the above result holds true for any continuous r.v. X!

Now we use the fact that $X \sim N(0,1)$ then

$$f_X(\sqrt{y}) = f_X(-\sqrt{y}) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(\sqrt{y})^2} = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}y}$$

$$\Rightarrow f_Y(y) = \frac{1}{\sqrt{y}}f_X(\sqrt{y}) = \frac{1}{\sqrt{y}\sqrt{2\pi}}e^{-\frac{y}{2}}.$$

Part 3.7.3: General idea

Final solution: If $X \sim N(0,1)$, and $Y = X^2$, then the p.d.f. of Y, denoted by $f_Y(y)$ is given by

$$f_Y(y) = \begin{cases} \frac{e^{-\frac{y}{2}}}{\sqrt{2\pi y}}, & \text{if } y > 0, \\ 0, & \text{if } y \leq 0. \end{cases}$$

Part 3.7.3: General idea

Note: The above p.d.f. is a special case of the Gamma Distributions discussed in 3.5.3. Indeed, if we let $\lambda = \alpha = \frac{1}{2}$, we can get that

$$f_Y(y) = \begin{cases} \frac{\frac{1}{2} e^{-\frac{x}{2}} \left(\frac{x}{2}\right)^{\frac{1}{2}-1}}{\Gamma\left(\frac{1}{2}\right)}, & \text{if } x > 0, \\ 0, & \text{if } x \leq 0. \end{cases}$$

By noting that $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$ and after some trivial algebra, we see that the above expression is just the one obtained in the above.

Hence, if $X \sim N(0,1)$ and $Y = X^2$, then $Y \sim \Gamma\left(\frac{1}{2}, \frac{1}{2}\right)$. This is a special case of the so-called χ^2-distribution, which will play essential role in Statistics.

The method used in the above examples has general meaning. It is usually called the <u>c.d.f. method</u>.

Part 3.7.4: A useful general result

1. **Theorem:** Let X be a continuous r.v. with p.d.f. $f_X(x)$. Let $g(x)$ be a strictly monotone (increasing or decreasing) differentiable function. Then $Y = g(X)$ has a p.d.f. given by

$$f_Y(y) = \begin{cases} f_X[g^{-1}(y)] \cdot \left|\frac{d}{dy} g^{-1}(y)\right|, & \text{if } y = g(x) \text{ for some } x, \\ 0, & \text{if } y \neq g(x) \text{ for all } x. \end{cases}$$

where $g^{-1}(y)$ is defined to be the value of x such that $g(x) = y$.

Part 3.7.4: A useful general result

2. Conditions:

   - the p.d.f. of X: $f_X(x)$;
   - function $g(x)$ either $\uparrow\uparrow$ or $\downarrow\downarrow$;
   - $\uparrow\uparrow$ means if $x_1 < x_2$ then $g(x_1) < g(x_2)$, i.e. strictly increasing;
   - $\downarrow\downarrow$ means if $x_1 < x_2$ then $g(x_1) > g(x_2)$, i.e. strictly decreasing;
   - Note that strictly monotone property guarantees that $g^{-1}(x)$ is uniquely determined and is also monotone;
   - Also, $g'(x)$ exists (and thus $g(x)$ is continuous).

Part 3.7.4: A useful general result

3. Procedure: (Assume that $y = g(x)$ for some x otherwise just 0)
   - First find $g^{-1}(y)$;
   - Obtain $\frac{d}{dy}g^{-1}(y)$ [This is guaranteed];
   - Obtain $\left|\frac{d}{dy}g^{-1}(y)\right|$ [In order to guarantee $f_Y(\cdot) \geq 0$];
   - Find $f_X\left(g^{-1}(y)\right)$ [$f_X(\cdot)$ and $g^{-1}(\cdot)$ are given];
   - Finally $f_X(y)$ is the product of $f_X\left(g^{-1}(y)\right)$ and $\left|\frac{d}{dy}g^{-1}(y)\right|$. If $y \neq g(x)$ for all x, then $f_Y(y) = 0$.

4. Proof: Let $F_X(x)$ and $F_Y(y)$ be the c.d.f.'s of X and Y, then

$$F_Y(y) = P\{Y \le y\} = P\{g(X) \le y\}.$$

Now consider the ω-set

$$A = \{\omega \in \Omega; g(X(\omega)) \le y\}.$$

(i) If $g(\cdot) \uparrow\uparrow$, then A happens iff $\{\omega \in \Omega; X(\omega) \le g^{-1}(y)\}$ happens;

(ii) if $g(\cdot) \downarrow\downarrow$, then A happens iff $\{\omega \in \Omega; X(\omega) \ge g^{-1}(y)\}$ happens.

Therefore, if $g(\cdot) \uparrow\uparrow$, then

$$F_Y(y) = P\{g(X) \le y\} = P\left\{X \le g^{-1}(y)\right\} = F_X\left(g^{-1}(y)\right).$$

Part 3.7.4: A useful general result

$$\frac{dF_Y(y)}{dy} = \frac{d}{dy}F_X\left(g^{-1}(y)\right) = F'_X\left(g^{-1}(y)\right) \cdot \frac{d}{dy}g^{-1}(y) \quad \text{(chain rule)}$$

$$\Rightarrow f_Y(y) = f_X\left(g^{-1}(y)\right) \cdot \frac{d}{dy}g^{-1}(y),$$

If $g(\cdot) \downarrow\downarrow$, then

$$\begin{aligned} F_Y(y) &= P\{g(X) \le y\} = P\left\{X \ge g^{-1}(y)\right\} \\ &= 1 - P\left\{X < g^{-1}(y)\right\} \quad \text{(Think why here!)} \\ &= 1 - P\left\{X \le g^{-1}(y)\right\} \quad (X \text{ continuous}) \\ &= 1 - F_X\left(g^{-1}(y)\right). \end{aligned}$$

Part 3.7.4: A useful general result

$$\Rightarrow \frac{dF_Y(y)}{dy} = (-1) \cdot F_X' \left(g^{-1}(y)\right) \cdot \frac{d}{dy} g^{-1}(y)$$
$$= (-1) \cdot f_X \left(g^{-1}(y)\right) \frac{d}{dy} g^{-1}(y),$$

i.e.,

$$f_Y(y) = f_X \left(g^{-1}(y)\right) \cdot (-1) \cdot \frac{d}{dy} g^{-1}(y).$$

In a uniformed form:

$$f_Y(y) = f_X \left(g^{-1}(y)\right) \cdot \left| \frac{d}{dy} g^{-1}(y) \right| .$$

The proof is completed.

Part 3.7.4: A useful general result

5. Remarks: By the proof we see that

   (1) If $g(\cdot) \uparrow\uparrow$, then

$$f_Y(y) = f_X\left(g^{-1}(y)\right) \cdot \frac{d}{dy} g^{-1}(y).$$

   Note that in this case,

$$g^{-1}(y) \uparrow\uparrow \quad \Rightarrow \quad \frac{d}{dy} g^{-1}(y) > 0,$$

   which implies $f_Y(y) \geq 0$.

Part 3.7.4: A useful general result

(2) If $g(\cdot) \downarrow\downarrow$, then

$$f_Y(y) = f_X\left(g^{-1}(y)\right) \cdot (-1) \cdot \frac{d}{dy} g^{-1}(y).$$

Note that

$$g^{-1}(y) \downarrow\downarrow \quad \Rightarrow \quad \frac{d}{dy} g^{-1}(y) < 0,$$

which implies $f_Y(y) = (-1)\frac{d}{dy} g^{-1}(y) \cdot f_X\left(g^{-1}(y)\right) \geq 0$.

Part 3.7.4: A useful general result

6. Notes.

   (i) This theorem requires the condition that $\forall x \in \mathbb{R}$, $g(x)$ is strictly monotone (either $\uparrow\uparrow$ or $\downarrow\downarrow$).

   However, it is easy to see that it is still true that it $g(x)$ is piecewise strictly monotone. Only in different intervals, the form is different, but can use the uniformed form

$$f_X\left(g^{-1}(y)\right) \cdot \left|\frac{d}{dy} g^{-1}(y)\right|.$$

   (ii) In practice, you do not need to remember the formula, just use the"c.d.f. method".

Part 3.7.4: A useful general result

(iii) Why if $y \neq g(x)$ for all x, we have $f_Y(y) = 0$?

Recall $Y(\omega) = g(X(\omega))$.

If $y \neq g(x)$ for all x, which means there does not exist any $\omega \in \Omega$ such that $Y(\omega) = g(X(\omega))$, i.e. $Y(\omega)$ does not take this value. Hence $f_Y(y) = 0$.

Example 1: Let X be a non-negative r.v. with p.d.f. $f_X(\cdot)$. Find the p.d.f. $f_Y(\cdot)$ of $Y = X^n$ (where $n \geq 1$ is a positive integer).

Solution: The function form is $y = x^n$.

$$X(\omega) \geq 0 \Rightarrow Y(\omega) = X^n(\omega) \geq 0.$$

Thus, $\underline{\text{for } y \leq 0,\ f_Y(y) = 0}$. For $y > 0$,

$$F_Y(y) = P(Y \leq y) = P(X^n \leq y) = P(X \leq \sqrt[n]{y}) = F_X(\sqrt[n]{y})$$

$$\Rightarrow f_Y(y) = F_Y'(y) = F_X'(\sqrt[n]{y}) \cdot \frac{1}{n} y^{\frac{1}{n}-1} = \frac{1}{n} y^{\frac{1}{n}-1} f_X(\sqrt[n]{y}).$$

Therefore,

$$f_Y(y) = \begin{cases} \frac{1}{n} y^{\frac{1}{n}-1} \cdot f_X\left(y^{\frac{1}{n}}\right), & \text{if } y \geq 0, \\ 0, & \text{if } y < 0. \end{cases}$$

Part 3.7.5: Further Examples and Distributions

Example 2: Let $X \sim N(\mu, \sigma^2)$. Find a p.d.f. of $Y = e^X$.

Solution: The function form is $y = e^x$, i.e. $g(x) = e^x$.

$$Y(\omega) = e^{X(\omega)} \geq 0, \quad \forall \omega \in \Omega.$$

Thus for $y < 0$, $f_Y(y) = 0$. For $y > 0$,

$$F_Y(y) = P(Y \leq y) = P(e^X \leq y) = P(X \leq \ln y) = F_X(\ln y)$$

$$\Rightarrow f_Y(y) = F_Y'(y) = F_X'(\ln y) \cdot \frac{1}{y} = \frac{1}{y} f_X(\ln y).$$

But $X \sim N\left(\mu, \sigma^2\right)$

$$\therefore f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty.$$

Therefore, for $y > 0$,

$$f_Y(y) = \frac{1}{y} f_X(\ln y) = \frac{1}{y} \cdot \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\ln y-\mu)^2}{2\sigma^2}}.$$

Finally,

$$f_Y(y) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma y} e^{-\frac{(\ln y-\mu)^2}{2\sigma^2}}, & \text{if } y > 0, \\ 0, & \text{if } y \leq 0. \end{cases}$$

Hence, we have

$$\frac{1}{\sqrt{2\pi}\sigma} \int_0^{+\infty} \frac{e^{-\frac{(\ln y-\mu)^2}{2\sigma^2}}}{y} dy = 1.$$

Example 3 (The Cauchy Distribution):

(1) **Definition:** A r.v. Y is called to follow a Cauchy Distribution with parameter μ $(-\infty < \mu < \infty)$, if its p.d.f. $f_Y(y)$ is given by

$$f_Y(y) = \frac{1}{\pi} \cdot \frac{1}{[1 + (y - \mu)^2]}, \quad -\infty < y < \infty.$$

If $\mu = 0$, we usually called it the standard Cauchy Distribution.

It is easy to see that, the above $f_Y(y)$ is indeed a p.d.f.. Indeed, $f_Y(y) \geq 0$ is clear. Also,

$$\frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{dy}{1 + (y - \mu)^2} = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{dz}{1 + z^2} = \frac{1}{\pi} \cdot \pi = 1.$$

Part 3.7.5: Further Examples and Distributions

(2) The standard Cauchy Distribution (i.e. $\mu = 0$) has an interesting explanation and also, has a close link with the uniform distribution $U\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$.

Let X be uniformly distributed over the interval $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ (we use open interval here). Hence the p.d.f. is

$$f_X(x) = \begin{cases} \frac{1}{\pi}, & \text{if } -\frac{\pi}{2} < x < \frac{\pi}{2}, \\ 0, & \text{otherwise.} \end{cases}$$

Let $Y = \tan(X)$, i.e., $g(x) = \tan(x)$. $g(x) \uparrow \left(-\frac{\pi}{2} < x < \frac{\pi}{2}\right)$.
Now $g^{-1}(y) = \tan^{-1}(y) \equiv \arctan(y)$,

$$\Rightarrow \frac{d}{dy} g^{-1}(y) = \frac{d}{dy} \tan^{-1}(y) = \frac{1}{1+y^2}.$$

By the general formula, we get that the p.d.f. of Y is

$$\begin{aligned} f_Y(y) &= f_X\left(g^{-1}(y)\right) \cdot \left|\frac{d}{dy} g^{-1}(y)\right| \\ &= f_X\left(g^{-1}(y)\right) \frac{d}{dy} g^{-1}(y) \quad (\because g(\cdot) \uparrow\uparrow) \\ &= f_X\left(g^{-1}(y)\right) \cdot \frac{1}{1+y^2}. \end{aligned}$$

Part 3.7.5: Further Examples and Distributions

Note $g^{-1}(y) = \tan^{-1}(y)$ taking values $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$. But

$$f_X(x) = \begin{cases} \frac{1}{\pi}, & \text{if } -\frac{\pi}{2} < x < \frac{\pi}{2}, \\ 0, & \text{otherwise.} \end{cases}$$

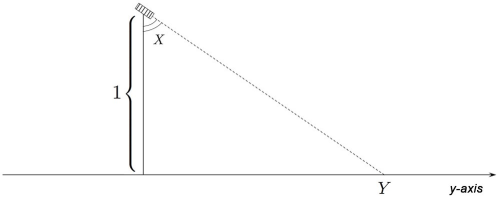
and thus $f_X\left(g^{-1}(y)\right) = \frac{1}{\pi}$. Therefore the p.d.f. of Y is

$$f_Y(y) = \frac{1}{\pi} \cdot \frac{1}{1+y^2}, \quad -\infty < y < +\infty.$$

(Thus the standard Cauchy distribution)

The conclusion can also be obtained by the c.d.f. method, try it by yourself!

(3) "Experiment": Suppose that a narrow beam flashlight is spun around its center which is located a unit distance from the y-axis. When the flashlight has stopped spinning, the point Y (a r.v.) at which the beam intersects the y-axis is just the standard Cauchy Distribution. See Figure below.



Clearly: $Y = g(X) = \tan(X)$ and X is uniformly distributed over $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$.