Ilkka Mellin
Todennäköisyyslaskenta

Osa 3: Todennäköisyysjakaumia
Diskreettejä jakaumia

Diskreettejä jakaumia

>> **Diskreetti tasainen jakauma**

Bernoulli-jakauma

Binomijakauma

Geometrinen jakauma

Negatiivinen binomijakauma

Hypergeometrinen jakauma

Poisson-jakauma

Diskreetti tasainen jakauma ja sen pistetodennäköisyysfunktio 1/2

- Olkoon X *diskreetti satunnaismuuttuja*, jonka mahdolliset arvot ovat

$$x_1, x_2, \dots , x_n$$

- Oletetaan, että satunnaismuuttujan X mahdollisiin arvoihin $x_1, x_2, \dots , x_n$ liittyvät todennäköisyydet ovat *yhtä suuria*:

$$\Pr(X = x_k) = \frac{1}{n}, \ k = 1, 2, \dots, n$$

- Huomautus:

  Diskreetti tasainen jakauma liittyy sellaisiin otosavaruuksiin, joissa alkeistapaukset ovat *symmetrisiä*.

Diskreetti tasainen jakauma ja sen pistetodennäköisyysfunktio 2/2

- Satunnaismuuttujan X **pistetodennäköisyysfunktio** on

$$f(x) = \Pr(X = x) = \frac{1}{n}, \; x = x_k$$

$$k = 1, 2, \ldots, n$$

- Sanomme, että satunnaismuuttuja X noudattaa **diskreettiä tasaista jakaumaa**.
- Huomautus:

Funktio $f(x)$ määrittelee todennäköisyysjakauman, koska

$$\sum_{k=1}^{n} f(x_k) = n \cdot \frac{1}{n} = 1$$

Odotusarvo, varianssi ja standardipoikkeama

- Diskreetin tasaisen jakauman **odotusarvo**:

$$\mathrm{E}(X) = \mu_X = \overline{x} = \frac{1}{n}\sum_{k=1}^{n} x_k$$

- Diskreetin tasaisen jakauman **varianssi**:

$$\mathrm{Var}(X) = \mathrm{D}^2(X) = \sigma_X^2 = \frac{1}{n}\sum_{k=1}^{n}(x_k - \overline{x})^2$$

- Diskreetin tasaisen jakauman **standardipoikkeama**:

$$\mathrm{D}(X) = \sigma_X = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(x_k - \overline{x})^2}$$

Diskreetti tasainen jakauma
Odotusarvon ja varianssin johto

- Suoraan diskreetin satunnaismuuttujan odotusarvon ja varianssin määritelmistä saadaan:

$$
\begin{aligned}
\mathrm{E}(X) &= \mu \\
&= \sum_{k=1}^{n} x_k f(x_k) \\
&= \frac{1}{n}\sum_{k=1}^{n} x_k = \overline{x}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Var}(X) &= \mathrm{D}^2(X) = \sigma^2 \\
&= \sum_{k=1}^{n} (x_k - \mu)^2 f(x_k) \\
&= \frac{1}{n}\sum_{k=1}^{n} (x_k - \overline{x})^2
\end{aligned}
$$

Odotusarvon ominaisuuksia

- Diskreetin tasaisen jakauman odotusarvo

$$\mathrm{E}(X) = \mu_X = \frac{1}{n}\sum_{k=1}^{n} x_k = \bar{x}$$

on satunnaismuuttujan X mahdollisten arvojen $x_1, x_2, \ldots , x_n$ **aritmeettinen keskiarvo**.

Odotusarvon ja varianssin laskeminen: Esimerkki

- Olkoon satunnaismuuttujan X pistetodennäköisyysfunktio

$$f(k) = \Pr(X = k) = p_k = 1/6 \text{ , } k = 1, 2, 3, 4, 5, 6$$

- Odotusarvo:

$$\mathrm{E}(X) = \sum_{k=1}^{6} k f(k) = \frac{1}{6}\sum_{k=1}^{6} k$$

$$= \frac{1}{6}(1+2+3+4+5+6) = 3.5$$

- Varianssi:

$$\mathrm{D}^2(X) = \sum_{k=1}^{6} (k - \mathrm{E}(x))^2 f(k) = \frac{1}{6}\sum_{k=1}^{6} (k - \mathrm{E}(x))^2$$

$$= \frac{1}{6}\left[(1-3.5)^2 + (2-3.5)^2 + \cdots + (6-3.5)^2\right] = \frac{35}{12} \approx 2.917$$

- Standardipoikkeama:

$$\mathrm{D}(X) = \sqrt{2.917} \approx 1.708$$

Diskreetti tasainen jakauma
Pistetodennäköisyysfunktion kuvaaja

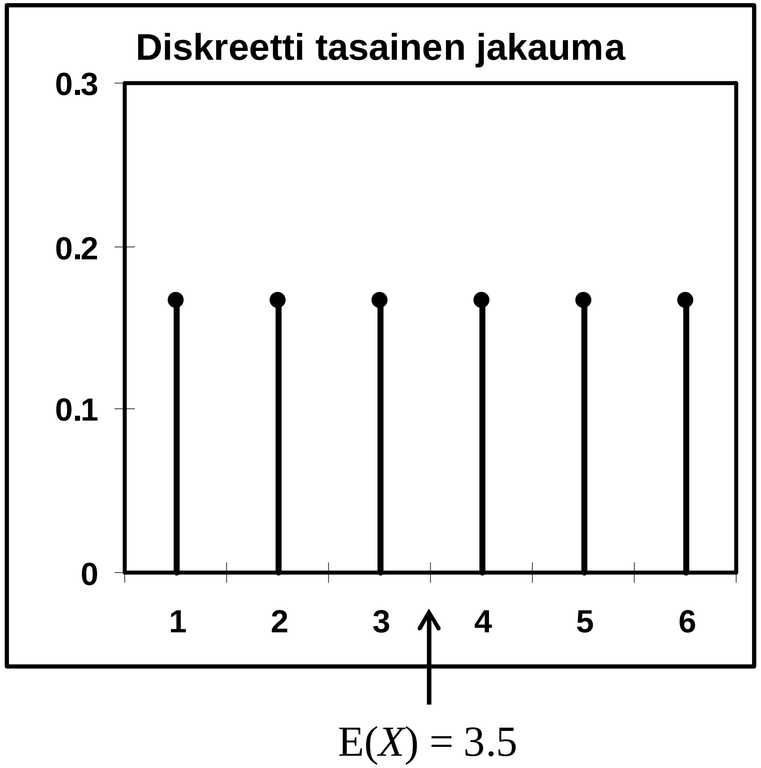
- Kuva oikealla esittää diskreetin tasaisen jakauman

$$f(x) = \frac{1}{6},\ x = 1, 2, 3, 4, 5, 6$$

pistetodennäköisyysfunktiota.

- Jakauman *odotusarvo*:

$$\mathrm{E}(X) = \frac{1}{6}\sum_{k=1}^{6} k = 3.5$$



Diskreettejä jakaumia

Diskreetti tasainen jakauma
>> Bernoulli-jakauma
Binomijakauma
Geometrinen jakauma
Negatiivinen binomijakauma
Hypergeometrinen jakauma
Poisson-jakauma

Bernoulli-jakauma ja sen pistetodennäköisyysfunktio 1/2

- Olkoon A otosavaruuden S *tapahtuma* ja $\Pr(A) = p$.
- Tällöin tapahtuman A *komplementtitapahtuman* A^c *todennäköisyys* on

$$\Pr(A^c) = 1 - \Pr(A) = 1 - p = q$$

- Määritellään *diskreetti satunnaismuuttuja* X:

$$X = \begin{cases} 1, & \text{jos tapahtuma } A \text{ sattuu} \\ 0, & \text{jos tapahtuma } A \text{ ei satu} \end{cases}$$

- Tällöin satunnaismuuttujan X *jakauma* on

$$\Pr(X = 1) = p$$

$$\Pr(X = 0) = 1 - p = q$$

Bernoulli-jakauma ja sen pistetodennäköisyysfunktio 2/2

- Satunnaismuuttujan X **pistetodennäköisyysfunktio** on

$$f(x) = \Pr(X = x) = p^x q^{1-x} \,,\ 0 < p < 1 \,,\ q = 1 - p$$

$$x = 0,1$$

- Sanomme, että satunnaismuuttuja X noudattaa **Bernoulli-jakaumaa parametrinaan *p***.
- Merkintä:

$$X \sim \text{Bernoulli}(p)$$

- Huomautus:

Funktio $f(x)$ määrittelee todennäköisyysjakauman, koska

$$f(0) + f(1) = q + p = 1$$

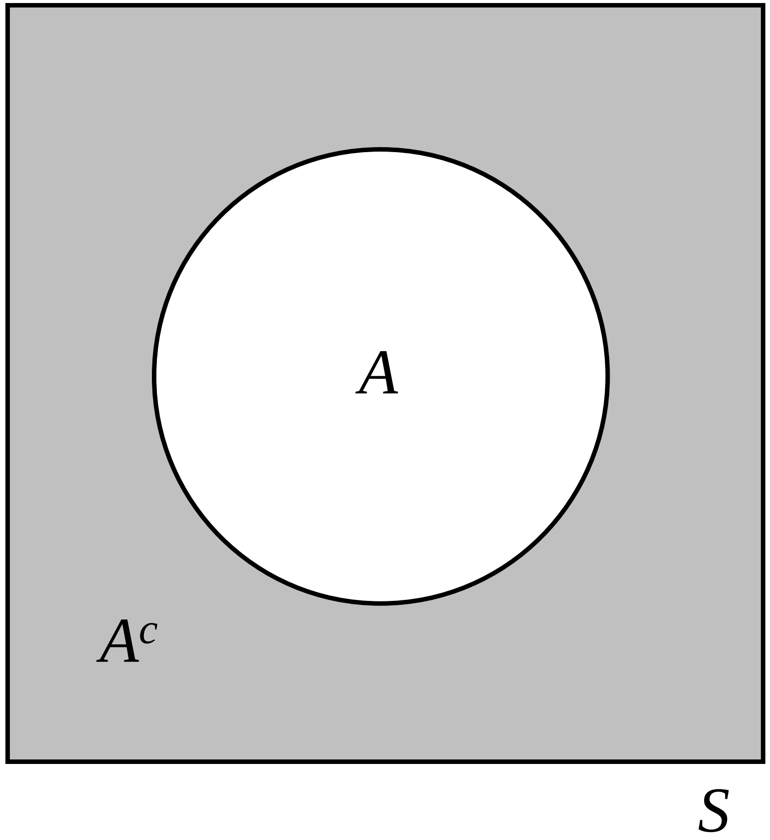
Komplementtitapahtuman todennäköisyys

- Olkoon

  $\Pr(A) = p$

- Tällöin

  $\Pr(A^c)$
  $= 1 - \mathrm{P}(A)$
  $= 1 - p$
  $= q$



Odotusarvo, varianssi ja standardipoikkeama

- Olkoon

$$X \sim \text{Bernoulli}(p)$$

- **Odotusarvo**:

$$\mathrm{E}(X) = p$$

- **Varianssi** ja **standardipoikkeama**:

$$\mathrm{Var}(X) = \mathrm{D}^2(X) = pq$$

$$\mathrm{D}(X) = \sqrt{pq}$$

Odotusarvon ja varianssin johto

- Suoraan diskreetin satunnaismuuttujan odotusarvon ja varianssin määritelmistä saadaan:

$$
\begin{aligned}
\mathrm{E}(X) &= 1\times\Pr(X=1)+0\times\Pr(X=0) \\
&= 1\times p+0\times q \\
&= p \\
\mathrm{E}(X^2) &= 1^2\times\Pr(X=1)+0^2\times\Pr(X=0) \\
&= 1\times p+0\times q \\
&= p \\
\mathrm{Var}(X) &= \mathrm{E}(X^2)-[\mathrm{E}(X)]^2 \\
&= p-p^2 \\
&= p(1-p) \\
&= pq
\end{aligned}
$$

Bernoulli-jakauma

Odotusarvon ja varianssin ominaisuuksia

- Olkoon

  $X \sim \text{Bernoulli}(p)$

- Bernoulli-jakauman odotusarvo

  $\text{E}(X) = p$

  yhtyy tapahtuman A todennäköisyyteen $\Pr(A) = p$.

- Bernoulli-jakauman varianssi

  $\text{Var}(X) = pq = p(1 - p) = p - p^2$

  saavuttaa maksiminsa

  $1/4$

  kun $p = q = 1/2$.

Bernoulli-jakauma
Pistetodennäköisyysfunktion kuvaaja

- Kuva oikealla esittää Bernoulli-jakauman

  Bernoulli(0.8)

  pistetodennäköisyysfunktiota

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

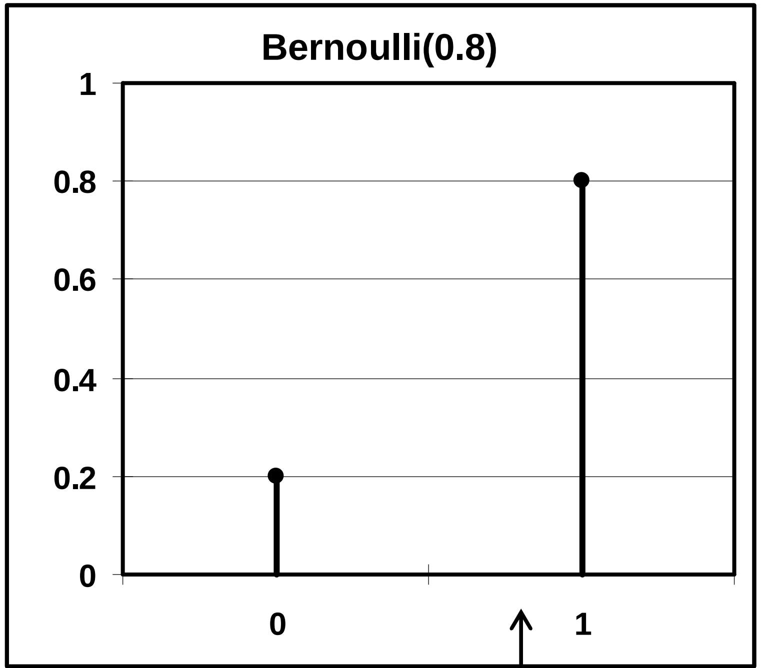
  $$p = 0.8\,,\ q = 1 - p$$

  pisteissä

  $x = 0, 1$

- Jakauman *odotusarvo*:

  $$\mathrm{E}(X) = p = 0.8$$



$\mathrm{E}(X) = 0.8$

Bernoulli-jakaumaa noudattavien satunnaismuuttujien summan jakauma 1/2

- Olkoot

  $$X_1, X_2, \ldots, X_n$$

  riippumattomia satunnaismuuttujia, jotka noudattavat *samaa Bernoulli-jakaumaa* parametrilla p:

  $$X_1, X_2, \ldots, X_n \perp$$

  $$X_i \sim \mathrm{Bernoulli}(p)\ ,\ i = 1, 2, \ldots, n$$

- Tällöin satunnaismuuttujien $X_1, X_2, \ldots, X_n$ **summa**

  $$Y = X_1 + X_2 + \cdots + X_n$$

  noudattaa binomijakaumaa parametrilla $(n\ , p)$:

  $$Y \sim \mathrm{Bin}(n\ , p)$$

Bernoulli-jakaumaa noudattavien satunnaismuuttujien summan jakauma 2/2

- Tulos perustellaan luvussa **Satunnaismuuttujien muunnosten jakaumat.**
- Huomautuksia:
  - Kaikilla Bernoulli-jakaumilla on oltava *sama* tapahtuman A todennäköisyyttä kuvaava parametri p.
  - Tarkoitamme *satunnaismuuttujien riippumattomuudella* sitä, että yhdenkään satunnaismuuttujan saamat arvot eivät riipu siitä, mitä arvoja muut satunnaismuuttujat saavat; käsite täsmennetään luvussa **Moniulotteiset satunnaismuuttujat ja jakaumat.**

Bernoulli-kokeet ja diskreetit todennäköisyysjakaumat 1/2

- Toistetaan *toisistaan riippumatta samaa Bernoulli-koetta* ja tarkastellaan tapahtuman A sattumista toistojen aikana:

  (i) **Binomijakauma** saadaan määräämällä todennäköisyys sille, että tapahtuma A sattuu x kertaa, kun koetta toistetaan n kertaa, jossa n on kiinteä, etukäteen päätetty luku.

  (ii) **Geometrinen jakauma** saadaan määräämällä todennäköisyys sille, että tapahtuma A sattuu ensimmäisen kerran x. koetoistossa.

  (iii) **Negatiivinen binomijakauma** saadaan määräämällä todennäköisyys sille, että tapahtuma A sattuu r. kerran x. koetoistossa.

Bernoulli-kokeet ja diskreetit todennäköisyysjakaumat 2/2

- **Poisson-jakauma** voidaan johtaa binomijakauman raja-arvona, kun koetoistojen lukumäärän annetaan tiettyjen ehtojen vallitessa kasvaa rajatta.

  Siten Poisson-jakauma kuvaa *harvinaisten* tapahtumien todennäköisyyksiä *pitkissä* toistokoesarjoissa.

Diskreettejä jakaumia

Diskreetti tasainen jakauma
Bernoulli-jakauma
>> Binomijakauma
Geometrinen jakauma
Negatiivinen binomijakauma
Hypergeometrinen jakauma
Poisson-jakauma

Binomijakauma ja sen pistetodennäköisyysfunktio 1/3

- Toistetaan *samaa* Bernoulli-koetta n kertaa, jossa n on *kiinteä*, *etukäteen päätetty* luku.
- Oletetaan, että koetoistot ovat *riippumattomia*.
- Tarkastellaan otosavaruuden S tapahtuman A sattumista koetoistojen aikana.
- Oletetaan, että

$$\Pr(A) = p$$

$$\Pr(A^c) = 1 - \Pr(A) = 1 - p = q$$

- Määritellään *diskreetti satunnaismuuttuja* X:

$X =$ Tapahtuman A esiintymisten lukumäärä n-kertaisessa Bernoulli-kokeessa

Binomijakauma ja sen pistetodennäköisyysfunktio 2/3

- Satunnaismuuttujan X **pistetodennäköisyysfunktio** on

$$f(x) = \Pr(X = x) = \binom{n}{x} p^x q^{n-x} \ , \ 0 < p < 1 \ , \ q = 1 - p$$

$$x = 0, 1, 2, \ldots, n$$

- Sanomme, että satunnaismuuttuja X noudattaa **binomijakaumaa parametreinaan *n* ja *p***.
- Merkintä:

$$X \sim \mathrm{Bin}(n, p)$$

Binomijakauma ja sen pistetodennäköisyysfunktio 3/3

- Huomautus:

Funktio $f(x)$ määrittelee todennäköisyysjakauman, koska *binomikaavan* mukaan

$$\sum_{x=0}^{n} f(x) = \sum_{x=0}^{n} \binom{n}{x} p^x q^{n-x} = (p+q)^n = 1^n = 1$$

Siten binomijakauman *pistetodennäköisyydet*

$$p_x = f(x) = \binom{n}{x} p^x q^{n-x} \text{ , } x = 0, 1, 2, \ldots, n$$

toteuttavat yhtälön

$$p_0 + p_1 + p_2 + \cdots + p_n = \sum_{x=0}^{n} \binom{n}{x} p^x q^{n-x} = 1$$

Binomijakauma
Pistetodennäköisyysfunktion johto 1/2

- Toistetaan *samaa* Bernoulli-koetta n kertaa.
- Oletetaan, että koetoistot ovat *riippumattomia* ja tarkastellaan tapahtuman A sattumista koetoistojen aikana.
- Oletetaan, että toistokoesarjan tuloksena saadaan tapahtumajono

$$A\,A\,A^c\,A\,A^c \ldots A$$

jossa on x kpl tapahtumia A ja $(n - x)$ kpl tapahtumia A^c .
- Koska

$$\Pr(A) = p$$

$$\Pr(A^c) = 1 - \Pr(A) = 1 - p = q$$

tarkasteltavan tapahtumajonon todennäköisyydeksi saadaan *riippumattomien tapahtumien tulosäännön* nojalla

$$ppqpq \ldots p = p^x q^{n-x}$$

Binomijakauma
Pistetodennäköisyysfunktion johto 2/2

- *Erilaisia* jonoja, joissa on x kpl tapahtumia A ja $(n-x)$ kpl tapahtumia A^c, on

$$\binom{n}{x} \text{ kpl}$$

- *Erilaiset* tapahtumajonot ovat *toisensa poissulkevia*.
- *Toisensa poissulkevien tapahtumien yhteenlaskusäännön* mukaan todennäköisyys saada sellainen jono, jossa on x kpl tapahtumia A ja $(n-x)$ kpl tapahtumia A^c saadaan laskemalla *erilaisten* tällaisten jonojen todennäköisyydet yhteen.
- Siten kysytyksi todennäköisyydeksi saadaan

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

$$x = 0, 1, 2, \ldots, n$$

Odotusarvo, varianssi ja standardipoikkeama

- Olkoon

$$X \sim \mathrm{Bin}(n, p)$$

- **Odotusarvo**:

$$\mathrm{E}(X) = np$$

- **Varianssi** ja **standardipoikkeama**:

$$\mathrm{Var}(X) = \mathrm{D}^2(X) = npq$$

$$\mathrm{D}(X) = \sqrt{npq}$$

Binomijakauma

Odotusarvon johto 1/2

- Olkoon

$$X \sim \mathrm{Bin}(n, p)$$

- Tällöin

$$
\begin{aligned}
\mathrm{E}(X) = \sum_{x=0}^{n} x f(x) &= \sum_{x=0}^{n} x \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \\
&= \sum_{x=1}^{n} x \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \\
&= \sum_{x=1}^{n} \frac{n!}{(x-1)!(n-x)!} p^x (1-p)^{n-x} \\
&= np \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} (1-p)^{n-x} \\
&= np
\end{aligned}
$$

Binomijakauma

Odotusarvon johto 2/2

- Kalvon 1/2 yhtälöketjun viimeinen yhtälö perustuu siihen, että

$$\sum_{x=1}^{n}\frac{(n-1)!}{(x-1)!(n-x)!}p^{x-1}(1-p)^{n-x}$$

$$=\sum_{x=0}^{n-1}\frac{(n-1)!}{x!(n-1-x)!}p^{x}(1-p)^{n-1-x}=1$$

- Tämä seuraa siitä, että jälkimmäisessä summassa lasketaan yhteen *kaikki* binomijakauman

  Bin(n – 1, p)

  pistetodennäköisyydet

$$f(x)=\frac{(n-1)!}{x!(n-1-x)!}p^{x}(1-p)^{n-1-x}\ ,\ x=0,1,2,\ldots,n-1$$

Binomijakauma

Odotusarvon ja varianssin ominaisuuksia

- Olkoon

$$X \sim \mathrm{Bin}(n, p)$$

- Binomijakauman odotusarvo

$$\mathrm{E}(X) = np$$

on suoraan verrannollinen sekä *toistokeiden lukumäärään* n että *tapahtuman A todennäköisyyteen* $\Pr(A) = p$.

- Binomijakauman varianssi

$$\mathrm{Var}(X) = npq = np(1-p)$$

saavuttaa maksiminsa

$$n/4$$

kun $p = q = 1/2$.

Binomijakauma
Pistetodennäköisyysfunktion kuvaaja

- Kuva oikealla esittää Binomi-jakauman

  Bin(12, 1/3)

  pistetodennäköisyysfunktiota

  $$f(x) = \binom{n}{x} p^x q^{n-x}$$

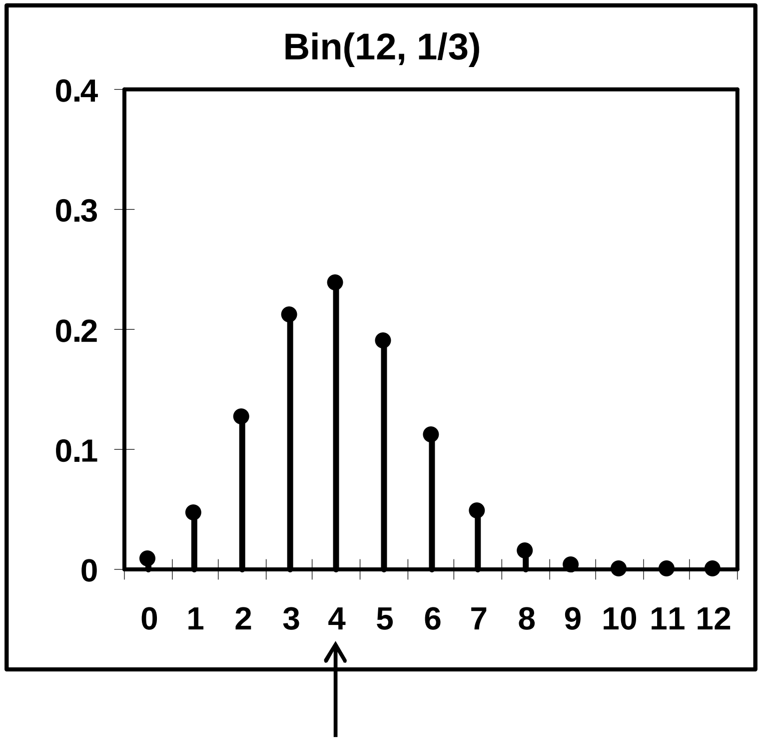
  $$n = 12\,,\; p = 1/3\,,\; q = 1 - p$$

  pisteissä

  $x = 0, 1, 2, \ldots , 12$

- Jakauman *odotusarvo*:

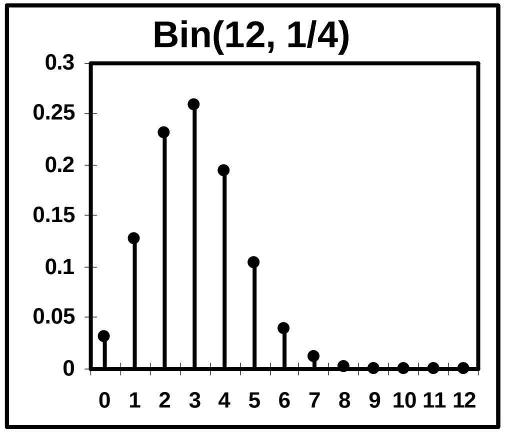
  $$\mathrm{E}(X) = np = 4$$

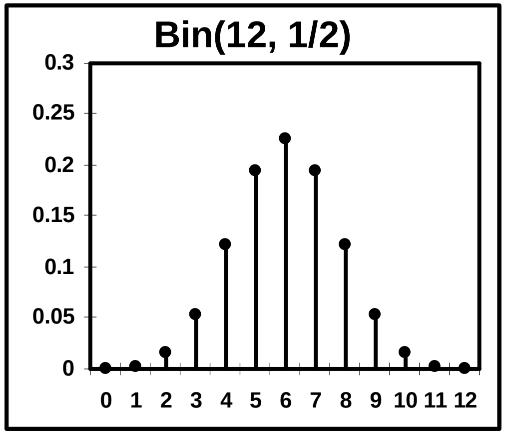


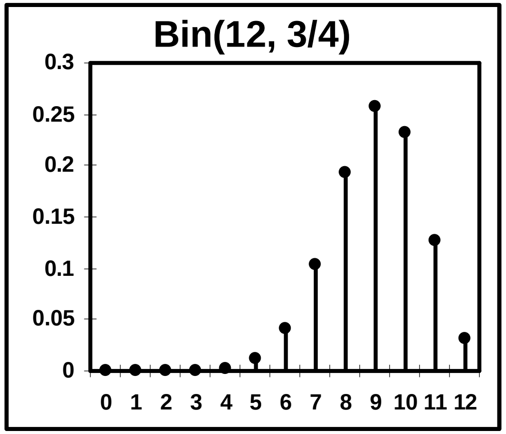
E(X) = 4

Binomijakauma

Pistetodennäköisyysfunktion kuvaaja: Tapaukset *p* < 1/2, *p* = 1/2, *p* > 1/2







- $p < 1/2$: Binomijakauma on *vino oikealle*.
- $p = 1/2$: Binomijakauma on *symmetrinen*.
- $p > 1/2$: Binomijakauma on *vino vasemmalle*.

Binomijakauma ja Bernoulli-jakauma 1/3

- Toistetaan *samaa* Bernoulli-koetta n kertaa, jossa n on *kiinteä*, *etukäteen päätetty* luku.
- Oletetaan, että koetoistot ovat *riippumattomia* ja tarkastellaan tapahtuman A sattumista koetoistojen aikana.
- Oletetaan, että

$$\Pr(A) = p$$

$$\Pr(A^c) = 1 - p = q$$

Binomijakauma ja Bernoulli-jakauma 2/3

- Määritellään *diskreetit satunnaismuuttujat* X_i , $i = 1, 2, \ldots , n$:

$$X_i = \begin{cases} 1, \text{ jos } A \text{ tapahtuu kokeessa } i \\ 0, \text{ jos } A \text{ ei tapahdu kokeessa } i \end{cases}$$

- Tällöin

$$X_i \sim \text{Bernoulli}(p),\ i = 1, 2, \ldots , n.$$

- Määritellään *diskreetti satunnaismuuttuja* X :

$$X = \text{Tapahtuman } A \text{ esiintymisten lukumäärä } n\text{-kertaisessa Bernoulli-kokeessa}$$

- Tällöin

$$X \sim \text{Bin}(n, p)$$

Binomijakauma ja Bernoulli-jakauma 3/3

- Selvästi

$$X = \sum_{i=1}^{n} X_i$$

koska luku 1 esiintyy summassa $\sum X_i$ *täsmälleen* yhtä monta kertaa kuin tapahtuma A sattuu n:n koetoiston aikana.

- Tämä merkitsee sitä, että *binomijakautunut satunnais-muuttuja* voidaan esittää *riippumattomien Bernoulli-jakautuneiden satunnaismuuttujien summana*.
- Huomautus:

  Binomi- ja Bernoulli-jakauman yhteyttä voidaan käyttää apuna binomijakauman *odotusarvon* ja *varianssin* määräämisessä; ks. >.

Binomijakauman odotusarvon ja varianssin johto sekä Bernoulli-jakauma 1/2

- Olkoot X_i , $i = 1,2, \ldots , n$ *riippumattomia* satunnaismuuttujia, jotka noudattavat *samaa Bernoulli-jakaumaa* parametrilla p:

  $X_1 , X_2 , \ldots , X_n \perp$

  $X_i \sim \text{Bernoulli}(p),\ i = 1,2, \ldots , n$
- Olkoon

  $$X = \sum_{i=1}^{n} X_i$$
- Tällöin

  $X \sim \text{Bin}(n, p)$
- Huomautus:

  Tarkoitamme *satunnaismuuttujien riippumattomuudella* sitä, että yhdenkään satunnaismuuttujan saamat arvot eivät riipu siitä, mitä arvoja muut satunnaismuuttujat saavat; käsite täsmennetään luvussa **Moniulotteiset satunnaismuuttujat ja jakaumat**.

Binomijakauman odotusarvon ja varianssin johto sekä Bernoulli-jakauma 2/2

- Satunnaismuuttujan $X = \sum X_i$ *odotusarvo* on

$$\mathrm{E}(X) = \mathrm{E}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{E}(X_i) = \sum_{i=1}^{n} p = np$$

koska *satunnaismuuttujien summan odotusarvo on satunnaismuuttujien odotusarvojen summa*.

- Satunnaismuuttujan $X = \sum X_i$ *varianssi* on

$$\mathrm{D}^2(X) = \mathrm{D}^2\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{D}^2(X_i) = \sum_{i=1}^{n} pq = npq$$

koska *riippumattomien satunnaismuuttujien summan varianssi on satunnaismuuttujien varianssien summa*.

Binomijakaumaa noudattavien satunnaismuuttujien summan jakauma 1/2

- Olkoot

$$X_1, X_2, \ldots , X_k$$

riippumattomia satunnaismuuttujia, jotka noudattavat *binomijakaumia* parametrein $(n_1 , p), (n_2 , p), \ldots , (n_k , p)$:

$$X_1, X_2, \ldots , X_k \perp$$

$$X_i \sim \mathrm{Bin}(n_i , p) \ , i = 1, 2, \ldots , k$$

- Tällöin satunnaismuuttujien $X_1, X_2, \ldots , X_k$ **summa**

$$Y = X_1 + X_2 + \cdots + X_k$$

noudattaa binomijakaumaa parametrein $(n_1 + n_2 + \cdots + n_k , p)$:

$$Y \sim \mathrm{Bin}(n_1 + n_2 + \cdots + n_k , p)$$

Binomijakaumaa noudattavien satunnaismuuttujien summan jakauma 2/2

- Tulos perustellaan luvussa **Satunnaismuuttujien muunnosten jakaumat**.
- Huomautuksia:
  - Kaikilla binomijakaumilla on oltava *sama* tapahtuman A todennäköisyyttä kuvaava parametri p, mutta sen sijaan toistokokeiden lukumäärää kuvaava parametri saa vaihdella jakaumasta toiseen.
  - Tarkoitamme *satunnaismuuttujien riippumattomuudella* sitä, että yhdenkään satunnaismuuttujan saamat arvot eivät riipu siitä, mitä arvoja muut satunnaismuuttujat saavat; käsite täsmennetään luvussa **Moniulotteiset satunnaismuuttujat ja jakaumat**.

Binomijakauma ja otanta takaisinpanolla 1/5

- Olkoon perusjoukon S alkioiden lukumäärä

$$n(S) = N$$

- *Poimitaan* perusjoukosta S *satunnaisesti osajoukko* B, jonka alkioiden lukumäärä on

$$n(B) = n$$

käyttämällä poiminnassa *otantaa takaisinpanolla*.

Binomijakauma ja otanta takaisinpanolla 2/5

- **Otanta takaisinpanolla**:

  (i) Perusjoukosta S poimitaan alkiot osajoukkoon B yksi kerrallaan *arpomalla*.

  (ii) Jokainen poimittu alkio **palautetaan** ennen seuraavan alkion arpomista *takaisin* perusjoukkoon S.

  (iii) *Jokaisella perusjoukon S alkiolla* on *jokaisessa arvonnassa sama todennäköisyys*

  $$1/N$$

  tulla poimituksi osajoukkoon B.

- Osajoukko B muodostaa **yksinkertaisen satunnais-otoksen** perusjoukosta S.

Binomijakauma ja otanta takaisinpanolla 3/5

- Otannassa takaisinpanolla arvonta voidaan toteuttaa seuraavalla tavalla:

  (1) Pannaan *uurnaan* jokaista perusjoukon *S* alkiota vastaava arpalippu.

  (2) *Sekoitetaan* arvat huolellisesti.

  (3) *Nostetaan* uurnasta arpalippu, jota vastaava alkio valitaan otokseen *B*.

  (4) **Palautetaan** nostettu arpalippu uurnaan.

  (5) Palataan vaiheeseen (2), kunnes haluttu otoskoko *n* on saavutettu.

Binomijakauma ja otanta takaisinpanolla 4/5

- Huomautuksia otannasta takaisinpanolla:

  (i) Jokaisen perusjoukon *S alkion todennäköisyys* tulla valituksi otokseen *säilyy samana* koko poiminnan ajan.

  (ii) Jokaisella perusjoukon *S* samankokoisella *osajoukolla* on *sama* todennäköisyys tulla valituksi otokseksi.

  (iii) Sama perusjoukon *S* alkio voi tulla valituksi *useita kertoja* otokseen.

Binomijakauma ja otanta takaisinpanolla 5/5

- Olkoon A perusjoukon osajoukko, jonka alkioiden lukumäärä on

$$n(A) = r$$

- Tällöin todennäköisyys poimia alkio joukosta A on

$$\Pr(A) = p = \frac{r}{N}$$

- Otannassa takaisinpanolla otokseen poimittujen A-tyyppisten alkioiden lukumäärä X on diskreetti satunnaismuuttuja, joka noudattaa *binomijakaumaa parametreilla* n ja p:

$$X \sim \text{Bin}(n, p)$$

Diskreettejä jakaumia

Diskreetti tasainen jakauma
Bernoulli-jakauma
Binomijakauma
>> Geometrinen jakauma
Negatiivinen binomijakauma
Hypergeometrinen jakauma
Poisson-jakauma

Geometrinen jakauma ja sen pistetodennäköisyysfunktio 1/2

- Toistetaan *samaa* Bernoulli-koetta.
- Oletetaan, että koetoistot ovat *riippumattomia*.
- Tarkastellaan otosavaruuden S tapahtuman A sattumista koetoistojen aikana.
- Oletetaan, että

$$\Pr(A) = p$$

$$\Pr(A^c) = 1 - \Pr(A) = 1 - p = q$$

- Määritellään *diskreetti satunnaismuuttuja* X:

$X =$ Tehtyjen Bernoulli-kokeiden lukumäärä, kun A sattuu *ensimmäisen* kerran

Geometrinen jakauma ja sen pistetodennäköisyysfunktio 2/2

- Satunnaismuuttujan X **pistetodennäköisyysfunktio** on

$$f(x) = \Pr(X = x) = q^{x-1}p\,,\ 0 < p < 1\,,\ q = 1 - p$$

$$x = 1, 2, 3, \ldots$$

- Sanomme, että satunnaismuuttuja X noudattaa **geometrista jakaumaa parametrinaan *p***.
- Merkintä:

$$X \sim \mathrm{Geom}(p)$$

- Huomautus:

Funktio $f(x)$ määrittelee todennäköisyysjakauman, koska *geometrisen sarjan summan kaavan* mukaan

$$\sum_{x=1}^{\infty} f(x) = \sum_{x=1}^{\infty} q^{x-1}p = p\sum_{x=0}^{\infty} q^{x} = p\frac{1}{1-q} = p\frac{1}{p} = 1$$

Geometrinen jakauma
Pistetodennäköisyysfunktion johto 1/2

- Toistetaan *samaa* Bernoulli-koetta.
- Oletetaan, että koetoistot ovat *riippumattomia* ja tarkastellaan tapahtuman A sattumista koetoistojen aikana.
- Tarkastellaan toistokoesarjaa, jossa tapahtuma A sattuu ensimmäisen kerran x:nnessä kokeessa.
- Toistokoesarjan tuloksena on tällöin ollut tapahtumajono

$$A^c\ A^c\ A^c\ \ldots\ A^c\ A$$

jossa on $(x-1)$ kpl tapahtumia A^c ja jossa tapahtuma A on *viimeisenä*.

Geometrinen jakauma
Pistetodennäköisyysfunktion johto 2/2

- Koska

$$\Pr(A) = p$$

$$\Pr(A^c) = 1 - \Pr(A) = 1 - p = q$$

tarkasteltavan tapahtumajonon todennäköisyydeksi saadaan *riippumattomien tapahtumien tulosäännön* nojalla

$$qqq \cdots qp = q^{x-1}p$$

$$x = 1, 2, 3, \ldots$$

mikä on kysytty todennäköisyys.

Geometrinen jakauma
Odotusarvo, varianssi ja standardipoikkeama

- Olkoon

$$X \sim \mathrm{Geom}(p)$$

- **Odotusarvo**:

$$\mathrm{E}(X) = \frac{1}{p}$$

- **Varianssi** ja **standardipoikkeama**:

$$\mathrm{Var}(X) = \mathrm{D}^2(X) = \frac{q}{p^2}$$

$$\mathrm{D}(X) = \frac{\sqrt{q}}{p}$$

Geometrinen jakauma
Odotusarvon johto 1/4

- Olkoon

$$X \sim \mathrm{Geom}(p)$$

- Satunnaismuuttujan X *pistetodennäköisyysfunktio* on

$$f(x) = q^{x-1}p \text{ , } q = 1 - p$$

$$x = 1, 2, 3, \ldots$$

- Pistetodennäköisyyksien $f(x)$, $x = 1, 2, 3, \ldots$ *summa* on

$$S(p) = \sum_{x=1}^{\infty} f(x) = \sum_{x=1}^{\infty} (1-p)^{x-1} p = 1$$

- Satunnaismuuttujan X *odotusarvo* on

$$\mathrm{E}(X) = \sum_{x=1}^{\infty} xf(x) = \sum_{x=1}^{\infty} x(1-p)^{x-1} p$$

Geometrinen jakauma
Odotusarvon johto 2/4

- Summan $S(p)$ derivaatta muuttujan p suhteen on

$$
\begin{aligned}
\frac{\partial S(p)}{\partial p} &= \sum_{x=1}^{\infty}\left[-(x-1)(1-p)^{x-2}p+(1-p)^{x-1}\right] \\
&= -\sum_{x=1}^{\infty} x(1-p)^{x-2}p \\
&\qquad +\sum_{x=1}^{\infty}(1-p)^{x-2}p+\sum_{x=1}^{\infty}(1-p)^{x-1} \\
&= -\frac{1}{1-p}\sum_{x=1}^{\infty} x(1-p)^{x-1}p \\
&\qquad +\frac{1}{1-p}\sum_{x=1}^{\infty}(1-p)^{x-1}p+\frac{1}{p}\sum_{x=1}^{\infty}(1-p)^{x-1}p
\end{aligned}
$$

Geometrinen jakauma
Odotusarvon johto 3/4

- Ottamalla huomioon yhtälöt

$$\sum_{x=1}^{\infty}(1-p)^{x-1}p = S(p) = 1$$

$$\sum_{x=1}^{\infty}x(1-p)^{x-1}p = \mathrm{E}(X)$$

saadaan yhtälö

$$\begin{aligned}\frac{\partial S(p)}{\partial p} &= -\frac{1}{1-p}\mathrm{E}(X) + \frac{1}{1-p} + \frac{1}{p} \\ &= -\frac{1}{1-p}\mathrm{E}(X) + \frac{1}{p(1-p)} \\ &= 0\end{aligned}$$

Geometrinen jakauma

Odotusarvon johto 4/4

- Geometrisen jakauman odotusarvo E(x) toteuttaa siis yhtälön

$$-\frac{1}{1-p}\mathrm{E}(X)+\frac{1}{p(1-p)}=0$$

- Siten *geometrisen jakauman odotusarvo* on

$$\mathrm{E}(X)=\frac{1}{p}$$

Odotusarvon ominaisuuksia

- Olkoon

$$X \sim \mathrm{Geom}(p)$$

- Geometrisen jakauman odotusarvo

$$\mathrm{E}(X) = \frac{1}{p}$$

  on *kääntäen verrannollinen tapahtuman A todennäköisyyteen* $\Pr(A) = p$.
- Siten tapahtumaa A saa odottaa keskimäärin sitä *kauemmin* mitä *pienempi* on tapahtuman A todennäköisyys.

Pistetodennäköisyysfunktion kuvaaja

- Kuva oikealla esittää geometrisen jakauman

  Geom(1/3)

  pistetodennäköisyysfunktiota

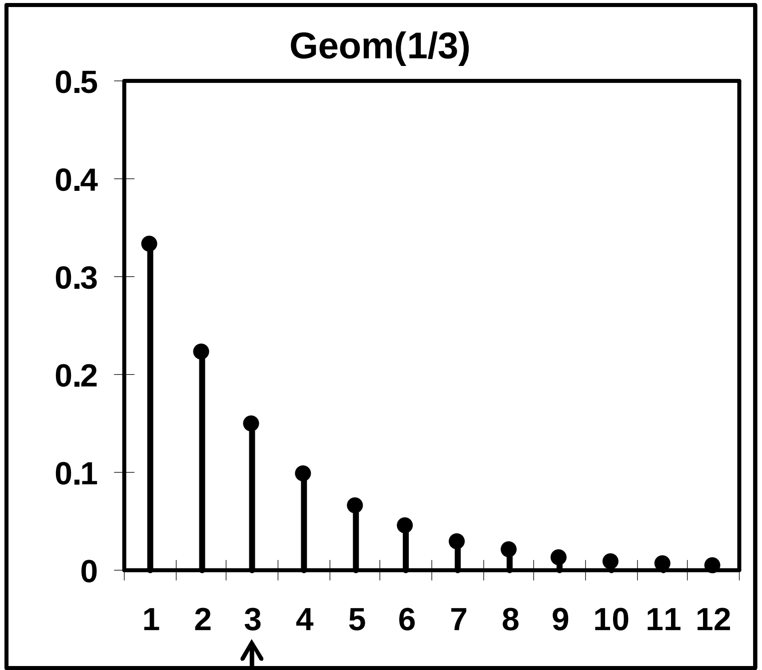
  $$f(x) = q^{x-1} p$$

  $$p = 1/3\,,\ q = 1 - p$$

  pisteissä

  x = 1, 2, … , 12
- Jakauman *odotusarvo*:

  $$\mathrm{E}(X) = \frac{1}{p} = 3$$



E(X) = 3

Geometrisen jakauman unohtamisominaisuus

- Olkoon

  $$X \sim \mathrm{Geom}(p)$$

- Tällöin

  $$\Pr(X \geq a + b \mid X \geq a) = \Pr(X \geq 1 + b)$$

- Siten geometrisella jakaumalla on seuraava *unohtamis-ominaisuus*:

  Se, että tapahtuman A sattumista on jouduttu odottamaan a koetoistoa, *ei vaikuta* todennäköisyyteen joutua odottamaan b koetoistoa lisää.

Diskreettejä jakaumia

Diskreetti tasainen jakauma
Bernoulli-jakauma
Binomijakauma
Geometrinen jakauma
>> Negatiivinen binomijakauma
Hypergeometrinen jakauma
Poisson-jakauma

Negatiivinen binomijakauma ja sen pistetodennäköisyysfunktio 1/2

- Toistetaan *samaa* Bernoulli-koetta.
- Oletetaan, että koetoistot ovat *riippumattomia*.
- Tarkastellaan otosavaruuden S tapahtuman A sattumista koetoistojen aikana.
- Oletetaan, että

$$\Pr(A) = p$$

$$\Pr(A^c) = 1 - \Pr(A) = 1 - p = q$$

- Määritellään *diskreetti satunnaismuuttuja* X:

$X =$ Tehtyjen Bernoulli-kokeiden lukumäärä, kun A sattuu r. kerran

Negatiivinen binomijakauma ja sen pistetodennäköisyysfunktio 2/2

- Satunnaismuuttujan X **pistetodennäköisyysfunktio** on

$$f(x) = \Pr(X = x) = \binom{x-1}{r-1} q^{x-r} p^r \ , 0 < p < 1 \, , q = 1 - p$$

$$r = 1, 2, 3, \ldots \ ; x = r, r+1, r+2, \ldots$$

- Sanomme, että satunnaismuuttuja X noudattaa **negatiivista binomijakaumaa parametreinaan *r* ja *p***.
- Merkintä:

$$X \sim \mathrm{NegBin}(r, p)$$

Negatiivinen binomijakauma

Pistetodennäköisyysfunktion johto 1/3

- Toistetaan *samaa* Bernoulli-koetta.
- Oletetaan, että koetoistot ovat *riippumattomia* ja tarkastellaan tapahtuman A sattumista koetoistojen aikana.
- Tarkastellaan toistokoesarjaa, jossa tapahtuma A sattuu r. kerran x. kokeessa.
- Olkoon toistokoesarjan tuloksena ollut tapahtumajono

$$A\,A\,A^c A\,A^c \ldots A^c A$$

jossa on r kpl tapahtumia A ja $(x-r)$ kpl tapahtumia A^c ja jossa tapahtuma A on *viimeisenä*.

Negatiivinen binomijakauma
Pistetodennäköisyysfunktion johto 2/3

- Koska

$$\Pr(A) = p$$

$$\Pr(A^c) = 1 - \Pr(A) = 1 - p = q$$

tarkasteltavan tapahtumajonon todennäköisyydeksi saadaan *riippumattomien tapahtumien tulosäännön* nojalla

$$ppqpq \cdots qp = q^{x-r}p^r$$

- *Erilaisten* sellaisten jonojen, joissa on r kpl tapahtumia A ja $(x-r)$ kpl tapahtumia A^c ja joissa A on *viimeisenä*, lukumäärä on

$$\binom{x-1}{r-1}$$

Negatiivinen binomijakauma
Pistetodennäköisyysfunktion johto 3/3

- *Erilaiset* tapahtumajonot ovat *toisensa poissulkevia*.
- *Toisensa poissulkevien tapahtumien yhteenlaskusäännön* mukaan todennäköisyys saada sellainen jono, jossa on r kpl tapahtumia A ja $(x-r)$ kpl tapahtumia A^c ja jossa tapahtuma A on *viimeisenä*, saadaan laskemalla *erilaisten* tällaisten jonojen todennäköisyydet yhteen.
- Koska ko. jonojen lukumäärä on

$$\binom{x-1}{r-1}$$

saadaan kysytyksi todennäköisyydeksi

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

$$r = 1, 2, 3, \ldots \ ; \ x = r, r+1, r+2 \ldots$$

Odotusarvo, varianssi ja standardipoikkeama

- Olkoon

$$X \sim \mathrm{NegBin}(r, p)$$

- **Odotusarvo**:

$$\mathrm{E}(X) = \frac{r}{p}$$

- **Varianssi** ja **standardipoikkeama**:

$$\mathrm{Var}(X) = \mathrm{D}^2(X) = \frac{rq}{p^2}$$

$$\mathrm{D}(X) = \frac{\sqrt{rq}}{p}$$

Negatiivinen binomijakauma
Odotusarvon johto 1/4

- Olkoon

  X ~ NegBin(r, p)
- Satunnaismuuttujan X *pistetodennäköisyysfunktio* on

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

$$x = r, r+1, r+2, \ldots$$

- Pistetodennäköisyyksien $f(x)$, $x = r$, $r + 1$, $r + 2$, … *summa* on

$$S(p) = \sum_{x=r}^{\infty} f(x) = \sum_{x=r}^{\infty} \binom{x-1}{r-1} (1-p)^{x-r} p^r = 1$$

- Satunnaismuuttujan X *odotusarvo* on

$$\mathrm{E}(X) = \sum_{x=r}^{\infty} x f(x) = \sum_{x=r}^{\infty} x \binom{x-1}{r-1} (1-p)^{x-r} p^r$$

Odotusarvon johto 2/4

- Summan $S(p)$ *derivaatta* muuttujan p suhteen on

$$\frac{\partial S(p)}{\partial p} = \sum_{x=r}^{\infty}\left[-(x-r)\binom{x-1}{r-1}(1-p)^{x-r-1}p^{r} + r\binom{x-1}{r-1}(1-p)^{x-r}p^{r-1}\right]$$

$$= -\sum_{x=r}^{\infty} x\binom{x-1}{r-1}(1-p)^{x-r-1}p^{r}$$

$$+r\sum_{x=r}^{\infty}\binom{x-1}{r-1}(1-p)^{x-r-1}p^{r} + r\sum_{x=r}^{\infty}\binom{x-1}{r-1}(1-p)^{x-r}p^{r-1}$$

$$= -\frac{1}{1-p}\sum_{x=r}^{\infty} x\binom{x-1}{r-1}(1-p)^{x-r}p^{r}$$

$$+\frac{r}{1-p}\sum_{x=r}^{\infty}\binom{x-1}{r-1}(1-p)^{x-1}p^{r} + \frac{r}{p}\sum_{x=r}^{\infty}\binom{x-1}{r-1}(1-p)^{x-r}p^{r}$$

Odotusarvon johto 3/4

- Ottamalla huomioon yhtälöt

$$\sum_{x=r}^{\infty}\binom{x-1}{r-1}(1-p)^{x-r}p^r = S(p) = 1$$

$$\sum_{x=r}^{\infty}x\binom{x-1}{r-1}(1-p)^{x-r}p^r = \mathrm{E}(X)$$

saadaan yhtälö

$$\begin{aligned}\frac{\partial S(p)}{\partial p} &= -\frac{1}{1-p}\mathrm{E}(X) + \frac{r}{1-p} + \frac{r}{p} \\ &= -\frac{1}{1-p}\mathrm{E}(X) + \frac{r}{p(1-p)} \\ &= 0\end{aligned}$$

Negatiivinen binomijakauma

Odotusarvon johto 4/4

- Negatiivisen binomijakauman odotusarvo E(x) toteuttaa siis yhtälön

$$-\frac{1}{1-p}\mathrm{E}(X)+\frac{r}{p(1-p)}=0$$

- Siten *negatiivisen binomijakauman odotusarvo* on

$$\mathrm{E}(X)=\frac{r}{p}$$

Odotusarvon ominaisuuksia

- Olkoon

$$X \sim \mathrm{NegBin}(r, p)$$

- Negatiivisen binomijakauman odotusarvo

$$\mathrm{E}(X) = \frac{r}{p}$$

  on *suoraan verrannollinen lukuun r* ja *kääntäen verrannollinen tapahtuman A todennäköisyyteen* $\Pr(A) = p$.
- Siten r. tapahtumaa A saa odottaa keskimäärin sitä *kauemmin* mitä *suurempi* on r ja mitä *pienempi* on tapahtuman A todennäköisyys.

Negatiivinen binomijakauma
Pistetodennäköisyysfunktion kuvaaja

- Kuva oikealla esittää negatiivisen binomijakauman

  NegBin(3, 1/3)

  pistetodennäköisyysfunktiota

$$f(x) = \binom{x-1}{r-1} q^{x-r} p^r$$

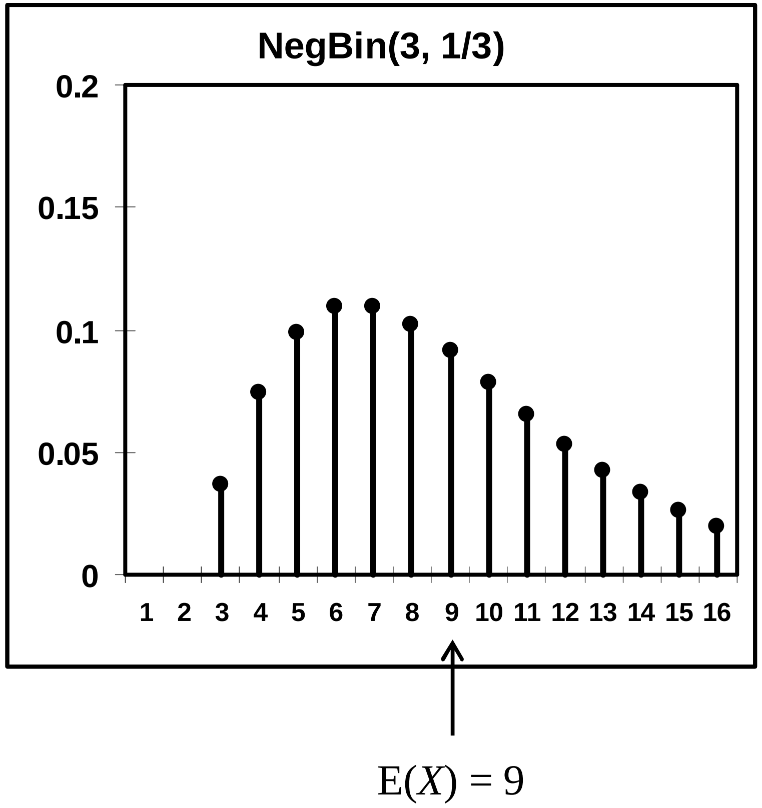
$$r = 3\,,\ p = 1/3\,,\ q = 1 - p$$

  pisteissä

  $x = 3, 4, \ldots, 16$

- Jakauman *odotusarvo*:

$$\mathrm{E}(X) = \frac{r}{p} = 9$$



Negatiivinen binomijakauma ja geometrinen jakauma

- Olkoon

$$X \sim \mathrm{NegBin}(r, p)$$

- Jos $r = 1$, niin satunnaismuuttuja X noudattaa **geometrista jakaumaa** Geom(p):

$$X \sim \mathrm{Geom}(p)$$

- Geometrinen jakauma on siten negatiivisen binomi-jakauman *erikoistapaus*.

Diskreettejä jakaumia

Diskreetti tasainen jakauma
Bernoulli-jakauma
Binomijakauma
Geometrinen jakauma
Negatiivinen binomijakauma
>> Hypergeometrinen jakauma
Poisson-jakauma

Hypergeometrinen jakauma ja sen pistetodennäköisyysfunktio 1/3

- Olkoon perusjoukon S alkioiden lukumäärä

$$n(S) = N$$

- Tarkastellaan perusjoukon S *ositusta* joukkoihin A ja A^c .
- Oletetaan, että joukossa $A \subset S$ on

$$n(A) = r$$

alkiota.

- Tällöin joukon A komplementissa A^c on

$$n(A^c) = N - r$$

alkiota.

Hypergeometrinen jakauma ja sen pistetodennäköisyysfunktio 2/3

- Poimitaan perusjoukosta S *satunnaisesti* osajoukko B, jonka alkioiden lukumäärä on

  $$n(B) = n$$

  käyttämällä poiminnassa *otantaa ilman takaisinpanoa*.
- Määritellään *diskreetti satunnaismuuttuja* X:

  X = Osajoukkoon B tulleiden A:n alkioiden lukumäärä

Hypergeometrinen jakauma ja sen pistetodennäköisyysfunktio 2/2

- Satunnaismuuttujan X **pistetodennäköisyysfunktio** on

$$f(x) = \Pr(X = x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$$

$$\max[0, n-(N-r)] \le x \le \min(n, r)$$

- Sanomme, että satunnaismuuttuja X noudattaa **hypergeometrista jakaumaa parametreilla *N*, *r* ja *n***.
- Merkintä:

$$X \sim \text{HyperGeom}(N, r, n)$$

Hypergeometrinen jakauma

Pistetodennäköisyysfunktion johto 1/4

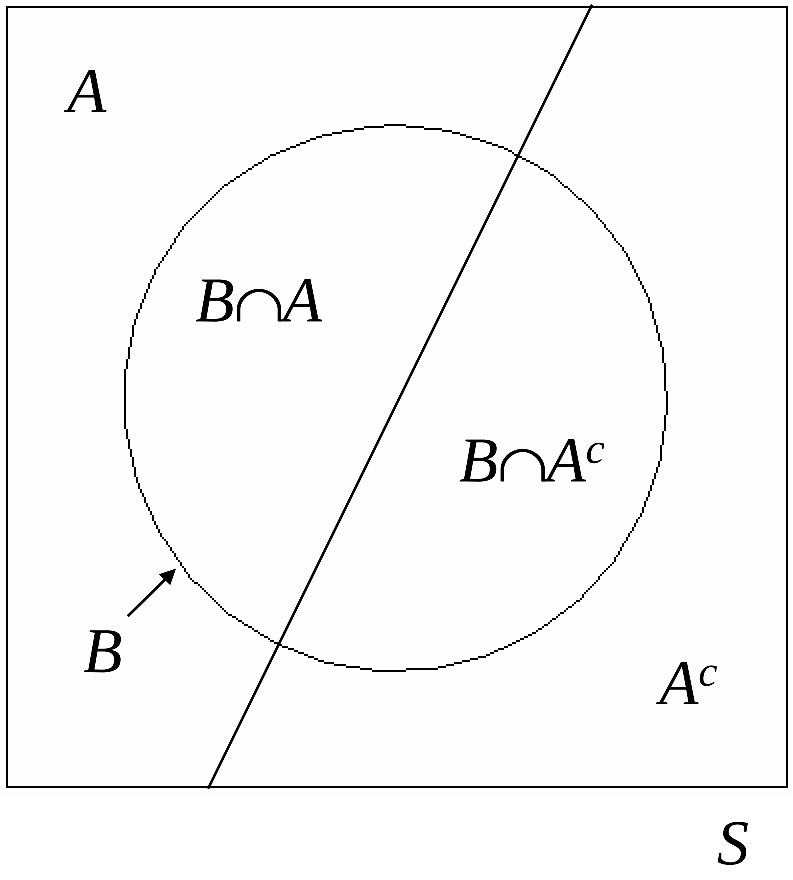
- Olkoon S otosavaruus ja

  $n(S) = N$
- Olkoon $A \subset S$
- Tällöin $\{A, A^c\}$ on otosavaruuden S *ositus*.
- Olkoon $n(A) = r$ ja $n(A^c) = N - r$
- Olkoon $B \subset S$ ja $n(B) = n$
- Otosavaruuden S ositus $\{A, A^c\}$ *indusoi* osituksen joukkoon B:

  $B = (B\cap A)\cup(B\cap A^c)$
- Olkoon

  $n(B\cap A) = x$

  $n(B\cap A^c) = n - x$



Pistetodennäköisyysfunktion johto 2/4

- N:n alkion joukosta S voidaan poimia n:n alkion osajoukko B

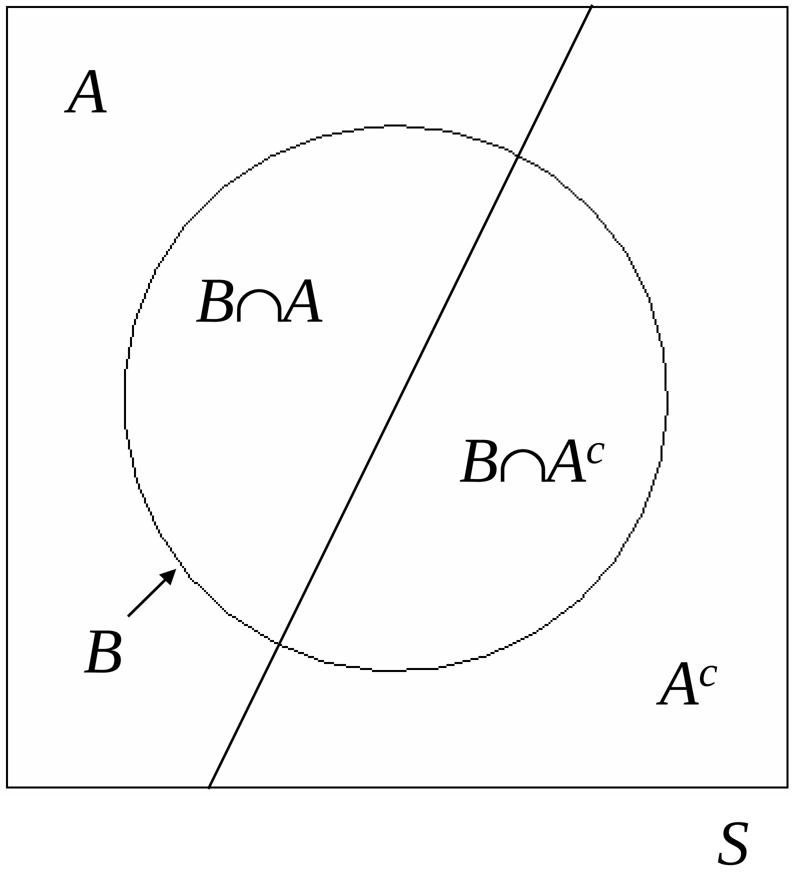
  $\binom{N}{n}$ eri tavalla.

- r:n alkion joukosta A voidaan poimia x alkiota

  $\binom{r}{x}$ eri tavalla.

- $(N-r)$:n alkion joukosta A^c voidaan poimia $n-x$ alkiota

  $\binom{N-r}{n-x}$ eri tavalla.

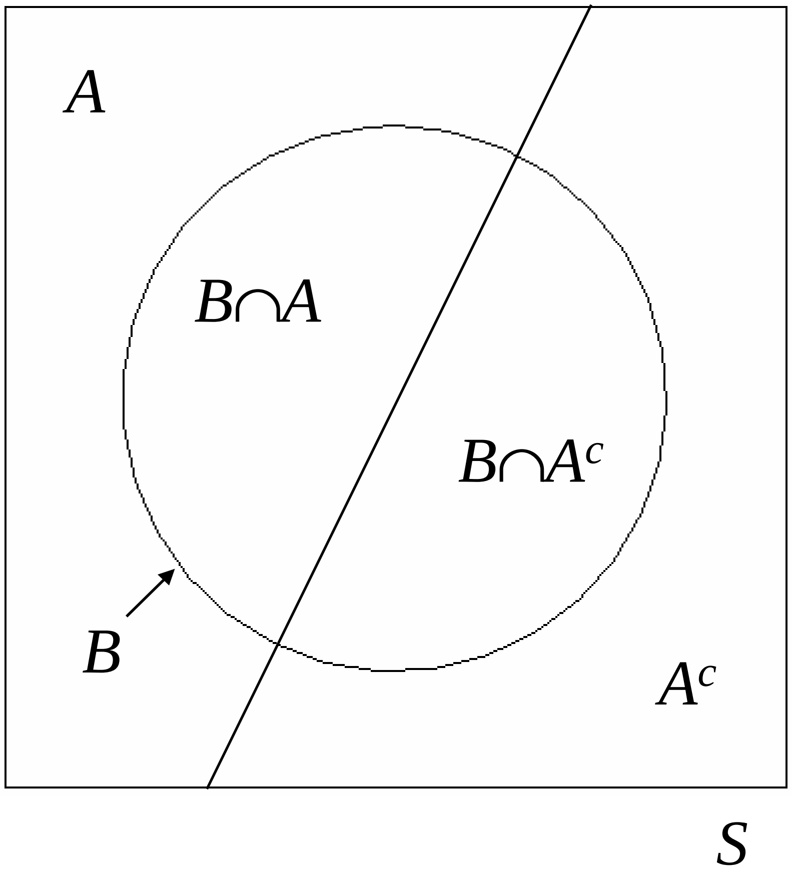


Hypergeometrinen jakauma
Pistetodennäköisyysfunktion johto 3/4

- r:n alkion joukosta A voidaan poimia x alkiota *riippumatta* siitä, mitkä $n - x$ alkiota poimitaan $(N - r)$:n alkion joukosta A^c.
- *Kertolaskuperiaatteen* nojalla n alkiota voidaan poimia joukosta S niin, että saadaan r alkiota joukosta A ja $(N - r)$ alkiota joukosta A^c

$$\binom{r}{x}\binom{N-r}{n-x}$$

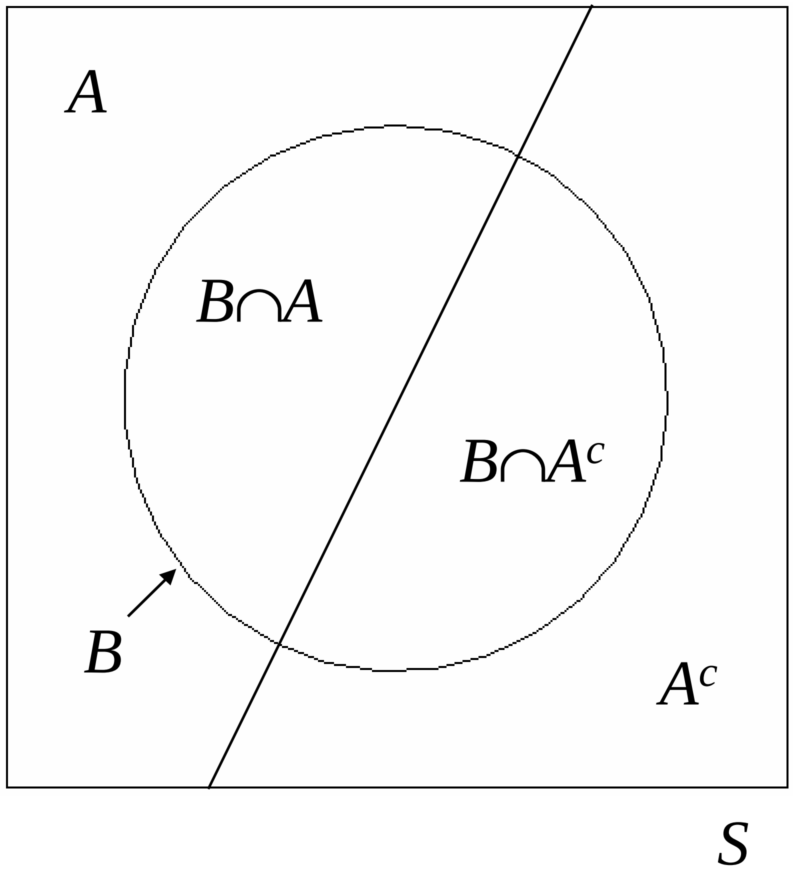
eri tavalla.



Pistetodennäköisyysfunktion johto 4/4

- Soveltamalla *klassisen todennäköisyyden* määritelmää saadaan:

$$\Pr(X = x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$$



Odotusarvo, varianssi ja standardipoikkeama

- Olkoon

$$X \sim \mathrm{HyperGeom}(N, r, n)$$

- **Odotusarvo**:

$$\mathrm{E}(X) = n\frac{r}{N}$$

- **Varianssi** ja **standardipoikkeama**:

$$\mathrm{Var}(X) = \mathrm{D}^2(X) = n\frac{r}{N}\left(1-\frac{r}{N}\right)\left(\frac{N-n}{N-1}\right)$$

$$\mathrm{D}(X) = \sqrt{n\frac{r}{N}\left(1-\frac{r}{N}\right)\left(\frac{N-n}{N-1}\right)}$$

Hypergeometrinen jakauma
Odotusarvon johto 1/3

- Olkoon

$$X \sim \mathrm{HyperGeom}(N, r, n)$$

- Koska

$$x\binom{r}{x} = x\frac{r!}{x!(r-x)!} = r\frac{(r-1)!}{(x-1)!(r-x)!} = r\binom{r-1}{x-1}$$

niin

$$\mathrm{E}(X) = \sum_{x=0}^{n} xf(x) = \sum_{x=0}^{n} x\frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}} = r\sum_{x=1}^{n}\frac{\binom{r-1}{x-1}\binom{N-r}{n-x}}{\binom{N}{n}}$$

Odotusarvon johto 2/3

- Koska

$$\binom{N}{n} = \frac{N!}{n!(N-n)!} = \frac{N}{n} \cdot \frac{(N-1)!}{(n-1)!(N-n)!} = \frac{N}{n}\binom{N-1}{n-1}$$

niin

$$\mathrm{E}(X) = r\sum_{x=1}^{n} \frac{\binom{r-1}{x-1}\binom{N-r}{n-x}}{\frac{N}{n}\binom{N-1}{n-1}} = \frac{nr}{N}\sum_{x=1}^{n} \frac{\binom{r-1}{x-1}\binom{N-r}{n-x}}{\binom{N-1}{n-1}} = \frac{nr}{N}$$

Hypergeometrinen jakauma
Odotusarvon johto 3/3

- Kalvon 2/3 yhtälöketjun viimeinen yhtälö perustuu siihen, että

$$\sum_{x=1}^{n}\frac{\binom{r-1}{x-1}\binom{N-r}{n-x}}{\binom{N-1}{n-1}}=\sum_{x=0}^{n-1}\frac{\binom{r-1}{x}\binom{N-r}{n-1-x}}{\binom{N-1}{n-1}}=1$$

- Tämä seuraa siitä, että jälkimmäisessä summassa lasketaan yhteen *kaikki* hypergeometrisen jakauman

  HyperGeom(N – 1, r – 1, n – 1)

  pistetodennäköisyydet

$$f(x)=\frac{\binom{r-1}{x}\binom{N-r}{n-1-x}}{\binom{N-1}{n-1}}\,,\; x=0,1,2,\ldots,n-1$$

Odotusarvon ja varianssin ominaisuuksia 1/3

- Olkoon

$$X \sim \text{HyperGeom}(N, r, n).$$

- Hypergeometrisen jakauman odotusarvo on

$$\mathrm{E}(X) = n\frac{r}{N}$$

- Odotusarvo on *suoraan verrannollinen* sekä perusjoukon *S osajoukon B* (= otos) *alkioiden lukumäärään* (= n) että *tyypin A alkioiden lukumäärään perusjoukossa S* (= r).
- Odotusarvo on *kääntäen verrannollinen perusjoukon S alkioiden lukumäärään* (= N).

Odotusarvon ja varianssin ominaisuuksia 2/3

- Olkoon perusjoukon S alkioiden lukumäärä

$$n(S) = N$$

- Olkoon joukon $A \subset S$ alkioiden lukumäärä

$$n(A) = r$$

- Poimitaan perusjoukosta S *otannalla ilman takaisinpanoa* osajoukko B, jonka alkioiden lukumäärä on

$$n(B) = n$$

- Tällöin *diskreetti satunnaismuuttuja*

$$X = \text{Osajoukkoon } B \text{ tulleiden } A\text{:n alkioiden lukumäärä}$$

noudattaa hypergeometrista jakaumaa parametreilla N, r, n:

$$X \sim \mathrm{HyperGeom}(N, r, n)$$

Hypergeometrinen jakauma

Odotusarvon ja varianssin ominaisuuksia 3/3

- Todennäköisyys poimia *yksi* alkio joukosta *A* on

$$\Pr(A) = \frac{n(A)}{n(S)} = \frac{r}{N} = p$$

- Hypergeometrisen jakauman *odotusarvo* voidaan kirjoittaa todennäköisyyden *p* avulla muotoon

$$\mathrm{E}(X) = np$$

- Hypergeometrisen jakauman *varianssi* voidaan kirjoittaa todennäköisyyden *p* avulla muotoon

$$\mathrm{D}^2(X) = np(1-p)\left(\frac{N-n}{N-1}\right)$$

Hypergeometrinen jakauma
Pistetodennäköisyysfunktion kuvaaja

- Kuva oikealla esittää hypergeometrisen jakauman

  HyperGeom(100, 12, 20)

  pistetodennäköisyysfunktiota

$$f(x)=\frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$$

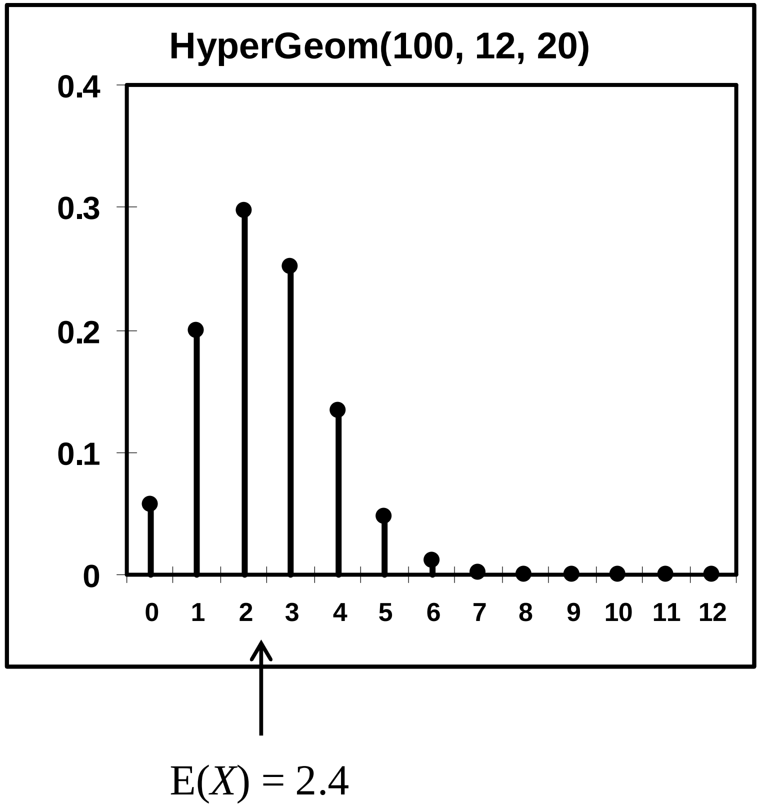
$$N=100\,,\ r=12\,,\ n=20$$

  pisteissä

  x = 0, 1, 2, … , 12

- Jakauman *odotusarvo*:

$$\mathrm{E}(X)=n\frac{r}{N}=2.4$$



Hypergeometrinen jakauma ja otanta ilman takaisinpanoa 1/5

- Olkoon perusjoukon S alkioiden lukumäärä

$$n(S) = N$$

- *Poimitaan* perusjoukosta S *satunnaisesti osajoukko* B, jonka alkioiden lukumäärä on

$$n(B) = n$$

käyttämällä poiminnassa *otantaa ilman takaisinpanoa*.

Hypergeometrinen jakauma ja otanta ilman takaisinpanoa 2/5

- **Otanta ilman takaisinpanoa**:
  - (i) Perusjoukosta S poimitaan alkiot osajoukkoon B yksi kerrallaan *arpomalla*.
  - (ii) Poimittuja alkioita **ei palauteta** takaisin perusjoukkoon S.
  - (iii) Kun otokseen poimitaan alkiota k, $k = 1, 2, \ldots , n$ *jokaisella perusjoukossa* S *jäljellä olevalla alkiolla on sama todennäköisyys*

$$1/(N - k + 1)$$

    tulla poimituksi osajoukkoon B.
- Osajoukko B muodostaa **yksinkertaisen satunnais-otoksen** perusjoukosta S.

Hypergeometrinen jakauma ja otanta ilman takaisinpanoa 3/5

- Otannassa ilman takaisinpanoa arvonta voidaan toteuttaa seuraavalla tavalla:
  (1) Pannaan *uurnaan* jokaista perusjoukon *S* alkiota vastaava arpalippu.
  (2) *Sekoitetaan* arvat huolellisesti.
  (3) *Nostetaan* uurnasta arpalippu, jota vastaava alkio valitaan otokseen *B*.
  (4) **Ei palauteta** nostettua arpalippua uurnaan.
  (5) Palataan vaiheeseen (3), kunnes haluttu otoskoko *n* on saavutettu.

Hypergeometrinen jakauma ja otanta ilman takaisinpanoa 4/5

- Huomautuksia otannasta ilman takaisinpanoa:

  (i) Perusjoukon *S* alkion todennäköisyys tulla valituksi otokseen *muuttuu* poiminnan aikana.

  (ii) Jokaisella perusjoukon *S* samankokoisella *osajoukolla* on kuitenkin *sama* todennäköisyys tulla valituksi otokseksi.

  (iii) Sama perusjoukon *S* alkio voi tulla valituksi *vain kerran* otokseen.

Hypergeometrinen jakauma ja otanta ilman takaisinpanoa 5/5

- Olkoon A perusjoukon S osajoukko, jonka alkioiden lukumäärä on

$$n(A) = r$$

- Todennäköisyys poimia *yksi* alkio joukosta A on

$$\Pr(A) = \frac{n(A)}{n(S)} = \frac{r}{N} = p$$

- Otannassa takaisinpanolla otokseen, jonka koko on n, poimittujen A-tyyppisten alkioiden lukumäärä X on diskreetti satunnaismuuttuja, joka noudattaa *hyper-geometristä jakaumaa parametreilla* N, r, n:

$$X \sim \text{HyperGeom}(N, r, n)$$

Hypergeometrinen jakauma *vs* binomijakauma 1/3

- Hypergeometrisen jakauman todennäköisyydet ovat lähellä binomitodennäköisyyksiä, jos *otantasuhde*

$$\frac{n}{N} \approx 0$$

- Otantasuhde ≈ 0, jos otoskoko n on *pieni* perusjoukon kokoon N nähden.

Hypergeometrinen jakauma *vs* binomijakauma 2/3

- Olkoon

$$X \sim \mathrm{HyperGeom}(N, r, n)$$

- Merkitään $p = r/N$, jolloin $r = Np$.
- Siten

$$X \sim \mathrm{HyperGeom}(N, Np, n)$$

- Annetaan $N \to +\infty$.
- Tällöin hypergeometrisen jakauman HyperGeom(N, Np, n) pistetodennäköisyydet *lähestyvät* binomijakauman Bin(n, p) pistetodennäköisyyksiä:

$$\lim_{N\to+\infty} f_{\mathrm{HyperGeom}(N,Np,n)}(x) = f_{\mathrm{Bin}(n,p)}(x)\ ,\ x = 0,1,2,\ldots,n$$

Hypergeometrinen jakauma *vs* binomijakauma 3/3

- Hypergeometrisen jakauman ja binomijakauman yhteys näkyy myös siinä, että jakaumilla on *sama* odotusarvo ja varianssit *eroavat vain multiplikatiivisella tekijällä*

$$\frac{N-n}{N-1}$$

jota sanotaan *äärellisen perusjoukon korjaustekijäksi*.

- Korjaustekijä vaikuttaa hypergeometrisen jakauman varianssiin sitä *vähemmän* mitä *pienempi* on otantasuhde n/N:

$$\frac{N-n}{N-1} \approx 1 \text{ , jos } \frac{n}{N} \approx 0$$

Otanta takaisinpanolla *vs* otanta ilman takaisinpanoa

- *Binomijakauma* muodostaa todennäköisyysmallin otannalle *takaisinpanolla*.
- *Hypergeometrinen* jakauma muodostaa todennäköisyysmallin otannalle *ilman takaisinpanoa*.
- *Ero* otannan takaisinpanolla ja otannan ilman takaisinpanoa välillä *on merkityksetön*, jos *otantasuhde* n/N on *pieni* tai perusjoukko on *ääretön*.
- Käytännössä otanta tehdään lähes aina *ilman takaisinpanoa*, mutta *laskutoimituksissa* käytetään silti usein kaavoja, jotka perustuvat otantaan *takaisinpanolla*.
- Edellä esitetyn mukaan tästä johtuva *virhe on* kuitenkin yleensä *merkityksetön*.

Diskreettejä jakaumia

Diskreetti tasainen jakauma
Bernoulli-jakauma
Binomijakauma
Geometrinen jakauma
Negatiivinen binomijakauma
Hypergeometrinen jakauma
>> Poisson-jakauma

Poisson-jakauma ja sen pistetodennäköisyysfunktio 1/3

- Toistetaan *samaa satunnaiskoetta*.
- Oletetaan, että toistot ovat toisistaan *riippumattomia*.
- Tarkastellaan jonkin *tapahtuman* A sattumista toistojen aikana.
- Oletetaan, että A-tapahtumien *keskimääräinen lukumäärä aika*- tai *tilavuusyksikköä* kohden on λ.
- Määritellään *diskreetti satunnaismuuttuja* X:

X = Tapahtuman A esiintymisten lukumäärä aika- tai tilavuusyksikköä kohden

Poisson-jakauma ja sen pistetodennäköisyysfunktio 2/3

- Jos tietyt oletukset pätevät (ks. >), satunnaismuuttujan *X* **pistetodennäköisyysfunktio** on muotoa

$$f(x) = \Pr(X = x) = \frac{e^{-\lambda}\lambda^{x}}{x!}, \lambda > 0$$

$$x = 0, 1, 2, \ldots$$

- Sanomme, että satunnaismuuttuja *X* noudattaa **Poisson-jakaumaa parametrinaan λ.**
- Merkintä:

$$X \sim \text{Poisson}(\lambda)$$

Poisson-jakauma ja sen pistetodennäköisyysfunktio 3/3

- Huomautus:

  Funktio $f(x)$ määrittelee todennäköisyysjakauman, koska *eksponenttifunktion määritelmän* mukaan

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

Pistetodennäköisyysfunktion johto 1/3

- Tarkastellaan jonkin tapahtuman A sattumista saman satunnaiskokeen toistojen aikana.
- Oletukset:

  (1) Toistot ovat toisistaan *riippumattomia*.

  (2) Pr(Yksi tapahtuma A *lyhyellä aikavälillä* dt) = vdt

  (3) Aikaväli on dt on *niin lyhyt*, että todennäköisyys

  Pr(k kpl tapahtumia A aikavälillä dt, $k > 1$)

  on häviävän pieni eli kertaluokkaa $o(t)$.
- Merkitään:

  $f(x; t)$ = Pr(x kpl tapahtumia A aikavälillä $[0, t]$)

Pistetodennäköisyysfunktion johto 2/3

- Oletusten (1)-(3) pätiessä aikavälillä

  $$[0, t + dt]$$

  voi sattua x kpl tapahtumia A kahdella *toisensa poissulkevalla tavalla* (tn = todennäköisyys):

  (1) x kpl tapahtumia A ajanhetkeen t mennessä; $tn = f(x; t)$
  Ei tapahtumia A aikavälillä dt; $tn = 1 - \nu dt$
  Lisäksi nämä ovat tapahtumina toisistaan *riippumattomia*.

  (2) $(x - 1)$ kpl tapahtumia A ajanhetkeen t mennessä; $tn = f(x - 1; t)$
  Yksi tapahtuma A aikavälillä dt; $tn = \nu dt$
  Lisäksi nämä ovat tapahtumina toisistaan *riippumattomia*.

Poisson-jakauma
Pistetodennäköisyysfunktion johto 3/3

- *Riippumattomien tapahtumien tulosäännön* ja *toisensa poissulkevien tapahtumien yhteenlaskusäännön* mukaan
$$f(x;\, t + dt) = f(x;\, t)(1 - \nu dt) + f(x - 1;\, t)\nu dt$$
- Järjestämällä termit uudelleen saadaan *erotusosamäärä*
$$\frac{f(x;t+dt) - f(x;t)}{dt} = \nu\left[f(x-1;t) - f(x;t)\right]$$
- Antamalla $dt \to 0$, saadaan (x:n suhteen) *differenssiyhtälö*
$$\frac{df(x;t)}{dt} = \nu\left[f(x-1;t) - f(x;t)\right]$$
- Voidaan osoittaa, että tämän differenssiyhtälön *ratkaisu* on muotoa
$$f(x) = \frac{e^{-\nu t}(\nu t)^x}{x!}\,,\ x = 0, 1, 2, \ldots$$
- Merkitsemällä $\nu t = \lambda$ saadaan *Poisson-jakauman pistetodennäköisyys-funktio*.

Poisson-jakauma

Odotusarvo, varianssi ja standardipoikkeama

- Olkoon

$$X \sim \text{Poisson}(\lambda)$$

- **Odotusarvo**:

$$\text{E}(X) = \lambda$$

- **Varianssi** ja **standardipoikkeama**:

$$\text{Var}(X) = \text{D}^2(X) = \lambda$$

$$\text{D}(X) = \sqrt{\lambda}$$

Poisson-jakauma

Odotusarvon johto

- Olkoon

$$X \sim \text{Poisson}(\lambda)$$

- Siten

$$\begin{aligned}
\mathrm{E}(X) = \sum_{x=0}^{\infty} x f(x) = \sum_{x=0}^{\infty} x \frac{e^{-\lambda}\lambda^{x}}{x!} &= e^{-\lambda} \sum_{x=1}^{\infty} x \frac{\lambda^{x}}{x!} \\
&= \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} \\
&= \lambda e^{-\lambda} e^{\lambda} \\
&= \lambda
\end{aligned}$$

Poisson-jakauma
Pistetodennäköisyysfunktion kuvaaja

- Kuva oikealla esittää Poisson-jakauman

  Poisson(5)

  pistetodennäköisyysfunktiota

  $$f(x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

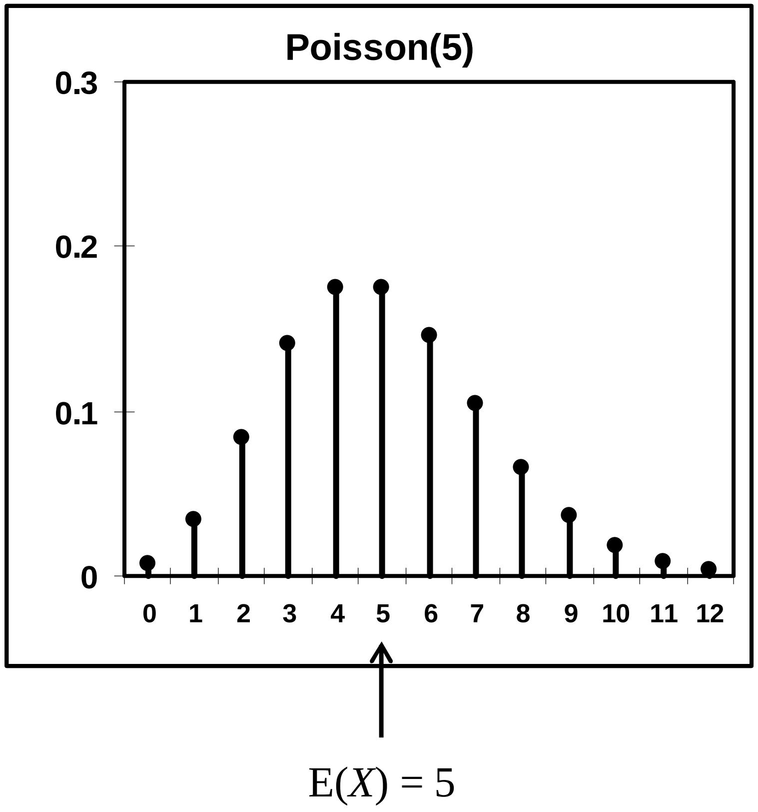
  $$\lambda = 5$$

  pisteissä

  $$x = 0, 1, 2, \dots , 12$$

- Jakauman *odotusarvo*:

  $$\mathrm{E}(x) = \lambda = 5$$



Poisson-jakaumaa noudattavien satunnaismuuttujien summan jakauma 1/2

- Olkoot

$$X_1, X_2, \ldots , X_k$$

riippumattomia satunnaismuuttujia, jotka noudattavat *Poisson-jakaumia* parametrein $\lambda_1 , \lambda_2 , \ldots , \lambda_k$:

$$X_1, X_2, \ldots , X_k \perp$$

$$X_i \sim \text{Poisson}(\lambda_i) \ , \ i = 1, 2, \ldots , k$$

- Tällöin satunnaismuuttujien $X_1, X_2, \ldots , X_k$ **summa**

$$Y = X_1 + X_2 + \cdots + X_k$$

noudattaa Poisson-jakaumaa parametrilla $\lambda_1 + \lambda_2 + \cdots + \lambda_k$:

$$Y \sim \text{Poisson}(\lambda_1 + \lambda_2 + \cdots + \lambda_k)$$

Poisson-jakaumaa noudattavien satunnaismuuttujien summan jakauma 2/2

- Tulos perustellaan luvussa **Satunnaismuuttujien muunnosten jakaumat.**
- Huomautuksia:
  - Jokaisella Poisson-jakaumalla saa olla *eri* parametri.
  - Tarkoitamme *satunnaismuuttujien riippumattomuudella* sitä, että yhdenkään satunnaismuuttujan saamat arvot eivät riipu siitä, mitä arvoja muut satunnaismuuttujat saavat; käsite täsmennetään luvussa **Moniulotteiset satunnaismuuttujat ja jakaumat.**

Binomijakauma ja Poisson-jakauma 1/3

- Binomitodennäköisyydet ovat lähellä Poisson-todennäköisyyksiä, jos n on suuri ja p on pieni.
- Siten Poisson-jakauma kuvaa *harvinaisten tapahtumien todennäköisyyksiä pitkissä toistokoesarjoissa*.

Binomijakauma ja Poisson-jakauma 2/3

- Olkoon

$$X \sim \mathrm{Bin}(n, p)$$

- Olkoon $p = \lambda/n$, jolloin $\lambda = np$.
- Annetaan $n \to +\infty$ ja $p \to 0$ niin, että $np = \lambda$.
- Tällöin binomijakauman $\mathrm{Bin}(n, p)$ pistetodennäköisyydet *lähestyvät* Poisson-jakauman $\mathrm{Poisson}(\lambda)$ pistetodennäköisyyksiä:

$$\lim_{\substack{n\to\infty \\ p\to 0 \\ np=\lambda}} f_{\mathrm{Bin}(n,p)}(x) = f_{\mathrm{Poisson}(\lambda)}(x)\ ,\ x = 0, 1, 2, \ldots$$

- Huomautus:

Ehto $np = \lambda$ voidaan korvata *lievemmällä* ehdolla $np \to \lambda$.

Binomijakauma ja Poisson-jakauma 3/3

- Poisson-jakauman ja binomijakauman välinen yhteys näkyy myös siinä, että jakaumien odotusarvot ovat lähellä toisiaan, jos *n* on *suuri* ja *p* on *pieni*:

$$\mathrm{E}(X) = \mu = \lambda \approx np$$

- Tällöin myös jakaumien varianssit ovat lähellä toisiaan:

$$\mathrm{D}^2(X) = \mu = \lambda = np \approx npq$$

koska

$$p \approx 0 \Rightarrow q = 1 - p \approx 1$$

Binomijakauma ja Poisson-jakauma: Todistus 1/3

- Olkoon X ~ Bin(n, p).
- Tällöin

$$f_X(x) = \binom{n}{x} p^x q^{n-x}, \; q = 1 - p \,, \; x = 0, 1, 2, \ldots, n$$

- Oletetaan, että

$$n \to +\infty$$

ja samaan aikaan

$$p \to 0$$

niin, että

$$np = \lambda$$

jossa $\lambda > 0$ on *vakio*.

Binomijakauma ja Poisson-jakauma: Todistus 2/3

- Ottamalla huomioon, että $np = \lambda$ ja $q = 1 - p$, voimme kirjoittaa

$$
\begin{aligned}
f_X(x) &= \binom{n}{x} p^x q^{n-x} \\
&= \frac{n(n-1)(n-2)\cdots(n-x+1)}{n^x x!}(np)^x(1-p)^{n-x} \\
&= \frac{n(n-1)(n-2)\cdots(n-x+1)}{n^x}\cdot\frac{\lambda^x}{x!}\cdot\frac{(1-p)^n}{(1-p)^x} \\
&= 1\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{x-1}{n}\right)\cdot\frac{\lambda^x}{x!}\cdot\frac{(1-p)^n}{(1-p)^x}
\end{aligned}
$$

Binomijakauma ja Poisson-jakauma: Todistus 3/3

- Kirjoitetaan

$$(1-p)^n = [(1-p)^{-1/p}]^{-np} = [(1-p)^{-1/p}]^{-\lambda}$$

- Luvun e määritelmän mukaan

(1) $$\lim_{p\to 0}[(1-p)^{-1/p}]^{-\lambda} = e^{-\lambda}$$

- Lisäksi pätee:

(2) $$\lim_{n\to\infty}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{x-1}{n}\right)=1$$

(3) $$\lim_{p\to 0}(1-p)^x = 1$$

- Yhdistämällä tulokset (1), (2) ja (3) saadaan haluttu lopputulos:

$$\lim_{\substack{n\to\infty \\ p\to 0 \\ np=\lambda}} f_X(x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

Poisson-prosessi

- Tarkastellaan jonkin tapahtuman A sattumista *jatkuvalla* aikavälillä, jonka pituus on t aikayksikköä.
- Määritellään *diskreetti satunnaismuuttuja* X:

  $X =$ Niiden tapahtumien lukumäärä, jotka sattuvat aikavälillä $[0, t]$

- Sopivin oletuksin (ks. edellä) satunnaismuuttuja X noudattaa **Poisson-jakaumaa parametrinaan** $\boldsymbol{\nu t}$:

  $X \sim \text{Poisson}(\nu t)$

- Parametri νt kuvaa *tapahtumaintensiteettiä* eli *tapahtumien keskimääräistä lukumäärää aikavälillä, jonka pituus on t aikayksikköä*.

Poisson-prosessi ja eksponenttijakauma

- Olkoon

$$X \sim \text{Poisson}(\nu t)$$

- Määritellään *jatkuva satunnaismuuttuja* Y:

$$\begin{aligned} Y &= \text{Ensimmäisen tapahtuman sattumisaika} \\ &= \text{Tapahtumien väliaika} \end{aligned}$$

- Satunnaismuuttuja Y noudattaa **eksponenttijakaumaa parametrinaan** $\boldsymbol{\nu}$.

  Ks. tarkemmin lukua **Jatkuvia jakaumia**.

Poisson-prosessi: Esimerkki

- Tarkastellaan *radioaktiivista hajoamista*.
- Olkoon satunnaismuuttuja

  X = aikavälillä $[0, t]$ hajoavien atomien lukumäärä
- Tällöin

  $$X \sim \mathrm{Poisson}(\nu t)$$

  jossa ν on *alkuainekohtainen* parametri, joka kuvaa *keskimäärin aikayksikköä kohden hajoavien atomien lukumäärää*.