Applied Statistics
for Computer Science BSc, Exam

Probability Theory and Mathematical Statistics
for Computer Science Engineering BSc, Term grade

István Fazekas
University of Debrecen

2020/21 fall

This work was supported by the construction EFOP-3.4.3-16-2016-00021. The project was supported by the European Union, co-financed by the European Social Fund.

Main topics

1. Probability theory
2. Statistics

Mathematical tools: combinatorics, calculus
Computer tool: Matlab
Book:
Yates, Goodman:
Probability and Stochastic Processes: A Friendly Introduction for Electrical and Computer Engineers

Lecture 9

Limit theorems

Markov's inequality

Theorem. Let $Y \geq 0$ be a random variable and let $\delta > 0$ be a number. Then

$$P\big(Y \geq \delta\big) \leq \mathbb{E}(Y)/\delta.$$

Proof. We prove for the absolute continuous case. As $Y \geq 0$, so for its PDF we have $f(x) = 0$ as $x < 0$.

$$\mathbb{E}Y = \int_{-\infty}^{\infty} xf(x)\,dx = \int_{0}^{\infty} xf(x)\,dx \geq \int_{\delta}^{\infty} xf(x)\,dx \geq$$

$$\geq \delta \int_{\delta}^{\infty} f(x)\,dx = \delta P(Y \geq \delta).$$

Exercise. Prove Markov's inequality for discrete random variables.

Chebyshev's inequality

Theorem. Assume that the variance of X is finite. Then for any $\varepsilon > 0$ we have

$$P\big(|X - \mathbb{E}X| \geq \varepsilon\big) \leq \mathrm{Var}(X)/\varepsilon^2.$$

Proof. Let $Y = (X - \mathbb{E}X)^2$ and $\delta = \varepsilon^2$. Apply Markov's inequality.

Weak law of large numbers, WLLN

Let $X_1, X_2, \ldots$ be random variables.

$$S_n = X_1 + \cdots + X_n, \quad n = 1, 2, \ldots$$

will denote the partial sums. WLLN's claim the stochastic convergence of S_n/n.

Definition. We say that the sequence $Y_1, Y_2, \ldots$ converges stochastically (in probability) to Y if $\forall\, \varepsilon > 0$

$$\lim_{n\to\infty} P\big(|Y_n - Y| > \varepsilon\big) = 0.$$

Notation: $P - \lim_{n\to\infty} Y_n = Y$

The weak law of large numbers

Theorem. Let $X_1, X_2, \dots$ be pairwise independent and identically distributed random variables with $\mathbb{E}X_i^2 < \infty$. Let $m = \mathbb{E}X_i$ be their expectation. Then

$$P - \lim_{n\to\infty} \frac{S_n}{n} = m.$$

Proof. By Chebyshev's inequality, for any $\varepsilon > 0$

$$P\left(\left|\frac{S_n}{n} - m\right| > \varepsilon\right) = P\left(\left|\frac{S_n}{n} - \mathbb{E}\left(\frac{S_n}{n}\right)\right| > \varepsilon\right) \leq$$

$$\leq \frac{1}{\varepsilon^2}\mathrm{Var}\left(\frac{S_n}{n}\right) = \frac{1}{\varepsilon^2 n^2}\sum_{i=1}^{n}\mathrm{Var}X_i = \frac{1}{\varepsilon^2 n}\mathrm{Var}X_1 \to 0$$

as $n \to \infty$. We applied that the variance is additive for independent r.v.'s.

The weak law of large numbers

The meaning of the WLLN is the following.
$X_1, X_2, \ldots$ are independent observations of X.
So S_n/n is the average of the observations. m is the theoretical mean.
So the the average of the observations converges to the theoretical mean.
The meaning of the stochastic convergence in the WLLN is the following.
For large n with large probability S_n/n is close to m.

Remark. Khintchine proved that the above WLLN if true if instead of $\mathbb{E}X_i^2 < \infty$ we assume the weaker condition $\mathbb{E}|X_i| < \infty$.

Bernoulli's weak law of large numbers

Consider an experiment, and in the experiment an event A with probability p.
Repeat the experiment n times independently.
Let $X_i = 1$ if A occurs in the ith repetition of the experiment, and $X_i = 0$ otherwise.
Then X_i has Bernoulli distribution: $P(X_i = 1) = p$, $P(X_i = 0) = 1 - p$.
So $\mathbb{E}X_i = p$.
Moreover, $k_A = X_1 + \cdots + X_n$ is the frequency of A.
As $X_1, \ldots, X_n$ are independent, by the WLLN, we have

$$P - \lim_{n\to\infty} \frac{k_A}{n} = p.$$

So the relative frequency of an event converges to its probability.

Kolmogorov's strong law of large numbers, SLLN

Theorem. Let $X_1, X_2, \ldots$ be independent and identically distributed random variables with $\mathbb{E}|X_i| < \infty$. Let $m = \mathbb{E}X_i$ be their expectation. Then

$$\lim_{n\to\infty} \frac{S_n}{n} = m \quad \text{almost surely.}$$

The meaning of the above limit is the following

$$P\left(\lim_{n\to\infty} \frac{S_n}{n} = m\right) = 1 = 100\%$$

Etemadi proved that the above result is true if we assume only pairwise independence.

Visualization of the SLLN

We generated 200 random variables with mean zero and calculated their average.

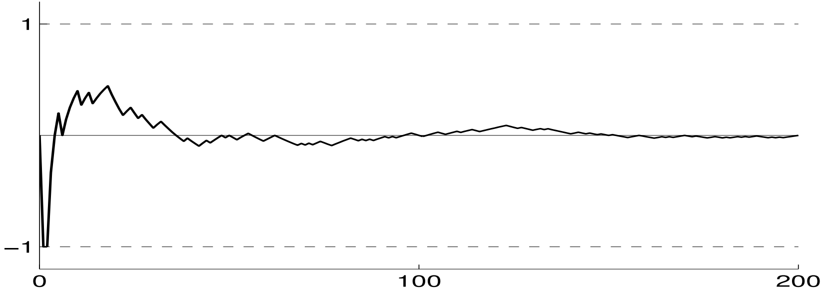

Figure: $\frac{S_1}{1}, \frac{S_2}{2}, \ldots, \frac{S_{199}}{199}, \frac{S_{200}}{200}$

Stochastic simulation (Monte Carlo methods)

These are numerical methods based on the SLLN.

Example. Let $f : [0,1] \longrightarrow [0,1]$.

Calculate numerically $\int_0^1 f(x)\,dx$. Let $X_1, \eta_1, X_2, \eta_2, \dots$ be independent and uniformly distributed on $[0,1]$.

Let

$$\varrho_i = \begin{cases} 1, & \text{if} \quad f(X_i) > \eta_i \\ 0, & \text{if} \quad f(X_i) \le \eta_i \end{cases} .$$

Then $\varrho_1, \varrho_2, \dots$ are independent identically distributed and $\mathbb{E}\varrho_i = P\left(f(X_i) > \eta_i\right) = \int_0^1 f(x)\,dx$. Then, by the SLLN,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \varrho_i = \int_0^1 f(x)\,dx$$

almost surely. The left hand side of (12) can be calculated using random number generators.

Stochastic simulation (Monte Carlo methods)

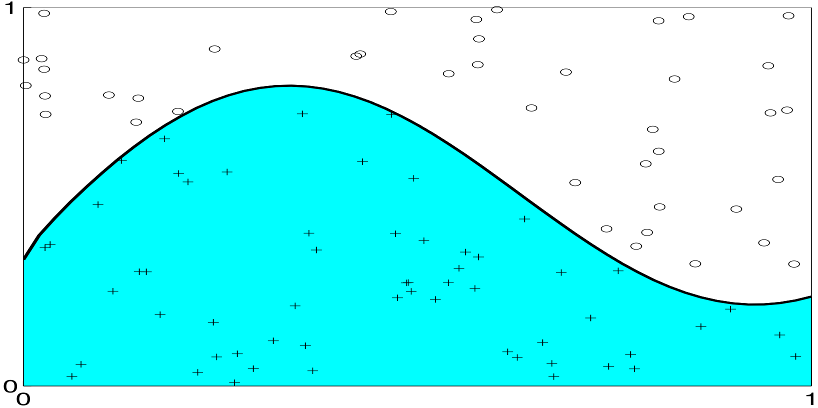

Figure: Calculating the integral by Monte Carlo method

Central limit theorems, CLT

The local version of the de Moivre-Laplace theorem

Let S_n be a r.v. with binomial distribution:

$$P_n(k) = P(S_n = k) = \binom{n}{k} p^k \, q^{n-k}, \quad k = 0, 1, \ldots, n \, .$$

Let $0 < p < 1$. Then

$$P_n(k) \sim \frac{1}{\sqrt{2\pi npq}} \exp\left\{ -\frac{(k-np)^2}{2npq} \right\} . \tag{1}$$

More precisely

$$\sup_{\{k \,:\, |k-np| \le g(n)\}} \left| \frac{P_n(k)}{\frac{1}{\sqrt{2\pi npq}} \exp\left\{ -\frac{(k-np)^2}{2npq} \right\}} - 1 \right| \to 0,$$

if $n \to \infty$, and $\lim_{n\to\infty} g(n)/(npq)^{2/3} = 0$.

Visualization of the local de Moivre-Laplace theorem

On the right hand side of (1) there is the PDF of $\mathcal{N}(np, npq)$ at value k.

So the binomial distribution can be approximated by the normal density.

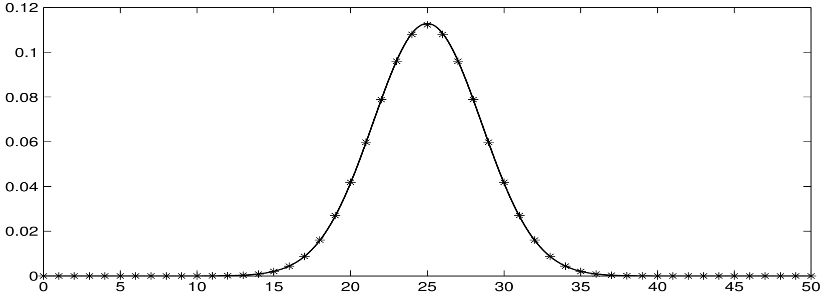

Figure: The binomial distribution (denoted by $*$) and the normal PDF

The integral version of the de Moivre-Laplace theorem

Theorem. Let S_n be a r.v. with binomial distribution.
Let $\Phi(x)$ denote the standard normal CDF.
Then

$$\lim_{n\to\infty} \sup_{-\infty\leq x\leq\infty} \left| P\left(\frac{S_n - np}{\sqrt{npq}} < x\right) - \Phi(x)\right| = 0. \qquad (2)$$

So the CDF of a standardized binomial random variable converges to the standard normal CDF.

Visualization of the integral version of the de Moivre-Laplace theorem

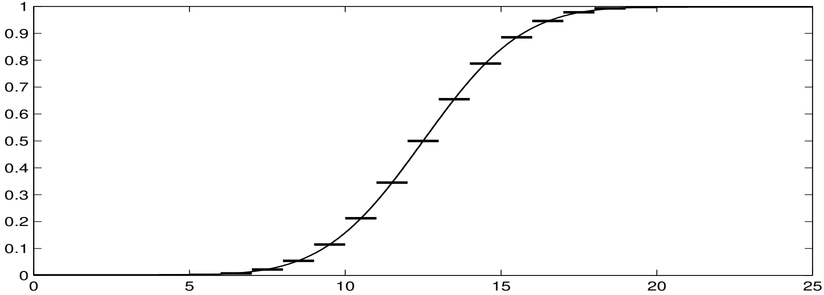

Figure: The step function is a binomial CDF, the continuous function is the CDF of the normal distribution having the same expectation and variance as those of the binomial one

The general form of the CLT

Theorem. Let $X_1, X_2, \dots$ be independent identically distributed random variables.
Let $S_n = X_1 + \cdots + X_n$.
Assume that $\sigma^2 = \mathrm{Var} X_1$ is finite and positive.
Let $m = \mathbb{E} X_1$.
Then

$$\boxed{\lim_{n \to \infty} P\left(\frac{S_n - nm}{\sqrt{n}\sigma} < x \right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt,}$$

$\forall x \in \mathbb{R}$.
That is the CDF of the standardized S_n converges to the standard normal CDF as $n \to \infty$.

Visualization of the CLT

Let $X_1, X_2, \ldots$ be independent random variables with $P(X_i = 1) = 0.5$ and $P(X_i = -1) = 0.5$. Let $S_n = X_1 + \cdots + X_n$. Then $S_1, S_2, \ldots$ is called symmetric random walk (because each second we make one step either to the right or to the left direction). As X_i has expectation 0 and variance 1, so the standardized random walk is $S_n/\sqrt{n}$.

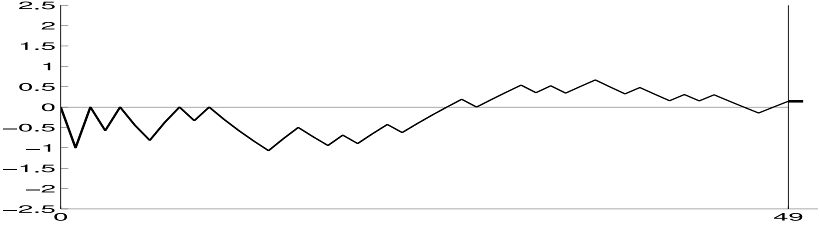

Figure: The standardized random walk $S_1/\sqrt{1}, S_2/\sqrt{2}, \ldots, S_{49}/\sqrt{49}$.

Visualization of the CLT

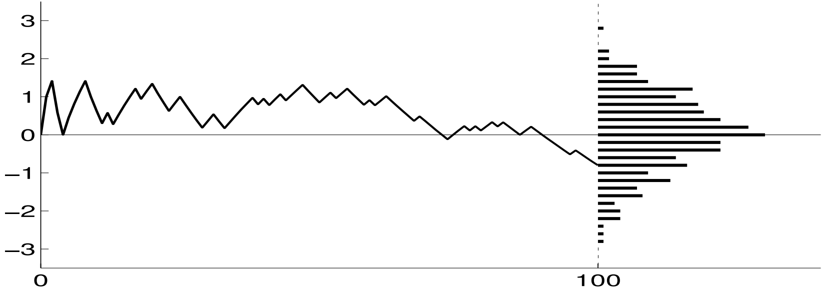

Figure: The standardized random walk $S_1/\sqrt{1}, S_2/\sqrt{2}, \ldots, S_{100}/\sqrt{100}$.

The histogram on the right hand side of the figure shows the results of 300 repetitions of the 100-step random walk. We can see that the histogram is close to the bell shaped curve of the normal PDF.