Statistical Inference

Topic 2. Fundamentals of Statistics

Fall 2019

Instructor: Dr. Jiangdian Wang

Outline

Sampling Distribution

Exponential Family

Sufficient Statistics and Complete Statistics

Today's Topic

- Sufficient statistic
  - Factorization theorem

- Complete statistic
  - Complete statistic in exponential family
  - Property of complete statistic

Sufficient Statistics

- Let $\boldsymbol{X} = (X_1, X_2, \dots, X_n)$ be a random sample from a probability distribution with unknown parameter θ, $f(x;\theta)$.

- A statistic $T(\boldsymbol{X})$ is a function of the sample (data) that does not depend on unknown parameters

- A statistic is to simplify (summarize) the information about θ in the sample
  - Higher degree of simplification
  - Less information loss

- If $T(\boldsymbol{X})$ extract all the information in the sample, i.e., there is no information loss, then it is said to be sufficient. We can use $T(\boldsymbol{X})$ instead of the n sample point to infer θ without information loss

- Proposed by R. A. Fisher in 1922. (Fisher, R. A. (1922).
  - On the Mathematical Foundations of Theoretical Statistics. Philos. Trans. R. Soc. London, Ser. A. 222A: 309–368.)

Example 1. Let X be a Bernoulli population, i.e., $X \sim B(1, \theta)$, and let $\boldsymbol{X} = (X_1, \cdots, X_n)$ be a random sample from X.

Goal: inference on θ

$$P(X_i = 1) = \theta, \qquad P(X_i = 0) = 1 - \theta, \qquad 0 < \theta < 1$$

Let $T(\boldsymbol{X}) = \sum_{i=1}^{n} X_i \sim B(n, \theta)$

Intuitively: The number of success contains all the information about θ, while no information in the order, so $T(\boldsymbol{X})$ capture all the information

Mathematically: Let $t = \sum_{i=1}^{n} x_i$

$$f_{\boldsymbol{X}}(\boldsymbol{x}; \theta) = P(X_1 = x_1, \cdots, X_n = x_n) = \theta^t (1-\theta)^{n-t},$$
$$f_T(t; \theta) = C_n^t \theta^t (1-\theta)^{n-t}, \qquad t = 0,1,\cdots,n, 0 < \theta < 1$$

- **Intuitively**

information (about θ) in $\boldsymbol{X}$ = information in $T(\boldsymbol{X})$ + remaining information in $\boldsymbol{X}$ after $T(\boldsymbol{X})$ is known

- **Definition 1.** A statistic $T(\boldsymbol{X})$ is said to be sufficient for θ if the conditional distribution of $\boldsymbol{X} = (X_1, \cdots, X_n)$ given T, does not depends on θ. That is,

$$g(\boldsymbol{x} \mid t, \theta) = g(\boldsymbol{x} \mid t)$$

Remark for Sufficient Statistic

- The definition says that a sufficient statistic T contains all the information there is in the sample about θ

- If T is a sufficient statistic of θ, and g is a real-valued one-to-one measurable function, then $T^* = g(T)$ is also a sufficient statistic since knowing/given T is equivalent to knowing/given T^*

- If T is a sufficient statistic of θ, and T_1 is another statistic, then (T, T_1) is sufficient for θ

- θ can be k-dimensional vector. Then we need a multidimensional sufficient statistic, usually, with dimensionality equal to the number of parameters.

Revisit Example 1.

$\boldsymbol{X} = (X_1, \cdots, X_n)$ be a random sample from $B(1, \theta)$, then $T(\boldsymbol{X}) = \sum_{i=1}^{n} X_i$ is a sufficient statistic for θ.

Discussion.

Example 2. $\boldsymbol{X} = (X_1, \cdots, X_n)$ be a random sample from $N(\mu, \sigma^2)$ where σ^2 is known. Then $T(\boldsymbol{X}) = \bar{X} = (\sum_{i=1}^{n} X_i)/n$ is a sufficient statistic for μ.

Discussion.

In **Example 2,** $T(\boldsymbol{X}) = X_1$ is **not** a sufficient statistic。

$T(\boldsymbol{X})$ only uses one observation in the sample and could not contain any information about $(X_2, \cdots, X_n)$, so it is not a sufficient statistic.

$$f(x_1, \cdots, x_n \mid x_1) = \frac{f(x_1, \cdots, x_n)}{f_T(x_1)}$$

$$= f(x_2, \cdots, x_n) = (2\pi)^{-\frac{n-1}{2}} \exp\left\{-\frac{1}{2}\sum_{i=2}^{n}(x_i - \mu)^2\right\}$$

Factorization Theorem (因子分解定理)

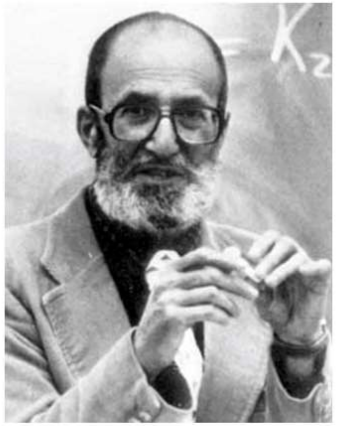
Theorem 6 (Factorization Theorem)**.** Let $f(\boldsymbol{x};\theta)$ be the p.d.f. of a random sample $\boldsymbol{X} = (X_1, \cdots, X_n)$ and $T(\boldsymbol{X})$ be a statistic, then $T(\boldsymbol{X})$ is a sufficient statistic for θ if and only if there exit two non-negative functions g and h such that the p.d.f. $f(\boldsymbol{x};\theta)$ can be written

$$f(\boldsymbol{x};\theta) = g(t(\boldsymbol{x});\theta)h(\boldsymbol{x}),$$

where $t(\boldsymbol{x})$ is the observed value of $T(\boldsymbol{X})$, g depends only on the sample through T, and h does not depends on θ.

- Proposed by R. A. Fisher in 1920s, its most general form and rigorous mathematical proof can be found in Halmos and Savage (1949)
  - Halmos, P. R., and Savage, L. J. (1949). Applications of the Radon-Nikodym Theorem to the Theory of Sufficient Statistics. Ann. *Math. Statist*. **20**, 225-241.

R. A. Fisher



P. R. Halmos

L. J. Savage

Examples

Example 3. Let $\boldsymbol{X} = (X_1, \cdots, X_n)$ be a random sample from Poisson distribution with parameter λ>0, find a sufficient statistic for λ.

Example 4. Let X_1, X_2, ..., X_n be a random sample from a normal distribution with mean μ and variance 1. Prove $t = (\sum_{i=1}^{n} x_i)/n$ is the sufficient statistic for the parameter μ.

Example 5. Let $\boldsymbol{X} = (X_1, \cdots, X_n)$ be a random sample from uniform distribution, $U(0, \theta)$, then $T(\boldsymbol{X}) = X_{(n)}$ is a sufficient statistic for θ.

Discussion.

Sufficiency in an Exponential Family

Example 6. Let $\boldsymbol{X} = (X_1, \cdots, X_n)$ be a random sample from an exponential family, then $T(\boldsymbol{X}) = (T_1(\boldsymbol{X}), \cdots, T_k(\boldsymbol{X}))$ is sufficient for θ.

$$f(\boldsymbol{x};\theta) = C(\theta)\exp\left\{\sum_{i=1}^{k} Q_i(\theta)t_i(\boldsymbol{x})\right\} h(\boldsymbol{x}) = g(t(\boldsymbol{x});\theta)h(\boldsymbol{x})$$

Proof.

Minimal Sufficient Statistic (极小充分统计量)

- Sufficient statistic is *not unique*
  - In example 2, $T(\boldsymbol{X}) = \bar{X} = (\sum_{i=1}^{n} X_i)/n$ is a sufficient statistic, so is the $T_1(\boldsymbol{X}) = (\sum_{i=1}^{m} X_i\,,\ \sum_{i=m+1}^{n} X_i)$
  - $T(\boldsymbol{X})$ is a function of $T_1(\boldsymbol{X})$ and is simpler than $T_1(\boldsymbol{X})$

- **Definition 2.** A statistic is *minimal sufficient* if it can be expressed as a function of every other sufficient statistic.

Complete Statistic (完全/完备统计量)

Definition 3 . Let $\mathcal{F} = \{f(x;\theta),\ \theta \in \Theta\}$ be a distribution family where Θ is the parametric space. Let $T = T(X)$ be a statistic. If for any function φ satisfying

$$E_\theta \varphi(T(X)) = 0, \quad \forall\, \theta \in \Theta,$$

it holds that

$$P_\theta(\varphi(T(X)) = 0) = 1,, \quad \forall\, \theta \in \Theta,$$

then $T(X)$ is said to be a complete statistic for θ.

- Remark. For any measurable function δ, $\delta(T)$ is also complete.

- Remark. Suppose T has **density $\boldsymbol{g(t;\theta)}$**, in the definition

$$\int \varphi(t)g(t;\theta)dt = 0, \qquad \forall\,\theta \in \Theta$$

- φ is orthogonal to $\{g(t;\theta), \theta \in \Theta\}$ implies $\varphi = 0\ a.s.$

- The function class $\{g(t;\theta), \theta \in \Theta\}$ is complete – equivalent to say T is complete

Revisit Example 1. Let $\boldsymbol{X} = (X_1, \cdots, X_n)$ be a random sample from $B(1, \theta)$, then $T(\boldsymbol{X}) = \sum_{i=1}^{n} X_i$ is a complete statistic for θ.

Discussion.

Revisit Example 2. Let $\boldsymbol{X} = (X_1, \cdots, X_n)$ be a random sample from $N(\theta, 1)$, then $T(\boldsymbol{X}) = \bar{X} = (\sum_{i=1}^{n} X_i)/n$ is a complete statistic for θ.

Discussion.

Revisit Example 3. Let $\boldsymbol{X} = (X_1, \cdots, X_n)$ be a random sample fromuniform distribution, $U(0, \theta)$, then$T(\boldsymbol{X}) = X_{(n)}$is a complete statistic for θ.

Discussion.

Completeness in an Exponential Family

Theorem (Complete statistics in the exponential family) *Let $X_1, \dots, X_n$ be iid observations from an exponential family with pdf or pmf of the form*

$$f(x|\boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta}) \exp\left(\sum_{j=1}^{k} w(\theta_j)t_j(x)\right), \tag{6.2.7}$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_k)$. Then the statistic

$$T(\mathbf{X}) = \left(\sum_{i=1}^{n} t_1(X_i), \sum_{i=1}^{n} t_2(X_i), \dots, \sum_{i=1}^{n} t_k(X_i)\right)$$

is complete as long as the parameter space Θ contains an open set in $\Re^k$.

Proof: 陈希孺. 数理统计引论. 北京: 科学出版社, 1981, 1998. (Theorem 1.6.1 on Page 80)

Example 7. Let $\boldsymbol{X} = (X_1, \cdots, X_n)$ be a random sample from $N(\mu, \sigma^2)$, and the parameter space $\Theta = \{\theta = (\mu, \sigma^2): -\infty < \mu < \infty, \sigma^2 > 0\}$, then $(\bar{X}, S^2)$ is complete statistic for θ.

Discussion.

Example 8 (not complete). Let $\boldsymbol{X} = (X_1, \cdots, X_n)$ be a random sample from uniform distribution, $U(\theta - 1/2, \theta + 1/2)$, then $T(\boldsymbol{X}) = (X_{(1)}, X_{(n)})$ is sufficient statistic but not complete statistic for θ.
(Hint: Not complete. Find a function φ such that $E_\theta \varphi(T) = 0$ but $P_\theta(\varphi(T) = 0) \neq 1$.)

Discussion.

Theorem *If a minimal sufficient statistic exists, then any complete statistic is also a minimal sufficient statistic.*

Ancillary Statistic

Definition 4 A statistic $S(\mathbf{X})$ whose distribution does not depend on the parameter θ is called an *ancillary statistic.*

Example. (Uniform ancillary statistic) let $X_1, X_2, \ldots, X_n$ be iid uniform observations on the interval $(\theta, \theta+1)$, $-\infty < \theta < +\infty$. Let $X_{(1)} < \ldots < X_{(n)}$ be the order statistics from the sample. The range statistic, R = $X_{(n)} - X_{(1)}$, is an ancillary statistic.

Basu Theorem

Theorem (Basu's Theorem) *If $T(\mathbf{X})$ is a complete and minimal sufficient statistic, then $T(\mathbf{X})$ is independent of every ancillary statistic.*

Revisit Example 2. Let $\boldsymbol{X} = (X_1, \cdots, X_n)$ be a random sample from $N(\theta, 1)$. Let $R(\boldsymbol{X}) = X_{(n)} - X_{(1)}$ and $T(\boldsymbol{X}) = \bar{X} = (\sum_{i=1}^{n} X_i)/n$, then $R(\boldsymbol{X})$ and $T(\boldsymbol{X})$ are independent.

Discussion.