Things that make you say 'hmmm'.

- Flip a coin 3 times $S=$

HHH	HHT	HTH	THH
HTT	THT	TTH	TTT

$P(H) = p$

$P(T) = 1-p = q$

$$\begin{cases} P(HHH) = p^3 \\ P(HHT) = P(HTH) = P(THH) = p^2q \\ P(HTT) = P(THT) = P(TTH) = pq^2 \\ P(TTT) = q^3 \end{cases}$$

$S=$

x=3 x=2

HHH	HHT	HTH	THH
HTT	THT	TTH	TTT

x=1 X=0

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

Assume $X=2$ → restrict our sample space.
(2 Heads)

$S=$

HHH	HHT	HTH	THH
HTT	THT	TTH	TTT

Hmmm

Conditional on $X=2$ the dist'n is $\perp$ of P(Head)

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

$P(HHT \mid X=2) = P(HHT \text{ and } X=2)/P(X=2)$

$= P(HHT)/P(X=2)$

$= [p^2q]/[\binom{3}{2}p^2q]$

$= 1/3$

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

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

$P(TTT \mid X=2) = 0$

Hmmm

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

Quick Review

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: $P_\theta(\underset{\sim}{X}=x \mid T(\underset{\sim}{X})=T(\underset{\sim}{x})) = P(\underset{\sim}{X}=x \mid T(\underset{\sim}{X})=T(\underset{\sim}{x}))$ $(\perp \theta)$

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.*

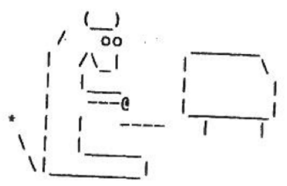
Goal: Find sufficient statistic using Thm 6.2.2

'intuition'

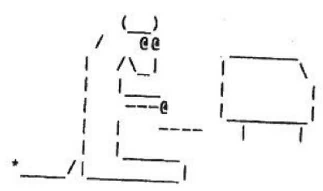
Step 1) "Guess" statistic $T(\underset{\sim}{X})$ to be sufficient

Step 2) Find dist'n of $T(\underset{\sim}{X})$ — may require tedious analysis.

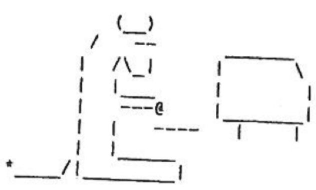
Step 3) Check ratio to determine if independent of θ



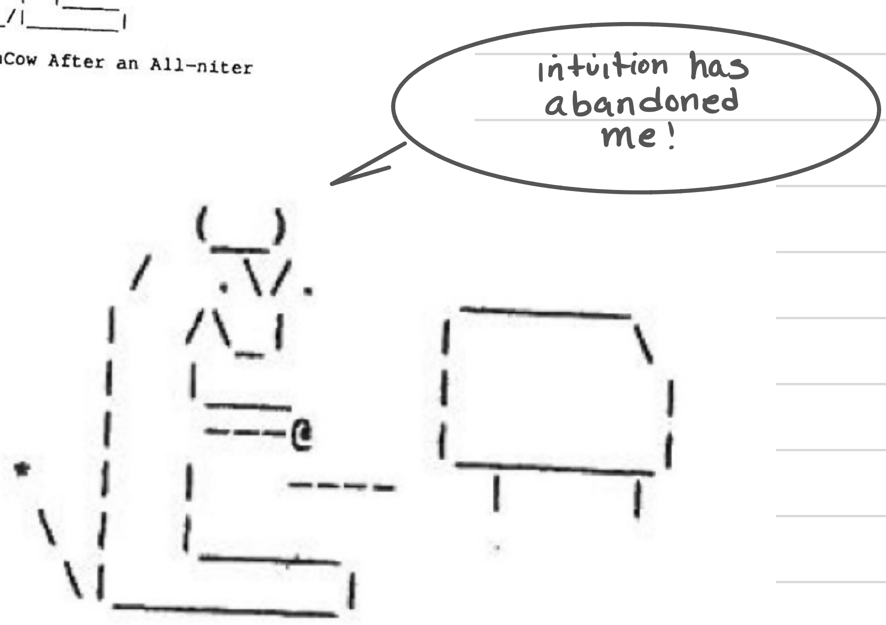
CompuCow



CompuCow After an All-niter



CompuCow Asleep at the Keyboard



Compu Cow 'guesses' wrong sufficient statistic.

Sufficiency is basically a property of partitioning the sample space.

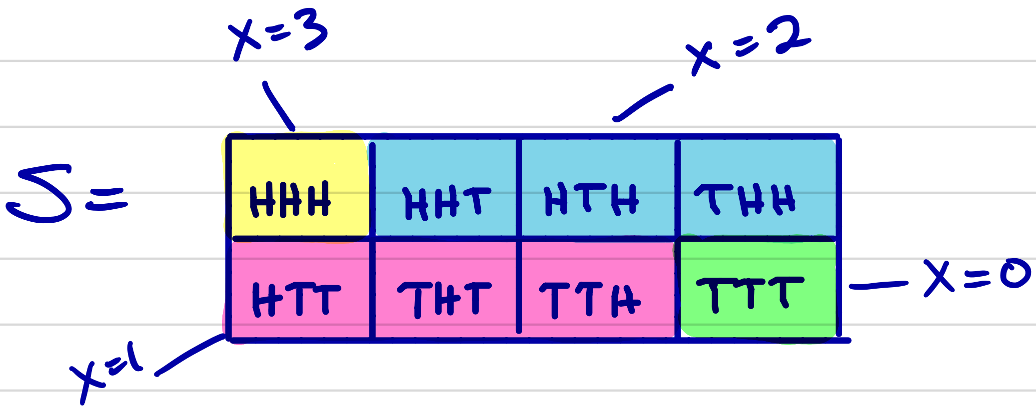
- Given the partition, the conditional dist'n of $\underset{\sim}{X}$ does not depend on θ. ("Hmmm")
- Same info in $T(\underset{\sim}{x})$ (partition) as in $\underset{\sim}{x}$ (sample point).

For Coin flip example: assume $X = \#$ heads

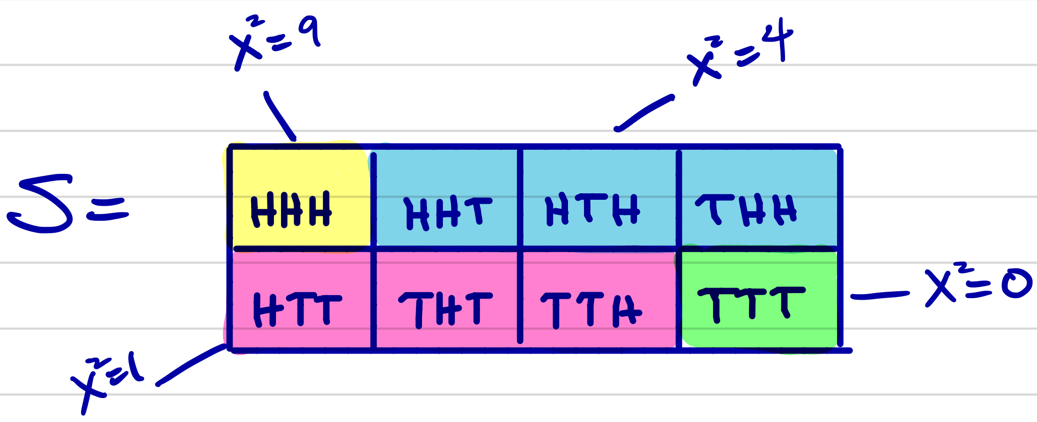
$X_i = \begin{cases} 0 & \text{if tails} \\ 1 & \text{if heads} \end{cases} \Big\} \; X = \Sigma X_i$

Coin Flip

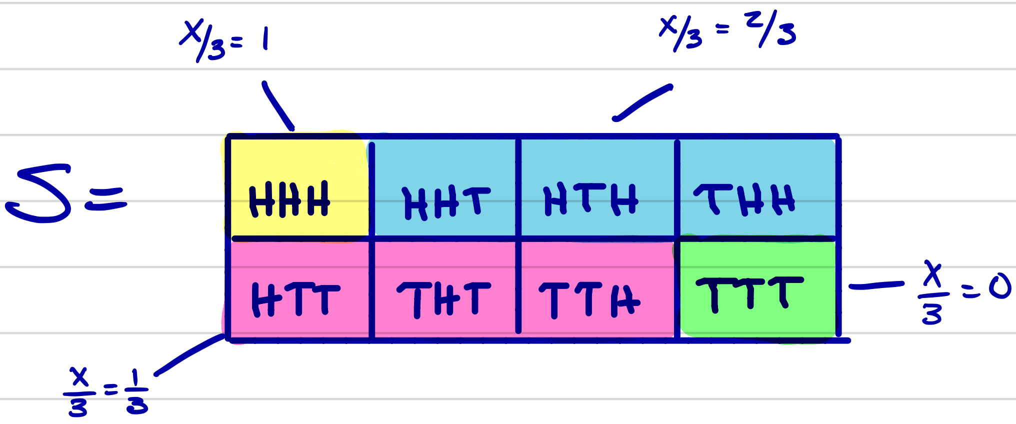
Partition Sample space by $T(\underset{\sim}{x}) = \Sigma X_i$



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



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



Any one-to-one transformation of $T(\underset{\sim}{x})$ defines same partion.

Theorem allows us to find a sufficient statistic by "simple" inspection of $f(\underline{x}|\theta)$.

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,*

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

$$T(\underline{x}) \text{ sufficient} \underset{\Longleftarrow}{\overset{\text{iff}}{\Longrightarrow}} f(\underline{x}|\theta) = g(T(\underline{x})|\theta)\,h(\underline{x})$$

must prove $T(\underline{x}) \Rightarrow f(\underline{x}|\theta)\ g(T(\underline{x})|\theta)\,h(\underline{x})$

and

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

prove for the discrete case

Prove: $T(\underline{x})$ sufficient $\Rightarrow f(\underline{x}|\theta) = g(T(\underline{x})|\theta)\,h(\underline{x})$

(discrete)

(HHT is subset of X=2, where X=# heads)

$$f(\underline{x}|\theta) = P_\theta(\underline{X}=\underline{x}) = P_\theta(\underline{X}=\underline{x} \text{ and } T(\underline{X})=T(\underline{x}))$$

$\underline{X}=\underline{x}$ is subset of partition $T(\underline{X})=T(\underline{x})$.

$$= P_\theta(\underline{X}=\underline{x} \mid T(\underline{X})=T(\underline{x})) \cdot P_\theta(T(\underline{X})=T(\underline{x}))$$

since $T(\underline{X})$ is sufficient

$$= P(\underline{X}=\underline{x} \mid T(\underline{X})=T(\underline{x})) \cdot P_\theta(T(\underline{X})=T(\underline{x}))$$

$\perp\theta$, by def'n of sufficiency

$$= \underbrace{P_\theta(T(\underline{X})=T(\underline{x}))}_{g(T(\underline{x})|\theta)} \cdot \underbrace{P(\underline{X}=\underline{x} \mid T(\underline{X})=T(\underline{x}))}_{h(\underline{x})}$$

Choose $g(t|\theta) = P_\theta(T(\underline{X})=t)$

$h(\underline{x}) = P(\underline{X}=\underline{x} \mid T(\underline{X})=T(\underline{x}))$

not a function of θ

Because $T(\underline{X})$ is sufficient, independent of θ (by def'n)

// half-way done.

Proof (discrete) cont.

Prove $\underset{\text{sufficient}}{T(X)} \Leftarrow f(\underline{x}|\theta) = g(T(\underline{x})|\theta)\, h(\underline{x})$

sufficient $\overset{\text{prove}}{\Leftarrow}$ Assume: factorization exists

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.*

Assume $T(X) \sim q(t|\theta)$

examine ratio: $f(\underline{x}|\theta) / q(T(\underline{x})|\theta)$

HHH	HHT	HTH	THH
HTT	THT	TTH	TTT

Define set $A_{T(\underline{x})} = \{y : T(y) = T(\underline{x})\}$ ← Partion which contains $T(\underline{X}) = t$ → realized value

Then
$$\frac{f(\underline{x}|\theta)}{q(T(\underline{x})|\theta)} = \frac{g(T(\underline{x})|\theta)\, h(\underline{x})}{\sum_{A_{T(\underline{x})}} g(T(y)|\theta)\, h(y)}$$
← $P_\theta(T(\underline{X}) = T(\underline{x}))$ – sum over all points $\underline{x}$ s.t. $T(\underline{X}) = t$. (Probability of observing the partition.)

$$= \frac{\cancel{g(T(\underline{x})|\theta)}\, h(\underline{x})}{\cancel{g(T(\underline{x})|\theta)} \sum_{A_{T(\underline{x})}} h(y)}$$
← since T is constant on $A_{T(\underline{x})}$

$$= h(\underline{x}) \Big/ \sum_{A_{T(\underline{x})}} h(y)$$
which is $\perp$ of θ.

$\Rightarrow T(\underline{X})$ is sufficient statistic // Q.E.D.

Example: Consider a random sample $X_1, X_2, \ldots X_n$ where each observation is geometric (p) $0 < p < 1$.

$$f(x|p) = p^x (1-p)\, I_{[0,1,2\ldots)}^{(x)}$$

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,*

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

$$f(\underline{x}|p) = \prod_{i=1}^{n} p^{x_i}(1-p)\, I_{[0,1,2\ldots)}^{(x_i)} = (1-p)^n p^{\sum x_i} \prod_{i=1}^{n} I_{[0,1,2,\ldots)}^{(x_i)}$$

$$\overset{?}{=} g(T(\underline{x})|p)\, h(\underline{x})$$

$$f(\underline{x}|p) = \underbrace{(1-p)^n p^{\sum x_i}}_{g(T(\underline{x})|p)} \underbrace{\prod_{i=1}^{n} I_{[0,1,2,\ldots)}^{(x_i)}}_{h(\underline{x})}$$

$\Rightarrow T(\underline{x}) = \sum x_i$

$\therefore$ $t = \sum x_i$ is sufficient for p.

Example (Classic): $X_1, \dots X_n$ iid $U(1,\theta)$ ← discrete θ = positive integer.

3-sided dice

Let $N = \{1,2,3,\dots\}$ = set of positive integers (possible θ)

$$I_N(x_i) = \begin{cases} 1 & \text{if } x_i \in N \\ 0 & \text{else} \end{cases}$$ indicator that x_i is positive integer

$N_\theta = \{1,2,3,\dots,\theta\}$ = possible values for $X_1,\dots,X_n$ given θ.

Sample space is ft'n of θ.

$$I_{N_\theta}(x_i) = \begin{cases} 1 & \text{if } x_i \in N_\theta \\ 0 & \text{else} \end{cases}$$ ← indicator x_i integer between 1 and θ

$$f(\underline{x}|\theta) = \prod_{i=1}^{n} \theta^{-1} I_{N_\theta}(x_i) = \theta^{-n} \prod_{i=1}^{n} I_{N_\theta}(x_i)$$

$$= \theta^{-n} \underbrace{\prod_{i=1}^{n} I_N(x_i)}_{\text{guarantees positive integer}} \underbrace{\prod_{i=1}^{n} I_{N_\theta}(x_i)}_{\text{if all positive integers then this will be one if all } x_i \le \theta}$$

$$= \underbrace{\theta^{-n} \underbrace{\prod_{i=1}^{n} I_N(x_i)}_{\text{positive integer}} \underbrace{I_{N_\theta}(X_{(n)})}_{\max \le \theta}}_{\text{positive integer and } \max(x_i) \le \theta}$$

$$\therefore f(\underline{x}|\theta) = \underbrace{\theta^{-n} I_{N_\theta}(X_{(n)})}_{g(T(\underline{x})|\theta)} \cdot \underbrace{\prod_{i=1}^{n} I_N(x_i)}_{h(\underline{x})}$$

$\therefore T(\underline{x}) = X_{(n)}$

$X_{(n)} = \max(x_i)$ is sufficient for θ.

Example $X_1, \ldots, X_n$ iid $N(\mu, \sigma^2)$ and assume σ^2 is known.

ICBST (It Can Be Shown That) (homework)

$$f(\underset{\sim}{x}|\mu) = (2\pi\sigma^2)^{-n/2} \exp\left(-\left(\Sigma(x_i-\bar{x})^2 + n(\bar{x}-\mu)^2\right)/2\sigma^2\right) \cdot \prod_{i=1}^{n} I_{(-\infty,\infty)}(x_i)$$

$$= \underbrace{e^{-n(\bar{x}-\mu)^2/2\sigma^2}}_{\substack{g(T(\underset{\sim}{x})|\mu) \\ (\sigma^2 \text{ known})}} \underbrace{e^{-\Sigma(x_i-\bar{x})^2/2\sigma^2} (2\pi\sigma^2)^{-n/2} \cdot \prod_{i=1}^{n} I_{(-\infty,\infty)}(x_i)}_{\substack{h(\underset{\sim}{x}) \\ (\sigma^2 \text{ known})}}$$

$\bar{X}$ is sufficient by factorization Thm.

Example: Sufficient statistic can be a vector

say $T(\underset{\sim}{X}) = (T_1(\underset{\sim}{X}), \ldots, T_r(\underset{\sim}{X}))$ is a vector.

Often: $\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_s)$ is also a vector.

$X_1, \ldots, X_n$ iid $N(\mu, \sigma^2)$ – σ^2 <u>unknown</u>

$$f(\underset{\sim}{x}|\mu) = (2\pi\sigma^2)^{-n/2} \exp\left(-\left(\Sigma(x_i-\bar{x})^2 + n(\bar{x}-\mu)^2\right)/2\sigma^2\right) \cdot \prod_{i=1}^{n} I_{(-\infty,\infty)}(x_i)$$

$$= e^{-n(\bar{x}-\mu)^2/2\sigma^2} e^{-\Sigma(x_i-\bar{x})^2/2\sigma^2} (2\pi\sigma^2)^{-n/2} \cdot \prod_{i=1}^{n} I_{(-\infty,\infty)}(x_i)$$

$$= \underbrace{e^{-n(\bar{x}-\mu)^2/2\sigma^2} e^{-(n-1)s^2/2\sigma^2} (2\pi\sigma^2)^{-n/2}}_{\substack{g(T(\underset{\sim}{x})|\mu,\sigma^2) \\ \sigma^2 \text{ unknown}}} \cdot \underbrace{\prod_{i=1}^{n} I_{(-\infty,\infty)}(x_i)}_{h(\underset{\sim}{x})}$$

$$f(\underset{\sim}{x}|\mu,\sigma^2) = g(T_1(\underset{\sim}{x}), T_2(\underset{\sim}{x}) | \mu, \sigma^2)\, h(\underset{\sim}{x})$$

$\therefore$ $(\bar{X}, S^2)$ is sufficient for μ, σ^2

→ in the normal dist'n.

If we know $f(\underset{\sim}{x}|\theta)$ is a member of an exponential family, we can easily find a set of sufficient statistics.

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 $\boldsymbol{\theta}$.

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

where $\underset{\sim}{\theta} = \theta_1, \theta_2, \ldots \theta_d$ $\quad d \le k$ — includes curved exponential family like $N(\theta, \theta^2)$

$$T(\underset{\sim}{X}) = \sum_{j=1}^{n} t_1(X_j), \ldots, \sum_{j=1}^{n} t_k(X_j)$$

is sufficient for $\underset{\sim}{\theta}$

- Return to coin flip example: Find sufficient statistics

- Flip a coin 3 times $\quad S =$

HHH	HHT	HTH	THH
HTT	THT	TTH	TTT

$X_i = 1$ H $\quad 0$ else

$P(H) = p$

$P(T) = 1-p = q$

$$\begin{cases} P(HHH) = p^3 \\ P(HHT) = P(HTH) = P(THH) = p^2 q \\ P(HTT) = P(THT) = P(TTH) = pq^2 \\ P(TTT) = q^3 \end{cases}$$

$S =$

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 $S \Rightarrow \mathbb{R}$)

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.

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

$S=$

HHH	HHT	HTH	THH
HTT	THT	TTH	TTT

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

$P(HHT \mid \Sigma X_i=2) = P(HHT \,\&\, \Sigma X_i = 2)/P(\Sigma X_i = 2)$

$= P(HHT)/P(\Sigma X = 2)$

$= [p^2 q]/[\binom{3}{2} p^2 q]$

$= 1/3$

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

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

Factorization Thm:

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

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

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

$\therefore$ ΣX_i is sufficient

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

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

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

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

$\therefore \sum_{i=1}^{3} X_i$ 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: $f(\underline{x}|\theta) = \underbrace{f(T(\underline{x})|\theta)}_{T(\underline{x}) = \underline{x}}\ \underbrace{h(\underline{x})}_{1 \text{ or Indicator}}$

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 $T(\underline{x}) = \sum x_i$

$S =$

$x=3$	$x=2$		
HHH	HHT	HTH	THH
HTT	THT	TTH	TTT

$x=1$ (HTT, THT, TTH); $x=0$ (TTT)

$(\sum x_i)^2$ defines same partition

$S =$

$x^2=9$	$x^2=4$		
HHH	HHT	HTH	THH
HTT	THT	TTH	TTT

$x^2=1$ (HTT, THT, TTH); $x^2=0$ (TTT)

$\frac{\sum x_i}{3}$ defines same partition

$S =$

$x/3=1$	$x/3=2/3$		
HHH	HHT	HTH	THH
HTT	THT	TTH	TTT

$\frac{x}{3}=\frac{1}{3}$ (HTT, THT, TTH); $\frac{x}{3}=0$ (TTT)

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) $\sum X_i$ is minimal sufficient

or $\sum X_i/n$ (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

$T(\underset{\sim}{x})$ minimal sufficient $\overset{\text{iff}}{\Longleftrightarrow}$ $\dfrac{f(\underset{\sim}{x}|\theta)}{f(\underset{\sim}{y}|\theta)}$ constant wrt θ if $T(\underset{\sim}{x}) = T(\underset{\sim}{y})$

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

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

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

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

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

unless $\Sigma x_i = \Sigma y_i$

constant wrt p 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

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

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

$$\begin{aligned} = {} & 1 + y_1^2 + y_2^2 + y_1^2 y_2^2 \\ & + \theta[-2y_1 - 2y_2 - 2y_2y_1^2 - 2y_1y_2^2] \\ & + \theta^2[1+1 + y_1 + 4y_1y_2 + y_2] \\ & + \theta^3[-2y_1 - 2y_2] \\ & + \theta^4[1] \end{aligned}$$

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 $U(\theta, \theta+1)$ $\quad -\infty < \theta < \infty$

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

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

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

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

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

Appendix: Proof Thm 6.3.13

Master's Theory

Lecture 6/5

Somewhat unsatisfying proof.
— really need measure theory

Proof: Assume $f(\underline{x}|\theta) > 0 \ \forall \underline{x} \in \mathcal{X}$ and θ — makes proof simpler, won't need to worry about dividing by zero.

Step 1) Show $T(X)$ is sufficient	Bernoulli, 3 trials example
	$T(x) = X_1 + X_2 + X_3 = 2$
Let $\mathcal{T} = \{t : t = T(\underline{x})$ for some $\underline{x} \in \mathcal{X}\}$	ex: $\mathcal{T} = \{t : t = 2$ for $\underline{x} \in \{(0,1,1)(1,0,1)(1,1,0)\}$
	$T(\underline{x}) = \{0,1,2,3\}$
Define $A_t = \{\underline{x} : T(\underline{x}) = t\}$ ↳ partition induced by $T(\underline{x})$	$A_0 = \{\{(0,0,0)\} : T(\underline{x}) = 0\}$ $A_1 = \{\{(0,0,1),(0,1,0),(1,0,0)\} : T(\underline{x}) = 1\}$ $A_2 = \{\{(0,1,1),(1,0,1),(1,1,0)\} : T(\underline{x}) = 2\}$ $A_3 = \{\{(1,1,1)\} : T(\underline{x}) = 3\}$ ↑ partition induced by $T(\underline{x}) = x_1 + x_2 + x_3$
For each partition, A_t, choose & fix one element $\underline{x}_t \in A_t$ — ie the 'first' member of the set	
For any $\underline{x} \in \mathcal{X}$, $\underline{x}_{T(\underline{x})}$ is the fixed element that is in the same set, A_t, as $\underline{x}$.	$\underline{x} = \{(0,0,0)\} \rightarrow x_{T(\underline{x})} = \{(0,0,0)\}$ $\underline{x} = \{(0,0,1),(0,1,0),(1,0,0)\} \rightarrow x_{T(\underline{x})} = \{(0,0,1)\}$ $\underline{x} = \{(0,1,1),(1,0,1),(1,1,0)\} \rightarrow x_{T(\underline{x})} = \{(0,1,1)\}$ $\underline{x} = \{(1,1,1)\} \rightarrow x_{T(\underline{x})} = \{(1,1,1)\}$
$\underline{x}$ + $\underline{x}_{T(\underline{x})}$ are in same partition A_t $\therefore T(\underline{x}) = T(\underline{x}_{T(\underline{x})})$	partitions all $\underline{x} \in \mathcal{X}$ with $f(\underline{x}|\theta) > 0$
$\therefore f(\underline{x}|\theta)/f(\underline{x}_{T(\underline{x})}|\theta)$ is a constant wrt θ	ie $\underline{x} = (1,1,0)$ + $x_{T(\underline{x})} = (0,1,1)$
define $h(\underline{x}) = f(\underline{x}|\theta)/f(\underline{x}_{T(\underline{x})}|\theta)$, $\underline{x} \in \mathcal{X}$	$\dfrac{f(\underline{x}|\theta)}{f(\underline{x}_{T(\underline{x})}|\theta)} = \dfrac{p^2(1-p)}{p^2(1-p)} = 1$
define $g(t|\theta) = f(\underline{x}_t|\theta)$, $t \in \mathcal{T}$	$\therefore h(\underline{x}) = 1$ $g(t|\theta) = p^t(1-p)$ $\quad t \in \{0,1,2,3\}$
Then: $f(\underline{x}|\theta) = \dfrac{f(\underline{x}_{T(\underline{x})}|\theta)}{f(\underline{x}_{T(\underline{x})}|\theta)} f(\underline{x}|\theta) = g(T(\underline{x})|\theta)h(\underline{x})$	$f(\underline{x}|\theta) = p^{T(\underline{x})}(1-p)$
By Factorization $T(\underline{X})$ is sufficient	

Appendix: Proof Thm 6.2.13 cont.

Master's Theory

Lecture 6/6

Proof Cont.

minimal sufficiency Proof cont: $\Bigg[$ - $\underset{\sim}{X} \sim f(\underset{\sim}{x}|\theta)$, $\exists$ ft'n $T(\underset{\sim}{x})$ s.t. $f(\underset{\sim}{x}|\theta)/f(\underset{\sim}{y}|\theta)$ is constant wrt θ iff $T(\underset{\sim}{x}) = T(\underset{\sim}{y})$

$\therefore$ $T(\underset{\sim}{X})$ is minimal sufficient

Step 2: Show $T(\underset{\sim}{X})$ is minimal: $\Big[$ Def'n 6.2.11 $T(\underset{\sim}{x})$ is a ft'n of $T'(\underset{\sim}{x})$ = any other suff. stat.

Let $T'(\underset{\sim}{X})$ be any other sufficient statistic for θ

By the factorization theorem $\exists$ ft'ns $g' + h'$ s.t.

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

Let $\underset{\sim}{x}$ + $\underset{\sim}{y}$ be any two sample points s.t.

$T'(\underset{\sim}{x}) = T'(\underset{\sim}{y})$. Then

$$\frac{f(\underset{\sim}{x}|\theta)}{f(\underset{\sim}{y}|\theta)} = \frac{g'(T'(\underset{\sim}{x})|\theta)\,h'(\underset{\sim}{x})}{g'(T'(\underset{\sim}{y})|\theta)\,h'(\underset{\sim}{y})} = \frac{h'(\underset{\sim}{x})}{h'(\underset{\sim}{y})}$$ independent of θ.

By assumptions of theorem $\Rightarrow$ $T(\underset{\sim}{x}) = T(\underset{\sim}{y})$

$\Bigg[$? Why ? $T(\underset{\sim}{X})$ is a function of $T'(\underset{\sim}{x})$

i.e. $T(\underset{\sim}{X}) = f(T'(\underset{\sim}{x}))$

f is for function not pdf

if $T'(\underset{\sim}{x}) = T'(\underset{\sim}{y})$

$\Rightarrow T(\underset{\sim}{x}) = f(T'(\underset{\sim}{x})) = f(T'(\underset{\sim}{y})) = T(\underset{\sim}{y})$

$T(\underset{\sim}{x}) = T(\underset{\sim}{y}) \Rightarrow$ ratio independent of θ

$\therefore \Rightarrow T(\underset{\sim}{X})$ minimal sufficient.

// Q.E.D.