Approaches to prove completeness of suff statistic

(I) s-parameter exponential family argument

Def^n

X has a dist^n of s-parameter exponential family if it's p.d.f. or p.m.f. is of the form

$$f(x) = h(x) \exp\left(\sum_{i=1}^{s} \eta_i(\theta) T_i(x) - \beta(\theta) \right)$$

or $$f(x) = h(x) \exp\left(\sum_{i=1}^{s} \eta_i T_i(x) - A(\eta) \right)$$

in the reparametrized form (in terms of η parametrization)

$\{ \underset{\sim}{\eta} : \underset{\sim}{\theta} \in \Theta \}$ – is called the natural parameter space

Remark: Many of the common dist^{n}s follow exponential family dist^n setup for some 's'.

e.g. Normal, exponential (scale), gamma, chi-square, log-normal, beta, Bernoulli, Binomial (n known), Poisson, geometric, negative binomial, etc.

Remark: Distributions for which range is dependent on parameter, do not ~~be~~ belong to exponential family dist^n setup.

e.g. $U(0, \theta)$, $U(\theta - \frac{1}{2}, \theta + \frac{1}{2})$, $U(-\theta, \theta)$ expenential (location), exponential (location-scale).

Remark: If the natural parameter space associated with an s-parameter exponential family dist^n contains an s-dimensional open rectangle (open interval for $s=1$), then the s-parameter exponential family dist^n is said to be of "full rank".

An important result:

If an s-parameter exponential family dist^n is of full rank, then the associated minimal sufficient statistic is complete.

i.e. $T(\underline{X}) = \left(\sum_{j=1}^{n} T_1(X_j), \sum_{j=1}^{n} T_2(X_j), \ldots, \sum_{j=1}^{n} T_s(X_j) \right)$

is complete sufficient

<u>Remark</u>: The above result can be used to prove completeness of minimal sufficient for all such distributions.

<u>Examples</u>

(i) $X \sim P(\theta)$ $\theta > 0$ $\Theta = \{\theta : \theta > 0\}$

p.m.f. $f_\theta(x) = \dfrac{e^{-\theta}\,\theta^x}{x!}$ $x = 0, 1, \ldots$

$= \dfrac{1}{x!}\, e^{x \log \theta - \theta}$

$h(x) = \dfrac{1}{x!}$; $T_1(x) = x$; $\eta_1(\theta) = \log \theta$; $\beta(\theta) = \theta$

i.e. $f_\eta(x) = \dfrac{1}{x!} \exp\left(\eta_1 T_1(x) - A(\eta)\right)$

This is 1-parameter exponential family form with natural parameter space as

$$\{\eta : \eta \in \mathbb{R}\}$$

The above natural parameter space contains open rectangles and hence the 1-parameter exponential family is of full rank

$\Rightarrow T(\underline{X}) = \sum_{i=1}^{n} X_i$ is complete suff stat

(2) $X \sim B(1,\theta)$ $1 > \theta > 0$ $\quad \Theta = \{\theta : 0 < \theta < 1\}$

p.m.f $\quad f_\theta(x) = \theta^x (1-\theta)^{1-x} \quad x = 0, 1$

$$= \exp\left(x \log\left(\frac{\theta}{1-\theta}\right) + \log(1-\theta) \right)$$

with $h(x) = 1$, $T_1(x) = x$, $\eta(\theta) = \log \frac{\theta}{1-\theta}$; $\beta(\theta) = -\log(1-\theta)$

the above is 1-parameter exponential family

$$f_\eta(x) = \exp(x\eta - A(\eta))$$

Natural parameter space : $\{\eta : \eta \in \mathbb{R}\}$ contains open intervals

$\Rightarrow$ The above 1-param expo family distⁿ is of full rank

$\Rightarrow$ $T(\underline{X}) = \sum_{i=1}^{n} X_i$ is complete suff statistic

(3) $X \sim N(\mu, \sigma^2)$ $\quad \underset{\sim}{\theta} = (\mu, \sigma)'$

$\Theta = \{(\mu,\sigma) : \mu \in \mathbb{R}, \sigma > 0\}$

p.d.f. $\quad f_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{1}{2\sigma^2}(x^2 + \mu^2 - 2\mu x)\right)$

$$= \left(\frac{1}{\sqrt{2\pi}}\right) \exp\left[-\frac{1}{2\sigma^2}x^2 + \frac{\mu}{\sigma^2}x - \frac{\mu^2}{2\sigma^2} - \log\sigma\right)$$

$$= h(x) \exp\left(\sum_{i=1}^{2} T_i(x)\eta_i(\underset{\sim}{\theta}) - \beta(\underset{\sim}{\theta}) \right)$$

$h(x) = \frac{1}{\sqrt{2\pi}}$; $T_1(x) = x^2$; $\eta_1(\underset{\sim}{\theta}) = -\frac{1}{2\sigma^2} \ (= \eta_1)$

$T_2(x) = x$; $\eta_2(\underset{\sim}{\theta}) = \frac{\mu}{\sigma^2} \ (= \eta_2)$

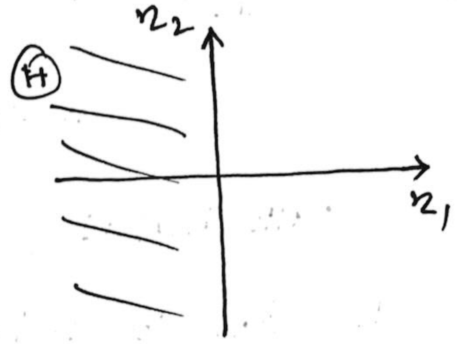
$\beta(\underset{\sim}{\theta}) = \frac{\mu^2}{2\sigma^2} + \log\sigma$

The above is 2-parameter exponential family representation

$$f_{\underset{\sim}{\eta}}(x) = \frac{1}{\sqrt{2\pi}} \exp\left(\eta_1 T_1(x) + \eta_2 T_2(x) - A(\underline{\eta}) \right)$$

with natural parameter space as

$$\{ (\eta_1, \eta_2) : \eta_1 < 0, \eta_2 \in \mathbb{R} \}$$



which contains 2-dim open rectangles

$$\Rightarrow T(\underline{X}) = \left(\sum_{i=1}^{n} X_i , \sum X_i^2 \right) \text{ is complete sufficient stat}$$

$$\Leftrightarrow \left(\bar{X}, \frac{1}{n-1} \sum (X_i - \bar{X})^2 = S^2 \right) \text{ is complete suff stat}$$

(4) example of non-full rank s-parameter expo family distⁿ

$X \sim N(\theta, \theta) \qquad \theta > 0 \; ; \; \Theta = \{ \theta : \theta > 0 \}$

p.d.f. $\quad f_\theta(x) = \frac{1}{\sqrt{2\pi}\,\theta} \exp\left(-\frac{1}{2\theta^2} (x^2 + \theta^2 - 2\theta x) \right)$

$$= \underbrace{\left(\frac{e^{-1/2}}{\sqrt{2\pi}} \right)}_{h(x)} \exp\left(-\frac{1}{2\theta^2} x^2 + \frac{1}{\theta} x - \log \theta \right)$$

$T_1(x) = x^2 \quad \eta_1(\theta) = -\frac{1}{2\theta^2} \; (=\eta_1)$

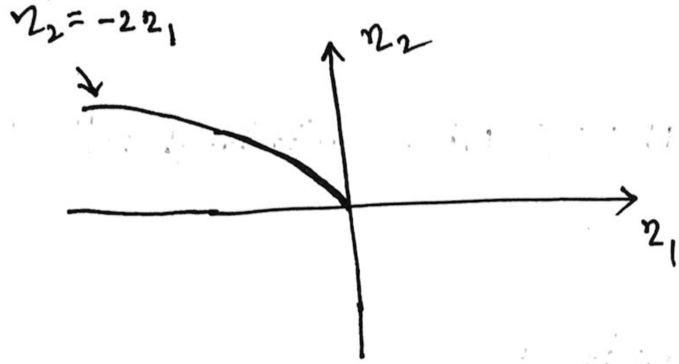
$T_2(x) = x \; ; \; \eta_2(\theta) = \frac{1}{\theta} \; (=\eta_2)$

The above is 2-parameter exponential family

$$f_{\underset{\sim}{\eta}}(x) = \left(\frac{e^{-1/2}}{\sqrt{2\pi}} \right) \exp\left(T_1(x)\eta_1 + T_2(x)\eta_2 - A(\underline{\eta}) \right)$$

The natural parameter space is

$$\left\{(\eta_1, \eta_2) : \eta_2^2 = -2\eta_1 , \eta_1 < 0, \eta_2 > 0\right\}$$



The natural parameter space is a curve and does not contain an open 2-dim rectangle

$\Rightarrow$ The 2-parameter expo family is not of full rank

Remark: We can show that here $\underset{\sim}{T} = (T_1, T_2)$; $T_1 = \sum X_i$ & $T_2 = \sum X_i^2$

is not complete

$$E\, T_2 = n\, E X_1^2 = n\, 2\theta$$

$$E\, T_1^2 = V(T_1) + (E\, T_1)^2 = n\theta^2 + (n\theta)^2$$

$$= \theta^2 n(n+1)$$

$$\Rightarrow E\left(\frac{T_1^2}{n(n+1)} - \frac{T_2}{2n}\right) = \theta^2 - \theta^2 = 0 \quad \forall \theta \in \Theta \qquad -(*)$$

$$\underline{\underline{\not\Rightarrow}} \quad \frac{T_1^2}{n(n+1)} = \frac{T_2}{2n} \quad \text{w.p. } 1 \text{ (a.e.)}$$

In fact $P\left(\frac{T_1^2}{n(n+1)} = \frac{T_2}{2n}\right) = 0$!!

(note that $\frac{T_1^2}{n(n+1)} - \frac{T_2}{2n}$ is a cont r.v.)