Sveučilište J. J. Strossmayera u Osijeku
Odjel za matematiku

Ivana Petrović

Normalna slučajna varijabla

Završni rad

Osijek, 2019.

Sveučilište J. J. Strossmayera u Osijeku
Odjel za matematiku

Ivana Petrović

Normalna slučajna varijabla

Završni rad

Mentor: izv. prof. dr. sc. Dragana Jankov Maširević

Osijek, 2019.

Sadržaj

Sažetak

U ovom radu upoznat ćemo se sa slučajnim varijablama. Nakon toga, detaljnije ćemo promatrati neprekidnu slučajnu varijablu i njene karakteristike te ćemo se bazirati na normalnu slučajnu varijablu i njen standardiziran oblik. Također ćemo izvesti očekivanje i varijancu te slučajne varijable, povezati ćemo ju i s Laplaceovom funkcijom te se upoznati s njezinim računanjem pomoću tablica. Naposljetku ćemo navesti i dokazati centralni granični teorem te ga primijeniti na aritmetičku sredinu.

Ključne riječi

slučajna varijabla, normalna slučajna varijabla, standardna normalna slučajna varijabla, očekivanje, varijanca, Laplaceova funkcija, centralni granični teorem

Abstract

In this paper we introduce random variables. After that, we will thoroughly observe continuous random variable and its characteristics and we will especially deal with normal random variable and its standardized form. Furthermore, we will derive expectation and variance for such random variable and see its connection with Laplace function and also how it can be calculated using tables. Finally, we will state and prove central limit theorem and apply it to arithmetic mean.

Key words

random variable, normal random variable, standard normal random variable, expectation, variance, Laplace function, central limit theorem

1 Uvod

Normalna slučajna varijabla je funkcija sa skupa Ω, odnosno prostora elementarnih događaja, u skup realnih brojeva koja ima određena svojstva. Kako bi bolje shvatili pojam i pripadna svojstva normalne slučajne varijable najprije ćemo se uvesti u osnovne pojmove teorije vjerojatnosti. Upoznat ćemo se s pojmom σ-algebre i vjerojatnosti te sa slučajnim varijablama od kojih ćemo posebno obraditi jednu od najvažnijih, odnosno već spomenutu normalnu slučajnu varijablu, pomoću koje ćemo doći do nekih važnih rezultata teorije vjerojatnosti.

2 Osnovni pojmovi

Kako bi definirali vjerojatnost kao funkciju, potrebno je identificirati domenu, odnosno objekte čiju vjerojatnost računamo. Te objekte nazivamo elementarnim događajima, a skup koji ih sadrži prostor elementarnih događaja. Međutim, nije uvijek potreban cijeli prostor elementarnih događaja niti ga je praktično koristiti zbog opsežnosti. Tada uvodimo pojam σ-algebre koja će sadržavati sve događaje od interesa.

Definicija 1. *Neka je dan neprazan skup Ω. Familija $\mathcal{F}$ podskupova skupa Ω jest σ-algebra skupova na Ω ako vrijedi:*

1. $\emptyset \in \mathcal{F}$,
2. *ako je $A \in \mathcal{F}$, onda je i $A^c \in \mathcal{F}$,*
3. *ako je dana prebrojiva familija skupova $(A_n,\ n \in \mathbb{N}) \subseteq \mathcal{F}$, onda $\mathcal{F}$ sadrži i njihovu uniju, tj.*

$$\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}.$$

Na osnovu toga definiramo vjerojatnost:

Definicija 2. *Neka je Ω neprazan prostor elementarnih događaja i $\mathcal{F}$ σ-algebra skupova na njemu. Funkciju*

$$P : \mathcal{F} \to \mathbb{R}$$

zovemo vjerojatnost na Ω ako zadovoljava sljedeća svojstva:

1. *nenegativnost vjerojatnosti: $P(A) \geq 0$ za sve $A \in \mathcal{F}$,*

2. *normiranost vjerojatnosti:* $P(\Omega) = 1$,

3. *σ-aditivnost vjerojatnosti: ako je dana prebrojiva familija međusobno disjunktnih skupova* $(A_i,\ i \in I) \subseteq \mathcal{F},\ I \subseteq \mathbb{N}$, *tj.* $A_i \cap A_j = \emptyset$, *čim je* $i \neq j$, *tada vrijedi*
$$P\Big(\bigcup_{i\in I} A_i\Big) = \sum_{i\in I} P(A_i).$$

Vjerojatnost ovisi i o prostoru na kojem ju zadajemo. Ω može biti diskretan skup (konačan ili prebrojiv), ali i neki interval. Na diskretnom prostoru glavnu ulogu ima niz vjerojatnosti $(p_i,\ i \in \mathbb{N})$ koji zadovoljava određena svojstva, dok kod vjerojatnosti na $\mathbb{R}, \mathbb{R}^2, \mathbb{R}^3, \ldots$ glavnu ulogu ima vjerojatnosna mjera na danom prostoru.

Obzirom da je skup Ω nekad zahtjevno odrediti, uvodimo pojam slučajne varijable, koja će nam omogućiti da promatramo isključivo karateristike od interesa.

Definicija 3. *Neka je* $(\Omega, \mathcal{F}, P)$ *vjerojatnosni prostor. Funkcija* $X : \Omega \to \mathbb{R}$ *jest slučajna varijabla (na* Ω*) ako je* $X^{-1}(B) \in \mathcal{F}$ *za proizvoljan* $B \in \mathcal{B}$ *tj.* $X^{-1}(\mathcal{B}) \subset \mathcal{F}$, *gdje* $\mathcal{B}$ *predstavlja* σ*-algebru Borelovih skupova na* $\mathbb{R}$, *koja je generirana familijom svih otvorenih intervala u* $\mathbb{R}$.

Slučajne varijable dijele se na diskretne i neprekidne u ovisnosti o prostoru elementarnih događaja. Detaljnije o diskretnim slučajnim varijablama može se pronaći u [1].

U nastavku definiramo neprekidnu slučajnu varijablu, koja nam je od interesa u ovom radu.

Definicija 4. *Neka je dan vjerojatnosni prostor* $(\Omega, \mathcal{F}, P)$ *i funkcija* $X : \Omega \to \mathbb{R}$ *za koju vrijedi:*

1. $\{\omega \in \Omega : X(\omega) \leq x\} = \{X \leq x\} \in \mathcal{F}$, *za svaki* $x \in \mathbb{R}$,

2. *postoji nenegativna realna funkcija realne varijable f, takva da vrijedi:*
$$P(\{\omega \in \Omega : X(\omega) \leq x\}) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt,$$
za svaki $x \in \mathbb{R}$.

Funkciju X zovemo apsolutno neprekidna slučajna varijabla na Ω *ili, kraće, neprekidna slučajna varijabla. Funkciju f tada zovemo funkcija gustoće vjerojatnosti slučajne varijable X ili, kraće, funkcija gustoće slučajne varijable X.*

Kako bi opisali svojstva neprekidne slučajne varijable definiramo funkciju distribucije. Naime, iz funkcije distribucije bilo koje slučajne varijable možemo izračunati i iščitati sva bitna svojstva i karakteristike te slučajne varijable.

Definicija 5. *Neka je $(\Omega, \mathcal{F}, P)$ vjerojatnosni prostor i neka je X slučajna varijabla na njemu. Funkcija $F : \mathbb{R} \to [0,1]$ koja realnom broju x pridružuje vjerojatnost da dana slučajna varijabla bude manja ili jednaka tom broju, tj. funkciju*

$$F(x) = P(\{\omega \in \Omega : X(\omega) \leq x\}) = P(X \leq x),$$

zovemo funkcija distribucije slučajne varijable X.

Računanje funkcije distribucije neprekidne slučajne varijable svodi se na računanje integrala:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt, \ x \in \mathbb{R},$$

pri čemu je f funkcija gustoće dane slučajne varijable.

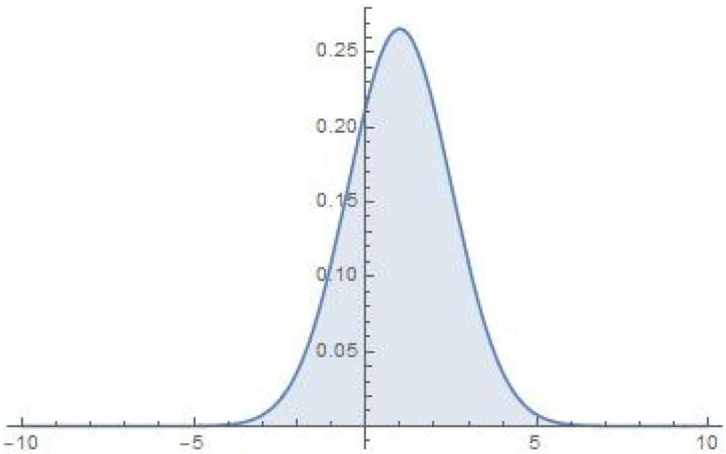
3 Normalna slučajna varijabla

Jedna od najčešće korištenih slučajnih varijabli je svakako normalna slučajna varijabla, odnosno slučajna varijabla sa normalnom distribucijom. Razlog tome je što ona opisuje mjerenja koja su u svakodnevnici često od interesa poput: bodovi na ispitu, IQ testovi, visina ljudi te druge prirodne pojave.
U nastavku definiramo normalnu slučajnu varijablu i njenu funkciju distribucije.

Definicija 6. *Za neprekidnu slučajnu varijablu kažemo da ima Gaussovu ili normalnu distribuciju s parametrima μ i σ^2 ako je njezina funkcija gustoće dana izrazom*

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \ x \in \mathbb{R},$$

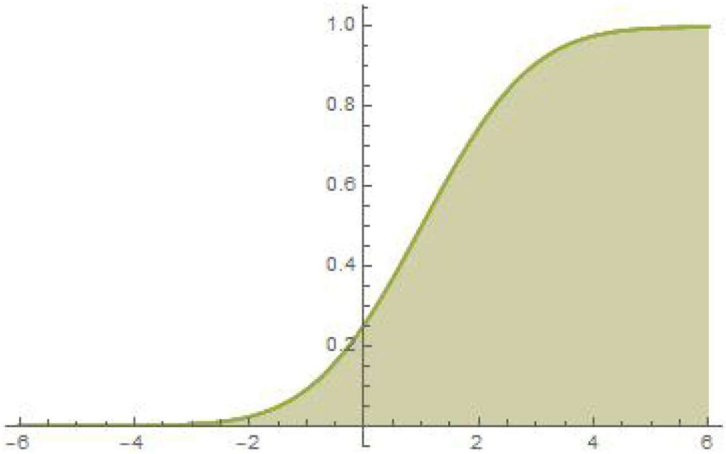
gdje su μ i σ realni brojevi i $\sigma > 0$. Ako slučajna varijabla X ima normalnu distribuciju s parametrima μ i σ^2, koristimo oznaku $X \sim \mathcal{N}\ (\mu, \sigma^2)$.



Slika 1: Graf funkcije gustoće normalne slučajne varijable s očekivanjem 1 i standardnom devijacijom 1.5

Iz funkcije gustoće lako se izvede i funkcija distribucije normalne slučajne varijable:

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt, \ x \in \mathbb{R}.$$



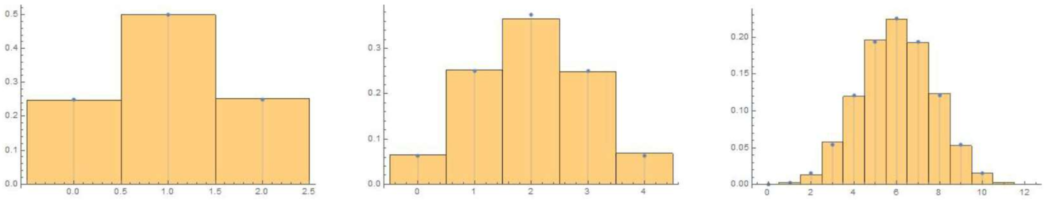
Slika 2: Graf funkcije distribucije normalne slučajne varijable s očekivanjem 1 i standardnom devijacijom 1.5

Normalnu slučajnu varijablu prvi je otkrio de Moivre kao aproksimaciju binomne slučajne varijable što je diskretna slučajna varijabla koja prima vrijednosti iz skupa $\{0, 1, 2, \ldots, n\}$, pri čemu su pripadne vjerojatnosti:

$$p_i = P(X = i) = \binom{n}{i} p^i (1-p)^{n-i},\ i \in \{0, 1, \ldots, n\},$$

gdje je $p \in \langle 0, 1\rangle$ vjerojatnost uspjeha u jednom izvođenju pokusa, a $n \in \mathbb{N}$ je broj nezavisnih ponavljanja pokusa. Također, uvodimo oznaku $q = 1 - p$. Slučajnu varijablu X, koja ima binomnu distribuciju s parametrima n i p, označavamo $X \backsim \mathcal{B}(n, p)$.

De Moivre je uočio, promatrajući pokus bacanja novčića, da povećavanjem broja ponavljanja funkcija gustoće binomne slučajne varijable poprima oblik zvona. Zaključio je da ako uzmemo $\mu = np$ i $\sigma = \sqrt{npq}$, gdje su n, p, q parametri binomne slučajne varijable, dobijemo normalnu slučajnu varijablu koja je dobra aproksimacija te binomne slučajne varijable. Koristili su ju i Laplace, za mjerenje pogrešaka te Gauss za astronomske podatke.



(a) $n = 2$ (b) $n = 4$ (c) $n = 12$

Slika 3: De Moivreov pokus bacanja novčića, odnosno binomna slučajna varijabla s parametrima $p = 0.5$ i $n = 2,\ 4$ i 12

Postupkom standardizacije normalne slučajne varijable (vidi [1, str. 103]) dobivamo novu slučajnu varijablu

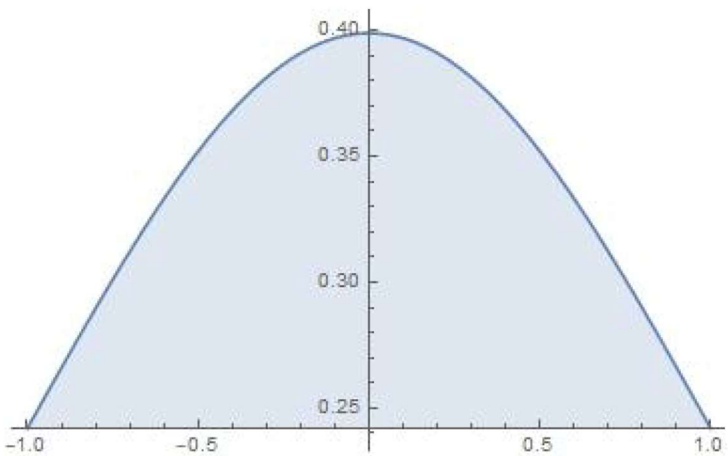
$$X^* = \frac{X - \mu}{\sigma}$$

i to će ponovno biti normalna slučajna varijabla, no s parametrima $\mu = 0$ i $\sigma^2 = 1$. Takva normalna slučajna varijabla, naziva se jedinična ili standardna normalna slučajna varijabla, označavamo je s $X^* \sim \mathcal{N}(0, 1)$ te joj funkcija gustoće ima oblik:

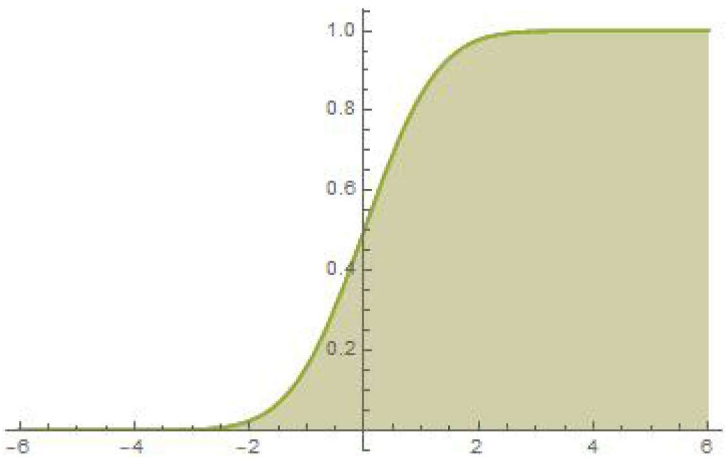
$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}},\ x \in \mathbb{R}.$$

Njena funkcija distribucije dana je s:

$$F(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-\frac{t^2}{2}} dt,\ x \in \mathbb{R}.$$



Slika 4: Graf funkcije gustoće standardne normalne varijable



Slika 5: Graf funkcije distribucije standardne normalne varijable

3.1 Numeričke karakteristike slučajne varijable

Promatrajući graf funkcije gustoće lako se može uočiti kako je on simetričan s obzirom na pravac paralelan s y-osi. Taj pravac također siječe x-os u točki čija vrijednost predstavlja očekivanje slučajne varijable.

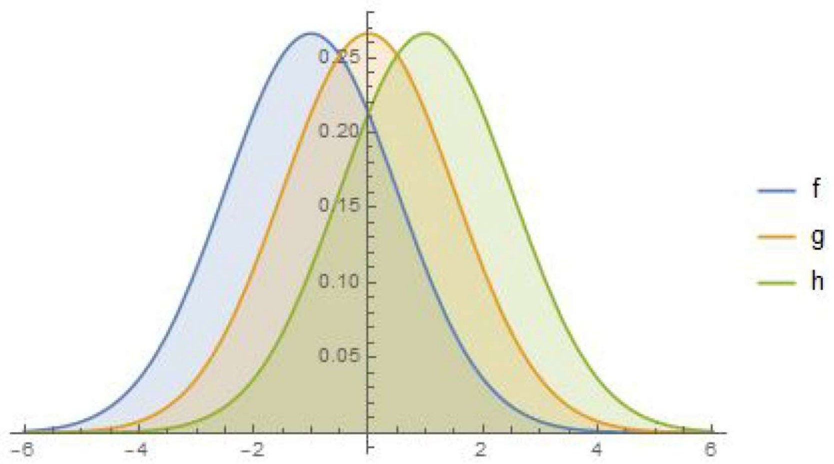
Definicija 7. *Neka je X neprekidna slučajna varijabla s funkcijom gustoće f. Ako je integral*

$$\int_{-\infty}^{\infty} |x| f(x) dx < \infty,$$

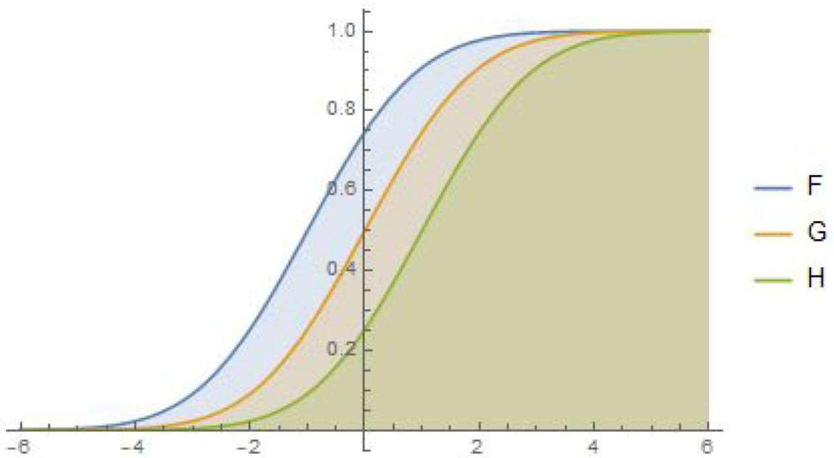
onda kažemo da slučajna varijabla X ima očekivanje i broj

$$\mu = EX = \int_{-\infty}^{\infty} x f(x) dx$$

zovemo matematičko očekivanje (ili samo očekivanje) neprekidne slučajne varijable X.



Slika 6: Graf funkcije gustoće normalne slučajne varijable s očekivanjima -1, 0, 1 (koja su prikazana funkcijama f, g i h, redom) i standardnom devijacijom 1.5



Slika 7: Graf funkcije distribucije normalne slučajne varijable s očekivanjima -1, 0, 1 (koja su prikazana funkcijama F, G i H, redom) i standardnom devijacijom 1.5

Sa prethodnih slika možemo uočiti kako promjenom vrijednosti očekivanja normalne slučajne varijable translatiramo graf funkcije gustoće i funkcije distribucije te slučajne varijable ulijevo, odnosno udesno po x-osi (u ovisnosti o parametru μ).

Da bismo izveli očekivanje normalne slučajne varijable koristit ćemo Gama funkciju koju označavamo s Γ. Gama funkcija se često još naziva i Eulerov integral druge vrste, Faktorijel ili Pi funkcija što proizlazi iz njezinih svojstava i oznaka. Za pozitivne realne brojeve možemo ju definirati kao:

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt, \ x > 0.$$

Za prirodne brojeve vrijedi:

$$\Gamma(n) = (n-1)!, \ n \in \mathbb{N}$$

odnosno

$$\Gamma(n) = (n-1)\Gamma(n-1).$$

Vrijednosti od $\Gamma\left(\frac{k}{2}\right)$, $k \in \mathbb{Z}$, dobijemo koristeći dvostruke faktorijele kao:

$$\Gamma\left(n + \frac{1}{2}\right) = \frac{(2n-1)!!}{2^n}\sqrt{\pi} = \frac{(2n)!}{4^n n!}\sqrt{\pi}$$

i

$$\Gamma\left(\frac{1}{2} - n\right) = \frac{(-2)^n}{(2n-1)!!}\sqrt{\pi} = \frac{(-4)^n n!}{(2n)!}\sqrt{\pi},$$

gdje je dvostruki faktorijel definiran s:

$$n!! = \begin{cases} 1, & n = -1, 0 \\ n \cdot (n-2) \cdot (n-4) \cdots 5 \cdot 3 \cdot 1, & n = 1, 3, 5, \dots \\ n \cdot (n-2) \cdot (n-4) \cdots 6 \cdot 4 \cdot 2, & n = 2, 4, 6, \dots \end{cases} .$$

Više o Gama funkciji i njezinom računanju možete pogledati u [4, str. 435], a o dvostrukom faktorijelu i njegovoj vezi s jednostrukim faktorijelom u [4, str. 25].

Izvedimo sada očekivanje normalne slučajne varijable. Neka je X normalna slučajna varijabla s parametrima μ i σ^2, odnosno

$$X \backsim \mathcal{N}(\mu, \sigma^2).$$

Kako je njena funkcija gustoće dana s:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \ \mu \in \mathbb{R}, \ x \in \mathbb{R}, \ \sigma > 0,$$

koristeći se definicijom očekivanja slijedi:

$$EX = \int_{-\infty}^{\infty} xf(x)dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} xe^{-\frac{(x-\mu)^2}{2\sigma^2}} dx.$$

Primijenimo li supstituciju $t = \frac{x-\mu}{\sigma}$, iz čega je i $dt = \frac{dx}{\sigma}$, slijedi da je očekivanje:

$$EX = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma t + \mu)e^{-\frac{t^2}{2}} dt = \frac{1}{\sqrt{2\pi}} \Big(\sigma \int_{-\infty}^{\infty} te^{-\frac{t^2}{2}} dt + \mu \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt\Big).$$

Uočimo da je $te^{-\frac{t^2}{2}}$ neparna funkcija pa koristimo svojstvo da je integral neparne funkcije na simetričnom intervalu jednak 0. Također, $e^{-\frac{t^2}{2}}$ je parna funkcija na simetričnom intervalu pa ju možemo računati kao $2\int_0^{\infty} e^{-\frac{t^2}{2}} dt$. U nastavku računa imamo da je

$$EX = \frac{1}{\sqrt{2\pi}}(0 + 2\mu \int_0^{\infty} e^{-\frac{t^2}{2}} dt) = \frac{\sqrt{2}\mu}{\sqrt{\pi}} \int_0^{\infty} e^{-\frac{t^2}{2}} dt.$$

Novom supstitucijom $u = \frac{t^2}{2}$, odakle je $tdt = du$ i $t = \sqrt{2u}$ dobijemo:

$$EX = \frac{\mu}{\sqrt{\pi}} \int_0^{\infty} e^{-u} \frac{du}{\sqrt{u}} = \frac{\mu}{\sqrt{\pi}} \int_0^{\infty} u^{-\frac{1}{2}} e^{-u} du = \frac{\mu}{\sqrt{\pi}} \int_0^{\infty} u^{\frac{1}{2}-1} e^{-u} du = \frac{\mu}{\sqrt{\pi}} \Gamma\Big(\frac{1}{2}\Big),$$

gdje je Γ Gama funkcija te slijedi:

$$EX = \mu \frac{1}{\sqrt{\pi}} \sqrt{\pi} = \mu.$$

Širina zvonolike krivulje, koja predstavlja graf funkcije gustoće normalne slučajne varijable, ovisi i o brojčanoj vrijednosti koja se naziva varijanca.

Definicija 8. *Neka je X slučajna varijabla na vjerojatnosnom prostoru $(\Omega, \mathcal{F}, P)$ i $r > 0$.*

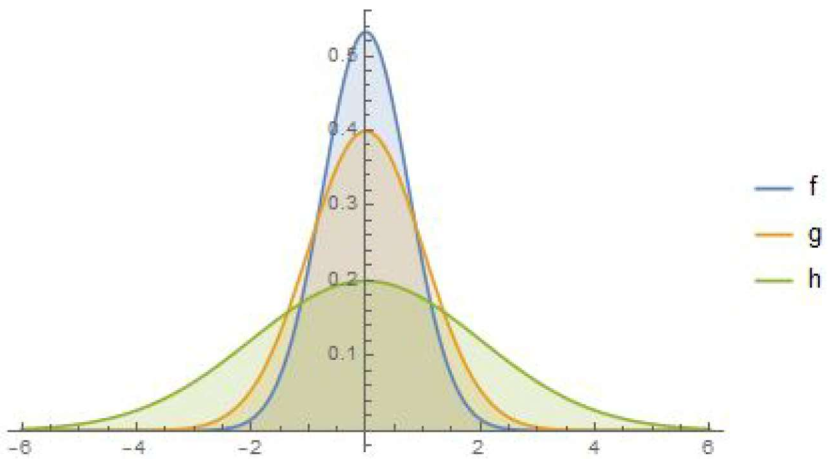
- *Ako postoji $E(X^r)$, onda broj $\mu_r = E(X^r)$ zovemo moment r-tog reda ili r-ti moment od X.*
- *Ako postoji $E(|X|^r)$, onda broj $E(|X|^r)$ zovemo apsolutni moment r-tog reda ili r-ti apsolutni moment od X.*

- *Ako postoji* EX *i* $E(|X - EX|^r)$*, onda broj* $E(|X - EX|^r)$ *zovemo r-ti centralni moment od* X*.*

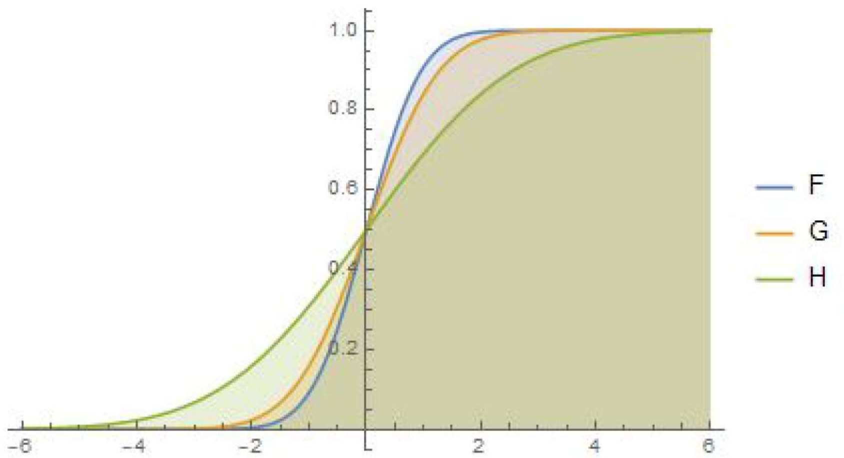
- *Ako postoji* $E(X - EX)^2$*, onda taj nenegativan broj zovemo varijanca slučajne varijable* X *i označavamo* $VarX$*,* σ_x^2 *ili* σ^2*. Korijen iz varijance (u oznaci* $\sqrt{VarX}$*,* σ_x *ili* σ*) nazivamo standardna devijacija slučajne varijable.*

Možemo uočiti da je kod neprekidne slučajne varijable varijanca dana s:

$$VarX = \int_{-\infty}^{\infty} (x - EX)^2 f(x) dx.$$



Slika 8: Graf funkcije gustoće normalne slučajne varijable s očekivanjem 0 i standardnim devijacijama 0.75, 1 i 2 (koje su prikazane funkcijama f, g i h, redom)



Slika 9: Graf funkcije distribucije normalne slučajne varijable s očekivanjem 0 i standardnim devijacijama $0.75, 1$ i 2 (koje su prikazane funkcijama F, G i H, redom)

Za razliku od očekivanja, promjena vrijednosti varijance normalne slučajne varijable mijenja oblik grafa. Naime, na slikama 8 i 9 možemo uočiti kako povećavanjem vrijednosti varijance graf funkcije gustoće postaje *širi*, a graf funkcije distribucije ima manji nagib, odnosno u tom slučaju se podatci *raspršuju* od očekivanja. S druge strane, smanjenjem vrijednosti varijance graf funkcije gustoće postaje *uži*, a graf funkcije distribucije ima veći nagib. U nastavku slijedi izvod formule za varijancu normalne slučajne varijable. Kako od prije znamo da je $EX = \mu$, slijedi:

$$VarX = \int_{-\infty}^{\infty} \frac{(x-\mu)^2}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx.$$

Istom supstitucijom koju smo koristili i kod računanja očekivanja, $t = \frac{x-\mu}{\sigma}$, dobivamo:

$$VarX = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t^2 e^{-\frac{t^2}{2}} dt.$$

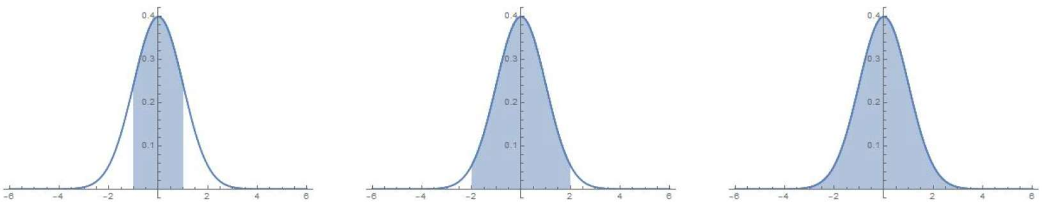
Ponovno koristeći svojstvo parne podintegralne funkcije imamo:

$$VarX = \frac{\sqrt{2}\sigma^2}{\sqrt{\pi}} \int_{0}^{\infty} t^2 e^{-\frac{t^2}{2}} dt.$$

Sljedećom supstitucijom $u = \frac{t^2}{2}$, odakle je $tdt = du$, a $t = \sqrt{2u}$ dobivamo:

$$
\begin{aligned}
VarX &= \frac{2\sigma^2}{\sqrt{\pi}} \int\limits_0^\infty u^{\frac{1}{2}} e^{-u} du = \frac{2\sigma^2}{\sqrt{\pi}} \int\limits_0^\infty u^{\frac{3}{2}-1} e^{-u} du \\
&= 2\sigma^2 \frac{1}{\sqrt{\pi}} \Gamma\left(\frac{3}{2}\right) = 2\sigma^2 \frac{1}{\sqrt{\pi}} \frac{\sqrt{\pi}}{2} = \sigma^2.
\end{aligned}
$$

Kod grafa funkcije gustoće normalne slučajne varijable aritmetička sredina, medijan (vrijednost u skupu podataka od koje je 50% podataka veće ili jednako i 50% podataka manje ili jednako) i mod (vrijednost koja se najčešće pojavljuje u skupu podataka, odnosno ima najveću frekvenciju) su isti. Većina podataka nalazi se u središtu grafa. Također unutar jedne standardne devijacije od očekivanja nalazi se 68% podataka, unutar dvije standardne devijacije 95%, a unutar tri 99.7% podataka. 50% podataka je manje od očekivanja, a 50% veće. Ove podatke možemo dobiti korištenjem programa Statistica ali i uz pomoć statističkih tablica, pomoću Laplaceove fukcije koju ćemo definirati u sljedećem poglavlju.



(a) Podatci unutar jedne standardne devijacije

(b) Podatci unutar dvije standardne devijacije

(c) Podatci unutar tri standardne devijacije

Slika 10: Količina podataka standardne normalne distribucije unutar jedne, dvije i tri standardne devijacije

3.2 Primjena normalne slučajne varijable

Kako bi računali vjerojatnost normalne slučajne varijable $X \sim \mathcal{N}(\mu, \sigma^2)$ na nekom intervalu $\langle a, b \rangle$, najčešće nećemo koristiti funkciju distribucije i gustoće ove slučajne varijable nego standardne normalne slučajne varijable. Kao što smo već vidjeli, standardna normalna slučajna varijabla ima funkciju distribucije danu s:

$$
F(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-\frac{t^2}{2}} dt, \; x \in \mathbb{R}.
$$

Laplaceova funkcija definirana je s (vidi [3, str. 41]):

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_0^x e^{-\frac{t^2}{2}} dt$$

pa je odnos prethodnih dviju funkcija dan izrazom: $\phi(x) = F(x) - 0.5$. Prema tome, vjerojatnost opisanu na početku ovog poglavlja možemo dobiti kao:

$$P(a < X < b) = \phi\left(\frac{b-\mu}{\sigma}\right) - \phi\left(\frac{a-\mu}{\sigma}\right),$$

gdje vrijednost $\phi(x)$ čitamo iz tablica koje se mogu pronaći u [3].
Na primjer, primjenom prethodne formule za $X \backsim \mathcal{N}(0,1)$ računamo količinu podataka unutar jedne standardne devijacije kao:

$$P(-1 < X < 1) = \phi\left(\frac{1-0}{1}\right) - \phi\left(\frac{-1-0}{1}\right) = 0.34134 + 0.34134 = 0.68268,$$

unutar dvije standardne devijacije:

$$P(-2 < X < 2) = \phi\left(\frac{2-0}{1}\right) - \phi\left(\frac{-2-0}{1}\right) = 0.47725 + 0.47725 = 0.9545$$

te unutar tri standardne devijacije:

$$P(-3 < X < 3) = \phi\left(\frac{3-0}{1}\right) - \phi\left(\frac{-3-0}{1}\right) = 0.49865 + 0.49865 = 0.9973.$$

Rezultati koje smo dobili korištenjem Laplaceove funkcije potvrđuju prethodno navedene rezultate sa str. 13 koje vidimo i na slici 10.

Osim Laplaceove funkcije, funkcija distribucije normalne slučajne varijable povezana je i s Eror funkcijom. Naime, Eror funkcija definirana je izrazom:

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int\limits_0^x e^{-t^2} dt,$$

a njezina komplementarna funkcija, koju jednostavno nazivamo komplementarna Eror funkcija, dana je s:

$$\operatorname{erfc}(x) = \frac{2}{\sqrt{\pi}} \int\limits_x^\infty e^{-t^2} dt$$

pa očigledno vrijedi:

$$\text{erf}\,(x) + \text{erfc}\,(x) = 1.$$

Prema tome, za funkciju distribucije F normalne slučajne varijable, gdje je f njena funkcija gustoće, vrijedi:

$$F(x) = \int_{-\infty}^{x} f(t)dt = \frac{1}{2} + \frac{1}{2}\text{erf}\left(\frac{X-\mu}{\sqrt{2}\sigma}\right) = 1 - \frac{1}{2}\text{erfc}\left(\frac{X-\mu}{\sqrt{2}\sigma}\right).$$

U tom se kontekstu $|X - \mu|$ često naziva pogreška ishoda, a σ standardna pogreška. Više detalja o Eror funkciji može se pronaći u [4].

3.3 Centralni granični teorem

Centralni granični teorem govori o graničnom ponašanju niza slučajnih varijabli. Naime, za dovoljno velik broj varijabli, suma niza tih slučajnih varijabli će konvergirati po distribuciji prema standardnoj normalnoj slučajnoj varijabli. Pojmove konvergencije po distribuciji te niza nezavisnih i jednako distribuiranih slučajnih varijabli, koji će biti potrebni za bolje razumijevanje ovog teorema objašnjavamo u nastavku.

Definicija 9. *Kažemo da je niz slučajnih varijabli* $(X_n, n \in \mathbb{N})$ *niz nezavisnih i jednako distribuiranih slučajnih varijabli ako vrijedi:*

1. *sve slučajne varijable* $X_1, X_2, \ldots$ *imaju istu distribuciju i*
2. *za sve* $n \in \mathbb{N}$ *slučajne varijable* $X_1, \ldots, X_n$ *su nezavisne.*

Definicija 10. *Neka je* $(X_n,\ n \in \mathbb{N})$ *niz slučajnih varijabli s pripadnim nizom funkcija distribucije* $(F_n,\ n \in \mathbb{N})$. *Ako postoji funkcija distribucije* F *takva da* $F_n(x) \to F(x)$ *za svaki* x *u kojemu je* F *neprekidna, tada kažemo da niz* $(X_n,\ n \in \mathbb{N})$ *konvergira po distribuciji prema slučajnoj varijabli* X *čija je funkcija distribucije* F *i pišemo:*

$$X_n \xrightarrow{D} X.$$

Sada možemo iskazati Levy-ev centralni granični teorem:

Teorem 3.1. *Neka je* $(X_n,\ n \in \mathbb{N})$ *niz nezavisnih i jednako distribuiranih slučajnih varijabli s očekivanjem* μ *i varijancom* σ^2 *i neka je* S_n *njegova n-ta parcijalna suma. Tada*

$$\frac{S_n - ES_n}{\sqrt{VarS_n}} \xrightarrow{D} Z \sim \mathcal{N}(0,1).$$

Kako bi dokazali ovaj teorem bit će nam potrebni neki osnovni pojmovi i teoremi vezani za karakterističnu funkciju, koje navodimo bez dokaza.

Definicija 11. *Karakteristična funkcija od F jest funkcija φ definirana s*

$$\varphi(t) = \int_{-\infty}^{\infty} e^{itx} dF(x) = \int_{-\infty}^{\infty} cos(tx) dF(x) + i \int_{-\infty}^{\infty} sin(tx) dF(x), \ t \in \mathbb{R}.$$

Definicija 12. *Neka je X slučajna varijabla s funkcijom distribucije F_X. Karakteristična funkcija φ_X od X je karakteristična funkcija od F_X.*

Teorem 3.2. *Ako je $E(|X|^n) < \infty$ za neki $n \in \mathbb{N}$, tada φ_X ima k-tu derivaciju za $k \leq n$ i vrijedi*

$$\varphi_X(t) = \sum_{k=0}^{n} \frac{(it)^k}{k!} E(X^k) + o(t^n), \ t \in \mathbb{R}.$$

Teorem 3.3. *Ako su $X_1, X_2, \ldots, X_n$ nezavisne slučajne varijable, tada vrijedi*

$$\varphi_{\sum\limits_{k=1}^{n} X_k}(t) = \prod_{k=1}^{n} \varphi_{X_k}(t), \ t \in \mathbb{R}.$$

Teorem 3.4. *Neka je $g : \mathbb{N} \to \mathbb{R}$ takva funkcija da je $\lim\limits_{n \to \infty} g(n) = \lambda, \ \lambda \in \mathbb{R}$. Tada je*

$$\lim_{n \to \infty} \left[1 + \frac{g(n)}{n}\right]^n = e^{\lambda}.$$

Definicija 13. *Kažemo da niz $(F_n, \ n \in \mathbb{N})$ funkcija distribucije slabo konvergira prema funkciji distribucije F ako vrijedi*

$$F(x) = \lim_{n \to \infty} F_n(x),$$

za svaki x u kojemu je F neprekidna i označavamo $F_n \xrightarrow{w} F$ za $n \to \infty$ ili $F = w - \lim\limits_{n \to \infty} F_n$.

Teorem 3.5. *Neka je $(F_{X_n}, \ n \in \mathbb{N})$ niz funkcija distribucije i $(\varphi_{X_n}, \ n \in \mathbb{N})$ odgovarajući niz karakterističnih funkcija. Ako za svaki $t \in \mathbb{R}$ postoji $\lim\limits_{n \to \infty} \varphi_{X_n}(t) = \varphi_X(t)$ i ako je funkcija φ_X neprekidna u $t = 0$, tada je φ_X karakteristična funkcija funkcije distribucije F_X i vrijedi $F_{X_n} \xrightarrow{w} F_X$.*

Dokaze ovih teorema i detaljnije o karakterističnoj funkciji može se pronaći u [2]. Iz navedenog možemo dokazati centralni granični teorem:

Dokaz. Označimo $Z_k = \frac{X_k - \mu}{\sigma}$, $k \in \mathbb{N}$. Niz $(Z_k,\ k \in \mathbb{N})$ je niz nezavisnih, jednako distribuiranih slučajnih varijabli te za svaki k vrijedi $EZ_k = 0$ i $E(Z_k^2) = VarZ_k = 1$. Primjenom teorema 3.2 slijedi $\varphi_{Z_k}(t) = 1 - \frac{t^2}{2} + o(t^2)$. Karakteristične funkcije su iste za jednako distribuirane slučajne varijable te ako uvedemo oznaku

$$Y_n = \frac{S_n - ES_n}{\sigma\sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} Z_k,\ n \in \mathbb{N},$$

koristeći teorem 3.3 dobivamo

$$\varphi_{Y_n}(t) = \prod_{k=1}^{n} \varphi_{Z_k}\Big(\frac{t}{\sqrt{n}}\Big) = \Big[1 - \frac{t^2}{2n} + o\Big(\frac{t^2}{n}\Big)\Big]^n,\ t \in \mathbb{R}.$$

Prema teoremu 3.4 vrijedi $\lim_{n\to\infty} \varphi_{Y_n}(t) = e^{-\frac{t^2}{2}}$, $t \in \mathbb{R}$ pa iz teorema 3.5 slijedi tvrdnja. □

Centralni granični teorem može se primijeniti i na specifične slučajne varijable kao što ćemo to navesti u de Moivre-Laplaceovom teoremu:

Teorem 3.6. *Neka je $S_n \backsim \mathcal{B}(n,p)$, gdje je $n \in \mathbb{N}$, $0 < p < 1$. Tada vrijedi*

$$\frac{S_n - np}{\sqrt{np(1-p)}} \xrightarrow{D} \mathcal{N}(0,1).$$

Dokaz. Slijedi direktno iz Levy-evog centralnog graničnog teorema. □

U nastavku uvedimo pojam konvergencije po vjerojatnosti. On će nam biti potreban za teorem koji slijedi, a govori nam o primjeni centralnog graničnog teorema na aritmetičku sredinu.

Definicija 14. *Neka je $(X_n,\ n \in \mathbb{N})$ niz slučajnih varijabli. Ako postoji slučajna varijabla X takva da je*

$$\lim_{n\to\infty} P(|X_n - X| < \varepsilon) = 1,$$

za svaki $\varepsilon > 0$, kažemo da niz X_n slučajnih varijabli konvergira prema slučajnoj varijabli X po vjerojatnosti i označavamo

$$X_n \xrightarrow{P} X.$$

Definicija 15. *Za niz slučajnih varijabli* $(X_n,\ n \in \mathbb{N})$ *kažemo da zadovoljava slabi zakon velikih brojeva ako postoji konstanta* c *takva da vrijedi*

$$\lim_{n\to\infty} P(|X_n - c| < \varepsilon) = 1,$$

za svaki $\varepsilon > 0$ *i to označavamo s*

$$X_n \xrightarrow{P} c.$$

Teorem 3.7. *Neka je* $(X_n,\ n \in \mathbb{N})$ *niz nezavisnih i jednako distribuiranih slučajnih varijabli, koje imaju ograničenu varijancu* $Var(X_i) = \sigma^2 \leq M,\ M > 0$ *i* $E(X_i) = \mu,\ i = 1, \dots, n$. *Tada za aritmetičku sredinu*

$$\overline{X}_n = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$$

vrijedi

$$\lim_{n\to\infty} P(|\overline{X}_n - \mu| < \varepsilon) = 1,$$

za svaki $\varepsilon > 0$ *i*

$$\overline{X}_n \xrightarrow{P} \mu.$$

Dokaz. Očigledno je $E(\overline{X}_n) = \mu$ i $Var(\overline{X}_n) = \frac{\sigma^2}{n}$, a koristeći Čebiševljevu nejednakost na koju smo primijenili svojstvo vjerojatnosti suprotnog događaja (vidi [1]), odnosno

$$P(|X - EX| < k\sigma) \geq 1 - \frac{1}{k^2}$$

dobivamo

$$P(|\overline{X}_n - E(\overline{X}_n)| < \varepsilon) \geq 1 - \frac{Var(\overline{X}_n)}{\varepsilon^2},$$

odakle je

$$P(|\overline{X}_n - \mu| < \varepsilon) \geq 1 - \frac{\sigma^2}{n\varepsilon^2}.$$

Djelujući limesom na prethodnu nejednakost slijedi da je

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}(X_1 + \cdots + X_n) - \mu\right| < \varepsilon\right) = 1.$$

□

Literatura

[1] M. Benšić, N. Šuvak, Uvod u vjerojatnost i statistiku, Sveučilište J. J. Strossmayera, Odjel za matematiku, Osijek, 2014.

[2] N. Sarapa, Teorija vjerojatnosti, Školska knjiga, Zagreb, 1986.

[3] T. Pogány, Z. Zenzerović, Statističke tablice s uputama za primjenu, Sveučilište u Rijeci, Pomorski fakultet, Rijeka, 1993.

[4] K. Oldham, J. Myland, J. Spanier, An atlas of functions -with Equator, the Atlas Function Calculator, Springer Science+Business Media (drugo izdanje), New York, 2009.

[5] M. Spajić, Karakteristične funkcije, Odjel za matematiku, Sveučilište u Osijeku, Završni rad, Osijek, 2015.

[6] V. Čuljak, Vjerojatnost i statistika, Građevinski fakultet, Sveučilište u Zagrebu, Zagreb, 2011. (radni materijal)