Uporabna statistika

Gregor Dolinar

Fakulteta za elektrotehniko
Univerza v Ljubljani

2. oktober 2013

Slučajni poskus

Poskus:

- meritev
- matematični, fizikalni model
- analiza

Vedno je prisotna slučajnost

Vzorčni prostor

Definicija

Vzorčni prostor S je množica vseh možnih izidov slučajnega poskusa.

Primer. V podjetju izdelujejo sestavni del električne naprave. Zanima nas debelina sestavnega dela:

- $S = \mathbb{R}^+$
- $S = \{x : 10 < x < 11\}$
- $S = \{\text{pretanek}, \text{dober}, \text{predebel}\}$
- $S = \{\text{dober}, \text{slab}\}$

Definicija

Vzorčni prostor slučajnega poskusa je:

- diskreten, če je vseh možnih izidov končno ali kvečjemu Števno neskončno.
- zvezen, če lahko vse možne izide predstavimo kot točke "unije intervalov"(interval je lahko končen ali neskončen)

Dogodek

Definicija

Dogodek je podmnožica vzorčnega prostora slučajnega poskusa.

Dogodek E vzorčnega prostora S se zgodi, če je rezultat slučajnega poskusa eden izmed možnih izidov iz množice E.

Primer. Izberemo 2 sestavna dela in jima izmerimo debelino.
Zanima nas, ali sta ustrezna.
Vzorčni prostor $S_1 = \{yy, yn, ny, nn\}$.

- $E_1 = \{yy, yn, ny\} \subset S_1$
  E_1 je dogodek, da je vsaj en izmed izbranih delov ustrezen.
- $E_2 = \{yn, ny\} \subset S_1$
  E_2 je dogodek, da je natanko en izmed izbranih delov ustrezen.

Vzorčni prostor $S_2 = \{0, 1, 2\}$.

- $E_1 = \{0, 1\} \subset S_2$
  E_1 je dogodek, da je vsaj en izmed izbranih delov ustrezen.
- $E_2 = \{1\} \subset S_2$
  E_2 je dogodek, da je natanko en izmed izbranih delov ustrezen.

- $E = S$, gotov dogodek
- $E = \emptyset$, nemogoč dogodek
- $|E| = 1$, elementarni dogodek
- $E_1 \subset E_2$, E_1 je način dogodka E_2

Operacije na dogodkih

- Unija dogodkov: $E_1 \cup E_2$
  $E_1 \cup E_2$ je množica vseh možnih izidov dogodkov E_1 in E_2
- Presek dogodkov: $E_1 \cap E_2$
  $E_1 \cap E_2$ je množica vseh tistih možnih izidov, ki so hkrati v E_1 in E_2
- Komplement dogodka (nasprotni dogodek): E_1^C
  E_1^C je množica vseh možnih izidov, ki niso v množici izidov dogodka E_1

Definicija

Dogodka E_1 in E_2 sta nezdružljiva (izključujoča), če je $E_1 \cap E_2 = \emptyset$.

Definicija verjetnosti

- klasična definicija verjetnosti (končno mnogo enako verjetnih izidov)

  $$\text{verjetnost dogodka} = \frac{\text{Število ugodnih izidov}}{\text{Število vseh izidov}}$$

Definicija verjetnosti

- statistična definicija verjetnosti
  N Število ponovitev slučajnega poskusa,
  k Število ponovitev dogodka,
  $\frac{k}{N}$ relativna frekvenca
  Verjetnost dogodka je približno enaka relativni frekvenci za velike N.

Primer

Buffonova igla, mečemo iglo na površino z narisanimi vzporednimi črtami, ki so na enaki razdalji, dolžina igle pa je manjša od razdalje med sosednjima črtama. Verjetnost P, da igla seka črto, je enaka

$$P = \frac{2l}{t\pi},$$

kjer je l dolžina igle in t razdalja med črtama.
Leta 1901 je italijanski matematik Mario Lazzarini vrgel iglo 3408-krat in dobil oceno $\frac{355}{113}$ za konstanto π, kar je zelo velika natančnost

$$355/113 - \pi = 2.7 \cdot 10^{-7}.$$

Lazzarini je "goljufal", saj je hotel dobiti zelo natančno oceno za π s pomočjo ulomka $\frac{355}{133}$. Za dolžino igle je izbral $5/6$ razdalje med črtama, poskus pa je ponavljal v paketih po 213 ($3408 : 213 = 16$) toliko časa, da je dobil ocena za π s pomočjo ulomka $\frac{355}{113}$.
Naj bo n število metov in x število metov, kjer je igla sekala črto. Statistična ocena verjetnosti je potem

$$P \doteq \frac{x}{n},$$

torej je

$$\pi = \frac{2l}{tP} = \frac{2 \cdot 5/6}{P} \doteq \frac{5/3n}{x} = \frac{5/3 \cdot 213 \cdot k}{x} = \frac{355k}{x}.$$

Definicija verjetnosti

- matematična (aksiomatična) definicija verjetnosti

Definicija

Verjetnost je preslikava, ki vsakemu dogodku E vzorčnega prostora S priredi neko število $P(E)$, tako da velja:

- $P(S) = 1$,
- $0 \leq P(E) \leq 1$,
- Za poljubna nezdružljiva dogodka, $E_1 \cap E_2 = \emptyset$, je $P(E_1 \cup E_2) = P(E_1) + P(E_2)$.

Lastnosti:

- $P(\emptyset) = 0$
- $P(E^C) = 1 - P(E)$
- Za $E_1 \subset E_2$ je $P(E_1) \leq P(E_2)$
- $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$

Pogojna verjetnost

Definicija

Pogojna verjetnost dogodka B pri pogoju A je verjetnost

$$P(B|A) = \frac{P(A \cap B)}{P(A)},$$

pri čemer $P(A) \neq 0$.

Primer. Vsi izidi so enako verjetni.

$$P(A) = \frac{\text{možni izidi dogodka } A}{n}$$

$$P(A \cap B) = \frac{\text{možni izidi dogodka } A \cap B}{n}$$

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{\text{možni izidi dogodka } A \cap B}{\text{možni izidi dogodka } A}$$

Neodvisnost dogodkov

Definicija

Dogodka A in B sta neodvisna, če velja

$$P(B|A) = P(B)$$

ali

$$P(A|B) = P(A)$$

Sledi, da sta dogodka A in B neodvisna, če je

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

Definicija

Dogodki $E_1, E_2, \ldots, E_n$ tvorijo popoln sistem dogodkov, če so paroma nezdružljivi ($E_i \cap E_j = \emptyset, i \neq j$), njihova unija pa je gotov dogodek.

Izrek (formula o popolni verjetnosti)

Naj bo $E_1, E_2, \ldots, E_n$ popoln sistem dogodkov. Potem je

$$P(A) = \sum_{i=1}^{n} P(A|E_i)P(E_i).$$

Izrek (Bayesova formula)

Naj bo $E_1, E_2, \ldots, E_n$ popoln sistem dogodkov. Potem je

$$P(E_i|A) = \frac{P(A|E_i)P(E_i)}{\sum_{i=1}^{n} P(A|E_i)P(E_i)}.$$

Primer. Test bolezni.

- Okužen, verjetnost, da je test pozitiven, je 0.99.
- Ni okužen, verjetnost, da je test negativen, je 0.95.
- Delež ljudi v populaciji, ki so okuženi, 0.0002.
  (npr. delež okuženih s HIV v podsaharski Afriki je 0.05, v Sloveniji 0.0002)

Oseba opravi test, test je pozitiven. Kolikšna je verjetnost, da je oseba okužena?

- B dogodek, da oseba okužena
- T_p dogodek, da test pozitiven
- $P(T_p|B) = 0.99$, $P(T_p^C|B^C) = 0.95$, $P(B) = 0.0002$

$$P(B|T_p) = ?$$

Bayesova formula

$$P(B|T_p) = \frac{P(T_p|B)P(B)}{P(T_p|B)P(B) + P(T_p|B^C)P(B^C)}$$

$P(T_p|B^C) = \frac{P(T_p \cap B^C)}{P(B^C)} = ?$
Ker je $P(T_p \cap B^C) + P(T_p^C \cap B^C) = P(B^C)$, je

$$P(T_p|B^C) = \frac{P(B^C) - P(T_p^C \cap B^C)}{P(B^C)} = 1 - P(T_p^C|B^C).$$

$$\begin{aligned} P(B|T_p) &= \frac{0.99 \cdot 0.0002}{0.99 \cdot 0.0002 + (1 - 0.95) \cdot (1 - 0.0002)} \\ &\doteq 0.0039. \end{aligned}$$

Denimo, da je delež ljudi v populaciji, ki so okuženi 0.05. Potem je dobljena pogojna verjetnost ustrezno višja.

Opomba

Izračunana verjetnost velja, če je oseba, ki je delala test, naključno izbrana. Običajno se testirajo osebe iz rizičnih skupin, v katerih je bolezen dosti bolj razširjena. V tem primeru moramo vzeti za populacijo ustrezno skupino in delež okuženih v tej skupini (npr, v ZDA leta 1990 v skupini odvisnikov 18 %, leta 2009 pa 9 % okuženih).