Probability & Statistics

Dongjiao Ge

djge@cityu.edu.mo

Session 04

Chapter 3 Random variables

3.1 What are random variables?

The Holy Roman Empire was, in the words of the historian Voltaire, "*neither holy, nor Roman, nor an empire*".

Similarly, a **random variable** is **neither random nor a variable**:

A random variable is a function defined on a sample space.

- The values of the function can be anything at all, but for us they will always be numbers.
- The standard abbreviation for 'random variable' is r.v.

Example I select at random a student from the class and measure his or her height in centimetres.

Here, the ***sample space*** is *the set of students*; the ***random variable*** is '*height*', which is a function from the set of students to the real numbers: $h(S)$ is the height of student S in centimetres.

r.v. height is the function $h(S) : S \to R$

Here the domain set is the sample space S, the set of students in the class, and the target space is the set of real numbers R.

Example I throw a six-sided die twice; I am interested in the sum of the two numbers. Here the sample space is

$$\mathcal{S} = \{(i, j) : 1 \leq i, j \leq 6\},$$

and the random variable F is given by $F(i, j) = i + j$. The target set is the set $\{2,3, \ldots, 12\}$.

The *two random variables in the above examples* are representatives of the **two types of random variables** that we will consider.

*A **random variable** F is **discrete** if the values it can take are separated by gaps.*

For example, F is discrete if

- it can take **only finitely many values** (as in the second example above, where the values are the integers from 2 to 12), or
- the values of F are **integers** (for example, the number of nuclear decays which take place in a second in a sample of radioactive material – the number is an integer but we can't easily put an upper limit on it.)

*A **random variable** is **continuous** if there are no gaps between its possible values.*

- In the first example, the height of a student could in principle be any real number between certain extreme limits.
- A random variable whose values range over an interval of real numbers, or even over all real numbers, is continuous.

3.2 Probability mass function

F --- a ***discrete random variable***.

The most basic question we can ask is: given any value a in the target set of F, what is the probability that F takes the value a?

Alternatively, if we consider the event

$$A = \{x \in \mathcal{S} : F(x) = a\}$$

What is the probability $P(A)$?

we simplify the notation: we write $P(F = a)$ in place of $P(\{x \in \mathcal{S} : F(x) = a\})$.

The ***probability mass function*** of a *discrete random variable* F is the function, formula or table which gives the value of $P(F = a)$ for each element a in the target set of F.

- If F takes only a few values, it is convenient to list it in a table;
- Otherwise, we should give a formula if possible.

The standard abbreviation for ‘probability mass function’ is **p.m.f.**

Example I toss a fair coin three times. The random variable X gives the number of heads recorded. The possible values of X are 0,1,2, 3, and its p.m.f. is

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

Basically, we have the sample space is $\{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$, and each outcome is equally likely.

- The event $X = 1$, when written as a set of outcomes, is equal to $\{TTT\}$, and has probability 1/8.
- The event $X = 1$, when written as a set of outcomes, is equal to $\{HTT, THT, TTH\}$, and has probability 3/8.
- The event $X = 2$, when written as a set of outcomes, is equal to $\{HHT, HTH, THH\}$, and has probability 3/8.
- The event $X = 2$, when written as a set of outcomes, is equal to $\{HHH\}$, and has probability 1/8.

Q1 An urn contains 20 balls numbered 1,2,…,20. Select 5 balls at random, without replacement. Let X be the largest number among selected balls.
Determine its p. m. f. and the probability that at least one of the selected numbers is 15 or more.

Two random variables X and Y are said to have **<u>the same distribution</u>** if the values they take and ***their probability mass functions are equal***. We write $X \sim Y$ in this case.

In the above example, if Y is the number of tails recorded during the experiment, then X and Y have the same distribution, even though their actual values are different (indeed, $Y = 3 - X$).

3.3 Expected value and variance

Let X be a discrete random variable which takes the values $a_1, \ldots, a_n$. The expected value or mean of X is the number $E(X)$ given by the formula

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

That is, we multiply each value of X by the probability that X takes that value, and sum these terms.

- The expected value is a kind of 'generalised average': if each of the values is equally likely, so that each has probability $1/n$, then $E(X) = (a_1 + \cdots + a_n)/n$, which is just the average of the values.

There is an interpretation of the expected value in terms of mechanics.
If we put a mass p_i on the axis at position a_i for $i = 1, \ldots, n$, where $p_i = P(X = a_i)$, then the centre of mass of all these masses is at the point $E(X)$.

If the random variable X takes *infinitely many values*, say $a_1, a_2, a_3, \ldots$, then we define the expected value of X to be the infinite sum

$$E(X) = \sum_{i=1}^{\infty} a_i P(X = a_i).$$

We won't discuss the convergence of the above series.

Usually, discrete random variables will only have finitely many values; in the few examples we consider where there are infinitely many values, the series will usually be a geometric series or something similar, which we know how to sum.

In the following proofs, we assume that the number of values is finite.

The variance of X is the number $Var(X)$ given by

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

Here, X^2 is just the random variable whose values are the squares of the values of X. Thus

$$E(X^2) = \sum_{i=1}^{n} a_i^2 P(X = a_i)$$

(It would be an infinite sum, if necessary).

The next proposition shows that, if $E(X)$ is a kind of average of the values of X, then $Var(X)$ is a measure of how spread-out the values are around their average.

Proposition 3.1 *Let X be a discrete random variable with $E(X) = \mu$. Then*

$$\mathrm{Var}(X) = E((X-\mu)^2) = \sum_{i=1}^{n} (a_i - \mu)^2 P(X = a_i).$$

The second term is equal to the third by definition, and the third is

$$\begin{aligned}
&\sum_{i=1}^{n} (a_i - \mu)^2 P(X = a_i) \\
= \quad & \sum_{i=1}^{n} (a_i^2 - 2\mu a_i + \mu^2) P(X = a_i) \\
= \quad & \left(\sum_{i=1}^{n} a_i^2 P(X = a_i) \right) - 2\mu \left(\sum_{i=1}^{n} a_i P(X = a_i) \right) + \mu^2 \left(\sum_{i=1}^{n} P(X = a_i) \right).
\end{aligned}$$

Continuing, we find

$$\begin{aligned}
E((X-\mu)^2) &= E(X^2) - 2\mu E(X) + \mu^2 \\
&= E(X^2) - E(X)^2,
\end{aligned}$$

($E(X) = \mu$, and that $\sum_{i=1}^{n} P(X = a_i) = 1$ since the events $X = a_i$ form a partition.)

Example I toss a fair coin three times; X is the number of heads. What are the expected value and variance of X?

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

$$E(X) = 0 \times (1/8) + 1 \times (3/8) + 2 \times (3/8) + 3 \times (1/8) = 3/2,$$

$$\mathrm{Var}(X) = 0^2 \times (1/8) + 1^2 \times (3/8) + 2^2 \times (3/8) + 3^2 \times (1/8) - (3/2)^2 = 3/4.$$

We can also use Proposition 3.1 to calculate the variance

$$\mathrm{Var}(X) = \left(-\frac{3}{2}\right)^2 \times \frac{1}{8} + \left(-\frac{1}{2}\right)^2 \times \frac{3}{8} + \left(\frac{1}{2}\right)^2 \times \frac{3}{8} + \left(\frac{3}{2}\right)^2 \times \frac{1}{8} = \frac{3}{4}.$$

Remarks

- The expected value of X always lies between the smallest and largest values of X.
- The variance of X is ***never negative***. (For the formula in Proposition 3.1 is a sum of terms, each of the form $(a_i \ - \mu)$^2 (a square, hence non-negative) times $P(X \ = \ a_i)$ (a probability, hence non-negative).

Q2 Let X be the number shown on a rolled fair die. Compute $EX, E(X^2)$, and $Var(X)$.

3.4 Joint p.m.f. of two random variables

Let X be a random variable taking the values $a_1, \ldots, a_n$, and let Y be a random variable taking the values $b_1, \ldots, b_m$. We say that X and Y are ***independent*** if, for any possible values i and j, we have

$$P(X = a_i, Y = b_j) = P(X = a_i) \cdot P(Y = b_j).$$

$P(X = a_i, Y = b_j)$ --- the probability of the event that X takes the value a_i and Y takes the value b_j. So we could re-state the definition as follows:

The random variables X and Y are *independent* if, for any value a_i of X and any value b_j of Y, the events $X = a_i$ and $Y = b_j$ are independent (events).

Please be aware the difference between ‘independent events’ and ‘independent random variables’.

Example In Chapter 2, we saw the following: I have two red pens, one green pen, and one blue pen. I select two pens without replacement.

Then the events 'exactly one red pen selected' and 'exactly one green pen selected' turned out to be independent.
X --- the number of red pens selected, and Y --- the number of green pens selected. Then,

$$P(X = 1, Y = 1) = P(X = 1) \cdot P(Y = 1).$$

Are X and Y independent random variables?

The answer is **No**. The reason is $P(X = 2) = 1/6$, $P(Y = 1) = 1/2$, but $P(X = 2, Y = 1) = 0$ (it is impossible to have two red and one green in a sample of two).

If I roll a die twice, and X and Y are the numbers that come up on the first and second throws, then X and Y will be independent, even if the die is not fair (so that the outcomes are not all equally likely).

If we have more than two random variables (for example X, Y, Z), we say that they are *mutually independent* if the events that the random variables take specific values (for example, $X = a, Y = b, Z = c$) are mutually independent.

What about the expected values and variance of random variables? For expected value, it is easy, but for variance it helps if the variables are independent:

Theorem 3.2 *Let X and Y be random variables.*

(a) $E(X+Y) = E(X) + E(Y)$.

(b) If X and Y are independent, then $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$.

Let' look at the definition of marginal distributions before giving the proof of the above theorem.

If two random variables X and Y are **not independent**, then knowing the p.m.f. of each variable does not tell the whole story.

The joint probability mass function (or joint p.m.f.) of X and Y is the table giving, for each value a_i of X and each value b_j of Y, the probability that $X = a_i$ and $Y = b_j$.

- We arrange the table so that the rows correspond to the values of X and the columns to the values of Y.
- Note that summing the entries in the row corresponding to the value a_i gives the probability that $X = a_i$; that is, the row sums form the p.m.f. of X. Similarly the column sums form the p.m.f. of Y.

The row and column sums are sometimes called the ***marginal distributions*** *or* ***marginals****.*

In particular, X and Y are independent r.v.s if and only if each entry of the table is equal to the product of its row sum and its column sum.

Example I have two red pens, one green pen, and one blue pen, and I choose two pens without replacement. Let X be the number of red pens that I choose and Y the number of green pens. Then the joint p.m.f. of X and Y is given by the following table:

$X \backslash Y$	0	1
0	0	$\frac{1}{6}$
1	$\frac{1}{3}$	$\frac{1}{3}$
2	$\frac{1}{6}$	0

The row and column sums give us the p.m.f.s for X and Y:

a	0	1	2
$P(X = a)$	$\frac{1}{6}$	$\frac{2}{3}$	$\frac{1}{6}$

b	0	1
$P(Y = b)$	$\frac{1}{2}$	$\frac{1}{2}$

Q3 A car dealership sells 0, 1, or 2 luxury cars on any day. When selling a car, the dealer also tries to persuade the customer to buy an extended warranty for the car. Let *X* denote the number of luxury cars sold in a given day, and let *Y* denote the number of extended warranties sold.

$$P[X = 0,\ Y = 0] = 1/6$$
$$P[X = 1,\ Y = 0] = 1/12$$
$$P[X = 1,\ Y = 1] = 1/6$$
$$P[X = 2,\ Y = 0] = 1/12$$
$$P[X = 2,\ Y = 1] = 1/3$$
$$P[X = 2,\ Y = 2] = 1/6$$

Calculate the variance of *X*.

proof of Theorem 3.2.

We consider the joint p.m.f. of X and Y.
The random variable $X + Y$ takes the values $a_i + b_j$ for $i = 1,\ldots,n$ and $j = 1,\ldots,m$.
Now the probability that it takes a given value c_k is the sum of the probabilities $P(X = a_i, Y = b_j)$ over all i and j such that $a_i + b_j = c_k$. Thus,

$$\begin{aligned} E(X+Y) &= \sum_k c_k P(X+Y=c_k) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{m}(a_i+b_j)P(X=a_i,Y=b_j) \\ &= \left(\sum_{i=1}^{n} a_i \sum_{j=1}^{m} P(X=a_i,Y=b_j)\right) + \left(\sum_{j=1}^{m} b_j \sum_{i=1}^{n} P(X=a_i,Y=b_j)\right). \end{aligned}$$

$\sum_{j=1}^{m} P(X = a\_i, Y = b\_j)$ is a row sum of the joint p.m.f. table, so is equal to $P(X = a\_i)$, and similarly $\sum_{i=1}^{n} P(X = a_i, Y = b_j)$is a column sum and is equal to $P(Y = b_j)$. So

$$\begin{aligned} E(X+Y) &= \sum_{i=1}^{n} a_i P(X=a_i) + \sum_{j=1}^{m} b_j P(Y=b_j) \\ &= E(X)+E(Y). \end{aligned}$$

(a) is proved.

For the variance, first, we calculate

$$E((X+Y)^2) = E(X^2+2XY+Y^2) = E(X^2)+2E(XY)+E(Y^2),$$

using part (a) of the Theorem. We have to consider the term $E(XY)$. For this, we have to use the assumption that X and Y are independent, that is,

$$P(X=a_1, Y=b_j) = P(X=a_i)\cdot P(Y=b_j).$$

So, we have

$$
\begin{aligned}
E(XY) &= \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j P(X=a_i, Y=b_j) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i b_j P(X=a_i) P(Y=b_j) \\
&= \left(\sum_{i=1}^{n} a_i P(X=a_i)\right)\cdot\left(\sum_{j=1}^{m} b_j P(Y=b_j)\right) \\
&= E(X)\cdot E(Y).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\mathrm{Var}(X+Y) &= E((X+Y)^2) - (E(X+Y))^2 \\
&= (E(X^2) + 2E(XY) + E(Y^2)) - (E(X)^2 + 2E(X)E(Y) + E(Y)^2) \\
&= (E(X^2) - E(X)^2) + 2(E(XY) - E(X)E(Y)) + (E(Y^2) - E(Y)^2) \\
&= \mathrm{Var}(X) + \mathrm{Var}(Y).
\end{aligned}
$$

Proposition 3.3 *Let C be a constant random variable with value c. Let X be any random variable.*

(*a*) $E(C) = c$, $\mathrm{Var}(C) = 0$.

(*b*) $E(X + c) = E(X) + c$, $\mathrm{Var}(X + c) = \mathrm{Var}(X)$.

(*c*) $E(cX) = cE(X)$, $\mathrm{Var}(cX) = c^2\,\mathrm{Var}(X)$.

Proof

(a) The random variable C takes the single value c with $P(C = c) = 1$. So $E(C) = c \cdot 1 = c$. Also,

$$\mathrm{Var}(C) = E(C^2) - E(C)^2 = c^2 - c^2 = 0.$$

(For C^2 is a constant random variable with value c^2.)

(b) This follows immediately from Theorem 3.2, once we observe that the constant random variable C and any random variable X are independent. (This is true because $P(X = a, C = c) = P(X = a) \cdot 1$.) Then

$$E(X+c) = E(X) + E(C) = E(X) + c,$$
$$\mathrm{Var}(X+c) = \mathrm{Var}(X) + \mathrm{Var}(C) = \mathrm{Var}(X).$$

(c) If $a_1, \ldots, a_n$ are the values of X, then $ca_1, \ldots, ca_n$ are the values of cX, and $P(cX = ca_i) = P(x = a_i)$. So

$$\begin{aligned} E(cX) &= \sum_{i=1}^{n} ca_i P(cX = ca_i) \\ &= c \sum_{i=1}^{n} a_i P(X = a_i) \\ &= cE(X). \end{aligned}$$

$$\begin{aligned} \mathrm{Var}(cX) &= E(c^2X^2) - E(cX)^2 \\ &= c^2E(X^2) - (cE(X))^2 \\ &= c^2(E(X^2) - E(X)^2) \\ &= c^2\,\mathrm{Var}(X). \end{aligned}$$

Exercises

EX1 An urn contains 11 balls, 3 white, 3 red, and 5 blue balls. Take out 3 balls at random, without replacement. You win \$1 for each red ball you select and lose \$1 for each white ball you select. Determine the p. m. f. of X, the amount you win.

EX2 Rolling a fair die twice. Let the random variable X be the maximum of the two numbers obtained, and let Y be the modulus of their difference (that is, the value of Y is the larger number minus the smaller number).
(a) Write down the joint p.m.f. of (X, Y).
(b) Write down the p.m.f. of X, and calculate its expected value and its variance.
(c) Write down the p.m.f. of Y, and calculate its expected value and its variance.
(d) Are the random variables X and Y independent?

EX3 Let X and Y be discrete random variables with joint probability function

$$p(x, y) = \begin{cases} \dfrac{1}{21}, & \text{for } x = 0,1,\ldots,5 \text{ and } y = 0,\ldots,x \\ 0, & \text{otherwise.} \end{cases}$$

Calculate the variance of Y.