2 Statistical Inference Principles

2.1 Statistical Model and Background

Definition 1. A statistical model is a class of possible probability measures $\mathcal{P}$ on the sample space $\mathcal{X}$, parameterized by parameter(s) θ:

$$\mathcal{P} \quad = \quad \{\mathrm{P}_\theta : \ \theta \in \Theta\},$$

where P_θ is the distribution with parameter θ such that $X \sim \mathrm{P}_\theta$, and Θ is the parameter space.
A sample $\boldsymbol{X} = (X_1, ..., X_n)$ is a collection of independent random variables where X_i is distributed according to a distribution $\mathrm{P}_{i,\theta}$. The sample size n is the number of the random variables.
A statistic T is a function of the sample

$$T : \boldsymbol{x} \in \mathcal{X} \quad \rightarrow \quad T(\boldsymbol{x}) = t \in \mathcal{T},$$

where $\mathcal{T}$ is some set. With random variable $\boldsymbol{X}$, the function $T(\boldsymbol{X})$ is also a random variable, the distribution of which is given by

$$\mathrm{P}_\theta^T(B) \quad = \quad \mathrm{P}_\theta(\{\boldsymbol{x} : T(\boldsymbol{x}) \in B\}).$$

An estimator of θ is a statistic $\delta(X)$ that we went to use to estimate unknown parameter θ.

Example 1. If we assume $X \sim \mathrm{Bernoulli}(\theta)$, then $\mathrm{P}_\theta = \mathrm{Bernoulli}(\theta)$, and our statistical model is

$$\mathcal{P} \quad = \quad \{\mathrm{Bernoulli}(\theta) : \ \theta \in (0,1)\}.$$

If we observe a sample X_1, ..., X_n from it, one statistic of interest is $T(x) = \sum_{i=1}^n x_i$ or $T(x) = \bar{X} \equiv n^{-1} \sum_{i=1}^n x_i$.

2.2 Sufficient Statistic

Any statistic $T : x \in \mathcal{X} \rightarrow T(x) = t \in \mathcal{T}$ generates a partition of the sample space

$$\mathcal{X} = \bigcup_t \mathcal{X}_t \quad \text{with} \quad \mathcal{X}_t = \{\boldsymbol{x} : T(\boldsymbol{x}) = t\}.$$

Example 2. Suppose that we toss a coin 2 times independently. For $i = 1, 2$, let $X_i = 1$ if we get a head, and 0 otherwise.

- The sample space is $\mathcal{X} = \{(0,0), (0,1), (1,0), (1,1)\}$.
- Define the statistic $T(X_1, X_2) = X_1 + X_2$.
- The partition is

$$\mathcal{X} \quad = \quad \{(0,0)\} \cup \{(0,1), (1,0)\} \cup \{(1,1)\}.$$

  Taking any partition incurs loss of information.

- For example, we only know we get 1 head, but don't know whether it is the first toss or the second.

But the lost information may be irrelevant for inference.

Definition 2 (Sufficient Statistic)**.** A statistic T is said to be sufficient for the statistical model $\{\mathrm{P}_\theta : \theta \in \Theta\}$ if the conditional distribution of X given $T = t$ is independent of θ for all t. Such T is called a sufficient statistic.

Note: the distribution of X depends on θ, the distribution of T depends on θ, but the distribution of $\boldsymbol{X} \mid T$ does not depend on θ. If T is sufficient for θ, then T captures all information about θ included in X.

Example 3 (Sufficient Statistic)**.** Let X_1, ..., X_n be i.i.d. Bernoulli random variables with parameter $\theta \in (0,1)$. Define $T = \sum_{i=1}^n X_i$. Then,

$$\begin{aligned} P\left(X_1 = x_1, ..., X_n = x_n \mid T = t\right) &= \frac{P\left(X_1 = x_1, ..., X_n = x_n, T = t\right)}{P\left(T = t\right)} = \frac{P\left(X_1 = x_1, ..., X_n = x_n\right)}{P\left(T = t\right)} \\ &= \frac{\prod_{i=1}^n \theta^{x_i}\left(1-\theta\right)^{1-x_i}}{\binom{n}{t}\theta^t\left(1-\theta\right)^{n-t}} = \frac{1}{\binom{n}{t}}, \end{aligned}$$

that does not depend on θ.

Theorem 1 (Factorization theorem)**.** *Let $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ be a family of distributions dominated by μ. A necessary and sufficient condition for a statistic T to be sufficient is that there exist functions $g\left(T\left(x\right), \theta\right) \geq 0$ and $h\left(x\right) \geq 0$ such that the densities satisfy*

$$p\left(x \mid \theta\right) = g\left(T\left(x\right), \theta\right) h\left(x\right), \text{ almost everywhere under } \mu.$$

Example 4. Let X_1, ..., X_n be i.i.d. random variables with parameter θ.

1. Bernoulli random variables with parameter $\theta \in (0,1)$. Then,

$$p\left(x \mid \theta\right) = \theta^{\sum_{i=1}^n x_i}\left(1-\theta\right)^{n-\sum_{i=1}^n x_i}.$$

   Hence, $T\left(x\right) = \sum_{i=1}^n x_i$.

2. $N\left(\mu, \sigma^2\right)$ with $\theta = \left(\mu, \sigma^2\right)$. Then,

$$p\left(x \mid \theta\right) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{\left(x_i-\mu\right)^2}{2\sigma^2}\right\} = \frac{1}{\left(2\pi\sigma^2\right)^{n/2}} \exp\left\{-\frac{n\mu^2}{2\sigma^2}\right\} \exp\left\{\frac{n\mu}{\sigma^2}\bar{x} - \frac{1}{2\sigma^2}\sum_{i=1}^n x_i^2\right\}.$$

   Hence $T\left(x\right) = \left(\bar{x}, \sum_{i=1}^n x_i^2\right)$.

Theorem 2 (Sufficiency in Exponential Family)**.** *Consider exponential family*

$$p\left(x \mid \theta\right) = \exp\left\{\sum_{i=1}^s \eta_i\left(\theta\right) T_i\left(x\right) - A\left(\eta\right)\right\} h\left(x\right).$$

Then, the statistic $T\left(x\right) = \left(T_1\left(x\right), ..., T_s\left(x\right)\right)$ is sufficient.

Proof. The factorization theorem directly implies that $T\left(x\right) = \left(T_1\left(x\right), ..., T_s\left(x\right)\right)$ is sufficient. □

Example 5. Let X_1, ..., X_n be i.i.d. Bernoulli with parameter θ. Then,

$$p\left(x \mid \theta\right) = \exp\left\{n\bar{X}\log\theta + n\left(1-\bar{X}\right)\log\left(1-\theta\right)\right\},$$

belongs to exponential family. Hence, $T\left(x\right) = \left(n\bar{X}, n\left(1-\bar{X}\right)\right)$ is sufficient.

2.3 Minimal Sufficiency

Sufficient statistic is not unique. For the normal example, we can rewrite the density as

$$\begin{aligned} p\left(x \mid \theta\right) &= \frac{1}{\left(2\pi\sigma^2\right)^{n/2}} \exp\left\{-\frac{n\mu^2}{2\sigma^2}\right\} \exp\left\{\frac{n\mu}{\sigma^2}\bar{x} - \frac{1}{2\sigma^2}\left[\sum_{i=1}^n x_i^2 - n\left(\bar{x}\right)^2\right] - \frac{n}{2\sigma^2}\left(\bar{x}\right)^2\right\} \\ &= \frac{1}{\left(2\pi\sigma^2\right)^{n/2}} \exp\left\{-\frac{n\mu^2}{2\sigma^2}\right\} \exp\left\{\frac{n\mu}{\sigma^2}\bar{x} - \frac{n-1}{2\sigma^2}S^2 - \frac{n}{2\sigma^2}\left(\bar{x}\right)^2\right\} \end{aligned}$$

Hence, $T\left(x\right) = \left(\bar{X}, S^2\right)$ is also sufficient. If T is sufficient for a family of distributions $\mathcal{P}$ and $T = f\left(\tilde{T}\right)$, then $\tilde{T}$ is also sufficient for $\mathcal{P}$.

Example 6. Any sufficient statistic is a function of x. Hence, the original data x is sufficient.

Definition 3 (Minimal Sufficiency)**.** A sufficient statistic T is called a minimal sufficient statistic if T is a measurable function of any other sufficient statistic (almost everywhere under $\mathcal{P}$). The set where this definition fails is a null set for every $\mathrm{P}_\theta = \mathcal{P}$.

Theorem 3. *Let $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ be a family of statistical models with density $p(x \mid \theta)$ with respected to a measure μ. Consider two independent samples x and y from P_θ. If $p(x \mid \theta)/p(y \mid \theta)$ does not depend on θ if and only if $T(x) = T(y)$ [we do not know T is sufficient yet]. Then T is sufficient and minimal sufficient.*

Example 7. Let X_1, ..., X_n be i.i.d. random variables with parameter θ.

1. Bernoulli random variables with parameter $\theta \in (0,1)$. Then,

$$p(x \mid \theta) = \theta^{\sum_{i=1}^n x_i} (1-\theta)^{n-\sum_{i=1}^n x_i}.$$

   Consider the ratio

$$\begin{aligned} \frac{p(x \mid \theta)}{p(y \mid \theta)} &= \frac{\theta^{\sum_{i=1}^n x_i} (1-\theta)^{n-\sum_{i=1}^n x_i}}{\theta^{\sum_{i=1}^n y_i} (1-\theta)^{n-\sum_{i=1}^n y_i}} = \theta^{\sum_i x_i - \sum_i y_i} (1-\theta)^{\sum_{i=1}^n y_i - \sum_{i=1}^n x_i} \\ &= \left(\frac{\theta}{1-\theta}\right)^{\sum_i x_i - \sum_i y_i} \end{aligned}$$

   The minimal sufficient statistic is $T = \sum_i x_i$.

2. Cauchy distribution with density

$$p(x \mid \theta) = \frac{1}{\pi\left[1 + (x-\theta)^2\right]}.$$

   The ratio is

$$\frac{p(x \mid \theta)}{p(y \mid \theta)} = \frac{\prod_{i=1}^n \left[1 + (y_i - \theta)^2\right]}{\prod_{i=1}^n \left[1 + (x_i - \theta)^2\right]} = \frac{\prod_{i=1}^n \left[1 + y_i^2 - 2y_i\theta + \theta^2\right]}{\prod_{i=1}^n \left[1 + x_i^2 - 2x_i\theta + \theta^2\right]}.$$

   The minimal sufficient statistic is the order statistic.

2.4 Completeness

Definition 4 (Completeness)**.** A statistical model $\{\mathrm{P}_\theta : \theta \in \Theta\}$ is called complete, if for any (Borel) measurable function $h : \mathcal{X} \to \mathbb{R}$, the statement

$$\mathrm{E}\left[h(X)\right] = 0 \text{ for all } \theta \in \Theta$$

implies the result $\mathrm{P}\{h(X) = 0 \mid \theta\} = 1$ for all $\theta \in \Theta$. A statistic $T \sim \mathrm{P}_\theta^T$ is called complete if the statistical model $\left\{\mathrm{P}_\theta^T : \theta \in \Theta\right\}$ is complete. If completeness holds for all bounded (Borel) measurable function $h()$, then we say it is boundedly complete.

Remark 1. By the definition, a complete statistic is also boundedly complete.

Example 8. Completeness means that as long as the expectation is zero for all θ, we must have $h = 0$ as the only unbiased estimator of 0. In order to show it is not complete, we just need to find a nonzero h such that its expectation is 0 for all θ.

1. Consider the statistical model $\{\text{Uniform}(0,\theta), \theta \in \mathbb{R}_+\}$. Let h be an arbitrary function. The statement $E[h(X)] = 0$ is equivalent to

$$\int_0^\theta h(x)\frac{1}{\theta}dx = 0 \text{ for any } \theta.$$

The statistical model is complete. See P6 of the book Theoretical Statistics for a more detailed discussion.

2. Let X_1, ..., X_n be i.i.d. random variable from Bernoulli(θ) where $\theta \in (0,1)$. Then, $T = \sum_{i=1}^n X_i$ follows Binomial. Then,

$$\begin{aligned} 0 = E[h(T)] &= \sum_{t=0}^n h(t)\binom{n}{t}\theta^t(1-\theta)^{n-t} \\ &= (1-\theta)^n \sum_{t=0}^n h(t)\binom{n}{t}\left(\frac{\theta}{1-\theta}\right)^t. \end{aligned}$$

The statistic T is a complete statistic.

3. Let X_1, ..., X_n be i.i.d. random variable from $N(\mu,\mu^2)$. The statistic $T = \left(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2\right)$ is not complete. This is because we let

$$h(T) = \left(\sum_{i=1}^n X_i\right)^2 - \frac{n+1}{2}\sum_{i=1}^n X_i^2$$

Note that $\mathrm{E}[\sum_{i=1}^n X_i] = n\mu$, $\mathrm{Var}[\sum_{i=1}^n X_i] = n\mu^2$, $\mathrm{E}\left[\sum_{i=1}^n X_i^2\right] = 2n\mu^2$. Then,

$$\mathrm{E}[h(T)] = \left(n+n^2\right)\mu^2 - \frac{n+1}{2}\cdot 2n\mu^2 = 0,$$

but $h(T) \neq 0$.

Proposition 1. *If T is a complete statistic and $S = \psi(T)$ for a continuous function ψ, then S is also complete. The proposition holds if we change both complete statistic to boundedly complete statistic.*

Proof. Since T is complete, then

$$\mathrm{E}[h(T)] = 0 \text{ for all } \theta \in \Theta$$

implies the result $\mathrm{P}\{h(T) = 0 \mid \theta\} = 1$ for all $\theta \in \Theta$. Assume that

$$\mathrm{E}[h(S)] = \mathrm{E}[h(\psi(T))] = 0 \text{ for all } \theta \in \Theta.$$

Since T is complete, we must have $\mathrm{P}\{h(\psi(T)) = 0 \mid \theta\} = 1$ for all $\theta \in \Theta$, as long as $h(\psi(\cdot))$ is measurable. That is $\mathrm{P}\{h(S) = 0 \mid \theta\} = 1$. □

Theorem 4 (Bahadur's theorem: Completeness and Minimal Sufficiency)**.** *If a minimal sufficient statistic exits, then*

1. *any sufficient and complete statistic is minimal sufficient.*

2. *any sufficient and boundedly complete statistic is minimal sufficient*

Proof. We show the first part of the theorem. Suppose that S is a minimal sufficient statistic, and that T is sufficient and complete. Then, $T - \mathrm{E}[T \mid S] = T - \mathrm{E}[T \mid S(T)]$ is a function of T, since S is minimal sufficient and is a function of any other sufficient statistic (including T). Note that

$$\mathrm{E}(T - \mathrm{E}[T \mid S]) = \mathrm{E}[T] - \mathrm{E}(\mathrm{E}[T \mid S]) = 0.$$

Since T is complete, then we must have $T = \mathrm{E}[T \mid S]$ with probability 1, that is, T is also a function of S. Since S is a function of any sufficient statistic, then T is a function of any sufficient statistic, which means that T is minimal sufficient. ☐

Theorem 5. *For the exponential family in canonical form*

$$p(x \mid \eta) = \exp\left\{\sum_{i=1}^{s} \eta_i T_i(x) - A(\eta)\right\} h(x),$$

consider the statistic $T = (T_1(x), \cdots, T_d(x))$. If the exponential family is of full rank, T is also complete.

Proof. If you are interested in the proof, see Brown (1986), Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory, Page 44. ☐

2.5 Ancillary Statistic

Definition 5 (Ancillary Statistic)**.** A statistic is called ancillary if its distribution does not depend on θ.

Example 9. Suppose that X_1, X_2 are iid $N(\theta, 1)$. Then, $X_1 - X_2 \sim N(0, 2)$ is an ancillary statistic.

Theorem 6 (Basu Theorem)**.** *If a statistic T is complete and sufficient for $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$, and if V is an ancillary statistic, then T and V are independent under P_θ for any $\theta \in \Theta$.*

Example 10. Suppose that X_1, ..., X_n are iid $N(\mu, \sigma^2)$. A well-known result is that $\bar{X}$ and $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ are independent. To show this result, we note that

$$p(x_1, ..., x_n \mid \theta) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2 + \frac{n\mu}{\sigma^2}\bar{x} - \frac{n\mu^2}{2\sigma^2}\right\}.$$

Hence, we have a full rank two-parameter exponential family. Thus, $T = \left(\bar{X}, \sum_{i=1}^{n} x_i^2\right)$ is sufficient and complete for the normal family with unknown μ. Define $Y_i = X_i - \mu \sim N(0, \sigma^2)$. Then,

$$X_i - \bar{X} = Y_i - \bar{Y}, \quad \forall i,$$

that only depends on σ^2, not μ. Further

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

only depends on σ^2, not μ. That is, S^2 is an ancillary statistic [no μ is involved]. By the Basu theorem, $\bar{X}$ and S^2 are independent.