STAT 611-600

Theory of Statistics - Inference
Lecture 2: More on Sufficient Stats

Tiandong Wang

Department of Statistics
Texas A&M University

How to check sufficiency?

- **Conditional Probability**(discrete case)
  For any x and t, we have

$$P_\theta(\mathbf{X} = \mathbf{x} | T(\mathbf{X}) = t) = \begin{cases} P_\theta(\mathbf{X} = \mathbf{x} | T(\mathbf{X}) = T(\mathbf{x})) & \text{when } T(\mathbf{x}) = t \\ 0 & \text{when } T(\mathbf{x}) \neq t \end{cases}$$

- Let the distribution of data $\mathbf{X}$ be $p(\mathbf{x}; \theta)$ and the distribution of T be $q(t; \theta)$. The t-th conditional is $p(\mathbf{x}; \theta)/q(t; \theta)$ on A_t.
  - This should be free of θ (but may depend on $\mathbf{x}$) for all t, if T is sufficient for θ.
  - And if $p(\mathbf{x}; \theta)/q(T(\mathbf{x}); \theta)$ is free of θ for all x and θ, then T is a sufficient statistic for θ.

How to check sufficiency: Examples

- Example. $X_1, \cdots, X_n$ iid $Poisson(\lambda). T = \sum_{i=1}^{n} X_i$.
- Remarks: conditional rule is not convenient to apply.
  - Need to guess the form of a sufficient statistic.
  - Need to figure out the distribution of T.

How to check sufficiency: Examples

- Example. $X_1, \cdots, X_n$ iid $Poisson(\lambda). T = \sum_{i=1}^{n} X_i$.
- Remarks: conditional rule is not convenient to apply.
  - Need to guess the form of a sufficient statistic.
  - Need to figure out the distribution of T.

How to find a sufficient statistic

Theorem

(Neyman-Fisher) Factorization theorem.
*Let $p_{\boldsymbol{X}}(\boldsymbol{x};\theta)$ be the pdf or pmf of a sample $\boldsymbol{X}$. $T(\boldsymbol{X})$ is sufficient **if and only if** there exists functions $g(t;\theta)$ and $h(\boldsymbol{x})$ s.t. $\forall \boldsymbol{x}, \theta$,*

$$p(\boldsymbol{x};\theta) = g(T(\boldsymbol{x});\theta)h(\boldsymbol{x}).$$

Note:
The first factor depends on $\mathbf{x}$ only though $T(\mathbf{x})$ and the second factor is free of θ.

How to find a sufficient statistic: Examples

- Example1 $X_1, \cdots, X_n$ iid $N(\theta, 1)$.
- Example2 $X_1, \cdots, X_n$ iid $Binomial(1, \theta)$
- Example3 $X_1, \cdots, X_n$ iid $Poisson(\theta)$.
- Example4 $X_1, \cdots, X_n$ iid Exponential(β)

How to find a sufficient statistic: Examples

- Example. Uniform. iid ($U(0, \theta)$).
  Note: When the range of X depends on θ, should be more careful about factorization. Must use indicator functions explicitly.
- Two-dimensional Examples.
  Normal. iid $N(\mu, \sigma^2).\theta = (\mu, \sigma^2)$ (both unknown)
  Gamma. iid $Ga(\alpha, \beta).\theta = (\alpha, \beta)$.

Proof of the Theorem

Sufficiency⇒Factorization:

Suppose $T(\mathbf{X})$ is sufficient, then

$$\begin{aligned}
p_{\mathbf{X}}(\mathbf{x};\theta) &= \mathbb{P}(\mathbf{X}=\mathbf{x};\theta) \\
&= \mathbb{P}(\mathbf{X}=\mathbf{x}, T(\mathbf{X})=T(\mathbf{x});\theta) \\
&= \mathbb{P}(T(\mathbf{X})=T(\mathbf{x});\theta)\mathbb{P}(\mathbf{X}=\mathbf{x}|T(\mathbf{X})=T(\mathbf{x});\theta) \\
&=: g(T(\mathbf{x});\theta)h(\mathbf{x}),
\end{aligned}$$

where the last step follows from the definition of a sufficient stat.

Sufficiency$\Leftarrow$Factorization:

Suppose the factorization holds, i.e.

$$p(\mathbf{x};\theta) = g(T(\mathbf{x});\theta)h(\mathbf{x}).$$

Then

$$\begin{aligned}
\mathbb{P}(\mathbf{X}=\mathbf{x}|T(\mathbf{X})=T(\mathbf{x});\theta) &= \frac{\mathbb{P}(\mathbf{X}=\mathbf{x}, T(\mathbf{X})=T(\mathbf{x});\theta)}{\sum_{\mathbf{x}':T(\mathbf{x}')=T(\mathbf{x})} p_{\mathbf{X}}(\mathbf{x}';\theta)} \\
&= \frac{g(T(\mathbf{x});\theta)h(\mathbf{x})}{\sum_{\mathbf{x}':T(\mathbf{x}')=T(\mathbf{x})} g(T(\mathbf{x});\theta)h(\mathbf{x}')} \\
&= \frac{h(\mathbf{x})}{\sum_{\mathbf{x}':T(\mathbf{x}')=T(\mathbf{x})} h(\mathbf{x}')},
\end{aligned}$$

which does not depend on θ.

Good News: Exponential family

Exponential Family: Recall the density function of an exponential family

$$f(x;\theta) = c(\theta)h(x)\exp[\sum_{j=1}^{k} w_j(\theta)t_j(x)], \theta = (\theta_1, \cdots, \theta_d).$$

Theorem

Let $X_1, \cdots, X_n$ be a random sample from the exponential family. Then

$$T(\boldsymbol{X}) = (\sum_{i=1}^{n} t_1(\boldsymbol{X}_i), \cdots, \sum_{i=1}^{n} t_k(\boldsymbol{X}_i))$$

is sufficient for $\theta = (\theta_1, \cdots, \theta_d)$.

- Applies to many standard families discussed above such as binomial, Poisson, normal, exponential, gamma.

Example

Show that the gamma distribution belongs to the exponential family and find the sufficient stats.

Minimality about $T(X)$

An exponential family is referred to as **minimal** if there are no linear constraints among the components of the parameter vector $(\{w_j(\theta) : 1 \le j \le k\})$ nor are there linear constraints among the components of the sufficient statistic $(\{t_j(\mathbf{X}) : 1 \le j \le k\})$, i.e.

- $\sum_{i=1}^{k} \alpha_k w_i(\theta) = \alpha_0 \Rightarrow \alpha_i = 0$, for all $i = 0, 1, \ldots, k$.
- $\sum_{i=1}^{k} \alpha_k t_i(x) = \alpha_0 \Rightarrow \alpha_i = 0$, for all $i = 0, 1, \ldots, k$.

Checking if an exponential family is **full rank** is equivalent to checking the following two conditions:

- It is minimal.
- The parameter set $\{w_j(\theta) : 1 \le j \le k\}$ contains a k-dimensional open rectangle.

Examples:

1. Consider $N(\mu, \sigma^2)$, where $w_1 = 1/(2\sigma^2)$, $w_2 = \mu/\sigma^2$, $t_1(x) = -x^2$, $t_2(x) = x$.
   1. Non-minimal: e.g. when $\mu = \sigma^2$, $w_1 = 1/(2\sigma^2)$, $w_2 = 1$.
   2. Minimal & Curved: e.g. when $\mu = \sqrt{\sigma^2}$, $w_1 = 1/(2\sigma^2)$, $w_2 = 1/\sqrt{\sigma^2}$ so that $2w_1 = w_2^2$.
   3. Minimal & Full-Rank: e.g., no extra constraint.
2. Multinomial distribution.

Examples:

1. Consider $N(\mu, \sigma^2)$, where $w_1 = 1/(2\sigma^2)$, $w_2 = \mu/\sigma^2$, $t_1(x) = -x^2$, $t_2(x) = x$.
    1. Non-minimal: e.g. when $\mu = \sigma^2$, $w_1 = 1/(2\sigma^2)$, $w_2 = 1$.
    2. Minimal & Curved: e.g. when $\mu = \sqrt{\sigma^2}$, $w_1 = 1/(2\sigma^2)$, $w_2 = 1/\sqrt{\sigma^2}$ so that $2w_1 = w_2^2$.
    3. Minimal & Full-Rank: e.g., no extra constraint.
2. Multinomial distribution.

More Complicated Examples

1. Zero-Inflated Poisson: $\{X_i : i = 1, \ldots, n\}$, iid with common pmf

$$f(x; \lambda, p) = p\mathbf{1}_{\{x=0\}} + (1-p)\frac{\lambda^x e^{-\lambda}}{x!}, \qquad x = 0, 1, \ldots.$$

   Full rank with suff stats: $\left(\sum_{i=1}^n \mathbf{1}_{\{X_i \neq 0\}}, \sum_{i=1}^n X_i \mathbf{1}_{\{X_i \neq 0\}}\right)$.

2. Censored Exponential: Hwk1 Q3.

More Complicated Examples

1. Zero-Inflated Poisson: $\{X_i : i = 1, \ldots, n\}$, iid with common pmf

$$f(x; \lambda, p) = p\mathbf{1}_{\{x=0\}} + (1-p)\frac{\lambda^x e^{-\lambda}}{x!}, \qquad x = 0, 1, \ldots.$$

   Full rank with suff stats: $\left(\sum_{i=1}^n \mathbf{1}_{\{X_i \neq 0\}}, \sum_{i=1}^n X_i \mathbf{1}_{\{X_i \neq 0\}}\right)$.

2. Censored Exponential: Hwk1 Q3.

Sufficient Statistics: Remarks

Universal Cases. $X_1, \cdots, X_n$ are iid with density f.

- The original data $X_1, \cdots, X_n$ are always sufficient for θ (They are trivial statistics, since they do not lead to any data reduction)
- Order statistics $T = (X_{(1)}, \cdots, X_{(n)})$ are always sufficient for θ (Using factorization theorem, the sample density $\prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} f(x_{(i)})$ (The dimension of order statistics is n, the same as the dimension of the data. Still this is a nontrivial reduction as n! different values of data corresponds to one value of T.)