then $\prod_{i=1}^{n} I(y_i \in y^*)$ could be part of $g[T(\underset{\sim}{y})|\underset{\sim}{\theta})$.

Typically y^* is an interval with endpoints a and b, not necessarily finite. For pdfs and $n>1$

$$\prod_{i=1}^{n} I[y_i \in (a,b)] = \underbrace{I[a < y_{(1)} \le y_{(n)} < b]}_{\text{i) use if a and b are unknown}} = \underbrace{I[a < y_{(1)}]\, I[y_{(n)} < b]}_{\text{ii) use if one of a or b is unknown}}$$

iii) use either i) or ii) if both a and b are known.

$$\prod_{i=1}^{n} I(y_i \in [a,b]) = I(a \le y_{(1)} \le y_{(n)} \le b) = I[a \le y_{(1)}]\, I[y_{(n)} \le b]$$

$$\prod_{i=1}^{n} I[y_i \in (-\infty, b)] = I(y_{(n)} < b)$$

$$\prod_{i=1}^{n} I(y_i \in [a,\infty)) = I[y_{(1)} \ge a] \quad \text{etc.}$$

ex] ex 4.10 $Y_1, \ldots, Y_n$ iid $U(\theta_1, \theta_2)$. Find a suff stat $\underset{\sim}{T}$ for (θ_1, θ_2).

soln $f(\underset{\sim}{y}) = \prod_{i=1}^{n} f(y_i) = \prod_{i=1}^{n} \frac{1}{\theta_2-\theta_1} I[\theta_1 \le y_i \le \theta_2]$

$$= \underbrace{\frac{1}{(\theta_2-\theta_1)^n} I[\theta_1 \le y_{(1)} \le y_{(n)} \le \theta_2]}_{g(T(\underset{\sim}{y})|\underset{\sim}{\theta})} \quad (\underbrace{1}_{h(\underset{\sim}{y})})$$

So $\underset{\sim}{T}(\underset{\sim}{Y}) = (Y_{(1)}, Y_{(n)})$ by Factorization

If θ_1 is known, $f(\underset{\sim}{y}) = \underbrace{\frac{1}{(\theta_2-\theta_1)^n} I(y_{(n)} \le \theta_2)}_{g(T(\underset{\sim}{y})|\theta_2)} \underbrace{I[\theta_1 \le y_{(1)}]}_{h(\underset{\sim}{y})}$

and $Y_{(n)}$ is suff for θ_2.

If θ_2 is known, $f(\underset{\sim}{y}) = \underbrace{\frac{1}{(\theta_2-\theta_1)^n} I(\theta_1 \le y_{(1)})}_{g(T(\underset{\sim}{y})|\theta_1)} \underbrace{I(y_{(n)} \le \theta_2)}_{h(\underset{\sim}{y})}$

and $Y_{(1)}$ is suff for θ_1 by factorization.

16] p114 If $\underset{\sim}{T}(\underset{\sim}{Y})$ is a sufficient statistic for $\underset{\sim}{\theta}$, then any statistic $\underset{\sim}{S}(\underset{\sim}{Y})$ that is a one to one function of $\underset{\sim}{T}(\underset{\sim}{Y})$ (so $\underset{\sim}{S}(\underset{\sim}{Y}) = k(\underset{\sim}{T}(\underset{\sim}{Y}))$ where k is one to one) is also a sufficient statistic for $\underset{\sim}{\theta}$.

ex] know $Y_1, \ldots, Y_n$ iid $N(\mu, \sigma^2)$ $\underset{\sim}{T}(\underset{\sim}{Y}) = (\Sigma Y_i, \Sigma Y_i^2)$ is sufficient for (μ, σ^2), so

$$\underset{\sim}{S}(\underset{\sim}{Y}) = (\overline{Y}, S^2) = \left(\frac{1}{n}\Sigma Y_i, \frac{1}{n-1}\left(\Sigma Y_i^2 - n\left(\frac{\Sigma Y_i}{n}\right)^2\right)\right)$$

is also suff for (μ, σ^2).

17] p113-4 If $\underset{\sim}{\eta} = \underset{\sim}{\tau}(\underset{\sim}{\theta})$ is one to one, then the parameter has just been renamed and the distribution of $Y_1, \ldots, Y_n$ is not changed. Hence if $\underset{\sim}{T}(\underset{\sim}{Y})$ is sufficient for $\underset{\sim}{\theta}$, then $\underset{\sim}{T}(\underset{\sim}{Y})$ is sufficient for $\underset{\sim}{\eta}$.

proof of Th4.5e

ex] one to one is crucial: If $Y_1, \ldots, Y_n$ are iid $N(\mu, \sigma^2)$, let $k(\Sigma Y_i, \Sigma Y_i^2) \equiv 1 = T'(\underset{\sim}{Y})$. Then $T'(\underset{\sim}{Y})$ is not sufficient for (μ, σ^2) since $\underset{\sim}{Y} | T'(\underset{\sim}{Y}) \sim \underset{\sim}{Y} \sim N_n(\mu \underset{\sim}{1}, \sigma^2 I_n)$

which depends on $\underset{\sim}{\theta} = (\mu, \sigma^2)$.

18] Know for final p111 A sufficient statistic $\underset{\sim}{T}(\underset{\sim}{Y})$ for $\underset{\sim}{\theta}$ is a minimal sufficient statistic for $\underset{\sim}{\theta}$ if

$\underset{\sim}{T}(\underset{\sim}{Y})$ is a function of $\underset{\sim}{S}(\underset{\sim}{Y})$ for any other suff stat $\underset{\sim}{S}(\underset{\sim}{Y})$. So $\underset{\sim}{T}(\underset{\sim}{Y}) = k_S(\underset{\sim}{S}(\underset{\sim}{Y}))$ for some function

k_S.

memory aid

19] Mnemonic: if $Y_1, \ldots, Y_n$ have pdf or pmf $\underbrace{f(y_1, \ldots, y_n \mid \underset{\sim}{\theta})}_{g(T(\underset{\sim}{y}) \mid \underset{\sim}{\theta})} \cdot \underbrace{1}_{h(\underset{\sim}{y})}$ then

$\underset{\sim}{S}(\underset{\sim}{Y}) = \underset{\sim}{Y}$ is sufficient by factorization, and $\underset{\sim}{T} \equiv \underset{\sim}{T}(\underset{\sim}{Y})$ is a function of $\underset{\sim}{S}$.

20] Want to use a minimal sufficient statistic which reduces the data as far as possible without losing information about $\underset{\sim}{\theta}$.

Note] $\underset{\sim}{T_1}(\underset{\sim}{Y}) = (Y_1, \ldots, Y_n)$ and $\underset{\sim}{T_2}(\underset{\sim}{Y}) = (Y_{(1)}, \ldots, Y_{(n)})$ are sufficient, but $\underset{\sim}{T_1}$ is not a function of $\underset{\sim}{T_2}$ for $n > 1$. So $\underset{\sim}{T_1}$ is not minimal suff.

21] Know p114 Any one to one function of a minimal sufficient statistic is a minimal sufficient statistic.

22) Know for final p 112 Suppose statistic

$\underset{\sim}{T}(\underset{\sim}{Y})$ has pdf or pmf $f(\underset{\sim}{t}|\underset{\sim}{\theta})$. Then

$\underset{\sim}{T}(\underset{\sim}{Y})$ is a complete sufficient statistic for $\underset{\sim}{\theta}$ if

$$E_{\underset{\sim}{\theta}}\left[g(\underset{\sim}{T}(\underset{\sim}{Y}))\right] = 0 \quad \forall \underset{\sim}{\theta} \in \Theta$$

implies that $P_{\underset{\sim}{\theta}}\left[g(\underset{\sim}{T}(\underset{\sim}{Y})) = 0\right] = 1 \ \forall \underset{\sim}{\theta} \in \Theta$.

23) Know p 114 Bahadur's theorem Suppose k does not depend on the sample size n. Then a k dimensional complete suff stat $\underset{\sim}{T}(\underset{\sim}{Y})$ is also minimal suff.

24) Know for final Lehmann Scheffé LSM theorem: Let iid sample $\underset{\sim}{Y}$ have pdf or pmf $f(\underset{\sim}{y}|\underset{\sim}{\theta})$. Let $c_{\underset{\sim}{x},\underset{\sim}{y}}$ be a constant. If there exists a function $\underset{\sim}{T}(\underset{\sim}{y})$ such that for any two sample points $\underset{\sim}{x}$ and $\underset{\sim}{y}$, the ratio $R_{\underset{\sim}{x},\underset{\sim}{y}}(\underset{\sim}{\theta}) = \dfrac{f(\underset{\sim}{x}|\underset{\sim}{\theta})}{f(\underset{\sim}{y}|\underset{\sim}{\theta})} = c_{\underset{\sim}{x},\underset{\sim}{y}} \ \forall \underset{\sim}{\theta} \in \Theta$

iff $\underset{\sim}{T}(\underset{\sim}{x}) = T(\underset{\sim}{y})$, then $\underset{\sim}{T}(\underset{\sim}{Y})$ is a minimal sufficient statistic for $\underset{\sim}{\theta}$.

Note: Want $R_{\underset{\sim}{x},\underset{\sim}{y}}(\underset{\sim}{\theta})$ constant as a function of $\underset{\sim}{\theta}$ iff $\underset{\sim}{T}(\underset{\sim}{x}) = \underset{\sim}{T}(\underset{\sim}{y})$.

Actually want $f(\underset{\sim}{x}|\underset{\sim}{\theta}) = c_{\underset{\sim}{x},\underset{\sim}{y}}\, f(\underset{\sim}{y}|\underset{\sim}{\theta}) \ \forall \underset{\sim}{\theta} \in \Theta$

so ignore $\underset{\sim}{\theta}$ such that $R_{\underset{\sim}{x},\underset{\sim}{y}}(\underset{\sim}{\theta}) = \frac{0}{0}$.

But $R_{\underset{\sim}{x},\underset{\sim}{y}}(\underset{\sim}{\theta}) = 0$ or $R_{\underset{\sim}{x},\underset{\sim}{y}}(\underset{\sim}{\theta}) = \infty$ means $\underset{\sim}{T}$ is not min suff.

25) 11/12/5 Know Th 4.5 Let $Y_1, \ldots, Y_n$ be iid from a k parameter exp family

$$f(y|\underline{\theta}) = h(y)\, c(\underline{\theta}) \exp[w_1(\underline{\theta})\, t_1(y) + \cdots + w_k(\underline{\theta})\, t_k(y)]$$

with natural parameterization

$$f(y|\underset{\sim}{\eta}) = h(y)\, b(\underset{\sim}{\eta}) \exp(\eta_1 t_1(y) + \cdots + \eta_k t_k(y)).$$

As usual, assume $\Omega = \{\underset{\sim}{\eta} \mid \eta_i = w_i(\underline{\theta}),\ \underline{\theta} \in \Theta\}$ except for the χ^2_p and inverse Gaussian distributions.

Let $\underset{\sim}{T}(\underset{\sim}{y}) = \left(\sum_{i=1}^n t_1(y_i), \ldots, \sum_{i=1}^n t_k(y_i)\right) = (T_1(\underset{\sim}{Y}), \ldots, T_k(\underset{\sim}{Y}))$.
Then $\underset{\sim}{T}(\underset{\sim}{Y})$ is minimal sufficient for $\underline{\theta}$ and $\underset{\sim}{\eta}$ if the $\eta_i = w_i(\underline{\theta})$ do not satisfy a linearity constraint.
$\underset{\sim}{T}(\underset{\sim}{Y})$ is complete sufficient for $\underline{\theta}$ and $\underset{\sim}{\eta}$ if Ω contains a k dimensional rectangle.

26) Know Cor 4.6 p134) If $Y_1, \ldots, Y_n$ are iid from a kP-REF, then $\underset{\sim}{T}(\underset{\sim}{Y})$ is sufficient, minimal sufficient and complete sufficient for $\underline{\theta}$ and $\underset{\sim}{\eta}$.

proof min suff and complete suff follow from Th 4.5 since η_i do not satisfy a linearity constraint and the k dim open set Ω contains a k dim rectangle

proof of sufficiency $f(\underline{y}|\underline{\theta}) = \prod_{j=1}^n f(y_j|\underline{\theta}) =$

$$\underbrace{\left[\prod_{j=1}^n h(y_j)\right]}_{h_s(\underline{y}) \geq 0} \underbrace{[c(\underline{\theta})]^n \exp\left[w_1(\underline{\theta}) \sum_{j=1}^n t_1(y_j) + \cdots + w_k(\underline{\theta}) \sum_{j=1}^n t_k(y_j)\right]}_{g(\underline{T}(\underline{y})|\underline{\theta}) \geq 0}$$

So true by factorization.

proof of min suff for Th 4.5 like HW6 problem 1 HW6# 28.5 6

$$R_{\underset{\sim}{x},\underset{\sim}{y}}(\underset{\sim}{\eta}) = \frac{f(\underset{\sim}{x}|\underset{\sim}{\eta})}{f(\underset{\sim}{y}|\underset{\sim}{\eta})} = \frac{\prod_j h(x_j)}{\prod_j h(y_j)} \frac{[b(\underset{\sim}{\eta})]^n}{[b(\underset{\sim}{\eta})]^n} \frac{\exp\left[\sum_{i=1}^k \eta_i T_i(\underset{\sim}{x})\right]}{\exp\left[\sum_{i=1}^k \eta_i T_i(\underset{\sim}{y})\right]}$$

$$= \frac{\prod_j h(x_j)}{\prod_j h(y_j)} \exp\left[\sum_{i=1}^k \eta_i \left(T_i(\underset{\sim}{x}) - T_i(\underset{\sim}{y})\right)\right]$$

is equal to a constant wrt $\underset{\sim}{\eta}$ iff

$$\sum_{i=1}^k \eta_i \left[T_i(\underset{\sim}{x}) - T_i(\underset{\sim}{y})\right] = \sum_{i=1}^k \eta_i a_i = c_{\underset{\sim}{x},\underset{\sim}{y}} \ \forall \underset{\sim}{\eta}$$

Since the η_i do not satisfy a linearity constraint, $R_{\underset{\sim}{x}\underset{\sim}{y}}(\underset{\sim}{\eta}) = c_{\underset{\sim}{x},\underset{\sim}{y}} \ \forall \underset{\sim}{\eta}$ iff all

$a_i = 0$ iff $T_i(\underset{\sim}{x}) = T_i(\underset{\sim}{y})$ iff $\underset{\sim}{T}(\underset{\sim}{x}) = \underset{\sim}{T}(\underset{\sim}{y})$,

So $\underset{\sim}{T}(\underset{\sim}{Y})$ is minimal sufficient by LSM.

27] **Know** Tips for finding a minimal sufficient statistic.

i) Use Factorization to find a candidate suff stat $\underset{\sim}{T}$ with k as small as possible. Check that $\underset{\sim}{T}$ is min suff with LSM.

ii) Recognize or prove that the family is a kP-REF and find $\underset{\sim}{T}(\underset{\sim}{Y})$.

iii) Recognize or prove that the family is a K par exp fam (not necessarily REF)

and a) apply LSM to $\underline{T}(\underline{Y}) = \left(\sum_j t_1(Y_j), \ldots, \sum_j t_K(Y_j)\right)$

or b) apply Th 4.5 by showing that $\eta_1, \ldots, \eta_K$ do not satisfy a linearity constraint.

iv) If the support depends on unknown parameter θ, often $Y_{(1)}$ and/or $Y_{(n)}$ will be a component of the min suff stat.

v) If the family is not an exp family and after factorization $k = n$, try LSM on the order statistics $(Y_{(1)}, \ldots, Y_{(n)})$ (often very hard).

ex) $Y \sim$ Weibull (ϕ, λ). min suff stat is $(Y_{(1)}, \ldots, Y_{(n)})$. But if ϕ is known

$$f(\underline{y}|\lambda) = \prod_{i=1}^n f(y_i) = \prod_{i=1}^n \frac{\phi}{\lambda} y_i^{\phi-1} e^{-y_i^\phi/\lambda} I(y_i > 0)$$

$$= \left(\frac{\phi}{\lambda}\right)^n \prod_{i=1}^n y_i^{\phi-1} e^{-\frac{1}{\lambda}\sum_{i=1}^n y_i^\phi} \prod I(y_i > 0)$$

$$\text{so } R_{\underline{x}\underline{y}}(\lambda) = \frac{f(\underline{x}|\lambda)}{f(\underline{y}|\lambda)} = \underbrace{\frac{\prod I(x_i>0) \prod x_i^{\phi-1}}{\prod I(y_i>0) \prod y_i^{\phi-1}}}_{\text{constant wrt } \lambda \text{ if } \phi \text{ is known}} e^{-\frac{1}{\lambda}\left(\sum x_i^\phi - \sum y_i^\phi\right)} \stackrel{\text{set}}{=} c_{\underline{x}\underline{y}} \ \forall \lambda$$

iff $e^{-\frac{1}{\lambda}\left(\sum x_i^\phi - \sum y_i^\phi\right)} \equiv c' \ \forall \lambda$ iff

$\sum x_i^\phi = \sum y_i^\phi$. $\therefore T(\underline{Y}) = \sum_{i=1}^n Y_i^\phi$ is

min suff by LSM. Note that $f(y|\lambda)$

$$= \underbrace{y^{\phi-1} I(y>0)}_{h(y)} \underbrace{\frac{\phi}{\lambda}}_{c(\lambda)} \exp\Big[\underbrace{\frac{-1}{\lambda}}_{w(\lambda)} \underbrace{y^\phi}_{t(y)}\Big]$$

is a 1PREF if ϕ is known.

So $T(\underline{Y}) = \sum_{i=1}^{n} t(Y_i) = \sum_{i=1}^{n} Y_i^{\phi}$ is min suff and complete suff.

29] In using LSM, it is crucial to treat $\underline{x}$ and $\underline{y}$ as constants. The only variable in $R_{\underline{x},\underline{y}}(\underline{\theta})$ is $\underline{\theta}$.

Analogy MLE $L(\theta|\underline{x})$

ex] $f(x) = \dfrac{\log(\theta)}{\theta - 1} \theta^x \quad 0 < x < \infty, \ \theta > 1$

$f(x) = \underbrace{\dfrac{\log(\theta)}{\theta - 1}}_{c(\theta) > 0} \underbrace{I(x > 0)}_{h(x) \geq 0} \exp(\underbrace{\log(\theta)}_{w(\theta)} \underbrace{x}_{t(x)})$ $n = 1$

$\eta = \log(\theta), \theta > 1$ so $\Omega = \underbrace{(0, \infty)}_{\text{contains 1dim rectangle = interval}}$

So $\sum_{i=1}^{n} X_i$ is complete suff by Th 4.5.

For exp fam try LSM on $T(\underline{Y})$

29] Tips for proving $\underline{T}(\underline{Y})$ is complete.

i) Note that g can depend on n since n is known, but g can't depend on $(\underline{\theta})$.

ii) If $\underline{Y}$ is a kP-REF with $\dim \Theta = k$, then $\underline{T}(\underline{Y}) = \left(\sum_{j=1}^{n} t_1(Y_j), \ldots, \sum_{j=1}^{n} t_k(Y_j) \right)$ is complete.

iii) If $Y_1, \ldots, Y_n$ are iid $U(\theta_1, \theta_2)$,