Estimating the mean

Andreas Jakobsson

Estimating the mean

Generally, one does not know the statistical properties of the process, and will need to estimate these from the observed realization. Under the assumption that the process is ergodic, the most natural estimator of the mean is

$$\hat{m}_y = \frac{1}{N}\sum_{t=1}^{N} y_t$$

This is an *unbiased* estimate of the true mean, m_y, as

$$E\left\{\hat{m}_y\right\} = \frac{1}{N}\sum_{t=1}^{N} E\left\{\hat{y}_t\right\} = m_y$$

The estimate is also *consistent*. This is a highly desirable property of an estimator, implying that the estimate is (asymptotically) unbiased and that the variance decrease to zero as the data length grows.

Estimating the mean

Generally, one does not know the statistical properties of the process, and will need to estimate these from the observed realization. Under the assumption that the process is ergodic, the most natural estimator of the mean is

$$\hat{m}_y = \frac{1}{N}\sum_{t=1}^{N} y_t$$

This is an *unbiased* estimate of the true mean, m_y, as

$$E\left\{\hat{m}_y\right\} = \frac{1}{N}\sum_{t=1}^{N} E\left\{\hat{y}_t\right\} = m_y$$

The estimate is also *consistent*. This is a highly desirable property of an estimator, implying that the estimate is (asymptotically) unbiased and that the variance decrease to zero as the data length grows.

The variance of the mean may be expressed as

$$V\left\{\hat{m}_y\right\} = \frac{1}{N^2}\sum_{t=1}^{N}\sum_{s=1}^{N} r_y(t-s) = \frac{r_y(0)}{N}\sum_{k=-N+1}^{N-1}\frac{N-|k|}{N}\rho_y(k) \to 0, \quad N \to \infty$$

Estimating the mean

In case y_t is a Gaussian process, with mean m_y and ACF $r_y(k)$, then

$$\lim_{N\to\infty} N V\{\hat{m}_y\} = \sum_{k=-\infty}^{\infty} r_y(k)$$

Thus, $\hat{m}_y$ is a consistent estimate to m_y. For large N, an often useful approximation is

$$V\{\hat{m}_y\} \approx \frac{1}{N}\sum_{k=-\infty}^{\infty} r_y(k)$$

In the special case of a white process,

$$V\{\hat{m}_y\} \approx r_y(0)/N$$

Estimating the mean

In case y_t is a Gaussian process, with mean m_y and ACF $r_y(k)$, then

$$\lim_{N\to\infty} NV\{\hat{m}_y\} = \sum_{k=-\infty}^{\infty} r_y(k)$$

Thus, $\hat{m}_y$ is a consistent estimate to m_y. For large N, an often useful approximation is

$$V\{\hat{m}_y\} \approx \frac{1}{N} \sum_{k=-\infty}^{\infty} r_y(k)$$

In the special case of a white process,

$$V\{\hat{m}_y\} \approx r_y(0)/N$$

This result may be used to determine the confidence in the estimate. For example, for a Gaussian process, the 95% quantile is 1.96, suggesting the confidence interval for $\hat{m}_x$ is

$$\hat{m}_x \pm \frac{1.96\sigma_x}{\sqrt{N}}$$