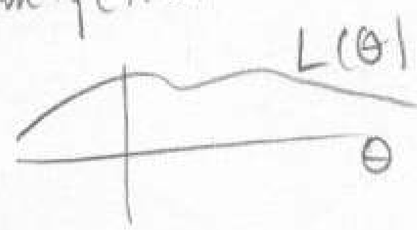
Sufficiency

Def. Let $Y_1 \ldots Y_n$ denote a random sample from a probability distribution with unknow parameter θ. Then, the statistic $U = g(Y_1 \ldots Y_n)$ is said to be sufficient for θ if the conditional distribution of $Y_1 \ldots Y_n$ given U does not depend on θ.

Def Let $y_1 \ldots y_n$ be sample observations taken from $f(y|\theta)$



The likelihood of the sample is

$$L(y_1 \ldots y_n|\theta) = f(y_1|\theta) \ldots f(y_n|\theta)$$

In the continuous case, $f(y_i|\theta)$ is a density

In the discrete case, $f(y_i|\theta)$ is the probability function

We want to estimate θ. We reduce the information from $(x_1 \ldots x_n)$ into $U = g(Y_1 \ldots Y_n)$. We donot want to loose any information about θ.

A sufficient statistic for a parameter θ is a statistic that in a certain sense captures all the information about θ contained in the sample

Theorem (Factorization theorem). U is a sufficient statistic for θ if and only if the likelihood can be factored into two nonnegative functions

$$L(y_1 \ldots y_n|\theta) = g(u, \theta)\, h(y_1 \ldots y_n)$$

where $g(u, \theta)$ is a function only of u and θ and $h(y_1 \ldots y_n)$ is a function only of $y_1 \ldots y_n$ (not of θ)

If $g(x_1 \ldots x_n) = g(y_1 \ldots y_n) = u$, then

$$\frac{L(y_1 \ldots y_n|\theta)}{h(y_1 \ldots y_n)} = g(u,\theta) = \frac{L(x_1 \ldots x_n|\theta)}{h(x_1 \ldots x_n)}$$

$L(y_1 \ldots y_n|\theta)$ (We make the same inference

$L(x_1 \ldots x_n|\theta)$ under $(y_1 \ldots y_n)$ than under $(x_1 \ldots x_n)$

So, any inference depends on the sample $(y_1 \ldots y_n)$ only through $g(y_1 \ldots y_n)$

A sufficient statistic contains all the information about θ which is contained in the data $y_1 \ldots y_n$. Once the value of $u = g(y_1 \ldots y_n)$ is known, we cannot squeeze any more information out the $y_1 \ldots y_n$ regarding to θ.

9.31. Let $Y_1, \dots Y_n$ denote a random sample from a Poisson distribution with parameter λ. Show by conditioning that $\sum_{i=1}^{n} Y_i$ is sufficient for λ

The probability function of the sample is

$$\prod_{i=1}^{n} p(y_i|\theta) = \prod_{i=1}^{n} e^{-\theta} \frac{\theta^{y_i}}{y_i!} = e^{-n\theta} \frac{\theta^{\sum_{i=1}^{n} y_i}}{\prod y_i!}$$

The probability function of $\sum_{i=1}^{n} Y_i \sim$ Poisson $(n\theta)$ is

$$q(u|\theta) = e^{-n\theta} \frac{(n\theta)^u}{u!}, \text{ where } u = \sum_{i=1}^{n} y_i$$

The conditional probability function of the sample given $\sum_{i=1}^{n} Y_i$ is

$$\frac{\prod_{i=1}^{n} p(y_i|\theta)}{q(u|\theta)} = \frac{e^{-n\theta} \frac{\theta^{\sum_{i=1}^{n} y_i}}{\prod y_i!}}{e^{-n\theta} \frac{(n\theta)^u}{u!}} = \frac{u!}{\prod y_i!} \frac{\theta^{\sum y_i}}{(n\theta)^{\sum y_i}} = \frac{u!}{\prod y_i! \; n^{\sum y_i}}$$

which does not depend on θ.

So, $\sum Y_i$ is a sufficient statistic for θ

9.30 Let $Y_1 \dots Y_n$ denote a random sample from a normal distribution with mean μ and variance σ^2?

a. If μ is unknown and σ^2 is known, show that $\bar{Y}$ is sufficient for μ

$$L(Y_1 \dots Y_n \mid \mu) = \prod_{i=1}^{n} \frac{e^{-\frac{(y_i-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\,\sigma} = \frac{e^{-\sum_{i=1}^{n}\frac{(y_i-\mu)^2}{2\sigma^2}}}{(\sqrt{2\pi})^n \sigma^n} = \frac{e^{-\frac{\sum y_i^2}{2\sigma^2} + \frac{\sum y_i \mu}{\sigma^2} - \frac{n\mu^2}{2\sigma^2}}}{(\sqrt{2\pi})^n\ \sigma^n}$$

$$= \underbrace{\frac{e^{\frac{\bar{Y} n\mu}{\sigma^2} - \frac{n\mu^2}{2\sigma^2}}}{\sqrt{2\pi}\ \sigma^n}}_{g(\bar{Y},\mu)} \quad \underbrace{e^{-\frac{\sum y_i^2}{2\sigma^2}}}_{h(y_1 \dots y_n)}$$

So, $\bar{Y}$ is sufficient for μ

b. If μ is known and σ^2 is unknown, show that $\sum_{i=1}^{n}(Y_i-\mu)^2$ is sufficient for σ^2?

$$L(y_1 \dots y_n \mid \sigma^2) = \underbrace{\frac{e^{-\frac{\sum(y_i-\mu)^2}{2\sigma^2}}}{(\sqrt{2\pi})^n\ \sigma^n}}_{g(\sum(y_i-\mu)^2,\ \sigma^2)} \cdot \underbrace{1}_{h(y_1 \dots y_n)}$$

So, $\sum_{i=1}^{n}(Y_i-\mu)^2$ is sufficient for σ^2.

c. If μ and σ^2 are both unknown, show that $\sum_{i=1}^{n} Y_i$ and $\sum_{i=1}^{n} Y_i^2$ are jointly sufficient for μ and σ^2? (Thus, it follows that $\bar{Y}$ and $\sum_{i=1}^{n}(Y_i-\bar{Y})^2$ or $\bar{Y}$ and S^2 are also jointly sufficient for μ and σ^2

$$L(y_1 \ldots y_n \mid \mu, \sigma^2) = \underbrace{\frac{e^{-\frac{\sum y_i^2}{2\sigma^2} + \frac{\sum y_i \mu}{\sigma^2} - \frac{\mu^2 n}{2\sigma^2}}}{(\sqrt{2\pi})^n \, \sigma^n}}_{g\left(\sum_{i=1}^{n} y_i,\ \sum_{i=1}^{n} y_i^2,\ \mu,\ \sigma^2\right)} \cdot \underbrace{1}_{h(y_1 \ldots y_n)}$$

9.34 If $Y_1, \ldots Y_n$ denote a random sample from a geometric distribution with parameter p, show that $\bar{Y}$ is sufficient for p

$$f(y|p) = p\,(1-p)^{y-1}, \quad y = 1, 2, \ldots$$

$$L(y_1 \ldots y_n|p) = \prod_{i=1}^{n} f(y_i|p) = p^n (1-p)^{\sum y_i - n} = \underbrace{p^n (1-p)^{n\bar{y}-n}}_{g(\bar{y},p)} \cdot \underbrace{1}_{h(y_1 - y_n)}$$

9.32 Let $Y_1 - Y_n$ denote independent and identically distribution from a Pareto distribution with parameters α and β, i.e

$$f(y|\alpha,\beta) = \begin{cases} \alpha \beta^{\alpha} y^{-(\alpha+1)} & y \geq \beta \\ 0 & \text{else} \end{cases}$$

If β is known, show that $\prod_{i=1}^{n} Y_i$ is sufficient for α

$$L(y_1 - y_n|\alpha) = \prod_{i=1}^{n} \frac{\alpha\beta^{\alpha}}{y_i^{\alpha+1}} I(y_i \geq \beta) = \underbrace{\frac{\alpha^n \beta^{n\alpha}}{\left(\prod_{i=1}^{n} y_i\right)^{\alpha+1}}}_{g(\prod y_i, \alpha)} \underbrace{I(Y_{(1)} \geq \beta)}_{h(y_1 - y_n)}$$

9.41 Let $Y_1 \ldots Y_n$ denote a random sample from the uniform distribution over the interval $(0, \theta)$. Show that $Y_{(n)} = \max(Y_1 \ldots Y_n)$ is sufficient for θ.

$$f(y|\theta) = \begin{cases} \frac{1}{\theta} & \text{if } 0 < y < \theta \\ 0 & \text{else} \end{cases} = \frac{1}{\theta} I(0 < y < \theta)$$

$$L(y_1 \ldots y_n | \theta) = \prod_{j=1}^{n} \frac{1}{\theta} I(0 < y_j < \theta) = \frac{1}{\theta^n} I(0 < Y_{(1)}, Y_{(n)} < \theta)$$

$$= \underbrace{\frac{1}{\theta^n} I(Y_{(n)} < \theta)}_{g(Y_{(n)}, \theta)} \underbrace{I(0 < Y_{(1)})}_{h(y_1 \ldots y_n)}$$

9.39. Let $Y_1 \ldots Y_n$ denote a random sample from the probability density function

$$f(y|\theta) = \begin{cases} e^{-(y-\theta)}, & \text{if } y \geq \theta \\ 0 & \text{elsewhere} \end{cases}$$

Show that $Y_{(1)} = \min(Y_1 \ldots Y_n)$ is sufficient for θ.

$$f(y|\theta) = e^{-y+\theta} I(y \geq \theta)$$

$$L(y_1 \ldots y_n | \theta) = \prod_{i=1}^{n} e^{-y_i+\theta} I(y_i \geq \theta) = e^{-\sum y_i + n\theta} I(Y_{(1)} \geq \theta)$$

$$= \underbrace{I(Y_{(1)} \geq \theta) e^{n\theta}}_{g(Y_{(1)}, \theta)} \underbrace{e^{-\sum y_i}}_{h(y_1 \ldots y_n)}$$

9.47. Let $Y_1 \ldots Y_n$ denote independent and identically distributed random samples from a Pareto distribution with parameters α and β.

$$f(y|\alpha,\beta) = \begin{cases} \alpha \beta^{\alpha} y^{-(\alpha+1)} & , \text{if } y \geq \beta \\ 0 & \text{else} \end{cases}$$

Show that $\prod_{i=1}^{n} Y_i$ and $\min(Y_1 \ldots Y_n)$ are jointly sufficient for α and β.

$$f(y|\alpha,\beta) = \frac{\alpha\beta^{\alpha}}{y^{\alpha+1}} I(y \geq \beta)$$

$$L(y_1 \ldots y_n|\alpha,\beta) = \prod_{j=1}^{n} \frac{\alpha\beta^{\alpha} I(y_j \geq \beta)}{y_j^{\alpha+1}} = \underbrace{\frac{\alpha^n \beta^{n\alpha} I(Y_{(1)} \geq \beta)}{\left(\prod_{j=1}^{n} Y_j\right)^{\alpha+1}}}_{g(Y_{(1)}, \prod Y_j, \alpha, \beta)} \cdot \underbrace{1}_{h(y_1 \ldots y_n)}$$