$R(\theta,\delta)=E_\theta(L(\theta,\delta(X)))$.

Sufficiency: $P(X|T)$ free of θ.

NFFC: T sufficient for $\theta \iff P_\theta(x)=g_\theta(T(x))h(x)$.

$\Leftarrow$ Suppose $P_\theta(x)=g_\theta(T(x))h(x)\ \forall x\in\mathcal{X}$. Need to show that T is sufficient.

$$P_\theta(X=x\mid T(X)=t)=\frac{P_\theta(X=x,T(X)=t)}{P(T(X)=t)}=\begin{cases}0 & \text{if } T(X)\neq t\\ \frac{P(X=x)}{P(T(X)=t)} & T(X)=t\end{cases}$$

$$=\begin{cases}0 & \text{if } T(X)\neq t\\ \frac{\boxed{g_\theta(t)}h(x)}{\sum_{y\in\mathcal{X}:T(y)=t}\boxed{g_\theta(t)}h(y)} & T(X)=t\end{cases}$$

$$=\begin{cases}0 & \text{if } T(X)\neq t\\ \frac{h(x)}{\sum_{y\in\mathcal{X}:T(y)=t}h(y)} \perp\!\!\!\perp \theta & T(X)=t\end{cases}$$

$(X|T)$ has a distribution which is free of θ and so T is sufficient.

$\Rightarrow$ Suppose T is sufficient for θ, so

$$P_\theta(X=x)=P_\theta(X=x,T(X)=t)=\underline{P_\theta(X=x\mid T(X)=t)}\,P_\theta(T(X)=t)\ \hat{=}\ h(x)\,g_\theta(t)$$ as $P_\theta(X=x\mid T(X)=t)$ is free of θ by def. $\square$

Exp family: $P_\theta(x)=\exp\left(\sum_{i=1}^{s}\eta_i(\theta)T_i(x)-B(\theta)\right)h(x)$.

$\left(\sum_{j=1}^n T_1(x_j),\dots,\sum_{j=1}^n T_s(x_j)\right)$ is sufficient if $X_1,\dots,X_n\sim P_\theta(x)$

Canonical exp family: $P(X;\eta)=\exp\left(\sum_{i=1}^{s}\eta_iT_i(x)-A(\eta)\right)h(x)$

Natural parameter space. $\Theta_{nat}=\{\eta: 0<\int\exp(\sum_{i=1}^{s}\eta_iT_i(x))h(x)\,d\mu(x)<\infty\}$.

Full rank: $\mathcal{P}=\{P_\eta:\eta\in H\}$ is s-dim minimal exp. family. H contains an open s-dim rectangle. ($\{\eta_i\}$ and $\{T_i\}$ linearly independent)

Property 1: $G(f,\eta)=\int f(x)\exp\{\sum_{i=1}^{s}\eta_iT_i(x)\}h(x)\,d\mu(x)$. is infinitely differentiable w.r.t. η and the derivatives can be obtained by differentiating under the integral sign.

Property 2: Let $f(x)=1$, $G(f,\eta)=\exp(A(\eta))$

$$\frac{\partial G}{\partial\eta_i}=\frac{\partial A(\eta)}{\partial\eta_i}\exp(A(\eta));\quad \frac{\partial A(\eta)}{\partial\eta_i}=E_\eta\{T_i(x)\}.$$

$$\frac{\partial^2A(\eta)}{\partial\eta_i\partial\eta_j}=\mathrm{Cov}_\eta(T_i(X),T_j(X)).$$

- **Minimal suff.** T is minimal if for every suff. T', $\exists f$, s.t. $T(x)=f(T'(x))$ for any $x\in\mathcal{X}$.
- **Verify minimal:** T is minimal suff. if for $\forall x,y\in\mathcal{X}$ and θ, $P(x;\theta)=C(x,y)P(y;\theta)\iff T(x)=T(y)$.

Proof. *First prove that T is sufficient and then T is minimal.*

1. *(T is sufficient) For all $t\in T(X)$(the image of T), consider the preimage $A_t=T^{-1}(t)$. For each A_t, we denote x_t as a representative. Then for any $y\in X$, we have $y\in A_{T(y)}$ and $x_{T(y)}\in A_{T(y)}$. From the assumption of T, we have*

$$p(y;\theta)=c(y,x_{T(y)})p(x_{T(y)};\theta)=h(y)g_\theta(T(y)) \tag{13}$$

Therefore, by NFFC, T is sufficient.

2. *(T is minimal) Consider another sufficient statistic T'. By NFFC,*

$$p(x;\theta)=\tilde{g}_\theta(T'(x))\tilde{h}(x) \tag{14}$$

Take any x and y such that $T'(x)=T'(y)$, then

$$p(x;\theta)=\tilde{g}_\theta(T'(x))\tilde{h}(x)=\tilde{g}_\theta(T'(y))\tilde{h}(y)\frac{\tilde{h}(x)}{\tilde{h}(y)}=p(y;\theta)C(x,y) \tag{15}$$

By the assumption of T, $T(x)=T(y)$. Therefore, we've proved that for any sufficient statistics T' and any x and y, $T'(x)=T'(y)$ implies $T(x)=T(y)$. T is minimal.

- For any minimal s-dim exp. family, $\left(\sum_{i=1}^n T_1(X_i),\dots,\sum_{i=1}^n T_s(X_i)\right)$ is **complete sufficient.**
- A is ancillary if the dist. of $A(X)$ free of θ.
- A is first-order ancillary if $E_\theta(A(X))$ free of θ.
- T is complete if no non-constant function of T is first order ancillary. ($E_\theta(f(T(X)))=0$ for all θ, then $f(T(X))=0$ with prob 1 for all θ)
- If T is complete sufficient, then T is minimal sufficient.
- **Basu's Thm:** If T complete sufficient. V ancillary, then $T(X)\perp\!\!\!\perp V(X)$.

Proof 2 *Define $q_A(t)=P_\theta(V\in A\mid T=t)$ or $q_A(T)=P_\theta(V\in A\mid T)$ and $p_A=P_\theta(V\in A)$. By sufficiency and ancillarity, neither p_A nor $q_A(t)$ depends θ. By smoothing,*

$$(P_A=P_\theta(V\in A)=E_\theta(P_\theta(V\in A\mid T))=E_\theta(q_A(T))$$

and so by completeness, $q_A(T)=p_A$ a.e. for $\mathcal{P}$. Again, by smoothing/tower expecation,

$$\begin{aligned}P_\theta(T\in B,V\in A)&=E_\theta(1_B(T)1_A(V))\\&=E_\theta(E_\theta(1_B(T)1_A(v)\mid T))\\&=E_\theta(1_B(T)E_\theta(1_A(v)\mid T))\\&=E_\theta(1_B(T)q_A(T))\\&=E_\theta(1_B(T)\cdot p_A)\\&=P_\theta(T\in B)\cdot P_\theta(V\in A)\end{aligned}$$

Hence, T and V ane independent as A and B are arbitrary Borel sets.

Jensen: f convex $\Rightarrow f(E(x))\le E(f(x))$.

Rao-Blackwell: Suppose T is sufficient, $\delta(x)$ is an estimator of $g(\theta)$, $R(\theta,\delta)=E_\theta(L(\theta,\delta(x)))<\infty$. If $L(\theta,\cdot)$ is convex, then $R(\theta,\eta)\le R(\theta,\delta)$ for $\eta(T(x))=E(\delta(x)|T(x))$

By Jensen's inequality,

$$E_\theta\left(L(g(\theta),\delta(X))\mid T\right)\ge L\left(g(\theta),E_\theta(\delta(X)\mid T)\right)=L(g(\theta),\eta(T))$$

Taking another expectation gives

$$E_\theta\left(L(g(\theta),\delta(X))\right)\ge E_\theta\left(L(g(\theta),\eta(T))\right)$$

$$\iff R(g(\theta),\delta)\ge R(g(\theta),\eta)$$

- **Unbiased estimator:** $E_\theta(\delta(x))=g(\theta)\ \forall\theta\in\Theta$.
- $E_\theta(g(\theta)-\delta(X))^2=(E_\theta\delta(X)-g(\theta))^2+E_\theta(\delta(X)-E\delta(X))^2$
- $g(\theta)$ is U-estimable if unbiased estimator exists.

An unbiased estimator δ is UMVU if

$$\mathrm{Var}_\theta(\delta)\le\mathrm{Var}(\delta')\quad\forall\theta\in\Theta,$$

where δ' is any other competing unbiased estimator.

(Lehmann-Scheffé Theorem) If T is a complete and sufficient statistic, and $E_\theta(h(T(X)))=g(\theta)$, ie. $h(T(X))$ is unbiased for $g(\theta)$, then $h(T(X))$ is

1) the only function of $T(X)$ that is unbiased for $g(\theta)$
2) an UMRUE under any convex loss function
3) the unique UMRUE (hence UMVUE), up to a P-null set, under any strictly convex loss function.

Pf. 1) Suppose $E_\theta(\tilde{h}(T(X)))=g(\theta)$, then

$$E_\theta\left(\underbrace{\tilde{h}(T(X))-h(T(X))}_{f(T)\equiv0}\right)=0\quad\forall\theta\in\Theta$$

Thus, $\tilde{h}(T(X))=h(T(X))$ a.s. for all $\theta\in\Theta$ by completeness

2) Consider any unbiased estimator $\delta(X)$ and let $\tilde{h}(X)=E_\theta(\delta(X)|T(X))$, then

$$E_\theta(\tilde{h}(T(X)))=E_\theta\left(E_\theta(\delta(X)|T(X))\right)=E_\theta(\delta(X))=g(\theta)$$

by the tower property of conditional expectation. By (1) $\tilde{h}(T(X))\overset{a.s.}{=}h(T(X))$ and by Rao-Blackwell Theorem, $R(\theta,\tilde{h}(T(X)))\le R(\theta,\delta)$ for all $\theta\in\Theta$ if the loss function is convex. Hence, $h(T(X))$ is an UMRUE under any convex loss function.

3) If the loss function is strictly convex $R(\theta,\tilde{h}(T(X)))<R(\theta,\delta)$ unless $\delta(X)\overset{a.s.}{=}h(T(X))$. $\Rightarrow$ .. UMVUE if the loss function is the squared loss function.

Some useful strategies ("educated guesses") for finding UMVUE:

1) Rao-Blackwellization.
2) Solve for the (unique) δ satisfying $E_\theta(\delta(T(X)))=g(\theta)\ \forall\theta\in\Theta$
3) Guess (the right form of unbiased function of $T(X)$).

① RB: $X_i\overset{iid}{\sim}\mathrm{Bernoulli}(\theta)$. Estimate θ^2. $T(X)=\sum_{i=1}^n X_i$ is complete sufficient. $\delta(X)=X_1X_2$. $E(\delta(X))=\theta^2$. Find UMVUE via $E(\delta(X)|T(X))=\frac{P_\theta(X_1=X_2=1,\sum_{i=3}^n X_i=t-2)}{P(T(X)=t)}$

$$=\frac{\theta^2\binom{n-2}{t-2}\theta^{t-2}(1-\theta)^{n-t}\,\mathbb{I}(t\ge2)}{\binom{n}{t}\theta^t(1-\theta)^{n-t}}=\frac{t(t-1)}{n(n-1)}$$

② RB: $X_i\sim U(0,\theta)$. $T(X)=X_{(n)}$ is complete suff. $\delta(X)=2X_1$ is unbiased.

$$E_\theta(\delta(X)|T(X))=2\left(\tfrac{1}{n}T(X)+\left(1-\tfrac{1}{n}\right)\int_0^{T(X)}\frac{x_1\,dx_1}{T(X)}\right)=\left(\frac{n+1}{n}\right)T(X).$$

③ (Solve for δ): $X\sim\mathrm{Poisson}(\theta)$. X is complete sufficient. Want to estimate $g(\theta)=e^{-a\theta}$.

$$E_\theta(\delta(X))=\sum_{x\ge0}^{\infty}\delta(x)\frac{e^{-\theta}\theta^x}{x!}=e^{-a\theta},\ \forall\theta>0.\ \delta(X)=(1-a)^X$$ is UMVUE for $g(\theta)$. X is UMVUE for θ.

④ (Example) (Guess) Let $X_1\dots X_n\overset{iid}{\sim}N(\mu,\sigma^2)$.

1) The UMVUE for $\sigma^2=\frac{1}{n-1}\sum_{i=1}^n(X_i-\bar{X})^2=S^2$ [Ex]
2) How about the UMVUE for σ?
3) What is the UMVUE for μ^2?

observe that $X_i-\bar{X}_n\sim N(0,\frac{n-1}{n}\sigma^2)$

$\Rightarrow E(|X_i-\bar{X}_n|)=\sigma\sqrt{\frac{2}{\pi}}\sqrt{\frac{n-1}{n}}$.

$\Rightarrow \delta'(X)=\frac{\sqrt{\pi n}}{\sqrt{2(n-1)}}|X_i-\bar{X}_n|$ is unbiased for σ

$E(\delta'(X)|T(X))$ Blackwellize ...

We may instead observe another fact that

$$S_*^2=\sum_{i=1}^n(X_i-\bar{X}_n)^2=(n-1)S^2$$

and $S_*^2\sim\sigma^2\chi^2_{n-1}$. Hence $E(S_*)=\sigma E(\chi_{n-1})$

$\Rightarrow E(S_*)/E(\chi_{n-1})=\sigma$, meaning that $\frac{S_*}{E(\chi_{n-1})}$ is unbiased for σ and hence the UMVU.

3). $E(\bar{X}_n^2)=\mu^2+\frac{\sigma^2}{n}$ $\quad \mu^2$

So, $\delta_n(X)=\underbrace{\bar{X}_n^2-\frac{S_*^2}{n(n-1)}}_{\text{can be neg!}}$ is the UMVUE.

$\to E(\max(0,\delta_n(X))\neq\max(E(\delta(X)),0))$.

- **Characteristic of UMVUE:** Let $\Delta=\{\delta:E(\delta^2)<\infty\}$. Then $\delta_0\in\Delta$ is UMVU for $g(\theta)=E(\delta_0)$ iff. $E(\delta_0(\theta)U)=0$ for every $U\in\mathcal{U}=\{U:E(U)=0\}$
- **Fisher information:** $I(\theta)=E_\theta\left(\frac{\partial\log P_\theta(X)}{\partial\theta}\right)^2=\mathrm{Var}\left(\frac{\partial\log P_\theta(X)}{\partial\theta}\right)=-E_\theta\left(\frac{\partial^2\log P_\theta(X)}{\partial\theta^2}\right)$

$$E_\theta\psi=E_\theta\left(\frac{\partial\log P_\theta(X)}{\partial\theta}\right)=0$$

Cramer-Rao bound: Let $\psi=\frac{\partial\log P_\theta(X)}{\partial\theta}$, $g(\theta)=E_\theta(\delta)$, $g'(\theta)=E_\theta(\delta\psi)$ holds for $\forall\theta\in\Theta$, $E_\theta(\psi)=0$, $E_\theta(\delta^2)$, then $\mathrm{Var}_\theta(\delta)\ge\frac{(g'(\theta))^2}{I(\theta)}$, $\theta\in\Omega$.

Exp family: $P_\eta(x)=\exp(\eta T(x)-A(\eta))h(x)$. $I(\eta)=\mathrm{Var}_\eta(T(X))=A''(\eta)$.

If $\mu=A'(\eta)=E_\eta(T(X))$, then $A''(\eta)=I(\mu)(A''(\eta))^2$, $I(\mu)=\frac{1}{\mathrm{Var}(T)}$.

- Suppose $X_1,\dots,X_n\overset{iid}{\sim}P_\theta$, $I_X(\theta)=nI_{X_1}(\theta)$. $\mathrm{Var}_\theta(\delta)\ge\frac{g'(\theta)^2}{nI(\theta)}$.

Multivar. Fisher information:

$$\{I(\theta)\}_{i,j}=E_\theta\left(\frac{\partial\log p_\theta(x)}{\partial\theta_i}\frac{\partial\log p_\theta(x)}{\partial\theta_j}\right)$$
$$=\mathrm{Cov}_\theta\left(\frac{\partial\log p_\theta(x)}{\partial\theta_i},\frac{\partial\log p_\theta(x)}{\partial\theta_j}\right)=-E_\theta\left(\frac{\partial^2\log p_\theta(x)}{\partial\theta_i\partial\theta_j}\right),\ \mathrm{Var}_\theta(\delta)\geq(\nabla g(\theta))^T I^{-1}(\theta)\nabla g(\theta).$$

- average risk: $r(\Lambda,\delta)=\int R(\theta,\delta)\,d\Lambda(\theta)=E_\Theta(R(\Theta,\delta))=E_\Theta(E_X(L(\Theta,\delta(X)))\mid\Theta)=E_{(X,\Theta)}(L(\Theta,\delta(X)))$.

δ_Λ is Bayes estimator if δ_Λ minimizes $r(\Lambda,\delta)$ w.r.t Λ.

Thm. Suppose $\Theta\sim\Lambda$, and $X\mid\Theta=\theta\sim P_\theta$. If
i) there exists δ_0, an estimator of $g(\theta)$ with finite risk for all θ.
ii) there exists a value $\delta_\Lambda(x)$ that minimizes $E(L(\Theta,\delta_\Lambda(X))\mid X=x)$ for almost every x.
then δ_Λ is a Bayes estimator w.r.t. Λ.

Proof:
If (i) and (ii) hold, for any estimator, and for almost every x,
$E(L(\Theta,\delta_\Lambda(x))\mid X=x)\leq E(L(\Theta,\delta'(x))\mid X=x)$
Taking expectation on X,
$E(L(\Theta,\delta_\Lambda(X)))\leq E(L(\Theta,\delta'(X)))$, $\forall\delta$.

Ex. Consider $L(\theta,d)=(\theta-d)^2$. we need to minimize
$E((g(\Theta)-\delta(X))^2\mid X=x)$
$=E((g(\Theta)-E(g(\Theta)\mid X))^2\mid X=x)$. (I)
$+E((E(g(\Theta)\mid X)-\delta(X))^2\mid X=x)$ (II).
Equivalently, we need to minimize (II). i.e.
The Bayes estimator is $E(g(\Theta)\mid X)$ which is posterior mean of $E(g(\Theta))$ given X.

- Let $X_1,\dots,X_n\overset{iid}{\sim}N(\Theta,\sigma^2)$ with σ^2 known. $\Theta\sim N(\mu,b^2)$. Then $\pi(\theta\mid X)\propto\exp(-\frac{1}{2\tilde\sigma^2}(\theta-\tilde\mu)^2)$.

$$\tilde\mu=\frac{n\bar x/\sigma^2+\mu/b^2}{n/\sigma^2+1/b^2},\quad\tilde\sigma^2=\frac{1}{n/\sigma^2+1/b^2}$$

- Consider $L(\theta,d)=w(\theta)(g(\theta)-d)^2$, $\min_d(L(\theta,d)\mid X=x)(x)$, $\frac{\partial(*)}{\partial d}=0$:
$$\delta_\Lambda(x)=d^*=\frac{E(w(\Theta)g(\Theta)\mid X=x)}{E(w(\Theta)\mid X=x)}$$

Thm: If δ is unbiased for $g(\theta)$, then δ is not Bayes under the squared loss function unless
$E_{(X,\Theta)}(\{\delta(X)-g(\Theta)\}^2)=0$.

Proof 2 *Let δ be an unbiased estimator under the squared loss function. Then we know that δ is the posterior mean, which is*

$$\delta(X)=E(g(\Theta)\mid X),$$

almost surely. Thus, we have

$$\begin{aligned}E(\delta(X)g(\Theta))&=E(E(\delta(X)g(\Theta)\mid X))\\&=E(\delta(X)E(g(\Theta)\mid X))\\&=E(\delta^2(X)).\end{aligned}\tag{7}$$

Also,

$$\begin{aligned}E(\delta(X)g(\Theta))&=E(E(\delta(X)g(\Theta)\mid\Theta))\\&=E(g(\Theta)E(\delta(X)\mid\Theta))\\&=E(g^2(\Theta)).\end{aligned}\tag{8}$$

Observe that

$$\begin{aligned}E(\{\delta(X)-g(\Theta)\}^2)&=E(\delta^2(X))-2E(\delta(X)g(\Theta))+E(g^2(\Theta))\\&=E(\delta^2(X))-E(\delta(X)g(\Theta))+E(g^2(\Theta))-E(\delta(X)g(\Theta))\\&=E(\delta^2(X))-E(\delta^2(X))+E(g^2(\Theta))-E(g^2(\Theta))\text{ (due to (7) and (8))}\\&=0.\end{aligned}$$

Thus we have that $E(\{\delta(X)-g(\Theta)\}^2)=0$, which means the average risk is zero. The claim is thus proved.

Example 5 (Application of Theorem 2) *Suppose $X_1,\dots,X_n\overset{iid}{\sim}N(\Theta,\sigma^2)$, with σ^2 known. Is $\bar X$ Bayes under the squared loss function for some choice of the prior distribution?*
Observe that $E(\bar X\mid\theta)=\theta$, hence $\bar X$ is unbiased for θ. The corresponding average risk under the squared loss function is given by

$$E_{(X,\Theta)}(\{\bar X-\Theta\}^2)=\frac{\sigma^2}{n}\neq0.$$

So $\bar X$ is not Bayes estimator under any prior distribution.

$X\sim\mathrm{Unif}(0,\theta)$ $T(X)=\max X_i$
$P(X_1=x_1\mid T(X))=\frac{1}{n}1\{T(X)=x_1\}+\frac{(1-\frac{1}{n})1\{0<x_1<T(X)\}}{T(X)}$
UMVUE
$E[\delta(X)\mid T(X)]=2E[X_1\mid T(X)]$
$=\frac{n+1}{n}T(X)$

① $X_i\overset{iid}{\sim}N(\theta,\sigma^2)$
- σ^2 fixed $g(\theta)=\theta$
$l_\theta(X)=\log p_\theta(X)=-\frac{1}{2\sigma^2}\sum_{i=1}^n(X_i-\theta)^2+\text{const}$
- $\dot l_\theta(X)=\frac{\partial}{\partial\theta}\log p_\theta(X)=\frac{1}{\sigma^2}\sum_{i=1}^n(X_i-\theta)$
$\ddot l_\theta(X)=-\frac{n}{\sigma^2}$
$I(\theta)=E_\theta[-\ddot l_\theta(X)]=\frac{n}{\sigma^2}$
CRB is $\mathrm{Var}_\theta(\delta)\geq\frac{\sigma^2}{n}$, achieved by $\bar X$

② Exponential family
$I(\theta)=E_\theta[-\ddot l_\theta(X)]=n\nabla^2A(\theta)=n\mathrm{Cov}_\theta(T)$
$\mu(\theta)=E_\theta(T)=A'(\theta)$
$\mathrm{Var}_\theta(\delta)\geq\frac{[\mu'(\theta)]^2}{I(\theta)}=A''(\theta)=\mathrm{Var}_\theta(T)$

General case:
$\bar T=\frac{1}{n}\sum_i T(X_i)$ attains CRB for the mean parameter.

3. $X_i\overset{iid}{\sim}\mathrm{Poi}(\lambda)$
- Exponential family $T(X)=X$ and mean parameter λ.
- sample mean $\delta(X)=\frac{1}{n}\sum_i X_i$ achieves the CRB for λ
- $g(\theta)=\lambda^2$
$I(\lambda)=n/\mathrm{Var}_\lambda(X_1)=n/\lambda$
$T_1=\frac{1}{n}\sum_i X_i(X_i-1)$ unbiased for λ^2
$\mathrm{Var}_\lambda(T_1)=4\lambda^3/n+2\lambda^2/n>$ CRB
$S=\sum_i X_i$ sufficient statistic
Rao Blackwellized $T_2=E[T_1\mid S]=S(S-1)/n^2$ is UMVUE
- CRB is not attached since $\mathrm{Var}_\lambda(T_2)=4\lambda^3/n+2\lambda^2/n^2>$ CRB

UMVUE:

- $\mathrm{Ber}(X\mid p)$: $\hat p=\frac{\sum_{i=1}^n X_i}{n}=\bar x$, $\widehat{p^2}=\frac{\bar x(\bar x-1)}{n(n-1)}$
- $X\sim\mathrm{Unif}(0,\theta)$ $\hat\theta=\frac{n+1}{n}X_{(n)}$

CDF: $F_T(t)=(\frac{t}{\theta})^n1\{t\in(0,\theta)\}+1\{t\geq\theta\}$ $X_{(n)}/\theta\sim\mathrm{Beta}(n,1)$
Density: $t\mapsto n\theta^{-n}t^{n-1}1\{t\in(0,\theta)\}$
Suppose $E_\theta[f(T)]=0$ $\forall\theta>0$
$\Rightarrow f(t)=0\Rightarrow T$ complete

- $X\sim N(\mu,\sigma^2)$ $\hat\mu:\bar X$ $\hat\sigma^2:\sum_{i=1}^n\frac{(X_i-\bar X)^2}{n-1}=\frac{S^2}{n-1}$ $\sigma:\frac{S}{E[\chi_{n-1}]}$

$T(X)=\begin{pmatrix}x\\x^2\end{pmatrix}$ $\eta=\begin{pmatrix}\mu/\sigma^2\\-\frac{1}{2\sigma^2}\end{pmatrix}$

$X_i-\bar X\sim N(0,\frac{n-1}{n}\sigma^2)$

$\widehat{\mu^2}=\bar X^2-\frac{S^2}{n(n-1)}$

$S^2=\sum_{i=1}^n(X_i-\bar X)^2$
$ES=\sigma E(\chi_{n-1})$

since $E(\bar X^2)=\mu^2+\frac{\sigma^2}{n}$.

- $X\sim\mathrm{Poi}(\lambda)$: $\hat\lambda=\bar X$.

$X\sim\mathrm{Exp}(\theta)=\frac{1}{\theta}e^{-\frac{1}{\theta}x}$ $\sum_{i=1}^n X_i\sim\Gamma(n,\theta)$
$\bar X$ is UMVUE for $\frac{1}{\theta}$. $E[\frac{1}{\bar X}]=\frac{n}{n-1}\theta$
$\frac{n-1}{n}\frac{1}{\bar X}$ is UMVUE for θ.

Theorem 2 *For any minimal, s-dimensional exponential family, the statistic $(\sum_{i=1}^n T_1(X_I),\dots,\sum_{i=1}^n T_s(X_i))$ is a minimal sufficient statistic. [Example 3.12 ?]*

Proof 1 *Let $p(x;\theta)=\exp\{\eta(\theta)T(x)-B(\theta)\}h(x)$ be the density of an s-dimensional exponential family, where $\theta\in\Omega$. By NFFC, T is sufficient.*

Suppose $p(x;\theta)\propto_\theta p(y;\theta)$, then

$$e^{\eta(\theta)T(x)}\propto_\theta e^{\eta(\theta)T(y)},$$

which implies that

$$\eta(\theta)\cdot T(x)=\eta(\theta)\cdot T(y)+C,$$

where the constant C may depend on both x and y (but is independent of θ).

If θ_0 and θ_1 are any two points in Ω,

$$\{\eta(\theta_0)-\eta(\theta_1)\}\cdot T(x)=\{\eta(\theta_0)-\eta(\theta_1)\}\cdot T(y)$$

if and only if

$$\{\eta(\theta_0)-\eta(\theta_1)\}\cdot\{T(x)-T(y)\}=0$$

This shows that $T(x)-T(y)$ is orthogonal to eery vector in

$$\eta(\Omega)\ominus\eta(\Omega)\equiv\{\eta(\theta_0)-\eta(\theta_1):\theta_0\in\Omega,\theta_1\in\Omega\},$$

so it must lie in the orthogonal complement of the linear span of $\eta(\Omega)\ominus\eta(\Omega)$. In particular, if the linear span of $\eta(\Omega)\ominus\eta(\Omega)$ is all of $\mathbb{R}^s$, then $T(x)$ must equal $T(y)$ in which case T is minimal sufficient.

$E_\theta[T]=\nabla A(\theta)$ $\mathrm{Cov}_\theta(T)=\nabla^2A(\theta)$
- MGF of T: $M_T(u)=M_T(u;\theta)=E_\theta[e^{\langle u,T\rangle}]$
$=\int e^{\langle u,T(x)\rangle}e^{\langle\theta,T(x)\rangle-A(\theta)}d\tilde\nu(x)$
$=e^{A(u+\theta)-A(\theta)}$
If $\theta\in\mathrm{int}\,\Omega_0$, M_T is finite in the neighborhood of zero
$M_T(u)<\infty$ for $\|u\|_2\leq\varepsilon$
- DCT $\Rightarrow M_T$ is C^∞ in a neighborhood of 0
interchange the order of – –
- $M_T(u)\frac{\partial A}{\partial u_i}(u+\theta)=\frac{\partial}{\partial u_i}M_T(u)=E_\theta[\frac{\partial}{\partial u_i}e^{\langle u,T\rangle}]$
$=E_\theta[T_ie^{\langle u,T\rangle}]$
$u=0\Rightarrow M_T(0)=1$

Example 8 *Let $X_1,\dots,X_n\overset{iid}{\sim}N(\mu,\sigma^2)$, where both of μ and σ^2 are unknown. Then $\bar X_n\perp\!\!\!\perp n^{-1}\sum_{i=1}^n(X_i-\bar X_n)^2$ with $\bar X_n=n^{-1}\sum_{i=1}^n X_i$.*

Fix any $\sigma>0$ and consider the submodel $\mathcal{P}_\sigma=\{N(\mu,\sigma^2):\mu\in\mathbb{R}\}$ In each submodel, $\bar X_n$ is complete and sufficient, and $n^{-1}\sum_{i=1}^n(X_i-\bar X)^2$ is ancillary

$$X_i=Z_i+\mu\qquad X_i-\bar X_n\to Z_i-\bar Z$$

By Basu's theorem, $\bar X_n\perp\!\!\!\perp\sum_{i=1}^n(x_i-\bar x_n)^2$ under $N(\mu,\sigma^2)$ for any μ. Since σ is arbitrary, we can conclude that $\bar X_n\perp\!\!\!\perp\sum_{i=1}^n(x_i-\bar x_n)^2$ hold for the full model $\mathcal{P}=\{N(\mu,\sigma^2):\mu\in\mathbb{R},\sigma^2>0\}$

Example 1. If $X\sim p(x;\theta)\propto h(x)e^{\theta x}$, then the statistic $T(X)=X$ is complete.

Proof. Outline of steps:

(1) Suppose $\int f(x)h(x)e^{\theta x}dx=0$ for all θ.

(2) Then we can decompose f as $f(x)=f_+(x)-f_-(x)$, with $f_+\geq0,f_-\geq0$

(3) f_+ and f_- can be viewed as unnormalized densities $p_+(x)$, $p_-(x)$

(4) (1) implies that the moment generating functions of $p_+(x)$ and $p_-(x)$ are equal: $\int p_+(x)e^{\theta x}dx=\int p_-(x)e^{\theta x}dx,\forall\theta$. Hence $p_+(x)$ and $p_-(x)$ represent the same distribution, so $p_+(x)=p_-(x)$, $f_+=f_-$, and $f\equiv0$.

Example 5. Let $X_1,\dots,X_n\overset{iid}{\sim}\mathrm{Ber}(\theta)$, $\theta\in(0,1)$. Then $T(X)=\sum_iX_i$ is sufficient. Suppose $\mathbb{E}_\theta[f(T(X))]=0$ for all $\theta\in(0,1)$. This means

$$\sum_{j=0}^n f(j)\binom{n}{j}\theta^j(1-\theta)^{n-j}=0,\qquad\forall\theta\in(0,1).$$

Dividing through by θ^n and using $\beta=\frac{\theta}{1-\theta}$, we get

$$\sum_{j=1}^n f(j)\binom{n}{j}\beta^j=0,\qquad\forall\beta>0.$$

If f are non-zero, then the quantity on the left is a polynomial of degree at most n. However, an nth-degree polynomial can have at most n roots. Hence it impossible for the LHS to equal 0 for every $\beta>0$ unless $f=0$. Hence T is complete.

Example 6. Let $X_1,\dots,X_n\overset{iid}{\sim}\mathcal{N}(\theta,\sigma^2)$ with unknown $\mu\in\mathbb{R}$ and known $\sigma^2>0$. Is $\overline{X}_n=\frac{1}{n}\sum_iX_i$ complete for this model? (We already know that it is minimal sufficient.) The answer is yes, but to keep the algebra simple we will show that it is complete in the special case of $n=1$ and $\sigma=1$, so that $T(X)=X\sim\mathcal{N}(\theta,1)$. Suppose

$$\mathbb{E}_\theta[f(X)]=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty f(x)\exp\left(-\frac{(x-\theta)^2}{2}\right)dx=0\qquad\forall\theta\in\mathbb{R}$$

Multiplying through by $\sqrt{2\pi}e^{\frac{\theta^2}{2}}$ gives

$$\int_{-\infty}^\infty f(x)\exp\left(-\frac{x^2}{2}\right)\exp(\theta x)\,dx=0\qquad\forall\theta\tag{3.1}$$

Now decompose f into its positive and negative part as $f(x)=f_+(x)-f_-(x)$, where $f_+(x)=\max(f(x),0)$ and $f_-(x)=\max(-f(x),0)$. Then $f_+(x)\geq0$ and $f_-(x)\geq0$ for all $x\in\mathbb{R}$, and $f_+(x)=f_-(x)$ if and only if $f_+(x)=f_-(x)=0$.

If $f(x)\geq0$ a.e. or $f(x)\leq0$ a.e., then (3.1) implies that $f(x)=0$ a.e. because setting $\theta=0$ gives us an integral of a nonnegative (or nonpositive) function being zero. This is completeness.

In the other case, f_+ and f_- have non-zero components and we may write

$$\frac{\int_{-\infty}^\infty f_+(x)e^{-\frac{x^2}{2}}e^{\theta x}dx}{\int_{-\infty}^\infty f_+(x)e^{-\frac{x^2}{2}}dx}=\frac{\int_{-\infty}^\infty f_-(x)e^{-\frac{x^2}{2}}e^{\theta x}dx}{\int_{-\infty}^\infty f_-(x)e^{-\frac{x^2}{2}}dx},\tag{3.2}$$

where the equality of the numerators follows from (3.1), and the equality of the denominators follows from setting $\theta=0$. The quantity

$$\frac{f_+(x)e^{-\frac{x^2}{2}}}{\int_{-\infty}^\infty f_+(x)e^{-\frac{x^2}{2}}dx}$$

defines a probability density, and the left-hand-side of (3.2) is the moment generating function of this density. Similarly, the right-hand-side is the moment generating function of the density

$$\frac{f_-(x)e^{-\frac{x^2}{2}}}{\int_{-\infty}^\infty f_-(x)e^{-\frac{x^2}{2}}dx}.$$

The equality of the moment generating functions implies equality of the densities, which in turn implies $f_+(x)=f_-(x)$ a.e. Then $f_+(x)=f_-(x)=0$ a.e. so $f(x)=0$ a.e. Hence T is complete.

Fisher information reparame
$\theta=h(\xi)$. $\tilde I(\xi)=(h'(\xi))^2I(\theta)$

$P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}$
$P(X=k)=(1-p)^{k-1}p^k$
$P(X=k)=\binom{k-1}{r-1}p^r(1-p)^{k-r}$
$P(X;\alpha,\beta)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}$
$P(X)=\frac{1}{\pi}\frac{\lambda}{\lambda^2+(x-\mu)^2}$

$P(X;\lambda)=\lambda e^{-\lambda x}1(x\geq0)$, $\beta=\frac{1}{\lambda}$
$P(x;\alpha,\beta)=\frac{x^{\alpha-1}e^{-\beta x}\beta^\alpha}{\Gamma(\alpha)}1(x\geq0)$

判断: By Hölder's inequality:
$E|X_n-X|^q\leq(E|X_n-X|^{q\cdot\frac{r}{q}})^{\frac{q}{r}}[E(1^{\frac{r}{r-q}})]^{\frac{r-q}{r}}\leq(E|X_n-X|^r)^{\frac{q}{r}}$
since $\frac{q}{r}<1$ $(\frac{r}{q})^{-1}+(\frac{r}{r-q})^{-1}=1$
Q3: $\frac{n!}{(i-1)!(n-i)!}=\frac{\Gamma(n+1)}{\Gamma(i)\Gamma(n-i+1)}$

Hölder $|EXY|\leq E|XY|\leq(E|X|^p)^{1/p}\cdot(E|Y|^q)^{1/q}$ $(\frac{1}{p}a^p+\frac{1}{q}b^q\geq ab,\ \frac{1}{p}+\frac{1}{q}=1)$
Cauchy Schwarz & Covariance Ineq $(p,q=2)$ $E(g(X)h(X))\leq Eg(X)h(X)$ ↑↓ $\geq$ ↑↑ or ↓↓
Minkowski $(E|X+Y|^p)^{1/p}\leq(E|X|^p)^{1/p}+(E|Y|^p)^{1/p}$ $p\geq1$
Jensen $Eg(X)\geq g(EX)$ g convex $M_H\leq M_G\leq M_A\leq M_Q\leq\mathrm{Norm}\ (p\geq2)$
Chebyshev $rP(\varphi(X)\geq r)\leq E\varphi(X)$ $P(|X-EX|\geq\varepsilon)\leq\frac{\mathrm{Var}X}{\varepsilon^2}$ Ee^{tX}
By integrating the tails of any dist / By MGF: $P(|X|\geq a)\leq e^{-at}M_X(t)$
$p(X=k)=\binom{n}{k}p^k(1-p)^{n-k}=\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}$

Bernoulli(p): $\{p,p(1-p),(1-p)+pe^t\}\Rightarrow$ Binomial(n,p): $\{n\cdot\cdot,[\cdot\cdot]^n\}$
Poisson(λ): $\{\lambda,\lambda,\exp[\lambda(e^t-1)]\}$. $P(X=x+1)=\frac{\lambda}{x+1}P(X=x)$ Lehman
Geometric(p): $\{\frac{1}{p},\frac{1-p}{p^2},pe^t/(1-(1-p)e^t)\}$, $P(X>s\mid X>t)=P(X>s-t)$
$H(N,M,K)$: $\{\frac{KM}{N},\frac{KM}{N}\frac{(N-M)(N-K)}{N(N-1)},\}$
$NB(r,p)$: $\{\frac{r(1-p)}{p},\frac{r(1-p)}{p^2},[p/(1-(1-p)e^t]^n\ t<-\log(1-p)\}$
Beta(α,β): $\{\frac{\alpha}{\alpha+\beta},\alpha\beta/(\alpha+\beta)^2(\alpha+\beta+1),\cdot\cdot\}$ $\Gamma(\alpha)=\int_0^\infty t^{\alpha-1}e^{-t}dt$
$\Gamma(\alpha+1)=\alpha\Gamma(\alpha)$, $\Gamma(1)=1$, $\Gamma(\frac{1}{2})=\sqrt{\pi}$, $\Gamma(n)=(n-1)!$
Cauchy(θ,σ): $\frac{1}{\pi\sigma}\frac{1}{1+(\frac{x-\theta}{\sigma})^2}$ $\{\infty,\infty,NA\}$ fat tail ~~exp family~~ $\frac{N}{N}$
χ_p^2: $\{p,2p,(\frac{1}{1-2t})^{p/2}\ t<\frac{1}{2}\}$ $\sum\chi_{p_i}^2=\chi_{\sum p_i}^2$ $\sum N^2$ $\mathrm{Exp}(2)=\chi_2^2$
DExp(μ,σ): $\{\mu,2\sigma^2,e^{\mu t}/(1-(\sigma t)^2)\}$ $|t|<\frac{1}{\sigma}$
Exp(β): $\{\beta,\beta^2,\frac{1}{1-\beta t}\ t<\frac{1}{\beta}\}$ memoryless $\approx$ Geometric.
$F_{m,n}$: $\{\frac{n}{n-2},2(\frac{n}{n-2})^2(m+n-2)/m(n-4),NA\}$ $\frac{1}{m}\chi_m^2/\frac{1}{n}\chi_n^2$
Gamma(α,β): $\{\alpha\beta,\alpha\beta^2,(1/(1-\beta t))^\alpha\ t<\frac{1}{\beta}\}$ α shape β scale
Logistic(μ,β): $\{\mu,\pi^2\beta^2/3,\cdot\cdot\}$ cdf $F(x\mid\mu,\beta)=1/(1+\exp[-(x-\mu)/\beta])$
$N(\mu,\sigma^2)$: $\{\mu,\sigma^2,\exp\{\mu t+\frac{1}{2}\sigma^2t^2\}\}$
t_n: $\{0,\frac{n}{n-2},NA\}$ $t_n^2=F_{1,n}$
Uniform(a,b): $\{(b+a)/2,(b-a)^2/12,(e^{bt}-e^{at})/(b-a)t$

Order Statistic
Cont. $F_{X_{(k)}}(x)=\sum_{j=k}^n\binom{n}{j}[F_X(x)]^j[1-F_X(x)]^{n-j}$
$f_{X_{(k)}}(x)=\frac{n!}{(k-1)!(n-k)!}f_X(x)[F_X(x)]^{k-1}[1-F_X(x)]^{n-k}$

$f_{X_{(i)}X_{(j)}}(u,v)=\frac{n!}{(i-1)!(j-1-i)!(n-j)!}f_X(u)f_X(v)[F_X(u)]^{i-1}[F_X(v)-F_X(u)]^{j-1-i}[1-F_X(v)]^{n-j}$ $u<v$
$f_{X_{(1)},\dots,X_{(n)}}(x_1,\dots,x_n)=\begin{cases}n!f_X(x_1)\cdots f_X(x_n)&-\infty<x_1<\cdots<x_n<\infty\\0&\text{otherwise.}\end{cases}$

Disc. $P(X_{(k)}\leq x_i)=\sum_{j=k}^n\binom{n}{k}P_i^j(1-P_i)^{n-j}$
$P(X_{(k)}=x_i)=\sum_{j=k}^n\binom{n}{k}[P_i^j(1-P_i)^{n-j}-P_{i-1}^j(1-P_{i-1})^{n-j}]$
$\begin{cases}P_0=0\\P_1=p_1\\\vdots\\P_i=\sum_{j=1}^ip_j\end{cases}$ $P(X=x_i)=p_i$ $x_1<x_2<\cdots<x_n$