Math 170S

Lecture Notes Section 6.7 [*] [†]

Sufficient statistics

Instructor: Swee Hong Chan

NOTE: The notes is a summary for materials discussed in the class and is not supposed to substitute the textbook. Please send me an email if you find typos.

[*]Version date: Monday 19th October, 2020, 11:10.

[†]This notes is based on Hanbaek Lyu's and Liza Rebrova's notes from the previous quarter, and I would like to thank them for their generosity. "*Nanos gigantum humeris insidentes* (I am but a dwarf standing on the shoulders of giants)".

1 Recap: MLE

Recall what we learn about MLE from before.

- **Problem:** X is an unknown random variable with distribution f_θ and unknown parameter θ.

- **Input:** Samples $x_1, x_2, \ldots, x_n$ of X.

- **Output:** An estimate $\widehat{\theta}$ for θ.

2 Example: Bernoulli

Let X be a Bernoulli random variable with parameter p.

The MLE is

$$\widehat{\mathrm{p}} = \frac{x_1 + \ldots + x_n}{n},$$

the sample mean. Other aspects (e.g., median, maximum, minimum) of $x_1, \ldots, x_n$ is irrelevant in computing $\widehat{\mathrm{p}}$.

Intuitively, this means that $\overline{\mathrm{x}}$ is a **sufficient statistics** for $\widehat{\mathrm{p}}$.

3 Example: Uniform

Let X be the uniform random variable on the interval $[0, \theta]$. The MLE is

$$\widehat{\theta} = \max(x_1, \ldots, x_n).$$

Note that you cannot compute $\widehat{\theta}$ by only knowing the sample mean.

Intuitively, this means that $\overline{\mathrm{x}}$ is **not a sufficient statistics** for $\widehat{\theta}$.

4 Statistics

A **statistic** $u := u(x_1, \ldots, x_n)$ is a real number whose values depend only on the samples $x_1, \ldots, x_n$ (and implicitly, on n).

For example, the functions

$$\frac{x_1 + \ldots + x_n}{n} \qquad \text{and} \qquad x_1^2 + \ldots + x_n^2$$

are statistics of X, as they depend only on $x_1, \ldots, x_n$. On the other hand, the function

$$\theta + x_1 + \ldots + x_n$$

is **not** a statistic as it also depends on (unknown) θ.

5 Sufficient statistics (rigorous)

A statistic $u := u(x_1, \ldots, x_n)$ is a **sufficient statistics**, if there exists functions ϕ and h such that

$$f_\theta(x_1) \ldots f_\theta(x_n) \;=\; \phi(u, \theta)\, h(x_1, \ldots, x_n),$$

where $\phi : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is a function that depends only on u and θ, and $h := h(x_1, \ldots, x_n)$ is a statistics that depend only on $x_1, \ldots, x_n$.

My mnemonic to remember the formula above is that

> "The u and θ are the favorite children of the family, so they are separated from the rest of the children (x's)."

6 Example: Bernoulli (rigorous)

Let X be the Bernoulli random variable with unknown parameter p. Let $u := u(x_1, \ldots, x_n)$ be the statistic

$$u = \frac{x_1 + \ldots + x_n}{n},$$

the sample mean. Then (BT)

$$f_p(x_1) \ldots f_p(x_n) \ = p^{nu}(1-p)^{n-nu}.$$

To check u is a sufficient statistics, let ϕ and h be

$$\phi(u,p) := p^{nu}(1-p)^{n-nu}; \qquad h(x_1, \ldots, x_n) := \quad 1.$$

Then we see that

$$f_p(x_1) \ldots f_p(x_n) \ = \ \phi(u,p)\, h(x_1, \ldots, x_n)\,.$$

7 Example: Poisson

Let X be the Poisson random variable with unknown parameter λ,

$$f_\lambda(x) = \frac{\lambda^x e^{-\lambda}}{x!} \qquad (x \in \{0, 1, 2, \ldots\}).$$

Let $u := u(x_1, \ldots, x_n)$ be the statistic

$$u = \frac{x_1 + \ldots + x_n}{n}.$$

We then have (BT)

$$\begin{aligned} f_\lambda(x_1) \ldots f_\lambda(x_n) &= \frac{\lambda^{x_1} e^{-\lambda}}{x_1!} \ldots \frac{\lambda^{x_n} e^{-\lambda}}{x_n!} \\ &= \lambda^{x_1 + \ldots + x_n} e^{-n\lambda} \frac{1}{x_1! x_2! \ldots x_n!} \\ &= \lambda^{nu} e^{-n\lambda} \frac{1}{x_1! x_2! \ldots x_n!}. \end{aligned}$$

To see that u is a sufficient statistics, let ϕ and h be

$$\begin{aligned}\phi(u,\lambda) :=& \lambda^{nu} e^{-n\lambda};\\ h(x_1,\ldots,x_n) :=& \frac{1}{x_1!x_2!\ldots x_n!}.\end{aligned}$$

Then we see that

$$f_\lambda(x_1)\ldots f_\lambda(x_n) \;=\; \phi(u,\lambda)\, h(x_1,\ldots,x_n),$$

as desired.

8 Example: Uniform (rigorous)

Let X be the uniform random variable on the interval $[0, \theta]$,

$$f_\theta(x) = \begin{cases} \frac{1}{\theta} & \text{if } 0 < x \leq \theta; \\ 0 & \text{otherwise.} \end{cases}$$

Let $u := u(x_1, \ldots, x_n)$ be the statistic

$$u = \frac{x_1 + \ldots + x_n}{n}.$$

We then have

$$f_\theta(x_1) \ldots f_\theta(x_n) = \begin{cases} \frac{1}{\theta}^n & \text{if } \theta \geq \max(x_1, \ldots, x_n), \\ 0 & \text{if } \theta < \max(x_1, \ldots, x_n). \end{cases}$$

This function cannot be written as *products* that separate u and θ from x's, so u is not a sufficient statistic.

On the other hand, let w be the statistics

$$w := \max(x_1, \ldots, x_n).$$

Let ϕ and h be defined by

$$\phi(w, \theta) := \begin{cases} \frac{1}{\theta}^n & \text{if } \theta \geq w, \\ 0 & \text{if } \theta < w. \end{cases};$$

$$h(x_1, \ldots, x_n) := 1.$$

Then we see that

$$f_\theta(x_1) \ldots f_\theta(x_n) \;=\; \phi(w, \theta)\, h(x_1, \ldots, x_n),$$

so w is a sufficient statistics.

9 Joint sufficient statistics

Recall that sometimes X is a random variable with two unknown parameters θ_1, θ_2.

Two statistics $u_1 := u_1(x_1, \ldots, x_n)$ and $u_2 := u_2(x_1, \ldots, x_n)$ are **joint sufficient statistics**, if there exists ϕ and h such that

$$f_{\theta_1,\theta_2}(x_1) \ldots f_{\theta_1,\theta_2}(x_n) \; = \; \phi(u_1, u_2, \theta_1, \theta_2)\, h(x_1, \ldots, x_n),$$

where $\phi : \mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is a function that depends only on $u_1, u_2, \theta_1, \theta_2$, and $h := h(x_1, \ldots, x_n)$ is a statistics that depend only on $x_1, \ldots, x_n$.

10 Example: Normal

Let X be a normal random variable with unknown mean μ and unknown variance σ^2,

$$f_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

Let $u_1 := u(x_1, \ldots, x_n)$ and $u_2 := u(x_1, \ldots, x_n)$ be the statistics

$$\begin{aligned} u_1 :=& x_1 + \ldots + x_n; \\ u_2 :=& x_1^2 + \ldots + x_n^2; \end{aligned}$$

We then have (BT)

$$
\begin{aligned}
& f_{\mu,\sigma^2}(x_1)\dots f_{\mu,\sigma^2}(x_n) \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_1-\mu)^2}{2\sigma^2}\right) \dots \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_n-\mu)^2}{2\sigma^2}\right) \\
&= (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right).
\end{aligned}
$$

Now note that

$$
\begin{aligned}
\sum_{i=1}^{n}(x_i-\mu)^2 &= \sum_{i=1}^{n}(x_i^2 - 2x_i\mu + \mu^2) \\
&= \left(\sum_{i=1}^{n} x_i^2\right) - 2\mu\left(\sum_{i=1}^{n} x_i\right) + \left(\sum_{i=1}^{n}\mu^2\right) \\
&= u_2 - 2\mu\, u_1 + n\mu^2.
\end{aligned}
$$

Combining the two equation above,

$$
f_{\mu,\sigma^2}(x_1)\dots f_{\mu,\sigma^2}(x_n) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{u_2 - 2\mu\, u_1 + n\mu^2}{2\sigma^2}\right).
$$

To see that u is a sufficient statistics, let ϕ and h be

$$\phi(u_1, u_2, \mu, \sigma^2) := (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{u_2 - 2\mu\, u_1 + n\mu^2}{2\sigma^2}\right);$$
$$h(x_1, \ldots, x_n) := 1.$$

Then we see that

$$f_{\mu,\sigma^2}(x_1) \ldots f_{\mu,\sigma^2}(x_n) \;=\; \phi(u_1, u_2, \mu, \sigma^2)\, h(x_1, \ldots, x_n),$$

as desired.