Lastnosti in računanje verjetnosti

A...$dogodek$
$\bar{A}$ ... $naspritni\ dogodek$
$A \cap B$ ... $presek\ dogodkov$
$A \cup B$...$unija\ dogodkov$
N...$nemogoč\ dogodek$
G...$gotov\ dogodek$
$P(A)$ ... $verjetnost\ dogodka\ (realno\ število\ med\ 0\ in\ 1)$
$P(N) = 0\ ; P(G) = 1$

IZREK

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$
$P(\bar{A}) = 1 - P(A)$

DEFINICIJA

Dogodka A in B sta **nezdružljiva**, če $A \cap B = N\ ; P(A \cap B) = 0$

TRDITEV

$\overline{(A \cap B)} = \bar{A} \cup \bar{B}$
$\overline{(A \cap B)} = \bar{A} \cup \bar{B}$

Pogojna verjetnost

$P(A|B)$ ... verjetnost, da se zgodi dogodek A, pri pogoju, da se zgodi dogodek B.
Pri uporabi tega zapisa zahtevamo, da je $P(B) > 0$.

IZREK

$P(A|B) = \frac{P(A \cap B)}{P(B)}$ $, to\ lahko\ vzamemo\ za\ def.če\ je\ P(B) > 0$
$P(A \cap B) = P(B) * P(A|B), če\ je\ P(B) > 0$
$P(A \cap B) = P(A) * P(B|A), če\ je\ P(A) > 0$

Neodvisni dogodki

A in B sta **neodvisna** ntk.

1. $P(A) = P(A|B)$
2. $P(B) = P(B|A)$

IZREK

Dogodka A in B sta neodvisna ntk.

$$P(A \cap B) = P(A) * P(B)$$

- To lahko vzamemo tudi za **definicijo** neodivnosti
- Ta formula dela tudi če $P(A) = 0$ ali $P(B) = 0$

OPAZKA

- Če je $B = N$ (ali $A = N$), potem zgornji izrek velja.
  To pomeni, da je N neodvisen od vseh dogodkov.
- Če je $B = G$ (ali $A = G$), potem zgornji izrek velja.
  To pomeni, da je G neodvisen od vseh dogodkov.

DOKAZ

Dvofazni poskus in Bayesov obrazec

Popoln sistem dogodkov je množica dogodkov: B_1, B_2, B_3, … , B_k

- $B_1 \cup B_2 \cup \ldots \cup B_k = G$
- $B_i \cap B_j = N \ (i \neq j)$
- $P(B_i) > 0 \ (za\ vse\ i)$

1. faza : Hipoteze H_1, H_2, … , H_k , ki sestavljajo popoln sistem dogodkov
   Poznamo $P(H_i)$ za vse i

2. faza: Opazujemo dogodek A
   Poznamo $P(A|H_i)$ za vse i
   Iščemo P(A)

FORMULA O POLNI VERJETNOSTI

$$P(A) = P(A|H_1) * P(H_1) + P(A|H_2) * P(H_2) + \cdots + P(A|H_k) * P(H_k)$$

BAYESOV OBRAZEC

$$P(H_i|A) = \frac{P(A|H_i) * P(H_i)}{P(A)}$$

<u>Zgled</u>: Kolikšna je verjetnost, da smo v 1. Metu vrgli enico, če smo v drugem metu vrgli več kot v prvem?

Zaporedja neodvisnih poskusov

- Veliko krat ponovimo isti poskus, ki lahko uspe ali ne uspe (zgodi se dogodek ali pa ne)
- V vsakem poskusu se dogodek A zgodi z isto verjetnostjo p
- V vsakem poskusu se dogodek A zgodi neodvisno od izidov/rezultatov prejšnjih/vseh poskusov

BERNULLIJEV OBRAZEC

$$P(n, p, k) = \binom{n}{k} * p^k * (1-p)^{n-k}$$

n … n-krat ponovimo isti poskus
p … dogodek A se v poskusu zgodi z verjetnostjo p
k … verjetnost, da se v n poskusih dogodek A zgodi k-krat

Zgled: $P(10000, \frac{1}{2}, 5000)$ – verjetnost, da v 10000 metih kovanca, grb pade 5000 krat

Slučajne spremenljivke

So funkcije, katerih rezultat ni odvisen od podatkov, temveč od slučaja. Pravimo, da je rezultat slučajne spremenljivke realno število.

Opis slučajne spremenljivke X:

a) Zalogo vrednosti
b) Za vsak X_k iz zaloge vrednosti, naj bo p_a verjetnost, da slučajna spremenljivka X zavzame vrednost x_k

VERJETNOSTNA SHEMA

$$X \sim \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ p_1 & p_2 & \dots & p_n \end{pmatrix}$$

$$p_i \geq 0\,; \sum_{j=1}^{n} p_i = 1$$

Matematično upanje

Matematično upanje oz. pričakovana vrednost označujemo z $E(X)$. Pove nam kakšna je vrednost spremenljivke X v povprečju.

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

Disperzija

Disperzija slučajne spremenljivke X, $D(X)$, meri, kako »niha« vrednost X okoli »povprečja« $E(X)$.

$$D(X) = E(X^2) - E(X)^2$$

Standardni odklon

Standardni odklon, označen z $\sigma(X)$, meri povprečno odstopanje od povprečne vrednosti in ima iste enote kot slučajna spremenljivka.

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

IZREK

Če je $X \sim b(n, p)$, potem je:

$$E(X) = n * p$$
$$q = 1 - p$$
$$D(X) = n * p * (1 - p) = n * p * q$$
$$\sigma(X) = \sqrt{n * p * (1 - p)} = \sqrt{n * p * q}$$

Neenakost Markova

IZREK

Naj bo slučajna spremenljivka $X \geq 0$. Potem je:

$$P[X \geq a] \leq \frac{E(x)}{a}$$

Zgled: Ocenite verjetnost, da pri metu dveh kock pade vsaj 10!

Neenakost Čebiševa

Neenačkba Čebiševa ocenjuje kakšna je verjetnost, da se slučajna spremenljivka veliko razlikuje od matematičnega upanja.

IZREK

X slučajna spremenljivka

$$P[|X - E(X)| \geq t] \leq \frac{D(X)}{t^2} \; ; t > 0$$

Zgled: Kovanec vržemo 1000 krat. Oceni verjetnost da, da bo število grbov med 400 in 600.

Lastnosti matematičnega upanja in disperzije

TRDITEV

Če je neka konstanta in $P[X = a] = 1$, potem je $E(X) = a, D(X) = 0$

IZREK

Matematično upanje je linearno, če sta X in Y slučajni spremenljivki in $a, b \in \mathbb{R}$, je

$$E(aX + bY) = a * E(X) + b * E(Y)$$

PROBLEM: Za disperzijo takšna trditev **ne** velja.

Kovarianca slučajnih spremenljivk

Naj bosta X in Y slučajni spremenljivki. **Kovarianca** slučajnih spremenljivk X in Y je količina, definirana z:

$$K(X, Y) := E\Big(\big(X - E(X)\big) * \big(Y - E(Y)\big)\Big) = E(X * Y) - E(Y) * E(X)$$

Komentar: $K(X, X) = D(X) = \sigma(X)^2$

Velja:

$$|K(X,Y)| \leq \sigma(X) * \sigma(Y)$$
$$r(X,Y) := \frac{K(X,Y)}{\sigma(X) * \sigma(Y)}$$

r … korelacijski koeficient slučajnih spremenljivk X in Y

IZREK:

Za slučajni spremenljivki X in Y velja

$$D(X+Y) = D(X) + D(Y) + 2 * K(X,Y)$$

Komentar: Če sta X in Y neodvisni slučajni spremenljivki, potem sta tudi nekorelirani! $K(X,Y) = 0\,, r(X,Y) = 0$
Obratno ni nujno res. Slučajni spremenljivki X in Y sta lahko nekolerirani in nista neodvisni.

IZREK:

Za neodvisni (dovolj nekorelirani) slučajni spremenljivki X in Y velja:

$$D(X+Y) = D(X) + D(Y)$$
$$D(a * X) = a^2 * D(X)\,; a \in \mathbb{R}$$
$$\sigma(a * X) = a * \sigma(X)\,; a > 0$$

Slučajne spremenljivke z neskončno mnogo vrednostmi

$$E(X) = \sum_{i=1}^{\infty} x_i * p_i$$
$$D(X) = E(X^2) - E(X)^2 = \sum_{i=1}^{\infty} {x_i}^2 * p_i - E(X)^2$$

Če vrsta ne konvergira matematično upanje te spremenljivke ne obstaja . Lahko se zgodi da disperzija ne obstaja, upanje pa. Obratno ni mogoče.

POTENČNA VRSTA:

$$E(X) = \sum_{i=1}^{\infty} i * \frac{1}{2^i}$$
$$D(X) = \sum_{i=1}^{\infty} i^2 p_i - E(X)$$

<u>Zgledi:</u> Mečemo kovanec ; X … zaporedno številko meta v katerem prvič pade grb
Igralnica: stavimo 1€ na rdeče ; dobimo ali izgubimo ; če izgubimo stavimo 2€ na rdeče

Zvezno porazdeljene slučajne spremenljivke

X je opisana z gostoto verjetnosti $g_x(x)$

- $g_x(X)$ je funkcija : $\mathbb{R} \rightarrow \mathbb{R}$
- $g_x(X) \geq 0$
- $\int_{-\infty}^{\infty} g_x(X)\,dx = 1$
- Odsekana zvezna (zvezna povsod, razen morda v nekaj točkah)

Računanje verjetnosti:

$$P[a < X < b] = \int_a^b g_x(X)\,dx \;; a, b \in \mathbb{R} \,; a < b$$

$$P[X = a] = P[X = b] = 0$$

<u>Matematično upanje:</u>

$$E(X) = \int_{-\infty}^{\infty} x * g_x(X) * dx$$

<u>Disperzija:</u>

$$D(X) = \int_{-\infty}^{\infty} x^2 * g_x(X) * dx - E(X)^2$$

Enakomerna zvezna slučajna spremenljivka na [a,b]

$$E(X) = \frac{a+b}{2}$$

$$D(X) = \frac{1}{12}(b-a)^2$$

$$\sigma(X) = \frac{1}{\sqrt{12}}(b-a)$$

Normalna porazdelitev

Normalna porazdelitev je »limita« binomske porazdelitve b(n,p).

$E(X) = n * p =: a$

$D(X) = n * p * (1-p)$

$\sigma(X) = \sqrt{n * p * (1-p)} =: \sigma$

Je t.i. normalna porazdelitev $N(a, \sigma)$ porazdeljena s porazdelitveno gostoto:

$$g(X) = \frac{1}{\sigma\sqrt{2\pi}} * e^{-\frac{(X-a)^2}{2\sigma^2}}$$

Standardna normalna slučajna spremenljivka

$Z \sim N(0,1)$ je standardna (standardizirana) normalna slučajna spremenljivka. Njena porazdelitvena gostota je enaka:

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

IZREK:

Naj bo $X \sim N(a, \sigma)$. Potem je slučajna spremenljivka $Z := \frac{X-a}{\sigma}$ porazdeljena standardno normalno $Z \sim N(0,1)$.

Nauk:

Če vemo, da je slučajna spremenljivka porazdeljena normalno $[X \sim N(a, \sigma)]$, lahko problem verjetnosti dogodkov v zvezi z X prevedemo na računanje verjetnosti v zvezi **s standardno normalno porazdeljeno slučajno spremenljivko.**

Centralno limitni izrek

Centralni limitni izrek pravi, da vsako zaporedje enako porazdeljenih neodvisnih slučajnih spremenljivk z istim matematičnim upanjem in disperzijo teži k normalni porazdelitvi.

Če vzamemo naključni vzorec n opazovanj iz katerekoli populacije dobimo (n > 30) distribucijo povprečij katera je porazdeljena približno normalno s povprečjem enakim povprečju populacije in standardnim odklonom enakim $\frac{1}{\sqrt{n}}$ standardnega odklona celotne populacije.

Statistikom omogoča aproksimacijo zaporedja podatkov z neznanimi porazdelitvami kot porazdeljene.

X_1, X_2, X_3, … zaporedje **neodvisnih** slučajnih spremenljivk, ki imajo vse isto:

- Matematično upanje $E(X_i) = a$
- Disperzijo $D(X_i) = \sigma^2$

Opazujemo S_1, S_2, S_3, …
$S_n = X_1+X_2+X_3+\ldots+X_n$ ($N \geq 30$ če želimo uporabiti CLI)

<u>Vemo:</u>

- $E(S_n) = n * a$
- $D(S_n) = n * \sigma^2$

Za zaporedje S_n velja:

$$\lim_{n\to\infty} P\left(\frac{S_n - E(S_n)}{\sigma(S_n)} < X\right) = \int_{-\infty}^{X} \frac{1}{\sqrt{2\pi}} * e^{-\frac{t^2}{2}} dt \;; rešitev: \; \leq 1$$

<u>**BERI:**</u>

Za velike n je

$$\frac{S_n - E(S_n)}{\sigma(S_n)} \sim N(0,1)$$
$$S_n \sim N(n * a, \sqrt{n} * \sigma)$$

CLI velja tudi če:

X1,X2,X3, … nimajo istega matematičnega upanja in iste disperzije morajo pa biti še vedno neodvisne in »primerljive« po velikost

$$S_n = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$$

$$S_n = N(a, \frac{\sigma}{\sqrt{n}})$$

Zgledi:

- Kovanec vržemo 1000 krat. Oceni verjetnost, da grb pade med 400 in 600 krat.