

Analysis

[Matrix] $c^T c = \|c\|^2 = c_1^2 + \dots + c_k^2$, cc^T is $k \times k$ matrix with (i, j) th element as $c_i c_j$,

[Max, Min] $\max(a, b) = \frac{1}{2}(a + b + |a - b|)$, $\min(a, b) = \frac{1}{2}(a + b - |a - b|)$

[Real Convergence] $\lim_{n \rightarrow \infty} (1 + \frac{x}{n})^n = \exp(x)$

[Linear Algebra] ① $[[a, b], [c, d]]^{-1} = \frac{1}{\det(A)} [[d, -b], [-c, a]]^T$, ② $\frac{d}{dX} X^T A X = 2AX$, $\frac{d}{dX} A^T X A = A A^T$

③ $(x - \mu)^T \Sigma^{-1} (x - \mu) = x^T \Sigma^{-1} x - \mu^T \Sigma^{-1} x - x^T \Sigma^{-1} \mu + \mu^T \Sigma^{-1} \mu$ ④ $(A + B)^T = A^T + B^T$

[Binomial Theorem] $(x + y)^n = \sum_{k=0}^n \binom{n}{k} x^{n-k} y^k$

Probability

[Ordered statistics] $P(X_{(n)} \leq x) = P(X_1 \leq x)^n$ and $P(X_{(n)} = x) = nP(X_1 \leq x)^{n-1}P(X_1 = x)$,

$P(X_{(1)} \leq x) = 1 - [1 - P(X_1 \leq x)]^n$ and $P(X_{(1)} = x) = n[1 - P(X_1 \leq x)]^{n-1}P(X_1 = x)$

[Moments] $\mu^k = E(X^k) = \int x^k f(x) dx$

[Deduce $X = 0$] If $X \geq 0$ a.s. and $EX = 0$ then $X = 0$ a.s.

[Variance, Covariance] $Var(X) = E[(X - EX)(X - EX)^T]$, $Cov(X, Y) = E[(X - EX)(Y - EY)^T]$, $Corr(X, Y) = Cov(X, Y) / (\sigma_X \sigma_Y)$,
 $E(a^T X) = a^T EX$, $Var(a^T X) = a^T Var(X) a$, $Var(X) = E(X^2) - E^2(X)$, $Var(aX + bY) = a^b Var(X) + b^2 Var(Y) + 2ab Cov(X, Y)$

[CHF] $\phi_X(t) = E[\exp(\sqrt{-1}t^T X)] = E[\cos(t^T X) + \sqrt{-1} \sin(t^T X)] \forall t \in \mathcal{R}^d$, well defined with $|\phi_X| \leq 1$

[MGF] $\psi_X(t) = E[\exp(t^T X)] \forall t \in \mathcal{R}^d$,

[MGF properties] $\psi_{-X}(t) = \psi_X(-t)$, if $\psi(t) < \infty \forall \|t\| < \delta \Rightarrow E|X|^a < \infty \forall a > 1$ and $\phi_X(t) = \psi_X(\sqrt{-1}t)$

[Conditional Exp] $f_{X|Y}(x|Y = y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{f_{Y|X}(y|X=x)f_X(x)}{f_Y(y)}$

[Bias] $E(\hat{\mu} - \mu)$

[Symmetric distribution] $Y =^D -Y$, $E_{-Y}(Y) = E_Y(-Y)$, mean = median = mode

[Radon-Nikodym] $\lambda \ll \nu$, there exist unique f s.t. $\lambda(A) = \int_A f d\nu$, $A \in \mathcal{F}$ and $f(x, \theta) = \frac{d\lambda}{d\nu}$

[Gamma family] $E(0, \theta) = \Gamma(1, \theta)$, $\Gamma(\frac{n}{2}, 2) \sim \chi_n^2$, $X \sim U(0, 1) \Rightarrow -\log X \sim E(0, 1)$, $Y \sim \Gamma(n, \theta_0)$, $\frac{2Y}{\theta_0} \sim \Gamma(n, 2) \sim \chi_{2n}^2$

[Common result] ① $X \sim \Gamma(\alpha, \gamma)$, $E(X) = \alpha\gamma$, $Var(X) = \alpha\gamma^2$ ② $Z \sim N(0, 1)$, $W \sim \chi_p^2$, $\frac{Z}{\sqrt{W/p}} \sim t_p$ ③ $t^2 \sim \chi_a^2 / \chi_b^2 \sim F_{a,b}$

Integration Convergence

[MCT] $0 \leq f_1 \leq f_2 \leq \dots \leq f_n$ and $\lim_n f_n = f$ a.e. $\Rightarrow \int \lim_n f_n d\nu = \lim_n \int f_n d\nu$

[Fatou] $f_n \geq 0 \Rightarrow \int \liminf_n f_n d\nu \leq \liminf_n \int f_n d\nu$

[DCT] $\lim_{n \rightarrow \infty} f_n = f$ and $|f_n| \leq g$ a.e. $\Rightarrow \int \lim_n f_n d\nu = \lim_n \int f_n d\nu$. g is an integrable function.

[Interchange Diff and Int] ① $\partial f(\omega, \theta) / \partial \theta$ exists in (a, b) ② $|\partial f(\omega, \theta) / \partial \theta| \leq g(\omega)$ a.e. \Rightarrow

① $\partial f(\omega, \theta) / \partial \theta$ integrable in (a, b) ② $\frac{d}{d\theta} \int f(\omega, \theta) d\nu(\omega) = \int \frac{\partial f(\omega, \theta)}{\partial \theta} d\nu(\omega)$

[Change of Var] $Y = g(X)$, $X = g^{-1}(Y) = h(Y)$ and A_i disjoint, $f_Y(y) = \sum_{j:1 \leq j \leq m, y \in g(A_j)} \left| \det \left(\frac{\partial h_j(y)}{\partial y} \right) \right| f_X(h_j(y))$. Simple version:
 $f_Y(y) = |\det(\partial h(y) / \partial y)| f_X(h(y))$

Inequalities

[Cauchy-Schwarz] $Cov(X, Y)^2 \leq Var(X)Var(Y)$, and $E^2[XY] \leq EX^2 EY^2$

[Jensen] φ is convex $\Rightarrow \varphi(EX) \leq E\varphi(X)$ e.g. $(EX)^{-1} < E(X^{-1})$ and $E(\log X) < \log(EX)$

[Chebyshev] If $\varphi(-x) = \varphi(x)$, and φ non-decreasing on $[0, \infty) \Rightarrow \varphi(t)P(|X| \geq t) \leq \int_{\{|X| \geq t\}} \varphi(X) dP \leq E\varphi(X) \forall t \geq 0$. e.g.

$P(|X - \mu| \geq t) \leq \frac{\sigma_X^2}{t^2}$ and $P(|X| \geq t) \leq \frac{E|X|}{t}$

[Hölder] $p, q > 0$ and $1/p + 1/q = 1$ or $q = p/(p-1) \Rightarrow E|XY| \leq (E|X|^p)^{1/p} (E|Y|^q)^{1/q}$. Equality $\Leftrightarrow |X|^p$ and $|Y|^q$ linearly dependent

[Young] $ab \leq \frac{a^p}{p} + \frac{b^q}{q}$, equality $\Leftrightarrow a^p = b^q$

[Minkowski] $p \geq 1$, $(E|X + Y|^p)^{1/p} \leq (E|X|^p)^{1/p} + (E|Y|^p)^{1/p}$

[Lyapunov] for $0 < s < t$, $(E|X|^s)^{1/2} \leq (E|X|^t)^{1/t}$

[KL] $K(f_0, f_1) = E_0 \log \frac{f_0(X)}{f_1(X)} = \int \log \left(\frac{f_0(x)}{f_1(x)} \right) f_0(x) d\nu(x) \geq 0$ equality $\Leftrightarrow f_1(\omega) = f_0(\omega)$

Convergence

[a.s.] $X_n \xrightarrow{\text{a.s.}} X$ if $P(\lim_{n \rightarrow \infty} X_n = X) = 1$. Can show $\forall \epsilon > 0, \sum_{i=1}^{\infty} P(|X_n - X| > \epsilon) < \infty$ via BC lemma

[Infinity often] $\{A_n \text{ i.o.}\} = \cap_{j \geq 1} \cup_{n \geq j} A_n := \limsup_{n \rightarrow \infty} A_n$

[Borel-Cantelli lemmas]

(First BC) If $\sum_{n=1}^{\infty} P(A_n) < \infty$, then $P(A_n \text{ i.o.}) = 0$

(Second BC) Given pairwise independent events $\{A_n\}_{n=1}^{\infty}$, if $\sum_{n=1}^{\infty} P(A_n) = \infty$, then $P(A_n \text{ i.o.}) = 1$

[L^p] $X_n \xrightarrow{L^p} X$ if $\lim_{n \rightarrow \infty} E|X_n - X|^p = 0$, given $p > 0$, $E|X|^p < \infty$ and $E|X_n|^p < \infty$

[Probability] $X_n \xrightarrow{P} X$ if $\forall \epsilon > 0 \lim_{n \rightarrow \infty} P(|X_n - X| > \epsilon) = 0$. Can show $E(X_n) = X$, $\lim_{n \rightarrow \infty} Var(X_n) = 0$

[Distribution] $X_n \xrightarrow{D} X$ if $\lim_{n \rightarrow \infty} F_n(x) = F(x)$ for every $x \in \mathcal{R}$ at which F is continuous

[Relationships between convergence]

① $L^p \Rightarrow L^q \Rightarrow P$, $p > q$ ② $a.s. \Rightarrow P \Rightarrow D$ ③ $X_n \xrightarrow{D} C \Rightarrow X_n \xrightarrow{P} C$ ④ If $X_n \xrightarrow{P} X \Rightarrow \exists$ sub-seq s.t. $X_{n_j} \xrightarrow{\text{a.s.}} X$.

[Continuous mapping] If $g: \mathcal{R}^k \rightarrow \mathcal{R}$ is continuous and $X_n \xrightarrow{*} X$, then $g(X_n) \xrightarrow{*} g(X)$, where $*$ is either ① a.s. ② P ③ D .

[Convergence properties]

① Unique in limit: $X = Y$ if $X_n \rightarrow X$ and $X_n \rightarrow Y$ for ① a.s., ② P , ③ L^p . ④ If $F_n \rightarrow F$ and $F_n \rightarrow G$, then $F(t) = G(t) \forall t$

② Concatenation: $(X_n, Y_n) \rightarrow (X, Y)$ when ① P ② a.s. ③ $(X_n, Y_n) \xrightarrow{D} (X, c)$ only when c is constant.

③ Linearity: $(aX_n + bY_n) \rightarrow aX + bY$ when ① a.s. ② P ③ L^p ④ NOT for distribution.

④ Cramér-Wold device: for k -random vectors, $X_n \xrightarrow{D} X \Leftrightarrow c^T X_n \xrightarrow{D} c^T X$ for every $c \in \mathcal{R}^k$

[Lévy continuity] $X_n \xrightarrow{D} X \Leftrightarrow \phi_{X_n} \rightarrow \phi_X$ pointwise

[Scheffé's theorem] If $\lim_{n \rightarrow \infty} f_n(x) = f(x) \Rightarrow \lim_{n \rightarrow \infty} \int |f_n(x) - f(x)| d\nu = 0$ and $P_{f_n} \rightarrow P_f$. Useful to check pdf converge in distribution.

[Slutsky's theorem] If $X_n \xrightarrow{D} X$ and $Y_n \xrightarrow{D} c$ for constant c . Then $X_n + Y_n \xrightarrow{D} X + c$, $X_n Y_n \xrightarrow{D} cX$, $X_n / Y_n \xrightarrow{D} X/c$ if $c \neq 0$

[Skorohod's theorem] If $X_n \xrightarrow{D} X$, then $\exists Y, Y_1, Y_2, \dots$ s.t. $P_{Y_n} = P_{X_n}$, $P_Y = P_X$ and $Y_n \xrightarrow{\text{a.s.}} Y$

[δ-method - first order] If $\{a_n\} > 0$ and $\lim_{n \rightarrow \infty} a_n = \infty$ and $a_n(X_n - c) \xrightarrow{D} Y$ and $c \in \mathcal{R}$ and $g'(c)$ exists at c , then $a_n[g(X_n) - g(c)] \xrightarrow{D} g'(c)Y$

[δ-method - higher order] If $g^{(j)}(c) = 0$ for all $1 \leq j \leq m-1$ and $g^{(m)}(c) \neq 0$. Then $a_n^m[g(X_n) - g(c)] \xrightarrow{D} \frac{1}{m!}g^{(m)}(c)Y^m$

[δ-method - multivariate] If X_i, Y are k -vectors rvs and $c \in \mathcal{R}^k$ and $a_n[g(X_n) - g(c)] \xrightarrow{D} \nabla g(c)^T Y$

[Stochastic order - Real] for a constant $c > 0$ and all n , ① $a_n = O(b_n) \Leftrightarrow |a_n| \leq c|b_n|$ ② $a_n = o(b_n) \Leftrightarrow \lim_{n \rightarrow \infty} a_n/b_n = 0$

[Stochastic order - RV] ① $X_n = O_{\text{a.s.}}(Y_n) \Leftrightarrow P\{|X_n| = O(|Y_n|)\} = 1$ ② $X_n = o_{\text{a.s.}}(Y_n) \Leftrightarrow X_n/Y_n \xrightarrow{\text{a.s.}} 0$, ③ $\forall \epsilon > 0, \exists C_\epsilon > 0, n_\epsilon \in \mathcal{N} \text{ s.t. } X_n = O_P(Y_n) \Leftrightarrow \sup_{n \geq n_\epsilon} P(\{\omega \in \Omega : |X_n(\omega)| \geq C_\epsilon |Y_n(\omega)|\}) < \epsilon$ ④ If $X_n = O_P(1)$, $\{X_n\}$ is bounded in probability. ⑤ $X_n = o_P(Y_n) \Leftrightarrow X_n/Y_n \xrightarrow{P} 0$

[Stochastic Order Properties] ① If $X_n \xrightarrow{\text{a.s.}} X$, then $\{\sup_{n \geq k} |X_n|\}_k$ is $O_p(1)$. ② If $X_n \xrightarrow{D} X$ for a rvs, then $X_n = O_P(1)$ (tightness). ③ If $E|X_n| = O(a_n)$, then $X_n = O_P(a_n)$ ④ If $E|X_n| = o(a_n)$, then $X_n = o_P(a_n)$

[SLLN, iid] $E|X_1| < \infty \Leftrightarrow \frac{1}{n} \sum_{i=1}^n X_1 \xrightarrow{\text{a.s.}} EX_1$

[SLLN, non-identical but independent] If $\exists p \in [1, 2]$ s.t. $\sum_{i=1}^\infty \frac{E|X_i|^p}{i^p} < \infty$, then $\frac{1}{n} \sum_{i=1}^n (X_i - EX_i) \xrightarrow{\text{a.s.}} 0$

[USLLN, iid] Suppose ① $U(x, \theta)$ is continuous in θ for any fixed x ② for each θ , $\mu(\theta) = EU(X, \theta)$ is finite ③ Θ is compact ④ There exists function $M(x)$ s.t. $EM(X) < \infty$ and $|U(x, \theta)| \leq M(x)$ for all x, θ . Then $P\left\{\lim_{n \rightarrow \infty} \sup_{\theta \in \Theta} \left|\frac{1}{n} \sum_{i=1}^n U(X_j, \theta) - \mu(\theta)\right| = 0\right\} = 1$

[WLLN, iid] $a_n = E(X_1 I_{\{|X_1| \leq n\}}) \in [-n, n]$ $nP(|X_1| > n) \rightarrow 0 \Leftrightarrow \frac{1}{n} \sum_{i=1}^n X_i - a_n \xrightarrow{P} 0$

[WLLN, non-identical but independent] If $\exists p \in [1, 2]$ s.t. $\lim_{n \rightarrow \infty} \frac{1}{n^p} \sum_{i=1}^n E|X_i|^p = 0$, then $\frac{1}{n} \sum_{i=1}^n (X_i - EX_i) \xrightarrow{P} 0$

[Weak Convergency] $\int f d\nu_n \rightarrow \int f d\nu$ for every bounded and continuous real function f . $X_n \xrightarrow{D} X \Leftrightarrow E[h(X_n)] \rightarrow E[h(X)]$

[CLT, iid] Suppose $\Sigma = \text{Var} X_1 < \infty$, then $\frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - EX_i) \xrightarrow{D} N(0, \Sigma)$, or $\sqrt{n}(\bar{X} - E(X_1)) \xrightarrow{D} N(0, \Sigma)$

[CLT, non-identical but independent] Suppose ① $k_n \rightarrow \infty$ as $n \rightarrow \infty$ ② (Lindeberg's condition) $0 < \sigma_n^2 = \text{Var}\left(\sum_{j=1}^{k_n} X_{n_j}\right) < \infty$. ③

If for any $\epsilon > 0$, $\frac{1}{\sigma_n^2} \sum_{j=1}^{k_n} E\{(X_{n_j} - EX_{n_j})^2 I_{\{|X_{n_j} - EX_{n_j}| > \epsilon \sigma_n\}}\} \rightarrow 0$. Then $\frac{1}{\sigma_n} \sum_{j=1}^{k_n} (X_{n_j} - EX_{n_j}) \xrightarrow{D} N(0, 1)$

[Check Lindeberg condition] Option ① (Lyapunov condition) $\frac{1}{\sigma_n^{2+\delta}} \sum_{j=1}^{k_n} E|X_{n_j} - EX_{n_j}|^{2+\delta} \rightarrow 0$ for some $\delta > 0$

Option ② (Uniform boundedness) If $|X_{n_j}| \leq M$ for all n and j and $\sigma_n^2 = \sum_{j=1}^{k_n} \text{Var}(X_{n_j}) \rightarrow \infty$

[Feller's condition] Ensures Lindeberg's condition is sufficient and necessary (else only sufficient). $\lim_{n \rightarrow \infty} \max_{j \leq k_n} \frac{\text{Var}(X_{n_j})}{\sigma_n^2} = 0$

Exponential Families

[NEF] $f_\eta(X) = \exp\{\eta^T T(X) - \mathcal{C}(\eta)\} h(x)$, where $\eta = \eta(\theta)$ and $\mathcal{C}(\eta) = \log\left\{\int_\Omega \exp\{\eta^T T(X)\} h(X) dX\right\}$. NEF is full rank if Ξ contains open set in \mathcal{R}^p , $\Xi = \{\eta(\theta) : \theta \in \Theta\} \subset \mathcal{R}^p$. Suppose $X_i \sim f_i$ independently with f_i Exp Fam, then joint distribution X is also Exp Fam.

[Showing non Exp Fam] For an exp fam P_θ , there is nonzero measure λ s.t. $\frac{dP_\theta}{d\lambda}(\omega) > 0$ λ -a.e. and for all θ . Consider $f = \frac{dP_\theta}{d\lambda} I_{(t, \infty)}(x)$, $\int f d\lambda = 0, f \geq 0 \Rightarrow f = 0$. Since $\frac{dP_\theta}{d\lambda} > 0$ by assumption, then $I_{(t, \infty)}(x) = 0 \Rightarrow v([t, \infty)) = 0$. Since t is arbitrary, consider $v(\mathcal{R}) = 0$ (contradiction)

[NEF MGF] Suppose η_0 is interior point on Ξ , then $\psi_{\eta_0}(t) = \exp\{\mathcal{C}(\eta_0 + t) - \mathcal{C}(\eta_0)\}$ and is finite in neighborhood of $t = 0$.

[Normal MGF] $X \sim N(\mu, \sigma^2)$, $E(X - \mu) = 0$, $E(X - \mu)^2 = \sigma^2$, $E(X - \mu)^3 = 0$, $E(X - \mu)^4 = 3\sigma^4$

[NEF Moments] Let $A(\theta) = \mathcal{C}(\eta_0(\theta))$, $\frac{dA(\theta)}{d\theta} = \frac{d\mathcal{C}(\eta_0(\theta))}{d\eta_0(\theta)} \cdot \frac{d\eta_0(\theta)}{d\theta}$, $T(x) = \frac{\partial \mathcal{C}(\eta)}{\partial \eta}$ ① $E_{\eta_0} T = \frac{d\psi_{\eta_0}}{dt}|_{t=0} = \frac{d\mathcal{C}}{d\eta_0} = \frac{A'(\theta)}{\eta'_0(\theta)}$, ② $E_{\eta_0} T^2 = \mathcal{C}''(\eta_0) + \mathcal{C}'(\eta_0)^2$, ③ $\text{Var}(T) = \mathcal{C}''(\eta_0) = \frac{A''(\theta)}{[\eta'_0(\theta)]^2} - \frac{\eta_0(\theta)'' A'(\theta)}{[\eta_0(\theta)']^3} = \frac{\partial^2 \mathcal{C}(\eta)}{\partial \eta \partial \eta^T} = -\frac{\partial^2 \log \ell(\eta)}{\partial \eta \partial \eta^T}$

[NEF Differential] $G(\eta) := E_\eta(g) = \int g(\omega) \exp\{\eta^T T(\omega) - \mathcal{C}(\eta)\} h(\omega) d\nu(\omega)$ for η in interior of Ξ_g ① G is continuous and has continuous derivatives of all orders. ② Derivatives can be computed by differentiation under the integral sign. $\frac{dG(\eta)}{d\eta} = E_\eta\left[g(\omega) \left(T(\omega) - \frac{\partial}{\partial \eta} \mathcal{C}(\eta)\right)\right]$

where Ξ_g is set η such that $\int |g(\omega)| \exp\{\eta^T T(\omega) - \mathcal{C}(\eta)\} h(\omega) d\nu(\omega) < \infty$

[NEF Min Suff] ① If there exists $\Theta_0 = \{\theta_0, \theta_1, \dots, \theta_p\} \subset \Theta$ s.t. vectors $\eta_i = \eta(\theta_i) - \eta(\theta_0), i \in [1, p]$ are linearly independent in \mathcal{R}^p , then T is also minimal sufficient. Check $\det([\eta_1, \dots, \eta_p])$ is non-zero ② $\Xi = \{\eta(\theta) : \theta \in \Theta\}$ contains $(p+1)$ points that do not lie on the same hyperplane ③ Ξ is full rank.

[NEF complete and sufficient] If \mathcal{P} is NEF of full rank then $T(X)$ is complete and sufficient for $\eta \in \Xi$

[NEF MLE] $\hat{\theta} = \eta^{-1}(\hat{\eta})$ or solution of $\frac{\partial \eta(\theta)}{\partial \theta} T(x) = \frac{\partial \xi(\theta)}{\partial \theta}$

[NEF Fisher Info] If $\underline{I}(\eta)$ is fisher info natural parameter η , then $\text{Var}(T) = \underline{I}(\eta)$. Let $\psi = E[T(X)]$. Suppose $\bar{I}(\psi)$ is fisher info matrix for parameter ψ , then $\text{Var}(T) = [\bar{I}(\psi)]^{-1}$

[NEF RLEs] RLE regularity condition (1, 2, 3, 4) holds due to proposition 3.2 and theorem 2.1, and result for (3). Only need to check condition on Fisher Info, then when n is large, there exists $\hat{\eta}_n$ s.t. $g(\hat{\eta}_n) = \hat{\mu}_n$ and $\hat{\eta}_n \rightarrow_{\text{a.s.}} \eta$ $\sqrt{n}(\hat{\eta}_n - \eta) \rightarrow_D N\left(0, \left[\frac{\partial^2}{\partial \eta \partial \eta^T} \mathcal{C}(\eta)\right]^{-1}\right)$

Where $g(\eta) = \frac{\partial \mathcal{C}(\eta)}{\partial \eta}$ and $\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^n T(X_i)$

[UMP NEF] ① UMP $T(Y) = I(Y > c)$ ① $\eta(\theta)$ increasing and $H_1 : \theta \geq \theta_0$ ② $\eta(\theta)$ decreasing and $H_1 : \theta \leq \theta_0$ ③ Reverse inequalities $T(Y) = I(Y < c)$ ① $\eta(\theta)$ increasing and $H_1 : \theta \leq \theta_0$ ② $\eta(\theta)$ decreasing and $H_1 : \theta \geq \theta_0$

[UMP Normal results] Given $X_i \sim N(\mu, \sigma^2)$ and $H_0 : \sigma^2 = \sigma_0^2$ ① $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ independent to \bar{X} ② $V = \frac{1}{\sigma_0^2} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$ ③ $t = \frac{\sqrt{n}(\bar{X} - \mu)/\sigma}{\sqrt{V/(n-1)}} = \frac{Z}{\sqrt{V/(n-1)}} \sim t_{(n-1)}$ (only if $X_i \sim N$)

[UMPU NEF $\eta(\theta) = \theta$] Require: ① suff stat Y for θ ② suff and complete U for φ such that φ is full-rank, when θ is fixed and φ is unknown, use $E_\theta(\cdot|U)$ to denote expectation wrt $f_{\theta, \varphi}$ as U is sufficient for φ . $f_{\theta, \varphi}(x) = \exp\{\theta Y(x) + \varphi^T U(x) - \mathcal{C}(\theta, \varphi)\}$, ③ $T(Y, U) = I(c_1(U) < Y < c_2(U))$ s.t. $E_{\theta_1}[T|U = u] = E_{\theta_2}[T|U = u] = \alpha$

[UMPU NEF $H_0 : \theta \leq \theta_1$ or $\theta \geq \theta_2$ $H_1 : \theta_1 < \theta < \theta_2$] $T(Y, U) = I(Y < c_1(U) \text{ or } Y > c_2(U))$ s.t. $E_{\theta_1}[T|U = u] = E_{\theta_2}[T|U = u] = \alpha$

[UMPU NEF $H_0 : \theta = \theta_0$ $H_1 : \theta \neq \theta_0$] $T(Y, U) = I(Y < c_1(U) \text{ or } Y > c_2(U))$ s.t. $E_{\theta_0}[T|U = u] = \alpha$ and $E_{\theta_0}[TY|U = u] = \alpha E_{\theta_0}(Y|U = u)$

[UMPU NEF $H_0 : \theta \leq \theta_0$ $H_1 : \theta > \theta_0$] $T(Y, U) = I(Y > c(U))$ s.t. $E_{\theta_0}[T|U = u] = \alpha$

[UMPU Normal] Require UMPU NEF ①, ② and ③ $V(Y, U)$ independent of U under H_0

[UMPU Normal $H_0 : \theta \leq \theta_1$ or $\theta \geq \theta_2$ $H_1 : \theta_1 < \theta < \theta_2$] ④ V to be increasing in $Y \Rightarrow T(V) = I(c_1 < V < c_2)$ s.t. $E_{\theta_1}[T] = E_{\theta_2}[T] = \alpha$

[UMPU Normal $H_0 : \theta_1 \leq \theta \leq \theta_2$ $H_1 : \theta < \theta_1$ or $\theta > \theta_2$] ④ V to be increasing in $Y \Rightarrow T(V) = I(V < c_1 \text{ or } V > c_2)$ s.t. $E_{\theta_1}[T] = \alpha$

$E_{\theta_0}[T] = \alpha$
UMPU Normal $H_0 : \theta = \theta_0 \quad H_1 : \theta \neq \theta_0$ ④ $V(y, u) = a(u)y + b(u)$ and $a(u) > 0 \Rightarrow T(V) = I(V < c_1 \text{ or } V > c_2)$ s.t. $E_{\theta_0}[T] = \alpha$ and $E_{\theta_0}[TV] = \alpha E_{\theta_0}(V)$
UMPU Normal $H_0 : \theta \leq \theta_0 \quad H_1 : \theta > \theta_0$ ④ V to be increasing in $Y \Rightarrow T(V) = I(V > c)$ s.t. $E_{\theta_0}[T] = \alpha$
MLR for one-param exp fam $\eta(\theta)$ nondecreasing in $\theta \Rightarrow \eta'(\theta) > 0$.
UMPU results - one sample normal ① $H_0 : \sigma^2 = \sigma_0^2, H_1 : \sigma^2 \neq \sigma_0^2 \quad \theta = -(2\sigma^2)^{-1}, \varphi = n\mu/\sigma^2, Y = \sum X_i^2, U\bar{X} \quad V = (n-1)S^2 = Y - nU^2, S = \frac{1}{n-1} \sum (X_i - \bar{X}), V/\sigma^2 \sim \chi_{n-1}^2$ ② $H_0 : \mu \leq \mu_0, H_1 : \mu > \mu_0, Y = \bar{X}, U = \sum (X_i - \mu_0)^2, \theta = -n(\mu - \mu_0)/\sigma^2, \varphi = -(2\sigma^2)^{-1}, V = (\bar{X} - \mu_0)/\sqrt{U}$ increasing in Y , under $H_0, V \perp U, t(X) = \sqrt{n}(\bar{X} - \mu_0)/S = \sqrt{(n-1)nV(X)}/\sqrt{1-n[V(X)]^2} \sim t_{n-1}, t(X)$ increasing in V ③ $H_0 : \mu = \mu_0, H_1 : \mu \neq \mu_0, V = (\bar{X} - \mu_0)/\sqrt{U}$ symmetric about 0 under $H_0 : \mu = \mu_0$, reject H_0 when $|V| > d$ where $P_{H_0}(|V| > d) = \alpha$. consider same t as before.
Statistics
Sufficiency $T(X)$ is sufficient for $P \in \mathcal{P} \Leftrightarrow P_X(x|Y=y)$ is known and does not depend on P . T sufficient for \mathcal{P}_0 but not necessarily \mathcal{P}_1 . $\mathcal{P}_0 \subset \mathcal{P} \subset \mathcal{P}_1$.
Factorization theorem $T(X)$ is sufficient for $P \in \mathcal{P} \Leftrightarrow$ there are non-negative Borel functions h with ① $h(x)$ does not depend on P ② $g_P(t)$ which depends on P s.t. $\frac{dP}{d\nu}(x) = g_P(T(x))h(x)$
Minimal sufficiency T is minimal sufficient $\Leftrightarrow T = \psi(S)$ for any other sufficient statistics S . Min suff is unique and usually exist.
Min Suff-Method 1 (Theorem A) Suppose $\mathcal{P}_0 \subset \mathcal{P}$ and \mathcal{P}_0 -a.s. implies \mathcal{P} -a.s. If T is sufficient for $P \in \mathcal{P}$ and minimal sufficient for $P \in \mathcal{P}_0$, then T is minimal sufficient for $P \in \mathcal{P}$ (Theorem B) Suppose \mathcal{P} contains PDFs f_0, f_1, \dots w.r.t a σ -finite measure. ① Define $f_\infty(x) = \sum_{i=0}^\infty c_i f_i(x)$ and $T_i(x) = f_i(x)/f_\infty(x)$, then $T(X) = (T_0(X), T_1(X), \dots)$ is minimal sufficient for \mathcal{P} . Where $c_i > 0, \sum_{i=0}^\infty c_i = 1, f_\infty(x) > 0$. ② If $\{x : f_i(x) > 0\} \subset \{x : f_0(x) > 0\}$ for all i , then $T(X) = (f_1(x)/f_0(x), f_2(x)/f_0(x), \dots)$ is minimal sufficient for \mathcal{P}
Min Suff-Method 2 (Theorem C) If ① $T(X)$ is sufficient, and ② $\exists \phi$ s.t. for $\forall x, y. f_P(x) = f_P(y)\phi(x, y) \quad \forall P \in \mathcal{P} \Rightarrow T(x) = T(y)$. Then $T(X)$ is minimal sufficient for \mathcal{P}
Ancillary statistics A statistics $V(X)$ is ancillary for \mathcal{P} if its distribution does not depend on population $P \in \mathcal{P}$ (First-order ancillary) if $E_P[V(X)]$ does not depend on $P \in \mathcal{P}$
Completeness $T(X)$ is complete for $P \in \mathcal{P} \Leftrightarrow$ for any Borel function $g, E_P g(T) = 0$ implies $g(T) = 0$, boundedly complete $\Leftrightarrow g$ is bounded. Completeness + Sufficiency \Rightarrow Minimal Sufficiency
Basu's theorem If V is ancillary and T is boundedly complete and sufficient, then V and T are independent w.r.t any $P \in \mathcal{P}$
Completeness for Varying Support $\int_0^\theta g(x)x^{n-1}dx = 0 \Rightarrow g(\theta)\theta^{n-1} = 0, \Rightarrow g(\theta) = g(X_{(n)}) = 0$ and thus $X_{(n)}$ is complete
Fisher information $I(\theta) = E\left(\frac{\partial}{\partial \theta} \log f_\theta(X)\right)^2 = \int \left(\frac{\partial}{\partial \theta} \log f_\theta(X)\right)^2 f_\theta(X) d\nu(x) = E\left\{\frac{\partial}{\partial \theta} \log f_\theta(X) \left[\frac{\partial}{\partial \theta} \log f_\theta(X)\right]^T\right\}$
Parameterization If $\theta = \psi(\eta)$ and ψ' exists, $\tilde{I}(\eta) = \psi'(\eta)^2 I(\psi(\eta))$
Twice differentiable Suppose f_θ is twice differentiable in θ and $\int \frac{\partial^2}{\partial \theta^2} f_\theta(x) I_{f_\theta(x) > 0} d\nu = 0$, then $I(\theta) = -E\left[\frac{\partial^2}{\partial \theta \partial \theta^T} \log f_\theta(X)\right]$
Independent samples If $\int \frac{\partial}{\partial \theta} f_\theta(x) d\nu = 0$ holds, then $I_{(X,Y)}(\theta) = I_X(\theta) + I_Y(\theta)$, and $I_{(X_1, \dots, X_n)}(\theta) = nI_{X_1}(\theta)$
Comparing decision rules
Compare decision rules ① as good as if $R_{T_1}(P) \leq P_{T_2}(P). \quad \forall P \in \mathcal{P}$ ② better if $R_{T_1}(P) < R_{T_2}(P)$ for some $P \in \mathcal{P}$ (and T_2 is dominated by T_1). ③ equivalent if $R_{T_2}(P) = R_{T_2}(P)$ for all $P \in \mathcal{P}$
Optimal T_* is \mathcal{J} -optimal if T_* is as good as any other rule in \mathcal{J} ,
Admissibility $T \in \mathcal{J}$ is \mathcal{J} -admissible if no $S \in \mathcal{J}$ is better than T in terms of the risk.
Minimaxity $T_* \in \mathcal{J}$ is \mathcal{J} -minimax if $\sup_{P \in \mathcal{P}} R_{T_*}(P) \leq \sup_{P \in \mathcal{P}} R_T(P)$ for any $T \in \mathcal{J}$
Bayes Risk A form of averaging $R_T(P)$ over $P \in \mathcal{P}$. Bayes risk $r_T(\Pi) = \int_{\mathcal{P}} R_T(P) d\Pi(P)$, $R_T(\Pi)$ is Bayes risk of T wrt a known probability measure Π .
Bayes rule T_* is \mathcal{J} -Bayes rule wrt Π if $r_{T_*}(\Pi) \leq r_T(\Pi)$ for any $T \in \mathcal{J}$.
Finding Bayes rule Let $\tilde{\theta} \sim \pi, X|\tilde{\theta} \sim P_{\tilde{\theta}}$, then $r_\pi(T) = E\left[L(\tilde{\theta}, T(X))\right] = E\left[E\left[L(\tilde{\theta}, T(X))\right] | X\right]$ where E is taken jointly over $(\tilde{\theta}, X)$. Then find $T_*(x)$ that minimises the conditional risk.
Rao-Blackwell ① Suppose $L(P, a)$ is convex and T is sufficient and S_0 is decision rule satisfying $E_P[|S_0|] < \infty$ for all $P \in \mathcal{P}$. Let $S_1 = E[S_0(X)|T]$, then $R_{S_1}(P) \leq R_{S_0}(P)$. ② If $L(P, a)$ is strictly convex in a , and S_0 is not a function of T , then S_0 is inadmissible and dominated by S_1 .
MOM
MoM $\mu_j = E_\theta X^j = h_j(\theta), \Rightarrow \hat{\theta} = h_j^{-1}(\hat{\mu}_j)$. Provided h_j^{-1} exists and $\hat{\mu}_j = \frac{1}{n} \sum_{i=1}^n X_i^j$.
MOM asymptotic θ_n is unique if $h^{-1}(X)$ exists. Strongly consistent if h^{-1} is continuous via SLLN and continuous mapping. If h^{-1} is differentiable and $E|X_1|^{2k} < \infty$ then use CLT and δ -method. V_μ is $k \times k$ with $(i, j) = \mu_{i+j} - \mu_i \mu_j \quad \sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{D} N(0, [\nabla g]^T V_\mu \nabla g)$ MOM is \sqrt{n} -consistent, and if $k = 1 \text{ amse}_{\hat{\theta}_n}(\theta) = g'(\mu_1)^2 \sigma^2/n, \sigma^2 = \mu_2 - \mu_1^2$
MLE
MLE $\hat{\theta} = \arg \max_\theta L(\theta)$. Consider (a) boundary opint (b) $\partial L(\theta)/\partial \theta = 0$ and $\partial^2 L(\theta)/\partial \theta^2 < 0$ (Concave), note MLE may not exist
MLE Consistency Suppose ① Θ is compact ② $f(x|\theta)$ is continuous in θ for all x ③ There exists a function $M(x)$ s.t. $E_{\theta_0}[M(X)] < \infty$ and $|\log f(x|\theta) - \log f(x|\theta_0)| \leq M(x)$ for all x, θ ④ identifiability holds $f(x|\theta) = f(x|\theta_0) \quad \nu$ -a.e. $\Rightarrow \theta = \theta_0$. Then MLE estimate $\hat{\theta}_n \xrightarrow{\text{a.s.}} \theta_0$
RLE [Roots of the Likelihood Equation] θ that solves $\frac{\partial}{\partial \theta} \log L_n(\theta) = 0$
RLE regularity conditions Suppose ① Θ is open subset of \mathcal{R}^k ② $f(x|\theta)$ is twice continuously differentiable in θ for all x , and $\frac{\partial}{\partial \theta} \int f(x|\theta) d\nu = \int \frac{\partial}{\partial \theta} f(x|\theta) d\nu, \frac{\partial}{\partial \theta} \int \frac{\partial}{\partial \theta^T} f(x|\theta) d\nu = \int \frac{\partial^2}{\partial \theta \partial \theta^T} f(x|\theta) d\nu$. ③ $\Psi(x, \theta) = \frac{\partial^2}{\partial \theta \partial \theta^T} \log f(x|\theta)$, there exists a constant c and non-negative function H s.t. $EH(X) < \infty$ and $\sup_{\|\theta - \theta_*\| < c} \|\Psi(x, \theta)\| \leq H(x)$. ④ Identifiable
RLE consistency Under RLE regularity conditions, there exists a sequence of $\hat{\theta}_n$ s.t. $\frac{\partial}{\partial \theta} \log L_n(\hat{\theta}_n) = 0$ and $\hat{\theta}_n \rightarrow_{\text{a.s.}} \theta_*$.
RLE asymptotic normality Assume RLE regularity conditions, and $I(\theta) = \int \frac{\partial}{\partial \theta} \log f(x|\theta) \left[\frac{\partial}{\partial \theta} \log f(x|\theta)\right]^T d\nu(x)$ is positive definite and $\theta = \theta_*$ (for single X). Then any consistent sequence $\{\tilde{\theta}_n\}$ of RLE it holds $\sqrt{n}(\tilde{\theta}_n - \theta_*) \xrightarrow{D} N\left(0, \frac{1}{I(\theta_*)}\right)$
One-step MLE Often asym efficient, useful to adjust an non asym efficient estimators provided $\hat{\theta}_n^{(0)}$ is \sqrt{n} -consistent. $\hat{\theta}_n^{(1)} = \hat{\theta}_n^{(0)} -$

$$\left[\nabla s_n(\hat{\theta}_n^{(0)})\right]^{-1} s_n(\hat{\theta}_n^{(0)})$$

Unbiased Estimators

UMVUE $T(X)$ is UMVUE for $\theta \Leftrightarrow \text{Var}(T(X)) \leq \text{Var}(U(X))$ for any $P \in \mathcal{P}$ and any other unbiased estimator $U(X)$ of θ

Lehmann-Scheffé If $T(X)$ is sufficient and complete for θ . If θ is estimable, then there is a unique unbiased estimator of θ that is of the form $h(T)$.

UMVUE method1 Using Lehmann-Scheffé, suppose T is sufficient and complete manipulate $E(h(T)) = \theta$ to get $\hat{\theta}$.

UMVUE method2 Using Rao-Blackwellization. Find ① unbiased estimator of $\theta = U(X)$ ② sufficient and complete statistics $T(X)$ ③ then $E(U|T)$ is the UMVUE of θ by Lehmann-Scheffé.

UMVUE method3 Useful when no complete and sufficient statistics. Can use to find UMVUE, check if estimator is UMVUE, show nonexistence of UMVUE. $T(X)$ is UMVUE $\Leftrightarrow E[T(X)U(X)] = 0$

① T is unbiased estimator of η with finite variance, \mathcal{U} is set of all unbiased estimators of 0 with finite variances. ② $T = h(S)$, where S is sufficient and h is Borel function, \mathcal{U}_S is subset of \mathcal{U} consisting of Borel functions of S .

Using method3 ① Find $U(x)$ via $E[U(x)] = 0$ ② Construct $T = h(S)$ s.t. T is unbiased ③ Find T via $E[TU] = 0$

Corollary If T_j is UMVUE of η_j with finite variances, then $T = \sum_{j=1}^k c_j T_j$ is UMVUE of $\eta = \sum_{j=1}^k c_j \eta_j$. If T_1, T_2 are UMVUE of η with finite variances, then $T_1 = T_2$ a.s. $P, P \in \mathcal{P}$

Cramér-Rao Lower Bound Suppose ① Θ is an open set and P_θ has pdf f_θ ② f_θ is differentiable and $\frac{\partial}{\partial \theta} \int f_\theta(x) d\nu = \int \frac{\partial}{\partial \theta} f_\theta(x) d\nu = 0$. ③ $g(\theta)$ is differentiable and $T(X)$ is unbiased estimator of $g(\theta)$ s.t. $g'(\theta) = \frac{\partial}{\partial \theta} \int T(x) f_\theta(x) d\nu = \int T(x) \frac{\partial}{\partial \theta} f_\theta(x) d\nu, \theta \in \Theta$. Then $\text{Var}(T(X)) \geq \frac{g'(\theta)^2}{I(\theta)} = \left[\frac{\partial}{\partial \theta} g(\theta)\right]^T [I(\theta)]^{-1} \frac{\partial}{\partial \theta} g(\theta)$

CR LB for biased estimator $\text{Var}(T) \geq \frac{[g'(\theta) + b'(\theta)]^2}{I(\theta)}$

CR LB iff CR achieve equality ① $\Leftrightarrow T = \left[\frac{g'(\theta)}{I(\theta)}\right] \frac{\partial}{\partial \theta} \log f_\theta(X) + g(\theta)$ ② $\Leftrightarrow f_\theta(X) = \exp(\eta(\theta)T(x) - \xi(\theta))h(x)$, s.t. $\xi'(\theta) = g(\theta)\eta'(\theta)$ and $I(\theta) = \eta'(\theta)g'(\theta)$

UMVUE asymptotic Typically consistent, exactly unbiased, ratio of mse over Cramér-Rao LB converges to 1 (asym they are the same).

Other estimators

Upper semi-continuous (usc) $\lim_{\rho \rightarrow 0} \left\{ \sup_{\|\theta' - \theta\| < \rho} f(x|\theta') \right\} = f(x|\theta)$

USC in θ Suppose (1) Θ is compact with metric $d(\cdot, \cdot)$ (2) $f(x|\theta)$ is usc in θ and for all x (3) there exists a function $M(x)$ s.t. $E_{\theta_0}|M(X)| < \infty$ and $\log f(x|\theta) - \log f(x|\theta_0) \leq M(x)$ for all x and θ (4) for all $\theta \in \Theta$ and sufficiency small $\rho > 0$, $\sup_{d(\theta', \theta) < \rho} f(x|\theta')$ is measurable in x (5) identifiable $f(x|\theta) = f(x|\theta_0)$ ν -a.e. $\Rightarrow \theta = \theta_0$. Then $d(\hat{\theta}_n, \theta_0) \rightarrow_{\text{a.s.}} 0$

Asym Covariance Matrix $V_n(\theta)$ is $k \times k$ positive definite matrix called asym covariance matrix. $V_n(\theta)$ is usually in form of $n^{-\delta}V(\theta)$, higher δ means faster convergence. $[V_n(\theta)]^{-1/2}(\hat{\theta}_n - \theta) \rightarrow_D N_k(0, I_k)$

Information Inequalities $A \preccurlyeq B$ means $B - A$ is positive semi-definite. Suppose two estimators $\hat{\theta}_{1n}, \hat{\theta}_{2n}$ satisfy asym covariance matrix with $V_{1n}(\theta), V_{2n}(\theta)$. $\hat{\theta}_{1n}$ is asym more efficient than $\hat{\theta}_{2n}$ if (1) $V_{1n}(\theta) \preccurlyeq V_{2n}(\theta)$ for all $\theta \in \Theta$ and all large n (2) $V_{1n}(\theta) \prec V_{2n}(\theta)$ for at least one $\theta \in \Theta$ But note $\hat{\theta}_n$ is asym unbiased but CR LB might not hold even if regularity condition is satisfied.

M-estimators General method to find $\hat{\theta}_n$ maximises criterion function $S_\theta(x)$, for MLE $s_\theta(x) = \log f(x|\theta)$. $E_{\theta_0}s_\theta(X) < E_{\theta_0}s_{\theta_0}(X) \forall \theta \neq \theta_0$. $\theta \mapsto S_n(\theta) = \frac{1}{n} \sum_{i=1}^n s_\theta(X_i)$

Consistency of M-estimators $S_n(\theta)$ is random function while $S(\theta)$ is fixed s.t. $\sup_{\theta \in \Theta} |S_n(\theta) - S(\theta)| \rightarrow_P 0$ and for every $\rho > 0$ $\sup_{\theta: d(\theta, \theta_0) \geq \rho} S(\theta) < S(\theta_0)$. Then any sequence of estimators $\hat{\theta}_n$ with $S_n(\hat{\theta}_n) \geq S_n(\theta_0) - o_P(1)$ converges in probability to θ_0

Hodges' estimator $X_i \sim N(\theta, 1)$, $\hat{\theta}_n = \bar{X}_n$ if $\bar{X}_n \geq n^{-1/4}$ and $t\bar{X}_n$ otherwise. $V_n(\theta) = 1/n$ if $\theta \neq 0$ and t^2/n otherwise. if $\theta \neq 0$: $\sqrt{n}(\hat{\theta} - \theta) = \sqrt{n}(\bar{X}_n - \theta) - (1-t)\sqrt{n}\bar{X}_n I_{|\bar{X}_n| < n^{-1/4}}$ if $\theta = 0$: $= t\sqrt{n}(\bar{X}_n - \theta) + (1-t)\sqrt{n}\bar{X}_n I_{|\bar{X}_n| \geq n^{-1/4}}$

Super-efficiency Point where UMVUE failed Hodeges' estiamtor in information inequality (2). But under the basic regularity condition and if Fisher Information is positive definite at $\theta = \theta_*$, if $\hat{\theta}_n$ satisfies Asym covariance matrix, then there is a $\Theta_0 \subset \Theta$ with Lebesgue measure 0 s.t. information inequality (2) holds for any $\theta \notin \Theta_0$

Asym efficiency Assume Fisher Info $I_n(\theta)$ is well-defined and positive definite for every n , seq of estimators $\{\hat{\theta}_n\}$ satisfies asym cov matrix is asym efficient or asym optimal if and only if $V_n(\theta) = [I_n(\theta)]^{-1}$.

Asymptotics

Consistency of point estimators ① consistent $T_n(X) \xrightarrow{P} \theta$ ② strongly consistent $T_n(X) \xrightarrow{\text{a.s.}} \theta$ ③ a_n -consistent $a_n(T_n(X) - \theta) = O_P(1)$, $\{a_n\} > 0$ and diverge to ∞ ④ L_r -consistent $T_n(X) \xrightarrow{L^P} \theta$ for some fixed $r > 0$.

Remark on consistency A combination of LLN, CLT, Slutsky's, continuous mapping, δ -method are used. If T_n is (strongly) consistent for θ and g is continuous at θ then $g(T_n)$ is (strongly) consistent for $g(\theta)$

Affine estimator Consider $T_n = \sum_{i=1}^n c_{ni} X_i$ ① If $c_{ni} = c_i/n$ s.t. $\frac{1}{n} \sum_{i=1}^n c_i \rightarrow 1$ and $\sup_i |c_i| < \infty$ then T_n is strongly consistent. ② If population variance is finite, then T_n is consistent in mse $\Leftrightarrow \sum_{i=1}^n c_{ni} \rightarrow 1$ and $\sum_{i=1}^n c_{ni}^2 \rightarrow 0$

Asymptotic distribution $\{a_n\} > 0$ and either ① $a_n \rightarrow \infty$ ② $a_n \rightarrow a > 0$, s.t. $a_n(T_n - \theta) \xrightarrow{D} Y$. When estimator's expectations or second moment are not well defined, we need asymptotic behaviours.

Asymptotic bias $\bar{b}_{T_n} = EY/a_n$, asymptotically unbiased if $\lim_{n \rightarrow \infty} \bar{b}_{T_n}(P) = 0$, $b_{T_n}(P) := ET_n(X) - \theta$

Asymptotic expectation If $a_n \xi_n \xrightarrow{D} \xi$, $E|\xi| < \infty$, then asymptotic expectation of ξ_n is $E\xi/a_n$

Asymptotic MSE asymptotic expectation of $(T_n - \theta)^2$ or $\text{amse}_{T_n}(P) = EY^2/a_n^2$ (Remark) $EY^2 \leq \liminf_{n \rightarrow \infty} E[a_n^2(T_n - v)^2]$ (amse is no greater than exact mse)

Asymptotic variance $\sigma_{T_n}^2(P) = \text{Var}(Y)/a_n^2$

Asym Relative Efficiency $e_{T_{1n}, T_{2n}} = \text{amse}_{T_{2n}(P)} / \text{amse}_{T_{1n}(P)}$. Note efficiency of estimator T refers to $1/[I(\theta)MSE_T(\theta)]$

δ -method corollary If $a_n \rightarrow \infty$, g is differentiable at θ , $U_n = g(T_n)$. Then ① amse of U_n is $[g'(\theta)^2 EY^2]/a_n^2$ ② asym var of U_n is $[g'(\theta)^2 \text{Var}(Y)]/a_n^2$

Quantiles asymptotic $F(\theta) = \gamma \in (0, 1)$ and $\hat{\theta}_n := \lfloor \gamma n \rfloor$ -th order statistics, $F'(\theta) > 0$ and exists. $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{D} N\left(0, \frac{\gamma(1-\gamma)}{[F'(\theta)]^2}\right)$

Hypothesis testing

Hypothesis tests Let \mathcal{P} be a family of distributions, $\mathcal{P}_0 \subset \mathcal{P}, \mathcal{P}_1 = \mathcal{P} \setminus \mathcal{P}_0$. Hypothesis testing decides between $H_0 : P \in \mathcal{P}_0, H_1 : P \in \mathcal{P}_1$. Action space $\mathcal{A} = \{0, 1\}$, decision rule is called a test $T : \mathcal{X} \rightarrow \{0, 1\} \Rightarrow T(X) = I_C(X)$ for some $C \subset \mathcal{X}$. C is called the region/critical

region.

[0 - 1 loss] Common loss function for hypo test, $L(P, j) = 0$ for $P \in \mathcal{P}_j$ and $= 1$ for $P \in \mathcal{P}_{1-j}, j \in \{0, 1\}$ Risk $R_T(P) = P(T(X) = 1) = P(X \in C)$ if $P \in \mathcal{P}_0$ or $P(T(X) = 0) = P(X \notin C)$ if $P \in \mathcal{P}_1$

[Type I and II errors] Type I: H_0 is rejected when H_0 is true. $\beta_T(\theta_0) = E_{H_0}(T) \leq \alpha$ (within controlled with size α) Error rate: $\alpha_T(P) = P(T(X) = 1), P \in \mathcal{P}_0$ Type II: H_0 is accepted when H_0 is false. $1 - \beta_T(\theta)$ for $\theta \in \Theta_1$ Error rate: $1 - \alpha_T(P) = P(T(X) = 1), P \in \mathcal{P}_1$

[Power function of T] $\alpha_T(P)$, Type I and Type II error rates cannot be minimized simultaneously.

[Significance level] Under Neyman-Pearson framework, assign pre-specified bound α (significance level of test): $\sup_{P \in \mathcal{P}_0} P(T(X) = 1) \leq \alpha$

[size of test] α' is the size of the test $\sup_{P \in \mathcal{P}_0} P(T(X) = 1) = \alpha'$

[NP Test] Steps ① Find joint distribution $f(X)$ and determine MLR and/or NEF ② Formulate hypothesis H_0, H_1 - simple/composite about θ and not $f(\theta)$ ③ Form N-P test structure T_* ④ Find test distribution and rejection region.

[Generalised NP] ϕ is the T (Test framework) $\max_{\phi} \int \phi f_{m+1} d\nu$ s.t. $\int \phi f_i d\nu \leq t_i \forall i \in (1, m)$, (Required condition) If $\exists c_1, \dots, c_m$ s.t. $\phi_*(x) = I[f_{m+1}(x) > \sum_{i=1}^m c_i f_i(x)]$, then ϕ_* maximises objective function with equality constraint. If $c_i \geq 0$ then ϕ_* maximises with inequality constraint.

[Generalised NP - working example] $H_0 : \lambda = 1, \lambda = 2, H_1 : \lambda \in (1, 2)$ ① $\max \int \varphi(x) f_{\lambda}(x) dx$ with $\int \varphi f_{\lambda=1} dx = \int \varphi f_{\lambda=2} dx = \alpha$ ② by generalised NP lemma, $\varphi^*(x) = I(f_{\lambda} > k_1 f_{\lambda=1} + k_2 f_{\lambda=2}) = I(c_1 g(x) + c_2 g(x) < 1)$ ③ show c_i are positives. If c_i are both negative then test always reject H_0 . If c_i have opposite signs, or one of them equals zero, LHS of inequalities is monotone function of x and test will be one-sided test. Power will be monotone and unable to satisfy constraints for power function. So c_i must be positive. ④ Then $c_1 f_{\lambda=1} + c_2 f_{\lambda=2}$ is convex, and φ^* is two-sided test with form $\varphi^*(x) = I_{b_1, b_2}(x)$ ⑤ Find b_1, b_2 s.t. $\int \varphi^* f_{\lambda=1} dx = \int \varphi^* f_{\lambda=2} dx = \alpha$

[UMP] ① $H_0 : P = p_0, H_1 : P = p_1 \Rightarrow T(X) = I(p_1(X) > c p_0(X)), \beta_T(p_0) = \alpha$ ② $H_0 : \theta \leq \theta_0, H_1 : \theta > \theta_0 \Rightarrow T(Y) = I(Y > c), \beta_T(\theta_0) = \alpha$ ③ $H_0 : \theta \leq \theta_1$ or $\theta \geq \theta_2, H_1 : \theta_1 < \theta < \theta_2, \Rightarrow T(Y) = I(c_1 < Y < c_2), \beta_T(\theta_1) = \beta_T(\theta_2) = \alpha$

UMP Satisfy (1) pre-set size $\alpha = E_{H_0}(T)$ (2) max power $\beta_T(P) = E_{H_1}(T)$

[No UMP] $H_0 : \theta = \theta_1, H_1 : \theta \neq \theta_1$ and $H_0 : \theta \in (\theta_1, \theta_2), H_1 : \theta \notin (\theta_1, \theta_2)$

[N-P lemma] NP test has non-trivial power $\alpha < \beta_{H_1}(T)$ unless $P_0 = P_1$, and is unique up to γ (randomised test)

[Show T_* is UMP in simple hypothesis] UMP when $E_1[T_*] - E_1[T] \geq 0$, key equation: $(T_* - T)(f_1 - c f_0) \geq 0. \Rightarrow \int (T_* - T)(f_1 - c f_0) = \beta_{H_1}(T_*) - \beta_{H_1}(T) \geq 0$.

[UMP unique up to randomised test in simple hypothesis] $(T_* - T)(f_1 - c f_0) \geq 0, \int (T_* - T)(f_1 - c f_0) = 0 \Rightarrow (T_* - T)(f_1 - c f_0) = 0$ and $T_* = T$

[Composite hypothesis] Simple \Rightarrow Composite when $\beta_T(\theta_0) \geq \beta_T(\theta \in H_0)$ and/or $\beta_T(\theta_0) \leq \beta_T(\theta \in H_1)$ (or does not depend on θ). For MLR this is satisfied, others need to check.

[Monotone Likelihood Ratio] $\theta_2 > \theta_1$, increasing likelihood ratio in Y if $g(Y) = \frac{f_{\theta_2}(Y)}{f_{\theta_1}(Y)} > 1$ or $g'(Y) > 0$.

[Simultaneous Interval] $C_t(X), t \in \mathcal{T}$ are $1 - \alpha$ simultaneous confidence intervals for $\theta_t, t \in \mathcal{T} \Leftrightarrow \inf_{P \in \mathcal{P}} P(\theta_t \in C_t(X) \text{ for all } t \in \mathcal{T}) \geq 1 - \alpha$ asymptotic CI if $\lim_{n \rightarrow \infty} P(\theta_t \in C_t(X) \text{ for all } t \in \mathcal{T}) \geq 1 - \alpha$ **[Simultaneous methods]** (Bonferroni) adjust each paramter level to $\alpha_t = \alpha/k$ (Bootstrap) Monte Carlo percentile estimate (Multivariate Normal) $\|(X - \mu)/\sigma\|^2 < \chi_p^2$

[UMPU] Exists for one-param,

Asymptotic test

[LR test] $\lambda(X) = \frac{\sup_{\theta \in \Theta_0} \ell(\theta)}{\sup_{\theta \in \Theta} \ell(\theta)}$ Rejects $H_0 \Leftrightarrow \lambda(X) < c \in [0, 1]$. 1-param Exp Fam LR test is also UMP.

Assume MLE regularity condition, under $H_0, -2 \log \lambda(X) \rightarrow \chi_r^2$, where $r := \dim(\theta)$ $T(X) = I[\lambda(X) < \exp(-\chi_{r, 1-\alpha}^2/2)]$ where $\chi_{r, 1-\alpha}^2$ is the $(1 - \alpha)$ th quantile of χ_r^2 .

[Wald's test] $W_n = R(\hat{\theta})^T \{C(\hat{\theta})^T I_n^{-1}(\hat{\theta}) C(\hat{\theta})\}^{-1} R(\hat{\theta})$, where $C(\theta) = \partial R(\theta)/\partial \theta, I_n(\theta)$ is fisher info for $X_1, \dots, X_n, \hat{\theta}$ is unrestricted MLE/RLE of θ . **[Wald's test - easy case]** if $H_0 : \theta = \theta_0 \Rightarrow R(\theta) = \theta - \theta_0$, and $W_n = (\hat{\theta} - \theta_0)^T I_n(\hat{\theta}) (\hat{\theta} - \theta_0)$

[Rao's score test] $Q_n = s_n(\tilde{\theta})^T I_n^{-1}(\tilde{\theta}) s_n(\tilde{\theta})$. where score function $s_n(\theta) = \partial \log \ell(\theta)/\partial \theta, \tilde{\theta}$ is MLE/RLE of θ under $H_0 : R(\theta) = 0$.

[Asymptotic Tests] Same test structure for LR, Wald', Rao's score test. $H_0 : R(\theta) = 0, \lim_{n \rightarrow \infty} W_n, Q_n \sim \chi_r^2, T(X) = I(W_n > \chi_{r, 1-\alpha}^2)$ or $I(Q_n > \chi_{r, 1-\alpha}^2)$

Non-param tests

[Sign test] $X_i \sim^{iid} F, u$ is fixed constant, $p = F(u), \Delta_i = I(X_i - u \leq 0), P(\Delta_i = 1) = p, p_0 \in (0, 1) H_0 : p \leq p_0, H_1 : p > p_0 \Rightarrow T(Y) = I(Y > m), Y = \sum_{i=1}^n \Delta_i \sim \text{Bin}(n, p), m, \gamma$ s.t. $\alpha = E_{p_0}[T(Y)] H_0 : p = p_0, H_1 : p \neq p_0 \Rightarrow T(Y) = I(Y < c_1 \text{ or } Y > c_2), E_{p_0}[T] = \alpha$ and $E_{p_0}[TY] = \alpha n p_0$

[Permutation test] $X_{i1}, \dots, X_{in_i} \sim^{iid} F_i, i = 1, 2 H_0 : F_1 = F_2, H_1 : F_1 \neq F_2, \Rightarrow T(X)$ with $\frac{1}{n!} \sum_{z \in \pi(x)} T(z) = \alpha$ $\pi(x)$ is set of $n!$ points obtained from x by permuting components of x E.g. $T(X) = I(h(X) > h_m), h_m := (m + 1)^{th}$ largest $\{h(z) : z \in \pi(x)\}$ e.g $h(X) = |\bar{X}_1 - \bar{X}_2|$ or $|S_1 - S_2|$

[Rank test] $X_i \sim^{iid} F, \text{Rank}(X_i) = \#\{X_j : X_j \leq X_i\}, H_0 : F$ symm and 0, $H_1 : H_0$ false, R_+^o vector of ordered R_+ . (Wilcoxon) $T(X) = I[W(R_+^o) < c_1 \text{ or } W(R_+^o) > c_2], W(R_+^o) = J(R_{+1}^o/n) + \dots + J(R_{+n}^o/n)$ c_1, c_2 are $(m + 1)^{th}$ smallest/largest of $\{W(y) : y \in \mathcal{Y}\}, \gamma = \alpha 2^n / 2 - m$

[KS test] $X_i \sim^{iid} F H_0 : F = F_0, H_1 : F \neq F_0, \Rightarrow T(X) = I(D_n(F_0) > c), D_n(F) = \sup_{x \in \mathcal{R}} |F_n(x) - F(x)|$ With F_n Emp CDF, and for any $d, n > 0, P(D_n(F) > d) \leq 2 \exp(-2nd^2)$,

[Cramer-von test] Modified KS with $T(X) = I(C_n(F_0) > c), C_n(F) = \int \{F_n(x) - F(x)\}^2 dF(x) n C_n(F_0) \xrightarrow{D} \sum_{j=1}^{\infty} \lambda_j \chi_{1j}^2$, with $\chi_{1j}^2 \sim \chi_1^2$ and $\lambda_j = j^{-2} \pi^{-2}$

[Empirical LR] $X_i \sim^{iid} F, H_0 : \Lambda(F) = t_0, H_1 : \Lambda(F) \neq t_0, \Rightarrow T(X) = I(ELR_n(X) < c) ELR_n(X) = \frac{\ell(\hat{F}_0)}{\ell(\hat{F})}, \ell(G) = \prod_{i=1}^n P_G(\{x_i\}), G \in \mathcal{F}. (\mathcal{F} := \text{collection of CDFs}, P_G := \text{measure induced by CDF } G)$

Confidence set $C(X) : X \rightarrow \mathcal{B}(\Theta)$, Require $\inf_{P \in \mathcal{P}} P(\theta \in C(X)) \geq 1 - \alpha$, that is confidencen coeff should be more than level

[Pratt's theorem] Suppose $\text{vol}(C(x)) = \int_C(x) d\theta'$ is finite, then expected volume of $C(X) E[\text{vol}(C(X))] = \int_{\theta' \neq \theta} P(\theta' \in C(x)) d\theta'$

[Uniformly most accurate (UMA)] $\theta \in \Theta$ and $\Theta' \subset \Theta$ that does not contain true $\theta, C(X)$ is Θ' -UMA $\Leftrightarrow P(\theta' \in C(X)) \leq P(\theta' \in C_1(X))$ for any other $C_1(X)$ $C(X)$ is UMA \Leftrightarrow it is Θ' -UMA with $\Theta' = \{\theta\}^c \Rightarrow$ inverting $H_0 : \theta = \theta_0$ and $H_1 : \theta \neq \theta_0$

[CI via pivotal qty] $C(X) = \{\theta : c_1 \leq \mathcal{R}(X, \theta) \leq c_2\}$, not dependent on P common pivotal qty: $(X_i - \mu)/\sigma$

[invert accept region] $C(X) = \{\theta : x \in A(\theta)\}$, Acceptance region $A(\theta) = \{x : T_{\theta_0}(x) \neq 1\}. H_0 : \theta = \theta_0$, any H_1 satisfy

[Shortest CI] require unimodal wrt x at $x_0: f'(x_0) = 0, f'(x) < 0, x < x_0$ and $f'(x) > 0, x > x_0$

Pivotal $(T - \theta)/U$, f unimodal at x_0 Interval $[T - b_*U, T - a_*U]$, shortest when $f(a_*) = f(b_*) > 0$ $a_* \leq x_0 \leq b_*$

Pivotal T/θ , $x^2f(x)$ unimodal at x_0 Interval $[b_*^{-1}T, a_*^{-1}T]$ shortest when $a_*^2f(a_*) = b_*^2f(b_*) > 0$ $a_* \leq x_0 \leq b_*$

General CI Require $f > 0$, integrable, unimodal at x_0 , (Objective) $\min b - a$ s.t. $\int_a^b f(x)dx$ and $a \leq b$ (Solution) a_*, b_* satisfy ① $a_* \leq x_0 \leq b_*$ ② $f(a_*) = f(b_*) > 0$ ③ $\int_{a_*}^{b_*} f(x)dx = 1 - \alpha$ forms the shortest CI, note it has to exactly the formulation above.

Asymptotic CI Require $\lim_{n \rightarrow \infty} P(\theta \in C(X)) \geq 1 - \alpha$,

Asymptotic pivotal $\mathcal{R}_n(X, \theta) = \hat{V}_n^{-1/2}(\hat{\theta}_n - \theta)$ does not depend on P in limit

Asymptotic LR CI $C(X) = \left\{ \theta : \ell(\theta, \hat{\varphi}) \geq \exp(-\chi_{r,1-\alpha}^2 - \alpha/2)\ell(\hat{\theta}) \right\}$

Asymptotic Wald CI $C(X) = \left\{ \theta : (\hat{\theta} - \theta)^T \left[C^T \left(I_n(\hat{\theta}) \right)^{-1} C \right]^{-1} (\hat{\theta} - \theta) \leq \chi_{r,1-\alpha}^2 \right\}$

Asymptotic Rao CI $C(X) = \left\{ \theta : [s_n(\theta, \hat{\varphi})]^T [I_n(\theta, \hat{\varphi})]^{-1} [s_n(\theta, \hat{\varphi})] \leq \chi_{r,1-\alpha}^2 \right\}$

Bayesian

Bayes formula $\frac{dP_{\theta|X}}{d\Pi} = \frac{f_{\theta}(X)}{m(X)}$.

Bayes action $\delta(x)$ $\arg \min_a E[L(\theta, a)|X = x]$, when $L(\theta, a) = (\theta - a)^2$, $\delta(x) = E(\theta|X = x)$, and bayes risk $r_{\delta}(\theta) = Var(\theta|X)$

Generalised Bayes action $\arg \min_a \int_{\Theta} L(\theta, a) f_{\theta}(x) d\Pi$, works for improper prior where $\Pi(\Theta) \neq 1$

Interval estimation - Credible sets $P_{\theta|X}(\theta \in C) = \int_C p_X(\theta) d\lambda \geq 1 - \alpha$

HPD highest posterior density $C(x) = \{\theta : p_X(\theta) \geq c_{\alpha}\}$, often shortest length credible set. Is a horizontal line in the posterior density plot. Might not have exact confidence level $1 - \alpha$.

Hierachical Bayes With hyper-priors as hyper-parameters on the priors.

Empirical Bayes Estimate hyper-paramter via data using MoM (no MLE as not independent). $X_i \sim N(\mu, \sigma^2)$, $\mu|\xi \sim N(\mu_0, \sigma_0^2)$, σ^2 known, $\xi = (\mu_0, \sigma_0^2)$, Using MoM $E_{\xi}(X|\xi) = E_{\xi}(E[X|\mu, \xi]) = E_{\xi}(\mu|\xi) = \mu_0 \approx \bar{X}$, $E_{\xi}(X^2|\xi) = E_{\xi}(\mu^2 + \sigma^2|\xi) = \sigma^2 + \mu_0^2 + \sigma_0^2 \approx \frac{1}{n} \sum X_i^2 \Rightarrow \sigma_0^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 - \sigma^2$

Normal posterior Normal posterior $N(\mu_*(x), c^2)$ with prior unknown μ and known σ^2 : $\mu_*(x) = \frac{\sigma^2}{\sigma^2 + n\sigma_0^2} \mu_0 + \frac{n\sigma_0^2}{\sigma^2 + n\sigma_0^2} \bar{x}$, $c^2 = \frac{\sigma_0^2 \sigma^2}{n\sigma_0^2 + \sigma^2}$

$C(x) = [\mu_*(x) - cz_{1-\alpha/2}, \mu_*(x) + cz_{1-\alpha/2}]$.

Bayesian Bias Under squared error loss, $\delta(X)$ is biased unless $r_{\delta}(\Pi) = 0$. *Not applicable to improper priors.*

Bayesian minimax If T is (unique) Bayes estimator under Π and $R_T(\theta) = \sup_{\theta'} R_T(\theta')$ π -a.e., then T is (unique) minimax. *Limit of Bayes estimators* If T has constant risk and $\liminf_j r_j \geq R_T$, then T is minimax. $\sup_{\theta \in \Theta} R_S(\theta) \geq \int_{\Theta_{\pi}} R_S(\theta) d\pi \geq \int_{\Theta_{\pi}} R_T(\theta) d\pi = \sup_{\theta \in \Theta} R_T(\theta)$

Bayesian admissibility (1) $\delta(X)$ unique \Rightarrow admissible, (2, 3) $r_{\delta}(\Pi) < \infty$, $\Pi(\theta) > 0$ for all θ and δ is Bayes action with respect to $\Pi \Rightarrow$ admissible. *Not true for improper priors unless it is limit of Bayes rule*, Improper priors require excessive risk ignorable, take limit and observe if risk is admissible.

Simultaneous credible set Simultaneous estimate vector-valued \mathcal{V} with e.g. squared loss $L(\theta, a) = \|a - \theta\|^2 = \sum_{i=1}^p (a_i - \theta_i)^2$

Bayes Asymptotic Property (Posterior Consistency) $X \sim P_{\theta_0}$ and $\Pi(U|X_n) \xrightarrow{P_{\theta_0}} 1$ for all open U containing θ_0 . (Wald type consistency) Assume $p_{\theta}(x)$ is continuous, measurable, θ_* is unique maximizer then MLE converge to true parameter θ^* P_* a.s. Furthermore, if θ^* is in the support of the prior, then posterior converges to θ^* in probability. (Posterior Robustness) all priors that lead to consistent posteriors are equivalent.

Bernstein-von Mises Assume MLE regularity conditions, posterior $T_n = \sqrt{n}(\tilde{\theta}_n - \hat{\theta}_n) \sim \mathcal{N}(0, V^*)$ asymptotically. (Well-specified) $V^* = I(\theta^*)^{-1} = E_*[-\nabla_{\theta}^2 \log p_{\theta^*}(Y)]^{-1}$ (same as MLE, with θ^* as true parameter, CI = CR) $\sqrt{n}(\hat{\theta}_n - E_{\theta}[\theta|X_1, \dots, X_n]) \xrightarrow{P} 0$ (If MLE has asym normality, so is posterior mean) (Mis-specified) $V^* = E_*[-\nabla_{\theta}^2 \log p_{\theta^*}(Y)]^{-1}$, θ_* is projection of θ^* onto parameter space, or unique maximizer of $\ell^*(\theta) = E_*[\log p_{\theta}(Y)]$

MLE asymptotic variance under model misspecification $E_*[-\nabla_{\theta}^2 \log p_{\theta^*}(Y)]^{-1} Var_*(\nabla \log p_{\theta^*}(Y)) E_*[-\nabla_{\theta}^2 \log p_{\theta^*}(Y)]^{-1}$ (differ from bayesian CR, θ_* is the projection of P_* to parameter space)

Linear Model

Linear Model $X = Z\beta + \epsilon$ (or $X_i = Z_i^T \beta + \epsilon_i$) Estimate with $b = \min_b \|X - Zb\|^2 = \|X - Z\hat{\beta}\|^2$, **Generalised inverse** Moore-Penrose inverse $A^+AA^+ = A^+$, $A = (Z^T Z)$ **Projection matrix** $P_Z = Z(Z^T Z)^{-1}Z^T$, $P_Z^2 = P_Z$, $P_Z Z = Z$, $rank(P_Z) = tr(P_Z) = r$

LM Solution (solution = normal equation) $Z^T Z \hat{\beta} = Z^T X$ (when Z is full rank): $\hat{\beta} = (Z^T Z)^{-1}Z^T X$ (when Z is not full rank): $\hat{\beta} = (Z^T Z)^- Z^T X$

LM MLE $X_i = \beta_0 + \beta_1 t_i + \epsilon_i$, $D := \sum (t_i - \bar{t})^2$, $\hat{\beta}_1 = \frac{1}{D}(\sum t_i(X_i - \bar{X}))$ and $\hat{\beta}_0 = \bar{X} - \hat{\beta}_1 \bar{t}$ and $\hat{\sigma}^2 = \frac{1}{n} \sum (X_i - \bar{X})^2 - \frac{D}{n} \hat{\beta}_1^2$ (?)

LM solution property ① $E(\hat{\beta}) = \beta$, ② $Var((Z^T Z)^{-1}Z^T X) = (Z^T Z)^{-1}Z Var(X)Z^T (Z^T Z)^{-1}$ ③ $Var(\ell^T \hat{\beta}) = \sigma^2 \ell^T (Z^T Z)^{-1} \ell$ ④ $Cov(\hat{\beta}) = \sigma^2 \begin{pmatrix} n & \sum t_i \\ \sum t_i & \sum t_i^2 \end{pmatrix}^{-1} = \frac{\sigma^2}{nD} \begin{pmatrix} \sum t_i^2 & -\sum t_i \\ -\sum t_i & n \end{pmatrix}$

LM tricks $X - Z\hat{\beta} = P_{Z\perp} X$, $Z\hat{\beta} = P_Z X$. $\exists W \in \mathcal{R}^{n \times (n-r)}$ s.t. $W^T W = I_{n-r}$ and $WW^T = P_{Z\perp} = I_n$

LM assumptions (A1 Gaussian noise) $\epsilon \sim N_n(0, \sigma^2 I_n)$ (A2 homoscedastic noise) $E(\epsilon) = 0$, $Var(\epsilon) = \sigma^2 I_n$ (A3 general noise) $E(\epsilon) = 0$, $Var(\epsilon) = \Sigma$

Estimable $\ell\beta$ Estimate linear combination of coefficient (General) necessary and Sufficient condition: $\ell \in R(Z) = R(Z^T Z)$ (under A3) LSE $\ell^T \hat{\beta}$ is unique and unbiased (under A1) if $\ell \notin R(Z)$, $\ell^T \beta$ not estimable

LM property under A1 ① LSE $\ell^T \hat{\beta}$ is UMVUE of $\ell^T \beta$, ② UMVUE of $\hat{\sigma}^2 = (n - r)^{-1} \|X - Z\hat{\beta}\|^2$, r is rank of Z ③ $\ell\hat{\beta}$ and $\hat{\sigma}^2$ are independent, $\ell^T \hat{\beta} \sim N(\ell^T \beta, \sigma^2 \ell^T (Z^T Z)^{-1} \ell)$, $(n - r)\hat{\sigma}^2/\sigma^2 \sim \chi_{n-r}^2$

LM property under A2 LSE $\ell^T \hat{\beta}$ is BLUE (Best Linear Unbiased Estimator, best as in min var)

LM property under A3 Following are equivalent: ① $\ell^T \hat{\beta}$ is BLUE for $\ell^T \beta$ (also UMVUE), ② $E[\ell^T \hat{\eta}^T X] = 0$, any η is s.t. $E[\eta^T X] = 0$ ③ $Z^T var(\epsilon)U = 0$, for U s.t. $Z^T U = 0$, $R(U^T) + R(Z^T) = R^n$ ④ $Var(\epsilon) = Z\Lambda_1 Z^T + U\Lambda_2 U^T$, for some Λ_1, Λ_2, U s.t. $Z^T U = 0$, $R(U^T) + R(Z^T) = R^n$ ⑤ $Z(Z^T Z)^- Z^T Var(\epsilon)$ is symmetric

LM consistency $\lambda_+[A]$ is the largest eigenvalue of $A_n = (Z^T Z)^-$. Suppose $\sup_n \lambda_+[Var\epsilon] < \infty$ and $\lim_{n \rightarrow \infty} \lambda_+[A_n] = 0$, $\ell^T \hat{\beta}$ is consistent in MSE.

[LM asymptotic normality] $\ell^T(\hat{\beta} - \beta)/\sqrt{\text{Var}(\ell^T \hat{\beta})} \xrightarrow{D} N(0, 1)$. sufficient condition: $\lambda_+[A_n] \rightarrow 0$ and $Z_n^T A_n Z_n \rightarrow 0$ as $n \rightarrow \infty$ and there exist $\{a_n\}$ s.t. $a_n \rightarrow \infty$, $a_n/a_{n+1} \rightarrow 1$ and $Z^T Z/a_n$ converge to positive definite matrix.

[LM Hypothesis testing] Under A1, $\ell \in R(Z)$, θ_0 fixed constant

[LM hypothesis testing - simple] $\ell \in R(Z)$, \textcircled{a} $H_0 : \ell^T \beta \leq \theta_0$, $H_1 : \ell^T \beta > \theta_0$, \textcircled{b} $H_0 : \ell^T \beta = \theta_0$, $H_1 : \ell^T \beta \neq \theta_0$, Under $H_0 :$

$$t(X) = \frac{\ell^T \hat{\beta} - \theta_0}{\sqrt{\ell^T (Z^T Z)^{-1} \ell} \sqrt{SSR/(n-r)}} \sim t_{n-r}, \text{ UMPU reject } t(X) > t_{n-r, \alpha} \text{ or } |t(X)| > t_{n-r, \alpha/2}$$

[LM hypothesis testing - multiple] $L_{s \times p}$, $s \leq r$ and all rows $= \ell_j \in R(Z)$ \textcircled{a} $H_0 : L\beta = 0$, $H_1 : L\beta \neq 0$ Under H_0 : $W = \frac{(\|X - Z\hat{\beta}_0\|^2 - \|X - Z\hat{\beta}\|^2)/s}{\|X - Z\hat{\beta}\|^2/(n-r)} \sim F_{s, n-r}$ with non-central param $\sigma^{-2} \|Z\beta - \Pi_0 Z\beta\|^2$, reject $W > F_{s, n-r, 1-\alpha}$

[LM confidence set] Pivotal qty: $\mathcal{R}(X, \beta) = \frac{(\hat{\beta} - \beta)^T Z^T Z (\hat{\beta} - \beta)/p}{\|X - Z\hat{\beta}\|^2/(n-p)} \sim F_{p, n-p}$, where $\hat{\beta}$ is LSE of β , $C(X) = \{\beta : \mathcal{R}(X, \beta) \leq F_{p, n-p, 1-\alpha}\}$

[CI for $H_0 : \theta = \theta_0$, $H_1 : \theta < \theta_0$] $A(\theta_0) = \left\{ X : \ell^T \hat{\beta} - \theta_0 > -t_{n-r, \alpha} \sqrt{\ell^T (Z^T Z)^{-1} \ell SSR/(n-r)} \right\}$

[CI For $H_0 : \theta = \theta_0$, $H_1 : \theta \neq \theta_0$] $A(\theta_0) = \left\{ X : |\ell^T \hat{\beta} - \theta_0| < t_{n-r, \alpha/2} \sqrt{\ell^T (Z^T Z)^{-1} \ell SSR/(n-r)} \right\}$

[Asymptotic CI] Does not require normality of noise $C(X) = \left\{ \beta : (\hat{\beta} - \beta)^T (Z^T Z) (\hat{\beta} - \beta) \leq \chi_{p, \alpha}^2 SSR/(n-p) \right\}$ $SSR = \|X - Z\hat{\beta}\|^2$

[Linear Estimator] Linear estimator for linear model $X = Z\beta + \epsilon$ is linear function of X . e.g. $\ell^T \hat{\beta} = \ell^T (Z^T Z)^{-1} Z^T X = C^T X$, $\text{Var}(C^T X) = C^T \text{Var}(X) C = C^T \text{Var}(\epsilon) C$

[Bivariate Normal density] X_i are iid from bivariate normal $f(X) = \frac{1}{(2\pi\sigma_1\sigma_2\sqrt{1-\rho^2})^n} \exp \left\{ -\frac{\|Y_1 - \mu_1 1_n\|^2}{2\sigma_1^2(1-\rho^2)} + \frac{\rho(Y_1 - \mu_1 1_n)^T (Y_2 - \mu_2 1_n)}{\sigma_1\sigma_2(1-\rho^2)} - \frac{\|Y_2 - \mu_2 1_n\|^2}{2\sigma_2^2(1-\rho^2)} \right\}$

Testing $H_0 : \rho = 0$, $H_1 : \rho \neq 0$ with $\theta = \frac{\rho}{\sigma_1\sigma_2(1-\rho^2)}$, $Y = \sum_{i=1}^n X_{i1}X_{i2}$, $U = (\sum_{i=1}^n X_{i1}^2, \sum_{i=1}^n X_{i2}^2, \sum_{i=1}^n X_{i1}, \sum_{i=1}^n X_{i2})$, Sample correlation coefficient: $R = \frac{\sum_{i=1}^n (X_{i1} - \bar{X}_1)(X_{i2} - \bar{X}_2)}{\{\sum_{i=1}^n (X_{i1} - \bar{X}_1)^2 \sum_{i=1}^n (X_{i2} - \bar{X}_2)^2\}^{1/2}}$, $T = \sqrt{n-2}R/\sqrt{1-R^2} \sim t_{n-2}$ under $H_0 : \rho = 0$, UMPU test reject $|T| > t_{n-2, \alpha/2}$

[Conditional normal] $x \sim N_x(\mu, \Sigma)$, $x = [x_a, x_b]^T$, $\mu = [\mu_a, \mu_b]^T$, $\Sigma = [[\Sigma_a, \Sigma_c], [\Sigma_c^T, \Sigma_b]]^T$, $p(x_a|x_b) = N_{x_a}(\hat{\mu}_a, \hat{\Sigma}_a)$, $\hat{\mu}_a = \mu_a + \Sigma_c \Sigma_b^{-1}(x_b - \mu_b)$, $\hat{\Sigma}_a = \Sigma_a - \Sigma_c \Sigma_b^{-1} \Sigma_c^T$