Statistics 519, Winter Quarter 2020

Problem Set 2

Problem 4 (3 points for each for the 3 parts). Suppose that $\{X_t : t \in \mathbb{Z}\}$ is a stationary process with zero mean and with ACVF $\gamma_X(\cdot)$ and that we subsample it by taking every other value to form a new process defined by $Y_t = X_{2t}$; i.e., the process $\{\ldots, Y_{-2}, Y_{-1}, Y_0, Y_1, Y_2, \ldots\}$ is given by $\{\ldots, X_{-4}, X_{-2}, X_0, X_2, X_4, \ldots\}$

a. Show that $\{Y_t\}$ is a stationary process, and determine its ACVF $\gamma_Y(\cdot)$.

b. If $\{X_t\}$ is an AR(1) process with parameters ϕ and σ^2 (excluding the possibilities $\phi = 0$ and/or $\sigma^2 = 0$), does the process $\{Y_t\}$ have an ACVF matching that of either an AR(1) process, an MA(1) process or a white noise process? If so, determine the parameters of the matching process in terms of ϕ and σ^2.

c. If $\{X_t\}$ is an MA(1) process with parameters θ and σ^2 (excluding the possibilities $\theta = 0$ and/or $\sigma^2 = 0$), does the process $\{Y_t\}$ have an ACVF matching that of either an AR(1) process, an MA(1) process or a white noise process? If so, determine the parameters of the matching process in terms of θ and σ^2.

Problem 5 (2 points for each of the 5 parts). (Here we expand upon a point made on overhead II–65, namely, that values in the sample autocorrelation function are not true correlations.) Suppose that we have two sets of numbers u_t and v_t, $t = 1, \ldots, m$, and that we form their sample correlation, for which the standard definition is

$$\frac{\sum_{t=1}^{m}(u_t - \bar{u})(v_t - \bar{v})}{\sqrt{\left[\sum_{t=1}^{m}(u_t - \bar{u})^2 \sum_{t=1}^{m}(v_t - \bar{v})^2\right]}},$$

where $\bar{u} \stackrel{\text{def}}{=} \frac{1}{m}\sum_{t=1}^{m} u_t$ and $\bar{v} \stackrel{\text{def}}{=} \frac{1}{m}\sum_{t=1}^{m} v_t$ are sample means. Given a time series x_t, $t = 1, \ldots, n$, a proper estimate of the lag $h > 0$ autocorrelation would be gotten by letting $u_t = x_{t+h}$, $v_t = x_t$ and $m = n - h$, leading to

$$\tilde{\rho}_X(h) \stackrel{\text{def}}{=} \frac{\sum_{t=1}^{n-h}(x_{t+h} - \bar{x}_{h+1:n})(x_t - \bar{x}_{1:n-h})}{\sqrt{\left[\sum_{t=1}^{n-h}(x_{t+h} - \bar{x}_{h+1:n})^2 \sum_{t=1}^{n-h}(x_t - \bar{x}_{1:n-h})^2\right]}}, \text{ where } \bar{x}_{j:k} \stackrel{\text{def}}{=} \frac{1}{k-j+1}\sum_{t=j}^{k} x_t.$$

By contrast, the sample autocorrelation is

$$\hat{\rho}_X(h) \stackrel{\text{def}}{=} \frac{\sum_{t=1}^{n-h}(x_{t+h} - \bar{x})(x_t - \bar{x})}{\sum_{t=1}^{n}(x_t - \bar{x})^2} \text{ with } \bar{x} \stackrel{\text{def}}{=} \frac{1}{n}\sum_{t=1}^{n} x_t,$$

which, in the context of stationary processes, has some intuitive appeal over $\tilde{\rho}_X(h)$ (e.g., on average, $\bar{x}$ should be a better estimate of the process mean than either $\bar{x}_{h+1:n}$ or $\bar{x}_{1:n-h}$).

Suppose now that $x_t = \alpha + \beta t$, $t = 1, \ldots, n$, where α and $\beta \neq 0$ are constants.

a. Derive an expression that describes how the scatter plot of x_{t+h} versus x_t depends upon α, β and h. Make a plot of x_{t+70} versus x_t, $t = 1, \ldots, 30$, for the specific case $\alpha = 1$, $\beta = 2$ and $n = 100$, and verify that your theoretical expression matches the plot.

b. Derive an expression for $\hat{\rho}_X(h)$ that is valid for the assumed $\{x_t\}$ and $h = 0, 1, \ldots, n-1$. Use the function `acf` in `R` (or equivalent software) to create a plot of $\hat{\rho}_X(h)$ versus $h = 0, 1, \ldots, 99$ for the specific case $\alpha = 1$, $\beta = 2$ and $n = 100$, and verify that your theoretical expression matches the plot. Two facts that might prove useful are $\sum_{t=1}^{m} t = m(m+1)/2$ and $\sum_{t=1}^{m} t^2 = m(m+1)(2m+1)/6$.

c. Based upon the expression derived in part b, argue that, for large n, $\hat{\rho}_X(h)$ achieves a minimum value at approximately $h = n/\sqrt{2}$ and that the minimum value is approximately $1 - \sqrt{2} \doteq -0.41$. How well do these approximations match up with the plot called for in part b?

d. Show that $\tilde{\rho}_X(h) = 1$ for $0 \le h \le n-1$. Hint: argue that $x_{t+h} - \bar{x}_{h+1:n} = x_t - \bar{x}_{1:n-h}$.

e. For the specific case considered in part a, how does $\hat{\rho}_X(70)$ compare to $\tilde{\rho}_X(70)$? Which one is the appropriate summary of the scatter plot requested in part a?

Problem 6 (3 points for each of the 2 parts). Given a time series $\{x_t\}$, let $\{w_t\}$ be the output from a two-sided moving average filter:

$$w_t = \frac{1}{2q+1} \sum_{j=-q}^{q} x_{t-j},$$

where q is a nonnegative integer (see overhead III–26).

a. Show that, if the time series is locally linear at t, i.e., $x_{t-j} = a + b(t-j)$ for $j = -q, \ldots, q$, then $w_t = x_t$.

b. Suppose now that x_t is a realization of $X_t = m_t + Z_t$, where $\{m_t\}$ is a deterministic trend and $\{Z_t\} \sim \mathrm{WN}(0, \sigma^2)$. Under the assumption that $m_{t-j} \approx a + b(t-j)$ for $j = -q, \ldots, q$, show that w_t is a realization of an approximately unbiased estimator of m_t, and determine the variance of this estimator.

Solutions are due **Friday**, January 24, at the beginning of the class.