Math 577 Monte Carlo Methods – Spring 2019
Comments on Lectures 12 (continued)

Kevin K. Lin*

February 22, 2019

Main topics:

- Estimating $\tau_{\exp}$ and τ_{int} from data
- Estimating autocorrelation functions

I am largely following [S]. See also the Wikipedia pages on estimating autocorrelations and on periodograms.

Standard estimator for autocorrelation functions. We assume the same set-up as in the previous notes. Given a sequence of measurements $\varphi_1, \cdots, \varphi_N$, the standard estimator for autocorrelation functions is

$$\widehat{C}_\varphi(n) = \frac{1}{N-|n|}\sum_{k=1}^{N-|n|}(\varphi_k - \widehat{\varphi}_N)\cdot(\varphi_{k+|n|} - \widehat{\varphi}_N). \tag{1}$$

Notice that when $n = 0$, this is just the sample variance:

$$\widehat{C}_\varphi(0) = \frac{1}{N}\sum_{k=1}^{N}(\varphi_k - \widehat{\varphi}_N)^2. \tag{2}$$

Similarly, for $n > 0$, this is just the sample covariance between X_k and $X_{k+|n|}$. For fixed n, the above estimator clearly converges to $C_\varphi(n)$ as $N \to \infty$, i.e., it is consistent.

One of the problems with the standard estimator is that while we know $C_\varphi(n) \to 0$ as $n \to \pm\infty$, the estimator actually gets worse as n increases. The reason is that there are fewer and fewer samples as one heads toward the tail of the autocorrelation function. The "double

*E-mail: `klin@math.arizona.edu`.

whammy" of small $C_\varphi(n)$ and small number of samples means the *relative* error is terrible. The biased estimator (the "periodogram" estimator)

$$\widehat{C}_\varphi^{\mathrm{per}}(n) = \frac{1}{N} \sum_{k=1}^{N-|n|} (\varphi_k - \widehat{\varphi}_N) \cdot (\varphi_{k+|n|} - \widehat{\varphi}_N) \tag{3}$$

tends to be more accurate in the tails. The reason is that

$$\widehat{C}_\varphi^{\mathrm{per}}(n) = \left(1 - \frac{|n|}{N}\right) \cdot \widehat{C}_\varphi(n). \tag{4}$$

The multiplicative factor forces the tails of the estimator to go to 0. As $N \to \infty$, the factor tends to 1 (for fixed n), so it is once again consistent.

The naive way of implementing these estimators involves $O(N^2)$ operations, which is relatively slow. The sums can be re-expressed as a convolution, which in turn can be implemented using the fast Fourier transform in $O(N \log N)$ time. In Matlab and in many other environments, this is as simple as calling the `conv()` function. See the accompanying Jupyter notebooks for examples.

Estimating τ_{exp}. To estimate τ_{exp}, a relatively simple procedure is to:

1) Estimate the autocorrelation function using, e.g., the estimator $\widehat{C}_\varphi^{\mathrm{per}}(n)$.

2) Make a log-linear plot and check that the estimated autocorrelation function does indeed decay exponentially.

3) By fitting a line to the linear part of the estimated autocorrelation function, estimate τ_{exp}.

It is important that the estimated autocorrelation function does exhibit exponential decay over 1 to 2 *decades*. Otherwise, something is wrong: there may not be enough data, or a bug in the code.

Note that estimating the autocorrelation function requires that we have stationary data to start with. So, there's a chicken-and-egg problem. One way is to proceed iteratively: first guess a burn-in period N_{init} (based perhaps by simply plotting φ_n and guessing a big enough cutoff), then estimate τ_{exp}. One then compares the initial N_{init} with τ_{exp}. If it is not a sufficiently large multiple (say $10\times$) of τ_{exp}, then increase N_{init} and try again.

Estimating τ_{int}. To estimate the integrated autocorrelation time, one does *not* simply sum up the estimated autocorrelation function. The reason is that most of the tail has large relative error (no matter which estimator is used), and there are *lots* of terms in the tail. A simple procedure is to introduce a cutoff M, and to use the estimator

$$\widehat{\tau}_{\mathrm{int}}(M) = \sum_{n=-M}^{M} \widehat{C}_\varphi^{\mathrm{per}}(n). \tag{5}$$

The qusetion then becomes: how do we pick M? [S] proposes the following heuristic: since samples become effectively independent after τ_{exp} steps, we expect $C_\varphi(n)$ to be effectively 0 for $n \geqslant c\tau_{int}$ for some $c > 0$, say $c = 6$ or $c = 10$. So, we look for the smallest cutoff M such that $M \geqslant c\widehat{\tau}_{int}(M)$. This can be done by plotting M/c and $\widehat{\tau}_{int}(M)$ and looing for the smallest intersection of the two curves.

Alternatives to estimating τ_{exp} and τ_{int}. The above are not the only methods for reducing initialization bias and approximating estimator variance.

1) As an alternative to estimating τ_{exp}, one can assess initialization bias by applying a test for stationarity. A number of such tests have been developed by statisticians.

2) As a "cheap" alternative to the Kubo formula, one can use the "method of batch means" to estimate variance. This is described in [AG], and will be discussed in the next lecture.