Chapter 1: Probability Theory

1.1 Counting

Experiment, Model and Event

Rule 1: If an experiment consists of n trials where each trial may result in one of k possible outcomes, there are k^n possible outcomes of the entire experiment.

Example 1, page 7.

Rule 2: There are n! (n factorial) ways of arranging *n* distinguishable objects into a row**.**

For example, *n!=n(n-1)(n-2)….3.2.1,* and *5!=5.4.3.2.1=120*

Example 2, page 8.

Rule 3: If a group of n objects is composed on n_1 identical objects of type 1, n_2 identical objects of type 2,…..…, n_r identical objects of type *r*, the number of distinguishable arrangements into a row, denoted by $\begin{bmatrix} n \\ n_i \end{bmatrix}$ is

$$\begin{bmatrix} n \\ n_i \end{bmatrix} = \frac{n!}{n_1!n_2!\ldots n_r!}$$

If a group of *n* objects is composed of k identical objects of one kind and the remaining *(n-k)* objects of a second kind, the number of distinguishable arrangement of the n objects into a row, denoted by $\binom{n}{k}$, is given by

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Note $\binom{n}{0} = 1, \binom{n}{n} = 1$ and $\binom{n}{k} = 0$ if $k > n$.

Binomial Coefficient:

The term, $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ is known as the binomial coefficient .

Example 4, page 9.

Binomial series

$$(x+y)^n = \sum_{i=0}^{n} \binom{n}{i} x^i y^{n-i}$$

Problem 2, page 13.

1.2 Probability

Sample Space:

Definition 1: The *sample space* is the collection of all possible different outcomes of an experiment.

Definition 2: A *point in the sample space* is a possible outcome of an experiment.

Event:

Definition 3: An event is any set of points in the sample space.

Example 1, page 13.

Probability:

Definition 4: If an experiment is repeated a large number of times and the event A is observed n_A times, the probability of A is approximately

$$P(A) \approx \frac{n_A}{n}$$

As n becomes infinitely large, the fraction $\frac{n_A}{n}$ approaches P(A). This is

$$P(A) = \lim_{n \to \infty} \frac{n_A}{n}$$

This is called statistical or empirical probability**.**

Definition: If a trial results in *n* exhaustive mutually exclusive and equally likely cases and n_A of them are favorable to the happening of an event *A* then the probability ``P'' of happening A is given by

$$P(A) = \frac{\text{Favouralble \# of Cases}}{\text{Exhasutive \# of Cases}} = \frac{n_A}{n}$$

This is called mathematical or classical probability**.**

Probability Function:

Definition 5: A probability function is a function that assigns probabilities to the various events in the sample space**.**

Conditional Probability

Definition 6: If A and B are two events in a sample space S, then the joint event A and B is denoted by $P(A \cap B)$ or P(AB).

Definition 7: If A and B are two events in a sample space S. The conditional probability of A given B is the probability that A occurs given that B occurred and defined as follows

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(AB)}{P(B)}, \qquad \text{P(B)} > 0$$

Exercise 11, page 21.

Independent Events

Definition 8: Two events A and B are said to be independent if

$P(A \mid B) = P(A)$ or $P(B \mid A) = P(B)$ or $P(AB) = P(A)P(B)$

Example 6, page 18.

Independent Experiments:

Definition 9: Two experiments are independent if for very event A associated with one experiment and very event B associated with the second experiment,

$$P(AB)=P(A)P(B)$$

Definition 10: *n* experiments are mutually independent if for every set of *n* events, formed by considering one event from each of the *n* experiments, the following equation is true,

$$P(A_1, A_2, \ldots A_n) = P(A_1)P(A_2),\ldots,P(A_n)$$

where A_i represents an outcome of the *ith* experiment, for i=1,2,…,n

1.3 Random Variables

Random Variables

Definition 1: A random variable is a function that assigns real numbers to the points in a sample space.

Extra Example 1: Toss a coin twice and observe the number of heads.

Definition 2: The conditional probability of X given Y, written as $P(X = x \mid Y = y)$ is the probability that the random variable X assumes the value x, given that the random variable Y has assumed the value y,

$$P(X = x \mid Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}, \quad P(Y = y) > 0$$

Example 4, page 24

Probability Function

Definition 3: The probability function of the random variable X usually denoted by *f(x),* is the function that gives the probability of X taking the value x, for any real number x. In symbol,

$$f(x)=P(X=x), \quad f(x) \text{ lies between 0 and 1.}$$

Distribution Function

Definition 4: The distribution function of a random variable X, usually denoted by $F(x)$ and is defined as follows**:**

$$F(x) = P(X \le x) = \sum_{t \le x} f(t)$$

See figure (for discrete random variable) on page 27. From the figure

$$F(2) = P(X \le 2) = 0.7$$

Binomial Distribution

Definition 5: Let X be a random variable which follows a binomial distribution with parameters n and p. The the probability function can be written as

$$p(x) = P(X = x) = \binom{n}{x} p^x q^{n-x}, 0 \le p \le 1 \text{ and } \mathrm{q} = 1 - \mathrm{p}$$

The mean and variance of Binomial distribution are respectively

$$E(X) = \mu = np \text{ and } V(X) = \sigma^2 = npq$$

The distribution function is

$$F(x) = P(X \le x) = \sum_{i \le x}^{x} \binom{n}{i} p^i q^{n-i}$$

where the summation extends overall possible values of I less than or equal to x. The binomial table (Table A3, page 511) will give cumulative probabilities and will be used a lot.

Exercise 1, page 32.

Discrete Uniform Distribution

Definition 6: Let X be a random variable. The probability distribution function of x is

$$f(x) = \frac{1}{N} \quad \mathrm{x} = 1,2,\ldots,\mathrm{N}$$

Example 6, page 28.

Joint Distribution

Definition 7: The joint probability function $f(x_1, x_2, \ldots, x_n)$ of the random variables $X_1, X_2, \ldots, X_n$, is the probability of the joint occurrence of $X_1 = x_1, X_2 = x \ ,\ldots, X_n = x_n$, which is defined as follows

$$f(x_1, x_2, \ldots, x_n) = P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$$

Definition 8: The joint distribution function $F(x_1, x_2, \ldots, x_n)$ of the random variables $X_1, X_2, \ldots, X_n$, is the probability of the joint occurrence of $X_1 \le x_1, X_2 \le x \ ,\ldots, X_n \le x_n$, which is defined as follows

$$F(x_1, x_2, \ldots, x_n) = P(X_1 \le x_1, X_2 \le x_2, \ldots, X_n \le x_n)$$

Example 7, page 29.

Definition 9: The conditional probability function of X given Y, $f(X \mid Y)$ is

$$f(x \mid y) = P(X = x \mid Y = y) = \frac{P(X = x \cap Y = y)}{P(Y = y)} = \frac{f(x, y)}{f(y)}, \quad \mathrm{f(y)} > 0$$

where *f(x,y)* is the joint probability function of X and Y and *f(y)* is the probability function of Y.

Example 8, page 30.

Definition 11: Let $X_1, X_2,...,X_n$ be random variables with the respective probability functions $f_1(x_1), f_2(x_2),..., f_n(x_n)$ and with the joint probability function $f(x_1, x_2,...,x_n)$. Then $X_1, X_2,...,X_n$ are mutually independent if

$f_1(x_1, x_2,...,x_n) = f_1(x_1), f_2(x_2),..., f_n(x_n)$ for all combinations values of $x_1, x_2,...,x_n$.

Example 9, page 31.

1.4 Some Properties of Random Variables

Quantiles

Definition 1: The number x_p, for a given value of *p* between 0 and 1, is called the *pth* quantile of the random variable X, if $P(X < x_p) \le p$ and $P(X > x_p) \le 1 - p.$

Example 1, page 34.

Expected Value

Definition 2: Let X be a random variable with the probability function *f(x)* and let *u(X)* be a real valued function of *X*. The expected value of *u(X),* written as *E[u(X)],* is

$$E(u(X)] = \sum_x u(x) f(x).$$

We are mainly interested for the following two expected values

Definition 3: Let X be a random variable with the probability function *f(x)*. The expected value of X, usually denoted by μ, is

$$\mu = E(X) = \sum_x x f(x).$$

Variance

Definition 4: Let X be a random variable with the probability function *f(x)*. The variance of X, usually denoted by σ^2 is

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

Example 4, page 38.

Definition 5: Let $X_1, X_2, \ldots, X_n$ be random variables with the joint probability function $f(x_1, x_2, \ldots, x_n)$, and let $u(X_1, X_2, \ldots, X_n)$ be a real valued function of $X_1, X_2, \ldots, X_n$. Then the expected value of $u(X_1, X_2, \ldots, X_n)$ is

$$E(u(X_1, X_2, \ldots, X_n)] = \sum_x u(x_1, x_2, \ldots, x_n) f(x_1, x_2, \ldots, x_n).$$

where the summation extends over all possible combinations of values of $x_1, x_2, \ldots, x_n$

Theorem 1: Let $X_1, X_2, \ldots, X_n$ be random variables and let $Y = X_1 + X_2 + \ldots + X_n$ then $E(Y) = E(X_1) + E(X_2) + \ldots + E(X_n)$

Example 6, page 40

Lemma 1

$$\sum_{i=a}^{N} i = \frac{(N+a)(N-a+1)}{2} \quad \text{and} \quad \sum_{i=1}^{N} i = \frac{N(N+1)}{2}$$

Example 7, page 41.

Covariance

Definition 6: Let X_1 and X_2 be two random variables with means μ_1 and μ_2, probability functions $f(x_1)$ and $f(x_2)$, respectively, and joint probability function $f(x_1, x_2)$.The covariance of X_1 and X_2 **is**

$$\begin{aligned} Cov(X_1, X_2) &= E[(X_1 - E(X_1))(X_2 - E(X_2))] \\ &= E[(X_1 - \mu_1)(X_2 - \mu_2)] \\ &= E(X_1 X_2) - \mu_1 \mu_2 \end{aligned}$$

Correlation Coefficient

Definition 7: The correlation coefficient between two random variables is

$$\rho = \frac{Cov(X_1, X_2)}{\sqrt{Var(X_1)Var(X_2)}} \qquad -1 \le \rho \le 1$$

Lemma 2:

$$\sum_{i=1}^{N} i^2 = \frac{N(N+1)(2N+1)}{6}$$

Theorem 5: Let *X* be the sum of n integers selected at random, without replacement, from the first *N* intergers *1* to *N*. Then the mean and variance of X are respectively

$$E(X) = \frac{n(N+1)}{2}$$

$$V(X) = \frac{n(N+1)(N-n)}{12}$$

Example 13, page 49. (Application of Theorem 5)

1.5 Continuous Random Variables

Discrete Random Variable

Definition 1: A random variable X is discrete if there exists a one-to-one correspondence between the possible values of X and some or all of the positive integers.

Continuous Random Variable

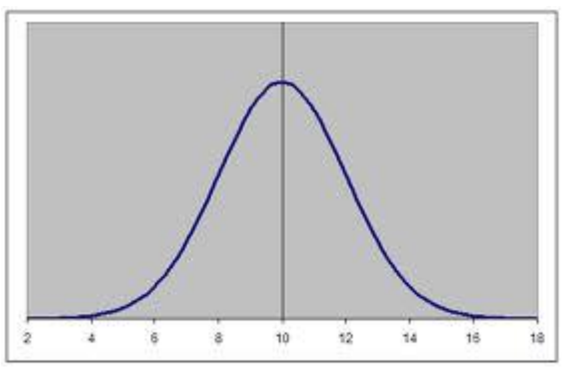
Definition 2: A continuous random variable (RV) is one that can assume the infinitely many values corresponding to the points on a line interval. A random variable X is continuous if $P(X \le x) = P(X < x)$, which also implies $P(X = x) = 0$.

Normal Distribution

Definition 3: The normal distribution is the most useful distribution in both theory and application of statistics. If X is a normal random variable, then the probability distribution function of X is defined as follows.

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad -\infty \le \mu \le \infty \text{ and } \sigma > 0$$

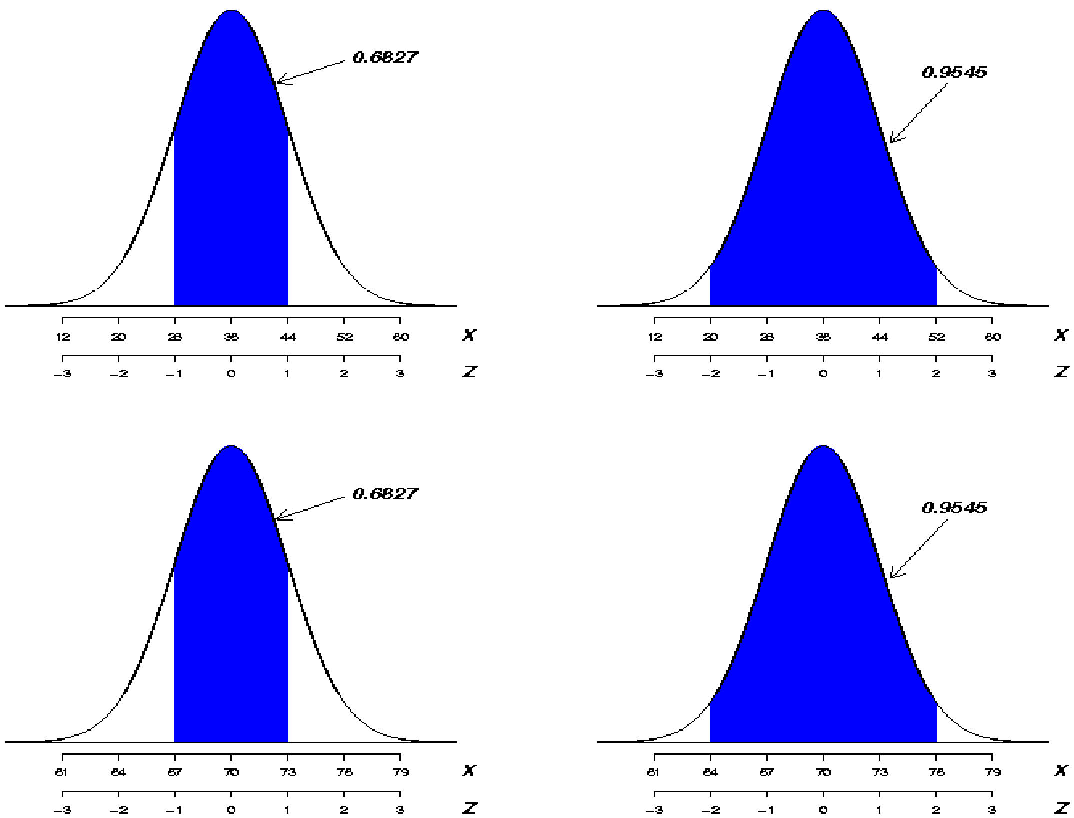
where μ is the mean and σ is the standard deviation of X**.** The pdf of a normal distribution with mean, μ =10 is given below



Standard Normal: If X is distributed as normal with mean μ and standard deviation, σ. That is $X \sim N(\mu, \sigma)$. Then

$$Z = \frac{x-\mu}{\sigma} \sim N(0,1)$$

is called a standard normal variable. That is, E(z)= 0 and V(z)=1.



The standard normal (or z) table is given in Table A1 (page 506)

Theorem 1: For a given value of p, let x_p be the *pth* quantile of a normal random variable with mean μ and variance σ^2. Let z_p be the *pth* quantile of a standard normal random variable. The qunatile x_p may be obtained from z_p by using the following relationship

$$z_p = \frac{x_p - \mu}{\sigma} \Rightarrow x_p = \mu + \sigma \times z_p$$

where the value of $z_p = -1.645$ when p=0.05 and 1.645 for p=0.95 **.**

Using R

```
> qnorm(0.05)
[1] -1.644854
> qnorm(0.95)
[1] 1.644854
```

Extra Example 2: A traffic study conducted at one point on an interstate highway shows that vehicle speeds are normally distributed with a mean of 63.5 miles per hour and a standard deviation of 4.8 mph.

a. If a vehicle is randomly selected, what is the probability that its speed is between 55 and 65 mph?
b. 60% of the speeds in the distribution exceed what value?.

Central Limit Theorem:

Theorem 2: If random sample of n observations are drawn from any population (non-normal, skewed, unknown etc) with finite mean μ and standard deviation σ, then when n is large $(n \geq 30)$, the sampling distribution of the sample mean $\bar{x}$ is approximately normally distrubuted, with mean and standard deviation (SD) are respectively

$$E(\bar{x}) = \mu_{\bar{x}} = \mu \quad and \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}.$$

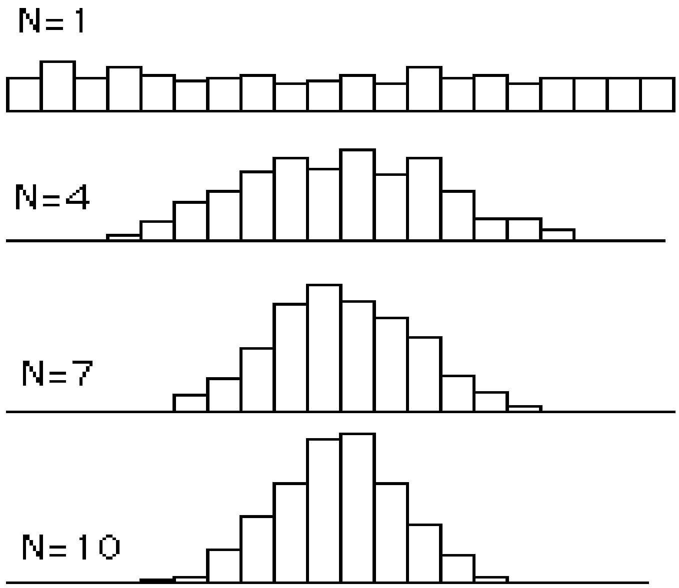
That is, for large n,

$$\bar{x} \approx N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

The approximation will become more and more accurate as n becomes larger and larger. The symbol "$\approx$" stands for *approximately distributed*.

Example 7, page 58

Comments: When the population distribution of x (measurements) is symmetrical about the mean μ, the CLT will apply very well to small sample size, $n = 10$. However, if the population is skewed, the larger the sample sizes are required to yield an effective approximation to the distribution of $\bar{x}$ by the normal probability distribution. As a rule of thumb, large sample size implies, $n \geq 30$.



Chi-squared (χ^2)Distribution

Definition 4: A random variable X has the chi-squared distribution with k degrees of freedom if the pdf of X is given by

$$f(x) = \frac{e^{-\frac{x}{2}} x^{\frac{k}{2}-1}}{2^{\frac{k}{2}} \Gamma\left(\frac{k}{2}\right)}$$

The mean and variance of chi-squared distribution are respectively.

$$E(X) = k \text{ and } V(X) = 2k$$

Theorem 3: Let $X_1, X_2, ..., X_n$ be k independent and identically distributed standard random variables. Let Y be the sum of the squares of the X_i.

$$Y = X_1^2 + X_2^3 + ... + X_k^2$$

Then Y has the chi-squared distribution with k degrees of freedom.

Exercise 7, page 63.

Exercise 8, page 63.