6.434/16.391 Statistics for Engineers and Scientists

Lecture 8 09/30/2013

Laboratory for Information and Decision Systems
Massachusetts Institute of Technology

Lecture 8 09/30/2013

EQUIVALENCE BETWEEN SAMPLE VALUES

The set $\mathcal{D}_\text{o}$

- Let $\{P_\theta : \theta \in \Theta\}$ be a family of distributions. Let

$$\mathcal{D}_\text{o} = \{\mathbf{x} : f(\mathbf{x}|\theta) = 0,\ \forall\theta\}$$

  - Note that $\mathcal{D}_\text{o}$ does not contain points that

    $f(\mathbf{x}|\theta) = 0$ for some θ

    but $f(\mathbf{x}|\theta) \neq 0$ for other θ.

    It only contains points that $f(\mathbf{x}|\theta) = 0$ for <u>all</u> θ

Equivalence between sample values

- Define: For $\mathbf{x},\, \mathbf{y} \notin \mathcal{D}_{\rm o}$, $\mathbf{x} \sim \mathbf{y}$ if and only if there exists a function $k(\mathbf{x},\, \mathbf{y}) \in (0,\, \infty)$ such that

$$f(\mathbf{x}|\theta) = k(\mathbf{x},\, \mathbf{y}) f(\mathbf{y}|\theta)$$

  Note the "likelihood ratio"

$$\frac{f(\mathbf{x}|\theta)}{f(\mathbf{y}|\theta)} = k(\mathbf{x},\, \mathbf{y})$$

  is independent of θ

- We will call this $\sim$ relationship as " $\mathbf{x}$ is equivalent to $\mathbf{y}$ "
- This likelihood equation would not make a good definition since $f(\mathbf{y}|\theta)$ may be equal to zero for some θ even though $f(\mathbf{y}|\theta) \neq 0$, $\forall\,\theta$, so $\mathbf{y} \notin \mathcal{D}_{\rm o}$, but have problem with dividing by zero

The equivalent notation (1 of 2)

- Lemma: $\sim$ is equivalent notation
- Proof: we need to show
  - a) $\mathbf{x} \sim \mathbf{x}$
  - b) $\mathbf{x} \sim \mathbf{y} \Rightarrow \mathbf{y} \sim \mathbf{x}$
  - c) $\mathbf{x} \sim \mathbf{y},\ \mathbf{y} \sim \mathbf{z} \Rightarrow \mathbf{x} \sim \mathbf{z}$
- a) Trivial: $k(\mathbf{x},\, \mathbf{x}) = 1$
- b) If $\mathbf{x} \sim \mathbf{y}$, then there exists $k(\mathbf{x},\, \mathbf{y}) \in (0,\, \infty)$, such that

$$f(\mathbf{x}|\theta) = k(\mathbf{x},\, \mathbf{y}) f(\mathbf{y}|\theta)$$

This implies that

$$f(\mathbf{y}|\theta) = k(\mathbf{y},\, \mathbf{x}) f(\mathbf{x}|\theta)$$

where

$$k(\mathbf{y},\, \mathbf{x}) = \frac{1}{k(\mathbf{x},\, \mathbf{y})} \in (0,\, \infty)$$

Therefore, $\mathbf{y} \sim \mathbf{x}$

The equivalent notation (2 of 2)

- c) If $\mathbf{x} \sim \mathbf{y}$, then there exists $k(\mathbf{x},\, \mathbf{y}) \in (0,\, \infty)$, such that

$$f(\mathbf{x}|\theta) = k(\mathbf{x},\, \mathbf{y}) f(\mathbf{y}|\theta)$$

  If $\mathbf{y} \sim \mathbf{z}$, then there exists $k(\mathbf{y},\, \mathbf{z}) \in (0,\, \infty)$, such that

$$f(\mathbf{y}|\theta) = k(\mathbf{y},\, \mathbf{z}) f(\mathbf{z}|\theta)$$

  Therefore,

$$f(\mathbf{x}|\theta) = \underbrace{k(\mathbf{x},\, \mathbf{y}) k(\mathbf{y},\, \mathbf{z})}_{=k(\mathbf{x},\, \mathbf{z})} f(\mathbf{z}|\theta)$$

  where $k(\mathbf{x},\, \mathbf{z}) \in (0,\, \infty)$, since $k(\mathbf{x},\, \mathbf{z}) = k(\mathbf{x},\, \mathbf{y}) k(\mathbf{y},\, \mathbf{z})$

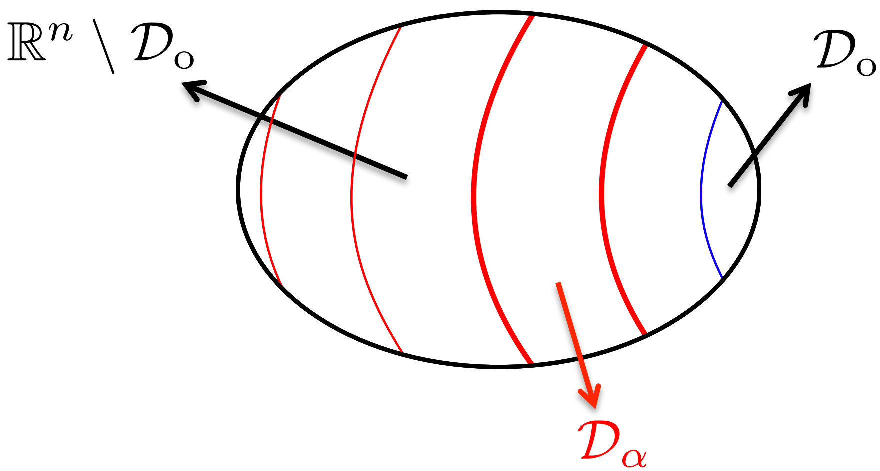
- Thus, $\mathbf{x} \sim \mathbf{y}$ is equivalent notation

Lecture 8 09/30/2013

CONSTRUCTING A MINIMAL SUFFICIENT STATISTIC

Constructing a minimal sufficient statistic (1 of 3)

- We can now partition $\mathbb{R}^n \setminus \mathcal{D}_\mathrm{o}$ into equivalent classes. For each class we can choose (by axiom of choice) a representative $\mathbf{x}(\alpha)$



Constructing a minimal sufficient statistic (2 of 3)

- We define the statistic based on the partition

$$T(\mathbf{x}) = \begin{cases} \mathbf{x}(\alpha), & \text{if } \mathbf{x} \in \mathcal{D}_\alpha \\ 0, & \text{if } \mathbf{x} \in \mathcal{D}_\text{o} \end{cases}$$

- In abstract sense all $\mathbf{x} \in \mathcal{D}_\alpha$ are represented by one "object" $\mathbf{x}(\alpha)$

- This partition induced by $\{f(\mathbf{x}|\theta) : \theta \in \Theta\}$ is called the "likelihood ratio" partition

Constructing a minimal sufficient statistic (3 of 3)

- Claim: $T(\mathbf{X})$ is sufficient statistic for θ
- Proof: For $\mathbf{x} \in \mathbb{R}^n \setminus \mathcal{D}_\mathrm{o}$,

  $\mathbf{x} \in \mathcal{D}_\alpha$ for some α with $\mathcal{D}_\alpha \subseteq \mathbb{R}^n \setminus \mathcal{D}_\mathrm{o}$

  we have $\mathbf{x} \sim \mathbf{x}_\alpha$. Therefore,

$$f(\mathbf{x}|\theta) = \underbrace{k(\mathbf{x},\, \mathbf{x}_\alpha)}_{h(\mathbf{x})}\,\underbrace{f(\mathbf{x}_\alpha|\theta)}_{g(T(\mathbf{x})|\theta)}$$

  For $\mathbf{x} \in \mathcal{D}_\mathrm{o}$,

  $f(\mathbf{x}_\alpha|\theta) = 0$.

  Thus, $h(\mathbf{x}) = 0$ and $g\,(T(\mathbf{x})|\theta)$ can be anything

Example 1 (1 of 3)

- Consider n i.i.d. observations of binomial, i.e.,

$$X \sim \text{binomial}(m,\, \theta)$$

Therefore

$$f(\mathbf{x}|\theta) = \left(\prod_{i=1}^{n} \binom{m}{x_i} \mathbb{I}_{\{0,1,\cdots,m\}}(x_i) \right) e^{mn \ln(1-\theta)} \exp\left\{ \sum_{i=1}^{n} x_i \ln\left(\frac{\theta}{1-\theta} \right) \right\}$$

Possible outcomes

$$x_i \in \{0,\, 1,\, \ldots,\, m\}$$

$$\mathbf{x} \in \{0,\, 1,\, \ldots,\, m\}$$

and

$$\mathcal{D}_{\mathrm{o}} = \mathbb{R}^n \setminus \{0,\, 1,\, \ldots,\, m\}$$

Example 1 (2 of 3)

- For $\mathbf{x},\, \mathbf{y} \notin \mathcal{D}_0$,

$$\frac{f(\mathbf{x}|\theta)}{f(\mathbf{y}|\theta)} = \prod_{i=1}^{n} \frac{\binom{m}{x_i}}{\binom{m}{y_i}} \exp\left\{ \ln\left(\frac{\theta}{1-\theta}\right) \left[\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i \right] \right\}$$

Thus, $\mathbf{x} \sim \mathbf{y}$ if and only if $\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$. Then

$$k(\mathbf{x},\, \mathbf{y}) = \prod_{i=1}^{n} \frac{\binom{m}{x_i}}{\binom{m}{y_i}}$$

- Equivalent class

$$\mathcal{D}_l = \left\{ \mathbf{x} : \sum_{i=1}^{n} x_i = l \right\}$$

for $l = 0,\, 1,\, 2,\, \ldots,\, mn$

Example 1 (3 of 3)

- Therefore, $T(\mathbf{x}) = \mathbf{x}_l$ if $\mathbf{x} \in \mathcal{D}_l$. This is equivalent to

$$T(\mathbf{x}) = \sum_{i=1}^{n} x_i$$

in a sense that partition stays the same

$$\mathcal{D}_l = \{\mathbf{x} : T(\mathbf{x}) = l\}$$

Lecture 8   09/30/2013

A SECOND DEFINITION OF MINIMAL SUFFICIENT STATISTIC

Partition of any sufficient statistic (1 of 3)

- Theorem: If $\widetilde{T}$ is any sufficient statistic for θ. Consider the partition

$$\widetilde{\mathcal{D}}_\alpha = \left\{\mathbf{x} : \widetilde{T}(\mathbf{x}) = \alpha\right\}$$

  Then for all α, there exists a β such that

$$\left(\widetilde{\mathcal{D}}_\alpha \setminus \mathcal{D}_\mathrm{o}\right) \subseteq \mathcal{D}_\beta$$

- Remark: This says that likelihood ratio partition is minimal

Partition of any sufficient statistic (2 of 3)

- Proof: Consider $\mathbf{x}, \mathbf{y} \in \widetilde{\mathcal{D}}_\alpha \setminus \mathcal{D}_\text{o}$. From sufficiency

$$f(\mathbf{x}|\theta) = h(\mathbf{x})g\left(\widetilde{T}(\mathbf{x})|\theta\right)$$

  We have $0 < h(\mathbf{x}) < \infty$:

  Since $\mathbf{x} \notin \mathcal{D}_\text{o} \Rightarrow$ For some θ_0 such that $f(\mathbf{x}|\theta_0) > 0$ $\qquad (\star)$

  $h(\mathbf{x}) = 0$ would imply $f(\mathbf{x}|\theta) = 0,\ \forall\theta$, contradicts $(\star)$

  $h(\mathbf{x}) = \infty$ would imply $g\big(\widetilde{T}(\mathbf{x})|\theta\big) = 0$, $\forall\theta$

$$\Rightarrow f(\mathbf{x}|\theta) = 0,\ \forall\theta \text{ contradicts } (\star)$$

- Similarly, this is true for $\mathbf{y}$. Then

$$\frac{f(\mathbf{x}|\theta)}{f(\mathbf{y}|\theta)} = \frac{h(\mathbf{x})}{h(\mathbf{y})} \frac{g\big(\widetilde{T}(\mathbf{x})|\theta\big)}{g\big(\widetilde{T}(\mathbf{y})|\theta\big)}$$

  Since $\mathbf{x}, \mathbf{y} \in \widetilde{\mathbf{D}}_\alpha$, we have $\widetilde{T}(\mathbf{x}) = \widetilde{T}(\mathbf{y})$

Partition of any sufficient statistic (3 of 3)

- Therefore,

$$\frac{f(\mathbf{x}|\theta)}{f(\mathbf{y}|\theta)} = \frac{h(\mathbf{x})}{h(\mathbf{y})} \triangleq k(\mathbf{x},\, \mathbf{y})$$

In particular, $k(\mathbf{x},\, \mathbf{y}) \in (0,\, \infty)$, since $h(\mathbf{x}),\, h(\mathbf{y}) \in (0,\, \infty)$. Thus,

$$\mathbf{x} \sim \mathbf{y} \;\;\Rightarrow\;\; \mathbf{x},\, \mathbf{y} \in \mathcal{D}_\beta \text{ for some } \beta.$$

Since this is true $\forall\, \mathbf{x},\, \mathbf{y} \in \widetilde{\mathcal{D}}_\alpha \setminus \mathcal{D}_\mathrm{o}$, we have

$$\left(\widetilde{\mathcal{D}}_\alpha \setminus \mathcal{D}_\mathrm{o}\right) \subseteq \mathcal{D}_\beta$$

A second definition of sufficient statistic

- By the previous theorem, one can define minimal sufficient statistic as follows
- Definition: A sufficient statistic T is <u>minimal</u> if the partition generated by T is likelihood ratio partition on $\mathbb{R}^n \setminus \mathcal{D}_o$

- Recipe for finding minimal sufficient statistic
  - Step 1: Look at the likelihood ratio
  - Step 2: Find the condition on $\mathbf{x}$ and $\mathbf{y}$ to make this ratio independent of θ

Example list

- Sufficient statistic: Binomial, Normal, Order statistic, Uniform [Casella & Berger]

- Minimal sufficient statistic: Normal minimal sufficient statistic, Uniform minimal sufficient statistic [Casella & Berger]

- Complete statistic: Binomial, Uniform, exponential family [Casella & Berger]

Lecture 8 09/30/2013

MINIMALITY AND EXPONENTIAL FAMILY

Minimality and exponential family (1 of 2)

- Consider

$$f(\mathbf{x}|\boldsymbol{\theta}) = h(\mathbf{x})c(\boldsymbol{\theta}) \exp\left\{\sum_{i=1}^{k} w_i(\boldsymbol{\theta})T_i(\mathbf{x})\right\}$$

  Let $\mathbf{x}, \mathbf{y} \notin \mathcal{D}_{\mathrm{o}}$,

$$\frac{f(\mathbf{x}|\boldsymbol{\theta})}{f(\mathbf{y}|\boldsymbol{\theta})} = \frac{h(\mathbf{x})}{h(\mathbf{y})} \exp\left\{\sum_{i=1}^{k} w_i(\boldsymbol{\theta})\left[T_i(\mathbf{x}) - T_i(\mathbf{y})\right]\right\}$$

  We want $\sum_{i=1}^{k} w_i(\boldsymbol{\theta})\left[T_i(\mathbf{x}) - T_i(\mathbf{y})\right]$ to be independent of $\boldsymbol{\theta}$.
  This is not straight forward in general.

- Remark: In general you cannot just do this by inspection

Minimality and exponential family (2 of 2)

- If $T_i(\mathbf{x}) = T_i(\mathbf{y})$, then

$$\sum_{i=1}^{k} w_i(\boldsymbol{\theta}) \left[T_i(\mathbf{x}) - T_i(\mathbf{y}) \right]$$

  is independent of $\boldsymbol{\theta}$.

- We <u>need</u> $T_i(\mathbf{x}) = T_i(\mathbf{y})$ <u>if and only if</u>

$$\sum_{i=1}^{k} w_i(\boldsymbol{\theta}) \left[T_i(\mathbf{x}) - T_i(\mathbf{y}) \right]$$

  is independent of $\boldsymbol{\theta}$.

Simple case

- $w_i(\boldsymbol{\theta}) = \theta_i$, Θ contains open ball in $\mathbb{R}^k$. Then

$$\sum_{i=1}^{k} \theta_i \left[T_i(\mathbf{x}) - T_i(\mathbf{y})\right]$$

is linear in $\theta_1, \theta_2, \ldots, \theta_k$. Since Θ contains an open ball in $\mathbb{R}^k$, we may look at the derivatives in every direction. If the above sum is independent of $\boldsymbol{\theta}$, then all the derivatives must be zero,

$$T_i(\mathbf{x}) = T_i(\mathbf{y}) \quad i = 1, 2, \ldots, k$$

Minimal sufficient statistic for exponential family

- If $w_i(\boldsymbol{\theta}) = \theta_i$, $\boldsymbol{\theta}$ is called natural parameterization.

- Proposition: For an exponential family in its natural parameterization and $\Theta \subseteq \mathbb{R}^k$ contains open ball in $\mathbb{R}^k$. The natural sufficient statistics are minimal.

Lecture 8 09/30/2013

COMPLETENESS

Complete family of distributions

- Definition: The family of probability distributions $f(\mathbf{x}|\theta)$, $\theta \in \Theta$ is called complete, if for every function $u(\mathbf{x})$, the identity

$$\mathbb{E}\left\{u(\mathbf{x})\right\} = 0$$

  implies that $u(\mathbf{x}) = 0$ except on the set of measure zero for each $\theta \in \Theta$, i.e., the set of all $\mathbf{x}$ such that $f(\mathbf{x}|\theta) > 0$ for some θ
- In other words, “the only unbiased estimator of zero is zero itself zero”

Complete sufficient statistic

- Definition: If a sufficient statistic T has a family of distribution $f(t|\theta)$ that is complete, then T is called a complete sufficient statistic

Example 1 (1 of 4)

- Let $X_1, X_2, \ldots, X_n$ be i.i.d. Poisson with unknown parameter $\mu > 0$. Find a minimal sufficient statistic for μ and check if it is complete.

$$f(x|\mu) = \begin{cases} \frac{e^{-\mu}\mu^{x_i}}{x_i!}, & x_i = 0,\ 1,\ \ldots \\ 0, & \text{otherwise} \end{cases}$$

- First, we have

$$f(\mathbf{x}|\mu) = \left(\prod_{i=1}^{n} \mathbb{I}_{\mathbb{N}_0}(x_i)\frac{1}{x_i!}\right)\exp\{-n\mu\}\exp\left\{\ln\mu\sum_{i=1}^{n} x_i\right\}$$

so $\sum_{i=1}^{n} x_i$ is sufficient statistic (We can use Theory of Exponential family or straight Neyman Factorization Theorem)

Example 1 (2 of 4)

- To show that $\sum_{i=1}^{n} x_i$ is minimal sufficient statistic for μ, we consider the ratio

$$\frac{f(\mathbf{x}|\mu)}{f(\mathbf{y}|\mu)}$$

This ratio to be independent of μ if and only if

$$\ln \mu \left[\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i \right]$$

is independent of μ, which holds if and only if

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$$

Thus, $T(\mathbf{X}) = \sum_{i=1}^{n} X_i$ is minimal

Example 1 (3 of 4)

- Now we show completeness. We know the pmf of T is again Poisson (in general it will be in exponential family again)

$$f(t|\mu) = \begin{cases} \frac{e^{-n\mu}(n\mu)^t}{t!}, & t = 0,\ 1,\ 2,\ \ldots \\ 0, & \text{otherwise} \end{cases}$$

Let $u(T)$ be any function of T. Then $\mathrm{E}\{u(T)\} = 0$, $\forall \mu$ implies

$$\sum_{t=0}^{\infty} u(t) \frac{e^{-n\mu}(n\mu)^t}{t!} = 0$$

Since $e^{-n\mu} > 0$,

$$\sum_{t=0}^{\infty} \frac{u(t) n^t}{t!} \mu^t = 0$$

Example 1 (4 of 4)

- Recall that two power series $\sum_{j=1}^{\infty} a_j x^j = \sum_{j=1}^{\infty} b_j x^j$ have the same value for every x in some interval if and only if $a_i = b_i$, $i = 0,\ 1,\ 2,\ \ldots$

  Since the two "power series" are equal, their coefficients are equal.

- In particular, we have

$$\frac{u(t)n^t}{t!} = 0, \quad \text{for } t = 0,\ 1,\ 2,\ \ldots$$

  and thus $u(t) = 0$. Thus

$$T(\mathbf{X}) = \sum_{i=1}^{n} X_i$$

  is complete sufficient statistic

Ancillary Statistic

- Sufficient statistic, in a sense, contain all the information about θ that is available in the sample. Ancillary statistic has a complementary purpose.

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

- An ancillary statistic alone contains no information about θ. However, when used in conjunction with other statistic, it contains information for inferences about θ.