

Session 1

PROBABILITY MASS FUNCTION

Contents

Introduction

1.1 Definition

1.2 Properties of pmf

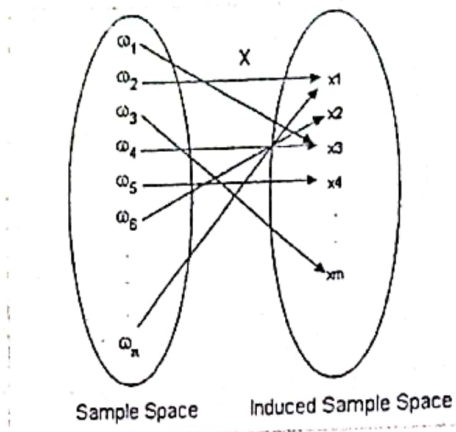
Summary

Learning outcomes

Introduction

The probability is the measure of the likelihood of the occurrence of an event. Mathematically, any event, say E , is a subset of a sample space S of some experiment. The probability of an event is denoted by $\Pr(E)$. Mathematically, $\Pr$ is a set function, defined from the set of all events of S onto the interval $[0,1]$. It is called the probability function. It satisfies the axioms of probability (or the Kolmogorov Axioms).

Some sample spaces contain quantitative (numerical) outcomes; others contain qualitative outcomes. Often it is convenient to work with sample spaces containing numerical outcomes. Hence, we map the original sample space into a new sample space containing numerical values. This is more useful when the original sample space contains qualitative outcomes. A function that maps the original sample space into a set of real numbers is called a random variable. The new sample space thus created is called the induced sample space or the range of the random variable. It is customary to denote random variables by upper case English letters such as X , Y , Z .



Let S be a sample space, X be a random variable, and S_X be the induced sample space. The total probability of the sample space S is 1. Hence, the total probability associated with the induced sample space S_X is also 1. Usually, it is possible to describe how the total probability associated with the induced sample space is distributed within its subsets, using a mathematical function. Such a function is called a probability distribution of the random variable.

There are two types of probability distributions, depending on the type of the variable; the probability mass function for a discrete random variable and the probability density function for a continuous random variable. In this session we discuss about the probability mass function.

1.1 Definition

Let X be a discrete random variable. Then, the induced sample space (range) of X is countable. The probability distribution of X that specifies the probability associated with each individual value of the induced sample space is called the probability mass function (pmf) of X . It is denoted by P_X .

Definition 1.1. The probability distribution or probability mass function (pmf) of a discrete random variable X is the function P_X defined as

$$P_X(x) = \Pr(X = x), x \in \mathbb{R}.$$

Note: In the above definition, ' $X = x$ ' is an abbreviation for the set $\{\omega \in S | X(\omega) = x\}$. It is an event. Similarly, we use abbreviations ' $X \leq x$ ', ' $X \geq x$ ' etc. for the events $\{\omega \in S | X(\omega) \leq x\}$, $\{\omega \in S | X(\omega) \geq x\}$ etc. respectively.

1.2 Properties of a pmf

The function P_X has the following properties:

1. For any $x \in \mathbb{R}$, $0 \leq P_X(x) \leq 1$.

2. Let E be an event and $\mathcal{E} = \{X(\omega) : \omega \in E\}$ be the corresponding induced event. Then,
 $\Pr(E) = \sum_{x \in \mathcal{E}} P_X(x)$.

3. Let S be the sample space and $S_X = \{X(\omega) : \omega \in S\}$ be the induced sample space. Then,
 $\sum_{x \in S_X} P_X(x) = 1$.

Example 1.1.

A curd vendor has observed that 70% of the customers buy 1 container, 15% of the customers buy 2 containers, 10% of the customers buy 3 containers and 5% of the customers buy 4 containers.

- (i) Let X be the number of containers bought by a randomly chosen customer. What is the probability mass function of X ?
- (ii) Let N be the number of customers up to the first customer to buy 4 containers on a given day. What is the probability mass function of N ?

Solution 1.1

- (i) The probability mass function of X can be written as

x	1	2	3	4
$P_X(x)$	0.70	0.15	0.10	0.05

or

$$P_X(x) = \begin{cases} 0.70 & \text{if } x = 1 \\ 0.15 & \text{if } x = 2 \\ 0.10 & \text{if } x = 3 \\ 0.05 & \text{if } x = 4 \end{cases}$$

Note: It is implied that $P_X(x) = 0$ if $x \neq 0, 1, 2, 3, 4$.

- (ii) Let X_i be the number of containers bought by the i th customer ($i = 1, 2, 3, \dots$). Let n be any integer $1, 2, \dots$. Then, the event $N = n$ is equivalent to the event $(X_1 \leq 3) \cap (X_2 \leq 3) \cap \dots \cap (X_{n-1} \leq 3) \cap (X_n = 4)$. Hence,

$$\begin{aligned} P_N(n) &= \Pr(N = n) \\ &= \Pr[(X_1 \leq 3) \cap (X_2 \leq 3) \cap \dots \cap (X_{n-1} \leq 3) \cap (X_n = 4)] \end{aligned}$$

Since the events $(X_1 \leq 3), (X_2 \leq 3), \dots, (X_{n-1} \leq 3), (X_n = 4)$ are independent,

$$P_N(n) = \Pr(X_1 \leq 3) \Pr(X_2 \leq 3) \dots \Pr(X_{n-1} \leq 3) \Pr(X_n = 4)$$

All X_i 's have the same distribution as that of X . Hence for any $i = 1, 2, \dots$,

$$\begin{aligned}\Pr(X_i \leq 3) &= \Pr(X \leq 3) \\ &= \Pr(X = 1) + \Pr(X = 2) + \Pr(X = 3) \\ &= P_X(1) + P_X(2) + P_X(3) \\ &= 0.70 + 0.15 + 0.10 \\ &= 0.95\end{aligned}$$

or

$$\begin{aligned}\Pr(X_i \leq 3) &= \Pr(X \leq 3) \\ &= 1 - \Pr(X = 4) \\ &= 1 - P_X(4) \\ &= 1 - 0.05 \\ &= 0.95\end{aligned}$$

So, the pmf of N is

$$P_N(n) = (0.95)^{n-1} 0.05, n = 1, 2, \dots$$

Activity 1.1.

Consider Example 1.1.

- (i) Show that $P_N(n)$ is a pmf.
- (ii) What is the probability that the first customer to buy 4 containers on a given day is within the first 3 customers of the day?

Activity 1.2.

A novice player is playing the game of darts using the dart board shown in figure 1.1. The circles have radii 0.5, 1, 1.5 and 2 units respectively. It has numbers from 1 to 4. The player receives a score corresponding to the number assigned to the region in which the dart lands. Let X be the score of the player on a randomly selected throw. Assuming that the player hits on the board with probability 1, find the pmf of X .

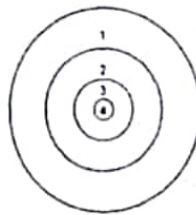


Figure 1.1: Dart board

Example 1.2.

A production process, when in-control, produces 95% products that conform to the standards.

In order to check whether the process is in-control, a sample of 25 products is examined every hour. Let X be the number of defective products in a sample. Find the probability mass function of X , when the process is in control.

Solution 1.2

Notice that $X = 0$ if and only if all the products in the sample are non-defective. When the process is in-control, the probability that any randomly selected product is non-defective is 0.95. Therefore the probability that all the products in the sample are non-defective is $(0.95)^{25} = 0.2774$.

$$P_X(0) = \Pr(X = 0) = 0.2774.$$

Next notice that $X = 1$ if and only if the sample contains exactly 1 defective product. There are $\binom{25}{1} = 25$ mutually exclusive ways of occurring this. For example, the first item is defective and all others are non-defective, or the second item is defective and all others are non-defective, and so on. The probability of any one of these ways is $(0.05)(0.95)^{24} = 0.0146$. Hence, $P_X(1) = \Pr(X = 1) = 25(0.0146) = 0.365$.

Likewise, $X = 2$ if and only if the sample contains exactly 2 defective products. There are $\binom{25}{2} = 300$ mutually exclusive ways of occurring this. For example, the first and the second items are defective and all others are non-defective, or the first and the third items are defective and all others are non-defective, and so on. The probability of any one of these ways is $(0.05)^2(0.95)^{23} = 0.0008$. Hence, $P_X(2) = \Pr(X = 2) = 300(0.0008) = 0.24$.

Similarly, $X = x$ if and only if the sample contains exactly x ($x = 0, 1, 2, \dots, 25$) defective products. There are $\binom{25}{x}$ mutually exclusive ways of occurring this. The probability of any one of these ways is $(0.05)^x(0.95)^{25-x}$. Hence,

$$P_X(x) = \Pr(X = x) = \binom{25}{x}(0.05)^x(0.95)^{25-x}, x = 0, 1, 2, \dots, 25.$$

Activity 1.3.

Consider Example 1.2. If the sample contains 0 or 1 defective products, it is decided that the process is in-control; otherwise it is decided that the process is out-of-control. What is the probability of deciding that the process is out-of-control when actually the process is in-control, that is, when the actual proportion of conforming products is 0.95?

Example 1.3.

In a box there are k balls of which m are black. A random sample of n balls are taken from the box. Let X be the number of black balls in the sample. Supposing $n \leq m$, what is the probability mass function of X ?

Solution 1.3

Using the theory of classical probability it is easy to see that

$$P_X(x) = \frac{\binom{m}{x}\binom{k-m}{n-x}}{\binom{k}{n}}, \quad x = 0, 1, \dots, n$$

Activity 1.4.

Five animals from a population of animals in a certain region were caught, tagged, and released back into the population. After they have had enough time to mix in, 10 animals were observed.

There were 2 animals with tags. What is the probability of such an event if there were n animals in the population?

Activity 1.5.

Consider Example 2.1. Calculate the above probability for $n=10, 11, 12, \dots$. Plot the probability versus n . For what value of n is the above event most likely?

There are a number of commonly used models for probability mass functions. Binomial distribution, Geometric distribution, Negative binomial distribution, Poisson distribution are some examples. You shall learn these models and their properties in the subsequent sessions.

Summary

In this session, we discussed about a function that describes the distribution of probability related to a discrete random variable. It is called the probability mass function. It directly gives the probability of each possible value of the random variable.

Learning outcomes

By now you should be able to

- Describe what a pmf is.
- Determine whether a given function is a pmf or not.
- Derive the pmf for relatively simple random variables.
- Calculate the probability of an event using a pmf.

Session 2

PROBABILITY DENSITY FUNCTION

Contents

Introduction

2.1 Definition

2.2 Existence of pdf

2.3 Calculation of probability using pdf

Summary

Learning outcomes

Introduction

Let X be a continuous random variable. Then, the induced sample space as well as any induced event is uncountable. Hence, the expressions $\Pr(E) = \sum_{x \in \mathcal{E}} P_X(x)$ and $\sum_{x \in \mathcal{S}_X} P_X(x) = 1$ are meaningless because the summation cannot be taken over an uncountable set. This shows that it is not possible to define a pmf P_X for a continuous random variable. Instead, we define a probability density function. In this session, we discuss about probability density function for a continuous random variable.

2.1 Definition

The probability density function f_X of a continuous random variable X is a function with the following properties.

1. For any $x \in \mathbb{R}$, $f_X(x) \geq 0$.
2. If E is an event and $\mathcal{E} = \{X(\omega) : \omega \in E\}$, then, $\Pr(E) = \int_{\mathcal{E}} f_X(x) dx$.
3. If $S_X = \{X(\omega) : \omega \in S\}$, then $\int_{S_X} f_X(x) = 1$.

2.2 Existence of pdf

To see the existence of such a function, consider a continuous random variable X and suppose that we have a very large number N of observations of X , measured with high accuracy (correct to a large number of decimal places). Suppose that the following grouped frequency table and the histogram in figure 2.1(a) have been constructed from those data. The height of the bar on a class interval of this histogram is equal to the relative frequency per unit length in that class interval.

interval	class boundaries	class frequency	height of the bar	area of the bar
I_1	$x_1 - \delta x/2, x_1 + \delta x/2$	n_1	$\frac{n_1}{\delta x \cdot N}$	$\frac{n_1}{N}$
I_2	$x_2 - \delta x/2, x_2 + \delta x/2$	n_2	$\frac{n_2}{\delta x \cdot N}$	$\frac{n_2}{N}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
I_k	$x_k - \delta x/2, x_k + \delta x/2$	n_k	$\frac{n_k}{\delta x \cdot N}$	$\frac{n_k}{N}$
	Total	N		1

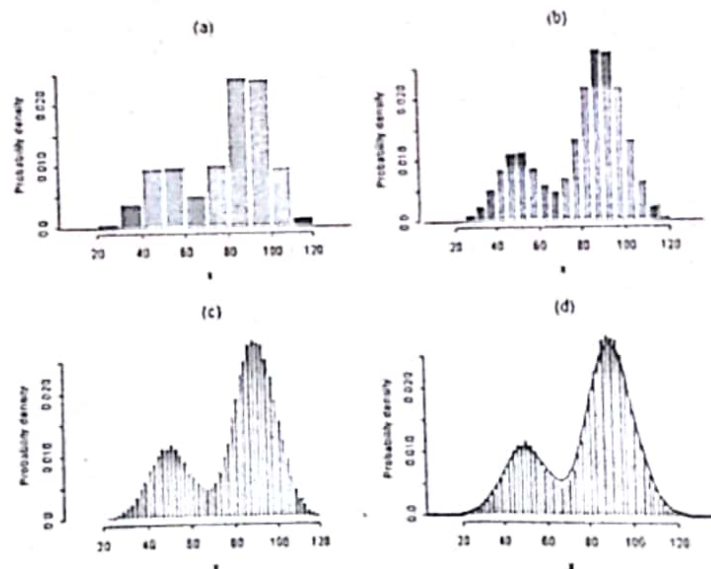


Figure 2.1:

Then, for the i th interval,

$$\Pr(x_i - \delta x/2 \leq X \leq x_i + \delta x/2) \approx \text{Area of the bar}$$

and therefore

$$\text{Height of the bar} = \frac{\text{Area of the bar}}{\delta x} \approx \frac{\Pr(x_i - \delta x/2 \leq X \leq x_i + \delta x/2)}{\delta x}$$

So, it is clear that the height of a bar represents the 'probability density' in that class interval.

What will happen if we let $\delta x \rightarrow 0$? (number of class intervals are increased). Then we may be able to approximate the histogram by a smooth curve like in figure 2.1(d). Because the area under each histogram is 1, the area under the curve is also 1. For any point x ,

$$\text{the height of the curve} \approx \lim_{\delta x \rightarrow 0} \frac{\Pr(x - \delta x/2 \leq X \leq x + \delta x/2)}{\delta x}$$

will represent the 'probability density at point x '. Let the above smooth curve be denoted by f_X . Then, f_X has the properties 1 to 3 described in section 2.1. This function is called the 'probability density function (pdf) of X '.

It should be clearly understood that $f_X(x)$ is the 'probability density at point x '; Not the 'probability at point x '.

2.3 Calculation of probability using pdf

Let $c, d \in \mathbb{R}$ such that $c \leq d$. Then,

$$\Pr(c \leq X \leq d) = \int_c^d f_X(x) dx$$

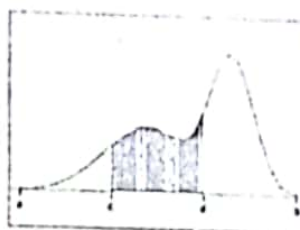


Figure 2.2: $\Pr(c \leq X \leq d) = \int_c^d f_X(x) dx$

Note: If X is a continuous random variable with the p.d.f. f_X , then for any $k \in \mathbb{R}$,

$$\Pr(X = k) = \Pr(k \leq X \leq k) = \int_k^k f_X(x) dx = 0.$$

Therefore, for a continuous random variable X ,

$$\Pr(c < X < d) = \Pr(c \leq X < d) = \Pr(c < X \leq d) = \Pr(c \leq X \leq d) = \int_c^d f_X(x) dx.$$

There are a large number of commonly used mathematical functions that are used as models for probability density functions. Uniform distribution, normal distribution gamma distribution, exponential distribution are few examples. A suitable model for a given random variable can be chosen based on the shape of the histogram. Sometimes, a suitable model can be chosen based on prior knowledge, or literature on similar situations. You shall learn about few commonly used pdf models in the subsequent sessions. Here we consider two examples.

Example 2.1.

Table 2.1 gives the writing times of 40 randomly selected ballpoint pens of a certain brand. Let X be the writing time of a randomly selected pen. Suggest a suitable model for the pdf of X .

31.7	26.2	35.0	34.8	33.1	30.8	34.1	34.4	31.4	29.5
29.0	29.7	32.5	25.3	30.6	37.7	28.7	23.1	25.7	26.7
31.0	28.7	30.7	36.4	27.3	32.1	29.2	30.1	26.3	25.9
27.5	28.6	37.2	31.3	22.7	27.4	31.7	30.0	33.6	29.6

Table 2.1: writing time

Figure 2.3 shows the histogram of the above data.

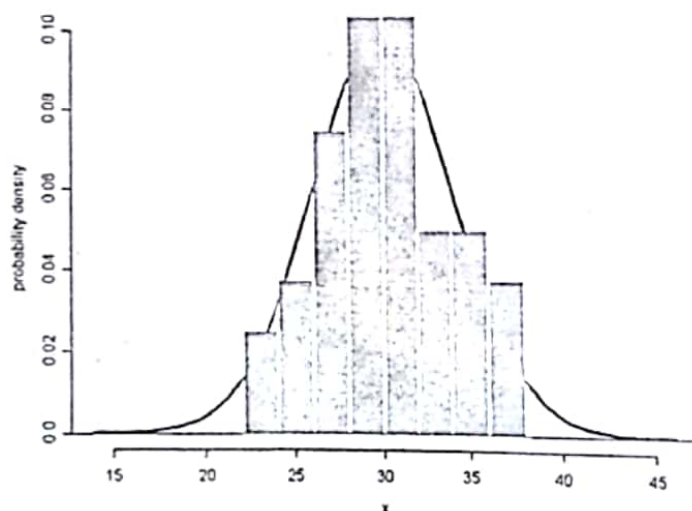


Figure 2.3: Histogram of writing time data and a possible model for the pdf

Histogram is roughly symmetric and bell-shaped. Therefore, a bell shaped symmetric curve is suitable as the model for pdf of X . There are a number of mathematical functions that has the symmetric bell shape. One such function is given by

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2 \right\}, -\infty < x < \infty$$

This is called the normal distribution. It contains two parameters μ and σ . In order to determine suitable values for these parameters, we use the well known mathematical properties of this function. This function is symmetric around the value μ and $f(x) \approx 0$ when $x < \mu - 3\sigma$ and $x > \mu + 3\sigma$. Matching these properties with the above histogram and curve, we choose $\mu = 30$ and $\sigma \approx (42 - 18)/6 = 4$. Hence, we may take

$$f_X(x) = \frac{1}{4\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left(\frac{x - 30}{4} \right)^2 \right\}, -\infty < x < \infty$$

as the pdf of X . This curve is shown in Figure 2.3. It seems a fairly good approximation.

Now, suppose we want to find the proportion of pens that can write more than 25 hours. In other words, we need $\Pr(X > 25)$. Using our model, this probability can be calculated as

$$\Pr(X > 25) = \int_{25}^{\infty} \frac{1}{4\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left(\frac{x - 30}{4} \right)^2 \right\} dx$$

This calculation is somewhat difficult. There is an easier method to calculate such probabilities of normal distributions. We shall discuss this method later.

Example 2.2.

The following data are the times (in days) between accidents occurred at a certain factory. Figure 2.4 shows the histogram of these data. Fit a suitable pdf.

Suppose there has been an accident today. Using your model, calculate the probability that there will be no accidents within the next 30 days.

26 47 168 63 20 112 2 25 51 19 145 23 4 2 0 76 80 12 97 68 2 21 34 181 89
9 9 7 3 11 104 11 40 1 58 10 3 22 31 49 28 49 15 19 38 89 41 7 12 1

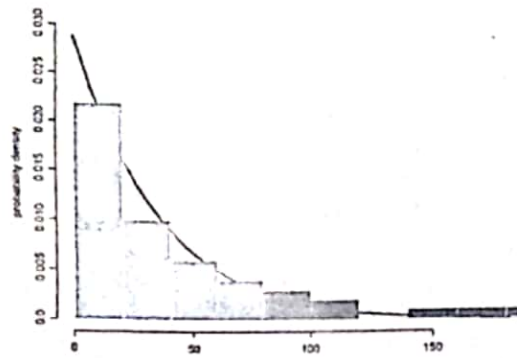


Figure 2.4: Histogram of times between accidents

According to the histogram, the distribution is positively skewed and has a positive density when $x = 0$. A model that is commonly used in such situations is

$$f(x) = \lambda e^{-\lambda x}, x > 0, \lambda > 0$$

This is called the exponential distribution with parameter λ . To be able to apply the model, we should know the values of the parameter. There are several methods to estimate λ . We shall employ a simple method here.

The sample median of the data is 24. Hence, it is reasonable to expect that $\Pr(X \leq 24) \approx 0.5$. This implies

$$\begin{aligned} \Pr(X \leq 24) &= \int_0^{24} \lambda e^{-\lambda x} \approx 0.5 \\ \lambda &\approx -\ln(0.5)/24 \approx 0.03 \end{aligned}$$

The density function obtained with $\lambda = 0.03$ is shown in Figure 2.4. It seems to approximate the shape of the histogram well. Hence, we may take

$$f_X(x) = 0.03e^{-0.03x}, x > 0$$

as the pdf of X .

There will be no accident within the next 30 days if and only if the time up to the next accident is greater than 30 days. Therefore, the required probability is

$$\begin{aligned} \Pr(X > 30) &= \int_{30}^{\infty} 0.03e^{-0.03x} dx \\ &= 0.4066 \end{aligned}$$

Activity 2.1.

For coming to work, I must take two buses. From the past experience, I know that a bus can

come at any time within 6 minutes. Also, the probability that a bus comes within any period of the same length is the same. Hence, it is reasonable to assume the following probability density function for the waiting time W (in minutes) for a bus.

$$f_W(w) = 1/6, 0 < w < 6.$$

Using a method that you will learn later, it can be shown that the total waiting time T at the two bus-stops has the density function

$$f_T(t) = \begin{cases} \frac{t}{36} & 0 \leq t \leq 6 \\ \frac{1}{3} - \frac{t}{36} & 6 \leq t \leq 12. \end{cases}$$

- (i) What is the probability that the waiting time at the first bus-stop will be less than 2 minutes?
- (ii) What is the probability that the waiting time at the first bus-stop will be less than 2 minutes and the waiting time at the second bus-stop will be greater than 2.5 minutes?
- (iii) What is the probability that the total waiting time will exceed 5 minutes?

Summary

In this session, we discussed about a function that describes the distribution of probability related to a continuous random variable. It is called the probability density function (pdf). The shape of the pdf is approximately equal to the smoothed histogram of a large data set. It provides the probability density at each value; not the probability. The probability in any interval is given by the integration of the pdf over the interval.

Learning outcomes

By now you should be able to

- Describe what a pdf is.
- Explain the difference between the functions $\Pr$, P_X , and f_X .
- Check whether a given function is a probability density function or not.
- Calculate probability when a pdf is given.

Session 3

DESCRIPTIVE PROPERTIES OF DISTRIBUTIONS

Contents

Introduction

3.1 Mean of a random variable

3.2 Variance and standard deviation of a random variable.

Summary

Learning outcomes

Introduction

We can describe a sample of data using few descriptive statistics. For example, the sample mean and sample median describe the location or the average of the data set, while the sample standard deviation and inter-quartile range describe the dispersion of the data set. Note that these measures vary from sample to sample from the same population.

Similarly, we can fairly well describe a population of values of a random variable X (simply called the population), using few parameters like the population mean, and the population standard deviation. Unlike the sample mean and the sample standard deviation, these values are fixed and usually unknown. The population of interest may be an existing one or an imaginary one. Sometimes, the variable of interest X is some characteristic of the existing units. Then, the population is the set of values that we would get if we measure X of all the units. For example, suppose X is the monthly mobile-phone bill of a randomly selected student in your university. Then the population of X is the set of all values that we would get if we

ask the monthly mobile-phone bills from all the students in the university. Note that this is an existing population. We may want to find the mean and the standard deviation of this population. Sometimes, the population of interest can be the population of values that will be generated if we observe the random variable repeatedly. For example, suppose X is the time between accidents in a workplace. Then, the population of X is the set of all the values we will get if repeatedly observe the times between accidents, infinitely many times. In this case, the population of interest is imaginary. However, we may be interested in finding the mean and standard deviation of this population.

We can calculate the summary measures for both of these populations without observing the values of the whole population. All we need to know is the probability distribution of X .

Suppose we know the distribution of X . In this session, you learn how to calculate the mean and the standard deviation of a population of values of a random variable X . We simply call them the mean and the standard deviation of X .

3.1 Mean of a random variable

It is easy to understand the definition of the mean of a random variable by looking at the definition of the mean of a data set. We calculate the mean of a data set by dividing the sum of the values by the number of values (or the sample size) in the sample.

Now, suppose X is a discrete random variable. If X is related to an existing population, think of all the values of X for the units of the existing population. If X is not related to an existing population, think of the population of values that will be generated if we repeatedly observe X . In any case, the proportion of each possible value will be equal to the probability assigned to that value by the pmf P_X . In case of an existing population, suppose the population size is N . Otherwise, think of a very large number N of values that will be generated by repeatedly observing X .

Suppose that the different possible values of X are $x_1, x_2, x_3, \dots$. Then, out of the N observations, approximately $NP_X(x_1)$ number of values will be x_1 , $NP_X(x_2)$ number of values will be x_2 and so on. Therefore the population mean is $(x_1NP_X(x_1) + x_2NP_X(x_2) + \dots)/N = x_1P_X(x_1) + x_2P_X(x_2) + \dots$. Hence, the mean of a discrete random variable X (or the mean of the population of values of X) is defined as follows.

Definition 3.1.

Let X be a discrete random variable with pmf P_X . The mean (expected value) of X is defined as

$$E(X) = \sum_x xP_X(x)$$

Example 3.1.

Consider Example 1.1. Calculate the mean number of containers of curd bought by a randomly selected customer.

Solution 3.1

$$E(X) = 1(0.7) + 2(0.15) + 3(0.10) + 4(0.05) = 1.5$$

This is the mean of the population of values that would be generated if we observe X infinitely many times. It can be interpreted as the number of containers that a customer buys on average.

Example 3.2.

Consider Example 1.1. Calculate $E(N)$.

Solution 3.2

$$\begin{aligned} E(N) &= 1(0.95)^0(0.05) + 2(0.95)^1(0.05) + 3(0.95)^2(0.05) + 4(0.95)^3(0.05) + \dots \\ &= 0.05[1 + 2(0.95) + 3(0.95)^2 + 4(0.95)^3 + \dots] \\ &= 20 \end{aligned}$$

This means, on average, the number of customers up to the first one to buy 4 containers is 20. To understand this suppose you count the number of customers up to the first customer to buy 4 containers. Once a customer buys 4 containers, you re-start the experiment again and from the next customer, count the number of customers up to the first customer to buy 4 containers. In other words, assume that you observe Y repeatedly an infinitely many times. Then, the mean of all such observed values will be 20. So, here, we find the mean of the population of values that would be created if we observe Y infinitely many times.

In general terms, this can be interpreted as every 20th customer, on average, buys 4 containers. This answer is intuitively meaningful as well. According to the pmf, 5% of the customers buy 4 containers. That means, $1/20$ customers buy 4 containers. Therefore we can expect that every 20th customer, on average, would buy 4 containers.

Activity 3.1.

Consider Activity 1.2. What is the mean score that can be expected from the player? Interpret the value that you obtain.

Definition 3.2.

Let X be a discrete random variable with pmf P_X . The mean (expected value) of any function $g(X)$ is defined as

$$E[g(X)] = \sum_x g(x)P_X(x)$$

Example 3.3.

Consider Example 1.1. Calculate $E(X^2)$.

Solution 3.3

Here, $g(X) = X^2$. Then,

$$E(X^2) = 1^2(0.7) + 2^2(0.15) + 3^2(0.10) + 4^2(0.05) = 3$$

Even though $E(X^2)$ is not a directly useful quantity, as you will see later, it is required for the calculation of the variance.

Definition 3.3 (Mean of a Continuous Random Variable).

Let X be a continuous random variable with pdf $f_X(x)$. The mean (expected value) of X is defined as

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

Let $g(X)$ be any function of X . Then, the mean of $g(X)$ is given by

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx$$

Example 3.4.

Let X be a random variable with pdf

$$f_X(x) = \frac{3}{4}(1 - x^2), \quad -1 \leq x \leq 1$$

Find $E(X)$ and $E(X^2)$.

Solution 3.4

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f_X(x) dx \\ &= \int_{-\infty}^{-1} x(0) dx + \int_{-1}^1 x \left[\frac{3}{4}(1 - x^2) \right] dx + \int_1^{\infty} x(0) dx \\ &= 0 \end{aligned}$$

Similarly,

$$\begin{aligned} E(X^2) &= \int_{-1}^1 x^2 \left[\frac{3}{4}(1 - x^2) \right] dx \\ &= 0.2 \end{aligned}$$

Activity 3.2.

Consider Activity 2.1. Calculate the expected total waiting time on a randomly selected day.

Properties of the function E

1. If a is a constant, then $E(a) = a$.
2. If a , b and c are constants, X is a random variable, and g and h are functions, then $E[a + b g(X) + c h(X)] = a + b E[g(X)] + c E[h(X)]$.

Activity 3.3.

Prove the above properties.

Example 3.5.

Let X be the number of cylinders in the engine of a car that comes to the tune-up facility of a car sales company. The cost C of a tune-up is related to X as $C = 1000 + 300X + 5X^2$. According to the past data, the pmf of X is approximately given by the following function.

$$P_X(x) = \begin{cases} 0.75 & \text{if } x = 4 \\ 0.15 & \text{if } x = 6 \\ 0.10 & \text{if } x = 8 \end{cases}$$

The company undertakes 25 tune-ups per day. The company wants to charge Rs. kx for a car with x cylinders so that the daily expected profit is Rs 50000.00. What should be the value of k ?

Solution 3.5

The profit from a tune-up of a car with X cylinders is

$$P = kX - (1000 + 300X + 5X^2).$$

Therefore, the expected profit from a tune-up of a car with X cylinders is

$$\begin{aligned} E(P) &= E(kX - 1000 - 300X - 5X^2) \\ &= kE(X) - 1000 - 300E(X) - 5E(X^2) \end{aligned}$$

Note that

$$E(X) = 4(0.75) + 6(0.15) + 8(0.1) = 4.7$$

and

$$E(X^2) = 4^2(0.75) + 6^2(0.15) + 8^2(0.1) = 23.8.$$

Therefore, the expected profit from a tune-up of a car with X cylinders is

$$E(P) = k(4.7) - 1000 - 300(4.7) - 5(23.8)$$

Note even though the profit P from tune-ups of different cars are different (depends on X), the expected profit from a tune-up of a randomly selected car is fixed.

Since the expected profit from 25 tune-ups is Rs. 50000.00, the expected profit from one tune-up is Rs. 2000.00. Thus,

$$\begin{aligned} k(4.7) - 1000 - 300(4.7) - 5(23.8) &= 2000 \\ k &\approx 963.62 \end{aligned}$$

It is possible that the $E(X)$ does not exist for a random variable X . For example, if the random variable X has the density function

$$f_X(x) = \frac{1}{\pi(1+x^2)}, \quad -\infty < x < \infty,$$

then $E(X)$ does not exist. However, such situations are rare in practice. Most of the random variables that we come across in practice have finite expected values.

3.2 Variance of a random variable

Recall that the variance of a data set of size n is defined as

$$S^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

So, it is the mean of squared distances between individual values in the sample and their sample mean. It is a measure of the total distance from individual values to the mean of the values.

Similarly, we define the variance of a random variable as the population mean of squared distances between the X values and the population mean of X values.

Definition 3.4.

Let X be a random variable (discrete or continuous). Then, the variance of the random variable X is defined as

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

The positive square root of $V(X)$ is called the standard deviation of X .

$$SD(X) = \sqrt{Var(X)}$$

Theorem 3.1.

For any random variable X with finite $E(X)$,

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

Activity 3.4.

Prove the above Theorem.

Example 3.6.

Consider Example 1.1. Calculate the standard deviation of the number of containers of curd bought by a randomly selected customer.

Solution 3.6

We have already seen that $E(X) = 1.5$ and $E(X^2) = 3$ in Examples 3.1 and 3.3.

$$\begin{aligned} V(X) &= E(X^2) - [E(X)]^2 \\ &= 3 - 1.5^2 = 0.75 \\ \therefore SD(X) &= \sqrt{0.75} \approx 0.87 \end{aligned}$$

So, the number of containers of curd bought by customers is deviated from the mean number of containers bought by a customer by 0.87, on average.

Example 3.7.

Consider a repeatable experiment where the result is either a 'success' or a 'failure'. Let

$$X = \begin{cases} 0 & \text{if the experiment results in a failure} \\ 1 & \text{if the experiment results in a success} \end{cases}$$

Suppose that the probability of 'success' in an experiment is θ . Obtain expressions for $E(X)$ and $V(X)$.

Solution 3.7

$$E(X) = 1(\theta) + 0(1 - \theta) = \theta$$

$$E(X^2) = 1^2(\theta) + 0^2(1 - \theta) = \theta$$

$$V(X) = E(X^2) - [E(X)]^2 = \theta - \theta^2 = \theta(1 - \theta)$$

Example 3.8.

Consider Example 3.4. Calculate $SD(X)$.

Solution 3.8

In Example 3.4, We have calculated that $E(X) = 0$ and $E(X^2) = 0.2$.

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

$$= 0.2 - (0)^2 = 0.2$$

$$\therefore SD(X) = \sqrt{0.2} \approx 0.447$$

Activity 3.5.

Consider Activity 2.1. Calculate the standard deviation of the total waiting time on a randomly selected day.

Properties of the function V

1. If a is a constant, then $V(a) = 0$.
2. If a and b are constants, and X is a random variable then
 $V(a + bX) = b^2 V(X)$.

Activity 3.6.

Prove the above properties.

Activity 3.7.

The weights W (in kg) of papaya sold by a fruit seller are distributed with the pdf

$$f_W(w) = e^{-(w-0.5)}, w > 0.5.$$

He has bought 500 papaya at a fixed rate of Rs. 25.00 per papaya and sells at Rs. 35.00 per kg. What is the expected total profit and standard deviation of total profit from selling the 500 papaya?

Summary

In this session, you learned about two descriptive measures of a random variable; namely, the mean and the standard deviation. They are also called the mean and the standard deviation

of the corresponding distribution. In order to calculate the standard deviation, you must first calculate the variance.

It should be clearly understood that the mean and standard deviation of a random variable X are not the same as sample mean $\bar{x}$ and sample standard deviation s . The mean and standard deviation of a random variable (or a distribution) are the mean and standard deviation of the population of values of X . This population can be either the set of values that would be generated if we observe the value of X of all the units in an existing population or the set of values that would be generated if we observe the value of X in infinitely many repetitions of an experiment.

Sample mean and sample standard deviation are calculated using a sample of data from such a population. The population mean and the population standard deviation are calculated using the probability distribution; not using the data.

Learning outcomes

By now, you should be able to

- Define the mean and the standard deviation of a random variable
- Calculate the mean and the standard deviation of a random variable when the pmf or pdf is given
- Interpret the values of the mean and the standard deviation of a random variable.

Session 4

BINOMIAL DISTRIBUTION

Contents

Introduction

4.1 Binomial distribution

4.2 Derivation of pmf

4.3 Mean and variance of Binomial distribution

Summary

Learning outcomes

Introduction

In the previous sessions you learned about the distributions of random variables and their descriptive properties. When we want to calculate the probability of some event, first we define a suitable random variable and express the event of interest in terms of that random variable. Then, we employ a suitable model for the probability distribution to calculate the necessary probability. In addition, using the same model, we can calculate the descriptive measures related to the variable of interest.

There are a large number of commonly used models for probability distributions. In this session we discuss one important model that is used when the problem of interest is related to the number of 'successes' in a sample of fixed size. It is called the Binomial Distribution.

4.1 Definition

Bernoulli trial

A random experiment of which the outcome can be classified into two categories is called a Bernoulli trial. For example, recording the gender of a person (male, female), result of an exam (pass, fail), result of a medical test (positive, negative), quality of a product (defective, non-defective), result of a coin toss (head, tail) are Bernoulli trials.

In general, the possible outcomes from a Bernoulli trial are called 'success' and 'failure'. We denote these results by S and F respectively.

Binomial experiment

A situation or an experiment that conforms to the following list of requirements is called a 'binomial experiment'.

1. The experiment consists of a sequence of n trials, where n is fixed in advance of the experiment.
2. Each trial can result in one of the two possible outcomes, which we call 'success' (S) or 'failure' (F)
3. The outcomes in different trials are independent; i.e., the outcome of any particular trial does not influence the outcome of any other trial.
4. The probability of 'success' is the same for each trial; suppose this probability is θ .

Consider a binomial experiment with n trials and probability θ of a 'success' from any trial. Let

X = the number of 'successes' in n trials.

Then, X is said to have a binomial distribution with probability mass function

$$P_X(x) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}; \quad x = 0, 1, 2, \dots, n \quad (4.1)$$

This is denoted as $X \sim \text{Bin}(n, \theta)$.

4.2 Derivation of the pmf

Let x be any value of the set $\{0, 1, 2, \dots, n\}$. The event $X = x$ happens when there are x successes in n trials. This can happen in $\binom{n}{x}$ number of mutually exclusive ways. Therefore, $\Pr(X = x)$ is equal to the sum of the probabilities of those $\binom{n}{x}$ outcomes. Since the probability of each of those outcomes is $\theta^x (1 - \theta)^{n-x}$, $\Pr(X = x) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}$. Hence, the pmf of X is given by (4.1).

Binomial distribution can be used to model the distribution of the number of 'successes' in a sample of size n when the sample is taken with replacement from a finite population. Then, the situation fits exactly to a binomial experiment. It is also used as the approximate probability distribution of the number of 'successes' in a sample of size n when the sample is taken without replacement from a very large population. In this situation, the conditions 3 and 4 above are not satisfied exactly; they are satisfied approximately.

Example 4.1.

A multiple choice exam consists of 10 questions, each having 5 possible answers to choose among. There is only one correct choice to a question. Each correct answer is given 10 marks, and each wrong answer is given 0 marks. To pass, a mark of 50% or more is required. To get an A, a mark of 80% or more is required. Each question carry equal marks.

What is the chance that you will pass if you go to the exam without knowing a thing, and have to resort to pure guessing?

Solution 4.1

Let X be the number of correct answers. You will pass if and only if $X \geq 5$. Hence we need $\Pr(X \geq 5)$.

The answer to each question is either correct or wrong. Since the variable of interest is the number of correct answers, we shall consider getting a correct answer as a 'success'. Then, answering a question is a Bernoulli trial and answering the 10 questions is a binomial experiment with $n = 10$ and $\theta = 0.2$. Hence, $X \sim \text{Bin}(10, 0.2)$ and

$$P_X(x) = \binom{10}{x} (0.2)^x (0.8)^{10-x}; \quad x = 0, 1, 2, \dots, 10.$$

Therefore

$$\begin{aligned} \Pr(X \geq 5) &= 1 - \Pr(X \leq 4) \\ &= 1 - [P_X(0) + P_X(1) + P_X(2) + P_X(3) + P_X(4)] \\ &= 1 - [0.1074 + 0.2684 + 0.3020 + 0.2013 + 0.0881] \\ &= 0.0328 \end{aligned}$$

Activity 4.1.

Consider Example 4.1. Suppose you have studied enough so that on each question, 3 wrong answers can be eliminated. What is the probability that you will get an A?

Example 4.2.

According to the past data, 5% of all TV's of a certain brand are submitted for service while under warranty. Of these, 75% can be repaired whereas the other 25% must be replaced with new units. A shop has sold 15 of these TV's last week. What is the probability that at least one of those TV's will be returned for service during the warranty period?

Let X be the number of TV's (out of the 15 sold last week) that will be returned for service during the warranty period. We need $\Pr(X \geq 1)$.

Any TV will either be returned or not returned for service during the warranty period. Since the variable of interest is the number of TV's that will be returned for service during the warranty period, we shall consider returning a TV as a 'success'. The 15 TV's of interest can be considered as a sample of TV's taken without replacement from a large population. Then, we have a binomial experiment with $n = 15$ and $\theta = 0.05$. Hence $X \sim \text{Bin}(15, 0.05)$ and

$$P_X(x) = \binom{15}{x} (0.05)^x (0.95)^{15-x}; \quad x = 0, 1, 2, \dots, 15.$$

Therefore

$$\begin{aligned} \Pr(X \geq 1) &= 1 - \Pr(X = 0) \\ &= 1 - P_X(0) \\ &= 1 - \binom{15}{0} (0.05)^0 (0.95)^{15-0} \\ &= 1 - 0.4633 \\ &= 0.5367 \end{aligned}$$

Activity 4.2.

Consider Example 4.2. What is the probability that at most 1 of those TV's sold last week will have to be replaced with a new unit free of charge?

Activity 4.3.

Very large lots of products are arriving at a dealer. The dealer is willing to accept lots with less than 5% defective products. However, he cannot check each and every item in a lot to decide the actual percentage of defective products in a lot. Therefore, he uses the following method to decide whether to accept or reject a lot: He examines 20 randomly selected products from the batch and accepts the batch only if the number of defective products in the sample is less than 2.

- (i) What is the probability that a batch with 5% defective products (unknowingly to the dealer) will be accepted?
- (ii) What is the probability that a batch with 10% defective products (unknowingly to the dealer) will be accepted?
- (iii) What is the probability that a batch with 2% defective products (unknowingly to the dealer) will be rejected?
- (iv) Comment about the method that the dealer is using.

4.3 Mean and variance of Binomial distribution

Theorem 4.1.

Suppose that $X \sim \text{Bin}(n, \theta)$. Then, $E(X) = n\theta$ and $V(X) = n\theta(1 - \theta)$.

Proof

$$\begin{aligned}
 E(X) &= \sum_{x=0}^n x \binom{n}{x} \theta^x (1-\theta)^{n-x} \\
 &= \sum_{x=1}^n x \binom{n}{x} \theta^x (1-\theta)^{n-x} \\
 &= \sum_{x=1}^n x \frac{n!}{x!(n-x)!} \theta^x (1-\theta)^{n-x} \\
 &= n\theta \sum_{x=1}^n \frac{(n-1)!}{(x-1)!(n-x)!} \theta^{x-1} (1-\theta)^{n-x}
 \end{aligned}$$

Let $y = x - 1$. Then,

$$E(X) = n\theta \sum_{y=0}^{n-1} \frac{(n-1)!}{y!(n-1-y)!} \theta^y (1-\theta)^{n-1-y}$$

Now note that $\frac{(n-1)!}{y!(n-1-y)!} \theta^y (1-\theta)^{n-1-y}$ is the pmf of a $\text{Bin}(n-1, \theta)$ distribution. Hence, $\sum_{y=0}^{n-1} \frac{(n-1)!}{y!(n-1-y)!} \theta^y (1-\theta)^{n-1-y} = 1$. Therefore,

$$E(X) = n\theta$$

Recall that $V(X) = E(X^2) - [E(X)]^2$. Since we already know $E(X)$, let us now find $E(X^2)$.

$$\begin{aligned}
 E(X^2) &= \sum_{x=0}^n x^2 \binom{n}{x} \theta^x (1-\theta)^{n-x} \\
 &= \sum_{x=1}^n x^2 \binom{n}{x} \theta^x (1-\theta)^{n-x} \\
 &= \sum_{x=1}^n x^2 \frac{n!}{x!(n-x)!} \theta^x (1-\theta)^{n-x} \\
 &= n\theta \sum_{x=1}^n x \frac{(n-1)!}{(x-1)!(n-x)!} \theta^{x-1} (1-\theta)^{n-x}
 \end{aligned}$$

Let $y = x - 1$. Then,

$$\begin{aligned}
 E(X^2) &= n\theta \sum_{y=0}^{n-1} (y+1) \frac{(n-1)!}{y!(n-1-y)!} \theta^y (1-\theta)^{n-1-y} \\
 &= n\theta \left[\sum_{y=0}^{n-1} y \frac{(n-1)!}{y!(n-1-y)!} \theta^y (1-\theta)^{n-1-y} + \sum_{y=0}^{n-1} \frac{(n-1)!}{y!(n-1-y)!} \theta^y (1-\theta)^{n-1-y} \right]
 \end{aligned}$$

Now note that $\sum_{y=0}^{n-1} y \frac{(n-1)!}{y!(n-1-y)!} \theta^y (1-\theta)^{n-1-y}$ is the expected value of a $\text{Bin}(n-1, \theta)$ distribution. According to the previous derivation, this should be $(n-1)\theta$. As explained above, $\sum_{y=0}^{n-1} \frac{(n-1)!}{y!(n-1-y)!} \theta^y (1-\theta)^{n-1-y} = 1$. Therefore,

$$E(X^2) = n\theta[(n-1)\theta + 1]$$

and

$$\begin{aligned} V(X) &= E(X^2) - [E(X)]^2 \\ &= n\theta[(n-1)\theta + 1] - (n\theta)^2 \\ &= n\theta(1-\theta) \end{aligned}$$

Note: Another way to find $E(X^2)$ is to use the fact that $E(X^2) = E[X(X-1)] + E(X)$. $E[X(X-1)]$ can be found as follows.

$$\begin{aligned} E[X(X-1)] &= \sum_{x=0}^n x(x-1) \binom{n}{x} \theta^x (1-\theta)^{n-x} \\ &= \sum_{x=2}^n x(x-1) \binom{n}{x} \theta^x (1-\theta)^{n-x} \\ &= \sum_{x=2}^n x(x-1) \frac{n!}{x!(n-x)!} \theta^x (1-\theta)^{n-x} \\ &= n(n-1)\theta^2 \sum_{x=2}^n \frac{(n-2)!}{(x-2)!(n-x)!} \theta^{x-2} (1-\theta)^{n-x} \end{aligned}$$

Let $y = x - 2$. Then,

$$\begin{aligned} E[X(X-1)] &= n(n-1)\theta^2 \sum_{y=0}^{n-2} \frac{(n-2)!}{y!(n-2-y)!} \theta^y (1-\theta)^{n-2-y} \\ &= n(n-1)\theta^2 \end{aligned}$$

Example 4.3.

An airport transport service provided by a hotel can accommodate up to 10 passengers on one trip and any passenger who wishes to use this facility must make a reservation by paying Rs. 1000.00. According to previous records, 20% of all those making reservations do not appear for the trip. From those appear for the trip, the hotel charges additional Rs. 1000.00 per passenger. The hotel returns Rs. 500.00 to each person who does not appear for the trip. The total amount the hotel spends for a trip is Rs. 3000.

Suppose that 10 reservations have been made for the trip that leaves the hotel at 8.00 A.M. tomorrow. What is the expected profit from that trip?

Solution 4.3

Let X be the number of passengers that will appear for the trip (out of those 10 who have made reservations). Then, the profit from the trip is

$$Y = 1000(10) + 1000X - 500(10 - X) - 3000 = 1500X - 2000$$

Therefore, the expected profit is

$$E(Y) = 1500E(X) - 2000$$

Since $X \sim \text{Bin}(10, 0.8)$, $E(X) = 10(0.8) = 8$. Hence,

$$E(Y) = 1500(8) - 2000 = 10000$$

The expected profit from the trip is Rs 10000.00

Summary

In this session we discussed about the Binomial distribution. It is used as the model for the distribution of the number of successes in a sample of fixed size. It is also used under sampling with replacement from a finite population or sampling without replacement from a very large population. The situation when the sample is taken without replacement from a finite population is discussed in the next session.

Learning outcomes

By now you should be able to

- Identify binomial experiments,
- Calculate probabilities using binomial distribution, and
- Calculate expected values and standard deviations.

Session 5

HYPERGEOMETRIC DISTRIBUTION

Contents

Introduction

5.1 Definition

5.2 Derivation of pmf

5.2 Mean and variance of Hypergeometric distribution

Summary

Learning outcomes

Introduction

There are situations where we are interested in the distribution of the number of successes in a sample, yet the situation does not fit in to a binomial experiment exactly or approximately. This happens when the sample is taken without replacement from a finite population.

5.1 Definition

Suppose that a population of size k contains m items of a certain kind (say, 'successes'). A sample of n items is taken from this population without replacement and the number of successes in the sample are counted. We shall call such an experiment a 'hypergeometric experiment'.

Let X be the number of S's in the sample. Then, X is said to have a hypergeometric distribution. The pmf. of X is given by

$$P_X(x) = \frac{\binom{m}{x} \binom{k-m}{n-x}}{\binom{k}{n}}, \quad x = 0, 1, \dots, \min(n, m) \quad (5.1)$$

This is denoted as $X \sim HG(n, m, k)$.

5.2 Derivation of the pmf

Consider the experiment of taking a sample of n items from a population of k items, out of which m items are 'S's. The sample space contains $\binom{k}{n}$ outcomes, disregarding the order of the items in an outcome. All these outcomes are equally likely.

Let x be any value in the set $\{0, 1, 2, \dots, \min(m, n)\}$. How many samples contain x S's and $(n-x)$ F's? There are $\binom{m}{x}$ ways of selecting x S's out of m S's. There are $\binom{k-m}{n-x}$ ways of selecting $(n-x)$ F's out of $(k-m)$ F's. Therefore, the total number of ways to select a sample containing x S's and $(n-x)$ F's is $\binom{m}{x} \times \binom{k-m}{n-x}$.

Then, by the classical definition of probability, $\Pr(X = x) = \frac{\binom{m}{x} \times \binom{k-m}{n-x}}{\binom{k}{n}}$. This gives (5.1).

Example 5.1.

An English lecturer has assigned his students an essay of 10 pages. He has the following policy on marking an essay: The lecturer randomly selects 4 pages. If more than 2 of those 4 pages contain spelling or grammar errors, the essay will be returned to the student without marking. Suppose a certain essay contains 3 pages with spelling or grammar errors. What is the probability that the essay will not be marked?

Let us consider having an error on a page as a 'success'. Let X be the number of pages with errors in the sample of 4 pages. We need $\Pr(X > 2)$.

Here, the lecturer samples from a finite population without replacement. Therefore, $X \sim HG(4, 3, 10)$. Hence,

$$P_X(x) = \frac{\binom{3}{x} \binom{7}{4-x}}{\binom{10}{4}}, \quad x = 0, 1, 2, 3$$

$$\begin{aligned} \Pr(X > 2) &= \Pr(X = 3) \\ &= P_X(3) \\ &= \frac{\binom{3}{3} \binom{7}{4-3}}{\binom{10}{4}} \\ &= 0.0333 \end{aligned}$$

Activity 5.1.

A worker in an auto parts store has mixed 5 improperly assembled (defective) starters with a stock of 50 starters in working condition, by a mistake. In response to a purchase order, 10 randomly selected starters were sent to a garage. What is the probability that the garage receives at least 1 improperly assembled starter?

Activity 5.2.

The proportion of non-defective products produced by a certain production process is 95%. These products are put in boxes of 4. Before sealing the boxes, two randomly selected products are examined from each box. If at least one of those examined products are defective, all 4 products are examined and all the defective products are replaced with non-defective products before sealing the box. If the two examined products are non-defective, they are put back in the box and sealed. What is the overall proportion of outgoing boxes with zero defective products?

Mean and Variance of Hypergeometric Distribution

Theorem 5.1.

Suppose that $X \sim HG(n, m, k)$. Let $\theta = \frac{m}{k}$. Then, $E(X) = n\theta$ and $V(X) = n\theta(1 - \theta) \left(\frac{k-n}{k-1} \right)$.

Proof

For convenience, assume that $n > m$.

$$\begin{aligned} E(X) &= \sum_{x=0}^m x \frac{\binom{m}{x} \binom{k-m}{n-x}}{\binom{k}{n}} \\ &= \sum_{x=1}^m x \frac{\binom{m}{x} \binom{k-m}{n-x}}{\binom{k}{n}} \\ &= \sum_{x=1}^m x \frac{m!}{x!(m-x)!} \frac{(k-m)!}{(n-x)!(k-m-n+x)!} \frac{n!(k-n)!}{k!} \\ &= \frac{m}{k} \sum_{x=1}^m \frac{(m-1)!}{(x-1)!(m-x)!} \frac{(k-m)!}{(n-x)!(k-m-n+x)!} \frac{(n-1)!(k-n)!}{(k-1)!} \end{aligned}$$

Let $y = x - 1$. Then,

$$\begin{aligned} E(X) &= n \sum_{y=0}^{m-1} \frac{(m-1)!}{y!(m-1-y)!} \frac{(k-m)!}{(n-y-1)(k-m-n+y+1)!} \frac{(n-1)!(k-n)!}{(k-1)!} \\ &= n \sum_{y=0}^{m-1} \frac{(m-1)!}{y!(m-1-y)!} \frac{(k-1-(m-1))!}{(n-1-y)((k-1)-(m-1)-(n-1)+y)!} \frac{(n-1)!(k-1)-(n-1)!}{(k-1)!} \\ &= n \sum_{y=0}^{m-1} \frac{\binom{m-1}{y} \binom{(k-1)-(m-1)}{(n-1)-y}}{\binom{(k-1)}{(n-1)}} \end{aligned}$$

It is clear that the term inside the summation is a pmf of a $HG((n-1), (m-1), (k-1))$ distribution. Hence, the sum is 1. Therefore,

$$E(X) = n \frac{m}{k}$$

Note: Let θ = the population proportion of successes = $\frac{m}{k}$. Then, $E(X) = n\theta$, which is the same result for a $Bin(n, \theta)$ distribution.

It can be shown that

$$E(X^2) = n \frac{m}{k} \left[(n-1) \frac{(m-1)}{(k-1)} + 1 \right]$$

Hence

$$\begin{aligned} V(X) &= n \frac{m}{k} \left[(n-1) \frac{(m-1)}{(k-1)} + 1 \right] - \left(n \frac{m}{k} \right)^2 \\ &= n \frac{m}{k} \left(1 - \frac{m}{k} \right) \left(\frac{k-n}{k-1} \right) \end{aligned}$$

Letting $\theta = \frac{n}{k}$,

$$V(X) = n\theta(1-\theta) \left(\frac{k-n}{k-1} \right)$$

This differs from the variance of $Bin(n, \theta)$ distribution by the factor $\left(\frac{k-n}{k-1} \right)$. This is called the 'finite population correction (fpc) factor'. The fpc factor is close to one when n is small relative to k . So, the Binomial model can be used as an approximation to Hypergeometric distribution if k is large and n is small relative to k .

If $X \sim HG(n, m, k)$, k is large, n/k is small and $m/k = \theta$, then, $X \simeq Bin(n, \theta)$.

Activity 5.3.

If $X \sim HG(n, m, k)$, prove that $E(X^2) = n \frac{m}{k} \left[(n-1) \frac{(m-1)}{(k-1)} + 1 \right]$.

Help: Use the fact that $E(X^2) = E[X(X-1)] + E(X)$.

Activity 5.4.

In a certain company that produces milk powder, '1 kg' bag is supposed to contain 998 g to 1002 g of milk powder. When the process is in-control, 95% of the bags conform to this standard. The

other 5% are non-conforming. The quality control supervisor is planning to set up a chart called 'p-chart' to monitor whether the process is in-control. The p-chart is an on-line process control technique. It contains three lines, namely, the center line (CL), lower control limit (LCL) and upper control limit (UCL). These are given by

$$\begin{aligned} CL &= E(P) \\ UCL &= E(P) + 3SD(P) \\ LCL &= \max(0, E(P) - 3SD(P)) \end{aligned}$$

Figure 5.1 shows a typical p-chart.

The quality control supervisor is going to take a random sample of 10 bags from each batch of 100 bags produced and determine the number of non-conforming bags in the sample. Let X be the number of non-conforming bags in a sample and $P = X/10$. If the observed value of P is between the UCL and the LCL, the process is considered to be in-control. Otherwise, it is considered to be out-of-control and suitable remedial actions are taken after an inspection.

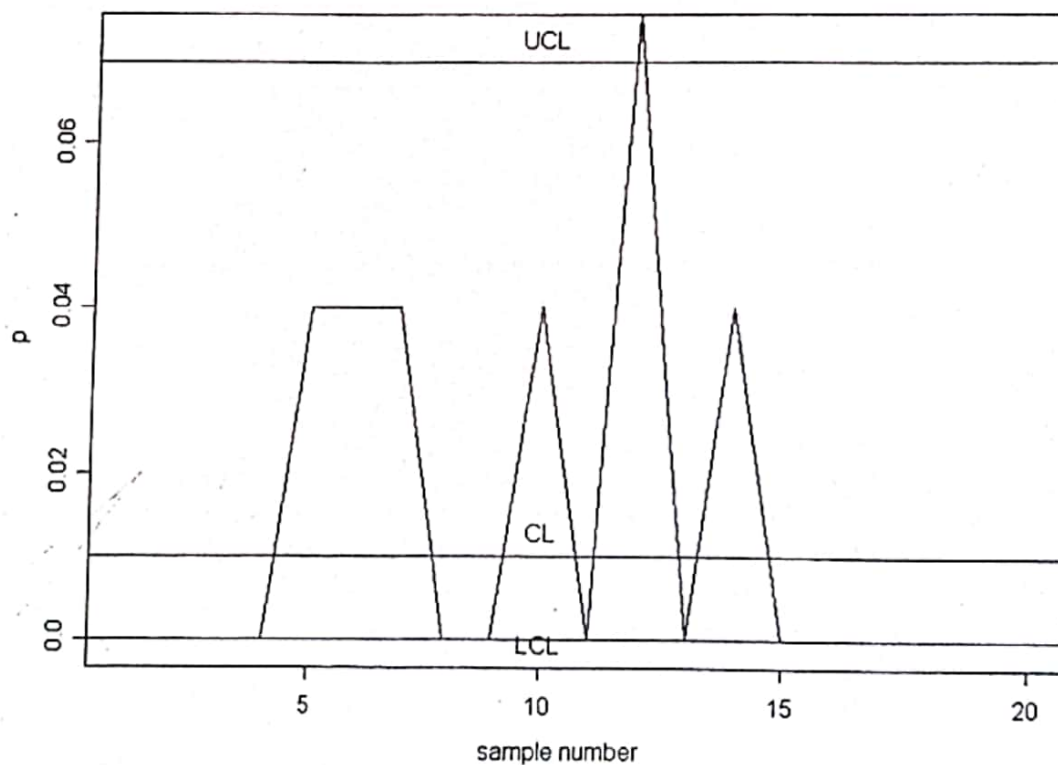


Figure 5.1: A typical p-chart.

Determine the CL, UCL and LCL for this process.

Summary

In this session, we discussed about the Hypergeometric distribution. It is used as a model for the distribution of the number of successes in a sample of fixed size when the sampling is done without replacement from a finite population. When the sample size n and the population proportion of successes θ are fixed, Binomial and Hypergeometric distributions have the same mean. The variance of the Hypergeometric distribution is equal to the variance of the Binomial distribution times a finite population correction factor. Hypergeometric distribution can be approximated by the Binomial distribution when the population size is large and the sample size is small relative to the population size.

Learning outcomes

By now you should be able to

- Identify Hypergeometric experiments,
- Calculate probabilities using Hypergeometric distribution, and
- Calculate expected values and standard deviations of random variables with Hypergeometric distributions.

Session 6

GEOMETRIC DISTRIBUTION

Contents

Introduction

6.1 Definition

6.2 Derivation of pmf

6.3 Mean and variance of Geometric distribution

Summary

Learning outcomes

Introduction

In the previous session, the variable of interest was the number of successes in a fixed number of trials, or the number of successes in a sample of fixed size. In this session, the variable of interest is the number of trials or the number of failures up to the first success in a sequence of Bernoulli trials. The distribution of such a variable is called the Geometric distribution.

6.1 Definition

Consider a sequence of independent Bernoulli trials with probability θ of 'success' for each trial.

Let X = number of failures before the first 'success'.

Then, X is said to have a Geometric distribution with probability θ of success. The pmf of X is given by

$$P_X(x) = (1 - \theta)^x \theta; \quad x = 0, 1, 2, \dots \quad (6.1)$$

6.2 Derivation of the pmf

Let x be any value of the set $\{0, 1, 2, 3, \dots\}$. The event $X = x$ occurs if and only if the first x outcomes are F 's and the $(x + 1)$ st outcome is a S . Since the outcomes are independent, $\Pr(X = x) = (1 - p)(1 - p) \dots (1 - p)p = (1 - p)^x p$. This gives (6.1).

Note:

Sometimes, the variable of interest is $Y =$ number of trials up to (including) the first 'success'.

The pmf of Y is given by

$$P_Y(y) = (1 - \theta)^{y-1} \theta; \quad y = 1, 2, \dots \quad (6.2)$$

Some authors call this the geometric distribution with probability θ of success.

In order to distinguish between the two distributions (6.1) and (6.2), we shall denote these as $X \sim G_f(\theta)$ and $Y \sim G_t(\theta)$. (G to denote the Geometric distribution, subscript f or t to denote the variable of interest; number of failures by f and the number of trials by t). Note that $Y = X + 1$.

Example 6.1.

A politician contesting for a provincial council claims that he will get at least 40 percent of all the votes. A reporter interviewed individuals until he finds 1 individual who support this politician. He had to interview 5 individuals. What is the probability of happening such an event

(i) if the actual proportion of supporters is 0.4?

(ii) if the actual proportion of supporters is 0.5?

(iii) if the actual proportion of supporters is 0.6?

What do you think about the candidate's claim?

Solution 6.1

Let Y be the number of individuals interviewed until the first supporter is found.

(i) If the actual proportion of supporters is 0.4, then, $Y \sim G_t(0.4)$. Hence,

$$\Pr(Y = 5) = (1 - 0.4)^4 (0.4) = 0.0518$$

(ii) If the actual proportion of supporters is 0.5, then, $Y \sim G_t(0.5)$. Hence,

$$\Pr(Y = 5) = (1 - 0.5)^4 (0.5) = 0.0313$$

(iii) If the actual proportion of supporters is 0.6, then, $Y \sim G_t(0.6)$. Hence,

$$\Pr(Y = 5) = \frac{(1 - 0.6)^4 (0.6)}{(1 - 0.6)^4 \times (0.6)} = 0.0154$$

36

So, if the candidate's claim is correct, the maximum probability of occurring an event like above is 0.0518. In other words, if the candidate's claim is correct, an event like above is very unlikely to occur. However, the event has occurred. Therefore, the candidate's claim is unlikely to be correct.

Activity 6.1.

According to the past data, the probability of rain in a certain area on a given day during August is 0.01. A one-day cricket tournament is going to be held for 10 days in this area in next August. What is the probability that there will be no rain during the tournament? Answer this question using two suitable distributions.

6.3 Mean and Variance of a Geometric Distribution

Theorem 6.1.

Suppose that $X \sim G_f(\theta)$. Then,

$$(a) E(X) = \frac{1-\theta}{\theta}$$

$$(b) V(X) = \frac{1-\theta}{\theta^2}$$

Proof

$$\begin{aligned} E(X) &= \sum_{x=0}^{\infty} x(1-\theta)^x \theta \\ &= \sum_{x=1}^{\infty} x(1-\theta)^x \theta \\ &= \theta(1-\theta) \sum_{x=1}^{\infty} x(1-\theta)^{x-1} \end{aligned}$$

We may use the following fact to calculate the above series.

$$S = \sum_{x=1}^{\infty} k^x = 1 + k + k^2 + \dots$$

$$S = \frac{1}{1-k}$$

$$\sum_{x=1}^{\infty} x k^{x-1} = \frac{dS}{dk} = \frac{1}{(1-k)^2}$$

Accordingly,

$$\begin{aligned} E(X) &= \theta(1-\theta) \frac{1}{(1-(1-\theta))^2} \\ &= \frac{(1-\theta)}{\theta} \end{aligned}$$

Activity 6.2.

Prove part (b) of Theorem 6.1.

Activity 6.3.

Suppose $Y \sim G_t(\theta)$. Find $E(Y)$ and $V(Y)$.

Example 6.2.

A newly married couple have decided to have children until they get a son. Suppose that the probability of a male birth is 0.45. What is the expected family size when they achieve their objective?

Solution 6.2

Let X be the number of daughters until the first son. Then, The family size is $S = X + 1 + 2 = X + 3$.

Since $X \sim G_f(0.45)$, $E(X) = (1 - 0.45)/0.45 = 1.22$ Therefore, $E(S) = E(X) + 3 = 1.22 + 3 = 4.22$.

Activity 6.4.

The p -chart is an on-line process control technique. It contains three lines, namely, the center line (CL), lower control limit (LCL) and upper control limit (UCL). (See figure 5.1).

The production engineer wants to establish a mechanism to control the proportion of defective products in a production process at 0.01 level. So he maintains a p -chart with the following control lines: $CL=0.01$, $LCL=0$, $UCL=0.07$.

The procedure that the engineer follows is as follows. A random sample of 25 products is taken from the process every one hour and the sample proportion of defective products is determined. The sample proportion is plotted against the sample number in the p -chart. If the sample proportion is above the UCL, an 'out of control signal' is given. Then, the process is stopped and checked for possible 'assignable causes' for the process to go out of control.

Suppose that the actual proportion of defective products in the production process has increased to 0.05 (the process has gone out-of-control) unknown to the engineer at 9.30 A.M. The next sample is due to be taken at 10.00 A.M.

- (i) What is the probability of receiving an 'out of control signal' from the sample taken at 10.00 A.M.?
- (ii) What is the probability of receiving an 'out of control signal' for the first time from the sample taken at 1.00 P.M.?
- (iii) What is the expected time to receive the first 'out of control signal'?

Summary

In this session we discussed about the distribution for the number of failures up to the first success in a sequence of Bernoulli trials. It is called the Geometric distribution. The number of trials up to the first success also follows a geometric distribution.

Then, X is said to have a negative binomial distribution with parameter θ . The pmf of X is given by

$$P_X(x) = \binom{x+r-1}{r-1} \theta^r (1-\theta)^x; \quad x = 0, 1, 2, \dots \quad (7.1)$$

7.2 Derivation of the pmf

Let x be any value of the set $\{0, 1, 2, 3, \dots\}$. The event $X = x$ occurs if and only if there are x F 's and $(r-1)$ S 's in any order from the first $x+r-1$ trials and a S from the $(x+r)$ th trial.

The probability of any such outcome is $(1-\theta)^x \theta^{r-1} \times \theta$. There are $\binom{x+r-1}{r-1}$ number of ways of getting such outcomes. Hence $\Pr(X = x) = \binom{x+r-1}{r-1} \theta^r (1-\theta)^x$.

Note:

Sometimes, the variable of interest is Y = total number of trials required to get r 'successes'. The pmf of Y is given by

$$P_Y(y) = \binom{y-1}{r-1} \theta^r (1-\theta)^{y-r}; \quad y = r, r+1, \dots \quad (7.2)$$

Some authors call this the Negative Binomial distribution with probability θ of success.

In order to distinguish between the two distributions (7.1) and (7.2), we shall denote these as $X \sim NB_f(r, \theta)$ and $Y \sim NB_t(r, \theta)$. (NB to denote the Negative Binomial distribution, subscript f or t to denote the variable of interest; number of failures by f and the number of trials by t). Note that $Y = X + r$.

Example 7.1.

A specialist in obesity wishes to recruit 5 obese patients to participate in a weight loss experiment. From past experience he knows that about 1 in 6 patients are willing to participate in such experiments. He examines 25 obese patients per day in his clinic. What is the probability that he can recruit 5 patients in a day?

Let Y be the total number of patients to be asked until 5 patients agree. Then, $Y \sim NB_t(5, 1/6)$. The specialist can recruit 5 patients in a day if and only if $Y \leq 25$.

$$\begin{aligned} \Pr(Y \leq 25) &= \sum_{y=5}^{25} \binom{y-1}{5-1} \theta^5 (1-\theta)^{y-5} \\ &= 0.5757 \end{aligned}$$

Activity 7.1.

The sales person of a pharmacy near a children's hospital notices on one morning that there are only 10 bottles of syrup A left. He knows that approximately 1 in 5 prescriptions require syrup A. What is the probability that he will be able fill at least 30 prescriptions before he gets a new stock of syrup A? (Assume that all other drugs are available).

7.3 Relationship between Geometric distribution and Negative binomial distribution

Theorem 7.1.

Sum of r independent, identically distributed $G_f(\theta)$ random variables has a $NB_f(\theta)$ distribution.

Proof

Suppose you repeat a Bernoulli experiment until you get r successes. Suppose the probability of a success in any trial is θ . Let X_1 be the number of failures before the first success. Then, $X_1 \sim G_f(\theta)$. Let X_2 be the number of failures after the first success but before the second success. Then, $X_2 \sim G_f(\theta)$. Similarly, let X_i be the number of failures after the $(i-1)$ st success but before the i th success ($i = 1, 2, \dots, r$). Then, $X_i \sim G_f(\theta)$. Also note that X_i 's are independent. Now, let X be the number of failures before the r th success. It is clear that $X = X_1 + X_2 + \dots + X_r$. Since $X \sim NB_f(r, \theta)$, it is clear that the sum of r independent, identically distributed $G_f(\theta)$ random variables has a $NB_f(r, \theta)$ distribution.

A more mathematical proof is given in Theorem 15.8.

Mean and Variance of a Negative Binomial Distribution

Theorem 7.2.

Suppose that $X \sim G_f(\theta)$. Then,

$$(a) E(X) = r \left(\frac{1-\theta}{\theta} \right)$$

$$(b) V(X) = r \left(\frac{1-\theta}{\theta^2} \right)$$

Proof

Theorem 7.2 is directly implied by Theorems 6.1 and 7.1.

Example 7.2.

A university wants to estimate the proportion of color blind students. Instead of the direct method of taking a sample and finding out the sample proportion of color blind students, it has used the indirect method. The university doctor has examined randomly selected students one by one, until he finds 5 color blind students. He had to examine 503 students. Estimate the proportion of color blind students in the university.

Solution 7.2

Let θ be the proportion of color blind students in the university.

Let Y be the number of students that have to be examined until 5 color blind students are found. Then, $Y \sim NB_f(5, \theta)$ and $E(Y) = 5 \left(\frac{1-\theta}{\theta} \right) + 5$ (why?)

Assuming that the observed number of Y is close to $E(Y)$,

$$5 \left(\frac{1-\theta}{\theta} \right) + 5 \approx 503$$

$$\theta \approx 0.0099$$

Learning Outcomes

By now, you should be able to

- Recognize situations that the Geometric distribution can be applied.
- Calculate probabilities using the Geometric distribution.
- Calculate the mean and the variance.

Session 7

NEGATIVE BINOMIAL DISTRIBUTION

Contents

Introduction

7.1 Definition

7.2 Derivation of pmf

7.3 Relationship between Geometric distribution and Negative binomial distribution

7.3 Mean and variance of Negative binomial distribution

Summary

Learning outcomes

Introduction

In the last session, the variable of interest was the number of failures or the number of trials up to the first success in a sequence of Bernoulli trials. An extension arises when the variable of interest is the number of failures or the number of trials up to a given number of successes in a sequence of Bernoulli trials. The distribution of such a variable is called the Negative binomial distribution.

7.1 Definition

Consider a sequence of independent Bernoulli trials with probability of 'success' θ for each trial. Let X = number of failures before the r th 'success', where $r > 0$ is a fixed integer.

Summary

In this session we discussed about the distribution for the number of failures up a given number r of successes in a sequence of Bernoulli trials. It is called the Negative Binomial distribution. Geometric distribution is the special case when $r = 1$. The distribution of the number of trials up to the r th success is also called the Negative binomial distribution.

Learning Outcomes

By now, you should be able to

- Recognize situations that the Negative Binomial distribution can be applied.
- Calculate probabilities using the Negative Binomial distribution.
- Calculate the mean and the variance.

Session 8

POISSON DISTRIBUTION

Contents

Introduction

8.1 Definition

8.2 Derivation of pmf

8.3 Mean and variance of Poisson distribution

8.4 Poisson approximation to the binomial

Summary

Learning outcomes

Introduction

So far, we have discussed the distributions for the number of successes in a fixed number of trials, and the number of trials up to a fixed number of successes. The distribution of the number of events of some kind during a given period of time is another variable of interest to many people. For example, the number of traffic accidents per week in a given city, the number of radioactive particle emissions per unit of time, the number of telephone calls coming into the reception of a company per hour, the number of meteorites that collide with a test satellite during a single orbit, etc. are all variables of that category. The Poisson distribution often provides a realistic probability model for such variables. The Poisson distribution can be used for counts of some sort in a given area or volume as well. For example, the number of bacteria per unit volume of some fluid, the number of defects per unit length of some material, the number of flaws per unit length of some wire, etc.

8.1 Definition

Let $N(t)$ be the number of events occur during an 'interval' of size t , where the 'interval' can be a period of time, a length of some material, an area or a volume of some substance. In addition, let $o(h)$ be a function such that $\lim_{h \rightarrow 0} \frac{o(h)}{h} \rightarrow 0$. For example, h^2, h^3 are $o(h)$ while h is not $o(h)$. As h goes to zero, $o(h)$ functions go to zero faster than h .

Suppose that the following assumptions are reasonable:

1. The probability of an event within any interval of length h is the same. That is, the probability of an event during $(t_1, t_1 + h)$ is the same as the probability of an event during $(t_2, t_2 + h)$, for any t_1, t_2 .
2. The probability of occurring one event over a small interval of length h is approximately proportional to the length of that interval. That is, $\Pr(N(h) = 1) = \lambda h + o(h)$. Here λ is the proportionality constant.
3. The probability of two or more events occurring in the same narrow interval is negligible. That is, $\Pr(N(h) \geq 2) = o(h)$.
4. The probability of an event in an interval is independent of the probability of an event in any other non-overlapping interval.

Then, the pmf of $N(t)$ is given by

$$P_{N(t)}(n) = \frac{e^{-\lambda t} (\lambda t)^n}{n!}; \quad n = 0, 1, 2, \dots \quad (8.1)$$

This is said to have a Poisson distribution with rate λt per 'interval' of size t . λ is the rate of events per unit 'interval'. This is denoted as $N(t) \sim \text{Poisson}(\lambda t)$.

Note:

Usually, N is used instead of $N(t)$ for convenience. Then, the pmf is written as

$$P_N(n) = \frac{e^{-\lambda t} (\lambda t)^n}{n!}; \quad n = 0, 1, 2, \dots$$

8.2 Derivation of the pmf

We show how to derive expressions for $\Pr(N(t) = 0)$ and $\Pr(N(t) = 1)$. The rest can be shown by induction.

Consider a small time interval of length h . The first three of above assumptions can be expressed mathematically as follows.

$$\Pr(N(h) = k) = \begin{cases} \lambda h + o(h) & k = 1 \\ o(h) & k \geq 2 \\ 1 - \lambda h + o(h) & k = 0. \end{cases}$$

Let $t \geq 0$. We obtain $\Pr(N(t) = n)$ for each $n = 0, 1, \dots$, through a differential equation. For this purpose, let

$$g_n(t) = \Pr[N(t) = n].$$

Derivation of $\Pr(N(t) = 0)$

$$\begin{aligned}
 g_0(t+h) &= \Pr[N(t+h) = 0] \\
 &= \Pr(0 \text{ events during an interval of length } t+h) \\
 &= \Pr(0 \text{ events during the interval } 0 \text{ to } t+h) \text{ (a special case of assumption 1)} \\
 &= \Pr(0 \text{ events during } 0 \text{ to } t \text{ and } 0 \text{ events during } t \text{ to } t+h) \\
 &= \Pr[N(t) = 0] \Pr[N(h) = 0] \text{ (assumption 4)} \\
 &= g_0(t) [1 - \lambda h + o(h)]
 \end{aligned}$$

$$\begin{aligned}
 \lim_{h \rightarrow 0} \frac{g_0(t+h) - g_0(t)}{h} &= -\lambda g_0(t) + \lim_{h \rightarrow 0} \frac{o(h)}{h} \\
 g_0'(t) &= -\lambda g_0(t) \\
 g_0(t) &= ce^{-\lambda t} \text{ for some constant } c
 \end{aligned}$$

Now, in order to determine the value of c , we use the following boundary condition $\Pr(N(0) = 0) = g_0(0) = 1$. By substituting this in the above equation, we get $c = 1$. Hence

$$g_0(t) = e^{-\lambda t}.$$

This means,

$$\Pr(N(t) = 0) = e^{-\lambda t}$$

This proves 8.1 for $n = 0$.

Derivation of $\Pr(N(t) = 1)$

$$\begin{aligned}
 g_1(t+h) &= \Pr[N(t+h) = 1] \\
 &= \Pr(1 \text{ event during an interval of length } t+h) \\
 &= \Pr(1 \text{ event during the interval } 0 \text{ to } t+h) \text{ (a special case of assumption 1)} \\
 &= \Pr(1 \text{ event during } 0 \text{ to } t \text{ and } 0 \text{ events during } t \text{ to } t+h \\
 &\quad \text{or } 0 \text{ event during } 0 \text{ to } t \text{ and } 1 \text{ event during } t \text{ to } t+h) \\
 &= \Pr(1 \text{ event during } 0 \text{ to } t \text{ and } 0 \text{ events during } t \text{ to } t+h) \\
 &\quad + \Pr(0 \text{ events during } 0 \text{ to } t \text{ and } 1 \text{ event during } t \text{ to } t+h) \\
 &= \Pr[N(t) = 1] \Pr[N(h) = 0] \\
 &\quad + \Pr[N(t) = 0] \Pr[N(h) = 1] \\
 &= g_1(t) [1 - \lambda h + o(h)] + g_0(t) [\lambda h + o(h)]
 \end{aligned}$$

$$\begin{aligned}
\lim_{h \rightarrow 0} \frac{g_1(t+h) - g_1(t)}{h} &= \lim_{h \rightarrow 0} \frac{-\lambda h g_1(t)}{h} + \lim_{h \rightarrow 0} \frac{\lambda h g_0(t)}{h} + \lim_{h \rightarrow 0} \frac{o(h)}{h} \\
g_1'(t) &= -\lambda g_1(t) + \lambda g_0(t) \\
g_1'(t) &= -\lambda g_1(t) + \lambda e^{-\lambda t} \\
e^{\lambda t} [g_1'(t) + \lambda g_1(t)] &= \lambda \\
\frac{d}{dt} [e^{\lambda t} g_1(t)] &= \lambda \\
e^{\lambda t} g_1(t) &= \int \lambda dt + c \\
&= \lambda t + c
\end{aligned}$$

Since $g_1(0) = 0$, $c = 0$. Hence,

$$g_1(t) = \lambda t e^{-\lambda t}$$

This means,

$$\Pr(N(t) = 1) = \lambda t e^{-\lambda t}$$

This proves 8.1 for $n = 1$.

Derivation of $\Pr(N(t) = n)$

By induction, it can be shown that $g_n(t) = \frac{e^{-\lambda t} (\lambda t)^n}{n!}$. We omit this derivation.

Example 8.1.

A towing service receives requests for towing vehicles at a rate of 2 per hour.

Compute the probability of receiving more than 3 requests during a two-hour period

Solution 8.1

Let N be the number of requests during two hours. Since the above assumptions are reasonable, we may assume that $N \sim \text{Poisson}(4)$. Then,

$$\begin{aligned}
\Pr(N > 3) &= 1 - \Pr(N \leq 2) \\
&= 1 - (P_N(0) + P_N(1) + P_N(2)) \\
&= 1 - \left(\frac{e^{-2}(2)^0}{0!} + \frac{e^{-2}(2)^1}{1!} + \frac{e^{-2}(2)^2}{2!} \right) \\
&= 1 - 0.2381 \\
&= 0.7619
\end{aligned}$$

Activity 8.1.

Consider Example 8.1. If the operators of the towing service take a 30 minute break for lunch, what is the probability that they do not miss any calls for assistance?

Example 8.2.

A bag of 200 chocolate chips is dumped in to a batch of cookie dough and mixed well. 40 cookies are made from such a batch of dough. What is the probability that a randomly selected cookie has at least 4 chocolate chips?

Solution 8.2

Let X be the chocolate chips in a randomly selected cookie. Then, the above assumptions are reasonable with time interval replaced by area. Hence, we may assume that $X \sim \text{Poisson}(5)$.

$$\Pr(X \geq 4) = 1 - \Pr(X \leq 3)$$

We leave the reader to complete the calculation.

8.3 Mean and Variance of a Poisson Distribution

Theorem 8.1.

Suppose that $X \sim \text{Poisson}(\lambda)$. Then,

(a) $E(X) = \lambda$

(b) $V(X) = \lambda$

Proof

$$\begin{aligned} E(X) &= \sum_{x=0}^{\infty} x \frac{e^{-\lambda}(\lambda)^x}{x!} \\ &= \sum_{x=1}^{\infty} x \frac{e^{-\lambda}(\lambda)^x}{x!} \\ &= \lambda \sum_{x=1}^{\infty} \frac{e^{-\lambda}(\lambda)^{x-1}}{(x-1)!} \end{aligned}$$

Let $y = x - 1$. Then,

$$E(X) = \lambda \sum_{y=0}^{\infty} \frac{e^{-\lambda}(\lambda)^y}{y!}$$

Note that $\frac{e^{-\lambda}(\lambda)^y}{y!}$ is the pmf of a $\text{Poisson}(\lambda)$ distribution. Hence, $\sum_{y=0}^{\infty} \frac{e^{-\lambda}(\lambda)^y}{y!} = 1$. Therefore,

$$E(X) = \lambda$$

Note that

$$\begin{aligned} E(X(X-1)) &= \sum_{x=0}^{\infty} x(x-1) \frac{e^{-\lambda}(\lambda)^x}{x!} \\ &= \sum_{x=2}^{\infty} x(x-1) \frac{e^{-\lambda}(\lambda)^x}{x!} \\ &= \lambda^2 \sum_{x=2}^{\infty} \frac{e^{-\lambda}(\lambda)^{x-2}}{(x-2)!} \end{aligned}$$

Let $y = x - 1$. Then,

$$\begin{aligned} E(X(X-1)) &= \lambda^2 \sum_{y=0}^{\infty} \frac{e^{-\lambda}(\lambda)^y}{y!} \\ &= \lambda^2 \\ E(X^2) - E(X) &= \lambda^2 \\ E(X^2) &= \lambda^2 + E(X) = \lambda^2 + \lambda \end{aligned}$$

Hence,

$$\begin{aligned} V(X) &= E(X^2) - [E(X)]^2 \\ &= \lambda^2 + \lambda - \lambda^2 \\ &= \lambda \end{aligned}$$

Example 8.3.

The *c*-chart is a on-line quality control technique. It is used to control the average number of nonconformities per unit in a production line. In a typical control chart there are three horizontal lines. Let X be the number of nonconformities in a unit. Then, the standard control limits of a *c*-chart are given by

$$\begin{aligned} CL &= E(X) \\ LCL &= \max(0, E(X) - 3SD(X)) \\ UCL &= E(X) + 3SD(X). \end{aligned}$$

A production process assembles electric clocks. The average (mean) number of nonconformities per clock is estimated to be 0.5. The quality engineer wishes to maintain a *c*-chart for the number of nonconformities per unit. The engineer has decided to take six clocks as one unit.

(i) Find the control limits for the *c*-chart.

(ii) The OC curve of a control chart is a graphical display of its sensitivity. The OC curve for a *c*-chart is a graph of $\beta(\lambda) = \Pr(LCL \leq X \leq UCL | X \sim \text{Poisson}(\lambda))$ versus λ . Draw the OC-curve of the above *c*-chart. (You will learn the practical use of the OC curve in due course).

Solution 8.3

Since X is the number of nonconformities per 6 clocks, $X \sim \text{Poisson}(3)$. Therefore,

$$\begin{aligned} CL &= E(X) \\ LCL &= \max(0, E(X) - 3SD(X)) = \max(0, 3 - 3\sqrt{3}) = 0 \\ UCL &= E(X) + 3SD(X) = 3 + 3\sqrt{3} \approx 8.2. \end{aligned}$$

$$\begin{aligned} \beta(1) &= \Pr(0 \leq X \leq 8.2 | X \sim \text{Poisson}(1)) \\ &= \sum_{x=0}^8 \frac{e^{-1}(1)^x}{x!} \\ &= 0.9999 \end{aligned}$$

$$\begin{aligned} \beta(2) &= \Pr(0 \leq X \leq 8.2 | X \sim \text{Poisson}(2)) \\ &= \sum_{x=0}^8 \frac{e^{-2}(2)^x}{x!} \\ &= 0.9997 \end{aligned}$$

$$\begin{aligned} \beta(3) &= \Pr(0 \leq X \leq 8.2 | X \sim \text{Poisson}(3)) \\ &= \sum_{x=0}^8 \frac{e^{-3}(3)^x}{x!} \\ &= 0.996 \end{aligned}$$

$$\begin{aligned} \beta(4) &= \Pr(0 \leq X \leq 8.2 | X \sim \text{Poisson}(4)) \\ &= \sum_{x=0}^8 \frac{e^{-4}(4)^x}{x!} \\ &= 0.979 \end{aligned}$$

Similarly, calculating $\beta(\lambda)$ for other values of λ , the OC curve can be drawn as in Figure 8.1.

8.4 Poisson approximation to the Binomial distribution

Manual calculation of Binomial probabilities is tedious when n is large, especially, due to the difficulty in calculating $n!$. The following theorem makes such calculations easier under certain conditions.

Theorem 8.3.

Suppose that $X \sim \text{Bin}(n, p)$. If $n \geq 100$, $p \leq 0.01$ and $np \leq 20$, then $X \approx \text{Poisson}(np)$.

$$\Pr(X = x) \approx \frac{e^{-np} (np)^x}{x!}$$

Proof

Beyond the scope of this course.

Summary

In this session, we have seen that the Poisson distribution is a useful distribution for the number of events during a given period of time. The underlying assumptions are reasonable in most situations. However, it is important to notice that there are situations where such assumptions are not reasonable. So, you must think of the appropriateness of the four assumptions before using the Poisson model. Poisson model can also be used to calculate Binomial probabilities under certain conditions.

Learning Outcomes

By now, you should be able to

- Check whether the Poisson assumptions are reasonable for a given situation.
- When suitable, calculate probabilities using the Poisson model.
- Calculate the mean and variance.
- Calculate Binomial probabilities using Poisson approximation when suitable.

Activity 8.3.

During the time between 7.00 A.M. and 9.00 A.M, buses come to my bus-stop at a rate of 6 per hour. I get up to a bus only if there is a vacant seat. The probability that there is a vacant seat in a randomly selected bus when it comes to my bus-stop is 0.4.

I go to the bus-stop at 7.00 A.M.

- (i) What is the probability that at least one bus will come within 10 minutes?
- (ii) What is the probability that I will be able to get in to a bus within one hour from my arrival to the bus-stop?
- (iii) What is the probability that I will have to wait until at least 5 buses come to my bus-stop to get in to a bus?

There are situations, where we are interested in the number of events during a given period of time, yet the Poisson distribution is not suitable. The following are two such examples

Example 8.4.

The infection rate at the IDH hospital is typically expressed as a number of infections per patient days. This is obviously counting a number of events across both time and patients. Does this data follow a Poisson distribution?

Solution 8.4

To assume that the number of infections per day has a Poisson distribution, we need to assume the following:

1. We need to assume that the probability of infection does not change over time. In other words, the probability of infection early in the stay is the same as the probability of infection later in the stay.
2. We need to assume that the probability of getting an infection over a short time period is proportional to the length of the time period. For example, a patient who stays one hour has twice the risk of a single infection as a patient who stays 30 minutes.
3. We need to assume that for a small enough interval, the probability of getting two or more infections is negligible.
4. We also need to assume that the probability of seeing an infection in one time period does not increase or decrease the probability of seeing an infection in another period.

All of these assumptions are suspect, but especially the first and the last. If one patient gets an infection it increases the chance that other patient will get the same infection, the infection rate changes from early in the stay to later in the stay. Therefore, it is not reasonable to assume that the number of infection per day has a Poisson Distribution.

Activity 8.4.

For a class assignment, a student counts the number of cars that pass by a junction on a busy street during one minute intervals. He computes a mean of 10.3, and a variance of only 5.3. Is the Poisson model suitable for the number of cars that pass by during one minute?

9.1 Uniform Distribution

Uniform distribution is a suitable model for the density function of a continuous random variable X , if the following assumptions are reasonable.

1. The possible values of X are in a finite interval (a, b) .
2. The probability that X being in any interval of the same length, within (a, b) , is the same. In other words, there is no reason to think that any interval of given length has a higher or lower probability than any other interval of the same length.

Random variable X is said to have a uniform distribution (also called rectangular distribution) in (a, b) if its probability density function is given by

$$f_X(x) = \frac{1}{b-a}; \quad a < x < b.$$

This is denoted by $X \sim U(a, b)$ or $X \sim \text{Unif}(a, b)$.

Example 9.1.

At a well-known Chinese restaurant, the preparation time for a special dish on the menu is listed as 30 minutes. The chef has recorded the preparation time for 100 such dishes for 100 randomly chosen orders. The minimum and maximum preparation times are 25.3 and 34.9 (in minutes). The histogram of the preparation times is presented in figure 9.1.

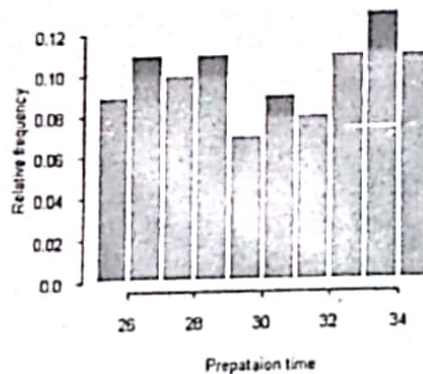


Figure 9.1:

Let X be the preparation time for a randomly chosen order. Assumption 1 is naturally reasonable. In the histogram, the frequencies in different intervals of equal widths are slightly

different. However, there is no particular pattern. There is no reason to think that some values are more likely than the other values, in general. Hence, assumption 2 seems reasonable. Therefore, it may be reasonable to assume that $X \sim U(25.3, 34.9)$.

9.2 Mean and variance of a Uniform distribution

Suppose $X \sim U(a, b)$. Then,

$$\begin{aligned} E(X) &= \int_a^b x \frac{1}{b-a} dx \\ &= \frac{(a+b)}{2} \end{aligned}$$

$$\begin{aligned} E(X^2) &= \int_a^b x^2 \frac{1}{b-a} dx \\ &= \frac{(a^2 + ab + b^2)}{3} \end{aligned}$$

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

9.3 Normal Distribution

The normal distribution (also called the Gaussian distribution) is probably the most often used probability density function in the world of Statistics. The shape of the the normal density is commonly known as 'bell shape'.

Random variable X is said to have a normal distribution with location parameter μ and scale parameter σ if its probability density function given by

$$\begin{aligned} f_X(x) &= \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{1}{2} \left(\frac{x-\mu}{\sigma} \right)^2 \right\}; -\infty < x < \infty \\ &\quad -\infty < \mu < \infty, \sigma > 0 \end{aligned}$$

This is denoted by $X \sim N(\mu, \sigma^2)$. Carefully note that the second value within brackets is the σ^2 ; not σ .

Figure 9.2 shows normal density functions with different values of μ and σ .

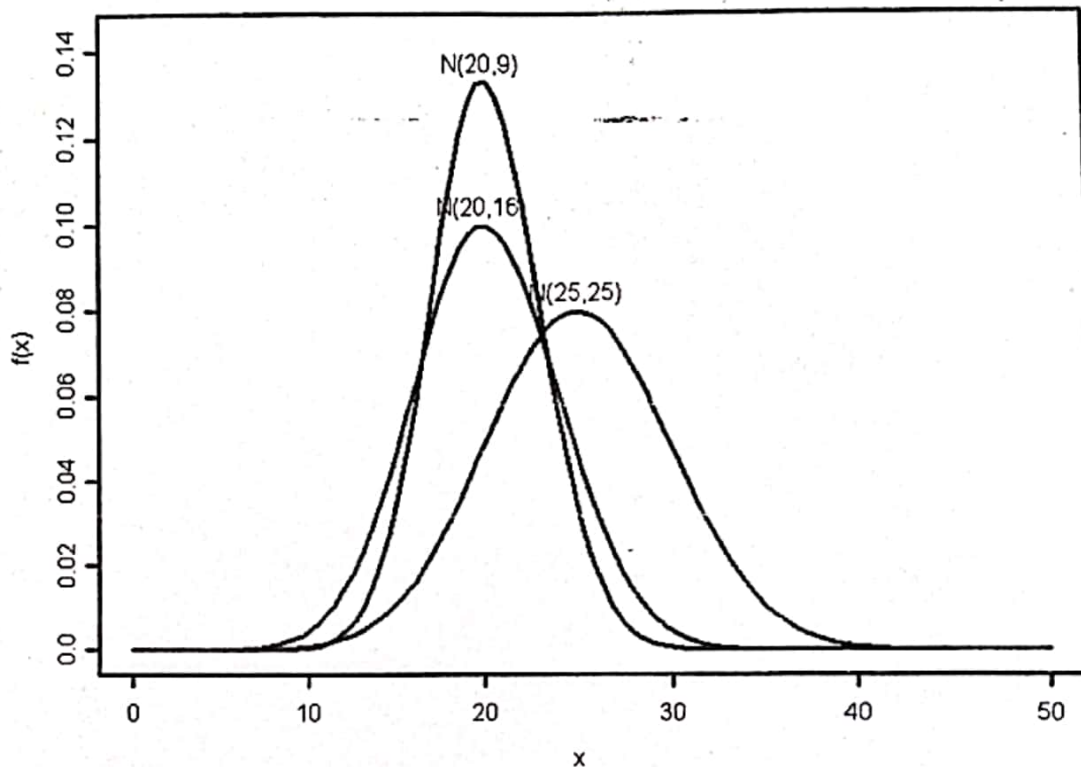


Figure 9.2: Normal density functions

It is clear from the above graphs that the normal density function is symmetric around the location parameter μ . The dispersion (or the spread) of the distribution depends on the scale parameter σ .

Eventhough the theoretical domain of this distribution is the entire real line, the density comes very close to zero when $x = \mu \pm 3\sigma$. Therefore, the normal distribution is often used for random variables with approximately bell shaped distributions in finite intervals.

9.4 Mean and Variance of a Normal Distribution

Suppose that $X \sim N(\mu, \sigma^2)$. Then,

$$E(X) = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

Let $z = \frac{x-\mu}{\sigma}$. Then,

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} (\sigma z + \mu) \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}z^2} \sigma dz \\ &= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z e^{-\frac{1}{2}z^2} dz + \frac{\mu}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2} dz \end{aligned}$$

We use the following mathematical identities to simplify the above integrals.

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2} dz = \sqrt{2\pi} \quad (9.1)$$

$$\int_{-\infty}^{\infty} z e^{-\frac{1}{2}z^2} dz = 0 \quad (9.2)$$

Using these identities, it is easy to show that

$$E(X) = \mu$$

Now, note that

$$E\left[\left(\frac{X-\mu}{\sigma}\right)^2\right] = \int_{-\infty}^{\infty} \left(\frac{x-\mu}{\sigma}\right)^2 \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

Let $z = \frac{x-\mu}{\sigma}$. Then,

$$E\left[\left(\frac{X-\mu}{\sigma}\right)^2\right] = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} z^2 e^{-\frac{1}{2}z^2} \sigma dz$$

Using the mathematical identity

$$\int_{-\infty}^{\infty} z^2 e^{-\frac{1}{2}z^2} dz = \sqrt{2\pi}, \quad (9.3)$$

$$E\left[\left(\frac{X-\mu}{\sigma}\right)^2\right] = 1$$

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

$$V(X) = \sigma^2$$

$$SD(X) = \sigma$$

The location parameter μ and the scale parameter σ are the same as the mean and the standard deviation respectively. Hereafter we call $N(\mu, \sigma^2)$ the normal distribution with mean μ and standard deviation σ .

Example 9.2.

The following are the capacities (in liters) the fuel tanks of 50 randomly selected cars of a certain model. Figure 9.3 is a histogram of these data.

30.1 29.6 30.6 29.4 28.8 29.6 29.8 30.9 29.4 30.8 30.2 29.6 30.1 30.7 29.9 31.5 29.7 29.6
29.5 30.2 29.4 29.5 29.0 30.1 30.4 29.2 29.5 30.0 29.6 30.6 29.7 29.9 29.5 30.6 30.9 30.5 30.1
30.3 29.8 29.8 29.6 30.0 30.4 28.9 30.1 29.5 29.8 29.2 29.9 29.8

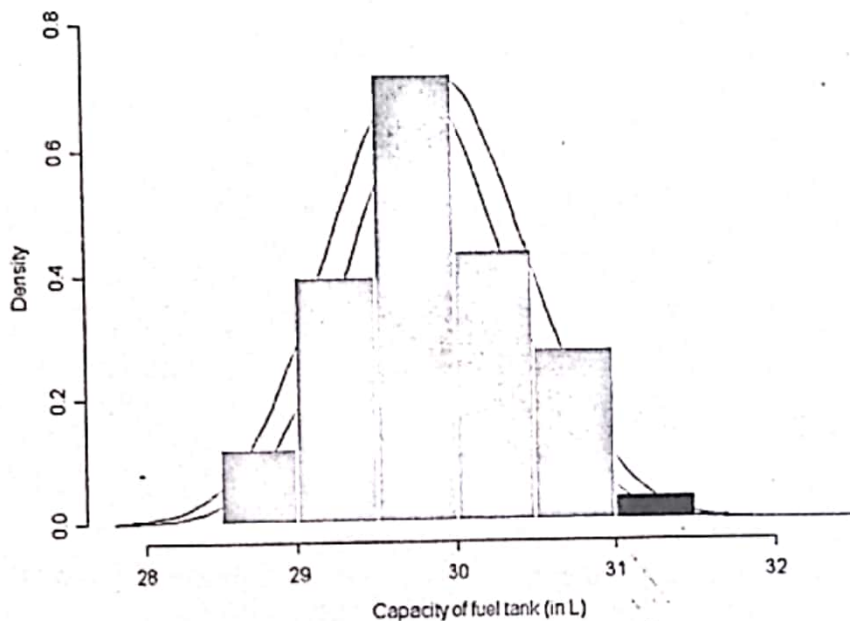


Figure 9.3:

According to the histogram, a normal distribution seems a good model. The figure shows two possible normal densities. In order to decide suitable values for μ and σ , we may use the fact that $E(X) = \mu$ and $V(X) = \sigma^2$. As suitable approximate values for μ and σ we usually use the sample mean $\bar{x}$ and the sample standard deviation s respectively. Accordingly, we may use the $N(29.9, 0.55^2)$ model for the density of X .

Example 9.3.

Consider Example 9.2. Assuming that $X \sim N(29.9, 0.55^2)$ calculate the probability that the fuel tank of a randomly selected car of the above type will hold more than 31 litres.

$$\begin{aligned}\Pr(X > 31) &= \int_{31}^{\infty} f_X(x) dx \\ &= \int_{31}^{\infty} \frac{1}{\sqrt{2\pi}0.55} e^{-\frac{1}{2}\left(\frac{x-29.9}{0.55}\right)^2} dx\end{aligned}$$

Calculation of this integration is tedious. Alternatively, one method to obtain this probability is to use a statistical package like Minitab. The following is the Minitab output for $\Pr(X \leq 31)$.

Cumulative Distribution Function

Normal with mean = 29.9 and standard deviation = 0.55

x	P(X <= x)
31	0.977250

Hence, $\Pr(X > 31) = 1 - 0.977250 = 0.02275$.

Another method is to standardize X and use a standard normal probability table to calculate the probability. This is discussed in the next session.

9.5 Calculation of Normal Probability

Theorem 9.1.

Suppose that $X \sim N(\mu, \sigma^2)$. Let $Z = \frac{X - \mu}{\sigma}$. Then, $Z \sim N(0, 1)$.

Proof

See Theorem 15.1 and Activity 15.1.

Transforming X into Z using the above transformation is called the standardization. Normal distribution with $\mu = 0$ and $\sigma = 1$ is called the standard normal distribution.

Example 9.4.

Suppose $X \sim N(29.9, 0.55^2)$. Express $\Pr(X > 31)$ in terms of Z .

$$\begin{aligned}\Pr(X > 31) &= \Pr\left(\frac{X - 29.9}{0.55} > \frac{31 - 29.9}{0.55}\right) \\ &= \Pr(Z > 2)\end{aligned}$$

Probabilities related to Z can be found by using Table 20.1 in the Appendix. This table gives probabilities $\Pr(Z > z)$ for non-negative z values containing up to 2 decimal places. z value, up to the first decimal value is read from the first column. The next decimal value is read from the first row (top row). The value at the intersection of the corresponding row and column gives $\Pr(Z > z)$. For example, to find $\Pr(Z > 2.00)$, we find the value at the

intersection corresponding to 2.0 from the first column and 0.0 from the first row (see Figure 9.4). Then, we find that $\Pr(Z > 2.00) = 0.0228$.

z	0.00	0.01
0	0.5	0.496
0.1	0.4602	0.4562
0.2	0.4207	0.4168
0.3	0.3821	0.3783
0.4	0.3446	0.3409
0.5	0.3085	0.305
0.6	0.2743	0.2709
0.7	0.242	0.2389
0.8	0.2119	0.209
0.9	0.1841	0.1814
1	0.1587	0.1562
1.1	0.1357	0.1335
1.2	0.1151	0.1131
1.3	0.0968	0.0951
1.4	0.0808	0.0793
1.5	0.0668	0.0655
1.6	0.0548	0.0537
1.7	0.0446	0.0436
1.8	0.0359	0.0351
1.9	0.0287	0.0281
2	0.0228	0.0222

Figure 9.4:

Example 9.5.

Suppose that $X \sim N(10, 4)$. Calculate $\Pr(X \geq 15)$.

$$\begin{aligned}
 P(X \geq 15) &= P\left(\frac{X - 10}{2} \geq \frac{15 - 10}{2}\right) \\
 &= P(Z \geq 2.5) \\
 &= 0.0062
 \end{aligned}$$

Example 9.6.

Suppose that $X \sim N(10, 4)$. Calculate $\Pr(X < 6)$.

$$\begin{aligned}
 P(X < 6) &= P\left(\frac{X - 10}{2} < \frac{6 - 10}{2}\right) \\
 &= P(Z < -2)
 \end{aligned}$$

Note: The distribution of Z is symmetric around 0. Hence, $\Pr(Z < -2) = \Pr(Z > 2)$. From Table 20.1, $\Pr(Z > 2) = 0.0228$.

Activity 9.1.

Calculate the probabilities given in the following table. Check the accuracy of your answers.

No.	μ	σ	Calculate this probability	Answer
1	58	10	$\Pr(75.19 < X < 82.1)$	0.0348
2	49	7	$\Pr(X > 48.52)$	0.5273
3	103	17	$\Pr(73.97 \leq X < 138.28)$	0.9371
4	99	24	$\Pr(X < 82.8)$	0.2498
5	52	10	$\Pr(X \leq 53.58)$	0.5628
6	72	8	$\Pr(70.45 \leq X \leq 93.5)$	0.5732
7	82	20	$\Pr(48.14 < X \leq 99.49)$	0.7639
8	94	15	$\Pr(91.93 \leq X < 98.55)$	0.1741
9	45	4	$\Pr(42.36 \leq X \leq 50.59)$	0.6643
10	73	1	$\Pr(X > 72.38)$	0.7324

Table 9.1:

Activity 9.2.

Suppose $X \sim N(\mu, \sigma^2)$. Find the following probabilities:

(i) $\Pr(\mu - \sigma \leq X \leq \mu + \sigma)$

(ii) $\Pr(\mu - 2\sigma \leq X \leq \mu + 2\sigma)$

(iii) $\Pr(\mu - 3\sigma \leq X \leq \mu + 3\sigma)$

Summary

In this session, we discussed about two symmetric distributions; namely, the uniform distribution and the Normal distribution. The uniform distribution is rectangular while the normal distribution is bell shaped. The uniform distribution is suitable when the probability within any interval of the same length, within the range of data, is the same. Normal distribution is probably the most often used distribution. It is symmetric around the mean. The density becomes almost zero at three standard deviations away from the mean. When the normal probabilities are calculated manually, we first standardize the distribution and use the z-table to find the probabilities.

Learning outcomes

By now, you should be able to

- Describe the Uniform distribution
- Calculate Uniform probabilities
- Describe the Normal distribution
- Calculate Normal probabilities

Session 10

CONTINUOUS SYMMETRIC DISTRIBUTIONS CONTINUED

Contents

Introduction

10.1 Empirical rule for a normal distribution

10.2 Calculation of Normal quantiles

10.3 Normal Approximations

10.4 Student's t distribution

10.5 Quantiles of the t distribution

Summary

Learning outcomes

Introduction

In the previous session, we discussed how to calculate normal probabilities. In this session, we further discuss useful properties and applications of the normal distribution. In addition, we discuss about another useful symmetric distribution called the student's t distribution.

10.1 Empirical rule for a normal distribution

Approximately 68% of the values in any normal distribution lie within one standard deviation, approximately 95% lie within two standard deviations, and approximately 99.7% lie within three standard deviations from the mean.

10.2 Calculation of Normal quantiles

z_α Notation

z_α denotes the value such that $\Pr(Z \geq z_\alpha) = \alpha$. Here, α represents a probability ($0 \leq \alpha \leq 1$). This notation is used often in formula for confidence intervals and decision-rules of tests of hypotheses. For example, a 95% confidence interval for the mean of a normal distribution with known standard deviation σ is given by $\bar{x} \pm z_{0.025} \frac{\sigma}{\sqrt{n}}$. Here $\bar{x}$ is the mean of n random observations from the normal distribution.

Example 10.1.

Find $z_{0.025}$.

$$\Pr(Z \geq 1.96) = 0.025.$$

$$\therefore z_{0.025} = 1.96.$$

See Figure 10.1.

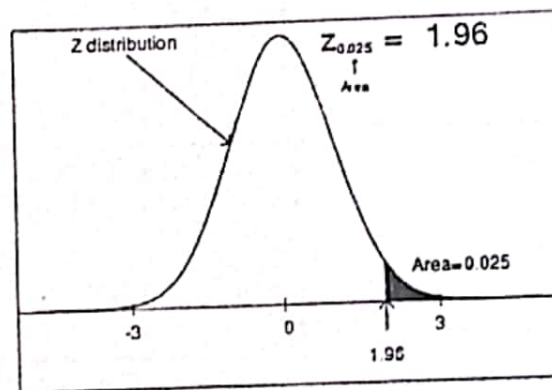


Figure 10.1:

Activity 10.1.

Find the following values:

$z_{0.01}$, $z_{0.05}$, $z_{0.9}$, $z_{0.975}$, $z_{0.85}$.

Activity 10.2.

In setting warranties on appliances, manufacturers want to set the time limit in such a way that only a few appliances will have to be repaired at the manufacture's expense, yet they want the buyer to have some degree of protection against malfunctions over a period of time after the purchase. A manufacturer of a certain appliance knows that the distribution of lifetime of appliances is normal with a mean of 38 months and a standard deviation of 9 months. The manufacturer would like to set the expiration time of the warranty at such a level that 90% of appliances made will remain in working condition throughout the warranty period. What should be the warranty period?

10.3 Normal Approximations

Normal approximation to Binomial

- Suppose $X \sim \text{Bin}(n, \theta)$
- When both $n\theta$ and $n(1 - \theta)$ are greater than 5, the $\text{Bin}(n, \theta)$ distribution can be approximated by a $N(n\theta, n\theta(1 - \theta))$ distribution. (see Figure 10.2).

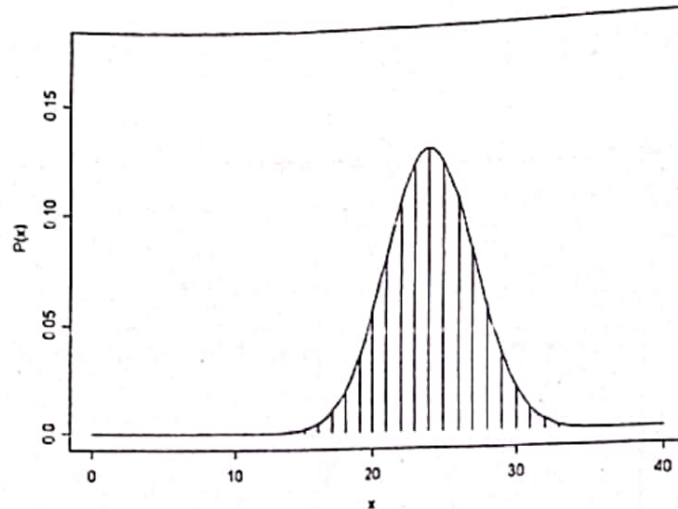


Figure 10.2: $\text{Bin}(40, 0.6)$ distribution (pmf is shown by vertical lines) and the approximate normal distribution $N(24, 9.6)$ by the solid curve.

Note that the Binomial distribution is discrete and the normal distribution is continuous. Therefore, when we use the normal distribution to calculate a Binomial probability, we should use a 'continuity correction'. Let x be an integer. The integer x in the discrete scale corresponds to the $(x - 0.5, x + 0.5)$ interval in the continuous scale. This is called the continuity correction'.

Recall that if $X \sim \text{Bin}(n, \theta)$, then, $E(X) = n\theta$ and $V(X) = n\theta(1 - \theta)$. Then, for the approximate normal distribution, $\mu = n\theta$ and $\sigma = \sqrt{n\theta(1 - \theta)}$.

$$\begin{aligned} \Pr(X = x)_{\text{Binomial}} &\approx \Pr(x - 0.5 \leq X \leq x + 0.5)_{\text{Normal}} \\ &= \Pr\left(\frac{x - 0.5 - \mu}{\sigma} \leq Z \leq \frac{x + 0.5 - \mu}{\sigma}\right), \end{aligned}$$

$$\begin{aligned} \Pr(X \leq x)_{\text{Binomial}} &\approx \Pr(X \leq x + 0.5)_{\text{Normal}} \\ &= \Pr\left(Z \leq \frac{x + 0.5 - \mu}{\sigma}\right), \end{aligned}$$

$$\begin{aligned}\Pr(X \geq x)_{\text{Binomial}} &\approx \Pr(X \geq x - 0.5)_{\text{Normal}} \\ &= \Pr\left(Z \geq \frac{x - 0.5 - \mu}{\sigma}\right)\end{aligned}$$

Activity 10.3.

Suppose X has a binomial distribution with $n = 40$ and $p = 0.6$. Find

(i) $\Pr(X \leq 20)$

(ii) $\Pr(X \geq 20)$

(iii) $\Pr(X = 30)$

(iv) $\Pr(X < 25)$

(v) $\Pr(X > 15)$

Note: $\Pr(X < 25) = \Pr(X \leq 24)$, $\Pr(X > 15) = \Pr(X \geq 16)$

Normal approximation to Poisson

- Suppose $X \sim \text{Poisson}(\lambda)$
- When λ is greater than 10, $\text{Poisson}(\lambda)$ distribution can be approximated by $N(\lambda, \lambda)$ distribution (see Figure 10.3).

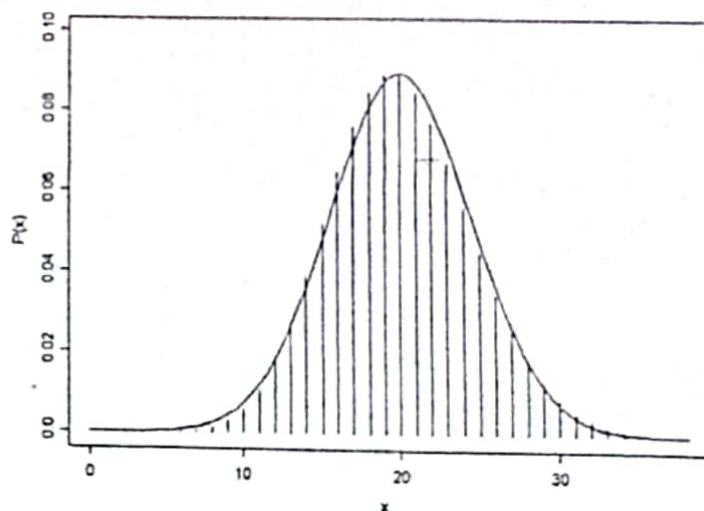


Figure 10.3: Poisson(20) distribution (pmf is shown by vertical lines) and the approximate normal distribution $N(20, 20)$.

Since the Poisson distribution is discrete and the normal distribution is continuous, the continuity correction should be applied when the probabilities are calculated. The results are similar to those for the Binomial distribution, except the fact that $\mu = \lambda$ and $\sigma = \sqrt{\lambda}$ now. For example,

$$\begin{aligned}\Pr(X \leq x)_{\text{Poisson}} &\approx \Pr(X \leq x + 0.5)_{\text{Normal}} \\ &= \Pr\left(Z \leq \frac{x + 0.5 - \mu}{\sigma}\right),\end{aligned}$$

Activity 10.4.

Suppose X has a Poisson distribution with $m = 40$. Find

- (i) $\Pr(X \leq 20)$
- (ii) $\Pr(X \geq 20)$
- (iii) $\Pr(X = 30)$
- (iv) $\Pr(X < 25)$
- (v) $\Pr(X > 15)$

There are other symmetric probability density functions. Some of the well known symmetric distributions are:

- Student's t -distribution (or simply, t -distribution)
- Logistic distribution

Out of the above distributions, the Student's t -distribution is often used for statistical calculations; especially for tests of hypotheses and confidence intervals. Therefore, we briefly describe the Student's t -distribution.

10.4 Student's t Distribution

Student's t distribution is not usually used as a model for the density of natural random variables. Therefore, calculation of probability of a t distribution does not arise often. However, the quantiles of the t distribution are often required when computing probabilities related to normally distributed data based on sample information. The t distribution arises as the distribution of $T = \frac{\sqrt{n}(\bar{X} - \mu)}{S}$, where $\bar{X}$ and S are respectively the mean and the standard deviation of a random sample of n observations from a $N(\mu, \sigma^2)$ distribution. For a partial proof of this result see Theorem 15.12 and Theorem 16.5.

A continuous random variable T is said to have a Student's t distribution with ν degrees of freedom if its density function is given by

$$f_T(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \quad -\infty < t, \infty, \nu > 0.$$

This is denoted as $T \sim t_\nu$. Usually in practical applications, ν is an integer.

The t_ν distribution is symmetric around 0, and bell shaped. It resembles the standard normal distribution. However, the tails of the t distribution are longer than those of the standard normal distribution. The t distribution converges to the standard normal distribution as ν increases. Figure 10.4 shows the density functions of t distributions with different degrees of freedoms.

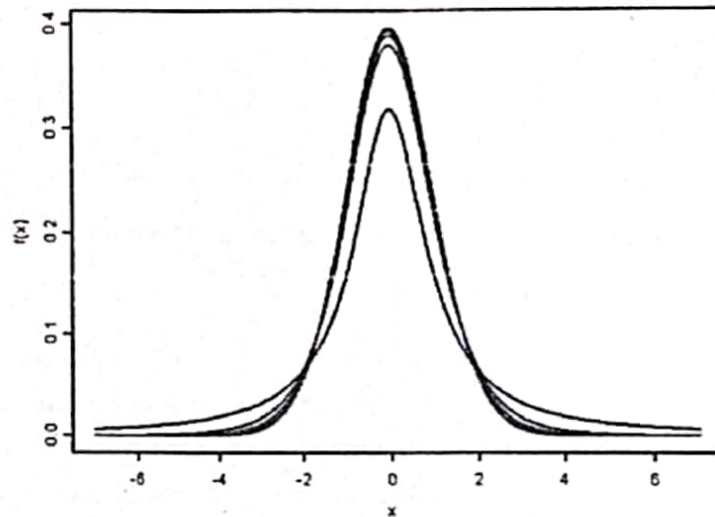


Figure 10.4: The density functions of the Student's t distribution. From innermost to outermost (at $x = 0$), the t_ν densities with $\nu = 1, 5, 10, 20, 30, 40$.

10.5 Quantiles of the t distribution

Suppose $T \sim t_\nu$. Often we need to find the quantiles of T . In particular, the upper 1, 2.5, 5, and 10 percentiles are often required for calculations. Let α be a probability. Then, $t_{\alpha, \nu}$ denotes the upper $\alpha \times 100\%$ percentile of the t_ν distribution. In other words, $\Pr(T > t_{\alpha, \nu}) = \alpha$.

The $t_{\alpha, \nu}$ values are given in Table 20.2 in Appendix. For example, $t_{0.025, 10} = 2.201$. This means, if $T \sim t_{10}$, then $\Pr(X > 2.201) = 0.025$.

Summary

In this session we discussed about the empirical property of the normal distribution. Further, we learned how to calculate the quantiles of a normal distribution and the notation z_α . The normal approximations for the Binomial and Poisson distributions are useful when we calculate

probabilities of those distributions manually. We discussed the need to apply the continuity correction when we approximate a discrete distribution by a continuous distribution.

We learned that the Student's t -distributions is useful in statistical calculations. Particularly, the quantiles of the t distribution are required for confidence intervals and tests of hypotheses. Student's t distribution approaches the standard normal distribution when the degrees of freedom increases. Therefore, often $t_{p, \nu}$ are replaced by z_p when $\nu > 30$.

Learning outcomes

By now, you should be able to

- Explain the empirical rule for the normal distribution.
- Calculate quantiles of any normal distribution.
- Apply normal model for suitable situations and calculate probabilities
- Calculate Binomial and Poisson probabilities using normal approximations when suitable
- Calculate t quantiles

Session 11

POSITIVELY SKEWED DISTRIBUTIONS

Contents

Introduction

11.1 Gamma distribution

11.2 Mean and variance of a Gamma distribution

11.3 Calculation of Gamma probabilities

11.4 Normal approximation to Gamma distribution

Summary

Learning outcomes

Introduction

There are many practical situations in which the variable of interest has a skewed distribution. A family of pdf's that yield a wide variety of positively skewed distributions is the gamma family. The Chi-square distribution and the exponential distribution are two members of this family.

11.1 Gamma Distribution

A continuous random variable X is said to have a gamma distribution with shape parameter α and scale parameter λ if its probability density function is given by

$$f_X(x; \alpha, \lambda) = \frac{\lambda^\alpha e^{-\lambda x} x^{\alpha-1}}{\Gamma(\alpha)}; \quad x > 0.$$

This is denoted as $X \sim \text{Gamma}(\alpha, \lambda)$. Here, $\Gamma(\cdot)$ is the gamma function defined as

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

It has the property $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$. When α is an integer, $\Gamma(\alpha) = (\alpha - 1)!$. In addition, $\Gamma(1/2) = \sqrt{\pi}$ is a useful fact.

Figure 11.1 illustrates the pdf's of gamma distributions for several pairs of (α, λ) . It is clear that by changing the values of α and λ we can obtain a variety of skewed curves.

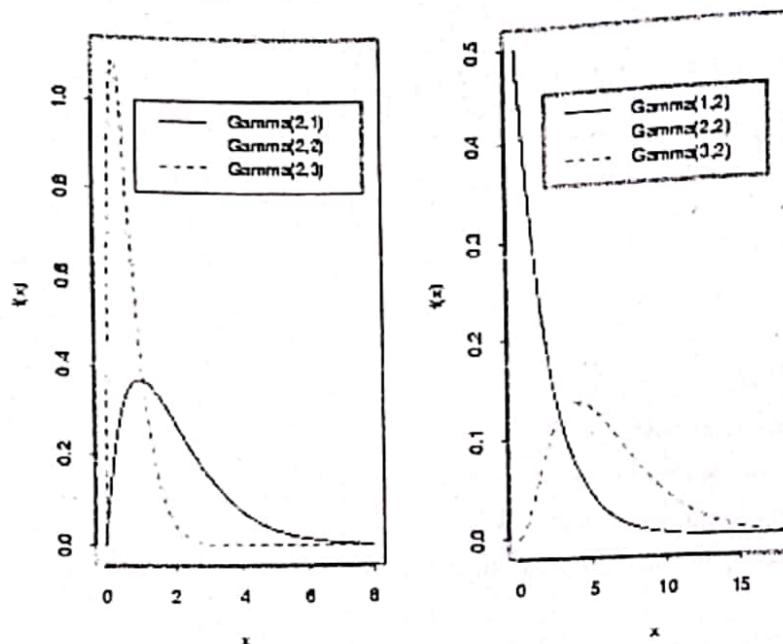


Figure 11.1: Gamma density functions

Example 11.1.

The following figure is the histogram of the Minitab data set named 'Tiles.MTW'. These data are the warping in 100 floor tiles (a data set used in Process Capability examples in on-line Help). Let X be the warping of a randomly selected tile. According to the histogram, the Gamma distribution seems a possible model for the density of X .

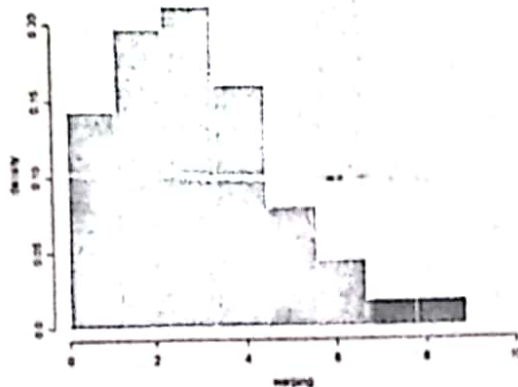


Figure 11.2: Histogram of warping data

11.2 Mean and Variance of a Gamma Distribution

Theorem 11.1.

Let $X \sim \text{Gamma}(\alpha, \lambda)$. Then,

$$E(X) = \alpha/\lambda$$

$$V(X) = \alpha/\lambda^2$$

Proof

$$\begin{aligned} E(X) &= \int_0^{\infty} x \frac{\lambda^{\alpha} e^{-x\lambda} x^{\alpha-1}}{\Gamma(\alpha)} dx \\ &= \frac{\lambda^{\alpha}}{\Gamma(\alpha)} \int_0^{\infty} e^{-x\lambda} x^{\alpha} dx \end{aligned}$$

Notice that

$$\begin{aligned} \int_0^{\infty} \frac{\lambda^{\alpha} e^{-x\lambda} x^{\alpha-1}}{\Gamma(\alpha)} dx &= 1 \\ \therefore \int_0^{\infty} e^{-x\lambda} x^{\alpha-1} dx &= \frac{\Gamma(\alpha)}{\lambda^{\alpha}} \end{aligned}$$

Analogously,

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

Hence

$$E(X) = \frac{\lambda^\alpha \Gamma(\alpha+1)}{\Gamma(\alpha) \lambda^{\alpha+1}} = \frac{\alpha}{\lambda}$$

$$\begin{aligned} E(X^2) &= \int_0^\infty x^2 \frac{\lambda^\alpha e^{-x\lambda} x^{\alpha-1}}{\Gamma(\alpha)} dx \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty e^{-x\lambda} x^{\alpha+1} dx \\ &= \frac{\lambda^\alpha \Gamma(\alpha+2)}{\Gamma(\alpha) \lambda^{\alpha+2}} = \frac{\alpha(\alpha+1)}{\lambda^2} \end{aligned}$$

Now,

$$\begin{aligned} V(X) &= E(X^2) - [E(X)]^2 \\ &= \frac{\alpha(\alpha+1)}{\lambda^2} - \frac{\alpha^2}{\lambda^2} = \frac{\alpha}{\lambda^2} \end{aligned}$$

There are several methods to find suitable values for α and λ of a gamma model. One method is called the method of moments. In this method, we use the idea that the sample mean and sample variance should be approximately equal to the population mean and population variance respectively.

So, we solve the following equations to find suitable values for α and λ .

$$\begin{aligned} \alpha/\lambda &\approx \bar{x} \\ \alpha/\lambda^2 &\approx s^2 \end{aligned}$$

Example 11.2.

Consider Example 11.1. The sample mean and sample variance of the warping data are 2.92 and 3.19 respectively. Fit a suitable gamma distribution.

$$\begin{aligned} \alpha/\lambda &\approx 2.92 \\ \alpha/\lambda^2 &\approx 3.19 \end{aligned}$$

These give

$$\begin{aligned} \lambda &\approx 0.92 \\ \alpha &\approx 2.69 \end{aligned}$$

The fitted Gamma (2.69, 0.92) density is shown in figure 11.3. The fit does not seem to be that good. Better estimates may be obtained using other methods like the method of maximum likelihood. However, those methods are beyond the scope of this course.

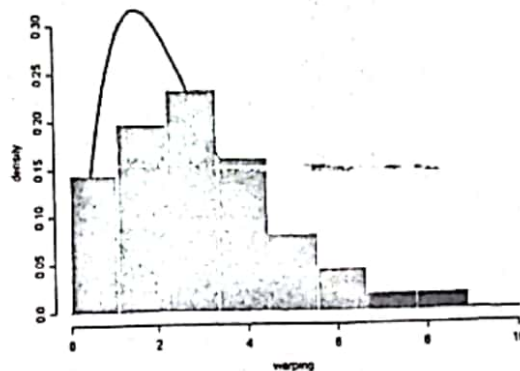


Figure 11.3: Histogram of warping data and the fitted Gamma distribution

11.3 Calculating Gamma probabilities

Suppose $X \sim \text{Gamma}(\alpha, \lambda)$ and we want to calculate $\Pr(X \leq x)$. We can calculate this by integrating the gamma density function from 0 to x . However, this integration is tedious to do manually. The following Theorem and the 'incomplete gamma table' is useful instead.

Theorem 11.2.

Suppose that $X \sim \text{Gamma}(\alpha, \lambda)$. Then, $\lambda X \sim \text{Gamma}(\alpha, 1)$.

For the proof, see Theorem 15.3 and Activity 15.2.

Standard gamma distribution and Incomplete gamma function

A $\text{Gamma}(\alpha, 1)$ distribution is called a standard gamma distribution with shape parameter α .

Suppose $W \sim \text{Gamma}(\alpha, 1)$. Then, the function γ defined as

$$\gamma(w; \alpha) = \Pr(W \leq w) = \int_0^w \frac{y^{\alpha-1} e^{-y}}{\Gamma(\alpha)} dy$$

is called the incomplete gamma function. The values of $\gamma(w, \alpha)$ have been tabulated for different values of w and α . Tables 20.3 and 20.4 in Appendix provide those values. The following example shows how to use Theorem 11.2 and the incomplete gamma table to calculate gamma probabilities.

Example 11.3.

Suppose that $X \sim \text{Gamma}(2, 4)$. Calculate $\Pr(X \geq 0.5)$.

$$\begin{aligned}\Pr(X \geq 0.5) &= 1 - \Pr(X \leq 0.5) \\ &= 1 - \Pr(4X \leq 2)\end{aligned}$$

Since $W = 4X \sim \text{Gamma}(2, 1)$, from Table 20.3, $\Pr(W \leq 2) = 0.59399$. Hence,

$$\Pr(X \geq 0.5) = 1 - 0.59399 = 0.40601$$

Activity 11.1.

Claim amounts (in Rs. 1000's) of a certain insurance plan are modeled by $\text{Gamma}(5, 1/2)$. What is the probability that a randomly selected claim is more than Rs. 10000?

11.4 Normal approximation to Gamma distribution

The $\text{Gamma}(\alpha, \lambda)$ distribution converges to the $N(\alpha/\lambda, \alpha/\lambda^2)$ as α increases. This approximation can be used when $\alpha > 10$. Figure 11.4 illustrates this phenomenon.

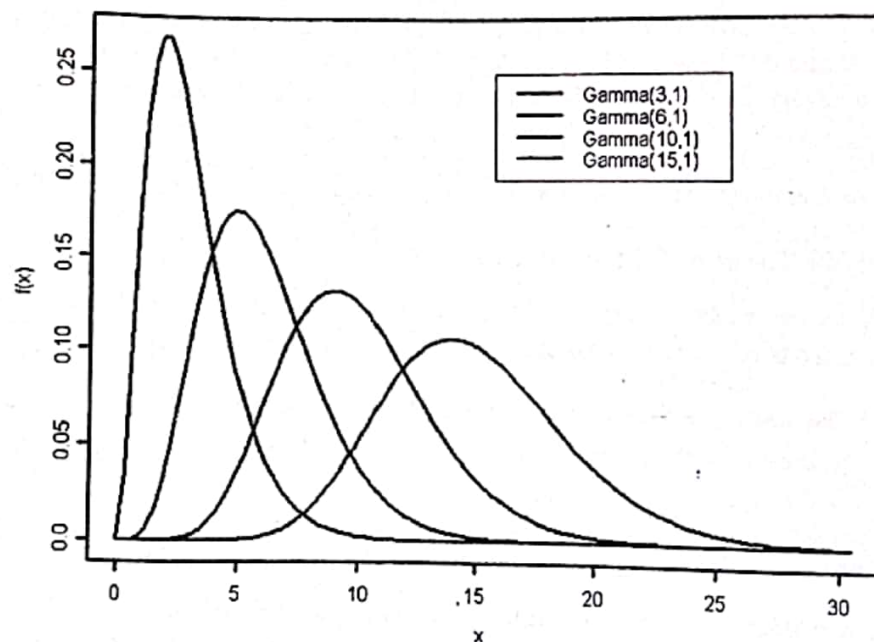


Figure 11.4: Convergence of gamma distribution to normal distribution as α increases.

Summary

In this session, we discussed about a very useful positively skewed distribution called the gamma distribution. We discussed how to use the tables of incomplete gamma function to calculate gamma probabilities manually. We also learned that the gamma distribution converges to a normal distribution as the value of the shape parameter increases.

Learning Outcomes

By now, you should be able to

- Identify situations for which the Gamma model is suitable
- Calculate Gamma probabilities using the tables of incomplete gamma function
- Calculate Gamma probabilities using the normal approximation when suitable.

Session 12

POSITIVELY SKEWED DISTRIBUTIONS CONTINUED

Contents

Introduction

12.1 χ^2 distribution

12.2 Mean and variance of χ^2 distribution

12.3 Quantiles of χ^2 distribution

12.4 Exponential Distribution

12.5 Mean and variance of exponential distribution

12.6 F distribution

12.7 Quantiles of F distribution

Summary

Learning outcomes

Introduction

In this session, we discuss about two particular members of the Gamma family. Namely, the chi-square distribution and the exponential distribution. In addition, we discuss about another useful distribution called the F distribution.

12.1 χ^2 Distribution

A gamma distribution with $\alpha = \nu/2$ and $\lambda = 1/2$ is called a 'Chi-square' distribution (χ^2 distribution) with ν 'degree of freedom'.

If random variable X has a χ^2 distribution with ν degrees of freedom, it is denoted by $X \sim \chi_\nu^2$.

12.2 Mean and Variance of χ^2 Distribution

Theorem 12.1.

Suppose $X \sim \chi_\nu^2$. Then,

$$E(X) = \nu$$

$$V(X) = 2\nu$$

Activity 12.1.

Prove Theorem 12.1 using Theorem 11.1.

χ^2 distribution arises often in theoretical statistical calculations. The following are such situations:

- If S^2 is the sample variance of n random observations from a normal distribution with variance σ^2 , the distribution of $(n-1)S^2/\sigma^2 \sim \chi_{n-1}^2$.
- If $Z \sim N(0, 1)$, then $Z^2 \sim \chi_1^2$.
- Suppose n is an integer. If $X \sim \text{Gamma}(n, \lambda)$, then $2\lambda X \sim \chi_{2n}^2$.

12.3 Quantiles of χ^2 Distribution

The quantiles of the χ^2 distribution are required often for various statistical calculations. Let α be a probability. Then $\chi_{\alpha, \nu}^2$ denotes the upper $\alpha(100)\%$ quantile of a χ_ν^2 distribution. In other words, if $X \sim \chi_\nu^2$, then $\Pr(X > \chi_{\alpha, \nu}^2) = \alpha$. Table 20.5 in the Appendix gives the χ^2 quantiles.

12.4 Exponential Distribution

The family of exponential distributions provides probability models that are very widely used in engineering and science disciplines. This distribution is often used to model lifetimes of various items. When the number of events in a time interval has a Poisson distribution, the length of time interval between successive events can be modelled by an exponential distribution.

Random variable X is said to have an exponential distribution with rate parameter $\lambda(> 0)$ if its probability density function is given by

$$f_X(x) = \lambda e^{-\lambda x}; \quad x > 0.$$

This is denoted by $X \sim \exp(\lambda)$. Note that exponential distribution is a special case of the gamma distribution. It is easy to see that $X \sim \exp(\lambda) \Leftrightarrow X \sim \text{Gamma}(1, \lambda)$

Figure 12.1 shows exponential density functions for several values of λ .

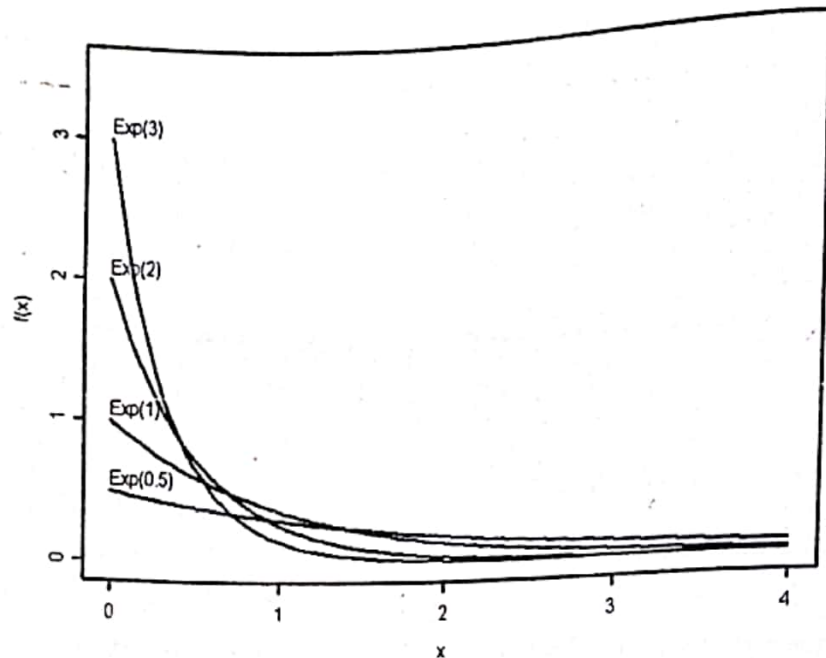


Figure 12.1: Exponential density functions

Example 12.1. Consider Example 2.2.

The histogram has the shape of an exponential distribution.

12.5 Mean and variance of Exponential distribution

Theorem 12.1.

Let $X \sim \exp(\lambda)$. Then,

$$E(X) = 1/\lambda$$

$$V(X) = 1/\lambda^2$$

A rough check to see whether an exponential distribution is suitable for a random variable is to check whether the sample mean and the sample standard deviations are roughly the same (population mean and population standard deviation should be exactly the same). For the accident data (Example 2.2), the sample mean and the sample standard deviations are 40.7 and 43.9 respectively. This is a good indication that an exponential model may be suitable.

In order to find a suitable value for λ , one may use one of the following two approximations. These approximations are based on the assumptions that the sample mean and the sample standard deviation are close to the population mean and population standard deviation respectively.

$$\frac{1}{\lambda} \approx \bar{x} \text{ or } \frac{1}{\lambda} \approx s.$$

Example 12.2.

Consider Example 2.2. Fit a suitable exponential distribution. Using the fitted distribution, calculate the probability that the time between two accidents is more than 60 days.

For the accident data, the above approximations give

$$\frac{1}{\lambda} \approx 40.7 \text{ or } \frac{1}{\lambda} \approx 43.9.$$

These give $\lambda \approx 0.025$ or $\lambda \approx 0.023$. We decided to take $\lambda = 0.024$ (the middle value of the two). Figure 12.2 shows the fitted density function. It seems a very good fit.

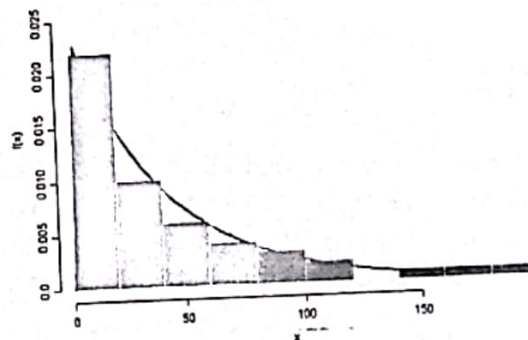


Figure 12.2: Fitted density function for the accident data

Then, the density function of X is

$$f_X(x) = 0.24e^{-0.24x}; \quad x > 0.$$

Hence,

$$\begin{aligned} \Pr(X > 60) &= \int_{60}^{\infty} 0.24e^{-0.24x} dx \\ &\approx 0.24 \end{aligned}$$

Activity 12.2.

A company that produces a certain pain killer claims that the time to relieve headache after taking a tablet of their pain killer is exponentially distributed with $\lambda = 0.1$ minutes.

1. What is the median time for relief from headache?
2. What is the probability that a randomly selected patient with headache will get relief within 10 minutes after taking a tablet?
3. If 5 patients with headache are given a tablet each, what is the probability that at least 4 patients will get relief within 10 minutes?

The following are few other commonly used positively skewed distributions:

- Log-normal distribution
- Pareto distribution
- Weibull distribution
- F distribution

The first three distributions above are commonly used to model distributions with long right tails. For example, they are often used to model the insurance claims. The F distribution, is not often used as a model for the distribution of a natural variable. However, its quantiles are used often for Statistical tests. Therefore, we briefly describe the F distribution below.

12.6 F Distribution

F distribution arises from the ratio of two independent χ^2 distributions, each divided by the corresponding degrees of freedom. Suppose that $U_1 \sim \chi^2_{\nu_1}$ and $U_2 \sim \chi^2_{\nu_2}$ are two independent random variables. Then, the distribution of $W = \frac{U_1/\nu_1}{U_2/\nu_2}$ is called the F distribution with ν_1 numerator degree of freedom and ν_2 denominator degree of freedom. It is denoted by $W \sim F_{\nu_1, \nu_2}$.

12.7 Quantiles of F distribution

F_{α, ν_1, ν_2} Notation

Let α be a probability. Suppose $W \sim F_{\nu_1, \nu_2}$. Then, the upper $\alpha(100)\%$ percentile of F_{ν_1, ν_2} is denoted by F_{α, ν_1, ν_2} .

$$\Pr(W > F_{\alpha, \nu_1, \nu_2}) = \alpha.$$

Tables 20.6-20.13 in the Appendix provide F_{α, ν_1, ν_2} values. Unlike z , t and χ^2 tables, there is a separate F table for each α value, of a set of selected α values.

Example 13.1.

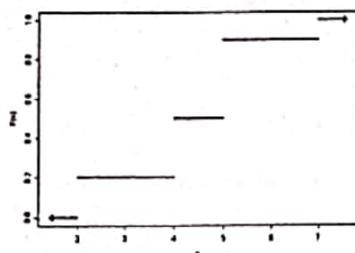
Consider pmf given below.

$$P_X(x) = \begin{cases} 0.2 & \text{if } x = 2 \\ 0.3 & \text{if } x = 4 \\ 0.4 & \text{if } x = 5 \\ 0.1 & \text{if } x = 7 \end{cases}$$

Find the cdf of X . Draw the cdf of X .

Solution 13.1

$$F_X(x) = \begin{cases} 0 & \text{if } x < 2 \\ 0.2 & \text{if } 2 \leq x < 4 \\ 0.5 & \text{if } 4 \leq x < 5 \\ 0.9 & \text{if } 5 \leq x < 7 \\ 1 & \text{if } x \geq 7 \end{cases}$$



The cdf of a discrete random variable is a non decreasing step function.

Example 13.2.

$$f_W(w) = \begin{cases} \frac{1}{25}w & 0 \leq w < 5 \\ \frac{2}{5} - \frac{1}{25}w & 5 \leq w \leq 10 \\ 0 & \text{otherwise} \end{cases}$$

Find the cdf of W .

Solution 13.2

$$F_W(w) = \begin{cases} 0 & \text{if } w \leq 0 \\ F_W(0) + \int_0^w \frac{1}{25}x dx & \text{if } 0 < w \leq 5 \\ F_W(5) + \int_5^w \left(\frac{2}{5} - \frac{1}{25}x\right) dx & \text{if } 5 < w \leq 10 \\ F_W(10) & \text{if } w > 10 \end{cases}$$

Activity 13.1.

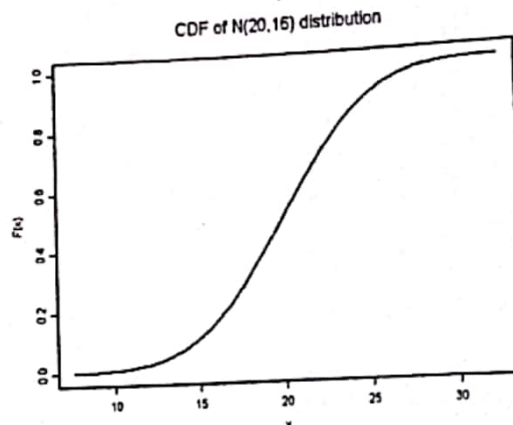
Let X be the distance between a point target and a shot aimed at the point in a coin operated target game. X has the following density function:

$$f_X(x) = \frac{3}{4}(1-x^2), \quad -1 \leq x \leq 1$$

Find the cdf of X and draw it. Compute $\Pr(-0.5 < X \leq 0.7)$ using the cdf.

Activity 13.2. Suppose that $X \sim N(20, 16)$. Find $F_X(18)$ and $F_X(25)$.

The cdf of a continuous random variable is a non decreasing continuous function. The following Figure shows the cdf of the $N(20, 16)$ distribution.



Activity 13.3. A random variable X has the following pdf.

$$f_X(x) = 0.15e^{-0.15(x-0.5)}, \quad x \geq 0.5$$

Find the cdf of X .

For example, from Table 20.8, $F_{0.05,5,6} = 4.387$.

The following is an important formula when calculating the quantiles using F tables:

$$F_{1-\alpha, \nu_1, \nu_2} = \frac{1}{F_{\alpha, \nu_2, \nu_1}}$$

For example,

$$F_{0.95,6,5} = F_{1-0.05,6,5} = \frac{1}{F_{0.05,5,6}} = \frac{1}{4.387} = 0.2279$$

Summary

In this session we discussed about the Chi-square distribution and the Exponential distribution. Both are particular cases of the Gamma distribution. The Chi-square distribution is particularly useful in statistical calculations like confidence intervals and test of hypotheses. We discussed how to calculate the quantiles of the Chi-square distribution and their notation. In addition, we discussed how to find the quantiles of the F distribution.

Learning Outcomes

By now, you should be able to

- Find chi-square quantiles and explain the notation $\chi^2_{\alpha, \nu}$.
- Apply exponential distribution to suitable situations and calculate probabilities
- Find F quantiles and explain the notation F_{α, ν_1, ν_2} .

Session 13

CUMULATIVE DISTRIBUTION FUNCTION

Contents

Introduction

13.1 Definition

13.2 Relationship between cdf and pdf

13.3 CDF technique to derive distributions

Summary

Learning outcomes

Introduction

In this session, we discuss another very useful function that describes the distribution of probabilities of a random variable. It is called the cumulative distribution function (cdf). In addition to its use in calculation of probability, it is useful for determining the distribution of functions of random variables as well.

13.1 Definition

The cumulative distribution function (c.d.f.) of a random variable X (regardless of discrete or continuous) is denoted by F_X . F_X is defined as

$$F_X(x) = \Pr(X \leq x), x \in \mathbb{R}$$

1.2 Relationship between cdf and pdf

Let X be a continuous random variable. Then,

$$F_X(x) = \int_{-\infty}^x f_X(x) dx$$

Hence,

$$\frac{dF_X(x)}{dx} = f_X(x)$$

3.3 CDF technique to derive distributions

Example 13.3. Let $X \sim N(0,1)$. Find the distribution of $Y = X^2$.

Solution 13.3

$$\begin{aligned} F_Y(y) &= \Pr(Y \leq y) \\ &= \Pr(X^2 \leq y) \\ &= \Pr(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= \int_{-\sqrt{y}}^{\sqrt{y}} f_X(x) dx \\ &= \int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx \\ &= \int_{-\sqrt{y}}^0 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx + \int_0^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx \\ &= 2 \int_0^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du \end{aligned}$$

Now let $z = u^2$. Then

$$\begin{aligned} F_Y(y) &= \frac{2}{\sqrt{2\pi}} \int_0^y e^{-\frac{1}{2}z} \frac{dz}{2\sqrt{z}} \\ &= \int_0^y \frac{e^{-\frac{1}{2}z} z^{-\frac{1}{2}}}{\sqrt{2\pi}} dz \\ &= \int_0^y \frac{e^{-\frac{1}{2}z} z^{-\frac{1}{2}}}{\Gamma(\frac{1}{2})\sqrt{2}} dz \end{aligned}$$

Now recall that the integrand is the gamma density with $\alpha = \frac{1}{2}$ and $\lambda = 1/2$. This implies that Y has a χ^2 distribution with 1 degree of freedom. The density function of Y is,

$$f_Y(y) = \frac{e^{-\frac{1}{2}y} y^{-\frac{1}{2}}}{\Gamma(\frac{1}{2})\sqrt{2}}$$

Example 13.4. Let $X \sim U(0,1)$. Find the distribution of $Y = -\ln X$.

Solution 13.4

$$\begin{aligned}
 F_Y(y) &= \Pr(Y \leq y) \\
 &= \Pr(-\ln X \leq y) \\
 &= \Pr(\ln X \geq -y) \\
 &= \Pr(X \geq e^{-y}) \\
 &= 1 - \Pr(X < e^{-y}) \\
 &= 1 - \int_0^{e^{-y}} f_X(x) dx \\
 &= 1 - \int_0^{e^{-y}} 1 dx \\
 &= 1 - e^{-y}
 \end{aligned}$$

The density function of Y can be found by differentiating $F_Y(y)$ with respect to y .

$$\begin{aligned}
 f_Y(y) &= \frac{d}{dy} F_Y(y) \\
 &= e^{-y}
 \end{aligned}$$

This is the density function of an exponential random variable with mean 1.

Example 13.5 (Inter-arrival Time Distribution for a Poisson Process). Consider the Poisson process $\{N(t)\}$ with rate λ per unit time. Let

$T =$ the time between two consecutive events

Find the distributions of T .

Solution 13.5

Let $t > 0$.

$$F_T(t) = 1 - \Pr(T > t)$$

$$\begin{aligned}
 \Pr(T > t) &= \Pr(\text{the number of events within } t \text{ period of time is zero}) \\
 &= \frac{e^{-\lambda t} (\lambda t)^0}{0!} \\
 &= e^{-\lambda t}
 \end{aligned}$$

So,

$$F_T(t) = 1 - e^{-\lambda t}.$$

Hence,

$$\begin{aligned} f_T(t) &= \frac{\partial F_T(t)}{\partial t} \\ &= \lambda e^{-\lambda t}. \end{aligned}$$

This implies that T has an exponential distribution with mean $1/\lambda$.

Summary

In this session we discussed about the cumulative distribution function. In addition to its use for calculating probabilities, it is useful for deriving the distributions of functions of variables.

Learning Outcomes

By now, you should be able to

- Derive the cdf of random variables
- Derive distributions of functions of variables using cdf technique

Session 14

GENERATING FUNCTIONS

Contents

Introduction

14.1 Moment generating function

14.2 Properties of moment generating function

14.3 Cumulant generating function

14.4 Probability generating function

Summary

Learning outcomes

Introduction

So far we have considered only two descriptive measures of a random variable, namely, the mean and the variance. The mean and the variance of a random variable contain important information about the distribution function of that variable. However, the mean and the variance do not contain all the information about the distribution. It is possible to give examples of different distribution functions that have the same mean and the same variance. The following are two such distributions. (Source: www.dartmouth.edu/~chance/teaching_aids/.../Chapter10.pdf)

Suppose X and Y are random variables, with distributions

x	1	2	3	4	5	6
$P_X(x)$	0	0.25	0.5	0	0	0.25

y	1	2	3	4	5	6
$P_Y(y)$	0.25	0	0	0.5	0.25	0

It is easy to see that $E(X) = E(Y) = 7/2$ and $V(X) = V(Y) = 9/4$. However, the pmf's P_X and P_Y are quite different.

Then, is there any measure that describes the distribution of a random variable completely? The answer is yes. The set of 'moments' that we describe in this session completely describes a distribution, when they exist.

In this session, you learn about a function that generates all the moments of a random variable, namely the moment generating function (mgf). In addition you learn about two other functions that are related to the mgf and are useful in studying about distribution functions.

14.1 Moment Generating Function

Definition 14.1.

- Let X be a random variable and x_0 be a fixed value. Let $k > 0$ be an integer. Then,

$$E[(X - x_0)^k]$$

is called the ' k th moment of X around x_0 '.

- In particular, the k th moment around 0 (central moment) is denoted by μ_k

$$\mu_k = E[X^k]$$

Recall that the variance of a random variable is defined as

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

So, the variance is the second moment around the mean.

Example 14.1.

Consider the above two random variables X and Y which have equal means and equal variances. Derive the first 3 central moments of X and Y .

Solution 14.1

The first 3 central moments of X are,

$$\begin{aligned}\mu_1 = E(X) &= 1(0) + 2(0.25) + 3(0.5) + 4(0) + 5(0) + 6(0.25) = 3.5, \\ \mu_2 = E(X^2) &= 1^2(0) + 2^2(0.25) + 3^2(0.5) + 4^2(0) + 5^2(0) + 6^2(0.25) = 14.5, \text{ and} \\ \mu_3 = E(X^3) &= 1^3(0) + 2^3(0.25) + 3^3(0.5) + 4^3(0) + 5^3(0) + 6^3(0.25) = 69.5\end{aligned}$$

The first 3 central moments of Y are,

$$\begin{aligned}\mu_1 = E(Y) &= 1(0.25) + 2(0) + 3(0) + 4(0.5) + 5(0.25) + 6(0) = 3.5, \\ \mu_2 = E(Y^2) &= 1^2(0.25) + 2^2(0) + 3^2(0) + 4^2(0.5) + 5^2(0.25) + 6^2(0) = 14.5, \text{ and} \\ \mu_3 = E(Y^3) &= 1^3(0.25) + 2^3(0) + 3^3(0) + 4^3(0.5) + 5^3(0.25) + 6^3(0) = 63.5\end{aligned}$$

Note: Even though the two random variables have the same first moment and same second moment, the third moments are different.

Example 14.2.

Let X be a random variable with the pdf

$$f_X(x) = 30x^2(1-x)^2, \quad 0 < x < 1$$

Calculate the first 3 central moments of X .

Solution 14.2

$$\begin{aligned}\mu_1 &= E(X) = \int_0^1 x f_X(x) dx = \int_0^1 30x^3(1-x)^2 dx = 1/2 \\ \mu_2 &= E(X^2) = \int_0^1 x^2 f_X(x) dx = \int_0^1 30x^4(1-x)^2 dx = 2/7 \\ \mu_3 &= E(X^3) = \int_0^1 x^3 f_X(x) dx = \int_0^1 30x^5(1-x)^2 dx = 5/28\end{aligned}$$

Activity 14.1.

Consider the pmf given below.

$$P_X(x) = \begin{cases} 0.2 & \text{if } x = 2 \\ 0.3 & \text{if } x = 4 \\ 0.4 & \text{if } x = 5 \\ 0.1 & \text{if } x = 7 \end{cases}$$

Calculate μ_k for $k = 1, 2, 3$

Activity 14.2.

Suppose X has the following density function:

$$f_X(x) = \theta e^{-\theta x}; x > 0, \theta > 0.$$

Calculate μ_k for $k = 1, 2, 3$.

Definition 14.2.

Let X be a random variable and $t \in \mathbb{R}$. The moment generating function (mgf) of X , denoted by M_X , is defined as

$$M_X(t) = E(e^{Xt}),$$

when it exists.

The mgf exists for most random variables. However there are random variables for which the moments do not exist. Since such variables are rare in practice, we do not discuss them here.

Example 14.3.

Consider the pmf given in Activity 14.1. derive the mgf.

Solution 14.3

$$\begin{aligned}M_X(t) &= E(e^{Xt}) \\&= e^{2t}(0.2) + e^{4t}(0.3) + e^{5t}(0.4) + e^{7t}(0.1) \\&= 0.2e^{2t} + 0.3e^{4t} + 0.4e^{5t} + 0.1e^{7t}\end{aligned}$$

Example 14.4.

Consider the discrete random variable with pmf

$$P_X(x) = \frac{4}{5^{x+1}}, \quad x = 0, 1, 2, \dots$$

Derive the mgf.

Solution 14.4

$$\begin{aligned}M_X(t) &= E(e^{Xt}) \\&= \sum_{x=0}^{\infty} e^{xt} \frac{4}{5^{x+1}} \\&= \frac{4}{5} \sum_{x=0}^{\infty} \left(\frac{e^t}{5}\right)^x \\&= \frac{4}{5} \left(\frac{1}{1 - \left(\frac{e^t}{5}\right)} \right) \\&= \frac{4}{5 - e^t}, \quad e^t < 5.\end{aligned}$$

Example 14.5.

Suppose $X \sim \exp(\lambda)$. Derive the mgf.

Solution 14.5

$$\begin{aligned}M_X(t) &= E(e^{Xt}) \\&= \int_0^{\infty} e^{xt} \lambda e^{-\lambda x} dx \\&= \int_0^{\infty} \lambda e^{-(\lambda-t)x} dx \\&= \lambda \int_0^{\infty} e^{-(\lambda-t)x} dx \\&= \lambda \int_0^{\infty} e^{-(\lambda-t)x} dx \\&= \frac{\lambda}{\lambda - t}, \quad t < \lambda\end{aligned}$$

Activity 14.3.

Let $X \sim \text{Bernoulli}(p)$. Show that $M_X(t) = pe^t + q$, where $q = 1 - p$.

The Tables 20.14 and 20.15 in the Appendix present the moment generating functions of some of the commonly used distributions.

Activity 14.4.

Derive the moment generating functions given in Tables 20.14 and 20.15 in the Appendix.

The moment generating function, when exists, generates the moments $\mu_k, k = 1, 2, 3, \dots$. To see this, note that

$$\begin{aligned} M_X(t) &= E(e^{Xt}) \\ &= E\left(1 + Xt + \frac{X^2 t^2}{2} + \frac{X^3 t^3}{3!} + \frac{X^4 t^4}{4!} + \dots\right) \quad (\text{Taylor expansion}) \\ &= 1 + E(X)t + \frac{E(X^2)t^2}{2} + \frac{E(X^3)t^3}{3!} + \frac{E(X^4)t^4}{4!} + \dots \\ &= 1 + \mu_1 t + \frac{\mu_2 t^2}{2} + \frac{\mu_3 t^3}{3!} + \frac{\mu_4 t^4}{4!} + \dots \end{aligned}$$

Now, note that

$$\begin{aligned} \left. \frac{dM_X(t)}{dt} \right|_{t=0} &= \mu_1 \\ \left. \frac{d^2 M_X(t)}{dt^2} \right|_{t=0} &= \mu_2 \\ \left. \frac{d^3 M_X(t)}{dt^3} \right|_{t=0} &= \mu_3 \\ \left. \frac{d^4 M_X(t)}{dt^4} \right|_{t=0} &= \mu_4 \\ &\vdots \end{aligned}$$

Example 14.6.

Consider Example 14.4. Using the moment generating function derive $E(X)$ and $E(X^2)$.

Solution 14.6

$$\begin{aligned} M_X(t) &= \frac{\lambda}{\lambda - t}, \quad t < \lambda \\ \frac{dM_X(t)}{dt} &= \frac{\lambda}{-(\lambda - t)^2} \\ \left. \frac{dM_X(t)}{dt} \right|_{t=0} &= \frac{1}{\lambda} \\ \therefore E(X) &= \frac{1}{\lambda} \end{aligned}$$

$$\begin{aligned}\frac{d^2 M_X(t)}{dt^2} &= \frac{2\lambda}{(\lambda - t)^3} \\ \left. \frac{d^2 M_X(t)}{dt^2} \right|_{t=0} &= \frac{2}{\lambda^2} \\ \therefore, E(X^2) &= \frac{2}{\lambda^2}\end{aligned}$$

Activity 14.5.

A certain random variable X has the moment generating function

$$M_X(t) = \frac{1}{(1 - 10t)^2}$$

Derive $E(X)$ and $E(X^2)$.

Activity 14.6.

Using the moment generating functions given in Tables 20.14 and 20.15, derive $E(X)$ and $V(X)$.

14.2 Properties of moment generating function

1. The moment generating function of a random variable X , when exists, is unique. (There is a unique moment generating function for a distribution). Sometimes, this uniqueness property can be used to identify the distribution of a function of a random variable.
2. Suppose a and b are two constants and X is a random variable. Then,

$$M_{aX+b}(t) = e^{bt} M_X(at)$$

Proof

$$\begin{aligned}M_{aX+b}(t) &= E(e^{(aX+b)t}) \\ &= E(e^{bt} e^{X(at)}) \\ &= e^{bt} E(e^{X(at)}) \\ &= e^{bt} M_X(at)\end{aligned}$$

3. Suppose a and b are two constants and X_1 and X_2 are independent random variables. Then,

$$M_{aX_1+bX_2}(t) = M_{X_1}(at) M_{X_2}(bt)$$

Proof

$$\begin{aligned}
 M_{aX_1+bX_2}(t) &= E(e^{(aX_1+bX_2)t}) \\
 &= E(e^{X_1(at)}e^{X_2(bt)}) \\
 &= E(e^{X_1(at)})E(e^{X_2(bt)}), \text{ Since } X_1 \text{ and } X_2 \text{ are independent} \\
 &= M_{X_1}(at)M_{X_2}(bt)
 \end{aligned}$$

4. Suppose $X_1, X_2, \dots, X_n$ is a random sample (iid) from a distribution with mgf M_X . Let $Y = X_1 + X_2 + \dots + X_n$. Then,

$$M_Y(t) = [M_X(t)]^n$$

Activity 14.7.

Prove the last property mentioned above.

The applications of the above properties will be clear in the next session.

14.3 Cumulant generating function

Sometimes it is easy to work with the natural logarithm of the mgf; especially if the mgf is an exponential function.

Definition 14.3.

The natural logarithm of the moment generating function is called the cumulant generating function. It is denoted by C_X .

$$C_X(t) = \ln M_X(t)$$

Suppose that the Taylor series expansion of $C_X(t)$ around the origin is

$$C_X(t) = \sum_{r=0}^{\infty} \frac{\kappa_r t^r}{r!}.$$

Then the coefficients κ_r , $r = 0, 1, 2, \dots$, are called the cumulants.

Important relationships between cumulants and moments.

Let μ_r and κ_r be the r th moment and r th cumulant ($r = 0, 1, 2, \dots$) of a random variable X , respectively. Then,

$$\begin{aligned}
 \kappa_1 &= \mu_1 = E(X) \\
 \kappa_2 &= \mu_2 - \mu_1^2 = V(X)
 \end{aligned}$$

Example 14.7.

The cumulant generating function of a random variable X is given by

$$C_X(t) = 2 \left(\frac{1}{(1-t)^{10} - 1} \right).$$

Determine the mean and the variance of X .

Solution 14.7

$$\begin{aligned} C_X(t) &= 20t + 220 \frac{t^2}{2} + O(t^3) \\ E(X) &= \kappa_1 = 20 \\ V(X) &= \kappa_2 = 220 \end{aligned}$$

14.4 Probability generating function

Definition 14.4.

Suppose X is a random variable that takes only non-negative integer values with probabilities $\Pr(X = x) = P_X(x)$. The probability generating function is defined as

$$G_X(s) = E(s^X) = \sum_{x=0}^{\infty} s^x P_X(x) \quad \text{for } 0 \leq s \leq 1. \quad (14.1)$$

Note:

- If $s = e^t$, then, $G_X(e^t) = E((e^t)^X) = E(e^{Xt}) = M_X(t)$.
- If $t = \ln s$, then, $M_X(\ln s) = E(e^{X \ln s}) = E((e^{\ln s})^X) = E(s^X) = G_X(s)$.

Example 14.8.

Show that the probability generating function for a binomial (n, p) distribution is $G_X(s) = (sp + 1 - p)^n$. Deduce the moment generating function.

Solution 14.8

$$\begin{aligned} G_X(s) &= E(s^X) \\ &= \sum_{x=0}^n s^x C_x p^x (1-p)^{n-x} \\ &= \sum_{x=0}^n C_x (sp)^x (1-p)^{n-x} \end{aligned}$$

Using the mathematical identity $(a+b)^n = \sum_{x=0}^n C_x (a)^x (b)^{n-x}$, it is easy to see that

$$G_X(s) = (sp + 1 - p)^n$$

The moment generating function can be obtained by substituting $s = e^t$. Then,

$$M_X(t) = (pe^t + 1 - p)^n$$

Example 14.9.

The moment generating function of a discrete random variable taking non-negative values is

$$M_X(t) = [1 - e^{k\lambda}(1 - e^{-t})]^{-1}$$

Calculate $\Pr(X = 1)$, $\Pr(X = 2)$.

Solution 14.9

Letting $e^t = s$,

$$M_X(\ln s) = G_X(s) = [1 - e^\lambda(1 - s^{-1})]^{-1}$$

The Taylor expansion of $G_X(s)$ is given by

$$G_X(s) = e^{-\lambda}s + \frac{e^\lambda - 1}{e^{2\lambda}}s^2 + \dots$$

Then,

$$\begin{aligned} P_X(1) = \Pr(X = 1) &= e^{-\lambda} \\ P_X(2) = \Pr(X = 2) &= \frac{e^\lambda - 1}{e^{2\lambda}} \end{aligned}$$

Summary

In this session we discussed three generating functions, namely, the moment generating function, the probability generating function and the cumulant generating function. The moment generating function, when it exists, generates the central moments. In addition, its uniqueness is useful in identifying the distributions of functions of variables. The probability generating function can be used to generate the probabilities of discrete random variables taking integer values. There are relationships between these generating functions. Depending on the problem, we may work with a convenient generating function and use the inter-relations to derive moments or probabilities.

Learning Outcomes

By now, you should be able to

- Derive the moment generating function, cumulant generating function and the probability generating function of a random variable when the distribution function is given.
- Derive moments or probabilities using a given generating function.

Session 15

DISTRIBUTIONS OF FUNCTIONS OF RANDOM VARIABLES

Contents

Introduction

15.1 Distributions functions of single random variable

15.2 Distributions of linear combinations of several random variables

15.3 Distributions of other functions of several variables

Summary

Learning outcomes

Introduction

We often need to calculate probabilities related to functions of one or more variables. A linear transformation of a single variable, linear combination of several variables, are two such examples. In this section, we discuss the distributions of some important functions of variables.

15.1 Distributions of Functions of Single Random Variables

Theorem 15.1 (Distribution of a linear transformation of a normal random variable).

Suppose that $X \sim N(\mu, \sigma^2)$. Let a and b be constants and $Y = aX + b$, be a linear transformation of X . Then,

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

Proof

We shall prove this using the uniqueness of the moment generating function, when it exists.

$$\begin{aligned}M_Y(t) &= M_{aX+b}(t) \\&= e^{bt} M_X(at)\end{aligned}$$

Since $X \sim N(\mu, \sigma^2)$, $M_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$. Therefore,

$$\begin{aligned}M_Y(t) &= e^{bt} \left[e^{\mu(at) + \frac{1}{2}\sigma^2(at)^2} \right] \\&= e^{(a\mu+b)t + \frac{1}{2}(a\sigma)^2 t^2}\end{aligned}$$

This is the moment generating function of a $N(a\mu + b, a^2\sigma^2)$ distribution. Hence, by the uniqueness of the mgf, it implies that $Y \sim N(a\mu + b, a^2\sigma^2)$.

Activity 15.1.

Suppose that $X \sim N(\mu, \sigma^2)$. Let $Z = \frac{X-\mu}{\sigma}$. Show that $Z \sim N(0, 1)$.

Example 15.1.

The distribution of Statistics marks of a certain class are normally distributed with mean 50 and standard deviation 15. Those who have less than 25 marks get grade F and those who have more or equal to 85 marks get A+. The lecturer thinks that the lowest 5% should get grade F and highest 5% should get A+. The lecturer wants to linearly transform the marks so that the transformed marks satisfy the above condition. What should be the transformation?

Let X be the current marks. We are given that $X \sim N(50, 15^2)$. Suppose that the transformation is $Y = aX + b$. The lecturer wants $\Pr(Y < 25) = 0.05$ and $\Pr(Y \geq 85) = 0.05$.

According to Theorem 15.1, $Y \sim N(50a + b, (15a)^2)$. The first condition gives

$$\begin{aligned}\Pr(Y < 25) &= 0.05 \\ \Pr\left(Z < \frac{25 - (50a + b)}{15a}\right) &= 0.05 \\ \Rightarrow \frac{25 - (50a + b)}{15a} &= -1.645\end{aligned}$$

Similarly, the second condition gives

$$\begin{aligned}\Pr(Y \geq 85) &= 0.05 \\ \Pr\left(Z > \frac{85 - (50a + b)}{15a}\right) &= 0.05 \\ \Rightarrow \frac{85 - (50a + b)}{15a} &= 1.645\end{aligned}$$

The above two equations give $a = 0.8225$, and $b = 13.875$. Hence, the transformation is, $Y = 0.8225X + 13.875$.

Theorem 15.2 (Distribution of the square of a standard normal random variable).
 Suppose that $Z \sim N(0, 1)$. Let $U = Z^2$. Then

$$U \sim \chi_1^2.$$

Proof

$$\begin{aligned} M_U(t) &= E(e^{Ut}) \\ &= E(e^{Z^2 t}) \\ &= \int_{-\infty}^{\infty} e^{z^2 t} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-z^2(\frac{1}{2}-t)} dz \\ &= 2 \frac{1}{\sqrt{2\pi}} \int_0^{\infty} e^{-z^2(\frac{1}{2}-t)} dz \\ &= 2 \frac{1}{\sqrt{2\pi}} \int_0^{\infty} e^{-u(\frac{1}{2}-t)} \frac{1}{2\sqrt{u}} du \quad (\text{letting } u = z^2) \\ &= \frac{1}{\sqrt{2\pi}} \int_0^{\infty} e^{-u(\frac{1}{2}-t)} u^{-\frac{1}{2}} du \end{aligned}$$

Notice that $\frac{(\frac{1}{2}-t)^{1/2}}{\Gamma(1/2)} e^{-u(\frac{1}{2}-t)} u^{-\frac{1}{2}}$ is the density of a $\text{Gamma}(1/2, (1/2-t))$ distribution. Therefore,

$$M_U(t) = \frac{1}{\sqrt{2\pi}} \frac{\Gamma(1/2)}{(\frac{1}{2}-t)^{1/2}}$$

As mentioned in section 11.1, $\Gamma(1/2) = \sqrt{\pi}$. Then,

$$M_U(t) = \left(\frac{1}{1-2t} \right)^{1/2}$$

This is the moment generating function of a χ_1^2 distribution. Hence, by the uniqueness of the mgf, $U \sim \chi_1^2$.

Theorem 15.3 (Distribution of a multiple of a gamma random variable).
 Suppose that $X \sim \text{gamma}(\alpha, \lambda)$. Let $Y = aX$, where a is a constant. Then

$$Y \sim \text{Gamma}\left(\alpha, \frac{\lambda}{a}\right).$$

Proof

$$\begin{aligned} M_Y(t) &= M_{aX}(t) \\ &= M_X(at) \end{aligned}$$

Since $X \sim \text{Gamma}(\alpha, \lambda)$, $M_X(t) = \left(\frac{\lambda}{\lambda - t}\right)^\alpha$. Therefore,

$$\begin{aligned} M_Y(t) &= \left(\frac{\lambda}{\lambda - at}\right)^\alpha \\ &= \left(\frac{\lambda/a}{\lambda/a - t}\right)^\alpha \end{aligned}$$

This is the mgf of a Gamma $(\alpha, \lambda/a)$ distribution. It implies that $Y \sim \text{Gamma}(\alpha, \lambda/a)$.

Activity 15.2.

Suppose that $X \sim \text{Gamma}(\alpha, \lambda)$. Using Theorem 15.3, deduce that $\lambda X \sim \text{Gamma}(\alpha, 1)$ and $2\lambda X \sim \chi_{2\alpha}^2$.

15.2 Distributions of Linear Combinations of Several Random Variables

Let $X_1, X_2, \dots, X_n$ be random variables and $k_0, k_1, \dots, k_n$ be constants. Then, $k_0 + k_1X_1 + k_2X_2 + \dots + k_nX_n$ is called a linear combination of $X_1, X_2, \dots, X_n$. In this section, we consider the distributions of some important linear combinations of several random variables.

Theorem 15.4 (Distribution of a linear combination of independent normal random variables).

Suppose that $X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$ are two independent random variables. Let a, b and c be constants and $Y = aX_1 + bX_2 + c$. Then,

$$Y \sim N(a\mu_1 + b\mu_2 + c, a^2\sigma_1^2 + b^2\sigma_2^2).$$

Proof

$$\begin{aligned} M_Y(t) &= M_{aX_1 + bX_2 + c}(t) \\ &= e^{ct} [M_{aX_1 + bX_2}(t)] \\ &= e^{ct} [M_{aX_1}(t) M_{bX_2}(t)] \\ &= e^{ct} [M_{X_1}(at) M_{X_2}(bt)] \\ &= e^{ct} \left[e^{\mu_1(at) + \frac{1}{2}\sigma_1^2(at)^2} e^{\mu_2(bt) + \frac{1}{2}\sigma_2^2(bt)^2} \right] \\ &= e^{(a\mu_1 + b\mu_2 + c)t + \frac{1}{2}(a^2\sigma_1^2 + b^2\sigma_2^2)t^2} \end{aligned}$$

This is the moment generating function of a $N(a\mu_1 + b\mu_2 + c, a^2\sigma_1^2 + b^2\sigma_2^2)$ distribution. Hence, by the uniqueness of the mgf, it implies that $Y \sim N(a\mu_1 + b\mu_2 + c, a^2\sigma_1^2 + b^2\sigma_2^2)$.

Example 15.2.

The time it takes for me to reach the university from my home has a normal distribution with mean 45 minutes and standard deviation 15 minutes.

I have a lecture starting at 9.00 A.M. and I have to get some handouts prepared before the lecture. The time it takes to prepare the handouts has a normal distribution with mean 30 minutes and standard deviation 5 minutes. If I want to be 99% sure that I will be able to go to the lecture with the handouts, without being late, what is the latest time that I can leave home?

Solution 15.2

Let X_1 (in minutes) be the time to reach the university, X_2 (in minutes) be the time to prepare handouts, and $Y = X_1 + X_2$. According to given information, $X_1 \sim N(45, 15^2)$, $X_2 \sim N(30, 5^2)$. It is reasonable to assume that X_1 and X_2 are independent. Suppose I have to leave home k minutes before 9.00 A.M. so that $\Pr(Y < k) = 0.99$.

According Theorem 15.1, $Y \sim N(75, 250)$. We need k such that

$$\begin{aligned}\Pr(Y < k) &= 0.99 \\ \Pr\left(Z < \frac{k - 75}{\sqrt{250}}\right) &= 0.99 \\ \Rightarrow \frac{k - 75}{\sqrt{250}} &\approx 2.33 \\ k &\approx 2.33\sqrt{250} + 75 \\ &\approx 112 \text{ minutes.}\end{aligned}$$

Hence the latest time I can leave home is 7.08 A.M.

Theorem 15.5 (Distribution of the sum of independent, identically distributed (i.i.d) Bernoulli random variables).

Suppose that $X_1, X_2, \dots, X_n$ are independent, identically distributed (i.i.d.) random variables with Bernoulli (θ) distribution. Let $Y = X_1 + X_2 + \dots + X_n$. Then

$$Y \sim \text{Bin}(n, \theta).$$

Proof

This can be shown through the mgf. Proof is left for the reader as an activity.

Theorem 15.6 (Distribution of the sum of independent, Binomial random variables).

Suppose that $X_1, X_2, \dots, X_n$ are independent, random variables with $X_i \sim \text{Bin}(m_i, \theta)$, $i = 1, 2, \dots, n$. Let $Y = X_1 + X_2 + \dots + X_n$. Then

$$Y \sim \text{Bin}(m_1 + m_2 + \dots + m_n, \theta).$$

Proof

This can be shown through the mgf. Proof is left for the reader as an activity.

Note: X_i can be considered as a sum of m_i i.i.d Bernoulli (θ) random variables, ($i = 1, 2, \dots, n$). Hence, the above Y can be considered as a sum of $m_1 + m_2 + \dots + m_n$ i.i.d Bernoulli (θ) random variables. So, according to Theorem 15.5, Y should have a $\text{Bin}(m_1 + m_2 + \dots + m_n, \theta)$ distribution.

Theorem 15.7 (Distribution of a sum of independent Poisson random variables).
 Suppose that $X_1, X_2, \dots, X_n$ are independent, random variables with $X_i \sim \text{Poisson}(\lambda_i)$. Let $Y = X_1 + X_2 + \dots + X_n$. Then

$$Y \sim \text{Poisson}(\lambda_1 + \lambda_2 + \dots + \lambda_n).$$

Proof

This can be shown through the mgf. Proof is left for the reader as an activity.

Theorem 15.8 (Distribution of a sum of i.i.d. Geometric random variables).
 Suppose that $X_1, X_2, \dots, X_n$ are independent, identically distributed (i.i.d.) random variables with $G_f(\theta)$ distribution. Let $Y = X_1 + X_2 + \dots + X_n$. Then

$$Y \sim NB_f(n, \theta).$$

Proof

This can be shown through the mgf. Proof is left for the reader as an activity.

Theorem 15.9 (Distribution of a sum of independent normal random variables).
 Suppose that $X_1, X_2, \dots, X_n$ are independent random variables with $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, 2, \dots, n$. Let $Y = X_1 + X_2 + \dots + X_n$. Then

$$Y \sim N\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right)$$

Proof

This can be shown through the mgf. Proof is left for the reader as an activity.

Theorem 15.10 (Distribution of a sum of i.i.d. exponential random variables).
 Suppose that $X_1, X_2, \dots, X_n$ are independent identically distributed random variables with $X_i \sim \exp(\lambda)$, $i = 1, 2, \dots, n$. Let $Y = X_1 + X_2 + \dots + X_n$. Then

$$Y \sim \text{Gamma}(n, \lambda).$$

Proof

This can be shown through the mgf. Proof is left for the reader as an activity.

Theorem 15.11 (Distribution of a sum of independent gamma random variables).
 Suppose that $X_1, X_2, \dots, X_n$ are independent random variables with $X_i \sim \text{Gamma}(\alpha_i, \lambda)$, $i = 1, 2, \dots, n$. Let $Y = X_1 + X_2 + \dots + X_n$. Then

$$Y \sim \text{Gamma}\left(\sum_{i=1}^n \alpha_i, \lambda\right).$$

Proof

This can be shown through the mgf. Proof is left for the reader as an activity.

Example 15.3.

The lifetime of a certain type of bulb is known to have a Gamma distribution with mean 1000 hours and standard deviation 200 hours. I have bought 2 of these bulbs. I am going to install the second bulb immediately after the first bulb is burned out. Calculate the probability that I will be able to use the two bulbs for more than 3000 hours.

Solution 15.3

Let X_1 and X_2 be the lifetimes of the two bulbs. It is given that $X_i \sim \text{Gamma}(\alpha, \lambda), i = 1, 2$. According to the given information, $E(X_i) = 1000$ and $SD(X_i) = 200, i = 1, 2$. Therefore,

$$\begin{aligned}\alpha/\lambda &= 1000 \\ \alpha/\lambda^2 &= 200^2\end{aligned}$$

These two equations imply that $\lambda = 0.025$ and $\alpha = 25$.

According to Theorem 15.11, $Y = X_1 + X_2 \sim \text{Gamma}(2\alpha, \lambda) = \text{Gamma}(50, 0.025)$. We need $\Pr(Y > 3000)$. We leave it for the reader to complete the calculation. (Help: use the normal approximation to Gamma distribution, introduced in section 11.4.)

15.3 Distributions of Other Functions of Several Random Variables

The following are two important theorems worth knowing. We omit the proofs since they are beyond the scope of this course.

Theorem 15.12 (t distribution).

Suppose that $Z \sim N(0, 1)$ and $U \sim \chi^2_\nu$. Also suppose that Z and U are independent. Let $T = \frac{Z}{\sqrt{U/\nu}}$. Then $T \sim t_\nu$.

Theorem 15.13 (F distribution).

Suppose that $U \sim \chi^2_{\nu_1}$ and $V \sim \chi^2_{\nu_2}$. Also suppose that U and V are independent. Let $W = \frac{U/\nu_1}{V/\nu_2}$. Then,

$$W \sim F_{\nu_1, \nu_2}.$$

Summary

In this session we discussed several important theorems regarding functions of random variables. Most of those theorems can be easily proved using the uniqueness of the mgf. This shows the usefulness of the mgf.

Learning outcomes

By now, you should be able to

- Derive the distribution of a linear transformation of a normal random variable
- Derive the distribution of a square of a standard normal random variable
- Derive the distribution of a multiple of a gamma distribution
- Derive the distribution of a linear combination of several independent normal random variables
- Derive the distribution of the sum of iid Bernoulli random variables
- Derive the distribution of the sum of independent Binomial random variables with the same probability of success.
- Derive the distribution of the sum of independent Poisson random variables.
- Derive the distribution of the sum of independent normal random variables
- Derive the distribution of the sum of iid exponential random variables
- Derive the distribution of the sum of independent Gamma random variables with the same rate parameter
- State how the t distribution and F distribution arise.
- Use the theorems discussed in this session to solve problems

Session 16

SAMPLING DISTRIBUTIONS

Contents

Introduction

16.1 Distribution of sample mean

16.2 Distribution of sample variance of normal data

16.3 Distributions of two important pivotal quantities.

Summary

Learning outcomes

Introduction

In this session, we discuss about the distributions of some widely used functions of data collected on random samples. These are called sampling distributions. These are very useful in Statistical inference.

16.1 Distribution of the sample mean

Theorem 16.1 (Distribution of the sample mean of normal data).

Suppose that $X_1, X_2, \dots, X_n$ is a random sample from a normal distribution with mean μ and variance σ^2 . Let $\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$ be the sample mean. Then,

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

Proof

This can be shown through the mgf. Proof is left for the reader as an activity.

The following is a very important theorem in Statistics. It is often used for making inferences regarding the population mean. We omit the proof since it is beyond the scope of this course unit.

Theorem 16.2 (Large sample distribution of sample mean - Central Limit Theorem). Suppose that $X_1, X_2, \dots, X_n$ is a random sample from any distribution with mean μ and variance σ^2 . Let $\bar{X}_n = \frac{X_1 + X_2 + \dots + X_n}{n}$ be the sample mean, and $Z_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$. Then, the distribution of Z_n approaches the standard normal distribution as n approaches ∞ .

The central limit theorem implies that the distribution of $\bar{X}$ approaches the $N(\mu, \sigma^2/n)$ distribution as n is increased. The following figures illustrate this phenomenon. Figure 16.1 shows the central limit behavior of $\bar{X}$, when the data are from a Poisson distribution. Notice that the parent distribution (distribution of X) is somewhat skewed. However, the distribution of $\bar{X}$ has approached the normal distribution as the sample size is increased to 30. Figure 16.2 shows the central limit behavior of $\bar{X}$, when the data are from an exponential distribution. Here, the parent distribution is very skewed. The distribution of $\bar{X}$ seems to converge to the normal distribution. However, the distribution of $\bar{X}$ contains some skewness even when $n = 30$. A larger sample size is required for the distribution of $\bar{X}$ to become normal.

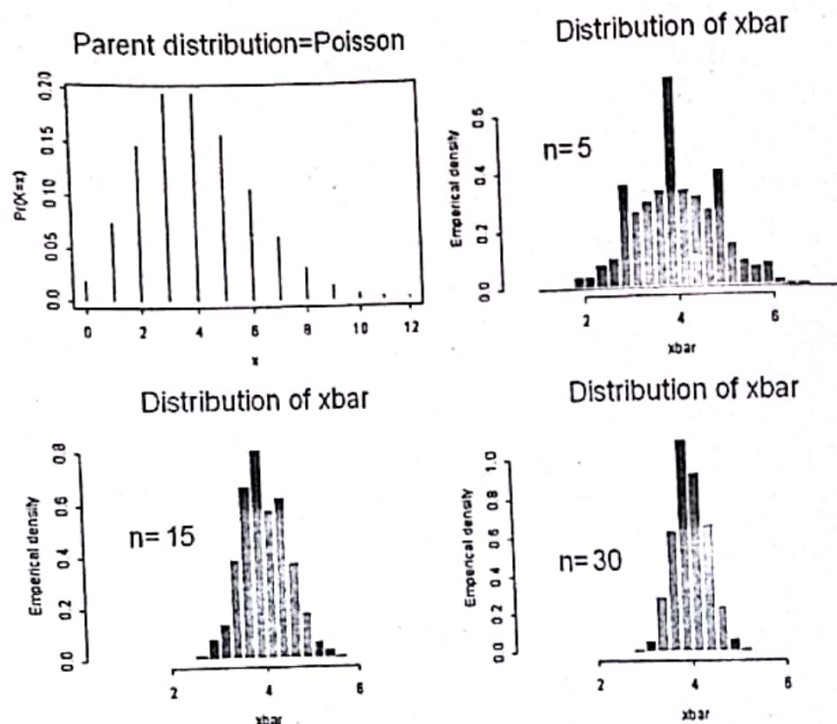


Figure 16.1: Central limit behavior of $\bar{X}$ when data are from a Poisson distribution

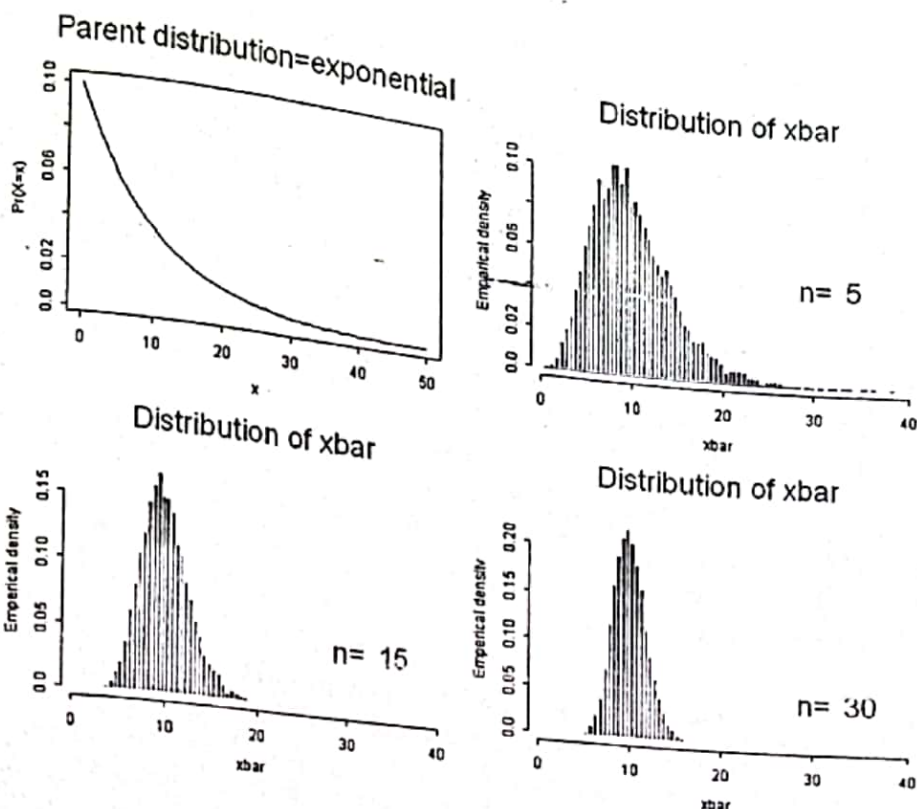


Figure 16.2: Central limit behavior of $\bar{X}$ when data are from an exponential distribution

The central limit theorem works well with $n \geq 30$ in for most practical situations. If the distribution of X_i 's is fairly symmetric, the theorem may work even with $n = 10$. However, the practitioner should be cautious when the distribution of X_i 's are skewed. If the distribution is very skewed, the required sample size for the central limit theorem to work may be quite larger than 30.

Example 16.1.

The movement of a stock price is modeled as follows. In each time period, the stock either goes up by 1 with probability 0.35, stays the same with probability 0.40 and goes down by 1 with probability 0.25. Calculate the approximate probability that after 500 time periods, the stock will be up by more than 70 from where it started.

Solution 16.1

Let X_i be the movement on the i th time period. According to the given information, for all i ,

$$E(X_i) = 1(0.35) + 0(0.4) - 1(0.25) = 0.1$$

$$E(X_i^2) = 1^2(0.35) + 0^2(0.4) + (-1)^2(0.25) = 0.6$$

$$V(X_i) = E(X_i^2) - [E(X_i)]^2 = 0.6 - 0.1^2 = 0.59$$

Let $Y = X_1 + X_2 + \dots + X_{500}$. It is required to find $\Pr(Y > 70)$.

Let $\bar{X} = Y/500$. According to the central limit theorem, $\bar{X} \sim N(0.1, 0.6/500)$. We leave the rest of the calculation for the reader to complete. (Help: Find the distribution of Y noting that $Y = 500\bar{X}$).

Example 16.2.

Suppose that a fraction θ of products from a production line is defective. A large random sample of n items is taken from this batch of products. Let P be the proportion of defective products in the sample. Derive the distribution of P .

Solution 16.2

Let

$$X_i = \begin{cases} 1 & \text{if the } i\text{th item in the sample is defective} \\ 0 & \text{otherwise} \end{cases}; i = 1, 2, \dots, n.$$

and $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$. Then it is clear that $P = \bar{X}$. Therefore, the distribution of P is the same as that of $\bar{X}$. The approximate distribution of $\bar{X}$ is given by the central limit theorem.

Note that $X_i \sim \text{Bernoulli}(\theta)$, $i = 1, 2, \dots, n$. Hence $E(X_i) = \theta$ and $V(X_i) = \theta(1-\theta)$. Hence by the central limit theorem, $\bar{X} \sim N(\theta, \theta(1-\theta)/n)$. It implies that $P \sim N(\theta, \theta(1-\theta)/n)$.

Activity 16.1.

An airline checks baggage without weighing each bag. However, the total weight of baggage is important in determining the fuel consumption and therefore has to be estimated. For this purpose, the airline has made a study of weight X (in pounds) of such bags and has found that for a randomly chosen bag, X is a random variable with mean 35 and variance 300. The cargo area has capacity for 100 pieces of baggage. If the cargo area of a plane is completely filled, what is the probability that the total weight of the bags is no more than 4000 pounds?

Activity 16.2.

Independent measurements $X_1, X_2, \dots, X_n$ are taken of the true distance δ in light-years between an observatory and a distant star. These measurements are known to have a common distribution with mean δ and standard deviation 2. An astronomer decides to use the sample mean of measurements as an estimate for δ .

If he uses 100 measurements, what is the probability that the estimate will be within 0.25 light-years of the actual distance δ ?

16.2 Distribution of sample variance of normal data

Theorem 16.3.

Suppose that $X_1, X_2, \dots, X_n$ is a random sample from a normal distribution with mean μ and variance σ^2 . Let $\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$ be the sample mean, and $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ be the sample variance. Then,

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2.$$

Proof

The proof is omitted because it is beyond the scope of this course.

Theorem 16.4 (Independence of $\bar{X}$ and S).
 Suppose that $X_1, X_2, \dots, X_n$ is a random sample from a normal distribution with mean μ and variance σ^2 . Let $\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$ be the sample mean, and $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ be the sample variance. Then, $\bar{X}$ and S are independent.

Proof

The proof is omitted because it is beyond the scope of this course.

16.3 Distributions of two important pivotal quantities.

Let $X_1, X_2, \dots, X_n$ be a random sample from a distribution with parameter θ . Let $Q = q(X_1, X_2, \dots, X_n; \theta)$ be a function of the data and θ such that the distribution of Q does not contain θ . Such a quantity Q is called a pivotal quantity. Pivotal quantities are very useful in constructing confidence intervals. Here we mention two important pivotal quantities that arise from normal data.

Theorem 16.5.

Suppose that $X_1, X_2, \dots, X_n$ is a random sample from a normal distribution with mean μ and variance σ^2 . Let $\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$ be the sample mean, $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ be the sample variance, and $Q = \frac{\sqrt{n}(\bar{X} - \mu)}{S}$. Then, Q is a pivotal quantity with

$$Q \sim t_{n-1}.$$

Proof

From Theorem 16.1,

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1).$$

From Theorem 16.3,

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2.$$

From Theorem 16.4, $\bar{X}$ and S are independent. These results, together with Theorem 15.12 imply that

$$\begin{aligned} \frac{\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2} / (n-1)}} &\sim t_{n-1} \\ \Rightarrow \frac{\sqrt{n}(\bar{X} - \mu)}{S} &\sim t_{n-1} \end{aligned}$$

Theorem 16.6.

Suppose that $X_1, X_2, \dots, X_m$ is a random sample from a normal distribution with mean μ_x

and variance σ_x^2 . Also suppose that $Y_1, Y_2, \dots, Y_n$ is another random sample from a normal distribution with mean μ_y and variance σ_y^2 . Let $S_x^2 = \frac{1}{m-1} \sum_{i=1}^m (X_i - \bar{X})^2$, $S_y^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2$ and $Q = \frac{S_x^2 \sigma_y^2}{S_y^2 \sigma_x^2}$. Then, Q is a pivotal quantity with

$$Q \sim F_{m-1, n-1}.$$

Activity 16.3.

Using Theorem 15.13 and other necessary theorems, prove Theorem 16.6.

Summary

In this session, we discussed about some important sampling distributions. One of the most important sampling distribution is given by the central limit theorem. This can be used in making inferences concerning the population mean. Most of the theorems discussed here are useful in deriving confidence intervals and testing hypotheses that you will learn under Statistical Inference.

Learning Outcomes

By now you should be able to

- Write down the distribution of the sample mean
- Use the distribution of the sample mean to find probabilities related to sample mean
- Write down the distribution of the sample variance of data that are normally distributed
- Write down the pivotal quantities and their distributions mentioned in section 16.3.

Session 17

JOINT PROBABILITY DISTRIBUTIONS

Contents

Introduction

17.1 Joint Probability Mass function

17.2 Joint Probability Density function

Summary

Learning outcomes

Introduction

In the previous sessions we studied distributions for a single random variable. Many problems involve several random variables simultaneously and often they are correlated. Therefore, we need to take into account the joint behaviour of the variables. In this session we discuss probability distributions that take this fact into account.

17.1 Joint probability mass function

Definition 17.1.

Let X and Y be two discrete random variables. The joint probability mass function $P_{X,Y}$ of X and Y is defined as

$$P_{X,Y}(x,y) = \Pr(X = x, Y = y)$$

113...

Note: By $(X = x, Y = y)$ we mean the intersection of the two events $(X = x)$ and $(Y = y)$.

Example 17.1.

A certain shop sells two brands of VCR (say, brand A and brand B). Let X denote the number of brand A VCR sold per day, and Y denote the number of brand B VCR sold per day. The following table shows the joint probabilities, according to the past data.

$P_{X,Y}(x,y)$		y		
		0	1	2
x	0	0.10	0.04	0.02
	1	0.30	0.14	0.06
	2	0.08	0.20	0.06

Table 17.1:

Let A be any set consisting of pairs of values (x,y) . Then

$$\Pr[(X,Y) \in A] = \sum_{(x,y) \in A} P_{X,Y}(x,y)$$

Example 17.1.

Consider Example 17.1. What is the probability that on a randomly selected day the number of brand A VCRs sold is more than that of brand B??

Solution 17.1

$$\begin{aligned} \Pr(X > Y) &= P_{X,Y}(1,0) + P_{X,Y}(2,0) + P_{X,Y}(2,1) \\ &= 0.30 + 0.08 + 0.20 = 0.58 \end{aligned}$$

Example 17.2 (Multinomial Distribution).

Suppose that there are k distinct possible outcomes of a trial (This is an extension of a Bernoulli trial). Denote these outcomes by $s_1, s_2, s_3, \dots, s_k$. Let $p_i = \Pr(s_i), i = 1, \dots, k$ such that $p_1 + p_2 + \dots + p_k = 1$. Suppose that we repeat the trial n times. Let X_i denote the number of times outcome s_i occurs in n trials, $i = 1, 2, \dots, k$. If the trials are independent, then, the joint pmf of the random variables $X_1, X_2, \dots, X_k$ is

$$P_{X_1, X_2, \dots, X_k}(x_1, x_2, \dots, x_k) = \frac{n!}{(x_1! x_2! \dots x_k!)} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k}$$

where $x_i = 0, 1, \dots, n$ ($i=1, 2, \dots, k$) such that $x_1 + x_2 + \dots + x_k = n$ and $p_1 + p_2 + \dots + p_k = 1$.

Example 17.3 (Binomial Distribution). Binomial distribution is a special case of the multinomial distribution. It arises when $k = 2$. The Binomial distribution can be written as

$$P_{X_1, X_2}(x_1, x_2) = \frac{n!}{(x_1! x_2!)} p_1^{x_1} p_2^{x_2}$$

where $x_1 = 0, \dots, n$, and $x_1 + x_2 = n$, $p_1 + p_2 = 1$.

The following table shows the joint pmf when $n = 3$ and $p_1 = 0.4$.

	$x_2=0$	$x_2=1$	$x_2=2$	$x_2=3$
$x_1=0$	0	0	0	0.216
$x_1=1$	0	0	0.432	0
$x_1=2$	0	0.288	0	0
$x_1=3$	0.064	0	0	0

Example 17.4.

Twenty five cylindrical shafts are selected from a shipment. It is found that 19 shafts have both diameter and length within the tolerance limits, 3 have satisfactory lengths but unsatisfactory diameters, 2 have satisfactory diameters but unsatisfactory lengths, and 1 is unsatisfactory with regard to both measurements. The manufacturer claims that the population proportions in each of the above categories are 0.9, 0.04, 0.04, and 0.02 respectively. What is the probability of occurring an event like above if the manufacturer's claim is correct?

Solution 17.4

$$\Pr(X_1 = 19, X_2 = 3, X_3 = 2, X_4 = 1) = \frac{25!}{(19! 3! 2! 1!)} (0.9)^{19} (0.04)^3 (0.04)^2 (0.02)^1 = 0.003$$

Activity 17.1.

Consider Example 17.1. According to past data, sixty percent of all customers who purchase brand B VCR's also buy an extended warranty. Let N denote the number of customers who buy an extended warranty on a given day.

- Determine the joint pmf of Y and N .
- What is $\Pr(Y = 3, N = 1)$?
- Calculate $\Pr(Y = N)$

17.2 Joint probability density function

Definition 17.2.

Let X and Y be two continuous random variables. Then, $f_{X,Y}(x, y)$ is called the joint probability density function of X and Y if for any two-dimensional set A ,

$$\Pr[(X, Y) \in A] = \int_{x \in A} \int_{y \in A} f_{X,Y}(x, y) dx dy$$

Example 17.5 (Bivariate normal distribution).

Two random variables X_1 and X_2 are said to have a bivariate normal distribution if the joint

probability density function is given by

$$f_{X_1, X_2}(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{x_1-\mu_1}{\sigma_1} \right)^2 - 2\rho \left(\frac{x_1-\mu_1}{\sigma_1} \right) \left(\frac{x_2-\mu_2}{\sigma_2} \right) + \left(\frac{x_2-\mu_2}{\sigma_2} \right)^2 \right] \right\}$$

$$-\infty < x_1 < \infty, -\infty < x_2 < \infty,$$

$$-\infty < \mu_1 < \infty, -\infty < \mu_2 < \infty,$$

$$\sigma_1 > 0, \sigma_2 > 0,$$

$$-1 < \rho < 1.$$

Two bivariate normal density functions are shown in Figures 17.1 and 17.2.



Figure 17.1: Bivariate normal density function with $\mu_x = 20, \sigma_x = 3, \mu_y = 10, \sigma_y = 2, \rho = 0$.



Figure 17.2: Bivariate normal density function with $\mu_x = 20, \sigma_x = 3, \mu_y = 10, \sigma_y = 2, \rho = 0.5$.

Example 17.6.

A company markets cans of mixed nuts containing almonds, cashews, and peanuts. Suppose that the net weight of each can is meant to be one pound, but the weight contribution of each type of nut is random. Let X be the weight of cashew and Y be the weight of peanuts in a randomly selected can. The following is the joint frequency table of X and Y from 40 randomly selected cans.

x	y				
	0.0-0.2	0.2-0.4	0.4-0.6	0.6-0.8	0.8-1.0
0.0-0.2	0	2	1	1	0
0.2-0.4	2	2	4	7	0
0.4-0.6	2	4	8	0	0
0.6-0.8	1	6	0	0	0
0.8-1.0	0	0	0	0	0

Table 17.2: Joint frequency distribution of X and Y from a sample of size 40 cans

Suggest a suitable joint probability density function for (X, Y) .

Solution 17.6

Note that because the three weights sum to 1, only two weights are free to vary. Therefore, it is enough to consider the joint probability distribution of two weights.

The joint probability density function of X and Y is defined on the region

$$D = \{(x, y) : 0 \leq x \leq 1, 0 \leq y \leq 1, 0 \leq x + y \leq 1\}$$

This region is shown in figure 17.3.

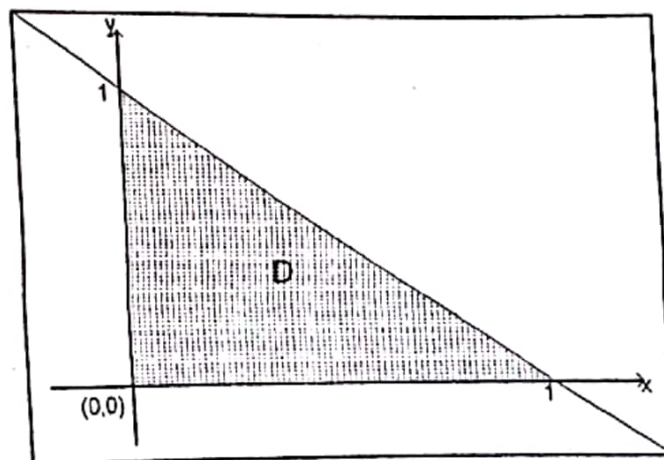


Figure 17.3: Domain $D = \{(x, y) : 0 \leq x \leq 1, 0 \leq y \leq 1, 0 \leq x + y \leq 1\}$

The following figure graphically shows the frequency distribution given in Table 17.2.

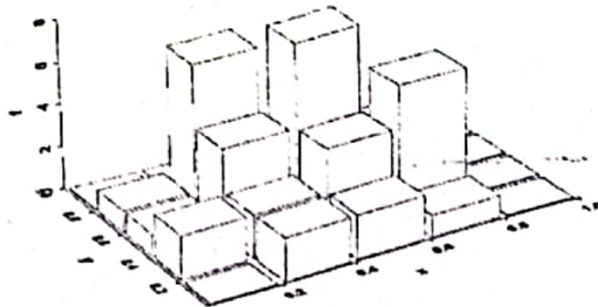


Figure 17.4: Joint frequency distribution of X and Y from a sample of 40 cans

According to this figure, a joint probability density function of the following form seems reasonable for X and Y .

$$f_{X,Y}(x,y) = \begin{cases} kxy; & 0 \leq x \leq 1, \quad 0 \leq y \leq 1, \quad x+y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

In order to determine the value of k , we use the fact that

$$\int_D f_{X,Y}(x,y) dx dy = 1$$

Note that

$$\begin{aligned} \int_D f_{X,Y}(x,y) dx dy &= \int_{x=0}^1 \int_{y=0}^{1-x} kxy dy dx \\ &= k \int_{x=0}^1 x \left[\frac{y^2}{2} \right]_{y=0}^{y=1-x} dx \\ &= k \int_{x=0}^1 x \frac{(1-x)^2}{2} dx \\ &= \frac{k}{2} \int_{x=0}^1 (x - 2x^2 + x^3) dx \\ &= \frac{k}{2} \left[\frac{x^2}{2} - \frac{2x^3}{3} + \frac{x^4}{4} \right]_{x=0}^{x=1} \\ &= \frac{k}{24} \end{aligned}$$

Since this should be equal to 1, $k = 24$. So, the joint pdf is

$$f_{X,Y}(x,y) = \begin{cases} 24xy; & 0 \leq x \leq 1, \quad 0 \leq y \leq 1, \quad 0 \leq x+y \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad (17.1)$$

Example 17.7.

Consider Example 17.6. Assuming that the joint probability density function of (X, Y) is given by equation (17.1),

- (a) calculate the percentage of cans containing more than $1/2$ pounds of cashew.
- (b) calculate the percentage of cans containing more than $1/3$ pounds of almond

Solution 17.7

- (a) In order to calculate $\Pr(X > 1/2)$, we should integrate the joint probability density function over the region E shown in Figure 17.5.

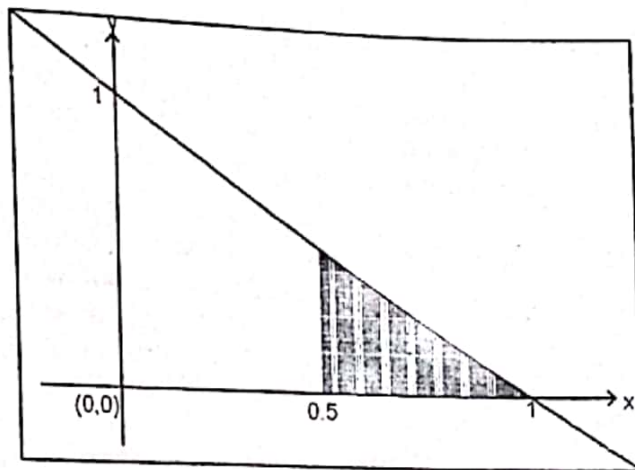


Figure 17.5: Region $E = \{(x, y) : 0.5 \leq x \leq 1, 0 \leq y \leq 1, 0 \leq x + y \leq 1\}$

$$\begin{aligned}
\Pr(X > 1/2) &= \int_E f_{X,Y}(x,y) dy dx \\
&= \int_{x=0.5}^1 \int_{y=0}^{1-x} f_{X,Y}(x,y) dy dx \\
&= \int_{x=0.5}^1 \int_{y=0}^{1-x} 24xy(x,y) dy dx \\
&= 24 \int_{x=0.5}^1 x \left[\frac{y^2}{2} \right]_0^{1-x} dx \\
&= 24 \int_{x=0.5}^1 x \left[\frac{(1-x)^2}{2} \right] dx \\
&= 0.3125
\end{aligned}$$

So, 31.25% cans contain more than 1/2 pounds of cashew.

- (b) A can containing more than 1/3 pounds of almond contains less than 2/3 pounds of cashew and peanut. Therefore, in order to find the percentage of cans containing more than 1/3 pounds of almond, we shall find the probability that the total weight of cashew and peanut is less than 2/3 pounds. For this, we should integrate the joint probability density function over the region F shown in Figure 17.6.

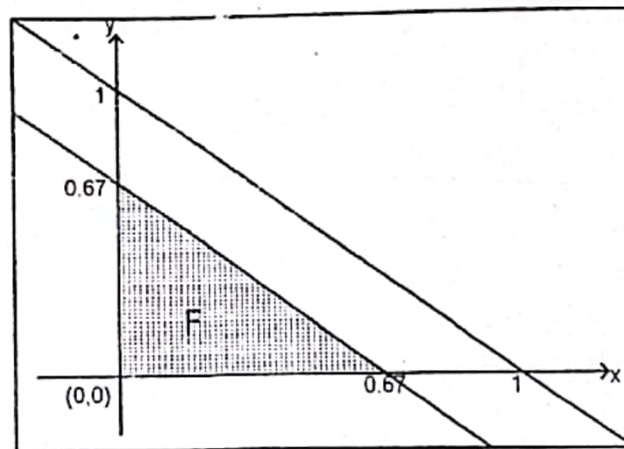


Figure 17.6: Region $F = \{(x,y) : 0 \leq x \leq 1, 0 \leq y \leq 1, 0 \leq x+y \leq 2/3\}$

$$\begin{aligned}
\Pr(X + Y < 2/3) &= \int_F f_{X,Y}(x,y) dy dx \\
&= \int_{x=0}^{2/3} \int_{y=0}^{1-x} f_{X,Y}(x,y) dy dx \\
&= \int_{x=0}^{2/3} \int_{y=0}^{1-x} 24xy(x,y) dy dx \\
&= 24 \int_{x=0}^{2/3} x \left[\frac{y^2}{2} \right]_0^{1-x} dx \\
&= 24 \int_{x=0}^{2/3} x \left[\frac{(1-x)^2}{2} \right] dx \\
&= 8/9 \approx 0.89
\end{aligned}$$

So, 89% cans contain more than 1/3 pound of almonds.

Summary

In this session, we discussed about joint probability mass function and joint probability density function. These can be considered as extensions of the probability mass function and probability density function to deal with several random variables that are correlated. The joint probabilities related to the random variables involved can be found using the joint probability densities or joint probability mass functions similar to that for single random variables.

Learning outcomes

By now, you should be able to

- Explain what is a joint pmf is
- Write down properties of joint pmf
- Explain what is a joint pdf is
- Write down properties of joint pdf
- Calculate probabilities using joint pmf or joint pdf.

Session 18

JOINT PROBABILITY DISTRIBUTIONS CONTINUED

Contents

Introduction

18.1 Marginal Distribution

18.2 Conditional Distribution

Summary

Learning outcomes

Introduction

Suppose we are given the joint distribution of two variables. When we want to find a probability or a moment related to a single variable, it is convenient to work with the distribution of that variable only. The distribution of a single variable is called the marginal distribution. In this session, we consider the marginal distribution of a variable when the joint distribution of two variables are known.

Another important problem is the calculation of probabilities and moments of a variable when the value of the other variable is known. That is, calculating the probabilities and moments of the conditional distribution. In this session we discuss how to derive the conditional distribution of one variable when the value of the other variable is known.

18.1 Marginal Distribution

Definition 18.1 (Marginal probability mass function).

Let X and Y be two discrete random variables with joint probability mass function $P_{X,Y}(x, y)$.

The marginal probability mass functions of X and Y are given by

$$P_X(x) = \sum_y P_{X,Y}(x,y) \text{ and } P_Y(y) = \sum_x P_{X,Y}(x,y)$$

respectively.

Example 18.1.

Consider Example 17.1. Tables 18.1 and 18.2 show the marginal distributions of X and Y respectively.

x	0	1	2
$P_X(x)$	0.16	0.50	0.34

Table 18.1: The marginal probability mass function of X

y	0	1	2
$P_Y(y)$	0.48	0.38	0.14

Table 18.2: The marginal probability mass function of Y

Activity 18.1.

Consider Example 17.3. Show that the marginal distributions of X_1 and X_2 are $\text{Bin}(3, 0.4)$ and $\text{Bin}(3, 0.6)$ distributions respectively.

Definition 18.2 (Marginal probability density function).

Let X and Y be two continuous random variables with joint probability density function $f_{X,Y}(x,y)$. Then, the marginal probability density functions of X and Y are given by

$$f_X(x) = \int_y f_{X,Y}(x,y) dy \text{ and}$$

$$f_Y(y) = \int_x f_{X,Y}(x,y) dx.$$

Example 18.2.

Suppose X_1 and X_2 are jointly distributed with the bivariate normal distribution given in Example 17.5. Find the marginal distributions of X_1 and X_2 .

Solution 18.2

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} f_{X_1, X_2}(x_1, x_2) dx_2$$

$$= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{x_1 - \mu_1}{\sigma_1} \right)^2 - 2\rho \left(\frac{x_1 - \mu_1}{\sigma_1} \right) \left(\frac{x_2 - \mu_2}{\sigma_2} \right) + \left(\frac{x_2 - \mu_2}{\sigma_2} \right)^2 \right] \right\} dx_2$$

Let $u = \frac{(x_1 - \mu_1)}{\sigma_1}$ and $v = \frac{(x_2 - \mu_2)}{\sigma_2}$. Then,

$$\begin{aligned} f_{X_1}(x_1) &= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}(u^2 - 2\rho uv + v^2)\right] \sigma_2 dv \\ &= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_1\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}(u^2 - 2\rho uv + v^2)\right] dv \end{aligned}$$

Now let $w = \frac{1}{\sqrt{1-\rho^2}}(v - \rho u)$. Then, $dw = \frac{dv}{\sqrt{1-\rho^2}}$, and

$$\begin{aligned} f_{X_1}(x_1) &= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_1} \exp\left[-\frac{1}{2}(w^2 + u^2)\right] dw \\ &= \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left[-\frac{1}{2}u^2\right] \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}w^2\right] dw \\ &= \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left\{-\frac{1}{2}u^2\right\} \\ &= \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left\{-\frac{1}{2}\left(\frac{x_1 - \mu_1}{\sigma_1}\right)^2\right\} \end{aligned}$$

This shows that $X_1 \sim N(\mu_1, \sigma_1^2)$. Similarly, it can be shown that $X_2 \sim N(\mu_2, \sigma_2^2)$.

Example 18.3.

Consider the joint pdf

$$f_{X,Y}(x,y) = \frac{1}{384} x^2 y^4 e^{-y-(x/2)}, \quad x > 0, y > 0$$

Find the marginal distributions of X and Y .

Solution 18.3

$$\begin{aligned} f_X(x) &= \int_0^{\infty} f_{X,Y}(x,y) dy \\ &= \int_0^{\infty} \frac{1}{384} x^2 y^4 e^{-y-(x/2)} dy \\ &= x^2 e^{-x/2}, \quad x > 0 \end{aligned}$$

$$\begin{aligned} f_Y(y) &= \int_0^{\infty} f_{X,Y}(x,y) dx \\ &= \int_0^{\infty} \frac{1}{384} x^2 y^4 e^{-y-(x/2)} dx \\ &= \frac{1}{384} y^4 e^{-y}, \quad y > 0 \end{aligned}$$

Example 18.4.
Consider the joint pdf

$$f_{X,Y}(x,y) = 6xy^2, 0 < x < 1, 0 < y < 1$$

Find the marginal distributions of X and Y .

Solution 18.4

$$\begin{aligned} f_X(x) &= \int_0^1 f_{X,Y}(x,y) dy \\ &= \int_0^1 6xy^2 dy \\ &= 2x, 0 < x < 1 \end{aligned}$$

$$\begin{aligned} f_Y(y) &= \int_0^1 f_{X,Y}(x,y) dx \\ &= \int_0^1 6xy^2 dx \\ &= 3y^2, 0 < y < 1 \end{aligned}$$

Example 18.5.
Consider the joint pdf

$$f_{X,Y}(x,y) = e^{-y}, 0 < x < y < \infty$$

Find the marginal distributions of X and Y .

Solution 18.5

$$\begin{aligned} f_X(x) &= \int_x^\infty f_{X,Y}(x,y) dy \\ &= \int_x^\infty e^{-y} dy \\ &= e^{-x}, x > 0 \end{aligned}$$

$$\begin{aligned} f_Y(y) &= \int_0^y f_{X,Y}(x,y) dx \\ &= \int_0^y e^{-y} dx \\ &= e^{-y}y, y > 0 \end{aligned}$$

Activity 18.2.

Consider Example 17.6. Show that the marginal distributions of X and Y are given by

$$f_X(x) = 12x(1-x)^2, 0 \leq x \leq 1$$

and

$$f_Y(y) = 12y(1-y)^2, 0 \leq y \leq 1$$

respectively

Independence of two random variables

Definition 18.3.

The two random variables X and Y are said to be independent if and only if their joint probability distribution (joint pmf or joint pdf) can be written as the product of the two marginal distributions.

That is, if and only if

$$P_{X,Y}(x,y) = P_X(x)P_Y(y) \quad \text{when } X \text{ and } Y \text{ are discrete}$$

$$f_{X,Y}(x,y) = f_X(x)f_Y(y) \quad \text{when } X \text{ and } Y \text{ are continuous}$$

Example 18.6.

Consider the joint distributions in Examples 17.1, 17.3, 17.5, 18.3, 18.4, and 18.5. In each case, determine whether or not the two random variables are independent.

Solution 18.6

In Example 17.1,

$$P_{X,Y}(0,0) = 0.10 \neq P_X(0)P_Y(0) = 0.16(0.48)$$

Therefore X and Y are not independent.

Note: If $P_{X,Y}(x,y) \neq P_X(x)P_Y(y)$ for at least one pair of (x,y) then X and Y are not independent. The random variables X and Y are independent only if $P_{X,Y}(x,y) = P_X(x)P_Y(y)$ all possible pairs of (x,y) .

In Example 17.3,

$$P_{X,Y}(0,0) = 0 \neq P_X(0)P_Y(0) = 0.216(0.064)$$

Therefore X and Y are not independent. This is also clear from the fact that $X_1 + X_2 = 3$.

In Example 17.5, the marginal densities are univariate normal distributions. It is easy to see that $P_{X,Y}(x,y) \neq P_X(x)P_Y(y)$. Therefore X and Y are not independent. Note that $P_{X,Y}(x,y) = P_X(x)P_Y(y)$ if and only if $\rho = 0$.

In Example 18.3, it is easy to see that $P_{X,Y}(x,y) = P_X(x)P_Y(y)$. Therefore, X and Y are independent.

In Example 18.4, it is easy to see that $P_{X,Y}(x,y) = P_X(x)P_Y(y)$. Therefore, X and Y are independent.

In Example 18.5, it is clear that $P_{X,Y}(x,y) \neq P_X(x)P_Y(y)$. Therefore, X and Y are not independent. This is also clear from the fact that $X < Y$.

Joint distribution of independent random variables

Let $X_1, X_2, \dots, X_n$ be independent random variables. Then, the joint pmf or the joint pdf of $X_1, X_2, \dots, X_n$ is given by the product of the individual pmf's or the pdf's. If the random variables are discrete,

$$P_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = P_{X_1}(x_1)P_{X_2}(x_2) \dots P_{X_n}(x_n).$$

If the random variables are continuous,

$$f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = f_{X_1}(x_1)f_{X_2}(x_2) \dots f_{X_n}(x_n).$$

18.2 Conditional Distributions

Definition 18.4.

1. Suppose that X and Y are two discrete random variables with joint pmf $P_{X,Y}(x,y)$. Let $P_X(x)$ and $P_Y(y)$ be the marginal pmf's of X and Y respectively. Then the conditional probability mass function of Y given that $X = x$ is denoted by $P_{Y|X}(y|x)$ and

$$P_{Y|X}(y|x) = \frac{P_{X,Y}(x,y)}{P_X(x)},$$

provided $P_X(x) \neq 0$.

2. Suppose that X and Y are two continuous random variables with joint p.d.f $f_{X,Y}(x,y)$. Let $f_X(x)$ and $f_Y(y)$ be the marginal p.d.f's of X and Y respectively. Then the conditional probability density function of Y given that $X = x$ is denoted by $f_{Y|X}(y|x)$ and

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)},$$

provided $f_X(x) \neq 0$.

Example 18.7.

Consider Example 17.1.

- (i) Derive $P_{X|Y}(x|2)$
- (ii) Calculate $\Pr(X > 0|Y = 2)$
- (iii) Calculate $E(X|Y = 2)$
- (iv) Derive $P_{Y|X}(y|1)$
- (v) Calculate $\Pr(Y \leq 1|X = 1)$
- (vi) Calculate $E(Y|X = 1)$

Solution 18.7

$$(i) P_{X|Y}(x|2) = \frac{P_{X,Y}(x,2)}{P_Y(2)}$$

Note that $P_Y(2) = 0.14$.

x	$P_{X,Y}(x,2)$	$P_{X Y}(x 2) = P_{X,Y}(x,2)/0.14$
0	0.02	1/7
1	0.06	3/7
2	0.06	3/7

$$(ii) \Pr(X > 0|Y = 2) = 6/7.$$

$$(iii) E(X|Y = 2) = 0(1/7) + 1(3/7) + 2(3/7) = 1$$

$$(iv) P_{Y|X}(y|1) = \frac{P_{X,Y}(1,y)}{P_X(1)}$$

Note that $P_X(1) = 0.5$.

y	$P_{X,Y}(1,y)$	$P_{Y X}(y 1) = P_{X,Y}(1,y)/0.5$
0	0.30	15/25
1	0.14	7/25
2	0.06	3/25

$$(v) \Pr(Y \leq 1|X = 1) = 22/25.$$

$$(vi) E(Y|X = 1) = 0(15/25) + 1(7/25) + 2(3/25) = 13/25.$$

Activity 18.1.

Calculate $V(X|Y = 2)$ and $V(Y|X = 1)$

Example 18.8.

Consider Example 17.6.

$$(i) \text{ What is } f_{X|Y}(x|0.2)?$$

$$(ii) \text{ What is } f_{Y|X}(0.3|0.2)?$$

Solution 18.8

(i)

$$\begin{aligned} f_{X|Y}(x|0.2) &= \frac{f_{X,Y}(x,0.2)}{f_Y(0.2)} \\ &= \frac{24(x)(0.2)}{12(0.2)(1-0.2)^2} \\ &= 3.125x \end{aligned}$$

(ii)

or

$$f_{X|Y}(0.3|0.2) = 3.125(0.3) = 0.9375$$

$$\begin{aligned} f_{X|Y}(0.3|0.2) &= \frac{f_{X,Y}(0.3, 0.2)}{f_Y(0.2)} \\ &= \frac{24(0.3)(0.2)}{12(0.2)(1-0.2)^2} \\ &= 0.9375 \end{aligned}$$

Example 18.9.

Suppose X_1 and X_2 are jointly distributed with the bivariate normal distribution given in Example 17.5. Show that the conditional distribution of X_1 given $X_2 = x_2$ is normal with mean $\mu_1 + \rho \frac{\sigma_1}{\sigma_2}(x_2 - \mu_2)$, and variance $\sigma_1^2(1 - \rho^2)$.

Solution 18.9

The conditional distribution of X_1 given $X_2 = x_2$ is given by

$$f_{X_1|X_2}(x_1|x_2) = \frac{f_{X_1, X_2}(x_1, x_2)}{f_{X_2}(x_2)}$$

From Example 18.2, we know that $X_2 \sim N(\mu_2, \sigma_2^2)$. Therefore,

$$f_{X_1|X_2}(x_1|x_2) = \frac{\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{x_1-\mu_1}{\sigma_1} \right)^2 - 2\rho \left(\frac{x_1-\mu_1}{\sigma_1} \right) \left(\frac{x_2-\mu_2}{\sigma_2} \right) + \left(\frac{x_2-\mu_2}{\sigma_2} \right)^2 \right] \right\}}{\frac{1}{\sqrt{2\pi}\sigma_2} \exp \left\{ -\frac{1}{2} \left(\frac{x_2-\mu_2}{\sigma_2} \right)^2 \right\}}$$

To make calculations simpler, let $u = \frac{(x_1-\mu_1)}{\sigma_1}$ and $v = \frac{(x_2-\mu_2)}{\sigma_2}$. Then,

$$\begin{aligned} f_{X_1|X_2=x_2}(x_1) &= \frac{\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} [u^2 - 2\rho uv + v^2] \right\}}{\frac{1}{\sqrt{2\pi}\sigma_2} \exp \left\{ -\frac{1}{2} v^2 \right\}} \\ &= \frac{1}{\sqrt{2\pi}\sigma_1\sqrt{(1-\rho^2)}} \exp \left\{ -\frac{1}{2(1-\rho^2)} (u^2 - 2\rho uv + \rho^2 v^2) \right\} \\ &= \frac{1}{\sqrt{2\pi}\sigma_1\sqrt{(1-\rho^2)}} \exp \left\{ -\frac{1}{2(1-\rho^2)} (u - \rho v)^2 \right\} \\ &= \frac{1}{\sqrt{2\pi}\sigma_1\sqrt{(1-\rho^2)}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\frac{\sigma_2(x_1 - \mu_1) - \sigma_1\rho(x_2 - \mu_2)}{\sigma_1\sigma_2} \right]^2 \right\} \\ &= \frac{1}{\sqrt{2\pi}\sigma_1\sqrt{(1-\rho^2)}} \exp \left\{ -\frac{1}{2\sigma_1^2(1-\rho^2)} \left[(x_1 - \mu_1) - \frac{\sigma_1}{\sigma_2}\rho(x_2 - \mu_2) \right]^2 \right\} \\ &= \frac{1}{\sqrt{2\pi}\sigma_1\sqrt{(1-\rho^2)}} \exp \left\{ -\frac{1}{2\sigma_1^2(1-\rho^2)} \left[x_1 - \left(\mu_1 + \frac{\sigma_1}{\sigma_2}\rho(x_2 - \mu_2) \right) \right]^2 \right\} \end{aligned}$$

This shows the above result.

Activity 18.3.

Suppose X_1 and X_2 jointly have a bivariate normal distribution with $\mu_1 = 10$, $\mu_2 = 20$, $\sigma_1 = 2$, $\sigma_2 = 3$ and $\rho = 0.6$. Calculate $\Pr(X_1 > 12 | X_2 = 15)$.

Summary

In this session we we learned about marginal distributions and conditional distributions. We discussed how to derive the marginal distributions and conditional distributions from the joint distributions. In particular, we derived the marginal and conditional distributions of bivariate normal distributions. In addition, we discussed about the independence of two random variables. We learned that two random variables are independent if and only if the joint distribution can be written as the product of respective marginal distributions.

Learning Outcomes

By now, you should be able to

- Derive marginal and conditional distributions when the joint distribution is given
- Calculate probabilities and moments related to the marginal and conditional distributions.
- Determine whether or not two random variables are independent, given the joint distribution

Session 19

DESCRIPTIVE PROPERTIES OF JOINT DISTRIBUTIONS

Introduction

19.1 Mean of a random vector

19.2 Mean of a function of random variables

Summary

Learning outcomes

Introduction

Suppose that $X_1, \dots, X_p$ are p random variables. Let

$$\underline{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_j \\ \vdots \\ X_p \end{pmatrix}$$

be the p variate random variable with component random variables $X_1, \dots, X_p$. In this session we discuss the descriptive properties of the joint distribution of $\underline{X}$. In particular, we consider the case $p = 2$. The main descriptive measures of a joint distribution are the mean vector, variance covariance matrix or the correlation matrix with variances of component random variables.

19.1 Mean of a random vector

Definition 19.1.

Let

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

be a random vector. Then, its mean is defined as

$$E(X) = \begin{pmatrix} E(X_1) \\ E(X_2) \end{pmatrix}$$

If X_1 and X_2 are discrete with joint pmf $P_{X_1, X_2}(x_1, x_2)$, then $E(X_1)$ and $E(X_2)$ are calculated using the corresponding marginal pmf's. That is, $E(X_1) = \sum_{x_1} x_1 P_{X_1}(x_1)$ and $E(X_2) = \sum_{x_2} x_2 P_{X_2}(x_2)$.

Similarly, if X_1 and X_2 are continuous with joint pdf $f_{X_1, X_2}(x_1, x_2)$, then $E(X_1)$ and $E(X_2)$ are calculated using the corresponding marginal pmf's. That is, $E(X_1) = \int_{x_1} x_1 f_{X_1}(x_1) dx_1$ and $E(X_2) = \int_{x_2} x_2 f_{X_2}(x_2) dx_2$.

19.2 Mean of a function of random variables

Definition 19.2.

Let X and Y be two random variables. Let $g(X, Y)$ be any function of X and Y .

If X and Y are discrete with joint pmf $P_{X, Y}(x, y)$, then, the mean (expected value) of $g(X, Y)$ is given by

$$E[g(X, Y)] = \sum_x \sum_y g(x, y) P_{X, Y}(x, y).$$

If X and Y are continuous with joint pdf $f_{X, Y}(x, y)$, then, the mean (expected value) of $g(X, Y)$ is given by

$$E[g(X, Y)] = \int_x \int_y g(x, y) f_{X, Y}(x, y) dy dx$$

Example 19.1.

Consider Example 17.1. The profit from a brand A VCR is Rs. 500 while the profit from a brand B VCR is Rs. 750. What is the mean profit per day?

Solution 19.1

The profit is $500X + 750Y$. It is a function of X and Y . Therefore, let $g(X, Y) = 500X + 750Y$. Then, the mean profit is given by

$$\begin{aligned} E[500X + 750Y] &= [500(0) + 750(0)]0.10 + [500(0) + 750(1)]0.04 + [500(0) + 750(2)]0.02 \\ &\quad + [500(1) + 750(0)]0.30 + [500(1) + 750(1)]0.14 + [500(1) + 750(2)]0.06 \\ &\quad + [500(2) + 750(0)]0.08 + [500(2) + 750(1)]0.20 + [500(2) + 750(2)]0.06 \\ &= 1085 \end{aligned}$$

Note: Since the above $g(X, Y)$ is a linear combination of X and Y ,

$$E[g(X, Y)] = E(500X + 750Y) = 500E(X) + 750E(Y)$$

Since

$$E(X) = \sum_x xP_X(x) = 0(.16) + 1(.50) + 2(.34) = 1.18$$

$$E(Y) = \sum_y yP_Y(y) = 0(.48) + 1(.38) + 2(.14) = 0.66,$$

$$E[g(X, Y)] = 500(1.18) + 750(0.66) = 1085.$$

Example 19.2.

Consider Example 17.1. Calculate $E(XY)$. Is $E(XY) = E(X)E(Y)$?

Solution 19.2

Let $g(X, Y) = XY$. Then,

$$\begin{aligned} E[XY] &= [(0)(0)]0.10 + [(0)(1)]0.04 + [(0)(2)]0.02 \\ &\quad [(1)(0)]0.30 + [(1)(1)]0.14 + [(1)(2)]0.06 \\ &\quad [(2)(0)]0.08 + [(2)(1)]0.20 + [(2)(2)]0.06 \\ &= 0.66 \end{aligned}$$

$$E(XY) = 0.66 \neq E(X)E(Y) = 1.18(0.66)$$

Example 19.3.

Consider Example 17.6. Calculate $E(XY)$. Is $E(XY) = E(X)E(Y)$?

Solution 19.3

$$\begin{aligned} E(XY) &= \int_D xy f_{X,Y}(x, y) dy dx \\ &= \int_{x=0}^1 \int_{y=0}^{1-x} xy(24xy) dy dx \\ &= 24 \int_{x=0}^1 x^2 \left[\int_{y=0}^{1-x} y^2 dy \right] dx \\ &= 24 \int_{x=0}^1 x^2 \frac{(1-x)^3}{3} dx \\ &= 2/15 \end{aligned}$$

We leave it for the reader to check whether $E(XY) = E(X)E(Y)$.

Activity 19.1.

Consider Example 18.4. Calculate $E(XY)$. Is $E(XY) = E(X)E(Y)$?

Theorem 19.1.

Let X and Y be random variables. If the two random variables are independent, then $E(XY) = E(X)E(Y)$

Activity 19.2.

Prove Theorem 19.1.

Summary

In this session we discussed about the mean vector of a random vector as well as how to calculate the mean of a function of two variables using the joint distribution. In particular, we discussed how to calculate $E(XY)$ which is required to calculate the covariance between two random variables X and Y . We learned that $E(XY)$ is not equal to $E(X)E(Y)$ in general.

Learning Outcomes

By now, you should be able to

- Calculate the mean vector of a random vector
- Calculate the mean of a function of random variables

Session 20

DESCRIPTIVE PROPERTIES OF JOINT DISTRIBUTIONS CONTINUED

Introduction

20.1 Variance covariance matrix of a random vector

20.2 Correlation between two random variables

Summary

Learning outcomes

=====

Introduction

In this session, we will learn about the variance covariance matrix and discuss how to derive the variance covariance matrix of a random vector as well as the correlation between two random variables, when the joint distribution is known.

20.1 Variance covariance matrix of a random vector

Definition 20.1.

Let X and Y be two random variables. The covariance between X and Y is defined as

$$\text{Cov}(X, Y) = E\{[X - E(X)][Y - E(Y)]\}$$

Theorem 20.1.

Let X and Y be two random variables.

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$$

Activity 20.1.

Prove Theorem 20.1.

Theorem 20.2.

Let X and Y be two random variables. If the two random variables are independent, then $\text{Cov}(X, Y) = 0$

Activity 20.2.

Prove Theorem 20.2.

Note:

The converse of Theorem 20.2 is not true in general. The covariance between two random variables can be zero even if the two random variables are not independent.

Example 20.1.

Suppose $X \sim U(-1, 1)$ and $Y = X^2$. Then,

$$\begin{aligned}\text{Cov}(X, Y) &= E(XY) - E(X)E(Y) \\ &= E(XX^2) - E(X)E(X^2) \\ &= E(X^3) - E(X)E(X^2).\end{aligned}$$

It is easy to see that $E(X^3) = 0$ and $E(X) = 0$. Therefore,

$$\text{Cov}(X, Y) = 0.$$

However, it is obvious that X and Y are not independent, since $Y = X^2$.

Definition 20.2.

Let $\underline{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ be a random vector. The variance covariance matrix of $\underline{X}$ is defined as

$$V(\underline{X}) = E\{[\underline{X} - E(\underline{X})][\underline{X} - E(\underline{X})]^T\}$$

Here, $[\underline{X} - E(\underline{X})]^T$ is the transpose of vector $[\underline{X} - E(\underline{X})]$. More simply, $V(\underline{X})$ is the following matrix.

$$V(\underline{X}) = \begin{pmatrix} V(X_1) & \text{Cov}(X_1, X_2) \\ \text{Cov}(X_1, X_2) & V(X_2) \end{pmatrix}$$

Notice that the variance covariance matrix is always symmetric.

Example 20.2.

Consider Example 17.1. Calculate the variance covariance matrix of the random vector $\begin{pmatrix} X \\ Y \end{pmatrix}$.

Solution 20.2

In Example 17.1, we have found the marginal distributions. Using those, we get

$$\begin{aligned}E(X) &= 0(0.16) + 1(0.50) + 2(0.34) = 1.18 \\ E(X^2) &= 0^2(0.16) + 1^2(0.50) + 2^2(0.34) = 1.86 \\ V(X) &= E(X^2) - [E(X)]^2 = 1.86 - 1.18^2 = 0.467\end{aligned}$$

$$\begin{aligned}
 E(Y) &= 0(0.48) + 1(0.38) + 2(0.14) = 0.66 \\
 E(Y^2) &= 0^2(0.48) + 1^2(0.38) + 2^2(0.14) = 0.94 \\
 V(Y) &= E(Y^2) - [E(Y)]^2 = 0.94 - 0.66^2 = 0.5044
 \end{aligned}$$

In Example 19.2, we found that $E(XY) = 0.66$. Therefore,

$$\begin{aligned}
 \text{Cov}(X, Y) &= E(XY) - E(X)E(Y) \\
 &= 0.66 - 1.18(0.66) \\
 &= -0.1188
 \end{aligned}$$

Hence, the variance covariance matrix is

$$\begin{pmatrix} 0.467 & -0.1188 \\ -0.1188 & 0.5044 \end{pmatrix}$$

Example 20.3.

Consider Example 18.5. Calculate the variance covariance matrix of the random vector $\begin{pmatrix} X \\ Y \end{pmatrix}$.

Solution 20.3

Using the marginal density functions found on page 125, we can calculate the variances. From marginal density of X we can see that $X \sim \text{Exp}(1)$. Hence $E(X) = 1$ and $V(X) = 1$. Similarly, from the marginal density of Y we can see that $Y \sim \text{Gamma}(2, 1)$. Hence $E(Y) = 2$ and $V(Y) = 2$. In order to calculate the Covariance, we need $E(XY)$.

$$\begin{aligned}
 E(XY) &= \int_{x=0}^{\infty} \int_{y=x}^{\infty} xye^{-y} dy dx \\
 &= 3
 \end{aligned}$$

Therefore,

$$\begin{aligned}
 \text{Cov}(X, Y) &= E(XY) - E(X)E(Y) \\
 &= 3 - 1(2) = 1
 \end{aligned}$$

Hence, the variance covariance matrix is

$$\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$$

Theorem 20.3.

Let X and Y be two random variables, a, b and c are constants. Then,

$$V(aX + bY + c) = a^2V(X) + b^2V(Y) + 2ab\text{Cov}(X, Y)$$

Activity 20.3.

Prove Theorem 20.3.

20.2 Correlation between two random variables

Definition 20.3.

Let X and Y be two random variables. The correlation between X and Y is defined as

$$\text{Cor}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{V(X)V(Y)}}$$

It is a value between -1 and 1 , inclusive; i.e., $-1 \leq \text{Cor}(X, Y) \leq 1$.

Correlation is a measure of the strength of the linear relationship between two random variables. $|\text{Cor}(X, Y)| = 1$ if and only if there is a perfect linear relationship between the two random variables. That is if $Y = a + bX$ for some constants a and b .

Note that the correlation between the same variable is 1. That is, for any random variable X , $\text{Cor}(X, X) = 1$.

The Correlation matrix of the random vector $\underline{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ is given by

$$\text{Cor}(\underline{X}) = \begin{pmatrix} \text{Cor}(X_1, X_1) & \text{Cor}(X_1, X_2) \\ \text{Cor}(X_1, X_2) & \text{Cor}(X_2, X_2) \end{pmatrix} = \begin{pmatrix} 1 & \text{Cor}(X_1, X_2) \\ \text{Cor}(X_1, X_2) & 1 \end{pmatrix}.$$

Summary

In this session, we learned about the variance covariance matrix as well as the correlation matrix. We discussed how to compute the variance covariance matrix and the correlation between two random variables.

Learning Outcomes

by now, you should be able to

- Define the covariance and correlation between two random variables
- Calculate the covariance and correlation between two random variables