Uporabna statistika

Gregor Dolinar

Fakulteta za elektrotehniko
Univerza v Ljubljani

23. oktober 2013

Definicija

Zvezna slučajna spremenljivka X je normalno porazdeljena s parametroma μ, $-\infty < \mu < \infty$, in $\sigma > 0$, če je njena gostota verjetnosti

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

Trditev

Če je X normalna slučajna spremenljivka s parametroma μ in σ, potem je

$$E(X) = \mu.$$

Varianca je

$$V(X) = \sigma^2.$$

Kako vpliva vrednost μ in σ na obliko funkcije, ki predstavlja gostoto verjetnosti?

Definicija

Zvezna slučajna spremenljivka X je standardizirana normalna spremenljivka, če je $\mu = 0$ in $\sigma = 1$. Označimo jo z Z.

Trditev

Če je X normalna slučajna spremenljivka s parametroma μ in σ, potem je

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

standardizirana normalna spremenljivka.

Izrek

(Centralni limitni izrek) Če so X_1, X_2, ..., X_n enako porazdeljene slučajne spremenljivke z matematičnim upanjem μ in varianco σ^2, potem ima v limiti, ko gre $n \to \infty$, slučajna spremenljivka

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

standardizirano normalno porazdelitev.

Primer

Pri sprejemu digitalnega signala imamo šum, ki je normalno porazdeljen, $\mu = 0$ V, $\sigma = 0.45$ V. Sistem potrdi sprejem signala 1, če je napetost večja od 0.9 V.

- Kolikšna je verjetnost, da napačno potrdimo signal?
- Kolikšna je verjetnost, da zgrešimo signal?

Velja si zapomniti:

$$P[\mu - \sigma \leq X \leq \mu + \sigma] = 68\%$$

$$P[\mu - 2\sigma \leq X \leq \mu + 2\sigma] = 95\%$$

$$P[\mu - 3\sigma \leq X \leq \mu + 3\sigma] = 99.7\%$$

Aproksimacija binomske porazdelitve z normalno porazdelitvijo

Spomnimo se.
Naj bo poskus sestavljen iz n Bernoullijevih poskusov. Slučajno spremenljivko, ki je enaka številu uspešnih Bernoullijevih poskusov, imenujemo binomska slučajna spremenljivka.
Vemo: $f(x) = \binom{n}{x} p^x (1-p)^{n-x}$,
$E(X) = np$, $V(X) = np(1-p)$.

Trditev

Naj bo X binomska slučajna spremenljivka. Potem je

$$Z = \frac{X - np}{\sqrt{np(1-p)}}$$

približno standardno normalno porazdeljena spremenljivka. Aproksimacija je dobra, če je $np > 5$ in $n(1-p) > 5$.

Opomba

Če je np ali $n(1-p)$ majhno število, potem je porazdelitev zelo nesimetrična. Torej normalna porazdelitev ne more biti dober približek.

Primer

Pri digitalnem prenosu podatkov privzamemo, da lahko število prispelih bitov, ki so slabi, opišemo z binomsko porazdelitvijo. Verjetnost, da je prispeli bit slab, je 10^{-5}. Kolikšna je verjetnost, da je pri prenosu 16 milijonov bitov več kot 150 prispelih bitov slabih?

Točno:

$P[X > 150]$

$= 1 - \sum_{x=0}^{150} \binom{16 \cdot 10^6}{x} (10^{-5})^x (1 - 10^{-5})^{16 \cdot 10^6 - x}$

Približno:

$P[X > 150] \doteq P[\frac{X-160}{\sqrt{160(1-10^{-5})}} > \frac{150-160}{\sqrt{160(1-10^{-5})}}]$

$= P[Z > -0.79] = P[Z < 0.79] = 0.785.$

Aproksimacija Poissonove porazdelitve z normalno porazdelitvijo

V povprečju na danem intervalu λ pojavov. Interval razdelimo na enake podintervale:

- verjetnost, da na podintervalu več kot 1 pojav, je nič
- verjetnost je enaka za vse podintervale in sorazmerna z dolžino podintervala
- verjetnosti na podintervalih med sabo neodvisne

Poissonova slučajna spremenljikva X s parametrom λ je enaka številu pojavov na danem intervalu.

Trditev

Naj bo X Poissonova slučajna spremenljivka. Potem je

$$Z = \frac{X - \lambda}{\sqrt{\lambda}}$$

približno standardno normalno porazdeljena spremenljivka. Aproksimacija je dobra, če je $\lambda > 5$.

Eksponentna porazdelitev

Denimo, da nas zanima:

- čas, ki mine med dvema zaporednima telefonskima klicema
- čas, ki mine med dvema zaporednima prijavama na strežnik
- razdalja na žici med dvema zaporednima napakama
- razdalja na cesti med dvema zaporednima luknjama

Označimo z X slučajno spremenljivko, ki je enaka razdalji med zaporednima dogodkoma pri Poissonovem poskusu. Naj bo N slučajna spremenljivka, ki označuje število dogodkov pri Poissonovem poskusu na razdalji x. Potem je verjetnost, da je razdalja med zaporednima dogodkoma večja od razdalje x, enaka verjetnosti, da se na razdalji x ni zgodil noben dogodek, torej

$$P[X > x] = P[N = 0] = \frac{e^{-\lambda x}(\lambda x)^0}{0!} = e^{-\lambda x}.$$

Torej je

$$F(x) = P[X \leq x] = 1 - e^{-\lambda x},$$

oziroma

$$f(x) = \lambda e^{-\lambda x}.$$

Definicija

Zvezno slučajno spremenljivko X, ki opisuje razdaljo med zaporednima pojavoma pri Poissonovem procesu s parametrom λ, imenujemo eksponentna slučajna spremenljivka s parametrom $\lambda > 0$. Njena gostota verjetnosti je

$$f(x) = \lambda e^{-\lambda x}, \ \ 0 \leq x.$$

Trditev

Če je X eksponentna slučajna spremenljivka s parametrom λ, potem je

$$\mu = E(X) = \frac{1}{\lambda}.$$

Varianca je

$$V(X) = \frac{1}{\lambda^2}.$$

Primer. Prijavljanje uporabnikov na informacijski sistem lahko modeliramo s Poissonovo porazdelitvijo. V povprečju se prijavi 25 uporabnikov na uro.

- Kolikšna je verjetnost, da se nihče ne prijavi v 6 minutah?
- Kolikšna je verjetnost, da je naslednja prijava med 2 in 3 minutami?
- Kolikšen je časovni interval, da je verjetnost, da ne bo nobene prijave, enaka 0.9?

Eksponentna slučajna spremenljivka nima spomina:

$$P[X < t_1 + t_2 | X > t_1] = P[X < t_2]$$