Review of Cramér - Rao inequality :

Under some mild technical conditions on the probability distribution $f(x;\theta)$

↳ key condition:
support does not depend on the unknown parameter!

fulfilled by all examples of relevance to us:

- Bernoulli
- Binomial
- geometric
- negative Binomial
- Poisson
- exponential
- gamma
- normal

For any unbiased estimator $\hat{\theta}$ of the unknown parameter θ, we have:

Cramér - Rao lower bound

Cramér - Rao inequality →

$$\boxed{Var(\hat{\theta}) \geq \frac{1}{I(\theta)}}$$

where $I(\theta)$ is the Fisher information, given by

$$I(\theta) = n E\left[\left(\frac{\partial}{\partial\theta} \ln f(X,\theta)\right)^2\right] = -n E\left[\frac{\partial^2}{\partial\theta^2} \ln f(X,\theta)\right]$$

is a measure of the amount of information about θ contained in $f(x,\theta)$

CONSEQUENCES

① If we happen to have an unbiased estimator Y of θ s.t

$$\mathrm{Var}(Y) = \frac{1}{I(\theta)}$$

Var(Y) reflects the spread of Y about its mean $\underbrace{E[Y]=\theta}_{\text{because } Y \text{ is unbiased}}$

then Y is a minimum variance unbiased estimator of θ

② We can define the efficiency of an unbiased estimator $\hat{\theta}$ of θ as

$$e(\hat{\theta}) = \frac{1/I(\theta)}{\mathrm{Var}(\hat{\theta})} \in (0,1]$$

If $e(\hat{\theta})$ then $\mathrm{Var}(\hat{\theta}) = \frac{1}{I(\theta)}$ and $\hat{\theta}$ is a MVUE of θ

More on the meaning of $I(\theta) = n\, E\left[\underbrace{\left(\frac{\partial}{\partial\theta}\ln f(x,\theta)\right)^2}\right] = n \cdot \mathrm{Var}\left(\frac{\partial}{\partial\theta}\ln f(x,\theta)\right)$

multiplying n interpretation:
the larger the sample size n is, the more information we may acquire from the distribution by sampling

$$\frac{\partial}{\partial\theta}\ln f(x,\theta) = \underbrace{\frac{\frac{\partial}{\partial\theta} f(x,\theta)}{f(x,\theta)}}_{\text{relative rate of change of } f \text{ w.r.t } \theta}$$

since $E\left[\frac{\partial}{\partial\theta}\ln f(x,\theta)\right] = 0$, then

$$\mathrm{Var}\left(\frac{\partial}{\partial\theta}\ln f(x,\theta)\right) = E\left[\left(\frac{\partial}{\partial\theta}\ln f(x,\theta)\right)^2\right] - \overbrace{E\left[\frac{\partial}{\partial\theta}\ln f(x,\theta)\right]}^{=0}$$

Let $f: I \longrightarrow \mathbb{R}$ be a real-valued differentiable function, $I \subseteq \mathbb{R}$ open interval.

Def : the relative rate of change of f at x is $\boxed{\dfrac{f'(x)}{f(x)}} = \boxed{\dfrac{d}{dx} \ln f(x)}$

EXAMPLE : City has population size given by $f(t)$, where t is years since the year 2000.

instantaneous rate of change of the population when $t=20$ (year 2020) $\rightarrow$ $f'(20) = 1000$ people/year $\Big\}$ interpretation : population is increasing at a rate of 1000 people per year at time $t=20$

how large the rate of change actually is depends on the size of $f(20)$

If $f(20) = 10{,}000$, then $\dfrac{f'(20)}{f(20)} = \dfrac{1000 \text{ people/year}}{10\,000 \text{ people}} = 0.1 \text{ /year}$

↳ rate of growth of 10% per year

If $f(20) = 10{,}000{,}000$, then

$$\frac{f'(20)}{f(20)} = \frac{1000}{10\,000\,000} = \frac{1}{10000} = 0.01\% \text{ /year}$$

↳ rate of growth

Math 4501 - Probability and Statistics II

6.7 - Sufficient Statistics

Sufficient Statistic

STATISTIC: function of random sample that DOES NOT depend on any unknown parameter.

Definition

Let $X_1, \ldots, X_n$ be a random sample from a distribution with pdf/pmf $f(\cdot;\theta)$ depending on a parameter θ. ← θ is the unknown parameter

A statistic $Y = u(X_1, \ldots, X_n)$ is said to be *sufficient for* θ if the conditional distribution of $X_1, \ldots, X_n$ given $Y = y$ does not depend on θ for any value y of Y.

Interpretation:

Y summarizes ALL the relevant info about θ in the random sample

- A sufficient statistic is a function of the random sample whose value contains all the information needed to compute any estimate of the parameter, i.e. there is no additional information about the unknown parameter left in the remaining (conditional) distribution.
- The joint probability distribution of the data is conditionally independent of the parameter given the value of the sufficient statistic for the parameter.

Example

Let $X_1, X_2, \dots, X_n$ be a random sample from the Bernoulli distribution with pmf

$$f(x;p) = p^x(1-p)^{1-x}\ , \quad x = 0, 1\ , \quad p \in (0,1)\ .$$

Use the definition to show that

$$Y = \sum_{i=1}^{n} X_i \longrightarrow \text{pmf of binomial } (n, p)$$

$$\binom{n}{y} p^y (1-p)^{n-y}, \quad y = 0, 1, 2, \dots, n$$

is a sufficient statistic for p.

binomial with $n=1$ and prob. of succ p

Note that since $X_1, X_2, \dots, X_n$ are Bernoulli(p) and independent, then

$$Y = \sum_{i=1}^{n} X_i \sim \text{Binomial}(n, p) \leftarrow \text{takes values in } \{0, 1, 2, \dots, n\}$$

To show that Y is a sufficient statistic for p, we need to check that

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n \mid Y = y)$$ does not depend on p for any $y \in \{0, 1, \dots, n\}$

Let us compute this conditional probability

$$\rightsquigarrow P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n \mid Y = y) = \frac{P(\{X_1 = x_1, X_2 = x_2, \dots, X_n = x_n\} \cap \{Y = y\})}{P(Y=y)}$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$= \frac{P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)}{P(Y=y)}$$

since $Y = \sum_{i=1}^{n} X_i$, then

$$y = \sum_{i=1}^{n} x_i$$

in which case

$$\{X_1 = x_1, \dots, X_n = x_n\} \subseteq \{Y = y\}$$

and so $\{X_1 = x_1, \dots, X_n = x_n\} \cap \{Y = y\} = \{X_1 = x_1, \dots, X_n = x_n\}$

$X_1, \dots, X_n$ are independent

because $X_1, \dots, X_n$ is a random sample!

$$= \frac{P(X_1 = x_1) \cdot P(X_2 = x_2) \cdots P(X_n = x_n)}{P(Y=y)}$$

$= \cdots$ (next slide)

using the respective pmfs we obtain

$$= \frac{\underbrace{p^{x_1}(1-p)^{1-x_1}}_{P(X_1=x_1)} \cdot \underbrace{p^{x_2}(1-p)^{1-x_2}}_{P(X_2=x_2)} \cdots\cdots \underbrace{p^{x_m}\cdot(1-p)^{1-x_m}}_{P(X_m=x_m)}}{\underbrace{\binom{m}{y} p^y (1-p)^{m-y}}_{P(Y=y)}}$$

$$= \frac{p^{x_1+x_2+\cdots+x_m}\cdot(1-p)^{1-x_1+1-x_2+\cdots+1-x_m}}{\binom{m}{y}p^y(1-p)^{m-y}} = \frac{p^{\overbrace{x_1+x_2+\cdots+x_m}^{y}}\cdot(1-p)^{m-\overbrace{(x_1+x_2+\cdots x_m)}^{y}}}{\binom{m}{y}p^y(1-p)^{m-y}}$$

$$= \frac{p^y(1-p)^{m-y}}{\binom{m}{y}p^y(1-p)^{m-y}} = \frac{1}{\binom{m}{y}}$$

does not depend on p no matter which value y takes

$\Longrightarrow Y = \sum_{i=1}^{m} X_i$ is a sufficient statistic for p.

Factorization Theorem

Theorem (Fisher-Neyman Factorization Theorem)

given

Let $X_1, X_2, \ldots, X_n$ denote random variables with joint pdf/pmf $f(x_1, x_2, \ldots, x_n; \theta)$ depending on the parameter θ.

The statistic $\boxed{Y = u(X_1, X_2, \ldots, X_n)}$ is sufficient for θ if and only if

$$f(x_1, x_2, \ldots, x_n; \theta) = \phi(\underbrace{u(x_1, x_2, \ldots, x_n)}_{Y}; \theta)\, \underbrace{h(x_1, x_2, \ldots, x_n)}_{}\,,$$

where ϕ depends on $x_1, x_2, \ldots, x_n$ only through $\underbrace{u(x_1, \ldots, x_n)}_{Y}$ and $h(x_1, \ldots, x_n)$ does not depend on θ.

Note: it is often easier to check sufficiency using the Factorization Theorem than it is using the definition.

Example (same example as before!)

Let $X_1, X_2, \ldots, X_n$ be a random sample from the Bernoulli distribution with pmf

$$f(x;p) = p^x(1-p)^{1-x} \;, \quad x = 0, 1 \;, \quad p \in (0,1) \;.$$

Use the Factorization Theorem to show that

$$Y = \sum_{i=1}^{n} X_i \quad \leftarrow \quad Y \sim \text{Binomial}(n, p)$$

is a sufficient statistic for p.

Note that we may write the joint pmf of $X_1, \ldots, X_n$ as

$$\rightsquigarrow \quad f_{\text{joint}}(x_1, x_2, \ldots, x_n; p) = \prod_{i=1}^{n} f(x_i; p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\overbrace{\sum_{i=1}^{n} x_i}^{y}} (1-p)^{n - \overbrace{\sum_{i=1}^{n} x_i}^{y}}$$

since $X_1, X_2, \ldots, X_n$ random sample $\Rightarrow X_1, \ldots, X_n$ iid

$$p^{x_1} \cdot (1-p)^{1-x_1} \cdot p^{x_2} \cdot (1-p)^{1-x_2} \cdots p^{x_n}(1-p)^{1-x_n}$$

We got that

$$f(x_1, \dots, x_n; p) = \overbrace{p^{y} (1-p)^{n-y}}^{\phi(y,p)} \cdot \underbrace{1}_{h(x_1, \dots, x_n)}$$

is of the form

$$f(x_1, \dots, x_n; p) = \phi(y, p) \cdot h(x_1, \dots, x_n)$$

where $\phi(y, p) = p^{y} (1-p)^{n-y}$ with $y = \sum_{i=1}^{n} x_i$

and $h(x_1, x_2, \dots, x_n) = 1$

By the factorization theorem, $Y = \sum_{i=1}^{n} X_i$ is a sufficient statistic for p.

Example

Let $X_1, X_2, \ldots, X_n$ denote a random sample from a Poisson distribution with parameter $\lambda > 0$.

Show that the sample mean $\bar{X}$ is a sufficient statistic for λ.

Recall that pmf of Poisson λ is $\boxed{f(x,\lambda) = \dfrac{\lambda^{x} e^{-\lambda}}{x!}, \quad x = 0, 1, 2, \ldots, \quad \lambda > 0}$

$$\rightarrow \quad f_{joint}(x_1, x_2, \ldots, x_n; \lambda) = \prod_{i=1}^{n} f(x_i; \lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} = \frac{\lambda^{\overbrace{\sum_{i=1}^{n} x_i}^{n\bar{x}}} e^{-n\lambda}}{\prod_{i=1}^{n} (x_i!)}$$

$X_1, \ldots, X_n$ are iid being a random sample

$$\frac{\lambda^{x_1} e^{-\lambda}}{x_1!} \cdot \frac{\lambda^{x_2} e^{-\lambda}}{x_2!} \cdots \frac{\lambda^{x_n} e^{-\lambda}}{x_n!}$$

Note that $\sum_{i=1}^{n} x_i = n\bar{x}$

We got that

$$f_{joint}(x_1, x_2, \dots, x_n; \lambda) = \underbrace{\lambda^{n\bar{x}} e^{-n\lambda}}_{\phi(\bar{x}, \lambda)} \cdot \underbrace{\frac{1}{\prod_{i=1}^{n} x_i!}}_{h(x_1, x_2, \dots, x_n)}$$

is of the form

$$f_{joint}(x_1, \dots, x_n; \lambda) = \phi(y, \lambda) \cdot h(x_1, x_2, \dots, x_n)$$

with $\underbrace{\phi(y, \lambda) = \lambda^{ny} \cdot e^{-ny}}_{\text{depends on } x_1, \dots, x_n \text{ only through } y = \bar{x}}$, $y = \bar{x}$, and $\overbrace{h(x_1, \dots, x_n) = \frac{1}{\prod_{i=1}^{n} (x_i!)}}^{\text{does not depend on } \lambda}$

By the factorization theorem $\bar{X}$ is a sufficient statistic for λ