- Density Function of the i-th order statstics:

$$f_{Y_{(i)}}(y_{(i)}) = \frac{n!}{(i-1)!(n-i)!} F_Y(y_{(i)})^{n-i!} f_Y(y_{(i)})[1-F_Y(y_{(i)})]^{n-i}$$

- Density Function of the i-th and the j-th order statistics:

$$f_{Y_{(i)},Y_{(j)}}(y_{(i)},y_{(j)}) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!} F_Y(y_{(i)})^{n-i!} f_Y(y_{(i)}) [F_Y(y_{(j)}) - F_Y(y_{(i)})]^{j-i-1} \cdot f_Y(y_{(j)}) [1-F_Y(y_{(j)})]^{n-j}$$

- The Camér-Rao Bound — works only if support of Y_i depends on θ:

$$Var(\hat{\tau}(Y_1,\dots,Y_n)) \geq \frac{\left(\frac{d\tau(\theta)}{d\theta}\right)^2}{-E\left[\frac{\partial^2}{\partial\theta^2}\log f_{Y_1,\dots,Y_n}(y_1,\dots,y_n)\right]}$$

$\leftarrow =1$ if $\tau(\theta)=\theta$.

$\leftarrow = -nE(\log f_Y(y))$ if iid.

"=" is achievable iff $f_{Y_1,\dots,Y_n}(y_1,\dots,y_n)$ is of the "exponential family" form:

$$f_{Y_1,\dots,Y_n}(y_1,\dots,y_n) = h(y_1,\dots,y_n)\, e^{C(\theta)+D(\theta)\cdot\hat{\tau}(Y_1,\dots,Y_n)}$$

. It's achieved by

$$\tau(\theta) = -\frac{\frac{d}{d\theta}C(\theta)}{\frac{d}{d\theta}D(\theta)} =: -\frac{A(\theta)}{B(\theta)}. \quad Var(\hat{\tau}) = -\frac{1}{B(\theta)}\frac{d}{d\theta}\left(\frac{A(\theta)}{B(\theta)}\right).$$

- $W(Y_1,\dots,Y_n)$ is a sufficient statistic for θ if the conditional distribution $f_{Y_1,\dots,Y_n}(y_1,\dots,y_n|w;\theta)$ is independent on θ. Equivalently:

$$f_{Y_1,\dots,Y_n}(y_1,\dots,y_n|w;\theta) = f_W(w;\theta)\cdot h(y_1,\dots,y_n)$$

If the support of Y is indep. of θ, $= \underbrace{g(w,\theta)}_{\text{any function of } w \text{ and } \theta} \cdot \underbrace{h(y_1,\dots,y_n)}$ $\to$ any function of Y_i's independent of θ.

- Transformation Theory
  - One-RV Case ($y := g(x)$):
    - When $y=g(x)$ is monotonic: $f_Y(y) = \left[f_X(x)\left|\frac{dx}{dy}\right|\right]_{x=g^{-1}(y)}$

      or $f_Y(y) = \left[f_X(x)\frac{1}{\left|\frac{dy}{dx}\right|}\right]_{x=g^{-1}(y)}$

    - When multiple monotonic regions exist: $f_Y(y) = \sum_{i\in reg.}\left[f_{X_i}(x_i)\left|\frac{dX_i}{dy}\right|\right]_{x_i=X_i(y)}$

      or $f_Y(y) = \sum_{i\in reg.}\left[f_{X_i}(x_i)\frac{1}{\left|\frac{dy}{dx_i}\right|}\right]_{x_i=X_i(y)}$

Multiple Transformation

$$f_{Y_1,Y_2}(y_1,y_2) = \Big[f_{X_1,X_2}(x_1,x_2) \cdot \overset{\text{absolute!}}{|J|} \Big]_{\substack{x_1 = x_1(y_1,y_2) \\ x_2 = x_2(y_1,y_2)}} \quad w/ \quad J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix}$$

$$= \Big[f_{X_1,X_2}(x_1,x_2) \cdot \frac{1}{|J^*|} \Big]_{\substack{x_1 = x_1(y_1,y_2) \\ x_2 = x_2(y_1,y_2)}} \quad w/ \quad J^* = \begin{vmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} \end{vmatrix}$$

— The Newman-Pearson Lemma:

If the "likelihood ratio" (of H_0 over H_1) is sufficiently small, reject H_0.

The ratio is $\lambda = \dfrac{f_{Y_1,\dots,Y_n}(y_1,\dots,y_n;\theta_0)}{f_{Y_1,\dots,Y_n}(y_1,\dots,y_n;\theta_1)}$ or $= \dfrac{\max_{\text{comp. param.}} \text{Prob}(\text{data}|H_0)}{\max_{\text{comp. param.}} \text{Prob}(\text{data}|H_1)}$ if any hypo. is composite.

Compare with the Smith-Jones Method to find the sufficient statistics:

$$\text{ratio} = \frac{f_{Y_1,\dots,Y_n}(y_{11},\dots,y_{1n};\theta)}{f_{Y_1,\dots,Y_n}(y_{21},\dots,y_{2n};\theta)}$$ — goal is to find a $h(y_1,\dots,y_n)$ st $h(y_{11},\dots,y_{1n}) = h(y_{21},\dots,y_{2n})$ iff ratio is indep. of θ.

— Distributions:

— Normal Distribution:

$$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y-\mu)^2} \sim N(\mu,\sigma^2) \qquad \mu \qquad \sigma^2$$

— Gamma Dist.:

$$f_Y(y) = \frac{1}{\Gamma(\alpha)\beta^\alpha} e^{-\frac{y}{\beta}} y^{\alpha-1} \sim \Gamma(\alpha,\beta) \qquad \alpha\beta \qquad \alpha\beta^2$$

— χ^2 Dist.:

$$f_Y(y) = \frac{1}{\Gamma(\frac{\nu}{2})2^{\frac{\nu}{2}}} e^{-\frac{y}{2}} y^{\frac{\nu}{2}-1} \sim \chi^2(\nu) \qquad \nu \qquad 2\nu$$

— Binomial Dist.:

$$\text{Prob}(Y=y) = \binom{N}{y}\theta^y(1-\theta)^{N-y} \sim \text{Binom}(N,\theta) \quad N\theta \quad N\theta(1-\theta)$$

— Poisson Dist.:

$$\text{Prob}(Y=y) = e^{-\lambda}\frac{\lambda^y}{y!} \sim \text{Poisson}(\lambda) \qquad \lambda \quad \lambda$$

— t dist.:

$$T = \frac{\sqrt{n}(\overset{NID(\mu,\sigma^2)}{\bar{Y}}-\mu)}{S} \quad w/ \quad S^2 = \frac{\Sigma(Y_i-\bar{Y})^2}{n-1} \sim T(n-1)$$

— f dist.:

$$F = \frac{S_1^2}{S_2^2} \qquad S_1^2 = \frac{\Sigma(Y_{1i}-\bar{Y}_1)^2}{n-1} \qquad S_2^2 = \frac{\Sigma(Y_{2i}-\bar{Y}_2)^2}{m-1} \sim F(n-1,m-1)$$

− χ^2 dist.:

$$\left\{\begin{array}{ll} \frac{1}{\sigma^2}\Sigma(Y_i-\bar{Y})^2 \sim \chi^2(n-1) & \\ \frac{1}{\sigma^2}\Sigma(Y_i-\mu)^2 \sim \chi^2(n) & \left.\right\} \text{assuming } Y_i\text{'s NID}(\mu,\sigma^2) \\ \Sigma Y_i^2 \sim \chi^2(n) & \text{assuming } Y_i\text{'s NID}(0,1) \end{array}\right.$$

− Gamma Functions

$\Gamma(n) = (n-1)\cdot\Gamma(n-1)$

$\Gamma(1) = \int_0^\infty e^{-x}dx = 1$

$\Gamma(2) = \int_0^\infty xe^{-x}dx = 1$

$\Gamma(\frac{1}{2}) = \int_0^\infty \frac{1}{\sqrt{x}}e^{-x}dx = \sqrt{\pi}$

$\Gamma(n) = \int_0^\infty x^{n-1}e^{-x}dx$

$\int_0^\infty x^{n-1}e^{-cx}dx = c^{-n}\Gamma(n)$

yī shēng xuán mìng miǎn qiáng
一生悬命勉强
With all my effort I study.

− Meta Theory

− Theory basis: p-value, as a RV, is uniformly distributed on [0, 1] and is the CDF of our test statistic when Ho is true.

− Usage: Suppose we have n p-values: $p_1, \dots, p_n$, then calculate $-2\sum_{i=1}^{n}\log p_i$. This follows a $\chi^2(\nu=2n)$ distribution. Look up from χ^2 table with this $\nu=2n$ and a specified α. If this value is large enough, then we reject Ho.

− Mann-Whitney ("Witcoxon 2-Sample") Test

1. Sort the union set of $\{y_i\}$ and $\{x_i\}$: $x_7, x_3, y_9, x_{10}, \dots, y_7, x_1$
2. Assign ordinal ranks to each element : ①, ②, ③, ④, …, $(n+m-1)$, $(n+m)$
3. $U :=$ sum of ranks of all x_i's : $1+2+\quad 4+\dots+\quad(n+m)$
4. This "test statistic" U should obey : $\text{Prob}(U=u) = \dfrac{\text{\# of ways } U \text{ attaining } u}{\binom{m+n}{n}}$.
5. When m and n are large, $U \sim N(\mu_u, \sigma_u^2)$, where μ_u and σ_u^2 are ~~always~~ exactly $\mu_u = \frac{1}{2}n(n+m+1)$ and $\sigma_u^2 = \frac{1}{12}mn(n+m+1)$, under Ho.
6. Suppose probability of Type-I error is chosen to be α, and the corresponding

threshold value is h, then our TS U has to satisfy

$$Z := \frac{U-\mu_u}{\sqrt{\sigma_u^2}} \geq \frac{h'-\mu_u}{\sqrt{\sigma_u^2}} \quad \text{where } h' = \underbrace{h \pm \tfrac{1}{2}}_{\text{"continuity correction"}}$$

- Asymptotic Relative Efficiency (ARE)

The ARE of the Wilcoxon Test relative to the 2-sample t-test is:

$$ARE(\text{wilcoxon} \mid \text{2-sample } t) = 12\sigma^2 \left[\int_{-\infty}^{+\infty} \{f_Y(y)\}^2 dy \right]^2$$

→ max val.: 100% when Y_i's are unifo[rm] (i.e. completel[y] NOT NID.)

→ ~~min~~ one val.: 95.5% when Y_i's are NI[D]

where $f_Y(y)$ is the true density function of the iid RVs Y_i's.

※ ARE is the relative efficiency of Test Method #2 rejecting H_0 relative to Test Method #1 when the true mean of ~~observed data~~ distribution μ is infinitely close to the claimed ("specified") value μ_0 in H_0 ("$\lim_{\mu \to \mu_0}$").

The idea is that, to demonstrate identical amount of confidence in rejecting H_0, the more efficient Test Method should require less amount of observations:

$$ARE = \lim_{\mu \to \mu_0} \frac{n_2(\mu)}{n_1(\mu)}$$

amount of observations required by Test Method #2 to demonstrate a power specified.

- "$-2\log\lambda$ χ^2" Approximation

- Requirement
  - ① Camér-Rao Requirement
  - ② H_0 is "nested" within H_1

- Usage: $-2\log\lambda$ approximately $\sim \chi^2(\nu)$ with $\nu =$ # of unspecified parameters in H_1 − # of that in H_0.

Then you can use $-2\log\lambda$ as the test statistic and look up values from a χ^2 chart:

$$\text{Prob}(-2\log\lambda \geq \boxed{}) = \alpha$$

looked up from a χ^2 table

manually specified