Nonparametric hypothesis tests and permutation tests

1.7 & 2.3. Probability Generating Functions
3.8.3. Wilcoxon Signed Rank Test
3.8.2. Mann-Whitney Test

Prof. Tesler

Math 283
Fall 2018

Probability Generating Functions (pgf)

- Let Y be an integer-valued random variable with a lower bound (typically $Y \geqslant 0$).
- The *probability generating function* is defined as

$$\mathbb{P}_Y(t) = E(t^Y) = \sum_y P_Y(y) t^y$$

Simple example

Suppose $P_X(x) = x/10$ for $x = 1, 2, 3, 4$, $P_X(x) = 0$ otherwise. Then

$$\mathbb{P}_X(t) = .1t + .2t^2 + .3t^3 + .4t^4$$

Poisson distribution

Let X be Poisson with mean μ. Then

$$\mathbb{P}_X(t) = \sum_{k=0}^{\infty} \frac{e^{-\mu}\mu^k}{k!} \cdot t^k = \sum_{k=0}^{\infty} \frac{e^{-\mu}(\mu\, t)^k}{k!} = e^{-\mu} e^{\mu\, t} = e^{\mu(t-1)}$$

Properties of pgfs

- Plugging in $t = 1$ gives total probability=1:

$$\mathbb{P}_Y(1) = \sum_y P_Y(y) = 1$$

- Differentiating and plugging in $t = 1$ gives $E(Y)$:

$$\mathbb{P}_Y'(t) = \textstyle\sum_y P_Y(y) \cdot y\, t^{y-1}$$
$$\mathbb{P}_Y'(1) = \textstyle\sum_y P_Y(y) \cdot y = E(Y)$$

- Variance is $\mathrm{Var}(Y) = \mathbb{P}_Y''(1) + \mathbb{P}_Y'(1) - (\mathbb{P}_Y'(1))^2$:

$$\mathbb{P}_Y''(t) = \textstyle\sum_y P_Y(y) \cdot y(y-1)\, t^{y-2}$$
$$\mathbb{P}_Y''(1) = \textstyle\sum_y P_Y(y) \cdot y(y-1) = E(Y(Y-1)) = E(Y^2) - E(Y)$$
$$\mathrm{Var}(Y) = E(Y^2) - (E(Y))^2 = \mathbb{P}_Y''(1) + \mathbb{P}_Y'(1) - (\mathbb{P}_Y'(1))^2$$

Example of pgf properties: Poisson

Properties

$$\begin{aligned}
\mathbb{P}_Y(t) &= \textstyle\sum_y P_Y(y)t^Y \\
\mathbb{P}_Y(1) &= 1 \\
E(Y) &= \mathbb{P}_Y'(1) \\
\mathrm{Var}(Y) &= E(Y^2) - (E(Y))^2 = \mathbb{P}_Y''(1) + \mathbb{P}_Y'(1) - (\mathbb{P}_Y'(1))^2
\end{aligned}$$

- For X Poisson with mean μ, we saw $\mathbb{P}_X(t) = e^{\mu(t-1)}$.
- $\mathbb{P}_X(1) = e^{\mu(1-1)} = e^0 = 1$
- $\mathbb{P}_X'(t) = \mu\, e^{\mu(t-1)}$ and $\mathbb{P}_X'(1) = \mu\, e^{\mu(1-1)} = \mu$
  Indeed, $E(X) = \mu$ for Poisson.
- $\mathbb{P}_X''(t) = \mu^2\, e^{\mu(t-1)}$
  $\mathbb{P}_X''(1) = \mu^2\, e^{\mu(1-1)} = \mu^2$
  $\mathrm{Var}(X) = \mathbb{P}_X''(1) + \mathbb{P}_X'(1) - (\mathbb{P}_X'(1))^2 = \mu^2 + \mu - \mu^2 = \mu$
  Indeed, $\mathrm{Var}(X) = \mu$ for Poisson.

Probability generating function of $X + Y$

Consider adding rolls of two biased dice together:

$$X = \text{roll of biased 3-sided die}$$
$$Y = \text{roll of biased 5-sided die}$$

$$\begin{aligned}
P(X+Y=2) &= P_X(1)P_Y(1) \\
P(X+Y=3) &= P_X(1)P_Y(2) + P_X(2)P_Y(1) \\
P(X+Y=4) &= P_X(1)P_Y(3) + P_X(2)P_Y(2) + P_X(3)P_Y(1) \\
P(X+Y=5) &= P_X(1)P_Y(4) + P_X(2)P_Y(3) + P_X(3)P_Y(2) \\
P(X+Y=6) &= P_X(1)P_Y(5) + P_X(2)P_Y(4) + P_X(3)P_Y(3) \\
P(X+Y=7) &= P_X(2)P_Y(5) + P_X(3)P_Y(4) \\
P(X+Y=8) &= P_X(3)P_Y(5)
\end{aligned}$$

Probability generating function of $X + Y$

$$\mathbb{P}_X(t) = P_X(1)t + P_X(2)t^2 + P_X(3)t^3$$
$$\mathbb{P}_Y(t) = P_Y(1)t + P_Y(2)t^2 + P_Y(3)t^3 + P_Y(4)t^4 + P_Y(5)t^5$$

$$\begin{aligned}
\mathbb{P}_X(t)\mathbb{P}_Y(t) = \quad & \big(P_X(1)P_Y(1)\big)t^2 \; + \\
& \big(P_X(1)P_Y(2) + P_X(2)P_Y(1)\big)t^3 \; + \\
& \big(P_X(1)P_Y(3) + P_X(2)P_Y(2) + P_X(3)P_Y(1)\big)t^4 \; + \\
& \big(P_X(1)P_Y(4) + P_X(2)P_Y(3) + P_X(3)P_Y(2)\big)t^5 \; + \\
& \big(P_X(1)P_Y(5) + P_X(2)P_Y(4) + P_X(3)P_Y(3)\big)t^6 \; + \\
& \big(P_X(2)P_Y(5) + P_X(3)P_Y(4)\big)t^7 \; + \\
& \big(P_X(3)P_Y(5)\big)t^8 \\
= \quad & P(X+Y=2)t^2 \; + \; \cdots \; + \; P(X+Y=8)t^8 \\
= \quad & \mathbb{P}_{X+Y}(t)
\end{aligned}$$

Probability generating function of $X + Y$

Suppose X and Y are independent random variables. Then

$$\mathbb{P}_{X+Y}(t) = \mathbb{P}_X(t) \cdot \mathbb{P}_Y(t)$$

Proof.

$$\mathbb{P}_{X+Y}(t) = E(t^{X+Y}) = E(t^X\, t^Y) = E(t^X)E(t^Y) = \mathbb{P}_X(t)\mathbb{P}_Y(t) \qquad \square$$

Second proof.

- $$\mathbb{P}_X(t) \cdot \mathbb{P}_Y(t) = \left(\textstyle\sum_x P(X = x)t^x\right)\left(\textstyle\sum_y P(Y = y)t^y\right)$$
- Multiply that out and collect by powers of t. The coefficient of t^w is
  $$\textstyle\sum_x P(X = x)P(Y = w - x)$$
- Since X, Y are independent, this simplifies to $P(X + Y = w)$, which is the coefficient of t^w in $\mathbb{P}_{X+Y}(t)$. $\square$

Binomial distribution

- Suppose $X_1, \ldots, X_n$ are i.i.d. with $P(X_i = 1) = p$, $P(X_i = 0) = 1 - p$ *(Bernoulli distribution)*.
- $\mathbb{P}_{X_i}(t) = (1-p)t^0 + pt^1 = 1 - p + pt$
- The Binomial(n, p) distribution is $X = X_1 + \cdots + X_n$.
- $\mathbb{P}_X(t) = \mathbb{P}_{X_1}(t) \cdots \mathbb{P}_{X_n}(t) = (1 - p + pt)^n$
- Check:

$$((1-p)+pt)^n = \sum_{k=0}^{n} \binom{n}{k} (1-p)^{n-k} p^k \cdot t^k = \sum_{k=0}^{n} P_Y(k) t^k$$

  where Y is the Binomial(n, p) distribution.
- Note: If X and Y have the same pgf, then they have the same distribution.

Moment generating function (mgf) in Chapter 1.1 & 2.3

- Let Y be a continuous or discrete random variable.
- The *moment generating function* (mgf) is $\mathbb{M}_Y(\theta) = E(e^{\theta Y})$.
- Discrete: Same as the pgf with $t = e^{\theta}$, and not just for integer-valued variables:
$$\mathbb{M}_Y(\theta) = \sum_y P_Y(y) e^{\theta y}$$
- Continuous: It's essentially the "2-sided Laplace transform" of $f_Y(y)$:
$$\mathbb{M}_Y(\theta) = \int_{-\infty}^{\infty} f_Y(y) e^{\theta y}\, dy$$
- The derivative tricks for pgf have analogues for mgf:
$$\frac{d^k}{d\theta^k}\mathbb{M}_Y(\theta) = E(Y^k\, e^{\theta Y})$$
$$\mathbb{M}_Y^{(k)}(0) = E(Y^k) = k\text{th moment of } Y$$

$$\mathbb{M}_Y(0) = E(1) = 1 = \text{Total probability}$$
$$\mathbb{M}_Y'(0) = E(Y) = \text{Mean}$$
$$\mathbb{M}_Y''(0) = E(Y^2) \quad \text{so} \quad \mathrm{Var}(Y) = \mathbb{M}_Y''(0) - (\mathbb{M}_Y'(0))^2$$

Non-parametric hypothesis tests

- *Parametric* hypothesis tests assume the random variable has a specific probability distribution (normal, binomial, geometric, ...). The competing hypotheses both assume the same type of distribution but with different parameters.
- A *distribution free* hypothesis test (a.k.a. *non-parametric* hypothesis test) doesn't assume any particular type of distribution. So it can be applied even if the distribution isn't known.
- If the type of distribution is known, a parametric test that takes it into account can be more precise (smaller Type II error for same Type I error) than a non-parametric test that doesn't.

Wilcoxon Signed Rank Test

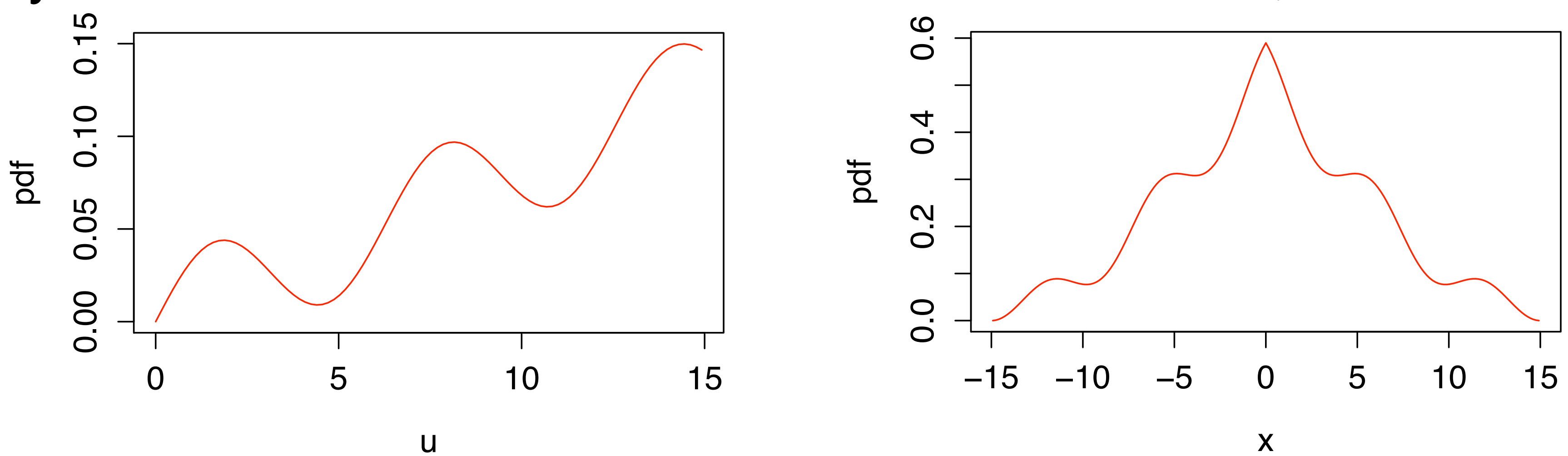
- Let X be a continuous random variable with a symmetric distribution.
- Let M be the median of X:

$$P(X > M) = P(X < M) = 1/2, \text{ or } F_X(M) = .5.$$

- Note that if the pdf of X is symmetric, the median equals the mean. If it's not symmetric, they usually are not equal.
- We will develop a test for

$$H_0 : M = M_0 \qquad \text{vs.} \qquad H_1 : M \neq M_0 \text{ (or } M < M_0 \text{ or } M > M_0\text{)}$$

  based on analyzing a sample $x_1, \ldots, x_n$ of data.
- Example: If U, V have the same distribution, then $X = U - V$ has a symmetric distribution centered around its median, 0.



Computing the Wilcoxon test statistic

Is median $M_0 = 5$ plausible, given data $1.1,\ 8.2,\ 2.3,\ 4.4,\ 7.5,\ 9.6$?

- Get a sample $x_1, \ldots, x_n$: $1.1,\ 8.2,\ 2.3,\ 4.4,\ 7.5,\ 9.6$
- Compute the following:
  - Compute each $x_i - M_0$.
  - Order $|x_i - M_0|$ from smallest to largest and assign ranks $1, 2, \ldots, n$ (1=smallest, n=largest).
  - Let r_i be the rank of $|x_i - M_0|$ and $z_i = \begin{cases} 0 & \text{if } x_i - M_0 < 0 \\ 1 & \text{if } x_i - M_0 > 0. \end{cases}$

    Note: Since X is continuous, $P(X - M_0 = 0) = 0$.
- Compute test statistic $w = z_1 r_1 + \cdots + z_n r_n$ (sum of r_i's with $x_i > M_0$)

i	x_i	$x_i - M_0$	r_i	sign	z_i
1	1.1	-3.9	5	$-$	0
2	8.2	3.2	4	$+$	1
3	2.3	-2.7	3	$-$	0
4	4.4	$-.6$	1	$-$	0
5	7.5	2.5	2	$+$	1
6	9.6	4.6	6	$+$	1

$n = 6$

$|x_i - M_0|$ in order:
$.6,\ 2.5,\ 2.7,\ 3.2,\ 3.9,\ 4.6$

$w = 4 + 2 + 6 = 12$

Computing the pdf of W

- The variable whose rank is i contributes either 0 or i to W. Under the null hypothesis, both of those have probability $1/2$. Call this contribution W_i, either 0 or i with prob. $1/2$. Then
$$W = W_1 + \cdots + W_n$$
- The W_i's are independent because the signs are independent.
- The pgf of W_i is
$$\mathbb{P}_{W_i}(t) = E(t^{W_i}) = \frac{1}{2}t^0 + \frac{1}{2}t^i = \frac{1+t^i}{2}$$
- The pgf of W is
$$\mathbb{P}_W(t) = \mathbb{P}_{W_1+\cdots+W_n}(t) = \mathbb{P}_{W_1}(t) \cdots \mathbb{P}_{W_n}(t) = 2^{-n} \prod_{i=1}^{n} (1+t^i)$$
- Expand the product. The coefficient of t^w is $P(W=w)$, the pdf of W.

Distribution of W for $n = 6$

- $$\begin{aligned}\mathbb{P}_W(t) &= \tfrac{1}{2^6}\left(1+t^1\right)\left(1+t^2\right)\left(1+t^3\right)\left(1+t^4\right)\left(1+t^5\right)\left(1+t^6\right)\\ &= \tfrac{1}{64}\big(1+t+t^2+2\,t^3+2\,t^4+3\,t^5+4\,t^6+4\,t^7\\ &\qquad +4\,t^8+5\,t^9+5\,t^{10}+5\,t^{11}+5\,t^{12}+4\,t^{13}\\ &\qquad +4\,t^{14}+4\,t^{15}+3\,t^{16}+2\,t^{17}+2\,t^{18}+t^{19}+t^{20}+t^{21}\big)\end{aligned}$$
- Example: $P(W = 6) = 4/64 = 1/16 = .0625$

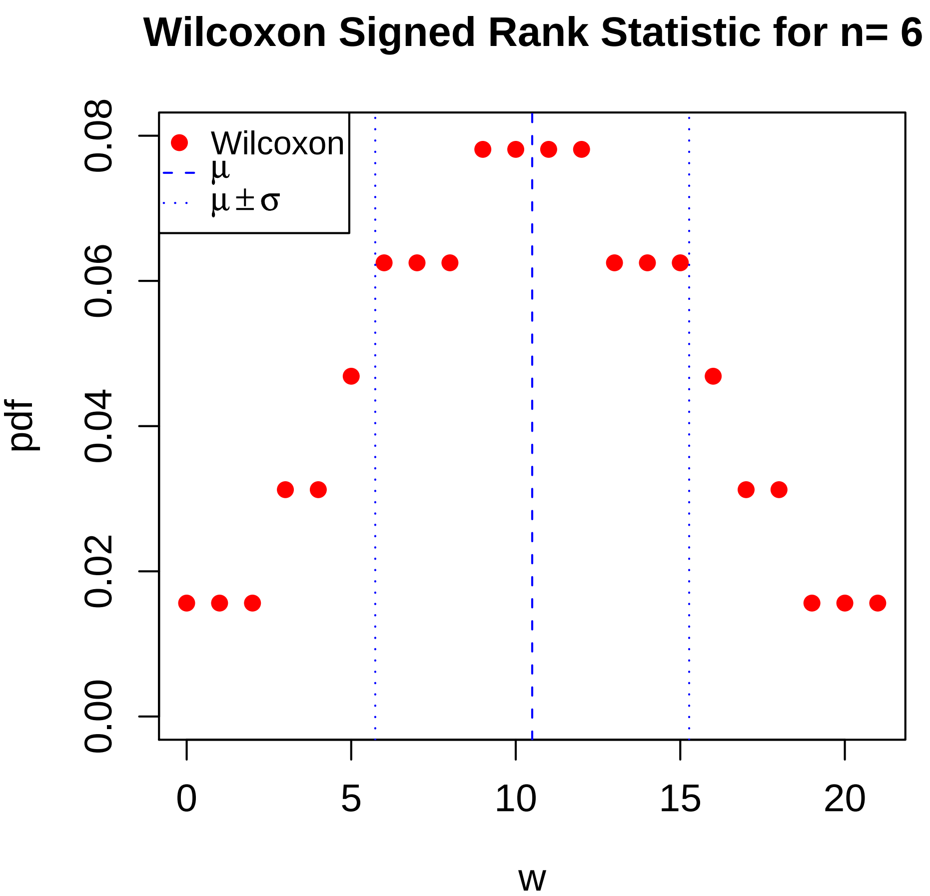
Cumulative distribution of W

w	$P(W \leqslant w)$	w	$P(W \leqslant w)$	w	$P(W \leqslant w)$
0	$1/64 = 0.015625$	8	$22/64 = 0.343750$	16	$57/64 = 0.890625$
1	$2/64 = 0.031250$	9	$27/64 = 0.421875$	17	$59/64 = 0.921875$
2	$3/64 = 0.046875$	10	$32/64 = 0.500000$	18	$61/64 = 0.953125$
3	$5/64 = 0.078125$	11	$37/64 = 0.578125$	19	$62/64 = 0.968750$
4	$7/64 = 0.109375$	12	$42/64 = 0.656250$	20	$63/64 = 0.984375$
5	$10/64 = 0.156250$	13	$46/64 = 0.718750$	21	$64/64 = 1.000000$
6	$14/64 = 0.218750$	14	$50/64 = 0.781250$		
7	$18/64 = 0.281250$	15	$54/64 = 0.843750$		

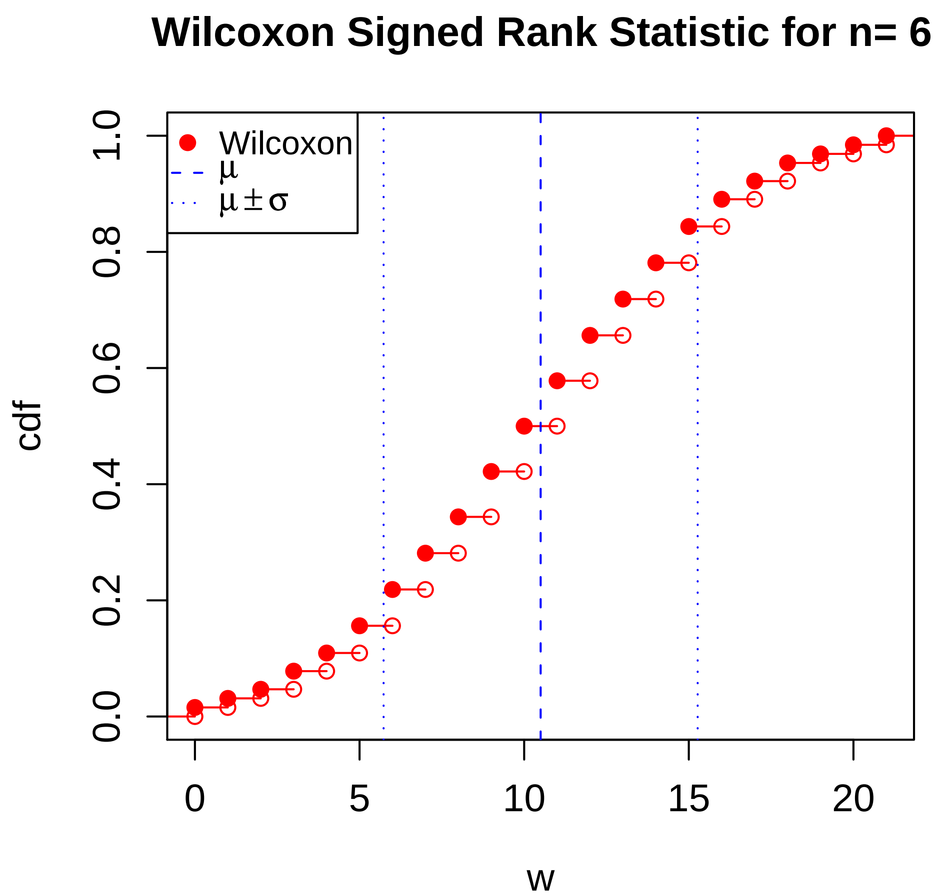
(The cdf is defined at all reals. It jumps at $w = 0, \ldots, 21$ and is constant in-between.)

Distribution of W for $n = 6$

PDF



CDF

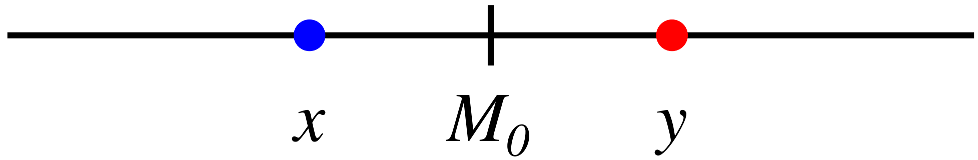


Properties of W (assuming H_0: $M = M_0$)

Range

- When all signs are negative, $w = 0 + 0 + \cdots = 0$.
- When all signs are positive, $w = 1 + 2 + \cdots + n = n(n+1)/2$.
- w ranges from 0 to $n(n+1)/2$.

Properties of W (assuming H_0: $M = M_0$)



Reflecting a point

Reflecting point x around M_0 gives $M_0 - x = y - M_0$, so $y = 2M_0 - x$.

Symmetry

If H_0 is correct, then reflecting all data in the sample around M_0 by setting $y_i = 2M_0 - x_i$ for all i:

- gives new values $y_1, \ldots, y_n$ equally probable to $x_1, \ldots, x_n$;
- keeps same magnitudes $|x_i - M_0| = |y_i - M_0|$ and same ranks;
- inverts all signs, switching whether a rank is / isn't included in w;
- sends w to $\frac{n(n+1)}{2} - w$.

So the pdf of W is symmetric about the center value $w = \frac{n(n+1)}{4}$.

Properties of W (assuming H_0: $M = M_0$)

Mean and variance

Mean: $E(W) = \frac{1}{4}n(n+1)$ **Variance:** $\mathrm{Var}(W) = \frac{1}{24}n(n+1)(2n+1)$

Central Limit Theorem

When $n > 12$, the Z-score of W is approximately standard normal:

$$Z = \frac{W - n(n+1)/4}{\sqrt{n(n+1)(2n+1)/24}} \qquad F_W(w) \approx \Phi(z) \text{ for } n > 12$$

- $W_1, W_2, \ldots$ are independent but not identically distributed.
- A generalization of CLT by Lyapunov applies; see “Lyapunov CLT” in the Central Limit Theorem article on Wikipedia.

Computing P-value

- Note that $P(W \geqslant w) = P(W \leqslant \frac{n(n+1)}{2} - w)$ by symmetry of the pdf.
  Let $w_1 = \min\left\{w, \frac{n(n+1)}{2} - w\right\}$ and $w_2 = \max\left\{w, \frac{n(n+1)}{2} - w\right\}$.
- Intuitively, w is close to $n(n+1)/4$ when H_0 is true, and much smaller or much larger when H_0 is false.
- **Two-sided test:** H_0: $M = 5$ vs. H_1: $M \neq 5$.
  Values "more extreme than w" are those farther away from $n(n+1)/4$ than w *in either direction*:
  $$P = P(W \leqslant w_1) + P(W \geqslant w_2) = 2P(W \leqslant w_1)$$
- In the example, $w = 12$ and $\frac{n(n+1)}{2} = \frac{6 \cdot 7}{2} = 21$, giving
  $P = P(W \geqslant 12) + P(W \leqslant 9) = 2P(W \leqslant 9) = 2(27/64) = 0.843750.$

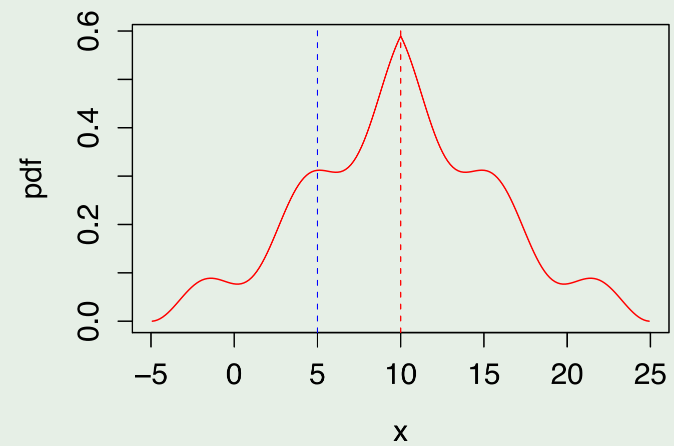
Performing the Wilcoxon Signed Rank Test

- Hypotheses: H_0: $M = 5$ vs. H_1: $M \neq 5$
- Choose a significance level α: $\alpha = 5\%$
- Get a sample $x_1, \ldots, x_n$: $1.1, 8.2, 2.3, 4.4, 7.5, 9.6$
- Compute test statistic w: $w = 12$
- Compute P-value: $P = 0.843750$
- Decision:
  Reject H_0 if $P \leqslant \alpha$.
  Accept H_0 if $P > \alpha$. $.843750 > .05$ so accept H_0.

One-sided tests

Example: Test H_0: $M = 5$ but true median=10

- $> \frac{1}{2}$ chance for $x_i - M = x_i - 5$ to be positive and $< \frac{1}{2}$ chance to be negative.
- This increases the chance of including each rank in the sum for W, and leads to higher values of W.



- **One-sided test:** H_0: $M = 5$ vs. H_1: $M > 5$.
  Higher medians lead to higher values of w, so values "more extreme than w" are $\geqslant w$:

$$P = P(W \geqslant w) = P(W \geqslant 12) = 1 - P(W \leqslant 11) = 27/64 = 0.421875$$

- **One-sided test:** H_0: $M = 5$ vs. H_1: $M < 5$.
  Lower medians lead to lower values of w, so values "more extreme than w" are $\leqslant w$:

$$P = P(W \leqslant w) = P(W \leqslant 12) = 42/64 = 0.656250$$

Computing w and P-value in Matlab or R

Matlab

```
>> x = [1.1,8.2,2.3,4.4,7.5,9.6];
>> M0 = 5;
>> signrank(x,M0)
    0.8438
>> [p,h,stats] = signrank(x,M0)
p = 0.8438
h = 0
stats =
    signedrank: 9

>> stats.signedrank
     9
```

Note $stats.signedrank = 9$ is our w_1, which is not necessarily w.

R

```
> x = c(1.1,8.2,2.3,4.4,7.5,9.6)
> test = wilcox.test(x,mu=5)
> test$statistic
 V
12
> test$p.value
[1] 0.84375
```

Critical region for a given significance level α

Cumulative distribution of W

w	$P(W \leqslant w)$	w	$P(W \leqslant w)$	w	$P(W \leqslant w)$
0	$1/64 = 0.015625$	8	$22/64 = 0.343750$	16	$57/64 = 0.890625$
1	$2/64 = 0.031250$	9	$27/64 = 0.421875$	17	$59/64 = 0.921875$
2	$3/64 = 0.046875$	10	$32/64 = 0.500000$	18	$61/64 = 0.953125$
3	$5/64 = 0.078125$	11	$37/64 = 0.578125$	19	$62/64 = 0.968750$
4	$7/64 = 0.109375$	12	$42/64 = 0.656250$	20	$63/64 = 0.984375$
5	$10/64 = 0.156250$	13	$46/64 = 0.718750$	21	$64/64 = 1.000000$
6	$14/64 = 0.218750$	14	$50/64 = 0.781250$		
7	$18/64 = 0.281250$	15	$54/64 = 0.843750$		

Significance level $\alpha = .05$

- $P \leqslant .05$ for "$w \leqslant 0$ or $w \geqslant 21$"
- The *critical region* (where H_0 is rejected) is $w = 0$ or 21.
- The *acceptance region* (where H_0 is accepted) is $1 \leqslant w \leqslant 20$.
- The Type I error rate is really $2/64 = 0.031250$.
  Discrete distributions will often have Type I error rate $< \alpha$.

Critical region for a given significance level α

Cumulative distribution of W

w	$P(W \leqslant w)$	w	$P(W \leqslant w)$	w	$P(W \leqslant w)$
0	$1/64 = 0.015625$	8	$22/64 = 0.343750$	16	$57/64 = 0.890625$
1	$2/64 = 0.031250$	9	$27/64 = 0.421875$	17	$59/64 = 0.921875$
2	$3/64 = 0.046875$	10	$32/64 = 0.500000$	18	$61/64 = 0.953125$
3	$5/64 = 0.078125$	11	$37/64 = 0.578125$	19	$62/64 = 0.968750$
4	$7/64 = 0.109375$	12	$42/64 = 0.656250$	20	$63/64 = 0.984375$
5	$10/64 = 0.156250$	13	$46/64 = 0.718750$	21	$64/64 = 1.000000$
6	$14/64 = 0.218750$	14	$50/64 = 0.781250$		
7	$18/64 = 0.281250$	15	$54/64 = 0.843750$		

Other significance levels

- $\alpha = .01$: $P \geqslant 2(.015625) = .031250$ for all w.
  So we never have $P \leqslant .01$. Thus, H_0 is always accepted.
- $\alpha = .10$: Accept H_0 for $3 \leqslant w \leqslant 18$.

Mann-Whitney Test, a.k.a. "Wilcoxon two-sample test"

- Let X, Y be random variables whose distributions are the same except for a possible shift, $Y \sim X + C$ for some constant C.
- We will test the hypotheses
  H_0: X and Y have the same median (i.e., $C = 0$).
  H_1: X and Y do not have the same median (i.e., $C \neq 0$).
- This is a non-parametric test.
  In practice, it's used if the plots look similar but possibly shifted.
  However, if there are other differences in the distributions than just the shift, the P-values will be off.
- Two sets of authors (Mann-Whitney vs. Wilcoxon) developed essentially equivalent tests for this; we'll do the one due to Wilcoxon.

Computing the statistic U

Wilcoxon's definition

- **Data:**
  Sample $x_1, \ldots, x_m$ for X: $11, 13 \ (m = 2)$
  Sample $x_{m+1}, \ldots, x_{m+n}$ for Y: $12, 15, 14 \ (n = 3)$
- **Replace data by ranks from smallest (1) to largest ($m + n$):**
  Ranks for X: $1, 3$
  Ranks for Y: $2, 5, 4$
- ***U* is the sum of the *X* ranks:** $U_0 = 1 + 3 = 4$

- Ties may happen in discrete case. If there's a tie for 2nd and 3rd smallest, use 2.5 for both of them.
- This is a *two sample test*.
  The Wilcoxon Signed Rank test previously covered is a *one sample test*.

Computing the statistic U

Mann-Whitney's definition

- We'll call Mann-Whitney's statistic $\widetilde{U}$, although they called it U.
- $\widetilde{U}$ is the number of pairs (x, y) with x in the X sample, y in the Y sample, and $x < y$.
- **Data:**

  Sample $x_1, \ldots, x_m$ for X: $11, 13\ (m = 2)$

  Sample $x_{m+1}, \ldots, x_{m+n}$ for Y: $12, 15, 14\ (n = 3)$
- $11 < 12,\ 11 < 15,\ 11 < 14,\ 13 < 15,\ 13 < 14$ so $\widetilde{U} = 5$.
- The statistics are related by $\widetilde{U} = mn + m(m+1)/2 - U$.
- We'll stick with Wilcoxon's definition and ignore this one.

Computing the distribution of U: permutation test

- Under H_0, X and Y have the same distribution. So we are just as likely to have seen any $m = 2$ of those numbers for the X sample and the other $n = 3$ for Y. *Resample* them as follows:
- Permute the $m + n = 2 + 3 = 5$ numbers in all $(m + n)! = 120$ ways.
- Treat the first m of them as a new sample of X and the last n as a new sample of Y, compute U for each.

X	Y	U
11, 13	12, 15, 14	4
11, 13	12, 14, 15	4
11, 13	14, 12, 15	4
11, 13	14, 15, 12	4
11, 13	15, 12, 14	4
11, 13	15, 14, 12	4
13, 11	12, 15, 14	4
13, 11	12, 14, 15	4
13, 11	14, 12, 15	4
13, 11	14, 15, 12	4
13, 11	15, 12, 14	4
13, 11	15, 14, 12	4
11, 12	13, 15, 14	3
11, 12	13, 14, 15	3
...	...	...

- $m!\,n! = 2!\,3! = 2 \cdot 6 = 12$ of the permutations give the same partition of numbers for X and Y.
- So it would suffice to list partitions instead of permutations.
- There are $\frac{(m+n)!}{m!\,n!} = \binom{m+n}{n}$ partitions; $\binom{5}{2} = 10$ partitions in this case.

Computing the distribution of U: permutation test

- *Resample* the data by partitioning the numbers between X & Y in all $\binom{m+n}{m} = \binom{2+3}{2} = \binom{5}{2} = 10$ possible ways. Compute U for each. As a short cut, we can just work with the ranks:

X ranks	Y ranks	U
$1,2$	$3,4,5$	3
$1,3$	$2,4,5$	4
$1,4$	$2,3,5$	5
$1,5$	$2,3,4$	6
$2,3$	$1,4,5$	5
$2,4$	$1,3,5$	6
$2,5$	$1,3,4$	7
$3,4$	$1,2,5$	7
$3,5$	$1,2,4$	8
$4,5$	$1,2,3$	9

- Compute the PDF and CDF of U from this (all 10 cases are equally likely):

U	$P_U(u)$	$F_U(u)$
< 3	$0/10$	$0/10$
3	$1/10$	$1/10$
4	$1/10$	$2/10$
5	$2/10$	$4/10$
6	$2/10$	$6/10$
7	$2/10$	$8/10$
8	$1/10$	$9/10$
9	$1/10$	$10/10$

- *P-value of $U_0 = 4$:* The mirror image of 4 is 8.
  $P = P(U \leqslant 4) + P(U \geqslant 8) = 2P(U \leqslant 4) = 2(.2) = .4.$

Computing P-value and U in Matlab or R

Matlab

```
>> ranksum([11,13],[12,15,14])
    0.4000

>> [p,h,stats] = ...
    ranksum([11,13],[12,15,14])

p =    0.4000
h =    0
stats =
    ranksum: 4

>> stats.ranksum
     4
```

Note: "..." lets you break a command onto two lines, both at the command line and in scripts. If you type it on one line, don't use "..."

R

```
> test = wilcox.test(c(11,13),
+                    c(12,15,14))
> test$p.value
[1] 0.4
> test$statistic
W
1
```

Notes:

- R computes a different statistic "W" instead of U.
- $W = U - m(m+1)/2$
  In this case, $W = 4 - 2(2+1)/2 = 1$.
- The $+$ prompt is given when you break a command onto two lines at the command line. Don't type it in.

Properties of U

- **Minimum:** $1+2+\cdots+m = m(m+1)/2$
  Maximum: $(n+1)+(n+2)+\cdots+(n+m) = m(2n+m+1)/2$

- *Assuming H_0:*
  Expected value: $E(U) = m(m+n+1)/2$
  Variance: $\mathrm{Var}(U) = mn(m+n+1)/12$

- **Symmetry of PDF:** In the sample data, switch the ith least and ith largest elements for all i.

  The ranks added together are replaced by the complementary ranks, so U goes to its mirror image around $m(m+n+1)/2$.

Expected value of U

- Each rank has probability $\frac{m}{m+n}$ to be in the X group and hence in the rank sum.
- Let $U_j = \begin{cases} 0 & \text{prob. } n/(m+n); \\ j & \text{prob. } m/(m+n) \end{cases}$ and $U = U_1 + \cdots + U_{m+n}$.
- The U_j's are dependent!
- $E(U_j) = 0 \cdot \frac{n}{m+n} + j \cdot \frac{m}{m+n} = j \cdot \frac{m}{m+n}$
- Expectation is still additive, even though the U_j's are dependent:

$$\begin{aligned} E(U) &= E(U_1) + \cdots + E(U_{m+n}) \\ &= (1 + 2 + \cdots + (m+n))\frac{m}{m+n} \\ &= \frac{(m+n)(m+n+1)}{2} \cdot \frac{m}{m+n} = \frac{m(m+n+1)}{2} \end{aligned}$$

- Variance is harder: it is *not additive* since the U_j's are dependent.

Covariance

- Let X and Y be random variables, possibly dependent.
- Let $\mu_X = E(X)$, $\mu_Y = E(Y)$
- $$\begin{aligned}\operatorname{Var}(X+Y) &= E((X+Y-\mu_X-\mu_Y)^2) = E\left(\left((X-\mu_X)+(Y-\mu_Y)\right)^2\right)\\ &= E\left((X-\mu_X)^2\right) + E\left((Y-\mu_Y)^2\right) + 2E\left((X-\mu_X)(Y-\mu_Y)\right)\\ &= \operatorname{Var}(X) + \operatorname{Var}(Y) + 2\operatorname{Cov}(X,Y)\end{aligned}$$
  where the *covariance* of X and Y is defined as
  $$\operatorname{Cov}(X,Y) = E\left((X-\mu_X)(Y-\mu_Y)\right)$$
- Expanding gives an alternate formula
  $\operatorname{Cov}(X,Y) = E(XY) - E(X)E(Y)$:
  $$\begin{aligned}\operatorname{Cov}(X,Y) &= E\left((X-\mu_X)(Y-\mu_Y)\right)\\ &= E(XY) - \mu_X E(Y) - \mu_Y E(X) + \mu_X\mu_Y = E(XY) - E(X)E(Y)\end{aligned}$$
- $$\operatorname{Var}(X_1+X_2+\cdots+X_n) = \operatorname{Var}(X_1)+\cdots+\operatorname{Var}(X_n) + 2\sum_{1\leqslant i<j\leqslant n}\operatorname{Cov}(X_i,X_j)$$

Variance of U

Variance of U_j

- Let $U_j = \begin{cases} 0 & \text{prob. } n/(m+n); \\ j & \text{prob. } m/(m+n) \end{cases}$ and $U = U_1 + \cdots + U_{m+n}$.
- $E(U_j) = j \cdot \frac{m}{m+n}$ and $E({U_j}^2) = j^2 \cdot \frac{m}{m+n}$
- $\mathrm{Var}(U_j) = E({U_j}^2) - (E(U_j))^2 = j^2 \frac{m}{m+n} - j^2 \frac{m^2}{(m+n)^2} = j^2 \frac{mn}{(m+n)^2}$

Covariance between U_i and U_j for $i \neq j$

- $U_i\,U_j$ is 0 if the rank i and/or j element is in the Y sample.
  It's $i \cdot j$ if both are in the X sample, which has prob. $\frac{m(m-1)}{(m+n)(m+n-1)}$.
- $E(U_i\,U_j) = ij \cdot \frac{m(m-1)}{(m+n)(m+n-1)}$
- $$\begin{aligned} \mathrm{Cov}(U_i, U_j) &= E(U_i\,U_j) - E(U_i)E(U_j) \\ &= ij \cdot \left(\frac{m(m-1)}{(m+n)(m+n-1)} - \frac{m^2}{(m+n)^2} \right) = -ij \frac{mn}{(m+n)^2(m+n-1)} \end{aligned}$$

Variance of U

Variance computation

- $\operatorname{Var}(U_j) = j^2 \frac{mn}{(m+n)^2}$ and $\operatorname{Cov}(U_i, U_j) = -ij \frac{mn}{(m+n)^2(m+n-1)}$ (if $i \neq j$)
- $\operatorname{Var}(U) =$ sum of variances + twice the sum of covariances:

$$\sum_{j=1}^{m+n} j^2 \frac{mn}{(m+n)^2} - 2 \sum_{1 \leqslant i < j \leqslant m+n} ij \cdot \frac{mn}{(m+n)^2(m+n-1)} = \cdots = \boxed{\frac{mn(m+n+1)}{12}}$$

Details

Plug in these identities (at $k = m+n$) and simplify:

- $1 + 2 + \cdots + k = k(k+1)/2$
- $1^2 + 2^2 + \cdots + k^2 = k(k+1)(2k+1)/6$
- $2 \sum_{1 \leqslant i < j \leqslant k} i \cdot j = (1+2+\cdots+k)^2 - (1^2+2^2+\cdots+k^2) = k(k-1)(k+1)(3k+2)/12$

Variations

Unpaired data

- Let $f([x_1, \ldots, x_m], [x_{m+1}, \ldots, x_{m+n}])$ be any test statistic on two vectors of samples (a *two sample test statistic*).
- Follow the same procedure as for computing U and its P-value, but compute f instead of U on each permutation of the x's.
- Ewens & Grant explains this for the t-statistic, pages 141 & 464.

Paired data

- **Unpaired:** If m subjects are measured who do not have a condition and n subjects are measured who do have it, and these are independent, then the Mann-Whitney test could be used.
- **Paired:** Suppose there are n subjects, with

  $x_i =$ measurement before treatment

  $y_i =$ measurement after treatment, $i = 1, \ldots, n$.
- Mann-Whitney on $[x_1, \ldots, x_n], [y_1, \ldots, y_n]$ ignores the pairing.
- Use Wilcoxon Signed Rank test on $x_1 - y_1, \ldots, x_n - y_n$: median=0?