Osnove verjetnosti in statistika

Gašper Fijavž

Fakulteta za računalništvo in informatiko
Univerza v Ljubljani

Ljubljana, 19. marec 2010

Slučajne spremenljivke

Slučajna spremenljivka je funkcija, katere unkcijska vrednost ni odvisna od vhodnih podatkov/parametrov, temveč od slučaja.

Privzeli bomo da slučajne spremenljivke za funkcijske vrednosti *vračajo* realna števila.

Zgledi:

- met poštene igralne kocke, rezultat je število iz $\{1, 2, 3, 4, 5, 6\}$.
- vsota pik pri metu dveh kock, dobimo $\{2, 3, 4, \ldots, 12\}$.

Slučajne spremenljivke

Slučajno spremenljivko X popolnoma opišemo z

- *zalogo vrednosti $\mathcal{Z}_X$* in
- za vsak $x_k \in \mathcal{Z}_X$ moramo poznati $p_k = P(X = x_k)$.

$$X \sim \left(\begin{smallmatrix} x_1 & x_2 & x_3 & \dots & x_n \\ p_1 & p_2 & p_3 & \dots & p_n \end{smallmatrix}\right)$$

Zgornjemu opisu slučajne spremenljivke X pravimo *verjetnostna shema*.

Pri tem je $p_i \geq 0$, za vse $i = 1, \dots, n$, in velja $\sum_{i=1}^{n} p_i = 1$.

Slučajne spremenljivke

Dve pomembni porazdelitvi:

- *Enakomerna porazdelitev* je slučajna spremenljivka, pri kateri imajo vsi izidi enako verjetnosti.

$$X \sim \begin{pmatrix} x_1 & x_2 & x_3 & \dots & x_n \\ \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \end{pmatrix}$$

- *Binomska porazdelitev* $b(n, p)$

$$X \sim \begin{pmatrix} 0 & 1 & 2 & \dots & n \\ P(n;p;0) & P(n;p;1) & P(n;p;2) & \dots & P(n;p;n) \end{pmatrix}$$

$$P(n; p; k) = P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

Matematično upanje

$X \sim \left(\begin{smallmatrix} x_1 & x_2 & x_3 & \cdots & x_n \\ p_1 & p_2 & p_3 & \cdots & p_n \end{smallmatrix}\right)$

Matematično upanje slučajne spremenljivke X, $E(X)$, definiramo kot

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

Matematično upanje

Zgledi:

- Kakšno je matematično upanje za število pik na kocki?
- V klobuku sta dva kovanca, za 2 in 5 centov. Slučajno potegnemo enega. Kakšno je matematično upanje za vrednost denarja v roki?
- Na loteriji je na voljo 1000 srečk po 10€. Organizator obljublja 100 dobitkov po 5€ , 20 dobitkov po 20€, 2 dobitka za 500€ in glavni dobitek 5000€.
  Se nam splača kupiti srečko?

Matematično upanje

- Štirikrat vržemo kovanec. Kolikšno je matematično upanje za število grbov?
- Kakšno je matematično upanje spremenljivke $x \sim b(n, p)$?
- Slepar meče tri kocke. Stava na posamezno številko k je vredna 1€. Po metu lahko vložek izgubimo lahko pa zaslužimo toliko €, kolikor številk k je padlo na treh kockah. Se nam igra splača?

Disperzija

Disperzija slučajne spremenljivke X, $D(X)$, meri kako slučajna spremenljivka X odstopa od matematičnega upanja.
$X \sim \left(\begin{smallmatrix} x_1 & x_2 & x_3 & \dots & x_n \\ p_1 & p_2 & p_3 & \dots & p_n \end{smallmatrix}\right)$

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

Izrek

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

Standardni odklon

Standardni odklon (tudi *standardna deviacija*) slučajne spremenljivke X, $\sigma(X)$, definiramo kot

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

Izrek

Naj bo slučajna spremenljivka x porazdeljena z binomsko porazdelitvijo $b(n, p)$. Potem je

$$\begin{aligned} E(X) &= n \cdot p \\ D(X) &= n \cdot p \cdot (1-p) \\ \sigma(X) &= \sqrt{n \cdot p \cdot (1-p)} \end{aligned}$$

Zgledi

Opiši slučajno spremenljivko in izračunaj matematično upanje, disperzijo, standardni odklon.

- Met kovanca, pri čemer uporabljaš *indikatorsko slučajno spremenljivko* za grb.
- Met kocke.
- Iz klobuka izvlečemo en kovanec, pri čemer
  (a) so v klobuku trije kovanci po 2 centa.
  (b) je v klobuku en kovanec za 10 cntov in 8 kovancev za 1 cent.

Zgledi

- Podjetje ima 5 strojev, ki se neodvisno en od drugega kvarijo z verjetnostjo 0,1.
  Slučajna spremenljivka X meri število pokvarjenih strojev.
  Seviser vsak dan pregleda stroje po vrsti. Če naleti na pokvarjen stroj ga nemudoma začne popravljati in ne konča do koca delovnega dne. Slučajna spremenljivka Y naj označuje zaporedno številko zadnjega pregledanega stroja.
- V igralnici prvič stavimo 10€, zadanemo z verjetnostjo $1/2$. Stavimo *vse* dokler (a) ne izgubimo ali (b) ne dobimo trikrat zapored.
  Slučajna spremenljivka X označuje naš izplen po koncu igre, slučajna spremenljivka Y pa število stav.
- Kocko mečemo dokler ne pade šestica. X označuje število metov.

Neenakost Markova

Izrek

Naj velja $X \geq 0$. Za vsak $a > 0$ velja neenakost

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

Zgled: Oceni verjetnost, da pri metu dveh kock pade vsaj 10 pik.

Neenakost Čebiševa

Izrek

Naj bo X slučajna spremenljivka. Za vsak $t > 0$ velja neenakost

$$P(|X - E(X)| \geq t) \ \leq \ \frac{D(X)}{t^2}$$

Uporaba neenakosti Čebiševa

Nalogi:

- 1000-krat vržemo kovanec. Oceni vejetnost, da bo
  - število grbov med 400 in 600.
- Kocko vržemo 120-krat. Oceni verjetnost, da pade med 15 in 25 šestic.

Lastnosti upanja in disperzije

Izrek

Če je $a \in \mathbb{R}$ in $P(X = a) = 1$, potem je $E(X) = a$ in $D(X) = 0$.

Izrek

Matematično upanje je linearno. Naj bosta X in Y slučajni spremenljivki, $E(X)$ in $E(Y)$ njuni matematični upanji in a, b poljubni realni števili. Potem je

$$E(aX + bY) = aE(X) + bE(Y)$$

Lastnosti upanja in disperzije

- Kocko mečemo po stekleni mizi. X ... število pik na zgornji ploskvi, Y ... število pik na spodnji ploskvi.
- Iz *vreče* izvlečemo slučajnega Slovenca ter izmerimo njegovo višino X in težo Y.

Kovarianca

Naj bosta X in Y slučajni spremenljivki. Količini

$$K(X,Y) = E((X - E(X))(Y - E(Y)))$$

pravimo *kovarianca slučajnih spremenljivk* X in Y.
Velja

$$K(X,Y) = E(XY) - E(X)E(Y)$$

in

$$|K(X,Y)| \leq \sigma(X)\sigma(Y)$$

Korelacija

Naj bosta X in Y slučajni spremenljivki. Izraz

$$r(X,Y) = \frac{K(X,Y)}{\sigma(X)\sigma(Y)}$$

je *korelacijski koeficient slučajnih spremenljivk* X in Y.

Velja $-1 \leq r(X,Y) \leq 1$.

Kdaj je in kaj pomeni:

- $r(X,Y) = 1$
- $r(X,Y) = -1$
- $r(X,Y) = 0$
- $r(X,Y) > 0$
- $r(X,Y) < 0$

Lastnosti disperzije

Izrek

Za slučajni spremenljivki X in Y velja

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

Izrek

Za neodvisni slučajni spremenljivki X in Y velja:

- $D(X+Y) = D(X) + D(Y)$
- $D(a \cdot X) = a^2 D(X)$ *in*
- $\sigma(a \cdot X) = \sqrt{D(a \cdot X)} = \sqrt{a^2 D(X)} = |a|\sigma(X)$

Lastnosti disperzije

Zgled: $X \sim b(n,p)$

Lastnosti disperzije

Posledica

Slučajne spremenljivke $X_1, X_2, \ldots, X_n$ naj bodo enako porazdejene, neodvisne, z matematičnim upanjem $E(X_i) = a$ in disperzijo $D(X) = \sigma^2$. Za slučajno spremenljivko

$$Y = \frac{1}{n}(X_1 + X_2 + \ldots + X_n)$$

velja

$$E(Y) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n} \cdot n \cdot a = a$$

$$D(Y) = \frac{1}{n^2}\sum_{i=1}^{n} D(X_i) = \frac{1}{n^n} \cdot n \cdot \sigma^2 = \frac{\sigma^2}{n}$$

Slučajne spremenljivke z neskončno zalogo

Zgledi:

- Mečemo kovanec, dokler ne pade grb, X je število metov.
- Igralnica, stavimo 1€ na rdeče. Če izgubimo, stavimo $2\times$ toliko kot v prejšnjem krogu. Dokler ne dobimo.
  Koliko je naš pričakovani dobitek, kaj pa disperzija?

Zvezno porazdeljene slučajne spremenljivke

Iščemo slučajno spremenljivko X, z zalogo vrednosti $[0, 1]$, ki ima naslednjo lastnost: če je $J \subseteq [0, 1]$ interval, potem je $P(X \in J)$ enaka dolžini intervala J.
Kako opisati X?

Gostota verjetnosti

Funkcija $g : \mathbb{R} \to \mathbb{R}$ je *gostota verjetnosti* neke slučajen spremenljivke, če

(G1) $g(x) \geq 0$ za vse $x \in \mathbb{R}$,

(G2) $\displaystyle\int_{-\infty}^{\infty} g(x)dx = 1$ in

(G3) $g(x)$ je odsekoma zvezna funkcija.

Gostota verjetnosti

Zgled:

$$g(x) = \begin{cases} 0 & x < 0 \\ 2x & 0 \leq x \leq 1 \\ 0 & x > 1 \end{cases}$$

Preveri, da g določa gostoto verjetnosti slučajne spremenljivke X in izračunaj $P(1/3 < X \leq 2/3)$.

Matematično upanje in disperzija

Slučajna spremenljivka X naj bo opisana z gostoto verjetnosti $g_X(x)$.

- $E(X) = \int_{-\infty}^{\infty} t\, g_X(t) dt$
- $D(X) = \int_{-\infty}^{\infty} (t - E(X))^2 g_X(t) dt = \int_{-\infty}^{\infty} t^2\, g_X(t) dt - E(X)^2$

Enakomerna porazdelitev

Zvezna slučajna spremenljivka X je razporejena *enakomerno* na intervalu $[a, b]$, $X \sim \mathcal{U}[a, b]$, če je njena gostota verjetnosti g_X enaka

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