Distribution	p.d.f.	mean	variance	C.F.	F. Info	M.S/C.S	UMVUE	Prior	Posterior
Normal(θ, σ^2)	$\frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(x-\theta)^2}{2\sigma^2})$	θ	σ^2	$\exp(i\theta t - \frac{1}{2}\sigma^2 t^2)$	$\begin{pmatrix} \sigma^{-2} & 0 \\ 0 & \frac{1}{2}\sigma^{-4} \end{pmatrix}$	$(\sum X_i, \sum X_i^2)$	$(\bar{X}, \frac{\sum(X_i-\bar{X})^2}{n-1})$	$\theta \sim N(\mu, \tau^2)$	$N(\frac{\frac{\mu}{\tau^2}+\frac{n\bar{x}}{\sigma^2}}{\frac{1}{\tau^2}+\frac{n}{\sigma^2}}, \frac{1}{\frac{1}{\tau^2}+\frac{n}{\sigma^2}})$
Poisson(λ)	$\lambda^x e^{-\lambda}/x!$	λ	λ	$\exp(\lambda(e^{it}-1))$	λ^{-1}	$\sum X_i$	$\bar{X}$	$\Gamma(\alpha, \beta)$	$\Gamma(\alpha + \sum x_i, \beta + n)$
Binomial(n, p)	$\binom{n}{x} p^x (1-p)^{n-x}$	np	$np(1-p)$	$(1-p+pe^{it})^n$	$\frac{n}{p(1-p)}$	X or $\sum X_i$	$\bar{X}$ or X/n	Beta(α, β)	$\alpha + \sum x_i, \beta + n - \sum x_i$
Gamma(α, β)	$\frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$	α/β	$\frac{\alpha}{\beta^2}$	$(1-\frac{it}{\beta})^{-\alpha}$	-			$\Gamma(\alpha_0, \beta_0)$	$\alpha_0 + n\alpha, \beta_0 + \sum x_i$
Beta(α, β)	$\frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}$	$\frac{\alpha}{(\alpha+\beta)}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	-	-				
Uniform(θ)	$\frac{1}{\theta}\mathbb{I}_{x\in(0,\theta)}$	$\frac{1}{2}\theta$	$\frac{1}{12}\theta^2$	$\frac{e^{it\theta}-1}{it\theta}$		$X_{(n)}$	$\frac{n+1}{n} X_{(n)}$	Pa(α, c)	Pa$(\alpha + n, \max(x_{(n)}, c))$
Uniform(a, b)	$\frac{1}{b-a}\mathbb{I}_{x\in(a,b)}$	$\frac{1}{2}(b+a)$	$\frac{1}{12}(b-a)^2$	$\frac{e^{itb}-e^{ita}}{it(b-a)}$					
U$\{1, \cdots, N\}$	$\frac{1}{N}\mathbb{I}_{x\in\{1,\cdots,N\}}$	$\frac{N+1}{2}$	$\frac{(N+1)(N-1)}{12}$						
Pareto(α, x_m)	$\frac{\alpha x_m^\alpha}{x^{\alpha+1}}, x \geq x_m$	$\frac{\alpha x_m}{\alpha-1}$	$\frac{x_m^2 \alpha}{(\alpha-1)^2(\alpha-2)}$	-	$\begin{pmatrix} \frac{\alpha}{x_m^2} & -\frac{1}{x_m} \\ -\frac{1}{x_m} & \frac{1}{\alpha^2} \end{pmatrix}$				
NB(r, p)	$\binom{x+r-1}{x}(1-p)^r p^x$	$\frac{pr}{1-p}$	$\frac{pr}{(1-p)^2}$	$(\frac{1-p}{1-pe^{it}})^r$	$\frac{r}{(1-p)^2 p}$	$\sum X_i$			
Geom(p)	$(1-p)^{x-1}p$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$\frac{pe^{it}}{1-(1-p)e^{it}}$					
Inv. $\Gamma(\alpha, \beta)$	$\frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} \exp(-\frac{\beta}{x})$	$\frac{\beta}{\alpha-1}$	$\frac{\beta^2}{(\alpha-1)^2(\alpha-2)}$	If	$X \sim \Gamma(\alpha, \beta)$	Then	$\frac{1}{X} \stackrel{D}{=} Y$	where	$Y \sim$ Inv. $\Gamma(\alpha, \beta)$
Cauchy(x_0, γ)	$\frac{1}{\pi\gamma[1+(\frac{x-x_0}{\gamma})^2]}$	NA	NA	$\exp(x_0 it - \gamma\|t\|)$					
Weibull(λ, k)	$\frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{(x/\lambda)^k}$	$\lambda\Gamma(\frac{k+1}{k})$	σ^2 where	$\sigma^2 = \lambda^2[\Gamma(1+$	$2/k) - (\Gamma(1+1/k))^2]$	cdf =	$1 - e^{-(x/\lambda)^k}$		
HyperGeom	$\frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$	$n\frac{K}{N}$	$n\frac{K}{N}\frac{N-K}{N}\frac{N-n}{N-1}$						

UNBIASEDNESS

Setup. Consider a set of probability measures $\{P_\theta, \theta \in \Theta\}$ on a sample space $(\mathcal{X}, \mathcal{F})$, dominated by a σ0finite measure (this assumption holds throughout, unless explicitly stated). Observe $X \sim P_\theta$ for some $\theta \in \Theta$, and infer θ. Let $L(\theta, \delta(X))$, be the loss function from estimating θ with $\delta(X)$.

Def (Dominate in measure). We say P dominates Q if $P(A) = 0 \implies Q(A) = 0, \forall A \in \mathcal{F}$.

Def (Risk fn). $R(\theta, \delta) = \mathbb{E}_\theta[L(\theta, \delta(\mathbf{X}))]$.

Def (Unbiased). An estimator $\delta(\mathbf{X})$ is unbiased for $g(\theta)$ if $\mathbb{E}_\theta \delta(X) = g(\theta), \forall \theta \in \Theta$.

Def (Minimax). An estimator δ_0 is minimax for estimating $g(\theta)$ if, for all other estimaors δ, $\sup_{\theta\in\Theta} R(g(\theta), \delta_0) \leq \sup_{\theta\in\Theta} R(g(\theta), \delta)$. The minimax risk of any estimator δ is $\sup_{\theta\in\Theta} R(g(\theta), \delta)$.

Def (Bayes Risk). Under a prior model $\theta \sim \pi(\theta)$, $r(\pi, \delta) = \mathbb{E}_{\theta\sim\pi}[R(\theta, \delta)] = \mathbb{E}[L(\theta, \delta(\mathbf{X}))]$.

Def (Statistic). A measurable function $T : (\mathcal{X}, \mathcal{F}) \to (\mathbb{R}^k, \mathcal{B})$.

Def (Sufficient) A statistic T is called sufficient for θ (or for $\{P_\theta, \theta \in \Theta\}$) if the conditional distribution of $X|T$ is independent of θ.

Thm (Neyman-Fisher Factorization Criterion). Suppose $\{P_\theta, \theta \in \Theta\}$ is a collection of probability measures on $(\mathcal{X}, \mathcal{F})$, which are dominated by a σ-finite measure γ. Let $X \sim P_\theta$ for some $\theta \in \Theta$. Then T is sufficient for θ iff $p_\theta(x) = g_\theta(T(x))h(x)$ a.s. γ, for some g_θ, h, where $p_\theta(\cdot) = dP_\theta/d\gamma$.

Def (Exp. Fam.) $\{P_\theta, \theta \in \Theta\}$ (dominated by σ-finite

measure) is said to form a k-dimensional exponential family if the corresponding pdfs are of the form
$p_\theta(x) = \exp\{\sum_{i=1}^k \eta_i(\theta)T_i(x) - B(\theta)\}h(x)$,
where $h, T_1, \cdots, T_j : \mathcal{X} \to \mathbb{R}$ and $B, \eta_1, \cdots, \eta_k : \Theta \to \mathbb{R}$.

Def (Support). The support of a density is the set where the density is strictly positive.

Thm (Pitman-Koopman-Darmois). Suppose $X_1, ..., X_n$ are iid with density $\{p_\theta, \theta \in \Theta\}$, which are continuous in x for fixed θ and supported on an interval $I \subseteq \mathbb{R}$. Suppose there exists a sufficient statistic $(T_1, \cdots, T_k)$ with continuous components.
(i) If $k = 1$, then $p_\theta(x) = e^{\eta(\theta)T(x)-B(\theta)}h(x)$.
(ii) If $n > k > 1$, and $x \mapsto p_\theta(x)$ is C^1, then $p_\theta(x) = \exp\{\sum_{i=1}^k \eta_i(\theta)T_i(x) - B(\theta)\}h(x)$.

Def (M.S.). Let S be sufficient for θ. S is Minimal Sufficient if, given any other sufficient statistic T, there exists a measurable fn. h s.t. $S(x) = h(T(x))$ a.s. P_θ, $\forall \theta \in \Theta$.

Thm (Bahadur). Let $X \sim P_\theta, \theta \in \Theta$ be an $\mathbb{R}^n$-valued RV. Then a MS statistic exists.

Thm (M.S.). If $\Theta_0 = \{\theta_0, ..., \theta_k\}$ and p_θ have common support $I \subseteq \mathcal{X}$, then
$T(x) = \left(\frac{p_{\theta_1}(X)}{p_{\theta_0}(X)}, ..., \frac{p_{\theta_k}(X)}{p_{\theta_0}(X)}\right)$ is M.S.

Thm (M.S.). Let $\{P_\theta : \theta \in \Theta\}$ be a collection of dominated probability measures with common support, and $\Theta_0 \subseteq \Theta$. If T is sufficient for $\{P_\theta : \theta \in \Theta\}$, and M.S for $\{P_\theta : \theta \in \Theta_0\}$, then T is M.S for $\{P_\theta : \theta \in \Theta\}$.

Thm (Lehman-Scheffe Partitions). Suppose $T(x) = T(y)$ iff the ratio $p_\theta(x)/p_\theta(y)$ is independent of θ. Then T is M.S.
Rigorous formulation: Suppose $\{P_\theta : \theta \in \Theta\}$ is a dominated by a σ-finite measure ν. Suppose $T(x) = T(y)$ iff $\exists \alpha, \beta > 0$ (depending on x, y) s.t. $\alpha p_\theta(x) = \beta p_\theta(y)$, $\forall \theta$. Then T is M.S.

Thm (M.S. for Exp. Fam.). Let $\{P_\theta, \theta \in \Theta\}$ be an exponential familhy of the form
$p_\theta(x) = \exp\{\sum_{i=1}^k \eta_i(\theta)T_i(x) - B(\theta)\}h(x)$, and let $\overline{\eta} = \{(\eta_1(\theta), ..., \eta_k(\theta)) : \theta \in \Theta\} \subseteq \mathbb{R}^k$.
(a) If $\exists \mathbf{v_0}, \mathbf{v_1}, \cdots, \mathbf{v_k} \in \overline{\eta}$ s.t. $\{\mathbf{v_1} - \mathbf{v_0}, \cdots, \mathbf{v_k} - \mathbf{v_0}\}$ are lin. indep., then $(T_1, \cdots, T_k)$ is M.S.
(b) If $(\overline{\eta})^0 \neq \emptyset$, then $(T_1, \cdots, T_k)$ is M.S.

Def (C.S.) A suff. stat. T is **complete** for θ if $\mathbb{E}_\theta f(T) = 0, \forall \theta \in \Theta \implies f(T) = 0$ a.s. $P_\theta, \forall \theta$.

Lemma (MGF). If $\mathbb{E}e^{tX} = \mathbb{E}e^{tY}, \forall t \in (-\delta, \delta)$, then $X \overset{D}{=} Y$.

Thm (C.S. for Exp. Fam.). In the previous setting, if $(\overline{\eta})^0 \neq \emptyset$, then $(T_1, ..., T_k)$ is C.S.

Thm (C.S. & M.S.). If $\exists$ a CS statistic T and $\exists$ an MS statistic U, then T is M.S. (and U C.S.).

Def (Ancillary). A statistic S is ancillary for θ if the distribution of S is free of θ.

Thm (Basu). If T is C.S and V is ancillary, then T and V are independent (under $P_\theta, \forall \theta \in \Theta$).

Def. $\mathcal{U} = \{U : \mathcal{X} \to \mathbb{R} : \mathbb{E}_\theta U(X)^2 < \infty, \mathbb{E}_\theta U(X) = 0, \forall \theta \in \Theta\}$
$\Delta = \{\delta : \mathcal{X} \to \mathbb{R} : \mathbb{E}_\theta \delta(X) = g(\theta), \mathrm{Var}(\delta(X)) < \infty\}$.
Note if $\Delta \neq \emptyset$, then $\mathcal{U} + \delta = \Delta, \forall \delta \in \Delta$.

Def (UMVUE). An estimator $\delta_0 \in \Delta$ is UMVUE if, $\forall \delta \in \Delta$, $\mathrm{Var}_\theta \delta_0(X) \leq \mathrm{Var}_\theta \delta(X), \forall \theta \in \Theta$.

Thm. δ_0 is UMVUE iff $\mathbb{E}_\theta \delta_0(X)U(X) = 0, \forall U \in \mathcal{U}$.

Def (Convexity). $C \subseteq \mathbb{R}^k$ is convex if
$x \in C, y \in C \implies \alpha x + (1-\alpha)y \in C, \forall \alpha \in (0, 1)$.
A function $f : C \to \mathbb{R}$ is convex if
$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y)$, $\forall x, y \in C$ and $\alpha \in (0, 1)$. Change to $<$ for strictly convex.

Remark. If $\nabla\phi$ exist, then ϕ is convex iff $\phi(y) \geq \phi(x) + (y-x)^T\nabla\phi(x), \forall x \neq y$ ($>$ for strictly convex).
If ϕ is twice differentiable, then ϕ is (strictly) convex if $H = \left[\frac{\partial^2\phi}{\partial x_i \partial x_j}\right]_{i,j}$ exists and is +ve (semi)definite.

Thm (Jensen's Inequality). Let $\phi : I \to \mathbb{R}$ be convex, where $I \subseteq \mathbb{R}$ is an interval.
(a) If $\mathbb{E}|X| < \infty$, then $\mathbb{E}\phi(x) \geq \phi(\mathbb{E}X)$.
(b) If ϕ is strictly convex, strict inequality holds above, unless $X = \mathbb{E}X$ a.s.
Note this also holds for conditional expectations

Thm (Rao-Blackwell). Let T be sufficient for θ.
(a) If $\delta(X)$ is unbiased for $g(\theta)$ and $a \mapsto L(g(\theta), a)$ is convex, then $\eta(T) = \mathbb{E}_\theta[\delta(X)|T]$ is unbiased and $R(g(\theta), \eta(T)) \leq R(g(\theta), \delta(X)), \forall \theta \in \Theta$.
(b) If δ_0 is unbiased and has finite risk $\forall \theta$, and $a \mapsto L(g(\theta), a)$ is strictly convex, then $R(g(\theta), \eta(T)) < R(g(\theta), \delta(X)), \forall \theta$, unless δ is a function of T a.s. $P_\theta, \forall \theta \in \Theta$.

Corollary (UMVUE) If $\delta(X)$ is an unbiased estimate of $g(\theta)$ and T is C.S., then $\mathbb{E}_\theta[\delta(X)|T]$ is the UMVUE.

Defn (Score func). For $\Theta = \mathbb{R}^k$,
$\mathbf{S} = (\partial_{\theta_1} \log p_\theta(X), \cdots, \partial_{\theta_k} \log p_\theta(X))^T$.

Defn (Fisher Info). $I(\theta) = \mathbb{E}_\theta(\partial_\theta \log p_\theta(X))^2$.
On $\mathbb{R}^k$, $I(\theta) = [\mathbb{E}_\theta(\partial_{\theta_i} \log p_\theta(X))(\partial_{\theta_j} \log p_\theta(X))]_{i,j}$.

Remark. If $\eta = \tau(\theta) : \tau \in C^1, \tau'(\theta) \neq 0$, then $I(\tau(\theta)) = I(\theta)/\tau'(\theta)^2$.
On $\mathbb{R}^k$, the information matrix is +ve semi-definite (symmetry is obvious) because $I(\theta) = \mathbb{E}[\mathbf{S}\mathbf{S}^T]$.

Thm (CRLB/Information Inequality). Suppose
(a) $\Theta \subseteq \mathbb{R}$ is an open interval.
(b) $\{p_\theta(x), \theta \in \Theta\}$ have common support.
(c) $p'_\theta(x) = \frac{\partial}{\partial\theta}p_\theta(x)$ exists and is finite for all x and θ.
(d) $\partial_\theta \int_\mathcal{X} p_\theta(x)d\mu = \int_\mathcal{X} \partial_\theta p_\theta(x)d\mu$.
Let $\delta(X)$ be an estimator s.t. $\mathbb{E}[\delta(X)^2] < \infty$, and $I(\theta) \in (0, \infty)$, and $\int_\mathcal{X} \delta(x)\partial_\theta p_\theta(x)d\mu = \partial_\theta \int_\mathcal{X} \delta(x)p_\theta(x)d\mu$.
Then $\mathrm{Var}(\delta(X)) \geq [\partial_\theta \mathbb{E}\delta(X)]^2/I(\theta)$.
(this is just $g'(\theta)^2/I(\theta)$ for unbiased estimators).

Remark. If equality holds, $p_\theta(x)$ is a 1-parameter exp. fam. and $\delta(X)$ is the natural sufficient stat.

Lemma (Fisher info). Assume (a) - (d) and $I(\theta) < \infty$. Then $I(\theta) = \mathrm{Var}(\partial_\theta \log(p_\theta(x)))$.
If, in addition, $p''_\theta(x)$ exists $\forall \theta, x$, and $\partial^2_\theta \int p_\theta(x)d\mu = \int \partial^2_\theta p_\theta(x)d\mu$, then $I(\theta) = -\mathbb{E}[\partial^2_\theta \log(p_\theta(X))]$.

Thm. Let $p_\theta(x) = e^{\eta(\theta)T(x)-B(\theta)}h(x)$ and $\theta \in \Theta$, an open interval. Let $\tau(\theta) = \mathbb{E}_\theta[T(X)]$, and assume T is not constant. Then
(a) $\tau'(\theta) \neq 0$ and $I(\tau(\theta)) = 1/\mathrm{Var}_\theta(T)$.
(b) $I(h(\theta)) = [\eta'(\theta)/h'(\theta)]^2\mathrm{Var}_\theta(T)$.

Propn (regularity of exp. fam.). Let μ be a σ-finite measure on $\mathcal{X}$ and $t_1, \cdots, t_n : \mathcal{X} \to \mathbb{R}$. Define $G(\theta_1, \cdots, \theta_n) = \int_\mathcal{X} \exp\{\sum \theta_i t_i(x)\}h(x)d\mu$, and $\Omega = \{\theta : G(\theta_1, \cdots, \theta_n) < \infty\}$. Then
(a) Ω is convex and $\theta \mapsto \log G(\theta)$ is convex on Ω.
(b) Let Ω_0 be the interior of Ω and assume $\Omega_0 \neq \emptyset$. Then, on Θ_0, $\theta \mapsto G(\theta)$ is infinitely differentiable and the derivatives can be taken inside the integral, e.g. $\partial_{\theta_i} G = \int_\mathcal{X} t_i(x) \exp\{\sum \theta_i t_i(x)\}h(x)d\mu$.

Remark. Similar conclusions hold with the normalizing constant $e^{-B(\theta)}$. Moreover, $B(\theta) \in \mathcal{C}^\infty$.

Remark. For a general function $\eta : \Omega \to \mathbb{R}$, all conclusions hold at $\theta = \theta_0$, provided $\eta(\theta_0)$ is an interior point if $\overline{\eta} = \{\eta : \int e^{\eta t(x)}h(x)d\mu < \infty\}$ and $\eta \in \mathcal{C}^\infty$.

Propn. If $p_\eta(x) = e^{\sum_{i=1}^k \eta_i T_i(x) - A(\eta)}h(x)$, and $\eta \in \overline{\eta}$,

- $E_\eta(T_j) = \frac{\partial}{\partial \eta_j} A(\eta)$
- $Cov_\eta(T_j, T_k) = \frac{\partial^2}{\partial\eta_j \partial\eta_k} A(\eta)$.

If $p_\theta(x) = \exp\{\sum_{i=1}^k \eta_i(\theta)T_i(x) - B(\theta)\}h(x)$, $\eta(\theta_0) \in \overline{\eta}$,
- If $k = 1$, then $E_{\theta_0}(T(X)) = B'(\theta_0)/\eta'(\theta_0)$ and $Var(T(X)) = \frac{B''(\theta)}{\eta'(\theta)^2} - \frac{\eta''(\theta)B'(\theta)}{\eta'(\theta)^3}$.
- If $k > 1$, then $E_\theta(T(X)) = J^{-1}\nabla B$, where $J = \{\frac{\partial \eta_j}{\partial \theta_i}\}_{ij}$ and $\nabla B = \{\frac{\partial}{\partial\theta_i} B(\theta)\}_i$.

Propn (regularity of the estimator). Let $\delta(X)$ be an estimator s.t. $\mathrm{Var}(\delta(X)) < \infty$. Then $\partial_\theta \int \delta(x)p_\theta(x)d\mu = \int \delta(X)\partial_\theta p_\theta(x)d\mu$, at any $\theta_0 \in (\Omega)^0$, provided $\exists b(x)$ s.t. $|\frac{p_{\theta_0+h}(x) - p_{\theta_0}(x)}{hp_{\theta_0}(x)}| \le b(x)$ for all sufficiently small h, and $\int b(x)|\delta(x)|p_\theta(x)d\mu < \infty$ (in particular, this will hold if $\mathbb{E}_{\theta_0}[b(X)^2] < \infty$, by Cauchy-Schwarz).

Propn (regularity of estimator in exp. fam.). Let $p_\theta(x) = e^{\eta(\theta)t(x)-B(\theta)}h(x)$ and $\eta \in \mathcal{C}^\infty$ (so that $B \in C^\infty$). If $\delta(X)$ is an estimator with $\mathrm{Var}(\delta(X)) < \infty$, then $\partial_\theta \int \delta(x)p_\theta(x)d\mu = \int \delta(x)\partial_\theta p_\theta(x)d\mu$.

Thm (Multi-parameter CRLB). Suppose
(a) $\Theta \subseteq \mathbb{R}^k$ is an open set.
(b) $\{p_\theta(x), \theta \in \Theta\}$ have common support.
(c) $\partial_{\theta_i} p_\theta(x)$ exists, $\forall i, x, \theta$, and is finite.
(d) $\partial_{\theta_i} \int_{\mathcal{X}} p_\theta(x)d\mu = \int_{\mathcal{X}} \partial_{\theta_i} p_\theta(x)d\mu$.
(e) $\partial_{\theta_i} \int_{\mathcal{X}} \delta(x)p_\theta(x)d\mu = \int_{\mathcal{X}} \delta(x)\partial_{\theta_i} p_\theta(x)d\mu$.
(f) $I(\theta)$ is finite and +ve definite.
Then we have $\mathrm{Var}(\delta(X)) \ge \alpha^T I(\theta)^{-1}\alpha$, where $\alpha_i = \partial_{\theta_i}\mathbb{E}_\theta\delta(X)$. In particular, if $\delta(X)$ is unbiased for $g(\theta)$, $\alpha_i = \partial_{\theta_i} g(\theta)$.

AVERAGE RISK OPTIMALITY

Setup. Suppose $\{P_\theta, \theta \in \Theta\}$ is a collection of probability measures on $\mathcal{X}$ dominated by a σ-finite measure μ. Assume now that θ is a random variable on Θ, with prior distn. π. Suppose we want to estimate $g(\theta)$. The risk function is still $R(g(\theta), \delta) = \mathbb{E}_{X\sim P_\theta} L(g(\theta), \delta(X)) = \mathbb{E}[L(g(\theta), \delta(X))|\theta]$.

Def (Bayes risk) of δ: $r(\pi, \delta) = \mathbb{E}_{\theta\sim\pi}[R(g(\theta), \delta)]$

Def (Bayes estimator). δ_0 is a Bayes estimator if $r(\pi, \delta_0) \le r(\pi, \delta)$ for any other estimator δ.

Def (Bayes risk of a Prior). $r(\pi) = \inf_\delta(r(\pi, \delta))$.

Remark. The joint distribution of (X, θ) is $p_\theta(x)\pi(\theta)$. The marginal distribution of X is $m(x) = \int_\Theta p_\theta(x)\pi(d\theta)$. The *posterior distn.* is $\pi(\theta|x) = p_\theta(x)\pi(\theta)/m(x) \propto p_\theta(x)\pi(\theta)$.

Thm (Bayes estimator for sq. err. loss). If $L(g(\theta), \delta(X)) = (g(\theta) - \delta(X))^2$, and $\mathbb{E}[g(\theta)^2] < \infty$,
(i) $\delta_0 = \mathbb{E}[g(\theta)|X]$ is a Bayes estimator with Bayes risk $\mathbb{E}[\mathrm{Var}(g(\theta)|X)]$.
(ii) If $\delta(X)$ is any other Bayes estimator, then $\delta_0(X) = \delta(X)$ a.s. under the joint distn. of (X, θ).

Remark. (ii) also implies $\delta_0(X) = \delta(X)$ a.s. under the marginal of X. If the marginal dominates the conditional, this will further imply that $\delta_0(X) = \delta(X)$ a.s. $P_\theta, \forall\theta \in \Theta$, i.e. we have uniqueness under the conditionals.

Lemma (Bias of Bayes estimator). Under squared error loss, a Bayes estimator cannot be unbiased, unless $\delta(X) = g(\theta)$ a.s.

Def (Conjugate Prior). A non-trivial class of probability distributions F is called a conjugate family of priors for a model $\{P_\theta : \theta \in \Theta\}$ if the posterior distribution $\pi(\theta|x)$ also belongs to F.

Example. For $p_\theta(x) = \exp\{\sum_{i=1}^k \eta_i(\theta)T_i(x) - B(\theta)\}h(x)$, the conjugate family is $\pi(\theta) = \exp\{\sum_{i=1}^k s_i\eta_i(\theta) - s_0B(\theta)\}\psi(s_0, ..., s_k)$

Def (least favourable). A prior π is least favourable if, for all other distributions π' on Θ, $r(\pi) \ge r(\pi')$. A sequence of priors $\{\pi_n\}_{n\ge1}$ is least favourable if $\lim_{n\to\infty} r(\pi_n) = \sup_\pi r(\pi)$.

Thm (minimax from Bayes). Suppose π is a distribution on Θ with Bayes estimator δ_π, s.t. $r(\pi) = r(\pi, \delta_\pi) = \sup_{\theta\in\Theta} R(g(\theta), \delta_\pi)$. Then:
(a) δ_π is minimax
(b) If δ_π is the unique (w.r.t. the conditionals) Bayes estimate w.r.t. π, then δ_π is unique minimax.
(c) π is least favourable.

Corollary. A Bayes estimator with constant risk is minimax.

Thm (minimax from L.F.). Suppose $\{\pi_n\}_{n\ge1}$ is a sequence of priors s.t. $\lim_{n\to\infty} r(\pi_n) = \sup_{\theta\in\Theta} R(g(\theta), \delta_0)$ for some estimate δ_0. Then:
(a) δ_0 is minimax.
(b) $\{\pi_n\}_{n\ge1}$ is least favourable.

Lemma (minimax on subset). Suppose $\delta(X)$ is minimax for $g(\theta)$ on the parameter set $\Theta_0 \subseteq \Theta$. If $\sup_{\theta\in\Theta_0} R(g(\theta), \delta) = \sup_{\theta\in\Theta} R(g(\theta), \delta)$, then δ is minimax for $\theta \in \Theta$.

Def (Admissible). An estimator δ is *inadmissible* if $\exists\delta'$ s.t. $R(g(\theta), \delta') \le R(g(\theta, \delta)$, with strict inequality for some $\theta \in \Theta$. Otherwise, δ is *admissible.*

Remark. If the loss is strictly convex, any estimator which is not a function of the M.S. statistic is inadmissible (Rao-Blackwell).

Lemma. If the loss is strictly convex, δ is admissible and $R(g(\theta), \delta) = R(g(\theta), \delta'), \forall\theta \in \Theta$, then $\delta = \delta'$ a.s. $P_\theta, \forall\theta \in \Theta$.

Lemma. Any unique (w.r.t. the conditionals) Bayes estimator is admissible.

Lemma. An admissible estimator with constant risk is minimax. If the loss function is strictly convex, it is also *unique* minimax.

Lemma. If δ is unique minimax, then δ is admissible.

Thm (Karlin). Suppose $\{P_\theta, \theta \in \Theta\}$ is a one-parameter exponential family $p_\theta(x) = e^{\theta T(x)-B(\theta)}h(x)$, for $\theta \in (a, b)$ (possibly unbounded). Let $\delta_{\lambda,\nu}(X) = \frac{1}{1+\lambda}T(X) + \frac{\nu\lambda}{1+\lambda}, \lambda \ge 0, \nu \in \mathbb{R}$. If $\exists\theta_0 \in \Theta$ s.t. $\int_a^{\theta_0} e^{-\nu\lambda\theta+\lambda B(\theta)}d\theta = \int_{\theta_0}^b e^{-\nu\lambda\theta+\lambda B(\theta)}d\theta = \infty$, then $\delta(X)$ is admissible for estimating $g(\theta) = \mathbb{E}_\theta T(X)$, w.r.t squared error loss.

Corollary If $(a, b) = (-\infty, \infty)$, then T is admissible for $\mathbb{E}_\theta T$.

Def (improper prior). A measure π on the parameter space Θ s.t. $\pi(\Theta) = \infty$.
If $m(x) := \int_\Theta p_\theta(x)\pi(d\theta) < \infty, \forall x \in \mathcal{X}$, we can define a probability measure $\pi(\cdot|x)$ on Θ by $\pi(A|x) = \int_A p_\theta(x)\pi(d\theta)/m(x)$.

Def (generalized Bayes estimate). A minimizer of $\int_{\Theta\times\mathcal{X}} L(g(\theta), \delta(x))p_\theta(x)\pi(d\theta)d\mu$, where π is an improper prior.

Thm (generalized Bayes estimate). If $m(x) < \infty, \forall x$, a generalized Bayes estimate, w.r.t squared error, is the posterior mean $\int_\Theta g(\theta)\pi(\mathrm{d}\theta|x)$, provided $\int_\Theta g(\theta)^2\pi(\mathrm{d}\theta) < \infty$.

Remark (Jeffrey's Prior). One common "vague"/improper prior is $\pi(\theta) \propto \sqrt{I(\theta)}$. In the multi-parameter case, $\pi(\theta) \propto \sqrt{\det(I(\theta))}$

Def (hierarchical Bayes). The prior distribution on the parameter θ has a *hyper-parameter*, λ, which itself has a *hyper-prior*. We have, $X|\theta \sim p_\theta(x)$, $\theta|\lambda \sim \pi_\lambda(\theta)$, $\lambda \sim \psi(\lambda)$.

Thm. Writing $\pi(\theta) = \int \pi_\lambda(\theta)\psi(\lambda)d\lambda$, we have that $D(\pi(\theta|x)||\pi(\theta)) \geq D(\psi(\lambda|x)||\psi(\lambda))$. (HW5 q5)

Def (K-L divergence). $D(P||Q) = \int p(x)\log\frac{p(x)}{q(x)}dx$.

Remark. It always exists and is ≥ 0 (maybe = infinity), with equality iff $p = q$.

Def (empirical Bayes estimate). Assume the hyperparameter λ is now fixed. An estimator derived from the posterior $\theta|x$ (e.g. the posterior mean) now also depends on λ. Substituting λ with a non-trivial estimator of λ derived from the marginal of X yields an *empirical Bayes* estimate for θ.

James Stein Estimator. Let $g(\mathbf{x}) = \frac{(n-2)\sigma^2}{||\mathbf{x}||_2^2}\mathbf{x}$. Then $\delta_{JS} = \mathbf{x} - g(\mathbf{x})$ and has a uniformly better risk than the UMVUE estimator ($\delta = \mathbf{x}$) for $n \geq 3$. (HW5 Q2)

ASYMPTOTIC OPTIMALITY

Setup. Consider a candidate estimator $\delta_n(X_1, ..., X_n)$ for estimating $g(\theta)$.

Def (Consistency). $\delta_n(X)$ is consistent for $g(\theta)$ if $\delta_n(X) \xrightarrow{p} g(\theta)$, under P_θ $\forall\theta \in \Theta$.

Def (Likelihood). $L(\theta|\mathbf{X}) = \prod_{i=1}^n p_\theta(X_i)$. If $\eta = g(\theta)$, the likelihood of η is $\tilde{L}(\eta|\mathbf{X}) = \sup_{\theta:g(\theta)=\eta} L(\theta|\mathbf{X})$.

Def (MLE). If there exists a unique $\hat{\theta}_n$ which is a global maximizer of $\theta \mapsto L(\theta|\mathbf{X})$, then $\hat{\theta}_n$ is the MLE.

Def (Asymptotic efficiency). for a sequence of estimators $\tilde{\theta}_n$: $\sqrt{n}(\tilde{\theta}_n - \theta_0) \xrightarrow[\theta_0]{\mathcal{D}} N(0, I(\theta_0)^{-1})$

Def (Tightness). A sequence of RVs $\{Y_n\}_{n\geq1}$ is tight if $\forall\epsilon > 0$, $\exists K_\epsilon < \infty$ s.t. $\sup_{m\geq1} P(|Y_n| > K_\epsilon) \leq \epsilon$.

Thm. If $Y_n \xrightarrow{\mathcal{D}} Y$, then $\{Y_n\}_{n\geq1}$ is tight.

Def ($\sqrt{n}$-consistent). An estimator $\tilde{\theta}_n$ is $\sqrt{n}$-consistent for θ if $\sqrt{n}(\tilde{\theta}_n - \theta_0)$ is tight under P_{θ_0}, $\forall\theta_0 \in \Theta$.

Thm. If $\tilde{\theta}_n$ is $\sqrt{n}$-consistent for θ, then $\tilde{\theta}_n \xrightarrow{p} \theta$.

Asymptotic Risk Thm (MLE) $X_1, ..., X_n \overset{iid}{\sim} P_\theta, \theta \in \Theta$, with pdf $p_\theta(\cdot)$. Consider the hypotheses:
(A0) Identifiability: $P_{\theta_1} \neq P_{\theta_2}$ whenever $\theta_1 \neq \theta_2$.
(A1) $\{p_\theta(\cdot), \theta \in \Theta\}$ have common support.
(A2) $\Theta \subseteq \mathbb{R}$ and θ_0 is an interior point of Θ.
(A3) The function $\theta \mapsto p_\theta(x)$ is 3 times differentiable and $\sup_{\theta\in[\theta_0-\delta,\theta_0+\delta]} |\partial_\theta^3 \log p_\theta(x)| \leq M(x)$, with $\mathbb{E}_{\theta_0}[M(X_1)] < \infty$, for some $\delta > 0$.
(A4) $\theta \mapsto \int_\mathcal{X} p_\theta(x)d\mu(x)$ can be differentiated twice through the integral. Further, $0 < I(\theta_0) < \infty$.
(A2*) Θ is an open interval.
(A3*) The map $\theta \mapsto p_\theta(x)$ is $\mathcal{C}^2$ and $\sup_{\theta\in[\theta_0-\delta,\theta_0+\delta]} |\partial_\theta^2 \log p_\theta(x)| \leq M(x)$, with $\mathbb{E}[M(X_1)] < \infty$, for some $\delta > 0$.

- Under A0 and A1, $P_{\theta_0}(l_n(\theta_0|\mathbf{X}) > l_n(\theta|\mathbf{X})) \to 1$ as $n \to \infty$, $\forall\theta \neq \theta_0$.
- Under A0 and A1, if Θ is finite, the MLE $\hat{\theta}_n$ exists with high probability (i.e. the probability that the likelihood function has a unique maximizer goes to 1), and $P_{\theta_0}(\hat{\theta}_n = \theta_0) \to 1$ as $n \to \infty$.
- Under A0-2, if $\theta \mapsto p_\theta(x)$ is $\mathcal{C}^1$ (differentiable with continuous derivative), there exists a sequence of roots $\hat{\theta}_n$ of the likelihood equation $l_n'(\theta) = 0$ which is consistent for θ_0 (though $\hat{\theta}_n$ depends on θ_0 so is not an estimator).
- Under A0-2, if $\theta \mapsto p_\theta(x)$ is differentiable and the likelihood equation $l_n'(\theta) = 0$ has a unique root $\hat{\theta}_n$, then $\hat{\theta}_n \xrightarrow{p} \theta_0$ under P_{θ_0}, and $\hat{\theta}_n$ is the MLE w.h.p.
- (Asymptotic normality of MLE). Under A0-4, for any consistent sequence of roots $\hat{\theta}_n$ of $l_n'(\theta) = 0$, we have $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} N(0, I(\theta_0)^{-1})$.
- Under A0, A1, A4, A2* and A3*, if $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} N(0, V(\theta_0))$, then the set $\{\theta : V(\theta) < I(\theta_0)^{-1}\}$ has Lebesgue measure 0.
- Under A0-4, if $\tilde{\theta}_n$ is $\sqrt{n}$-consistent for θ, then $\delta_n := \tilde{\theta}_n - l_n'(\tilde{\theta}_n)/l_n''(\tilde{\theta}_n)$ is asymptotically efficient.

Remark. $\frac{l_n'(\theta_0)}{\sqrt{n}} \xrightarrow{\mathcal{D}} N(0, I(\theta_0))$, $\frac{l_n''(\theta_0)}{n} \xrightarrow{p} -I(\theta_0)$, $|\frac{l_n'''(\xi_n)}{n}| \leq \frac{1}{n}\sum M(X_i) \xrightarrow{p} EM(X_1) < \infty$ where $\xi_n \in (\theta_0, \hat{\theta}_n)$.

Propn. If the MLE is consistent and conditions A0 through A4 hold, then the MLE is asymptotically efficient (HW6 Q6).

Example (Exp. Fam.). Let $p_\theta(x) = e^{\theta T(x) - B(\theta)}h(x)$, $\theta \in \Theta$, an open interval. Let $l_n(\theta) = \log\prod p_\theta(x_i)$. Then $l_n''(\theta) = -nB''(\theta) = -n\text{Var}(T(X)) < 0$, so $\theta \to l_n(\theta)$ is strictly concave so $l_n'(\theta) = 0$ can have at most 1 root.

Thm (Slutsky). Suppose $X_n \xrightarrow{\mathcal{D}} X$, $A_n \xrightarrow{p} a$, $B_n \xrightarrow{p} b$. Then $A_nX_n + B_n \xrightarrow{\mathcal{D}} aX + b$.

Thm (Invariance of MLE). (a) If $\hat{\theta}$ is a global maximizer of $\theta \mapsto L(\theta|\mathbf{X})$, then $\hat{\eta} = g(\hat{\theta})$ is a global maximizer of $\eta \mapsto \tilde{L}(\eta|\mathbf{X})$.
(b) If $\hat{\theta}$ is the MLE and $\forall\eta$, $|\{\theta : g(\theta) = \eta\}| < \infty$, then $\hat{\eta}$ is the MLE for η.

Thm (Δ-Method). If $\sqrt{n}(X_n - \mu) \xrightarrow{\mathcal{D}} N(0, \sigma^2)$, and $g \in \mathcal{C}^1$ s.t. $g'(\mu) \neq 0$, then $\sqrt{n}(g(X_n) - g(\mu)) \xrightarrow{\mathcal{D}} N(0, \sigma^2 g'(\mu)^2)$.

Remark. Multivariate result holds $\sqrt{n}(g(X_n) - g(\mu)) \xrightarrow{\mathcal{D}} N(0, \xi^T\Sigma\xi)$ where $\xi_i = \frac{\partial g}{\partial x_i}|_{x=\mu}$

Thm (Modified Δ-Method). If $\sqrt{n}(X_n - \mu) \xrightarrow{\mathcal{D}} N(0, \sigma^2)$, and $g \in \mathcal{C}^2$ s.t. $g'(\mu) = 0$, then $n(g(X_n) - g(\mu)) \xrightarrow{\mathcal{D}} \frac{\sigma^2}{2}g''(\mu)\chi_1^2$.

Thm (Uniform integrability). If $X_n \xrightarrow{\mathcal{D}} X$ and $\sup_{n\geq1} \mathbb{E}[|X_n|^{1+\delta}] < \infty$ for some $\delta > 0$, then $\mathbb{E}X_n \to \mathbb{E}X$.

Thm (Multivariate CLT for MLE). Under A0, A1, and:
(A2) $\Theta \subseteq \mathbb{R}^p$ and $\theta_0 \in \Theta$ is an interior point.
(A3) The function $\theta \mapsto p_\theta(x)$ is 3 times partially differentiable and $\sup_{||\theta-\theta_0||_2<\delta} \left|\frac{\partial^3 \log p_\theta(x)}{\partial\theta_i\partial\theta_j\partial\theta_k}\right| \leq M_{ijk}(x)$, where $\mathbb{E}_{\theta_0} M_{ijk}(\mathbf{X}) < \infty$, $\forall i, j, k$.
(A4) $\mathbb{E}_{\theta_0}\partial_{\theta_i} \log p_\theta(X) = 0$ and $\mathbb{E}_{\theta_0}[\frac{\partial \log p_\theta(X)}{\partial\theta_i}\frac{\partial \log p_\theta(X)}{\partial\theta_j}] = -\mathbb{E}_{\theta_0}\left[\frac{\partial^2 \log p_\theta(X)}{\partial\theta_i\partial\theta_j}\right] = I_{ij}(\theta_0)$, with the matrix $I(\theta_0)$ finite and +ve definite.

- Then there exists a consistent sequence of roots of the likelihood equation $\frac{\partial \log p_\theta(x)}{\partial\theta_i} = 0$, $1 \leq i \leq p$.
- Further, this sequence is asymptotically efficient, i.e. $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} N(\mathbf{0}, I(\theta_0)^{-1})$.

HYPOTHESIS TESTING

Setup. Let $\{P_\theta, \theta \in \Theta\}$ be a collection of probability measures on $\mathcal{X}$ dominated by a σ-finite measure μ. Let $p_\theta(\cdot) = \frac{dP_\theta}{d\mu}$. Let Θ_0 and Θ_1 be disjoint subsets of Θ. Given $X \sim P_\theta$ for some $\theta \in \Theta$, we want to test whether $\theta \in \Theta_0$ or $\theta \in \Theta_1$.

Def (Test function). A function $\phi : \mathcal{X} \to \{0, 1\}$ is called a non-randomized test function.

Def. Types of errors of a test. If $\theta \in \Theta_0$, then $\phi = 1$ is *Type I error.* If $\theta \in \Theta_2$, then $\phi = 0$ is *Type II error.*

Def (Power). The power of a test ϕ is 1 - Probability of type II error; $\beta(\theta) = P_\theta(\phi = 1)$ for $\theta \in \Theta_1$, a function of θ.

Def (Size). The size of a test ϕ is $\sup_{\theta\in\Theta_0} P_\theta(\phi=1)$. Let $\alpha\in(0,1)$. A test ϕ is called level α if $\sup_{\theta\in\Theta_0} P_\theta(\phi=1)\le\alpha$.

Def (UMP). A test ϕ is called uniformly most powerful level α if, given any other level α test ψ, we have $P_\theta(\phi=1)\ge P_\theta(\phi=1)\ \forall\theta\in\Theta_1$.

Def. A function $\phi:\mathcal{X}\to[0,1]$ is called a *randomized test function.* If $\phi=p$, toss a coin w prob heads p. If heads choose Θ_1, else Θ_0. In all previous definitions, replace $P_\theta(\phi=1)$ by $\mathbb{E}_\theta[\phi]$, and $P_\theta(\phi=0$ by $1-\mathbb{E}_\theta[\phi]$.

Thm (NP lemma). Suppose we want to test $H_0:\theta=\theta_0$ vs $H_1:\theta=\theta_1$ at level α.
(i) There exists a test ϕ satisfying
. (a) $\mathbb{E}_{\theta_0}[\phi]=\alpha$
. (b) There exists $k\in[0,\infty]$ such that

$$\phi(X)=1 \text{ if } p_{\theta_1}(X)>kp_{\theta_0}(X)$$
$$=0 \text{ if } p_{\theta_1}(X)<kp_{\theta_0}(X)$$

(ii) If a test ϕ satisfies (a) and (b), then ϕ is a Most Powerful test for testing $\theta=\theta_0$ vs $\theta=\theta_1$.
(iii) If ϕ is Most Powerful level α, it must satisfy (b) for some k. It also satisfies (a), unless $\mathbb{E}_{\theta_1}[\phi]=1$, in which case $\mathbb{E}_{\theta_0}[\phi]\le\alpha$.

Remark. If the boundary $\{X:p_{\theta_1}(X)=kp_{\theta_0}(X)\}$ has measure 0, then the MP test is unique.

Corollary. Let $\beta=\beta(\theta_1)$ denote the power of the MP test for testing $\theta=\theta_0$ vs $\theta=\theta_1$ at level $\alpha\in(0,1)$. Then $\beta\ge\alpha$. Further, $\beta>\alpha$ unless $p_{\theta_1}=p_{\theta_0}$.

Def (MLR). Suppose Θ is an interval (*Keener only requires that* $\Theta\subseteq\mathbb{R}$). We say that $\{p_\theta(\cdot),\theta\in\Theta\}$ have the Monotone Likelihood Ratio property in a statistic $T(X)$, if $\forall\theta_1<\theta_2\in\Theta$, $p_{\theta_2}(x)/p_{\theta_1}(x)$ is a non-decreasing function of $T(X)$.
Keener: Natural conventions concerning division by zero are used here, with the likelihood ratio interpreted as ∞ when $p_{\theta_2}>0$ and $p_{\theta_1}=0$. On the null set where both densities are zero the likelihood ratio is not defined and monotonic dependence on T is not required.

Thm. Let $\{p_\theta(\cdot),\theta\in\Theta\}$ be MLR in $T(X)$, Θ an interval, and $p_{\theta_1}\ne p_{\theta_2}$ if $\theta_1\ne\theta_2$.
(i) For testing $H_0:\theta\le\theta_0$ vs $H_1:\theta>\theta_0$ at level $\alpha\in(0,1)$, there exists a UMP test ϕ of the form

$$\phi(X)=1 \text{ if } T(X)>c$$
$$=\nu \text{ if } T(X)=c$$
$$=0 \text{ if } T(X)<c,$$

and $\mathbb{E}_{\theta_0}\phi(X)=\alpha$.
(ii) The power function $\beta(\theta)=\mathbb{E}_\theta\phi$ is strictly increasing on the set $\{\theta:0<\beta(\theta)<1\}$.
(iii) For all $\theta'\in\Theta$, the test of part (i) is UMP for testing $H_0:\theta\le\theta'$ vs $H_1:\theta>\theta'$ at level $\alpha'=\beta(\theta')$.
(iv) For any $\theta<\theta_0$, ϕ minimises $\beta(\theta)$ among all tests satisfying $\mathbb{E}_{\theta_0}\psi(X)=\alpha$.

Lemma. Let $\{p_\theta(.),\theta\in\Theta\}$ be MLR in $T(X)$, and Θ an interval.
(i) If $\psi:\mathbb{R}\to\mathbb{R}$ is non-decreasing, then so is $\theta\mapsto E_\theta\psi(T)$.
(ii) If ψ has a simple change of sign, i.e. $\exists x_0\in\mathbb{R}$ s.t
$T(x)<x_0 \implies \psi(T(x))\le 0$
$T(x)>x_0 \implies \psi(T(x))\ge 0$
Then one of three things happen:
a. $E_\theta\psi(T)\ge 0,\ \forall\theta\in\Theta$
b. $E_\theta\psi(T)\le 0,\ \forall\theta\in\Theta$
c. $\exists\theta_0$ s.t. $E_\theta\psi(T)\le 0,\ \forall\theta<\theta_0$, $E_\theta\psi(T)\ge 0,\ \forall\theta>\theta_0$.
(iii) Suppose $p_\theta(x)>0,\forall x\in\mathcal{X},\theta\in\Theta$ and the function $p_{\theta'}(x)/p_\theta(x)$ is strictly increasing in $T(x)$ for $\theta'>\theta$.
Let ψ be as in (ii) and further assume $P_\theta(\psi(T)\ne 0)>0$. If $E_{\theta_0}\psi(T)=0$, then $E_\theta\psi(T)>0$ for $\theta>\theta_0$, $E_\theta\psi(T)<0$ for $\theta<\theta_0$.

Lemma. Assume $p_\theta(x)>0,\forall\theta\in\Theta,x\in\mathcal{X}$, Θ an interval, and $p_{\theta'}(x)/p_\theta(x)$ is strictly increasing in $T(X),\forall\theta<\theta'$. Then there is a unique test function ϕ, which is a function of T, of the form:
$\phi(X)=1$ if $T(X)\in(c_1,c_2)$
$=\nu_i$ if $T(X)=c_i$
$=0$ if $T(X)\notin[c_1,c_2]$
such that $E_{\theta_1}\phi=\alpha_1$ and $E_{\theta_2}\phi=\alpha_2$, for some $\theta_1\ne\theta_2$, $\alpha_1,\alpha_2\in(0,1)$.
That is to say, if $\phi^*(X)$ is such that
$\phi^*(X)=1$ if $T(X)\in(c_1^*,c_2^*)$
$=\nu_i^*$ if $T(X)=c_i^*$
$=0$ if $T(X)\notin[c_1^*,c_2^*]$
and $E_{\theta_1}\phi^*=\alpha_1$ and $E_{\theta_2}\phi^*=\alpha_2$, then $\phi=\phi^*$ a.s.

Thm (Generalized NP). Let $f_1,...,f_{m+1}$ be real-valued integrable functions w.r.t μ. Let $(c_1,...,c_m)\in\mathbb{R}^m$ and set $\mathcal{C}_0=\{\phi:\int\phi f_i d\mu=c_i,1\le i\le m,\phi\text{ is a test fn}\}$ and assume $\mathcal{C}_0$ is not empty.
(i) Among all $\phi\in\mathcal{C}_0$, there exists a test ϕ_0 which maximizes $\int\phi f_{m+1}d\mu$.
(ii) A sufficient condition for $\phi_0\in\mathcal{C}_0$ to maximize $\int\phi f_{m+1}d\mu$ is that $\exists(K_1,...,K_m)$ s.t.
$\phi_0=1$ if $f_{m+1}>K_1f_1+...+K_mf_m$ $(*)$
$\phi_0=0$ if $f_{m+1}<K_1f_1+...+K_mf_m$ $(*)$
(iii) If $\phi_0\in\mathcal{C}_0$ satisfies $(*)$ for some $K_1,...,K_m\ge 0$, then ϕ_0 maximizes $\int\phi f_{m+1}d\mu$ among all tests ϕ satisfying $\int\phi f_i d\mu\le c_i$, for $1\le i\le m$.
(iv) The set $M=\{(\int\phi f_1d\mu,...,\int\phi f_md\mu),\phi\text{ is a test fn}\}$, a subset of $\mathbb{R}^m$, is closed and convex. If $(c_1,...,c_m)$ is an interior point of M, then $\exists K_1,...,K_m$ and $\phi_0\in\mathcal{C}_0$ such that $(*)$ holds, and a necessary condition for $\phi_0\in\mathcal{C}_0$ to maximize $\int\phi f_{m+1}d\mu$ is that $(*)$ holds a.s. (for some $K_1,\cdots,K_m$).

Propn. If ϕ is MP, and T is sufficient, then $\psi:=\mathbb{E}[\phi|T]$ is MP for the same test.

Thm. Suppose we want to test $H_0:\theta\le\theta_1$ or $\theta\ge\theta_1$ vs $H_1:\theta_1<\theta<\theta_2$ at level α, where $X\sim p_\theta(x)=e^{\eta(\theta)T(x)-B(\theta)}h(x)$, and η strictly increasing.
(i) There exists a UMP test ϕ which satisfies:
$\phi(X)=1$ if $c_1<T<c_2$
$=\nu_i$ if $T=c_i$
$=0$ otherwise.
and $\mathbb{E}_{\theta_1}\phi=\mathbb{E}_{\theta_2}\phi=\alpha$.
(ii) Among all tests ψ satisfying $E_{\theta_1}\psi=E_{\theta_2}\psi=\alpha$, ϕ minimizes type I error $E_{\theta'}\psi$ for any $\theta'\le\theta_1$ or $\theta'\ge\theta_2$.

Setup (Least Favorable π). Consider the problem of testing $H_0:\theta\in\Theta_0$ vs $H_1:\theta=\theta_1$. Let π be a distribution on Θ_0 and let $m(x)=\int_{\Theta_0}p_\theta(x)\pi(d\theta)$. Consider the modified problem $H_0':X\sim m(\cdot)$ vs $H_1:X\sim p_{\theta_1}(\cdot)$. Let ϕ_π be the NP test (MP) at level α with power β_π.

Theorem. Assume ϕ_π is level α for the original problem. Then:
(i) ϕ_π is MP for the original problem.
(ii) If ϕ_π is unique MP for the modified problem, then ϕ_π is unique MP for the original problem.
(iii) $\beta_\pi\le\beta_{\pi'},\forall\pi'$. (i.e π is least favorable).

Remark. To find a UMP under a composite null, use a Least Favourable Prior (including point masses)! (unless we can apply our standard MLR/exp. fam. results).

Def (p-value). Suppose we want to test H_0 vs H_1 at level α. Let ϕ_α be a non-randomized test function at level α. Let $S_\alpha=\{X:\phi_\alpha(X)=1\}$ be the rejection region, and assume these are nested: $\alpha_1<\alpha_2\implies S_{\alpha_1}\subseteq S_{\alpha_2}$. The p-value is $\hat{p}(X)=\inf\{u:X\in S_u\}$.
Intuitively, given the p-value, you can construct a level α test by rejecting H_0 if $\hat{p}(X)<\alpha$, accepting otherwise.

Lemma. Suppose $X\sim p_\theta$ for some $\theta\in\Theta$, and we want to test $H_0:\theta\in\Theta_0$ vs $H_1:\theta\in\Theta_1$ at level α. Let $\{\phi_\alpha\}_{\alpha\in(0,1)}$ be a collection of nested level α tests.

(i) Then $P_\theta(\hat{p}(X) \le u) \le u, \forall u \in (0,1), \theta \in \Theta_0$
(ii) If $\exists \theta_0 \in \Theta_0$ such that $P_{\theta_0}(X \in S_\alpha) = \alpha, \forall \alpha$ then $P_{\theta_0}(\hat{p}(X) \le u) = u$.

Def (Confidence Interval). Let $X \sim P_\theta$ for some $\theta \in \Theta$. For every $x \in \mathcal{X}$, let $\mathcal{S}(x)$ be a subset of Θ. We say the collection of sets $\{\mathcal{S}(x), x \in \mathcal{X}\}$ is a $(1-\alpha)$ confidence region if $P_\theta(\theta \in \mathcal{S}(X)) \ge 1-\alpha$, $\forall \theta \in \Theta$.
Assume $\Theta \subseteq \mathbb{R}$. If $\mathcal{S}(x) = [l(x), \infty)$, then we call it a lower confidence interval. If $\mathcal{S}(x) = (-\infty, u(x)]$, an upper CI. If $\mathcal{S}(x) = [l(x), u(x)]$, a 2-sided CI.

Remark. Suppose for every $\theta_0 \in \Theta$, ϕ_{θ_0} is a non-randomized level α test for $H_0 : \theta = \theta_0$ vs H_1.
Let $\mathcal{S}(x) = \{\theta : \phi_\theta(X) = 0\}$. Then $\{\mathcal{S}(x) : x \in \mathcal{X}\}$ is a $(1-\alpha)$ confidence region.

Remark (Asymptotic CI). In practice, suppose $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, V^2(\theta))$ where V is continuous. Then, by Slutsky's (and cts. mapping thm), $\sqrt{n}\frac{\hat{\theta}-\theta}{V(\hat{\theta})} \xrightarrow{d} N(0,1)$, and therefore, $(\hat{\theta} - \frac{1}{\sqrt{n}} z_{1-\alpha/2} V(\hat{\theta}), \hat{\theta} + \frac{1}{\sqrt{n}} z_{1-\alpha/2} V(\hat{\theta}))$ is a $1-\alpha$ C.I. for θ.

Def (Unbiased Test). Suppose we want to test $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_1$ at level α. We say a test ϕ is level α unbiased if
(i) $\sup_{\theta \in \Theta_0} E_\theta \phi \le \alpha$
(ii) $\inf_{\theta \in \Theta_1} E_\theta \phi \ge \alpha$

Def (UMPU). We say ϕ is Uniformly Most Powerful Unbiased at level α, if ϕ is unbiased at level α and for any other unbiased test ψ, $E_\theta \phi \ge E_\theta \psi, \forall \theta \in \Theta_1$.

Remark. If ϕ is UMP, it is also UMPU.

Lemma (UMPU). Suppose $\{p_\theta, \theta \in \Theta\}$ is a collection of prob. measures, s.t. $\theta \mapsto E_\theta \phi$ is continuous in θ (metric on Θ implicit). If ϕ_0 is a test such that:
(i) ϕ_0 is UMP among the class of tests satisfying $E_\theta \phi = \alpha, \forall \theta \in \partial\Theta_0 \cap \partial\Theta_1$. ($\partial S$ = boundary of S).
(ii) ϕ_0 is level α for $\theta \in \Theta_0$.
Then ϕ_0 is UMPU for $\theta \in \Theta_0$ vs $\theta \in \Theta_1$ at level α.

Theorem. Let $X \sim p_\theta(x) = e^{\eta(\theta)T(x) - A(\theta)} h(x)$, η strictly increasing and continuous, and Θ an open interval. For the test $H_0 : \theta \in [\theta_1, \theta_2]$ vs $H_1 : \theta \notin [\theta_1, \theta_2]$, there exists a UMPU level α test ϕ given by:
$\phi = 1$ if $T(X) \notin [c_1, c_2]$
$= \nu_i$ if $T(X) = c_i$
$= 0$ otherwise.
and $E_{\theta_1}\phi = E_{\theta_2}\phi = \alpha$.

Theorem. $X \sim p_\theta(x) = e^{\eta(\theta)T(x) - A(\theta)} h(x)$, Θ is an open interval, $\eta \in \mathcal{C}^1$ and $\eta'(\theta) > 0$. We want to test $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$ at level α. There exists a UMPU of the form:
$\phi = 1$ if $T(X) \notin [c_1, c_2]$
$= \nu_i$ if $T(X) = c_i$
$= 0$ if $T(X) \in (c_1, c_2)$,
where $E_{\theta_0}\phi = \alpha$ and $E_{\theta_0}\{\phi(X)T(X)\} = \alpha E_{\theta_0}\{T(X)\}$.

Lemma. Let $M = \{(E_{\theta_0}[\phi], E_{\theta_0}[\phi T]), \phi \text{ is a test fn}\} \subseteq \mathbb{R}^2$. Then for any $\alpha \in (0,1)$, $(\alpha, \alpha E_{\theta_0} T)$ is an interior point of M. *(consider $\phi = \alpha \pm \varepsilon I(T > E_{\theta_0} T)$)* (hw3 q3)

Lemma. Suppose ϕ is a test of the form
$\phi = 1$ if $T(x) > c$
$= \nu$ if $T(x) = c$
$= 0$ if $T(x) < c$
Then $E_{\theta_0}\phi = \alpha$ and $E_{\theta_0}\phi T = \alpha E_{\theta_0} T$ cannot hold simultaneously. *(consider $(\phi - \alpha)(T - c) \ge 0$)*

Lemma. There is at most one test of the form:
$\phi = 1$ if $T \notin [c_1, c_2]$
$= 0$ if $T \in (c_1, c_2)$
$= \nu_i$ if $T = c_i$
such that $E_{\theta_0}\phi = \alpha$, $E_{\theta_0}\phi T = \alpha E_{\theta_0} T$. (HW3 Q4)

Theorem. Suppose $X \sim p_{\theta,\eta}(x) = e^{\theta U(x) + \sum_{i=1}^K \eta_i T_i(x) - A(\theta,\eta)} h(x)$ where $(\theta, \eta) \in \Theta \times \Omega$ is open. Suppose we want to test $H_0 : \theta \le \theta_0$ vs $H_1 : \theta > \theta_0$ at level α. In this case, there exists a UMPU of the form
$\phi = 1$ if $U > K(\mathbf{T})$
$= \nu(\mathbf{T})$ if $U = K(\mathbf{T})$
$= 0$ if $U < K(\mathbf{T})$
where $E_{\theta_0,\eta}(\phi(U, \mathbf{T})|\mathbf{T}) = \alpha$ a.s.

Remark. The conditional distribution of U given $T = t$ is an exponential family of the form $\tilde{p}(u|t) = e^{\theta u - A_t(\theta)} h_t(u)$, $\theta \in \Theta$.

Remark. Similarly, you can find UMPU in the exponential family $p_{\theta,\eta}(x) = \exp\{\theta U(x) + \sum_{i=1}^k \eta_i T_i(x) - A(\theta,\eta)\} h(x)$ for these problems:
(ii) $H_0 : \theta \notin (\theta_1, \theta_2)$ vs $H_1 : \theta \in (\theta_1, \theta_2)$.
(iii) $H_0 : \theta \in [\theta_1, \theta_2]$ vs $H_1 : \theta \notin [\theta_1, \theta_2]$.
(take $\mathcal{C} = \{\psi : E_{\theta_1,\eta}(\psi|T) = \alpha \text{ a.s.}, E_{\theta_2,\eta}(\psi|T) = \alpha \text{ a.s.}\}$)
(iv) $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$
(take $\mathcal{C} = \{\psi : E_{\theta_0,\eta}(\psi|T) = \alpha \text{ a.s.}, E_{\theta_0,\eta}(\psi U|T) = \alpha E_{\theta_0}(\psi|T) \text{ a.s.}\}$)

Def (LRT). Suppose $X_1, \cdots, X_n$ are iid from $p_\theta(\cdot)$, and you want to test $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_1$. The LRT statistic is $\Lambda(X_1, \cdots, X_n) = \frac{\sup_{\theta \in \Theta_0} p_\theta(X_1, \cdots, X_n)}{\sup_{\theta \in \Theta_0 \cup \Theta_1} p_\theta(X_1, \cdots, X_n)}$.

Remark. In many examples $-2\log\Lambda(X_1, \cdots, X_n)$ has an asymptotic χ^2 distribution with $\dim(\Theta_0 \cup \Theta_1) - \dim(\Theta_0)$ degrees of freedom.

Thm (Wilks). Suppose A0-A4 hold, MLE is consistent, $\Theta \subseteq \mathbb{R}^k$ open. Suppose we want to test $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$. Then $-2\log\Lambda(X_1, \cdots, X_n) \xrightarrow{d} \chi^2_k$.

Wald's Test. $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$, A0-A4 and MLE consistent. Thus, $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, I(\theta_0)^{-1})$ under H_0. Reject H_0 if $|\hat{\theta}_n - \theta_0| > \frac{z_{1-\alpha/2}}{\sqrt{nI(\theta_0)}}$. For general k, reject if $n(\hat{\theta}_n - \theta_0)^T I(\theta_0)(\hat{\theta}_n - \theta_0) > \chi^2_{k,1-\alpha}$. Can replace $I(\theta_0)$ by $I(\hat{\theta}_n)$ and still have this asymptotic distn.

Rao Score Test. Let $U_\theta(X_i) = \frac{\partial}{\partial\theta}\log p_\theta(X_i)$. We know $\mathbb{E}_{\theta_0} U_{\theta_0}(X_i) = 0$, $Var_{\theta_0} U_{\theta_0}(X_i) = I(\theta_0)$, so $\frac{1}{\sqrt{n}}\sum U_{\theta_0}(X_i) \underset{\theta_0}{\xrightarrow{d}} N(0, I(\theta_0))$. So reject $H_0 : \theta = \theta_0$ if $|\frac{1}{\sqrt{n}}\sum U_{\theta_0}(X_i)| > \frac{z_{1-\alpha/2}}{\sqrt{I(\theta_0)}}$.

M-ESTIMATION

Setup. $X_1, \cdots, X_n \overset{iid}{\sim} P$ on $(\mathcal{X}, \mathcal{A})$. Family of *criterion functions* $m_\theta(x), m_\theta : \mathcal{X} \to \mathbb{R}$, $\theta \in \Theta$ (e.g. $-L(\theta, X)$).

Def (M-estimator). $\hat{\theta}_n = \arg\max_{\theta \in \Theta} \frac{1}{n}\sum m_\theta(x_i)$.
- e.g. mean minimizes $\frac{1}{n}\sum_{i=1}^n (X_i - \theta)^2$
- e.g. median minimizes $\frac{1}{n}\sum_{i=1}^n |X_i - \theta|$

Def (Z-estimator). $\hat{\theta}_n$ such that $\sum M_\theta(x_i) = 0$.
- e.g. MLE often solves $\sum_{i=1}^n \nabla_\theta \log p_\theta(X_i) = 0$

Setup. $K \subseteq \mathbb{R}^p$ compact. $\mathcal{C}(K)$ is the space of continuous functions $K \to \mathbb{R}$. $\mathcal{C}(K)$ is a Banach space with norm $||w||_\infty = \sup_{t \in K} |w(t)|$, and it is separable (has a countable dense subset) $W_1, W_2, \cdots$ are iid random functions on $\mathcal{C}(K)$ (e.g. $W_i(t) = m_t(X_i)$).

Thm. Suppose W is a random function in $\mathcal{C}(K)$, K compact. Let $\mu(t) = \mathbb{E}W(t)$, $t \in K$. If $\mathbb{E}||W||_\infty < \infty$, then
(i) μ is continuous.
(ii) Define $M_\varepsilon(t) := \sup_{s:||t-s||<\varepsilon} |W(s) - W(t)|$. Then $\sup_{t \in K} \mathbb{E}M_\varepsilon(t) \to 0$ as $\epsilon \downarrow 0$

Thm. $W_1, W_2, \cdots$ iid random functions in $\mathcal{C}(K)$, K compact. Let $\mu(t) = \mathbb{E}W(t)$, $\overline{W}_n(\cdot) = \frac{1}{n}\sum W_i(\cdot)$. If $\mathbb{E}||W||_\infty < \infty$, then $||\overline{W}_n - \mu||_\infty \xrightarrow{p} 0$ as $n \to \infty$.

Thm. $\{G_n\}_{n \ge 1}$ random functions in $\mathcal{C}(K)$, K compact. Suppose $||G_n - g||_\infty \xrightarrow{p} 0$, g non-random in $\mathcal{C}(K)$. Then
(i) If $\{t_n\}_{n \ge 1} \subseteq K$ are random vectors s.t. $t_n \xrightarrow{p} t^* (\in K)$, then $G_n(t_n) \xrightarrow{p} g(t^*)$.

(ii) If g achieves its maximum at a unique t^* and if $\{t_n\}_{n\geq 1}$ are random vectors maximizing G_n, i.e. $G_n(t_n) = \sup_{t\in K} G_n(t)$, then $t_n \xrightarrow{p} t^*$.
(iii) *(from Keener, 9.4.3)* If $K \subseteq \mathbb{R}$ and $g(t) = 0$ has a unique solution t^*, and if t_n are RVs solving $G_n(t_n) = 0$, then $t_n \xrightarrow{p} t^*$.

Remark (MLE). $X_1, \cdots, X_n$ iid p_θ, $\theta \in \Theta$, θ_0 denotes the truth. $\hat\theta_n = \arg\max_{\theta\in\Theta}[l_n(\theta) - l_n(\theta_0)]$, $l_n(\theta) = \sum \log p_\theta(x_i)$. Here $\overline{W}_n = l_n(\theta) - l_n(\theta_0)$, where $W_i(\theta) = \log\frac{p_\theta(x_i)}{p_{\theta_0}(x_i)}$; $\mathbb{E}W_i(\theta) = -I(\theta_0,\theta) = -\int \log\frac{f_{\theta_0}(x)}{f_\theta(x)} f_{\theta_0}(x)d\mu(x)$, (KL divergence). Have $\theta_0 = \arg\max_{\theta\in\Theta}\mathbb{E}W(\theta)$, by lemma.

Lemma. If $P_\theta \neq P_{\theta_0}$, then $I(\theta_0,\theta) > 0$ and $I(\theta_0,\theta_0) = 0$.

Thm. $\Theta \subseteq \mathbb{R}^p$ compact. $\mathbb{E}_{\theta_0}||W||_\infty < \infty$ where $W(\theta) = \log\frac{p_\theta(X)}{p_{\theta_0}(X)}$. $p_\theta(\cdot)$ is a continuous function in θ for almost all x. $p_\theta \neq p_{\theta_0}$ $\forall\theta \neq \theta_0$. Then, under P_{θ_0}, $\hat\theta_n \xrightarrow{p} \theta_0$.

Thm. Let $\Theta = \mathbb{R}^p$, $W(\theta) = \log\frac{p_\theta(X)}{p_{\theta_0}(X)}$. Suppose
(i) $\theta \mapsto p_\theta(x)$ is cts.
(ii) $\theta \neq \theta_0 \implies p_\theta \neq p_{\theta_0}$
(iii) $\forall K$ compact, $K \subseteq \Theta$, $\mathbb{E}_{\theta_0}\sup_{\theta\in K}|W(\theta)| < \infty$
(iv) $\exists a > 0$ s.t. $\mathbb{E}_{\theta_0}\sup_{||\theta||>a} W(\theta) < \infty$.
(v) $p_\theta(x) \to 0$ as $||\theta||_2 \to \infty$.
Then $\hat\theta_n \xrightarrow{p} \theta_0$ under P_{θ_0}, where $\hat\theta_n$ denotes the MLE, if it exists.

Remark. The weaker condition $\mathbb{E}_{\theta_0}\sup_{\theta\in\Theta} W(\theta) < \infty$ is sufficient. Also, $\Theta \subseteq \mathbb{R}^p$ can be any open set.

Remark. Let $\hat\theta_n$ be a global maximizer of $\overline{W}_n(\theta)$. Assume A0-A4, and $\hat\theta_n$ is consistent. Then $\sqrt{n}(\hat\theta_n - \theta_0) \xrightarrow{d} N(0, I(\theta_0)^{-1})$ under P_{θ_0}. (pf: check whp $l_n'(\hat\theta) = 0$)

Thm. Let $W(\theta) = \log\frac{p_\theta(X)}{p_{\theta_0}(X)}$. Suppose
(i) $\mathbb{E}_{\theta_0}\sup_{\theta\in\Theta} W(\theta) < \infty$
(ii) $\theta \mapsto p_\theta(x)$ is upper semi cts
(iii) $\theta \neq \theta_0 \implies P_\theta \neq P_{\theta_0}$
(iv) $\Theta = \cup_{l\geq 1} K_l$, K_l compact, increasing, s.t. $\lim_{l\to\infty}\sup_{\theta\in K_l^C} W(\theta) = -\infty$ a.s. (w.r.t. $p_\theta, \forall\theta$).
Then $\hat\theta_n \xrightarrow{p} \theta_0$. (HW3 Q1)
a) $\exists l$ s.t. $\theta_0, \hat\theta_n \in K_l$ whp. b) Fix $\delta > 0$. $\forall\theta \in B_l := K_l \cap \{\theta \in \Theta : d(\theta,\theta_0) \geq \delta\}$, $\exists$ neighborhood V_θ s.t. $E_{\theta_0}\sup_{\theta\in V_\theta} W(\theta, X_1) < E_{\theta_0}W(\theta_0, X_1)$ (by u.s.c.). c) B_l is compact + WLLN $\implies$ $\sup_{\theta\in B_l}\overline{W}_n(\theta, X) < W_n(\theta_0, X)$ whp

Prop. Let Θ be an interval and $Z_n(\theta)$ a random func s.t.
(i) $\theta \mapsto Z_n(\theta)$ is non-decreasing with $Z_n(\hat\theta_n) = o_p(1)$
(ii) $Z_n(\theta) \xrightarrow{p} Z(\theta), \forall\theta$, where $Z(\theta)$ is non-random.
(iii) $\theta \mapsto Z(\theta)$ is strictly increasing with $Z(\theta_0) = 0$.
Then $\hat\theta_n \xrightarrow{p} \theta_0$. (HW3 Q2).

CONTIGUITY AND LAN

Def (absolute continuity of measure). Let P and Q be two probability measures on $(\mathcal{X},\mathcal{F})$. We say P is absolutely continuous w.r.t Q (noted $P \ll Q$) if $Q(A) = 0 \implies P(A) = 0$.
By Radon-Nikodyn theorem, $P \ll Q$ iff $P(A) = \int_A h\mathrm{d}Q$ for some non-negative measurable $h : (\mathcal{X},\mathcal{F}) \to (\mathbb{R},\mathcal{B})$, i.e. $dP/dQ = h$.

Prop. $P \ll Q$ iff $Q(A_n) \to 0 \implies P(A_n) \to 0, \forall\{A_n\}$.

Def (contiguity). Let P_n and Q_n be prob measures on $(\mathcal{X}_n,\mathcal{F}_n)$. P_n is contiguous to Q_n (noted $P_n \triangleleft Q_n$) if $Q_n(A_n) \to 0 \implies P_n(A_n) \to 0$.

Prop. $P_n \triangleleft Q_n$ iff $T_n \underset{Q_n}{\xrightarrow{p}} 0 \implies T_n \underset{P_n}{\xrightarrow{p}} 0$ $\forall$ RVs T_n on $\mathcal{X}_n$

Def (total variation distance). $||P-Q||_{TV} = \sup_{A\in\mathcal{F}}|P(A) - Q(A)|$. If μ is a dominating measure for P and Q, and $p = \mathrm{d}P/\mathrm{d}\mu$, $q = \mathrm{d}Q/\mathrm{d}\mu$, then $||P-Q||_{TV} = \frac{1}{2}\int_{\mathcal{X}}|p(x) - q(x)|\mathrm{d}\mu$.
Also $||P-Q||_{TV} = |P(A) - Q(A)|$ where $A = \{\frac{p}{q} \geq 1\}$.

Prop. If $||P_n - Q_n||_{TV} \to 0$ then $P_n \triangleleft\triangleright Q_n$. Note the converse is <u>not</u> true (e.g. $P_n = N(0,1)$, $Q_n = N(1,1)$).

Thm (Portmanteau). Let S be a metric space, with a Borel σ-algebra. Let P_n, P be prob measures on S. Then TFAE:
(i) $\lim_{n\to\infty}\int g\mathrm{d}P_n = \int g\mathrm{d}P$, $\forall g$ bounded continuous.
(ii) $\limsup_{n\to\infty}\int g\mathrm{d}P_n \leq \int g\mathrm{d}P$, $\forall g$ u.s.c. bounded above.
(iii) $\liminf_{n\to\infty}\int g\mathrm{d}P_n \geq \int g\mathrm{d}P$, $\forall g$ l.s.c. bounded below.
(iv) $P_n(A) \to P(A)$, $\forall A$ s.t $P(\partial A) = 0$.

Remark. Can change $\int gdP_n$ to $E_{P_n}g(X_n)$ and $\int gdP$ to $E_P g(X)$ in (i) - (iv).
Note $E_{P_n}[g(X_n)] = \int g(X_n)dP_n = \int g(X_n(\omega))P_n(d\omega) = \int g(x)P^{X_n}(dx) = E_{P^{X_n}}[g]$, where $P^{X_n}(A) = P_n(X \in A)$ is the distribution function of X_n.
Note also that $X_n \underset{P_n}{\xrightarrow{d}} X$ means $E_{P_n}g(X_n) \to X_P g(X)$ for all g bdd. cts., or equivalently that the distn. funcs P^{X_n} converge weakly to P^X.

Remark. If U is open, 1_U is l.s.c, and if K is closed, 1_K is u.s.c. Moreover, for (ii) and (iii), we can equivalently take g just of this form.

Def. f is l.s.c. at x_0 if $\forall\varepsilon > 0\exists\delta > 0 : ||x - x'|| < \delta \implies f(x') \geq f(x_0) - \varepsilon$, when $f(x) < \infty$ (and $f(x') \to \infty$ as $x' \to x_0$ if $f(x) = \infty$). Equivalently, $\liminf_{x\to x_0} f(x) \geq f(x_0)$. Change to $f(x') \leq f(x_0) + \varepsilon$ for u.s.c.

Lemma. Let μ be a dominating measure of P and Q, and $p = \mathrm{d}P/\mathrm{d}\mu$, $q = \mathrm{d}Q/\mathrm{d}\mu$. Then TFAE:
(i) $P << Q$
(ii) $P(q = 0) = 0$
(iii) $\int p/q\mathrm{d}Q = 1$

Def. Let $\mathrm{d}P/\mathrm{d}Q = p/q$ if $q > 0$ and $= 0$, otherwise. In general $\int hdP \geq \int h\frac{dP}{dQ}dQ$, with equality if $P \ll Q$.

Le Cam's first lemma. Let (P_n, Q_n) be prob measures on $(\mathcal{X}_n,\mathcal{F}_n)$. The following are equivalent:
(i) $P_n \triangleleft Q_n$
(ii) $\frac{\mathrm{d}Q_n}{\mathrm{d}P_n} \underset{P_n}{\xrightarrow{d}} U$ along a subsequence, then $\Pr(U = 0) = 0$
(iii) If $\frac{\mathrm{d}P_n}{\mathrm{d}Q_n} \underset{Q_n}{\xrightarrow{d}} V$ along a subsequence, then $\mathbb{E}V = 1$.

Remark. If $\frac{\mathrm{d}Q_n}{\mathrm{d}P_n} \underset{P_n}{\xrightarrow{d}} U$, such that $\Pr(U = 0) = 0$ and $\mathbb{E}U = 1$, then $P_n \triangleleft\triangleright Q_n$.

Cor. Suppose $\frac{\mathrm{d}Q_n}{\mathrm{d}P_n} \underset{P_n}{\xrightarrow{d}} e^{N(\mu,\sigma^2)}$ such that $\mu + \frac{\sigma^2}{2} = 0$. Then $P_n \triangleleft\triangleright Q_n$.

Remark. If $\frac{\mathrm{d}Q_n}{\mathrm{d}P_n} \underset{P_n}{\xrightarrow{d}} e^{N(\mu,\sigma^2)}$ and $P_n \triangleleft Q_n$, then $\mu + \frac{\sigma^2}{2} = 0$.

Le Cam's third lemma. Let $P_n \triangleleft Q_n$. Assume $\left(X_n, \frac{\mathrm{d}P_n}{\mathrm{d}Q_n}\right) \underset{Q_n}{\xrightarrow{d}} (X, R)$ with distribution $F_{X,R}(x,r) = P(X \leq x, R \leq r)$, then $\left(X_n, \frac{\mathrm{d}P_n}{\mathrm{d}Q_n}\right)$converges in distribution under P_n and $\mathbb{E}_{P_n}f(X_n, \mathrm{d}P_n/\mathrm{d}Q_n) \to \mathbb{E}\{Rf(X,R)\}$, $\forall f$ bounded cts.

Corollary. Assume $\left(X_n, \log\frac{\mathrm{d}P_n}{\mathrm{d}Q_n}\right) \underset{Q_n}{\xrightarrow{d}} (X, Z) \sim N\left(\begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix}, \begin{bmatrix}\sigma_1^2 & \sigma_{12}\\ \sigma_{12} & \sigma_2^2\end{bmatrix}\right)$ where $\sigma_{12} = \sigma_1\sigma_2\rho$, such that $\mu_2 + \frac{\sigma_2^2}{2} = 0$, then $X_n \underset{P_n}{\xrightarrow{d}} N(\mu_1 + \sigma_{12}, \sigma_1^2)$.

Remark. The same holds with vector-valued R.V. $\mathbf{X}_n$. Note that in this case, μ_1 would be a vector, σ_1^2 would be a matrix, and σ_{12} would ve a vector.

Corollary. Under previous corollary, we also have jointly:
$(X_n, \log\frac{dP_n}{dQ_n}) \xrightarrow[P_n]{d} N\left(\begin{bmatrix}\mu_1+\sigma_{12}\\ \mu_2+\sigma_2^2\end{bmatrix}, \begin{bmatrix}\sigma_1^2 & \sigma_{12}\\ \sigma_{12} & \sigma_2^2\end{bmatrix}\right)$ (hw4 q6)

Definition (LAN). Let Θ be open. For every θ_0, let $P^n_{\theta_0}$ be a prob measure on $(\mathcal{X}_n, \mathcal{F}_n)$. LAN holds at θ_0 if there exists a positive sequence $\{\phi_n\}_{n\geq 1}$ converging to 0, s.t $\forall h$ fixed, $\log\frac{dP^n_{\theta_0+h\phi_n}}{dP^n_{\theta_0}} = h\Delta_n - \frac{h^2}{2}I(\theta_0) + \varepsilon_n(h)$, for some $I(\theta_0) > 0$, s.t.
(i) $\Delta_n \xrightarrow[P^n_{\theta_0}]{d} N(0, I(\theta_0))$
(ii) $\varepsilon_n \xrightarrow[P^n_{\theta_0}]{p} 0$.

Remark. LAN $\implies P^n_{\theta_0+\frac{h}{\sqrt{n}}} \triangleleft\triangleright P^n_{\theta_0}$.

Remark. If in IID set-up $A_0 - A_4$ hold, LAN holds with $\phi_n = 1/\sqrt{n}$.

Thm. Suppose LAN holds at θ_0, for all $\theta_0 \in \Theta$. If $\frac{T_n-\theta_0}{\phi_n} \xrightarrow[P^n_\theta]{d} N(0, \sigma^2(\theta))$, $\forall\theta_0 \in \Theta$ then $\sigma^2(\theta_0) \geq 1/I(\theta_0)$ for a.e. θ_0 (under Lebesgue measure).

Cor. If $\sigma(\theta)$ and $I(\theta)$ are both continuous, then $\sigma^2(\theta_0) \geq 1/I(\theta_0)$, $\forall\theta_0 \in \Theta$.

Lemma. Suppose LAN holds at θ_0, $\forall\theta_0 \in \Theta$. Let T_n be s.t. $\frac{T_n-\theta_0}{\phi_n} \xrightarrow[P^n_{\theta_0}]{d} N(0, \sigma^2(\theta_0))$, and $\liminf_{n\to\infty} P^n_{\theta_0+\phi_n}(T_n \leq \theta_0 + \phi_n) \leq 1/2$.
Then $\sigma^2(\theta_0) \geq \frac{1}{I(\theta_0)}$.

Theorem. Suppose LAN holds at θ_0. Let T_n be a sequence of rv's, such that $T_n \xrightarrow{d} G_h$ under $P^n_{\theta_0+h\phi_n}$, $\forall h$ fixed. Then $G_h \overset{d}{=} F(Z, U)$ where $Z \sim N(h, I(\theta_0)^{-1})$, $U \sim \mathrm{U}(0,1)$. Also Z and U are independent, and F is a non-random measurable function free of h.

Theorem. Suppose LAN holds at θ_0. Let ψ_n be a sequence of asymptotically level α tests for $\theta = \theta_0$ vs. $\theta > \theta_0$ i.e. $\limsup_{n\to\infty}\mathbb{E}_{\theta_0}\psi_n \leq \alpha$. Then $\forall h > 0$, $\lim_{n\to\infty}\sup \mathbb{E}_{\theta_0+h\phi_n}\psi_n \leq 1 - \Phi(z_{1-\alpha} - h/\sqrt{I(\theta_0)})$.
Pf: On subsequence, $\limsup E_{\theta_0+hr_n}\phi_n = \lim E_{\theta_0+hr_{n_k}}\phi_{n_k}$. On further subsequence, $(\phi_n, \frac{dP_{\theta_0+hr_n}}{dP_{\theta_0}}) \xrightarrow[P_{\theta_0}]{d} (V, R)$ by jt. tightness. $\therefore \phi_n \xrightarrow[P_{\theta_0+hr_n}]{d} V(h)$ (le Cam). By thm, $V(h) = F(Z,U)$, $Z \sim N(h, H(\theta_0)^{-1}), U \sim U(0,1)$. Also $E_{\theta_0+hr_n}\phi_n \to EF(U,V)$ (UI), so $E_{h=0}F(Z,U) \leq \alpha$. Now compare $F(U,V)$ to MP test □

Remark. A test that achieves this bound is locally asymptotically optimal.

Lemma. (i) Given a real-valued r.v X, there is a non-random measurable function F such that $X \overset{d}{=} F(U)$, $U \sim \mathrm{U}(0,1)$.
(ii) Given real-valued r.v.s (X, Y), there is non-random measurable F s.t. $(X, Y) \overset{d}{=} (X, F(X,U))$, and $X \perp U$.

PROJECTIONS

Def (Projection). Let $(\Omega, \mathcal{F}, P)$ be a prob space. Let $\mathcal{L}^2$ be the vector space of all r.v.'s X in this space such that $\mathbb{E}X^2 < \infty$. $\hat{X}$ is the the projection of $X \in \mathcal{L}^2$ onto the sub-vector space S if
(i) $\hat{X} \in S$
(ii) $\mathbb{E}(X - \hat{X})^2 \leq \mathbb{E}(X - Y)^2, \forall Y \in S$.

Prop. (i) $\hat{X} \in S$ is a projection iff $\mathbb{E}(X - \hat{X})Y = 0$, $\forall Y \in S$.
(ii) Projection, if it exists, is unique.
(iii) If $1 \in S$, then $\mathrm{Var}(\hat{T}) \leq \mathrm{Var}(T)$ and $\mathbb{E}(\hat{T}) = \mathbb{E}(T)$

Def. S is *closed* if $\{Y_n\}_{n\geq 1} \in S$ and $E(Y_n - Y)^2 \to 0$ implies $Y \in S$.

Prop. If S is closed, then a projection exists.

Remark. Let S be the space of all X such that $\mathbb{E}X^2 < \infty$ and X is $\mathcal{G}$-measurable, where $\mathcal{G} \subseteq \mathcal{F}$. Then $\hat{X} = \mathbb{E}[X|\mathcal{G}]$.

Lemma (Hájek Projection). Let $X_1, ..., X_n$ be independent, and let S be the set of all rv's of the form $\sum_{j=1}^n g_j(X_j)$ where $\mathbb{E}g_j(X_j)^2 < \infty$ (equivalently those of the form $\sum_{i=1}^n Y_j$, where $EY_j^2 < \infty$, Y_j is X_j-measurable).
If $T \in \mathcal{L}^2$, its projection is $\hat{T} = \sum_{i=1}^n \mathbb{E}(T|X_i) - (n-1)\mathbb{E}T$.

Remark. In general $E[T|X_j]$ will depend on j. However, if T is symmetric in $(X_1, \cdots, X_n)$, and $(X_1, \cdots, X_n)$ are independent, then $\mathbb{E}[T|X_j]$ does not depend on j, i.e. $\mathbb{E}[T|X_j] = g(X_j)$, for some function g free of j.

Thm. Let $(\Omega_n, \mathcal{F}_n, P_n)$ be a prob space for each n, and let S_n, with $1 \in S_n$, be a subspace of $\mathcal{L}^2(\Omega_n, \mathcal{F}_n, P_n)$ for each n. Suppose $T_n \in \mathcal{L}^2$ has a projection $\hat{T}_n$, such that $\frac{\mathrm{Var}(T_n)}{\mathrm{Var}(\hat{T}_n)} \xrightarrow[n\to\infty]{} 1$. Then $\frac{T_n - \mathbb{E}T_n}{\sqrt{\mathrm{Var}(T_n)}} - \frac{\hat{T}_n - \mathbb{E}\hat{T}_n}{\sqrt{\mathrm{Var}(\hat{T}_n)}} \xrightarrow[n\to\infty]{\mathcal{L}^2/p} 0$.

Setup (U-Stats). Suppose $(X_1, \cdots, X_n)$ are iid cts rvs on $\mathcal{X}$. Let $h: \mathcal{X}^k \to \mathbb{R}$ be a measurable function. Want to estimate $\theta := \mathbb{E}h(X_1, \cdots, X_k)$, and $h(X_1, \cdots, X_k)$ is an unbiased estimator.
Define $U := \mathbb{E}[h(X_1, \cdots, X_k)|X_{(1)}, \cdots, X_{(n)}]$. Then $\mathbb{E}U = \theta$ and $Var(U) \leq Var(h(X_1, \cdots, X_k))$ as U is a projection (or by Rao-Blackwell).
WLOG assume h is symmetric in its arguments, so that $U = \frac{1}{\binom{n}{k}}\sum_{1\leq l_1<\cdots<l_k\leq n} h(X_{l_1}, \cdots, X_{l_k}) = \frac{1}{\binom{n}{k}}\sum_{\mathbf{i}\in[\binom{n}{k}]} h(X_{\mathbf{i}})$.

Prop. (i) $\mathbb{E}U = \theta$.
(ii) $Var(U) = \sum_{c=1}^k \binom{k}{c}\binom{n-k}{k-c}\xi_c / \binom{n}{k}$, where $\xi_c = Cov(h(X_{i_1}, \cdots, X_{i_k}), h(X_{j_1}, \cdots, X_{j_k}))$, where $|\{i_1, \cdots, i_k\} \cap \{j_1, \cdots, j_k\}| = c$.

Remark. If $\xi_1 \neq 0$, $Var(U) \sim \frac{k\binom{n-k}{k-1}}{\binom{n}{k}}\xi_1 \sim \frac{k^2}{n}\xi_1$, since $\binom{n}{r} \sim \frac{n^r}{r!}$ for r fixed, $n \to \infty$.

Thm. If $\mathbb{E}h^2(X_1, \cdots, X_k) < \infty$, then $\sqrt{n}(U_n - \theta) \xrightarrow[n\to\infty]{d} N(0, k^2\xi_1)$, provided $\xi_1 \neq 0$. Moreover:
- The Hájek projection of $U_n - \theta$ is $\hat{U}_n = \frac{k}{n}\sum_{i=1}^n g(X_i)$, where $g(x) = \mathbb{E}[h(x, X_2, \cdots, X_k) - \theta]$.
- $\mathrm{Var}(g(X_1)) = \xi_1$.

Setup (2-sample U stats). Suppose $X_1, \cdots, X_m \overset{iid}{\sim} F$ and $Y_1, \cdots, Y_n \overset{iid}{\sim} G$.
Let $U_{m,n} = \frac{1}{\binom{m}{r}\binom{n}{s}}\sum_{\mathbf{i}\in[\binom{m}{r}],\mathbf{j}\in[\binom{n}{s}]} h(X_{\mathbf{i}}, Y_{\mathbf{j}})$, where $h: \mathcal{X}^r \times \mathcal{Y}^s \to \mathbb{R}$. Also assume h is symmetric between X variables with Y fixed, and viceversa, i.e. $h(X_{\pi(\mathbf{i})}, Y_{\pi(\mathbf{j})}) = h(X_{\mathbf{i}}, Y_{\mathbf{j}})$. We assume $N = m + n \to \infty$ s.t. $\frac{m}{N} \to \lambda$, $\frac{n}{N} \to 1 - \lambda$, for some $\lambda \in (0,1)$. Let $\theta = \mathbb{E}h(X_{\mathbf{i}}, Y_{\mathbf{j}})$.

Thm. If $\mathbb{E}h^2(X_{\mathbf{i}}, Y_{\mathbf{j}}) < \infty$, then
$\sqrt{N}(U_{m,n} - \theta) \xrightarrow{d} N(0, \frac{r^2}{\lambda}\xi_{1,0} + \frac{s^2}{1-\lambda}\xi_{0,1})$, where $\xi_{1,0} = Cov(h(X_{\mathbf{i}}, Y_{\mathbf{j}}), h(X_{\mathbf{i}'}, Y_{\mathbf{j}'}))$, where $|\mathbf{i} \cap \mathbf{i}'| = 1, |\mathbf{j} \cap \mathbf{j}'| = 0$.
- The Hájek projection of $U_{m,n} - \theta$ is
$\hat{U}_{m,n} = \frac{r}{m}\sum_{i=1}^m g_{1,0}(X_i) + \frac{s}{n}\sum_{j=1}^n g_{0,1}(Y_j)$, where
$g_{1,0}(x) = \mathbb{E}h(x, X_2, \cdots, X_r, Y_1, \cdots, Y_s) - \theta$,
$g_{0,1}(y) = \mathbb{E}h(X_1, \cdots, X_r, y, Y_2, \cdots, Y_s) - \theta$.
- $\mathrm{Var}(g_{1,0}(X_1)) = \xi_{1,0}$, $\mathrm{Var}(g_{0,1}(Y_1)) = \xi_{0,1}$

DISTRIBUTIONAL RESULTS

- $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$
- $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$
- $\Gamma(k) = (k-1)!$, for $k \in \mathbb{Z}^+$.
- $\Gamma(\frac{1}{2}) = \sqrt{\pi}$
- If $a_n \to a$, then $(1 + \frac{a_n}{n})^n \to e^a$

• If $X \geq 0$, then $\mathbb{E}[X] = \int_0^\infty P(X > x)dx$

• Suppose $X_i \sim N(\theta, \sigma^2)$:
- $E(\sum X_i) = n\theta$
- $E(\sum X_i^2) = n\sigma^2 + n\theta^2$
- $E((\sum X_i)^2) = n^2\sigma^2 + n^2\theta^2$
- $(n-1)S^2 = \sum(X_i - \overline{X})^2 \sim \sigma^2\chi^2_{n-1}$
- $\frac{\overline{X}-\mu}{\sqrt{S^2/n}} \sim t_{n-1}$
- $E(\frac{1}{\sum X_i^2}) = \frac{1}{\sigma^2(n-2)}$
- MLE is $(\overline{X}, \frac{1}{n}\sum(X_i - \overline{X})^2)$

• **Def (Sample variance).** $s^2 := \frac{1}{n-1}\sum(x_i - \overline{x})^2$

• $\sum(x_i - \overline{x})^2 = \sum x_i^2 - n\overline{x}^2$

• $\sum(X_i - \mu)^2 = n(\overline{X} - \mu)^2 + \sum(X_i - \overline{X})^2$

• $\text{Var}(\sum_i X_i) = \sum_i \text{Var}(X_i) + 2\sum_{i<j}\text{Cov}(X_i, X_j)$

• $\chi^2_k = Gamma(\alpha = \frac{k}{2}, \beta = \frac{1}{2})$

• $Exp(\lambda) = Gamma(\alpha = 1, \beta = \lambda)$

• If $U \sim U(0,1)$, then $-\log(U) = Exp(1)$

• If $X_i \overset{iid}{\sim} U(0,\theta)$, then $n(1 - \frac{X_{(n)}}{\theta}) \overset{d}{\to} Exp(1)$. In particular, $X(n) \overset{p}{\to} \theta$.

• If $X_i \overset{iid}{\sim} Bin(1, \theta/n)$, then $\sum_{i=1}^n X_i \overset{d}{\to} Poisson(\theta)$.

• If $X_n \sim Bin(n, p_n)$ and $np_n \to \lambda$, then $X_n \overset{d}{\to} Pois(\lambda)$

• If $X \sim P_0(\lambda)$ and $Y \sim P_0(\mu)$ independently, then $X + Y \sim P_0(\lambda + \mu)$

• If $X \sim Gamma(\alpha, \theta)$ and $Y \sim Gamma(\beta, \theta)$ independently, then $X + Y \sim Gamma(\alpha + \beta, \theta)$, and $\frac{X}{X+Y} \sim Beta(\alpha, \beta)$.

• If $X \sim Gamma(\alpha, \beta)$, then $\sigma X \sim Gamma(\alpha, \beta/\sigma)$.

• If $X_1, X_2 \overset{iid}{\sim} N(\theta, 1)$,
then $X_1|\{X_1 + X_2 = t\} \sim N(t/2, 1/2)$.

• If $X_1, \cdots, X_n \overset{iid}{\sim} N(\theta, 1)$, $T = \sum X_i$, then

$$(X_1, \cdots, X_n | T = t) \sim N(\begin{pmatrix} t/n \\ ... \\ t/n \end{pmatrix}, \begin{pmatrix} 1 - \frac{1}{n} & -\frac{1}{n} & \cdots \\ -\frac{1}{n} & 1 - \frac{1}{n} & \cdots \\ \cdots & \cdots & \cdots \end{pmatrix})$$

• If $X \sim Pois(\lambda)$, $Y \sim Pois(\mu)$ independently, then $X|\{X + Y = t\} \sim Bin(t, \frac{\lambda}{\lambda+\mu})$.

• **MVN** (Multi-variate normal). If $\mathbf{X} \sim N(\mu, \Sigma)$, then
- $f(\mathbf{x}) = (2\pi|\det\Sigma|)^{-n/2}\exp(-\frac{1}{2}(\mathbf{x} - \mu)^T\Sigma^{-1}(\mathbf{x} - \mu))$.
- $\mathbb{E}e^{\mathbf{v^t X}} = e^{\mathbf{v^t}\mu + \frac{1}{2}\mathbf{v^t \Sigma v}}$.

• In particular, in the bivariate case with correlation ρ, $f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \times$
$\exp\left(-\frac{1}{2(1-\rho^2)}\left[(\frac{x-\mu_x}{\sigma_x})^2 - 2\rho(\frac{x-\mu_x}{\sigma_X})(\frac{y-\mu_U}{\sigma_Y}) + (\frac{y-\mu_Y}{\sigma_Y})^2\right]\right)$

• In the standardized case with correlation ρ, (i.e. $X, Y \sim N(0,1)$, $EXY = \rho$), we have $Y = \rho X + \sqrt{1-\rho^2}Z$, where $Z \perp X$.

• If $U \sim N(0,1)$ and $V \sim \chi^2_p$ independently, then $\frac{U}{\sqrt{V/p}} \sim t_p$

• If $U \sim \chi^2_p$ and $V \sim \chi^2_q$ independently, then $\frac{U/p}{V/q} \sim F_{p,q}$

Order Statistics. If $X_1, \cdots, X_n \overset{iid}{\sim} f(x)$, then
• $f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!}f(x)F(x)^{j-1}(1 - F(x))^{n-j}$
• $F_{X_{(j)}}(x) = \sum_{k=j}^n \binom{n}{k}F(x)^k(1 - F(x))^{n-k}$
• $f_{X_{(i)},X_{(j)}}(u, v) = \frac{n!}{(i-1)!(j-1-i)!(n-j)!} \times f(u)f(v) \times F(u)^{i-1}(F(v) - F(u))^{j-1-i}(1 - F(v))^{n-j}$, for $u < v$, $i < j$
• $f_{X_{(1)},\cdots,X_{(n)}}(\mathbf{x}) = n!f(x_1)\cdots f(x_n)$, for $x_1 < \cdots < x_n$
• If $U_1, \cdots, U_n \overset{iid}{\sim} U[0,1]$, then $U_{(k)} \sim Beta(k, n - k + 1)$
• The conditional distribution of $X_{(i)}|X_{(j)} = t$ is that of the ith order statistic from $j - 1$ samples of the original distribution truncated at t.
• $(X_1|X_{(n)} = t) \overset{d}{=} \frac{1}{n}\delta_t + \frac{n-1}{n}U(0,t)$ (HW2 Q4)
• Order statistics are independent of rank statistics

Propn (Asymptotic distribution of ordered statistics). If $X_1, ..., X_n$ are i.i.d from continuous strictly positive density f, then, for $p \in (0,1)$,
$\sqrt{n}(X_{(\lceil np \rceil)} - F^{-1}(p)) \overset{\mathcal{D}}{\to} N\left(0, \frac{p(1-p)}{f_X(F^{-1}(p))^2}\right)$

• If $X_1, \cdots, X_n$ have continuous cdf F, then $F(X_1), \cdots, F(X_n) \sim U[0,1]$, and if $U_1, \cdots, U_n \sim U[0,1]$, then $F^{-1}(U_1), \cdots, F^{-1}(U_n) \overset{d}{=} X_1, \cdots, X_n$.

• If $X_1, \cdots, X_n \overset{iid}{\sim} N(\theta, \sigma^2)$ and $\theta \sim N(\mu, \tau^2)$, then
- $\theta|\mathbf{X} \sim N(\frac{\mu\sigma^2 + n\tau^2\overline{X}}{\sigma^2 + n\tau^2}, \frac{\sigma^2\tau^2}{\sigma^2 + n\tau^2})$
- $\mathbf{X} \sim N(\mu\mathbf{1}, \sigma^2\mathbf{I}_n + \tau^2\mathbf{1}\mathbf{1}^T)$ (marginally) (HW3 q5)

• If $X_1, \cdots, X_n \sim B(1,p)$ and $p \sim B(\sqrt{n}/2, \sqrt{n}/2)$, then $\delta(X) = \frac{\sum X_i + \sqrt{n}/2}{n + \sqrt{n}}$ is the unique Bayes estimator. It has constant risk $\frac{1}{4(1+\sqrt{n})^2}$, so it's unique minimax and L.F.

• MLE for Normal, Poisson, and Bernoulli is $\bar{X}$. For uniform it is $X_{(n)}$.

• Cauchy Distribution verifies conditions A3 and A4.

• If X is negative binomial (r, p), and $Y = 2pX$, then $Y \overset{d}{\to} \chi^2_{2r}$ as $p \to 0$.

• If $X \sim Gamma(\alpha, \beta)$ and $Y \sim Poisson(x\beta)$, then $P(X \leq x) = P(Y \geq \alpha)$.

• If $X \sim Bin(m, p)$, $Y \sim Bin(n, p)$ independently, then $P(X = k|X + Y = t) = \frac{\binom{m}{k}\binom{n}{t-k}}{\binom{m+n}{t}}$ (HyperGeometric)

INEQUALITIES

Triangle: $\Big|||x|| - ||y||\Big| \leq ||x + y|| \leq ||x|| + ||y||$

• $||f||_p = \left(\int |f|^p d\mu\right)^{\frac{1}{p}}$ or $||X||_p = \left(E|X|^p\right)^{\frac{1}{p}}$ are norms

Holder's: Suppose $p, q \in [1, \infty]$ s.t. $\frac{1}{p} + \frac{1}{q} = 1$. Then $||fg||_1 \leq ||f||_q||g||_p$. In particular,
• $\int |f(x)g(x)|dx \leq \left(\int |f(x)|^p dx\right)^{\frac{1}{p}}\left(\int |g(x)|^q dx\right)^{\frac{1}{q}}$
• $E|XY| \leq (E|X|^p)^{1/p}(E|Y|^q)^{1/q}$

Cauchy-Schwarz. Setting $p = q = 2$ in Holder's,
• $E|XY| \leq \sqrt{EX^2EY^2}$
• $Cov(X,Y)^2 \leq Var(X)Var(Y)$, with $=$ iff $Y = aX + b$

Pinsker's: $||P - Q||_{TV} \leq \sqrt{2D_{KL}(P||Q)}$.

Markov's: $P(|X| \geq M) \leq \frac{E|X|}{M}$

Jensen's: Under *UNBIASEDNESS*.

Cosh. $\cosh(x) = \frac{e^x + e^{-x}}{2} \leq e^{x^2/2}$

Log. $\log(1 + x) \leq x - \frac{x^2}{2}$ if $x \geq 0$ (Taylor expansion)
• $\log(1 + x) \leq x - 2\frac{x^2}{2}$ if $x \geq -0.5$
• $\log(1 + x) \geq x - \frac{x^2}{2} + \frac{x^3}{4}$ iff $x \in [0, 0.45...]$ ($\leq$ elsewhere)
• $\log(1 + x) \geq x - \frac{x^2}{2} + \frac{x^3}{2}$ iff $x \in [-0.43, 0]$ ($\leq$ elsewhere)

MISCELLANEOUS

Sterling's Approx. $n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$.

O notation.
• $f(x) = o(g(x))$ as $x \to \infty$ iff $\frac{|f(x)|}{g(x)} \to 0$ as $x \to \infty$.
• $X_n = o_p(a_n)$ if $X_n/a_n \overset{p}{\to} 0$.
• $f(x) = O(g(x))$ as $x \to \infty$ iff $\exists x_o, M$ such that $|f(x)| < Mg(x)$ for all $x > x_0$.
• $X_n = O_p(a_n)$ if X_n/a_n is stochastically bounded, i.e. $\forall\varepsilon > 0$ $\exists M, N$ s.t. $P(|X_n| \geq Ma_n) < \varepsilon$ for all $n \geq N$.

Thm (joint convergence).
• Suppose $X_n \overset{p}{\to} X$ and $Y_n \overset{p}{\to} Y$. Then $(X_n, Y_n) \overset{p}{\to} (X, Y)$.

• Suppose $X_n \xrightarrow{\mathcal{D}} X$ and $Y_n \xrightarrow{\mathcal{D}} Y$, and X_n is independent of Y_n for all n. Then $(X_n, Y_n) \xrightarrow{\mathcal{D}} (X, Y)$.
• $(X_n, Y_n) \xrightarrow{d} (X, Y)$ iff $\forall k_1, k_2 \in \mathbb{R}, k_1X_n + k_2Y_n \xrightarrow{d} k_1X + k_2Y$. $\mathbf{X}_n \xrightarrow{d} \mathbf{X}$ iff $\langle \mathbf{t}, \mathbf{X}_n\rangle \xrightarrow{d} \langle \mathbf{t}, \mathbf{X}\rangle, \forall \mathbf{t} \in \mathbb{R}^d$.

Thm (Cts. Mapping). If f is cts. and $X_n \to X$, then $f(X_n) \to f(X)$ (holds for convergence a.s., in $\mathbb{P}$ or in $\mathcal{D}$)

Thm (Continuous Mapping). Let g be a function, such that the set of discontinuity points has prob. measure 0. Then
• $X_n \to X$ implies $g(X_n) \to g(X)$ for convergence in distribution, prob. and a.s. respectively.

Def (L_p convergence). $X_n \xrightarrow{L_p} X$ if $E|X_n - X|^p \to 0$
• For $s \geq r \geq 1$, $X_n \xrightarrow{L_s} X \implies X_n \xrightarrow{L_r} X$ (Jensen's)
• For $p \geq 1$, $X_n \xrightarrow{L_p} X \implies X \xrightarrow{p} X$
• If X_n is UI and $X \xrightarrow{p} X$, then $X_n \xrightarrow{L_1} X$
• If $X_n \xrightarrow{L_p} X$, then $EX_n^p \to EX^p$ (reverse Δ inequality)

Uniform Integrability. A sequence $(X_n)_{n\geq 1}$ is UI if $\forall \varepsilon > 0, \exists M > 0$ s.t. $\sup_{n\geq 1} \mathbb{E}|X_n|I_{(|X_n|>M)} < \varepsilon$
• If $X_n \xrightarrow{\mathcal{D}} X$ and $\sup_{n\geq 1} \mathbb{E}[|X_n|^{1+\delta}] < \infty$ for some $\delta > 0$, then $\mathbb{E}X_n \to \mathbb{E}X$.

Tightness. We say $\{V_n\}_{n\geq 1}$ is tight if given $\varepsilon > 0$, $\exists K_\varepsilon < \infty$ such that $P(V_n \in [-K_\varepsilon, K_\varepsilon]) \geq 1 - \varepsilon, \forall n \geq 1$.
Alternatively, $\sup_{n\geq 1} P(|V_n| > M) \to 0$ as $M \to \infty$.
Also written as $V_n = O_p(1)$ or 'bounded in probability'.
• Marginal tightness implies joint tightness. This in turn implies convergence in distribution along a subsequence.
• If $X_n \xrightarrow{d} X$, then $\{X_n\}_{n\geq 1}$ is tight.
• If $\{X_n\}_{n\geq 1}$ is UI, then $\{X_n\}_{n\geq 1}$ is tight.

Def $(X_n)_{n\geq 1}$ is bounded in L_p for $p \geq 1$ if $\sup_{n\geq 1} \mathbb{E}[|X_n|^p] < \infty$.
• For $p \geq 1$, this implies tightness.
• For $p > 1$, this implies UI. (Counterexample for $p = 1$; $X_n = nI(0, 1/n]$). Conversely, UI $\implies$ bounded in L_1 (but NOT bounded in L_p for $p > 1$).

Prohorov's thm. If V_n is tight, there exists a subsequence along which it converges in distribution.

Lagrange Multipliers. Let $f : \mathbb{R}^d \to \mathbb{R}$, $h = (h_1, \cdots, h_k)^T$, $h_i : \mathbb{R}^d \to \mathbb{R}$, $f, h \in \mathcal{C}^1$. Let $\mathcal{L}(x, \lambda) = f(x) - \langle \lambda, h(x)\rangle$. If $\exists(x^*, \lambda^*)$ s.t.
i) $\mathcal{L}(x^*, \lambda^*) = \max_{x\in\mathbb{R}^d} \mathcal{L}(x, \lambda^*)$
ii) $h(x^*) = 0$
Then x^* maximizes $f(x)$ subject to $h(x) = 0$. Therefore:
1. Maximize $\mathcal{L}(x, \lambda)$ in x to find $x^*(\lambda)$.
2. Find λ^* s.t. $x^*(\lambda^*)$ satisfies $h(x^*) = 0$.

KKT. Consider $\max_{x\in\mathbb{R}^d} f(x)$ subject to $h(x) = 0$ and $g(x) \leq 0$, $g = (g_1, \cdots, g_m)^T$ $g_i \leq 0$. Let $\mathcal{L}(x, \lambda, \mu) = f(x) - \langle \mu, g(x)\rangle - \langle \lambda, h(x)\rangle$. If x^* is a solution, $\exists \lambda^*, \mu^*$ s.t.
Stationarity: $\nabla\mathcal{L}(x^*, \lambda^*, \mu^*) = 0$
Primal feasibility: $g_i(x^*) \leq 0$, $h_i(x^*) = 0$
Dual feasibility: $\mu_i^* \geq 0$
Complementary slackness: $\mu_i^* g_i(x^*) = 0$

KKT (sufficiency). Consider:
$(*)$ $\min_{x\in\mathbb{R}^d} f(x)$ s.t. $g(x) \leq 0$ and $h(x) = 0$.
Let $\mathcal{L}(x, \lambda, \mu) = f(x) + \langle \lambda, h(x)\rangle + \langle \mu, g(x)\rangle$.
Suppose $\exists(x^*, \lambda^*, \mu^*)$ s.t. $g(x^*) \leq 0$, $h(x^*) = 0$, $\mu^* \geq 0$, $\langle \mu^*, g(x^*)\rangle = 0$ and $\mathcal{L}(x^*, \lambda^*, \mu^*) = \min_{x\in\mathbb{R}^d} \mathcal{L}(x, \lambda^*, \mu^*)$.
Then x^* solves $(*)$. Therefore:
1. Minimize $\mathcal{L}(x, \lambda, \mu)$ in x to find $x^*(\lambda, \mu)$.
2. Maximize $\mathcal{L}(x^*(\lambda, \mu), \lambda, \mu)$ over $\mu \geq 0$ to find $\mu^*(\lambda)$.
3. Find λ^* s.t. $h(x^*(\lambda^*, \mu^*(\lambda^*))) = 0$.
4. Check $\langle \mu^*, g(x^*)\rangle = 0$ (automatic for 'nice' convex problems).

KKT (inequalities only). Consider:
$(*)$ $\min_{x\in\mathbb{R}^d} f(x)$ s.t. $g(x) \leq 0$.
Let $\mathcal{L}(x, \mu) = f(x) + \langle \mu, g(x)\rangle$.
Suppose $\exists(x^*, \mu^*)$ s.t. $g(x^*) \leq 0$, $\mu^* \geq 0$, $\langle \mu^*, g(x^*)\rangle = 0$ and $\mathcal{L}(x^*, \mu^*) = \min_{x\in\mathbb{R}^d} \mathcal{L}(x, \mu^*)$.
Then x^* solves $(*)$. Therefore:
1. Minimize $\mathcal{L}(x, \mu)$ in x to find $x^*(\mu)$.
2. Maximize $\mathcal{L}(x^*(\mu), \mu)$ over $\mu \geq 0$ to find μ^*.
3. Check $\langle \mu^*, g(x^*(\mu^*))\rangle = 0$.

Def (compactness). A set K is compact if every open cover has a finite subcover.
Usually: closed and bounded.

Def (Characteristic function). $\phi_X(u) = \mathbb{E}e^{i\langle u, X\rangle}$

Cumulant generating function: $\log(Ee^{tX})$
• If it exists, it is convex and infinitely differentiable.

Weighted loss. If $L(g(\theta), \delta(X)) = w(\theta)(\delta(X) - g(\theta))^2$, the Bayes estimator is $\delta_0(X) = \frac{\mathbb{E}[\theta w(\theta)|X]}{\mathbb{E}[w(\theta)|X]}$.
An admissible estimator w.r.t sq. err. is also admissible w.r.t. weighted loss.

Absolute error loss. If $L(g(\theta), \delta(X)) = |\delta(X) - g(\theta)|^2$, the Bayes estimator is $\delta_0(X) = \text{median}(\theta|X)$.

0-1 loss. If $L(g(\theta), \delta(X)) = I(\delta(X) \neq g(\theta))$, the Bayes estimator is $\delta_0(X) = \text{mode}(\theta|X)$.

Scaled/shifted Bayes/minimax. If $\delta(X)$ is Bayes/minimax for $g(\theta)$, then $a\delta(X)+b$ is Bayes/minimax for $ag(\theta) + b$.

• Under sq. err. loss, $aX + b$ is inadmissible for EX if:
  - $a > 1$ (dominated by X)
  - $a < 0$ (dominated by $-\frac{b}{a-1}$)
  - $a = 1, b \neq 0$ (dominated by X)

Cochran's Thm. Suppose $Z \sim N(0, \Sigma)$ and $\Sigma^2 = \Sigma$. Then $Z^TZ \sim \chi^2_{tr(\Sigma)} = \chi^2_{r(\Sigma)}$.

Convexity characterizations. f is convex iff
$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$, $\forall x, y$, $\forall \lambda \in (0, 1)$
iff $\frac{f(x_2)-f(x_1)}{x_2-x_1} \leq \frac{f(x_3)-f(x_2)}{x_3-x_2}$, $\forall x_1 < x_2 < x_3$
iff $\frac{f(x_2)-f(x_1)}{x_2-x_1} \leq \frac{f(x_3)-f(x_1)}{x_3-x_1}$, $\forall x_1 < x_2 < x_3$

Lyapunov CLT. Suppose $X_1, X_2, \cdots$ are independent with means μ_i and variances σ_i^2. Let $s_n^2 = \sum_{i=1}^n \sigma_i^2$. If, for some $\delta > 0$, $\lim_{n\to\infty} \frac{1}{s_n^{2+\delta}} \sum_{i=1}^n E|X_i - \mu_i|^{2+\delta} = 0$ (Lyapunov's condition). Then $\frac{1}{s_n}\sum_{i=1}^n (X_i - \mu_i) \xrightarrow{d} N(0, 1)$.

Binomial theorem. $(x + y)^n = \sum_{i=0}^n \binom{n}{i} x^i y^{n-i}$

Sherman-Morrison (Woodbury) formula.
$(A + uv^T)^{-1} = A^{-1} - \frac{A^{-1}uv^TA^{-1}}{1+v^TA^{-1}u}$.

U stats. $h(x, y) = \frac{1}{2}(x-y)^2 \implies U_n = \frac{1}{n-1}\sum(X_i - \overline{X})^2$

RANDOM FACTS FROM EXERCISES

• $\sum_{i=1}^n i^2 = \frac{1}{6}n(n + 1)(2n + 1)$
• Let $\alpha > 0$. Then $x^\alpha \log(x) \to 0$ as $x \to 0^+$.
• Let $P(X_i = \pm 1) = \frac{1}{2}$, $S_n = \sum_{i=1}^n X_i$.
Then $Ee^{\lambda S_n} = (\cosh(\lambda))^n$ and $P(|S_n| > nt) \leq 2e^{-nt^2/2}$
• Suppose $P_{n,\beta}(Y_i = y_i) = \frac{1}{Z_n(\beta)} \exp(\frac{\beta}{n-1} \sum_{1\leq i<j\leq n} y_iy_j)$.
Then $\frac{Z_n(\beta)}{2^n} \to \exp(-\frac{\beta}{2})(1 - \beta)^{-\frac{1}{2}}$ (HW4 Q2)
Also $\sqrt{n}\overline{Y} \xrightarrow[P_{n,\beta}]{d} N(0, \frac{1}{1-\beta})$ (HW4 Q3)

• Suppose $X \sim p_\theta$, $\Theta_0 \subseteq \Theta_1$, $\delta_0(X)$ is unique UMVUE for $\theta \in \Theta_0$, and Θ_1 also has a UMVUE, and Θ_0, Θ_1 have the same null sets. If δ_0 is unbiased for Θ_1, then δ_0 is also a UMVUE for $\theta \in \Theta_1$ (midterm 1).

TAYLOR SERIES

• $e^x = \sum_{k=0}^\infty \frac{x^k}{k!} = 1 + x + \frac{x^2}{2!} + \cdots$
• $\log(1 + x) = \sum_{k=0}^\infty (-1)^{k+1}\frac{x^k}{k} = x - \frac{x^2}{2} + \ldots$ for $|x| < 1$
• $\sin(x) = \sum_{k=0}^\infty \frac{(-1)^k}{(2k+1)!} x^{2k+1} = x - \frac{x^3}{3!} + \frac{x^5}{5!} + \cdots$
• $\cos(x) = \sum_{k=0}^\infty \frac{(-1)^k}{(2k)!} x^2k = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} + \cdots$