MATEMATIKA S STATISTIKO

UNIVERZITETNA ŠTUDIJSKA PROGRAMA LABORATORIJSKA BIOMEDICINA IN KOZMETOLOGIJA

1. LETNIK

OSNOVE TEORIJE VERJETNOSTI

V tednu je sedem dni. Kolikšna je verjetnost, da bo jutri petek?

Verjetnost, da sta na letalu dve bombi je neprimerno manjša kot verjetnost, da je na letalu ena bomba. Za koliko se zmanjša verjetnost, da je na letalu bomba, če eno bombo prinesemo s seboj?

Polovici razreda se pouk zaključi ob dvanajstih, polovici pa ob dveh. Torej se jim pouk v povprečju zaključi ob sedmih ($(12+2)/2=7$).

Kolikšna je verjetnost, da pri 100 metih kovanca dobimo 50 cifer? 1, 0.5 ali kaj drugega?

Statistično je dokazano, da večja, ko je teža mladostnika, višja je njegova stopnja izobrazbe. Torej čim več jejte!

Teorija verjetnosti obravnava situacije, ki jim pravimo poskusi in pri katerih je izid odvisen od naključja.

Prostor izidov je množica vseh izidov nekega poskusa.

Med vožnjo na faks pelje študent mimo treh semaforjev. Pri vsakem se bodisi ustavi (R) ali pa pelje brez ustavljanja (Z). Prostor izidov je { ZZZ , ZZR , ZRZ , RZZ , ZRR , RZR , RRZ ,RRR }.

Uvrstitev tekmovalca na kolesarski dirki 'Franja' je izid pri poskusu - tekmi - in za prostor izidov vzamemo množico $\{1,2,...,N\}$ kjer je N število udeležencev. Ker se število udeležencev iz leta v leto spreminja, je bolj smiselno vzeti za prostor izidov množico vseh naravnih števil $\{1,2,3,...\}$

Letna količina padavin v nekem kraju je zelo odvisna od naključja. Če jo gledamo kot izid poskusa je prostor izidov množica vseh pozitivnih realnih števil $\{t \mid t \geq 0\}$.

Podmnožicam prostora izidov pravimo dogodki.

Dogodek, da se študent ustavi pri drugem semaforju je {ZRZ,ZRR,RRZ,RRR}.
Dogodek,da se kolesar uvrsti med prvih deset je $\{1,2,3,4,5,6,7,8,9,10\}$.
Interval $[500,1200]$ ustreza dogodku, da pade med 500 in 1000 milimetrov dežja.

Na dogodkih izvajamo iste operacije kot na množicah (unija, presek,komplement...), le da jih drugače imenujemo.

SLOVAR

element	→	izid	ZZZ,ZZR,ZRZ,RZZ,ZRR,RZR,RRZ,RRR
množica	→	dogodek	**A je dogodek, da se študent ustavi na prvem, B pa, da se ustavi na drugem semaforju:** A={RZZ,RZR,RRZ,RRR}, B={ZRZ,ZRR,RRZ,RRR}
unija	→	vsota	**$A+B$ je dogodek, da se študent ustavi na prvem, ali na drugem semaforju ali pa na obeh:** $A+B$={RZZ,RZR,ZRZ,ZRR,RRZ,RRR}
presek	→	produkt	**AB je dogodek, da se študent ustavi na prvem in na drugem semaforju:** AB={RRZ,RRR}
komplement	→	nasprotni dogodek	**$\bar{A}$ je dogodek, da se študent ne ustavi na prvem semaforju $\bar{A}$ ={ZRZ,ZZR,ZZZ,ZRR}.**
prazna množica	→	nemogoč dogodek	$N=\emptyset$
cela množica	→	gotov dogodek	G={ZZZ,ZZR,ZRZ,RZZ,ZRR,RZR,RRZ,RRR}
tuji množici	→	nezdružljiva dogodka	**dogodka sta nezdružljiva, če je njun produkt nemogoč dogodek: npr., da se študent hkrati ustavi in ne ustavi na prvem semaforju.**

Verjetnost je funkcija, ki vsakemu dogodku A priredi število $P(A)\in[0,1]$ tako, da velja:

- $P(G)=1$

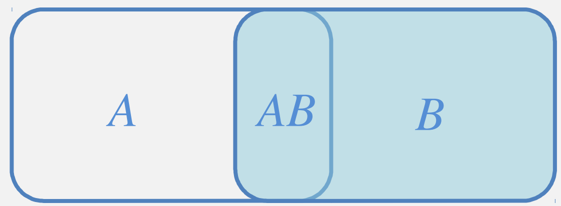
Verjetnost gotovega dogodka je 1.

- $AB=N \;\Rightarrow\; P(A+B)=P(A)+P(B)$

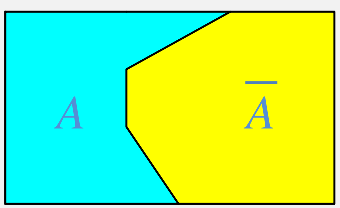
Verjetnost vsote nezdružljivih dogodkov je enaka vsoti njunih verjetnosti

Računska pravila:

- $P(A+B)=P(A)+P(B)-P(AB)$



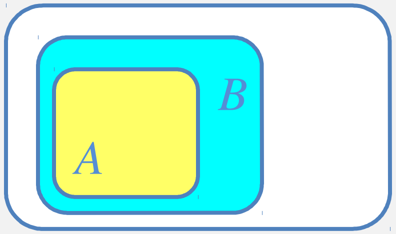
- $P(\bar{A})=1-P(A)$



$$P(A)+P(\bar{A})=P(A+\bar{A})=P(G)=1$$

- $P(N)=0$

- $A\subseteq B \;\Rightarrow\; P(A)\le P(B)$



$$P(B)=P(A)+P(B-A)\ge P(A)$$

Klasična definicija verjetnosti

Če ima poskus končno število enako verjetnih izidov, potem je

$$P(A) = \frac{\text{število izidov v dogodku } A}{\text{število vseh izidov}}$$

Statistična definicija verjetnosti

Frekvenca dogodka A pri n ponovitvah poskusa je

$$\frac{\text{število poskusov z izidom } A}{n}$$

$P(A)$ je limita frekvenc dogodka A pri velikem številu ponovitev poskusa.

Naj bo pri metu kocke A dogodek, da pade sodo število pik.

Po klasični definicija: $P(A)=½$, ker je $A=\{2,4,6\}$ v množici izidov $\{1,2,3,4,5,6\}$, za katere privzamemo, da so enako verjetni.

Statistična definicija: $P(A)$ je frekvenca metov s sodim številom pik pri velikem številu metov kocke.

Po 1000 metih kovanca dobimo 700 grbov

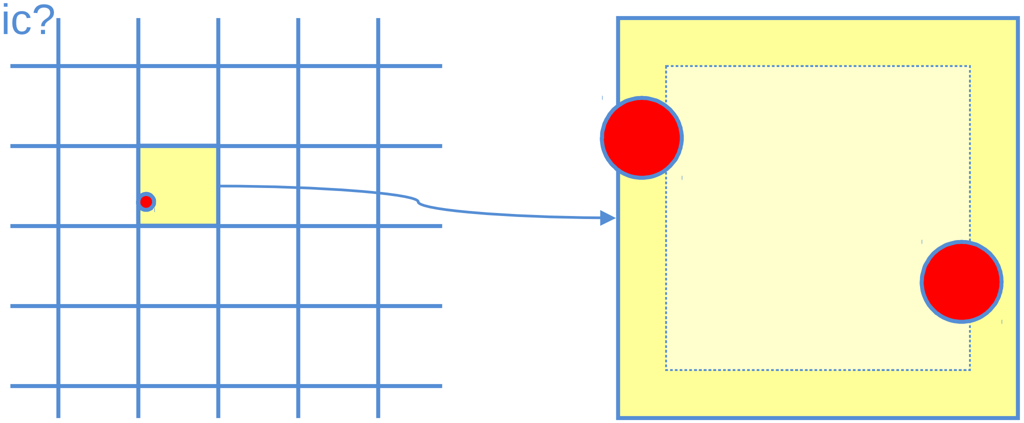
klasično → pri 1001. metu sta oba izida enako verjetna

statistično → pri 1001. metu je bolj verjetno, da pade grb

Za uporabo je odločilna verjetnost, 'izmerjena' po statistični definiciji. Klasična definicija je lahko kvečjemu dober približek.

Včasih izidov ne moremo prešteti, lahko pa jih predstavimo geometrično. V tem primeru je klasična definicija verjetnosti $P(A)$ opredeljena kot razmerje med velikostjo (dolžino, ploščino, prostornino...) množice A in velikostjo množice vseh izidov.

Kovanec s premerom 2 cm vržemo na tla pokrita s ploščicami s stranico 10 cm. Kolikšna je verjetnost dogodka A, da kovanec ne pade na stik dveh ploščic?



$P(A)=8^2/10^2=0.64=64$ %

Tudi pri tem pristopu se klasična in statistična definicija razlikujeta:

Kolikšna je verjetnost, da se bo voznik ustavil pri nekem semaforju?

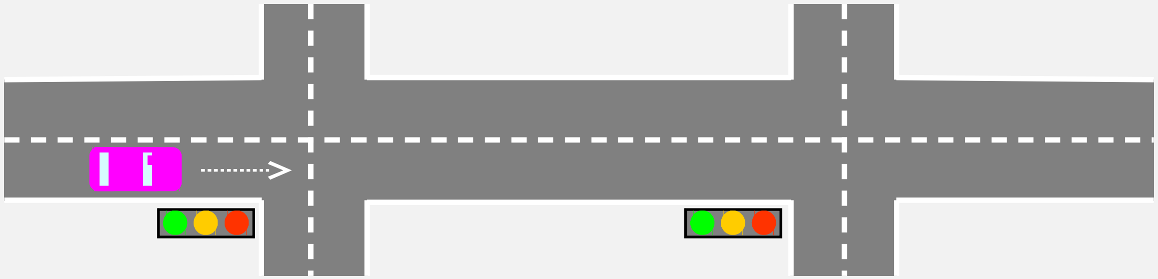
Klasično: če je r čas trajanja rdeče luči na semaforju, z pa čas trajanja zelene luči, potem je verjetnost, da se voznik ustavi enaka $r/(r+z)$.

Statistično: verjetnost je razmerje med številom ustavljanj in številom vseh voženj pri dovolj velikemu številu voženj.

POGOJNA VERJETNOST

Voznik se vsak dan vozi po isti poti in se jezi, da na nekem semaforju skoraj vsakič pripelje na rdečo luč. Sčasoma ugotovi, da v povprečju le enkrat na vsakih pet voženj pripelje na zeleno. Ali lahko sklepa, da je rdeči interval štirikrat daljši od zelenega?

Po opazovanju semaforja ugotovi, da sta rdeča in zelena prižgani enako dolgo časa. Kako je potem mogoče, da vedno pripelje na rdečo?



Izkaže se, da na svoji poti pelj mimo dveh semaforjev. Na prvega pripelje povsem naključno, mimo pa gre le pri zeleni luči. Semaforja sta pa tako (ne)vsklajena, da se v času, ko pripelje do drugega ravno prižge rdeča luč.

Izid na drugem semaforju je **pogojen** z izidom na prvem semaforju.

A, B dogodka ($P(B) \neq 0$)

Pogojna verjetnost dogodka A pri pogoju B je delež dogodka A med poskusi, pri katerih se zgodi dogodek B.

$$P(A \mid B) = \frac{P(AB)}{P(B)}$$

Pri kontroli kakovosti v tovarni 30% izdelkov ocenijo kot prvovrstne, 50% kot drugovrstne, ostale pa kot neuporabne. V trgovino seveda pošljejo le uporabne izdelke. Kolikšna je verjetnost, da je naključno izbrani izdelek v trgovini prvovrsten?

A: izdelek je prvovrsten
U: izdelek je uporaben
Zanima nas $P(A|U)$.

$\longrightarrow$

$P(AU) = P(A) = 30\ \%$
$P(U) = 80\ \%$
$P(A|U) = 30/80 = 0.375 = 37.5\ \%$

S pomočjo pogojne verjetnosti izrazimo verjetnost, da se dva dogodka zgodita hkrati:

$$P(AB) = P(A \mid B) \cdot P(B) = P(A) \cdot P(B \mid A)$$

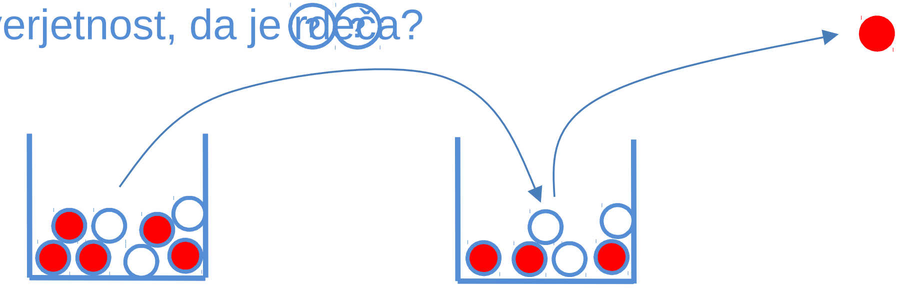
Iz vrečke, v kateri so 3 rdeče in 2 beli kroglici, zaporedoma brez gledanja izvlečemo dve kroglici . Kolikšna je verjetnost, da sta obe rdeči?

A: prvič izvlečemo rdečo kroglico
B: drugič izvlečemo rdečo kroglico

$$P(A) = \frac{3}{5}, \quad P(B \mid A) = \frac{2}{4} \longrightarrow P(AB) = \frac{3}{5} \cdot \frac{2}{4} = \frac{3}{10}$$

S pomočjo pogojne verjetnosti lahko izračunamo verjetnost dogodka, ki je rezultat dvo- ali večstopenjskega poskusa:

Iz škatle s petimi rdečimi in tremi belimi kroglicami na slepo prenesemo dve kroglici v škatlo, v kateri so tri rdeče in tri bele kroglice. Nato iz druge škatle izvlečemo eno kroglico. Kolikšna je verjetnost, da je rdeča?



možnosti na 1. koraku

možnosti na 2. koraku

$P = \frac{3}{8} \times \frac{2}{7} = \frac{6}{56}$ prenesemo dve beli kroglici

$P = \frac{5}{8} \times \frac{4}{7} = \frac{20}{56}$ prenesemo dve rdeči kroglici

$P = \frac{3}{8} \times \frac{5}{7} + \frac{5}{8} \times \frac{3}{7} = \frac{30}{56}$ prenesemo eno rdečo in eno belo kroglico

$P = \frac{3}{8}$, $P = \frac{5}{8}$, $P = \frac{4}{8}$

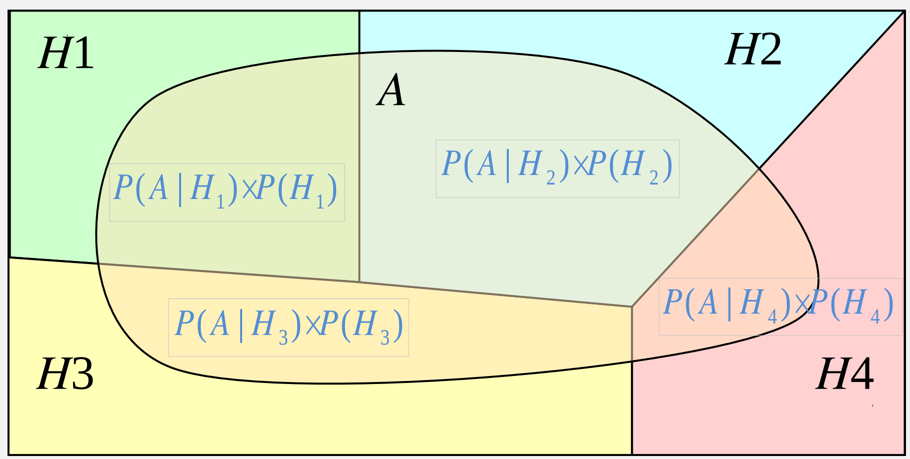
izvlečemo belo kroglico

izvlečemo rdečo kroglico

$$P(R) = \frac{6}{56} \times \frac{3}{8} + \frac{20}{56} \times \frac{5}{8} + \frac{30}{56} \times \frac{4}{8} = \frac{238}{448} \approx 0.531$$

Splošni postopek: najprej določimo vse možnosti na prvem koraku, $H1, H2, \ldots, Hn$ in njihove verjetnosti $P(H1), P(H2), \ldots, P(Hn)$.

Nato določimo pogojne verjetnosti, da se dogodek A zgodi pri vsaki od teh možnosti $P(A|H1), P(A|H2), \ldots, P(A|Hn)$.



$$P(A) = P(A|H_1)\times P(H_1) + P(A|H_2)\times P(H_2) + \ldots + P(A|H_n)\times P(H_n)$$

formula o popolni verjetnosti

Včasih želimo na podlagi opaženega izida ugotoviti verjetnosti izidov na prvi stopnji.

Iz škatle, v kateri je pet kovancev na slepo izberemo enega, ga vržemo trikrat in vsakič dobimo cifro, na kar kovance in škatlo pospravimo.

Čez čas spet pogledamo kovance in ugotovimo, da je en ponarejen in ima na obeh straneh cifro!

Kolikšna je verjetnost, da smo metali ponarejen kovanec?

$$P(B \mid A) = P(A \mid B) \times \frac{P(B)}{P(A)}$$

$$P(\text{ponarejen kovanec} \mid \text{tri cifre}) = P(\text{tri cifre} \mid \text{ponarejen kovanec}) \times \frac{P(\text{ponarejen kovanec})}{P(\text{tri cifre})}$$

$$P(H_i \mid A) = \frac{P(A \mid H_i) \times P(H_i)}{P(A \mid H_1) \times P(H_1) + \ldots + P(A \mid H_n) \times P(H_n)}$$

Bayesova formula

$$P(\text{ponarejen kovanec}) = \frac{1}{5} \qquad P(\text{tri cifre} \mid \text{ponarejen kovanec}) = 1$$

$$P(\text{pošten kovanec}) = \frac{4}{5} \qquad P(\text{tri cifre} \mid \text{pošten kovanec}) = \frac{1}{8}$$

$$P(\text{ponarejen kovanec} \mid \text{tri cifre}) = \frac{1 \times \frac{1}{5}}{1 \times \frac{1}{5} + \frac{1}{8} \times \frac{4}{5}} = \frac{2}{3}$$

Bayesovo formulo pogosto uporabljamo pri vrednotenju testov zdravil in presejalnih testov

V Veliki Britaniji je incidenca raka dojke v starostni skupini 40-49 let okoli 160 primerov na 100.000 prebivalcev. Zanesljivost standardnega presejalnega testa (mamografije) v primeru zdrave osebe je okoli 85%, v primeru obolele pa okoli 90%.

Če je mamografski test pozitiven, kolikšna je verjetnost, da ima testiranka raka?

$$P(\text{obolela}) = \frac{160}{100.000} = 0.16\% \qquad P(\text{test pozitiven} \mid \text{obolela}) = 90\%$$

$$P(\text{zdrava}) = 99.84\% \qquad P(\text{test pozitiven} \mid \text{zdrava}) = 15\%$$

$$P(\text{zdrava} \mid \text{test pozitiven}) = \frac{0.15 \times 0.9984}{0.15 \times 0.9984 + 0.90 \times 0.0016} = 99.04\%$$

99% testirank s pozitivnim mamografskim testom nima raka!

ODVISNOST IN NEODVISNOST DOGODKOV

A je neodvisen od B, če je $P(A|B)=P(A)$.

$$\left[\quad P(A|B)=P(A) \Longleftrightarrow P(AB)=P(A)P(B) \quad\right]$$

$$\boxed{A \text{ in } B \text{ sta neodvisna} \Longleftrightarrow \quad P(AB)=P(A).P(B)}$$

Iz škatle, v kateri imamo 7 polnih in 3 prazne baterije naključno vzamemo dve. Naj bo A dogodek, da je prva baterija polna, B pa dogodek, da je druga baterija polna. Ali sta dogodka A in B neodvisna?

$$\left.\begin{matrix} P(A)=\dfrac{7}{10},\ P(B|A)=\dfrac{6}{9} \\ P(\bar{A})=\dfrac{3}{10},\ P(B|\bar{A})=\dfrac{7}{9} \end{matrix}\right\} \Rightarrow P(B)=\frac{7}{10}\cdot\frac{6}{9}+\frac{3}{10}\cdot\frac{7}{9}=\frac{7}{10}$$

$$P(B)\neq P(B|A) \quad \Rightarrow \quad B \text{ ni neodvisen od } A$$

Zgled razkrije razliko med izbiranjem vzorca z vračanjem in izbiranjem brez vračanja.

SLUČAJNE SPREMENLJIVKE

Če vržemo dve kocki, dobimo za vsoto pik število med 2 in 12, vendar te vsote ne moremo vnaprej napovedati, saj je odvisna od naključja. Podobno velja za število šestic v dveh metih.

Primeri količin odvisnih od naključja:
- število potnikov mestnega avtobusa, ki izstopijo na postaji
- število metov potrebnih, da igralec z določene razdalje zadane koš
- število bonbonov v vrečki
- življenjska doba žarnice
- teža hlebca kruha

……

Slučajna spremenljivka je funkcija, katere vrednosti so odvisne od naključja.

Določata jo:

- zaloga vrednosti = nabor vrednosti, ki jih lahko zavzame, in
- porazdelitev = verjetnost, da zavzame eno ali več vrednosti iz zaloge

Pri metu dveh kock je možnih 36 različnih in enako verjetnih izidov. Če z V označimo vsoto pik, dobimo slučajno spremenljivko z vrednostmi 2,…,12 in porazdelitvijo:

$$P(V=2)=P(V=12)=\frac{1}{36}$$

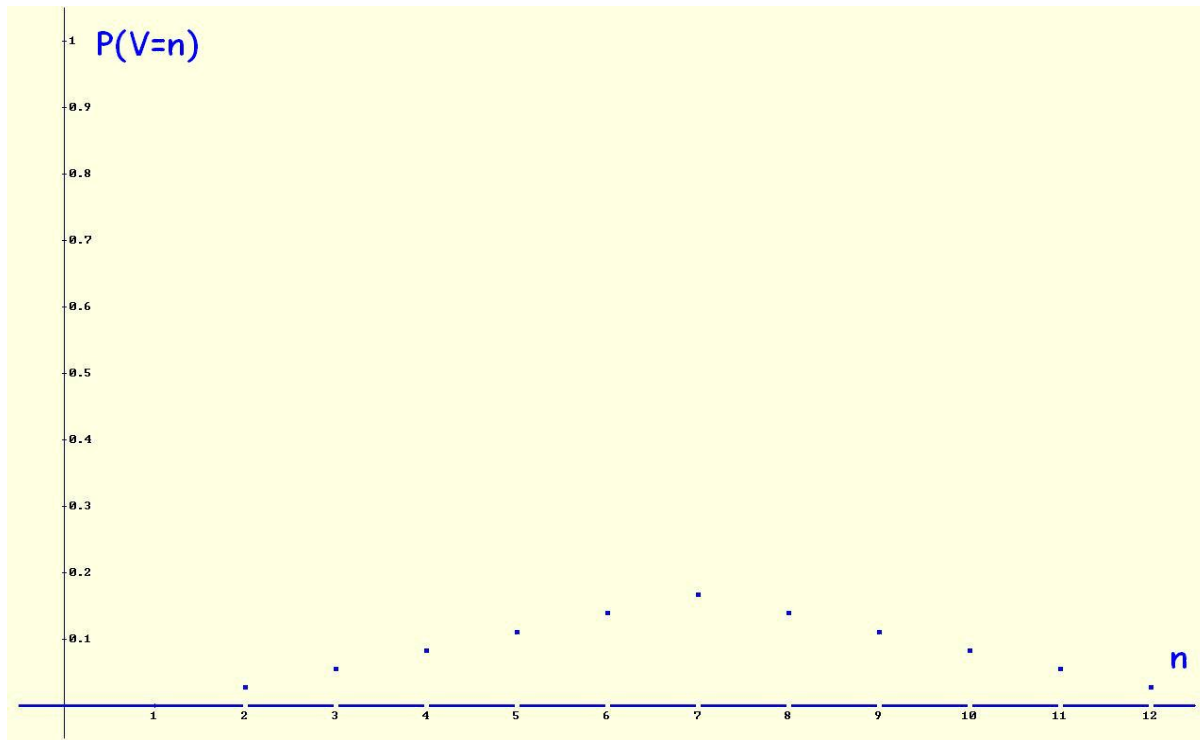
$$P(V=3)=P(V=11)=\frac{2}{36}$$

$$P(V=4)=P(V=10)=\frac{3}{36}$$

$$P(V=5)=P(V=9)=\frac{4}{36}$$

$$P(V=6)=P(V=8)=\frac{5}{36}$$

$$P(V=7)=\frac{6}{36}$$



Slučajna spremenljivka X je diskretna, če zavzame končno ali največ štetno mnogo vrednosti $x1$, $x2$, $x3$,...

Njena porazdelitev je povsem določena s funkcijo $pX(xi)=P(X=xi)$.
Običajno naštejemo le neničelne vrednosti: $p(x1), p(x2), p(x3)$,...

Velja: $0 \leq p(x_i) \leq 1 \quad \sum_i p(x_i) = 1$

Primeri diskretnih porazdelitev

enakomerna porazdelitev

- X zavzame vrednosti $x1$, $x2$,..., xn
- $pX(x)=1/n$, če je $x \in \{x1, x2, \ldots xn\}$

$pX(x)=0$, sicer

Število pik pri metu kocke je enakomerno porazdeljeno:
zaloga vrednosti je $\{1,2,3,4,5,6\}$, vse vrednosti imajo verjetnost $1/6$.

binomska porazdelitev

Poskus ponovimo n-krat: naj bo vsakič verjetnost uspeha enaka p (in verjetnost neuspeha 1-p).

(npr. žogo vržemo 10-krat na koš; zadanemo z verjetnostjo 70%)

Slučajna spremenljivka B naj bo število uspešnih poskusov. Kako je porazdeljena?

(tj. kolikšna je verjetnost, da bomo imeli k zadetkov?)

- Zaloga vrednosti spremenljivke B je $\{0,1,2,\ldots,n\}$
- Privzamemo, da so izidi poskusov medsebojno neodvisni.

Obstaja $\binom{n}{k}$ različnih zaporedij k uspešnih in $(n$-$k)$ neuspešnih poskusov;

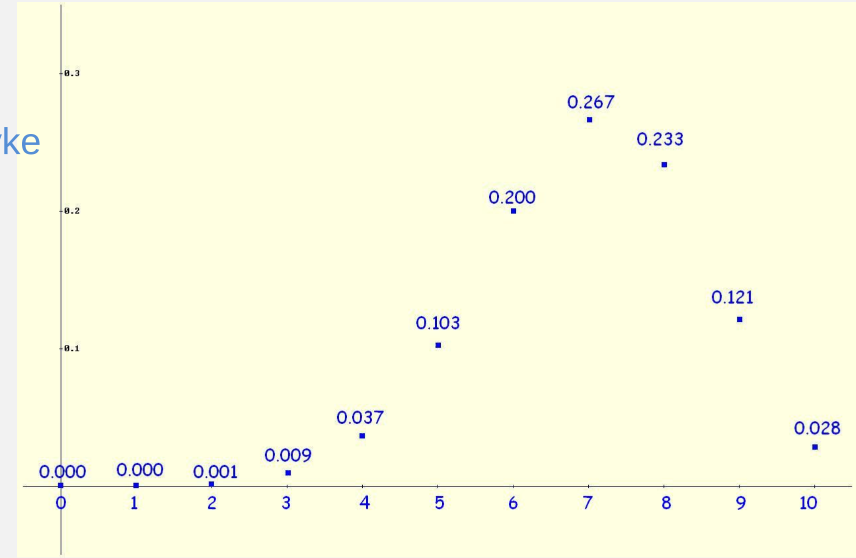
verjetnost vsakega zaporedja je $pk(1$-$p)n$-k .

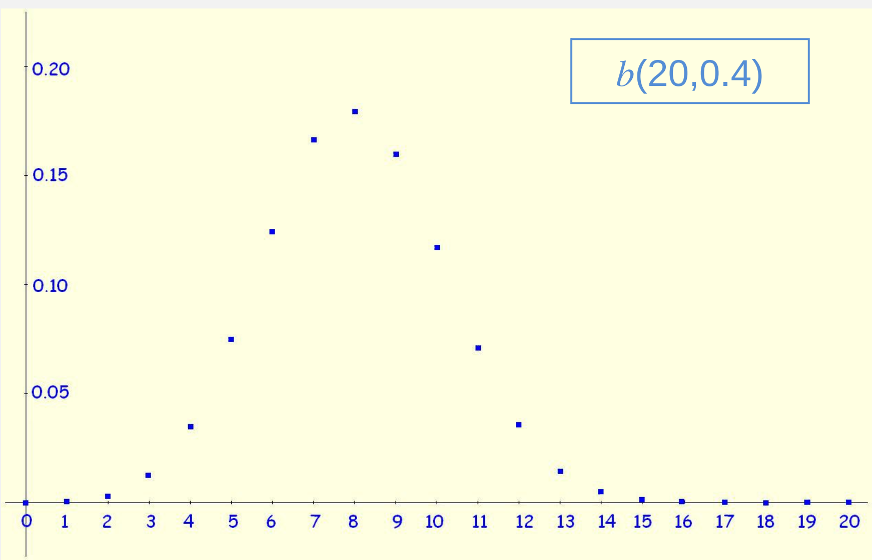
$$\Longrightarrow \quad p_B(k) = P(B=k) = \binom{n}{k} p^k (1-p)^{n-k}$$

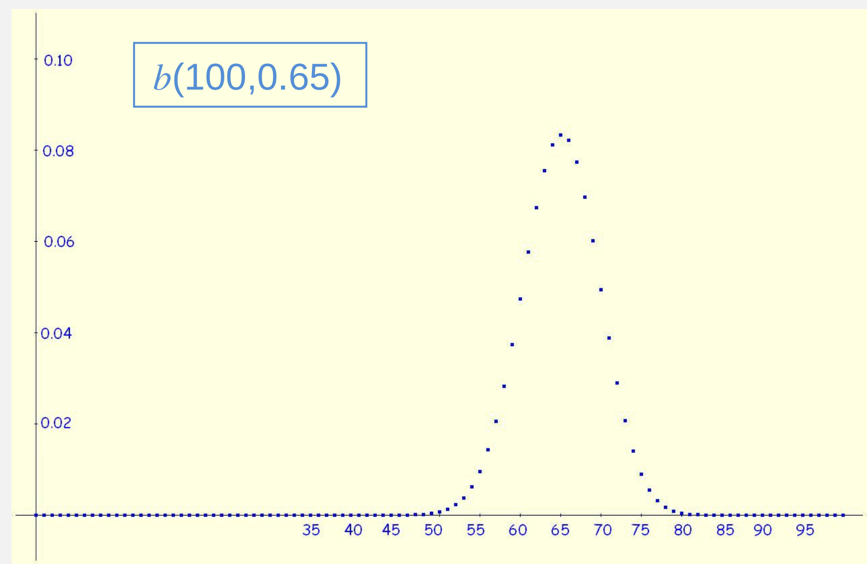
(npr. verjetnost, da koš zadanemo natanko 6-krat je $p(6) = \binom{10}{6} 0.7^6 0.3^4 = 0.200 = 20\%$)

binomska porazdelitev $b(n,p)$

Porazdelitev spremenljivke B za n=10 in p=0.7: b(10,0.7)







Lastnosti binomske porazdelitve $b(n,p)$:

- značilna zvonasta oblika grafa
- maksimum pri $n.p$ (približno)
- za velike n so vse verjetnosti zelo majhne ali celo zanemarljive
  - tedaj je bolj smiselno verjetnosti opazovati
    kumulativno: $P(B \le k)$
    ali
    intervalsko: $P(j \le B \le k)$

Žogo vržemo na koš 100-krat, pri čemer je verjetnost zadetka 70%. Kolikšna je verjetnost, da bomo zadeli več kot 65-krat?

$$P(65 < B \le 100) = \sum_{k=66}^{100} \binom{100}{k} 0.7^k\, 0.3^{100-k} \approx 0.837 = 83.7\%$$

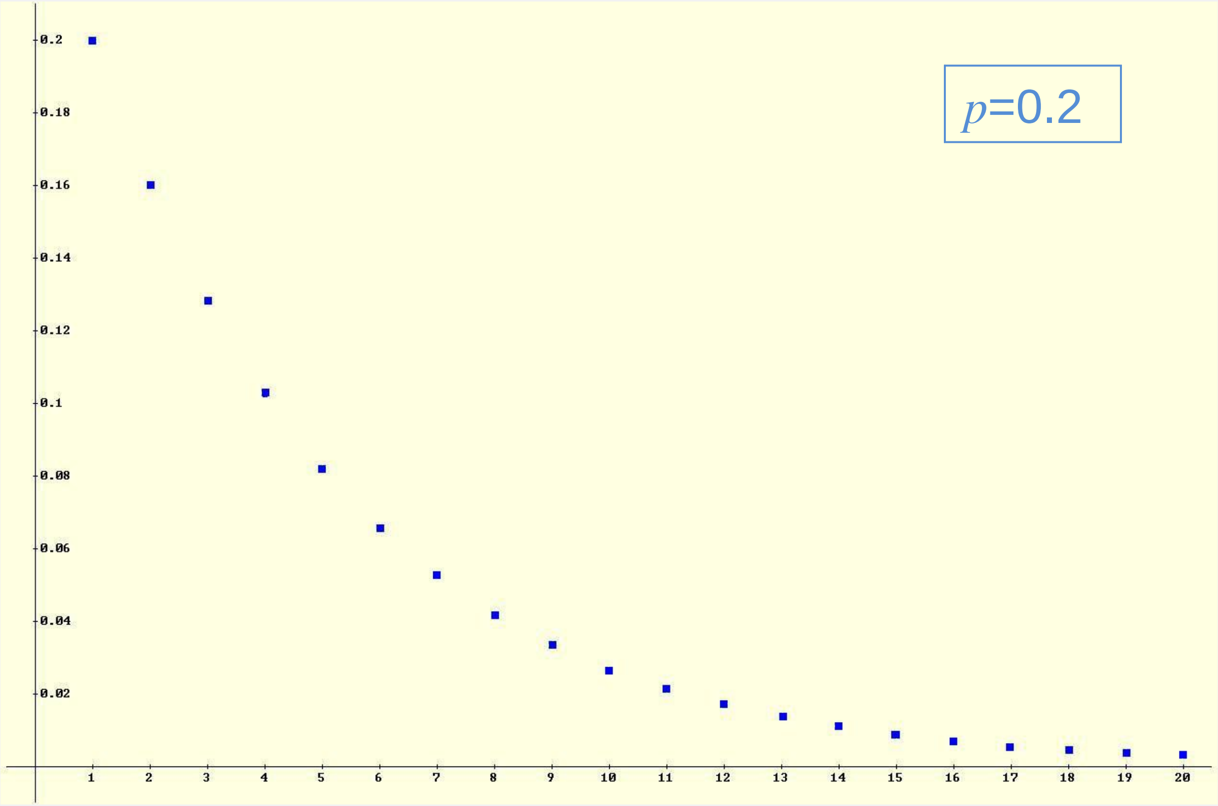
Kaj je bolj verjetno: da bomo v 10 metih zadeli 10-krat ali v 100 metih več kot 80-krat?

- računanje je zelo zamudno in numerično zahtevno

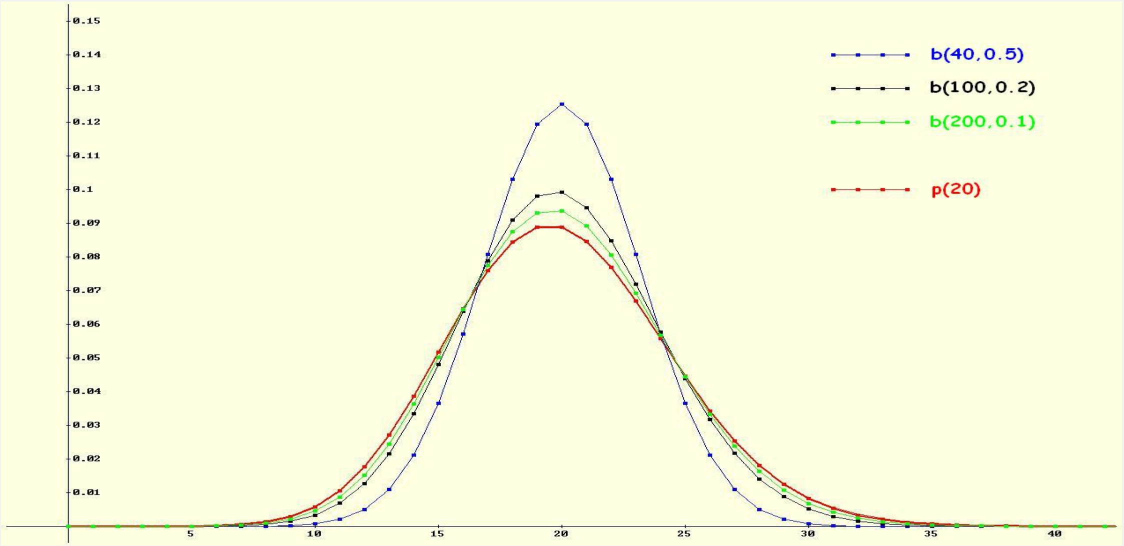
- geometrična porazdelitev

Ponavljamo poskus pri katerem je verjetnostjo uspeha p. Slučajna spremenljivka G naj bo število poskusov, potrebnih za prvi uspeh. Kako je porazdeljena?

- Zaloga vrednosti spremenljivke G je $\{1,2,3,\ldots\}$
- $P(G=k)=p.(1-p)k-1$



➢ Poissonova porazdelitev



Poissonova porazdelitev
P(a)

- zaloga: $\{0,1,2,3,\ldots\}$
- porazdelitev:

$$p(k) = \frac{a^k}{k!} \cdot e^{-a}$$

Če je $a=n.p$ majhen, je Poissonova porazdelitev zelo dober približek za binomsko porazdelitev.

Uporaba:

- modeliranje emisije α-delcev v danem časovnem intervalu
- modeliranje časovnih vrst (vrste pred bančnimi okenci, gostota prometa, obremenitve telefonskega omrežja)
- modeliranje redkih nesreč v zavarovalništvu (npr. čebelji piki, padci pod tušem)

. . . .

ZVEZNE SLUČAJNE SPREMENLJIVKE

Kadar je zaloga slučajne spremenljivke X neštevna (npr. življenjska doba žarnice), potem ne moremo našteti verjetnosti posameznih izidov in jim povrhu običajno sploh ne moremo pripisati pozitivne verjetnosti.

Pomagamo si s kumulativno verjetnostjo:

$P(X \le x)$ = verjetnost, da X zavzame vrednost največ x

(npr. da žarnica pregori po x urah)

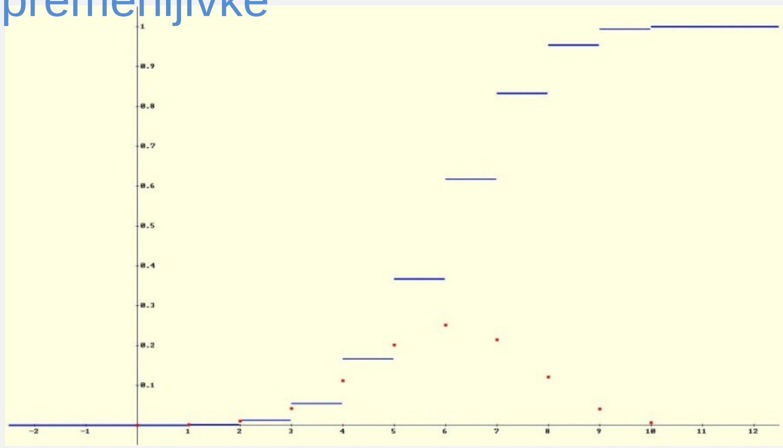
$F_X(x) = P(X \le x)$ je (kumulativna) porazdelitvena funkcija spremenljivke X

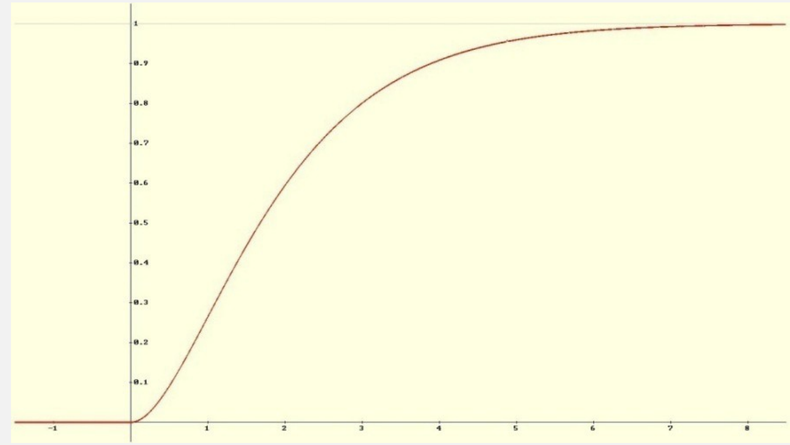
Porazdelitvena funkcija slučajne spremenljivke je

- naraščajoča
- na $(-\infty,\infty)$ zraste od 0 do 1

Spremenljivka X je zvezna če je njena porazdelitvena funkcija F_X zvezna.

Kumulativne porazdelitve diskretne in zvezne slučajne spremenljivke





Če je spremenljivka X zvezna, potem obstaja funkcija $pX(x)$, da je $F_X(x) = \int_{-\infty}^{x} p_X(t)\, dt$

$pX(x)$ je gostota slučajne spremenljivke X

Za gostoto slučajne spremenljivke velja: $\int_{-\infty}^{+\infty} p_X(t)\, dt = 1$

Kjer je pX zvezna je $pX=FX'$.

S pX računamo podobno, kot z diskretno gostoto, le da vsote nadomestimo z integrali:

$P(a \leq X \leq b)$ = verjetnost, da X zavzame vrednost med a in b (da je življenjska doba žarnice med a in b ur)

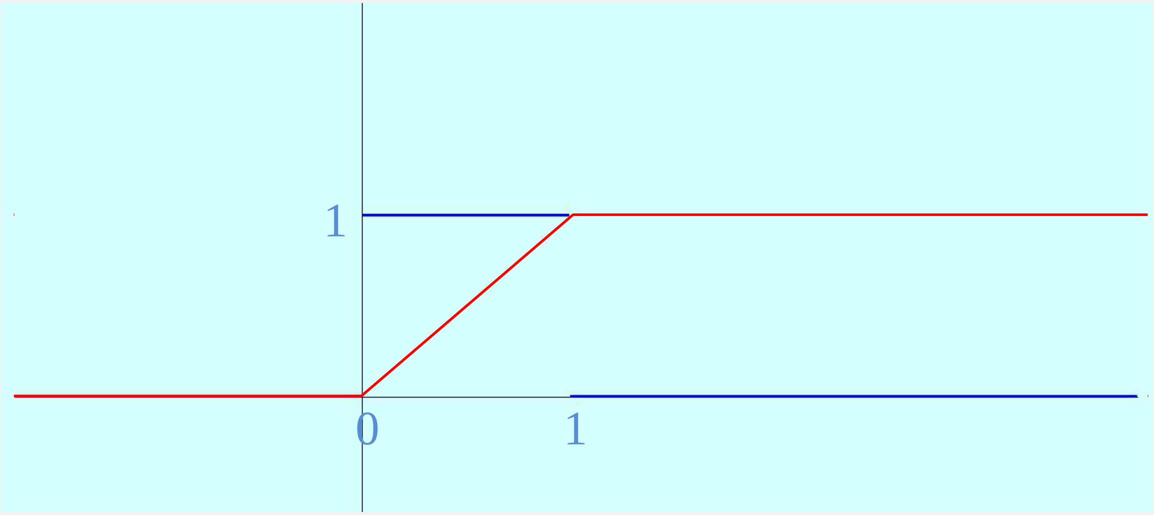
$$P(a \leq X \leq b) = \int_{a}^{b} p_X(x)\, dx$$

Primeri zveznih porazdelitev

- enakomerna porazdelitev

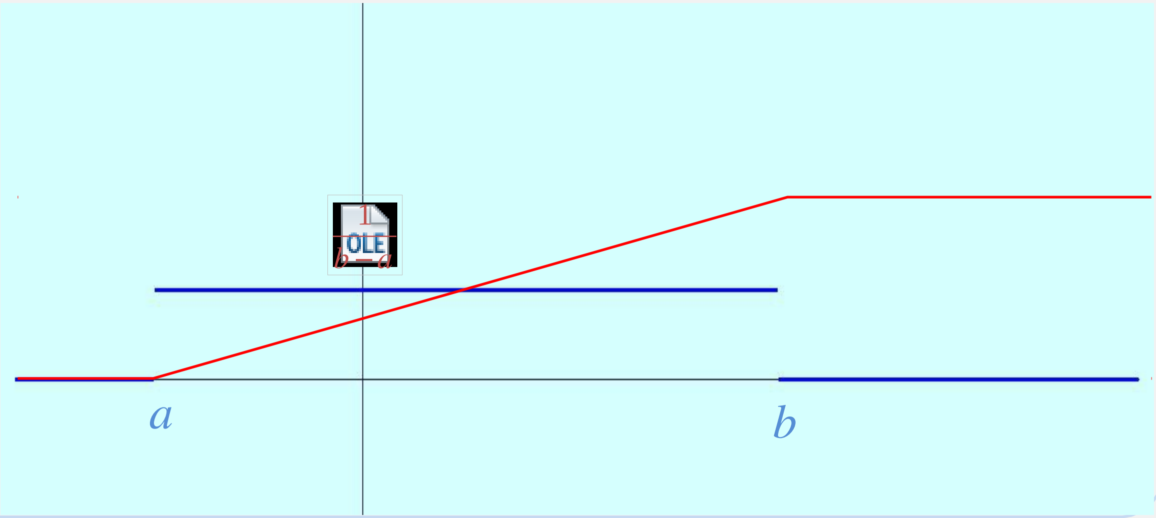
na $[0,1]$, gostota:

$$p(x) = \begin{cases} 1 & 0 \le x \le 1 \\ 0 & \text{sicer} \end{cases}$$



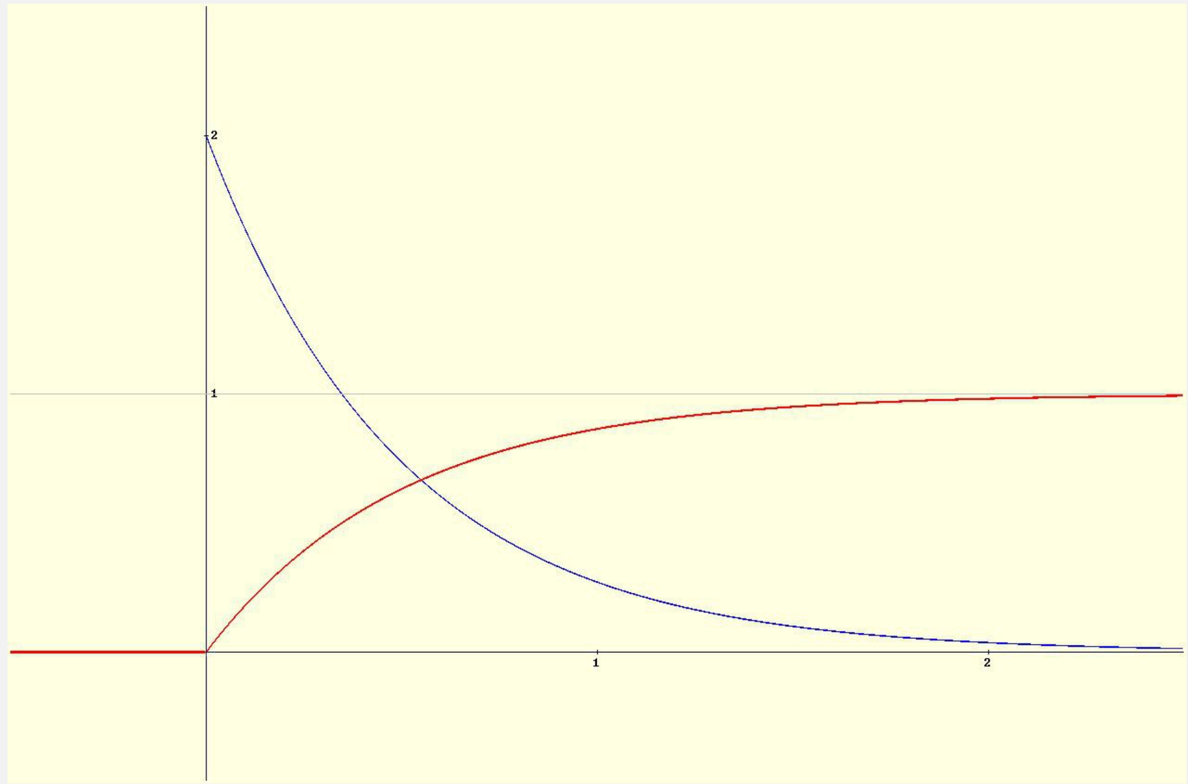
na $[a,b]$, gostota:

$$p(x) = \begin{cases} \dfrac{1}{b-a} & a \le x \le b \\ 0 & \text{sicer} \end{cases}$$



- eksponentna porazdelitev

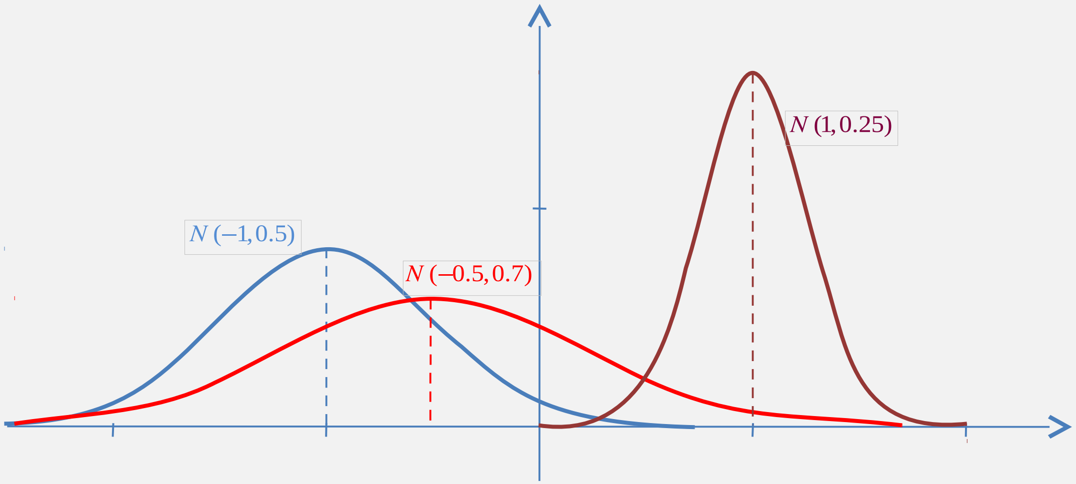
$$p(x) = \begin{cases} 0 & x < 0 \\ a \cdot e^{-ax} & 0 \leq x \end{cases}$$



Podobna Poissonovi; uporaba pri modeliranju življenjske dobe, modeliranju vpliva mamil na živčne receptorje, napovedovanju potresov...

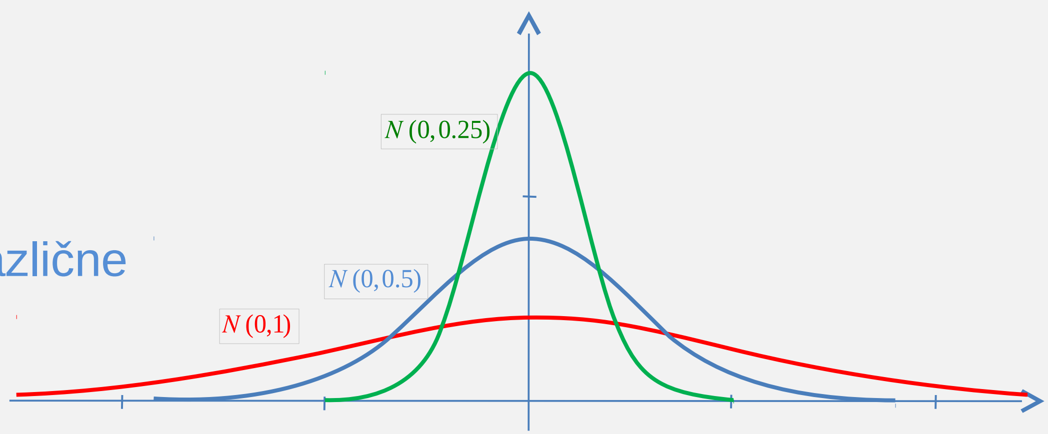
➢ Normalna porazdelitev $N(a,\sigma)$

gostota: $$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-a}{\sigma}\right)^2}$$



- zvonasta oblika
- maksimum pri a
- simetrična glede na a
- eksponetno pada proti 0

gostota porazdelitve N(0,σ) za različne σ:



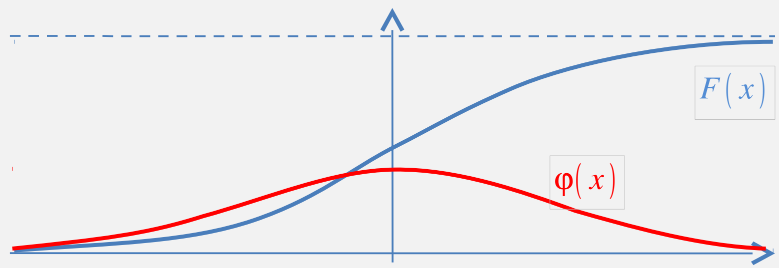
$N(0,1)$ je standardizirana normalna porazdelitev; njena gostota je

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

$$p_{N(a,\sigma)}(x) = \frac{1}{\sigma}\varphi\left(\frac{x-a}{\sigma}\right)$$

vse normalne porazdelitve lahko izrazimo s pomočjo standardizirane

Kumulativna porazdelitvena funkcija standardizirane normalne porazdelitve



$$F_{N(0,1)}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$$

Tudi vse kumulativne normalne porazdelitve lahko izrazimo s standardizirano:

$$F_{N(a,\sigma)}(x) = \int_{-\infty}^{x} p_{N(a,\sigma)}(t)\, dt = \frac{1}{\sigma}\int_{-\infty}^{x} \varphi\left(\frac{t-a}{\sigma}\right) dt =$$

$$= \int_{-\infty}^{\frac{x-a}{\sigma}} \varphi(u)\, du = F_{N(0,1)}\left(\frac{x-a}{\sigma}\right)$$

$$u = \frac{t-a}{\sigma}$$

$$du = \frac{1}{\sigma} dt$$

t	$-\infty$	x
u	$-\infty$	$\frac{x-a}{\sigma}$

Integral gostote ni elementarna funkcija – v praksi si pomagamo s tabelami za funkcijo

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt \longrightarrow F_{N(0,1)}(x) = \frac{1}{2} + \Phi(x)$$

Funkcija $\Phi(x)$ je liha, zato so tabelirane le njene vrednosti za pozitivne x.

$\Phi(1.02)=0.3461$

$\Phi(-0.89)= - \Phi(0.89)= - 0.3132$

$F(-0.89)=0.5+\Phi(- 0.89)=0.1868$

	0	1	2	3	4	5	6	7	8	9
0.0	0.0000	0.0039	0.0079	0.0119	0.0159	0.0199	0.0239	0.0279	0.0318	0.0358
0.1	0.0398	0.0437	0.0477	0.0517	0.0556	0.0596	0.0635	0.0674	0.0714	0.0753
0.2	0.0792	0.0831	0.0870	0.0909	0.0948	0.0987	0.1025	0.1064	0.1102	0.1140
0.3	0.1179	0.1217	0.1255	0.1293	0.1330	0.1368	0.1405	0.1443	0.1480	0.1517
0.4	0.1554	0.1590	0.1627	0.1664	0.1700	0.1736	0.1772	0.1808	0.1843	0.1879
0.5	0.1914	0.1949	0.1984	0.2019	0.2054	0.2088	0.2122	0.2156	0.2190	0.2224
0.6	0.2257	0.2290	0.2323	0.2356	0.2389	0.2421	0.2453	0.2485	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2703	0.2733	0.2763	0.2793	0.2823	0.2852
0.8	0.2881	0.2910	0.2938	0.2967	0.2995	0.3023	0.3051	0.3078	0.3105	0.3132
0.9	0.3159	0.3185	0.3212	0.3238	0.3263	0.3289	0.3314	0.3339	0.3364	0.3389
1.0	0.3413	0.3437	0.3461	0.3484	0.3508	0.3531	0.3554	0.3576	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3707	0.3728	0.3749	0.3769	0.3789	0.3809	0.3829
1.2	0.3849	0.3868	0.3887	0.3906	0.3925	0.3943	0.3961	0.3979	0.3997	0.4014
1.3	0.4031	0.4049	0.4065	0.4082	0.4098	0.4114	0.4130	0.4146	0.4162	0.4177
1.4	0.4192	0.4207	0.4221	0.4236	0.4250	0.4264	0.4278	0.4292	0.4305	0.4318
1.5	0.4331	0.4344	0.4357	0.4369	0.4382	0.4394	0.4406	0.4417	0.4429	0.4440
1.6	0.4452	0.4463	0.4473	0.4484	0.4494	0.4505	0.4515	0.4525	0.4535	0.4544
1.7	0.4554	0.4563	0.4572	0.4581	0.4590	0.4599	0.4607	0.4616	0.4624	0.4632
1.8	0.4640	0.4648	0.4656	0.4663	0.4671	0.4678	0.4685	0.4692	0.4699	0.4706
1.9	0.4712	0.4719	0.4725	0.4731	0.4738	0.4744	0.4750	0.4755	0.4761	0.4767
2.0	0.4772	0.4777	0.4783	0.4788	0.4793	0.4798	0.4803	0.4807	0.4812	0.4816
2.1	0.4821	0.4825	0.4829	0.4834	0.4838	0.4842	0.4846	0.4849	0.4853	0.4857
2.2	0.4860	0.4864	0.4867	0.4871	0.4874	0.4877	0.4880	0.4883	0.4886	0.4889
2.3	0.4892	0.4895	0.4898	0.4900	0.4903	0.4906	0.4908	0.4911	0.4913	0.4915
2.4	0.4918	0.4920	0.4922	0.4924	0.4926	0.4928	0.4930	0.4932	0.4934	0.4936
2.5	0.4937	0.4939	0.4941	0.4942	0.4944	0.4946	0.4947	0.4949	0.4950	0.4952
2.6	0.4953	0.4954	0.4956	0.4957	0.4958	0.4959	0.4960	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4971	0.4972	0.4973
2.8	0.4974	0.4975	0.4975	0.4976	0.4977	0.4978	0.4978	0.4979	0.4980	0.4980
2.9	0.4981	0.4981	0.4982	0.4983	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986
3.0	0.4986	0.4986	0.4987	0.4987	0.4988	0.4988	0.4988	0.4989	0.4989	0.4989

Če je X standardizirano normalna $N(0,1)$, je $P(x_1 \le X \le x_2) = \Phi(x_2) - \Phi(x_1)$

Če pa je X normalna $N(a,\sigma)$, je $$P(x_1 \le X \le x_2) = \Phi\left(\frac{x_2 - a}{\sigma}\right) - \Phi\left(\frac{x_1 - a}{\sigma}\right)$$

Slučajna spremenljivka X je porazdeljena po zakonu $N(1.5,0.2)$. Kolikšna je verjetnost, da X zavzame vrednost med 1 in 1.5?

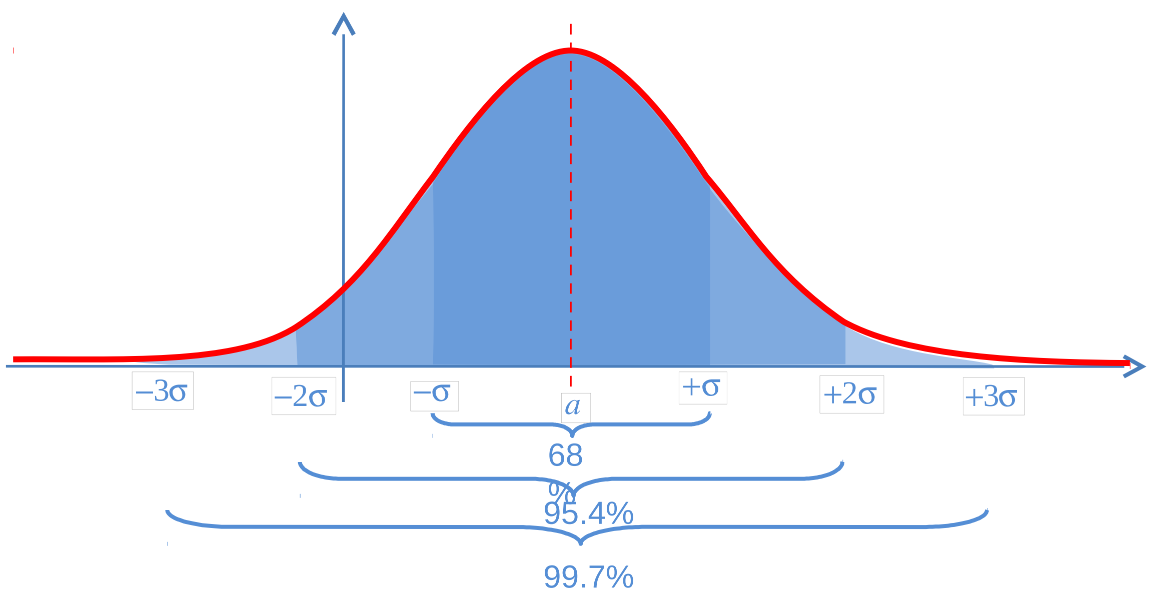
$$P(1<X<1.5)=\Phi\left(\frac{1.5-1.5}{0.2}\right)-\Phi\left(\frac{1-1.5}{0.2}\right)=\Phi(0)-\Phi(-2.5)=\Phi(2.5)=0.4937=49.37\%$$

X porazdeljena po $N(a,\sigma)$:

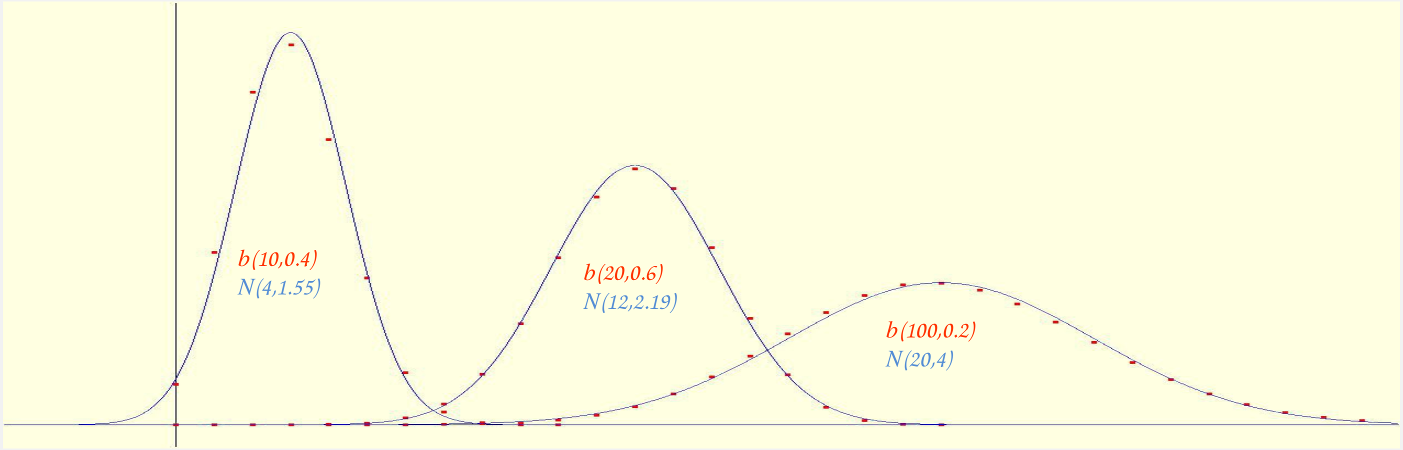
$$P(a-\sigma<X<a+\sigma)=\Phi\left(\frac{a+\sigma-a}{\sigma}\right)-\Phi\left(\frac{a-\sigma-a}{\sigma}\right)=\Phi(1)-\Phi(-1)=2\Phi(1)=0.6826$$

$$P(a-2\sigma<X<a+2\sigma)=2\Phi(2)=0.9544$$

$$P(a-3\sigma<X<a+3\sigma)=2\Phi(3)=0.9972$$



Normalna porazdelitev $N\left(np,\sqrt{np(1-p)}\right)$ je dober približek za binomsko porazdelitev $b(n,p)$



Laplaceova ocena za binomsko porazdelitev $b(n,p)$ ($q=1-p$):

lokalna: $\binom{n}{k}p^k q^{n-k} \approx \frac{1}{\sqrt{2\pi npq}}\, e^{-\frac{(k-np)^2}{2npq}} = \frac{1}{\sqrt{npq}}\,\varphi\left(\frac{k-np}{\sqrt{npq}}\right)$

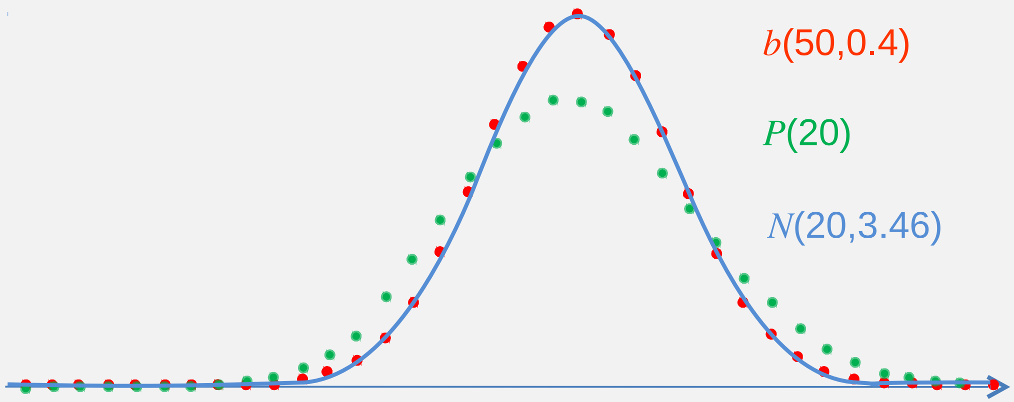
integralska: $P(x_1 \le X \le x_2) \approx \Phi\left(\frac{x_2-np}{\sqrt{npq}}\right) - \Phi\left(\frac{x_1-np}{\sqrt{npq}}\right)$

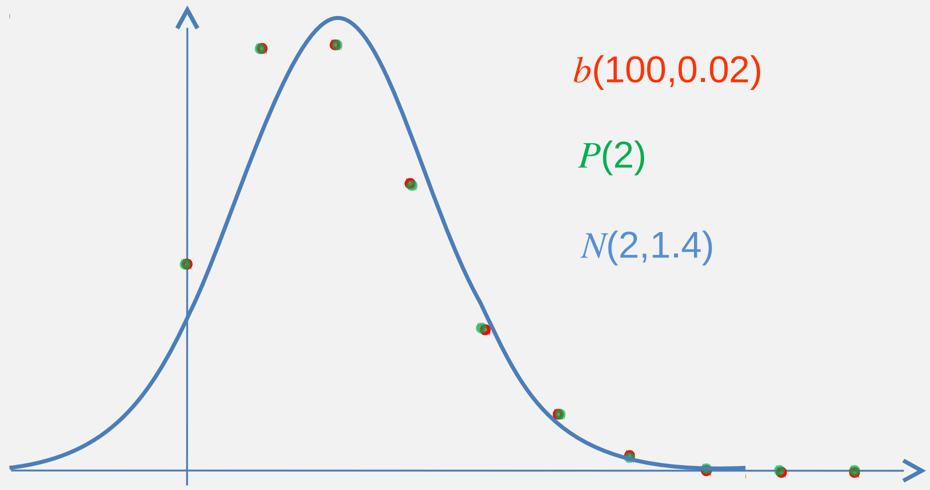
Žogo vržemo na koš 100-krat, pri čemer je verjetnost zadetka 70%. Kolikšna je verjetnost, da bomo zadeli več kot 65-krat?

$$P(65<B\le 100) = \sum_{k=66}^{100}\binom{100}{k}0.7^k\,0.3^{100-k} \approx 0.837 = 83.7\%$$

$$P(65<B\le 100) \approx \Phi\left(\frac{100-100\cdot 0.7}{\sqrt{100\cdot 0.7\cdot 0.3}}\right)-\Phi\left(\frac{66-100\cdot 0.7}{\sqrt{100\cdot 0.7\cdot 0.3}}\right)=\Phi(6.54)+\Phi(0.87)=0.5+0.3078=80.8\%$$

Primerjava binomske, Poissonove in normalne porazdelitve





Normalna porazdelitev je običajno boljši približek za binomsko kot Poissonova.

Ko je produkt $n.p$ majhen (in n dovolj velik) pa je Poissonov približek boljši.

POVPREČNA VREDNOST

X diskretna, vrednosti xk, gostota $p(xk)$

X zvezna, gostota $p(x)$

$$E(X) = \sum_k x_k \cdot p(x_k)$$

povprečna vrednost spremenljivke X

$$E(X) = \int_{-\infty}^{+\infty} x \cdot p(x)\, dx$$

Na ruleti so številke od 1 do 36 ter še 0 in 00. Kdor vloži 1 EUR na sode, dobi ali zgubi 1 EUR glede na to ali kroglica pade na sodo oziroma liho številko.

Dobiček X :+1 z verjetnostjo 18/38

-1 z verjetnostjo 20/38.

Povprečni dobiček:

$$E(X) = 1 \cdot \frac{18}{38} + (-1) \cdot \frac{20}{38} = -\frac{1}{19}$$

Kdor vloži 1 EUR na izbrano številko (npr. 25) dobi 36 EUR, če kroglica pade na 25, v nasprotnem pa zgubi 1 EUR.

Povprečni dobiček:

$$E(X) = 36 \cdot \frac{1}{38} + (-1) \cdot \frac{37}{38} = -\frac{1}{38}$$

Življenjska doba žarnice je porazdeljena eksponentno. Kolikšna je, v povprečju, njena življenjska doba?

$$p(x) = \begin{cases} 0 & x < 0 \\ 0.01 \cdot e^{-0.01x} & 0 \leq x \end{cases}$$

$$E(X) = 0.01 \int_0^{+\infty} x\, e^{-0.01x}\, dx =$$

$$= 0.01\left(-\frac{1}{0.01} x\, e^{-0.01x} \Big|_0^{+\infty} + \frac{1}{0.01} \int_0^{+\infty} e^{-0.01x}\, dx \right) =$$

$$= -\frac{1}{0.01} e^{-0.01x} \Big|_0^{+\infty} = \frac{1}{0.01} = 100 \text{ ur}$$

V neki tovarni je približno en izdelek od desetih pokvarjen. Vsak dan izdelke pregledujejo enega po enega dokler ne najdejo pokvarjenega. Koliko izdelkov morajo v povprečju pregledati?

X je geometrično porazdeljena s p=0.1:

$$E(X)=\sum_{k=1}^{\infty} k\cdot p\cdot(1-p)^{k-1}=$$

trik:

$$\sum_{k=0}^{\infty} x^k=\frac{1}{1-x}\Rightarrow\sum_{k=1}^{\infty} k\,x^{k-1}=\frac{1}{(1-x)^2}$$

$$\Rightarrow\sum_{k=1}^{\infty} k\,(1-p)^{k-1}=\frac{1}{(1-(1-p))^2}=\frac{1}{p^2}$$

$$=\frac{1}{p}$$

Povprečno morajo dnevno pregledati po 10 izdelkov.

Igralec na ruleti igra po naslednjem sistemu. Vsakič igra igro z verjetnostjo 0.5 (npr. stavi na rdeče, izidov 0 in 00 ne štejemo). Najprej vloži 1 EUR; če izgubi, podvoji vložek in to ponavlja, dokler ne zmaga; ob vsaki zmagi je na dobičku 1 EUR (zaporedja vložkov so 1-2, 1-2-4, 1-2-4-8, 1-2-4-8-16 itn.). Po zmagi spet začne z 1 EUR ...

Ali je to zanesljiva pot do zaslužka?

Naj bo X količina denarja vložena pri zadnji igri (tisti, v kateri igralec zmaga).

Zaloga vrednosti X je {1,2,4,8,...}, tj. {2k; k=0,1,2,3,...}; porazdelitev je $P(X=2k)=2-(k+1)$.

$$E(X)=\sum_{k=0}^{\infty} 2^k\cdot\frac{1}{2^{k+1}}=\infty$$

Povprečna vrednost slučajne spremenljivke X ni definirana! 'Sistem' zahteva neskončno zalogo denarja (in možnost za neomejene stave).

RAZPRŠENOST

Razpršenost je povprečje odklonov spremenljivke od njene povprečne vrednosti:

$$D(X) = E\left(\left(X - E(X)\right)^2\right)$$ razpršenost (varianca, disperzija)

$m=E(X)$

$$D(X) = \sum_k (x_k - m)^2 \cdot p(x_k)$$

$$D(X) = \int_{-\infty}^{+\infty} (x - m)^2 \cdot p(x)\, dx$$

$$D(X) = \int_{-\infty}^{+\infty} (x - m)^2 \cdot p(x)\, dx = \int_{-\infty}^{+\infty} (x^2 - 2mx + m^2) \cdot p(x)\, dx =$$

$$= \int_{-\infty}^{+\infty} x^2 \cdot p(x)\, dx - 2m \int_{-\infty}^{+\infty} x \cdot p(x)\, dx + m^2 \int_{-\infty}^{+\infty} p(x)\, dx =$$

$$= E(X^2) - 2m^2 + m^2 = E(X^2) - m^2$$

praktična formula: $$D(X) = E(X^2) - E(X)^2$$

Kako je razpršeno število pik pri metu kocke?

$$E(X)=1\cdot\frac{1}{6}+2\cdot\frac{1}{6}+3\cdot\frac{1}{6}+4\cdot\frac{1}{6}+5\cdot\frac{1}{6}+6\cdot\frac{1}{6}=\frac{21}{6}=3.5$$

$$E(X^2)=1\cdot\frac{1}{6}+4\cdot\frac{1}{6}+9\cdot\frac{1}{6}+16\cdot\frac{1}{6}+25\cdot\frac{1}{6}+36\cdot\frac{1}{6}=\frac{91}{6}$$

$$D(X)=\frac{91}{6}-\left(\frac{21}{6}\right)^2=\frac{35}{12}\approx 2.92$$

$$\sigma(X)=\sqrt{D(X)}$$

standardni odklon slučajne spremenljivke X

Standardni odklon pri metu kocke je $\sqrt{2.92}=1.71$

$$D(aX+b)=E\left((aX+b)^2\right)-E(aX+b)^2$$
$$=E\left(a^2X^2+2abX+b^2\right)-\left(aE(X)+b\right)^2$$
$$=a^2\left(E(X^2)-E(X)^2\right)$$

$$D(aX+b)=a^2\cdot D(X)$$
$$\sigma(aX+b)=|a|\cdot\sigma(X)$$

VPREČNA VREDNOST IN RAZPRŠENOST NEKATERIH POMEMBNIH PORAZDELITEV

Binomska porazdelitev $b(n,p)$: zaloga $\{0,1,2,\dots,n\}$, $\quad p_k=\binom{n}{k}p^kq^{n-k}$ $\quad$ (kjer je $q=1-p$)

Sešteti moramo $E(X)=\sum_{k=0}^{n}k\cdot\binom{n}{k}p^kq^{n-k}$ in $E(X^2)=\sum_{k=0}^{n}k^2\cdot\binom{n}{k}p^kq^{n-k}$

vpeljemo pomožno funkcijo $(px+q)^n=\sum_{k=0}^{n}\binom{n}{k}p^kq^{n-k}\cdot x^k$

$$\sum_{k=0}^{n}k\cdot\binom{n}{k}p^kq^{n-k}\cdot x^{k-1}=((px+q)^n)'=n(px+q)^{n-1}\cdot p$$

$$x=1\ \Rightarrow\ \sum_{k=0}^{n}k\cdot\binom{n}{k}p^kq^{n-k}=n(p\cdot 1+q)^{n-1}\cdot p=np$$

$$E(X)=n.p$$

$$\sum_{k=0}^{n}k(k-1)\cdot\binom{n}{k}p^kq^{n-k}\cdot x^{k-2}=((px+q)^n)''=n(n-1)(px+q)^{n-2}\cdot p^2$$

$$x=1\ \Rightarrow\ \sum_{k=0}^{n}k^2\cdot\binom{n}{k}p^kq^{n-k}=n(n-1)(p\cdot 1+q)^{n-1}\cdot p^2+np=n^2p^2+npq$$

$$D(X)=n.p.\ q$$

Poissonova porazdelitev $P(a)$: zaloga $\{0,1,2,3,\dots\}$, $\quad p_k=\frac{a^k}{k!}\cdot e^{-a}$

$$e^x=\sum_{k=0}^{\infty}\frac{x^k}{k!}\ \Rightarrow\ \sum_{k=0}^{\infty}k\cdot\frac{x^{k-1}}{k!}=(e^x)'=e^x\ \Rightarrow\ \sum_{k=0}^{\infty}k\cdot\left(\frac{a^k}{k!}e^{-a}\right)=a\cdot e^a\cdot e^{-a}=a$$

$$E(X)=a$$

$$e^x=\sum_{k=0}^{\infty}k(k-1)\cdot\frac{x^{k-2}}{k!}\ \Rightarrow\ \sum_{k=0}^{\infty}(k^2-k)\cdot\left(\frac{a^k}{k!}e^{-a}\right)=a^2\ \Rightarrow\ \sum_{k=0}^{\infty}k^2\cdot\left(\frac{a^k}{k!}e^{-a}\right)=a^2+a$$

$$D(X)=a$$

Normalna porazdelitev $N(a,\sigma)$: zaloga $(-\infty,+\infty)$, $p(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-a}{\sigma}\right)^2}=\frac{1}{\sigma}\varphi\left(\frac{x-a}{\sigma}\right)$

$$E(X)=\int_{-\infty}^{+\infty}x\cdot\frac{1}{\sigma}\varphi\left(\frac{x-a}{\sigma}\right)dx=\int_{-\infty}^{+\infty}(\sigma t+a)\cdot\varphi(t)\,dt=\sigma\underbrace{\int_{-\infty}^{+\infty}t\cdot\varphi(t)\,dt}_{=0\text{ (liha funkcija)}}+a\underbrace{\int_{-\infty}^{+\infty}\varphi(t)\,dt}_{=1}=a$$

$$t=\frac{x-a}{\sigma},\quad dt=\frac{1}{\sigma}dx$$

$\boxed{E(X)=a}$

$$\int_{-\infty}^{+\infty}x^2\cdot\frac{1}{\sigma}\varphi\left(\frac{x-a}{\sigma}\right)dx=\int_{-\infty}^{+\infty}(\sigma t+a)^2\cdot\varphi(t)\,dt=\sigma^2\int_{-\infty}^{+\infty}t^2\cdot\varphi(t)\,dt+2a\sigma\underbrace{\int_{-\infty}^{+\infty}t\cdot\varphi(t)\,dt}_{=0}+a^2\underbrace{\int_{-\infty}^{+\infty}\varphi(t)\,dt}_{=1}=$$

$$=\sigma^2\left(-t\,\varphi(t)\Big|_{-\infty}^{+\infty}+\int_{-\infty}^{+\infty}\varphi(t)\,dt\right)+a^2=\sigma^2+a^2$$

$$u=t,\quad du=dt$$
$$dv=t\,\varphi(t)\,dt,\quad v=-\varphi(t)$$

$\boxed{D(X)=\sigma 2}$

$\boxed{\sigma(X)=\sigma}$

porazdelitev	zaloga	gostota		$E(X)$	$D(X)$
$\sigma(X)$					
enakomerna	$\{1,2,\ldots,n\}$	$p(k)=\frac{1}{n}$	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$	$\sqrt{\frac{n^2-1}{12}}$
binomska $b(n,p)$	$\{0,1,2,\ldots,n\}$	$p(k)=\binom{n}{k}p^k(1-p)^{n-k}$	np	$np(1-p)$	$\sqrt{np(1-p)}$
geometrijska	$\{1,2,3,\ldots\}$	$p(k)=p(1-p)^{k-1}$	$\frac{1}{p}$	$\frac{1-p}{p}$	$\frac{\sqrt{1-p}}{p}$
Poissonova $P(a)$	$\{0,1,2,\ldots\}$	$p(k)=\frac{a^k}{k!}e^{-a}$	a	a	$\sqrt{a}$
enakomerna	$[a,b]$	$p(x)=\frac{1}{b-a}\ (x\in[a,b])$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{(b-a)}{2\sqrt{3}}$
eksponentna	$[0,+\infty)$	$p(x)=a\,e^{-ax}$	$\frac{1}{a}$	$\frac{1}{a^2}$	$\frac{1}{a}$
normalna $N(a,\sigma)$	$(-\infty,+\infty)$	$p(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-a}{\sigma}\right)^2}$	a	σ^2	σ