POLYTECHNIC UNIVERSITY
Department of Computer and Information Science

PROBABILITY AND STATISTICS

K. Ming Leung

Abstract: This chapter provides a brief coverage of the basics in the area of probability and statistics that are relevant to simulation.

Directory

Last Revision Date: February 4, 2003 Version 1.0

Table of Contents

1. Introduction

In a simulation study, probability and statistics are needed to understand how to model a probabilistic system, validate the simulation model, choose the input probability distributions, generate random samples from these distributions, perform statistical analysis of the simulation output data, and design the simulation experiment. This chapter provides only a minimal coverage of the basics in the area of probability and statistics that are relevant to simulation.

2. Discrete Random Variables and Their Properties

In probability theory, a process whose outcome cannot be predicted with certainty is called an *experiment*. The set of all possible outcomes of an experiment is called the *sample space* S. The outcomes themselves are specified as *sample points* in the sample space. In the *discrete* case, the number of sample points is countable (meaning that the number of sample points can be put in a one-to-one correspondence with the set of positive integers).

For example, the process of tossing a die is an experiment, whose outcome is either $1, 2, 3, 4, 5$ or 6, and so the sample space is

$$S = \{1, 2, 3, 4, 5, 6\} . \qquad (1)$$

If the experiment consists of rolling a pair of dice, then

$$S = \{(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (2, 1), \ldots, (6, 6)\} , \qquad (2)$$

where (i, j) means that i and j appeared on the first and second die, respectively.

A *random variable* X is defined as a function or rule that assigns a real number, x, to each point in the sample space S. A *random* *variable* X is discrete if it can only take on a countable (possibly infinite) number of possible values, x_1, x_2, ...

In the first experiment, if X is the random variable corresponding to the face value of the die, then the possible values that X can take on are $1, 2, 3, 4, 5, 6$. In the second example, if X is the random variable corresponding to the sum of the two die, then X assigns the value x of 7 to the outcome $(4, 3)$.

We will adopt the common notation of using capital letters such as X, Y, Z to denote random variables, and lower case letters such as x, y, z for the values taken on by these variables.

The *cumulative distribution function* $F(x)$ of the random variable X is defined for any real values of x by

$$F(x) = P(X \le x) \quad -\infty < x < \infty, \tag{3}$$

where $P(X \le x)$ is the probability associated with the event $\{X \le x\}$. Thus $F(x)$ is the probability that in an experiment the random variable X takes on a value no larger than x.

For example a life-insurance company, the age of a person, x, is usually treated as discrete. Let X be the random variable denoting the age when a certain person dies. Then the company is interested in the cumulative distribution function $P(X \leq x)$, which represents the probability that the person does not live past age x.

A cumulative distribution function $F(x)$ has the following properties:

1. $0 \leq F(x) \leq 1$ for all x, because $F(x)$ is a probability.
2. $F(x)$ is non-decreasing [*i.e.* if $x_1 < x_2$, then $F(x_1) \leq F(x_2)$].
3. $F(-\infty) = 0$ and $F(\infty) = 1$.

These properties can be easily proved from the definition of $F(x)$.

The probability that X takes on a particular value x is given by the probability mass function

$$p(x) = P(X = x), \tag{4}$$

which must satisfy the following 2 conditions:

1. $p(x) \geq 0$ for all x, since probabilities cannot be negative. We usually omit any value of x which has $p(x) = 0$, in that case we can assume $p(x) > 0$ for all x.

2. $\sum_x^\infty p(x) = 1$, where the sum is to be carried out over all the possible values of x. Each allowed value of x contributes exactly one term to the sum. This is referred to as the normalization condition.

Example 1: Consider a discrete random variable X that describe the outcome of tossing a die. The possible values of x and their respective probabilities are given by

x	1	2	3	4	5	6
$p(x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

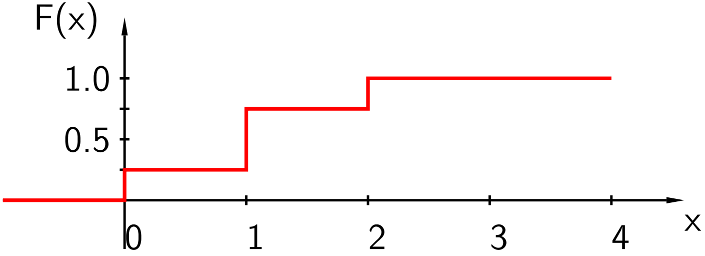
For any x, the cumulative distribution function can be written as

$$F(x) = \sum_{x' \leq x} p(x'). \tag{5}$$

Example 2: Consider a discrete random variable X that obeys the following distribution:

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

To obtain the cumulative probability distribution function $F(x)$, notice that $F(x)$ must be 0 for $x < 0$ since X does not have a negative value. At $x = 0$ we see that $F(x) = P(X \leq 0) = P(X = 0) = \frac{1}{4}$. $F(x)$ continues to have this value till $x = 1$. At $x = 1$, $F(x) = P(X \leq 1) = P(X = 0) + P(X = 1) = \frac{1}{4} + \frac{1}{2} = \frac{3}{4}$. $F(x)$ continues to have this value till $x = 2$. At $x = 2$, $F(x) = P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = \frac{1}{4} + \frac{1}{2} + \frac{1}{4} = 1$. Clearly $F(x) = 1$ for $x \geq 2$. So $F(x)$ remains constant except at each possible values of x where the function jumps up by an amount $p(x)$.



The *expected value* of X is defined as

$$E(X) = \sum_x xp(x) = \mu, \tag{6}$$

and represents the mean or average value.

For example 2, we have

$$E(X) = 0 \times \frac{1}{4} + 1 \times \frac{1}{2} + 2 \times \frac{1}{4} = 1. \tag{7}$$

It can be shown [1] that if $g(X)$ is a real-valued function of X, then the expected value of $g(X)$ is given by

$$E(g(X)) = \sum_x g(x)p(x). \tag{8}$$

However, unless g is a linear function,

$$E(g(X)) \neq g(E(X)). \tag{9}$$

Example 3: Consider next a discrete random variable X that obeys the following distribution:

x	-1.0	0.5	0	1.5	4.0
$p(x)$	0.1	0.2	0.2	0.4	0.1

$$E(X) = -1 \times 0.1 + 0.5 \times 0.2 + 0 \times 0.2 + 1.5 \times 0.4 + 4.0 \times 0.1 = 1. \quad (10)$$

Therefore the expected value is 1 just like example 2. However the two distributions are quite different. The values in examples 3 are distributed over a wider range than in example 2. A measure of the width of a given distribution is the *variance* of a random variable, which is defined by

$$V(X) = E((X - \mu)^2). \quad (11)$$

It measures the sum of the squares of the deviation of the possible values of X from its mean value. Notice that the variance is the expected value of the square of a quantity and therefore cannot be negative. It takes on its lowest possible value of zero for the case where X can take on only a single value. In that case, the distribution has no spread at all, and the variable X is really not random.

For example 2, since the mean is 1, we have

$$V(X) = (0-1)^2 \times \frac{1}{4} + (1-1)^2 \times \frac{1}{2} + (2-1)^2 \times \frac{1}{4} = \frac{1}{2}. \quad (12)$$

However, for example 3, the variance

$$\begin{aligned} V(X) &= (-1-1)^2 \times 0.1 + (0.5-1)^2 \times 0.2 + (0-1)^2 \times 0.2 \quad (13) \\ &+ (1.5-1)^2 \times 0.4 + (4-1)^2 \times 0.1 = 2.55, \end{aligned}$$

is substantially larger, as expected.

Another useful quantity is the *standard deviation* σ defined by

$$\sigma = \sqrt{V(X)}. \quad (14)$$

The following results involving the expected values can be easily proved: For an constant c

$$E(c) = \sum_x cp(x) = c \sum_y p(y) = c. \quad (15)$$

For any constant c and any function of the random variable X, $g(X)$,

$$E(cg(X)) = \sum_x cg(x)p(x) = c\sum_y g(x)p(y) = cE(g(X)). \tag{16}$$

If $g_1(X), g_2(X), \dots$ are functions of X, then

$$\begin{aligned} E(g_1(X) + g_2(X) + \dots) &= \sum_x (g_1(x) + g_2(x) + \dots)p(x) \qquad (17) \\ &= \sum_x g_1(x)p(x) + \sum_x g_2(x)p(x) + \dots \\ &= E(g_1(x)) + E(g_2(x)) + \dots . \end{aligned}$$

Using the above results, we have

$$\begin{aligned} V(X) &= E((X-\mu)^2) = E(X^2 - 2\mu X + \mu) \qquad (18) \\ &= E(X^2) + E(-2\mu X) + E(\mu^2) \\ &= E(X^2) - 2\mu E(X) + \mu^2. \end{aligned}$$

Thus we obtain the following important result for the variance:

$$V(X) = E(X^2) - \mu^2. \tag{19}$$

Next we consider the effect of shifting the random variable on the variance by computing $V(X+c)$ where c is a constant. First, we have $V(X+c) = E((X+c)^2) - (E(X+c))^2$. The second term $E(X+c) = E(X) + E(c) = \mu + c$. The first term $E((X+c)^2) = E(X^2) + 2cE(X) + c^2 = E(X^2) + 2c\mu + c^2$. Therefore $V(X+c) = E(X^2) + 2c\mu + c^2 - (\mu+c)^2 = E(X^2) - \mu^2 = V(X)$. This result is to be expected because a rigid shift in the distribution should not change the breath of the distribution.

Another result is

$$\begin{aligned} V(cX) &= E((cX)^2) - (E(cX))^2 \\ &= c^2(E(X^2) - (E(X))^2) = c^2 V(X). \end{aligned} \tag{20}$$

3. Continuous Random Variables and Their Properties

If the cumulative distribution function of a random variable X defined by

$$F(X) = P(X \leq x) \quad \mathit{for} \quad -\infty < x < \infty \tag{21}$$

is a continuous function of x, then X is a *continuous random variable*.

The *probability density function* $f(x)$ for X is defined as the derivative of the cumulative distribution function:

$$f(x) = \frac{dF}{dx}, \tag{22}$$

wherever the derivative exists. By writing

$$dF = f(x)dx, \tag{23}$$

changing the variable from x to t to get

$$dF = f(t)dt, \tag{24}$$

and integrating both sides of the equation from $t = -\infty$ to $t = x$, we have

$$\int_{-\infty}^{x} dF = \int_{-\infty}^{x} f(t)dt. \tag{25}$$

The integral on the left-hand side can be evaluated to give $F(x) - F(-\infty)$. Since $F(-\infty) = 0$, we now have the cumulative distribution function expressed in terms of the probability density function:

$$F(x) = \int_{-\infty}^{x} f(t)dt. \tag{26}$$

Thus for any x_0, the area underneath the curve of $f(x)$ from $-\infty$ to x_0 is precisely $F(x_0)$. We refer $f(x)$ as the probability *density* function because its integral gives a probability. This is analogous to the fact that in Physics, the integral of the mass density over a certain region gives the mass of that region.

The probability density function $f(x)$ obeys the following properties:

1. Since $F(x)$ is a monotonic non-decreasing function of x, we must have $f(x) = dF/dy \geq 0$ for any value of x.
2. Since $F(\infty) = 1$, therefore $\int_{-\infty}^{\infty} f(x)dx = 1$.

Since $F(x) = P(X \leq x)$ gives the probability that $X \leq x$, the probability that X has a value in the interval $a \leq y \leq b$ is

$$\begin{aligned} P(a \leq X \leq b) &= P(X \leq b) - P(X \leq a) = F(b) - F(a) \qquad (27) \\ &= \int_{-\infty}^{b} f(x)dx - \int_{-\infty}^{a} f(x)dx = \int_{a}^{b} f(x)dx. \end{aligned}$$

Graphically $P(a \leq X \leq b)$ is represented by the area of the region below the $f(x)$ curve from a to b. If we set $a = b = x_0$, we see the probability $P(X = x_0)$ of finding X having any particular value x_0 is exactly zero. This is in sharp contrast to the case of discrete random variables where $P(X = x_0)$ is non-zero for certain discrete values of x_0.

Supposed that a continuous random variable X has a probability distribution function

$$F(x) = \begin{cases} 0, & x < 0 \\ x, & 0 \leq x \leq 1 \\ 1, & x \geq 1. \end{cases} \qquad (28)$$

The probability density function is therefore given by

$$f(x) = \frac{dF}{dx} = \begin{cases} 0, & x < 0 \\ 1, & 0 < x < 1 \\ 1, & x \geq 1. \end{cases} \tag{29}$$

Notice that it is undefined at $x = 0$ and $x = 1$.

In general the probability distribution function for a continuous random variable must be continuous everywhere, but the probability density function need not be everywhere continuous.

The *expected* value of a continuous variable X is defined in terms of an integral by

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx = \mu, \tag{30}$$

provided that the integral exists. Since the integral is a limiting form of a sum, consequently the results that we obtained for the discrete case are also valid for the continuous case. Therefore we have

$$\begin{aligned} E(g(X)) &= \int_{-\infty}^{\infty} g(x)f(x)dx \\ E(c) &= c \\ E(cg(X)) &= cE(g(X)) \\ E(g_1(X) + g_2(X) + \dots) &= E(g_1(X)) + E(g_2(X)) + \dots . \end{aligned} \tag{31}$$

3.1. Uniform Probability Distribution

- **Expected Value**
- **Variance**
- **Mapping to Different Domain**

3.2. Normal Probability Distribution

The *normal probability distribution* is another often encountered distribution. It is especially important in understanding how computer simulations based on random sampling work.

A random variable X has a normal probability distribution if it has a probability density function

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \quad -\infty < x < \infty, \tag{32}$$

with real parameters μ and σ. In addition, σ must be positive. As will be shown below, the expected value $E(X)$ is μ, and the standard deviation is σ. Notice that $f(x)$ is properly normalized since

$$\begin{aligned} \int_{-\infty}^{\infty} f(x)dx &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left[-\frac{t^2}{2}\right] dt = 1, \end{aligned} \tag{33}$$

where we have a change of variable from x to $t = (x-\mu)/\sigma$.

The probability density function has the form of a bell-shape curve with a single peak at $x = \mu$. The value at the peak is $1/\sqrt{2\pi}\sigma$, and so the smaller σ is the higher is the peak.

Letting $x = \mu + h$ be the point at which $f(x)$ is equal to half its peak value, inserting it into Eq. (32) and setting f to a half, yields

$$\exp\left[-\frac{h^2}{2\sigma^2}\right] = \frac{1}{2}. \tag{34}$$

Solving for h, which is referred to as the half-width of the peak, gives $h = \pm\sqrt{2\ln 2}\,\sigma \approx \pm 1.177\,\sigma$. Decreasing σ decreases h and the peak becomes sharper. At the same time the height of the peak increases so that the total area underneath $f(x)$ remains unity.

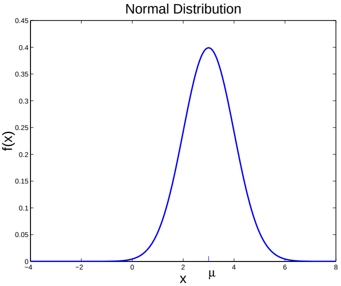


Figure 1: *Probability density function of a normal distribution. In this example,* $\mu = 3$ *and* $\sigma = 1$.

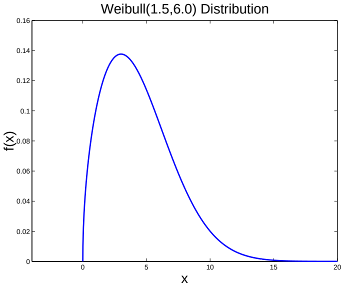


Figure 2: *Probability density function for the Weibull distribution with the shape parameter* $\alpha = 1.5$ *and the scale parameter* $\beta = 6.0$.

- **Expected Value**

The expected value can be easily calculated because

$$\begin{aligned} E(X) &= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} x \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx \\ &= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} (t+\mu) \exp\left[-\frac{t^2}{2\sigma^2}\right] dt \\ &= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} t\ \exp\left[-\frac{t^2}{2\sigma^2}\right] dt + \frac{\mu}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} \exp\left[-\frac{t^2}{2\sigma^2}\right] dt. \end{aligned} \tag{35}$$

The first integral is zero since the integrand is an odd function. The second integral is basically the same one encountered before for the normalization. The result shows that $E(X) = \mu$.

• The Variance

The variance is given by

$$\begin{aligned} V(X) &= E((X-\mu)^2) \\ &= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}(x-\mu)^2 \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx \\ &= \frac{2\sigma^2}{\sqrt{\pi}}\int_{-\infty}^{\infty} t^2 \exp\left[-t^2\right] dt \\ &= \sigma^2, \end{aligned} \tag{36}$$

where we made a change of the variable of integration from x to $t = (x-\mu)/(\sqrt{2}\sigma)$. Thus we see that σ is indeed the standard deviation.

- **Probability**

For any $x_2 \geq x_1$, the probability of find X having a value between x_1 and x_2 is

$$\begin{aligned} P(x_1 \leq X \leq x_2) &= \frac{1}{\sqrt{2\pi}\sigma} \int_{x_1}^{x_2} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx \\ &= \frac{1}{\sqrt{\pi}} \int_{t_1}^{t_2} \exp[-t^2] dt \\ &= \frac{1}{\sqrt{\pi}} \left\{ \int_0^{t_2} - \int_0^{t_1} \right\} \exp[-t^2] dt \\ &= \frac{1}{2} [\Phi(t_2) - \Phi(t_1)], \end{aligned} \tag{37}$$

where where we made a change of integration variable from x to $t = (x-\mu)/(\sqrt{2}\sigma)$, and so $t_1 = (x_1 - \mu)/(\sqrt{2}\sigma)$ and $t_2 = (x_2 - \mu)/(\sqrt{2}\sigma)$, and $\Phi(t)$ is the error function defined by

$$\Phi(t) = \frac{2}{\sqrt{\pi}} \int_0^t \exp[-x^2] dx. \tag{38}$$

Since the integrand is a positive function, $\Phi(t)$ is monotonic increasing function of t. At $t = 0$, clearly $\Phi(0) = 0$. At $t = \infty$, the integral can be evaluated exactly and one finds that $\Phi(\infty) = 1$. Moreover

$$\begin{aligned} \Phi(-t) &= \frac{2}{\sqrt{\pi}} \int_0^{-t} \exp[-x^2] dx \qquad (39) \\ &= -\frac{2}{\sqrt{\pi}} \int_0^{t} \exp[-u^2] du \\ &= -\Phi(t), \end{aligned}$$

where we made a change of integration variable from x to $u = -x$. Therefore $\Phi(t)$ is an odd function, which has the value of -1 at $t = -\infty$, rises monotonically to 0 at $t = 0$ and then to 1 at $t = \infty$. Here the change of integration variable is from x to $t = (x - \mu)/\sigma$.

- **The Rule of 3σs**

Letting $x_1 = \mu - 3\sigma$ and $x_2 = \mu + 3\sigma$ in Eq. (37), we see that

$$\begin{aligned} P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) &= \frac{1}{2}\left[\Phi\left(\frac{3}{\sqrt{2}}\right) - \Phi\left(-\frac{3}{\sqrt{2}}\right)\right] \\ &= \Phi\left(\frac{3}{\sqrt{2}}\right) \approx 0.9973. \end{aligned} \tag{40}$$

What this means is that in a single trial, it is highly unlikely (less than 0.3% of a chance) to obtain a value for the normal random variable X that differs from its mean value by more than 3σs.

- **The most probable error**

The most probable error, r, is defined so that

$$P(\mu - r \leq X \leq \mu + r) = \frac{1}{2}, \tag{41}$$

which becomes

$$\frac{1}{2}\left[\Phi\left(\frac{r}{\sqrt{2}\sigma}\right) - \Phi\left(-\frac{r}{\sqrt{2}\sigma}\right)\right] = \Phi\left(\frac{r}{\sqrt{2}\sigma}\right) = \frac{1}{2}, \tag{42}$$

by setting $t_{1,2} = [(\mu \pm r) - \mu]/(\sqrt{2}\sigma) = \mp r/(\sqrt{2}\sigma)$ in Eq. (37). Thus the most probable error is

$$r = \sqrt{2}\Phi^{-1}(1/2)\sigma \approx 0.6745\sigma, \tag{43}$$

where Φ^{-1} is the inverse function of Φ.

4. Sampling Distribution

We are interested in functions of N random variables $X_1, X_2, \ldots, X_N$ observed in a random sample selected from a population of interest. The variables $X_1, X_2, \ldots, X_N$ are assumed to be independent of each other and obeying the same common distribution. From a random sample of N observations, $x_1, x_2, \ldots, x_N$, we can estimate the population mean μ with the sample mean

$$\bar{x} = \frac{1}{N}\sum_{n=1}^{N} x_n. \tag{44}$$

The goodness of this estimate depends on the behavior of the random variables $X_1, X_2, \ldots, X_N$ and the effect that this behavior has on the random variable

$$\bar{X} = \frac{1}{N}\sum_{n=1}^{N} X_n. \tag{45}$$

Now suppose that the X_ns are independent, normally distributed variables, with a common mean $E(X_n) = \mu$ and a common variance $V(X_n) = \sigma^2$, for $n = 1, 2, \ldots, N$. It can be shown that $\bar{X}$ is a normally distributed random variable. Its mean is

$$E(\bar{X}) = E\left(\frac{1}{N}\sum_{n=1}^{N} X_n\right) = \frac{1}{N}\sum_{n=1}^{N} E(X_n) = \frac{1}{N}\sum_{n=1}^{N} \mu = \mu, \tag{46}$$

and its variance is

$$V(\bar{X}) = V\left(\frac{1}{N}\sum_{n=1}^{N} X_n\right) = \frac{1}{N^2}\sum_{n=1}^{N} V(X_n) = \frac{1}{N^2}\sum_{n=1}^{N} \sigma^2 = \frac{\sigma^2}{N}. \tag{47}$$

Therefore $\bar{X}$ has a mean $\bar{\mu} = \mu$ and a variance $\bar{\sigma}^2 = \sigma^2/N$.

4.1. Central Limit Theorem

It turns out that to a good approximation, $\bar{X}$ has a normal distribution with a mean $\bar{\mu} = \mu$ and a variance $\bar{\sigma}^2 = \sigma^2/N$ even when the X_ns are not normally distributed. The requirement is that the sample size has to be large enough and a few individual variables do not play too great a role in determining e sum. This is a consequence of the *Central Limit Theorem.*

In Nature, the behavior of a variable often depends on the accumulated effect of a large number of small random factors and so its behavior is approximately normal. This is why we call it "normal".

The Central Limit Theorem can be stated more formally as follows:

Let $X_1, X_2, \ldots, X_N$ be independent and identically distributed (iid) random variables with mean $E(X_n) = \mu$ and variance $V(X_n) = \sigma^2$, then as $N \to \infty$, the variable

$$\bar{X} = \frac{1}{N} \sum_{n=1}^{N} X_n \tag{48}$$

has a distribution function that converges to a normal distribution function with mean $\bar{\mu} = \mu$ and variance $\bar{\sigma}^2 = \sigma^2/N$.

In reality, the X_ns need not have the same identical distribution function, and N needs not be extremely large. Often $N = 5$ or more is large enough for a reasonably good approximate normal behavior for $\bar{X}$.

5. General Scheme of Monte Carlo Simulation Method

In a Monte Carlo simulation, we compute a quantity of interest, say q, by randomly sampling a population. We start by identifying a random variable X whose mean μ is supposed to be given by q, the quantity we want to compute. Let us denote the variance of X by σ^2.

Suppose we sample the population N times where N is supposed to be a large number. So we consider N independent random variables $X_1, X_2, \ldots, X_N$, each obeying the same distribution as X, with the same mean μ and variance σ^2. From the Central Limit Theorem, the variable

$$\bar{X} = \frac{1}{N}\sum_{n=1}^{N} X_n \tag{49}$$

is approximately normal with a mean $\bar{\mu} = \mu$ and a variance $\bar{\sigma}^2 = \sigma^2/N$.

Applying the Rule of the 3σs to $\bar{X}$, we see that

$$\begin{aligned} P(\mu - 3\bar{\sigma} \le \bar{X} \le \mu + 3\bar{\sigma}) &= P\left(\mu - \frac{3\sigma}{\sqrt{N}} \le \bar{X} \le \mu + \frac{3\sigma}{\sqrt{N}}\right) \quad (50) \\ &= P\left(\left|\frac{1}{N}\sum_{n=1}^{N} X_n - \mu\right| \le \frac{3\sigma}{\sqrt{N}}\right) \\ &\approx 0.9973. \end{aligned}$$

Therefore we can approximate the value of q by the sample mean

$$\bar{x} = \frac{1}{N}\sum_{n=1}^{N} x_n. \quad (51)$$

In all likelihood (with a 99.73% chance for being correct), the error does not exceed $3\sigma/\sqrt{N}$. In practice, the bound given by $3\sigma/\sqrt{N}$ is too loose (overly cautious), it is commonly replaced by the most probable error of $0.6745\sigma/\sqrt{N}$.

5.1. General Remarks on Monte Carlo Methods

Based on the above mathematical results, we can make the following three general remarks concerning Monte Carlo simulations:

1. Note that the most probable error decreases with increasing N as $N^{-\frac{1}{2}}$. That means that our estimation for the value of q from the sample mean gets better and better if more and more trials or samples are used in the simulation. However, the dependence of the error on N to the power of $-1/2$ means that the convergence of the Monte Carlo method is in general very slow. So for example, if we want to further decrease the probable error by a factor of 10, we will need to use 100 times more samples.

2. The most probable error is proportional to σ. Different ways of sampling the population lead to different values of σ. In the next month or so, we will find different sampling schemes that will reduce the value of σ and therefore reduce the probable error for the same number of trials, N. These techniques are referred to as *variance reduction techniques*.

3. The strong point about the Monte Carlo methods is that the

methods are in general very flexible and can be readily adapted to treat rather complex models.

References

[1] Wackerly, Mendenhall and Scheaffer, "Mathematical Statistics with Applications", 5th edition, (Duxburg, 1996).

[2] A proof can be found on page 82 of Wackerly, Mendenhall and Scheaffer, "Mathematical Statistics with Applications", 5th edition, (Duxburg, 1996).

9