Biostat 602 Winter 2017

Lecture Set 4

Principles of Data Reduction

Ancillary Statistics, Completeness

Ancillary Statistic

Reading: CB 6.2

- Sufficient statistics contain all information about θ.
- At the other extreme is a statistic which does not contain any information on θ.

Definition 6.2.11

A statistic $S(\mathbf{X})$ is an *ancillary statistic* if its distribution does not depend on θ.

Question: Why then bother about an ancillary statistic when making an inference on θ?

Examples

1. $X_1, \cdots, X_n$ iid $\mathcal{N}(\mu, \sigma^2)$ where σ^2 is known.

   - $X_1 - X_2 \sim \mathcal{N}(0, 2\sigma^2)$ is ancillary.

   - $(X_1 + X_2)/2 - X_3 \sim \mathcal{N}(0, 1.5\sigma^2)$ is ancillary.

   - $s^2_{\mathbf{X}} = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2$ is ancillary.

   - $\frac{(n-1)s^2_{\mathbf{X}}}{\sigma^2} \sim \chi^2_{n-1}$ is ancillary.

2. $X_1, \cdots, X_n$ iid $\mathcal{N}(0, \sigma^2)$ where σ^2 is unknown.

   - X_1/X_2 is ancillary.

   - $\overline{X}/S_{\mathbf{X}}$ is ancillary.

   - Is $\overline{X}/\sigma$ ancillary?

3. Let $X_1, \cdots, X_n$ iid $Uniform(\theta,\ \theta+1)$. Show that the range statistic

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

   is ancillary. What is its distribution?

Location-Scale Family of Distributions

Let $f(x)$ be any pdf *free of any parameter* and let $-\infty < \mu < \infty$ and $\sigma > 0$ be unknown constants. Then

$$g(x|\mu, \sigma) = \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$$

is a pdf.

Proof: Because $f(x)$ is a pdf, then $f(x) \geq 0$, and $g(x|\mu, \sigma) \geq 0$ for all x. Let $y = (x-\mu)/\sigma$, then $x = \sigma y + \mu$, and $dx/dy = \sigma$.

$$\int_{-\infty}^{\infty} \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right) dx \;=\; \int_{-\infty}^{\infty} \frac{1}{\sigma} f(y) \sigma dy = \int_{-\infty}^{\infty} f(y) dy = 1$$

Therefore, $g(x|\mu, \sigma)$ is also a pdf.

- The pdf g corresponds to a **location-scale** family of distribution with location $= \mu$ and scale $= \sigma$.

- When $\mu = 0$, g is the pdf of a scale family with **scale** parameter σ.

- When $\sigma = 1$, g is the pdf of a **location** family with location parameter μ.

How do you show a pdf belongs to a location-scale family?

Use the transformation $Y = (X-\mu)/\sigma$. If Y has a **parameter-free** pdf, then the original pdf belongs to a location-scale family.

Examples

1. $X \sim N(\mu, \sigma^2)$

2. $X \sim Exp(\theta)$

3. $X \sim Cauchy(\theta, 1)$

4. $X \sim Uniform(0, \theta)$

5. $X \sim Uniform(\theta, 2\theta)$

Ancillary Statistic for Location Family

Let $X_1, \cdots, X_n$ be iid from a location family with pdf $f(x-\mu)$ where $-\infty < \mu < \infty$. Show that the range $R = X_{(n)} - X_{(1)}$ is an ancillary statistic.

Solution: Since the original population distribution belongs to a location family, $Z_1 = X_1 - \mu, \cdots, Z_n = X_n - \mu$ are iid observations from pdf $f(x)$ and cdf $F(x)$, which are free of the parameter μ. Then the cdf of the range statistic R becomes

$$
\begin{aligned}
F_R(r|\mu) &= \Pr(R \le r|\mu) = \Pr(X_{(n)} - X_{(1)} \le r|\mu) \\
&= \Pr(Z_{(n)} + \mu - Z_{(1)} - \mu \le r|\mu) = \Pr(Z_{(n)} - Z_{(1)} \le r|\mu)
\end{aligned}
$$

which does not depend on μ because $Z_1, \cdots, Z_n$ does not depend on μ. Therefore, R is an ancillary statistic.

Ancillary Statistic for Scale Family

Let $X_1, \cdots, X_n$ be iid from a scale family with pdf $f(x/\sigma)/\sigma$ where $\sigma > 0$. Show that the statistic

$$
\mathbf{T}(\mathbf{X}) = (X_1/X_n, \cdots, X_{n-1}/X_n) \qquad \text{is ancillary.}
$$

Solution: Let $Z_1 = X_1/\sigma, \cdots, Z_n = X_n/\sigma$ be iid observations from pdf $f(x)$. Then the joint cdf of $\mathbf{T}(\mathbf{X})$ is

$$
\begin{aligned}
F_{\mathbf{T}}(t_1, \cdots, t_{n-1}|\sigma) &= \Pr(X_1/X_n \le t_1, \cdots, X_{n-1}/X_n \le t_{n-1}|\sigma) \\
&= \Pr(\sigma Z_1/\sigma Z_n \le t_1, \cdots, \sigma Z_{n-1}/\sigma Z_n \le t_{n-1}|\sigma) \\
&= \Pr(Z_1/Z_n \le t_1, \cdots, Z_{n-1}/Z_n \le t_{n-1}|\sigma)
\end{aligned}
$$

Because $Z_1, \cdots, Z_n$ does not depend on σ, $\mathbf{T}(\mathbf{X})$ is an ancillary statistic.

Ancillary vs Minimal Sufficient Statistic

- Ancillary statistic is free of θ.
- Minimal sufficient statistic contains minimal information related to θ.
- Are ancillary statistics independent of minimal sufficient statistics?

Example: For $X_1, \cdots, X_n \sim \text{Uniform}(\theta, \theta+1)$, $R = X_{(n)} - X_{(1)}$ and $M = (X_{(n)} + X_{(1)})/2$ are jointly minimal sufficient statistic (why?)

But R is ancillary statistic, so ancillary statistics are not always independent of minimal sufficient statistic.

However, how does R give any information about θ?

- If $M = 1$, then $0 < \theta < 1$ (why?).

- Suppose now $R = 0.8$. By itself, it does not provide any information about θ.

- In combination with the fact that $M = 1$, it yields that $X_{(1)} = 0.6$ and $X_{(n)} = 1.4$, and so the possible range of θ is narrowed down to $0.4 < \theta < 0.6$.

- Combination of ancillary statistic and another statistic can be more informative jointly than the other statistic alone.

- Thus, an ancillary statistic can provide additional precision about the parameter when combined with another statistic.

Completeness

In statistical inference, the ulterior objective is to identify a statistic that is a *good* estimator for the parameter. **Sufficiency** helps us identify statistics that contain information on the parameter. While somewhat counter-intuitive, **Ancillary** statistics enhance that information, while working in conjunction with a sufficient statistics. The final piece of the puzzle is the concept of **completeness**. Together, these three principles provide enough structure for us to pursue our quest for an efficient estimator in a systematic way.

Definition: Let $\{f_T(t|\theta), \theta \in \Omega\}$ be a family of pdfs or pmfs for a statistic $T(\mathbf{X})$. This family of probability distributions is called *complete* if

$$E[g(T)|\theta] = 0 \textit{ for all } \theta \;\text{ implies }\; \Pr[g(T) = 0|\theta] = 1 \text{ for all } \theta.$$

Remarks

- In other words, $g(T) = 0$ almost surely, i.e. only the zero function of T can have a mean of zero for all parameter values.
- Loosely $T(\mathbf{X})$ is called a *complete statistic*. However, as we shall see soon, completeness is the property of the family of distributions induced by T, and not that of T itself.
- Completeness implies 'no unnecessary part' conceptually. There is no non-trivial $g(T)$ whose expectation (or distribution) does not depend on θ.
- If an ancillary statistic could be made out of $T(\mathbf{X})$, it is NOT complete.
- This is a more stringent requirement than that is needed for minimal sufficient statistics.

Example 1: Let $X_1, \cdots, X_n$ be a random sample from a $Bern(p)$ population. Show that $T = \sum_{i=1}^n X_i$ is complete.

Example 2: Let $X_1, \cdots, X_n$ be a random sample from a $Uniform(0, \theta)$ population. Show that $T = \max_i X_i$ is complete.

Example 3: Let $X_1, \cdots, X_n$ be a random sample from a $Uniform(\theta, \theta+1)$ population. We know $\mathbf{T} = (X_{(1)}, X_{(n)})$ is minimal sufficient. Is $\mathbf{T}$ complete?

Example 4: Let $X_1, \cdots, X_n$ be a random sample from a $Pois(\lambda)$. Show that $T = \sum_{i=1}^{n} X_i$ is complete.

Example 5: Let $T \sim Pois(\lambda)$, where the parameter space of λ is given by

$$\Omega = \{\lambda : \lambda = \{1, 2\}\}.$$

Show that the family of distributions induced by T is NOT complete.

Proof: We need to find a counter example which is a function g such that $E[g(T)|\lambda] = 0$ for $\lambda = 1, 2$ but $g(T) \neq 0$. The function g must satisfy

$$E[g(T)|\lambda] = \sum_{t=0}^{\infty} g(t) \frac{\lambda^t e^{-\lambda}}{t!} = 0$$

for $\lambda \in \{1, 2\}$. Thus,

$$\begin{cases} E[g(T)|\lambda = 1] &= \sum_{t=0}^{\infty} g(t) \frac{1^t e^{-1}}{t!} = 0 \\ \\ E[g(T)|\lambda = 2] &= \sum_{t=0}^{\infty} g(t) \frac{2^t e^{-2}}{t!} = 0 \end{cases}$$

The above equation can be rewritten as

$$\begin{cases} \sum_{t=0}^{\infty} g(t)/t! &= 0 \\ \\ \sum_{t=0}^{\infty} 2^t g(t)/t! &= 0 \end{cases}$$

Define $g(t)$ as

$$g(t) = \begin{cases} 2 & t = 0 \text{ and } t = 2 \\ -3 & t = 1 \\ 0 & \text{otherwise} \end{cases}$$

Then

$$\begin{aligned} \sum_{t=0}^{\infty} g(t)/t! &= g(0)/0! + g(1)/1! + g(2)/2! = 2 - 3 + 2/2 = 0 \\ \sum_{t=0}^{\infty} 2^t g(t)/t! &= g(0)/0! + 2g(1)/1! + 2^2 g(2)/2! = 2 - 6 + 8/2 = 0 \end{aligned}$$

There exists a non-zero function g that satisfies $E[g(T)|\lambda] = 0$ for all $\lambda \in \Omega$. Therefore this family is NOT complete.

Question: Why is a complete statistic called 'complete'?

Note that requiring $g(T)$ to satisfy the definition of completeness puts a restriction on g. The larger the family of pdfs/pmfs, the greater is the restriction on g. When the family of pdfs/pmfs is augmented to the point that $E[g(T)] = 0$ for all θ rules out all g except for the trivial $g(T) = 0$, then the family is said to be complete. A common verbalization of this definition is that the family of distributions is complete if there is no *unbiased estimator* of zero except for the trivial estimator $g \equiv 0$.

As the Poisson example shows, 'completeness' is a property of the family of distributions rather than the random variable or its parametric form.

Ancillary and Complete Statistics

Fact 1: For a statistic $T(\mathbf{X})$, if a non-constant function of T, say $r(T)$ is ancillary, then $T(\mathbf{X})$ cannot be complete.

Proof: Define $g(T) = r(T) - E[r(T)]$, which does not depend on the parameter θ because $r(T)$ is ancillary. Then $E[g(T)|\theta] = 0$ for a non-zero function $g(T)$, and $T(\mathbf{X})$ is not a complete statistic.

Arbitrary Functions of Complete Statistics

Fact 2: If $T(\mathbf{X})$ is a complete statistic, then a non-constant function of T, say $T^* = r(T)$ is also complete.

Proof: We can write

$$E[g(T^*)|\theta] \;=\; E[g \circ r(T)|\theta]$$

Now assume that $E[g(T^*)|\theta] = 0$ for all θ. Then

$$E[g \circ r(T)|\theta] = 0$$

holds for all θ too. Since $T(\mathbf{X})$ is a complete statistic,

$$\Pr[g \circ r(T) = 0] = 1, \;\forall \theta \in \Omega.$$

Therefore $\Pr[g(T^*) = 0] = 1$, and T^* is a complete statistic.

Completeness and sufficiency

Theorem 6.2.28: If a minimal sufficient statistic exists, then any complete sufficient statistic is also a minimal sufficient statistic.

Proof: Known as *Bahadur's Theorem*, beyond the scope of the course. Book statement is inaccurate.

Remarks:

- With the exception of very unusual cases, under a mild assumption, minimal sufficient statistics always exist.
- The converse is NOT true. A minimal sufficient statistic is not necessarily complete. Recall the example of Uniform$(\theta, \theta + 1)$.

Basu's Theorem

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

Proof – for discrete case

Suppose that $S(\mathbf{X})$ is an ancillary statistic. We want to show that

$$\Pr(S(\mathbf{X}) = s|T(\mathbf{X}) = t) = \Pr(S(\mathbf{X}) = s),\ \forall t \in \mathcal{T} \qquad (*)$$

Now we have, using *law of total probability*,

$$\Pr(S(\mathbf{X}) = s|\theta) = \sum_{t\in\mathcal{T}} \Pr(S(\mathbf{X}) = s|T(\mathbf{X}) = t)\Pr(T(\mathbf{X}) = t|\theta) \qquad (1)$$

Since, $\sum_{t\in\mathcal{T}} \Pr(T(\mathbf{X}) = t|\theta) = 1$, we can write

$$\begin{aligned}\Pr(S(\mathbf{X}) = s|\theta) &= \Pr(S(\mathbf{X}) = s)\sum_{t\in\mathcal{T}} \Pr(T(\mathbf{X}) = t|\theta) \\ &= \sum_{t\in\mathcal{T}} \Pr(S(\mathbf{X}) = s)\Pr(T(\mathbf{X}) = t|\theta) \qquad (2)\end{aligned}$$

Define $g(t) = \Pr(S(\mathbf{X}) = s|T(\mathbf{X}) = t) - \Pr(S(\mathbf{X}) = s)$. Using (1) and (2),

$$\sum_{t\in\mathcal{T}} \big[\Pr(S(\mathbf{X}) = s|T(\mathbf{X}) = t) - \Pr(S(\mathbf{X}) = s)\big]\Pr(T(\mathbf{X}) = t|\theta) = 0$$

This implies

$$\sum_{t\in\mathcal{T}} g(t)\Pr(T(\mathbf{X}) = t|\theta) = E[g(T(\mathbf{X}))|\theta] = 0$$

$T(\mathbf{X})$ is complete, so $g(t) = 0$ almost surely for all possible $t \in \mathcal{T}$.

Therefore, $(*)$ is established and $S(\mathbf{X})$ is independent of $T(\mathbf{X})$.

Example 6: Let $X_1, X_2, \ldots, X_n$ be a random sample from a $Uniform(0, \theta)$ distribution. Let $X_{(1)} < X_{(2)} < \ldots < X_{(n)}$ be the corresponding order statistics.

(a) Show that $X_{(n)}$ and $X_{(1)}/X_{(n)}$ are independent random variables.

(b) Establish that

$$E\left[\frac{X_{(1)}}{X_{(n)}}\right] = \frac{E(X_{(1)})}{E(X_{(n)})} = \frac{1}{n}.$$

Example 7: Let $X_1, X_2, \ldots, X_n$ be a random sample from a $\mathcal{N}(\mu, \sigma^2)$ distribution. Conclude that $\overline{X}$ and S^2 are independent.

Example 8: *Exercise 6.19 CB* The random variable X takes the values 0, 1, 2, according to one of the following distributions:

	$\Pr(X=0)$	$\Pr(X=1)$	$\Pr(X=2)$	
Distribution 1	p	$3p$	$1-4p$	$0<p<\frac{1}{4}$
Distribution 2	p	p^2	$1-p-p^2$	$0<p<\frac{1}{2}$

In each case, determine whether the family of distribution of X is complete.

Solution - Distribution 1

Suppose that there exist $g(\cdot)$ such that $E[g(X)|p]=0$ for all $0<p<\frac{1}{4}$.

$$
\begin{aligned}
f_X(x|p) &= p^{I(x=0)}(3p)^{I(x=1)}(1-4p)^{I(x=2)} \\
E[g(X)|p] &= \sum_{x\in\{0,1,2\}} g(x)f_X(x|p) \\
&= g(0)\cdot p + g(1)\cdot(3p) + g(2)\cdot(1-4p) \\
&= p[g(0)+3g(1)-4g(2)] + g(2) = 0
\end{aligned}
$$

Therefore, $g(2)=0$, $g(0)+3g(1)=0$ must hold, and it is possible that g is a nonzero function that makes $\Pr[g(X)=0]<1$. For example, $g(0)=3, g(1)=-1, g(2)=0$. Therefore the family of distributions of X is not complete.

Solution - Distribution 2

Suppose that there exist $g(\cdot)$ such that $E[g(X)|p] = 0$ for all $0 < p < \frac{1}{4}$.

$$
\begin{aligned}
f_X(x|p) &= p^{I(x=0)}(p^2)^{I(x=1)}(1-p-p^2)^{I(x=2)} \\
E[g(X)|p] &= \sum_{x\in\{0,1,2\}} g(x) f_X(x|p) \\
&= g(0)\cdot p + g(1)\cdot p^2 + g(2)\cdot(1-p-p^2) \\
&= p^2[g(1)-g(2)] + p[g(0)-g(2)] + g(2) = 0
\end{aligned}
$$

$g(0) = g(1) = g(2) = 0$ must hold in order to $E[g(X)|p] = 0$ for all p. Therefore the family of distributions of X is complete.

Example 9: Let $X_1, \cdots, X_n$ *i.i.d.* $Pois(\lambda)$, where $\lambda > 0$ is unknown. Let $\overline{X},\ S^2$ denote the sample mean and variance, respectively. Show that

$$E\left[S^2 \middle| \overline{X}\right] = \overline{X} \ \text{ almost surely.}$$