MEDICINSKA STATISTIKA SA INFORMATIKOM

Slučajne promjenljive

Uvod

- U teoriji vjerovatnoće obilježju odgovara pojam slučajne promjenljive
  - Funkcija koja svakom elementarnom ishodu dodjeljuje broj
  - Obilježavaju se velikim slovima, a njihove realizovane vrijednosti malim
  - Diskretne slučajne promjenljive uzimaju prebrojivo mnogo vrijednosti, a neprekidne neprebrojivo
  - Osnovni je problem odrediti raspodjelu slučajne promjenljive

Primjer

- Ako je X broj na gornjoj strani kocke onda su moguće vrijednosti ove promjenljive 1, 2, 3, 4, 5, 6
- Ako je X broj palih šestica u tri bacanja kocke onda su moguće vrijednosti ove promjenljive 0, 1, 2, 3

Raspodjela slučajne promjenljive

- Osnovni je problem odrediti raspodjelu slučajne promjenljive, odnosno odrediti skup njenih vrijednosti i vjerovatnoće sa kojim se ove vrijednosti pojavljuju
- Diskretna slučajna veličina - tabela
- Neprekidna slučajna veličina – funkcija gustine φ
  - Vjerovatnoća $P(a < X < b)$ jednaka je površini ispod grafika funkcije gustine nad intervalom (a, b)

Raspodjela diskretne slučajne promjenljive

- Diskretna slučajna promjenljiva X određena je skupom vrijednosti

$$x_1, x_2, \dots, x_n, \dots$$

i odgovarajućim vjerovanoćama

$$p_1, p_2, \dots, p_n, \dots$$

gdje je

$$p_i = P\{X = x_i\}, \quad i = 1, 2, \dots,$$

što zapisujemo u obliku šeme

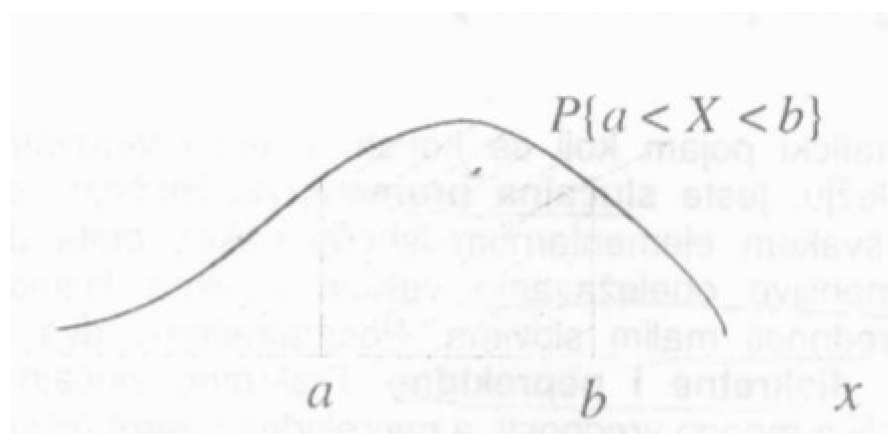
$$X : \begin{pmatrix} x_1 & x_2 & \cdots & x_n & \cdots \\ p_1 & p_2 & \cdots & p_n & \cdots \end{pmatrix}.$$

Raspodjela neprekidne slučajne promjenljive

- Neprekidna slučajna promjenljiva X određena je svojom funkcijom gustine φ_X, koja je nenegativna i važi

$$\int_{-\infty}^{\infty} \varphi_X(x)\,dx = 1$$

- Vjerovatnoća P{a < X < b} jednaka je površini ispod funkcije gustine nad intervalom (a, b), pri čemu su a, b є R



Numeričke karakteristike slučajnih promjenljivih

	Diskretne	Neprekidne
Srednja vrednost $m = E(X)$	$E(X) = \sum_{k=1}^{\infty} x_i p_i$	$E(X) = \int_{-\infty}^{\infty} x\varphi_X(x)dx$
Običan moment reda r, $r = 1,2,...$ $m_r = E(X^r)$	$E(X^r) = \sum_{k=1}^{\infty} x_i^r p_i$	$E(X^r) = \int_{-\infty}^{\infty} x^r \varphi_X(x)dx$
Centralni moment reda r, $r = 1,2,...$ $\mu_r = E((X-m)^r)$	$\mu_r = \sum_{k=1}^{\infty} (x_i - m)^r p_i$	$\mu_r = \int_{-\infty}^{\infty} (x-m)^r \varphi_X(x)dx$

Numeričke karakteristike slučajnih promjenljivih (2)

Disperzija, Centralni moment reda $r=2$ $\mu_2=\sigma^2$	$D(X)=\sum_{k=1}^{\infty}(x_i-m)^2 p_i$	$D(X)=\int_{-\infty}^{\infty}(x-m)^2\varphi_X(x)dx$
Standardna devijacija σ	$\sigma=\sqrt{D(X)}$	
Koeficijenat asimetrije	$K_A=\frac{\mu_3}{\mu_2^{3/2}}=\frac{E(X-E(X))^3}{[E(X-E(X))^2]^{3/2}}$	
Koeficijenat ekscesa	$K_E=\frac{\mu_4}{\mu_2^2}=\frac{E(X-E(X))^4}{[E(X-E(X))^2]^2}$	

Numeričke karakteristike slučajnih promjenljivih (3)

- Mod M_o
  - diskretne slučajne promjenljive je vrijednost sa najvećom vjerovatnoćom
  - neprekine slučajne promjenljive je vrijednost u kojoj funkcija gustine dostiže maksimum
- Medijana M_e
  - diskretne promjenljive je ona vrijednost za koju važi

$$P\{X < M_e\} \leq \frac{1}{2}, \quad \text{i} \quad P\{X > M_e\} \leq \frac{1}{2}.$$

  - neprekidne promjenljive je ona vrijednost za koju važi

$$P\{X < M_e\} = P\{X \geq M_e\} = \frac{1}{2}$$

Uniformna diskretna raspodjela

- Svaka vrijednost slučajne promjenljive ima istu vjerovatnoću pojavljivanja, ako ima n različitih vrijednosti pripadajuće vjerovanoće su 1/n
- Primjer, bacanje kocke za igru

$$X:\begin{pmatrix} x_1 & x_2 & \dots & x_n \\ \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \end{pmatrix}.$$

Binomna raspodjela

- Broj pojavljivanja događaja A koji ima vjerovatnoću p=P(A) u nizu od n nezavisnih eksperimenata
  - Oznaka A : B(n, p)
- Vrijednosti ove promjenljive su 0, 1, 2, …, n a odgovarajuće vjerovatnoće računaju se po formuli

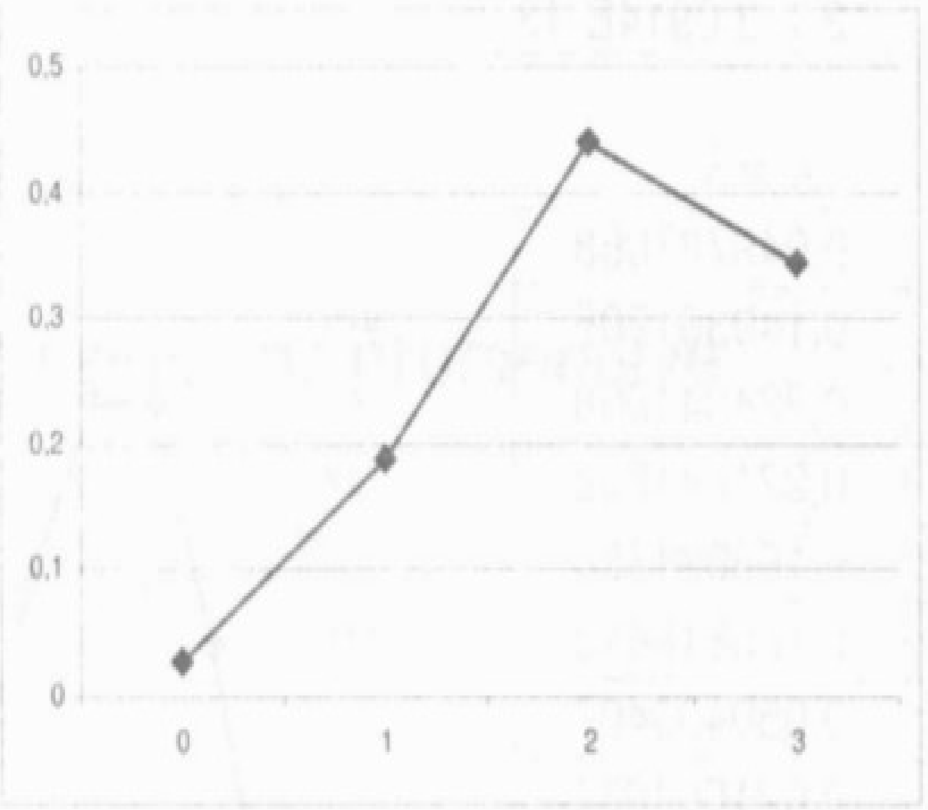
$$P\{X = k\} = \binom{n}{k} p^k q^{n-k}, \quad k = 0,1,2,\ldots,n$$

- Primjer, broj palih šestica u tri bacanja kocke za igru
- Važi E(A) = n*p, D(A) = n * p * (1 - p)

Binomna raspodjela 2

$B(3, 0.7)$

0	0,027
1	0,189
2	0,441
3	0,343



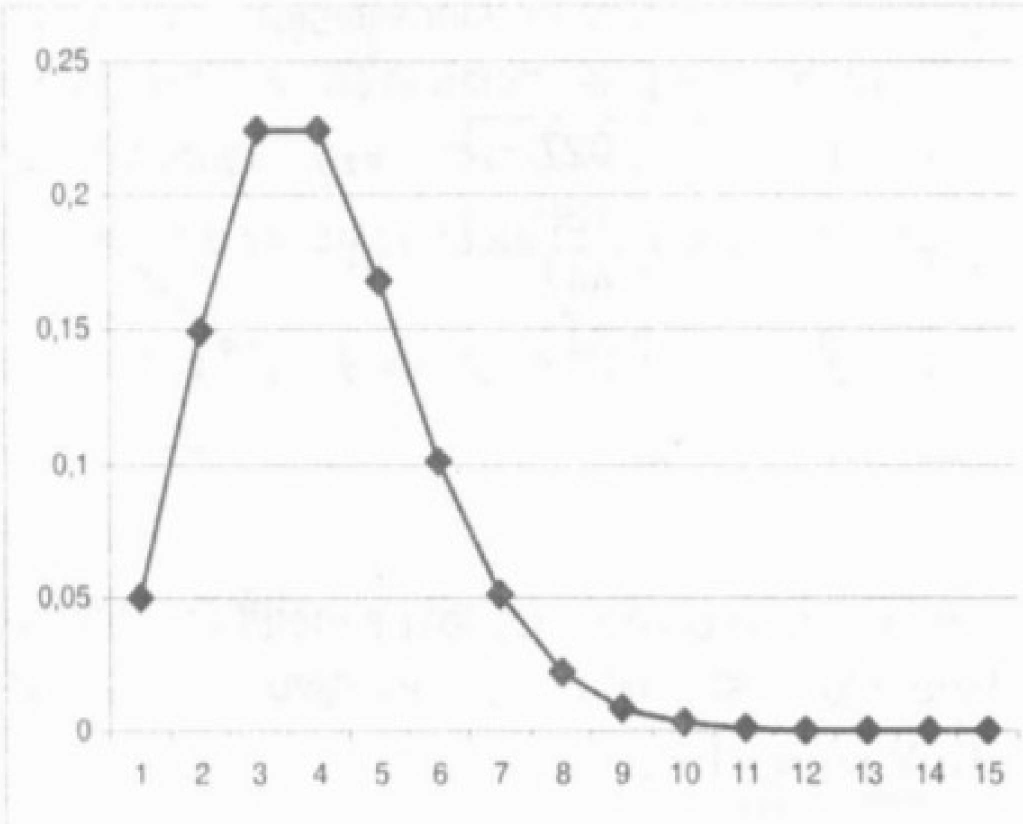
Puasonova raspodjela

- Slučajna promjenljiva A ima Puasonovu raspodjelu ako je P(A=k)=λ^k / k!, λ>0, k=0, 1, ..
- Primjer, broj poziva u telefonskoj centrali u toku jednog minuta, pri čemu je λ prosječan broj poziva u toku jednog minuta
- Važi E(A) = D(A) = λ

Puasonova raspodjela 2

$\lambda = 3$

0	0,049787068
1	0,149361205
2	0,224041808
3	0,224041808
4	0,168031356
5	0,100818813
6	0,050409407
7	0,021604031
8	0,008101512
9	0,002700504
10	0,000810151
11	0,00022095
12	5,52376E-05
13	1,27471E-05
14	2,73153E-06



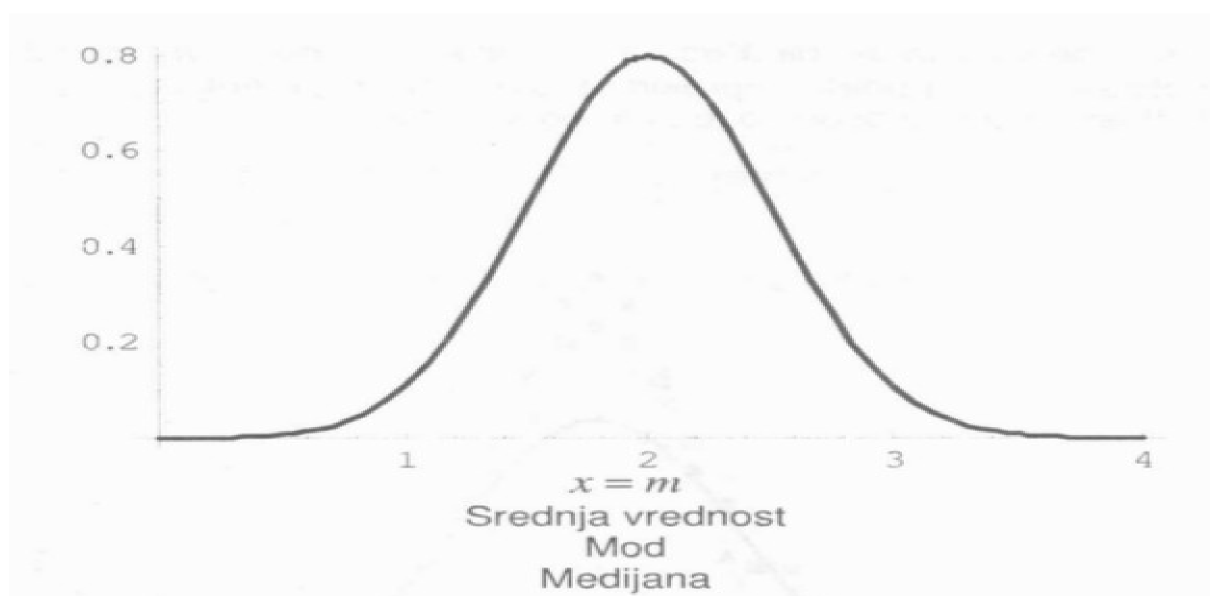
Normalna raspodjela

- Normalna raspodjela $N(m, \sigma^2)$ ima funkciju gustine

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

Srednja vrednost	m
Moda	m
Medijana	m
Disperzija	σ^2
K_A	0
K_E	3

- Gustina normalne raspodjele simetrična je oko prave x=m

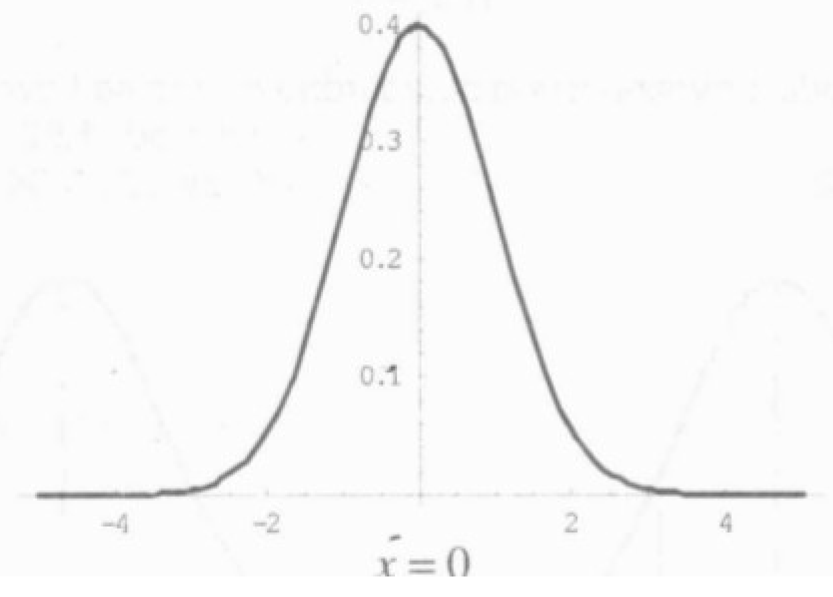


Standardizovana normalna raspodjela

- Normalna raspodjela N(0, 1) zadata je gustinom

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

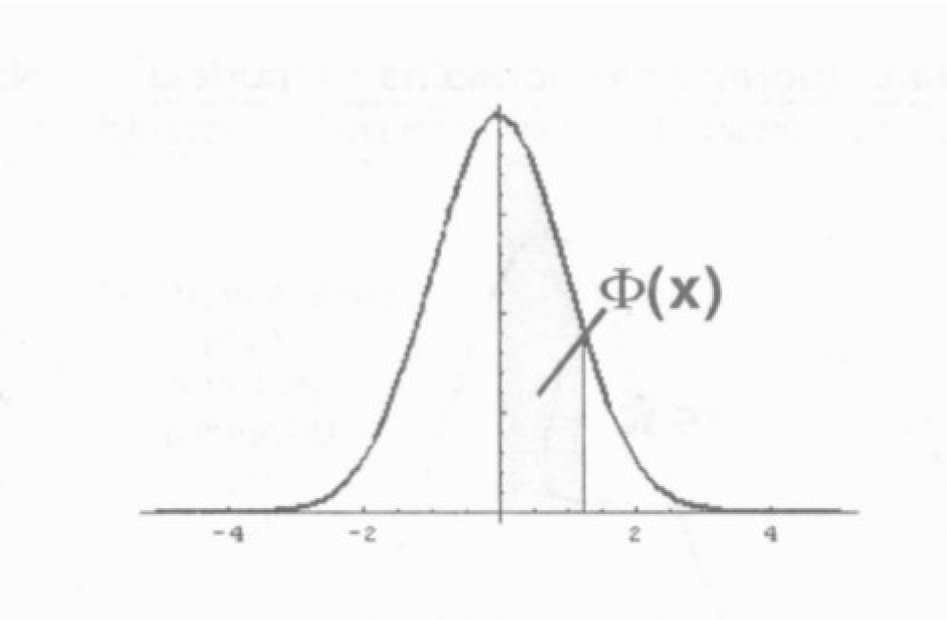
Srednja vrednost	0
Moda	0
Medijana	0
Disperzija	1
K_A	0
K_E	3



Standardizovana normalna raspodjela 2

- Za izračunavanje vjerovatnoća oblika P{a < X < b} koristi se funkcija

$$\Phi(x) = \int_0^x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx ,$$



z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,7	0,2580	0,2611	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3,0	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990
4,0	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000

Primjer 1

- Odrediti površinu ispod standardne normalne krive koja leži
  - Lijevo od 1, rješenje 0.8413
  - Desno od 1.32, rješenje 0.0934
  - Lijevo od -1.62, rješenje 0.0526
  - Između 0.34 i 1.15, rješenje 0.2418

Z-skor

- Ako slučajna promjenljiva X ima normalnu raspodjelu N(m, σ^2), tada njen z-skor, slučajna promjenljiva Z = (X – m) / σ ima normalnu raspodjelu N(0, 1)
- Ovo nam omogućava da određujemo vrijednosti P{a < X < b} po fomuli

$$P\{a < X < b\} = P\{\frac{a-m}{\sigma} < Z < \frac{b-m}{\sigma}\} = \Phi(\frac{b-m}{\sigma}) - \Phi(\frac{a-m}{\sigma})$$

Primjer 2

- Ako visina studenta ima normalnu raspodjelu sa parametrima $m = 162$ i $\sigma = 11.2$ odrediti
  - Koji procenat populacije ima visinu između 156cm i 171 cm, rješenje 0.4935
  - Koji je procenat populacije viši od 165cm, rješenje 0.3936
  - Od koje visine je niže 75% populacije, rješenje 169.504
  - U kom intervalu leži 95% populacije, rješenje (140.1, 183.9), ovakav interval nije jedinstven, obično se uzima onaj koji je simetričan u odnosu na srednju vrijednost m

Diskretne dvodimenzione slučajne promjenljive

- Često se u eksperimentu registruju dvije ili više karakteristika
- Diskretna slučajna promjenljiva (X, Y) zadaje se pomoću parova svojih vrijednosti (x_i, y_j), i, j = 1, 2, 3, … i odgovarajućih vjerovatnoća

$$p_{ij} = p(x_i, y_j) = P\{X = x_i, Y = y_j\},\ i, j = 1, 2, \ldots$$

Primjer 3

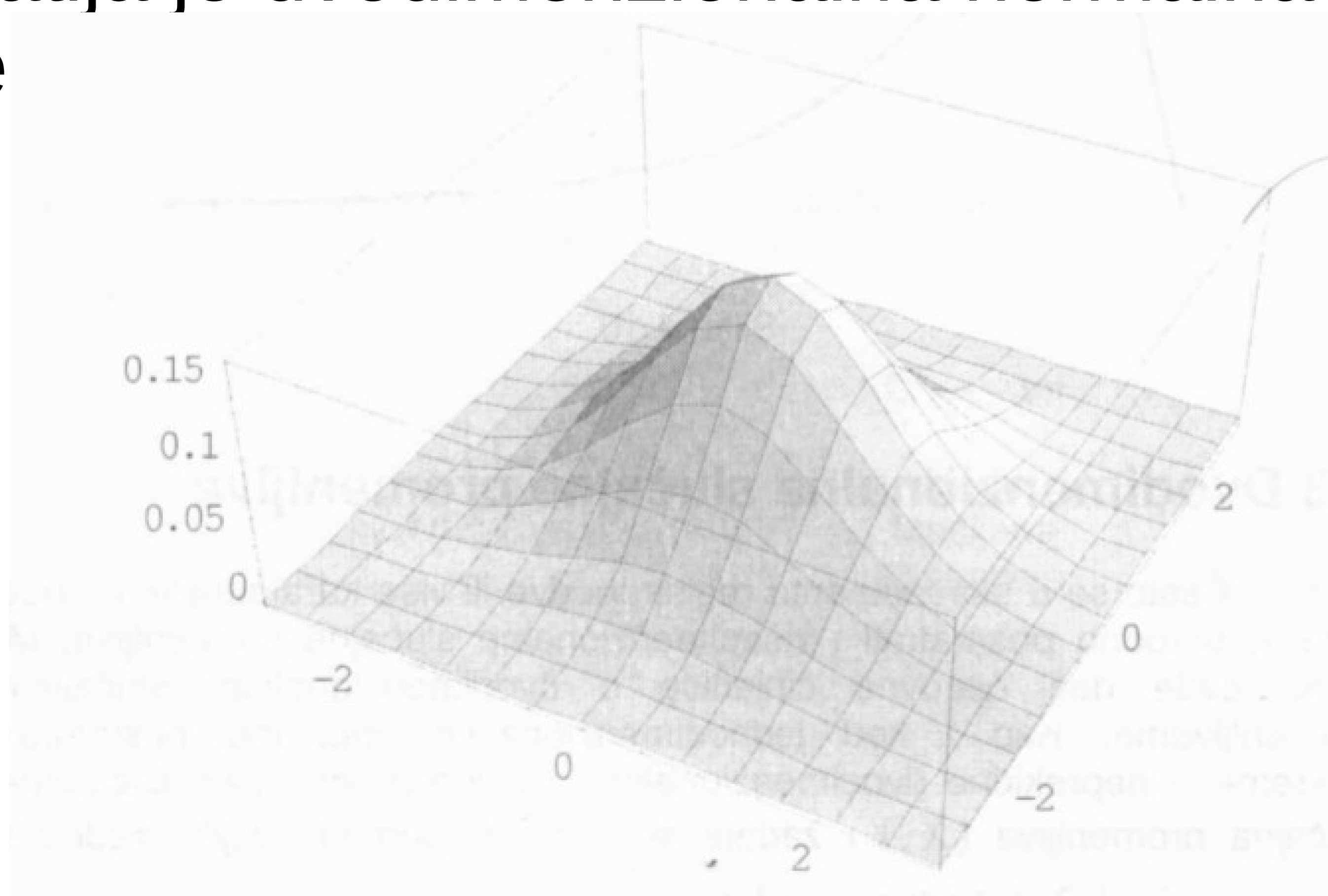
- U kutiji se nalaze 3 zelene i 3 crvene kuglice. Igrač bira 3 kuglice iz kutije i baca pravilnu kocku onoliko puta koliko je izvukao zelenih kuglica. Neka je X slučajna promjenljiva koja označava broj izvučenih zelenih kuglica, a Y slučajna promjenljiva koja označava broj pojavljivanja 6. Naći raspodjelu promjenljive (X, Y).

Rješenje

$X \setminus Y$	0	1	2	3
0	1/20	0	0	0
1	3/8	3/40	0	0
2	5/16	1/8	0	0
3	25/864	5/288	1/288	1/4320

Neprekidne dvodimenzione slučajne promjenljive

- Neprekidna slučajna promjenljiva (X, Y) zadaje se funkcijom gustine φ(x, y).
- Najpoznatija je dvodimenzionalna normalna raspodje



Koeficijent korelacije

- Koeficijent korelacije pokazuje stepen linearne povezanosti dvije slučajne promjenljive

$$\rho(X,Y)=\frac{E((X-m_X)(Y-m_Y))}{\sqrt{D(X)D(Y)}}=\frac{E(XY)-m_X m_Y}{\sqrt{D(X)D(Y)}}$$

- Osobine

  $-1 \leq \rho(X, Y) \leq 1$

  pozitivna korelacija – ako raste X raste i Y

  negativna korelacija – ako raste X opada Y

  ako je korelacija 0 – ne postoji linearna povezanost

  ako je $\rho(X, Y) = 1$, tada je $Y = a + bX$, $b > 0$

  ako je $\rho(X, Y) = -1$, tada je $Y = a + bX$, $b < 0$

  ako promjenljiva (X, Y) ima dvodimnzionalnu normalnu raspodjelu i ako je $\rho=0$, slučajne promjenljive X i Y su nezavisne

Slučajne promjenljive, zadaci za vježbu

1. Pravilan dinar baca se 2 puta. Ako oba puta padne ista strana, izvodi se još jedno bacanje. Naći raspodjelu slučajne promjenljive X koja predstavlja broj grbova.

2. Prvog dana cijena nekog proizvoda je 5. Svakog dana cijena može da poraste za 1 sa vjerovatnoćom 0.8 ili da se smanji za 1 sa vjerovatnoćom 0.2. Naći raspodjelu cijene S tog proizvoda trećeg dana. Naći vjerovatnoću da je cijena najviše 5.

3. Vjerovatnoća da je proizvod neispravan je 0.2. Radi kontrole kvaliteta iz serije se bira uzorak od 4 elementa, a serija se odbacuje ako je u uzorku bar dva neispravna proizvoda. Naći raspodjelu broja neipravnih proizvoda u uzorku, kao i vjerovatnoću da se serija odbaci.

4. Slučajna promjenljiva X ima normalnu raspodjelu sa parametrima $m = 0$ i $\sigma^2 = 1$. Odrediti vjerovatnoće:

a) $P\{X < 2\}$
b) $P\{0 < X < 1.42\}$
c) $P\{-1.79 < X < 0.54\}$.

5. Dimenzija prečnika šrafa je slučajna promjenljiva sa normalnom raspodjelom sa parametrima $m = 2.5$mm i $\sigma^2 = 0.01$mm^2. Kolika je vjerovatnoća da će prečnik biti između 2.35 i 2.8 mm?

6. Odrediti numeričke karakteristike slučajne promjenljive X koja predstavlja broj grbova u tri bacanja novčića.