Measure Theory

Measure Space: $(\Omega,\mathcal{F})$: $\mathcal{F}$ is a $\sigma-$field on Ω s.t. (1) $\Omega\in\mathcal{F}$, (2) $\forall A\in\mathcal{F}\Rightarrow A^c\in\mathcal{F}$, (3) $A_1,A_2,\cdots\in\mathcal{F}\Rightarrow\bigcup_{i=1}^{\infty}A_i\in\mathcal{F}$.

Measure: a measure μ on $\mathcal{F}$ is a non-negative extended real-valued function on $\mathcal{F}$ s.t. any disjointed sets $A_1,A_2,\cdots$, $\mu(\cup A_i)=\sum\mu(A_i)$. $\mu(\Omega)$ can be ∞. If $\mu(\Omega)=1$, it's probability measure and $(\Omega,\mathcal{F},\mu)$ is probability space.

Lebesgue Measure: Consider $\Omega=\mathbb{R}$, $\mathcal{A}=\{(a,b):-\infty<a<b<+\infty\}$ i.e. collection of open intervals. $\mathcal{B}=\sigma(\mathcal{A})$ is called Borel sets. On $\mathcal{B}(\mathbb{R})$, Lebesgue measure is the length of the interval, which can be ∞.

Counting Measure: Let Ω be a countable set, $\mathcal{F}=2^{\Omega}$, and for $A\in\mathcal{F}$, $\mu(A)=|A|$.

Measure Function: s.t. $g(x)=\lim S_n(x)$, where (1)$S_n(x)$ takes finite number of values $\{a_i\}_{i=1}^m$, (2)$\{x: S_n(x)=a_i\}\in\mathcal{F}$, (3)$S_n(x)$ is non-decreasing w.r.t n.

Integration: $\int g d\mu=\lim_n\sum_{i=1}^n a_i\mu(\{S_n(x)=a_i\})$.

Dominated: If $\mu(A)=0\Rightarrow\nu(A)$, $\forall A$, we say $\nu\ll\mu$.

Derivative: $\nu\ll\mu\Rightarrow\exists f$ s.t. $\nu(A)=\int_A f d\mu$. f is the derivative.

Probability Measure: Space $(\Omega,\mathcal{F},\mathbb{P})$, if $\mathbb{P}(\Omega)=1$, $\mathbb{P}$ is a probability measure.

- If $\exists\mu$ s.t. $\mathbb{P}\ll\mu$, then $\exists p\, s.t.\, \mathbb{P}(A)=\int_A p\,d\mu$, p is the density of $\mathbb{P}$ w.r.t. μ.
- A set of prob. measures $\mathcal{P}$ on $(\Omega,\mathcal{F})$. If $\forall\mathbb{P}\in\mathcal{P}$, $\mathbb{P}\ll\mu$, $\to$ a family of p.

Random Variable: A measure function $X:(\Omega,\mathcal{F})\to(E,\mathcal{G})$. E: sample space.

Support: $Supp(\mathbb{P})=\{x:\mathbb{P}[a,b]>0,\ a<x<b\}$. If $\exists p$, it's $\{x:p(x)>0\}$

Convex

Set: A is a convex set if $\forall x,y\in A$, $0<t<1\Rightarrow tx+(1-t)y\in A$.

Function: Real valued function $\phi(x)$ defined over open interval (a,b) is convex if

- $\forall a<x,y<b$, $\phi[rx+(1-r)y]\le r\phi(x)+(1-r)\phi(y)$, $0<r<1$
- If ϕ is differentiable: $\phi'(x)\le\phi\prime(y)$, $\forall a<x<y<b$
- If ϕ is twice differentiable: $\phi''(x)\ge0$, $\forall a<x<b$

Model

Exponential Family: A parametric family $\{P_\theta:\theta\in\Theta\}$ is said to be $s-$dimensional a Exp family if the distributions P_θ have densities of form: $p_\theta(x)=exp\{\sum_{i=1}^s\eta_i(\theta)T_i(x)-A(\theta)\}h(x)$, where $A(\theta)$ is to normalize the density as $A(\theta)=log\{\int h(x)exp[\sum_{i=1}^s\eta_i(\theta)T_i(x)]\}$.

Ex: $N(\mu,\sigma^2)$: $p(x)=exp\{\frac{\mu}{\sigma^2}x-\frac{1}{2\sigma^2}x^2-\frac{\mu^2}{2\sigma^2}-log(\sqrt{2\pi}\sigma)\}$.

If $\mu=1$ is known, η^TT cannot reduce to 1 term, but the dim of η is 1.

Natural Exponential Family: Reparameterize by $\eta=\eta(\theta)$, and there is the canonical form: $p_\eta(x)=exp\{\eta^TT(x)-B(\eta)\}h(x)$. Where η is the nature parameter, and the nature parameter space is $\Xi=\{\eta:\int p_\eta\,d\mu<\infty\}$

- Canonical form is not unique, as we can use $(c\eta,T/c)$ instead.
- Dim of θ and η can be different: $\eta_1=\eta_2=\frac{\theta}{2}$, $d(\theta)=1,d(\eta)=2,d(\Xi)=1$

If $\eta\in\mathbb{R}^s$ and $d(\Xi)=s$ i.e. Ξ contains a $s-$dim. open set, we say it's full rank.

Properties:

If $X_1,\cdots,X_n$ indep. exp. family r.v.s, the joint density is still exp. family.

Decompose: $T=(T_1,\cdots,T_s)=:(Y,U)$, then $Y,Y|U=u$ are still exp. family.

Any integrable function f and interior point $\eta_0\in\Xi$, $\mathbb{E}_\eta f(X)=\int f(x)p_\eta\,d\mu$ is infinitely differentiable w.r.t η in a neighborhood of η_0. The diff. can interchange with int. Ex: take $f=1$, $\frac{\partial}{\partial\eta_i}\mathbb{E}_\eta[1]=0=\mathbb{E}_\eta\{T_i-\frac{\partial}{\partial\eta_i}A(\eta)\}\Rightarrow\mathbb{E}_\eta T_i-\frac{\partial}{\partial\eta_i}A(\eta)$.

Similarly: $\frac{\partial^2}{\partial\eta_i\partial\eta_j}\mathbb{E}_\eta[1]=0\Rightarrow Cov(T_i,T_j)=\frac{\partial^2}{\partial\eta_i\partial\eta_j}A(\eta)$

MGF: Let $u=(u_1,\cdots,u_s)$, $M_T(u)=\mathbb{E}e^{u_1T_1+\cdots+u_sT_s}=e^{A(\eta+u)-A(\eta)}$. We need $\eta+u\in\Xi$, and $\mathbb{E}T_i^n=\frac{\partial^n}{\partial u_i^n}M_y(u)|_{u=0}$.

Cumulate generating function: $K(u)=log\,M(u)=A(\eta+u)-A(\eta)$

Prove NOT Exp family: Take $\frac{1}{2}e^{-|x-\mu|}$ as an example:

Assume it belongs to some exp family, and the jointed PDF of size $n>s$ is: $f(x;\mu)=2^{-n}exp\{-\sum|x_i-\mu|\}=exp\{\sum_{j=1}^s[\eta_j(\mu)\sum_{i=1}^nT_j(x_i)]-nA(\mu)\}$

Remove $h(x)$ by $log[f(x;\mu)/f(x;0)]=\sum|x_i|-\sum|x_i-\mu|=\cdots$ $*$

Note $\psi(x,\mu)=\sum|x_i|-\sum|x_i-\mu|$, $\tilde{\eta}_j(\mu)=\eta_j(\mu)-\eta_j(0)$, $\tilde{A}(\mu)=A(\mu)-A(0)$, $\tilde{T}_j(x)=\sum_{i=1}^nT_j(x_i)$ Above $*$ is $\psi(x,\mu)=\sum_{j=1}^s\tilde{\eta}_j(\mu)\tilde{T}_j(x)-n\tilde{A}(\mu)$.

- If the Exp assumption is correct: if $\exists x,y\,s.t.\,\tilde{T}(x)=\tilde{T}(y)\Rightarrow\forall\mu$ same R.H.S$\Rightarrow$ $\psi(x,\mu)=\psi(y,\mu),\forall\mu$. As a function of μ, $\psi(x,\mu)$ is not differentiable at x_i. Hence, $\tilde{T}(X)=\tilde{T}(Y)\Rightarrow(X_{(1)},\cdots,X_{(n)})=(Y_{(1)},\cdots,Y_{(n)})$.
- However, we can find liner independent $\tilde{\eta}(\mu_1),\cdots,\tilde{\eta}(\mu_s)$ by choosing μ. There is a full rank linear system. So $\psi(x,\mu_j)=\psi(y,\mu_j),\forall j\Rightarrow\tilde{T}(X)=\tilde{T}(Y)$

If we choose $min\{x_i\}>max\{\mu_j\}$ $min\{y_i\}>max\{\mu_j\}$, they can have same ψ. If the assumption is correct $\Rightarrow$ same $\tilde{T}\Rightarrow$ same order statistics, which is not necessary.

Sufficient: $X\sim P\in\mathcal{P}$, $T(X)$ is suff. for P if the distribution of $X|T$ doesn't depend on P. (parametric: not on θ). The family $\mathcal{P}$ or Θ need to be given.

Factorization Thm: $X\sim P\in\mathcal{P}$, and $P\ll\mu$. Then T is suff. iff the density can be written as $\frac{d}{d\mu}P(x)=g_p(T(x))h(x)$, i.e. $f(x)=g(T,\theta)h(x)$.

Not Unique: T is suff. and $\exists h(U)=T$, U is also suff.

Exp Family: T is always suff. by the Factorization Thm.

Minimal Sufficient Statistic: T is MSS iff for any other suff. statistic S exists a measurable function ϕ s.t. $T=\phi(S)$. (Function is always from more to less).

Unique: T_1,T_2 are MSS, by Def there is a $1-1$ mapping between them.

Existence: Usually exists, but exceptions are possible.

Check MSS: (1)Suppose $\mathcal{P}_0\subset\mathcal{P}$ with $a.s.\mathcal{P}_0\Rightarrow a.s.\mathcal{P}$. If T is suff. for $\mathcal{P}$ and MSS for $\mathcal{P}_0$, T is MSS for $\mathcal{P}$.

(2)Suppose $\mathcal{P}$ contains PDF's $f_0,f_1,\cdots$, w.r.t $\sigma-$finite measure μ. Let $f_\infty=\sum_{i=1}^\infty c_if_i$ where $c_i>0$, $\sum c_i=1$ and $T_i(X)=f_i(X)/f_\infty(X)$ where $f_\infty(X)>0$. Then $T=(T_0,T_1,\cdots)$ is MSS. If $\forall i\ge1$: $\{x:f_i(x)>0\}\subset\{x:f_0(x)>0\}$, use f_0 instead of f_∞. $T=(T_1,T_2,\cdots)$ is MSS.

Remark: f_∞ cover the union of the supports, with $\int f_\infty\,d\mu=1$

$\mathcal{P}$ only contains countable f_i. $\Rightarrow$ Choose countable from $\mathcal{P}$, then ues (1).

(3)Suppose $\mathcal{P}$ contains PDF's f_p w.r.t μ and T is suff. s.t. for any possible x,y $f_p(x)=f_p(y)\psi(x,y)$ for all $P\Rightarrow T(x)=T(y)$. Then T is MSS. i.e. $f_p(x)/f_p(y)$ doesn't depends on $p\Leftrightarrow T(x)=T(y)$.

Exp. family: If $\exists\eta_o,\cdots,\eta_s\in\Xi$, $s.t.\eta_1-\eta_0,\cdots,\eta_s-\eta_0$ linear indep. $T(x)-T(y)=0$ is the only root of $\eta^T(T(x)-T(y))=0\Rightarrow T$ is MSS.

Such η exist if it is full rank.

Ancillary: $V(X)$ is ancillary if it's distribution doesn't depends on P.

Complete: $T(X)$ is complete iff any measurable function f:

$\mathbb{E}_P[f(T)]=0\forall P\in\mathcal{P}\Rightarrow f=0\ a.s.\mathcal{P}$

- If T is complete, $S=\psi(T)$ is also complete.
- If T is complete and sufficient, T is MSS: If $\exists t$ is MSS, $t=g(T)$ by definition. Let $h(t)=\mathbb{E}_P(T|t)\to\mathbb{E}_P[h(t)-T]=0$, $\forall P\in\mathcal{P}$. As comp. $T=h(t)\,a.s.\mathcal{P}$. Hence there is $1-1$ mapping between T, t, T is also MSS. □
- Full Rank Exp. family: T is suff. & comp. $\Rightarrow$ MSS: Proof:

T is suff. by Factorization Thm. Let $p_\eta(t)=h(x)e^{\eta^Tt-A(\eta)}$

Suppose f s.t. $\mathbb{E}_\eta[f(T)]=\int f(t)p_\eta(t)\,d\lambda=0$, for all $\eta\in\Xi$.

Let η_0 be a interior point of Ξ, and there is a neighborhood $N_\epsilon(\eta_0)\subset\Xi$, with $\forall\eta\in N_\epsilon(\eta_0)s.t.\mathbb{E}_\eta f_+(T)=\mathbb{E}_\eta f_-(T)$. Let$\mathbb{E}_{\eta_0}f_+(T)=\mathbb{E}_{\eta_0}f_-(T)=c$

If $c=0$ trivial, if $c>0$: $\frac{1}{c}p_{\eta_0}(t)f_+(t)$ and $\frac{1}{c}p_{\eta_0}(t)f_-(t)$ are also PDF.

Let $a=\eta-\eta_0$: $\int e^{at}\frac{p_{\eta_0}(t)f_+(t)}{c}d\lambda=\int\frac{p_\eta(t)f_+(t)}{c}d\lambda=\int\frac{p_\eta(t)f_-(t)}{c}d\lambda=\int e^{at}\frac{p_{\eta_0}(t)f_-(t)}{c}d\lambda$. As such a can cover $N_\epsilon(\eta_0)$, they have same MGF in a neighborhood of $0.\Rightarrow p_{\eta_0}(t)f_+(t)=p_{\eta_0}(t)f_-(t)\,a.s.\Rightarrow f_-=f_+=0\,a.s.$ □

- If not full rank, but exists linear indep.: Check MSS (3)

Basu's Thm: If T is comp.& suff. any ancillary V: $V\perp T$ Proof:

If V is ancillary, $p_A=P[V\in A]$ doesn't depend on $P\in\mathcal{P}$. Let $\eta_A(t)=P[V\in A|T=t]$ also indep. of P. As $\mathbb{E}\eta_A(T)=p_A\Leftrightarrow\mathbb{E}[\eta_A(T)-p_A]=0$, $\forall P$.

As comp. $\eta_A(T)=\mathbb{P}[V\in A|T]=\mathbb{P}[V\in A]=p_A$, $a.s.\mathcal{P}$, i.e. $V\perp T$

Estimator

Point estimator: statistic $T(X)$ estimate τ. (Fun. of para., or non-para. dist.)

Bias: $\mathbb{E}T-\tau$ Unbiased: $\mathbb{E}T=\tau$.

Loss function: $L(\tau,T(X)):\Theta\times\{T(X),X\in\mathbb{R}^n\}\to[0,\infty)$, e.g. $(T-\tau)^2$

Risk function: $R(\tau,T)=\mathbb{E}_P[L(\tau,T)]$ $P\in\mathcal{P}$. Expectation w.r.t. T and P.

Admissibility: T is inadmissible if $\exists$ another estimator $U\,s.t.\,R(\tau,T)\ge R(\tau,U)$ for all $P\in\mathcal{P}$. And "$>$" for some P. If no such U, T is admissible.

UMVUE: Unbiased T of τ is ... if any other unbiased U:$Var\,U\ge Var\,T\,\forall P\in\mathcal{P}$.

Locally MVUE: $Var T\le Var U$ at some fixed $P\in\mathcal{P}$

- May not exist: Ex: $X\sim Binomial(n,\theta)$ $\tau=\frac{1}{\theta}$ If $T(X)$ unbiased: $\mathbb{E}_\theta T(X)=\sum T(k)\binom{n}{k}\theta^k(1-\theta)^{n-k}=\frac{1}{\theta}$ When $\theta\to0$ LHS$\to T(0)$ while RHS$\to\infty$. Hence, no unbiased estimator of τ.
- Estimable: if $\exists$ unbiased estimator of τ, it is called estimable.

Jensen's ineq.: If $\phi(x)$ is a convex function over open interval I, and $\mathbb{P}[X\in I]=1$. Then $\phi(\mathbb{E}X)\le\mathbb{E}[\phi(X)]$. If strictly convex, "$<$", unless $\mathbb{P}[X=c]=1$.

Rao-Blackwell Thm: Let $X\sim P\in\mathcal{P}$, and S is a suff. stat. Given loss fun. $L(\tau,a)$, convex in a for any $P\in\mathcal{P}$. Let $T(X)$ be the estimator of τ with finite risk $R(\tau,T)$, then $U=\mathbb{E}[T|S]$ s.t. $R(\tau,U)\le R(\tau,T)$. Proof: (Jensen's)

$R(\tau,T)=\mathbb{E}_S\{\mathbb{E}_{T|S}[L(\tau,T)|S]\}\ge\mathbb{E}_S[L(\tau,\mathbb{E}(T|S))]=R(\tau,U)$ □

Requirement: Convex loss function: e.g. 0-1 loss is not convex.

Lemman-Scheffe Thm: Suppose S is suff. and comp. for $P\in\mathcal{P}$, and τ estimable

- There is a unique unbiased estimator of the form $h(S)$, h is Borel function.
- $h(S)$ is the unique UMVUE for τ. Proof: estimable$\Rightarrow\exists T\,s.t.\mathbb{E}T=\tau\Rightarrow h(S)=\mathbb{E}(T|S)$. $\mathbb{E}h(S)=\mathbb{E}T=\tau\,\forall P\in\mathcal{P}$ Unbiased✓As S comp, if $\exists\mathbb{E}g(S)=\tau\,\forall P\in\mathcal{P}$,$h(S)=g(S)\,a.s.\mathcal{P}$.Uniqueness✓Any unbiased U, $h(S)=\mathbb{E}[U|S]$ is same(as unique). By Rao-B, $R(\tau,h(S))\le R(\tau,U)$. UMVUE ✓ □

Find UMVUE: get a suff & comp S first then: · Find h s.t. $\mathbb{E}h(S)=\tau$;

· Solve $\mathbb{E}_P(h(S))=\tau\,\forall P\in\mathcal{P}$ directly; · Find unbiased T, $h(S)=\mathbb{E}(T|S)$

EX1:$X_i\sim N(\mu,\sigma^2)\ \mu\in\mathbb{R},\sigma>0$: $T=(\bar{X},S^2)$ is comp & suff for $\theta=(\mu,\sigma^2)$, with $\bar{X}\sim N(\mu,\frac{\sigma^2}{n})$, $\frac{(n-1)S^2}{\sigma^2}\sim\chi^2_{n-1}$ and $\bar{X}\perp\frac{(n-1)S^2}{\sigma^2}$. For μ : $\mathbb{E}\bar{X}=\mu$, it's UMVUE.

For μ^2 : $\mathbb{E}[\bar{X}-\frac{S^2}{n}]=\mu^2$, it's unbiased, and function of T hence UMVUE.

σ^r: Let $Y\sim\chi^2_{n-1}$, $\mathbb{E}S^r=\mathbb{E}\left(\frac{\sigma^2Y}{n-1}\right)^{\frac{r}{2}}=2^{r/2}\frac{\sigma^r}{(n-1)^{r/2}}\frac{\Gamma[(n-1+r)/2]}{\Gamma[(n-1)/2]}$ When $r>1-n$, $\mathbb{E}\frac{(n-1)^{r/2}\Gamma[(n-1)/2]}{2^{r/2}\Gamma[(n-1+r)/2]}S^r=\mathbb{E}k_{n-1,r}S^r=\sigma^r$, it's unbiased and function of T, hence UMVUE. Where $k_{n,r}=[n^{r/2}\Gamma(n/2)]/[2^{r/2}\Gamma((n+r)/2)]$.

μ/σ : $\mu:\bar{X}$, $\sigma^{-1}:k_{n-1,-1}S^{-1}$ as indep, $\bar{X}k_{n-1,-1}S^{-1}$ unbiased and fun of T.

$\tau\,s.t.\mathbb{P}[X_1\le\tau]=p:\tau=\mu+\Psi^{-1}(p)\sigma$, sub result of μ,σ in.

EX2: $X_i\sim Uni(0,\theta)$: $X_{(n)}$ is comp. & suff. with $\mathbb{E}(\frac{n+1}{n}X_{(n)})=\theta$.

Approach 2: Let $\mathbb{E}_\theta h(S)=\tau$, expand both sides. As a function of θ, The coefficients should be same. Then we can get function h.

EX3: $X_i\sim Bernoulli(\theta),S=\sum X_i$ is comp & suff $\tau=\theta(1-\theta)$. Assume $\mathbb{E}h(S)=\tau$ for all $\theta\in(0,1)\Rightarrow\mathbb{E}h(S)=\sum h(k)\binom{n}{k}\theta^k(1-\theta)^{n-k}=\theta(1-\theta)$. Divide by $(1-\theta)^n$ on both sides, and let $\rho=\frac{\theta}{1-\theta}$: $\sum_{k=0}^nh(k)\binom{n}{k}\rho^k=\rho(1+\rho)^{n-2}=\sum_{k=1}^{n-1}\binom{n-2}{k-1}\rho^k$ $\Rightarrow h(0)=h(n)=0,h(k)=\binom{n-2}{k-1}/\binom{n}{k}=\frac{k(n-k)}{n(n-1)}\Rightarrow h(T)=\frac{T(n-T)}{n(n-1)}$.

EX4: Power series: $\mathbb{P}[X=x]=\gamma(x)\theta^x/c(\theta)$, $\gamma(x)$ known, θ unknown.

$Poisson(\theta)$:$\gamma(x)=\frac{1}{x!}$, $c(\theta)=e^\theta$, $Bino(n,p):\gamma(x)=\binom{n}{k}I_{\mathbb{N}}(x)$, $c(\theta)=(1+\theta)^n$

As full rank Exp Family: $T=\sim X_i$ is comp & suff. PMF of T is $\mathbb{P}[T=t]=\gamma_n(t)\theta^t/c^n(\theta)$, where $\gamma_n(t)=\sum_{x_1+,\cdots,+x_n=t}[\gamma(x_1)\cdots\gamma(x_n)]$ For $\tau=g(\theta)$:

$g(\theta)=\frac{\theta^r}{[c(\theta)]^p}$ assume $\mathbb{E}h(T)=\tau$ then $\sum_{t=1}^\infty h(t)\gamma_n(t)\theta^t=[c(\theta)]^{n-p}\theta^r=\sum_{t=1}^\infty\gamma_{n-p}(t)\theta^{t+r}=\sum_{t=r}^\infty\gamma_{n-p}(t-r)\theta^t$ where the second equality is expectation of real number w.r.t PMF of $n-p$. Then $h(T)=\frac{\gamma_{n-p}(T-r)}{\gamma_n(T)}I_{(T\ge r)}$.

EX 3rd approach: $X_i\sim f_\theta=\theta x^{-2}I_{(x>\theta)}$, $\tau=\mathbb{P}[X_1>t]$, where $t>0$ is a constant. $X_{(1)}$ is comp & suff while $T=I_{(X_1>t)}$ is unbiased. The UMVUE is $\mathbb{E}[T|X_{(1)}]$ i.e. $\mathbb{P}[X_1>t|X_{(1)}=x_{(1)}]=\mathbb{P}[\frac{X_1}{x_{(1)}}>\frac{t}{x_{(1)}}|X_{(1)}=x_{(1)}]=\mathbb{P}[\frac{X_1}{x_{(1)}}>\frac{t}{x_{(1)}}]$ as $\frac{X_1}{X_{(1)}}$ is ancillary.Let $S=\frac{t}{x_{(1)}}$. When $s<1$, as $\frac{X_1}{X_{(1)}}\ge1\,a.s.$, $\mathbb{P}[\cdot]=0$. When $s\ge1$ $\mathbb{P}[\frac{X_1}{X_{(1)}}>s]=\sum_{i=1}^n\mathbb{P}[\frac{X_1}{X_{(1)}}>s,X_{(1)}=x_i]=(n-1)\mathbb{P}[\frac{X_1}{X_{(1)}}>s,X_{(1)}=x_n]=(n-1)\mathbb{P}\{X_1>sX_n,X_2>X_n,\cdots,X_{n-1}>X_n\}=\mathbb{E}\{\mathbb{P}[X_1>sX_n,X_2>X_n,\cdots,X_{n-1}>X_n]|X_n\}$. Int. with the pdf given above, it is $\frac{n-1}{nt}X_{(1)}$. Hence, $h(X_{(1)})=\frac{n-1}{nt}X_{(1)}I_{X_{(1)}\le t}+I_{X_{(1)}>t}$ is UMVUE of τ.

Non-parametric: Order statistic $T=(X_{(1)},\cdots,X_{(n)})$ is comp & suff. Function $\psi(X_1,\cdots,X_n)$ is a function of T iff ψ is symmetric. Hence, unbiased U-statistic is UMVUE. e.g. $\bar{X}:\mathbb{E}X_1$; $S^2=\frac{1}{n}\sum(X_i-\bar{X})^2:VarX_1$; $F_n(t)=\sum I_{X_i\le t}:F(t)$.

Stein's Shrinkage: $X_i\sim N(\theta,I_k)$, where θ is an unknown $k\times1$ vector. We know $\bar{X}$ is UMVUE for θ. It is proved $\bar{X}$ is inadmissible when $k\ge3$. Assume $\hat{\theta}=\bar{X}+\frac{1}{n}g(\bar{X})$ with g to be determined s.t. $\mathbb{E}[\|\bar{X}-\theta\|^2]-\mathbb{E}[\|\hat{\theta}-\theta\|^2]>0$, for $\theta\in\Theta$. Rewrite as: $\mathbb{E}[\|\bar{X}-\theta\|^2]-\mathbb{E}[\|\bar{X}-\theta+\frac{1}{n}g(\bar{X})\|^2]=-\frac{1}{n^2}\mathbb{E}\|g(\bar{X})\|^2-\frac{2}{n}\mathbb{E}\{g^T(\bar{X})(\bar{X}-\theta)\}=-\frac{1}{n^2}\mathbb{E}\|g(\bar{X})\|^2-\frac{2}{n}\sum_{j=1}^k\mathbb{E}\{g_j(\bar{X})(\bar{X}_j-\theta_j)\}$. Note $Y=\bar{X}$, with pdf $f(y;\theta)$, then $(Y_j-\theta_j)=-\frac{1}{n}\frac{\partial}{\partial Y_j}log\,f(Y,\theta)$. Then integral by parts, there is $\mathbb{E}\{g_j(\bar{X})(\bar{X}_j-\theta_j)\}=\frac{1}{n}\mathbb{E}[\frac{\partial}{\partial Y_j}g_j(Y)]$. Question reduce to $\mathbb{E}\|g(\bar{X})\|^2+2\sum_{j=1}^k\mathbb{E}[\frac{\partial}{\partial X_j}g_j(\bar{X})]<0$.

Consider $\psi:\mathbb{R}^k\to\mathbb{R}$ s.t. $g_i(x)=\frac{\partial}{\partial x_i}log\psi(x)=\frac{1}{\psi(X)}\frac{\partial}{\partial x_i}\psi(x)$. If $\sum_{i=1}^k\frac{\partial^2}{\partial x_i^2}\psi(x)=0$, we can check the inequality holds, and such ψ is called harmonic function.

e.g. $\psi(x)=\|x\|^{-(k-2)}$, $k>3$, $g(x)=-\frac{k-2}{\|x\|^2}x$, $MSE(\hat{\theta})<MSE(\bar{X})$.

- James-Stein Estimator: Biased but better than all of the unbiased ones.

Information Inequality

Interested in Squared Error Loss. $T(X)$ estimate $\tau(P)$, which is fun. of $P\in\mathcal{P}$

Fisher Information Preparation: 1. Parametric family with PDF $p(x;\theta)\in\mathcal{P}_\theta$, and dominated by measure μ; 2. Support doesn't depend on θ, denoted as A; 3. $\frac{\partial}{\partial\theta}p(x;\theta)$ exists for all $x\in A$, $\theta\in\Theta$. 4. If T is any statistic with finite mean for all $\theta\in\Theta$, then the order of can be changed: $\frac{\partial}{\partial\theta}\int Tp(x;\theta)dx=\int T\frac{\partial}{\partial\theta}p(x;\theta)dx$.

Remark: all Exp family ✓, but $Uni(0,\theta)$ & $\frac{1}{b}e^{-(x-a)/b}I_{(x>a)}$ ×.

Fisher Information: Let X be a single sample from $P\in\mathcal{P}_\theta$, where parameter space Θ is an open set in $\mathbb{R}$. Suppose conditions above hold, the Fisher Information number is defined as: $I(\theta)=\mathbb{E}\{\frac{\partial}{\partial\theta}log\,p(X;\theta)\}^2=\int(\frac{\partial}{\partial\theta}log\,p(x;\theta))^2p(x;\theta)dx$

Multi-parameter: Fisher Information Matrix $I(\theta)=\mathbb{E}\{\frac{\partial}{\partial\theta}log\,f_\theta(X)[\frac{\partial}{\partial\theta}log\,f_\theta(X)]^T\}$

Remarks: · Fisher Information doesn't depend on estimator, but on parameterization.

- Let $\theta=\psi(\eta)$, FI of θ is $I(\theta)$, for η : $I_\eta(\eta)=[\psi'(\eta)]^2I(\psi(\eta))$
- Θ is open set: to make $\frac{\partial}{\partial\theta}p(x;\theta)$ always exists. In Exp fam. full rank is needed.
- Interpret: Larger $I(\theta)\Rightarrow$ more Information about $\theta\Rightarrow$ better estimated.

Properties: · If $X\perp Y$, $I_{X,Y}(\theta)=I_X(\theta)+I_Y(\theta)$. can be diff. dist. share same θ;

- In particular: $X_1,\cdots,X_n$ i.i.d $I_n(\theta)=nI_1(\theta)$;
- Suppose $p(x;\theta)$ twice differentiable in θ, and $\frac{\partial}{\partial\theta}\int\frac{\partial p(x;\theta)}{\partial\theta^T}dx=\int\frac{\partial^2p(x;\theta)}{\partial\theta\partial\theta^T}dx$, $\theta\in\Theta$, then $I(\theta)=-\mathbb{E}\{\frac{\partial^2}{\partial\theta\partial\theta^T}log\,p(x;\theta)\}$. Exp family satisfy this one.

Cramer-Rao Lower Bound: $T(X)$ is an estimator with $\mathbb{E}T=g(\theta)$ being a differentiable function of θ. Suppose P_θ has pdf $p(x;\theta)$ w.r.t. a measure μ for all $\theta\in\Theta$, and $p(x;\theta)$ is differentiable in θ, and s.t. $\frac{\partial}{\partial\theta}\int h(x)p(x;\theta)d\mu=\int h(x)\frac{\partial}{\partial\theta}p(x;\theta)d\mu$, $\theta\in\Theta$, for $h=1$ and $h(X)=T(X)$. Then $Var\,T\ge[\frac{\partial}{\partial\theta}g(\theta)]^T[I_n(\theta)]^-[\frac{\partial}{\partial\theta}g(\theta)]$.

Remark: If T is unbiased and $Var\,T=CRLB$, it is UMVUE.

Proof of $k=1$: use $Var\,T\cdot Var[\frac{\partial}{\partial\theta}log\,p(X;\theta)]\ge Cov(T,\frac{\partial}{\partial\theta}log\,p(X;\theta))$, it can be proved that $Cov = g'(\theta)$ and $Var = I_n(\theta)$ as $\mathbb{E}\frac{\partial}{\partial\theta}log\,p(X;\theta) = 0$;

Proof of multi: RHS$= \max_c \frac{(c^T\frac{\partial}{\partial\theta}g(\theta))^2}{c^T I_n(\theta)c}$. Similar with $k = 1$, use $c^T\frac{\partial}{\partial\theta}g(\theta)$ instead.

· CRLB is not affected by 1-1 Reparameterize. Similar to that in Fisher Information.

MLE

Definition: Let $X = (X_{1:n})$ be a sample with joint PDF $f(x;\theta)$ w.r.t measure μ when $\theta \in \Theta \subset \mathbb{R}^k$. For each outcome x, $f(x;\theta)$ is a function of θ called Likelihood: $L(\theta)$

Let $\bar{\Theta}$ be the closure of Θ, A $\hat{\theta} \in \bar{\Theta}$ s.t. $L(\hat{\theta}) = max_{\theta\in\Theta}L(\theta)$ is called a ML estimate of θ. If $\hat{\theta}$ is a Borel function, it's MLE of θ .

Let $g(\,\cdot\,)$ be a Borel function from $\theta \to \mathbb{R}^p$, $p \le k$, if g is not 1-1, $\hat{\nu} = g(\hat{\theta})$ is defined to be MLE of $\nu = g(\theta)$. If it is 1-1, by invariant of MLE, it's MLE of ν.

Computation: · If Θ is finite: Compare directly;

· Generally: Get $L(\theta) \to l(\theta)$, first derivative $= 0$, second < 0. Or Check by def.

Ex: $X_{1:n} = x_{1:n}$ observed, with $X_i \sim Bernoulli(p)$ $L(p) = p^{n\bar{x}}(1-p)^{n(1-\bar{x})}$. $\Theta = (0, 1)$, $\bar{\Theta} = [0, 1]$. If $0 < \bar{x} < 1$, $\bar{x}$ is the unique root with second < 0, and $l(p) \to 0$ when $p \to 0\,or\,1$. If $\bar{x} = 0$ $l(p) = (1-p)^n \searrow$, $\hat{p} = 0 = \bar{x}$. If $\bar{x} = 1$ $l(p) = p^n \nearrow$, $\hat{p} = 1 = \bar{x}$. Hence, $\bar{X}$ is the unique MLE on $\bar{\Theta}$.

MLE in Exp. fam.: $l(\eta) \propto \eta^T T - A(\eta)$, likelihood equation: $\frac{\partial}{\partial\eta}l(\eta) = T - \frac{\partial}{\partial\eta}A(\eta) = 0$ and $\frac{\partial^2}{\partial\eta\partial\eta^T}l(\eta) = -\frac{\partial^2}{\partial\eta\partial\eta^T}A(\eta) = -Var(T) \le 0$. If $T(X)$ in the range of $\frac{\partial}{\partial\eta}A(\eta)$, T is unique MLE of $\mu(\eta) = \frac{\partial}{\partial\eta}A(\eta)$. As each component of $\mu(\eta)$ is monotone decreasing, $\exists\,\mu^{-1}\,s.t.\,\eta = \mu^{-1}(\frac{\partial}{\partial\eta}A(\eta))$, and hence $\hat{\eta} = \mu^{-1}(T)$ is the MLE of η.

Asymptotic Properties:

Conditions: 1. $f(x;\theta)$ are distinct; 2.they have common support; 3.Observations $X = (X_{1:n})$ are iid with density $f(x_i;\theta)$ w.r.t μ; 4.Space Θ contains an open set, where true θ_0 is interior point.

Reasonability: With 1-3: For any fixed $\theta \ne \theta_0$ $\mathbb{P}_{\theta_0}[L(\theta_0|X) > L(\theta|X)] \to 1$, $n \to \infty$

Consistency: With 1-4: Suppose for almost all x, $f(x;\theta)$ is differentiable in the open set Θ, then, with probability 1 there is at least 1 seq. of $\hat{\theta_n}\,s.t.\,\forall\epsilon > 0$, $\mathbb{P}[|\hat{\theta}_n - \theta_0| > \epsilon] \to 0 \Leftrightarrow \hat{\theta}_n \to_P \theta_0$.

Efficiency: With 1-4, assuming Fisher Information exists and finite, together with:

· $\frac{\partial^3}{\partial\theta^3}f(x;\theta)$ exists and continuous in θ;

· $\int f(x;\theta)d\mu$ can be 3 times differentiated under integral sign;

· For all $\theta_0 \in \Theta$ there exists positive number c and $M(X)$ with $\mathbb{E}_\theta[M_{ijk}(x)] < \infty$, $s.t.\,\|\frac{\partial^3 log\,f(x;\theta)}{\partial\theta_i\partial\theta_j\partial\theta_k}\| \le M_{ijk}(x)$, for all $\|\theta - \theta_0\| < c$.

Then, any consistent seq. $\hat{\theta} : \sqrt{n}(\hat{\theta}_n - \theta_0) \to^L N(0, [I(\theta)]^{-1})$.

Achieves the CRLB for unbiased estimators when $n \to \infty$.

If the root of likelihood equation is unique, it's consistent, asymptotically efficient whether or not it's MLE.

Linear Model

Model Setting: Observations: $(X_1, Z_1), \cdots (X_n, Z_n)$, $Z_i : p \times 1$, $X_i : 1 \times 1$;

Model: $X_i = Z_i^T\beta + \epsilon_i$, $i = 1 : n$. $\beta_{p\times1}$: unknown parameter, ϵ_i : random error;

Matrix Form: $X_{n\times1} = Z_{n\times p}\beta + \epsilon_{n\times1}$, Z : design matrix (no ^{T} here).

Estimation: LSE: If $\hat{\beta}$ s.t. $\|X - Z\hat{\beta}\| = \min_b\|X - Zb\|$ it is LSE.

For any p vector a, $a^T\hat{\beta}$ is LSE of $a^T\beta$.

Solution: $\|X - Z\hat{b}\|^2 = Z^TX + b^TZ^TZb - 2X^TZb$, $\frac{\partial}{\partial b}\underline{\cdots} = 0 \Rightarrow Z^TZb = Z^TX$

As it is convex in b, any b s.t. $Z^TZb = Z^TX$ is an LSE of β. If full rank, $r(Z) = p$ and $\hat{\beta} = (Z^TZ)^{-1}Z^TX$, if not $\hat{\beta} = (Z^TZ)^-Z^TX$.

Non-Full Rank case: $\exists\beta_1 \ne \beta_2$, $s.t.\,Z\beta_1 = Z\beta_2$, the model is nor identifiable.

Reparameterize: If $Z_{n\times p}$ is of rank r, $\exists Q_{r\times p}\,s.t.\,Z = Z_*Q$, where $r(Z_*) = r$, and the model can be written as $X = Z\beta + \epsilon = Z_*Q\beta + \epsilon = Z_*\tilde{\beta} + \epsilon$.

To estimate $\nu a^T\beta$, only when $a = Q^Tc$ for some c it is meaningful, i.e.$a^T\beta = c^T\tilde{\beta}$.

Assumptions: A_1: $\epsilon \sim N(0, \sigma^2I_n)$, with unknown $\sigma^2 > 0$

A_2: $\mathbb{E}\epsilon = 0$, $Var\ \epsilon = \sigma^2I_n$, with unknown $\sigma^2 > 0$

A_1: $\mathbb{E}\epsilon = 0$, $Var\ \epsilon$ unknown. (Actually $A_1 \Rightarrow A_2 \Rightarrow A_3$)

Properties: Assume a linear model with A_3:

1. $a = Q^Tc$ for some $c \in \mathbb{R}^r \Leftrightarrow a \in R(Z) = R(Z^TZ)$, row space of Z;

2. If $a \in R(Z)$, the LSE $a^T\hat{\beta}$ is unique and unbiased for $a^T\beta$

Proof: $a \in R(Z) \Rightarrow \exists b\,s.t.\,a = Z^TZb \Rightarrow a^T\hat{\beta} = b^TZ^TZ(Z^TZ)^-Z^TX$, $\mathbb{E}a^T\hat{\beta} = b^TZ^TZ\beta = a^T\beta$. Unbiased✓. If $\bar{\beta}$ is also LSE $Z^TZ\bar{\beta} = Z^TX \Rightarrow a^T\hat{\beta} - a^T\bar{\beta} = b^TZ^TZ(\hat{\beta} - \bar{\beta}) = b^T(Z^TX - Z^TX) = 0$. □

3. If $a \notin R(Z)$, and A_1 holds, then $a^T\beta$ is not estimable.

Proof: Assume $\exists h(X, Z)\,s.t.\,\mathbb{E}h(X, Z) = a^T\beta$, then $a = \frac{\partial}{\partial\beta}a^T\beta = \frac{\partial}{\partial\beta}\mathbb{E}h(X, Z) = \frac{\partial}{\partial\beta}\int\cdots dx = Z^Tc$. Contradictory. □

Ex: One Way ANOVA: $n = \sum n_j$, with integers $n_1, \cdots, n_m > 0$, $X_{ij} = \mu_i + \epsilon_{ij}$. ϵ_{ij} are iid random error with $\mathbb{E}\epsilon = 0$, $Var\,\epsilon = \sigma^2$. Let J_k be the $k \times 1$ vector of ones, $X_i = (X_{i1}, \cdots, X_{in_i})^T$, and $X = (X_1, \cdots, X_m)^T$, $Z = diag(J_{n_1}, \cdots, J_{n_m})$, ϵ is similar to X, and $\beta = (\mu_1, \cdots, \mu_m)^T$. Then $Z^TZ = diag(n_1, \cdots, n_m)$, $(Z^TZ)^{-1} = diag(n_1^{-1}, \cdots, n_m^{-1})$ $\hat{\beta} = (\bar{X}_1\cdot, \cdots, \bar{X}_m\cdot)$.

Linear Estimator: a linear function of X i.e. c^TX for some fixed c.

· $a^T\hat{\beta}$ is linear estimator with $c = Z(Z^TZ)^-a$ · $Var(c^TX) = c^TVar(\epsilon)c$

· If $a \in R(Z)$ and $Var\epsilon = \sigma^2I_n$, $Var\hat{\beta} = \sigma^2a^T(Z^TZ)^-a$.

Properties With A_2: · $\exists$ linear unbiased estimator of $a^T\beta$ iff $a \in R(Z)$.

· Gauss-Markov Thm: If $a \in R(Z)$, then the LSE $a^T\hat{\beta}$ is best linear unbiased estimator (BLUE) of $a^T\beta$. As $Var\,a^T\hat{\beta}$ is min. among all the unbiased linear estimator.

Proof: 1. Assume $\exists c \in \mathbb{R}^p$ s.t. $a^T\beta = \mathbb{E}c^TX = c^TZ\beta$, $\forall\beta \in \mathbb{R}^p \Rightarrow a = Z^Tc$.

2. As $a \in R(Z)$, $\exists b\,s.t.\,a^T\hat{\beta} = b^T(Z^TZ)\hat{\beta} = b^TZ^TX$ by def of LSE, and hence $Var\,c^TX = Var(c^TX + a^T\hat{\beta} - a^T\hat{\beta}) = Var(c^TX - b^TZ^TX) + Var(a^T\hat{\beta}) + 2Cov(c^TX - b^TZ^TX, a^T\hat{\beta}) \ge Var(a^T\hat{\beta}) + 2\sigma^2\{c^TZb - b^TZ^TZb\}$ as shown above $a = Z^Tc$, $c^TZb - b^TZ^TZb = a^Tb - a^Tb = 0$. □

Properties with A_1: · LSE $a^T\hat{\beta}$ is UMVUE for all estimable $a^t\beta$

· UMVUE for σ^2 is $\hat{\sigma}^2 = (n-r)^{-1}\|X - Z\hat{\beta}\|^2$, r is rank of Z.

· Fir estimable $a^T\beta : a^T\hat{\beta} \sim N(a^T\beta, \sigma^2a^T(Z^TZ)^-Z)$, $(n-r)\hat{\sigma}^2/\sigma^2 \sim \chi^2_{n-r}$.

Proof: 1. As it is LSE $Z^TZ\hat{\beta} = Z^TX \Rightarrow \|X - Z\beta\|^2 = \|X - Z\hat{\beta}\|^2 + \|Z\hat{\beta} - Z\beta\|^2 = \|X - Z\hat{\beta}\|^2 - 2\beta^TZ^TX + \|Z\beta\|^2 + \|Z\hat{\beta}\|^2$, the pdf is $exp\{\frac{1}{\sigma^2}\beta^TZ^Tx - \frac{1}{2\sigma^2}[\|X - Z\hat{\beta}\|^2 + \|Z\hat{\beta}\|^2] + \cdots\}$. $(Z^TX, \|X - Z\hat{\beta}\|^2)$ is comp and suff for (β, σ^2). If estimable, $a\hat{\beta}$ is unbiased function of $T \Rightarrow$ UMVUE.

2. $\mathbb{E}\|X - Z\hat{\beta}\|^2 = tr\{VarX - VarZ\hat{\beta}\} = \sigma^2\{n - tr[Z(Z^TZ)^-ZT]\} = \sigma^2\{n - tr[(Z^TZ)^-ZTZ]\} = \sigma(n-r)$, hence unbiased function of T.

3. As it is linear function of normal, it still normal distribution.

Consistency of LSE: Model $X = Z\beta + \epsilon$ with A_3, consider LSE $a^T\hat{\beta}$ with $a \in R(Z)$. Let $\lambda_+[A]$ be the largest eigenvalue of A Suppose $\sup_n\lambda_+[Var\epsilon] < \infty$ and $\lim_n\lambda_+[(Z^TZ)^-] = 0$, then $a^T\hat{\beta} \to^{L^2} a^T\beta$. i.e. $\mathbb{E}\|a^T\hat{\beta} - a^T\beta\|^2 \to 0$, $a^T\hat{\beta} \to_P a^T\beta$.

Proof: $a \in R(Z) \Rightarrow \mathbb{E}[a^T\hat{\beta}] = a^T\beta$, only need to check $Var\,a^T\hat{\beta} = a^T(Z^TZ)^-Z^T[Var\,\epsilon]Z(Z^TZ)^-a$ which is less or equal $\lambda_+[Var\epsilon]a^T(Z^TZ)^-a \le \lambda_+[Var\epsilon]\lambda_+[(Z^TZ)^-]\|a\|^2 \to 0$.

Asymptotic Normality: Model $X = Z\beta + \epsilon$ with A_3 Suppose $\inf_n\lambda_-[Var\epsilon] > 0$ and $\lim_n\max_{1\le i\le n}Z_i^T(Z^TZ)^-Z_i = 0$. Suppose further $n = \sum m_j$ with all of m_j bounded by some m. Random error $\epsilon = (\xi_1, \cdots, \xi_k)$, $\xi_j \in \mathbb{R}^{m_j}$ and $\xi's$ are independent.

· If $\sup_i\mathbb{E}|\epsilon_i|^{2+\delta} < \infty$ for any $a \in R(Z)$, $\frac{a^T(\hat{\beta}-\beta)}{\sqrt{Var(a^T\hat{\beta})}} \to^d N(0, 1)$

· It still holds if $m_1 = \cdots = m_k$, $\xi's$ have same distribution.

Lemma: following is sufficient to $\lim_n\max_{1\le i\le n}Z_i^T(Z^TZ)^-Z_i = 0$:

a). $\lambda + [(Z^TZ)^-] \to 0$ and $Z_n^T(Z^TZ)^-Z_n \to 0$ as $n \to 0$

b). $\exists \nearrow$ seq $\{a_n\}\,s.t.\,\frac{a_n}{a_{n+1}} \to 1$, and $\frac{Z^TZ}{a_n}$ converge to a positive defined matrix.

Ex1: $X_i = \beta_0 + \beta_1t_i + \epsilon_i$ $i = 1 : n$. If $\frac{\sum t_i^2}{n} \to c$, $\frac{\sum t_i}{n} \to d$, $c > d^2$. b) in Lemma ✓.

Ex2: One-Way ANOVA: $\max_{1\le i\le n}Z_i^T(Z^TZ)^-Z_i = \lambda_+[(Z^TZ)^-] = max\frac{1}{n_j}$. If $\min n_j \to \infty$, Asymptotic Normality✓.

Decision Theory

X: a sample from population $P \in \mathcal{P}$ $\mathcal{X}$: the range of X

$\mathcal{A}$: the range of allowable actions $(\mathcal{A}, \mathcal{F}_\mathcal{A})$: the action space

Decision Rule: A measurable function $T : (\mathcal{X}, \mathcal{F}_\mathcal{X}) \to (\mathcal{A}, \mathcal{F}_\mathcal{A})$

Ex: · Point Estimation: $\mathcal{A}$ is the parameter space Θ

· Hypothesis testing: $\mathcal{A}$ is reject or accept H_0

Loss function: Evaluate action a: $L(P, a) : \mathcal{P} \times \mathcal{A} \to [0, +\infty)$.

Risk function: Evaluate rule $T_{(X)}$: $R_T(P) = \mathbb{E}_P[L(P, T(X))] = \int L(P, T(x))dP_X$

If $\mathcal{P}$ is parametric, θ is used instead.

Comparison of Two Decision Rules: T_1 is $\underline{\cdot\ \cdot}$ T_2 if:

· as good as: $R_{T_1}(P) \le R_{T_2}(P)$, $\forall P \in \mathcal{P}$;

· better than: T_1 is as good as T_2, and $R_{T_1}(P) < R_{T_2}(P)$ for some $P \in \mathcal{P}$;

· Equivalent: $R_{T_1}(P) = R_{T_2}(P)$, $\forall P \in \mathcal{P}$;

· Optimal: $T_* \in \mathcal{T}$ is $\mathcal{T}-$optimal if T_* is as good as any other $T \in \mathcal{T}$;

· Admissible: If no $U \in \mathcal{T}$ is better than $T \in \mathcal{T}$, T is $\mathcal{T}-$admissible.

· If there are two $\mathcal{T}-$admissible and not equivalent rules $\Rightarrow$ no optimal.

e.g. Very small risk at some P, no better rules, but not as good as others.

Randomized Decision Rule: δ is a function on $\mathcal{X} \times \mathcal{F}_\mathcal{A}$, s.t. $\forall x \in \mathcal{X}$, $\delta(x, \,\cdot\,)$ is a measure on $(\mathcal{A}, \mathcal{F}_\mathcal{A})$. i.e. If $X = x$ is observed, we have a distribution of actions.

· Non-Randomized Rules can be regarded as $\delta(x, \{a\}) = I_{\{a\}}(T(X))$

· To show δ is randomized rule, we need to show $\delta(x, \,\cdot\,)$ is a prob. measure.

Loss function: $L(P, \delta, x) = \int_\mathcal{A} L(P, a)d\delta(x, a)$.

Risk function: $R_\delta(P) = \mathbb{E}_P[L(P, \delta, X)] = \int_c X\int_\mathcal{A} L(P, a)d\delta(x, a)dP_X$

Randomized Rule with Discrete Dist.: $\delta(x, \cdot)$ assign $p_j(x)$ to non-randomized $T_j(x)$

Ex: Non-rand. $T_1(X) = \bar{X}$, under SEL: $L_1 = (\bar{X} - \theta)^2$, $R_1 = (\mu - \theta)^2 + \frac{\sigma^2}{n}$

Non-rand. $T_2(X) = c$, under Squared Error Loss: $L_2 = (c - \theta)^2$, $R_2 = (c - \theta)^2$

Randomized: $T = T_1$ with p, $T = T_2$ with $1 - p$. $L = pL_1 + (1 - p)L_2$, $R = \cdots$

Ex: Hypothesis Testing: $\mathcal{P}_0 \subset \mathcal{P}$, $\mathcal{P}_1 = \mathcal{P}_0^c$: $H_0 : P \in \mathcal{P}_0$ $v.s.$ $H_1 : P \in \mathcal{P}_1$.

With 0-1 Loss, and $\mathcal{A} = \{0, 1\}$, $T : \mathcal{X} \to \mathcal{A}$. Loss $L = 0$ if correct $L = 1$ if not.

Risk $R_T(P) = \mathbb{P}[T = 1]I_{P\in\mathcal{P}_0} + \mathbb{P}[T = 0]I_{P\in\mathcal{P}_1}$

Thm: Randomized & Non-Randomized: Suppose that $\mathcal{A}$ is convex and $L(P, a)$ is a convex function of a for any $P \in \mathcal{P}$. Let δ be a randomized rule s.t. $\int_\mathcal{A}\|a\|d\delta < \infty$ for $\forall x \in \mathcal{X}$. Let $T(X) = \int_\mathcal{A} ad\delta(x, a)$, then $L(P, T) \le L(P, \delta, x)$ for any $x \in \mathcal{X}$, $P \in \mathcal{P}$. Proof based on Jensen's inequality.

Squared Error Loss, Absolute Loss, etc are convex; 0-1 Loss is not convex.

Interpret: Non-randomized T get from δ will be better.

Bayes Risk: Average of risk function $R_T(P)$ over $\mathcal{P}$: $r_T(\Pi) = \int_\mathcal{P} R_T(P)d\Pi$. Where Π is a known prob. measure on $(\mathcal{P}, \mathcal{F}_\mathcal{P})$. $r_T(\Pi)$ is the Bayes risk of T w.r.t. Π. If $T_* \in \mathcal{T}$ s.t. $r_{T_*}(\Pi) \le r_T(\Pi)$, $\forall T \in \mathcal{T}$. T_* is called a $\mathcal{T}-$Bayes rule w.r.t Π.

Minimax Rule: If $T_* \in \mathcal{T}$ and $\sup_{P\in\mathcal{P}}R_{T_*}(P) \le \sup_{P\in\mathcal{P}}R_T(P)$, $\forall T \in \mathcal{T}$. T_* is called $\mathcal{T}-$minimax rule.

Recall Bayes Analysis:

X is from a population in a parametric family $\mathcal{P} = \{p_\theta : \theta \in \Theta\}$, where $\Theta \subset \mathbb{R}^k$ for some fixed $k \in \mathbb{N}^+$. Real valued θ is a realization of r.v. $\tilde{\theta} \sim \pi$, π is the prior dist.

Sample $X \in \mathcal{X}$ from $P_\theta = P_{X|\theta}$, it is conditional dist. of $X|\tilde{\theta} = \theta$.

Posterior: dist. of $\tilde{\theta}$ conditional on $X = x$: $\pi(\theta|x) = \int f(x_{1:n}|\theta)\pi(\theta)dx_{1:n}$

Marginal: dist. of $X = x$: $m(x) = \int f(x_{1:n}|\theta)\pi(\theta)d\theta$

Bayes Formula: Assume $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ is dominated by measure ν, and $f_\theta(x) = \frac{dP_\theta}{d\nu}$ is a Borel function on $(\mathcal{X} \times \Theta, \sigma(\mathcal{B}_\mathcal{X} \times \mathcal{B}_\theta))$. Let Π be a prior dist. on Θ. Suppose that $m(x) = \int_\Theta f_\theta(x)d\Pi > 0$, Π is another measure on $\mathcal{X}$. Then the Posterior dist $P_{\theta|x} \ll \Pi$ and $\frac{dP_{\theta|x}}{d\Pi} = \frac{f_\theta(x)}{m(x)}$. Further, if $\Pi \ll \lambda$ for a measure λ and $\frac{d\Pi}{d\lambda} = \pi(\theta)$, then $\frac{dP_{\theta|x}}{d\lambda} = f_\theta(x)\pi(\theta)/m(x)$.

Bayes Action: Let $\mathcal{A}$ be an action space in a decision space, and $L(\theta, a) > 0$ be a loss function. For any $x \in \mathcal{X}$, a Bayes action w.r.t. Π is any $\delta(x) \in \mathcal{A}$ s.t. $\mathbb{E}[L(\tilde{\theta}, \delta(x)|X = x)] = \min_{a\in\mathcal{A}}\mathbb{E}[L(\tilde{\theta}, a|X = x)]$, the $\mathbb{E}[\,\cdot\,]$ is w.r.t posterior $P_{\theta|x}$.

Remarks: · For each $x \in \mathcal{X}$ $\delta(x)$ minimize posterior expected loss, and hence we can get a mapping $\mathcal{X} \to \mathcal{A}$;

· If the mapping is a measurable function, it is a Bayes Rule;

· Bayes action depends on prior and loss function.

Properties: Assume conditions in Bayes Formula Thm satisfied, and loss function $L(\theta, a)$ is convex in a for any fixed θ. And for each $x \in \mathcal{X}$, $\mathbb{E}[L(\tilde{\theta}, a|X = x)] < \infty$ for some a

1) If $\mathcal{A} \subset \mathbb{R}^p$ is compact, a Bayes action exists for each $x \in \mathcal{X}$;

2) If $\mathcal{A} \subset \mathbb{R}^p$ and $L(\theta, a)$ goes to ∞ as $\|a\| \to \infty$ uniformly in $\theta \in \Theta_0 \subset \Theta$ with $\Pi(\Theta_0) > 0$, a Bayes action exists for each $x \in \mathcal{X}$;

3) If $L(\theta, a)$ is strictly convex for each fixed θ in a in 2), 3), the result will be unique.

Ex1: $X_i \sim N(\mu, \sigma^2)$, $\mu \sim N(a, b)$, σ^2 is known under Squared Error Loss:

$\pi(\mu|x) = f(x_{1:n}|\mu)\pi(\mu)/m(x) \propto f(x_{1:n}|\mu)\pi(\mu) \propto exp\{-\frac{nb+\sigma^2}{2b\sigma^2}[\mu^2 - 2\frac{b\sum x + a\sigma^2}{nb+\sigma^2}\mu]\}$

$\mu|x \sim N(\frac{b\sum x + a\sigma^2}{nb+\sigma^2}, \frac{b\sigma^2}{nb+\sigma^2})$, $\mathbb{E}[L|X = x] = \mathbb{E}[(\mu - \delta)^2|x] = (\delta - \frac{b\sum x + a\sigma^2}{nb+\sigma^2})^2 + \frac{b\sigma^2}{nb+\sigma^2}$

$\Rightarrow$ Under SEL: $\delta(x) = \mathbb{E}[\mu|X = x] = \frac{b\sum x + a\sigma^2}{nb+\sigma^2}$

Ex2: Same setting with Ex1, but for $g(\mu)$, e.g. $g(\mu) = \mu^2$: $\delta = \mathbb{E}[g(\mu)|X = x]$.

Ex3: $X_i \sim Poisson(\lambda)$, $\lambda \sim Gamma(a, b)$ find Bayes estimator of λ^j:

$f(x_{1:n}|\lambda) \propto \lambda^{\sum x}e^{-n\lambda}$, $\pi(\lambda) = \frac{b^a}{\Gamma(a)}\lambda^{a-1}e^{-b\lambda}$, $\pi(\lambda|x) \propto \lambda^{\sum x + a - 1}e^{-(b+n)\lambda}$

$\lambda|x \sim Gamma(\sum x + a, b + n) =: Gamma(a', b')$, the Bayes rule under SEL is:

$\mathbb{E}[\lambda^j|X = x] = \int_0^\infty \frac{b'^{a'}}{\Gamma(a')}\lambda^{a'+j-1}e^{-b'\lambda}d\lambda = \frac{b'^{a'}}{\Gamma(a')}\frac{\Gamma(a'+j)}{b'j + a'}$.

Ex4: Bayes Classifier: label $y_i = 1, \cdots, k$, $X_i|y_i = k \sim p_k(x_i)$. Assume labels have prior π and under 0-1 loss the $\hat{y} = \arg\max_y p(y)\prod p(x_i|y)$.

Conjugate Prior: If a prior is in the same parametric family as the posterior, it's ..

Exp. Family: $f(x;\eta) = h(x)exp\{\eta^TT(x) - A(\eta)\}$ always have a conjugate prior in the form of $\pi(\eta;\xi,\nu) = g(\xi,\nu)exp\{\eta^T\xi - \nu A(\eta)\}$, where ν is a scalar and ξ is a vector in the same length of η.

Admissibility: In a decision problem, let $\delta(X)$ be a Bayes rule w.r.t. a prior Π.

(i) If $\delta(X)$ is a unique Bayes rule, then $\delta(X)$ is admissible.

(ii) If Θ is a countable set, the Bayes risk $r_\delta(\Pi) < \infty$, and Π gives positive probability to each $\theta \in \Theta$, then $\delta(X)$ is admissible.

(iii) Let $\mathcal{T}$ be the class of decision rules with continuous risk fun. If $\delta(X) \in \mathcal{T}$, $r_\delta(\Pi) < \infty$, and Π gives positive prob. to any open subset of Θ, then $\delta(X)$ is $\mathcal{T}$-admissible.

Remark: If T is better$\Rightarrow$ T has same posterior risk as $\delta(X) \Rightarrow T$ is also Bayes rule.

Problem: strictly better on Θ_0, where $\Pi(\Theta_0) = 0$. No such Θ_0 in i) ii) iii).

Bias: If $\delta(X)$ is Bayes estimator of $\tau(\theta)$ under SEL, w.r.t. Π. If $\delta(X)$ is unbiased, Bayes risk $r_\delta(\Pi) = 0$. Proof: under SEL $\delta(X) = \mathbb{E}_\theta[\tau(\theta)|X]$. If unbiased $\mathbb{E}_X[\delta(X)|\theta] = \tau(\theta)$. Hence $\mathbb{E}[\delta(X)\tau(\theta)] = \mathbb{E}_\theta\{\mathbb{E}_X[\delta(X)\tau(\theta)|\theta]\} = \mathbb{E}_\theta[\tau^2(\theta)]$, Similarly $\mathbb{E}[\delta(X)\tau(\theta)] = \mathbb{E}_X\{\mathbb{E}_\theta[\delta(X)\tau(\theta)|X]\} = \mathbb{E}_X[\delta^2(X)]$. And the Bayes risk is

$r_\delta = \mathbb{E}_\theta\{\mathbb{E}_X[(\delta(X)-\tau(\theta))^2|\theta]\} = \mathbb{E}\delta^2(X) + \mathbb{E}\tau^2(\theta) - 2\mathbb{E}[\delta(X)\tau(\theta)] = 0.$ □

Remark: Usually biased, comes from prior. But usually vanish when $n \to \infty$.

Generalized Bayes Action: The minimization in the def. of Bayes action is same as: $\int_\Theta L(\theta,\delta(x))f_\theta(x)d\Pi = \min_a \int_\Theta L(\theta,a)f_\theta(x)d\Pi$. This def. also works when Π is not a prob, measure on Θ, in with case $m(x)$ may not finite, and we can not get $\pi(\theta|x)$. δ solved from above is called Generalized Bayes Action.

· Improper prior: $\Pi(\Theta) \neq 1$ · Proper prior: $\Pi(\Theta) = 1$

Ex: $X_i \sim N(\mu,\sigma^2)$, $\mu \in \Theta \subset \mathbb{R}$ unknown, $\sigma^2 > 0$ known, under SEL:

No Information: If $\Theta = [a,b]$ we can set $\Pi = Uni(a,b)$. But if $\Theta = \mathbb{R}$, let $\pi(\theta) = 1$ for all θ, it is improper. Minimize $\int_\mathbb{R}(\mu-a)^2(2\pi\sigma^2)^{-n/2}exp\{-\frac{\sum(x_i-\mu)^2}{2\sigma^2}\}d\mu$ is same with minimize: $\int_\mathbb{R}(\mu-a)^2 exp\{-\frac{1}{2\sigma^2}[\sum(x_i-\bar{x})^2 + n(\bar{x}-\mu)^2]\}d\mu$ Let $\frac{\partial}{\partial a}[\cdot] = 0$, the Bayes rule is $\delta(x) = \frac{\int_\mathbb{R}\mu exp\{-n(\bar{x}-\mu)^2/(2\sigma^2)\}d\mu}{\int_\mathbb{R} exp\{-n(\bar{x}-\mu)^2/(2\sigma^2)\}d\mu} = \bar{x}$. If $\Pi = N(a,b)$, it is $\frac{\sigma^2}{nb+\sigma^2}a + \frac{nb}{nb+\sigma^2}\bar{x}$, converge to $\bar{x}$ when $b \to \infty$.

Admissibility of Generalized Bayes Rules: Suppose that Θ is an open set of $\mathbb{R}^k$. Let $\mathcal{T}$ be the class of decision rules having continuous risk functions. A decision rule $T \in \mathcal{T}$ is $\mathcal{T}$-admissible if there exists a sequence $\{\Pi_j\}$ of (possibly improper) priors, which give positive measures to any open set, such that

(a) the generalized Bayes risks $r_T(\Pi_j)$ are finite for all j;

(b) for any $\theta_0 \in \Theta$ and $\eta > 0$ $\lim_j \frac{r_T(\Pi_j) - r_j^*(\Pi_j)}{\Pi_j(O_{\theta_0,\eta})} = 0$, where $r_j^*(\Pi_j) = \inf_T r_T(\Pi_j)$, and $O_{\theta_0,\eta} = \{\theta \in \Theta : \|\theta-\theta_0\| < \eta\}$ with $\Pi_j(O_{\theta_0,\eta}) < \infty$ for all j.

Proof: Suppose T is not $\mathcal{T}-$admissible. Then there exists $T_0 \in \mathcal{T}$ s.t. $R_{T_0}(\theta) \le R_T(\theta)$ for all θ and $R_{T_0}(\theta_0) < R_T(\theta_0)$ for a $\theta_0 \in \Theta$. From the continuous $R_{T_0}(\theta) < R_T(\theta) - \epsilon$ for $\theta \in O_{\theta_0,\eta}$, for some constant $\epsilon > 0, \eta > 0$. Then $r_T(\Pi_j) - r_j^*(\Pi_j) \ge r_T(\Pi_j) - r_{T_0}(\Pi_j) > \epsilon\Pi_j(O_{\theta_0,\eta})$, contradictory with (b). □

Ex: $X_i \sim N(\mu,\sigma^2)$, $\mu \sim N(a,b)$, σ^2 is known under SEL: $\delta(X) = \bar{X}$

Risk function is continuous in μ if the risk is finite. Let $\Pi_j = N(0,j)$,

$R_\delta(\mu) = Var\bar{X} = \frac{\sigma^2}{n}$ fixed, and $r_\delta(\Pi_j) = \frac{\sigma^2}{n}$. Consider Bayes rule w.r.t. Π_j: $\delta_j = \frac{nj}{nj+\delta^2}\bar{X}$, $R_{\delta_j}(\mu) = \frac{\sigma^2 nj^2 + \sigma^4\mu^2}{(nj+\sigma^2)^2}$, $r_j^*(\Pi_j) = \frac{\sigma^2 nj}{nj+\sigma^2}$, $\Pi_j(O_{\mu_0,\eta}) \approx \frac{2\eta}{\sqrt{j}}\Phi'(\xi_j)$ for some $\xi_j \in ((\mu_0-\eta)/\sqrt{j}, (\mu_0+\eta)/\sqrt{j})$. $j \to 0$, (b) satisfied $\Rightarrow \bar{X}$ admissible.

Empirical Bayes: Estimate the hyperparameter with historical data or the current data x, if historical data is not available. This method is called Empirical Bayes

View x from marginal dist. $P_\xi = \int_\Theta P_\theta(x)d\Pi_{\theta|\xi}$. $\Rightarrow$Find MLE of ξ.

$X_i \sim N(\mu,\sigma^2)$, $\mu \sim N(a,b)$, σ^2 is known under SEL, $\Pi_{\theta|\xi} = N(0,\sigma_0^2)$, with $\xi = \sigma_0^2$

$P_\xi = \int_\mathbb{R} f(x_{1:n}|\mu)\pi(\mu|\sigma_0^2)d\mu \cdots l(\sigma_0^2) \propto (n\sigma_0^2+\sigma^2)^{-1/2}exp\{-\frac{n\bar{x}^2}{2(n\sigma_0^2+\sigma^2)}\}$

MLE of σ_0^2 is $\hat{\sigma_0^2} = \max\{0, \bar{x} - \sigma^2/n\}$.

Hierarchical Bayes: put a prior on hyperparameter.

Computation issues: we need $\mathbb{E}_p(\tau)$ where the expectation is w.r.t. posterior $p(\theta)$

MCMC: · Generate iid $\theta^{1:m}$ from a pdf $h(\theta) > 0$ w.r.t ν

· By SLLN, as $m \to \infty$, $\hat{\mathbb{E}}_p(\tau) = \frac{1}{m}\sum_j \frac{\tau(\theta^j)p(\theta^j)}{h(\theta^j)} \to_{a.s.} \int \frac{\tau(\theta)p(\theta)}{h(\theta)} = \mathbb{E}_p(\tau)$.

Minimax Rule

Definition: $T_* = \arg\inf_T sup_{\theta\in\Theta}R_T(\theta)$

Control the worst case, but maybe not that good in other case.

Find a Minimax Rule: Let $\Theta_0 \subset \Theta$, and T is minimax of $\tau(\theta)$ when $\theta \in \Theta_0$. If $sup_{\theta\in\Theta}R_T(\theta) = sup_{\theta\in\Theta_0}R_T(\theta)$, T is minimax on Θ. Proof: By def.

$\forall T_0 \neq T : sup_{\theta\in\Theta}R_T(\theta) = sup_{\theta\in\Theta_0}R_T(\theta) \le sup_{\theta\in\Theta_0}R_{T_0}(\theta) \le sup_{\theta\in\Theta}R_{T_0}(\theta)$ □

Similarly, if T is unique minimax on Θ_0 also unique on Θ

Minimaxity of a Bayes rule: Let Π be a proper prior on Θ and T be a Bayes rule of $\tau(\theta)$ w.r.t. Π. If $R_T(\theta) \le \int R_T(\theta)d\Pi = r_T(\Pi), \forall\theta \in \Theta$. i.e. T has constant risk function or bounded by Bayes risk. Then: (1) T is minimax

(2) If in addition T is the unique Bayes rule, it's also unique Minimax rule.

Proof: Let δ be any other rule $\sup_\theta R_\delta(\theta) \ge \int R_\delta(\theta)d\Pi \ge \int R_T(\theta)d\Pi \ge \sup_\theta R_T(\theta)$

The last $\ge$ comes from "$\forall\theta \in \Theta$". If unique, second $\ge$ is $>$, i.e. unique minimax □

Corollary: Let Π be a proper prior on Θ and T be a Bayes rule of $\tau(\theta)$ w.r.t. Π. If $\exists\Theta_0$ s.t. $R_T(\theta)$ is constant on Θ_0 which equals to $sup_{\theta\in\Theta}R_T(\theta)$, then,

· If $\Pi(\Theta_0) = 1$, T is minimax; · If in addition T is the unique Bayes rule w.r.t. Π, it is also the unique minimax estimator. Proof:

$\Pi(\Theta_0) = 1 \Rightarrow T$ is minimax on Θ_0 (proved above) $\Rightarrow$ minimax on Θ. □

Ex: $X_i \sim Bernoulli(p)$ estimate p under SEL, with prior $p \sim Beta(\alpha,\beta)$:

$p|x \sim Beta(\alpha + \sum x, \beta + n - \sum x)$, $T = \mathbb{E}[p|x] = \frac{\alpha+\sum x}{\alpha+\beta+\sum x}$

$R_T(p) = \frac{np(1-p)+(\alpha-p\alpha-p\beta)^2}{(\alpha+\beta+n)^2}$. Let $\alpha = \beta = \sqrt{n}/2$, $R_T(p) = 1/[4(1+\sqrt{n})^2]$ is a constant. $\Rightarrow T = \frac{\sqrt{n}\bar{X}+1/2}{\sqrt{n}+1}$ is the unique minimax estimator.

Remark: Minimax estimators are irrelevant with prior, but depends on loss function.

Limit of Bayes rules: Let Π_j $j = 1,2,\cdots$ be a seq. of priors and r_j be the Bayes risk of a Bayes rule of $\tau(\theta)$ w.r.t. Π_j. If for a rule T with $sup_\theta R_T(\theta) < \infty$, $\liminf_j r_j \ge \sup_\theta R_T(\theta)$, then T is minimax.

Corollary: Let Π_j $j = 1,2,\cdots$ be a seq. of priors and r_j be the Bayes risk of a Bayes rule of $\tau(\theta)$ w.r.t. Π_j, If a rule T with constant risk function $R_T(\theta) = r < \infty$, and $\liminf_j r_j \ge r$, then T is minimax.

Ex: $X_i \sim N(\mu,\sigma^2)$, $\theta = (\mu,\sigma^2)$ is unknown under SEL:

First consider $\Theta = \mathbb{R}\times(0,c]$, $\Theta_0 = \mathbb{R}\times\{c\}$. On Θ_0 : $R_{\bar{X}} = \frac{c^2}{n} = r$ constant.

Recall: $\mu \sim N(0,j)$, $r_j = \frac{c^2 j}{nj+c^2} \to r$. Hence, $\bar{X}$ is minimax on $\Theta_0 \Rightarrow$ also on Θ.

If $\Theta = \mathbb{R}\times(0,\infty)$ sup goes to ∞, it's meaningless.

Hypothesis Testing

Sample $X_{1:n} \sim P \in \mathcal{P}$. Test: $H_0 : P \in \mathcal{P}_0$ $v.s.$ $H_1 : P \in \mathcal{P}_1$, $\mathcal{P}_0 \subset \mathcal{P}$, $\mathcal{P}_1 = \mathcal{P}\backslash\mathcal{P}_0$

If $\mathcal{P}_0$ contains only 1 element, we call it simple null hypothesis; otherwise composite.

Test is a statistic $T(X)$ takes value in $[0,1]$. When $X = x$ is observed, we reject H_0 with probability $T(x)$. If $T(X) \in \{0,1\}$ $a.s.\mathcal{P}$, it's non-randomized.

Errors: Type I error $\mathbb{P}[reject\ H_0|H_0\ is\ true]$ i.e. $\mathbb{E}_0[T]$

Type II error $\mathbb{P}[accept\ H_0|H_1\ is\ true]$ i.e. $\mathbb{E}_1[1-T]$

Power function: $\beta_T(P) = \mathbb{E}_P[T(X)]$, it is a function of $P \in \mathcal{P}$

Level α: $\sup_{P\in\mathcal{P}_0}\beta_T(P) \le \alpha$ Size α: $\sup_{P\in\mathcal{P}_0}\beta_T(P) = \alpha$

Uniformly Most Powerful Test (UMP)

Test T_* of size α is a UMP test iff $\beta_{T_*}(P) \ge \beta_T(P)$ for all $P \in \mathcal{P}_1$ and T of level α.

If $U(X)$ is a suff statistic for $P \in \mathcal{P}$, for any test $T(X)$, $E[T|U]$ is a test, with same power function as T. As $\mathbb{E}[E[T|U]] = \mathbb{E}[T]\,\forall P$. To find UMP, consider $\psi(U)$ only.

Neyman-Pearson lemma: Let $\mathcal{P}_0 = \{P_0\}$, $\mathcal{P}_1 = \{P_1\}$, f_j be the pdf of P_j

(i) (Existence). For every α, there exists a UMP test of size α, given by T_*, where $\gamma \in [0,1]$, $c \ge 0$ is to be determined as $\mathbb{E}_0[T_*(X)] = \alpha$.

$$T_*(X) = \begin{cases} 1 & f_1(X) > cf_0(X) \\ \gamma & f_1(X) = cf_0(X) \\ 0 & f_1(X) < cf_0(X) \end{cases} \qquad T_{**}(X) = \begin{cases} 1 & f_1(X) > cf_0(X) \\ 0 & f_1(X) < cf_0(X) \end{cases}$$

(ii) (Uniqueness). If $T_{**}(X)$ is a UMP test of size α, then $\uparrow$ $a.s.\mathcal{P}$

Can only differ on $B = \{x : f_1(x) = cf_0(x)\}$. T_* is the simplest form of randomized.

Remarks: · Both null and alternative are simple

· UMP exists, and unique except on B · If $\nu(B) = 0 \Rightarrow$ unique UMP

· If $\nu(B) > 0 \Rightarrow$ Random, but can be constant γ on B; · $\lambda = \frac{f_1(X)}{f_0(X)}$ is suff.

Proof: Assume $\alpha \in (0,1)$.

(1) γ, c exist: $\mathbb{E}_0T_* = \mathbb{P}_0[f_1(X) > cf_0(X)] + \gamma\mathbb{P}_0[f_1(X) = cf_0(X)]$. Let $\gamma(t) = \mathbb{P}_0[f_1(X) > tf_0(X)]$ it is non-increasing with $\gamma(0) = 1$, $\gamma(\infty) = 0$. Thus $\exists c \in (0,\infty)$ s.t. $\gamma(c) \le \alpha \le \gamma(c-)$. Set $\gamma = \frac{\alpha-\gamma(c)}{\gamma(c-)-\gamma(c)}I_{(\gamma(c-)\neq\gamma(c))}$. Note that $\gamma(c-) - \gamma(c) = \mathbb{P}_0[f_1(X) = cf_0(X)]$. Such γ, c satisfy.

(2) T_* is UMP: another T s.t. $\mathbb{E}_0T \le \alpha$: As $T_* > T \Rightarrow T_* > 0, f_1(X) \ge cf_0(X)$, and $T_* < T \Rightarrow T_* < 1, f_1(X) \le cf_0(X)$, there is $[T_* - T][f_1(X) - cf_0(X)] \ge 0$

$\int[T_* - T]f_1 d\nu = \beta_{T_*}(1) - \beta_T(1) \ge c\int[T_* - T]f_0 d\nu = c[\beta_{T_*}(0) - \beta_T(0)] \ge 0$

(3) Uniqueness: Define $A = \{x : f_1(x) \neq cf_0(x)\}$. Similarly $[T_* - T][f_1(X) - cf_0(X)] > 0$ when $X \in A$, $[\cdot][\cdot] = 0$ when $X \notin A$. As both UMP test with size α: $\int[T_* - T][f_1(X) - cf_0(X)]d\nu = \beta_{T_*}(1) - \beta_T(1) - c[\beta_{T_*}(0) - \beta_T(0)] = 0$.

$\Rightarrow \nu(A) = 0 \Rightarrow T_* \neq T_{**}$ only on B □.

Procedure: $\lambda = f_1(X)/f_0(X) \to$ if λ monotone in $U(X) \to U < d$ or $U > d$ instead.

Ex: X is a sample of size 1, $P_0 = N(0,1)$, P_1: $e^{-|x|/2}/4$. UMP test of level $\alpha < 1/3$: $\lambda(X) = \sqrt{\frac{\pi}{8}}exp\{\frac{x^2-|x|}{2}\}$ monotone in $(|x| - \frac{1}{2})^2$ As it is continuous, $\nu(B) = 0$. Then $\lambda > c \Leftrightarrow |x| > t$ or $|x| < 1 - t$.

Ex: $X_i \sim Bernoulli(p)$, $H_0 : p = p_0$ $v.s.$ $H_1 : p = p_1$, where $0 < p_0 < p_1 < 1$:

$f(x_{1:n;p}) = p^{\sum x}(1-p)^{n-\sum x}$ Let $Y = \sum X$, $\lambda = (\frac{p_1}{p_0})^Y(\frac{1-p_1}{1-p_0})^{n-Y}$ increase in Y.

Find γ, m s.t. $\alpha = \mathbb{P}_0[Y > m] + \gamma\mathbb{P}_0[Y = m]$

Remark: · T_* relies on p_0 only, not on p_1.

· For any $p_1 > p_0$, the test T_* has level α, and it is a UMP test for $H_1 : p = p_1$

· Therefore T_* is a UMP test for testing $H_0 : p = p_0$ $v.s.$ $H_1 : p > p_0$

Lemma: Suppose that there is a test T_* of size α s.t. for every $P_1 \in \mathcal{P}_1$, T_* is UMP for testing H_0 versus the hypothesis $P = P_1$. Then T_* is UMP for testing H_0 $v.s.$ H_1.

Extend to a family: $\downarrow$ satisfy this.

Monotone Likelihood Ratio Family: Suppose $X \sim P_\theta$ with $\theta \in \Theta$. Suppose that $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ is dominated by a measure ν, with PDF $f_\theta = dP_\theta/d\nu$. For a statistic $Y(X)$, $\mathcal{P}$ has monotone likelihood ratio in $Y(X)$ iff, for any $\theta_1 < \theta_2$:

On $supp(f_{\theta_1}) \cup supp(f_{\theta_2})$, $f_{\theta_2}(x)/f_{\theta_1}(x)$ is a non-decreasing function of $Y(x)$.

Remark: When monotone in Y, UMP given by NP-Lemma, can be defined by $Y > d$ and the calculation is based on θ_1, doesn't depends on θ_2.

Lemma: Suppose $X \sim P_\theta$ with $\theta \in \Theta$, and $\mathcal{P}$ has monotone likelihood ratio in $Y(X)$.

If ψ is a non-dec. function of Y , then $g(\theta) = \mathbb{E}[\psi(Y)]$ is a non-dec. function of θ.

Proof: Let $\theta_1 < \theta_2$, $h(y(x)) = f_{\theta_2}(x)/f_{\theta_1}(x)$, Let $A = \{x : f_{\theta_1}(x) > f_{\theta_2}(x)\} = \{x : h(y(x)) < 1\}$, $B = \{x : f_{\theta_1}(x) < f_{\theta_2}(x)\} = \{x : h(y(x)) > 1\}$. Since $h(y)$ is non-decreasing in y, $a = \sup_{x\in A}\psi(Y(x)) \le b = \inf_{x\in B}\psi(Y(x))$. $g(\theta_2) - g(\theta_1)$ is $\int\psi(Y(x))(f_{\theta_2}(x) - f_{\theta_1}(x))d\nu \ge a\int_A(f_{\theta_2}(x) - f_{\theta_1}(x))d\nu + b\int_B(f_{\theta_2}(x) - f_{\theta_1}(x))d\nu = (b-a)\int_B(f_{\theta_2}(x) - f_{\theta_1}(x))d\nu \ge 0$, as $\int_A(f_{\theta_2}(x) - f_{\theta_1}(x))d\nu + \int_B(f_{\theta_2}(x) - f_{\theta_1}(x))d\nu = \int f_{\theta_2}(x)d\nu - \int f_{\theta_1}(x)d\nu = 0$. □

Ex: Let $\theta \in \Theta \subset \mathbb{R}$, $\eta(\theta)$ non-decreasing function of θ. Then the one-parameter exponential family with $f_\theta(x) = exp\{\eta(\theta)T(x) - A(\theta)\}h(x)$ has monotone likelihood ratio in $T(X)$.

Ex: $X_i \sim Uni(0,\theta)$ where $\theta > 0$ PDF of $X_{1:n}$ is $f_\theta(x) = \theta^{-n}I_{(0,\theta)}(x_{(n)})$, for $\theta_1 < \theta_2$, only need to consider $(0,\theta_1) \cup (0,\theta_2) = (0,\theta_2)$ $f_{\theta_2}(x)/f_{\theta_1}(x)$ is non-dec. in $x_{(n)}$.

Theorem: UMP of Monotone Likelihood Ratio Family:

Suppose $X \sim P_\theta$ with $\theta \in \Theta$, and $\mathcal{P}$ has monotone likelihood ratio in $Y(X)$. Consider the testing $H_0 : \theta \le \theta_0$ $v.s.$ $H_1 : \theta > \theta_0$, where θ_0 is a given constant.

(1) There exists a UMP test of size α, which is given by $T_*(X) = \begin{cases} 1 & Y(X) > c \\ \gamma & Y(X) = c \\ 0 & Y(X) < c \end{cases}$ c and γ are from $\beta_{T_*}(\theta_0) = \alpha$, and $\beta_T(\theta) = \mathbb{E}_\theta T$ is the power function of a test T.

(2) $\beta_{T_*}(\theta)$ is strictly increasing for all θ's for which $0 < \beta_{T_*}(\theta) < 1$.

(3) For any $\theta < \theta_0$, T_* minimizes $\beta_T(\theta)$ (type I error of T) among T s.t. $\beta_T(\theta_0) = \alpha$.

(4) For any fixed θ_1, T_* is UMP for $H_0 : \theta \le \theta_1$ $v.s.$ $H_1 : \theta > \theta_1$, with size $\beta_T(\theta_1)$.

(5) Assume that $\mathbb{P}_\theta[f_\theta(X) = cf_{\theta_0}(X)] = 0$, for any $\theta > \theta_0$ and $c \ge 0$. If T is a test with $\beta_T(\theta_0) = \beta_{T_*}(\theta_0)$, then for $\forall\theta > \theta_0$ either $\beta_T(\theta) < \beta_{T_*}(\theta)$ or $T = T_*a.s.\mathcal{P}$

Remark: · optimal: $\theta < \theta_0$ minimize Type I; $\theta > \theta_0$ minimize Type II

· Uniqueness: When $\mathbb{P}_\theta[f_\theta(X) = cf_{\theta_0(X)}] = 0$ holds for any $\theta < \theta_0$ and $c > 0$, and the power at $\theta = \theta_0$ are equal.

Proof: (1) T_* is UMP for $H_0 : \theta = \theta_0$ $v.s.$ $H_1 : \theta > \theta_0$ from Lemma above, and β_{T_*} is non-decreasing in θ as T_* is non-decreasing in Y (Another Lemma). $\Rightarrow T_*$ is size α on $\{\theta \le \theta_0\}$. Meanwhile any level α, T, for $H_0 : \theta \le \theta_0$ $v.s.$ $H_1 : \theta > \theta_0$ is also level α, T, for $H_0 : \theta = \theta_0$ $v.s.$ $H_1 : \theta > \theta_0$. As T_* UMP in the " = " test $\Rightarrow$ more powerful on $\Theta_1 \Rightarrow$ also UMP of the " $\le$ " test.

(2)

(3)The result can be proved using Neyman-Pearson lemma with all inequalities reversed.

(4) Similar to (1)

(5)

Ex $X_i \sim Ubif(0,\theta)$ $\theta > 0$. Testing $H_0 : \theta \le \theta_0$ $v.s.$ $H_1 : \theta > \theta_0$

$Y = X_{(n)}$, monotone likelihood ratio, UMP is T_*, $\alpha = \beta_{T_*}(\theta_0) = \frac{n}{\theta_0^n}\int_c^{\theta_0}x^{n-1}dx = 1 - c^n\theta_0^{-n} \Rightarrow c = \theta_0(1-\alpha)^{1/n}$. For $\theta > \theta_0$, $\beta_{T_*}(\theta) = 1 - \theta_0^n\theta^{-n}(1-\alpha)$. Another test $T = \alpha I_{(X_{(n)}\le\theta_0)} + I_{(X_{(n)}>\theta_0)}$. Same power function when $\theta > \theta_0$.

As in this case $\mathbb{P}\{f_{\theta_1} = f_{\theta_0}\} = 1$, is not contradictory with the unique lemma.

One Parameter Exponential Family:

$f_\theta(x) = exp\{\eta(\theta)T(x) - A(\theta)\}h(x)$, η is strictly monotone function of θ.

· If η is increasing, then T_* given by Monotone Likelihood Ratio Theorem is UMP for testing $H_0 : \theta \le \theta_0$ $v.s.$ $H_1 : \theta > \theta_0$

· If η is decreasing or test is $H_0 : \theta \ge \theta_0$ $v.s.$ $H_1 : \theta < \theta_0$ the result is still valid by reversing inequalities in definition of T_*.

Ex: $X_i \sim N(\mu,\sigma^2)$, $\mu \in \mathbb{R}$ unknown, σ^2 is known. $H_0 : \mu \le \mu_0$ $v.s.$ $H_1 : \mu > \mu_0$:

$Y = \bar{X}$, $\eta = \frac{n\mu}{\sigma^2} \Rightarrow T_* = I_{(\bar{X}>C_\alpha)} \Rightarrow C_\alpha = \sigma Z_{1-\alpha}/\sqrt{n} + \mu_0$, where $Z_\alpha = \Phi^{-1}(\alpha)$.

· Discuss: dist of Y is needed. If it is continuous the test is non-randomized.

Ex: $X_i \sim Poisson(\theta)$ with unknown $\theta > 0$, $H_0 : \theta \le \theta_0$ $v.s.$ $H_1 : \theta > \theta_0$:

$Y = \sum X \sim Poisson(n\theta)$, $\eta = log\,\theta \nearrow \alpha = \sum_{j=c+1}^\infty \frac{e^{n\theta_0}(n\theta_0)^j}{j!} + \gamma\frac{e^{n\theta_0}(n\theta_0)^c}{c!}$, if $\alpha = \sum_{j=c+1}^\infty[e^{n\theta_0}(n\theta_0)^j/j!]$ for some integer c it is non-randomized.

Two Sided Tests: For fixed θ_0, $\theta_1 < \theta_2$:

(1)$H_0 : \theta \le \theta_1$ or $\theta \ge \theta_2$	v.s.	$H_1 : \theta_1 < \theta < \theta_2$	UMP in 1-para exp.
(2)$H_0 : \theta_1 \le \theta \le \theta_2$	v.s.	$H_1 : \theta > \theta_1$ or $< \theta_2$	Only UMPU
(3)$H_0 : \theta = \theta_0$	v.s.	$H_1 : \theta \neq \theta_0$	Only UMPU

Generalized Neyman-Pearson lemma: Define the class of tests:

Let $f_1, \cdots, f_{m+1}$ be measurable on $(\mathbb{R}^p, \mathcal{B})$ and also integrable w.r.t a measure ν. For given constants $t_1, \cdots, t_m$ let $\mathcal{T}$ be the class of measurable functions $\phi : \mathbb{R}^p \to [0,1]$ satisfying $\int\phi f_i d\nu \le t_i$, $i = 1, \cdots, m$, and $\mathcal{T}_0$ be the set of ϕ's in $\mathcal{T}$ satisfying the condition with all inequalities replaced by equalities.

Generalized Neyman-Pearson lemma: Result:

If there are constants $c_1, \cdots, c_m$ s.t.

$$\phi_*(x) = \begin{cases} 1 & f_{m+1}(x) > c_1f_1(x) + \cdots + c_mf_m(x) \\ 0 & f_{m+1}(x) < c_1f_1(x) + \cdots + c_mf_m(x) \end{cases}$$

is a member of $\mathcal{T}_0$, then ϕ_* maximizes $\int\phi f_{m+1}d\nu$ over $\psi \in \mathcal{T}_0$. If $c_i \ge 0$ for all i, ϕ_* maximizes $\int\phi f_{m+1}d\nu$ over $\psi \in \mathcal{T}$

Lemma: $f_1, \cdots, f_{m+1}$ and ν given by the generalized Neyman-Pearson lemma. Then the set $M = \{(\int\phi f_1 d\nu, \cdots, \int\phi f_m d\nu) : \phi : \mathbb{R}^p \to [0,1]\}$ is convex and closed. If $t_1, \cdots, t_m$ is an interior point of M, then there exist constants $c_1, \cdots, c_m$ s.t. the function $\phi_*(x)$ defined in the generalized Neyman-Pearson lemma is in $\mathcal{T}_0$.

Proof: Suppose $\phi_* \in \mathcal{T}_0$, $\forall\phi \in \mathcal{T}_0$ $(\phi_* - \phi)(f_{m+1} - \sum c_if_i) \ge 0$

Therefore $\int(\phi_* - \phi)(f_{m+1} - \sum c_if_i)d\nu \ge 0 \Rightarrow \int(\phi_* - \phi)f_{m+1}d\nu \ge c_i\int(\phi_* - \phi)f_id\nu$.

Hence, ϕ_* maximizes $\int\phi f_{m+1}d\nu$ over $\psi \in \mathcal{T}_0$, If $c_i > 0$ the first line still holds. □

UMP Tests for Two-Sided Hypothesis:

$X \sim f_\theta(x) = exp\{\eta(\theta)T(x) - A(\theta)\}h(x)$, 1-parameter exp. family.

(a) For hypothesis (1), a size α UMP is given as following, where c_i, γ_i are s.t. $\alpha = \beta_{T_*}(\theta_1) = \beta_{T_*}(\theta_2)$.

$$T_*(X) = \begin{cases} 1 & c_1 < Y(X) < c_2 \\ \gamma_i & Y(X) = c_i,\ i = 1, 2 \\ 0 & Y(X) < c_1 \ or\ c2 > Y(X) \end{cases}$$

(b) T_* minimizes $\beta_T(\theta)$ over $\theta < \theta_1$, $\theta > \theta_2$ and T s.t. $\alpha = \beta_T(\theta_1) = \beta_T(\theta_2)$

(c) If T_* and T_{**} are two tests given by (a), $\beta_{T_*}(\theta_1) = \beta_{T_{**}}(\theta_1)$, and if the region $\{T_{**} = 1\}$ is to the right of $\{T_* = 1\}$, then $\beta_{T_*}(\theta) < \beta_{T_{**}}(\theta)$ for $\theta > \Theta_1$ and $\beta_{T_*}(\theta) > \beta_{T_{**}}(\theta)$ for $\theta < \Theta_1$. If both T_* and T_{**} satisfy (a) and have power α at $\theta = \theta_1, \theta_2$, then $T_* = T_{**} a.s.P$.

Proof: (a) generalized Neyman-Pearson Lemma above. Start from $H_0 : \theta = \theta_1, \theta_2$ $v.s.$ $H_1 : \theta = \theta_3$, where $\theta_1 < \theta_3 < \theta_2$. $(\beta_t(\theta_1), \beta_t(\theta_1))$ is interior point. $\Rightarrow \tilde{c_1}, \tilde{c_2} \cdots T_*$ based on Y doesn't depends on θ_3 $\Rightarrow$ From 3 points$\rightarrow$ testing (1).

(b) Consider $\theta_3 < \theta_1$ similar with above. And $\theta_3 > \theta_2 \cdots$

Uniformly Most Powerful Unbiased (UMPU) Tests:

Given α. Test T for $H_0 : P \in \mathcal{P}_0$ $v.s.$ $H_1 : P \in \mathcal{P}_1$ is unbiased of level α iff

$$\beta_T(P) \leq \alpha,\ P \in \mathcal{P}_0 \quad and \quad \beta_T(P) \geq \alpha,\ P \in \mathcal{P}_1$$

A test of size α is UMPU iff it is UMP among the unbiased tests of level α.

Similarity: hypothesis: $H_0 : \theta \in \Theta_0$ $v.s.$ $H_1 : \theta \in \Theta_1$. Let α be a given level of significance, and $\bar{\Theta}_{01}$ be the common boundary of Θ_0 and Θ_1. i.e. common limit points of Θ_0 and Θ_1.

A test T is similar on $\bar{\Theta}_{01}$ if and only if $\beta_T(\theta) = \alpha$ for all $\theta \in \bar{\Theta}_{01}$.

Remark: · Transform P to θ to make it easier to find the boundary;

· Unbiased are usually similar. Work with similar tests are much easier.

Continuity of the power function: $\beta_T(\theta)$ is continuous in θ iff $\forall\{\theta_j\}_{j=1}^{\infty} \subset \Theta$, $\theta_j \to \theta$ implies $\beta_T(\theta_j) \to \beta_T(\theta)$, where $P_j \in \mathcal{P}$ and $\theta_j = \theta(P_j)$.

If parametric, β_T is just a function of θ, the continuous is just that of $\beta_T(\theta)$.

Lemma: Hypothesis: $H_0 : \theta \in \Theta_0$ $v.s.$ $H_1 : \theta \in \Theta_1$ Suppose that, for every T, $\beta_T(P)$ is continuous in θ. If T_* is uniformly most powerful among all similar tests and has size α, then T_* is a UMPU test.

Proof: continuous: $\{unbiased\} \subset \{similar\}$, T_* is size α and more powerful than fixed test $T = \alpha \Rightarrow T_*$ unbiased and UMP in a larger set $\Rightarrow$ UMPU □

Neyman structure: Let $U(X)$ be a suff statistic for the boundary $\bar{\mathcal{P}} = \{P : \theta \in \bar{\Theta}_{01}\}$ and let $\bar{P}_U$ be the distribution of U for $P \in \bar{\mathcal{P}}$. Test T is said to have Neyman structure w.r.t. U if $\mathbb{E}[T|U] = \alpha$, $a.s.\bar{\mathcal{P}}$.

· If T has Neyman structure, $\mathbb{E}T = \mathbb{E}[\mathbb{E}(T|U)] = \alpha\ \forall P \in \bar{\mathcal{P}} \Rightarrow T$ is similar on $\bar{\Theta}_{01}$.

· If all tests similar on $\bar{\Theta}_{01}$ have Neyman structure w.r.t. U, then working with tests having Neyman structure is the same as working with tests similar on $\bar{\Theta}_{01}$.

Lemma 6.6: Let $U(X)$ be a sufficient and complete statistic for $P \in \bar{\mathcal{P}}$, then all tests similar on $\bar{\Theta}_{01}$ have Neyman structure w.r.t. U.

Theorem: UMPU tests in exponential families:

In Exp family with PDF $f_{\theta,\phi}(x) = exp\{\theta Y(x) + \phi^T U(x) - \zeta(\theta, \phi)\}$

θ is real valued, ϕ can be a vector, $Y \in \mathbb{R}$ and vector U are statistics.

(1) For test $H_0 : \theta \leq \theta_0$ $v.s.$ $H_1 : \theta > \theta_0$ a UMPU test of size α is given as:

$$T_*(Y, U) \quad = \quad \begin{cases} 1 & Y > c(U) \\ \gamma(U) & Y = c(U) \\ 0 & Y < c(U) \end{cases}$$

where $c(u)$ and $\gamma(u)$ are Borel functions s.t $\mathbb{E}_{\theta_0}[T_*|U = u] = \alpha$ for each u

(2) For test $H_0 : \theta \leq \theta_1$ or $\theta_0 \geq \theta_2$ $v.s.$ $H_1 : \theta_1 < \theta < \theta_2$ a UMPU test of size α is given as:

$$T_*(Y, U) = \begin{cases} 1 & c_1(U) < Y < c_2(U) \\ \gamma_i(U) & Y = c_i(U) \\ 0 & otherwise \end{cases}$$

where $c(u)$ and $\gamma(u)$ are Borel functions s.t $\mathbb{E}_{\theta_1}[T_*|U = u] = \mathbb{E}_{\theta_2}[T_*|U = u] = \alpha$ for each u

(3) For test $H_0 : \theta_1 \leq \theta \leq \theta_2$ $v.s.$ $H_1 : \theta < \theta_1$ or $\theta > \theta_2$ a UMPU test of size α is given as:

$$T_*(Y, U) = \begin{cases} 1 & otherwise \\ \gamma_i(U) & Y = c_i(U) \\ 0 & c_{1(U)} < Y < c_{2(U)} \end{cases}$$

where $c(u)$ and $\gamma(u)$ are Borel functions s.t $\mathbb{E}_{\theta_1}[T_*|U = u] = \mathbb{E}_{\theta_2}[T_*|U = u] = \alpha$ for each u

(4) For test $H_0 : \theta = \theta_0$ $v.s.$ $H_1 : \theta \neq \theta_0$ a UMPU test of size α is given as that in (3), but with $\mathbb{E}_{\theta_0}[T_*|U = u] = \alpha$ and $\mathbb{E}_{\theta_0}[T_* Y|U = u] = \alpha\mathbb{E}_{\theta_0}[Y|U = u]$ for each u.

Remark: This result only for Exp family, and no Uniqueness is assured.

Proof: Given $U = u$, $Y \sim f()$ is 1-parameter. in (1)-(4) find $\bar{\Theta}_{01}$, and U comp & suff on it $\Rightarrow$ Neyman structure $\Rightarrow$ UMP among them $\Leftrightarrow$ UMP among similar $\Rightarrow$ UMPU

Ex1 Poisson: $X_1 \sim P(\lambda_1)$, $X_2 \sim P(\lambda_2)$, rewrite the density as:

$p = \frac{exp\{-(\lambda_1+\lambda_2)\}}{x_1!x_2!} exp\{x_2 log\frac{\lambda_2}{\lambda_1} + (x_1 + x_2) log\,\lambda_1\}$, and let $\theta = log\frac{\lambda_2}{\lambda_1}$ $Y = X_2$

Test $H_0 : \lambda_1 = \lambda_2$ $v.s.$ $H_1 : \lambda_1 \neq \lambda_2 \Leftrightarrow H_0 : \theta = 0$ $v.s.$ $H_1 : \theta \neq 0$, with $U = X_1 + X_2$, $\phi = log\,\lambda_1$ Find a UMPU: $\mathbb{P}[Y = y|U = u] = \binom{u}{y} p^y (1-p)^{u-y}$ where $p = \frac{e^\theta}{1+e^\theta}$ on the boundary $\theta = 0$ dist. of Y is known. We can find the UMPU.

Ex2: Binomial: $X_1 \perp\!\!\!\perp X_2$, $X_i \sim Binomial(n_i, p_i)$, n_1, n_2 are known, p_1, p_2 not. PMF is $\binom{n_1}{x_1}\binom{n_2}{x_2}(1-p_1)^{n_1}(1-p_2)^{n_2} exp\{x_2 log\frac{p_2(1-p_1)}{p_1(1-p_2)} + (x_1 + x_2) log\frac{p_1}{p_1+p_2}\}$

Hence, $\theta = \frac{p_2(1-p_1)}{p_1(1-p_2)}$, $Y + X_2$, $U = X_1 + X_2$. Transform the test into θ UMPU can be found.

Remark: UMP amd UMPU are very good, but may not exist is some cases.$\Rightarrow$ Find some test not bad and always exist.

Likelihood Ratio Test: Let $L(\theta) = f_\theta(X)$ be the likelihood function. The likelihood ratio is defined as $\lambda(X) = \sup_{\theta \in \Theta_0} L(\theta) / \sup_{\theta \in \Theta} L(\theta)$, where X is the sample with some size n. For test $H_0 : \theta \in \Theta_0$ $v.s.$ $H_1 : \theta \in \Theta_1$.

Likelihood ratio (LR) test is any test that rejects H_0 iff $\lambda(X) < c$, where $c \in [0, 1]$.

Remark: · If $\lambda(X)$ is well defined, then $\lambda(X) \leq 1$, and tends to 1 if H_0 is true;

· Let $\hat{\theta}$ be the MLE of θ, and $\hat{\theta}_0$ be MLE on Θ_0. Then $\lambda(X) = L(\hat{\theta}_0)/L(\hat{\theta})$;

· For given α if $\exists c_\alpha$ s.t. $\sup_{\theta \in \Theta_0} \mathbb{P}_\theta[\lambda(X) < c_\alpha] = \alpha$, size α test can be defined.

Properties: When a UMP or UMPU test exists, an LR test is often the same.

Suppose that X is in a 1-parameter exp family: $f_\theta(x) = \exp\{\eta(\theta)Y(x) - A(\theta)\}h(x)$, where η is a strictly increasing and differentiable function of θ.

(1) For test $H_0 : \theta \leq \theta_0$ $v.s.$ $H_1 : \theta > \theta_0$ there is an LR test whose rejection region is the same as that of the UMP test.

(2) For test $H_0 : \theta \leq \theta_1$ or $\theta_0 \geq \theta_2$ $v.s.$ $H_1 : \theta_1 < \theta < \theta_2$ there is an LR test whose rejection region is the same as that of the UMP test.

(3) For testing the other two-sided hypotheses, there is an LR test whose rejection region is equivalent to $Y(X) < c_1$ or $Y(X) > c_2$ for some constants c_1 and c_2.

Proof: Prove (1) only. (2) and (3) are very similar.

Let $\hat{\theta}$ be be the MLE of $\theta \in \Theta$. Recall for exp family dist, $\frac{\partial}{\partial\eta} log\,L(\theta) = Y(x) - B'(\eta)$, which is a strictly decreasing function shown before. Therefore, the MLE exists and is unique for η. Since η is strictly increasing of θ, so the MLE of θ exists and unique. And $L(\theta) \nearrow$ if $\theta < \hat{\theta}$, $L(\theta) \searrow$ if $\theta > \hat{\theta}$. Thus $\lambda = 1$ if $\hat{\theta} \leq \theta_0$ and $\lambda = L(\theta_0)/L(\hat{\theta})$ if $\hat{\theta} > \theta_0$. Then $\lambda < c$ is same as $\hat{\theta} > \theta_0$ and $L(\theta_0)/L(\hat{\theta}) < c$.

As $\hat{\eta}$ is s.t. $Y(X) = B'(\eta)$, with B' strictly $\nearrow \Rightarrow \hat{\eta} \nearrow$ in $Y \Rightarrow \hat{\theta} \nearrow$ in Y

Consequently, for any θ_0 $\frac{d}{dY}[log\,L(\hat{\theta}) - log\,L(\theta_0)] = \eta(\hat{\theta}) - \eta(\theta_0)$. Thus,

$log\,[L(\theta_0)/L(\hat{\theta})] \nearrow$ in Y when $\hat{\theta} > \theta_0$, and $\searrow$ when $\hat{\theta} \leq \theta_0$.

Hence, for any $c \in (0, 1)$ $\lambda < c \Leftrightarrow \hat{\theta} > \theta_0$ and $L(\theta_0)/L(\hat{\theta}) < c \Leftrightarrow Y > d$ □

Ex: $X_i \sim Uni(0\,\theta)$, $H_0 : \theta = \theta_0$ $v.s.$ $H_1 : \theta \neq \theta_0$: $\lambda(X) = [X_{(n)}/\theta_0]^n I_{(X_{(n)} \leq \theta_0)}$

Reject when $\lambda < c \Leftrightarrow X_{(n)} > \theta_0$ or $X_{(n)} < c^{-n}\theta_0$. Take $c = \alpha$ it's size α.

Ex: Normal Linear Models: $X \sim N_n(Z\beta, \sigma^2 I_n)$. $H_0 : L\beta = 0$ $v.s.$ $H_1 : L\beta \neq 0$: $\hat{\beta}_0$ is LSE under H_0, $\hat{\sigma}_0^2 = \|X - Z\hat{\beta}_0\|^2/n$. $sup_{\theta \in \Theta_0} L(\theta) = (2\pi\hat{\sigma}_0^2)^{-n/2} e^{-n/2}$

LR test is $\lambda = [\hat{\sigma}^2/\hat{\sigma}_0^2]^{n/2} = (\frac{\|X - Z\hat{\beta}\|^2}{\|X - Z\hat{\beta}_0\|^2})^{n/2}$. Select c s.t. $sup_{\theta \in \Theta_0} \mathbb{P}_\theta[\lambda < c] \leq \alpha$

Especially, we consider a two-sample problem. $n = n_1 + n_2$, $\beta = (\mu_1, \mu_2)$ and $Z = diag(J_{n_1}, J_{n_2})$ with $L = (1, -1)$ to test $\mu_1 = \mu_2$. $\lambda < c \Leftrightarrow |t| > c_0$:

$t(X) = \{(\bar{X}_1 - \bar{X}_2)/\sqrt{n_1^{-1} + n_2^{-1}}\}/\{[(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2]/(n_1 + n_2 - 2)\}$

Regularity conditions: Let $X_{1:n}$ iid from a PDF f_θ with a measure ν where $\theta \in \Theta$ and Θ is an open set in $\mathbb{R}^k$. The regularity conditions below for the asymptotic of MLE will be assumed:

(1) $f_\theta(x)$ is twice continuously differentiable in θ and s.t.: $\frac{\partial}{\partial\theta}\int g_\theta d\nu = \int \frac{\partial}{\partial\theta} g_\theta d\nu$. for $g_\theta = f_\theta(x)$ and $\partial f_\theta(x)/\partial\theta$;

(2) The Fisher information matrix, $I_1(\theta)$, based on X_1 is positive definite;

(3) For any given $\theta \in \Theta$, there exists a positive number c_θ and a positive function h_θ s.t. $\mathbb{E}h_\theta(X_1) < \infty$ and $\sup_{\gamma:\|\gamma - \theta\| < c_\theta} \|\frac{\partial^2 log\, f_\gamma(x)}{\partial\gamma\,\partial\gamma^T}\| \leq h_\theta(x)$ for all x in the range of X_1, where $\|A\| = \sqrt{tr(A^T A)}$ for any matrix A.

Thm: Asymptotic LRT: Regularity conditions ↑ hold, Suppose that $H_0 : \theta = g(\vartheta)$, where ϑ is a $(k - r)$-vector of unknown parameters and g is a continuously differentiable function from $\mathbb{R}^{k-r}$ to $\mathbb{R}^k$ with a full rank $\partial g(\vartheta)/\partial\vartheta$. Then, under H_0: $-2\log\lambda_n \to^d \chi_r^2$: $r = dim(\theta) - dim(\vartheta)$, and reject $\lambda_n < exp\{-\frac{1}{2}\chi_{r,\alpha}^2\}$

Wald Test: $H_0 : R(\theta) = 0$: $W_n = [R(\hat{\theta})]^T\{[C(\hat{\theta})]^T[I_n(\hat{\theta})]^{-1}C(\hat{\theta})\}^{-1}R(\hat{\theta})$. Where $C(\theta) = \partial R(\theta)/\partial\theta$, I_n is the Fisher Inf. Matrix of $X_{1:n}$, and $\hat{\theta}$ is the MLE or RLE.

Score Test: $R_n = [s_n(\tilde{\theta})]^T[I_n(\tilde{\theta})]^{-1}[s_n(\tilde{\theta})]$. Where $s_n(\theta) = \partial log L(\theta)/\partial\theta$ is the score function, and $\tilde{\theta}$ is an MLE or RLE under $H_0 : R(\theta) = 0$

Remark: They are asymptotically same, and reject when W, R is large.

Thm 6.6: Under regularity conditions:

(1) $R(\theta)$ continuously differentiable function from $\mathbb{R}^k$ to $\mathbb{R}^r$, $W_n \to^d \chi_r^2$, reject when $W_n > \chi_{r,\alpha}^2$, where $\chi_{r,\alpha}^2$ is the $1 - \alpha$ quantile of χ_r^2.

(2) Result for R_n is same with that of W_n above.

Confidence Set: Let X be sample from a population $P \in \mathcal{P}$. Let $\theta = \theta(P)$ be the parameter of interest. Let $C(X)$ be a random set determined by sample X. The random set $C(X)$ is said to be a confidence set for θ with confidence level $1 - \alpha$, or a level $1 - \alpha$ confidence set, if $\inf_{P \in \mathcal{P}} P[\theta \in C(X)] \geq 1 - \alpha$. The exact infimum $\inf_{P \in \mathcal{P}} P[\theta \in C(X)]$ is called the confidence coefficient of $C(X)$.

If $C(X)$ is of the form: $[\underline{\theta}(X), \bar{\theta}X]$, it's confidence interval.

$[\underline{\theta}(X), \infty)$, it's confidence lower bound. $(-\infty, \bar{\theta}X]$, it's confidence upper bound.

Construct a Confidence Interval:

Pivotal Quantity: If the dist. of $R(X, \theta)$ does not depend on θ, then it is a pivotal.

Thm: Pivotal Quantity: Suppose that $P \in \mathcal{P} = \{P_\theta\}$. Let $T(X)$ be a real-valued statistic with CDF $F_{T,\theta}(t)$ and let α_1 and α_2 be fixed positive constants s.t. $\alpha_1 + \alpha_2 = \alpha < 1/2$.

(1) Suppose $F_{T,\theta}(t)$ and $F_{T,\theta}(t-)$ are non-increasing in θ for each fixed t. Define: $\bar{\theta} = sup\{\theta : F_{T,\theta}(T) \geq \alpha_1\}$, and $\underline{\theta} = inf\{\theta : F_{T,\theta}(T-) \leq 1 - \alpha_2\}$. $[\underline{\theta}(X), \bar{\theta}X]$ is $1 - \alpha$ confidence interval;

(2) Suppose $F_{T,\theta}(t)$ and $F_{T,\theta}(t-)$ are non-decreasing in θ for each fixed t. Result is $\bar{\theta} = sup\{\theta : F_{T,\theta}(T-) \leq 1 - \alpha_2\}$, and $\underline{\theta} = inf\{\theta : F_{T,\theta}(T) \geq \alpha_1\}$;

(3) If continuous, $F_{T,\theta}(T)$ is a pivotal quantity. Result is same.

Proof: (1): $\theta > \bar{\theta} \Rightarrow F_{T,\theta}(T) < \alpha_1$, $\theta < \underline{\theta} \Rightarrow F_{T,\theta}(T-) > 1 - \alpha_2$. $\mathbb{P}[\underline{\theta} \leq \theta \leq \bar{\theta}] = 1 - \mathbb{P}[F_{T,\theta}(T) < \alpha_1] - \mathbb{P}[F_{T,\theta}(T-) > 1 - \alpha_2] \geq 1 - \alpha_1 - \alpha_2 = 1 - \alpha$. □

Ex: $X_i \sim Poisson(\theta)$, $T = \sum X \sim Poisson(n\theta)$ is comp and suff, and we can find $F_{T,\theta}(t) = \sum_{j=0}^{t} e^{-n\theta}(n\theta)^j/j!\ t = 0, 1, \cdots$, which is continuous in θ, $\bar{\theta}$ is the unique root of $F_{T,\theta}(T) = \alpha_1$. As $F_{T,\theta}(T-) = F_{T,\theta}(T - 1)$, $\underline{\theta}$ is the unique root of $F_{T,\theta}(T-1) = 1 - \alpha_2$ when $T > 0$ and $\underline{\theta} = 0$ when $T = 0$. As $\frac{1}{\Gamma(t)}\int_\lambda^\infty X^{t-1}e^{-x}dx = \sum_{j=0}^{t-1} e^{-\lambda}\lambda^j/j!$, $\bar{\theta} = (2n)^{-1}\chi^2_{2(T+1),\alpha_1}$, $\underline{\theta} = (2n)^{-1}\chi^2_{2(T),1-\alpha_2}$

Inverting acceptance regions of tests: Consider testing problem $H_0 : \theta = \theta_0$ $v.s.$ some H_1 T be a size α test, and the acceptance region is $A_T(\theta_0) = \{x : T(x) \neq 1\}$. For every $\theta \in \Theta$, $A_T(\theta)$ is a function from Θ to subsets of $\mathcal{X}$. "Inverse" $C(x) = \{\theta : x \in A_T(\theta)\}$. If all T_θ is level α, $C(x)$ is level $1 - \alpha$ CI.

The other direction: $C(X)$ be level $1 - \alpha$ CI, $A(\theta_0) = \{x : \theta_0 \in C(X)\}$ is subset of $\mathcal{X}$. $T = 1 - T_{A(\theta_0)}(X)$ is a level α test for H_0.

Ex: 1-parameter exp family: $f_\theta(x) = \exp\{\eta(\theta)Y(x) - A(\theta)\}h(x)$, $\eta \nearrow$ strictly.

· Testing $H_0 : \theta = \theta_0$ $v.s.$ $H_1 : \theta > \theta_0$ there is UMP T_* based on Y. accept set: $A(\theta_0) = \{x : Y(x) \leq c(\theta_0)\}$ $c(\theta)$ non-dec. in θ can be shown. Then $C(x) = \{\theta : c(\theta) \geq Y(x)\}$ is a lower bound. If Y is continuous, conf. coef. is $1 - \alpha$.

· For testing $H_0 : \theta = \theta_0$ $v.s.$ $H_1 : \theta < \theta_0$: Upper bound.

· For $H_0 : \theta = \theta_0$ $v.s.$ $H_1 : \theta \neq \theta_0$: Confidence Interval.

Evaluation: Better test should have better CI, but hard to say which is better.

Length Criterion: Consider CI's of a real-valued θ with the same conf. coef.

· The shorter the better · Uniformly shortest may not exists

· Find the best among a class of CI's.

Shortest CI for Pivotal: Consider real-valued parameter θ and statistic $T(X)$

(1) Let U be a positive statistic s.t. $(T - \theta)/U$ is a pivotal with pdf f that is unimodal at x_0. Consider CI's for θ: $\mathcal{C} = \{[T - bU, T - aU] : \int_a^b f dx = 1 - \alpha\}$. If $[T - b_* U, T - a_* U] \in \mathcal{C}$, with $f(a_*) = f(b_*) > 0, a_* < x_0 < b_*$, it's shortest in $\mathcal{C}$.

(2) Suppose that $T > 0$, $\theta > 0$, T/θ is a pivotal with PDF f, and that $x^2 f(x)$ is unimodal at x_0. Consider $\mathcal{C} = \{[T/b, T/a] : a, b > 0, \int_a^b f dx = 1 - \alpha\}$ If $[T/b_*, T/a_*] \in \mathcal{C}, a_*^2 f(a_*) = b_*^2 f(b_*) > 0, a_* < x_0 < b_*$, it's shortest in $\mathcal{C}$.

· Unimodal: non-decreasing when $x < x_0$, non-increasing when $x > x_0$

Proof: (1) length of CI in $\mathcal{C}$ is $(b - a)U$. When $a < b, b - a < b_* - a_*$, if $a < a_*$:

· $a < b \leq a_*$ by unimodal: $\int_a^b f dx \leq f(a_*)(b - a) < \int_{a_*}^{b_*} f dx = 1 - \alpha$

· $a \leq a_* < b < b_*$ and $a > a_*$ is similar. (2) change x to $1/y$ can be proved □

Ex: $X_i \sim N(\mu, \sigma^2)$, if σ^2 unknown $\sqrt{n}(\bar{X} - \mu)/S \sim t_{n-1}$ is the pivotal; if σ^2 known $\sqrt{n}(\bar{X} - \mu)/\sigma \sim N(0, 1)$. It is the shortest among that in $\mathcal{C}$.

UMA CI: Let $\theta \in \Theta$ be unknown parameter, and $\Theta' \subset \Theta$ where true $\theta \notin \Theta'$. $C(X)$ with conf coef $1 - \alpha$ is $\Theta'-$UMA iff for any other level $1 - \alpha$ set $C_1(X)$, $\forall\theta' \in \Theta'$,$\mathbb{P}[\theta' \in C(X)] \leq \mathbb{P}[\theta' \in C_1(X)]$. It is UMA iff $\Theta' = \{\theta\}^c$

Remark: Less prob. to cover false θ. For lower bound can use $\Theta' = \{\theta' \in \Theta : \theta' < \theta\}$

Thm UMA: $C(X)$ be conf set for θ by inverting acceptance regions of non-randomized tests T_{θ_0} for $H_0 : \theta = \theta_0$ $v.s.$ $H_1 : \theta \in \Theta_{\theta_0}$ where Θ_{θ_0} is a set related to θ_0.

If for each θ_0, T_{θ_0} is UMP of size α, then $C(X)$ is $\Theta'-$UMA with conf coef $1 - \alpha$, where $\Theta' = \{\theta' : \theta \in \Theta_{\theta'}\}$ region of θ' that reject true θ.

· In 1-para exp fam with MLR, UMP exists hence UMA exists.

Proof: Assume another level $1 - \alpha$ $C_1(X)$ test $T_{1\theta_0}(X) = 1 - T_{A_1(\theta_0)}(X)$ is also level α. For non-randomized UMP T: $\mathbb{P}[\theta' \in C] = 1 - \mathbb{P}[T_{\theta'} = 1] \leq 1 - \mathbb{P}[T_{1\theta'} = 1]$ □

UMAU CI: · Level $1 - \alpha$ conf set $C(X)$ is Θ'-unbiased iff $\mathbb{P}[\theta'] \leq 1 - \alpha$, $\forall\theta' \in \Theta'$.

· Let $C(X)$ be a Θ'-unbiased conf set with conf coef $1 - \alpha$ if for any other level $1 - \alpha$, Θ'-unbiased set $C_1(X)$, $\forall\theta' \in \Theta'$,$\mathbb{P}[\theta' \in C(X)] \leq \mathbb{P}[\theta' \in C_1(X)]$, it is $\Theta'-$UMAU

· $C(X)$ is UMAU iff $\Theta' = \{\theta\}^c$

Thm UMAU CI: $C(X)$ be conf set for θ by inverting AR of non-randomized tests T_{θ_0} for $H_0 : \theta = \theta_0$ $v.s.$ $H_1 : \theta \in \Theta_{\theta_0}$. If for each θ_0, T_{θ_0} is unbiased of size α, $C(X)$ is $\Theta'-$unbiased with conf coef $1 - \alpha$ where $\Theta' = \{\theta' : \theta \in \Theta_{\theta'}\}$.

If T_{θ_0} is also UMPU for each θ_0, $C(X)$ is $\Theta'-$UMAU.

· Proof is similar to UMA. Unbiased: always smaller prob. to cover false θ'.

Ex: Linear Model: $X \sim N(Z\beta, \sigma^2 T_n)$, with $\theta = a^T\beta$, with $a \in R(Z)$:

Non-rand. test AR is $A(\theta_0) = \{x : a^T\hat{\beta} - \theta_0 > t_{n-r,\alpha}\sqrt{a^T(Z^T Z)^- a\, SSR/(n - r)}\}$ is size α UMPU for $H_0 : \theta = \theta_0$ $v.s.$ $H_1 : \theta < \theta_0$. Inverting it, there is a Θ'-UMAU upper bound with conf coef $1 - \alpha$, and $\Theta' = \{\theta' : \theta \in \Theta_{\theta'}\} = \{\theta' : \theta < \theta'\} = (\theta, \infty)$.

The upper bound is $\bar{\theta} = a^T\hat{\beta} - t_{n-r,\alpha}\sqrt{a^T(Z^T Z)^- a\, SSR/(n - r)}$