Statističke i numeričke metode

Materijali za seminare iz kolegija:

Statističke i numeričke metode,

Numeričke i statističke metode,

Osnove statistike okoliša i numeričke metode.

Erna Begović Kovač

Miroslav Jerković

Zavod za matematiku

Fakultet kemijskog inženjerstva i tehnologije

Sveučilište u Zagrebu

Sadržaj

1 Deskriptivna statistika **1**
1.1 Tablica frekvencija i histogram . . . 3

2 Slučajne varijable **5**
2.1 Diskretna slučajna varijabla . . . 5
2.1.1 Binomna razdioba . . . 7
2.1.2 Poissonova razdioba . . . 8
2.2 Kontinuirana slučajna varijabla . . . 9
2.2.1 Normalna razdioba . . . 10
2.2.2 Eksponencijalna razdioba . . . 11

3 Testovi hipoteza **14**
3.1 Interval pouzdanosti za očekivanje . . . 14
3.2 Testovi hipoteza . . . 14
3.3 Naredba IF u Excelu . . . 17

4 Metoda najmanjih kvadrata **19**

5 Interpolacija i ekstrapolacija **22**
5.1 Interpolacijski polinom . . . 22

6 Jednadžbe i sustavi **24**
6.1 Metoda bisekcije . . . 25
6.2 Metoda sekante . . . 26
6.3 Metoda tangente . . . 26
6.4 Solver . . . 27
6.5 Sustavi jednadžbi . . . 28

7 Obične diferencijalne jednadžbe **30**
7.1 Eulerova metoda . . . 30
7.2 Runge-Kutta 4 metoda . . . 31

Poglavlje 1

Deskriptivna statistika

Primjeri u Excelu vezani za ovu cjelinu nalaze se u dokumentu *Deskriptivnastatistika.xlsx*.

Deskriptivna statistika je dio matematičke statistike koji se koristi za opisivanje i bolje razumijevanje izmjerenog (ili zadanog) skupa podataka. Upoznat ćemo se s osnovnim pojmovima deskriptivne statistike.

- **Duljina uzorka** je broj podataka.
- **Najmanji podatak**
- **Najveći podatak**
- **Raspon** je razlika najvećeg i najmanjeg podatka.
- **Aritmetička sredina** (prosjek)
- **Medijan** je srednji podatak. Pola podataka nalazi se iznad, a pola ispod medijana.
- **Prvi kvartil** (donji kvartil) je broj od kojeg je manje ili jednako 25% podataka.
- **Drugi kvartil** je broj od kojeg je manje ili jednako 50% podataka. Drugi kvartil je isto što i medijan.
- **Treći kvartil** (gornji kvartil) je broj od kojeg je manje ili jednako 75% podataka.
- **Mod** uzorka je podatak koji se u uzorku pojavljuje najviše puta.

Napomenimo da je nulti kvartil zapravo najmanji podatak, a četvrti kvartil je najveći podatak. Kvartili dijele skup podataka na četiri dijela. Analogno bi mogli podijeliti podatke i na neki drugi broj dijelova. Podjela na 100 dijelova je podjela na percentile. Onda je primjerice sedmi percentil broj od kojeg je manje ili jednako 7% podataka.

U donjoj tablici navedene su Excel formule koje se koriste za računanje navedenih pojmova. Umjesto riječi "uzorak" u desnom stupcu upisuje se raspon polja u kojima se nalaze podatci, npr. MIN(A1:A50).

POJAM	NAREDBA
duljina uzorka	COUNT(uzorak)
najmanji podatak	MIN(uzorak)
najveći podatak	MAX(uzorak)
raspon	MAX(uzorak) - MIN(uzorak)
aritmetička sredina	AVERAGE(uzorak)
medijan	MEDIAN(uzorak)
1. kvartil	QUARTILE(uzorak; 1)
3. kvartil	QUARTILE(uzorak; 3)
mod	MODE(uzorak)
7. percentil	PERCENTILE(uzorak; 0.07)

Za bolju interpretaciju podataka bitno je znati koliko su podatci raspršeni, odnosno koliko odstupaju od prosjeka. Pretpostavimo da skup podataka ima duljinu n i označimo njegovu aritmetičku sredinu s $\bar{x}$. **Suma apsolutnih odstupanja** podataka od aritmetičke sredine (SAO) definira se kao

$$\mathrm{SAO} = \sum_{i=1}^{n} |x_i - \bar{x}| = |x_1 - \bar{x}| + |x_2 - \bar{x}| + \cdots + |x_n - \bar{x}|.$$

Prosječno apsolutno odstupanje od aritmetičke sredine (PAO) dobije se tako što se suma apsolutnih odstupanja podijeli s brojem podataka,

$$\mathrm{PAO} = \frac{\mathrm{SAO}}{n}.$$

Umjesto da gledamo apsolutno odstupanje od prosjeka, možemo gledati kvadratno odstupanje. Prosječno kvadratno odstupanje od aritmetičke sredine naziva se **varijanca** uzorka ($(s')^2$) i definirana je sa

$$(s')^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n}.$$

Standardna devijacija (s') je korijen iz varijance. **Korigirana varijanca** uzorka (s^2) je veličina

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{(x_1-\bar{x})^2 + (x_2-\bar{x})^2 + \cdots + (x_n-\bar{x})^2}{n-1},$$

a **korigirana standardna devijacija** (s) je korijen iz korigirane varijance. Za ove ćemo pojmove također navesti odgovarajuće Excel formule.

POJAM	NAREDBA
prosječno apsolutno odstupanje	AVEDEV(uzorak)
suma apsolutnih odstupanja	AVEDEV(uzorak)·COUNT(uzorak)
varijanca	VARP(uzorak)
standardna devijacija	STDEVP(uzorak)
korigirana varijanca	VAR(uzorak)
korigirana standardna devijacija	STDEV(uzorak)

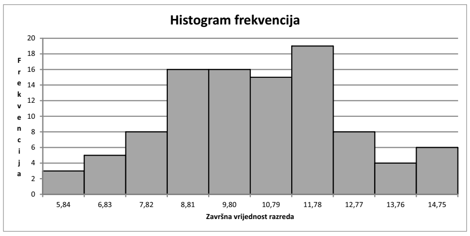
1.1 Tablica frekvencija i histogram

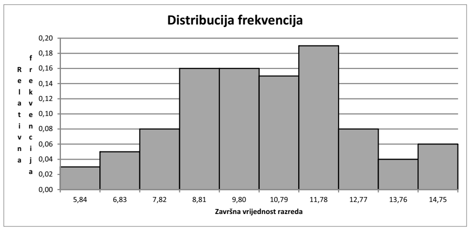
Zbog bolje interpretacije podataka, često je korisno podijeliti veliki skup podataka na podskupove koji se nazivaju **razredi**. Podjelu na n razreda radimo tako da interval od najmanjeg do najvećeg podatka podijelimo na n dijelova. Duljina svakog od tih n podintervala naziva se **širina razreda**, a dobije se kao kvocijent raspona i broja razreda n. (Svi su razredi jednake širine.)

Bitno je dobiti granice svakog razreda, a potom je lako rasporediti podatke po razredima. Početna vrijednost prvog razreda je najmanji podatak, a završna vrijednost jednaka je zbroju početne vrijednosti i širine razreda. Završna vrijednost prvog razreda ujedno je i početna vrijednost drugog razreda. Općenito, osim za prvi razred, početna vrijednost nekog razreda jednaka je završnoj vrijednosti prethodnog razreda. Završna vrijednost nekog razreda jednaka je zbroju njegove početne vrijednosti i širine razreda. Završna vrijednost zadnjeg razreda uvijek je jednaka najvećem podatku.

Na primjer, neka je zadan skup podataka među kojima je najmanji jednak 4, a najveći 14. Zadatak je podijeliti taj skup na 10 razreda. Tada je širina razreda jednaka $\frac{14-4}{10} = 1$. Prvi se razred proteže od 4 (što je najmanji podatak) do 5 ($4 + 1 = 5$), drugi razred od 5 do 6 ($5 + 1 = 6$), itd. Konačno, zadnji razred obuhvaća podatke od 13 do 14.

Frekvencija razreda je broj podataka u pojedinom razredu. U Excelu se ona dobije korištenjem naredbe





Slika 1.1: Graf frekvencija i graf relativnih frekvencija

FREQUENCY(uzorak; skup završnih vrijednosti razreda).

Relativna frekvencija razreda dobije se tako da se frekvencija razreda podijeli s ukupnim brojem podataka. Zbroj svih frekvencija razreda jednak je ukupnom broju podataka jer je svaki podatak ubrojen točno jednom i pripada samo jednom razredu. Zbroj svih relativnih frekvencija je 1.

RAZRED	ZAVRŠNA VRIJEDNOST	FREKVENCIJA	REL. FREKVENCIJA
1	5.84	3	0.03
2	6.83	5	0.05
3	7.82	8	0.08
4	8.81	16	0.16
5	9.80	16	0.16
6	10.79	15	0.15
7	11.78	19	0.19
8	12.77	8	0.08
9	13.76	4	0.04
10	14.75	6	0.06

Graf koji predočava frekvencije razreda naziva se **histogram frekvencija**. Crta se kao stupčasti graf gdje su na x-osi završne vrijednosti razreda, a na y-osi frekvencije. S druge strane, graf koji predočava relativne frekvencije naziva se **distribucija frekvencija**. To je također stupčasti graf, na x-osi su završne vrijednosti razreda, a na y-osi relativne frekvencije.

Poglavlje 2

Slučajne varijable

Primjeri u Excelu vezani za ovu cjelinu nalaze se u dokumentu *Slucajnevarijable.xlsx*.

Slučajna varijabla je funkcija koja svakom mogućem ishodu nekog pokusa pridružuje realni broj. Jednostavnijim riječima, vrijednost slučajne varijable možemo promatrati kao numerički ishod slučajnog eksperimenta. Glavna podjela slučajnih varijabli je na diskretne i kontinuirane. **Diskretna** slučajna varijabla je ona kojoj je skup vrijednosti konačan ili prebrojiv, a **kontinuirana** je ona kojoj je skup vrijednosti neprebrojiv. Za početak se bavimo diskretnim slučajnim varijablama.

2.1 Diskretna slučajna varijabla

Primjer 2.1
Bacamo tri novčića. (S jedne strane svakog novčića je pismo, s druge glava.) Slučajna varijabla X bilježi broj dobivenih pisama. Pogledajmo koje se sve mogućnosti mogu dogoditi.

Ishod	Broj pisama
GGG	0
GGP	1
GPG	1
PGG	1
GPP	2
PGP	2
PPG	2
PPP	3

Na svakom novčiću imamo dvije mogućnosti, a imamo tri novčića. Stoga je broj svih mogućih ishoda jednak $2 \cdot 2 \cdot 2 = 8$. Skup vrijednosti slučajne varijable X je $\{0, 1, 2, 3\}$. Uočimo da se u tablici neke vrijednosti pojavljuju više puta, tj. nekim je ishodima pridružena ista numerička vrijednost. Svakoj vrijednosti slučajne varijable, x_i, pridružena je njena vjerojatnost, $P(X = x_i)$. Ta je vjerojatnost jednaka

$$P(X = x_i) = \frac{\text{broj ishoda koji daju vrijednost } x_i}{\text{broj svih mogućih ishoda}}.$$

Nula pisama možemo dobiti na samo jedan način (GGG) pa je $P(X = 0) = \frac{1}{8}$. Jedno pismo možemo dobiti na tri načina (GGP, GPG, PGG) pa je $P(X = 1) = \frac{3}{8}$. Na isti se način dobiju vjerojatnosti za preostale vrijednosti slučajne varijable.

X	$P(X = x_i)$
0	$\frac{1}{8}$
1	$\frac{3}{8}$
2	$\frac{3}{8}$
3	$\frac{1}{8}$

Vjerojatnosti $P(X = x_i)$ su brojevi između 0 i 1. Možemo ih interpretirati i kao postotke. Ako će se nešto sigurno dogoditi, to ima vjerojatnost 1 (100%), a ako se sigurno neće dogoditi, ima vjerojatnost 0 (0%). Zbroj vjerojatnosti svih vrijednosti slučajne varijable je 1.

Očekivanje diskretne slučajne varijable X s vrijednostima $x_1, x_2, \ldots, x_n$ definira se kao

$$E(X) = \sum_{i=1}^{n} x_i P(x_i) = x_1 P(x_1) + x_2 P(x_2) + \cdots + x_n P(x_n),$$

a njena **varijanca** (disperzija) iznosi

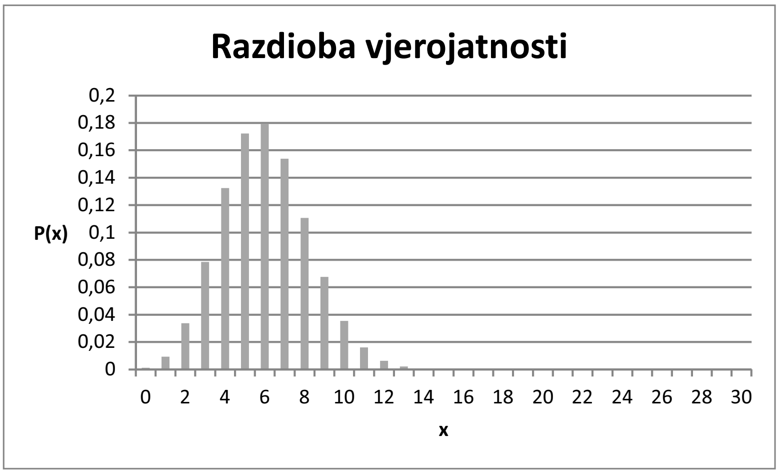
$$V(X) = \sum_{i=1}^{n} (x_i - E(X))^2 P(x_i) = (x_1 - E(X))^2 P(x_1) + \cdots + (x_n - E(X))^2 P(x_n).$$

U prethodnom primjeru očekivanje iznosi

$$E(X) = 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = 1.5,$$

a varijanca

$$V(X) = (0 - 1.5)^2 \cdot \frac{1}{8} + (1 - 1.5)^2 \cdot \frac{3}{8} + (2 - 1.5)^2 \cdot \frac{3}{8} + (3 - 1.5)^2 \cdot \frac{1}{8} = 0.75.$$



Slika 2.1: Graf razdiobe vjerojatnosti binomne slučajne varijable s parametrima $n = 30$, $p = 0.2$.

2.1.1 Binomna razdioba

Binomna slučajna varijabla je diskretna slučajna varijabla koja se generira tako što se n puta ponavlja isti pokus. Slučajna varijabla registrira koliko je puta pokus uspio, odnosno, koliko se puta dogodio neki fiksirani događaj A koji ima vjerojatnost p. Parametri n i p određuju binomnu slučajnu varijablu X i pišemo

$$X \sim B(n, p).$$

Kako ova slučajna varijaba registrira koliko se puta, od n pokušaja, dogodio A, lako je vidjeti da su njene vrijednosti cijeli brojevi između 0 i n. U najgorem se slučaju događaj A pojavio 0 puta, a u najboljem slučaju svih n puta.

Vjerojatnost svakog pojedinog ishoda i iznosi

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

U Excelu za računanje vjerojatnosti $P(X = i)$ koristimo naredbu

=BINOMDIST(i; n; p; FALSE).

Pomoću iste naredbe možemo računati i vjerojatnost da je slučajna varijabla manja ili jednaka od neke vrijednosti i, $P(X \leq i)$,

=BINOMDIST(i; n; p; TRUE).

Očekivanje binomne slučajne varijable X dano je formulom

$$E(X) = np,$$

a varijanca

$$V(X) = np(1-p).$$

Primjer 2.2
Bacamo kocku 20 puta. (Brojevi na kocki su od 1 do 6.) Slučajna varijabla X bilježi broj dobivenih jedinica. Ovo je binomna slučajna varijabla. Odredimo njene parametre. S obzirom da se pokus ponavlja 20 puta, $n = 20$. Vjerojatnost da dobijemo jedinicu u jednom bacanju je $\frac{1}{6}$ jer je jedinica jedna od ukupno 6 mogućnosti. Stoga je $p = \frac{1}{6}$, tj. $X \sim B(20, \frac{1}{6})$.

Primjer 2.3
Bacamo kockicu 5 puta. Slučajna varijabla Y bilježi broj dobivenih parnih brojeva. Ovo je također binomna slučajna varijabla. Parametri su joj $n = 5$ i $p = \frac{3}{6} = \frac{1}{2}$, jer su tri od šest brojeva na kocki parni.

2.1.2 Poissonova razdioba

Poissonova slučajna varijabla je diskretna slučajna varijabla koja broji koliko se puta pojavio fiksni događaj A. Međutim, za razliku od binomne, Poissonova slučajna varijabla ima prebrojiv (beskonačan) skup vrijednosti

$$\{0, 1, 2, 3, \ldots\}.$$

Poissonova slučajna varijabla određena je parametrom a,

$$X \sim P(a),$$

za koji vrijedi

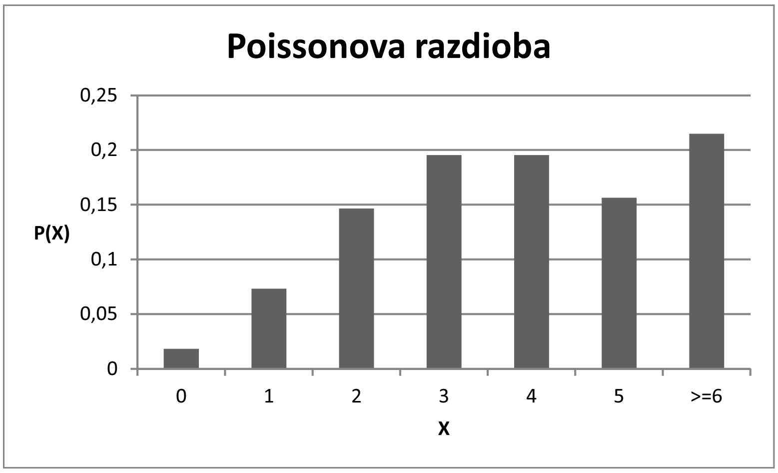
$$P(X=i) = e^{-a}\frac{a^i}{i!}, \qquad i = 0, 1, 2, 3, \ldots$$

Računanje u Excelu je slično kao za binomnu slučajnu varijablu. Da dobijemo vrijednost $P(X=i)$ u Excelu koristimo naredbu

=POISSON(i; a; FALSE),

a za računanje $P(X \leq i)$,

=POISSON(i; a; TRUE).



Slika 2.2: Graf razdiobe vjerojatnosti Poissonove slučajne varijable s parametrom $a = 4$.

Za očekivnje i varijancu Poissonove slučajne varijable X vrijedi

$$E(X) = a, \qquad V(X) = a.$$

Primjer 2.4
Pretpostavimo da Marija prosječno dobije 4 poruke na sat. Slučajna varijabla koja bilježi broj poruka koje je Marija dobila unutar jednog sata je Poissonova s parametrom 4, tj. $X \sim P(4)$.

2.2 Kontinuirana slučajna varijabla

Skup vrijednosti kontinuirane slučajne varijable je neprebrojiv. Ovakva slučajna varijabla poprima sve vrijednosti u nekom intervalu $[a, b]$. **Razdioba vjerojatnosti** na tom intervalu zadana je funkcijom f za koju vrijedi

(i) $f(x) \geq 0, \quad x \in [a, b]$,

(ii) $\int_a^b f(x)\, dx = 1$.

Vjerojatnost da slučajna varijabla poprimi vrijednost iz intervala $[a_0, b_0]$ jednaka je

$$P(a_0 < X < b_0) = \int_{a_0}^{b_0} f(x)\, dx.$$

Funkciju f nazivamo **funkcija gustoće** vjerojatnosti.

Uz funkciju gustoće definira se i **funkcija distribucije** vjerojatnosti normalne slučajne varijable X u x. To je vjerojatnost da je $X < x$. Funkcija distribucije je primitivna funkcija funkcije gustoće vjerojatnosti. Točnije, vrijedi $F'(x) = f(x)$ osim za zanemarivo mnogo vrijednosti x. Računa se po formuli

$$F(x) = \int_{-\infty}^{x} f(t)\, dt,$$

Očekivanje kontinuirane slučajne varijable X je integral

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

a njena **varijanca** je

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

2.2.1 Normalna razdioba

Normalna (Gaussova) slučajna varijabla je kontinuirana slučajna varijabla. Normalna slučajna varijabla X određena je s dva parametra, očekivanjem μ i standardnom devijacijom σ (odnosno varijancom σ^2), te pišemo

$$X \sim N(\mu, \sigma^2).$$

Funkcija gustoće vjerojatnosti normalne slučajne varijable dana je izrazom

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

a u Excelu se računa naredbom

=NORMDIST(x; μ; σ; FALSE).

Funkcija distribucije dobije se naredbom

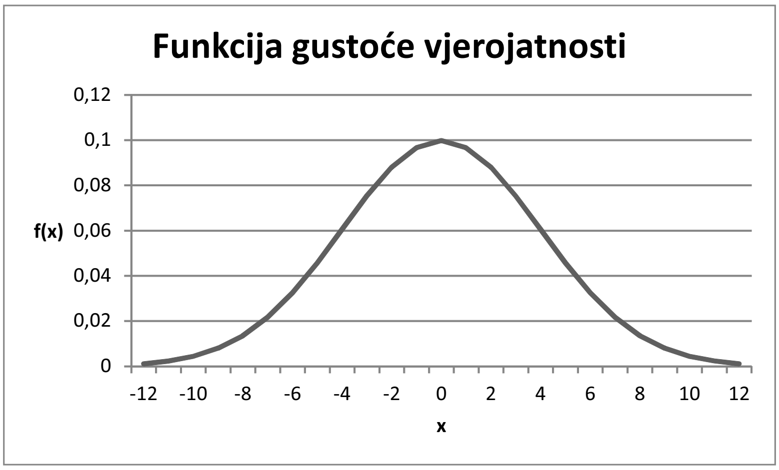
=NORMDIST(x; μ; σ; TRUE).

Interval tri sigme je interval oko očekivanja μ, od 3σ lijevo od očekivanja do 3σ desno od očekivanja, tj.

$$[\mu - 3\sigma, \mu + 3\sigma].$$

Unutar tog intervala nalaze se gotovo sve vrijednosti normalno distribuirane slučajne varijable X, njih 99.7%. Intervali jedna sigma i dvije sigme dobiju se analogno kao $[\mu - \sigma, \mu + \sigma]$, odnosno $[\mu - 2\sigma, \mu + 2\sigma]$.

Uobičajeni primjeri normalne slučajne varijable su varijable koje registriraju rezultate nekog mjerenja ili grešku mjerenja.



Slika 2.3: Graf funkcije gustoće vjerojatnosti normalne slučajne varijable s parametrima $\mu = 0$, $\sigma = 4$.

2.2.2 Eksponencijalna razdioba

Eksponencijalna slučajna varijabla je kontinuirana slučajna varijabla određena parameterom λ,

$$X \sim E(\lambda).$$

Ona poprima sve vrijednosti u iz skupa $\langle 0, \infty\rangle$. Funkcija gustoće vjerojatnosti eksponencijalne slučajne varijable dana je izrazom

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0, \\ 0, & x \leq 0. \end{cases}$$

Iz toga se izračuna da je funkcija distribucije vjerojatnosti

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & x > 0, \\ 0, & x \leq 0. \end{cases}$$

U Excelu se vrijednosti ovih funkcija računaju naredbom

=EXPONDIST(x; λ; FALSE),

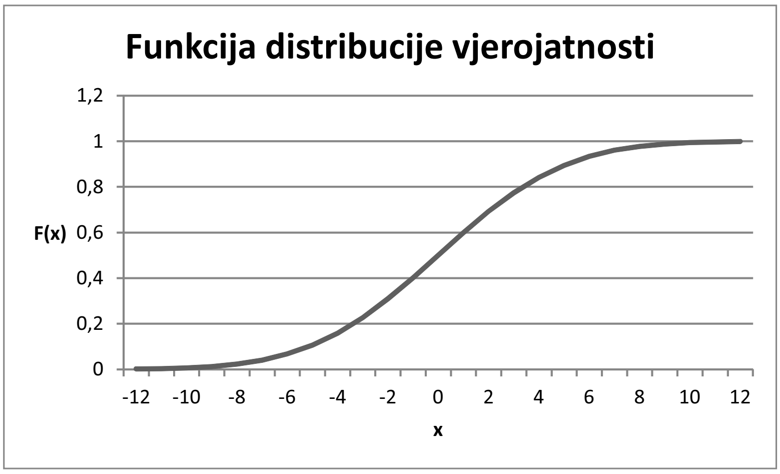
odnosno

=EXPONDIST(x; λ; TRUE).

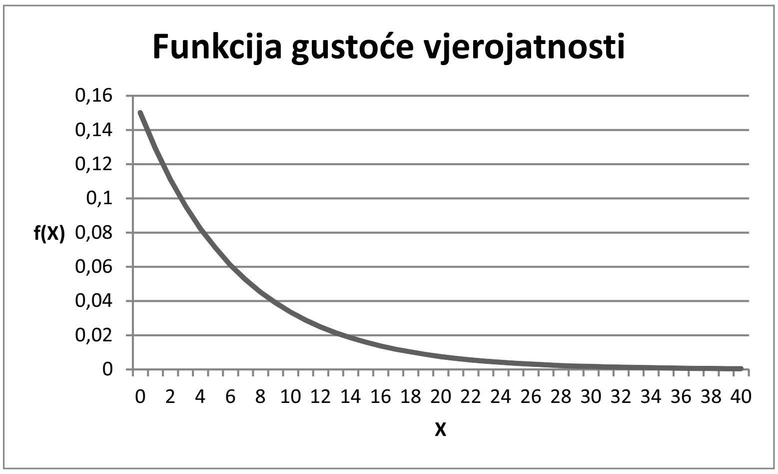
Za očekivnje i varijancu Poissonove slučajne varijable X vrijedi

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

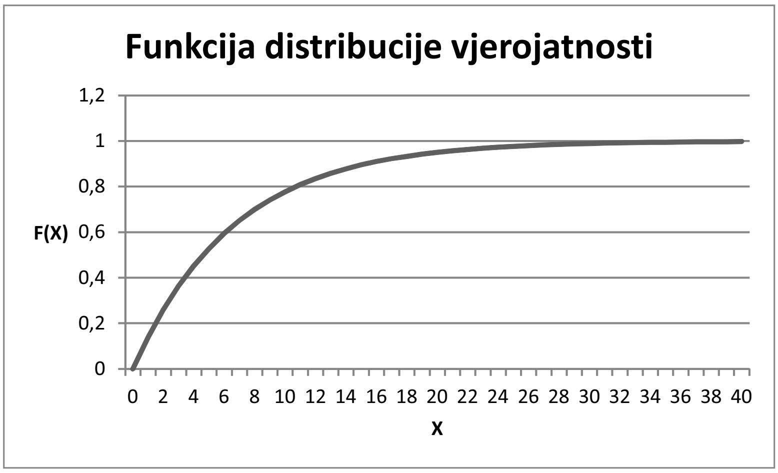
Primjer eksponencijalne slučajne varijale je varijabla koja bilježi vrijeme proteklo do pojave nekog događaja.



Slika 2.4: Graf funkcije distribucije vjerojatnosti normalne slučajne varijable s parametrima $\mu = 0$, $\sigma = 4$.



Slika 2.5: Graf funkcije gustoće vjerojatnosti eksponencijalne slučajne varijable s parametrom $\lambda = 0.15$.



Slika 2.6: Graf funkcije distribucije vjerojatnosti eksponencijalne slučajne varijable s parametrom $\lambda = 0.15$.

Poglavlje 3

Testovi hipoteza

Primjeri u Excelu vezani za ovu cjelinu nalaze se u dokumentu *Testovihipoteza.xlsx*.

3.1 Interval pouzdanosti za očekivanje

Očekivanje neke slučajne varijable procjenjujemo aritmetičkom sredinom dobivenih podataka $x_1, x_2, \ldots, x_n$. Tražimo interval oko aritmetičke sredine $\bar{x}$ unutar kojeg se, uz određenu vjerojatnost, nalazi očekivanje μ. Takav interval nazivamo **interval pouzdanosti** za očekivanje slučajne varijable.

Da bismo iz poznatih vrijednosti $x_1, x_2, \ldots, x_n$ normalno distribuirane slučajne varijable odredili njen interval pouzdanosti uz pouzdanost $1 - 2p$ potrebno je znati broj podataka n, njihovu aritmetičku sredinu $\bar{x}$ te korigiranu standardnu devijaciju s. Tada interval pouzdanosti zapisujemo kao

$$\langle a, b\rangle = \langle \bar{x} - t_p(n-1)\frac{s}{\sqrt{n}},\ \bar{x} + t_p(n-1)\frac{s}{\sqrt{n}}\rangle,$$

pri čemu je broj $t_p(n-1)$ vrijednost **t-razdiobe** s $n-1$ stupnjeva slobode. U Excelu $t_p(n-1)$ računamo na sljedeći način

$$=\text{TINV}(1-p;\, n-1).$$

Što je veća pouzdanost koju zahtijevamo, širi je interval pouzdanosti.

3.2 Testovi hipoteza

Kod statističkih testova razlikujemo dva slučaja. Prva mogućnost je da testiramo rezultate dobivene mjerenjem u odnosu na neki kontrolni uzorak. Tada testiramo hipotezu o jednakosti varijance i jednakosti očekivanja

$$\mathbf{H_0}: \quad \sigma^2 = \sigma_0^2, \quad \mu = \mu_0, \tag{3.1}$$

gdje su σ_0 i μ_0 poznate (deklarirane) vrijednosti vezane uz kontrolni uzorak, a σ i μ su nepoznate vrijednosti. Druga mogućnost je testiranje rezultata dobivenih u dva odvojena skupa mjerenja. Tu testiramo hipotezu o jednakosti varijance i očekivanja dobivenih u ta dva skupa mjerenja,

$$\mathbf{H_0}: \quad \sigma_1^2 = \sigma_2^2, \quad \mu_1 = \mu_2. \tag{3.2}$$

Statistička testiranja ne daju 100% sigurnost u dobiveni rezultat. Prilikom testiranja unaprijed zadajemo dozvoljenu pogrešku, tzv. **nivo signifikantnosti**. Uobičajeno se uzima signifikantnost $\alpha = 0.05$, što znači da je vjerojatnost odbacivanja istinite hipoteze 5%. Naravno, može se uzeti i neki drugi nivo signifikantnosti.

(a) Neka je dan kontrolni uzorak s očekivanjem μ i standardnom devijacijom σ, te neka je u n mjerenja dobiven prosječni rezultat $\bar{x}$ uz korigiranu standardnu devijaciju s.

(i) Testiramo hipotezu

$$\mathbf{H_0}: \quad \sigma^2 = \sigma_0^2,$$

uz alternativnu hipotezu

$$\mathbf{H_a}: \quad \sigma^2 > \sigma_0^2.$$

Za testiranje ove hipoteze koristimo **hi-kvadrat test** s $n-1$ stupnjeva slobode, $\chi^2(n-1)$. Test se zasniva na činjenici iz teorije vjerojatnosti da je

$$(n-1)\frac{s^2}{\sigma_0^2} \sim \chi^2(n-1).$$

Računamo $\chi^2(n-1)$, u Excelu za to koristimo naredbu

=CHIINV(signifikantnost; $n-1$),

i veličinu

$$\chi^2_{\text{exp}} = (n-1)\frac{s^2}{\sigma_0^2}.$$

Hipotezu $\mathbf{H_0}$ prihvaćamo ako je

$$\chi^2_{\text{exp}} < \chi^2_\alpha(n-1). \tag{3.3}$$

(ii) Testiramo hipotezu

$$\mathbf{H_0}: \quad \mu = \mu_0,$$

uz alternativnu hipotezu

$$\mathbf{H_a}: \quad \mu \neq \mu_0.$$

Ovdje koristimo **t-test** s $n-1$ stupnjeva slobode, $t(n-1)$, a testiranje se zasniva na činjenici iz teorije vjerojatnosti da je

$$\frac{\bar{x}-\mu_0}{\frac{s}{\sqrt{n}}} \sim t(n-1).$$

Za računanje vrijednosti $t(n-1)$ koristimo naredbu

=TINV(signifikantnost; $n-1$),

te računamo

$$t_{\exp} = \frac{\bar{x}-\mu_0}{\frac{s}{\sqrt{n}}}.$$

Hipotezu $\mathbf{H_0}$ prihvaćamo ako je

$$|t_{\exp}| < t_{\frac{\alpha}{2}}(n-1). \tag{3.4}$$

(b) Neka je u n_1 mjerenja jedne normalno distribuirane slučajne varijable dobiven prosječni rezultat $\bar{x}_1$ uz korigiranu standardnu devijaciju s_1, a u n_2 mjerenja druge normalno distribuirane slučajne varijable dobiven prosječni rezultat $\bar{x}_2$ uz korigiranu standardnu devijaciju s_2. Indekse ćemo odabrati tako da je $s_1^2 > s_2^2$.

(i) Testiranje hipoteze

$$\begin{aligned} \mathbf{H_0}: &\quad \sigma_1^2 = \sigma_2^2, \\ \mathbf{H_a}: &\quad \sigma_1^2 > \sigma_2^2. \end{aligned}$$

Koristimo **F test** s (n_1-1, n_2-1) stupnjeva slobode. Test se zasniva na činjenici iz teorije vjerojatnosti da je

$$\frac{s_1^2}{s_2^2} \sim F(n_1-1, n_2-1).$$

Stoga računamo $F(n_1-1, n_2-1)$ koristeći naredbu

=FINV(signifikantnost; n_1-1; n_2-1), i veličinu

$$F_{\text{exp}} = \frac{s_1^2}{s_2^2}.$$

Slično kao do sad, hipotezu $\mathbf{H_0}$ prihvaćamo ako je

$$F_{\text{exp}} < F_\alpha(n_1 - 1, n_2 - 1). \tag{3.5}$$

(ii) Testiranje hipoteze

$$\begin{aligned}\mathbf{H_0}: &\quad \mu_1 = \mu_2,\\ \mathbf{H_a}: &\quad \mu_1 \neq \mu_2.\end{aligned}$$

Za testiranje hipoteze o jednakosti očekivanja opet koristimo **t-test**, ali sada s n_1+n_2-2 stupnjeva slobode. Testiranje se zasniva na činjenici iz teorije vjerojatnosti da je

$$\frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}\sqrt{\frac{n_1+n_2}{n_1 n_2}}} \sim t(n_1 + n_2 - 2).$$

Izračunamo

$$t_{\text{exp}} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}\sqrt{\frac{n_1+n_2}{n_1 n_2}}},$$

a hipotezu prihvaćamo ako je

$$|t_{\text{exp}}| < t_{\frac{\alpha}{2}}(n_1 + n_2 - 2). \tag{3.6}$$

3.3 Naredba IF u Excelu

Kada u Excelu testiramo vrijedi li neka od relacija (3.3)– (3.6), to radimo na sljedeći način. Recimo da u polju B1 imamo vrijednost $\chi^2(n-1)$, a u polju B2 vrijednost χ^2_{exp}. Tada u polje u kojem testiramo relaciju (3.3) upisujemo

=B2 < B1.

Rezultat će biti TRUE ili FALSE, ovisno o tome je li vrijednost u polju B2 manja od one u polju B1 ili nije. S ovakvim se poljima može dalje računati jer TRUE ima numeričku vrijednost 1, a FALSE 0.

Da bi konačni rezultat testiranja (3.1) bio pozitivan, moraju vrijediti relacije (3.3) i (3.4). Dakle, kao rezultate tih pojedinačnih testiranja moramo dobiti TRUE i TRUE. U bilo kojem drugom slučaju, rezultat ukupnog testiranja je negativan. Isto vrijedi i kod testiranja hipoteze (3.2) gdje provjeravamo vrijede li relacije (3.5) i (3.6).

Za ukupni test možemo korisiti naredbu IF,

=IF(uvjet; ako uvjet vrijedi; ako uvjet ne vrijedi).

Ako u slučaju da uvjet koji provjeravamo vrijedi želimo u polju ispisati *da*, a u slučaju da ne vrijedi *ne*, onda pišemo

=IF(uvjet; “da”; “ne”).

Poglavlje 4

Metoda najmanjih kvadrata

Primjeri u Excelu vezani za ovu cjelinu nalaze se u dokumentu *MNK.xlsx*.

Pretpostavimo da su zadana dva skupa podataka,

$$x = \{x_1, x_2, \dots, x_n\} \quad \text{i} \quad y = \{y_1, y_2, \dots, y_n\}.$$

Zanima nas jesu li ta dva skupa podataka međusobno povezana (korelirana) i ako da, na koji način. U slučaju da su veličine x i y korelirane, onda ako znamo jednu od njih, x_i, možemo procijeniti drugu, y_i. Najjednostavnija veza među podatcima je linearna. To znači da su vrijednosti x i y povezane linearnom funkcijom,

$$y := f(x) = ax + b.$$

Podatke prikazujemo kao točke u koordinatnom sustavu pri čemu je

$$T_1 = (x_1, y_1), T_2 = (x_2, y_2), \dots, T_n = (x_n, y_n).$$

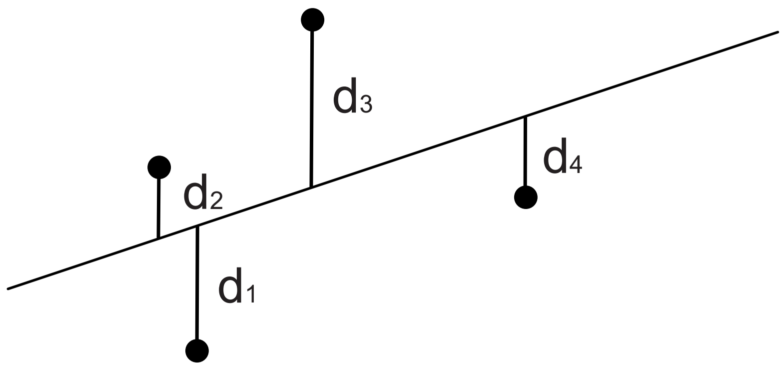
Imajući u vidu da je graf linearne funkcije pravac, vezu među podatcima grafički prikazujemo kao pravac. Taj se pravac naziva **regresijski pravac** ili pravac linearne regresije.

U općem slučaju nije moguće povući jedan pravac koji će prolaziti kroz sve točke. Stoga, tražimo pravac koji prolazi "dovoljno blizu" svim točkama. Biramo ga tako da promatramo udaljenosti d_i, (slika 4.1)

$$d_i = y_i - (ax_i + b),$$

gdje su a i b nepoznati parametri. Potrebno je da suma kvadrata ovih udaljenosti bude što manja,

$$\sum_{i=1}^{n} d_i^2 = d_1^2 + d_2^2 + \ldots + d_n^2 \to \min.$$



Slika 4.1: Metoda najmanjih kvadrata

Iz gornjeg zahtjeva dobiju se parametri a i b koji se potom uvrste u jednadžbu $y = ax + b$. Ovdje nećemo raditi izvod tih parametara, nego ćemo za dobivanje jednadžbe regresijskog pravca korisiti mogućnosti Excela.

Odgovor na pitanje koliko dobro regresijski pravac aproksimira zadane podatke daje **koeficijent linearne korelacije** R. Njegova vrijednost je između -1 i 1. Koeficijent R je pozitivan ako je regresijska funkcija rastuća, a negativan ako je padajuća. Što je R po apsolutnoj vrijednosti bliže 1, to je aproksimacija bolja. Ako bi imali situaciju u kojoj sve zadane točke leže na regresijskom pravcu, vrijednost koeficijenta R bila bi 1 ili -1. Što ćemo uzeti za dovoljno dobru aproksimaciju ovisi o problemu. Uglavnom možemo smatrati da su podatci jako linearno korelirani ako je

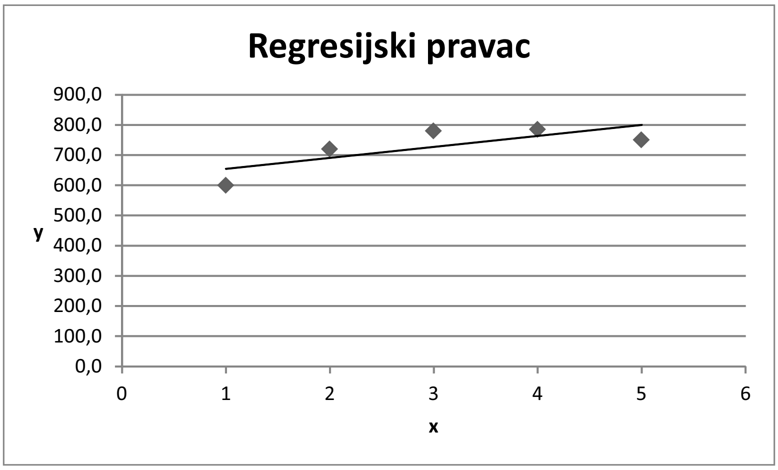
$$|R| \geq 0.9.$$

U Excelu se koeficijent R računa naredbom

$$\text{CORREL(vrijednosti } x\text{; vrijednosti } y\text{).}$$

Kada znamo vezu podataka x i y, za dodatne vrijednosti x_k pripadne vrijednosti y_k računamo jednostavnim uvrštavanjem u jednadžbu regresijskog pravca. Ovdje treba imati na umu da su dobivene vrijednosti približne, te više vjerodostojne što je bolji koeficijent R.

Također, potrebno je znati da koreliranost podataka ne uvjetuje nužno i njihovu uzročnost. Odnosno, moguće je da su dva skupa podataka dobro korelirana, ali su u stvarnosti neovisni. Stoga uvijek treba biti pažljiv pri interpretaciji rezultata.



Slika 4.2: Primjer regresijskog pravca

Poglavlje 5

Interpolacija i ekstrapolacija

Primjeri u Excelu vezani za ovu cjelinu nalaze se u dokumentu *Interpolacija.xlsx*.

Problem interpolacije i ekstrapolacije javlja se u praksi kada imamo zadane vrijednosti neke funkcije samo na diskretnom skupu podataka. Tada je pomoću poznatih podataka potrebno aproksimirati (približno izračunati) nepoznate vrijednosti. Računanje vrijednosti funkcije unutar nekog poznatog intervala naziva se **interpolacija**, a izvan tog intervala **ekstrapolacija**.

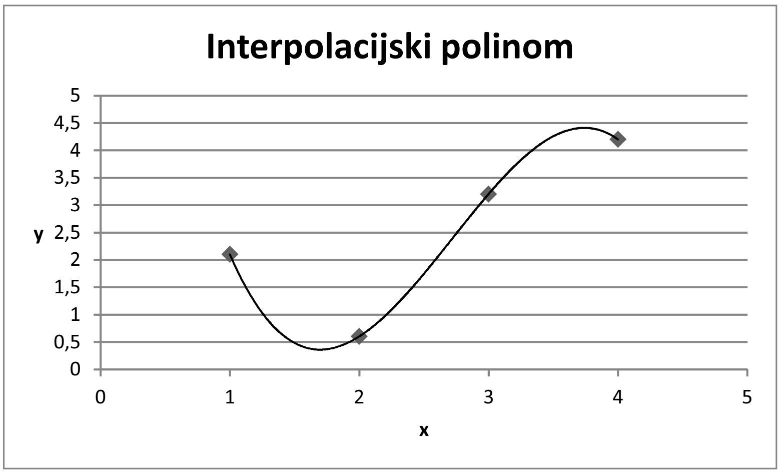
5.1 Interpolacijski polinom

Slično kao kod metode najmanjih kvadrata tražimo vezu između dva skupa podataka

$$x = \{x_1, x_2, \ldots, x_n\} \quad \text{i} \quad y = \{y_1, y_2, \ldots, y_n\}.$$

Ovdje ta veza neće biti linearna funkcija nego polinom. Već smo rekli da u općem slučaju pravac (graf polinoma 1. stupnja) nije moguće izabrati tako da prolazi kroz n točaka (gdje je $n \geq 3$). Da bi graf sigurno prolazio kroz svih n točaka, to mora biti graf polinoma stupnja barem $n-1$. Takav se polinom naziva **interpolacijski polinom**. Interpolacijski polinom je jednstven uz uvjet da je stupnja najviše $n-1$. Stoga, radimo s polinomima stupnja točno $n-1$. Za dobivanje jednadžbe interpolacijskog polinoma koristimo mogućnosti Excela pa nećemo ulaziti u izvod koeficijenata.

Napomenimo da bi polinom stupnja višeg od $n-1$ također mogli izabrati tako da prolazi kroz svih n točaka, no ne želimo dobiti kompliciraniju funkciju nego što je to nužno. Također, u praksi je preporučljivo koristiti interpolacijski polinom najviše trećeg stupnja jer se kod polinoma višeg stupnja



Slika 5.1: Primjer interpolacijskog polinoma stupnja 3

pri računanju na računalu mogu javiti velike greške nastale zbog greške zaokruživanja koju radi računalo. U takvim se situacijama uglavnom prelazi na tzv. interpolaciju splajnovima.

Poglavlje 6

Numeričko rješavanje jednadžbi i sustava

Primjeri u Excelu vezani za ovu cjelinu nalaze se u dokumentu *Jednadzbe.xlsx*.

Često se nelinearne jednadžbe ne mogu riješiti egzaktno. U tom slučaju tražimo **približno rješenje** jednadžbe.

Svaku jednadžbu s jednom nepoznanicom možemo zapisati u obliku

$$f(x) = 0.$$

Npr. ako je zadana jednadžba $e^x - x = 2$, vrijedi $e^x - x - 2 = 0$ pa je $f(x) = e^x - x - 2$. Za svako rješenje jednadžbe x^* vrijedi $f(x^*) = 0$. Stoga se traženje rješenja dane jednadžbe svodi na traženje nultočki funkcije f.

Za daljnju analizu pretpostavljamo da je funkcija f neprekidna i da su joj nultočke izolirane. Određivanje nultočki funkcije f provodi se u dva koraka.

(1) Odredimo interval na kojem se nalazi nultočka.

(2) Koristimo neku iterativnu metodu za nalaženje nultočke.

Prvi se korak radi analizom toka funkcije. Iz grafa možemo vidjeti koliko nultočki ima i gdje se one (približno) nalaze. U drugom koraku možemo koristiti različite iterativne metode. Te metode generiraju niz aproksimacija $x_0, x_1, x_2, \ldots$ Zaustavljamo se kod neke aproksimacije x_N koja je "dovoljno blizu" egzaktnog rješenja x^*.

S obzirom da ne možemo mjeriti koliko je x_N daleko od x^* jer ne znamo x^*, odstupanje od egzaktnog rješenja mjerimo odstupanjem $f(x_N)$ od 0 jer za egzaktno rješenje x^* vrijedi $f(x^*) = 0$. Vrijednost $f(x_N)$ nazivamo **greška aproksimacije**. Bitna svojstva svake metode su da li konvergira (vodi prema rješenju) te brzina konvergencije.

Slika 6.1: Izbor početnih točaka kod metode bisekcije

6.1 Metoda bisekcije

Najjednostavnija metoda za traženje nultočke funkcije f je **metoda bisekcije** (raspolavljanja). Polazimo od nekog intervala $I = [x_0, x_1]$ koji sadrži nultočku. U svakom koraku taj interval raspolavljamo i time sužavamo područje oko nultočke. Pretpostavka koja mora biti zadovoljena da bismo mogli korisiti ovu metodu je

$$f(a)f(b) < 0, \tag{6.1}$$

gdje je $a = x_0$, $b = x_1$. Drugim riječima, mora vrijediti

$$f(x_0) > 0,\ f(x_1) < 0 \quad \text{ili} \quad f(x_0) < 0,\ f(x_1) > 0.$$

Činjenica da je funkcija na jednom rubu intervala pozivitvna, a na drugom negativna osigurava postojanje nultočke unutar intervala, slika 6.1. (Zbog potpunosti, napomenimo da ovo vrijedi jer smo još na početku pretpostavili da je f neprekidna. Osim toga, metoda bisekcije koristi se samo za nultočke neparnog reda jer kod nultočki parnog reda relacija (6.1) ne može vrijediti.)

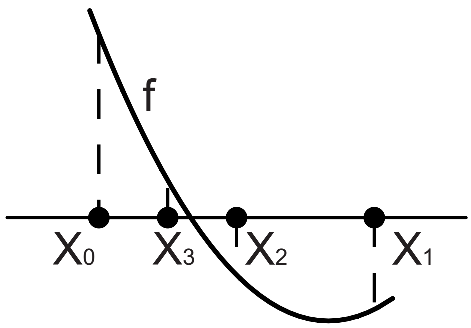
Nakon što pravilno izaberemo početne točke x_0 i x_1, za sljedeću točku, x_2, uzimamo aritmetičku sredinu točaka x_0 i x_1,

$$x_2 = \frac{x_0 + x_1}{2}.$$

Tako smo prepolovili početni interval $[x_0, x_1]$ i dalje nastavljamo ili s intervalom $[x_0, x_2]$ ili $[x_2, x_1]$. Uzimamo onaj od ta dva koji sadrži nultočku. To se provjeri iz uvjeta (6.1). Dakle, uzet ćemo onaj podinterval za kojeg je funkcija pozitivna na jednom rubu, a negativna na drugom, slika 6.2.

Opisani postupak raspolavljanja početnog intervala ponavlja se dalje na isti način. U svakom koraku područje oko nultočke postaje sve uže te je time nultočka sve preciznije određena. Postupak završavamo kada je interval oko nultočke dovoljno mali ili kada je ispunjen zadani broj koraka.

Uz pretpostavku da početni interval sadrži točno jednu nultočku, metoda bisekcije je konvergentna, uvijek će nakon dovoljnog broja koraka dovesti do rješenja, ali je spora.



Slika 6.2: Grafički prikaz metode bisekcije

6.2 Metoda sekante

Metoda sekante temelji se na aproksimaciji grafa funkcije njegovom sekantom. Kao kod metode bisekcije, uzimamo dvije početne točke, x_0 i x_1, samo što sada interval $[x_0, x_1]$ ne mora sadržavati nultočku.

Sekanta (pravac) se povlači kroz točke $(x_0, f(x_0))$ i $(x_1, f(x_1))$. Sljedeća točka, x_2, nalazi se na presjeku dobivene sekante i x-osi, slika 6.4. Općenito, točka x_{n+1} dobije se kao sjecište sekante kroz $(x_{n-1}, f(x_{n-1}))$ i $(x_n, f(x_n))$ i x-osi. Njena formula izvodi se iz formule za jednadžbu pravca kroz dvije točke i glasi

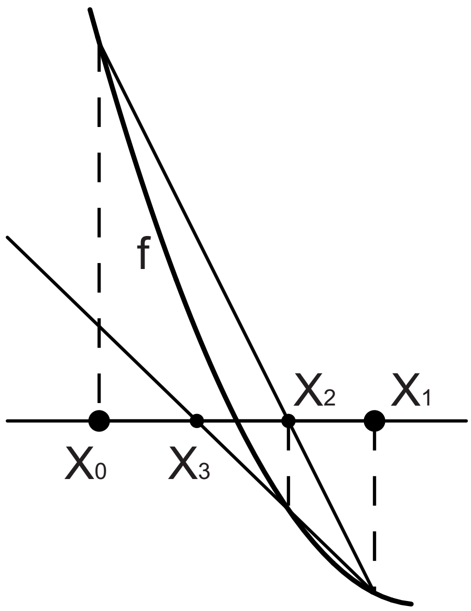
$$x_{n+1} = x_n - f(x_n)\frac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1})}.$$

Metodu zaustavljamo kada je vrijednost funkcije f u trenutnoj aproksimaciji x_N, $f(x_N)$, dovoljno mala ili kada je ispunjen zadani broj koraka. Metoda sekante nije nužno konvergentna. Konvergencija ovisi o izboru početnih točaka što predstavlja problem. S druge strane, ako konvergira, onda je brža od metode bisekcije.

6.3 Metoda tangente

Metoda tangente (Newtonova metoda) ima dosta sličnosti s metodom sekante. Razlika je u tome što ovdje imamo jednu početnu točku, x_0. Kroz točku $(x_0, f(x_0))$ povlači se tangenta na graf funkcije f, a sljedeće točka, x_1, nalazi se na presijeku te tangente i x-osi. Općenito, točka x_{n+1} nalazi se na presijeku tangente na f kroz točku $(x_n, f(x_n))$ i x-osi. Formula za točku x_{n+1} glasi

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}.$$



Slika 6.3: Grafički prikaz metode sekante

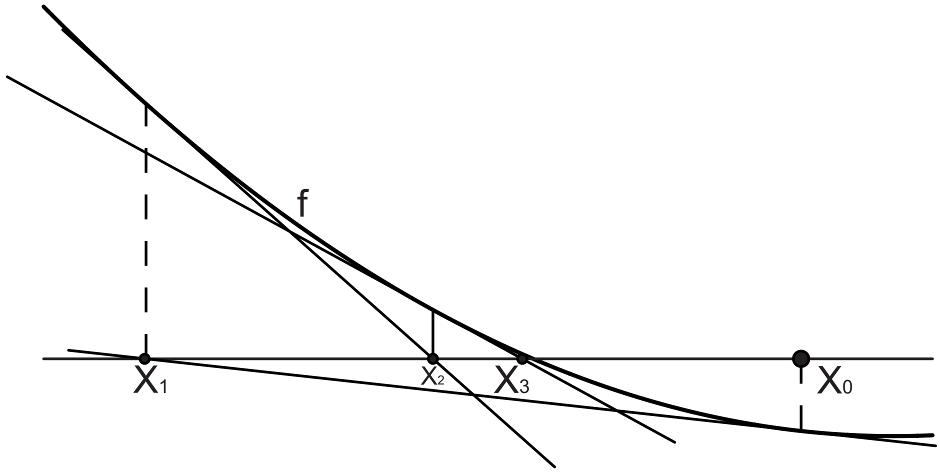
Svojstva konvergencije metode tangente slična su kao kod metode sekante. Ona će konvergirati ako je početna aproksimacija x_0 dovoljno blizu egzaktnog rješenja.

6.4 Solver

Solver je alat u Excelu koji se koristi za numeričko rješavanje jednadžbi. Prije prvog korištenja Solvera potrebno je instalirati dodatak Solver u Excel. To se radi na sljedeći način. U izborniku *File* kliknete na *Options*, te potom *Add-Ins*. Dalje, pod *Manage* izaberete *Excel Add-ins* i kliknete *Go*. U novom prozoru označite *Solver Add-in* i *OK*. Nakon što je Solver uključen, pojavit će se u izborniku *Data*.

Da bi riješili jednadžbu korištenjem Solvera, potrebno je uzeti početnu procjenu rješenja, označimo ga x_0, i izračunati vrijednost funkcije $f(x_0)$. Procjenu x_0 uglavnom biramo koristeći grafičku interpretaciju jednadžbe. Potrebno je modificirati x_0 kako bi vrijedilo $f(x_0) = 0$. Otvorimo Solver te u njemu postavimo sljedeće zahtjeve:

$$\begin{aligned} \textit{Set objective} \rightarrow &\ f(x_0) \\ \textit{Value of} \rightarrow &\ 0 \\ \textit{By changing variable cells} \rightarrow &\ x_0 \end{aligned}$$



Slika 6.4: Grafički prikaz metode tangente

i kliknemo *Solve* te *Keep solver solution*. U polju u kom je prije bila upisana procjena x_0 sada dobijemo novi broj koji je približno rješenje jednadžbe.

6.5 Sustavi jednadžbi

Sustav jednadžbi također možemo riješiti korištenjem Solvera. Pokazat ćemo to na primjeru dvije jednadžbe s dvije nepoznanice, x i y.

Dvije zadane jednadžbe možemo poistovijetiti s dvije funkcije dvije varijable, $F_1(x, y)$ i $F_2(x, y)$. Npr., ako je zadan sustav

$$\begin{aligned} 3x^2 - y^2 &= 3, \\ 2x - y &= 5, \end{aligned}$$

imamo

$$\begin{aligned} F_1(x, y) &= 3x^2 - y^2 - 3, \\ F_2(x, y) &= 2x - y - 5. \end{aligned}$$

Stoga su rješenja sustava (x^*, y^*) ujedno nultočke funkcija F_1 i F_2.

Sada su potrebne početne procjene za obe nepoznanice, x_0 i y_0. Njih modificiramo u Solveru. Budući da u polje *Set objective* u Solveru nije moguće istovremeno upisati i $F_1(x_0, y_0)$ i $F_2(x_0, y_0)$, uvodimo pomoćno polje

$F_1(x_0, y_0) + F_2(x_0, y_0)$. Postavljamo sljedeće zahtjeve:

$$\begin{aligned}
\textit{Set objective} \rightarrow\ & F_1(x_0, y_0) + F_2(x_0, y_0) \\
\textit{Value of} \rightarrow\ & 0 \\
\textit{By changing variable cells} \rightarrow\ & x_0;\ y_0 \\
\textit{Subject to the constrains} \rightarrow\ & F_1(x_0, y_0) = 0;\ F_2(x_0, y_0) = 0.
\end{aligned}$$

Kao i u primjeru jedne jednadžbe, kliknemo *Solve* te *Keep solver solution.* U poljima u kojima su bile upisane procjene x_0 i y_0 dobijemo rješenja sustava.

Poglavlje 7

Numeričko rješavanja običnih diferencijalnih jednadžbi

Primjeri u Excelu vezani za ovu cjelinu nalaze se u dokumentu *ODJ.xlsx*.

U različitim primjenama susrećemo se s problemima koji uključuju rješavanje diferencijalnih jednadžbi. Često diferencijanlu jednadžbu nije moguće riješiti egzaktno, ili je pak postupak rješavanja dug i težak. Tada pribjegavamo nekoj numeričkoj metodi za rješavanje.

Ovdje se bavimo rješavanjem običnih diferencijalnih jednadžbi (ODJ) prvog reda,

$$y'(x) = f(x, y(x)),$$

uz početni uvjet $y(x_0) = y_0$. Ovakvu vrstu problema nazivamo Cauchyjev ili inicijalni problem. Jednadžbe rješavamo na intervalu $[a, b]$ koji je podijeljen točkama

$$a = x_0, x_1, x_2, \ldots, x_n = b.$$

Udaljenost svake dvije susjedne točke iznosi h, tj. vrijedi $x_{i+1} = x_i + h$, $i = 0, \ldots, n-1$.

7.1 Eulerova metoda

Eulerova metoda je najjednostavnija metoda za numeričko rješavanje Cauchyjevog problema

$$y' = f(x, y), \quad y(x_0) = y_0.$$

Prvu derivaciju y' možemo, uz određenu grešku, aproksimirati izrazom

$$y' \approx \frac{y(x+h) - y(x)}{h}.$$

Uvrštavanjem u zadanu jednadžbu dobijemo

$$\frac{y(x+h)-y(x)}{h} \approx f(x,y),$$

iz čega se lako izvede relacija

$$y(x+h) \approx y(x) + hf(x,y).$$

Iz gornjeg izraza slijedi rekurzivna formula za Eulerovu metodu

$$y_{i+1} = y_i + hf(x_i, y_i), \quad i = 0, \ldots, n-1. \tag{7.1}$$

Vrijednost y_0 zadana je početnim uvjetom. Vrijednosti y_i dobivene formulom(7.1) su aproksimacije rješenja diferencijalne jednadžbe u točkama x_i, $i = 1, \ldots, n$. Rješenje je točnije što je manji korak h.

Formulu (7.1) ponekad zapisujemo u obliku

$$\begin{aligned} y_{i+1} &= y_i + k, \\ k &= hf(x_i, y_i), \qquad i = 0, \ldots, n-1. \end{aligned}$$

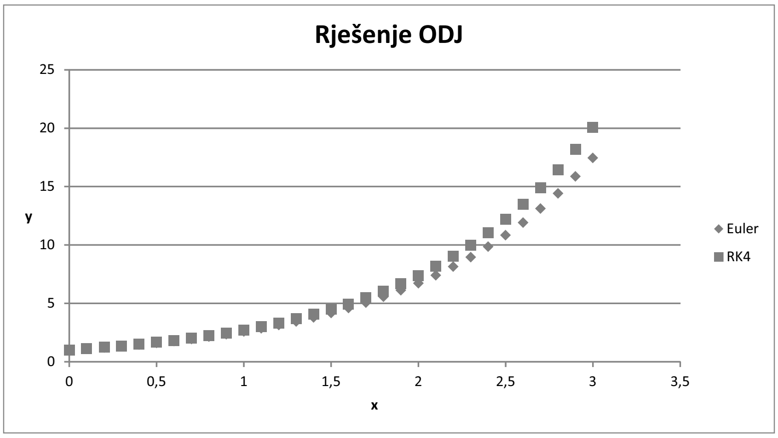
7.2 Runge-Kutta 4 metoda

Runge-Kutta metode zasnivaju se na istoj ideji kao i Eulerova metoda. Uz poznatu vrijednost vrijednost y_i u točki x_i računamo vrijednost y_{i+1} u točki $x_{i+1} = x_i + h$. Ove su metode gotovo uvijek točnije od Eulerove metode, ali su složenije za računanje. Runge-Kutta metode razlikuju se međusobno po broju stadija, a najpopularnija je ona s četiri statdija, tzv. **Runge-Kutta 4** metoda.

Runge-Kutta 4 metoda određena je formulom

$$\begin{aligned} y_{i+1} &= y_i + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4), \\ k_1 &= hf(x_i,\, y_i), \\ k_2 &= hf(x_i + \frac{1}{2}h,\, y_i + \frac{1}{2}k_1), \\ k_3 &= hf(x_i + \frac{1}{2}h,\, y_i + \frac{1}{2}k_2), \\ k_4 &= hf(x_i + h,\, y_i + k_3), \qquad i = 0, \ldots, n-1. \end{aligned}$$

Izvod formule ovdje preskačemo.



Slika 7.1: Usporedba rješenja dobivenih Eulerovom i Runge-Kutta 4 metodom za jednadžbu $y' = y$.