MAST90105: Lab and Workshop Problems for Week 11

The Lab and Workshop this week covers problems arising from Modules 7 and 8.

1 Lab

1. How good are confidence intervals? If we repeat the experiment a large number of times we expect 95% of the confidence intervals for contain the parameter values. We can check this using simulations. Enter the following commands:

```
x = t.test(rnorm(10))
x
names(x)
x$conf.int
```

   You should use the help function or your tutor to understand what each command does. Note that rnorm simulates values from $N(0,1)$ so we know the true mean is zero. Then automate the process

```
 f=function(t){x=t.test(rnorm(t));as.vector(x$conf.int)};
f(10);
 f(20);
t <- as.matrix(rep(10,100));
C <- t(apply(t,1,f)); #this is a trick so we don't have to program
 matplot(C,type="l");#a matrix plot
abline(0,0)#includes a line at 0
```

   Each column of the matrix C is the lower and upper bounds of a 95% confidence interval. From your plot determine how many of these intervals contain the true mean zero. Is it close to 95%? You can check as follows:

```
num = (C[, 1] < 0) & (C[, 2] > 0)
sum(num)/nrow(C)
```

2. In class, we discussed the Newspoll outcomes from March 20 and April 3 2017. The March 20 poll reported that 675 of 1824 voters would vote first for the Government if an election were held then, and on April 3 it was 615 out of 1708 voters.

   a. Starting with a uniform distribution over (0,1), find the posterior distribution for the population proportion after the March 20 .

   b. Use this posterior as a prior distribution for the April 3 Newspoll and find the resulting posterior distribution.

c. Plot this density with the posterior density obtained in lectures from a uniform prior.

d. Find a 95% posterior probability interval from your posterior distribution and compare this to the one from lectures.

e. Construct a Beta distribution as a prior for the data that arrived on April 3 based on your Bayes estimates from the previous poll so that there is 99% probability that the true proportion is less than (a) 50% (b) 40%. Compute the posterior in each case.

2 Workshop

3. Let $X \sim \text{binomial}(1, p)$ and let $X_1, \ldots, X_{10}$ be a random sample of size 10. Consider a test of $H_0 : p = 0.5$ against $H_1 : p = 0.25$. Let $Y = \sum_{i=1}^{10} X_i$. Define the critical region as $C = \{y : y < 3.5\}$.

   a. Find the value of α the probability of a Type I error. Do not use a normal approximation. (Hint: Use pbinom).

   b. Find the value of β, the probability of a Type II error. Do not use a normal approximation.

   c. Simulate 200 observations on Y when $p = 0.5$. Find the proportion of cases when H_0 was rejected. Is this close to α?

   d. Simulate 200 observations on Y when $p = 0.25$. Find the proportion of cases when H_0 was not rejected. Is this close to β?

4. A ball is drawn from one of two bowls. Bowl A contains 100 red balls and 200 white balls; Bowl B contains 200 red balls and 100 white balls. Let p denote the probability of drawing a red ball from the bowl. Then p is unknown as we don't know which bowl is being used.To test the simple null hypothesis $H_0 : p = 1/3$ against the simple alternative that $p = 2/3$, three balls are drawn at random with replacement from the selected bowl. Let X be the number of red balls drawn. Let the critical region be $C = \{x : x = 2, 3\}$. Using R, what are the probabilities α and β respectively of Type I and Type II errors?

5. Let $Y \sim \text{binomial}(100, p)$. To test $H_0 : p = 0.08$ against $H_1 : p < 0.08$, we reject H_0 and accept H_1 if and only if $Y \leq 6$. Using R,

   a. Determine the significance level α of the test.

   b. Find the probability of a Type II error if in fact $p = 0.04$.

6. Let p be the probability a tennis player's first serve is good. The player takes lessons to increase p. After the lessons he wishes to test the null hypothesis $H_0 : p = 0.4$ against the alternative $H_1 : p > 0.4$. Let y be the number out of $n = 25$ serves that are good, and let the critical region be defined by $C = \{y : y \geq 13\}$.

a. Define the power function to be $K(p) = P(Y \geq 13; p)$. Graph this function for $0 < p < 1$.

b. Find the value of $\alpha = K(0.40)$

c. Find the value of β when $p = 0.6$, $(\beta = 1 - K(0.6))$

7. Let $X_1, \ldots, X_{10}$ be a random sample of size $n = 10$ from a distribution with p.d.f. $f(x;\theta) = \exp(-(x-\theta))$, $\theta \leq x < \infty$.

   a. Show that $Y_1 = \min(X_i)$ is the maximum likelihood estimator of θ.

   b. Find the p.d.f. of Y_1 and show that $E(Y_1) = \theta + 1/10$ so that $Y_1 - 1/10$ is an unbiased estimator of θ.

   c. Compute $P(\theta \leq Y_1 \leq \theta + c)$ and use this to construct a 95% confidence interval for θ.

8. A random variable X is said to have a Pareto distribution with parameters, x_0 and β, if its cdf is

$$F_X(x) = \begin{cases} 1 - \left(\frac{x_0}{x}\right)^\beta & x > x_0 \\ 0 & x \leq x_0 \end{cases}$$

   a. What is the pdf of X?

   b. Suppose $U_1, \cdots, U_n$ are a random sample from the uniform distribution on $(0, X)$ where X is the unknown parameter. Suppose that X has a Pareto prior distribution with parameters x_0, β. Calculate the posterior distribution of X. (Hint: Consider carefully the values of the posterior pdf which are strictly positive, noting that both the joint distribution of the sample and the prior distribution pdf's have to be positive.)

   c. Find a $100(1-\alpha)$ % posterior probability interval for X.

9. If a newborn baby has a birth weight that is less than 2500 grams we say the baby has a low birth weight. The proportion of babies with birth weight is an indicator of nutrition for the mothers. In the USA approximately 7% of babies have a low birth weight. Let p be the proportion of babies born in the Sudan with low birth weight. Test the null hypothesis $H_0 : p = 0.07$ against the alternative $H_1 : p > 0.07$. If $y = 23$ babies out of a random sample of $n = 209$ babies had low birth weight, , using a suitable approximation, what is your conclusion at the significance levels

   a. $\alpha = 0.05$?

   b. $\alpha = 0.01$?

   c. Find the p-value of this test. (Recall the p-value is the probability of the observed value or something more extreme when the null hypothesis is true).

   Helpful R output

```r
qnorm(c(0.95, 0.99))

## [1] 1.644854 2.326348

pnorm(2.269)

## [1] 0.9883658
```

10. Let p_m and p_f be the respective proportions of male and female white crowned sparrows that return to their hatching site. Give the endpoints for a 95% confidence interval for $p_m - p_f$, given that 124 out of 894 males and 70 out of 700 females returned. (*The Condor*, 1992 pp.117-133.). Does this agree with the conclusion of the test of $H_0 : p_m = p_f$ against $H_1 : p_m \neq p_f$ with $\alpha = 0.05$?