Theory of Statistical Inference - Lecture III.4
STA422 and STA2162

Michael Evans
University of Toronto
https://utstat.utoronto.ca/mikevans/sta422/sta4222026.html

2026

Definition III.4.1 A model $\{f_\theta : \theta \in \Theta\}$ is of *exponential form* if

$$f_\theta(x) = \exp\{\eta^t(\theta)t(x) - A(\theta)\}h(x)$$

where $t : \mathcal{X} \to \mathbb{R}^k$, $h : \mathcal{X} \to [0, \infty)$ and $\eta : \Theta \to \mathbb{R}^k$. ■

Example III.4.1 $N(\mu, \sigma^2)$ where $\theta = (\mu, \sigma^2) \in \Theta = \mathbb{R} \times [0, \infty)$

- then

$$\begin{aligned} f_{(\mu,\sigma^2)}(x) &= (2\pi\sigma^2)^{-1/2} \exp\{-(x-\mu)^2/2\sigma^2\} \\ &= (2\pi)^{-1/2} \exp\left\{ \begin{pmatrix} -1/2\sigma^2 & \mu/\sigma^2 \end{pmatrix} \begin{pmatrix} x^2 \\ x \end{pmatrix} - (\log\sigma^2 + \mu^2/\sigma^2)/2 \right\} \end{aligned}$$

is of exponential form with $\eta^t(\theta) = \begin{pmatrix} -1/2\sigma^2 & \mu/\sigma^2 \end{pmatrix}$, $t(x) = \begin{pmatrix} x^2 & x \end{pmatrix}^t$, $A(\theta) = (\log\sigma^2 + \mu^2/\sigma^2)/2$, $h(x) = (2\pi)^{-1/2}$ ■

- note - if $x = (x_1, \ldots, x_n)$ is iid from f_θ having exponential form, then the model for x has exponential form

$$f_\theta(x) = \prod_{i=1}^{n} f_\theta(x_i) = \exp\left\{\eta^t(\theta) \sum_{i=1}^{n} t(x_i) - A(\theta)\right\} \prod_{i=1}^{n} h(x_i)$$

and so it is immediate that $T(x) = \sum_{i=1}^{n} t(x_i)$ is a sufficient statistic

Definition III.4.2 The functions $\eta_1, \ldots, \eta_k : \Theta \to \mathbb{R}$ are *affinely independent* if, whenever $a + \sum_{i=1}^{k} b_i \eta_i(\theta) = 0$ for every $\theta \in \Theta$, then we must have $a = b_1 = \cdots = b_k = 0$.

- consider $\eta = (\eta_1, \ldots, \eta_k)$ in Def. III.4.1 and suppose that $a + \sum_{i=1}^{k} b_i \eta_i(\theta) = 0$ for every $\theta \in \Theta$ and $b_k \neq 0$

- then we have that $b_k\eta_k(\theta) = -a - \sum_{i=1}^{k-1} b_i\eta_i(\theta)$ and so

$$\begin{aligned}
\eta^t(\theta)t(x) &= \sum_{i=1}^{k} \eta_i(\theta)t_i(x) \\
&= \sum_{i=1}^{k-1} \eta_i(\theta)t_i(x) - b_k\left(a + \sum_{i=1}^{k-1} b_i\eta_i(\theta)\right)t_k(x) \\
&= \sum_{i=1}^{k-1} \eta_i(\theta)\left(t_i(x) - b_k b_i t_k(x)\right) - ab_k t_k(x) \\
&= \sum_{i=1}^{k-1} \eta_i(\theta)t_i^*(x) - ab_k t_k(x)
\end{aligned}$$

Proposition II.4.1 A model having exponential form can always be specified so that the η function has affinely independent components.

Proposition II.4.2 If a model has exponential form and the components of η are affinely independent, then t is a mss.
Proof: Suppose $L(\theta \mid x_1) = c(x_1, x_2) L(\theta \mid x_2)$ for all $\theta \in \Theta$ and so x_1, x_2 are in the same mss equivalence class. This implies that

$$c(x_1, x_2) = \frac{L(\theta \mid x_1)}{L(\theta \mid x_2)} = \exp\left(\eta^t(\theta)(t(x_1) - t(x_2))\right) \frac{h(x_1)}{h(x_2)}$$

for every θ and so $\eta^t(\theta)(t(x_1) - t(x_2))$ is constant in θ. Since the components of η are affinely independent this implies that $t(x_1) = t(x_2)$ which, by the sufficiency of t implies that t is a mss. ■

Example III.4.1 (continued)

- for a sample of n we have that $T(x) = \left(\sum_{i=1}^n x_i^2 \quad \sum_{i=1}^n x_i \right)^t$ is a sufficient statistic and since $(\bar{x}, \sum_{i=1}^n (x_i - \bar{x})^2)$ is a 1-1 function of T, then this is also a sufficient statistic

- now suppose, for every (μ, σ^2)

$$a + b_1(-1/2\sigma^2 + b_2(\mu/\sigma^2) = 0$$

- letting $\sigma^2 \to \infty$ implies $a = 0$, then putting $\mu = 0$ implies $b_1 =$ and then choosing (μ, σ^2) so that $\mu/\sigma^2 \neq 0$ implies $b_2 = 0$ so $\eta_1(\mu, \sigma^2) = -1/2\sigma^2$ and $\eta_2(\mu, \sigma^2) = \mu/\sigma^2$ are affinely independent which implies $(\bar{x}, \sum_{i=1}^n (x_i - \bar{x})^2)$ is a mss for this model ■

Exercise III.4.1 Show that the following models are of exponential form and determine a mss under iid sampling:
(i) $\{\text{Poisson}(\lambda) : \lambda > 0\}$,
(ii) $\{\text{gamma}(\alpha, \lambda) : \lambda > 0\}$,
(iii) $\{\text{multinomial}(n, p_1, \ldots, p_k) : p_1 \geq 0, \ldots, p_k \geq 0\}$,
(iv) $\{N_2(\mu, \Sigma) : \mu \in \mathbb{R}, \Sigma \in \mathbb{R}^{2\times 2} \text{ p.d.}\}$.

Proposition II.4.3 A model $\{f_\theta : \theta \in \Theta\}$ has exponential form iff the dimension of $\mathcal{L}\{\log f_\theta : \theta \in \Theta\}$ is finite.
Proof:
$\Longrightarrow$) We have that
$\log f_\theta(x) = \sum_{i=1}^{k} \eta_i(\theta) t_i(x) - A(\theta) + \log h(x) \in \mathcal{L}\{1, t_1, \ldots, t_k, \log h\}$ which has finite dimension.
$\Longleftarrow$) Then $\mathcal{L}\{\log f_\theta : \theta \in \Theta\}$ has a finite basis $\log f_{\theta_1}, \ldots, \log f_{\theta_k}$ and so $\log f_\theta(x) = \sum_{i=1}^{k} \eta_i(\theta) \log f_{\theta_i} = \sum_{i=1}^{k} \eta_i(\theta) t_i(x)$ for some $\eta : \Theta \to \mathbb{R}^k$. ■

note - when $\mathcal{L}\{\log f_\theta : \theta \in \Theta\}$ is finite dimensional with basis $\log f_{\theta_1}, \ldots, \log f_{\theta_k}$, then $t(x) = (\log f_{\theta_1}(x), \ldots, \log f_{\theta_k}(x))$ is sufficient and so the mss is finite dimensional so "maybe" the best definition for a mss to be finite dimensional is to say $\mathcal{L}\{\log f_\theta : \theta \in \Theta\}$ is finite dimensional which is equivalent to saying the model is of exponential form

Exercise III.4.2 Show that for an exponential family $f_\theta(x) = \exp\{\eta^t(\theta) t(x) - A(\theta)\} h(x)$, if there exists $\theta_0 \in \Theta$ and for some $\delta > 0$, the ball $B_\delta(\eta(\theta_0)) \subset \{\eta(\theta) : \theta \in \Theta\}$, then t is a mss. Hint: can the η_i be affinely dependent in $B_\delta(\eta(\theta_0))$?

Example III.4.2 *A model not of exponential form*

- suppose $f_\theta(x) = 1/\pi(1 + (x - \theta)^2)$ for $\theta \in \mathbb{R}$ (location Cauchy model)

- what is the mss when sampling from the location Cauchy?

- note that

$$\pi^{-n}\prod_{i=1}^{n}(1 + (x_i - \theta)^2)^{-1} = \pi^{-n}\prod_{i=1}^{n}(1 + (x_{(i)} - \theta)^2)^{-1}$$

and so the order statistic $(x_{(1)}, \ldots, x_{(n)})$ is sufficient

- consider the polynomial $\prod_{i=1}^{n}(1 + (x_{(i)} - \theta)^2)$ of degree $2n$ which has no real roots and complex roots $x_{(i)} \pm \sqrt{-1}$ which are distinct wp1

- therefore knowing the likelihood function we can compute the order statistic $(x_{(1)}, \ldots, x_{(n)})$ which proves that the order statistic is mss

- so there is no real reduction in this case

- if the dimension of $\mathcal{L}\{\log f_\theta : \theta \in \Theta\}$ is finite then this model would be of exponential from Proposition II.4.3

- and then the mss under sampling would be of the form $T(x) = \sum_{i=1}^n t(x_i)$ for some function $t : \mathcal{X} = \mathbb{R} \to \mathbb{R}^k$ where $k = \dim \mathcal{L}\{\log f_\theta : \theta \in \Theta\}$ and so $T : \mathbb{R}^n \to \mathbb{R}^k$

- but the order statistic $(x_{(1)}, \ldots, x_{(n)}) : \mathbb{R}^n \to \mathbb{R}^n$ and the probability that it lies in a lower dimensional subspace is 0 because the probability the sample $x = (x_1, \ldots, x_n)$ is in a lower dimensional subspace is 0 when sampling from the Cauchy (all lower dimensionsal subspaces have probability content 0)

- as such this model cannot be of exponential form and so $\dim \mathcal{L}\{\log f_\theta : \theta \in \Theta\} = \infty$ ■

- for more on exponential families see book Efron (2023) Exponential Families in Theory and Practice.