DISKRETNE SLUČAJNE VARIJABLE

Diskretne slučajne varijable

UVOD

DISKRETNE SLUČAJNE VARIJABLE

BERNOULLIJEVI POKUSI

BINOMNA RASPODJELA

POISSONOVA RASPODJELA

HIPERGEOMETRIJSKA RASPODJELA

UVOD

•***slučajna varijabla*** je veličina čije vrijednosti ovise o ishodu slučajnog pokusa, tj. ovisno o pojedinom slučaju dobiva različite vrijednosti

• slučajnoj varijabli je pridružena neka ***raspodjela vjerojatnosti***

primjer: bacanje novčića

prostor elementarnih događaja

$$\Omega = \{G, P\}$$

slučajna varijabla

$$X(\omega) = \begin{cases} 0, & \omega = G \\ 1, & \omega = P \end{cases}$$

DISKRETNE SLUČAJNE VARIJABLE

• slučajne varijable mogu biti ***diskretne*** *ili* ***kontinuirane***

• ***diskretna slučajna varijabla*** može poprimati konačno ili prebrojivo mnogo vrijednosti

• diskretna slučajna varijabla **karakterizirana** je svojom **raspodjelom**, odnosno **vrijednostima** x_1 , x_2 , … koje može poprimiti i **odgovarajućim vjerojatnostima** $p(x_1)$, $p(x_2)$, općenito $p_i = P(X = xi)$, koje te vrijednosti imaju

• skup svih parova $\{x_i, p(x_i)\}$ tvori **razdiobu (distribuciju) slučajne varijable *x***

• ***funkcija vjerojatnosti diskretne slučajne varijable*** je jednoznačno preslikavanje skupa x_i u skup p_i, a opisuje kako se vjerojatnost ishoda mijenja s njegovom brojčanom vrijednošću

DISKRETNE SLUČAJNE VARIJABLE

• za diskretnu slučajnu varijablu koja može poprimiti vrijednosti x_1 , x_2 , … **funkcija vjerojatnosti** je funkcija za koju vrijedi:

$$f(x_i) \geq 0$$

$$\sum_{i=1}^{n} f(x_i) = 1$$

$$f(x_i) = P(X = x_i)$$

• **kumulativna funkcija raspodjele** diskretne slučajne varijable *X* je funkcija sa sljedećim svojstvima

$$F(x_0) = \sum_{x_i \leq x_0} p(x_i)$$

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} f(x_i)$$

$$0 \leq F(x) \leq 1$$

$$x \leq y \Rightarrow F(x) \leq F(y)$$

ta funkcija pokazuje kolika je vjerojatnost da slučajna varijabla *x* poprimi bilo koju vrijednost manju od x_0 ili jednako x_0

DISKRETNE SLUČAJNE VARIJABLE

• **srednja ili očekivana vrijednost** diskretne slučajne varijable je

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \overline{x} = \sum_{i=1}^{r} f_i x_i$$

• zbroj udaljenosti svih slučajnih varijabli od njihove srednje ili očekivane vrijednosti jednak je nuli *(analogija sa osnovama statistike i srednjom vrijednosti gdje smo rekli da je srednja vrijednost ona za koju vrijedi da je oko nje suma odstupanja nula)*

$$\sum_{i=1}^{r} f_i \left(x_i - \overline{x} \right) = 0$$

Teorem $\overline{y} = a \cdot \overline{x} + b$

PRIMJER.

Označimo s *x* ishod bacanja kocke. Tada slučajna varijabla *x* može poprimiti vrijednosti 1, 2, 3, 4, 5, 6, svaka ima vjerojatnost 1/6.

Kolika je očekivana vrijednost?

$$\overline{x} = \sum_i x_i \cdot p(x_i) = 1 \cdot 1/6 + 2 \cdot 1/6 + 3 \cdot 1/6 + 4 \cdot 1/6 + 5 \cdot 1/6 + 6 \cdot 1/6 = 3{,}5$$

Očekivana vrijednost je 3,5.

Uvedemo novi varijablu *y* koju definiramo kao *y* = 2*x* + 1.

Kolika je njena očekivana vrijednost?

$$\overline{y} = 3 \cdot 1/6 + 5 \cdot 1/6 + \ldots + 13 \cdot 1/6 = 8$$

Isti rezultat dobijemo koristeći relaciju

$$\overline{y} = 2 \cdot \overline{x} + 1$$

DISKRETNE SLUČAJNE VARIJABLE

- ***medijan*** *diskretne slučajne varijable*
- ukoliko je broj podataka neparan, za medijan se uzima vrijednost srednjeg podatka
- ukoliko je broj podataka paran, za medijan se uzima srednja vrijednost dva srednja podatka

20 19 17 **15** 3 2 1

10 9 **8** **4** 3 2

6

DISKRETNE SLUČAJNE VARIJABLE

varijanca diskretne slučajne varijable

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

$$\sigma^2 = \sum_{i=1}^{r} f_i (x_i - \overline{x})^2$$

varijanca govori o raspršenju slučajnih varijabli oko njihove srednje vrijednosti, a naziva se još i srednji kvadrat odstupanja

DISKRETNE SLUČAJNE VARIJABLE

standardna devijacija je drugi korijen varijance

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2} = \sqrt{\sum_{i=1}^{r} f_i\,(x_i - \overline{x})^2}$$

standardna devijacija je dimenzijski jednaka diskretnoj slučajnoj varijabli *x*

BERNOULLIJEVI POKUSI

- ***Bernoullijevi pokusi*** - **neovisni pokusi** koji se ponavljaju i kod kojih postoje **samo dva moguća ishoda** za svaki pokus te se vjerojatnosti tih ishoda **ne mijenjaju** tijekom pokusa (npr. niz neovisnih bacanja novčića, npr. bacanje kocke uz pitanje "dobivam 6", …)

- vjerojatnost koja odgovara povoljnom ishodu pokusa se označava s ***p***, vjerojatnost koja odgovara nepovoljnom ishodu pokusa se označava s ***q***

- p i q moraju biti neke **nenegativne vrijednosti** te mora vrijediti: $p + q = 1$, $q = 1 - p$

BERNOULLIJEVI POKUSI

- **prostor elementarnih događaja** za svaki pojedini pokus se sastoji od ukupno **dvije točke** (povoljni ishod, engl. *success (S) i nepovoljni* ishod, engl. *Failure (F)*

- prostor elementarnih događaja za **ukupno *n*** *Bernoullijevih pokusa* sastoji od ukupno **2^n točaka**, a budući da su **pokusi međusobno neovisni,** ukupna vjerojatnost je u**množak njihovih pojedinih vjerojatnosti**

$$P\{(\mathrm{S},\mathrm{S},\mathrm{F},\mathrm{S})\} = ppqp = pp(1-p)p = p^3(1-p)^1$$

$$P\{(\mathrm{F},\mathrm{S},\mathrm{F},\mathrm{S},\mathrm{F},\mathrm{S})\} = qpqpqp = (1-p)p(1-p)p(1-p)p$$
$$= p^3(1-p)^3$$

BERNOULLIJEVI POKUSI

Općenito, za Bernoullijev pokus vrijedi da je vjerojatnost nastupanja povoljnog ishoda ***x*** puta u ***n*** Berrnoullijevih pokusa (ALI S TOČNO ZADANIM REDOSLIJEDOM NASTUPANJA POVOLJNIH I NEPOVOLJNIH ISHODA) dana izrazom:

$$P_{x,n-x}^{n} = p^{x} q^{n-x} = p^{x} (1-p)^{n-x}$$

Bernoullijeva shema pokusa je samo teorijski model i jedino iskustvo može pokazati da li je takva shema prikladna za opis određenih opažanja.

BERNOULLIJEVI POKUSI

Primjer.
Kolika je vjerojatnost da u tri bacanja novčića padne grb, a zatim dva pisma?

$$P_{x,n-x}^{n} = p^{x} q^{n-x} = p^{x} (1 - p)^{n-x}$$

GPP

$P = 0{,}5^{1} * 0{,}5^{2} = 0{,}125$

ŠTO AKO NAS NE ZANIMA REDOSLIJED DOGAĐAJA, NEGO SAMO ISHOD?

Primjer.
Kolika je vjerojatnost da u tri bacanja novčića padne grb i dva pisma?

GPP PGP PPG

- svaka od tih kombinacija ima istu vjerojatnost od 0,125

$P = 0{,}125 * 3 = 0{,}375$

BINOMNA RASPODJELA

Ukoliko nas **ne zanima** redoslijed nastupanja povoljnih i nepovoljnih ishoda, onda je vjerojatnost nastupanja povoljnog ishoda ***x*** puta u **n Bernoullijevih pokusa** dana izrazom:

$$P^{n}_{x,n-x} = \binom{n}{x} p^{x} q^{n-x}$$

$\binom{n}{x}$ predstavlja broj načina na koji se dana kombinacija uspjeha i neuspjeha može ostvariti

n – broj Bernoullijevih pokusa

x – broj pokusa s uspješnim ishodom

n-x – broj pokusa s neuspješnim ishodom

BINOMNA RASPODJELA

primjer:

Ako se simetrični novčić baca 4 puta zaredom, koliko puta se u tih 4 bacanja može dogoditi da novčić padne na grb točno 2 puta (poredak nije bitan)?

Rješenje:

(G,G,P,P) (G,P,G,P) (G,P,P,G)

(P,G,G,P) (P,G,P,G) (P,P,G,G)

-imamo 6 kombinacija

Rješenje:

$$\binom{4}{2} = \frac{4!}{(4-2)! \cdot 2!} = 6$$

BINOMNA RASPODJELA

primjer:

Ako se simetrični novčić baca 4 puta zaredom, koliko je vjerojatnost da u tih 4 bacanja padne na grb točno 2 puta (poredak nije bitan)?

Rješenje:

(G,G,P,P) (G,P,G,P) (G,P,P,G)

(P,G,G,P) (P,G,P,G) (P,P,G,G)

-imamo 6 kombinacija, svaka ima vjerojatnost

$0{,}5^4$

- ukupna vjerojatnost je $P=6*0{,}5^4 = 0{,}375$

Rješenje:

$$P^{n}_{x,n-x} = \binom{n}{x} p^{x} q^{n-x}$$

$$P^{4}_{2,2} = \binom{4}{2} 0{,}5^{2} \cdot 0{,}5^{2} = 0{,}375$$

BINOMNA RASPODJELA

• često se za neki događaj traži samo **vjerojatnost za ukupan broj** povoljnih ishoda u ***n*** uzastopnih Bernoullijevih pokusa, ali njihov poredak nije bitan

• **broj načina** na koji se takav događaj kod kojeg će *n* neovisnih pokusa rezultirati sa *x* povoljnih *i n – x* nepovoljnih ishoda može se dogoditi je:

$$\binom{n}{x}$$

BINOMNA RASPODJELA

• budući da su svi pokusi međusobno neovisni, a vjerojatnost povoljnog događaja je ***p*** (ukupno će se dogoditi ***x*** povoljnih događaja) i vjerojatnost nepovoljnog događaja je ***q* = *1 – p*** (ukupno će se dogoditi ***n- x*** nepovoljnih događaja), svaki od tih $\binom{n}{x}$ načina imat će vjerojatnost:

$$P = p^x q^{n-x} = p^x (1-p)^{n-x}$$

• neka je vjerojatnost da ***n*** Bernoullijevih pokusa, svaki s vjerojatnošću ***p*** povoljnog ishoda i s vjerojatnošću ***q= 1-p*** nepovoljnog ishoda, rezultira s ***x*** povoljnih ishoda i ***n-x*** nepovoljnih ishoda, tada je S_n broj povoljnih ishoda u ukupno *n* pokusa i predstavlja slučajnu varijablu, dok je s ***P(x; n,p)*** zadana njena **BINOMNA RASPODJELA**

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

<u>BINOMNA RASPODJELA</u>

• opisuju je dva parametra: broj jedinki u uzorku ili broj ponavljanja pokušaja - ***n***, i vjerojatnost uspjeha za svaki jedinku ili za svaki pokušaj - ***p***

• **srednja ili očekivana vrijednost** kod binomne raspodjele je vrijednost slučajne varijable koju očekujemo ako promatrano ***n*** pokušaja i iznosi ***n·p***;

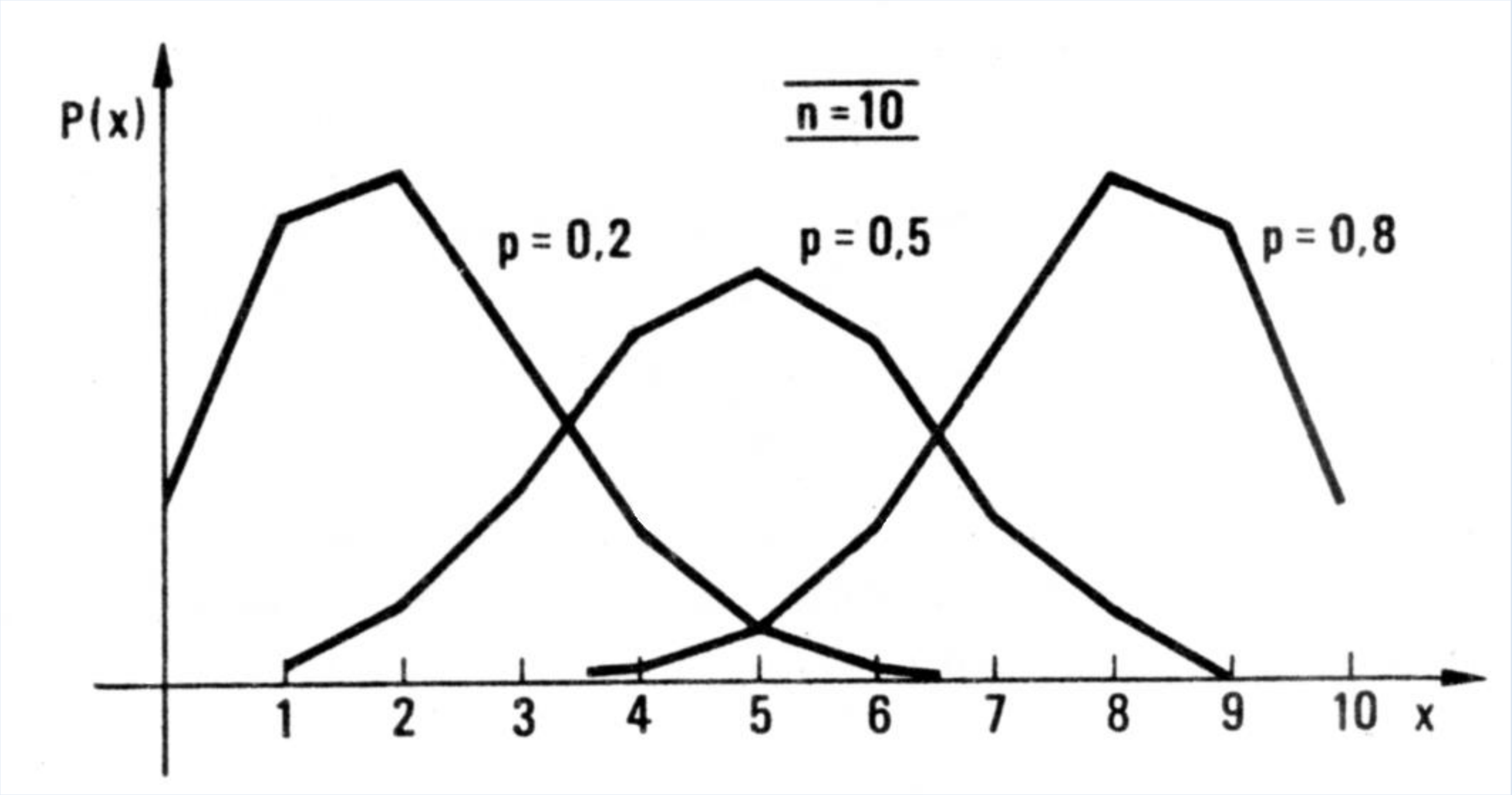
• **varijanca** kod binomne distribucije iznosi ***n·p·(1-p);***

• binomna raspodjela je **simetrična** za $p=q=0,5$, pozitivno asimetrična za $p<q$, negativno asimetrična za $p>q$

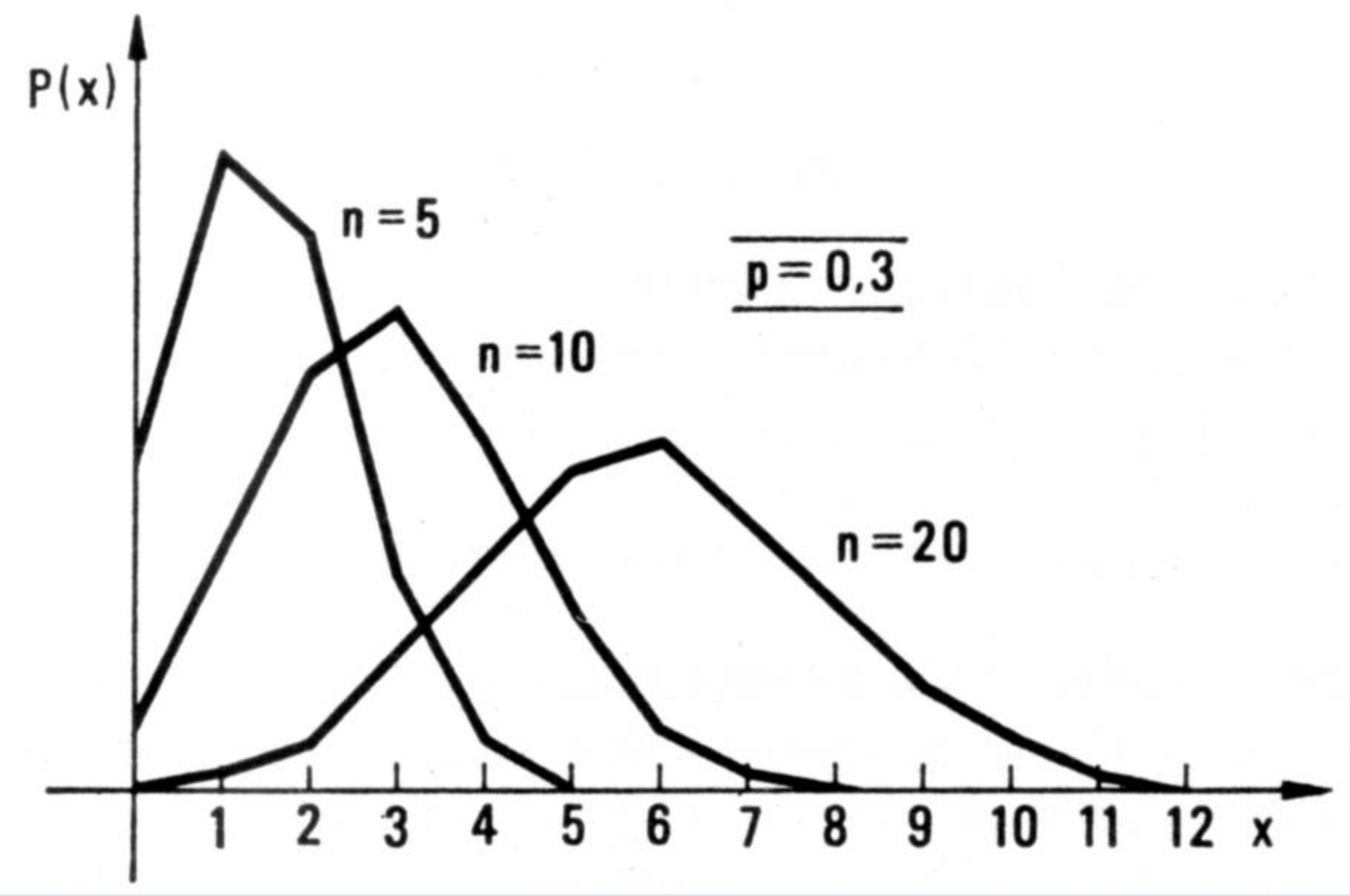
• kada ***n*** raste u beskonačnost, asimetričnost se smanjuje bez obzira na odnos p i q

• upotrebljava se pri zaključivanju o proporcijama, značajna je od statističkih testova, postoje tablice binomne raspodijele (računanje bionomnih koeficijenta za velike ***n*** i ***x***), izvod za rekurzivnu formulu, …

BINOMNA RASPODJELA



BINOMNA RASPODJELA



BINOMNA RASPODJELA

TEOREM: Pri binomnoj raspodjeli *P(x; n,p)* najveća vjerojatnost pripada onoj vrijednosti $\boldsymbol{x_0}$ varijable $\boldsymbol{x}$ koja zadovoljava nejednakost:

$$np - q \leq x_0 \leq np + p$$

DOKAZ

OBJAŠNJENJE

Primjer:
Bacamo istovremeno 7 novčića. Broj grbova koji će pasti na novčićima označimo s x. Koji broj grbova ima najveća vjerojatnost ?

Rješenje:
P(3) = P(4) = 0,273

*n.b. Očekivana vrijednost je n*p=3,5!*

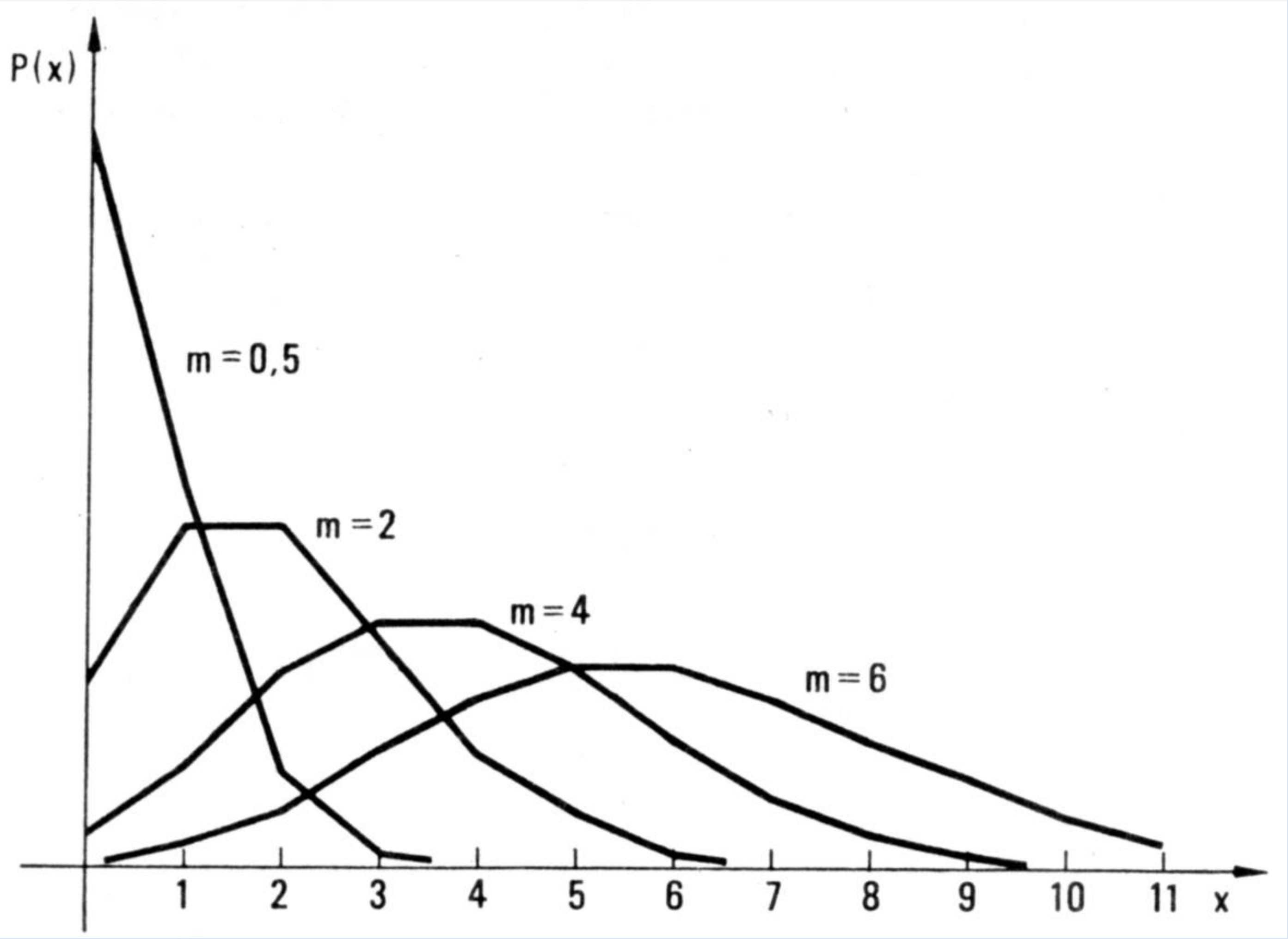
POISSONOVA RASPODJELA

- do Poissonove raspodjele možemo doći graničnim prijelazom na binomnoj raspodjeli
- pustimo neka ***n*** teži u beskonačnosti, uz uvjet da je produkt ***np*** konstantan!
- zbog navedenog uvjeta, ***p*** će težiti nuli (jedino u tom slučaju će produkt ***np*** ostati konstantan)
- u tom slučaju, navedena raspodjela bit će određena izrazom:

$$P(x;m)=\frac{m^x}{x!}e^{-m}$$

- Poissonova raspodjela određena je parametrom ***m (m=np)*** i njen će oblik i karakteristike ovisiti o njemu
- sa slike se vidi da Poissonova raspodjela za svaki ***m*** pozitivno asimetrična, ali porastom ***m*** se smanjuje njena asimetričnost
- *n.b. m>0 (jer je m=np)!*

POISSONOVA RASPODJELA



POISSONOVA RASPODJELA

- IZVOD POISSONOVE RASPODJELE IZ BINOMNE

- IZVOD ZA OČEKIVANU VRIJEDNOST

- računanje vjerojatnosti po Poissonovoj raspodjeli je **jednostavnije** nego po binimonoj raspodjeli te se u situacijama u kojima je kod bionomne raspodjele ***p*** malen, a ***n*** velik, može načiniti **aproksimacija** i računati po Poissonovoj raspodjeli (aproksimacija je to bolja što je ***p*** manji i što je ***n*** veći!)

POISSONOVA RASPODJELA

- računanje vjerojatnosti po Poissonovoj raspodjeli je jednostavnije nego po binimonoj raspodjeli te se u situacijama u kojima je kod bionomne raspodjele ***p*** malen, a ***n*** velik, može načiniti aproksimacija i računati po Poissonovoj raspodjeli (aproksimacija je to bolja što je ***p*** manji i što je ***n*** veći!)

Primjer:
Pri proizvodnji pipeta nastaje 4% neispravnih pipeta. Kolika je vjerojatnost da u uzorku od 10 pipeta ne nađemo niti jednu ili samo jednu neispravnu pipetu?

BIONOMNA RASPODJELA:
$P(x \leq 1) = P(0) + P(1) = 0{,}6648 + 0{,}2270 = 0{,}9418$

POISSONOVA RASPODJELA:
$P(x \leq 1) = P(0) + P(1) = e^{-0{,}4} + 0{,}4 * e^{-0{,}4} = 0{,}9384$

- rezultati se ne razlikuju JAKO puno, iako je n relativno malen!

HIPERGEOMETRIJSKA RASPODJELA

- u osnovnom skupu imamo ***N*** elemenata
- njih ***M*** ima neko obilježje ***A***, što implicira da preostalih ***N-M*** elementa to obilježje nema

- uzmemo iz osnovnog skupa ***n*** elemenata slučajnim odabirom
- među tim elementima naći će se i onih koji imaju i onih koji nemaju obilježje ***A***
- broj elemenata koji imaju obilježje ***A*** u novom skupu elemenata označit ćemo s ***x***
- očito, ***x*** može poprimiti bilo koju vrijednost između 0 i ***n*** (ako je ***n ≤ M***)

- **Kolika je vjerojatnost da *x* poprimi neku vrijednost (odnosno, koja je vjerojatnost da u nasumično izabranom uzorku elemenata imamo *x* elemenata s obilježjem *A*)?**

HIPERGEOMETRIJSKA RASPODJELA

- općenito, uzorak od ***n*** elemenata od ukupno ***N*** elemenata moguće je složiti na slijedeći broj načina:

$$K_n^{(N)} = \binom{N}{n}$$

- nadalje, ***x*** elemenata s obilježjem ***A*** (od ukupno njih ***M***) moguće je složiti na slijedeći broj načina:

$$K_x^{(M)} = \binom{M}{x}$$

- ostalih ***n-x*** elemenata u ukupnom uzorku nemaju obilježje ***A***, te je njih moguće složiti na slijedeći broj načina:

$$K_{n-x}^{(N-M)} = \binom{N-M}{n-x}$$

HIPERGEOMETRIJSKA RASPODJELA

-svaka grupa od ***x*** elemenata s obilježjem ***A***, ujedinjena s bilo kojom grupom od ***n-x*** elemenata koji nemaju obilježje ***A***, također čini traženi uzorak (uzorak koji sadrži ***x*** elemenata s obilježjem ***A***, čiju vjerojatnost tražimo)

- vjerojatnost takvog uzorka je dana slijedećim izrazom:

$$P(x) = \frac{K_x^{(M)} \cdot K_{n-x}^{(N-M)}}{K_n^{(N)}} = \frac{\binom{M}{x} \cdot \binom{N-M}{n-x}}{\binom{N}{n}}$$

HIPERGEOMETRIJSKA RASPODJELA

- skup uređenih parova {*x, P(x)*} tvori **hipergeometrijsku raspodjelu** koja je jednoznačno određena parametrima: ***N, M, n***
- očekivana vrijednost: ***n·(M/N);*** varijanca: ***n·(M/N)·((N-M)/N)·((N-n)/(N-1))***
- rekurzivna formula može se izvesti na sličan način kao i kod binomne raspodijele
- ukoliko ***N*** teži beskonačnosti, ali na način da pri tome ***M/N*** teži realnom, nenegativnom broju manjem od 1 (vjerojatnosti ***p***), i uz uvjet da ***n*** ostane **konstantan**, hipergeometrijska raspodjela **teži bionomnoj raspodjeli** (IZVOD)

- problem kojim se bavi hipergeometrijska raspodjela sličan je Bernoullijevom pokusu, razlika je u tome da svako pojedino izvlačenje mijenja sastav skupa iz kojeg se vrši sljedeće izvlačenje (shodno tome mijenja se i vjerojatnost "uspješnog" izvlačenja)