Stat 425 HW3

Fritz Scholz

1. Describe in words what the following function does, in particular the distinction between `flag=T` and `flag=F`.

```
Initialize=function(flag=T,Nsim=10000){
if(flag==T){
   out=NULL}else{
   out=rep(0,Nsim)
}
for(i in 1:Nsim){
   out[i]=sum(rnorm(20))
}
}
```

Import this function into your R workspace (cut and paste should do). Time this function by executing the command

$$\texttt{system.time(Initialize(flag} = \texttt{F,Nsim} = \texttt{10000))}$$

and do the same with `flag=T`. Repeat this with `Nsim=20000, 30000, 50000`. Plot the vector `y` of the respective execution times for `flag = T` against the `x` vector holding the corresponding `Nsim` values. You can add a second set of points with coordinate vectors `x1` and `y1` to an existing plot by `points(x1,y1)`. Thus you can plot the points corresponding to `flag=T` and `flag=F` on the same graph.

In another graph plot the square root (`sqrt(y)`) of the execution times obtained for `flag=T` against `x`. You can fit a line to a plot pattern created by plot(x,y) by doing out=lsfit(x,y), then out$coef gives you a vector of intercept and slope. The command abline(out) would add the fitted line to your plot. Only fit a line to those point patterns that look approximately linear.

Project how long you would have to wait to run each (`flag=T` and `flag=F`) for `Nsim=100000, 1000000`. What have you learned from this, in particular with respect to calculating estimated p-values?

You can add R plots to your Word (or free Open Office Writer) document by activating the graphics window in R that shows the plot (i.e., click on it), go to File then Copy to the Clipboard, then choose as Metafile and then do a CTRL V in your Word (or free Open Office Writer) document.

2. Write a function `Problem2=function(x,y,Nsim=30000)`$\{\ldots\}$ that computes the p-value for the two-sided Wilcoxon rank-sum test of the hypothesis of no treatment effect as indicated on slides 70-71 (Chapter 1). Write this as a function of two response vectors `x` and `y` (control and treatment) and of `Nsim` (see below).
Allow for possible ties and obtain an exact p-value by using `combn` as long as $\binom{m+n}{n}$ is not too large, say $\leq \texttt{Nsim} = 300000$. When the full enumeration exceeds `Nsim`, let the function estimate the p-value by sampling `Nsim` splits of the mid-rank vector and evaluating the treatment rank-sum each time.

In addition to the exact or estimated p-value this function should also produce the corresponding p-value obtained by normal approximation. Use the continuity correction if there are no ties, otherwise don't use it. You have no ties when `length(c(x,y))==length(unique(c(x,y))`.

Apply this test to the two vectors

```
x0=c(91, 94, 99, 99, 99, 100, 101, 102, 105, 106, 108)
```

```
y0=c(97, 102, 103, 104, 105, 107, 108, 111, 117)
```

Now run `wilcox.test(x0,y0)`. How do the results compare?

Repeat this with `wilcox.test(x1,y1)` and `Problem2(x1,y1,Nsim=300000)`, where

```
x1=c(94.1, 96.4, 101.55, 103.6, 100.51, 97.4, 110.69, 106.13, 101.61, 97.3, 100.33)
```

```
y1=c(104.59, 101.87, 105.84, 105.31, 108.72, 95.86, 102.35, 104.68, 102.21)
```