Chapter6

Principles of Data Reduction

Contents

Sufficient statistics

Data is collected to make inference about a parameter θ.

Whereas large samples make accurate inference concerning θ more feasible, in many situations much of the information given by the sample is unnecessary for inference.

In such situations, storing only relevant statistics (functions of the data) is more practical.

Examples

Suppose that we observe independent Bernoulli trials $X_1, \ldots, X_n$ and the success probability θ is unknown.

Do I really need to store the complete sample, or is the total number of successes sufficient information concerning θ?

Sufficient statistics

Let $\mathbf{X} \equiv X_1, \ldots, X_n$.

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

The idea is that if I know $T(\mathbf{X})$, inference concerning θ depends on $\mathbf{X}$ only through the value of $T(\mathbf{X})$. Thus given $T(\mathbf{X})$, I can discard the actual data.

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

Proof (discrete case):

$$P_\theta\left(\mathbf{X}=\mathbf{x}\middle|T(\mathbf{X})=T(\mathbf{x})\right)=\frac{P_\theta\left(\mathbf{X}=\mathbf{x},T(\mathbf{X})=T(\mathbf{x})\right)}{P_\theta\left(T(\mathbf{X})=T(\mathbf{x})\right)}$$

$$=\frac{P_\theta(\mathbf{X}=\mathbf{x})}{P_\theta\left(T(\mathbf{X})=T(\mathbf{x})\right)}=\frac{f(\mathbf{x}|\theta)}{q(T(\mathbf{x})|\theta)}.$$

Example 6.2.3

Let $X_1, \ldots, X_n$ be i.i.d. Bernoulli(θ) and $T(\mathbf{x}) = \sum_i x_i$.

By definition, $T \sim bin(n, \theta)$.

So, with $t = \sum x_i$,

$$\frac{f(\mathbf{x}|\theta)}{q(t|\theta)} = \frac{\prod_i \theta^{x_i}(1-\theta)^{1-x_i}}{\binom{n}{t}\theta^t(1-\theta)^{n-t}} = \frac{1}{\binom{n}{t}},$$ which

does not depend on θ. The total number of 1's is all the information about θ that is in the sample.

To use **Theorem 6.2.2,** we must guess a statistic $T(\mathbf{x})$ that is sufficient, determine its distribution $f(\mathbf{x}|\theta)$, and check that the ratio $f(\mathbf{x}|\theta)/q(T(\mathbf{x})|\theta)$ does not depend on θ.

The next theorem is useful because it allows us to determine a sufficient statistic by writing down the joint pdf or pmf and factoring it.

Factorization theorem: A statistic $T(\mathbf{x})$ is sufficient for θ if and only if there exists functions $g(t|\theta)$ and $h(\mathbf{x})$ such that, for all sample points $\mathbf{x}$ and parameter values θ, $f(\mathbf{x}|\theta) = g(T(\mathbf{x})|\theta)h(\mathbf{x})$.

Proof (discrete case)

Suppose that $T(\mathbf{x})$ is sufficient for θ. Choose $g(t|\theta) = P_\theta(T(\mathbf{X}) = t)$ and $h(\mathbf{x}) = P(\mathbf{X} = \mathbf{x}|T(\mathbf{X}) = T(\mathbf{x}))$ (this does not depend on θ by the sufficiency of $T(\mathbf{x})$).

Then

$$f\left(\mathbf{x}\middle|\theta\right)=P_\theta\left(\mathbf{X}=\mathbf{x}\right)=P_\theta\left(\mathbf{X}=\mathbf{x},T\left(\mathbf{X}\right)=T\left(\mathbf{x}\right)\right)$$

$$=P_\theta\left(T\left(\mathbf{X}\right)=T\left(\mathbf{x}\right)\right)P\left(\mathbf{X}=\mathbf{x}\middle|T\left(\mathbf{X}\right)=T\left(\mathbf{x}\right)\right)$$

$$g\left(T\left(\mathbf{x}\right)\middle|\theta\right)h\left(\mathbf{x}\right).$$

Now assume that the factorization $f\left(\mathbf{x}\middle|\theta\right)=g\left(T\left(\mathbf{x}\right)\middle|\theta\right)h\left(\mathbf{x}\right)$ exists. Then

$$\frac{f\left(\mathbf{x}\middle|\theta\right)}{q\left(T\left(\mathbf{x}\right)\middle|\theta\right)}=\frac{g\left(T\left(\mathbf{x}\right)\middle|\theta\right)h\left(\mathbf{x}\right)}{\sum_{\mathbf{y}:T(\mathbf{y})=T(\mathbf{x})}g\left(T\left(\mathbf{y}\right)\middle|\theta\right)h\left(\mathbf{y}\right)}$$

$$\frac{g\left(T\left(\mathbf{x}\right)\middle|\theta\right)h\left(\mathbf{x}\right)}{g\left(T\left(\mathbf{x}\right)\middle|\theta\right)\sum_{\mathbf{y}:T(\mathbf{y})=T(\mathbf{x})}h\left(\mathbf{y}\right)}=\frac{h\left(\mathbf{x}\right)}{\sum_{\mathbf{y}:T(\mathbf{y})=T(\mathbf{x})}h\left(\mathbf{y}\right)}$$

(The ratio doesn’t depend on θ.)

Example

Let $X_1, \ldots, X_n$ be i.i.d. Bernoulli(θ).

$$f(\mathbf{x}|\theta) = \theta^{\sum x_i}(1-\theta)^{n-\sum x_i} = \left(\frac{\theta}{1-\theta}\right)^{\sum x_i}(1-\theta)^n \cdot 1$$

Set $g(T(\mathbf{x})|\theta) = \left(\frac{\theta}{1-\theta}\right)^{\sum x_i}(1-\theta)^n$.

h(x)=1

Then $T(\mathbf{x}) = \sum x_i$ is sufficient for θ.

Example

Let $X_1, \ldots, X_n$ be i.i.d. Poisson(θ).

$$f(\mathbf{x}|\theta) = \prod_i \frac{e^{-\theta}\theta^{x_i}}{x_i!} = \frac{e^{-n\theta}\theta^{\sum x_i}}{\prod_i x_i!} = \frac{e^{-n\theta}\theta^{n\bar{x}}}{\prod_i x_i!} \ .$$

Set $g(T(\mathbf{x})|\theta) = e^{-n\theta}\theta^{\sum x_i}$ (or $g(T(\mathbf{x})|\theta) = e^{-n\theta}\theta^{n\bar{x}}$); $h(\mathbf{x}) = \dfrac{1}{\prod_i x_i!}$.

Thus $T(\mathbf{x}) = \sum x_i$ and $T(\mathbf{x}) = \bar{x}$ are both sufficient for θ. (Any one-to-one transformation of $T(\mathbf{x}) = \sum x_i$ is sufficient).

Example 6.2.7

Let $X_1, \ldots, X_n$ be i.i.d. $\mathrm{N}\left(\mu, \sigma^2\right)$, where σ^2 is known.

$$f\left(\mathbf{x} \middle| \mu\right) = \prod \left(2\pi\sigma^2\right)^{-1/2} \exp\left[-\left(x_i - \mu\right)^2 / 2\sigma^2\right]$$

$$= \left(2\pi\sigma^2\right)^{-n/2} \exp\left[-\sum \left(x_i - \mu\right)^2 / 2\sigma^2\right]$$

$$= \left(2\pi\sigma^2\right)^{-n/2} \exp\left[-\sum \left(x_i^2 - 2x_i\mu + \mu^2\right) / 2\sigma^2\right]$$

$$= \left(2\pi\sigma^2\right)^{-n/2} \exp\left(-\frac{\sum x_i^2}{2\sigma^2}\right) \exp\left(\frac{\mu}{\sigma^2}\sum x_i - \frac{n\mu^2}{2\sigma^2}\right)$$

Set $g\left(T\left(\mathbf{x}\right) \middle| \mu\right) = \exp\left(\frac{\mu}{\sigma^2}\sum x_i - \frac{n\mu^2}{2\sigma^2}\right)$;

$h\left(\mathbf{x}\right) = \left(2\pi\sigma^2\right)^{-n/2} \exp\left(-\frac{\sum x_i^2}{2\sigma^2}\right)$. So again, $T\left(\mathbf{x}\right) = \sum x_i$ is sufficient for μ.

Example 6.2.9

Let $X_1, \ldots, X_n$ be i.i.d. N$\left(\mu, \sigma^2\right)$, where both $\theta = \left(\mu, \sigma^2\right)$ is unknown.

$$f\left(\mathbf{x}|\theta\right) = \left(2\pi\sigma^2\right)^{-n/2} \exp\left[-\sum\left(x_i - \mu\right)^2 / 2\sigma^2\right]$$

$$= \left(2\pi\sigma^2\right)^{-n/2} \exp\left[-\sum\left(x_i^2 - 2x_i\mu + \mu^2\right) / 2\sigma^2\right]$$

$$= \left(2\pi\sigma^2\right)^{-n/2} \exp\left(-\frac{\sum x_i^2}{2\sigma^2} + \frac{\mu}{\sigma^2}\sum x_i - \frac{n\mu^2}{2\sigma^2}\right).$$

Set

$$g\left(T\left(\mathbf{x}\right)|\mu\right) = \left(\sigma^2\right)^{-n/2} \exp\left(-\frac{\sum x_i^2}{2\sigma^2} + \frac{\mu}{\sigma^2}\sum x_i - \frac{n\mu^2}{2\sigma^2}\right)$$

and $h\left(\mathbf{x}\right) = \left(2\pi\right)^{-n/2}$. So $T\left(\mathbf{x}\right) = \left(\sum x_i, \sum x_i^2\right)$ is sufficient for $\theta = \left(\mu, \sigma^2\right)$.

Note that

$$s^2 = (n-1)^{-1} \sum (x_i - \overline{x})^2 = (n-1)^{-1} \left(\sum x_i^2 - n\overline{x}^2 \right)$$
$$= (n-1)^{-1} \left[\sum x_i^2 - \left(\sum x_i \right)^2 / n \right]$$

and thus the one-to-one transformation $T(\mathbf{X}) = \left(\bar{X}, S^2 \right)$ is also sufficient.

Examples

Let $X_1,\ldots,X_n$ be i.i.d. $Unif\left(\theta,\theta+1\right)$.

$$f\left(\mathbf{x}\middle|\theta\right)=I\left(\max_i x_i-1<\theta<\min_i x_i\right)=g\left(T\left(\mathbf{x}\right)\middle|\theta\right)$$

where $I\left(\bullet\right)$ is the indicator function.

$T\left(\mathbf{x}\right)=\left(\min_i x_i,\max_i x_i\right)$ is sufficient for θ.

Example 6.2.8 (Discrete uniform sufficient statistic)

All of the examples other than above two are special cases of what is called an exponential family.

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

$$f(x|\theta) = h(x)c(\theta)\exp\left[\sum_{j=1}^{k} w_j(\theta)t_j(x)\right].$$

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

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

is sufficient for θ.

Proof: Write

$$f(\mathbf{x}|\theta) = \left[\prod_i h(x_i)\right] \underbrace{c(\theta)\exp\left[\sum_{j=1}^{k} w_j(\theta)\sum_{i=1}^{n} t_j(x_i)\right]}_{g(T(\mathbf{X})|\theta)}$$

and use the Factorization Theorem.

Example

Let $X_1,\ldots,X_n$ be i.i.d. $\mathrm{N}(\theta, a\theta)$, where $a>0$ is known. (The variance is proportional to the mean.)

$$f\left(x_i \middle| \theta\right)=\left(2\pi a\theta\right)^{-1/2}\exp\left[-\left(x_i-\theta\right)^2/2a\theta\right]$$

$$=\left(2\pi a\theta\right)^{-1/2}\exp\left[-\left(x_i^2-2x_i\theta+\theta^2\right)/2a\theta\right]$$

$$=\left(2\pi a\theta\right)^{-1/2}\exp\left[-\frac{x_i^2}{2a\theta}+\frac{x_i}{a}-\frac{\theta}{2a}\right].$$

Set

$h\left(x_i\right)=\exp\left(x_i/a\right)$,

$c\left(\theta\right)=\left(2\pi a\theta\right)^{-1/2}\exp\left(-\theta/2a\right)$,

$w_1\left(\theta\right)=-1/\left(2a\theta\right)$ and $t_1\left(x_i\right)=x_i^2$. So $T\left(\mathbf{x}\right)=\sum x_i^2$ is sufficient for θ.

It is always true that the complete sample is a sufficient statistic (take $g\left(T(\mathbf{x})\middle|\theta\right)= f\left(\mathbf{x}\middle|\theta\right)$, $T(\mathbf{x})=\mathbf{x}$, and $h(\mathbf{x})=1$ in the Factorization Theorem).

However, this supplies no data reduction.

Also, it follows that any one-to-one function of a sufficient statistic is a sufficient statistic (e.g. $\sum x_i$ and $\overline{x}$).

Proof: Suppose $T(\mathbf{X})$ is a sufficient statistic and define $T^*(\mathbf{X}) = r(T(\mathbf{X}))$ for all $\mathbf{x}$, where r is a one-to-one function with inverse r^{-1}. Then by the Factorization Theorem there exist g and h such that

$$f(\mathbf{x}|\theta) = g(T(\mathbf{x})|\theta)h(\mathbf{x}) = g\left(r^{-1}\left[T^*(\mathbf{x})\right]|\theta\right)h(\mathbf{x})$$

Defining $g^*(t|\theta) = g\left(r^{-1}(t)|\theta\right)$, we see that

$f(\mathbf{x}|\theta) = g^*\left(T^*(\mathbf{x})|\theta\right)h(\mathbf{x})$, and by the Factorization Theorem, $T^*(\mathbf{X})$ is also sufficient for θ.

Being that there are multiple sufficient statistics for a parameter, we want to consider whether one is better than another.

By "better" we mean the statistic that achieves the most data reduction while retaining all the information about θ.

Minimal Sufficient Statistics

Definition: 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 $T'(\mathbf{X})$.

Example 6.2.12

Let $X_1,\ldots,X_n$ be i.i.d. $\mathrm{N}\left(\mu,\sigma^2\right)$, where σ^2 is known.

$$f\left(\mathbf{x}\middle|\theta\right)=\left(2\pi\sigma^2\right)^{-n/2}\exp\left(-\frac{\sum x_i^2}{2\sigma^2}\right)\exp\left(\frac{\mu}{\sigma^2}\sum x_i-\frac{n\mu^2}{2\sigma^2}\right)$$

$$=\left(2\pi\sigma^2\right)^{-n/2}\exp\left(-\frac{\sum x_i^2}{2\sigma^2}\right)\exp\left(\frac{\mu}{\sigma^2}\sum x_i-\frac{n\mu^2}{2\sigma^2}\right)$$

Using the Factorization Theorem, $T\left(\mathbf{X}\right)=\sum X_i$ is sufficient for μ, and so is $T'\left(\mathbf{X}\right)=\left(\sum X_i,\sum X_i^2\right)$. $T\left(\mathbf{X}\right)$ achieves greater data reduction, and $T\left(\mathbf{x}\right)$ is a function of $T'\left(\mathbf{x}\right)$.

But how do we show that $T\left(\mathbf{X}\right)$ is minimal?

Finding minimal sufficient statistics

<u>Lehmann-Scheffe Theorem</u> $T(\mathbf{X})$ is a minimal sufficient statistic for θ if 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 $T(\mathbf{x}) = T(\mathbf{y})$.

Example

Let $X_1,\ldots,X_n$ be i.i.d. $\mathrm{N}\left(\mu,\sigma^2\right)$, where σ^2 is known.

$$\frac{f\left(\mathbf{x}|\mu\right)}{f\left(\mathbf{y}|\mu\right)}=\frac{\left(2\pi\sigma^2\right)^{-n/2}\exp\left(-\frac{\sum x_i^2}{2\sigma^2}\right)\exp\left(\frac{\mu}{\sigma^2}\sum x_i-\frac{n\mu^2}{2\sigma^2}\right)}{\left(2\pi\sigma^2\right)^{-n/2}\exp\left(-\frac{\sum y_i^2}{2\sigma^2}\right)\exp\left(\frac{\mu}{\sigma^2}\sum y_i-\frac{n\mu^2}{2\sigma^2}\right)}$$

$$=\exp\left(\frac{\sum y_i^2-\sum x_i^2}{2\sigma^2}\right)\exp\left[\frac{\mu}{\sigma^2}\left(\sum x_i-\sum y_i\right)\right].$$

This ratio is constant as a function of μ if and only if $\sum x_i=\sum y_i$, and $T\left(\mathbf{X}\right)=\sum X_i$ is a minimal sufficient statistic. (So is $\bar{X}$ and any one-to one function of $T\left(\mathbf{X}\right)$.)

Example 6.2.14

Let $X_1,\ldots,X_n$ be i.i.d. $\mathrm{N}\left(\mu,\sigma^2\right)$, where both μ and σ^2 are unknown.

$$\frac{f\left(\mathbf{x}|\mu\right)}{f\left(\mathbf{y}|\mu\right)}=\frac{\left(2\pi\sigma^2\right)^{-n/2}\exp\left(-\frac{\sum x_i^2}{2\sigma^2}\right)\exp\left(\frac{\mu}{\sigma^2}\sum x_i-\frac{n\mu^2}{2\sigma^2}\right)}{\left(2\pi\sigma^2\right)^{-n/2}\exp\left(-\frac{\sum y_i^2}{2\sigma^2}\right)\exp\left(\frac{\mu}{\sigma^2}\sum y_i-\frac{n\mu^2}{2\sigma^2}\right)}$$

$$=\exp\left(\frac{\sum y_i^2-\sum x_i^2}{2\sigma^2}\right)\exp\left[\frac{\mu}{\sigma^2}\left(\sum x_i-\sum y_i\right)\right].$$

This ratio is constant as a function of $\theta=\left(\mu,\sigma^2\right)$ if and only if $\sum x_i=\sum y_i$ and $\sum x_i^2=\sum y_i^2$, and $T\left(\mathbf{X}\right)=\left(\sum X_i,\sum X_i^2\right)$ is a minimal sufficient statistic.

Ancillary statistics

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

An ancillary statistic by itself contains no information about θ.

However, at times, an ancillary statistic in conjunction with other statistics does contain information for inference about θ.

General location family

Let $X_1, \ldots, X_n$ be i.i.d. from a location family with cdf $F(x-\theta)$, $-\infty < x < \infty$. Also let $Z_1, \ldots, Z_n$ be i.i.d. with cdf $F(x)$, i.e. $X_i = Z_i + \theta$.

Then the cdf of the range $R = X_{(n)} - X_{(1)}$ is

$$F_R(r|\theta) = P_\theta(R \le r) = P_\theta\left(\max(X_i) - \min(X_i) \le r\right)$$
$$= P_\theta\left(\max(Z_i + \theta) - \min(Z_i + \theta) \le r\right)$$

$$= P_\theta\left(\max(Z_i) - \min(Z_i) \le r\right).$$

As this doesn't depend on θ, the range is ancillary in a location family.

Example 6.2.17

Let $X_1, \ldots, X_n$ be i.i.d. $Unif\left(\theta, \theta+1\right)$.

Now for a sample from the $Unif\left(\theta, \theta+1\right)$ distribution, $f\left(x|\theta\right)=1,\ \theta < x < \theta+1$ and

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

is a location family. The Range R is an ancillary statistic.

However, remember that $\left(X_{(1)}, X_{(n)}\right)$ is minimal sufficient, and so is the one-to-one transformation $\left(R, X_{(1)}\right)$.

So R by itself gives no information, but combined with $X_{(1)}$ is minimal sufficient.

Scale family

Let $X_1,\ldots,X_n$ be i.i.d. from a scale family with cdf $F(x/\sigma)$, $\sigma > 0$. Also let $Z_1,\ldots,Z_n$ be i.i.d. with cdf $F(x)$, i.e. $X_i = \sigma Z_i$.

A statistic that depends on ratios, such as $T(\mathbf{X}) = X_1 / X_n$, is ancillary.

$$P(T(\mathbf{X}) \le t) = P\left(X_1 / X_n \le t\right) = P\left(\sigma Z_1 / \sigma Z_n \le t\right)$$

$= P\left(Z_1 / Z_n \le t\right)$. This doesn't depend on σ and thus $T(\mathbf{X})$ is ancillary.

Example

Let $X_1,\ldots,X_n$ be i.i.d. $N\left(\mu,\sigma^2\right)$, where σ^2 is known. Recall that $S^2 = \dfrac{\sum\left(X_i-\bar{X}\right)^2}{n-1}$.

Also recall that $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$. So in this case S^2 is ancillary.

More generally, for any location invariant statistic such that $T(x_1+c,\ldots,x_n+c)=T(x_1,\ldots,x_n)$, or *scale family where*

$T(ax_1,\ldots,ax_n) = T(x_1,\ldots,x_n)$, *or*

location-scale family where $T(ax_1+c,\ldots,ax_n+c)=T(x_1,\ldots,x_n)$, T is ancillary.

Completeness

Completeness indicates when a minimal sufficient statistic is independent of any ancillary statistic.

The property of completeness will also be used to construct "good" estimators.

Definition: Let $f\left(\mathbf{t}\,\middle|\theta\right)$ be a family of distributions for a statistic $T\left(\mathbf{X}\right)$. The family is called *complete* if $E_\theta g\left(T\right)=0$ for all θ implies $P_\theta\left(g\left(T\right)=0\right)=1$ for all θ.

Completeness is a property of an entire family of probability distributions, indexed by θ.

For a complete family, no non-constant function of T can have an expectation that doesn't depend on θ.

This is true because if $E_\theta g(T) = c$, i.e. does not depend on θ, then $g^*(t) = g(t) - c$ has expectation zero, a contradiction.

Example 6.2.22

Let $X_1, \ldots, X_n$ be i.i.d. $Bernoulli(p)$. We know that $T(\mathbf{X}) = \sum X_i$ is sufficient for the parameter p $(0 < p < 1)$. We claim that it is a complete as well.

Now $T(\mathbf{X}) \sim Bin(n, p)$. We need to show that there does not exist a $g(T)$ such that $E_p(g(T)) = 0$ for all p except $g(T) = 0$ (the zero function).

Suppose that $E_p(g(T)) = 0$. Then

$$0 = E_p\left(g\left(T\right)\right) = \sum_{t=0}^{n} g\left(t\right)\binom{n}{t} p^t \left(1-p\right)^{n-t}$$

$$= \left(1-p\right)^n \sum_{t=0}^{n} g\left(t\right)\binom{n}{t}\left(\frac{p}{1-p}\right)^t$$

$$= \left(1-p\right)^n \sum_{t=0}^{n} g\left(t\right)\binom{n}{t} r^t$$

for all r, where $r = \dfrac{p}{1-p}$ satisfies $0 < r < \infty$. This implies that the nth degree polynomial in r above is zero for all r, and thus all of its coefficients are zero. Since $\binom{n}{t} \neq 0$, $g\left(t\right) = 0$, $t = 0, 1, \ldots, n$.

Thus T is complete.

Theorem: Let $X_1, \ldots, X_n$ be i.i.d. observations from an exponential family given by

$$f(x|\theta) = h(x)c(\theta)\exp\left[\sum_{j=1}^{k} w_j(\theta)t_j(x)\right].$$

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

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

is complete if $\{w_1(\theta), \ldots, w_k(\theta) : \theta \in \Theta\}$ contains an open set in $\mathbb{R}^k$.

Example

Let $X_1, \dots, X_n$ be i.i.d. $N\left(\mu, \sigma^2\right)$.

$$f\left(x \middle| \theta\right) = \left(2\pi\sigma^2\right)^{-1/2} \exp\left(-\frac{x^2}{2\sigma^2} + \frac{\mu}{\sigma^2}x - \frac{\mu^2}{2\sigma^2}\right).$$

$$w_1\left(\mu, \sigma^2\right) = \frac{\mu}{\sigma^2},\ t_1\left(x\right) = x$$

$$w_2\left(\mu, \sigma^2\right) = -\frac{1}{2\sigma^2},\ t_2\left(x\right) = x^2$$

Since $-\infty < \mu < \infty$, $0 < \sigma^2 < \infty$ we have $-\infty < w_1 < \infty$ and $-\infty < w_2 < 0$. As $\left(w_1\left(\mu, \sigma^2\right), w_2\left(\mu, \sigma^2\right)\right)$ contains an open set in $\mathbb{R}^2$, $T\left(\mathbf{X}\right) = \left(\sum x_i, \sum x_i^2\right)$ is a complete sufficient statistic.

Example

Let $X_1, \ldots, X_n$ be i.i.d. $N\left(\theta, \theta^2\right)$.

$$f\left(x\middle|\theta\right) = \left(2\pi\theta^2\right)^{-1/2} \exp\left(-\frac{x^2}{2\theta^2} + \frac{x}{\theta} - \frac{1}{2}\right).$$

$$w_1\left(\theta\right) = \frac{1}{\theta},\ t_1\left(x\right) = x$$

$$w_2\left(\theta\right) = -\frac{1}{2\theta^2},\ t_2\left(x\right) = x^2$$

Since $w_2\left(\theta\right) = -\frac{1}{2}\left[w_1\left(\theta\right)^2\right]$, the parameter space is on a parabola, which is not an open set on $\mathbb{R}^2$. However $T\left(\mathbf{X}\right) = \left(\sum x_i, \sum x_i^2\right)$ is not complete in this case. (p301: Exercise 6.15)

Note that $E\left(\bar{X}\right)=\theta$, and by Cochran's theorem, the square of $\sqrt{n-1}\, s/\sigma$ has a chi-square distribution with $n-1$ degrees of freedom.

It can be shown that $E\left(cS\right)=\theta$, where

$c=\dfrac{\sqrt{n-1}\,\Gamma\left(\dfrac{n-1}{2}\right)}{\sqrt{2}\,\Gamma\left(\dfrac{n}{2}\right)}$, and thus

$E\left(\bar{X}-cS\right)=0$.

Hence $T(\mathbf{X})=\left(\sum x_i, \sum x_i^2\right)$ is not complete.

$$E(\bar{X}^2)=\mu^2+\frac{\sigma^2}{n}=\frac{n+1}{n}\theta^2;\quad E(S^2)=\theta^2$$

$$\Rightarrow\quad E\left(n\bar{X}^2-(n+1)S^2\right)=0$$

Example

Let $X_1, \ldots, X_n$ be i.i.d. $N(\theta, \theta)$, $\theta > 0$. $f(x|\theta) = (2\pi\theta^2)^{-1/2} \exp\left(-\frac{x^2}{2\theta} + x - \frac{\theta}{2}\right)$. Here

$w_1(\theta) = -\frac{1}{2\theta}$, $t_1(x) = x^2$, and $-\infty < w_1(\theta) < 0$, and $\sum x_i^2$ is a complete sufficient statistic.

Example 6.2.23

Let $X_1,\ldots,X_n$ be i.i.d. $U(0,\theta)$. Then $f(\mathbf{x}|\theta)=\frac{1}{\theta^n}I\left(X_{(n)}<\theta\right)$ and $T(\mathbf{X})=X_{(n)}$ is sufficient.

Let $g(t)$ be a function such that $E\left(g(T)\right)=0$ for all θ. To find $E_\theta\left(g(T)\right)$ we compute $\int_0^\theta g(t)f_{X_{(n)}}(t)dt$.

Now, $F_{X_{(n)}}(t)=P\left(X_{(n)}\le t\right)=\left[P(X_1\le t)\right]^n=\left(\frac{t}{\theta}\right)^n$, $0<t<\theta$, so that $f_{X_{(n)}}(t)=nt^{n-1}/\theta^n$, $0<t<\theta$.

Then

$$0 = E_\theta\left(g\left(T\right)\right) = \int_0^\theta g\left(t\right) n t^{n-1} \theta^{-n} dt = n\theta^{-n} \int_0^\theta g\left(t\right)\, t^{n-1} dt \Rightarrow$$

$\int_0^\theta g\left(t\right)\, t^{n-1} dt = 0$ for all θ.

Differentiating with respect to θ on both sides of the last equation, $g\left(\theta\right)\theta^{n-1} = 0$ for all $\theta > 0$ (Fundamental Theorem of Calculus), which implies that $g\left(\theta\right) = 0$ for all θ.

Hence $T\left(\mathbf{X}\right) = X_{(n)}$ is a complete sufficient statistic.

Some facts:

1. If T is complete and $S=\Psi(T)$ is a function of T, then S is also complete.

Proof: If not, then $E_\theta\left[g(S)\right]=0$ for all θ. But then $E_\theta\left[g\left(\Psi(T)\right)\right]=0$ for all θ, and T cannot be complete, a contradiction. (Consider the composite function $g^*=g(\Psi)$).

2. Complete sufficient statistic is also minimal sufficient statistic

3. An ancillary statistic cannot be complete.

Proof: If S is ancillary then its distribution does not depend on θ.

So $E_\theta\left[g(S)\right]=E\left[g(S)\right]=C$ is constant in θ. So $g^*(S)=g(S)-C$ has $E\left[g^*(S)\right]=0$ for all θ. Thus S is not complete.

4. If a function of T is ancillary, then T cannot be complete.

Proof: Suppose that $\Psi(T)$ is ancillary, whereas T is complete. Completeness of T implies that $\Psi(T)$ must be complete as well. Since an ancillary statistic cannot be complete, we have a contradiction.

Note: This implies that no complete sufficient statistic exists for the $U(\theta, \theta+1)$ family. This follows since $T(\mathbf{X}) = \left(X_{(1)}, X_{(n)} - X_{(1)}\right)$ is minimal sufficient, and thus a function of every sufficient statistic, while the function of T, $X_{(n)} - X_{(1)}$, is ancillary.

5. An estimator $g(T)$ for a parameter $\Psi(\theta)$ (based on T only) is called *unbiased* if $E_\theta\left(g(T)\right)=\Psi(\theta)$ for all θ.

If T is complete then only one unbiased estimators based on T is possible.

Proof: If there were two unbiased estimator based on T, then $E_\theta\left(g(T)\right)=E_\theta\left(h(T)\right)=\Psi(\theta)$, but then $E_\theta\left(g^*(T)\right)=0$ for $g^*(t)=g(t)-h(t)$.

In Chapter 7, we will show that this estimator based on a complete sufficient statistic is the best possible unbiased estimator in a certain sense.

Basu's Theorem

If $T(\mathbf{X})$ is a complete sufficient statistic, then $T(\mathbf{X})$ is independent of every ancillary statistic.

(Note: The book says that $T(\mathbf{X})$ is minimal sufficient, but the minimal part is not a necessary assumption.)

Proof (discrete case)

Let S be ancillary and T a complete sufficient statistic. To show independence, we will show that

$P\left(S(\mathbf{X})=s \middle| T(\mathbf{X})=t\right)=P\left(S(\mathbf{X})=s\right)$ for all t.

First note that $P\left(S(\mathbf{X})=s \middle| T(\mathbf{X})=t\right)$ does not depend on θ because $T(\mathbf{X})$ is sufficient. Also, $P\left(S(\mathbf{X})=s\right)$ does not depend on θ because S is ancillary.

Now,

$$P\left(S(\mathbf{X})=s\right)=\sum_t P\left(S(\mathbf{X})=s \middle| T(\mathbf{X})=t\right)P_\theta\left(T(\mathbf{X})=t\right)$$

Also, $P\left(S(\mathbf{X})=s\right)=P\left(S(\mathbf{X})=s\right)\sum_t P_\theta\left(T(\mathbf{X})=t\right)$

$=\sum_t P\left(S(\mathbf{X})=s\right)P_\theta\left(T(\mathbf{X})=t\right)$.

Thus,

$$\sum_t \left[P\left(S(\mathbf{X}) = s\right) - P\left(S(\mathbf{X}) = s \middle| T(\mathbf{X}) = t\right) \right] P_\theta\left(T(\mathbf{X}) = t\right) = 0.$$

Now $g(t) = P\left(S(\mathbf{X}) = s\right) - P\left(S(\mathbf{X}) = s \middle| T(\mathbf{X}) = t\right)$ is not a function of θ, and $\sum_t g(t) P_\theta\left(T(\mathbf{X}) = t\right) = E_\theta g(T) = 0$. But since T is complete, $g(T) = 0$ for all t.

Note: Basu's Theorem depends heavily on completeness. For example, for the $U(\theta, \theta+1)$ distribution, minimal sufficient $T = \left(X_{(1)}, X_{(n)} - X_{(1)}\right)$ is not independent of $X_{(n)} - X_{(1)}$, which is ancillary.

Using Basu's theorem: Example

Let $X_1,\ldots,X_n$ be i.i.d. $\exp(\theta)$. We compute $E\left[X_{(1)}/\sum x_i\right]$, i.e. the expected proportion of the total that is accounted for by the smallest observation.

$$f\left(x_i|\theta\right)=\frac{1}{\theta}\exp\left[-x_i/\theta\right],$$

which is an exponential family with $t\left(x_i\right)=x_i$ (and also a scale family). Now $T\left(\mathbf{X}\right)=\sum X_i$ is a complete sufficient statistic whereas $g\left(\mathbf{X}\right)=X_{(1)}/\sum x_i$ is scale invariant (ancillary).

So $g(\mathbf{X})$ is independent of $T(\mathbf{X})$.

$$E\left[g(\mathbf{X})\right]=\frac{E\left[g(\mathbf{X})\right]E\left[T(\mathbf{X})\right]}{E\left[T(\mathbf{X})\right]}$$

$$=\frac{E\left[g(\mathbf{X})T(\mathbf{X})\right]}{E\left[T(\mathbf{X})\right]}=\frac{E\left(X_{(1)}\right)}{E\left(\sum X_i\right)}=\frac{?}{n\theta}.$$

$$F_{X_{(1)}}(x) = P\left[X_1 \le x\right] = 1 - P\left(\text{all } X_i > x\right)$$

$$= 1 - \left[1 - F_X(x)\right]^n = 1 - \left[\exp\left(x/\theta\right)\right]^n$$

$$= 1 - \exp\left(\frac{x}{\theta/n}\right).$$ Thus $X_{(1)} \sim \exp\left(\theta/n\right)$,

$E\left(X_{(1)}\right) = \theta/n$, and $E\left[g\left(\mathbf{X}\right)\right] = 1/n^2$.

(For $X \sim \exp\left(\theta/n\right)$,

$F_X(x) = \int_0^x \left(1/\theta\right) e^{-t/\theta} dt = 1 - e^{-x/\theta},\ \ 0 < x < \infty$.)

Example 6.2.26

6.3 The likelihood principle

We typically think of joint densities or probability mass functions $f(\mathbf{x}|\theta)$ as allowing us to quantify probabilities relative to x for a given θ.

Instead, we can think of the likelihood $L(\theta|\mathbf{x}) = f(\mathbf{x}|\theta)$ as a function of θ. It's not that θ is random, but that different values of θ give different probabilities for $\mathbf{x}$.

Idea: Consider $L(\theta_1|\mathbf{x})$ and $L(\theta_2|\mathbf{x})$ for two possible values of θ. If $L(\theta_1|\mathbf{x}) > L(\theta_2|\mathbf{x})$, then the sample that we observed is more likely to have occurred if θ_1 was the parameter than if θ_2 was the parameter.

So we say that θ_1 is a more plausible value than θ_2.

6.4 The equivariance principle

The equivariance principle is a data reduction technique that restricts inference by prescribing what other inferences must be made at related sample points.

This principle suggests certain inference procedures over others.

For a location family: $X_1,\ldots,X_n$ i.i.d. $F(\mathbf{x}-\theta)$. Suppose I transform the data so that $(X_1,\ldots,X_n)\to(X_1+b,\ldots,X_n+b)$.

Location invariant statistics: $T(x_1+b,\ldots,x_n+b)=T(x_1,\ldots,x_n)$, so the statistic has no information about θ (e.g. when taking differences).

Location equivariant statistic: $T(x_1+b,\ldots,x_n+b)=T(x_1,\ldots,x_n)+b$ (e.g. $\bar{X}$, $X_{(n)}$). The idea is that if θ shifts by b, my statistic should do the same if it is a good estimator of θ.

For a scale family: $X_1, \ldots, X_n$ i.i.d. $F(\mathbf{x}/\theta)$. Suppose I make the transformation $(X_1, \ldots, X_n) \to (aX_1, \ldots, aX_n)$.

Scale invariant statistics: $T(ax_1, \ldots, ax_n) = T(x_1, \ldots, x_n)$, so the statistic has no information about scale (e.g. anything based on ratios).

Scale equivariant statistic: $T(ax_1, \ldots, ax_n) = aT(x_1, \ldots, x_n)$ (e.g. S, $X_{(n)}$).

Homework: p300-309

6.3, 6.12, 6.15, 6.19, 6.21, 6.40