1 Random variables and distributions

In this chapter we consider real valued functions, called random variables, defined on the sample space.

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

The set of possible values of X is denoted by the set $\mathbb{R}_{\mathbb{X}}$. Random variables constitute the quantities of interest in most experiments.

If $\mathbb{R}_{\mathbb{X}}$ is countable, we say X is discrete. If on the other hand, $\mathbb{R}_{\mathbb{X}}$ is an interval or a union of intervals, we say X is continuous.

Example Toss a balanced coin 3 times. Our interest is the number of heads obtained and not necessarily on the order of the outcomes. Each point in the sample space is equally likely.

Sample space	X =Number of heads	Number of tails	Probability
(HHH)	3	0	$\frac{1}{8}$
(HHT)	2	1	$\frac{1}{8}$
(HTH)	2	1	$\frac{1}{8}$
(HTT)	1	2	$\frac{1}{8}$
(THH)	2	1	$\frac{1}{8}$
(THT)	1	2	$\frac{1}{8}$
(TTH)	1	2	$\frac{1}{8}$
(TTT)	0	3	$\frac{1}{8}$

Here, $\mathbb{R}_{\mathbb{X}} = \{0, 1, 2, 3\}$and hence X is discrete. The pmf is given by the table

x	0	1	2	3
f(x)	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Example Let X be the number of e-mails received during the time interval $(0, t)$ and let Y be the time between e-mails. The random variable X is discrete and takes values $0, 1, 2, \ldots.$ On the other hand, the random variable Y is continuous and takes values in an interval of length at most t.

Defiinitions

- The probability mass function or probability distribution of the discrete random variable X is the function $f(x) = P(X = x)$ for all possible values x. It satisfies the following properties

  i) $f(x) \geq 0$

  ii) $\sum f(x) = 1$

The cumulative distribution function $F(x)$ of a discrete random variable X is

$$F(x) = \sum_{t \leq x} f(t), -\infty < x < \infty$$

Associated with a random variable is the probability mass function which attaches to each value. We see for example that the value 2 is taken 3 times and the probability for each instance is $\frac{1}{8}$. Hence the probability of the random variable taken the value 3 is $3\left(\frac{1}{8}\right)$

Example1 Coin tossing

i) Probability mass function plot for coin tossing

ii) Probability histogram

iii) Discrete cumulative distribution function

$$F\left(x\right)=\begin{cases}0 & x<0\\ \frac{1}{8} & 0\leq x<1\\ \frac{4}{8} & 1\leq x<2\\ \frac{7}{8} & 2\leq x<3\\ 1 & 3\leq x\end{cases}$$

Example Find the constant c which makes the following a proper density

$$f(x) = cx, x = 1, 2, 3$$

We must have $1 = \sum f(x) = c \sum_{x=1}^{3} x = 6c$. Hence $c = \frac{1}{6}$

1.1 Continuous probability distributions

If a sample space contains an infinite number of possibilities equal to the number of points on a line segment, it is called a continuous sample space. For such a space, probabilities can no longer be assigned to individual values. Instead, probabilities must be assigned to sub-intervals

$$P(a < X < b)$$

Definition The function $f(x)$ is a probability density function for the continuous random variable X defined over the set of real numbers if

i) $f(x) \geq 0, all\ real\ x$

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

iii) $P(a < X < b) = \int_a^b f(t)\,dt$

iv) the cumulative distribution function

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

Example1 $f(x) = 1, 0 < x < 1$

Example2 $f(x) = e^{-x}, 0 < x < \infty$. $F(t) = \int_0^t e^{-x} dx = 1 - e^{-t}, t > 0$

Example Find the constant c which makes $f(x) = cx, 0 < x < 1$ a proper density.

We must have $1 = \int f(x)\, dx = c \int_0^1 x dx = c \frac{x^2}{2} |_0^1 = \frac{c}{2}$. Hence, $c = 2$.

We may calculate

$$P\left(\frac{1}{4} < X < \frac{3}{4}\right) = \int_{\frac{1}{4}}^{\frac{3}{4}} 2x dx = \frac{1}{2}$$

Example (Exercise 2.7) Let

$$f(x) = \begin{cases} x & 0 < x < 1 \\ 2 - x & 1 \leq x < 2 \\ 0 & elsewhere \end{cases}$$

a) $P(X < 1.2) = \int_0^{1.2} f(x)\, dx = \int_0^1 x dx + \int_1^{1.2} (2 - x)\, dx = 0.68$

b) $P(0.5 < X < 1) = \int_{0.5}^1 x dx = \frac{3}{8}$

1.2 Joint probability distributions

We have situations where more than characteristic is of interest. Suppose we toss a pair of dice once. The discrete sample space consists of the pairs

$$\{(x, y) : x = 1, ..., 6; y = 1, ..., 6\}$$

where X, Y are the random variables representing the results of the first and second die respectively. For two electric components in a series connection, we may be interested in the lifetimes of each.

Definition The function $f\left(x,y\right)$ is a joint probability distribution or probability mass function o$\vartheta f$the discrete random variables X,Y if

i) $f\left(x,y\right)\geq 0, all\ \ \left(x,y\right)$

ii) $\sum_{x}\sum_{y}f\left(x,y\right)=1$

iii) $P\left(X=x,Y=y\right)=f\left(x,y\right)$

iv) $P\left(\left(X,Y\right)\epsilon A\right)=\sum\sum_{\left(x,y\right)\epsilon A}f\left(x,y\right)$

Example One toss of a pair of dice: $P\left(X = x, Y = y\right) = \frac{1}{36}$

				Y			
		1	2	3	4	5	6
	1	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
	2	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
X	3	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
	4	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
	5	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
	6	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$

Definition The function $f(x,y)$ is a joint probability distribution of the continuous random variables X, Y if

i) $f(x,y) \geq 0, all \ \ (x,y)$

ii) $\int \int f(x,y)\, dxdy = 1$

iii) $P((X,Y)\, \epsilon A) = \int \int_A f(x,y)\, dxdy$

Example1 Find the constant c which makes $f(x,y) = cxy^2, 0 < x < 1, 0 < y < 1$ a proper density

We must have $1 = \int\int f(x,y)\, dxdy = c\int_0^1 \int_0^1 xy^2 dxdy = c\frac{x^2}{2}|_0^1 \frac{y^3}{3}|_0^1 = \frac{c}{6}$

Hence, $c = 6$

Example1 (continued) We calculate

$$P\left(0 < X < \frac{1}{2}, 0 < Y < \frac{1}{2}\right) = 6\int_0^{\frac{1}{2}}\int_0^{\frac{1}{2}} xy^2 dxdy = \frac{1}{4}$$

Example1 (continued) Calculate

$$\begin{aligned} P\left(0 < X < \frac{1}{2} | 0 < Y < \frac{1}{2}\right) &= \frac{P\left(0 < X < \frac{1}{2}, 0 < Y < \frac{1}{2}\right)}{P\left(0 < Y < \frac{1}{2}\right)} \\ &= \frac{\frac{1}{4}}{1} = 0.25 \end{aligned}$$

It is possible to retrieve the individual distribution of a random variable from the joint distribution. This is done through the notion of marginal distributions.

Definition The marginal distributions of X alone and of Y alone are

$$g(x) = \sum_y f(x,y), h(y) = \sum_x f(x,y)$$

for the discrete case and

$$g(x) = \int f(x,y)\,dy, h(y) = \int f(x,y)\,dx$$

for the continuous case.

Example1 Suppose that the joint density of X, Y is given as $f(x,y) = 8xy, 0 < x < y < 1$. Find the marginal densities of X and Y. We calculate

$$g_X(x) = \int_x^1 8xydy = 4x\left(1 - x^2\right), 0 < x < 1$$

$$g_Y(y) = \int_0^y 8xydx = 4y^3, 0 < y < 1$$

Example2 Suppose we draw two balls from an urn containing 5 white, 2 blue and 4 red balls. Let X and Y be the number of white and red balls drawn respectively. The joint density of X, Y is given by

$$f(x,y) = \frac{\binom{5}{x}\binom{4}{y}\binom{2}{2-x-y}}{\binom{11}{2}}, x + y \leq 2$$

		y			$g_X(x)$
		0	1	2	
x	0	$\frac{1}{55}$	$\frac{8}{55}$	$\frac{6}{55}$	$\frac{15}{55}$
	1	$\frac{10}{55}$	$\frac{20}{55}$	0	$\frac{30}{55}$
	2	$\frac{10}{55}$	0	0	$\frac{10}{55}$
$g_Y(y)$		$\frac{21}{55}$	$\frac{28}{55}$	$\frac{6}{55}$	1

The marginal densities are expressed in the "margins".

1.3 Conditional distributions

Definition The conditional distribution of the random variable Y given $X = x$ is

$$f(y|x) = \frac{f(x,y)}{g(x)}, g(x) > 0$$

Similarly the conditional distribution of X given $Y = y$

$$f(x|y) = \frac{f(x,y)}{h(y)}, h(y) > 0$$

Example Suppose that the joint density of X, Y is given as $f(x,y) = 8xy, 0 < x < y < 1$.

$$\begin{aligned} f(y|x) &= \frac{8xy}{4x(1-x^2)}, 0 < x < y < 1 \\ &= \frac{2y}{1-x^2}, 0 < x < y < 1 \end{aligned}$$

We may easily verify that $f(y|x)$ is a proper density since $\int_x^1 f(y|x)\,dy = 1$

Example2 (continued) We can calculate

$$f(y|x) = \begin{cases} \frac{1}{15} & y = 0 \\ \frac{8}{15} & y = 1 \\ \frac{6}{15} & y = 2 \end{cases}$$

The variables are said to be independent if

$$f(x,y) = g(x)\,h(y)\,,all\ \ (x,y)$$

For random variables $X_1, ..., X_n$ with joint density $f(x_1, ..., x_n)$ and marginals $f_1(x_1), ..., f_n(x_n)$ we say they are mutually independent if and only if

$$f(x_1, ..., x_n) = f_1(x_1)\,...f_n(x_n)$$

for all tuples $(x_1, ..., x_n)$ within their range.

Examples In the urn example, X and Y are not independent. Similarly for the continuous density example.

1.4 Properties of random variables

Definition The mean or expected value of a function L of a random variable X , say $L(X)$ is

$$\mu = E[L(X)] = \begin{cases} \sum L(x) f(x) & X \ discrete \\ \int L(x) f(x) \, dx & X \ continuous \end{cases}$$

Definition The mean or expected value of a function L of random variables X, Y , say $L(X, Y)$ is

$$\mu = E[L(X, Y)] = \begin{cases} \sum L(x, y) f(x, y) & X \ Y, discrete \\ \int L(x, y) f(x, y) \, dxdy & X \ Y, continuous \end{cases}$$

Example (Urn example) $E(X) = 0\left(\frac{15}{55}\right) + 1\left(\frac{30}{55}\right) + 2\left(\frac{10}{55}\right) = \frac{50}{55} = 0.909.$

$E(Y) = 0\left(\frac{21}{55}\right) + 1\left(\frac{28}{55}\right) + 2\left(\frac{6}{55}\right) = \frac{40}{55} = 0.727.$

$E\left(X^2\right) = 0\left(\frac{15}{55}\right) + 1^2\left(\frac{30}{55}\right) + 2^2\left(\frac{10}{55}\right) = \frac{70}{55} = 1.272.$

$E\left(Y^2\right) = 0\left(\frac{15}{55}\right) + 1^2\left(\frac{28}{55}\right) + 2^2\left(\frac{6}{55}\right) = \frac{52}{55} = 0.945.$

$E(XY) = 1(1)\left(\frac{20}{55}\right) = \frac{20}{55}$

Definition The covariance of two random variables X, Y is $\sigma_{XY} = E\left[\left(X - \mu_X\right)\left(Y - \mu_Y\right)\right] = E\left(XY\right) - \mu_X \mu_Y$.

Definition The variance of a random variable X is $\sigma_X^2 = E\left(X - \mu_X\right)^2 = E\left(X\right)^2 - \mu_X^2$

Definition The correlation between X, Y is $\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$

The variance measures the amount of variation in a random variable taking into account the weighting given to the values of the random variable. The correlation measures the "interaction" between the variables. A positive correlation indicates that an increase in X results in an increase in Y. A negative correlation indicates that an increase in$X$ results in a decrease in Y.

Example (urn example) $\sigma_X^2 = \frac{70}{55} - \left(\frac{50}{55}\right)^2 = \frac{54}{121} = 0.446$,$\sigma_{XY} = \frac{20}{55} - \left(\frac{50}{55}\right)\left(\frac{40}{55}\right) = -\frac{36}{121} = -0.2975$ $\rho = - - .690$

Example Suppose that the joint density of X, Y is given as $f(x, y) = 8xy, 0 < x < y < 1$. Then

$$
\begin{aligned}
\mu_X &= \frac{8}{15} \\
\mu_Y &= \frac{4}{5} \\
\sigma_X^2 &= \frac{11}{225} \\
\sigma_Y^2 &= \frac{2}{75} \\
\sigma_{XY} &= \frac{4}{225} \\
\rho &= 0.49237
\end{aligned}
$$

In this section, we assemble several results on the calculation of expectation for linear combinations of random variables

Theorem Let X, Y be two random variables and let a, b, c be three arbitrary constants. Also, let $h_1(x), h_2(x), h_3(x, y), h_4(x, y)$ be real valued functions. Then,

$$
\begin{aligned}
E(aX + b) &= aE(X) + b \\
E[h_1(X) \pm h_2(X)] &= E[h_1(X)] \pm E[h_2(X)] \\
E[h_3(X, Y) \pm h_4(X, Y)] &= E[h_3(X, Y)] \pm E[h_4(X, Y)] \\
\sigma^2_{aX+bY+c} &= a^2\sigma^2_X + b^2\sigma^2_Y + 2ab\sigma_{XY}
\end{aligned}
$$

Theorem Let X, Y be two independent random variables and let a, b be two arbitrary constants. Then

$$\begin{aligned} E(XY) &= E(X)\,E(Y) \\ \sigma_{XY} &= 0 \\ \sigma^2_{aX+bY} &= a^2\sigma^2_X + b^2\sigma^2_Y \end{aligned}$$

This theorem generalizes to several independent random variables $X_1, ..., X_n$

Theorem Let $X_1, ..., X_n$ be random variables and let $a_1, ..., a_n$ be arbitrary constants. Then

i) $E\left[\sum a_i X_i\right] = \sum a_i E\left[X_i\right]$

ii)If in addition, $X_1, ..., X_n$ are independent random variables

$$\sigma^2_{\sum a_i X_i} = \sum a_i^2 \sigma^2_{X_i}$$

Example Flip a coin with probability of heads p, n times and let $X_i = 1$ if we observe heads on the i^{th} toss and $X_i = 0$, if we observe tails. Assume the results of the tosses are independent and performed under indentical conditions. Then, $X_1, ..., X_n$ are independent and identically distributed random variables. The sum, $\sum X_i$ represents the number of heads in n tosses. It follows from the theorem

i) $E\left[\frac{1}{n}\sum X_i\right] = \sum \frac{1}{n} E\left[X_i\right] = \frac{1}{n}(np) = p$

ii) $\sigma^2_{\frac{1}{n}\sum X_i} = \sum \left(\frac{1}{n}\right)^2 \sigma^2_{X_i} = \frac{\sigma^2}{n} = \frac{p(1-p)}{n}$

The example demonstrates that the average of a set of independent random variables preserves the mean and importantly reduces the variance by a factor equal to the size of the sample. .