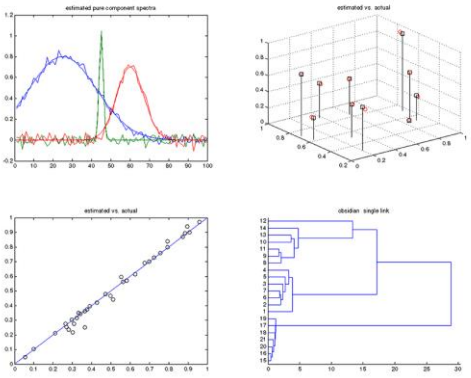
ANALYTICKÁ CHEMOMETRIE



Patrik Kania
Ústav analytické chemie, VŠCHT Praha

16. 09. 2025

CHEMOMETRIE (angl. *chemometrics*)

- chemická disciplína, která využívá matematických a statistických metod:
  1) k volbě optimálního experimentálního postupu,
  2) k získávání maxima relevantních informací z experimentálních dat,
  3) k prezentaci analytických výsledků,
  4) k zhodnocení analytických metod.

- vznik na přelomu 60. a 70. let s masivním nástupem počítačů do vědy
- pojem poprvé použil Wold v roce 1971 na univerzitě v Umea (SWE)

- v roce 1974 zakládá Bruce Kowalski a Svant Wold
  International Chemometrics Society

70. léta – určování koeficientů známých modelů
od 80. let – analýza vícerozměrných souborů dat, faktorová analýza, atd.

ABSOLVOVÁNÍ PŘEDMĚTU

zápočet:

	test č. 1	test č. 2	test č. 3
maximum bodů	15	20	15

pro získání zápočtu je minimum 15 z 50 bodů

zkouška: zkouškový test (maximum 50 bodů)

výsledek zkoušky = zkouškový test + zápočtové testy + bonusy

A	B	C	D	E	F
100–89	88–79	78–69	68–59	58–50	<50

známku lze zlepšit i nepovinnou ústní zkouškou (max. o dva stupně)

3

Tematické okruhy pro zkoušku (příp. i pro SZZ)

TO3: Management jakosti a chemometrika

(vychází z předmětu M402008 Analytická chemometrika, M402016 Vícerozměrné statistické metody)

Náhodná veličina a její charakterizace, číselné charakteristiky rozdělení, normální rozdělení a jeho standardizace, náhodný výběr a jeho realizace, bodové a intervalové odhady parametrů základního souboru, testování hypotéz (chyba I. a II. druhu, síla testu, volba hypotézy), parametrické testy středních hodnot a rozptylů, vylučování odlehlých výsledků, neparametrické testy středních hodnot, analýza rozptylu (mnohonásobné porovnávání, Kruskalův-Wallisův test), faktoriální pokusy, optimalizace experimentu, kontingenční tabulky, regresní a korelační analýza, kalibrace analytické metody, nejistoty měření.

Předzpracování vícerozměrných dat, vícerozměrné průzkumné techniky, regresní metody, klasifikační metody, validace statistických metod.

studijní program Analytická chemie

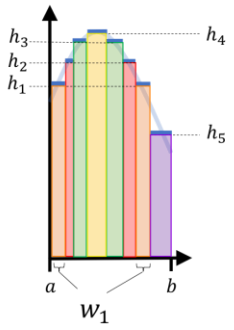
specializace: N403A (Analýza léčiv)
N403B (Forenzní chemie)
N403C (Analytická chemie a jakostní inženýrství)
N403D (Molekulární chemická analýza)

studijní program Datové inženýrství v chemii N405

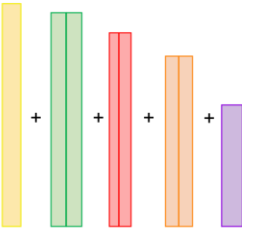
6

ANALYTICKÁ CHEMOMETRIE

základy teorie pravděpodobnosti



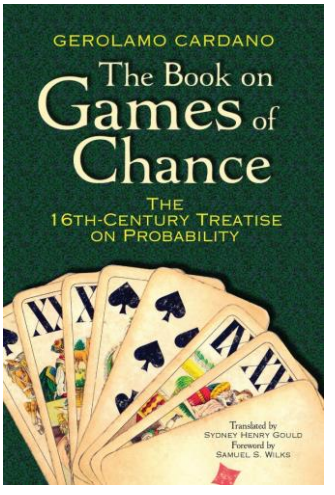
$$\int_{(a,b)} f \approx \sum_{i=1}^{5} w_i h_i =$$

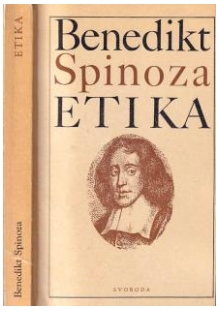


- základy teorie pravděpodobnosti až v 16. století
  (teorie hazardu, později aplikace v astronomii, demografii a obchodě)

Gerolamo Cardano (1501–1576)

- italský matematik, lékař a filosof
- roku 1545 publikace řešení rovnic 3. a 4. řádu
- používal pravděpodobnost při hazardu





Baruch Spinoza (1632–1677)

- nizozemský filozof
- jediná substance × Descartes

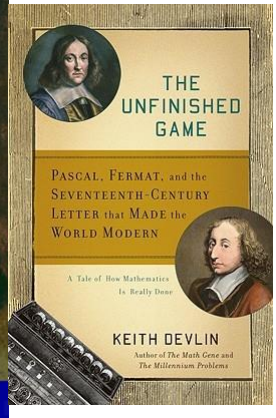
tvrzení 29 (str. 88):

V přírodě neexistuje nic náhodného, nýbrž všechny věci jsou přirozeností Boha nutně determinovány k určitému modu existence a působení.

poznámka 1 (str. 92– 93):

… Jako náhodnou však označujeme věc jen z důvodů tkvících v nedostatečnosti našeho poznání. Ta věc, o níž nevíme, zda její esence nezahrnuje protiklad, nebo o níž bezpečně víme, že žádný protiklad nezahrnuje, a přesto o její existenci nemůžeme nic s jistotou tvrdit, protože řád příčin je nám skryt, taková věc se nám nemůže jevit ani jako nutná, ani jako nemožná, a proto ji nazýváme náhodnou nebo možnou.

- rok 1654 všeobecně považován za počátek teorie pravděpodobnosti
  → vyřešena úloha o rozdělení sázky



HROMADNÉ JEVY - za stejných podmínek opakovaně nastávají

JISTÝ JEV - stejné podmínky → stejné výsledky
(např. pohyb vodiče v magnetickém poli indukuje proud)

NÁHODNÝ JEV - stejné podmínky → různé výsledky (např. při měření)
- může a nemusí za určitých okolností nastat
- nemůžeme předpovědět s jistotou, zda kdykoli v budoucnu jev nastane či nenastane

Každé pozorování náhodného jevu či každé jeho změření je náhodný pokus

výsledkem náhodného pokusu - číslo, vlastnost či konstatování A × N

náhodný pokus - např. házení kostkou

výsledky - 1, 2, 3, 4, 5 a 6 tvoří množinu Ω

podmnožiny množiny Ω - náhodné jevy

padne číslo 1	$A = \{1\}$
padne sudé číslo	$B = \{2, 4, 6\}$
číslo větší než 6	$C = \{\}$

Náhodný jev M se nazývá elementární, jestliže neexistují jiné náhodné jevy X a Y takové, že $M = X \cup Y$

Prostor elementárních jevů náhodného pokusu Ω - množina všech elementárních jevů

konečný nebo spočetný počet prvků
DISKRÉTNÍ prostor

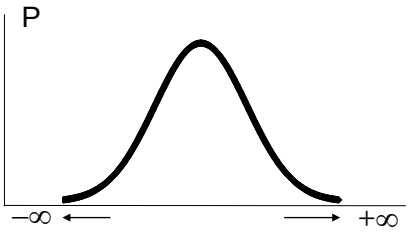
nespočetné množství prvků
SPOJITÝ prostor

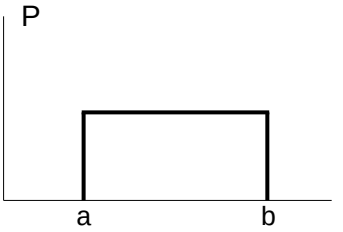
Číselná veličina, jejíž hodnota závisí na působení náhodných jevů se nazývá náhodná veličina (angl. *random variable*)

- musí být jednoznačně určena jednou hodnotou
- pro její charakterizaci musí být předem známy její meze (definiční obor)

příklady náhodných veličin, jejich mezí a pravděpodobnosti výskytu hodnot:

hmotnostní obsah, *w*	(mg/kg)	$0–10^6$
teplota, *t*	(°C)	$> -273{,}15$
koncentrace, *c*	(mol/l)	≥ 0

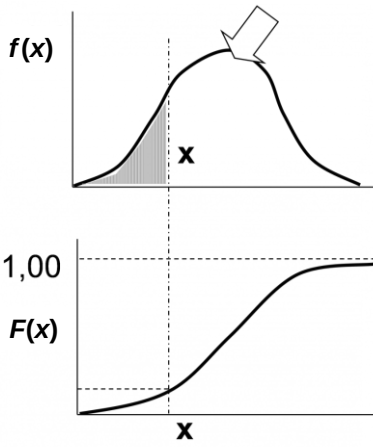




Funkce rozdělení pravděpodobnosti

Náhodná veličina - popsána funkcí udávající rozdělení pravděpodobnosti

plocha = kumulovaná četnost



pravděpodobnostní funkce
nebo
hustota pravděpodobnosti

udává pravděpodobnost, že určitá náhodná veličina ξ nabude právě konkrétní hodnoty x

$$f(x) = \mathrm{P}\{\xi = x\}$$

distribuční funkce
(relativní kumulovaná četnost)

udává pravděpodobnost, že určitá náhodná veličina ξ nabude nejvýše konkrétní hodnoty x

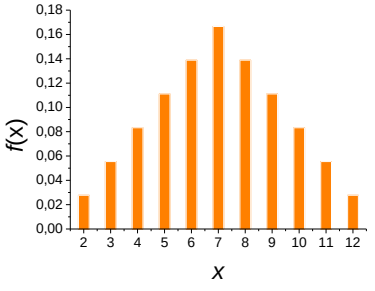
$$F(x) = \mathrm{P}\{\xi \leq x\}$$

Náhodná nespojitá (diskrétní) veličina (angl. *discrete random variable*)

- nabývá vybraných hodnot v definičním intervalu
- má schodovitý tvar
- skoky DF v bodech x_i mají velikost $\mathrm{P}\{\xi = x_i\}$

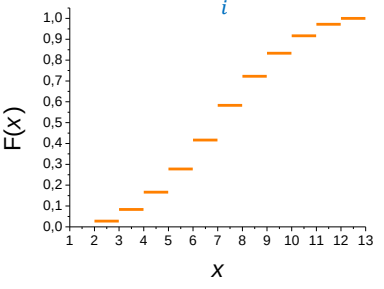
pravděpodobnostní funkce
(angl. *probability mass function*)

$$f(x) = \mathrm{P}\{\xi = x\}$$



distribuční funkce (DF)
(angl. *cumulative distribution function*)

$$F(x) = \mathrm{P}\{\xi \leq x\} = \sum_i \mathrm{P}\{\xi = x_i\}, x_i \leq x$$

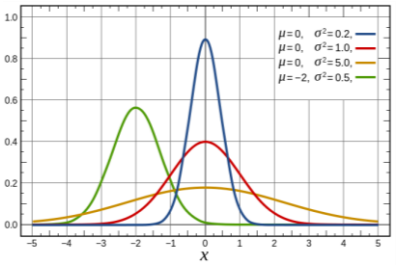


Náhodná spojitá veličina (ang. *continuous random variable*)

- nabývá jakýchkoliv hodnot v definičním intervalu
- má spojitý tvar

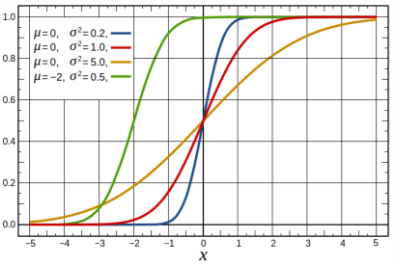
hustota pravděpodobnosti
(angl. *probability density function*)

$$f(x)\mathrm{d}x = \mathrm{P}\{x \le \xi \le x + \mathrm{d}x\}$$



distribuční funkce
(angl. *cumulative distribution function*)

$$F(x) = \mathrm{P}\{\xi \le x\} = \int_{-\infty}^{x} f(x)dx$$



Vlastnosti distribuční funkce $F(x)$ náhodné veličiny ξ:

$0 \le F(x) \le 1$

$F(x_1) \le F(x_2)$, pro všechna $x_1 < x_2$

je to zprava spojitá funkce

$F(-\infty) = 0$ a $F(\infty) = 1$

$\mathrm{P}(x_1 \le \xi \le x_2) = F(x_2) - F(x_1)$

$\mathrm{P}(\xi > x) = 1 - F(x)$

v případě intervalu $\langle a, b \rangle$ platí:

$F(a) \ge 0$ a $F(b) = 1$

v případě normovaného normálního rozdělení platí:

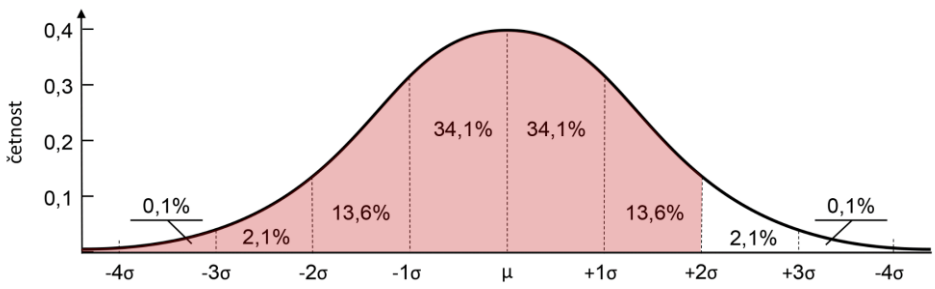
$\Phi(-z) = 1 - \Phi(z)$

Kvantily normálního rozdělení

Kvantily normálního rozdělení

- body rozdělující obor hodnot náhodné veličiny v určitém poměru
- jsou mírou polohy rozdělení prsti náhodné veličiny
- jsou tabelovány (platí pouze pro normované normální rozdělení)
- 100·p% kvantil je hodnota x_p, která splňuje nerovnosti:

$$\mathrm{P}(\xi \leq x_p) = p, \qquad \text{tedy } F(x_p) = p$$
$$\mathrm{P}(\xi > x_p) = 1 - p$$



Distribuční funkce normálního rozdělení N(0, 1) Excel: **NORMSDIST(*z*)**

(pro hodnoty *z* jsou uvedeny příslušné hodnoty F(*z*))

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,50000	0,50399	0,50798	0,51197	0,51595	0,51994	0,52392	0,52790	0,53188	0,53586
0,1	0,53983	0,54380	0,54776	0,55172	0,55567	0,55962	0,56356	0,56749	0,57142	0,57535
0,2	0,57926	0,58317	0,58706	0,59095	0,59483	0,59871	0,60257	0,60642	0,61026	0,61409
0,3	0,61791	0,62172	0,62552	0,62930	0,63307	0,63683	0,64058	0,64431	0,64803	0,65173
0,4	0,65542	0,65910	0,66276	0,66640	0,67003	0,67364	0,67724	0,68082	0,68439	0,68793
0,5	0,69146	0,69497	0,69847	0,70194	0,70540	0,70884	0,71226	0,71566	0,71904	0,72240
0,6	0,72575	0,72907	0,73237	0,73565	0,73891	0,74215	0,74537	0,74857	0,75175	0,75490
0,7	0,75804	0,76115	0,76424	0,76730	0,77035	0,77337	0,77637	0,77935	0,78230	0,78527
0,8	0,78814	0,79103	0,79389	0,79673	0,79955	0,80234	0,80511	0,80785	0,81057	0,81327
0,9	0,81594	0,81859	0,82121	0,82381	0,82639	0,82894	0,83147	0,83398	0,83646	0,83891
1,0	0,84134	0,84375	0,84614	0,84850	0,85083	0,85314	0,85543	0,85769	0,85993	0,86214
1,1	0,86433	0,86650	0,86864	0,87076	0,87286	0,87493	0,87698	0,87900	0,88100	0,88298
1,2	0,88493	0,88686	0,88877	0,89065	0,89251	0,89435	0,89617	0,89796	0,89973	0,90147
1,3	0,90320	0,90490	0,90658	0,90824	0,90988	0,91149	0,91309	0,91466	0,91621	0,91774
1,4	0,91924	0,92073	0,92220	0,92364	0,92507	0,92647	0,92786	0,92922	0,93056	0,93189
1,5	0,93319	0,93448	0,93574	0,93699	0,93822	0,93943	0,94062	0,94179	0,94295	0,94408
1,6	0,94520	0,94630	0,94738	0,94845	0,94950	0,95053	0,95154	0,95254	0,95352	0,95449
1,7	0,95543	0,95637	0,95728	0,95818	0,95907	0,95994	0,96080	0,96164	0,96246	0,96327
1,8	0,96407	0,96485	0,96562	0,96638	0,96712	0,96784	0,96856	0,96926	0,96995	0,97062
1,9	0,97128	0,97193	0,97257	0,97320	0,97381	0,97441	0,97500	0,97558	0,97615	0,97670
2,0	0,97725	0,97778	0,97831	0,97882	0,97932	0,97982	0,98030	0,98077	0,98124	0,98169
2,1	0,98214	0,98257	0,98300	0,98341	0,98382	0,98422	0,98461	0,98500	0,98537	0,98574
2,2	0,98610	0,98645	0,98679	0,98713	0,98745	0,98778	0,98809	0,98840	0,98870	0,98899
2,3	0,98928	0,98956	0,98983	0,99001	0,99036	0,99061	0,99086	0,99111	0,99134	0,99158
2,4	0,99180	0,99202	0,99224	0,99245	0,99266	0,99286	0,99305	0,99324	0,99343	0,99361
2,5	0,99379	0,99396	0,99413	0,99430	0,99446	0,99461	0,99477	0,99492	0,99506	0,99520
2,6	0,99534	0,99547	0,99560	0,99573	0,99586	0,99598	0,99609	0,99621	0,99632	0,99643
2,7	0,99653	0,99664	0,99674	0,99683	0,99693	0,99702	0,99711	0,99720	0,99728	0,99736
2,8	0,99744	0,99752	0,99760	0,99767	0,99774	0,99781	0,99788	0,99795	0,99801	0,99807
2,9	0,99813	0,99819	0,99825	0,99830	0,99836	0,99841	0,99846	0,99851	0,99856	0,99860
3,0	0,99865	0,99869	0,99874	0,99878	0,99882	0,99886	0,99889	0,99893	0,99896	0,99900
3,1	0,99903	0,99906	0,99910	0,99913	0,99916	0,99918	0,99921	0,99924	0,99926	0,99929
3,2	0,99931	0,99934	0,99936	0,99938	0,99940	0,99942	0,99944	0,99946	0,99948	0,99950
3,3	0,99952	0,99953	0,99955	0,99957	0,99958	0,99960	0,99961	0,99962	0,99964	0,99965
3,4	0,99966	0,99968	0,99969	0,99970	0,99971	0,99972	0,99973	0,99974	0,99975	0,99976
3,5	0,99977	0,99978	0,99978	0,99979	0,99980	0,99981	0,99981	0,99982	0,99983	0,99983

V přírodních vědách jsou téměř všechny hypotézy ověřovány na hladině významnosti $\alpha = 0{,}05$ (5 %), tedy na hladině spolehlivosti 95 %.

$\alpha =$ hladina významnosti $1 - \alpha =$ hladina spolehlivosti

počítáme tedy 95% kvantil

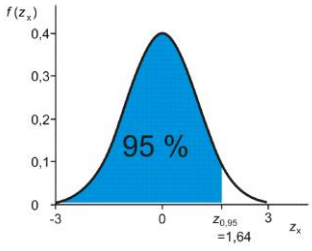
Dva příklady, které jsou zásadní v teorii měření:

1. Veličina ξ má rozdělení N(μ,σ^2). Vypočtěte hodnotu kvantilu, pro kterou platí rovnice:

$P\{\xi \le \mu + k \cdot \sigma\} = 1 - \alpha = 0{,}95$

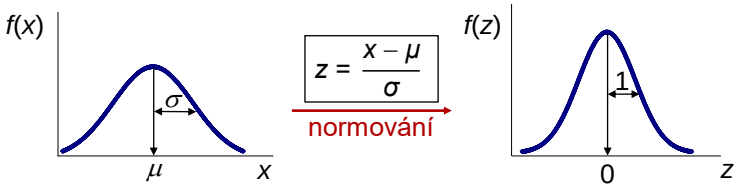
$P\{\zeta \le k\} = 0{,}95 \Rightarrow \Phi(k) = 0{,}95 \Rightarrow k = 1{,}64$

95% jednostranný kvantil je tedy hodnota, kterou 95 % výsledků analýzy nepřekročí.



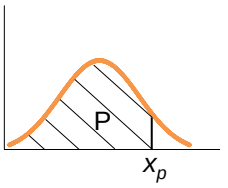
náhodná veličina ξ
normální rozdělení N(μ,σ^2)

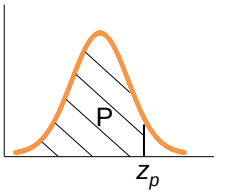
náhodná veličina ζ
normované normální rozdělení N(0,1)



distribuční funkce $F(x_p) = \mathrm{P}\{\xi \le x_p\}$
Excel: NORMDIST(x_p, μ, σ, 1)

distribuční funkce $\Phi(z_p) = \mathrm{P}\{\zeta \le z_p\}$
Excel: NORMSDIST(z_p)





$100 \cdot p$% kvantil: $F(x_p) = p$
Excel: NORMINV(p, μ, σ)

$100 \cdot p$% kvantil: $\Phi(z_p) = p$
Excel: NORMSINV(p)

V přírodních vědách jsou téměř všechny hypotézy ověřovány na hladině významnosti $\boldsymbol{\alpha} = 0{,}05$ (5 %), tedy na hladině spolehlivosti 95 %.

α = hladina významnosti $1 - \alpha$ = hladina spolehlivosti

počítáme tedy 95% kvantil

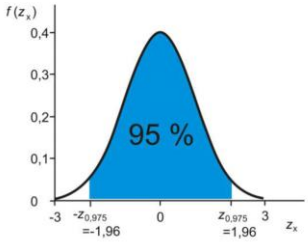
Dva příklady, které jsou zásadní v teorii měření:

2. Veličina ξ má rozdělení N(μ,σ^2). Jaké jsou hodnoty kvantilů, aby byla splněna rovnice:

$\mathrm{P}\{\mu - k \cdot \sigma \le \xi \le \mu + k \cdot \sigma\} = 1 - \alpha = 0{,}95$

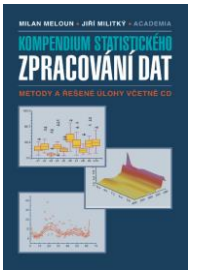
$\mathrm{P}\{-k \le \zeta \le k\} = 0{,}95 \Rightarrow \Phi(k) = 0{,}975 \Rightarrow k = 1{,}96$

95% oboustranný kvantil je tedy hodnota, kterou 95 % výsledků analýzy nepřekročí.



DOPORUČENÁ LITERATURA

Suchánek, M.
Kvalimetrie 16 - Statistické metody v metrologii a analytické chemii
EURACHEM-ČR 2009
http://www.eurachem.cz/publikace-detail.php?clanek=10

Meloun, M.; Militký, J.
Statistické zpracování experimentálních dat, Plus, Praha 1994

Meloun, M.; Militký, J.
Kompendium statistického zpracování dat
Academia, Praha 2002