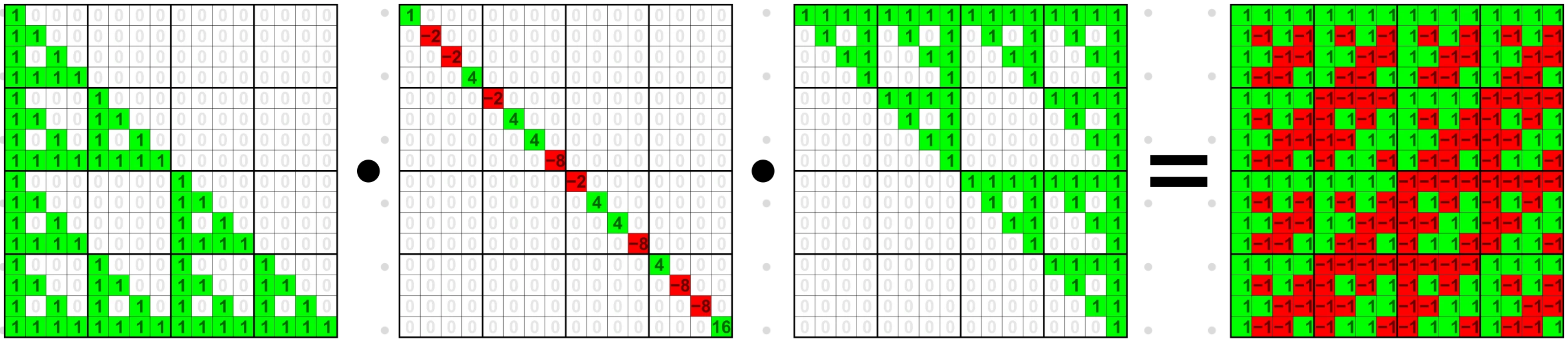

Statistics

Lecture 7

a general introduction to inference.

MATH 414

In a statistics problem, for a given random variable X, we are unaware of its p.m.f. $p(x)$ or p.d.f. $f(x)$.

In a statistics problem, for a given random variable X, we are unaware of its p.m.f. $p(x)$ or p.d.f. $f(x)$.

① We do not know $f(x)$ or $p(x)$ at all.

In a statistics problem, for a given random variable X, we are unaware of its p.m.f. $p(x)$ or p.d.f. $f(x)$.

① We do not know $f(x)$ or $p(x)$ at all.

② We know the form of $f(x)$ or $p(x)$ but not some parameter θ.

In a statistics problem, for a given random variable X, we are unaware of its p.m.f. $p(x)$ or p.d.f. $f(x)$.

① We do not know $f(x)$ or $p(x)$ at all.

② we know the form of $f(x)$ or $p(x)$ but not some parameter θ.

we focus on this case mostly.

Examples.

① $X \sim \text{Exp}(\theta)$ but we don't know θ.

Examples.

1. $X \sim \mathrm{Exp}(\theta)$ but we don't know θ.

2. $X \sim$ Binomial (n, p). we know n but not p.

Examples.

1. $X \sim \mathrm{Exp}(\theta)$ but we don't know θ.

2. $X \sim$ Binomial (n, p). we know n but not p.

3. $X \sim$ Gamma(α, β), both α and β are unknown.

Examples.

(1) $X \sim Exp(\theta)$ but we don't know θ.

(2) $X \sim$ Binomial (n, p). we know n but not p.

(3) $X \sim$ Gamma(α, β), both α and β are unknown.

(4) $X \sim$ Normal (μ, σ^2), both μ and σ^2 are unknown

We write $f(x;\theta)$ or $p(x;\theta)$ where $\theta \in \Omega$ for a specified set Ω.

The quantity θ is called a parameter of X.

One part of statistics is to estimate parameters.

Estimation is done through samples. Let $X_1, X_2, \ldots X_n$ be i.i.d. r.v.s. Then they constitute a sample of size n from the common distribution X.

Let $X_1, X_2, \ldots, X_n$ denote a sample on a random variable X. Let $T = T(X_1, X_2, \ldots, X_n)$ be a function of the sample. Then T is called a statistic.

Let $X_1, X_2, \dots, X_n$ denote a sample on a random variable X. Let $T = T(X_1, X_2, \dots, X_n)$ be a function of the sample. Then T is called a statistic.

Once the sample is drawn, then t is called the realization of T, where $t = T(x_1, x_2, \dots, x_n)$ and $x_1, x_2, \dots, x_n$ is the realization of the sample.

Point Estimators

The function T of the random sample $X_1, X_2, \ldots, X_n$ is called a point estimator for θ. A realization t of T is called an estimate.

Point Estimators

The function T of the random sample $X_1, X_2, \dots, X_n$ is called a point estimator for θ. A realization t of T is called an estimate.

Let $X_1, X_2, \dots, X_n$ be a sample on an r.v. X with p.d.f. $f(x, \theta)$, $\theta \in \Omega$. Let $T = T(X_1, X_2 \dots X_n)$ be a statistic. We say that T is unbiased estimator of θ if $E[T] = \theta$.

maximum likelihood Estimator (m.l.e)

Maximum likelihood Estimator (m.l.e)

The joint distribution of the random sample $X_1, \ldots, X_n$ is written as $\prod_{i=1}^{n} f(x_i; \theta)$.

since it is i.i.d.

Consider this as a function of θ

$$L(\theta) = L(\theta; x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f(x_i; \theta).$$

maximum likelihood Estimator (m.l.e)

The joint distribution of the random sample $X_1, \ldots, X_n$ is written as $\prod_{i=1}^{n} f(x_i; \theta)$.

since it is i.i.d.

Consider this as a function of θ

$$L(\theta) = L(\theta; x_1, x_2, \ldots, x_n) = \prod_{i=1}^{\theta} f(x_i; \theta).$$

This is called the likelihood function of the random sample.

In order to estimate θ, we might want to find the parameter θ that maximizes $\mathcal{L}(\theta)$

$$\boxed{\hat{\theta} = \operatorname{Argmax} L(\theta)}$$

= the value of θ that maximizes $L(\theta)$.

In order to estimate θ, we might want to find the parameter θ that maximizes $\mathcal{L}(\theta)$

$$\boxed{\hat{\theta} = \operatorname{Argmax} L(\theta)}$$

$= $ the value of θ that maximizes $L(\theta)$.

Sometimes one works with the log of the likelihood function since log function is monotone increasing. We check for maximality by checking $\frac{\partial l(\theta)}{\partial \theta} = 0$ for $l(\theta) = \log L(\theta)$

Examples

① Exponential Distribution

Suppose a random sample $X_1, X_2, \ldots, X_n$ has the identical distribution $f(x) = \theta^{-1} e^{-x/\theta}$

$$L(\theta) = \prod_{i=1}^{n} f(x_i ; \theta) = \theta^{-n} e^{-\sum_{i=1}^{n} x_i/\theta}$$

$$\therefore \ l(\theta) = \log L(\theta) = -n \log \theta - \theta^{-1} \sum_{i=1}^{n} x_i .$$

$$\frac{\partial l}{\partial \theta}(\theta) = \frac{-n}{\theta} + \frac{1}{\theta^2} \sum_{i=1}^{n} x_i = 0 \Rightarrow \boxed{\theta = \frac{\sum_{i=1}^{n} x_i}{n}} .$$

Examples

① Exponential Distribution

Suppose a random sample $X_1, X_2, \ldots, X_n$ has the identical distribution $f(x) = \theta^{-1} e^{-x/\theta}$

$$L(\theta) = \prod_{i=1}^{n} f(x_i;\theta) = \theta^{-n} e^{-\sum_{i=1}^{n} x_i/\theta}$$

$$\therefore\ l(\theta) = \log L(\theta) = -n\log\theta - \theta^{-1}\sum_{i=1}^{n} x_i .$$

$$\frac{\partial l}{\partial \theta}(\theta) = \frac{-n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n} x_i = 0 \Rightarrow \boxed{\theta = \frac{\sum_{i=1}^{n} x_i}{n}}.$$

mle is sample mean which is unbiased.

Examples

① **Normal Distribution** Let X have a $N(\mu, \sigma^2)$ distribution. In this case $\underline{\theta} = (\mu, \sigma)$ is a vector.

The m.l. function

$$L(\underline{\theta}) = L(\mu, \sigma) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)$$

$$\therefore \ l(\mu, \theta) = -\frac{n}{2}\log 2\pi - n\log\sigma - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2.$$

Examples

① **Normal distribution** Let X have a $N(\mu, \sigma^2)$ distribution. In this case $\underline{\theta} = (\mu, \sigma)$ is a vector.

The m.l. function

$$L(\underline{\theta}) = L(\mu, \sigma) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)$$

$$\therefore \ l(\mu, \sigma) = -\frac{n}{2}\log 2\pi - n\log\sigma - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2.$$

$$\frac{\partial l}{\partial \mu} = -\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)\left(-\frac{1}{\sigma}\right) \ ; \ \frac{\partial l}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2.$$

Examples

① **Normal Distribution** Let X have a $N(\mu, \sigma^2)$ distribution. In this case $\underline{\theta} = (\mu, \sigma)$ is a vector.

The m.l. function

$$L(\underline{\theta}) = L(\mu, \sigma) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)$$

$$\therefore\ l(\mu, \sigma) = -\frac{n}{2}\log 2\pi - n\log\sigma - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2.$$

$$\frac{\partial l}{\partial \mu} = -\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)\left(-\frac{1}{\sigma}\right) = 0 \; ; \quad \frac{\partial l}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 = 0.$$

$$\Rightarrow \boxed{\hat{\mu} = \bar{X}} \sim \boxed{\text{unbiased}} \qquad \boxed{\text{biased}} \sim \boxed{\hat{\sigma} = \frac{\sum (X_i - \bar{X})^2}{n}}$$

Exercises for assignment

① m.l.e. for binomial, uniform distributions etc.