STAT 824 sp 2023 Lec 12 slides

Wilcoxon rank-sum test

Karl B. Gregory

University of South Carolina

These slides are an instructional aid; their sole purpose is to display, during the lecture, definitions, plots, results, etc. which take too much time to write by hand on the blackboard. They are not intended to explain or expound on any material.

Table of Contents

Two-sample. Control vs Treatment

Does treatment increase the values?

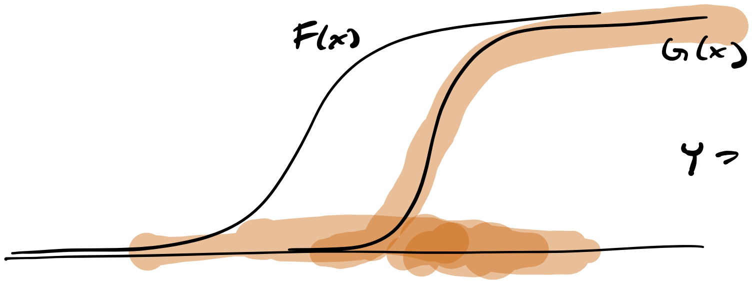
$X_1, \dots, X_m \overset{ind}{\sim} F$ "control"

$Y_1, \dots, Y_n \overset{ind}{\sim} G$ "treatment"

$H_0: F = G$

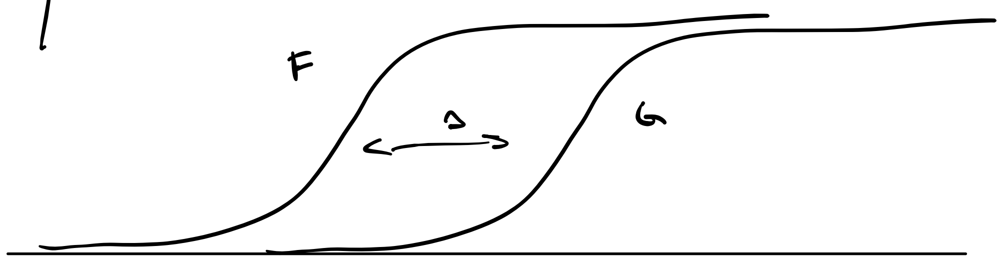
$\boxed{H_1: \text{Y's tend to be greater than X's.}}$

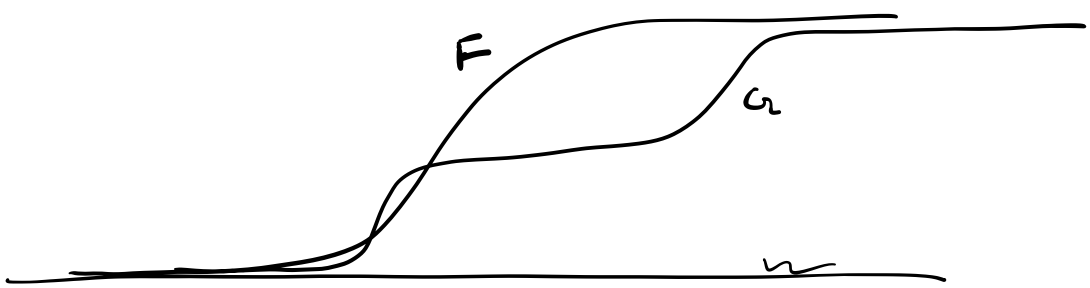
<u>**Stochastically greater:**</u> $F(x) > G(x) \quad \forall x$, Y is stochastically greater than X.



$Y > X$ with prob.

<u>**Location shift.**</u> $G(x) = F(x - \Delta)$





Suppose we collect random samples from "control" and "treatment" populations:

$$X_1, \dots, X_m \overset{\text{ind}}{\sim} F \qquad \text{"control"}$$

$$Y_1, \dots, Y_n \overset{\text{ind}}{\sim} G \qquad \text{"treatment"}$$

We wish to test for treatment effectiveness (are Y's bigger than X's?).

Wilcoxon rank sum test (quintessential nonparametric test)

The *Wilcoxon rank sum test* concludes a "positive treatment effect" if

counts the # pairs (x, y) such that x < y.

$$W_{XY} = \boxed{\sum_{i=1}^{m} \sum_{j=1}^{n} \mathbf{1}(X_i < Y_j)} \geq c,$$

where c can be calibrated to control the Type I error rate.

Can modify to find a "negative" or "either direction" treatment effect.

If $G = F$, the (null) distribution of W_{XY} is the same for any continuous F.

Rank-sum form of Wilcoxon rank sum statistic

An alternate way of computing W_{XY}:

1. Sort all the data $(X_1, \dots, X_m, \boxed{Y_1, \dots, Y_n})$.
2. Obtain the ranks.

   1 2 .. n+m
3. Keep the ranks corresponding to $Y_1, \dots, Y_n$, calling these $\boxed{S_1, \dots, S_n.}$

Then $W_{XY} = \underbrace{S_1 + \cdots + S_n}_{} - \boxed{n(n+1)/2.}$

Let $W_S = S_1 + \cdots + S_n$.

Exercise: Show that $W_{XY} = W_S - n(n+1)/2$.

Minimum value of W_{XY}? Can be 0.

$$W_{XY} = \sum_{i=1}^{m} \sum_{j=1}^{n} \mathbb{1}(X_i < Y_j)$$

$$= \sum_{j=1}^{n} \Big[\underbrace{\sum_{i=1}^{m} \mathbb{1}(X_i < Y_j) + \sum_{j'=1}^{n} \mathbb{1}(Y_{j'} \le Y_j)}_{S_j} - \sum_{j'=1}^{n} \mathbb{1}(Y_{j'} \le Y_j) \Big]$$

$$= \sum_{j=1}^{n} \Big[S_j - \sum_{j'=1}^{n} \mathbb{1}(Y_{j'} \le Y_{(j)}) \Big]$$

$$= S_1 + \dots + S_n - \sum_{j=1}^{n} \underbrace{\sum_{j'=1}^{n} \mathbb{1}(Y_{j'} \le Y_{(j)})}_{j}$$

$$= S_1 + \dots + S_n - \sum_{j=1}^{n} j$$

$$= S_1 + \dots + S_n - \frac{n(n+1)}{2}$$

$$= W_S - \frac{n(n+1)}{2}.$$

Mann-Whitney

S_1 S_2 S_n

X Y X X Y X X ... Y

1 2 3 4 ... N

$\binom{N}{n}$

Theorem

$F=G$

Let $X_1, \dots, X_m, Y_1, \dots, Y_n$ *be continuous iid rvs and set* $N = n + m$. *Then*

sets of n ranks

$$P(S_1 = s_1, \dots, S_n = s_n) = \frac{1}{\binom{N}{n}} \quad \text{for all } \cancel{n\text{-tuples}}\ \{s_1, \dots, s_n\} \subset \{1, \dots, N\}.$$

$P(\{S_1, \dots, S_n\} = \{s_1, \dots, s_n\})$

Exercise: Tabulate the null distribution of W_{XY} under $\boxed{N = 5,\ n = 2.}$

$\binom{N}{n} = \binom{5}{2} = 10$

$N = 5$ $\quad n = 2$ $\quad Y_1\ Y_2$ $\quad X_1\ X_2\ X_3$

$W_{XY} = S_1 + \dots + S_n - \frac{1}{2}n(n+1)$

$= S_1 + S_2 - 3$

$Y_1\ Y_2\ X_1\ X_2\ X_3$

$Y_1\ X_1\ Y_2\ X_2\ X_3$

s_1	s_2	W_S	W_{XY}
1	2	3	0
1	3	4	1
1	4	5	2
1	5	6	3
2	3	5	2
2	4	6	3
2	5	7	4
3	4	7	4
3	5	8	5
4	5	9	6

w	0	1	2	3	4	5	6
$P(W_{XY}=w)$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{2}{10}$	$\frac{2}{10}$	$\frac{1}{10}$	$\frac{1}{10}$

$\alpha = 0.05$

Say, reject H_0 when $W_{XY} \geq 6$.

Gives size $= \frac{1}{10}$

$$P\left(W_{XY} \geq W_{XY}^{obs}\right) = 1 - P\left(W_{XY} < W_{XY}^{obs}\right) = 1 - \text{pwilcox}$$
$$= 1 - P\left(W_{XY} \leq W_{XY}^{obs} - 1\right)$$

Example code:

```
m <- 25
n <- 20
X <- rnorm(m,1,1)
Y <- rnorm(n,1,1)
N <- n + m

U <- c(X,Y)
id <- c(rep(1,m),rep(2,n))
id_ord <- id[order(U)]
S <- c(1:N)[which(id_ord == 2)]
Ws <- sum(S)
Wxy <- Ws - n*(n+1)/2

# note that n and m are interchangeable:
1 - pwilcox(Wxy-1,m = n,n = m) # these are slow for large n, m
1 - pwilcox(Wxy-1,m = m,n = n)

EWs <- n*(N+1)/2
VarWs <- n*(N-n)*(N+1)/12
1 - pnorm(Ws,EWs,sqrt(VarWs))
1 - pnorm(Ws,EWs + 1/2,sqrt(VarWs)) # with continuity correction

wilcox.test(x = Y, y = X, alternative = "greater")
```

On computing the exact distribution of W_{XY} when N and n are large...

Theorem (Asymptotic Normality of rank sum under the null)

Under H_0: $F = G$ we have

$S_1 + \dots + S_n$

$$\frac{W_S - \mathbb{E}W_S}{\sqrt{\operatorname{Var} W_s}} \xrightarrow{D} \text{Normal}(0,1)$$

as $N \to \infty$, provided $n \to \infty$ and $N - n \to \infty$.

Exercise: Show $\mathbb{E}W_S = \boxed{\frac{1}{2}n(N+1)}$ and $\operatorname{Var} W_S = \frac{1}{12}n(N-n)(N+1)$.

$$\mathbb{E} W_s = \mathbb{E} \sum_{j=1}^{n} S_j \qquad P(S_1 = s_1 \cap \dots \cap S_n = s_n) = \frac{1}{\binom{N}{n}}$$

$$= \sum_{j=1}^{n} \mathbb{E} \overset{Y_j}{S_j} \qquad \boxed{P(S_1 = s_1)} = P(Y_1 \text{ in position } s_1) = \frac{1}{N}$$

$$= n \, \mathbb{E} S_1$$

$$= n \sum_{s_1=1}^{N} s_1 \cdot P(S_1 = s_1)$$

$$= n \sum_{s_1=1}^{N} s_1 \frac{1}{N}$$

$$= \frac{n}{N} \frac{N(N+1)}{2}$$

$$= \frac{1}{2} n (N+1)$$

$$\mathrm{Var}\, W_s = \mathrm{Var}\left(\sum_{j=1}^{n} S_j \right)$$

$$= \sum_{j=1}^{n} \mathrm{Var}\, S_j + \sum_{j \neq j'} \mathrm{Cov}(S_j, S_{j'})$$

$$= n \, \mathrm{Var}\, S_1 + n(n-1) \, \mathrm{Cov}(S_1, S_2).$$

$$P(S_1 = s_1) = \frac{1}{N}$$

$$\mathrm{Var}\, S_1 = \mathbb{E} S_1^2 - (\mathbb{E} S_1)^2$$

$$= \sum_{s_1=1}^{N} s_1^2 \cdot \frac{1}{N} - \left(\frac{N(N+1)}{2}\right)^2$$

$$= \frac{1}{N} \frac{N(N+1)(2N+1)}{6} - \left(\frac{N(N+1)}{2}\right)^2$$

$\vdots$

$$= \frac{(N-1)(N+1)}{12}.$$

Find $\mathrm{Cov}(S_1, S_2)$ by crazy trick.

let $n = N$, then $S_1 + \dots + S_n = \underline{S_1 + \dots + S_N} \Rightarrow \mathrm{Var}\, W_S = 0$

$$\mathrm{Var}\, W_S = N \,\mathrm{Var}\, S_1 + N(N-1)\,\mathrm{Cov}(S_1, S_2) = 0$$

$$\Leftrightarrow \quad N \,\mathrm{Var}\, S_1 = -N(N-1)\,\mathrm{Cov}(S_1, S_2)$$

$$\Leftrightarrow \quad -\frac{N}{N(N-1)} \mathrm{Var}\, S_1 = \mathrm{Cov}(S_1, S_2)$$

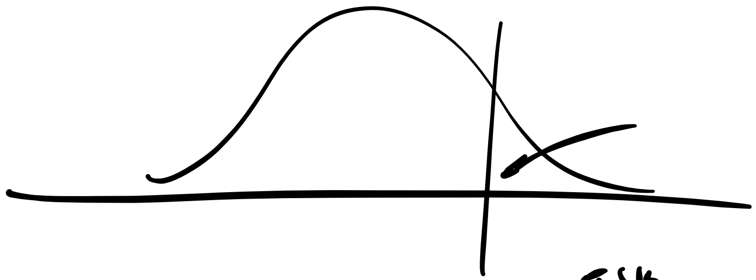
$$\Leftrightarrow \quad -\frac{N}{N(N-1)} \frac{(N+1)(N-1)}{12} = \mathrm{Cov}(S_1, S_2)$$

$$''$$

$$\mathrm{Cov}(S_1, S_2) = -\frac{(N+1)}{12}.$$

$$\text{Var } W_s = \frac{n(N-1)(N+1)}{12} - \frac{n(n-1)(N+1)}{12}$$

$$= \frac{n}{12}(N-n)(N+1).$$



$$\frac{W_S - \mathbb{E} W_S}{\sqrt{\mathrm{Var} W_S}}$$

Corollary

An asymptotic p*-value for testing* H_0: $F = G$ *versus the "right-sided" alternative is*

$$1 - \Phi\left(\frac{W_S - \frac{1}{2}n(N+1) + \frac{1}{2}}{\sqrt{\frac{1}{12}n(N-n)(N+1)}}\right),$$

where the extra $\frac{1}{2}$ *is a "continuity correction".*

$W_S = S_1 + \dots + S_n$

Sketch of asymptotic Normality proof

N $n = 4$

Assume H_0: $F = G$ and introduce $U_1, \dots, U_{\cancel{n}\,N} \overset{\text{ind}}{\sim}$ Uniform$(0,1)$. Then:

1. Write W_S as a sum of *dependent* rvs: $W_S = \sum_{i=1}^{N} i \cdot J_i$, $J_i = \mathbf{1}(U_i \leq U_{(n)})$.
2. Introduce approximator $\tilde{W}_S$, which is a sum of *independent* rvs:
$$\tilde{W}_S = \sum_{i=1}^{N}\left(i - \frac{N+1}{2}\right)K_i + \frac{n(N+1)}{2}, \quad \boxed{K_i = \mathbf{1}(U_i \leq n/N).}$$
3. Show that $\frac{\tilde{W}_S - \mathbb{E}\tilde{W}_S}{\sqrt{\text{Var}\,\tilde{W}_S}} \xrightarrow{\text{D}}$ Normal$(0,1)$ as $n \to \infty$.
4. Argue same convergence holds for W_S since $\frac{\mathbb{E}(\tilde{W}_S - W_S)^2}{\text{Var}\,\tilde{W}_S} \to 0$ as $\overset{N}{n} \to \infty$.

Exercise:

1. Show $\mathbb{E}\tilde{W}_S = \mathbb{E}W_S$ and $\boxed{\text{Var}\,\tilde{W}_S = \frac{(N-1)}{N}\text{Var}\,W_S.}$
2. Establish $\boxed{\frac{\tilde{W}_S - \mathbb{E}\tilde{W}_S}{\sqrt{\text{Var}\,\tilde{W}_S}}} \xrightarrow{\text{D}}$ Normal$(0,1)$ as $\cancel{n}\,N \to \infty$ with the Lindeberg CLT.

$N \to \infty$ and $n \to \infty$, $N - n \to \infty$.

$$\mathbb{E}\,\tilde{W}_s = \mathbb{E}\left[\sum_{i=1}^{N}\left(i-\frac{N+1}{2}\right)K_i + \frac{n(N+1)}{2}\right]$$

$$= \sum_{i=1}^{N}\left(i-\frac{N+1}{2}\right)\frac{n}{N} + \frac{n(N+1)}{2}$$

$$= \left(\frac{N(N+1)}{2} - \frac{N(N+1)}{2}\right)\frac{n}{N} + \underline{\frac{1}{2}n(N+1)}$$

$$= \frac{1}{2}n(N+1)$$

$$= \mathbb{E}\,W_s$$

$$\mathrm{Var}\,\tilde{W}_s = \mathrm{Var}\left(\sum_{i=1}^{N}\left(i-\frac{N+1}{2}\right)K_i + \frac{n(N+1)}{2}\right)$$

$$= \sum_{i=1}^{N}\left(i-\frac{N+1}{2}\right)^2 \underline{\mathrm{Var}\,K_i}$$

$$= \sum_{i=1}^{N}\left(i-\frac{N+1}{2}\right)^2 \frac{n}{N}\left(1-\frac{n}{N}\right)$$

$$\frac{\tilde{W}_s - \mathbb{E}\,\tilde{W}_s}{\sqrt{\mathrm{Var}\,\tilde{W}_s}} = \frac{\left[\sum_{i=1}^{N}\left(i-\frac{N+1}{2}\right)K_i + \frac{n(N+1)}{2}\right] - \frac{n(N+1)}{2}}{\sqrt{\sum_{i=1}^{N}\left(i-\frac{N+1}{2}\right)^2 \frac{n}{N}\left(1-\frac{n}{N}\right)}}$$

$$\sum_{i=1}^{N}\left(i-\frac{N+1}{2}\right)\frac{n}{N} = 0$$

$$= \frac{\displaystyle\sum_{i=1}^{N} \left(i - \frac{N+1}{2}\right) \underbrace{\frac{\left(K_i - \frac{n}{N}\right)}{\sqrt{\frac{n}{N}\left(1-\frac{n}{N}\right)}}}_{\text{mean zero, variance } = 1}}{\sqrt{\displaystyle\sum_{i=1}^{N} \left(i - \frac{N+1}{2}\right)^2}}$$

$$= \frac{\displaystyle\sum_{i=1}^{n} a_i q_i}{\sqrt{\displaystyle\sum_{i=1}^{n} a_i^2}} \xrightarrow{D} N(0,1).$$

(Lindeberg CLT).

if

$$\frac{\max |a_i|}{\sqrt{\displaystyle\sum_{i=1}^{n} a_i^2}} \longrightarrow 0 \quad \text{as} \quad N \to \infty.$$

☑

$X_1, \ldots, X_m \overset{ind}{\sim} F$ control

$Y_1, \ldots, Y_n \overset{ind}{\sim} G$ trt.

$$\underline{X_1}\ \ Y_1\ \ X_2\ \ Y_2\ \ Y_3\ \ \underline{X_4}\ \ \cdots\ \ \underline{X}\ \ \underline{Y}\ \ \underline{Y}\ \ \underline{X}$$

$$1\ \ 2\ \ 3\ \ 4\ \ 5\ \ 6\ \ \cdots\ \ N-3\ \ N-2\ \ N-1\ \ N = n+m$$

S_1 S_2 S_3

$$W_s = S_1 + \ldots + S_n$$

1. Wilcoxon rank sum test

$$\frac{W_s - \mathbb{E} W_s}{\sqrt{\operatorname{Var} W_s}} \xrightarrow{D} N(0,1)$$

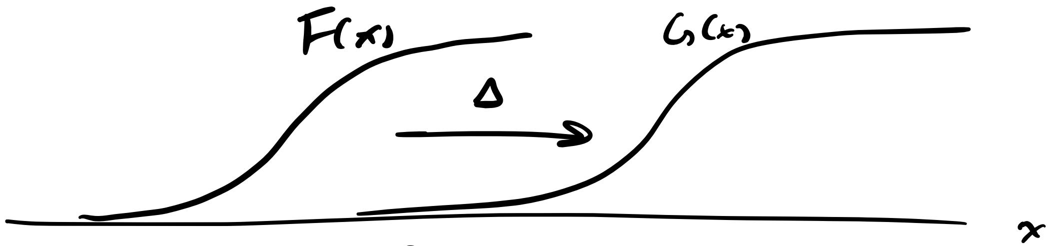
$$W_s = \sum_{i=1}^{N} i \cdot J_i$$

$$J_i = \begin{cases} 1 & U_i \leq U_{(n)} \\ 0 & \text{o.w.} \end{cases}$$

where $U_1, \ldots, U_N \overset{ind}{\sim} \text{Unif}(0,1)$.

2. Power comparisons

$H_0: F = G$ vs treatment tends to increase values of Y.



To analyze the power of the WXRS we must specify an alternative to H_0: $F = G$.

Location shift model

In the *location-shift* model we assume $\boxed{G(x)} = F(x - \Delta)$ for some $\Delta \in \mathbb{R}$.

We will consider the right-sided test H_0: $\Delta \leq 0$ vs H_1: $\Delta > 0$.

$W_{XY} = \sum_{i=1}^{m}\sum_{j=1}^{n} \mathbb{1}(X_i < Y_j)$

Exercise: Show that the power of the rule $W_{XY} \geq c$ is nondecreasing in Δ.

Power: $\gamma(\Delta) = P_\Delta(W_{XY} \geq c)$

$= P_\Delta\left(\sum_{i=1}^{m}\sum_{j=1}^{n} \mathbb{1}(X_i < Y_j) \geq c\right)$

$X_1, \ldots, X_m \overset{ind}{\sim} F$

Introduce $X_1', \ldots, X_n' \overset{ind}{\sim} F$

$Y_j \overset{D}{=} X_j' + \Delta$ for $j = 1, \ldots, n$

$$= P_\Delta\left(\sum_{i=1}^{m}\sum_{j=1}^{n}\mathbb{1}(X_i < X_j' + \Delta) \geq c\right)$$

$$= P_\Delta\left(\sum_{i=1}^{m}\sum_{j=1}^{n}\mathbb{1}(X_i - X_j' < \Delta) \geq c\right)$$

is non-decreasing in Δ.

It is convenient to use a Normal approximation to the power:

Theorem (Approximate power of WXRS in location-shift model)

In the location-shift model the power of $W_{XY} \geq c$ admits the approximation

$$\gamma(\Delta) \approx 1 - \Phi\left(\frac{c - \overbrace{nmp_1(\Delta)}^{\mathbb{E} W_{XY}}}{\underbrace{\sqrt{\vartheta(\Delta)}}_{\mathrm{Var}\, W_{XY}}}\right)$$

provided N, n, and $N - n$ are all large, where $p_1(\Delta) = P(X_1 < Y_1)$ and

$$\vartheta(\Delta) = mnp_1(\Delta)[1 - p_1(\Delta)] + mn(n-1)[p_2(\Delta) - p_1^2(\Delta)] + \underline{nm(m-1)[p_3(\Delta) - p_1^2(\Delta)]}$$

with $p_{\cancel{2}\,3}(\Delta) = P(X_1 < Y_1, X_2 < Y_1)$ and $p_{\cancel{3}\,2}(\Delta) = P(X_1 < Y_1, X_1 < Y_2)$.

Exercise:

1. Establish the above result.
2. Find the value of c such that the test has size approximately equal to α.

$$W_{xy} = \overbrace{S_1 + \dots + S_n}^{W_S} - \tfrac{1}{2}n(n+1)$$

$$\mathbb{E} W_{xy} = \mathbb{E} W_S - \tfrac{1}{2}n(n+1)$$

$$\mathrm{Var}\, W_{xy} = \mathrm{Var}\, W_S$$

$$\gamma(\Delta) = P_\Delta\left(W_{xy} \geq c \right)$$

$$= P_\Delta\left(\underbrace{\frac{W_{xy} - \mathbb{E} W_{xy}}{\sqrt{\mathrm{Var}\, W_{xy}}}} \geq \frac{c - \mathbb{E} W_{xy}}{\sqrt{\mathrm{Var}\, W_{xy}}} \right)$$

$$= P_\Delta\left(\frac{W_S - \mathbb{E} W_S}{\sqrt{\mathrm{Var}\, W_S}} \geq \frac{c - \mathbb{E} W_{xy}}{\sqrt{\mathrm{Var}\, W_{xy}}} \right)$$

$$\left(\text{for large } N, N-m, n\right) \quad \approx P_\Delta\left(Z \geq \frac{c - \mathbb{E} W_{xy}}{\sqrt{\mathrm{Var}\, W_{xy}}} \right), \quad \text{where } Z \sim N(0,1).$$

$$= 1 - \Phi\left(\frac{c - \mathbb{E} W_{xy}}{\sqrt{\mathrm{Var}\, W_{xy}}} \right).$$

$$\mathbb{E} W_{xy} = \mathbb{E} \sum_{i=1}^{m} \sum_{j=1}^{n} \mathbb{1}(X_i < Y_j)$$

$$= \sum_{i=1}^{m} \sum_{j=1}^{n} P(\underline{X_i < Y_j}) \qquad Y_j := X_j' + \Delta$$

$$= \sum_{i=1}^{m} \sum_{j=1}^{n} P(X_i < X_j' + \Delta)$$

$$= \sum_{i=1}^{m} \sum_{j=1}^{n} P(X_i - X_j' < \Delta)$$

$$= mn\, P(X_1 - X_1' < \Delta) \qquad X_1, X_1' \overset{\text{iid}}{\sim} F$$

$$= mn\, p_1(\Delta), \qquad p_1(\Delta) = F^*(\Delta), \quad F^* \text{ is cdf of } X_1 - X_1'.$$

$$\mathrm{Var}\, W_{XY} = \mathrm{Var}\left(\sum_{i=1}^{m} \sum_{j=1}^{n} \mathbb{1}(X_i < Y_j) \right)$$

$$= \sum_{i=1}^{m} \mathrm{Var}\left(\sum_{j=1}^{n} \mathbb{1}(X_i < Y_j) \right)$$

$$+ \sum_{i \neq i'} \mathrm{Cov}\left(\sum_{j=1}^{n} \mathbb{1}(X_i < Y_j), \sum_{j'=1}^{n} \mathbb{1}(X_{i'} < Y_{j'}) \right)$$

$$= \sum_{i=1}^{m} \left\{ \sum_{j=1}^{n} \underbrace{\mathrm{Var}\left(\mathbb{1}(X_i < Y_j) \right)}_{p_1(\Delta)[1 - p_1(\Delta)]} + \sum_{j \neq j'} \mathrm{Cov}\left(\mathbb{1}(X_i < Y_j), \mathbb{1}(X_i < Y_{j'}) \right) \right\}$$

$$+ \underbrace{\sum_{i \neq i'} \sum_{j=1}^{n} \sum_{j'=1}^{n} \mathrm{Cov}\left(\mathbb{1}(X_i < Y_j), \mathbb{1}(X_{i'} < Y_{j'}) \right)}_{\sum_{i \neq i'} \sum_{j=1}^{n} \mathrm{Cov}\left(\mathbb{1}(X_i < Y_j), \mathbb{1}(X_{i'} < Y_j) \right)}$$

$$= n \cdot m \cdot p_1(\Delta)[1 - p_1(\Delta)] + \sum_{i=1}^{m} \sum_{j \neq j'} \underbrace{\mathrm{Cov}\left(\mathbb{1}(X_i < Y_j), \mathbb{1}(X_i < Y_{j'}) \right)}_{\underbrace{P(X_1 < Y_1 \cap X_1 < Y_2)}_{p_2(\Delta)} - p_1^2(\Delta)}$$

$$+ \sum_{i \neq i'} \sum_{j=1}^{n} \underbrace{\mathrm{Cov}\left(\mathbb{1}(X_i < Y_j), \mathbb{1}(X_{i'} < Y_j) \right)}_{\underbrace{\mathbb{E}\left(\mathbb{1}(X_i < Y_j) \cdot \mathbb{1}(X_{i'} < Y_j) \right)}_{\underbrace{P(X_1 < Y_2 \cap X_2 < Y_2)}_{p_3(\Delta)}} - \underbrace{\mathbb{E}\, \mathbb{1}(X_i < Y_j)}_{p_1(\Delta)} \underbrace{\mathbb{E}\, \mathbb{1}(X_{i'} < Y_j)}_{p_1(\Delta)}}$$

$$= n m \, p_1(\Delta)[1 - p_1(\Delta)] + m \, n(n-1)[p_2(\Delta) - p_1^2(\Delta)]$$

$$+ n \, m(m-1)[p_3(\Delta) - p_1^2(\Delta)]$$

② Find c so the test has size α:

$$\text{size} = \beta(0) = 1 - \Phi\left(\frac{c - mn\, p_1(0)}{\sqrt{\vartheta(0)}}\right),$$

set equal to α

$$p_1(0) = P(X_1 < Y_1) = \tfrac{1}{2}$$

$$p_2(0) = P(X_1 < Y_1 \cap X_1 < Y_2) = P(X_1 \text{ is smallest among } X_1, Y_1, Y_2) = 1/3$$

$$p_3(0) = P(X_1 < Y_1 \cap X_2 < Y_1) = P(Y_1 \text{ is greatest among } X_1, X_2, Y_1) = 1/3$$

$\Rightarrow$

$$\vartheta(0) = nm\, p_1(s)[1 - p_1(s)] + m\, n(n-1)[p_2(s) - p_1^2(s)] + n\, m(m-1)[p_3(s) - p_1^2(s)]$$

$$= nm\, \tfrac{1}{2}\left(1 - \tfrac{1}{2}\right) + mn(n-1)\left[\tfrac{1}{3} - \left(\tfrac{1}{2}\right)^2\right] + nm(m-1)\left[\tfrac{1}{3} - \left(\tfrac{1}{2}\right)^2\right]$$

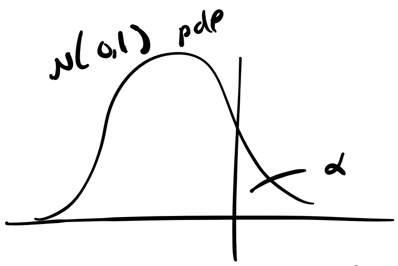
$$= \frac{3nm}{12} + \frac{1}{12}\left[mn(n-1) + nm(m-1)\right]$$

$$= \frac{nm}{12}\left[3 + n - 1 + m - 1\right]$$

$$= \frac{nm}{12}(\underbrace{n+m}_{N} + 1)$$

$$\boxed{= \frac{n}{12}(N-n)(N+1)} \leftarrow \frac{nm(N+1)}{12}$$

$$\gamma(\Delta) \approx 1 - \Phi\left(\frac{c - mn\frac{1}{2}}{\sqrt{\frac{nm}{12}(N+1)}}\right) = \alpha$$



$$z_\alpha = \frac{c - mn\,^1\!/_2}{\sqrt{\frac{nm(N+1)}{12}}}$$

$$\Leftrightarrow \quad \boxed{c_\alpha = z_\alpha\sqrt{\frac{nm(N+1)}{12}} + \frac{mn}{2}}$$

So size-α WRS is reject when

$$W_{XY} \geq c_\alpha = z_\alpha\sqrt{\frac{nm(N+1)}{12}} + \frac{mn}{2}.$$

(approx) Power under c_α:

$$\tilde{\gamma}_\alpha(\Delta) = 1 - \Phi\left(\frac{z_\alpha\sqrt{\frac{nm(N+1)}{12}} + \frac{mn}{2} - mn\,p_1(\Delta)}{\sqrt{v(\Delta)}}\right)$$

$p_1(\Delta) = p_1(X_1 < Y_1) = P(X_1 < X_1' + \Delta) = P(X_1 - X_1' < \Delta) = F^*(\Delta)$,

Let f^* be pdf of $X_1 - X_1'$. $X_1 - X_1' \sim F$

Then $p_1(\Delta) = F^*(\Delta) \approx \underbrace{F^*(0)}_{P(X_1 < X_1') = 1/2} + f^*(0)(\Delta - 0)$

Exercise: Show that making the substitutions

1. $c = c_\alpha$
2. $p_1(\Delta) = \boxed{1/2 + \Delta f^*(0)}$, f^* the density of $X_1 - X_2$
3. $\vartheta(\Delta) = \vartheta(0)$

leads to the approximate power curve for the size-α test given by

$$\tilde{\gamma}_\alpha(\Delta) = 1 - \Phi\left(z_\alpha - \sqrt{\frac{12nm}{N+1}} \cdot \Delta \cdot f^*(0)\right).$$

$$\tilde{\gamma}_\alpha(\Delta) = 1 - \Phi\left(\frac{z_\alpha \sqrt{\frac{nm(N+1)}{12}} + \frac{mn}{2} - mn\,p_1(\Delta)}{\sqrt{\underbrace{v(\Delta)}_{\text{replace with } v(0)}}} \right)$$

$$\approx 1 - \Phi\left(\frac{z_\alpha \sqrt{\frac{nm(N+1)}{12}} + \cancel{\frac{mn}{2}} - mn\left[\frac{1}{2} + \Delta \cdot f^*(0)\right]}{\sqrt{\frac{nm(N+1)}{12}}} \right)$$

$$= 1 - \Phi\left(z_\alpha - \sqrt{\frac{12 \cdot m \cdot n}{N+1}}\, \Delta \cdot f^*(0) \right)$$

If $n = m$, $N+1 \approx 2n$

$$\tilde{\gamma}_\alpha(\Delta) = 1 - \Phi\left(z_\alpha - \sqrt{\frac{12 \cdot n^2}{2n}}\, \Delta \cdot f^*(0) \right)$$

$$= 1 - \Phi\left(z_\alpha - \sqrt{6n}\, \Delta \cdot \underline{f^*(0)} \right)$$

If F is Normal(μ, σ^2) then $X_1, X_1' \sim$ Normal(μ, σ^2)

f^* is pdf of $X_1 - X_1' \sim$ Normal$(0, 2\sigma^2)$

$$f^*(x) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{2\sigma^2}} e^{-\frac{x^2}{2 \cdot 2\sigma^2}} \Rightarrow f^*(0) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{2\sigma^2}} = \frac{1}{2\sqrt{\pi}} \frac{1}{\sigma}$$

$$\tilde{\gamma}_\alpha(\Delta) = 1 - \Phi\left(z_\alpha - \frac{\sqrt{6n}}{2\sqrt{\pi}\,\sigma}\Delta\right)$$

Find smallest n such that

$$\iff \quad \tilde{\gamma}_\alpha(\Delta^*) \geq \gamma^*$$

$$1 - \Phi\left(z_\alpha - \frac{\sqrt{6n}}{2\sqrt{\pi}\,\sigma}\Delta\right) \geq \gamma^*$$

$$z_{\gamma^*} \geq z_\alpha - \frac{\sqrt{6n}}{2\sqrt{\pi}\,\sigma}\Delta^*$$

$$\iff$$

$$z_{\gamma^*} \geq z_\alpha - \frac{\sqrt{6n}}{2\sqrt{\pi}\,\sigma}\Delta$$

$$\frac{\sqrt{6n}}{2\sqrt{\pi}\,\sigma}\Delta^* \geq z_\alpha - z_{\gamma^*}$$

$$(z_{\gamma^*} = -z_{1-\gamma^*}$$

$$\sqrt{n} \geq \frac{2\sqrt{\pi}\,\sigma\,(z_\alpha + z_{1-\gamma^*})}{\sqrt{6}\,\Delta^*}$$

$$n \geq \frac{2}{6}\pi\,\frac{2\sigma^2(z_\alpha + z_{1-\gamma^*})^2}{(\Delta^*)^2}$$

WXRS $\rightarrow$ $$n \geq \underbrace{\frac{\pi}{3}}_{1.05}\,\frac{2\sigma^2(z_\alpha + z_{1-\gamma^*})^2}{(\Delta^*)^2}$$

What if $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$

$Y_1, \dots, Y_n \overset{iid}{\sim} N(\mu + \Delta, \sigma^2)$

Reject $H_0: \Delta \le 0$ in favor of $H_1: \Delta > 0$ if

$$\frac{\bar{Y} - \bar{X}}{\sigma\sqrt{\frac{1}{n} + \frac{1}{n}}} > z_\alpha .$$

To reject H_0 when $\Delta = \Delta^*$ with prob $\ge \gamma^*$ we need

$$n \ge \boxed{\cancel{\tfrac{\pi}{3}} \frac{2\sigma^2 (z_\alpha + z_{1-\gamma^*})^2}{(\Delta^*)^2}}$$

Z-test $\nearrow$

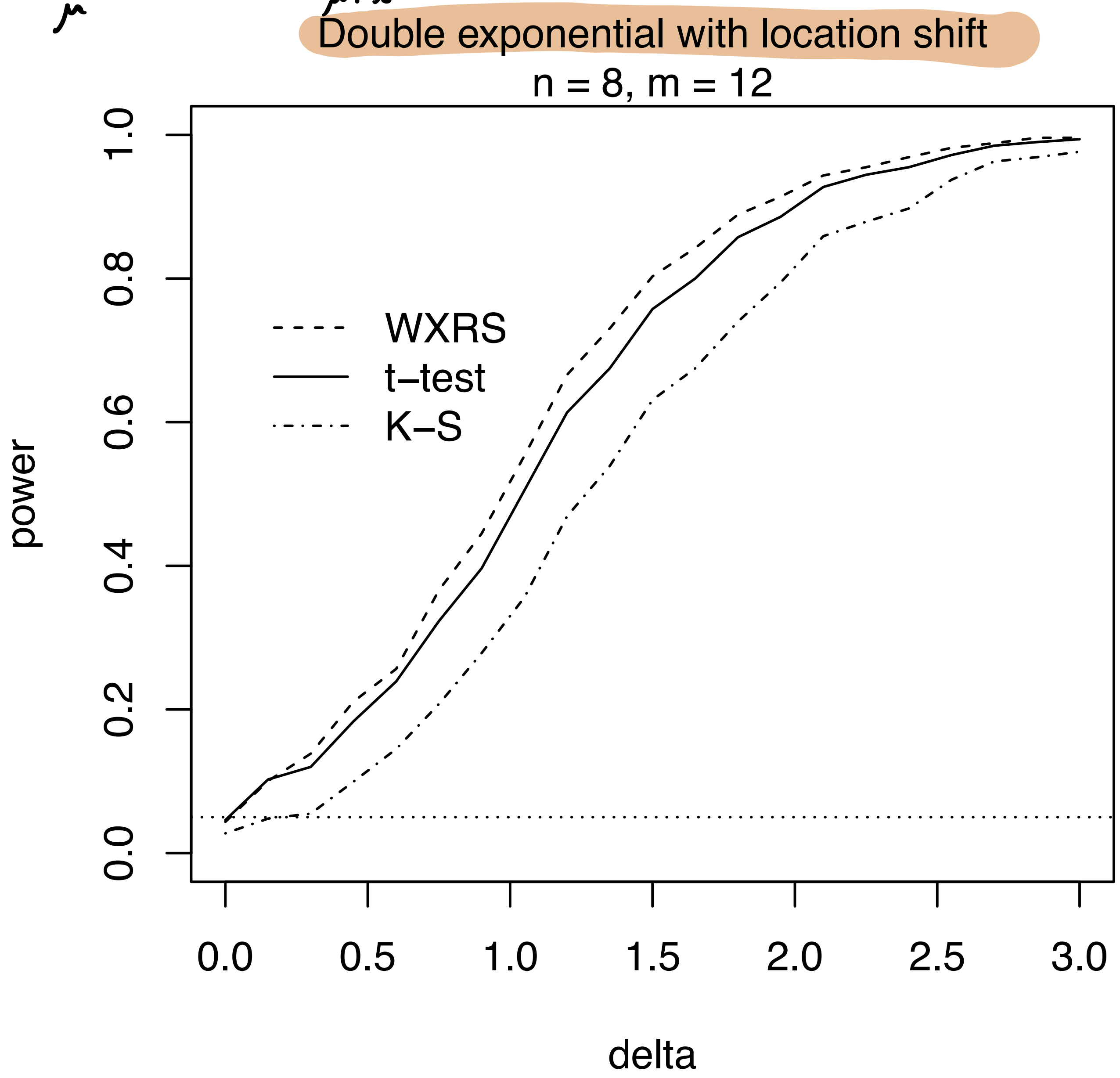
Exercise:

1. Show that if F is Normal, $n = m$, and $N + 1$ is replaced by $2n$, we obtain

$$\tilde{\gamma}_\alpha(\Delta) = 1 - \Phi\left(z_\alpha - \frac{\sqrt{6n}}{2\sigma\sqrt{\pi}} \cdot \Delta\right).$$

2. Find the smallest n such that the WXRS has power $\geq \gamma^*$ when $\Delta \geq \Delta^*$.
3. Compare to n needed for the equal-variances two-sample z-test.


Δ
μ
μ+Δ


Double exponential with location shift
n = 8, m = 12
WXRS
t–test
K–S
power
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
delta