Density Estimation for Analysis of Simulation Output and Input using Independent and Correlated Processes

E. JACK CHEN[1] and W. DAVID KELTON[2]
[1] *BASF Corporation, 333 Mount Hope Avenue, Rockaway, NJ 07866-0909, USA*
E-mail: e.jack.chen@basf.com
[2] *Department of Quantitative Analysis and Operations Management, University of Cincinnati, Cincinnati, OH 45221-0130, USA*
E-mail: david.kelton@uc.edu

This paper evaluates sequential procedures for estimating the steady-state density of a stochastic process, typically (though not necessarily) observed by simulation, with or without intra-process independence. The procedure computes sample densities at certain points and uses Lagrange interpolation to estimate the density $f(x)$. Even though the proposed sequential procedure is a heuristic, it does have a strong basis. Our empirical results show that the procedure gives density estimates that satisfy a pre-specified precision requirement. Though our primary focus is on simulation output analysis, we also show how our methods can be used to develop robust non-parametric input distributions for simulation model building. An experimental performance evaluation demonstrates the validity of using the procedure to estimate densities of steady-state stochastic processes.

Keywords: Simulation, Output Analysis, Density Estimation, Histogram Estimation

1. Introduction

Simulation studies have been used to investigate the characteristics of systems, for example the mean and the variance of certain system performance measures like waiting times in queue. The probability density function f gives a natural description of the distribution of a stationary continuous output random variable X produced by a simulation. The probability density function associated with X satisfies

$$P(a < X \leq b) = \int_a^b f(x)dx \text{ for all } a < b.$$

Related to the density function is the cumulative distribution function (cdf) associated with X, which satisfies

$$P(X \leq b) = \int_{-\infty}^b f(x)dx.$$

In this paper, we investigate the performance of a technique proposed by Chen and Kelton (2008) to estimate the density of a simulation output random variable. Direct density estimates are more sensitive to the bin (window) widths of the underlying histogram.

Density estimation from observed data is a useful tool for data exploration. Silverman (1986, p. 5) points out that "density estimates are ideal for presentation of data to provide explanation and illustration of conclusions, since they are fairly easily comprehensible to non-mathematicians." One approach to density estimation is *parametric*, assuming that the data are drawn from a known parametric family of distributions, for example the normal distribution with mean μ and variance σ^2. The density f underlying the data is then estimated simply by estimating the values of μ and σ^2 from the data and substituting these estimates into the formula for the normal density. Another approach is *nonparametric*, where less rigid assumptions are made about the distribution of the observed data. We consider the nonparametric approach since it is more robust for the wide variety of data behavior possible in simulation output. Furthermore, the procedure is data-based, i.e., it can be embodied in a software package whose input is the simulation output data $(X_1, \ldots, X_n)$, and whose output is the density estimate. Several different approaches have received extensive treatment; see Silverman (1986), Scott and Sain (2004), and the references therein.

The most widely used density estimator is the *histogram*, basically a graphical estimate of the underlying probability density function that reveals all the essential distributional features of a simulation output random variable, such as skewness and multimodality. Hence, a histogram is often used in the informal investigation of the properties of a given set of data. Given an origin g_0 and a bin width w, the bins of the histogram are the intervals $[g_0 + mw, g_0 + (m+1)w]$ for positive and negative integers m. Suppose that we have any division of the real line into bins; then the histogram density estimator is

$$\hat{f}_h(x) = \frac{1}{n} \times \frac{\text{no. of } X_i \text{ in same bin as } x}{\text{width of bin containing } x} = \frac{\text{proportion of } X_i \text{ in same bin as } x}{\text{width of bin containing } x}.$$

Hence, to construct the histogram, we need to choose both an origin and a bin width. It is the bin width that, primarily, controls the amount of smoothing inherent in the procedure. Note that the optimal smoothing parameter that minimizes the mean integrated square error (MISE) can be computed if the true underlying sampling density f is known; see Section 2.2.

A histogram can be constructed with a properly selected set of quantiles. For both independent and identically distributed (i.i.d.) and *ϕ-mixing* sequences, sample quantiles will be asymptotically unbiased if certain conditions are satisfied; see Sen (1972). Intuitively, a stochastic process is ϕ-mixing if its distant future is essentially independent of its present and past (Billingsley 1999).

In Section 2 we discuss some theoretical bases of density estimation in the context of simulation output analysis. In Section 3 we present our implementations of procedures to estimate densities. In Section 4 we present our empirical-experimental results of density estimation. In Section 5 we give concluding remarks. Earlier versions of some of the results in this paper appeared in Chen and Kelton (2006).

2. Theoretical Basis

In this section, we review the definition of probability density functions and the basis of density estimation. From the definition of a probability density, if the random variable X has density f, then

$$f(x) = \lim_{h \to 0} \frac{1}{2h} P(x - h < X < x + h).$$

Here $h > 0$ is a real number.

2.1 The Histogram Density Estimator

A natural estimator by a histogram $\hat{f}_h(x)$ of the density is given by choosing a small number h $(= w/2)$ and setting

$$\hat{f}_h(x) = \frac{1}{2nh}[\text{no. of } X_1, \ldots, X_n \text{ falling in } (x - h, x + h)] = \frac{1}{nh} \sum_{i=1}^{n} W\left(\frac{x - X_i}{h}\right).$$

Here the weight function $W(\cdot) = I(\cdot)/2$ and $I(\cdot)$ is the indicator function for the interval $(-1, 1)$.

The estimator $\hat{f}_h(x)$ is based on a transformation of the output sequence $\{X_i\}$ to the sequence $\{I_i(h)\}$, $i = 1, 2, \ldots, n$:

$$\hat{f}_h(x) = \frac{1}{2nh} \sum_{i=1}^{n} I_i(h).$$

For data that are i.i.d., the following properties of $I_i(h)$ are well known (Hogg and Craig 1995, pp. 116-117): $E(I_i(h)) = p$ and $\text{Var}(I_i(h)) = p(1 - p)$, where $p = P(-h < x - X < h)$. Since $\hat{f}_h(x)$ is based on the mean of the random variable $I_i(h)$, we can use any method developed for estimating the variance of the mean to estimate $\text{Var}(\hat{f}_h(x))$. By elementary manipulations, for each x,

$$E(\hat{f}_h(x)) = \frac{1}{nh} \sum_{i=1}^{n} E\left(W\left(\frac{x - X_i}{h}\right)\right) = \frac{1}{2h} \int I\left(\frac{x - y}{h}\right) f(y) dy = \frac{p}{2h}$$

and

$$\text{Var}(\hat{f}_h(x)) = \frac{n}{(nh)^2} \text{Var}\left(W\left(\frac{x - X_i}{h}\right)\right) = \frac{1}{4nh^2} \text{Var}\left(I\left(\frac{x - X_i}{h}\right)\right) = \frac{p(1 - p)}{4nh^2}.$$

Note that $\hat{f}_h(x)$ has a binomial distribution. It follows from the definition that $\hat{f}_h$ is not a continuous function, but has jumps at the points $x_i \pm h$ and has zero derivative everywhere else. This gives the estimate a somewhat ragged character.

The histogram density estimator can also be expressed as the average of the Dirac delta function at the n data points, $\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} \delta_{x,h}(X_i)$, where

$$\delta_{x,h}(X) = \begin{cases} \frac{1}{2h} & \text{if } |x - X| < h \\ 0 & \text{otherwise.} \end{cases}$$

It follows that as $h \to 0$, $\delta_{x,h}(X)$ becomes the "idealized" unit impulse function

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

and $E(\hat{f}_h(x)) \to \int \delta_x(y) f(y) dy = f(x)$.

2.2 The Kernel Density Estimator

To overcome the difficulties stemming from the ragged character of $\hat{f}_h(x)$, one can replace the weighting function by a *kernel function* K, which satisfies the condition $\int_{-\infty}^{\infty} K(x)dx = 1$. For simplicity, the kernel K usually is a symmetric function satisfying $\int xK(x)dx = 0$, and $\int x^2 K(x)dx = k_2 \neq 0$; an example is the normal density. The kernel method has been studied extensively in the area of nonparameteric statistics and has been applied to estimate densities. The kernel density estimator with kernel K is defined by

$$\hat{f}_{K,h}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right).$$

The weighting function W of the histogram density estimate satisfies the conditions for a Kernel function. Consequently, the histogram density estimate is just a special case of the kernel density estimate with the kernel function $K(X) = W(X)$. Silverman (1986, pp. 15-17) points out that "$\hat{f}_{K,h}$ will inherit all the continuity and differentiability properties of the kernel K, so that if K is the normal density function, then $\hat{f}_{K,h}$ will be a smooth curve having derivatives of all orders." However, the kernel method often under-estimates the density at the boundary when the domain of the density being estimated is not the whole real line but an interval bounded on one or both sides; see Silverman (1986, p. 29).

It can be shown that

$$E(\hat{f}_{K,h}(x)) = \frac{1}{h}\int K\left(\frac{x - y}{h}\right) f(y) dy;$$

$$\begin{aligned}\mathrm{Var}(\hat{f}_{K,h}(x)) &= \frac{1}{n}\left[\frac{1}{h^2}\int K\left(\frac{x-y}{h}\right)^2 f(y)dy - \left[\frac{1}{h}\int K\left(\frac{x-y}{h}\right) f(y)dy\right]^2\right] \\ &\approx \frac{1}{nh}f(x)\int K(y)^2 dy,\end{aligned}$$

and

$$\begin{aligned}\mathrm{bias}_h(x) &= E(\hat{f}_{K,h}(x) - f(x)) \\ &= \frac{1}{2}h^2 f''(x)k_2 + \text{ higher-order terms in } h.\end{aligned}$$

See Silverman (1986, pp. 37-40) for details. Let $N(\mu, \sigma^2)$ denote the normal distribution with mean μ and variance σ^2. In the special case that the kernel is $N(0, 1)$ and the true density is $N(\mu, \sigma^2)$, then $E(\hat{f}_{K,h})$ is $N(\mu, \sigma^2 + h^2)$; see Fryer (1976) for details.

The approximation of bias and variance indicates one of the fundamental difficulties of density estimation. To eliminate the bias, a small value of h should be used, but then the variance will become large. On the other hand, a large value of h will reduce the variance, but will increase the bias. The mean square error (MSE) is widely used to evaluate the quality of estimates and addresses the trade-off between variance and bias. Note that MSE = Variance + Bias^2. Since the shape of the true density is of most interest, a relevant criterion is the integrated mean squared error (IMSE) (Rosenblatt 1971). Scott and Sain (2004) point out that by Fubini's theorem, IMSE is the same as mean integrated square error (MISE):

$$\mathrm{IMSE} = \int E[\hat{f}(x) - f(x)]^2 dx = E\int [\hat{f}(x) - f(x)]^2 dx = \mathrm{MISE}.$$

IMSE is a function of h, f, and K. The ideal window width h should minimize IMSE and can be computed for a chosen K if f is known. Devroye and Lugosi (2001, p. 83) argue that the density estimate with h that minimizes $\int(\hat{f} - f)^2$ would not be universally useful. They suggest that "squaring tends to squash errors in the tails and make them unimportant." Note that f is what we want to estimate and is unknown. Nevertheless, the ideal window width h should satisfy:

$$\lim_{n\to\infty} h = 0, \lim_{n\to\infty} nh = \infty. \tag{1}$$

That is, h should converge to zero as the sample size increases but more slowly than n. Furthermore, smaller values of h will be appropriate for more rapidly fluctuating densities. Let $J_n = \int |\hat{f}_n - f|$, where $\hat{f}_n(x) = f_n(x; X_1, \ldots, X_n)$ is a real-valued Borel-measurable function of its arguments. Conditions (1) imply that there exists $r \in R$ such that $P(J_n \geq \epsilon) \leq e^{-rn\epsilon^2}$ for all $\epsilon \in (0, 1)$ and all $n \geq n_f$, where n_f depends upon f and ϵ; see Devroye and Györfi (1985).

There has been extensive research on the selection of an optimal kernel function to minimize the IMSE. Scott and Factor (1981) indicate that many symmetric uni-modal kernel functions are nearly optimal, so we use the normal kernel. Moreover, it can be shown that the asymptotically optimal smoothing parameter is

$$h = \alpha(K)\beta(f)n^{-1/5}, \tag{2}$$

where

$$\alpha(K) = \left[\int K(y)^2 dy\right]^{1/5} \left[\int K(y)y^2 dy\right]^{-2/5} \text{ and } \beta(f) = \left[\int f''(x)^2 dx\right]^{-1/5}.$$

With this choice, the IMSE decreases in proportion to $n^{-4/5}$; see Scott and Factor (1981).

It is known that window widths in the tails should be relatively large in order to dampen "wiggles" since there are less data available in the tails. More sophisticated density-estimation algorithms have deployed a variable window-width function that allows different amounts of smoothing depending on the various characteristics of the data and the density being estimated. However, these variable window-width functions often are based on the very function we want to estimate: the density function. Devroye and Lugosi (2001, pp. 163-166) and Sain and Scott (2002) explore how the variable window width can be optimized. Of course, these procedures require more computation.

2.3 The Complication of Lack of Independence

The density estimator (obtained by the methods described above) would be unbiased provided that the observations are independent. However, simulation output data are generally correlated and consequently the estimator maybe biased when the sample size is not sufficiently large. Hence, to estimate the density of stochastic processes, the procedures need to determine the sample sizes dynamically to ensure that the density estimator is unbiased. That is, we assume that the simulation output sequence satisfies the ϕ-mixing conditions. Furthermore, we assume that the underlying process is stationary; i.e., the joint distribution of the X_i's is insensitive to time shifts (in a simulation context, this would mean that the model has been adequately "warmed up").

To determine whether data are independent, we use the von Neumm test. We briefly review the von Neumann test (Fishaman 2001, von Neumann 1941) for the hypothesis H_0: the observations $X_1, X_2, \ldots, X_n$ are uncorrelated. The von Neumann ratio is

$$C_n = 1 - \frac{\sum_{j=2}^{n}(X_j - X_{j-1})^2}{2\sum_{j=1}^{n}(X_j - \hat{\mu}(n))^2},$$

where $\hat{\mu}(n) = \frac{1}{n}\sum_{i=1}^{n} X_i$. Note that C_n is an estimator of the lag-one autocorrelation $w_1 \equiv \mathrm{Corr}(X_j, X_{j+1})$, with an adjustment for end effects that diminishes in importance as the number of observations n increases. If $\{X_i, i \geq 0\}$ has a monotone non-increasing autocorrelation function, then w_1 is positive and decreases monotonically to zero. If for given n, H_0 is true for $X_1, X_2, \ldots, X_n$, then $w_1 = 0$.

The von Neumann test statistic for H_0 is

$$Z = \sqrt{\frac{n^2 - 1}{n - 2}} C_n.$$

Under H_0, $Z \sim N(0, 1)$, so one rejects H_0 at level $1 - \alpha_{\text{ind}}$ if $Z > z_{1-\alpha_{\text{ind}}}$, where $z_{1-\alpha_{\text{ind}}}$ is the $1 - \alpha_{\text{ind}}$ quantile of the standard normal distribution.

Fishman (2001) points out that the von Neumann test of independence is unlikely to reject H_0 when $\{X_i\}$ has an autocorrelation function that is negatively correlated and exhibits damped harmonic behavior around zero.

3. An Implementation

This section presents our implementation of procedures to estimate density. A flow chart of the procedure is in Figure 1. An imbedded pilot run is executed to set up the bin points; see Section 3.1 for details. On each iteration, the algorithm operates as follows. The simulation outputs are funneled into bins. The number of observations in each bin is updated dynamically as the observation is produced during the simulation run. The systematic samples are obtained through lag-l observations and are stored in a buffer. The initial value of l is 1. Let $l' = 1, 2, 3$ denote the lag of the systematic samples stored in the buffer. If lag-l' systematic samples appear to be dependent, then the lag l is doubled every other iteration and the process is repeated until the lag-l' systematic samples appear to be independent. The initial value of l' is 0 and will be updated each iteration by the following rule: "if $l' < 3$, then $l' = l' + 1$; else $l' = 2$."

3.1 Determine the Window Width

Scott and Factor (1981) point out that "the great potential of nonparametric density estimators in data analysis is not being fully realized, primarily because of the practical difficulty associated with choosing the smoothing parameter given only data $X_1, X_2, \ldots, X_n$." There are various data-based algorithms for determining the window width h for the kernel density estimate. Duin (1976) uses a modified maximum likelihood approach, Scott et al. (1977) use an iterative algorithm based

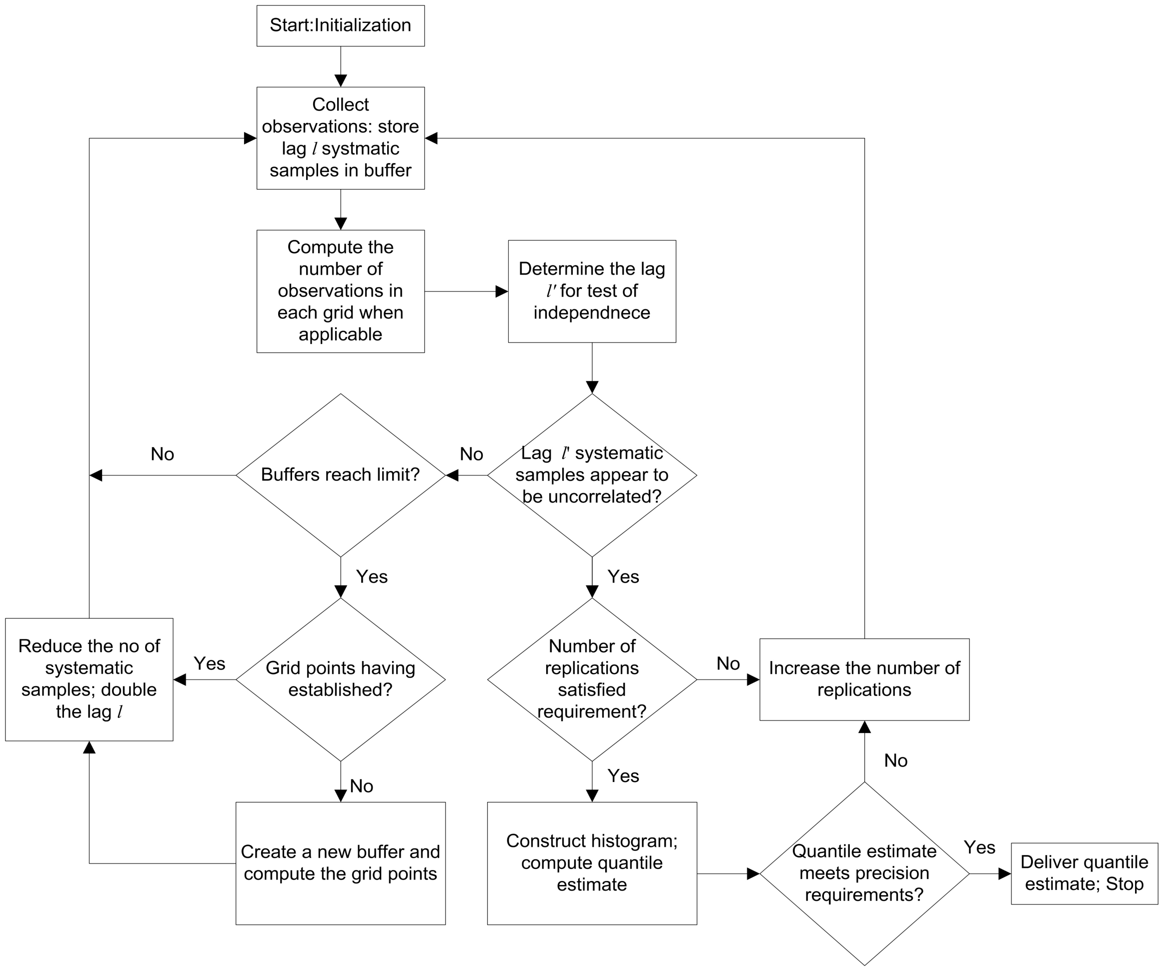

Figure 1: Flow Chart of the Procedure

on (2). Hearne and Wegman (1994) use random window widths. However, these algorithms are computationally intensive.

Silverman (1986, p. 47) suggests that the window width of the kernel estimator be $h = 0.9An^{-1/5}$, where $A = \min(\text{standard deviation, inter-quartile range}/1.34)$. For many purposes this will be an adequate choice of window width in terms of obtaining a small IMSE. For others, it will be a good starting point for subsequent fine tuning. Let x_p be the p sample quantile. In our procedure, we set $A = \min(\text{standard error}, (x_{0.75} - x_{0.25})/2.68)$. Let $x_{[1]}$ and $x_{[n]}$, respectively, denote the minimum and maximum of the initial n_0, $2n_0$, or $3n_0$ observations, depending on the correlation of the output sequences. Note that the sample sizes of the first three iterations are, respectively, $n_0, 2n_0$, and $3n_0$ (which is the size of the buffer). If $(x_{[n]} - x_{[1]})/(2h) < 25)$, then h will be halved. The choice of 25 to be the minimal initial number of windows is somewhat arbitrary and is based on our empirical results. This adjustment is needed for distributions that have relatively large variance with a small range of the initial observations, so too large a window width. For example, if the estimated number of bins is 10, the procedure increases the number of bins to 20 and reduces the window width by a factor of two.

We use the following strategy to determine the bin points. The are two categories of bins: main bins and auxiliary bins. Main bins are constructed based on the initial observations that "anchor" the bin of the simulation-generated histogram, while auxiliary bins are extensions of main bins to ensure that the bins cover future observations. The number of main bin points is $G_m = \lceil (x_{[n]} - x_{[1]})/(2h) \rceil$, and the number of auxiliary bin points is $G_a = 2\lceil \delta G_m \rceil$, where $0 < \delta < 1$. The total number of bin points is thus $G = G_m + G_a + 1$. Let the beginning indices of the main bin point (i.e., the origin) be $b = \lceil \delta G_m \rceil + 1$. The procedure sets $g_{b+i} = x_{[1]} + 2ih$, for $i = 0, 1, \ldots, G_m + G_a/2 - 1$, and $g_{b-i} = x_{[1]} - 2ih$, for $i = 1, 2, \ldots, G_a/2 - 1$. Bin point g_1 is set to $-\infty$ and g_G is set to ∞.

It is straightforward to compute the histogram density estimator. The array $n_i, i = 2, 3, \ldots, G$ stores the number of observations between bin points g_{i-1} and g_i, so the density of $x_i = (g_{i-1} + g_i)/2$ can be estimated by $\hat{f}(x_i) = (n_i/n)/(g_i - g_{i-1})$, where $n = \sum_{i=2}^{G} n_i$ is the total number of observations. To obtain the kernel estimator, the procedure needs to read through the output sequence again.

Scott (1985) shows that the smoothness of the frequency polygon can result in reduced bias and variance compared to a histogram. He also suggests that piece-wise quadratic estimates can achieve even closer approximation. Our procedure computes the point density estimator via the histogram by (four-point) Lagrange interpolation (Knuth, 1998). That is, for some k such that $x_{k-1} < x \leq x_k$, the x density point estimator can be computed as follows. Let

$$w_j = \prod_{j'=1, j' \neq j}^{4} \frac{x - x_{k+j'-3}}{x_{k+j-3} - x_{k+j'-3}}, \text{ for } j = 1, 2, 3, 4,$$

then $\hat{f}(x) = \sum_{j=1}^{4} w_j \hat{f}(x_{k+j-3})$. In two extreme cases, $x_1 < x \leq x_2$ or $x_{G-1} < x \leq x_G$, linear interpolation will be used.

Since the procedure uses interpolation to obtain point estimates, it eliminates the ragged character of the histogram. Hence, the density estimates for different points within the same bin can have different values.

3.2 Determine the Sample Size

The asymptotic validity of the density estimate is reached as the sample size or simulation run length gets large. However, in practical situations simulation experiments are restricted in time and it is not known in advance what the required simulation run length might be for the estimator to become essentially unbiased. Moreover, estimating the variance of the density estimator is needed to evaluate its precision. Therefore, a workable finite sample size must be determined dynamically for the precision required.

We use an initial sample size of $n_0 = 600$, which is somewhat arbitrary but is tested in our empirical results below. If the underlying sequence is only slightly correlated and high precision is desired, a larger initial sample size should be used. For correlated sequences, the sample size n will be replaced with $N = nl$. Here l will be chosen sufficiently large so that systematic samples that are lag-l observations apart are essentially uncorrelated; see Chen and Kelton (2003). This is possible because we assume the underlying process satisfies the property that the autocorrelation approaches zero as the lag approaches infinity. Consequently, the final sample size N increases as the auto-correlation increases. In this procedure, we use the von Neumann (1941) test for absence of correlation instead of the runs test. We can apply the von Neumann test with a smaller sample size, but it has less power. Nevertheless, it serves the purpose well, as demonstrated below in our empirical results.

Since we need to process the sequence again to obtain the kernel estimator, we re-compute the window width h with the final sample size N and the number of bin points with the new sample range. We need to allocate only the main bins because the minimum and maximum are known. Furthermore, the sample error and the quantiles $x_{0.25}$ and $x_{0.75}$ will be estimated through the histogram constructed while calculating the natural estimator. That is, the variance is conservatively estimated by

$$S_H^2 = \sum_{i=2}^{G} \max((g_{i-1} - \hat{\mu}(N))^2, (g_j - \hat{\mu}(N))^2) P_i.$$

Note that $N = nl = \sum_{i=2}^{G} n_i$, $\hat{\mu}(N) = \frac{1}{N}\sum_{j=1}^{N} X_j$, and $P_i = n_i/N$.

To estimate the error, the IMSE is approximated by

$$\overline{IMSE} = 2\sum_{r=1}^{\Gamma}\sum_{i=2}^{G-1}[\hat{f}_r(g_i) - f(g_i)]^2 h/\Gamma,$$

where Γ is the number of replications and $\hat{f}_r(\cdot)$ is the estimate in the r^{th} replication. The density of g_1 and g_G is not included in the calculation because they could be $-\infty$ and ∞, respectively. Furthermore, if the true minimum (ω) or the true maximum (Ω) are known, the values $g_i < \omega$ or $\Omega < g_i$ will not be included in the calculation.

3.3 Density Confidence Interval

An approximate point-wise confidence interval (c.i.) for the density $f(x)$ can be obtained using the binomial distribution from the histogram density estimate. The usual unbiased estimator of the variance of $\hat{f}_h(x)$ is $S_b^2 = \text{Var}(\hat{f}_h(x)) = p(1-p)/(4h^2N)$. This would then lead to the $100(1-\alpha)\%$ c.i. for $f(x)$, $\hat{f}_h(x) \pm z_{1-\alpha/2}S_b$. On the other hand, the distribution of $\hat{f}_{K,h}(x)$ is unknown, hence, a c.i. cannot be constructed through one replication of $\hat{f}_{K,h}(x)$.

Let $\hat{f}_r(x)$ denote the (histogram or kernel) estimator of $f(x)$ in the r^{th} replication. We use

$$\bar{f}(x) = \frac{1}{R}\sum_{r=1}^{R}\hat{f}_r(x)$$

as a point estimator of $f(x)$. Assuming $\bar{f}(x)$ has a limiting normal distribution, by the central limit theorem a c.i. for $f(x)$ using the i.i.d. $\hat{f}_r(x)$'s can be approximated using standard statistical procedures. That is, the ratio

$$T = \frac{\bar{f}(x) - f(x)}{S/\sqrt{R}}$$

would have an approximate t distribution with $R-1$ d.f. (degrees of freedom), where

$$S^2 = \frac{1}{(R-1)}\sum_{r=1}^{R}\left(\hat{f}_r(x) - \bar{f}(x)\right)^2$$

is the usual unbiased estimator of the variance of $f(x)$. This would then lead to the $100(1-\alpha)\%$ c.i., for $f(x)$,

$$\bar{f}(x) \pm t_{R-1,1-\alpha/2}\frac{S}{\sqrt{R}}, \tag{3}$$

where $t_{R-1,1-\alpha/2}$ is the $1-\alpha/2$ quantile for the t distribution with $R-1$ d.f. ($R \geq 2$).

Let the half-width H be $t_{R-1,1-\alpha/2}S/\sqrt{R}$. The final step in the procedure is to determine whether the c.i. meets the user's half-width requirement, a maximum absolute half-width ϵ' or a maximum relative fraction γ of the magnitude of the final point density estimator $\bar{f}(x)$. If the relevant requirement $H \leq \epsilon'$ or $H \leq \gamma|\bar{f}(x)|$ for the precision of the confidence interval is satisfied, then the procedure terminates, returns the point density estimator $\bar{f}(x)$, and the c.i. with half-width H. If the precision requirement is not satisfied with R replications, then the procedure will increase the number of replications to

$$(H/\epsilon')^2R \text{ or } (H/(\gamma\bar{f}(x)))^2R. \tag{4}$$

This step will be executed repeatedly until the half-width is within the specified precision.

Table 1: Properties of the QI subsequence at each iteration

Iteration	0	1_A	1_B	2_A	2_B	3_A	3_B	...	$k_A(k > 0)$	$k_B(k > 0)$
Total Observations	n	$2n$	$3n$	$4n$	$6n$	$8n$	$12n$	...	$2^k n$	$2^{k-1}3n$
Samples in Buffer	n	$2n$	$3n$	$2n$	$3n$	$2n$	$3n$	...	$2n$	$3n$
l_0	2^0	2^0	2^0	2^1	2^1	2^2	2^2	...	2^{k-1}	2^{k-1}
l'	1	2	3	2	3	2	3	...	2	3
l	$2^0 1$	$2^0 2$	$2^0 3$	$2^1 2$	$2^1 3$	$2^2 2$	$2^2 3$	...	$2^{k-1}2$	$2^{k-1}3$

3.4 The Density-Estimation Procedure

The procedure progressively increases the simulation run length N until a pre-determined number of systematic samples appear to be uncorrelated, as assessed by the von Neumann test. We allocate a buffer, QI, with size $t_1 = 3n$ to store our systematic samples y_i, $1 \leq i \leq t_1$ that will be used by the von Neumann test. Note that lag l' (=1,2,3) of the systematic samples is used to refer to systematic samples $y_{kl'+1}$, for $k = 0, 1, 2, \ldots, n-1$. Table 1 shows the total number of observations and other properties at each iteration. The *Total Observations* row shows the total number of observations at a certain iteration. The *Samples in Buffer* row shows the number of systematic samples stored in the buffer. The l_0 row shows the lag used to obtain the systematic samples stored in the buffer. The l' row shows the lag of the systematic samples that will be used for the von Neumann test. The l row shows the lag l at which the sequence appears to be uncorrelated if the lag-l' systematic samples passed the von Neumann test at that iteration.

For example, at the end of iteration 1_B, the total number of observations is $3n$, there are $3n$ systematic samples in the buffer, and the systematic samples are the lag-1 observations. At the beginning of iteration 2_A, we reduce the number of systematic samples in the buffer from $3n$ to $3n/2$ by discarding the even systematic samples; consequently, the systematic samples are lag-2 observations apart. We then generate another n observations and store the $n/2$ lag-2 observations at the buffer during iteration 2_A.

Note that the size of the buffer used to store the systematic samples is $t_1 = 3n$, l_0 is the lag used to obtain systematic samples, δ is the incremental sample size, and k is the index of iterations. Each iteration k contains two sub-iterations k_A and k_B. Note that we limit the number of systematic samples used in the von Neumann test to 600.

The quasi-independent-density-estimation algorithm:

1. Initialization: Set $n = 600$, $l_0 = 1$, $\delta = n$, and $k = 0$.

2. Generate δ systematic samples, which are lag-l_0 observations apart. If $k > 1$, record the number of observations in each bin.

3. If this is the initial iteration, set $l' = 1$. If this is a k_A iteration, set $l' = 2$. If this is a k_B iteration, set $l' = 3$.

4. Carry out the von Neumman test to assess whether lag-l' systematic samples appear to be uncorrleated.

5. If the lag-l' systematic samples appear to be uncorrleated, go to step 12.

6. If $k = 0$, set $k = k + 1$ and start the 1_Ath iteration by going to step 2.

7. If this is the 2_Ath iteration, then compute the bin points and the number of observations in each bin.

8. If this is a k_B iteration, set $k = k + 1$ and start a k_A iteration. If this is a k_A iteration, start a k_B iteration.

9. If this is a k_A iteration ($k > 1$), then discard the even systematic samples in the buffer, and re-index the rest of the $3n/2$ systematic samples in the first half of the buffer. Set $l_0 = 2^{k-1}$, $\delta = n/2$.

10. If this is a k_B iteration ($k > 1$), set $\delta = n$.

11. Go to step 2.

12. If the number of replications is less than specified, go to step 1.

13. Construct the confidence interval for $f(x)$ according to (3).

14. Let ϵ' be the desired absolute half-width criterion, or let $\gamma|\bar{f}(x)|$ be the desired relative half-width criterion. If the half-width of the c.i. is greater than ϵ' or $\gamma|\bar{f}(x)|$, compute R', the required number of independent replications according to (4), set $R = R'$, and go to step 1; otherwise the procedure returns the c.i. estimator and terminates.

3.5 Generating Random Variates from $\hat{f}$ and Estimating F

While our focus in this paper is on simulation *output* analysis, our methods can also be used for simulation *input* analysis, i.e., specifying input distributions for the simulation model. In order to carry out a simulation using random inputs such as service times, we have to specify their probability distributions. One of the approaches to specifying a distribution is as follows. First, we use the observations themselves to define an empirical density function $\hat{f}$ in some way. Then, we sample from this empirical density when an observation is needed.

In previous sections, we have discussed how to specify $\hat{f}$ from observations $(X_1, \ldots, X_n)$. In this section, we show how to generate a random variate Y from $\hat{f}$. Suppose $\hat{f}$ has been constructed by the ordinary kernel method with kernel K and window width h. Random variate Y from $\hat{f}$ can be generated as follows:

1. Generate I from discrete uniform $\{1, \ldots, n\}$ distribution.
2. Generate ξ from the kernel function K.
3. Return $Y = X_I + h\xi$.

This algorithm allows random values to be generated beyond the smallest or the largest observations. The bootstrap is an appealing approach to the assessment of errors and related quantities in statistical estimation. The method is described and explored in detail by Efron (1982). The bootstrap approach draws samples from the original values $X_1, \ldots, X_n$ with replacement and nearly every sample will contain repeated values. If n is small, most samples will contain some values repeated several times. The algorithm described in this section can be used to perform the smoothed bootstrap.

The cdf is $F(t) = \int_{-\infty}^{t} f(u)du$. Let $H(u)$ be the cdf of the kernel, i.e., $H(u) = \int_{-\infty}^{u} K(v)dv$. Then $\hat{F}(t)$, the estimate of $F(t)$, can be computed by

$$\hat{F}(t) = \frac{1}{n}\sum_{i=1}^{n} H\left(\frac{t - X_i}{h}\right).$$

In the case that $K(X) = \delta_t(X)$, then $H((t - X_i)/h) = 1$ when $X \leq t$. Note that $K((t - X_i)/h) = 1$ when $X_i = t$. Consequently, $\hat{F}(t) =$ the proportion of the samples that are equal to or less than t. Chen and Kelton (2008) have developed quantile-estimation procedures based on this property. Note that $\hat{f}(x)$ is an estimate of $f(x) = F'(x)$ and is a measure of the sensitivity of $F(x)$ as the value of x changes.

4. Empirical experiments

We tested the proposed procedure with several i.i.d. and correlated sequences. In these experiments, we used $R = 3$ independent replications to construct c.i.'s. We constructed density c.i.'s at four

Table 2: Coverage of 90% Confidence Density Estimators for the Tri-modal Distribution

avg N	666			
stdev N	119			
x	0.5	5	7.5	10
$f(x)$	0.1226	0.0665	0.0363	0.1360
Histogram Estimator (0.001856, 0.000756)				
coverage	88.1%	91.5%	91.0%	84.3%
avg γ	0.0515	0.0677	0.1007	0.0536
stdev γ	0.0389	0.0503	0.0749	0.0391
avg hw	0.0186	0.0150	0.0115	0.0192
stdev hw	0.0098	0.0078	0.0061	0.0106
Kernel Estimator (0.001207,0.000530)				
coverage	78.7%	88.7%	86.6%	63.3%
avg γ	0.0621	0.0574	0.1003	0.0808
stdev γ	0.0395	0.0418	0.0731	0.0442
avg hw	0.0147	0.0118	0.0090	0.0151
stdev hw	0.0076	0.0060	0.0048	0.0117

points for each distribution. The confidence level $1-\alpha$ of the density c.i. (i.e. (3)) is set to 0.90. Moreover, the confidence level of the von Neumann test for no correlation is set to 0.90 as well.

We tested the following independent sequences:

- Observations are i.i.d. from the tri-modal density $f(x) = \frac{1}{3\sqrt{2\pi}}(e^{-x^2/2}+\frac{1}{2}e^{-(x-5)^2/8}+e^{-(x-10)^2/2})$.
- Observations are i.i.d. from the exponential density

$$f(x) = \begin{cases} e^{-x} & \text{if } x \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

Tables 2 and 3 list the experimental results using the tri-modal and exponential distributions, respectively. Each design point was based on 1000 replications. The *avg N* row lists the average of the sample size of each independent run. The *stdev N* row lists the standard deviation of the sample size. The x row lists the point where we want to estimate the density. The $f(x)$ row lists the true density. The values after each of the estimation methods are the $\overline{IMSE}$ and the standard error of the integrated mean squared error. The *coverage* row lists the percentage of the c.i.'s that cover the true $f(x)$. The *avg* γ row lists the average of the relative precisions of the density estimators. Here, the relative precision is defined as $\gamma = |\hat{f}(x) - f(x)|/f(x)$. The *stdev* γ row lists the standard deviation of the relative precision of the density estimators. The *avg hw* row lists the average of the c.i. half-widths. The *stdev hw* row lists the standard deviation of the c.i. half-width.

As expected, the $\overline{IMSE}$ from the kernel estimator is better than from the histogram estimator. Even though the kernel estimator requires more computation, the additional computational cost

Table 3: Coverage of 90% Confidence Density Estimators for the expon(1) Distribution

avg N	667			
stdev N	124			
x	0.5	2.0	3.0	5.0
$f(x)$	0.6065	0.1353	0.0498	0.0067
Histogram Estimator (0.007100, 0.003581)				
coverage	90.4%	91.5%	90.0%	73.2%
avg γ	0.0443	0.0950	0.1643	0.3612
stdev γ	0.0340	0.0699	0.1240	0.7372
avg hw	0.0885	0.0426	0.0263	0.0057
stdev hw	0.0491	0.0226	0.0137	0.0141
Kernel Estimator (0.003114, 0.001654)				
coverage	88.7%	90.4%	89.1%	86.5%
avg γ	0.0316	0.0759	0.1304	0.4824
stdev γ	0.0248	0.0563	0.0989	2.9082
avg hw	0.0627	0.0339	0.0207	0.0100
stdev hw	0.0343	0.0182	0.0109	0.0572

is minimal with today's computers. In these experiments, no relative or absolute precisions were specified, so the half-width of the c.i. is the result of the default precision. Even though the coverages are close to the nominal values of 90%, the relative precision of the half-width is large. Note that the half-width can be shortened with a larger sample size; see step 14 of the procedure in Section 3.4. In general, the histogram estimator has larger variance, so better c.i. coverage. However, the histogram estimators are biased high around the tail area. This is because the histogram estimators often result in a bounded distribution, i.e., the tail of the distribution is truncated. With $\alpha = 0.10$, the independent sequences will fail the test for no correlation 10% of the time. The average sample sizes, 666 and 667, are close to the theoretical value, i.e., $\sum_{i=0}^{\infty} n_0 \alpha^i$, where $n_0 = 600$.

Figures 2 and 3, respectively, show the empirical and true densities of the tri-modal and exponential distributions, generated from the first run of the kernel estimate of our experiments. These figures indicate good approximation and reveal the essential characteristics of the underlying density functions. The empirical tri-modal distribution correctly reveals that there are three modes. Even though the estimates do not exactly reflect that the heights of the two greater modes are equal, the accuracy and precision of the approximation is still good, and can be increased by increasing the sample size. The exponential distribution has a steep slope, so has a smaller window width h and has a more ragged empirical distribution curve. Furthermore, the kernel estimate over-smooths the bounded tail density of the exponential distribution. Even though it is not plotted in the figure, the empirical density curve actually continues in the left tail into the negative region of x. To deal

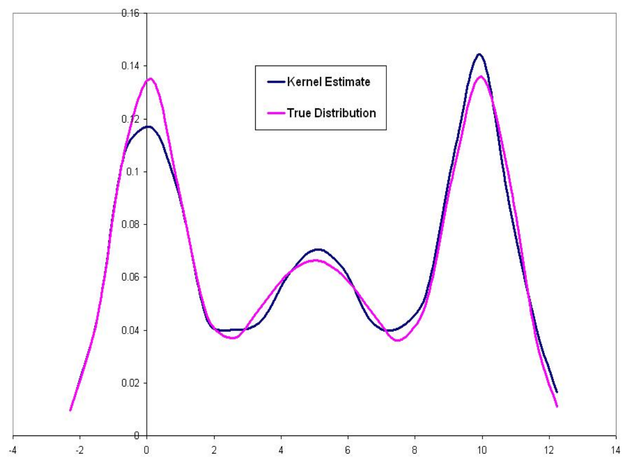

Figure 2: Empirical Density of the Tri-modal Distribution

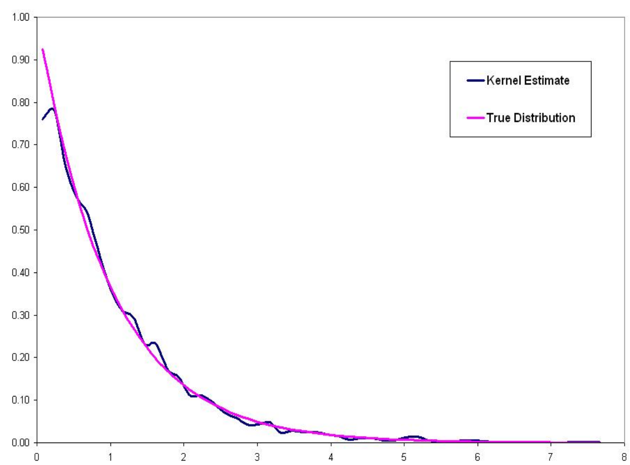

Figure 3: Empirical Density of the Exponential Distribution

Table 4: Coverage of 90% Confidence Density Estimators for the AR1(0.9) Process

avg N	20965			
stdev N	3836			
x	-0.5	0.0	1.0	2.0
$f(x)$	0.1698	0.1739	0.1581	0.1189
Histogram Estimator (0.000221, 0.000133)				
coverage	89.6%	89.7%	89.1%	90.4%
avg γ	0.0138	0.0132	0.0147	0.0192
stdev γ	0.0107	0.0105	0.0111	0.0148
avg hw	0.0075	0.0076	0.0077	0.0074
stdev hw	0.0041	0.0042	0.0041	0.0038
Kernel Estimator (0.000209, 0.000128)				
coverage	88.7%	88.9%	90.7%	90.4%
avg γ	0.0142	0.0133	0.0151	0.0190
stdev γ	0.0107	0.0103	0.0110	0.0142
avg hw	0.0074	0.0076	0.0077	0.0074
stdev hw	0.0041	0.0042	0.0041	0.0038

with this difficulty, various adaptive methods have been proposed; see Silverman (1986, pp. 19-29).

We also tested the following correlated sequences:

- Observations are from the AR1 (first-order auto-regressive) process $X_i = \mu + \varphi(X_{i-1} - \mu) + \epsilon_i$ for $i = 1, 2, \ldots$, where

  $$E(\epsilon_i) = 0, \quad E(\epsilon_i \epsilon_j) = \begin{cases} \sigma^2 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases},$$

  $$-1 < \varphi < 1.$$

  The ϵ_i's are commonly called *error terms.*

- Observations are the delays in queue (exclusive of service times) from the M/M/1 queuing model.

The AR1 process shares many characteristics observed in simulation output processes, including asymptotic first- and second-order stationarity, and autocorrelations that decline exponentially with increasing lag. We set φ to 0.90 and set μ to zero for this experiment. In order to eliminate the initial bias, X_0 is set to a random variate drawn from the steady-state distribution $N(0, \frac{1}{1-\varphi^2})$.

Table 4 lists the experimental results of the AR1 process. The c.i. coverage of these four design points are near the specified 90% confidence level for both estimators. The simulation run length generally increases as the correlation coefficient φ of the AR1 process increases. The simulation run length of the AR1 process with $\varphi = 0.9$ is much larger than for independent sequences and

Table 5: Coverage of 90% Confidence Density Estimators for the MM1(0.9) Process

avg N	418496			
stdev N	92729			
x	0.5	2.5	5.0	10.0
$f(x)$	0.0856	0.0701	0.0546	0.0331
Histogram Estimator (0.000021, 0.000020)				
coverage	85.0%	90.8%	92.6%	91.2%
avg γ	0.0295	0.0102	0.0082	0.0093
stdev γ	0.0232	0.0083	0.0064	0.0070
avg hw	0.0050	0.0023	0.0015	0.0010
stdev hw	0.0034	0.0013	0.0008	0.0006
Kernel Estimator (0.0000021, 0.000017)				
coverage	0.0%	91.2%	91.3%	90.4%
avg γ	0.9355	0.0086	0.0082	0.0096
stdev γ	0.0291	0.0063	0.0063	0.0076
avg hw	0.0066	0.0020	0.0015	0.0010
stdev hw	0.0037	0.0011	0.0008	0.0006

consequently produces much smaller $\overline{IMSE}$ and smaller values of the relative precisions of the half-width.

The waiting-time density of the stationary M/M/1 delay in queue is $f(x) = (\nu - \lambda)\frac{\lambda}{\nu}e^{-(\nu-\lambda)x}$ for $x \geq 0$ and is discontinuous at $x = 0$, where λ is the arrival rate and ν is the service rate. A summary of our experimental results of the M/M/1 delay-in-queue process is in Table 5. Except for the kernel density estimate of $x = 0.5$, the c.i. coverages are above or close to the specified 90%. The kernel method encounters difficulty when estimating $f(0.5)$ for the M/M/1 queuing process because the value 0.5 is close to the discontinuity point 0. However, how close is "too close" for the method to encounter difficulty depends on many factors, such as the underlying distribution and the kernel function. Various adaptive methods are available to handle this situation.

Figures 4 and 5, respectively, show the empirical distributions of the AR1 process with $\phi = 0.9$ and the M/M/1 delay-in-queue process with $\rho = 0.90$, generated from the first run of our experiments. The theoretical steady-state distributions of this AR1 process and this M/M/1 queuing process are, respectively, $N(0, 1/0.18)$ and $1 - 0.9e^{-0.1x}$, where $x \geq 0$. Again, our experimental results show that these density estimates provide excellent approximations to the underlying steady-state distributions. However, the kernel estimator over-smooths the density around the discontinuity point. Because the observations are correlated, the procedure allocates more samples. Consequently, the accuracy and precision of the empirical distributions are better.

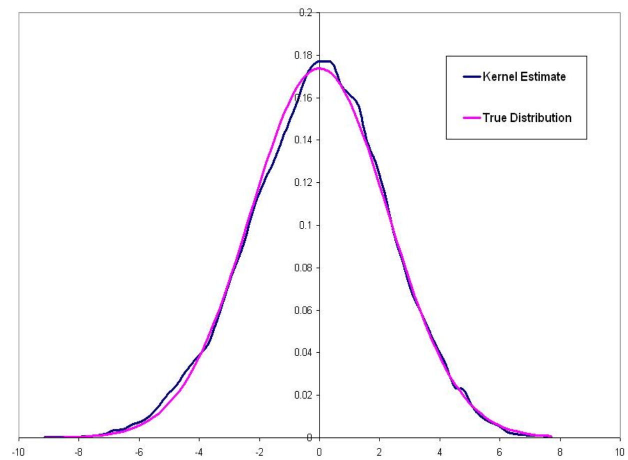

Figure 4: Empirical Density of the AR1 Process

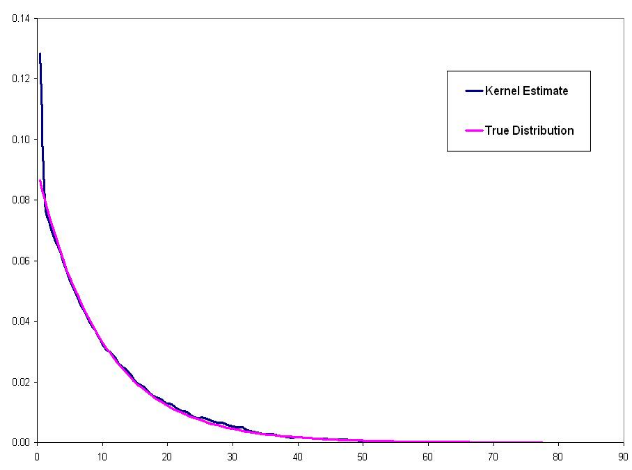

Figure 5: Empirical Density of the M/M/1 Process

5. Concluding remarks

We have evaluated two algorithms for estimating the density $f(x)$ of a stationary stochastic process, with or without intra-process independence. Since, to obtain the kernel estimator, the procedure needs to compute the histogram estimator, a prudent course is to choose any reasonable estimate based on these two estimates that are consistent with prior belief about the true sampling density. However, the histogram procedure is more suitable as a generic density-estimation procedure since it requires less computation, delivers a valid c.i., and has no difficulty estimating the density around a bounded tail or discontinuity point, though its $\overline{IMSE}$ is generally larger. While our focus in this paper is on simulation output analysis, our methods can also be used for simulation input analysis, i.e., specifying input distributions for the simulation model.

Some density estimates require more observations than others before the asymptotics necessary for density estimates become approximately valid. Our algorithm works well in determining the required simulation run length for the asymptotic approximation to become valid. The results from our empirical experiments show that the procedure is excellent in achieving the pre-specified accuracy. Our proposed histogram-approximation algorithm computes quantiles only at bin points and uses Lagrange interpolation to estimate the density at certain points. Consequently, the density at different points within the same bin can have different values. The algorithm also generates an empirical distribution (histogram) of the output sequence, which can provide insights into the underlying stochastic process.

Our approach has the desirable properties that it is a sequential procedure and it does not require users to have *a priori* knowledge of values that the data might take on. This allows the user to apply this method without having to execute a separate pilot run to determine the range of values to be expected, or guess and risk having to re-run the simulation. The main advantage of our approach is that, by using a straightforward test for lack of correlation to determine the simulation run length and obtain quantiles at bin points, we can apply classical statistical techniques directly and do not require more advanced statistical theory, thus making it easy to understand, and simple to implement.

References

Billingsley P. *Convergence of Probability Measures*. 2nd ed. New York: John Wiley & Sons, Inc., 1999.

Chen EJ, Kelton WD. Determining Simulation Run Length with the Runs Test. *Simulation Modelling Practice and Theory* 2003;11(3-4):237-250.

Chen EJ, Kelton WD. Empirical Evaluation of Data-Based Density Estimation. *Proceedings of the 2006 Winter Simulation Conference*. 2006; 333-341.

Chen EJ, Kelton WD. Estimating Steady-State Distributions via Simulation-Generated Histograms. *Computers and Operations Research* 2008;35:1003-1016.

Devroye L, Györfi L. *Nonparametric Density Estimation: the L_1 View.* New York: John Wiley & Sons, Inc., 1985.

Devroye L, Lugosi G. *Combinatorial Methods in Density Estimation.* New York: Springer-Verlag, 2001.

Duin, RPW. On the Choice of Smoothing Parameters for Parzen Estimators of Probability Density Functions. *IEEE Transactions on Computers* 1976; C-25:1175-1179.

Efron B. *The Jackknife, the Bootstrap and Other Resampling Plans.* Philadelphia:SIAM. 1982.

Fishman GS. *Discrete-Event Simulation: Modeling Programming and Analysis.* New York: Springer-Verlag, 2001.

Fryer MJ. Some Errors Associated with the Nonparametric Estimation of Density Functions. *J. Inst. Maths. Applics.* 1976; 18:371-380.

Hearne LB, Wegman EJ. Fast multidimensional density estimation based on random-width bins. *Computing Science and Statistics* 1994; 26:150-155.

Hogg RV, Craig AT. *Introduction to Mathematical Statistics.* Fifth ed. New Jersey:Prentice Hall, 1995.

Knuth DE. *The Art of Computer Programming*, Vol. 2, Third Edition. Reading, Mass.: Addison-Wesley, 1998.

Rosenblatt M. Curve Estimates. *Annals of Mathematical Statistics* 1971;42:1815-1842.

Sain SR, Scott DW. Zero-Bias Bandwidths for Locally Adaptive Kernel Density Estimation. *Scandinavian Journal of Statistics* 2002; 29:441–460.

Scott DW. On optimal and data-based frequency ploygons. *J. Amer. Statist. Assoc.* 1985;80:348-354.

Scott DW, Factor LE. Monte Carlo Study of Three Data-Based Nonparametric Probability Density Estimators. *Journal of the American Statistical Association* 1981;76 (373):9–15.

Scott DW, Sain SR. 2004. "Multi-dimensional Density Estimation" in Handbook of Statistics. Vol 23: Data Mining and Computational Stattics. eds: CR Rao and EJ Wegman, Elsevier, Amsterdam.

Scott DW, Tapia RA, and Thompson JR. Kernel Density Estimation Revisit. *Journal of Nonlinear Analysis, Theory, Methods and Applications* 1977;1:339-372.

Sen PK. On the Bahadur Representation of Sample Quantiles for Sequences of ϕ-mixing Random Variables. *Journal of Multivariate Analysis* 1972;2(1):77–95.

Silverman BW. *Density Estimation for Statistics and Data Analysis.* New York: Chapman and Hall. 1986.

von Neumann J. Distribution of the Ratio of the Mean Square Successive Difference and the Variance. *Annals of Mathematical Statistics* 1941;12:367–395.

Biographies

E. Jack Chen is a Senior Staff Specialist with BASF Corporation. He received a Ph.D. in Quantitative Analysis from the University of Cincinnati. His research interests are in the areas of computer simulation, statistical data analysis, and stochastic processes.

W. David Kelton is a Professor in the Department of Quantitative Analysis and Operations Management at the University of Cincinnati. He received a B.A. in mathematics from the University of Wisconsin-Madison, an M.S. in mathematics from Ohio University, and M.S. and Ph.D. degrees in industrial engineering from Wisconsin. His research interests and publications are in the probabilistic and statistical aspects of simulation, applications of simulation, and stochastic models. He served as Editor-in-Chief of the *INFORMS Journal on Computing* from 2000 through 2006.