Ancillary statistics

A statistic $S(\underline{X})$ is called ancillary if it provides no information about θ.

Example: $X_1, \dots, X_n \overset{iid}{\sim} n(\mu, 1)$

let $Z_i = \underbrace{X_i - \bar{X}}_{1}$

no information about μ from Z_is.

What is the distribution of Z_i?

Since Z_i is normal, its distribution is specified by its mean and variance

$$E[Z_i] = E[X_1 - \bar{X}]$$
$$= E[X_1] - E[\bar{X}]$$
$$= \mu - \mu = 0$$

$$Var(Z_i) = Var(X_i - \bar{X}) = var(X_1) + var(\bar{X}) - cov(X, \bar{X})$$
$$= Var\left(X_1 - \frac{1}{n}\sum_{i=1}^{n} X_i\right)$$
$$= Var\left(X_1 - \frac{1}{n}X_1 - \sum_{i=2}^{n} X_i\right)$$
$$= var\left(\underbrace{\left(\frac{n-1}{n}\right)X_1} - \underbrace{\sum_{i=2}^{n} X_i}\right)$$
$$= \left(\frac{n-1}{n}\right)^2 Var(X_1) + var\left(\sum_{i=2}^{n} X_i\right)$$
$$= \left(\frac{n-1}{n}\right)^2 \cdot 1 + (n-1) \cdot 1$$ doesn't depend on μ.

$$y_i \overset{iid}{\sim} U(0,1)$$
$$X_i = y_i + \theta$$

Another example $X_1, \dots, X_n \overset{iid}{\sim} U(\theta, \theta+1)$

$X_{(n)} - X_{(1)}$ is ancillary for θ.

Generally for location families,

$X_{(n)} - X_{(1)}$ is ancillary for θ

For scale familes

$$\frac{X_1}{X_n}, \dots, \frac{X_{n-1}}{X_n}$$

or $$\frac{X_1}{X_{(n)}}, \dots, \frac{X_n}{X_{(n)}}$$

are ancillary.

For $U(0,\theta)$ case

let $X_1, \dots X_n \overset{iid}{\sim} U(0,\theta)$

~~let~~ $X_i = \theta U_i$, $U_i \sim (0,1)$

$$\frac{X_i}{X_n} = \frac{\theta U_i}{\theta U_n} = \frac{U_i}{U_n}$$

exponentials are also scale families

If $X \sim \exp(1)$

$\theta X \sim$ exp w/mean θ.

Intuitively, ancillary statistics are "opposite" of sufficient statistics.

You might expect ancillary statistics to be independent of sufficient statistics. Often true, but not always.

The concept of complete statistics helps fill the gap. Ancillary statistics are independent of complete, sufficient statistics.

<u>Defn.</u> 6.2.21 Let $f(t|\theta)$ be a family of distributions for $T(\underline{X})$. The family is called <u>complete</u> if

$$E_\theta[g(T)] = 0 \text{ implies } P_\theta[g(T)=0] = 1$$

for all θ.

<u>Example</u> (check for completeness)

Let $X_1, \ldots, X_n \overset{iid}{\sim} U(0,\theta)$

$T(\underline{X}) = X_{(n)}$. We know $T(\underline{X})$ is sufficient (previous examples). We want to check for completeness.

$$f(t|\theta) = \begin{cases} \frac{nt^{n-1}}{\theta^n} & 0 < t < \theta \\ 0 & \text{otherwise} \end{cases}$$

we want to show that $T(\underline{X}) = X_{(n)}$ is a C.S.S.

Suppose g is a function satisfying

$$E[g(T)] = 0.$$

Since $E[g(T)]$ is constant w.r.t. θ

$$\frac{d}{d\theta} E[g(T)] = \frac{d}{d\theta} 0 = 0.$$

also $E[g(T)] = \int g(t) f(t|\theta)\, dt$

so

$$0 = \frac{d}{d\theta} 0 = \frac{d}{d\theta} E[g(T)] = \frac{d}{d\theta} \int_0^\theta g(t) f(t|\theta)\, d\theta$$

p. 69 Leibniz's Rule

$$= \frac{d}{d\theta} \int_0^\theta g(t)\, n t^{n-1} \theta^{-n}\, dt$$

$$\left[\frac{d}{d\theta} \int_0^\theta f(x,\theta)\, dx = f(x,\theta)\Big|_{x=\theta} + \int_0^\theta \frac{\partial}{\partial\theta} f(x,\theta)\, dx \right]$$

$$= g(\theta)\, n\, \theta^{n-1} \theta^{-n} + \theta \int_0^\theta \frac{\partial}{\partial\theta} g(t)\, n t^{n-1}$$

missing θ^{-n} ?

$$= g(\theta)\, n\, \theta^{n-1} = 0 \Rightarrow g(\theta) = 0.$$

Basu Thm. 6.2.24

If $T(\underline{X})$ is a complete s.s. then it is independent of every ancillary statistic.

Example $\bar{X}$ and S^2 ~~are~~ for a $n(\mu, 1)$ are independent because

$$S^2 = \frac{1}{n-1}\sum (X_i - \bar{X})^2 = \frac{1}{n-1}\sum Z_i^2$$

so S^2 is ancillary for μ.

$\bar{X}$ is a c.s.s. $\Rightarrow$ $\bar{X}$ is ind. of S^2.

Ex. let $X_1, \ldots, X_n \overset{iid}{\sim}$ ~~iid~~ Exp w. mean β.

Then $\dfrac{X_n}{\sum X_i}$ is ancillary for θ

$\dfrac{X_n}{\sum X_i}$ is the proportion of $\sum X_i$ contributed by X_n.

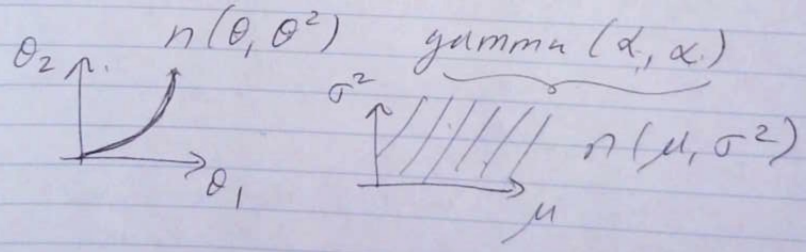
$\sum_{i=1}^{n} X_i$ is a c.s.s.

so $\sum X_i$ is ind. of $\dfrac{X_n}{\sum X_i}$

Thm 6.2.~~24~~25

If $X_1, \dots, X_n$ is an iid sample from an exp. fam., ~~[illegible]~~ with $\theta = (\theta_1, \dots, \theta_k)$ ~~[illegible]~~ then $T(\underline{X}) = \left(\sum_{i=1}^{n} t_1(X_i), \dots, \sum_{i=1}^{n} t_k(X_i)\right)$ is a c.s.s. as long as the parameter space Θ contains an open set in $\mathbb{R}^k$.

This thm doesn't apply to curved exponential families such as



Thm 6.2.28 If a minimal s.s. exists, then any c.s.s. is a m.s.s.

§ 6.3. Likelihood

If $f(\underline{x}/\theta)$ = the joint density or pmf for a sample

$L(\theta/\underline{x})$ is the likelihood.

$$L(\theta/\underline{x}) = \underbrace{f(\underline{x}/\theta)}$$

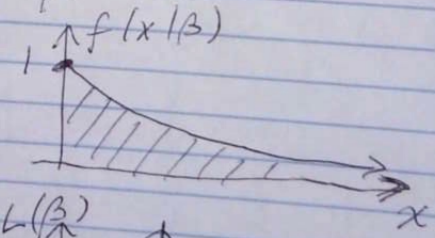
think of data as fixed parameter as variable

think of parameter as fixed, data as variable

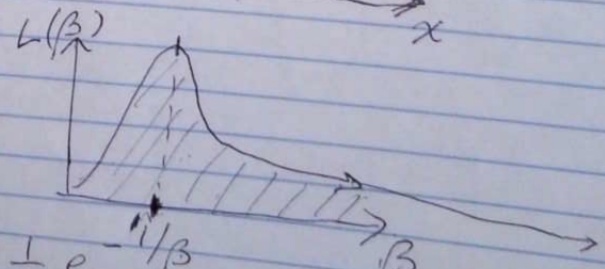
Suppose you have one observation of an exponential w. mean β.

$$f(x/\beta) = \frac{1}{\beta} e^{-x/\beta}.$$

let $\beta = 1$



$L(\beta/x)$



Suppose $x = 1$

$$L(\beta/x=1) = \frac{1}{\beta} e^{-1/\beta}$$

$$\int \frac{1}{\beta} e^{-1/\beta} \, d\beta$$