MATH11400 Statistics 1 2010–11

Homepage http://www.stats.bris.ac.uk/%7Emapjg/Teach/Stats1/

Problem Sheet 6

*1. Let $X_1, \ldots, X_n$ be a random sample from the Uniform$(0, \theta)$ distribution, for which the population median is $\tau = \theta/2$.

(a) The method of moments estimate of τ is $\bar{X}$. Find $\mathrm{E}(\bar{X})$ and $\mathrm{Var}(\bar{X})$, and hence show that the method of moments estimator is unbiased as an estimator of the population median τ and has mean square error $\theta^2/12n$.

(b) The maximum likelihood estimate of θ is $Y = \max\{X_1, \ldots, X_n\}$. You are given that Y has probability density function $f_Y(y; \theta) = ny^{n-1}/\theta^n$ for $0 < y < \theta$ (and $f_Y(y; \theta) = 0$ otherwise). Show that $\mathrm{E}(Y) = n\theta/(n + 1)$.

The maximum likelihood estimator of the population median $\tau = \theta/2$ is $\hat{\tau}_{\text{mle}} = \max\{X_1, \ldots, X_n\}/2$. Use the results for Y to show that $\hat{\tau}_{\text{mle}}$ has bias $-\theta/2(n + 1)$.

*2. The methods and **R** commands required for this question are similar to those in the handout (section 5.7/5.8) for the simulation from the Uniform(0,1) distribution, but with the obvious adjustments to the names and to the formulae for the various estimators. You may find it helpful to consult the handout before answering the question.

(a) Use the **R** commands below to construct a matrix `xsamples` with 1000 rows and 10 columns. The 10 data values in each row can be thought of as a random sample of size 10 from an Exp(θ) distribution with rate $\theta = 1$ and the 1000 rows can be thought of as $B = 1000$ independent repeated samples.

```
> xvalues <- rexp(10000,rate=1)
> xsamples <- matrix(xvalues,nrow=1000)
```

(b) Calculate the maximum likelihood estimate $\hat{\theta} = 1/\bar{x}$ for each sample, and plot a histogram of the relative frequencies of the 1000 estimates of θ obtained.

(c) You may *assume* that the Exp(θ) distribution has median $\log(2)/\theta$. For each of the $B = 1000$ samples in part (a) above, calculate both the sample median and the maximum likelihood estimate of the population median $\tau(\theta) = \log(2)/\theta$.

Produce a single plot containing a boxplot of the 1000 values of the sample median and a boxplot of the 1000 values of the mle of the population median.

(d) Since the observations above were drawn from a population distribution with $\theta = 1$, add a horizontal line at height $\log(2)$ to your plot showing the true value of the median for this population and use it to compare the sample median and the mle as estimators of the population median.

(e) Calculate the mean and the variance of the 1000 values of the sample median and the 1000 values of the maximum likelihood estimate of the population median and use these numerical values to compare the bias, variance and mean square error of the two estimators.

3. Let T be the total number of heads obtained when a fair coin is tossed 10 times. Let $A = \{T \leq 1\}$ be the event that at most one head is obtained and let $B = \{T \geq 6\}$ be the event that at least 6 heads are obtained.

   (a) Calculate the exact values of $P(A)$ and $P(B)$.

   (b) Calculate the approximate values of $P(A)$ and $P(B)$ given by applying the central limit theorem without any continuity correction.

   (c) Calculate the approximate values of $P(A)$ and $P(B)$ given by applying the central limit theorem, but now with a continuity correction.

   (d) Comment on the accuracy of the approximations in (b) and (c).

*4. An architect is designing the car park for a new apartment block, which has 200 apartments. She believes that the residents in 20% of the apartments will require 2 parking places, that 70% will require 1 place, and that the remaining 10% will not have a car.

   (a) Let X be the number of parking places required by the residents of a randomly chosen apartment. Find the mean and variance of X.

   (b) If the architect provides 230 car parking places for residents, what is the probability that this will not be enough? How many places would she need to provide for there to be a 99% chance that there will be enough places to satisfy all the residents' demands?

5. Opinion polls indicate that support for the government has been running at about 37% of the electorate, but it is thought that this may have changed in the light of recent events. Assume a random sample of n electors is interviewed. Let $X_i = 1$ if the ith elector sampled supports the government and let $X_i = 0$ otherwise, so that $T_n = X_1 + \cdots + X_n$ is the total number in the sample that say they support the government. Assume throughout that $X_1, X_2, X_3, \ldots$ are independent random variables and that $P(X_i = 1) = 0.37 = 1 - P(X_i = 0)$.

   Assume the pollsters take a sample of size $n = 1500$. Use the central limit theorem to find the probability that the proportion in the sample supporting the government will differ from 0.37 by no more than 0.02, i.e. find $P(|T_n/n - 0.37| \leq 0.02)$.

6. At one time, a common method of simulating an observation from a $N(0, 1)$ distribution was to simulate a random sample $U_1, \ldots, U_{12}$ of size 12 from a U$(0, 1)$ distribution and set $Z \equiv (U_1 + \cdots + U_{12}) - 6$. The central limit theorem says that, for large sample sizes, the sum of the observations will have a Normal distribution. Here the sample size is only 12, but nevertheless we will see that the distribution of Z is very close to $N(0, 1)$.

   (a) Show that, for Z defined above, $\mathrm{E}(Z) = 0$ and $\mathrm{Var}(Z) = 1$.

   (b) Using the methods of Section 5 of your notes, generate 1000 random samples, each of size 12, from a U$(0, 1)$ distribution and calculate the sample value of Z for each sample. Plot the histogram of relative frequencies for the Z values (you will need to specify `probability = T` in the `hist` command). Add a graph of the $N(0, 1)$ density over the range $(-3, 3)$ to your plot, with the commands

   ```
   range <- seq(-3,3,0.01); lines(range, dnorm(range))
   ```

   and visually compare the fit of the histogram to the normal density. Alternatively, you can make the comparison with a Q-Q plot.