Analysis results [Matrix operations] $c^T c = c_1^2 + \cdots + c_k^2$, cc^T is $k \times k$ matrix with (i,j)th element as $c_i c_j$
[Max function] $\max(a,b) = \frac{a+b-|a-b|}{2}$

Probability theory

[positive measure] on measurable space $(\Omega, \mathcal{F})$ $\nu : \mathcal{F} \to \mathcal{R}$ s.t. (1. non-negativity) $0 \le \nu(A) \le \infty$ $\forall$ $A \in \mathcal{F}$ (2. empty is zero) $\nu(\emptyset) = 0$ (3. σ-additivity) $\sum_{i=1}^{\infty} \nu(A_i)$ $\nu(\cup_{i=1}^{\infty})A_i$ if $A_i \in \mathcal{F}$ are disjoint

[measure properties] (1. Monotonicity) $A \subset B \Rightarrow \nu(A) \le \nu(B)$ (2. Sub-additivity) any sequence of potentially non-disjoint set A_n, $\nu(\cup_{n=1}^{\infty} A_n) \le \sum_{n=1}^{\infty} \nu(A_n)$ (3. Continuity of Increasing sequences) $\lim_{n\to\infty} A_n := \cup_{n=1}^{\infty} A_n$ and $\nu(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} \nu(A_n)$ (4. Continuity of Decreasing sequences) $\lim_{n\to\infty} A_n := \cap_{n=1}^{\infty} A_n$, and if $\nu(A_1) < \infty$ then $\nu(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} \nu(A_n)$

[Integration] $f = f_+ - f_-$, $f_+ = \max\{f(x), 0\}$, $f_- = \max\{-f(x), 0\}$ $\int f d\nu := \int f_+ d\nu - \int f_- d\nu$

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

[MCT] if $0 \le f_1 \le f_2 \le \cdots$ and $\lim_n f_n = f$ a.e. then $\int \lim_n f_n d\nu = \lim_n \int f_n d\nu$

[Fatou's lemma] If $f_n \ge 0$ $\int \liminf_n f_n d\nu \le \liminf_n \int f_n d\nu$

[DCT] If $\lim_{n\to\infty} f_n = f$ and $\exists$ integrable function g s.t. $|f_n| \le g$ a.e. $\int \lim_n f_n d\nu = \lim_n \int f_n d\nu$

[Interchange diff and Int] (1) Suppose $\exists$ $(a,b) \subset \mathcal{R}$ which $\partial f(\omega,\theta)/\partial\theta$ exists a.e. (2) There is an integrable function g on ω s.t. $|\partial f(\omega, \theta/\partial\theta| \le g(\omega)$ a.e. $\frac{d}{d\theta} \int f(\omega,\theta) d\nu(\omega) = \int \frac{\partial f(\omega,\theta)}{\partial\theta} d\nu(\omega)$

[Change of Var Formula] $Y = g(X)$, A_i disjoint, h_j is inverse function of g on A_j.

$$f_Y(y) = \sum_{j:1\le j\le m, y\in g(A_j)} \left| \det\left(\frac{\partial h_j(y)}{\partial y}\right) \right| f_X(h_j(y))$$

[Fubini's Theorem] Suppose $f \ge 0$ or $\int |f| d(\nu_1 \times \nu_2) < \infty$ then $g(\omega_2) = \int_{\Omega_1} f(\omega_1, \omega_2) d\nu_1(\omega_1)$
$\int_{\Omega_1\times\Omega_2} f d(\nu_1 \times \nu_2) = \int_{\Omega_1} \left[\int_{\Omega_2} f(\omega_1,\omega_2) d\nu_1(\omega_1)\right] d\nu_2(\omega_2)$

[Absolutely continuity] $\lambda << \nu$ iff for any $A \in \mathcal{F}$, $\nu(A) = 0 \Rightarrow \lambda(A) = 0$

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

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

[Cauchy-Schewarz ineq] $Cov(X,Y)^2 \le Var(X)Var(Y)$ $(EXY)^2 \le EX^2 EY^2$

[Jensen's inequality] A is a convex set in $\mathcal{R}^d$, φ is a convex function on A and $X \in A$ is a d-random vector $\varphi(EX) \le E\varphi(X)$
If φ is strictly convex and $\varphi(X)$ is not a constant, then $\varphi(EX) < E\varphi(X)$
$(EX)^{-1} < E(X^{-1})$ $E(logX) < log(EX)$ $\int f \log\left(\frac{f}{g}\right) d\nu \ge 0$

[Chebyshev's inequality] X is R.V, φ is nonnegative and symmetric function ($\varphi(-x) = \varphi(x)$) and is non-decreasing on $[0, \infty)$, then for each constant $t \ge 0$ $\varphi(t) P(|X| \ge t) \le \int_{\{|X|\ge t\}} \varphi(X) dP \le E\varphi(X)$
Common results $P(|X - \mu| \ge t) \le \frac{\sigma_X^2}{t^2}, P(|X| \ge t) \le \frac{E|X|}{t}$

[Hölder's inequality] suppose $p, q > 0$ are Hölder's conjugate s.t. $1/p + 1/q = 1 \Rightarrow q = p/(p-1)$
$E|XY| \le (E|X|^p)^{1/p} (E|Y|^q)^{1/q}$ If both $E|X|^p$ and $E|Y|^q$ are finite, equality holds if and only if $|X|^p$ and $|Y|^q$ are linearly dependent

[Young's inequality] equality if and only if $a^p = b^q$ $ab \le \frac{a^p}{p} + \frac{b^q}{q}$

[Minkowski's inequality] $p \ge 1$, $(E|X+Y|^p)^{1/p} \le (E|X|^p)^{1/p} + (E|Y|^p)^{1/p}$

[Lyapunov's inequality] for $0 < s < t$, $(E|X|^s)^{1/2} \le (E|X|^t)^{1/t}$

[KL Information] $K(f_0, f_1) = E_0 \log \frac{f_0(X)}{f_1(X)} = \int \log\left(\frac{f_0(x)}{f_1(x)}\right) f_0(x) d\nu(x) \ge 0$ with equality if and only if $f_1(\omega) = f_0(\omega)$ ν-a.e.

[info equality] $K(f_0, f_1) \ge 0$ with equality if and only if $f_1(\omega) = f_0(\omega)$ ν-a.e.

[CHF] $\forall$ $t \in \mathcal{R}^d$ $|\phi_X| \le 1$, $\phi_{-X} = \overline{\phi_X(t)}$ $\phi_X(t) = E\left[exp(\sqrt{-1} t^T X\right] = E\left[\cos(t^T X) + \sqrt{-1}\sin(t^T X)\right]$

[MGF] $\psi_{-X}(t) = \psi_X(-t)$, $\psi_X(t) = E\left[exp(t^T X)\right]$ if ψ is finite in neighborhood of $\mathbf{0} \in \mathcal{R}^d$, then moments of X of any order are finite, and $\phi_X(t) = \psi_X(\sqrt{-1} t)$

[Conditional Exp] Simple function Y, disjoint A_i A_i disjoint and $\cup A_i = \Omega$, $P(A_i) > 0$, $Y = \sum_{i\ge 1} c_i I_{A_i}$
$E(X|Y) = \sum_{i=1}^{\infty} \frac{\int_{A_i} X dP}{P(A_i)} I_{A_i}$

[a.s. convergence]	$X_n \to^{\text{a.s.}} X$ if $P(\lim_{n\to\infty} X_n = X) = 1$. Can show $\forall \epsilon > 0, \sum_{i=1}^{\infty} P(\lvert X_n - X\rvert > \epsilon) < \infty$ via BC lemmas
[Infinity often]	$\{A_n\ i.o.\} = \cap_{n\geq 1} \cup_{j\geq n} A_j := \limsup_{n\to\infty} A_n$
[Borel-Cantelli lemmas]	[First BC] If $\sum_{n=1}^{\infty} P(A_n) < \infty$, then $P(A_n\ i.o.) = 0$ [Second BC] pairwisely independent events $\{A_n\}_{n=1}^{\infty}$, if $\sum_{n=1}^{\infty} P(A_n) = \infty$, then $P(A_n\ i.o.) = 1$
[Convergence in L^p]	A sequence of $\{X_n\}_{n=1}^{\infty}$ of rvs converges to a random variable X in the L^p sense for some $p > 0$ if $E\lvert X\rvert^p < \infty$ and $E\lvert X_n\rvert^p < \infty$ and $\lim_{n\to\infty} E\lvert X_n - X\rvert^p = 0$
[Con in prob]	A sequence $\{X_n\}_{n=1}^{\infty}$ of rvs converges to a rando variable X in probability if for all $\epsilon > 0$ $\lim_{n\to\infty} P(\lvert X_n - X\rvert > \epsilon) = 0$ denoted by $X_n \to^P X$. Can show $E(X_n) = X$, $\lim_{n\to\infty} Var(X_n) = 0$
[Con in dist]	$X_n \to^D X$ or $F_n \Rightarrow F$ if $\lim_{n\to} F_n(x) = F(x)$ for every $x \in \mathcal{R}$ at which F is continuous
[RS between Con]	$L^p \Rightarrow L^q \Rightarrow P$, $a.s. \Rightarrow P$, $P \Rightarrow D$. $X_n \to_D C \Rightarrow X_n \to_P C$. If $X_n \to_P X \Rightarrow \exists$ sub-seq s.t. $X_{n_j} \to_{\text{a.s.}} X$.
[Continuous mapping]	Let $\{X_n\}_{n=1}^{\infty}$ be seq of random $k-$vectors and X is random $k-$vector in the same probability space. Let $g: \mathcal{R}^k \to \mathcal{R}$ be continuous. Then If $X_n \to^* X$, then $g(X_n) \to^* g(X)$, where * is either a.s., P or D.
[Convengence properties]	1. Unique in limit: $X = Y$ if $X_n \to X$ and Y when a.s., P, L^p. If $F_n \Rightarrow F$ and G, then $F(t) = G(t)\ \forall\ t$ 2. Concatenation: $(X_n, Y_n) \to (X, Y)$ when P or a.s., $(X_n, Y_n) \to_D (X, c)$ only for constant. 3. Linearity: $(aX_n + bY_n) \to aX + bY$ when a.s., P, L^p NOT for distribution. 4. Cramér-Wold device: for k-random vectors, $X_n \to_D X \Leftrightarrow c^T X_n \to_D c^T X$ for every $c \in \mathcal{R}^k$
[Lévy continuity]	$\{X_n\}$ converges in dist to X iff corresponding characteristic functions $\{\phi_n\}$ converges pointwise to ϕ_X
[Scheffés theorem]	If $\lim_{n\to\infty} f_n(x) = f(x)$ a.e. ν where $f(x)$ is pdf. Then $\lim_{n\to\infty} \int \lvert f)n(x) - f(x)\rvert d\nu = 0$ and $P_{f_n} \Rightarrow P_f$. Useful for checking convergence in distribution via pdfs.
[Slutsky's theorem]	If $X_n \to^D X$, $Y_n \to^D$ constant c. Then $X_n + Y_n \to^D X + c$, $X_n Y_n \to^D cX$, $X_n/Y_n \to^D X/c$ if $c \neq 0$
[Skorohod's theorem]	If $X_n \to^D X$, then there are some random vectors $Y, Y_1, Y_2, \cdots$ defined on a common probability space such that $P_{Y_n} = P_{X_n}, n = 1, 2, \cdots$, $P_Y = P_X$ and $Y_n \to^{\text{a.s.}} Y$
[δ-method]	$\{a_n\} > 0$, $\lim_{n\to\infty} a_n = \infty$ and $a_n(X_n - c) \to^D Y$, $c \in \mathcal{R}$. If $g'(c)$ exists at c, then $a_n[g(X_n) - g(c)] \to^D g'(c)Y$ If $g^{(j)}(c) = 0$ for all $1 \leq j \leq m-1$ and $g^{(m)}(c) \neq 0$. Then $a_n^m[g(X_n) - g(c)] \to^D \frac{1}{m!} g^{(m)}(c) Y^m$ If X_i, Y are k-vectors rvs and $c \in \mathcal{R}^k$ $a_n[g(X_n) - g(c)] \to_D [\nabla g(c)]^T Y = N\left(0, g(c)^T \Sigma g(c)\right)$ if Y is normal
[Stochastic order]	[real numbers] $\{a_n\}, \{b_n\}$, const c and all n, $a_n = O(b_n) \Leftrightarrow \lvert a_n\rvert \leq c\lvert b_n\rvert$, $a_n = o(b_n) \Leftrightarrow \lim_{n\to\infty} a_n/b_n = 0$ [rvs] $\{X_n\}, \{Y_n\}$, $X_n = O_{\text{a.s.}}(Y_n) \Leftrightarrow P\{\lvert X_n\rvert = O(\lvert Y_n\rvert)\} = 1$, $X_n = o_{\text{a.s.}}(Y_n) \Leftrightarrow X_n/Y_n \to^{\text{a.s.}} 0$, $\forall\ \epsilon > 0, \exists C_\epsilon > 0, n_\epsilon \in \mathcal{N} s.t.\ X_n = O_P(Y_n) \Leftrightarrow \sup_{n\geq n_\epsilon} P\left(\{\omega \in \Omega : \lvert X_n(\omega)\rvert \geq C_\epsilon \lvert Y_n(\omega)\rvert\}\right) < \epsilon$ If $X_n = O_P(1)$, $\{X_n\}$ is bounded in probability. $X_n = o_P(Y_n) \Leftrightarrow X_n/Y_n \to^P 0$
[Properties]	If $X_n \to_{\text{a.s.}} X$, then $\{\sup_{n\geq k} \lvert X_n\rvert\}_k$ is $O_p(1)$. If $X_n \to_D X$ for a rvs, then $X_n = O_P(1)$ (tightness). If $E\lvert X_n\rvert = O(a_n)$, then $X_n = O_P(a_n)$; If $E\lvert X_n\rvert = o(a_n)$, then $X_n = o_P(a_n)$
[SLLN, iid]	If X_i are identical, let $c := EX_1$, $E\lvert X_1\rvert < \infty \Leftrightarrow \frac{1}{n}\sum_{i=1}^n X_1 \to^{\text{a.s.}} c$
[SLLN, non-idential]	If there is a constant $p \in [1, 2]$ s.t. $\sum_{i=1}^{\infty} E\lvert X_i\rvert^p / i^p < \infty$, then $\frac{1}{n}\sum_{i=1}^n (X_i - EX_i) \to^{\text{a.s.}} 0$
[USLLN, idd]	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 $\lvert U(x, \theta) \leq M(x)\rvert$ for all x, θ. Then $P\left\{\lim_{n\to\infty} \sup_{\theta\in\Theta} \left\lvert \frac{1}{n}\sum_{i=1}^n U(X_j, \theta) - \mu(\theta)\right\rvert = 0\right\} = 1$
[WLLN]	If X_i are identical, $\{a_n\}$ exist and take $a_n = E(X_1 I_{\{\lvert X_1\rvert \leq n\}}) \in [-n, n]$ $nP(\lvert X_1\rvert > n) \to 0 \Leftrightarrow \frac{1}{n}\sum_{i=1}^n X_i - a_n \to^{\mathcal{P}} 0$
[WLLN, non-identical]	If there is a constant $p \in [1, 2]$ s.t. $\lim_{n\to\infty} \frac{1}{n^p}\sum_{i=1}^n E\lvert X_i\rvert^p = 0$, then $\frac{1}{n}\sum_{i=1}^n (X_i - EX_i) \to^P 0$
[Weak Convergency]	$\int f d\nu_n \to \int f d\nu$ for every bounded and continous real function f. $X_n \to_D X \Leftrightarrow E[h(X_n)] \to E[h(X)]$
[CLT, iid]	Let $\{X_n\}_{n=1}^{\infty}$ be seq of iid random $k-$vectors. Suppose $\Sigma = Var X_1 < \infty$, then $\frac{1}{\sqrt{n}}\sum_{i=1}^n (X_i - EX_i) \to^D N(0, \Sigma)$
[CLT, non-identical]	X_i independent, suppose (1) $k_n \to \infty$ as $n \to \infty$ (2) $0 < \sigma_n^2 = Var\left(\sum_{j=1}^{k_n} X_{nj}\right) < \infty, n = 1, 2, \cdots$. [Lindeberg's condition] (3) If for any $\epsilon > 0$, $\frac{1}{\sigma_n^2}\sum_{j=1}^{k_n} E\left\{(X_{nj} - EX_{nj})^2 I_{\{\lvert X_{nj} - EX_{nj}\rvert > \epsilon\sigma_n\}}\right\} \to 0$. Then $\frac{1}{\sigma_n}\sum_{j=1}^{k_n} (X_{nj} - EX_{nj}) \to^D N(0, 1)$

[Lindeberg's condition] Check [Lyapunov condition] $\frac{1}{\sigma_n^{2+\delta}}\sum_{j=1}^{k_n} E|X_{nj}-EX_{nj}|^{2+\delta} \to 0$ for some $\delta > 0$
[Uniform boundedness] If $|X_{nj}| \leq M$ for all n and j and $\sigma_n^2 = \sum_{j=1}^{k_n} Var(X_{nj}) \to \infty$
[Feller's condition] In general, Lindeberg's condition is not necessary for convergence result. However, if Feller's condition is met then it is sufficient and necessary. $\lim_{n\to\infty} \max_{j\leq k_n} \frac{Var(X_{nj})}{\sigma_n^2} = 0$

Elements of Stats [Ordered Statistics] $X_(k)$ which is the kth smallest value of $X_1, \cdots, X_n$.
$X_{(n)} = [F(x)]^n, f_{X_{(n)}} = nf(x)[F(x)]^{n-1}$, $X_{(1)} = 1-[1-F(x)]^n, f_{X_{(1)}} = nf(x)[1-F(x)]^{n-1}$
[Empirical variance] $\frac{1}{n}\sum_i (X_i - \bar{X})^2$

[NEF] Since exp fam representation is not unique, consider $\eta = \eta(\theta)$, $f_\eta(\omega) = \exp\left\{\eta^T T(\omega) - \mathcal{C}(\eta)\right\} h(\omega)$, $\mathcal{C}(\eta) = \log\left\{\int_\Omega \exp\left\{\eta^T T(\omega)\right\} h(\omega)d\nu(\omega)\right\}$. η is called natural parameter and natural parameter space $\Xi = \{\eta(\theta) : \theta \in \Theta\} \subset \mathcal{R}^p$. Full rank if Ξ contains open set in $\mathcal{R}^p$

[Joint Exp Fam] Suppose $X_i \sim f_i$ independently with f_i Exp Fam, then joint distribution $X_1, \cdots, X_n$ 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 $\frac{dP_\theta}{d\lambda} > 0$ (assume), then $I_{(t,\infty)}(x) = 0 \Rightarrow v([t,\infty)) = 0$. Since t is arbitary, consider $v(\mathcal{R}) = 0$ (contradiction)

[Separate statistics T] Let $T = (Y, U)$ and $\eta = (\nu, \varphi)$ where Y and ν have same dimension. Then Y has PDF $f_\eta(y) = \exp\left\{\nu^T y - \mathcal{C}(\eta)\right\}$, w.r.t σ-finite measure depending on φ. If T has a PDF in NEF, the conditional distirbution of Y given $U = u$ has PDF (w.r.t σ-finite measure depending on u), $f_{\nu,u}(y) = \exp\left\{\nu^T y - \mathcal{C}_u(\nu)\right\}$, which is in a NEF indexed by ν

[MGF of NEFs] If η_0 is an interior point on natural parameter space, then MGF $\phi_{\eta_0}(t)$ of T (with $P = P_{\eta_0}$ is finite in neighborhood of $t = 0$ and is given by $\psi_{\eta_0}(t) = \exp\{\mathcal{C}(\eta_0 + t) - \mathcal{C}(\eta_0)\}$. Let $A(\theta) = \mathcal{C}(\eta_0(\theta))$, $\frac{dA(\theta)}{d\theta} = \frac{d\mathcal{C}(\eta_0(\theta))}{d\eta_0(\theta)} \cdot \frac{d\eta_0(\theta)}{d\theta}$, $E_{\eta_0}T = \frac{d\psi_{\eta_0}}{dt}|_{t=0} = \frac{d\mathcal{C}}{d\eta_0} = \frac{A'(\theta)}{\eta_0'(\theta)}$, $E_{\eta_0}T^2 = \mathcal{C}''(\eta_0) + \mathcal{C}'(\eta_0)^2$, $Var(T) = \mathcal{C}''(\eta_0) = \frac{A''(\theta)}{[\eta_0(\theta)]^2} - \frac{\eta_0(\theta)'' A'(\theta)}{[\eta_0(\theta)']^3}$

[NEFs Differential id] For a Borel function g, let Ξ_g be set of values of η such that
$\int |g(\omega)| \exp\left\{\eta^T T(\omega) - \mathcal{C}(\eta)\right\} h(\omega)d\nu(\omega) < \infty$
Define G on Ξ_g by
$G(\eta) := \int g(\omega) \exp\left\{\eta^T T(\omega) - \mathcal{C}(\eta)\right\} h(\omega)d\nu(\omega)$
Then for η in interior of Ξ_g
(1) G is continuous and has continuous derivatives of all orders.
(2) These derivatives can be computed by differentiation under the integral sign.
$\frac{dG(\eta)}{d\eta} = E_\eta\left[g(\omega)\left(T(\omega) - \frac{\partial}{\partial\eta}\xi(\eta)\right)\right]$

[Sufficiency] Let X be a sample from an unknown population $P \in \mathcal{P}$. Statistics $T(X)$ is sufficient for $P \in \mathcal{P}$ iff $P_X(x|Y)$ is known and does not depend on P. If $\mathcal{P}$ is parametric family, we can also say $T(X)$ is sufficient for θ. Suppose T is sufficient for $\mathcal{P}_0$, $\mathcal{P}_0 \subset \mathcal{P} \subset \mathcal{P}_1$. Then $T(X)$ is sufficient for $\mathcal{P}_0$ but not ncessarily $\mathcal{P}_1$. $P(X = x|T = t)$ does not depend on θ

[Factorization thm] $T(X)$ is sufficient for $P \in \mathcal{P}$ iff there are non-negative Borel functions
(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] Let T be a sufficient statistics for $P \in \mathcal{P}$. T is called minimal sufficient statistics iff for any other statistics S sufficient for $P \in \mathcal{P}$, there is a measurable function ψ s.t. $T = \psi(S)$ $\mathcal{P}$-a.s.

[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 $f_0, f_1, \cdots$ w.r.t a σ-finite measure.
(1) Define $f_\infty(x) = \sum_{i=0}^\infty c_i f_i(x)$, $T_i(x) = f_i(x)/f_\infty(x)$, then $T(X) = (T_0(X), T_1(X), \cdots)$ 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 $T(X) = (f_1(x)/f_0(x), f_2(x)/f_0(x), \cdots$ is minimal sufficient for $\mathcal{P}$

[Min Suff-Method 2] [Theorem C] Suppose $\mathcal{P}$ contains PDFs f_P w.r.t. σ-finite measure ν. If
(a) $T(X)$ is a sufficient statistics, and
(b) There is a measurable function ϕ s.t. for any possible values x, y of X, or $x, y \in \{x : h(x) > 0\}$ for NEF.

$$f_P(x) = f_P(y)\phi(x,y) \forall P \in \mathcal{P} \Rightarrow T(x) = T(y)$$

Then $T(X)$ is minimal sufficient for $\mathcal{P}$

[min suff for NEF] If there exists $\Theta_0 = \{\theta_0, \theta_1, \cdots, \theta_p\} \subset \Theta$ 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 $det([\eta_1, \cdots, \eta_p])$ is non-zero OR $\Xi = \{\eta(\theta) : \theta \in \Theta\}$ contains $(p+1)$ points that do not lie on the same hyperplane OR Ξ is full rank.

[Completeness] [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] Statistics $T(X)$ is complete for $P \in \mathcal{P}$ iff for any Borel function f, $E_P f(T) = 0$ for all $P \in \mathcal{P}$ implies $f(T) = 0$ $\mathcal{P}$-a.s. T is boundedly complete iff statements holds for bounded Borel functions f.
[Completeness + Sufficiency $\Rightarrow$ Minimal Sufficiency]
Suppose X is a sample from unknown $P \in \mathcal{P}$, and suppose a minmal sufficient statistics exists. If a statistics U is sufficient and boundedly complete, then U is minimal sufficient
[Complete sufficient statistics for NEF]
If $\mathcal{P}$ is NEF of full rank then $T(X)$ is complete and sufficient for $\eta \in \Xi$

[Basu's theorem] Let V and T be two statistics of X from a population $P \in \mathcal{P}$. If V is ancillary and T is boundedly complete and sufficient for $P \in \mathcal{P}$, then V and T are independent w.r.t any $P \in \mathcal{P}$

Evaluation

[Hypothesis tests] Let $\mathcal{P}$ be a family of distributions, $\mathcal{P}_0 \subset \mathcal{P}, \mathcal{P}_1 = \mathcal{P} \backslash \mathcal{P}_0$. 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 $T : \mathcal{X} \to \{0, 1\} \Rightarrow T(X) = I_C(X)$ 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. 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. 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):

$$\sup_{P \subset \mathcal{P}_0} P(T(X) = 1) \leq \alpha$$

[size of test] α' is the size of the test

$$\sup_{P \subset \mathcal{P}_0} P(T(X) = 1) = \alpha'$$

[Comparing decision rules]
[Compare decision rules] T_1 is ... as T_2 if ...: as good as if $R_{T_1}(P) \leq P_{T_2}(P)$. $\forall\ P \in \mathcal{P}$
better if $R_{T_1}(P) < R_{T_2}(P)$ for some $P \in \mathcal{P}$ (and T_2 is dominated by T_1).
equivalent if $R_{T_2}(P) = R_{T_2}(P)$ for all $P \in \mathcal{P}$

[Optimal] Let $\mathcal{J}$ be collection of decision rules in consideration. T_* is $\mathcal{J}$-optional if T_* is as good as any other rule in $\mathcal{J}$, Optimal if T_* is as good as any other possible rule

[Admissibility] Let $\mathcal{J}$ be a class of decision rules. A decision rule $T \in \mathcal{J}$ is called $\mathcal{J}$-admissible if no $S \in \mathcal{J}$ is better than T in terms of the risk.

[Minimaxity] Let $\mathcal{J}$ be a class of decision rules. A decision rule $T_* \in \mathcal{J}$ is called $\mathcal{J}$-minimax if $\sup_{P \subset \mathcal{P}} R_{T_*}(P) \leq \sup_{P \subset \mathcal{P}} R_T(P)$ for any $T \in \mathcal{J}$

[Bayes Risk and Rule] A form of averaging $R_T(P)$ over $P \in \mathcal{P}$. Bayes risk $r_T(\Pi) = \int_{\mathcal{P}} R_T(P) d\Pi(P)$, Π is known probability measure. $R_T(\Pi)$ is Bayes risk of T wrt Π. If $T_* \in \mathcal{J}$, $r_{T_*}(\Pi) \leq r_T(\Pi)$ for any $T \in \mathcal{J}$, then T_* is called $\mathcal{J}$-Bayes rule wrt Π.

[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] Require convex loss $L(P, a)$ and sufficient statistics T for $P \in \mathcal{P}$. Suppose 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 $R_{S_1}(P) \leq R_{S_0}(P)$. If $L(P, a)$ is strictly convex in a, and S_0 is not a funciton of T, then S_0 is inadmissible and dominated by S_1.

Estimators

[MLE for Exp Fam] NEF: $\ell(\eta) = \exp\left[\eta^T T(x) - \mathcal{C}(\eta)\right] h(x)$
$T(x) = \frac{\partial \mathcal{C}(\eta)}{\partial \eta}, Var(T) = \frac{\partial^2 \mathcal{C}(\eta)}{\partial \eta \partial \eta^T}$
General: $\ell(\theta) = \exp\left[\eta(\theta)^T T(x) - \xi(\theta)\right] h(x)$, note $\xi(\theta) = \mathcal{C}(\eta(\theta))$
$\hat{\theta} = \eta^{-1}(\hat{\eta})$, or solution of $\frac{\partial \eta(\theta)}{\partial \theta} T(x) = \frac{\partial \xi(\theta)}{\partial \theta}$

[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 $f(x|\theta) = f(x|\theta_0)$ ν-a.e. $\Rightarrow \theta = \theta_0$. Then for any sequence of maximum likelihood-likelihood estimates $\hat{\theta}_n$ of θ

$$\hat{\theta}_n \to^{\text{a.s}} \theta_0$$

[Unbiased Estimators]

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

[Lehmann-Scheffé] Suppose there exists sufficient and complete statistic $T(X)$ for $P \in \mathcal{P}$, and θ is related to P. If θ is estimable, then there is a unique unbiased estimator of θ that is of the form $h(T)$ with a Borel function h. Furthermore, $h(T)$ is the unique UMVUE of θ.

[UMVUE method1] Using Lehmann-Scheffé, manipulate $E(h(T)) = \theta$ to get $\hat{\theta}$ where T is sufficient and complete. Useful when $E(h(T))$ is easy to solve.

[UMVUE method2] Using Rao-Blackwellization. Find (1) unbiased estimator of $\theta = U(X)$, (2) sufficient and complete statistics $T(X)$, then $E(U|T)$ is the UMVUE of θ by Lehmann-Scheffé. Useful if $E(U|T)$ is easy to solve.

[UMVUE method3] [necessary and sufficient condition] Useful when no complete and sufficient statistics. Can use to find UMVUE, check if estimator is UMVUE, show nonexistence of UMVUE.
Let T is an unbiased estimator of eta with finite variance, $\mathcal{U}$ is set of all unbiased estimators of 0 with finite variances. $T(X)$ is UMVUE $\Leftrightarrow E[T(X)U(X)] = 0$ for any $U \in \mathcal{U}$ and any $P \in \mathcal{P}$.
Suppose $T = h(S)$, where S is sufficient statistics for $P \in \mathcal{P}$ and h is a Borel function. Let $\mathcal{U}_S$ be the subset of $\mathcal{U}$ consisting of Borel functions of S. $T(X)$ is UMVUE $\Leftrightarrow E[T(X)U(X)] = 0$ for any $U \in \mathcal{U}_S$ and any $P \in \mathcal{P}$

[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. $P, P \in \mathcal{P}$

[Fisher information] Suppose fixed support, for any $\theta \in \Theta$, $\frac{\partial f_\theta(x)}{\partial \theta}$ exists and is finite $P_\theta-$a.s., X is a sample from $P_\theta \in \mathcal{P}$.
Amount of information from X is
$I(\theta) = E\left(\frac{\partial}{\partial \theta} \log f_\theta(X)\right)^2 = \int \left(\frac{\partial}{\partial \theta} \log f_\theta(X)\right)^2 f_\theta(X) d\nu(x) = E\left\{\frac{\partial}{\partial \theta} \log f_\theta(X) \left[\frac{\partial}{\partial \theta} \log f_\theta(X)\right]^T\right\}$

[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
$I(\theta) = -E\left[\frac{\partial^2}{\partial \theta^2} \log f_\theta(X)\right] = -E\left[\frac{\partial^2}{\partial \theta \partial \theta^T} \log f_\theta(X)\right]$

[Independent samples] If regularity condition $\int \frac{\partial}{\partial \theta} f_\theta(x) d\nu = 0$ holds, then

$$I_{(X,Y)}(\theta) = I_X(\theta) + I_Y(\theta)$$

[iid samples] If regularity condition holds
$I_{(X_1, \cdots, X_n)}(\theta) = n I_X(X_1)(\theta)$

[Exp fam] For any S with $E[S(X)] < \infty$, it holds that $\frac{\partial}{\partial \theta} \int S(x) f_\theta(x) d\nu = \int S(x) \frac{\partial}{\partial \theta} f_\theta(x) d\nu$ and
$I(\theta) = -E\left[\frac{\partial^2}{\partial \theta \partial \theta^T} \log f_\theta(X)\right]$
If $\underline{I}(\eta)$ is fisher information matrix for natural parameter η, then covariance matrix $Var(T) = \underline{I}(\eta)$.
Let $\psi = E[T(X)]$. Suppose $\bar{I}(\psi)$ is fisher info matrix for parameter ψ, then $Var(T) = [\bar{I}(\psi)]^{-1}$

[Cramér-Rao Lower Bound] Suppose (1) Θ is an open set; P_θ has pdf f_θ (2) f_θ is differentiable and
$0 = \frac{\partial}{\partial \theta} \int f_\theta(x) d\nu = \int \frac{\partial}{\partial \theta} f_\theta(x) d\nu, \theta \in \Theta$.
Suppose $g(\theta)$ is differentiable. $T(X)$ 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
$Var(T(X)) \geq \frac{g'(\theta)^2}{I(\theta)} = \left[\frac{\partial}{\partial \theta} g(\theta)\right]^T [I(\theta)]^{-1} \frac{\partial}{\partial \theta} g(\theta)$
where $I(\theta) > 0$ for any $\theta \in \Theta$

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

[CR LB equality] CR achieve equality iff $T = \left[\frac{g'(\theta)}{I(\theta)}\right] \frac{\partial}{\partial\theta} \log f_\theta(X) + g(\theta)$ a.s. One such example is exp fam.

Asymptotics

[Consistency of point estimators] $X = (X_1, \cdots, X_n)$ is sample from $P \in \mathcal{P}$ and $T_n(X)$ be estimator of θ for P.
[consistent] $\Leftrightarrow T_n(X) \to^P \theta$
[strongly consistent] $\Leftrightarrow T_n(X) \to^{\text{a.s.}} \theta$
[a_n-consistent] $\Leftrightarrow a_n(T_n(X) - \theta) = O_P(1)$, $\{a_n\} > 0$ and diverge to ∞
[L_r-consistent] $T_n(X) \to^{L^P} \theta$ for some fixed $r > 0$
A combination of LLN, CLT, Slustky'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 $c_{ni} = c_i/n$ satisfy (1) $\frac{1}{n}\sum_{i=1}^n c_i \to 1$ 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 $\Leftrightarrow \sum_{i=1}^n c_{ni} \to 1$ and $\sum_{i=1}^n c_{ni}^2 \to 0$

[Asympotics bias, variance, MSE] [Approximate unbiased] Estimator $T_n(X)$ for θ is approximately unbiased if $b_{T_n}(P) \to 0$ as $n \to \infty$, $b_{T_n}(P) := ET_n(X) - \theta$

....

When estimator's expectations or second moment are not well defined, we need asymptotic behaviours.
[Asymptotic statistics conditions] $\{a_n\} > 0$ and either (a) $a_n \to \infty$ or (b) $a_n \to a > 0$. If

$$a_n(T_n - \theta) \to^D Y$$

[Asymptotic expectation] If $a_n \xi_n \to^D \xi$, $E|\xi| < \infty$, then asymptotic expectation of ξ_n is $E\xi/a_n$
[Asymptotic bias] $\tilde{b}_{T_n} = EY/a_n$, asymptotically unbiased if $\lim_{n\to\infty} \tilde{b}_{T_n}(P) = 0$ for any $P \in \mathcal{P}$.
[Asymptotic MSE] amse is the asymptotic expectation of $(T_n - \theta)^2$ or $\text{amse}_{T_n}(P) = EY^2/a_n^2$
[Asymptotic Variance] $\sigma^2_{T_n}(P) = Var(Y)/a_n^2$
[Remark] $EY^2 \leq \liminf_{n\to\infty} E[a_n^2(T_n - v)^2]$ (amse is no greater than exact mse)

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

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

[Properties of MOM] θ_n is unique if h^{-1} 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 by CLT and δ-method. V_μ is $k \times k$ with $(i,j) = \mu_{i+j} - \mu_i \mu_j$

$$\sqrt{n}(\hat{\theta}_n - \theta) \to_D N(0, [\nabla g]^T V_\mu \nabla g)$$

MOM is $\sqrt{n}$-consistent, and if $k = 1$ $amse_{\hat{\theta}_n}(\theta) = g'(\mu_1)^2 \sigma^2/n$, $\sigma^2 = \mu_2 - \mu_1^2$

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

[Asym sample quantiles] $X_1, X_2, \cdots$ iid rvs with CDF F, $\gamma \in (0,1)$, $\hat{\theta}_n := \lfloor \gamma n \rfloor$-th order statistics. Suppose $F(\theta) = \gamma$ and $F'(\theta) > 0$ and exists.

$$\sqrt{n}(\hat{\theta}_n - \theta) \to^D N\left(0, \frac{\gamma(1-\gamma)}{[F'(\theta)]^2}\right)$$

[Cons and Asym eff MLEs, RLEs]
[Continuous in θ] 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 and θ (4) identifiable $f(x|\theta) = f(x|\theta_0)$ ν-a.e. $\Rightarrow \theta = \theta_0$. Then for any sequence of MLE $\hat{\theta}_n \to_{\text{a.s.}} \theta_0$

[Upper semi-continuous (usc)]

$$\lim_{\rho\to 0} \left\{ \sup_{||\theta' - \theta|| < \rho} f(x|\theta') \right\} = f(x|\theta)$$

[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 $d(\hat{\theta}_n, \theta_0) \to_{\text{a.s.}} 0$
[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. $\sup_{\theta\in\Theta}\|S_n(\theta) - S(\theta)\| \to_P 0$ 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
[RLE]	[Roots of the Likelihood Equation] θ that solves $\frac{\partial}{\partial\theta}\log L_n(\theta) = 0$
[Basic 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
[Consistency of RLEs]	Under basic regularity conditions, there exists a sequence of $\hat{\theta}_n$ s.t. $\frac{\partial}{\partial\theta}\log L_n(\hat{\theta}_n) = 0$ and $\hat{\theta}_n \to_{\text{a.s.}} \theta_*$. More useful if likelihood is concave or unique.
[Asymptotic Normality of RLEs]	Assume basic 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 $\theta = \theta_*$. Then any consistent sequence $\{\tilde{\theta_n}\}$ of RLE it holds $$\sqrt{n}(\tilde{\theta_n} - \theta_*) \to_D N\left(0, \frac{1}{I(\theta_*)}\right)$$
[NEF RLEs]	Basic 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 $\hat{\eta}_n \to_{\text{a.s.}} \eta$ $$\sqrt{n}(\hat{\eta}_n - \eta) \to_D N\left(0, \left[\frac{\partial^2}{\partial\eta\partial\eta^T}\mathcal{C}(\eta)\right]^{-1}\right)$$ Where $g(\eta) = \frac{\partial\mathcal{C}(\eta)}{\partial\eta}$ and $\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^n T(X_i)$
[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. $$[V_n(\theta)]^{-1/2}(\hat{\theta}_n - \theta) \to_D N_k(0, I_k)$$
[Information Inequalities]	AB 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 thant $\hat{\theta}_{2n}$ if (1) $V_{1n}(\theta)V_{2n}(\theta)$ 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.
[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 $\theta \neq 0$: $\sqrt{n}(\hat{\theta} - \theta) = \sqrt{n}(\bar{X}_n - \theta) - (1-t)\sqrt{n}\bar{X}_n I_{\|\bar{\theta}_n\|<n^{-1/4}}$ 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 $V_n(\theta) = [I_n(\theta)]^{-1}$.
[One-step MLE]	Often asym efficient, useful to adjust an non asym efficient estimators provided $\hat{\theta}_n^{(0)}$ is $\sqrt{n}$-consistent. $$\hat{\theta}_n^{(1)} = \hat{\theta}_n^{(0)} - \left[\nabla s_n(\hat{\theta}_n^{(0)})\right]^{-1} s_n(\hat{\theta}_n^{(0)})$$