08 – přednáška – význačná rozdělení psti

Když nám matematika sestaví ty (diskrétní/spojité) pravděpodobnostní modely, často není jednoduché vytvořit danou pstní funkci (u diskrétní veličiny) nebo hustotu psti (u spojité veličiny). Za některými z těchto modelů je tolik práce, že ty modely samotné dostaly svůj název. Dnes si projdeme ty základní z nich – veškerou teorii jsme dělali minulý týden, tedy dnešní přednáška bude jen řadou příkladů; ALE VŠECHNY TYTO PŘÍKLADY MAJÍ SVÉ JMÉNO A DANOU PSTNÍ FUNKCI/HUSTOTU PSTI MUSÍTE BUĎ UMĚT ODVODIT, NEBO SI PAMATOVAT VZOREC!!

Jedná se o jádro celého kursu, věnujte prosím této přednášce pozornost. Vzorce i celý postup popisu najdete na slajdu 08 přednáška.pptx, zde budu jen kreslit obrázky, které ve slajdech chybí.

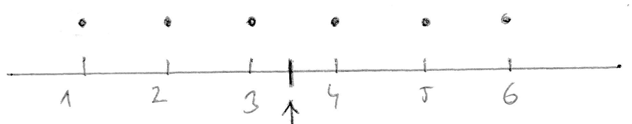
D1: diskrétní rovnoměrné rozdělení psti – slajdy 07, 08, 09, 10

Příklad D1 (slajd 11) X = co padne na kostce .... veličina má rozdělení diskrétní rovnoměrné, značíme $X \sim Ro(1,2,3,4,5,6)$

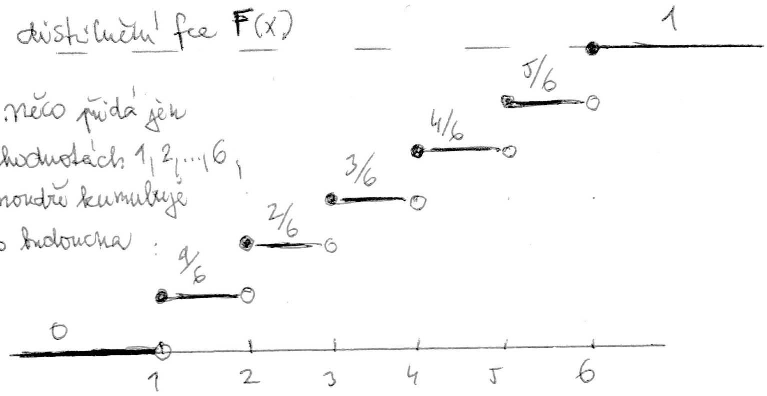
pstní funkce $p(k)$:
všechny psti jsou rovny $\frac{1}{6}$, vlastně se jedná o model klasické psti

distribuční fce $F(x)$

něco přidá jen v hodnotách $1, 2, \dots, 6$, moudře kumuluje do budoucna



$EX = 3{,}5$ ... vyšlo by výpočtem



DX = výpočet ... $= 2{,}9166667$

$\sqrt{DX} = 1{,}707825$

Takhle lze popsat výsledek hodu kostkou, popis působí bizarně, protože otázky ohledně hodu kostkou byste zodpověděli i bez tohoto popisu (nicméně všechny pstní pojmy existují i v tomto případě)

Příklad D2: alternativní rozdělení psti – slajdy 12, 13, 14, 15

X = počet šestek při jednom hodu kostkou ... veličina má rozdělení $Alt(p = \frac{1}{6})$

označujeme $X \sim Alt(\frac{1}{6})$

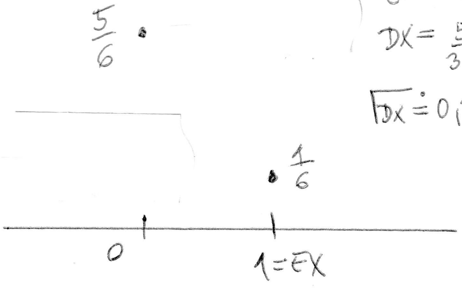
pstní funkce $p(k)$:

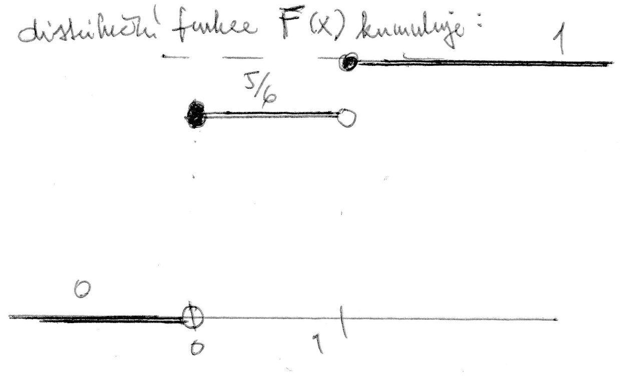
distribuční funkce $F(x)$ kumuluje:

$EX = 1$

$DX = \frac{5}{36}$

$\sqrt{DX} \doteq 0{,}3727$





D3: binomické rozdělení pstí – slajdy 17, 18, 19, 20, 21, 22, 23

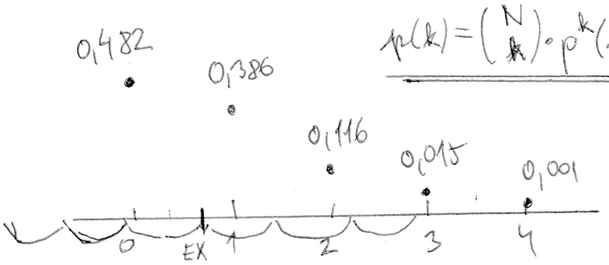
binomické rozdělení už jsme měli, jen si k němu přidáme $F(x)$, EX, DX

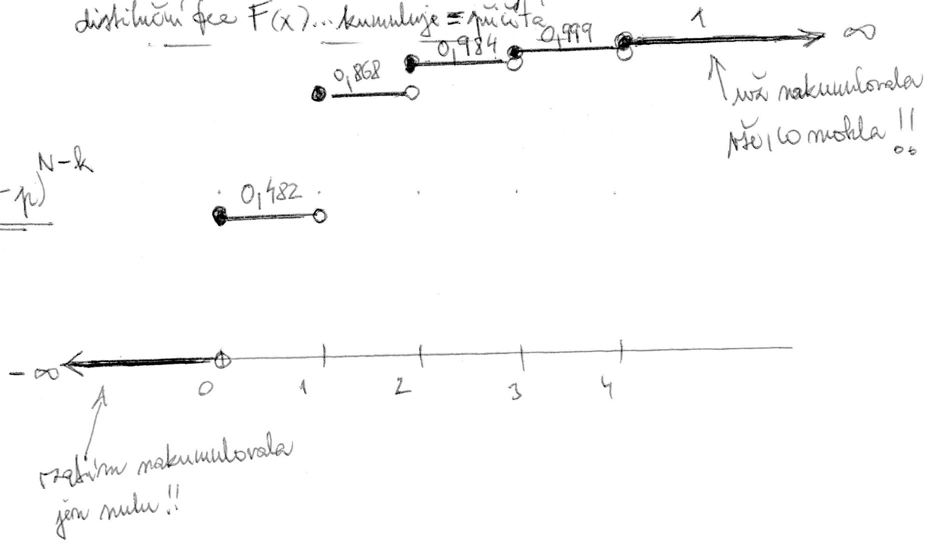
Příklad D3 (slajd 21): X = počet šestek ze 4 hodů kostkou ... značíme $X \sim Bi(N=4, p=\frac{1}{6})$

distribuční fce $F(x)$... kumuluje = přičítá !

pstní fce $p(k)$: výpočet viz skripta BMA3-staré, str. 169

$$p(k) = \binom{N}{k} \cdot p^k (1-p)^{N-k}$$





$$EX = N \cdot p = 4 \cdot \frac{1}{6} = 0{,}6666$$

$$DX = N \cdot p(1-p) = 4 \cdot \frac{1}{6} \cdot \frac{5}{6} = 0{,}5555$$

$$\sqrt{DX} = \underline{0{,}745}$$

$3 \cdot \sqrt{DX} = 2{,}235$, $EX \pm 3 \cdot \sqrt{DX} = 0{,}6667 \pm 2{,}235 = \langle -1{,}568 \, ; \, 2{,}9017 \rangle$ ... počet šestek ze 4 hodů bude nejčastěji naměřený v tomto intervalu

D4: geometrické rozdělení pstí – slajdy 24 až 30

Příklad D4: (vlastně příklad (velmi podobný příkladu) ze prednaska02.pptx, slajd 18), jen si zde přidáme $F(x)$, EX, DX

X = počet hodů potřebný před prvním padnutím šestky na hrací kostce

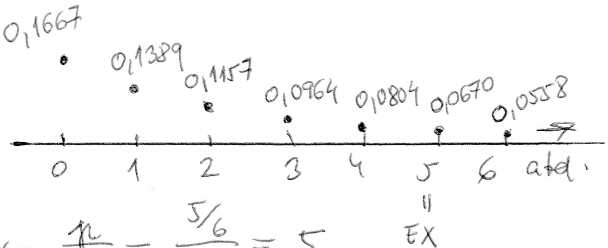
pstní funkce $p(k) = \left(\frac{5}{6}\right)^k \cdot \frac{1}{6}$ pro $k = 0, 1, 2, \ldots$

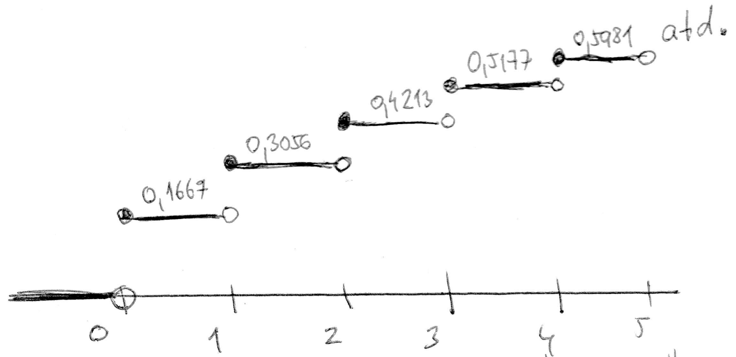
distribuční fce $F(x)$: kumuluje !!

schody se blíží k hodnotě 1

důležité je, že $\sum_{0}^{\infty} p(k) = 1$

$p(k)$:





nekonečně mnoho schodů, ale nestoupají „do nebe", výš k hodnotě 1
(maximální kumulace pstí může být 1)

$$EX = \frac{p}{1-p} = \frac{5/6}{1-5/6} = \underline{\underline{5}}$$

$$DX = \frac{p}{(1-p)^2} = \frac{5/6}{(1/6)^2} = 30$$

$\sqrt{DX} = 5{,}477 \rightarrow EX \pm 3\sqrt{DX} = 5 \pm 16{,}\text{něco} \doteq \langle -11 ; 21 \rangle$ ... zde padne většina měření veličiny X.

Spočteme pst, že před první šestkou padne 4 a více hodů:

$$P(X \geq 4) = p(4) + p(5) + p(6) + \ldots = 1 - p(0) - p(1) - p(2) - p(3) = 1 - 0{,}5177 = \underline{\underline{0{,}4823}}$$

lze spočítat i pomocí distribuční fce: 1 − součet výšek prvních 4 schodů $= 1 - 0{,}5177 = \underline{\underline{0{,}4823}}$

D5. Poissonovo rozdělení psti → čteme francouzsky: „POASONOVO" rozdělení psti !

(odvození vzorců viz BMA3-nove.pdf, str. 204-206 ... nepovinné, ale k ústní části zk. se vzorce naučte)

Př. D5 X = počet zákazníků restaurace (mimo období koronaviru) za jednotku času 1 hod
(= počet nových zákazníků, kteří za 1 hod do restaurace přijdou) v době otevření

značíme $\boxed{X \sim Po(\lambda)}$

λ … udává průměrný počet příchodů za 1 hod

ze zadání je vidět, že se jedná o malinkou restauraci TUKAN v Řečkovicích :-)
kde $\boxed{\lambda = 20 \text{ zákazníků za hod}}$

tento údaj je ještě součástí zadání příkladu (změřili jsme, že průměrně $\lambda = 20$ zák./hod)

pstní funkce: $p(k) = \dfrac{\lambda^k}{k!} \cdot e^{-\lambda}$ pro $k = 0, 1, 2, 3 \dots$ tedy POZOR, definujeme označení $0! = 1$
(setkáte se s ním možná i v analýze 1,2 při rozvoji e^x v nekonečnou řadu pomocí faktoriálu v Taylorově větě)

lze odvodit pomocí šachování se součty nekonečných řad, že

$EX = \lambda$
$DX = \lambda$
} nemusíte odvození umět, jen si pamatujte výsledek

$p(k)$ … pstní funkce, $F(x)$ distribuční funkce … tentokrát vás poprosím, abyste si grafy nakreslili pomocí jazyka R
viz slajd 36 !

S1: exponenciální rozdělení psti – slajdy 39 až 44

Př. S₁: X = doba mezi dvěma nejbližšími příchody zákazníka do restaurace TUKAN, když průměrně přijde $\lambda = 20$ zákazníků za hodinu

(tato veličina je spojitá, proto je označení modelu S1, nikoli D6 !)

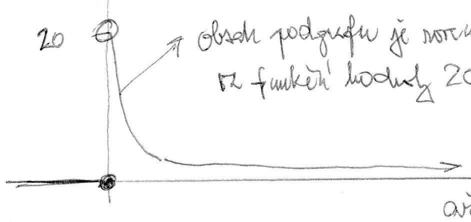
(odvození vzorců viz BMA3-nove.pdf, str. 220-221 ... odvození nepovinné, znát výsledný vzorec je POVINNÉ)

značíme $\boxed{X \sim Exp(\lambda)}$

hustota psti: $f(x) = \begin{cases} 0 \dots & x \le 0 \\ \lambda \cdot e^{-\lambda x} \dots & x > 0 \end{cases}$

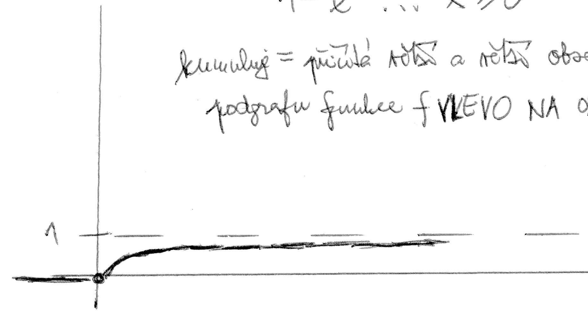
ad př. TUKAN:
$f(x) = \begin{cases} 0 \dots & x \le 0 \\ 20 \cdot e^{-20x} \dots & x > 0 \end{cases}$

distribuční fce $F(x) = \begin{cases} 0 \dots & x \le 0 \\ 1 - e^{-20x} \dots & x \ge 0 \end{cases}$



obsah podgrafu je roven 1, tj. z funkční hodnoty 20 klesá celkem rychle k nule

až do nekonečna, asymptota je osa x



kumuluje = počítá větší a větší obsah podgrafu funkce f VLEVO NA OBRÁZKU když chceme třeba spočítat, že v restauraci TUKAN budou na dalšího zákazníka čekat 5 a více minut,

máme dvě možnosti:

a) $P(X \geq 5\,min) = P(X \geq \frac{1}{12}\,hod) = \int_{1/12}^{\infty} f(t)dt$, ale byli bychom hloupí, kdybychom integrál počítali, když známe funkci $F(x)$.

naší jednotkou času je HODINA!

b) $P(X \geq 5\,min) = P(X \geq \frac{1}{12}\,hod) = P(X \in \langle \frac{1}{12}; \infty)) = F(\infty) - F(\frac{1}{12}) = 1-(1-e^{-20\cdot\frac{1}{12}}) \doteq \underline{0,1889}$

$F(\infty)$ je vždy rovna 1, protože v nekonečnu máme už nakumulováno vše, co bylo možné, čili 100 % měření, vždy psti

S2: Rovnoměrné spojité rozdělení psti – slajdy 47 až 51, značíme $\boxed{X \sim Ro(\langle a;b\rangle)}$

Př. S2 X = doba příchodu člověka, o kterém nevíme žádnou informaci, pouze že v daném intervalu určitě přijde (jakýkoli okamžik příchodu je stejně pravděpodobný)

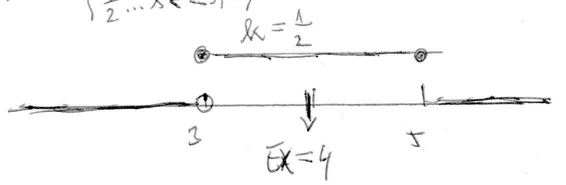
jedná se vlastně o geometrickou pst z přednášky 02.pptx, jen si přidáme $F(x)$, EX, DX

Př.: Víme o někom, že přijde mezi 3 a 5 hodinou odpoledne, X = doba jeho příchodu

hustota psti $f(x)$ je ta nejjednodušší, co může být, konstanta:

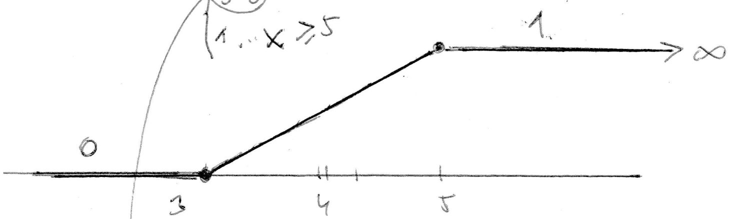
hodnota konst je taková, aby $\int_{-\infty}^{\infty} f(x) = \int_3^5 k\,dx = 1$

$$f(x) = \begin{cases} 0 \dots x \notin \langle 3;5\rangle \\ \frac{1}{2} \dots x \in \langle 3;5\rangle \end{cases}$$



pozor, distribuční funkce $F(x)$ stále kumuluje

$$F(x) = \begin{cases} 0 \dots x \leq 3 \\ \frac{x-3}{5-3} \dots x \in (3;5) \\ 1 \dots x \geq 5 \end{cases}$$



tohleto je „divné" vyjádření – my sami bychom asi našli rovnici přímky ve tvaru $y = \frac{x}{2} - \frac{3}{2}$, ale jedná se stále o tutéž funkci

$$EX = \frac{b+a}{2} = 4$$

DX = spočítali bychom podle vzorce

$$DX = \int_3^5 x^2 \cdot f(x)dx - (EX)^2$$

Jaká je pst, že člověk přijde mezi půl pátou a pátou? Ano, $\frac{1}{4}$. Kromě toho, že odpověď je jednoduchá, mohli bychom psát i:

a) $P(X > 4,5) = \int_{4,5}^{5} \frac{1}{2}\,dx = \underline{\frac{1}{4}} = 0,25$

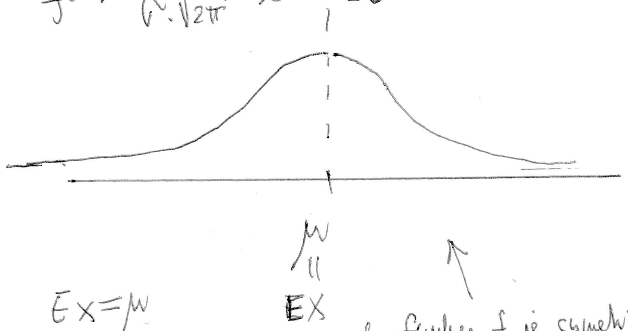
nebo b) $P(X > 4,5) = F(5) - F(4,5) = 1 - \frac{1,5}{2} = \underline{0,25}$
$= F(\infty) - F(4,5) =$

S3: Normální rozdělení pstí – slajdy 54 až 59

označení $\boxed{X \sim No(\mu, \sigma^2)}$

hustota pstí $f(x)$: μ, σ^2 jsou konstanty, které vystupují ve vzorci fce $f(x)$

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



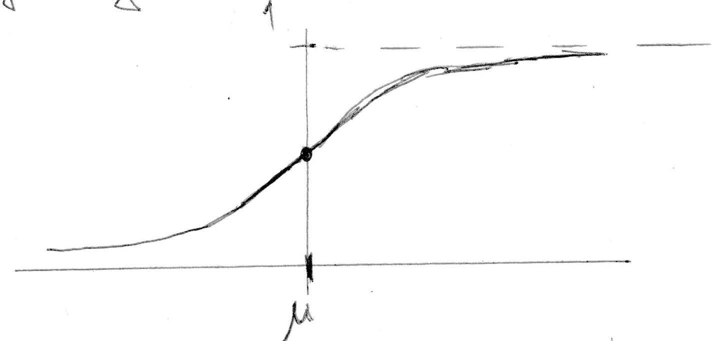
$EX = \mu$

$DX = \sigma^2$

$\sqrt{DX} = \sigma$

graf funkce f je symetrická množina v ose souměrnosti podle přímky $x = \mu$

distribuční funkce $F(x)$ jen kumuluje obsah, jako vždy:



- v bodě μ má F nakumulovánu polovinu obsahu, tj. $F(\mu) = \frac{1}{2}$
- $x = \mu$ je inflexní bod fce F

(pro pamětníky, nebo spíše nepamětníky: v inflexním bodě se F mění z konvexní na konkávní)

Hustota psti $f(x)$ je slavná Gaussova funkce, kterou pan Gauss odvodil pomocí součtů nekonečných řad

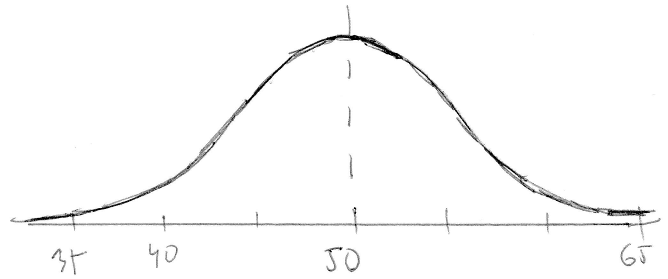
Př. 53 X = výška konkrétního stromu v daném lese;

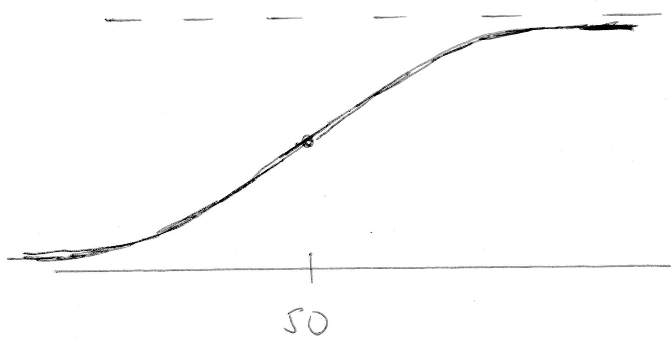
abychom tuto veličinu popsali, musíme zhruba odhadnout $\underset{\mu}{EX}$, $\underset{\sigma^2}{DX}$ (nebo $\sqrt{DX} = \sigma$)

výšky stromů se střední hodnotou $\mu = 50$ m a směrodatnou odchylkou $\sigma = 5$ m (tedy $\sigma^2 = 25$) lze popsat pomocí $No(\mu = 50, \sigma^2 = 25)$:

$$f(x) = \frac{1}{5 \cdot \sqrt{2\pi}} \cdot e^{-\frac{(x-50)^2}{2 \cdot 25}}$$

distribuční funkce $F(x) = \int_{-\infty}^{x} f(t)\,dt$ stále kumuluje více a více, až bude mít vše nakumulováno





Př. Pokud bychom chtěli nyní určit pst, že náhodně vybraný strom v daném lese má výšku menší než 45 m, použijeme $P(X \le 45) = \int_{-\infty}^{45} f(t)\,dt = F(45) - \underbrace{F(-\infty)}_{\text{toto je vždy } 0,} = \dots$

toto je vždy 0, ještě jsme v $(-\infty)$ nic nenakumulovali

nechejme si na příští týden