Chapter. 6. Principle. of. Data. Reduction.

Principles of. Data. Reduction:

1. Sufficient principle: use. a statistics, that contain more. information about θ.
2. Likelihood. principle: A function of parameters determined by data $\Rightarrow$, what. value. of θ. makes the observed data most. likely.
3. Equivalent. principle: Are. the, data reduction method processing, those. important. features.

Definition. Sufficient. Statistics

A. Statistics $T(x)$ is sufficient. statistics for θ if conditional. distribution of. sample. x. given the. value. of $T(x)$ does not depends on θ.

Example:. Sufficient. Statistics.

$\{X_i\}$ denotes number of accident. in NYC on a particular day

Assume. independent. with. poission distribution.

$X_i \sim$ poission(λ).

$$f(\underset{\sim}{x}|\lambda) = f(x_1 \dots x_n|\lambda) \quad \text{Joint. PDF}$$
$$= \frac{e^{-\lambda}\lambda^{x_1}}{x_1!} \cdot \frac{e^{-\lambda}\lambda^{x_2}}{x_2!} \cdots \frac{e^{-\lambda}\lambda^{x_n}}{x_n!}$$
$$= \frac{e^{-n\lambda}\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}$$

$T(x) = \sum_{i=1}^{n} x_i$ is the sufficient. statistics.

proof. $T(x) = \sum_{i=1}^{n} x_i$ is a sufficient. statistics. by. definition of. sufficient.

$$p(X_1 = x_1, X_2 = x_2 \dots X_n = x_n \mid T(x) = t).$$
$$= \frac{p(X_1 = x_1 \dots X_n = x_n, T(x) = t)}{p(T(x) = t)} = \frac{p(X_1 = x_1 \dots X_n = x_n, \sum_{i=1}^{n} X_i)}{p(\sum_{i=1}^{n} x_i)}.$$
$$= \frac{p(X_1 = x_1 \dots X_n = x_n)}{p(\sum_{i=1}^{n} x_i)}. \quad \text{Since. } \{X_1 = x_1 \dots X_n = x_n\} \subseteq \{T(\underset{\sim}{x}) = t(\underset{\sim}{x})\}$$
$$= \frac{e^{-\lambda}\lambda^{x_1}, \dots, e^{-\lambda}\lambda^{x_n}}{e^{-n\lambda}\lambda^{\sum_{i=1}^{n} x_i}} = \sim \text{ independent. of. } \lambda.$$

Theorem 6.2.2.

If $f(\underset{\sim}{x}|\theta)$ is a joint PDF of $\underset{\sim}{X}$ and $q(t|\theta)$ is PDF or PMF of $T(x)$

Then $T(\underset{\sim}{X})$ is a sufficient statistic for θ if for every x in the sample space the ratio of $\dfrac{f(\underset{\sim}{x}|\theta)}{q(T(\underset{\sim}{x})|\theta)}$ is constant as a function of θ.

proof:

$P_\theta(X_1=x_1 \dots X_n=x_n) = P_\theta(X_1=x_1 \dots X_n=x_n, T(\underset{\sim}{x})=t(\underset{\sim}{x}))$. $\{\underset{\sim}{x}\} \subseteq \{T(\underset{\sim}{x})\}$.

$P_\theta(\underset{\sim}{X}=\underset{\sim}{x}) = P_\theta(\underset{\sim}{X}=\underset{\sim}{x} \mid T(\underset{\sim}{X})=t(\underset{\sim}{x})) \cdot P_\theta(T(\underset{\sim}{x})=t(\underset{\sim}{x}))$.

$$\frac{P_\theta(\underset{\sim}{X}=\underset{\sim}{x})}{P_\theta(T(\underset{\sim}{x})=t(\underset{\sim}{x}))} = \underbrace{P_\theta(\underset{\sim}{X}=\underset{\sim}{x} \mid T(\underset{\sim}{X})=t(\underset{\sim}{x}))}_{\text{independent of } \theta \text{ if } T(\underset{\sim}{x}) \text{ is a sufficient statistics}}$$

Theorem: ~~Factorization~~ Factorization Theorem (SB 6.6.2).

Let $f(\underset{\sim}{x}|\theta)$ denote the joint PDF or PMF of $\underset{\sim}{X}$. A statistics $T(\underset{\sim}{X})$ is a sufficient statistics for θ iif there exists function $g(t|\theta)$ and $h(x)$ s.t. for all sample points x and all parameter points θ,

$f(\underset{\sim}{x}|\theta) = g(T(x)|\theta)\,h(x)$.

proof: discrete case. prove both necessary & sufficient.

Suppose $T(x)$ is a sufficient statistics. Choose $g(t|\theta) = P_\theta(T(x)=t)$ and $h(x) = P(\underset{\sim}{X}=\underset{\sim}{x} \mid T(\underset{\sim}{X})=t(\underset{\sim}{x}))$. Because $T(x)$ is sufficient, the conditional probability defining $h(x)$ does not depends on θ. Thus the choice of $h(x)$ & $g(t|\theta)$ is legitimate, and for this choice, we have

$$\begin{aligned} f(\underset{\sim}{x}|\theta) &= P_\theta(\underset{\sim}{X}=\underset{\sim}{x}) \\ &= P_\theta(\underset{\sim}{X}=\underset{\sim}{x}, T(\underset{\sim}{x})=t(\underset{\sim}{x})) \\ &= P_\theta(T(\underset{\sim}{X})=t(\underset{\sim}{x})) \cdot P_\theta(\underset{\sim}{X}=\underset{\sim}{x} \mid T(\underset{\sim}{X})=t(\underset{\sim}{x})) \\ &= g(T(\underset{\sim}{x})|\theta)\,h(\underset{\sim}{x}) \end{aligned}$$

The above exhibited the factorization theorem. We also see from the last two lines above that

$P_\theta(T(\underset{\sim}{X})=T(\underset{\sim}{x}))=g(T(\underset{\sim}{x})|\theta)$. so, $g(T(\underset{\sim}{x})|\theta)$ is PDF or PMF of $T(x)$.

Now, we assume factorization theorem exists. let $q(t|\theta)$ be the PMF of $T(x)$. Show that $T(x)$ is sufficient, we examine the ratio $\frac{f(\underset{\sim}{x}|\theta)}{q(T(x)|\theta)}$

Define $A_{T(x)}=\{\underset{\sim}{y}: T(\underset{\sim}{y})=T(x)\}$. Then,

$$\frac{f(\underset{\sim}{x}|\theta)}{q(T(\underset{\sim}{x})|\theta)}=\frac{g(T(\underset{\sim}{x})|\theta)h(\underset{\sim}{x})}{q(T(\underset{\sim}{x})|\theta)}$$ Since factorization theorem exists.

$$=\frac{g(T(\underset{\sim}{x})|\theta)h(\underset{\sim}{x})}{\sum_{A_{T(x)}} f(\underset{\sim}{y}|\theta)}$$ $\left(T(\underset{\sim}{x})=\sum x_i. \quad \begin{matrix} x_1=1 & x_1=2 \\ x_2=2 & x_2=1 \end{matrix}\right)$

$P(T(x))=f(x_1=1, x_2=2)+f(x_1=2, x_2=1)$.

$$=\frac{g(T(\underset{\sim}{x})|\theta)h(\underset{\sim}{x})}{\sum_{A_{T(x)}} g(T(\underset{\sim}{x})|\theta)h(\underset{\sim}{y})}$$

$$=\frac{g(T(\underset{\sim}{x})|\theta)h(\underset{\sim}{x})}{g(T(\underset{\sim}{x})|\theta)\sum_{A_{T(x)}} h(\underset{\sim}{y})}$$ Since T is constant on $A_{T(x)}$

$$=\frac{h(\underset{\sim}{x})}{\sum_{A_{T(x)}} h(\underset{\sim}{y})}$$ $\sim$ not depends on θ.

Since the ratio does not depends on θ. by theorem 6.2.2.

$T(x)$ is a sufficient statistics for θ.

Various version of this proved by Fisher & Savage.

See "Testing statistical Hypothesis by F. Lehman

Also see Billingsly P471 for general framework.

Example: Poission Re-visit.

$$\left.\begin{aligned} f_\lambda(x_1 \ldots x_n) &= \frac{e^{-n\lambda}\cdot\lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} \\ &= h(\underset{\sim}{x})\, e^{-n\lambda}\cdot\lambda^{\sum_{i=1}^n x_i} \end{aligned}\right\} \Rightarrow$$

$\sum_{i=1}^n x_i$ is a sufficient Statistics.

$\updownarrow$ one-to-one function.

$\frac{1}{n}\sum_{i=1}^n x_i=\bar{x}$ is also a sufficient statistics.

This form holds for any distribution in general.

Example. Uniform distribution. German Tank Problem

Use serial number on Tank to estimate total number of tank θ

Set $\{x_i\} \sim$ uniform distribution

$f(x|\theta) = \frac{1}{\theta}$ for $x = 1, 2, \cdots \theta$.

Select n serial number without replacement We will have C_n^θ possible Samples each with PDF $\frac{1}{C_n^\theta}$

$$f(\underset{\sim}{x}|\theta) = \frac{1}{C_n^\theta} \cdot I(1 \le x_1 \le \theta,\ 1 \le x_2 \le \theta \ldots 1 \le x_n \le \theta).$$

$$= (C_n^\theta)^{-1} \cdot I(\max(x_i) \le \theta).$$

$$= (C_n^\theta)^{-1} \cdot I(x_{(n)}) \quad n^{th} \text{ order statistic.}$$

let $h(\underset{\sim}{x}) = 1$. by factorization theorm.

$X_{(n)}$ is a sufficient Statistic for θ.

However, $E[X_{(n)}] < \theta$. so it's a biased sufficient statistic

So it need to be rescaled to get the unbiased estimator.

Note: Sufficient Statistics with extreme value Statistics.

Typically, in examples where support of distribution depends on some unknown parameters (e.g. uniform $(0, \theta)$) The sufficient statistics will involve some extreme order statistics.

Example: Sufficient Statistics for Normal distribution.

$\{x_i\} \sim N(\mu, \sigma^2)$. Both μ & σ^2 are unknown.

$$f(\underset{\sim}{x}|\mu, \sigma^2) = (2\pi)^{-\frac{n}{2}} (\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2} \frac{\sum_{i=1}^n (x_i-\mu)^2}{\sigma^2}\right\}$$

$$= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \bar{x} + \bar{x} - \mu)^2\right\}.$$

$$= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i-\bar{x})^2 + n(\bar{x}-\mu)^2 + 2 \underbrace{\sum_{i=1}^n (x_i-\bar{x})}_{=0}(\bar{x}-\mu)\right\}.$$

$$= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i-\bar{x})^2 + n(\bar{x}-\mu)^2\right\}. \quad S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i-\bar{x})^2.$$

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

By factorization Theorem, $T(x) = (\bar{x}, S^2)$ are sufficient statistics for (μ, σ^2).

Example: Sufficient Statistics for Multi-variate Normal distribution.

$$\underset{\sim}{x_i} = \begin{pmatrix} x_{i1} \\ \vdots \\ x_{ip} \end{pmatrix} \sim N(\underset{\sim}{\mu}, \underset{\sim}{\Psi})$$

$$f(\underset{\sim}{x_1}, \underset{\sim}{x_2} \dots \underset{\sim}{x_n} \mid \mu, \Psi) = (2\pi)^{-\frac{np}{2}} |\underset{\sim}{\Psi}|^{-\frac{n}{2}} \exp\left\{ -\frac{1}{2} \sum_{i=1}^{n} (\underset{\sim}{x_i} - \underset{\sim}{\mu})' \Psi^{-1} (\underset{\sim}{x_i} - \underset{\sim}{\mu}) \right\}$$

$$= (2\pi)^{-\frac{np}{2}} |\underset{\sim}{\Psi}|^{-\frac{n}{2}} \exp\left\{ -\frac{1}{2} tr\left(\sum_{i=1}^{n} (\underset{\sim}{x_i} - \underset{\sim}{\mu})' \Psi^{-1} (\underset{\sim}{x_i} - \underset{\sim}{\mu}) \right) \right\}$$

$$= (2\pi)^{-\frac{np}{2}} |\underset{\sim}{\Psi}|^{-\frac{n}{2}} \exp\left\{ -\frac{1}{2} \sum_{i=1}^{n} tr\left(\underset{\sim}{\Psi}^{-1} (\underset{\sim}{x_i} - \underset{\sim}{\mu})' (\underset{\sim}{x_i} - \underset{\sim}{\mu}) \right) \right\}.$$

$$= (2\pi)^{-\frac{np}{2}} |\underset{\sim}{\Psi}|^{-\frac{n}{2}} \exp\left\{ -\frac{1}{2} tr\left(n \underset{\sim}{\Psi}^{-1} \sum_{i=1}^{n} (\underset{\sim}{x_i} - \underset{\sim}{\mu})' (\underset{\sim}{x_i} - \underset{\sim}{\mu}) \right) \right\}$$

$$= (2\pi)^{-\frac{np}{2}} |\underset{\sim}{\Psi}|^{-\frac{n}{2}} \exp\left\{ -\frac{1}{2} tr\left(n \underset{\sim}{\Psi}^{-1} \sum_{i=1}^{n} (\underset{\sim}{x_i} - \underset{\sim}{\bar{x}} + \underset{\sim}{\bar{x}} - \underset{\sim}{\mu})' (\underset{\sim}{x_i} - \underset{\sim}{\bar{x}} + \underset{\sim}{\bar{x}} - \underset{\sim}{\mu}) \right) \right\}.$$

$$= (2\pi)^{-\frac{np}{2}} |\underset{\sim}{\Psi}|^{-\frac{n}{2}} \exp\left\{ -\frac{1}{2} tr\left(n \underset{\sim}{\Psi}^{-1} \sum_{i=1}^{n} (\underset{\sim}{x_i} - \underset{\sim}{\bar{x}})' (\underset{\sim}{x_i} - \underset{\sim}{\bar{x}}) + n (\underset{\sim}{\bar{x}} - \underset{\sim}{\mu})' (\underset{\sim}{\bar{x}} - \underset{\sim}{\mu}) + 2 \sum_{i=1}^{n} \underbrace{(\underset{\sim}{x_i} - \underset{\sim}{\bar{x}})'}_{=0} (\underset{\sim}{\bar{x}} - \underset{\sim}{\mu}) \right) \right\}$$

$$= (2\pi)^{-\frac{np}{2}} |\underset{\sim}{\Psi}|^{-\frac{n}{2}} \exp\left\{ -\frac{1}{2} tr\left(n \underset{\sim}{\Psi}^{-1} (\underset{\sim}{S}^2 + n (\underset{\sim}{\bar{x}} - \underset{\sim}{\mu})' (\underset{\sim}{\bar{x}} - \underset{\sim}{\mu})) \right) \right\}.$$

By factorization theorem, $(\underset{\sim}{\bar{x}}, \underset{\sim}{S}^2)$ are sufficient statistics for $(\underset{\sim}{\mu}, \underset{\sim}{\Psi})$.

Definition: Minimal Sufficient Statistics.

A sufficient statistics $T(x)$ is called minimal sufficient statistics iff for any other sufficient statistics $T'(x)$, $T(x)$ is a function of $T'(x)$.

Theorem. Minimal Sufficient Statistics. 6.2.13.

let $f(x|\theta)$ be the PMF or PDF of a sample $\mathbf{x}$. Suppose there exists a function $T(x)$ s.t. for every two sample points x & y. The ratio of $\frac{f(x|\theta)}{f(y|\theta)}$ is constant as function of θ iff $T(x) = T(y)$. Then, $T(x)$ is a minimal sufficient statistics for θ.

Example: Minimal Sufficient Statistics for Normal distribution.

$\{x_i\} \sim N(\mu, \sigma^2)$.

let $\{x_1\} \sim N(\bar{x}_1, s_1^2)$ & $\{x_2\} \sim N(\bar{x}_2, s_2^2)$ are two sample points.

Apply theoerm 6.2.13

$$\frac{f(\underset{\sim}{x_1}\mid \mu, \sigma^2)}{f(\underset{\sim}{x_2}\mid \mu, \sigma^2)} = \frac{\cancel{(2\pi\sigma^2)}^{\frac{p}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_{1i}-\mu)^2\right\}}{\cancel{(2\pi\sigma^2)}^{\frac{p}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_{2i}-\mu)^2\right\}} = \exp\left\{-\frac{1}{2\sigma^2}\left[\frac{S_1^2}{\sigma^2}(n-1)-\frac{S_2^2}{\sigma^2}(n-1)+\frac{n(\bar{x}_2^2-\bar{x}_1^2)}{\sigma^2}\right]\right\}$$

is constant as function of (μ, σ^2) iif $\bar{x}_1 = \bar{x}_2$, $S_1^2 = S_2^2$

proving that $(\bar{x}, s^2)$ is minimal sufficient statistics for (μ, σ^2)

Note: Dimension of Minimal Sufficient Statistics

Dimension of Minimal Sufficient Statistics $\geq$ Dimension of θ.

Example: Dimension of minimal sufficient statistics > Dimension of θ

$\{X_i\} \sim$ Uniform $(\theta, \theta+1)$ where θ is one dimensional

But $(X_{(1)}, X_{(n)})$ is minimal sufficient statistic (2 dimensional)

Note: Uniqueness of Minimal Sufficient Statistics.

If T_1 & T_2 are both minimal sufficient statistics, Then by definition, each of them is a measurable function of the other.

So it's unique in the sense that they must be 1-to-1 function of each other.

Definition: Ancillary Statistics.

$S(\underset{\sim}{x})$ is ancillary statistics if its distribution doesn't dependent on θ

Example: Ancillary Statistics.

Suppose θ is a location parameter with CDF $F(x-\theta)$ F doesn't dependent on θ. (e.g. $\{X_i\} \sim N(\mu, 1)$ $F \sim \Phi \sim$ Standard Normal

$\Rightarrow X_i = Z_i + \theta$, Where Z_i is a standard Normal).

When we look at the CDF of Range $R = X_{(n)} - X_{(1)}$

$P(X_{(n)} - X_{(1)} \leq r) = P(R \leq r)$.

$$P(Z_{(n)} + \theta - Z_{(1)} - \theta \leq r) = P(\underbrace{Z_{(n)} - Z_{(1)}}_{\text{independent of } \theta} \leq r).$$

We have R as an ancillary statistics for location family.

Similar results for scale parameter.

$x_i = \sigma z_i$, if we look at the ratio. Any statistics that dependents on $\frac{x_1}{x_n}, \frac{x_2}{x_n}, \dots, \frac{x_{n-1}}{x_n}$ is ancillary statistics.

Note: Ancillary statistics.

Ancillary statistics in some cases gives information for inference on θ.

Definition: Complete Statistics. C & B. 6.2.21.

let $f(t|\theta)$ be a family of PDF or PMF for a statistics $T(x)$. The family of probability distribution is called complete if $E_\theta(g(T)) = 0$ for all θ implies $P_\theta(g(T)=0) = 1$ for all θ. Equivalently, $T(x)$ is called complete statistics.

Example. Complete statistics.

$\{x_i\} \sim N(98.6, 0.2)$ represent Human body Tempreture

let $g(x) = x - 98.6$. $E[g(x)] = E[x - 98.6] = 0$.

$T(\underset{\sim}{x}) = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$, if we don't assume that μ is known.

we have $T(\underset{\sim}{x}) = \bar{x}$ as a complete statistics for $\mu \in \mathbb{R}$.

Note: Origin of the name of complete statistics.

A set of vector is complete if they span the whole space.

i.e. $v \in$ vector space. Then $\exists$ coeffecients $a_1 \dots a_p$.

s.t. $v = a_1 v_1 + \dots a_p v_p$.

if $u \in$ vector space & $u \perp v_j$. $j = 1, 2 \dots p$. $\Rightarrow u = 0$.

Now consider a discrete example of complete statistic T.

$$E_\theta[g(T)] = 0 = \sum_{j=1}^{p} g(t_j) P_\theta(t_j) = 0 \Rightarrow g(\cdot) = 0,$$

$$= \left(g(t_1) \dots g(t_p)\right)\begin{pmatrix} P_\theta(t_1) \\ \vdots \\ P_\theta(t_p)\end{pmatrix} = 0 \Rightarrow \text{orthogonal}.$$

Thus $P_\theta(t_j) = \begin{pmatrix} P_\theta(t_1) \\ \vdots \\ P_\theta(t_p) \end{pmatrix}$ is complete in vector-space context.

Similarly, when T is continuous,

$$E[g(T)] = \int_{-\infty}^{\infty} g(T) f(T|\theta) \cdot dt = 0.$$

= inner-product on function space, is complete.

Example: Complete statistic for poisson distribution

$\{x_i\} \sim$ poisson (λ). $f(x|\lambda) = \frac{e^{-\lambda} \lambda^x}{x!}$

$T(x) = \sum_{i=1}^{n} x_i$ is a sufficient statistics. $E[T(x)] = n\lambda$.

Suppose there exists a measurable function g s.t. $E_\lambda[g(T(x))] = 0$ for all λ.

$T(x) = S \sim$ poisson $(n\lambda)$.

$$E[g(s)] = \sum_{S=0}^{\infty} g(s) P(S=s)$$

$$= \sum_{S=0}^{\infty} g(s) \cdot e^{-n\lambda} \cdot \frac{(n\lambda)^s}{s!}$$

$$= e^{-n\lambda} \cdot \sum_{S=0}^{\infty} \frac{g(s) n^s}{S!} \lambda^s = 0,$$

if $\sum_{S=0}^{\infty} \frac{g(s) n^s}{S!} \lambda^s = 0, \Rightarrow \frac{g(s) n^s}{s!} = 0 \Rightarrow g(s) = 0$.

since $S = 0, 1, 2 \dots \Rightarrow g(\cdot) = 0$.

Thus S is a complete statistics.

We will see using complete, $\frac{1}{n}S$ is the unbiased statistics for λ with smallest variance

Theorem. Basu's theorem. C & B. 6.2.24

If $T(x)$ is a complete & minimal sufficient statistics, then $T(x)$ is independent of every ancillary statistics.

Proof: discrete case.

let $S(x)$ be ancillary statistic. Then $P(S(x)=s)$ does not depends on θ since $S(x)$ is ancillary. Also, The conditional probability,

$P(S(x)=s \mid T(x)=t) = P(X \in \{x: S(x)=s\} \mid T(x)=t)$.

does not depends on θ. Because $T(x)$ is a sufficient statistics.

Thus to show that $S(x)$ & $T(x)$ are independent, it's sufficient to show that, $P(S(x)=s \mid T(x)=t) = P(S(x)=s)$ for all possible values $t \in \mathcal{T}$. Now,

$$P(S(x)=s) = \sum_{t \in \mathcal{T}} P(S(x)=s \mid T(x)=t) \cdot P_\theta(T(x)=t).$$

Furthermore, since $\sum_{t \in \mathcal{T}} P_\theta(T(x)=t) = 1$. We can write

$$P(S(x)=s) = \sum_{t \in \mathcal{T}} P(S(x)=s) P_\theta(T(x)=t).$$

Therefore, if we define the statistics $g(t) = P(S(x)=s \mid T(x)=t) - P(S(x)=s)$ the above two equations show that

$$E_\theta(g(T)) = \sum_{t \in \mathcal{T}} g(t) P_\theta(T(x)=t) = 0. \text{ for all } \theta$$

Since $T(x)$ is a complete statistic, this implies that $g(t)=0$ for all possible values $t \in \mathcal{T}$. Hence verified.

Lemma: Distribution of sufficient statistics in exponential family.

$\{x_i\}$ belongs to exponential family.

$$f(x|\theta) = h(x)\, c(\theta) \exp\left\{ \sum_{j=1}^{k} w(\theta_j)\, t_j(x_i) \right\}.$$

Then $T(x) = \left\{ \sum_{j=1}^{k} t_1(x_j) \ldots \sum_{j=1}^{k} t_k(x_j) \right\}$ has a distribution also belongs to exponential family.

Theorem: Find complete statistic in the exponential family.

Let $x_1 \ldots x_n$ be iid observations from an exponential family with PDF or PMF of the form $f(x|\theta) = h(x)\, c(\theta) \exp\left\{ \sum_{j=1}^{k} w(\theta_j)\, t_j(x) \right\}$

Where $\theta = (\theta_1, \theta_2 \ldots \theta_k)$. Then the statistic

$T(x) = \left(\sum_{j=1}^{n} t_1(x_j) \ldots \sum_{j=1}^{n} t_k(x_j) \right)$ is complete as long as the parameter space Θ contains a open set in $\mathbb{R}^k$.

Example: Poission distribution.

We saw $S = \sum_{i=1}^{n} X_i$ ~ complete & sufficient statistics.

$f(x) = \frac{e^{-\lambda} \cdot \lambda^{x}}{x!} = \left(\frac{1}{x!}\right) \cdot \exp\{x \log \lambda - \lambda\}$.

$T(x) = \sum_{i=1}^{n} X_i$ is complete & sufficient statistic by theorem 6.2.25.

Theorem (Minimal sufficient & complete statistics. C&B 6.2.28.)

If a minimal sufficient statistics exists, then any complete statistics is also a minimal sufficient statistics.

Definition: Likelihood Principle.

Let $f(\underset{\sim}{x} | \theta)$ denotes the joint pdf or pmf of sample, $X = (X_1, \dots, X_n)$.

Then, given that $X = x$ is observed, the function of θ, defined by

$L(\theta | x) = f(x | \theta)$ is called Likelihood function.

If X is a discrete random vector, then $L(\theta | x) = P_\theta(X = x)$. if we compare the likelihood function at two parameter points and find that

$P_{\theta_1}(X = x) = L(\theta_1 | x) > L(\theta_2 | x) = P_{\theta_2}(X = x)$

Then, the sample we actually observed is more likely to occure if $\theta = \theta_1$.