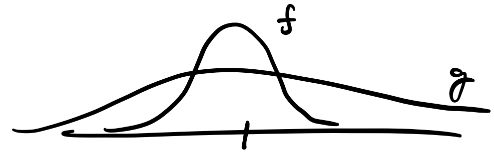
STAT 824 sp 2025 Lec 12 slides

Wilcoxon rank-sum test

Karl 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.



$X_1, \dots, X_n \overset{iid}{\sim} F$

$Y_1, \dots, Y_m \overset{iid}{\sim} G$

WRST : Test $H_0 : F = G$

versus H_1 : "The Y_i tend to be greater than the X_i"

Compute

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

Reject H_0 if $W_{XY} \geq c$, for some $c > 0$.

Aside:

$P(X < Y)$.

$H_0: F = G$

$$\mathbb{E}\left[\frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} \mathbb{1}(X_i \le Y_j)\right] = \mathbb{E}\, \mathbb{1}(X_1 \le Y_1) = \frac{1}{2}$$

$$= \mathbb{E}\left(\mathbb{E}\left[\mathbb{1}(\underbrace{X_1 \le Y_1}) \mid Y_1\right]\right)$$

$$= \mathbb{E}\left(F(Y_1)\right)$$

$g(y)$ the density of G

$$= \int_{-\infty}^{\infty} F(y)\, g(y)\, dy$$

$$= \int_0^1 F(G^{-1}(u)) \frac{g(G^{-1}(u))}{g(G^{-1}(u))}\, du$$

$u = G(y)$ $\quad \frac{du}{dy} = g(y)$

$y = G^{-1}(u)$ $\quad \frac{dy}{du} = \frac{1}{g(y)} = \frac{1}{g(G^{-1}(u))}$

$$= \int_0^1 F(G^{-1}(u))\, du$$

$H_0: F = G$

$$= \int_0^1 u \, du$$

$$= \frac{1}{2} .$$

Could set $\quad H_0: P(X \leq Y) \leq 1/2$

$\quad H_1: P(X \leq Y) > 1/2$

Reject H_0 if $W_{XY} \geq c$, some $c > 0$.

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

$$X_1, \ldots, X_n \overset{\text{ind}}{\sim} F \qquad \text{“control”}$$
$$Y_1, \ldots, Y_m \overset{\text{ind}}{\sim} G \qquad \text{“treatment”}$$

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 (WXRS)* concludes a "positive treatment effect" if

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

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.

Exercise: For $X \sim F$ and $Y \sim G$, both continuous, show

1. $P(X < Y) = \int_0^1 F(G^{-1}(u))du$.
2. $P(X < Y) = 1/2$ if $F = G$.

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

$N = n+m$

Rank-sum form of Wilcoxon rank sum statistic

An alternate way of computing W_{XY}: $(Z_1,\dots,Z_N) = (X_1,\dots,X_n,Y_1,\dots,Y_m)$.

1. Sort all the data $(X_1,\dots,X_n,Y_1,\dots,Y_m)$
2. Obtain the ranks. R_j is rank of Y_j. Then $R_j = \sum_{i=1}^{N} \mathbb{1}(Z_i \le Y_j)$
3. Keep the ranks corresponding to $Y_1,\dots,Y_m$, calling these $R_1,\dots,R_m$

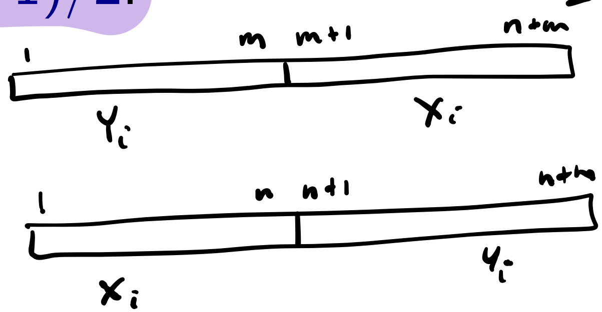
Then $W_{XY} = \underbrace{R_1+\cdots+R_m}_{W_R} - m(m+1)/2.$

Count # obs $\le Y_j$.

Let $W_R = R_1+\cdots+R_n$.

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

What values does $\underline{W_R}$ take?



$\to W_R = \frac{m(m+1)}{2}$

$\to W_R = mn + \frac{m(m+1)}{2}$

eliminate probability of ties

Theorem $(H_0: F = G)$

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

$$P(\{R_1, \dots, R_m\} = \{r_1, \dots, r_m\}) = \frac{1}{\binom{N}{m}}.$$

for all sets of m ranks $\{r_1, \dots, r_m\} \subset \{1, \dots, N\}$.

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

$N = 5$

$n = 3$, $m = 2$

Y_2, X_1, Y_1, X_2, X_3

$\dfrac{m(m+1)}{2} = \dfrac{2(2+1)}{2} = 3$

20

r_1	r_2	W_R
1	2	3
1	3	4
1	4	5
1	5	6
2	1	3
2	3	5
2	4	6
2	5	7
3	1	4
3	2	5
3	4	7
3	5	8
4	1	5
4	2	6
4	3	7
4	5	9
5	1	6
5	2	7
5	3	8
5	4	9

W_{XY}

subtract 3

w	3	4	5	6	7	8	9
$P(W_R = w)$	$2/20$	$2/20$	$4/20$	$4/20$	$4/20$	$2/20$	$2/20$

Reject H_0 if $W_{XY} \geq 6$.

Type I error rate $= 0.10$.

w	0	1	2	3	4	5	6
$P(W_{XY} = w)$	$2/20$	$2/20$	$4/20$	$4/20$	$4/20$	$2/20$	$2/20$

```
# generate some data
n <- 20
m <- 25
X <- rnorm(n,1,1)
Y <- rnorm(m,1,1)

# compute WR and WXY
Z <- c(X,Y)
id <- c(rep(1,n),rep(2,m))
R <- rank(Z)[id == 2]
WR <- sum(R)
WXY <- sum(R) - m*(m+1)/2

# must subtract 1, since we reject when WXY >= c
pval <- 1 - pwilcox(WXY-1,m = m, n = n)

# see that the wilcox.test() function gives the same values
wilcox.test(x=Y,y=X,alternative="greater",exact = TRUE) # switch X and Y
```

Reject $H_0: F=G$ when

$W_{xy} \geq c$

$$\text{p-val} = P\left(W_{xy} \geq W_{xy}^{obs}\right) = 1 - P\left(W_{xy} \leq W_{xy}^{obs} - 1\right)$$

slow for large n, m.

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

From now on, focus on $W_R = R_1 + \cdots + R_m$

Theorem (Asymptotic Normality of rank sum under the null)

Under H_0: $F = G$ we have

$$\frac{W_R - \mathbb{E}W_R}{\sqrt{\operatorname{Var} W_R}} \xrightarrow{d} \text{Normal}(0, 1)$$

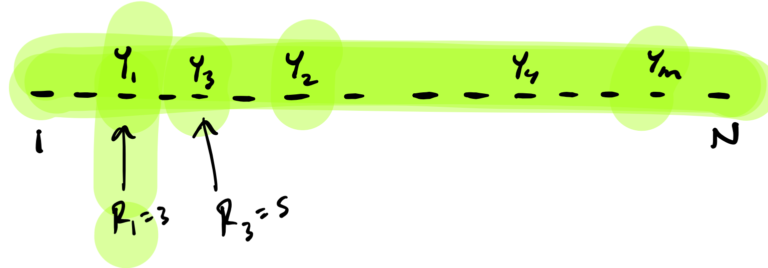
as $n, m \to \infty$.

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

$$\mathbb{E}\, W_R = \mathbb{E} \sum_{j=1}^{m} R_j$$

$$= m\, \mathbb{E} R_1$$

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



$R_1 \sim$ Discrete Unif on $1, \dots, N$.

$$\mathbb{E}\, R_1 = \frac{N+1}{2}$$

$$\mathrm{Var}\, R_1 = \frac{N^2-1}{12} = \frac{(N+1)(N-1)}{12}$$

$$\mathrm{Var}\, W_R = \mathrm{Var}\left(\sum_{j=1}^{m} R_j \right)$$

$$= \sum_{j=1}^{m} \mathrm{Var}\, R_j + 2 \sum_{j<k} \mathrm{Cov}(R_j, R_k)$$

$j<k \to \binom{m}{2}$

$$= \frac{m(N+1)(N-1)}{12} + 2\, \frac{m(m+1)}{2}\, \underbrace{\mathrm{Cov}(R_1, R_2)}_{?}$$

Trick to obtain $\mathrm{Cov}(R_1, R_2)$: Set $m = N$. $(n = 0)$

Then $W_R = \dfrac{N(N+1)}{2}$, $\left(\begin{array}{l} Y_1, \dots, Y_N \\ \text{are the} \\ \text{only observations} \\ \text{if } m = N \end{array} \right)$

so $\mathrm{Var}\, W_R = 0$.

We have, if $m = N$,

$$0 = \mathrm{Var}\, W_R = \frac{N(N+1)(N-1)}{12} + 2\, \frac{N(N+1)}{2}\, \mathrm{Cov}(R_1, R_2)$$

$$\iff -\frac{(N-1)}{12} = \mathrm{Cov}(R_1, R_2)$$

$$\text{Var}\, W_R = \frac{m(N+1)(N-1)}{12} + 2\,\frac{m(m+1)}{2}\left[-\frac{(N-1)}{12}\right]$$

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

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

Corollary

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

$$1 - \Phi\left(\frac{W_R^{obs} - \frac{1}{2}m(N+1) \cancel{+} - \frac{1}{2}}{\sqrt{\frac{1}{12}m(N-m)(N+1)}}\right),$$

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

Want to show:

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

Strategy:

① Make approximation $\tilde{W}_R$ to W_R.

Construct $\tilde{W}_R$ as a sum of <u>independent</u> r.v.s.

② show

$$\frac{\tilde{W}_R - \mathbb{E}\, \tilde{W}_R}{\sqrt{\operatorname{Var} \tilde{W}_R}} \xrightarrow{D} N(0,1). \quad (☆)$$

③ Argue that (☆) implies (☼), i.e. $\tilde{W}_R$ is a good approx. to W_R.

(1): Under $H_0: F = G$. $R_1, \ldots, R_m$ are a draw without repl. from $\{1, \ldots, N\}$.

One way to draw m indices from $1, \ldots, N$ without repl. is:

(i) Generate $U_1, \ldots, U_N \overset{iid}{\sim} U(0,1)$.

(ii) Order them $U_{(1)} < \cdots < U_{(N)}$

(iii) Keep indices $\{i : U_i \le U_{(m)}\}$

So, do this, and write

$$W_R \overset{D}{=} \sum_{i=1}^{N} i J_i \quad = \sum_{j=1}^{m} R_j$$

$$\boxed{J_i} = \mathbb{1}(U_i \le U_{(m)}) \qquad P(J_i = 1) = \frac{m}{N}$$

dependent.

$$\sum_{i=1}^{N} J_i = m$$

Let $\boxed{K_i} = \mathbb{1}\left(U_i \le \frac{m}{N}\right)$. $\quad K_i \sim \text{Bernoulli}\left(\frac{m}{N}\right)$

independent

$$\sum_{i=1}^{N} K_i \sim \text{Binomial}\left(N, \frac{m}{N}\right)$$

$$\tilde{W}_R = \sum_{i=1}^{N}\left(i - \frac{N+1}{2}\right)K_i + \frac{1}{2}m(N+1)$$

$$\frac{\tilde{W}_R - \mathbb{E}\tilde{W}_R}{\sqrt{\text{Var}\,\tilde{W}_R}} \xrightarrow{d} N(0,1)$$

Let's find $\mathbb{E}\,\tilde{W}_R$:

$K_i \sim \text{Bernoulli}\left(\frac{m}{N}\right)$

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

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

$$= \frac{\cancel{N}(N+1)}{2}\frac{m}{\cancel{N}} - \frac{\cancel{N}(N+1)}{2}\frac{m}{\cancel{N}} + \frac{1}{2}m(N+1)$$

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

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

$$\tilde{W}_R = \sum_{i=1}^{N}\left(i - \frac{N+1}{2}\right)K_i + \frac{1}{2}m(N+1)$$

$$\operatorname{Var}\tilde{W}_R = \sum_{i=1}^{N}\left(i - \frac{N+1}{2}\right)^2 \frac{m}{N}\left(1 - \frac{m}{N}\right) \overset{\text{skip}}{= \cdots =} \frac{N-1}{N}\operatorname{Var}W_R$$

$$\frac{\tilde{W}_R - \mathbb{E}\,\tilde{W}_R}{\sqrt{\operatorname{Var}\tilde{W}_R}} = \frac{\sum_{i=1}^{N}\left(i - \frac{N+1}{2}\right)K_i}{\sqrt{\sum_{i=1}^{N}\left(i - \frac{N+1}{2}\right)^2 \frac{m}{N}\left(1 - \frac{m}{N}\right)}}$$

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

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

$$= \frac{\sum_{i=1}^{N} a_i g_i}{\sqrt{\sum_{i=1}^{N} a_i^2}}$$

$$g_i = \frac{K_i - \frac{m}{N}}{\sqrt{\frac{m}{N}\left(1 - \frac{m}{N}\right)}}$$

$$\uparrow$$

$\mathbb{E}\, g_i = 0$

$\mathbb{E}\, g_i^2 = 1$

$$a_i = i - \frac{N+1}{2}$$

$$\xrightarrow{D} N(0,1)$$

provided

$$\frac{\max_{1 \le i \le N} |a_i|}{\sqrt{\sum_{i=1}^{N} a_i^2}} \longrightarrow 0 \qquad \text{as} \quad n, m \to \infty.$$

We have

$$\max_{1 \le i \le N} |a_i| = \max\left\{ \left|1 - \frac{N+1}{2}\right|, \left|2 - \frac{N+1}{2}\right|, \dots, \left|N - \frac{N+1}{2}\right| \right\}$$

$$= \max\left\{ \left|1 - \frac{N+1}{2}\right|, \left|N - \frac{N+1}{2}\right| \right\}$$

$$= \left|\frac{N-1}{2}\right|$$

$$\frac{2 - N + 1}{2}$$

$$\frac{1-N}{2}$$

$$\frac{2N - N - 1}{2} = \frac{N-1}{2}$$

Then

$$\sum_{i=1}^{N} a_i^2 = \sum_{i=1}^{N} \left(i + \frac{N+1}{2}\right)^2 = \overset{\text{skip}}{\cdots} = \frac{N(N^2-1)}{12}$$

Check

$$\frac{\frac{N-1}{2}}{\sqrt{\frac{N(N^2-1)}{12}}} = \frac{\frac{1}{2}(N-1)}{\sqrt{\frac{1}{12}N(N-1)(N+1)}} = O\left(N^{-\frac{1}{2}}\right) ✓$$

$\to 0.$

$$\Rightarrow \quad \frac{\tilde{W}_R - \mathbb{E}\tilde{W}_R}{\sqrt{\operatorname{Var}\tilde{W}_R}} \xrightarrow{D} N(0,1).$$

③

Now:

$\mathbb{E}W_R = \mathbb{E}\tilde{W}_R$

$$\frac{W_R - \mathbb{E}W_R}{\sqrt{\operatorname{Var}W_R}} = \left(\frac{W_R - \tilde{W}_R}{\sqrt{\operatorname{Var}\tilde{W}_R}} + \underbrace{\frac{\tilde{W}_R - \mathbb{E}\tilde{W}_R}{\sqrt{\operatorname{Var}\tilde{W}_R}}}_{\xrightarrow{D} N(0,1)}\right)\boxed{\frac{\sqrt{\operatorname{Var}\tilde{W}_R}}{\sqrt{\operatorname{Var}W_R}}}$$

Now show that this $\xrightarrow{P} 0.$

$$= \sqrt{\frac{N-1}{N}} \to 1$$

Suff. to show

$$\frac{\mathbb{E}(W_R - \tilde{W}_R)^2}{\operatorname{Var}\tilde{W}_R} \longrightarrow 0 \quad \text{as} \quad n,m \to \infty.$$

see eBook or Lehmann (1975).

$$J_i = \mathbb{1}(U_i \le U_{(m)})$$

$$W_R - \tilde{W}_R = \sum_{i=1}^{N} i J_i - \sum_{i=1}^{N}\left(i - \frac{N+1}{2}\right) K_i - \frac{1}{2} m (N+1)$$

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

because $\sum_{i=1}^{N} J_i = m$

$$\mathbb{E}\left(W_R - \tilde{W}_R\right)^2 = \mathrm{Var}\left(W_R - \tilde{W}_R\right) = \mathrm{Var}(W_R) + \mathrm{Var}(\tilde{W}_R) - 2\,\mathrm{Cov}(W_R, \tilde{W}_R)$$

$$= \mathbb{E}\left(\mathrm{Var}\left(W_R - \tilde{W}_R \mid U_{(1)}, \dots, U_{(N)}\right)\right)$$

$$+ \mathrm{Var}\left(\mathbb{E}\left(W_R - \tilde{W}_R \mid U_{(1)}, \dots, U_{(N)}\right)\right)$$

$\vdots$

Complication: W_R is <u>not</u> a sum of independent r.v.s.

Sketch of asymptotic Normality proof

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

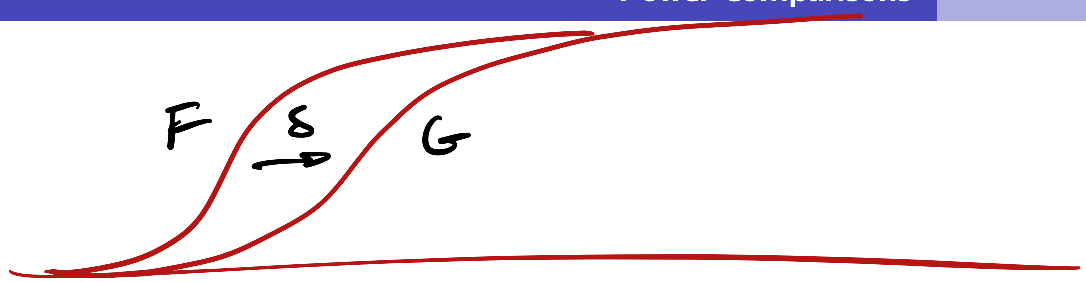
1. Write W_R as a sum of *dependent* rvs: $W_R = \sum_{i=1}^N i \cdot J_i, \quad J_i = \mathbf{1}(U_i \leq U_{(N)})$.
2. Introduce approximator $\boxed{\tilde{W}_R,}$ which is a sum of *independent* rvs:

$$\tilde{W}_R = \sum_{i=1}^{N} \left(i - \frac{N+1}{2} \right) K_i + \frac{m(N+1)}{2}, \quad K_i = \mathbf{1}(U_i \leq m/N).$$

3. Show that $\dfrac{\tilde{W}_R - \mathbb{E}\tilde{W}_R}{\sqrt{\text{Var}\, \tilde{W}_R}} \xrightarrow{\text{d}} \text{Normal}(0,1)$ as $n, m \to \infty$.
4. Argue same holds for W_R since $\dfrac{\mathbb{E}(W_R - \tilde{W}_R)^2}{\text{Var}\, \tilde{W}_R} \to 0$ as $n, m \to \infty$.

Exercise:

1. Show $\mathbb{E}\tilde{W}_R = \mathbb{E}W_R$ and $\text{Var}\,\tilde{W}_R = \frac{(N-1)}{N}\text{Var}\,W_R$.
2. Show $\dfrac{\tilde{W}_R - \mathbb{E}\tilde{W}_R}{\sqrt{\text{Var}\,\tilde{W}_R}} \xrightarrow{\text{d}} \text{Normal}(0,1)$ as $n, m \to \infty$ with Lindeberg CLT.



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 $G(x) = F(x - \Delta)$ for some $\Delta \in \mathbb{R}$.

$\Delta = 0 \iff F = G$

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

Exercise: Show that the power of the rule $W_{XY} \geq c$ is nondecreasing in Δ. > 0

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

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

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

$$= P\left(\sum_{i=1}^{n}\sum_{j=1}^{m} \mathbb{1}\left(X_i - \Delta \le X_j'\right) \ge c\right)$$

Let $X_1', \dots, X_m' \overset{ind}{\sim} F$, independent of $X_1, \dots, X_n$. $\quad Y_j - \Delta \overset{D}{=} X_j'$

$$= P\left(\sum_{i=1}^{n}\sum_{j=1}^{m} \mathbb{1}\left(X_i - X_j' \le \Delta\right) \ge c\right)$$

Let $\Delta_1 \le \Delta_2$. Then $\gamma(\Delta_1) \le \gamma(\Delta_2)$. ✓

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

$$\gamma(\Delta) \approx 1 - \Phi\left(\frac{c - nmp_1(\Delta)}{\sqrt{\vartheta(\Delta)}}\right)$$

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

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

with $p_2(\Delta) = P(X_1 < Y_1, X_2 < Y_1)$ and $p_3(\Delta) = P(X_1 < Y_1, X_1 < Y_2)$.

Exercise:

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

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

$$= \frac{W_R - \mathbb{E} W_R}{\sqrt{\operatorname{Var} W_R}} \xrightarrow{d} N(0,1)$$

Must use U-statistics to prove this under H_1.

$$= P_\Delta\left(\frac{W_{XY} - \mathbb{E}_\Delta W_{XY}}{\sqrt{\operatorname{Var}_\Delta W_{XY}}} \geq \frac{c - \mathbb{E}_\Delta W_{XY}}{\sqrt{\operatorname{Var}_\Delta W_{XY}}}\right)$$

$$\approx P\left(Z \geq \frac{c - \mathbb{E}_\Delta W_{XY}}{\sqrt{\operatorname{Var}_\Delta W_{XY}}}\right)$$

$$= 1 - \Phi\left(\frac{c - \mathbb{E}_\Delta W_{XY}}{\sqrt{\operatorname{Var}_\Delta W_{XY}}}\right)$$

see above

Find $\underline{\text{size-}\alpha}$ test:

$$W_{XY} = W_R - \frac{m(m+1)}{2}$$

$$\mathbb{E} W_{XY} = \mathbb{E} W_R - \frac{m(m+1)}{2}$$

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

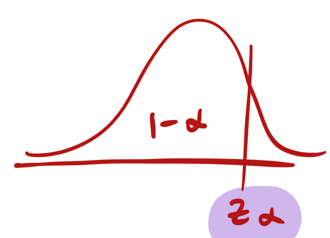
$$= \frac{mn}{2}$$

$$n = N - m$$

$$\alpha = 1 - \Phi\left(\frac{c_\alpha - \mathbb{E}_{\Delta=0} W_{XY}}{\sqrt{\operatorname{Var}_{\Delta=0} W_{XY}}}\right)$$

$$= 1 - \Phi\left(\frac{c_\alpha - \frac{nm}{2}}{\sqrt{\frac{1}{12} m(N-m)(N+1)}}\right)$$

$$\iff 1 - \alpha = \Phi\left(\;''\;\right)$$



$r = s$

$$Z_\alpha = \frac{c_\alpha - \frac{nm}{2}}{\sqrt{\frac{1}{12} m (N-m)(N+1)}}$$

$$\Leftrightarrow \quad c_\alpha = \frac{nm}{2} + Z_\alpha \sqrt{\frac{1}{12} m (N-m)(N+1)} \,.$$

$p_1(\Delta) = P(X_1 \le Y_1)$

$= P(X_1 - \Delta \le Y_1 - \Delta)$

$= P(X_1 - \Delta \le X_1')$

$= P(X_1 - X_1' \le \Delta)$

$= F^*(\Delta)$

$X_1 - X_1' \sim F^*$

$X_1 - X_1' \sim f^*$

symm. around 0.

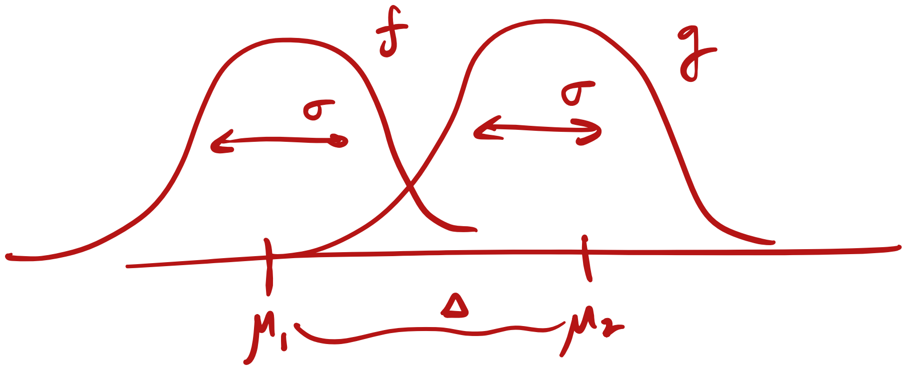
Exercise: Show that making the substitutions

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

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

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

$F^*(\Delta) \approx \underbrace{F^*(0)}_{= 1/2} + f^*(0)\,\Delta$



$f^*(0) = \frac{1}{2\sqrt{\pi}} \frac{1}{\sigma}$

Exercise:

1. Show that if F is Normal, $\boxed{n = m,}$ 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 $\geq \gamma^*$ for all $\Delta \geq \Delta^*$.
3. Compare to n needed for the equal-variances two-sample z-test.

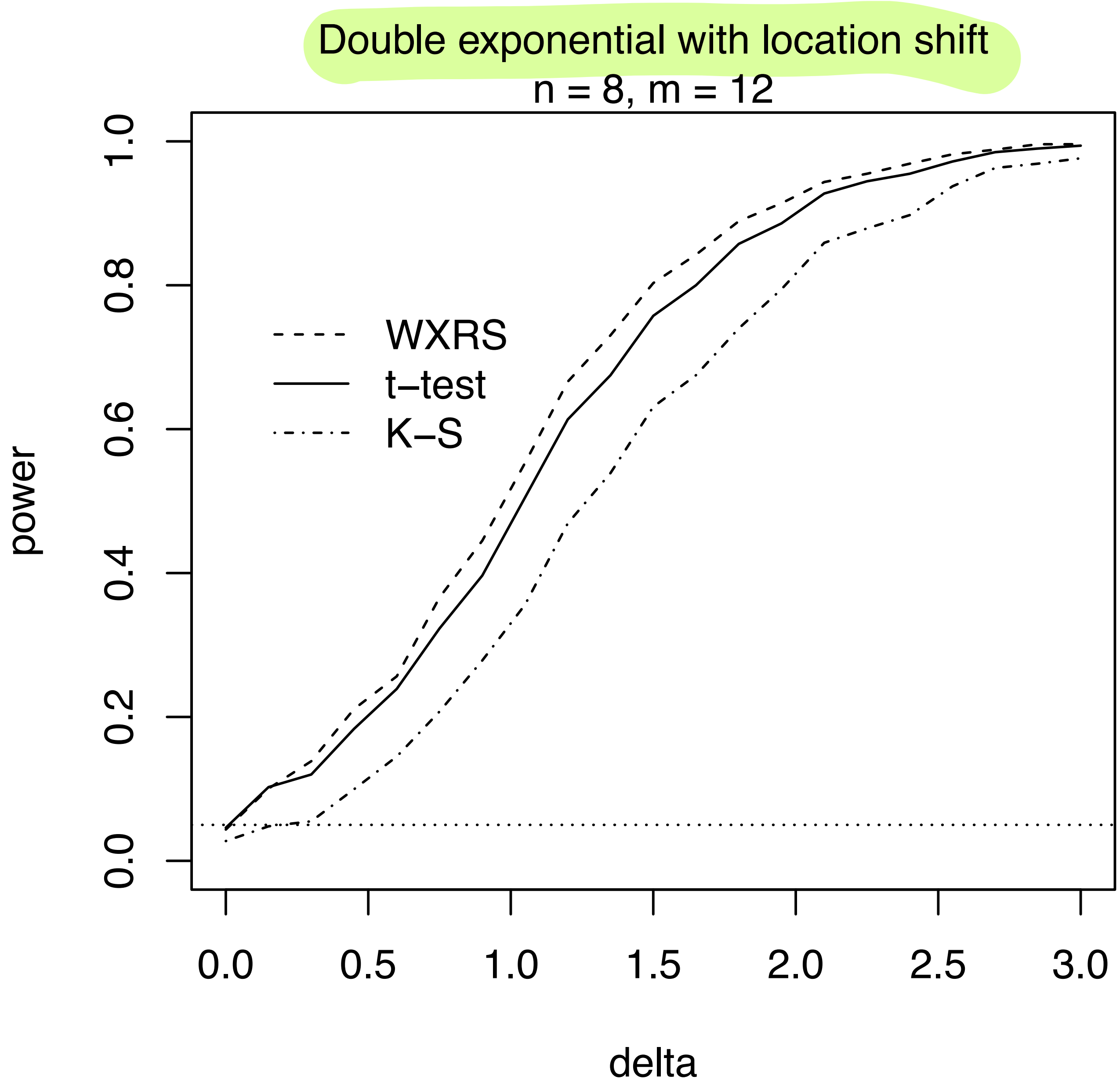
Take

$$n \geq \underbrace{\frac{\pi}{3}}_{\approx 1.05} 2\sigma^2 \left(\frac{z_\alpha + z_{\beta^*}}{\Delta^*}\right)^2$$

$\beta^* = 1 - \gamma^*$

For two-sample t-test (z-test), need

$$n \geq 2\sigma^2 \left(\frac{z_\alpha + z_{\beta^*}}{\Delta^*}\right)^2$$


Double exponential with location shift
n = 8, m = 12
WXRS
t–test
K–S
power
delta
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