Activity Solution: The Cross–Correlogram

Recall for a bivariate stochastic process $\{X(t), Y(t)\}$ we defined the *cross–covariance function* at lag k to be

$$\begin{aligned}\gamma_{XY}(k) &= \operatorname{Cov}(X(t), Y(t+k)) \\ &= E\left((X(t)-\mu_X)(Y(t+k)-\mu_Y)\right).\end{aligned}$$

The *cross–correlation* at lag k is defined as

$$\rho_{XY}(k) = \frac{\gamma_{XY}(k)}{\sqrt{\gamma_X(0)\gamma_Y(0)}} = \frac{\gamma_{XY}(k)}{\sigma_X \sigma_Y}.$$

1. *Clickers question:* Let

   $$X(t) = 0.9X(t-1) + Z_1(t)$$

   where $Z_1(t)$ is white noise with variance 1, and let

   $$Y(t) = 0.9Y(t-1) + Z_2(t)$$

   where $Z_2(t)$ is white noise with variance 1, independent of $Z_1(t)$. What is $\rho_{XY}(k)$?
   As $X(t)$ and $Y(t)$ are independent, $\rho_{XY}(k) = 0$ for all k.

2. *Clickers question:* Given bivariate data $(x(1), y(1)), \ldots, (x(N), y(N))$, we can estimate $\gamma_{XY}(k)$ using the *sample cross–covariance* function, defined as

   $$c_{XY}(k) := \frac{1}{N}\sum_{t=1}^{N-k}(x(t)-\bar{x})(y(t+k)-\bar{y})$$

   for $k = 0, 1, \ldots, N$ and

   $$c_{XY}(k) := \frac{1}{N}\sum_{t=1-k}^{N}(x(t)-\bar{x})(y(t+k)-\bar{y})$$

   for $k = -1, \ldots, -(N-1)$. The *sample cross–correlation* estimates $\rho_{XY}(k)$ and is

   $$r_{XY}(k) = \frac{c_{XY}(k)}{s_X s_Y}.$$

A plot of $r_{XY}(k)$ against k is known as the *cross–correlogram.* Suppose $\{x(t)\}$ and $\{y(t)\}$ are realisations from $X(t)$ and $Y(t)$ with

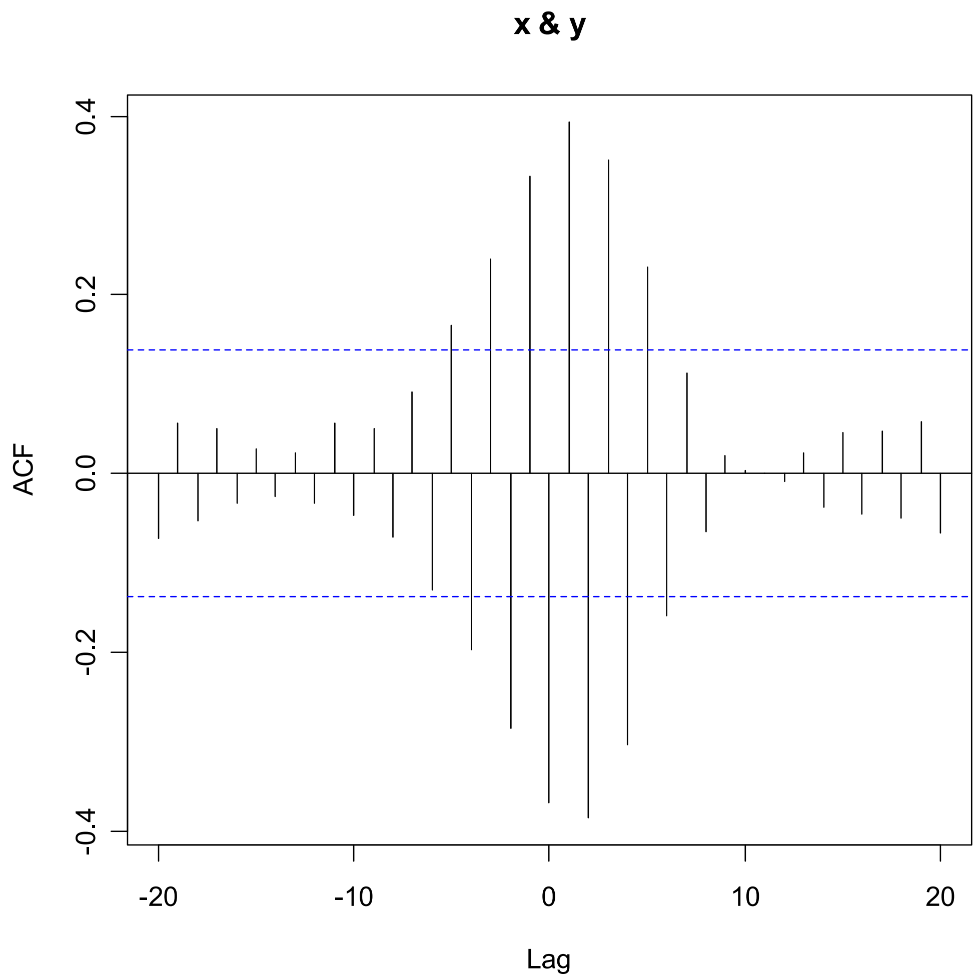
$$X(t) = -0.9X(t-1) + Z_1(t)$$

where $Z_1(t)$ is white noise with variance 1, and

$$Y(t) = -0.9Y(t-1) + Z_2(t)$$

where $Z_2(t)$ is white noise with variance 1, independent of $Z_1(t)$. Consider the sample cross–correlation function from 200 values for the process $\{X(t), Y(t)\}$. How would you expect the cross–correlogram to look?

An example of the cross-correlogram in this case is given below:



r_{XY} for independent samples of 200 from AR(1), $\alpha = -0.9$.

The two series will tend to alternate, sometimes in and sometimes out of phase, making absolute values of r_{XY} *large. However, the two series are completely independent. The above suggests the two realisations were (mostly) out of phase, noting the large negative correlation at lag zero.*

3. *Clickers question:* Let

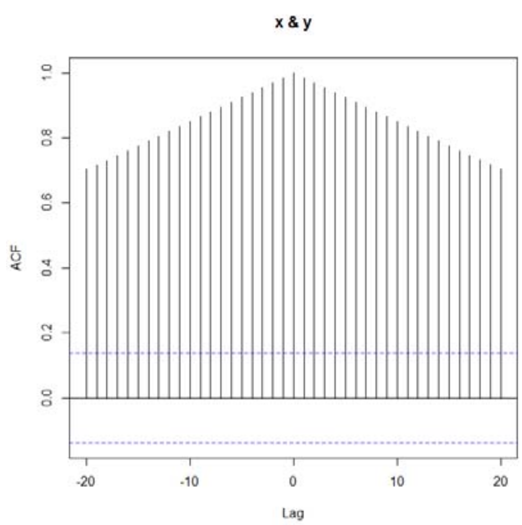
$$\begin{aligned} X(t) &= 2 + 5t + Z_1(t) \\ Y(t) &= 2 + 5t + Z_2(t) \end{aligned}$$

where $Z_1(t)$ and $Z_2(t)$ are independent white noise processes each with variance 1. What is $\rho_{XY}(0)$? What is $\rho_{XY}(k)$?
With $(X(t), Y(t))$ as above,

$$\begin{aligned} \gamma_{XY}(0) &= \mathrm{Cov}(X(t), Y(t)) \\ &= \mathrm{Cov}(2 + 5t + Z_1(t), 2 + 5t + Z_2(t)) \\ &= 0 \end{aligned}$$

as Z_1 and Z_2 are independent. Moreover, in general $\rho_{XY}(k) = 0$.

4. Consider the sample cross–correlation function from 200 values from the process $\{X(t), Y(t)\}$ above. How would you expect the cross–correlogram to look?
Although actually independent, observations from $X(t)$ will tend to closely resemble those from $Y(t)$, giving (spurious) high values in the cross-correlogram.



5. Re-cap what you have done in this activity. What did you learn?
We introduced a natural estimator for the cross–correlation function,

a plot of which against the lag is known as the correlogram. We have observed that in two special cases where samples are from a bivariate process $\{X(t), Y(t)\}$, *although* $X(t)$ *and* $Y(t)$ *were independent the correlogram suggested a strong dependence between the two. This tells us that the correlogram must be interpreted with extreme care, as there is a danger of inferring causation between two series when none exists. Learning outcomes encountered in the activity include:*

(a) *Define the sample cross–correlation function and explain its interpretation in practice.*

(b) *Recall key sampling properties of the sample cross–correlation function.*