## Lecture 7, Review

## Data Reduction

We should think about the following questions carefully before the "simplification" process:

• Is there any loss of information due to summarization?

• How to compare the amount of information about $\theta$ in the original data **X** and in $T(\boldsymbol{X})$?

• Is it sufficient to consider only the "reduced data" $T$?

## 1. Sufficient Statistics

A statistic T is called sufficient if the conditional distribution of **X** given $T$ is free of $\theta$ (that is, the conditional is a completely known distribution).

**Example.** Toss a coin n times, and the probability of head is an unknown parameter θ. Let $T$ = the total number of heads. Is T sufficient for $\theta$?

### Sufficiency Principle

If T is sufficient, the "extra information" carried by **X** is worthless as long as $\theta$ is concerned. It is then only natural to consider inference procedures which do not use this extra irrelevant information. This leads to the Sufficiency Principle :

*Any inference procedure should depend on the data only through sufficient statistics.*

**Definition: Sufficient Statistic (in terms of Conditional Probability)**

(discrete case):

For any **x** and t, if the conditional pdf of X given T:

$$P_\theta(\mathbf{X}=\mathbf{x}|T(\mathbf{X})=t)=\frac{P_\theta(\mathbf{X}=\mathbf{x},T(\mathbf{X})=t)}{P_\theta(T(\mathbf{X})=t)}=\frac{P_\theta(\mathbf{X}=\mathbf{x})}{P_\theta(T(\mathbf{X})=t)}$$

does not depend on θ then we say $T(\mathbf{X})$ is a sufficient statistic for θ.

## Sufficient Statistic, the general definition (for both discrete and continuous variables):

Let the pdf of data **X** is $f(\mathbf{x};\theta)$ and the pdf of $T$ be $q(t;\ \theta)$. *If*

$f(\mathbf{x};\ \theta)/q(T\ (\mathbf{x});\ \theta)$ is free of $\theta$, (may depend on **x**) $(*)$

for all **x** and $\theta$, then $T$ is a sufficient statistic for $\theta$.

**Example:** Toss a coin n times, and the probability of head is an unknown parameter θ. Let $T$ = the total number of heads. Is T sufficient for $\theta$?

$X_i\ i.i.d.$ Bernoulli: $f(x)=\theta^x(1-\theta)^{1-x}, x=0,1$

$$f(\boldsymbol{x};\ \theta)=f(x_1,\dots,x_n)=\theta^{\sum_i x_i}(1-\theta)^{n-\sum_i x_i}$$

$T=\sum_i X_i\ \sim$ B(n,θ):

$$q(t;\ \theta)=\ q(\textstyle\sum_i x_i)=\begin{bmatrix} n \\ \sum_i x_i\end{bmatrix}\theta^{\sum_i x_i}(1-\theta)^{n-\sum_i x_i}$$

Thus

$$\frac{f(\mathbf{x};\ \theta)}{q(T\ (\mathbf{x});\ \theta)}=1/\begin{bmatrix} n \\ t\end{bmatrix}$$

is free of $\theta$,

So by the definition, $\sum_i X_i$ is a sufficient statistic for $\theta$.

Example. $X_1, \dots, X_n$ iid $N\,(\theta, 1). T = \bar{X}$.

Remarks: The definition (*) is not always easy to apply.

- Need to guess the form of a sufficient statistic.
- Need to figure out the distribution of $T$.

## How to find a sufficient statistic?

### 2. (Neyman-Fisher) Factorization theorem.

T is sufficient if and only if $f(\mathbf{x};\ \theta)$ can be written as the product $g(T(\mathbf{x});\ \theta)h(\mathbf{x})$, where the first factor depends on x only though $T(\mathbf{x})$ and the second factor is free of $\theta$.

**Example.** Binomial. iid bin(1, $\theta$)

**Solution 1:**

Bernoulli: $f(x) = \theta^x(1-\theta)^{1-x}, x = 0, 1$

$$\begin{aligned} f(\mathbf{x};\ \theta) = f(x_1, \dots, x_n) &= \theta^{\sum_i x_i}(1-\theta)^{n-\sum_i x_i} \\ &= \left[\theta^{\sum_i x_i}(1-\theta)^{n-\sum_i x_i}\right] \cdot [1] \\ &= g\left(\sum_i x_i, \theta\right) \cdot h(x_1, \dots, x_n) \end{aligned}$$

So according to the factorization theorem, $T = \sum_i X_i$ is a sufficient statistic for $\theta$.

**Solution 2:**

$$f(\mathbf{x};\ \theta) = f(x_1, x_2, \dots, x_n|\theta)$$
$$= \begin{cases} \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i}, & if\ x_i = 0,1, i = 1,2, \dots, n \\ 0, & otherwise \end{cases}$$

$$= \begin{cases} \theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i}, & if\ x_i = 0,1, i = 1,2, \dots, n \\ 0, & otherwise \end{cases}$$

$$= \theta^t(1-\theta)^{n-t}h(x_1, x_2, \dots, x_n)$$

$$= g(t,\theta)h(x_1, \dots, x_n),$$

where $g(t,\theta) = \theta^t(1-\theta)^{n-t}$ and

$$h(x_1, \dots, x_n) = \begin{cases} 1, & if\ x_i = 0,1, i = 1, 2, \dots, n \\ 0, & otherwise. \end{cases}$$

Hence $T$ is a sufficient statistic for $\theta$.

**Example. Exp($\lambda$).**

Let $X_1, \dots, X_n$ be a random sample from an exponential distribution with rate $\lambda$. And Let $T = X_1 + X_2 + \cdots + X_n$ and $f$ be the joint density of $X_1, X_2, \dots, X_n$.

$$f(\mathbf{x};\ \lambda) = f(x_1, x_2, \dots, x_n|\lambda)$$
$$= \begin{cases} \prod_{i=1}^{n} \lambda e^{-\lambda x_i}, & if\ x_i > 0, i = 1,2, \dots, n \\ 0, & otherwise \end{cases}$$

$$= \begin{cases} \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}, & if\ x_i > 0, i = 1,2, \dots, n \\ 0, & otherwise \end{cases}$$

$$= \lambda^n e^{-\lambda t} h(x_1, \dots, x_n)$$

$$= g(t,\lambda)h(x_1, \dots, x_n)$$

where $g(t,\lambda) = \lambda^n e^{-\lambda t}$, and

$$h(x_1, \dots, x_n) = \begin{cases} 1, & if\ x_i > 0, i = 1, 2, \dots, n \\ 0, & otherwise. \end{cases}$$

Hence $T$ is a sufficient statistic for $\lambda$.

**Example. Normal. iid N($\theta$,1).**

**Please derive the sufficient statistic for $\theta$ *by yourself.***

**When the range of $X$ depends on $\theta$, should be more careful about factorization. Must use indicator functions explicitly.**

**Example. Uniform. iid $U(0, \theta)$.**

**Solution 1:**

Let $X_1, \dots, X_n$ be a random sample from an uniform distribution on $(0, \theta)$. And Let $T = X_{(n)}$ and $f$ be the joint density of $X_1, X_2, \dots, X_n$.

Then

$$
\begin{aligned}
f(\mathbf{x};\ \theta) &= f(x_1, x_2, \dots, x_n|\theta) \\
&= \begin{cases} \prod_{i=1}^{n} \frac{1}{\theta}, & if\ \theta > x_i > 0, i = 1,2, \dots, n \\ 0, & otherwise \end{cases} \\
&= \begin{cases} \frac{1}{\theta^n}, & if\ \theta > x_i > 0, i = 1,2, \dots, n \\ 0, & otherwise \end{cases} \\
&= \begin{cases} \frac{1}{\theta^n}, & if\ \theta > x_{(n)} \geq \cdots \geq x_{(1)} > 0 \\ 0, & otherwise \end{cases} \\
&= g(t, \theta) h(x_1, \dots, x_n)
\end{aligned}
$$

where

$$
g(t, \theta) = \begin{cases} \frac{1}{\theta^n}, & if\ \theta > t > 0 \\ 0, & otherwise \end{cases},
$$

and

$$
h(x_1, \dots, x_n) = \begin{cases} 1, & if\ x_{(1)} > 0 \\ 0, & otherwise. \end{cases}
$$

Hence $T$ is a sufficient statistic for $\theta$.

*** I personally prefer this approach because it is most straightforward. Alternatively, one can use the indicator function and simplify the solution as illustrated next.

**Definition: Indicator function**

$$I_A(x) = \begin{cases} 1, & if\ x \in A \\ 0, & if\ x \notin A \end{cases}$$

**Solution 2 (in terms of the indicator function):**

Uniform: $f(x) = \frac{1}{\theta}, x \in (0, \theta)$

$$\begin{aligned} f(\mathbf{x};\ \theta) &= f(x_1, \dots, x_n) = \left(\frac{1}{\theta}\right)^n, x_i \in (0, \theta), \forall i \\ &= \left(\frac{1}{\theta}\right)^n \prod_i I_{(0,\theta)}(x_i) \\ &= \left(\frac{1}{\theta}\right)^n I_{(0,+\infty)}\big(x_{(1)}\big) \cdot I_{(0,\theta)}\big(x_{(n)}\big) \\ &= [\left(\frac{1}{\theta}\right)^n I_{(0,\theta)}\big(x_{(n)}\big)] \cdot [I_{(0,+\infty)}\big(x_{(1)}\big)] \\ &= g\big(x_{(n)}, \theta\big) \cdot h(x_1, \dots, x_n) \end{aligned}$$

So by factorization theorem, $x_{(n)}$ is a sufficient statistic for $\theta$.

**Example:** Please derive the sufficient statistics for θ, when given a random sample of size n from $U\ (\theta, \theta\ +\ 1)$.

**Solution:**

1. **Indicator function approach:**

Uniform: $f(x) = 1, x \in (\theta, \theta + 1)$

$$\begin{aligned} f(x_1, \dots, x_n|\theta) &= (1)^n, x_i \in (\theta, \theta + 1), \forall i \\ &= (1)^n \prod_i I_{(\theta,\theta+1)}(x_i) \\ &= I_{(\theta,+\infty)}\big(x_{(1)}\big) \cdot I_{(-\infty,\theta+1)}\big(x_{(n)}\big) \\ &= [I_{(\theta,+\infty)}\big(x_{(1)}\big) \cdot I_{(-\infty,\theta+1)}\big(x_{(n)}\big)] \cdot [1] \\ &= g\big(x_{(1)}, x_{(n)}, \theta\big) \cdot h(x_1, \dots, x_n) \end{aligned}$$

So, $T = \big(X_{(1)}, X_{(n)}\big)$ is a SS for $\theta$.

2. **Do not use the indicator function:**

$$f(x_1, x_2, \dots, x_n|\theta)$$
$$= \begin{cases} \prod_{i=1}^{n} 1, & if\ \theta+1 > x_i > \theta, i = 1,2,\dots,n \\ 0, & otherwise \end{cases}$$
$$= \begin{cases} 1, & if\ \theta+1 > x_i > \theta, i = 1,2,\dots,n \\ 0, & otherwise \end{cases}$$
$$= \begin{cases} 1, & if\ \theta+1 > x_{(n)} \geq \cdots \geq x_{(1)} > \theta \\ 0, & otherwise \end{cases}$$
$$= g(x_{(1)}, x_{(n)}, \theta) h(x_1, \dots, x_n)$$

where

$$g\left(x_{(1)}, x_{(n)}, \theta\right) = \begin{cases} 1, & if\ \theta+1 > x_{(n)}\ and\ x_{(1)} > \theta \\ 0, & otherwise \end{cases},$$

and

$$h(x_1, \dots, x_n) = 1$$

So $T = \left(X_{(1)}, X_{(n)}\right)$ is a SS for $\theta$.

## Two-dimensional Examples.

**Example. Normal. iid $N(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$. $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$ (both unknown).**

Let $X_1, \dots, X_n$ be a random sample from a normal distribution $N(\mu, \sigma^2)$. And Let

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n} (X_i - \bar{X})^2,$$

and let $f$ be the joint density of $X_1, X_2, \dots, X_n$.

$$f(\mathbf{x};\ \boldsymbol{\theta}) = f(x_1, x_2, \dots, x_n|\mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2}$$

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

Now

$$\sum_{i=1}^{n}(x_i-\mu)^2 = \sum_{i=1}^{n}(x_i - \bar{x} + \bar{x} - \mu)^2$$

$$= \sum_{i=1}^{n}(x_i-\bar{x})^2 + 2\sum_{i=1}^{n}(x_i-\bar{x})(\bar{x}-\mu) + \sum_{i=1}^{n}(\bar{x}-\mu)^2$$

$$= (\mathrm{n}-1)\mathrm{s}^2 + 2(\bar{x}-\mu)\sum_{i=1}^{n}(x_i-\bar{x}) + n(\bar{x}-\mu)^2$$

$$= (\mathrm{n}-1)\mathrm{s}^2 + n(\bar{x}-\mu)^2.$$

Thus,

$$f(\mathbf{x};\ \boldsymbol{\theta}) = f(x_1, x_2, \ldots, x_n \mid \mu, \sigma^2)$$
$$= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp(-\frac{1}{2\sigma^2}((n-1)s^2 + n(\bar{x}-\mu)^2))$$

$$= g(\bar{x}, s^2, \mu, \sigma^2)h(x_1, \ldots, x_n),$$

where

$g(\bar{x}, s^2, \mu, \sigma^2)$

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

and

$h(x_1, \ldots, x_n) = 1.$

In this case we say $(\bar{X}, S^2)$ is sufficient for $(\mu, \sigma^2)$.

## 3. (Regular) Exponential Family

The density function of a regular exponential family is:

$$f\ (x;\ \boldsymbol{\theta}) = c(\boldsymbol{\theta})h(x)\exp\left[\sum_{j=1}^{k} w_j(\boldsymbol{\theta})t_j(x)\right],\ \ \boldsymbol{\theta}\ =\ (\theta_1, \dots, \theta_k)$$

**Example. Poisson(θ)**

$$f\ (x;\ \theta) = \exp(-\theta)\frac{1}{x!}\exp[\ln(\theta) * x]$$

**Example. Normal.** $\boldsymbol{N(\mu, \sigma^2)}.\ \boldsymbol{\theta}\ =\ (\mu, \sigma^2)$ **(both unknown).**

$$fx;\ \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

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

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

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

## 4. Theorem (Exponential family & sufficient Statistic).

Let $X_1, \dots, X_n$ be a random sample from the regular exponential family.

Then

$$\boldsymbol{T}(\boldsymbol{X}) = \left(\sum_{i=1}^{n} t_1(X_i), \dots, \sum_{i=1}^{n} t_k(X_i)\right)$$

is sufficient for $\boldsymbol{\theta}\ =\ (\theta_1, \dots, \theta_k)$.

**Example. Poisson(θ)**

Let $X_1, \dots, X_n$ be a random sample from **Poisson(θ)**

Then

$$T(\boldsymbol{X}) = \sum_{i=1}^{n} X_i$$

is sufficient for $\theta$ .

**Example. Normal. $\boldsymbol{N(\mu, \sigma^2). \theta} = (\mu, \sigma^2)$ (both unknown).**

Let $X_1, \dots, X_n$ be a random sample from $\boldsymbol{N}(\mu, \sigma^2)$

Then

$$\boldsymbol{T}(\boldsymbol{X}) = (\sum_{i=1}^{n} X_i , \sum_{i=1}^{n} X_i^2)$$

is sufficient for $\boldsymbol{\theta} = (\mu, \sigma^2)$.

**Exercise.**

Apply the general exponential family result to all the standard families discussed above such as binomial, Poisson, normal, exponential, gamma.

**A Non-Exponential Family Example.**

Discrete uniform.

$P\ (X = x) = 1/\theta,\ \ x = 1, \dots, \theta,\ \ \theta$ is a positive integer.

**Another Non-exponential Example.**

$X_1, \dots, X_n$ iid $U\ (0, \theta), T = X_{(n)}$.

## Universal Cases.

$X_1, \dots, X_n$are iid with density $f$ .

- The original data $X_1, \dots, X_n$ are always sufficient for $\theta$.

(They are trivial statistics, since they do not lead any data reduction)

- Order statistics $T \;=\; \left(X_{(1)}, \dots, X_{(n)}\right)$ are always sufficient for $\theta$.

( The dimension of order statistics is $n$, the same as the dimension of the data. Still this is a nontrivial reduction as $n!$ different values of data corresponds to one value of $T$ . )

## 5. Theorem (Rao-Blackwell)

Let $X_1, \dots, X_n$ be a random sample from the population with pdf $f\ (\boldsymbol{x};\ \boldsymbol{\theta})$. Let $\boldsymbol{T}(\boldsymbol{X})$ be a sufficient statistic for θ, and $\boldsymbol{U}(\boldsymbol{X})$ be any unbiased estimator of **θ.**

Let $\boldsymbol{U}^*(\boldsymbol{X}) = E[\boldsymbol{U}(\boldsymbol{X})|\boldsymbol{T}]$, then

**(1)** $\boldsymbol{U}^*(\boldsymbol{X})$ is an unbiased estimator of $\boldsymbol{\theta}$,

**(2)** $\boldsymbol{U}^*(\boldsymbol{X})$ is a function of **T,**

**(3)** $\boldsymbol{Var}(\boldsymbol{U}^*) \le \boldsymbol{Var}(\boldsymbol{U})$ for every $\boldsymbol{\theta}$**,** and $\boldsymbol{Var}(\boldsymbol{U}^*) < Var(U)$ for some $\boldsymbol{\theta}$ unless $\boldsymbol{U}^* = \boldsymbol{U}$ with probability 1 **.**

**Rao-Blackwell theorem tells us that in searching for an unbiased estimator with the smallest possible variance (i.e., the best estimator, also called the uniformly minimum variance unbiased estimator – UMVUE, which is also referred to as simply the MVUE), we can restrict our search to only unbiased functions of the sufficient statistic T(X).**

## 6. Transformation of Sufficient Statistics

1. If $T$ is sufficient for $\theta$ and $T = c(U)$, a mathematical function of some other statistic, then $U$ is also sufficient.

2. If $T$ is sufficient for $\theta$, and $U = G(T)$ with $G$ being one-to-one, then $U$ is also sufficient.

Remark: When one statistic is a function of the other statistic and vice versa, then they carry exactly the same amount of information.

Examples:

- If $\sum_{i=1}^{n} X_i$ is sufficient, so is $\bar{X}$.

- If $(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2)$ are sufficient, so is $(\bar{X}, S^2)$.

- If $\sum_{i=1}^{n} X_i$ is sufficient, so is $(\sum_{i=1}^{m} X_i, \sum_{i=m+1}^{n} X_i)$ is sufficient, and so is $\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2)$.

**Examples of non-sufficiency.**

Ex. $X_1, X_2$ iid Poisson$(\lambda)$. $T = X_1 - X_2$ is not sufficient.

Ex. $X_1, \ldots, X_n$ iid pmf $f(x; \theta)$. $T = (X_1, \ldots, X_{n-1})$ is not sufficient.

## 7. Minimal Sufficient Statistics

It is seen that different sufficient statistics are possible. Which one is the "best"? Naturally, the one with the maximum reduction.

• For $N(\theta, 1)$, $\bar{X}$ is a better sufficient statistic for $\theta$ than $(\bar{X}, S^2)$.

**Definition:**

$T$ is a minimal sufficient statistic if, given any other sufficient statistic $T'$, there is a function $c(\bullet)$ such that $T = c(T')$.

Equivalently, $T$ is minimal sufficient if, given any other sufficient statistic $T'$, whenever $\mathbf{x}$ and $\mathbf{y}$ are two data values such that $T'(\mathbf{x}) = T'(\mathbf{y})$, then $T(\mathbf{x}) = T(\mathbf{y})$.

**Partition Interpretation for Minimal Sufficient Statistics:**

• Any sufficient statistic introduces a partition on the sample space.

• The partition of a minimal sufficient statistic is the coarsest.

**• Minimal sufficient statistic has the smallest dimension among possible sufficient statistics. Often the dimension is equal to the number of free parameters (exceptions do exist).**

### Theorem (How to check minimal sufficiency).

A statistic T is minimal sufficient if the following property holds: For any two sample points $\mathbf{x}$ and $\mathbf{y}$, $f(\boldsymbol{x};\ \theta)/f(\boldsymbol{y};\ \theta)$ does not depend on $\theta$ (i.e. $f(\boldsymbol{x};\ \theta)/f(\boldsymbol{y};\ \theta)$ is a constant function of $\theta$) if and only if $T(\boldsymbol{x}) = T(\boldsymbol{y})$.

## 8. Exponential Families & Minimal Sufficient Statistic:

For a random sample from the regular exponential family with probability density $f\ (\mathbf{x};\ \boldsymbol{\theta})\ =\ c(\boldsymbol{\theta})h(\mathbf{x})\ \exp[\sum_{j=1}^{k} w_j(\boldsymbol{\theta})t_j(\mathbf{x})]$, where $\boldsymbol{\theta}$ is $k$ dimensional, the statistic

$$T(\mathbf{X}) = (\sum_{i=1}^{n} t_1(\mathbf{X}_i)\,, \dots, \sum_{i=1}^{n} t_k(\mathbf{X}_i))$$

is minimal sufficient for $\boldsymbol{\theta}$.

**Example. Poisson(θ)**

Let $X_1, \dots, X_n$ be a random sample from **Poisson(θ)**

Then

$$T(X) = \sum_{i=1}^{n} X_i$$

is minimal sufficient for $\theta$ .

**Example. Normal. $\boldsymbol{N(\mu, \sigma^2)}. \boldsymbol{\theta}\ =\ (\mu, \sigma^2)$ (both unknown).**

Let $X_1, \dots, X_n$ be a random sample from $\boldsymbol{N}(\mu, \sigma^2)$

Then

$$\boldsymbol{T}(\boldsymbol{X}) = (\sum_{i=1}^{n} X_i\,, \sum_{i=1}^{n} X_i^2)$$

Is minimal sufficient for $\boldsymbol{\theta}\ =\ (\mu, \sigma^2)$.

**Remarks:**

• Minimal sufficient statistic is not unique. Any two are in one-to-one correspondence, so are equivalent.

## 9. Complete Statistics

Let a parametric family $\{f(x,\theta), \theta \in \Theta\}$ be given. Let $T$ be a statistic. Induced family of distributions $f_T(t,\theta), \theta \in \Theta$.

A statistic $T$ is complete for the family $\{f(x,\theta), \theta \in \Theta\}$, or equivalently, the induced family $f_T(t,\theta), \theta \in \Theta$ is called complete, if $E_\theta(g(T)) = 0$ for all $\theta \in \Theta$ implies that $g(T) = 0$ with probability 1.

**Example. Poisson(θ)**

Let $X_1, \ldots, X_n$ be a random sample from **Poisson(θ)**

Then

$$T(X) = \sum_{i=1}^{n} X_i$$

is minimal sufficient for $\theta$ . Now we show that T is also complete.

We know that $T(X) = \sum_{i=1}^{n} X_i \sim Poisson(\lambda = n\theta)$

Consider any function $u(T)$. We have

$$E[u(T)] = \sum_{t=0}^{\infty} \frac{u(t)e^{-\lambda}\lambda^t}{t!}$$

Because $e^{-\lambda} \neq 0$, setting $E[u(T)] = 0$ requires all the coefficient

$$\frac{u(t)}{t!}$$

to be zero, which implies $u(T) = 0$.

**Example.** Let $X_1, \ldots, X_n$ be iid from $Bin(1,\theta)$. Show $T = \sum_{i=1}^{n} X_i$ is a complete statistic. **(*Please read our text book for more examples – but the following result on the regular exponential family is the most important.)**

## 10. Exponential Families & Complete Statistics

**Theorem.** Let $X_1, \ldots, X_n$ be iid observations from the regular exponential family, with the pdf
$f\ (\mathbf{x};\ \theta) = c(\theta)h(\mathbf{x}) \exp[\sum_{j=1}^{k} w_j(\theta)t_j(\mathbf{x})]$, and $\theta = (\theta 1, \ldots, \theta_k)$. Then

$$T(\mathbf{X}) = (\sum_{i=1}^{n} t_1(\mathbf{X}_i), \ldots, \sum_{i=1}^{n} t_k(\mathbf{X}_i))$$

is complete if the parameter space $\{(w_1(\theta), \ldots, w_k\ (\theta)) : \theta \in \Theta\}$ contains an open set in $R^k$.

(*This is only a sufficient condition, not a necessary condition*)

Example. $X_1, \ldots, X_n \sim N\ (\theta, 1), -\infty < \theta < \infty$.

**Example. Poisson(θ);** $0 < \theta < \infty$.

$$f\ (x;\ \theta) = \exp(-\theta)\frac{1}{x!}\exp[\ln(\theta) * x]$$

**Example. Normal.** $\boldsymbol{N(\mu, \sigma^2)}$. $\boldsymbol{\theta} = (\mu, \sigma^2)$ **(both unknown).**
$-\infty < \mu < \infty,\ 0 < \sigma^2 < \infty$.

$$fx;\ \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

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

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

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

Example. $X_1, \ldots, X_n \sim N\ (\theta, 1), \theta = 1, 2$, is not complete.

Example. $X_1, \ldots, X_n \sim N\ (\theta, 1), -\infty < \theta < \infty$, is complete.

Example. {Bin(2,$p$), $p$ = 1/2, $p$ = 1/4} is not complete.

Example. The family {Bin(2,p), 0 < p < 1} is complete.

## Properties of the Complete Statistics

(i) If $T$ is complete and $S = \psi(T)$, then $S$ is also complete.

(ii) If a statistic T is complete and sufficient, then any minimal sufficient statistic is complete.

(iii) Trivial (constant) statistics are complete for any family.

## 11. Theorem (Lehmann-Scheffe). (Complete Sufficient Statistic and the Best Estimator)

**If T is complete and sufficient, then $\boldsymbol{U = h(T)}$ is the Best Estimator (also called UMVUE or MVUE) of its expectation.**

**Example. Poisson(θ)**

Let $X_1, \dots, X_n$ be a random sample from **Poisson(θ)**

Then

$$T(X) = \sum_{i=1}^{n} X_i$$

is complete sufficient for $\theta$ . Since

$$U = \frac{T(X)}{n} = \frac{\sum_{i=1}^{n} X_i}{n}$$

is an unbiased estimator of θ – by the **Lehmann-Scheffe theorem we know that U is a best estimator (UMVUE/MVUE) for θ.**

## 12. Theorem (Basu)

A complete sufficient statistic **T** for the parameter **θ** is independent of any **ancillary statistic** – that is, a statistic whose distribution does not depend on **θ**

**Example.** Consider a random sample of size n from a normal distribution $\boldsymbol{N(\mu, \sigma^2)}$. $\boldsymbol{\theta} = (\mu, \sigma^2)$**.**

**Consider the MLEs** $\boldsymbol{for\ \mu, \sigma^2} \Rightarrow \begin{cases} \hat{\mu} = \bar{X} \\ \hat{\sigma}^2 = \dfrac{\sum (X_i - \bar{X})^2}{n} \end{cases}$

It is easy to verify that $\bar{X}$ is a complete sufficient statistic for $\mu$, for fixed values of $\sigma^2$. Also:

$$\frac{\boldsymbol{n\hat{\sigma}^2}}{\boldsymbol{\sigma^2}} \sim \boldsymbol{\chi^2(n-1)}$$

which does not depend on $\mu$. It follows from the Basu Theorem that the two MLEs are independent to each other.