STAT 346/446 Lecture 6

Principles of Data Reduction

CB Sections 6.1, 6.2, and DS Section 7.7

1. Sufficient Statistic
2. Minimal Sufficient Statistic
3. Ancillary Statistic
4. Complete Statistic

Finding the "best" point estimator

- We have seen that the MLE of σ^2 in $N(\mu, \sigma^2)$ has a smaller MSE than the sample variance S^2.
- Could we find an estimator of σ^2 that has the smallest possible MSE?
- Hard to answer in general, but if we restrict the space of estimators to e.g.
  - all unbiased estimators, or
  - all linear estimators, or
  - all linear and unbiased estimators

  we can sometimes find the estimator with the smallest possible MSE in that space.
- UMVUE = Minimum Variance
- BLUE = Best Linear Unbiased Estimator

Data reduction

- Want to use a sample $X_1, X_2, \ldots, X_n$ to infer about an unknown parameter θ
  - In practice: have data points $x_1, x_2, \ldots, x_n$
- A statistic $T(X_1, X_2, \ldots, X_n)$ is a method of summarizing the sample data
  - In practice: $T(x_1, x_2, \ldots, x_n)$
- Is there a statistic $T(\cdot)$ (or statistics $T_1(\cdot), \ldots, T_k(\cdot)$) that gives the same amount of information about θ as the sample $X_1, X_2, \ldots, X_n$ does?
  - Then we could store observed statistics only, instead of the whole dataset, i.e. get *data reduction*.
- Main use for STAT 346: Add tools to find UMVUEs (Section 7.3.3)

Statistic as a partition of sample space

- A statistic $T(X_1, X_2, \ldots, X_n)$ can be thought of as a *partition* of the sample space $\mathcal{X}$ of all possible outcomes for $\mathbf{X} = (X_1, X_2, \ldots, X_n)$

  $\uparrow$ $\underline{x} \in \mathcal{X}$ e.g. $\mathcal{X} = \mathbb{R}^n$

- Let

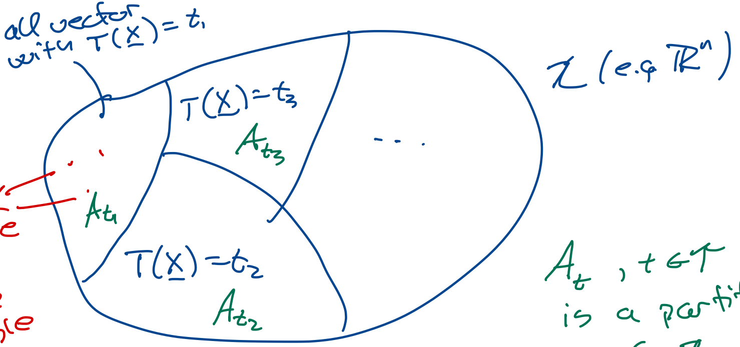
$$\mathcal{T} = \{t : t = T(x_1, x_2, \ldots, x_n) \text{ for some } \mathbf{x} \in \mathcal{X}\}$$

  - $\mathcal{T}$ contains all possible outcomes of $T(X_1, X_2, \ldots, X_n)$
  - Often has lower dimension than $\mathcal{X}$
    - E.g. if $T(X_1, X_2, \ldots, X_n) = \overline{X}$ then $\mathcal{X} = \mathbb{R}^n$ and $\mathcal{T} = \mathbb{R}$

- Partition of $\mathcal{X}$:

$$A_t = \{\mathbf{x} : T(\mathbf{x}) = t\} \subset \mathcal{X} \qquad t \in \mathcal{T}$$

note: can have an uncountable index space



all vectors in here have same sample mean

A_t, $t \in \mathcal{T}$ is a partition of $\mathcal{X}$

e.g. if $T(\underline{X}) = \overline{X} \implies \mathcal{T} = \mathbb{R}$

$$\mathcal{T} = \{ t : T(\underline{x}) = t \text{ for some } \underline{x} \in \mathcal{X} \}$$

Sufficient data reduction

- A **sufficient statistic $T(\mathbf{X})$ for θ** (for a given distribution of $\mathbf{X}$) is a statistic that contains (in some way) *all* the information about θ in our sample $X_1, X_2, \ldots, X_n$

Sufficiency Principle

If $T(\mathbf{X})$ is a sufficient statistic for θ, then any inference about θ should depend on the sample $\mathbf{X}$ *only* through the value $T(\mathbf{X})$.

- I.e. if $\mathbf{x}_1$ and $\mathbf{x}_2$ are two sample outcomes such that $T(\mathbf{x}_1) = T(\mathbf{x}_2)$ then the inference about θ should be the same whether $\mathbf{X} = \mathbf{x}_1$ or $\mathbf{X} = \mathbf{x}_2$ was observed.

- Can have more than one sufficient statistic for the same parameter

Sufficient statistic

Def: Sufficient Statistic

A statistic $T(\mathbf{X})$ is a **sufficient statistic for θ** if

$$f(\mathbf{x} \mid T(\mathbf{X}) = t) \qquad \text{does not depend on } \theta$$

i.e. the conditional distribution of the sample $\mathbf{X}$ given the value of $T(\mathbf{X})$ does not depend on θ.

- Conditional probability = change of sample space
  - $f(\mathbf{x} \mid T(\mathbf{X}) = t)$ has support inside A_t
- A sufficient statistic does not have to be one dimensional

Making sense of the definition of a Sufficient Statistic

Example: X_1, X_2, X_3 i.i.d. $\text{Bernoulli}(\theta)$ and $T(\mathbf{X}) = X_1 + X_2 + X_3$

- The joint pdf:

$$f(\mathbf{x}) = \overbrace{\theta^{x_1}(1-\theta)^{1-x_1}}^{f(x_1)} \overbrace{\theta^{x_2}(1-\theta)^{1-x_2}}^{f(x_2)} \overbrace{\theta^{x_3}(1-\theta)^{1-x_3}}^{f(x_3)}$$

  for $x_1, x_2, x_3 \in \{0,1\} \times \{0,1\} \times \{0,1\}$ (x_1, x_2, x_3)

- What is the distribution of (X_1, X_2, X_3) *given* that $T(\mathbf{X}) = 2$?

i.e. what is $f(\underline{x} \mid T(X) = 2)$?

$$\underset{P(\underline{X} = \underline{x} \mid T(\underline{X}) = 2)}{f(\underline{x} \mid T(\underline{x}) = 2)} = \frac{P(\underline{X} = \underline{x}, T(\underline{X}) = 2)}{P(T(\underline{X}) = 2)}$$

Given that $T(\underline{x}) = 2$ we know that the new sample space is $A_2 = \{(1,1,0), (1,0,1), (0,1,1)\}$

Given that $T(\underline{x}) = 2$ we know that the new sample space is $A_2 = \{(1,1,0), (1,0,1), (0,1,1)\}$

$f(\underline{x})$	x_1	x_2	x_3	$T(\underline{x})$	
$(1-\theta)^3$	0	0	0	0	A_0
$\theta(1-\theta)^2$	0	0	1	1	
$\theta(1-\theta)^2$	0	1	0	1	A_1
$\theta(1-\theta)^2$	1	0	0	1	
$\theta^2(1-\theta)$	0	1	1	2	
$\theta^2(1-\theta)$	1	0	1	2	A_2
$\theta^2(1-\theta)$	1	1	0	2	
θ^3	1	1	1	3	A_3

$$f((0,1,1) \mid T(\underline{x}) = 2) = \frac{\theta^2(1-\theta)}{P(T(\underline{x}) = 2)}$$

= same for $(1,0,1)$ and $(1,1,0)$

$\Rightarrow$ The three outcomes are equally likely, no matter what θ is!

$\Rightarrow T(X) = X_1 + X_2 + X_3$ is a suff. stat for θ.

In general : The outcomes in A_t (the part of the sample space we are conditioning on) will not always be equally likely for a sufficient statistic

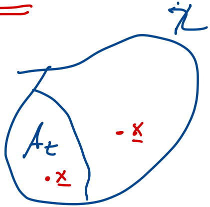
Point is : we can find the conditional probability of all outcomes in A_t without knowing the value of θ.

$\downarrow$

$f(\underline{x} \mid T(\underline{X}) = t)$ does not depend on θ.

Identifying a Sufficient statistic - Discrete case

- Let's take a closer look at $f(\mathbf{x} \mid T(\mathbf{X}) = t)$ for discrete random samples:



$$f(\mathbf{x} \mid T(\mathbf{X}) = t) = P(\underline{X} = \underline{x} \mid T(\underline{X}) = t)$$

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

$$= \frac{P(\underline{X} = \underline{x},\ \underline{X} \in A_t)}{P(T(\underline{X}) = t)}$$

$$= \begin{cases} \dfrac{P(\underline{X} = \underline{x})}{P(T(\underline{X}) = t)} = \dfrac{f(\underline{x})}{f_T(t)} & \text{if } \underline{x} \in A_t \\ 0 & \text{if } \underline{x} \notin A_t \end{cases}$$

Identifying a Sufficient statistic

Theorem: Identifying a sufficient statistic

- Let $p(\mathbf{x} \mid \theta)$ be the joint pmf or pdf of $\mathbf{X} = (X_1, \ldots, X_n)$ and
- Let $q(t \mid \theta)$ be the pmf or pdf of a statistic $T(\mathbf{X})$.

If for every $\mathbf{x} \in \mathcal{X}$ the ratio

i.e. does not depend on θ

$$\frac{p(\mathbf{x} \mid \theta)}{q(T(\mathbf{x}) \mid \theta)}$$

is a constant as a function of θ, then $T(\mathbf{X})$ is a *sufficient statistic for θ*

- Need to find the pmf/pdf of $T(\mathbf{X})$ (the sampling distribution) to use this result

Example: Poisson

- Let $X_1, X_2, \ldots, X_n$ be a random sample from $\mathrm{Poisson}(\lambda)$. Show that $T(\mathbf{X}) = \sum_{i=1}^n X_i$ is a sufficient statistic for λ.

$$p(\underline{x} \mid \lambda) = \prod_{i=1}^n f(x_i \mid \lambda) = \prod_{i=1}^n \frac{e^{-\lambda} \lambda^{x_i}}{x_i!} = e^{-n\lambda} \lambda^{\overbrace{\sum_{i=1}^n x_i}^{=t}} \prod_{i=1}^n \frac{1}{x_i!}$$

Also need the pmf of $T(\underline{X}) = X_1 + X_2 + \cdots + X_n$

Know that $T(\underline{X}) \sim \text{Poisson}(n\lambda)$

$$\Rightarrow q(t \mid \lambda) = \frac{e^{-n\lambda} (n\lambda)^t}{t!} \quad \text{where } t = \sum_{i=1}^n x_i$$

$$\Rightarrow \frac{p(\underline{x} \mid \lambda)}{q(t \mid \lambda)} = \frac{e^{-n\lambda} \lambda^t \prod_{i=1}^n \frac{1}{x_i!}}{e^{-n\lambda} n^t \lambda^t \frac{1}{t!}} = \frac{t!}{n^t} \prod_{i=1}^n \frac{1}{x_i!}$$

= constant as a function of λ

$\Rightarrow T(\underline{X}) = \sum_{i=1}^n X_i$ is a sufficient statistic for λ.

Example: Gamma[1]

$Expo(\theta)$

$\parallel$

- Let $X_1, X_2, \ldots, X_n$ be a random sample from $\mathrm{Gamma}(1, \theta)$. Show that $T(\mathbf{X}) = \sum_{i=1}^{n} X_i$ is a sufficient statistic for θ.

$$p(\underline{x} \mid \theta) = \prod_{i=1}^{n} \frac{1}{\theta} e^{-x_i/\theta} = \theta^{-n} e^{-\frac{1}{\theta}\sum_{i=1}^{n} x_i} \quad = t$$

$$T = X_1 + X_2 + \cdots + X_n \sim \mathrm{Gamma}(n, \theta)$$

$$q(\underline{x} \mid \theta) = \frac{1}{\Gamma(n)\theta^n} t^{n-1} e^{-t/\theta} \quad \text{where } t = \sum_{i=1}^{n} x_i$$

$$\Rightarrow \frac{p(\underline{x} \mid \theta)}{q(t \mid \theta)} = \frac{\theta^{-n} e^{-t/\theta}}{\frac{1}{(n-1)!}\theta^{-n} t^{n-1} e^{-t/\theta}} = \frac{(n-1)!}{t^{n-1}}$$

= a constant as a function of θ

$\Rightarrow$ T is a sufficient statistic for θ.

[1] See book for Binomial and Normal

More about sufficient statistics

- The original sample is it self a sufficient statistic
  $T(\underline{X}) = (X_1, X_2, \cdots, X_n)$
- The vector of the n *order* statistics

$$T(\mathbf{X}) = (X_{(1)}, X_{(2)}, \ldots, X_{(n)})$$

is always a sufficient statistic for a random sample:

  - Joint pdf/pmf of $(X_{(1)}, X_{(2)}, \ldots, X_{(n)})$ is

$$n! f(x_1) f(x_2) \cdots f(x_n) \qquad \text{for } -\infty < x_1 < \cdots < x_n < \infty$$

  so

$$\frac{p(\mathbf{x} \mid \theta)}{q(T(\mathbf{x}) \mid \theta)} = \frac{f(x_1) f(x_2) \cdots f(x_n)}{n! f(x_1) f(x_2) \cdots f(x_n)} = \frac{1}{n!}$$

  - For some distributions the order statistics is as far as we can go with data reduction.

Finding a sufficient statistic

- Finding $q(T(\mathbf{x}) \mid \theta)$ can be difficult. There are ways around it!

Factorization Theorem

Let $f(\mathbf{x} \mid \theta)$ be the joint pmf or pdf of $\mathbf{X} = (X_1, \ldots, X_n)$. A statistic $T(\mathbf{X})$ is a *sufficient statistic for* θ if and only if $f(\mathbf{x} \mid \theta)$ can be written as

$$f(\mathbf{x} \mid \theta) = \underbrace{g(T(\mathbf{x}) \mid \theta)}\ \underbrace{h(\mathbf{x})}$$

function of θ and of $\underline{x}$ only through $T(\underline{x})$

i.e. could evaluate $g(T(x) \mid \theta)$ only knowing $T(\underline{x})$

function of $\underline{x}$ only, no θ

Poisson example:

$$f(\underline{x} \mid \lambda) = \underbrace{e^{-n\lambda} \lambda^{\sum_{i=1}^n x_i}}_{g(t \mid \lambda)} \cdot \underbrace{\prod_{i=1}^n \frac{1}{x_i!}}_{h(\underline{x})}$$

where $t = \sum_{i=1}^n x_i$

set $h(\underline{x}) = \prod_{i=1}^n \frac{1}{x_i!}$

and $g(T(\underline{x}) \mid \lambda) = e^{-n\lambda} \lambda^{T(\underline{x})}$

$\Rightarrow$ By the factorization theorem $T(\underline{X})$ is a sufficient statistic for λ

Example: Normal distribution

- Let $X_1, X_2, \ldots, X_n$ be a random sample from $\mathrm{N}(\mu, \sigma^2)$. Use the factorization theorem to show that the sample mean and sample variance are a sufficient statistic for (μ, σ^2) (both unknown)

$$f(\underline{x} \mid \mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)$$

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

$$= \underbrace{(2\pi)^{-n/2}}_{h(\underline{x})} \underbrace{\sigma^{-n} \exp\left(-\frac{1}{2\sigma^2} \left((n-1)s^2 + n(\bar{x}-\mu)^2\right)\right)}_{g(\bar{x}, s^2 \mid \mu, \sigma^2)}$$

$\Rightarrow$ by factorization theorem, $\bar{x}$ and s^2 are sufficient statistics for μ and σ^2.

Good fact to know: $$\sum_{i=1}^{n} (x_i - \mu)^2 = (n-1)s^2 + n(\bar{x} - \mu)^2$$

where $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$ and $s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$

Example: Uniform distribution

- Let $X_1, X_2, \ldots, X_n$ be a random sample from $\mathrm{Uniform}(0, \theta)$. Use the factorization theorem to show that the nth order statistic $X_{(n)}$ is a sufficient statistic for θ

$$f(x_i) = \frac{1}{\theta} I_{(0,\theta)}(x_i)$$

$$f(\underline{x} \mid \theta) = \prod_{i=1}^{n} \frac{1}{\theta} I_{(0,\theta)}(x_i)$$

$$= \theta^{-n} \underbrace{I_{(0,\theta)}(x_{(1)})\, I_{(0,\theta)}(x_{(n)})}_{g(x_{(1)},\, x_{(n)} \mid \theta)} \Rightarrow (x_{(1)}, x_{(n)}) \text{ are suff.}$$

$$= \theta^{-n} \underbrace{I_{(-\infty,\theta)}(x_{(n)})}_{g(x_{(n)} \mid \theta)} \underbrace{I_{(0,\infty)}(x_{(1)})}_{h(\underline{x})}$$

$\Leftarrow$ $X_{(n)}$ is a suff. stat for θ.

Sufficient statistics for exponential families

- Let $X_1, X_2, \ldots, X_n$ be a random sample from pdf or pmf of the form

$$f(x \mid \theta) = h(x)c(\theta)\exp\left(\sum_{j=1}^{k} w_j(\theta)t_j(x)\right)$$

- The statistic

$$T(\mathbf{X}) = \left(\sum_{i=1}^{n} t_1(X_i), \sum_{i=1}^{n} t_2(X_i), \ldots, \sum_{i=1}^{n} t_k(X_i)\right)$$

  is a sufficient statistic for θ

$$f(x \mid \boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta}) \exp\left(\sum_{j=1}^{k} w_j(\boldsymbol{\theta}) t_j(x)\right)$$

for each i

$i = 1, \ldots, n$

$$f(x_i \mid \underline{\theta}) = h(x_i) c(\underline{\theta}) \exp\left(t_1(x_i) w_1(\underline{\theta}) + \cdots + t_k(x_i) w_k(\underline{\theta})\right)$$

Joint pdf/pmf:

$$f(\underline{x} \mid \underline{\theta}) = \prod_{i=1}^{n} h(x_i) c(\underline{\theta}) \exp\left(t_1(x_i) w_1(\underline{\theta}) + \cdots + t_k(x_i) w_k(\underline{\theta})\right)$$

$$= \prod_{i=1}^{n} h(x_i) * c(\underline{\theta})^n$$

$$* \exp\left(w_1(\underline{\theta}) \sum_{i=1}^{n} t_1(x_i) + \cdots + w_k(\underline{\theta}) \sum_{i=1}^{n} t_k(x_i)\right)$$

$h(\underline{x})$

$$= g\left(\sum_{i=1}^{n} t_1(x_i), \ldots, \sum_{i=1}^{n} t_k(x_i) \mid \underline{\theta}\right)$$

Many sufficient statistics

- For many distributions there are many different sufficient statistics
- The whole sample $T(\mathbf{X}) = (X_1, X_2, \ldots, X_n)$ is a sufficient statistic
- The set of all order statistics $T(\mathbf{X}) = \left(X_{(1)}, X_{(2)}, \ldots, X_{(n)}\right)$ is a sufficient statistic
- Any one-to-one function of a sufficient statistic is a sufficient statistic

say, T is sufficient, let $c(T) = w \Rightarrow T = c^{-1}(w)$
& one-to-one

$f(\underline{x} \mid \theta) = h(\underline{x})\, g(T \mid \theta) = h(\underline{x})\, \underbrace{g(c^{-1}(w) \mid \theta)}_{\text{funct. of } w}$

Can we find a sufficient statistics that gives us most "data reduction" possible?

Minimal Sufficient Statistic

Def: Minimal sufficient statistic

A sufficient statistic $T(\mathbf{X})$ is called a **minimal sufficient statistic** if for any other sufficient statistic $T^*(\mathbf{X})$, $T(\mathbf{X})$ is a function of $T^*(\mathbf{X})$.

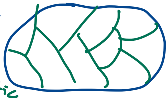
E.g. $T^* = (X_{(1)}, \dots, X_{(n)})$ and $T = \sum_{i=1}^{n} X_i$

then $T = \sum_{i=1}^{n} X_{(i)}$ i.e. T is a function of T^*

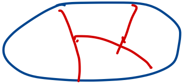
- Obtains the *coarsest* partition of the sample space as possible, without loosing any information about the parameter θ

$\mathcal{X}$

part. for a statistic



$\mathcal{X}$



$\Leftarrow$ coarser partition

- Can still have many different minimal statistics

→ one-to-one function

Finding a Minimal Sufficient Statistic

Theorem

Let $f(\mathbf{x} \mid \theta)$ be the pmf or pdf of a sample $\mathbf{X}$. Suppose there exists a function $T(\mathbf{X})$ such that:

- For every two outcomes $\mathbf{x}$ and $\mathbf{y}$ the ratio any $\underset{\sim}{x}, \underset{\sim}{y} \in \mathcal{X}$

$$\frac{f(\mathbf{x} \mid \theta)}{f(\mathbf{y} \mid \theta)} \quad *$$

is a constant as a function of θ *if and only if* $T(\mathbf{x}) = T(\mathbf{y})$

then $T(\mathbf{X})$ is a minimal sufficient statistic

① If $T(\underset{\sim}{x}) = T(\underset{\sim}{y})$ then $*$ is a const. as a funct. of θ

② If $*$ is a const. as a function of θ => $T(\underset{\sim}{x}) = T(\underset{\sim}{y})$

Example: $X_1, \dots, X_n$ are iid Poisson(λ)

$$f(\underline{x} \mid \lambda) = \prod_{i=1}^{n} \frac{e^{-\lambda} \lambda^{x_i}}{x_i!} = e^{-n\lambda} \lambda^{n\bar{x}} \prod_{i=1}^{n} \frac{1}{x_i!}$$

$$* = \frac{f(\underline{x} \mid \lambda)}{f(\underline{y} \mid \lambda)} = \frac{e^{-n\lambda} \lambda^{n\bar{x}} \prod_{i=1}^{n} \frac{1}{x_i!}}{e^{-n\lambda} \lambda^{n\bar{y}} \prod_{i=1}^{n} \frac{1}{y_i!}}$$

$$= \lambda^{n(\bar{x}-\bar{y})} \prod_{i=1}^{n} \frac{1}{x_i} \Big/ \prod_{i=1}^{n} \frac{1}{y_i}$$

$A \Rightarrow B$
equiv.:
$\neg B \Rightarrow \neg A$

1) If $\bar{x} = \bar{y}$ then $* = \prod_{i=1}^{n} \frac{1}{x_i} \Big/ \prod_{i=1}^{n} \frac{1}{y_i}$ = const. as a function of λ

2) If $\bar{x} \neq \bar{y}$ the $*$ does depend on λ.

$\Rightarrow$ $T(\underline{X}) = \bar{X}$ is a minimal suff. statistic.

Example: Normal distribution

$n\bar{x}^2 - 2n\bar{x}\mu + n\mu^2$
$n\bar{y}^2 - 2n\bar{y}\mu + n\mu^2$

- Let $X_1, X_2, \ldots, X_n$ be a random sample from $\mathrm{N}(\mu, \sigma^2)$, both μ and σ^2 are unknown. Find a minimal sufficient statistic for (μ, σ^2).

Recall: $f(\underline{x} \mid \mu, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\left[(n-1)s^2 + n(\bar{x}-\mu)^2\right]\right)$

$$\frac{f(\underline{x} \mid \mu, \sigma^2)}{f(\underline{y} \mid \mu, \sigma^2)} = \frac{\cancel{(2\pi\sigma^2)^{-n/2}} \exp\left(-\frac{1}{2\sigma^2}\left[(n-1)s_x^2 + n(\bar{x}-\mu)^2\right]\right)}{\cancel{(2\pi\sigma^2)^{-n/2}} \exp\left(-\frac{1}{2\sigma^2}\left[(n-1)s_y^2 + n(\bar{y}-\mu)^2\right]\right)}$$

where $s_x^2 = \frac{1}{n-1}\sum (x_i - \bar{x})^2$ and $s_y^2 = \frac{1}{n-1}\sum (y_i - \bar{y})^2$

$$= \exp\left(-\frac{1}{2\sigma^2}\left[(n-1)(s_x^2 - s_y^2) + n(\bar{x}^2 - \bar{y}^2) - 2n\mu(\bar{x}-\bar{y})\right]\right)$$

$=$ constant as a function of μ and σ^2
if and only if $s_x^2 = s_y^2$ and $\bar{x} = \bar{y}$

$\Rightarrow$ S^2 and $\bar{X}$ are minimal suff. stat.

Ancillary statistic

will not use much

Def: Ancillary statistic

A statistic $S(\mathbf{X})$ is called an **ancillary statistic** if its distribution does not depend on the parameter θ. $\Rightarrow$ I.e. no information about θ in $S(\underline{X})$

- Kind of an opposite to a sufficient statistic
- There are a few examples of ancillary statistics actually providing information about a parameter when combined with a sufficient statistic.
  - See examples in the textbook
- But often sufficient and ancillary statistics are *statistically independent*
  - At least if the sufficient statistic is *complete*

Examples of Ancillary Statistics

- Say $X_1, X_2, \ldots, X_n$ is a random sample from $\mathrm{N}(\mu, \sigma^2)$ where σ^2 is known. i.e. μ is the parameter of interest
- We have seen that

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

OR: $S^2 = \frac{\sigma^2}{n-1} W$ where $W \sim \chi^2_{n-1}$

distr. of S^2 does not depend on μ $\Rightarrow$ S^2 is an ancillary statistic for μ.

Complete statistic

* $g(T)$ is an unbiased estimator of zero.

Def: Complete statistic

Let $f(t \mid \theta)$ be a family of pdfs or pmfs for a statistic $T(\mathbf{X})$. The family is called **complete** if the following holds : any function $g(t)$:

- If $E(g(T)) \overset{*}{=} 0$ for all θ then $P(g(T) = 0) = 1$ for all θ

Also, the statistic $T(\mathbf{X})$ is called a **complete statistic**

- Property of the *family* of distributions $f(t \mid \theta)$ belongs to. distr. of the statistic
- Can be hard to verify sometimes
  - See examples 6.2.22 and 6.2.23 in the textbook

Completeness of exponential families

Theorem 6.2.25

Let $X_1, X_2, \ldots, X_n$ be a random sample from pdf or pmf of the form

$$f(x \mid \boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta}) \exp\left(\sum_{j=1}^{k} w_j(\boldsymbol{\theta}) t_j(x)\right)$$

The statistic

$$T(\mathbf{X}) = \left(\sum_{i=1}^{n} t_1(X_i), \sum_{i=1}^{n} t_2(X_i), \ldots, \sum_{i=1}^{n} t_k(X_i)\right)$$

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

- The condition means: the theorem does not hold for *curved exponential families* like $N(\theta, \theta^2)$

Example

- Let $X_1, \ldots, X_n$ be a random sample from $\mathrm{Poisson}(\lambda)$. Use theorem 6.2.25 to find a complete statistic.

pmf: $f(x|\lambda) = \dfrac{e^{-\lambda}\lambda^x}{x!} = \dfrac{1}{x!} e^{-\lambda} e^{\ln \lambda^x}$

$= \underbrace{\dfrac{1}{x!}}_{h(x)} \underbrace{e^{-\lambda}}_{c(\lambda)} e^{\overbrace{x}^{t(x)} \underbrace{\ln\lambda}_{w(\lambda)}}$ $\Rightarrow$ Exponential family

$\Rightarrow \sum_{i=1}^{n} t(x_i) = \sum_{i=1}^{n} X_i$ is a <u>complete</u> statistic.

Theorem regarding complete statistics

redundant condition

Basu's Theorem

If $T(\mathbf{X})$ is a complete and minimal sufficient statistic, then $T(\mathbf{X})$ is independent of every ancillary statistic.

E.g. $\bar{X}$ is a complete and minimal sufficient statistic for μ. (Given a random sample from $N(\mu,\sigma^2)$) S^2 is ancillary for μ $\Rightarrow$ By Basu, $\bar{X}$ and S^2 are independent.

Theorem 6.2.28

If a minimal sufficient statistic exists, then any complete statistic is also a minimal sufficient statistic

Complete $\Rightarrow$ minimal suff.

$\Uparrow$

strongest.

Summary

Let $X_1, \dots, X_n$ be a random sample from $f(x|\theta)$

Let $T = T(\underline{X})$

How to show that T is:

does not depend on θ

* Sufficient ?

① : Def: $f(\underline{x} \mid T(\underline{X}) = t)$ is a constant as a func. of θ

② : $\dfrac{f(\underline{x}|\theta)}{q(t|\theta)}$ ← pdf/pmf of T, $t = T(\underline{x})$ — " —

easiest → ③ : Factorization Theorem:

$$f(\underline{x}|\theta) = g\big(T(\underline{x})|\theta\big)\, h(\underline{x})$$

* Minimal sufficient?

for any $\underline{x}, \underline{y} \in \mathcal{X}$:

$\dfrac{f(\underline{x}|\theta)}{f(\underline{y}|\theta)}$ a constant as a function of θ if and only if $T(\underline{x}) = T(\underline{y})$

* Complete?

Exponential family.