Todennäköisyyslaskenta sivuaineopiskelijoille

- **Tentit:** 4.11.2013 ja 2.12.2013. Loput kaksi tenttiä (vuonna 2014) ilmoitetaan myöhemmin.
- Tentissä on 4 tehtävää á 8 pistettä, aikaa 4 tuntia.
- Arvostelu 0 – 5.
- Tenttiin on ilmoittauduttava vähintään viikko ennen tenttiä nettiopsun kautta.
- Tentissä saa olla mukana kynä, kumi, matematiikan kaavakokoelma, todennäköisyyslaskennan kaavakokoelma (päivätty 2010) ja laskin, joka ei kykene symboliseen eikä graafiseen laskentaan.
- Muista laittaa tenttipaperiin myös opiskelijanumerosi.
- Kaikki kussille osallistuneet ovat saaneet tenttioikeuden. Demotehtäviä tekemällä on voinut saada hyvityspisteitä tenttiin.
- Hyvityspisteet ja tenttioikeus säilyvät tämän lukuvuoden.

Klassinen todennäköisyys

Merkitään Ω:lla kaikkien alkeistapausten joukkoa eli *otosavaruutta* ja A:lla jotain alkeistapausten joukkoa eli *tapausta*. *Klassinen todennäköisyys* tapahtumalle A määritellään kaavalla

$$P(A) = \frac{|A|}{|\Omega|},$$

missä $|A|$=suotuisten alkeistapausten lukumäärä ja $|\Omega|$=kaikkien alkeistapausten lukumäärä.

Esimerkki: Korttipakasta (52 korttia) nostetaan umpimähkään yksi kortti. Millä todennäköisyydellä saadaan a) kuningas; b) hertta?

Ratkaisu: Nyt siis $|\Omega| = 52$.
a) Olkoon tapahtuma A= “kortti on kuningas”, jolloin $|A| = 4$.
$\Rightarrow \quad P(A) = |A|/|\Omega| = 4/52 = 1/13 \approx 0.0769$

b) Olkoon tapahtuma B= “kortti on hertta”, jolloin $|B| = 13$.
$\Rightarrow \quad P(B) = 13/52 = 1/4 = 0.25.$

Komplementti

Olkoon A^c tapauksen A *komplementtitapaus*, eli se sisältää ne alkeistapaukset, jotka eivät kuulu tapaukseen A. Klassiselle todennäköisyydelle pätee

$$
\begin{aligned}
&0 \leq P(A) \leq 1, \\
&P(A) = 1 - P(A^c).
\end{aligned}
$$

Esimerkki: Korttipakasta (52 korttia) nostetaan umpimähkään yksi kortti. Millä todennäköisyydellä kortti ei ole kuningas?

Ratkaisu: Olkoon tapahtuma A= "kortti on kuningas". Edellisestä esimerkistä tiedetään, että $P(A) = 1/13 \approx 0.0769$. Nyt

$$P(\text{"kortti ei ole kuningas"}) = P(A^c) = 1 - P(A) = 1 - 1/13 = 12/13 \approx 0.9231.$$

Joukko-oppi & todennäköisyys

Joukko	**Merkitys**	**Todennäköisyys**
A	tapahtuma A esiintyy	$P(A) = \|A\|/\|\Omega\|$
A^c	tapahtuma A ei esiinny	$P(A^c) = 1 - P(A)$
Ω	otosavaruus, varma tapaus	$P(\Omega) = 1$
$\emptyset$	mahdoton tapaus	$P(\emptyset) = 0$
$A \cap B$	A ja B esiintyvät	$P(A \cap B) = \|A \cap B\|/\|\Omega\|$
$A \cup B$	A tai B esiintyy (tai molemmat)	$P(A \cup B) =$ $P(A) + P(B) - P(A \cap B)$
$A - B$	A esiintyy mutta B ei	$P(A - B) =$ $P(A) - P(A \cap B)$
$A \subset B$	jos A esiintyy, niin myös B esiintyy	$P(A) \leq P(B)$

de Morganin kaavat:

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

Joukko-oppi & todennäköisyys jatkuu

Esimerkki: Korttipakasta (52 korttia) nostetaan umpimähkään yksi kortti. Millä todennäköisyydellä a) kortti on herttakuningas; b) kortti on hertta tai kuningas; c) kortti ei ole hertta eikä kuningas? d) kortti on kuningas mutta ei hertta; e) kortti on jokeri?

Ratkaisu: Olkoon tapahtuma A= “kortti on kuningas” ja B= “kortti on hertta”. Edellisestä esimerkistä tiedetään, että $P(A) = 1/13 \approx 0.0769$ ja $P(B) = 1/4 = 0.25$. Nyt

a) $P(A \cap B) = 1/52 \approx 0.0192$

b) $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 1/13 + 1/4 - 1/52 = 16/52 \approx 0.3077$

c) $P(A^c \cap B^c) = 1 - P((A^c \cap B^c)^c) = 1 - P(A \cup B) = 1 - 16/52 = 36/52 \approx 0.6923$

d) $P(A - B) = P(A) - P(A \cap B) = 1/13 - 1/52 = 3/52 \approx 0.0577$

e) $P(\emptyset) = 0$

Tuloperiaate, permutaatiot ja kombinaatiot

Klassisen tn:n laskemisessa ongelmaksi voi muodostua kaikkien mahdollisten alkeistapausten määrittäminen. Seuraavana tarkastellaan muutamaa apuneuvoa:

Tuloperiaate: Jos jokin operaatio on mahdollista suorittaa k eri vaiheessa ja kussakin vaiheessa on n_i, $i = 1, 2, \ldots, k$ eri valinnan mahdollisuutta, niin eri vaihtoehtoja on

$$\Pi_{i=1}^{k} n_i = n_1 \cdot n_2 \cdot \ldots \cdot n_k$$

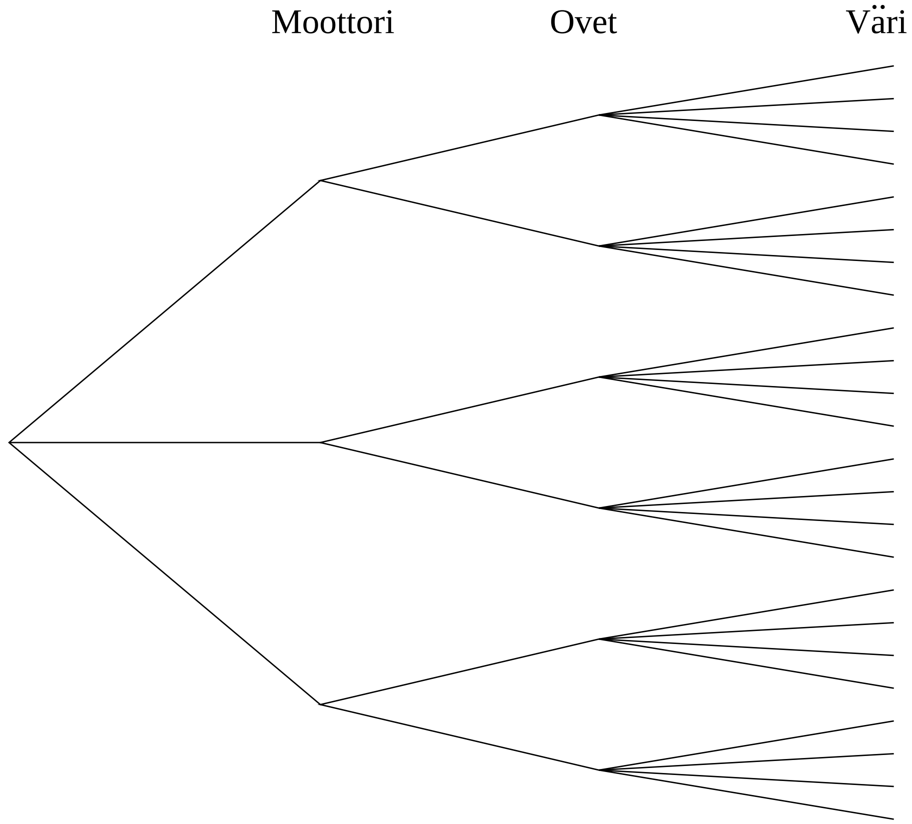
kappaletta.

Jos $n_i = n$ kaikilla i, niin tapoja on n^k kappaletta.

Tuloperiaate, permutaatiot ja kombinaatiot (jatkuu)

Esimerkki: Autotehdas valmistaa autoja kolmella eri moottorivaihtoehdolla, kaksi- tai neliovisena ja neljällä eri värivaihtoehdolla. Montako erilaista vaihtoehtoa tehtaalla on tarjota asiakkaalle?

Ratkaisu: Ajatellaan, että asiakas valitsee haluamansa mallin vaiheittain eri vaihtoehdoista: ensin moottorin koko ($n_1 = 3$), sitten ovien määrä ($n_2 = 2$)ja lopuksi väri ($n_3 = 4$). Eri mahdollisuuksista voidaan piirtää seuraava puurakenne:



Jokainen polku kuvaa aina yhtä mallia. Eri malleja on siis kaikkiaan $n_1 \cdot n_2 \cdot n_3 = 3 \cdot 2 \cdot 4 = 24$ kpl.

Tuloperiaate, permutaatiot ja kombinaatiot (jatkuu)

Otanta: Joukosta voidaan ottaa alkioita eli suorittaa *otanta* kahdella tavalla. Otanta tehdään *palauttamatta*, jos joukosta otetaan pois yksi alkio kerrallaan. Otanta tehdään *palauttaen*, jos jokaisen alkion oton jälkeen alkio palautetaan joukkoon.

Järjestetty joukko: Joukko on *järjestetty*, jos alkioiden järjestyksen muuttuessa myös joukko muuttuu. Toisin sanoen, on siis $(a, b, c) \neq (b, a, c)$.

Järjestämätön joukko: Joukko on *järjestäytymätön*, jos alkioiden järjestyksellä ei ole väliä; on siis $\{a, b, c\} = \{b, a, c\}$.

k **-permutaatio:** k -permutaatio muodostetaan ottamalla joukosta k:n alkion otos (palauttaen tai palauttamatta) ja muodostamalla siitä järjestetty joukko.

k **-kombinaatio:** k -kombinaatio muodostetaan ottamalla joukosta k:n alkion otos (palauttaen tai palauttamatta) ja muodostamalla siitä järjestämätön joukko.

Tuloperiaate, permutaatiot ja kombinaatiot (jatkuu)

Järjestetty otanta palauttamatta: Joukosta, jossa on n erilaista alkiota, voidaan ottaa k:n alkion otos palauttamatta $n(n-1)(n-2)\cdot\ldots\cdot(n-k+1)$ eli $n!/(n-k)!$ eri tavalla (k -permutaatioiden lkm).

Jos joukossa on n erilaista alkiota, niin alkiot voidaan järjestää jonoon $n!$ eri tavalla.

Järjestetty otanta palauttaen: Joukosta, jossa on n erilaista alkiota, voidaan ottaa k:n alkion otos palauttaen n^k eri tavalla (k -permutaatioiden lkm).

Järjestämätön otanta palauttamatta: Joukosta, jossa on n erilaista alkiota, voidaan ottaa k:n alkion otos palauttamatta

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

eri tavalla (k -kombinaatioiden lkm).

Tuloperiaate, permutaatiot ja kombinaatiot (jatkuu)

	Otos palauttamatta	Otos palauttaen
Järjestetty otos	$\frac{n!}{(n-k)!}$	n^k
Järjestämätön otos	$\binom{n}{k} = \frac{n!}{k!(n-k)!}$	$\binom{n+k-1}{k}$

Esimerkki: 25 oppilaan luokka päätti lähettää rehtorin puheille nelihenkisen lähetystön. Montako erilaista lähetystöä luokasta voidaan koota?

Ratkaisu: Erilaisten lähetystöjen määrä on

$$\binom{25}{4} = \frac{25!}{4!(21)!} = \frac{25 \cdot 24 \cdot 23 \cdot 22}{4 \cdot 3 \cdot 2 \cdot 1} = 12650$$

kappaletta.

Esimerkki: Kuinka monella tavalla 7 ihmistä voidaan mennä jonoon?

Ratkaisu: Tapoja on $7! = 5040$ kappaletta.

Riippumattomuus ja toisensa poissulkevuus

Toisensa poissulkevuus: Tapaukset A ja B ovat *toisensa poissulkevat*, jos

$$A \cap B = \emptyset.$$

Riippumattomuus: Tapaukset A ja B ovat *riippumattomat*, jos

$$P(A \cap B) = P(A)P(B).$$

Jos tapaukset eivät ole riippumattomat, ne ovat *riippuvat*.

Esimerkki: Korttipakasta (52 korttia) nostetaan umpimähkään yksi kortti. Olkoon tapahtuma A= "kortti on kuningas" ja B= "kortti on hertta". Ovatko tapahtumat riippumattomat?

Ratkaisu: Aiemmista esimerkeistä tiedetään, että $P(A) = 1/13$, $P(B) = 1/4$ ja $P(A \cap B) = 1/52$. Nyt $P(A)P(B) = 1/13 \cdot 1/4 = 1/52 = P(A \cap B)$, joten tapaukset A ja B ovat riippumattomat.

Todennäköisyyksien laskeminen

Seuraa luentomonisteen luvun 1.4.4 (s. 15–18) ohjeita. Myös tärkeimmät todennäköisyyksiä koskevat kaavat löytyvät näiltä sivuilta.

Esimerkki: Tietoverkossa lähetetään testisignaali alkupisteestä S loppupisteeseen F. Reititin ohjaa signaalin kuormituksesta riippuen reitille R_1, R_2, R_3 tai R_4 vastaavilla todennäköisyyksillä 12.5%, 50.0%, 25% ja 12.5%. Reitillä R_1 tapahtuu koodausvirhe 2%:n, reitillä R_2 5%:n, reitillä R_3 3% ja reitillä R_4 1% todennäköisyydellä. Millä todennäköisyydellä tietoverkon läpi kulkeneessa signaalissa on koodausvirhe?

Ratkaisu:

1) Annetaan tapauksille helposti muistettavat *symbolit*:

Merkitään

$A =$ "signaalissa on koodausvirhe",

$R_i =$ "signaali kulkee reittiä R_i pitkin", $i = 1, \ldots, 4$.

2) Kirjoitetaan *annetut todennäköisyydet* symbolien avulla:

$$P(R_1) = 0.125, \quad P(R_2) = 0.5, \quad P(R_3) = 0.25, \quad P(R_4) = 0.125,$$
$$P(A|R_1) = 0.02, \quad P(A|R_2) = 0.05, \quad P(A|R_3) = 0.03, \quad P(A|R_4) = 0.01.$$

Todennäköisyyksien laskeminen (jatkuu)

3) ja 4) Kirjoita ja laske kysytty todennäköisyys symbolien avulla, käytä vain todennäköisyyslaskennan *tunnettuja tuloksia*:

Käytetään *kokonaistodennäköisyyslausetta* (R_i:t muodostavat Ω:n partition)

$$\begin{aligned} P(A) &= \sum_{i=1}^{4} P(A|R_i)P(R_i) \\ &= P(A|R_1)P(R_1) + P(A|R_2)P(R_2) + P(A|R_3)P(R_3) + P(A|R_4)P(R_4). \end{aligned}$$

5) ja 6) Sijoita saatuun lausekkeeseen annetut todennäköisyydet, laske ja sievennä (anna mahdollisuuksien mukaan myös tarkka arvo):

$$\begin{aligned} P(A) &= \sum_{i=1}^{4} P(A|R_i)P(R_i) \\ &= 0.02 \cdot 0.125 + 0.05 \cdot 0.50 + 0.03 \cdot 0.25 + 0.01 \cdot 0.125 \\ &= 0.03625 \approx 3.6\%. \end{aligned}$$

Todennäköisyyksien laskeminen (jatkuu)

Esimerkki: Edellisen esimerkin signaalissa todettiin koodausvirhe. Millä todennäköisyydellä signaali on kulkenut reittiä R_2?

Ratkaisu: Merkinnät ja todennäköisyydet kuten edellä. Käytetään *Bayesin lausetta*:

$$P(R_2|A) = \frac{P(R_2)P(A|R_2)}{P(A)} = \frac{0.50 \cdot 0.05}{0.03625} = 0.6897 \approx 69.0\%.$$

Satunnaismuuttujat ja jakaumat

Satunnaismuuttuja: Satunnaismuuttuja X otosavaruudessa Ω on kuvaus $X : \Omega \to \mathbb{R}$. Satunnaismuuttuja siis liittää kuhunkin alkeistapaukseen ω_i jonkin reaaliluvun $X(\omega_i)$.

Jakaumafunktio: Satunnaismuuttujan X *jakaumafunktio* on

$$F(x) = P(X \leq x), \quad x \in \mathbb{R}.$$

Lause: Jakaumafunktiolla $F(x)$ on seuraavat ominaisuudet:

- $0 \leq F(x) \leq 1 \quad \forall x \in \mathbb{R}$,
- $x_1 \leq x_2 \Rightarrow F(x_1) \leq F(x_2)$, eli $F(x)$ on ei-vähenevä,
- $lim_{x \to -\infty} F(x) = 0$,
- $lim_{x \to \infty} F(x) = 1$.

Diskreetti Satunnaismuuttuja

Diskreetti satunnaismuuttuja: Satunnaismuuttuja X on *diskreetti*, jos se voi saada vain äärellisen tai numeroituvasti äärettömän määrän arvoja.

Todennäköisyysfunktio: Diskreetin satunnaismuuttujan X *todennäköisyysfunktio* on

$$p(k) = P(X = k), \quad k \in \mathcal{K},$$

missä $\mathcal{K}$ on diskreetin satunnaismuuttujan saamien arvojen joukko.

Todennäköisyysfunktiolla $p(k)$ on seuraavat ominaisuudet:

- $0 \leq p(k) \leq 1$,
- $\sum_{k \in \mathcal{K}} p(k) = 1$.

Todennäköisyysfunktio $p(k)$ ja jakaumafunktio $F(x)$ vastaavat toisiaan seuraavasti:

- $F(x) = \sum_{k \leq x} p(k)$,
- $p(k) = F(k) - F(k-1)$.

Jakaumafunktion $F(x)$ avulla voidaan helposti laskea puoliavoimen välin $]a, b]$ todennäköisyyksiä:

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

Diskreetti satunnaismuuttuja (jatkuu)

Esimerkki: Laatikossa on neljä sinistä ja kuusi valkoista palloa. Nostetaan kolme palloa palauttamatta. Merkitään satunnaismuuttujalla X sinisten pallojen lukumäärää. Määritä X:n todennäköisyys- ja jakaumafunktiot ja laske $P(1 \leq X \leq 2)$.

Ratkaisu: Satunnaismuuttujan X arvojoukko on $\mathcal{K} = \{0, 1, 2, 3\}$. Alkeistapauksia eli kolmen pallon osajoukkoja on

$$\binom{10}{3} = 120$$

kappaletta. Todennäköisyysfunktion $p(k) = P(X = k)$ määrittämiseksi lasketaan todennäköisyydet kaikilla $k \in \mathcal{K}$.

$$P(X = 0) = \frac{\binom{6}{3}}{120} = \frac{1}{6} \approx 16.7\%, \qquad P(X = 1) = \frac{\binom{4}{1}\binom{6}{2}}{120} = \frac{1}{2} \approx 50.0\%,$$

$$P(X = 2) = \frac{\binom{4}{2}\binom{6}{1}}{120} = \frac{3}{10} \approx 30.0\%, \qquad P(X = 3) = \frac{\binom{4}{3}}{120} = \frac{1}{30} \approx 3.3\%.$$

Diskreetti satunnaismuuttuja (jatkuu)

Näin ollen

$$p(k) = P(X = k) = \begin{cases} 1/6, & \text{kun } k = 0, \\ 1/2, & \text{kun } k = 1, \\ 3/10, & \text{kun } k = 2, \\ 1/30, & \text{kun } k = 3. \end{cases}$$

Jakaumafunktio $F(x)$ saadaan summaamalla todennäköisyysfunktion arvot:

$$F(x) = \sum_{k \leq x} p(k) = \begin{cases} 0, & \text{kun } x < 0, \\ 1/6, & \text{kun } 0 \leq x < 1, \\ 2/3, & \text{kun } 1 \leq x < 2, \\ 29/30, & \text{kun } 2 \leq x < 3, \\ 30/30, & \text{kun } 3 \leq x. \end{cases}$$

Diskreetti satunnaismuuttuja (jatkuu)

Kysytty todennäköisyys $P(1 \leq X \leq 2)$ voidaan laskea joko todennäköisyysfunktiota tai jakaumafunktiota hyödyntäen. Todennäköisyysfunktiota käytettäessä saadaan yhtälö

$$P(1 \leq X \leq 2) = P(X = 1) + P(X = 2) = 1/2 + 3/10 = 4/5 = 80\%.$$

Jakaumafunktion avulla puolestaan voidaan laskea puoliavoimen välin todennäköisyyksiä suoraan. Näin ollen

$$P(1 \leq X \leq 2) = P(0 < X \leq 2) = F(2) - F(0) = 29/30 - 1/6 = 4/5.$$

Diskreetti satunnaismuuttuja (jatkuu)

Odotusarvo: Diskreetin satunnaismuuttujan X *odotusarvo* on

$$E(X) = \sum_{k \in \mathcal{K}} kp(k),$$

jos $\sum_{k \in \mathcal{K}} |k|p(k) < \infty$.

Jos X saa arvoja $0, 1, 2, \ldots$, niin odotusarvo voidaan laskea myös kaavalla $E(X) = \sum_{k=0}^{\infty} [1 - F(k)]$. X:n odotusarvoa merkitään usein kirjaimelle μ.

Jos X on diskreetti satunnaismuuttuja ja g jokin funktio, niin myös $g(X)$ on satunnaismuuttuja ja $E[g(X)] = \sum_{k \in \mathcal{K}} g(k)p(k)$. Esimerkiksi $E[X^2] = \sum_{k \in \mathcal{K}} k^2 p(k)$.

Varianssi: Diskreetin satunnaismuuttujan X *varianssi* on

$$Var(X) = E\left[(X - \mu)^2\right].$$

Varianssi voidaan laskea myös kaavoilla

$$\begin{aligned} Var(X) &= E(X^2) - \mu^2 \quad \text{ja} \\ Var(X) &= E\left[(X(X-1)\right] + \mu - \mu^2. \end{aligned}$$

Tärkeimmät diskreetit jakaumat

Jakaumien todennäköisyydet sekä odotusarvojen ja varianssin laskukaavat löytyvät kaavakokoelmasta.

- **tasainen jakauma** $X \sim DU(n)$:
  - X:llä on n yhtä todennäköistä arvoa $1, 2, \ldots, n$.
- **binomijakauma** $X \sim Bin(n, p)$:
  - toistetaan koetta, joka onnistuu tn:llä p, riippumattomasti n kertaa, X on onnistuneitten kokeiden lkm.
  - X:llä on likimain $Po(np)$-jakauma jos n on suuri ja p pieni.
- **Poisson-jakauma** $X \sim Po(\lambda)$:
  - X noudattaa Poisson-jakaumaa, jos $P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}$, $k = 0, 1, 2, \ldots$.
  - Poisson-jakauma sopii esim. seuraavien ongelmien tilastolliseen mallintamiseen: Liikennepisteen sivuttaa keskimäärin 300 autoa tunnissa. Millä tn:llä autojen lkm ylittää 1000 kpl tunnissa? Ilman siitepölypitoisuus on keskimäärin 450 hiukkasta/cm^3. Millä tn:llä satunnaisesti otetussa näytteessä on alle 200 hiukkasta/cm^3?

Tärkeimmät diskreetit jakaumat (jatkuu)

- **Bernoulli-jakauma** $X \sim Ber(p)$:
  - koe onnistuu tn:llä p. X saa arvon 1, jos koe onnistuu, ja 0 muutoin.
- **geometrinen jakauma** $X \sim Geom(p)$:
  - koetta, joka onnistuu tn:llä p, toistetaan riippumattomasti, kunnes se onnistuu ensimmäisen kerran. X on tarvittavien kokeiden lkm.
  - geometrinen jakauma on ainoa positiivinen kokonaislukuarvoinen jakauma, jolla on muistamattomuusominaisuus.
- **hypergeometrinen jakauma** $X \sim Hyp(N, M, n)$:
  - laatikossa on N palloa, joista M on mustia ja $N - M$ valkoisia. Laatikosta otetaan n palloa palauttamatta. X on mustien pallojen lkm.
  - X:llä on likimain $Bin(n, M/N)$-jakauma jos N on suuri ja n/N pieni.

Jatkuvat satunnaismuuttujat ja jakaumat

Jatkuva satunnaismuuttuja: Satunnaismuuttuja X on *jatkuva*, jos on olemassa sellainen ei-negatiivinen funktio $f(x)$, että

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

missä $a, b \in \mathbb{R}$ ja $a \leq b$. Funktio $f(x)$ on satunnaismuuttujan X *tiheysfunktio*.

Lause: Jatkuvan satunnaismuuttujan tiheys- ja jakaumafunktioilla on seuraavat ominaisuudet:

- $f(x) \geq 0$,
- $\int_{-\infty}^{\infty} f(x)dx = 1$,
- $F(x) = \int_{-\infty}^{\boldsymbol{x}} f(t)dt$,
- $F(x)$ on jatkuva kaikilla x,
- $f(x) = F'(x)$ kaikilla x, joilla $f(x)$ on jatkuva.

Jatkuva jakauma eroaa diskreetistä mm. siinä, että aina on

$$P(X = a) = P(a \leq X \leq a) = \int_a^a f(x)dx = 0.$$

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

Jatkuvat satunnaismuuttujat ja jakaumat (jatkuu)

Esimerkki: Henkilö lupaa soittaa ystävälleen klo 12:00 – 14:00 välisenä aikana. Mitkä ovat soittoajan tiheys- ja jakaumafunktiot? Millä todennäköisyydellä henkilö soittaa ennen klo 12:40? Entäpä välillä 13:05 – 13:17?

Ratkaisu: Soittoaika X on nyt jatkuva satunnaismuuttuja. Tehdään kellonajoille muunnos $12:00 \to 0$ ja $14:00 \to 120$ (yksikkönä siis minuutti). Soittoajan X tiheysfunktio on

$$f(x) = \begin{cases} 0, & \text{kun } x < 0 \text{ tai } x > 120, \\ 1/120, & \text{kun } 0 \leq x \leq 120. \end{cases}$$

Jakaumafunktio $F(x)$ saadaan integroimalla tiheysfunktio

$$F(x) = \begin{cases} 0, & \text{kun } x < 0, \\ \frac{1}{120}x, & \text{kun } 0 \leq x \leq 120, \\ 1, & \text{kun } x > 120. \end{cases}$$

Todennäköisyys, että henkilö soittaa ennen 12:40 on

$$P(X < 40) = \int_0^{40} \frac{1}{120} dx = \frac{40}{120} = \frac{1}{3} = F(40).$$

Todennäköisyys, että soitto tulee klo 13:05 – 13:17 on

$$P(65 \leq X \leq 77) = \int_{65}^{77} \frac{1}{120} dx = \frac{1}{10} = F(77) - F(65).$$

Jatkuvat satunnaismuuttujat ja jakaumat (jatkuu)

Odotusarvo: Jatkuvan satunnaismuuttujan X *odotusarvo* on

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

jos $\int_{-\infty}^{\infty} |x| f(x) dx < \infty$.

Varianssi: Jatkuvan satunnaismuuttujan X *varianssi* on

$$Var(X) = E\left[(X - \mu)^2\right].$$

Varianssi voidaan laskea myös kaavalla $Var(X) = E(X^2) - \mu^2$.

Vastaavasti kuin diskreetissä tapauksessa voidaan määrittää satunnaismuuttuja $g(X)$ ja sen odotusarvo $E(g(X)) = \int_{-\infty}^{\infty} g(x) f(x) dx$. Näin ollen varianssi voidaan laskea kaavoilla

$$Var(X) = \begin{cases} \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx, \\ \int_{-\infty}^{\infty} x^2 f(x) dx. \end{cases}$$

Tärkeimmät jatkuvat jakaumat

Jakaumien todennäköisyydet sekä odotusarvojen ja varianssin laskukaavat löytyvät kaavakokoelmasta.

- **standardoitu normaalijakauma** $X \sim N(0,1)$
  - X:n tiheysfunktio on muotoa $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$.
  - jakauma-funktion $\Phi(x)$ arvot luetaan normaalijakaumataulukosta.
  - jakauma on symmetrinen, joten $\Phi(-x) = 1 - \Phi(x)$.

- **yleinen normaalijakauma** $X \sim N(\mu, \sigma^2)$
  - X:n tiheysfunktio on muotoa $\phi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$.
  - Huom! tällä kurssilla jakauman toisena parametrinä annetaan varianssi σ^2 ei hajontaa σ.
  - jos varianssi (tai hajonta) on pieni, on jakauma suppea ja korkea; jos taas varianssi (hajonta) on suuri, jakauma on matala ja leveä.
  - jos $X \sim N(\mu, \sigma^2)$, niin $\frac{X-\mu}{\sigma} \sim N(0,1)$; näin ollen jakauman arvot voidaan selvittää standardoidun normaalijakauman avulla.
  - tyypillinen $N(\mu, \sigma^2)$-jakautunut satunnaismuuttuja on jonkin suureen annostelu- tai mittausvirhe.

Tärkeimmät jatkuvat jakaumat (jatkuu)

Esimerkki: Pakattaessa tavaraa rasioihin on tavaran paino X normaalijakautunut odotusarvon ollessa 1200g ja hajonnan ollessa 3g. Millä todennäköisyydellä satunnaisesti valitussa rasiassa tavaran paino on välillä $[1199, 1201]$ (yksikkönä gramma)? Entä mikä on todennäköisyys, että paino alittaa 1195g?

Ratkaisu: $\mu = 1200$, $\sigma^2 = 3^2$ ja $X \sim N(1200, 3^2) \Rightarrow F(x) = \Phi\left(\frac{x-1200}{3}\right)$.

$$\begin{aligned}
P(1199 \leq X \leq 1201) &= F(1201) - F(1199) \\
&= \Phi\left(\frac{1201-1200}{3}\right) - \Phi\left(\frac{1199-1200}{3}\right) \\
&= \Phi\left(\frac{1}{3}\right) - \Phi\left(-\frac{1}{3}\right) \\
&= 2\Phi\left(\frac{1}{3}\right) - 1 \approx 2 \cdot 0.6306 - 1 = 0.2612 \approx 26.1\%
\end{aligned}$$

$$\begin{aligned}
P(X < 1195) = F(1195) &= \Phi\left(\frac{1195-1200}{3}\right) = \Phi\left(-\frac{5}{3}\right) = 1 - \Phi\left(\frac{5}{3}\right) \\
&\approx 1 - 0.9520 = 0.048 = 4.8\%
\end{aligned}$$

Tärkeimmät jatkuvat jakaumat (jatkuu)

- **tasainen jakauma** $X \sim U(a, b)$:
  - tiheysfunktio on tasainen eli vakio välillä $]a, b[$.

- **eksponenttijakauma** $X \sim Exp(\lambda)$:
  - tiheysfunktio on muotoa

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{kun } x > 0, \\ 0, & \text{muutoin.} \end{cases}$$

  - eksponenttijakaumalla on ns. muistamattomuusominaisuus:

$$P(X > t + h | X > t) = P(X > h)$$

  - tyypillinen $Exp(\lambda)$-jakautunut satunnaismuuttuja on kahden peräkkäisen tapahtuman välinen aika tai laitteen kestoikä.
  - eksponentti- ja Poisson jakauman välillä on yhteys: jos jokin tapahtuma sattuu eksponenttijakautunein aikavälein, parametrinä λ, ja aikavälit ovat riippumattomat, on aikavälillä t sattuvien tapahtumien lukumäärä Poisson-jakautunut parametrillä λt. Tällaista prosessia kutsutaan *Poissonin prosessiksi*.

Tärkeimmät jatkuvat jakaumat (jatkuu)

Esimerkki: Eräs laite tarvitsee huoltoa keskimäärin 3kk välein. Millä todennäköisyydellä huoltoväli on a) alle 3kk; b) kahdesta viiteen kuukauteen; c) yli 5 kk?

Ratkaisu: Oletetaan, että huoltoväli T on eksponentti jakautunut. Valitaan satunnaismuuttujan T odotusarvoksi 3kk. Eksponenttijakautumaa noudattavan satunnaismuuttujan odotusarvo $E(T) = 1/\lambda$, joten $\lambda = 1/3$ (yksikkönä 1/kk). Nyt

a) $P(T < 3) = \int_0^3 \frac{1}{3} e^{-\frac{1}{3}\cdot t} dt = F(3) = 1 - e^{-\frac{1}{3}\cdot 3} = 1 - e^{-1} = 0.6321$

b) $P(2 \leq T \leq 5) = F(5) - F(2) = 1 - e^{-\frac{1}{3}\cdot 5} - 1 + e^{-\frac{1}{3}\cdot 2} = e^{-\frac{2}{3}} - e^{-\frac{5}{3}} = 0.3245$

c) $P(T > 5) = 1 - P(T \leq 5) = 1 - F(5) = 1 - 1 + e^{-\frac{1}{3}\cdot 5} = e^{-\frac{5}{3}} = 0.1889.$

Keskeinen raja-arvolause

Keskeinen raja-arvolause: Olkoon satunnaismuuttujat X_1, X_2, $\ldots X_n$ riippumattomia ja samoin jakautuneita odotusarvona $E(X_i) = \mu$ ja varianssina $Var(X_i) = \sigma^2$. Tällöin niiden summa $S_n = \sum_{i=1}^{n} X_i$ ja keskiarvo $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ ovat satunnaismuuttujia siten, että suurilla n

- S_n:llä on likimain $N(n\mu, n\sigma^2)$ -jakauma ja
- $\bar{X}_n$:llä on likimain $N(\mu, \sigma^2/n)$ -jakauma.

Sanotaan, että S_n ja $\bar{X}_n$ ovat *asymptoottisesti normaalisti jakautuneet* ja merkitään

$$S_n \sim AsN(n\mu, n\sigma^2) \qquad \text{ja} \qquad \bar{X}_n \sim AsN(\mu, \sigma^2/n).$$

Esimerkki: Eräässä 300-sivuisessa kirjassa on keskimäärin yksi painovirhe viittä sivua kohti. Millä todennäköisyydellä kirjassa on yli 50 virhettä?

Ratkaisu: Olkoon X_i sivulla i olevien painovirheiden lukumäärä ja $X = \sum_{i=1}^{300} X_i$ koko kirjan painovirheiden lukumäärä. Nyt $X_i \sim Po(0.2)$ kaikilla i, joten keskeisen raja-arvolauseen mukaan $X \sim AsN(300 \cdot 0.2, 300 \cdot 0.2) = AsN(60, 60)$, koska $E(X_i) = Var(X_i) = \lambda = 0.2$. On siis

$$P(X > 50) = 1 - P(X \leq 50) \approx 1 - \Phi\left(\frac{50-60}{\sqrt(60)}\right) = \Phi(1.29) = 0.9015.$$

Keskeinen raja-arvolause (jatkuu)

Keskeinen raja-arvolause mahdollistaa jakaumien approksimoinnin ja normaalijakauman laajan käytön erilaisissa sovelluksissa. Erityisesti on voimassa:

Lause:

a) Olkoon $X \sim Bin(n, p)$. Parametrin n kasvaessa X:n jakauma lähenee normaalijakaumaa $N(np, npq)$, missä $q = 1 - p$, eli

$$X \sim AsN(np, npq),$$

jos n on suuri ($npq > 10$).

b) Olkoon $X \sim Po(\lambda)$. Parametrin λ kasvaessa X:n jakauma lähenee normaalijakaumaa $N(\lambda, \lambda)$ eli

$$X \sim AsN(\lambda, \lambda),$$

jos λ on suuri ($\lambda > 15$).

Jos diskreettiä jakaumaa approksimoidaan jatkuvalla jakaumalla approksimaation tarkkuutta voidaan hiukan parantaa käyttämällä ns. *jatkuvuuskorjausta*.

Keskeinen raja-arvolause (jatkuu)

Esimerkki: Eräässä 300-sivuisessa kirjassa on keskimäärin yksi painovirhe viittä sivua kohti. Koko kirjan painovirheiden lukumäärä X on $Po(60)$ -jakautunut. Millä todennäköisyydellä kirjassa on yli 50 virhettä?

Ratkaisu: Poisson-jakauman parametri $\lambda = 60 > 15$, joten jakaumaa voidaan approksimoida normaalijakaumalla. On siis $X \sim AsN(60, 60)$. Näin ollen

$$\begin{aligned}
P(X > 50) &= 1 - P(0 \leq X \leq 50) \\
&\approx 1 - \Phi\left(\frac{50.5 - 60}{\sqrt(60)}\right) + \Phi\left(\frac{-0.5 - 60}{\sqrt(60)}\right) \\
&= 1 - \Phi(-1.226) + \Phi(-7.811) \\
&= 1 - (1 - \Phi(1.226)) + 1 - \Phi(7.811) \\
&= 1 - 1 + \Phi(1.226)) + 1 - 1 \\
&\approx \Phi(1.23) = 0.8907.
\end{aligned}$$

Muita jatkuvia jakaumia

- t-**jakauma** $T \sim t(v)$:
  - tärkeä testauksiin liittyvä jakauma,
  - jakaumia on itseasiassa äärettömän monta (riippuu vapausasteista v),
  - symmetrinen, $E(T) = 0$ ja $Var(T) = v/(v-2)$.
  - suurilla v:n arvoilla t-jakauma muistuttaa suuresti standardi normaalijakaumaa,
  - t-jakauman kriittinen arvo $t_{v;\alpha}$ on arvo, jota suuremman arvon $t(v)$-jakautunut satunnaismuuttuja T saa todennäköisyydellä α. Ts.

$$P(T > t_{v;\alpha}) > \alpha,$$

  - t-jakauman kriittisiä arvoja $t_{v;\alpha}$ on taulukoitu eri v:n ja α:n arvoilla.

Parametrien estimointi

Parametrien estimoinnilla tarkoitetaan perusjoukon tunnuslukujen arvioiden eli *estimaattien* määrittämistä. Estimointia varten otetaan perusjoukosta otos ns. *satunnaisotos* $\{X_1, \ldots, X_n\}$ ja estimaatti lasketaan otoksesta *estimaattorin* eli estimoitavaa parametriä kuvaavan satunnaismuuttujan avulla.

Parametrin Θ hyvä estimaattori $\hat{\Theta}$ on

- *harhaton:* $E(\hat{\Theta}) = \Theta$,
- *tarkentuva:* $Var(\hat{\Theta}) \to 0$, kun otoskoko kasvaa.

Tavallisimpia estimoitavia parametrejä ovat odotusarvo μ ja varianssi σ^2.

Otoskeskiarvo ja otosvarianssi: Olkoon $\{X_1, \ldots, X_n\}$ satunnaisotos jakaumasta, jonka odotusarvo on μ ja varianssi σ^2. Satunnaisotoksen *otoskeskiarvo* lasketaan kaavalla

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

ja *otosvarianssi* on

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

Otoskeskiarvo $\bar{X}$ on odotusarvon μ ja otosvarianssi S^2 varianssin σ^2 harhaton ja tarkentuva estimaattori.

Parametrien estimointi (jatkuu)

Koska estimaattori on satunnaismuuttuja (esim. otoskeskiarvo $\bar{X}$ odotusarvon μ estimaattorina), sen arvo vaihtelee satunnaisotoksesta riippuen. Todelliselle parametrille voidaan kuitenkin tietyin edellytyksin laskea väli (a,b), jolla parametri sijaitsee tietyllä varmuudella.

Odotusarvon luottamusväli: Jos otos on normaalijakaumasta, niin odotusarvon μ $100 \cdot (1-\alpha)\%$ -luottamusväli on

$$P\left(\bar{X} - t_{n-1;\alpha/2}\frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{n-1;\alpha/2}\frac{S}{\sqrt{n}}\right) = 1 - \alpha.$$

Jos jakauma ei ole normaali, niin luottamusvälin kaava pätee likimäärin ja sitä paremmin, mitä suurempi on n.

Esimerkki: Oletetaan, että kun tapahtumasta saatiin 140 havaintoa, niin otoskeskiarvo oli $\bar{x} = 54.7$ ja otosvarianssi $s^2 = 90$. Millä välillä odotusarvo μ on 95% varmuudella?

Ratkaisu: Halutaan 95% luottamusväli, joten $\alpha = 0.05$. Tällöin $t_{n-1;\alpha/2} = t_{139;0.05/2} \approx t_{150;0.025} = 1.976$. Odotusarvon luottamusväli lasketaan kaavasta

$$P\left(54.7 - 1.976\frac{\sqrt{90}}{\sqrt{140}} < \mu < 54.7 + 1.976\frac{\sqrt{90}}{\sqrt{140}}\right) = 0.95$$

eli

$$P(53.12 < \mu < 56.28) = 0.95.$$

Lähteet:

- Heikki Ruskeepää: Luentomoniste, Turun yliopisto, 2010
- Antti Niemi: Todennäköisyyslaskennan ja tilastomatematiikan perusteet, Jyväskylän teknillinen oppilaitos, 1995.