Stat 301

$(P_\theta : \theta \in \Theta)$ $\quad X_1, \cdots, X_n \overset{iid}{\sim} P_\theta$

Def: $T = T(X_1, \cdots, X_n)$ is sufficient iff $X \mid T$ does not depend on θ. for all $\theta \in \Theta$.

Def (Bayesian): T is sufficient iff $\theta \to T \to X$ is Markov chain.

$$(\theta \perp\!\!\!\perp X \mid T)$$

Theorem (factorization). suppose $(P_\theta : \theta \in \Theta)$ is continuous or discrete. (has pdf or pmf).

then T is sufficient $\Longleftrightarrow$ $p(x|\theta) = g_\theta(T(x))\, h(x)$. for some g_θ and h.

Proof (discrete): suppose $\underbrace{p(x|\theta)}_{\text{pmf}} = g_\theta(T(x))\, h(x)$.

$$\frac{p(x|\theta_1)}{p(x|\theta_0)} = \frac{g_{\theta_1}(T(x))}{g_{\theta_0}(T(x))}$$

$$\mathbb{P}(X=x \mid T=t) = \frac{\mathbb{P}(X=x, T=t)}{\mathbb{P}(T=t)}$$

$$\mathbb{P}(X=x, T=t) = \begin{cases} \mathbb{P}(X=x) & T(x)=t \\ 0 & T(x)\neq t \end{cases} = \mathbb{1}\{T(x)=t\}\,\mathbb{P}(X=x).$$

$$= \mathbb{1}\{T(x)=t\}\, g_\theta(T(x))\, h(x).$$

$$= \mathbb{1}\{T(x)=t\}\, g_\theta(t)\, h(x).$$

$$\mathbb{P}(T=t) = \sum_{x': T(x')=t} p(x'\mid\theta). = \sum_{x': T(x')=t} g_\theta(T(x'))\, h(x')$$

$$= \sum_{x': T(x')=t} g_\theta(t)\, h(x') = g_\theta(t) \sum_{x': T(x')=t} h(x')$$

$$\mathbb{P}(X=x \mid T=t) = \frac{\mathbb{1}\{T(x)=t\}\, h(x)}{\sum_{x': T(x')=t} h(x')}$$ does not depend on $\theta \Rightarrow T$ is sufficient.

Suppose T is sufficient.

$$p(x|\theta) = \mathbb{P}_\theta(X = x) = \mathbb{P}_\theta(X = x, T(X) = T(x)).$$

$$= \underbrace{\mathbb{P}_\theta(X = x \mid T(X) = T(x))}_{\text{does not depend on } \theta} \mathbb{P}_\theta(T(X) = T(x)),$$

$$= h(x) \cdot g_\theta(T(x))$$

$\square$

e.g. $X_1, \ldots, X_n \overset{iid}{\sim} N(\theta, 1)$.

$$p(x|\theta) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x_i-\theta)^2} = \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum_{i=1}^n (x_i-\theta)^2}$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum_{i=1}^n x_i^2 - \frac{1}{2}n\theta^2 + \theta\sum_{i=1}^n x_i}$$

$$= \left(\left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum_{i=1}^n x_i^2}\right)\left(e^{-\frac{1}{2}n\theta^2 + \theta n \overline{x}}\right) \Rightarrow \overline{X} \text{ is sufficient.}$$

e.g. $X_1, \dots, X_n \overset{iid}{\sim} \mathrm{Unif}(0, \theta)$.

$$p(x|\theta) = \prod_{i=1}^{n} \left(\frac{1}{\theta} \mathbb{1}\{0 < x_i < \theta\} \right).$$

$$= \theta^{-n} \prod_{i=1}^{n} \mathbb{1}\{0 < x_i < \theta\} = \theta^{-n} \mathbb{1}\{0 < \min_{1\le i\le n} x_i, \max_{1\le i\le n} x_i < \theta\}$$

$$= \left(\mathbb{1}\{0 < \min_{1\le i\le n} x_i\} \right) \left(\mathbb{1}\{\max_{1\le i\le n} x_i < \theta\} \theta^{-n} \right).$$

$\Rightarrow \max_{1\le i\le n} x_i$ is sufficient.

exponential family.

$$p(x|\theta) = \exp\left(\sum_{j=1}^{d} \underbrace{\eta_j(\theta)}_{\text{natural parameter}} \underbrace{T_j(x)}_{\text{sufficient statistic}} - \underbrace{B(\theta)}_{\text{log-partition function}} \right) \underbrace{h(x)}_{\text{base measure}}.$$

pdf or pmf

$$B(\theta) = \log \int e^{\sum_{j=1}^{d} g_j(\theta) T_j(x)} h(x)\, d\mu(x).$$

e.g. exponentiated distribution $\mathrm{Exp}(\theta)$.

$$p(x|\theta) = \theta e^{-x\theta} \mathbb{1}\{x \geq 0\}.$$

$$= \exp(-\theta x + \log\theta)\, \mathbb{1}\{x \geq 0\}$$

e.g. $N(\mu, \sigma^2)$. $\theta = (\mu, \sigma^2)$.

$$p(x|\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}(x-\mu)^2/\sigma^2}.$$

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

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

$$X_1, \cdots, X_n \overset{iid}{\sim} p(x|\theta) = e^{\sum_{j=1}^{d} \eta_j(\theta) T_j(x) - B(\theta)} h(x).$$

$$\Rightarrow p(x_1, \cdots, x_n|\theta) = \exp\left(\sum_{j=1}^{d} \eta_j(\theta) \left(\sum_{i=1}^{n} T_j(x_i)\right) - n B(\theta)\right) \prod_{i=1}^{n} h(x_i).$$

sufficient statistic is $T = \left(\sum_{i=1}^{n} T_1(x_i), \cdots, \sum_{i=1}^{n} T_d(x_i)\right)$.

canonical form $p(x|\eta) = \exp\left(\sum_{j=1}^{d} \eta_j T_j(x) - A(\eta)\right) h(x).$

$$A(\eta) = \log \int e^{\sum_{j=1}^{d} \eta_j T_j(x)} h(x)\, d\mu(x).$$

Def: an exponential family $(P_\eta : \eta \in H)$ (of canonical form) is minimal if its dimension cannot be reduced.

(sufficient stats are linearly independent, natural parameters are linearly independent) e.g. (non-minimal examples).

$$p(x|\eta) = \exp\left(\eta_1 T(x) + \eta_2 (3T(x) + 2) - A(\eta)\right).$$

$$= \exp\left((\eta_1 + 3\eta_2) T(x) + 2\eta_2 - A(\eta)\right).$$

$$p(x|\eta) = \exp\left(\eta T_1(x) + (4 - 5\eta) T_2(x) - A(\eta)\right).$$

$$= \exp\left(\eta \left(T_1(x) - 5T_2(x)\right) - A(\eta)\right) \exp\left(4T_2(x)\right)$$

two types of minimal exponential family $(P_\eta : \eta \in H)$.

① full rank: H contains an open d-dimensional rectangle.

② curved: $\eta_1, \dots, \eta_d$ are related in non-linear ways.

e.g. $N(\mu, \sigma^2)$ $\quad p(x|\mu,\sigma^2) = \exp\left(-\frac{X^2}{2\sigma^2} + \frac{\mu}{\sigma^2}X - \frac{\mu^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2)\right)$

$$\begin{cases} T_1(x) = -X^2 \\ T_2(x) = X \end{cases} \qquad \begin{cases} \eta_1 = \frac{1}{2\sigma^2} \\ \eta_2 = \frac{\mu}{\sigma^2} \end{cases}$$

* $\mu = \sigma^2$, $N(\sigma^2, \sigma^2)$ $\quad \eta_2 = 1$. $\quad$ non-minimal.

* $\mu = \sqrt{\sigma^2}$ $\quad \begin{cases} \eta_1 = \frac{1}{2\sigma^2} \\ \eta_2 = \frac{1}{\sqrt{\sigma^2}} \end{cases} \quad \eta_2^2 = 2\eta_1$ $\quad$ minimal & curved.

* μ & σ^2 do not have constraints, minimal & full rank.

$$H = (0, \infty) \times \mathbb{R}$$

minimal sufficiency.

$X_1, \dots, X_n \overset{iid}{\sim} N(\theta, 1)$

$$\begin{cases} T_1 = (X_1, \dots, X_n) \\ T_2 = (X_1 + X_2, X_3 + X_4, \dots, X_{n-1} + X_n) \\ T_3 = \left(\sum_{i \le n/2} X_i, \sum_{i > n/2} X_i \right) \end{cases}$$

$$T_4 = \sum_{i=1}^{n} X_i$$

Def: S is minimal sufficient iff it is sufficient & for every sufficient T,

S is a function of T.

how to find minimal sufficient statistic?

① sub-family method.

Lemma: suppose $\Theta_0 \subseteq \Theta$, S is minimal sufficient for $(P_\theta : \theta \in \Theta_0)$
and sufficient for $(P_\theta : \theta \in \Theta)$
it is minimal sufficient for $(P_\theta : \theta \in \Theta)$.

Proof: suppose there is a sufficient T for $(P_\theta : \theta \in \Theta)$.

$\Rightarrow$ T is sufficient for $(P_\theta : \theta \in \Theta_0)$.

since S is minimal sufficient for $(P_\theta : \theta \in \Theta_0)$, S is a function of T.

$\Rightarrow$ S is minimal sufficient for $(P_\theta : \theta \in \Theta)$ $\square$

Theorem: for $(P_\theta : \theta \in \{\theta_0, \theta_1, \cdots, \theta_d\})$ (common support).

$$T(x) = \left(\frac{P_{\theta_1}(x)}{P_{\theta_0}(x)}, \frac{P_{\theta_2}(x)}{P_{\theta_0}(x)}, \cdots, \frac{P_{\theta_d}(x)}{P_{\theta_0}(x)} \right)$$

is minimal sufficient.

Proof:

$$\begin{cases} P_{\theta_0}(x) = P_{\theta_0}(x) \\ P_{\theta_j}(x) = T_j(x) \, P_{\theta_0}(x), \quad j = 1, \cdots, d. \end{cases}$$

$$g_{\theta_j}(T(x)) = \begin{cases} 1 & j = 0 \\ T_j(x) & j = 1, \cdots, k. \end{cases} \qquad h(x) = P_{\theta_0}(x).$$

$\Rightarrow$ T is sufficient by factorization.

for any T' sufficient $\Rightarrow$ likelihood ratio is function of T'

$\Rightarrow$ T is function of T' $\Rightarrow$ T is minimal $\square$

e.g. $X_1, \dots, X_n \overset{iid}{\sim} \mathrm{Bern}(\theta)$ $\theta \in [0,1]$.

$\sum_{i=1}^{n} X_i$ is sufficient

Consider $\theta_0 = 0.5$, $\theta_1 = 0.6$

$$\frac{p(x|\theta_1)}{p(x|\theta_0)} = \frac{\theta_1^{\sum_{i=1}^n X_i}(1-\theta_1)^{n-\sum_{i=1}^n X_i}}{\theta_0^{\sum_{i=1}^n X_i}(1-\theta_0)^{n-\sum_{i=1}^n X_i}} = \left(\frac{\theta_1(1-\theta_0)}{\theta_0(1-\theta_1)}\right)^{\sum_{i=1}^n X_i}\left(\frac{1-\theta_1}{1-\theta_0}\right)^n$$

$$= \left(\frac{3}{2}\right)^{\sum_{i=1}^n X_i}\left(\frac{4}{5}\right)^n.$$ equivalent $\sum_{i=1}^n X_i$

$\Rightarrow$ $\sum_{i=1}^n X_i$ is minimal sufficient.