6.434/16.391 Statistics for Engineers and Scientists

Lecture 7 9/25/2013

Laboratory for Information and Decision Systems
Massachusetts Institute of Technology

Lecture 7 9/25/2013

EXAMPLES OF NEYMAN FACTORIZATION THEOREM

Example 1

- Let $X_1, X_2, \ldots, X_n$ be random sample from displaced exponential distribution, i.e.,

$$f(x|\theta) = e^{-(x-\theta)}\mathbb{I}_{\{X>\theta\}}(x)$$

Then we have

$$\begin{aligned} f(\mathbf{x}|\theta) &= e^{-\sum_{i=1}^{n}(x_i-\theta)}\prod_{i=1}^{n}\mathbb{I}_{\{X_i>\theta\}}(x_i) \\ &= \underbrace{e^{-\sum_{i=1}^{n}x_i}}_{=h(\mathbf{x})}\underbrace{e^{n\theta}\mathbb{I}_{\{Y_1>\theta\}}(y_1)}_{=g(T(\mathbf{x})|\theta)} \\ &= h(\mathbf{x})g\left(T(\mathbf{x})|\theta\right) \end{aligned}$$

where $T(\mathbf{X}) = Y_1 = \min\{X_i\}$. By Neyman Factorization Theorem, $T(\mathbf{X}) = \min\{X_i\}$ is sufficient statistic for θ

Example 2

- Let $X_1, X_2, \ldots, X_n$ denote a random sample from a distribution with pdf

$$f(x|\theta) = \begin{cases} \theta x^{\theta-1}, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

Then we have

$$f(\mathbf{x}|\theta) = \theta^n \prod_{i=1}^{n} x_i^{\theta-1} \mathbb{I}_{\{0<X_i<1\}}(x_i)$$

$$= \underbrace{\frac{\prod_{i=1}^{n} \mathbb{I}_{\{0<X_i<1\}}(x_i)}{\prod_{i=1}^{n} x_i}}_{=h(\mathbf{x})} \cdot \underbrace{\theta^n \left(\prod_{i=1}^{n} x_i \right)^{\theta}}_{=g(T(\mathbf{x})|\theta)}$$

By Neyman Factorization Theorem, $T(\mathbf{x}) = \prod_{i=1}^{n} X_i$ is sufficient for θ

Example 3

- Let $X_1, X_2, \ldots, X_n$ be i.i.d. $N\left(\mu, \sigma^2\right)$ with $\theta = \left(\mu, \sigma^2\right)$. Then

$$f\left(\mathbf{x}|\theta\right) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}\left(x_i - \mu\right)^2\right\}$$

$$= \underbrace{\left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{-\frac{\sum_{i=1}^{n} x_i^2}{2\sigma^2} + \frac{\mu \sum_{i=1}^{n} x_i}{\sigma^2}\right\} \exp\left\{-\frac{n\mu^2}{2\sigma^2}\right\}}_{g(T(\mathbf{x}|\theta))}$$

Thus, we have $h(\mathbf{x}) = 1$ and

$$T(\mathbf{X}) = \left(\sum_{i=1}^{n} X_i^2, \sum_{i=1}^{n} X_i\right)$$

is sufficient for $\theta = \left(\mu, \sigma^2\right)$. Moreover, it can be verified that

$$T(\mathbf{X}) = \left(\bar{X}, \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2\right)$$

is also sufficient for θ, where $\bar{X}$ and $\frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2$ are sample mean and sample variance, respectively

Example 4

- Suppose $X_1, X_2, \ldots, X_n$ are i.i.d. with distribution $N(\mu, 1)$. We want to find sufficient statistic for μ. First,

$$f(\mathbf{x}|\theta) = \left(\frac{1}{2\pi}\right)^{n/2} \exp\left\{-\frac{\sum_{i=1}^n x_i^2}{2} + \mu\sum_{i=1}^n x_i - \frac{n\mu}{2}\right\}$$

Apparently, $T(\mathbf{X}) = \left(\sum_{i=1}^n X_i^2, \sum_{i=1}^n X_i\right)$ is sufficient, but one could get "smaller" sufficient statistic for μ as

$$T_2(\mathbf{X}) = \sum_{i=1}^n X_i$$

and since $f(\mathbf{x}|\theta)$ can be factorized with

$$h(\mathbf{x}) = \exp\left\{-\frac{\sum_{i=1}^n x_i^2}{2}\right\} \text{ and } g(T(\mathbf{x})|\mu) = \left(\frac{1}{2\pi}\right)^{n/2} \exp\left\{\mu\sum_{i=1}^n x_i - \frac{n\mu}{2}\right\}$$

Thus, $T_2(\mathbf{X})$ is "smaller" sufficient statistic for μ. Moreover, $\bar{X}$ is also sufficient statistic for μ.

Example 5

- Let $X_1, X_2, \ldots, X_n$ be random sample from distribution with pdf $f(x|\theta)$. Due to independence,

$$f(\mathbf{x}|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$$

Now order X_i's as $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$, we have

$$f(\mathbf{x}|\theta) = \underbrace{\prod_{i=1}^{n} f\left(x_{(i)}|\theta\right)}_{\text{think of this as } g(T(\mathbf{x})|\theta)}$$

Therefore, the ordered data are already sufficient statistic for θ

Example 6 (1 of 3)

- Recall the definition of exponential family: A family of pdfs or pmfs is called exponential family if it can be expressed as

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

where $h(\mathbf{x}) \geq 0$; $t_1(x), t_2(x), \ldots, t_k(x)$ are real valued function of observation x; $c(\boldsymbol{\theta}) \geq 0$; and $w_1(\boldsymbol{\theta}), w_2(\boldsymbol{\theta}), \ldots, w_k(\boldsymbol{\theta})$ are real valued function of possibly vector-valued $\boldsymbol{\theta}$

Example 6 (2 of 3)

- Let $X_1, X_2, \ldots, X_n$ be i.i.d. observations from a pdf or pmf $f(x|\theta)$ that belongs to an exponential family given by

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

where $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_d)$, $d \leq k$. Then we have

$$f(\mathbf{x}|\boldsymbol{\theta}) = \underbrace{\prod_{j=1}^{n} h(x_j)}_{=h(\mathbf{x})} \underbrace{c^n(\boldsymbol{\theta}) \exp\left\{\sum_{i=1}^{k} w_i(\boldsymbol{\theta}) \sum_{j=1}^{n} t_i(x_j)\right\}}_{=g(T(\mathbf{x})|\boldsymbol{\theta})}$$

Therefore,

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

is a sufficient statistic for $\boldsymbol{\theta}$

Example 6 (3 of 3)

- In fact, $T(\mathbf{X})$ is called natural sufficient statistic.

- Natural sufficient statistic is derived specifically for exponential family.

Lecture 7 9/25/2013

SEVERAL FACTS ABOUT SUFFICIENT STATISTIC

Function of sufficient statistic

- Lemma: Function of a sufficient statistic is also a sufficient statistic if the function is invertible

- Proof: Let $T(\mathbf{X})$ be sufficient statistic for θ. Let

$$S(\mathbf{X}) = u\left(T(\mathbf{X})\right)$$

where $u(\,\cdot\,)$ is any invertible function. By Neyman Factorization Theorem,

$$\begin{aligned} f(\mathbf{x}|\theta) &= h(\mathbf{x}) g\left(T(\mathbf{X})|\theta\right) \\ &= h(\mathbf{x}) \underbrace{g\left(u^{-1}\left(S(\mathbf{X})\right)|\theta\right)}_{\triangleq \tilde{g}(S(\mathbf{x})|\theta)} \end{aligned} \qquad \tilde{g}(\cdot) = g \circ u^{-1}(\cdot)$$

By Factorization Theorem, $S(\mathbf{X})$ is also sufficient for θ

Sufficient statistic and maximum likelihood estimator (1 of 2)

- Lemma: If a sufficient statistic $T(\mathbf{X})$ exists and a unique maximum likelihood estimator $\hat{\theta}$ of θ exists, then $\hat{\theta}$ is a function of $T(\mathbf{X})$
- Proof: Since $T(\mathbf{X})$ is sufficient statistic of θ,
$$f(\mathbf{x}|\theta) = h(\mathbf{x})g\left(T(\mathbf{x})|\theta\right)$$
The likelihood function is
$$L(\theta|\mathbf{X}) = f(\mathbf{x}|\theta) = h(\mathbf{x})g\left(T(\mathbf{x})|\theta\right)$$
The maximum likelihood estimator (MLE) is given by
$$\begin{aligned}\hat{\theta} &= \arg\max_{\theta} L(\theta|\mathbf{X}) \\ &= \arg\max_{\theta} h(\mathbf{x})g\left(T(\mathbf{x})|\theta\right) \\ &= h(\mathbf{x}) \arg\max_{\theta} g\left(T(\mathbf{x})|\theta\right)\end{aligned}$$

Sufficient statistic and maximum likelihood estimator (2 of 2)

- Therefore, MLE can be calculated as

$$\hat{\theta} = \arg\max_{\theta} g\left(T(\mathbf{x})|\theta\right)$$

which is a function of $T(\mathbf{x})$. Since MLE is unique (by hypothesis), we conclude that MLE $\hat{\theta}$ must be a function of $T(\mathbf{x})$

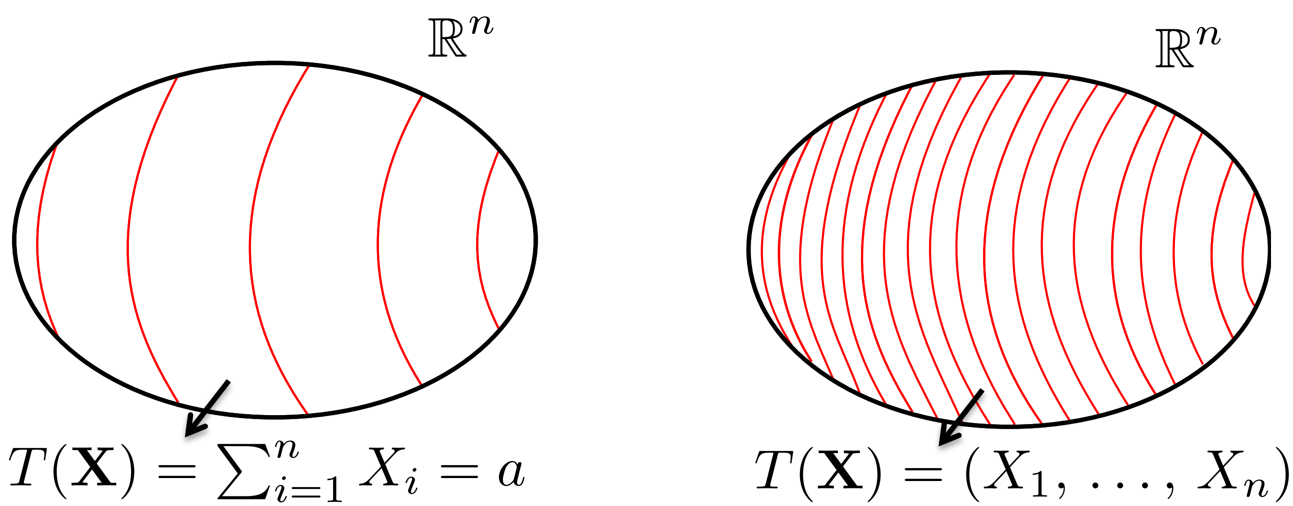
- Note: This lemma does <u>not</u> imply that any MLE is a function of sufficient statistic. However, if MLE exists, then <u>one</u> MLE can be found as a function of sufficient statistic

Lecture 7 9/25/2013

MINIMAL SUFFICIENT STATISTIC

Motivation and intuition

- We saw that there are more than one sufficient statistic. How should we choose?
- We adopt a "Geometric view"



- Sufficient statistic can be thought of as partition of the space
- Minimal sufficient statistic is the sufficient statistic with "coarsest" partition

Minimal sufficient statistic

- Definition: $T(\mathbf{X})$ is a minimal sufficient statistic if $T(\mathbf{X})$ is a sufficient statistic and a function of every other sufficient statistic

Equivalent sufficient statistic

- Definition: The sufficient statistic T_1 and T_2 are equivalent if for $\forall \mathbf{x}, \mathbf{y}$

$$T_1(\mathbf{x}) = T_1(\mathbf{y}) \Leftrightarrow T_2(\mathbf{x}) = T_2(\mathbf{y})$$

- Example: Let $T_1(\mathbf{X}) = \sum_{i=1}^{n} X_i$ and $T_2(\mathbf{X}) = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then T_1 and T_2 are equivalent, since $\forall \mathbf{x}, \mathbf{y}$,

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

- Sufficiency is preserved under equivalence
  - This fact can be verified by showing that two equivalent sufficient statistic are invertible functions of each other

- Next we will use the notion of "equivalence" more formally