

Statistika

Vanja Wagner

3. Osnove vjerojatnosti

Neprekidna distribucija.

Promotrimo sada slučajne varijable koje mogu poprimiti beskonačno neprebrojivo mnogo vrijednosti. U takvu klasu spada slučajna varijabla koja mjeri (nezaokruženu) visinu slučajno odabranog studenta Biologije.

Neprekidna distribucija.

Promotrimo sada slučajne varijable koje mogu poprimiti beskonačnoeprebrojivo mnogo vrijednosti. U takvu klasu spada slučajna varijabla koja mjeri (nezaokruženu) visinu slučajno odabranog studenta Biologije.

Za slučajnu varijablu X kažemo da je **neprekidna** ukoliko postoji funkcija f_X takva da je

$$F_X(x) = P(X \leq x) = \int_1^x f_X(t) dt:$$

Neprekidna distribucija.

Promotrimo sada slučajne varijable koje mogu poprimiti beskonačno neprebrojivo mnogo vrijednosti. U takvu klasu spada slučajna varijabla koja mjeri (nezaokruženu) visinu slučajno odabranog studenta Biologije.

Za slučajnu varijablu X kažemo da je **neprekidna** ukoliko postoji funkcija f_X takva da je

$$F_X(x) = P(X \leq x) = \int_1^x f_X(t) dt:$$

Gornja vjerojatnost je zadana površinom ispod krivulje funkcije f_X :

Vjerojatnosna funkcija gustoće

Funkcija f_X se naziva **vjerojatnosna funkcija gustoće** slučajne varijable X i zadovoljava sljedeća svojstva:

Vjerojatnosna funkcija gustoće

Funkcija f_X se naziva **vjerojatnosna funkcija gustoće** slučajne varijable X i zadovoljava sljedeća svojstva:

$f(x) > 0$ - vjerojatnost je uvijek pozitivna

Vjerojatnosna funkcija gustoće

Funkcija f_X se naziva **vjerojatnosna funkcija gustoće** slučajne varijable X i zadovoljava sljedeća svojstva:

$f(x) > 0$ - vjerojatnost je uvijek pozitivna

$\int_{-\infty}^{\infty} f(t) dt = 1$ - vjerojatnost sigurnog događaja je 1.

Vjerojatnosna funkcija gustoće

Funkcija f_X se naziva **vjerojatnosna funkcija gustoće** slučajne varijable X i zadovoljava sljedeća svojstva:

$f(x) > 0$ - vjerojatnost je uvijek pozitivna

$\int_1^Z f(t) dt = 1$

- vjerojatnost sigurnog događaja je 1.

Očekivanje neprekidne slučajne varijable:

$$E(X) = \int_1^Z t f(t) dt$$

Vjerojatnosna funkcija gustoće

Funkcija f_X se naziva **vjerojatnosna funkcija gustoće** slučajne varijable X i zadovoljava sljedeća svojstva:

$f(x) > 0$ - vjerojatnost je uvijek pozitivna

$\int_1^Z f(t) dt = 1$

- vjerojatnost sigurnog događaja je 1.

Očekivanje neprekidne slučajne varijable:

$$E(X) = \int_1^Z t f(t) dt$$

Varijanca:

$$\text{Var}(X) = E(X^2) - (E(X))^2 = \int_1^Z t^2 f(t) dt - (E(X))^2.$$

Normalna distribucija - motivacija

Næ glavni primjer neprekidne distribucije bitcenormalna ili Gaussova distribucija, odnosno distribucija kojoj graf funkcije gustæe odgovara tzv. Gaussovoj krivulji (krivulji zvona).

Normalna distribucija - motivacija

Næ glavni primjer neprekidne distribucije bitcenormalna ili Gaussova distribucija, odnosno distribucija kojoj graf funkcije gustæe odgovara tzv. Gaussovoj krivulji (krivulji zvona).

Promotrimo kao motivaciju gracki prikaz distribucija standardizirane binomne slučajne varijable.

Normalna distribucija - motivacija

Næ glavni primjer neprekidne distribucije bitcenormalna ili Gaussova distribucija, odnosno distribucija kojoj graf funkcije gustæe odgovara tzv. Gaussovoj krivulji (krivulji zvona).

Promotrimo kao motivaciju gracki prikaz distribucija standardizirane binomne slučajne varijable. Neka je $X \sim B(n; \frac{1}{2})$ i

$$Y = \frac{X - E(X)}{\sqrt{E(X)}} = \frac{X - \frac{n}{2}}{\sqrt{\frac{n}{4}}}$$

Normalna distribucija - motivacija

Næ glavni primjer neprekidne distribucije bitcenormalna ili Gaussova distribucija, odnosno distribucija kojoj graf funkcije gustæe odgovara tzv. Gaussovoj krivulji (krivulji zvona).

Promotrimo kao motivaciju gracki prikaz distribucija standardizirane binomne slucajne varijable. Neka je $X \sim B(n; \frac{1}{2})$ i

$$Y = \frac{X - E(X)}{\sqrt{Var(X)}} = \frac{X - \frac{n}{2}}{\sqrt{\frac{n}{4}}}$$

Promotrimo gracki prikaz distribucije od Y za $n = 10; 100; 500; 10000$.

Normalna distribucija - definicija

Distribucija neprekidne slučajne varijable je određena pripadnom vjerojatnosnom funkcijom gustoće.

Normalna distribucija - definicija

Distribucija neprekidne slučajne varijable je određena pripadnom vjerojatnosnom funkcijom gustoće. Tako kažemo da slučajna varijabla X ima **normalnu distribuciju** s parametrima $\mu \in \mathbb{R}$ i $\sigma^2 > 0$ ako je pripadna vjerojatnosna funkcija gustoće oblika

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

i koristimo oznaku $X \sim N(\mu; \sigma^2)$.

Normalna distribucija - definicija

Distribucija neprekidne slučajne varijable je određena pripadnom vjerojatnosnom funkcijom gustoće. Tako kažemo da slučajna varijabla X ima **normalnu distribuciju** s parametrima $\mu \in \mathbb{R}$ i $\sigma^2 > 0$ ako je pripadna vjerojatnosna funkcija gustoće oblika

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

i koristimo oznaku $X \sim N(\mu; \sigma^2)$. Parametri μ i σ^2 imaju direktnu interpretaciju:

$$E(X) = \mu; \quad \text{Var}(X) = \sigma^2:$$

Jedinčna normalna distribucija

Jedinčna ili **standardna** normalna distribucija je normalna distribucija s parametrima

$$\mu = 0; \quad \sigma = 1:$$

Jedinčna normalna distribucija

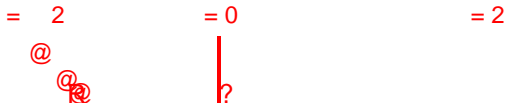
Jedinčna ili **standardna** normalna distribucija je normalna distribucija s parametrima

$$\mu = 0; \quad \sigma = 1:$$

Graf pripadne funkcije gustoće izgleda ovako:

Kako se mijenja graf funkcije gustoće ako promijenimo parametar očekivanja?

Kako se mijenja graf funkcije gustoće ako promijenimo parametar očekivanja?
 Promotrimo grafove funkcija gustoće za normalne distribucije s parametrima $\mu = 2; 0; 2$
 $\sigma = 1$:



Kako se mijenja graf funkcije gustoće ako promijenimo parametar varijance²?

Kako se mijenja graf funkcije gustoće ako promijenimo parametar varijance²?
Promotrimo grafove funkcija gustoća za normalne distribucije s parametrima $\mu = 0$ i $\sigma^2 = 0.5; 1; 2$.

— = 0.5

= 1

= 2

Aproksimacija binomne distribucije normalnom distribucijom

Podsjetimo se našeg motivacijskog primjera - za $X \sim B(n; p)$ definiramo pripadnu standardiziranu slučajnu varijablu

$$Y = \frac{X - E(X)}{\sqrt{E(X)}} = \frac{X - np}{\sqrt{npq}}$$

Aproksimacija binomne distribucije normalnom distribucijom

Podsjetimo se našeg motivacijskog primjera - za $X \sim B(n; p)$ definiramo pripadnu standardiziranu slučajnu varijablu

$$Y = \frac{X - E(X)}{\sqrt{E(X)}} = \frac{X - np}{\sqrt{npq}}$$

i promotrimo grafički prikaz distribucije od Y za $n = 10; 100; 500; 10000$:

Aproksimacija binomne distribucije normalnom distribucijom

Podsjetimo se našeg motivacijskog primjera - za $X \sim B(n; p)$ definiramo pripadnu standardiziranu slučajnu varijablu

$$Y = \frac{X - E(X)}{\sqrt{E(X)}} = \frac{X - np}{\sqrt{npq}}$$

i promotrimo grafički prikaz distribucije od Y za $n = 10; 100; 500; 10000$:

Uočimo da se distribucija ove standardizirane varijable stabilizira kada n je velik.

Aproksimacija binomne distribucije normalnom distribucijom

Podsjetimo se našeg motivacijskog primjera - za $X \sim B(n; p)$ definiramo pripadnu standardiziranu slučajnu varijablu

$$Y = \frac{X - E(X)}{\sqrt{E(X)}} = \frac{X - np}{\sqrt{npq}}$$

i promotrimo grafički prikaz distribucije od Y za $n = 10; 100; 500; 10000$:

Uočimo da se distribucija ove standardizirane varijable Y stabilizira kada n je velik.

Za veliki n je distribucija standardizirane varijable Y slična distribuciji standardne normalne slučajne varijable.

Aproksimacija binomne distribucije normalnom distribucijom

Podsjetimo se našeg motivacijskog primjera - za $X \sim B(n; p)$ definiramo pripadnu standardiziranu slučajnu varijablu

$$Y = \frac{X - E(X)}{\sqrt{\text{Var}(X)}} = \frac{X - np}{\sqrt{npq}}$$

i promotrimo grafički prikaz distribucije od Y za $n = 10; 100; 500; 10000$:

Uočimo da se distribucija ove standardizirane varijable Y stabilizira kada n je velik.

Za veliki n je distribucija standardizirane varijable Y slična distribuciji standardne normalne slučajne varijable. Izbor parametara $p = 0$ i $p = 1$ ima smisla jer je $E(Y) = 0$; $\text{Var}(Y) = 1$.

Aproksimacija binomne distribucije normalnom distribucijom

Ovu opservaciju možemo matematički formulirati na sljedeći način: Ako je $X \sim B(n; p)$ za $n \gg N$ velik onda je pripadna standardizirana slučajna varijabla približno normalno distribuirana s parametrima $\mu = np$ i $\sigma^2 = np(1-p)$,

Aproksimacija binomne distribucije normalnom distribucijom

Ovu opservaciju možemo matematički formulirati na sljedeći način: Ako je $X \sim B(n; p)$ za $n \gg N$ velik onda je pripadna standardizirana slučajna varijabla približno normalno distribuirana s parametrima $\mu = np$ i $\sigma^2 = npq$, odnosno

$$\frac{X - np}{\sqrt{npq}} \sim N(0; 1):$$

Ovaj je rezultat poznat pod nazivom de Moivre-Laplaceov centralni granični teorem .

Aproksimacija binomne distribucije normalnom distribucijom

Ovu opservaciju možemo matematički formulirati na sljedeći način: Ako je $X \sim B(n; p)$ za $n \gg N$ velik onda je pripadna standardizirana slučajna varijabla približno normalno distribuirana s parametrima $\mu = np$ i $\sigma^2 = npq$, odnosno

$$\frac{X - np}{\sqrt{npq}} \sim N(0; 1):$$

Ovaj je rezultat poznat pod nazivom de Moivre-Laplaceov centralni granični teorem.

Alternativno, možemo ga formulirati i kao:

$$X \sim N(np; npq)$$

i u ovom ga obliku najčešće i koristimo.

Funkcija distribucije standardne normalne razdiobe

Promotrimo sada kako računamo vjerojatnosti vezane uz normalne slučajne varijable. Za početak promotrimo $X \sim N(0; 1)$. Kako bismo odredili vjerojatnost da je X najviše 2?

Funkcija distribucije standardne normalne razdiobe

Promotrimo sada kako računamo vjerojatnosti vezane uz normalne slučajne varijable. Za početak promotrimo $X \sim N(0; 1)$. Kako bismo odredili vjerojatnost da je X najviše 2?

$$P(X \leq 2) = F_X(2) = \int_{-\infty}^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt:$$

Funkcija distribucije standardne normalne razdiobe

Promotrimo sada kako računamo vjerojatnosti vezane uz normalne slučajne varijable. Za početak promotrimo $X \sim N(0; 1)$. Kako bismo odredili vjerojatnost da je X najviše 2?

$$P(X \leq 2) = F_X(2) = \int_{-\infty}^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt:$$

Ovaj se integral općenito ne može dobiti u zatvorenoj formi, već se računa aproksimacijski.

Funkcija distribucije standardne normalne razdiobe

Promotrimo sada kako računamo vjerojatnosti vezane uz normalne slučajne varijable. Za početak promotrimo $X \sim N(0; 1)$. Kako bismo odredili vjerojatnost da je X najviše 2?

$$P(X \leq 2) = F_X(2) = \int_{-\infty}^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt:$$

Ovaj se integral općenito ne može dobiti u zatvorenoj formi, već se računa aproksimacijski. ! vrijednosti za F_X koji često označavamo s Φ su dane tablicno.

Funkcija distribucije standardne normalne razdiobe

Promotrimo sada kako računamo vjerojatnosti vezane uz normalne slučajne varijable. Za početak promotrimo $X \sim N(0; 1)$. Kako bismo odredili vjerojatnost da je X najviše 2?

$$P(X \leq 2) = F_X(2) = \int_{-\infty}^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt:$$

Ovaj se integral općenito ne može dobiti u zatvorenoj formi, već se računa aproksimacijski. ! vrijednosti za F_X koji često označavamo s Φ su dane tablicno.

! vrijednosti za F_X mogu se odrediti korištenjem statističkih alata (Excel, R, ...).

Funkcija distribucije standardne normalne razdiobe

Promotrimo sada kako računamo vjerojatnosti vezane uz normalne slučajne varijable. Za početak promotrimo $X \sim N(0; 1)$. Kako bismo odredili vjerojatnost da je X najviše 2?

$$P(X \leq 2) = F_X(2) = \int_{-\infty}^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt:$$

Ovaj se integral općenito ne može dobiti u zatvorenoj formi, već se računa aproksimacijski. ! vrijednosti za F_X koji često označavamo s Φ su dane tablično.

! vrijednosti za F_X mogu se odrediti korištenjem statističkih alata (Excel, R, ...).

Ako želimo odrediti $F_X(2)$ u Excelu pozivamo naredbu

$$= \text{NORM.DIST}(2, 0, 1, \text{TRUE}).$$

VAZNO! Treći parametar u ovoj naredbi je standardna devijacija, a ne varijanca ²!

Neka svojstva normalne razdiobe

Ako je X normalno distribuirana slučajna varijabla, onda je $aX + b$ također normalno distribuirana slučajna varijabla, za sve $a; b \in \mathbb{R}$, $a \neq 0$.

Neka svojstva normalne razdiobe

Ako je X normalno distribuirana slučajna varijabla, onda je $aX + b$ također normalno distribuirana slučajna varijabla, za sve $a; b \in \mathbb{R}$, $a \neq 0$.

Neka je $X \sim N(0; 1)$ i $Y = aX + b$. Tada je

$$Y \sim N(b; a^2):$$

Neka svojstva normalne razdiobe

Ako je X normalno distribuirana slučajna varijabla, onda je $aX + b$ također normalno distribuirana slučajna varijabla, za sve $a; b \in \mathbb{R}$, $a \neq 0$.

Neka je $X \sim N(0; 1)$ i $Y = aX + b$. Tada je

$$Y \sim N(b; a^2):$$

Korštenjem ovog rezultata dobivamo da za $X \sim N(\mu; \sigma^2)$ vrijedi

$$\frac{X - \mu}{\sigma} \sim N(0; 1):$$

Racunanje vjerojatnosti vezanih uz normalnu distribuciju - nastavak

Primjer. Ako slučajna varijabla X ima normalnu razdiobu $N(3; 4)$, kolika je vjerojatnost da će vrijednost slučajne varijable biti manja od 5?

Računanje vjerojatnosti vezanih uz normalnu distribuciju - nastavak

Primjer. Ako slučajna varijabla X ima normalnu razdiobu $N(3; 4)$, kolika je vjerojatnost da će vrijednost slučajne varijable biti manja od 5?

Rješenje. $\mu = 3$; $\sigma^2 = 4$; $\sigma = 2$

Računanje vjerojatnosti vezanih uz normalnu distribuciju - nastavak

Primjer. Ako slučajna varijabla X ima normalnu razdiobu $N(3; 4)$, kolika je vjerojatnost da će vrijednost slučajne varijable biti manja od 5?

Rješenje. $\mu = 3$; $\sigma^2 = 4$; $\sigma = 2$

$$P(X < 5) = ?$$

Računanje vjerojatnosti vezanih uz normalnu distribuciju - nastavak

Primjer. Ako slučajna varijabla X ima normalnu razdiobu $N(3; 4)$, kolika je vjerojatnost da će vrijednost slučajne varijable biti manja od 5?

Rješenje. $\mu = 3$; $\sigma^2 = 4$; $\sigma = 2$

$$P(X < 5) = ?$$

= **NORM.DIST(5, 3, 2, TRUE).**

Računanje vjerojatnosti vezanih uz normalnu distribuciju - nastavak

Primjer. Ako slučajna varijabla X ima normalnu razdiobu $N(3; 4)$, kolika je vjerojatnost da će vrijednost slučajne varijable biti manja od 5?

Rješenje. $\mu = 3$; $\sigma^2 = 4$; $\sigma = 2$

$$P(X \leq 5) = ?$$

$= \text{NORM.DIST}(5, 3, 2, \text{TRUE})$.

0.8413447

Računanje vjerojatnosti vezanih uz normalnu distribuciju - nastavak

Kako izračunati $P(a < X < b)$?

Računanje vjerojatnosti vezanih uz normalnu distribuciju - nastavak

Kako izračunati $P(a < X < b)$?

Računanje vjerojatnosti vezanih uz normalnu distribuciju - nastavak

Kako izračunati $P(a < X < b)$?

$$P(a < X < b) = P(X < b) - P(X < a) = F_X(b) - F_X(a):$$

Računanje vjerojatnosti vezanih uz normalnu distribuciju - nastavak

Kako izračunati $P(a < X < b)$?

$$P(a < X < b) = P(X < b) - P(X < a) = F_X(b) - F_X(a):$$

Napomena. Uočite da je kod neprekidnih distribucija $P(X = a) = 0$ za svaki a .

Primjer. Kolika je vjerojatnost da će slučajna varijabla s normalnom razdiobom $N(0; 1)$ poprimiti vrijednost između -1 i 2?

Primjer. Kolika je vjerojatnost da će slučajna varijabla s normalnom razdiobom $N(0; 1)$ poprimiti vrijednost između -1 i 2?

Rješenje. $\Phi(2) - \Phi(-1) = 0.9772 - 0.2420 = 0.7352$

Primjer. Kolika je vjerojatnost da će slučajna varijabla s normalnom razdiobom $N(0; 1)$ poprimiti vrijednost između -1 i 2?

Rješenje. $\Phi(0) = 0.5$; $\Phi(2) = 0.9772499$; $\Phi(-1) = 0.2420381$

$$P(-1 \leq X \leq 2) = P(X \leq 2) - P(X \leq -1) = F_X(2) - F_X(-1):$$

$\Rightarrow = \text{NORM.DIST}(2, 0, 1, \text{TRUE}) - \text{NORM.DIST}(-1, 0, 1, \text{TRUE})$

0.9772499 - 0.2420381 = 0.7352118

Primjer. Kolika je vjerojatnost da će slučajna varijabla s normalnom razdiobom $N(0; 1)$ poprimiti vrijednost između -1 i 2?

Rješenje. $\mu = 0$; $\sigma^2 = 1$; $\sigma = 1$

$$P(-1 \leq X \leq 2) = P(X \leq 2) - P(X \leq -1) = F_X(2) - F_X(-1):$$

$> = \text{NORM.DIST}(2, 0, 1, \text{TRUE})$

0.9772499

$> = \text{NORM.DIST}(-1, 0, 1, \text{TRUE})$

0.1586553

Primjer. Kolika je vjerojatnost da će slučajna varijabla s normalnom razdiobom $N(0; 1)$ poprimiti vrijednost između -1 i 2?

Rješenje. $\Phi(2) = 0.97725$; $\Phi(-1) = 0.24205$; $\Phi(2) - \Phi(-1) = 0.7352$

$$P(-1 \leq X \leq 2) = P(X \leq 2) - P(X \leq -1) = F_X(2) - F_X(-1):$$

$> = \text{NORM.DIST}(2, 0, 1, \text{TRUE})$

0.9772499

$> = \text{NORM.DIST}(-1, 0, 1, \text{TRUE})$

0.2420533

$$\begin{aligned} P(-1 \leq X \leq 2) &= P(X \leq 2) - P(X \leq -1) = \\ &= 0.977250 - 0.242053 = 0.735197 \end{aligned}$$

Aproksimacija binomne distribucije normalnom - nastavak

Kako koristimo ranije spomenuti rezultat gdje za velika ($np > 5$ i $nq > 5$) distribuciju binomne slučajne varijable $X \sim B(n; p)$ možemo aproksimirati normalnom distribucijom,

$$X \sim N(np; npq)?$$

Aproksimacija binomne distribucije normalnom - nastavak

Kako koristimo ranije spomenuti rezultat gdje za velika n ($np > 5$ i $nq > 5$) distribuciju binomne slučajne varijable $X \sim B(n; p)$ možemo aproksimirati normalnom distribucijom,

$$X \sim N(np; npq)?$$

Neka je $F = F_{N(np; npq)}$ funkcija distribucije normalne razdiobe $N(np; npq)$. Tada

$$P(X \leq b) \approx F(b) \text{ ili } F(b + \frac{1}{2});$$

$$P(a \leq X \leq b) \approx F(b) - F(a) \text{ ili } F(b + \frac{1}{2}) - F(a + \frac{1}{2});$$

gdje je crveni izraz **bez korekcije** a plavi izraz **korekcijom** po neprekidnosti. Obe aproksimacije su korisne, korekcija daje nešto bolju aproksimaciju uzimajući u obzir činjenicu da je binomna razdioba diskretna a normalna razdioba neprekidna.

Aproksimacija binomne distribucije normalnom - nastavak

Kako koristimo ranije spomenuti rezultat gdje za velike n ($np > 5$ i $nq > 5$) distribuciju binomne slučajne varijable $X \sim B(n; p)$ možemo aproksimirati normalnom distribucijom,

$$X \sim N(np; npq)?$$

Neka je $F = F_{N(np; npq)}$ funkcija distribucije normalne razdiobe $N(np; npq)$. Tada

$$P(X \leq b) \quad F(b) \text{ ili } F(b + \frac{1}{2});$$

$$P(a \leq X \leq b) \quad F(b) - F(a) \text{ ili } F(b + \frac{1}{2}) - F(a + \frac{1}{2});$$

gdje je crveni izraz *bez korekcije* a plavi izraz *s korekcijom* po neprekidnosti.

Aproksimacija binomne distribucije normalnom - nastavak

Kako koristimo ranije spomenuti rezultat gdje za velike n ($np > 5$ i $nq > 5$) distribuciju binomne slučajne varijable $X \sim B(n; p)$ možemo aproksimirati normalnom distribucijom,

$$X \sim N(np; npq)?$$

Neka je $F = F_{N(np; npq)}$ funkcija distribucije normalne razdiobe $N(np; npq)$. Tada

$$P(X \leq b) \quad F(b) \text{ ili } F(b + \frac{1}{2});$$

$$P(a < X < b) \quad F(b) - F(a) \text{ ili } F(b + \frac{1}{2}) - F(a + \frac{1}{2});$$

gdje je crveni izraz *bez korekcije* a plavi izraz *s korekcijom* po neprekidnosti. Obe aproksimacije su korisne, korekcija daje nešto bolju aproksimaciju uzimajući u obzir činjenicu da je binomna razdioba diskretna a normalna razdioba neprekidna.