Midterm II - Solutions

Applied Statistics and Econometrics II
Spring 2018, NYU
Ercan Karadas

1. Let $X \sim \text{Poisson}(\theta)$ with the p.m.f.

$$P(X = x) = \frac{\theta^x e^{-\theta}}{x!}$$

   (a) Find the mle of θ.

   The standard procedure gives the mle of θ as $\hat{\theta} = \bar{x}$.

   (b) Let the table represent a summary of a random sample of size 60 from the Pois-

x	0	1	2	3	4	5	6	7
Frequency	7	14	12	13	6	3	3	2

   son distribution above. Find the maximum likelihood estimate of $P(X = 3)$.

   For the given data we compute $\hat{\theta} = \bar{x} = 2.47$. Then we compute the mle of $P(X = 2)$ as

$$P(X = 3) = \frac{\hat{\theta}^3 e^{-\hat{\theta}}}{3!} = \frac{\bar{x}^3 e^{-\bar{x}}}{3!} = \frac{(2.47)^3 e^{-2.47}}{3!} \approx 0.21$$

2. Let $\{X_1, \ldots, X_5\}$ be an iid random sample from a uniform distribution with the support $(0, \theta)$. In this problem, you are asked to construct the likelihood ratio test to test the null hypothesis that the true parameter value is $\theta_0 = 10$. As data suppose we have a random sample of size $n = 5$ at hand with the maximum value of 8.

   (a) State the null and alternative hypotheses.

$$H_0 : \theta = 10, \qquad H_1 : \theta \neq 10$$

   (b) Calculate the likelihood ratio test statistic

$$\Lambda = \frac{L(\theta_0)}{L(\hat{\theta})}$$

   where $\hat{\theta}$ is the mle of θ.

We know that the mle $\hat{\theta} = \max_{x_i} = 8$ *and that*

$$L(\theta; \mathbf{x}) = \left(\frac{1}{\theta}\right)^n$$

provided $\theta \geq \max_{x_i}$. *Thus*

$$\Lambda = \frac{L(\theta_0)}{L(\hat{\theta})} = \left(\frac{(1/\theta_0)^n}{(1/\hat{\theta})^n}\right)^n = \left(\frac{\max_{x_i}}{\theta_0}\right)^n$$

and

$$\lambda = \left(\frac{8}{10}\right)^5 = 0.327$$

(c) Conclude the test using the following decision rule,

$$\text{Reject } H_0 \text{ in favor of } H_1 \text{ if } \Lambda \leq c,$$

where c is such that $\alpha = P_{\theta_0}[\Lambda \leq c]$ and take $\alpha = 0.1$.

Now we need to determine the p-value, that is, $P(\Lambda \leq \lambda \mid H_0$ *is true*). *In this example, we can do this fairly easily since we know the distribution of* $\max_{x_i}$*:*

$$\begin{aligned} P(\Lambda \leq \lambda) &= P\left(\left(\frac{\max_{x_i}}{\theta_0}\right)^n \leq \lambda\right) \\ &= P\left(\max_{x_i} \leq \theta_0(\lambda)^{1/n}\right) \\ &= P\left(x_i \leq \theta_0(\lambda)^{1/n}, \forall i\right) \\ &= \left((\lambda)^{1/n}\right)^n = \lambda = 0.327 \end{aligned}$$

With such a large p-value, we do not have enough evidence to reject the null hypothesis. Our data are consistent with the hypothesis that $\theta = 10$.

3. A researcher uses a sample of 200 quarterly observations on Y_t, the number (in 1000s) of unemployed people, to model the time series behaviour of the series and to generate predictions. First, he computes the sample autocorrelation and the sample partial autocorrelation functions, respectively, with the following results:

k	1	2	3	4	5	6	7	8
$\hat{\rho}_k$	0.83	0.71	0.60	0.45	0.44	0.35	0.29	0.20
$\hat{\theta}_{kk}$	0.83	0.16	-0.09	0.05	0.04	-0.05	0.01	0.01

(a) What do we mean by the sample autocorrelation and the sample partial autocorrelation functions? Why is the first partial autocorrelation equal to the first autocorrelation coefficient (0.83)

The autocorrelation function describes the evolution of the sample autocorrelation coefficient across different lag lengths. The current pattern indicates a slow decay in the sample autocorrelation coefficient and is consistent with an autoregressive process (with a small lag length) rather than a moving average process.
The sample partial autocorrelation function depicts the estimated coefficients of lag length k in an $AR(k)$ model for different lag lengths k. The first order sample autocorrelation coefficient and the sample partial autocorrelation coefficient are identical. Both measure the correlation between Y_t and Y_{t-1}.

(b) Does the above pattern indicate that an autoregressive or a moving average representation is more appropriate? Why?

The pattern is consistent with a (first or second order) autoregressive process, because the sample PACF is close to zero after lag 2.

Suppose the researcher decides to estimate, as a first attempt, an AR(1) model

$$Y_t = \delta + \theta Y_{t-1} + \varepsilon_t$$

and the OLS produces an estimated value of 0.83 for θ with a standard error of 0.07.

(c) Can OLS produce a consistent estimate of θ in this set up? Explain breifly why.

The first-order autoregressive model can be estimated by OLS. It is consistent because Y_{t-1} is uncorrelated with ε_t.

(d) The researcher wants to test for a unit root. What is meant by 'a unit root'? What are the implications of the presence of a unit root? Why are we interested in it? (Give statistical or economic reasons.)

A unit root arises when the autoregressive lag polynomial $\theta(L)$ has a root of unity. In this case $\theta(1) = 1 - \theta = 0$ and the autoregressive process is nonstationary. In a nonstationary process, shocks have a permanent effect. Statistically, nonstationary series have very different properties than stationary ones. For example, a series with a unit root does not have a well-defined mean and has an infinite variance.

(e) Formulate the hypothesis of a unit root and perform a unit root test based on the above regression.

The unit root hypothesis says $H_0 : \theta = 1$. Based on the provided results, the Dickey-Fuller test statistic is

$$DF = \frac{0.83 - 1}{0.07} = -2.482$$

Using the critical values of DF Table (with no trend), this does not allow us to reject the null hypothesis.

Next, the researcher extends the model to an AR(2), with the following results (standard errors in parentheses):

$$\begin{array}{ccccccc} Y_t & = & 50 & + 0.74\ Y_{t-1} + & 0.16\ Y_{t-2} + & e_t \\ & & (5.67) & (0.07) & (0.07) & \end{array}$$

(f) Would you prefer the AR(2) model to the AR(1) model? How would you check whether an ARMA(2, 1) model may be more appropriate?

Based on the given information, the only criterion to choose between the AR(1) and the AR(2) model is the statistical significance of the coefficient for the 2nd lag in the AR(2) model. With a t-statistic of 0.17/0.07 = 2.43, the null hypothesis that $\theta_2 = 0$ *is rejected (at 95% confidence), and the AR(2) model is preferable to the AR(1). One can check the adequacy of an ARM A(2, 1) specification by estimating it and then testing the significance of the MA coefficient. Alternatively, one could use the AIC or BIC criteria, although they are not provided here.*

(g) How would you test for the presence of a unit root in the AR(2) model?

In the AR(2) model, one can test for a unit root using an augmented Dickey-Fuller test. Based on the estimation results, the ADF test statistic is

$$ADF(1) = \frac{0.74 - 1}{0.07} = -3.714$$

which rejects the null hypothesis of a unit root.

(h) From the above estimates, compute an estimate for the average number of unemployed, $E[Y_t]$.

The unconditional mean of Y_t *is*

$$\mu = \frac{\delta}{1 - \theta_1 - \theta_2}$$

which gives the estimated value as

$$\mu = \frac{0.5}{1 - 0.74 - 0.16} = 500$$

(i) Suppose the last two quarterly unemployment levels for 2017:III and 2017:IV were 550 and 600, respectively. Compute forecasts for 2018:I and 2018:II.

The one-period ahead forecast is given by

$$Y_{T+1|T} = \mu + \theta_1(Y_T - \mu) + \theta_2(Y_{T-1} - \mu)$$

where T denotes the final sample period. The two-period ahead forecast is

$$Y_{T+2|T} = \mu + \theta_1(Y_{T+1|T} - \mu) + \theta_2(Y_T - \mu)$$

The forecast for 2018:I and 2018:II are calculated as

$$\hat{Y}_{2018:I} = 500 + 0.74(600 - 500) + 0.16(550 - 500) = 582$$
$$\hat{Y}_{2018:II} = 500 + 0.74(582 - 500) + 0.16(600 - 500) = 568.68$$

4. Consider the process

$$Y_t = \delta + \phi Y_{t-1} + w_t,$$

where w_t is a white noise process with variance σ_w^2 and let $|\phi| < 1$ be a constant.

(a) Show that

$$Y_t = \frac{\delta}{1-\phi} + \sum_{j=0}^{t} \phi^j w_{t-j}$$

This is essentially the same problem as Problem 3.2 in PS7, just define

$$x_t = Y_t - \mu,$$

where $\mu = \delta/(1-\phi)$. Use induction or insert the solution into the equation, i.e.,

$$\sum_{j=0}^{t} \phi^j w_{t-j} = \phi\left(\sum_{j=0}^{t-1} \phi^j w_{t-1-j}\right) + w_j$$

to show

$$x_t = \sum_{j=0}^{t} \phi^j w_{t-j}$$

then the result immediately follows.

(b) Find $E[Y_t]$.

$E[Y_t] = \mu + \sum_{j=0}^{t} \phi^j E(w_{t-j}) = \mu$

(c) Show that,

$$\text{Var}[Y_t] = \frac{\sigma_w^2}{1-\phi^2}\left(1 - \phi^{2(t+1)}\right), \qquad t = 0, 1, \ldots$$

$Var[Y_t] = Var[x_t] = \sum_{j=0}^{t} \phi^{2j}\, Var(w_{t-j}) = \sigma_w^2 \sum_{j=0}^{t} \phi^{2j} = \frac{\sigma_w^2}{1-\phi^2}\left(1 - \phi^{2(t+1)}\right)$ *using the fact that w_t is noise.*

(d) Show that, for $h \geq 0$,

$$\text{Cov}(Y_t, Y_{t+h}) = \phi^h \text{Var}(Y_t)$$

$Cov(Y_t, Y_{t+h}) = Cov(x_t, x_{t+h})$ so this part is the same. Notice that

$$\begin{aligned} x_{t+h} &= \sum_{j=0}^{t+h} \phi^j w_{t+h-j} \\ &= \sum_{j=0}^{h-1} \phi^j w_{t+h-j} + \sum_{j=h}^{t+h} \phi^j w_{t+h-j} \\ &= \sum_{j=0}^{h-1} \phi^j w_{t+h-j} + \phi^h \sum_{k=0}^{t} \phi^k w_{t-k} \\ &= \sum_{j=0}^{h-1} \phi^j w_{t+h-j} + \phi^h x_t \end{aligned}$$

Alternately, just iterate x_{t+h} back h time units.
Since x_t involves the w_s for $s \leq t$,

$$Cov(x_t, x_{t+h}) = Cov\left(\sum_{j=0}^{h-1} \phi^j w_{t+h-j} + \phi^h x_t, x_t\right) = \phi^h\, Var(x_t)$$

(e) Is Y_t stationary?

Y_t is not stationary because the (co)variance depends on time t.

(f) Argue that, as $t \to \infty$, the process becomes stationary, so in a sense, Y_t is "asymptotically stationary".

As $t \to \infty$, $Var(Y_t) \to \frac{\sigma_w^2}{1-\phi^2}$ by part (c), and hence by (d), the autocovariance function is independent of t.

	Without trend		With trend	
Sample size	1 %	5 %	1 %	5 %
$T = 25$	−3.75	−3.00	−4.38	−3.60
$T = 50$	−3.58	−2.93	−4.15	−3.50
$T = 100$	−3.51	−2.89	−4.04	−3.45
$T = 250$	−3.46	−2.88	−3.99	−3.43
$T = 500$	−3.44	−2.87	−3.98	−3.42
$T = \infty$	−3.43	−2.86	−3.96	−3.41

Figure 1: 1% and 5% critical values for Dickey-Fuller tests