ECON 4130
2004 H

Written Paper II (compulsary)

(To be submitted, Monday 25 October, at Ekspedisjonskontoret 12th floor ES.)

Exercise 1

a. Suppose that $X \sim \Gamma(\alpha, \lambda)$. Show that $Y = \lambda X \sim \Gamma(\alpha, 1)$.

b. Problem 6 in Rice chapter 5. Use the method of example 4 in *Lecture Notes II to Rice chapter 5.*

c. Suppose that $X \sim \Gamma(20, 4)$. Calculate, using STATA, $P(X \leq x)$ for x = 3, 4, 5, 6, 7, and compare with the corresponding approximating normal probabilities. [**Hint:** Make a column in STATA consisting of 3,4,5,6,7, and use the functions *gammap* and *norm*. Note that *gammap* only calculates the cdf of $\Gamma(\alpha, 1)$.]

Exercise 2

Some people (particularly gamblers) prefer to think in terms of odds in stead of probabilities to judge the chance of an event. Let $p = P(A)$ be the probability of an event, *A*. The odds for *A* is defined as

(1) $$\theta = \frac{p}{1-p}$$

and can be interpreted as the expected number of wins for each loss in a series of trials of a game. I.e., if *X* is the number of wins in *n* independent games (trials), then $\mathrm{E(No.\ of\ wins)} / \mathrm{E(No.\ of\ losses)} = np/(n(1-p)) = \theta$.

a. One game on a slot machine consists of initially paying a fixed amount, then push a button to make some figures spin before they stop in a random pattern. Certain predetermined patterns lead to a win, small or big, while the rest of the possible patterns lead to loss. Let *p* denote the probability of a win in a single game for a given slot machine (this *p* may vary between machines and is usually unknown for the player). Let *X* be the number of wins in *n* games (trials) on this particular slot machine. Based on the following data we wish to estimate *p* and θ. Data: n = 64 trials gave 8 wins.

Let X_i be 1 or 0 if the *i*-th trial gives a win or a loss respectively. The probability mass function (pmf) for X_i is $P(X_i = x) = f(x \mid p) = p^x(1-x)^x$. Find the maximum likelihood estimators (mle), $\hat{p}, \hat{\theta}$, for p and θ, based on $X_1, X_2, \ldots, X_n$, and calculate the estimates.

b. In the introductory statistics course an approximately $1-\alpha$ confidence interval (CI) for *p* as $\hat{p} \pm z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$ where $z_{\alpha/2}$ is the upper $\alpha/2$ -point in *N*(0, 1). Verify that this is the approximately $1-\alpha$ CI obtained from the large sample mle theory. I.e. find first the so called Fisher information, $I(p) = -\mathrm{E}\left(\frac{\partial^2 \ln f(X_i \mid p)}{\partial p^2}\right)$. Then, use theorem B in section 8.5.2 to show that $\sqrt{n}\frac{(\hat{p}-p)}{\sqrt{p(1-p)}} \xrightarrow[n\to\infty]{D} Z \sim N(0,1)$. Finally, use the fact that $\hat{p}$ is consistent and Slutzky's lemma. Calculate the 95% CI for *p* based on the data.

c. There are usually several ways to approximate things, as illustrated by the following three ways to construct approximate CI's for θ:

Method 1: Since, from **b.,** we have $\sqrt{n}(\hat{p}-p) \xrightarrow[n\to\infty]{D} N(0,\ p(1-p))$, we may use the delta-method from supplementary exercise 6 to develop the asymptotic distribution of $\sqrt{n}(\hat{\theta}-\theta)$ and from there set up the CI for θ using Slutzky. Do this and calculate the CI with degree of confidence 95%.

d. **Method 2:** Since the transformation, $\theta = g(p) = p/(1-p)$, is an increasing function of *p*, we can determine a CI for θ directly from the CI of *p* in **b.** in the following way: Let *L* and *U* be observable r.v.'s such that $P(L \le p \le U) \approx 1-\alpha$. Then the interval (*L*, *U*) is an approximate $1-\alpha$ CI for *p*. Explain why the interval $(g(L), g(U))$, is an approximate $1-\alpha$ CI for θ with exactly the same degree of confidence as (L, U). Calculate this interval for the given data (degree of confidence 95%).

e. **Method 3 (Parametric Bootstrap):** A weakness with asymptotic methods is that they require large samples to apply. Given a finite sample we cannot always be sure that an interpretation of data based on asymptotic approximations is justified. Only in exceptional cases there exist analytical ways to evaluate the exact statistical properties of an estimator based on a finite small sample. Therefore the most common way to obtain insight into the small sample properties of estimators is by simulation techniques (of which the "Bootstrap" family has gained a lot of importance in the later years).

In the binomial case we have the rule of thumb that the approximation to the normal distribution for the binomial variable is considered satisfactory if $np \ge 10$ and $n(1-p) \ge 10$. In our case $n\hat{p} = 8$ which indicates that we may be in a border case. It is to be expected that the CI's in **c.** and **d.** are reasonably well justified, but we cannot be completely sure. To obtain evidence about this issue we arrange a parametric bootstrap

experiment. Based on a simulated sample of estimates of θ, we can calculate a so called bootstrap 95% CI for θ and compare with the two asymptotic CI's. If there is no great difference this can be taken as evidence that the asymptotic methods worked well. If the difference is substantial, it is reasonable to discard the asymptotic CI's and report the bootstrap CI for θ instead.

Read the text between example C and D in Rice section 8.5.3, in addition to example E. Use the method described there to calculate a 95% bootstrap CI for θ based on a bootstrap sample of size $B = 1000$. You then need to generate 1000 observations of a binomial r.v., X, with $n = 64$ trials and success probability $\hat{p}$. For each of the simulated X, calculate the mle estimates of p and θ. Thus you have obtained 1000 simulated observations of the mle's $\hat{p}$ and $\hat{\theta}$, that we may call p_i^*, θ_i^*, $i = 1, 2, \ldots, 1000$. To draw binomial observations is not direct in STATA and you need a small program to do that which you can find in the appendix.

f. Make histograms for both the p_i^*'s and the θ_i^* in **e.** Draw the best fitting normal density in both histograms. Which of the two distributions seems closest to a normal distribution?

Appendix to exercise 2

Suppose we want 1000 observations of $X \sim$ bin(64, 0.125). The following small do-file (can be written directly into STAT by the do-editor or read into STATA from an ASCII-file from the do-editor), produces a STATA data file containing the 1000 observations. I have called the data file, "bindat", here but you can choose your own name of course. The file is stored under the name, bindat.dta. For calculation of the p_i^*'s and the θ_i^*'s this file must be read into STATA by the use-command (e.g. use bindat).

```
capture program drop binsim
program define binsim
      tempname sim
      postfile `sim' x using bindat, replace
      quietly {
            local i=1
            while `i' <= 1000 {
                  drop _all
                  set obs 64
                  gen z=sum(uniform()<=.125)
                  post `sim' (z[_N])
                  local i=`i'+1
                        }
                  }
      postclose `sim'
      clear
end
```

Notes: The first line makes it possible to edit the program binsim and then run it again without STATA protesting. The single quotation mark ` I find on the top of my backslash (\) key. The closing single quotation mark, ‘ , I find under * on the *-key.

If z is a variable with numbers $z_1, z_2, \ldots, z_k$, using the function *sum(z*), creates a new column with the cumulative sums, z_1, $z_1 + z_2$, $z_1 + z_2 + z_3, \ldots$ The total sum of the z_i's is found as the last element of this column. Note that $\_N$ gives the number of observations in the current dataset, so $z[\_N]$ refers to the total sum.

1. Read the lines into the do-editor.
2. Run the do-file by pushing the run-key in the do editor (this only reads the program *binsim* into STATA but does not execute it).
3. Then run the program by writing *binsim* in the command window.
4. Then read into STATA the simulated data by the command, *use bindat*.