STA261 - Module 1

Statistics

Rob Zimmerman

University of Toronto

July 6-8, 2021

Data and samples

- *Data* is factual information collected for the purposes of inference (Merriam-Webster)

- *Inference* is the act of passing from statistical sample data to generalizations (as of the value of population parameters) usually with calculated degrees of certainty (also Merriam-Webster)

- We collect a *sample* of data from a *population* associated with some probability distribution, and we would like to infer unknown properties of that distribution

- Example 1.1: - # of courses STA261 students are taking

  $\sim N(\mu, \sigma^2)$

Random variables versus observed data (this is really important)

- Our data sample goes through two phases of life: first as a *random sample*, and then as *observed data* x_i X_i
- A random sample is a set of *random variables*; observed data is a set of *constants*; the same goes for functions thereof $Z \sim N(0,1)$. $P(Z < 0) = 1/2$
- We denote random variables using uppercase letters, and constants using lowercase letters: Observe $Z = z \in \mathbb{R}$ $P(z < 0) \in \{0, 1\}$
- Example 1.2: $\vec{X} = (X_1, X_2, \ldots, X_n)$ $\vec{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$
- It is **very** important to clearly distinguish between the two quantities. But why?

$\vec{X} = (X_1, \ldots, X_n)$ $\quad X_i$ = i'th response

$\sim f_\theta$

Normal(μ, σ^2) $\quad \theta = (\mu, \sigma^2)$

$f_\theta(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$, $x \in \mathbb{R}$

"$\vec{X} \sim f_\theta$" means "$X_i \sim f_\theta$, $i = 1, \ldots, n$"

$\vec{X} \sim f_\theta$ means $f_\theta(\vec{x}) = \prod_{i=1}^{n} f_\theta(x_i)$

$Exp(\lambda) \Rightarrow \theta = \lambda \in (0, \infty)$

$X \sim Exp(\lambda) \Rightarrow$ unknown parameter is $\theta = \lambda$

iid-ness

- "iid" stands for "**independent and identically distributed**"
- This term is used everywhere in statistics, because it saves a lot of time
- Instead of "let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with pdf/pmf f_θ"

we write "let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} f_\theta$"

Ex: "Let $Z_1, Z_2, \ldots, Z_n \overset{iid}{\sim} N(\mu, 1)$"

Statistics

- Definition 1.1: A **statistic** is a function of the (random) data sample $\vec{X}$
- Example 1.3: $T(\vec{X}) = \bar{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i$ $\quad T(\vec{X}) = 100$

  $T(\vec{X}) = 24X_1^3$ $\quad T(\vec{X}) = X_{(n)} = \max\{X_1, \ldots, X_n\}$
- A statistic is useful when it allows us to summarize the data sample in ways that helps us with inference
- Different statistics are useful for different models
- Example 1.4: $X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$

  Intuitively, $T(\vec{X}) = \bar{X}_n$ is useful if we're interested in μ.
- Ex: $X_1, \ldots, X_n \overset{iid}{\sim} \text{Bernoulli}(p)$, $p \in (0,1)$.

  Then $\bar{X}_n$ is also useful for p.

Parameters and Statistical Models

- Many classical probability distributions have *parameters* associated with them
- Example 1.5: $N(\mu, \sigma^2)$ (μ, σ^2: parameters) $\quad Exp(\lambda)$ (λ: parameter) $\quad Bin(n, p)$ (n, p: parameters)
- Definition 1.2: A **statistical model** is a set of probability distributions $\{F_\theta(\cdot) : \theta \in \Theta\}$ defined on the same sample space, where each θ is a fixed **parameter** in a known **parameter space** Θ. When $\Theta \subseteq \mathbb{R}^k$ for some $k \in \mathbb{N}$, the set is also called a **parametric model** (or **parametric family**).
- Example 1.6: $\{N(\mu, 1) : \mu \in \mathbb{R}\}$ $\quad \{\lambda e^{-\lambda x} : \lambda > 0\} = \{Exp(\lambda) : \lambda > 0\}$

  $\Theta = \{\theta_1, \theta_2\} \Rightarrow \{F_\theta : \theta \in \Theta\} = \{F_{\theta_1}, F_{\theta_2}\}$
- Statistical inference is classically concerned with figuring out which one of those distributions generated the data, based on the data sample we have available
- This amounts to inferring the particular parameter θ

Parameters and Statistical Models: More Examples

- Example 1.7:

$$\{N(\mu, \sigma^2) : \mu \in \mathbb{R},\ \sigma^2 > 0\}$$

$$= \{N(\mu, \sigma^2) : (\mu, \sigma^2) \in \mathbb{R} \times (0, \infty)\}$$

Maybe instead, $\{N(\mu, \sigma^2) : \mu \in (0, \infty),\ \sigma^2 \in (0, \infty)\}$

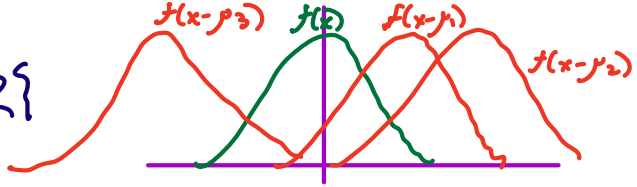
if we <u>know</u> that $\mu > 0$.

Important Parametric Families: Location-Scale Families

- Definition 1.3: A **location family** is a family of distributions $\{F(x-\mu) : \mu \in \mathbb{R}\}$ formed by translating a "standard" family member $F(\cdot)$.

  $\{f(x-\mu) : \mu \in \mathbb{R}\}$ ← pdf version

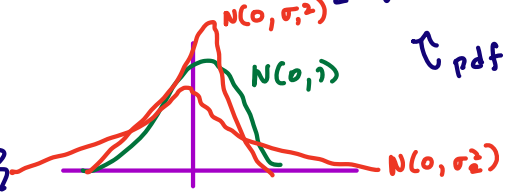
- Example 1.8: $\{N(\mu, 1) : \mu \in \mathbb{R}\}$



- Definition 1.4: A **scale family** is a family of distributions $\{F(x/\sigma) : \sigma > 0\}$ formed by rescaling a "standard" family member $F(x)$.

  $\{\frac{1}{\sigma} f(x/\sigma) : \sigma > 0\}$ ↑ pdf version

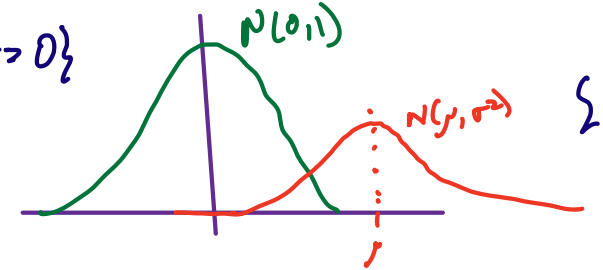
- Example 1.9: $\{N(0, \sigma^2) : \sigma^2 > 0\}$, $\{N(20, \sigma^2) : \sigma^2 > 0\}$, $\{Exp(\lambda) : \lambda > 0\}$



- Definition 1.5: A **location-scale family** is a family of distributions $\{F(\frac{\cdot-\mu}{\sigma}) : \mu \in \mathbb{R}, \sigma > 0\}$ formed by translating and rescaling a "standard" family member $F(\cdot)$.

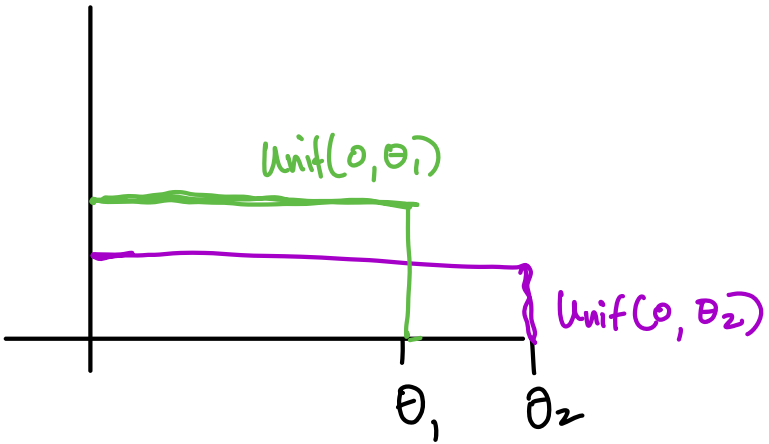
  $\{\frac{1}{\sigma} f(\frac{x-\mu}{\sigma}) : \mu \in \mathbb{R}, \sigma^2 > 0\}$ ↑ pdf version

- Example 1.10: $\{N(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\}$



Poll Time!

$N(\mu, 1) \Rightarrow f_\mu(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-(x-\mu)^2/2\right)$



$\{\text{Unif}(\theta, 2+\theta) : \theta \in \mathbb{R}\}$ is a location family w/ "standard" Unif(0,2).

Important Parametric Families: Exponential Families

- Definition 1.6: An **exponential family** is a parametric family of pdfs/pmfs of the form

$$f_\theta(x) = h(x) \cdot g(\theta) \cdot \exp\left(\sum_{j=1}^{k} w_j(\theta) \cdot T_j(x) \right),$$

where all functions of x and θ are *known*.

generic function of θ ✓ generic function of x ✓

If $\theta \in \mathbb{R}$, then usually just $\exp(w(\theta) \cdot T(x))$

eg, $T_j(x)$ eg, $w_j(\theta)$

- Lots of theory simplifies considerably if we assume our random sample comes from an exponential family

Ex: show in exp families

- Many of your favourite distributions are included

$\text{Bin}(n,p)$ n known — $\text{Poisson}(\lambda)$
$N(\mu, \sigma^2)$ — $\text{Multinomial}(n, p_1, \dots, p_n)$ n known
$\text{Gamma}(\alpha, \beta)$ — $\chi^2_{(n)}$
$\text{Beta}(\alpha, \beta)$

- Example 1.11:

$X \sim \text{Exp}(\lambda)$

$f_\lambda(x) = \lambda e^{-\lambda x}$

$= 1 \cdot \lambda \cdot \exp(x \cdot (-\lambda))$

$h(x)$ $g(\lambda)$ $T(x)$ $w(\lambda)$

Bernoulli(p): $f_\theta(x) = \theta^x (1-\theta)^{1-x}$

$x = \exp(\log(x))$

$= (1-\theta) \cdot \left(\frac{\theta}{1-\theta}\right)^x$

$= (1-\theta) \cdot \exp\left(x \cdot \log\left(\frac{\theta}{1-\theta}\right)\right)$

$h(x)=1$ $g(\theta)=1-\theta$ $T(x)=x$ $w(\theta)=\log\left(\frac{\theta}{1-\theta}\right)$

A Quick Review of Conditional Distributions

- $X|Y$ is a random variable, which has its own distribution called a conditional distribution
- Remember Bayes' rule: $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}, \quad \mathbb{P}(B) > 0$
- $X \mid Y = y$ $\quad f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$
- $X \mid X = x$ $\sim$ degenerate @ x
- Example 1.12: $\mathbb{E}[X|Y]$ is a random variable (random function of Y)
- Example 1.13: $\mathbb{E}[X|X] = X$. If $X \perp\!\!\!\perp Y$, $\mathbb{E}[X|Y] = \mathbb{E}[X]$. ($\perp\!\!\!\perp$ "independent")

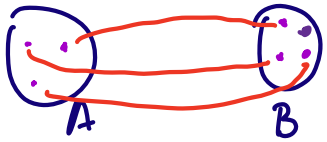
"Tower property": $\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|Y]]$

"Conditional variance": $\mathrm{Var}(X) = \mathbb{E}[\mathrm{Var}(X|Y)] + \mathrm{Var}(\mathbb{E}[X|Y])$ (Exercise)

A Quick Review of Functions

- Let $f : A \to B$ be a function
- If f is one-to-one, then $f(x) = f(y) \iff x = y$

  (injective)



- If f is onto, then

  (surjective) $\forall b \in B, \exists a \in A$ s.t. $f(a) = b$
- If f is a bijection, then

  f is one-to-one AND onto
- Example 1.14:

Freedom From θ

- Most of the functions $f_\theta(x)$ we will deal with have parameters involved in addition to the "independent variable"

- If the parameter θ can vary too, then $f_\theta(x)$ is really a function of both x and θ $\quad f_\theta(x) = g(\theta, x)$

- If $f_\theta(x)$ is actually *not* a function of θ (i.e., it doesn't change with θ), we might also say that it's "free of θ" or that it "does not depend on θ"

$f_\theta(x) = x^2$ free of θ

- Example 1.15: $f_\theta(x) = \theta x^2$, not free of θ

Say $X \sim N(\mu, 1)$, then $P_\mu(X - \mu \le x) = \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt$

- So if we say that the distribution of X is free of θ, we mean that the cdf of X (and hence the pdf/pmf) is the same for all $\theta \in \Theta$

- Example 1.16: $X \sim Exp(\lambda)$, the dist of λX is free of λ

Data Reduction: A Thought Experiment

- Is there a such thing as "more data than necessary"?
- Suppose that field researchers collect a sample $\mathbf{X} = (X_1, X_2, \ldots, X_n) \overset{iid}{\sim} f_\theta$, where n is astronomically large; they want us statisticians to do inference on θ, but sending us $\mathbf{X}$ would take weeks
- Wouldn't it be great if we didn't need the entire sample $\mathbf{X}$ to make inferences about θ, but rather a much smaller statistic $T(\mathbf{X})$ – perhaps just a single number – that still contained as much information about θ as $\mathbf{X}$ itself did?
- The researchers observe $\mathbf{X} = \mathbf{x}$, calculate $T(\mathbf{x}) = t$ on their end, and then text t over to us
- Example 1.17: $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, 1)$

  Instead $(X_1, \ldots, X_n)$, what if we could get away with $T(\vec{X}) = \bar{X}_n$?

Sufficiency

- How do we "encode" this idea?
- If we know that $T(\mathbf{X}) = t$, then there should be nothing else to glean from the data about θ
- Definition 1.7: A statistic $T(\mathbf{X})$ is a **sufficient statistic** for a parameter θ if the conditional distribution of $\mathbf{X} \mid T(\mathbf{X}) = t$ does not depend on θ.
- An interpretation: if the conditional distribution

$$\mathbb{P}(\mathbf{X} = \mathbf{x} \mid T(\mathbf{X}) = \overbrace{T(\mathbf{x})}^{t}) = \frac{\mathbb{P}_\theta(\mathbf{X} = \mathbf{x})}{\mathbb{P}_\theta(T(\mathbf{X}) = T(\mathbf{x}))}$$

  is really free of θ, then the information about θ in $\mathbf{X}$ and the information about θ in $T(\mathbf{X})$ are "equal"

- Example 1.18: $T(\vec{X}) = \vec{X}$. Then $P_\theta(\vec{X} = \vec{x} \mid \vec{X} = \vec{x}) = 1$

  Trivial; no data reduction!

Sufficiency

- Example 1.19: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Bernoulli}(\theta)$, where $\theta \in (0,1)$. Show that $T(\mathbf{X}) = \sum_{i=1}^{n} X_i$ is sufficient for θ.

$T(\vec{X}) \sim Bin(n,\theta) \Rightarrow P_\theta(T(\vec{X}) = t) = \binom{n}{t}\theta^t(1-\theta)^{n-t}$

$P_\theta(X_1 = x_1, \ldots, X_n = x_n \cap \sum_{i=1}^{n} X_i = t)$

$= P_\theta(X_1 = x_1, \ldots, X_n = x_n, \sum_{i=1}^{n} x_i = t)$ $\rightsquigarrow x_n = t - \sum_{i=1}^{n-1} x_i$

$= P_\theta(X_1 = x_1, \ldots, X_n = t - \sum_{i=1}^{n-1} x_i)$

$\overset{\text{indep}}{=} P_\theta(X_1 = x_1) \cdot \ldots \cdot P_\theta(X_n = t - \sum_{i=1}^{n-1} x_i)$

$= \theta^{x_1}(1-\theta)^{1-x_1} \cdots \theta^{x_{n-1}}(1-\theta)^{1-x_{n-1}} \cdot \theta^{t-\sum_{i=1}^{n-1} x_i}(1-\theta)^{1-t+\sum_{i=1}^{n-1} x_i}$

$= \theta^{\sum_{i=1}^{n-1} x_i + t - \sum_{i=1}^{n-1} x_i}(1-\theta)^{n-\sum_{i=1}^{n-1} x_i - t + \sum_{i=1}^{n-1} x_i}$

$= \theta^t(1-\theta)^{n-t}$

$\therefore P_\theta(\vec{X} = \vec{x} \mid T(\vec{X}) = t) = \dfrac{\theta^t(1-\theta)^{n-t}}{\binom{n}{t}\theta^t(1-\theta)^{n-t}} = \dfrac{1}{\binom{n}{t}}$, free of θ.

$\therefore T(\vec{X})$ is sufficient for θ.

Sufficiency

- Example 1.20: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$, where $\mu \in \mathbb{R}$ and σ^2 is known. Show that the sample mean $T(\mathbf{X}) = \bar{X}_n := \frac{1}{n}\sum_{i=1}^n X_i$ is sufficient for μ.

$\bar{x} \sim N(\mu, \sigma^2/n)$ so $f_T(t) = \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} \exp\left(\frac{-n(t-\mu)^2}{2\sigma^2}\right)$

Need joint distn of $(\vec{X}, T)$.

Observe $\sum (x_i-\mu)^2$
$= \sum (x_i - t + t - \mu)$
$= \sum \left[(x_i-t)^2 + 2(x_i-t)(t-\mu) + (t-\mu)^2\right]$
$= \sum (x_i-t)^2 + n(t-\mu)^2$

Hence $f_{\vec{X}}(\vec{x}) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-(x_i-\mu)^2}{2\sigma^2}\right)$
$= (2\pi\sigma)^{-n/2} \cdot \exp\left(-\frac{\sum (x_i-\mu)^2}{2\sigma^2}\right)$
$= (2\pi\sigma)^{-n/2} \cdot \exp\left(-\frac{\sum (x_i-t)^2 - n(t-\mu)^2}{2\sigma^2}\right)$
$= f_{\vec{X},T}(\vec{x}, t)$

So $f_{\vec{X}|T}(\vec{x}|t)$
$= \frac{f_{\vec{X},T}(\vec{x}, t)}{f_T(t)}$
$= \frac{1}{\sqrt{n}(2\pi\sigma^2)^{\frac{n-1}{2}}} \cdot \exp\left(-\frac{\sum (x_i-t)^2}{2\sigma^2}\right)$

which is free of μ.

Hence, $T(\vec{X})$ is sufficient for μ.

The Factorization Theorem

- Theorem 1.1 (**Factorization theorem**): Let $\mathbf{X} = (X_1, \ldots, X_n) \sim f_\theta(\mathbf{x})$, where $f_\theta(\mathbf{x})$ is a joint pdf/pmf. A statistic $T(\mathbf{X})$ is sufficient for θ if and only if there exist functions $g_\theta(t)$ and $h(\mathbf{x})$ such that

$$f_\theta(\mathbf{x}) = h(\mathbf{x}) \cdot g_\theta(T(\mathbf{x})) \quad \text{for all } \theta \in \Theta,$$

  where $h(\mathbf{x})$ is free of θ and $g_\theta(T(\mathbf{x}))$ only depends on $\mathbf{x}$ through $T(\mathbf{x})$.

- In other words, $T(\mathbf{X})$ is sufficient whenever the "part" of $f_\theta(\mathbf{x})$ that actually depends on θ is a function of $T(\mathbf{x})$, rather than $\mathbf{x}$ itself

Proof. (Discrete case). WTS $\dfrac{P_\theta(\vec{X} = \vec{x} \cap T(\vec{x}) = t)}{P_\theta(T(\vec{x}) = t)}$ free of θ

iff $P_\theta(\vec{X} = \vec{x}) = h(\vec{x}) \cdot g_\theta(t)$.

The Factorization Theorem

($\Rightarrow$) Assume T is sufficient.

$$P_\theta(\vec{X}=\vec{x}) = P_\theta(\vec{X}=\vec{x} \cap T(\vec{x})=t)$$
$$= P_\theta(\vec{X}=\vec{x} \mid T(\vec{x})=t) \cdot P_\theta(T(\vec{x})=t)$$
$$= \underbrace{P(\vec{X}=\vec{x} \mid T(\vec{x})=t)}_{h(\vec{x})} \cdot \underbrace{P_\theta(T(\vec{x})=t)}_{g_\theta(t)}$$

($\Leftarrow$) Assume $P_\theta(\vec{X}=\vec{x}) = h(\vec{x}) \cdot g_\theta(t)$.

Then $f_\theta(t) = \sum_{\vec{x}: T(\vec{x})=t} f_\theta(\vec{x}, t)$

$$= \sum_{\vec{x}: T(\vec{x})=t} f_\theta(\vec{x})$$
$$= \sum h(\vec{x}) \cdot g_\theta(t)$$
$$= \left(\sum_{\vec{x}: T(\vec{x})=t} h(\vec{x})\right) \cdot g_\theta(t)$$

Hence $f_\theta(\vec{x} \mid t)$

$$= \frac{f_\theta(\vec{x}, t)}{f_\theta(t)}$$
$$= \frac{f_\theta(\vec{x})}{f_\theta(t)}$$
$$= \frac{h(\vec{x}) \cdot g_\theta(t)}{\left(\sum_{\vec{x}: T(\vec{x})=t} h(\vec{x})\right) \cdot g_\theta(t)}$$
$$= \frac{h(\vec{x})}{\sum_{\vec{x}: T(\vec{x})=t} h(\vec{x})}$$ free of θ.

So $T(\vec{X})$ is sufficient for θ. $\square$

Poll Time!

$$f_0(x) = h(x) \cdot g_0(x)$$

The Factorization Theorem: Examples

- Example 1.21: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Bernoulli}(\theta)$, where $\theta \in (0,1)$. Show that $T(\mathbf{X}) = \sum_{i=1}^n X_i$ is sufficient for θ.

Let $t = \sum x_i$

$$f_\theta(\vec{x}) = \prod_{i=1}^n \theta^{x_i}(1-\theta)^{1-x_i}$$

$$= \theta^{\sum x_i}(1-\theta)^{n-\sum x_i}$$

$$= 1 \cdot \theta^{\sum x_i}(1-\theta)^{n-\sum x_i}$$

$$= \underbrace{1}_{h(\vec{x})} \cdot \underbrace{\theta^t(1-\theta)^{n-t}}_{g_\theta(t)}$$

By the Factorization Theorem, $T(\vec{x})$ is sufficient for θ.

- Example 1.22: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$, where $\mu \in \mathbb{R}$ and σ^2 is known. Show that the sample mean $T(\mathbf{X}) = \bar{X}_n := \frac{1}{n}\sum_{i=1}^n X_i$ is sufficient for μ.

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

$$f_\mu(\vec{x}) = \prod_{i=1}^n f_\mu(x_i)$$

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

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

$$= \underbrace{\frac{1}{(2\pi\sigma^2)^{n/2}} \cdot \exp\left(-\frac{\sum (x_i-t)^2}{2\sigma^2}\right)}_{h(\vec{x})} \cdot \underbrace{\exp\left(-\frac{n(t-\mu)^2}{2\sigma^2}\right)}_{g_\mu(t)}$$

$\sum (x_i-\mu)^2 = \sum (x_i^2 - 2\mu x_i + \mu^2)$

By the Factorization Theorem, $T(\vec{x}) = \frac{1}{n}\sum x_i$ is sufficient for μ.

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

The Factorization Theorem: Examples

- Example 1.23: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$, where $\mu \in \mathbb{R}$ and $\sigma^2 > 0$. Define the **sample variance** as $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$. Show that $T(\mathbf{X}) = (\bar{X}_n, S^2)$ is sufficient for (μ, σ^2).

Let $t_1 = \frac{1}{n}\sum x_i$, $t_2 = \frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x})^2$

$$f_{\mu,\sigma^2}(\vec{x}) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(- \frac{\sum (x_i - \bar{x})^2 - n(\bar{x} - \mu)^2}{2\sigma^2} \right)$$

$$= \underbrace{\frac{1}{(2\pi\sigma^2)^{n/2}} \cdot \exp\left(\frac{-(n-1)t_2 - n(t_1 - \mu)^2}{2\sigma^2} \right)}_{g_{\mu,\sigma^2}(t_1, t_2)} \cdot \underbrace{1}_{h(\vec{x})}.$$

By the FT, $(\bar{X}_n, S^2)$ is sufficient for (μ, σ^2).

- Example 1.24: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Unif}(0, \theta)$ where $\theta > 0$. ~~Show that~~ $\bar{X}_n$ is not sufficient for θ, ~~and~~ find a statistic that is.

Fact: $0 \le x_i \le \theta \ \forall i \iff 0 \le x_{(1)} \wedge x_{(n)} \le \theta$

$$f_\theta(\vec{x}) = \prod_{i=1}^n \frac{1}{\theta} \cdot \mathbb{1}_{0 \le x_i \le \theta}$$

$$= \theta^{-n} \cdot \mathbb{1}_{0 \le x_1 \le \theta \wedge 0 \le x_2 \le \theta \wedge \cdots \wedge 0 \le x_n \le \theta}$$

$\leftarrow$ no factorization in terms of $\frac{1}{n}\sum x_i$

$\mathbb{1}_A = \begin{cases} 1, & A \text{ is true} \\ 0, & A \text{ is not true} \end{cases}$

$$= \underbrace{\mathbb{1}_{0 \le x_{(1)}}}_{h(\vec{x})} \cdot \underbrace{\theta^{-n} \cdot \mathbb{1}_{x_{(n)} \le \theta}}_{g_\theta(x_{(n)})}$$

By the FT, $T(\vec{x}) = X_{(n)}$ is sufficient for θ.

The Factorization Theorem: Examples

- Theorem 1.2: Let $X_1, \ldots, X_n \overset{iid}{\sim} f_\theta$ be a random sample from an exponential family, where

$$f_\theta(x) = h(x) \cdot g(\theta) \cdot \exp\left(\sum_{j=1}^{k} w_j(\theta) \cdot T_j(x) \right).$$

$w(\theta) \cdot T(x) \rightarrow T(\vec{x}) = \sum_{i=1}^{n} T(x_i)$

Then $T(\mathbf{X}) = \left(\sum_{i=1}^{n} T_1(X_i), \ldots, \sum_{i=1}^{n} T_k(X_i) \right)$ is sufficient for θ. suff for θ.

Proof. $f_\theta(\vec{x}) = \prod_{i=1}^{n} h(x_i) \cdot g(\theta) \cdot \exp\left(\sum_{i=1}^{n} \left(\sum_{j=1}^{k} w_j(\theta) \cdot T_j(x_i) \right) \right)$

$= \left(\prod_{i=1}^{n} h(x_i) \right) \cdot g(\theta)^n \cdot \exp\left(\sum_{j=1}^{n} w_j(\theta) \cdot \left[\sum_{i=1}^{n} T_j(x_i) \right] \right)$ let $t_j = \sum_{i=1}^{n} T_j(x_i)$

$= \left(\prod_{i=1}^{n} h(x_i) \right) \cdot g(\theta)^n \cdot \exp\left(\sum_{j=1}^{n} w_j(\theta) \cdot t_j \right)$

$h(\vec{x})$ $g_\theta(t_1, \ldots, t_k)$

By the FT, $T(\vec{x}) = \left(\sum_{i=1}^{n} T_1(x_i), \ldots, \sum_{i=1}^{n} T_k(x_i) \right)$ is sufficient for θ. □

The Factorization Theorem: Examples

- Example 1.25: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$, where $\mu \in \mathbb{R}$ and $\sigma^2 > 0$. Show that $T(\mathbf{X}) = (\sum_{i=1}^n X_i^2, \sum_{i=1}^n X_i)$ is sufficient for $(\mu, \sigma^2) =: \theta$.

Exp family: $f_\theta(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$

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

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

$h(x)$ | $g(\theta)$ | $w_1(\theta)$ | $w_2(\theta)$ | $T_1(x)$ | $T_2(x)$

By Theorem 1.2, $T(\vec{X}) = \left(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2\right)$ is sufficient for $\Theta = (\mu, \sigma^2)$. □

The Factorization Theorem: Examples

- Example 1.26: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Unif}(\{1, 2, \ldots, \theta\})$, where $\theta \in \mathbb{N}$. Show that $T(\mathbf{X}) = X_{(n)}$ is sufficient for θ.

$$f_\theta(\vec{x}) = \prod_{i=1}^{n} f_\theta(x_i)$$

$$= \prod_{i=1}^{n} \frac{1}{\theta} \cdot \mathbb{1}_{x_i \in \{1,2,\ldots,\theta\}}$$

$$= \theta^{-n} \cdot \mathbb{1}_{x_i \in \{1,2,\ldots,\theta\}\ \forall i}$$

$$= \theta^{-n} \cdot \mathbb{1}_{x_i \in \mathbb{N}\ \forall i \text{ and } x_{(n)} \le \theta}$$

$$= \underbrace{\mathbb{1}_{x_i \in \mathbb{N}\ \forall i}}_{h(\vec{x})} \cdot \underbrace{\theta^{-n} \cdot \mathbb{1}_{x_{(n)} \le \theta}}_{g_\theta(x_{(n)})}.$$

By the Factorization Theorem, $T(\vec{x}) = X_{(n)}$ is sufficient for θ.

If There's One, There's More...

- If we have some sufficient statistic, we can always come up with (infinitely) many others...

- Theorem 1.3: Let $T(\mathbf{X})$ be sufficient for θ and suppose that $r(\cdot)$ is a bijection. Then $r(T(\mathbf{X}))$ is also sufficient for θ.

Proof. If $T(\vec{X})$ is sufficient for θ, then by FT,

$$f_\theta(\vec{x}) = h(\vec{x}) \cdot g_\theta(T(\vec{x})) \quad \text{for some } h(\cdot), g_\theta(\cdot).$$

$$= h(\vec{x}) \cdot g_\theta(r^{-1}(r(T(\vec{x}))))$$

$$= h(\vec{x}) \cdot \tilde{g}_\theta(T(\vec{x})) \quad \text{where } \tilde{g}_\theta(t) = g_\theta(r^{-1}(t)).$$

By the FT, $r(T(\vec{X}))$ is sufficient for θ. □

Too Many Sufficient Statistics

- So there are lots of sufficient statistics out there
- We saw that $T(\mathbf{X}) = \mathbf{X}$ is always sufficient – it's also pretty useless as far as data reduction goes
- There are usually "better" ones out there – how do we get the best bang for our buck?
- Another issue: the factorization theorem makes it easy to show that a statistic is sufficient (if it actually is), but less so to show that a statistic is *not* sufficient
- We will develop theory that takes care of both of these issues at once

Minimal Sufficiency

- Definition 1.8: A sufficient statistic $T(\mathbf{X})$ is called a **minimal sufficient statistic** if, for any other sufficient statistic $U(\mathbf{X})$, there exists a function h such that $T(\mathbf{X}) = h(U(\mathbf{X}))$.

- In other words, a minimal sufficient statistic is some function of *any other sufficient statistic*
  $N(\mu, 1)$: $\vec{X}$ is sufficient for μ
  $\bar{X}_n$ is also sufficient for μ $\quad \bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$
- A minimal sufficient statistic achieves the greatest reduction of data possible (while still maintaining sufficiency)

- Example 1.27: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$, where $\mu \in \mathbb{R}$ and σ^2 is known. Show that $T(\mathbf{X}) = (\bar{X}_n, S^2)$ is not minimal sufficient for μ.

  We saw $\bar{X}_n$ is sufficient for μ. But $T(\vec{X})$ is <u>not</u> a function of $\bar{X}_n$. Hence $T(\vec{X})$ cannot be minimal sufficient for μ.

  (Eg: if it were, say $h(\bar{X}) = (\bar{X}, S^2)$. Then we'd have, for $S^2 \in \{1, 2\}$, $h(t) = (t, 1)$ and $h(t) = (t, 2)$ — not a function!)

Poll Time!

$$T(\vec{x}) - X_n = X_1 + X_2 + \cdots + X_{n-1}$$

$$(T(\vec{x}),\ T(\vec{x}) + 2c_1)$$

$\downarrow$ choose first coordinate

$$T(\vec{x})$$

$\downarrow$

$$(T(\vec{x}),\ T(\vec{x}) + 2c_1)$$

A Criterion For Minimal Sufficiency

- It's usually not that hard to show that a statistic is not minimal sufficient
- But how can we possibly show that a statistic *is* minimal?
- Theorem 1.4: Let $f_\theta(\mathbf{x})$ be the pdf/pmf of a sample $\mathbf{X}$. Suppose there exists a function $T(\cdot)$ such that for any $\mathbf{x}, \mathbf{y} \in \mathcal{X}^n$, $T(\mathbf{x}) = T(\mathbf{y})$ if and only if the ratio $f_\theta(\mathbf{x})/f_\theta(\mathbf{y})$ is free of θ. Then $T(\mathbf{X})$ is minimal sufficient for θ.
- This criterion is easier to apply than it looks

$T(\vec{x}) = \sum_{i=1}^n x_i$ $(\vec{x} = (x_1, x_2, \ldots, x_n))$.

- Example 1.28: $X_1, \ldots, X_n \overset{iid}{\sim} \text{Bernoulli}(\theta)$, $\theta \in (0,1)$. Show $T(\vec{X}) = \sum_{i=1}^n X_i$ is min suff for θ.

We showed that $f_\theta(\vec{x}) = \theta^{\sum x_i}(1-\theta)^{n-\sum x_i}$. Let $\vec{x}, \vec{y} \in \mathcal{X}^n$. Then

$$\frac{f_\theta(\vec{x})}{f_\theta(\vec{y})} = \frac{\theta^{\sum x_i}(1-\theta)^{n-\sum x_i}}{\theta^{\sum y_i}(1-\theta)^{n-\sum y_i}} = \theta^{\sum x_i - \sum y_i}(1-\theta)^{\sum y_i - \sum x_i}$$

free of θ iff $\sum x_i = \sum y_i$ (i.e., $T(\vec{x}) = T(\vec{y})$).

By Theorem 1.4, $T(\vec{X})$ is minimal sufficient for θ.

Minimal Sufficiency: Examples

- Example 1.29: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$, where $\mu \in \mathbb{R}$ and $\sigma^2 > 0$. Show that $T(\mathbf{X}) = (\bar{X}, S^2)$ is minimal sufficient for $(\mu, \sigma^2) =: \Theta$.

Let $s_x^2 = \frac{1}{n-1}\sum (x_i - \bar{x})^2$.

From Ex 1.23,

$$f_\theta(\vec{x}) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(\frac{-(n-1)s_x^2 - n(\bar{x}-\mu)^2}{2\sigma^2}\right).$$

Let $\vec{x}, \vec{y} \in \mathcal{X}^n$. Then

$$\frac{f_\theta(\vec{x})}{f_\theta(\vec{y})} = \frac{\frac{1}{(2\pi\sigma^2)^{n/2}} \cdot \exp\left(\frac{-(n-1)s_x^2 - n(\bar{x}-\mu)^2}{2\sigma^2}\right)}{\frac{1}{(2\pi\sigma^2)^{n/2}} \cdot \exp\left(\frac{-(n-1)s_y^2 - n(\bar{y}-\mu)^2}{2\sigma^2}\right)}$$

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

is free of (μ, σ^2) iff $\bar{x} = \bar{y}$ and $s_x^2 = s_y^2$.

By Theorem 1.4, $T(\vec{x})$ is minimal sufficient for (μ, σ^2).

Minimal Sufficiency: Examples

- Example 1.30: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim}$ Poisson (λ), where $\lambda > 0$. Find a minimal sufficient statistic for λ.

$f_\lambda(\vec{x}) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} = \frac{1}{\prod_{i=1}^{n} x_i!} e^{-n\lambda} \lambda^{\sum x_i}$. So $T(\vec{X}) = \sum X_i$ is sufficient for λ, by the FT.

Let $\vec{x}, \vec{y} \in \mathcal{X}^n$.

Then $\frac{f_\lambda(\vec{x})}{f_\lambda(\vec{y})} = \frac{\frac{1}{\prod x_i!} \cdot e^{-n\lambda} \cdot \lambda^{\sum x_i}}{\frac{1}{\prod y_i!} \cdot e^{-n\lambda} \cdot \lambda^{\sum y_i}} = \left(\frac{\prod y_i!}{\prod x_i!} \right) \cdot \lambda^{\sum x_i - \sum y_i}$

is free of λ iff $\sum x_i = \sum y_i$.

By Theorem 1.4, $T(\vec{X})$ is minimal sufficient for λ.

Minimal Sufficiency: Examples

- A minimal sufficient statistic isn't always as minimal as you would expect...

- Example 1.31: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Unif}([\theta, \theta+1])$, where $\theta \in \mathbb{R}$. Show that $T(\mathbf{X}) = (X_{(1)}, X_{(n)})$ is minimal sufficient for θ.

$$f_\theta(\vec{x}) = \prod_{i=1}^n f_\theta(x_i) = \prod_{i=1}^n \mathbb{1}_{\theta \le x_i \le \theta+1}$$

$$= \mathbb{1}_{\theta \le x_1, \ldots, x_n \le \theta+1}$$

$$= \mathbb{1}_{\theta \le x_{(1)} \wedge x_{(n)} \le \theta+1}$$

$$= \mathbb{1}_{x_{(n)}-1 \le \theta \le x_{(1)}}.$$

Let $\vec{x}, \vec{y} \in \mathcal{X}^n$. Then $\dfrac{f_\theta(\vec{x})}{f_\theta(\vec{y})} = \dfrac{\mathbb{1}_{x_{(n)}-1 \le \theta \le x_{(1)}}}{\mathbb{1}_{y_{(n)}-1 \le \theta \le y_{(1)}}}$ is constant with respect to θ iff $x_{(1)} = y_{(1)}$ and $x_{(n)} = y_{(n)}$.

Hence $T(\vec{x})$ is minimal sufficient for θ, by Theorem 1.4.

Poll Time!

The "Opposite" of Sufficiency?

- We know that a sufficient statistic contains all the information about θ that the original sample has
- What about a statistic that contains *no* information about θ?
- Why would such a thing be useful?

Age in country i follows $N(\mu_i, 1)$.
Sample $X_i \sim N(\mu_i, 1)$, calculate $D_i(X_i)$.
If D_i contains no info about μ_i, then all D_i's should look similar.

- Definition 1.9: A statistic $D(\mathbf{X})$ is an **ancillary statistic** for a parameter θ if the distribution of $D(\mathbf{X})$ does not depend on θ

Eg, $X_i \sim N(\mu_i, 1)$.

$D_i = D_i(X_i) = X_i - \mu_i \sim N(0,1)$

free of μ_i.

Ancillarity

- Definition 1.10: A statistic $D(\mathbf{X})$ is an **ancillary statistic** for a parameter θ if the distribution of $D(\mathbf{X})$ does not depend on θ

- Example 1.32: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Unif}([\theta, \theta+1])$, where $\theta \in \mathbb{R}$. Show that the range statistic $R(\mathbf{X}) := X_{(n)} - X_{(1)}$ is ancillary for θ.

Let $Y_i = X_i - \theta$. Then $Y_i \sim \text{Unif}(0,1)$, and $Y_{(j)} = X_{(j)} - \theta$ also free of θ.

Then $P_\theta(R(\vec{X}) \le r)$

$= P_\theta(X_{(n)} - X_{(1)} \le r)$

$= P_\theta([X_{(n)} - \theta] - [X_{(1)} - \theta] \le r)$

$= P_\theta(Y_{(n)} - Y_{(1)} \le r)$

$= P(\text{Beta}(n,1) - \text{Beta}(1,n) \le r)$ is free of θ.

Hence $R(\vec{X})$ is free of θ.

Ancillarity: Examples

- Did we actually use the uniform distribution anywhere in the previous example?

In other words: the range statistic is always ancillary for a location parameter.

- Theorem 1.5: Let $X_1, \ldots, X_n$ be a random sample from a location family with cdf $F(x-\theta)$, for $\theta \in \mathbb{R}$. Then the range statistic is ancillary for θ

$R(\vec{X}) := X_{(n)} - X_{(1)}$.

Proof. Let $Y_i = X_i - \theta \sim F(x)$

Then $P_\theta(R(\vec{X}) \le r)$

$= P_\theta(X_{(n)} - X_{(1)} \le r)$

$= P_\theta([X_{(n)} - \theta] - [X_{(1)} - \theta] \le r)$

$= P_0(Y_{(n)} - Y_{(1)} \le r)$ is free of θ, since the distribution of $Y_{(n)}$ and $Y_{(1)}$ is free of θ.

□

Ancillarity: Examples

- Example 1.33: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathcal{N}\left(0, \sigma^2\right)$. Show that $D(\mathbf{X}) = \frac{X_1 + \cdots + X_{n-1}}{X_n}$ is ancillary for σ^2.

Let $Z_i = \frac{X_i}{\sigma} \sim N(0,1)$.

$$P_{\sigma^2}\left(\frac{X_1 + \cdots + X_{n-1}}{X_n} \leq x \right)$$

$$= P_{\sigma^2}\left(\frac{X_1}{X_n} + \cdots + \frac{X_{n-1}}{X_n} \leq x \right)$$

$$= P_{\sigma^2}\left(\frac{X_1/\sigma}{X_n/\sigma} + \cdots + \frac{X_{n-1}/\sigma}{X_n/\sigma} \leq x \right)$$

$$= P_{\sigma^2}\left(\frac{Z_1}{Z_n} + \cdots + \frac{Z_{n-1}}{Z_n} \leq x \right) \text{ is free of } \sigma^2.$$

Hence $D(\vec{X})$ is ancillary for σ^2.

- Theorem 1.6: Let $X_1, \ldots, X_n$ be a random sample from a scale family with cdf $F(\cdot/\sigma)$, for $\sigma > 0$. Then any statistic which is a function of the ratios $X_1/X_n, \ldots, X_{n-1}/X_n$ is ancillary for σ.

EXERCISE!

Ancillarity: Examples

- Recall that if $Z_1, \ldots, Z_n \overset{iid}{\sim} \mathcal{N}(0,1)$, then the distribution of $Y = \sum_{i=1}^{n} Z_i^2$ is called a **chi-squared distribution with n degrees of freedom**, which we write as $Y \sim \chi^2_{(n)}$.

- Theorem 1.7: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$ with $\mu \in \mathbb{R}$ and $\sigma^2 > 0$. Then $\frac{n-1}{\sigma^2} S^2 \sim \chi^2_{(n-1)}$.

Proof $(n = 2)$.

$(n-1)S^2 = \sum_{i=1}^{2} (X_i - \bar{X})^2 = \left(X_1 - \frac{1}{2}(X_1+X_2)\right)^2 + \left(X_2 - \frac{1}{2}(X_1+X_2)\right)^2$

$= \left(\frac{1}{2}X_1 - \frac{1}{2}X_2\right)^2 + \left(\frac{1}{2}X_2 - \frac{1}{2}X_1\right)^2$

$= \frac{1}{2}(X_1 - X_2)^2$

$\overset{d}{=} \frac{1}{2} \cdot 2\sigma^2 \cdot \chi^2_{(1)} \overset{d}{=} \sigma^2 \cdot \chi^2_{(1)}$

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

$X_1 - X_2 \sim N(0, 2\sigma^2)$

$\overset{d}{=} \sqrt{2}\sigma \cdot N(0,1)$

$\Rightarrow (X_1 - X_2)^2 \overset{d}{=} 2\sigma^2 \cdot N(0,1)^2$

$= 2\sigma^2 \cdot \chi^2_{(1)}$

- Example 1.34: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$ with $\mu \in \mathbb{R}$ and $\sigma^2 > 0$. Show that the sample variance S^2 is ancillary for μ.

We showed $S^2 \sim \frac{\sigma^2}{n-1} \chi^2_{(n-1)}$, free of μ. So it's ancillary for μ.

Poll Time!

Completeness: An Abstract Definition

- Everything so far has been about ways to reduce the amount of data we need while still retaining all information about θ
- We've seen that ancillary statistics are bad at it, sufficient statistics are good at it, and minimal sufficient statistics are very good at it
- We will study one more kind of statistic, but the definition isn't pretty
- Definition 1.11: A statistic $U(\mathbf{X})$ is **complete** if *any* function $h(\cdot)$ which satisfies $\mathbb{E}_\theta\left[h(U(\mathbf{X}))\right] = 0$ for all $\theta \in \Theta$ must also satisfy $\mathbb{P}_\theta\left(h(U(\mathbf{X})) \neq 0\right) = 1$ for all $\theta \in \Theta$.

$0 \overset{=}{\downarrow} \int_{\mathcal{U}} h(u) \cdot f_{U;\theta}(u)\, du$ (in cts case)

$0 = \sum_{u \in \mathcal{U}} h(u) \cdot p_{U;\theta}(u)$

$\forall\theta \Longrightarrow h(U(\vec{x})) = 0 \ \forall\theta.$

Completeness: An Abstract Definition

- The concept of completeness is notoriously unintuitive – probably the most abstract one in our course – but it will pay off later
- For now, you can think about the finite case a bit like a finite-dimensional basis from linear algebra
- If $\mathbf{v}_1, \ldots, \mathbf{v}_n$ span $\mathbb{R}^n$, then $\sum_{i=1}^n a_i \mathbf{v}_i = \mathbf{0}$ implies $a_i = 0$ for all i
- If $U(\mathbf{X})$ is complete and supported on $\{u_1, \ldots, u_n\}$, then $\sum_{i=1}^n h(u_i) \cdot \mathbb{P}_\theta \left(U(\mathbf{X}) = u_i\right) = 0$ implies $h(u_i) = 0$ for all i
  $u \in \mathcal{U}$
- The meaning will become clearer at the end of Module 2
- So why bring it up now?

Showing Completeness is Very Difficult In General...

$U(\vec{X}) \sim \text{Bin}(n, \theta)$

- Example 1.35: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Bernoulli}(\theta)$ with $\theta \in (0,1)$. Show that $U(\mathbf{X}) = \sum_{i=1}^{n} X_i$ is complete.

Suppose $h(\cdot)$ is any function s.t. $\mathbb{E}_\theta[h(U(\vec{X}))] = 0 \quad \forall \theta \in (0,1)$.

Then $0 = \sum_{j=0}^{n} h(j) \cdot \binom{n}{j} \theta^j (1-\theta)^{n-j}$

$= (1-\theta)^n \sum_{j=0}^{n} h(j) \cdot \left[\binom{n}{j} \left(\frac{\theta}{1-\theta}\right)^j \right]$ let $r = \frac{\theta}{1-\theta} \in (0, \infty)$

$= (1-\theta)^n \sum_{j=0}^{n} \tilde{h}(j) \cdot r^j$ where $\tilde{h}(j) = \binom{n}{j} \cdot h(j)$

which is a polynomial in r, which is $0 \ \forall r \in (0, \infty)$.

Therefore, $\tilde{h}(j) = 0 \ \forall j$

$\Rightarrow \binom{n}{j} \cdot h(j) = 0 \ \forall j$

$\Rightarrow h(j) = 0 \ \forall j.$

$\Rightarrow h(\cdot) = 0$ on $\{0, 1, \ldots, n\}$.

Hence $P_\theta(h(U(\vec{X})) = 0) = 1 \ \forall \theta \in \mathbb{R}$

Hence $U(\vec{X})$ is complete.

...But for Exponential Families, There's Nothing To It

- Theorem 1.8: Let $X_1, \ldots, X_n \overset{iid}{\sim} f_\theta$ be a random sample from an exponential family, where

$$f_\theta(x) = h(x) \cdot g(\theta) \cdot \exp\left(\sum_{j=1}^{k} w_j(\theta) \cdot T_j(x) \right).$$

Then $T(\mathbf{X}) = \left(\sum_{i=1}^{n} T_1(X_i), \ldots, \sum_{i=1}^{n} T_k(X_i) \right)$ is a complete statistic, as long as each component of Θ contains an open interval in $\mathbb{R}$.[1]

- Recall from Theorem 1.2 that in this case, $T(\mathbf{X})$ is also sufficient for θ

- So it's really easy to find complete sufficient statistics for exponential families

[1] More generally, Θ must contain an open set in $\mathbb{R}^k$ – this requirement is sometimes called the "open set condition"

Completeness: Examples

- Example 1.36: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$, where $\mu \in \mathbb{R}$ and $\sigma^2 > 0$ is known. Show that $\bar{X}_n$ is complete for μ.

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

$$= \underbrace{\frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left(\frac{-\mu^2}{2\sigma^2}\right)}_{g(\mu)} \underbrace{\exp\left(\frac{-x^2}{2\sigma}\right)}_{h(x)} \cdot \underbrace{\exp\left(\frac{\mu x}{2\sigma^2}\right)}_{\exp(w(\mu)\cdot T(x))}$$

$\frac{n\mu}{\sigma} \cdot \frac{1}{n} x$, with $w(\mu)$ and $T(x)$ marked.

Since $\Theta = \mathbb{R}$ contains an open interval,

$T(\vec{X}) = \sum_{i=1}^{n} \frac{1}{n} X_i$ is complete for μ by Theorem 1.8.

Completeness: Examples

- Example 1.37: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim}$ Poisson (λ), where $\lambda > 0$. Show that $\bar{X}_n$ is complete for λ.

$$f_\lambda(x) = \frac{\lambda^x e^{-\lambda}}{x!} = \underbrace{\frac{1}{x!}}_{h(x)} \cdot \underbrace{e^{-\lambda}}_{g(\lambda)} \cdot \exp\Big(\underbrace{x}_{T(x)} \cdot \underbrace{\log(\lambda)}_{w(\lambda)} \Big)$$

$\mathcal{H} = (0, \infty)$ contains an open interval, $T(\vec{x}) = \overline{X_n}$ is complete, by Theorem 1.8.

Completeness: Examples

- Example 1.38: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} f_{\mu,\sigma}$ where

$$f_{\mu,\sigma}(x) = \frac{1}{2\sigma} \exp\left(-\frac{|x-\mu|}{\sigma}\right), \quad x \in \mathbb{R},$$

where $\sigma > 0$ and μ is known. Find a complete statistic for σ.

$f_\sigma(x) = \frac{1}{2\sigma} \cdot \exp\left(|x-\mu| \cdot \left(-\frac{1}{\sigma}\right)\right)$ $\quad w(\sigma) = -\frac{1}{\sigma}$

$h(x) = 1$ $\quad g(\sigma) = \frac{1}{2\sigma}$ $\quad T(x) = |x-\mu|$

$\Omega = (0, \infty)$ contains an open interval.

By Theorem 1.8, $T(\vec{X}) = \sum_{i=1}^{n} |X_i - \mu|$ is complete for σ.

Complete Statistics Are Minimal Sufficient!

- There is nothing resembling sufficiency in the definition of completeness; the two concepts seem completely unrelated
- And yet, Theorem 1.8 says that for exponential families, certain complete statistics are sufficient
- What about in general? The answer might surprise you...
- Theorem 1.9 (**Bahadur's theorem**): If a minimal sufficient statistic and a complete statistic both exist, then the complete statistic must also be minimal sufficient. No proof.
- That's *not* the same as saying that all minimal sufficient statistics are complete (which is unfortunately not true)

Minimal Sufficient Statistics Are Not Always Complete

- But if a minimal sufficient statistic exists and it's not complete, then no complete statistic exists

- This is probably the simplest example of a minimal sufficient statistic that is not complete

- Example 1.39: Let $X_1 \sim \text{Unif}(\theta, \theta+1)$, where $\theta \in \mathbb{R}$. Show that $T(X_1) = X_1$ is minimal sufficient for θ, but not complete.

$f_\theta(x) = \mathbb{1}_{\theta \le x \le \theta+1}$.

Let $x, y \in \mathcal{X}$. Then

$$\frac{f_\theta(x)}{f_\theta(y)} = \frac{\mathbb{1}_{\theta \le x \le \theta+1}}{\mathbb{1}_{\theta \le y \le \theta+1}}$$

free of θ iff $x = y$.

Hence $T(X) = X$ is minimal sufficient, by Theorem 1.4.

However, consider $h(x) = \sin(2\pi x)$.

We have $\mathbb{E}_\theta[h(T(X_1))]$

$= \mathbb{E}_\theta[\sin(2\pi X_1)]$

$= \int_\theta^{\theta+1} \sin(2\pi x)\, dx$

$= 0 \quad \forall \theta \in \mathbb{R}$.

Clearly, $h(\cdot)$ is not identically 0.

So $T(X)$ can't be complete.

I wouldn't expect you to come up with a counterexample like this yourself on a quiz!

The Amazingly Useful Basu's Theorem

- Theorem 1.10 (**Basu's theorem**): Complete sufficient statistics are independent of *all* ancillary statistics.

Proof. (Discrete)

Let $T = T(\vec{X})$ be a complete sufficient statistic

Let $S = S(\vec{X})$ be ancillary for θ.

It suffices to show $\mathbb{P}(S=s \mid T=t) = \mathbb{P}(S=s)$.

By the law of total probability, $\mathbb{P}(S=s) = \sum_{t \in T} \mathbb{P}(S=s \mid T=t) \cdot \mathbb{P}_\theta(T=t)$. ①

Also, since $1 = \sum_{t \in T} \mathbb{P}_\theta(T=t)$, $\mathbb{P}(S=s) = \left(\sum_{t \in T} \mathbb{P}_\theta(T=t)\right) \cdot \mathbb{P}_\theta(S=s)$ ②

Therefore $0 = ① - ② = \sum_{t \in T} \left[\mathbb{P}(S=s \mid T=t) - \mathbb{P}(S=s)\right] \cdot \mathbb{P}_\theta(T=t)$

$h(t)$

$= \sum_{t \in T} h(t) \cdot \mathbb{P}_\theta(T=t) = \mathbb{E}_\theta[h(T)]$.

But T is complete $\Rightarrow h(t) = 0 \ \forall \theta \Rightarrow \mathbb{P}(S=s \mid T=t) = \mathbb{P}(S=s)$.

Hence $S \perp\!\!\!\perp T$. □

Poll Time!

Basu's Theorem: Examples

- Example 1.40: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathcal{N}\left(\mu, \sigma^2\right)$ where $\mu \in \mathbb{R}$ and $\sigma^2 > 0$. Show that the sample mean $\bar{X}$ is independent of the sample variance S^2.

Example 1.36 $\Rightarrow$ $\bar{X}_n$ is complete sufficient statistic for μ.

Example 1.39 $\Rightarrow$ S^2 is ancillary for μ.

By Basu's theorem, $\bar{X}_n \perp\!\!\!\perp S^2$.

- This is actually a characterizing property of the Normal distribution: $\bar{X} \perp S^2$ if and only if $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathcal{N}\left(\mu, \sigma^2\right)$ (Interesting fact)

Basu's Theorem: Examples

- Example 1.41: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Exp}(\theta)$, where $\theta > 0$. Use Basu's theorem to find $\mathbb{E}_\theta\left[\frac{X_1}{X_1+\cdots+X_n}\right]$.

$\{\text{Exp}(\theta): \theta > 0\}$ is a scale family; by Theorem 1.6, $\frac{X_1}{X_1+\cdots+X_n}$ is ancillary for θ.

$\{\text{Exp}(\theta): \theta > 0\}$ is an exponential family $T(x) = x$, so $\sum X_i$ is a CSS.

By Basu's theorem, $\frac{X_1}{X_1+\cdots+X_n} \perp\!\!\!\perp X_1+\cdots+X_n$.

Then $\mathbb{E}[X_1] = \mathbb{E}\left[\frac{X_1}{X_1+\cdots+X_n} \cdot (X_1+\cdots+X_n)\right]$

$\frac{1}{\theta}$ $\overset{\text{indep}}{=}$ $\mathbb{E}\left[\frac{X_1}{X_1+\cdots+X_n}\right] \cdot \underbrace{\mathbb{E}[X_1+\cdots+X_n]}_{n/\theta}$

$\Rightarrow \mathbb{E}\left[\frac{X_1}{X_1+\cdots+X_n}\right] = \frac{1}{\theta} \cdot \frac{\theta}{n} = \frac{1}{n}$.

Basu's Theorem: Examples

- Example 1.42: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} f_{\mu,\sigma}$ where

$$f_{\mu,\sigma}(x) = \frac{1}{2\sigma} \exp\left(-\frac{|x-\mu|}{\sigma}\right), \quad x \in \mathbb{R},$$

where $\sigma > 0$ and μ is known. Show that X_1/X_n is independent of $\sum_{i=1}^n |X_i - \mu|$.

This is a scale family, so $\frac{X_1}{X_n}$ is ancillary for σ by Theorem 1.6.

Moreover, $\sum_{i=1}^n |X_i - \mu|$ is a complete sufficient statistic by Example 1.38.

By Basu's theorem, $\frac{X_1}{X_n} \perp\!\!\!\perp \sum_{i=1}^n |X_i - \mu|$.

↑ "independent"