Example (Approximate mean + variance)

- assume X is R.V. $E[X] = \mu \neq 0$ [here $n=1$]
- wish to estimate $g(\mu)$
- First order Taylor's:

$$g(X) = g(\mu) + g'(\mu)(X-\mu)$$

$$E[g(X)] \approx g(\mu)$$

$$Var[g(X)] \approx (g'(\mu))^2 \, Var[X]$$

Suppose: $g(\mu) = 1/\mu$

estimate $1/\mu$ by $1/X$ since $E[1/X] \approx 1/\mu$

$$Var[1/X] \approx (1/\mu)^4 \, Var(X)$$

hmm... again ft'n of unknown μ.

- also what if don't know $Var(X)$?

Theorem 5.5.24 (Delta Method) *Let Y_n be a sequence of random variables that satisfies $\sqrt{n}(Y_n - \theta) \to n(0, \sigma^2)$ in distribution. For a given function g and a specific value of θ, suppose that $g'(\theta)$ exists and is not 0. Then*

(5.5.10) $$\sqrt{n}[g(Y_n) - g(\theta)] \to n(0, \sigma^2[g'(\theta)]^2) \text{ in distribution.}$$

Proof (Taylor's + Slutsky's).

Continuing above example (now have random sample, rather than $n=1$).

$E[\bar{X}] = \mu$ assuming finite $var(X_1) = \sigma^2$ $\quad (Var(\bar{X}) = \sigma^2/n)$

by CLT: $\sqrt{n}(\bar{X} - \mu) \xrightarrow{d} N(0, \sigma^2)$

by Delta: $\sqrt{n}\left(\frac{1}{\bar{X}} - \frac{1}{\mu}\right) \xrightarrow{d} N\left(0, \left(\frac{1}{\mu}\right)^4 \sigma^2\right)$

or $Var\left(\frac{1}{\bar{X}}\right) \approx \left(\frac{1}{\mu}\right)^4 \sigma^2$

know $\bar{X} \xrightarrow{P} \mu$ WLLN, $S^2 \to \sigma^2$ Cheby's $\Big\}$ Lecture 3

$$\Rightarrow \left(\frac{1}{\bar{X}}\right)^4 \xrightarrow{P} \left(\frac{1}{\mu}\right)^4$$

$$\left(\frac{1}{\bar{X}}\right)^4 S^2 \xrightarrow{P} \left(\frac{1}{\mu}\right)^4 \sigma^2$$

Theorem 5.5.4 *Suppose that $X_1, X_2, \ldots$ converges in probability to a random variable X and that h is a continuous function. Then $h(X_1), h(X_2), \ldots$ converges in probability to $h(X)$.*

Summary

$$\sqrt{n}\left(\frac{1}{\bar{X}}-\frac{1}{\mu}\right) \xrightarrow{\mathcal{L}} N\left(0, \left(\frac{1}{\mu}\right)^4\sigma^2\right)$$
$$\left(\frac{1}{\bar{X}}\right)^4 S^2 \xrightarrow{P} \left(\frac{1}{\mu}\right)^4\sigma^2$$

Convergence in Law $\xrightarrow{\mathcal{L}}$
Convergence in prob $\xrightarrow{P}$
Slutsky's

Theorem 5.5.17 (Slutsky's Theorem) *If $X_n \to X$ in distribution and $Y_n \to a$, a constant, in probability, then*

a. *$Y_n X_n \to aX$ in distribution.*

b. *$X_n + Y_n \to X + a$ in distribution.*

$$\frac{\sqrt{n}\left(\frac{1}{\bar{X}}-\frac{1}{\mu}\right)}{\left(\frac{1}{\bar{X}}\right)^2 S} = \underbrace{\frac{\sqrt{n}\left(\frac{1}{\bar{X}}-\frac{1}{\mu}\right)}{\left(\frac{1}{\mu}\right)^2\sigma}}_{\xrightarrow{\mathcal{L}} N(0,1)} * \underbrace{\frac{\left(\frac{1}{\mu}\right)^2\sigma}{\left(\frac{1}{\bar{X}}\right)^2 S}}_{\xrightarrow{P} 1} \xrightarrow{\mathcal{L}} N(0,1)$$

Extensions of Delta Method

If $g'(\theta)=0$ $\longrightarrow$

Theorem 5.5.26 (Second-order Delta Method) *Let Y_n be a sequence of random variables that satisfies $\sqrt{n}(Y_n-\theta) \to \mathrm{n}(0,\sigma^2)$ in distribution. For a given function g and a specific value of θ, suppose that $g'(\theta)=0$ and $g''(\theta)$ exists and is not 0. Then*

(5.5.13) $$n[g(Y_n)-g(\theta)] \to \sigma^2\frac{g''(\theta)}{2}\chi_1^2 \text{ in distribution.}$$

If $\underset{\sim}{\theta}$ is vector

vector of length p

n patients, each with p observations.

Theorem 5.5.28 (Multivariate Delta Method) *Let $\mathbf{X}_1,\ldots,\mathbf{X}_n$ be a random sample with $\mathrm{E}(X_{ij})=\mu_i$ and $\mathrm{Cov}(X_{ik},X_{jk})=\sigma_{ij}$. For a given function g with continuous first partial derivatives and a specific value of $\boldsymbol{\mu}=(\mu_1,\ldots,\mu_p)$ for which $\tau^2=\Sigma\Sigma\sigma_{ij}\frac{\partial g(\boldsymbol{\mu})}{\partial\mu_i}\cdot\frac{\partial g(\boldsymbol{\mu})}{\partial\mu_j}>0$,*

$$\sqrt{n}[g(\bar{X}_1,\ldots,\bar{X}_s)-g(\mu_1,\ldots,\mu_p)] \to \mathrm{n}(0,\tau^2) \text{ in distribution.}$$

$s=p$?

Var/Cov matrix symmetric $\sigma_{ij}=\sigma_{ji}$

If $\sigma_{ij}=\mathrm{Cov}(X_i,X_j)$
$\sigma_i^2=\mathrm{Var}(X_i)=\mathrm{Cov}(X_i,X_i)$

→ Define symmetric Variance / Covariance matrix

$$\tau^2 = \underbrace{\underset{(1\times p)}{\begin{bmatrix} g_1'(\theta) & \cdots & g_p'(\theta)\end{bmatrix}} \underset{(p\times p)}{\begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2p} \\ \vdots & & \ddots & \\ \sigma_{1p} & & & \sigma_p^2 \end{bmatrix}} \underset{(p\times 1)}{\begin{bmatrix} g_1'(\theta) \\ \vdots \\ g_p'(\theta)\end{bmatrix}}}_{1\times 1}$$

Note: This version of multivariate delta is ft'ns of $\bar{X}_1,\ldots,\bar{X}_s$.
n is sample size p is # parameters (# θs).

assumes $(\bar{X}_1,\ldots\bar{X}_p) \to MVN\left[\begin{pmatrix}\mu_1\\ \vdots \\ \mu_p\end{pmatrix}, \underset{\text{var/cov matrix}}{\Sigma}\right]$ by CLT...

Example (Moments of a Ratio Estimator)

Assume $\bar{X}, \bar{Y}$ with $E[\bar{X}] = \mu$ $\quad E[\bar{Y}] = \mu_y$ $\quad Cov(X,Y) = \sigma_{xy}$

$Var[\bar{X}] = \sigma_x^2/n \quad Var[\bar{Y}] = \sigma_y^2/n$

$Var[X] = \sigma_x^2 \quad Var[Y] = \sigma_y^2$

Further assume $g(\mu_x, \mu_y) = \frac{\mu_x}{\mu_y}$

$$\frac{\partial}{\partial \mu_x}\left[\frac{\mu_x}{\mu_y}\right] = \frac{1}{\mu_y}$$

$$\frac{\partial}{\partial \mu_y}\left[\frac{\mu_x}{\mu_y}\right] = \frac{-\mu_x}{\mu_y^2}$$

By Taylor's

$$E[X_1/Y_1] \cong \frac{\mu_x}{\mu_y}$$

$$Var[X_1/Y_1] \cong \frac{1}{\mu_y^2}\sigma_x^2 + \frac{\mu_x^2}{\mu_y^4}\sigma_y^2 - 2\frac{\mu_x}{\mu_y^3}\sigma_{xy}$$

$$= \left(\frac{\mu_x}{\mu_y}\right)^2\left[\frac{\sigma_x^2}{\mu_x^2} + \frac{\sigma_y^2}{\mu_y^2} - \frac{2\sigma_{xy}}{\mu_x \mu_y}\right]$$

for $n=1$

$$= \left[\frac{1}{\mu_y}, -\frac{\mu_x}{\mu_y^2}\right]\begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix}\begin{bmatrix} 1/\mu_y \\ -\mu_x/\mu_y^2 \end{bmatrix}$$

§5.6 Generating a Random Sample - Focus on Probability Integral Transform.

Recall Last semester (Lecture-5)

One of most useful transformations:

Theorem 2.1.10 (Probability integral transformation) *Let X have continuous cdf $F_X(x)$ and define the random variable Y as $Y = F_X(X)$. Then Y is uniformly distributed on $(0,1)$, that is, $P(Y \le y) = y, 0 < y < 1$.*

X continuous w/ cdf $F_X(x)$

$Y = F_X(x)$

$Y \sim U(0,1) \quad P(Y \le y) = y, \ 0 < y < 1.$

Example: Assume $X \sim$ exponential

$$f_X(x) = \frac{1}{\lambda} e^{-x/\lambda} I_{(0,\infty)}(x)$$

$$F_X(x) = \int_0^x \frac{1}{\lambda} e^{-t/\lambda} dt = 1 - e^{-x/\lambda}$$

[let $u = -t/\lambda$

$\frac{du}{dt} = -1/\lambda \quad du = -1/\lambda \, dt$

$\int_0^x \frac{1}{\lambda} e^{-t/\lambda} dt = \int -e^u du = -e^u \Big|$

$= -e^{-t/\lambda}\Big|_0^x = -e^{-x/\lambda} + 1$]

Let $Y = 1 - e^{-x/\lambda}$

Find dist'n of $Y = F_x(x)$

- Show monotonic ft'n:

$$\frac{d}{dx}(1 - e^{-x/\lambda}) = \frac{e^{-x/\lambda}}{\lambda} > 0 \Rightarrow \text{monotonic}$$

[$\frac{d}{dx}(-e^{-x/\lambda}) = (-)(-1/\lambda) e^{-x/\lambda}$]

- Find $g^{-1}(y)$: Solve for x in terms of y

$$x = -\lambda \log(1-y) = g^{-1}(y)$$

[$e^{-x/\lambda} = 1 - y$

$-\frac{x}{\lambda} = \log(1-y)$

$x = -\lambda \log(1-y)$]

- Determine sample space, $\mathcal{Y}$

$$0 \le Y \le 1$$

[If $x = 0 \quad y = 0$

$x \to \infty \quad y \to 1$]

- Calculate $\frac{d}{dy}(g^{-1}(y))$ (derivate old wrt new)

$$\frac{d(g^{-1}(y))}{dy} = \frac{d}{dy}(-\lambda \log(1-y))$$

[$\frac{d}{dy}(-\lambda \log(1-y))$

$= \frac{-\lambda}{(1-y)}(-1) = \frac{\lambda}{1-y}$]

- In $f_X(x)$ replace x by $g^{-1}(y)$ multiply by $\left|\frac{d}{dy} g^{-1}(y)\right|$ and identify $\mathcal{Y}$

pdf of Y $\Rightarrow f_Y(y) = 1 \, I_{(0,1)}(y)$

$$Y \sim U(0,1)$$

[$f_Y(y) = \frac{1}{\lambda} \exp\{-\cancel{\lambda} \log(1-y)/\cancel{\lambda}\} \Big| * \Big| \frac{\lambda}{1-y} \Big|$

$= \frac{1}{\lambda} (1-y) \Big| \frac{\lambda}{1-y} \Big| = 1$]

Go Backwards: Find dist'n $X = \lambda \log(1-Y) = F_X^{-1}(Y)$

If $Y \sim U(0,1)$, Find dist'n of $X = -\lambda \log(1-Y)$

switch $\overset{\text{new}}{X} = -\lambda \log(1-\overset{\text{old}}{Y})$

- Show monotonic ft'n:

$$\frac{d}{dy}(-\lambda \log(1-y)) = \frac{\lambda}{1-y} > 0 \quad \text{for } 0<y<1$$

[previous page $\frac{d}{dy}(-\lambda\log(1-y)) = \frac{d}{dy} g^{-1}(y)$]

- Find $g^{-1}(x)$: Solve for y in terms of x

$$y = 1 - e^{-x/\lambda}$$

[$y = F_X(x)$ previous page]

- Determine sample space, $\mathcal{X}$

$$0 \le X < \infty$$

[As expected if $y=0$ $x=0$; $y \to 1$ $x \to \infty$]

- Calculate $\frac{d}{dx}(g^{-1}(x))$

$$\frac{d}{dx}(1 - e^{-x/\lambda}) = \frac{e^{-x/\lambda}}{\lambda}$$

[previous page $\frac{d}{dx} g(x)$]

- In $f_Y(y)$ replace y by $g^{-1}(x)$ multiply by $\left|\frac{d}{dx} g^{-1}(x)\right|$ and identify $\mathcal{X}$

$$f_X(x) = 1 * \left|\frac{e^{-x/\lambda}}{\lambda}\right| * I_{(0,\infty)}(x)$$

$X \sim$ exponential (λ)

Why so important? If you can generate a $U(0,1)$, you can generate random variables from any continuous dist'n.

To generate exp (λ) $(n=10{,}000)$

generate 10,000 values of $Y \sim U(0,1)$

calculate $X = -\lambda \log(1-Y)$

$\Rightarrow$ 10,000 exp (λ).

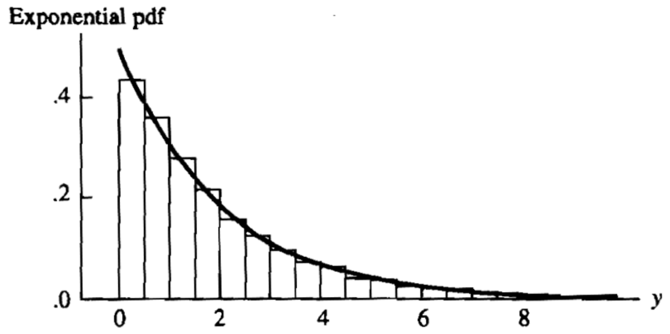


Figure 5.6.1. *Histogram of 10,000 observations from an exponential pdf with* $\lambda = 2$, *together with the pdf*

The relationship between the exponential and other distributions allows the quick generation of many random variables. For example, if U_j are iid uniform$(0,1)$ random variables, then $Y_j = -\lambda \log(u_j)$ are iid exponential (λ) random variables and

$$Y = -2\sum_{j=1}^{\nu} \log(U_j) \sim \chi^2_{2\nu}, \quad \text{— Note df} = 2\nu\text{; only generate even df}$$

(5.6.5)
$$Y = -\beta\sum_{j=1}^{a} \log(U_j) \sim \text{gamma}(a,\beta), \quad \leftarrow \text{Note } \alpha = \text{integer}$$

$$Y = \frac{\sum_{j=1}^{a} \log(U_j)}{\sum_{j=1}^{a+b} \log(U_j)} \sim \text{beta}(a,b). \quad \leftarrow \text{Note } a, b \text{ integer.}$$

Other methods covered in C&B.

Not in C&B:

Hmm... If software will generate $N(0,1)$ how can you use this to generate a $N(\mu,\sigma^2)$?

Things that make you say 'hmmm'.

- Flip a coin 3 times

$S =$

HHH	HHT	HTH	THH
HTT	THT	TTH	TTT

$P(H) = p$

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

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

$S =$

HHH	HHT	HTH	THH
HTT	THT	TTH	TTT

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

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

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

$$\begin{aligned} P(HHT \mid X=2) &= P(HHT \text{ and } X=2)/P(X=2) \\ &= P(HHT)/P(X=2) \\ &= [p^2 q] / [\tbinom{3}{2} p^2 q] \\ &= 1/3 \end{aligned}$$

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!!

'Hmmm' cont.

Assume $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, 1)$

easier to think about in coin flip example

for: $\underline{x}$ consistent w/ $\bar{x}$. different samples can have same $\bar{x}$ value.

find $f(\underline{x} \mid \bar{X} = \bar{x}) = \dfrac{f(\underline{x} \text{ and } \bar{x})}{f(\bar{x})} = \dfrac{f(\underline{x})}{f(\bar{x})} = \dfrac{\prod_{i=1}^{n} \left(\frac{1}{\sqrt{2\pi}} e^{-(x_i - \mu)^2/2} \right)}{\frac{\sqrt{n}}{\sqrt{2\pi}} e^{-\frac{n}{2}(\bar{x}-\mu)^2}}$

else: $f(\underline{x} \text{ and } \bar{x}) = 0$

$$\frac{f(\underline{x} \mid \mu)}{f(\bar{x} \mid \mu)} = \frac{\left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\Sigma(x_i^2 - 2\mu x_i + \mu^2)}}{\frac{\sqrt{n}}{\sqrt{2\pi}} e^{-\frac{1}{2}(n\bar{x}^2 - 2\mu\bar{x} + n\mu^2)}}$$

$$= \underbrace{\frac{\sqrt{2\pi}}{\sqrt{n}} \left(\frac{1}{\sqrt{2\pi}}\right)^n}_{K} e^{-(\Sigma x_i^2 - n\bar{x}^2)/2} \, e^{2\mu[\Sigma x_i - n\bar{x}]/2} \, e^{-n/2\mu^2 + n\mu^2/2}$$

$$= K e^{-(\Sigma x_i^2 - n\bar{x}^2)/2} \, e^{\mu[\overset{0}{\Sigma x_i - \Sigma x_i}]} \, e^{\overset{0}{-n\mu^2/2 + n\mu^2/2}}$$

$$= \underbrace{K e^{-(\Sigma x_i^2 - n\bar{x}^2)/2}}$$

'Hmmm' the conditional dist'n (given $\bar{X}$) is independent of μ!!

Chapter 6: Principles of Data Reduction

Sample $X_1, \dots, X_n$ and wish to make inferences about an unknown parameter: θ.

Summarize using statistics - ft'ns of observed data.

($\bar{X}, S^2, X_{(1)}, X_{(n)} \dots$)

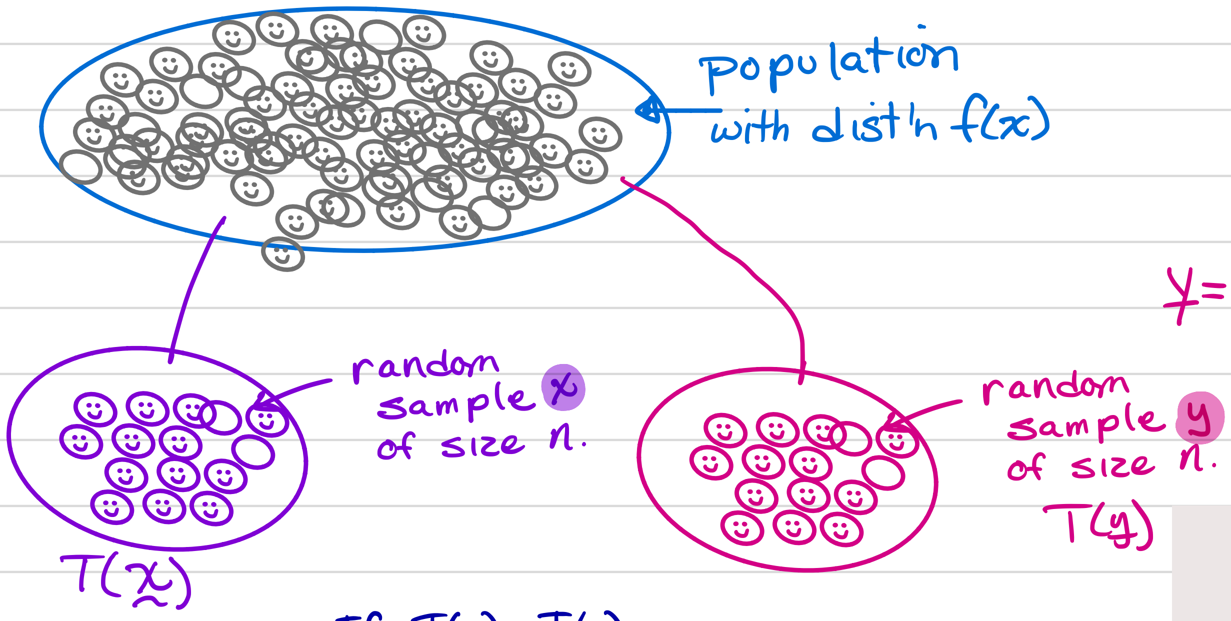
	Sample - 1	Sample - 2
$\underline{X} = X_1, \dots, X_n$	$\underline{x} = x_1, \dots, x_n$	$y_1, \dots, y_n$
$T(\underline{X})$ = statistic (random variable)	$T(\underline{x})$ = observed value.	$T(\underline{y})$ = observed value

If $T(\underline{x}) = T(\underline{y})$

make same conclusions even if $\underline{x}$ and $\underline{y}$ differ in some ways

Sufficiency: A sufficient statistic for unknown parameter θ is a statistic that captures <u>all</u> <u>of</u> <u>the</u> <u>information</u> <u>about</u> <u>θ</u> contained in the sample.

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.



$\underset{\sim}{X}$ = CNN political poll.

$\underset{\sim}{Y}$ = ABC News political poll

Wolf Blitzer

If $T(\underset{\sim}{x}) = T(\underset{\sim}{y})$
make same inference
whether observe $\underset{\sim}{x}$ or $\underset{\sim}{y}$.

Nate Silver

§6.2.1 Sufficient statistics

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.

(T is for Sufficient)

"Hmmm!!!"

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

$S =$

HHH	HHT	HTH	THH
HTT	THT	TTH	TTT

Given $X = 2$

HHT, HTH and THH are equally likely $= 1/3$. This is same if $p = 9/10$ or or $p = 1/100$ or ...

Consider discrete case:

Let t be a possible value of $T(\underset{\sim}{x})$ s.t. $P_\theta(T(\underset{\sim}{X}) = t) > 0$

In def'n the conditional dist'n of $\underset{\sim}{X}$ given t does not depend on θ.

If $T(\underset{\sim}{x}) \neq t$ for sample point $\underset{\sim}{x}$ $P(\underset{\sim}{X} = \underset{\sim}{x} \mid T(\underset{\sim}{X}) = t) = 0$

Interested in

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

By def'n $T(\underline{X})$ is a sufficient statistic, if this conditional probability is the same for all values of θ for any fixed values of $\underline{x}$ and t.

$$P_\theta(\underline{X}=\underline{x} \mid T(\underline{X})=t) \text{ is same for all } \theta.$$

Need to verify: $P_\theta(\underline{X}=\underline{x} \mid T(\underline{X})=T(\underline{x}))$ does not depend on θ.

Note: $\{\underline{X}=\underline{x}\}$ is a subset of $\{T(\underline{X})=T(\underline{x})\}$

HHT is subset of $X=2$

HHH	HHT	HTH	THH
HTT	THT	TTH	TTT

→ Find condition for which

$$P_\theta(\underline{X}=\underline{x} \mid T(\underline{X})=t) \text{ is same for all } \theta.$$

$$P_\theta(\underline{X}=\underline{x} \mid T(\underline{X})=t) = \frac{P_\theta(\underline{X}=\underline{x} \text{ and } T(\underline{X})=T(\underline{x}))}{P_\theta(T(\underline{X})=T(\underline{x}))}$$

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

$$= \frac{p(\underline{x}|\theta)}{q(T(\underline{x})|\theta)}$$

$p(\underline{x}|\theta)$ = joint pmf of $\underline{X}$

$q(T(\underline{x})|\theta)$ = pmf of $T(\underline{X})$.

Thus, $T(\underline{X})$ is a sufficient statistic for θ, iff for every $\underline{x}$

$$\frac{p(\underline{x}|\theta)}{q(T(\underline{x})|\theta)} \text{ is constant wrt } \theta.$$

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.*

Example: $X_1, \dots, X_n$ iid Bernoulli(θ) $0<\theta<1$.

Is $T(\underline{X}) = X_1 + \dots + X_n$ a sufficient statistic for θ?

know $\Sigma X_i \sim \text{bin}(n, \theta)$ and let $\Sigma x_i = t$

$$\therefore \frac{p(\underline{x}|\theta)}{q(T(\underline{x})|\theta)} = \frac{\prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} I_{[0,1]}(x_i)}{\binom{n}{t}\theta^{t}(1-\theta)^{n-t} I_{[0,1,\dots,n]}(t)}$$

$$= \frac{\theta^{\Sigma x_i}(1-\theta)^{n-\Sigma x_i} I_{[0,1,\dots,n]}(\Sigma x_i)}{\binom{n}{t}\theta^{t}(1-\theta)^{n-t} I_{[0,1,\dots,n]}(t)}$$

$$= \frac{\cancel{\theta^{t}}\,\cancel{(1-\theta)^{n-t}}\,\cancel{I_{[0,1,\dots,n]}(t)}}{\binom{n}{t}\cancel{\theta^{t}}\,\cancel{(1-\theta)^{n-t}}\,\cancel{I_{[0,1,\dots,n]}(t)}}$$

$$= 1/\binom{n}{t} = 1/\binom{n}{\Sigma x_i}$$

$$\therefore \frac{p(\underline{x}|\theta)}{q(T(\underline{x})|\theta)} \perp \theta \implies T(\underline{X}) = \Sigma X_i \text{ is a sufficient stat.}$$

$\implies$ Total number of successes in a Bernoulli sample contains all info about θ in the data.

$\Sigma x_i = 2$
Make same conclusion if HHT or HTH or THH is observed.

HHH	HHT	HTH	THH
HTT	THT	TTH	TTT

Chapter 6: Principles of Data Reduction

Example: Assume that we are interested in estimating the latency period distribution of Bovine Spongiform Encephalopathy (BSE) (a.k.a. Mad Cow Disease.)

Cow skiing a Black Diamond at Aspen

Experiment: Feed newborn calves brain from infected animals

Assume: All cows will eventually develop disease (i.e. no competing risks)

All cows receive same 'dose'. $\mathcal{X} = \{1, 2, 3 \ldots 9, \ldots\}$

fake Pseudo Data 1: $(n=50)$

Cow	Age	Cow	Age	Cow	Age	Cow	Age	Cow	Age
1	5	11	6	21	5	31	5	41	4
2	4	12	5	22	4	32	4	42	5
3	6	13	4	23	5	33	4	43	3
4	6	14	6	24	3	34	5	44	5
5	7	15	6	25	4	35	4	45	6
6	4	16	8	26	5	36	6	46	4
7	5	17	5	27	9	37	7	47	5
8	5	18	4	28	5	38	5	48	6
9	7	19	4	29	9	39	6	49	7
10	8	20	7	30	6	40	7	50	8

Pseudo Data 2: $(n=50)$

Cow	Age	Cow	Age	Cow	Age	Cow	Age	Cow	Age
1	3	11	4	21	5	31	6	41	7
2	3	12	4	22	5	32	6	42	7
3	4	13	4	23	5	33	6	43	7
4	4	14	4	24	5	34	6	44	7
5	4	15	5	25	5	35	6	45	7
6	4	16	5	26	5	36	6	46	8
7	4	17	5	27	5	37	6	47	8
8	4	18	5	28	5	38	6	48	8
9	4	19	5	29	5	39	6	49	9
10	4	20	5	30	6	40	7	50	9

Cow sheltering from English Weather

```
        (__)
        (oo)
 /-------\/
/ |     ||
* ||----||
  ~~    ~~
```

Cownt Dracula Cow

Let $X_1, \dots, X_n$ be a sample from a dist'n with unknown parameter $\underset{\sim}{\theta}$.
Assume cows: $\underset{\sim}{\theta} = (p_1, p_2, \dots p_9)$ from a multinomial dist'n.

Data Reduction: summarize information in a sample using statistics (counts).
Goal: Fewest number of summary stats, without losing info.

Data Reduction on Pseudo Data 1:

Age	1	2	3	4	5	6	7	8	9
# Dx	0	0	2	12	15	10	6	3	2
Prop Dx	0	0	.04	.24	.30	.20	.12	.06	.04

```
 /          \
( deja mooo  )
 \__  ______/
    ) )
   //
  /   (__)
      (oo)
 /-------\/
/ |     ||
* ||----||
  ~~    ~~
```

Data Reduction on Pseudo Data 2:

Age	1	2	3	4	5	6	7	8	9
# Dx	0	0	2	12	15	10	6	3	2
Prop Dx	0	0	.04	.24	.30	.20	.12	.06	.04

The 'raw' data and the 'reduced' data have the same information about the latency dist'n.

This implies that Data sets 1 + 2 have identical information.

and any other data set that has identical summary stats.

$\rightarrow$ Draw same conclusion from 'pseudo data 1' & 'pseudo data 2'

All data sets that fall into same partition of the sample space will have same information (make same inferences or conclusions).

- Think of summary stats as a way of grouping or partioning the sample space $\mathcal{X}$.
- Let $\mathcal{T} = \{t : t = T(\underset{\sim}{x}) \text{ for some } \underset{\sim}{x} \in \mathcal{X}\}$ be image of $\mathcal{X}$ under $T(\underset{\sim}{x})$.
- Partition into sets A_t, $t \in \mathcal{T}$

  s.t. $A_t = \{\underset{\sim}{x} : T(\underset{\sim}{x}) = t\}$

Sufficiency is basically a property of partitioning the sample space.

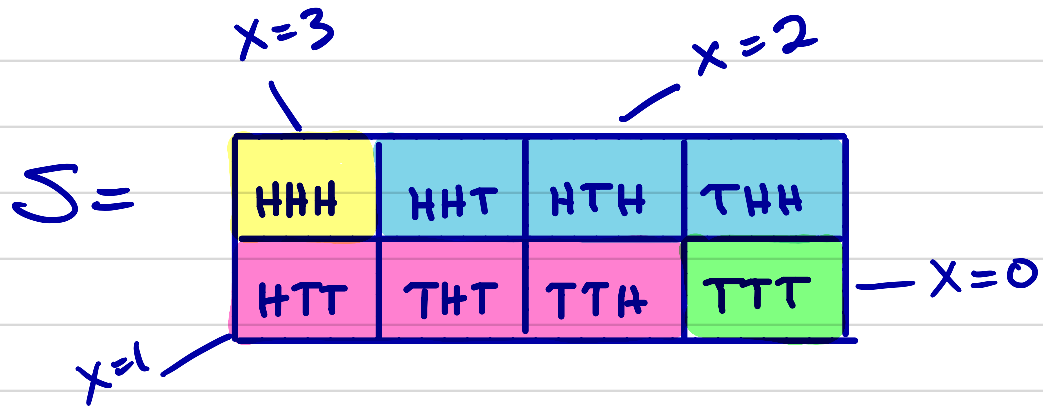
- Given the partition, the conditional dist'n of $\underline{X}$ does not depend on θ. ("Hmmm")
- Same info in $T(\underline{x})$ (partition) as in $\underline{x}$ (sample point).

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

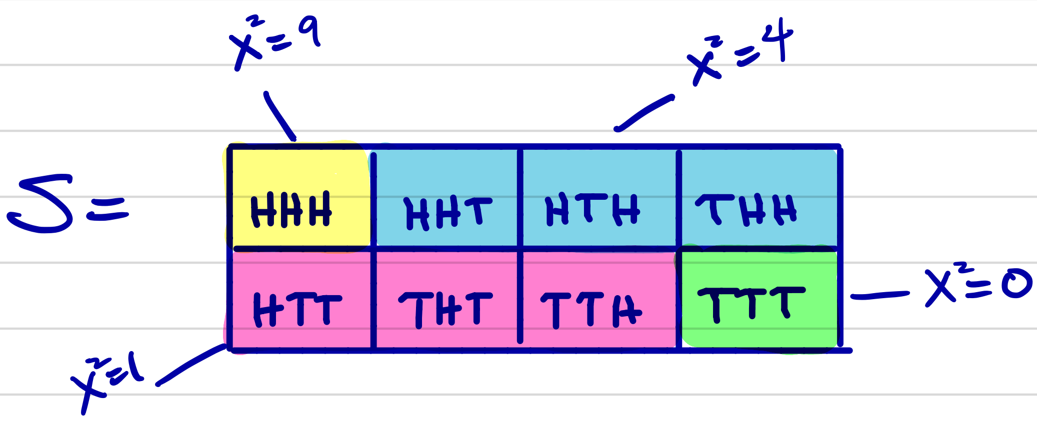
$$X_i = \begin{cases} 0 & \text{if tails} \\ 1 & \text{if heads} \end{cases} \Bigg\} \; X = \Sigma X_i$$

Coin Flip

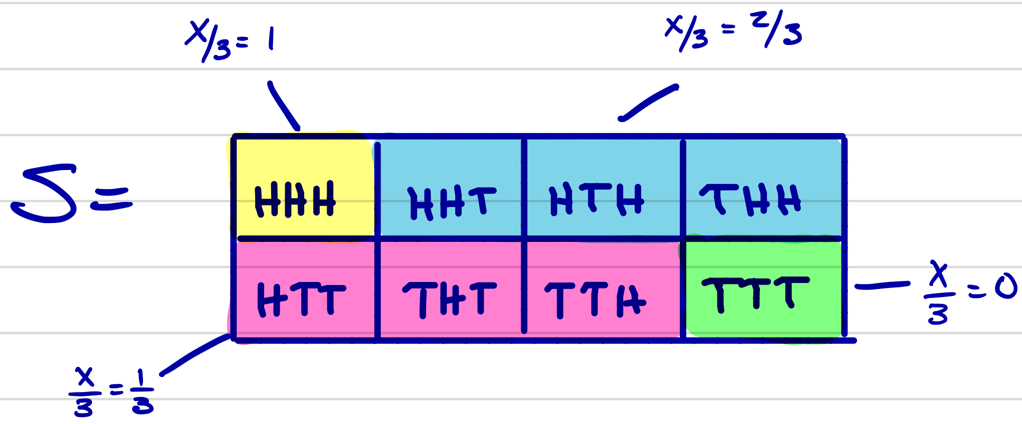
Partition Sample space by $T(\underline{x}) = \Sigma X_i$



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



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



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

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.

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.*

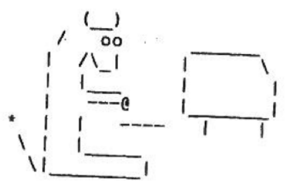
Goal: Find sufficient statistic using Thm 6.2.2

'intuition'

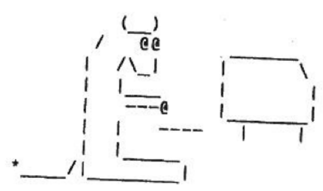
Step 1) "Guess" statistic $T(\underline{X})$ to be sufficient

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

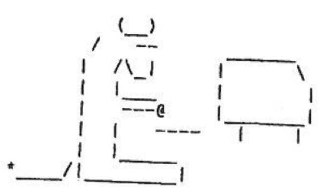
Step 3) Check ratio to determine if independent of θ



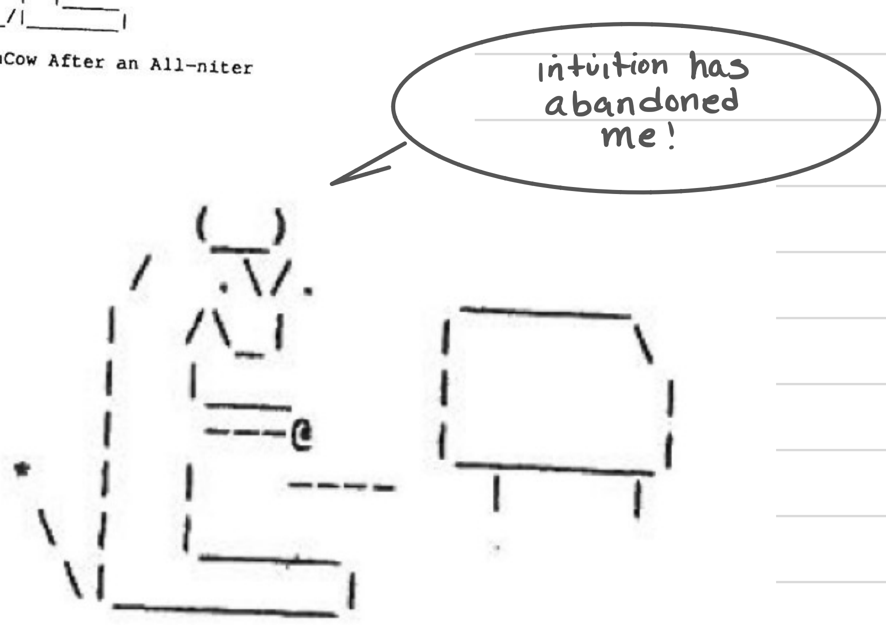
CompuCow



CompuCow After an All-niter



CompuCow Asleep at the Keyboard



Compu Cow 'guesses' wrong sufficient statistic.

Next time:

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