Stat 8.UE

(2.) Let $X_1,\dots,X_n$ random sample with pdf $f(x|\vartheta) = \vartheta/x^2 \, \mathbb{1}_{[\vartheta,\infty]}(x)$ with unknown $\vartheta > 0$.
Use factorization theorem to find a sufficient statistic for ϑ.

$$f_X(x) = f_X(x_1,\dots,x_n|\vartheta) = \prod_{i=1}^{n} f(x_i|\vartheta) = \vartheta^n \left(\prod_{i=1}^{n} \frac{1}{x_i}\right)^2 \mathbb{1}_{[\vartheta,\infty]}(\min_{i=1,\dots,n} x_i),$$

so $T(X) = \left(\left(\prod \frac{1}{X_i}\right)^2, \min X_i\right)$ is sufficient by the factorization theorem $(h \equiv 1)$.

(3.) $X_1,\dots,X_n \sim \mathcal{N}(\mu,\mu)$ random sample, $\mu > 0$ unknown.

(a) Show $\sum X_i^2$ is minimal sufficient statistic in $\mathcal{N}(\mu,\mu)$ family.

$$\frac{f(x|\mu)}{f(y|\mu)} = \frac{\prod f(x_i|\mu)}{\prod f(y_i|\mu)} = \frac{\frac{1}{(2\pi\mu)^{n/2}} e^{-\frac{1}{2\mu}\sum_{i=1}^{n}(x_i-\mu)^2}}{\frac{1}{(2\pi\mu)^{n/2}} e^{-\frac{1}{2\mu}\sum_{i=1}^{n}(y_i-\mu)^2}} = e^{-\frac{1}{2\mu}\sum_{i=1}^{n}((x_i-\mu)^2-(y_i-\mu)^2)}$$

$$= e^{-\frac{1}{2\mu}\left[\sum x_i^2 - \sum y_i^2 - 2\mu\left(\sum x_i - \sum y_i\right)\right]} = e^{-\frac{1}{2\mu}\left(\sum x_i^2 - \sum y_i^2\right)} e^{-\left(\sum x_i + \sum y_i\right)},$$

which is constant in μ $\Leftrightarrow$ $\sum x_i^2 = \sum y_i^2$.

(b) Show that $(\sum X_i, \sum X_i^2)$ is sufficient but not minimal sufficient.

It is trivial that the statistic is sufficient, since its second component is by (a).

If it would be minimal, then, since $\sum X_i^2$ is sufficient, there would be a function $\tilde{r}$ s.t. $\tilde{r}(\sum X_i^2) = (\sum X_i, \sum X_i^2)$ and therefore a function r s.t. $r(\sum X_i^2) = \sum X_i$.

But such a function does not exist, because e.g. for $x = (1,-1)$ and $x' = (1,1)$ we have $x_1^2 + x_2^2 = x_1'^2 + x_2'^2 = 1$, but $0 = x_1 + x_2 \neq x_1' + x_2' = 2$. $\square$

(4.) $X_1,\dots,X_n$ random sample, pdf $f(x|\vartheta) = 2x/\vartheta \, \mathbb{1}_{(0,\vartheta)}(x)$, $\vartheta > 0$ unknown.
Find minimal sufficient statistic for ϑ.

$$f(x|\vartheta) = \prod_{i=1}^{n} f(x_i|\vartheta) = \left(\prod_{i=1}^{n} x_i\right) \mathbb{1}_{(0,\infty)}(\min x_i)\, \mathbb{1}_{(-\infty,\vartheta)}(\max x_i) \left(\frac{2}{\vartheta}\right)^n.$$

We define $\boxed{T(x) := \max_{i=1,\dots,n} x_i}$, which is sufficient by the Fisher-Neyman theorem with $h(x) := \left(\prod x_i\right) \mathbb{1}_{(0,\infty)}(\min x_i)$ and $g(T(x)|\vartheta) := \mathbb{1}_{(0,\vartheta)}(T(x)) \cdot \left(\frac{2}{\vartheta}\right)^n$. It is minimal because

$$\frac{f(x|\vartheta)}{f(y|\vartheta)} = \frac{h(x)\, \mathbb{1}_{(0,\vartheta)}(\max x_i)}{h(y)\, \mathbb{1}_{(0,\vartheta)}(\max y_i)}$$ is constant in ϑ if $\max x_i = \max y_i$, and if

$\max x_i \neq \max y_i$, there is a ϑ' with $\max x_i < \vartheta' < \max y_i$ or $\max y_i < \vartheta' < \max x_i$ and a ϑ'' with $0 < (\max x_i, \max y_i) < \vartheta''$. We obtain

$$\frac{f(x|\vartheta')}{f(y|\vartheta')} = 0 \text{ or undefined}, \quad \text{but} \quad \frac{f(x|\vartheta'')}{f(y|\vartheta'')} = 1.$$

⑤ $X_1,\dots,X_n \sim Poi(\lambda)$ iid. with unknown $\lambda > 0$. $\left[\mathbb{P}(X=x) = \frac{1}{x!}\lambda^x e^{-\lambda} \right]$ $x \geq 0$

(a) Show that $Y = \sum_{i=1}^n X_i$ is a sufficient statistic for λ.

$f(x|\lambda) \overset{\text{iid}}{=} \prod_{i=1}^n f(x_i|\vartheta) = \prod_{i=1}^n \frac{1}{x_i!}\lambda^{x_i} e^{-\lambda} = \left(\prod_{i=1}^n \frac{1}{x_i!}\right) e^{-\lambda} \lambda^{\sum_{i=1}^n x_i}$, so by the factorization theorem, $Y(X)$ is sufficient.

We have $\frac{f(x|\lambda)}{f(y|\lambda)} = \frac{(\prod x_i)\lambda^{\sum x_i}}{(\prod y_i)\lambda^{\sum y_i}} = \left(\prod \frac{x_i}{y_i}\right)\lambda^{(\sum x_i)-(\sum y_i)}$, which is obviously constant in λ iff $\sum x_i = \sum y_i$, so Y is minimal sufficient.

(b) • Find unbiased est. of $\mathbb{P}(X=r) =: p_r$ which only depends on X_1:

We choose $T(x) = \mathbb{1}_{\{r\}}(x_1)$ and obtain $\mathbb{E}_{\mathbb{P}(X=r)=p_r}(\mathbb{1}_{\{r\}}(x_1)) = \mathbb{P}(x_1 = r) = p_r$.

• Find $\mathbb{P}(X_1 = r \mid Y = k)$ for $k \geq r$ and $k < r$.

$k<r$: $\mathbb{P}(X_1 = r \mid Y = k) \leq \mathbb{P}(X_1 = r \mid \sum_i X_i < r) = 0$, as $X_i \geq 0 \ \forall i$.

$k \geq r$: We already know that the sum of n $Poi(\lambda)$ r.v. is $Poi(\lambda n)$-distributed.

Hence, $\mathbb{P}(X_1 = r \mid Y = k) = \frac{\mathbb{P}(\sum_{i=1}^n X_i = k \mid X_1 = r)\,\mathbb{P}(X_1 = r)}{\mathbb{P}(\sum_{i=1}^n X_i = k)} = \frac{\mathbb{P}(\sum_{i=2}^n X_i = k - r)\frac{1}{r!}e^{-\lambda}\lambda^r}{\frac{1}{k!}e^{-n\lambda}(n\lambda)^k} =$

$= \frac{\frac{1}{(k-r)!}e^{-(n-1)\lambda}((n-1)\lambda)^{k-r}\frac{1}{r!}e^{-\lambda}\lambda^r}{\frac{1}{k!}e^{-n\lambda}(n\lambda)^k} = \frac{k!}{r!(k-r)!}(\lambda n)^{-k}((n-1)\lambda)^{k-r}\lambda^r =$

$= \binom{k}{r}\cancel{\lambda^{-k}} n^{-k}(n-1)^{k-r}\cancel{\lambda^{k-r}}\cancel{\lambda^r} = \binom{k}{r} n^{-k}(n-1)^{k-r}$.

• Hence, use the Rao-Blackwell theorem to improve your estimator of μ.

<u>Rao-Blackwell</u>: W unbiased est. of $\gamma(\vartheta)$, T sufficient stat. for ϑ. Define $\varphi(T) := \mathbb{E}(W|T)$.
Then, $\mathbb{E}_\vartheta(\varphi(T)) = \gamma(\vartheta)$ and $\mathbb{V}_\vartheta(\varphi(T)) \leq \mathbb{V}_\vartheta(W)$.

$\vartheta = \lambda$, $\gamma(\vartheta) = p_r = \frac{1}{r!}e^{-\lambda}\lambda^r$, $T(X) = Y = \sum_{i=1}^n X_i$, $W(X) = \mathbb{1}_{\{r\}}(x_1)$;

by what we showed above, $\mathbb{E}(X_1 = r \mid Y = k) = \binom{k}{r} n^{-k}(n-1)^{k-r}$,

so $\varphi(T) := \mathbb{E}(W(X)|T) = \binom{Y}{r} n^{-Y}(n-1)^{Y-r}\mathbb{1}_{[0,Y]}(r)$ is our „better" estimator of p_r.