Activity: Properties of the Sample Mean

Suppose we would like to estimate the mean μ of our process $\{X_t\}$ using some data $X_1, \ldots, X_n$. We would like to know to what extent the mean of a sample,

$$\bar{X} = \frac{1}{n} \sum_{t=1}^{n} X_t$$

constitutes a useful estimator of μ. There are several theorems showing how this estimator behaves asymptotically as $n \to \infty$, but these do not tell us the variance of our estimator for a finite amount of data. For i.i.d. data of size n from a random variable X with variance σ_X^2 we know

$$\operatorname{Var}\left(\bar{X}\right) = \frac{\sigma_X^2}{n}.$$

We start by deriving the variance of the sample mean when the data are correlated.

1. Consider the case of the variance of the sum of three random variables X_1, X_2 and X_3. Show that

$$\operatorname{Var}\left(X_1 + X_2 + X_3\right) = \sum_{i=1}^{3} \operatorname{Var}(X_i) + 2 \sum_{i,j\,:\,i<j} \operatorname{Cov}(X_i, X_j).$$

2. What is the general result for the variance of $X_1 + X_2 + \cdots + X_n$?

3. Using your result from above, show that

$$\operatorname{Var}\left(\bar{X}\right) = \frac{\sigma_X^2}{n} \left(1 + 2 \sum_{k=1}^{n-1} \left(1 - \frac{k}{n}\right) \rho\left(k\right)\right),$$

where the data are from a stationary process with variance σ_X^2 and acf $\rho\left(\cdot\right)$. (Hint: Try the case $n = 3$ first if you have difficulties.)

4. Now consider the stochastic process

$$X_t = \frac{1}{2} X_{t-1} + Z_t$$

where $\{Z_t\}$ is white noise as usual. Using the result from above, show that for this process the variance of the mean of a sample $X_1, \ldots, X_n$ from $\{X_t\}$ would be

$$\mathrm{Var}\left(\bar{X}\right) = \frac{\sigma_X^2}{n}\left(1 + 2\left(\sum_{k=1}^{n-1}\left(\frac{1}{2}\right)^k - \frac{1}{n}\sum_{k=1}^{n-1} k\left(\frac{1}{2}\right)^k\right)\right).$$

5. What happens to

$$\frac{1}{n}\sum_{k=1}^{n-1} k\left(\frac{1}{2}\right)^k$$

as n grows large? What about $\sum_{k=1}^{n-1}\left(\frac{1}{2}\right)^k$ for large n?

6. Using the expression for $Var(\bar{X})$ from Question 4 and asymptotic results from Question 5, approximate the variance of the mean of a sample from $\{X_t\}$ for n large. Comment on the result.

7. Consider now a general AR(1) process:

$$X_t = \alpha X_{t-1} + Z_t, \qquad |\alpha| < 1.$$

Similar to Question 6, derive an asymptotic approximation for $Var(\bar{X})$. What does this tell us when α is positive? What about when α is negative? Compare the above with the result for $\mathrm{Var}\left(\bar{X}\right)$ when the data are i.i.d., and try to provide some intuition to the results.