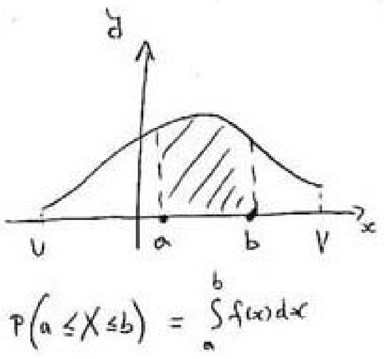
5. lekcija. Kontinuirane slučajne varijable.

Diskretne slučajne varijable povezane su s prebrojavanjem u nekom pokusu. One primaju konačan skup vrijednosti (ili možda beskonačan, ali je tada nužno prebrojiv i diskretan). Kod tih slučajnih varijabla možemo zamišljati da je jedinična masa (ukupna vjerojatnost) raspoređena na brojevnom pravcu u tih konačno točaka (odnosno prebrojivo mnogo točaka). Međutim, u pokusima se prirodno javlja i **mjerenje**. Skup brojeva kojima se (teoretski) zapisuju rezultati mjerenja nije ni konačan niti diskretan, već neki interval u skupu realnih brojeva . Tipični primjeri su III i IV iz 4. lekcije.
Ako slučajna varijabla X nešto mjeri, onda ona načelno može postići svaki broj ili svaki broj u nekom intervalu (u svakom slučaju neprebrojivo mnogo brojeva). Zato je intuitivno jasno da će vjerojatnost postizanja bilo kojega konkretnog broja biti jednaka nuli, tj.
$p(X=c)=0$, za svaki konkretan broj c.
Takve slučajne varijable nazivat ćemo kontinuiranim ili neprekinutim . Zato imamo sljedeću, matematički ne baš preciznu definiciju.
Kažemo da je slučajna varijabla X neprekinuta (kontinuirana) ako prima svaku vrijednost na nekom intervalu (i svaku s vjerojatnošću 0).
Za kontinuiranu slučajnu varijablu X isprazno je pitanje o tome kolika je vjerojatnost da postigne neki konkretan broj. Pravo je pitanje o vjerojatnosti da X postigne rezultat u nekom intervalu, na primjer u intervalu $[x,x+\Delta x]$. Ako je Δx mali broj, intuitivno je jasno da će ta vjerojatnost biti približno proporcionalna duljini intervala Δx, a da će koeficijent proporcionalnosti biti promjenjiv, tj. ovisan o x. Drugim riječima
$p(x \le X \le x+ \Delta x) \approx f(x)\, \Delta x$, (*)
za neku funkciju f. Matematički je ispravnije reći da se funkcija f definira ovako:

$$f(x) = \lim_{\Delta x \to 0} \frac{p(x \le X \le x + \Delta x)}{\Delta x} \quad (**)$$

Desna strana u (*) ima značenje površine pravokutnika duljine Δx i visine $f(x)$. Iz definicije odredjenog integrala slijedi da se vjerojatnost da slučajna varijabla X poprimi vrijednost unutar intervala [a,b] dobije integriranjem. Točnije:

$$p(a \le X \le b) = \int_a^b f(x)dx \ .$$



Nekoliko komentara.

1. U našem je primjeru vjerojatnost bila raspoređena prema nekom pravilu na intervalu. To je mogao biti i otvoreni interval i poluinterval i beskonačni intervali i, konačno, najveći

interval: cijeli skup realnih brojeva. Od sada ćemo smatrati da je **uvijek tako,** tj. da je vjerojatnost na **cijelom skupu realnih brojeva** (tj. da je funkcija *f* zadana na čitavom skupu realnih brojeva; to napravimo tako da na onom dijelu gdje nema vjerojatnosti stavimo da je vrijednost funkcije *f* jednaka 0). Zato će od sada biti:

(i) $f(x) \geq 0$, za sve realne brojeve x (f je pozitivna funkcija)

(ii) $\int_{-\infty}^{\infty} f(x)dx = 1$ (površina ispod grafa funkcije f je 1).

Pri tom je vjerojatnost da slučajna varijabla X poprimi vrijednost u intervalu [a,b] :

$$p(a < X < b) = \int_a^b f(x)dx .$$

2. Funkciju f s ovim dvama svojstvima zovemo **funkcijom gustoće vjerojatnosti** kontinuirane slučajne varijable X.
   Dakle, svaka ovakva funkcija definira neku razdiobu vjerojatnosti kontinuirane slučajne varijable.

3. Vjerojatnost da slučajna varijabla *X,* koja ima funkciju gustoće *f,* poprimi vrijednost na intervalu [a,b], jednaka je površini ispod grafa funkcije *f* od x=a do x=b (integral).

4. Iako ima puno takvih funkcija *f,* nas će u praksi zanimati samo nekoliko takvih funkcija (odnosno nekoliko tipova ili klasa takvih funkcija, što ćemo vidjeti poslije).

5. Možemo zamišljati da *f* opisuje kako je jedinična masa razmazana po *x*-osi (na početku je to bilo samo po intervalu). Razumno je pretpostaviti da dok kist vučemo neprekinuto, onda će i pripadna funkcija koja opisuje gustoću namaza biti neprekinuta. Također, razumno je pretpostaviti da ćemo, možda nekoliko puta zastati, pa potom nastaviti. Na tim mjestima funkcija f će možda biti prekinuta jer ćemo nastaviti razmazivanje po drugom pravilu. Tako će funkcija f biti **po dijelovima neprekinuta**, a kako je ukupna masa koju razmazujemo konačna (jedinična), ona će imati integral na svakom podintervalu, što znači da na njima možemo računati vjerojatnosti.

6. Formule (i) i (ii) koje definiraju funkciju gustoće treba gledati u **analogiji s diskretnom slučajnom varijablom**. Tako je formula (i) kontinuirani analogon uvjeta:

(i)'. $p_i>0$, za sve i (vjerojatnosti su pozitivne; eventualno se može dopustiti da budu 0),
a formulu (ii) . kao analogon formule

(ii)' $\sum p_i = 1$ (ukupna vjerojatnost je 1).

Jasno je da se formula za vjerojatnost da slučajna varijabla *X* poprimi vrijednost na nekom intervalu može smatrati analogijom one za diskretnu varijablu (integral od donje do gornje granice zamjenjuje prijašnje zbrajanje vjerojatnosti). Poslije ćemo vidjeti da se ta analogija provlači i dalje: na očekivanje, varijancu i sl.

Primjer 1. (jednolika kontinuirana razdioba na segmentu [u,v]). Zamislimo da rezultati mjerenja mogu ravnopravno biti brojevi u intervalu [u,v] i da slučajna varijabla X registrira taj rezultat. Tada je vjerojatnost (jedinična masa) jednoliko raspoređena na intervalu [0,2]. **Matematički model** za ovakvu razdiobu vjerojatnosti je konstantna funkcija gustoće:

$f(x) := \frac{1}{v-u}$, za x u intervalu [u,v]; izvan tog intervala je f(x)=0.

Funkcija distribucije vjerojatnosti.

Viodimo da bi nam nam za računanje vjerojatnosti **dobro bilo** znati primitivnu funkciju funkcije *f*. Tako za računanje vjerojatnosti **ne bismo trebali integrirati**, već **samo oduzimati**. Kako znamo, primitivna je funkcija određena do na konstantu, što znači da kada znamo jednu, sve se ostale dobiju dodavanjem konstante. Od svih tih, **izabrat ćemo jednu** i dati joj **posebno ime**.

Definicija 1. **Funkcija distribucije vjerojatnosti** slučajne varijable *X* jest, prema definiciji, funkcija F: $R \to R$, zadana uvjetom:
F(x):=p(X<x)
(vrijednost te funkcije u realnom broju *x* jest vjerojatnost da ta slučajna varijabla poprimi rezultat manji od tog *x*).

Pokazat ćemo da je upravo ta funkcija **jedna od primitivnih funkcija** funkcije *f*. Prije toga navedimo nekoliko njenih svojstava koje proizlaze iz njene definicije.

1. očito je F rastuća (jer se povećanjem broja *x* ne može ukupna vjerojatnost do tog mjesta smanjiti),

2. očito je da F prima vrijednosti između 0 i 1 (jer su njene vrijednosti neke vjerojatnosti)

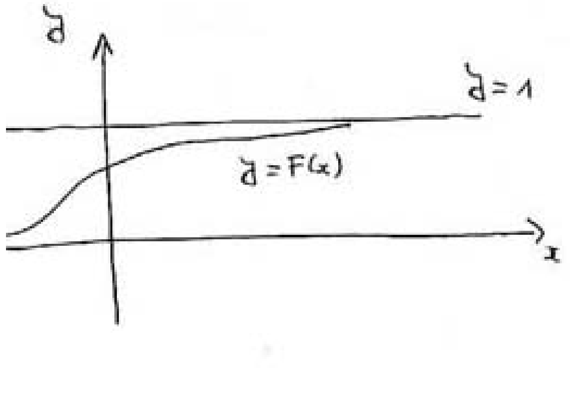
3. razumno je pretpostaviti da joj je limes kad *x* ide u *beskonačnost* upravo 1 (jer je ukupna vjerojatnost 1), a da joj je limes kad *x* ide u *–beskonačnost* jednaka 0.

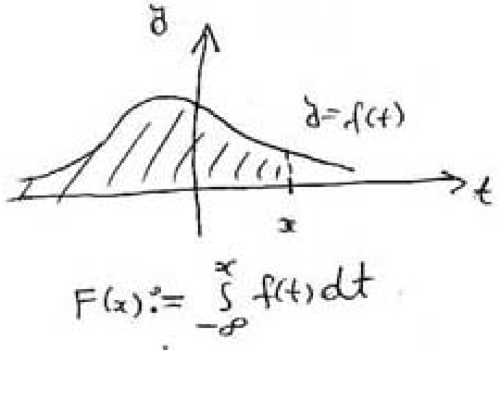
4. razumno je pretpostaviti i da je ta funkcija neprekinuta (jer malim promjenama vrijednosti *x*, malo se mijenja i vjerojatnost).

Sva ta svojstva izravno se dokazuju iz sljedeće veze:

$$F(x) = p(X < x) = p(-\infty < X < x) = \int_{-\infty}^{x} f(t)dt .$$

Tu smo ispod integrala stavili varijablu *t*, jer je *x* **gornja granica** integrala (pa da ne dođe do zabune).





Tako smo funkciju *F* opisali pomoću funkcije *f*. Obratno, iz definicije primitivne funkcije, sada slijedi da je:
F'(x)=f(x)
za sve *x* osim onih u kojima funkcija *f* ima prekid (a takvih je mjesta konačno mnogo; dopušta se i općenitija situacija, ali mi se ograničavamo na ovu).
Dakle *F* je primitivna funkcija funkcije *f* (odnosno to je barem po dijelovima).
Sada je.

$$p(a < X < b) = \int_a^b f(x)dx.$$
$$=F(b)-F(a).$$

Očekivanje i varijanca kontinuirane slučajne varijable.

Ove dvije važne karakteristike definiramo **prema uzoru na diskretan slučaj**. Tamo je razdioba vjerojatnosti bila u analogiji sa sustavom čestica na pravcu, očekivanje je odgovaralu težištu sustava čestica, a varijanca momentu inercije oko težišta. Sad umjesto sustava čestica imamo jediničnu masu raspoređenu po pravcu (skupu realnih brojeva), a način kako je ona raspoređena zadan je funkcijom gustoće *f* (koja se inače, obično u mehanici, označava slovom ρ). Dakle:

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx$$

(u usporedbi s diskretnim slučajem, suma se zamjenjuje integralom, x_i s x, a p_i s f(x)dx).

$$V(X) = \int_{-\infty}^{\infty} (x - E(X))^2 f(x)dx$$

U ovim formulama, a i inače kad je riječ o slučajnim varijablama, treba lučiti oznaku *X* (za slučajnu varijablu) od oznake *x* (za realnu varijablu tj. bilo koji realni broj).

Napomene:

1. Kao i prije, očekivanje ima značenje točke u kojoj je ravnoteža. Taj broj može biti i pozitivan i negativan.

2. Varijanca, kao i prije, mjeri stupanj rasipanja oko očekivanja. Iz formule se vidi da je uvijek pozitivna (jer integriramo funkciju koja je umnožak jedne funkcije koja je kvadrat i druge koja je pozitivna, prema definiciji, tj. računamo površinu ispod grafa funkcije koji je iznad osi *x*). Zato, kao i prije, možemo definirati **standardnu devijaciju** slučajne varijable X:
   s(X) = $\sqrt{V(X)}$.

3. Za računanje u konkretnim slučajevima pogodnija je druga formula za varijancu (a dobije se lako iz ove, kao i prije):
   $$V(X) = \int_{-\infty}^{\infty} x^2 f(x)dx - E^2(X).$$

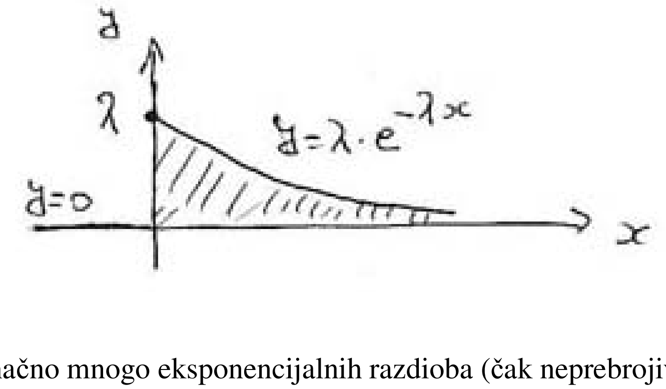
Tu, kao i obično, $E^2(X)$ znači $(E(X))^2$.

Eksponencijalna razdioba.

Vratimo se na primjer III iz 5. lekcije, tj. na opis slučajne varijable X koja opisuje vrijeme izmedju dviju uzastopnih poruka na nekoj adresi. Slično bi bilo da razmatramo vrijeme izmedju dvaju kvarova na nekom uredjaju (odnosno vrijeme rada bez kvara), vrijeme rada neke žarulje do pregorijevanja i sl.
Vrijeme izmedju dvaju kvarova na nekom uređaju ili stroju, nije egzaktno određeno (odnosno mi, sa sadašnjim znanjem, nismo sposobni to odrediti), već je slučajno. Slično je s vremenom izmedju dvaju posjeta neke adrese, poziva na nekom telefonu, izmedju dvaju radioaktivnih raspada, vremenom rada do kvara neke žarulje i sl. Zato ima smisla govoriti o slučajnoj varijabli koja registrira vrijeme izmedju dvaju kvarova, trajanja neke žarulje i sl..
Razmatranje nastavimo sa žaruljama.
Da bismo situaciju pojasnili možemo testirati mnogo takvih žarulja i tako doći do predodžbe o **prosječnom** vijeku trajanja takvih
žarulja, a i o distribuciji vremena trajanja (time se bavi **statistika** i o tome ćemo više govoriti poslije).
Neka je *X* slučajna varijabla koja registrira vrijeme trajanja žarulje. U idealnim uvjetima X će biti opisana samo s jednim parametrom – prosječnim vijekom trajanja. Nadalje, ako je f funkcija gustoće vjerojatnosti od X, onda je f(x)=0, za x<0 (naime, vrijeme trajanja ne može biti negativno). U proširenim lekcijama izveli smo da za funkciju gustoće f od X vrijedi.
$f(x)=\lambda e^{-\lambda x}$, za x>0.
Pri tom je parameter λ obrnuto proporcionalan očekivanju E(X) od X . To ćemo preciznije izvesti poslije.

To nas upućuje na sljedeću definiciju.

Definicija 2. Kažemo da slučajna varijabla *X* ima **eksponencijalnu** razdiobu s **parametrom** λ>0, ako joj je funkcija gustoće vjerojatnosti
$f(x)=\lambda e^{-\lambda x}$, za x>0
= 0, inače.
Ako je tako, pišemo X ~E(λ).



Dakle, ima beskonačno mnogo eksponencijalnih razdioba (čak neprebrojivo mnogo), ali sve su slične i parametrizirane su pozitivnim parametrom λ. Nabrojili smo više različitih (ali sličnih) pojava koje se približno ponašaju prema eksponencijalnom zakonu. Parametar λ razlikuje medjusobno te pojave (žarulje različitih kvaliteta, stranice različitih posjećenosti, različite radioaktivne materije i sl.).

Izvod očekivanja.
Ako je X~E(λ), onda je E(X) = 1/λ.
Naime:

$$E(X)=\int_0^\infty x\cdot\lambda e^{-\lambda x}dx=(-xe^{-\lambda x}-\frac{e^{-\lambda x}}{\lambda})\Big|_0^\infty=0-(-0-1/\lambda)=\frac{1}{\lambda}.$$

(pažljivo provjerite preskočene korake u gornjem računu; uputa: riječ je o parcijalnoj integraciji i L'Hospitalovu pravilu; gdje se u računu primjenjuje da je $\lambda>0$?).
Slično, ali tehnički nešto teže dobije se:

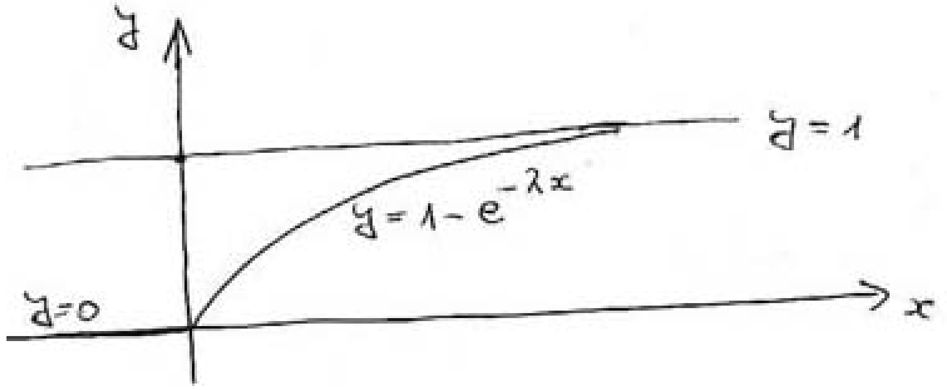
$$V(X)=\frac{1}{\lambda^2}.$$

Funkcija distribucije *F* eksponencijalne razdiobe. Vrijedi, F(x) =0 za negativne x, dok je za pozitivne x (i za nulu)

$$F(x)=\int_0^x\lambda e^{-\lambda t}dt=1-e^{-\lambda x}$$

Dakle, za pozitivne *x* je F(x) = 1 − $e^{-\lambda x}$.

Nacrtajmo sliku.



Vidimo da graf funkcije usporeno raste. To znači da se vjerojatnost dosta rano potroši (objasnite što to praktički znači).

Primjer 2. Prosječno vrijeme između dvaju poziva na nekoj adresi je 80 sekunda. Odredimo vjerojatnost:
a) da to vrijeme bude najmanje 80 s.
b) da do sljedeće poruke dođe za manje od 600 s.
c) da poruka stigne između 50-te i 100-te s.

Prije rješavanja pokušajte na ova pitanja odgovoriti odoka.
Da bismo riješili ovaj problem prihvatit ćemo predpostavku da je vrijeme između dviju poruka eksponencijalno distribuirano. Dakle, ako slučajna varijabla *X* registrira to vrijeme onda je X~E(1/80).
Tu smo izabrali λ=1/80 jer je prosječno vrijeme između poruka 80, a taj broj odgovara očekivanju slučajne varijable *X*, a ono je 1/λ.   Sad se problemi mogu shvatiti ovako:
a) $p(X\geq 80) = p(80<X<\infty) = 1-F(80) = 1-(1-e^{-80/80})=0.3679$ (na 4 decimalna mjesta)
b) $p(X<60) = p(-\infty<X<60) = F(60)-0 = 1-e^{-60/80}= 0.5276$
c) $p(50<X<100) = F(1\ 00)-F(50) = (1-e^{-100/80})-(1-e^{-50/80}) = 0.2488$.

Normalna (Gaussova) razdioba.

To je najvažnija razdioba u teoriji, ali i u primjeni vjerojatnosti. Tu razdiobu (približno) imaju mnoge slučajne varijable koje nastaju u praksi, primjerice slučajna varijabla koja

a) registrira grješku pri mjerenju
b) registrira rezultat mjerenja (na pr. mase, visine, postotka, inteligencije, …)
c) registrira rezultate pri dobro odmjerenom pismenom ispitu itd.

S druge strane, do ove razdiobe može se doći statistički (o čemu će više biti riječ poslije). Naime kad se izvodi veći broj nezavisnih mjerenja neke veličine, pa se gleda prosječan rezultat, dolazi se do normalne razdiobe (približno, a u limesu točno; to se može točno matematički opisati).

Razmotrimo nekoliko **heurističkih razloga** koji nam daju naslutiti da će se grješke pri mjerenju ponašati prema normalnom zakonu (poslije ćemo točno reći o kakvoj je razdiobi riječ).

Neka slučajna varijabla X registrira grješku pri mjerenju (koju ne možemo egzaktno opisati jer nastaje slučajno, čak i kod najpreciznijih instrumenata) i neka je f funkcija gustoće od X. Tada je prirodno, uz činjenicu da je površina ispod grafa od f jednaka 1, predpostaviti sljedeće:

1. grješke teoretski mogu biti po volji velike, pa je f definirana i pozitivna za sve realne brojeve x.

2. grješke su simetrično raspoređene oko 0, pa je f **parna funkcija**.

3. najvjerojatnije su grješke oko nule, a kako idemo dalje vjerojatnost da greška bude u nekom malom intervalu stalne duljine, smanjuje se. Zato je f padajuća funkcija za x>0.

4. Trebala bi postojati familija takvih f, ovisnih o nekom parametru, koje bi opisivale grješke (jer jedna je funkcija za instrument koji radi male, a druga za neki koji radi velike grješke; ipak te funkcije trebaju biti slične).

Iz ovih, vrlo razumnih zahtijeva, naslućujemo da je pripadna funkcija gustoće vjerojatnosti **zvonolika krivulja**. Koristeći poznatu činjenicu da je:

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\,dx = 1 \quad (*)$$

dolazimo do toga da se f može zapisati kao:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}, \qquad \sigma > 0 \quad (**)$$

(izvod je u proširenoj verziji lekcija).

Graf te funkcije zove se **gaussova krivulja**; ona je

1. **zvonolika** oblika s tjemenom u x=0.
2. simetrična s obzirom na y-os,
3. ima točke infleksije u $x = \pm\sigma$.
4. Odoka vidimo da je površina između tih dviju vrijednosti veća od 1/2, poslije ćemo to izračunati preciznije.
   To ćemo približiti na primjeru.

Pretpostavimo da je grješka pri očitavanju rezultata na nekom mjernom istrumentu normalno distribuirana uz σ =0.001. Tu činjenicu možemo tumačiti kao: vjerojatnost da grješka bude između –0.001 i 0.001 je veća od 1/2 (poslije ćemo vidjeti da je ona oko 2/3). To znači da je vjerojatnost da grješka bude veća od 0.001 ili manja od –0.001, manja od 1/2 (poslije ćemo vidjeti da je ona oko 1/3).

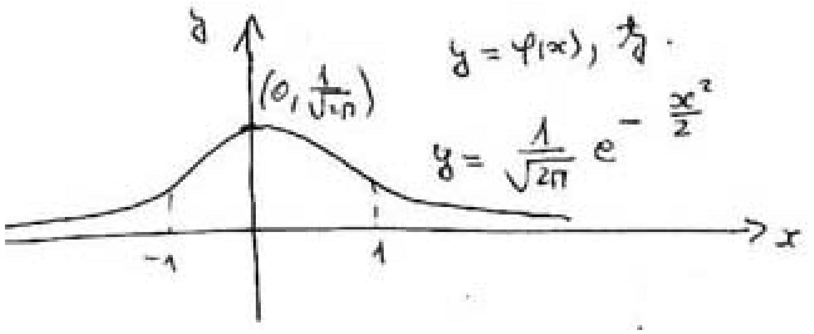
5. Što je σ veći, krivulja je spljoštenija (površina je rasturenija); to govori o tome da su grješke raspršenije, dakle veće. Obratno, što je σ manje, krivulja je uža i viša; znači grješke su manje, koncentrirane su oko 0.
6. Za slučajnu varijablu X koja ima ovakvu funkciju gustoće kažemo da je **normalna** i pišemo: X~N(0, σ^2).

   Specijalno, ako je σ =1, kažemo da je razdioba **jedinična normalna** (ili **standardna**) i tada pišemo:

   X~N(0,1^2).

   Često jediničnu razdiobu označavamo s T, a njenu funkciju gustoće s φ. Dakle:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$



Da je sa (**) zadana funkcija gustoće (tj. da je površina ispod krivulje 1) proizlazi iz (*), samo se kod integriranja primjeni očita zamjena varijable.
Također je očito da sve ove razdiobe imaju očekivanje 0 (zbog simetričnosti); tko hoće to može i dokazati (izravno ili korištenjem svojstva neparnosti).
Lako se može dokazati (parcijalna integracija, zamjena varijable i formula (*)) da je varijanca ove razdiobe σ^2.

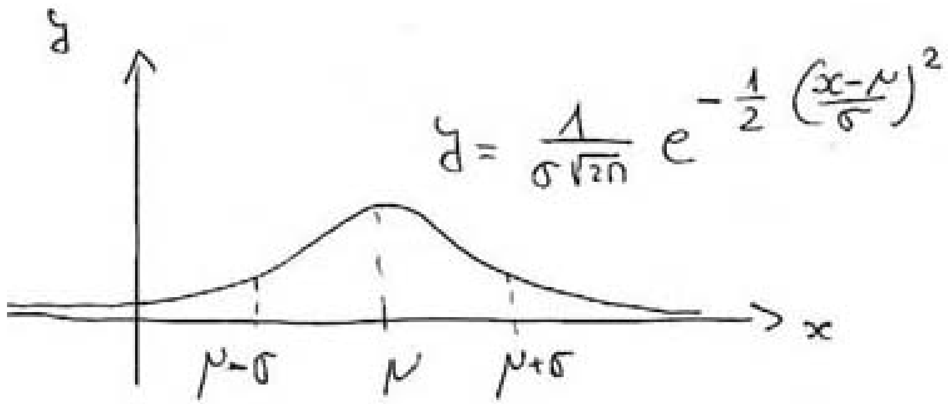
Dakle:

Ako je X~N(0, σ^2), onda je
E(X) = 0, i
V(X) = σ^2.

Napomena.

1. Ako graf funkcije f pomaknemo za μ onda se tjeme pomakne za taj isti μ (pomak će biti udesno ako je μ>0, u suprotnom bit će ulijevo). Pomicanjem za μ funkcija postaje:

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



2. Ovako pomaknuta funkcija također je funkcija gustoće (pomakom se površina ne mijenja). Slučajnu varijablu *X* koja ima takvu funkciju gustoće pišemo kao:
   X~N(μ,σ^2).
   Pritom se očekivanje također pomakne za μ pa je.
   E(X) = μ.
   Varijanca se ne mijenja (krivulja po izgledu ostaje potpuno ista) pa je:
   V(X) = σ^2.
   Sve se to može potvrditi i računom.
3. Pomaknuta krivulja ima slična svojstva kao i ona početna, samo što je os simetrije pravac s jednadžbom x=μ (a ne više x-os).

   Sada imamo konačnu definiciju.
   Definicija 3. Kažemo da kontinuirana slučajna varijabla *X* ima normalnu razdiobu s parametrima μ i σ^2, ako joj je funkcija gustoće:
   $$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$
   Pri tom je parametar μ očekivanje, a parametar σ^2 varijanca te slučajne varijable.

Računanje vjerojatnosti kod normalne razdiobe.

Već smo vidjeli da se vjerojatnost kod kontinuirane razdiobe najlakše računa ako nam je poznata funkcija distribucije (tada se vjerojatnost da slučajna varijabla poprimi rezultat na nekom intervalu svodi na oduzimanje vrijednosti te funkcije u rubovima tog intervala).
U slučaju normalne razdiobe **ne može se funkcija distribucije zapisati** pomoću poznatih elementarnih funkcija (već samo pomoću beskonačnih redova potencija ili nekih drugih funkcija čije se vrijednosti ne mogu lako računati, na primjer, nisu dostupne na svakom kalkulatoru; jasno je da se, koristeći na primjer Simphsonovu metodu, može napraviti program koji će po volji precizno računati vrijednosti te funkcije).
Zato su prije razvoja kompjutora vrijednosti funkcije distribucije normalne razdiobe bile **tabelirane**. Tu je bila sreća da **nije trebalo tabelirati vrijednosti za sve** normalne razdiobe (kojih ima neprebrojivo mnogo i ovisni su o dvama spomenutim parametrima), već je bilo

dovoljno tablicu načiniti za jediničnu normalnu razdiobu, a vjerojatnost u općem slučaju svodila se na jedinični. Pokažimo to.

Neka je X~N(μ,σ^2), i neka su *a, b* realni brojevi takvi da je *a*<*b*. Tada je:

$$p(a<X<b) = \int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \quad \text{(nakon zamjene varijable } \frac{x-\mu}{\sigma} = t\text{)}$$

$$= \int_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \quad \text{(ispod integrala je funkcija } \varphi\text{)}$$

$$= Z(\frac{b-\mu}{\sigma}) - Z(\frac{a-\mu}{\sigma}),$$

gdje smo sa Z označili funkciju distribucije jedinične normalne razdiobe.

Dakle, računanje vjerojatnosti po volji odabrane normalne razdiobe, svodi se na računanje vjerojatnosti jedinične normalne razdiobe tj. do oduzimanja vrijednosti funkcije Z. Ovaj smo rezultat izveli pomoću zamjene varijable u određenom integralu, međutim on ima i druga objašnjenja, na primjer geometrijska (lako se vidi da bilo koja normalna razdioba pomakom suprotno od očekivanja, potom dijeljenjem sa standardnom devijacijom postaje jedinična normalna razdioba; to grafovi funkcija gustoće lijepo prate)

Danas se vrijednosti funkcije distribucije normalne razdiobe lako računaju u mnogim kompjutorskim paketima.

Primjer 9. Težina neke populacije normalno je distribuirana s prosječnom težinom 66 kg i standardnom devijacijom 5kg. Odredimo vjerojatnost da slučajno odabrani predstavnik te populacije ima:

a) težinu između 65 i 70 kilograma.
b) težinu veću od 72 kg.

Neka X označava slučajnu varijablu koja registrira težinu populacije. Tada je

X~N(66,5^2). Neka je F pripadna funkcija distrubicije.

Zadatak a) može se zapisati kao

p(65<X<70) = F(70)-F(65) =0.78 815-0.42 074 = 0.36 741 (prema Excelu)

Takodjer, za zadatak b):

p(X>70)=p(70<X<+∞)=F(+∞)-F(70)=1- 0.78 815 =0.21 185.

Pravilo "tri sigme".

Najvažnije svojstvo normalne razdiobe u primjenama, jest to da interval unutar kojega se nalazi gotovo 100% svih vrijednosti slučajne varijable, **ovisi samo o očekivanju i standardnoj devijaciji** (to je za po 3 standardne devijacije lijevo i desno od očekivanja).

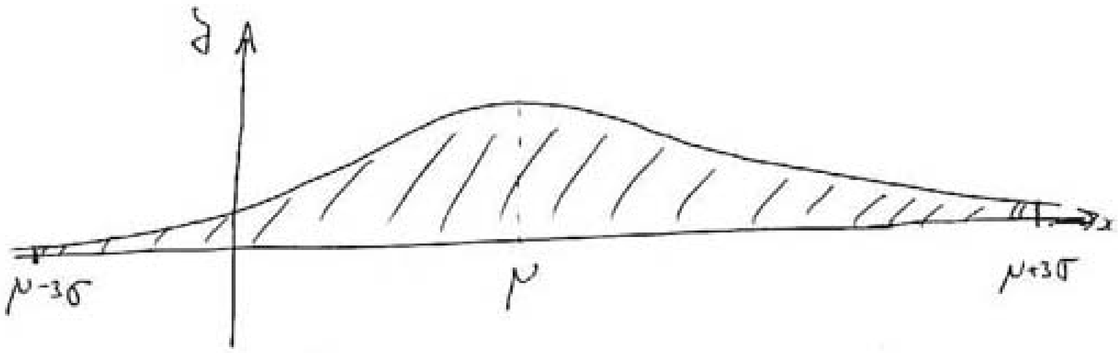
Preciznije, neka je X~N(μ,σ^2). Tada je

$p(\mu - 3\sigma < X < \mu + 3\sigma) = 0.9973$

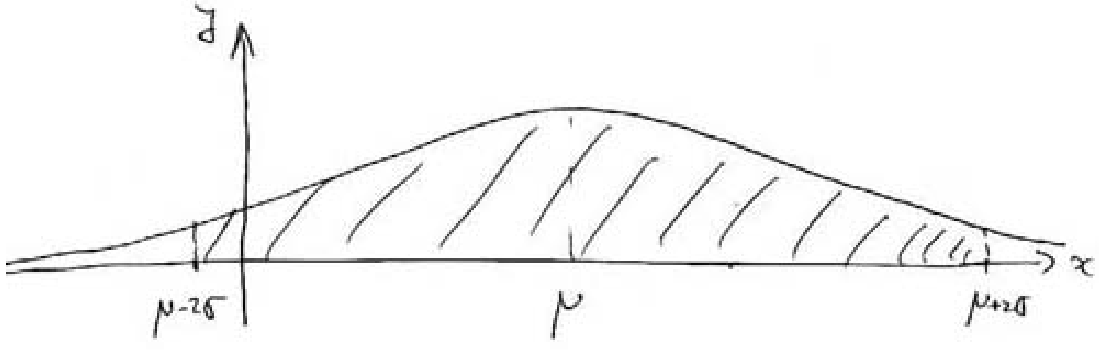
Također je.

$p(\mu - 2\sigma < X < \mu + 2\sigma) = 0.9544$

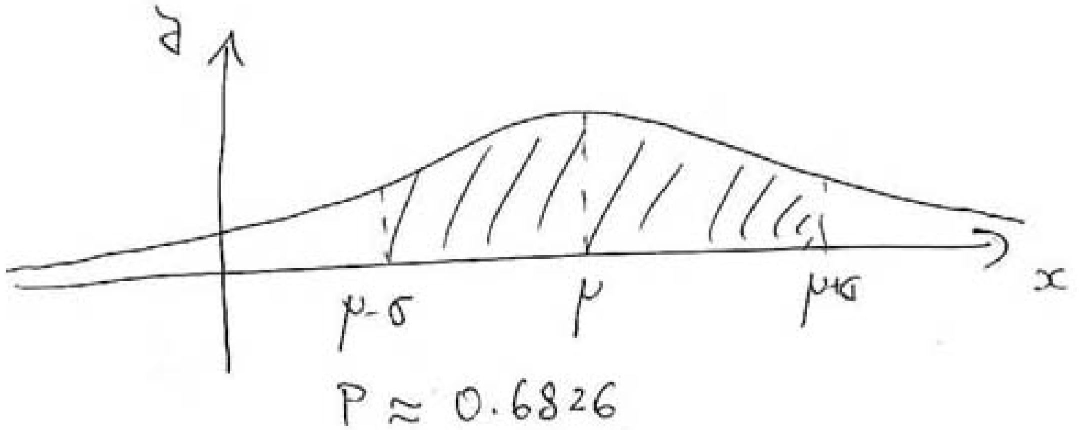
$p(\mu - \sigma < X < \mu + \sigma) = 0.6826$



$P \approx 0.9973$



$P \approx 0.9544$



$P \approx 0.6826$

Znači s 95% sigurnošću možemo tvrditi da se neka veličina normalno distribuirana nalazi za najviše 2 standardne devijacije lijevo ili desno od očekivanja (to znači da je u tom intervalu više od 95% površine ispod pripadne krivulje.
Slično, vidimo da za 1 standardnu devijaciju lijevo i desno od oćekivanja ima oko 2/3 površine ispod te krivulje (to je ono što smo ranije odoka procjenjivali da je više od jedne polovine).

Primjer 10. Odredimo interval unutar kojeg se nalazi težina slučajno odabrane osobe iz Primjera 9:

(i) s vjerojatnošću oko 0.68
(ii) s vjerojatnošću oko 0.95
(iii) gotovo 100%

U Primjeru 9 je riječ o normalnoj razdiobi s μ=66 i σ=5, pa je, prema pravilima "jedan sigma", "dva sigma" i "tri sigma":

(i) [66-5,66+5]=[61,71], tj.
p(težina slučajno izabrane osobe je između 61 i 71) $\approx$ 0.68
(ii) $[66-2\cdot 5,\ 66+2\cdot 5]$=[56, 76], tj.
p(težina slučajno izabrane osobe je između 56 i 76) $\approx$ 0.95
(ii) $[66-3\cdot 5,\ 66+3\cdot 5]$=[51, 81], tj.
gotovo sve osobe imaju težinu izmedju 51 i 81.