# Stat 425 HW4

Fritz Scholz

1. The purpose of this homework is to understand the power behavior of the two-sample Wilcoxon test when sampling from normal populations which may differ from each other by a shift parameter $\Delta$, i.e., $\mathcal{N}(\mu,\sigma^2)$ and $\mathcal{N}(\mu+\Delta,\sigma^2)$, respectively.
In particular, we want to compare the power function of the rank-sum test against that of the two-sample $t$-test. We also want to understand to what extent the asymptotic relative efficiency (ARE) $e_{W,t} = 3/\pi$ is reflected for finite sample sizes $m$ and $n$. We want to use both normal approximations for the power function and explore their quality in relation to $m$ and $n$.
This exercise offers opportunity for extra credit (to make up for previous losses) by extending the breadth of your investigation (other $\alpha, m$ and $n$). Provide your function codes, plots and a narrative that explains coherently what you have learned.
First note that the ranks of samples

$$X_1,\ldots,X_m \sim \mathcal{N}(\mu,\sigma^2) \qquad \text{and} \qquad Y_1,\ldots,Y_n \sim \mathcal{N}(\mu+\Delta,\sigma^2)$$

are the same as the ranks of the transformed samples

$$X_i' = (X_i-\mu)/\sigma \sim \mathcal{N}(0,1),\ i=1,\ldots,m \quad \text{and} \quad Y_j' = (Y_j-\mu)/\sigma \sim \mathcal{N}(\Delta/\sigma,1) = \mathcal{N}(\Delta',1),\ j=1,\ldots,n$$

since the common transformation $(\cdot-\mu)/\sigma$ does not alter the joint order relationships among $X$'s and $Y$'s. Hence the distribution of the rank-sum is the same, whether we sample from $\mathcal{N}(\mu,\sigma^2)$ and $\mathcal{N}(\mu+\Delta,\sigma^2)$ or from $\mathcal{N}(0,1)$ and $\mathcal{N}(\Delta',1)$ with $\Delta'=\Delta/\sigma$. Thus the power of the rank-sum test does not depend on $\mu$ and it depends on $\Delta$ and $\sigma$ only through the ratio $\Delta'=\Delta/\sigma$. A corresponding property holds for the two-sample $t$-test, namely its power depends on $\mu$, $\Delta$ and $\sigma$ only through $\Delta'=\Delta/\sigma$. Note however, that in both cases (Wilcoxon and $t$-test) the sample sizes $m$ and $n$ affect the power.

Write a function `Ranksum.sim=function(m=10,n=10,alpha=.05,Nsim=10000,Delta.p=.5){...}` that simulates the distribution of the Wilcoxon rank-sum statistic $W_s$ for samples of sizes $m$ and $n$ from $\mathcal{N}(0,1)$ and $\mathcal{N}(\Delta',1)$, respectively ($\Delta' \equiv$ `Delta.p`). By distribution is meant a vector `Ws.vec` of length `Nsim`, containing the results from calculating the rank-sums $W_s$ for `Nsim` simulations of independent samples of sizes `m` and `n` from $\mathcal{N}(0,1)$ and $\mathcal{N}(\Delta',1)$, respectively. Run these simulations in a loop (`for(i in 1:Nsim){...}` ) with appropriate initialization of `Ws.vec` (remember HW3).
We consider one-sided rank-sum tests which reject $H_0:\Delta=0$ whenever $W_s \geq c_\alpha$, where $c_\alpha$ is the lowest integer value such that $P_{H_0}(W_s \geq c_\alpha) \leq \alpha$. To find the appropriate $c_\alpha$ you may use `qwilcox` but understand that `qwilcox(p,m,n)` returns the smallest $L$ such that $P_{H_0}(W_{XY} \leq L) \geq p$ and realize the appropriate relationship between $W_{XY}$ and $W_s$. Explain your reasoning in coming up with $c_\alpha$.
`Ranksum.sim` should produce a named vector[1] with components representing

$$\texttt{Nsim},\ \texttt{m},\ \texttt{n},\ \alpha,\ c_\alpha,\ \alpha_c,\ \Delta',\ P_{\Delta'}(W_s \geq c_\alpha)$$

[1]For example, you name a vector `out = c(x,y,z)` via `names(out) = c("x.name","name.y","z")`.

where $\alpha_c = P_0(W_s \geq c_\alpha)$ is the achieved significance level ($\leq \alpha$) when using $c = c_\alpha$ as critical point. $P_{\Delta'}(W_s \geq c_\alpha)$ represents the power of the test at the alternative $\Delta'$, the quantity of main interest to us. While building this function use `Nsim` $= 100$ for faster debugging.
As a check run `Ranksum.sim` for `Nsim` $= 10000$ and $\Delta' = 0$. Your power should then be close to the achieved significance level $\alpha_c$, which of course depends on $m$ and $n$ though `qwilcox`.
Next, write a function

$$\texttt{power.fun} = \texttt{function(Nsim} = 10000, \texttt{alpha} = .05, \texttt{m} = 10, \texttt{n} = 10, \texttt{fac} = 3/\texttt{pi})\{\ldots\}$$

that evaluates `Ranksum.sim` for `Delta.p` in `Delta.vec` = `seq(0,2,length.out = 21)` and then plots $P_{\Delta'}(W_s \geq c_\alpha)$ against `Delta.p` over the grid vector `Delta.vec`. In a loop store the calculated values of $P_{\Delta'}(W_s \geq c_\alpha)$ in a vector `power.vec` of same length as `Delta.vec`. Superimposed on this plot

```
plot(Delta.vec,power,type="l",xlab=expression(Delta*minute==Delta/sigma),
              ylab=expression(Pi(Delta*minute)==Pi(Delta/sigma)),ylim=c(0,1))
```

add the power function of the two-sample $t$-test, evaluated over the same grid. Do this by using the `lines(x,y)` command for appropriate vectors `x` and `y`. The power function values for the $t$-test can be obtained in vectorized mode (since we use the vector argument `Delta.vec`) via

$$\texttt{power.t} = 1 - \texttt{pt(qt(}1 - \texttt{alpha.c,m} + \texttt{n} - 2, 0), \texttt{m} + \texttt{n} - 2, \texttt{Delta.vec/sqrt(}1/\texttt{m} + 1/\texttt{n}))$$

Explain this last command in terms of the fact that the distribution of the two-sample $t$-statistics is a noncentral $t$-distribution with $m+n-2$ degrees of freedom and noncentrality parameter

$$\delta = \frac{\Delta'}{\sqrt{1/m + 1/n}} = \frac{\Delta}{\sigma\sqrt{1/m + 1/n}}.$$

We expect the power of the $t$-test to be slightly higher than the power of the Wilcoxon rank-sum test. To get a better match of the power functions recompute the power of the $t$-test when $m$ and $n$ are reduced by the factor `fac` $= 3/$`pi` $= 3/\pi$, which represents the ARE of the Wilcoxon test relative to the $t$-test. This adjustment (only for the power of the $t$-test) is possible since `pt` and `qt` allow non-integer degrees of freedom. However, non-integer sample sizes don't make sense in the application of the $t$-test. What can you say about the quality of the match-up? Note that you can make both comparisons by using `fac` $= 1$ and `fac` $= 3/$`pi` in the argument sequence to `power.fun`.
In spite of the quality of the match-up what aspect makes the rank-sum test preferable? Does the above match-up of power suggest a way to plan the sample sizes for the rank-sum test when dealing with normal shift alternatives (without simulating the $W_s$ distribution for $\Delta$)?
Now add to this plot the power as computed by the two normal approximations and add a legend in the upper left corner using the `legend(...)` command, e.g.,

```
legend(0,1,c("simulated power of Ws",
      paste("non-central t power (fac =",round(fac,3),")"),
     "power: normal approx. 1","power: normal approx. 2"),
      col=c("black","blue","red","orange"),lty=1:4,bty="n")
```

Make sure the various `lines(...)` commands use the appropriate colors and `lty` parameters. Also add the following annotation to your plot.

```
text(max(Delta.vec),0,substitute(N[sim]==xNsim~", "~m ==xm~", "
  ~n ==xn~", "~alpha==xalpha~", "~alpha[c]==xalpha.c,
  list(xNsim=Nsim,xm=m,xn=n,xalpha=alpha,xalpha.c=round(alpha.c,3))),adj=1)
```

Show these plots for $m = 10$ and $n = 10$ and $\alpha = .05$. Discuss your results.

Note that I have given you two instances of writing mathematical expressions (Greek) in your plot via `expression` and `substitute`. For more on this see the link to "An Approach to Providing Mathematical Annotation in Plots" by Paul Murrell and Ross Ihaka that I provided on the class web page.

Optional (no need to do all): While the plots should look fine for $m = n = 10$ they could use some scaling improvement for $m = n = 5$ or $m = n = 30$. Try to implement this in an automatic fashion by using the second normal approximation to find an appropriate $U$ (corresponding to approximate power .99) to get an adjustable grid vector `Delta.vec=seq(0,U,length.out=21)`. What about two-sided tests? What would you have to change if you were to compare the power of $t$-test and Wilcoxon test for non-normal shift alternatives. Note that the non-central $t$-distribution no longer applies.