MTH211A: Theory of Statistics

Module 2: Principles of Data Reduction

February 16, 2024

Recall the (parametric) inference problem: Let $X_1, \cdots, X_n$ be a random sample from some distribution F, which is parameterized by some parameter vector $\boldsymbol{\theta}$, $\boldsymbol{\theta} \in \Theta$, where Θ is the parameter space. Our goal is to infer about F (which is equivalent to infer about $\boldsymbol{\theta}$, or some function of $\boldsymbol{\theta}$) using the random sample $X_1, \cdots, X_n$. Usually a statistician summarizes the data using some statistics (functions of the samples), for example, mean, SD, mode, maximum, minimum, etc.

Recall that, the sample space, say $\mathcal{X}$ is a subset of $\mathbb{R}^n$, and a statistic $T(\{X_1, \cdots, X_n\}) = T(\mathbf{X})$ is a function from $\mathcal{X} \to R$. Any statistics $T(\mathbf{X})$ defines a form of data reduction, in the sense that the possible values of $T(\mathbf{X})$ induces a partition in the sample space. Suppose the statistic $T(\mathbf{X})$ has the realization t, then there exists a collection of sample realizations $A_t = \{\mathbf{x} = (x_1, \ldots, x_n)' : T(\mathbf{x}) = t\}$ which leads to the functional value t. Now, let $\mathcal{T} = \{t = T(\mathbf{x}) : \ \mathbf{x} \in \mathcal{X}\}$ be the range of the function T with domain $\mathcal{X}$. Then it is not difficult to see that $\cup\{A_t : t \in \mathcal{T}\} = \mathcal{X}$, and if $t \neq t'$, then $A_t \cap A_{t'} = \phi$. Therefore, the collection of sets $\{A_t : t \in \mathcal{T}\}$ defines a partition on $\mathcal{X}$, and it is called the partition induced by the statistic T. Thus the reduction due to the statistic T is equivalent to the partition.

Example. Let X, Y be a random sample from $\texttt{unifrom}(0, 1)$. Consider two statistics $T_1 = \max\{X, Y\}$ and $T_2 = \mathbb{I}(X > Y)$ where $\mathbb{I}$ is the indicator function. It is easy to see that $A_{1,t} = \{(x, y) : \max\{x, y\} = t\}$ is the collection of all points in the two line segments $\{x = t, 0 < y \leq t\}$ and $\{y = t, 0 < x \leq t\}$, whereas $A_{2,0} = \{(x, y) : x \leq y\}$ and $A_{2,1} = \{(x, y) : x > y\}$. Thus, T_2 induces higher level of data reduction.

Note that, higher level of reduction may lead to over-summarizing of the data resulting in loss of important information about the population. On the other hand, no reduction or very low level of reduction may lead to storing unimportant information. The goal of a statistician is to employ highest level of reduction as long as no *important* information is lost. What information is important? In parametric inference, all information relevant to the parameter $\boldsymbol{\theta}$ is important.

1 Sufficient Statistic

Definition 1 (Sufficient Statistic)**.** *Let $\mathbf{X} = \{X_1, \cdots, X_n\}$ be a random sample from the distribution $\{F_\theta; \theta \in \Theta\}$. A statistic $T(\mathbf{X})$ is a sufficient statistic for θ, if the conditional distribution of the random sample $\mathbf{X}$ given the value of $T(\mathbf{X})$ does not depend on θ.*

Understanding the abstract definition. First, consider an example. Let $X_1, \cdots, X_n$ be a random sample from $\texttt{Uniform}(0, \theta)$. Suppose a statistician (S1) observes a realization of $X_1, \cdots, X_n$ as $\{x_1, \ldots, x_n\}$ with $\max\{x_1, \ldots, x_n\} = t$. Now if s/he conveys only the value t to a second statistician (S2) as the largest order statistic, then will S2 have less information about θ compared to S1? The answer is 'no', because the highest order statistics, $X_{(n)}$, is sufficient for θ. All the realizations with $X_{(n)} = t$ has same amount of information about θ.

(One may skip the following in the first reading) To see in more detail, let $\mathbf{X}$ be a discrete (or continuous) random variable with pmf (or, pdf) $f_{\boldsymbol{\theta}}$. Then the conditional distribution of $\mathbf{X}$ given $T(\mathbf{X}) = t$ is

$$f_{\mathbf{X}|T(\mathbf{X})=t}(\mathbf{x}) = \frac{f_{\boldsymbol{\theta};\mathbf{X},T}(\mathbf{x}, t)}{f_{\boldsymbol{\theta};T}(t)} = \begin{cases} f_{\boldsymbol{\theta};\mathbf{X}}(\mathbf{x})/f_{\boldsymbol{\theta};T}(t) & \text{if} \quad \mathbf{x} \in A_t, \\ 0 & \text{otherwise.} \end{cases}$$

By definition of sufficient statistic, $f_{\mathbf{X}|T(\mathbf{X})=t}(\mathbf{x})$ is free of $\boldsymbol{\theta}$, and hence is completely known. Thus, it is (theoretically) possible to simulate from this distribution. Suppose the random variable $\mathbf{Y} \mid T(\mathbf{X}) = t$ is distributed as this conditional distribution. Then the unconditional distribution of $\mathbf{Y}$ is same as unconditional distribution of $\mathbf{X}$ regardless of the value of $\boldsymbol{\theta}$, i.e., for any (measurable) subset $A \subseteq \mathcal{X}$, $P_{\boldsymbol{\theta}}(\mathbf{X} \in A) = P_{\boldsymbol{\theta}}(\mathbf{Y} \in A)$, regardless of the value of $\boldsymbol{\theta}$, as

$$P_{\boldsymbol{\theta}}(\mathbf{Y} \in A) = \int_A \left\{ \int_t f_{\boldsymbol{\theta};\mathbf{X}|T=t}(\mathbf{x}) f_T(t) dt \right\} d\mathbf{x} = \int_A f_{\boldsymbol{\theta};\mathbf{X}}(\mathbf{x}) d\mathbf{x} = P_{\boldsymbol{\theta}}(\mathbf{X} \in A).$$

Thus. $\mathbf{Y} \stackrel{D}{=} \mathbf{X}$.

This implies that S2, without knowing the distribution of $\mathbf{X}$, is able to generate realizations from the distribution of $\mathbf{X}$, regardless of the value of $\boldsymbol{\theta}$.

Example 1. Let $X_1, \cdots, X_n$ be a random sample from `Bernoulli`(θ) distribution. Then the statistic $T(\mathbf{X}) = \sum_{i=1}^n X_i$ is sufficient.

Example 2. Let $X_1, \ldots, X_n$ be a random sample from `Gamma`$(2, \theta)$ distribution. Then the statistic $T(\mathbf{X}) = \sum_{i=1}^n X_i$ is sufficient.

Example 3. Let $X_1, \ldots, X_n$ be a random sample from `Normal`$(\mu, 1)$ distribution. Then the statistic $T(\mathbf{X}) = \sum_{i=1}^n X_i$ is sufficient.

[Alternative proof using orthogonal transformation]

Example 4. Let $X_1, \cdots, X_n$ be a random sample from some absolutely continuous distribution with parameter vector $\boldsymbol{\theta}$. Then $\{X_{(1)}, \cdots, X_{(n)}\}$ is jointly sufficient for $\boldsymbol{\theta}$. The complete sample $\{X_1, \cdots, X_n\}$ is also sufficient for $\boldsymbol{\theta}$.

Theorem 1. *Let $X_1, \cdots, X_n$ denote a random sample from a discrete or absolutely continuous distribution that has a joint pmf or joint pdf $f_{\mathbf{X}}(\cdot; \boldsymbol{\theta})$, $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. The statistic $T = T(\mathbf{X})$ is sufficient for θ if and only if (iff) there exists functions $g(t; \boldsymbol{\theta})$ and $h(\mathbf{x})$ such that, for all sample points $\mathbf{x}$ and all parameter values $\boldsymbol{\theta} \in \Theta$*

$$f_{\mathbf{X}}(\mathbf{x}; \boldsymbol{\theta}) = g\left(T(\mathbf{x}); \boldsymbol{\theta}\right) h(\mathbf{x}).$$

[Proof]

Remark 1. 1. *Let $X_1, \cdots, X_n$ be a random sample from some distribution with parameter vector $\boldsymbol{\theta}$, and $T(\mathbf{X})$ be a sufficient statistics of $\boldsymbol{\theta}$. If $T(\mathbf{X})$ is a function of another statistics $U(\mathbf{X})$, then $U(\mathbf{X})$ is also sufficient for $\boldsymbol{\theta}$. However, the converse is not true, in general.*

[Proof/counter example]

2. *Let $X_1, \cdots, X_n$ be a random sample from some distribution with parameter vector $\boldsymbol{\theta}$, and $T(\mathbf{X})$ be a sufficient statistics of $\boldsymbol{\theta}$. If $U(\mathbf{X})$ is a bijective function of $T(\mathbf{X})$. Then $U(\mathbf{X})$ is also sufficient for $\boldsymbol{\theta}$. [Homework]*

Example 5. Let $X_1, \ldots, X_n$ be a random sample from `Uniform`$(0, \theta)$ distribution. Then the statistic $T(\mathbf{X}) = X_{(n)}$ is sufficient for θ.

Example 6. Let $X_1, \ldots, X_n$ be a random sample from discrete uniform distribution with equal probability mass on each point of $\{1, 2, \ldots, \theta\}$ distribution. Then the statistic $T(\mathbf{X}) = X_{(n)}$ is sufficient for θ.

Example 7. Let $X_1, \ldots, X_n$ be a random sample from `Normal`(μ, σ^2) distribution. Then the statistic $\mathbf{T}(\mathbf{X}) = \left\{\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2\right\}$ is jointly sufficient for $\boldsymbol{\theta} = (\mu, \sigma^2)$.

Exponential family. A family of pmfs of pdfs is called a d-parameter exponential family if it can be expressed as

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

where $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_d)$. Here $h(x) \geq 0$ for all x, and $t_1, \ldots, t_k$ are real valued functions of x, not depending on $\boldsymbol{\theta}$. Further, $c(\boldsymbol{\theta}), w_1(\boldsymbol{\theta}), \ldots, w_k(\boldsymbol{\theta})$ are real-valued functions of $\boldsymbol{\theta}$, not depending on x.

Many common distributions belong to the the exponential family. Examples include (i) `binomial`(n, p) with n known, (ii) `Poisson`(λ), (iii) `normal`(μ, σ^2), (iv) `exponential`(λ), (v) `Beta`(α, β), (vi) `Gamma`(α, β), etc.

Theorem 2 (Sufficient statistics for exponential family of distributions.)**.** *Let $X_1, \ldots, X_n$ be a random sample from a distribution with pmf or pdf $f_X(\cdot;\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^d$ which belongs to an exponential family given by (1) with $d \leq k$. Then*

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

is jointly sufficient for $\boldsymbol{\theta}$.

2 Minimal Sufficient Statistics

For any family of distributions $\{F_\theta : \theta \in \Theta\}$, a sufficient statistic for θ captures the maximum information about θ from the sample, but often fails to have highest level of data reduction. For example, in the `Uniform`$(0, \theta)$ class of distributions, both $\mathbf{T}_1(\mathbf{X}) = \{X_{(1)}, \ldots, X_{(n)}\}$ and $T_2(\mathbf{X}) = X_{(n)}$ are sufficient for θ. However, $T_2(\mathbf{X})$ has better level of data reduction compared to $\mathbf{T}_1(\mathbf{X})$. Naturally, one would like to find the sufficient statistic which is most precise, among the pool of sufficient statistics. This leads to the concept of *minimal sufficiency.*

Definition 2 (Minimal Sufficient Statistic)**.** *A sufficient statistic $T(\mathbf{X})$ is called minimal sufficient if, for any other sufficient statistic $U(\mathbf{X})$, $T(\mathbf{x})$ is a function of $U(\mathbf{x})$.*

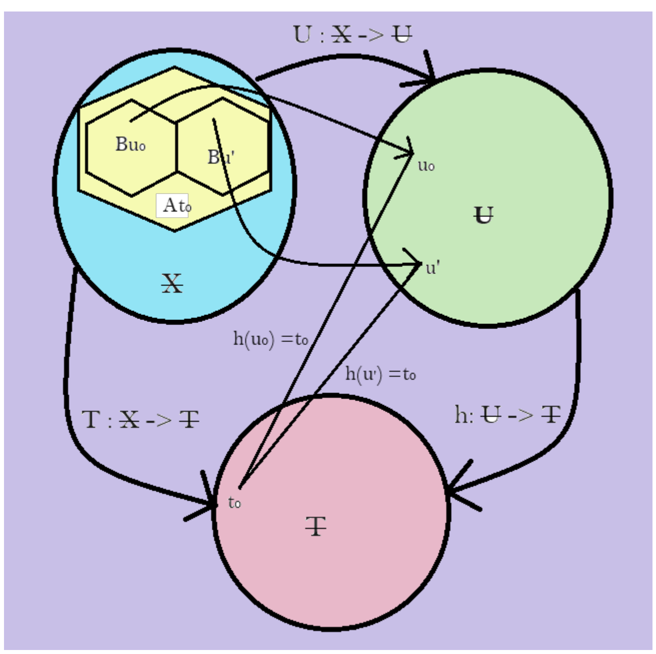

Figure 1: Understanding minimal sufficiency

Understanding minimal sufficiency. Suppose $T(\mathbf{X})$ and $U(\mathbf{X})$ are two different sufficient statistics for the class of distributions F_θ, $\theta \in \Theta$. Let $\mathcal{T} = \{t = T(\mathbf{z}) : \mathbf{z} \in \mathcal{X}\}$, and $\mathcal{U} = \{u = U(\mathbf{z}) : \mathbf{z} \in \mathcal{X}\}$ be the ranges of U and T, respectively. For any $t \in \mathcal{T}$ and $u \in \mathcal{U}$, consider the pre-images of t and u as $A_t = \{\mathbf{z} : T(\mathbf{z}) = t\} \subseteq \mathcal{X}$, and $B_u = \{\mathbf{z} : U(\mathbf{z}) = u\} \subseteq \mathcal{X}$.

Note that,both $\{A_t : t \in \mathcal{T}\}$ and $\{B_u : u \in \mathcal{U}\}$ are partitions of $\mathcal{X}$. As both T and U as sufficient, each point $\mathbf{z} \in A_t$, has same information about θ (as the conditional distribution of $f_{\mathbf{X}|T=t}$ is free of θ). Similarly, each $\mathbf{z} \in B_u$ has same information about θ.

Now, suppose $T(\mathbf{X})$ be a function of $U(\mathbf{X})$, i.e., there exists a function $h : \mathcal{U} \to \mathcal{T}$ such that for each realization $\mathbf{z} \in \mathcal{X}$ (sample space), $T(\mathbf{z}) = t = h(u) = h(U(\mathbf{z}))$ where $U(\mathbf{z}) = u$. Define $C_t = \{u : h(u) = t\} \subseteq \mathcal{U}$ as the pre-image of t.

Consider a $\mathbf{z} \in B_{u_0}$, then $U(\mathbf{z}) = u_0$, and $T(\mathbf{z}) = h(U(\mathbf{z})) = h(u_0) = t_0$, which implies $\mathbf{z} \in A_{t_0}$, where $h(u_0) = t_0$. Thus, $B_{u_0} \subseteq A_{t_0}$, where $h(u_0) = t_0$. Now, suppose there exists another u' such that $h(u') = t_0$. Then, similarly we have $B_{u'} \subseteq A_{t_0}$. So, more generally, $B_u \subseteq A_{t_0}$, for all $u \in C_{t_0}$.

As all $\mathbf{z} \in A_{t_0}$ contains same information about θ, B_{u_0} has same information about θ as A_{t_0}. However, A_{t_0}, being a larger subset of $\mathcal{X}$, does a better level of data reduction.

Thus, a minimal sufficient statistic T, being functions of all other sufficient statistics, provides the highest level of data reduction among the class of sufficient statistics.

How do we know if a sufficient statistic is minimal sufficient?

Theorem 3. *Let $f(\mathbf{x};\theta)$ be the pmf or pdf of a sample $\mathbf{X}$. Suppose there exists a function $T(\mathbf{x})$ such that, for every two points $\mathbf{x}$ and $\mathbf{y}$, the ratio $f(\mathbf{x};\theta)/f(\mathbf{y};\theta)$ is constant as a function of θ iff $T(\mathbf{x}) = T(\mathbf{y})$. Then $T(\mathbf{X})$ is a minimal sufficient statistic for θ.*

[Proof]

Remark 2.

1. *Let $T(\mathbf{X})$ be a minimal sufficient statistic and $U(\mathbf{X})$ is a bijective function of $T(\mathbf{X})$. Then $U(\mathbf{X})$ is also minimal sufficient. [Proof]*
2. *Let both $T(\mathbf{X})$ and $U(\mathbf{X})$ be minimal sufficient statistics. Then $U(\mathbf{X})$ is a bijective function of $T(\mathbf{X})$. [Proof]*
3. *Minimal sufficient statistic is not unique.*

Example 8. Let $X_1, \ldots, X_n$ be a random sample from `Normal`(μ, σ^2) distribution. Then the statistic $(\bar{X}, S^2)$ is minimal sufficient, where $S^2 = \text{var}(X)$.

Example 9. Let $X_1, \ldots, X_n$ be a random sample from `Uniform`$(\theta, \theta + 1)$ distribution. Then the statistic $T(\mathbf{X}) = (X_{(1)}, X_{(n)})$ is minimal sufficient.

Example 10. Let $X_1, \ldots, X_n$ be a random sample from `binomial`(m, p) distribution, where m is known. Then the statistic $T(\mathbf{X}) = \sum_{i=1}^n X_i$ is minimal sufficient.

Ancillary Statistic. *Ancillary statistic* plays a complementary role compared to that of a sufficient statistic. That is, while a sufficient statistic contains all the information about θ, which could be obtained from the sample $\{X_1, \ldots, X_n\}$, an ancillary statistic contains no information about θ.

Definition 3 (Ancillary Statistic)**.** *A statistic $S(\mathbf{X})$ whose distribution does not depend on the parameter θ is called an ancillary statistic.*

Example 11. Let $X_1, \ldots, X_n$ be a random sample from `normal`$(\mu, 1)$ distribution. Then the statistic S^2 is ancillary for μ.

Example 12. Let $X_1, \ldots, X_n$ be a random sample from `uniform`$(\theta, \theta + 1)$ distribution. Then the statistic $S(\mathbf{X}) = (X_{(n)} - X_{(1)})$ is ancillary for θ.

Some Special Families of Distributions.

1. **Location family of distributions:** Let $X_1, \dots, X_n$ be a random sample, where

$$X_i = \theta + W_i, \quad i = 1, \dots, n,$$

where W_is are i.i.d. from some distribution with CDF F (does not depend on θ). This type of distribution of X is called location family of distributions, and θ is called the location parameter. Let $S(\mathbf{X})$ be a statistic such that

$$S(\mathbf{x} + d\mathbf{1}) = S(x_1 + d, \dots, x_n + d) = S(\mathbf{x}),$$

for all real d. Then $S(\mathbf{X})$ is ancillary for θ, and is called *location-invariant statistic.*

Example: Range, mean deviation about mean, standard deviation are location invariant statistics.

2. **Scale family of distributions:** Let $X_1, \dots, X_n$ be a random sample, where

$$X_i = \theta W_i, \quad i = 1, \dots, n,$$

where W_is are i.i.d. from some distribution with CDF F (does not depend on θ). This type of distribution of X is called scale family of distributions, and θ is called the scale parameter. Let $S(\mathbf{X})$ be a statistic such that

$$S(c\mathbf{x}) = S(cx_1, \dots, cx_n) = S(\mathbf{x}),$$

for all $c > 0$. Then $S(\mathbf{X})$ is ancillary for θ, and is called *scale-invariant statistic.*

Example: The statistics $X_1^2 / \sum_{i=1}^n X_i^2$, $\min_i X_i / \max_i X_i$ are scale invariant statistics.

Example 13. Let $X_1, \cdots, X_n$ be a random sample from `uniform`$(0, \theta)$. Then $X_{(n)}/X_{(1)}$ is an ancillary statistic for θ.

Example 14. Let $X_1, \cdots, X_n$ be a random sample from `unifrom`$(\theta - 1, \theta + 1)$. Then $X_1 - \bar{X}_n$ is an ancillary statistic.

Complete family of distributions. So far we came across the concepts of minimal sufficient and ancillary statistics. Intuitively, it seems interesting to know if these two statistics are unrelated (independent). In fact, for two statistics $S(\mathbf{X})$ and $T(\mathbf{X})$ are independently distributed, and $T(\mathbf{X})$ is sufficient for θ, then $S(\mathbf{X})$ must be ancillary for θ. (Why?)

Remark 3. *The converse of the above statement is not true in general, i.e., if $S(\mathbf{X})$ is sufficient for θ and $T(\mathbf{X})$ is ancillary for θ, then it is not necessarily true that $S(\mathbf{X})$ and $T(\mathbf{X})$ are independently distributed. For example, let X_1, X_2 be a random sample from* `normal`$(\theta, 1)$ *distribution. Then $\mathbf{S}(\mathbf{X}) = (X_1, X_2)'$ is jointly sufficient for θ and $T(\mathbf{X}) = X_1 - X_2$ is ancillary for θ. Now, observe that $\mathbf{S}$ and T are not independent. [To see this, you may verify that $P(T > 0 \mid \mathbf{S} = s) \neq P(T > 0)$.]*

Remark 4. *If fact,if $S(\mathbf{X})$ is minimal sufficient for θ and $T(\mathbf{X})$ is ancillary for θ, then it is not necessarily true that $S(\mathbf{X})$ and $T(\mathbf{X})$ are independently distributed. For example, let $X_1, \cdots, X_n$ be a random sample from* `uniform`$(\theta, \theta + 1)$. *We have previously seen that $\mathbf{S}(\mathbf{X}) = [X_{(1)}, X_{(n)}]'$ is minimal sufficient for θ. Being a bijective function, $\mathbf{S}^\star(\mathbf{X}) = [X_{(1)} + X_{(n)}, X_{(n)} - X_{(1)}]'$ is also minimal sufficient. But $T(\mathbf{X}) = X_{(n)} - X_{(1)}$ is ancillary for θ. Thus, here the ancillary statistic is a function of minimal sufficient statistic.*

In all the above cases, observe that, there exists a non-zero function of the sufficient (or, minimal sufficient) statistics, the expectation of which is a constant c (free of θ). For example in Remark 3, $T(\mathbf{X}) = X_1 - X_2$ satisfies $E_\theta(T(\mathbf{X})) = 0$ but $P_\theta(T(\mathbf{X}) = 0) = 0 \neq 1$. In Remarks 4, if we take the function $g : \mathbb{R}^2 \to \mathbb{R}$ as $g((x, y)) = y$, then $E_\theta(g(\mathbf{S}(\mathbf{X}))) = E\left[X_{(n)} - X_{(1)}\right] = c$ (free of θ), but $P_\theta\left(X_{(n)} - X_{(1)} = c\right) = 0 \neq 1$. It turns out that, if one rules out the possibility of having zero expectation of a non-zero function of the sufficient statistics $T = T(\mathbf{X})$, then T becomes independent of all ancillary statistics of θ. This special property, which ensures independence of minimal sufficient and ancillary statistics, is called *completeness.*

Definition 4 (Complete Family of Distributions)**.** *Let $T(\mathbf{X})$ be a statistic with pdf or pmf $f_T(\cdot;\theta)$. The family of distribution $f_T(\cdot;\theta), \theta \in \Theta$ is called complete if $E_\theta(g(T)) = 0$ for all $\theta \in \Theta$ implies $P_\theta(g(T) = 0) = 1$ for all $\theta \in \Theta$. If the family of a statistic $T(\mathbf{X})$ is complete, then $T(\mathbf{X})$ is called a complete statistic.*

Remark 5. *$E_\theta(g(T)) = 0$ for all $\theta \in \Theta$ does not in general imply that $P_\theta(g(T) = 0) = 1$ for all $\theta \in \Theta$. For example, let $\{X_1, X_2\}$ be a random sample from $N(\theta, 1)$, $T = X_1 - X_2$ and $g(T) = T$. Then for any θ, $E_\theta(g(T)) = 0$. However, $P(g(T) = 0) = P(X_1 = X_2) = 0$.*

Remark 6. *Completeness is a property of the family of distribution of a statistic $T(\mathbf{X})$. For example, the $N(\theta, 1), \theta \in \mathbb{R}$, family is complete. (Proof after midsem)*

Example 15. The `Poisson`$(\theta), \theta > 0$ family is complete. (Proof)

Example 16. The `exponential`$(\theta), \theta > 0$ family is complete. (Homework after midsem)

Example 17. The `binomial`$(n, \theta), 0 < \theta < 1$ family is complete. (Homework)

Thus, $T(\mathbf{X}) = \sum_i X_i$ is a complete sufficient statistic of θ, given $\{X_1, \cdots, X_n\}$ is a random sample from `Bernoulli`(θ).

Example 18. $\{X_1, \cdots, X_n\}$ be a random sample from `uniform`$(0, \theta)$, $\theta > 0$. Then the family of distributions of $X_{(n)}$ is complete. (Proof)

Example 19. The distribution `normal` $\left(0, \sigma^2\right)$ is not complete. However, if $X \sim$ `normal` $\left(0, \sigma^2\right)$, the the distribution of $T(X) = X^2$ is complete.

Thus, if $X_1, \cdots, X_n$ is a random sample `normal` $\left(0, \sigma^2\right)$, then $T_1(\mathbf{X}) = \sum_{i=1}^n X_i$ is not complete, but $T_2(\mathbf{X}) = \sum_{i=1}^n X_i^2$ is complete.

Theorem 4 (Basu's theorem)**.** *Let $\{X_1, \cdots, X_n\}$ be a sample from the family of distributions with pmf or pdf $f_{\mathbf{X}}(\cdot;\theta)$, $\theta \in \Theta$. If $T(\mathbf{X})$ is a complete-sufficient statistics for θ, then $T(\mathbf{X})$ is independent of every ancillary statistic of θ. [Proof after midsem]*

Remark 7. *Basu's theorem provides us a way to verify independence of two statistics without explicitly deriving the joint/conditional distributions. Let us see an example.*

Example 20. Let $\{X_1, \cdots, X_n\}$ be a random sample from `normal`$(\mu, 1)$, $\mu \in \mathbb{R}$. Then the sample mean and sample variance, $\bar{X}$ and S^2, are independent.

We conclude this section with a theorem which would be proved later.

Theorem 5. *A complete-sufficient statistic is minimal sufficient.*