Sufficiency Summary / Tools

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 $T(\mathbf{X})$. That is, if $\mathbf{x}$ and $\mathbf{y}$ are two sample points such that $T(\mathbf{x}) = T(\mathbf{y})$, then the inference about θ should be the same whether $\mathbf{X} = \mathbf{x}$ or $\mathbf{X} = \mathbf{y}$ is observed.

Definition 6.2.1 A statistic $T(\mathbf{X})$ is a *sufficient statistic for* θ if the conditional distribution of the sample $\mathbf{X}$ given the value of $T(\mathbf{X})$ does not depend on θ.

That is: $P_\theta(\underline{X}=x \mid T(\underline{X})=T(\underline{x})) = P(\underline{X}=x \mid T(\underline{X})=T(\underline{x}))$ $(\perp \theta)$

Theorem 6.2.2 *If $p(\mathbf{x}|\theta)$ is the joint pdf or pmf of $\mathbf{X}$ and $q(t|\theta)$ is the pdf or pmf of $T(\mathbf{X})$, then $T(\mathbf{X})$ is a sufficient statistic for θ if, for every $\mathbf{x}$ in the sample space, the ratio $p(\mathbf{x}|\theta)/q(T(\mathbf{x})|\theta)$ is constant as a function of θ.*

Theorem 6.2.6 (Factorization Theorem) *Let $f(\mathbf{x}|\theta)$ denote the joint pdf or pmf of a sample $\mathbf{X}$. A statistic $T(\mathbf{X})$ is a sufficient statistic for θ if and only if there exist functions $g(t|\theta)$ and $h(\mathbf{x})$ such that, for all sample points $\mathbf{x}$ and all parameter points θ,*

$$f(\mathbf{x}|\theta) = g(T(\mathbf{x})|\theta)h(\mathbf{x}). \tag{6.2.3}$$

Theorem 6.2.10 *Let $X_1, \ldots, X_n$ be iid observations from a pdf or pmf $f(x|\boldsymbol{\theta})$ that belongs to an exponential family given by*

$$f(x|\boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta}) \exp\left(\sum_{i=1}^{k} w_i(\boldsymbol{\theta}) t_i(x)\right),$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_d)$, $d \le k$. Then

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

is a sufficient statistic for $\boldsymbol{\theta}$.

Show (def'n 6.2.1) $P(\underline{X}=\underline{x} \mid T(\underline{X})=T(\underline{x})) \perp \theta$

ΣX_i defines the following partition of the Sample Space.

$S=$

HHH	HHT	HTH	THH
HTT	THT	TTH	TTT

$\Sigma X=3$ $\quad \Sigma X=2 \quad \Sigma X=1 \quad \Sigma X=0$

Let ΣX_i be # heads (Random Var. map $S \Rightarrow \mathbb{R}$)

Hmmm
Conditional on ΣX_i the dist'n is $\perp$ of P(Head)

Find $P(HHH \mid \Sigma X_i=2) = 0$

$$P(HHT \mid \Sigma X_i=2) = P(HHT \,\&\, \Sigma X_i=2)/P(\Sigma X_i=2)$$
$$= P(HHT)/P(\Sigma X=2)$$
$$= [p^2 q]/[\tbinom{3}{2} p^2 q]$$
$$= 1/3$$

Similarly $P(HTH \mid \Sigma X_i=2) = P(THH \mid \Sigma X_i=2) = 1/3$

$P(HTT \mid \Sigma X=2) = P(THT \mid \Sigma X_i=2) = P(TTH \mid \Sigma X_i=2) = 0$

$P(HHH \mid \Sigma X=2) = 0$

Hmmm
Similarly we can show that the probability of any outcome in S given the number of heads, X, is independent of p!!

Ratio

$T(\underline{X})$ is sufficient if $\dfrac{p(\underline{x} \mid \theta)}{q(T(\underline{X}) \mid \theta)} \perp \theta$

Choose $T(\underline{X}) = \Sigma X_i \qquad \Sigma X_i \sim \text{bin}(3,p)$

$$\frac{\cancel{p^{\Sigma X_i}(1-p)^{3-\Sigma X_i}}}{\binom{3}{\Sigma X_i} \cancel{p^{\Sigma X_i}(1-p)^{3-\Sigma X_i}}} = \frac{1}{\binom{3}{\Sigma X_i}} \perp \theta$$

Factorization Thm: $f(\underline{x} \mid \theta) = \underbrace{g(T(\underline{X}) \mid \theta)}_{\text{f'n } \underline{x} \,\&\, \theta} \underbrace{h(\underline{x})}_{\text{f'n } \underline{x}} \Rightarrow T(\underline{X})$ sufficient

$$f(\underline{x} \mid p) = \prod_{i=1}^{3} p^{x_i}(1-p)^{1-x_i} I_{[0,1]}(x_i) = p^{\Sigma X_i}(1-p)^{3-\Sigma X_i} \prod_{i=1}^{3} I_{[0,1]}(x_i)$$

$$= \underbrace{\left(\frac{p}{1-p}\right)^{\Sigma X_i}(1-p)^3}_{g(T(\underline{X}) \mid \theta)} \underbrace{\prod_{i=1}^{3} I_{[0,1]}(x_i)}_{h(\underline{x})}$$

$\therefore \Sigma X_i$ is sufficient

Exponential family $f(x|\theta) = h(x)\,c(\theta)\exp\big(t(x)\,\omega(\theta)\big)$

$$f(x|\theta) = p^{x}(1-p)^{1-x}\, I_{[0,1]}(x) \longleftarrow$$ Note this is for 1 x.

$$p^{x}(1-p)^{1-x} I_{[0,1]}(x) = (1-p)\left(\frac{p}{1-p}\right)^{x} I_{[0,1]}(x)$$

$$= \underbrace{(1-p)}_{c(\theta)}\ \underbrace{I_{[0,1]}(x)}_{h(x)}\ \exp\Big\{\underbrace{x}_{t(x)}\ \underbrace{\log\left(\frac{p}{1-p}\right)}_{\omega(\theta)}\Big\}$$

$\therefore\ \sum_{i=1}^{3} X_i$ is sufficient

---

§6.2.2 Minimal Sufficient Statistics

In any problem there are many suff. stats.

→ The complete sample $X_1, \dots, X_n = \underline{X}$ is sufficient

by factorization Thm: $f(\underline{x}|\theta) = \underbrace{f(T(\underline{x})|\theta)}_{T(\underline{X}) = \underline{X}}\ \underbrace{h(\underline{x})}_{1}$ or Indicator

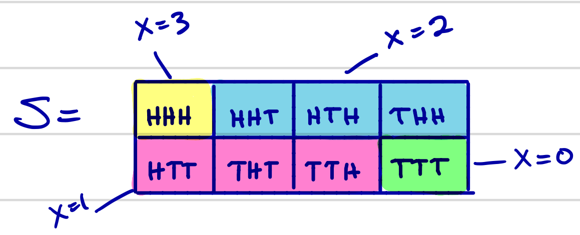
Flip coin $f(\underline{x}|\theta) = \underbrace{\prod_{i=1}^{3} p^{x_i}(1-p)^{1-x_i} I_{[0,1]}(x_i)}_{g(T(\underline{x})|\theta)} \cdot \underbrace{1}_{h(\underline{x})}$

Also any one-to-one ft'n of a sufficient stat is also sufficient

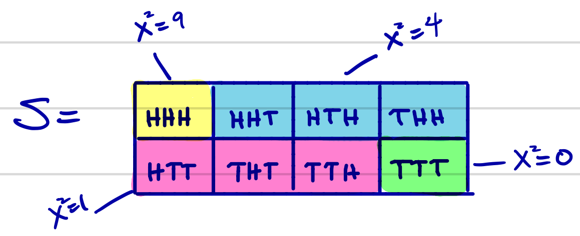
- defines same partition

Coin Flip

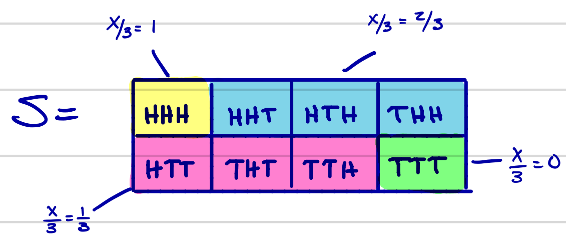
Partition Sample space by $T(\underline{X}) = \Sigma X_i$

$S=$ 

$(\Sigma X_i)^2$ defines same partition

$S=$ 

$\frac{\Sigma X_i}{3}$ defines same partition

$S=$ 

Any one-to-one transformation of $T(\underline{X})$ defines same partion.

- There will be numerous sufficient stats for any problem.

- Purpose of suff stat is to achieve data reduction without losing info about θ, our parameter of interest.

- Want most data reduction (statistic with smallest dimension).
  - But no loss of information

Definition 6.2.11 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 $T'(\mathbf{x})$.

Coin flip example: X_1, X_2, X_3 is sufficient (the observed sample)

ICBST (soon) ΣX_i is minimal sufficient
or $\Sigma X_i/n$ (a one-to-one transformation)

A practical Theorem for identifying minimal suff stats:

Theorem 6.2.13 *Let $f(\mathbf{x}|\theta)$ be the pmf or pdf of a sample $\mathbf{X}$. Suppose there exists a function $T(\mathbf{x})$ such that, for every two sample points $\mathbf{x}$ and $\mathbf{y}$, the ratio $f(\mathbf{x}|\theta)/f(\mathbf{y}|\theta)$ is 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 for θ.*

A "somewhat unsatisfying proof" - with coin flips in Appendix

iff

$T(\underset{\sim}{x})$ minimal sufficient $\Longrightarrow$ / $\Longleftarrow$ $\dfrac{f(\underset{\sim}{x}|\theta)}{f(\underset{\sim}{y}|\theta)}$ constant wrt θ if $T(\underset{\sim}{x}) = T(\underset{\sim}{y})$

$\rightarrow T(\underset{\sim}{x}) = T(\underset{\sim}{y}) \rightarrow$ same partion of sample space

$\underset{\sim}{x}$ & $\underset{\sim}{y}$ on

Coin flip example: $f(x|p) = p^x(1-p)^{1-x} I_{[0,1]}(x)$

$$\frac{f(\underline{x}|p)}{f(\underline{y}|p)} = \frac{p^{\Sigma x_i}(1-p)^{3-\Sigma x_i}\prod_{i=1}^{3} I_{[0,1]}(x_i)}{p^{\Sigma y_i}(1-p)^{3-\Sigma y_i}\prod_{i=1}^{3} I_{[0,1]}(y_i)}$$

$$= \frac{\cancel{(1-p)^3}\left(\frac{p}{1-p}\right)^{\Sigma x_i}\prod_{i=1}^{3} I_{[0,1]}(x_i)}{\cancel{(1-p)^3}\left(p/1-p\right)^{\Sigma y_i}\prod_{i=1}^{3} I_{[0,1]}(y_i)}$$

$$= \underbrace{\underbrace{\left(\frac{p}{1-p}\right)^{\Sigma x_i - \Sigma y_i}}_{\text{ft'n of } p \text{ unless } \Sigma x_i = \Sigma y_i} \underbrace{\frac{\prod_{i=1}^{3} I_{[0,1]}(x_i)}{\prod_{i=1}^{3} I_{[0,1]}(y_i)}}_{\text{Not ft'n of } p}}_{\text{constant wrt } p \text{ if } \Sigma x_i = \Sigma y_i}$$

$\therefore$ Σx_i is minimal sufficient

Example: $X_1, \dots, X_n$ iid $N(\mu, \sigma^2)$ μ & σ^2 both unknown

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

$= \dots$

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

How do we make this constant wrt (μ, σ^2)

$\bar{X} = \bar{Y} \Rightarrow (\bar{x}^2 = \bar{y}^2)$

$S_x^2 = S_y^2$

$\therefore$ $\bar{X}, S^2$ minimal sufficient example: $f(x|\theta) = \frac{1}{\pi}\frac{1}{(1+(x-\theta)^2)} I_{(-\infty,\infty)}(x)$ — Cauchy

$$\frac{f(\underline{x}|\theta)}{f(\underline{y}|\theta)} = \frac{\prod_{i=1}^{n} \frac{1}{\pi}\frac{1}{(1+(x_i-\theta)^2)} I_{(-\infty,\infty)}(x_i)}{\prod_{i=1}^{n} \frac{1}{\pi}\frac{1}{(1+(y_i-\theta)^2)} I_{(-\infty,\infty)}(y_i)}$$

$\underline{x}, \underline{y}$ lie on same partition iff ratio $\perp \theta$ for $T(\underline{x}) = T(\underline{y})$

True iff $(x_1, \dots, x_n)$ is permutation of $(y_1, \dots y_n)$
(same order statistics)

Cauchy $n=2$
(homework).

Assume $n=2$

Find $\prod_{i=1}^{n} 1 + (y_i-\theta)^2 \leftarrow$ polynomial of degree $2n$ in θ.

$$(1+(y_1-\theta)^2)(1+(y_2-\theta)^2)$$

$$= 1 + (y_1^2 - 2\theta y_1 + \theta^2) + (y_2^2 - 2\theta y_2 + \theta^2)$$
$$+ [(y_1^2 - 2\theta y_1 + \theta^2)(y_2^2 - 2\theta y_2 + \theta^2)]$$

Aside

$$[\,] = y_1^2 y_2^2 - 2\theta y_2 y_1^2 + y_1\theta^2 - 2\theta y_1 y_2^2 + 4\theta^2 y_1 y_2 - 2\theta^3 y_1 + \theta^2 y_2 - 2\theta^3 y_2 + \theta^4$$

$$(1+(y_1-\theta)^2)(1+(y_2-\theta)^2)$$

$$\left.\begin{aligned} = &\ 1 + y_1^2 + y_2^2 + y_1^2 y_2^2 \\ &+ \theta[-2y_1 - 2y_2 - 2y_2 y_1^2 - 2y_1 y_2^2] \\ &+ \theta^2[1+1 + y_1 + 4y_1 y_2 + y_2] \\ &+ \theta^3[-2y_1 - 2y_2] \\ &+ \theta^4[1] \end{aligned}\right\}$$

need $\frac{f(\underline{x}|\theta)}{f(\underline{y}|\theta)} \perp$ of θ

Need coefficients of $\theta, \theta^2, \theta^3, \theta^4$ to be same
true if $X_{(1)} = Y_{(1)}$ and $X_{(2)} = Y_{(2)}$

For general n need $X_{(1)} = Y_{(1)} \dots X_{(n)} = Y_{(n)}$.

minimal suff statistic for Cauchy $\underbrace{X_{(1)}, X_{(2)}, \dots X_{(n)}}_{\text{order stats.}}$

Example: $X_1, \dots, X_n$ iid $U(\theta, \theta+1)$ $\quad -\infty < \theta < \infty$

$\theta < X_{(1)} < X_{(n)} < \theta+1$

$$f(\underline{x}|\theta) = \begin{cases} 1 & \prod_{i=1}^{n} I_{[\theta,\theta+1]}(x_i) \\ 0 & \text{else} \end{cases} = 1 \underbrace{I_{[\theta,\infty)}(X_{(1)})}_{\substack{\text{indicator} \\ \theta > X_{(1)}}} \cdot \underbrace{I_{[-\infty,\theta+1]}(X_{(n)})}_{\substack{\text{indicator} \\ X_{(n)} < \theta+1}}$$

$$\underbrace{I_{[\theta < X_{(1)} < X_{(n)} < \theta+1]}(X_{(1)}, X_{(n)})}$$

or $I_{[X_{(n)}-1 < \theta < X_{(1)}]}(X_{(1)}, X_{(n)})$

$$\frac{f(\underline{x}|\theta)}{f(\underline{y}|\theta)} = \frac{I_{[X_{(n)}-1 < \theta < X_{(1)}]}(x_{(1)}, x_{(n)})}{I_{[y_{(n)}-1 < \theta < y_{(1)}]}(y_{(1)}, y_{(n)})}$$

indicators must be same
$\rightarrow X_{(1)} = Y_{(1)}$
$+ \; X_{(n)} = Y_{(n)}$

$\therefore X_{(1)}, X_{(n)}$ is minimal sufficient.

Finally § 6.2.3 Ancillary Statistics

Definition 6.2.16 A statistic $S(\mathbf{X})$ whose distribution does not depend on the parameter θ is called an *ancillary statistic.*

- sufficient stats contain all info about θ, which is available from sample space

- Ancillary statistic has "a complementary purpose".

Back to $U(\theta,\theta+1)$ example: $X_{(1)}, X_{(n)}$ minimal sufficient

any one-to-one transformation is minimal sufficient

$\therefore \underbrace{X_{(n)} - X_{(1)}}_{\substack{\text{we can show} \\ \text{(homework)}}}, \; \frac{(X_{(1)}+X_{(n)})}{2}$ is also min. sufficient

$\underbrace{X_{(n)} - X_{(1)}}_{\text{ancillary.}} \sim \text{beta}(n-1, 2) \longleftarrow$ does ~~not~~ depend on θ !

Consider situation T' is minimal sufficient stat for θ and

- $\dim(T') > \dim(\theta)$ ⟨ like $U(\theta, \theta+1)$ example
- Then sometimes happens write $T' = (T, S)$ where S is ancillary (marginal dist'n of $S \perp \theta$).
- We can think of T as 'conditionally sufficient', (Cox & Hinkley)
  T sufficient $\mid S=s$ (T sufficient conditional on $S=s$).

- Choose S to have maximum dimension.

Uniform $(\theta, \theta+1)$ example cont. $f(\underline{x}|\theta) = I^{(x_{(1)}, x_{(n)})}_{[x_{(n)}-1 < \theta < x_{(1)}]}$

know $(X_{(1)}, X_{(n)})$ is minimal sufficient

one-to-one transformations of $(X_{(1)}, X_{(n)})$ are also minimal sufficient

$\left(X_{(n)} - X_{(1)}, \frac{X_{(1)}+X_{(n)}}{2}\right) = (\text{Range, Midrange})$ is one-to-one

$R = X_{(n)} - X_{(1)}$

$M = (X_{(1)} + X_{(n)})/2$

or $X_{(1)} = M - R/2$

$X_{(n)} = M + R/2$

[solve for old in terms of new

$X_{(1)} = X_{(n)} - R \rightarrow M = \frac{X_{(n)} - R + X_{(n)}}{2} = X_{(n)} - R/2$

$X_{(n)} = M + R/2 \rightarrow X_{(1)} = M + \frac{R}{2} - \frac{2R}{2} = M - \frac{R}{2}$]

(M, R) is minimal sufficient, but R is ancillary (see Appendix, $R \sim \text{beta}(n-1, 2)$)

Does R give info about θ?

R gives info about θ; even though R ancillary!

If $M = 4.5$ what are possible θ?

$\frac{X_{(n)} + X_{(1)}}{2} = 4.5$

[possible θ? If $\left.\begin{array}{l} X_{(n)} = \theta+1 \\ X_{(1)} = \theta \end{array}\right\rangle$

If $X_{(n)} = \theta+1$ & $X_{(1)} = \theta$

$M = \frac{\theta+\theta+1}{2} = \theta + 1/2$ $\quad \theta = M - 1/2$]

know $\theta = 4$ given $R = 1$

Note R alone gives no info about θ.

given $R=0$

$3.5 < \theta < 4.5$

[If $X_{(n)} = X_{(1)}$ $\quad M = \frac{2X_{(1)}}{2} = X_{(1)}$ (or $X_{(n)}$)

$\theta < X_{(1)}$ $\quad \theta+1 > X_{(n)}$ $\quad \theta+1 > 4.5$

$\theta > 3.5$]

given $R = .5$

$3.75 < \theta < 4.25$

[$X_{(1)} = 4.5 - .25 = 4.25$ $\quad \theta < 4.25$

$X_{(n)} = 4.5 + .25 = 4.75$ $\quad \theta+1 > 4.75$ $\quad \theta > 3.75$]

- The ancillarity of R does not depend on Uniform dist'n of X_i's but because θ is a location parameter

$$U(0,1) \quad f(x|\theta) = 1\, I_{[0,1]}(x)$$

$$U(\theta,\theta+1) \quad f(x|\theta) = 1\, I_{[0,1]}(x-\theta)$$

where ever have x replace by $x-\theta$

$$I_{[0,1)}(x-\theta) = I_{[\theta,\theta+1]}(x)$$

$f(x-\theta)$ θ is location parameter.

Example : Random sample size. (N is a random variable) (Cox + Hinkley)

- Let N be a random variable w/ a known dist'n.

$$p_n = P(N=n) \quad n = 1,2,3,\ldots \qquad (N \sim \text{Pois}(\lambda) \text{ or } N \sim \text{negbin}(p), \ldots)$$

- Let $Y_1, \ldots, Y_N$ be iid with density from exponential family (random)

- Then

$$f_{N,y}(n,\underset{\sim}{y}) = f(n)\, f(\underset{\sim}{y}|n)$$

$$= (p_n)\left[\left[\prod_{i=1}^{n} h(y_i)\right][c(\theta)]^n \exp\left[\sum_{i=1}^{n} w(\theta)\, t(y_i)\right]\right]$$

by ratio thm

$$\frac{p_n\left[\prod_{i=1}^{n} h(y_i)\right][c(\theta)]^n \exp\left[w(\theta)\sum_{i=1}^{n} t(y_i)\right]}{p_m\left[\prod_{i=1}^{m} h(z_i)\right][c(\theta)]^m \exp\left[w(\theta)\sum_{i=1}^{m} t(z_i)\right]}$$

$$= \frac{p_n}{p_m}\,\frac{\prod_{i=1}^{n} h(y_i)}{\prod_{i=1}^{m} h(z_i)}\,[c(\theta)]^{n-m} \exp\left[w(\theta)\left[\sum_{i=1}^{n} t(y_i) - \sum_{i=1}^{m} t(z_i)\right]\right]$$

$\perp \theta$ if $n=m$ and $\sum_{i=1}^{n} t(y_i) = \sum_{i=1}^{m} t(z_i)$

$\therefore N, \sum_{i=1}^{N} t(y_i)$ are minimal sufficient

N is ancillary

$\sum_{i=1}^{N} t(y_i)$ is sufficient conditional on N.

Any sample size not fixed in advance, but with known dist'n $\perp$ of θ, is ancillary.

<u>Example</u>: Mixture of Normal dist'ns (Cox + Hinkley)

$$Y \begin{cases} \xrightarrow{} p = 1/2 \sim N(\mu, \sigma_1^2) & \sigma_1^2 \text{ known} \\ \xrightarrow{} q = 1/2 \sim N(\mu, \sigma_2^2) & \sigma_2^2 \text{ known} \end{cases} \qquad S = \begin{cases} 1 & \text{if } N(\mu, \sigma_1^2) \\ 2 & \text{if } N(\mu, \sigma_2^2) \end{cases}$$

$$f_{S,Y}(s,y) = \left(\frac{1}{2}\right)(2\pi\sigma_s^2)^{-1/2} \exp\left\{-\frac{1}{2}\left(\frac{y-\mu}{\sigma_s}\right)^2\right\} = f(s) f(y|s)$$

find minimal sufficient:

$$f_{S,Y}(s,y) = \frac{\cancel{\left(\frac{1}{2}\right)}(\cancel{2\pi}\sigma_s^2)^{-1/2} \exp\left\{-\frac{1}{2}\left(\frac{y-\mu}{\sigma_s}\right)^2\right\}}{\cancel{\left(\frac{1}{2}\right)}(\cancel{2\pi}\sigma_r^2)^{-1/2} \exp\left\{-\frac{1}{2}\left(\frac{z-\mu}{\sigma_r}\right)^2\right\}}$$

μ unknown
σ_1^2, σ_2^2 known

$$= \left(\frac{\sigma_s^2}{\sigma_r^2}\right)^{-1/2} \frac{\exp\left\{-\frac{1}{2}\left(\frac{y^2 - 2\mu y + \mu^2}{\sigma_s^2}\right)\right\}}{\exp\left\{-\frac{1}{2}\left(\frac{z^2 - 2\mu z + \mu^2}{\sigma_r^2}\right)\right\}}$$

$$= \left(\frac{\sigma_s^2}{\sigma_r^2}\right)^{-1/2} e^{-\frac{1}{2}(y^2 - z^2)} \, e^{\mu\left(y/\sigma_s^2 - \frac{z}{\sigma_r^2}\right)} \, e^{-\frac{1}{2}\mu^2\left[1/\sigma_s^2 - 1/\sigma_r^2\right]}$$

Need $y = z$ and $\underbrace{\sigma_s^2 = \sigma_r^2}_{\text{Know } S}$ to be $\perp$ of μ

$\therefore$ Y, S minimal sufficient for μ

$P(S=1) = P(S=2) = 1/2 \perp \mu \rightarrow$ ancillary

i.e. $P(\sigma_s^2 = \sigma_1^2) = P(\sigma_s^2 = \sigma_2^2) = 1/2 \perp \mu$

Y is sufficient conditional on S.

→ Random sample size and mixture of Normals, suggest inference should be conditional on ancillary statistic.

→ Observed ancillary statistic, S, describes the part of the total sample space relevant to problem.

Theorem 3.5.6 *Let $f(\cdot)$ be any pdf. Let μ be any real number, and let σ be any positive real number. Then X is a random variable with pdf $(1/\sigma)f((x-\mu)/\sigma)$ if and only if there exists a random variable Z with pdf $f(z)$ and $X = \sigma Z + \mu$.*

Example 6.2.18 (Location parameter family)

cdf $F(x-\theta)$ $-\infty<\theta<\infty$.

3.5.6

Show range is ancillary statistic. (based on theorem ~~3.4.2~~)

$Z_1,\dots,Z_n$ iid observations from $F(x)$ (corresponding to $\theta=0$)

w/ $X_1 = Z_1+\theta,\dots, X_n = Z_n+\theta$.

Thus cdf of range statistic. R is

$$
\begin{aligned}
F_R(r|\theta) &= P_\theta(R\le r)\\
&= P_\theta(X_{(n)}-X_{(1)}\le r)\\
&= P_\theta\left(\max_i(Z_i+\theta) - \min_i(Z_i+\theta)\le r\right)\\
&= P_\theta\left(\max_i Z_i - \min_i Z_i + \theta - \theta \le r\right)\\
&= P_\theta(Z_{(n)}-Z_{(1)}\le r)
\end{aligned}
$$

← does not depend on θ! $Z_1,\dots Z_n$ doesn't depend on θ.

cdf of R does not depend on θ

$\Rightarrow$ R is ancillary.

Example 6.2.19 (Scale parameter families)

$X_1,\dots,X_n$ iid cdf $F(x/\sigma)$ $\sigma>0$

Any statistic depends on sample through $(n-1)$ values:

$\dfrac{X_1}{X_n}, \dfrac{X_2}{X_n},\dots,\dfrac{X_{n-1}}{X_n}$ is ancillary

ie $\dfrac{X_1+X_2+\dots+X_n}{X_n} = \dfrac{X_1}{X_n}+\dfrac{X_2}{X_n}+\dots+\dfrac{X_{n-1}}{X_n}+1$ is ancillary.

details (homework).

§ 6.2.4 Sufficient, Ancillary + Complete Statistics

" A minimal sufficient statistic is a statistic that has achieved the maximal amount of data reduction possible, while still retaining all of the information about the parameter θ."

→ But examples $U(\theta, \theta+1)$ + $N(\mu, \sigma_3^2)$ show that ancillary statistic can be an important component of the minimal sufficient statistic.

→ Intuition correct that ideally minimal sufficient statistic is independent of any ancillary statistic.

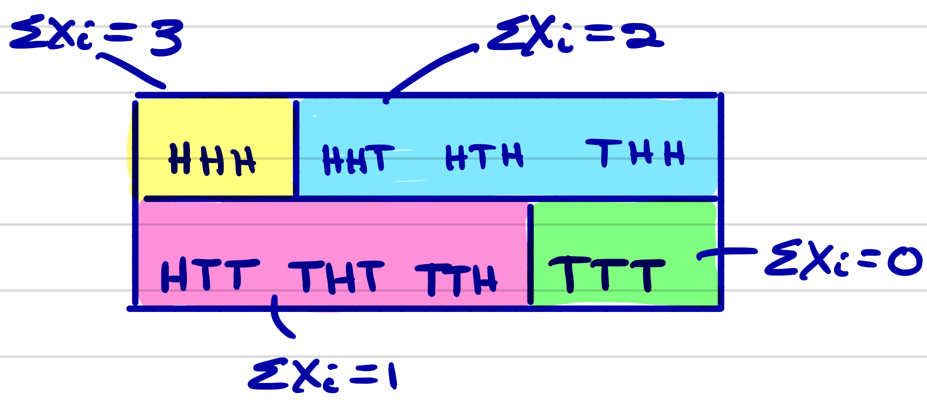
→ Happens with complete statistic:

Definition 6.2.21 Let $f(t|\theta)$ be a family of pdfs or pmfs for a statistic $T(\mathbf{X})$. The family of probability distributions is called *complete* if $E_\theta g(T) = 0$ for all θ implies $P_\theta(g(T) = 0) = 1$ for all θ. Equivalently, $T(\mathbf{X})$ is called a *complete statistic*.

Heuristic Interpretation:

Flipping coin:

Partition sample space based on # heads is complete.



Not Complete:

If separate data partitions that have same $T(\underline{x})$ (sufficient statistic)
→ Not Complete.

Definition 6.2.21 Let $f(t|\theta)$ be a family of pdfs or pmfs for a statistic $T(\mathbf{X})$. The family of probability distributions is called *complete* if $E_\theta g(T) = 0$ for all θ implies $P_\theta(g(T) = 0) = 1$ for all θ. Equivalently, $T(\mathbf{X})$ is called a *complete statistic.*

Completeness property of family of dist'n (dist'n of $T(\underset{\sim}{x})$)

Example:

$T(\underset{\sim}{x}) \sim \text{bin}(n,p) \quad 0<p<1$

Let g be a ft'n s.t. $E_p[g(T)] = 0 \ \forall p$

That is $0 = E_p[g(T)] \overset{?}{=} \sum_{t=0}^{n} g(t)\binom{n}{t} p^t (1-p)^{n-t}$

$$0 = \underbrace{(1-p)^n}_{\substack{\neq 0 \\ \text{for } 0<p<1}} \underbrace{\sum_{t=0}^{n} g(t)\binom{n}{t}\left(\frac{p}{1-p}\right)^t}_{\text{must be zero}} \quad \text{for all } p \quad 0<p<1$$

$r = \frac{p}{1-p}$

$$0 = \underbrace{\sum_{t=0}^{n} \underbrace{g(t)\binom{n}{t}}_{\text{coefficient of poly}} r^t}_{}$$

polynomial of degree n in r

$\binom{n}{t} \neq 0 \quad \binom{n}{t} > 0 \quad r^t \neq 0 \quad r^t > 0$

for any t $\binom{n}{t} + r^t > 0$

For polynomial to be zero $\forall r \ 0<r<\infty$ each coefficient $(g(t)\binom{n}{t})$ must be zero.

$\therefore g(t) = 0$ for $t = 0, 1, \ldots n$

$\therefore$

$$E[g(T)] = 0 \Rightarrow P_p(g(T) = 0) = 1 \quad \forall T$$

$\Rightarrow T$ is a complete statistic.

$\rightarrow$ We can't find any function $g(t)$, except $g(t) = 0 \ \forall t$, that has $E[g(T)] = 0$ under the binomial dist'n.

Let $X_1, \dots, X_n$ be iid $U(0,\theta)$ $0<\theta<\infty$; $f(x|\theta) = \underbrace{\frac{1}{\theta} I_{(0,\theta)}(x_{(n)})}$ ← Continuous

We can show by ratio $X_{(n)}$ is minimal sufficient $T(\underset{\sim}{X}) = X_{(n)}$

Is the dist'n of $\overbrace{T(\underset{\sim}{X})}^{\text{not dist'n of } T(\underset{\sim}{X}) \text{ not } X.} = X_{(n)}$ complete ?

$$f(t|\theta) = n t^{n-1} \theta^{-n} I_{(0,\theta)}(t)$$

Complete $\underbrace{E_\theta[g(T)] = 0}_{E[g(T)] \text{ constant wrt } \theta}$ $\forall \theta \Rightarrow P(g(T)=0) = 1$

$\Rightarrow$ clever trick know $\frac{d}{d\theta} E[g(T)] = 0 = \frac{d}{d\theta} \int_0^\theta g(t) n t^{n-1} \theta^{-n} dt$

$$= \frac{d}{d\theta}\left[(\theta^{-n}) \cdot \int_0^\theta g(t) n t^{n-1} dt\right]$$

by product rule

can't interchange $\int$ + $\frac{d}{d\theta}$

$$= \underbrace{-n\theta^{-n-1} \int_0^\theta g(t) n t^{n-1} dt}_{-\frac{n}{\theta} \underbrace{\int_0^\theta g(t) n t^{n-1} \theta^{-n} dt}_{-\frac{n}{\theta} E[g(T)] = 0}} + \underbrace{\theta^{-n} \frac{d}{d\theta} \int_0^\theta g(t) n t^{n-1} dt}$$

Use Leibnitz's Rule

Theorem 2.4.1 (Leibnitz's Rule) *If $f(x,\theta)$, $a(\theta)$, and $b(\theta)$ are differentiable with respect to θ, then*

$$\frac{d}{d\theta}\int_{a(\theta)}^{b(\theta)} f(x,\theta)\,dx = f(b(\theta),\theta)\frac{d}{d\theta}b(\theta) - f(a(\theta),\theta)\frac{d}{d\theta}a(\theta) + \int_{a(\theta)}^{b(\theta)} \frac{\partial}{\partial\theta} f(x,\theta)\,dx.$$

Notice that if $a(\theta)$ and $b(\theta)$ are constant, we have a special case of Leibnitz's Rule:

$$\frac{d}{d\theta}\int_a^b f(x,\theta)\,dx = \int_a^b \frac{\partial}{\partial\theta} f(x,\theta)\,dx.$$

$$= \theta^{-n}\left[n g(\theta) \theta^{n-1}\right] = n g(\theta) \theta^{-1}$$

Need $\frac{n}{\theta} g(\theta) = 0$ $\forall \theta$ $0<\theta<\infty$

$\frac{n}{\theta} \neq 0 \Rightarrow g(\theta) = 0$ $\forall \theta > 0$

$\Rightarrow T(\underset{\sim}{X}) = X_{(n)}$ is a complete statistic

Why do we care about completeness? ancillary? minimal sufficient?

Theorem 6.2.24 (Basu's Theorem) *If $T(\mathbf{X})$ is a complete and minimal sufficient statistic, then $T(\mathbf{X})$ is independent of every ancillary statistic.*

→ Completeness : Condition under which a minimal sufficient statistic is independent of every ancillary statistic.

(proof appendix)

Most problems covered by exponential family

Theorem 6.2.25 (Complete statistics in the exponential family) *Let $X_1, \ldots, X_n$ be iid observations from an exponential family with pdf or pmf of the form*

$$f(x|\boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta}) \exp\left(\sum_{j=1}^{k} w(\theta_j) t_j(x)\right), \tag{6.2.7}$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$. Then 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 as long as the parameter space Θ contains an open set in $\Re^k$.

(C+B errata)

Theorem 6.2.28 *If a minimal sufficient statistic exists, then any complete sufficient statistic is also a minimal sufficient statistic.*

not needed for proof.

Theorem 6.2.24 (Basu's Theorem) *If $T(\mathbf{X})$ is a complete and ~~minimal~~ sufficient statistic, then $T(\mathbf{X})$ is independent of every ancillary statistic.*

For problems we consider a sufficient statistic will be complete only if it is minimal sufficient.
(separate same → lose completeness)

That is: complete + sufficient $\Rightarrow$ minimal

Example: $X_1, \ldots, X_n$ iid exponential $f(x|\theta) = \frac{1}{\theta} e^{-x/\theta} I_{[0,\infty)}(x)$ (Note θ is a scale parameter)

Find Expected Value $g(X) = \dfrac{X_n}{X_1 + \cdots + X_n}$ } note since scale parameter $g(X)$ is ancillary stat.

BPOEF (By properties of exponential families)

$\sum_{i=1}^{n} X_i$ is a complete, sufficient statistic. (Thm 6.2.10, 6.2.25)

By Basu's $\Sigma X_i = T(\underline{X}) \perp g(\underline{X})$

Easy to get $E[g(x)]$

$$\theta = E[X_n] = E[T(\underline{X}) g(\underline{X})] = E[T(\underline{X})]\, E[g(\underline{X})] = n\theta\, E[g(\underline{X})]$$

$$E[g(\underline{X})] = \frac{\theta}{n\theta} = \frac{1}{n}.$$

Appendix Show $X \sim U(\theta, \theta+1)$
$R = X_{(n)} - X_{(1)}$ is ancillary

C+B page 282-283

of R does not depend on θ. Recall that the cdf of each X_i is

$$F(x|\theta) = \begin{cases} 0 & x \le \theta \\ x - \theta & \theta < x < \theta + 1 \\ 1 & \theta + 1 \le x. \end{cases}$$

Thus, the joint pdf of $X_{(1)}$ and $X_{(n)}$, as given by (5.5.7), is

$$g(x_{(1)}, x_{(n)}|\theta) = \begin{cases} n(n-1)(x_{(n)} - x_{(1)})^{n-2} & \theta < x_{(1)} < x_{(n)} < \theta + 1 \\ 0 & \text{otherwise.} \end{cases}$$

Making the transformation $R = X_{(n)} - X_{(1)}$ and $M = (X_{(1)} + X_{(n)})/2$, which has the inverse transformation $X_{(1)} = (2M - R)/2$ and $X_{(n)} = (2M + R)/2$ with Jacobian 1, we see that the joint pdf of R and M is

$$h(r, m|\theta) = \begin{cases} n(n-1)r^{n-2} & 0 < r < 1, \theta + (r/2) < m < \theta + 1 - (r/2) \\ 0 & \text{otherwise.} \end{cases}$$

(Notice the rather involved region of positivity for $h(r, m|\theta)$.) Thus, the pdf for R is

$$\begin{aligned} h(r|\theta) &= \int_{\theta+(r/2)}^{\theta+1-(r/2)} n(n-1)r^{n-2}dm \\ &= n(n-1)r^{n-2}(1-r), \quad 0 < r < 1. \end{aligned}$$

This is a beta pdf with $\alpha = n - 1$ and $\beta = 2$. More important, the pdf is the same for all θ. Thus, the distribution of R does not depend on θ, and R is ancillary. ||

Appendix: Basu's Proof.

THEOREM 6.2.24 (Basu's Theorem): If $T(X)$ is a complete and minimal sufficient statistic, then $T(X)$ is independent of every ancillary statistic.

* Completeness → condition under which a minimal sufficient statistic is independent of every ancillary stat.
* "Since complete sufficient statistics are particularly effective in reducing the data, it is not surprising that a complete sufficient statistic is always minimal."
  Lehmann p.46

Proof: (discrete dist'n only)

$S(\underline{X})$ be any ancillary statistic

$P(S(\underline{X})=s)$ does not depend on θ, by def'n of ancillary.

Also

$$P(S(\underline{X})=s \mid T(\underline{X})=t) = P(\underline{X} \in \{\underline{x} : S(\underline{x})=s\} \mid T(\underline{X})=t)$$

does not depend on θ. — because $T(\underline{X})$ is sufficient.

DEFINITION 6.1.1: A statistic $T(X)$ is a *sufficient statistic for θ* if the conditional distribution of the sample X given the value of $T(X)$ does not depend on θ.

To show $S(\underline{X})$ and $T(\underline{X})$ are independent, show

$$P(S(\underline{X})=s \mid T(\underline{X})=t) = P(S(\underline{X})=s) \text{ for all } t \in T$$

$$* \quad P(S(\underline{X})=s) = \sum_{t\in T} P(S(\underline{X})=s \mid T(\underline{X})=t)\, P_\theta(T(\underline{X})=t).$$

is a fct'n of θ

Furthermore, since $\sum_{t\in T} P_\theta(T(\underline{X})=t) = 1$

$$P(S(\underline{X})=s) = \sum_{t\in T} P(S(\underline{X})=s)\, P_\theta(T(\underline{X})=t)$$

$$** \quad = P(S(\underline{X})=s) \sum_{t\in T} P_\theta(T(\underline{X})=t)$$

have shown $P(S(\underline{X}) = s) = \sum_{t \in T} P(S(\underline{X}) = s \mid T(\underline{X}) = t)\, P_\theta(T(\underline{X}) = t)$ $*$

and

$$P(S(\underline{X}) = s) = \sum_{t \in T} P(S(\underline{X}) = s)\, P_\theta(T(\underline{X}) = t \qquad **$$

Therefore if we define

$$g(t) = P(S(\underline{X}) = s \mid T(\underline{X}) = t) - P(S(\underline{X}) = s)$$

$$E_\theta[g(T)] = \sum_{t \in T} g(t)\, P_\theta(T(\underline{X}) = t) = 0 \text{ for all } \theta.$$

Since $T(\underline{X})$ is a complete statistic, this implies that $g(t) = 0$ for all $t \in T$.

$g(t) = 0 \rightarrow P(S(\underline{X}) = s \mid T(\underline{X}) = t) = P(S(\underline{X}) = s)$

$\rightarrow S(\underline{X})$ & $T(\underline{X})$ are independent.

// Q.E.D.

Note: "Minimality" not used in proof.
Theorem true if word "minimal" omitted.

THEOREM 6.2.24 (Basu's Theorem): If $T(X)$ is a complete and ~~minimal~~ sufficient statistic, then $T(X)$ is independent of every ancillary statistic.

For problems we consider, a sufficient statistic will be complete only if it is minimal.