Statistical Inference

Lecture 04a

ANU - RSFAS

Last Updated: Mon Mar 12 18:15:14 2018

Minimal Sufficient

Definition 2.6: A sufficient statistic $T(\boldsymbol{X})$ is called a minimal sufficient statistic if, for any other sufficient statistic $T'(\boldsymbol{X})$, $T(\boldsymbol{X})$ is a function of $T'(\boldsymbol{X})$.

- Not easy to use the definition to find a minimal sufficient statistic!

Lemma 2.3: Let $f(\boldsymbol{x};\theta)$ be the pdf or pmf of a sample $\boldsymbol{X}$. Suppose there exists a function $T(\boldsymbol{x})$ such that, for every two sample points $\boldsymbol{x}$ and $\boldsymbol{y}$ the ratio

$$L(\theta;\boldsymbol{x})/L(\theta;\boldsymbol{y}) \sim 1$$

is constant as function of θ [note: this can be a vector] if and only if

$$T(\boldsymbol{x}) = T(\boldsymbol{y}).$$

Then $T(\boldsymbol{X})$ is a minimal sufficient statistic.

Minimal Sufficient

Proof: Consider:

$$g(\boldsymbol{x}|\boldsymbol{t}) = \frac{f(x_1, x_2, \dots, x_n)}{h(\boldsymbol{t})} = \frac{L(\boldsymbol{\theta}; \boldsymbol{x})}{\sum_{\boldsymbol{y} \in \tau} L(\boldsymbol{\theta}; \boldsymbol{y})}$$

- Where τ is the set of $\boldsymbol{y}$s such that $\boldsymbol{T} = \boldsymbol{t}$, that is such that:

$$\frac{L(\boldsymbol{\theta}; \boldsymbol{y})}{L(\boldsymbol{\theta}; \boldsymbol{x})} = m(\boldsymbol{x}, \boldsymbol{y})$$

$$g(\boldsymbol{x}|\boldsymbol{t}) = \frac{L(\boldsymbol{\theta}; \boldsymbol{x})}{\sum_{\boldsymbol{y} \in \tau} L(\boldsymbol{\theta}; \boldsymbol{x}) m(\boldsymbol{x}, \boldsymbol{y})} = \frac{L(\boldsymbol{\theta}; \boldsymbol{x})}{L(\boldsymbol{\theta}; \boldsymbol{x}) \sum_{\boldsymbol{y} \in \tau} m(\boldsymbol{x}, \boldsymbol{y})}$$

- Therefore $\boldsymbol{T}$ is a sufficient statistic.

Minimal Sufficient

- Now supose that $\boldsymbol{U(x)}$ is any other sufficient statistic and that $\boldsymbol{U(x)} = \boldsymbol{U(y)}$.

sufficient => factorize

$$\frac{L(\theta;\boldsymbol{x})}{L(\theta;\boldsymbol{y})} = \frac{K_1[\boldsymbol{u(x)};\theta]K_2[\boldsymbol{x}]}{K_1[\boldsymbol{u(y)};\theta]K_2[\boldsymbol{y}]} = \frac{K_2[\boldsymbol{x}]}{K_2[\boldsymbol{y}]}$$

constant for θ.

Minimal Sufficient

Example:

- Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} n(\mu, \sigma^2)$, with both μ, σ^2 unknown.
- Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be two sample points.
- Let $(\bar{x}, s_{\boldsymbol{x}}^2)$ and $(\bar{y}, s_{\boldsymbol{y}}^2)$ be the sample means and sample variances for the samples $\boldsymbol{x}$ and $\boldsymbol{y}$.

$$
\begin{aligned}
\frac{f(\boldsymbol{x}|\mu,\sigma^2)}{f(\boldsymbol{y}|\mu,\sigma^2)} &= \frac{(2\pi\sigma^2)^{-n/2} exp(-[n(\bar{x}-\mu)^2 - (n-1)s_{\boldsymbol{x}}^2]/(2\sigma^2))}{(2\pi\sigma^2)^{-n/2} exp(-[n(\bar{y}-\mu)^2 - (n-1)s_{\boldsymbol{y}}^2]/(2\sigma^2))} \\
&= exp([-n(\bar{x}^2 - \bar{y}^2) + 2n\mu(\bar{x}-\bar{y}) - (n-1)(s_{\boldsymbol{x}}^2 - s_{\boldsymbol{y}}^2)]/(2\sigma^2))
\end{aligned}
$$

- This ratio will not depend on μ and σ^2 if and only if $\bar{x} = \bar{y}$ and $s_{\boldsymbol{x}}^2 = s_{\boldsymbol{y}}^2$.
- $(\bar{X}, S^2)$ are minimally sufficient for μ, σ^2.

Minimal Sufficient

- Note: Minimal sufficient statistics are not unique. Any one-to-one function of a minimal sufficient statistic is also minimal sufficient.
- In the previous example $(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2)$ is also a set of minimal sufficient statistics for (μ, σ^2)

Can Sufficient Statistics Help Us with MVUEs?

Theorem 2.2 (Rao-Blackwell):

- Let W be any unbiased estimator of $\tau(\theta)$.
- Let T be a sufficient statistic for θ.

KEY: SUFFICIENCY

- Define $\phi(T) = E[W|T]$.
- Then

The new estimator is ① also unbiased ② may have smaller variance.

$$E[\phi(T)] = \tau(\theta)$$

$$V[\phi(T)] \leq V[W]$$

- So if we have unbiased estimator and condition it on a sufficient statistic, our new statistic $\phi(T)$ has the same or smaller variance!!

Proof: Recall, that if X and Y are any two random variables:

$$E[X] = E[E(X|Y)]$$

$$V[X] = V[E(X|Y)] + E[V(X|Y)]$$

- Show that $\phi(T)$ is unbiased for $\tau(\theta)$:

$$\begin{aligned} E[W] &= \tau(\theta) \\ E[W] &= E[\underbrace{E[W|T]}_{\phi(T)}] = E[\phi(T)] = \underline{\tau(\theta)} \end{aligned}$$

- Show that $V[\phi(T)] \leq V[W]$:

$$
\begin{aligned}
V[W] &= V[E(W|T)] + E[V(W|T)] \\
&= V[\phi(T)] + E[V(W|T)] \\
&\geq V[\phi(T)]
\end{aligned}
$$

- As $V(W|T) \geq 0$

- So the whole idea seems quite cool. We can potentially get better estimators. But the key seems to be that idea of sufficiency.
- What happens if we don't condition on a sufficient statistic?

Example: $X_1, X_2 \overset{\text{iid}}{\sim} n(\theta, 1)$. Consider the statistic $\bar{X}$:

$$E[\bar{X}] = \theta \qquad V(\bar{X}) = 1/2$$

- Now let's condition on X_1. This is not a sufficient statistic! Recall our new estimator is $\phi(X_1) = E[\bar{X}|X_1]$ (note the expectation):

$$= E\left[\frac{1}{2}(X_1 + X_2) \mid X_1\right]$$

$$\begin{aligned}
\phi(X_1) &= E[\bar{X}|X_1] \\
&= \frac{1}{2}E[X_1|X_1] + \frac{1}{2}E[X_2|X_1] \quad \Rightarrow P(X_2|X_1) = P(X_2) \\
&= \frac{1}{2}E[X_1|X_1] + \frac{1}{2}E[X_2] \quad \text{b/c independent} \\
&= \frac{1}{2}E[X_1|X_1] + \frac{1}{2}\theta
\end{aligned}$$

$\neq \theta$ $\therefore$ not unbiased.

some unknown parameter here.

we have a problem here.

- As $\phi(X_1)$ depends on an unknown parameter it is not even an estimator (statistic).
- Recall, conditioning on a sufficient statistic removes the parameter!

Complete Statistics

Definition 2.9: Let $f_T(t;\theta)$ be a family of of pdfs or pmfs for a statistic $T(\boldsymbol{x})$. The family of probability distributions is called complete if

$$E[h(T)] = \int h(t) f_T(t) dt = 0$$

for all θ implies that

$$P(g(T) = 0) = 1$$

for all θ.

Complete Statistic

Example:

- Suppose that T has a binomial (n, p) distribution, $0 < p < 1$.
- Let h be a function such that $E_\theta[h(T)] = 0$.

$$\begin{aligned} 0 = E[h(T)] &= \sum_{t=0}^{n} h(t) \binom{n}{t} p^t (1-p)^{n-t} \\ &= (1-p)^n \sum_{t=0}^{n} h(t) \binom{n}{t} \left(\frac{p}{(1-p)} \right)^t \\ &= \sum_{t=0}^{n} h(t) \binom{n}{t} \left(\frac{p}{(1-p)} \right)^t \\ &\Rightarrow \quad 0 = \sum_{t=0}^{n} h(t) \binom{n}{t} r^t \end{aligned}$$

we want

$r = \frac{p}{1-p}$

Complete Statistic

$$0 = \sum_{t=0}^{n} h(t) \binom{n}{t} r^t \quad \forall\ r$$

- The only way for this to occur is that $g(t) = 0 \ \ \forall\ t$.
- So we have:

$$P_p(h(T) = 0) = 1$$

- T is a complete statistic.

Lehman - Scheffe Theorem

Lemma 2.6: Let $X_1, \ldots, X_n$ be a random sample from a distribution with density function $f(x;\theta)$. If $T = T(\boldsymbol{X})$ is a complete and sufficient statistic, and $\phi(T)$ is an unbiased estimator of $\tau(\theta)$, then $\phi(T)$ is the unique MVUE of $\tau(\theta)$.

$\phi(T) = E[W|T]$

Proof:

- Let U be any other unbiased estimator of $\tau(\theta)$.
- Let $U^* = E[U|T]$.
- Consider $h(T) = U^* - \phi(T)$. Recall: $\phi(T) = E[W|T]$. This means:

both are $\tau(\theta)$

$$E[h(T)] = \overset{=\tau(\theta)}{E[U^*]} - \overset{=\tau(\theta)}{E[\phi(T)]} = 0, \quad \forall \quad \theta$$

$\Downarrow$ completeness

- We know that T is complete. So:

so what's inside is also zero

$$h(T) = U^* - \phi(T) = 0 \quad \Rightarrow \quad U^* = \phi(T)$$

There is only one unbiased estimator of $\tau(\theta)$ that is a function of T!

See also Lemma 2.7.

- How to find MVUEs? It seems we have an approach:
    1. Find or construct a sufficient and complete statistic T.
    2. Find an unbiased estimator W for $\tau(\theta)$.
    3. Compute $\phi(T) = E[W|T]$, then $\phi(T)$ is the ~~U~~MVUE.
- Or:
    1. Find or construct a sufficient and complete statistic T.
    2. Find a function $h(T)$, where $E[h(T)] = \tau(\theta)$ (i.e. it is unbiased).
    3. Then $h(T)$ is the MVUE.

Method 1

Example: Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Bernoulli}(\theta)$.

- $T = \sum_{i=1}^{n} X_i$ is a sufficient and complete statistic for θ.
- Let's consider $W = X_1$. $E[W] = \theta$. So W is unbiased.
- Compute $\phi(T) = E[W|T]$.

a really bad estimator (not really need to be a good one)

Find the expectation of a conditional.

Note: W is 0 or 1. $E[W] = 1P(X_1 = 1) + 0P(X_1 = 0)$.

$$
\begin{aligned}
E[W|T] &= P\left(X_1 = 1 \middle| T = t\right) \\
&= \frac{P\left(X_1 = 1,\ T = t\right)}{P\left(T = t\right)} \\
&= \frac{P\left(X_1 = 1,\ \sum_{i=1}^{n} X_i = t\right)}{P\left(\sum_{i=1}^{n} X_i = t\right)} \\
&= \frac{P\left(X_1 = 1,\ \sum_{i=2}^{n} X_i = (t-1)\right)}{P\left(\sum_{i=1}^{n} X_i = t\right)} \\
&= \frac{P\left(X_1 = 1\right) \times P\left(\sum_{i=2}^{n} X_i = (t-1)\right)}{P\left(\sum_{i=1}^{n} X_i = t\right)} \\
&= \frac{[\theta] \times [\binom{n-1}{t-1}\theta^{t-1}(1-\theta)^{(n-1)-(t-1)}]}{\binom{n}{t}\theta^{t}(1-\theta)^{n-t}} \\
&= \frac{t}{n} \Rightarrow \frac{T}{n} = \bar{X}
\end{aligned}
$$

$\bar{X}$ is the ~~U~~MVUE of θ.

Method 2

- It would be great if we could automatically pick out sufficient and complete statistics . . .

Exponential Families