

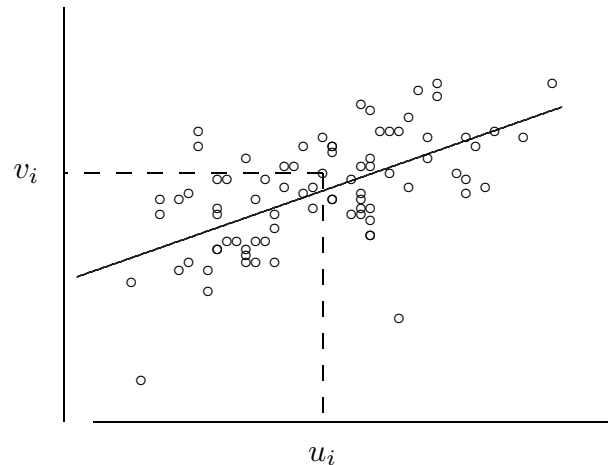
In a discrete-event simulation model it is often the case that two random variables are “co-related”. That is, the natural variation in these two variables is somehow coupled. The statistical term for this is *correlation*. For example, in a FIFO single-server service node there is (positive) correlation between a job’s wait in the service node and the wait time of the preceding job. This section examines two types of correlation: *paired* and *serial*.

4.4.1 PAIRED CORRELATION

Given a *paired sample* (u_i, v_i) for $i = 1, 2, \dots, n$, we begin by considering how to compute a statistic that measures the extent to which this sample exhibits correlation.

Definition 4.4.1 A display of the paired data (u_i, v_i) for $i = 1, 2, \dots, n$ as illustrated in Figure 4.1.1 is called a *bivariate scatterplot*.

Figure 4.4.1.
Bivariate
scatterplot.



When a bivariate scatterplot is created, it is sometimes the case that the (u_i, v_i) points lie primarily clustered around a line, as is the case in Figure 4.4.1. If this is the case then it is natural to ask what line “best fits” the scatterplot data. It is in this sense that what we are discussing is called *linear* correlation. Consider the line in the (u, v) plane defined by the equation $au + bv + c = 0$ and for each point (u_i, v_i) , let d_i be the *orthogonal* distance from this point to the line, as illustrated below.

$$d_i = \frac{|au_i + bv_i + c|}{\sqrt{a^2 + b^2}}$$

From calculus recall that the equation for d_i is as indicated. Thus we choose the (a, b, c) line parameters that *minimize* the mean-square orthogonal distance

$$D = \frac{1}{n} \sum_{i=1}^n d_i^2 = \frac{1}{n(a^2 + b^2)} \sum_{i=1}^n (au_i + bv_i + c)^2.$$

Three-Parameter Minimization

To determine the line that best fits the data, or equivalently to determine the choice of the (a, b, c) line parameters that will minimize D , we begin with the observation that D can be written as*

$$D = \frac{1}{n(a^2 + b^2)} \sum_{i=1}^n \left(a(u_i - \bar{u}) + b(v_i - \bar{v}) \right)^2 + \frac{1}{n(a^2 + b^2)} \sum_{i=1}^n (a\bar{u} + b\bar{v} + c)^2$$

where

$$\bar{u} = \frac{1}{n} \sum_{i=1}^n u_i \quad \text{and} \quad \bar{v} = \frac{1}{n} \sum_{i=1}^n v_i$$

are the sample means of the u and v data respectively. Both terms in the equation for D are non-negative since they involve sums of squares. The first term is independent of c ; the second term is not. Therefore, to minimize D the line parameter c should be chosen so as to minimize the second term. By inspection, the minimum value of the second term is zero and this is achieved by choosing c so that

$$a\bar{u} + b\bar{v} + c = 0,$$

which eliminates the second term. For any choice of the (a, b) line parameters, if $a\bar{u} + b\bar{v} + c = 0$ it follows that the line must pass through the point $(u, v) = (\bar{u}, \bar{v})$. In particular, this geometric property is true for the choice of the (a, b) line parameters that will minimize D . The three-parameter minimization problem has been reduced to a two-parameter minimization problem.

Two-Parameter Minimization

Because $c = -a\bar{u} - b\bar{v}$ the equation $au + bv + c = 0$ can be written equivalently as $a(u - \bar{u}) + b(v - \bar{v}) = 0$. Moreover, since the (a, b) line parameters cannot both be zero, we can simplify the equation for D by assuming that (a, b) are normalized so that $a^2 + b^2 = 1$. The following theorem summarizes this discussion.

Theorem 4.4.1 The line that best fits the data (u_i, v_i) for $i = 1, 2, \dots, n$ in a mean-squared orthogonal distance sense is given by the equation

$$a(u - \bar{u}) + b(v - \bar{v}) = 0$$

where the (a, b) line parameters are chosen to minimize

$$D = \frac{1}{n} \sum_{i=1}^n \left(a(u_i - \bar{u}) + b(v_i - \bar{v}) \right)^2$$

subject to the constraint $a^2 + b^2 = 1$.

* The details of this derivation are left as an exercise.

Covariance and Correlation

We now re-write the algebraic expression for D in Theorem 4.4.1. We do this by introducing two important (related) statistical measures of how the u 's and v 's are “co-related”.

Definition 4.4.2 Given the bivariate sample (u_i, v_i) for $i = 1, 2, \dots, n$ the (linear) *sample covariance* is

$$c_{uv} = \frac{1}{n} \sum_{i=1}^n (u_i - \bar{u})(v_i - \bar{v})$$

and, provided both s_u and s_v are not zero, the (linear) *sample correlation coefficient* is

$$r = \frac{c_{uv}}{s_u s_v}$$

where \bar{u} , \bar{v} ,

$$s_u^2 = \frac{1}{n} \sum_{i=1}^n (u_i - \bar{u})^2, \quad \text{and} \quad s_v^2 = \frac{1}{n} \sum_{i=1}^n (v_i - \bar{v})^2$$

are the sample means and sample variances of the u and v data respectively.*

As we will see, the correlation coefficient r measures the “spread” (dispersion) of the u, v data about the line that best fits the data. By using Definition 4.4.2 the expression for D can be written in terms of s_u^2 , s_v^2 , and r as

$$\begin{aligned} D &= \frac{1}{n} \sum_{i=1}^n \left(a(u_i - \bar{u}) + b(v_i - \bar{v}) \right)^2 \\ &= \frac{1}{n} \sum_{i=1}^n \left(a^2(u_i - \bar{u})^2 + 2ab(u_i - \bar{u})(v_i - \bar{v}) + b^2(v_i - \bar{v})^2 \right) \\ &= \frac{a^2}{n} \left(\sum_{i=1}^n (u_i - \bar{u})^2 \right) + \frac{2ab}{n} \left(\sum_{i=1}^n (u_i - \bar{u})(v_i - \bar{v}) \right) + \frac{b^2}{n} \left(\sum_{i=1}^n (v_i - \bar{v})^2 \right) \\ &= a^2 s_u^2 + 2ab r s_u s_v + b^2 s_v^2. \end{aligned}$$

Note that $r = 1$ if and only if $D = (a s_u + b s_v)^2$ and that $r = -1$ if and only if $D = (a s_u - b s_v)^2$. Therefore, if $|r| = 1$ then it is possible to choose (a, b) in such a way that $D = 0$. Indeed, as we will see, the following three conditions are equivalent

$$|r| = 1 \quad \iff \quad D = 0 \quad \iff \quad \text{all the points } (u_i, v_i) \text{ lie on a line.}$$

* The covariance is a generalization of the variance in the sense that $c_{uu} = s_u^2$ and $c_{vv} = s_v^2$. Note that the covariance derives its “dimensions” from u and v . The correlation coefficient is dimensionless.

The covariance equation in Definition 4.4.2 is a two-pass expression. As in Section 4.1 it can be shown that an equivalent one-pass expression for the covariance is

$$c_{uv} = \frac{1}{n} \left(\sum_{i=1}^n u_i v_i \right) - \bar{u}\bar{v}.$$

The derivation is left as an exercise. We will consider the computational significance of this result later in the section.

One-Parameter Minimization

Let θ be the angle of the line $a(u - \bar{u}) + b(v - \bar{v}) = 0$ measured counterclockwise relative to the u -axis, as illustrated in Figure 4.4.2.

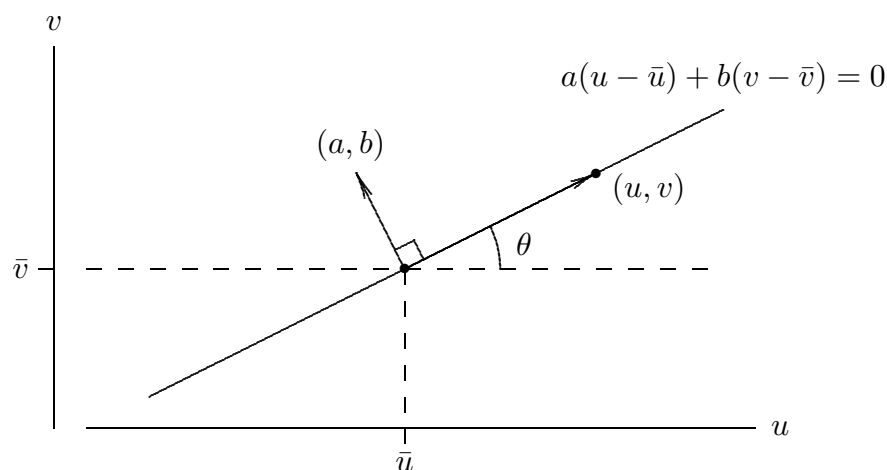


Figure 4.4.2.
Defining the
angle θ .

Because $a(u - \bar{u}) + b(v - \bar{v}) = 0$ and $a^2 + b^2 = 1$, the (a, b) line parameters form a unit vector that is *orthogonal* to the line. Therefore the relation between (a, b) and θ using elementary trigonometry is

$$a = -\sin \theta \quad \text{and} \quad b = \cos \theta.$$

Consistent with Theorem 4.4.1, minimizing D is accomplished by rotating the line about the point (\bar{u}, \bar{v}) to find the angle θ for which D is smallest. To find this angle, we can write D in terms of θ as

$$\begin{aligned} D &= a^2 s_u^2 + 2abc_{uv} + b^2 s_v^2 \\ &= s_u^2 \sin^2 \theta - 2c_{uv} \sin \theta \cos \theta + s_v^2 \cos^2 \theta \\ &= s_u^2 (1 - \cos^2 \theta) - c_{uv} \sin 2\theta + s_v^2 \cos^2 \theta \\ &\vdots \\ &= \frac{1}{2}(s_u^2 + s_v^2) - c_{uv} \sin 2\theta - \frac{1}{2}(s_u^2 - s_v^2) \cos 2\theta. \end{aligned}$$

The value of θ that minimizes

$$D = \frac{1}{2}(s_u^2 + s_v^2) - c_{uv} \sin 2\theta - \frac{1}{2}(s_u^2 - s_v^2) \cos 2\theta$$

solves the equation

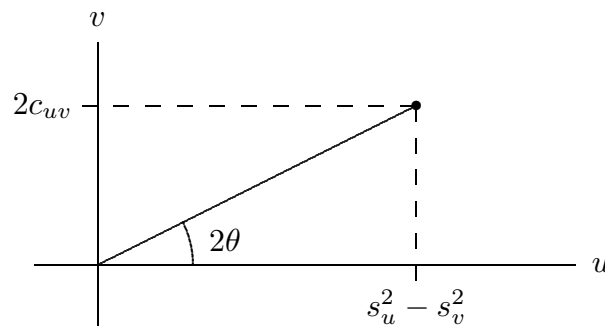
$$\frac{dD}{d\theta} = -2c_{uv} \cos 2\theta + (s_u^2 - s_v^2) \sin 2\theta = 0.$$

This equation can be solved for $\tan 2\theta$ to yield

$$\tan 2\theta = \frac{2c_{uv}}{s_u^2 - s_v^2} \quad (s_u \neq s_v)$$

as illustrated in Figure 4.4.3.

Figure 4.4.3.
Finding the
value of θ
that minimizes D .



Therefore, the angle which minimizes D is

$$\theta = \frac{1}{2} \tan^{-1}(s_u^2 - s_v^2, 2c_{uv}),$$

where $\tan^{-1}(u, v)$ is the usual 4-quadrant inverse tangent function, measured counterclockwise from the positive u -axis.* This discussion is summarized by the following theorem.

Theorem 4.4.2 The line that best fits the data (u_i, v_i) for $i = 1, 2, \dots, n$ in a mean-squared orthogonal distance sense passes through the point (\bar{u}, \bar{v}) at the angle

$$\theta = \frac{1}{2} \tan^{-1}(s_u^2 - s_v^2, 2c_{uv})$$

measured counterclockwise relative to the positive u -axis. The equation of the line is

$$v = (u - \bar{u}) \tan(\theta) + \bar{v}$$

provided $\theta \neq \pi/2$. (By convention $-\pi < \tan^{-1}(u, v) \leq \pi$ so that $-\pi/2 < \theta \leq \pi/2$.)

* The function `atan2(v, u)` in the ANSI C library `<math.h>` represents the mathematical function $\tan^{-1}(u, v)$. Note the (u, v) switch.

Definition 4.4.3 The line that best fits the data (u_i, v_i) for $i = 1, 2, \dots, n$ is known as the (mean-square orthogonal distance) *linear regression* line.*

As a corollary to Theorem 4.4.2, it can be shown that the smallest possible value of D is

$$2D_{\min} = (s_u^2 + s_v^2) - \sqrt{(s_u^2 - s_v^2)^2 + 4r^2 s_u^2 s_v^2},$$

which can be written equivalently as

$$D_{\min} = \frac{2(1 - r^2)s_u^2 s_v^2}{s_u^2 + s_v^2 + \sqrt{(s_u^2 - s_v^2)^2 + 4r^2 s_u^2 s_v^2}}.$$

The details of this derivation are left as an exercise. There are three important observations that follow immediately from this equation and the fact that D_{\min} cannot be negative.

- The correlation coefficient satisfies the inequality $-1 \leq r \leq 1$.
- The closer $|r|$ is to 1, the smaller the dispersion of the (u, v) data about the regression line, and the better the (linear) fit.
- All the (u_i, v_i) points lie on the regression line if and only if $D_{\min} = 0$ or equivalently if and only if $|r| = 1$.

Example 4.4.1 The scatterplot in Definition 4.4.1, reproduced in Figure 4.4.4, corresponds to 82 student scores on two standardized tests of English verbal skills: the Test of English as a Foreign Language (TOEFL) and Graduate Record Examination (GRE). Although one might hope that the two test scores would be highly correlated with r close to 1, in this case $r = 0.59$. That is, consistent with the considerable scatter that is evident about the linear regression line, the correlation is not particularly high. The consistency between the two tests is certainly less than desirable.

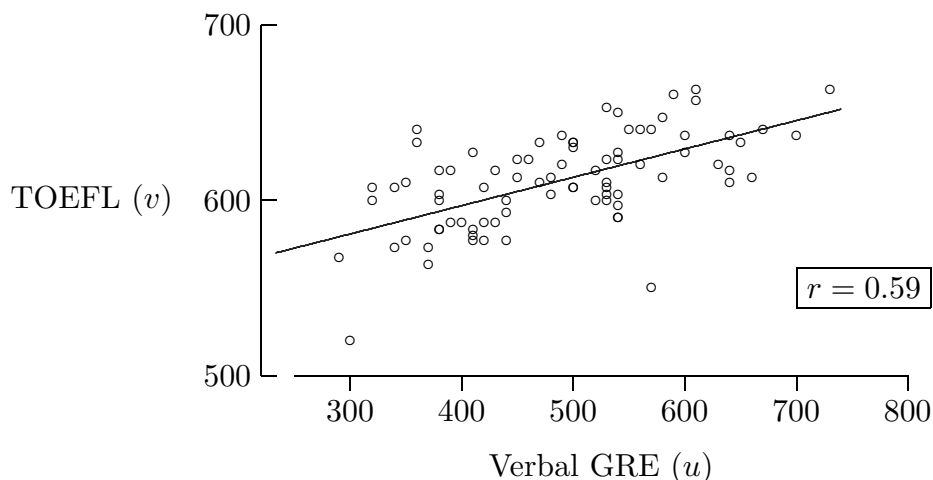


Figure 4.4.4.
English verbal
measures
scatterplot.

* The derivation of an alternate mean-square *non-orthogonal* distance linear regression line is outlined in the exercises.

Significance

The magnitude of $|r|$ is a measure of the extent to which there is a *linear* relation between the u and v data. Associated terminology is provided by the following definition.

Definition 4.4.4 If $r \neq 0$ then the slope of the regression line is positive ($\theta > 0$) if and only if $r > 0$ and the slope of the regression line is negative ($\theta < 0$) if and only if $r < 0$. If r is close to $+1$ the data is said to be *positively correlated*. If r is close to -1 the data is said to be *negatively correlated*. If r is close to 0 then the data is said to be *uncorrelated*.

Example 4.4.2 A modified version of program `ssq2` was used to generate interarrival, service, delay, and wait times for a steady-state sample of 100 jobs passing through an $M/M/1$ service node with arrival rate 1.0 and service rate 1.25. An $M/M/1$ service node has *Exponential* interarrival time, *Exponential* service times, and a single server. By pairing these times, a total of six bivariate scatterplots could be formed, four of which are illustrated in Figure 4.4.5.

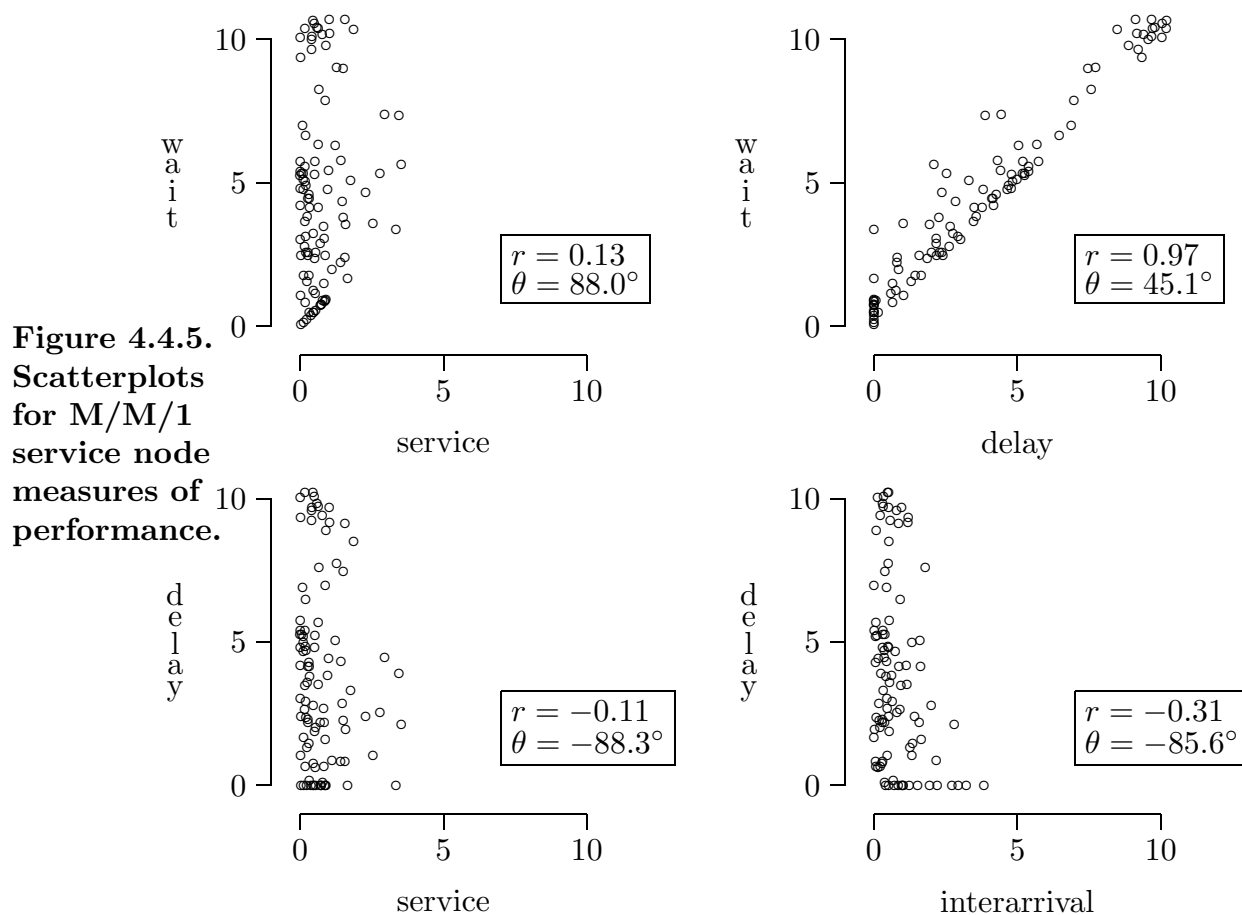


Figure 4.4.5. Scatterplots for $M/M/1$ service node measures of performance.

In this case we see, as expected, that the strongest positive correlation is between a job's delay and wait. For the other three pairings illustrated, the correlation is weak, if it is non-zero at all.

The statistical question of how far $|r|$ has to be from 0 to conclude that the bivariate data is actually correlated is a difficult one made more complicated by the fact that the decision depends on the sample size and the joint probability distribution of the two random variables. The smaller the sample size, the larger $|r|$ needs to be before one can safely conclude that the data is correlated.

If the sample size is small then the value of r is uncertain in the sense that another sample of the same size could produce a significantly different value of r . For example, for an $M/M/1$ service node, a job's delay and service time are *uncorrelated*. Why? Thus the weak $r = -0.11$ negative correlation indicated in Figure 4.4.5 is not statistically different from zero — another sample of size 100 will produce a *positive* correlation with probability $1/2$.

Computational Considerations

To return to the covariance equation in Definition 4.4.2, recall from Section 4.1 that there are two ways to calculate a variance (or standard deviation). One method involves *two* passes through the data, the first to evaluate the mean and then the second to sum the squares of the deviations about the mean. The other method involves just *one* pass through the data. An analogous result applies to the calculation of the covariance and, therefore, the correlation coefficient. Indeed, we have already observed that there are two mathematically equivalent expressions for the covariance

$$\underbrace{c_{uv} = \frac{1}{n} \sum_{i=1}^n (u_i - \bar{u})(v_i - \bar{v})}_{\text{two-pass}} \quad \text{and} \quad \underbrace{c_{uv} = \frac{1}{n} \sum_{i=1}^n u_i v_i - \bar{u}\bar{v}}_{\text{one-pass}}.$$

For the same reasons discussed in Section 4.1, the one-pass algorithm is virtually always preferred in discrete-event simulation. Moreover, there is an extension to Welford's algorithm (Algorithm 4.1.2) that applies in this case, based on the following theorem.

Theorem 4.4.3 Let \bar{u}_i and \bar{v}_i denote the sample means of u_1, u_2, \dots, u_i and v_1, v_2, \dots, v_i respectively and define

$$w_i = (u_1 - \bar{u}_i)(v_1 - \bar{v}_i) + (u_2 - \bar{u}_i)(v_2 - \bar{v}_i) + \dots + (u_i - \bar{u}_i)(v_i - \bar{v}_i)$$

for $i = 1, 2, \dots, n$ where w_i/i is the covariance of the first i data pairs. Then, with the initial condition $w_0 = 0$,

$$w_i = w_{i-1} + \left(\frac{i-1}{i} \right) (u_i - \bar{u}_{i-1})(v_i - \bar{v}_{i-1}) \quad i = 1, 2, \dots, n$$

which provides a one-pass recursive algorithm to compute $c_{uv} = w_n/n$.

Program bvs

The program **bvs** is based upon the extended version of Welford's algorithm in Theorem 4.4.3. This program illustrates the calculation of the *bivariate* sample statistics \bar{u} , s_u , \bar{v} , s_v , r , and the linear regression line angle θ .

4.4.2 SERIAL CORRELATION

It is frequently the case that one is interested in the extent to which a set of data is *auto-correlated* (e.g., self-correlated). This is particularly true, for example, in a steady-state analysis of the waits experienced by consecutive jobs entering a service node. Intuitively, particularly if the utilization of the service node is high, there will be a high positive correlation between the wait w_i experienced by the i^{th} job and the wait w_{i+1} experienced by the next job. Indeed, there will be a statistically significant positive correlation between w_i and w_{i+j} for some range of small, positive j values.

In general, let x_1, x_2, \dots, x_n be data which is presumed to represent n consecutive observations of some stochastic process whose serial correlation we wish to characterize. In the (u_i, v_i) notation used previously in this section, we pick a (small) fixed positive integer $j \ll n$ and then associate u_i with x_i and v_i with x_{i+j} as illustrated

$$\begin{array}{cccccccccccc} u & : & & & x_1 & x_2 & x_3 & \cdots & x_i & \cdots & x_{n-j} & x_{n-j+1} & \cdots & x_n \\ v & : & x_1 & \cdots & x_j & x_{1+j} & x_{2+j} & x_{3+j} & \cdots & x_{i+j} & \cdots & x_n & & \end{array}$$

The integer $j > 0$ is called the *autocorrelation lag* (or *shift*). Although the value $j = 1$ is generally of primary interest, it is conventional to calculate the serial correlation for a range of lag values $j = 1, 2, \dots, k$ where $k \ll n$.*

Because of the lag we must resolve how to handle the “non-overlap” in the data at the beginning and end. The standard way to handle this non-overlap is to do the obvious — ignore the extreme data values. That is, define the sample autocovariance for lag j , based only on the $n - j$ overlapping values, as

$$c_j = \frac{1}{n-j} \sum_{i=1}^{n-j} (x_i - \bar{x})(x_{i+j} - \bar{x}) \quad j = 1, 2, \dots, k,$$

where the sample mean, based on all n values, is

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$$

The associated autocorrelation is then defined as follows.

Definition 4.4.5 The *sample autocorrelation* for lag j is

$$r_j = \frac{c_j}{c_0} \quad j = 1, 2, \dots, k$$

where the sample variance is

$$c_0 = s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2.$$

* Serial statistics are commonly known as *auto-statistics*, e.g., *autocovariance* and *autocorrelation*.

Computational Considerations

The problem with the “obvious” definition of the sample autocovariance is that an implementation based on this definition would involve a two-pass algorithm. For that reason, it is common to use the following alternate definition of the sample autocovariance. Although this definition is *not* algebraically equivalent to the “obvious” definition, if $j \ll n$ then the numerical difference between these two autocovariance definitions is slight. Because it can be implemented as a one-pass algorithm, Definition 4.4.6 is preferred (in conjunction with Definition 4.4.5).

Definition 4.4.6 The *sample autocovariance* for lag j is

$$c_j = \left(\frac{1}{n-j} \sum_{i=1}^{n-j} x_i x_{i+j} \right) - \bar{x}^2 \quad j = 1, 2, \dots, k.$$

Example 4.4.3 A modified version of program `ssq2` was used to generate a sample of waits and services experienced by 10 000 consecutive jobs processed through an $M/M/1$ service node, in steady-state, with arrival rate 1.0, service rate 1.25, and utilization $1/1.25 = 0.8$. Definitions 4.4.5 and 4.4.6 were used to compute the corresponding sample autocorrelations r_j for $j = 1, 2, \dots, 50$, in what is commonly known as a *sample autocorrelation function*, or *correlogram*, illustrated in Figure 4.4.6.

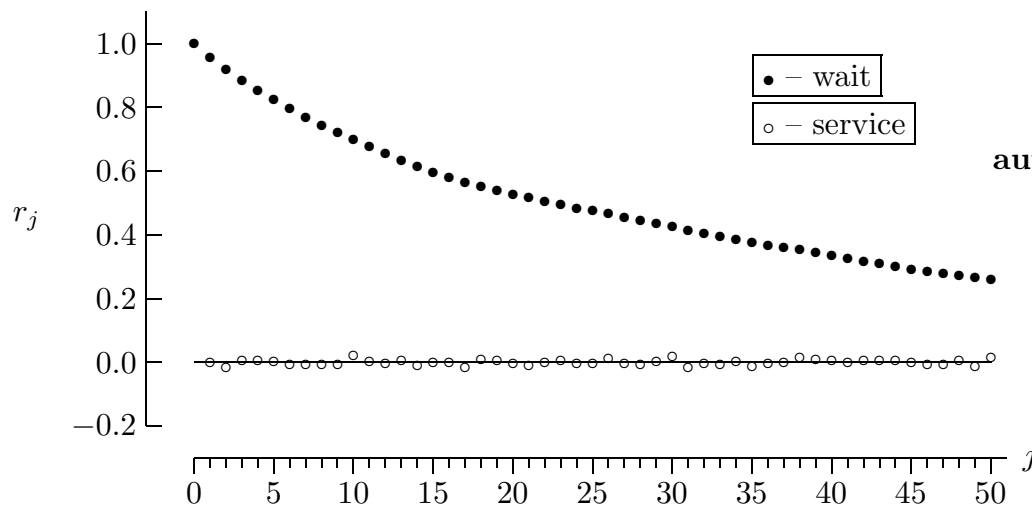


Figure 4.4.6.
Sample autocorrelation functions for the $M/M/1$ queue.

As expected, the sample *wait* autocorrelation is positive and high for small values of j , indicating that each job’s wait is strongly (auto)correlated with the wait of the next few jobs that follow. Also, as expected, the sample autocorrelation decreases monotonically toward zero as j increases. The rate of decrease may be slower than expected; if the utilization were smaller (larger), the rate of decrease would be higher (lower). It is quite surprising that the wait times of two jobs separated by 49 intervening jobs have a moderately strong positive correlation. Also, as expected, the sample *service* autocorrelation is essentially zero for all values of j , consistent with the stochastic independence of the service times.

We select the first three values of the the sample autocorrelation function for the wait times in order to interpret the magnitude of the r_j values. First, $r_0 = 1$ means that there is perfect correlation between each observation and itself, since the lag associated with r_0 is zero. The next sample autocorrelation value, $r_1 = 0.957$, indicates that adjacent (lag 1) jobs, such as job number 15 and job number 16, have a statistically significant strong positive correlation. If the 15th job has a long wait, then the 16th job is almost certain to also have a long wait. Likewise, if the 15th job has a short wait, then the 16th job is almost certain to also have a short wait. Anyone who has waited in a busy queue recognizes this notion intuitively. Finally, consider the estimated lag-two sample autocorrelation $r_2 = 0.918$. This autocorrelation is not quite as strong as the lag-one autocorrelation due to the increased temporal distance between the wait times. The positive value of r_2 indicates that wait times two jobs apart (e.g., the 29th and the 31st wait times) *tend* to be above the mean wait time together or below the mean wait time together.

Graphical Considerations

Several formats are common for displaying the sample autocorrelation function. In Figure 4.4.6, we plot the r_j values as points. Another common practice is to draw “spikes” from the horizontal axis to the r_0, r_1, \dots, r_k values. It is certainly *not* appropriate to connect the points to produce a piecewise-linear function. This would imply that r_j is defined for non-integer values of j — which it is not.

Statistical Considerations

The previous example indicated that jobs separated by 50 lags have wait times that are positively correlated. But how do we know that r_{50} differs significantly from 0. Leaving out the details, Chatfield (2004, page 56) indicates that an r_j value will fall outside the limits $\pm 2/\sqrt{n}$ with approximate probability 0.95 when the lag j values are uncorrelated. In the previous example with $n = 10\,000$, for instance, these limits are at ± 0.02 , indicating the *all* of the wait time sample autocorrelation values plotted differ significantly from 0. For the service times, only $r_{10} = 0.022$ falls outside of these limits. Experience dictates that this is simply a function of random sampling variability rather than some relationship between service times separated by 10 jobs. We have set up our service time model with independent service times, so we expect a flat sample autocorrelation function for service times. The spurious value can be ignored.

The high autocorrelation that typically exists in the time-sequenced stochastic data produced by a simulation makes the statistical analysis of the data a challenge. Specifically, if we wish to make an *interval* estimate of some steady-state statistic like, for example, the average wait in a service node, we must be prepared to deal with the impact of autocorrelation on our ability to make an accurate estimates of the standard deviation. Most of so-called “classical” statistics relies on the assumption that the values sampled are drawn independently from a population. This is often not the case in discrete-event simulation and appropriate measures must be taken in order to compute appropriate interval estimates.

Program acs

To implement Definition 4.4.6 as a one-pass algorithm for a fixed lag-value j involves nothing more than storing the values x_i, x_{i+j} , accumulating the x_i sum, and accumulating the x_i^2 and $x_i x_{i+j}$ “cosums.” It is a greater challenge to construct a one-pass algorithm that will compute c_j for a *range* of lags $j = 1, 2, \dots, k$. In addition to the accumulation of the x_i sum, the simultaneous computation of c_0, c_1, \dots, c_k involves storing the $k + 1$ consecutive values $x_i, x_{i+1}, \dots, x_{i+k}$ and accumulating the $k + 1$ (lagged) $x_i x_{i+j}$ cosums for $j = 0, 1, 2, \dots, k$. The $k + 1$ cosums can be stored as an array of length $k + 1$. A more interesting queue data structure is required to store the values $x_i, x_{i+1}, \dots, x_{i+k}$. This queue has been implemented as a *circular* array in the program **acs**. A circular array is a natural choice here because the queue length is fixed at $k + 1$ and efficient access to *all* the elements in the queue, not just the head and tail, is required. In the following algorithm the box indicates the rotating head of the circular queue. An array index p keeps track of the current location of the rotating head; the initial value is $p = 0$.

Algorithm 4.4.1 Program **acs** is based on the following algorithm. A circular queue is initially filled with $x_1, x_2, \dots, x_k, x_{k+1}$, as illustrated by the boxed elements below. The lagged products $x_1 x_{1+j}$ are computed for all $j = 0, 1, \dots, k$ thereby initializing the $k + 1$ cosums. Then the next data value is read into the (old) head of the queue location, p is incremented by 1 to define a new head of the queue location, the lagged products $x_2 x_{2+j}$ are computed for all $j = 0, 1, \dots, k$, and the cosums are updated. This process is continued until all the data has been read and processed. (The case $n \bmod (k + 1) = 2$ is illustrated.)

$(i = k + 1)$	x_1	x_2	x_3	\cdots	x_{k-1}	x_k	x_{k+1}	$(p = 0)$
$(i = k + 2)$	x_{k+2}	x_2	x_3	\cdots	x_{k-1}	x_k	x_{k+1}	$(p = 1)$
$(i = k + 3)$	x_{k+2}	x_{k+3}	x_3	\cdots	x_{k-1}	x_k	x_{k+1}	$(p = 2)$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
$(i = 2k)$	x_{k+2}	x_{k+3}	x_{k+4}	\cdots	x_{2k}	x_k	x_{k+1}	$(p = k)$
$(i = 2k + 1)$	x_{k+2}	x_{k+3}	x_{k+4}	\cdots	x_{2k}	x_{2k+1}	x_{k+1}	$(p = k + 1)$
$(i = 2k + 2)$	x_{k+2}	x_{k+3}	x_{k+4}	\cdots	x_{2k}	x_{2k+1}	x_{2k+2}	$(p = 0)$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
$(i = n)$	x_{n-1}	x_n	x_{n-k}	\cdots	x_{n-4}	x_{n-3}	x_{n-2}	$(p = 2)$

After the last data value, x_n , has been read, the associated lagged products computed, and the cosums updated, all that remains is to “empty” the queue. This can be accomplished by effectively reading k additional 0-valued data values. For more details, see program **acs**.

4.4.3 EXERCISES

Exercise 4.4.1 Prove that the orthogonal distance from the point (u_i, v_i) to the line $au + bv + c = 0$ is in fact

$$d_i = \frac{|au_i + bv_i + c|}{\sqrt{a^2 + b^2}}.$$

Hint: consider the squared distance $(u - u_i)^2 + (v - v_i)^2$ from (u_i, v_i) to a point (u, v) on the line and show that d_i^2 is the smallest possible value of this distance.

Exercise 4.4.2 (a) If $u'_i = \alpha_u u_i + \beta_u$ and $v'_i = \alpha_v v_i + \beta_v$ for $i = 1, 2, \dots, n$ and constants $\alpha_u, \alpha_v, \beta_u,$ and β_v how does the covariance of the u', v' data relate to the covariance of the u, v data? (b) Same question for the correlation coefficients? (c) Comment.

Exercise 4.4.3^a The orthogonal distance regression derivation presented in this section treats both variables equally — there is no presumption that one variable is “independent” and the other is “dependent.” Consider the more common regression approach in which the equation of the regression line is $v = au + b$, consistent with a model that treats u as independent and v as dependent. That is, given the data (u_i, v_i) for $i = 1, 2, \dots, n$ and the line defined by the equation $v = au + b$, the *conventional* (non-orthogonal) distance from the point (u_i, v_i) to the line is

$$\delta_i = |v_i - (au_i + b)|.$$

(a) What choice of the (a, b) parameters will minimize the conventional mean-square distance

$$\Delta = \frac{1}{n} \sum_{i=1}^n \delta_i^2 = \frac{1}{n} \sum_{i=1}^n (v_i - au_i - b)^2.$$

(b) Prove that the minimum value of Δ is $(1 - r^2)s_v^2$.

Exercise 4.4.4 Prove Theorem 4.4.3.

Exercise 4.4.5 To what extent are these two definitions of the autocovariance different?

$$c'_j = \frac{1}{n-j} \sum_{i=1}^{n-j} (x_i - \bar{x})(x_{i+j} - \bar{x}) \quad \text{and} \quad c_j = \left(\frac{1}{n-j} \sum_{i=1}^{n-j} x_i x_{i+j} \right) - \bar{x}^2$$

Exercise 4.4.6 (a) Generate a figure like the one in Figure 4.4.6 but corresponding to a utilization of 0.9. Do this for three different **rngs** streams. Comment. (b) Repeat for a utilization of 0.7. (You can ignore the service autocorrelations.) Comment.

Exercise 4.4.7 If Definition 4.4.6 is used in conjunction with Definition 4.4.5, there is no *guarantee* that $|r_j| \leq 1$ for all $j = 1, 2, \dots, k$. If it is important to guarantee that this inequality is true, then how should the two definitions be modified?