Probability Distributions

Probability and probability distributions

1. Discrete probability distributions: the binomial distribution

A **random variable** is a variable whose values are associated with some probability of being observed. **A discrete** (as opposed to *continuous)* random variable is one that can assume only and distinct values. A **discrete variable** is a variable that can take only selected values. Some examples are the number of students in a class and the number of rooms in a house. There cannot be 18¼ students in the class, but there can be 18 or 19 students. *A discrete variable is mostly the result of a count.*

The set of all possible values of a random variable and its associated probabilities is called a **probability distribution.** The sum of all probabilities equals 1 (see Example 1).

Example 1

The possible outcomes in 2 tosses of a balanced coin are TT, TH, HT, and HH. Thus,

P(0H) = 1/4 P(1H) = 1/2 P(2H) = 1/4

The number of heads is therefore a discrete random variable, and the set of all possible outcomes with associated probabilities is a discrete probability distribution (see Table 1).

Table 1 Probability Distribution of Heads in Two Tosses of a fair Coin

Number of Heads	Possible Outcomes	Probability of Heads
0	TT	0.25
I	TH,HT	0.50
2	HH	0.25
		1.00

Binomial distribution

One discrete probability distribution is the **binomial distribution**. This is used to find the probability of X number of occurrences or successes of an event, **P(*X*),** in *n* trials of the same experiment

(1) there are only *two* possible and mutually **exclusive outcomes**,

(2) the *n* trials are **independent**

(3) the probability of occurrence or success, *p,* remains **constant** in each trial.

Then

$$P(X) = \frac{n!}{X!(n-X)!} P^X (1-p)^{n-X}$$

where *n!* (read "*n* factorial") = *n (n* - 1)· *(n* - 2) ... 3 ·2· 1, and 0! = 1 by definition.

The mean of the binomial distribution is $\mu = np$ and

The standard deviation is σ = $\sqrt{np(1-p)}$

If *P* = 1 – *P* = 0.5, the binomial distribution is symmetrical; if

***P* < 0.5, it is skewed to the right; and**

If *P* > 0.5, it is skewed to the left.

Example 2

Using the binomial distribution, we can find the probability of 4 heads in 6 flips of a balance coin as follows:

$$P(4) = \frac{6!}{4!(6-4)!}(1/2)^4(1/2)^2 = \frac{6\times5\times4\times3\times2\times1}{4\times3\times2\times1}\ (1/16)(1/4) = 15(1/64) = 0.23$$

The <u>expected</u> number (mean) of heads in 6 flips

μ= np = (6)(1/2) = 3 heads.

The standard deviation (σ) of the probability distribution of 6 flips is

$$\sigma = \sqrt{np(1-p)} = \sqrt{\frac{6}{4}} = \sqrt{1.5} = 1.22 \text{ heads}$$

Because P = 0.5, this probability distribution is symmetrical. If we were not dealing with a coin and the trials were not independent (as in sampling without replacement), we would have had to use the **hyper geometric distribution**

The Poisson distribution

The **Poisson distribution** is another discrete probability distribution. It is used to determine the probability of a designated number of successes **per unit of time**, when the events or successes are independent and the average number of successes per unit of time remains constant. Then

$$P(X) = (\lambda^{x} e^{-\lambda}) / X!$$

Where

X = designated number of successes

λ = average number of successes per unit of time P (X) = Probability of X number of successes

e = base of the natural logarithmic system, or 2.71828

Given the value of X (the expected value or mean **and** variance of the Poisson distribution), we can. Find e^{λ} and substitute in $P(X) = (\lambda^{x} e^{-\lambda}) / X!$, and find $P(X)$.

Example 3

A police department receives an average of 5 calls per hour. The probability of receiving 2 calls in a randomly selected hour is

$P(X) = (\lambda^{x} e^{-\lambda}) / X! = (5^{2} e^{-5})/2! = \{(25)(0.00674)\}/2 = 0.08425$

The Poisson distribution can be used as an approximation to the binomial distribution when n

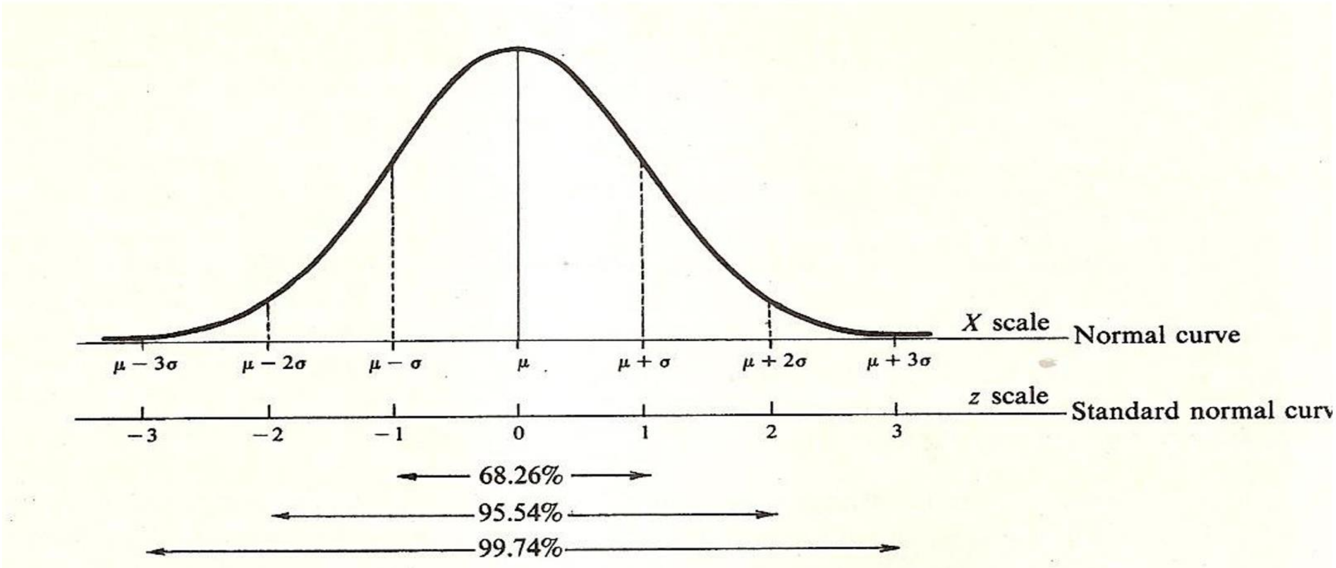
is large and *p* or 1 – *P* is small [say; *n* > 30 and *np* < 5 or n (l - *p)* < 5].

Continuous probability distributions: the normal distribution

A continuous random variable *X* is one that can assume an infinite number of values within any given interval. The probability that *X* falls within any interval is given by the area under the probability distribution (or density function) within that interval. The total area (probability) under the curve is 1. For example. Height of a student in the class, the time required to complete a task, weight of a shipment, the length of time before the first failure of a device, the profits or revenues of a company, are some examples. The height of a student cannot jump from 1.6m to 1.7m. It must assume every value between 1.6m and 1.7m at some time. *A continuous variable is mostly the result of a measurement.*

The normal distribution is a continuous probability distribution and the most commonly used distribution in statistical analysis. The normal' curve is bell-shaped and symmetrical about its mean. It extends indefinitely in both directions, but most of the area (probability) is clustered around the mean.

The standard normal distribution is a normal distribution with a mean of 0 and a standard deviation of I (that is, μ = 0 and σ = 1). Any normal distribution *(X* scale in Fig. 3-4) can be converted into a standard normal distribution by letting μ = 0 and expressing deviations from μ in standard deviation units *(z* scale). Under such conditions, 68.26% of the area (probability) under the standard normal curve is included within one standard deviation of the mean (i.e., within μ ± I *σ),* 95.54% within $\mu \pm 2\sigma$, and 99.74% within $\mu \pm 3\sigma$



Example 3b

Indicate which of the following are **continuous** and which are **discrete**:

(a) The number of children in a family.

(b) The weight of a person.

(c) The speed of a car.

(d) The number of plants in a garden.

(e) The capacity of an oil tank in liters.

(f) Number of units of an item held as inventory.

Solution:

(a) discrete (b) continuous (c) continuous

(d) discrete (e) continuous (f) discrete

To find probabilities (areas) for problems involving the normal distribution, we first convert the *X* value into its corresponding *z* value, as follows:

$$z = \frac{x - \mu}{\sigma}$$

Then we look up the *z* value in Appendix 3. This gives the proportion of the area (probability) include under the curve between the mean and that *z* value.

Example 4:

The area (probability) under the standard normal curve between $z = 0$ and $z = 1.96$ is obtain by leaking up the value *of* 1.96 in Appendix 3. We move down the z column in the table *to* 1.9 and then across un we are below the column headed .06. The value that we get is 0.4750. This means that 47.50% of the total area *(*of 1, or 100%) under the curve lies between $z = 0$ and $z = 1.96$ (the shaded area in the figure above the table Because of symmetry, the area between $z = 0$ and

$z = -1.96$ (not given in the table) is also 0.4750, or 47.509%.

Example 5

Suppose X is a normally distributed random variable with $\mu = 10$ and $\sigma^2 = 4$ and we want find the probability of X assuming a value between 8 and 12. We first calculate the' z values corresponding to the X values of 8 and 12 and then leak up these z values in Appendix 3:

$$z = \frac{X-\mu}{\sigma} = \frac{8-10}{2} = -1$$

And

$$z_2 = \frac{X-\mu}{\sigma} = \frac{12-10}{2} = 1$$

For z =1, we get 0.3413 farm Appendix 3. Then, z = ± 1 equals 2(0.3413), or 0.6826.

This means that the probability of X assuming a value between 8 and 12, or $P(8 < X < 12)$, is 68.26% .

Example 6

Suppose again that X is a normally distributed random variable with $\mu = 10$ and $\sigma^2 = 4$

The probability that X will assume a value between 7 and 14 can be found as follows:

$$z_1 = \frac{X-\mu}{\sigma} = \frac{7-10}{2} = -1.5$$

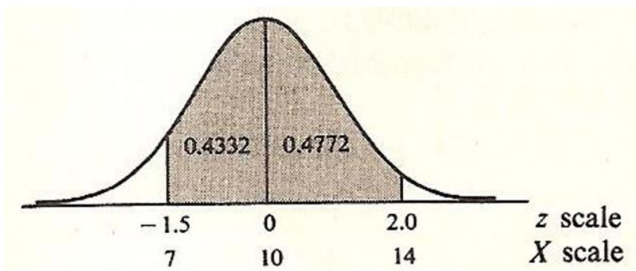
And

$$z_2 = \frac{X-\mu}{\sigma} = \frac{14-10}{2} = 2$$

For $Z_l = -1.5$, we *leak* up 1.50 in App. 3 and get 0.4332. For $Z_2 = 2$, we get 0.4772.

Therefore, P(*7* < X < 14) = 0.4332 + 0.4772 = 0.9104, or 91.04% Therefore, the probability of X assuming a value: *smaller than* 7 or *larger than* 14 (the un-shaded tail *areas* in fig) is 1 – 0.9104 = 0.0896, 8.96%.

The normal distribution approximates to binomial distribution when *n*≥ 30 and both *n p* > 5 and n(1 – *p)* > 5, and approximates the Poisson distribution when λ ≥10.



Another continuous probability distribution is the **exponential distribution**.

Chebyshev's theorem, or inequality, states that regardless of the shape of a distribution, the proportion of the observations or area falling within *K* standard deviations the mean is at least

$1 – 1/ K^2$, for $K \geq 1$.