Math Prob

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

[Matrix] [Ops] $x^Tx \in \mathcal{R}$, $xx^T \in \mathcal{R}_{k\times k}$ $g(b)=\|X-Zb\|=(X-Zb)^T(X-Zb)$, $\frac{\partial}{\partial b}(b^TAb)=2Ab$, $\frac{\partial}{\partial b}(b^TAc)=Ac$,
[Generalised inverse] A^- is inverse of A if $AA^-A=A$ [Moore-Penrose inverse] $A^-AA^-=A^-$
[Projection] $P_z=Z(Z^TZ)^-Z^T$, $P_z^2=P_z$, $P_zZ=Z$, $rank(P_z)=r$

[Sets] [Mapping] (a) $f\left(\cup_{\alpha\in I}A_\alpha\right)=\cup_{\alpha\in I}f(A_\alpha)$ (b) $f^{-1}\left(\cup_{\alpha\in I}B_\alpha\right)=\cup_{\alpha\in I}f^{-1}(B_\alpha)$ (c) $f^{-1}(B^c)=(f^{-1}(B))^c$
[Ops] $\cup_i(C\cap A_i)=C\cap(\cup_iA_i)$, $f(\omega)=\int I_{[0,f(\omega)]}t\,dm(t)$

[Expectation] [Linearity] $E(aX+bY)=aEX+bEY$ [Abs finite] EX is finite $\Leftrightarrow E|X|$ is finite [Order] $EX\le EY$, equality if $X=Y$ a.s. [Deduce $X=0$] If $X\ge 0$ a.s. and $EX=0$ then $X=0$ a.s.

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

[Inequalities] [Cauchy-Schewarz] $Cov(X,Y)^2\le Var(X)Var(Y)$, $(EXY)^2\le EX^2EY^2$
[Jensen] A is convex set, φ is convex function $\varphi(EX)\le E\varphi(X)$, $(EX)^{-1}<E(X^{-1})$, $E(logX)<log(EX)$
[Chebyshev] $\varphi>0$, $\varphi(-x)=\varphi(x)$ and is non-decreasing on $[0,\infty)\Rightarrow\varphi(t)P(|X|\ge t)\le\int_{\{|X|\ge t\}}\varphi(X)dP\le E\varphi(X)$ E.g. $P(|X-\mu|\ge t)\le\frac{\sigma_X^2}{t^2}$, $P(|X|\ge t)\le\frac{E|X|}{t}$
[Hölder] $p,q>0$, $1/p+1/q=1\Rightarrow q=p/(p-1)$ $E|XY|\le(E|X|^p)^{1/p}(E|Y|^q)^{1/q}$, equality $\Leftrightarrow |X|^p=c|Y|^q$
[Young] $ab\le\frac{a^p}{p}+\frac{b^q}{q}$ equality $\Leftrightarrow a^p=b^q$
[Minkowski] $p\ge 1$, $(E|X+Y|^p)^{1/p}\le(E|X|^p)^{1/p}+(E|Y|^p)^{1/p}$
[Lyapunov] for $0<s<t$, $(E|X|^s)^{1/2}\le(E|X|^t)^{1/t}$
[Kullback-Leibler] $K(f_0,f_1)=E_0\log\frac{f_0(X)}{f_1(X)}=\int\log\left(\frac{f_0(x)}{f_1(x)}\right)f_0(x)d\nu(x)\ge 0$ with equality $\Leftrightarrow f_1(\omega)=f_0(\omega)$ ν-a.e.

[Char, MGF] $\forall\, t\in\mathcal{R}^d$ [Char func] $|\phi_X|\le 1$, $\phi_{-X}=\overline{\phi_X(t)}$ $\phi_X(t)=E\left[exp(\sqrt{-1}t^TX\right]=E\left[\cos(t^TX)+\sqrt{-1}\sin(t^TX)\right]$
[MGF] $\psi_{-X}(t)=\psi_X(-t)=E\left[exp(t^TX)\right]$, if ψ finite near $\mathbf{0}\in\mathcal{R}^d$, then $E[X^p]$ finite for all p, $\phi_X(t)=\psi_X(\sqrt{-1}t)$

[Convergence] [a.s.] $P\left(\lim_{n\to\infty}X_n=X\right)=1$, use B.C. show $\forall\,\epsilon>0,\sum_{i=1}^\infty P(|X_n-X|>\epsilon)<\infty$
[L^p] if $E|X|^p<\infty$ and $E|X_n|^p<\infty$ and $\lim_{n\to\infty}E|X_n-X|^p=0$
[Prob] for all $\epsilon>0$, $\lim_{n\to\infty}P(|X_n-X|>\epsilon)=0$, can be showned by $E(X_n)=X$, $\lim_{n\to\infty}Var(X_n)=0$
[dist (weak convergence)] $\lim_{n\to}F_n(x)=F(x)$
[RS] $L^p\Rightarrow L^q\Rightarrow P$, $a.s.\Rightarrow P$, $P\Rightarrow D$. If $X_n\xrightarrow{D}C$, then $X_n\xrightarrow{P}C$. If $X_n\xrightarrow{P}X$, $\exists$ sub-sequence s.t. $X_{n_j}\xrightarrow{\text{a.s.}}X$.
[Continuous mapping] If $g:\mathcal{R}^k\to\mathcal{R}$ be continuous. If $X_n\xrightarrow{*}X$, then $g(X_n)\xrightarrow{*}g(X)$ [a.s., P or D]
[Lévy continuity] $\{X_n\}$ converges in distribution to $X\Leftrightarrow$ char $\{\phi_n\}$ converges pointwise to ϕ_X
[Scheffés] (check pdf converge) If $\lim_{n\to\infty}f_n(x)=f(x)$, then $\lim_{n\to\infty}\int|f_n(x)-f(x)|d\nu=0$ and $P_{f_n}\Rightarrow P_f$
[Slutsky's theorem] If $X_n\xrightarrow{D}X$, $Y_n\xrightarrow{D}c$ (constant). Then $X_n+Y_n\xrightarrow{D}X+c$, $X_nY_n\xrightarrow{D}cX$, $X_n/Y_n\xrightarrow{D}X/c$ $(c\ne 0)$
[Skorohod's theorem] If $X_n\xrightarrow{D}X$, then $\exists Y,Y_1,Y_2,\cdots$ on common probability space s.t. $P_{Y_n}=P_{X_n},n=1,2,\cdots$, $P_Y=P_X$ and $Y_n\to^{\text{a.s.}}Y$

[δ-method] If $\{a_n\}>0$, $\lim_{n\to\infty}a_n=\infty$, c is constant, Then $a_n(X_n-c)\xrightarrow{D}Y\Rightarrow a_n[g(X_n)-g(c)]\xrightarrow{D}g'(c)Y$
If $g^{(j)}(c)=0$ for all $1\le j\le m-1$ and $g^{(m)}(c)\ne 0$. Then $a_n^m[g(X_n)-g(c)]\xrightarrow{D}\frac{1}{m!}g^{(m)}(c)Y^m$
If X_i,Y are k-vectors rvs and $c\in\mathcal{R}^k$ $a_n[g(X_n)-g(c)]\xrightarrow{D}[\nabla g(c)]^TY=N\left(0,g(c)^T\Sigma g(c)\right)$ *only if Y is normal*

[SLLN] [X_i are identical] $E|X_i|<\infty\Leftrightarrow\frac{1}{n}\sum_{i=1}^nX_i\xrightarrow{\text{a.s.}}E(X_i)$
[non-identical] If $\exists$ constant $p\in[1,2]$ s.t. $\sum_{i=1}^\infty E|X_i|^p/i^p<\infty$, then $\frac{1}{n}\sum_{i=1}^n(X_i-EX_i)\to^{\text{a.s.}}0$
[Uniform SLLN, iid samples] If (1) $U(x,\theta)$ continuous in θ for fixed x (2) $\mu(\theta)=E[U(X,\theta)]$ is finite (3) Θ compact (4) $\exists M(x)$ s.t. $EM(X)<\infty$, $|U(x,\theta)\le M(x)|\ \forall\,x,\theta$. Then $P\left\{\lim_{n\to\infty}\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{i=1}^nU(X_j,\theta)-\mu(\theta)\right|=0\right\}=1$

[Weak LLN] [X_i identical] If $\{a_n\}$ exist, $a_n:=E\left(X_1I_{\{|X_1|\le n\}}\right)\in[-n,n]$, then $\lim_{n\to\infty}nP\left(|X_1|>n\right)\to 0\Leftrightarrow\frac{1}{n}\sum_{i=1}^nX_i-a_n\xrightarrow{\mathcal{P}}0$
[WLLN, non-identical] If $\exists$ constant $p\in[1,2]$ s.t. $\lim_{n\to\infty}\frac{1}{n^p}\sum_{i=1}^nE|X_i|^p=0$, then $\frac{1}{n}\sum_{i=1}^n(X_i-EX_i)\to^P 0$

[CLT] [iid] $\{X_n\}_{n=1}^\infty$ iid. If $\Sigma=VarX_1<\infty$, then $\frac{1}{\sqrt{n}}\sum_{i=1}^n(X_i-EX_i)\to^D N(0,\Sigma)$
[Lindeberg's non-identical] $\{X_{nj},j=1,\cdots,k_n\}$ independent, if (1) $k_n\to\infty$ as $n\to\infty$ (2, Lin cond) $0<\sigma_n^2=Var\left(\sum_{j=1}^{k_n}X_{nj}\right)<\infty,n=1,2,\cdots$. (3) for any $\epsilon>0$, $\frac{1}{\sigma_n^2}\sum_{j=1}^{k_n}E\left\{(X_{nj}-EX_{nj})^2I_{\{|X_{nj}-EX_{nj}|>\epsilon\sigma_n\}}\right\}\to 0$.
Then $\frac{1}{\sigma_n}\sum_{j=1}^{k_n}(X_{nj}-EX_{nj})\to^D N(0,1)$
[Checking Lindeberg's condition] (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 $|X_{nj}|\le M$ for all n and j and $\sigma_n^2=\sum_{j=1}^{k_n}Var(X_{nj})\to\infty$
[Feller's condition] If $\lim_{n\to\infty}\max_{j\le k_n}\frac{Var(X_{nj})}{\sigma_n^2}=0$, Lindeberg $\Leftrightarrow$ convergence
[Berry-Esseen] There $\exists$ constant C s.t. $\sup_{y\subset\mathcal{R}}|F_n(y)-\Phi(y)|\le\frac{C\rho}{\sigma^3\sqrt{n}}$

Stat Est

[Common Terms] [Identifiable] Parametric family is identifiable iff $\forall\,\theta_1,\theta_2\in\Theta$ and $\theta_1\ne\theta_2\Rightarrow P_{\theta_1}\ne P_{\theta_2}$
[Sample Variance] $S^2=\frac{1}{n-1}\sum_i(X_i-\bar{X})^2$ [Empirical variance] $\frac{1}{n}\sum_i(X_i-\bar{X})^2$
[Ordered Statistics] $X_{(n)}=[F(x)]^n$, $f_{X_{(n)}}=nf(x)[F(x)]^{n-1}$, $X_{(1)}=1-[1-F(x)]^n$, $f_{X_{(1)}}=nf(x)[1-F(x)]^{n-1}$
[Empirical dist] $P_n(A)=\frac{1}{n}\#\{i:X_i\in A\}$, $\forall\,A\in B$ [Empirical CDF] F_n $F_n(t)=\frac{1}{n}\sum_{i=1}^n1_{(-\infty,t)}(X_i),t\in\mathcal{R}$

[Exp Fam]

$f_\theta(x) = \exp\left[\eta(\theta)Y(x) - \xi(\theta)\right] h(x) = \exp\left\{\eta^T T - \mathcal{C}(\eta)\right\}$ $A(\theta) := \mathcal{C}(\eta_0(\theta))$, $\frac{dA(\theta)}{d\theta} = \frac{d\mathcal{C}(\eta_0(\theta))}{d\eta_0(\theta)} \cdot \frac{d\eta_0(\theta)}{d\theta}$,
[MGF] $\psi_{\eta_0}(t) = \exp\left\{\mathcal{C}(\eta_0 + t) - \mathcal{C}(\eta_0)\right\}$ $E_{\eta_0}T = \frac{d\psi_{\eta_0}}{dt}|_{t=0} = \frac{d\mathcal{C}}{d\eta_0} = \frac{A'(\theta)}{\eta_0'(\theta)}$, $E_{\eta_0}T^2 = \mathcal{C}''(\eta_0) + \mathcal{C}'(\eta_0)^2$,
$Var(T) = \mathcal{C}''(\eta_0) = \frac{A''(\theta)}{[\eta_0(\theta)]^2} - \frac{\eta_0(\theta)''A'(\theta)}{[\eta_0(\theta)']^3}$
[Differnetial] $G(\eta) := \int g(\omega)\exp\left\{\eta^T T(\omega) - \mathcal{C}(\eta)\right\} h(\omega) d\nu(\omega)$ $\frac{dG(\eta)}{d\eta} = E_\eta\left[g(\omega)\left(T(\omega) - \frac{\partial}{\partial\eta}\xi(\eta)\right)\right]$

[Min Suff]

[Method 1: A + B] [A] Suppose $\mathcal{P}_0 \subset \mathcal{P}$, $\mathcal{P}_0$-a.s. $\Rightarrow \mathcal{P}$-a.s. If T is suff for $P \in \mathcal{P}$ and min suff for $P \in \mathcal{P}_0$, then T is min suff for $P \in \mathcal{P}$ [B] Suppose $\mathcal{P}$ contains PDFs $f_0, f_1, \cdots$ w.r.t a σ-finite measure.
(B1 $T(X) = \{T_i(X)\}$). $c_i > 0$, $\sum_{i=0}^\infty c_i = 1$ $f_\infty(x) := \sum_{i=0}^\infty c_i f_i(x) > 0$, $T_i(x) = f_i(x)/f_\infty(x)$,
(B2 $T(X) = \{f_i(x)/f_0(x)\}$) If $\{x : f_i(x) > 0\} \subset \{x : f_0(x) > 0\} \forall i$
[Method 2: C] Suppose $\mathcal{P}$ contains PDFs f_P w.r.t. σ-finite measure ν. If (a) $T(X)$ is a suff stat, and (b) $\exists\phi$ s.t. $f_P(x) = f_P(y)\phi(x,y) \forall P \in \mathcal{P} \Rightarrow T(x) = T(y)$ for any possible values x, y of X (NEF: $x, y \in \{x : h(x) > 0\}$)
[NEF special] T is min suff if $\eta_i = \eta(\theta_i) - \eta(\theta_0)$ are linearly independent. (a) Check $det([\eta_1, \cdots, \eta_p])$ is non-zero (b) $\Xi = \{\eta(\theta) : \theta \in \Theta\}$ contains $(p+1)$ points that do not lie on the same hyperplane (c) Ξ is full rank.

[Completeness, Basu]

$T(X)$ is complete for $P \in \mathcal{P} \Leftrightarrow$ for any Borel function f, $E_P f(T) = 0 \Rightarrow f(T) = 0$ (bounded if f is bounded)
[Bounded Complete + Suff $\Rightarrow$ Min Suff] [NEF] full rank $\Rightarrow T(X)$ 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 $P \in \mathcal{P}$

[Decision rule, Rao Blackwell]

[Risk] $R_T(P) = E[L(P, T(X))]$ [Optimal] $R_T(P) \le R_{T_*}(P) \forall P$ [Admissibility] $\nexists R_T(P) > R_{T_*}(P)$ for any P
[MiniMax] $\sup_{P \subset \mathcal{P}} R_{T_*}(P) \le \sup_{P \subset \mathcal{P}} R_T(P)$ [Bayes] $r_T(\Pi) = \int_{\mathcal{P}} R_T(P) d\Pi(P)$
[MSE] $E(||\hat\theta - \theta||^2) = E||\hat\theta - E(\hat\theta)||^2 + (E\hat\theta - \theta)^2 = Var(\hat\theta) + \text{Bias}^2$
[Rao-Blackwell] Given convex $L(P, a)$ in a, suff stat T, $S_1 = E[S_0(X)|T]$, then $R_{S_1}(P) \le R_{S_0}(P)$. If $L(P, a)$ strictly convex, S_0 not a fn of T, then S_0 is inadmissible and dominated by S_1.

[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 $|\log f(x|\theta) - \log f(x|\theta_0)| \le M(x)$ 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 $\hat\theta_n \to^{\text{a.s}} \theta_0$

[MLE regularity]

(1) Θ is open set (2) $f_\theta(x)$ is twice continuously differentiable in θ (3) integration is interchangeable $\frac{\partial}{\partial\theta}\int = \int\frac{\partial}{\partial\theta}$ (4) Fisher information is positive definite

[UMVUE]

Hypo testing
NP Test

[Steps] (1) Find joint distribution $f(X_1, \cdots, X_n)$ - MLR/NEF (2) Hypothesis H_0, H_1 - simple/composite, must be θ and not $f(\theta)$ (3) Form N-P test structure T_* (4) Find test dist, rejection/acceptance region.
[Type I error] reject H_0 when H_0 is correct. $\beta_T(\theta_0) = E_{H_0}(T) \le \alpha$ (within controlled with size α)
[Type II error] do not reject H_0 when H_1 is correct. $1 - \beta_T(\theta)$ for $\theta \in \Theta_1$
[N-P lemma] NP test has non-trival power $\alpha < \beta_{H_1}(T)$ unless $P_0 = P_1$, and is unique up to γ (randomised test)
[Show T_* is UMP] UMP when $E_1[T_*] - E_1[T] \ge 0$, *key equation*: $(T_* - T)(f_1 - cf_0) \ge 0$.
$\Rightarrow \int(T_* - T)(f_1 - cf_0) = \beta_{H_1}(T_*) - \beta_{H_1}(T) \ge 0$.
[Composite hypothesis] Simple $\Rightarrow$ Composite when $\beta_T(\theta_0) \ge \beta_T(\theta \in H_0)$
and/or $\beta_T(\theta_0) \le \beta_T(\theta \in H_1)$ (or does not depend on θ. For MLR this is satisfied, others need to check.

[Monoton Likelihood]

$\theta_2 > \theta_2$, increasing likelihood ratio in Y if $g(Y) = \frac{f_{\theta_2}(Y)}{f_{\theta_1}(Y)} > 1$ or $g'(Y) > 0$. For NEF, check $\eta'(\theta) > 0$.

[UMP]

(1) $H_0 : P = p_0$ $H_1 : P = p_1 \Rightarrow T(X) = I(p_1(X) > cp_0(X))$, $\beta_T(p_0) = \alpha$
(2) $H_0 : \theta \le \theta_0$ $H_1 : \theta > \theta_0 \Rightarrow T(Y) = I(Y > c)$, $\beta_T(\theta_0) = \alpha$
(3) $H_0 : \theta \le \theta_1$ or $\theta \ge \theta_2$ $H_1 : \theta_1 < \theta < \theta_2$, $\Rightarrow T(Y) = I(c_1 < Y < c_2)$, $\beta_T(\theta_1) = \beta_T(\theta_2) = \alpha$
[No UMP] $H_0 : \theta = \theta_1, H_1 : \theta \ne \theta_1$ and $H_0 : \theta \in (\theta_1, \theta_2)$ $H_1 : \theta \notin (\theta_1, \theta_2)$

[UMP Exp fam]

[$\eta(\theta)$ increasing, $H_0 : \theta \le \theta_0$] [$\eta(\theta)$ decreasing, $H_0 : \theta \ge \theta_0$] Same UMP $T(Y) = I(Y < c)$
[$\eta(\theta)$ increasing, $H_0 : \theta \ge \theta_0$] [$\eta(\theta)$ decreasing, $H_0 : \theta \le \theta_0$] Reverse inequalities $T(Y) = I(Y > c)$

[Normal results]

$X_i \sim N(\mu, \sigma^2)$, under $H_0 : \sigma^2 = \sigma_0^2$, note $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar X)^2$ independent to $\bar X$
$V = \frac{1}{\sigma_0^2}\sum_{i=1}^n (X_i - \bar X)^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$
$t = \frac{\sqrt{n}(\bar X - \mu)/\sigma}{\sqrt{V/(n-1)}} = \frac{Z}{\sqrt{V/(n-1)}} \sim t_{(n-1)}$ [(only if $X_i \sim N$]

[Simultaneous]

[Bonferroni] adjust each paramter level to $\alpha_t = \alpha/k$ [Bootstrap] Monte Carlo percentile estimate

[UMPU NEF $\eta(\theta) = \theta$]

Require: (1) suff stat Y for θ (2) suff and complete U for φ (2a) U complete when φ to be full-rank
(1) $H_0 : \theta \le \theta_1$ or $\theta \ge \theta_2$ $H_1 : \theta_1 < \theta < \theta_2$
$\Rightarrow T(Y, U) = I(c_1(U) < Y < c_2(U))$, $E_{\theta_1}[T(Y, U)|U = u] = E_{\theta_2}[T(Y, U)|U = u] = \alpha$
(2) $H_0 : \theta_1 \le \theta \le \theta_2$ $H_1 : \theta < \theta_1$ or $\theta > \theta_2$
$\Rightarrow T(Y, U) = I(Y < c_1(U)$ or $Y > c_2(U))$, $E_{\theta_1}[T(Y, U)|U = u] = E_{\theta_2}[T(Y, U)|U = u] = \alpha$
(3) $H_0 : \theta = \theta_0$ $H_1 : \theta \ne \theta_0 \Rightarrow T(Y, U) = I(Y < c_1(U)$ or $Y > c_2(U))$,
$E_{\theta_0}[T_*(Y, U)|U = u] = \alpha$ and $E_{\theta_0}[T_*(Y, U)Y|U = u] = \alpha E_{\theta_0}(Y|U = u)$
(4) $H_0 : \theta \le \theta_0$ $H_1 : \theta > \theta_0 \Rightarrow T(Y, U) = I(Y > c(U))$, $E_{\theta_0}[T(Y, U)|U = u] = \alpha$

[UMPU Normal]

Assume $V(Y,U)$ independent of U under H_0
(1) $H_0: \theta \leq \theta_1$ or $\theta \geq \theta_2$ $H_1: \theta_1 < \theta < \theta_2$ Require V to be increasing in Y.
$\Rightarrow T(V) = I(c_1 < V < c_2)$, $E_{\theta_1}[T(V)] = E_{\theta_2}[T(V)] = \alpha$
(2) $H_0: \theta_1 \leq \theta \leq \theta_2$ $H_1: \theta < \theta_1$ or $\theta > \theta_2$ Require V to be increasing in Y.
$\Rightarrow T(V) = I(V < c_1$ or $V > c_2)$, $E_{\theta_1}[T(V)] = E_{\theta_2}[T(V)] = \alpha$
(3) $H_0: \theta = \theta_0$ $H_1: \theta \neq \theta_0$ Require $V(Y,U) = a(u)Y + bU$
$\Rightarrow T(V) = I(V < c_1$ or $V > c_2)$, $E_{\theta_0}[T(V)] = \alpha$, $E_{\theta_0}[T(V)V] = \alpha E_{\theta_0}(V)$
(4) $H_0: \theta \leq \theta_0$ $H_1: \theta > \theta_0$ Require V to be increasing in Y. $\Rightarrow T(V) = I(V > c)$, $E_{\theta_0}[T(V)] = \alpha$

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.*

Asym test

Assume MLE regularity condition, under H_0, $-2\log\lambda(X) \to \chi^2_r$, where $r := dim(\theta)$
$T(X) = I\left[\lambda(X) < \exp(-\chi^2_{r,1-\alpha}/2)\right]$ where $\chi^2_{r,1-\alpha}$ is the $(1-\alpha)$th quantile of χ^2_r.

[Asymptotic Tests]

$H_0: R(\theta) = 0$, $\lim_{n\to\infty} W_n, Q_n \sim \chi^2_r$, $T(X) = I(W_n > \chi^2_{r,1-\alpha})$ or $I(Q_n > \chi^2_{r,1-\alpha})$
[Wald's test] $W_n = R(\hat\theta)^T\{C(\hat\theta)^T I_n^{-1}(\hat\theta)C(\hat\theta)\}^{-1}R(\hat\theta)$
$C(\theta) = \partial R(\theta)/\partial\theta$, $I_n(\theta)$ is fisher info for $X_1,\cdots,X_n$, $\hat\theta$ is unrestricted MLE/RLE of θ.
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] $Q_n = s_n(\tilde\theta)^T I_n^{-1}(\tilde\theta)s_n(\tilde\theta)$.
$s_n(\theta) = \partial\log\ell(\theta)/\partial\theta$ is score function, $\tilde\theta$ is MLE/RLE of θ under $H_0: R(\theta) = 0$ (under H_0).

Non-param tests

[Sign test]

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

[Permutation test]

$X_{i1},\cdots,X_{in_i} \sim^{iid} F_i$, $i = 1,2$ $H_0: F_1 = F_2$ $H_1: F_1 \neq F_2$, $\Rightarrow T(X)$ 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]

$X_i \sim^{iid} F$, $Rank(X_i) = \#\{X_j : X_j \leq X_i\}$, $H_0: F$ symm ard 0, $H_1: H_0$ false, R^o_+ vector of ordered R_+.
[Wilcoxon] $T(X) = I[W(R^o_+) < c_1$ or $W(R^o_+ > c_2)]$, $W(R^o_+) = J(R^o_{+1}/n) + \cdots + J(R^o_{+n_*}/n)$
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 $T(X) = I(C_n(F_0) > c)$, $C_n(F) = \int\{F_n(x) - F(x)\}^2 dF(x)$
$nC_n(F_0) \xrightarrow{D} \sum_{j=1}^\infty \lambda_j\chi^2_{1j}$, with $\chi^2_{1j} \sim \chi^2_1$ and $\lambda_j = j^{-2}\pi^{-2}$

[Empirical LR]

$X_i \sim^{iid} F$, $H_0: \Lambda(F) = t_0$ $H_1: \Lambda(F) \neq t_0$, $\Rightarrow T(X) = I(ELR_n(X) < c)$
$ELR_n(X) = \frac{\ell(\hat F_0)}{\ell(\hat F)}$, $\ell(G) = \prod_{i=1}^n P_G(\{x_i\})$, $G \in \mathcal{F}$. ($\mathcal{F}$:= collection of CDFs, P_G := measure induced by CDF G)

Confidence set

$C(X): X \to \mathcal{B}(\Theta)$, Require $\inf_{P\in\mathcal{P}} P(\theta \in C(X)) \geq 1-\alpha$. Conf coeff more than level
[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 $A(\theta) = \{x: T_{\theta_0}(x) \neq 1\}$. $H_0: \theta = \theta_0$, H_1 any

[Shortest CI]

[unimodal] $f'(x_0) = 0$ $f'(x) < 0, x < x_0$ and $f'(X) > 0, x > x_0$
[Pivotal $(T-\theta)/U$, f unimodal at x_0] $[T - b_*U, T - a_*U]$, shortest when $f(a_*) = f(b_*) > 0$ $a_* \leq x_0 \leq b_*$
[Pivotal T/θ, $x^2 f(x)$ unimodal at x_0] $[b_*^{-1}T, a_*^{-1}T_*]$ shortest when $a_*^2 f(a_*) = b_*^2 f(b_*) > 0$ $a_* \leq x_0 \leq b_*$
[General] Suppose $f > 0$, integrable, unimodal at x_0, want: $\min b - a$ s.t. $\int_a^b f(x)dx$ and $a \leq b$
sol: a_*, b_* satisfy (1) $a_* \leq x_0 \leq b_*$ (2) $f(a_*) = f(b_*) > 0$ (3) $\int_{a_*}^{b_*} f(x)dx = 1-\alpha$

[asym]

require $\lim_{n\to} P(\theta \in C(X)) \geq 1-\alpha$,
[asym pivotal] $\mathcal{R}_n(X,\theta) = \hat V_n^{-1/2}(\hat\theta_n - \theta)$ does not depend on P in limit
[LR] $C(X) = \left\{\theta: \ell(\theta,\hat\varphi) \geq exp(-\chi^2_{r,1-\alpha}\alpha/2)\ell(\hat\theta)\right\}$
[Wald] $C(X) = \left\{\theta: (\hat\theta - \theta)^T\left[C^T\left(I_n(\hat\theta)\right)^{-1}C\right]^{-1}(\hat\theta - \theta) \leq \chi^2_{r,1-\alpha}\right\}$
[Rao] $C(X) = \left\{\theta: [s_n(\theta,\hat\varphi)]^T[I_n(\theta,\hat\varphi)]^{-1}[s_n(\theta,\hat\varphi)] \leq \chi^2_{r,1-\alpha}\right\}$

Bayesian

[Method]

[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 $L(\theta,a) = (\theta - a)^2$, $\delta(x) = E(\theta|X = x)$.
[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 dentsity)] $C(x) = \{\theta: p_x(\theta) \geq c_\alpha\}$, often shortest length credible set. Is a horizontal line in the posterior density plot. Might not have 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 with prior unknown μ and known σ^2 $N(\mu_*(x), c^2)$:
$\mu_*(x) = \frac{\sigma^2}{\sigma^2 + n\sigma_0^2}\mu_0 + \frac{n\sigma_0^2}{\sigma^2 + n\sigma_0^2}\bar{x}$, $c^2 = \frac{\sigma_0^2\sigma^2}{n\sigma_0^2 + \sigma^2}$
$C(x) = [\mu_*(x) - cz_{1-\alpha/2},\ \mu_*(x) + cz_{1-\alpha/2}]$.

[Decision theory] [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*, Improper priors require excessive risk ignorable, take limit and observe if risk is admissible.
[Bias] Under squared error loss, $\delta(X)$ is biased unless $r_\delta(\Pi) = 0$. *No applicable to improper priors.*
[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.

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

[Asymptotic] [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.

[BM] Bernstein-von Mises: assume regularity conditions, posterior $T_n = \sqrt{n}(\tilde{\theta_n} - \hat{\theta_n}) \sim \mathcal{N}(\hat{\theta}_n, V^*/n)$ asymptotically.
[Well-specified] $V^* = E_*\left[-\nabla_\theta^2 \log p_{\theta^*}(Y)\right]^{-1}$ (same as MLE, with θ^* as true parameter, CI = CR)
[Mis-specified] $V^* = \mathbb{E}_*\left[-\nabla_\theta^2 \log p_{\theta_*}(Y)\right]^{-1} = \mathbb{E}_*\left[-\nabla_\theta^2 \log p_{\theta^*}(Y)\right]^{-1} \mathrm{Var}_*\left(\nabla \log p_{\theta^*}(Y)\right) \mathbb{E}_*\left[-\nabla_\theta^2 \log p_{\theta^*}(Y)\right]^{-1}$ (differ from MLE, with θ_* the projection of P_* to parameter space)
[Result] $\sqrt{n}\left(\hat{\theta}_n - E_\theta[\theta|X_1, \cdots, X_n]\right) \xrightarrow{P} 0$ (If MLE has asym normality, so is posterior mean)

Linear Model

[Linear Model] $X = Z\beta + \epsilon$ (or $X_i = Z_i^T\beta + \epsilon_i$)
Estimate with $b = \min_b \|X - Zb\|^2 = \|X - Z\hat{\beta}\|^2$, [solution = normal equation] $Z^Z b = Z^T X$
[Full rank]: $\hat{\beta} = (Z^T Z)^{-1} Z^T X$ [Non-full rank]: $\hat{\beta} = (Z^T Z)^{-} Z^T X$
[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] $E(\epsilon) = 0$, $Var(\epsilon) = \Sigma$

[Inference] Estimate linear combination of coefficient
[General] Necce and Suff condition: $\ell\ inR(Z) = R(Z^T Z)$
[A3] LSE $\ell^T\hat{\beta}$ is unique and unbiased
[A1] if $\ell \notin R(Z)$, $\ell^T\beta$ not estimable

[Properties] Require $\ell \in R(Z) = R(Z^T Z)$
[A1] (i) LSE $\ell^T\hat{\beta}$ is UMVUE of $\ell^T\beta$, (ii) UMVUE of $\hat{\sigma}^2 = (n-r)^{-1}\|X - Z\hat{\beta}\|^2$, r is rank of Z (iii) $\ell\hat{\beta}$ and $\hat{\sigma}^2$ are independent, $\ell^T\hat{\beta} \sim N(\ell^T\beta, \sigma^2\ell^T(Z^T Z)^{-}\ell)$, $(n-r)\hat{\sigma}/^2 \sim \chi^2_{n-r}$
[A2] LSE $\ell^T\hat{\beta}$ is BLUE (Best Linear Unbiased Estimator, best as in min var)
[A3] Following are equivalent: (a) $\ell^T\hat{\beta}$ is BLUE for $\ell^T\beta$ (also UMVUE), (b) $E[\ell^T\hat{\eta}^T X) = 0]$, any η is s.t. $E[\eta^T X] = 0$ (c) $Z^T var(\epsilon) U = 0$, for U s.t. $Z^T U = 0$, $R(U^T) + R(Z^T) = R^n$ (d) $Var(\epsilon) = Z\Lambda_1 Z^T + U\Lambda_2 U^T$, for some Λ_1, Λ_2, U s.t. $Z^T U = 0$, $R(U^T) + R(Z^T) = R^n$ (e) $Z(Z^T Z)^{-} Z^T Var(\epsilon)$ is symmetric

[Asymptotic] $\lambda_+[A]$ is the largest eigenvalue of $A_n = (Z^T Z)^{-}$.
[Consistency] Suppose $\sup_n \lambda_+[Var\epsilon)] < \infty$ and $\lim_{n\to\infty} \lambda_+[A_n] = 0$, $\ell^T\hat{\beta}$ is consistent in MSE.
[Asym Normality] $\ell^T(\hat{\beta} - \beta)/\sqrt{Var(\ell^T\hat{\beta})} \to_d N(0, 1)$ suff cond: $\lambda_+[A_n] \to 0$, $Z_n^T A_n Z_n \to 0$ as $n \to \infty$, and there exist $\{a_n\}$ s.t. $a_n \to \infty$, $a_n/a_{n+1} \to 1$, $Z^T Z/a_n$ converge to positive definite matrix.

[Testing] Under A1, $\ell \in R(Z)$, θ_0 fixed constant,
[Hypothesis testing]
(simple) $\ell \in R(Z)$, $H_0 : \ell^T\beta \leq \theta_0$, $H_1 : \ell^T\beta > \theta_0$, or $H_0 : \ell^T\beta = \theta_0$, $H_1 : \ell^T\beta \neq \theta_0$, $t(X) = \frac{\ell^T\hat{\beta} - \theta_0}{\sqrt{\ell^T(Z^T Z)^{-}\ell}\sqrt{/(n-r)}} \sim t_{n-r}$ under H_0, UMPU reject $t(X) > t_{n-r,\alpha}$ or $|t(X)| > t_{n-r,\alpha/2}$
(multiple) $L_{s\times p}$, $s \leq r$ and all rows $= \ell_j \in R(Z)$ $H_0 : L\beta = 0$, $H_1 : L\beta \neq 0$ $W = \frac{(\|X - Z\hat{\beta}_0\|^2 - \|X - Z\hat{\beta}\|^2)/s}{\|X - Z\hat{\beta}\|^2/(n-r)} \sim F_{s,n-r}$ with non-central param $\sigma^{-2}\|Z\beta - \Pi_0 Z\beta\|^2$, reject $W > F_{s,n-r,1-\alpha}$
[Confidence set] Pivotal qty: $\mathcal{R}(X, \beta) = \frac{(\hat{\beta} - \beta)^T Z^T Z(\hat{\beta} - \beta)/p}{\|X - Z\hat{\beta}\|^2/(n-p)} \sim F_{p,n-p}$, $\hat{\beta}$ is LSE of β, $C(X) = \{\beta : \mathcal{R}(X, \beta) \leq F_{p,n-p,1-\alpha}\}$