pst 06 – prověrka

PST A STATISTIKA VYSOKOŠKOLSKY

pst 07 [předn. / cvičení } v tomto textu jsou odkazy na příslušné slajdy – slajdy jsem ratovsil, byly příliš nekvalitní

Následující tři týdny uvedou do popisu modelů psti 03 a 04 na přednáškách v 02 předn.,

slajd 17 (tzv. diskrétní pst – Ω má konečně mnoho elementárních výsledků, nebo výsledků stejně jako množina přirozených čísel
– tyto výsledky mají různou šanci nastat)

slajd 19 (tzv. spojitá pst – $\Omega = \langle a, b \rangle$ nebo $\Omega = \mathbb{R}$ ... můžeme naměřit výsledek z množiny reálných čísel
– tyto elementární výsledky mají různou šanci nastat)

Oba tyto modely se opírají o jediný pojem, a sice pojem distribuční funkce $F(x)$

(pozor, nepleťte si se slovem DISTRIBUTIVNÍ ZÁKON z algebry 1, všimněte si, že mezi přídavnými jmény je rozdíl!!)

Podobně jako teorie fyzikové se snaží o jakýsi sjednocující pohled na svět, i matematika se snažila o sjednocující pohled na náhodnost – a do velké míry se jí to podařilo právě pojmem DISTRIBUČNÍ FUNKCE $F(x)$

(pozor, písmeno F je důležité, protože f bude označovat něco jiného – ale vzpomeňte si na analýzu 2, kde $\int f(x)dx = F(x) + c$ ... tohoto označení vlastně se budeme držet a tedy bude i vztah mezi funkcemi $f(x)$, $F(x)$).

Dalším důležitým pojmem je pojem náhodná veličina X – to není nic nového, to je veličina, jejíž náhodnost měříme a chceme popsat, ovšem formálně se jedná o zobrazení, které elementárním výsledkům $\omega \in \Omega$ (malé omega z množiny Ω) přiřazuje reálná čísla.

Abychom mohli dospět k definici distribuční funkce a sjednocenému pohledu, budeme uvádět dva rozdílné příklady, které popíšeme různými pojmy, a nakonec je těmi sjednocujícími pojmy vyřešíme se sjednocující teorií

Příklad 7.1. Jeden hráč basketbalu má pst 0,7, že trefí trestný hod na koš. Háže šestkrát za sebou; náhodná veličina X = počet tref ze šesti pokusů. Popište tuto náhodnou veličinu matematicky.

Příklad 7.2. Náhodná veličina X měří dobu pozdního příchodu studenta na cvičení PST v minutách po začátku hodiny. Popište tuto náhodnou veličinu matematicky za těchto předpokladů:

– ~~v intervalu~~ $\langle -5$ min, 0 min$\rangle$ přijde 75 % studentů, každý okamžik má stejnou šanci

– v intervalu $\langle 0$ min, 10 min$\rangle$ po začátku šance příchodu rovnoměrně klesá až ~~k~~ k nule

– větší zpoždění než 10 minut vyučující zakázal a studenti si jej nedovolí

---

popišme nyní pst v obou situacích – ten popis bude v každé ze těchto situací jiný: v případě př. 7.1 to bude popis pomocí pstní funkce $p(x)$ (moc se neliší od psti, kterou jsme až dosud používali, ale označujeme ji zde malé písmeno p), v případě př. 7.2 to bude pomocí tzv. hustoty psti $f(x)$ (malé f !!!)

Ad př. 7.1. $\Omega = \{NNTTTN, NTTNTT, TTTNTT, \dots\}$ množina možných sekvencí trefil – netrefil ze šesti pokusů

$\downarrow$ $X = 3$ trefy, $\downarrow$ $X = 4$ trefy, $\downarrow$ $X = 5$ tref … veličina X počítá trefy v každé ze sekvencí

$p(0) = P(X=0) = $ … všimněte si klasického střídání p, P … pst P, že $X=0$, označíme jako funkční hodnotu $p(0)$; velké P je proto sjednocující označení, bude vystupovat v obou příkladech

$= 0{,}3^6 \doteq 0{,}0007$

$p(1) = P(X=1) = 0{,}7 \cdot 0{,}3^5 \cdot$ počet všech sekvencí, ve kterých vystupuje jedno T $= 0{,}7 \cdot 0{,}3^5 \cdot \binom{6}{1} \doteq 0{,}0102$

$p(2) = P(X=2) = 0{,}7^2 \cdot 0{,}3^4 \binom{6}{2} \doteq 0{,}05955$, atd. $p(k) = \binom{6}{k} \cdot 0{,}7^k \cdot 0{,}3^{6-k}$

podobně – blíže viz binomická pst, cvičení 5: $p(3) \doteq 0{,}1852$, $p(4) \doteq 0{,}3241$, $p(5) \doteq 0{,}3025$, $p(6) \doteq 0{,}11765$

Aby tato pstní funkce splňovala vlastnosti psti, musí platit:

1) $\sum_{0}^{6} p(k) = 1$

2) $p(k) \geq 0 \quad \forall k \in \{0,1,2,3,4,5,6\}$

VLASTNOSTI PRAVDĚPODOBNOSTNÍ FUNKCE

3) $P(X \in (a;b\rangle) = \sum_{k \in (a;b\rangle} p(k)$

když chceme určit pst, že veličina X nabude hodnot ze nějakého intervalu, prostě sečteme $p(k)$ pro ty hodnoty k, které $\in (a;b\rangle$.

Z určitého důvodu, který bude patrný při pokusu o sjednocení, bereme interval $(a;b\rangle$ bez levého krajního bodu

---

Ad př. 7.2 U spojité veličiny máme matematický popis rozdílně po částech spojitá funkce zvaná hustota psti.

$\Omega = \langle -5; 10\rangle$ … doba příchodu vzhledem k začátku hodiny; další hodnoty neuvažujeme (tj. NEMOHOU V NAŠEM MODELU NASTAT, zanedbáme je)

náhodná veličina X zde bude identita, protože prvky množiny Ω už jsou reálná čísla vyjadřující dobu příchodu (takže je už do reálných čísel nemusíme zobrazovat).

Nyní musíme VYMYSLET či nějakým způsobem zkonstruovat hustotu psti $f(x)$, že

1) $\int_{-\infty}^{\infty} f(x)\,dx = 1$

2) $f(x) \geq 0$ pro všechna $x \in R$

3) $P(X \in (a;b\rangle) = \int_a^b f(t)\,dt$

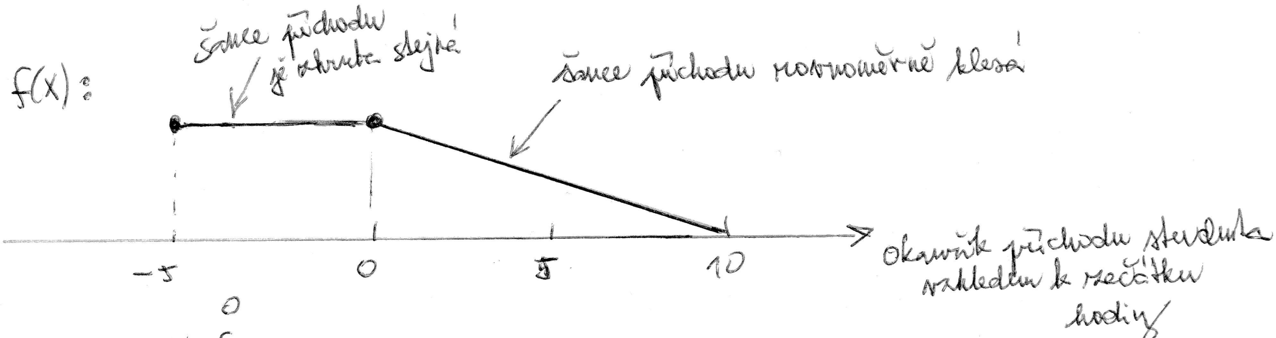
VLASTNOSTI HUSTOTY PRAVDĚPODOBNOSTI

Všimněte si, že u pravděpodobnosti zde nyní budeme funkci f(x) integrovat; nestačí – jen sečítat funkční hodnoty – jednalo by se o součet nekonečně mnoha kladných hodnot, který je vždy roven ∞, tj. prostým sčítáním funkčních hodnot bychom nic rozumného nedostali.

Nebo ještě lépe, ze souvislosti určitého integrálu jako limity součtu obsahů dílčích obdélníčků: když je hodnot x tak moc, že vyplní nějaký interval, tak suma v limitě přechází v určitý integrál

(proto je předmět pst zařazen až za analýzu 2 ;-)

Pokud začínáte připouštět, že u spojitých veličin by se měla pst vyjadřovat pomocí obsahu plochy, mohli bychom navrhnout třeba následující hustotu pravděpodobnosti f(x):



Dokonce ještě víme, že $\int_{-5}^{0} f(x)\,dx = 0{,}75$, a odtud bychom mohli spočítat hodnotu konstantní funkce na tomto intervalu:

$$c[x]_{-5}^{0} = c\cdot\int_{-5}^{0} dx = \int_{-5}^{0} c\,dx = 0{,}75 \quad \ldots\ c \text{ při integraci}$$

$$5c = 0{,}75 \Rightarrow \underline{\underline{c = 0{,}15}}$$

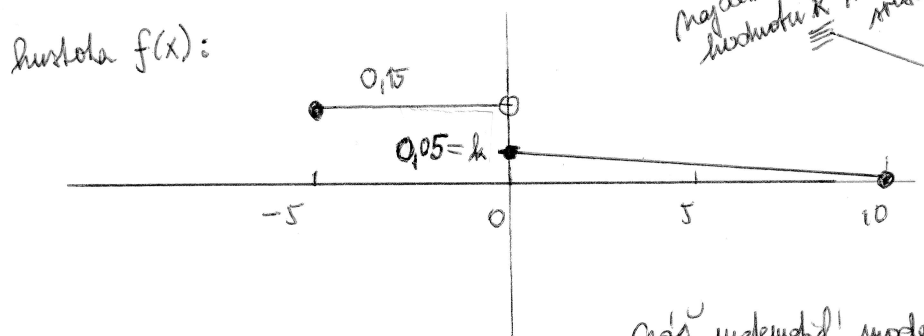
Problém našeho modelu je zatím ten (zkuste přemýšlet se mnou), že obsah Δ na intervalu ⟨0; 10⟩ je také roven 0,75, tedy celkem

$$\int_{-\infty}^{\infty} f(t)\,dt = \int_{-5}^{10} f(t)\,dt = 0{,}75 + 0{,}75 = \underline{1{,}5}$$

protože mimo interval ⟨-5; 10⟩ je hustota rovna 0, můžeme zúžit meze

A TO JE PROBLÉM, PROTOŽE MODEL PSTI MUSÍ BÝT VŽDY NASTAVEN NA 100 % (tj. $\int_{-\infty}^{\infty} f(x)$ se musí rovnat 1)

Musíme tedy trochu poupravit svůj model f(x):



najdeme hodnotu k na svislé ose:

$S_\Delta = 0{,}25$

$\frac{1}{2}\cdot 10\cdot k = 0{,}25$

$k = 0{,}50 : 10 = \underline{0{,}05}$

náš matematický model je hotov!!

(respektive, musíme ještě určit vzorcem rovnici přímky procházející body [0; 0,05] a [10; 0], zde doporučuji vyjádřit si rovnici ve tvaru $y = a \cdot x + 0{,}05$

absolutní člen je vždy roven průsečíku přímky se svislou osou, jak víme ze základů matematiky

a dosadíme bod $[10; 0]$, dostaneme:

$$0 = a \cdot 10 + 0{,}05$$

$$-0{,}005 = -\frac{0{,}05}{10} = a \Rightarrow y = -0{,}005x + 0{,}05$$

a celkem můžeme psát

$$f(x) = \begin{cases} 0 \dots x < -5 \\ 0{,}15 \dots x \in \langle -5; 0) \\ -0{,}005x + 0{,}05 \dots x \in \langle 0; 10 \rangle \\ 0 \dots x > 10 \end{cases}$$

A) Prvním sjednocujícím prvkem v obou příkladech byl začátek vlastnosti 3) obou popisných pojmů:

$$\boxed{P(X \in (a;b\rangle) = \begin{cases} \text{diskr.: } = \sum_{k \in (a;b\rangle} p(k) \\ \text{spoj.: } = \int_a^b f(t)\,dt \end{cases}}$$

výpočet psti, že veličinu X naměříme v intervalu $(a;b\rangle$

zde se ovšem vzoreček větví, pamatujte si:

- u diskrétní veličiny píšeme – sumu
  – pstní funkci $p(k)$
- u spojité veličiny píšeme – integrál
  – hustotu psti $f(x)$

B) Dalším sjednocujícím pojmem v obou situacích bude STŘEDNÍ HODNOTA veličiny X. Budeme ji označovat písmenem E z anglického slova EXPECTED.

EX = expected value of X = očekávaná hodnota veličiny X

Zjednodušeně řečeno, EX je jakýsi „teoretický průměr" dané veličiny X, neboli průměrná hodnota veličiny X, kterou bychom spočítali z velkého počtu jejích měření, kdyby se chovala přesně podle našeho teoretického popisu $p(k)$/nebo $f(x)$. !

Pokud střední hodnota má fungovat jako průměr, hledejme cestu k jejímu výpočtu ve vzorci pro průměr měření ze začátku semestru (z popisné statistiky):

$$\boxed{\bar{x} = \frac{1}{n}\sum_1^n x_i = \frac{1}{n}\sum_{x_i} n_i \cdot x_i = \sum_{x_i} \frac{n_i}{n} \cdot x_i}$$

... tyto tři vzorce pro výpočet průměru jsme si řekli na začátku semestru

sčítáme přes třídy četností, tj. přes všechna RŮZNÁ x_i

s hodnotou $\frac{1}{n}$ „skočíme" do sumy

tohleto jsou relativní četnosti, které lze chápat jako jakési „naměřené" psti.

Kdybychom místo nich dosadili teoretické psti, dostaneme teoretický průměr !!

Tato úvaha nám pomůže pochopit a zapamatovat si vzorec pro EX :

$$EX = \begin{cases} \text{diskr.:} & \sum_{x_i} p(x_i)\cdot x_i \\ \text{spoj.:} & \int_{\Omega} f(x)\cdot x \, dx \end{cases}$$

... měli byste chápat z předchozí strany

... vlastně totéž, ale převedeno na spojitý případ INFINITEZIMÁLNÍHO POČTU (sčítáme nekonečně mnoho nekonečně malých hodnot):

- místo sumy – píšeme určitý integrál přes interval Ω
- místo $p(k)$ – píšeme hustotu $f(x)$
- místo x_i – index i vynecháváme, hodnoty bereme „velmi nahusto naházené" z celého intervalu Ω

Ad příklad 7.1: průměrný = očekávaný počet tref ze šesti hodů:

$$EX = \sum_{0}^{6} p(x_i)\cdot x_i = 0{,}0007\cdot 0 + 0{,}0102\cdot 1 + 0{,}05955\cdot 2 + 0{,}1852\cdot 3 + 0{,}3241\cdot 4 + 0{,}3025\cdot 5 + 0{,}11765\cdot 6 \doteq \underline{4{,}2}$$

košů trefených ze šesti pokusů

Ad příklad 7.2: průměrný okamžik příchodu studenta vzhledem k času 0 = začátek hodiny:

$$EX = \int_{-5}^{10} f(x)\cdot x\,dx = \int_{-5}^{0} 0{,}15\cdot x\,dx + \int_{0}^{10} (-0{,}005x + 0{,}05)\cdot x\,dx =$$

před integrací roznásobte !!!!!!!!!

$$= \left[0{,}15\frac{x^2}{2}\right]_{-5}^{0} + \left[-0{,}005\cdot\frac{x^3}{3} + 0{,}05\cdot\frac{x^2}{2}\right]_{0}^{10} = -0{,}15\cdot 12{,}5 - \frac{5}{3} + \frac{5}{2} = -1{,}0417$$

průměrný student přijde asi 1,04 minut před začátkem hodiny

(C) Třetím sjednocujícím pojmem, který nás zajímá v obou situacích, bude tzv. rozptyl veličiny X, který, jak už název napovídá, bude udávat míru rozptýlení veličiny X kolem své střední hodnoty. Budeme označovat písmenem D z anglického DISPERSION = rozptýlení.

(v některé literatuře bývá označováno jako VAR(X), z anglického VARIANCE = rozptýlenost, variabilita)

Cesta k výpočtu bude opět spočívat v nahrazení „naměřených" proti teoretickým, nyní ve vzorci rozptylu měření

$$\boxed{s^2 = \frac{1}{n}\sum_{1}^{n}(x_i - \bar{x})^2 = (\dots) = \left(\sum_{x_i} \frac{n_i}{n}\cdot x_i^2\right) - (\bar{x})^2}$$

... už jsme si říkali na začátku semestru? Teď si nejsem jistý

relativní četnosti

(a ještě jedno nahrazení ve vzorci provedeme, místo skutečného průměru $\bar{x}$, který nemáme k dispozici, budeme psát „teoretický průměr" EX)

tenhle mezikrok je

úprava vzorce umocněním závorky na druhou a využitím četností n_i

$$= \left(\frac{1}{n}\sum_{1}^{n} x_i^2\right) - (\bar{x})^2 = \left(\sum_{x_i} \frac{n_i}{n}\cdot x_i^2\right) - (\bar{x})^2$$

různé hodnoty, suma přes třídy četností

$$DX = \begin{cases} \text{diskr.: } \left(\sum_{x_i} p(x_i)\cdot x_i^2\right) - (EX)^2 \\ \text{spoj.: } \int_\Omega f(x)\cdot x^2\,dx - (EX)^2 \end{cases}$$

... pozor, studenti zapomínají na to odečtení $(EX)^2$

... vlastně totéž, ale převedeno na spojitý případ

jednoduchost vzorců je opět dána tím, že při převodu na spojitý případ

- sumu nahrazujeme integrálem
- $p(x_i)$ nahrazujeme pomocí $f(x)$
- x_i přechází spojitě na x v intervalu Ω

No a abychom mohli vyjádřit rozptýlení ve stejných, neumocněných jednotkách, zavádíme pojem

$\sigma = \sqrt{DX}$ ... směrodatná odchylka

Ad př. 7.1. Kvadratické odchýlení hodnot X od střední hodnoty EX vypočteme

$$DX = \sum_0^6 p(x_i)\cdot x_i^2 - 4{,}2^2 = 0{,}0007\cdot 0^2 + 0{,}0102\cdot 1^2 + 0{,}05955\cdot 2^2 + 0{,}1852\cdot 3^2 + 0{,}3241\cdot 4^2 + 0{,}3025\cdot 5^2 + 0{,}11765\cdot 6^2 - 4{,}2^2 = 1{,}26$$

řecké písmeno sigma

$$\sigma = \sqrt{DX} = \sqrt{1{,}26} \doteq 1{,}1225$$

Z důvodu, když vyrážím na 14 dní, naneseme $\sqrt{DX}$ třikrát na každou stranu od střední hodnoty 4,2 a říkáme: většina měření veličiny (= většina hodnot veličiny), asi 99 %, leží v intervalu $EX \pm 3\cdot\sqrt{DX} = 4{,}2 \pm 3\cdot 1{,}1225 = \langle 0{,}525;\ 7{,}875\rangle$

většinou naměříme 1 až 6 trefů, že se ze šesti pokusů hráč netrefí vůbec je velmi málo pravděpodobné

Ad př. 7.2 Kvadratické odchýlení hodnot X od střední hodnoty EX vypočteme

$$DX = \int_{-5}^{10} f(x)\cdot x^2\,dx - (-1{,}0417)^2 =$$

$$= \int_{-5}^{0} 0{,}15\cdot x^2\,dx + \int_0^{10} (-0{,}005x + 0{,}05)\cdot x^2\,dx - 1{,}085 =$$

nejprve roznásobte !!!!!!!!!!!!!!!!

čtvrté MINUS je z Newton-Leibnizova vzorce

Taky bych měl zapomenout

$$= \left[0{,}15\cdot\frac{x^3}{3}\right]_{-5}^{0} - \left[0{,}005\cdot\frac{x^4}{4}\right]_0^{10} + \left[0{,}05\cdot\frac{x^3}{3}\right]_0^{10} - 1{,}085 = +\frac{0{,}15\cdot 125}{3} - \frac{50}{4} + \frac{50}{3} - 1{,}085 =$$

$$= 9{,}331667$$

Tedy $\sigma = \sqrt{DX} = \sqrt{9{,}331667} = 3{,}0548$

Většina příchodů studenta vzhledem k 0 = začátek hodiny

leží v intervalu $EX \pm 3\cdot\sqrt{DX} = -1{,}0417 \pm 3\cdot 3{,}0548 = \langle -10{,}206\,\text{min};\ 8{,}1227\,\text{min}\rangle$

D ten poslední a nejdůležitější pojem (distribuční funkce) jsme nestihli, necháme do cvičení po obědě.