Chap 6

§6.1

- Any statistic, $T(\underline{X})$, defines a form of data reduction or data summary

- We treat two samples as equals if $T(\underline{x}) = T(\underline{y})$

- Let $\mathcal{X}$ be the sample space
- Let $\mathcal{T} = \{ t \mid t = T(\underline{x})$ for some $x \in \mathcal{X} \}$ be the image of $\mathcal{X}$ under T
- Then $T(\underline{x})$ partitions the sample space into sets $A_t,\ t \in \mathcal{T}$ defined by $A_t = \{\underline{x} \mid T(\underline{x}) = t\}$

insert drawings from notes

§6.2

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

* If $T(\underline{X})$ has a continuous distribution then $P_\theta(T(\underline{X}) = t) = 0$ for all values of t

* Let t be a possible value of $T(\underline{X})$ ie $\ni$ $P_\theta(T(\underline{X}) = t) > 0$. We wish to consider the conditional probability $P_\theta(\underline{X} = x \mid T(\underline{X}) = t)$

- * if $T(\underline{x}) \neq t \rightarrow P(\underline{X} = \underline{x} \mid T(\underline{x}) = t) = 0$
- * thus we are truly interested in $P_\theta(\underline{X} = \underline{x} \mid T(\underline{X}) = T(\underline{x}))$
  - * Sufficient if the above probability is the same for all values of θ

HAPPY STUDYING!
☺ -GEORGINA

* A sufficient statistic captures all the information about θ in this sense.
* To show sufficient we must show $P_\theta(\underline{X}=\underline{x} \mid T(\underline{X})=T(\underline{x}))$ does not depend on θ.

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

Example 6.2.3: Let $X_1, X_2, \dots X_n \overset{iid}{\sim}$ Bernoulli(θ), $0<\theta<1$

- Show $T(\underline{X}) = X_1 + \cdots + X_n$ is sufficient for θ

$T(\underline{X}) = \#$ of X_i's equal to one

$\therefore T(\underline{X}) \sim$ binomial(n,θ)

- To show sufficiency look at:

$$\frac{p(x|\theta)}{q(T(\underline{x})|\theta)} = \frac{\prod \theta^{x_i}(1-\theta)^{1-x_i}}{\binom{n}{t}\theta^t(1-\theta)^{n-t}} \qquad \left(t = \sum x_i\right)$$

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

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

$$= \frac{1}{\binom{n}{t}} = \frac{1}{\binom{n}{\sum x_i}}$$

* Since this ratio doesn't depend on θ, by theorem 6.2.2, $T(\underline{X})$ is a sufficient statistic for θ
i.e. The total # of 1s in this Bernoulli sample contains all the information about θ in this data

If $\underset{\sim}{X} \sim N(\mu, \sigma^2)$
$\bar{X} \sim N(\mu, \sigma^2/\ )$

Example 6.2.4 Let $X_1, X_2, \dots X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ where σ^2 is known
Show $T(\underset{\sim}{X}) = \bar{X} = \frac{1}{n}\sum_1^n (X_1 + X_2 + \cdots + X_n)/n$ sufficient for μ

$$f_{\underset{\sim}{x}}(\underset{\sim}{x} \mid \mu, \sigma^2) = \prod_1^n \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} I(x_i \in \mathbb{R})$$
$$= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\sum_1^n \left(\frac{(x_i-\mu)^2}{2\sigma^2}\right)} I(x_i \in \mathbb{R}\ \forall\ i = 1, 2, 3, \dots n)$$

$$f_{T(x)}(t) = \left(\frac{1}{\frac{\sigma}{\sqrt{n}}\sqrt{2\pi}}\right) e^{-\frac{(t-\mu)^2}{2\sigma^2}} \quad I(t \in \mathbb{R})$$

$$\frac{f_{\underset{\sim}{x}}(t)}{f_{T(x)}(t)} = \frac{\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\sum_1^n \left(\frac{(x_i-\mu)^2}{2\sigma^2 n}\right)}}{\left(\frac{1}{\frac{\sigma}{\sqrt{n}}\sqrt{2\pi}}\right) e^{-\frac{(t-\mu)^2}{2(\sigma^2/n)}}} \qquad \bar{X} \sim N(\mu, \sigma^2/n)$$

$$= \left(\frac{1}{1/\sqrt{n}}\right) \frac{\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\sum_1^n \left(\frac{(x_i-\mu)^2}{2\sigma^2}\right)}}{\left(\frac{1}{\sigma\sqrt{2\pi}}\right) e^{-\left(\frac{(t-\mu)^2}{2(\sigma^2/n)}\right)}}$$

$$= \frac{1}{\sqrt{n}} \cdot \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^{n-1} e^{\left(\frac{(t-\mu)^2}{2(\sigma^2/n)} - \sum_1^n \left(\frac{(x_i-\mu)^2}{2\sigma^2}\right)\right)}$$

* as σ is known we want to get rid of μ to show sufficiency of T for μ

How w/ this part?

$$= \frac{1}{\sqrt{n}} \cdot \left(\frac{1}{\sigma\sqrt{2\pi}}\right) e^{\left(\frac{n(\bar{x}-\mu)^2}{2\sigma^2}\right) - \sum_1^n \left(\frac{(x_i-\bar{x})^2 + (\bar{x}-\mu)^2}{2\sigma^2}\right)}$$

$$= \frac{1}{\sqrt{n}} \left(\frac{1}{\sigma\sqrt{2\pi}}\right) e^{-\sum_1^n \frac{(x_i-\bar{x})^2}{2\sigma^2}} \quad I(t, x_i \in \mathbb{R})$$

$\therefore$ T is sufficient for μ

Theorem 6.2.6

Factorization theorem - Let $f(\underline{x}|\theta)$ denote the joint pdf or pmf of a sample $\underline{X}$. A statistic $T(\underline{X})$ is a sufficient statistic for θ iff $\exists$ $g(t|\theta)$ and $h(x)$ $\ni$ $\forall$ sample points $\underline{x}$ and all parameters θ

$$f_{\underline{x}}(\underline{x}|\theta) = g(T(\underline{x})|\theta)h(\underline{x})$$

Proof p. 276

Example 6.2.7 for $X_1, X_2, \ldots X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ w/ known σ^2

we can factor $f_{\underline{x}}(\underline{x}|\theta)$

$$f_{\underline{x}}(\underline{x}|\theta) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\left(\sum_1^n (x_i-\bar{x})^2/(2\sigma^2)\right) - \left(n(\bar{x}-\mu)^2/2\sigma^2\right)}$$

$$h(x) = e^{-\left(\sum_1^n (x_i-\bar{x})^2/2\sigma^2\right)}$$

$$g(t|\theta) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-n(\bar{x}-\mu)^2/2\sigma^2}$$

$T(x) = \frac{1}{n}\sum_1^n x_i$ is sufficient for μ

Example 6.2.8 · Let $X_1, X_2, \ldots X_n \overset{iid}{\sim}$ Discreet Uniform on $1, 2, \ldots \theta$

· Thus, $f_x(x|\theta) = \frac{1}{\theta} I(x = 1, 2, \ldots \theta)$

· Thus, $f_{\underline{x}}(\underline{x}|\theta) = \frac{1}{\theta^n} I(x_i = 1, 2, \ldots \theta \ \forall i = 1, 2, \ldots n)$

$= \frac{1}{\theta^n} I(x_i = 1, 2, \ldots \ \forall i = 1, 2, \ldots n) I(X_{(n)} \le \theta)$

$h(x) = I(x_i = 1, 2, \ldots \ \forall i = 1, 2, \ldots n)$

$g(t|\theta) = \frac{I(X_{(n)} \le \theta)}{\theta} = \frac{I(t \le \theta)}{\theta}$ for $T(\underline{X}) = X_{(n)}$

$\therefore$ By Factorial theorem $X_{(n)}$ is sufficient for θ

Example 6.2.9 · Let $X_1, X_2, \ldots X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ $\underline{\theta} = (\mu, \sigma^2)$

· Thus $f_x(x|\theta) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$

· Thus $f_{\underline{x}}(\underline{x}|\theta) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\sum(x_i-\mu)^2/2\sigma^2} = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\sum_1^n (x_i-\bar{x}+\bar{x}-\mu)^2/2\sigma^2}$

$= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{\left(-\sum_1^n (x_i-\bar{x})^2\right) - n(\bar{x}-\mu)^2/2\sigma^2}$

$$g(t|\theta) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{\frac{-(n-1)t_1 - n(t_2-\mu)^2}{2\sigma^2}}$$

$$T(\underline{X}) = \begin{bmatrix} S^2 \\ \bar{X} \end{bmatrix}$$

Theorem 6.2.10 Let $X_1, X_2, \ldots X_n$ be iid observations from a pdf or pmf $f(x|\theta)$ that belongs to an exponential family given by

$$f(x|\underline{\theta}) = h(x)\, c(\underline{\theta}) \exp\left(\sum_i^k w_i(\underline{\theta})\, t_i(x)\right) \quad \text{where } \dim(\theta) < k$$

THEN

$$T(\underline{X}) = \left(\sum_{j=1}^n t_1(x_j), \sum_j^n t_2(x_j), \ldots, \sum_j^n t_k(x_j)\right) \text{ is sufficient for } \theta$$

** $T(\underline{X}) = \underline{x}$ is always a sufficient statistic **

** Any 1-1 function of a sufficient statistic is ** also sufficient

Def 6.2.11 A sufficient statistic $T(X)$ is called a minimal sufficient statistic if, for any other sufficient statistics T', $T(X)$ is a function of T'. * minimal stat provides coarsest partition of $\mathcal{X}$

$$\text{ie: if } T'(x) = T'(y) \rightarrow T(x) = T(y)$$

Theorem 6.2.13 Let $f(\underline{x}|\theta)$ be the pmf or pdf of a sample $\underline{X}$. Suppose $\exists$ a function $T(\underline{x})$, $\ni$ $\forall$ two sample points x and y, the ratio $f(\underline{x}|\theta)/f(\underline{y}|\theta)$ is constant as a function of θ iff $T(\underline{x}) = T(\underline{y})$. Then $T(\underline{x})$ is a minimal sufficient stat for θ

Proof p 281

Example 6.2.14 $X_1, X_2, \ldots X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ let $\underline{x}$ and $\underline{y}$ denote two sample points w/ means and variances $(\bar{x}, S_x^2)$, $(\bar{y}, S_y^2)$

Consider

$$\frac{f_{\underline{x}}(\underline{x}|\mu,\sigma^2)}{f_{\underline{x}}(\underline{y}|\mu,\sigma^2)} = \frac{\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\Sigma(x_i-\mu)^2/2\sigma^2}}{\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\Sigma(y_i-\mu)^2/2\sigma^2}} = e^{\frac{\Sigma(y_i-\mu)^2 - \Sigma(x_i-\mu)^2}{2\sigma^2}}$$

$$= e^{\frac{\Sigma(y_i-\bar{y})^2 + n(\bar{y}-\mu)^2 - \Sigma(x_i-\bar{x})^2 - n(\bar{x}-\mu)^2}{2\sigma^2}}$$

$$= e^{\frac{n\left[\Sigma(\bar{y}-\mu)^2 - \Sigma(\bar{x}-\mu)^2\right] + \Sigma(y_i-\bar{y})^2 - \Sigma(x_i-\bar{x})^2}{2\sigma}}$$

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

$\therefore$ as this is only constant when $\bar{y} = \bar{x}$ and $S_y^2 = S_x^2$

$T(x) = \begin{bmatrix} \bar{x} \\ S_x^2 \end{bmatrix}$ is minimally sufficient

Example 6.2.15 : $X_1, X_2, \ldots X_n \overset{iid}{\sim}$ uniform$(\theta, \theta+1)$

$$f_X(x|\theta) = 1\, I(\theta < x < \theta+1)$$
$$f_{\underline{X}}(\underline{x}|\theta) = 1\, I(\theta < X_{(1)})\, I(\theta+1 > X_{(n)})$$

Notice : $h(x) = 1$ $\quad g(t|\theta) = I(\theta < X_{(1)})\, I(\theta+1 > X_{(n)})$

$\Rightarrow T(x) = [X_{(1)}, X_{(n)}]$ is a sufficient stat

Consider: $$\frac{f_X(\underline{x}|\theta)}{f_Y(\underline{y}|\theta)} = \frac{1\, I(\theta < X_{(1)})\, I(X_{(n)} < \theta+1)}{1\, I(\theta < Y_{(1)})\, I(Y_{(n)} < \theta+1)}$$

Thus, this is free of θ if $X_{(1)} = Y_{(1)}$ & $X_{(n)} = Y_{(n)}$

$\therefore$ $T(\underline{X}) = \begin{bmatrix} X_{(1)} \\ X_{(n)} \end{bmatrix}$ is minimally sufficient

** Any 1-1 function of a minimally sufficient statistic is **
also a minimally sufficient statistic

Def 6.2.16 A Statistic $S(\underline{X})$ whose distribution does not depend on the parameter θ is called an <u>ancillary statistic</u>

* An ancillary stat:
  - contains no information about θ
  - an observation on a RV whose distribution is fixed and known
  - unrelated to θ
  - in conjunction w/ other statistics, however, ancillary stats can provide info for inferences about θ

Example 6.2.17 $X_1, X_2, \dots X_n \overset{iid}{\sim}$ uniform $(\theta, \theta+1)$

Let $X_{(1)}, X_{(2)}, \dots X_{(n)}$ be ordered stats

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

$f_x(x) = \frac{1}{\theta+1-\theta} I(\theta < x < \theta+1) = 1\, I(\theta < x < \theta+1$

$F_x(x) = \int_\theta^{x_i} 1\, dx = x|_\theta^x = x - \theta \quad I(\theta < x_i < \theta+1)$

By theorem 5.4.6: $f_{x_{(1)}, x_{(n)}}(x_1, x_n) = \left[\frac{n!}{(n-2)!}(1)(1)(1)\right] * \left[x_1 - \theta - x_n + \theta\right]^{n-2} (1)$

$$\boxed{= n(n-1) * [x_1 - x_n]^{n-2} \quad I(\theta < x_1 < x_n < \theta+1}$$

$U = X_{(n)} - X_{(1)} \qquad V = X_{(1)}$

$X_{(1)} = V$

$X_{(n)} = U + V$

$$J = \begin{vmatrix} \frac{dx_{(1)}}{dv} & \frac{dx_{(1)}}{du} \\ \frac{dx_{(n)}}{dv} & \frac{dx_{(n)}}{du} \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ 1 & 1 \end{vmatrix} = 1$$

fix support

help!

$f_{u,v}(u,v) = n(n-1)[v - (u+v)]^{n-2}$

$= n(n-1)[-u]^{n-2}$

$f_u(u) = \int n(n-1)[-u]^{n-2} dv = n(n-1)[-u]^{n-2} v|$

$= n(n-1)(-u)^{n-2}(u-1)$

$U \sim$ beta$(n-1, 2)$

Note the distribution of U doesn't depend on θ thus it is ancillary.

Example 6.2.18 · Let $X_1, \dots X_n \overset{iid}{\sim}$ location family w/ CDF $F(x-\theta)$
$-\infty < \theta < \infty$

· Show $R = X_{(n)} - X_{(1)}$ is ancillary

· By theorem 3.5.6 $\exists\ Z$ w/ pdf $f(z)$ and $Z = X - \theta$
$\downarrow X = Z + \theta$

Thus

$$F_R(r|\theta) = P(R \le r) = P(X_{(n)} - X_{(1)} \le r)$$
$$= P((Z_{(n)} + \theta) - (Z_{(1)} + \theta) \le r)$$
$$= P(Z_{(n)} - Z_{(1)} \le r)$$

↑ doesn't depend on θ as Z doesn't depend on θ

$\therefore$ R is an ancillary statistic

Example 6.2.19 $X_1, X_2, \dots X_n \overset{iid}{\sim}$ scale family w/ CDF $F(\frac{x}{\sigma})$ $\sigma > 0$

· Show $R = \frac{X_{(1)} + X_{(2)} + \dots + X_{(n)}}{X_{(n)}}$ is ancillary

· By theorem 3.5.6 $\exists\ Z$ w/ pdf $f(z)$ and $Z = \sigma X$
$X = Z/\sigma$

Thus

$$F_R(r|\theta) = P(R \le r) = P\left(\frac{X_{(1)} + X_{(2)} + \dots + X_{(n)}}{X_{(n)}} \le r\right)$$
$$= P\left(\frac{\frac{1}{\sigma} Z_{(1)} + \frac{1}{\sigma} Z_{(2)} \cdots \frac{1}{\sigma} Z_{(n)}}{\frac{1}{\sigma} Z_{(n)}} \le r\right) = P\left(\frac{Z_{(1)} + Z_{(2)} + Z_{(n)}}{Z_{(n)}} \le r\right)$$

this doesn't depend on σ as Z doesn't depend on σ

$\therefore$ R is an ancillary statistic

Minimal Statistic - a statistic that has achieved maximal data reduction while retaining all the information in the sample

Ancillary Precision - Let X_1 and X_2 be iid observations from a discrete distribution that satisfies

$P_\theta(X = \theta) = P_\theta(X = \theta + 1) = P(X = \theta + 2) = \frac{1}{3}$ w/ unknown $\theta \in \mathbb{Z}$

· let $X_{(1)} \le X_{(2)}$ be the order stats for the sample

· $(R = X_{(2)} - X_{(1)},\ M = (X_{(1)} + X_{(2)})/2)$ is a minimally sufficient stat

· As this is a location family R is also ancillary

· The idea is that although R contains no information on θ it could provide valuable information toward the inference of θ

Def 6.2.21 Let $f(t|\theta)$ be a family of pdfs or pmfs for a statistic $T(\underset{\sim}{X})$. The family of probability distribution is called <u>complete</u> if $E_\theta(g(T))=0 \ \forall \theta$ implies $P_\theta(g(T)=0)=1 \ \forall \theta$, $T(x)$ is called a complete statistic

<u>Example 6.22</u>
- suppose $T \sim bin(n,p)$ $\quad 0<p<1$
- Let g be a function $\ni E_p(g(T))=0$

Then:
$$0 = E_p(g(T)) = \sum_{t=0}^{n} g(t)\binom{n}{t}p^t(1-p)^{n-t}$$
$$= (1-p)^n \sum_{t=0}^{n} g(t)\binom{n}{t}\left(\frac{p}{1-p}\right)^t \quad \forall \quad 0<p<1$$
$$\therefore \sum_{t=0}^{n} g(t)\underbrace{\binom{n}{t}}_{>0}\underbrace{\left(\frac{p}{1-p}\right)^t}_{>0} = 0$$
$$\therefore g(t) = 0 \quad \forall t$$
$\therefore T$ is a complete statistic

<u>Example 6.2.23</u>
- Let $X_1, X_2, \dots X_n \sim U(0,\theta)$ $\quad 0<\theta<\infty$
- $T(\underset{\sim}{X}) = \max_i X_i$
- $f(t|\theta) = nt^{n-1}\theta^{-n} I(0<t<\theta)$
- Suppose $g(t)$ is a function $\ni E_\theta(g(T))=0 \ \forall \theta$

$$0 = E_\theta(g(T)) = \frac{d}{d\theta}E_\theta(g(T)) = \frac{d}{d\theta}\int_0^\theta g(t)nt^{n-1}\theta^{-n}dt$$
$$= \theta^{-n}\frac{d}{d\theta}\left(\int_0^\theta g(t)nt^{n-1}dt\right) + \left(\frac{d}{d\theta}(\theta^{-n})\right)\int_0^\theta ng(t)t^{n-1}dt$$
$\uparrow E_\theta(g(t))$
$$= \theta^{-n}ng(\theta)\theta^{n-1} + 0$$
$$= \underbrace{\theta^{-1}}_{>0}\underbrace{n}_{>0}g(\theta)$$
$$\therefore g(\theta) = 0$$
$\therefore T$ is complete

Theorem 6.2.24 — **Basu's Theorem** If $T(X)$ is complete and minimal sufficient then $T(X)$ is independent of every ancillary statistic

Proof p 287

Theorem 6.2.25 — **Complete stats of exp family:**
- $X_1, X_2, \dots X_n$ be iid from an exponential family ie its pdf or pmf
$$f(x|\underline{\theta}) = h(x)c(\underline{\theta})\exp\left(\sum_1^k w_j(\theta)t_j(x)\right)$$
where $\underline{\theta} = \theta_1, \theta_2, \dots \theta_k$
- Then the statistic
$$T(\underline{x}) = \left[\sum_1^n t_1(x_i), \sum_1^n t_2(x_i), \dots \sum_1^n t_k(x_i)\right]$$
is complete if $\{w_1(\underline{\theta}), w_2(\underline{\theta}), \dots w_k(\underline{\theta})\} \; \theta \in \Theta$ contains an open set in $\mathbb{R}^k$

Example 6.2.26 - Using Basu's Theorem - I

- $X_1, X_2, \dots X_n$ iid exponential(θ)
- Consider $g(\underline{x}) = \frac{x_n}{x_1 + \dots x_n}$ is ancillary
- $\frac{1}{\theta}e^{-x/\theta} = \frac{1}{\theta}e^{-x(\frac{1}{\theta})} \Rightarrow T(x) = \Sigma x$ is a complete statistic 6.2.25
  is a sufficient stat 6.2.10
- $\theta = E_\theta(X_n) = E_\theta T(\underline{x})g(\underline{x}) = E_\theta(T(\underline{x}))(E_\theta g(\underline{x})) = n\theta E_\theta(g(\underline{x}))$

$\therefore E_\theta(g(\underline{x})) = \frac{1}{n}$

Example 6.2.27
- Consider independence of $\bar{x}$ and s^2, the sample mean and variance when sampling from $N(\mu, \sigma^2)$
- Let σ^2 be fixed and known, $-\infty < \mu < \infty$
- We know $\bar{x}$ is sufficient
- $6.2.5 \rightarrow N(\mu, \sigma^2/n)$ is complete which is the pdf of $\bar{x}$
- s^2 is ancillary (can be shown using location pdf

$\therefore$ By Basu's $\bar{x} \perp\!\!\!\perp s^2$

*ONLY FOR σ UNKNOWN]

theorem
6.2.28

If a minimal sufficient statistic exists then any complete statistic is also a minimal sufficient ~~stat~~,