VJEROJATNOST I STATISTIKA U SREDNJOJ ŠKOLI

doc. dr. sc. Matija Bašić

10. Kongres nastavnika matematike, 2. srpnja 2024.

CILJEVI OBRAZOVANJA I KURIKULUM PITANJA

- CILJEVI OBRAZOVANJA:
  - PROFESIONALNO CERTIFICIRANJE VS. PRIPREMA ZA ŽIVOT

- PARADIGME POUČAVANJA:
  - POSJEĆIVANJE SPOMENIKA VS. PROPITIVANJE SVIJETA

- POTREBA ZA POSTAVLJANJEM RELEVANTNIH PITANJA I PROBLEMA, INFORMIRANJE O SUVREMENIM ALATIMA

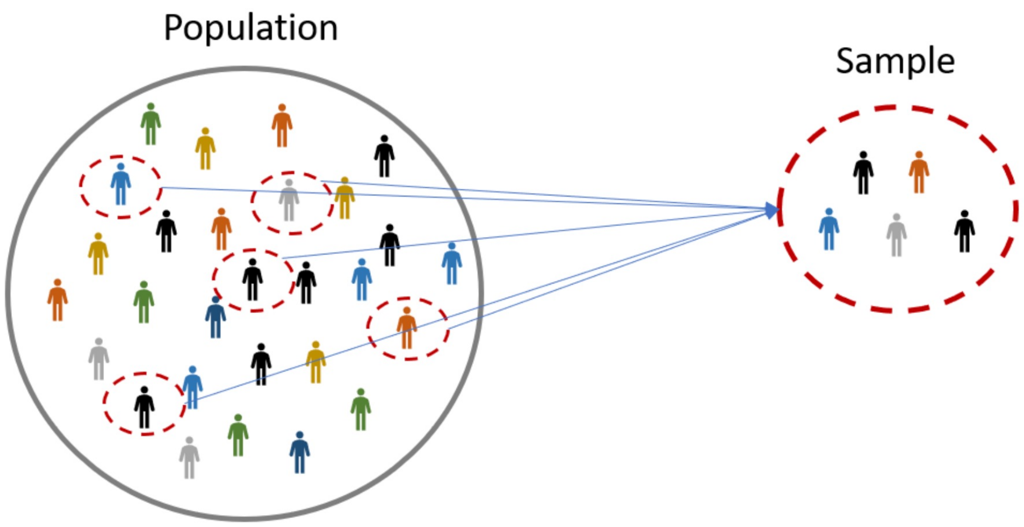
- TEK SMO ZAPOČELI PRILAGODBU ZA POUČAVANJE

STATISTIČKO MIŠLJENJE

Promatramo populaciju za koju nas zanimaju određena obilježja/svojstva.

U situaciji u kojoj ne možemo testirati cijelu populaciju obilježja možemo procijeniti na temelju uzorka (podskupa populacije).

Reprezentativni uzorak je uzorak koji ima ista obilježja (parametre) kao i čitava populacija.



Slučajni uzorak je niz nezavisnih jednako distribuiranih (n.j.d.) slučajnih varijabli. Interpretiramo da svaka slučajna varijabla predstavlja jednu jedinku iz populacije i da ima istu distribuciju kao populacija.

STATISTIKA

1) DESKRIPTIVNA

- različiti načini prikazivanja podataka: tablice frekvencija, histogram, kružni dijagram…
- mjere srednje vrijednosti: prosječna (aritmetička sredina), središnja (medijan) i dominantna vrijednost (mod)
- mjere raspršenosti
- brkata kutija

2) INFERENCIJALNA

- normalna i druge razdiobe
- standardiziranje i z-vrijednosti
- procjena populacijskih parametara
- testiranje hipoteza
- …

Nema nikad 100%–nih zaključaka!
Vrlo pažljivo treba interpretirati!

SIMPSONOV PARADOKS

Petar je oba dana postigao bolji rezultat, ali Šimun je ukupno ostvario bolji rezultat. Tko ima bolji rezultat?

	Petar	**Šimun**
Prvi dan	**2/2=100%**	6/7=86%
Drugi dan	**4/8=50%**	1/3=33%
Ukupno	6/10=60%	**7/10=70%**

TRI VRSTE PRISTRANOSTI (eng. bias)

1) ORIJENTACIJA NA ISHOD

- zanemarivanje nasumičnosti i vjerojatnosti
- meteorolog: "70% vjerojatnost da će padati kiša"
- zaključak: svaki dan pada kiša ili 7 od 10 dana pada kiša?

2) PRISTRANOST JEDNAKOJ VJEROJATNOSTI

- "kiša će padati ili neće padati, vjerojatnost je 50%"
- je li vjerojatnije da će zbroj brojeva na dvije kockice biti 6 ili 7?

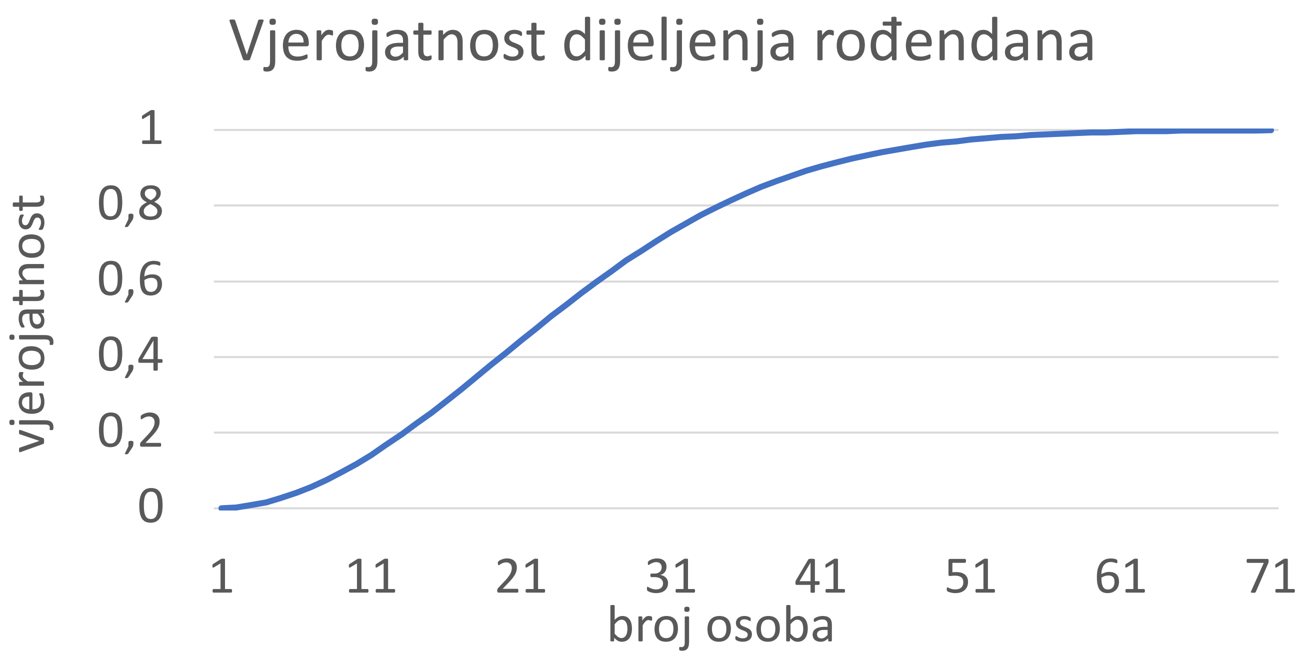
3) REPREZENTATIVNOST

- pravilno je manje vjerojatno: PPPGGG ima manju vjerojatnost nego PGGPPG

KAD IMATE ROĐENDAN?

Kolika je vjerojatnost da su dvije osobe iz ove grupe rođene istog datuma?

Broj učenika	Vjerojatnost
10	11,7 %
20	41,1 %
23	50,7 %
30	70,6 %
75	99,97 %



$$P = 1 - \frac{365!}{(365-n)!\,365^n}.$$

PRIMJER

Statistički je utvrđeno da u populaciji 1% ljudi ima rijetku bolest.

Test za tu bolest za 2% bolesnih daje negativan odgovor (osoba nema bolest), a u 5% zdravih daje pozitivan odgovor (da osoba ima bolest).

Ako test daje pozitivan odgovor, koja je vjerojatnost da ste bolesni?

	Bolesni	Zdravi
Pozitivan test	98%	5%
Negativan test	2%	95%

Uočimo važnost „lažno pozitivnih” i ”lažno negativnih” osoba, te potencijal za interdisciplinarnost teme!

KLASIČNA EKSPERIMENTALNA (A POSTERIORI) DEFINICIJA VJEROJATNOSTI

Provodimo slučajan pokus (eksperiment) čiji ishod varira:

- ishod nazivamo još i elementarni događaj
- događaj = skup nekih ishoda (*elementarnih događaja*)
- relativna frekvencija događaja = broj pojavljivanja događaja od ukupnog broja ponavljanja

Osnovne pretpostavke:

Problem: ishodi i njihove vjerojatnosti?

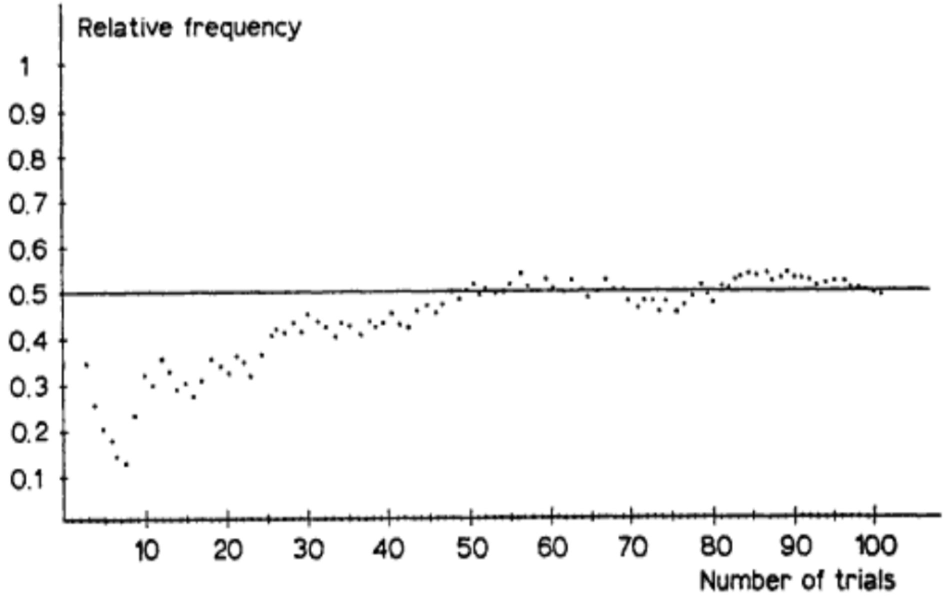
- svi ishodi su jednako mogući
  - dolazi od pretpostavljene simetrije objekta kojim izvodimo eksperiment (novčić, kocka itd.)
- relativne frekvencije se grupiraju oko realnog broja (između 0 i 1)

Neformalno!

PONAVLJANJE EKSPERIMENTA

Koliko su nam veliki uzorci?
Imamo li istu relativnu frekvenciju za dani broj ponavljanja?
Zašto povećanjem dobivamo brojeve koji su sve bliži 50%?

Broj bacanja	Relativna frekvencija pojavljivanja pisma (neki primjeri izmjerenih uzoraka)
5	60%
5	80%
10	40 %
10	70 %
100	50 %
100	54 %
100	49 %
1000	51 %



Hans Freudenthal, The ‘empirical law of large numbers’ or ‘the stability of frequencies’, 1972.

Potrebno posvetiti više vremena statističkom značaju Zakona velikih brojeva.

KLASIČNA TEORIJSKA (A PRIORI) DEFINICIJA

Što znači jednako vjerojatno? Cirkularna definicija!

Pretpostavimo li da imamo ***konačno mnogo elementarnih događaja i svi su jednako vjerojatni*** govorimo o Laplaceovom modelu vjerojatnosti.

Vjerojatnost događaja $A \subseteq \Omega$ je $P(A) = \frac{k(A)}{k(\Omega)}$.

Ova definicija se može koristiti samo uz gornje pretpostavke!

Kažemo da je vjerojatnost omjer broja elementarnih događaja *povoljnih za A* i broja svih elementarnih događaja.

UVJETNA VJEROJATNOST – DEFINICIJA

Uvjetna vjerojatnost događaja A uz uvjet da se dogodio događaj B ***pozitivne vjerojatnosti*** definiramo formulom

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

Ne zaboravite slučaj P(B)=0.

Dva puta bacamo kocku. Neka je A={prvi put je dobiven broj 2}

i B={zbroj je najviše 5}. Tada je P(A|B)=3/10 i P(B|A)=1/2.

	Prvi broj 2	**Prvi broj nije 2**
Zbroj najviše 5	(2,1), (2,2), (2,3)	(1,1), (1,2), (1,3), (1,4), (3,1), (3,2), (4,1)
Zbroj veći od 5	(2,4), (2,5), (2,6)	Još 23 para

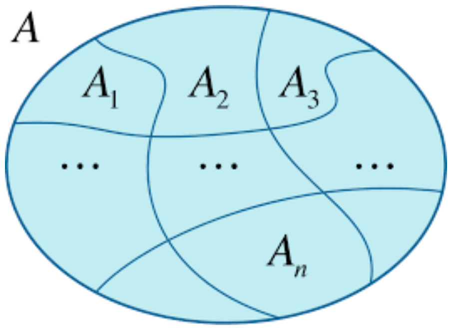
UVJETNA VJEROJATNOST – BAYESOVA FORMULA

Neka je $\Omega = H_1 \cup H_2 \cup \cdots \cup H_k$ particija prostora elementarnih događaja.

Podskupove H_i nazivamo hipotezama.

Za svaki događaj $A \subseteq \Omega$ vrijedi formula potpune vjerojatnosti

$$P(A) = \sum_i P(A \cap H_i) = \sum_i P(A|H_i) \cdot P(H_i)$$



Bayesova formula: $P(H_i|A) = \frac{P(A|H_i)P(H_i)}{P(A)}$.

UVJETNA VJEROJATNOST - PRIMJER

Statistički je utvrđeno da u populaciji 1% ljudi ima rijetku bolest.

Test za tu bolest za 2% bolesnih daje negativan odgovor (osoba nema bolest), a u 5% zdravih daje pozitivan odgovor (da osoba ima bolest).

Ako test daje pozitivan odgovor, koja je vjerojatnost da ste bolesni?

	Bolesni	Zdravi
Pozitivan test	98%	5%
Negativan test	2%	95%

Na 20 000 ljudi	Bolesni (200)	Zdravi (19 800)
Pozitivan test	196	990
Negativan test	4	18810

Je li test pouzdan?

$$P(bol|poz) = \frac{P(poz|bol)P(bol)}{P(poz)} = \frac{196}{196 + 990} \approx 16\%$$

https://www.mathsisfun.com/data/probability-false-negatives-positives.html

FERMATOVO PISMO PASCALU

U igri se baca novčić. Fermat pobjeđuje ako se u nizu pojavi 10 puta glava, a Pascal pobjeđuje ako se u nizu pojavi 10 puta pismo. Pobjednik treba dobiti 100 franaka. Igra se morala završiti u trenutku kad se pojavilo 8 puta pismo i 7 puta glava. U kojem omjeru je pravedno podijeliti nagradu?

PP	PGP GPP	PGGP GPGP GGPP	GGG	PGGG GPGG GGPG	Sustavno ispisivanje i prebrojavanje
1/4	2/8	3/16	1/8	3/16	
	11/16			5/16	

NEZAVISNOST

U prethodnim primjerima smo ponavljali pokus (bacanje novčića ili kockice) i računali vjerojatnost događaja u kojem su se određeni ishodi dogodili u nizu takvih pokusa. Pritom smo koristili da ishod u jednom ponavljanju ni na koji način ne utječe na ishod u sljedećem ponavljanju – **nezavisnost**.

Definicija. Kažemo da su dva događaja A i B nezavisni ako vrijedi

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

Primjer. Tako smo računali da je $P(pp) = P(p) \cdot P(p) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$.

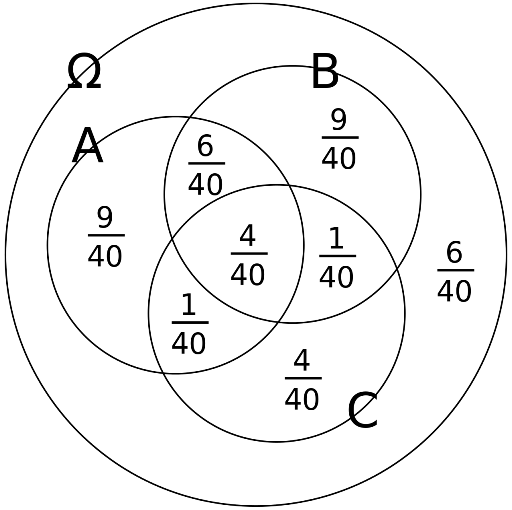
NEZAVISNOST VIŠE DOGAĐAJA

Nije dovoljno uzeti samo svih n niti svaka dva!

Kažemo da su događaji $A_1, \dots, A_n$ nezavisni ako vrijedi

$$P\left(A_{i_1} \cap \cdots \cap A_{i_k}\right) = P\left(A_{i_1}\right) \cdots P\left(A_{i_k}\right)$$

za svaki odabir podskupa indeksa $\{i_1, \dots, i_k\} \subseteq \{1, \dots, n\}$.



Na slici su dane vjerojatnosti za tri događaja.

Svaka dva događaja su nezavisna, ali sva tri nisu.

https://en.wikipedia.org/wiki/Independence_(probability_theory)

Nađite primjer tri skupa za koje vrijedi $P(A \cap B \cap C) = P(A)P(B)P(C)$, ali tako da nikoja dva među njima nisu nezavisna.

VAŽNOST ODABIRA MODELA

Kako uvjeriti učenike? Aktivnost mjerenja relativnih frekvencija!

Bacamo dva novčića. Kolika je vjerojatnost da ćemo dobiti dva pisma?

PRVI MODEL: Razlikujemo koji je novčić prvi, a koji drugi (npr. crveni i plavi).
Tada je $\Omega = \{pp, pg, gp, gg\}$. Svi događaji su jednako vjerojatni.
Tražena vjerojatnost je ¼. Ovo je Laplaceov model.

Vjerojatnost nije 1/3!

DRUGI MODEL: Ne razlikujemo novčiće. Tada je $\Omega = \{\{p, p\}, \{p, g\}, \{g, g\}\}$.
Nisu svi događaji jednako vjerojatni: P({p,p})=1/4, P({p,g})=1/2, P({g,g})=1/4.
Tražena vjerojatnost je ¼. Ovaj model nije Laplaceov.

UREĐENI I NEUREĐENI PAROVI

	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Bacamo dvije kocke.

Kolika je vjerojatnost da će zbroj dobivenih brojeva biti 8?

PRVI MODEL: Modeliramo uređenim parovima (i,j). Laplaceov model.

Broj povoljnih ishoda je 5, a ukupan broj je 36. Vjerojatnost je 5/36.

DRUGI MODEL: Modeliramo neuređenim parovima {i,j}. Ishod {3,5} je dvostruko vjerojatniji od ishoda {4,4}. Model NIJE Laplaceov.

Ukupan broj neuređenih parova je 21, od toga ih 6 ima vjerojatnost 1/36, a 15 ih ima vjerojatnost 2/36. Povoljnih ishoda je 3 i njihove vjerojatnosti su 2/36, 2/36 i 1/36, pa je tražena vjerojatnost 5/36 (ne 3/21).

TKO JE U PRAVU?

U teniskom klubu je 12 dječaka i 13 djevojčica. Među djevojčicama su tri sestre. Kolika je vjerojatnost da ako odaberemo dvije osobe da će one biti sestre?

Treba li modelirat pomoću uređenih ili neuređenih parova?

OBA NAČINA SU ISPRAVNA I OBA SU LAPLACEOVA JER LJUDE RAZLIKUJEMO!
NEMA NEUREĐENOG PARA KOJI JE MANJE VJEROJATAN OD DRUGIH.

POSTAVLJANJE LAPLACEOVOG MODELA

U kutiji je jedna crvena i pet bijelih kuglica označenih brojevima 1, 2, 3, 4, 5.

Ne gledajući
= slučajno!

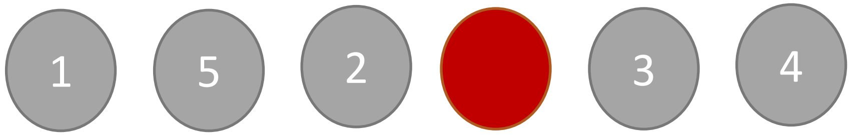
Ne gledajući, Domagoj izvlači jednu po jednu kuglicu sve dok ne izvuče crvenu kuglicu, nakon čega prekida izvlačenje. Izvučene kuglice se ne vraćaju u kutiju.

Kolika je vjerojatnost da je zbroj brojeva na izvučenim bijelim kuglicama barem 10?

Prvi način: uvjetna vjerojatnost – vjerojatnost da će crvena kuglica biti prva nije ista kao vjerojatnost da će biti druga itd.

POSTAVLJANJE LAPLACEOVOG MODELA

Drugi način: Odgovor se ne mijenja ako nastavimo kuglice izvlačiti do kraja! (Zašto?)



Prostor svih elementarnih događaja se sastoji od svih permutacija šest kuglica.

Svaki je jednako vjerojatan, tj. imamo Laplaceov model.

Trebamo prebrojiti u koliko permutacija je zbroj prije crvene kuglice barem 10.

Ispisujemo sve mogućnosti prema broju kuglica, pa kao u prethodnom primjeru za svaki neuređeni skup pribrojnika prebrojavamo na koliko načina ih možemo poredati. Ukupno imamo 288 povoljnih slučajeva od 6!=720 permutcija.

Tražena vjerojatnost je 288/720=2/5.

Kombinatorika: prebrojavanje permutacija

FERMATOVO PISMO PASCALU

U igri se baca novčić. Fermat pobjeđuje ako se u nizu pojavi 10 puta glava, a Pascal pobjeđuje ako se u nizu pojavi 10 puta pismo. Pobjednik treba dobiti 100 franaka. Igra se morala završiti u trenutku kad se pojavilo 8 puta pismo i 7 puta glava. U kojem omjeru je pravedno podijeliti nagradu?

PPPP	PGPP	PGGP	GGGP	PGGG
PPGP	PGPG	GPGP	GGGG	GPGG
PPPG	GPPP	GGPP		GGPG
PPGG	GPPG			

Sustavno ispisivanje i prebrojavanje

11/16 5/16

MALO OPĆENITIJA DEFINICIJA

Promotrimo (matematičku, hipotetsku) situaciju u kojoj je zadan

- konačan ili prebrojiv skup elementarnih događaja Ω
- realan broj $p_i = P(a_i) \in [0,1]$ za svaki element $a_i \in \Omega$

za koje vrijedi

$$\sum_i P(a_i) = 1.$$

Tada za svaki podskup $A \subseteq \Omega$, možemo definirati vjerojatnost

$$P(A) = \sum_{a \in A} P(a)\,.$$

Koja je vjerojatnost praznog skupa?

Svojstva. $P(A^c) = 1 - P(A), P(A \cup B) = P(A) + P(B) - P(A \cap B).$

GEOMETRIJSKA VJEROJATNOST

Što sve pretpostavljamo?

Primjer. Dva prijatelja su se dogovorila da će se naći na trgu između 12 i 13 sati. Tko prvi stigne, čeka drugog 15 minuta i nakon toga odlazi.

Kolika je vjerojatnost da će se prijatelji sresti?

Izazov za učenike: postavljanje vjerojatnosnog prostora.

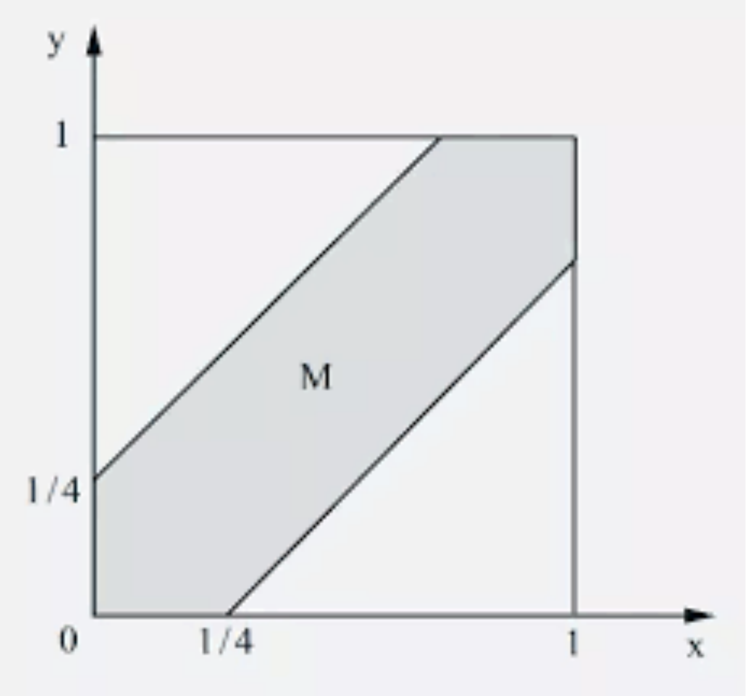
Koje je prosječno vrijeme čekanja?

Neka su x i y vremena kojima su prijatelji došli na trg, računajući da je 12 sati u 0, a 13 sati u 1.

Uvjet da su se sreli glasi $|x - y| \leq \frac{1}{4}$.

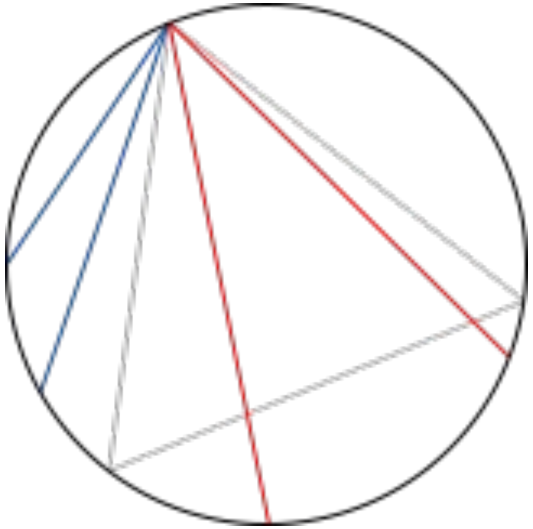
Vjerojatnost da je uvjet zadovoljen je

$$P(M) = \frac{\mu(M)}{\mu(\Omega)} = 1 - 2 \cdot \frac{1}{2} \cdot \frac{3}{4} \cdot \frac{3}{4} = \frac{7}{16}.$$



BERTRANDOV PARADOKS

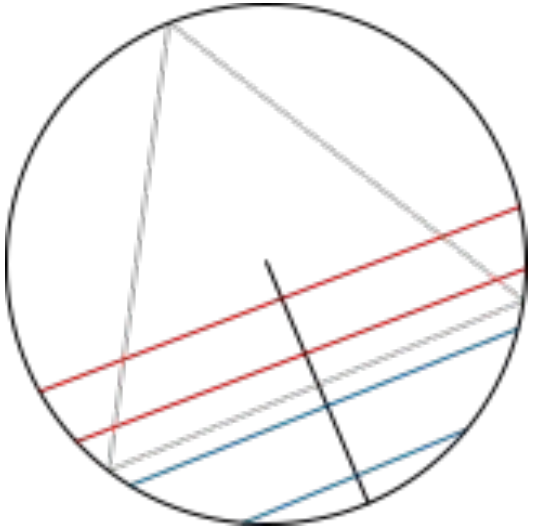
Koja je vjerojatnost da je tetiva dulja od stranice upisanog jednakostraničnog trokuta?



DVIJE TOČKE - KUT

$\Omega = [0, 180]$

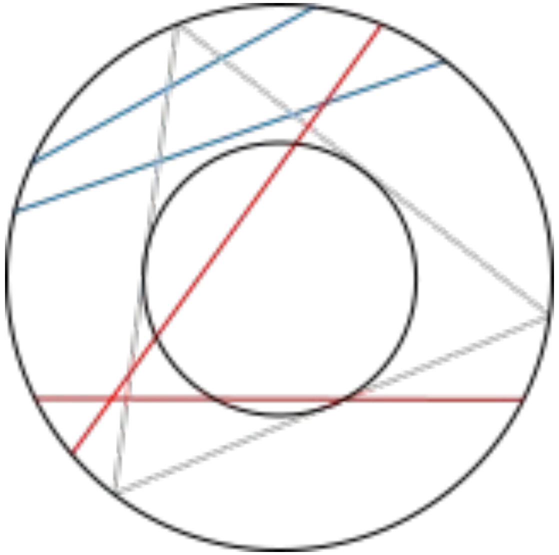
$P = \frac{1}{3}$



TOČKA NA POLUMJERU

$\Omega = [0, r]$

$P = \frac{1}{2}$



POLOVIŠTE – MANJI KRUG

$\Omega = [0, r^2\pi]$

$P = \frac{1}{4}$

POTREBA ZA JOŠ BOLJOM DEFINICIJOM

U prošlom primjeru smo vidjeli definiciju geometrijske vjerojatnosti:

$$P(M) = \frac{\mu(M)}{\mu(\Omega)}.$$

Izraz $\mu(M)$ označava površinu, a općenito može označavati bilo koju mjeru.

Imamo potrebu modelirati ovisnosti o varijablama koje nisu diskretne, već kontinuirane. Također, prostor elementarnh događaja je beskonačan, i to neprebrojivo!

Problem: nema svaki podskup kvadrata površinu, svaki podskup dužine duljinu.

Primjer. *Vitalijev skup* je podskup od [0,1] pri čemu smo odabrali po jednu točku iz svake klase $\mathbb{R}/\mathbb{Q}$.

AKSIOMATSKA DEFINCIJA VJEROJATNOSTI

Definicija (Kolmogorov 1933.) Vjerojatnosni prostor se sastoji od

- skupa elementarnih događaja Ω
- σ-algebre $\mathcal{A}$ na skupu Ω
- funkcije vjerojatnosti $P: \mathcal{A} \to [0,1]$

za koje vrijedi

(normiranost) $P(\Omega) = 1$ i

(σ-aditivnost) za svaki niz u parovima disjunktnih događaja $(A_n)_n$ u $\mathcal{A}$ vrijedi

$$P\left(\bigcup_n A_n\right) = \sum_n P(A_n).$$

(Familija $\mathcal{A}$ podskupova od Ω je σ-algebra ako sadrži Ω, te je zatvorena na komplemente i prebrojive unije.)

KLASIČNE DEFINICIJE VJEROJATNOSTI

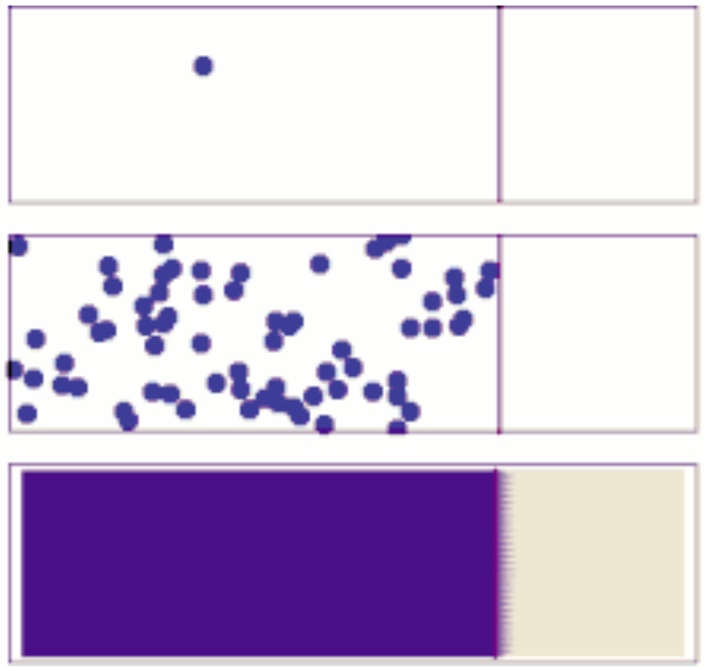
Eksperimentalna (a posteriori) definicija

- *Zakon velikih brojeva* je teorem koji možemo dokazati na temelju aksiomatske definicije.
- Ako promatramo ishode slučajne varijable koja ima binomnu razdiobu, onda relativne frekvencije $\frac{n_A}{n}$ teže P(A).

Teorijska (a priori) definicija

- Laplaceov model
- Konačno mnogo jednakovjerojatnih elementarnih događaja u Ω
- $\mathcal{A} = \wp(\Omega)$ partitivni skup

ZAKON VELIKIH BROJEVA



Neka su $X_1, X_2, \ldots$ nezavisne jednako distribuirane slučajne varijable s očekivanjem μ. Tada vrijedi

$$S_n = \frac{X_1 + \cdots + X_n}{n} \to \mu$$

KONVERGENCIJA PO VJEROJATNOSTI (SLABI ZAKON)

$$\lim_n P(|S_n - \mu| < \varepsilon) = 1 \text{ za svaki } \varepsilon > 0$$

KONVERGENCIJA GOTOVO SIGURNO (JAKI ZAKON)

$$P(\lim_n S_n = \mu) = 1$$

HVALA NA PAŽNJI!