SOLUTIONS TO EXERCISES ON BAYESIAN ESTIMATION

PAWEL ZABCZYK

Centre for Central Banking Studies
Bank of England

1. A Poisson example

For independent Poisson observations the joint density will simply equal the product of the individual densities. Accordingly

$$p(x,\ \theta^*) = \prod_{i=1}^{10} p(x_i,\ \theta^*) = \frac{e^{-10\theta^*}\,(\theta^*)^{\sum_{i=1}^{10} x_i}}{\prod_{i=1}^{n}(x_i!)} = \frac{e^{-10\theta^*}\,(\theta^*)^{20}}{207,360}$$

and so the likelihood function, which is a function of θ (rather than the true parameter value θ^*) can be written down as

$$L(\theta) = \frac{e^{-10\theta}\theta^{20}}{207,360}.$$

The plot in Figure 1 below is generate in the Excel file Solutions.xlsx. Of course, the plot suggests

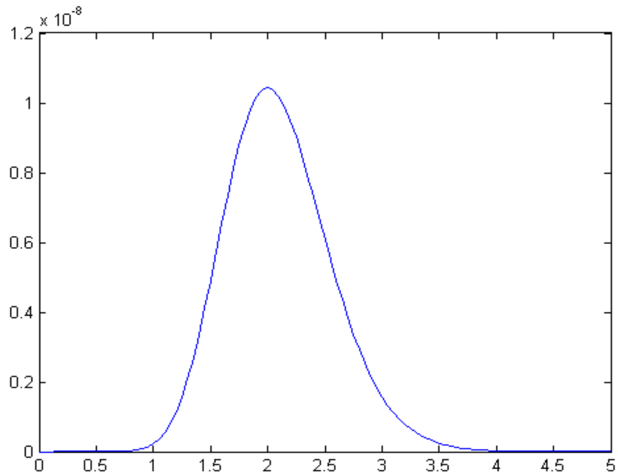

Figure 1. Poisson distribution likelihood function

a unique maximum at $\theta = 2$. To verify whether that is exactly right, we can compute the ML estimator by calculus. Since the log of the likelihood equals

$$\ln L(\theta) = -10\theta + 20\ln\theta - \ln(207,360)$$

therefore the corresponding first order condition is[1]

$$\frac{\partial \ln L(\theta)}{\partial \theta} = -10 + \frac{20}{\theta} \Rightarrow \theta^* = \frac{20}{10} = 2$$

and so the algebra confirms what the chart suggested, i.e. that our ML estimate of θ^* equals 2. Notably, 2 also happens to be the value of the sample mean, so we could also have guessed the estimate of θ^* by averaging across all observations (we have 10 and they sum to 20). Equally, we could have found that 2 is the maximum by simply inspecting the function values computed in Excel.

2. Maximum Likelihood Linear Regressions in Eviews

The walk-through for this exercise can be found in the solution manual.

3. Estimating a fat-tailed regression by ML in Eviews

The OLS coefficients equal:

Equation: UNTITLED Workfile: ML_EXERCISE::Untitled\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: YFT
Method: Least Squares
Date: 04/11/16 Time: 21:05
Sample: 1 320
Included observations: 320
YFT = C(1) + C(2)*X1 + C(3)*X2

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	2.939537	3.650431	0.805257	0.4213
C(2)	1.987793	0.012025	165.3019	0.0000
C(3)	-4.076560	2.667386	-1.528297	0.1274

R-squared	0.993760	Mean dependent var	93.59198
Adjusted R-squared	0.993721	S.D. dependent var	20.67807
S.E. of regression	1.638564	Akaike info criterion	3.834848
Sum squared resid	851.1106	Schwarz criterion	3.870176
Log likelihood	-610.5757	Hannan-Quinn criter.	3.848955
F-statistic	25242.68	Durbin-Watson stat	1.998446
Prob(F-statistic)	0.000000		

[1] Note: as before, we take logs only to simplify algebra, and doing so will *not* affect the final value of the ML estimate. Furthermore, if we hadn't plotted the chart, we could double check that we've found a maximum rather than minimum by verifying that the sign of the second derivative at the estimate is negative

$$\frac{\partial^2 \ln L(\theta)}{\partial \theta^2} = -\frac{20}{\theta^2} < 0.$$

Despite the high R^2 the coefficients $c(1)$ and $c(3)$ are actually very imprecisely estimated. This is because most of the volatility is coming from `x2` (with the the corresponding coefficient estimated very well indeed).

To estimate the specification by Maximum Likelihood we proceed as before and define a new likelihood object given by

```
@logl LL2
param c(1) 0 c(2) 0 c(3) 0
res2 = yft - c(1) - c(2)*x1 - c(3)*x2
v =3
pi = @acos(-1)
const =@gamma((v+1)/2)/(@sqrt(v*pi)*@gamma(v/2))
LL2= log(const)  - ((v+1)/2)*log(1+(res2^2/v))
```

Proceeding as before we see that we get a better estimate of $c(2)$, but the problem with $c(1)$ and $c(3)$ has not gone away – i.e. both remain imprecisely estimated. Of course, to get around this we could have chosen a smaller value of $c(2)$ when computing y^{ft}.

4. The Steepest Ascent Method

To apply the method of steepest ascent we need to compute the gradient vector. The elements of the gradient vector are

$$\frac{\partial \mathcal{L}(\theta)}{\partial \theta_1} = -3\theta_1 \qquad \frac{\partial \mathcal{L}(\theta)}{\partial \theta_2} = -4\theta_2.$$

The gradient vector evaluated at the initial guess $\theta^0 = (-1, 1)'$ therefore equals

$$\left.\frac{\partial \mathcal{L}(\theta)}{\partial \theta_1}\right|_{\theta=\theta^0} = 3 \qquad \left.\frac{\partial \mathcal{L}(\theta)}{\partial \theta_2}\right|_{\theta=\theta^0} = -4.$$

We thus see that an increase in θ_1 would increase the likelihood, while an increase in θ_2 would decrease it. Since the gradient vector equals $(3, -4)'$ therefore the optimal step should be proportional to $(3, -4)'$. For a step of length k the new guess would be

$$\theta_1^1 - \theta_1^0 = 3k \qquad \theta_2^1 - \theta_2^0 = -4k.$$

Plugging in $k = 1$ gives us

$$\theta_1^1 = -1 + 3 = 2 \qquad \theta_2^1 = 1 - 4 = -3$$

while for $k = 1/5$ we get

$$\theta_1^1 = -1 + 3/5 = -2/5 \qquad \theta_2^1 = 1 - 4/5 = 1/5.$$

The value of the likelihood corresponding to $k = 1$ is

$$\mathcal{L}(\theta) = -1.5 \times 4 - 2 \times 9 = -24$$

while the value of the likelihood for $k = 1/5$ equals

$$\mathcal{L}(\theta) = -1.5 \times 4/25 - 2 \times 1/25 = -8/25.$$

So we see that picking a large step length ($k = 1$) actually lowers the likelihood, while the smaller step ($k = 1/5$) results in a higher likelihood value and so would be preferred.

5. Hypothesis Testing (taken from Hamilton, Ch.5)

To find the restricted likelihood, we replace θ_2 by $\theta_1 + 1$ and maximise the resulting expression with respect to θ_1:

$$\mathcal{L}(\theta^r) = -1.5\theta_1^2 - 2(\theta_1 + 1)^2.$$

The first order condition for the maximisation of $\mathcal{L}(\theta^r)$ is

$$-3\theta_1 - 4(\theta_1 + 1) = 0,$$

or $\theta_1 = -4/7$. The restricted MLE is thus $\theta^r = (-4/7, 3/7)'$ and the maximum value attained for the log likelihood while satisfying the restriction is

$$\mathcal{L}(\theta^r) = (-1/2) \times (-4/7)^2 - (4/2) \times (3/7)^2 = -6/7.$$

As discussed previously, the unrestricted MLE is $\theta = 0$, at which $\mathcal{L}(\theta) = 0$.
Accordingly, the value of the likelihood ratio test statistic would be

$$2|\mathcal{L}(\theta) - \mathcal{L}(\theta^r)| = 12/7 \approx 1.71.$$

The test involves a single restriction, i.e. $m = 1$ and so the appropriate probability for testing is that related to the $\chi^2(1)$, reported in the exercise. Since $1.71 < 3.84$ we therefore accept the null hypothesis that $\theta_2 = \theta_1 + 1$ at the 5% significance level.