Statistika i osnovna mjerenja

Binomna i Poissonova raspodjela

M. Makek

2017/2018

Raspodjele vjerojatnosti

Želimo matematički i statistički opisati raspodjele varijabli koje se javljaju u prirodi u raznim područjima fizike i izvan nje

- Diskretne raspodjele: Bernoullijeva, Binomna, Poissonova, Hipergeometrijska, idr.

- Kontinuirane raspodjele: Gaussova, Gama, idr.

Binomna raspodjela

Primjeri:

- Bacanje novčića
  - Dva moguća ishoda P i G
  - Za pošten novčić p(P) = p(G) = 0.5
- Galtonova daska: http://www.geogebra.org/m/49479
  - Kuglica pada i na svakom koraku može skrenuti lijevo ili desno
  - p(L) = p(D) = 0.5

Što je zajedničko ovim pokusima?

→ Imaju dva moguća ishoda

- ➢ P ili G
- ➢ L ili D
- ➢ Uspjeh ili neuspjeh

Binomna raspodjela

Binomni pokus:

1. Sastoji se od ***n*** (Bernoullijevih) pokušaja, gdje je ***n*** unaprijed zadan
2. Pokušaji imaju dva moguća ishoda: uspjeh (***A***) ili neuspjeh ($\bar{A}$)
3. Vjerojatnost uspjeha (***p***) i neuspjeha (***q=1-p***) je jednaka za sve pokušaje
4. Pokušaji su nezavisni (ishod jednog ne utječe na drugi)

Binomna slučajna varijabla:

Za binomni pokus sa ***n*** pokušaja definiramo slučajnu varijablu:

X = broj uspjeha u ***n*** pokušaja, **X~Bin(n,p)**

$X = 0 \ldots n$

Binomna raspodjela

- Def. Binomna raspodjela
  - Za binomni eksperiment i binomnu slučajnu varijablu ***X~Bin(n,p)***, njenu raspodjelu vjerojatnosti nazivamo binomnom raspodjelom i označavamo: ***b(x; n,p)***
- Uzmimo da niz od n Bernoullijevih eksperimenata završi:

  $AA\bar{A}A\bar{A}A\ldots\ \bar{A}A$

  - A se pojavljuje x puta, $\bar{A}$ n-x puta
  - Koja je vjerojatnost jednog takvog ishoda? → $p^x q^{n-x}$
  - Koliko ima takvih istih ishoda (koji su svi jednako vjerojatni)? → $\binom{n}{x}$
  - Vjerojatnost da se pojavi bilo koji ishod koji ima **x** uspjeha je:

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

za x=0,1,2,…n

b(x; n, p) = 0 za sve ostale x

→ vjerojatnost da u seriji od ***n*** pokusa događaj ***A*** nastupi ***x*** puta

Binomna raspodjela

- <u>Primjer</u>: Galtonova daska

$$b(x;4{,}0.5)=\binom{4}{x}0.5^{x}0.5^{4-x}$$

$$b(x;4{,}0.5)=\binom{4}{x}\frac{1}{16}$$

slučajna varijabla x	putevi	vjerojatnost
0	LLLL	q^4
1	LLLD LLDL LDLL DLLL	$p \cdot q^3$
2	LLDD LDLD LDDL DLLD DLDL DDLL	$p^2 \cdot q^2$
3	LDDD DLDD DDLD DDDL	$p^3 \cdot q$
4	DDDD	p^4

Binomna raspodjela

Rekurzivna formula

- Pomaže u računu, posebice kod velikih faktorijela
- Koja je vjerojatnost da varijabla poprimi vrijednost x-1?

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

- Iz omjera **b(x)/b(x-1)** dobivamo rekurzivnu formulu:

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

→ Vjerojatnost za bilo koji x možemo dobiti rekurzijom iz:

$$b(0;n,p) = q^n$$

Binomna raspodjela

Primjer korištenja rekurzije: za koju vrijednost x je b(x) maksimalan?

- Koristimo rekurzivnu formulu

  Mora biti zadovoljeno:

$$b(x-1;n,p) \leq b(x;n,p) \geq b(x+1;n,p)$$

Nakon raspisivanja (domaća zadaća) po rekurzivnoj formuli slijedi:

$$\boxed{np - q \leq x_{max} \leq np + p}$$

Binomna raspodjela

- Vrijedi binomni poučak:

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

Funkcija raspodjele diskretne slučajne varijable sa binomnom raspodjelom:

$$F(x) = \begin{cases} \sum_{k=0}^{x} \binom{n}{k} p^k q^{n-k}, & 0 \leq x < n \\ 1, & x = n \end{cases}$$

Binomna raspodjela

Očekivanje diskretne slučajne varijable sa binomnom raspodjelom je: $\boxed{E(X) = np}$

- To možemo pokazati na sljedeći način:
  - Definiramo funkciju $g(t) = (q + pt)^n$
  - Razvoj po binomnom poučku:

$$(q + pt)^n = \sum_{x=0}^{n} \binom{n}{x} p^x q^{n-x} t^x = \sum_{x=0}^{n} b(x) t^x$$

  - Deriviranjem po t dobivamo:

$$np(q + pt)^{n-1} = \sum_{x=0}^{n} x\, b(x) t^{x-1}$$

  - Za t=1:

$$np(q + p)^{n-1} = \sum_{x=0}^{n} x\, b(x)$$

  - Što po definiciji daje:

$$np = E(X)$$

Funkcija izvodnica

- Općenito za diskretnu slučajnu varijablu definiramo funkciju izvodnicu:

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

- Iz definicije slijedi:

$$g'(t=0) = \sum_{x \in D} x \cdot p(x) = m_1$$

$$g''(t=0) = \sum_{x \in D} x^2 \cdot p(x) = m_2$$

$$g^{(r)}(t=0) = \sum_{x \in D} x^r \cdot p(x) = m_r$$

→ pomoću funkcije izvodnice mogu se dobiti svi pomoćni, a posredno i centralni momenti slučajne varijable

Binomna raspodjela

- Drugi način da odredimo očekivanje (a i varijancu) je pomoću funkcije izvodnice
- Funkcija izvodnica za binomnu raspodjelu:

$$g_X(t) = E(e^{tX}) = \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x q^{n-x}$$

$$= \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x q^{n-x}$$

- Prema binomnom poučku slijedi:

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

Binomna raspodjela

- Prva derivacija funkcije izvodnice:

$$g'_X(t) = npe^t(pe^t + q)^{n-1}$$

$$g'_X(0) = npe^0(pe^0 + q)^{n-1} = np \qquad (= \boldsymbol{m_1})$$

$$\Rightarrow \boxed{E(X) = np}$$

- Druga derivacija funkcije izvodnice:

$$g''_X(t) = npe^t(pe^t + q)^{n-1} + n(n-1)(pe^t)^2(pe^t + q)^{n-2}$$

$$g''_X(0) = np + n(n-1)p^2 \qquad (= \boldsymbol{m_2})$$

Binomna raspodjela

- Varijanca je jednaka centralnom momentu drugog reda
- Otprije imamo: $M_2 = m_2 - m_1^2$
- Slijedi Varijanca diskretne slučajne varijable sa binomnom raspodjelom:

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

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

Treba još primijetiti da se za n=1 binomna raspodjela svodi na Bernoullijevu, što je i za očekivati budući da ona opisuje Bernoullijev pokus koji se ponavlja n puta!

Binomna raspodjela

Primjer: kontrola uzoraka

- Neki uređaj izrađuje 8% neispravnih proizvoda. Proizvodi se pakiraju u kutije od 50 komada.

  Koja je vjerojatnost da se u kutiji nađe 2 defektna komada?

  - n=50, p=0.08
  - $b(2) = \binom{50}{2} p^2 q^{50-2}$

    → **b(2) = 0.14**

  Koji je najvjerojatniji broj defektnih komada u kutiji?

  - n=50, p=0.08
  - np-q <= x_M <= np+p → 3.08 < = x_M <= 4.08

    → **x_M = 4**

Poissonova raspodjela

- Uzmimo da je vjerojatnost uspjeha pojedinog Bernoullijevog pokušaja u binomnom pokusu jako mala, te taj pokus ponavljamo jako puno puta (**n >>, p <<**)

  pri tome očekivana vrijednost konstantna **E(x)=np= const.**

- Primjer: raspad radioaktivnog uzorka:
  - Vjerojatnost pojedinog raspada je iznimno mala
  - Uzorak sadrži jako puno izotopa koji se nezavisno mogu raspasti

  → Srednji broj raspada u nekom vremenskom intervalu je konstantan.

Poissonova raspodjela

- Kao granični slučaj binomne:

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

$$= \frac{n(n-1)\dots(n-x+1)}{x!} p^x q^{n-x}$$

$$= \frac{n(n-1)(n-x+1)}{x!} \frac{\lambda^x}{n^x} \left(1-\frac{\lambda}{n}\right)^{n-x}$$

$$= \frac{\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\dots\left(1-\frac{x-1}{n}\right)}{x!} \lambda^x \left(1-\frac{\lambda}{n}\right)^n \left(1-\frac{\lambda}{n}\right)^{-x}$$

$$p(x) = \lim_{n\to\infty} b(x) = \frac{\lambda^x}{x!} \lim_{n\to\infty} \left(1-\frac{\lambda}{n}\right)^n$$

$$\Rightarrow \boxed{p(x) = \frac{\lambda^x}{x!} e^{-\lambda}}$$

Poissonova raspodjela

- Očekivanje je prema definiciji:

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

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

$$\Rightarrow \boxed{E(X) = \lambda}$$

- Prema definiciji varijance također izlazi da je:

$\boxed{V(X) = \lambda}$ odn. $\boxed{\sigma_X = \sqrt{\lambda}}$

- Normiranost raspodjele slijedi iz definicije:

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

Poissonova raspodjela

- 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:

$$g'(t) = \lambda e^t e^{\lambda(e^t - 1)} \rightarrow \ m_1 = \lambda$$

$$g''(t) = \lambda e^t (1 + \lambda e^t) e^{\lambda(e^t - 1)} \rightarrow m_2 = \lambda^2 + \lambda$$

→ ***E(X)*** i ***V(X)*** možemo elegantno dobiti preko funkcije izvodnice

Poissonova raspodjela

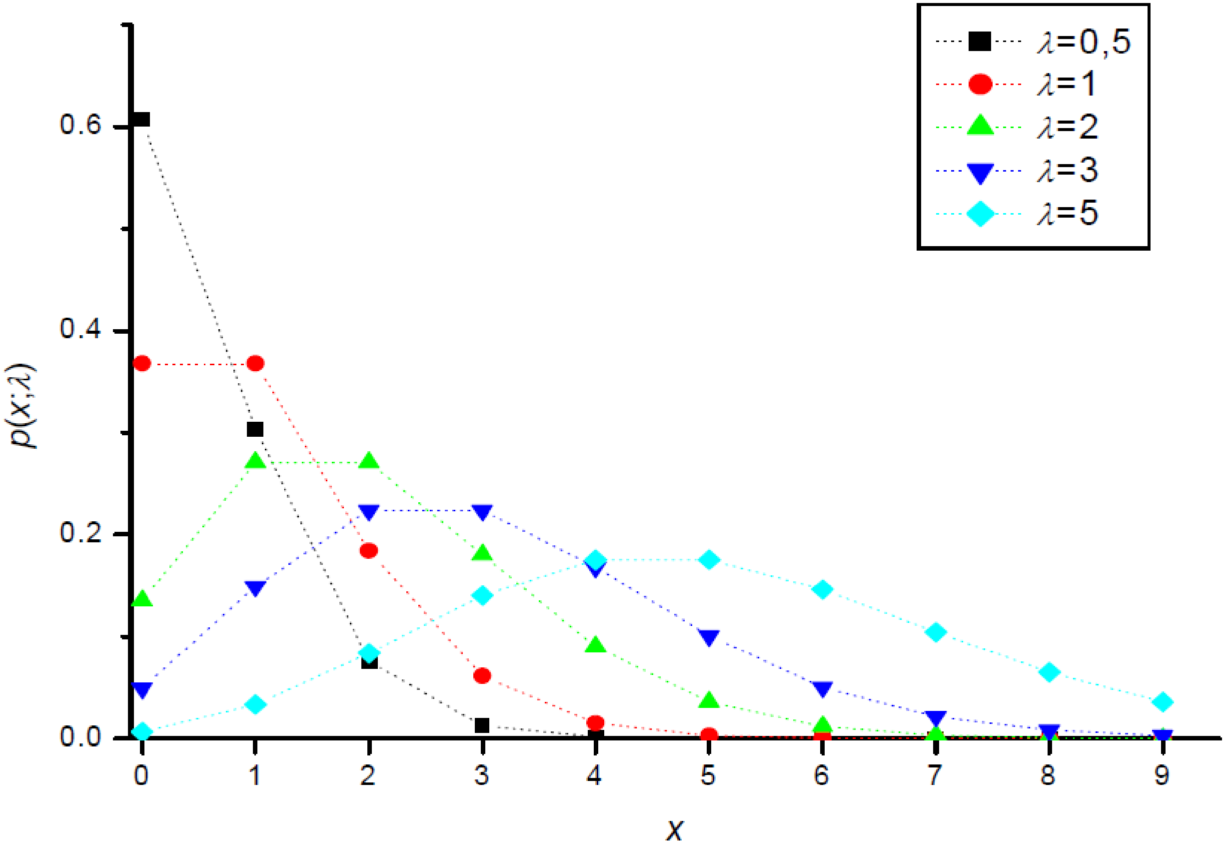
- Rekurzija
  - Iz omjera p(x)/p(x-1) slijedi:

$$\boxed{p(x) = \frac{\lambda}{x} p(x-1)}$$

- Najvjerojatnija vrijednost
  - Iz uvjeta: $p(x_M - 1) \le p(x_M) \ge p(x_M + 1)$
  - Slijedi: $\boxed{\lambda - 1 \le x_M \le \lambda}$

Poissonova raspodjela

- P.r. je potpuno određena parametrom λ



Suma dvije nezavisne Poissonove varijable

- Za dvije slučajne varijable ***X*** i ***Y*** s Poissonovim raspodjelama ***p(x,m)*** i ***p(y,n)*** vrijedi da je raspodjela slučajne varijable ***Z=X+Y*** jednaka ***p(z, m+n)***
- Dokaz:

$$P(Z=z)=\sum_{k=0}^{z}P(X=k)P(Y=z-k)$$

$$=\sum_{k=0}^{z}e^{-m}\frac{m^k}{k!}\,e^{-n}\frac{n^{z-k}}{(z-k)!}$$

$$=\frac{e^{-(m+n)}}{z!}\sum_{k=0}^{z}\binom{z}{k}m^k n^{z-k}$$

$$=\frac{e^{-(m+n)}}{z!}(m+n)^z=p(z,m+n)$$

Poissonova raspodjela
(primjene)

<u>Primjer</u>: **Poissonova raspodjela kao granični slučaj binomne**

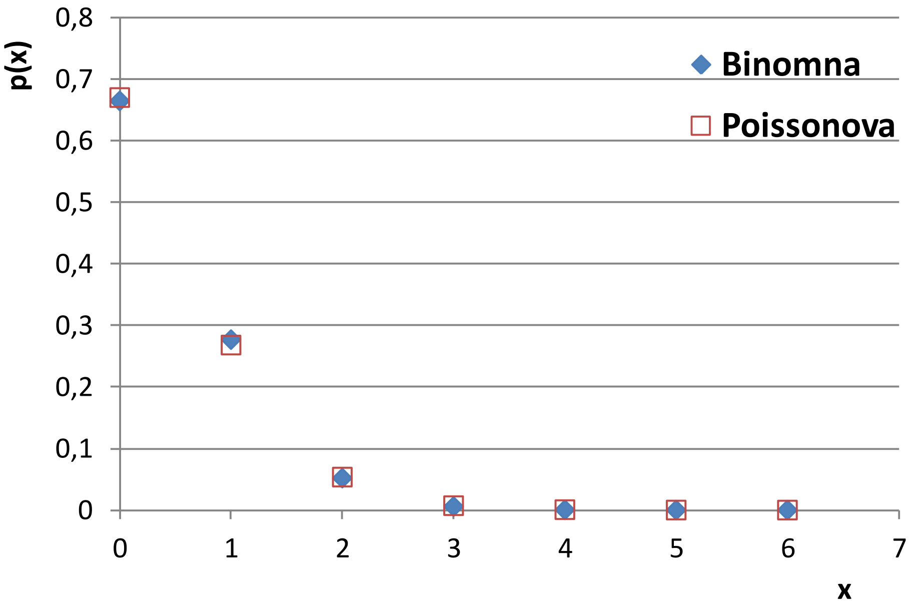
- Serija proizvoda sadrži 4% loših
- Kolika je vjerojatnost da u seriji od 10 proizvoda nadjemo najviše jedan loš proizvod?
  - N=10
  - P=0.04

    b(0;10,0.04) + b(1;10,0.04) = 0.9418

  - Np = 0.4

    p(0;0.4) + p(1;0.4) = 0.9384

→ Poissonova r. je dobra aproksimacija binomne!



Poissonova raspodjela
(primjene)

Primjer: **Poissonova raspodjela za proučavanje rijetkih događaja**

- Konstanta raspada nekog izotopa je λ=7,3 10^{-10} s^{-1}
- Uzorak sadrži 4,8 10^{22} radioaktivnih atoma
- Treba odrediti raspodjelu vjerojatnosti ***k*** raspada u 1 s.
- Zakon radioaktivnog raspada daje broj atoma u vremenu:

$$N(t) = N_0 e^{-t/\tau}$$

- I aktivnost uzorka: $a = \left|\frac{dN}{dt}\right| = \frac{N}{\tau}$
- Slijedi da je vjerojatnost k raspada u intevalu Δt:

$$P_k\left(\Delta t\right) = b(k; N, \frac{N\Delta t}{\tau}\frac{1}{N}) = p(k; a\Delta t)$$

- Vjerojatnost događaja proporcionalna je ***veličini intervala***

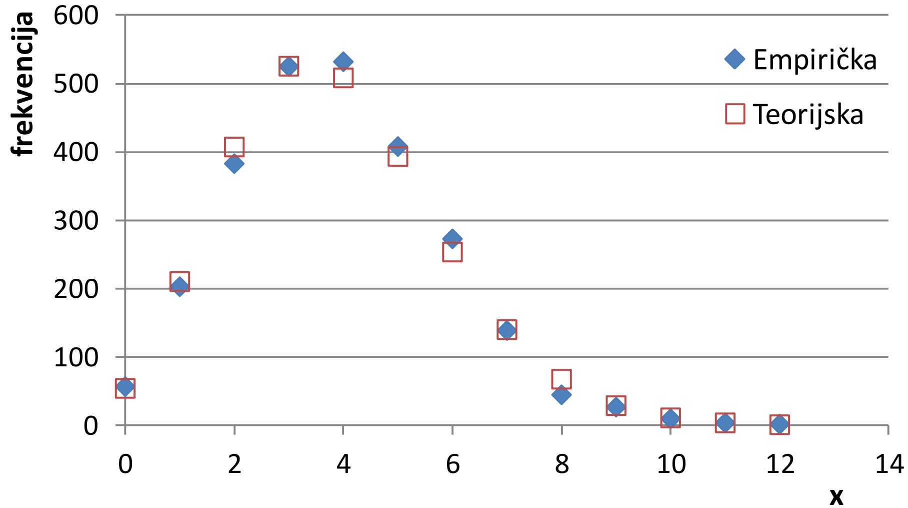
→ Takve procese nazivamo Poissonovim procesima!

Prilagodba empiričkim podacima

- Primjer Poissonova raspodjela
- Mjeren je broj beta raspada ***x*** iz nekog uzorka u intervalu od 10 sekundi. Broj slučajeva u kojima je emitirano točno ***x*** čestica je označen kao f_x
  → smatramo da bi broj opaženih čestica trebao slijediti Poiss. raspodjelu.

x	f_x	p(x)	f_{tx}
0	57	0,0209	54,4
1	203	0,0807	210,5
2	383	0,1562	407,4
3	525	0,2015	525,5
4	532	0,1949	508,4
5	408	0,1509	393,5
6	273	0,0973	253,8
7	139	0,0538	140,3
8	45	0,0260	67,9
9	27	0,0112	29,2
10	10	0,0043	11,3
11	4	0,0015	4,0
12	2	0,0005	1,3

- Očekivanje raspodjele možemo aproksimirati aritmetičkom sredinom uzorka: $\lambda = \bar{x}$ =3,87
- Teorijske frekvencije računamo prema: $f_{tx} = Np(x;\lambda)$

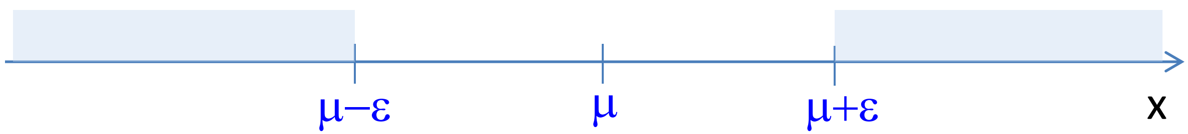


Čebiševljev teorem

- Za slučajnu varijablu ***X*** sa raspodjelom vjerojatnosti ***P(x)*** i očekivanjem ***E(X)=*** μ i varijancom ***V(X)=*** σ^2 vrijedi teorem:

  Vjerojatnost da ***X*** poprimi vrijednost izvan intervala **(μ–ε, μ+ε)** je manja ili jednaka $\boldsymbol{\sigma^2/\varepsilon^2}$, za bilo koji **ε>0**.

$$\boxed{P\{|x-\mu| \geq \epsilon\} \leq \frac{\sigma^2}{\epsilon^2}}$$



<u>Dokaz</u>: po definiciji je varijanca jednaka: $\sigma^2 = \sum_i (x_i - \mu)^2 P(x_i)$

Posebno je za sve x iz osjenčanog intervala:

$$\sigma^2 \geq \sum_{|x_i-\mu|\geq\epsilon} (x_i - \mu)^2 P(x_i)$$

Obzirom da je: $|x_i - \mu| \geq \epsilon$ onda vrijedi:

$$\sigma^2 \geq \epsilon^2 \sum_{|x_i-\mu|\geq\epsilon} P(x_i)$$

Budući da se x_i međusobno isključuju vrijedi: $\sum_{|x_i-\mu|\geq\epsilon} P(x_i) = P\{|x-\mu| \geq \epsilon\}$

Odavde slijedi tvrdnja

Zakon velikih brojeva

- Za Bernoullijev događaj s vjerojatnošću ***P(A)=p***, koji se ponavlja u ***n*** pokusa, vjerojatnost slijedi binomnu raspodjelu sa očekivanjem $\boldsymbol{\mu=np}$ i varijancom $\boldsymbol{\sigma^2=npq}$. Empirijski je relativna frekvencija tog događaja: $f_{rx}=\frac{x}{n}$ (gdje je x broj nastupa u n pokušaja)

Bernoullijev teorem: ako broj pokušaja teži u beskonačno onda relativna frekvencija teži *po vjerojatnosti* stvarnoj vjerojatnosti događaja:

$$\boxed{\lim_{n\to\infty} P\{|f_{rx}-p|\geq\epsilon\}=0}$$ (za bilo koji ε)

<u>Dokaz</u>

$$P\{|f_{rx}-p|\geq\epsilon\}=P\left\{\left|\frac{x}{n}-p\right|\geq\epsilon\right\}=P\{|x-np|\geq n\epsilon\}$$

Prema Čeb. Tm (uz np=μ): $P\{|x-np|\geq n\epsilon\}\leq\frac{\sigma^2}{n^2\epsilon^2}$

$$P\{|f_{rx}-p|\geq\epsilon\}\leq\frac{\sigma^2}{n^2\epsilon^2}=\frac{npq}{n^2\epsilon^2} \Rightarrow P\{|f_{rx}-p|\geq\epsilon\}\leq\frac{pq}{n\epsilon^2}$$

Za lim(n-> beskonačno:) $P\{|f_{rx}-p|\geq\epsilon\}=0$ (Q.E.D.)

Druge raspodjele diskretnih varijabli

- **Hipergeometrijska raspodjela**
  - Populacija od N elemenata (konačna populacija)
  - Svaki element može biti uspješan (A) ili neuspješan ($\bar{A}$)
  - Postoji M uspješnih elemenata (fiksno!)
  - Izvlačenje bilo kojeg uzorka od ***n*** elemenata je jednako vjerojatno
  - Slučajna varijabla ***X=broj uspješnih elemenata u uzorku***
  - **h(x)**=broj mogućih uzoraka s ***x*** uspjeha/ukupni broj mogućih uzoraka

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

Druge raspodjele diskretnih varijabli

- **Hipergeometrijska raspodjela**
  - Očekivanje $E(X) = n\frac{M}{N}$
  - Varijanca $V(X) = n\frac{M}{N}\frac{N-M}{N}\frac{N-n}{N-1}$
  - Uz definicije: p=M/N i q=(N-M)/N slijedi:

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

- Primjer: skup ima 50 proizvoda od kojih 40 dobrih i 10 loših. Slučajno odaberemo 5 proizvoda. Koja je vjerojatnost da se u uzorku nađe x dobrih proizvoda? -> h(x; 5,40,50)

Druge raspodjele diskretnih varijabli

- **Pascalova raspodjela**
  - Niz nezavnisih pokušaja
  - Svaki pokušaj je Bernoullijev pokus koji može biti uspješan (A) ili neuspješan (Ā)
  - Vjerojatnost uspjeha ***p*** jednaka je za sve pokušaje
  - Pokušaji se izvode sve dok se ne opazi ***r*** uspjeha (r unaprijed određen)
  - Slučajna varijabla ***X = broj neuspjeha koji prethode r-tom uspjehu***
  - Za razliku od binomne broj pokušaja je varijabilan, a broj uspjeha je zadan

$$nb(x;r,p) = P(r-1\,A\;u\;x+r-1\;pokušaja)P(uspjeh)$$

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

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