NMAI059 Probability and statistics 1
Class 10

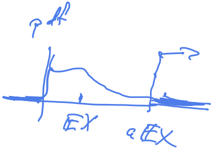
Robert Šámal

Inequalities we know from the last time (and one more)

- Markov:

rand. var.

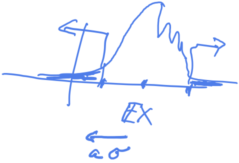
$$X \geq 0 \Rightarrow P(X \geq a\mathbb{E}(X)) \leq \frac{1}{a}$$



- Chebyshev

any r.v. with E(X) & var(X)

$$P(|X - \mathbb{E}(X)| \geq a\sigma_X) \leq \frac{1}{a^2}$$



- Chernoff ($\sigma_X = \sqrt{n}$)

$$X = \sum_{i=1}^{n} X_i, X_i = \pm 1 \Rightarrow P(|X - \mathbb{E}(X)| > a\sigma_X) \leq 2e^{-a^2/2}$$

ind. r.v.

Overview

Limit theorems – approximation

Statistics – an introduction

Strong law of large numbers

Theorem

Let $X_1, \dots, X_n$ be i.i.d. with expectation μ and variance σ^2. Let $S_n = (X_1 + \cdots + X_n)/n$ be the sample mean. Then we have

$$\lim_{n\to\infty} S_n = \mu \quad \textit{almost surely (i.e. with probability 1).}$$

We say that sequence S_n converges to μ almost surely. and write $S_n \xrightarrow{a.s.} \mu$).

10 rolls

10^6 rolls

80% of 1

16.6% of 1

$10+10^6$

Monte Carlo integration

How to compute $\int_{x \in A} g(x)dx$?

In particular

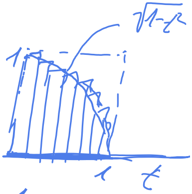
$$g(x) = \begin{cases} 1 & \text{for } x \in S \\ 0 & \text{otherwise} \end{cases}$$

... area of a circle

$\frac{\pi}{4}$ unit circle

S

$A = [-1,1]^2$



$\int_0^1 \sqrt{1-t^2}\, dt$

$A = [0,1]$

$\frac{\pi}{4}$

$\omega_1, \omega_2, \ldots$

ω = sequence of random points in A

$X_n(\omega) = g(\omega_n)$

$$\mathbb{E} X_n = \frac{\int_{x \in A} g(x)}{\text{area of } A} = \mu$$

Law of Large Numbers $\to$ $S_n \to \mu$ a.s.

Weak law of large numbers

fix ε — error bound

Theorem

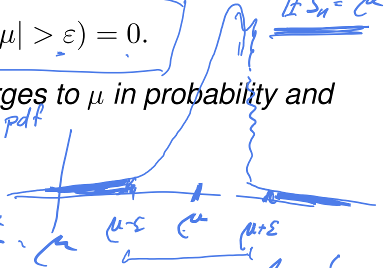
Let $X_1, \dots, X_n$ be i.i.d. with expectation μ and variance σ^2. Let $S_n = (X_1 + \cdots + X_n)/n$ be the sample mean. Then for every $\varepsilon > 0$ we have

$$\boxed{\lim_{n\to\infty} P(|S_n - \mu| > \varepsilon) = 0.}$$

Chebyshev

$\mathbb{E}S_n = \mu$

We say that sequence S_n converges to μ in probability and write $S_n \xrightarrow{P} \mu$).



Proof

$$\mathbb{E}(S_n) = E\,\frac{X_1+\dots+X_n}{n} = \frac{\overbrace{EX_1+\dots+EX_n}^{n}}{n} = \frac{n\mu}{n} = \mu$$

indep.!

$$\mathrm{var}(S_n) = \mathrm{var}\left(\frac{X_1+\dots+X_n}{n}\right) = \frac{1}{n^2}\,\mathrm{var}(X_1+\dots+X_n) = \frac{1}{n^2}\left(\mathrm{var}(X_1)+\dots+\mathrm{var}(X_n)\right) = \frac{\sigma^2}{n}$$

$$P(|S_n - \mu| > \underbrace{a\,\sigma_{S_n}}_{\varepsilon}) \le \frac{1}{a^2} = \frac{\sigma^2}{\varepsilon^2 \cdot n} \xrightarrow[n\to\infty]{} 0$$

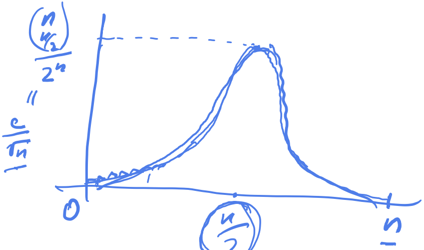
$$\Rightarrow a = \frac{\varepsilon}{\sigma}\sqrt{n}$$

$$\sigma_{S_n} = \sqrt{\sigma^2/n} = \frac{\sigma}{\sqrt{n}}$$

Law of Large numbers → Central Limit Theorem

$X_1, \dots, X_n \sim \mathrm{Bern}(\tfrac{1}{2})$

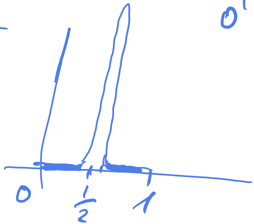
$X = X_1 + \dots + X_n \sim \mathrm{Bin}(n, \tfrac{1}{2})$

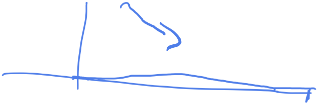


LLN: $S_n = \frac{X}{n}$

$E S_n \to \frac{1}{2}$

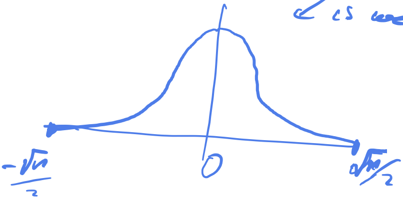
$S_n \to \frac{1}{2}$





this shape is well-centered

$$Y_n = \frac{X - \frac{n}{2}}{\sqrt{n}}$$



Central Limit Theorem

$E Y_n = \frac{E(X_1+\dots+X_n) - n\mu}{\sqrt{n}\,\sigma} = 0$

$\mathrm{var}(Y_n) = \frac{\mathrm{var}(X_1+\dots+X_n)}{n\sigma^2} = 1$

Theorem

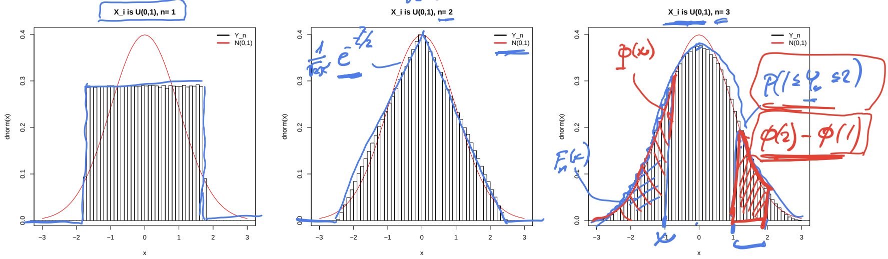
Let $X_1, \dots, X_n$ be i.i.d. with expectation μ and variance σ^2. Put $Y_n := ((X_1 + \dots + X_n) - n\mu)/(\sqrt{n}\cdot\sigma)$.
Then $Y_n \xrightarrow{d} N(0,1)$. This means, that if F_n is the cdf of Y_n, then

$$\lim_{n\to\infty} F_n(x) = \Phi(x) \quad \textit{for every } x \in \mathbb{R}.$$

We say that the sequence Y_n converges to $N(0,1)$ in distribution.

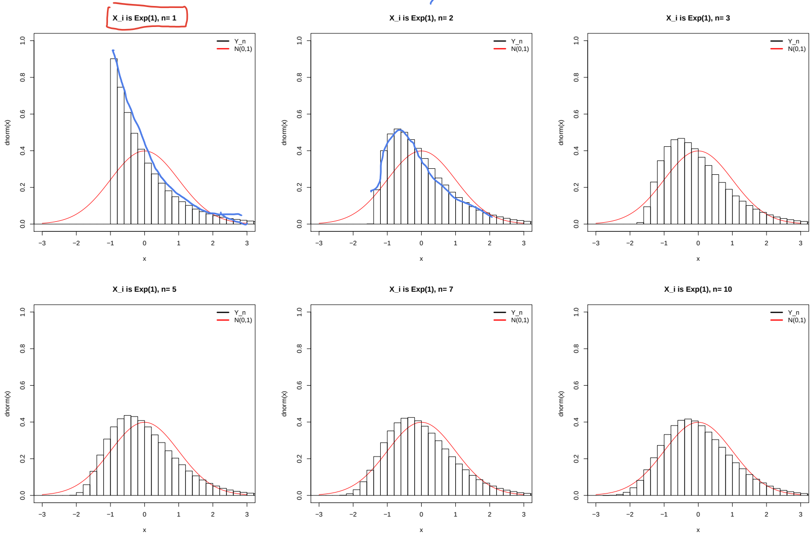
pdf

$\frac{X_1+X_2-2\mu}{\sqrt{2}\,\sigma}$



CLT another illustration

Gamma distr.



Bonus: Moment generating function

Definition

For a random variable X we let

$$M_X(t) = \mathbb{E}\big(e^{tX}\big).$$

Function $M_X(t)$ is called the moment generating function.

- $M_X(t) = \sum_{n=0}^{\infty} \mathbb{E}(X^n)\frac{t^n}{n!}$.
- $M_{Bern(p)}(t) = p \cdot e^t + (1-p)$.
- $M_{X+Y}(t) = M_X(t)M_Y(t)$, jsou-li X, Y n.n.v.
- $M_{Bin(n,p)} = (pe^t + 1 - p)^n$
- $M_{N(0,1)} = e^{t^2/2}$
- $M_{Exp(\lambda)} = \frac{1}{1-t/\lambda}$
- If $M_X(t) = M_Y(t)$ on $(-a, a)$ for some $a > 0$, then $X = Y$ a.s.

Overview

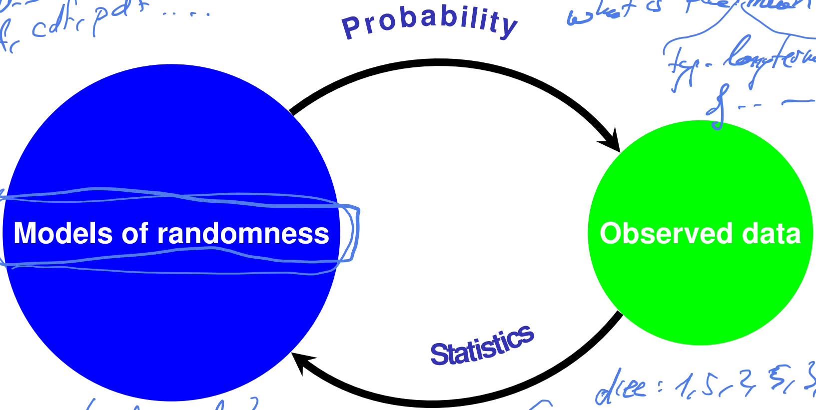
Limit theorems – approximation

Statistics – an introduction

Lecture overview

prob. spaces
r.v.'s
pmf, cdf, pdf ....

how likely it is that ... -
what is the mean
typ. longterm
& ... -



dice: 1, 5, 2, 5, 3, 5, 5, ...
# of emails: 51, 67, 32, ...

Q: is the fair die?

Poi(λ = 50)

1st illustration – number of left-handed people

#L : 0 0% | 14%

#R : 4 100% |

estimate in Czechia:

4–8 %

small sample → estimates about a larger population

caveats : → small sample

→ unrepresentative sample

→ misleading questions

2nd illustration – running time of a program

- $X_1, \ldots, X_n \sim F$ i.i.d., F is their CDF
- **Definition:** *Empirical CDF* is defined by

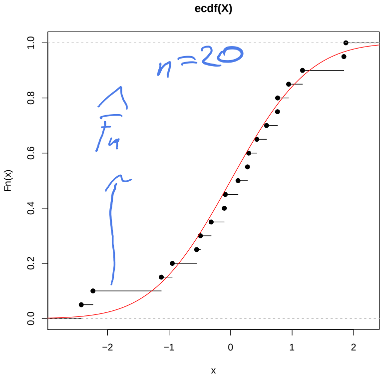
N(0,1)

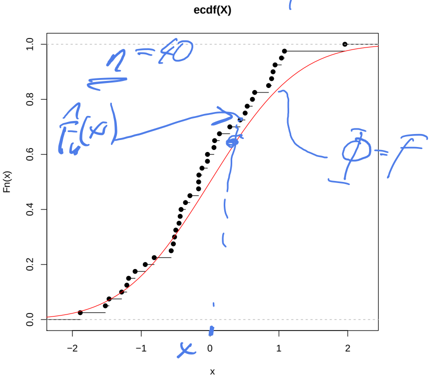
$F = \Phi$

a random variable

$$\widehat{F}_n(x) = \frac{\sum_{i=1}^{n} I(X_i \le x)}{n},$$

where $I(X_i \le x) = 1$ if $X_i \le x$ and 0 otherwise.





Empirical CDF – properties

Theorem

For a fixed x

- $\mathbb{E}(\widehat{F}_n(x)) = F(x)$
- $var(\widehat{F}_n(x)) = \frac{F(x)(1-F(x))}{n}$
- $\widehat{F}_n(x)$ *converges to* $F(x)$ *in probability,* $\widehat{F}_n(x) \xrightarrow{P} F(x)$.

$P(|\hat{F}_n(x) - F(x)| > \varepsilon) \longrightarrow 0$

~~Důkaz~~. Proof

Weak law of large numbers.

Note that $n\widehat{F}(x) \sim Bin(n, F(x))$

S_n $\qquad \mu = EY_i$

$E\,\underbrace{I(X_i \le x)}_{Y_i} = P(X_i \le x) = F(x)$

$Y_i \sim Bern(F(x))$

$var(Y_i) = F(x)(1-F(x))$

$\hat{F}_n(x) := S_n = \frac{Y_1 + \dots + Y_n}{n}$

$ES_n = EY_1 = F(x)$

$var(S_n) = \frac{F(x)(1-F(x))}{n}$

Empirical CDF – Dvoretzky-Kiefer-Wolfowitz (DKW)

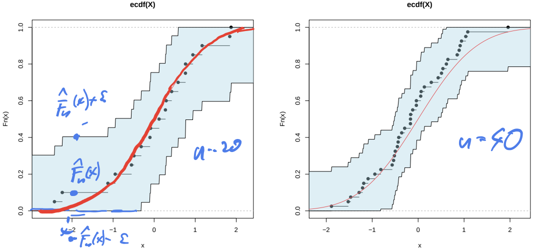
Theorem

Let $X_1, \ldots, X_n \sim F$ *be i.i.d., let* $\widehat{F}_n$ *be their empirical CDF. Let* $\mathbb{E}(X_i)$ *be finite. Choose* $\alpha \in (0,1)$ *and let* $\varepsilon = \sqrt{\frac{1}{2n} \log \frac{2}{\alpha}}$*. Then we have*

$\varepsilon = \frac{c_\alpha}{\sqrt{n}}$

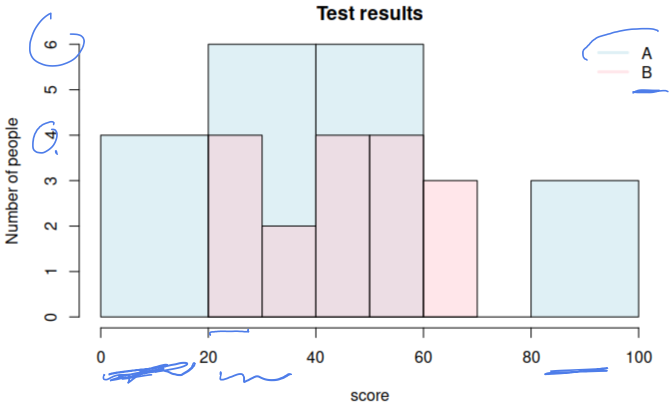
$$P(\widehat{F}_n(x) - \varepsilon \leq F(x) \leq \widehat{F}_n(x) + \varepsilon) \geq 1 - \alpha.$$

$\alpha = 0.05$



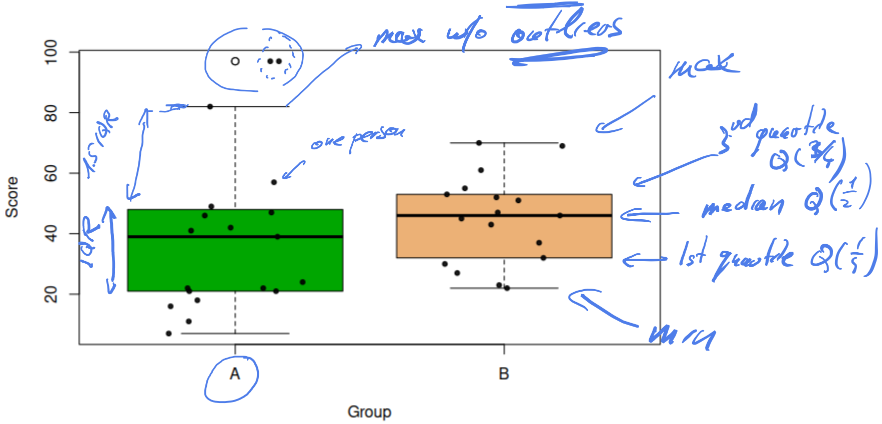
Intro – exploratory data analysis

- we collect data (and pay attention to systematic errors – independence, bias, . . .)
- we make various tables
- any appropriate charts: histogram, boxplot, etc.



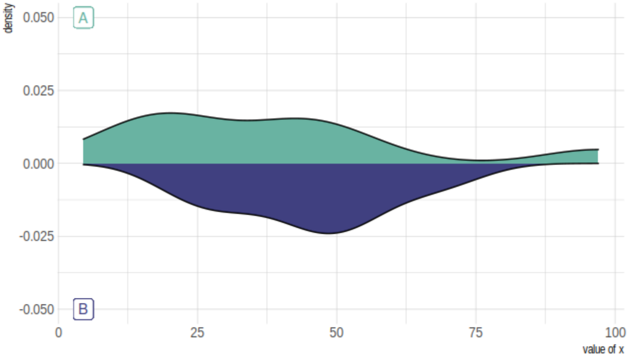
Intro – exploratory data analysis

- we collect data (and pay attention to systematic errors – independence, bias, . . .)
- we make various tables
- any appropriate charts: histogram, boxplot, etc.



Intro – exploratory data analysis

- we collect data (and pay attention to systematic errors – independence, bias, . . .)
- we make various tables
- any appropriate charts: histogram, boxplot, etc.



Goals of confirmatory data analysis

- point estimates
- interval estimates
- hypothesis testing
- (linear) regression

$\mu = 1.8\ m$

$\mu \in (1.5, 2)$ w 95% chance

Examples:

- We assume human height follows $N(\mu, \sigma^2)$. What are μ and σ?
- Is our coin/dice fair?
- Is a medical treatment beneficial?
- Is new version of a program faster?
- How does running time depend on size of input?