PREDAVANJA IZ

OSNOVA TEORIJE VJEROJATNOSTI I MATEMATIČKE STATISTIKE

Održana 2004/05

Miroslav Požek

Literatura:

- J. L. Devore: Probability & Statistics for Engineering and the Sciences, Brooks/Cole, Monterey 2000.
- I. Šošić, V. Serdar: Uvod u Statistiku, Školska knjiga, Zagreb 1994. (Ekonomski fakultet).
- V. Vranić: Vjerojatnost i statistika, Tehnička knjiga, Zagreb 1971.
- I. Pavlić: Statistička teorija i primjena, Tehnička knjiga, Zagreb 1970. (Strojarski fakultet)
- N. Sarapa: Teorija vjerojatnosti, Školska knjiga, Zagreb 1992. (PMF- Matematika)
- D. Ugrin-Šparac: Primijenjena teorija vjerojatnosti I., II., Sveučilišna naklada - Liber, Zagreb, 1975/1976. (FER)
- W. Feller: An Introduction to Probability Theory and its Applications I., II., Willey, New York 1968/1971.

Kolokviji:

I. kolokvij nakon 7. predavanja

II. kolokvij na kraju

Ako su ocjene 4 ili 5 $\Rightarrow$ oslobađanje od pismenog ispita

Predavanja se mogu naći na adresi http://www.phy.hr/otvims

Konzultacije: petak 10-11, soba 215

Čemu služi teorija vjerojatnosti i statistika?

- igre na sreću
- politika i marketing
- gospodarstvo
- zdravstvo
- genetika
- fizika
  - teorijska fizika
    - statistička fizika
    - kvantna fizika
  - eksperimentalna fizika
    - obrada rezultata mjerenja - procjena pogrešaka
    - provjera hipoteza

I. UVOD

Svijet u kojem živimo i koji želimo razumjeti pun je različitosti i neodređenosti. Kad to ne bi bilo tako, bilo bi malo prostora za granu znanosti zvanu Statistika. Statistika nas uči kako donositi ispravne zaključke i odluke u svijetu neodređenosti.
U svakodnevnom životu prikupljamo mnoštvo podataka. Statistika nam daje metode za njihovo organiziranje i sažeto prikazivanje te izvlačenje zaključaka na osnovu informacija sadržanih u tim podacima.
Ime 'statistika' dolazi od latinskog *status* - stanje, situacija.
Povijesno, državni interes - popisi vojnih obveznika u Kini, Perziji, Grčkoj,...

Državni zavod za statistiku http://www.dzs.hr

http://www.dzs.hr/StatInfo/Zaposl.htm

http://www.dzs.hr/StatInfo/RADSNAGA.htm *(anketa na reprezentativnom uzorku)*

primjeri u meteorologiji: http://meteo.hr

OSNOVNI STATISTIČKI POJMOVI

Populacija - dobro definiran skup objekata koji nas zanima.

Populacija može biti konačna ili beskonačna, realna ili hipotetična.
Primjeri:
1. Svi zaposleni na području Krapinsko-Zagorske županije u 2001.
2. Svi stanovnici Republike Hrvatske 31.3.2001.
3. Sve baterije proizvedene u nekoj tvornici u jednom mjesecu.
4. Svi automobili koji će se proizvesti na osnovu 5 prototipova.
5. Svi mogući rezultati nekog mjerenja.

Poznavanje cijele populacije - **cenzus**

Rijetko možemo imati cenzus (preskupo ili neizvedivo)

Uzorak - podskup populacije koji uzimamo na unaprijed određen način

Primjeri:
1. Iz populacije svih studenata uzimamo samo jednu godinu.
2. U skladištu cigala uzimamo jedan red.
3. Od automobila neke marke uzimamo samo one prodane u Zagrebu 1999.
4. Obavimo 10 mjerenja mase nekog predmeta.

Da bismo dobili reprezentativan uzorak, potreban je oprez.

Varijabla - obilježje čija vrijednost se može mijenjati od objekta do objekta.

Možemo promatrati jedno obilježje (npr. visina studenta) pa govorimo o jednodimenzionalnom skupu podataka ili više obilježja. Npr. promatramo li visinu i težinu studenata, imat ćemo dvodimenzionalan skup podataka.

Statistiku dijelimo na opisnu i induktivnu.

Opisna statistika

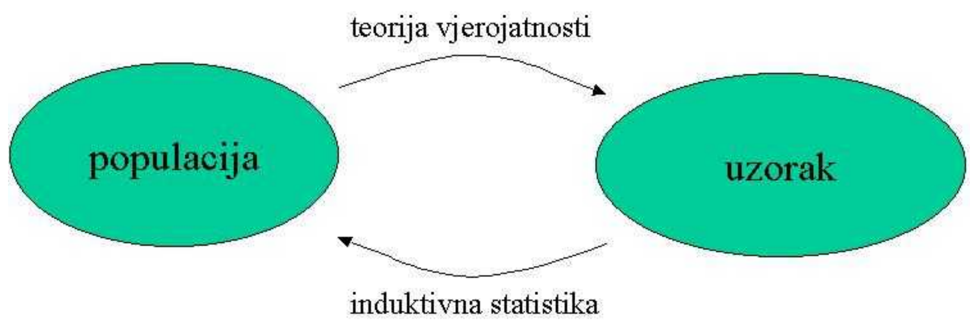
Istraživač je skupio podatke i želi ih organizirati te sažeto opisti važna svojstva.
Grafičke metode: histogrami, krivulje, kružni dijagrami,...
Numeričke vrijednosti: prosjek, standardna devijacija, korelacijski koeficijent,...

Induktivna statistika (statističko zaključivanje)

Istraživač je proučio uzorak i iz njegovih svojstava želi izvesti zaključke o cijeloj populaciji.
Da bismo to razumjeli, moramo najprije razumjeti neodređenosti povezane s uzimanjem uzorka iz populacije.

Teorija vjerojatnosti

Prepostavljamo da znamo svojstva populacije i želimo predvidjeti svojstva uzorka



U ovom kolegiju najprije ćemo ukratko naučiti metode opisne statistike, zatim ćemo proučavati teoriju vjerojatnosti, a tek nakon toga induktivnu statistiku.

II. METODE OPISNE STATISTIKE

Grafičke i tablične metode

S-L dijagram (Stem and leaf, stabljika i list)

primjer: Na 140 američkih koledža provedena je anketa o 'tulum-pijancima'. Kao tulum-pijanac opisan je onaj student koji u nizu popije 4 ili više alkoholnih pića. Postotak tulum-pijanaca na koledžima dan je u sljedećem popisu:

37	27	36	43	49	57
61	27	37	44	49	46
11	28	68	55	50	26
14	29	37	44	50	58
17	29	64	45	51	59
29	4	37	66	51	60
18	29	38	45	52	42
51	30	59	46	52	61
19	64	57	59	52	62
21	31	16	57	18	64
22	39	38	46	53	65
66	31	39	13	37	43
22	32	39	46	53	65
23	56	39	47	54	46
24	32	41	67	47	66
52	33	41	47	55	67
25	48	42	58	56	48
56	44	61	48	56	39
28	34	42	49	41	36
26	35	46	33	56	44
61	35	42	58	48	33
48	35	18	26	57	
26	36	27	38	38	
34	36	58	42	15	

Da bismo iz ove nepregledne gomile brojeva stekli dojam o raspro-stranjenosti pojave, izradimo S-L dijagram:

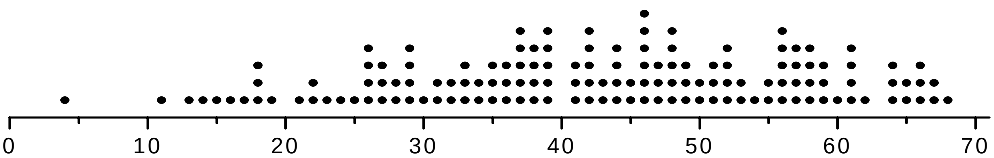
1. odaberemo jednu ili više početnih znamenki za 'stabljiku'
2. u stupac popišemo sve moguće stabljike
3. desno od svake stabljike popišemo sve pripadajuće 'listove'

0	4
1	1345678889
2	1223456666777889999
3	011223334455566667777888899999
4	111222223344445566666677788888999
5	00111222233455666667777888899
6	01111244455666778

Prikaz pomoću točaka

1. Nacrtamo dio brojevnog pravca koji obuhvaća sve moguće podatke
2. Svako opažanje prikažemo točkom iznad odgovarajućeg položaja

Za podatke iz gornjeg primjera točkasti prikaz izgleda ovako:



Histogrami

Postupak crtanja histograma različit je za diskretne i kontinuirane varijable.

Diskretna varijabla - ona za koji je skup mogućih vrijednosti konačan ili prebrojiv

-najčešće se radi o prebrojavanju

Kontinuirana varijabla - skup mogućih vrijednosti jest cijeli interval na brojevnom pravcu

-najčešće rezultat mjerenja

Primjer diskretne varijable:

Bilježio sam starost djece koja se igraju u parku. Varijabla je X=starost djeteta.

Rezultati ankete za 48 djece (starost u godinama):

5	5	8	9
10	11	11	12
3	5	8	9
4	6	18	10
9	11	9	9
11	6	8	3
4	4	1	7
14	10	8	10
4	9	9	5
10	6	11	12
5	7	9	5
8	10	4	20

Moguće vrijednosti varijable X su 1,2,3,...18,19,....,70,...130,.... Dakle, radi se o diskretnoj varijabli.

Def.: **Frekvencija** f_i neke određene vrijednosti x_i varijable X jest broj pojavljivanja te vrijednosti u promatranom skupu podataka.

Vrijedi: $\sum_{\text{svi } i} f_i = N$

gdje je N broj svih podataka u skupu.

Tablični prikaz svih frekvencija jest **raspodjela frekvencija**.

U našem primjeru raspodjela frekvencija dana je tablicom:

x_i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
f_i	1	0	2	5	6	3	2	5	8	6	5	2	0	1	0	0	0	1	0	1

Def.: **Relativna frekvencija** f_{ri} neke određene vrijednosti x_i varijable X jest frekvencija podijeljena s ukupnim brojem podataka $N = \sum_{\text{svi } i} f_i$:

$$f_{ri} = \frac{f_i}{N}$$

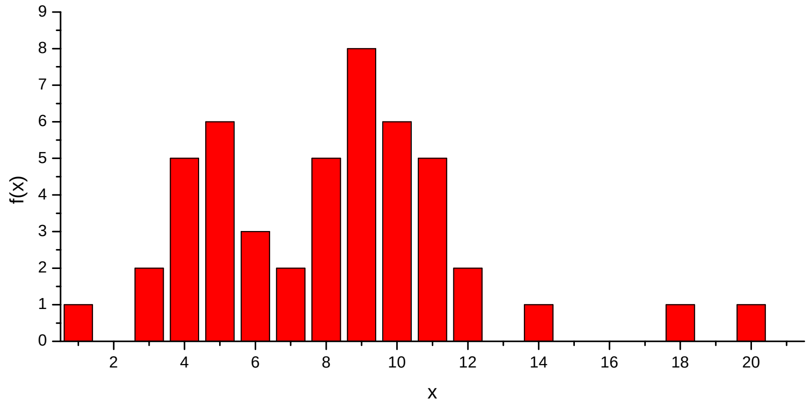
U našem primjeru je $N = 48$ pa je raspodjela relativnih frekvencija dana tablicom:

x_i	f_{ri}	f_{ri} (%)
1	0,0208	2,08
2	0	0,00
3	0,0417	4,17
4	0,1042	10,42
5	0,125	12,50
6	0,0625	6,25
7	0,0417	4,17
8	0,1042	10,42
9	0,1667	16,67
10	0,125	12,50
11	0,1042	10,42
12	0,0417	4,17
13	0	0,00
14	0,0208	2,08
15	0	0,00
16	0	0,00
17	0	0,00
18	0,0208	2,08
19	0	0,00
20	0,0208	2,08

Crtanje histograma (za diskretnu raspodjelu):

1. odredi frekvencije ili relativne frekvencije
2. na apscisi označi moguće vrijednosti varijable X
3. nacrtaj pravokutnik visine f_i *ili* f_{ri}.

Podaci iz našeg primjera prikazani su histogramom:



Konstrukcija histograma za kontinuiranu varijablu

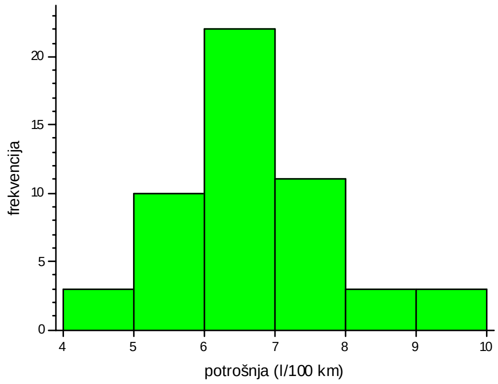
Primjer: potrošnja benzina (u litrama na 100 km):

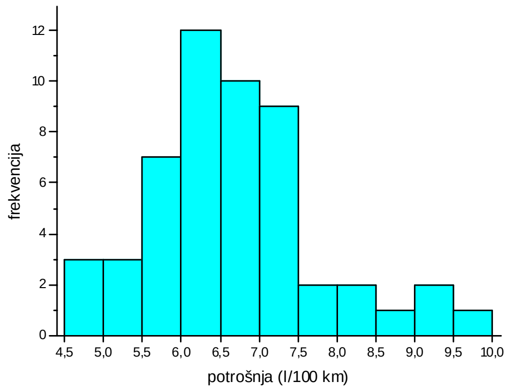
6,25	5,93	7,8	4,95	9,2	8,57
6,82	7,43	5,78	5,46	6,54	7,02
6,78	4,75	5,32	7,11	5,66	5,99
6,87	8,35	7,66	7,23	6,58	6,92
6,32	7,08	5,98	6,25	5,45	6,72
6,38	6,9	9,87	6,23	6,52	6,43
6,12	5,81	6,37	7,23	7,46	8,06
6,09	5,82	4,99	6,32	6,51	6,49
9,49	6,39				

Konstrukcija histograma:

1. Podijelimo apscisu na prikladan broj razreda (ekvidistantno ili neekvidistantno)
2. odredimo frekvencije
3. crtamo pravokutnike
   - za ekvidistantne razrede visina = f_{ri}
   - za neekvidistantne razrede visina = f_{ri}/širina

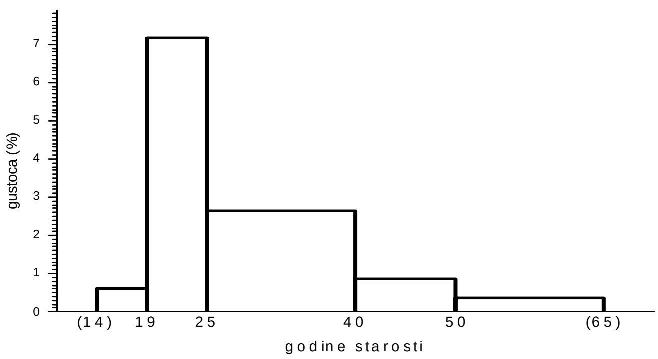
Primjer potrošnje goriva - ekvidistantni razredi





Primjer neekvidistantnih razreda:
Osobe prijavljene zavodima za zapošljavanje u Hrvatskoj 1990:

dob	broj osoba (u tisućama)	f_{ri} (%)
(14)-19	6,1	3,12
19-25	84,0	42,99
25-40	77,5	39,66
40-50	16,6	8,50
50-(65)	11,2	5,73
UKUPNO	195,4	100,00



Napomene:
-površina pravokutnika je frekvencija danog razreda
-ukupna površina relativnih frekvencija je 1

Prema obliku histogrami mogu biti:
- -unimodalni
- -bimodalni
- -višemodalni

ili:
- -simetrični
- -pozitivno nagnuti
- -negativno nagnuti

Mjere položaja

Grafičkim prikazom stekli smo impresiju, ali želimo izvući neke brojevekoji nam karakteriziraju skup podataka.

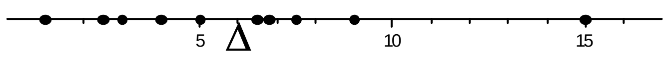
Neka se naš skup sastoji od brojeva $x_1, x_2, \dots x_n$ (neki mogu biti i isti). Najprije nas zanima položaj tiog skupa, posebno središte.

Neka skup $x_1, x_2, \dots x_n$ predstavlja uzorak iz veće populacije!

Def.: **Srednja vrijednost** $\bar{x}$ uzorka $x_1, x_2, \dots x_n$ je

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Fizikalno značenje - težište!



Def.: **Srednju vrijednost populacije** označavamo s μ.

U induktivnoj statistici najčešće želimo procijeniti μ na osnovu uzorka.

Najveći nedostatak srednje vrijednosti kao mjere položaja jest taj da podliježe utjecaju ekstremnih vrijednosti u uzorku.

Def.: **Medijan** $\tilde{x}$ određujemo tako da sva opažanja poredamo po veličini. Ako je broj opažanja neparan, medijan je vrijednost $\left(\frac{n+1}{2}\right)$-tog opažanja. Ako je broj opažanja paran, medijan je srednja vrijednost $\left(\frac{n}{2}\right)$-tog i $\left(\frac{n}{2}+1\right)$-og opažanja.

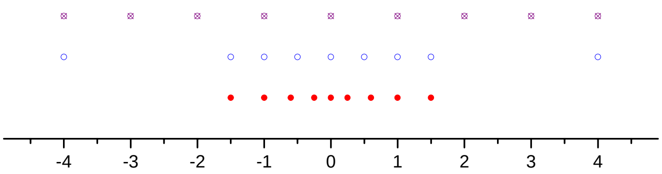
Kvartili dijele skup opažanja na četiri dijela.

Percentili dijele skup opažanja na 100 dijelova.

"Ošišani" prosjek - odbacimo jednak broj ekstrema s obje strane pa računamo srednju vrijednost.

Mjere raspršenja

Uzorci s istim mjerama položaja (prosjek, medijan) mogu se bitno razlikovati. Slika prikazuje primjer tri različita skupa istog prosjeka i medijana:



Osim središta uzorka, zanima nas i koliko je raspršen. Najjednostavnija mjera raspršenja je

Def.: **Raspon uzorka** - razlika između najveće i najmanje vrijednosti u uzorku

Mana: ovisi samo o ekstremima.

Def.: **Odstupanje od prosjeka** pojedine vrijednosti u uzorku: $x_i - \bar{x}$

Želimo izraziti mjeru svih odstupanja u jednom broju.
Prva ideja: zbrojimo sva odstupanja:

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})$$

Nažalost, to je jednako nuli (pozitivna i negativna odstupanja se poništavaju).

Druga ideja: zbrojimo apsolutne vrijednosti svih odstupanja

$$\text{prosječno apsolutno odstupanje} = \frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}|$$

To će nam dati mjeru odstupanja, ali stvoriti matematičke komplikacije.

Treća ideja: zbrojimo sve kvadrate odstupanja

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Def.: **Varijanca uzorka**

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Def.: **Standardna devijacija uzorka**

$$s = \sqrt{s^2}$$

Zašto smo uzeli *n-1*, a ne *n*?

Kad proučavamo cijelu populaciju od *N* elemenata definiramo varijancu populacije

Def.: **Varijanca populacije**

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$$

gdje je μ prosjek populacije.

Vrlo rijetko poznajemo μ pa pokušavamo iz uzorka odrediti μ i σ. Međutim, prosječni kvadrat odstupanja $(x_i - \bar{x})^2$ u uzorku je manji od nepoznatog $(x_i - \mu)^2$. Da bismo to korigirali, u nazivnik izraza za varijancu stavljamo *n-1* umjesto *n*. Na taj način možemo s pomoću varijance uzorka procijeniti varijancu populacije.
Korekcija dolazi zbog umanjenog broja stupnjeva slobode u uzorku.

To ćemo pokazati pred kraj semestra.

III. TEORIJA VJEROJATNOSTI

Proučavanje nasumičnosti i neodređenosti.
Jezik vjerojatnosti u svakodnevnom životu:

- *Cijena Plivinih dionica vjerojatno će porasti.*
- *Imamo 50% šansi da ne izgubimo utakmicu.*
- *Ovdje ima toliko osigurača da će sigurno bar jedan pregorjeti.*
- *Naš predsjednički kandidat ima 80% šanse za pobjedu.*

III.1. OSNOVNI POJMOVI I DEFINICIJE

Terminologija

Pokus= bilo koji postupak ili proces koji rezultira opažanjem.
pokus 1: bacanje dvije kocke
pokus 2: mjerenje napona baterije. Ponavljanje pokusa.
pokus 3: izvlačenje kuglice iz kutije. Ponavljanje pokusa
slučajni pokus
Ishod= **rezultat pokusa. Ishodi su nerazloživi i međusobno se isključuju!**
primjer 1: izvučena je crvena kuglica.
primjer 2: dvaput je palo pismo.

Prostor elementarnih događaja Ω=skup svih ishoda nekog pokusa.

primjer 1: bacamo dva različita novčića. Prostor elementarnih događaja je
Ω={(P,P),(P,G),(G,P),(G,G)}

primjer 2: bacamo novčić dok ne padne pismo. Prostor elementarnih događaja je
Ω={(P),(GP),(GGP),(GGGP),...} beskonačan, ali prebrojiv

Događaj=svaki podskup od Ω.

Elementaran događaj=jednočlani podskup
Složeni događaj=višečlani podskup

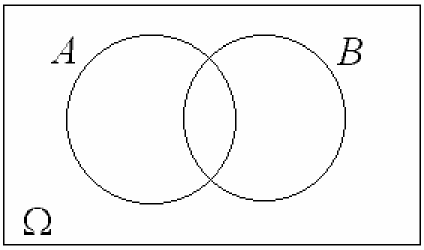
Primjer 1: Pokus bacanja dva novčića. Definiramo događaje A i B:
A='pali su jednako'={(P,P);(G,G)}
B='pao je barem jedan grb'={(P,G);(G,P);(G,G)}
Događaji A i B se međusobno ne isključuju.

Primjer 2: Pokus bacanja novčića dok ne padne pismo. Definiramo događaj A:
A='bacali smo barem četiri puta'={(GGGP),(GGGGP),(GGGGGP),....}

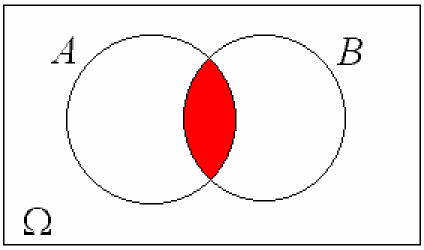
Siguran događaj = Ω
Nemoguć događaj = ∅ (prazan skup)

Relacije iz teorije skupova

Budući da smo događaje definirali kao skupove, moći ćemo primjenjivati relacije i operacije iz teorije skupova. Za zorno prikazivanje tih relacija upotrebljavamo Vennove dijagrame. Prostor elementarnih događaja nekog pokusa označimo s Ω (pravokutnik), a ishodi tog pokusa predstavljeni su točkama unutar pravokutnika.

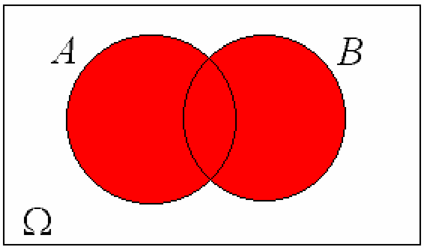


Složeni događaji *A* i *B*



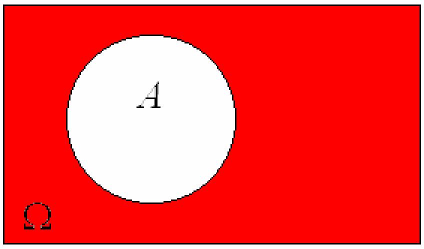
Presjek događaja *A* i *B*

$A \cap B$



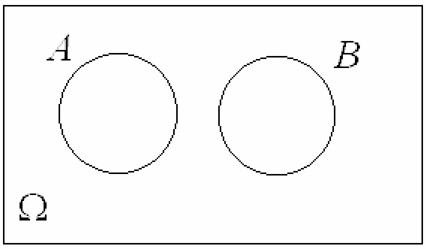
Unija događaja *A* i *B*

$A \cup B$



Komplement događaja *A*

$\overline{A}$



Međusobno isključivi događaji *A* i *B*

$A \cap B = \emptyset$

Definicija vjerojatnosti

Definicija **a priori**

Def:

Neka imamo slučajni pokus s konačno mnogo elementarnih događaja i neka su svi ti elementarni događaji jednako mogući. Tada je vjerojatnost proizvoljnog događaja A vezanog uz taj pokus dana brojem elementarnih događaja povoljnih za taj događaj n_A podijeljenim s ukupnim brojem elementarnih događaja n:

$$n = k(\Omega) \quad ; \quad n_A = k(A) \quad \Rightarrow \quad P(A) = \frac{n_A}{n}$$

Manjkavosti ove definicije:

- primjenjiva je samo na konačne skupove
- definicija je kružna. Rabimo izraz 'jednako mogući' što znači 'jednako vjerojatni' da bismo definirali vjerojatnost.

Primjer: Bacamo novčić. Kolika je vjerojatnost da padne pismo?

1/2 !

Na osnovu čega to zaključujemo?

- iskustvo
- pretpostavljena simetrija novčića

Primjer 2: Bacamo kocku. A priori vjerojatnost da padne trica je P(3)=1/6 . Oprez! Još u staroegipatskim piramidama pronađene su "nepoštene" kocke.

Postoji i druga definicija vjerojatnosti. Da bismo do nje došli, definirat ćemo još dva pojma.

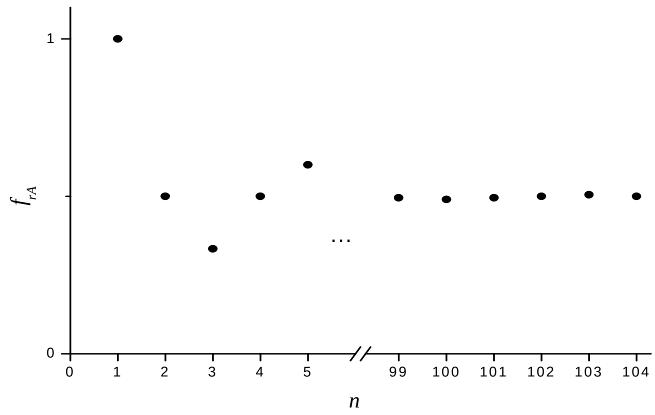
Frekvencija

Ponovimo slučajni pokus n puta. Neka se događaj A pojavio točno n_A puta. Tada broj n_A nazivamo **frekvencijom** događaja A (f_A), a broj $f_{rA} = \frac{n_A}{n} = \frac{f_A}{n}$ je **relativna frekvencija** tog događaja. Vrijedi

$$0 \leq f_{rA} \leq 1$$

Statistička stabilnost relativnih frekvencija

Ako se prilikom velikog broja ponavljanja slučajnog pokusa, relativne frekvencije događaja A grupiraju oko nekog broja, kažemo da je relativna frekvencija statistički stabilna.



<u>Definicija vjerojatnosti **a posteriori**</u>

Def:

Ako slučajni pokus zadovoljava uvjet statističke stabilnosti relativnih frekvencija, onda se **vjerojatnost a posteriori** proizvoljnog događaja A definira kao realan broj $P(A) \in [0,1]$ oko kojeg se grupiraju relativne frekvencije tog događaja.

Ovu definiciju zovemo još i 'Zakon velikih brojeva'.
Problemi:

- kako provjeriti stabilnost?
- Kako odrediti vjerojatnost jednog događaja? (Velika je vjerojatnost da sutra neće pasti kiša)

<u>Aksiomatsko zasnivanje teorije vjerojatnosti</u>

A. N. Kolmogorov, 1933.

Poznajemo li prostor elementarnih događaja Ω za neki pokus, svrha teorije vjerojatnosti je da svakom događaju $A \subseteq \Omega$ pridruži broj $P(A)$ koji će biti precizna mjera šanse da se A ostvari. Svako takvo pridruživanje mora zadovoljavati sljedeće aksiome:

A1. Za svaki događaj $A \Rightarrow P(A) \geq 0$

A2. $P(\Omega) = 1$

A3. a) Ako se konačan broj događaja $A_1, A_2, \dots A_n$ međusobno isključuje (tj. ako su u parovima disjunktni), vrijedi:

$$P(A_1 \cup A_2 \cup \ldots \cup A_n) = \sum_{i=1}^{n} P(A_i)$$

b) Ako se prebrojivo beskonačan broj događaja $A_1, A_2, \dots$ međusobno isključuje, vrijedi:

$$P(A_1 \cup A_2 \cup \ldots) = \sum_{i=1}^{\infty} P(A_i)$$

Ovim aksiomima nije jednoznačno određeno pridruživanje. Aksiomi samo isključuju pridruživanja koja nisu konzistentna s našim poimanjem vjerojatnosti.

Primjer: Pošten i nepošten novčić.

<u>Svojstva vjerojatnosti</u>
Propozicije:

P1. $\forall A \Rightarrow P(A) = 1 - P(\overline{A})$

dokaz:

Iz trećeg aksioma. Neka je n=2. Neka je $A_1 = A \quad \Rightarrow \quad A_2 = \overline{A}$

$$A \cup \overline{A} = \Omega$$

$$1 = P(\Omega) = P(A \cup \overline{A}) = P(A) + P(\overline{A}) \qquad \text{Q.E.D.}$$

Primjer:
Znamo da je 85% ljudi Rh-pozitivno. Liječnik čini krvni test na novorođenčadi tako dugo dok ne nađe jedno Rh-negativno. Kolika je vjerojatnost da će morati učiniti barem dva testa?
Rješenje:
Zanima nas događaj B='barem dva testa'={(+-),(++-),(+++-),(++++-),....}
Njegov komplement je $\overline{B}$ ='prvi je Rh-negativan'={(-)}
Lako izračunamo vjerojatnost $P(\overline{B})$=0,15.
Dakle, $P(B)$=0,85.

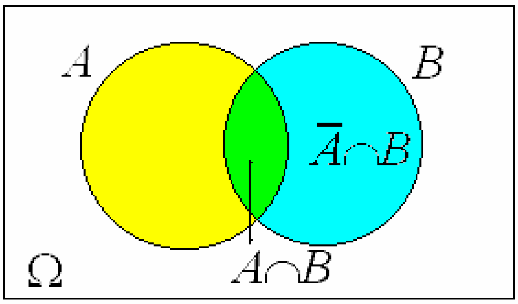
P2. Ako se A i B isključuju, onda je $P(A \cap B) = 0$.

dokaz:

$$A \cap B = \emptyset \quad \Rightarrow \quad \overline{A \cap B} = \Omega \quad \Rightarrow \quad 1 = P(\Omega) = P(\overline{A \cap B}) = 1 - P(A \cap B)$$

P3. Za bilo koja dva događaja A i B vrijedi $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Dokaz:
Ako su disjunktni slijedi iz aksioma.
Ako nisu disjunktni:



$$A \cup B = A \cup (B \cap \overline{A})$$

$$B = (B \cap A) \cup (B \cap \overline{A})$$

Oduzimanjem druge jenadžbe od prve

$$\Rightarrow \quad \begin{aligned} P(A \cup B) &= P(A) + P(B \cap \overline{A}) \\ P(B) &= P(B \cap A) + P(B \cap \overline{A}) \end{aligned}$$

Q.E.D.

P3a. Vrijedi i za više događaja:

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - \\ - P(A \cap B) - P(A \cap C) - P(B \cap C) \\ + P(A \cap B \cap C)$$

Sustavno određivanje vjerojatnosti

Neka neki pokus ima velik broj mogućih ishoda (elementarnih događaja) E_i. Za određivanje vjerojatnosti nekog složenog događaja A, najbolje je najprije odrediti vjerojatnosti svih elementarnih događaja E_i. Mora vrijediti:

$$\forall E_i \quad \Rightarrow \quad P(E_i) \geq 0$$

$$\sum_{\text{svi } i} P(E_i) = 1 \ .$$

Tada je vjerojatnost bilo kojeg složenog događaja A dana zbrajanjem

$$P(A) = \sum_{\text{svi } E_i \text{ iz } A} P(E_i) \ .$$

Primjer:
Nepoštena kocka

$P(E_1) = P(E_3) = P(E_5) = 1/9$

$P(E_2) = P(E_4) = P(E_6) = 2/9$

Zanima nas događaj A='broj manji od 4'

$P(A) = P(E_1) + P(E_2) + P(E_3) = 4/9$

Jednako vjerojatni ishodi

Primjeri:
Izvlačenje karata iz novog špila, bacanje poštene kocke, popunjavanje energetskih nivoa,...

Imamo n mogućih ishoda, a vjerojatnost svakog od njih je

$$P(E_i) = \frac{1}{n} \quad , \quad \forall i$$

Vjerojatnost događaja A je

$$P(A) = \frac{n(A)}{n}$$ (vratili smo se na definiciju vjerojatnosti a priori)

III.2. KOMBINATORIKA

Kad su svi elementarni događaji jednako vjerojatni, određivanje vjerojatnosti složenog događaja svodi se na prebrojavanje.
Kad imamo velik broj elementarnih događaja, moramo upotrebljavati pravila prebrojavanja. Time se bavi kombinatorika.

1. Teorem o uzastopnom prebrojavanju

Proučavamo uređene parove. Neka prvi element uređenog para možemo izabrati na n_1 načina, a za već izabrani prvi element, drugi možemo izabrati na n_2 načina. Tada uređeni par možemo izabrati na $N=n_1 \cdot n_2$ načina.

Primjer 1: Bacamo dvije kockice različite boje. Na koliko različitih načina mogu pasti?

$$n=6 \cdot 6=36$$

Primjer 2: U gradu ima 85 osnovnih škola ravnomjerno raspoređenih u 17 gradskih četvrti. Obitelj s jednim školskim djetetom kupuje stan. Na koliko načina može odabrati četvrt i školu ako: a) nije bitno da škola bude u istoj četvrti; b) škola mora biti u istoj četvrti kao i stan.

a) $N=17 \cdot 85=1445$
b) $N=17 \cdot 5=85$

Teorem vrijedi općenito:

Proučavamo uređene k-torke. Neka prvi element uređene k-torke možemo izabrati na n_1 načina, a za već izabrani prvi element, drugi možemo izabrati na n_2 načina i tako dalje do k-tog koji možemo izabrati na n_k načina. Tada uređenu k-torku možemo izabrati na $N=n_1 \cdot n_2 \cdots n_k$ načina.

Primjer: Obitelj iz prošlog primjera ima još i jedno predškolsko dijete, a u blizini svake škole postoje dva vrtića. Na koliko načina mogu odabrati četvrt, školu i vrtić.

a) $N=17 \cdot 85 \cdot 170=245.650$
b) $N=17 \cdot 5 \cdot 2=170$

2. Permutacije

Iz skupa od n različitih elemenata sastavljamo uređene n-torke. Svaka uređena n-torka zove se **permutacija** n-tog razreda. Broj svih permutacija n-tog razreda je

$$P_n = n \cdot (n-1) \cdot (n-2) \cdots 2 \cdot 1 = n! \quad (en\ faktorijela)$$

Dokaz: Prema teoremu o uzastopnom prebrojavanju, prvi element možemo izabrati na n načina, za već odabrani prvi element, drugi možemo izabrati na (n-1) načina itd.

Pokus: Iz kutije u kojoj su kuglice različitih boja izvlačim jednu po jednu i slažem po redu.

Primjer 2: Na koliko se načina u klupe mogu smjestiti učenici u razredu ako ima 30 mjesta i 30 učenika? ($2{,}65 \cdot 10^{32}$)

Dogovorno: $0! = 1$ (često će nam trebati)

3. Varijacije

Iz skupa od n različitih elemenata sastavljamo uređene k-torke ($k \leq n$). Svaka uređena k-torka zove se **varijacija** k-tog razreda od n elemenata. Broj svih varijacija k-tog razreda je

$$V_n^{(k)} = n \cdot (n-1) \cdot (n-2) \cdots (n-k+1) = \frac{n!}{(n-k)!}$$

Pokus: Iz kutije u kojoj je šest kuglica različitih boja izvlačim tri, jednu po jednu, i slažem po redu.

Primjer 2: Na koliko se načina u razredu u kojem je 30 učenika može odabrati glumačka družina za "Crvenkapicu"? Likovi su Crvenkapica, vuk, baka i lovac.

Rješenje: $V_{30}^{(4)} = \frac{30!}{26!} = 30 \cdot 29 \cdot 28 \cdot 27 = 657.720$

4. Kombinacije

Iz skupa od n različitih elemenata odabiremo k-člane podskupove ($k \leq n$) (redoslijed elemenata u skupu nije bitan). Svaki k-člani podskup zove se **kombinacija** k-tog razreda od n elemenata. Broj svih kombinacija k-tog razreda je

$$K_n^{(k)} = \frac{n!}{k!(n-k)!} = \binom{n}{k}$$

Pokus: Iz kutije u kojoj je šest kuglica različitih boja izvlačim tri odjednom. Ne slažem ih nego stavljam u tanjurić.

Primjer 2: Koliko postoji različitih načina ispunjavanja listića lota 7/39?

Rješenje: $K_{39}^{(7)} = \frac{39!}{7! \cdot 32!} = \frac{39 \cdot 38 \cdot 37 \cdot 36 \cdot 35 \cdot 34 \cdot 33}{7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2} = 15.380.937$

Primjer 3:

Binomni razvoj:

$$(a+b)^n = (a+b) \cdot (a+b) \cdot (a+b) \cdots (a+b)$$

Članovi u razvoju imat će elemente oblika a^n, $a^{n-1}b$, $a^{n-2}b^2$,..., b^n. Prvi član možemo dobiti na $\binom{n}{0} = 1$ načina (broj načina da se ni iz jedne zagrade ne uzme b). Drugi član možemo dobiti na $\binom{n}{1} = n$ načina (broj načina da se iz točno jedne zagrade uzme b), itd. Binomni razvoj (razvoj n-te potencije binoma) možemo pisati:

$$\boxed{(a+b)^n = \sum_{i=0}^{n} \binom{n}{i} a^{n-i} b^i}$$

Dosad smo radili permutacije, varijacije i kombinacije bez ponavljanja. Ako je dopušteno više puta izvući isti element, imamo dodatne mogućnosti.

5. Varijacije s ponavljanjem

Iz skupa od n različitih elemenata sastavljamo uređene k-torke, ali tako da se elementi mogu i ponavljati (k može biti i veće od n). Svaka takva uređena k-torka zove se **varijacija s ponavljanjem** k-tog razreda od n elemenata. Broj varijacija s ponavljanjem je

$$\overline{V}_n^{(k)} = n^k$$

Pokus: Iz kutije u kojoj je šest kuglica različitih boja izvlačim jednu, zapišem koja je bila, vratim u kutiju i ponovno izvlačim. Tako tri puta.

Primjer 2: Sportska prognoza (Toto 13). Koliko mogućnosti?

Rješenje: $V_3^{(13)} = 3^{13} = 1.594.320$

6. Permutacije s ponavljanjem

Imamo skup od n elemenata od kojih je m_1 jedne vrste, m_2 druge vrste, ... , m_k k-te vrste. Sastavljamo uređene n-torke. Svaka takva uređena n-torka zove se **permutacija s ponavljanjem**. Broj permutacija s ponavljanjem je

$$\overline{P}_n^{m_1, m_2, \ldots, m_k} = \frac{n!}{m_1! m_2! \cdots m_k!}$$

Pokus: U kutiju stavim 7 kuglica i to 3 crvene, 2 zelene, jednu crnu i jednu bijelu. Izvlačim ih jednu po jednu i slažem po redu.

Primjer 2: Koliko različitih riječi se može napraviti od slova u riječi BANANA?

Rješenje: $\overline{P}_6^{3,2} = \frac{6!}{3! \cdot 2!} = 60$

7. Kombinacije s ponavljanjem

Pokus: U kutiji imam 4 kuglice različite boje. Izvlačim kuglicu, zapišem i vratim u kutiju. Tako tri puta. Nije važan redoslijed!

Imamo skup od n različitih elemenata i neki broj k. Svaka neuređena k-torka elemenata iz skupa u kojoj se elementi mogu i ponavljati (redoslijed elemenata u k-torci nije bitan) zove se **kombinacija s ponavljanjem** k-tog razreda od n elemenata. Broj svih kombinacija s ponavljanjem k-tog razreda od n elemenata je

$$\overline{K}_n^{(k)} = \binom{n+k-1}{k} = K_{n+k-1}^{(k)}$$

"Dokaz":
Pokažimo da ova formula vrijedi za gornji pokus (kombinacije s ponavljanjem trećeg razreda od četiri elementa)! Popišimo sve kombinacije i to tako da brojevi budu u poredani po iznosu:

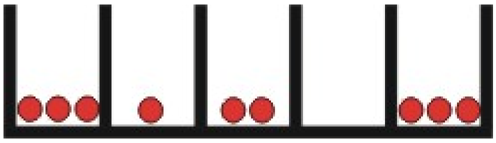
[1,1,1]	[1,2,2]	[1,3,4]	[2,2,4]	[3,3,3]
[1,1,2]	[1,2,3]	[1,4,4]	[2,3,3]	[3,3,4]
[1,1,3]	[1,2,4]	[2,2,2]	[2,3,4]	[3,4,4]
[1,1,4]	[1,3,3]	[2,2,3]	[2,4,4]	[4,4,4]

Tih kombinacija ima 20. U gore popisanim trojkama, dodajmo prvom članu nula, drugom članu 1, a trećem članu 2. Na taj način dobit ćemo 20 tročlanih skupova s različitim elementima:

{1,2,3}	{1,3,4}	{1,4,6}	{2,3,6}	{3,4,5}
{1,2,4}	{1,3,5}	{1,5,6}	{2,4,5}	{3,4,6}
{1,2,5}	{1,3,6}	{2,3,4}	{2,4,6}	{3,5,6}
{1,2,6}	{1,4,5}	{2,3,5}	{2,5,6}	{4,5,6}

To su upravo sve kombinacije bez ponavljanja trećeg razreda od 6 elemenata. Dakle, malo prije smo definirali bijekciju između skupa svih kombinacija s ponavljanjem 3. razreda od 4 elementa i skupa svih kombinacija bez ponavljanja 3. razreda od 6 elemenata. Budući da postoji bijekcija, ti su skupovi jednakobrojni pa vrijedi $\overline{K}_4^3 = K_6^3$. Općenito, za bilo koje n i k može se pokazati da postoji bijekcija između skupa kombinacija s ponavljanjem k-tog razreda od n elemenata i skupa kombinacija bez ponavljanja k-tog razreda od $n+k$-1 elementa.

Pokus 2:
Umjesto da iz jedne kutije u kojoj je n različitih kuglica odabirem k kuglica, ali s vraćanjem u kutiju, pokus možemo napraviti tako da imamo k jednakih kuglica i n različitih kutija. Bacam kuglice u kutije i bilježim u koju kutiju je pala kuglica. (U stvari, ne moram bilježiti, dovoljno je da kuglice na kraju prebrojim u kutijama.) Svaki način kako se k kuglica može rasporediti u n kutija predstavlja jednu kombinaciju s ponavljanjem. To možemo iskoristiti da na drugi način izvedemo broj kombinacija s ponavljanjem. Zamislimo da imamo n kutija poredanih u niz i da u njih bacamo k kuglica. Primjer na slici prikazuje n=5 kutija i k=9 kuglica.



Problem možemo razmatrati kao da slažemo kuglice i pregrade među kutijama. Imamo n-1=4 jednake pregrade i k=9 jednakih kuglica. Dakle, imamo permutacije $n+k$-1=13 elemenata od kojih su 4 jedne vrste i 9 druge vrste pa je njihov broj:

$$\overline{P}_{13}^{9,4} = \frac{13!}{9!\cdot 4!}\ .$$

Općenito, vrijedi

$$\overline{K}_n^k = \binom{n+k-1}{k} = \frac{(n+k-1)!}{k!(n-1)!} = \overline{P}_{n+k-1}^{k,n-1}\ .$$

Stirlingova formula

Za približno izračunavanje faktorijela velikih brojeva možemo primijeniti aproksimativnu Stirlingovu formulu:

$$\boxed{n! \approx n^n e^{-n} \sqrt{2\pi n}}$$

Tu formulu ćemo izvesti u jednom od kasnijih predavanja.

Primjeri iz kombinatorike

1. Kolika je vjerojatnost da ćete u podjeli 13 karata u bridžu dobiti samo crvene?
   Rješenje:

   Broj svih mogućih podjela 13 karata je $\binom{52}{13}$.

   Crvenih karata ima 26. Broj mogućih podjela 13 crvenih karata je $\binom{26}{13}$.

   Vjerojatnost svih crvenih karata u ruci je:

   $$P(c) = \frac{\text{broj povoljnih kombinacija}}{\text{ukupni broj kombinacija}} = \frac{\binom{26}{13}}{\binom{52}{13}} = \frac{39!\cdot 26!}{52!\cdot 13!} = 1{,}64\cdot 10^{-5}$$

   Ako svaki dan igrate jednu partiju bridža, to će vam se dogoditi jednom u 167 godina.
2. Na koliko načina možemo poredati 10 ljudi oko okruglog stola?
   a) ako postoji jedan 'kraljevski' stolac?
   b) ako su svi stolci ekvivalentni?
   c) ako imamo 5 plesnih parova?

Rješenje:
   a) 10!
   b) 10!/10=9!
   c) 5!

3. Koliko cijelih brojeva manjih od 10^8 sadrži točno 5 sedmica? Kolika je vjerojatnost takvog broja?
   Rješenje:

   Pet mjesta za znamenku '7' možemo odabrati na $\binom{8}{5} = 56$ načina.

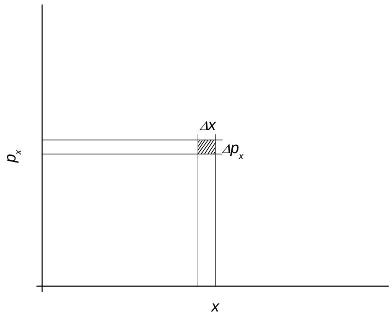
   Na preostala tri mjesta stavljamo ostalih 9 znamenaka. To možemo učiniti na $\overline{V}_9^{(3)} = 9^3 = 729$ načina.

   Brojeva s 5 sedmica ima 56·729=40.824 .

   Vjerojatnost takvog broja je $\frac{40.824}{10^8 - 1} = 0{,}000408$

Malo statističke fizike:

4. Kinetička teorija plinova.
   Energija neke molekule plina određena je njezinim položajem i impulsom
   $E = E_p(x, y, z) + E_k(p_x, p_y, p_z)$
   Čestice koje se u šesterodimenzionalnom faznom prostoru nalaze u istoj točki, imaju istu energiju. Da bismo to ilustrirali, nacrtajmo dvodimenzionalni fazni prostor x, p_x (za jednodimenzionalni problem, npr. harmonički oscilator):



   Za molekule koje se nalaze u istom "pretincu" fazne površine $\Delta x \cdot \Delta p_x$ kažemo da su u istom energetskom stanju. Neka u našem faznom prostoru ima n pretinaca, a u plinu koji proučavamo ima r molekula i neka te molekule možemo označiti brojevima (tj. možemo ih razlikovati). Tada imamo $\overline{V}_n^r = n^r$ raznih raspodjela. <u>Maxwell-Boltzmannova pretpostavka</u> je da su sve te raspodjele jednako vjerojatne. Vjerojatnost jedne takve raspodjele je $p = \frac{1}{n^r}$.

5. Bozoni
   U kvantnoj fizici čestice ne možemo razlikovati. Ako imamo n kvantnih stanja (pretinaca) i r čestica koje sve mogu ući u isti pretinac (takve čestice nazivamo bozonima), onda je broj raspodjela dan s
   $$\overline{K}_n^r = \binom{n+r-1}{r}$$
   <u>Bose-Einsteinova pretpostavka</u> je da su sve te raspodjele jednako vjerojatne. Vjerojatnost jedne takve raspodjele je $p = \frac{1}{\binom{n+r-1}{r}}$. Fotoni su takve čestice.

6. Fermioni
   Ako imamo n kvantnih stanja (pretinaca) i r čestica koje ne mogu ući u isti pretinac s nekom drugom česticom (takve čestice nazivamo fermionima), onda je broj raspodjela dan s
   $$K_n^r = \binom{n}{r}$$
   <u>Fermi-Diracova pretpostavka</u> je da su sve te raspodjele jednako vjerojatne. Vjerojatnost jedne takve raspodjele je $p = \frac{1}{\binom{n}{r}}$. Takve čestice su protoni, neutroni, elektroni,...

III.3. UVJETNA VJEROJATNOST I NEZAVISNOST

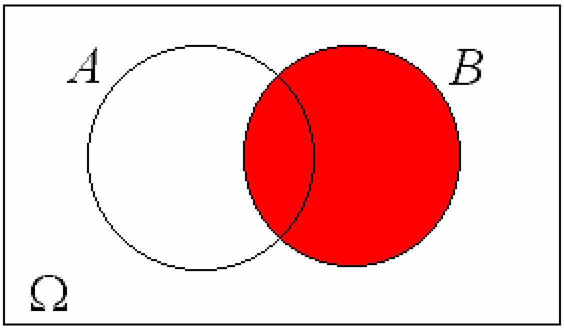
Uvjetna vjerojatnost

Često vjerojatnost nekog događaja ovisi o tome što već znamo o tijeku pokusa. Zamislimo da u nekom pokusu promatramo događaje A i B, čije vjerojatnosti su $P(A)$ i $P(B)$. Međutim, ako smo već saznali da se je događaj A dogodio, onda vjerojatnost događaja B ne mora biti ista kao kad ne znamo tu informaciju. Tu novu vjerojatnost događaja B nazivamo

Uvjetna vjerojatnost događaja B ako se je dogodio A i pišemo

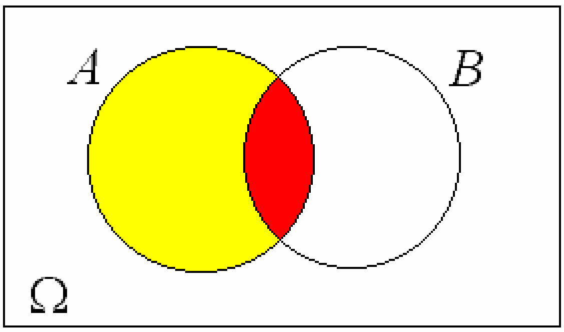
$$P(B|A)$$

Prije no što znamo ishod događaja A, vjerojatnost događaja B je $P(B)$. To je vjerojatnost svih elementarnih događaja u skupu B podijeljena s vjerojatnošću sigurnog događaja $P(\Omega)$:



$$P(B)$$

Međutim, ako se je A već dogodio, onda on predstavlja novi siguran događaj pa je vjerojatnost događaja B u stvari vjerojatnost svih elementarnih događaja iz skupa $A \cap B$ podijeljena s vjerojatnošću događaja A:



$$P(B|A)$$

Tako dobivamo izraz za uvjetnu vjerojatnost:

$$\boxed{P(B|A)=\frac{P(A\cap B)}{P(A)}}$$

Primjer: Dva stroja proizvela su 19 proizvoda i bacila ih u istu kutiju. Stroj A proizveo je 9 ispravnih i 1 neispravan proizvod, dok je stroj B proizveo 7 ispravnih i 2 neispravna proizvoda.
Kontrolor izvlači jedan proizvod iz kutije. a) Kolika je vjerojatnost da izvuče proizvod iz stroja B? b) Ako je izvukao neispravan proizvod, koliko je vjerojatno da je on iz stroja B?
Rješenje:

Napravimo tablicu:

	I	N
A	9	1
B	7	2

a) $P(B)=\frac{9}{19}=47\,\%$

b) $P(B|N)=\frac{P(B\cap N)}{P(N)}=\frac{2}{3}=67\,\%$

Potpun sistem događaja

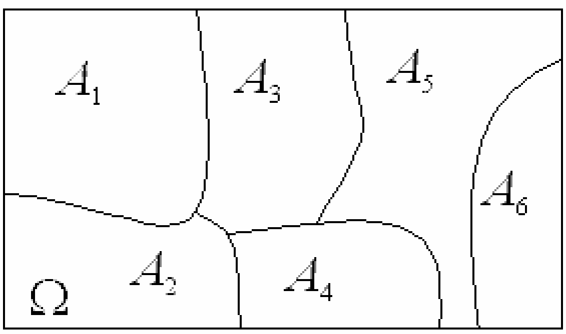
Neka za događaje A_i, (i=1,2,...) vrijedi

$A_i \neq \emptyset \quad , \quad \forall i$

$A_i \cap A_j = \emptyset \quad , \quad \forall i,j \quad i\neq j$ (međusobno se isključuju)

$\bigcup\limits_{\text{svi } i} A_i = \Omega$ (prekrivaju cijeli prostor).

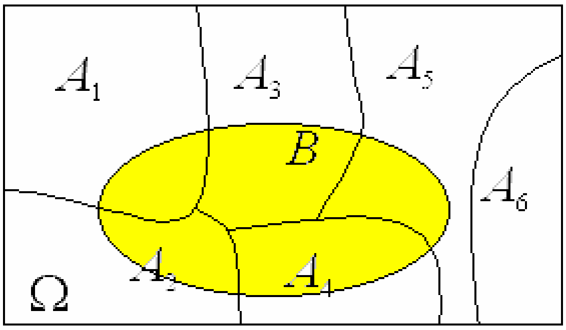
Tada kažemo da skupovi A_i čine **potpun sistem događaja**.



Zakon totalne vjerojatnosti

Neka A_i čine potpun sistem događaja. Tada za bilo koji događaj B vrijedi
$$P(B) = \sum_i P(B|A_i) \cdot P(A_i)$$

Dokaz:



Iz slike vidimo da je $P(B) = \sum_i P(A_i \cap B)$ pa iz definicije uvjetne vjerojatnosti slijedi gornja tvrdnja.

Bayesov teorem

Neka A_i čine potpun sistem događaja. Neka je B neki događaj. Tada vrijedi

$$P(A_k|B) = \frac{P(A_k \cap B)}{P(B)} = \frac{P(B|A_k) \cdot P(A_k)}{\sum_i P(B|A_i) \cdot P(A_i)}$$

Dokažite sami!

Primjer: Za neku rijetku bolest koja pogađa svakog tisućitog čovjeka razvijen je krvni test koji je pozitivan s vjerojatnošću 99% ako je čovjek uistinu bolestan, ali će i za zdrave ljude u 2% slučajeva test biti pozitivan. Ako nasumično odaberemo čovjeka s ulice i načinimo test pa test bude pozitivan, kolika je vjerojatnost da je čovjek uistinu bolestan?

Rješenje:

Definirajmo najprije događaje i pripadne vjerojatnosti:

A=’čovjek je bolestan’ $P(A)$=0,001

B=’test je pozitivan’ $P(B|A) = 0{,}99$ $P(B|\overline{A}) = 0{,}02$

Zanima nas $P(A|B)$.

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B|A) \cdot P(A) + P(B|\overline{A}) \cdot P(\overline{A})} = \frac{0{,}99 \cdot 0{,}001}{0{,}99 \cdot 0{,}001 + 0{,}02 \cdot 0{,}999} = 0{,}047$$

Nezavisnost

Def:

Događaji A i B su nezavisni onda i samo onda ako vrijedi

$$P(A|B) = P(A)$$

Lako se pokaže da onda vrijedi i $P(B|A) = P(B)$ **!**

Primjer:
Bacamo kocku i definiramo događaje:

A={1,2,3}
B={2,4,6}
C={1,2,3,4}

Koji od tih događaja su međusobno nezavisni?
Rješenje:

$P(A)$=1/2 ; $P(A|B)$=1/3 ; $P(A|C)$=3/4
$P(B)$=1/2 ; $P(B|C)$=1/2

Zaključak: Međusobno su nezavisni samo događaji B i C.

Propozicija:

Događaji A i B su nezavisni onda i samo onda ako vrijedi

$$P(A \cap B) = P(A) \cdot P(B)$$

OPREZ! Događaji koji se međusobno isključuju nisu nezavisni

IV. SLUČAJNA VARIJABLA

Pokus: bacam kocku s obojenim plohama. Svakoj boji plohe pridružujem neki broj:

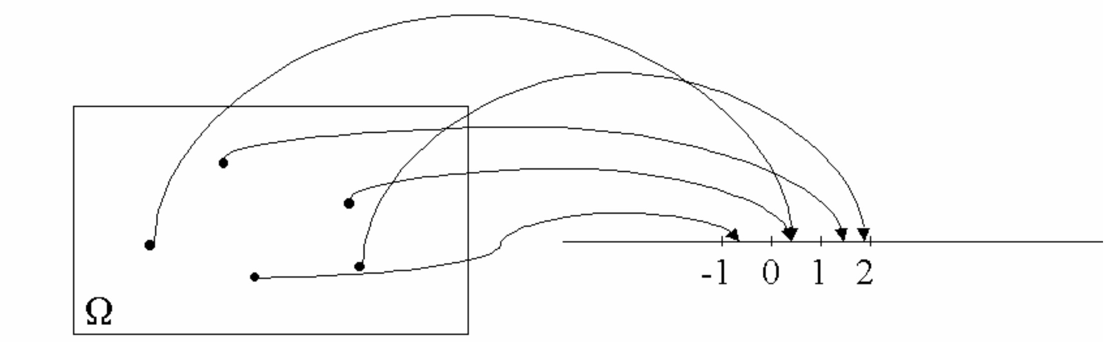
bijela ⟶ 1
žuta ⟶ 2
crvena ⟶ 3
narančasta ⟶ 4
zelena ⟶ 5
plava ⟶ 6

Def.:

Za dani prostor događaja Ω nekog pokusa, **slučajna varijabla** jest bilo koje pravilo kojim svakom ishodu pokusa pridružujemo neki broj.

$$X: \Omega \to \Re$$
$$X: s \in \Omega \mapsto x \in \Re$$

Označimo s D skup svih mogućih vrijednosti slučajne varijable X. Vrijedi $D \subseteq \Re$.



Primjeri:

1. Bacamo obojenu kocku. Svakoj boji pridružimo neki broj. Npr. žuta ⟶ 2
   ***Ω* = {bijela, žuta, crvena, narančasta, zelena, plava}; *D*={1,2,3,4,5,6}**
   X(bijela)=1; X(žuta)=2 itd.

2. Bacamo dvije kocke, crvenu i zelenu. Prostor elementarnih događaja je
   Ω={(1,1),(1,2),(1,3),...,(6,5),(6,6)}={(a,b): a,b∈{1,2,3,4,5,6}}

   Definiramo tri različite slučajne varijable

   a) X = zbroj točkica na kockama
   D={2,3,4,...12}; npr. X((1,3)) = 4

   b) Y = razlika broja točkica na crvenoj i zelenoj kocki
   D={-5,-4,-3,...5}; npr. Y((1,3)) = -2

   c) Z = umnožak broja točkica na crvenoj i zelenoj kocki
   D={1,2,3,4,5,6,8,9,10,12,15,16,18,20,24,25,30,36};
   npr. Z((5,3)) = 15

3. Studentima u predavaonici možemo npr. pridružiti slučajne varijable
   a) X = mjesec rođenja
   b) Y = visina zaokružena na centimetre

4. Bacanje novčića
   a) jednom $\Omega=\{P,G\}$; $D=\{0,1\}$
   $X(P)=1$, $X(G)=0$

   Def:
   Događaj koji ima samo dva moguća ishoda zove se **Bernoullijev događaj**, a slučajna varijabla koja ga opisuje zove se **Bernoullijeva slučajna varijabla**.

   b) dok ne padne pismo
   $\Omega = \{(P),(GP),(GGP),....\}$; $D = \mathbf{N} = \{1,2,3,...\}$
   slučajna varijabla: Y = broj bacanja dok ne padne pismo
   $Y((P))=1$ $Y((GP))=2$ $Y((GGP))=3$ ....

5. Nadmorska visina svih položaja u Hrvatskoj.
   Ω={skup svih položaja u Hrvatskoj}; $D=[0,1831]$

Peti primjer razlikuje se bitno od prva četiri. Slučajna varijabla definirana u tom primjeru je kontinuirana. Ona u konačnom intervalu može poprimiti beskonačno mnogo vrijednosti.
Slučajna varijabla u primjeru 4.b) može također poprimiti beskonačno mnogo vrijednosti, ali su te vrijednosti prebrojive, tj. mogu se poredati tako da se uvijek zna koja je prva, koja druga itd. U konačnom intervalu moguć je konačan broj vrijednosti.

Def.:

Diskretna slučajna varijabla je ona s.v. za koju je skup mogućih vrijednosti konačan ili prebrojiv.
Kontinuirana slučajna varijabla je ona s.v. za koju je skup mogućih vrijednosti cijeli interval na brojevnom pravcu

Za proučavanje diskretnih varijabli potreban nam je samo alat diskretne matematike: zbrajanje i oduzimanje. Za proučavanje kontinuiranih varijabli trebamo višu matematiku: integriranje i diferenciranje.

IV.1. DISKRETNA SLUČAJNA VARIJABLA I RASPODJELA VJEROJATNOSTI

Def:

Raspodjela vjerojatnosti diskretne slučajne varijable X definirana je za svaki broj x relacijom:

$$p(x) = P(X=x) = P(\text{svi ishodi } s \in \Omega : X(s) = x)$$

Ovdje se $P(X=x)$ čita "vjerojatnost da slučajna varijabla X poprimi vrijednost x"

Raspodjele vjerojatnosti možemo prikazati tablično ili funkcionalno.

Primjeri: Raspodjele vjerojatnosti za slučajne varijable iz primjera 1, 2a, 2b, 2c, 4a i 4b. To su *a priori* vjerojatnosti uz pretpostavku jednako vjerojatnih ishoda.

1.

$$p(x) = \begin{cases} 1/6 & , \quad x \in D = \{1,2,3,4,5,6\} \\ 0 & , \quad x \notin D \end{cases}$$

2.a)

$$p(x) = \begin{cases} \frac{1}{36}\left(6 - |x-7|\right) & , \quad x \in D = \{2,3,4,\ldots,12\} \\ 0 & , \quad x \notin D \end{cases}$$

2.b)

$$p(y) = \begin{cases} \frac{1}{36}\left(6 - |y|\right) & , \quad y \in D = \{-5,-4,\ldots,0,\ldots,5\} \\ 0 & , \quad y \notin D \end{cases}$$

2.c)

z	$p(z)$
1	1/36
2	2/36
3	2/36
4	3/36
5	2/36
6	4/36
8	2/36
9	1/36
10	2/36
12	4/36
15	2/36
16	1/36
18	2/36
20	2/36
24	2/36
25	1/36
30	2/36
36	1/36

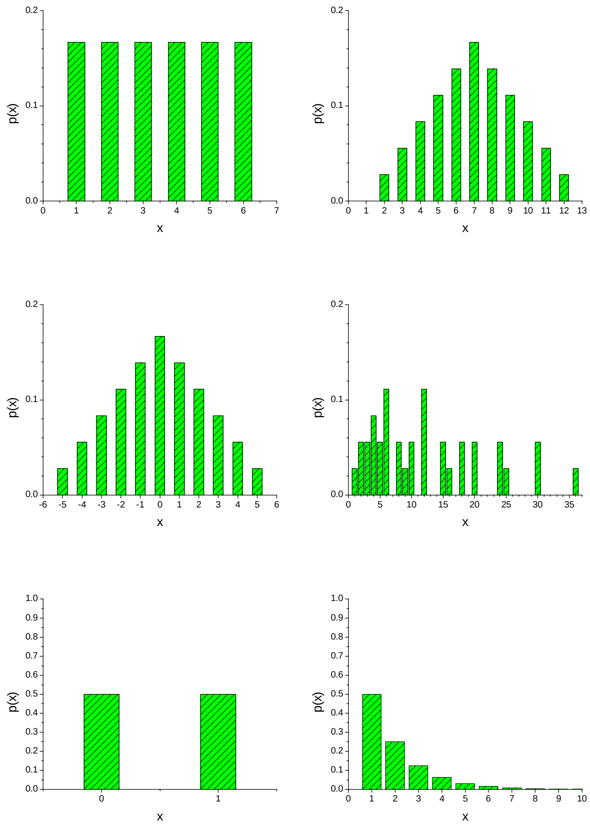
4.a)

$$p(x)=\begin{cases}1/2 & , \quad x\in D=\{0,1\}\\ 0 & , \quad x\notin D\end{cases}$$

4.b)

$$p(y)=\begin{cases}1/2^{y} & , \quad y\in \mathbf{N}=\{1,2,3,4,.....\}\\ 0 & , \quad y\notin \mathbf{N}\end{cases}$$

Raspodjele vjerojatnosti za slučajne varijable iz primjera 1, 2a, 2b, 2c, 4a i 4b prikazane su na sljedećoj slici.



Važno je napomenuti: $\boxed{\sum_{x\in D} p(x)=1}$

Funkcija raspodjele (kumulativna funkcija distribucije)

Def:

Za diskretnu slučajnu varijablu koja ima raspodjelu vjerojatnosti $p(x)$ definira se **funkcija raspodjele** $F(x)$ relacijom:

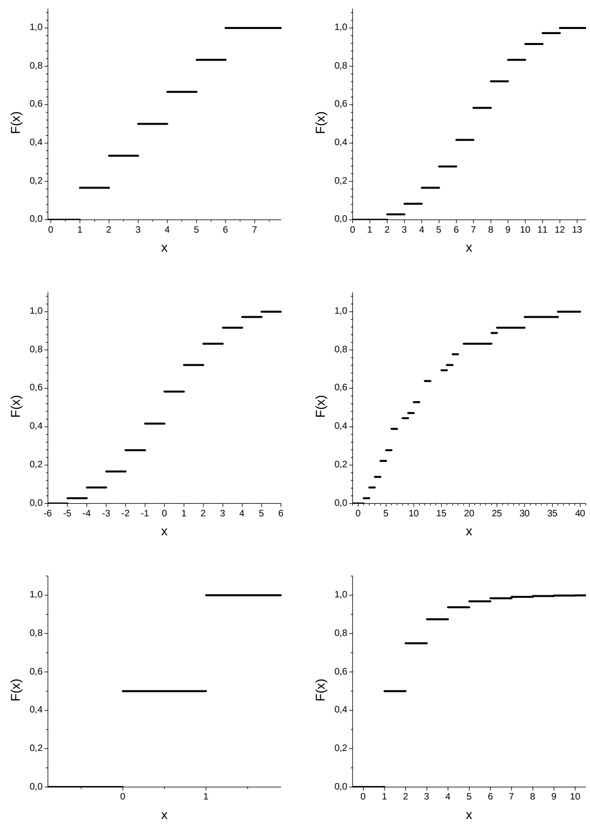
$$F(x) = P(X \leq x) = \sum_{y \leq x} p(y)$$

Za svaki x, $F(x)$ predstavlja vjerojatnost da X poprimi vrijednost ne veću od x.

Vrijedi:

$$\boxed{F(-\infty)=0 \quad ; \quad F(+\infty)=1}$$

Primjer:
Funkcije raspodjele za slučajne varijable iz primjera 1, 2a, 2b, 2c, 4a i 4b prikazane su na sljedećoj slici.



Za bilo koja dva broja a i b ($a \leq b$) vrijedi

$$P(a \leq X \leq b) = F(b) - F(a-),$$

gdje "a-" predstavlja najveću vrijednost varijable X koja je strogo manja od a.

Primjer 4.b)
Kolika je vjerojatnost da ćemo morati baciti najmanje dva, a najviše 5 puta?

$$P(2 \le Y \le 5) = F(5) - F(1)$$

$$F(1) = \sum_{y'=1}^{1} \frac{1}{2^{y'}} = \frac{1}{2},$$

$$F(5) = \sum_{y'=1}^{5} \frac{1}{2^{y'}} = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \frac{1}{32} = \frac{31}{32}$$

$$P(2 \le Y \le 5) = \frac{15}{32}$$

Karakterizacija raspodjela

Očekivana vrijednost diskretne slučajne varijable

Kada imamo neku raspodjelu, najčešći podatak koji o njoj želimo znati jest njezina srednja vrijednost.

Def.:

Neka je X diskretna slučajna varijabla sa skupom mogućih vrijednosti D i neka je $p(x)$ njezina funkcija vjerojatnosti. Tada je **srednja vrijednost** ili **očekivanje** varijable x dano s

$$E(X) = \mu_X = \sum_{x \in D} x \cdot p(x)$$

Očekivanja za raspodjele vjerojatnosti iz primjera 1, 2a, 2b, 2c, 4a i 4b:

1. $E(X) = \sum_{x \in D} x \cdot p(x) = \sum_{x=1}^{6} \frac{1}{6} x = \frac{21}{6} = 3.5$

2.a) $E(X) = 2 \cdot \frac{1}{36} + 3 \cdot \frac{2}{36} + ... + 7 \cdot \frac{6}{36} + ... + 12 \cdot \frac{1}{36} = 7$

2.b) $E(Y) = -5 \cdot \frac{1}{36} - 4 \cdot \frac{2}{36} + ... + 0 \cdot \frac{6}{36} + ... + 5 \cdot \frac{1}{36} = 0$

2.c) $E(Z) = 12.25$

4.a) $E(X) = \sum_{x=0}^{1} x \cdot \frac{1}{2} = \frac{1}{2}$

4.b) $E(Y) = \sum_{y=1}^{\infty} y \cdot \frac{1}{2^y} = 2$

Očekivanje funkcije slučajne varijable

Često nas ne zanima očekivanje same slučajne varijable nego neke njezine funkcije.

Primjer (nastavak primjera 4b):
U bacanju novčića dok ne padne pismo, zarada se izračunava tako da se kvadrira broj bacanja u kojima je pao grb prije no što je palo pismo. Definiramo novu slučajnu varijablu:

$$Z = \text{zarada u takvom bacanju} = Z(Y) = (Y-1)^2$$

Skup mogućih vrijednosti varijable Z je $D_Z=\{0,1,4,9,...\}$, a vjerojatnosti $p(z)$ su dane odgovarajućim vjerojatnostima za $p(y)$: $p(z=0) = p(y=1)$, $p(z=1) = p(y=2)$, $p(z=4) = p(y=3)$ itd.
Očekivana zarada je

$$E(Z) = \sum_{z \in D_Z} z \cdot p(z) = \sum_{z \in D_Z} z \cdot p(y) = \sum_{y \in D} (y-1)^2 \cdot p(y) = 3\,kn$$

Općenito:

Ako je $g(X)$ neka funkcija slučajne varijable X onda je njezino očekivanje

$$E(g(X)) = \sum_{x \in D} g(x)p(x) \quad .$$

Uočimo da očekivanje funkcije <u>općenito nije funkcija očekivanja</u>.
U našem primjeru je bilo $E(Z)=3$, a $(E(Y)-1)^2=1$.

Poseban slučaj je <u>linearna funkcija</u>:

$$g(X) = aX + b$$

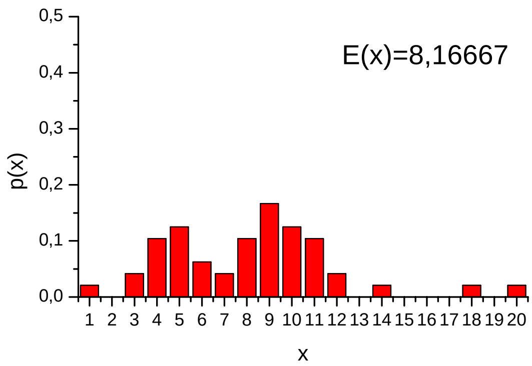
$$E(aX+b) = \sum_{x \in D} (ax+b)p(x) = a\sum_{x \in D} xp(x) + b\sum_{x \in D} p(x) = aE(X) + b \quad .$$

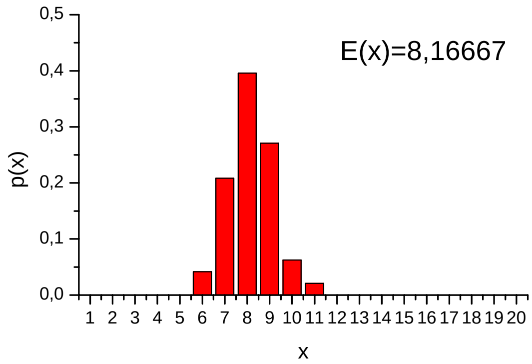
$$E(aX+b) = aE(X) + b$$
$$E(aX) = aE(X)$$
$$E(X+b) = E(X) + b$$

Samo za linearnu funkciju vrijedi da je očekivanje funkcije jednako funkciji očekivanja!

Varijanca diskretne slučajne varijable

U primjeru djece iz parka (prvo predavanje), prosječna starost djece je 8,17 godina. Međutim, iz tog broja ne možemo zaključiti je li starost djece grupirana oko tog broja (jer su, npr., igračke u parku takve da su primjerene osmogodišnjacima) ili se u parku igraju djeca svih uzrasta (jer su igračke primjerene širokom rasponu dječjih uzrasta). Zanima nas raspršenost dječjih uzrasta u parku. Raspodjela starosti djece prikazana je u gornjem dijelu sljedeće slike. U donjem dijelu iste slike prikazana je jedna druga raspodjela koja ima isto očekivanje.





Vidimo da nam očekivanje raspodjele nije dovoljan podatak. Zanima nas i koliko su vjerojatnosti raspršene.

Srednja vrijednost ima posebnu važnost u razmatranju neke raspodjele. Rasipanje raspodjele oko srednje vrijednosti važno je za njezinu karakterizaciju. Međutim, kako je srednje odstupanje od srednje vrijednosti uvijek jednako nuli, moramo pronaći neku drugu veličinu koja će nam karakterizirati rasipanje.

Zato proučavamo srednje kvadratično odstupanje, tj. aritmetičku sredinu kvadrata odstupanja:

Def.:

Neka je X diskretna slučajna varijabla i neka je $p(x)$ njezina raspodjela vjerojatnosti. Tada je **varijanca** od X dana s očekivanjem kvadrata odstupanja od μ.

$$V(X) = \sigma_X^2 = \sum_{x \in D} (x - \mu_X)^2 \, p(x) = E[(X - \mu_X)^2]$$

Veličinu $\sigma_X = \sqrt{V(X)}$ nazivamo **standardnim odstupanjem** ili **standardnom devijacijom** slučajne varijable X.

Izračunamo li standardne devijacije za raspodjele s prethodne slike, dobivamo σ=3,64 za gornju raspodjelu i σ=1,03 za donju raspodjelu.

Važno svojstvo varijance koje često olakšava račune je:

$$\boxed{V(X) = \sigma^2 = \sum_{x \in D} x^2 p(x) - \mu_X^2 = E(X^2) - [E(X)]^2}$$

Varijanca funkcije slučajne varijable

$$\boxed{V(g(X)) = \sum_{x \in D} \left(g(x) - E(g(X))\right)^2 p(x) = E\left(g^2(X)\right) - \left(E\left(g(X)\right)\right)^2}$$

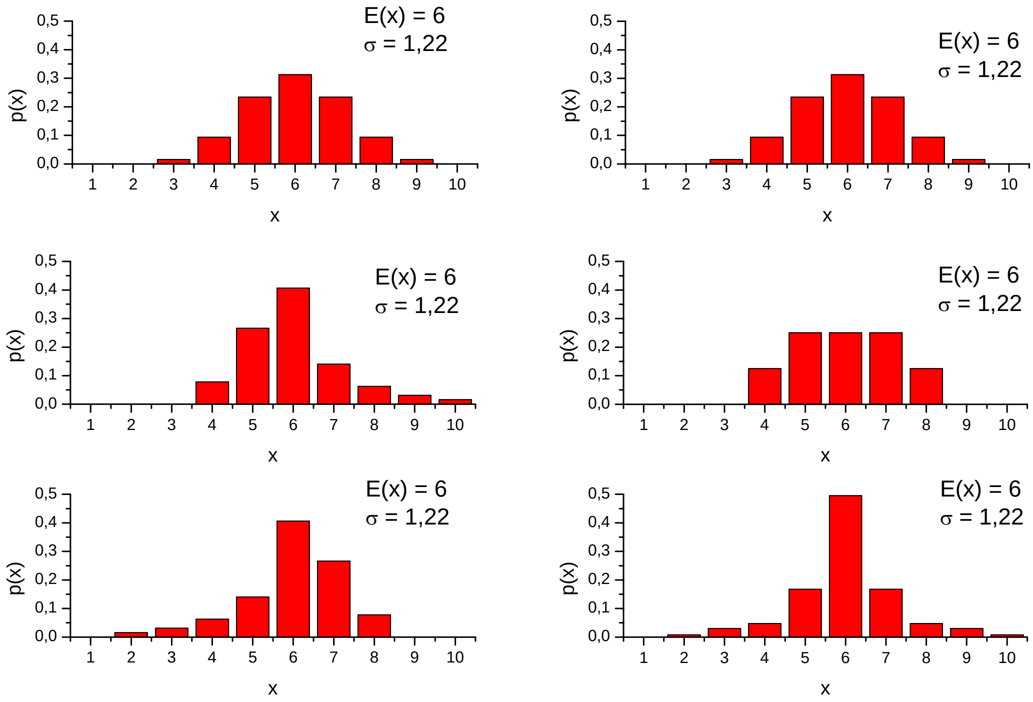
Poseban slučaj: linearna funkcija

$g(X) = aX + b$

$\boxed{V(aX + b) = a^2 V(X)}$ (dokažite sami)

Momenti višeg reda

Možemo uvesti dodatne veličine koje nam još detaljnjije opisuju neku raspodjelu. Na sljedećoj slici prikazano je nekoliko raspodjela koje sve imaju istu srednju vrijednost i standardnu devijaciju, ali su ipak različite. U prvom stupcu su raspodjele



različite simetrije. Najgornja raspodjela je simetrična, ispod nje je raspodjela 'nagnuta udesno', a na dnu je raspodjela 'nagnuta ulijevo'. U drugom stupcu su tri raspodjele različite spljoštenosti. Da bismo te razlike kvantitativno opisali uvodimo momente višeg reda.

Def:

Središnji (centralni) moment r-tog reda jest

$$M_r = \sum_{x \in D} (x-\mu)^r \, p(x)$$

Def:

Pomoćni (ishodišni) moment r-tog reda jest

$$m_r = \sum_{x \in D} x^r \, p(x)$$

Prvih nekoliko momenata predstavlja veličine koje smo već upoznali:

- Moment nultog reda predstavlja totalnu vjerojatnost:
$$M_0 = m_0 = 1$$
- Središnji moment prvog reda predstavlja srednje odstupanje od aritmetičke sredine:
$$M_1 = 0$$

- Pomoćni moment prvog reda predstavlja očekivanje ili aritmetičku sredinu:

$$m_1 = E(X) = \bar{x} = \mu$$

- Središnji moment drugog reda predstavlja varijancu:

$$M_2 = V(X) = \sigma^2$$

- Za izračun središnjeg momenta drugog reda možemo upotrijebiti relaciju:

$$M_2 = m_2 - m_1^2$$

Općenita formula za središnji moment *r*-tog reda je:

$$\boxed{M_r = \sum_k (-1)^k \binom{r}{k} m_{r-k} m_1^k}$$

(izvod na ploči !!!)

Moment trećeg reda

.....govori o asimetriji. Primjeri iz lijevog stupca prethodne slike imaju momente trećeg reda: $M_3 = 0$

$M_3 = 1{,}594 > 0$

$M_3 = -1{,}594 < 0$

Definiramo veličinu **koeficijent asimetrije** kao

$$\alpha_3 = \frac{M_3}{\sigma^3}$$

Ako je $\alpha_3 = 0$ kažemo da je raspodjela simetrična, ako je $\alpha_3 > 0$ kažemo da je raspodjela nagnuta udesno, a ako je $\alpha_3 < 0$ kažemo da je raspodjela nagnuta ulijevo.

Gornji primjeri imaju

$\alpha_3 = 0$

$\alpha_3 = 0{,}868 > 0$

$\alpha_3 = -\ 0{,}868 < 0$

Moment četvrtog reda

.....govori o spljoštenosti. Primjeri iz desnog stupca prethodne slike imaju momente četvrtog reda: $M_4 = 6$

$M_4 = 4{,}5$

$M_4 = 10{,}67$

Definiramo veličinu **koeficijent spljoštenosti** kao

$$\alpha_4 = \frac{M_4}{\sigma^4}$$

Ako je $\alpha_4 = 3$ kažemo da je raspodjela normalno spljoštena, ako je $\alpha_4 > 3$ kažemo da je raspodjela šiljata, a ako je $\alpha_4 < 3$ kažemo da je raspodjela široka.

Gornji primjeri imaju

$\alpha_4 = 2{,}67$; $\alpha_4 = 2$; $\alpha_4 = 4{,}74$

Funkcija izvodnica

Mnoge teorijske raspodjele mogu se *jednostavno* definirati pomoću **funkcije izvodnice** ili **generatrise.** Za bilo koju diskretnu raspodjelu slučajne varijable definiramo:

$$\boxed{g(t)=\sum_{x\in D} e^{t\cdot x}\,p(x)}$$

Vidimo da je generatrisa kontinuirana funkcija varijable *t*, razvijena u diskretan red. Budući da je kontinuirana (i analitička), možemo ju derivirati:

$$\frac{dg}{dt}=g'(t)=\sum_{x\in D} e^{t\cdot x}x\cdot p(x)$$

Vrijednost prve derivacije u $t = 0$ predstavlja pomoćni moment prvog reda:

$$g'(0)=\sum_{x\in D} x\cdot p(x)=\mu=m_1\ .$$

Vrijednost druge derivacije u $t = 0$ predstavlja pomoćni moment drugog reda:

$$g''(0)=m_2\ .$$

Općenito, vrijednost *r*-te derivacije u $t = 0$ predstavlja pomoćni moment *r*-tog reda:

$$g^{(r)}(0)=m_r\ .$$

Vidimo da nam poznavanje funkcije izvodnice za neku raspodjelu znatno olakšava pronalaženje pomoćnih, a time i središnjih momenata.

IV.1.1. BINOMNA RASPODJELA

ponovimo:

<u>Bernoullijev pokus</u>

... je pokus koji ima samo dva moguća ishoda:

"uspjeh"= A ili "neuspjeh"= $\overline{A}$

Vjerojatnost uspjeha označavamo s p, a vjerojatnost neuspjeha s q:

$p = P(A)$; $q = P(\overline{A})$

Vrijedi $\boxed{p+q=1}$.

<u>Binomni pokus</u>

Def.:

Binomni pokus (eksperiment) je pokus koji zadovoljava uvjete:

1. Sastoji se od n pokušaja, a n je određen prije početka pokusa.
2. Pokušaji su međusobno identični Bernoullijevi pokusi s mogućim ishodima '*uspjeh*' (A) i '*neuspjeh*' ($\overline{A}$).
3. Pokušaji su nezavisni. Ishod bilo kojeg pokušaja ne utječe na ishod drugog.
4. Vjerojatnost '*uspjeha*' jednaka je za sve pokušaje i označavamo ju s p .

Primjer 1:

Bacamo novčić četiri puta (ili bacamo četiri jednaka novčića). Uspjeh je kad padne 'grb':

$n = 4$; $p = 1/2$

Primjer 2:

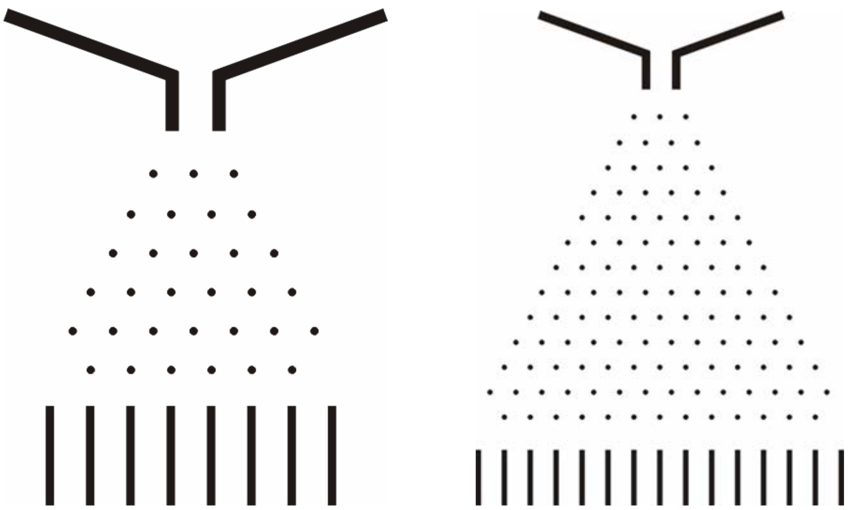
Bacamo kocku 10 puta. Uspjehom smatramo kad padne šestica:

$n = 10$; $p = 1/6$

Primjer 3:

Galtonova daska. Kuglicu spuštamo na čavliće koji su složeni u pravilnu trokutastu rešetku. Padom na čavlić, kuglica može skrenuti ulijevo (*neuspjeh*) ili udesno (*uspjeh*). Ako je daska pravilna i vodoravna, ti su ishodi jednako vjerojatni.

n = broj redova čavlića ; $p = 1/2$



Binomna slučajna varijabla

Def:

Ako imamo binomni eksperiment koji se sastoji od n pokušaja, slučajnu varijablu X definiranu kao

$$X = \text{broj uspjeha u } n \text{ pokušaja}$$

nazivamo **binomna slučajna varijabla** i označavamo je

$$X \sim \mathrm{Bin}(n,p)$$

Binomna raspodjela vjerojatnosti

Ako imamo binomni eksperiment i binomnu slučajnu varijablu $X \sim \mathrm{Bin}(n,p)$, njezinu raspodjelu vjerojatnosti $p(x)$ nazivamo **binomna raspodjela** i označavamo ju s

$$b(x;n,p)$$

Odredimo vrijednosti te raspodjele! Razmotrimo Galtonovu dasku s n=4 i pogledajmo sve načine na koje kuglica može padati.
(Animacija 'Wege einer Kugel' (put jedne kugle) s Internet adrese:
http://www.uni-koeln.de/ew-fak/Mathe/Projekte/VisuPro/galton/galton.html)

Tablica: Svi mogući ishodi binomnog pokusa s n=4:

ishod	slučajna varijabla x	vjerojatnost
$\bar{A}\bar{A}\bar{A}\bar{A}$	0	q^4
$\bar{A}\bar{A}\bar{A}A$	1	$p \cdot q^3$
$\bar{A}\bar{A}A\bar{A}$	1	
$\bar{A}A\bar{A}\bar{A}$	1	
$A\bar{A}\bar{A}\bar{A}$	1	
$\bar{A}\bar{A}AA$	2	$p^2 \cdot q^2$
$\bar{A}A\bar{A}A$	2	
$\bar{A}AA\bar{A}$	2	
$A\bar{A}\bar{A}A$	2	
$A\bar{A}A\bar{A}$	2	
$AA\bar{A}\bar{A}$	2	
$\bar{A}AAA$	3	$p^3 \cdot q$
$A\bar{A}AA$	3	
$AA\bar{A}A$	3	
$AAA\bar{A}$	3	
$AAAA$	4	p^4

Vidimo da slučajna varijabla X poprima vrijednost 0 za jedan ishod, tj. $\binom{4}{0}$;

vrijednost 1 poprima za 4 ishoda, tj. $\binom{4}{1}$; vrijednost 2 poprima za 6 ishoda, tj. $\binom{4}{2}$;

vrijednost 3 poprima za 4 ishoda, tj. $\binom{4}{3}$; a vrijednost 4 poprima za 1 ishod, tj. $\binom{4}{4}$.

Općenito možemo reći da će raspodjela vjerojatnosti u binomnom pokusu biti dana s:

$$b(x;n,p) = \begin{Bmatrix} \text{broj} & \text{ishoda} \\ \text{koji} & \text{imaju} \\ x & \text{uspjeha} \end{Bmatrix} \cdot \begin{Bmatrix} \text{vjerojatnost} \\ \text{svakog} \quad \text{takvog} \\ \text{ishoda} \end{Bmatrix} = K_n^x p^x q^{n-x}$$

Dakle,

Binomna raspodjela vjerojatnosti je

$$b(x;n,p) = \begin{cases} \binom{n}{x} p^x q^{n-x} & \text{za} \quad x = 0,1,2,\ldots,n \\ 0 & \text{in ace} \end{cases}$$

Primjer 1:
(Animacija 'Römischer Brunnen' (Rimska vrela) s Internet adrese:
http://www.uni-koeln.de/ew-fak/Mathe/Projekte/VisuPro/galton/galton.html)

Primjer 2:
(Animacija 'Galton Brett' (Galtonova daska) za n=6; N=50; p=0,5 s Internet adrese:
http://www.uni-koeln.de/ew-fak/Mathe/Projekte/VisuPro/galton/galton.html)

Primjer 3:
(Animacija 'Galton Brett' (Galtonova daska) za n=6; N=50; p=0,3 s Internet adrese:
http://www.uni-koeln.de/ew-fak/Mathe/Projekte/VisuPro/galton/galton.html)

Funkcija raspodjele

Funkciju raspodjele (kumulativnu) označavamo velikim slovom B:

$$F(x)=B(x;n,p)=\sum_{y=0}^{x} b(y;n,p)$$

Rekurzivna formula

U izračunavanju vjerojatnosti za binomnu raspodjelu često je vrlo korisna rekurzivna formula:

$$\frac{b(x;n,p)}{b(x-1;n,p)} = \frac{\binom{n}{x} p^x q^{n-x}}{\binom{n}{x-1} p^{x-1} q^{n-x+1}} = \frac{n-x+1}{x} \frac{p}{q}$$

$$\boxed{b(x;n,p)=\frac{n-x+1}{x}\cdot\frac{p}{q}\cdot b(x-1;n,p)}$$

Primjer:
Vjerojatnost da prilikom bacanja sedam kockica nijedna ne pokaže 6 (n=7; p=1/6):

$$b(0;7,1/6)=(5/6)^7=0,279$$

Vjerojatnost da padne jedna šestica je:

$$b(1;7,1/6)=7(1/5)\ b(0;7,1/6)=0,391$$

Vjerojatnost da ima dvije sedmice je:

$$b(2;7,1/6)=(6/2)(1/5)\ b(1;7,1/6)=0,234$$

itd.

Pomoću rekurzivne formule možemo pronaći vrijednost x_M slučajne varijable X za koju je vjerojatnost najveća. Mora vrijediti:

$$b(x_M-1;n,p) \le b(x_M;n,p) \ge b(x_M+1;n,p)$$

Lakim računom pokazujemo da to vrijedi kada je zadovoljen uvjet:

$$np-q \le x_M \le np+p$$

Za poseban slučaj $p=q=1/2$ vrijedi:

$$(n-1)/2 \le x_M \le (n+1)/2$$

Ako je n paran, imamo jedno rješenje, a ako je neparan, imamo dva rješenja.

Očekivanje i varijanca

Izračunajmo očekivanje i varijancu za binomnu raspodjelu. Prema definiciji je

$$E(X)=\sum_{x=0}^{n} x\cdot b(x;n,p)=\sum_{x=0}^{n}\binom{n}{x}p^x q^{n-x}x=\sum_{x=1}^{n}\binom{n}{x}p^x q^{n-x}x\ .$$

Možemo iskoristiti relaciju $\binom{n}{x}=\frac{n}{x}\binom{n-1}{x-1}$ i pisati

$$E(X)=p\sum_{x=1}^{n}\frac{n}{x}\binom{n-1}{x-1}p^{x-1}q^{n-x}x\ .$$

Sada možemo pokratiti x i uvesti supstitucije $y=x-1$ i $m=n-1$. Tada je

$$E(X)=np\sum_{y=0}^{m}\binom{m}{y}p^y q^{m-y}\ .$$

Sumacija predstavlja upravo totalnu vjerojatnost za binomnu raspodjelu Y~Bin(m,p) i iznosi 1. Stoga za očekivanje slučajne varijable X~Bin(n,p) vrijedi

$$\boxed{E(X) = np} \,.$$

Za izračunavanje varijance iskoristit ćemo relaciju

$$V(X) = E(X^2) - [E(X)]^2 = E(X^2) - n^2 p^2 \,.$$

Prvi član je

$$E(X^2) = \sum_{x=0}^{n} x^2 \cdot b(x;n,p) = \sum_{x=1}^{n} \binom{n}{x} p^x q^{n-x} x^2 = np \sum_{x=1}^{n} \binom{n-1}{x-1} p^{x-1} q^{n-x} x \,.$$

Sada možemo uvesti supstitucije $y=x-1$ i $m=n-1$. Tada je

$$E(X^2) = np \sum_{y=0}^{n} \binom{m}{y} p^y q^{m-y} (y+1) = np \left[\sum_{y=0}^{n} y \binom{m}{y} p^y q^{m-y} + 1 \right] .$$

Sumacija u uglatoj zagradi je očekivanje varijable Y~Bin(m,p) i iznosi $E(Y)=mp$ pa je

$$E(X^2) = np[mp+1] = np[(n-1)p+1] = n^2 p^2 - np^2 + np \,.$$

Varijanca je tada

$$V(X) = np - np^2 = np(1-p) = npq$$

ili

$$\boxed{V(X) = npq}$$

<u>Funkcija izvodnica</u>

Tvrdim: Za binomnu raspodjelu funkcija izvodnica je

$$\boxed{g(t) = (q + pe^t)^n}$$

Dokaz:
Prema definiciji je

$$g(t) = \sum_{x=0}^{n} e^{t \cdot x} p(x)$$

Stoga sumaciju možemo pisati:

$$g(t) = \sum_{x=0}^{n} e^{tx} b(x;n,p) = \sum_{x=0}^{n} \binom{n}{x} e^{tx} p^x q^{n-x} = \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x q^{n-x} = (pe^t + q)^n$$

Kada znademo funkciju izvodnicu, lako izračunavamo pomoćne momente:

$$m_0 = g(0) = 1$$

$$m_1 = g'(0) = n(q + pe^0)pe^0 = np$$

$$m_2 = g''(0) = n^2 p^2 - np^2 + np$$

Varijanca je tada

$$M_2 = m_2 - m_1^2 = npq$$

Prilagođavanje empirijskih podataka binomnoj raspodjeli

Primjer 1:
(Animacija 'Galton Brett' (Galtonova daska) za n=5; N=300; p=0,5 s Internet adrese: http://www.uni-koeln.de/ew-fak/Mathe/Projekte/VisuPro/galton/galton.html)

Za vrijeme trajanja animacije izračunamo teorijske frekvencije koje očekujemo za slučajnu varijablu X:

X	$B(x;5,1/2)$	$f_{teor}=N\cdot b(x;5,1/2)$	f_{exp}
0	$(0,5)^5$=0,03125	9	
1	0,15625	47	
2	0,3125	94	
3	0,3125	94	
4	0,15625	47	
5	0,03125	9	
Σ	1	300	300

Kad završi animacija, izračunamo $\bar{x} = \frac{1}{N}\sum_{x=0}^{5} x \cdot f_{\exp}(x)$. Odredimo *a posteriori* novi p:

$$p = \frac{\bar{x}}{n}$$

Primjer 2:
U nekom tvorničkom skladištu imamo velik broj gotovih proizvoda (npr. 100.000). Želimo procijeniti postotak škarta pa uzimamo slučajni uzorak od n = 20 komada i bilježimo broj neispravnih proizvoda. Pokus ponavljamo N = 100 puta. Zanima nas kolika je vjerojatnost škarta (neispravnog proizvoda). Opažene su frekvencije:

x	0	1	2	3	4	5	6	≥7	Ukupno
$f(x)$	14	25	27	23	7	3	1	0	100

Najprije izračunavamo srednju vrijednost:

x	$f(x)$	$x\cdot f(x)$
0	14	0
1	25	25
2	27	54
3	23	69
4	7	28
5	3	15
6	1	6
Σ	100	197

Srednja vrijednost nam određuje *a posteriori* vjerojatnost škarta:

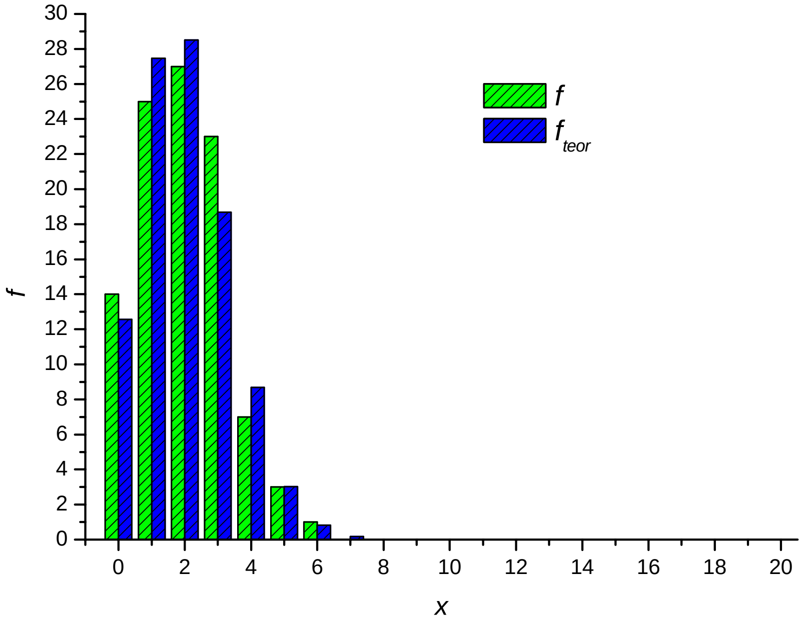
$$\bar{x} = \frac{1}{N}\sum_{x} x \cdot f = 1{,}97\,;\quad p = \frac{\bar{x}}{n} = 0{,}0985$$

Izračunavamo teorijske frekvencije prema relaciji:

$$f_{teor}(x) = N \cdot b(x;n,p) = 100\binom{20}{x} p^{x} q^{20-x}$$

i zaokružujemo na cjelobrojne vrijednosti:

x	$f(x)$	$f_{teor}(x)$	$f_{teor}(x)$
0	14	12,57	13
1	25	27,47	27
2	27	28,51	28
3	23	18,69	19
4	7	8,68	9
5	3	3,03	3
6	1	0,83	1
Σ	100	<100	100



primjer 3:
Zakon velikih brojeva.
(Animacija u programu s Internet adrese:
ftp://didaktik.physik.uni-wuerzburg.de/pub/physik/krahmer/galton.zip)

Pokazujemo da se usporedbom teorijske i opažene frekvencije dobivaju to bolja slaganja što je broj pokusa veći.