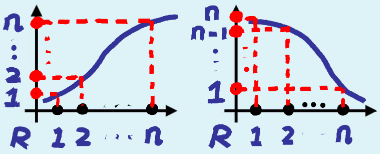
- Ranks are invariant under any monotonic transformation of data, i.e.,

$$R(X_1,\ldots,X_n)=R(H(X_1),\ldots,H(X_n)),$$

if H is a monotone increasing function and

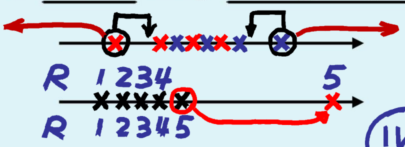
$$R(X_1,\ldots,X_n)=(n+1)-R(H(X_1),\ldots,H(X_n)),$$

if H is a monotone decreasing function. ($\overset{\text{cf.}}{\longleftrightarrow}$ z- or t-tests may change significantly under monotonic transformations of data).



$(X_1,\cdots,X_n) \to (R_1,\cdots,R_n)$

- ⦿ Replacing the data by their ranks also has the effect of moderating the influence of outliers.



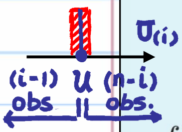
check graphs in LNp.25

- Many nonparametric methods are based on order statistics and/or ranks. 1/1
- **Q**: Why are many nonparametric methods based on replacement of the data by ranks? What information of data are contained in their ranks?

nondecreasing

TBp.63 (exercise) – **Recall**. Let $X_1,\ldots,X_n$ be i.i.d. from a continuous cdf F, and let $U_i=F(X_i)$, $i=1,\ldots,n$. Then, $U_1,\ldots,U_n$ are i.i.d. from $U(0,1)$.

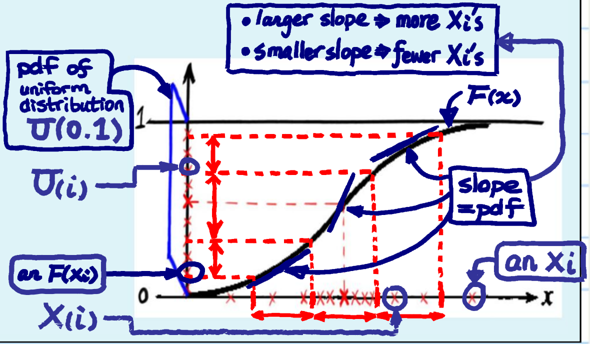
TBp.105 (exercise) – **Recall**. If $U_1,\ldots,U_n \sim$ i.i.d. $U(0,1)$, the pdf of the ith-order statistic $U_{(i)}$ is



$$f_{U_{(i)}}(u)=\frac{n!}{(i-1)!(n-i)!}u^{i-1}(1-u)^{n-i},$$

Beta(i, $n-i+1$)

for $0<u<1$ and zero, otherwise. Note that $E(U_{(i)})=i/(n+1)$. rank



– $U_i=F(X_i)$ is not a statistic because F is an unknown function.

– But, statistics, data

Review3 (LNp.28)

$$X_i = X_{(R_i)} \to U_{(R_i)} = F(X_{(R_i)}) \to R_i=(n+1)\frac{R_i}{n+1} \leftrightarrow (n+1)E[\,U_{(R_i)}|\,R_i].$$

unknown distribution

cf. pivotal quantity

cf.

Question 6.

How to use ranks to compare two samples? Under the nonparametric model (☐) in LNp.27, for the null and alternative hypotheses:

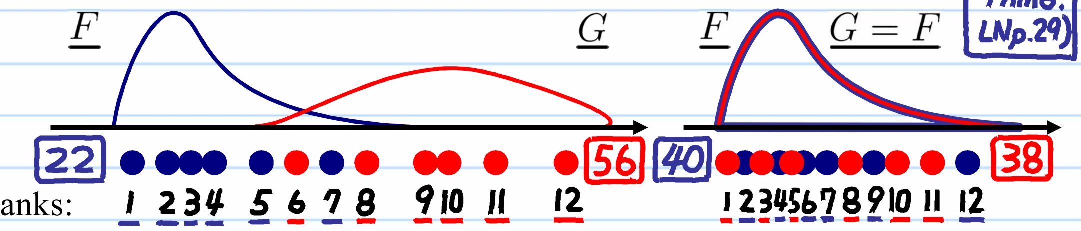
$$H_0: F=G \quad \text{vs.} \quad H_A: F\neq G$$

what data are "more extreme," i.e., cast more doubts on H_0?

Q: Why ranks are useful in this case? (cf., the information of ranks is useless in one sample case. Thm8, LNp.29)

Intuition.

$1+2+\cdots+12 = \frac{12\times 13}{2} = 78$



Theorem 9 (Mann-Whitney test or Wilcoxon rank sum test)

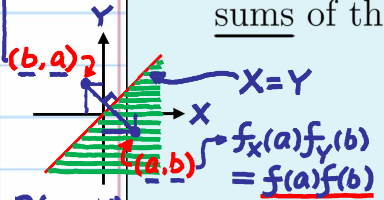
Consider the nonparametric model (☐) in LNp.27.

- Pool all $m+n$ observations (i.e., $X_1,\ldots,X_n,Y_1,\ldots,Y_m$) together and rank them in order of increasing size, i.e.,

X ← | → Y

$$R(X_1,\ldots,X_n,Y_1,\ldots,Y_m)=(R_1,\ldots,R_n,R_{n+1},\ldots,R_{m+n}).$$

$f_X(b)f_Y(a) = f(b)f(a)$



$f_X(a)f_Y(b) = f(a)f(b)$

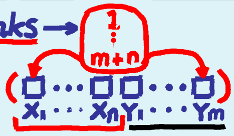
$P(X<Y) = P(X>Y) = 1/2$

- Test statistic W_X (or W_Y)
  - Let $W_X = \sum_{i=1}^{n} R_i$ and $W_Y = \sum_{j=1}^{m} R_{n+j}$. They are respectively the sums of the ranks of X_i's and Y_j's in the pooled data. Notice that

$$W_X + W_Y = 1 + 2 + \cdots + (m+n) = \frac{(m+n)(m+n+1)}{2}$$

$$\Rightarrow W_Y = \frac{(m+n)(m+n+1)}{2} - W_X.$$

  - Data with larger or smaller W_X are more extreme ⇒ tend to reject H_0
- Null distribution of W_X
  - Under H_0 ($F = G$),

ranks →


1-sample model with m+n observations

Why irrelevant to F? Check Thm 8 (LN p.29) ⇒ ranks carry no information of F

$$\begin{array}{ccccccccc} X_1, & \ldots, & X_n, & Y_1, & \ldots, & Y_m & \sim & \text{i.i.d. } F \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & & \\ R_1, & \ldots, & R_n, & R_{n+1} & \ldots, & R_{m+n} & \sim & ? \end{array}$$

cf

equal probability $1/(m+n)!$

X	Y		Prob.
(R_1, R_2)			
(1, 2)	← X<Y		1/2
(2, 1)	← X>Y		1/2

  - Any assignments of the ranks $\{1, \ldots, m+n\}$ to the pooled $m+n$ data are equally likely, and the total number of different assignments is $(m+n)!$.
  - Joint distribution of $R_1, \ldots, R_n$:

$R(X_1, X_2, Y_1) = (R_1, R_2, R_3)$	$a<b<c$	Prob.
(1, 2, 3)	← $X_1 < X_2 < Y_1$	1/6
(1, 3, 2)	← $X_1 < Y_1 < X_2$	1/6
(2, 1, 3)	← $X_2 < X_1 < Y_1$	1/6
(2, 3, 1)	← $Y_1 < X_1 < X_2$	1/6
(3, 1, 2)	← $X_2 < Y_1 < X_1$	1/6
(3, 2, 1)	← $Y_1 < X_2 < X_1$	1/6

    - ∗ Consider an urn containing $m+n$ balls, labelled by $1, 2, \ldots, m+n$, respectively.
    - ∗ Sequentially draw n balls without replacement from the urn ⇒ there are $\binom{m+n}{n} \times n!$ different outcomes, each with equal probability

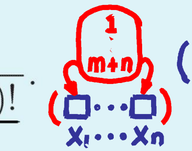
$Ri = \begin{cases} 1, & 1/m{+}n \\ 2, & 1/m{+}n \\ \vdots & \vdots \\ m{+}n, & 1/m{+}n \end{cases}$ Prob

$E(Ri) = \frac{\cancel{(m+n)}(m+n+1)}{2\,\cancel{(m+n)}}$

- ∗ Let $r_1, r_2, \ldots, r_n$ be the numbers on the 1st, 2nd, …, nth balls drawn, respectively. Then, all permutations of $R_{n+1}, \cdots, R_{n+m}$

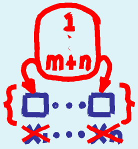
$$P(R_1 = r_1, \ldots, R_n = r_n) = \frac{1}{\binom{m+n}{n} \times n!} = \frac{m!}{(m+n)!}.$$

← $(m+n-1)!/(m+n)!$



$\binom{m+n}{n} \times n!$

- The null distribution of $W_X = R_1 + \cdots + R_n$ (W_X is the sum of the numbers on the n balls) can be obtained from the joint distribution of $R_1, \ldots, R_n$.



$\binom{m+n}{n}$ ← $R_{(1)}, \cdots, R_{(n)}$

- Rejection region ← 2-sided test ⇒ reject if W_X small (↔ W_Y large) or W_X large (↔ W_Y small)
  - Let $n_1 = \min(n, m)$ be the smaller sample size, and W be the rank sum from that sample (i.e., $W = W_X$ if $n \le m$ and $W = W_Y$ if $n > m$).
    - ∗ Note that under H_0,
      - · $E(W) = \begin{cases} E(R_1) + \cdots + E(R_n), & \text{if } n \le m \\ E(R_{n+1}) + \cdots + E(R_{n+m}), & \text{if } n > m \end{cases} = \frac{n_1(m+n+1)}{2}$. (n m)
      - · the null distribution of W is symmetric around $E(W)$ (exercise).
    - ∗ Let $W' = n_1(m+n+1) - W$. → $\frac{W+W'}{2} = \frac{n_1(m+n+1)}{2}$
    - ∗ Let $W^* = \min(W, W')$. cf. $|Z|$ $|T|$
  - Reject H_0 when W^* is small, i.e., $W^* \le w$.
  - Table 8 of Appendix B in the textbook gives critical values w for W^*.

For example, $n<m$, $n=3$, {□□□}

$W_X = 6$

1 = 1+0	m+n-0
2 = 1+1	m+n-1
3 = 1+2	m+n-2

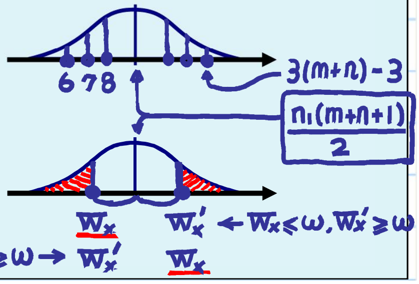
$W_X = 7$

1 = 1+0	m+n-0
2 = 1+1	m+n-1
4 = 1+3	m+n-3

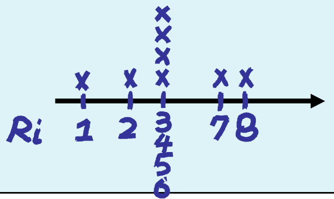
$W_X = 8$

1 = 1+0	m+n-0
3 = 1+2	m+n-2
4 = 1+3	m+n-3

1 = ⋯
2 = ⋯
5 = ⋯



- We have assumed here that there are no ties among the observations. If there are only a small number of ties; tied observations are assigned average ranks.



$$\frac{3+4+5+6}{4} = 4.5$$

Example 4 (Mann-Whitney test, heat of fusion of ice, cont. Ex.1 in LNp.3)

- The ranks are (ties ⇒ average rank)

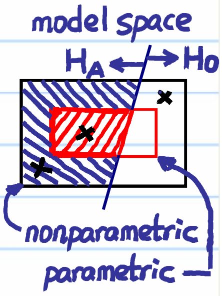
$n=13$	Method A	7.5	19.0	11.5	19.0	15.5	15.5	19.0	4.5	
∨		21.0	15.5	11.5	9.0	11.5				
$m=8$	Method B	11.5	1.0	7.5	4.5	4.5	15.5	2.0	4.5	← Sum=51

- $n_1 = 8$, $W = W_B = 51$, $W' = 8(8+13+1) - W = 125$, $W^* = \min(W, W') = 51$
- two-sided test at level $\alpha = 0.01$, critical value $= 53$
  two-sided test at level $\alpha = 0.05$, critical value $= 60$ } Table 8 (TBp.A21)
- Therefore, the Mann-Whitney test rejects the null hypothesis at $\alpha = 0.01$.

a comparison of parametric and nonparametric models

	model space	data reduction	robustness	power on H_A^p	power on $H_A^{np} \setminus H_A^p$
parametric models	small	low-dim	worse	higher	(usually) lower
nonparametric models	large	high-dim	better	lower	(usually) higher

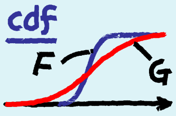
parametric
nonparametric
dim<∞
dim=∞
e.g., sensitive to outlier

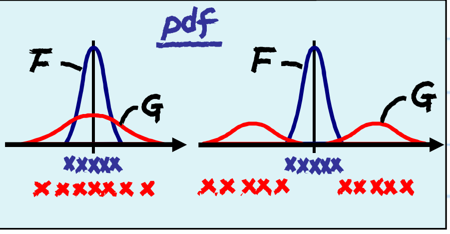


Question 7.

Does Mann-Whitney test have reasonably good powers over the whole $H_A : F \neq G$? Note that

$$H_0 \cup H_A = \{(F,G) \mid F, G \in \Omega\},$$
$$H_0 = \{(F,G) \mid F \in \Omega, G = F\}.$$



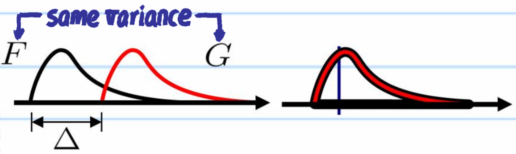


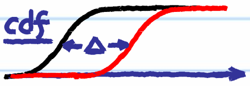
⊙ Assume that the distributions (cdfs) $F, G \in \Omega$ and F, G have same shape.

– If $X \sim F$ and $Y = X + \Delta$, where Δ is an unknown constant, then for the cdf $G(y)$ of Y, we have

check LNp.2 case (b)

a parameter



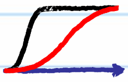


$$G(y) \equiv P(Y \leq y) = P(X + \Delta \leq y) = P(X \leq y - \Delta) = F(y - \Delta),$$

and for the pdfs $f(x)$ of X and $g(y)$ of Y, we have

$$g(y) = \frac{d}{dy} G(y) = \frac{d}{dy} F(y - \Delta) = f(y - \Delta).$$

Alternative setup for $F \neq G$: G is stochastically larger ($G(x) < F(x)$) or smaller ($G(x) > F(x)$) than F.



model (□) in LNp.27

cf.

- Thus, the statistical model is:

$$\left.\begin{array}{lll} \text{1st sample:} & X_1, \ldots, X_n & \sim \text{ i.i.d. from } F \\ \text{2nd sample:} & Y_1, \ldots, Y_m & \sim \text{ i.i.d. from } G \end{array}\right\} \Leftarrow \text{independent} \qquad (\Diamond)$$

where $F \in \Omega$ and $G(x) = F(x - \Delta)$.

Δ (dim=1): parameter of main interest
shape of F (dim=∞): nuissance parameter

information about the shape of F is in order statistics, not ranks

– This model contains *infinitely many* parameters because $\dim(\Omega) = \infty$.

- Under this model, the null $H_0 : F = G$ become $H_0 : \Delta = 0$, and the alternative $H_A : F \neq G$ becomes $H_A : \Delta \neq 0$, i.e.,

$$\circledast \quad H_0 \cup H_A = \{(F, G) \,|\, F \in \Omega,\ G(y) = F(y-\Delta),\ \Delta \in \mathbb{R} \ (\text{or } \pi_\Delta \in [0,1])\}$$

$$\circledast \quad H_0 = \{(F, G) \,|\, F \in \Omega,\ G(y) = F(y-\Delta),\ \Delta = 0 \ (\text{or } \pi_\Delta = 1/2)\}$$

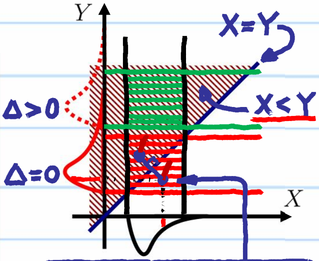
$H_A: \Delta \neq 0$ $(\pi_\Delta \neq 1/2)$

Theorem 10 (An alternative formulation of H_0 and H_A)

1 observation from F, 1 observation from G

- Suppose that (1) $X \sim F \in \Omega$, (2) $Y \sim G$, where $G(x) = F(x-\Delta)$, and (3) X, Y are independent. The joint pdf of (X, Y) is $f(x)g(y) = f(x)f(y-\Delta)$.
- Define $\pi_\Delta = P_\Delta(X < Y)$. Clearly, $0 \leq \pi_\Delta \leq 1$.
- Then, $\pi_\Delta = 1/2$ if and only if $\Delta = 0$.

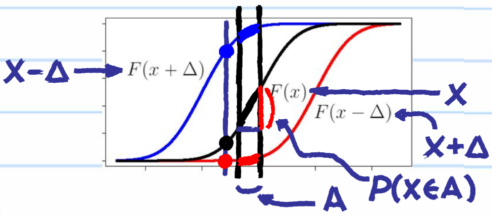
what if Δ<0?



check graph in LNp.32

Proof.

(*) $$P_\Delta(X < Y) = \int_{-\infty}^{\infty}\int_{x}^{\infty} f(x)\, f(y-\Delta)\, dy\, dx$$



$$= \int_{-\infty}^{\infty} f(x)\, \left[F(y-\Delta)\big|_{x}^{\infty}\right] dx$$

$$= \int_{-\infty}^{\infty} [1 - F(x-\Delta)]\, f(x)\, dx = 1 - \int_{-\infty}^{\infty} F(x-\Delta)\, f(x)\, dx$$

$P_{\Delta=0}(X\in A)$

It is impossible that they equal for any x

- If $\Delta > 0$, then $F(x-\Delta) \leqslant F(x) \leqslant F(x+\Delta)$, $\forall\, x$, and there must exist a region A of x in which the inequalities are strict and $\int_A f(x)\, dx > 0$.

What if Δ<0?

- Thus, for $\Delta > 0$,

$$\underbrace{\int_{-\infty}^{\infty} F(x-\Delta)\, f(x)\, dx}_{1 - P_\Delta(X<Y)} < \underbrace{\int_{-\infty}^{\infty} F(x)\, f(x)\, dx}_{1 - P_{\Delta=0}(X<Y)} < \underbrace{\int_{-\infty}^{\infty} F(x+\Delta)\, f(x)\, dx}_{1 - P_{-\Delta}(X<Y)}$$

By (*) in LNp.36

$\boxed{>1/2}$ $\boxed{=1/2}$ $\boxed{<1/2}$

- Then, the results follow from: $\displaystyle\int_{-\infty}^{\infty} F(x)\, f(x)\, dx = \int_0^1 z\, dz = \frac{1}{2} z^2 \Big|_0^1 = \frac{1}{2}.$

Let $z = F(x) \Rightarrow dz/dx = f(x)$

Theorem 11 (An alternative view of Mann-Whitney test)

$X_1, \ldots, X_n$ i.i.d. X

$Y_1, \ldots, Y_m$ i.i.d. Y

Consider the nonparametric model $(\Diamond)$ in LNp.35.

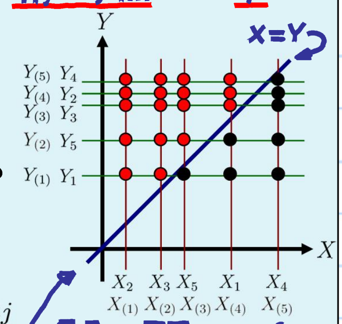
- Estimation of π_Δ: the parameter $\pi_\Delta = P_\Delta(X < Y)$ can be estimated by the proportion of the comparisons for which X was less than Y, i.e.,

  mn = # of ● & ●

  - consider any pairs (X_i, Y_j), $1 \leq i \leq n$, $1 \leq j \leq m$,
  - let $Z_{ij} = \begin{cases} 1, & \text{if } X_i < Y_j, \\ 0, & \text{otherwise}, \end{cases} \Rightarrow \hat{\pi}_\Delta = \dfrac{1}{mn}\displaystyle\sum_{i=1}^{n}\sum_{j=1}^{m} Z_{ij}$

    $\text{Bin}(1, \pi_\Delta) \sim$

    $\boxed{\text{not independent}}$ ← Note that $Z_{11}, \ldots, Z_{n1} | Y_1 = y \overset{\text{i.i.d.}}{\sim} \text{Bin}(1, F(y))$



cf. $\sum Z_{ij} = \sum V_{ij}$ = # of ● graph in Thm10 (LNp.36)

  - an alternative expression: consider the mn pairs $(X_{(i)}, Y_{(j)})$, and let

    $V_{ij} = \begin{cases} 1, & \text{if } X_{(i)} < Y_{(j)}, \\ 0, & \text{otherwise}, \end{cases} \Rightarrow \hat{\pi}_\Delta = \dfrac{1}{mn}\displaystyle\sum_{i=1}^{n}\sum_{j=1}^{m} V_{ij}$

    $\text{Bin}(1, \cancel{\pi_\Delta}) \sim$

$$\begin{aligned} \hat{\pi}_\Delta &= \tfrac{1}{mn}(\#\{X_i < Y_j\}) \\ &= \tfrac{1}{mn}(\#\{X_{(i)} < Y_{(j)}\}) \end{aligned}$$

e.g., $P(X_{(1)} < Y_{(j)}) > P(X_{(n)} < Y_{(j)})$