STAT 210A Crib Sheet - DIVIT RAWAL

$T(x)$ **sufficient** for P if $P_\theta(X|T)$ does not depend on θ.

Factorization Thm. P is a model w/ densities $p_\theta(x)$ w.r.t. μ. $T(x)$ is sufficient for P iff $\exists$ nonnegative g_θ, h s.t.
$p_\theta(x) = g_\theta(T(x))h(x)$ for ~~are~~ almost every (no measure 0) x under μ.

If $T(x)$ sufficient & $T(x) = f(S(x))$, $S(x)$ is sufficient.

$T(x)$ is **minimal sufficient** if $T(x)$ is sufficient & for any other sufficient stat $S(x)$, we have $T(x) = f(S(x))$.

Lehman-Scheffé Criterion: $T(x)$ minimal sufficient iff $\frac{L(\theta|y)}{L(\theta|x)}$ is indep. of θ iff $T(x) = T(y)$.

Exponential Families: $P = \{P_\eta \mid \eta \in \Xi\}$ is an s-param exp. fam. if
$p_\eta(x) = \exp(\eta^T T(x) - A(\eta))h(x)$.
$T: \mathcal{X} \to \mathbb{R}^s$ is suff. stat. $h: \mathcal{X} \to [0,\infty)$ is base density
$\eta \in \Xi \subseteq \mathbb{R}^s$ is nat. param. $A: \Xi \to \mathbb{R}$ is log partition fn.
$A(\eta) = \log\left(\int_{\mathcal{X}} \exp(\eta^T T(x))h(x)\,d\mu(x)\right) \le \infty$
$\Xi_1 = \{\eta \mid A(\eta) < \infty\} \subseteq \mathbb{R}^s$ is convex set.
$\nabla A(\eta) = E_\eta[T(x)]$; $\nabla^2 A(\eta) = \mathrm{Var}_\eta(T(x))$.
Assume $\nexists\, \alpha \in \mathbb{R}$ & $\beta \in \mathbb{R}^s \setminus \{0\}$ s.t. $\beta^T T(x) \overset{P\text{-a.s.}}{=} \alpha$.
If Ξ contains an open set $\Rightarrow$ full rank, else curved.

Score: $S_\theta(x) = \nabla \ell(\theta; x)$.
$\ell(\theta_0 + \eta; x) - \ell(\theta_0; x) \approx \eta^T S_{\theta_0}(x)$ for η small.
$E_\theta[S_\theta(x)] = 0$ if $\theta = \theta_0$.
$\mathrm{Var}_\theta[S_\theta(x)] = E_\theta[-\nabla^2 \ell(\theta; x)] \doteq J(\theta) \succeq 0$ (PSD).

Cramér-Rao Lower Bound: Let $\delta(x)$ be unbiased for $g(\theta) \in \mathbb{R}$.
$$\mathrm{Var}_\theta[\delta(x)] = \frac{\dot g(\theta)^2}{J(\theta)\,\mathrm{Corr}^2_\theta(\delta, S_\theta)} \ge \frac{\dot g(\theta)^2}{J(\theta)} \text{ where}$$
$$\mathrm{Corr}^2_\theta(\delta(x), S_\theta(x)) = \frac{\dot g(\theta)^2}{\mathrm{Var}_\theta[\delta(x)]\,J(\theta)}.$$
Multivariate Case: $\mathrm{Var}_\theta[\delta(x)] \ge \nabla g(\theta)^T J(\theta)^{-1} \nabla g(\theta)$

James-Stein Estimator: $\delta_{JS}(x) = \left(1 - \frac{d-2}{\|X\|^2}\right)X$ for $d \ge 3$. $MSE(\theta, \delta_{JS}) \le MSE(\theta, X)$ $\forall \theta \in \mathbb{R}^d$

Stein's Lemma: Assume $X \sim \mathcal{N}_d(\theta, \sigma^2 I_d)$ and $h: \mathbb{R}^d \to \mathbb{R}^d$ differentiable with $E\|Dh(x)\|_F < \infty$. Then
$$E[(X-\theta)^T h(x)] = \sigma^2 E[\mathrm{Tr}[Dh(x)]] = \sigma^2 \sum_i E\frac{\partial h_i}{\partial x_i}(x)$$
SURE: $\delta(x) = x + h(x)$.
SURE(δ) = $d\sigma^2 + \|h(x)\|^2 + 2\sigma^2 \nabla \cdot h(x)$
$E_\theta[\mathrm{SURE}(\delta)] = R(\theta, \delta)$. $E_\mu[\mathrm{SURE}(\delta)] = MSE[h]$

Minimax Estimation: goal: $\min_\delta \sup_{\theta \in \Theta} R(\theta, \delta)$.
$r^* = \inf_\delta \sup_{\theta \in \Theta} R(\theta; \delta)$. δ^* minimax if $\sup_\theta R(\theta, \delta) = r^*$
$r_\Lambda = \inf_\delta \int_\Theta R(\theta, \delta)\,d\Lambda(\theta) \le \inf_\delta \sup_\theta R(\theta; \delta) = r^*$
$\sup_\Lambda r_\Lambda \le r^*$. If Λ^* attains sup, it is least favorable.

Thm. $r_\Lambda, \delta_\Lambda$ are Bayes risk & estimator for prior Λ.
Ass. $r_\Lambda = \sup_\theta R(\theta; \delta_\Lambda)$. Then δ_Λ is minimax, Λ is least favorable, $r_\Lambda = r^*$
& if δ_Λ is unique Bayes estimator for Λ (up to a.s.), it is the unique minimax estimator.

Push-forward measure: $\nu(B) = \mu(f^{-1}(B))$ for $f^{-1}(B)$ preimage.

Completeness: $T(x)$ is complete for $P = \{P_\theta \mid \theta \in \Theta\}$ if
$$E_\theta f(T(x)) = 0 \ \forall \theta \in \Theta \Rightarrow f(T) \overset{P\text{-a.s.}}{\equiv} 0.$$
complete sufficient $\Rightarrow$ minimal sufficient.

Ancillary: $V(x)$ is ancillary if distribution of $V(x)$ does not depend on θ.
Conditionality Principle: All inference should be conditional on $V(x)$.

Basu's Thm. If $T(x)$ complete sufficient, $V(x)$ ancillary,
$V(x) \perp\!\!\!\perp T(x)\ \forall \theta \in \Theta$

Unbiased Estimation: $E_\theta \delta = g(\theta)\ \forall \theta$. If $T(x)$ is complete sufficient, $\exists!$ unbiased estimator of the form $\delta(T(x))$ & if an unbiased est. exists, it uniformly minimizes risk for any convex loss function.

Convex Loss Functions: $f(y)$ is convex if $\forall x_1, x_2$ & all $\gamma \in [0,1]$:
$f(\gamma x_1 + (1-\gamma)x_2) \le \gamma f(x_1) + (1-\gamma) f(x_2)$. Strictly convex if strict inequality for $x_1 \ne x_2$.

Jensen's Ineq. For convex f, $\forall$ random vars X, $f(E[X]) \le E[f(X)]$.
$MSE_\theta \delta = (B_\theta \delta)^2 + \mathrm{Var}_\theta(\delta(x))$.

Rao-Blackwell Thm. Let $T(x)$ be sufficient for P & let $\delta(x)$ be any estimator for $g(\theta)$. Define $\bar\delta(T(x)) = E[\delta(x) \mid T(x)]$. Then for any convex loss $L(\theta, d)$, $R(\theta, \bar\delta) \le R(\theta, \delta)\ \forall \theta$. Strict dominance for L strictly convex.

UMVU Estimators: if $T(x)$ is complete sufficient, $\exists \le$ unbiased estimator based on $T(x)$. If L is convex, we need only consider estimators based on $T(x)$.
Thm. Assume $T(x)$ is complete sufficient for P. Then
(i) For any U-estimable $g(\theta)$ $\exists$ an unbiased estimator $\delta(T(x))$.
(ii) For a (strictly) convex loss, that estimator (strictly) dominates any other estimator $\tilde\delta(x)$ unless $\tilde\delta(x) \overset{a.s.}{=} \delta(T(x))$.

Finding UMVUE: either (i) solve for unbiased estimator based on $T(x)$ or (ii) Find any unbiased estimator & Rao-Blackwellize

Bayes Estimation: $r_\Lambda(\delta) = \int_\Theta R(\theta, \delta)\,d\Lambda(\theta)$. If δ_Λ minimizes r_Λ, it is Bayes. If $\Lambda(\theta) = \infty$, the prior is improper, otherwise assume normalized to 1.
$= E_{\theta \sim \Lambda}[R(\theta, \delta)]$
$= E[L(\theta, \delta(x))]$.

$p(\theta|x) \propto_\theta p(\theta) \cdot p^{(x|\theta)}$

Assume $r_\Lambda(\delta_0) < \infty$ for δ_0. Then δ_Λ is Bayes with $r_\Lambda(\delta_\Lambda) < \infty$ iff $\delta_\Lambda(x) \in \arg\min E[L(\theta, d) \mid X = x]$ for a.e. x.

Bayes Rule: $\lambda(\theta|x) = \frac{\lambda(\theta)p_\theta(x)}{g(x)}$

Bayesian Computation: Gibbs Sampler: $\pi(x) = \pi(x_1, \dots, x_d)$ $\quad \pi(x_j \mid x_{-j})\ j = 1, \dots, d$
Set $x^{(0)} = (x_1^{(0)}, \dots, x_d^{(0)})$ For $t \in [T]$ $x_1^{(t)} \sim \pi(x_1 \mid x_2^{(t-1)}, \dots, x_d^{(t-1)})$
$\pi(x) = \int \pi(x_j \mid x_{-j})\,\pi(x_{-j})\,dx_{-j}$
$x_d^{(t)} \sim \pi(x_d \mid x_1^{(t-1)}, \dots, x_{d-1}^{(t-1)})$

t-Distribution: $t_{n-1} = \frac{\bar x - \mu}{s/\sqrt n}$ $\quad$ χ^2 Distribution $\chi^2 = \sum \frac{(O-E)^2}{E}$

$F_{k,n-p} = \frac{(RSS_{restr} - RSS_{full})/k}{(RSS_{full})/(n-p)}$ $\quad$ $\frac{RSS}{\sigma^2} \sim \chi^2_{n-p}$
$= \frac{\chi^2_k / k}{\chi^2_{n-p}/(n-p)}$ $\quad$ $F_{1,n-p} = t^2_{n-p}$ $\quad$ $\frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2_{n-1}$
k = # constraints
$n-p$ = residual DoF (data pts - remaining params)

MLE solves $E_\theta T = \bar T$
Under sq. err loss Bayes is posterior mean
Testing w/ nuisance params: condition on suff. stat of nuis. params.
$T(x)$ ranges over open interval & exp. fam. $\Rightarrow$ full rank
$I(\theta) = \mathrm{Var}[S(\theta)] = -E[\ell''(\theta)]$.
Full rank exp fam has a complete suff stat.
$\bar X \perp\!\!\!\perp S^2$ for normal vars
$E[(Z-a)^2] = \mathrm{Var}[Z] + (a - E[Z])^2 \Rightarrow \delta(x) = E[\theta|x]$.
For Exp Fam, MLE is $T(x) - E[T(x)] = 0$

Hypothesis Testing:

$H_0: \theta \in \Theta_0$ are disjoint &
$H_1: \theta \in \Theta_1$ exhaustive.

Critical Function:

$$\phi(x) = \begin{cases} 0 & \text{accept } H_0 \\ \gamma \in (0,1) & \text{reject w.p. } \gamma \\ 1 & \text{reject } H_0 \end{cases}$$

T test stat.

$$\equiv \phi(x) = \begin{cases} 0 & T(x) < c \\ \gamma \in (0,1) & T(x) = c \\ 1 & T(x) > c \end{cases}$$

Type I err: false positive
↳ reject H_0 but is true

Type II err: false negative
↳ fail to reject H_0 but is false.

Goal: minimize $\Pr[\text{Type II}]$ while keeping Type I below $\alpha \in [0,1]$. Def $\beta(\theta) = \mathbb{E}_\theta[\phi(X)] = \Pr_\theta[\text{rej } H_0]$.

$\max_\phi \beta_\phi(\theta)$ for $\theta \in \Theta$ subject to $\beta_\phi(\theta) \le \alpha$ for $\theta \in \Theta_0$.

ϕ is level-α if $\sup_{\theta \in \Theta_0} \beta_\phi(\theta) \le \alpha$.

Likelihood Ratio Test

$H_0: X \sim P_0$, $H_1: X \sim P_1$. Optimal level α test is

$$\Lambda(x) \doteq \frac{P_1(x)}{P_0(x)} \Rightarrow \phi(x) = \begin{cases} 1 & \text{if } \Lambda(x) > c \\ \gamma & \text{if } \Lambda(x) = c \\ 0 & \text{if } \Lambda(x) < c \end{cases}$$

for γ, c chosen so $\mathbb{E}_0\, \phi(X) = \alpha$ exactly.

$c_\alpha = \min\{c \in \mathbb{R} \mid \Pr_{\theta_0}[T(x) > c] \le \alpha\}$.

Testing w/ Nuisance Parameters

Observe $X \sim P_{\theta,\lambda}$ from $\mathcal{P} = \{P_{\theta,\lambda} \mid (\theta,\lambda) \in \Omega\}$.
λ is nuisance parameter
→ soln: condition out via a suff. stat. of nuisance param.

Testing in Linear Models

$Z \sim N(0,1)$, $V \sim \chi^2_\nu$; $Z \perp\!\!\!\perp V$ $\Rightarrow T = \frac{Z}{\sqrt{V/\nu}} \sim t_\nu$

$Y \sim N(X\beta, \sigma^2 I_n)$. Exp. fam. in (β, σ^2) w/ suff. stats $X^T Y$ for β, $Y^T Y$ for σ^2.

Consider hypotheses $H_0: \beta \in \mathcal{L}_0$; $H_1: \beta \in \mathcal{L}_1$ where $\mathcal{L}_0, \mathcal{L}_1$ are linear subspaces of $\mathbb{R}^p$.

Trick: $Q^T Y = \begin{pmatrix} z_0 \\ z_1 \end{pmatrix}$ ← spans H_0; ← spans orthogonal space

Under H_0, test stat is

$$F = \frac{\|\hat{Y}_1\|^2 / \dim(\mathcal{L}_1^\perp)}{\|\text{residuals}\|^2 / \dim(\mathcal{L}_0^\perp)}.$$

For $H_0: \beta_j = 0$ → reduces to t-stat.

F-test: $H_0: C\beta = 0$. $H_1: C\beta \ne 0$.

$$F = \frac{(RSS_0 - RSS_1)/q}{RSS_1/(n-p)} \quad q = \text{rank}(C) \quad F \sim F_{q,n-p}$$

O-Notation

$X_n = O_p(a_n)$:
$\forall \varepsilon > 0\ \exists M_\varepsilon > 0\ \&\ N_\varepsilon$ s.t. $\forall n \ge N_\varepsilon$

$$\Pr[|X_n| > M_\varepsilon a_n] < \varepsilon \iff \frac{X_n}{a_n} = O_p(1)$$

$X_n = o_p(a_n)$: $\frac{X_n}{a_n} \xrightarrow{P} 0$, i.e.
$\forall \varepsilon > 0$, $\Pr[|X_n| > \varepsilon a_n] \to 0$

$X_n = o_p(a_n) \Rightarrow X_n = O_p(a_n)$

Consistency of MLE Conditions

(i) Identification: $\theta \ne \theta_0 \iff f(\cdot|\theta) \ne f(\cdot|\theta_0)$
(ii) Continuity: $\Pr[\ln f(x|\theta) \in C^0(\theta)] = 1$
(iii) Dominance: $|\ln f(x|\theta)| < D(x)\ \forall \theta \in \Theta$
$\exists D(x)$ integrable w.r.t. $f(x|\theta_0)$

Uniformly Most Powerful Tests

Def. ϕ^* is UMP if it is a valid level α test and for any other valid ϕ $\beta_{\phi^*}(\theta) \ge \beta_\phi(\theta)\ \forall \theta \in \Theta_1$.

Def. P has monotone likelihood ratios (MLR) in $T(x)$ if $P_{\theta_2}(x)/P_{\theta_1}(x)$ is non-dec. fn of $T(x)$ for any $\theta_1 < \theta_2$.

Thm. Assume P has MLR in $T(x)$ and testing $H_0: \theta \le \theta_0$ against $H_1: \theta > \theta_0$ for some $\theta_0 \in \Theta \subseteq \mathbb{R}$. If ϕ^* rejects for large $T(x)$, then ϕ^* is UMP @ level $\alpha = \mathbb{E}_{\theta_0} \phi^*(x)$

Score Test: LRT for θ_0 vs. $\theta_0 + \varepsilon$ rejects for large values of

$$\log \frac{P_{\theta_0+\varepsilon}(x)}{P_{\theta_0}(x)} = \ell(\theta_0+\varepsilon; x) - \ell(\theta_0; x) \approx \varepsilon \dot{\ell}(\theta_0; x).$$

2-Sided Tests: $H_0: |\theta - \theta_0| \le \delta$ vs. $H_1: |\theta - \theta_0| > \delta$. Two tailed test

$$\phi(x) = \begin{cases} 1 & \text{if } T(x) < c_1 \text{ or } T(x) > c_2 \\ 0 & \text{if } c_1 < T(x) < c_2 \\ \gamma_i & \text{if } T(x) = c_i;\ i \in [2]. \end{cases}$$

Thm. (UMPU) Test $H_0: |\theta - \theta_0| \le \delta$, $H_1: |\theta - \theta_0| > \delta$ for $X \sim \exp(\theta^T T(x) - A(\theta)) h(x)$ for $\delta \ge 0$.
Supp. ϕ^* rejects for extreme values of $T(x)$ w/ $c_1, c_2, \gamma_1, \gamma_2$ s.t.
(i) ϕ^* has power α at boundary of null; i.e., $\beta_{\phi^*}(\theta_0 - \delta) = \beta_{\phi^*}(\theta_0 + \delta) = \alpha$.
(ii) if $\delta > 0$, $\dot{\beta}_{\phi^*}(\theta_0) = 0$. Then ϕ^* is UMPU.

p-val & interp.

$$p(x) = \sup_{\theta \in \Theta_0} \Pr_\theta[T(X) \ge T(x)].\ P(x) \text{ is p-val}$$

Confidence Regions: $C(x)$ is a $1-\alpha$ confidence region for $g(\theta)$ iff:

$$\Pr_\theta[C(X) \ni g(\theta)] \ge 1 - \alpha \quad \forall \theta \in \Theta$$

$C(x) = \{a \mid \phi(X; a) < 1\}$

UMA is inv UMP
UMAU is inv UMPU

Asymptotics

conv. in probability $X_n \xrightarrow{P} c \equiv \Pr[|X_n - c| \ge \varepsilon] \to 0\ \forall \varepsilon > 0$.
conv. in distribution $X_n \xrightarrow{d} X \equiv F_{X_n}(t) \to F_X(t)$

Cts Mapping Thm: $X_n \xrightarrow{d} X$ & g continuous, $g(X_n) \xrightarrow{d} g(X)$.

Slutsky's Thm: If $X_n \xrightarrow{d} X$ & $Y_n \xrightarrow{P} c$, $X_n + Y_n \xrightarrow{d} X + c$, $X_n Y_n \xrightarrow{d} cX$, $\frac{X_n}{Y_n} \xrightarrow{d} \frac{X}{c}$ for $c \ne 0$.

Central Lim Thm: iid $X_1, \dots, X_n$ w/ mean μ & var σ^2: $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$

Delta Method: If $\sqrt{n}(T_n - \theta) \xrightarrow{d} N(0, \tau^2)$, & g differentiable @ θ w/ $g'(\theta) \ne 0$,

$$\sqrt{n}(g(T_n) - g(\theta)) \xrightarrow{d} N(0, (g'(\theta))^2 \tau^2).$$

Asymptotic Normality of MLE: MLE $\hat{\theta}_n$ satisfies $\hat{\theta}_n \xrightarrow{P} \theta_0$ and

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, (I_1(\theta_0))^{-1}).\ \text{for } I(\theta_0) \text{ Fisher info.}$$

For Delta Meth. $\text{Var} = \nabla g^T (\text{Var}(g)) \nabla g$

Under regularity, $\nabla \ell_n(\theta_0) \approx N(0, nI(\theta_0))$.

$\ell_n(\theta) \approx \ell_n(\theta_0) + (\theta - \theta_0)^T \nabla \ell_n(\theta_0) - \frac{n}{2}(\theta - \theta_0)^T I(\theta_0)(\theta - \theta_0)$.

$\Delta_n = 2(\ell_n(\hat{\theta}) - \ell_n(\theta_0)) \xrightarrow{d} \chi^2_p$

$S_n = \nabla \ell_n(\theta_0)^T I(\theta_0)^{-1} \nabla \ell_n(\theta_0) \xrightarrow{d} \chi^2_p$

$W_n = (\hat{\theta} - \theta_0)^T (\widehat{\text{Var}}(\hat{\theta}))^{-1} (\hat{\theta} - \theta_0) \xrightarrow{d} \chi^2_p$.

Wald Test: Reject H_0 if $(\hat{\theta}_n - \theta_0)^T (\widehat{\text{Var}}(\hat{\theta}_n))^{-1} (\hat{\theta}_n - \theta_0) > \chi^2_d(\alpha)$ where d is # params tested

ARE: ARE of 1 rel to 2 is $\frac{V_2}{V_1}$ (V_i is asymptotic var).

Generalized LRT

$$L_g(x) = \frac{\sup_{\theta \in \Theta} f_\theta(x)}{\sup_{\theta \in \Theta_0} f_\theta(x)} \underset{H_0}{\overset{H_1}{\gtrless}} L_0$$

← MLEs over each other ← likelihood vals @ MLEs.

equiv $G = -2 \log \Lambda = 2(\sup_{\theta \in \Theta} \ell(\theta) - \sup_{\theta \in \Theta_0} \ell(\theta))$

Score Test

$$U(\theta) = \frac{\partial \ell(\theta|x)}{\partial \theta} \quad I(\theta) = -\mathbb{E}\left[\frac{\partial^2}{\partial \theta^2} \log f(x|\theta) \,\middle|\, \theta\right]$$

$$S(\theta_0) = \frac{U(\theta)^2}{I(\theta)} = U(\hat{\theta}_0)^T I(\hat{\theta}_0)^{-1} U(\hat{\theta}_0)$$

Bootstrapping

$X^{*(b)} \sim \hat{F}_n$ $\quad \hat{F}_n = \frac{1}{n} \sum_{i=1}^n \delta_{x_i}$ $\quad T^{*(b)} = t(X^{*(b)})$ $\quad b \in [B]$.

$$\widehat{SE}_{boot}(T) = \sqrt{\frac{1}{B-1} \sum_{b=1}^B (T^{*(b)} - \bar{T}^*)^2} \quad \widehat{Bias}_{boot} = \bar{T}^* - T \quad \mathcal{L}_{\hat{F}_n}(T^*) \approx \mathcal{L}_F(T)$$