

IV.1.2. POISSONOVA RASPODJELA

Def:

Diskretna slučajna varijabla ima **Poissonovu raspodjelu** ako joj je raspodjela vjerojatnosti dana s

$$p(x; \lambda) = \frac{e^{-\lambda} \lambda^x}{x!} \quad x = 0, 1, 2, \dots$$

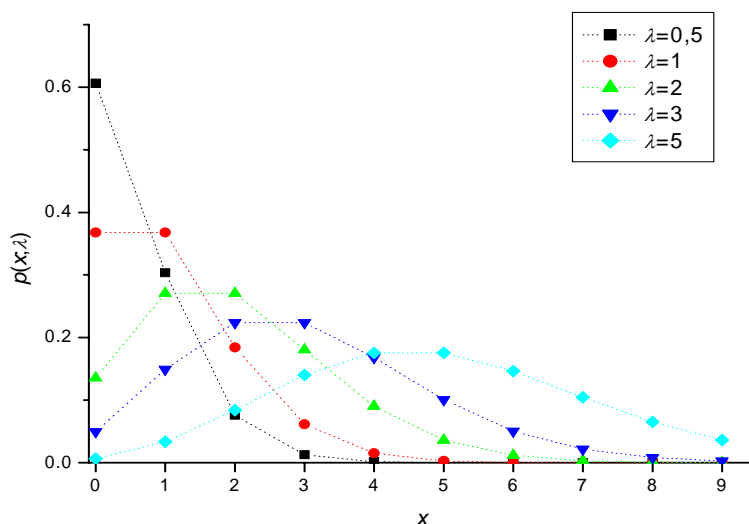
za neki $\lambda > 0$.

Poissonovu slučajnu varijablu označavamo:

$$X \sim Po(\lambda)$$

Pokažimo da je $p(x; \lambda)$ raspodjela vjerojatnosti (tj., da zadovoljava aksiome):

$$\sum_{x=0}^{\infty} p(x; \lambda) = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{-\lambda} e^{\lambda} = 1$$



Slika 1: Poissonova raspodjela za nekoliko različitih λ .

Očekivanje i varijanca

$$E(X) = \sum_{x=0}^{\infty} x e^{-\lambda} \frac{\lambda^x}{x!} = \sum_{x=1}^{\infty} x e^{-\lambda} \frac{\lambda^x}{x!} = \sum_{x=1}^{\infty} e^{-\lambda} \frac{\lambda^x}{(x-1)!} = \lambda \sum_{y=0}^{\infty} e^{-\lambda} \frac{\lambda^y}{y!} = \lambda$$

$$E(X^2) = \sum_{x=0}^{\infty} x^2 e^{-\lambda} \frac{\lambda^x}{x!} = \sum_{x=1}^{\infty} x e^{-\lambda} \lambda \frac{\lambda^{x-1}}{(x-1)!} = \lambda e^{-\lambda} \left[\sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} + \sum_{x=1}^{\infty} (x-1) \frac{\lambda^{x-1}}{(x-1)!} \right] =$$

$$\lambda e^{-\lambda} \left[e^{\lambda} + \lambda \sum_{x=2}^{\infty} (x-2) \frac{\lambda^{x-2}}{(x-2)!} \right] = \lambda e^{-\lambda} \left[e^{\lambda} + \lambda e^{\lambda} \right] = \lambda + \lambda^2$$

$$V(X) = E(X^2) - (E(X))^2 = \lambda$$

Rekurzivna formula

$$\frac{p(x+1; \lambda)}{p(x; \lambda)} = \frac{e^{-\lambda} \frac{\lambda^{x+1}}{(x+1)!}}{e^{-\lambda} \frac{\lambda^x}{x!}} = \frac{\lambda}{x+1} \quad \Rightarrow \quad p(x+1; \lambda) = \frac{\lambda}{x+1} p(x; \lambda)$$

Najvjerojatnija vrijednost

$$p(x_M - 1) \leq p(x_M) \geq p(x_M + 1) \quad \Rightarrow \quad \lambda - 1 \leq x_M \leq \lambda$$

Funkcija izvodnica

$$g(t) = \sum_{x=0}^{\infty} e^{tx} e^{-\lambda} \frac{\lambda^x}{x!} = e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t - 1)}$$

Pomoćni momenti

$$\begin{aligned} g'(t) &= \lambda e^t e^{\lambda(e^t - 1)} & m_1 &= \lambda \\ g''(t) &= \lambda e^t (1 + \lambda e^t) e^{\lambda(e^t - 1)} & m_2 &= \lambda^2 + \lambda \end{aligned}$$

Funkcija raspodjele

$$F(x; \lambda) = \sum_{y=0}^x e^{-\lambda} \frac{\lambda^y}{y!} \quad \rightarrow \quad \text{tablice}$$

Primjena Poissonove raspodjele

- A. Aproksimacija binomne raspodjele
- B. Proučavanje rijetkih događaja

A. Poissonova raspodjela kao granični slučaj binomne

Pretpostavimo da u $Bin(n, p)$ stavimo $n \rightarrow \infty$, a $p \rightarrow 0$ tako da $np = \lambda = \text{const.} > 0$. Tada $b(x; n, p) \rightarrow p(x; \lambda)$.

Dokaz:

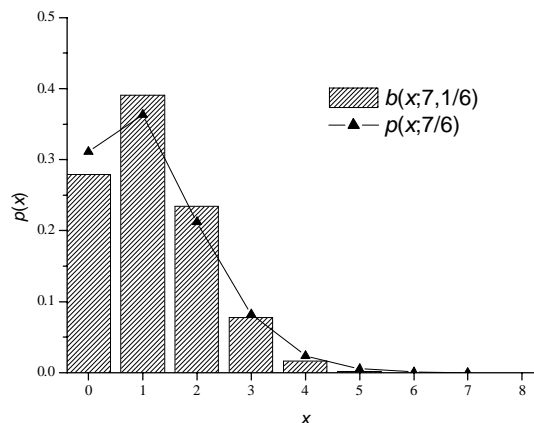
$$\begin{aligned} b(x; n, p) &= \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} = \frac{n!}{x!(n-x)!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} = \\ &= \frac{n(n-1) \cdots (n-x+1)}{n^x} \left(1 - \frac{\lambda}{n}\right)^{n-x} \frac{\lambda^x}{x!} = \\ &= 1 \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{x-1}{n}\right) \left(1 - \frac{\lambda}{n}\right)^n \frac{\lambda^x}{x!} \left(1 - \frac{\lambda}{n}\right)^{-x} \end{aligned}$$

$$\lim_{n \rightarrow \infty} b(x; n, p) = 1 \cdot 1 \cdots 1 \cdot e^{-\lambda} \cdot \frac{\lambda^x}{x!} = p(x; \lambda)$$

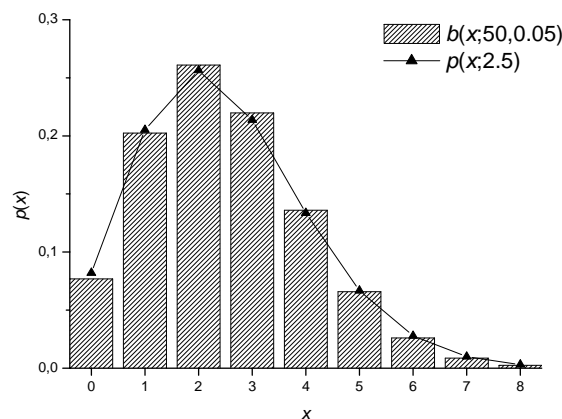
Upotrijebili smo relaciju $\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}$

Praktično binomnu raspodjelu možemo aproksimirati Poissonovom kad je:
 $n \geq 50$; $p \leq 0,1$

Slijede dva primjera, najprije loše, a onda dobre aproksimacije:



Usporedba binomne i Poissonove raspodjele za $n=7$ i $p=1/6$ (broj šestica prilikom bacanja 7 kockica).



Usporedba binomne i Poissonove raspodjele za $n=50$ i $p=0,05$.

B. Poissonova raspodjela kao zakon malih brojeva

Rijetki događaji:

- prometne nesreće na određenoj dionici autoceste u jednome danu
- telefonski pozivi centrali u jednoj minuti
- defekti po jedinici duljine bakrene žice
- broj čestica kozmičkog zračenja detektiranih u sekundi.
- broj oboljelih stabala po aru šume.

Primjer 1:

Odredi raspodjelu vjerojatnosti k raspada tijekom jedne milisekunde ako imamo $N=4,8 \cdot 10^{12}$ radioaktivnih atoma joda, a konstanta raspada je $\lambda=7,3 \cdot 10^{-10} \text{s}^{-1}$.

Broj radioaktivnih atoma u vremenu: $N = N_0 e^{-\lambda t}$;

Aktivnost = broj raspada u jedinici vremena: $a = \left| \frac{dN}{dt} \right| = N\lambda = 3,4 \cdot 10^3 \text{s}^{-1}$

$$P_k(t) = b(k; 1000, p(\Delta t)) = p(k; at)$$

$$E(k) = at = 3,4$$

primjer 2:

Knjiga od 750 stranica ima 400 tiskarskih pogrešaka. Uz pretpostavku da su pogreške nezavisne nađi vjerojatnost da neka stranica ima

a) nula pogrešaka
$$p\left(0; \frac{400}{750}\right) = e^{-\lambda} \frac{\lambda^0}{0!} = 0,59$$

b) više od jedne pogreške
$$1 - p(0) - p(1) = 0,10$$

c) točno 4 pogreške
$$p(4) = 0,002$$

Vidimo: U pitanju su vremenski ili prostorni intervali. Vjerojatnost događaja proporcionalna je veličini intervala. To su **Poissonovi procesi**.

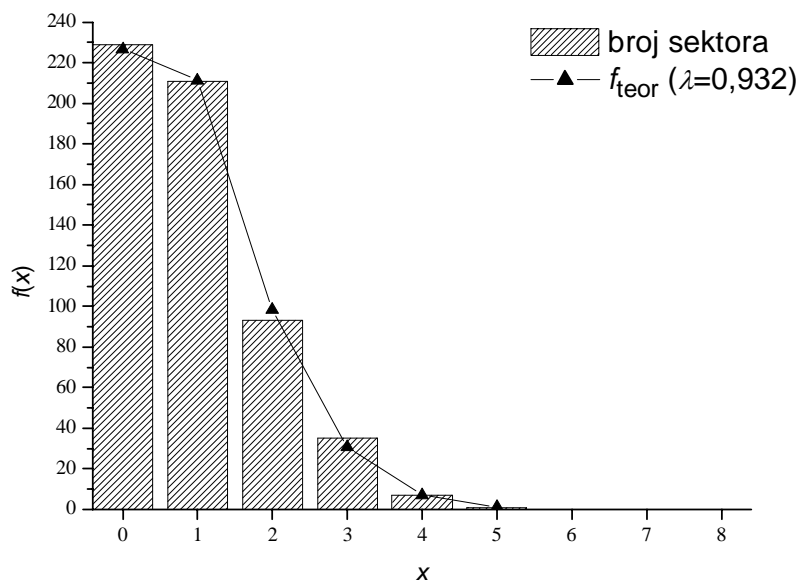
Prilagođavanje empirijskih podataka Poissonovoj raspodjeli

primjer:

Bombe koje su padale po Londonu u II svjetskom ratu.

Broj bombi po sektoru x	broj sektora $f_{\text{opaženo}}$	broj bombi $x \cdot f$	$f_{\text{teorijski}}$
0	229	0	227
1	211	211	211
2	93	186	99
3	35	105	31
4	7	28	7
>5	1	5	1
Σ	576	535	576

$$\bar{x} = \frac{535}{576} = 0,929 = E(x) = \lambda$$



Slika 4: uz primjer prilagodbe (bombe po Londonu)

Raspodjela dvije nezavisne Poissonove varijable

Imamo slučajne varijable $X \sim Po(m)$ i $Y \sim Po(n)$.

Tvrdnja: Za slučajnu varijablu $(X+Y)$ vrijedi $(X+Y) \sim Po(m+n)$.

dokaz:

definirajmo slučajnu varijablu $Z = X + Y$

$$\begin{aligned}
 P(Z = z) &= \sum_{w=0}^z P(X = w) \cdot P(Y = z - w) = \sum_{w=0}^z e^{-m} \frac{m^w}{w!} \cdot e^{-n} \frac{n^{z-w}}{(z-w)!} = \\
 &= \frac{e^{-(m+n)}}{z!} \sum_{w=0}^z \binom{z}{w} m^w n^{z-w} = \frac{e^{-(m+n)}}{z!} (m+n)^z
 \end{aligned}$$

Q.E.D.

IV.1.3. HIPERGEOMETRIJSKA RASPODJELA

Pretpostavke:

1. Populacija se sastoji od N elemenata. (konačna populacija)
2. Svaki element ima svojstvo koje možemo označiti kao 'uspješan' (A) ili 'neuspješan' (\bar{A}). Postoji M uspješnih elemenata.
3. Izvlači se uzorak od n elemenata, a izvlačenje bilo kojeg uzorka jednako je vjerojatno.

Slučajna varijabla:

X = broj uspješnih elemenata u uzorku

$$P(X = x) = h(x; n, M, N)$$

$$h(x; n, M, N) = \frac{\text{broj mogućih uzoraka s } x \text{ uspjeha}}{\text{ukupni broj mogućih uzoraka}}$$

$$h(x; n, M, N) = \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}} \quad x = 0, 1, 2, \dots, \min(n, M)$$

Očekivanje i varijanca

$$E(X) = n \frac{M}{N} \quad V(X) = \left(\frac{N-n}{N-1} \right) n \frac{M}{N} \left(1 - \frac{M}{N} \right)$$

Za lakše pamćenje stavimo $p = \frac{M}{N}$; $q = 1 - p$

$$E(X) = np \quad V(X) = \left(\frac{N-n}{N-1} \right) npq$$

Faktor $\left(\frac{N-n}{N-1} \right)$ nazivamo **korekcijskim faktorom** konačne populacije.

Primjer iz prvog dijela predavanja: Imali smo $N=100.000$; $n=20$. Izvlačili smo 100 puta. Raspodjelu smo prilagođavali binomnoj. Korekcija do hipergeometrijske bila bi vrlo mala.

Primjer hipergeometrijske raspodjele:

25 tigrova ugrožene vrste nalazi se u nekoj prašumi. Znanstvenici su već ranije uspjeli obilježiti 5 tigrova radi promatranja. U novom promatranju ulovimo 10 tigrova. Kolika je vjerojatnost da će manje od tri biti obilježena?

$$P(X < 3) = h(0;10,5,25) + h(1;10,5,25) + h(2;10,5,25) = 0,057 + 0,128 + 0,385 = 0,570$$

IV.1.4. NEGATIVNA BINOMNA (PASCALOVA) RASPODJELA

Uvjeti pokusa:

1. Sastoji se od niza nezavisnih pokušaja.
2. Pokušaji su međusobno identični Bernoullijevi pokusi s mogućim ishodima 'uspjeh' (A) i 'neuspjeh' (\bar{A}).
3. Vjerojatnost 'uspjeha' jednaka je za sve pokušaje i označavamo je s p .
4. Pokus traje (pokušaji se izvode) sve dok se ne opazi r uspjeha, a r je unaprijed određen prirodni broj.

Slučajna varijabla:

X = broj neuspjeha koji prethode r -tom uspjehu

Za razliku od binomne, ovdje je broj pokušaja varijabilan, a broj uspjeha unaprijed određen. (zato se zove negativna binomna)

Npr. : Ako je $r=5$, a $x=10$ znači da u prvih 14 pokušaja imamo 4 uspjeha i da je 15. Pokušaj uspješan

$$P(X = x) = nb(x; r, p)$$

$$nb(x; r, p) = P(r-1 \text{ uspjeha u } x+r-1 \text{ pokušaja}) \cdot P(\text{uspjeh})$$

$$nb(x; r, p) = \binom{x+r-1}{r-1} p^{r-1} q^x p$$

$$nb(x; r, p) = \binom{x+r-1}{r-1} p^r q^x \quad x = 0, 1, 2, \dots$$

primjer:

Pedijatar želi naći 5 parova koji očekuju prvo dijete da ih uključi u novi program prirodnog poroda. Ako je vjerojatnost pristanka $p=0,2$, kolika je vjerojatnost da će morati upitati 15 parova?

$$r + x = 15$$

$$nb(10; 5, 0,2) = \binom{14}{4} (0,2)^5 (0,8)^{10} = 0,034$$

Očekivanje i varijanca

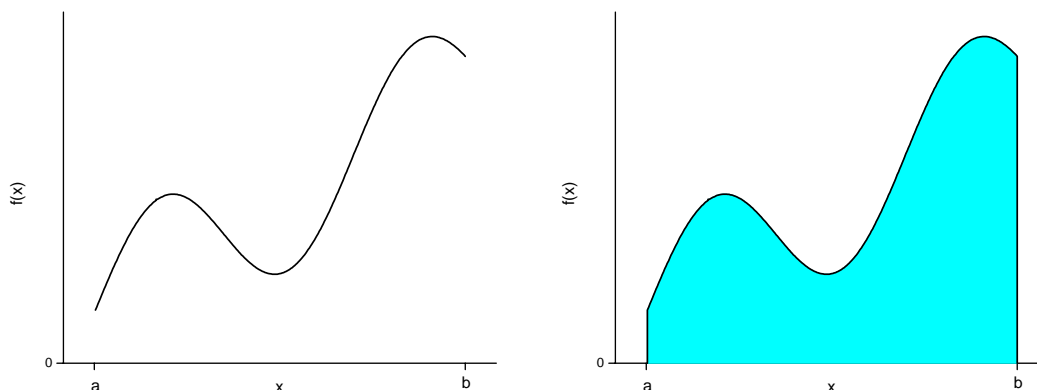
Bez izvoda:

$$E(X) = \frac{rq}{p}$$

$$V(X) = \frac{rq}{p^2}$$

POJAM INTEGRALA

Prilikom razmatranja kontinuiranih raspodjela često ćemo se susretati s problemom određivanja površine ispod funkcije.

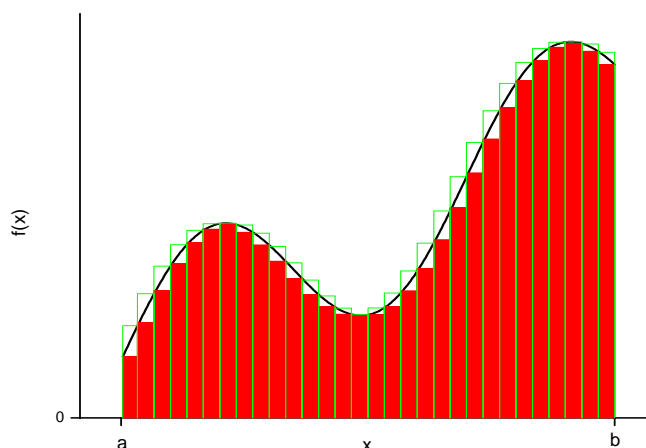


Budući da je funkcija prikazana zakrivljenom crtom, nema jednostavnog geometrijskog načina da točno izračunamo obojenu površinu.

Mi znademo izračunavati površine pravokutnika pa ćemo pokušati promatranu plohu podijeliti na manje dijelove koje ćemo tada aproksimirati pravokutnicima. Podijelimo promatrani interval $[a, b]$ na n ekvidistantnih segmenata $[x_{k-1}, x_k]$. Pritom je $x_0 = a$, a $x_n = b$. Na svakom segmentu pronađemo najveću (M_k) i najmanju (m_k) vrijednost funkcije f . Definirajmo sume:

$$s = \sum_{k=1}^n m_k (x_k - x_{k-1}) = \sum_{k=1}^n m_k \Delta x$$

$$S = \sum_{k=1}^n M_k (x_k - x_{k-1}) = \sum_{k=1}^n M_k \Delta x$$



Očito je da je naša tražena površina I veća od s , a manja od S . Što je veći broj segmenata n na koje smo podijelili promatrani interval, to je razlika između S i s manja pa je procjena iznosa površine I to bolja. Ako je promatrana funkcija f neprekinuta i konačna na promatranom intervalu, onda će se u granici $n \rightarrow \infty$ sume s i S izjednačiti.

$$\lim_{n \rightarrow \infty} s_n = \lim_{n \rightarrow \infty} S_n = I$$

Pišemo:

$$I = \lim_{n \rightarrow \infty} \sum_{k=1}^n f(x_k) \Delta x = \int_{[a,b]} f(x) dx = \int_a^b f(x) dx$$

Def:

Primitivna funkcija od f je svaka funkcija F sa svojstvom

$$F'(x) = f(x)$$

Sve primitivne funkcije neke funkcije f razlikuju se samo za konstantu!

Za neprekidne funkcije vrijedi **Leibniz-Newtonova formula**:

$$\int_a^b f(x) dx = F(b) - F(a)$$

Navodim nekoliko «tabličnih» primitivnih funkcija:

$f(x)$	$F(x)$
$x^n, \quad n \neq -1$	$\frac{x^{n+1}}{n+1}$
x^{-1}	$\ln x $
e^x	e^x
a^x	$\frac{a^x}{\ln a}$
$\sin x$	$-\cos x$
$\cos x$	$\sin x$
$\frac{1}{\cos^2 x}$	$\operatorname{tg} x$
$\frac{1}{\sin^2 x}$	$-\operatorname{ctg} x$
\vdots	\vdots

IV.2. KONTINUIRANA SLUČAJNA VARIJABLA I RASPODJELA VJEROJATNOSTI

Dosad smo razmatrali diskretne slučajne varijable. Moguće vrijednosti diskretne slučajne varijable čine konačan skup ili se mogu složiti u uređen beskonačan niz. U slučaju kontinuirane slučajne varijable, moguće vrijednosti čine cijeli interval brojeva.

Def: Slučajna varijabla X je **kontinuirana** ako skup njezinih mogućih vrijednosti čini cijeli interval brojeva, t.j. ako je za neke A i B ($A < B$), svaki $x \in [A, B]$ moguć.

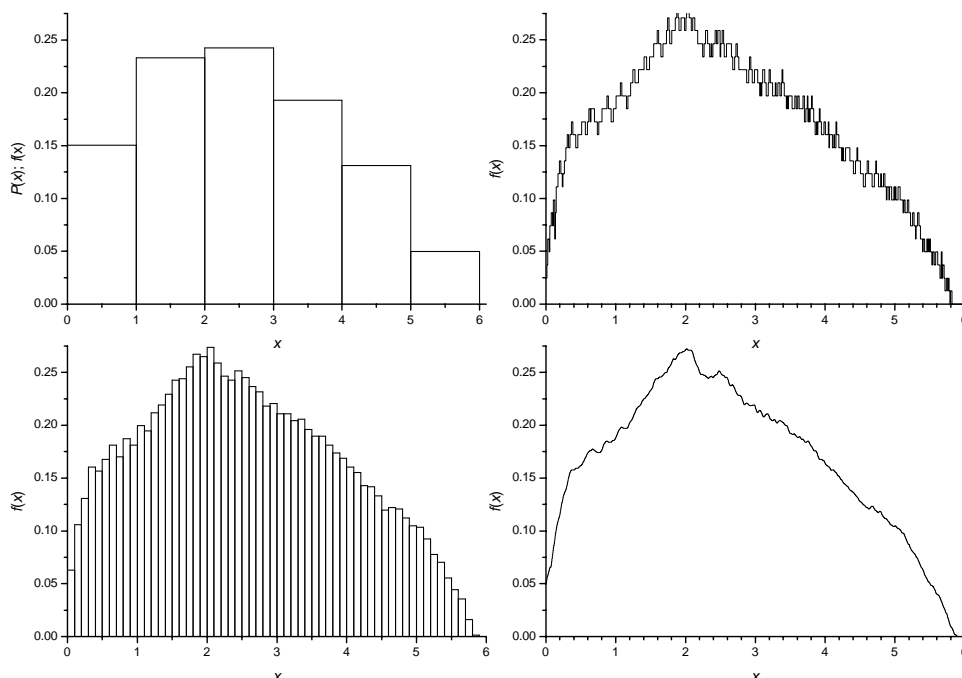
Kontinuirana slučajna varijabla je teorijska reprezentacija kontinuiranih varijabli kao što su duljina, masa, vrijeme,...

Primjer: mjerimo dubinu nekog jezera na 8115 mjesta. Slike prikazuju: 1. raspodjelu dubina zaokruženih na cijele metre; 2. zaokruženih na cijele decimetre. 3. raspodjelu dubina zaokruženih na cijele centimetre; 4. kontinuiranu raspodjelu dubina.

Za svaki interval Δx definiramo funkciju $f(x)$ s pomoću vjerojatnosti

$P(x \leq X \leq x + \Delta x) = f(x)\Delta x$. Na četvrtoj slici intervali su infinitezimalni ($\Delta x \rightarrow dx$) pa imamo $P(x \leq X \leq x + dx) = f(x)dx$

Površina ispod svih grafova jednaka je jedinici!



Funkcija gustoće vjerojatnosti

Def: Neka je X kontinuirana slučajna varijabla. **Raspodjela vjerojatnosti** ili **funkcija gustoće vjerojatnosti** varijable X je funkcija $f(x)$ takva da za bilo koja dva broja $a \leq b$ vrijedi

$$P(a \leq X \leq b) = \int_a^b f(x) dx \quad .$$

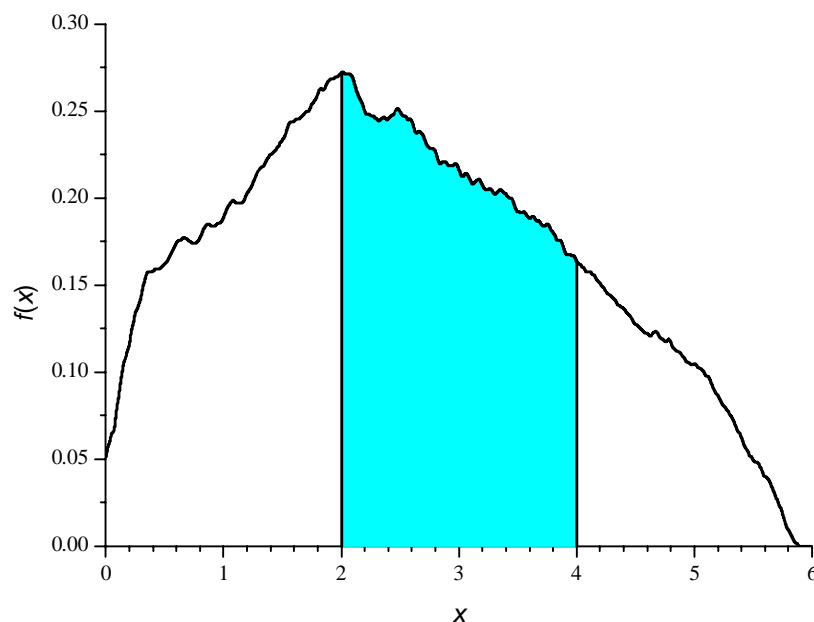
Riječima: Vjerojatnost da X poprimi vrijednost u intervalu $[a, b]$ dana je površinom ispod funkcije gustoće vjerojatnosti u tom intervalu.

Uvjeti na $f(x)$:

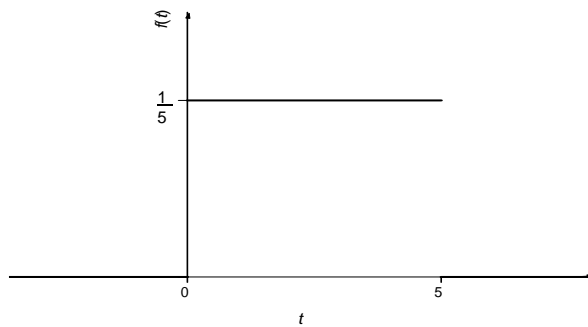
1. $f(x) \geq 0 \quad , \quad \forall x$

2. $\int_{-\infty}^{\infty} f(x) dx = 1$

Primjer: za raspodjelu dubina jezera: $P(2 \leq X \leq 4) = \int_2^4 f(x) dx$



Primjer: Čekanje tramvaja koji vozi svakih 5 minuta.



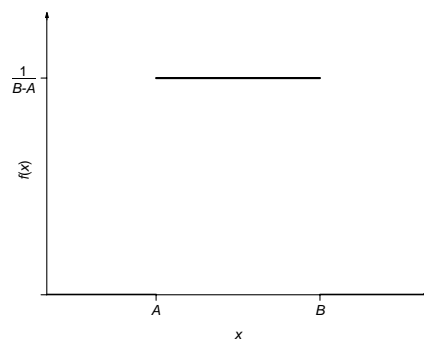
$$P(X \leq 3) = \int_{-\infty}^3 f(x) dx = \frac{3}{5}$$

To je primjer uniformne ili pravokutne raspodjele.

Def: Kontinuirana slučajna varijabla X ima **uniformnu** raspodjelu u intervalu $[A, B]$ ako je funkcija gustoće vjerojatnosti dana s

$$f(x) = \begin{cases} \frac{1}{B-A} & A \leq x \leq B \\ 0 & \text{inace} \end{cases} .$$

Ilustracija pravokutne raspodjele:



Važno: Za razliku od diskretne raspodjele, ovdje vrijedi

$$P(X = c) = 0 \quad , \quad \forall c .$$

To znači:

$$P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b)$$

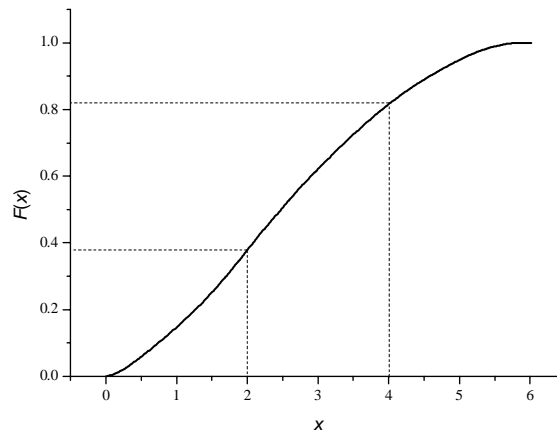
Funkcija raspodjele

Def: Funkcija raspodjele $F(x)$ za kontinuiranu slučajnu varijablu X definirana je:

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(y) dy .$$

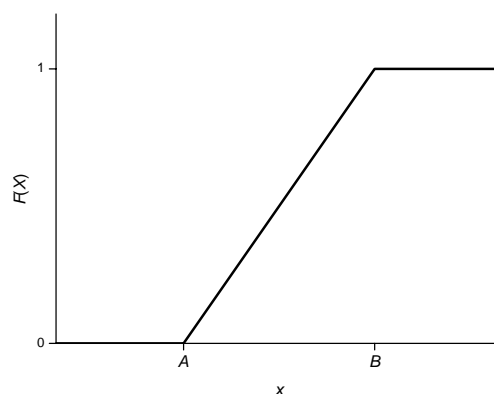
To je površina ispod funkcije gustoće vjerojatnosti.

Primjer: funkcija raspodjele za dubinu jezera



Za pravokutnu raspodjelu vrijedi:

$$F(x) = \begin{cases} 0 & , \quad x < A \\ \frac{x-A}{B-A} & , \quad A \leq x \leq B \\ 1 & , \quad x > B \end{cases}$$



Općenito:

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

Ako poznajemo $F(x)$, onda $f(x)$ lako određujemo:

$$f(x) = \frac{dF(x)}{dx} .$$

Očekivanje kontinuirane slučajne varijable

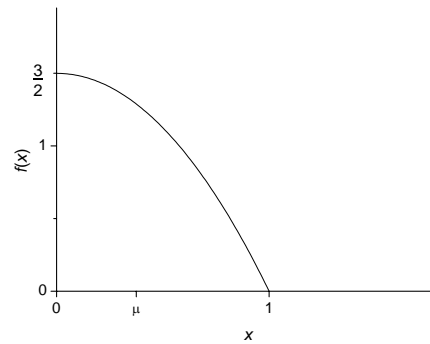
Def: Ako je X kontinuirana slučajna varijabla i $f(x)$ njezina funkcija gustoće vjerojatnosti, onda je njezino **očekivanje**

$$\mu_x = E(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx \quad .$$

Primjer: pravokutna: $E(x) = \frac{A+B}{2}$

Primjer: Neka je

$$f(x) = \begin{cases} \frac{3}{2}(1-x^2) & 0 \leq x \leq 1 \\ 0 & \text{inace} \end{cases}$$



Očekivanje je:

$$E(X) = \int_0^1 xf(x)dx = \frac{3}{2} \int_0^1 (x - x^3) dx = \frac{3}{4} - \frac{3}{8} = \frac{3}{8} \quad .$$

Očekivanje funkcije

Def: Ako je X kontinuirana slučajna varijabla i $f(x)$ njezina funkcija gustoće vjerojatnosti, a $h(X)$ bilo koja funkcija od X , onda je

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

Za linearnu funkciju $h(X) = aX + b$ vrijedi

$$E(aX + b) = aE(X) + b \quad .$$

Varijanca kontinuirane slučajne varijable

Def: Ako je X kontinuirana slučajna varijabla, $f(x)$ njezina funkcija gustoće vjerojatnosti, a μ njezino očekivanje, onda je njezina **varijanca** dana relacijom

$$V(X) = \sigma_x^2 = \int_{-\infty}^{\infty} (x - \mu)^2 \cdot f(x) dx = E\left((X - \mu)^2\right).$$

Standardna devijacija je $\sigma_x = \sqrt{V(X)}$.

Vrijedi:
$$V(X) = E(X^2) - (E(X))^2$$

Varijanca funkcije:

$$V(h(X)) = E(h^2(X)) - (E(h(X)))^2$$

primjer: linearna funkcija $h(X) = aX + b$

$$V(aX + b) = a^2 V(X)$$

Momenti

$$m_r = \int_{-\infty}^{\infty} x^r f(x) dx$$

$$M_r = \int_{-\infty}^{\infty} (x - \mu)^r f(x) dx$$

Funkcija izvodnica

$$g(t) = \int_{-\infty}^{\infty} e^{tx} f(x) dx$$

IV.2.1. NORMALNA (GAUSSOVA) RASPODJELA

Premda postoje mnoge druge raspodjele, normalna raspodjela objašnjava najveći broj statističkih opažanja.

Dva primjera:

- Raspodjela visina odraslih ljudi
- Rezultati mjerenja fizikalnih veličina

Kao objašnjenje Gaussove prirode raspodjele ljudskih visina pomislite na mnoštvo genetskih utjecaja i utjecaja okoliša na nečiju visinu.

Svaki takav utjecaj možemo zamisliti kao jedan čavlić u Galtonovoj dasci koji određuje hoće li netko biti malo višji ili malo niži. (*Galton je bio antropolog*)

Zbroj svih tih utjecaja određuje nečiju konačnu visinu.

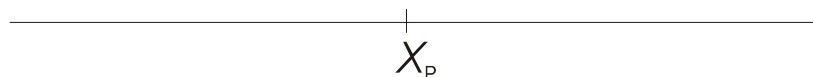
Spušta li se mnoštvo kuglica na Galtonovu dasku, a broj redova (odluka) se povećava, raspodjela kuglica sve će više ličiti Gaussovoj.

<http://www.uni-koeln.de/ew-fak/Mathe/Projekte/VisuPro/galton.html>

Galtonova daska je prejednostavna za taj opis. Neki utjecaji imaju jače djelovanje od drugih, dok neki djeluju udruženo, a ne nezavisno.

Ali pokazuje se da ako raznih čimbenika ima jako mnogo, Gaussova krivulja dobro opisuje raspodjelu.

Razmotrimo mjerenje neke fizikalne veličine. Neka je prava vrijednost te veličine x_p :



Prilikom mjerenja te veličine dolazi do pogrešaka.

Pogreške mogu biti **grube**, **sistematske** ili **slučajne**.

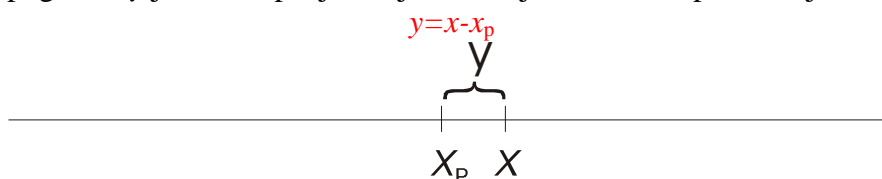
Ako smo pažljivim mjerenjem otklonili grube i sistematske, ostaju nam samo slučajne pogreške. Na njih ne možemo utjecati i one dovode do rasipanja rezultata.

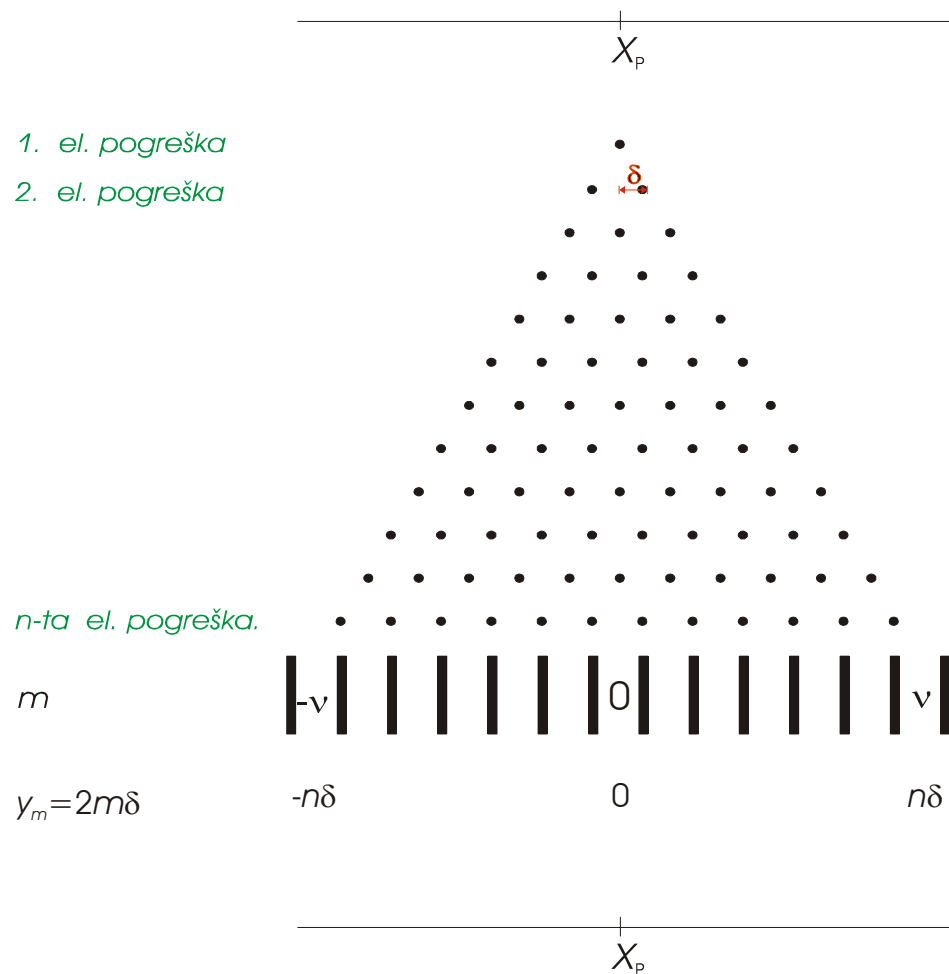
Pretpostavke za **Hagenov izvod Gaussove distribucije**:

- ukupna pogreška sastoji se od niza **elementarnih pogrešaka** koje su sve istog iznosa δ
- jednaka je vjerojatnost da elementarna pogreška bude pozitivna ili negativna:
 $p_+ = p_- = 1/2$
- ukupna pogreška sastoji se od n elementarnih pogrešaka, $n \rightarrow \infty$

Radi jednostavnosti stavimo da je n paran broj ($n = 2\nu$)

Ukupna pogreška y jest odstupanje izmjerene vrijednosti x od prave vrijednosti x_p :





vrijedi: $p_m = \binom{2\nu}{\nu+m} \frac{1}{2^{2\nu}}$

Da bismo prešli na kontinuiranu raspodjelu najprije ćemo odrediti diferencije, a zatim u limesu preći na diferencijale pa integrirati.

$$\Delta y_m = y_{m+1} - y_m = 2\delta$$

$$\Delta p_m = p_{m+1} - p_m = p_m \left[\frac{2\nu - \nu - m}{\nu + m + 1} - 1 \right] \approx p_m \frac{-2m}{\nu} = p_m \frac{-y_m}{\nu\delta}$$

pa konačno dobivamo:

$$\frac{\Delta p_m}{\Delta y_m} = -\frac{p_m y_m}{2\nu\delta^2}$$

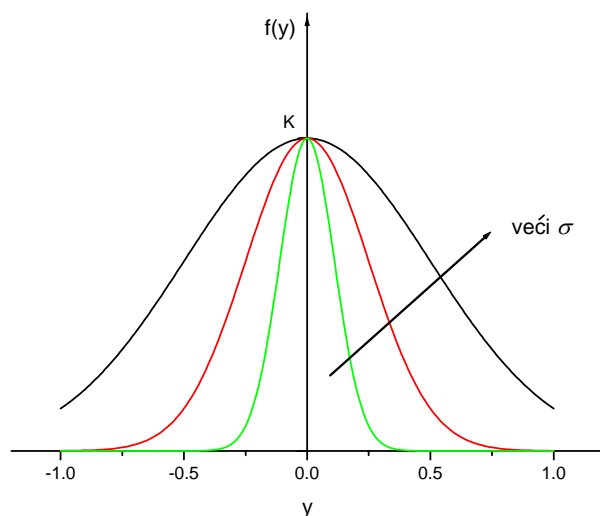
Uvedimo oznaku $\sigma^2 = 2\nu\delta^2$ pa konačno imamo

$$\frac{dp}{dy} = \lim_{\substack{\Delta y \rightarrow 0 \\ n\delta^2 = \text{const}}} \frac{\Delta p}{\Delta y} = -\frac{1}{\sigma^2} py$$

Razdvojimo varijable pa integriramo:

$$\frac{dp}{p} = -\frac{1}{\sigma^2} y dy$$

$$\ln p = -\frac{1}{2\sigma^2} y^2 + C \quad \Rightarrow \quad p = K e^{-\frac{y^2}{2\sigma^2}}$$



Da bi ova funkcija mogla biti funkcija gustoće vjerojatnosti, mora biti normirana.

$$\int_{-\infty}^{\infty} f(y) dy = 1$$

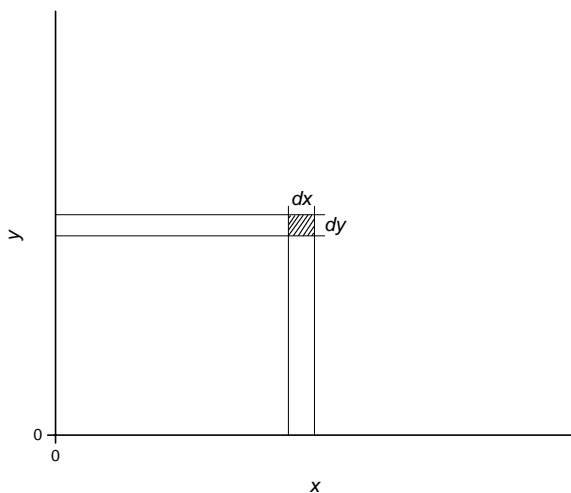
$$\int_{-\infty}^{\infty} K e^{-\frac{y^2}{2\sigma^2}} dy = 2K \int_0^{\infty} e^{-\frac{y^2}{2\sigma^2}} dy = 2\sqrt{2}\sigma K \int_0^{\infty} e^{-u^2} du = 2\sqrt{2}\sigma K \cdot I = 1$$

Izračun integrala $I = \int_0^{\infty} e^{-u^2} du$

Izračunat ćemo najprije kvadrat tog integrala:

$$\begin{aligned} I^2 &= \int_0^{\infty} e^{-x^2} dx \cdot \int_0^{\infty} e^{-y^2} dy = \\ &= \int_0^{\infty} \int_0^{\infty} e^{-(x^2+y^2)} dx dy = \end{aligned}$$

To je dvostruki (površinski) integral u kojem varijabla integracije ide po cijelom prvom kvadrantu. Diferencijalni element površine je $dx dy$:



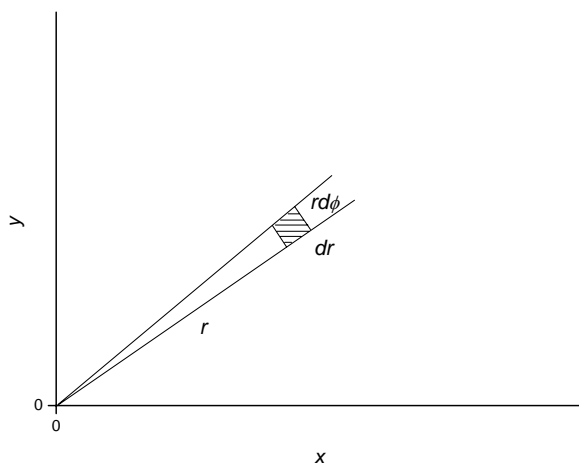
Možemo prijeći na polarne koordinate:

$$x = r \cos \varphi$$

$$y = r \sin \varphi$$

$$x^2 + y^2 = r^2$$

Diferencijalni element površine postaje $rdrd\varphi$:



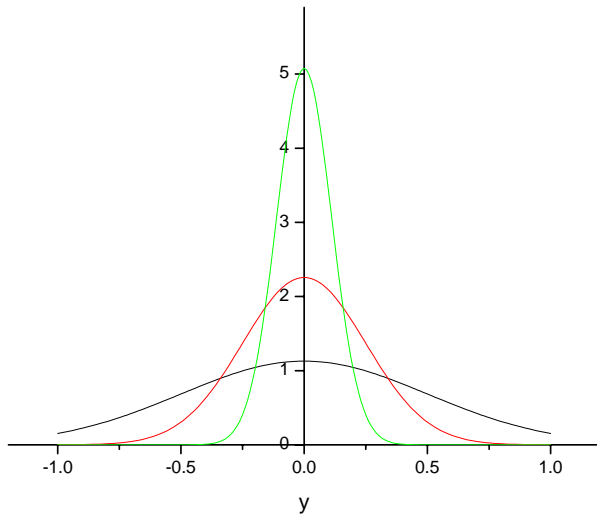
Sada naš integral u polarnim koordinatama izgleda ovako:

$$I^2 = \int_{r=0}^{\infty} \int_{\varphi=0}^{\pi/2} e^{-r^2} r dr d\varphi = \frac{\pi}{2} \int_{r=0}^{\infty} e^{-r^2} r dr = (\text{supstitucija } z = r^2) = \frac{\pi}{4} \int_0^{\infty} e^{-z} dz = \frac{\pi}{4}$$

pa je $I = \sqrt{I^2} = \frac{\sqrt{\pi}}{2}$.

Izračunali smo faktor normiranja $\Rightarrow K = \frac{1}{\sigma\sqrt{2\pi}}$

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



Uvrstimo li $y = x - x_p$, funkcija gustoće vjerojatnosti postaje:

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

Možemo pokazati da je **očekivanje**:

$$\mu = E(X) = x_p. \text{ (pokažite sami)}$$

Možemo pokazati da je **varijanca**:

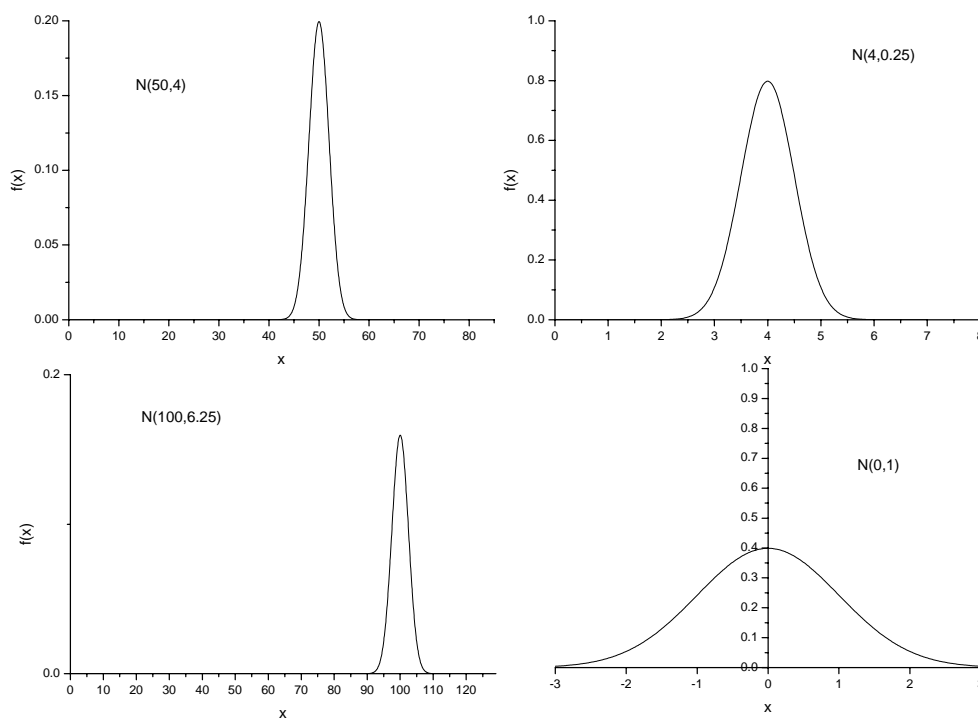
$$V(X) = \sigma^2. \text{ (pokažite sami)}$$

Normalnu raspodjelu obično pišemo

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

Normalnu slučajnu varijablu označavamo:

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



IV.2.1.1. Standardna normalna raspodjela

Ako je $X \sim N(\mu, \sigma^2)$ uvodimo slučajnu varijablu $Z = \frac{X - \mu}{\sigma}$.

Tada je $Z \sim N(0,1)$.

Funkcija gustoće vjerojatnosti standardne normalne raspodjele:

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

Funkcija raspodjele za standardnu normalnu varijablu:

$$\Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy \rightarrow \text{tablice}$$

Upotreba tablica

Primjer 1: vjerojatnost $P(Z > 1,38) = 1 - P(Z < 1,38) = 0,0838$

Primjer 2: Imamo normalnu raspodjelu s $\mu = 50$ i $\sigma = 2$. Zanima nas vjerojatnost da X bude između 46 i 53.

Rješenje:

$$X \sim N(50,4) \quad , \quad P(46 \leq X \leq 53) = ?$$

standardiziramo: $Z = \frac{X - \mu}{\sigma} = \frac{X - 50}{2}$

$\Rightarrow P(-2 \leq Z \leq 1,5) = \Phi(1,5) - \Phi(-2) = 0,9332 - 0,0228 = 0,9104$

Važno svojstvo:

$P(\mu - \sigma \leq X \leq \mu + \sigma) = 0,6826 \approx 68\%$ $P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = 0,9544 \approx 95\%$ $P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = 0,9974 \approx 99,7\%$
--

Gaussova aproksimacija binomne raspodjele

Ako je $X \sim \text{Bin}(n, p)$, za veliki n i ne premali p vrijedi približno $X \sim N(np, npq)$.

Primjer: U 12 bacanja novčića nađi vjerojatnost da će pasti između 4 i 7 glava (uključivo).

a) binomnom:

$$X \sim \text{Bin}(12, \frac{1}{2})$$

$$P(X = 4) = \binom{12}{4} \frac{1}{2^{12}} = 0,121$$

$$P(X = 5) = \binom{12}{5} \frac{1}{2^{12}} = 0,193$$

$$P(X = 6) = \binom{12}{6} \frac{1}{2^{12}} = 0,226$$

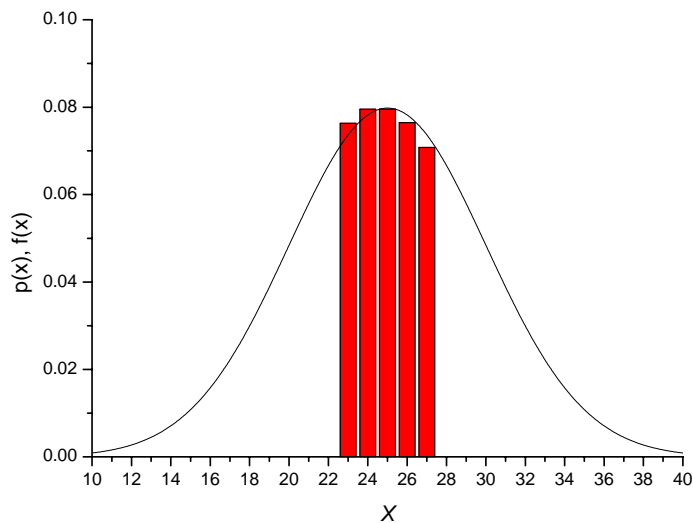
$$P(X = 7) = \binom{12}{7} \frac{1}{2^{12}} = 0,193$$

$$P(4 \leq X \leq 7) = 0,733$$

b) normalnom:

$$X \sim N(6,3)$$

$$P(3,5 \leq X \leq 7,5) = P\left(\frac{3,5 - 6}{\sqrt{3}} \leq Z \leq \frac{7,5 - 6}{\sqrt{3}}\right) = \Phi(0,87) - \Phi(-1,44) = 0,8078 - 0,0749 = 0,733$$



Gaussova aproksimacija Poissonove raspodjele

Ako je $X \sim \text{Po}(\lambda)$, za veliki λ vrijedi približno $X \sim N(\lambda, \lambda)$.

Primjer: Geigerov brojač opaža raspade koji slijede Poissonovu raspodjelu s 25 raspada u sekundi. Nađi vjerojatnost da u jednoj sekundi broj raspada bude $23 \leq X \leq 27$.

Rješenje:

a) Poissonovom:

$$X \sim \text{Po}(25)$$

$$P(X = 23) = e^{-25} \frac{25^{23}}{23!} = 0,0763$$

$$P(X = x + 1) = \frac{25}{x + 1} P(X = x)$$

$$P(X = 24) = \frac{25}{24} P(X = 23) = 0,0795$$

$$P(X = 25) = P(X = 24) = 0,0795$$

$$P(X = 26) = \frac{25}{26} P(X = 25) = 0,0765$$

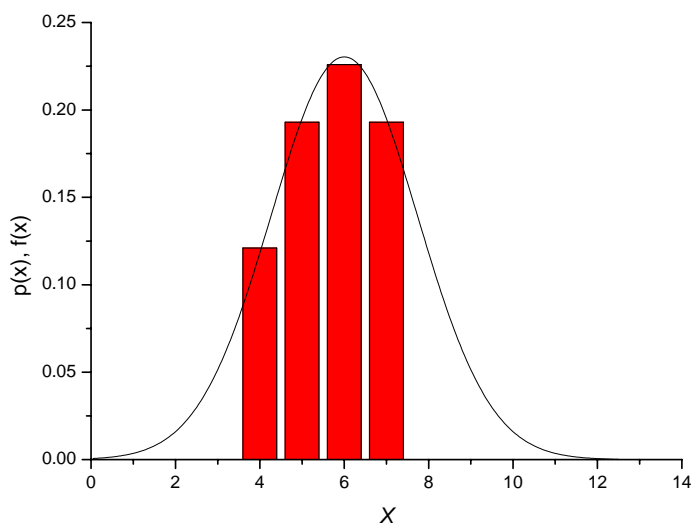
$$P(X = 27) = \frac{25}{27} P(X = 26) = 0,0708$$

$$P(23 \leq X \leq 27) = 0,3826$$

b) normalnom:

$$X \sim N(25, 25)$$

$$P(22,5 \leq X \leq 27,5) = P\left(\frac{22,5 - 25}{5} \leq Z \leq \frac{27,5 - 25}{5}\right) = \Phi(0,5) - \Phi(-0,5) = 0,6915 - 0,3085 = 0,3830$$



Kada možemo aproksimirati?

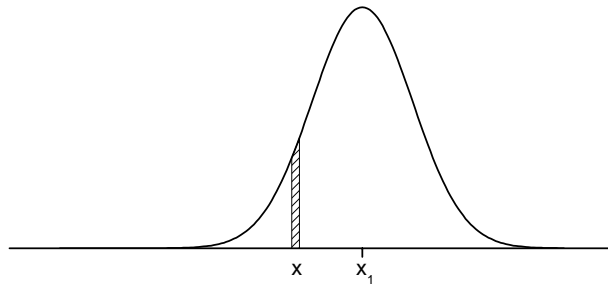
Raspodjela	Ograničenja	Aproksimacija
$X \sim \text{Bin}(n, p)$	n velik (>50) p malen ($<0,1$)	$X \sim \text{Po}(np)$
$X \sim \text{Bin}(n, p)$	$n > 10, p \approx 1/2$ ili $n > 30, p \neq 1/2$	$X \sim N(np, npq)$
$X \sim \text{Po}(\lambda)$	$\lambda > 20$	$X \sim N(\lambda, \lambda)$

IV.2.1.1. Princip najmanjih kvadrata

Mjerimo veličinu X , a njezina prava vrijednost je x_p koju ne poznajemo. Obavimo n mjerenja i rezultati su x_1, x_2, \dots, x_n . Tražimo najvjerojatniju vrijednost za x_p .

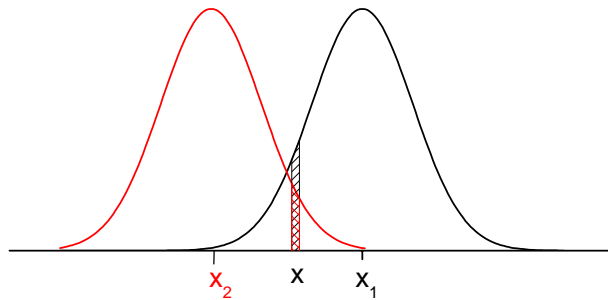
Da smo obavili samo jedno mjerenje, rezultat bi bio x_1 . Vjerojatnost da je x_p u intervalu $(x, x+\Delta x)$ iznosi:

$$\Delta P_1 = P(x \leq x_p \leq x + \Delta x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-x_1)^2}{2\sigma^2}} \cdot \Delta x$$



Da smo obavili dva mjerenja, rezultat bi bili x_1 i x_2 . Vjerojatnost da je x_p u intervalu $(x, x+\Delta x)$ iznosi:

$$\Delta P = \Delta P_1 \cdot \Delta P_2 = P(x \leq x_p \leq x + \Delta x) = \frac{1}{(\sigma\sqrt{2\pi})^2} \cdot e^{-\frac{(x-x_1)^2 + (x-x_2)^2}{2\sigma^2}} \cdot (\Delta x)^2$$



Obavimo n mjerenja i rezultati su x_1, x_2, \dots, x_n . Vjerojatnost da je x_p u intervalu $(x, x+\Delta x)$ iznosi:

$$\Delta P = \Delta P_1 \cdot \Delta P_2 \cdots \Delta P_n = \frac{1}{(\sigma\sqrt{2\pi})^n} \cdot e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x-x_i)^2} \cdot (\Delta x)^n$$

Budući da ne znamo pravu vrijednost x_p , zanima nas najvjerojatnija vrijednost. To je ona vrijednost x_p za koju gornja funkcija (produkt) ima maksimum. Nađimo ga!

Da bi ova funkcija imala maksimum, mora vrijediti

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

To zovemo “**princip najmanjih kvadrata**”.

Nadimo sada x_p^* za koji je zbroj kvadrata odstupanja minimalan:

$$\frac{d}{dx} \left(\sum_{i=1}^n (x_i - x)^2 \right) = 0$$

$$-2 \sum_{i=1}^n (x_i - x) = 0$$

Za najvjerojatniji x_p^* vrijedi

$$\sum_{i=1}^n (x_i - x_p^*) = \sum_{i=1}^n x_i - n x_p^* = 0$$

$$x_p^* = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}$$

Osnove teorije pogrešaka

1. Najvjerojatnija vrijednost na osnovi n mjerenja jest ona pri kojoj je zbroj kvadrata odstupanja mjerenih veličina najmanji.
2. Ako je prava vrijednost x_p neke veličine nepoznata, onda se kao najvjerojatnija vrijednost koju x_p poprima uzima $x_p = \bar{x}$.

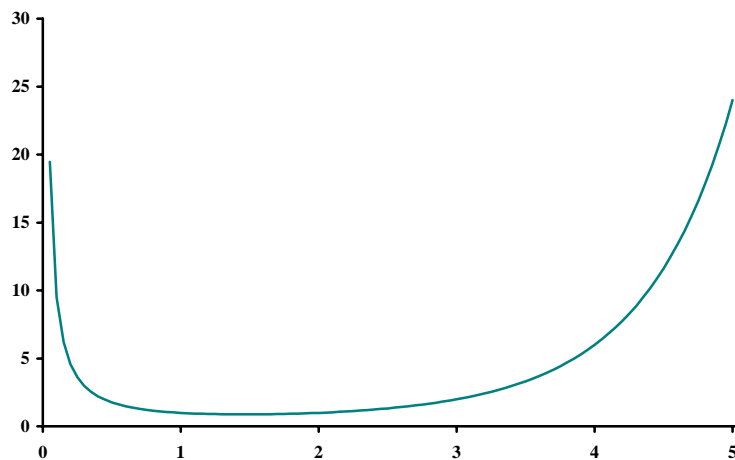
IV.2.2. OBITELJ Γ -RASPODJELA

Prije no što nastavimo kontinuirane raspodjele, napraviti ćemo malu digresiju. Definirati ćemo gama funkciju:

Γ -funkcija

Def:

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx \quad \alpha > 0$$



Tri najvažnija svojstva:

1. $\forall \alpha > 1 \Rightarrow \Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$
2. $\forall n \in \mathbf{N} \Rightarrow \Gamma(n) = (n - 1)!$
3. $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$

Dokazi:

1. Primijenimo pravilo parcijalne integracije $\int u dv = uv - \int v du$. Stavimo $u = x^{\alpha-1}$,

$dv = e^{-x} dx$, $du = (\alpha - 1)x^{\alpha-2}$, $v = -e^{-x}$. Integriramo:

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx = -x^{\alpha-1} e^{-x} \Big|_0^{\infty} + (\alpha - 1) \int_0^{\infty} x^{\alpha-2} e^{-x} dx = 0 + (\alpha - 1)\Gamma(\alpha - 1).$$

2. Izračunajmo prvih nekoliko vrijednosti. Dalje primjenjujemo rekurziju iz 1.

$$\Gamma(1) = \int_0^{\infty} x^0 e^{-x} dx = -e^{-x} \Big|_0^{\infty} = 1 = 0!$$

$$\Gamma(2) = 1 \cdot \Gamma(1) = 1 \cdot 0! = 1!$$

$$\Gamma(3) = 2 \cdot \Gamma(2) = 2 \cdot 1! = 2!$$

$$\Gamma(4) = 3 \cdot \Gamma(3) = 3 \cdot 2! = 3!$$

3. Uz supstituciju $t = \sqrt{x}$, $dt = \frac{1}{2} x^{-1/2} dx$ imamo

$$\Gamma\left(\frac{1}{2}\right) = \int_0^{\infty} x^{-1/2} e^{-x} dx = 2 \int_0^{\infty} e^{-t^2} dt = \sqrt{\pi}.$$

Izvod Stirlingove formule

$$n! = \Gamma(n+1) = \int_0^{\infty} x^n e^{-x} dx$$

uvedemo supstituciju:

$$x = z\sqrt{n} + n$$

$$dx = \sqrt{n} dz$$

računamo:

$$n! = \int_{-\sqrt{n}}^{\infty} (z\sqrt{n} + n)^n e^{-(z\sqrt{n} + n)} \sqrt{n} \cdot dz =$$

$$= \sqrt{n} e^{-n} \int_{-\sqrt{n}}^{\infty} e^{n \ln(z\sqrt{n} + n) - z\sqrt{n}} dz =$$

$$= \sqrt{n} e^{-n} \int_{-\sqrt{n}}^{\infty} e^{n \ln[n(1+z/\sqrt{n})] - z\sqrt{n}} dz =$$

$$= n^n \sqrt{n} e^{-n} \int_{-\sqrt{n}}^{\infty} e^{n \ln(1+z/\sqrt{n}) - z\sqrt{n}} dz =$$

Za veliki n možemo u granice integrala staviti $n \rightarrow \infty$ i razviti logaritam u red:

$$\approx n^n \sqrt{n} e^{-n} \int_{-\infty}^{\infty} e^{n \left(\frac{z}{\sqrt{n}} - \frac{z^2}{2n} + \dots \right) - z\sqrt{n}} dz = 2n^n \sqrt{n} e^{-n} \int_0^{\infty} e^{-z^2/2} dz =$$

napravimo supstituciju $u = z/\sqrt{2}$

$$= 2n^n \sqrt{2n} \cdot e^{-n} \int_0^{\infty} e^{-u^2} du = 2n^n \sqrt{2n} \cdot e^{-n} \frac{\sqrt{\pi}}{2} = \sqrt{2\pi n} \cdot n^n \cdot e^{-n}$$

Vraćamo se na kontinuirane raspodjele slučajne varijable.

Γ raspodjele

Standardna Γ raspodjela

Def:

Kontinuirana slučajna varijabla X ima **standardnu gama raspodjelu** ako je njezina funkcija gustoće vjerojatnosti dana s

$$f(x; \alpha) = \begin{cases} \frac{x^{\alpha-1} e^{-x}}{\Gamma(\alpha)} & , \quad x \geq 0 \\ 0 & , \quad x < 0 \end{cases}$$

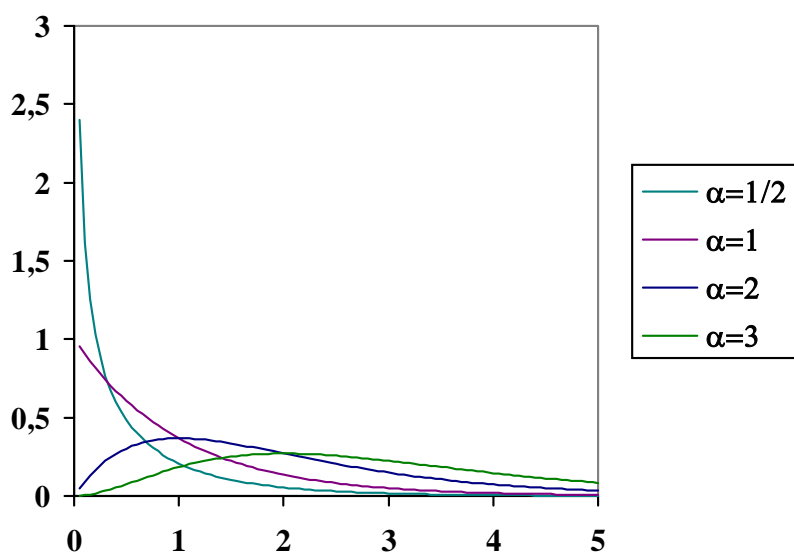
gdje je α parametar raspodjele ($\alpha > 0$).

Lako vidimo da ova raspodjela zadovoljava osnovne uvjete za raspodjelu vjerojatnosti:

1. $f(x; \alpha) \geq 0$, $\forall x$
2. $\int_{-\infty}^{\infty} f(x; \alpha) dx = \frac{1}{\Gamma(\alpha)} \int_0^{\infty} x^{\alpha-1} e^{-x} dx = \frac{\Gamma(\alpha)}{\Gamma(\alpha)} = 1$

Očekivanje i varijanca:

$$E(X) = \alpha \quad V(X) = \alpha$$



Primjena:

Vrijeme reakcije na neki podražaj....

Općenita Γ raspodjela

Def:

Kontinuirana slučajna varijabla X ima **gama raspodjelu** ako je njezina funkcija gustoće vjerojatnosti dana s

$$f(x; \alpha, \beta) = \begin{cases} \frac{x^{\alpha-1} e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)} & , \quad x \geq 0 \\ 0 & , \quad x < 0 \end{cases}$$

gdje su α i $\beta > 0$.

Očekivanje i varijanca:

$$E(X) = \alpha\beta \quad V(X) = \alpha\beta^2 \quad (\text{Tko želi, neka sam izvede!})$$

Općenita gama raspodjela postaje standardna za $\beta = 1$.

poseban slučaj je

Eksponencijalna raspodjela

$$\alpha = 1 \quad , \quad \beta = \frac{1}{\lambda}$$
$$f(x; \lambda) = \begin{cases} \lambda e^{-\lambda x} & , \quad x \geq 0 \\ 0 & , \quad x < 0 \end{cases}$$

Vjerojatnosti u eksponencijalnoj raspodjeli lako se računaju.

Očekivanje i varijanca:

$$E(X) = \frac{1}{\lambda} \quad V(X) = \frac{1}{\lambda^2}$$

Funkcija raspodjele:

$$F(x; \lambda) = \begin{cases} 1 - e^{-\lambda x} & , \quad x \geq 0 \\ 0 & , \quad x < 0 \end{cases}$$

Primjena:

Kad promatramo vrijeme između dva događaja u Poissonovim procesima.

Ako imamo Poissonovu raspodjelu s parametrom λt , onda su vremena između događaja raspodijeljena eksponencijalno s parametrom λ .

χ^2 raspodjela

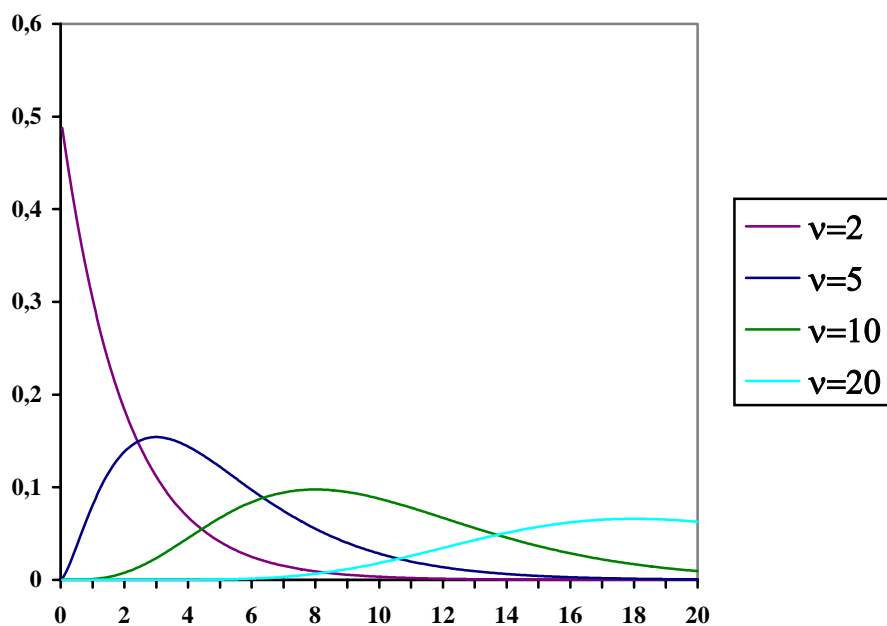
Def:

Neka je $n \in \mathbb{N}$. Kontinuirana slučajna varijabla X ima χ^2 **raspodjelu** s parametrom n ako je njezina funkcija gustoće vjerojatnosti dana s

$$f(x;n) = \begin{cases} \frac{x^{n/2-1} e^{-x/2}}{2^{n/2} \Gamma(n/2)} & , \quad x \geq 0 \\ 0 & , \quad x < 0 \end{cases}$$

Parametar n zove se “broj stupnjeva slobode” varijable X .

χ^2 raspodjela izuzetno je važna za primjenu statističkih testova. Na to ćemo se vratiti kasnije.



DODATAK (nije ispredavano)

Veza normalne i gama raspodjele

Neka imamo $N(\mu, \sigma^2)$. Vjerojatnost da će normalna slučajna varijabla pasti u interval $(x, x+dx)$ jest

$$dP = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \cdot dx$$

Definirajmo novu slučajnu varijablu U koja predstavlja kvadrat odstupanja slučajne varijable X od očekivane vrijednosti:

$$U = \frac{(X - \mu)^2}{\sigma^2}.$$

Nađimo vjerojatnost da U bude u intervalu $(u, u+du)$. Treba razmotriti dva slučaja: $x > \mu$ i $x < \mu$.

1. Kad je $x > \mu$ imamo $x - \mu = \sigma\sqrt{u}$ i $dx = \frac{\sigma}{2\sqrt{u}} du$ pa je vjerojatnost da se X nađe u intervalu $(x, x+dx)$ jednaka

$$dP = \frac{1}{2\sqrt{2\pi}} e^{-u/2} u^{-1/2} \cdot du$$

2. Kad je $x < \mu$ imamo $x - \mu = -\sigma\sqrt{u}$ i $dx = \frac{-\sigma}{2\sqrt{u}} du$ pa je vjerojatnost da se X nađe u intervalu $(x-dx, x)$ jednaka

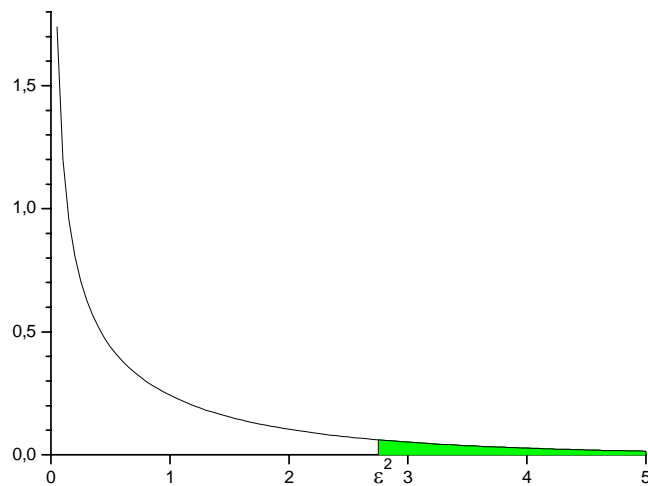
$$dP = \frac{1}{2\sqrt{2\pi}} e^{-u/2} u^{-1/2} \cdot du$$

Ukupna vjerojatnost za $(u, u+du)$ je dakle dvostruka (zbroy dvaju slučajeva) pa imamo

$$dP = \frac{1}{\sqrt{2\pi}} e^{-u/2} u^{-1/2} \cdot du = \frac{1}{2^{1/2} \Gamma(1/2)} u^{1/2-1} e^{-u/2} \cdot du$$

Zaključak: Ako je X normalna slučajna varijabla s očekivanjem μ i standardnom devijacijom σ , onda slučajna varijabla $U = \frac{(X - \mu)^2}{\sigma^2}$ ima gama raspodjelu s parametrima $\alpha = 1/2$ i $\beta = 2$.

Značenje ove veze: Vjerojatnost da veličina $\left| \frac{x - \mu}{\sigma} \right|$ bude veća od nekog broja ε određena je površinom ispod repa gama raspodjele s parametrima $\alpha = 1/2$ i $\beta = 2$ koji se nalazi iza ε^2 . (osjenčani dio na slici)



Neka imamo n nezavisnih normalnih slučajnih varijabli koje sve imaju isti μ i različite σ_i , tj. $X_i \sim N(\mu, \sigma_i^2)$. Definirajmo veličinu

$$\chi^2 = \sum_{i=1}^n \frac{(X_i - \mu)^2}{\sigma_i^2}.$$

Može se pokazati da je to slučajna varijabla raspodijeljena prema χ^2 raspodjeli s n stupnjeva slobode.

IV.3. DVODIMENZIONALNE RASPODJELE

U mnoštvu eksperimentalnih slučajeva, istraživača zanima više od jedne slučajne varijable. Za početak, mi ćemo razmotriti raspodjele vjerojatnosti kad promatramo dvije slučajne varijable definirane na prostoru elementarnih događaja nekog pokusa.

Def: Neka su X i Y dvije diskretne slučajne varijable definirane na prostoru elementarnih događaja nekog eksperimenta. **Združena raspodjela vjerojatnosti** $p(x,y)$ definirana je za svaki par točaka (x,y) kao vjerojatnost da istodobno varijabla X poprimi vrijednost x i varijabla Y poprimi vrijednost y :

$$p(x,y)=P(X=x \text{ i } Y=y)$$

primjer:

Na PMF-u studente razvrstamo prema područjima studiranja (varijabla X) i prema spolu (varijabla Y). Združenu raspodjelu vjerojatnosti za dva obilježja (područje i spol) možemo prikazati tablicom:

	M	Ž	
Matematika	0,151	0,105	
Fizika	0,125	0,079	
Kemija	0,069	0,074	
Biologija	0,062	0,153	
Geo-znanosti	0,124	0,058	

Def: **Rubne (marginalne) raspodjele vjerojatnosti** za varijable X i Y označavamo $p_X(x)$ i $p_Y(y)$, a dane su izrazima:

$$p_X(x) = \sum_{y \in D_Y} p(x, y) \quad i \quad p_Y(y) = \sum_{x \in D_X} p(x, y)$$

primjer:

U gornjem primjeru je, npr., rubna vjerojatnost za matematiku $p_X(\text{mat})=0,256$, a rubna vjerojatnost za muški spol $p_Y(\text{M})=0,531$.

Analogno definiramo združene vjerojatnosti za kontinuirane raspodjele:

Def: Neka su X i Y dvije kontinuirane slučajne varijable. Funkcija $f(x,y)$ je njihova **združena funkcija gustoće vjerojatnosti** ako za bilo koji dvodimenzionalni skup $A \subseteq \mathbb{R} \times \mathbb{R}$ vrijedi

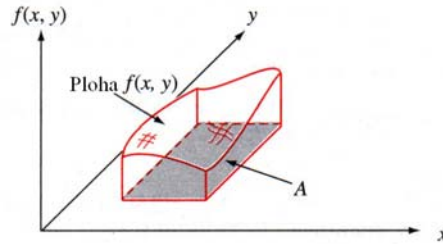
$$P((X,Y) \in A) = \iint_A f(x, y) dx dy$$

Posebno, ako je A dvodimenzionalni pravokutnik $A = \{(x, y) : a \leq x \leq b, c \leq y \leq d\}$, tada je

$$P((X,Y) \in A) = P(a \leq x \leq b, c \leq y \leq d) = \int_a^b \int_c^d f(x, y) dx dy$$

Takva funkcija mora zadovoljavati uvjete: $f(x, y) \geq 0$; $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1$

Zamislamo li plohu u trodimenzionalnom koordinatnom sustavu čija visina iznad točke $(x,y) \in A$ iznosi $f(x,y)$, tada je vjerojatnost $P[(x,y) \in A]$ određena volumenom ispod te plohe.



primjer:

Neka je združena funkcija gustoće vjerojatnosti za kontinuirane varijable X i Y dana s

$$f(x, y) = \begin{cases} \frac{6}{5}(x + y^2) & x \in [0, 1], y \in [0, 1] \\ 0 & \text{inace} \end{cases}$$

Provjerite je li to ispravna funkcija gustoće vjerojatnosti! Izračunajte vjerojatnost da je $X \leq \frac{1}{4}$ i $Y \leq \frac{1}{4}$!

Def: Rubne (marginalne) funkcije gustoće vjerojatnosti za varijable X i Y označavamo $f_X(x)$ i $f_Y(y)$, a dane su izrazima:

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy \quad i \quad f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx$$

primjer: U gornjem primjeru izračunajte vjerojatnost da je $Y \leq \frac{1}{4}$ bez obzira na X !

Nezavisne slučajne varijable

Def: Slučajne varijable X i Y su **nezavisne** ako za svaki par vrijednosti x i y vrijedi

$$p(x, y) = p_X(x) \cdot p_Y(y) \quad (\text{za diskretne varijable})$$

ili

$$f(x, y) = f_X(x) \cdot f_Y(y) \quad (\text{za kontinuirane varijable}).$$

Ako ti uvjeti nisu ispunjeni za sve (x, y) , kažemo da su X i Y **zavisne**.

Primjer: U primjeru studenata na PMF-u provjerite nezavisnost!

Uvjetne raspodjele

Def: Neka su X i Y dvije kontinuirane slučajne varijable sa združenom funkcijom gustoće vjerojatnosti $f(x, y)$ i rubnom funkcijom gustoće vjerojatnosti za varijablu X danom s $f_X(x)$. Tada za bilo koju vrijednost varijable X čija rubna vjerojatnost je različita od nule definiramo **uvjetnu funkciju gustoće vjerojatnosti za Y ako je poznat $X=x$:**

$$f_{Y|X}(y|x) = \frac{f(x, y)}{f_X(x)}.$$

Analogno definiramo **uvjetnu raspodjelu vjerojatnosti** za diskretne varijable.

Primjer:

1. Ako znamo da je student PMF-a muškog spola, odredi uvjetnu raspodjelu vjerojatnosti smjerova!
2. Ako znamo da u primjeru kontinuirane raspodjele vrijedi $X=0,8$, odredi uvjetnu raspodjelu vjerojatnosti za Y !

Sva razmatranja mogu se poopćiti i na promatranje višedimenzionalnih raspodjela.

Momenti dvodimenzionalne raspodjele

Očekivanja

Def:

$$\begin{aligned}\mu_X &= \sum_{x \in D_X} x \cdot p_X(x) = \sum_{x \in D_X} \sum_{y \in D_Y} x \cdot p(x, y) \quad \text{ili} \\ \mu_X &= \int_{-\infty}^{\infty} x \cdot f_X(x) dx = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x \cdot f(x, y) dx dy\end{aligned}$$

Def:

$$\begin{aligned}\mu_Y &= \sum_{y \in D_Y} y \cdot p_Y(y) = \sum_{x \in D_X} \sum_{y \in D_Y} y \cdot p(x, y) \quad \text{ili} \\ \mu_Y &= \int_{-\infty}^{\infty} y \cdot f_Y(y) dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y \cdot f(x, y) dx dy\end{aligned}$$

Momenti

Pomoćni momenti dvodimenzionalne raspodjele:

$$\begin{aligned}m_{rs} &= \sum_{x \in D_X} \sum_{y \in D_Y} x^r y^s p(x, y) \quad \text{ili} \\ m_{rs} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^r y^s f(x, y) dx dy\end{aligned}$$

Vidimo da je $m_{10} = \mu_X$ i $m_{01} = \mu_Y$

Središnji momenti dvodimenzionalne raspodjele:

$$\begin{aligned}M_{rs} &= \sum_{x \in D_X} \sum_{y \in D_Y} (x - \mu_X)^r (y - \mu_Y)^s p(x, y) \quad \text{ili} \\ M_{rs} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)^r (y - \mu_Y)^s f(x, y) dx dy\end{aligned}$$

Vidimo da je

$$\begin{aligned}M_{20} &= V(X) = \sigma_X^2 \\ M_{02} &= V(Y) = \sigma_Y^2\end{aligned}$$

Kovarijanca

$$\text{Cov}(X, Y) = \sigma_{XY} = M_{11}$$

Primjer:

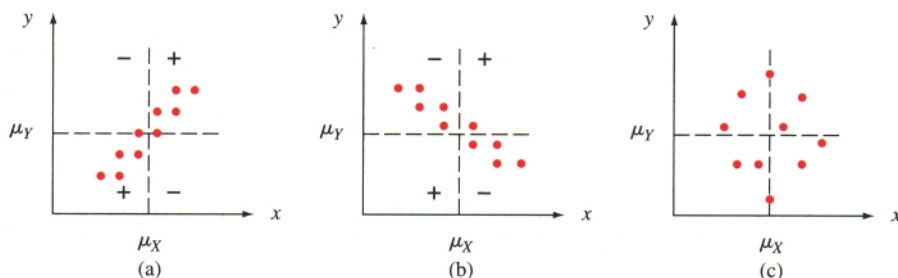


Figure 5.4 $p(x, y) = 1/10$ for each of ten pairs corresponding to indicated points; (a) positive covariance; (b) negative covariance; (c) covariance near zero

Skraćena formula za računanje:

$$\text{Cov}(X, Y) = m_{11} - m_{10} \cdot m_{01} = E(X \cdot Y) - E(X)E(Y)$$

Korelacija

Kovarijanca je veličina koja nam govori o povezanosti (zavisnosti) varijabli X i Y . Međutim, iznos kovarijance ovisi o jedinicama u kojima mjerimo. Da bismo mogli nešto reći stupnju zavisnosti, moramo tu povezanost izraziti bezdimenzionalnim veličinama. Stoga uvodimo pojam korelacije.

Def:

Koeficijent korelacije varijabli X i Y definiran je kao

$$\text{Corr}(X, Y) = \rho_{X,Y} = \rho = \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y}$$

Svojstva:

1. $\text{Corr}(aX + b, cY + d) = \text{Corr}(X, Y)$
2. $-1 \leq \rho \leq 1$
3. Ako su X i Y nezavisne varijable, onda je $\rho=0$. (Obrat ne vrijedi)
4. Ako je $Y=aX+b$ onda i samo onda je $\rho=1$ ili $\rho=-1$.

Korelacija koja iščezava ne isključuje povezanost veličina, nego samo linearnu povezanost.

Korelacija blizu 1 ne znači nužno da su porast veličine X uzrokuje porast veličine Y , nego samo da postoji veza među njima (primjer: karijes i rječnik djece).