Positive and negative dependence

In the two examples below, the gain of using either an average of nearby samples,

$$\hat{m}_1 = \frac{Y_t + Y_{t-1}}{2},$$

or an average from every other sample,

$$\hat{m}_2 = \frac{Y_t + Y_{t-2}}{2},$$

is investigated. We will show in two small examples that depending of what covariance structure the process has, either $\hat{m}_1$ or $\hat{m}_2$ will be shown to have the smallest variance.

Example 1: Positive dependence

The stationary process X_t, $t = 0, \pm 1, \pm 2, \ldots$ has mean value, $m_X = 0$, and covariance function

$$r_X(\tau) = \begin{cases} 5/4 & \tau = 0 \\ 1/2 & \tau = \pm 1 \\ 0 & \text{otherwise.} \end{cases}$$

A new process is defined as $Y_t = X_t + m_Y$, where the average level m_Y should be estimated. We have the already mentioned two choices for the estimation of m_Y,

$$\hat{m}_1 = \frac{Y_t + Y_{t-1}}{2} \quad \text{or} \quad \hat{m}_2 = \frac{Y_t + Y_{t-2}}{2}.$$

For this example, which of these two estimates is the most reliable, i.e., has the lowest variance?

Example 1: Positive dependence

The variances are given as

$$\begin{aligned} V[\hat{m}_1] &= C[\frac{Y_t + Y_{t-1}}{2}, \frac{Y_t + Y_{t-1}}{2}] = \\ &= \frac{1}{4}(V[Y_t] + 2C[Y_{t-1}, Y_t] + V[Y_{t-1}]) = \frac{1}{2}(r_Y(0) + r_Y(1)), \end{aligned}$$

and similarly

$$\begin{aligned} V[\hat{m}_2] &= C[\frac{Y_t + Y_{t-2}}{2}, \frac{Y_t + Y_{t-2}}{2}] = \\ &= \frac{1}{4}(V[Y_t] + 2C[Y_{t-2}, Y_t] + V[Y_{t-2}]) = \frac{1}{2}(r_Y(0) + r_Y(2)). \end{aligned}$$

Example 1: Positive dependence

The covariance function of Y_t is exactly the same as for X_t as the two processes only differ by a constant level.

Therefore,

$$V[\hat{m}_1] = \frac{1}{2}(r_X(0) + r_X(1)) = \frac{1}{2}(\frac{5}{4} + \frac{1}{2}) = \frac{7}{8},$$

and

$$V[\hat{m}_2] = \frac{1}{2}(r_X(0) + r_X(2)) = \frac{1}{2}(\frac{5}{4} + 0) = \frac{5}{8}.$$

We find that $V[\hat{m}_2] < V[\hat{m}_1]$, i.e. $\hat{m}_2$ is a more reliable estimator of the average level m_Y. The difference is that in $V[\hat{m}_1]$, the positive value of $r_X(1)$ contributes to the higher variance, where in $V[\hat{m}_2]$, the only contribution to the variance comes from $r_X(0)$ as $r_X(2)$ is zero.

Example 2: Negative dependence

The stationary process $X_t,\ t = 0, \pm 1, \pm 2, \ldots$ has mean value, $m_X = 0$, and a covariance function with negative dependence,

$$r_X(\tau) = \begin{cases} 5/4 & \tau = 0 \\ -1/2 & \tau = \pm 1 \\ 0 & \text{otherwise.} \end{cases}$$

A new process is defined as $Y_t = X_t + m_Y$, and the average level m_Y should be estimated. We have the same two choices for the estimation of m_Y,

$$\hat{m}_1 = \frac{Y_t + Y_{t-1}}{2} \quad \text{or} \quad \hat{m}_2 = \frac{Y_t + Y_{t-2}}{2}.$$

Which of these two estimates is the best in this case, i.e., has the lowest variance?

Example 2: Negative dependence

The variances are exactly the same as in example 1,

$$V[\hat{m}_1] = \frac{1}{2}(r_Y(0) + r_Y(1)),$$

and

$$V[\hat{m}_2] = \frac{1}{2}(r_Y(0) + r_Y(2)).$$

With $r_X(\tau) = r_Y(\tau)$,

$$V[\hat{m}_1] = \frac{1}{2}(r_X(0) + r_X(1)) = \frac{1}{2}(\frac{5}{4} - \frac{1}{2}) = \frac{3}{8},$$

and

$$V[\hat{m}_2] = \frac{1}{2}(r_X(0) + r_X(2)) = \frac{1}{2}(\frac{5}{4} + 0) = \frac{5}{8}.$$

In this case $V[\hat{m}_1] < V[\hat{m}_2]$, i.e. $\hat{m}_1$ is a more reliable estimator of the average level m_Y. In $V[\hat{m}_1]$, the negative value of $r_X(1)$ now contributes to a smaller variance, where, as in example 1, for $V[\hat{m}_2]$, the only contribution to the variance comes from $r_X(0)$.

Conclusion

For a process with positive dependence, i.e. a process that changes slowly, we get a more reliable estimate of the mean value if the samples that are averaged are far apart, if possible, as far as no positive covariances are included in the average. The positive dependence contributes to a larger variance of the mean value, so the best is to average as independent samples as possible. On the other hand, if the process has negative dependence, i.e. the process changes quickly, the negative covariance values contributes to a reduced variance of the mean value, compared to using only independent samples.