Gaussova krivka

Radoslav Harman

Katedra aplikovanej matematiky a štatistiky
FMFI UK

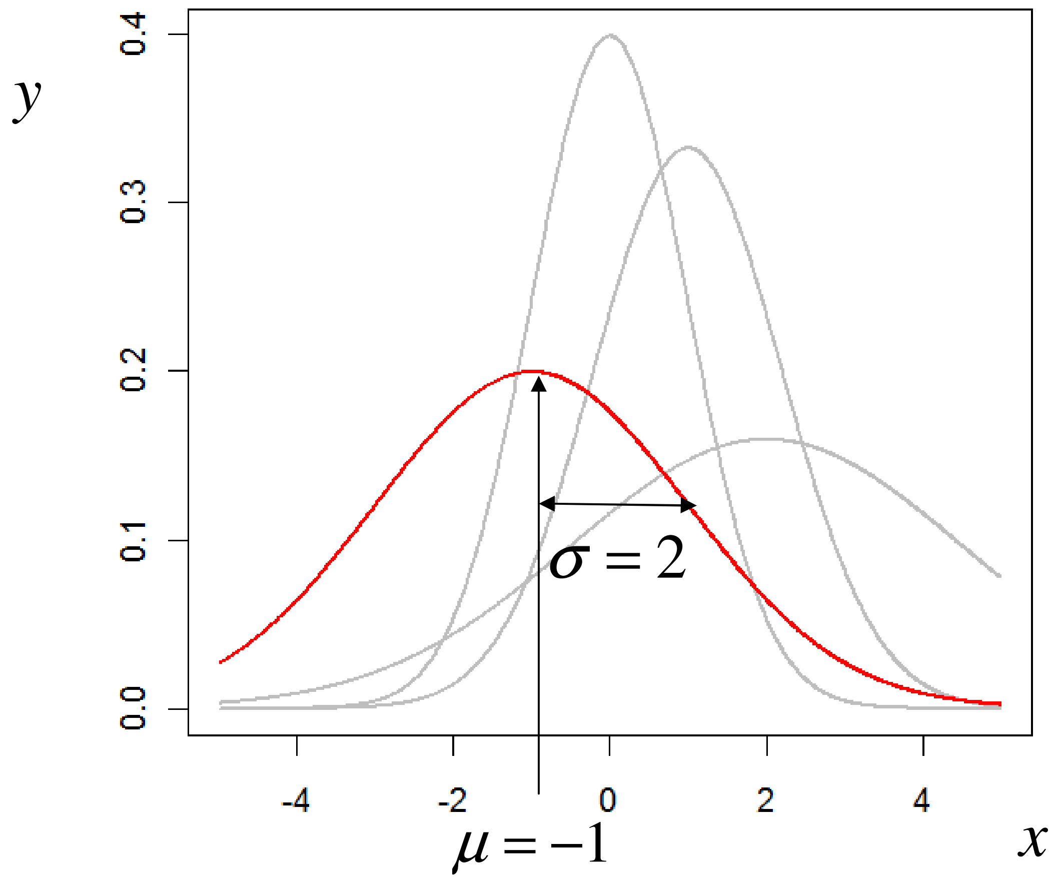
Akadémia Trojstenu 2010

Obsah

- Gaussova krivka a rozdelenie pravdepodobnosti
- Centrálna limitná veta
- Gaussova krivka a objemy
- Maxwellova charakterizácia
- Testovanie štatistických hypotéz

Gaussova krivka a
Gaussovo rozdelenie pravdepodobnosti

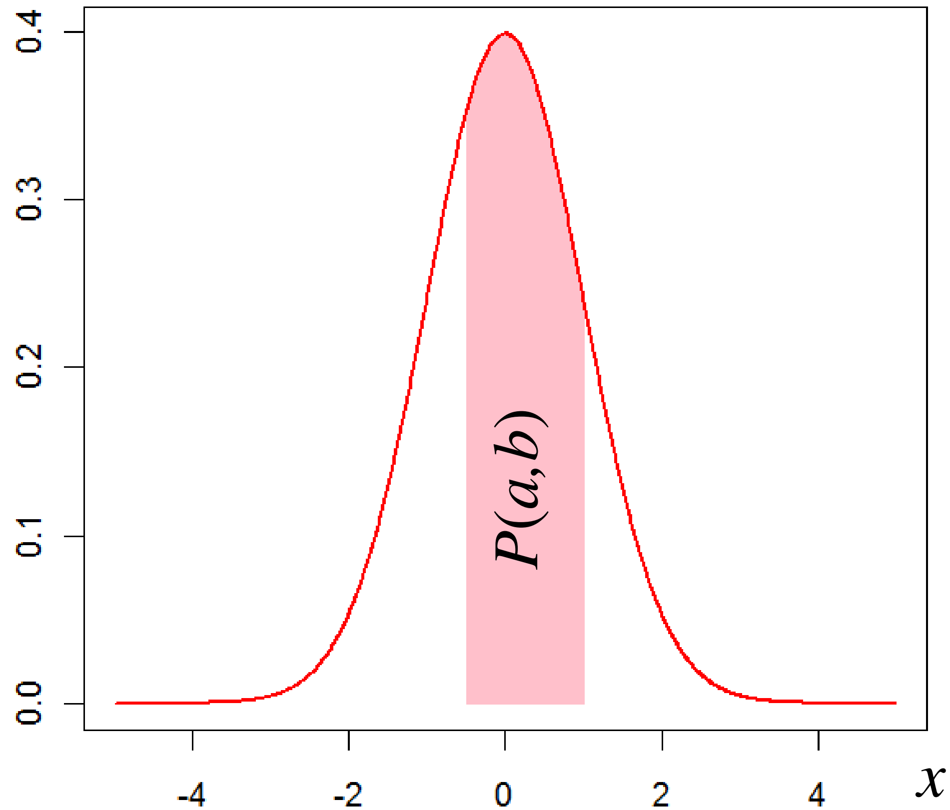
$$y = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$




GN4480100S8
DEUTSCHE BUNDESBANK
Banknote
ZEHN DEUTSCHE MARK
10
1777–1855 Carl Friedr. Gauß
$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
Deutsche Bundesbank
Frankfurt am Main
1. September 1999
GN4480100S8

Gaussova krivka a Gaussovo rozdelenie pravdepodobnosti

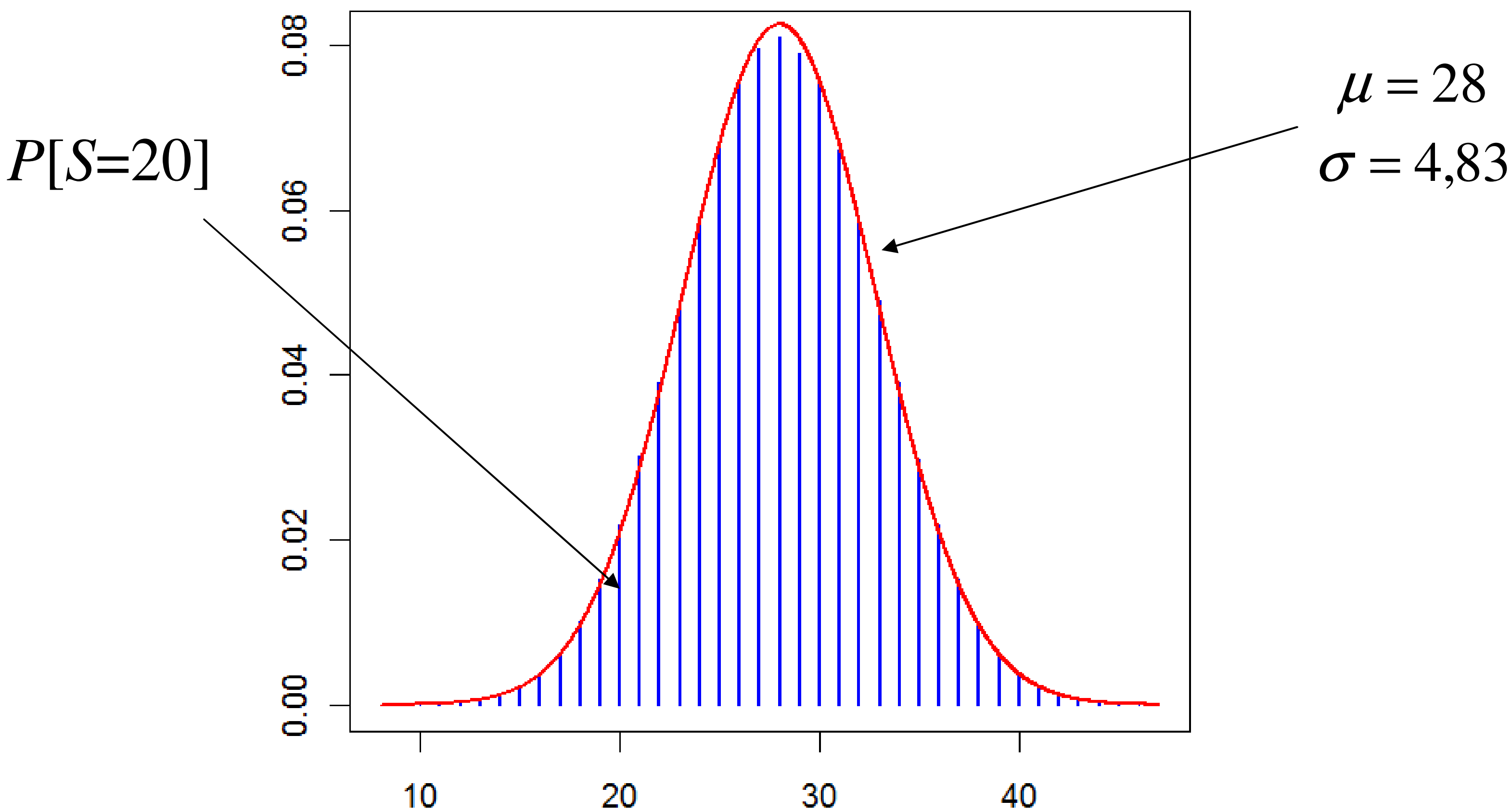
$$P(a,b) = \int_a^b \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$



$$\mu = 0, \sigma = 1, a = -0{,}5, b = 2$$

Centrálna limitná veta

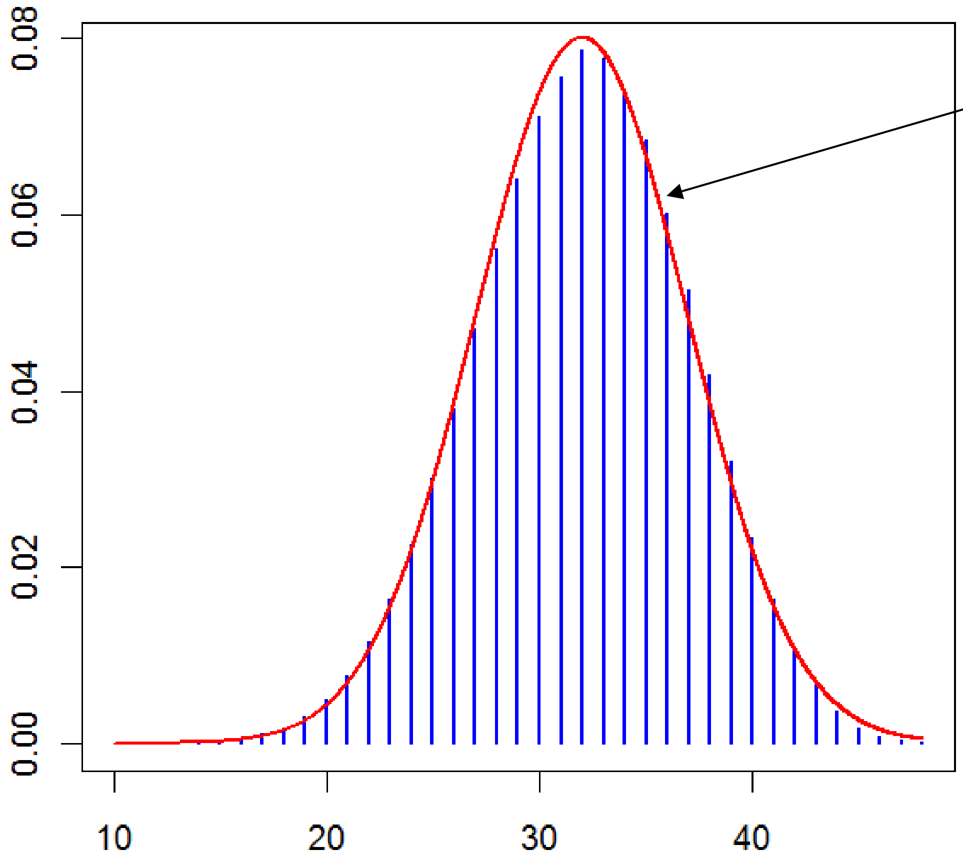
Súčet bodov na 8 hracích kockách



Centrálna limitná veta

Súčet bodov na 8 *falošných* hracích kockách

p_1=0,1; p_2=0,2; p_3=0,05; p_4=0,2; p_5=0,15; p_6=0,3



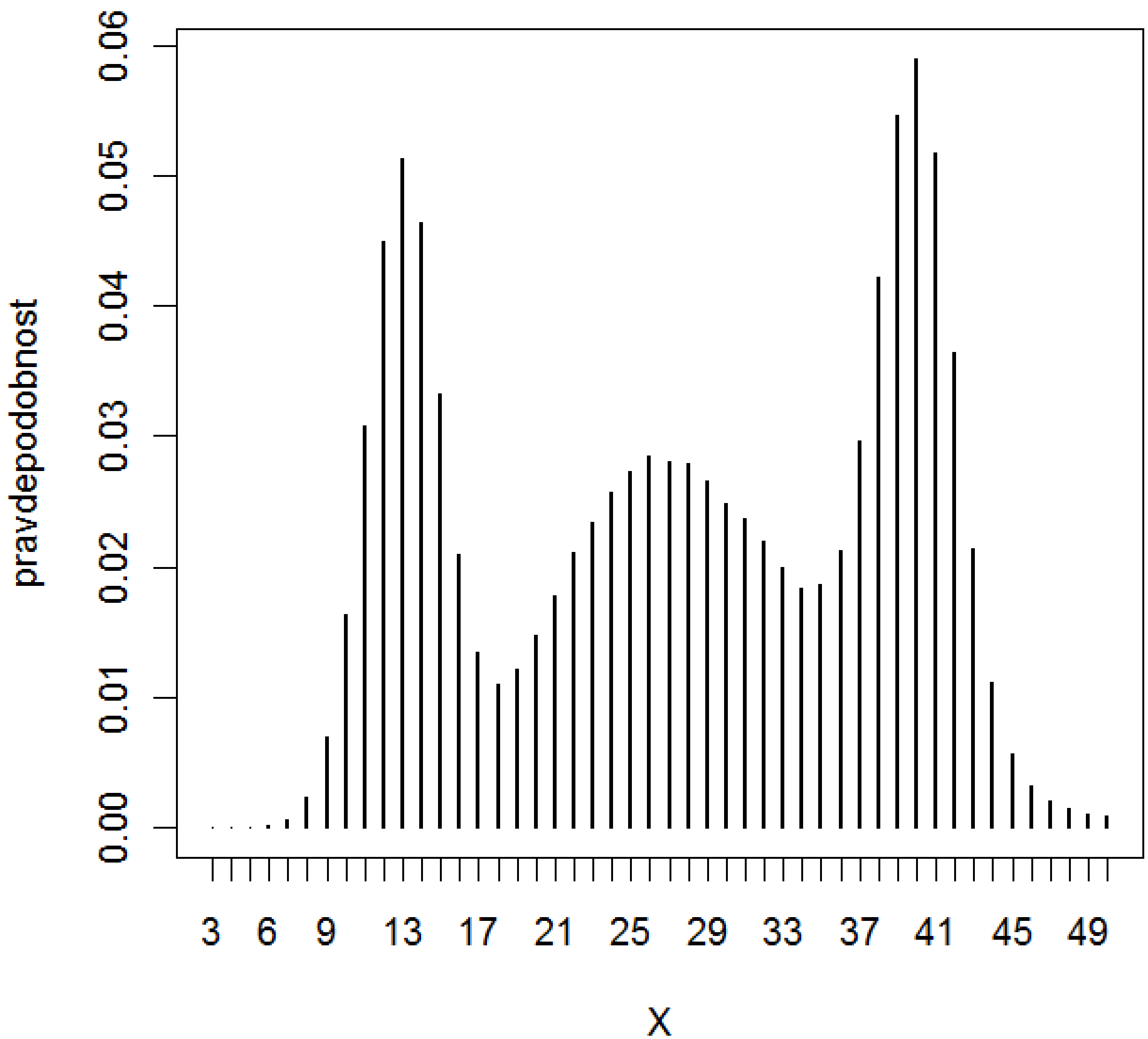
$\mu = 32$

$\sigma = 4{,}98$

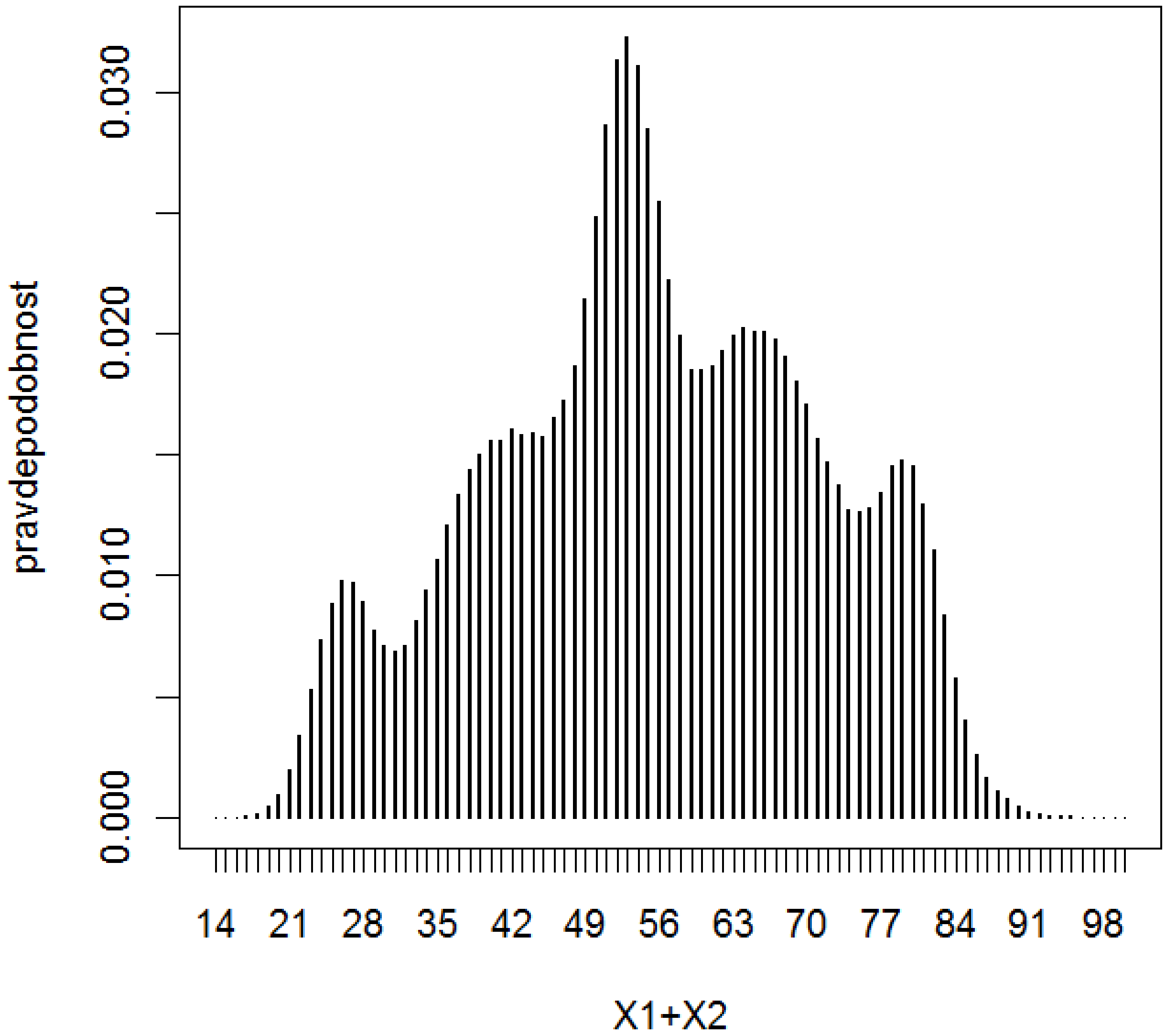
Centrálna limitná veta

Intuitívna formulácia:

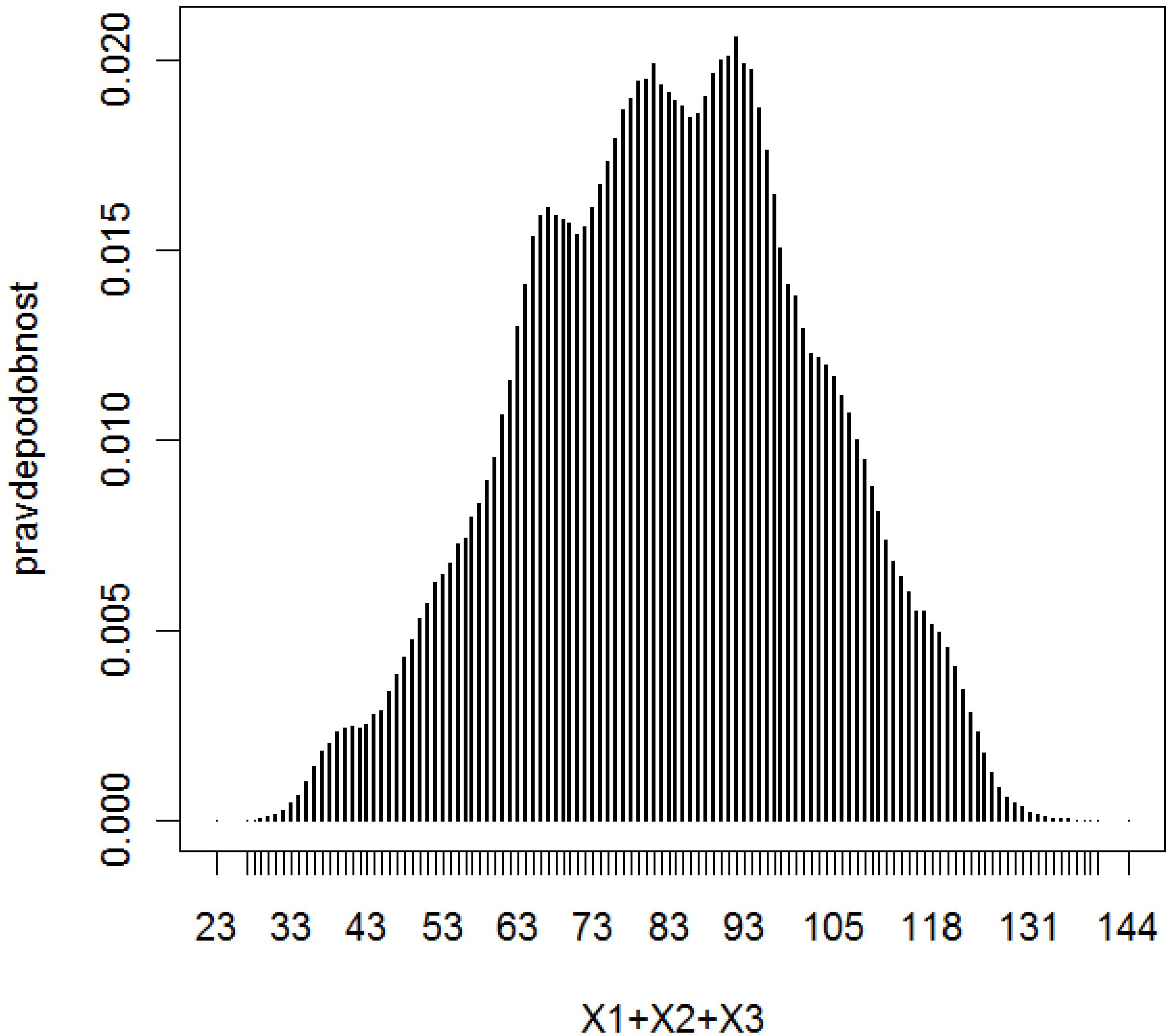
Súčet nezávislých náhodných výsledkov sa správa približne podľa Gaussovho rozdelenia pravdepodobnosti. Čím väčší je počet sčítancov, tým je aproximácia súčtu Gaussovým rozdelením presnejšia. Toto tvrdenie nezávisí na type náhodnosti jednotlivých sčítancov.



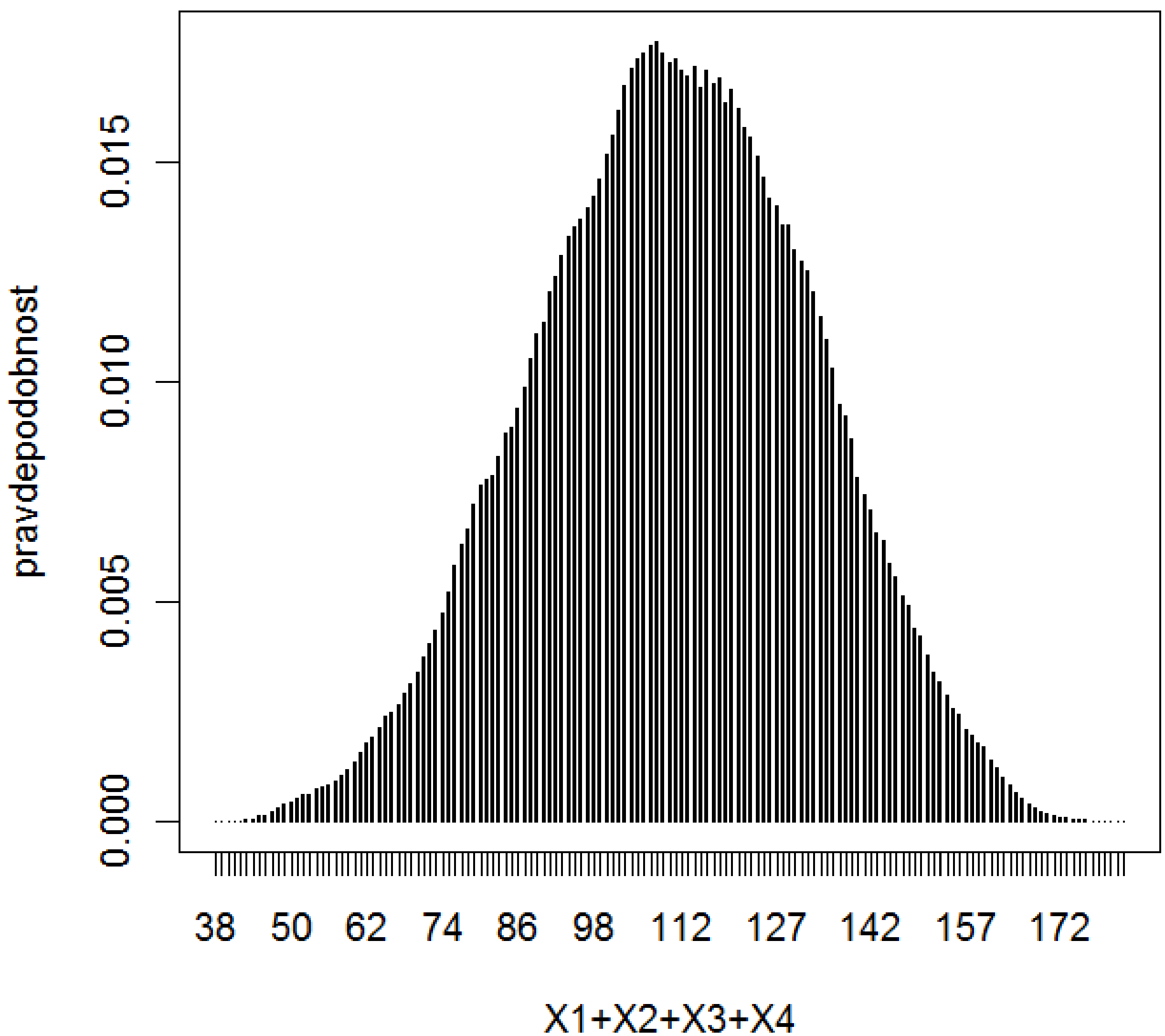
0.00
0.01
0.02
0.03
0.04
0.05
0.06
pravdepodobnost
3
6
9
13
17
21
25
29
33
37
41
45
49
X

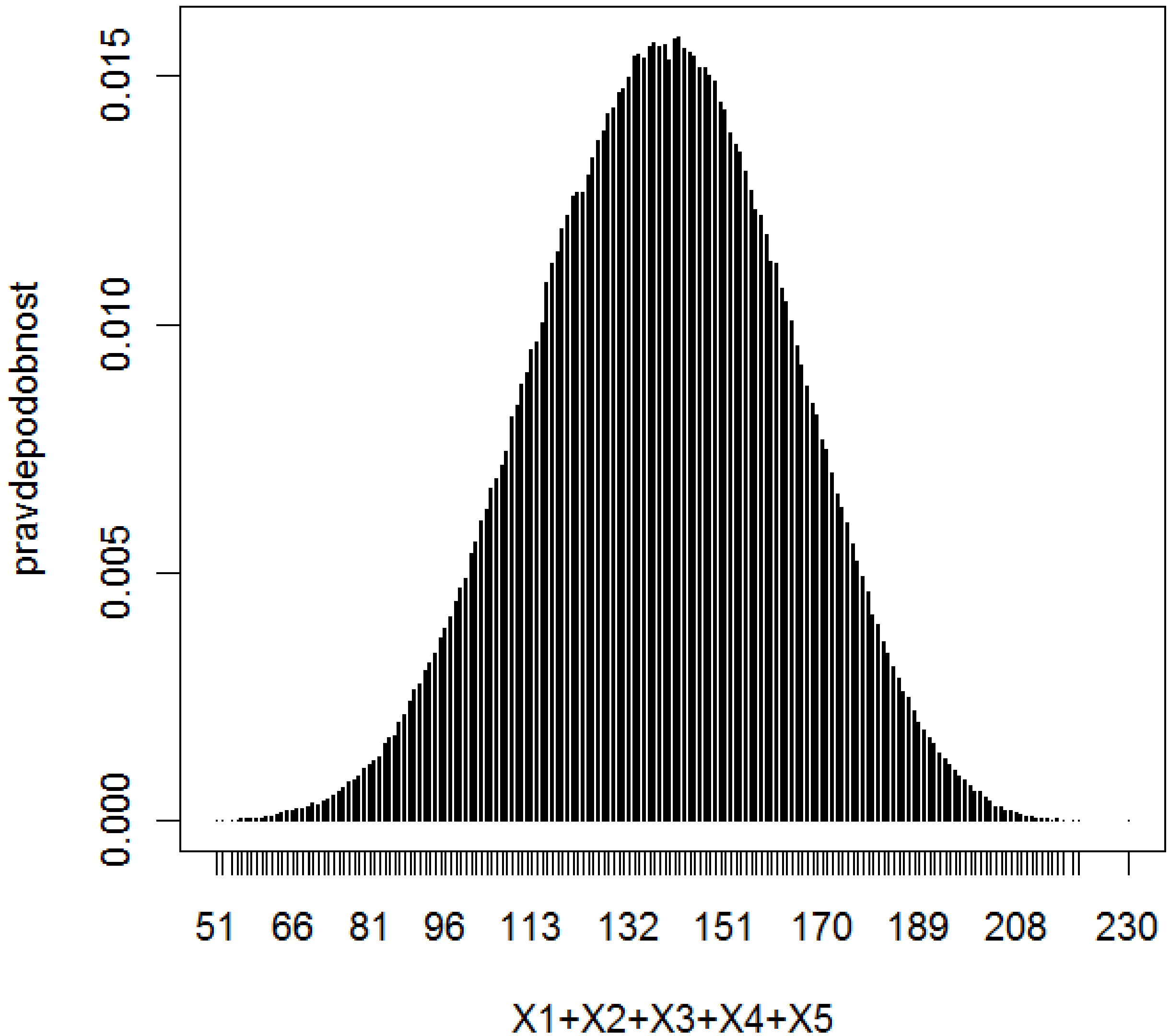


pravdepodobnost
0.000
0.010
0.020
0.030
14
21
28
35
42
49
56
63
70
77
84
91
98
X1+X2

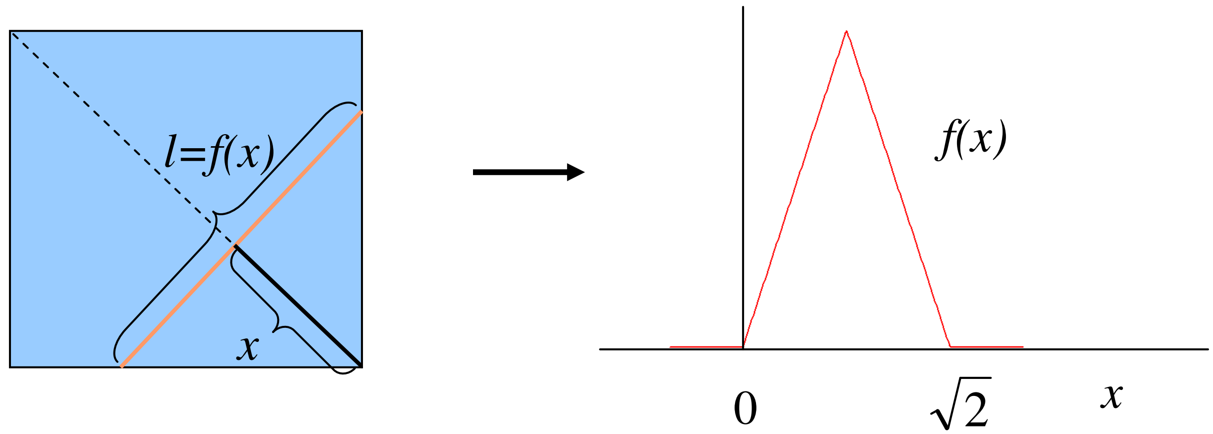


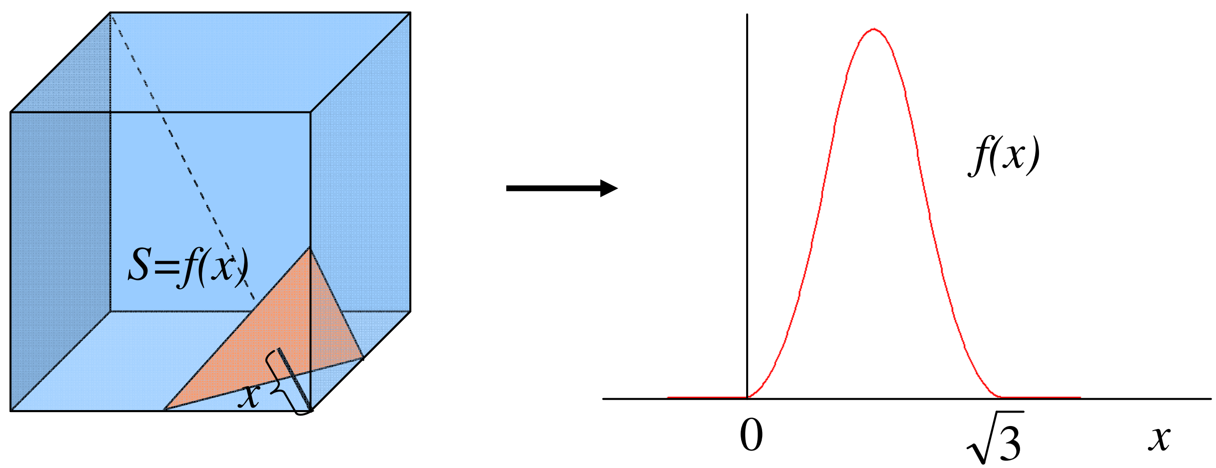
pravdepodobnost
0.000
0.005
0.010
0.015
0.020
23
33
43
53
63
73
83
93
105
118
131
144
X1+X2+X3


pravdepodobnost
0.000
0.005
0.010
0.015
38
50
62
74
86
98
112
127
142
157
172
X1+X2+X3+X4


0.000
0.005
0.010
0.015
pravdepodobnost
51
66
81
96
113
132
151
170
189
208
230
X1+X2+X3+X4+X5

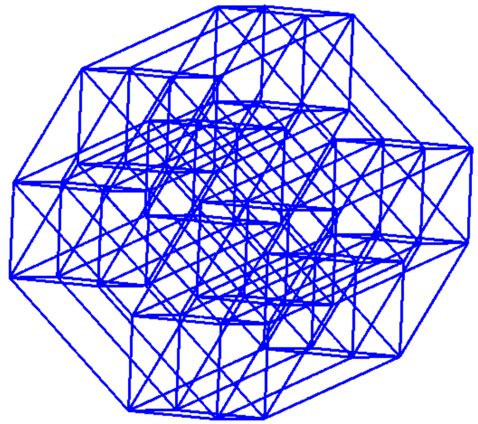
Gaussova krivka a objemy

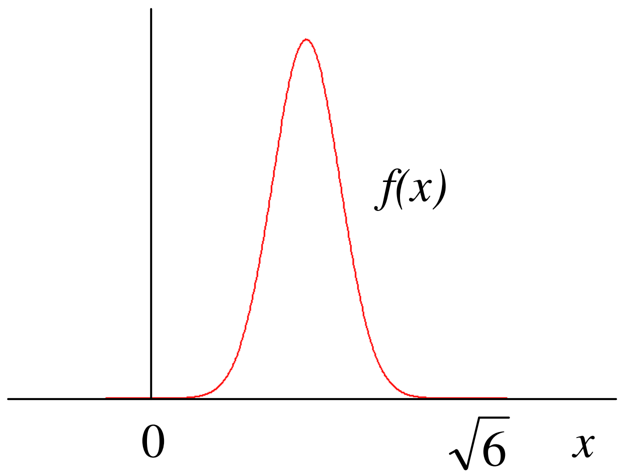




Gaussova krivka a objemy

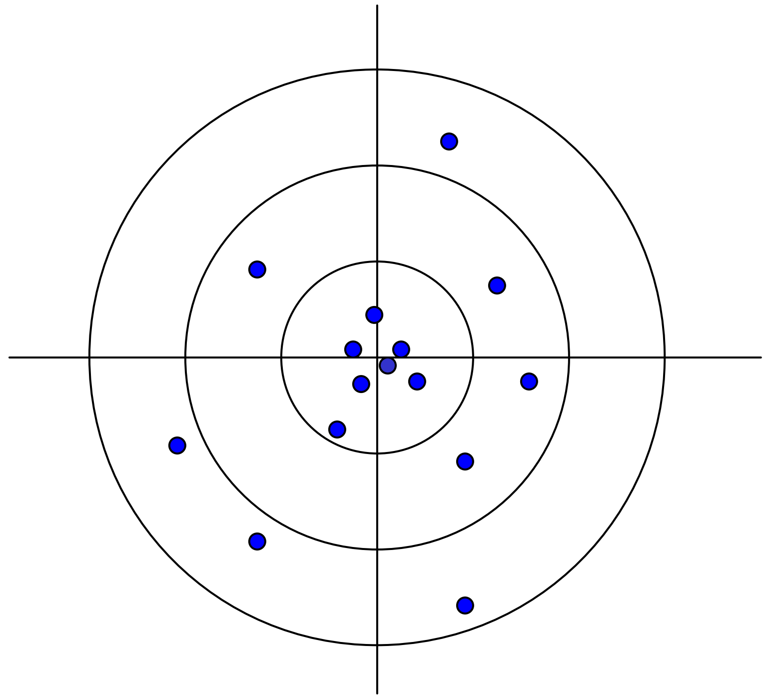
6-rozmerná kocka





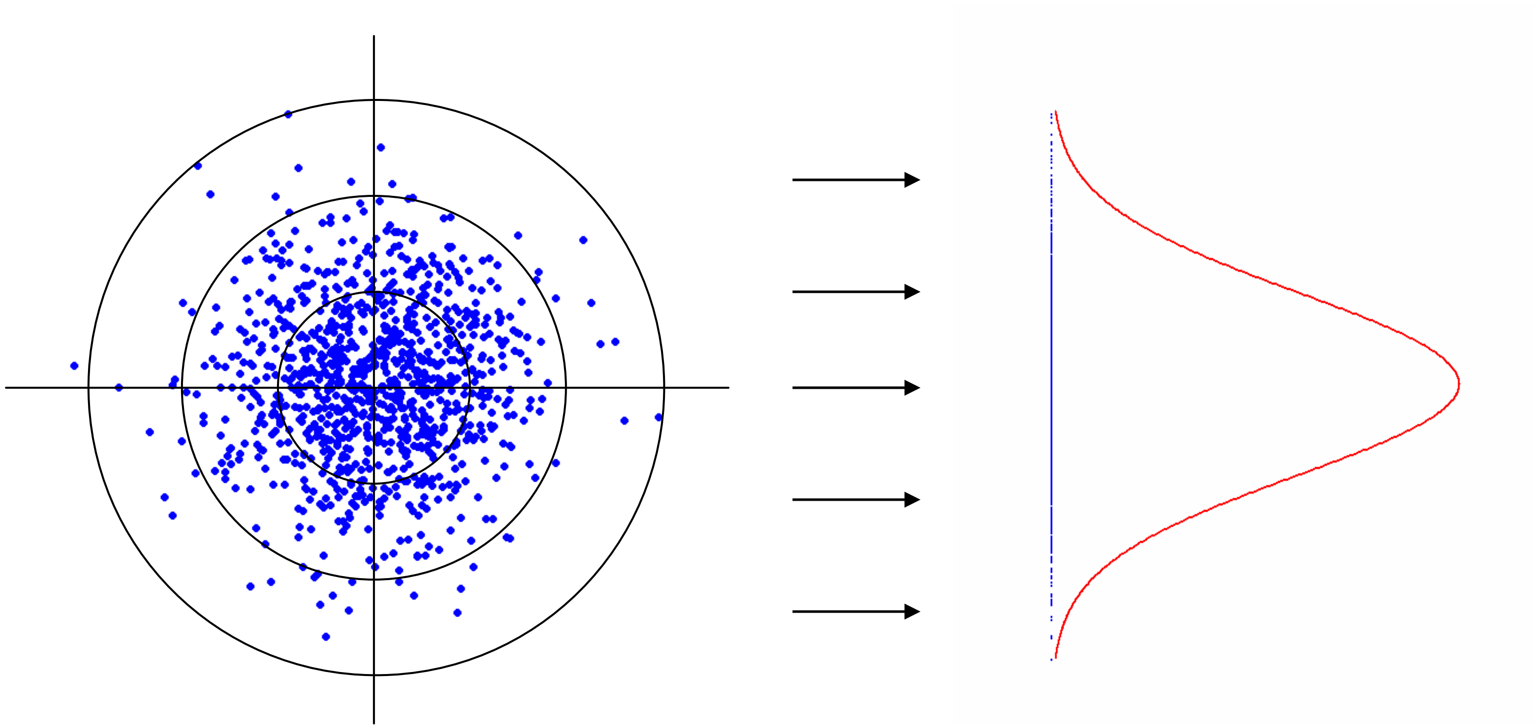
x ...vzdialenosť od vrchola v smere hlavnej uhlopriečky
V=f(x) ...5-rozmerný objem rezu kocky kolmého na uhlopriečku v bode *x*

Maxwellova charakterizácia



Spýtajme sa, či existuje spôsob generovania bodov v rovine, ktorý:
- správa sa rovnako pri rotácií osí (rotačná invariantnosť)
- informácia o súradnici bodu neposkytuje žiadnu informáciu o kolmej súradnici tohto bodu (stochastická nezávislosť zložiek)

Maxwellova charakterizácia



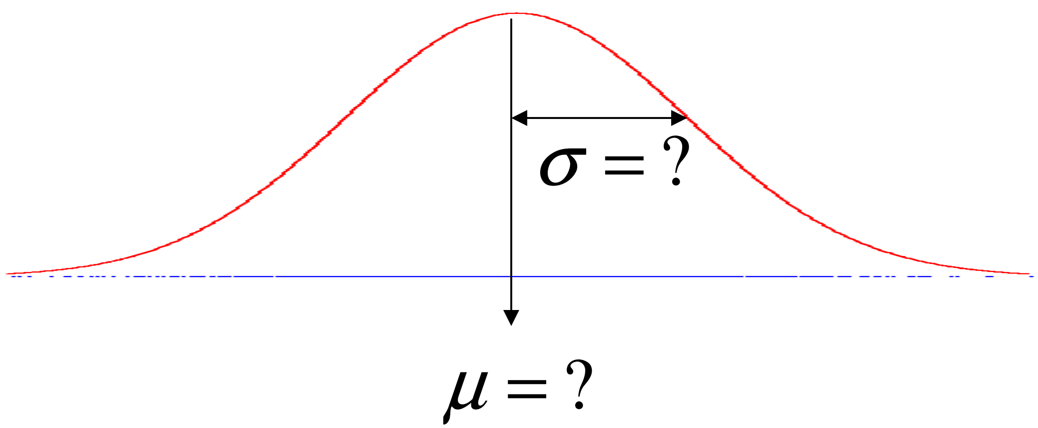
Taký mechanizmus existuje a je jediný (až na „roztiahnutie“):
V každom smere je náhodnosť bodov popísaná Gaussovou krivkou

Testovanie štatistických hypotéz

Číselné dáta

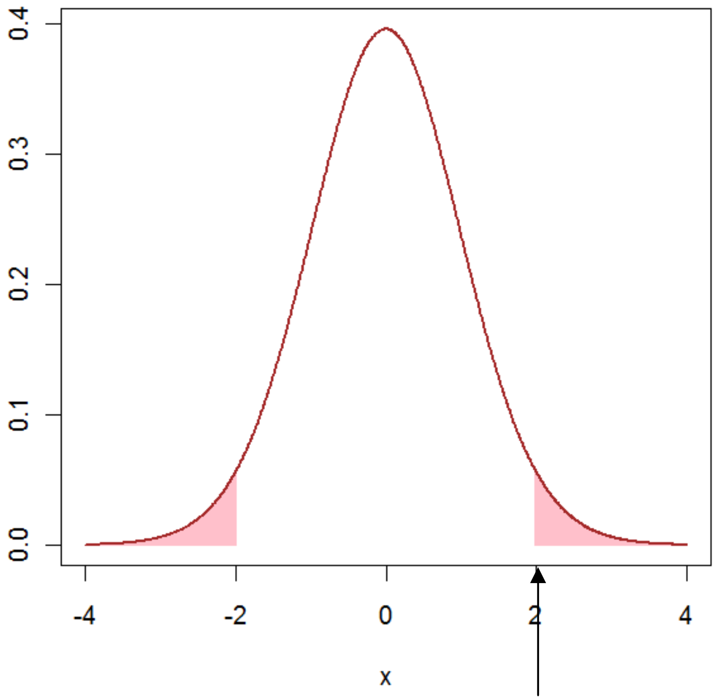
3.253	2.894	3.053	2.903	3.184	3.172	3.411	2.819	2.962	3.048	2.478	2.747	2.921
3.191	2.935	3.077	2.797	2.872	2.891	2.613	2.899	3.405	3.367	3.078	3.333	2.910
3.136	2.906	3.278	3.107	3.114	2.670	2.965	3.013	3.141	2.847	2.765	2.761	3.033

Predpoklad: Dáta „pochádzajú“ z rozdelenia pravdepodobnosti, sú dané nejakým pravdepodobnostným modelom.
Veľmi často je tento model založený na Gaussovej krivke.



Testovanie štatistických hypotéz

Nulová hypotéza: $\mu = \mu_0$ $\qquad \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}$



$$T = \sqrt{n}\,\frac{\overline{X} - \mu_0}{S}$$

p-hodnota (ružová plôška na obrázku): Pravdepodobnosť, že ak nulová hypotéza platí, tak dostaneme také „extremálne“ T, ako vyšlo z dát, alebo ešte extremálnejšie.

Ak je p-hodnota veľmi malá, tak nulovej hypotéze neveríme („zamietame“ ju).

Testovanie štatistických hypotéz

Problém: Gaussovo rozdelenie poskytuje veľké odchýlky s extrémne malými pravdepodobnosťami. Metódy založené na Gaussovom rozdelení prakticky vylučujú „odľahlé pozorovania“.

Napríklad ak $\mu = 0, \sigma = 1$ P[X>0] = 0,500… P[X>1] = 0,158…

P[X>2] = 0,0227… P[X>3] = 0,00134… P[X>4] = 0,0000316…

P[X>5] = 0,000000286… P[X>6] = 0,000000000986…

Problém: Pravda je často nudná a nepravda zaujímavá. Štatistické testy občas „náhodou“ potvrdia nepravdivé tvrdenie, ale vďaka tomu, že je zaujímavé, sa o ňom dozvieme.

Ďakujem za pozornosť!

Tento prednáška bude dostupná na adrese
www.radoslav-harman.blogspot.com