Mathematical Statistics

How to get information from the experiment?

Sergey Titov, Konstantin Grotov

Neapolis University Pafos, Pahpos

07.07.2023

Outline

Law of Large Numbers

Theorem (Markov's Inequality)

Let X be a random variable and let $g(x)$ be a nonnegative function. Then, for any $r > 0$

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

Law of Large Numbers

Theorem (Chebyshev's Inequality)

if the variance is small then X is unlikely to be too far from the mean

$$P(|X - \mu| \geq c) \leq \frac{\sigma^2}{c^2}$$

Law of Large Numbers

Theorem (Weak Law of Large Numbers)

Let $X_1, \dots, X_n$ be iid random variables with $\mathbb{E}X_i = \mu$ and $\operatorname{Var} X_i = \sigma^2 < \infty$. Define $\bar{X}_n = (1/n)\sum_{i=1}^{n} X_i$. Then for every $\epsilon > 0$,

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

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

The law of large numbers states that the sample mean converges to the distribution mean as the sample size increases, and is one of the fundamental theorems of probability. There are different versions of the law, depending on the mode of convergence.

Law of Large Numbers

Proof.

$$P(|\bar{X}_n - \mu| \geq \epsilon) = P((\bar{X}_n - \mu)^2 \geq \epsilon^2) \leq \frac{\mathbb{E}(\bar{X}_n - \mu)^2}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}$$

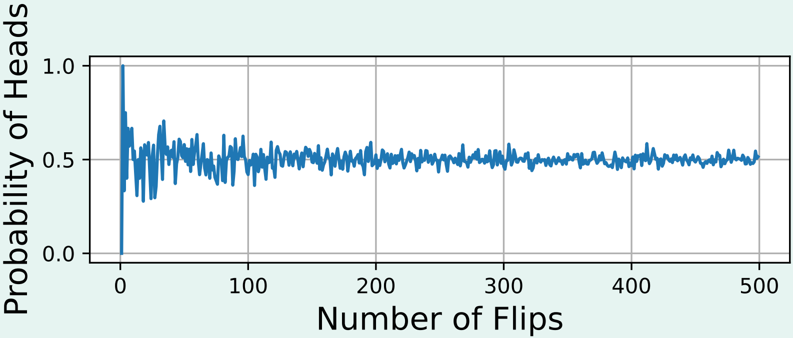
Hence,

$$P(|\bar{X}_n - \mu| < \epsilon) = 1 - P(|\bar{X}_n - \mu| \geq \epsilon) \geq 1 - \frac{\sigma^2}{n\epsilon^2} \to 1, \quad \text{as } n \to \infty$$

$\square$

Law of Large Numbers

Example (Tossing a coin)



Central Limit Theorem

Theorem (Central Limit Theorem)

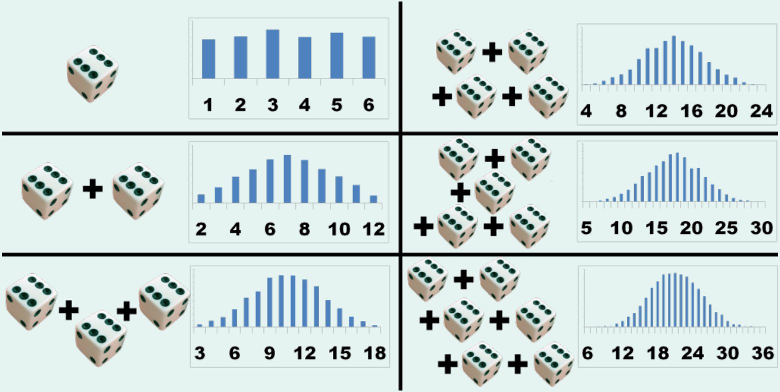
Let $X_1, \ldots, X_n$ be a sequence of iid random variables. Let $\mathbb{E}X_i = \mu$ and $\operatorname{Var} X_i = \sigma^2 > 0$ and finite. Define $\bar{X}_n = (1/n)\sum_{i=1}^{n} X_i$. Let $G_n(x)$ denote the cdf of $\sqrt{n}(\bar{X}_n - \mu)/\sigma$. Then for any $-\infty < x < \infty$,

$$\lim_{n\to\infty} G_n(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$$

that is, $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ has a limiting standard normal distribution.

Central Limit Theorem

Example



Central Limit Theorem

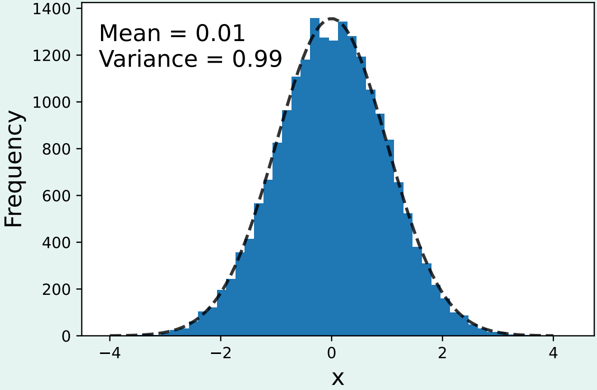
Theorem

Roughly, the central limit theorem states that the distribution of the sum (or average) of a large number of independent, identically distributed variables will be approximately normal, regardless of the underlying distribution.

Central Limit Theorem

Example

Let $X_i \sim U(0, 1)$



Random Sample

Definition

The random variables $X_1, \ldots, X_n$ are called a random sample of size n from the population $f(x)$ if $X_1, \ldots, X_n$ are mutually independent random variables and the marginal pdf or pmf of each X_i is the same function $f(x)$. Alternatively, $X_1, \ldots, X_n$ are called independent and identically distributed (iid) random variables with pdf or pmf $f(x)$. Another notation is $X_{[n]} = [X_1, \ldots, X_n]$

Definition

The list of $[x_1, \ldots, x_n] = x_{[n]}$ called a sample of size n of realization of **random element** X.

Population of Sample

for sample $x_{[n]}$ assume empirical distribution:

	x_1	x_2	$\dots$	x_n
$\mathbb{P}$	$\frac{1}{n}$	$\frac{1}{n}$	$\dots$	$\frac{1}{n}$

This is discrete distribution. Let A be an event, then

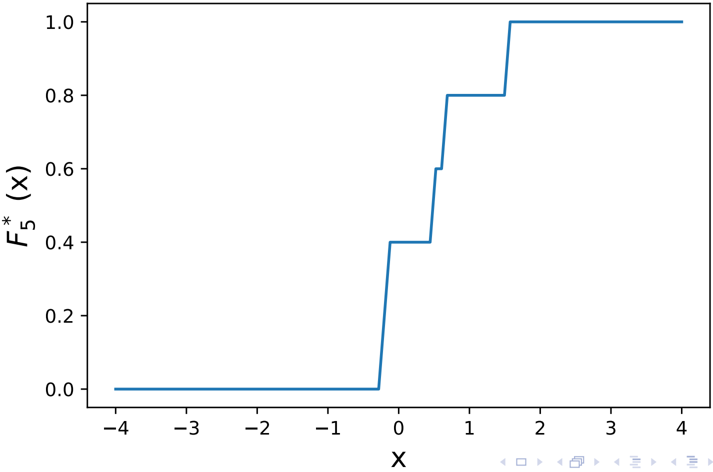
$$P(A) = \frac{1}{n}\sum_{i=1}^{n}[x_i \in A]$$

Definition

Empirical random variable X^* – r.v. that has distribution function F_n^*

Population of sample

For sample $x_{[n]}$ assume $F_n^*(x) = \frac{1}{n}\sum_{i=1}^{n}[x_i < x]$



Why Sample?

Theorem (Glivenko-Cantelli)

$$\sup_{x\in\mathbb{R}} |F_n^*(x) - F_X(x)| = \|F_n^* - F_X\|_\infty \xrightarrow{a.s.} 0$$

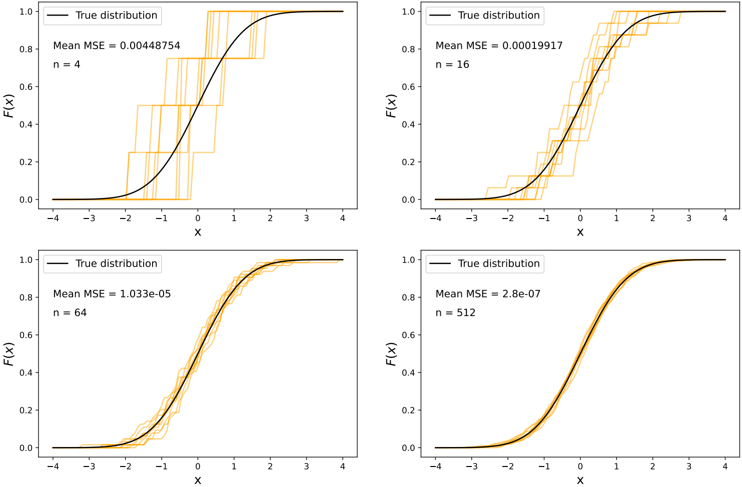
Theorem (Kolmogorov)

If F_X is continuous, then

$$\sqrt{n}\|F_n^* - F_X\|_\infty \xrightarrow{d} K$$

Where K - Kolmogorov's distribution

Why Sample?


True distribution
Mean MSE = 0.00448754
n = 4
True distribution
Mean MSE = 0.00019917
n = 16
True distribution
Mean MSE = 1.033e-05
n = 64
True distribution
Mean MSE = 2.8e-07
n = 512
F(x)
x

Statistics and Estimators

Definition

Let $X_1, \ldots, X_n$ be a random sample of size n from a population and let $T = (x_1, \ldots x_n)$ be a real-valued or vector-valued function whose domain includes the sample space of $(X_i, \ldots X_n)$. Then the random variable or random vector $Y = T(X_1, \ldots, X_n)$ is called a statistic. The probability function of a statistic Y is called the sampling distribution of Y.

Definition

A point estimator is any function $W(X_1, \ldots X_n)$ of a sample; that is, any statistic is a point estimator.

Estimate $= W(x_1, \ldots x_n)$

A Note About Statistic

Statistic is a function of sample! Statistic NOT a function of a unknown parameter.

$$T(x_1, \ldots, x_n) = x_1 + x_2 + \cdots + x_n - \text{ Statistic}$$

$$T(x_1, \ldots, x_n) = \mathbb{E}X_1 - \text{ NOT statistic}$$

$$\mathbb{E}X_i = \mu \quad T(x_1, \ldots, x_n) = \sum (x_i - \mu)^2 - \text{depends on } \mu \text{ status}$$

Statistics

Definition

The sample mean is the arithmetic average of the values in a random sample. It is usually denoted by

$$\bar{X} = \frac{X_1 + \cdots + X_n}{n} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Definition

The sample variance (or corrected sample variance) is the statistic defined by

$$S^2 = \frac{n}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

The Order Statistics

Definition

The order statistics of a random sample $X_1, \ldots, X_n$ are the sample values placed in ascending order. They are denoted by $X_{(1)}, \ldots X_{(n)}$

The order statistics are random variables that satisfy $X_{(1)} \leq \cdots \leq X_{(n)}$. In particular,

$$
\begin{cases}
X_{(1)} = \min\limits_{1 \leq 1 \leq n} X_i, \\
X_{(2)} = \text{second smallest } X_i \\
\vdots \\
X_{(1)} = \max\limits_{1 \leq 1 \leq n} X_i,
\end{cases}
$$

Quantiles, Quartiles, Median

Definition (α-Quantile)

A number x_α that $\mathbb{P}(X \leq x_\alpha) \geq \alpha$ and $\mathbb{P}(X \geq x_\alpha) \geq 1 - \alpha$. If continuous we can write $P(X < x_\alpha) = \alpha$

Definition (Sample α-Quantile)

- if $\frac{k}{n} < \alpha < \frac{k+1}{n}$, then $x^*_\alpha = x_{(k+1)}$
- if $\alpha = \frac{k}{n}$, then x^*_α any number in $[x_{(k)}, x_{(k+1)}]$

Definition (Sample Quartile)

$$Q_1 = x^*_{0.25} \quad Q_2 = x^*_{0.5} \quad Q_3 = x^*_{0.75}$$

Quantiles, Quartiles, Median

Definition (Sample Median m^*)

- if $n = 2k + 1$, then $m^* = x_{(k+1)}$
- if $n = 2k$, then $m^* = \frac{1}{2}(x_{(k)} + x_{(k+1)})$

Method of Moments

Let's remember definitions of the moments:

Definition

$$\mu_1 = \frac{(\sum x)}{n}$$
$$\mu_2 = \frac{\sum (x-\mu)^2}{n}$$
$$\mu_3 = \frac{1}{n}\frac{\sum (x-\mu)^3}{\sigma^3}$$
$$\mu_4 = \frac{1}{n}\frac{\sum (x-\mu)^4}{\sigma^4}$$

Method of Moments

Let's assume that we have a $X_1, \ldots, X_n$ sample form population with pdf $f(x|\theta_1, \ldots, \theta_k)$
The idea of the method of moments is to equate the first k sample moments to corresponding k population moments and then solve the system of equation.

Method of Moments

Example

Suppose $X_1, \ldots, X_n$ are iid $n(\theta, \sigma^2)$. $\theta_1 = \theta$, $\theta_2 = \sigma^2$. We have $m_1 = \bar{X}, m_2 = (1/n)\sum X_i^2, \mu_1 = \theta, \mu_2 = \theta^2 + \sigma^2$, so we must solve

$$\bar{X} = \theta, \quad \frac{1}{n}\sum X_i^2 = \theta^2 + \sigma^2$$

Solving this system for θ, σ^2 yields method of moments estimators

$$\tilde{\theta} = \bar{X}, \quad \sigma^2 = \frac{1}{n}\sum X_i^2 - \bar{X}^2 = \frac{1}{n}\sum(X_i - \bar{X})^2$$