Sum of $\perp\!\!\!\perp$ r.v.s.

- $Pois(\lambda_1) + Pois(\lambda_2) = Pois(\lambda_1+\lambda_2)$
- $\chi^2_{v_1} + \chi^2_{v_2} = \chi^2_{(v_1+v_2)}$ $\qquad \chi^2_k = Gamma(\frac{k}{2}, \frac{1}{2})$
- $Cauchy(\mu_1, \sigma_1) + Cauchy(\mu_2, \sigma_2) = Cauchy(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$
- $N(\mu_1, \sigma_1^2) + N(\mu_2, \sigma_2^2) = N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$
- Δ $\sum_{i=1}^{n} Gamma(\alpha_i, \beta) = Gamma(\sum_{i=1}^{n}\alpha_i, \beta)$
- $\sum_{i=1}^{n} N^2(0,1) = \chi^2_n$

- $Bern(p) = Bino(1, p)$

  $\sum_{i=1}^{n} Bern(p) = Bino(n, p)$

- Δ $exp(\beta) = Gamma(1, \beta)$

  $\sum_{i=1}^{n} exp(\beta) = Gamma(n, \beta)$

Beta p.d.f. $\quad \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot x^{\alpha-1}(1-x)^{\beta-1} \quad x \in [0,1]$

$\mathbb{E} = \frac{\alpha}{\alpha+\beta}. \quad Var = \frac{\alpha}{\alpha+\beta} \cdot \frac{\beta}{\alpha+\beta} \cdot \frac{1}{(\alpha+\beta+1)}$

A k-**parameter exponential family** has pdf

$$f_\theta(\mathbf{x}) = \exp\{\eta(\theta) \cdot T(\mathbf{x}) - B(\theta)\}\, h(\mathbf{x}),$$

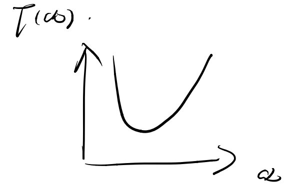
A canonical family is of **full rank** if

- it is minimal (i.e., neither T's nor η's satisfy a linear constraint), and
- Ξ contains a k-dimensional rectangle.

Exponential family includes

- Gaussian $\quad \chi^2$
- Gamma $\quad$ exp
- Binomial
- Poisson
- etc.

Gamma Function.



factorial function $\quad \Gamma(n) = (n-1)!$ $\quad$ **Def**

Gamma function $\quad \Gamma(\alpha) = \int_0^\infty x^{\alpha-1} \cdot e^{-x} \, dx$

NB!

for $\alpha = 1$, we can write $\Gamma(1) = \int_0^\infty e^{-x} dx = 1$

what if we let $x = \lambda y$?

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

Another form of gamma function.

$$\overset{x=\lambda y}{=} \int_0^\infty (\lambda y)^{\alpha-1} \cdot e^{-(\lambda y)} \cdot \lambda \cdot dy$$

$$= \lambda^{\alpha} \cdot \int_0^\infty y^{\alpha-1} e^{-(\lambda y)} dy. \qquad (\text{for } \alpha, \lambda > 0)$$

Moreover

$$\Gamma(\alpha+1) = \int_0^\infty x^{\alpha} \cdot e^{-x} \cdot dx = -\int_0^\infty x^{\alpha} \cdot d(e^{-x}) \overset{\text{分部}}{\underset{\text{积分}}{=}} -\Big[\underbrace{e^{-x} \cdot x^{\alpha}\Big|_0^\infty}_{0-0=0} - \int_0^\infty e^{-x} \cdot \alpha \cdot x^{\alpha-1} \cdot dx\Big]$$

$$= \alpha \cdot \int_0^\infty e^{-x} \cdot x^{\alpha-1} \cdot dx$$

$$= \alpha \, \Gamma(\alpha).$$

prop 1.

Try.

$$\int_0^\infty x^{\alpha-1} \cdot e^{-\lambda x} \cdot dx \underset{\lambda>0}{\overset{y=\lambda x}{=\!=}} \int_0^\infty \left(\frac{y}{\lambda}\right)^{\alpha-1} \cdot e^{-y} \cdot \frac{1}{\lambda} \cdot dy = \frac{1}{\lambda^{\alpha}} \cdot \int_0^\infty y^{\alpha-1} \cdot e^{-y} \cdot dy$$

$$= \lambda^{-\alpha} \cdot \Gamma(\alpha).$$

prop2.

Try

$$\Gamma\left(\tfrac{1}{2}\right) = \int_0^\infty \cdot x^{-\frac{1}{2}} \cdot e^{-x} \cdot dx \underset{dx = 2t\,dt}{\overset{t^2 = x}{=\!=}} \int_0^\infty \frac{1}{t} \cdot e^{-t^2} \cdot 2t \, dt = 2\int_0^\infty e^{-t^2} \cdot dt.$$

$$I = \int_0^\infty \cdot e^{-x^2} \cdot dx \qquad I^2 = \int_0^\infty \int_0^\infty e^{-(x^2+y^2)} dx\, dy = \int_0^{\frac{\pi}{2}} \int_0^\infty e^{-r^2} \cdot r \cdot dr\, d\theta$$

$$= -\frac{1}{2} \int_0^{\frac{\pi}{2}} \left(\int_{-\infty}^{0} e^{-r^2} \cdot d(-r^2) \right) d\theta$$

$$= \frac{1}{2} \frac{\pi}{2} \cdot 1 = \frac{\pi}{4}.$$

$$I = \frac{\sqrt{\pi}}{2}$$

$$\Gamma\left(\tfrac{1}{2}\right) = 2I = \sqrt{\pi}$$

prop3.

properties of the gamma. function.

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

2. $\int_0^\infty x^{\alpha-1} \cdot e^{-\lambda x} dx = \dfrac{\Gamma(\alpha)}{\lambda^\alpha}$

$X \sim$ Gamma (α, λ) if it's pdf is given by

$$f_X(x) = \begin{cases} \dfrac{\lambda^\alpha x^{\alpha-1} \cdot e^{-\lambda x}}{\Gamma(\alpha)} & x > 0 \\ 0 & \text{otherwise.} \end{cases}$$

let $\alpha = 1$ we obtain.

$$f_X(x) = \begin{cases} \lambda \cdot e^{-\lambda x} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

Gamma $(1, \lambda)$ = Exponential (λ)

Sum n independent Exponential (λ). r.v. then you will get a Gamma (n, λ)

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

$X \sim$ Gamma (α, λ) then $EX = \frac{\alpha}{\lambda}$. $Var = \frac{\alpha}{\lambda^2}$

Poission.

$X \sim P(\lambda)$

$f(k; \lambda) = Pr(X = k) = \dfrac{\lambda^k e^{-\lambda}}{k!}$

$\lambda = E(X) = Var(X)$

Exponential

$X \sim \mathcal{E}(\lambda)$

$$f(x;\lambda) = \begin{cases} \lambda \cdot e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

$\mathbb{E}X = \frac{1}{\lambda}$

$Var(X) = \frac{1}{\lambda^2}$

$\mathbb{E}(X^n) = \frac{n!}{\lambda^n}$

Student t

How student's distribution arises from. sampling.

$X_1, \cdots X_n \underbrace{\text{i.i.d.}} N(\mu, \sigma^2)$

$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ sample mean.

$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$ Bessel-corrected sample variance

$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

$\frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0, 1)$

What if we do not have the value of σ? Using S^2 as a substitution

$S \sim t_{n-1}$

$$f(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2})} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

$\mathbb{E}(t(k)) = \mu.$

$Var(t(k)) = \frac{k}{k-2} \cdot S^2$

Order statistics.

$X_1\ X_2 \cdots X_n \overset{i.i.d.}{\sim} F, f$ $\quad X_{(k)}$ is the k-th min

① $$F_{max}(x) = F_{X_{(n)}}(x) = \mathbb{P}(X_1 \le x, \cdots X_n \le x) = (F_X(x))^n$$

$$f_{max}(x) = \frac{\partial F_{max}(x)}{\partial x} = n \cdot (F_X(x))^{n-1} f_X(x)$$

② $$F_{min}(x) = F_{X_{(1)}}(x) = \mathbb{P}(X_{min} \le x) = 1 - \mathbb{P}(X_{min} > x)$$
$$= 1 - \mathbb{P}(X_1 > x \cdots X_n > x)$$
$$= 1 - (1 - F(x))^n$$
$$f_{min}(x) = \frac{d F_{min}(x)}{dx} = n(1 - F(x))^{n-1} \cdot f(x)$$

$f_X(x)\,dx = \mathbb{P}(X \in dx)$

③ $$f_{(k)}(x) \cdot dx = \mathbb{P}(\text{the k-th} \in dx \quad k-1 < x \quad n-k \ge x)$$
$$= n \cdot \mathbb{P}(\text{the } X_1 \text{ is k-th that} \in dx, \quad k-1 < x,\ n-k > x)$$
$$= n \cdot \mathbb{P}(X_1 \in dx) \cdot \binom{n-1}{k-1} \cdot F(x)^{k-1} (1 - F(x))^{n-k}$$
$$= n \cdot f(x) \cdot dx \binom{n-1}{k-1} F(x)^{k-1} \cdot (1 - F(x))^{n-k}.$$

eg.

$U_1 \cdots U_n \sim \text{Unif}(0,1)$ then. $\quad f(x) = \mathbb{1}(0 \le x \le 1)$

$\quad F(x) = x \cdot \mathbb{1}(0 \le x \le 1)$

$$\therefore\ u^{(k)} = n \cdot 1 \cdot \binom{n-1}{k-1} \cdot x^{k-1} \cdot (1-x)^{n-k} \cdot \mathbb{1}(0 \le x \le 1) \quad (*)$$

令. $\alpha - 1 = k-1 \qquad \beta - 1 = n-k$

$\quad \alpha = k \qquad \beta = n-k+1$

$$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} = \frac{\Gamma(n+1)}{\Gamma(k) \cdot \Gamma(n-k+1)} = \frac{n!}{(k-1)!(n-k)!}$$

The coefficient of $(*)$ is $n \cdot \frac{(n-1)!}{(k-1)!(n-k)!}$ $=$

$\therefore$ Lemma. $U_1\ U_2 \cdots \text{Unif}(0,1)$ then,

$$U_{(k)} \sim Beta(k, n-k+1) \quad 1 \le k \le n$$

Δ For $Beta(\alpha, \beta)$

$$\mathbb{E} = \frac{\alpha}{\alpha+\beta}. \quad Var = \frac{\alpha}{\alpha+\beta} \times \frac{\beta}{\alpha+\beta} \times \frac{1}{(\alpha+\beta+1)}$$

$\therefore \mathbb{E} U_{(k)} = \frac{k}{n+1}$

EG1

$\therefore X_1 \cdots X_n \overset{i.i.d.}{\sim} U(0,\theta) \quad Y_i \triangleq \frac{X_i}{\theta} \overset{i.i.d.}{\sim} U(0,1)$

$\therefore \mathbb{E} Y^{(k)} = \frac{k}{n+1} = \frac{\mathbb{E} X^{(k)}}{\theta} \Rightarrow \mathbb{E} X^{(k)} = \frac{k\theta}{n+1}$

EG2

$X_1 \cdots X_n \overset{i.i.d.}{\sim} (\theta, 1+\theta) \quad Y_i \triangleq X_i - \theta \overset{i.i.d.}{\sim} U(0,1)$

$\therefore \mathbb{E} Y^{(k)} = \frac{k}{n+1} = \mathbb{E} X^{(k)} - \theta \Rightarrow \mathbb{E} X^{(k)} = \frac{k}{n+1} + \theta$

How to prove $T(x)$ is sufficient for θ $\quad X \sim f_\theta(x)$

① According to def. Pf: $\mathbb{P}(X=x \mid T=t)$ is free of θ.

②. Factorization Thm. $T(x)$ is sufficient for θ iff

$$f_\theta(x) = g(T(x), \theta)\, h(x)$$

sufficient ✓

How to prove $T(x)$ is minimal sufficient for θ $\quad X \sim f_\theta(x)$

① According to def. Pf: $\forall$ sufficient $S(x)$

$T(x)$ is a func of $S(x)$

② Find the sufficient $T(x)$

Verify $\boxed{r(\theta) = \frac{f_\theta(x)}{f_\theta(y)}}$ is free of θ $\forall x, y$

iff $T(x) = T(y)$

aka only if $T(x) = T(y)$ we have $\square$ ✓

Δ For full rank exponential families.

minimal sufficient statistics. is the $T_1(x) \cdots T_k(x)$ also complete.

in. $f_\theta(x) = h(x) \cdot \exp\{\sum_{i=1}^{k} \eta_i(\theta) \cdot T_i(x) - B(\theta)\}$

Δ Basically, the sufficient $T(x)$ for $Unif(\cdots)$ is also minimal sufficient.

Δ A complete and sufficient statistics is minimal sufficient

Δ A bounded complete and sufficient $T(x) \perp\!\!\!\perp$ Ancillary A

Basu's Theorem.

Ancillary statistics. $S(X)$ is ancillary if it's distribution does not depend on θ

location family $X_1 \cdots \sim F(x-\theta)$

$z_i := X_i - \theta \sim F(x)$ is free of θ

$\therefore R = X^{(n)} - X^{(1)} = z^{(n)} - z^{(1)}$ is ancillary

scale family $X_1 \cdots \sim F(\frac{x}{\theta})$

$z_i := \frac{X_i}{\theta} \sim F(x)$ is free of θ.

$= \frac{z_n}{z_1} = \frac{x_n/\theta}{x_1/\theta}$.

1. T is *complete* for θ if

$$E_\theta g(T) = 0 \qquad \Longrightarrow \qquad g(T) = 0 \;\; a.s. \qquad \forall\theta.$$

2. T is *boundedly complete* if the previous statement holds for all bounded g.

Clearly, completeness implies bounded completeness, but not vice versa.

Characterization Thm.

① $\mathbb{E}(g(X)) = g(\theta)$

② $\forall\, u(X) \in \mathcal{U}\{0\}$. $\quad Cov(g(X), u(X)) = \mathbb{E}[g \cdot u] = 0$

① T is sufficient.

②. $\mathbb{E}(g(T)) = g(\theta)$

$\forall\, u(T) \in \mathcal{U}\{0\} = \{h(T) : \mathbb{E}h(T) = 0\}$ $\quad \mathbb{E}[g(T), u(T)] = 0$

Lehmann-Scheffe.

① T is C-S.

② $\mathbb{E}(g(T)) = g(\theta)$

$\Rightarrow g(T)$ is UMVUE

1. $\mathbb{E}g(T) = \mathbb{E}\delta$

2. $Var(g(T)) \le Var(\delta)$

3. $MSE(g) \le MSE(\delta)$

Rao-Blackwell.

① T is sufficient $\quad \delta(X)$ is an estimator of $g(\theta)$

② Define $g(T) = \mathbb{E}(\delta(X) \mid T)$

Thm.

①. T is CS

② $\mathbb{E}(\delta) = g(\theta)$

③ $g(T) \stackrel{\Delta}{=} \mathbb{E}[\delta \mid T]$ is the UMVUE for $g(\theta)$

Δ $X^{(n)}$ is C-S for θ in. $Unif(0, \theta)$

Δ $\sum_{i=1}^{n} X_i$ is C-S for λ in $Pois(\lambda)$

Δ $(\bar{X}, S^2)$ is C-S for μ, σ^2 in $N(\mu, \sigma^2)$

LSE $Y = X\beta + \varepsilon$

Aim: minimize. $(Y - X\beta)^T (Y - X\beta) \triangleq W = \varepsilon^T \varepsilon$

$$(Y^T - \beta^T X^T)(Y - X\beta)$$

$$Y^T Y + \beta^T X^T X \beta - Y^T X \beta - \beta^T X^T Y \triangleq W$$

$$\frac{\partial W}{\partial \beta} = (X^T X + X^T X)\beta - 2X^T Y = 0$$

$$X^T X \beta = X^T Y$$

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

$$E\hat{\beta} = (X^T X)^{-1} X^T E Y = (X^T X)^{-1} X^T X \beta = \beta.$$

$$\mathrm{Var}(\hat{\beta}) = (X^T X)^{-1} X^T \mathrm{Var}(Y) X (X^T X)^{-1}$$

$$= \sigma^2 \cdot \cancel{(X^T X)^{-1} X^T X} (X^T X)^{-1}$$

$$= \sigma^2 \cdot (X^T X)^{-1}.$$

设 $\tilde{\beta} = (X'X)^{-1} X' Y + CY = ((X'X)^{-1} X' + C) Y$

$$E(\tilde{\beta}) = E(\cancel{(X'X)^{-1} X' X}\beta) + E C X \beta$$

$$= \beta + CX\beta = \beta. \quad \boxed{CX = 0}$$

$$\mathrm{Var}(\tilde{\beta}) = ((X'X)^{-1}X' + C)((X'X)^{-1}X' + C)^T \sigma^2 I$$

$$= ((X'X)^{-1}X' + C)(X(X'X)^{-1} + C^T) \cdot \sigma^2 I$$

$$= [\cancel{(X'X)^{-1}X'X}(X'X)^{-1} + CC^T + (X'X)^{-1}\underbrace{X'C^T}_{0} + \underbrace{CX}_{0}(X'X)^{-1}] \sigma^2 I.$$

$$= (X'X)^{-1} \cdot \sigma^2 I + CC^T \sigma^2 I$$

semi positive definite.

MLE

invariant

always exist $\Rightarrow$ may not be unique.

Asymptotic efficiency

$$\tilde{L}(\eta) = \sup_{\{\theta\in\Theta : g(\theta)=\eta\}} L(\theta)$$

$$\eta \hat{=} g(\theta) \quad \text{pf} \quad \hat{\eta} = \widehat{g(\theta)} = g(\hat{\theta})$$

$$\tilde{L}(\hat{\eta}) = \sup_{\eta} \tilde{L}(\eta) = \sup_{\eta} \sup_{\{\theta\in\Theta : g(\theta)=\eta\}} L(\theta) = \sup_{\{\theta\in\Theta\}} L(\theta) = L(\hat{\theta})$$

Definition **4.2** *Suppose $g(\cdot)$: $\Theta \to \Lambda \in R^p$, and let $\Theta_\eta =: \{\theta \in \Theta : g(\theta) = \eta\}$.*

- *The* **induced likelihood function** *for $\eta = g(\theta)$ is defined as*

$$\widetilde{L}(\eta) = \sup_{\{\theta\in\Theta : g(\theta)=\eta\}} L(\theta).$$

- *For fixed $\mathbf{x}$, let $\hat{\eta} = \hat{\eta}(\mathbf{x})$ satisfy $\widetilde{L}(\hat{\eta}) = \sup_{\eta\in g(\Theta)} \widetilde{L}(\eta)$ i.e.,*

$$\hat{\eta} = \arg \sup_{\eta\in g(\Theta)} \widetilde{L}(\eta).$$

Such a $\hat{\eta}$ always exists in $\overline{g(\Theta)}$ (not necessarily in $g(\Theta)$), and is still called a **maximum likelihood estimator** *(MLE) of θ.*

Theorem **4.2 (MLE Invariance Theorem)** *Let $\mathbf{X} \sim f_\theta(\mathbf{x})$ and $\eta = g(\theta)$. If $\hat{\theta}$ is an MLE for θ, then $g(\hat{\theta})$ is an MLE for $\eta = g(\theta)$.*

Proof. It suffices to show that $\widetilde{L}(\hat{\eta}) = \widetilde{L}(g(\hat{\theta}))$. By definition, we have

$$\widetilde{L}(\hat{\eta}) = \sup_{\eta\in g(\Theta)} \widetilde{L}(\eta) = \sup_{\eta} \sup_{\{\theta\in\Theta : g(\theta)=\eta\}} L(\theta) = \sup_{\theta\in\Theta} L(\theta) = L(\hat{\theta}).$$

On the other hand, from the definition of $\tilde{L}(\eta)$, we have

$$\widetilde{L}(g(\hat{\theta})) = \sup_{\{\theta\in\Theta : g(\theta)=g(\hat{\theta})\}} L(\theta) = L(\hat{\theta}). \quad \blacksquare$$

$0 \le Var(\delta - \lambda^T Y) = cov(\delta - \lambda^T Y, \delta - \lambda^T Y) := g(\lambda)$

the minimum reaches. when $\lambda_0 = \frac{cov(\delta, Y)}{Var(Y)}$

$g(\lambda_0) = Var(\delta) - cov(\delta, Y) \cdot Var(Y)^{-1} cov(\delta, Y) \ge 0$

Lemma: $Var(\delta) \ge cov(\delta, Y) Var(Y)^{-1} \cdot cov(\delta, Y)$

The equality holds iff $\delta - \lambda_0^T Y = 0$.

where $\lambda_0 = \frac{cov(\delta, Y)}{Var(Y)}$

$Y := l'(\theta)$ $\quad EY = 0$. $\quad cov(\delta, Y) = E(\delta \cdot Y) = \int \delta \cdot \frac{f'_\theta(x)}{f_\theta(x)} \cdot f_\theta(x) \cdot dx$

$= \frac{\partial \log_\theta f(x)}{\partial \theta} = \frac{\frac{\partial f_\theta(x)}{\partial \theta}}{f_\theta(x)}$. $\quad = \frac{\partial}{\partial \theta} \cdot \int \delta \cdot f_\theta(x) \cdot dx$

$= \frac{\partial}{\partial \theta} \cdot E\delta(x)$.

$Var(Y) = I_x(\theta)$. $\quad = \frac{\partial}{\partial \theta} g(\theta) = g'(\theta)$.

Cramer-Rao Lower Bound.

$Var(\delta) \ge [E\delta(x)'_\theta]^T [I_x(\theta)]^{-1} [E\delta(x)']_\theta$

where the equality holds iff $\delta - \lambda_0^T l'(\theta) = c$

iff $\delta(x) - E\delta(x) = \lambda_0^T \cdot l'(\theta)$.

Information matrix 计算技巧

1. If $\mathbf{X} \perp \mathbf{Y}$, then $I_{(\mathbf{X},\mathbf{Y})}(\theta) = I_{\mathbf{X}}(\theta) + I_{\mathbf{Y}}(\theta)$.
2. If $\mathbf{X} = \{X_1, \ldots, X_n\} \sim_{i.i.d.} f_\theta(x)$, then $I_{\mathbf{X}}(\theta) = \sum_{i=1}^n I_{X_i}(\theta) = nI_{X_i}(\theta)$.
3. $I_{\mathbf{X}}(\theta) = -El''(\theta)$, where $\frac{\partial}{\partial \theta^\tau} \int \frac{\partial}{\partial \theta} f_\theta(\mathbf{x}) dx = \int \frac{\partial^2}{\partial \theta \partial \theta^\tau} f_\theta(\mathbf{x}) dx$, $\quad$ for $\theta \in \Theta$.

5.2 Properties of Fisher information matrix

1. If $\mathbf{X} \perp \mathbf{Y}$, then $I_{(\mathbf{X},\mathbf{Y})}(\theta) = I_{\mathbf{X}}(\theta) + I_{\mathbf{Y}}(\theta)$.
2. If $\mathbf{X} = \{X_1, \ldots, X_n\} \sim_{i.i.d.} f_\theta(x)$, then $I_{\mathbf{X}}(\theta) = \sum_{i=1}^n I_{X_i}(\theta) = nI_{X_i}(\theta)$.
3. $I_{\mathbf{X}}(\theta) = -El''(\theta)$, where $\frac{\partial}{\partial\theta^\tau} \int \frac{\partial}{\partial\theta} f_\theta(\mathbf{x})dx = \int \frac{\partial^2}{\partial\theta\partial\theta^\tau} f_\theta(\mathbf{x})dx, \quad$ for $\theta \in \Theta$.

   Proof. We only prove the case $k = 1$. Differentiating $1 = \int f_\theta(\mathbf{x})d\mathbf{x}$ w.r.t. θ, we get
   $$0 = \int l'(\theta) f_\theta(\mathbf{x})d\mathbf{x} = E\left(l'(\theta)\right).$$
   Differentiating this again w.r.t. θ gives
   $$0 = \int l''(\theta) f_\theta(\mathbf{x})d\mathbf{x} + \int \left(l'(\theta)\right)^2 f_\theta(\mathbf{x})d\mathbf{x} = E\left(l'(\theta)\right)^2 + El''(\theta).$$
4. Let $\theta = h(\eta)$ where $h : R^m \to R^k$ with $m, k \geq 1$. Then,
   $$I_{\mathbf{X}}(\eta) = h'(\eta)^\tau_{m\times k} I_{\mathbf{X}}(\theta) h'(\eta)_{k\times m}.$$
   Proof. From the definition,
   $$\begin{aligned} I_{\mathbf{X}}(\eta) &= E\left[\left(\frac{\partial l(\theta)}{\partial\eta}\right)^\tau \left(\frac{\partial l(\theta)}{\partial\eta}\right)\right] \\ &= E\left[\left(l'(\theta)h'(\eta)\right)^\tau \left(l'(\theta)h'(\eta)\right)\right] \\ &= \left(h'(\eta)\right)^\tau E\left[l'(\theta)^\tau l'(\theta)\right] h'(\eta) \\ &= h'(\eta)^\tau I_{\mathbf{X}}(\theta) h'(\eta) \end{aligned}$$
5. Let $\theta = \theta(\eta)$ is from R^m to R^k, where $m, k \geq 1$. Assume that $\theta(.)$ is differentiable. Then the Fisher information that $\mathbf{X}$ contains about η is
   $$I_{\mathbf{X}}(\eta) = \left(\frac{\partial\theta}{\partial\eta^\tau}\right)_{m\times k} I_{\mathbf{X}}(\theta) \left(\frac{\partial\theta^\tau}{\partial\eta}\right)_{k\times m}.$$
   Proof. From the definition,
   $$\begin{aligned} I_{\mathbf{X}}(\eta) &= E\left[\left(\frac{\partial l(\theta)}{\partial\eta}\right)^\tau \left(\frac{\partial l(\theta)}{\partial\eta}\right)\right] \\ &= E\left[\left(l'(\theta)\theta'(\eta)\right)^\tau \left(l'(\theta)\theta'(\eta)\right)\right] \\ &= \left(\theta'(\eta)\right)^\tau E\left[\left(l'(\theta)\right)^\tau \left(l'(\theta)\right)\right] \theta'(\eta) \\ &= \left(\theta'(\eta)\right)^\tau I_{\mathbf{X}}(\theta)\theta'(\eta) \end{aligned}$$
6. If $\theta = h(\eta)$ is 1-1 (from R^k to R^k), then the Cramer-Rao lower bound remains the same (i.e., reparametrization invariant).

   Proof. $X_1, ..., X_n \sim f_\theta(x)$, and $\theta = h(\eta)$ is 1-1. Now $E\delta(X) = g(\theta) = g(h(\eta)) = r(\eta)$. By C-R lower bound theorem,
   $$Var(\delta(\mathbf{X})) \geq \left(g'(\theta)\right)_{1\times k} \left[I_X(\theta)\right]^{-1}_{k\times k} \left(g'(\theta)\right)^\tau_{k\times 1} =: A,$$
   and also
   $$Var(\delta(\mathbf{X})) \geq \left(r'(\eta)\right)_{1\times k} \left[I_X(\eta)\right]^{-1}_{k\times k} \left(r'(\eta)\right)^\tau_{k\times 1} =: B.$$
   We shall show that $A = B$. Note that
   $$l'(\eta) = \frac{\partial l(\eta)}{\partial\eta} = \frac{\partial l(\eta)}{\partial\theta}\frac{\partial\theta^\tau}{\partial\eta} = \frac{\partial g(\theta)}{\partial\theta} h'(\eta)^\tau = g'(\theta)h'(\eta)^\tau,$$
   and that $\frac{\partial\theta^\tau}{\partial\eta}$ is $k \times k$ matrix of full rank. So from this and (4) above, we get
   $$B = \left(g'(\theta)h'(\eta)^\tau\right)\left(h'(\eta)^\tau\right)^{-1}\left[I_{\mathbf{X}}(\theta)\right]^{-1}\left(h'(\eta)\right)^{-1}\left(g'(\theta)h'(\eta)^\tau\right)^\tau = \left(g'(\theta)\right)\left[I_X(\theta)\right]^{-1}\left(g'(\theta)\right)^\tau = A.$$

5.6 MLE is asymptotically efficient

5.6.1 Asymptotically efficiency

Suppose that $\{\delta_n\}$ is a sequence of estimators $g(\theta)$. Then typically,

1. **(Consistency):** $\delta_n \to_p g(\theta)$.
2. **(Asymptotic normality):** $\sqrt{n}[\delta_n - g(\theta)] \to_d N(0, v(\theta))$.

If the lower bound in (4.5) is attained, δ_n is called *Asymptotic Efficient Estimators.*

Definition. A sequence $\{\delta_n\}$ is said to be **asymptotically efficient (AE)** if

$$\sqrt{n}[\delta_n - g(\theta)] \to_d N\left(0, \frac{[g'(\theta)]^2}{I(\theta)}\right).$$

1. *(Strong consistency)* $\quad \hat{\theta}_n \to_{a.s.} \theta_0$
2. *(Asymptotic efficiency)* $\quad \sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \to_d N\left(0, I_{X_1}^{-1}(\theta)\right).$

Remark **5.5** *To estimate $g(\theta)$ (g is smooth), we can use $g(\hat{\theta}_n)$. Consequently, we can apply the continuous mapping theorem and δ-method to obtain*

- *Strong consistency: $g(\hat{\theta}_n) \to g(\theta_0)$ with probability 1.*
- *Asymptotic efficiency: $\sqrt{n}\left(g(\hat{\theta}_n) - g(\theta_0)\right) \to_d N\left(0, \frac{[g'(\theta)]^2}{I_{X_1}(\theta)}\right).$*

Continuous Mapping Thm.

Slutsky's Thm. $X_n \xrightarrow{d} X \Rightarrow X_n = O_p(1)$

Stochastic Order Relation. $Y_n = O_p(1)$ $R_n = o(Y_n) \Rightarrow R_n = o_p(1)$

$X_n = O_p(1)$

$\Leftrightarrow \forall \varepsilon > 0 \ \exists M \ \text{s.t.} \ n > N \quad \mathbb{P}(|X_n| > M_\varepsilon) < \varepsilon$

$\Leftrightarrow \forall \varepsilon > 0 \ \exists M \quad \sup_n \mathbb{P}(|X_n| > M_\varepsilon) < \varepsilon$

$\Leftrightarrow \lim_{M \to \infty} \limsup_n \mathbb{P}(|X_n| > M_\varepsilon) = 0$

$X_n = o_p(1)$

$\Leftrightarrow X_n \xrightarrow{P} 0$

$\Leftrightarrow \lim_{n \to \infty} \mathbb{P}(|X_n - 0| > \varepsilon) \to 0 \quad \forall \varepsilon$

The δ-method

Thm1. $(X_1 \cdots X_n$ iid $\mathbb{E}X = \mu$ $Var(X) = \sigma^2 \in (0, \infty))$ $g'(\mu) \neq 0$ Then.

$g(\bar{x}) \xrightarrow{d} N(g(\mu), n^{-1}[g'(\mu)]^2 \sigma^2)$

CLT $\bar{X} \xrightarrow{d} N(\mu, \frac{\sigma^2}{n})$ $\frac{\sigma^2}{n} \to 0$ $(n \to \infty)$ $g'(\mu) \neq 0$

$g(\bar{x}) \to N(g(\mu), \frac{\sigma^2}{n} \cdot (g'(\mu))^2)$

Thm2. $T_n \xrightarrow{d} N(\mu, \sigma_n^2)$ with. $\underline{\sigma_n \to 0 \ (n \to \infty)}$ $g'(\mu) \neq 0$

$g(T_n) \xrightarrow{d} (g(\mu), g'(\mu)^2 \cdot \sigma_n^2)$

THEOREM **6.6** *Suppose T_n is $AN(\mu, \sigma_n^2)$ with $\sigma_n \to 0$. If $g(\cdot)$ is m times at μ differentiable with $g^{(j)}(\mu) = 0$ for $1 \le j \le m-1$ and $g^{(m)}(\mu) \neq 0$, then,*

$$\frac{g(T_n) - g(\mu)}{\frac{1}{m!} g^{(m)}(\mu) \sigma_n^m} \longrightarrow_d [N(0,1)]^m.$$

THEOREM **6.7** *If $X, X_1, ..., X_n$ are i.i.d. random k-vectors with $\mu = EX$ and $\Sigma = Cov(X, X)$. Assume that $g(\cdot) : R^k \to R^1$ is differentiable with $\nabla g(\mu) \neq 0$, where $\nabla g(x)_{k \times 1} = (\partial g(x)/\partial x_1, ..., \partial g(x)/\partial x_k)^\tau$. Then,*

$$g(\bar{X}) \longrightarrow_d N(g(\mu), v_n).$$

where $v_n = n^{-1} \nabla g(\mu)^\tau_{1 \times k} \Sigma_{k \times k} \nabla g(\mu)_{k \times 1}$. ∎

examples.

Example 2. Let $\theta = T(F) = \mu^2$, and $T_n = T(F_n) = (\bar{X})^2$. Then,

$$T_{n-1}^{(-i)} = \left(\frac{n\bar{X} - X_i}{n-1}\right)^2, \qquad T_{n-1}^{(\cdot)} = \frac{1}{n}\sum_{i=1}^{n} T_{n-1}^{(-i)}.$$

The i-th pseudo-value is $T_i^{pseudo} = nT_n - (n-1)T_{n-1}^{(-i)}$.

$$T_{n-1}^{(-i)} = \left(\frac{n\bar{X} - X_i}{n-1}\right)^2 \qquad T_{n-1}^{(\cdot)} = \frac{1}{n}\sum_{i=1}^{n} T_{n-1}^{(-i)}$$

$$T_i^{pseudo} = nT_n - (n-1)T_{n-1}^{(-i)}$$

$$T_n^{jack} = nT_n - (n-1)T_{n-1}^{(\cdot)} \qquad \text{where} \qquad T_{n-1}^{(\cdot)} = \frac{1}{n}\sum_{i=1}^{n} T_{n-1}^{(-i)}$$

$$\widehat{bias}(T_n) = T_n - T_n^{jack} = (n-1)T_{n-1}^{(\cdot)} - (n-1)T_n = (n-1)(T_{n-1}^{(\cdot)} - T_n)$$

$$T_{n-1}^{(-i)} = \frac{1}{(n-1)^2}\left[n^2\bar{X}^2 + X_i^2 - 2n\bar{X}X_i\right]$$

$$\sum_{i=1}^{n} T_{n-1}^{(-i)} = \frac{1}{(n-1)^2}\left[n^3\bar{X}^2 + \sum_{i=1}^{n} X_i^2 - 2n\bar{X}\sum_{i=1}^{n} X_i\right]$$

$$\frac{1}{n}\sum_{i=1}^{n} T_{n-1}^{(-i)} = \frac{1}{(n-1)^2}\left[n^2(\bar{X})^2 + \overline{X^2} - 2n(\bar{X})^2\right] = T_{n-1}^{(\cdot)}$$

$$\Rightarrow T_{n-1}^{(\cdot)} - T_n = \frac{1}{(n-1)^2}\left[\cancel{n^2(\bar{X})^2} - \cancel{2n(\bar{X})^2} + \overline{X^2} - \cancel{n^2(\bar{X})^2} + \cancel{2n(\bar{X})} - (\bar{X})^2\right]$$

$$= \frac{1}{(n-1)^2}\left[\overline{X^2} - (\bar{X})^2\right]$$

$$\widehat{bias}(T_n) = \frac{1}{(n-1)}\left[\overline{X^2} - (\bar{X})^2\right] = \frac{1}{n}\cdot\frac{1}{n-1}\left[n\overline{X^2} - n(\bar{X})^2\right] = \frac{1}{n}\cdot\hat{S}^2$$

$$\widehat{Var}_{jack}(T_n) = \frac{1}{n}\cdot\frac{1}{n-1}\cdot\sum_{i=1}^{n}\left(T_i^{pseudo} - T_n^{jack}\right)^2.$$

$$T_i^{pseudo} = nT_n - (n-1)\cdot T_{n-1}^{(-i)}$$

$$T_n^{jack} = nT_n - (n-1)\cdot T_{n-1}^{(\cdot)}$$

$$\left(T_i^{pseudo} - T_n^{jack}\right)^2 = (n-1)^2\cdot\left[T_{n-1}^{(\cdot)} - T_{n-1}^{(-i)}\right]^2$$

$$\sum_{i=1}^{n}\left(T_i^{pseudo} - T_n^{jack}\right)^2 = (n-1)^2\cdot\sum_{i=1}^{n}\left[(T_{n-1}^{(\cdot)})^2 - 2T_{n-1}^{(\cdot)}T_{n-1}^{(-i)} + (T_{n-1}^{(-i)})^2\right]$$

$$= (n-1)^2\left[n\cdot(T_{n-1}^{(\cdot)})^2 - 2n\cdot(T_{n-1}^{(\cdot)})^2 + \sum_{i=1}^{n}(T_{n-1}^{(-i)})^2\right]$$

$$= (n-1)^2\left[\sum_{i=1}^{n}(T_{n-1}^{(-i)})^2 - n\cdot(T_{n-1}^{(\cdot)})^2\right]$$

6.6.1 An example : the sample correlation coefficient

(X, Y) $(X_1 Y_1) \cdots (X_n Y_n)$ i.i.d. r. vectors. Assume $\mu_x = \mu_y = 0$

$$\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{cov(X Y)}{\sqrt{V(X) V(Y)}} = \frac{\mathbb{E} XY}{\sqrt{\mathbb{E}X^2 \, \mathbb{E}Y^2}}$$

$$\hat{\rho} = \frac{\frac{1}{n}\sum_{i=1}^{n} X_i Y_i}{\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right)^{1/2} \left(\frac{1}{n}\sum_{i=1}^{n} Y_i^2\right)^{1/2}}$$

(i) Find the limiting d.f of $\hat{\rho}$

$\theta = (\mathbb{E}X^2, \mathbb{E}Y^2 \ \mathbb{E}XY)$ $\qquad \rho = g(\theta)$

$\hat{\theta} = \left(\underbrace{\frac{1}{n}\sum X_i^2}_{z_1}, \underbrace{\frac{1}{n}\sum Y_i^2}_{z_2} \ \underbrace{\frac{1}{n}\sum_{i=1}^{n} X_i Y_i}_{z_3}\right)$ $\qquad , \hat{\rho} = g(\hat{\theta})$

$$g(z_1, z_2, z_3) = \frac{z_3}{(z_1 z_2)^{1/2}}$$

Note that

$\mathbb{E}\hat{\theta} = \theta$ Assume $\mathbb{E}X^4 < \infty$ $\mathbb{E}Y^4 < \infty$ from CLT

$\hat{\theta} \sim AN(\theta, \Sigma/n)$

where $\Sigma_{3\times3}$ is the covariance matrix of (X^2, Y^2, XY)

$$\Sigma = \begin{pmatrix} Var(X^2) & \cdots & \\ cov(X^2, Y^2) & \cdots & \\ cov(X^2, XY) & \cdots & V(XY) \end{pmatrix}$$

Therefore from the Thm. $\hat{\rho} = g(\hat{\theta}) \sim AN(\rho, n^{-1}[\nabla g(\theta)]^T \Sigma g(\theta))$

$$[\nabla g(\theta)]^T = \left(\frac{\partial g(\theta)}{\partial z_1} \cdots \frac{\partial g(\theta)}{\partial z_3}\right) = \left(-\frac{z_3}{2 z_1^{3/2} z_2^{1/2}}, -\frac{z_3}{2 z_1^{1/2} z_2^{3/2}}, \frac{1}{z_1^{1/2} z_2^{1/2}}\right)$$

$[\nabla g(\theta)]^T = (\text{———})$

$\hat{\theta} \xrightarrow{a.s.} \theta.$

$\hat{\rho} = g(\hat{\theta}) \xrightarrow{a.s.} g(\theta) = \rho.$

7.3 Asymptotic normality of the sample qunatiles

Asymptotic normality can be proved in several different ways, e.g. Bahadur representation. For simplicity, let $m = F^{-1}(1/2)$ and $\hat{m} = F_n^{-1}(1/2)$ be the population and sample medians, respectively. General quantiles can be treated similarly.

Theorem **7.4** *Assume $f(x) = F'(x)$ is continuous near m and $f(m) > 0$. Then*

$$\hat{m} \sim_{asymp.} N\left(m, \frac{1}{4f^2(m)n}\right).$$

Proof. WLOG, assume n is odd. Let $Y_{ni} = I\{X_i \le m + x/\sqrt{n}\}$. Then $\sum Y_{ni} \sim Bin(n, p_n)$, where $p_n = P(X_i \le m + x/\sqrt{n})$. Now

$$\begin{aligned}
P\left(\sqrt{n}(\hat{m} - m) \le x\right) &= P\left(\hat{m} \le m + x/\sqrt{n}\right) \\
&= P\left(\sum_{i=1}^{n} I\{X_i \le m + x/\sqrt{n}\} \ge \frac{n+1}{2}\right) \\
&= P\left(\sum_{i=1}^{n} Y_{ni} \ge \frac{n+1}{2}\right) \\
&= P\left(\frac{\sum_{i=1}^{n} Y_{ni} - np_n}{\sqrt{np_n(1-p_n)}} \ge \frac{\frac{n+1}{2} - np_n}{\sqrt{np_n(1-p_n)}}\right)
\end{aligned}$$

Since F is differentiable (hence continuous) at m, thus $F(m) = 1/2$. So

$$p_n = F(m + x/\sqrt{n}) = \frac{1}{2} + \frac{x}{\sqrt{n}} f\left(m + \frac{\delta x}{\sqrt{n}}\right) \to \frac{1}{2},$$

and $\sqrt{n}(p_n - 1/2) \to x f(m)$. Thus,

$$\frac{\frac{1}{2}(n+1) - np_n}{\sqrt{np_n(1-p_n)}} = \frac{\frac{1}{2\sqrt{n}} + \sqrt{n}(\frac{1}{2} - p_n)}{\sqrt{p_n(1-p_n)}} \to \frac{-xf(m)}{1/2} = -2xf(m)$$

Applying the CLT to the triangular array Y_{ni} and Slutsky's theorem, we have

$$P\left(\sqrt{n}(\hat{m} - m) \le x\right) \to 1 - \Phi(-2xf(m)) = \Phi(2xf(m)) = \Phi(x/(2f(m))^{-1}). \quad \blacksquare$$

THEOREM **9.1** *The Hoeffding-decomposition to 2-order U-statistics is*

$$U_n - \theta = \frac{2}{n}\sum_{i=1}^{n} g(X_i) + \frac{2}{n(n-1)}\sum_{i<j}\psi(X_i, X_j),$$

where $h_{ij} = h(X_i, X_j)$, $g_i = g(X_i) = E\left(h_{ij} \mid X_i\right) - \theta$ *and* $\psi(X_i, X_j) = h_{ij} - g_i - g_j + \theta$ *and all the terms on the RHS of* U_n *are uncorrelated.*

Proof. WLOG, assume that $\theta = 0$.

if $X_1 = x_1$ $X_i = x_1$

$$\begin{aligned} T_1 &= E(U_n|X_1) = \frac{2}{n(n-1)}\sum_{i<j} E[h(X_i, X_j)|X_1] \\ &= \frac{2}{n(n-1)}(n-1)E[h(X_1, X_j)|X_1] \\ &= \frac{2}{n}g_1, \\ E(U_n|X_1, X_2) &= \frac{2}{n(n-1)}\sum_{i<j} E[h(X_i, X_j)|X_1, X_2] \\ &= \frac{2}{n(n-1)}[h_{12} + (n-2)g_1 + (n-2)g_2] \end{aligned}$$

$\because X_1 \cdots X_n \overset{iid}{\sim}$.

$E[X|H] = \dfrac{\sum_{w\in H} X(w)}{|H|}$

$= \sum_{x\in\mathcal{X}} x \dfrac{\{w\in H \mid X(w)=x\}}{|H|}$

$= \dfrac{E\{I_H X\}}{P(H)}$

$= \int_{\mathcal{X}} x \cdot dP(x|H)$

Hence,

$$\begin{aligned} T_{12} &= E(U_n|X_1, X_2) - E(U_n|X_1) - E(U_n|X_2) \\ &= \frac{2}{n(n-1)}\{h_{12} + (n-2)[g_1 + g_2] - (n-1)[g_1 + g_2]\} \\ &= \frac{2}{n(n-1)}\left(h_{12} - g_1 - g_2\right) \\ &= \frac{2}{n(n-1)}\psi(X_1, X_2). \end{aligned}$$

To prove uncorrelatedness, we will show a few examples. First we note that $E[\psi(X_1, X_2)|X_1]\} = E[h(X_1, X_2) - g(X_1) - g(X_2)|X_1]\} = g(X_1) - g(X_1) - Eg(X_2) = 0.$ Hence,

$$Eg(X_1)\psi(X_1, X_2) = EE[g(X_1)\psi(X_1, X_2)|X_1] = E\{g(X_1)E[\psi(X_1, X_2)|X_1]\}$$

and

$$\begin{aligned} &E\psi(X_1, X_2)\psi(X_1, X_3) = EE[\psi(X_1, X_2)\psi(X_1, X_3)|X_1] \\ &= E\{E[\psi(X_1, X_2)|X_1]\}E\{E[\psi(X_1, X_3)|X_1]\} = 0. \quad \blacksquare \end{aligned}$$

9.3.1 Asymptotic Normality

THEOREM **9.2** *If* $Eh^2(X_1, X_2) < \infty$ *and* $\sigma_g^2 > 0$, *then*

$$\frac{\sqrt{n}(U_n - \theta)}{2\sigma_g} \longrightarrow_d N(0, 1).$$

Proof. Note $\sqrt{n}(U_n - \theta) = \frac{2}{\sqrt{n}}\sum_{i=1}^{n} g(X_i) + \sqrt{n}R_n$, where $R_n = \frac{2}{n(n-1)}\sum_{i<j}\psi(X_i, X_j)$. Now

$$Var(\sqrt{n}R_n) = n\frac{4}{n^2(n-1)^2}\sum_{i<j} Var(\psi(X_i, X_j)) = \frac{2n}{n(n-1)}Var(\psi(X_i, X_j)) \to 0,$$

implying $\sqrt{n}R_n \to_p 0$. Then apply the CLT and Slutsky theorem. $\blacksquare$

9.4 Jackknife variance estimates for U-statistics

Note that $Var(U_n) = 4Var[g(X_1)]/n + O(n^{-2})$ which is approximated by

$$\widetilde{Var}(U_n) \simeq \frac{4}{n}\frac{1}{n}\sum_{i=1}^{n}[g(X_i)]^2 = \frac{4}{n^2}\sum_{i=1}^{n} g^2(X_i).$$

Here $g(X_i) = E\{h(X_i, X_j)|X_i\}$ can be approximated by $g(X_i) \simeq \frac{1}{n-1}\sum_{j=1,j\neq i}^{n} h_{ij} - U_n$. Then an approximation to the variance of U_n is

$$\widehat{Var}(U_n) = \frac{4}{n^2}\sum_{i=1}^{n}\left(\frac{1}{n-1}\sum_{j=1,j\neq i}^{n} h_{ij} - U_n\right)^2.$$

This is almost the same as the jackknife variance estimator, $\widehat{Var}_{Jack}(U_n)$, given in the next theorem. In effect, $\widehat{Var}_{Jack}(U_n)$ estimates the variance of the dominating term in the H-decomposition of U_n.

THEOREM **9.3** *If U_n is a U-statistics of order 2 with kernel $h(x, y)$, then*

$$\widehat{Var}_{Jack}(U_n) = \frac{4(n-1)}{n(n-2)^2}\sum_{i=1}^{n}\left(\frac{1}{n-1}\sum_{j=1,i\neq j}^{n} h_{ij} - U_n\right)^2. \tag{4.1}$$

Furthermore, $\widehat{Var}_{Jack}(U_n)$ is a consistent estimator of $Var(U_n)$ in the sense that

$$\frac{\widehat{Var}_{Jack}(U_n)}{Var(U_n)} \longrightarrow_p 1, \qquad \text{as } n \to \infty. \tag{4.2}$$

Proof. We only derive (3.1) below (the proof for (3.2) is more involved and hence omitted). From Theorem 10.1, we have

$$\begin{aligned}
U_n &= \frac{1}{n(n-1)}\sum_{i\neq j} h_{ij} \\
U_{n-1}^{(-k)} &= \frac{1}{(n-1)(n-2)}\left(\sum_{i\neq j} h_{ij} - 2\sum_{j=1,j\neq k}^{n} h_{kj}\right) \\
U_{n-1}^{(\cdot)} &= \frac{1}{n}\sum_{k=1}^{n} U_{n-1}^{(-k)} \\
&= \frac{1}{(n-1)(n-2)}\left(\sum_{i\neq j}^{n} h_{ij} - \frac{2}{n}\sum_{k=1}^{n}\sum_{j=1}^{n} h_{kj} + \frac{2}{n}\sum_{k=1}^{n} h_{kk}\right) \\
&= \frac{1}{(n-1)(n-2)}\left(\sum_{i\neq j} h_{ij} - \frac{2}{n}\sum_{i\neq j} h_{ij}\right) \\
&= \frac{1}{(n-1)(n-2)}\left(\sum_{i\neq j} h_{ij} - 2(n-1)U_n\right) \\
\widehat{Var}_{Jack}(U_n) &= \frac{n-1}{n}\sum_{k=1}^{n}\left(U_{n-1}^{(-k)} - U_{n-1}^{(\cdot)}\right)^2
\end{aligned}$$

examples.

Example 2. Let $\theta = T(F) = \mu^2$, and $T_n = T(F_n) = (\bar{X})^2$. Then,

$$T_{n-1}^{(-i)} = \left(\frac{n\bar{X} - X_i}{n-1}\right)^2, \qquad T_{n-1}^{(\cdot)} = \frac{1}{n}\sum_{i=1}^{n} T_{n-1}^{(-i)}.$$

The i-th pseudo-value is $T_i^{pseudo} = nT_n - (n-1)T_{n-1}^{(-i)}$.

$$T_{n-1}^{(-i)} = \left(\frac{n\bar{X} - X_i}{n-1}\right)^2 \qquad T_{n-1}^{(\cdot)} = \frac{1}{n}\sum_{i=1}^{n} T_{n-1}^{(-i)}$$

$$T_i^{pseudo} = nT_n - (n-1)T_{n-1}^{(-i)}$$

$$T_n^{jack} = nT_n - (n-1)T_{n-1}^{(\cdot)} \quad \text{where} \quad T_{n-1}^{(\cdot)} = \frac{1}{n}\sum_{i=1}^{n} T_{n-1}^{(-i)}$$

$$\widehat{\text{bias}}(T_n) = T_n - T_n^{jack} = (n-1)T_{n-1}^{(\cdot)} - (n-1)T_n = (n-1)(T_{n-1}^{(\cdot)} - T_n)$$

$$T_{n-1}^{(-i)} = \frac{1}{(n-1)^2}[n^2\bar{x}^2 + x_i^2 - 2n\bar{x}x_i]$$

$$\sum_{i=1}^{n} T_{n-1}^{(-i)} = \frac{1}{(n-1)^2}\left[n^3\bar{x}^2 + \sum_{i=1}^{n} x_i^2 - 2n\bar{x}\sum_{i=1}^{n} x_i\right]$$

$$\frac{1}{n}\sum_{i=1}^{n} T_{n-1}^{(-i)} = \frac{1}{(n-1)^2}[n^2(\bar{x})^2 + \overline{x^2} - 2n(\bar{x})^2] = T_{n-1}^{(\cdot)}$$

$$\Rightarrow T_{n-1}^{(\cdot)} - T_n = \frac{1}{(n-1)^2}[\cancel{n^2(\bar{x})^2} - \cancel{2n(\bar{x})^2} + \overline{x^2} - \cancel{n^2(\bar{x})^2} + \cancel{2n(\bar{x})} - (\bar{x})^2]$$

$$= \frac{1}{(n-1)^2}[\overline{x^2} - (\bar{x})^2]$$

$$\widehat{\text{bias}}(T_n) = \frac{1}{(n-1)}[\overline{x^2} - (\bar{x})^2] = \frac{1}{n}\cdot\frac{1}{n-1}[n\overline{x^2} - n(\bar{x})^2] = \frac{1}{n}\cdot\hat{S}^2$$

$$\widehat{\text{Var}}_{jack}(T_n) = \frac{1}{n}\cdot\frac{1}{n-1}\cdot\sum_{i=1}^{n}(T_i^{pseudo} - T_n^{jack})^2.$$

$$T_i^{pseudo} = nT_n - (n-1)\cdot T_{n-1}^{(-i)}$$

$$T_n^{jack} = nT_n - (n-1)\cdot T_{(n-1)}^{(\cdot)}$$

$$(T_i^{pseudo} - T_n^{jack})^2 = (n-1)^2\cdot[T_{n-1}^{(\cdot)} - T_{n-1}^{(-i)}]^2$$

$$\sum_{i=1}^{n}(T_i^{pseudo} - T_n^{jack})^2 = (n-1)^2\cdot\sum_{i=1}^{n}[(T_{n-1}^{(\cdot)})^2 - 2T_{n-1}^{(\cdot)}T_{n-1}^{(-i)} + (T_{n-1}^{(-i)})^2]$$

$$= (n-1)^2\left[n\cdot(T_{n-1}^{(\cdot)})^2 - 2n\cdot(T_{n-1}^{(\cdot)})^2 + \sum_{i=1}^{n}(T_{n-1}^{(-i)})^2\right]$$

$$= (n-1)^2\left[\sum_{i=1}^{n}(T_{n-1}^{(-i)})^2 - n\cdot(T_{n-1}^{(\cdot)})^2\right]$$