Notes on exponentially correlated stationary Markov processes

John Kerl

March 29, 2010

Abstract

We make concrete various [CB, GS, Berg] ideas regarding autocorrelation of stationary Markov processes, with the particular goal of placing error bars on sample means. We focus on processes where the autocorrelation takes the form of a single exponential. We define a particular toy-model process, the *correlated-uniform Markov process*, which is exactly solvable. (This is in contrast to the typical Markov chain Monte Carlo process: in the MCMC field, one resorts to experimental methods only for systems which are *not* exactly solvable.) When a practitioner applies new methods to an MCMC process which is itself under examination, it can be difficult to identify computational problems which arise. Using this toy-model process, we elucidate strengths and shortcomings of autocorrelation and its estimators, clearly separating properties of the estimators themselves from the properties of the particular Markov process. The policies developed herein will be used to design and analyze MCMC experiments for the author's doctoral dissertation.

Contents

List of Figures

List of Tables

1 Problem statement

The following problem occurs throughout Markov chain Monte Carlo (MCMC) experiments. Let X_t be an identically distributed, but not necessarily independent, Markov process; let μ_X and σ_X^2 be the common mean and variance, respectively. (We will construct a specific process Y_t with the same properties. We reserve the notation X_t for a general process with these properties.) Given a time-series realization $X_0, \ldots, X_{N-1}$, the sole desired expressed in this paper is to to estimate μ_X, with an error bar on that estimate. The presence of correlations between the X_t's make this process more complicated than in the IID case.

The standard estimator for μ_X is the sample mean, $\overline{X}_N$. Given a time-series realization $X_0, \ldots, X_{N-1}$, we compute a single value of $\overline{X}_N$. Since the X_t's are random variables, $\overline{X}_N$ is itself a random variable. When we conduct M such experiments, we will get M different values of $\overline{X}_N$. (We will quantify below the dependence of the variance, or error bar, of $\overline{X}_N$, upon the autocorrelation of the process X_t.) Suppose for the sake of discussion that the autocorrelation is exponential: $\mathrm{Corr}(X_i, X_j) = \eta^{|i-j|}$ for some $\eta \in [0, 1)$. Then $\eta = 0$ is the IID case, and higher η's correspond to more highly correlated processes. A few such realizations are shown in figure 1.

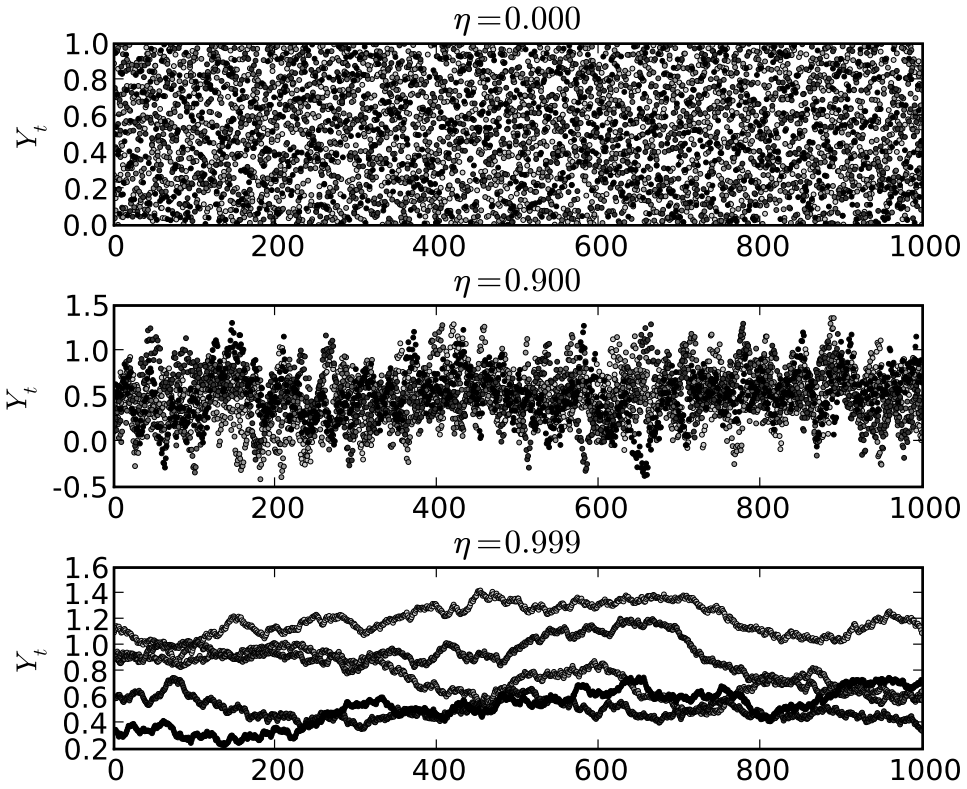

Figure 1: Five realizations each of the correlated-uniform Markov process Y_t with $\eta = 0.0, 0.9, 0.999$.

For any process $W_1, \ldots, W_K$, write $m_K(W)$ for the sample mean and $s_K^2(W)$ for the sample variance, the unbiased estimator of $\mathrm{Var}(W)$. Then:

- $m_N(X_t)$, which is $\overline{X}_N$, estimates μ_X. This is the *sample mean*, taken over N samples.
- $s_N^2(X_t)$ estimates σ_X^2. This is the *sample variance*, taken over N samples.

- $m_M(\overline{X}_N)$ estimates $\mu_{\overline{X}_N}$. This is also referred to as the sample mean; it is taken over MN samples.
- $s_M^2(\overline{X}_N)$ uses MN data points to estimate $\sigma^2_{\overline{X}_N}$, which is the *variance of the sample mean*.
- In the IID case, the true variance of the sample mean is $\sigma^2_{\overline{X}_N} = \sigma_X^2/N$; $t_N^2(X_t) = s_N^2(X_t)/N$ is the *naive estimator* of the variance of the sample mean, using N data points. It is an unbiased estimator only in the IID case.
- $u_N^2(X_t)$ is the *corrected estimator* of $\sigma^2_{\overline{X}_N}$. The estimators $t_N^2(X_t)$ and $u_N^2(X_t)$ will be discussed graphically, numerically, and theoretically below. The estimated *integrated autocorrelation time* $\hat{\tau}_{\text{int}}$ will be used to compute $u_N^2(X_t)$ from $t_N^2(X_t)$.
- $\text{Var}(u_N^2(X_t))$ is the *error of the error bar*. It turns out that $u_N^2(X_t)$ is a rough estimator for $\text{Var}(\overline{X}_N)$, and $\text{Var}(u_N^2(X_t))$ increases with η. The very name "error of the error bar" sounds overwrought; yet, it is a necessary consideration in MCMC experiments, and must be thought through.

The processes Y_t of figure 1, to be defined explicity in section 4, have $\mu_Y = 1/2$ and $\sigma_Y^2 = 1/12$, regardless of the autocorrelation exponent η. (Note that $\sqrt{1/12} \approx 0.2887$.) We observe the following behavior from the aforementioned estimators. (See figures 7 through 10 starting on page 15, and table 2 on page 16.)

- For all η, $\overline{Y}_N$ is unbiased for μ_Y. Its uncertainty widens visibly with autocorrelation exponent η. This uncertainty is the quantity of interest.
- Quantitatively, $s_M(\overline{Y}_N)$ gives a good idea of this increasing uncertainty. However, $s_M(\overline{Y}_N)$ requires M experiments, where M may be unacceptably large. If we were always willing to conduct such a large number of experiments, it would not be necessary to write this paper. We wish to estimate the variance of the sample mean *using only one experiment* $Y_0, \ldots, Y_{N-1}$. This is the rub.
- The corrected estimator $u_N^2(Y_t)$ corresponds roughly with $s_M(\overline{Y}_N)$, and moreover is computed from a single experiment $Y_0, \ldots, Y_{N-1}$. The roughness of the approximation of the error bar is acceptable: it is only an error bar.

To summarize, $s_M^2(\overline{X}_N)$ is a multi-experiment estimator for the variance of the sample mean; $u_N^2(X_t)$ is a single-experiment estimator. The former is of higher quality, but is more expensive to obtain; the latter carries its own uncertainty which worsens as the autocorrelation η increases.

Having motivated the problem, we now develop the notation and theory to make all of these ideas precise.

2 Autocovariance and autocorrelation

Definition 2.1. A Markov process X_t, $t = 0, 1, 2, \ldots$, is *stationary* if the X_t's have a common mean $\mu_X = \mathbb{E}[X_t]$ and variance $\sigma_X^2 = \text{Var}(X_t)$.

Definition 2.2. Let X_t be a stationary Markov process with $\mathbb{E}[X_t] = \mu_X$ and $\text{Var}(X_t) = \sigma_X^2$. The *autocovariance* and *autocorrelation* of X_t, respectively, are

$$C(t) = \text{Cov}(X_0, X_t) = \mathbb{E}[X_0 X_t] - \mathbb{E}[X_0]\mathbb{E}[X_t] = \mathbb{E}[X_0 X_t] - \mu_X^2$$

$$c(t) = \text{Corr}(X_0, X_t) = \frac{\mathbb{E}[X_0 X_t] - \mathbb{E}[X_0]\mathbb{E}[X_t]}{\sigma_{X_0}\sigma_X} = \frac{\mathbb{E}[X_0 X_t] - \mu_X^2}{\sigma_X^2}.$$

Remark 2.3. In the literature, what we call the autocovariance is often referred to as autocorrelation. This incorrect and misleading terminology is, sadly, quite widespread.

Remark 2.4. Recall that, as with all correlations, the autocorrelation takes values between -1 and 1.

3 The IID uniform process

Here we recall familiar [CB, GS] facts about random numbers U which are uniformly distributed on a closed interval $[a, b]$. These will be used as building blocks in section 4. Writing the probability density function of U as $f_U(x)$, we have

$$f_U(x) = \frac{1}{b-a} \cdot 1_{[a,b]}(x) \qquad \mu_U = \frac{1}{b-a}\int_a^b x\,dx = \frac{a+b}{2}$$

$$\mu_U^2 + \sigma_U^2 = \mathbb{E}[U^2] = \frac{1}{b-a}\int_a^b x^2\,dx = \frac{a^2+ab+b^2}{3} \qquad \sigma_U^2 = \frac{(b-a)^2}{12}.$$

Now consider an IID sequence $\{U_i\}$ of such random variables, indexed by the integers. We develop a particularly phrased formula which will simplify the calculations in section 4. Note that if X_1, X_2 are IID with common mean μ_X and variance σ_X^2, then $\mathbb{E}[X_1^2] = \mu_X^2 + \sigma_X^2$ whereas $\mathbb{E}[X_1 X_2] = \mu_X^2$. For a sequence of IID X_i's, including our particular uniform U_i's, this means

$$\mathbb{E}[X_i X_j] = \mu_X^2 + \delta_{ij}\sigma_X^2. \tag{3.1}$$

4 The correlated-uniform Markov process

This paper addresses correlated Markov processes, focusing in particular on those with exponential autocorrelation. Here we construct a simple process for which the mean, variance, and autocorrelation are exactly solvable. In particular, the autocorrelation will be controlled by a parameter $\eta \in [0, 1]$, while the mean and variance will be the same as for IID $U(0, 1)$.

Definition 4.1. Let U be uniformly distributed on $[a, b]$ as in the previous section, where $a < b$ are left variable for the moment. Let $0 \le \eta \le 1$ and $a < b$. The *correlated-uniform Markov process* Y_t is defined by $Y_0 \sim U(a, b)$, and for $t \ge 1$,

$$Y_t \quad = \quad \eta Y_{t-1} + (1-\eta)U_t \quad = \quad \eta^t U_0 + (1-\eta)\sum_{i=1}^{t} \eta^{t-i} U_i \tag{4.2}$$

where the first equality is a definition and the second equality follows by an easy induction argument.

Remark. Note that $\eta = 0$ is the IID case from the previous section; $\eta = 1$ would give a constant process with zero variance. The η parameter is the control knob with which we specify the autocorrelation of the process, as will be made precise in section 5.

Definition 4.3. Closely related to this is the *correlated-uniform stationary Markov process* (or asymptotic process)

$$Y_t = (1-\eta)\sum_{i=-\infty}^{t} \eta^{t-i} U_i. \tag{4.4}$$

In practice, we will run the original process for a number of time steps s until $\eta^s \approx 0$, such that the $\eta^s U_0$ term of equation (4.2) dies out, then consider the values of the process only from that time forward. In that regime, the process of definition 4.3 is an approximation to that of definition 4.1, but it is easier to manipulate algebraically.

We seek a, b such that the mean and variance of Y_t do not depend on η. The mean is immediate:

$$\mathbb{E}[Y_t] = (1-\eta) \sum_{i=-\infty}^{t} \eta^{t-i} \mathbb{E}[U_i] = \frac{a+b}{2}.$$

The variance $\mathrm{Var}(Y_t)$ is a special case of the covariance $\mathrm{Cov}(Y_t, Y_{t+k})$, which will be needed below. Equation (3.1) and expressions for geometric sums give us

$$\begin{aligned}
\mathbb{E}[Y_t Y_{t+k}] &= (1-\eta)^2 \eta^{2t+k} \sum_{i=-\infty}^{t} \eta^{-i} \sum_{j=-\infty}^{t+k} \eta^{-j}\, \mathbb{E}[U_i U_j] \\
&= (1-\eta)^2 \eta^{2t+k} \sum_{i=-\infty}^{t} \eta^{-i} \sum_{j=-\infty}^{t+k} \eta^{-j}\, (\mu_U^2 + \delta_{ij}\sigma_U^2) \\
&= \mu_U^2 (1-\eta)^2 \eta^{2t+k} \sum_{i=-\infty}^{t} \eta^{-i} \sum_{j=-\infty}^{t+k} \eta^{-j} + \sigma_U^2 (1-\eta)^2 \eta^{2t+k} \sum_{j=-\infty}^{t} \eta^{-2j} \\
&= \mu_U^2 + \sigma_U^2\, \eta^k \left(\frac{1-\eta}{1+\eta}\right).
\end{aligned}$$

Then

$$\mathrm{Var}(Y_t) = \sigma_U^2 \left(\frac{1-\eta}{1+\eta}\right) = \frac{(b-a)^2}{12} \left(\frac{1-\eta}{1+\eta}\right).$$

Now we may solve for a and b such that μ_U and σ_U are the same as for IID $U(0,1)$, namely, $1/2$ and $1/12$ respectively. Solving the pair of equations

$$\frac{a+b}{2} = \frac{1}{2} \qquad \text{and} \qquad \frac{(b-a)^2}{12}\left(\frac{1-\eta}{1+\eta}\right) = \frac{1}{12},$$

we obtain

$$a = \frac{1}{2}\left(1 - \sqrt{\frac{1+\eta}{1-\eta}}\right) \qquad \text{and} \qquad b = \frac{1}{2}\left(1 + \sqrt{\frac{1+\eta}{1-\eta}}\right). \tag{4.5}$$

Note in particular that for $\eta = 0$, the IID case, we recover $a = 0, b = 1$ as expected. Figure 2 shows some realizations for $\eta = 0.0, 0.5, 0.9$. For $\eta = 0.9$, correlations are clearly visible. Also note that there is a burn-in time required for the process to forget its initial state Y_0. In this figure, the asymptotic formula of definition 4.3 appears valid for $t > 50$ or so, at which point $\eta^t = 0.9^{50} \approx 0.005 \approx 0$. This burn-in phenomenon is discussed in more detail in section 8.

The following is pseudocode (technically, it is Python code, which is largely the same thing) for displaying N steps of Y_t, given the correlation-control parameter η and the number N_{therm} of burn-in iterates to be discarded:

```
s = sqrt((1+eta)/(1-eta)); a = 0.5 * (1 - s); b = 0.5 * (1 + s)

Y = random.uniform(a, b) # Burn-in iterates
for k in range(0, Ntherm):
    U = random.uniform(a, b)
    Y = eta * Y + (1-eta) * U
```

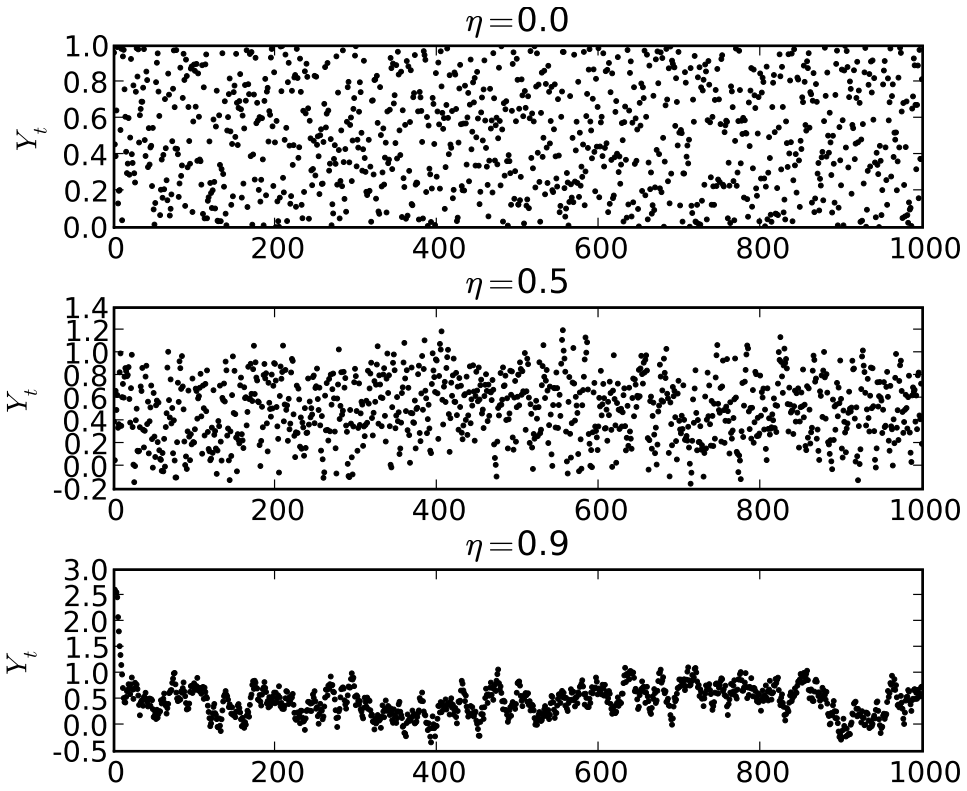

Figure 2: Realizations of the correlated-uniform Markov process Y_t with $\eta = 0.0, 0.5, 0.9$. Burn-in iterates are included.

```
for k in range(0, N):    # Iterates to be displayed
    U = random.uniform(a, b)
    Y = eta * Y + (1-eta) * U
    print Y
```

5 Statistics of the correlated-uniform Markov process

We now write all statistics of the correlated-uniform Markov process Y_t in terms of η. With a and b in terms of η (equation (4.5)), we have

$$\mu_U = \frac{a+b}{2} = \frac{1}{2}, \qquad \sigma_U^2 = \frac{(b-a)^2}{12} = \frac{1}{12}\left(\frac{1+\eta}{1-\eta}\right),$$

$$\mathbb{E}[Y_t] = \mu_Y = \frac{a+b}{2} = \frac{1}{2}, \qquad \mathrm{Var}(Y_t) = \sigma_Y^2 = \frac{(b-a)^2}{12}\left(\frac{1-\eta}{1+\eta}\right) = \frac{1}{12};$$

$$\mathbb{E}[Y_t Y_{t+k}] = \mu_U^2 + \sigma_U^2 \eta^k \left(\frac{1-\eta}{1+\eta}\right) = \frac{1}{4} + \frac{\eta^k}{12},$$

$$\mathrm{Cov}(Y_t, Y_{t+k}) = \mathbb{E}[Y_t Y_{t+k}] - \mathbb{E}[Y_t]\mathbb{E}[Y_{t+k}] = \frac{\eta^k}{12},$$

$$\mathrm{Corr}(Y_t, Y_{t+k}) = \frac{\mathrm{Cov}(Y_t, Y_{t+k})}{\sigma_Y \sigma_{Y_{t+k}}} = \eta^k.$$

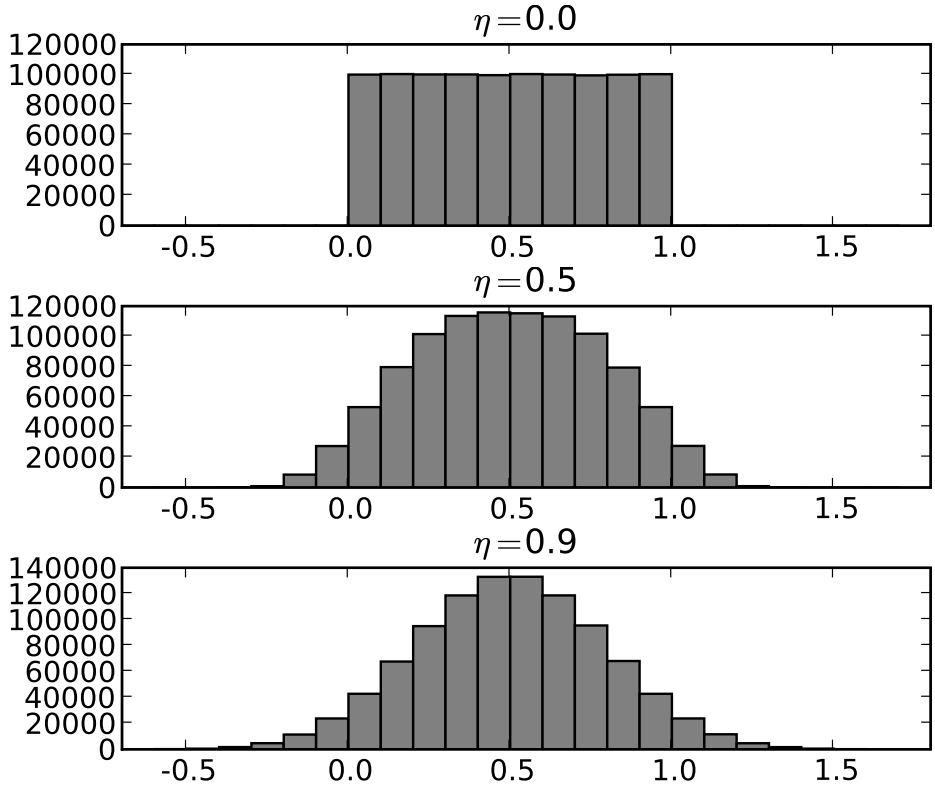

Figure 3: Histograms of the correlated-uniform Markov process Y_t with $\eta = 0.0, 0.5, 0.9$: 10^6 iterates, bins of 0.1 from -0.7 to 1.7. Burn-in iterates have been discarded.

The remaining step needed to completely specify the correlated-uniform Markov process is to write down the PDF of Y_t. This could be done using convolutions, since Y_t is a weighted sum (weighted by powers of η) of IID uniform random variables. The algebra is messy, though, and an expression for the PDF is not needed in this work. It is sufficient to point out the following: (*i*) For $\eta = 0$, the density is uniform on $[0, 1]$. (*ii*) For η close to 1, which is the case of interest in this work, the density closely resembles a normal with mean 1/2 and variance 1/12. The support is compact, so the density cannot be Gaussian, but the support is wide enough to include substantial tail mass. See figure 3 for empirical histograms.

6 The variance of the sample mean

When we use the data from an MCMC simulation to compute the sample mean of a random variable, the next order of business is to place an error bar on that sample mean.

As before, let X_t be a stationary Markov process with common mean μ_X, variance σ_X, and autocorrelation $\text{Corr}(X_t, X_{t+k}) = \eta^k$. Given $X_0, \ldots, X_{N-1}$, the sample mean $\overline{X}_N$ is an unbiased estimator of μ_X:

$$\overline{X}_N = \frac{1}{N} \sum_{i=0}^{N-1} X_i.$$

By linearity of expectation, $\mathbb{E}[\overline{X}_N] = \mu_X$. To find the variance of $\overline{X}_N$, we first need $\mathbb{E}[\overline{X}_N{}^2]$. This is

$$\mathbb{E}[\overline{X}_N{}^2] = \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \mathbb{E}[X_i X_j].$$

Since

$$\text{Corr}(X_i, X_j) = \eta^{|i-j|} = \frac{\mathbb{E}[X_i, X_j] - \mu_X^2}{\sigma_X^2},$$

we have

$$\mathbb{E}[X_i X_j] = \mu_X^2 + \sigma_X^2 \eta^{|i-j|}. \tag{6.1}$$

Then

$$\begin{aligned}
\mathbb{E}[\overline{X}_N{}^2] &= \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (\mu_X^2 + \sigma_X^2 \eta^{|i-j|}) = \mu_X^2 + \frac{\sigma_X^2}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \eta^{|i-j|} \\
&= \mu_X^2 + \frac{\sigma_X^2}{N^2} \left[\sum_{i=0}^{N-1} 1 + \sum_{i=0}^{N-2} \eta^{-i} \sum_{j=i+1}^{N-1} \eta^j + \sum_{i=1}^{N-1} \eta^i \sum_{j=0}^{i-1} \eta^{-j} \right].
\end{aligned}$$

Applying geometric-sum formulas and several lines of algebra, we get

$$\mathbb{E}[\overline{X}_N{}^2] = \mu_X^2 + \frac{\sigma_X^2}{N} + \frac{2\sigma_X^2 \eta}{N^2(1-\eta)} \left[(N-1) - \left(\frac{\eta - \eta^N}{1-\eta} \right) \right].$$

With $N \approx N - 1$ we have

$$\mathbb{E}[\overline{X}_N{}^2] \approx \mu_X^2 + \frac{\sigma_X^2}{N} \left(\frac{1+\eta}{1-\eta} \right) - \frac{2\sigma_X^2 \eta^2}{N^2(1-\eta)^2} (1 - \eta^{N-1}).$$

With $\eta^N \approx 0$ and a bit more algebra we have

$$\mathbb{E}[\overline{X}_N{}^2] \approx \mu_X^2 + \frac{\sigma_X^2}{N} \left(\frac{1+\eta}{1-\eta} \right) \qquad \text{and} \qquad \text{Var}(\overline{X}_N) \approx \frac{\sigma_X^2}{N} \left(\frac{1+\eta}{1-\eta} \right). \tag{6.2}$$

Recall that for the IID case ($\eta = 0$) we have $\text{Var}(\overline{X}_N) = \sigma_X^2/N$. This expression recovers that; furthermore, correlations enlarge the error bar on the sample mean.

7 Estimates of autocorrelation

Throughout this section, let X_t be a stationary Markov process with $\mathbb{E}[X_t] = \mu_X$, $\text{Var}(X_t) = \sigma_X^2$, and $\text{Corr}(X_t, X_{t+k}) = \eta^k$. (Without loss of generality, take $k \geq 0$.) The simple correlated-uniform Markov process of section 4 is one example of this; moreover, an MCMC process on a finite state space may take this form. (As described in [Berg], η is related to the second dominant eigenvalue of the transition matrix of the Markov process. If the third dominant eigenvalue is comparable with the second, then the autocorrelation will not take the simple exponential form described here.)

Remark 7.1. In the literature, one more often sees $\mathrm{Corr}(X_0, X_t) = \exp(-t/\tau_{\exp})$. Then $\tau_{\exp}$ and η are put into one-to-one correspondence by

$$\tau_{\exp} = -1/\log\eta \qquad \text{and} \qquad \eta = \exp(-1/\tau_{\exp}).$$

For the correlated-uniform process, the autocorrelation is already known; for a general MCMC process, one wishes to estimate η (or $\tau_{\exp}$) from realization data. Recall that

$$\mathrm{Corr}(X_t, X_{t+k}) = \frac{\mathbb{E}[X_t X_{t+k}] - \mathbb{E}[X_t]\mathbb{E}[X_{t+k}]}{\sigma_{X_t}\sigma_{X_{t+k}}} = \frac{\mathbb{E}[X_0 X_k] - \mu_X^2}{\sigma_X^2} \tag{7.2}$$

where the second equality holds by the stationarity of the process, and that we always have

$$-1 \le \mathrm{Corr}(X_t, X_{t+k}) \le 1. \tag{7.3}$$

(This holds for the correlation of any pair of random variables.) Also recall that

$$\sigma_X^2 = \mathbb{E}[X_t^2] - \mathbb{E}[X_t]^2. \tag{7.4}$$

Recall as well [CB] that, for M realizations $X_t^{(0)}, \ldots, X_t^{(M-1)}$ of X_t, the unbiased estimator for the variance of X_t is

$$s_{X_t}^2 = \frac{1}{M-1}\left[\sum_{i=0}^{M-1}(X_t^{(i)})^2 - \frac{1}{M}\left(\sum_{i=0}^{M-1} X_t^{(i)}\right)^2\right]. \tag{7.5}$$

Definition 7.6. Fix t and k. The *multi-realization estimator* of the autocorrelation $\mathrm{Corr}(X_t, X_{t+k})$, requiring M realizations $X_t^{(0)}, \ldots X_t^{(M-1)}$ of the process, is a straightforward combination of equations (7.2), (7.4), and (7.5). Namely,

$$\hat{c}_m(t,k) = \frac{\frac{1}{M}\sum_{i=0}^{M-1}\left(X_t^{(i)}X_{t+k}^{(i)}\right) - \frac{1}{M^2}\left(\sum_{i=0}^{M-1} X_t^{(i)}\right)\left(\sum_{j=0}^{M-1} X_{t+k}^{(j)}\right)}{\frac{1}{M-1}\left[\sum_{i=0}^{M-1}(X_t^{(i)})^2 - \frac{\left(\sum_{i=0}^{M-1} X_t^{(i)}\right)^2}{M}\right]^{1/2}\left[\sum_{j=0}^{M-1}(X_{t+k}^{(j)})^2 - \frac{\left(\sum_{j=0}^{M-1} X_{t+k}^{(j)}\right)^2}{M}\right]^{1/2}}.$$

Remark. Since the process is stationary, one may be tempted to reuse the X_t variance estimator for X_{t+k} — after all, they estimate the same quantity σ_X^2. In practice, however, doing so tends to produce autocorrelation estimates which fall (quite far) outside the range $[-1, 1]$, violating inequality 7.3. That is, the second equality in equation (7.2) holds theoretically but not at the estimator level. This same remark holds for the sliding-window estimator, to be defined next.

The difficulty with the multi-realization estimator is that realizations X_t can be expensive to compute. Rather than running M processes from $t = 0$ up to some N, which takes $O(MN)$ process-generation time, perhaps we can (carefully) use the stationarity of the process, estimating the autocorrelation using only a single realization. This will take only $O(N)$ process-generation time.

Definition 7.7. Given a single realization $X_0, \ldots, X_{N-1}$, take k from $0, 1, 2, \ldots, N-2$. The *sliding-window estimator* of the autocorrelation $\mathrm{Corr}(X_0, X_k)$, is

$$\hat{c}(k) = \frac{\frac{1}{N-k}\sum_{i=0}^{N-k-1}(X_i X_{i+k}) - \frac{1}{(N-k)^2}\left(\sum_{i=0}^{N-k-1} X_i\right)\left(\sum_{j=0}^{N-k-1} X_{j+k}\right)}{\left(\frac{1}{N-k-1}\right)\left[\sum_{i=0}^{N-k-1} X_i^2 - \frac{\left(\sum_{i=0}^{N-k-1} X_i\right)^2}{N-k}\right]^{1/2}\left[\sum_{j=0}^{N-k-1} X_{j+k}^2 - \frac{\left(\sum_{j=0}^{N-k-1} X_{j+k}\right)^2}{N-k}\right]^{1/2}}. \tag{7.8}$$

This formula is perhaps intimidating, but is made quite simple with the aid of the example below, wherein $N = 10$ and $k - 2$. Namely:

- We consider all pairs separated by k time steps: X_0X_k, X_1X_{k+1}, ..., $X_{N-k-1}X_{N-1}$. There are $N - k$ such pairs.
- The first elements in each pair form a window from X_0 to X_{N-k-1}.
- The second elements in each pair form a window from X_k to X_{N-1}.
- We estimate the mean and variance of X_0 by the sample mean and sample variance over the first window.
- We estimate the mean and variance of X_k by the sample mean and sample variance over the second window.
- We estimate the cross-moment $\mathbb{E}[X_tX_{t+k}]$ by the sample mean over pair products.

Example 7.9. $\triangleright$ There are $N = 10$ samples, X_0 through X_9:

X_0	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9

Picking $k = 2$, there are two windows of length $N - k = 8$:

X_0	X_1	X_2	X_3	X_4	X_5	X_6	X_7		
		X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9

Equation (7.8) has five distinct sums: the sum of X_0 through X_7, the sum of squares of X_0 through X_7, the sum of X_2 through X_9, the sum of squares of X_2 through X_9, and the cross sum $X_0X_2 + \ldots + X_7X_9$. $\triangleleft$

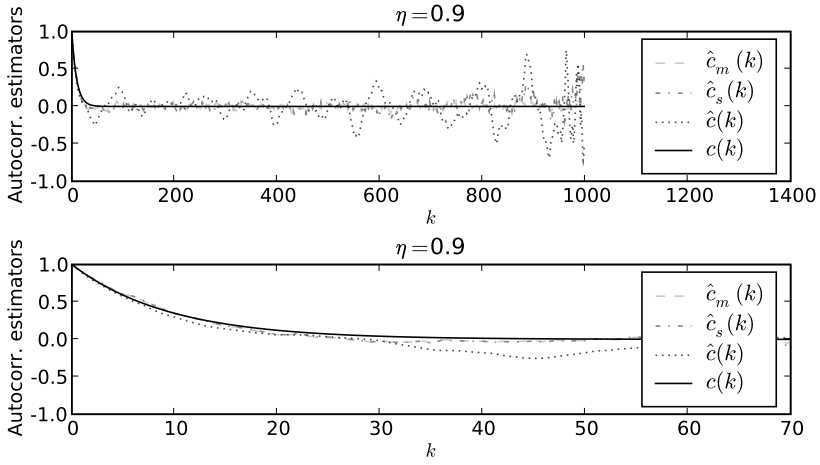

Figure 4: Autocorrelation and estimators thereof for Y_t with $\eta = 0.9$. Burn-in iterates have been discarded. The second plot zooms in on the first 50 samples of the first plot.

Remark 7.10. One would hope that $\hat{c}(t)$ is an unbiased estimator of $c(t)$. Finding its expectation using the definition is intimidating: we have a ratio of products of sums of correlated random variables. Taking an experimental approach instead, making multiple plots of the form of figure 4, one finds that $\hat{c}(t)$ does in fact fractionally underestimate $c(t)$. This affects the estimated integrated autocorrelation time, as discussed in remark 9.2.

This estimator has the benefit of making use of all the data in a single realization. Its drawback is that, for larger k, the sample size $N-k$ is small. Thus, the error in the estimator increases for larger k.

Figure 4 compares estimators against the true value $c(k) = \mathrm{Corr}(X_0, X_k) = \eta^k$ for $\eta = 0.9$. Here, $N = 1000$ time steps have been used; $M = 1000$ realizations for the multi-realization estimator $\hat{c}_m(k)$. Note that the decreasing sample size, $N-k$, of the sliding-window estimator $\hat{c}(k)$ increases the error of this estimator. For this reason, $\hat{c}_s(k)$ is also plotted. This is the same as $\hat{c}_m(k)$, but with $M = N-k$. The first plot shows the autocorrelation estimators for $k = 0$ to 998; the second zooms in on the first 50 values of k.

Remark 7.11. We observe the following:

- Comparing the full-length and short-length multi-realization estimators $\hat{c}_m(k)$ vs. $\hat{c}_s(k)$ shows that decreasing sample size does have an effect for larger k. Nonetheless, the sliding-window estimator $\hat{c}(k)$ shows markedly wilder behavior for larger k, which cannot be accounted for by small-sample-size effects alone.
- For all three estimators, errors are small when k is small, which is when the true autocorrelation $c(k) = \eta^k$ is non-negligible.
- Thus, one should examine estimators of the autocorrelation only for values of k until the estimators approach zero. Values past that point are neither accurate nor needed.

8 Integrated autocorrelation time

Following [Berg], we develop the notion of integrated autocorrelation time as follows. We reconsider the variance of the sample mean (see section 6) from a different point of view. Again, X_t is a stationary Markov process with common mean μ_X and common variance σ_X^2. We have

$$\begin{aligned}
\mathrm{Var}(\overline{X}_N) &= \mathbb{E}[(\overline{X}_N - \mu_X)^2] = \frac{1}{N^2}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\mathbb{E}[(X_i-\mu_X)(X_j-\mu_X)] \\
&= \frac{1}{N^2}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\mathbb{E}[X_iX_j - \mu_X X_i - \mu_X X_j + \mu_X^2] \\
&= \frac{1}{N^2}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\left(\mathbb{E}[X_iX_j] - \mu_X^2\right) = \frac{1}{N^2}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\mathrm{Cov}(X_i, X_j) \\
&= \frac{1}{N^2}\left[\sum_{i=0}^{N-1}\mathrm{Var}(X_i) + 2\sum_{t=1}^{N-1}(N-t)\,\mathrm{Cov}(X_0, X_t)\right] \\
&= \frac{\sigma_X^2}{N} + 2\sigma_X^2\sum_{t=1}^{N-1}(N-t)\,\mathrm{Corr}(X_0, X_t) \\
&= \frac{\sigma_X^2}{N}\left[1 + 2\sum_{t=1}^{N-1}\left(1-\frac{t}{N}\right)\,\mathrm{Corr}(X_0, X_t)\right] \approx \frac{\sigma_X^2}{N}\left[1 + 2\sum_{t=1}^{\infty}\mathrm{Corr}(X_0, X_t)\right].
\end{aligned}$$

If X_t is IID then we recover the familiar $\mathrm{Var}(\overline{X}_N) = \sigma_X^2/N$; otherwise we have

$$\mathrm{Var}(\overline{X}_N) = \frac{\sigma_X^2}{N}\,\tau_{\mathrm{int}} \tag{8.1}$$

where τ_{int} is the last bracketed expression above. Note as well that if $\text{Corr}(X_0, X_k) = \eta^k$, then

$$\tau_{\text{int}} = 1 + 2\sum_{t=1}^{\infty} \eta^t = 1 + \frac{2\eta}{1-\eta} = \frac{1+\eta}{1-\eta} \tag{8.2}$$

which is what we would have expected by comparing equations (6.2) and (8.1). As a consequence, when $c(t) = \eta^t$ we have

$$\tau_{\text{int}} = \frac{1+\eta}{1-\eta} \qquad \text{and} \qquad \eta = \frac{\tau_{\text{int}} - 1}{\tau_{\text{int}} + 1}. \tag{8.3}$$

Some values are shown for reference in table 1.

η	0	0.1	0.2	0.5	0.6	0.9	0.990	0.999
$(1+\eta)/(1-\eta)$	1	1.222	1.500	3.000	4.000	19	199	1999

Table 1: η vs. $(1+\eta)/(1-\eta)$.

Remark 8.4. If the process is IID, i.e. $\eta = 0$, then $c(0) = 1$, $c(t) = 0$ for all $t \geq 1$, and $\tau_{\text{int}} = 1$.

Definition 8.5. Recall that $s_N^2(X_t)$ (equation (7.5)) estimates σ_X^2. Using equation (8.1), the *naive estimator* and *corrected estimator* of $\text{Var}(\overline{X}_N)$ are

$$t_N^2(X_t) = \frac{s_N^2(X_t)}{N} \qquad \text{and} \qquad u_N^2(X_t) = \frac{s_N^2(X_t)}{N}\,\hat{\tau}_{\text{int}}, \tag{8.6}$$

as long as we have an estimator $\hat{\tau}_{\text{int}}$ of τ_{int}.

9 Estimation of the integrated autocorrelation time

Recall from remark 7.11 that $\hat{c}(t)$ is a rather wild estimator of $c(t)$ at high t. Since

$$\hat{\tau}_{\text{int}} = 1 + 2\sum_{t=1}^{\infty} \hat{c}(t)$$

is nothing more than a sum of $c(t)$, we can expect it to be ill-behaved as well.

Definition 9.1. The *running-sum estimator* of τ_{int} is

$$\hat{\tau}_{\text{int}}(t) = 1 + 2\sum_{k=1}^{t} \hat{c}(k).$$

The idea is to accumulate the reliable low-t values of $\hat{c}(t)$ until the sum becomes approximately constant at some s, then stop and declare $\hat{\tau}_{\text{int}}$ to be $\hat{\tau}_{\text{int}}(s)$. This is the *flat-spot estimator* or *turning-point estimator* for τ_{int}. See figure 5 for illustration, where s is approximately 24 for the blue realization and 29 for the red. From the plots, we estimate $\tau_{\text{int}} \approx 15$; using equation (8.3), we estimate $\eta = (15-1)/(15+1) = 0.875$. This is reasonable since the data were obtained with $\eta = 0.9$, for which the true τ_{int} is 19 by equation (8.2).

It is clear from the figure that estimators $\hat{\tau}_{\text{int}}$ can vary noticeably from one realization to the next. Our estimator for the variance of the sample mean, i.e. the error bar on the sample mean, is $u_N^2(X_t)$ (equation

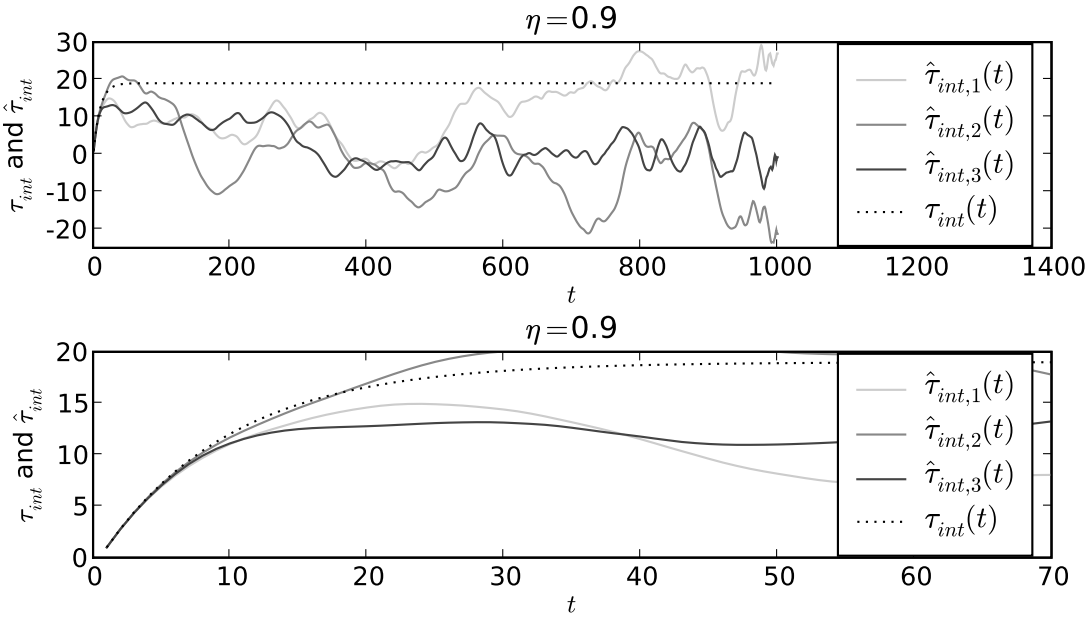

Figure 5: Estimated and exact integrated autocorrelation times for Y_t with $\eta = 0.9$, using three realizations similar to the one in figure 4. Burn-in iterates have been discarded. The second plot zooms in on the first 50 samples of the first plot. The flat-spot estimator $\hat{\tau}_{\text{int}}$ of τ_{int} is found by reading off the vertical coordinate of the first turning point of each solid-line plot; the true τ_{int} is the horizontal asymptote of the dotted-line plot. Two of the three turning points yield a $\hat{\tau}_{\text{int}}$ which is less than the true τ_{int}. This is the general case: we find that $\hat{\tau}_{\text{int}}$ underestimates more often than it overestimates. See also figure 8 on page 15.

(8.6)). Since $\hat{\tau}_{\text{int}}$ is a factor in $u_N^2(X_t)$, variations in $\hat{\tau}_{\text{int}}$ will result in error of the error bar. Figure 6 shows that variations in $\hat{\tau}_{\text{int}}$ increase with η.

At present I know of no solution to this problem other than the running of multiple experiments — larger M, using the notation of section 1. As long as τ_{int} is estimated based on a single experimental result $X_0, \ldots, X_{N-1}$, one must be aware that the error bars on the sample mean are crude.

Remark 9.2. As was noted in remark 7.10, $\hat{c}(t)$ underestimates $c(t)$. Since $\hat{\tau}_{\text{int}}$ is formed from a sum of $\hat{c}(t)$'s, $\hat{\tau}_{\text{int}}$ is also a fractional underestimator of τ_{int}, as will be seen in section 10.

10 Estimation of the variance of the sample mean

Given the flat-spot estimator $\hat{\tau}_{\text{int}}$ of τ_{int} from section 9 and the naive estimator of the variance of the sample mean from equation (8.6), we may now compute the corrected estimator of the the variance of the sample mean:

$$u_N^2(X_t) = \frac{s_N^2(X_t)}{N}\, \hat{\tau}_{\text{int}}.$$

We use the correlated-uniform Markov process to illustrate, since for this process all quantities have known theoretical values. As in section 1, we display standard deviations in our plots and tables, rather than variances: the units of measurement of the former match those of the mean, and they correspond visually to variations in the data.

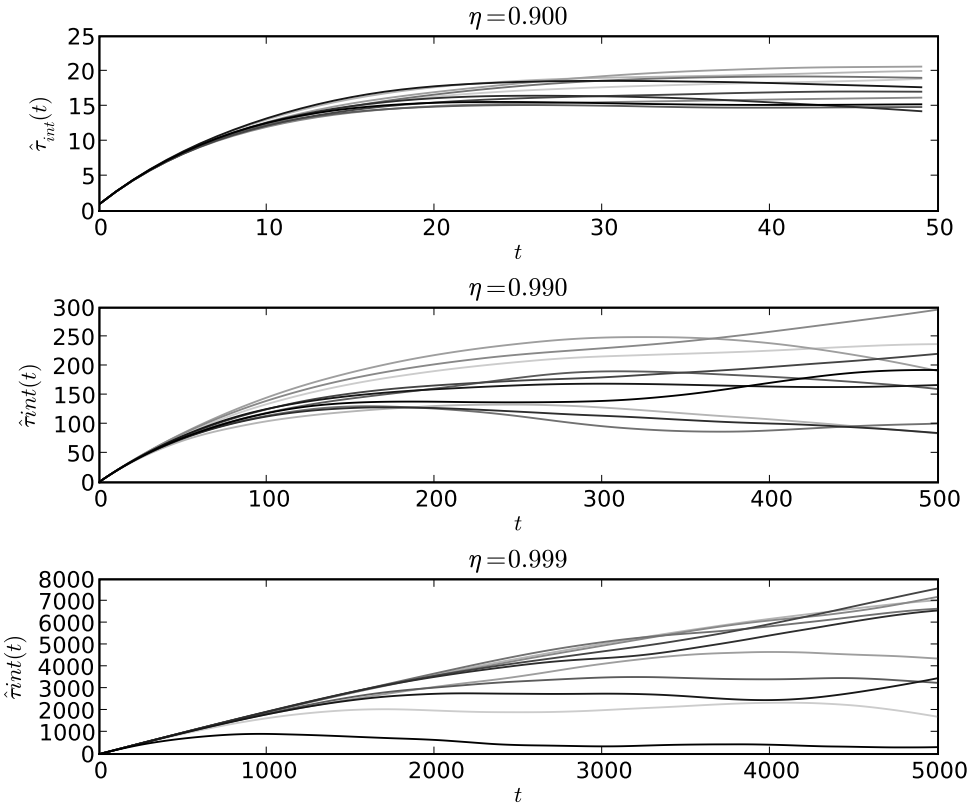

Figure 6: Estimated integrated autocorrelation times for Y_t with $\eta = 0.9, 0.99, 0.999$, using ten realizations each. N is 100,000; burn-in iterates have been discarded. Recall that true τ_{int} values are 19,199, and 1999, respectively. The variation in the vertical coordinate of the first flat spot in each plot, which increases with η, gives rise to the error of the error bar on the sample mean.

- The mean and variance of Y_t are $\mu_Y = 1/2$ and $\sigma_Y^2 = 1/12$; $\sigma_Y \approx 0.289$. Using $\eta = 0.0, 0.9, 0.999$, the true τ_{int} is $1, 19, 1999$, respectively. We conduct $M = 100$ experiments of collecting and analyzing $N = 10000$ time-series samples $Y_0, \ldots, Y_{N-1}$.
- The true mean is shown in row 1 of table 2. Estimators $\overline{Y}_N$ are shown in figure 7. The average of these over all M experiments is shown in row 2 of table 2.
- The true naive variance of the sample mean is σ_Y^2/N, with true naive standard deviation of the sample mean $\sigma_Y/\sqrt{N} \approx 0.00289$. The true corrected variance of the sample mean is $\sigma^2_{\overline{Y}_N} = \tau_{\text{int}}\,\sigma_Y^2/N = 1/120000, 19/120000, 1999/120000$. The true standard deviations of the sample means are then $\sigma_{\overline{Y}_N} \approx 0.0028868, 0.0125831, 0.1290672$. These are shown in row 3 of table 2.
- The multi-experiment estimator $s_M(\overline{Y}_N)$ of $\sigma_{\overline{Y}_N}$ is the sample standard deviation of the M values $\overline{X}_N{}^{(0)}, \ldots, \overline{X}_N{}^{(M-1)}$. These estimators are shown in row 4 of table 2. As expected, the multi-experiment estimator is a good one.
- Next we turn to single-experiment estimators of the variance of the sample mean. The estimated naive standard deviation of the sample mean is $t_N(Y_t) = s_N(Y_t)/\sqrt{N}$. These are not plotted for each experiment; their average over all M experiments is shown in row 5 of table 2. Note that they match the true variance of the sample mean only in the IID ($\eta = 0$) case.

- True values of τ_{int} for each η are shown in row 8 of the table. The flat-spot estimators $\hat{\tau}_{\text{int}}$ for all $M = 100$ experiments are shown in figure 8. Their average and sample standard deviation over all M experiments are shown in rows 9 and 10. As discussed in remark 9.2, we see that $\hat{\tau}_{\text{int}}$ fractionally underestimates τ_{int}.
- Using the $\hat{\tau}_{\text{int}}$ values, the corrected estimators $u_N(Y_t) = t_N(Y_t)\sqrt{\hat{\tau}_{\text{int}}}$ are shown, for all $M = 100$ experiments, in figure 9. Their averages over all M experiments are shown in row 6 of table 2. (Again, the corresponding true values are in row 2 of the table.) The fractional underestimation of $\hat{\tau}_{\text{int}}$ carries over to $u_N(Y_t)$. One trades the quality of the estimator for the feasibility of its computation.
- Standard deviations over M experiments of $u_N(Y_t)$ are shown in row 7 of the table. Figure 10 shows, for $\eta = 0.999$, the $M = 100$ values of $\overline{Y}_N$ along with their respective $u_N(Y_t)$'s. These show the error of the error bar.

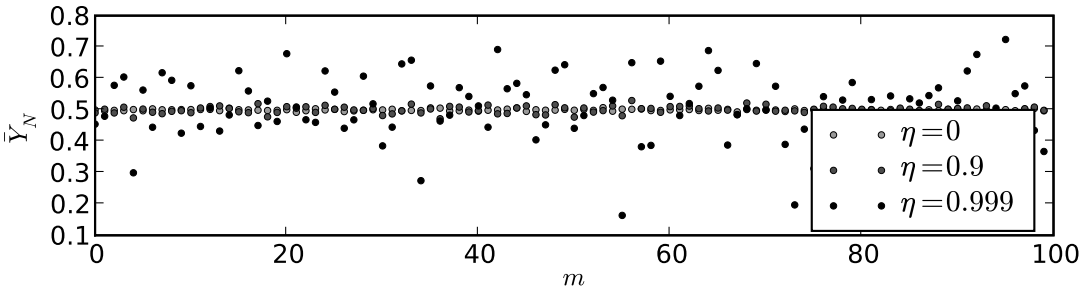

Figure 7: $\overline{Y}_N$ over $M = 100$ experiments, where the true value is $\mu_X = 0.5$. Variance of $\overline{Y}_N$ increases with autocorrelation factor η.

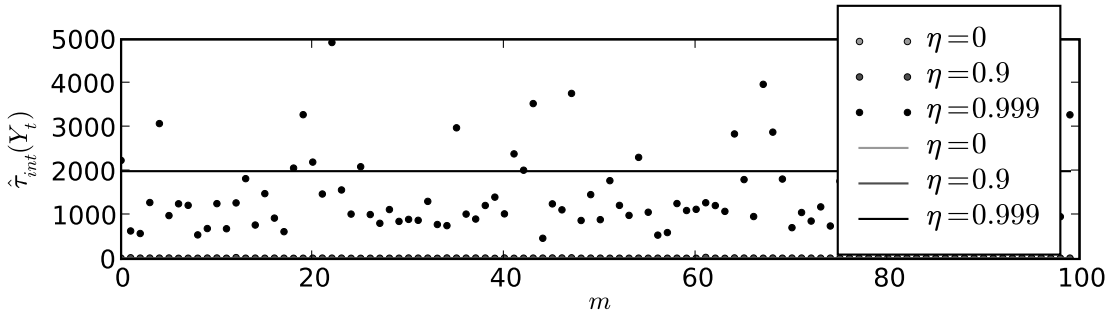

Figure 8: $\hat{\tau}_{\text{int}}(Y_t)$ over $M = 100$ experiments, along with true values. Note that $\hat{\tau}_{\text{int}}(Y_t)$ fractionally underestimates the true $\tau_{\text{int}}(Y_t)$.

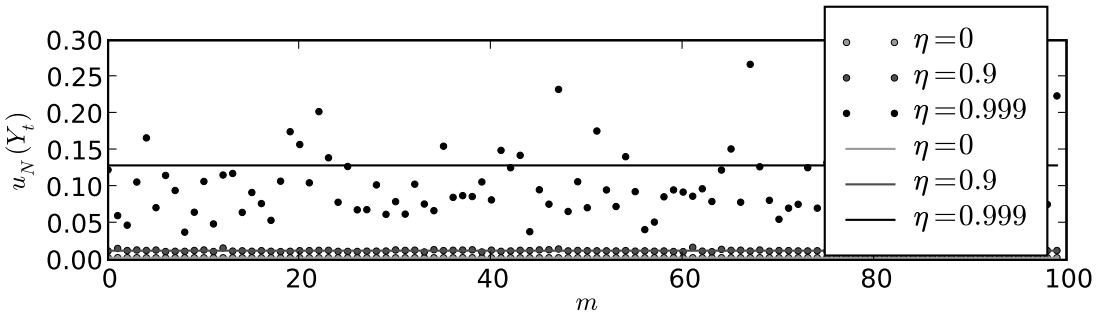

Figure 9: $u_N(Y_t)$ over $M = 100$ experiments, along with true values. Note that $u_N(Y_t)$ fractionally underestimates the true standard deviation of the sample mean, $\sigma_{\overline{Y}_N} = \sigma_Y/\sqrt{N}$.

Description	$\boldsymbol{\eta}$	**0**	**0.9**	**0.999**
1. True mean	$\mu_{\overline{Y}_N}$	0.50000	0.50000	0.50000
2. Sample mean	$m_M(\overline{Y}_N)$	0.49948	0.49952	0.51100
	$m_M(\overline{Y}_N)$	0.49987	0.50231	0.47341
	$m_M(\overline{Y}_N)$	0.49991	0.49895	0.47958
3. True standard deviation of sample mean	$\sigma_{\overline{Y}_N}$	0.00288	0.01258	0.12906
4. Multi-experiment estimator of $\sigma_{\overline{Y}_N}$	$s_M(\overline{Y}_N)$	0.00274	0.01166	0.10342
	$s_M(\overline{Y}_N)$	0.00274	0.01167	0.12303
	$s_M(\overline{Y}_N)$	0.00298	0.00986	0.11929
5. Averaged single-experiment naive estimators of $\sigma_{\overline{Y}_N}$	$m_M(t_N(Y_t))$	0.00288	0.00287	0.00263
	$m_M(t_N(Y_t))$	0.00288	0.00288	0.00260
	$m_M(t_N(Y_t))$	0.00288	0.00287	0.00250
6. Averaged single-experiment corrected estimators of $\sigma_{\overline{Y}_N}$	$m_M(u_N(Y_t))$	0.00288	0.01279	0.09957
	$m_M(u_N(Y_t))$	0.00289	0.01280	0.10037
	$m_M(u_N(Y_t))$	0.00289	0.01276	0.08947
7. Sample standard deviation of corrected estimators of $\sigma_{\overline{Y}_N}$	$s_M(u_N(Y_t))$	0.00004	0.00105	0.04402
	$s_M(u_N(Y_t))$	0.00005	0.00108	0.04665
	$s_M(u_N(Y_t))$	0.00006	0.00118	0.03851
8. True integrated autocorrelation time	τ_{int}	1	19	1999
9. Averages of estimated integrated autocorrelation time	$m_M(\hat{\tau}_{\text{int}})$	0.999	19.854	1442.627
	$m_M(\hat{\tau}_{\text{int}})$	1.002	19.763	1500.137
	$m_M(\hat{\tau}_{\text{int}})$	1.008	19.857	1279.173
10. Standard deviation across M experiments of $\hat{\tau}_{\text{int}}$	$s_M(\hat{\tau}_{\text{int}})$	0.028	3.162	865.992
	$s_M(\hat{\tau}_{\text{int}})$	0.031	3.092	1045.842
	$s_M(\hat{\tau}_{\text{int}})$	0.039	3.628	758.903

Table 2: Statistics for three trials of $M = 100$ experiments on $N = 10000$ samples of Y_t: $\eta = 0.0, 0.9, 0.999$.

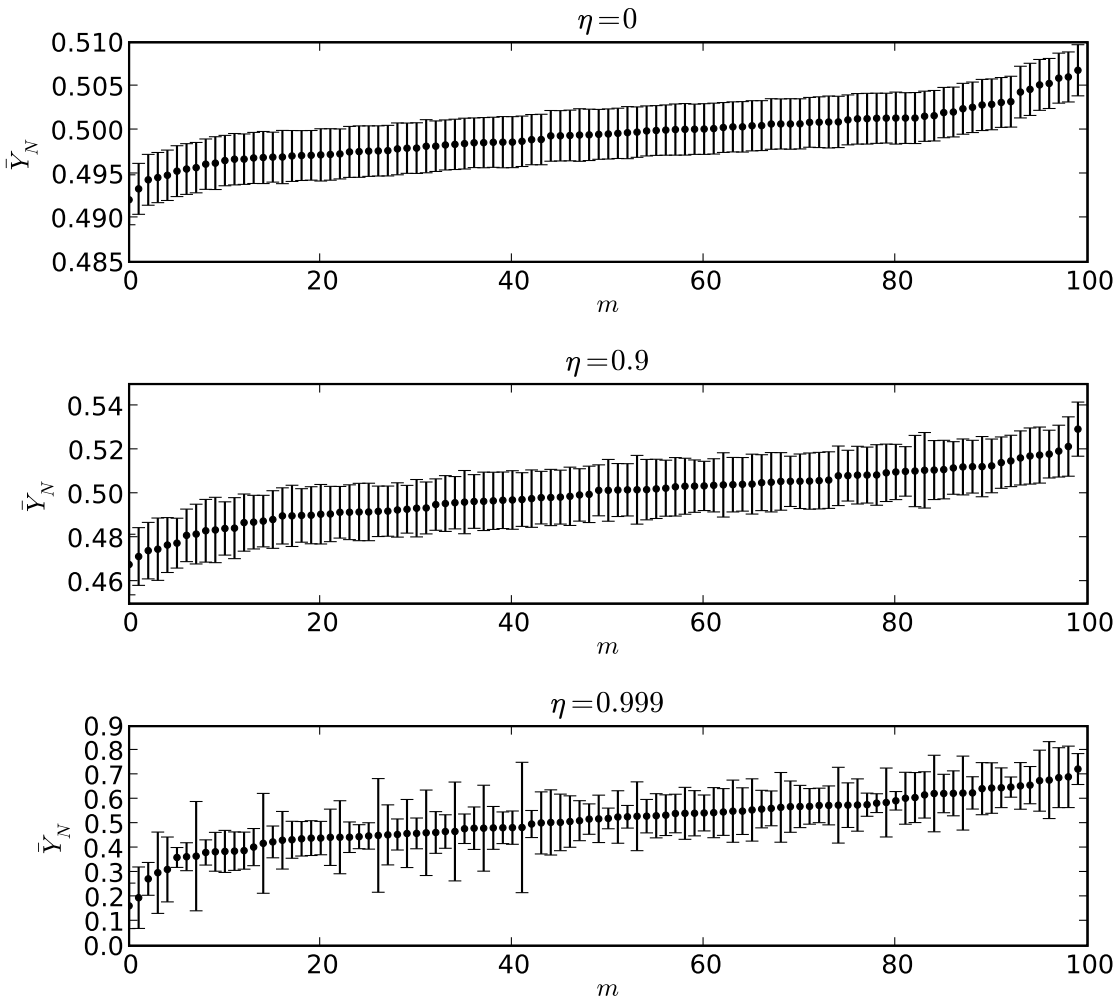

Figure 10: $\overline{Y}_N$ with single-sigma error bars, $\eta = 0, 0.9, 0.999$, $M = 100$ experiments, sorted by increasing $\overline{Y}_N$. The magnitude and the variation of the error bars both increase with η.

11 Integrated and exponential autocorrelation times

In remark 7.1 of section 7, we noted that if $\mathrm{Corr}(X_0, X_t) = \eta^t$ for $\eta \in [0, 1)$, then we may define an exponential autocorrelation time via

$$\tau_{\mathrm{exp}} = -1/\log \eta$$

such that $\mathrm{Corr}(X_0, X_t) = \exp(-t/\tau_{\mathrm{exp}})$. Yet section 8 gave us something similar: the integrated autocorrelation time τ_{int}. In particular, if $\mathrm{Corr}(X_0, X_t) = \eta^t$, then we had

$$\tau_{\mathrm{int}} = \frac{1+\eta}{1-\eta}.$$

Figure 11 compares these two.

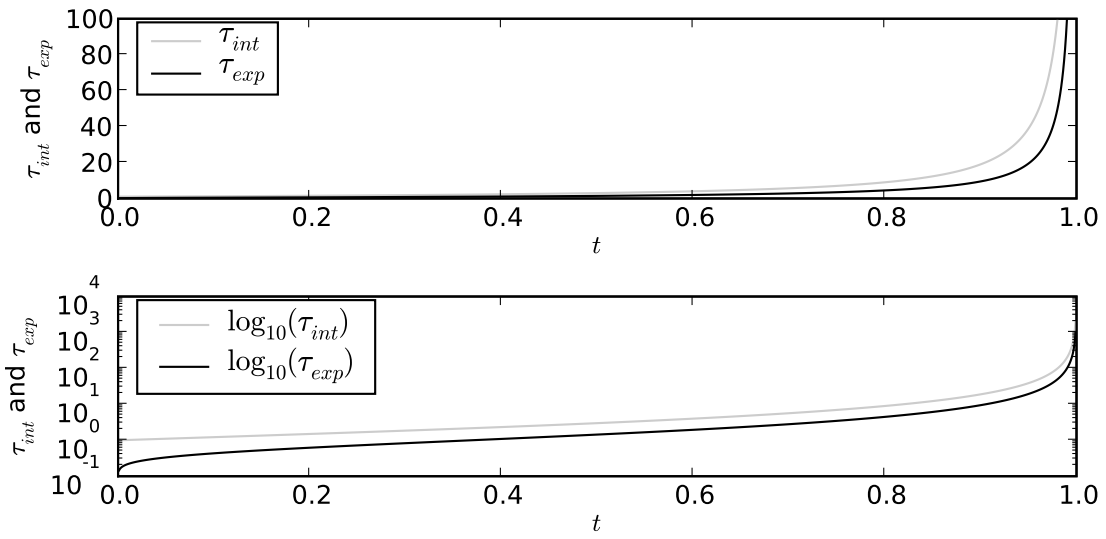

Figure 11: Integrated and exponential autocorrelation times as a function of η.

12 Batched means

Introductory statistics tends to deal with the analysis of IID samples. Yet, realization sequences from an MCMC experiment tend to be highly correlated. The sample mean estimates the true mean, since expectation is linear. But when one wishes to place an accurate error bar on the sample mean, correlations must be taken into account.

One approach (see for example [Berg], who calls this process *binning*) is to subdivide $X_0, \ldots, X_{N-1}$ into $K = N/B$ batches of size B. The K sample means over batches may be treated as IID samples. The independence of the K samples means that the variance of their sample mean will be reduced, but reducing the sample size from N to K will increase the variance. We will show that these two effects cancel: binning N samples down to K samples does not change the variance of the sample mean. (As shown in [Berg], batched means have their uses: they may be used to construct a method to estimate τ_{int}, as an alternative to the method of section 9.)

Definition 12.1. Given $X_0, \ldots, X_{N-1}$ with common mean μ_X and variance σ_X^2, let B divide N and $K = N/B$. Then B is the *batch size* and K is the *number of batches*. For $k = 0, \ldots, K-1$, the kth batch consists of $X_{kB}, \ldots, X_{(k+1)B-1}$. The sample mean of the kth batch is

$$A_k = \frac{1}{B} \sum_{i=0}^{B-1} X_{kB+i}.$$

We now consider the sequence $A_0, \ldots, A_{K-1}$. We define the *batched mean* to be

$$\overline{X}_{N,B} = \frac{1}{K} \sum_{k=0}^{K-1} A_k.$$

By linearity of expectation, we immediately have $\mathbb{E}[\overline{X}_{N,B}] = \mu_X$. We next inquire about the variance of the batched mean, then compare that to the variance of the (non-batched) sample mean.

13 Variance and covariance of batches

To compute $\mathrm{Var}(A_k)$ and $\mathrm{Corr}(A_0, A_k)$, we first need $\mathbb{E}[A_k A_\ell]$ for $k = \ell$ and $k \neq \ell$. In the $k = \ell$ case, the computation is the same as in section 6, with B playing the role of N. We have

$$\mathbb{E}[A_k^2] \approx \mu_X^2 + \frac{\sigma_X^2}{B}\left(\frac{1+\eta}{1-\eta}\right) \qquad \text{and} \qquad \mathrm{Var}(A_k) \approx \frac{\sigma_X^2}{B}\left(\frac{1+\eta}{1-\eta}\right).$$

For $k \neq \ell$, without loss of generality assume $k < \ell$. Using equation (6.1), we have

$$\begin{aligned}\mathbb{E}[A_k A_\ell] &= \frac{1}{B^2}\sum_{i=0}^{B-1}\sum_{j=0}^{B-1}\mathbb{E}[X_{kB+i}X_{\ell B+j}] &&= \frac{1}{B^2}\sum_{i=0}^{B-1}\sum_{j=0}^{B-1}(\mu_X^2 + \sigma_X^2\eta^{\ell B+j-kB-i}) \\ &= \mu_X^2 + \frac{\sigma_X^2\eta^{(\ell-k)B}}{B^2}\sum_{i=0}^{B-1}\eta^{-i}\sum_{j=0}^{B-1}\eta^j &&= \mu_X^2 + \frac{\sigma_X^2\eta^{(\ell-k)B}}{B^2}\left(\frac{1-\eta^B}{1-\eta}\right)^2.\end{aligned}$$

If the batch size is chosen so that η^B is negligible, then

$$\mathbb{E}[A_k A_\ell] = \mu_X^2.$$

Now we have (for $\eta^B \approx 0$)

$$\mathrm{Var}(A_k) \approx \frac{\sigma_X^2}{B}\left(\frac{1+\eta}{1-\eta}\right) \qquad \text{and} \qquad \mathrm{Corr}(A_k, A_\ell) = \frac{\mathbb{E}[A_k A_\ell] - \mu_X^2}{\sigma_{A_0}^2} = \delta_{k,\ell}. \tag{13.1}$$

This justifies the hope that batches can be constructed to form an IID sequence.

14 Variance of the batched mean

We now find out what effect batching has on the variance of the sample mean: batching produces an IID sequence, which will reduce the variance (equation (8.1)), yet it reduces the sample size from N down to $K = N/B$, which by central-limit reasoning should increase the variance.

For the non-batched mean, we have the random variables $X_0, \ldots, X_{N-1}$; parameters are mean μ_X, variance σ_X^2, autocorrelation η^k, and (from equation (8.3)) integrated autocorrelation time $\tau_{\mathrm{int}} = (1+\eta)/(1-\eta)$. Equation (6.2) gives

$$\mathrm{Var}(\overline{X}_N) = \frac{\sigma_X^2}{N}\left(\frac{1+\eta}{1-\eta}\right). \tag{14.1}$$

For the batched mean, we batch $X_0, \ldots, X_{N-1}$ into K IID batches of size B. We have the random variables $A_0, \ldots, A_{K-1}$, with mean μ_X, variance $(\sigma_X^2/B)(1+\eta)/(1-\eta)$ (equation (13.1)), autocorrelation $c(k) = \delta_{0,k}$ (since A_k is IID) and integrated autocorrelation time $\tau_{\mathrm{int}} = 1$ (remark 8.4). Then

$$\mathrm{Var}(\overline{X}_{N,B}) = \frac{\sigma_X^2}{KB}\left(\frac{1+\eta}{1-\eta}\right) = \frac{\sigma_X^2}{N}\left(\frac{1+\eta}{1-\eta}\right).$$

Thus, to first order in η and N, as long as B is large enough that η^B is negligibly small, we do not expect batching to change the variance of the sample mean.

Table 3 shows some sample results of these calculations for the correlated-uniform Markov process Y_t. There are $M = 100$ experiments of $N = 10000$ samples. Each experiment was analyzed as-is ($B = 1$), as well as with batch size $B = 64$, 512, and 4096. (Recall from table 1 that $\eta = 0, 0.9, 0.999$ correspond to $\tau_{\text{int}} = 1, 19, 1999$, respectively.) Now the process being analyzed is $A_0, \ldots, A_{K-1}$ where $K = N/B$. We note the following:

- For $B = 1$ (the original time series), the estimator u^2 of the variance of the sample mean employed was the corrected estimator of equation (8.6), while for $B = 64, 512, 4096$, the A_k's were treated as if they were IID. That is, for $B > 1$ we set $u^2 = t^2$.
- Batch size does not, of course, affect the sample mean (column 3 of the table). Likewise, it does not affect the multi-experiment estimator of the variance of the sample mean (column 4).
- The true variance of the sample mean, $\sigma_{\overline{A}_N} = \sqrt{\tau_{\text{int}} \sigma^2_{A_k}/K}$, is shown in column 5.
- The last two columns show the first two autocorrelation estimates. These show that for $\eta = 0$ (the IID case), Y_t samples are indeed approximately IID. For $\eta = 0.9$, the $B = 64$ batches are nearly independent, and the largest batches are quite weakly correlated. For $\eta = 0.999$, where $\tau_{\text{int}} = 1999$, batch sizes of 64 and 512 are too small, but batch size 4096 is large enough to produce weakly correlated batches.
- For $\eta = 0$, the average of the single-experiment estimator u of the variance of the sample mean is approximately constant with respect to batch size. For $\eta = 0.9$ and $\eta = 0.999$, once the batch size is large enough to get weakly correlated samples, the estimator u^2 on batches agrees with the estimator u^2 on the original time-series data.
- The multi-realization estimator and the averaged single-realization estimator of the variance of the sample mean (columns 4 and 6) roughly agree, for batch sizes large enough that batches are weakly correlated.
- The error of the error bar (column 7 of the table) is not improved by use of batched means.

η	B	$m_M(\overline{A}_N)$	$s_M(\overline{A}_N)$	$\sigma_{\overline{A}_N}$	$m_M(u_N(A_k))$	$s_M(u_N(A_k))$	$\hat{c}_{A_k}(0)$	$\hat{c}_{A_k}(1)$
0.000	1	0.5001	0.00113	0.00113	0.00113	0.000002	1.0000	0.0001
0.000	64	0.5001	0.00113	0.00113	0.00113	0.000027	0.9990	0.0378
0.000	512	0.5001	0.00113	0.00113	0.00114	0.000073	0.9922	0.0031
0.000	4096	0.5001	0.00113	0.00113	0.00111	0.000206	0.9375	0.2053
0.900	1	0.4996	0.00481	0.00492	0.00490	0.000034	1.0000	0.8982
0.900	64	0.4996	0.00481	0.00492	0.00451	0.000103	0.9990	0.1591
0.900	512	0.4996	0.00481	0.00492	0.00480	0.000280	0.9922	-0.0552
0.900	4096	0.4996	0.00481	0.00492	0.00476	0.000894	0.9375	0.1445
0.999	1	0.5112	0.04801	0.05042	0.04937	0.004180	1.0000	0.9987
0.999	64	0.5112	0.04801	0.05042	0.00874	0.000756	0.9990	0.9493
0.999	512	0.5112	0.04801	0.05042	0.02301	0.002276	0.9922	0.6462
0.999	4096	0.5112	0.04801	0.05042	0.04280	0.007354	0.9375	-0.0060

Table 3: Statistics for $M = 100$ batched experiments on $N = 65536$ samples of Y_t: $\eta = 0.0, 0.9, 0.999$.

15 Conclusions on error bars, autocorrelation, and batched means

Given an MCMC experimental result $X_0, \ldots, X_{N-1}$, we may compute the sample mean $\overline{X}_N$ and an estimator $s^2_N(X_t)$ of the sample variance.

Batched means improve neither the bias nor the variation of the error bar. The variance reduction obtained by (approximate) independence of batches cancels out the variance increase caused by reduced sample size.

Computing autocorrelations and summing them as described in section 9, we may obtain an estimate $\hat{\tau}_{\text{int}}$ of the integrated autocorrelation time τ_{int}. This is used to update the naive estimated variance of the sample mean $t_N^2(X_t) = s^2/N$ to the corrected estimator $u_N^2(X_t) = \hat{\tau}_{\text{int}} s^2/N$. With the understanding that $\hat{\tau}_{\text{int}}$ has itself a noticeable variance and a fractional underbias, u_N estimates the standard deviation of the sample mean.

References

[Berg] Berg, B. *Markov Chain Monte Carlo Simulations and Their Statistical Analysis.* World Scientific Publishing, 2004.

[CB] Casella, G. and Berger, R.L. *Statistical Inference* (2nd ed.). Duxbury Press, 2001.

[GS] Grimmett, G. and Stirzaker, D. *Probability and Random Processes*, 3rd ed. Oxford, 2001.

Index