Tilastollisen päättelyn jatkokurssi, sl 2010, Exercise 2, week 38

1. Let $(Y_1, X_1), ..., (Y_n, X_n)$ be an independent sample from a bivariate normal distribution with $\rho = \mathsf{Cor}(Y_1, X_1)$ $(|\rho| < 1)$ the theoretical correlation coefficient. It is known that the sample correlation coefficient (or the maximum likelihood estimator of ρ)

$$R_n = \frac{\sum_{i=1}^n \left(Y_i - \bar{Y}_n\right)\left(X_i - \bar{X}_n\right)}{\sqrt{\sum_{i=1}^n \left(Y_i - \bar{Y}_n\right)^2}\sqrt{\sum_{i=1}^n \left(X_i - \bar{X}_n\right)^2}},$$

satisfies $\sqrt{n}\,(R_n - \rho) \xrightarrow{d} \mathsf{N}(0, (1-\rho^2)^2)$. Consider the so–called Fisher's z–transformation

$$Z_n = \frac{1}{2}\log\left(\frac{1+R_n}{1-R_n}\right) \quad \text{and} \quad \zeta = \frac{1}{2}\log\left(\frac{1+\zeta}{1-\zeta}\right).$$

(i) Use the delta method to show that $\sqrt{n}\,(Z_n - \zeta) \xrightarrow{d} \mathsf{N}(0,1)$ (for simplicity the subscript 0 has been dropped here from ρ and ζ).

(ii) Use the preceding result and derive an approximate test (based on Z_n) for the null hypothesis $\rho = 0$ against the alternative $\rho \neq 0$.

Note: Because there is no need to estimate the variance of the limiting distribution the normal approximation works better for Z_n than for R_n which is useful when one constructs tests and confidence intervals for ρ.

2. Let $Y_1, ..., Y_n$ be an independent sample from a distribution with a finite fourth moment. Denote $\mathsf{E}(Y_1) = \mu$, $\mathsf{Var}(Y_1) = \sigma^2$ and $\mu_4 = \mathsf{E}[(Y_1 - \mu)^4]$ and, furthermore,

$$\tilde{S}_n^2 = \frac{1}{n}\sum_{i=1}^n (Y_i - \mu)^2, \quad \hat{S}_n^2 = \frac{1}{n}\sum_{i=1}^n \left(Y_i - \bar{Y}_n\right)^2 \quad \text{and} \quad S_n^2 = \frac{1}{n-1}\sum_{i=1}^n \left(Y_i - \bar{Y}_n\right)^2.$$

Show that

$\tilde{S}_n^2 \xrightarrow{p} \sigma^2$ and $\sqrt{n}(\tilde{S}_n^2 - \sigma^2) \xrightarrow{d} \mathsf{N}(0, \upsilon)$, where $\upsilon = \mathsf{Var}[(Y_1 - \mu)^2] = \mu_4 - \sigma^4$.

3. (Continuation for the preceding one) (i) Show that $\sqrt{n}(\tilde{S}_n^2 - \hat{S}_n^2) \xrightarrow{p} 0$ and conclude from this and the preceding exercise that $\hat{S}_n^2 \xrightarrow{p} \sigma^2$ and $\sqrt{n}(\hat{S}_n^2 - \sigma^2) \xrightarrow{d} \mathsf{N}(0, \upsilon)$.

(ii) Use arguments similar to those in (i) and show that the usual sample variance satisfies $S_n^2 \xrightarrow{p} \sigma^2$ and $\sqrt{n-1}\left(S_n^2 - \sigma^2\right) \xrightarrow{d} \mathsf{N}(0, \upsilon)$.

4. Let $\varepsilon_1, \varepsilon_2, ...$ be an independent sequence with a $\mathsf{N}(0,1)$–distribution and $Y_i = \sum_{j=1}^i \varepsilon_j$, $i = 1, 2, ...$

(i) Show that $\mathsf{Cov}(Y_i, Y_{i+h}) = \mathsf{E}(Y_i Y_{i+h}) = i$ for all $h \geq 0$ and, using this, that $\mathsf{Cor}(Y_i, Y_{i+h}) = 1/\sqrt{1 + h/i}$ $(h \geq 0,\ i \geq 1)$. What happens to the correlation coefficient $\mathsf{Cor}(Y_i, Y_{i+h})$, when $i \to \infty$ and h is fixed?

(continues on backside)

(ii) Show the identity

$$X_n := \frac{1}{n}\sum_{i=1}^{n} Y_{i-1}\varepsilon_i = \frac{1}{2n}\left(Y_n^2 - \sum_{i=1}^{n}\varepsilon_i^2\right),$$

where $Y_0 = 0$ and $:=$ defines X_n.

Hint: One possibility in (ii) is to use the identity $\varepsilon_i = Y_i - Y_{i-1}$ obtained from the definition of Y_i and calculate an expression for the sum $\sum_{i=1}^{n}\varepsilon_i^2$. The desired identity is then obtained from the resulting equation.

5. (Continuation for the preceding one) Show that $\mathsf{E}(Y_{i-1}\varepsilon_i) = 0$ for all $i \geq 1$ and that $X_n \xrightarrow{d} \frac{1}{2}\left(\chi_1^2 - 1\right)$. Deduce from this that the law of large numbers (LLN) and central limit theorem (CLT) do not hold for the sample mean formed from $Y_{i-1}\varepsilon_i$, $i = 1, ..., n$. In the last point strict mathematical justification is not required.

Note: This in conjunction with exercise 4 shows that no usual LLN and CLT hold because the dependence is "too strong" (the variables $Y_{i-1}\varepsilon_i$, $i = 1, ..., n$, are strongly dependent even though uncorrelated). Note also that Y_i solves the autoregressive equation $Y_i = \phi Y_{i-1} + \varepsilon_i$ $(i = 1, ..., n)$, when $\phi = 1$ and $Y_0 = 0$.