CS445: Review of Probability and Statistics

Travis Desell

Averill M. Law, Simulation Modeling & Analysis, Chapter 4
Introduction
Introduction to Probability Review

Successfully performing simulation involves more than creating flowcharts, developing a computer program based on those flowcharts and then running it a few times to gather some data.

In general, a team developing a simulation should have at least one person well trained in probability and statistics, as probability and statistics are needed to:

1. Model a probabilistic system
2. Validate the simulation model
3. Choose the input probability distributions
4. Generate random samples from these distributions
5. Perform statistical analysis of the simulation output data
6. Design the simulation experiments
Random Variables and their Properties
An *experiment* is a process whose outcome is not known with certainty. (If it was known, why would you need to run an experiment?)

The set of all possible outcomes of an experiment is a *sample space* and we will denote that as $S$.

The outcomes themselves are *sample points* in the sample space, $S$. 

Some Definitions
Some Definitions

If we have the standard coin toss experiment, then:

\[ S = \{H, T\} \]

where the symbol \{\} means the set consisting of H as heads, and T as tails. So the sample space of the coin toss experiment, S, is \{H, T\}, because the only two outcomes are heads or tails (barring cracks where a coin could land on its side).

If the experiment is rolling a six sided die, then:

\[ S = \{1, 2, 3, 4, 5, 6\} \]

Where 1,2,3,4,5,6 are the different sides of the die being rolled.
Some Definitions

A random variable is a function (or rule) that assigns a real number (any number greater than -infinity and less than infinity) to each point in the sample space S.

Consider rolling two dice:

\[ S = \{(1, 1), (1, 2), (1, 3), \ldots (2, 3), (2,4), \ldots, (6,6)\} \]

The two numbers between the ()s are the faces of the die that were rolled. If \( X \) is a random variable corresponding to the sum of two dice, then \( X \) assigns the value 7 to outcome (4,3), or (3,4).

Consider tossing two coins. If \( X \) is the random variable corresponding to the number of heads that occur, then \( X \) assigns the value 1 to the outcome (H,T) or (T,H).

We will denote random variables are denoted by capital letters, like \( X, Y, Z \) and the values they take by lowercase letters like \( x, y, z \).
The distribution function (or cumulative distribution function) $F(x)$ of the random variable $X$ is defined for each real number $x$ as follows:

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

Where $P(X \leq x)$ means the probability associated with the event $\{X \leq x\}$, i.e., $F(x)$ is the probability that, when the experiment is done, the random variable $X$ will have taken on a value no larger than the number $x$. 
Some Definitions

A distribution function has the following properties:

1. $0 \leq F(x) \leq 1$ for all $x$. (It is a probability so it should be between 0 and 1).
2. $F(x)$ is non-decreasing, i.e., if $x_1 < x_2$ then $F(x_1) \leq F(x_2)$. (If it was decreasing in some areas, that would mean some values of $X$ had a negative probability).
3. $\lim_{x \to \infty} F(X) = 1$ and $\lim_{x \to -\infty} F(X) = 0$ (since $X$ takes only finite values)
A random variable is said to be \textit{discrete} if it can take on at most a countable number of values. Countable here means that the set of possible values can be put on a one-to-one correspondence with the set of positive integers. (Note that there are more real numbers than integers, and real numbers cannot be put on a one-to-one correspondence with integers — see Cantor if you want the heavy math for that).

The coin toss, double coin toss, die roll, and double die roll experiments are examples of countable random variables.

If the values the random variable $X$ can take are $x_1, x_2, x_3, \ldots, x_n$, then the probability that the discrete random variable $X$ takes on the value $x_i$ is:

$$p(x_i) = P(X = x_i) \quad \text{for } i = 1, 2, \ldots$$
A random variable is said to be *discrete* if it can take on at most a countable number of values. Countable here means that the set of possible values can be put on a one-to-one correspondence with the set of positive integers. (Note that there are more real numbers than integers, and real numbers cannot be put on a one-to-one correspondence with integers — see Cantor if you want the heavy math for that).

The coin toss, double coin toss, die roll, and double die roll experiments are examples of countable random variables.

If the values the random variable $X$ can take are $x_1, x_2, x_3, \ldots, x_n$, then the probability that the discrete random variable $X$ takes on the value $x_i$ is:

$$p(x_i) = P(X = x_i) \quad \text{for } i = 1, 2, \ldots$$

And the sum of all $p(x_i)$ must equal 1.
Some Definitions

All probability statements about $X$ can be computed (theoretically) from $p(x)$, which is called the *probability mass function* for the discrete random variable $X$. If $I = [a, b]$ where $a$ and $b$ are real numbers such that $a \leq b$, then:

$$P(X \in I) = \sum_{a \leq x_i \leq b} p(x_i)$$

Where the E like symbol means “contained in” and the summation means add all the $p(x_i)$, such that $a \leq x_i \leq b$.

The distribution function $F(x)$ for the discrete random variable $X$ is:

$$F(X) = \sum_{x_i \leq x} p(x_i) \text{ for all } -\infty < x < \infty$$
An Example

Given the inventory example from previous lectures (and section 1.5). The size of the demand for the product was a discrete random variable $X$ that takes the values 1, 2, 3 or 4 with probabilities $1/6$, $1/3$, $1/3$, $1/6$.

If we want to determine the probability of a 2 or a 3 occurring, we can use the probability mass function:

$$P(2 \leq X \leq 3) = p(2) + p(3) = \frac{1}{3} + \frac{1}{3} = \frac{2}{3}$$
An Example

Given the inventory example from previous lectures (and section 1.5). The size of the demand for the product was a discrete random variable $X$ that takes the values 1, 2, 3 or 4 with probabilities $\frac{1}{6}$, $\frac{1}{3}$, $\frac{1}{3}$, $\frac{1}{6}$.

The probability mass function is:
An Example

Given the inventory example from previous lectures (and section 1.5). The size of the demand for the product was a discrete random variable $X$ that takes the values 1, 2, 3 or 4 with probabilities $1/6$, $1/3$, $1/3$, $1/6$.

The distribution function is:

$$
\begin{align*}
0 & \quad 0.1667 \quad 0.5 \quad 1 \quad 1.6667 \\
1 & \quad 0.1667 \quad 0.3333 \quad 0.5 \quad 1
\end{align*}
$$
A random variable \( X \) is said to be *continuous* if there exists a non-negative function \( f(x) \) such that for any set of real numbers \( B \) (e.g., \( B \) could be all real number between 1 and 2):

\[
P(X \in B) = \int_B f(x)\,dx \quad \text{and} \quad \int_{-\infty}^{\infty} f(x)\,dx = 1
\]

The total area under \( f(x) \) is 1. Also, if \( X \) is a non-negative random variable (which is usually the case in simulations) the second range of integration is from 0 to infinity.

All probability statements about \( X \) can (in principle) be computed from \( f(x) \), which for continuous random variables is called the *probability density function*. 
Continuous Random Variables

For a discrete random variable $X$, $p(x)$ is the actual probability associated with the value $x$. However, $f(x)$ is not the probability that a continuous random variable $X$ equals $x$. For any real number $x$:

$$P(X = x) = P(X \in [x, x]) = \int_{x}^{x} f(y) \, dy = 0$$

The integral under a point (i.e., the values under the curve from $x$ to $x$) is 0. This means the probability associated with any particular value $x$ is 0. This makes sense because there are an infinite number of possible values that $X$ can assume as it is continuous.
Continuous Random Variables

Instead, we use the following, where $x$ is any number and $\Delta x > 0$ then:

$$ P(X \in [x, x + \Delta x]) = \int_{x}^{x+\Delta x} f(y) dy $$

Where is the area under $f(x)$ between $x$ and $x + \Delta x$ (see my drawing).
Continuous Random Variables

The distribution function $F(x)$ for a continuous random variable $X$ is given by:

$$F(x) = P(X \in (-\infty, x]) = \int_{-\infty}^{x} f(x)\,dy \text{ for all } -\infty < x < \infty$$

Thus (under some mild assumptions), $f(x) = F'(x)$, i.e., the $f(x)$ equals the derivative of $F(x)$. Further, if $I = [a,b]$ for any real numbers $a$ and $b$ such that $a < b$, then:

$$P(X \in I) = \int_{a}^{b} f(y)\,dy = F(b) - F(a)$$

Where the last equality is an application of the fundamental theorem of calculus, since $F'(x) = f(x)$. This also makes sense if you graph it out.
Continuous Random Variables

Remember, $f(x)$ is the probability density function, and $F(x)$ is the distribution function (or cumulative distribution function).

For an example, a uniform random variable on the interval $[0, 1]$, has the following probability density function:

$$f(x) = \begin{cases} 1 & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Further, if $0 \leq x \leq 1$ then:

$$F(x) = \int_{0}^{x} f(y) \, dy = \int_{0}^{x} 1 \, dy = x$$
Continuous Random Variables

Remember, \( f(x) \) is the probability density function, and \( F(x) \) is the distribution function (or cumulative distribution function).

The probability density function for a uniform random variable is:
Continuous Random Variables

Remember, \( f(x) \) is the probability density function, and \( F(x) \) is the distribution function (or cumulative distribution function).

The cumulative distribution function for a uniform random variable is:
Continuous Random Variables

The probability density function and cumulative distribution function for an exponential random variable are (such as used for inter-arrival and service times in the queueing example and inter demand times in the inventory example):

![Probability Density Function](image1)

![Cumulative Distribution Function](image2)
Multiple Random Variables

So far, we’ve only dealt with one random variable at a time. What happens when we need to deal with multiple random variables? In simulations, we typically have multiple (n) random variables, e.g., $X_1, X_2, X_3, \ldots X_n$.

So, for simplicity, if we have two random variables $X$ and $Y$, then we can let:

$$p(x, y) = P(X = x, Y = y)$$

for all $x, y$

So $p(x,y)$ is the probability that of the value $x$ AND $y$ occur. $p(x,y)$ is called the joint probability mass function of $X$ and $Y$. 
Multiple Random Variables

In this case, X and Y are *independent* if:

\[ p(x, y) = p_x(x)p_y(y) \text{ for all } x, y \]

Where:

\[ p_X(x) = \sum_{\text{all } y} p(x, y) \]

\[ p_Y(y) = \sum_{\text{all } x} p(x, y) \]

Are the (marginal) probability mass functions of X and Y. Note that this makes sense. It’s saying that the probability of choosing \( y \) is the same for any \( x \); and vice versa. \( P_X(x) \) is the probability of X being \( x \), \( P_Y(y) \) is the probably of Y being \( y \).
Multiple Random Variables

For an example:

Suppose that $X$ and $Y$ are jointly discrete random variables with

$$p(x, y) = \begin{cases} \frac{xy}{27} & \text{for } x = 1, 2 \text{ and } y = 2, 3, 4 \\ 0 & \text{otherwise} \end{cases}$$

Then:

$$p_X(x) = \sum_{y=2}^{4} \frac{xy}{27} = \frac{x}{3} \quad \text{for } x = 1, 2$$

$$p_Y(y) = \sum_{x=1}^{2} \frac{xy}{27} = \frac{y}{9} \quad \text{for } y = 2, 3, 4$$

Since $p(x,y) = \frac{xy}{27} = p_X(x)p_Y(y)$ for all $x, y$, the random variables $X$ and $Y$ are independent.
Multiple Random Variables

For another example:

Suppose 2 cards can be dealt from a deck of 52 without replacement. Let the random variables \( X \) and \( Y \) be the number of aces and kings that occur, both which have possible values of 0, 1, 2. It can be shown that:

\[
p_X(1) = p_Y(1) = 2 \left( \frac{4}{52} \right) \left( \frac{48}{51} \right)
\]

And (because of no replacement):

\[
p(1, 1) = 2 \left( \frac{4}{52} \right) \left( \frac{4}{51} \right)
\]

Since:

\[
p(1, 1) = 2 \left( \frac{4}{52} \right) \left( \frac{4}{51} \right) \neq 4 \left( \frac{4}{52} \right)^2 \left( \frac{48}{51} \right)^2
\]

it follows that \( X \) and \( Y \) are not independent.
Multiple Random Variables

The random variables $X$ and $Y$ are *jointly continuous* if there exists a non-negative function $f(x,y)$ called the *joint probability density function* of $X$ and $Y$, such that for all sets of real numbers $A$ and $B$:

$$P(X \in A, Y \in B) = \int_B \int_A f(x, y) \, dx \, dy$$

In this case, $X$ and $Y$ are independent if:

$$f(x, y) = f_X(x)f_Y(y) \text{ for all } x, y$$

Where:

$$f_X(x) = \int_{-\infty}^{\inf} f(x, y) \, dy$$

$$f_Y(y) = \int_{-\infty}^{\inf} f(x, y) \, dx$$

are the (marginal) probability density functions of $X$ and $Y$, respectively.
Multiple Random Variables

An example:

Suppose that $X$ and $Y$ are jointly continuous random variables with:

$$f(x, y) = \begin{cases} 
24xy & \text{for } x \geq 0, y \geq 0, \text{ and } x + y \leq 1 \\
0 & \text{otherwise}
\end{cases}$$

Then:

$$f_X(x) = \int_0^{1-x} 24xydy = 12xy^2\big|_0^{1-x} = 12x(1-x)^2 \text{ for } 0 \leq x \leq 1$$

$$f_Y(y) = \int_0^{1-y} 24xydy = 12yx^2\big|_0^{1-y} = 12y(1-y)^2 \text{ for } 0 \leq y \leq 1$$

Since:

$$f\left(\frac{1}{2}, \frac{1}{2}\right) = 6 \neq \left(\frac{3}{2}\right)^2 = f_X\left(\frac{1}{2}\right)f_Y\left(\frac{1}{2}\right)$$

$X$ and $Y$ are not independent.
Intuitively, the random variables $X$ and $Y$ (whether discrete or continuous) are independent if knowing the value that one random variable takes on tells us nothing about the distribution of the other. Also, if $X$ and $Y$ are not independent, we say that they are dependent.
Again consider the case of $n$ random variables $X_1, X_2, \ldots, X_n$. We can discuss some characteristics of the single random variable $X_i$ and some measures of the dependence that may exist between two random variables $X_i$ and $X_j$.

The *mean* or *expected value* of the random variable $X_i$ (where $i = 1, 2, \ldots, n$) will be denoted by $\mu_i$ (the $u$ is a mu) or $E(X_i)$ and is defined by:

$$
\mu_i = \begin{cases} 
\sum_{j=1}^{\infty} x_j p_{X_i}(x_j) & \text{if } X_i \text{ is discrete} \\
\sum_{-\infty}^{\infty} x f_{X_i}(x) dx & \text{if } X_i \text{ is continuous}
\end{cases}
$$

The mean is one measure of central tendency in the sense that it is the center of gravity.
Mean, Median, Mode, Variance

For the demand-size random variable, the mean is given by:

$$\mu = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{3} + 3 \cdot \frac{1}{3} + 4 \cdot \frac{1}{6} = \frac{5}{2}$$

And the mean of the a uniform random variable is (remember, f(x) is 1 from 0 to 1):

$$\mu = \int_{0}^{1} x f(x) dx = \int_{0}^{1} x dx = \frac{1}{2}$$
Let $c$ or $c_i$ denote a constant (a real number). Then the following are important properties of means:

$$E(cX) = cE(X)$$

$$E\left(\sum_{i=1}^{n} c_i X_i\right) = \sum_{i=1}^{n} c_i E(X_i) \text{ even if the } X_i \text{'s are dependent}$$
The median \( x_{0.5} \) of the random variable \( X_i \), which is an alternative measure of central tendency, is defined to be the smallest value of \( x \) such that \( F_{X_i}(x) \geq 0.5 \). If \( X_i \) is a continuous random variable, then \( F_{X_i}(X_{0.5}) = 0.5 \). The median may be a better measure of central tendency than the mean when \( X_i \) can take on very large or very small values, since extreme values can greatly affect the mean even if they are very unlikely to occur. This does not effect the median.
The *mode* $m$ of a continuous (discrete) random variable $X_i$, is another measure of central tendency. It is defined as the value of $x$ that maximizes $f_{X_i}(x)[p_{X_i}(x)]$. Note that the mode may not be unique for some distributions.
The variance of a random variable $X_i$ is denoted by lower case sigma $\sigma^2_i$, or Var($X_i$) and is defined by:

$$\sigma^2_i = E[(X_i - \mu_i)^2] = E(X_i^2) - \mu_i^2$$

The variance is the measure of the dispersion of a random variable about it’s mean. The larger the variance, the more likely the random variable is to take on values far from the mean.
For the demand-size random variable, the variance is computed as:

$$E(X^2) = 1^2 \frac{1}{6} + 2^2 \frac{1}{3} + 3^2 \frac{1}{3} + 4^2 \frac{1}{6} = \frac{43}{6}$$

$$Var(X) = E(X^2) - \mu^2 = \frac{43}{6} - \left(\frac{5}{2}\right)^2 = \frac{11}{12}$$

The variance is the measure of the dispersion of a random variable about its mean. The larger the variance, the more likely the random variable is to take on values far from the mean.
Mean, Median, Mode, Variance

For the uniform random variable on $[0, 1]$, the variance is computed as follows:

\[
E(X^2) = \int_0^1 x^2 f(x) dx = \int_0^1 x^2 dx = \frac{1}{3}
\]

\[
Var(X) = E(X^2) - \mu^2 = \frac{43}{6} - \left(\frac{5}{2}\right)^2 = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12}
\]
Mean, Median, Mode, Variance

The variance has the following (useful) properties:

\[ Var(X) \geq 0 \]

\[ Var(cX) = c^2 Var(X) \]

\[ Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i) \text{ if the } X_i \text{'s are independent (or uncorrelated)} \]
Standard Deviation

The *standard deviation* of the random variable $X_i$ is defined to be:

$$\sigma_i = \sqrt{\sigma_i^2}$$

The standard deviation can be given the most definitive interpretation when $X_i$ has a normal distribution. In particular, if $X_i$ has a normal distribution with mean $\mu_i$ and standard deviation $\sigma_i$, the probability that $X_i$ is between $\mu_i - 1.96 \times \sigma_i$ and $\mu_i + 1.96 \times \sigma_i$ is 0.95.
The **covariance** between random variables $X_i$ and $X_j$ (where $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, n$) is a measure of their (linear) dependence, is denoted by $C_{ij}$ or $\text{Cov}(X_i, X_j)$. Formally:

$$C_{ij} = E[(X_i - \mu_i)(X_j - \mu_j)] = E(X_iX_j) - \mu_i\mu_j$$

Note that covariances are symmetric, i.e., $C_{ij} = C_{ji}$ and that if $i = j$, then $C_{ij} = C_{ji} = \mu_i\mu_i = \mu_i^2$. 

**Covariance**
Covariance

Using the the jointly continuous random variables $X$ and $Y$ from the multiple random variables slides:

$$f(x, y) = \begin{cases} 
24xy & \text{for } x \geq 0, y \geq 0, \text{ and } x + y \leq 1 \\
0 & \text{otherwise}
\end{cases}$$

$$E(XY) = \int_0^1 \int_0^{1-x} xyf(x, y)dydx$$
$$= \int_0^1 x^2 (\int_0^{1-x} 24y^2 dy)dx$$
$$= \int_0^1 8x^2 (1 - x)^3 dx$$
$$= \frac{2}{15}$$

$$E(X) = \int_0^1 xf_X(x)dx = \int_0^1 12x^2 (1 - x)^2 dx = \frac{2}{5}$$

$$E(Y) = \int_0^1 yf_Y(y)dy = \int_0^1 12y^2 (1 - y)^2 dy = \frac{2}{5}$$
Covariance

Therefore:

\[ \text{Cov}(X, Y) = E(XY) - E(X)E(Y) = \frac{2}{15} - \frac{2}{5} \cdot \frac{2}{5} = -\frac{2}{75} \]

If \( C_{ij} = 0 \), then the random variables \( X_i \) and \( X_j \) are said to be uncorrelated. If \( X_i \) and \( X_j \) are independent random variables, then \( C_{ij} = 0 \); however in general the converse is not true.

However, if \( X_i \) and \( X_j \) are jointly normally distributed random variables with \( C_{ij} = 0 \), then they are also independent.
Significance of Covariance

If $C_{ij} > 0$ then $X_i$ and $X_j$ are said to be *positively correlated*. In this case, $X_i > \mu_i$ and $X_j > \mu_j$ tend to occur together; as do $X_i < \mu_i$ and $X_j < \mu_j$. So for positively correlated random variables, if one is large then the other is likely to be large as well (likewise if one is small the other is likely to be small).

If $C_{ij} < 0$ then $X_i$ and $X_j$ are said to be *negatively correlated*. In this case, $X_i > \mu_i$ and $X_j < \mu_j$ tend to occur together; as do $X_i < \mu_i$ and $X_j > \mu_j$. So for positively correlated random variables, if one is large then the other is likely to be small as well (likewise if one is small the other is likely to be large).
Correlation

If $X_1, X_2, \ldots, X_n$ are simulation output data, then we often need to know not only the mean $\mu_i$ and variance $\sigma^2_i$, for $i = 1, 2, \ldots, n$, but also a measure of the dependence between $X_i$ and $X_j$ for $i \neq j$.

However, $C_{ij}$ is not dimensionless (for example, if $X_i$ and $X_j$ are in units of minutes, then $C_{ij}$ is in units of minutes squared - which doesn’t make much sense). As a result, correlation $\rho_{ij}$ is used:

$$\rho_{ij} = \frac{C_{ij}}{\sqrt{\sigma_i^2 \sigma_j^2}}$$

Where $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, n$. Correlation is the primary measure of the (linear) dependence between $X_i$ and $X_j$. As the correlation will always have the same sign as the covariance, the same properties of negative and positive correlation hold. If $\rho_{ij}$ is close to $+1$, then $X_i$ and $X_j$ are highly positively correlated; and if $\rho_{ij}$ is close to $-1$ then $X_i$ and $X_j$ are highly negatively correlated.
Simulation Output Data and Stochastic Processes
Most simulations use random variables as input, which means that the simulation output data is also random — which means you need to be careful in how you interpret it.

A **stochastic process** is a collection of “similar” random variables ordered over time, which are all defined on a common sample space. The set of all possible values these random variables can take on are their **state space**. If the collection is $X_1, X_2, \ldots, X_n$ then we have a **discrete-time** stochastic process. If the collection is $\{X(t), t \geq 0\}$ (a function) then we have a **continuous-time** stochastic process.
Example 1:

The single server queueing system (or M/M/1 queue), with IID inter-arrival times $A_1, A_2, \ldots, A_n$ and service times $S_1, S_2, \ldots, S_n$ and customers served in a FIFO manner. Relative to the experiment of generating the random variates $A_i, A_2, \ldots$ and $S_1, S_2, \ldots$; one can define the discrete-time stochastic process of delays in queue $D_1, D_2, \ldots$ as follows:

$D_1 = 0$

$D_{i+1} = \max\{D_i + S_i - A_{i+1}, 0\}$ for $i = 1, 2, \ldots$

Thus, the simulation maps the input random variables (the $A_i$s and $S_i$s) into the output stochastic process $D_1, D_2, \ldots$ of interest. This state space is a set of non-negative real numbers, and $D_i, D_{i+1}$ are positively correlated.
Example 2:

For the same M/M/1 queueing system, let $Q(t)$ be the number of customers in the queue at time $t$. Then $\{Q(t), t \geq 0\}$ is a continuous time stochastic process with state space $\{0, 1, 2, \ldots\}$. 
Covariance-Stationary Processes

To draw inferences about an underlying stochastic process from a set of simulation output data, one must sometimes make assumptions about the stochastic process that might not be strictly true.

An example is to assume that a stochastic process is covariance-stationary, which means:

\[
\mu_i = \mu \text{ for } i = 1, 2, \ldots \text{ and } -\infty < \mu < \infty
\]

\[
\sigma_i^2 = \sigma \text{ for } i = 1, 2, \ldots \text{ and } \sigma^2 < \infty
\]

And that \( C_{i,i+j} = \text{Cov}(X_i, X_{i+j}) \) is independent of \( i \) for \( j = 1, 2, \ldots \)

This means that the mean and variance do not change over the time of the simulation, and that the covariance between two observations \( X_i \) and \( X_{i+j} \) only depends on the separation \( j \) (or the lag), not on the actual time values. It is also possible to define this for continuous-time processes in a similar way.
Covariance-stationary Processes

For a covariance-stationary process, the covariance and correlation between $X_i$ and $X_{i+j}$ can be be denoted as:

$$
\rho_j = \frac{C_{i,i+j}}{\sqrt{\sigma_i^2 \sigma_{i+j}^2}} = \frac{C_j}{\sigma^2} = \frac{C_j}{C_0}
$$

For $j = 0, 1, 2, \ldots$
If $X_1, X_2$ is a stochastic process beginning at time 0 in a simulation, then it is quite likely not to be covariance stationary. However, some simulations $X_{k+1}, X_{k+2}, \ldots$ will be approximately covariance stationary if $k$ is large enough, where $k$ is the length of the warmup period.
Correlation function $\rho_j$ of the process $D_1, D_2, \ldots$ for the M/M/1 queue. Where $\rho$ = the arrival rate / the service rate. Don’t confuse the $\rho$ describing the arrival rate/service rate with the correlation function (the book uses the same variable for both).
The correlations $\rho_j$ are positive and monotonically decrease to 0 as $j$ increases. In particular, $\rho_1 = 0.99$ for $\rho = 0.9$ and $\rho_1 = 0.78$ for $\rho = 0.5$ (the graph isn’t perfect). Furthermore, the convergence of $\rho_j$ to zero is considerably slower for $\rho$ (arrival/service) = 0.9 ($\rho_{50} = 0.69$, for example).
Covariance-stationary Processes

For the inventory system, we can calculate the total cost $C_i$ for each month. Note that $p_2$ is positive as in this system orders are typically placed every other month, incurring a large cost at each time. On the other hand, if $p_1$ is negative if an order is placed in a particular month (which will be large cost) then it is likely that no order will be placed in the next month, (which will be a small cost).
Estimation of Means, Variances and Correlations
Sample Mean

Suppose that $X_1, X_2, \ldots, X_n$ are IID random variables (observations) with a finite population mean $\mu$ and finite population variance $\sigma^2$. The sample mean is:

$$\bar{X}(n) = \frac{\sum_{i=1}^{n} X_i}{n}$$

Which is an unbiased estimator of $\mu$; this means $E[\bar{X}(n)] = \mu$. Intuitively, this means that if we perform a very large number of independent experiments, each resulting in an $\bar{X}(n)$ the average of those $\bar{X}(n)$’s will be $\mu$. 
Sample Variance

Similarly, the sample variance is:

\[ S^2(n) = \frac{\sum_{i=1}^{n} [X_i - \bar{X}(n)]^2}{n-1} \]

Which is an unbiased estimator of \( \sigma^2 \), since \( E[S^2(n)] = \sigma^2 \). Note that the estimators \( \bar{X}(n) \) and \( S^2(n) \) are sometimes denoted by \( \hat{\mu} \) and \( \hat{\sigma}^2 \):

\[ \bar{X}(n) = \hat{\mu} \]

\[ S^2(n) = \hat{\sigma}^2 \]
The difficulty with using the sample mean as an estimator of the mean without any additional information is that we have no way of knowing how close the sample mean is to the mean. Because the sample mean is a random variable with variance $\text{Var}[\hat{X}(n)]$, on one experiment the sample mean may be close to the actual mean, while on another experiment it may be quite different. (Figure 4.12)
Sample Mean/Variance Issues

The usual way to assess the precision of the sample mean as an estimator of the mean is to construct a confidence interval for the mean. The first step in this is to estimate the variance of the sample mean (step 3 can be done because the $X_i$s are independent):

$$\text{Var} [\bar{X} (n)] = \text{Var} \left( \frac{1}{n} \sum_{i=1}^{n} X_i \right)$$
$$= \frac{1}{n^2} \text{Var} \left( \sum_{i=1}^{n} X_i \right)$$
$$= \frac{1}{n^2} \sum_{i=1}^{n} \text{Var}(X_i)$$
$$= \frac{1}{n^2} n \sigma^2 = \frac{\sigma^2}{n}$$
Sample Mean/Variance Issues

It should be clear (intuitively) that the bigger the sample size $n$ of independent runs, the closer the sample mean should be to the actual mean. (Figure 4.13)

**FIGURE 4.13**
Distributions of $\bar{X}(n)$ for small and large $n$. 
Sample Mean/Variance Issues

An unbiased estimator of $\text{Var}[\bar{X}(n)]$ (the variance of the sample mean) can be obtained by replacing $\sigma^2$ by $S^2(n)$:

$$\hat{\text{Var}}[\bar{X}(n)] = \frac{S^2(n)}{n} = \frac{\sum_{i=1}^{n} [X_i - \bar{X}(n)]^2}{n(n-1)}$$

Note that the expression for the estimated variance of the sample mean has both an $n$ and an $n-1$ in the denominator. Also, since the $X_i$'s are independent, they are uncorrelated and thus $\rho_{ij} = 0$ for $j = 1, 2, \ldots, n-1$. 
Sample Mean/Variance Issues

However, as the author notes - in their experience, simulation output data is almost always correlated. Which means the above is not directly applicable to analyzing the simulation output data. This can lead to some problems.

If the random variables $X_1, X_2, \ldots X_n$ are from a covariance-stationary stochastic process, then the sample mean is still an unbiased estimator of the mean; however the sample variance is no longer an unbiased estimator of the variance. In fact:

$$E[S^2(N)] = \sigma^2[1 - 2 \sum_{j=1}^{n-1} \frac{(1 - j/n) \rho_j}{n - 1}]$$

Remember the rho$_j$ is the correlation between $X_i$ and $X_{i+j}$. Because of this, if the correlation is positive, as is usually the case, the sample variance will have a negative bias $E[S^2(n)] < \sigma^2$. This is important to realize because several simulation software products use the sample variance to estimate the variance of a set of simulation output data which could lead to serious errors in analysis.
Sample Mean/Variance Issues

There is also an issue with estimating the variance of the sample mean (as opposed to estimating the sample variance in the previous slide). Again, with $X_1, X_2, \ldots, X_n$ being a covariance stationary process, it can be shown that:

$$\text{Var}[\bar{X}(n)] = \sigma^2 \frac{1 + 2 \sum_{j=1}^{n-1} (1 - j/n) \rho_j}{n}$$

Therefore, if the variance of the sample mean is estimated from $S^2(n)/n$ (the correct expression in the IID case) then there are two sources of possible error. First, the bias in the sample variance as an estimator of the variance, and the negligence of the correlation terms in the above equation.

In fact, combining these two equations:

$$E\left[ \frac{S^2(n)}{n} \right] = \frac{\frac{1 + 2 \sum_{j=1}^{n-1} (1 - j/n) \rho_j}{n} - 1}{n - 1} \text{Var}[\bar{X}(n)]$$
Sample Mean/Variance Issues

\[ E\left[ \frac{S^2(n)}{n} \right] = \frac{1 + 2 \sum_{j=1}^{n-1} (1 - j/n) \rho_j}{n - 1} \bar{X}(n) \]

So in this case, if the correlation is positive (the usual case), \( \rho_j > 0 \), and therefore \( E[S^2(n)/n] < \text{Var}[\bar{X}(n)] \). Basically, if the correlation is positive, then estimating the variance using the equations what would be valid in the IID case results in a lower than actual variance.

This could lead to serious problems, for example, if you were using the variance to determine thresholds for safety tolerances.
Example 1

Suppose we have D1, D2, … D10 from the process of delays D1, D2, … for a covariance stationary M/M/1 queue with \( \rho = 0.9 \) (where this rho is the arrival rate / service rate). Substituting the true correlations \( \rho_j \) (where \( j = 1, 2, \ldots, 9 \)) we get:

\[
E[S^2(10)] = 0.0328 \sigma^2
\]

\[
E\left[\frac{S^2(10)}{10}\right] = 0.0034 Var[\bar{D}(10)]
\]

where:

\[
\sigma^2 = Var(D_i)
\]

\[
\bar{D}(10) = \frac{\sum_{i=1}^{10} D_i}{10}
\]

\[
S^2(10) = \frac{\sum_{i=1}^{10} [D_i - \bar{D}(10)]^2}{9}
\]

Thus on average, \( S^2(10)/10 \) is a gross underestimate of \( Var[D\hat{(}10)] \) and there is a likelihood of being overly optimistic about the closeness of \( D\hat{(}10) \) to \( \mu = E(D_i) \) (the mean being the expected value of \( D_i \)).
Sample Mean/Variance Issues

Sometimes you might want to estimate the rho\(_j\)'s (or C\(_j\)'s) (the correlations and covariances) from the data \(X_1, X_2, \ldots, X_n\). In this case it can be estimated as:

\[
\hat{p}_j = \frac{\hat{C}_j}{S^2(n)}
\]

\[
\hat{C}_j = \frac{\sum_{i=1}^{n-j} [X_i - \bar{X}(n)][X_{i+j} - \hat{X}(n)]}{n - j}
\]

Other estimators are also used (such as substituting the n-j with n in the denominator).

One difficulty with this however, is that it is biased, and has a large variance unless \(n\) is very large, and it is correlated with other correlation estimators, that is:

\[
Cov(\hat{p}_j, \hat{p}_k) \neq 0
\]
Sample Mean/Variance Issues

In particular, $\hat{\rho}_{n-1}$ will be a poor estimator of $\rho_{n-1}$ since it is base on a single product $[X_1 - \hat{X}(n)][X_n - \hat{X}(n)]$. Thus in general good estimates of the $\rho_j$’s will be difficult to obtain unless $n$ is very large and $j$ is relatively small to $n$.

(Figure 4.14) shows that if you plot the data from $D_1, D_2, \ldots, D_{100}$ and $p_j$ for $j = 1, 2, \ldots, 10$; then:

Also note that correlation estimates will not necessarily be 0 when the $X_i$’s are independent, since the estimator $\hat{\rho}_j$ is a random variable.

The correlation estimates are very poor — and they report significant under-correlation, which could lead to significant problems in actually using these results.
Sample Mean/Variance Issues

The big takeaway from this is that simulation data is often correlated, and therefore formulas from classical statistics based on IID observations cannot be used directly for estimating variance (however means are usually just fine).

However, the book discusses later in Chapter 9 to group the simulation output data into new “observations” which the formulas based on IID observations can be applied. Thus these formulas based on IID observations can be indirectly applied to analyzing simulation output data.
Confidence Intervals and Hypothesis Tests for the Mean
Confidence Intervals

Let $X_1, X_2, \ldots X_n$ be IID random variables with finite mean $\mu$ and finite variance $\sigma^2$. And assume that $\sigma^2 > 0$, which means the $X_i$’s are not degenerate random variables.

This section deals with constructing a confidence interval for $\mu$, and also the complementary problem of testing the hypothesis that $\mu = \mu_0$. 

Central Limit Theorem

Let $Z_n$ be a random variable: 

$$ Z_n = \frac{[\bar{X}(n) - \mu]}{\sqrt{\frac{\sigma^2}{n}}} $$

and $F_n(z)$ be the distribution function of $Z_n$ for a sample size of $n$; that is $F_n(z) = P(Z_n \leq z)$. Note that $\mu$ and $\sigma^2/n$ are the mean and variance of $X$-hat$(n)$ (the sample mean).

The central limit theorem is that:

$$ F_n(z) \rightarrow \Phi(z) \text{ as } n \rightarrow \infty $$

where:

$$ \Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-y^2/2} dy \text{ for } -\infty < z < \infty $$

Here, $\Phi(z)$ is the distribution function of a normal random variable with $\mu$ (mean) $= 0$ and $\sigma^2$ (variance) $= 1$. 
Central Limit Theorem

The theorem essentially states that if $n$ is sufficiently large, the random variable $Z_n$ will be approximately distributed as a standard normal random variable, regardless of the underlying distribution of the $X_i$’s.

It can also be shown that for large $n$ the sample mean $(X\text{-bar}(n))$ is approximately distributed as a normal random variable with mean $\mu$ and variance $\sigma^2/n$. 
Central Limit Theorem

The difficulty with using the theorem in practice is that the variance is generally unknown (which is required to calculate $Z_n$). However, as the sample variance $S^2(n)$ converges to $\sigma^2$ as $n$ gets large, it can be shown that the central limit theorem holds if we replace $\sigma^2$ by $S^2(n)$ in the equation for $Z_n$.

With this change, the theorem says that if $n$ is sufficiently large, the random variable

$$t_n = \frac{[\bar{X}(n) - \mu]}{\sqrt{(S^2(n)/n)}}$$

is approximately distributed as a standard normal random variable. It follows that for large $n$ that:

$$P(-z_{1-\alpha/2} \leq \frac{\bar{X}(n) - \mu}{\sqrt{S^2(n)/n}} \leq z_{1-\alpha/2} = P[\bar{X}(n) - z_{1-\alpha/2}\sqrt{\frac{s^2(n)}{n}} \leq \mu \leq \bar{X}(n) + z_{1-\alpha/2}\sqrt{\frac{s^2(n)}{n}}$$

$$\approx 1 - \alpha$$

where $z_{\{1-\text{alpha}/2\}}$ (for $0 < \text{alpha} < 1$) is the upper 1 - alpha/2 critical point for a standard normal random variable (see next slide).
Central Limit Theorem

Therefore, if \( n \) is sufficiently large, an approximate 100(1-\( \alpha \)) percent confidence interval for \( \mu \) (the mean) is given by:

\[
\bar{X}(n) \pm z_{1-\alpha/2} \sqrt{\frac{S^2(n)}{n}}
\]
Central Limit Theorem

For a given data set of $X_1, X_2, \ldots, X_n$ the lower confidence interval endpoint, $l(n, \alpha)$ and the upper confidence interval endpoint, $u(n, \alpha)$:

$$ l(n, \alpha) = \bar{X}(n) - z_{1-\alpha/2} \sqrt{S^2(n)/n} $$

$$ u(n, \alpha) = \bar{X}(n) + z_{1-\alpha/2} \sqrt{S^2(n)/n} $$

are just numbers (or specific realizations of random variables) and the confidence interval between them either contains $\mu$ (the mean) or it does not. There is nothing probabilistic about the single confidence interval $[l(n,\alpha), u(n,\alpha)]$ after the data have been collected and the interval’s endpoints have been given numerical values.
The correct interpretation to give to the confidence interval is:

If one constructs a very large number of independent 100(1-alpha) percent confidence intervals, each based on \( n \) observations, where \( n \) is sufficiently large, the proportion of these confidence intervals that contain (cover) the mean should be 1 - alpha. We call this proportion the *coverage* for the confidence interval.
Central Limit Theorem

The problem here lies in knowing what “n sufficiently large” means. It turns out the more skewed (non-symmetric) the underlying distribution of $X_i$’s, the larger the value of $n$ needed for the distribution of $t_n$ to be closely approximated by $\Phi(z)$.

If $n$ is too small, the actual coverage of a desired $100(1 - \alpha)$ percentage confidence interval will generally be less than $1 - \alpha$. That is why the confidence interval presented here is stated to only be approximate.
We can now develop an alternate confidence interval expression. If the $X_i$’s are normal random variables, the random variable:

$$ t_n = \frac{[\bar{X}(n) - \mu]}{\sqrt{(S^2(n)/n)}} $$

has a $t$ distribution with $n - 1$ degrees of freedom (df) and an exact (for any $n \geq 2$) $100(1-\alpha)$ percent confidence interval for $\mu$ (the mean) is given by:

$$ \bar{X}(n) \pm t_{n-1,1-\alpha/2} \sqrt{S^2(n)/n} $$

Where $t_{n-1,1-\alpha/2}$ is the upper $1-\alpha/2$ critical point for the $t$ distribution with $n - 1$ df.

These critical points are given in the book.
t confidence interval

The following is a plot of the $t$ distribution with 4 df and for the standard normal distribution. Note that the $t$ distribution is less peaked and has longer tails.

For any finite $n$, $t_{n-1,1-\alpha/2} > z_{1-\alpha/2}$. That is called the $t$ confidence interval.
Generally speaking, the more degrees of freedom there are, the closer the t distribution becomes to the normal distribution.
The quantity that we add and subtract from $X$-bar(n) (the sample mean) to construct the confidence interval is called the *half-length* of the confidence interval. It's a measure of how precisely the actual mean ($\mu$) is known.

It can be shown that increasing the sample size from $n$ to $4n$, then the half-length is decreased by a factor of approximately 2. (However these improvements are not linear, and there is diminishing returns for larger sample sizes).
In practice however, the distribution of the $X_i$’s will rarely be normal, and the t confidence interval will also only be approximate in terms of coverage. However, since:

$$t_{n-1, 1-\alpha/2} > z_{1-\alpha/2}$$

The t confidence interval will be larger than the previous standard normal confidence interval. For this reason, the t confidence interval is recommended for constructing confidence intervals. Note:

$$t_{n-1, 1-\alpha/2} \rightarrow z_{1-\alpha/2} \text{ as } n \rightarrow \infty$$

Which the previous graph displays.
Example

Suppose that the 10 observations: 1.20, 1.50, 1.68, 1.89, 0.95, 1.49, 1.58, 1.55, 0.50 and 1.09 are drawn from a normal distribution with an unknown mean $\mu$ and that we want to construct a 90% confidence interval for $\mu$. From this data:

$$\bar{X}(10) = 1.34 \text{ and } S^2(10) = 0.17$$

which results in the following confidence interval for $\mu$:

$$\bar{X}(10) \pm t_{9,0.95} \sqrt{\frac{S^2(10)}{10}} = 1.34 \pm 1.83 \sqrt{\frac{0.17}{10}} = 1.34 \pm 0.24$$

The value for $t_{9,0.95}$ was taken from a table (T.1 in the book). Subject to this interpretation, we can claim with 90% confidence that $\mu$ (the mean) is within the interval $[1.10, 1.58]$. 
Confidence Interval Coverage

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Skewness (v)</th>
<th>n = 5</th>
<th>n = 10</th>
<th>n = 20</th>
<th>n = 40</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>0</td>
<td>0.910</td>
<td>0.902</td>
<td>0.989</td>
<td>0.900</td>
</tr>
<tr>
<td>Exponential</td>
<td>2</td>
<td>0.854</td>
<td>0.878</td>
<td>0.870</td>
<td>0.890</td>
</tr>
<tr>
<td>Chi-square</td>
<td>2.83</td>
<td>0.810</td>
<td>0.830</td>
<td>0.848</td>
<td>0.890</td>
</tr>
<tr>
<td>Lognormal</td>
<td>6.18</td>
<td>0.758</td>
<td>0.769</td>
<td>0.842</td>
<td>0.852</td>
</tr>
<tr>
<td>Hyperexponential</td>
<td>6.43</td>
<td>0.584</td>
<td>0.586</td>
<td>0.682</td>
<td>0.775</td>
</tr>
</tbody>
</table>

The above table shows estimated coverages based on 500 experiments from random variables various distribution functions. Chi-square has 1 df (which means its a standard normal random variable squared), lognormal is \((e^Y\), where \(Y\) is a standard normal random variable\), and hyper exponential is:

\[
F(x) = 0.9F_1(x) + 0.1F_2(x)
\]

Where \(F_1(x)\) and \(F_2(x)\) are distribution functions of exponential random variables with means 0.5 and 0.55, respectively.
Confidence Interval Coverage

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Skewness (v)</th>
<th>n = 5</th>
<th>n = 10</th>
<th>n = 20</th>
<th>n = 40</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>0</td>
<td>0.910</td>
<td>0.902</td>
<td>0.989</td>
<td>0.900</td>
</tr>
<tr>
<td>Exponential</td>
<td>2</td>
<td>0.854</td>
<td>0.878</td>
<td>0.870</td>
<td>0.890</td>
</tr>
<tr>
<td>Chi-square</td>
<td>2.83</td>
<td>0.810</td>
<td>0.830</td>
<td>0.848</td>
<td>0.890</td>
</tr>
<tr>
<td>Lognormal</td>
<td>6.18</td>
<td>0.758</td>
<td>0.769</td>
<td>0.842</td>
<td>0.852</td>
</tr>
<tr>
<td>Hyperexponential</td>
<td>6.43</td>
<td>0.584</td>
<td>0.586</td>
<td>0.682</td>
<td>0.775</td>
</tr>
</tbody>
</table>

For example, the value for the exponential distribution with n = 10 was calculated by generating 10 observations from an exponential random variable with a known mean, and a 90% confidence interval was constructed, and it was then determined if the confidence interval contained the actual mean mu.

This experiment was done 500 times, and the value is the proportion of the 500 confidence intervals that contained the mean. Note that these are random values, so in some cases the coverage is worse when n increases, simply due to the randomness, even though we would expect otherwise.
As expected, as $n$ gets larger the coverage generally gets closer to 0.90, which follows from the central limit theorem. For a particular $n$, coverage also decreases as the skewness $v$ increases, where the skewness is:

$$ v = \frac{E[(X - \mu)^3]}{(\sigma^2)^{3/2}} $$

Skewness is a measure of symmetry, and is equal to 0 for a symmetric distribution such as the normal distribution. The greater the skewness, the larger $n$ needs to be to obtain satisfactory coverage.
Hypothesis Tests

Assume $X_1, X_2, \ldots X_n$ are normally distributed and that we would like to test the null hypothesis $H_0$ that the mean $\mu$ is equal to $\mu_0$, where $\mu_0$ is a fixed, hypothesized value for $\mu$.

Intuitively, if $|\bar{X}(n) - \mu_0|$ (the absolute value of the difference between the sample mean and our hypothesized $\mu$) is high, then $H_0$ is not likely to be true.

To develop a test with known statistical properties, a statistic is needed (some function of the $X_i$’s) whose distribution is known when $H_0$ is true. It follows from above that if $H_0$ is true then the statistic:

$$t_n = \frac{[\bar{X}(n) - \mu_0]}{\sqrt{(S^2(n)/n)}}$$

will have a t distribution with $n - 1$ degrees of freedom.
T Tests

Therefore, the form of the (two-tailed) hypothesis test for $\mu = \mu_0$ is:

$$\text{if } |t_n| \begin{cases} > t_{n-1,1-\alpha/2} & \text{reject } H_0 \\ \leq t_{n-1,1-\alpha/2} & \text{”accept” } H_0 \end{cases}$$

The portion of the real line that corresponds to rejection of $H_0$, namely, the set of all $x$ such that $|x| > t_{n-1,1-\alpha/2}$ is called the critical region of the test, and the probability that the statistic $t_n$ falls in the critical region given that $H_0$ is true, which is equal to $\alpha$, is called the level (or size) of the test. Typically, experimenters choose a level equal to 0.05 or 0.1 (corresponding to 95% and 90% confidence).

This hypothesis is called the $t$ test.
Type I vs Type II Error

When doing t tests, two types of error can be made.

Type I error is when one rejects $H_0$ when it is actually true. The probability of this error occurring is equal to the level alpha, and is therefore under the control of the person running the experiment. This can also be called a *false negative*.

On the other hand, Type II error is when one accepts $H_0$ when it is actually false. This can also be called a *false positive*. For a fixed level alpha and sample size $n$, this is based on what is actually true (what the mean really is) as opposed to $H_0$, so it may be unknown. The probability of Type II error is commonly denoted by Beta.

The power of the test (denoted by $\delta = 1 - \beta$) is equal to the probability of rejecting the test $H_0$ when it is false. Obviously, a higher power is better. If alpha is fixed, the only way to increase the power of a test is by increasing the number of samples, $n$. 

Accepting vs Rejecting $H_0$

Since the power of a test may be low and unknown, generally it is said that you “fail to reject $H_0$” as opposed to “accept $H_0$” when $t_n$ does not lie in the critical region.

When $H_0$ is not rejected, we generally do not know with any certainty whether $H_0$ is true or whether $H_0$ is false, since our test might not be powerful enough to detect any difference between $H_0$ and what is actually true.
Accepting vs Rejecting $H_0$

It should be mentioned that there is an intimate relationship between between the confidence interval for the $t$ test:

$$\bar{X}(n) \pm t_{n-1,1-\alpha/2} \sqrt{S^2(n)/n}$$

and the hypothesis test:

$$\text{if } |t_n| \begin{cases} > t_{n-1,1-\alpha/2} & \text{reject } H_0 \\ \leq t_{n-1,1-\alpha/2} & \text{”accept” } H_0 \end{cases}$$

In particular, rejection of the null hypothesis $H_0$ that $\mu = \mu_0$ is equivalent to $\mu_0$ not being contained in the confidence interval for $\mu$, assuming the same value of alpha for both the hypothesis test and the confidence interval.
The Strong Law of Large Numbers
The second most important result in probability theory (after the central limit theorem) is arguably the strong law of large numbers.

Let $X_1, X_2, \ldots, X_n$ be IID random variables with finite mean $\mu$. Then the strong law of large numbers is:

$$\bar{X}(n) \rightarrow \mu \text{ w.p. 1 as } n \rightarrow \infty$$

This theorem basically states that if one performs an infinite number of experiments each resulting in an $X$-bar($n$) and $n$ is sufficiently large, then $X$-bar($n$) will be arbitrarily close to $\mu$ for almost all the experiments.
The Strong Law of Large Numbers

\[ \bar{X}(n) \text{ for various values of } n \text{ when the } X_i \text{'s are normal random variables with } \mu = 1 \text{ and } \sigma^2 = 0.01. \]
The Danger of Replacing a Probability Distribution by its Mean
Sometimes simulation analysts have sometimes replaced an input probability distribution by its mean in their simulation models. This may be because a lack of understanding on the part of the analyst or by lack of information on the form of the actual probability distribution.
Example

Considering a manufacturing system consisting of a single machine tool. Suppose that the “raw” parts arrive to the machine with exponential inter arrival times having a mean of 1 minute and that processing times at the machine are exponentially distributed with mean of 0.99 minute.

Thus, this system is a M/M/1 queue with utilization factor $\rho = 0.99$. Furthermore, it can be shown that the average delay in queue of a part in the long run is 98.01 minutes.

On the other hand, if we replace each distribution by it’s corresponding mean (i.e., customers arrive at minute 1, 2, 3, and so on, and parts have processing times of exactly 0.99 minute), then no part is ever delayed in the queue. In general, the variances as well as the means of the input distributions affect the output measures for queuing-type systems.
Comments on Covariance-Stationary Processes
Consider the process process \( \{D_i, i \geq 1\} \) for the M/M/1 queue when no customers are present at time 0. Clearly, \( D_1 = 0 \), but \( P(D_i > 0) > 0 \) for \( i = 2, 3, \ldots \). Therefore, \( E(D_1) = 0 \) and \( E(D_i) > 0 \) for \( i = 2, 3, \ldots \), which implies that \( \{D_i, i \geq 1\} \) is not covariance stationary. However, if \( \rho < 1 \), it can be shown that for all \( x \geq 0 \) that:

\[
P(D_i \leq x) \to (1 - \rho) + \rho(1 - e^{-(\omega - \lambda)x}) \quad \text{as} \quad i \to \infty
\]

It follows from this and the equation for \( D_{i+1} \) that if we delete the first \( k \) observations from \( D_1, D_2, \ldots \) and \( k \) is sufficiently large then the process \( D_{k+1}, D_{k+2}, \ldots \) will be (approximately) covariance-stationary. Therefore, when we say “consider the process \( \{D_i, i \geq 1\} \) for the covariance-stationary M/M/1 queue,” we mean that we let the M/M/1 queue “warm up” for some amount of time before observing the first delay.
Covariance Stationary Processes

Consider the process \( \{C_i, i \geq 1\} \) for the inventory system, when \( I_1 = S \). Since \( P(I_i = S) \) is not equal to 1 for \( i = 2, 3, \ldots \) it follows that \( \{C_i, i \geq 1\} \) is no covariance-stationary. However, it can be shown that \( p(C_i \leq x) \) converges to a limiting distribution function as \( i \to \infty \). Thus, \( C_{k+1}, C_{k+2}, \ldots \) will be (approximately) covariance-stationary for \( k \) large. Furthermore, the (previously shown) following correlations are for an inventory system warmed up for some amount of time before the first cost is observed.
Conclusions