6.434/16.391 Statistics for Engineers and Scientists

Lecture 6 09/23/2013

Laboratory for Information and Decision Systems
Massachusetts Institute of Technology

Lecture 6 09/23/2013

DEFINITION OF SUFFICIENT STATISTIC

A useful fact

- For general discrete random variables the following is true. Let $q(t|\theta)$ be the pmf of $T(\mathbf{X})$, where $\mathbf{X} = [X_1 \ X_2 \ \cdots \ X_n]^{\mathrm{T}}$ is random sample from distribution of discrete type $f(x|\theta)$. Then the conditional probability of $\mathbf{X} = \mathbf{x}$ given $T = t$ is given by

$$\mathbb{P}\{\mathbf{X} = \mathbf{x} | T(\mathbf{X}) = t\} = \begin{cases} \frac{\Pi_{i=1}^{n} f(x_i|\theta)}{q(t|\theta)} & \text{if } T(\mathbf{X}) = t \\ 0, & \text{otherwise} \end{cases}$$

Definition of sufficient statistic

- Definition: A statistic $T(\mathbf{X})$ is sufficient for θ if the conditional distribution of $\mathbf{X}$ given $T(\mathbf{X}) = t$ does not involve θ

- Remark: The use of terminology for "sufficient statistic" can be explained as follows:

$$\mathbb{P}\{\mathbf{X} = \mathbf{x} | T(\mathbf{X}) = t, \theta\} = \mathbb{P}\{\mathbf{X} = \mathbf{x} | T(\mathbf{X}) = t\}$$

$$\mathbb{P}\{Y = y | T(\mathbf{X}) = t, \theta\} = \mathbb{P}\{Y = y | T(\mathbf{X}) = t\}$$

  for any $Y = g(\mathbf{X})$.
  i.e., once $T(\mathbf{X})$ is known, the sample $\mathbf{X}$, or any other function of it, say $g(\mathbf{X})$, contains no further information about θ

Remarks

- In a sense, $T(\mathbf{X})$ exhausts all the information about θ that is contained in the sample $\mathbf{X}$
- Intuitively, if there exists a sufficient statistic $T(\mathbf{X})$ for θ, any estimator of θ should be a function only of $T(\mathbf{X})$
  - For formal proof, see Neyman Factorization Theorem

Example 1 (1 of 2)

- Let $X_1,\, X_2,\, \ldots,\, X_n$ be independent random variables each with uniform distribution, i.e., $X_i \sim \mathrm{uniform}(0,\, \theta)$. Let

$$Y_n = \max\{X_1,\, X_2,\, \ldots,\, X_n\}$$

$$Y_1 = \min\{X_1,\, X_2,\, \ldots,\, X_n\}$$

Then

$$\begin{aligned} f_{\mathbf{X}|Y_n}(\mathbf{x}|y_n,\, \theta) &= \frac{f_{\mathbf{X},Y_n}(\mathbf{x},\, y_n,\, \theta)}{f_{Y_n}(y_n,\, \theta)} \\ &= \frac{1/\theta^n}{n y_n^{n-1}/\theta^n} \\ &= \frac{1}{n y_n^{n-1}} \end{aligned}$$

which is independent of θ

Example 1 (2 of 2)

- On the other hand

$$f_{\mathbf{X}|Y_1}(\mathbf{x}|y_1,\,\theta) = \frac{f_{\mathbf{X},Y_1}(\mathbf{x},\,y_1,\,\theta)}{f_{Y_1}(y_1,\,\theta)} = \frac{1/\theta^n}{n(\theta-y_1)^{n-1}/\theta^n} = \frac{1}{n(\theta-y_1)^{n-1}}$$

which depends on θ

- Remark:
  - Y_1 and Y_n each reduces the data to a single number,
  - Y_1 does lose information about θ
  - Y_n does not lose any information about θ

Example 2

- $X_1, X_2, \ldots, X_n$ are i.i.d. Poisson random variable with parameter λ. Let $T(\mathbf{X}) = \sum_{i=1}^n X_i$. Then

$$\mathbb{P}_\lambda\{\mathbf{X} = \mathbf{x} | T(\mathbf{X}) = t\} = \begin{cases} (\frac{1}{n})^t \frac{t!}{x_1! x_2! \cdots x_n!}, & \text{if } \sum_{i=1}^n x_i = t \\ 0, & \text{otherwise} \end{cases}$$

which does not depend on λ. Therefore $T(\mathbf{X}) = \sum_{i=1}^n X_i$ is a sufficient statistic for λ. Moreover, the above expression is the multinomial distribution.

- On the other hand, let $T(\mathbf{X}) = \bar{X} = \sum_{i=1}^n X_i / n$, we have

$$\mathbb{P}_\lambda\{\mathbf{X} = \mathbf{x} | T(\mathbf{X}) = t\} = \begin{cases} (\frac{1}{n})^{nt} \frac{(nt)!}{x_1! x_2! \cdots x_n!}, & \text{if } \sum_{i=1}^n x_i = nt \\ 0, & \text{otherwise} \end{cases}$$

Thus, $\bar{X}$ is also a sufficient statistic. Therefore, sufficient statistics are not unique

Detailed derivation for Example 2: Multinomial distribution (1 of 2)

- Consider $T(\mathbf{X}) = \sum_{i=1}^{n} X_i$, we have

$$\mathbb{P}_\lambda\{\mathbf{X} = \mathbf{x} | T(\mathbf{X}) = t\} = \frac{\mathbb{P}_\lambda\{\mathbf{X} = \mathbf{x}, T(\mathbf{X} = t)\}}{\mathbb{P}_\lambda\{T(\mathbf{X} = t)\}}$$

in which

$$\mathbb{P}_\lambda\{\mathbf{X} = \mathbf{x}, T(\mathbf{X} = t)\} = \begin{cases} \mathbb{P}_\lambda\{X_1 = x_1, \cdots, X_n = x_n\}, & \text{if} \sum_{i=1}^{n} x_i = t \\ 0, & \text{otherwise} \end{cases}$$

where

$$\begin{aligned} \mathbb{P}_\lambda\{X_1 = x_1, \ldots, X_n = x_n)\} &= \prod_{i=1}^{n} \mathbb{P}\{X_i = x_i\} \\ &= \frac{e^{-n\lambda} \lambda^{\sum_{i=1}^{n} x_i}}{x_1! x_2! \cdots x_n!} \\ &= \frac{e^{-n\lambda} \lambda^{t}}{x_1! x_2! \cdots x_n!} \end{aligned}$$

Detailed derivation for Example 2: Multinomial distribution (2 of 2)

- Further,

$$P_\lambda\left\{T(\mathbf{X})=t\right\}=P_\lambda\left\{\sum_{i=1}^{n}X_i=t\right\}=\frac{e^{-n\lambda}(n\lambda)^t}{t!}$$

  – Note: the sum of independent Poisson r.v is also a Poisson r.v.

- Therefore, when $x_i \geq 0$ and $\sum_{i=1}^{n} X_i = t$,

$$\begin{aligned}P_\lambda\{\mathbf{X}=\mathbf{x}|T(\mathbf{X})=t\} &= \frac{e^{-n\lambda}\lambda^t}{x_1!x_2!\cdots x_n!}\frac{t!}{e^{-n\lambda}(n\lambda)^t} \\ &= \frac{t!}{x_1!\cdots x_n!}\left(\frac{1}{n}\right)^t \\ &= \frac{t!}{x_1!\cdots x_n!}\left(\frac{1}{n}\right)^{x_1}\cdots\left(\frac{1}{n}\right)^{x_n}\end{aligned}$$

which is a multinomial distribution

Lecture 6 09/23/2013

NEYMAN FACTORIZATION THEOREM

Motivation

- Direct verification of the definition can be quite difficult (although it is easy to understand intuitively)

- The Neyman Factorization Theorem gives another criterion

Neyman Factorization Theorem

- Theorem: A statistic $T(\mathbf{X})$ is sufficient statistic for θ if and only if there exists a function $g(t|\theta)$, $t \in \mathbb{T}$, $\theta \in \Theta$ and a function $h(\mathbf{x})$ such that

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

Proof for "if part" (1 of 3)

- Suppose $T(\mathbf{X})$ is sufficient. Let

$$\mathcal{S} = \{\mathbf{x} : \ f(\mathbf{x}|\theta) > 0, \ \text{for some } \theta \in \Theta\}$$

Then for $\mathbf{x} \in \mathcal{S}$, by the definition of sufficient statistic,

$$\mathbb{P}_\theta\{\mathbf{X} = \mathbf{x}|T(\mathbf{X}) = t\} = \mathbb{P}\{\mathbf{X} = \mathbf{x}|T(\mathbf{X}) = t\}$$

where the right hand side does not depend on θ.

Proof for "if part" (2 of 3)

- Then we have

$$\begin{aligned} f(\mathbf{x}|\theta) &= \mathbb{P}_\theta\{\mathbf{X}=\mathbf{x}\} \\ &\overset{(a)}{=} \mathbb{P}_\theta\left\{\mathbf{X}=\mathbf{x},\, T(\mathbf{X})=t\right\} \quad \text{where } t=T(\mathbf{x}) \\ &= \mathbb{P}_\theta\left\{\mathbf{X}=\mathbf{x}|T(\mathbf{X})=t\right\}\mathbb{P}_\theta\left\{T(\mathbf{X})=t\right\} \\ &\overset{(b)}{=} \underbrace{\mathbb{P}\left\{\mathbf{X}=\mathbf{x}|T(\mathbf{X})=t\right\}}_{\triangleq h(\mathbf{x})}\underbrace{\mathbb{P}_\theta\left\{T(\mathbf{X})=t\right\}}_{\triangleq g(T(\mathbf{x})|\theta\}} \end{aligned}$$

where (b) is because $T(\mathbf{X})$ is sufficient statistic, and (a) follows from the fact that $\{\mathbf{X}=\mathbf{x}\}=\{\mathbf{X}=\mathbf{x},\, T(\mathbf{X})=T(\mathbf{x})\}$

  – Proof is in the next slide

- Thus

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

Proof for “if part” (3 of 3)

- Claim: $\{\mathbf{X}=\mathbf{x}\} = \{\mathbf{X}=\mathbf{x}, T(\mathbf{X}) = T(\mathbf{x})\}$
- Proof: we need to show
  (i) $\{\mathbf{X}=\mathbf{x}\} \supseteq \{\mathbf{X}=\mathbf{x}, T(\mathbf{X}) = T(\mathbf{x})\}$
  (ii) $\{\mathbf{X}=\mathbf{x}\} \subseteq \{\mathbf{X}=\mathbf{x}, T(\mathbf{X}) = T(\mathbf{x})\}$
- i) is obvious (since RHS is a restricted set of LHS)
  ii) is proved as follows.
  For every $\omega \in \{\mathbf{X}=\mathbf{x}\}$, $\mathbf{X}(\omega) = \mathbf{x}$
  and we have $T\left(\mathbf{X}(\omega)\right) = T\left(\mathbf{x}\right)$. Thus,

$$\omega \in \{\mathbf{X}=\mathbf{x}, T(\mathbf{X}) = T(\mathbf{x})\}$$

  Therefore, $\{\mathbf{X}=\mathbf{x}\} = \{\mathbf{X}=\mathbf{x}, T(\mathbf{X}) = T(\mathbf{x})\}$

Proof for “only if part” (1 of 4)

- For $\mathbf{x} \in \mathcal{S}$, suppose

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

We need to show that

$$\mathbb{P}_\theta\{\mathbf{X} = \mathbf{x}|T(\mathbf{X}) = t\} = \frac{\mathbb{P}_\theta\{\mathbf{X} = \mathbf{x}, T(\mathbf{x}) = t\}}{\mathbb{P}_\theta\{T(\mathbf{X}) = t\}}$$

is independent of θ

Proof for "only if part" (2 of 4)

- Claim: A) $\mathbb{P}_\theta\{\mathbf{X}=\mathbf{x}, T(\mathbf{X})=t\} = \begin{cases} h(\mathbf{x})g(T(\mathbf{x})|\theta), & \text{for } T(\mathbf{x})=t \\ 0, & \text{otherwise} \end{cases}$

  B) $\mathbb{P}_\theta\{T(\mathbf{X})=t\} = \left[\sum_{\mathbf{x}:T(\mathbf{x})=t} h(\mathbf{x})\right] g\left(t|\theta\right)$

- Finally,

$$\mathbb{P}_\theta\{\mathbf{X}=\mathbf{x}|T(\mathbf{X})=t\} = \begin{cases} \frac{h(\mathbf{x})}{\sum_{\mathbf{x}:T(\mathbf{x})=t} h(\mathbf{x})}, & \text{for } T(\mathbf{x})=t \\ 0, & \text{otherwise} \end{cases}$$

which is independent of θ. Therefore, $T(\mathbf{X})$ is a sufficient statistic for θ

Proof for "only if part" (3 of 4)

- Proof (A): Numerator

$$\mathbb{P}_\theta\{\mathbf{X} = \mathbf{x}, T(\mathbf{X}) = t\} = \begin{cases} \text{Non-zero}, & \text{for } T(\mathbf{x}) = t \\ 0, & \text{otherwise} \end{cases}$$

For $T(\mathbf{x}) = t$

$$\begin{aligned} \mathbb{P}_\theta\{\mathbf{X} = \mathbf{x},\, T(\mathbf{X}) = t\} &\overset{(a)}{=} \mathbb{P}_\theta\{\mathbf{X} = \mathbf{x}\} \\ &= h(\mathbf{x}) g\left(T(\mathbf{x})|\theta\right) \quad \text{(by hypothesis)} \end{aligned}$$

otherwise,

$$\mathbb{P}_\theta\{\mathbf{X} = \mathbf{x},\, T(\mathbf{X}) = t\} = 0$$

Proof for "only if part" (4 of 4)

- Proof (B): Denominator

  $\mathbb{P}_\theta\{T(\mathbf{X}) = t\}$ can be obtained as

$$\begin{aligned}
\mathbb{P}_\theta\{T(\mathbf{X}) = t\} &= \sum_{\mathbf{x} \in \mathcal{S}} \mathbb{P}_\theta\{\mathbf{X} = \mathbf{x},\, T(\mathbf{X}) = t\} \quad \text{(marginalization)} \\
&= \sum_{\mathbf{x}:T(\mathbf{x})=t} \mathbb{P}_\theta\{\mathbf{X} = \mathbf{x},\, T(\mathbf{X}) = t\} + \sum_{\mathbf{x}:T(\mathbf{x})\neq t} \underbrace{\mathbb{P}_\theta\{\mathbf{X} = \mathbf{x},\, T(\mathbf{X}\} = t)}_{0} \\
&= \left[\sum_{\mathbf{x}:T(\mathbf{x})=t} h(\mathbf{x}) \right] g\left(t|\theta\right)
\end{aligned}$$