MATH1520 University Mathematics for Applications **Fall 2021**

Chapter 12: Probability

Learning Objectives:
(1) Define outcome, sample space, random variable, and other basic concepts of probability.
(2) Define and examine continuous probability density functions.
(3) Compute and use expected value.
(4) Interpret variance and standard deviation.

1 Basic Concepts of Probability

Some probability examples we learned before:

Example 1.1.

- Roll a dice.

$$\text{Probability of rolling a “two”} = \frac{1}{6}$$

- Toss a coin

$$\text{Probability of getting a tail} = \frac{1}{2}.$$

Our goal:

- Make things formal
- Change the setting from discrete to continuous.

Definition 1.1. For a random experiment:

1. **Possible outcome** : possible result of a single experiment.
2. **Sample space**: collection of all possible outcomes
3. **Event**: any collection of possible outcomes.
4. **Probability of an event**: a number between 0 and 1 which describes the possibility that the event occurs.

Example 1.2. Roll a dice once and record the score on the top face.

1. **Sample space:** $S = \{1, 2, 3, 4, 5, 6\}$ equally likely for each outcome.
2. **Events:** $\emptyset$, $\{1\}$,. . ., $\{1, 2\}$, . . ., $\{1, 2, 3\}$, . . ., . . ., $\{1, 2, 3, 4, 5, 6\}$.
3.
   - $P(\emptyset) = 0$: probability of getting "nothing" is 0, impossible!
   - Let event A be "getting a 3", then $A = \{3\}$, $P(A) = \frac{1}{6}$.
   - Let event B be "getting a score ≤ 3", then $B = \{1, 2, 3\}$, $P(B) = \frac{3}{6} = \frac{1}{2}$.
   - Let event C be "getting a score ≥ 1", then $C = \{1, 2, 3, 4, 5, 6\}$, $P(C) = 1$. certain!

Exercise 1.1. A family has 3 children. Denote a boy by B, a girl by G. Write down the sample space and find the probability for the event "at most 1 girl".

2 Discrete Random Variable

Definition 2.1. For a random experiment with sample space S, a **random variable** X is a function that assigns a real number to each possible outcome in S, i.e.

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

If image of X is finite or countably infinite, X is called a discrete random variable. Otherwise, if image of X is an interval, X is called a continuous random variable.

Example 2.1. Toss a coin 3 times.

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

Let X be the number of times a head comes out.

Image of $X = \{0, 1, 2, 3\}$, discrete random variable

$$X(HHH) = 3, \quad X(HHT) = 2 = X(HTH), \ldots$$

Example 2.2. Toss a coin until a head comes out.

$$S = \{H, TH, TTH, TTTH, TTTTH, \ldots\}.$$

Let X be the number of trials.

$$X(H) = 1, X(TH) = 2, X(TTH) = 3, \ldots$$

Image of X is $\{1, 2, 3, \ldots\}$ which is countably infinite, so X is discrete.

Example 2.3. Suppose AB is a rope of length 10cm and AB is cut into two pieces AC, CB randomly.

Let X be the length of AC.

Image of X is $(0, 100)$: an open interval from 0 to 100, so X is a continuous random variable.

2.1 Probability Distribution of Discrete Random Variable

Let S be a sample space, X be a discrete random variable with image $\{x_1, x_2, \ldots\}$.

For each value x_i, define

$$\boxed{p(x_i) = P(X = x_i) = \text{probability of the event } X = x_i}$$

$\{p(x_i), \quad i = 1, 2, \ldots\}$ is called probability distribution of X.

Theorem 1. *A* **probability distribution** *(or* **probability density function** *short form: pdf)* $\{p(x_i), \quad i = 1, 2, \ldots\}$ *for* X *satisfies:*

1. $0 \leq p(x_i) \leq 1$, $i = 1, 2, \ldots$
2. $\sum_i p(x_i) = 1$

Example 2.4. Toss a coin 3 times.

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

Let X be the number of times a head comes out. Then

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

x_i	0	1	2	3
$p(x_i)$	$1/8$	$3/8$	$3/8$	$1/8$

Note

$$p(1) + p(2) + p(3) = 1$$
$$P(1 \leq X \leq 2) = P(X = 1) + P(X = 2) = p(1) + p(2) = \frac{6}{8} = \frac{3}{4}$$

Example 2.5. Toss a coin until a head comes out.

$$S = \{H, TH, TTH, TTTH, \ldots\}$$

Let X be the number of trials.

$$p(1) = P(X = 1) = \frac{1}{2}$$
$$p(2) = P(X = 2) = \frac{1}{4}$$
$$p(3) = P(X = 3) = \frac{1}{8}$$
$$p(4) = P(X = 4) = \frac{1}{16}$$
$$\vdots$$

Generally

$$p(x) = \begin{cases} \frac{1}{2^x} & \text{if } x \text{ is a nonnegative integer} \\ 0 & \text{otherwise} \end{cases}$$

Again

$$\begin{aligned} \sum_{x:p(x)\neq 0} &= p(1) + p(2) + p(3) + \cdots \\ &= \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \cdots \\ &= \frac{1}{2} \times \frac{1}{1 - \frac{1}{2}} = 1. \end{aligned}$$

Here we use the formula

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \cdots$$

for $|x| < 1$.

2.2 Expected Value and Variance of a Discrete Random Variable

Example 2.6. Compare two shooting player whose performance are recorded as follows:

score x_i	8	9	10
probability $p(x_i)$	0.3	0.1	0.6

player A

score x_i	8	9	10
probability $p(x_i)$	0.2	0.5	0.3

player B

Whose performance is better?

Idea: compare the weighted average $\sum_i x_i p(x_i)$. (values with higher probability have larger distribution!)

$$\text{Player A: } 8 \times 0.3 + 9 \times 0.1 + 10 \times 0.6 = 9.3$$
$$\text{Player B: } 8 \times 0.2 + 9 \times 0.5 + 10 \times 0.3 = 9.1$$

Since $9.3 > 9.1$, player A is better!

Definition 2.2 (Expected value $E(X)$). Let X be a discrete random variable with probability distribution $\{p(x_i), i = 1, 2, \ldots\}$. The expected value of X , denoted by $E(X)$ is given by

$$\boxed{E(X) = \sum_i x_i p(x_i).}$$

We also call it mean and denoted by μ.

Example 2.7. Expected value reflects the long-run average of repetitions of experiments.

In Example 2.6, let Player A shoot N times, (N is sufficiently large),

So the total score is roughly :

$0.3N$	times	8
$0.1N$	times	9
$0.6N$	times	10

$$\Rightarrow \text{the long-run average is } \frac{8 \times 0.3N + 9 \times 0.1N + 10 \times 0.6N}{N} = 9.3 = E(X).$$

Example 2.8. Roll a dice.

Let X be the random variable that denotes the number facing up.

x_i	1	2	3	4	5	6
$p(x_i)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

$$\begin{aligned} E(x) &= \sum_i x_i p(x_i) \\ &= 1 \cdot p(1) + 2 \cdot p(2) + \cdots + 6 \cdot p(6) \\ &= \frac{1}{6} + \frac{2}{6} + \frac{3}{6} + \frac{4}{6} + \frac{5}{6} + \frac{6}{6} \\ &= 3.5 \end{aligned}$$

Definition 2.3 (Variance $\operatorname{Var}(X)$ and Standard Deviation σ). Let X be a discrete random variable with probability distribution $\{p(x_i), i = 1, 2, \ldots\}$. Then the variance of X, is

$$\boxed{\operatorname{Var}(X) = \sum_i (x_i - \mu)^2 p(x_i),}$$

where μ is the mean (expected value) of X.

Standard deviation is defined as

$$\boxed{\sigma = \sqrt{\operatorname{Var} X}.}$$

Remark. $\operatorname{Var}(X)$ measures how far the value of X spread out of the mean $E(X)$.

In example 2.8, the variance is

$$\begin{aligned} \operatorname{Var}(X) &= \sum_i (x_i - \mu)^2 p(x_i) \\ &= \frac{(1-3.5)^2}{6} + \frac{(2-3.5)^2}{6} + \frac{(3-3.5)^2}{6} + \frac{(4-3.5)^2}{6} + \frac{(5-3.5)^2}{6} + \frac{(6-3.5)^2}{6} \\ &= \frac{35}{12} \\ \sigma &= \sqrt{\frac{35}{12}} \end{aligned}$$

Theorem 2. *Let X be a discrete random variable with probability distribution $\{p(x_i), i = 1, 2, \ldots\}$, then*

$$\boxed{\mathrm{Var}(X) = E(X^2) - (E(X))^2.}$$

Proof.

$$\begin{aligned}
\mathrm{Var}(X) &= \sum_i (x_i - \mu)^2 p(x_i) \\
&= \sum_i (x_i^2 - 2x_i\mu + \mu^2) p(x_i) \\
&= \sum_i x_i^2 p(x_i) - 2\mu \underbrace{\sum_i x_i p(x_i)}_{=\mu} + \mu^2 \underbrace{\sum_i p(x_i)}_{=1} \\
&= \sum_i x_i^2 p(x_i) - \mu^2 \\
&= E(X^2) - (E(X))^2
\end{aligned}$$

□

In Example 2.6,

$$\begin{aligned}
\text{For Player A: } \mathrm{Var}(X) &= (8-9.3)^2 \times 0.3 + (9-9.3)^2 \times 0.1 + (10-9.3)^2 \times 0.6 = 0.81 \\
\text{or: } \mathrm{Var}(X) &= 8^2 \times 0.3 + 9^2 \times 0.1 + 10^2 \times 0.6 - (9.3)^2 = 0.81
\end{aligned}$$

3 Continuous Random variable

Example 3.1. A certain traffic light remains red for 50 seconds every time. Andy arrives (at random) at the light and finds it red. Let X be the random variable denote the waiting time of Andy.

Image of X : $[0, 50]$, continuous random variable

1. probability that Andy has to wait for **at most** 10 seconds:

$$P(0 \le X \le 10) = \frac{\text{length of } [0, 10]}{\text{length of } [0, 50]} = \frac{10}{50} = \frac{1}{5}.$$

2. probability that Andy has to wait between 20 seconds and 40 seconds:

$$P(20 \leq X \leq 40) = \frac{\text{length of } [20,40]}{\text{length of } [0,50]} = \frac{20}{50} = \frac{2}{5}.$$

Definition 3.1. Let X be a continuous random variable. A probability distribution (or probability density function short form: pdf) is a function f such that

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

Under this distribution, we have

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

Definition 3.2. Let $f(x)$ be a probability density function. The cumulative distribution function (short form: cdf) is

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

Theorem 3 (Properties of $F(x)$, $f(x)$)**.**

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

2. $\begin{cases} F(x) \textit{ is non-decreasing}, \quad 0 \leq F(x) \leq 1. \\ \lim_{x\to-\infty} F(x) = 0, \quad \lim_{x\to+\infty} F(x) = 1. \end{cases}$

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

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

$P(X \geq b) = \int_b^{+\infty} f(x)\,dx$

$P(X = a) = 0$

4. $\begin{cases} F(x) = \int_{-\infty}^{x} f(t)\,dt \\ f(x) = F'(x), \quad \textit{except at discontinuity points of } f(x). \end{cases}$

Example 3.2. Uniform Distribution
In Example 3.1,

For $x \in [0, 50]$, $P(X \le x) = \dfrac{x}{50}$.

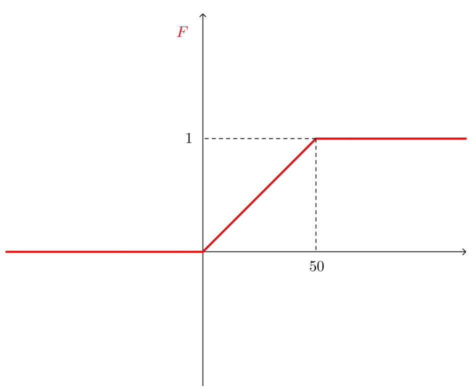
Let's extend the image from $[0, 50]$ to $\mathbb{R}$, define

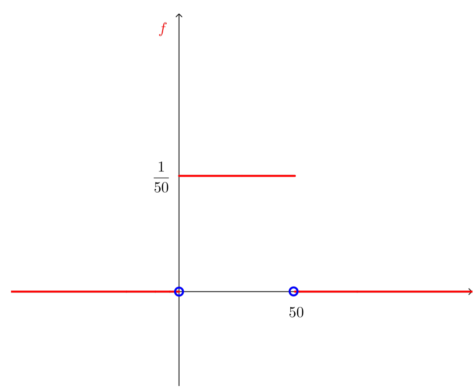
$$F(x) = P(X \le x) = \begin{cases} 0, & x < 0, \\ \frac{x}{50}, & 0 \le x \le 50, \\ 1, & x > 50. \end{cases} \quad \text{cumulative distribution function}$$

$$f(x) = \begin{cases} \frac{1}{50}, & 0 \le x \le 50, \\ 0, & \text{otherwise.} \end{cases} \quad \text{probability density function}$$

Check that it is a probability distribution:

$$\int_{-\infty}^{+\infty} f(x)dx = \int_{0}^{50} \frac{dx}{50} = \left[\frac{x}{50}\right]_0^{50} = 1.$$





$$F(x) = \int_{-\infty}^{x} f(t)\,dt$$
$$f(x) = F'(x) \quad (\text{ except at } x = 0, 50\text{, but it does not matter!})$$

Example 3.3. Let

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

be a probability density function.

Find C, $F(x)$, $P(-1 \leq x \leq 2)$.

Solution. We should have $\displaystyle\int_{-\infty}^{+\infty} f(x)\,dx = 1$. Therefore,

$$\begin{aligned} 1 &= \int_0^{+\infty} Ce^{-x}\,dx \\ &= \lim_{b\to+\infty} \int_0^b Ce^{-x}\,dx \\ &= \lim_{b\to+\infty} C(1 - e^{-b}) \\ &= C \end{aligned}$$

$$\begin{aligned} F(x) &= \int_{-\infty}^x f(t)\,dt \\ &= \begin{cases} \text{if } x < 0, & \displaystyle\int_{-\infty}^x 0\,dt = 0. \\ \text{if } x \geq 0, & \displaystyle\int_0^x e^{-t}\,dt = 1 - e^{-x}. \end{cases} \end{aligned}$$

$$P(-1 \leq X \leq 2) = \int_{-1}^2 f(x)\,dx = \int_0^2 e^{-x}\,dx = 1 - e^{-2}.$$

or

$$P(-1 \leq X \leq 2) = F(2) - F(-1) = 1 - e^{-2} - 0 = 1 - e^{-2}.$$

■

Let $G(x)$ be an antiderivative of $f(t)$. Then

$$\int_a^x f(t)dt = G(x) - G(a).$$

So

$$\frac{d}{dx}(G(x) - G(a)) = G'(x) = f(x).$$

Thus

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

is an antiderivative of $f(x)$, i.e.

$$f(x) = F'(x).$$

Example 3.4. Let

$$F(x) = \begin{cases} \frac{x}{x+1} & x \geq 0, \\ 0 & x < 0 \end{cases}$$

be a cumulative distribution function. Find the density distribution function.

Solution. The density distribution function is

For $x > 0$,

$$F'(x) = \frac{d}{dx}\frac{x}{x+1} = \frac{1}{(1+x)^2}.$$

For $x < 0$

$$F'(x) = 0$$

Thus the density distribution function is

$$f(x) = F'(x) = \begin{cases} \dfrac{1}{(1+x)^2} & x \geq 0, \\ 0 & x < 0. \end{cases}$$

■

3.1 Expected Value and Variance

Recall for discrete random variable, the expected value of X is

$$E(X) = \sum_i x_i p(x_i)$$

Now for a continuous random variable X. Let $f(x)$ be the probability distribution.

$$P(x_{i-1} \leq X \leq x_i) \approx f(x_i)\Delta x_i,$$

where $\Delta x_i = x_i - x_{i-1}$. Then

$$E(X) \approx \sum \text{value} \times \text{probability} = \sum_{i=1}^{n} x_i f(x_i)\Delta x_i.$$

This is the Riemann sum, let $\Delta x_i \to 0$, we have

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

Similarly

$$\text{Var}(X) \approx \sum_{i=1}^{n} (x_i - \mu)^2 f(x_i) \Delta x_i.$$

Let $\Delta x_i \to 0$,

$$\text{Var}(X) = E((X-\mu)^2) = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x).$$

Definition 3.3. Let X be a continuous random variable with density function $f(x)$

Expected value (mean): $\boxed{E(X) = \int_{-\infty}^{+\infty} x f(x)\, dx}$

Variance: $\boxed{\text{Var}(X) = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)\, dx}$

Standard deviation: $\boxed{\sigma = \sqrt{\text{Var}(X)}}$

Theorem 4. *Let X be a continuous random variable with probability density function $f(x)$, then*

$$\boxed{\text{Var}(X) = E(X^2) - (E(X))^2.}$$

Example 3.5. Uniform Distribution
Let

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

$$\begin{aligned} \mu = E(X) &= \int_{-\infty}^{+\infty} x f(x) dx \\ &= \int_a^b \frac{x}{b-a} dx \\ &= \frac{a+b}{2} \end{aligned}$$

$$\begin{aligned}\mathrm{Var}(X) &= \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)\,dx \\ &= \int_a^b (x - \frac{a+b}{2})^2 \frac{1}{b-a}\,dx \\ &= \frac{(b-a)^2}{12}\end{aligned}$$

or

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

$$\begin{aligned}\mathrm{Var}(X) = E(X^2) - E(X)^2 &= \frac{a^2+ab+b^2}{3} - \left(\frac{a+b}{2}\right)^2 \\ &= \frac{a^2-2ab+b^2}{12} = \frac{(a-b)^2}{12}.\end{aligned}$$

$$\sigma = \sqrt{\mathrm{Var}(X)} = \frac{b-a}{2\sqrt{3}}.$$

Example 3.6. Exponential Distribution
Let

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

with $\lambda > 0$.

Solution.

$$
\begin{aligned}
E(X) &= \int_{-\infty}^{+\infty} x f(x) dx = \int_{0}^{+\infty} \lambda x e^{-\lambda x} dx \\
&= \lim_{b\to+\infty} \int_0^b x d(-e^{\lambda x}) = \lim_{b\to+\infty} \left\{ \left[-xe^{-\lambda x} \right]_0^b + \int_0^b e^{-\lambda x} dx \right\} \\
&= \lim_{b\to+\infty} \int_0^b e^{-\lambda x} dx = \lim_{b\to+\infty} \left\{ \left[-\frac{1}{\lambda} e^{-\lambda x} \right]_0^b \right\} \\
&= \frac{1}{\lambda}
\end{aligned}
$$

$$
\begin{aligned}
E(X^2) &= \int_{-\infty}^{+\infty} x^2 f(x) dx = \int_{0}^{+\infty} \lambda x^2 e^{-\lambda x} dx \\
&= \lim_{b\to+\infty} \int_0^b x^2 d(-e^{\lambda x}) = \lim_{b\to+\infty} \left\{ \left[-x^2 e^{-\lambda x} \right]_0^b + \int_0^b 2xe^{-\lambda x} dx \right\} \\
&= \lim_{b\to+\infty} \int_0^b 2x d(-\frac{e^{-\lambda x}}{\lambda}) = \lim_{b\to+\infty} \left\{ \left[-\frac{2xe^{-\lambda x}}{\lambda} \right]_0^b + \int_0^b \frac{2e^{-\lambda x}}{\lambda} dx \right\} \\
&= \lim_{b\to+\infty} \left[-\frac{2e^{-\lambda x}}{\lambda^2} \right]_0^b = \frac{2}{\lambda^2}
\end{aligned}
$$

So

$$
\mathrm{Var}(X) = E(X^2) - E(X)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.
$$

■

Example 3.7. Let X be a random variable that measures that duration of cell phone calls in a certain city and assume that X has a density function

$$
f(t) = \begin{cases} 0.5e^{-0.5t} & \text{if } t \geq 0 \\ 0 & \text{if } t < 0 \end{cases}
$$

where t denotes the duration (in minutes) of a randomly selected call.

1. Find the probability that a randomly selected call last no more than 1 minute.
2. Find the probability that a randomly selected call last at least 2 minutes.

Solution. 1.

$$
\begin{aligned}
P(0 \leq X \leq 1) &= \int_0^1 f(t) dt = \int_0^1 0.5e^{-0.5t} dt \\
&= \left[-e^{-0.5t} \right]_0^1 = 1 - e^{-0.5} \approx 0.3935
\end{aligned}
$$

2.

$$\begin{aligned} P(X \geq 2) &= \int_2^{+\infty} f(t)dt = \lim_{b \to +\infty} \int_2^b 0.5e^{-0.5t}dt \\ &= \lim_{b \to +\infty} \left[-e^{-0.5t}\right]_2^b = \lim_{b \to +\infty} (e^{-1} - e^{-0.5b}) \\ &= e^{-1} \approx 0.3679 \end{aligned}$$

■

Remark

1. $P(X \geq 2) = P(X > 2)$. (Why?)
2. How to obtain the probability distribution? Even if we conduct a survey to collect all data, we have only finitely many calls (though is a big number) then the sample space $S =$ set of all calls. $X : S \to \mathbf{R}$ is a random variable that denotes the duration. The image of S is still finite. But it can be approximated by the X in the question.