STAT 713 sp 2023 Lec 03 slides

Data reduction part 3: Completeness

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* retains info about θ under any non-deg. transformation.

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

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

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

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 retain info about λ under any non-deg. transformation?

Let $X_1, \ldots, X_n \overset{ind}{\sim} \text{Gamma}(2, \beta)$, $\beta > 0$.

Find an ancillary stat.

$$f(x;\beta) = \frac{1}{\Gamma(2)\beta^2} x^{2-1} e^{-x/\beta} \mathbb{1}(x>0)$$

$$= \frac{1}{\beta}\frac{x}{\beta} e^{-x/\beta} \mathbb{1}(x>0)$$

$$= \frac{1}{\beta} f_Z\left(\frac{x}{\beta}\right) \qquad f_Z(z) = z e^{-z} \mathbb{1}(z>0)$$

So we have a scale family.

$$S_1(\underset{\sim}{X}) = \frac{X_{(1)}}{X_{(n)}} = \frac{\beta Z_{(1)}}{\beta Z_{(n)}} = \frac{Z_{(1)}}{Z_{(n)}} \qquad X_i = \beta Z_i,\ Z_i \sim f_Z \quad i=1,\ldots,n$$

$$S_2(\underset{\sim}{X}) = \frac{X_1}{X_2} \qquad S_3(\underset{\sim}{X}) = \frac{\bar{X}_n}{X_{(1)}}.$$

Complete statistic

A statistic $T(\mathbf{X})$ is called *complete* if for every function g such that $\mathbb{E}_\theta g(T(\mathbf{X})) = 0$ for all θ, we have $P_\theta(g(T(\mathbf{X})) = 0) = 1$ for all θ.

→ g(T(X)) is an unbiased estimator of 0

- If we can't construct an unbiased estimator of 0 from $T(\mathbf{X})$, it is complete.
- If we can find nonzero g st $\mathbb{E}_\theta g(T(\mathbf{X})) = 0$ for all θ, $T(\mathbf{X})$ is not complete.

This property becomes important later on, when we search for "best" estimators.

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

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

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

Determine which properties each statistic has among *sufficiency*, *minimality*, *ancillarity*, and *completeness*.

Exercise: Let $X_1, \ldots, X_n \overset{\text{ind}}{\sim} \text{Normal}(\theta, \theta^2)$, $\theta \in \mathbb{R}$.

✓ 1. Find a sufficient statistic for θ.
2. Check whether it is a complete statistic.

$$f(\underset{\sim}{x}; \theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \frac{1}{\theta} \exp\left[- \frac{(x_i - \theta)^2}{2\theta^2} \right]$$

$$\vdots$$

$$= (2\pi)^{-n/2} (\theta^2)^{-n/2} \exp\left[- \frac{(n-1)S_n^2 + n(\bar{X}_n - \theta)^2}{2\theta} \right]$$

So $T(\underset{\sim}{X}) = (\bar{X}_n, S_n^2)$ is a suff stat for θ.

Complete statistic

A statistic $T(\mathbf{X})$ is called *complete* if for every function g such that $\mathbb{E}_\theta g(T(\mathbf{X})) = 0$ for all θ, we have $P_\theta(g(T(\mathbf{X})) = 0) = 1$ for all θ.

Let g be any function s.t.

$$\mathbb{E}_\theta\, g(\bar{X}_n, S_n^2) = 0 \quad \forall\ \theta \in \mathbb{R}.$$

Does this imply $g(\bar{X}_n, S_n^2) = 0$ w.p. 1 for all $\theta \in \mathbb{R}$?

"with probability 1"

Play around to find a function g such that $\mathbb{E}_\theta g(\bar{X}_n, S_n^2) = 0$ $\forall\theta$

BUT $P_\theta(g(\bar{X}_n, S_n^2) = 0) \neq 1$ for some θ

$$\mathbb{E}\, S_n^2 = \theta^2$$

$$\mathbb{E}\, \bar{X}_n = \theta$$

$$\mathbb{E}\, \bar{X}_n^2 = \underbrace{\mathrm{Var}\, \bar{X}_n} + \underbrace{(\mathbb{E}\bar{X}_n)^2}_{\theta^2} = \frac{\theta^2}{n} + \theta^2 = \theta^2\left(\frac{1}{n} + 1\right) = \theta^2\left(\frac{n+1}{n}\right)$$

I propose

$$g(\bar{X}_n, S_n^2) = \frac{n}{n+1}\bar{X}_n^2 - S_n^2 .$$

Then $\mathbb{E}_\theta\, g(\bar{X}_n, S_n^2) = \frac{n}{n+1}\mathbb{E}\,\bar{X}_n^2 - \mathbb{E}\, S_n^2 = \frac{n}{n+1}\theta^2\frac{n+1}{n} - \theta^2 = 0$ $\forall\,\theta$

But $P_\theta(g(\bar{X}_n, S_n^2) = 0) \neq 1$.

So $T(\mathbf{X}) = (\bar{X}_n, S_n^2)$ is not complete.

Complete statistic

A statistic $T(\mathbf{X})$ is called *complete* if for every function g such that $\mathbb{E}_\theta g(T(\mathbf{X})) = 0$ for all θ, we have $P_\theta(g(T(\mathbf{X})) = 0) = 1$ for all θ.

Exercise: Let $X_1, \ldots, X_n \overset{\text{ind}}{\sim} \text{Bernoulli}(p)$, $0 < p < 1$. $f(x;p) = p^x(1-p)^{1-x}\mathbb{1}(x \in \{0,1\})$

1. Check whether X_1 is a complete statistic. Is it a sufficient statistic? ← yes!
2. Find a minimal sufficient statistic for p.
3. Check whether it is a complete statistic.

① Is $T(\underset{\sim}{X}) = X_1$ complete?

Let g be a function s.t. $\mathbb{E}_p\, g(X_1) = 0$ $\forall\, p \in (0,1)$.

Does this mean $g(X_1) = 0$ w.p. 1?

Write $0 = \mathbb{E}_p\, g(X_1) = \sum_{x=0}^{1} g(x)\, p^x (1-p)^{1-x}$

$= g(0)(1-p) + g(1)\, p$

$= 0 \quad \forall \quad p \in (0,1).$

$\Rightarrow$ $g(0) = g(1) = 0.$ To see this, write

$p = 1/2 \qquad g(0)\,\frac{1}{2} + g(1)\,\frac{1}{2} = 0$

$p = 1/3 \qquad g(0)\,2/3 + g(1)\,1/3 = 0$

$\Rightarrow \quad g(0) = -g(1)$

$\Rightarrow \quad -g(1)\,\frac{2}{3} + g(1)\,1/3 = 0$

$\Leftrightarrow \quad -\frac{1}{3}\, g(1) = 0$

$\Leftrightarrow \quad g(1) = 0$

So $\quad g(0) = g(1) = 0.$

This means $\quad P_p\left(g(X_1) = 0 \right) = 1.$

Therefore $T(\underset{\sim}{X}) = X_1$ is complete.

Is $T(\underset{\sim}{X}) = X_1$ suff?

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

$= p^{x_1} (1-p)^{1-x_1}\, \mathbb{1}(x_1 \in \{0,1\}) \prod_{i=2}^{n} p^{x_i} (1-p)^{1-x_i}\, \mathbb{1}(x_i \in \{0,1\})$

Cannot factorize, so $T(\underset{\sim}{X}) = X_1$ is not suff.

② $f(\underset{\sim}{x}; p) = \prod_{i=1}^{n} p^{x_i} (1-p)^{1-x_i} \mathbb{1}(x_i \in \{0,1\})$

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

So $T(\underset{\sim}{X}) = \sum_{i=1}^n X_i$ is suff. [Can show it is min. suff]

Complete statistic

A statistic $T(\mathbf{X})$ is called *complete* if for every function g such that $\mathbb{E}_\theta g(T(\mathbf{X})) = 0$ for all θ, we have $P_\theta(g(T(\mathbf{X})) = 0) = 1$ for all θ.

③ Is $T(\underset{\sim}{X}) = \sum_{i=1}^n X_i$.

Let g be fun. s.t. $\mathbb{E}_p\, g(\sum_{i=1}^n X_i) = 0 \quad \forall\, p \in (0,1)$.

Write

$Y \sim \text{Binomial}(n, p)$

$$0 = \mathbb{E}_p\, g\left(\sum_{i=1}^n X_i\right) = \sum_{y=0}^n g(y) \binom{n}{y} p^y (1-p)^{n-y}$$

$$= \underbrace{(1-p)^n}_{>0} \sum_{y=0}^n g(y) \binom{n}{y} \left(\frac{p}{1-p}\right)^y \quad \forall\, p \in (0,1)$$

$$\Rightarrow \sum_{y=0}^n g(y)\binom{n}{y} t^y = 0 \quad \forall\, t \in \mathbb{R}$$

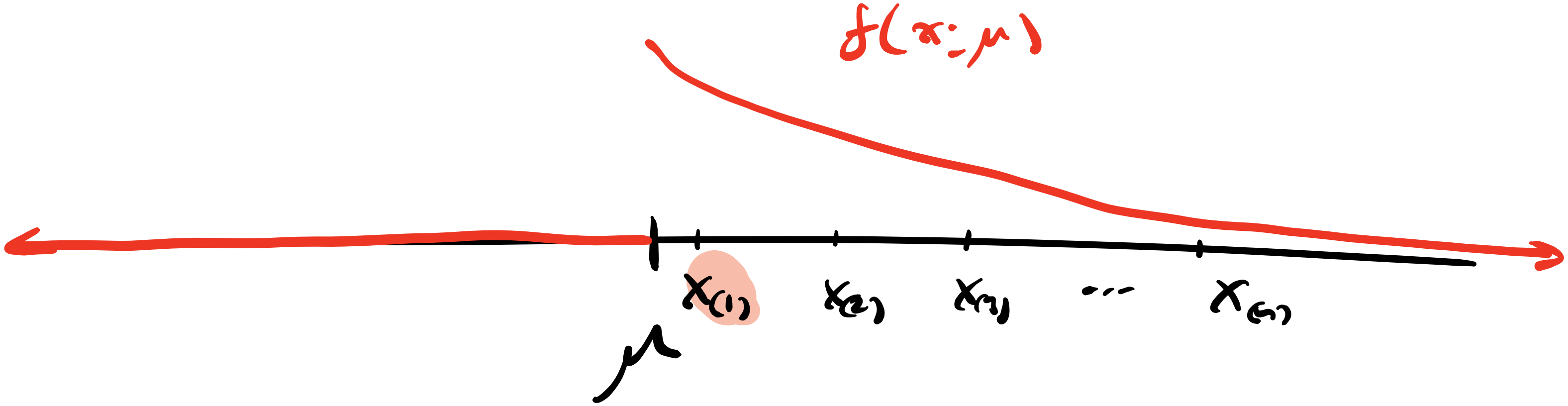
$$f(t) = a_0 + a_1 t + a_2 t^2 + \cdots + a_n t^n = 0 \quad \forall\, t \in \mathbb{R}$$

$$\Leftrightarrow a_0 = a_1 = \cdots = a_n = 0.$$

Therefore $g(y) = 0$ for $y = 0, 1, \ldots, n$.

Means $P_p\left(g\left(\sum_{i=1}^n X_i\right) = 0\right) = 1 \quad \forall\, p \in (0,1)$,

so $T(\underset{\sim}{X}) = \sum_{i=1}^n X_i$ is a complete stat.



Exercise: Let $X_1, \dots, X_n \stackrel{\text{ind}}{\sim} f(x;\mu) = e^{-(x-\mu)}\mathbf{1}(x > \mu)$, where $\mu \in \mathbb{R}$.

1. Find a minimal sufficient statistic for μ.
2. Check whether it is a complete statistic.

①

$$f(\underset{\sim}{x};\mu) = \prod_{i=1}^{n} e^{-(x_i-\mu)} \underbrace{\mathbf{1}(x_i > \mu)}$$

$$= e^{-\sum_{i=1}^{n}(x_i-\mu)} \mathbf{1}(x_{(1)} > \mu)$$

$$= \underbrace{e^{-\sum_{i=1}^{n} x_i}}_{h(\underset{\sim}{x})} \underbrace{e^{n\mu} \mathbf{1}(x_{(1)} > \mu)}_{g(x_{(1)};\mu)}$$

To check minimal suff: let $\underset{\sim}{x}, \underset{\sim}{y}$ be samples.

$$\frac{f(\underset{\sim}{x};\mu)}{f(\underset{\sim}{y};\mu)} = \frac{e^{-\sum x_i}\, \cancel{e^{n\mu}}}{e^{-\sum y_i}\, \cancel{e^{n\mu}}} \; \frac{\mathbb{1}(x_{(1)} \geq \mu)}{\mathbb{1}(x_{(1)} \geq \mu)}$$

num, denominator > 0 for same value of μ $\Leftrightarrow$ $x_{(1)} = y_{(1)}$.
And ratio is constant $\Leftrightarrow$ $x_{(1)} = y_{(1)}$.

So $T(\underset{\sim}{x}) = X_{(1)}$ is min suff.

Complete statistic

A statistic $T(\mathbf{X})$ is called *complete* if for every function g such that $\mathbb{E}_\theta g(T(\mathbf{X})) = 0$ for all θ, we have $P_\theta(g(T(\mathbf{X})) = 0) = 1$ for all θ.

③ Let g be a fn. s.t. $\mathbb{E}_\mu\, g(X_{(1)}) = 0 \quad \forall \mu$.

Want to show that as consequence $g(X_{(1)}) = 0$ w.p. 1 $\forall \mu$.

Write

$$0 = \mathbb{E}_\mu\, g(X_{(1)}) = \int_{\mathbb{R}} g(x)\, f_{X_{(1)}}(x;\mu)\, dx$$

$$f_X(x;\mu) = \underline{e^{-(x-\mu)}\, \mathbb{1}(x \geq \mu)}$$

For $x \geq \mu$,

$$F_X(x;\mu) = \int_\mu^x e^{-(t-\mu)}\, dt = e^{\mu} \int_\mu^x e^{-t}\, dt = e^{\mu}\left[-e^{-t}\right]\Big|_\mu^x = e^{\mu}\left[-e^{-x} + e^{-\mu}\right]$$

$$= \underline{1 - e^{-(x-\mu)}}$$

$$f_{X_{(1)}}(x;\mu) = n\left[1 - F_X(x;\mu)\right]^{n-1} f_X(x;\mu)$$

$$= n\left[1 - \left(1 - e^{-(x-\mu)}\right)\right]^{n-1} e^{-(x-\mu)} \quad \mathbb{1}(x \geq \mu)$$

$$= n \, e^{-(x-\mu)(n-1)} e^{-(x-\mu)} \quad \mathbb{1}(x \geq \mu)$$

$$= n \, e^{-n(x-\mu)} \quad \mathbb{1}(x \geq \mu).$$

So

$$0 = \mathbb{E}_\mu \, g(X_{(1)}) = \int_\mu^\infty g(x) \, n \, e^{-n(x-\mu)} \, dx \qquad \forall \mu \in \mathbb{R}$$

$$\Rightarrow \quad g(x) = 0.$$

To see this, write

$$0 = \frac{\partial}{\partial\mu} \mathbb{E}_\mu \, g(X_{(1)})$$

$$= \frac{\partial}{\partial\mu} \int_\mu^\infty g(x) \, n \, e^{-n(x-\mu)} \, dx$$

$$= \frac{\partial}{\partial\mu} \left(n e^{n\mu}\right) \left(\int_\mu^\infty g(x) \, e^{-nx} \, dx \right)$$

$$= \left(\frac{\partial}{\partial\mu} n e^{n\mu} \right) \int_\mu^\infty g(x) \, e^{-nx} \, dx$$

$$+ \; n e^{n\mu} \left(\frac{\partial}{\partial\mu} \int_\mu^\infty g(x) \, e^{-nx} \, dx \right)$$

$$= n^2 e^{n\mu} \int_\mu^\infty g(x) e^{-nx} dx$$

$$+ n e^{n\mu} \left(- g(\mu) e^{-n\mu} \right)$$

$$= n \underbrace{\int_\mu^\infty g(x)\, n e^{-n(x-\mu)} dx}_{\mathbb{E}_\mu g(X_{(1)}) = 0} - n g(\mu)$$

$$= -n g(\mu) \qquad \forall \mu \in \mathbb{R}$$

$$(= 0)$$

$$\Rightarrow g(\mu) = 0 \qquad \forall \mu \in \mathbb{R}.$$

Mean $P_\mu \left(g(X_{(1)}) = 0 \right) = 1 \quad \forall \mu \in \mathbb{R}$,

so $T(\underset{\sim}{X}) = X_{(1)}$ is complete.

Theorem (Functions of complete statistics)

Any known function of a complete statistic is complete.

Exercise: Prove the result.

Let $T' = T'(\underset{\sim}{x})$ be a complete statistic

Then define $T = a(T')$ for some function a.

Let g be some function s.t. $\mathbb{E}_\theta\, g(T) = 0 \;\; \forall \theta$.

Then $\mathbb{E}_\theta\, g(a(T')) = 0 \;\; \forall \theta$.

Since T' is complete, we have

$$P_\theta\big(g(\underbrace{a(T')}_{\text{is } T}) = 0\big) = 1 \quad \forall \theta.$$

$$P_\theta\big(g(T) = 0\big) = 1 \quad \forall \theta.$$

So T is complete.

In most situations it is hard to check completeness using the definition. Try this:

Theorem (Completeness via exponential family, cf. Thm 6.2.25 in CB)

Suppose $X_1, \dots, X_n \overset{\text{ind}}{\sim} f(x;\theta)$, *where*

$$f(x;\theta) = h(x)c(\theta)\exp\left(\sum_{j=1}^k w_j(\theta)t_j(x)\right), \quad x \in \mathbb{R}, \quad \theta \in \Theta,$$

and the range of $(w_1(\theta), \dots, w_k(\theta))$ *over* $\theta \in \Theta$ *contains an open set in* $\mathbb{R}^k$. *Then*

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

is a complete minimal sufficient statistic for θ.

- Use to find complete statistics in full exponential families.
- Completeness not guaranteed for curved exponential families.

Exercise: Check for completeness in family Normal(θ, θ^2), $\theta > 0$. $(\bar{X}_n, S_n^2)$

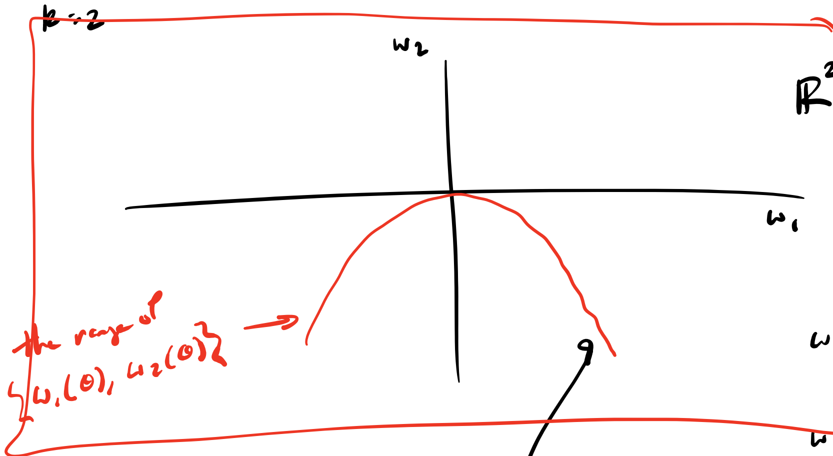
$$f(x;\theta) = \frac{1}{\sqrt{2\pi}} \frac{1}{\theta} \exp\left[-\frac{(x-\theta)^2}{2\theta^2}\right]$$

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

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

$t_1(x) = x$ $\quad$ $t_2(x) = x^2$

$w_1(\theta) = \frac{1}{\theta}$ $\quad$ $w_2(\theta) = -\frac{1}{2\theta^2}$



$\mathbb{R}^2 = \mathbb{R}^k$

$\{w_1(\theta), w_2(\theta) : \theta \in \Theta\}$ ← $\mathbb{R}$

$w_1 = \frac{1}{\theta} \iff \theta = \frac{1}{w}$

$w_2 = -\frac{1}{2\theta^2} \iff w_2 = -\frac{1}{2\left(\frac{1}{w_1}\right)^2} = -\frac{w_1^2}{2}$

$\{(w_1, w_2) : w_2 = -\frac{1}{2} w_1^2\}$

$\log(a \cdot b) = \log a + \log b$

$\sum \log X_i = \log\left(\prod X_i\right)$

Exercise: Let $X_1, \dots, X_n \overset{\text{ind}}{\sim} \text{Gamma}(\alpha, \beta)$, $\alpha > 0$, $\beta > 0$. Find a complete [minimal] sufficient statistic.

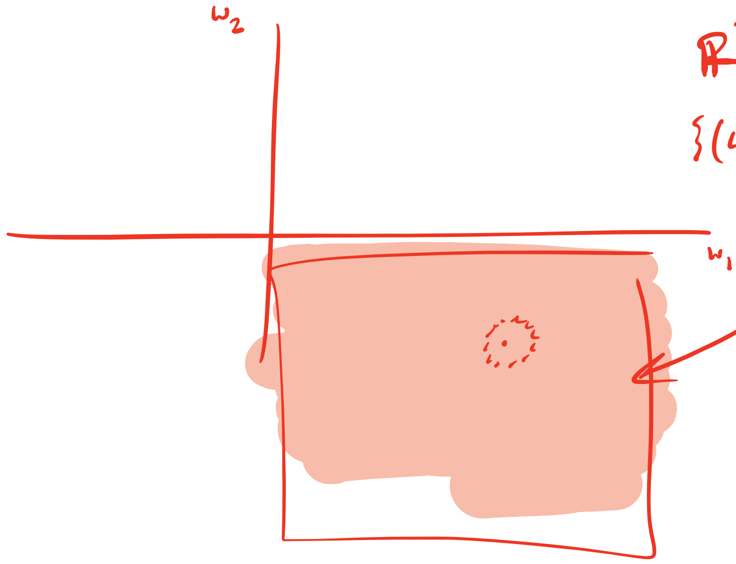
$$f(x; \alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} \mathbb{1}(x>0)$$

$$= \underbrace{\frac{1}{x}\mathbb{1}(x>0)}_{h(x)} \underbrace{\frac{1}{\Gamma(\alpha)\beta^\alpha}}_{c(\alpha,\beta)} \exp\left[\alpha \log x - \frac{x}{\beta}\right]$$

$w_1(\alpha,\beta) = \alpha$, $t_1(x) = \log x$

$w_2(\alpha,\beta) = -\frac{1}{\beta}$, $t_2(x) = x$

So $T(\underset{\sim}{X}) = \left(\sum_{i=1}^n \log X_i, \sum_{i=1}^n X_i\right)$ is a complete suff. stat. for (α,β)



$\mathbb{R}^2$

$$\{(w_1(\alpha,\beta),\ w_2(\alpha,\beta)) : \alpha > 0,\ \beta > 0\}$$

$$= \{(\alpha,\ -\tfrac{1}{\beta}) : \alpha > 0,\ \beta > 0\}$$

Exercise: Let $X_1, \ldots, X_n \overset{\text{ind}}{\sim} (2x/\theta)e^{-x^2/\theta} \cdot \mathbf{1}(x > 0)$, where $\theta > 0$.

Identify, if possible, a complete sufficient statistic.

$$f(x;\theta) = \frac{2x}{\theta} e^{-x^2/\theta} \mathbb{1}(x>0)$$

$$= \underbrace{\frac{2}{\theta}}_{c(\theta)} \cdot \underbrace{x \cdot \mathbb{1}(x>0)}_{h(x)} \quad e^{-\frac{1}{\theta} x^2}$$

$w_1(\theta) = -\frac{1}{\theta}$

$t_1(x) = x^2$

$\Rightarrow T(\underline{X}) = \sum_{i=1}^{n} X_i^2$ is complete min. suff.

Theorem

Let $T(\mathbf{X})$ be a complete sufficient statistic for a parameter θ. Then

1. *(Lehmann, cf. Thm 6.2.28 in CB): $T(\mathbf{X})$ is minimal sufficient if at least one minimal sufficient statistic exists.*
2. *(Basu, cf. Thm 6.2.24 in CB): $T(\mathbf{X})$ is indep. of every ancillary statistic.*

Lehmann's says only minimal sufficient statistics can be complete and sufficient.

Exercise: Go through proofs of the above results.

① Lehmann's: Let T be complete, suff.

Let U be min suff stat.

Want to show that T is a function of U.

Define a function $g(u) = \mathbb{E}[T \mid U = u]$. This is free of θ since U is suff.

Now write

$$\mathbb{E}_\theta[g(U) - T] = \mathbb{E}_\theta\, g(U) - \mathbb{E}_\theta T$$
$$= \mathbb{E}_\theta\left(\mathbb{E}[T \mid U]\right) - \mathbb{E}_\theta T$$
$$= \mathbb{E}_\theta T - \mathbb{E}_\theta T$$
$$= 0.$$

This implies, since T is complete, that $g(U) - T = 0$ w.p.-1.

That is $T = g(U)$ w.p.1.

So T is a fn. of U.

This means T is min suff.

② **Basu's**: Let T be complete suff.
Let S be ancillary.
Show $S \perp\!\!\!\perp T$.

Let $S \in \mathbb{R}^k$ and fix $\underset{\sim}{s} \in \mathbb{R}^k$.

We want to show

$$\underbrace{P(S_1 \le s_1, \dots, S_k \le s_k \mid T = t)}_{F_S(\underset{\sim}{s} \mid t)} = \underbrace{P(S_1 \le s_1, \dots, S_k \le s_k)}_{F_S(\underset{\sim}{s})} \quad \forall t, \ \forall \theta.$$

Define

$$g(t) = F_S(\underset{\sim}{s} \mid t) = P(S_1 \le s_1, \dots, S_k \le s_k \mid T = t)$$
$$= \mathbb{E}\left[\mathbb{1}(S_1 \le s_1, \dots, S_k \le s_k) \mid T = t \right].$$

Note $g(t)$ is free of θ since T is suff.

Write

$$\mathbb{E}_\theta\left[g(T) - F_S(\underset{\sim}{s})\right] = \mathbb{E}_\theta\, g(T) - \mathbb{E}_\theta\, F_S(\underset{\sim}{s})$$
$$= \mathbb{E}_\theta\left(\mathbb{E}\left[\mathbb{1}(S_1 \le s_1, \dots, S_k \le s_k) \mid T\right]\right) - F_S(\underset{\sim}{s})$$
$$= \mathbb{E}_\theta\, \mathbb{1}(S_1 \le s_1, \dots, S_k \le s_k) - F_S(\underset{\sim}{s})$$
$$= P(S_1 \le s_1, \dots, S_k \le s_k) - F_S(\underset{\sim}{s})$$
$$= F_S(\underset{\sim}{s}) - F_S(\underset{\sim}{s})$$
$$= 0,$$

which means, since T is complete, that

$$g(T) = F_S(s) \quad \text{w.p. } 1,$$

So we have

$$F_S(s \mid t) = F_S(s).$$

So this means $S \perp\!\!\!\perp T$.

Exercise: Let $X_1, \dots, X_n \overset{\text{ind}}{\sim} \text{Normal}(\mu, \sigma_0^2)$, where $\mu \in \mathbb{R}$ and σ_0^2 is known.

Use Basu's theorem to argue that $\bar{X}_n$ and S_n^2 are independent.

$\bar{X}_n$ is complete suff for μ. (Use exponential family result).

S_n^2 is ancillary for μ. $\left(\frac{(n-1) S_n^2}{\sigma_0^2} \sim \chi^2_{n-1} \right)$

$$f(x;\beta) = \frac{1}{\Gamma(\alpha_0)\beta^{\alpha_0}} x^{\alpha_0 - 1} e^{-x/\beta} \mathbb{1}(x > 0)$$

Exercise: Let $X_1, \ldots, X_n \overset{\text{ind}}{\sim} \text{Gamma}(\alpha_0, \beta)$, where $\beta > 0$ and α_0 is known.

1. Find a complete sufficient statistic for β.
2. Find a statistic that is ancillary to β.
3. Are these two statistics independent?

① $$f(x;\beta) = \underbrace{\frac{1}{\Gamma(\alpha_0)} x^{\alpha_0 - 1} \mathbb{1}(x > 0)}_{h(x)} \underbrace{\left(\frac{1}{\beta}\right)^{\alpha_0}}_{c(\beta)} \exp\left[x\left(-\frac{1}{\beta}\right)\right]$$

$t_1(x) = x$, $w_1(\beta) = -\frac{1}{\beta}$.

So $T(\underset{\sim}{X}) = \sum_{i=1}^{n} X_i$ is a comp. suff stat.

(2) $$f(x;\beta) = \frac{1}{\Gamma(\alpha_0)\beta^{\alpha_0}} x^{\alpha_0 - 1} e^{-x/\beta} \mathbb{1}(x>0)$$

$$= \frac{1}{\Gamma(\alpha_0)} \frac{1}{\beta} \left(\frac{x}{\beta}\right)^{\alpha_0 - 1} e^{-\frac{x}{\beta}} \mathbb{1}\left(\frac{x}{\beta} > 0\right)$$

$$= \frac{1}{\beta} f_Z\left(\frac{x}{\beta}\right), \quad f_Z(z) = \frac{1}{\Gamma(\alpha_0)} z^{\alpha_0 - 1} e^{-z} \mathbb{1}(z>0)$$

so β is a scale parameter.

propose $X_{(n)}$ as potential ancillary.

$X_i = \beta Z_i$, $Z_i \sim f_Z$, $i = 1, \dots, n$.

$X_{(n)} = \beta Z_{(n)}$

$$S_1(\underset{\sim}{X}) = \frac{X_{(1)}}{X_{(n)}}$$

$$S_2(\underset{\sim}{X}) = \frac{\bar{X}_n}{X_{(1)}}$$

both ancillary.

(3) Do we have

$$\underbrace{\frac{X_{(1)}}{X_{(n)}}}_{\text{ancill.}} \overset{?}{\perp\!\!\!\perp} \underbrace{\sum_{i=1}^{n} X_i}_{\text{comp suff}} \qquad \underbrace{\frac{\bar{X}_n}{X_{(1)}}}_{\text{ancill}} \overset{?}{\perp\!\!\!\perp} \underbrace{\sum_{i=1}^{n} X_i}_{\text{comp suff.}}$$

Yes, thanks to Basu

Exercise: Let $X_1, \ldots, X_n \overset{\text{ind}}{\sim} \text{Beta}(\theta, \theta)$, where $\theta > 0$.

1. Check sufficiency and completeness of $T(\mathbf{X}) = \prod_{i=1}^n X_i(1 - X_i)$.
2. Check whether $T'(\mathbf{X}) = (\prod_{i=1}^n X_i, \prod_{i=1}^n (1 - X_i))$ is complete, sufficient.

① $f(x;\theta) = \frac{\Gamma(2\theta)}{\Gamma(\theta)\Gamma(\theta)} x^{\theta-1}(1-x)^{\theta-1} \mathbb{1}(0<x<1)$

$= \underbrace{\frac{\Gamma(2\theta)}{\Gamma(\theta)\Gamma(\theta)}}_{c(\theta)} \underbrace{\frac{1}{x(1-x)} \mathbb{1}(0<x<1)}_{h(x)} \exp\left[\theta \log\left(x(1-x)\right)\right]$

$t_1(x) = \log(x(1-x))$

$w_1(\theta) = \theta$

So $T(\underset{\sim}{X}) = \sum_{i=1}^{n} \log(x_i(1-x_i))$ is compl. suff.

$T_2(\underset{\sim}{X}) = \prod_{i=1}^{n} x_i(1-x_i)$ is 1:1 fn of $T(\underset{\sim}{X})$.

② If $T'(\underset{\sim}{X}) = \left(\prod_{i=1}^{n} x_i , \prod_{i=1}^{n} (1-x_i)\right)$ is not min suff,

then it is not complete suff.

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

$$\frac{f(\underset{\sim}{x};\theta)}{f(\underset{\sim}{y};\theta)} = \frac{\prod_{i=1}^{n} \frac{\Gamma(2\theta)}{\Gamma(\theta)\Gamma(\theta)} x_i^{\theta-1} (1-x_i)^{\theta-1} \mathbb{1}(0<x_i<1)}{\prod_{i=1}^{n} \frac{\Gamma(2\theta)}{\Gamma(\theta)\Gamma(\theta)} y_i^{\theta-1} (1-y_i)^{\theta-1} \mathbb{1}(0<y_i<1)}$$

$$= \frac{\left(\prod_{i=1}^{n} x_i\right)^{\theta-1} \left(\prod_{i=1}^{n} (1-x_i)\right)^{\theta-1}}{\left(\prod_{i=1}^{n} y_i\right)^{\theta-1} \left(\prod_{i=1}^{n} (1-y_i)\right)^{\theta-1}}$$

~~const. in θ $\iff$ $(\prod x_i, \prod(1-x_i)) = (\prod y_i, \prod(1-y_i))$~~

So $T(\underset{\sim}{X}) = (\prod x_i, \prod(1-x_i))$

is **not min suff.**

Since it is not min suff, it cannot,

by Lehmann's theorem, be complete suff.

Theorem (Cannot construct an ancillary from a complete statistic)

If $U(\mathbf{X})$ is a known function of $T(\mathbf{X})$ with a non-degenerate distribution and $U(\mathbf{X})$ is ancillary, then $T(\mathbf{X})$ is not complete.

A *complete statistic* retains info about θ under any non-deg. transformation.

Exercise: Prove the above.

Let T be any statistic.

Suppose $U = a(T)$ is ancillary and non-degenerate (does not always take the same value)

Set $m = \mathbb{E}\,U = \mathbb{E}\,a(T)$. Note this free of θ.

Now write

$$\mathbb{E}_\theta\left(\underline{a(T) - m}\right) = \mathbb{E}_\theta\, a(T) - m$$

$$= m - m$$

$$= 0 .$$

If T is complete, then the above would imply

$$\underbrace{a(T)}_{U} - m = 0 \quad \text{w.p. } 1 \iff U = m \quad \text{w.p. } 1.$$

BUT, U is nondegenerate, so T cannot be complete.

Hopefully a useful summary:

To show that $T(\mathbf{X})$ is complete, you can:

1. Use the exponential family result (which also gives sufficiency).
2. Use the definition: Let g be any function such that $\mathbb{E}_\theta g(T(\mathbf{X})) = 0$ for all θ and show that this implies $g(T(\mathbf{X})) = 0$ with probability 1.
   (Try going this route if the support depends on the parameter).

To show that $T(\mathbf{X})$ is not complete, you can:

1. Show that a function of $T(\mathbf{X})$ is ancillary.
2. Use the definition: Let g be a function such that $\mathbb{E}_\theta g(T(\mathbf{X})) = 0$ for all θ and show that we can still have $g(T(\mathbf{X})) \neq 0$ with positive probability.

To determine the existence of a complete sufficient statistic, find a min. suff. and check if it's complete. If it is not, there exists no complete sufficient statistic.