MAT 3103: Computational Statistics and Probability

Chapter 4: Random Variable

Random variable:

A random variable is a variable whose possible values are numerical outcomes of a random experiments. Usually written as X.

There are two types of random variables,

a) Discrete random variable

b) Continuous random variable

Discrete random variable:

A discrete random variable is one which may take on only a countable number of distinct values such as 0, 1, 2, ... Discrete random variables are usually (but not necessarily) counts.

The probability distribution of a discrete random variable is a list of probabilities associated with each of its possible values. It is also sometimes called the probability function or the probability mass function.

Suppose a random variable X may take k different values, with the probability that $X = x$ defined to be $P(X = x) = p(x)$. The probabilities $p(x)$ must satisfy the following:

1. $0 \leq p(x) \leq 1$
2. $\sum_{x=0}^{\infty} p(x) = 1.$

Continuous random variable:

A continuous random variable is one which takes an infinite number of possible values. Continuous random variables are usually measurements. A continuous random variable is not defined at specific values. Instead, it is defined over an interval of values,

Suppose a random variable X may take all values over an interval of real numbers. Then the probability that X is in the set of outcomes A, $P(A)$, is defined to be the area above A and under a curve. The curve, which represents a function $f(x)$, must satisfy the following:

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

Thus, $f(x)$ *is* also sometimes called the probability density function.

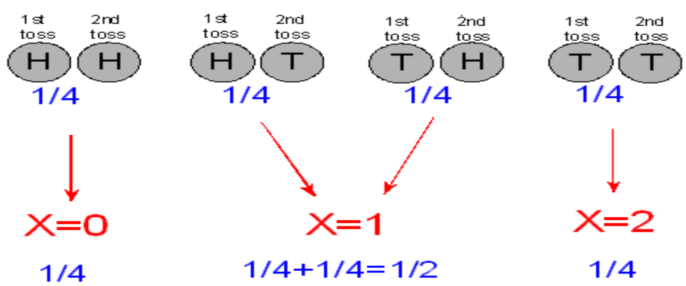
Application of random variable:

When we learned how to find probabilities by applying the basic principles, we mostly focused on just one particular outcome or event, like the probability of getting exactly one tail when a coin is tossed twice, or the probability of getting a 5 when a die is rolled. Now that we have mastered the solution of individual probability problems, we'll proceed to look at the big picture by considering all the possible values of a discrete random variable, along with their associated probabilities.

Explaining random variable with an example:

Consider an experiment where 2 fair coins are tossed. The sample space, S = {HH, HT, TH, TT}.
Let, X be the no. of tails. Then,

As X takes different values 0, 1 or 2 with associated probabilities, X is called random variable.



The probability distribution of the random variable X can be easily summarized in a table as:

X	0	1	2	Total
$p(x)$	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{1}{4}$	1

This $p(x)$ is also a probability function.

Joint probability function:

The joint probability function for X, Y, … is a probability distribution that gives the probability that each of X, Y, … falls in any particular range or discrete set of values specified for that variable. In the case of only two random variables, this is called a bivariate distribution, but the concept generalizes to any number of random variables, giving a multivariate distribution.

When X and Y are discrete, the probabilities $p(x, y)$ must satisfy the following:

1. $p(x, y) \geq 0$
2. $\sum_{x=0}^{\infty} \sum_{y=0}^{\infty} p(x, y) = 1$

When X and Y are continuous, the probabilities $f(x, y)$ must satisfy the following:

1. $f(x, y) \geq 0$
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, \mathrm{d}x\mathrm{d}y = 1$.

Mean and Variance of Random variable:

Mean: The average value of a random variable is called the expected value of the random variable. If X is a random variable, then its mean (expectation) is given by-

$$E(X) = \sum_{x=0}^{\infty} x\, p(x), \text{ if } X \text{ is discrete}$$

$$= \int_{-\infty}^{\infty} x\, f(x)\mathrm{d}x, \text{ if } X \text{ is continuous.}$$

Variance: If X is a random variable, then its variance is given by-

$$V(X) = \sigma^2 = E[X - E(X)]^2 = E(X^2) - [E(X)]^2$$

Where, $E(X^2) = \sum_{x=0}^{\infty} x^2\, p(x)$, if X is discrete

$$= \int_{-\infty}^{\infty} x^2\, f(x)\mathrm{d}x, \text{ if } X \text{ is continuous.}$$

Some properties regarding random variable:

- If A is any constant and X is a random variable, then E(A) = A, V(A) = 0.
- If A and B are any two constants and X is a random variable, then

  $E(AX \pm B) = AE(X) \pm B$,

  $V(AX \pm B) = A^2V(X)$.
- If A and B are any two constants and X and Y are two random variables, then

  $E(AX \pm BY) = AE(X) \pm BE(Y)$

  $V(AX \pm BY) = A^2V(X) + B^2V(Y) \pm 2AB\,cov(XY)$

  [if X and Y are not independent]

  $V(AX \pm BY) = A^2V(X) + B^2V(Y)$

  [if X and Y are independent, then $cov(XY) = 0$]
- If X and Y are two independent random variables, then $E(XY) = E(X)\,E(Y)$.

Note: Covariance of X and Y: $cov(XY) = E(XY) - E(X)E(Y)$

Example 4.1: A random sample of graduates was surveyed at AIUB. A question that was asked is: How many times did you change majors? The results are shown in a probability distribution as:

X	0	1	2	3	4	5	Total
P(X)	0.28	0.37	0.23	0.09	0.02	0.01	1

Find the probability that a randomly selected graduate changed majors (a) more than once, (b) less or two times, and (c) four or more times.

Solution:

(a) $P(X > 1) = P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5) = 0.23 + 0.09 + 0.02 + 0.01 = 0.35$.

(b) $P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = 0.28 + 0.37 + 0.23 = 0.88$.

(c) $P(X \geq 4) = P(X = 4) + P(X = 5) = 0.02 + 0.01 = 0.03$.

Example 4.2: A fair coin is tossed three times. Let the random variable X be the number of heads. Find the mean, variance and standard deviation of X. Also, find E (X + 1), E (2X - 1), V(X - 1), V (2X + 1).

Solution: Sample space, S = {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}. Then,

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

Mean: $E(X) = \sum_{x=0}^{\infty} x\,p(x) = (0 \times \frac{1}{8}) + (1 \times \frac{3}{8}) + (2 \times \frac{3}{8}) + (3 \times \frac{1}{8}) = \frac{12}{8} = \frac{3}{2} = 1.5.$

$E(X^2) = \sum_{x=0}^{\infty} x^2\,p(x) = (0^2 \times \frac{1}{8}) + (1^2 \times \frac{3}{8}) + (2^2 \times \frac{3}{8}) + (3^2 \times \frac{1}{8}) = 3,$

Variance: $V(X) = E(X^2) - [E(X)]^2 = 3 - (\frac{3}{2})^2 = \frac{3}{4} = 0.75$, Standard deviation, $\sigma = \sqrt{V(X)} = \sqrt{0.75}$.

$E(X + 1) = E(X) + 1 = 1.5 + 1 = 2.5.$

$E(2X - 1) = 2E(X) - 1 = 3 - 1 = 2.$

$V(X - 1) = V(X) = 0.75.$

$V(2X + 1) = 2^2\,V(X) = 4 \times 0.75 = 3.$

Example 4.3: Find the mean and variance of the numbers obtained on a single roll of a fair die.

Solution: Let X = numbers obtained on a single roll. Then,

X	1	2	3	4	5	6	Total
P(X)	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1

Mean, $E(X) = \sum_{x=0}^{\infty} x\,p(x) = (1 \times \frac{1}{6}) + (2 \times \frac{1}{6}) + (3 \times \frac{1}{6}) + (4 \times \frac{1}{6}) + (5 \times \frac{1}{6}) + (6 \times \frac{1}{6}) = \frac{21}{6},$

$E(X^2) = \sum_{x=0}^{\infty} x^2\,p(x) = (1^2 \times \frac{1}{6}) + (2^2 \times \frac{1}{6}) + (3^2 \times \frac{1}{6}) + (4^2 \times \frac{1}{6}) + (5^2 \times \frac{1}{6}) + (6^2 \times \frac{1}{6}) = \frac{91}{6},$

Variance, $V(X) = E(X^2) - [E(X)]^2 = \frac{91}{6} - (\frac{21}{6})^2 = \frac{105}{36}.$

Example 4.4: Walton's production line produces a variable number of defective parts in an hour, with probabilities shown in this table:

X	0	1	2	3	4	Total
P(X)	0.15	0.30	0.25	0.20	0.10	1

How many defective parts are expected to be produced in an hour on Walton's production line?

Solution: E(X) = $\sum_{x=0}^{\infty} x\,p(x)$ = (0 × 0.15) + (1 × 0.3) + (2 × 0.25) + (3 × 0.2) + (4 × 0.1) = 1.8.

Example 4.5: The length of time X, needed by students in a Math course at AIUB, to complete a 1-hour exam is a random variable with probability density function (pdf) given by:

$$f(x) = \begin{cases} k(x^2 + x);\ 0 < x < 1 \\ 0;\ otherwise \end{cases}$$

Estimate the value of k. Also, find the average time needed to complete a 1-hour exam.

Solution: A given pdf must integrate to 1. Hence,

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

$$= k\int_0^1 (x^2 + x)dx$$

$$= k\left(\frac{x^3}{3} + \frac{x^2}{2}\right)\Big|_0^1$$

$$= k\left(\frac{5}{6}\right)$$

So, $k = \frac{6}{5}$.

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} xf(x)dx$$

$$= \frac{6}{5}\int_0^1 x(x^2 + x)dx$$

$$= \frac{6}{5}\int_0^1 (x^3 + x^2)dx$$

$$= \frac{6}{5}\left(\frac{x^4}{4} + \frac{x^3}{3}\right)\Big|_0^1$$

$$= \frac{6}{5}\left(\frac{1}{4} + \frac{1}{3}\right)$$

$$= \frac{7}{10}.$$

Example 4.6: The probability density function of a random variable X is given by:

$$f(x) = \frac{x}{2};\ 0<x<2.$$

Find (a) $P(X<1)$, (b) $P(X>0.5)$ and (c) $P(1<X<2)$.

Solution: $P(X<1) = \int_0^1 f(x)\,dx = \int_0^1 \frac{x}{2}\,dx = [\ \frac{x^2}{4}\ ]_0^1 = \frac{1}{4}$.

$$P(X>0.5) = \int_{0.5}^2 f(x)\,dx = \int_{0.5}^2 \frac{x}{2}\,dx = [\ \frac{x^2}{4}\ ]_{0.5}^2 = [\frac{2^2}{4} - \frac{0.5^2}{4}] = 0.9375.$$

$$P(1<X<2) = \int_1^2 f(x)\ dx\ =\ \int_1^2 \frac{x}{2}\,dx = [\ \frac{x^2}{4}\ ]_1^2 = [\frac{2^2}{4} - \frac{1^2}{4}] = \frac{3}{4}.$$

Example 4.7: A University uses a certain software to check errors in any program. The number of errors found is denoted by a random variable *X* whose probability density function (pdf) given by:

$$f(x) = \begin{cases} \frac{2(x+2)}{5};\ 0 < x < 4 \\ 0; otherwise \end{cases}$$

Find the average number of errors a University expects to find in a given program.

Solution: The average number of errors a university expects to find in a given program is

$$\mathrm{E}\,(X) = \int_{-\infty}^{\infty} xf(x)dx = \frac{2}{5}\int_{0}^{4} x(x+2)dx = \frac{2}{5}\int_{0}^{4}(x^2+2x)dx = \frac{2}{5}\left(\frac{x^3}{3}+x^2\right)\Big|_0^4 = \frac{224}{15}.$$

Example 4.8: Find the variance of a random variable *X* whose probability density function is:

$$f(x) = \begin{cases} 2(1-\ x);\ 0 < x < 1 \\ 0; otherwise \end{cases}$$

Also, find E (*X* - 1), E (3*X* + 1), V(*X* + 1) and V(3*X* - 1).

Solution:

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_{0}^{1} x(2(1-x))dx = 2\int_{0}^{1}(x-x^2)dx = 2\left(\frac{x^2}{2}-\frac{x^3}{3}\right)\Big|_0^1 = \frac{1}{3}\,,$$

$$\mathrm{E}(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx = \int_{0}^{1} x^2(2(1-x))dx = 2\int_{0}^{1}(x^2-x^3)dx = 2\left(\frac{x^3}{3}-\frac{x^4}{4}\right)\Big|_0^1 = \frac{1}{6},$$

$$\mathrm{V}(X) = \mathrm{E}(X^2) - [\mathrm{E}(X)]^2 = \frac{1}{6} - \left(\frac{1}{3}\right)^2 = \frac{1}{18},$$

E (3*X* + 1) = 3E(*X*) + 1 = 1 + 1 = 2.

E $(X - 1) = \mathrm{E}(X) - 1 = \frac{1}{3} - 1 = -\frac{2}{3}$.

V(*X* + 1) = V(*X*) = $\frac{1}{18}$.

V(3*X* − 1) = 3^2 V(*X*) = $9 \times \frac{1}{18} = \frac{1}{2}$.

Example 4.9: The joint probability density function of two random variables X and Y is given by: $f(x, y) = 2x$, $0 < x < 1$, $0 < y < 1$. Find (a) marginal distribution of X, (b) marginal distribution of Y, (c) E(X – 2Y), and (d) V(2X – Y). Are X and Y independent?

Solution: Marginal distribution of X: $g(x) = \int_y f(x,y)dy = \int_0^1 2x\ dy = 2x\ (y)|_0^1 = 2x$.

Marginal distribution of Y: $h(y) = \int_x f(x,y)\ dx = \int_0^1 2x\ dx = \left(\frac{2x^2}{2}\right)\Big|_0^1 = 1$.

Hence, X and Y are independent random variables since $g(x) \times h(y) = 2x = f(x, y)$.

$E(X) = \int_x x\ g(x)\ dx = \int_0^1 x\ 2x\ dx = \int_0^1\ 2x^2\ dx = \left(\frac{2x^3}{3}\right)\Big|_0^1 = ,$

$E(X^2) = \int_x x^2 g(x)\ dx = \int_0^1 x^2\ 2x\ dx = \int_0^1\ 2x^3\ dx = [\left(\frac{2x^4}{4}\right)\Big|_0^1 = \frac{1}{2},$

$E(Y) = \int_y y\ h(y)\ dy = \int_0^1 y\ dy = \left(\frac{y^2}{2}\right)\Big|_0^1 = \frac{1}{2},$ $\quad E(Y^2) = \int_y y^2\ h(y)\ dy = \int_0^1 y^2\ dy = \left(\frac{y^3}{3}\right)\Big|_0^1 = \frac{1}{3},$

$V(X) = E(X^2) - [E(X)]^2 = \frac{1}{2} - (\frac{2}{3})^2 = \frac{1}{18}.$ $\quad V(Y) = E(Y^2) - [E(Y)]^2 = \frac{1}{3} - (\frac{1}{2})^2 = \frac{1}{12}.$

$E(X - 2Y) = E(X) - 2E(Y) = \frac{2}{3} - (2\times \frac{1}{2}) = -\frac{1}{3}.$ $\quad V(2X\text{-}Y) = 4V(X) + V(Y) = (4\times \frac{1}{18}) + \frac{1}{12} = \frac{11}{36}.$

Example 4.10: The joint probability function of two random variables *X* and *Y* is given as:

X \ *Y*	1	2	3	4	5	6	g(*X*)
1	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{3}$
2	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{3}$
3	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{3}$
h(*Y*)	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1

Find (a) marginal distribution of *X*, (b) marginal distribution of *Y*, (c) E(*X*) and (d) E(*Y*).

Solution: Marginal distribution of *X*:

X	1	2	3	Total
P(*X*)	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	1

E(*X*) = $\sum_{x=0}^{\infty} x\,p(x)$ = $(1 \times \frac{1}{3}) + (2 \times \frac{1}{3}) + (3 \times \frac{1}{3}) = 2$.

Marginal distribution of *Y*:

Y	1	2	3	4	5	6	Total
P(*Y*)	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1

E(*Y*) = $\sum_{y=0}^{\infty} y\,p(y)$ = $(1 \times \frac{1}{6}) + (2 \times \frac{1}{6}) + (3 \times \frac{1}{6}) + (4 \times \frac{1}{6}) + (5 \times \frac{1}{6}) + (6 \times \frac{1}{6}) = \frac{21}{6}$.

Exercise 4

4.1. Walton's production line produces a variable number of defective parts in an hour, with probabilities shown in this table:

X	1	2	3	4	Total
$P(X)$	0.10	0.30	0.45	0.15	1

How many defective parts are expected to be produced in an hour on Walton's production line?

4.2. A random sample of graduates was surveyed at AIUB. A question that was asked is: How many times did you change majors? The results are shown in a probability distribution as:

X	0	1	2	3	Total
$P(X)$	0.20	0.30	0.20	0.30	1

Find the probability that a randomly selected graduate changed majors (a) more than once, (b) two or less times, and (c) two or more times. Also estimate (d) E $(2X + 1)$ and (e) V $(4X - 3)$ mean, and (f) variance for the random variable X.

4.3. The probability density function of a random variable X is given by,

$$f(x) = \frac{x}{2}\ ;\ 0 < x < 2.$$

Find $P(X < 1)$, (b) $P(X > 0.5)$, (c) $P(1 < X < 2)$ (d) $E(X)$ and (e) $V(X)$ for the random variable X.

4.4. The probability density function of a random variable X is given by,

$$f(x) = 2x \; ; 0 < x < 1.$$

Estimate (a) $P(X > \frac{1}{2})$, (b) $P(\frac{1}{3} < X < \frac{1}{2})$, (c) mean and (d) variance for the random variable X.

4.5. The probability density function of a random variable Y is given by,

$$f(y) = 1; 0 < y < 1.$$

Find (a) mean and (b) variance for the random variable Y.

4.6. The joint probability density function of two random variables X and Y is given by,

$$f(x, y) = 4xy,\ 0 < x < 1,\ 0 < y < 1.$$

Show that, X and Y are independent. Also calculate, (a) $P\left(X > \frac{1}{2}\right)$, (b) $P\left(\frac{1}{3} < Y < \frac{1}{2}\right)$, (c) E $(2X + Y)$ and (d) V $(2X - 3Y)$.

4.7. The joint probability density function of two random variables X and Y is given by:

$$f(x, y) = 4x(1 - y),\ 0 < x < 1,\ 0 < y < 1.$$

Show that, X and Y are independent random variables.

4.8. The joint probability function of two random variables *X* and *Y* is given as:

X \ *Y*	0	1	2
0	0.10	0.20	0.00
1	0.10	0.20	0.10
2	0.00	0.10	0.10
3	0.10	0.00	0.00

Find (a) marginal distribution of *X*, (b) marginal distribution of *Y*, (c) E $(5X - Y)$ and (d) V$(Y - 2X)$.

4.9. The number of adults living in homes on a randomly selected block in Bashundhara residential area is described by the following probability distribution:

X	1	2	3	4	Total
P(*X*)	0.25	0.50	0.15	0.10	1

What is the standard deviation of the probability distribution?

4.10. Find the value of *c* if a random variable *X* has the probability density function (pdf) given by:

$$f(x) = \begin{cases} c(1 - x^2); & -1 < x < 1 \\ 0; & otherwise \end{cases}$$

Also calculate E (X) and P ($X > 0$).

4.11. The probability that 0, 1, 2, 3, or 4 people will be placed on hold when they call a radio talk show is given by the following probability distribution. Find the mean and variance for the data.

X	0	1	2	3	4	Total
$P(X)$	0.18	0.34	0.23	0.21	0.04	1

4.12. A fair coin is tossed twice. Then X is a random variable of the number of with the following probability function.

X	0	1	2	Total
$P(X)$	0.25	0.50	0.25	1

Find the expected number of heads.

4.13. Suppose that a certain region of a country the daily rainfall (in inches) is a continuous random variable X with probability density function $f(x)$ given by

$$f(x) = \frac{3}{4}(2x - x^2);\ 0 < x < 2.$$

Find the probability that, a given day in this region the rainfall is (a) not more than 1 inch, (b) more than 1.5 inches, (c) between 0.5 and 1.5 inches and (d) less than 1 inch. Also find the expected daily rainfall in that region.

Sample MCQs

1. The joint probability density function of X and Y is as follows: $f(x, y) = 2y$, $0 < x < 1$, $0 < y < 1$. Calculate $V(2X-Y)$.

a) 0.39 b) 0.31 c) 0.43 d) 0.16

2. Given, the probability function of X as follows:

X	0	1	2	3
$P(X)$	0.4	0.2	0.1	0.3

Calculate $P(1<X\leq 3)$.

a) 0.30 b) 0.50 c) 0.40 d) 0.90

3. Find the value of c if a random variable X has the probability density function given by $f(x) = c\,(2x - x^2)$; $0 < x < 2$.

a) $\frac{1}{4}$ b) $\frac{4}{3}$ c) $\frac{6}{10}$ d) $\frac{3}{4}$

4. The probability function of a continuous random variable X is given by, $f(x) = \frac{x}{8}$, $0 < x < 6$. Calculate $E(X+2)$.

a) 3 b) 11 c) 14 d) 9

5. The joint probability density function of two random variables X and Y is given by:
$f(x, y) = e^{x+y}$, $0 < x < 1$, $1 < y < 2$.
Are X and Y independent random variables?

a) Yes b) No c) both a and b d) None

6.Find variance of X. Let X be a continuous random variable with probability density function, $f(x) = 2x^{-2}$, $1 < x < 2$.

a) 0.3641 b) 0.7820 c) 0.0782 d) 0.0923