STAT 713 sp 2023 Lec 02 slides

Data reduction part 2: Minimality, ancillarity

Karl B. Gregory

University of South Carolina

These slides are an instructional aid; their sole purpose is to display, during the lecture, definitions, plots, results, etc. which take too much time to write by hand on the blackboard. They are not intended to explain or expound on any material.

Setup and notation

- Let $\mathbf{X} = (X_1, \ldots, X_n)$ represent the set of rvs from an experiment.
- Use $\mathbf{x} = (x_1, \ldots, x_n)$ to represent a specific set of values for the rvs in $\mathbf{X}$.
- Let $f(\mathbf{x}; \theta)$ or $p(\mathbf{x}; \theta)$ denote the joint pdf or pmf of the rvs in $\mathbf{X}$, resp.
- The distribution of $\mathbf{X}$ depends on a parameter (or some parameters) $\theta \in \Theta$.
- A function $T(\mathbf{X})$ of the rvs in $\mathbf{X}$ is called a *statistic*.

Goal: Learn about θ from a realization of $\mathbf{X}$ via the value of a statistic $T(\mathbf{X})$.

Key concepts of data reduction in the rough:

1. A *sufficient statistic* carries all the information about θ from $\mathbf{X}$.
2. A *minimal sufficient statistic* carries the above and no more than this.
3. An *ancillary statistic* carries no information about θ.
4. A *complete statistic* cannot be used to construct an unbiased estimator of 0.

We think of computing a statistic $T(\mathbf{X})$ as "reducing" or summarizing the data.

Example: Let $X_1, \dots, X_n \overset{\text{ind}}{\sim} \text{Exponential}(\lambda)$ and consider the statistics

$$T_1(\mathbf{X}) = X_{(1)}, \quad T_2(\mathbf{X}) = \bar{X}_n, \quad T_3(\mathbf{X}) = S_n^2,$$

$$T_4(\mathbf{X}) = \left(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2\right), \quad T_5(\mathbf{X}) = X_{(1)}/X_{(n)}, \quad T_6(\mathbf{X}) = (X_{(1)}, \dots, X_{(n)}).$$

Sufficiency, *minimality*, *ancillarity*, and *completeness* address (resp) the questions

1. Which reductions of the data do not discard any information about λ?
2. Which ones keep all relevant information about λ, but discard all else?
3. Which ones discard all information about λ?
4. Which ones cannot be used to construct an unbiased estimator of 0?

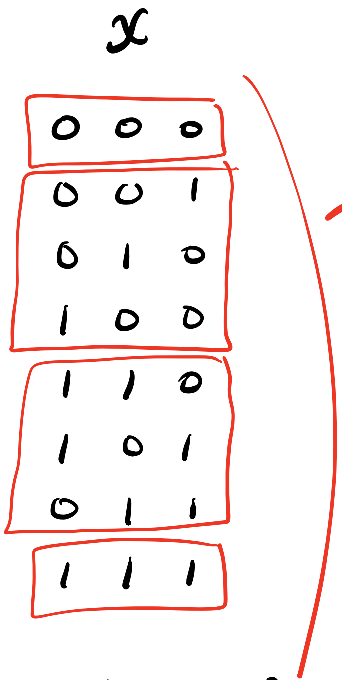
Table of Contents

Exercise: Let $X_1, X_2, X_3 \overset{\text{ind}}{\sim} \text{Bernoulli}(p)$, $p \in (0,1)$, and consider the statistics

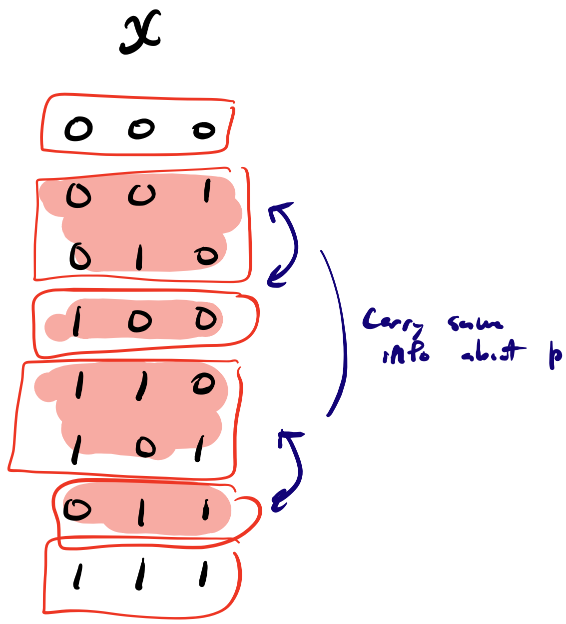
$$T(\mathbf{X}) = X_1 + X_2 + X_3 \quad \text{and} \quad T'(\mathbf{X}) = (\underline{X_1}, \underline{X_1 + X_2 + X_3}),$$

both of which are sufficient statistics for p. Now:

1. List the elements in the support $\mathcal{X}$ of $\mathbf{X} = (X_1, X_2, X_3)$.
2. Give the supports $\mathcal{T}$ and $\mathcal{T}'$ of $T(\mathbf{X})$ and $T'(\mathbf{X})$.
3. Identify the partition of the sample space $\mathcal{X}$ induced by each statistic.
4. Which suff. statistic corresponds to a coarser partition of the sample space?



$T(\underset{\sim}{X}) = X_1 + X_2 + X_3$ is "more minimal", induces a "coarser" partition of $\mathcal{X}$.



$\mathcal{T} = \{0, 1, 2, 3\}$

$\mathcal{T}' = \{ (0,0), (0,1), (0,2), (1,1), (1,2), (1,3) \}$

We want maximal data reduction without losing information about θ.

Minimal sufficient statistic

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

Hard to find minimal sufficient statistics using the definition. Use following result:

Theorem (Find a minimal suff. statistic, cf. Thm 6.2.13 in CB)

Let $\mathbf{X}$ *have joint pdf/pmf* $f(\mathbf{x};\theta)$ *with support not depending on* θ. *Let* $T(\mathbf{x})$ *be a function such that for any* $\mathbf{x}, \mathbf{y} \in \mathcal{X}$

$$\boxed{f(\mathbf{x};\theta)/f(\mathbf{y};\theta) \text{ is constant as a function of } \theta \iff T(\mathbf{x}) = T(\mathbf{y}).}$$

Then $T(\mathbf{X})$ *is a minimal sufficient statistic for* θ.

Exercise: Prove the result.

$$p(x;p) = p^x(1-p)^{1-x}\,\mathbb{1}(x\in\{0,1\})$$

Exercise: Let $X_1,\dots,X_n \overset{\text{ind}}{\sim} \text{Bernoulli}(p)$, where $p\in(0,1)$.

Check whether $T(\mathbf{X}) = \sum_{i=1}^n X_i$ is a minimal sufficient statistic for p.

$$f(\underset{\sim}{x};p) = \prod_{i=1}^n p^{x_i}(1-p)^{1-x_i}\,\mathbb{1}(x_i\in\{0,1\})$$

$$= p^{\sum x_i}(1-p)^{n-\sum x_i}\prod_{i=1}^n \mathbb{1}(x_i\in\{0,1\})$$

Consider $\underset{\sim}{x},\underset{\sim}{y}\in\mathcal{X}$.

$$\frac{f(\underset{\sim}{x};p)}{f(\underset{\sim}{y};p)} = \frac{p^{\sum x_i}(1-p)^{n-\sum x_i}\prod_{i=1}^n \mathbb{1}(x_i\in\{0,1\})}{p^{\sum y_i}(1-p)^{n-\sum y_i}\prod_{i=1}^n \mathbb{1}(y_i\in\{0,1\})}$$

$$= p^{\Sigma x_i - \Sigma y_i} (1-p)^{\Sigma y_i - \Sigma x_i}$$

is this free of p $\iff$ $\Sigma x_i = \Sigma y_i$?

Yes.

So $T(\underset{\sim}{x}) = \Sigma x_i$ is a min. suff. stat. for p.

Exercise: Let $X_1, \dots, X_n \overset{\text{ind}}{\sim} \text{Normal}(\mu, \sigma^2)$, where $\mu \in \mathbb{R}$, $\sigma > 0$.

Check if $T(\mathbf{X}) = (\bar{X}_n, \overset{S_x^2}{S_n^2})$ is a minimal sufficient statistic for (μ, σ^2).

For $\underset{\sim}{x}, \underset{\sim}{y} \in \mathcal{X}$

$$\frac{f(\underset{\sim}{x};\theta)}{f(\underset{\sim}{y};\theta)} = \frac{\prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}}{\prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} e^{-\frac{(y_i-\mu)^2}{2\sigma^2}}}$$

$$\begin{aligned} \sum (x_i-\mu)^2 &= \sum\left((x_i-\bar{x}) + (\bar{x}_n-\mu)\right)^2 \\ &= \sum (x_i-\bar{x})^2 + n(\bar{x}_n-\mu)^2 \\ &= (n-1)S_n^2 + n(\bar{x}_n-\mu)^2 \end{aligned}$$

$$= \frac{(2\pi)^{-n/2}(\sigma^2)^{-n/2}\exp\left[-\dfrac{(n-1)\overbrace{S_x^2}^{=\frac{1}{n-1}\sum(x_i-\bar{x}_n)^2} + n(\bar{x}_n-\mu)^2}{2\sigma^2}\right]}{(2\pi)^{n/2}(\sigma^2)^{-n/2}\exp\left[-\dfrac{(n-1)S_y^2 + n(\bar{y}_n-\mu)^2}{2\sigma^2}\right]}$$

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

is constant in (μ, σ^2) $\iff$ $(\bar{x}_n, S_n^2) = (\bar{y}_n, S_y^2)$.

So $T(\underset{\sim}{x}) = (\bar{x}_n, S_x^2)$

If the support of $\mathbf{X}$ depends on θ, we check minimal sufficiency more carefully:

Theorem (Minimal suff. if support depends on the parameter)

Let $\mathbf{X}$ have joint pdf/pmf $f(\mathbf{x};\theta)$ and let $T(\mathbf{x})$ be a function st for any $\mathbf{x}, \mathbf{y} \in \mathcal{X}$,

(i) $\{\theta : f(\mathbf{x};\theta) > 0\} = \{\theta : f(\mathbf{y};\theta) > 0\}$ *and*

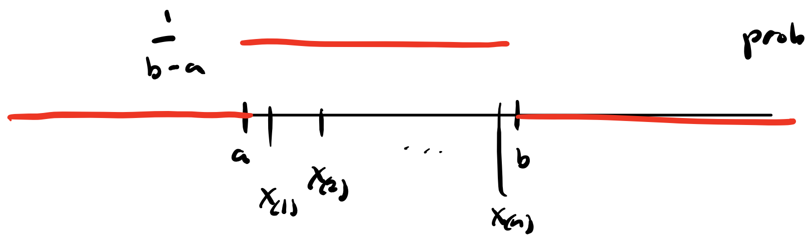
(ii) $f(\mathbf{x};\theta)/f(\mathbf{y};\theta)$ *is constant as a function of θ on* $\{\theta : f(\mathbf{y};\theta) > 0\}$

$\iff$ $T(\mathbf{x}) = T(\mathbf{y})$. *Then $T(\mathbf{X})$ is a minimal sufficient statistic.*

Exercise: Let $X_1, \dots, X_n \overset{\text{ind}}{\sim} \text{Uniform}(a, b)$, where $-\infty < a < b < \infty$.

Find a minimal sufficient statistic for (a, b).

$$f(x; a, b) = \frac{1}{b-a} \mathbb{I}(a < x < b)$$



probably $T(\underset{\sim}{X}) = (X_{(1)}, X_{(n)})$

For $\underset{\sim}{x}, \underset{\sim}{y} \in \mathcal{X}$

$$\frac{f(\underset{\sim}{x};\theta)}{f(\underset{\sim}{y};\theta)} = \frac{\prod_{i=1}^{n} \cancel{\frac{1}{b-a}}\, \mathbb{1}(a < x_i < b)}{\prod_{i=1}^{n} \cancel{\frac{1}{b-a}}\, \mathbb{1}(a < y_i < b)}$$

$$= \frac{\prod_{i=1}^{n} \mathbb{1}(a < x_i < b) \leftarrow \begin{cases} 1 & a < x_{(1)},\ x_{(n)} < b \\ 0 & \text{o.w.} \end{cases}}{\prod_{i=1}^{n} \mathbb{1}(a < y_i < b) \leftarrow \begin{cases} 1 & a < y_{(1)},\ y_{(n)} < b \\ 0 & \text{o.w.} \end{cases}}$$

$$= \frac{\mathbb{1}(a < x_{(1)})\ \mathbb{1}(x_{(n)} < b)}{\mathbb{1}(a < y_{(1)})\ \mathbb{1}(y_{(n)} < b)}$$

TRUE $\Longleftrightarrow$ $(x_{(1)}, x_{(n)}) = (y_{(1)}, y_{(n)})$.

(i) $\{\theta : f(\mathbf{x};\theta) > 0\} = \{\theta : f(\mathbf{y};\theta) > 0\}$ *and*

(ii) $f(\mathbf{x};\theta)/f(\mathbf{y};\theta)$ *is constant as a function of* θ *on* $\{\theta : f(\mathbf{y};\theta) > 0\}$

$\Longleftrightarrow$ $T(\mathbf{x}) = T(\mathbf{y})$. *Then* $T(\mathbf{X})$ *is a minimal sufficient statistic.*

So $T(\underset{\sim}{X}) = (X_{(1)}, X_{(n)})$ is min. suff stat for (a, b).

Theorem (Non-uniqueness of minimal sufficient statistics)

Any one-to-one function of a minimal sufficient statistic is also a minimal sufficient statistic.

Exercise: Prove the result.

Example: For $X_1, \dots, X_n \overset{\text{ind}}{\sim} \text{Normal}(\mu, \sigma^2)$, where $\mu \in \mathbb{R}$, $\sigma > 0$, both

$$T(\mathbf{X}) = (\bar{X}_n, S_n^2) \quad \text{and} \quad T'(\mathbf{X}) = \left(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2 \right)$$

are minimal sufficient statistics for (μ, σ^2).

Result:

If T' is min suff for θ and a is a 1:1 function then $T = a(T')$ is min suff for θ.

Proof:

(i) We have $T = a(T')$

Since this is min suff, it is a function of any suff. stat.

So T is also a function of any suff statistic.

Also must show that T is suff.

(ii) We have

$$f(\underset{\sim}{x};\theta) = g(T'(\underset{\sim}{x});\theta)\, h(\underset{\sim}{x})$$

$$= g\big(a^{-1}(\underbrace{a(T'(\underset{\sim}{x}))}_{T(\underset{\sim}{x})});\theta\big)\, h(\underset{\sim}{x})$$

$$= g\big(a^{-1}(T(\underset{\sim}{x}));\theta\big)\, h(\underset{\sim}{x})$$

$$= \tilde{g}(T(\underset{\sim}{x});\theta)\, h(\underset{\sim}{x}),$$

where $\tilde{g}(\cdot;\theta) = g(a^{-1}(\cdot);\theta)$.

So T is suff.

Minimality is not about the dimension of the statistic.

Exercise: Let $X_1, \dots, X_n$ be iid rvs equal to j with probability p_j for $j = 1, \dots, k$, where $p_1 + \cdots + p_k = 1$. Find two minimal sufficient statistics for $(p_1, \dots, p_k)$.

$$p(x; p_1, \dots, p_k) = \begin{cases} p_1 & x = 1 \\ \vdots & \\ p_k & x = k \\ 0 & \text{o.w.} \end{cases}$$

$X_1 \in \{1, \dots, k\}$

$p_1 \;\; \dots \;\; p_k$

$$= p_1^{\mathbb{1}(x=1)} \cdot p_2^{\mathbb{1}(x=2)} \cdot \ldots \cdot p_k^{\mathbb{1}(x=k)} \cdot \mathbb{1}(x \in \{1,\ldots,k\})$$

$$p(\underset{\sim}{x}; p_1,\ldots,p_k) = \prod_{i=1}^{n} p_1^{\mathbb{1}(x_i=1)} \cdot p_2^{\mathbb{1}(x_i=2)} \cdot \ldots \cdot p_k^{\mathbb{1}(x_i=k)} \cdot \mathbb{1}(x_i \in \{1,\ldots,k\})$$

$$= \underbrace{p_1^{N_1} \cdot p_2^{N_2} \cdot \ldots \cdot p_k^{N_k}}_{g(N_1,\ldots,N_k;\, p_1,\ldots,p_k)} \underbrace{\prod_{i=1}^{n} \mathbb{1}(x_i \in \{1,\ldots,k\})}_{h(\underset{\sim}{x})}$$

Let $N_j = \sum_{i=1}^{n} \mathbb{1}(x_i = j)$

$= \# \{x_i = j\}$

So $T(\underset{\sim}{X}) = (N_1,\ldots,N_k)$ is suff for $(p_1,\ldots,p_k)$.

Can also show that $(N_1,\ldots,N_k)$ is min. suff.

$$\frac{p(\underset{\sim}{x}; p_1,\ldots,p_k)}{p(\underset{\sim}{y}; p_1,\ldots,p_k)} = \frac{p_1^{N_1} \cdot p_2^{N_2} \cdot \ldots \cdot p_k^{N_k} \prod_{i=1}^{n} \mathbb{1}(x_i \in \{1,\ldots,k\})}{p_1^{M_1} \cdot p_2^{M_2} \cdot \ldots \cdot p_k^{M_k} \prod_{i=1}^{n} \mathbb{1}(y_i \in \{1,\ldots,k\})}$$

$M_j = \sum_{i=1}^{n} \mathbb{1}(y_i = j)$

const. in $(p_1,\ldots,p_k)$ $\iff$

$(N_1,\ldots,N_k) = (M_1,\ldots,M_k)$,

So $(N_1,\ldots,N_k)$ is min suff for $(p_1,\ldots,p_k)$.

Note $n = \sum_{j=1}^{k} N_j$, which means $N_k = n - (N_1 + \cdots + N_{k-1})$

So I can "throw away" N_k. Then $(N_1,\ldots,N_{k-1})$ is also a min. suff. stat.

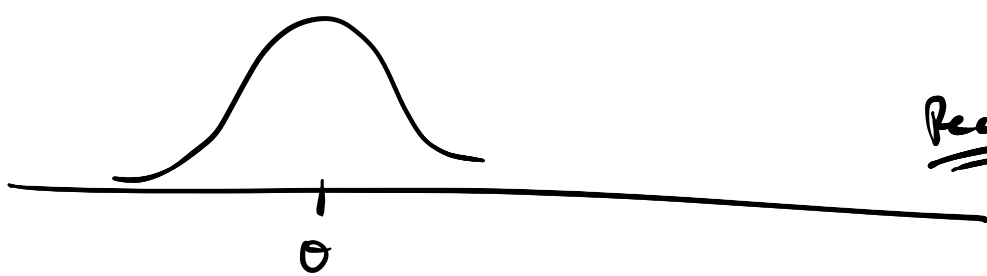
Table of Contents

Some statistics discard all information about θ.

Ancillary statistic

A statistic $S(\mathbf{X})$ is an *ancillary statistic* if its distribution does not depend on θ.

Exercise: Let $X_1, \ldots, X_n \overset{\text{ind}}{\sim} \text{Normal}(\theta, 1)$. Show that $S(\mathbf{X}) = S_n^2$ is ancillary to θ.



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

In this case $\sigma^2 = 1$, so $\boxed{(n-1)S_n^2 \sim \chi^2_{n-1}}$

I.e. dist. of S_n^2 is free of θ.

$X \sim f(x;\mu,\sigma) \Rightarrow Z = \frac{X-\mu}{\sigma} \sim f_Z$ and $\sigma Z + \mu \sim f(x;\mu,\sigma)$

Ancillary statistics are easy to find for location-scale families.

Recall location-scale families

The family of pdfs $\{f(x;\mu,\sigma) : \mu \in \mathbb{R}, \sigma > 0\}$ is a location-scale family if

$$f(x;\mu,\sigma) = \frac{1}{\sigma} f_Z\left(\frac{x-\mu}{\sigma}\right)$$

for some "standard" pdf f_Z for all $\mu \in \mathbb{R}$ and $\sigma > 0$.

Exercise: Find the location and/or scale parameters for the following families:

1. Exponential(λ), $\lambda > 0$.
2. Uniform($\mu - \theta, \mu + \theta$), $\mu \in \mathbb{R}$, $\theta > 0$.
3. Gamma(α_0, β), where $\alpha_0 > 0$ is known and $\beta > 0$.
4. Weibull(ν_0, β), where $\nu_0 > 0$ is known and $\beta > 0$.

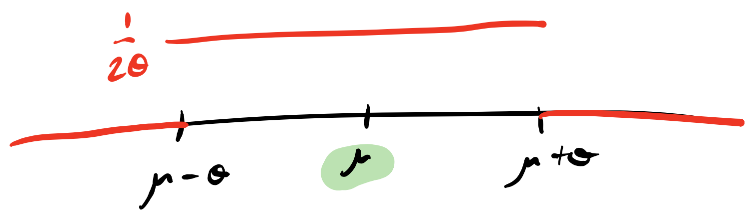
Exponential (λ): $$f(x;\lambda) = \frac{1}{\lambda} e^{-x/\lambda} \mathbb{1}(x > 0)$$

Put $f_Z(z) = e^{-z} \mathbb{1}(z > 0)$, "standard exponential"

Then $$\frac{1}{\lambda} f_Z\left(\frac{x-0}{\lambda}\right) = \frac{1}{\lambda} e^{-\left(\frac{x-0}{\lambda}\right)} \mathbb{1}\left(\frac{x-0}{\lambda} > 0\right)$$
$$= \frac{1}{\lambda} e^{-x/\lambda} \mathbb{1}(x > 0).$$

So this means Exponential (λ) densities are a scale family with scale parameter λ and standard density $f_Z(z) = e^{-z} \mathbb{1}(z > 0)$.

Uniform ($\mu-\theta, \mu+\theta$):



$$f(x;\mu,\theta) = \frac{1}{2\theta} \mathbb{1}(\mu-\theta < x < \mu+\theta)$$
$$= \frac{1}{\theta} \frac{1}{2} \mathbb{1}\left(-1 < \frac{x-\mu}{\theta} < 1\right)$$
$$= \frac{1}{\theta} f_Z\left(\frac{x-\mu}{\theta}\right), \quad f_Z(z) = \frac{1}{2} \mathbb{1}(-1 < z < 1)$$

So Uniform ($\mu-\theta, \mu+\theta$) is loc-scale with standard pdf and loc parameter μ, scale param. θ.

support of $X \sim f(x;\theta)$

is $\{ x \in \mathbb{R} : f(x;\theta) > 0 \}$

$a > 0 \quad b \in \mathbb{R}$

A function $S(x_1, \ldots, x_n)$ is *location-scale-invariant* if for any ~~$a \in \mathbb{R}$ and $b > 0$~~

$$S(x_1, \ldots, x_n) = S(ax_1 + b, \ldots, ax_n + b)$$

for all $(x_1, \ldots, x_n)$.

loc-inv. $S(x_1, \ldots, x_n) = S(ax_1, \ldots, ax_n)$

Theorem (Ancillary statistics for location-scale families)

If $\mathbf{X}$ *a rs from a L-S family, then* $S(\mathbf{X})$ *is ancillary if* $S(\mathbf{x})$ *is L-S-invariant.*

Exercise: Prove the above.

If S is L-S invariant, then

$$\underbrace{\tfrac{1}{\sigma}}_{a} x_1 + \underbrace{\tfrac{\mu}{\sigma}}_{b}$$

$$S(\underset{\sim}{x}) = S(x_1, \ldots, x_n) = S\left(\frac{x_1 - \mu}{\sigma}, \ldots, \frac{x_n - \mu}{\sigma} \right)$$

$$\overset{D}{=} S(Z_1, \ldots, Z_n), \quad Z_1, \ldots, Z_n \overset{ind}{\sim} f_Z$$

Exercise: Check for ancillarity of the statistics:

$\bar{X}_n = \frac{1}{n}\sum X_i = \frac{1}{n}\sum_{i=1}^{n} \theta Z_i$

$= \theta \bar{Z}_n$

1. For $X_1, \dots, X_n \overset{\text{ind}}{\sim} \text{Uniform}(0, \theta)$, check

$$S_1(\mathbf{X}) = \frac{X_{(1)}}{X_{(n)}} \checkmark \quad \text{and} \quad \underbrace{S_2(\mathbf{X}) = \frac{X_1 - X_2}{\bar{X}_n}}_{\text{ancillary.}} \checkmark = \frac{\cancel{\theta} Z_1 - \cancel{\theta} Z_2}{\cancel{\theta} \bar{Z}_n}$$

2. For $X_1, \dots, X_n \overset{\text{ind}}{\sim} \text{Normal}(\mu, \sigma^2)$, check

$X = \sigma Z + \mu$

$$\cancel{S_1(\mathbf{X}) = \left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2\right)} \quad \text{and} \quad S_2(\mathbf{X}) = \frac{X_2 - X_1}{X_3 - X_2}. \checkmark$$

$$= \frac{(\sigma Z_2 + \mu) - (\sigma Z_1 + \mu)}{(\sigma Z_3 + \mu) - (\sigma Z_2 + \mu)} = \frac{Z_2 - Z_1}{Z_3 - Z_2}$$

3. For $X_1, \dots, X_n \overset{\text{ind}}{\sim} f(x; \theta) = e^{-(x-\theta)} e^{-e^{-(x-\theta)}}$, check

$$\cancel{S_1(\mathbf{X}) = \frac{X_1}{X_1 + \dots + X_n}} \quad \text{and} \quad S_2(\mathbf{X}) = X_{(n)} - X_{(1)}. \checkmark$$

$f(x;\theta) = e^{-(x-\theta)} e^{-e^{-(x-\theta)}} = f_Z(x-\theta)$ for $f_Z(z) = e^{-z} e^{-e^{-z}}$ for all θ.

① Uniform $(0, \theta)$

$$f(x;\theta) = \frac{1}{\theta} \mathbb{1}(0 < x < \theta)$$

$$= \frac{1}{\theta} \mathbb{1}\left(0 < \frac{x}{\theta} < 1\right)$$

$$= \frac{1}{\theta} f_Z\left(\frac{x}{\theta}\right) \quad \text{for} \quad f_Z(z) = \mathbb{1}(0 < z < 1)$$

So Uniform $(0, \theta)$ is a scale family with standard pdf the Unif$(0,1)$ pdf and scale param. θ.

$$S_1(\underset{\sim}{x}) = S(x_1, \dots, x_n) = \frac{x_{(1)}}{x_{(n)}}$$

$$S_1(x_1, \dots, x_n) = S_1(a x_1, \dots, a x_n) = \frac{a\, x_{(1)}}{a\, x_{(n)}} = \frac{x_{(1)}}{x_{(n)}}$$

Since S_1 is scale-invariant, $\frac{X_{(1)}}{X_{(n)}}$ is ancillary.

Can write

$$X_1 = \theta Z_1, \dots, X_n = \theta Z_n, \quad Z_1, \dots, Z_n \overset{iid}{\sim} U(0,1).$$

$$\frac{X_{(1)}}{X_{(n)}} = \frac{\theta\, Z_{(1)}}{\theta\, Z_{(n)}} = \frac{Z_{(1)}}{Z_{(n)}} \text{ has a dist free of } \theta.$$

Questions:

1. Are ancillary statistics worthless? No; see discussion in Sec 6.2.3 in CB.
2. Are ancillary statistics independent of minimal sufficient statistics?

This leads us to consider the property of *completeness*.