Session 5: Computer-based Exercise

Question 1

(a) 40 events are observed among 100 subjects. Assuming a binomial model with parameter π, and using, say, 41 values of π from 0.20 to 0.60 (ie in steps of 0.01), use Stata to plot the log-likelihood ratio (*llr*) function against π.

(b) With the help of suggestions below, use the quadratic function $-(\pi - M)^2/2S^2$ to approximate this curve. What value of M should you choose? Plot the quadratic curves corresponding to different values of S, superimposed on the true log-likelihood ratio curve. For example, for $S = 0.1$ (shown immediately below), using your chosen value of M in place of `M`:

```
gen quad=-(M-pi)^2/(2*0.1^2)
twoway (line llr pi)(line quad pi)
```

[Assuming you called your log-likelihood ratio `llr`.]

[Note that since you do not need to set up separate conditions for the two superimposed plots, you can simplify this command to:

```
twoway (line llr quad pi)
```

and Stata will understand that the first two variables are to be plotted against the third.]

```
drop quad
```

[so that you can repeat for different values of S]

Try a number of values of S, and choose a value for which the quadratic curve is a good approximation to the true log-likelihood ratio.

(c) Compare your chosen value of S with one deduced theoretically.

(d) Visually compare the 95% likelihood ratio confidence interval for π derived from the true loglikelihood with that derived from the quadratic approximation.

Question 2

(a) The Stata command `bloglik` carries out the calculations for the exact binomial log-likelihood and its quadratic approximation. Type `bloglik 40 60` in Stata, and you should get results similar to those in question 1. Note that the default cut-off for the log-likelihood ratio is chosen to be -1:92 = 3.84/2; this produces a likelihood ratio confidence interval (sometimes termed supported range) which corresponds to a 95% confidence interval if the log-likelihood were quadratic.

(b) Use `bloglik` for 4 events out of 10 subjects, 40 out of 100, and 400 out of 1000. What do you conclude?

(c) Now use `bloglik` for 1 event out of 100 subjects, 10 out of 1000, and 100 out of 10000. What do you now conclude? Is there a particular problem with the first of these? What would happen for 99 events out of 100 subjects?

Question 3

(a) Assuming a Poisson model with rate λ for 8 events occurring in 160 person-years of observation, use algebra to obtain the MLE of $\log(\lambda)$ and its standard error.

(b) Use Stata to plot the true log-likelihood ratio against log(λ), and then plot also on the same graph its quadratic approximation, using, say, 81 values of λ between 0.02 and 0.10 (ie in steps of 0.001). Does the quadratic approximation seem appropriate?

(c) From your quadratic approximation, derive the approximate likelihood ratio confidence interval for log(λ), and hence for λ.