

# 1. Del

## 1. Slučajne spremenljivke

- primeri
  - Imamo nek poskus katerega izidi so števila (npr. met kocke). Se pravi da je poskusom prirejena neka količina ki mora zavzeti različne vrednosti.
    - Vrednost, ki jo zavzame v določeni ponovitvi poskusa je odvisno od slučaja.
  - Primer: Vrzemo kocko, pade lahko od 1-6 pik, vsi izidi so enako verjetni
  - $X \sim \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix}$
- definicija (zaloga vrednosti in porazdelitveni zakon)
  - **Zaloga vrednosti**  $Z_X$ : vrednosti ki jih slučajna spremenljivka lahko zavzame
  - **Porazdelitveni zakon** - predpis, ki določa verjetnost vsake izmed možnih vrednosti ali intervala vrednosti, ki ga s.s. lahko zavzame
- porazdelitvena funkcija
  - $F_X(x) = P(X \leq x)$
- lastnosti porazdelitvene funkcije
  1. Funkcija  $F$  je definirana na vsem  $\mathbb{R}$  in zanjo velja  $0 \leq F(x) \leq 1 \forall x \in \mathbb{R}$
  2.  $F$  je nepadajoca
  3.  $F(-\infty) := \lim_{x \rightarrow -\infty} F_X(x) = 0$
  4.  $F(\infty) := \lim_{x \rightarrow \infty} F_X(x) = 1$
  5. Funkcija je v vsaki točki zvezna z desne
  6. Funkcija ima lahko v nekaterih točkah skok
  7.  $P(x_1 < X \leq x_2) = F(x_2) - F(x_1)$
  8.  $P(X > x) = 1 - F(x)$
  9.  $P(X = x) = F(x) - F(x-)$
- diskretna, zvezna
  - **diskretna**: zaloga vrednosti je števna množica
  - **zvezna**: lahko zavzame vsako realno število znotraj določenega intervala (končnega/nekončnega)
- primer mešane slučajne spremenljivke
  - kombinirana: na nekem območju diskretna na nekem zvezna
  - primer: količina padavin v nekem kraju v določenem časovnem obdobju
    - npr: dan brez padavine ima vrednost 0
    - zvezno območje padavin za dneve ko so padavine

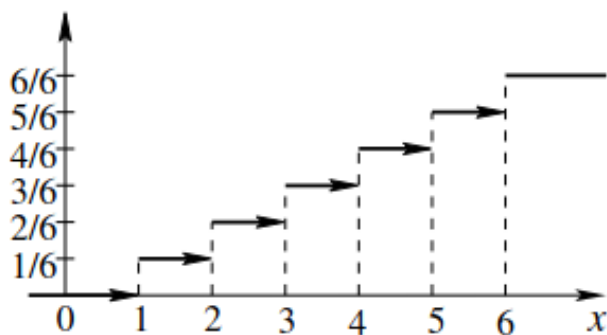
## 2. Diskretne slučajne spremenljivke

- primer tabele
  - $X \equiv$  izidi pri metu pravične kocke
  - $X \sim \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix}$
- primeri: enakomerna, binomska, Poissonova, Pascalova, Hipergeometrijska
  - bernoullijeva  $\sim B(p)$ 
    - V vsakem poskusu ima dogodek A verjetnost  $p$ ,  $X$  pa ima vrednost 1, če se je zgodil dogodek A in 0 sicer
    - $P(X = 1) = p$
    - $P(X = 0) = 1 - p$
  - binomska  $\sim B(n, p)$ 
    - Kovanec vzmemo  $n$ -krat, kolikšna je verjetnost da pade cifra z verjetnostjo  $p$
    - $p_k = P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$
  - Geometrijska  $\sim G(p)$ 
    - $X$  je število ponovitev poskusa do (vključno) prve pojavitve izida A
    - $X$  = št. poskusov da prvič pade 6
    - $P(X = k) = (1 - p)^{k-1} p$
  - Pascalova/Negativna binomska  $\sim P(n, p)$ 
    - $X$  je število ponovitev poskusa do (vključno)  $n$ -te pojavitve izida A
      - npr. koliko poskusov rabimo do 10. šestice

- $P(X = k) = \binom{k-1}{n-1} p^{k-n} p^n$
- Hipergeometrijska  $\sim H(R, B, n)$ 
  - X je število rdečih kroglic med izbranimi n kroglicami
  - $P(X = k) = \frac{\binom{R}{k} + \binom{B}{n-k}}{\binom{R+B}{n}}$
- Poissonova  $X \sim P(\lambda)$ 
  - V povprečju imamo na intervalu  $\lambda$  ponovitev dogodka A
  - X pa je število ponovitev dogodka A na danem intervalu
  - $P(X = k) = \frac{\lambda^k}{k!} \cdot e^{-\lambda}$
- definicija (tabela, zaloga vrednosti je števna množica)
  - **zaloga vrednosti** je števna množica  $\{x_1, x_2, \dots, x_n, \dots\}$
  - dogodki  $X = x_k \forall k$  predstavljajo popoln sistem dogodkov
  - verjetnost dogodka  $P(X = x_i) = p_i$
  - $\sum_i p_i = 1$
  - **verjetnostna tabela**, v prvi vrstici imamo vrednosti  $x_i \in Z_X$  v drugi vrstici  $P(X = x_i) = p_i$

$$X \sim \begin{pmatrix} x_1 & x_2 & \dots & x_m & \dots \\ p_1 & p_2 & \dots & p_m & \dots \end{pmatrix}$$

- porazdelitvena funkcija je stopnicasta funkcija
  - Porazdelitvena funkcija ni zvezna je stopnicasta



- izpeljava netrivialne pričakovane vrednosti ali odklona
  - **pričakovana vrednost E(X)** diskretne spremenljivke X je posplošitev povprečne vrednosti / tezišca
  - $\bar{X} = \frac{x_1 k_1 + \dots + x_m k_m}{N} = x_1 f_1 + \dots + x_m f_m = x_1 p_1 + \dots + x_m p_m = E(X)$

### 3. Poissonova porazdelitev

- primer
  - Neka roketna ekipa da v **povprečju 30 golov** na tekmo (tekma traja **60min**).
    - Koliko verjetno ekipa na naslednji tekmi v prvi minuti doseže vsaj en gol?
      - $t = 1\text{min}, \rightarrow \lambda = 0.5$
      - $P(X \geq 1) = 1 - F_X(0) = 1 - \frac{0.5^0 \cdot e^{-0.5}}{0!} = 1 - e^{-0.5}$
    - Koliko verjetno ekipa v zadnjih 3min tekme doseže natanko dva gola?
      - $t = 3\text{min} \rightarrow \lambda = \frac{3}{2}$
      - $P(X = 2) = \frac{(\frac{3}{2})^2 \cdot e^{-\frac{3}{2}}}{2!} \approx 0.251$
- zaloga vrednosti
  - **števna a neomejena** (teoretično se lahko zgodi neskončno mnogo dogodkov na danem intervalu)
- predpostavke
  - dogodki morajo biti porazdeljeni po Poissonu
    - dogodki se pojavijo **neodvisno** od časa, ki je potekel od zadnjega dogodka
    - povprečno število dogodkov ki se pojavijo na nekem intervalu je **konstantno** -  $\lambda$

- definicija - verjetnostna funkcija
  - $p_k = P(X = k) = \lambda^k \frac{e^{-\lambda}}{k!}$
- povezava z binomsko porazdelitvijo
  - Poissonova porazdelitev se lahko dobi kot limitni primer binomske porazdelitve (ce gre stevilo poskusov preko vseh mej)
  - Torej Poissonova porazdelitev je aproksimacija binomske porazdelitve ce je n dovolj velik in p dovolj majhen
    - $X \sim B(n, p) \approx P(np)$
  - **Dokaz:**
    - imamo  $P(X = k) = \lim_{n \rightarrow \infty} \left( \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \right)$ 
      - Vpeljemo  $\lambda = np \rightarrow p = \frac{\lambda}{n}$

$$P(X = k) = \lim_{n \rightarrow \infty} \left( \frac{n!}{k!(n-k)!} \cdot \frac{\lambda^k}{n^k} \cdot \left(1 - \frac{\lambda}{n}\right)^n \cdot \left(1 - \frac{\lambda}{n}\right)^{-k} \right) =$$

- Upostevamo:  $\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}$  ter  $\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^{-k} = 1$

$$P(X = k) = \lim_{n \rightarrow \infty} \left( \frac{n!}{k!(n-k)!} \cdot \frac{\lambda^k}{n^k} \cdot e^{-\lambda} \right) =$$

- Konstante premaknemo ven iz limite.

$$\begin{aligned} &= \frac{\lambda^k}{k!} e^{-\lambda} \lim_{n \rightarrow \infty} \left( \frac{n!}{(n-k)!} \cdot \frac{1}{n^k} \right) = \\ &= \frac{\lambda^k}{k!} e^{-\lambda} \lim_{n \rightarrow \infty} \left( \frac{n \cdot (n-1) \cdot (n-2) \cdots (n-k)!}{(n-k)! \cdot n^k} \right) = \\ &= \frac{\lambda^k}{k!} e^{-\lambda} \lim_{n \rightarrow \infty} \left( \frac{n(n-1)(n-2) \cdots (n-k+1)}{n^k} \right) = \\ &= \frac{\lambda^k}{k!} e^{-\lambda} \lim_{n \rightarrow \infty} \left( \left(\frac{n}{n}\right) \left(\frac{n-1}{n}\right) \cdots \left(\frac{n-k+1}{n}\right) \right) = \end{aligned}$$

- Vidimo da v limiti grejo vsi notranji členi  $\rightarrow 1$

$$= \frac{e^{-\lambda} \lambda^k}{k!}$$

- razlaga parametra  $\lambda$ 
  - st. ponovitev dogodka A, ki jih imamo v povprečju na nekem intervalu
- pričakovana vrednost in disperzija
  - $E(X) = \lambda$
  - $D(X) = \lambda$ 
    - $D(X) = E((X - E(X))^2) = E(X^2 + E^2(X) - 2XE(X)) = E(X^2) + E^2(X) - 2E(X)E(X) = E(X^2) - E^2$
    - Upostevali smo (linearnost in konstante):
    - $E(E^2(X)) = E^2(X)$ , **E(konstanta)=konstanta**
    - $E(-2XE(X)) = -2E(X)E(X)$  (konstanto premaknemo vn)
- Poissonov obrazec
  - $B(n, p) \approx P(np)$ 
    - dokaz zgoraj

## 4. Pascalova porazdelitev

- primeri (npr. geometrijska)
  - Naj bo X stevilo metov postenega kovanca, ki ga mecemo dokler ne pade cifra in takoj nato grb.
  - $X_1 \sim G\left(\frac{1}{2}\right) \dots$  pade prva cifra
  - $X_2 \sim G\left(\frac{1}{2}\right) \dots$  pade prvi grb (po cifri - torej neodvisen)
  - $X = X_1 + X_2$ 
    - $X \sim P\left(2, \frac{1}{2}\right)$
- definicija, verjetnostna funkcija
  - $X \sim P(n, p)$
  - X je stevilo ponovitev poskuso do (vključno) n-te pojavitve izida A

- $P(X = k) = \binom{k-1}{n-1} (1-p)^{k-n} p^n$
- zaloga vrednosti
  - $k = n, n+1, n+2, \dots$
- pricakovana vrednost in disperzija
  - $E(X) = \frac{n}{p}$
  - $D(X) = \frac{n \cdot (1-p)}{p^2}$
- izpeljava pricakovane vrednosti za geometrijsko porazdelitev
  - Če mecemo kovanec toliko časa, da pade grb in z  $X$  označimo število potrebnih metov, vključno z zadnjim, potem je slučajna spremenljivka  $X$  geometrijsko porazdeljena.
  - Izračunajmo pricakovano vrednost
  - $E(X) = \sum_{i=1}^{\infty} i p q^{i-1} = p \sum_{i=1}^{\infty} (q^i)' = p \left( \sum_{i=1}^{\infty} q^i \right)' = p \left( \frac{1}{1-q} \right)' = \frac{1}{p}$
- uporaba pri problemu "zbiranja kuponov"
  - V trgovini lahko kupimo Kraseve čokoladice kraljestvo zivali. Vsaka čokoladica ima priloženo nalepko določene zivali. Vseh nalepk je 250, posamezna pa stane okoli 0.4. Pricakovana vrednost za takosno zbiranje pride  $E(X) = \frac{250}{0.4} = 625 > 150$  (knjiga o zivalih).
  - Nauk zgodbe: Bolj se splaca kupiti knjigo o zivalih, kot zbirati slicice, oz. če se vseeno odločimo da bomo zbirali slicice, se nam jih proti koncu splaca izmenjati z drugimi.

## 5. Hipergeometrijska porazdelitev

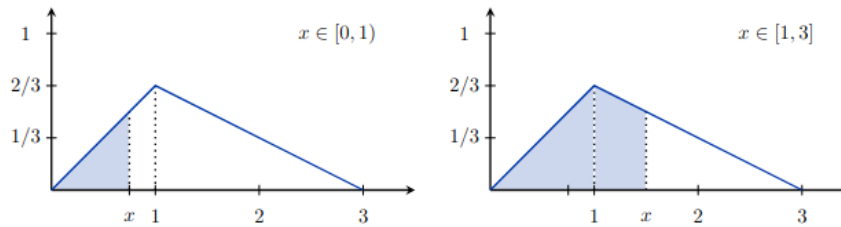
- zaloga vrednosti
  - $0, 1, 2, \dots$
- definicija (verjetnostna funkcija)
  - $X \sim (n; M, N)$
  - $X$  je število rdečih kroglic med izbranimi  $n$  kroglicami.
    - V posodi imamo  $M$  rdečih in  $N-M$  belih kroglic. Iz posode izvlecemo  $n$  kroglic.
  - $P(X = k) = \frac{\binom{M}{k} \cdot \binom{N-M}{n-k}}{\binom{N}{n}}$
- omejitve parametrov
  - $\max(0, n - (N - M)) \leq k \leq \min(M, n)$
  - $n \leq N$
- primer uporabe
  - Iz vreče, ki ima 4 modre in 5 rdečih potegnemo 3 kroglice.
  - Kolikšna je verjetnost da potegnemo 2 modri
    - $M=4, N-M=5, n=3$
  - $P(X = 2) = \frac{\binom{4}{2} \cdot \binom{5}{1}}{\binom{9}{3}} = 0.357$
- Standardni odklon in upanje
  - $E(X) = \frac{nM}{N}$
  - $D(X) = \frac{M(N-M)n(N-n)}{N^2(N-1)}$
- je več parametrov boljše ali slabše
  - Pri hipergeometrijski porazdelitvi imamo le dve kategoriji (ima lastnost, nima lastnosti). Ampak lahko bi imeli več kategorij, in posledično več informacij. Na splošno je pri merjenju bolje izbrati večje število parametrov/kategorij, saj na tak način pridobimo več informacij. Parametri določajo porazdelitev: nam pokažejo kje je povprečje(pricakovana vrednost), mediana, modus, oblika porazdelitve,...
- povezava z binomsko
  - Pri veliki seriji bi lahko vzeli binomsko porazdelitev (praktično vseeno ali izbiramo vzorec z vračanjem ali brez)
  - Imamo  $H(R, B, n)$ 
    - nastavimo  $p = \frac{R}{R+B} \rightarrow B(n, p)$

## 6. Zvezne slučajne spremenljivke

- primer - slika gostote verjetnosti

- Vsaka paleta paketov riza vsebuje 100kg. Slučajna spremenljivka ki steje kolicino prodanih palet riza ima gostoto

$$p_X(x) = \begin{cases} \frac{2x}{3}, & x \in [0, 1] \\ -\frac{x}{3} + 1, & x \in [1, 3] \\ 0, & x \notin [0, 3] \end{cases}$$



verjetnost je ploscina pod krivuljo, ki jo določa  $p_x(x)$

- opisi primerov: enakomerna, normalna, eksponentna (in Gama), Cauchyeva

- **enakomerna zvezna**  $\sim U[a, b]$

- vsi poskusi na intervalu  $[a, b]$  so enako verjetni

$$p_X(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b] \\ 0, & \text{sicer} \end{cases}$$

$$F_X(x) = \begin{cases} 0, & x \in [-\infty, a) \\ \frac{x-a}{b-a}, & x \in [a, b] \\ 1, & x \in (b, \infty] \end{cases}$$

- **eksponentna**  $\sim \text{Exp}(\lambda)$

- cas med dvema zaporednima dogodkoma na Poissonovem intervalu

$$p_X(x) = \begin{cases} 0 & x \leq 0 \\ \lambda e^{-x\lambda}, & x \geq 0 \end{cases}$$

$$F_X(x) = \begin{cases} 0, & x \leq 0 \\ 1 - e^{-\lambda x}, & x \geq 0 \end{cases}$$

- **Gamma**  $\sim \Gamma(n, \lambda)$

- cas med n zaporednimi dogodki na Poissonovem intervalu

$$p_X(x) = \begin{cases} 0, & x \leq 0 \\ \frac{\lambda^n x^{n-1} e^{-\lambda x}}{\Gamma(x)}, & x > 0 \end{cases}$$

$$\Gamma(x) = (x-1)\Gamma(x-1)$$

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$$

- **Normalna**  $\sim N(\mu, \sigma)$

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \text{ za } x \in \mathbb{R}$$

$$F_X(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

- definicija zvezne z gostoto verjetnosti

- Slučajna spremenljivka  $X$  je **zvezno porazdeljena** ce obstaja integrabilna funkcija  $p_X$  imenovana **gostota verjetnosti**

$$F(X) = P(X \leq x) = \int_{-\infty}^x p_X(t) dt, p_X(x) \geq 0$$

- $p_X(x)$  je integrabilna

$$p_X : \mathbb{R} \rightarrow [0, \infty)$$

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

- zveza med gostoto verjetnosti in porazdelitveno funkcijo

$$p_X(x) = F'_X(x)$$

$$F_X(x) = \int_{-\infty}^x p_X(x) dx$$

- racunanje verjetnosti na podintervalu

$$P(a < X < b) = \int_a^b p_X(x) dx$$

- izpeljava netrivialne pricakovane vrednosti ali odklona

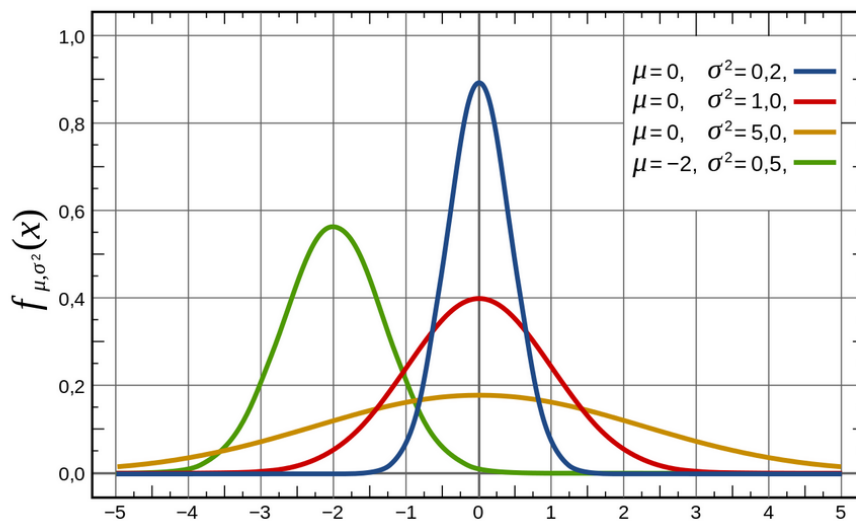
- TODO

## 7. Normalna porazdelitev

- slika: unimodalna, zvonasta krivulja

- $N(\mu, \sigma)$

- slika: unimodalna, zvonasta krivulja



- povezava z vsoto slučajnih spremenljivk in/ali CLI

- Naj bodo  $X_1, \dots, X_n$  **neodvisne** in **enako porazdeljene** slučajne spremenljivke s:

- $E(X_i) = \mu,$
- $D(X_i) = \sigma^2$

- Potem za dovolj velik  $n$  velja  $S = X_1 + \dots + X_n$

- $S \sim N(n \cdot \mu, \sqrt{n} \cdot \sigma)$

- Normalna porazdelitev je najpomembnejša oz. najpogostejše uporabljena porazdelitev v statistiki, saj marsikatera količina predstavlja vsoto mnogih drugih in je zato vsaj približno  $n$ . porazdeljena. Npr. Rezultat izpita, ki je sestavljen iz večjega števila kratkih vprašanj, je vsota posameznih vprašanj.

- zaloga vrednosti, gostota verjetnosti, funkcija napake

- $X \sim N(\mu, \sigma)$

- $Z_X = \mathbb{R}$

◦

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

- **funkcija napake**  $\phi(x) = F_X(x) = \int_{-\infty}^x e^{-\frac{y^2}{2}} dy$ , je liha, zvezno odvedljiva in strogo narasčujoča funkcija

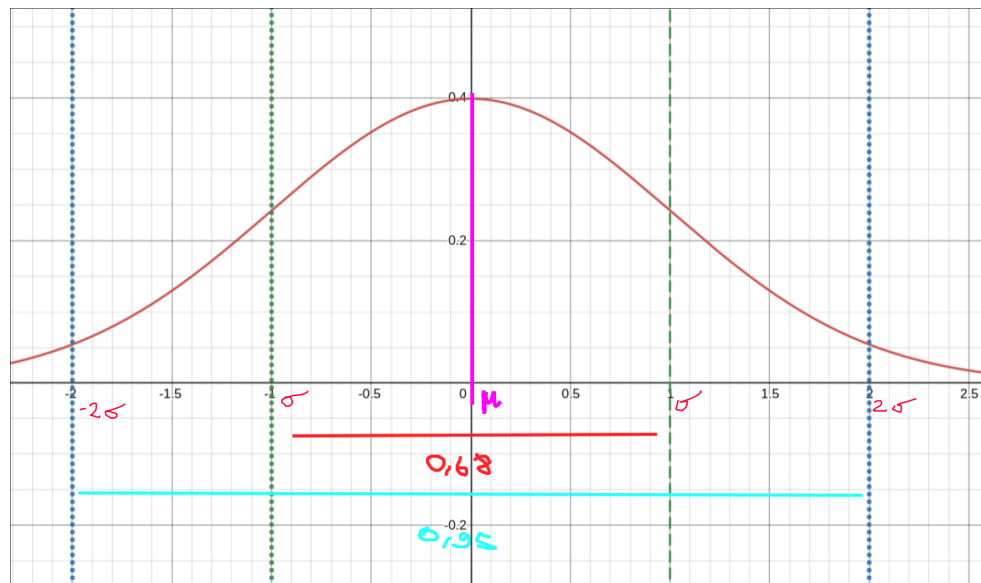
- posebne vrednosti funkcije napake:

- $\phi(0) = 0, \phi(\infty) = 0.5, \phi(-\infty) = -0.5$

- standardizacija in tabela za  $N(0, 1)$

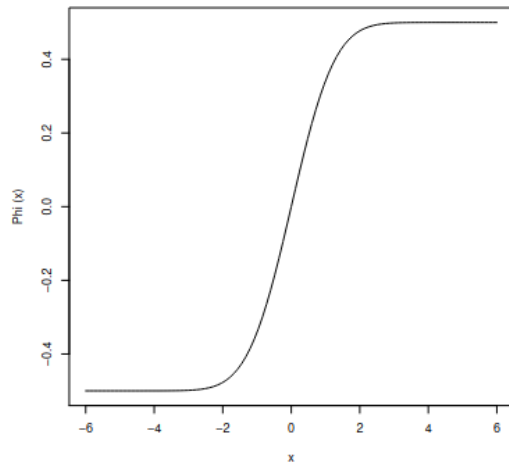
- Porazdelitev  $N(0, 1)$  je standardizirana normalna porazdelitev

- $N(0, 1)$



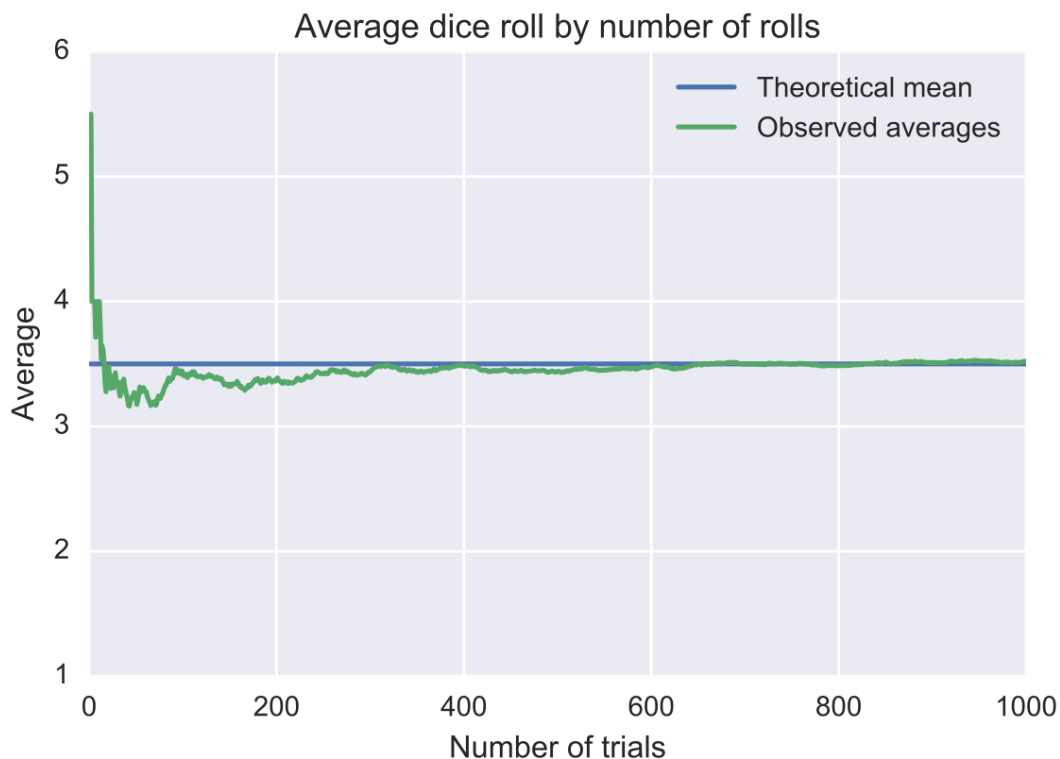
- Spremenljivko  $X \sim N(\mu, \sigma)$  pretvorimo s transformacijo  $Z = \frac{X - \mu}{\sigma}$  v standardno spremenljivko  $Z \sim N(0, 1)$
- kje sta  $\mu$  in  $\sigma$  na grafu  $y=p(x)$  in pravilo 68-95-99,7
  - $\mu$  se nahaja na simetrali zvonaste krivulje
  - $\mu + \sigma$  se nahaja za en standardni odklon naprej v pozitivni smeri x osi
    - Približno 68% površine pod krivuljo spada v en standardni odklon  $[\mu - \sigma, \mu + \sigma]$
    - Približno 95% površine pod krivuljo spada v dva standardna odklona  $[\mu - 2\sigma, \mu + 2\sigma]$
    - Približno 99% površine pod krivuljo spada v tri standardne odklone  $[\mu - 3\sigma, \mu + 3\sigma]$
- vecrazsezna gostota porazdelitve
  - $p(x) = \sqrt{\frac{\det A}{(2\pi)^n}} e^{-\frac{1}{2}(x-\mu)^T A(x-\mu)}$  kjer je matrika A kovariancijska matrika.
- vektorska oblika s kovariancijsko matriko
  - primer kovariancijske matrike za  $(X, Y) \sim N(\mu_x, \mu_y, \sigma_x, \sigma_y, \rho)$ 

$$\begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}$$
- Laplaceov tockovni obrazec - Funkcija napake
  - za p blizu 1/2 in velike n velja  $B(n, p) \approx N(np, \sqrt{npq})$
  - Zanima nas kolikšna je verjetnost  $P_n(k_1, k_2)$ , da se v Bernoullijevem zaporedju neodvisnih poskusov v n zaporednih poskusih zgodi dogodek A vsaj  $k_1$ -krat in manj kot  $k_2$  - krat
  - Oznacimo  $x_k = \frac{k-np}{\sqrt{npq}}$  in  $\Delta x_k = x_{k+1} - x_k = \frac{1}{\sqrt{npq}}$
  - Tedaj je ce upostevamo Laplaceov tockovni obrazec
  - $P_n(k_1, k_2) = \sum_{k=k_1}^{k_2-1} P_n(k) = \frac{1}{\sqrt{2\pi}} \sum_{k=k_1}^{k_2-1} e^{-\frac{1}{2}x_k^2} \Delta x_k \approx \frac{1}{\sqrt{2\pi}} \int_{x_{k_1}}^{x_{k_2}} e^{-\frac{1}{2}x^2} dx$
  - Funkcija napake je definirana:  $\phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{1}{2}t^2} dt$ 
    - je liha, zvezno odvedljiva in strogo narascujaca funkcija
    - $\phi(0) = 0$



## 8. Bernulliljev zakon velikih števil

- TODO str 58,59 skripta



- uvod
  - Zakon velikih števil je osnovni limitni izrek, ki opisuje rezultat izvajanja istega poskusa zelo velikokrat.
  - Po zakonu mora biti srednja vrednost rezultatov blizu pričakovane vrednosti (s številom poskusov se samo približuje)
  - $\overline{X}_n \rightarrow \mu$ , ko gre  $n \rightarrow \infty$
- podroben zapis izreka
  - Naj bo  $k$  frekvenca dogodka A v  $n$  neodvisnih ponovitvah danega poskusa, v katerem ima dogodek A verjetnost  $p$ .
  - Torej za  $\forall \epsilon > 0$  velja:

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{k}{n} - p\right| \leq \epsilon\right) = 1$$

$k$  = frekvenca dogodka A v  $n$  ponovitvah poskusa

- zgornja enačba: as the number of trials  $n$  goes to infinity, the average of the observations converges to the expected value

- skica dokaza
  - uporabimo **Chebyshevo neenakost**



- $P(|X - \mu| > a) \leq \frac{\text{Var}(X)}{a^2}$
- ta se dokazuje z **Markovo neenakostjo**
  - $P(|X| \geq a) \leq \frac{E(X)}{a}$ 
    - dokaz:  $aI_{|X| \geq a} \leq |X| \rightarrow aE(I_{|X| \geq a}) \leq E(X)$
- ter dejstvo:  $\text{Var}(\bar{X}_n) = \text{Var}(\frac{1}{n}(X_1 + \dots + X_n)) = \frac{\sigma^2 n}{n^2}$

$$P(|\bar{X}_n - \mu| > \epsilon) \leq \frac{\text{Var}(\bar{X}_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \rightarrow 0$$

- statistična in klasična definicija verjetnosti
  - del-0 (/teorija0.md/1#4.-popoln-sistem-dogodkov-in-definicije-verjetnosti)
- definicija funkcije napake
  - zgoraj (/1#7.-normalna-porazdelitev)
- uporaba/primer za izračun verjetnosti ali velikost n
  - Kolikokrat moramo vreci posten kovanec, da bo verjetnost dogodka, da se relativna frekvenca grba razlikuje od 0.5 za manj kot 0.05 večja od 0.997
  - Iz tabele vidimo  $2\phi(x) > 0.997$  za  $x = 3$
  - Poiščemo tak n, da bo  $\frac{n\epsilon}{\sqrt{npq}} > 3$
  - $\frac{0.05 \cdot \sqrt{n}}{\sqrt{0.5 \cdot 0.5}} > 3 \rightarrow \frac{0.0025n}{0.25} > 9 \rightarrow n = 900$
- dokaz z aproksimacijo prek zveze med binomsko in normalno porazdelitvijo
  - ker je n naravno število, lahko oba izraza v neenakosti iz zgornje verjetnosti pomnožimo z n, in z upoštevanjem da je tudi k celo število med 0 in n, dobimo oceno

$$P\left(-\epsilon \leq \left|\frac{k}{n} - p\right| \leq \epsilon\right) = P(np - n\epsilon \leq k \leq np + n\epsilon) = P_n(k_1) + P_n(k_1 + 1) + \dots + P_n(k_2)$$

- kjer so  $k_1 < k_1 + 1 < \dots < k_2$  vsa cela števila na intervalu  $[np - n\epsilon, np + n\epsilon]$
- Dobljeno vsoto označimo s  $P(k_1 - 1, k_2)$  in jo ocenimo s funkcijo napake kar nam da

$$P(k_1 - 1, k_2) \approx \phi\left(\frac{k_2 - np}{\sqrt{npq}}\right) - \phi\left(\frac{k_1 - 1 - np}{\sqrt{npq}}\right) \approx 2\phi\left(\frac{n\epsilon}{\sqrt{npq}}\right)$$

- Zakaj to smatramo za prvo verzijo CLI?
  - Pri velikem številu ponovitev poskusa, povprečje dobljenih vrednosti konvergira k pričakovani vrednosti.

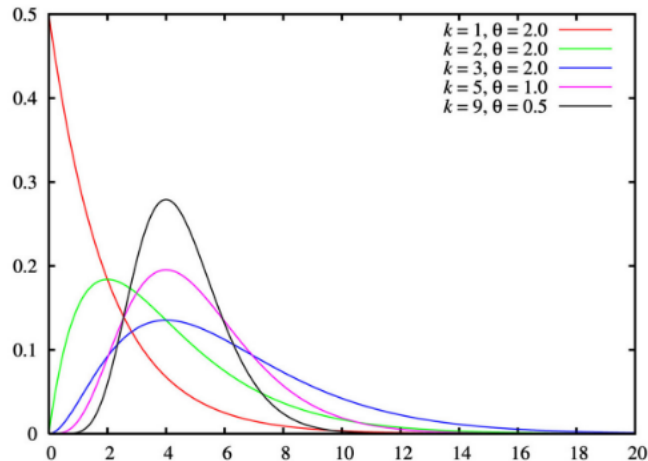
## 9. Eksponentna porazdelitev

- primer
  - Studenti prihajajo v klub porazdeljeni približno po Poissonu, s povprečno stopnjo 30 studentov na uro.
  - Kolikšna je verjetnost, da bo vratar čakal več kot 3 minute na naslednjega studenta?
    - $P(X > \frac{1}{20}) = 1 - F(\frac{1}{20}) = 1 - (1 - e^{-30 \cdot \frac{1}{20}}) = 0.223$
- definicija - povezava s Poissonovim procesom
  - $X \sim \text{Exp}(\lambda)$
  - Opisuje čas med dvema zaporednima dogodkoma v Poissonovem procesu
    - tj. proces kjer se dogodki pojavljajo zvezno in neodvisno pri povprečni hitrosti ponavljanja
    - hkrati je tudi analog geom. porazdelitve
- slika, zaloga vrednosti, gostota verjetnosti
  - $p(x) = \lambda e^{-\lambda x}, x \geq 0$
  - $Z_f = [0, \infty]$
- porazdelitvena funkcija
  - $F(X) = \int_0^x \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x}$
- razlaga parametra  $\lambda$ 
  - Povprečno število dogodka A na danem (casovnem) intervalu
- pričakovana vrednost in disperzija
  - $E(X) = \frac{1}{\lambda}$
  - $D(X) = \frac{1}{\lambda^2}$
- karakterizacija: zvezna slučajna spremenljivka brez spomina
  - Poissonova porazdelitev: pojavitev dogodka je nedovisna od tega kdaj se pojavi dogodek pred tem
  - $P(X > x + y | X > x) = P(X > y)$

- npr. življenska doba zarnice

## 10. Gama porazdelitev

- posebni primeri (npr. eksponentna, hi-kvadrat)
  - Eksponentno porazdelitev lahko se posplošimo: pri Poissonovem procesu merimo čas da se zgodi  $k$  zaporednih dogodkov.
  - Naj bosta  $k, \lambda > 0$ . Tedaj imamo Gama Porazdelitev  $X \sim \Gamma(k, \lambda)$ 
    - $k$  je število dogodkov, za katere čakamo.
    - $\lambda$  parameter raztega, pove nam kako se dogodki časovno odvijajo
    - na sliki  $\theta = \frac{1}{\lambda}$ , za  $k = 1$  dobimo eksponentno porazdelitev



- Kako lahko iz eksponentne pridemo do Gama porazdelitve
  - za parameter pri gama  $k = 1$  ali;
  - Če  $X_i \sim \text{Exp}(\lambda)$
  - potem  $\Gamma(k, \lambda) = X_1 + X_2 + \dots + X_k$
- slika, zaloga vrednosti, gostota verjetnosti in kako pridemo do zgornjih primerov iz Gama porazdelitve
  - $$p(x) = \begin{cases} \frac{\lambda^k}{\Gamma(k)} x^{k-1} e^{-\lambda x}, & x > 0 \\ 0, & \text{sicer} \end{cases}$$
  - $Z_X = \mathbb{R}^+$
- Gama funkcija (definicija, rekurzija in nekatere vrednosti, tudi za  $1/2$ )
  - Funkcijo gama definiramo z določenim integralom

$$\Gamma(z) = \int_0^{\infty} t^{z-1} e^{-t} dt$$

- $\Gamma(1) = 1$
- $\Gamma(\frac{1}{2}) = \sqrt{\pi}$
- Izražena rekurzivno (prek per partes):  $\Gamma(x) = (x-1)\Gamma(x-1)$
- Za naravno število dobimo tako  $\Gamma(n) = (n-1)!$
- Príakovana vrednost in disperzija
  - $E(X) = \frac{k}{\lambda}$
  - $D(X) = \frac{k}{\lambda^2}$
- Uporaba hi-kvadrat porazdelitve
  - Hi-kvadrat je poseben primer gama porazdelitve
  - $\chi^2(n) = \Gamma(\frac{n}{2}, \frac{1}{2})$
  - $\chi^2$  test se uporablja za ugotavljanje razlike kategoricnih spremenljivk (npr. ali je kocka obtežena). Uporabljamo jo v dveh primerih:
    - kako dobro se opazovana/izmerjena porazdelitev prilega pričakovani
    - ocenjevanje ali sta naključni spremenljivki neodvisni

Primer (ali je kocka obtežena?): 60x vzremo kocko in primerjamo rezultate s pričakovanimi vrednostmi. Če pride do velikega odstopanja, lahko sklepamo, da je kocka nepostena.

| st. pik         | 1  | 2  | 3  | 4  | 5  | 6  | skupaj |
|-----------------|----|----|----|----|----|----|--------|
| opazovana kocka | 14 | 12 | 16 | 8  | 5  | 5  | 60     |
| teoreticno      | 10 | 10 | 10 | 10 | 10 | 10 | 60     |

Uporabimo formulo  $\chi^2(n-1) = \frac{(E_1 - O_1)^2}{E_1} + \dots + \frac{(E_n - O_n)^2}{E_n}$ , kjer je  $E_i$  pričakovana vrednost posameznega izida,  $O_i$  pa st. pojavitev posameznega izida opazovane kocke. Če ustavimo podatke v formulo, dobimo rezultat 11 (odstopanje 5%).

Odstopanje ni zadostno (morali bi vzeti večji vzorec), tako da ne moremo sklepati ali je kocka postena ali ne.

- Dedna lastnost gama porazdelitve, skica dokaza za aditivnost, t.j.  $\Gamma(k, \lambda) + \Gamma(h, \lambda) = \Gamma(k + h, \lambda)$ 
  - Če sta  $X \sim \Gamma(k_1, \lambda)$  in  $Y \sim \Gamma(k_2, \lambda)$  neodvisni slučajni spremenljivki, je tudi njuna vsota:  $Z = X + Y$  porazdeljena po porazdelitvi  $Z \sim \Gamma(k_1 + k_2, \lambda)$