Copputer Intensive Statistical Methods (MSA100/MVE185)

Petter Mostad (mostad@math.chalmers.se)

October 2, 2008

1. We will look at the data presented in Examples 1.2 and 2.8 in the material copied from Davison: Bootstrap Methods and their Application. This exercise has also been adapted from Davison.

   The data represent the number of people living in 10 American cities, in 1920 (variable u) and 1930 (variable x). The data is part of the R package *boot*, which is already downloaded in all new releases of R. To read more about it, do, in R,

   ```
   > library(boot)
   > help(city)
   ```

   The data is imagined to be a sample from a bivariate distribution for (u, x). The statistic we want to investigate is (as in the example) $\overline{u}/\overline{x}$, where $\overline{u}$ and $\overline{x}$ represent the averages of the sampled u and x values, respectively.

   (a) Use R to find the parameters of a logbinormal distribution fitting the data. (In other words, find a binormal distribution fitting the logarithms of the data values).

   (b) Write an R function of type

   ```
   city.sim <- function(city, mle) { ... }
   ```

   that simulates from the logbinormal distribution above: The "mle" argument should be the parameters for the logbinormal distribution found above, and the "city" argument should only be used for counting its rows, so that the output of the function is an object with the same form as the city object, but where the rows are simulated logbinormal values.

   (c) Write an R function

   ```
   city.fun <- function(data)
   ```

   which, for data of the same type as "city" or as the objects simulated by "city.sim", produces a vector consisting of the statistic $\overline{u}/\overline{x}$ and its variance estimate (see page 24 of the material copied from Davies).

(d) Do a parametric bootstrap for this statistic, using a command like

```
city.para <- boot(city, city.fun, R = 999,
sim=''parametric'', ran.gen=city.sim, mle=city.mle)
```

Look at the help function for *boot* to get this to work, and to understand the result. Investigate the result using commands like

```
print(city.para)
plot(city.par)
names(city.para)
```

(e) Compare histograms of the logarithms of the simulated statistics with Figure 2.5 from Davies. Investigate the normality of these simulated values. Compute basic bootstrap confidence limits using Equation (2.10). Repeat with various values of R to see the effect of changing the number of simulated values.

2. We now look at the data contained in the object "co.transfer"; you may read about it using the command "help(co.transfer)". The statistic we would like to investigate is the mean of the differences between the values "week" and the values "entry"; we are studying whether the hospital stay has changed the carbon monoxide transfer factor.

   (a) Write an R function of type

   ```
   co.fun <- function(data, i) {...}
   ```

   where "data" is the same kind of object as co.transfer, and "i" is a vector of possible indices for the rows of "data". The output should be a vector containing the statistic mentioned above, and an estimate for its variance.

   (b) Do a bootstrap in this situation, with a command like

   ```
   co.boot <- boot(co.transfer, co.fun, R=999)
   ```

   Display the output of the bootstrap as above, and compare with an analysis where it is assumed that the differences "week minus entry" are normally distributed (use the function "t.test").