TEORIJSKE DISTRIBUCIJE VJEROJATNOSTI

Mirjana Kujundžić Tiljak i Davor Ivanković

Za razliku od empirijskih distribucija frekvencija opaženih podataka, ***teorijske distribucije vjerojatnosti*** opisane su matematičkim modelom.

Kada neka empirijska distribucija aproksimira, tj. slijedi određenu teorijsku distribuciju vjerojatnosti, možemo upotrijebiti teorijsko znanje o dotičnoj distribuciji kako bismo dobili odgovore na pitanja o podacima. To najčešće zahtjeva procjenu vjerojatnosti.

VJEROJATNOST DOGAĐAJA mjera je neizvjesnosti. Mjeri šansu da se određeni događaj dogodi. Pozitivan je broj i može poprimiti vrijednosti između 0 i 1.

Ako je vjerojatnost 0, onda je događaj nemoguć.
Ako je vjerojatnost 1 , onda je događaj nužan, tj. mora se dogoditi.
Vjerojatnost komplementarnog događaja (1-vjerojatnost) mjeri šansu da se promatrani događaj ne dogodi.

Postoje različiti pristupi u računanju vjerojatnosti:

- *subjektivan pristup* podrazumijeva osobni stupanj vjerovanja da će se događaj dogoditi (npr. da će svijet propasti 2050. godine);
- *frekvencijski pristup* temelji se na brojanju događaja pri nebrojenom ponavljanju eksperimenta (npr. koliko puta će novčić pasti na glavu ako ga 1000 puta bacimo);
- *a priori pristup* pretpostavlja poznavanje teorijskog modela, tj. distribucije svih mogućih vjerojatnosti nekog događaja (npr. boja očiju djeteta majke s plavim i oca sa smeđim očima)

Pravila vjerojatnosti:

- *Pravilo adicije*: ako se dva događaja (A i B) međusobno isključuju, vjerojatnost da se dogodi jedan od njih (A ili B) jednaka je sumi njihovih vjerojatnosti

  Vjerojatnost (A ili B) = Vjerojatnost (A) + Vjerojatnost (B)

- *Pravilo multiplikacije*: ako su dva događaja (A i B) međusobno nezavisni, vjerojatnost da se dogode oba događaja (A i B) jednaka je umnošku njihovih vjerojatnosti.

  Vjerojatnost (A i B) = Vjerojatnost (A) × Vjerojatnost (B)

SLUČAJNA VARIJABLA, *x*, je varijabla koja poprima pojedinačne vrijednosti s određenom vjerojatnošću. Dva su osnovna tipa:

- *diskretna ili diskontinuirana slučajna varijabla:* numeričke vrijednosti su cijeli brojevi (primjer: broj nezrelih stanica u nekom preparatu može biti 0, 1, 2, 3, ... k);

- *kontinuirana slučajna varijabla:* numeričke vrijednosti su realni brojevi (primjer: tjelesna težina 72,35 kg, glukoza u krvi 7,2 mmol/l).

DISTRIBUCIJA (RASPODJELA) VJEROJATNOSTI prikazuje način na koji je ukupna vjerojatnost (koja je jednaka 1) raspodijeljena na pojedine vrijednosti slučajne varijable.
Svaku distribuciju vjerojatnosti definiraju *parametri* (npr. prosjek, varijanca).
Zavisno od tipa slučajne varijable i distribucije dijelimo na *diskretne* i *kontinuirane.*

DISKRETNE DISTRIBUCIJE VJEROJATNOSTI

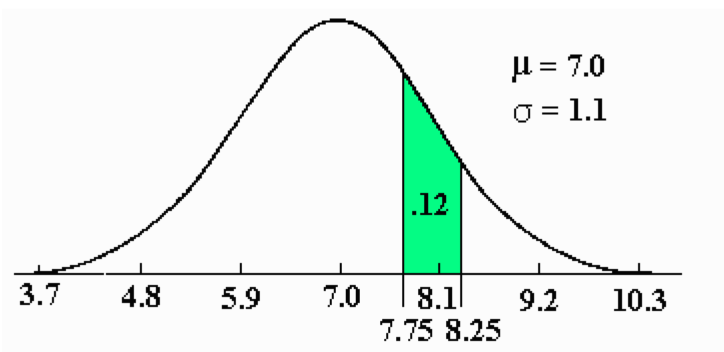
- primjeri: Binomna raspodjela, Poissonova raspodjela

Možemo izvesti vjerojatnost za svaku moguću vrijednosti slučajne varijable. *Suma svih mogućih vjerojatnosti slučajne varijable je 1.*

KONTINUIRANE DISTRIBUCIJE VJEROJATNOSTI

* primjeri: normalna raspodjela, χ^2 raspodjela; t raspodjela, F raspodjela
Možemo izvesti vjerojatnost za slučajnu varijablu, *x*, koja poprima vrijednosti u određenim razredima (budući da ima beskonačno mnogo vrijednosti *x*)
Ako horizontalna os predstavlja vrijednosti varijable *x,* prema jednadžbi distribucije može se nacrtati krivulja. Ova jednadžba zove se *funkcija gustoće vjerojatnosti*.

Ukupna površina ispod krivulje iznosi 1 i predstavlja vjerojatnost svih mogućih događaja. Vjerojatnost da x leži između dvije vrijednosti jednaka je površini ispod krivulje između te dvije vrijednosti.



Slika 1. Vjerojatnost da x leži između dvije vrijednosti
Izvor slike: http://www.psychstat.missouristate.edu/introbook/sbk11.htm

NORMALNA (GAUSSOVA) RASPODJELA

Najvažnija je distribucija kontinuiranih obilježja. Opisao ju je njemački matematičar C.F. Gauss. Većina bioloških mjerenja slijedi normalnu distribuciju. Primjenjuje se u mnogim modelima analize podataka.

Obilježja normalne raspodjele:

- u potpunosti ju opisuju dva parametra: aritmetička sredina (engl. mean, μ) i varijanca (engl. variance, σ^2) te se simbolički prikazuje kao $N(\mu,\sigma^2)$;

- područje vrijednosti slučajne varijable, x, je $(-\infty,+\infty)$;
- funkcija gustoće vjerojatnosti je:

  $f(x) = (1/\sigma\sqrt{2\pi})\, e^a$

  pri čemu je $a = -1/2\ ((x-\mu)/\sigma)^2$;

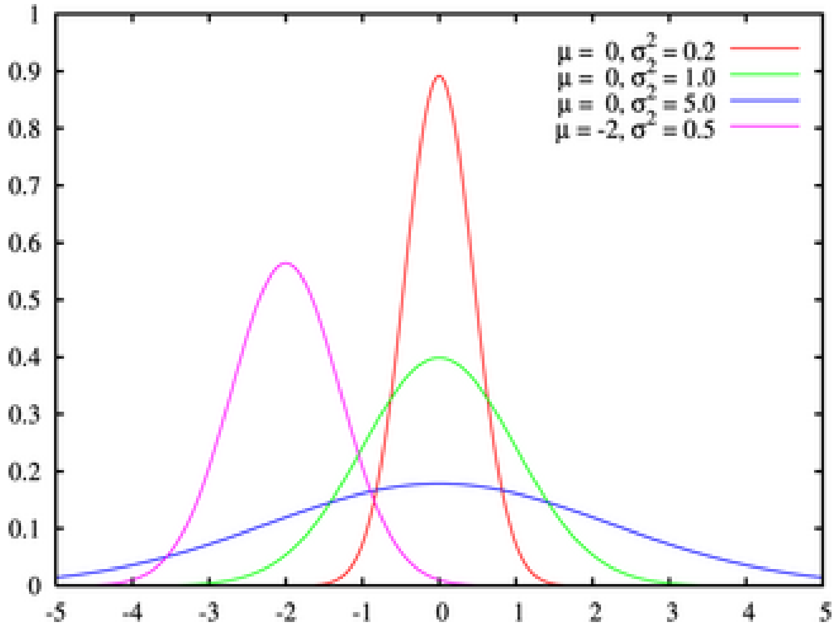
- zvonolikog je oblika i unimodalna;
- simetrična oko aritmetičke sredine;
- ukoliko se povećava vrijednost aritmetičke sredine krivulja se pomiče udesno, a ukoliko se vrijednost aritmetičke sredine smanjuje krivlja se pomiče ulijevo (uz pretpostavku jednake varijance);
- ukoliko se vrijednost varijance povećava krivulja se snižava se i širi, a ukoliko se vrijednost varijance smanjuje krivlja se povisuje i suzuje (uz nepromijenjenu aritmetičku sredinu);
- aritmetička sredina i medijan poprimaju istu vrijednost;

- vjerojatnost da će normalno distribuirana slučajna varijabla, x, s aritmetičkom sredinom, μ, i standardnom devijacijom, σ, poprimiti vrijednost između:

  $(\mu - \sigma)$ i $(\mu + \sigma)$ iznosi 0,68;

  $(\mu - 1{,}96\ \sigma)$ i $(\mu + 1{,}96\ \sigma)$ iznosi 0,95;

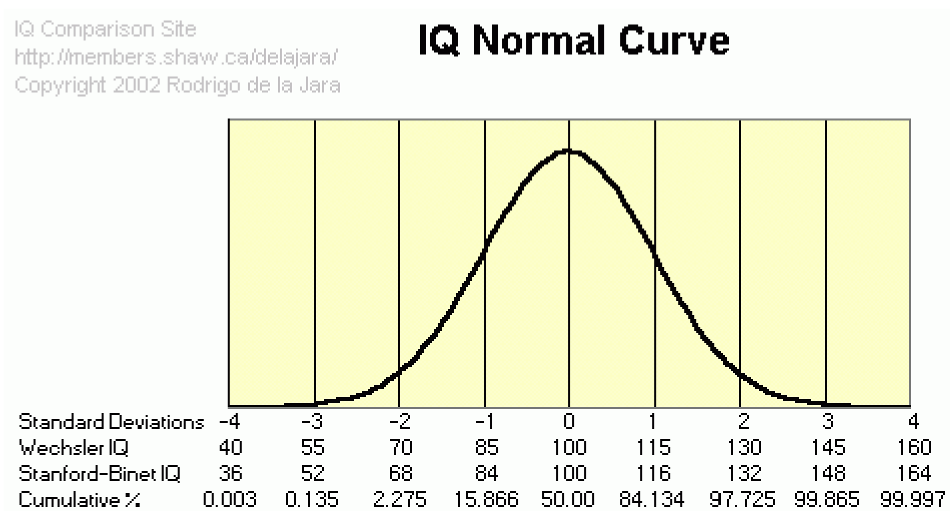
  $(\mu - 2{,}58\ \sigma)$ i $(\mu + 2{,}58\ \sigma)$ iznosi 0,99.



Slika 2. Gustoća vjerojatnosti normalne distribucije

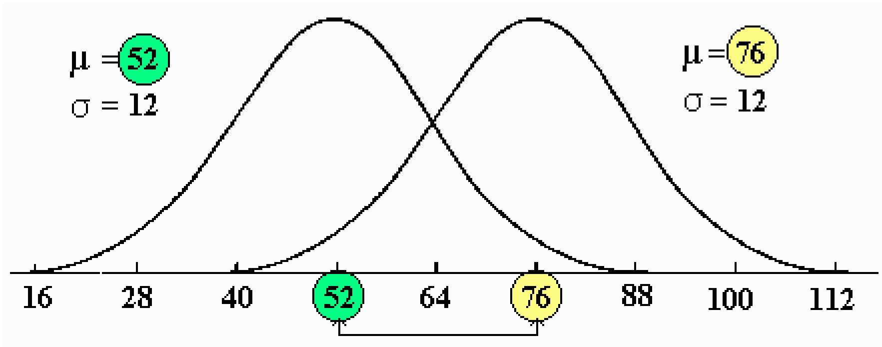
* zeleno = standardna normalna distribucija

Izvor slike: http://en.wikipedia.org/wiki/Normal_distribution

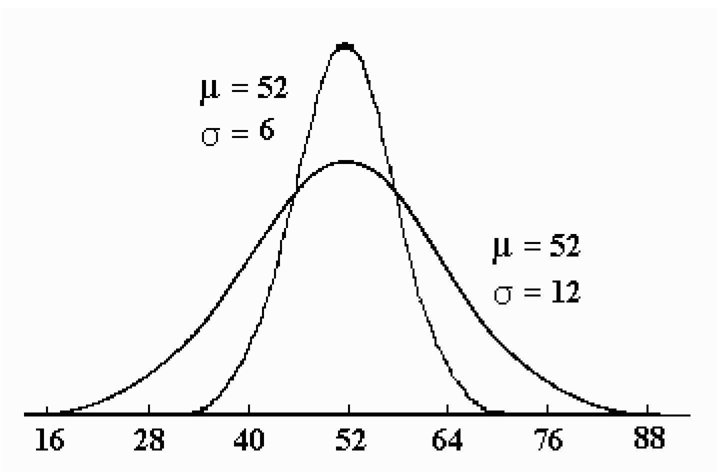


Slika 3.Gustoća vjerojatnosti normalne raspodjele kvocijenta inteligencije (IQ)

Izvor slike: http://members.shaw.ca/delajara/IQBasics.html



Slika 4.
Krivulje normalne raspodjele s različitim vrijednostima aritmetičke sredine uz jednake vrijednosti varijanci
Izvor slike: http://www.psychstat.missouristate.edu/introbook/sbk11.htm



Slika 5..
Krivulje normalne raspodjele s različitim vrijednostima varijanci uz jednake vrijednosti aritmetičke sredine
Izvor slike: http://www.psychstat.missouristate.edu/introbook/sbk11.htm

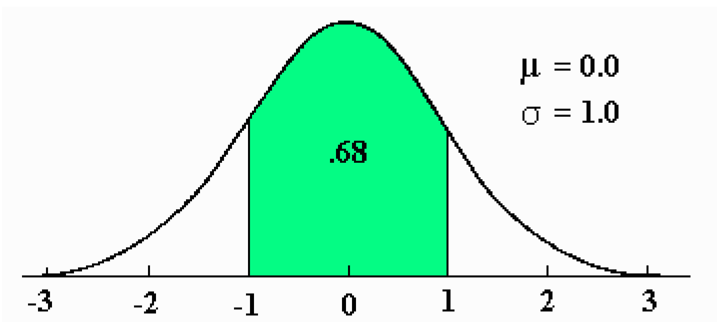
STANDARDNA NORMALNA RASPODJELA

Standardna normalna raspodjela ima vrijednost aritmetičke sredine 0, a vrijednost varijance 1. Simbolički se prikazuje N(0,1).

Vrijednosti na horizontalnoj osi su *standardizirane varijable, z - vrijednosti,* a računaju se:

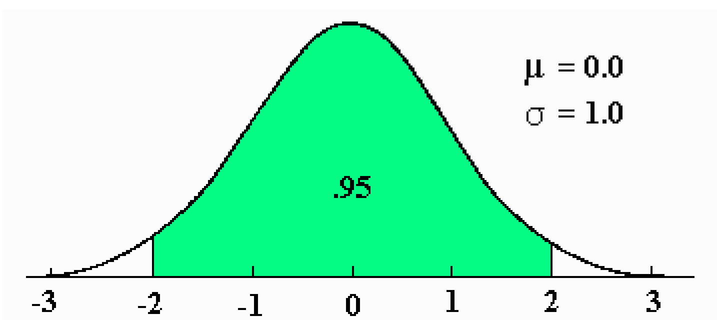
$$z = \frac{x - \mu}{\sigma}.$$

Z-vrijednosti određuju položaj pojedinog rezultata u nekoj normalnoj raspodjeli, i to tako da se ta vrijednost izrazi u dijelovima standardne devijacije.



Slika6. Krivulja standardne normalne raspodjele i vjerojatnost da će x poprimiti vrijednost između (μ-1σ) i (μ+1σ).

Izvor slike: http://www.psychstat.missouristate.edu/introbook/sbk11.htm



Slika6. Krivulja standardne normalne raspodjele i vjerojatnost da će x poprimiti vrijednost između (μ-1,96σ) i (μ+1,96σ).

Izvor slike: http://www.psychstat.missouristate.edu/introbook/sbk11.htm

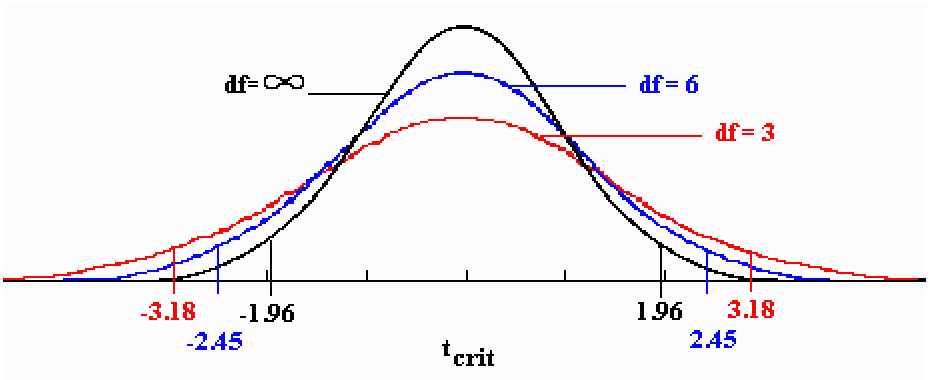
DRUGE KONTINUIRANE DISTRIBUCIJE VJEROJATNOSTI

t – DISTRIBUCIJA

Derivirao ju je W.S.Gosset, poznatiji pod pseudonimom „Student“ pa se distribucija zove i *Studentova t-distribucija.*

Obilježja t-distribucije:

- karakteriziraju je *stupnjevi slobode*;
- ima sličan oblik kao normalna distribucija samo što je šira i položenija;
- kako raste broj stupnjeva slobode oblikom je sve sličnija normalnoj raspodjeli;
- primjenjuje se u računanju intervala pouzdanosti i testiranju hipoteza o razlici između dva uzorka.



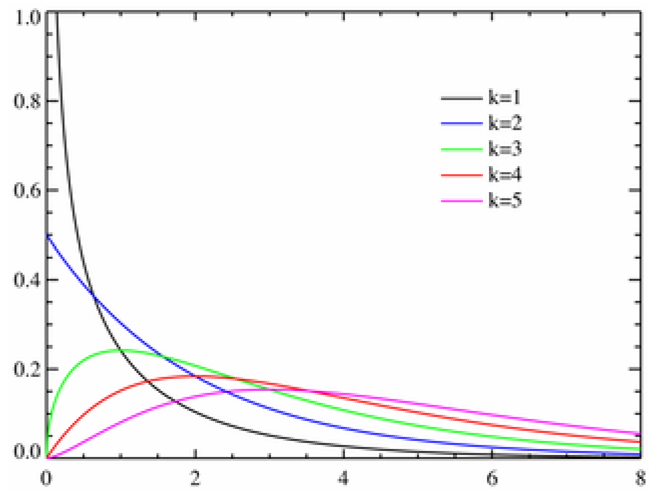
Slika 7. Krivulje t - distribucije

Izvor slike: http://www.psychstat.missouristate.edu/introbook/sbk24.htm

χ^2 – DISTRIBUCIJA

Obilježja χ^2-distribucije:

- distribucija je pozitivnih vrijednosti, zakrivljena u desno;
- karakteriziraju je *stupnjevi slobode*;
- oblik distribucije ovisi o broju stupnjeva slobode: kako raste broj stupnjeva slobode distribucija postaje sve više simetrična i sličnija normalnoj distribuciji
- primjenjuje se u analizi kategorijskih podataka.

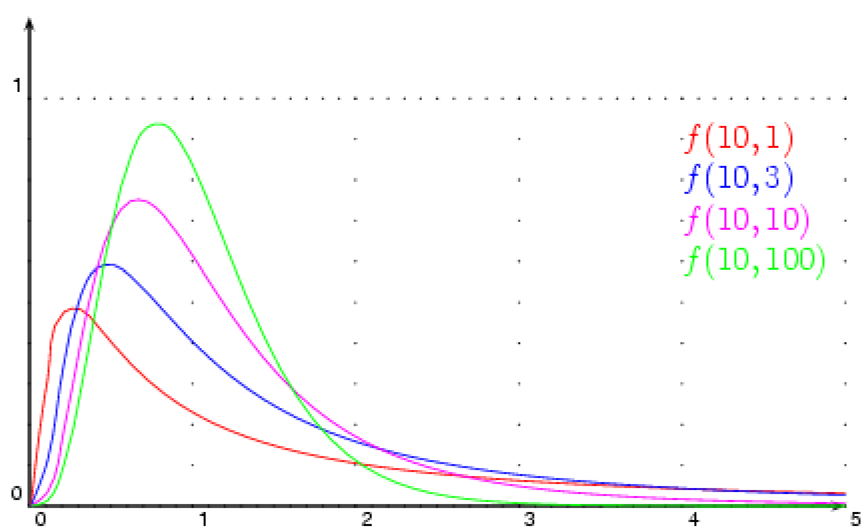


Slika 9. Krivulje χ^2- distribucije
Izvor slike: http://en.wikipedia.org/wiki/Chi-square_distribution

F – DISTRIBUCIJA

Obilježja F-distribucije:

- zakrivljena prema desno;
- distribucija je omjera dvaju varijanci izračunatih iz normalno distribuiranih podataka;
- karakteriziraju je stupnjevi slobode brojnika i nazivnika omjera varijanci;
- upotrebljava se za usporedbu dvije varijance, kao i za usporedbu više od dvije aritmetičke sredine analizom varijance (ANOVA).

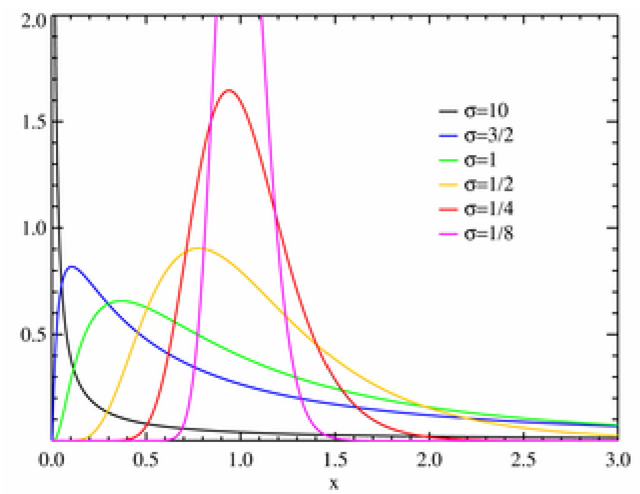


Slika 9. Krivulje F- distribucije
Izvor slike: http://planetmath.org/encyclopedia/FDistribution.html

LOGNORMALNA DISTRIBUCIJA

Obilježja lognormalne distribucije:

- distribucija vjerojatnosti slučajne varijable čiji logaritmi (baze 10 ili e) slijede normalnu distirbuciju;
- izrazito je zakrivljena u desno;
- kada logaritmiramo vrijednosti varijable čija distribucija je zakrivljena prema desno, dobiveni logaritmi slijede normalnu distribuciju;
- geometrijska sredina;
- kao mjera centralne tendencije upotrebljava se geometrijska sredina.



Slika 10. *Krivulje lognormalne distribucije*
Izvor slike: http://en.wikipedia.org/wiki/Lognormal_distribution

DISKRETNE DISTRIBUCIJE VJEROJATNOSTI

Slučajna varijabla koja definira distribuciju vjerojatnosti je diskretna. Zbroj vjerojatnosti svih mogućih, međusobno isključujućih događaja iznosi 1.

BINOMNA DIOSTRIBUCIJA

Binomnu distribuciju definirao je Jacob Bernuolli 1700. godine. To je teorijska distribucija za diskretne slučajne varijable kada imamo dva moguća ishoda, "uspjeh" i "neuspjeh", u n mogućih događaja.

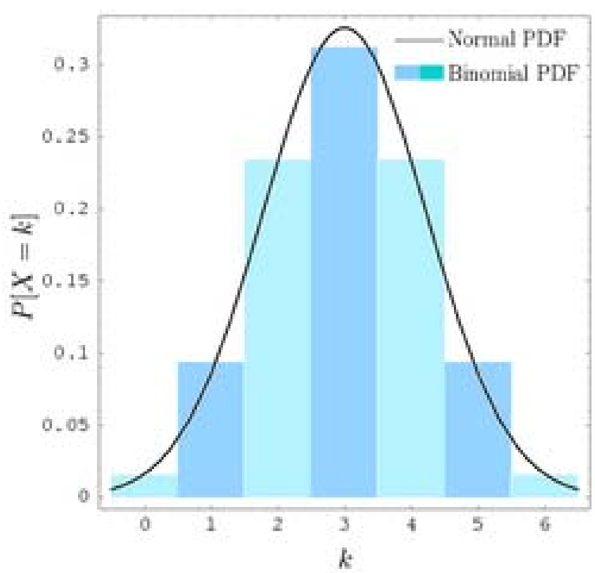
Primjer: Imamo 100 pokušaja IVF (in vitro fertilizacija) kod kojih ishod moža biti "uspjeh" (trudnoća) ili "neuspjeh".

$$P(x) = \binom{n}{x} p^x q^{n-x}$$ je vjerojatnost da u seriji od n pokusa događaj A nastupi točno x puta.

Skup svih parova $\{x, P(x)\}$, $x = 0, 1, 2, ..., n$ čini binomnu distribuciju, točnije *gustoću binomne distribucije.*

Obilježja binomne distribucije

- opisuju je dva parametra: broj jedinki u uzorku ili broj ponavljanja pokušaja, n, i stvarna vjerojatnost uspjeha za svaki jedinku ili za svaki pokušaj, π, te se simbolički prikazuje kao $B(n, \pi)$;
- *aritmetička sredina binomne distribucije* je vrijednost slučajne varijable koju očekujemo ako promatrano n jedinki ili pokušaj ponavljano n puta, i iznosi $n\pi$;
- *varijanca binomne distribucije* iznosi $n\pi(1-\pi)$;
- kada je n male vrijednosti, distribucija je zakrivljena u desno ako je $\pi<0{,}5$ a ako je $\pi>0{,}5$ distribucija je zakrivljena u lijevo;
- kako veličina uzorka raste distribucija postaje više simetrična;
- binomna distribucija aproksimira normalnu distribuciju ako su joj vrijednosti i aritmetičke sredine i varijance veće od 5;
- upotrebljava se pri zaključivanju o proporcijama.



Slika 11. Funkcija gustoće binomne distribucije s aproksimacijom normalne raspodjele za n=6 i p = 0,5

Izvor slike: http://en.wikipedia.org/wiki/Binomial_distribution

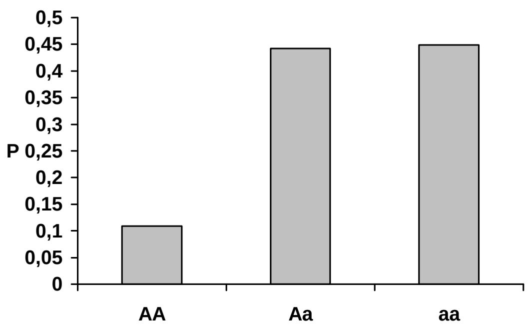
Primjeri binomne distribucije:

(1) *vjerojatnost genotipova*

- frekvencija gena A = 0,33
- frekvencija gena a = 0,67

$(p+q)^2 = (0{,}33 + 0{,}67)^2 = 0{,}332 + 2 \times 0{,}33 \times 0{,}67 + 0{,}672$

Prostor ishoda:
P (AA)= 0,332 = 0.109
P (Aa) = 0,33 ×0,67 = 0,221
P (aA) = 0,67 * 0,33 = 0,221
P (aa) = 0,672 = 0,449



Slika 12. Binomna distribucija vjerojatnosti genotipova

(2) *smrtni ishod kao binomna distribucija*

- letalitet od neke bolesti = 0,30
- vjerojatnost preživljenja = 0,70
- n = 5
- binom: $(0{,}30 + 0{,}70)^5$

Tablica 1. Prostor ishoda binomne distribucije smrtnog ishoda

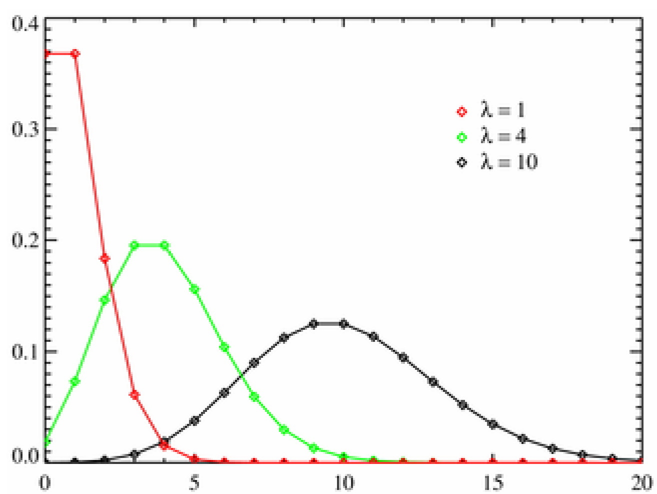
Broj umrlih	Razvoj binoma	Vjerojatnost
5 (svi)	p^5	0,003
4	$5p^4q$	0,028
3	$10p^3q^2$	0,132
2	$10p^2q^3$	0,309
1	$5pq^4$	0,360
0 (nijedan)	q	0,168
Ukupno		1,000

POISSONOVA DISTRIBUCIJA

Opisao ju je Siméon Denis Poisson početkom XIX stoljeća.
Poissonova slučajna varijabla je broj događaja koji se zbivaju neovisno i slučajno u vremenu ili prostoru, s prosječnom frekvencijom, μ (npr. rast bakterija na hranjivoj podlozi, broj prijema u bolnicu na dan).

Obilježja Poissonove distribucije:

- parametar koji opisuje Poissonovu distribuciju je aritmetička sredina, tj. prosječna frekvencija, μ;
- aritmetička sredina i varijanca imaju jednake vrijednosti;
- unimodalna je krivulja, zakrenuta u desno kada je vrijednost aritmetičke sredine mala;
- kako raste aritmetička sredina, asimetrija se smanjuje i na kraju aproksimira normalnu raspodjelu.



Slika 13. Krivulje Poissonove distribucije
(horizontalna os je indeks k;
napomena: funkcija je definirana cijelim brojevima k,
linija koja povezuje točke ne ukazuje na kontinuitet)
Izvor slike: http://en.wikipedia.org/wiki/Poisson_distribution

Literatura:

1. *Ivanković D, i sur. Osnove statističke analize za medicinare. Zagreb: Medicinski fakultet Sveučilišta u Zagrebu, 1989.*
2. *Petrie A, Sabin C. Medical Statistics at a Glance (2nd Ed). Oxford: Blackwell Science Ltd, 2005.*
3. *Glantz. SA. Primer of Biostatistics (4th Ed). New York: McGraww-Hill: 1997.*
4. *Altman DG. Practical Statistics for Medical Research. London. Chapman & Hall, 1991.*
5. *Bland M. An Introduction to Medical Statistics (3rd Ed). Oxford: Oxford University Press, 2005.*
6. *Armitage P, Berry P. Statistical Methods in Medical Research. Oxford: Blackwell Science Ltd, 1994.*