

## 2. del

### 1. Pricakovana vrednost slučajne spremenljivke (matematično upanje)

- primer: npr. igralna kocka (3.5) ali binomska (np)

- $X \sim \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix}$

- $E(X) = \sum_{i=1}^6 x_i \cdot p_i = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \dots + 6 \cdot \frac{1}{6} = 3.5$

Porazdelitev	E(X)
$B(p)$	$p$
$B(n, p)$	$np$
$G(p)$	$\frac{1}{p}$
$P(n, p)$	$\frac{n}{p}$
$H(R, B, n)$	$\frac{nR}{R+B}$
$E([a, b])$	$\frac{a+b}{2}$
$P(\lambda)$	$\lambda$
$\text{Exp}(\lambda)$	$\frac{1}{\lambda}$
$\Gamma(n, \lambda)$	$\frac{n}{\lambda}$
$N(\mu, \sigma)$	$\mu$
$\chi^2(n)$	$n$

- motivacija za definicijo (utezeno povprecje), tj.

tezisce

- povprecje vrednosti diskretne spremenljivke (verjetnost · vrednost)

- utezeno povprecje:  $k_1 + \dots + k_m = N$ ,  $f_i = \frac{k_i}{N}$

$$\bar{x} = \frac{x_1 k_1 + \dots + x_m k_m}{N} = x_1 f_1 + \dots + x_m f_m$$

- predstavlja  $\mu$  pri CLI

$$E(X) = \sum_{i=1}^n x_i \cdot p_i$$

- $P(X = a) = 1 \Rightarrow E(X) = a$

- definicija za diskretne slučajne spremenljivke (kdaj obstaja)

- diskretna slučajna spremenljivka je slučajna spremenljivka, ki ima stevno zalogo vrednosti

- $E(X) = \sum_{i=1}^n x_i \cdot p_i$

- **pogoj**  $\sum_{i=1}^n |x_i| \cdot p_i < \infty$

- Slučajna spremenljivka  $X$  ima pričakovano vrednost, ko ga ima slučajna sp.  $|X|$ . Očitno velja  $|E(X)| \leq E(|X|)$ .

- definicija za zvezne slučajne spremenljivke (kdaj obstaja)

- zvezna slučajna spremenljivka je slučajna spremenljivka ki lahko zavzameo katerokoli vrednost iz nekega intervala

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

- **pogoj:**  $\int_{-\infty}^{\infty} |x| p_X(x) dx < \infty$

- primer slučajne spremenljivke za katero ne obstaja  $E(X)$

- $X \sim p(x) = \frac{1}{\pi(1+x^2)}$  **Caucheyeva porazdelitev**

- $x_k = (-1)^{k+1} \frac{2^k}{k}$  in  $p_k = 2^{-k}$  (primer diskretne ko je  $E(X) = \infty$ ) V tem primeru bi morala biti

končna vsota:  $s = \frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \dots$ , opazimo  $\frac{1}{3} + \frac{1}{4} > \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$ , ter v splošnem

$$\frac{1}{2^n + 1} + \dots + \frac{1}{2^{n+1}} > \frac{2^n}{2^n + 1} = \frac{1}{2}$$

torej velja  $S > \sum_{n=1}^{\infty} \frac{n}{2} \rightarrow$  neskončna vsota.

- lastnosti: linearnost, z dokazom za homogenost

- **linearnost:**  $E(aX + bY) = aE(x) + bE(Y)$

- **homogenost:**  $E(aX) = aE(X)$ :

- Dokaz : homogenost in linearnost **integriranja** ter **vsote** (po definiciji)

- $\sum_{i=1}^{\infty} a \cdot x_i p_{X_i} + \sum_{i=1}^{\infty} b \cdot y_i p_{Y_i} = a \cdot \sum_{i=1}^{\infty} x_i p_{X_i} + y_i p_{Y_i}$

- $E(X)$  je definirana kot povprečje diskretnih spremenljivk, ce vse to pomnozis s konstanto se mnozi tudi povprečje

- Skica dokaz aditivnosti  $E(X+Y) = E(X)+E(Y)$  = ali Trditev 7.1:

$$E(|XY|) \leq \sqrt{E(X^2)E(Y^2)}$$

◦

$$E(X + Y) = E(Z) = \int_{-\infty}^{\infty} z p_z(z) dz$$

$$= \int_{-\infty}^{\infty} z \left( \int_{-\infty}^{\infty} p(x, z - x) \right) dz = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x + y) p(x, y) dx dy$$

$$\int_{-\infty}^{\infty} \left( \int_{-\infty}^{\infty} x p(x, y) dx \right) dy + \int_{-\infty}^{\infty} \left( \int_{-\infty}^{\infty} y p(x, y) dx \right) dy$$

$$= \int_{-\infty}^{\infty} xp_X(x)dx + \int_{-\infty}^{\infty} yp_X(y)dy$$

$$= E(X) + E(Y)$$

- lastnosti vsot/integralov

## 2. Disperzija (razpršenost oz. varianca) slučajne spremenljivke, odklon in standardizacija

- primeri: npr. binomska ( $np(1-p)$ ), enakomerna ( $(b-a)^2/12$ )

porazdelitev	D(X)
$B(p)$	$p(1-p)$
$B(n, p)$	$np(1-p)$
$G(p)$	$\frac{1-p}{p^2}$
$P(n, p)$	$\frac{n(1-p)}{p^2}$
$H(R, B, n)$	$\frac{nRB \cdot (R+B-n)}{(R+B)^2(R+B-1)}$
$P(\lambda)$	$\lambda$
$\text{Exp}(\lambda)$	$\frac{1}{\lambda^2}$
$\Gamma(n, \lambda)$	$\frac{n}{\lambda^2}$
$E([a, b])$	$\frac{(b-a)^2}{12}$
$\chi^2(n)$	$2n$

- definicija s pričakovano vrednostjo in obstoj
  - **Definicija:**  $D(X) = E((X - E(X))^2) = E(X^2) - E^2(X)$
  - $D(X) \geq 0$
  - **Pogoj:** nesme biti neskončna
  - $D(aX) = a^2 D(X)$  in  $D(x + a) = D(X)$
- $D(X) = 0 \iff X$  je konstanta
  - $D(X) = E((X - E(X))^2) = E((X - X)^2) = E(0) = 0$
  - Torej ce je npr  $P(X = a) = c$ , Bo odklon vseh ostalih posameznih vzorcev od povprecja = 0;
- lastnosti (aditivnost za neodvisni slučajni spremenljivki)
  - Aditivnost disperzije:  $D(X + Y) = D(X) + D(Y) + 2\text{Cov}(X, Y)$ 
    - ce sta X in Y neodvisni:  $D(X + Y) = D(X) + D(Y)$
- Standardizacija, slučajne spremenljivke in njena pričakovana vrednost oz. odklon

- Slučajno spremenljivko  $X$  **standardiziramo** s transformacijo

$$X_s = \frac{X - \mu}{\sigma}$$

- $E(X) = \mu$  in  $D(X) = \sigma^2$

- **Velja:**

- $E(X_s) = E\left(\frac{X - \mu}{\sigma}\right) = \frac{E(X - \mu)}{\sigma} = \frac{\mu - \mu}{\sigma} = 0$

- $D(X_s) = D\left(\frac{X - \mu}{\sigma}\right) = \frac{D(X - \mu)}{\sigma^2} = \frac{\sigma^2 + 0}{\sigma^2} = 1$

- Standardizacija slučajne spremenljivke povzroči da vsaka spr. enako vpliva na pričakovano vrednost (npr. ce smo neke rezultate zmerili z različnimi merili, to nam pomaga za primerjavo različnih tipov spremenljivk)

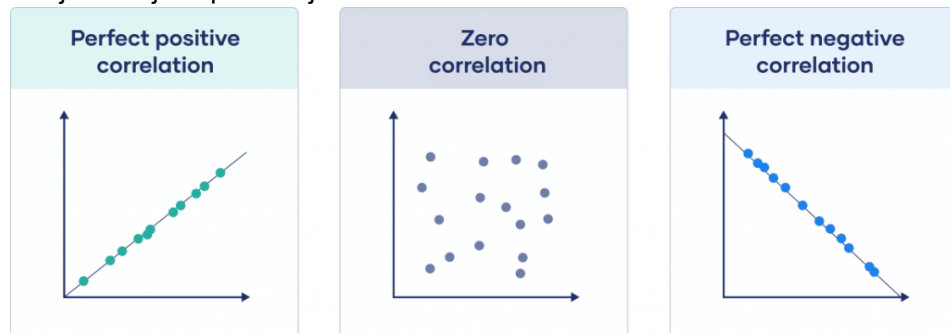
- Skica dokaza zveze "kosinusni izrek"

- $D(X + Y) = D(X) + D(Y) + 2Cov(X, Y) \sim (x + y)^2 = x^2 + y^2 + 2\cos\alpha$

- analogija: Disperzijo bi lahko primerjali s tipicnim geometrijskim kosinusnim izrekom. Ko sta si stranici pravokotni se izraz izracuna neodvisno od  $\cos(90) = 0$ . Ce pogledamo pri disperziji, ko sta si spremenljivki nedovisni je njuna kovarianca  $Cov(X, Y) = 0$ .

### 3. Korelacija in kovarianca

- meri algebraicno povezanost dveh številskih slučajnih spremenljivk
  - predstavlja mero povezanosti med dvema spremenljivkama
    - os  $y$  predstavlja slučajno spremenljivko  $Y$
    - os  $x$  predstavlja slučajno spremenljivko  $X$



Scribbr

- definicija kovariance in njen obstoj ( $CS E(|XY|) \leq \sqrt{E(X^2)E(Y^2)}$ )
  - **Kovarianca**  $Cov(X, Y) = E((X - E(X)) \cdot (Y - E(Y))) = E(XY) - E(X)E(Y)$
  - $|Cov(X, Y)| \leq \sqrt{D(X)D(Y)} = \sigma_X \sigma_Y$
  - spremenljivki za katere velja da:
    - $Cov(X, Y) \neq 0$  sta **korelirani**
      - $Cov(X, Y) > 0 \rightarrow X \uparrow Y \uparrow$  (pozitivno korelirani)
      - $Cov(X, Y) < 0 \rightarrow X \uparrow Y \downarrow$  (negativno korelirani)
    - $Cov(X, Y) = 0$  sta **nekorelirani**
  - **Primer**  $X \equiv$  telesna visina osebe,  $Y \equiv$  teza osebe
    - $X$  in  $Y$  sta korelirani (pozitivno)
- lastnosti korelacije

- Če sta slučajni spremenljivki  $X$  in  $Y$  neodvisni  $\rightarrow E(XY) = E(X)E(Y) \rightarrow \text{Cov}(X, Y) = 0$
- definicij korelacijskega koeficienta, vedno na  $[-1, 1]$ 
  - Korelacijski koeficient vpeljemo zato, ker je moc povezanosti med dvema s.s. težko ocentit preko kovariance, zato jo delimo s obema std. odklonoma
  - $$r(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma(X)\sigma(Y)} = \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X^2) - E^2(X)} \cdot \sqrt{E(Y^2) - E^2(Y)}}$$
  - $r(X, Y) = 0 \Leftrightarrow X$  in  $Y$  nekorelirani
  - $r(X, Y) = \pm 1 \Leftrightarrow X$  in  $Y$  sta v linearni zvezi
- kdaj lahko zaključimo linearno odvisnost
  - Dve slučajni spremenljivki sta **linearno odvisni** ce lahko eno zapisemo kot linearno funkcijo druge  $\rightarrow$  korelacijski koeficient med njima bo 1 ali -1.
- Ali lahko izračunamo korelacijo iz disperzij (tj.  $D(X)$ ,  $D(Y)$  in  $D(X+Y)$ )
  - $D(X + Y) = D(X) + D(Y) + 2\text{Cov}(X, Y)$
- kovariančna matrika
  - Naj bo  $X$  stolpčni vektor  $X = (X_1, X_2, \dots, X_n)^T$  (kjer so  $X_i$  slučajne spremenljivke)
  - Potem definiramo kovariančno matriko  $K_{XX}$ ,  $(i, j)$ -ti element je (kovarianca):
    - $K_{X_i X_j} = \text{cov}[X_i, X_j] = E[(X_i - E[X_i])(X_j - E[X_j])]$
  - Kovariančna matrika je **simetrična**
  - Diagonalne vrednosti so **disperzije**

$$K_{X_i X_i} = E((X_i - E(X_i))(X_i - E(X_i))) = E((X_i - E(X_i))^2) = D(X_i)$$
- povezava z regresijsko premico
  - $K_{YX} K_{XX}^{-1}$  je matrika regresijskih koeficientov

## 4. Slučajni vektorji 2D, 3D, nD

- definicija slučajnega vektorja (primer)
  - Slučajni vektor je  $n$ -terica slučajnih spremenljiv  $X = (X_1, \dots, X_n)$
  - Primer slučajni vektor  $X = (X_1, X_2)$ :
    - $X_1$  stevilo metov ko pade sestica, pri 3 metih kocke
    - $X_2$  stevilo metov ko pade stevilo manjše od 3, pri 3 metih kocke
- definicija porazdelitvene funkcije (primer)
  - $F(x_1, x_2, \dots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n)$
  - Primer za metanje kock
    - $$F_{X,Y}(x, y) = P(X \leq x, Y \leq y) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} p_{ij} \cdot 0/1 \quad (1 \text{ ce je } i \leq x \text{ in } j \leq y \text{ 0 sicer})$$
  - Za zvezni vektor uporabimo integrale
    - $$F_{X,Y}(x, y) = P(X \leq x, Y \leq y) = \int_{-\infty}^x \int_{-\infty}^y p_{X,Y}(x, y) dx dy$$
- verjetnostna in kontingencna tabela, verjetnostna funkcija (primer)
  - **verjetnostna tabela**

$X, Y$	$y_1$	$y_2$	$\dots$	$y_m$	$X$
$x_1$	$p_{11}$	$p_{12}$	$\dots$	$p_{1m}$	$p_1$
$x_2$	$p_{21}$	$p_{22}$	$\dots$	$p_{2m}$	$p_2$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$x_n$	$p_{n1}$	$p_{n2}$	$\dots$	$p_{nm}$	$p_n$
$Y$	$q_1$	$q_2$	$\dots$	$q_m$	1

▪ kjer:

- $P(X = x_i, Y = y_j) = p_{ij}$
- $P(X = x_i) = p_i$ , **robna porazdelitev** za  $X$
- $P(Y = y_i) = q_i$ , **robna porazdelitev** za  $Y$

◦ **kontingencna tabela** (modelira nakup avtomobila)

starost	>20	<20	sum
kupil avtomobil	80	20	100
ni kupil	100	50	150
sum	180	70	250

• gostota verjetnosti (primer)

- funkciji  $p_{X,Y}$  pravimo (dvorazsezna) gostota verjetnosti (določa vektor zveznih spremenljivk)
- npr:  $p_{X,Y}(x, y) = \frac{1}{4}$  z zalogo vrednosti  $x \in [0, 2], y \in [0, 2]$

• robne porazdelitvene funkcije

- Funkciji  $F_i(x_i) = F(\infty, \dots, \infty, x_i, \infty, \dots, \infty)$  pravimo **robna porazdelitvena funkcija** spremenljivke  $X_i$
- Npr za diskretne

$$\begin{aligned} \text{▪ } P(X = x_i) &= \sum_{j=1}^{\infty} p_{ji} \\ \text{▪ } P(Y = y_i) &= \sum_{j=1}^{\infty} p_{ij} \end{aligned}$$

• Ali se da iz verjetnostne funkcije slučajnega vektorja ugotoviti neodvisnost njegovih komponent?

◦ **DA**, npr za diskretni vektor:

$$\text{▪ } \forall x, y : P(X = x, Y = y) = P(X = x)P(Y = y)$$

◦ npr, za zvezni vektor

$$\text{▪ } \forall x, y, X \leq x \wedge Y \leq y : p_{X,Y}(x, y) = p_X(x)p_Y(y)$$

• zveza med gostoto verjetnosti in porazdelitveno funkcijo

◦ za dve spremenljivki (na  $n$  spremenljivk trivialen prehod):

$$\text{▪ } F_{X,Y}(x, y) = P(X \leq x, Y \leq y) = \int_{-\infty}^x \int_{-\infty}^y p_{X,Y}(x, y) dx dy$$

• definicija kvadranta in izpeljava formule pravokotnik

- Naj bo  $A(x, y) = \{(u, v) \in \mathbb{R}^2 : u \leq x \wedge v \leq y\}$  (levi spodnji kvadrant glede na  $(x, y)$ )
- Naj porazdelitvena funkcija opisuje verjetnost da je slučajna točka  $(X, Y)$  v množici  $A(x, y)$

$$F(x, y) = p(X \leq x, Y \leq y) = P(X, Y) \in A(x, y)$$

- Tedaj je verjetnost da je slučajna točka  $(X,Y)$  v pravokotniku  $(a, b] \times (c, d]$  enaka:

$$P(X, Y) \in (a, b] \times (c, d] = F(b, d) - F(a, d) - F(b, c) + F(a, c)$$

- neodvisnost
  - diskretne:  $P(X = x) \cdot P(Y = y) = P(X = x, Y = y)$
  - zvezne:  $P_{X,Y}(X, Y) = P_X(X) * P_Y(Y)$

## 5. Polinomska porazdelitev

- primeri
  - Imamo volitve z 3 izbirami (A,B,C). Kandidat A prejme 20% glasov, B 30%, C 50% glasov. Če so glasovalci izbrani randomly, kaksna je verjetnost da bomo izmed 6 izbranih izbrali natanko enega volivca za kandidata A, dva za B in tri za C.

$$X \sim P(6; 0.2, 0.3, 0.5)$$

$$P(A = 1, B = 2, C = 3) = \frac{6!}{1!2!3!} (0.2^1)(0.3^2)(0.5^3)$$

- Iz kupa igralnih kart (52) na slepo izberemo eno karto in jo nato vrnemo nazaj. Postopek ponovimo 5-krat. Kolikšna je verjetnost da bomo videli dvakrat srce, po enkrat pa pika kriza in karo

$$X \sim P(5, 0.25, 0.25, 0.25, 0.25)$$

$$P(X_1 = 2, X_2 = 1, X_3 = 1, X_4 = 1) = \frac{5!}{2!1!1!1!} 0.25^2 0.25^1 0.25^1 0.25^1 = 0.05859$$

- definicija
  - Polinomska porazdelitev  $\sim P(n; p_1, \dots, p_r)$  je določena s predpisom

$$\blacksquare P(X_1 = k_1, \dots, X_r = k_r) = \frac{n!}{k_1! \dots k_r!} p_1^{k_1} \dots p_r^{k_r}$$

- zaloga vrednosti

$$\circ \sum_{i=1}^r p_i = 1$$

$$\circ \sum_{i=1}^r k_i = n$$

- verjetnostna funkcija (zapisi  $p_i, j, \dots, k$ )

- Polinomska porazdelitev  $\sim P(n; p_1, \dots, p_r)$

$$\blacksquare \sum p_i = 1, \sum k_i = n$$

- spremenljivke  $X_i$  opisujejo število pojavitev rezultata i

$$\circ P(X_1 = k_1, \dots, X_r = k_r) = \frac{n!}{k_1! \dots k_r!} p_1^{k_1} \dots p_r^{k_r}$$

- povezava z binomsko

- je posplošitev binomske porazdelitve

- za  $r=2$  dobimo binomsko spremenljivko  $B(n, p) = P(n, p, q)$

- pričakovana vrednost in disperzija

$$\circ E(X_i) = np_i$$

$$\circ D(X_i) = np_i(1 - p_i)$$

## 6. Funkcije slučajnih spremenljivk

- primeri enostavnih funkcij

- Imamo  $X \sim \begin{pmatrix} -1 & 0 & 1 \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{6} \end{pmatrix}$
- Potem  $5X \sim \begin{pmatrix} -5 & 0 & 5 \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{6} \end{pmatrix}$
- Imamo podano porazdelitveno shemo

$Y, X$	0	1	2	$X$
0	$\frac{2}{50}$	$\frac{2}{50}$	$\frac{1}{50}$	$\frac{5}{50}$
1	$\frac{6}{50}$	$\frac{6}{50}$	$\frac{3}{50}$	$\frac{15}{50}$
2	$\frac{12}{50}$	$\frac{12}{50}$	$\frac{6}{50}$	$\frac{30}{50}$
$Y$	$\frac{20}{50}$	$\frac{20}{50}$	$\frac{10}{50}$	1

- $Z = X^2 \sim \begin{pmatrix} 0 & 1 & 4 \\ 0.1 & 0.3 & 0.6 \end{pmatrix}$
- $W = X + Y \sim \begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ 0.04 & 0.16 & 0.38 & 0.30 & 0.12 \end{pmatrix}$ 
  - $P(W = 0) = P(X = 0, Y = 0) = 0.04$
  - $P(W = 1) = P(X = 0, Y = 1) + P(X = 1, Y = 0) = 0.16$
  - $P(W = 2) = P(X = 0, Y = 2) + P(X = 1, Y = 1) + P(X = 2, Y = 0) = 0.38$
  - $P(W = 3) = P(X = 1, Y = 2) + P(X = 2, Y = 1) = 0.30$
  - $P(W = 4) = P(X = 2, Y = 2) = 0.12$

- definicija in povezava med ustreznima porazdelitvenima funkcijama

- naj bo  $Y = g(X)$
- $F_Y(y) = P(g(X) \leq y) = \begin{cases} P(X \leq g^{-1}(y)), & g^{-1} \text{ narascujoca} \\ P(X \geq g^{-1}(y)), & g^{-1} \text{ padajoca} \end{cases}$
- **Primer:** naj bo  $Y = X^2$ 
  - $P(X^2 \leq y) = P(|X| \leq \sqrt{y}) = P(-\sqrt{y} \leq X \leq \sqrt{y})$

- zveza med gostatami verjetnosti

- Če je  $X$  porazdeljena z zvezno gostoto  $p(x)$ , je  $F_Y(y) = \int_{-\infty}^{f^{-1}(y)} p(x) dx$ ,
- in če je  $f$  se odvedljiva velja  $p_Y(x) = p(f^{-1}(y)) f^{-1}(y)'$

- formula za pričakovano vrednost

- diskretna:

$$E(f(X)) = \sum_{k=0}^{\infty} f(x_k) \cdot p_k$$

- zvezna



$$E(f(X)) = \int_{-\infty}^{\infty} f(x) \cdot p_X(x) dx$$

- neodvisnost
  - Če so  $X_1, X_2, \dots, X_n$  neodvisne standardizirane normalne slučajne spremenljivke, je slučajna spremenljivka  $Y = X_1^2 + \dots + X_n^2$  porazdeljena po  $\chi^2(n)$
- izpeljava zveze med  $N(0, 1)$  in  $\chi^2(1)$ 
  - Naj bo  $X$  standardizirana normalna spremenljivka z  $p_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$
  - nastavimo  $Y = X^2$ 
    - $g^{-1}(x) = \sqrt{x}$
    - $\frac{dg^{-1}}{dx} = \frac{1}{2} x^{-\frac{1}{2}}$
  - $p_Y(x) = p_X(g^{-1}(x)) \cdot \frac{dg^{-1}}{dx} = \frac{1}{\sqrt{2\pi}} e^{-\frac{x}{2}} \cdot \frac{1}{2} x^{-\frac{1}{2}} = \frac{1}{\sqrt{2\pi x}} e^{-\frac{x}{2}}$
  - Dobili smo  $\chi^2(1)$ 
    - $X_i \sim N(0, 1)$
    - $\chi^2(k) = X_1^2 + \dots + X_k^2$

## 7. Funkcije slučajnih vektorjev

- primer
  - Izračun pričakovane vrednosti ( $E(X)$ )
- definicija
  - Naj bo  $f: (x, y) \rightarrow (u, v)$  transformacija slučajnega vektorja  $(X, Y)$  v slučajni vektor  $(U, V)$  določena z zvezama  $U = u(X, Y)$  in  $V = v(X, Y)$ . Porazdelitveni zakon za nov slučajni vektor  $(U, V)$  je
  - $F_{U,V}(u, v) = P(U < u, V < v) = P((U, V) \in A(u, v)) = P(X, Y \in f^{-1}(A(u, v)))$
- definicija konvolucije
  - Definiramo  $Z = X + Y$ , kjer je  $(X, Y)$  zvezno porazdeljen slučajni vektor z gostoto  $p(x, y)$  verjetnostna funkcija (diskretni)
  - $F_Z(z) = P(Z \leq z) = P(X + Y \leq z) = \int \int_{x+y \leq z} p(x, y) dx dy = \int_{-\infty}^{\infty} dx \int_{-\infty}^{z-x} p(x, y) dy$
  - za slučajni dobimo gostoto verjetnosti:

$$p_Z(z) = \int_{-\infty}^{\infty} f_{XY}(x, z-x) dx$$

- za diskretni pa dobimo

$$P(Z = z) = \sum_{k=-\infty}^{\infty} P(X = k) P(Y = z - k)$$

- ce sta neodvisni pa

$$p_Z(z) = \int_{-\infty}^{\infty} p_X(x)p_Y(z-x)dx$$

- Gostota  $p_Z = p_X + p_Y$  je **konvolucija** funkcij  $p_X$  in  $p_Y$ .
- uporaba za vsoto dveh neodvisnih normalnih porazdelitev
  - $X \sim N(\mu_X, \sigma_X^2)$
  - $Y \sim N(\mu_Y, \sigma_Y^2)$
  - $Z = X + Y \rightarrow Z \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$
- uporaba za vsoto dveh neodvisnih Gama porazdelitev
  - $X \sim \Gamma(n_1, \lambda)$ , in  $Y \sim \Gamma(n_2, \lambda)$  potem  $X + Y \sim \Gamma(n_1 + n_2, \lambda)$
- formula za pričakovano vrednost produkta
  - $E(XY) = E(E(XY|Y))$
  - ce sta neovidsni
    - $E(XY) = E(X) \cdot E(Y)$
- Jacobijeva determinanta in prehod na nove spremenljivke
  - Ce je funkcija  $f$  bijektivna z zveznimi parcialnimi odvodi lahko nadaljujemo

$$F_{U,V}(u, v) = \int \int_{A(u,v)} p(x(u, v), y(u, v)) \left| \det \begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{pmatrix} \right| dudv$$

## 8. Pogojna porazdelitev

- primer za diskretni slučajni vektor
  - Imamo porazdelitveno shemo. Zapisi pogojno verjetnostno porazdelitev slučajne spremenljivke  $X$ , glede na pogoj  $y=2$

$Y \setminus X$	1	2	3	4	$Y$
0	0	0.10	0.20	0.10	0.40
1	0.03	0.07	0.10	0.05	0.25
2	0.05	0.10	0.05	0	0.20
3	0	0.10	0.05	0	0.15
$X$	0.08	0.37	0.40	0.15	1

- Verjetnost v vrstici pri  $y = 2$ , moramo deliti s  $P(Y = 2) = 0.2$

$$X|y = 2 \sim \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0.25 & 0.5 & 0.25 & 0 \end{pmatrix}$$

- definicija v diskretnem primeru in v zveznem primeru
  - diskretni primer:
    - Naj bo  $B$  nek mogoc dogodek  $P(B) > 0$ . Potem lahko vpeljemo **pogojno porazdelitveno funkcijo**:

$$F(x|B) = P(X \leq x|B) = \frac{P(X \leq x, B)}{P(B)}$$

- zvezni primer:

- Postavimo  $B = (y < Y \leq y + h)$  za  $h > 0$  in zahtevajmo  $P(B) > 0$

$$F_X(x|B) = P(X \leq x|B) = \frac{P(X \leq x, y < Y \leq y + h)}{P(y \leq Y < y + h)} = \frac{F(x, y + h) - F(x, y)}{F_Y(y + h) - F_Y(y)}$$

- pogojna porazdelitvena funkcija v obeh primerih

- diskretni primer:

$$F_X(x|y_k) = F_X(x|Y = y_k) = P(X \leq x|Y = y_k) = \frac{P(X \leq x, Y = y_k)}{P(Y = y_k)} = \frac{1}{q_k} \sum_{x_i \leq x} p_{ik}$$

- zvezni primer:

- ce obstaja limita za ( $h \rightarrow 0$ )

$$F_X(x|y) = F_X(x|Y = y) = \lim_{h \rightarrow 0} \frac{F(x, y + h) - F(x, y)}{F_Y(y + h) - F_Y(y)}$$

- imenujemo jo pogojna porazdelitvena funkcija slučajne spremenljivke X glede na dogodek ( $Y=y$ )

- izpeljava pogojne verjetnostne funkcija v diskretnem primeru

- Vpeljimo pogojno verjetnostno funkcijo  $p_{i|k} = \frac{p_{ik}}{q_k}$ . Tedaj je  $F_X(x|y_k) = \sum_{x_i \leq x} p_{i|k}$

- izpeljava pogojne gostote v zveznem primeru

- Naj bosta gostoti  $p(x, y)$  in  $p_Y(y) > 0$  zvezni. Tedaj je:

$$F_X(x|y) = \lim_{h \rightarrow 0} \frac{\frac{F(x, y+h) - F(x, y)}{h}}{\frac{F_Y(y+h) - F_Y(y)}{h}} = \frac{\frac{\partial F}{\partial y}(x, y)}{F'_Y(y)} = \frac{1}{p_Y(y)} \int_{-\infty}^x p(u, y) du$$

- primer za zvezni slučajni vektor

## 9. Momenti in kvantili

- Momenti pokazuje lastnosti vzorca - povprečno vrednost, razprsitev, asimetrijo in sploščenost.

- katere momente poznaš

- **Zacetne** (merimo od 0)

- Consider the following dataset [12 14 14 17 18]
- Consider alternate dataset

1515151515

(enak prvi moment)

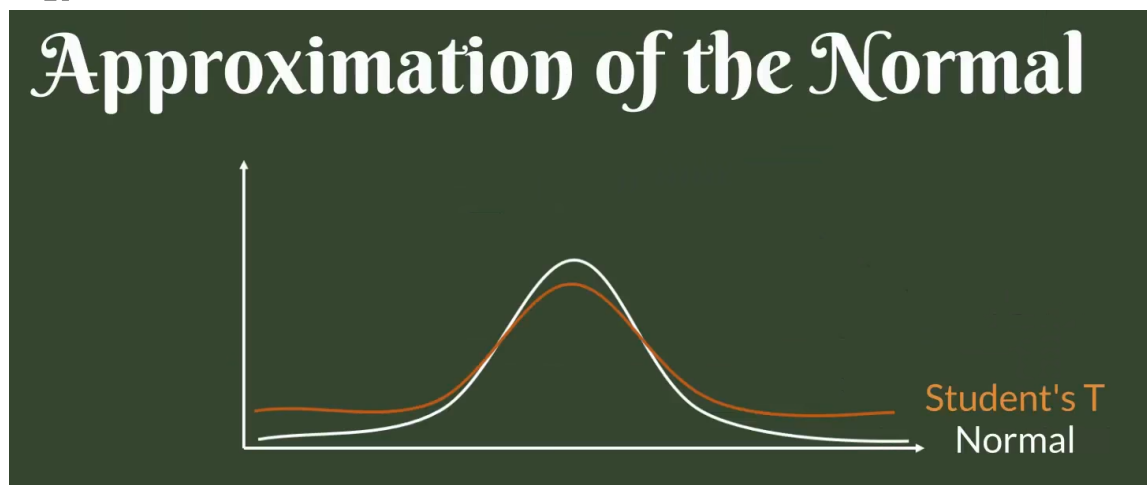
- Each item represents distance from 0
- average distance from zero:  $\frac{\sum x_i}{n} \rightarrow$  **prvi moment** =  $\mu_1 = 15$  predstavlja **povprečje**
- average square distance from zero:  $\frac{\sum x_i^2}{n} \rightarrow$  **drugi moment** =  $\mu_2 = 229.8$  (alternate data set ima  $\mu_2 = 225$ )
- $\frac{\sum x_i^3}{n}$  **tretji moment**

- $\frac{\sum x_i^4}{n}$  **četrti moment**
- **Centralne** (merimo od sredine)
  - average square distance from povprečna vrednost  $\frac{\sum (x_i - \mu_1)^2}{n} \rightarrow$  **centriran drugi moment**  
predstavlja varianco  $\sigma^2$
  - $\frac{\sum (x_i - \mu_1)^3}{n} \rightarrow$  **centriran tretji moment** (asimetrija skewness)
  - $\frac{\sum (x_i - \mu_1)^4}{n} \rightarrow$  **centriran četrti moment** (splosčenost, kurtosis)
- **Standardizirane**
  - $\frac{1}{n} \frac{\sum (x - \mu)^3}{\sigma^3}$  (standardiziran tretji moment)
  - $\frac{1}{n} \frac{\sum (x - \mu)^4}{\sigma^4}$  (standardiziran četrti moment)
- definicija momenta reda  $k$  glede na točko  $a$ 
  - Momenti so posplošitev pričakovane vrednosti in disperzije.
  - Momenti reda  $k \in \mathbb{N}$  glede na točko  $a \in \mathbb{R}$  imenujemo kolicino  $m_k(a) = E((X - a)^k)$
  - Moment obstaja, če obstaja pričakovana vrednost, ki ni neskončna  $E(|X - a|^k) < \infty$
  - Za  $a = 0$  dobimo začetni moment  $z_k = m_k(0)$
  - Za  $a = E(x)$  dobimo centralni moment  $m_k = m_k(E(X))$
- začetni moment, centralni moment
  - začetni moment ko je  $a = 0$
  - centralni moment ko je  $a =$  povprečna vrednost
- lastnosti
  - Če obstaja centralni moment reda  $n$ , potem obstaja vsi momenti reda  $k$ ,  $k \leq n$
  - Če obstaja začetni moment reda  $n$ , potem obstajajo tudi centralni momenti reda  $n$  za  $\forall a \in \mathbb{R}$
  - Če sta  $X$  in  $Y$  neodvisni velja:  $m_3(X + Y) = m_3(X) + m_3(Y)$
- definicija kvantila
  - Kadar spremenljivka nima momentov uporabimo kvantile.
  - Kvantili so "linije" ki razdelijo podatke v skupine enake velikosti
  - Mediana je drugi kvantil (ker polovica podatkov manše polovica večje)
  - Percentili so kvantili ki delijo podatke v 100 skupine enake velikosti
- povezava z inverzom porazdelitvene funkcije
  - TODO

## 10. Studentova porazdelitev t-test

- why (<https://www.youtube.com/watch?v=32CuxWdOlow>)
- Primer1
  - Povprečna teža 20 studentov je bila 165lbs z vzorcnim standardnim odklonom 4.5. Konstruiraj 95% interval zaupanja za populacijsko povprečje.
  - $\bar{x} = 165, n = 20, s = 4.5$
  - Ker **nimamo standardnega odklona populacije**  $\sigma$ , ne moremo izračunati normalne porazdelitve, ampak uporabimo studentovo
    - Poleg tega imamo tudi  $n \leq 20$  (za normalno rabimo  $n \geq 30$ )
  - $\mu \rightarrow \bar{x} \pm t_{n-1, \alpha/2}$ 
    - $n - 1 =$  degrees of freedom = 19

- $\frac{\alpha}{2} = \frac{0.05}{2} = 0.025$
- $t_{19,0.025} = \text{pogledamo v tabeli} = 2.093$
- $\mu \rightarrow 165 \pm 2.093\left(\frac{4.5}{\sqrt{20}}\right) = 165 \pm 2.106$
- Dobili smo 95% interval zaupanja  $I_\mu = [162.894, 167.106]$ 
  - z 95% lahko recemo da je povprečna teža v tem intervalu
- Kako lahko pridemo do te porazdelitve?
  - Uporablja se kadar imamo na voljo majhno število podatkov, in je približek normalne porazdelitve
  - Pridemo preko naslednjega obrazca:
- $$t_{n-1,\alpha} = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$
- Podoben normalni z "debelimi kraki" (večja dispersija - manj podatkov - manjša zanesljivost)
- za  $n > 30$  je že skor enaka z-statistiki (normalni porazdelitvi)
- $Z_X = \mathbb{R}$



- gostota verjetnosti

$$p(x) = \frac{\left(1 + \frac{x^2}{n-1}\right)^{-\frac{n}{2}}}{\sqrt{n-1} B\left(\frac{n-1}{2}, \frac{1}{2}\right)}$$

- definicija beta funkcije

$$B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)} = B(y, x)$$

- posebne vrednosti beta funkcije

$$B\left(\frac{1}{2}, \frac{1}{2}\right) = \pi$$

$$B(n_1, n_2) = \frac{(m-1)!(n-1)!}{(m+n-1)!} \text{ za } n_1, n_2 \in \mathbb{N}$$

- katero porazdelitev dobimo za eno prostostno stopnjo in katero, če je število prostostnih stopenj dovolj veliko
  - **ena prostostna stopnja:** Cauchyeva porazdelitev

- dovolj veliko število prostostnih stopenj: Normalna porazdelitev

## 11. Fisherjeva porazdelitev f-test

- kako lahko pridemo do te porazdelitve primer

Recimo da imamo dve populaciji. Populacijo1 bomo primerjali z Populacijo2. Recimo da naredimo **IQ** test na obeh populacijah. Na vzorcih iz obeh populacij izračunamo vzorčno povprečje ter (vzorčni odklon)  $s_1^2$  ter  $s_2^2$  - Primerjali bomo dva vzorčna odklona - **Testna statistika** (Fisherjeva) je  $F = \frac{S_1^2}{S_2^2} - s_1^2$  dobimo iz  $n$  vzorcev 1. populacije -  $s_2^2$  dobimo iz  $m$  vzorcev 2. populacije - F-distribution, parametri -  $F(n, m) = \frac{\chi^2(n)/n}{\chi^2(m)/m}$  -  $n$  prostostne stopnje prve  $\chi^2$  spremenljivke -  $m$  prostostne stopnje druge  $\chi^2$  spremenljivke -  $Z_F = \mathbb{R}^+$  - oblika odvisna od prostostnih stopenj obeh vzorcev - definicija - verjetnostna funkcija

$$p(x) = \frac{m^{\frac{3}{2}} n^{\frac{n}{2}}}{B(\frac{m}{2}, \frac{n}{2})} \cdot \frac{x^{\frac{(m-2)}{2}}}{(n + mx)^{\frac{m+n}{2}}}$$

- beta  
funkcije

$$B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x + y)} = B(y, x)$$

$$B(\frac{1}{2}, \frac{1}{2}) = \pi$$

- Primer statistike, ki se porazdeljuje po Fisherjevo - F-test za hipoteze o enakosti varianc v dveh normalno porazdeljenih statističnih populacijah in v regresijski analizi. - se kaksne

lastnosti fisherjeve porazdelitve - Za  $U \sim F(m, n)$  je  $\frac{1}{U} \sim F(n, m)$  - Za  $U \sim t(n)$  je  $U^2 \sim F(1, n)$

Statisticni parameter	enacba
Pričakovana vrednost $\mu$	$\frac{m}{m-2}, m > 2$
Modus	$\frac{n-2}{n} \cdot \frac{m}{m+2}$
varianca	$\frac{2m^2 \cdot (n+m-2)}{n(m-2)^2(m-4)}$