Verjetnostni račun, Bernoulijeva in Gaussova porazdelitev, centralni limitni izrek, izrek velikih števil

I.

Ponovili bomo osnove verjetnostnega računa, ki iz verjetnosti znanih elementarnih dogodkov nekega poskusa, računa verjetnosti sestavljenih dogodkov poskusa, nato opisali pojem porazdelitvenega zakona slučajne spremenljivke ter preučili centralni limitni izrek in zakon o velikih številih.

II.

Nato bomo predstavili osnove opisne oziroma deskriptivne statistike, ki se ukvarja z zbiranjem in opisovanjem podatkov ter nekatere elemente inferenčne oziroma matematične, analitične, sklepne, indukcijske statistike, ki nas uči, kako o celoti sklepamo iz podatkov dobljenih na manjših vzorcih, ki so le del celote.

Kazalo

I Verjetnostni račun

Začetki verjetnosti

Slučajna spremenljivka

Porazdelitveni zakoni slučajne spremenljivke

- **Homogena porazdelitev**
- **Bernoullijeva porazdelitev**
- **Gaussova normalna porazdelitev**

Centralni limitni izrek

Izrek o velikih številih

II Inferenčna statistika

I Verjetnostni račun

Začetki verjetnosti

Naključnost so poznale že stare kulture: Egipčani, Grki, . . . a je niso poskušale razumeti, razlagale so jo kot voljo bogov.

Leta 1662 je plemič Chevalier de Mere zastavil matematiku Blaise Pascalu vprašanje: zakaj določene stave prinašajo dobiček druge pa ne. Le-ta si je o tem začel dopisovati s Fermatom in iz tega so nastali začetki verjetnostnega računa.

Prvo razpravo je napisal že leta 1545 italijanski kockar in matematik Cardano, a ni bila širše znana.

Tudi leta 1662 je Anglež John Graunt sestavil na osnovi podatkov prve zavarovalniške tabele.

Ahil in Ajaks kockata, amfora, okrog 530 pr.n.š, Eksekias, Vatikan

Leta 1713 je Jakob Bernoulli objavil svojo knjigo Umetnost ugibanja s katero je verjetnostni račun postal resna in splošno uporabna veda. Njegov pomen je še utrdil Laplace, ko je pokazal njegov pomen pri analizi astronomskih podatkov (1812). Leta 1865 je avstrijski menih Gregor Mendel uporabil verjetnostno analizo pri razlagi dednosti v genetiki. V 20. stoletju se je uporaba verjetnostnih pristopov razširila skoraj na vsa področja.

Slučajna spremenljivka

Denimo, da imamo poskus, katerega izidi so števila (npr. pri metu kocke so izidi števila pik). Se pravi, da je poskusom prirejena neka količina, ki more imeti različne vrednosti. Torej je spremenljivka. Katero od mogočih vrednosti zavzame v določeni ponovitvi poskusa, je odvisno od slučaja. Zato ji rečemo slučajna spremenljivka.

Da je slučajna spremenljivka znana, je potrebno vedeti

1. kakšne vrednosti more imeti (zaloga vrednosti) in

2. kolikšna je verjetnost vsake izmed možnih vrednosti ali intervala vrednosti.

Predpis, ki določa te verjetnosti, imenujemo porazdelitveni zakon.

Slučajne spremenljivke označujemo z velikimi tiskanimi črkami iz konca abecede, vrednosti spremenljivke pa z enakimi malimi črkami. Tako je npr. $(X = x_i)$ dogodek, da slučajna spremenljivka X zavzame vrednost x_i. Porazdelitveni zakon slučajne spremenljivke X je poznan, če je mogoče za vsako realno število x določiti verjetnost

$F(x) = P(X < x)$

F(x) imenujemo porazdelitvena funkcija. Najpogosteje uporabljamo naslednji vrsti slučajnih spremenljivk:

1. diskretna slučajna spremenljivka, pri kateri je zaloga vrednosti neka števna (diskretna) množica;

2. zvezna slučajna spremenljivka, ki lahko zavzame vsako realno število znotraj določenega intervala.

A. Jurišić in V. Batagelj: Verjetnostni račun in statistika 15

Diskretne slučajne spremenljivke

Zaloga vrednosti diskretne slučajne spremenljivke X je števna množica $\{x_1, x_2, \ldots, x_m, \ldots\}$. Dogodki

$$X = x_k \quad k = 1, 2, \ldots$$

sestavljajo popoln sistem dogodkov. Označimo verjetnost posameznega dogodka s

$$P(X = x_i) = p_i.$$

Vsota verjetnosti vseh dogodkov je enaka 1:

$$p_1 + p_2 + \cdots + p_m + \cdots = 1.$$

Verjetnostna tabela

Verjetnostna tabela prikazuje diskretno slučajno spremenljivko s tabelo tako, da so v prvi vrstici zapisane vse vrednosti x_i, pod njimi pa so pripisane pripadajoče verjetnosti:

$$X : \begin{pmatrix} x_1 & x_2 & \cdots & x_m & \cdots \\ p_1 & p_2 & \cdots & p_m & \cdots \end{pmatrix}.$$

Porazdelitvena funkcija je v tem primeru

$$F(x_k) = P(X < x_k) = \sum_{i=1}^{k-1} p_i.$$

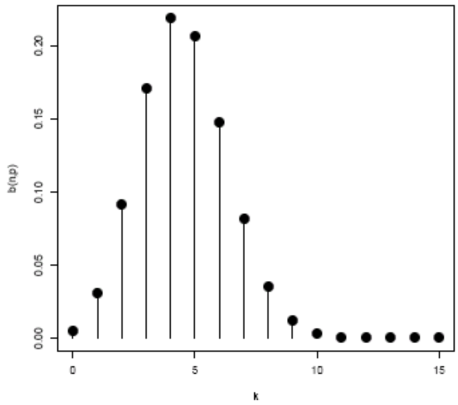
PORAZDELITVENI ZAKONI SLUČAJNE SPREMENLJIVKE

Homogena porazdelitev
Končna diskretna slučajna spremenljivka se porazdeljuje enakomerno, če so vse njene vrednosti enako verjetne. Primer take slučajne spremenljivke je število pik pri metu kocke.

$$X : \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \end{pmatrix}$$

Binomska porazdelitev

Binomska porazdelitev




```
> h <- dbinom(0:15,size=15,prob=0.3)
> plot(0:15,h,type="h",xlab="k",ylab="b(n,p)")
> points(0:15,h,pch=16,cex=2)
```

Binomska porazdelitev ima zalogo vrednosti $\{0, 1, 2, \cdots, n\}$ in verjetnosti, ki jih računamo po Bernoullijevem obrazcu:

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

$k = 0, 1, 2, \cdots, n$. Binomska porazdelitev je natanko določena z dvema podatkoma – parametroma: n in p. Če se slučajna spremenljivka X porazdeljuje binomsko s parametroma n in p, zapišemo:

$$X : B(n, p).$$

Binomska porazdelitev / Primer

Naj bo slučajna spremenljivka X določena s številom fantkov v družini s 4 otroki. Denimo, da je enako verjetno, da se v družini rodi fantek ali deklica:

$$P(F) = p = \frac{1}{2} \quad , \quad P(D) = q = \frac{1}{2}.$$

Spremenljivka X se tedaj porazdeljuje binomsko $B(4, \frac{1}{2})$ in njena verjetnostna shema je:

$$X : \begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ 1/16 & 4/16 & 6/16 & 4/16 & 1/16 \end{pmatrix}.$$

Npr.

$$P(X = 2) = P_4(2) = \binom{4}{2}\left(\frac{1}{2}\right)^2 (1 - \frac{1}{2})^{4-2} = \frac{6}{16}$$

Porazdelitev obravnavane slučajne spremenljivke je simetrična.
Pokazati se da, da je binomska porazdelitev simetrična, le če je $p = 0,5$.

Lahko pokažemo, da je $E[X]=np \in V[X]=npq$

Enakomerna porazdelitev zvezne slučajne spremenljivke

Verjetnostna gostota enakomerno porazdeljene zvezne slučajne spremenljivke je:

$$p(x) = \begin{cases} \dfrac{1}{b-a} & a \leq X \leq b \\ 0 & \text{drugod.} \end{cases}$$

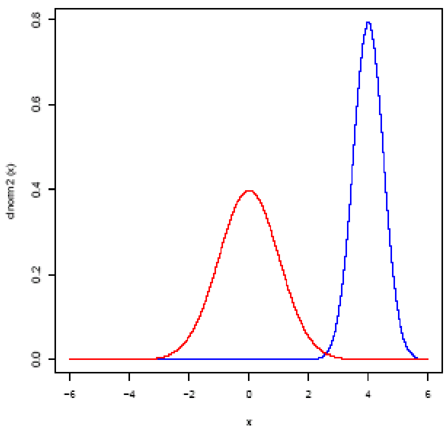
Grafično si jo predstavljamo kot pravokotnik nad intervalom (a, b) višine $\dfrac{1}{b-a}$.

Gaussova normalna porazdelitev

Leta 1738 je Abraham De Moivre (1667-1754) objavil aproksimacijo binomske porazdelitve, ki je normalna krivulja. Leta 1809 je Karl Frederic Gauss (1777-1855) raziskoval matematično ozadje planetarnih orbit, ko je izpeljal normalno porazdelitveno funkcijo.

A. Jurišić in V. Batagelj: Verjetnostni račun in statistika 1'

... Normalna porazdelitev




```
> d2 <- function(x){dnorm(x,mean=4,sd=0.5)}
> curve(d2,-6,6,col="blue")
> curve(dnorm,-6,6,col="red",add=TRUE)
```

Zaloga vrednosti **normalno porazdeljene** slučajne spremenljivke so vsa realna števila, gostota verjetnosti pa je:

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$

Normalna porazdelitev je natanko določena z parametroma: μ in σ.

Če se slučajna spremenljivka X porazdeljuje normalno s parametroma μ in σ zapišemo:

$$X : N(\mu, \sigma).$$

Glej geoGebra.Gauss

Porazdelitev $N(0,1)$ imenujemo standardizirana normalna porazdelitev. Spremenljivko X tipa $N(\mu,\sigma)$ pretvorimo v standardizirano spremenljivko $Z=\frac{x-\mu}{\sigma}$: N(0, 1).

Lahko pokažemo, da je normalna porazdelitev limita binomske porazdelitve. To je leta 1773 pokazal De Moivre. Najboljpi primer dobimo, ko je $p=\frac{1}{2}$.

$$B(n,p) \approx N(np, \sqrt{npq})$$

Glej Gauss.html

CENTRALNI LIMITNI IZREK

Izrek: Če imamo zaporedje neodvisnih slučajnih spremenljivk X_n, ki imajo enako porazdelitev z matematičnim upanjem μ in varianco σ^2, potem limita zaporedja slučajnih spremenljivk Y_n je slučajna spremenljivka, ki je porazdeljena normalno z upanjem μ in varianco σ^2.

$$Y_n = \frac{X_1 + X_2 + \ldots + X_n}{\sigma \sqrt{n}}$$

Izrek je velikega pomena in nam pokaže, zakaj je Gaussova porazdelitev tako pomembna. Namreč rezultati kompleksnih slučajnih pojavov so posledica velikega števila faktorjev, ki vplivajo istočasno in neodvisno na pojav. Posameznih faktorjev ne utegnemo kontrolirati, lahko pa sklepamo, da bo imel končni pojav Gaussovo porazdelite. Tako so na primer porazdeljene napake pri meritvah v fiziki, lastnosti velikih populacij in podobno. Poskusite preučiti porazdelitveni zakon meta treh poštenih kock ali šestih kovancev. Na kratko nam torej izrek pravi, da je vsota velikega števila neodvisnih slučajnih spremenljivk porazdeljena normalno. To nam omogoča tudi preprost način za generacijo slučajnih števil, ki so normalno porazdeljena z dano srednjo vrednostjo in varianco.

Glej Excel.kovanci

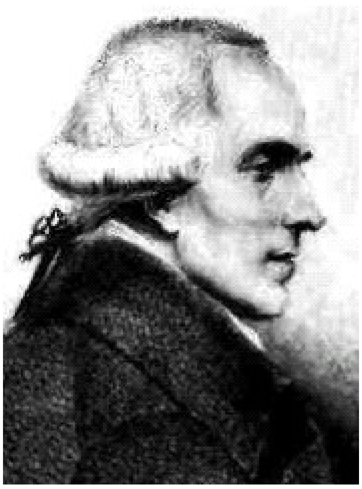
Centralni limitni zrek lahko vidimo tudi iz sledečega zornega kota: če imajo slučajne spremenljivke zaporedja X_n upanje $\mu=0$, potem je vsota velikega števila spremenljivk, ki imajo vedno manjše vrednosti

$$Y_n = \frac{X_1}{\sqrt{n}} + \frac{X_2}{\sqrt{n}} + \ldots + \frac{X_n}{\sqrt{n}}$$

porazdeljena normalno

Leta 1810 je Pierre Laplace (1749-1827) študiral anomalije orbit Jupitra in Saturna, ko je izpeljal razširitev De Moivrovega limitnega izreka,

"Vsaka vsota ali povprečje, če je število členov dovolj veliko,je približno normalno porazdeljena." ***Pierre Laplace***



IZREK O VELIKIH ŠTEVILIH

Izrek: Če imamo zaporedje neodvisnih slučajnih spremenljivk X_n, ki imajo enako porazdelitev z matematičnim upanjem μ, potem limita zaporedja slučajnih spremenljivk Y_n je slučajna spremenljivka, ki je porazdeljena konstantno z upanjem μ.

$$Y_n=\frac{X_1+X_2+...+X_n}{n}$$

Posledice tega izreka so koristne pri inferenčni statistiki, ko iz lastnosti vzorcev neke populacije želimo dobiti lastnosti populacije. Sledi, da za velike vzorce, ki imajo veliki *n,* je zelo velika verjetnost, da bo njihova lastnost tudi lastnost populacije. Zanimiva posledica tega izreka, je tudi dejstvo, da če nek poskus, recimo met kocke ali kovanca ponavljamo veliko krat, potem se bo relativna frekvenca izidov približevala (limita) teoritični verjetnosti izida: če vržem kocko tisočkrat, bo relativna frekvenca šestic zelo blizu številu $\frac{1}{6}$. Lahko smo še bolj natančni.

Glej.Excel.Kovanec

Izrek (Bernoullijeva neenakost): Naj bo $X_{1,}X_{2,...},X_n$ zaporedje neodvisnih slučajnih spremenljivk, ki imajo enako matematično upanje μ in enako varianco σ^2. Naj bo S_n slučajna spremenljivka delnih vsot $S_n=\frac{X_1+X_2+...+X_n}{n}$. Potem velja, da

$$P(|S_n-\mu|\geq k)\leq\frac{\sigma^2}{nk^2} \quad .$$

Vržemo kovanec, slučajna spremenljivka X_1 velja 0 ali 1. X_1 ima seveda upanje $\mu=\frac{1}{2}$ in varianco $\sigma^2=\frac{1}{2}\frac{1}{2}=\frac{1}{4}$. Če poskus ponovimo 3000 krat $X_{1,}X_{2,...},X_{3000}$, potem bo vsota $X_1+X_2+...+X_n$ enaka številu glav pri 3000 metih in $S_n=\frac{X_1+X_2+...+X_n}{n}$ bo predstavljala relativno frekvenco glav. Verjetnost, da se S_n razlikuje od pričakovane vrednosti $\frac{1}{2}$ za manj kot 0.2 je

$$P\left(\left|S_n-\frac{1}{2}\right|\geq 0.2\right)\leq\frac{\left(\frac{1}{4}\right)^2}{3000(0.2)^2}=0.00208$$

oziroma

$$P\left(\frac{1}{2}-0.2\geq S_n \quad ali \quad S_n\geq\frac{1}{2}+0.2\right)\leq 0.00208$$

oziroma

$$P(900\geq št.glav \quad ali \quad št.glav\geq 2100)\leq 0.00208 \quad .$$