Introductory Statistics

Tatiana Tyukina
tt51@leicester.ac.uk

Topic 2 - Estimators
2.2 - Properties of Estimators
Sufficiency and Consistency

- **Topic 0: Introduction**
- **Topic 1: Descriptive Statistics**
- **Topic 2: Estimators: Point estimation, Interval estimation**
  2.1 **Point estimation**,
  2.2 Properties of estimators,
  2.3 Interval estimation,
  2.4 Inference based on Normal Distribution
- Topic 3: Hypothesis Testing: Test for single population mean, test for two population mean
- Topic 4: Bayesian Estimation
- Topic 5: Goodness of Fit: The χ^2 test.

Desired properties of point estimator

- unbiasedness,

Desired properties of point estimator

- unbiasedness,
- efficiency (minimal variance),

Desired properties of point estimator

- unbiasedness,
- efficiency (minimal variance),
- sufficiency,

Desired properties of point estimator

- unbiasedness,
- efficiency (minimal variance),
- sufficiency,
- consistency

DEFINITION

Suppose that we learn that an event B has occurred and that we wish to compute the probability of another event A taking into account that we know that B has occurred. The new probability of A is called the ***conditional probability*** *of the event A given that the event B has occurred and is denoted* $P(A|B)$. *If* $P(B) > 0$, *we compute this probability as*

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

The conditional probability $P(A|B)$ *is not defined if* $P(B) = 0$.

DEFINITION

*Suppose that we learn that an event B has occurred and that we wish to compute the probability of another event A taking into account that we know that B has occurred. The new probability of A is called the **conditional probability** of the event A given that the event B has occurred and is denoted $P(A|B)$. If $P(B) > 0$, we compute this probability as*

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

The conditional probability $P(A|B)$ is not defined if $P(B) = 0$.

THEOREM (MULTIPLICATION RULE FOR CONDITIONAL PROBABILITIES)

Let A and B be events.
If $P(B) > 0$, then

$$P(A \cap B) = P(B)P(A|B).$$

If $P(A) > 0$, then

$$P(A \cap B) = P(A)P(B|A).$$

DEFINITION

Let $X_1, ..., X_n$ be a random sample from a probability distribution with unknown parameter θ. Then, the statistic $U = u(X_1, ..., X_n)$ is said to be ***sufficient*** *for θ if the conditional pdf $f_X(X_1, ..., X_n | U = u(x_1, x_2, ..., x_n))$ (or pmf $p_X(X_1, ..., X_n | U = u(x_1, x_2, ..., x_n)))$ does not depend on θ for any value of $u(x_1, x_2, ..., x_n)$.*

An estimator $\hat{\theta}$ that is a function of a sufficient statistic for θ is said to be a ***sufficient estimator*** *of θ.*

$$f_X(X_1, ..., X_n \mid U = u) = \frac{f(x_1, ..., x_n, u)}{f_U(U = u)}$$

EXAMPLE

Let $X_1, ..., X_5$ be a random sample of size 5 drawn from the Bernoulli pmf:
$p_X(x, p) = p^x(1-p)^{1-x}$, where $x = 0, 1$ and p is unknown parameter.
The maximum likelihood estimator for p is $\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i$
(then the MLE for p is $\hat{p}_l = \frac{1}{5}\sum_{i=1}^{5} x_i$) 11001
Show that $\hat{p}$ is a sufficient estimator for p.

EXAMPLE

Let $X_1, ..., X_5$ be a random sample of size 5 drawn from the Bernoulli pmf: $p_X(x,p) = p^x(1-p)^{1-x}$, where $x = 0, 1$ and p is unknown parameter.
The maximum likelihood estimator for p is $\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i$
(then the MLE for p is $\hat{p}_l = \frac{1}{5}\sum_{i=1}^{5} x_i$)
Show that $\hat{p}$ is a sufficient estimator for p.

Solution.

$$P(X_1 = x_1, ..., X_5 = x_5 |\, \hat{p} = \hat{p}_l) \quad = \frac{P(X_1=x_1,\ldots,X_5=x_5 \cap \hat{p}=\hat{p}_l)}{P(\hat{p}=\hat{p}_l)} = \frac{P(X_1=x_1,\ldots,X_5=x_5)}{P(\hat{p}=\hat{p}_l)}$$

EXAMPLE

Let $X_1, ..., X_5$ be a random sample of size 5 drawn from the Bernoulli pmf: $p_X(x,p) = p^x(1-p)^{1-x}$, where $x = 0, 1$ and p is unknown parameter.
The maximum likelihood estimator for p is $\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i$
(then the MLE for p is $\hat{p}_l = \frac{1}{5}\sum_{i=1}^{5} x_i$)
Show that $\hat{p}$ is a sufficient estimator for p.
Solution.

$$P(X_1 = x_1, ..., X_5 = x_5 | \hat{p} = \hat{p}_l) = \frac{P(X_1=x_1,\ldots,X_5=x_5 \cap \hat{p}=\hat{p}_l)}{P(\hat{p}=\hat{p}_l)} = \frac{P(X_1=x_1,\ldots,X_5=x_5)}{P(\hat{p}=\hat{p}_l)}$$

$$= \frac{p^{x_1}(1-p)^{1-x_1}\ldots p^{x_5}(1-p)^{1-x_5}}{P(\hat{p}=\hat{p}_l)} = \frac{p^{\sum_{i=1}^{5} x_i}(1-p)^{5-\sum_{i=1}^{5} x_i}}{P(\hat{p}=\hat{p}_l)} = \frac{p^{5\hat{p}_l}(1-p)^{5-5\hat{p}_l}}{P(\hat{p}=\hat{p}_l)}$$

EXAMPLE

Let $X_1, ..., X_5$ be a random sample of size 5 drawn from the Bernoulli pmf: $p_X(x,p) = p^x(1-p)^{1-x}$, where $x = 0, 1$ and p is unknown parameter.
The maximum likelihood estimator for p is $\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i$
(then the MLE for p is $\hat{p}_l = \frac{1}{5}\sum_{i=1}^{5} x_i$)
Show that $\hat{p}$ is a sufficient estimator for p.
Solution.

$$P(X_1 = x_1, ..., X_5 = x_5 | \, \hat{p} = \hat{p}_l) = \frac{P(X_1=x_1,\ldots,X_5=x_5 \cap \hat{p}=\hat{p}_l)}{P(\hat{p}=\hat{p}_l)} = \frac{P(X_1=x_1,\ldots,X_5=x_5)}{P(\hat{p}=\hat{p}_l)}$$

$$= \frac{p^{x_1}(1-p)^{1-x_1}\ldots p^{x_5}(1-p)^{1-x_5}}{P(\hat{p}=\hat{p}_l)} = \frac{p^{\sum_{i=1}^{5} x_i}(1-p)^{5-\sum_{i=1}^{5} x_i}}{P(\hat{p}=\hat{p}_l)} = \frac{p^{5\hat{p}_l}(1-p)^{5-5\hat{p}_l}}{P(\hat{p}=\hat{p}_l)}$$

EXAMPLE

Let $X_1, ..., X_5$ be a random sample of size 5 drawn from the Bernoulli pmf:
$p_X(x,p) = p^x(1-p)^{1-x}$, where $x = 0, 1$ and p is unknown parameter.
The maximum likelihood estimator for p is $\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i$
(then the MLE for p is $\hat{p}_l = \frac{1}{5}\sum_{i=1}^{5} x_i$)
Show that $\hat{p}$ is a sufficient estimator for p.

Solution.

$$P(X_1 = x_1, ..., X_5 = x_5 | \hat{p} = \hat{p}_l) = \frac{P(X_1=x_1,\dots,X_5=x_5 \cap \hat{p}=\hat{p}_l)}{P(\hat{p}=\hat{p}_l)} = \frac{P(X_1=x_1,\dots,X_5=x_5)}{P(\hat{p}=\hat{p}_l)}$$

$$= \frac{p^{x_1}(1-p)^{1-x_1}\dots p^{x_5}(1-p)^{1-x_5}}{P(\hat{p}=\hat{p}_l)} = \frac{p^{\sum_{i=1}^5 x_i}(1-p)^{5-\sum_{i=1}^5 x_i}}{P(\hat{p}=\hat{p}_l)} = \frac{p^{5\hat{p}_l}(1-p)^{5-5\hat{p}_l}}{P(\hat{p}=\hat{p}_l)}$$

$$P(\hat{p} = \hat{p}_l) = P\left(\frac{1}{5}\sum_{i=1}^{5} X_i = \hat{p}_l\right) = P\left(\underbrace{\sum_{i=1}^{5} X_i}_{\text{Binomial }(5,p)} = 5\hat{p}_l\right) = \binom{5}{5\hat{p}_l} p^{5\hat{p}_l}(1-p)^{5-5\hat{p}_l}$$

EXAMPLE

Let $X_1, ..., X_5$ be a random sample of size 5 drawn from the Bernoulli pmf:
$p_X(x,p) = p^x(1-p)^{1-x}$, where $x = 0, 1$ and p is unknown parameter.
The maximum likelihood estimator for p is $\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i$
(then the MLE for p is $\hat{p}_l = \frac{1}{5}\sum_{i=1}^{5} x_i$)
Show that $\hat{p}$ is a sufficient estimator for p.

Solution.

$$P(X_1 = x_1, ..., X_5 = x_5 | \hat{p} = \hat{p}_l) = \frac{P(X_1=x_1,\ldots,X_5=x_5 \cap \hat{p}=\hat{p}_l)}{P(\hat{p}=\hat{p}_l)} = \frac{P(X_1=x_1,\ldots,X_5=x_5)}{P(\hat{p}=\hat{p}_l)}$$

$$= \frac{p^{x_1}(1-p)^{1-x_1}\ldots p^{x_5}(1-p)^{1-x_5}}{P(\hat{p}=\hat{p}_l)} = \frac{p^{\sum_{i=1}^{5} x_i}(1-p)^{5-\sum_{i=1}^{5} x_i}}{P(\hat{p}=\hat{p}_l)} = \frac{p^{5\hat{p}_l}(1-p)^{5-5\hat{p}_l}}{P(\hat{p}=\hat{p}_l)}$$

$$P(\hat{p} = \hat{p}_l) = P\left(\frac{1}{5}\sum_{i=1}^{5} X_i = \hat{p}_l\right) = P\left(\sum_{i=1}^{5} X_i = 5\hat{p}_l\right) = \binom{5}{5\hat{p}_l} p^{5\hat{p}_l}(1-p)^{5-5\hat{p}_l}$$

$$P(X_1 = k_1, ..., X_5 = k_5 | \hat{p} = \hat{p}_l) = \left[\binom{5}{5\hat{p}_l}\right]^{-1}$$

$\hat{p} = \frac{1}{5}\sum_{i=1}^{n} X_i$

EXAMPLE

$X_1, \dots, X_n$ with Bernoulli pmf, show $\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i$

$\hat{p}_e = \frac{1}{n}\sum_{i=1}^{n} k_i$, $k_i = 1, 0$

$$P(X_1 = k_1, \dots, X_n = k_n \mid \hat{p} = \hat{p}_e) = \frac{P(X_1, \dots, X_n)}{P(\hat{p} = \hat{p}_e)} = \frac{\prod_{i=1}^{n} p^{k_i}(1-p)^{1-k_i}}{P_{\hat{p}}(\hat{p} = \hat{p}_e)} =$$

$$= f(k_1, \dots, k_n) = f(\hat{p}_e) \text{ independent on } p$$

If $\hat{\theta}$ is sufficient statistic for θ, then any one-to-one function of $\hat{\theta}$ is also sufficient statistic for θ.
For example, $\hat{p}^* = n\hat{p} = n\frac{1}{n}\sum_{i=1}^{n} X_i = \sum_{i=1}^{n} X_i$ is also sufficient.

THEOREM (NEYMAN-FISHER FACTORIZATION CRITERIA)

Let $\hat{\theta} = u(X_1, ..., X_n)$ be a statistic based on the random sample $X_1, ..., X_n$. Then, $\hat{\theta}$ is a sufficient statistic for θ if and only if the discrete joint pmf $p_X(x_1, ..., x_n, \theta)$ (which depends on the parameter θ) can be factored into two non-negative functions.

$$L(\hat{\theta}) = p_X(x_1, ..., x_n, \theta) = g(\overbrace{u(x_1, ..., x_n)}^{\hat{\theta}_c}, \theta) \cdot h(x_1, ..., x_n), \;\; \textit{for all } x_1, ..., x_n,$$

where $g(\hat{\theta}, \theta)$ is a function only of $\hat{\theta}$ and θ and $h(x_1, ..., x_n)$ is a function of only $x_1, ..., x_n$ and not of θ.
(A similar statement holds for continuous case.)

PROOF.

1) Suppose that $\hat{\theta} = u(X_1, ..., X_n)$ is sufficient for θ. Then, $P(X_1 = x_1, ..., X_n = x_n | \hat{\theta}) = \underline{P(X_1 = x_1, ..., X_n = x_n)}$ if and only if $\hat{\theta} = u(x_1, ..., x_n) = u$ does not depend on θ. Hence,

$$\begin{aligned} p_X(x_1, ..., x_n, \theta) = & \quad P_\theta(X_1 = x_1, ..., X_n = x_n \cap \hat{\theta} = u)) \\ = & \quad P_\theta(X_1 = x_1, ..., X_n = x_n | \hat{\theta} = u) P_\theta(\hat{\theta} = u) \end{aligned}$$

$$p_X(x_1, ..., x_n, \theta) = h(x_1, ..., x_n) \cdot g(u, \theta)$$

PROOF.

2) (*Converse*) Assume, that

$$p_X(x_1,...,x_n,\theta) = h(x_1,...,x_n)\cdot g(u,\theta)$$

Define a set A_u :

$$A_u = \left\{(x_1,...,x_n) : \hat{\theta} = u(x_1,...,x_n) = u\right\}$$

11001
10101
11100

$$P_\theta(X_1 = x_1,...,X_n = x_n|\hat{\theta} = u) = \begin{cases} \frac{P_\theta(X_1=x_1,...,X_n=x_n\cap\hat{\theta}=u)}{P_\theta(\hat{\theta}=u)} & \text{if } (x_1,...,x_n)\in A_u \\ 0, & \text{if } (x_1,...,x_n)\notin A_u \end{cases}$$

$$P_\theta(X_1 = x_1,...,X_n = x_n|\hat{\theta} = u) = \frac{P_\theta(X_1 = x_1,...,X_n = x_n)}{P_\theta(\hat{\theta} = u)} = \frac{p_X(x_1,...,x_n,\theta)}{P_\theta(\hat{\theta} = u)}$$

$$= \frac{h(x_1,...,x_n)\cdot g(u,\theta)}{\sum\limits_{(x_1,...,x_n)\in A_u} g(\hat{\theta} = u,\theta)h(x_1,...,x_n)} = \frac{h(x_1,...,x_n)\cdot \cancel{g(u,\theta)}}{\cancel{g(\hat{\theta} = u,\theta)}\cdot \sum\limits_{(x_1,...,x_n)\in A_u} h(x_1,...,x_n)}$$

$$P_\theta(X_1 = x_1,...,X_n = x_n|\hat{\theta} = u) = \frac{h(x_1,...,x_n)}{\sum\limits_{(x_1,...,x_n)\in A_u} h(x_1,...,x_n)}$$

□

EXAMPLE

Let $X_1 = k_1, ..., X_n = k_n$ be a random sample of size n from the Poisson pdf,

$$p_X(k;\lambda) = \frac{e^{-\lambda}\lambda^k}{k!},\ k = 0, 1, 2, ...$$

Show that $\hat{\lambda} = \bar{X}$ is a sufficient statistic for λ.

Solution:

$$P(X_1 = k_1, \dots, X_n = k_n, \hat{\lambda} = \lambda_e) = h(x_1, \dots, x_n) \cdot g(\hat{\lambda}_e, \lambda)$$

$$L(\lambda) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{k_i}}{k_i!} = e^{-n\lambda}\lambda^{\sum_{i=1}^n k_i} \frac{1}{\prod_{i=1}^n k_i!} = e^{-n\lambda}\lambda^{n\frac{\sum_{i=1}^n k_i}{n}} \left(\prod_{i=1}^n k_i!\right)^{-1}$$

$$= \underbrace{e^{-n\lambda}\lambda^{n\hat{\lambda}_e}}_{g(\hat{\lambda}_e, \lambda)} \cdot \underbrace{\left(\prod_{i=1}^n (k_i)!\right)^{-1}}_{h(k_1,\dots,k_n)}$$

$\Rightarrow \hat{\lambda} = \bar{X}$ is a sufficient estimator for λ

$\square$

EXAMPLE

Let $X_1, ..., X_n$ denote a random sample from a geometric population with parameter p:

$$p_X(x;p) = p(1-p)^{x-1}, \; x = 1, 2, 3...$$

Show that $\bar{X}$ is sufficient for p.

Solution:

$$L(p) = \prod_{i=1}^{n} p\,(1-p)^{x_i - 1} = p^n (1-p)^{\sum_{i=1}^{n} x_i - n} =$$

$$= \underbrace{p^n (1-p)^{n\bar{x}-n}}_{g(\hat{p}_x, p)} \cdot \underbrace{1}_{h(x_1,...,x_n)=1}$$

$\Rightarrow \hat{p} = \bar{x}$ is a sufficient estimator for p

$\square$

THEOREM

The two statistics $\hat{\theta}_1$ and $\hat{\theta}_2$ are jointly sufficient for θ_1 and θ_2 if and only if the likelihood function can be factored into two non-negative functions,

$$f_X(X_1 = x_1, X_2 = x_2, ..., X_n = x_n; \theta_1, \theta_2) = g(\hat{\theta}_1, \hat{\theta}_2; \theta_1, \theta_2) h(X_1 = x_1, ..., X_n = x_n)$$

where $g(\hat{\theta}_1, \hat{\theta}_2; \theta_1, \theta_2)$ is only a function of $\hat{\theta}_1, \hat{\theta}_2$, θ_1, and θ_2, and $h(X_1 = x_1, ..., X_n = x_n)$ is free of θ_1 or θ_2.

EXAMPLE

Let $X_1, ..., X_n$ be a random sample from $N(\mu, \sigma^2)$.
i) If μ is unknown and $\sigma^2 = \sigma_0^2$ is known, show that $\bar{X}$ is a sufficient statistic for μ.
ii) If $\mu = \mu_0$ is known and σ^2 is unknown, show that $\sum_{i=1}^n (X_i - \mu_0)^2$ is sufficient for σ^2.
iii) If μ and σ^2 are both unknown, show that $\sum_{i=1}^n X_i$ and $\sum_{i=1}^n X_i^2$ are jointly sufficient for μ and σ^2.

Solution:

$$L(\mu,\sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2} = (2\pi)^{-\frac{n}{2}} \sigma^{-n} \cdot e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

$$= (2\pi)^{-\frac{n}{2}} \sigma^{-n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2} \cdot e^{-\frac{1}{2\sigma^2}(-2\mu\sum_{i=1}^{n} x_i)} \cdot e^{-\frac{1}{2\sigma^2} n\mu^2}$$

EXAMPLE

$$L(\mu,\sigma^2) = (2\pi)^{-\frac{n}{2}}\,\sigma^{-n}\, e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n x_i^2} \cdot e^{-\frac{1}{2\sigma^2}(-2\mu\sum_{i=1}^n x_i)} \cdot e^{-\frac{1}{2\sigma^2}n\mu^2}$$

1) $\bar{x}$ for μ, $\sigma^2 = \sigma_0^2$

$$L(\mu,\sigma_0^2) = \underbrace{e^{-\frac{1}{2\sigma_0^2}(-2\mu n\bar{x})} \cdot e^{-\frac{1}{2\sigma^2}n\mu^2}}_{g(\bar{x},\mu)} \cdot \underbrace{(2\pi)^{-\frac{n}{2}} \cdot \sigma_0^{-n} \cdot e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n x_i^2}}_{h(x_1,\dots,x_n)}$$

$\bar{x}$ is sufficient for μ

2) $\sum_{i=1}^n (x_i-\mu_0)^2$ for σ^2, $\mu=\mu_0$

$$L(\mu_0,\sigma^2) = \underbrace{e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu_0)^2}\,\sigma^{-n}}_{g(\hat{\sigma}^2,\sigma^2)} \cdot \underbrace{(2\pi)^{-\frac{n}{2}}}_{h(x_1,\dots,x_n)}$$

$$L(\mu,\sigma^2) = (2\pi)^{-\frac{n}{2}} \sigma^{-n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2} e^{-\frac{1}{2\sigma^2}(-2\mu\sum_{i=1}^{n} x_i)} e^{-\frac{1}{2\sigma^2} n\mu^2}$$

3) $\sum_{i=1}^{n} x_i$, $\sum_{i=1}^{n} x_i^2$ jointly sufficient for μ, σ^2

$$L(\mu,\sigma^2) = \underbrace{\sigma^{-n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i \cdot (-2\mu)} e^{-\frac{1}{2\sigma^2} n\mu^2}}_{g(\sum x_i, \sum x_i^2; \mu, \sigma^2)} \cdot \underbrace{(2\pi)^{-\frac{n}{2}}}_{h(x_1,\ldots,x_n)}$$

$\sum_{i=1}^{n} x_i$, $\sum_{i=1}^{n} x_i^2$ are jointly sufficient

The Sufficiency Principle:
Any inference procedure should depend on the data only through sufficient statistics.

EXAMPLE

The estimator $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ is biased estimator for σ^2:

$$E(\hat{\sigma}^2) = \frac{n-1}{n}\sigma^2.$$

EXAMPLE

The estimator $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ is biased estimator for σ^2:

$$E(\hat{\sigma}^2) = \frac{n-1}{n}\sigma^2.$$

However, if the sample size $n \to \infty$

$$E(\hat{\sigma}^2) \to \sigma^2,$$

hence, $\hat{\sigma}^2$ is *asymptotically unbiased.*

DEFINITION

A sequence of random variables $X_1, X_2, \ldots$ ***converges in probability*** *to a random variable X if for every $\varepsilon > 0$*

$$\lim_{n\to\infty} P\left(|X_n - X| < \varepsilon\right) = 1 \;\Leftrightarrow\; \lim_{n\to\infty} P\left(|X_n - X| \geq \varepsilon\right) = 0$$

Denoted as $X_n \xrightarrow{p} X$

DEFINITION

*A sequence of random variables $X_1, X_2, \ldots$ **converges in probability** to a random variable X if for every $\varepsilon > 0$*

$$\lim_{n\to\infty} P\left(|X_n - X| < \varepsilon\right) = 1 \;\Leftrightarrow\; \lim_{n\to\infty} P\left(|X_n - X| \geq \varepsilon\right) = 0$$

Denoted as $X_n \xrightarrow{p} X$

DEFINITION

*A sequence $\hat{\theta}_n = u(X_1, X_2, \ldots, X_n)$, $n = 1, 2, \ldots$ is said to be **consistent sequence of estimators** for θ if it converges in probability to θ, i. e. for $\varepsilon > 0$*

$$\lim_{n\to\infty} P\left(|\hat{\theta}_n - \theta| < \varepsilon\right) = 1$$

DEFINITION

A sequence of random variables $X_1, X_2, ...$ ***converges in probability*** *to a random variable X if for every $\varepsilon > 0$*

$$\lim_{n\to\infty} P\left(|X_n - X| < \varepsilon\right) = 1 \;\Leftrightarrow\; \lim_{n\to\infty} P\left(|X_n - X| \geq \varepsilon\right) = 0$$

Denoted as $X_n \xrightarrow{p} X$

DEFINITION

A sequence $\hat{\theta}_n = u(X_1, X_2, ..., X_n)$, $n = 1, 2, ...$ is said to be ***consistent sequence of estimators*** *for θ if it converges in probability to θ, i. e. for $\varepsilon > 0$*

$$\lim_{n\to\infty} P\left(|\hat{\theta}_n - \theta| < \varepsilon\right) = 1$$

Consistency means that the probability of our estimator being within some small ε-interval of θ can be made as close to one as we like by making the sample size n sufficiently large.

EXAMPLE

Suppose that $X_1, X_2, ..., X_n$ is a random sample of size n from discrete pmf $p_X(x, \mu)$, where $E(X) = \mu$, $\mathrm{Var}(X) = \sigma^2 < \infty$. Let $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}_n$.
Is $\hat{\mu}$ a consistent estimator for μ?

EXAMPLE

Suppose that $X_1, X_2, ..., X_n$ is a random sample of size n from discrete pmf $p_X(x, \mu)$, where $E(X) = \mu$, $\text{Var}(X) = \sigma^2 < \infty$. Let $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}_n$.
Is $\hat{\mu}$ a consistent estimator for μ?

Solution.

(Markov's inequality)
Let X be a random variable and let $g(\cdot)$ be a non-negative function. Then, for any $k > 0$,

$$P(g(X) \geq k) \leq \frac{E(g(X))}{k}$$

EXAMPLE

Suppose that $X_1, X_2, ..., X_n$ is a random sample of size n from discrete pmf $p_X(x, \mu)$, where $E(X) = \mu$, $\text{Var}(X) = \sigma^2 < \infty$. Let $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}_n$.

$E(\bar{X}_n) = \mu$

$Var(\bar{X}_n) = \frac{\sigma^2}{n}$

Is $\hat{\mu}$ a consistent estimator for μ?

Solution.

(Markov's inequality)
Let X be a random variable and let $g(\cdot)$ be a non-negative function. Then, for any $k > 0$,

$$P(g(X) \geq k) \leq \frac{E(g(X))}{k}$$

$$P\left(|\bar{X}_n - \mu| < \varepsilon\right) = 1 - P\left(|\bar{X}_n - \mu| \geq \varepsilon\right) = 1 - P\Big(\underbrace{(\bar{X}_n - \mu)^2}_{= g(\bar{x}_n - \mu)} \geq \underbrace{\varepsilon^2}_{= k}\Big) \leftarrow \leq \frac{E(g(\cdot))}{k}$$

$$\geq 1 - \frac{E((\bar{X}_n - \mu)^2)}{\varepsilon^2} = 1 - \frac{\text{Var}(\bar{X}_n)}{\varepsilon^2} = 1 - \underbrace{\frac{\sigma^2}{n\varepsilon^2}}_{\xrightarrow{n\to\infty} 0} \to 1 \text{ as } n \to \infty$$

$1 \leq \lim_{n\to\infty} P(|\bar{x}_n - \mu| < \varepsilon) \leq 1$

$$P\left(|\bar{X}_n - \mu| < \varepsilon\right) \to 1 \text{ as } n \to \infty$$

THEOREM (WEAK LAW OF LARGE NUMBERS)

Let $X_1, X_2, \ldots$ be i.i.d random variables with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2 < \infty$. Define $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then for every $\varepsilon > 0$,

$$\lim_{n\to\infty} P\left(|\bar{X}_n - \mu| < \varepsilon\right) = 1)$$

that is, $\bar{X}_n$ converges in probability to μ

1. State main properties of point estimators;
2. Define the notions of bias and unbiasedness;
3. Define the notions of efficiency and relative efficiency of estimators;
4. What is the Mean Square Error, state the relation between MSE of an estimator, its variance and bias.
5. Define the notion of sufficiency of an estimator;
6. Define the notion of consistency of an estimator.