5 Tärkeitä yksiulotteisia jakaumia

- Jakaumista löytyy lisätietoja ja kuvaajia Wikipediasta.
- Kirjallisuudessa käytetään useille näistä jakaumista monia erilaisia parametrointeja. Kussakin lähteessä käytetty parametrointi paljastuu tavallisesti vasta tutkimalla lähemmin siinä käytettyjä kaavoja.
- Eri lähteet käyttävät jakaumille eri tunnuksia.

5.1 Diskreettejä jakaumia

Kaikilla tämän jakson jakaumilla on se ominaisuus, että niitä noudattavat satunnaismuuttujat voivat saada vain kokonaislukuarvoja. Tämä on tyypillistä sovelluksissa.

- Binomijakauma
- Hypergeometrinen jakauma
- Geometrinen jakauma
- Negatiivinen binomijakauma
- Poissonin jakauma

5.1.1 Binomijakauma

Synty Onnistumisten lkm n-kertaisessa toistetussa Bernoullin kokeessa, kun onnistumistn on p.

Tunnus $X \sim \mathrm{Bin}(n, p)$, jossa $n \geq 1$ kokonaisluku (otoskokoparametri) ja $0 \leq p \leq 1$ (onnistumistn; todennäköisyysparametri).

Ptnf

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

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

Binomijakauma (jatkoa)

Odotusarvo, varianssi ja momenttiemäfunktio

$$EX = np, \quad \operatorname{var} X = np(1-p), \quad M(t) = (pe^t + 1 - p)^n.$$

Yhteyksiä $\mathrm{Bin}(1, p)$ on $\mathrm{Bernoulli}(p)$.

Yhteenlaskuominaisuus Jos $X \sim \mathrm{Bin}(m, p)$ ja $Y \sim \mathrm{Bin}(n, p)$ (sama onnistumistn) ja $X \perp\!\!\!\perp Y$, niin

$$X + Y \sim \mathrm{Bin}(n + m, p).$$

Yhteenlaskuominaisuuden voi perustella joko toistokoetulkinnan avulla tai momenttiemäfunktiota tarkastelemalla.

Esimerkki: binomijakauma ja poiminta takaisinpanolla

- Olkoon laatikossa N palloa, joista $0 \leq K \leq N$ on valkoista ja loput mustia. Laatikosta poimitaan umpimähkään n palloa takaisinpanolla.
- Tällöin nostettujen valkoisten pallojen lukumäärällä X on jakauma $\mathrm{Bin}(n, K/N)$.

5.1.2 Hypergeometrinen jakauma

Synty Laatikossa on N palloa, joista $0 \leq K \leq N$ on valkoista ja loput ovat mustia. Laatikosta nostetaan umpimähkäisesti $n \leq N$ palloa ilman takaisinpanoa.

Ptnf Jos X kertoo, kuinka moni nostetuista palloista on valkoinen, niin X:n ptnf on

$$f(x) = f(x \mid N, K, n) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}, \qquad x = 0, 1, \ldots, n.$$

Tämä on hypergeometrinen jakauma parametreilla N, K ja n.

Muista, että $\binom{m}{k} = 0$ ellei $0 \leq k \leq m$. Tämän takia $f(x \mid N, K, n) \neq 0$ vain, kun $0 \leq x \leq \min(n, K)$ ja $n - x \leq N - K$.

Hypergeometrinen jakauma: ptnf:n johto kombinatoriikan tuloperiaattella

- Ajattele, että pallot on numeroitu, mutta että ainoastaan pallon väristä ollaan kiinnostuneita.
- Symmetrisiä alkeistapauksia ovat N pallon n-osajoukot, joiden lukumäärä on $\binom{N}{n}$.
- Sellaisen osajoukon valinta, jossa on x valkoista ja $n - x$ mustaa palloa, voidaan ajatella tehtävän kahdessa vaiheessa.
  1. valitaan x:n valkoisen pallon osajoukko K valkoisesta pallosta: eri mahdollisuuksia on $\binom{K}{x}$
  2. valitaan $n - x$ mustan pallon osajoukko $N - K$ mustasta pallosta: eri mahdollisuuksia on $\binom{N-K}{n-x}$.
- $f(x)$ saadaan jakamalla suotuisten alkeistapauksien lukumäärä kaikkien alkeistapausten lukumäärällä.

Huomautuksia hypergeometrisesta jakaumasta

- Hypergeometrisen jakauman käsittely on hankalahkoa.
- Suurten populaatioiden tapauksessa (N suuri ja $n \ll N$) sitä mielellään approksimoidaan vastaavalla binomijakaumalla $\mathrm{Bin}(n, K/N)$, joka syntyisi otannasta takaisinpanolla.
- Odotusarvo ja varianssi ovat (vrt. binomijakauman $\mathrm{Bin}(n, K/N)$ kaavoihin)

$$EX = n\frac{K}{N}, \qquad \operatorname{var} X = n\frac{K}{N}\left(1 - \frac{K}{N}\right)\frac{N-n}{N-1}.$$

Nämä tulokset johdetaan esim. Tuomisen kirjassa.

Lisää huomautuksia

- Binomijakauma ja hypergeometrinen jakauma ovat lähisukulaisia.
- Hypergeometrisella jakaumalla ja seuraavaksi esitettävällä geometrisella jakaumalla ei ole keskenään mitään ilmeistä yhteyttä.
- Hypergeometrisen jakauman nimen alkuperä on hämärän peitossa, mutta se saattaa johtua jakauman tietyistä löyhistä yhteyksistä ns. hypergeometriseen funktioon, joka on eräs klassinen erikoisfunktio.

5.1.3 Geometrinen jakauma

Synty Toistetaan riippumattomasti Bernoullin koetta, jossa onnistumistn on $0 < p < 1$. Olkoon

$$X = \text{“epäonnistumisten lkm ennen ensimmäistä onnistumista”}.$$

Tällöin X:llä on geometrinen jakauma parametrilla p, eli $X \sim \mathrm{Geom}(p)$.

Geometrinen jakauma (jatkoa)

Ptnf

$$f(x) = f(x \mid p) = P(X = x) = p\,(1{-}p)^x, \quad x = 0, 1, 2, \ldots$$

Perustelu Epäonnistumisia ennen ensimmäistä onnistumista on $X = x$ kpl jos ja vain jos koetta vastaavat riippumattomat Bernoullin muuttujat $Y_1, \ldots, Y_x, Y_{x+1}$ saavat arvot

$$Y_1 = 0, Y_2 = 0, \ldots, Y_x = 0 \quad \text{ja} \quad Y_{x+1} = 1.$$

Riippumattomuuden nojalla tämän tapahtuman tn on $(1 - p)^x\, p$.

Geometrinen jakauma (jatkoa)

Tarkistus Se, että $f(x) = p\,(1-p)^x$ on ptnf voidaan tarkistaa geometrisen sarjan avulla,

$$\sum_{j=0}^{\infty} r^j = 1 + r + r^2 + \cdots = \frac{1}{1-r}, \qquad |r| < 1,$$

jonka seurauksena

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

Lisäksi jokainen termeistä $f(x) = p\,(1-p)^x$ on positiivinen.

Momenttiemäfunktio saadaan geometrisen sarjan avulla,

$$M(t) = \frac{p}{1-(1-p)\mathrm{e}^t}, \qquad t < -\ln(1-p).$$

Geometrinen jakauma (jatkoa)

Odotusarvo ja varianssi lasketaan helposti momenttiemäfunktiosta,

$$EX = \frac{1-p}{p}, \qquad \operatorname{var} X = \frac{1-p}{p^2}.$$

Huomautus Toisinaan geometrinen jakauma määritellään edellisesti poikkeavasti niin, että se on satunnaismuuttujan Y jakauma, jossa

$$Y = \text{"sen toiston järjestysnumero, jolla onnistutaan ensimmäisen kerran"}.$$

Tällöin $Y = X + 1$, ja $EY = 1/p$.

5.1.4 Negatiivinen binomijakauma

Synty Toistetaan riippumattomasti Bernoullin koetta, jossa onnistumistn on $0 < p < 1$. Olkoon $r > 0$ kokonaisluku, ja tarkastellaan sm:a

$$X = \text{"epäonnistumisten lkm, ennen kuin onnistutaan } r\text{:nnen kerran".}$$

Parametrit ovat p ja r; geometrinen jakauma saadaan, kun $r = 1$.

Negatiivisen binomijakauman ptnf:n johto

- Tapahtuma $X = x$ eli että on x epäonnistumista ennen r:ttä onnistumista on sama kuin että
  - $r - 1 + x$ ensimmäisessä toistossa epäonnistutaan x kertaa ja onnistutaan $r - 1$ kertaa ja
  - toistossa numero $r + x$ onnistutaan.

Koska toistot ovat riippumattomia, on

$$f(x) = f(x \mid r, p) = \binom{r + x - 1}{x} p^r (1 - p)^x, \qquad x = 0, 1, 2, \ldots$$

Tässä $p^r(1 - p)^x$ on sellaisen jonon onnistumisia ja epäonnistumisia tn, joka toteuttaa ylläolevat ehdot, ja binomikerroin kertoo tällaisten jonojen lukumäärän.

Huomautuksia negatiivisesta binomijakaumasta

- Geometrinen jakauma on negatiivisen binomijakauman erikoistapaus.
- Sen sijaan binomijakaumalla ja negatiivisella binomijakaumalla ei ole mitään ilmeistä yhteyttä keskenään.

Lisää huomautuksia negatiivisesta binomijakaumasta

- Negatiivinen binomijakauma voidaan yleistää tilanteeseen, jossa $r > 0$ ei ole kokonaisluku. Tällöin toistokoetulkinta menetetään.
- Negatiivista binomijakaumaa käytetään usein Poissonin jakauman sijasta lukumäärien mallina, jos aineistossa niillä näyttää olevan suurempi varianssi kuin mitä Poissonin jakaumalla olisi.
- Laskemme myöhemmin kurssilla negatiivisen binomijakauman odotusarvon ja varianssin, mutta täysin erilaisella tekniikalla kuin mikä selitetään tässä jaksossa luentomonisteessa.

5.1.5 Poissonin jakauma

Tunnus $X \sim \mathrm{Poi}(\theta)$, jossa $\theta > 0$.

Ptnf

$$f(x) = f(x \mid \theta) = \mathrm{e}^{-\theta} \, \frac{\theta^x}{x!}, \quad x = 0, 1, 2, \ldots$$

Tarkistus Tämä on ptnf, sillä kaikki arvot ovat ei-negatiivisia, ja eksponenttifunktion sarjakehitelmän mukaan

$$\mathrm{e}^u = \exp(u) = \sum_{j=0}^{\infty} \frac{u^j}{j!}, \qquad \forall u \in \mathbb{R}\,.$$

Mikäli $u > 0$, niin sarjan kaikki termit ovat positiivisia. Siksi luvut $\mathrm{e}^{-\theta} \, \theta^j / j!$ muodostavat pistetodennäköisyysfunktion, kun $j = 0, 1, 2, \ldots$.

Poissonin jakauma (jatkoa)

Momenttiemäfunktio saadaan eksponenttifunktion sarjakehitelmän avulla,

$$M(t) = \sum_{x=0}^{\infty} \mathrm{e}^{xt}\, \mathrm{e}^{-\theta}\, \frac{\theta^x}{x!} = \exp(\theta(\mathrm{e}^t - 1)).$$

Odotusarvo ja varianssi saadaan tästä helposti derivoimalla,

$$EX = \operatorname{var} X = \theta.$$

Yhteenlaskuominaisuus Jos $X \sim \mathrm{Poi}(\theta_1)$ ja $Y \sim \mathrm{Poi}(\theta_2)$, ja $X \perp\!\!\!\perp Y$, niin

$$M_{X+Y}(t) = M_X(t)\, M_Y(t) = \exp((\theta_1 + \theta_2)(\mathrm{e}^t - 1)),$$

joten $X + Y \sim \mathrm{Poi}(\theta_1 + \theta_2)$.

Poissonin prosessi: johdattelua

Poissonin prosessilla mallinnetaan diskreettejä tapahtumia jatkuvalla aika- tai pituusvälillä tai tasossa tai avaruudessa. Sillä saateetaisiin mallintaa esimerkiksi sitä,

- kuinka monta asiakasta palvelupisteeseen saapuu tietyllä aikavälillä,
- kuinka monta fotonia osuu tiettyyn filmin alueeseen,
- kuinka monta valkosolua löytyy tietystä veripisarasta.

Poissonin prosessi

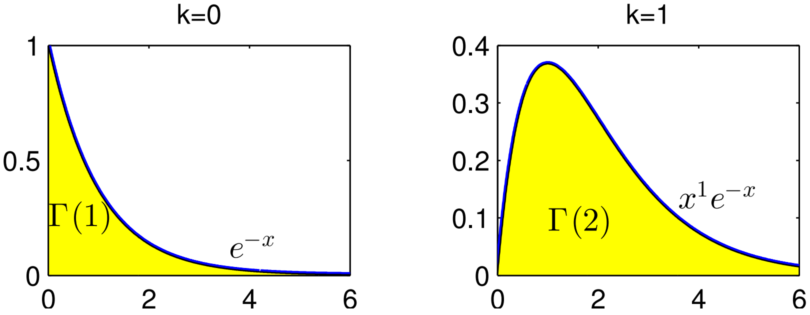
- Mikäli välin pituus (tai tasoalueen pinta-ala, avaruuden osan tilavuus) on s yksikköä, niin Poissonin prosessissa siinä havaittavien tapahtumien lukumäärällä on Poissonin jakauma $\mathrm{Poi}(s\lambda)$.
- Lisäksi erillisillä väleillä (tasoalueilla, avaruuden osilla) havaittavat lukumäärät ovat keskenään riippumattomia.
- Keskimääräinen tapahtumien lukumäärä yhtä (pituus-, pinta-ala- tai tilavuus-)yksikköä kohti on $s\lambda/s = \lambda$. Ts. $\lambda > 0$ eli Poissonin prosessin intensiteetti on tapahtumien odotusarvo yhtä (pituus-, pinta-ala-, tilavuus-)yksikköä kohti.

5.2 Gamma- ja beetafunktio

- Tutustumme seuraavaksi kahteen "erikoisfunktioon".
- "Erikoisfunktio" = funktio jota tarvitaan niin usein, että sille on annettu oma nimi (poislukien kaikkein tutuimmat, ns. "alkeisfunktiot" kuten sin ja exp)
- Tyypillisesti erikoisfunktioita tarvitaan erilaisten integraalien laskentaan.

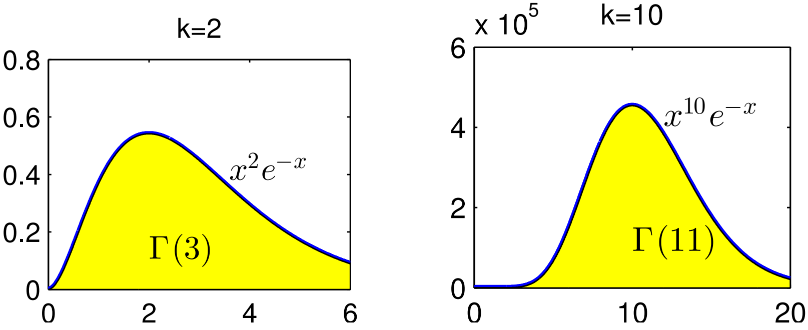
Johdattelua gammafunktioon

- Usein tarvitaan integraalia $\int_0^\infty x^k e^{-x}\,dx$.
- Tässä $k > -1$ on vakio, usein kokonaisluku.



Johdattelua gammafunktioon

- Usein tarvitaan integraalia $\int_0^\infty x^k e^{-x}\, dx$.
- Tässä $k > -1$ on vakio, usein kokonaisluku.



Gamma- ja beetafunktio

- Eulerin gammafunktio Γ määritellään positiivisilla argumenteilla ($t > 0$) integraalilla

$$\Gamma(t) = \int_0^\infty x^{t-1}\,\mathrm{e}^{-x}\,\mathrm{d}x. \tag{1}$$

- Jos $t \geq 1$, niin tämä integraali on epäoleellinen vain ylärjallaan, mutta jos $0 < t < 1$, niin integraali on epäoleellinen myös alarajalla, koska tällöin integrandi kasvaa rajatta, kun t lähestyy nollaa.
- Integraali (1) on äärellinen jos ja vain jos $t > 0$.

Gammafunktion $\Gamma(t) = \int_0^\infty x^{t-1}\,\mathrm{e}^{-x}\,\mathrm{d}x$ ominaisuuksia

- Osittaisintegroinnilla nähdään, että

$$\Gamma(t+1) = \int_0^\infty x^t\,\mathrm{e}^{-x}\,\mathrm{d}x = -\Big/_{\!0}^{\infty} x^t\,\mathrm{e}^{-x} + t\int_0^\infty x^{t-1}\,\mathrm{e}^{-x}\,\mathrm{d}x.$$

- Tämän takia

$$\Gamma(t+1) = t\,\Gamma(t), \qquad \text{kaikilla } t > 0. \tag{2}$$

- Koska lisäksi $\Gamma(1) = 1$, niin kokonaislukuargumenteilla n pätee

$$\Gamma(n) = (n-1)!, \qquad \text{kun } n = 1, 2, 3, \ldots$$

Tässä mielessä gammafunktio on kertoman yleistys.

Standardinormaalijakauman normalisointivakio

- Tarkastellaan integraalia

$$I = \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2}x^2}\,\mathrm{d}x.$$

- Ottamalla huomioon, että integrandi on parillinen funktio ja tekemällä sitten muuttujanvaihto $t = \frac{1}{2}x^2$, saadaan

$$I = 2\int_{0}^{\infty} \mathrm{e}^{-\frac{1}{2}x^2}\,\mathrm{d}x = 2\int_{0}^{\infty} \mathrm{e}^{-t}\,\frac{1}{\sqrt{2t}}\,\mathrm{d}t = \sqrt{2}\,\Gamma(\frac{1}{2}).$$

- Kohta nähdään, että $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.
- Tämän jälkeen on tarkistettu, että

$$\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}x^2}$$

on tiheysfunktio, standardinormaalijakauman tiheysfunktio.

Beetafunktion esitys gammafunktion avulla

- Luentomonisteessa näytetään napakoordinaattien ja muuttujanvaihtojen avulla, että

$$B(a,b) = \frac{\Gamma(a)\,\Gamma(b)}{\Gamma(a+b)},$$

jossa $B(a,b)$ on (Eulerin) beetafunktio, joka määritellään integraalina

$$B(a,b) = \int_0^1 u^{a-1}\,(1-u)^{b-1}\,\mathrm{d}u, \qquad a,b>0. \tag{3}$$

- Integraali (3) on äärellinen jos ja vain jos $a > 0$ ja $b > 0$.
- Erityisesti $B(\frac{1}{2},\frac{1}{2}) = (\Gamma(\frac{1}{2}))^2$.

$\Gamma(\frac{1}{2}) = \sqrt{\pi}$

- Määritelmän nojalla

$$B(\frac{1}{2}, \frac{1}{2}) = \int_0^1 u^{-\frac{1}{2}} (1-u)^{-\frac{1}{2}} \, \mathrm{d}u$$

- Sijoituksella $u = \sin^2\theta$ saadaan

$$B(\frac{1}{2}, \frac{1}{2}) = \int_0^{\frac{\pi}{2}} \frac{1}{\sqrt{\sin^2\theta}} \frac{1}{\sqrt{\cos^2\theta}} 2 \sin\theta \cos\theta \, \mathrm{d}\theta = \pi.$$

- Koska $\Gamma(\frac{1}{2}) = \sqrt{B(\frac{1}{2}, \frac{1}{2})}$, saadaan $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

5.3 Jatkuvia jakaumia

- Skaalaus ja siirto
- Tasajakauma
- Eksponenttijakauma
- Gammajakauma
- Beetajakauma
- Normaalijakauma
- Normaalijakaumasta johdettuja jakaumia:
  - χ^2-jakauma
  - t-jakauma
  - F-jakauma.

5.3.1 Skaalaus ja siirto

- Jos sm:lla Z on jatkuva jakauma tiheysfunktiolla f_0, ja sm X määritellään kaavalla

$$X = aZ + b, \qquad \text{missä} a > 0,$$

niin tiheysfunktion muuntokaavan nojalla sm:n X tf on

$$f_X(x \mid a, b) = \frac{1}{b} f_0 \left(\frac{x - b}{a} \right).$$

- Jos erityisesti $a = 1$ ja $b = 0$, niin X:n tf on f_0.
- Tällä tavoin saadaan perhe jakaumia, jossa parametria b kutsutaan sijaintiparametriksi (engl. *location parameter*) ja parametria $a > 0$ skaalaparametriksi (engl. *scale parameter*).

Skaalaus: skaalaparametri vai *rate*-parametri?

- Erityisesti jos $b = 0$, jolloin $X = aZ$, saadaan jakaumaperhe, jonka tiheysfunktiot ovat muotoa

$$\frac{1}{a} f_0(x/a), \qquad a > 0.$$

- Jotkin jakaumaperheet parametroidaan muodossa

$$\lambda f_0(\lambda x), \qquad \lambda > 0,$$

jossa f_0 on tf. Tällöin $a = 1/\lambda$ on skaalaparametri, ja parametria λ voidaan kutsua rate-parametriksi.

- Sana *rate* voidaan kääntää, asiayhteydestä riippuen, esim. sanoilla intensiteetti, vauhti, osuus, aste, taajuus jne.

Skaalaus ja siirto: Mitä momenteille tapahtuu?

- Olkoon $X = aZ + b$, missä $a > 0$.
- Koska odotusarvo on lineaarinen, on

$$EX = E(aZ + b) = a(EZ) + b.$$

- X:n k:s keskusmomentti on

$$\begin{aligned} E(X - EX)^k &= E[aZ + b - aEZ - b]^k \\ &= E[a(Z - EZ)]^k \\ &= a^k E(Z - EZ)^k, \end{aligned}$$

  siis a^k kertaa vastaava Z:n keskusmomentti.
- Erityisesti varianssi on $\operatorname{var} X = a^2 \operatorname{var} Z$.

Skaalaus ja siirtoparametrit momentteina

- Jos $EZ = 0$ ja $\operatorname{var} Z = 1$, niin muunnokselle $X = aZ + b$ saadaan

$$
\begin{aligned}
EX &= aEZ + b = 0 + b = b \\
\operatorname{var} X &= a^2 \operatorname{var} Z = a^2 \cdot 1 = a^2
\end{aligned}
$$

- Tällöin on sijainti- ja skaalaparametrille luontevaa käyttää kirjaimia μ ja σ.
- Varoitus: Näitä kirjaimia käytetään muunnosparametrien symboleina joskus muutenkin (siis vaikka olisi $EZ \neq 0$ tai $\operatorname{var} Z \neq 1$), mutta tällöin on varottava sekoittamasta niitä X:n momentteihin! (Ks. edellinen kalvo.)

5.3.2 Tasajakauma

- Jos $a < b$, niin välin (a, b) tasajakaumalla, $X \sim \mathrm{U}(a, b)$ on tiheysfunktio

$$f(x) = f(x \mid a, b) = \frac{1}{b-a} 1_{(a,b)}(x), \qquad x \in \mathbb{R} .$$

- Kaksi ensimmäistä momenttia lasketaan helposti integroimalla,

$$EX = \frac{1}{2}(a+b), \qquad EX^2 = \frac{1}{3}(a^2 + ab + b^2),$$

- Tästä nähdään, että

$$\operatorname{var} X = \frac{(b-a)^2}{12}.$$

Tasajakaumaperhe skaalauksella ja siirrolla

- Olkoon $Z \sim U(0,1)$ (ns. yksikkövälin tasajakauma).
- Mikä tahansa tasajakauma saadaan siitä skaalauksella ja siirrolla. Jos nimittäin

$$X = (b - a)Z + b,$$

  niin tf:n muuntokaavasta nähdään, että $X \sim U(a, b)$.
- Tätä voidaan hyödyntää mm. (pseudo)satunnaislukujen generoinnissa.
- Toinen sovellus on momenttien laskennassa: $EZ^k = 1/(k+1)$, josta saadaan helposti X:n momentit. Esim. tätä kautta voidaan laskea, että minkä tahansa tasajakauman vinous on olla.

5.3.3 Eksponenttijakauma

- X noudattaa eksponenttijakaumaa parametrilla $\lambda > 0$, eli $X \sim \mathrm{Exp}(\lambda)$, jos sen tf on

$$f(x) = f(x \mid \lambda) = \lambda \mathrm{e}^{-\lambda x}, \qquad x > 0.$$

  Tässä λ on *rate*-parametri.

- Jakauman odotusarvo ja varianssi ovat

$$EX = \frac{1}{\lambda}, \qquad \operatorname{var} X = \frac{1}{\lambda^2} = (EX)^2.$$

- Momenttiemäfunktio on

$$M(t) = \frac{\lambda}{\lambda - t}, \qquad t < \lambda.$$

- Joskus parametriksi valitaan λ:n sijasta odotusarvo, $\theta = 1/\lambda$. Tällöin varianssi on θ^2.

Eksponenttijakauman muistinmenetysominaisuus

- Eksponenttijakaumaa käytetään usein komponentin eliniän mallina. Tällöin tehdään se oletus, että komponentti ei kulu käytössä. Eksponenttijakaumalla on nimittäin ns. muistinmenetysominaisuus.
- Oletetaan, että komponentin elinikä $X \sim \mathrm{Exp}(\lambda)$.
- Lasketaan todennäköisyys, että elinikä on suurempi kuin $x + h$, jos se on suurempi kuin x, kun $x, h > 0$.

$$P(X > x + h \mid X > x) = \frac{P(X > x + h \text{ ja } X > x)}{P(X > x)} = \frac{P(X > x + h)}{P(X > x)}$$

$$= \frac{1 - F(x + h)}{1 - F(x)} = \frac{\mathrm{e}^{-\lambda(x+h)}}{\mathrm{e}^{-\lambda x}} = \mathrm{e}^{-\lambda h} = P(X > h).$$

Todennäköisyys ei riipu siitä, kuinka kauan komponenttia on jo käytetty.

5.3.4 Gammajakauma

- Usein tarvitaan jakaumaa, jonka tiheysfunktio on

$$g(x) = c \cdot x^{\alpha-1} e^{-x}, \qquad \text{kun } x > 0,$$

  missä $\alpha > 0$ on jokin vakio.

- Jotta saadaan tiheysfunktio, selvitetään normalisointivakio c. Se selviää gammafunktion määritelmän (1) perusteella: Jotta $1 = \int_0^\infty g(x)\mathrm{d}x = c \cdot \Gamma(\alpha)$, valitaan normalisointivakioksi

$$c = 1/\Gamma(\alpha).$$

- Kutsumme tätä gammajakaumaksi parametreilla α ja 1, ja merkitsemme sitä tunnuksella $\mathrm{Gam}(\alpha, 1)$.

Skaalattu gammajakauma

- Olkoon $Z \sim \mathrm{Gam}(\alpha, 1)$, ja $\alpha > 0$, $\lambda > 0$.
- Muodostetaan skaalaamalla uusi satunnaismuuttuja

$$X = Z/\lambda$$

jossa siis λ oli *rate*-parametri (skaala on $1/\lambda$).

- Sanomme nyt, että $X \sim \mathrm{Gam}(\alpha, \lambda)$ eli X:llä on gammajakauma parametrein α ja λ.
- α:aa kutsutaan jakauman muotoparametriksi.
- Helpolla laskulla saadaan X:n tiheysfunktio

$$f(x \mid \alpha, \lambda) = \lambda\, g(\lambda x) = \frac{\lambda^\alpha}{\Gamma(\alpha)}\, x^{\alpha-1}\, \mathrm{e}^{-\lambda x}, \qquad x > 0$$

jossa g on jakauman $\mathrm{Gam}(\alpha, 1)$ tiheysfunktio.

Integroidaan tilastotieteilijän tapaan

- Jakauman $\mathrm{Gam}(\alpha, \lambda)$ momenttiemäfunktio saadaan laskettua, kun huomataan, että siinä tarvittava integraali on erään toisen gammajakauman normalisointivakio,

$$
\begin{aligned}
M(t) = E\mathrm{e}^{Xt} &= \int_0^\infty \mathrm{e}^{xt} \, \frac{\lambda^\alpha}{\Gamma(\alpha)} \, x^{\alpha-1} \, \mathrm{e}^{-\lambda x} \, \mathrm{d}x \\
&= \frac{\lambda^\alpha}{\Gamma(\alpha)} \, \frac{\Gamma(\alpha)}{(\lambda - t)^\alpha} \int_0^\infty \frac{(\lambda - t)^\alpha}{\Gamma(\alpha)} \, x^{\alpha-1} \, \mathrm{e}^{-(\lambda - t)x} \, \mathrm{d}x \\
&= \left(\frac{\lambda}{\lambda - t} \right)^\alpha, \qquad t < \lambda.
\end{aligned}
$$

- Tässä laskussa integroimme kuten tilastotietielijä.
- Tunnistamme että integrandi on vakiota vaille tutun jakauman tiheysfunktio, jota integroidaan kantajansa yli. Tällöin osaamme suoraan kirjoittaa lausekkeen integraalin arvolle katsomalla ko. tiheysfunktion normalisointivakiota.

Gammajakauman $\mathrm{Gam}(\alpha, \lambda)$ ominaisuuksia

- Odotusarvo ja varianssi saadaan momenttiemäfunktiosta.

$$EX = \frac{\alpha}{\lambda}, \qquad \operatorname{var} X = \frac{\alpha}{\lambda^2}.$$

- Eräät usein käytetyt jakaumat ovat gammajakauman erikoistapauksia.
  - $\mathrm{Gam}(1, \lambda)$ on sama kuin $\mathrm{Exp}(\lambda)$.
  - $\mathrm{Gam}(n/2, 1/2)$ on sama kuin χ^2_n (khiin neliön jakauma n:llä vapausasteella).
- Yhteenlaskuominaisuus. Jos $X \perp\!\!\!\perp Y$ ja $X \sim \mathrm{Gam}(\alpha_1, \lambda)$ ja $Y \sim \mathrm{Gam}(\alpha_2, \lambda)$ (jälkimmäinen parametri sama), niin

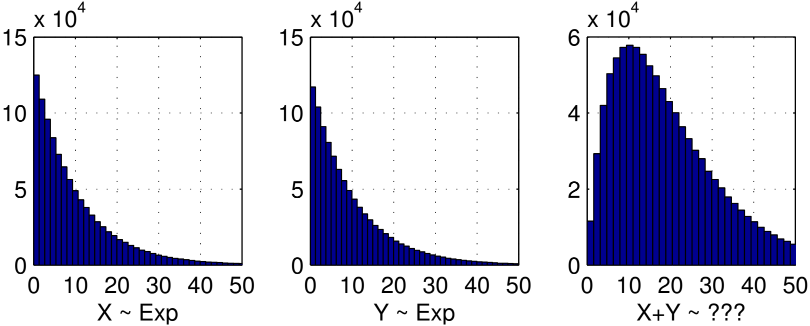
$$X + Y \sim \mathrm{Gam}(\alpha_1 + \alpha_2, \lambda).$$

Lisää gammajakauman ominaisuuksia

- Jos haluamme sellaisen gammajakauman, jolla on tietty odotusarvo ja varianssi, se onnistuu aina valitsemalla parametrit α ja λ sopivasti: koska
$$EX = \frac{\alpha}{\lambda}, \qquad \operatorname{var} X = \frac{\alpha}{\lambda^2}.$$
niin
$$\lambda = \frac{EX}{\operatorname{var} X}, \qquad \alpha = \lambda EX.$$
- Muotoparametri α säätelee jakauman muotoa, erityisesti sen huipun sijaintia.

Sovellusesimerkki: Eksponenttijakautuneiden summa

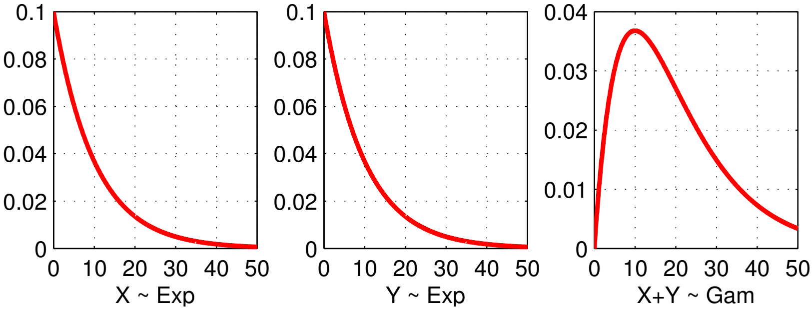
- Olkoot $X, Y \sim \mathrm{Exp}(\frac{1}{10})$ riippumattomia.
- Tutkitaan satunnaismuuttujaa $Z = X + Y$.
- Koska $EX = EY = 10$, on tietysti $EZ = 20$, mutta mikä on Z:n jakauma? Alla empiiriset histogrammit.



Sovellusesimerkki: Eksponenttijakautuneiden summa (2)

Koska

- eksponenttijakauma on gammajakauma: $X, Y \sim \mathrm{Gam}(1, \frac{1}{10})$
- ja gammajakaumalla on yhteenlaskuominaisuus (muotoparametrit lasketaan yhteen),
- niin $X + Y \sim \mathrm{Gam}(2, \frac{1}{10})$.



Varoitus gammajakauman parametroinneista

- Gammajakauma parametroidaan usein myös käyttämällä parametreja α ja skaalaparametria $1/\lambda$.
- Kummassakin parametroinnissa jälkimmäistä parametria (*rate*-parametri tai skaalaparametri) on tapana merkitä symbolilla β.
- Se, kummasta parametroinnista on kyse selviää, mikäli jakauman tiheysfunktion tai odotusarvon kaava annetaan.

5.3.5 Beetajakauma

- Usein tarvitaan jakaumaa, jonka tiheysfunktio on muotoa

$$f(x) = f(x \mid \alpha, \beta) = c \cdot x^{\alpha-1} (1-x)^{\beta-1}, \qquad 0 < x < 1,$$

  missä $\alpha, \beta > 0$ ovat vakioita.
- Tiheysfunktio sinänsä on varsin yksinkertaista muotoa, kunhan normalisointivakio c saadaan selville.
- Normalisointivakio selviää beetafunktion avulla: jotta $1 = \int_0^1 f(x) = c \cdot B(\alpha, \beta)$, valitaan $c = 1/B(\alpha, \beta)$.
- Tämä on beetajakauma parametrein α ja β, eli $\mathrm{Be}(\alpha, \beta)$.
- Yksikkövälin tasajakauma $\mathrm{U}(0, 1)$ on sama kuin $\mathrm{Be}(1, 1)$.

Beetajakauman ominaisuuksia

- Beetajakauman momentit on helppo laskea, sillä ne saadaan suoraan erään toisen beetajakauman normalisointivakiosta (ts. integroimalla kuten tilastotieteilijä).
- Jos $r > 0$, on
$$EX^r = \frac{B(\alpha + r, \beta)}{B(\alpha, \beta)}.$$
- Erityisesti
$$EX = \frac{\alpha}{\alpha + \beta}, \qquad EX^2 = \frac{\alpha(\alpha + 1)}{(\alpha + \beta)(\alpha + \beta + 1)}.$$
- Tästä
$$\operatorname{var} X = EX^2 - (EX)^2 = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}.$$

Beetajakauman sovelluksia

- Beetajakaumaa käytetään yleisesti ns. Bayes-päättelyssä tuntemattoman parametrin jakaumana: esim. epäreilu kolikonheitto tuntemattomalla kruunatn:llä T. Jos T:llä on ennen heittoja tasajakauma, niin heittojen jälkeen saatu T:n *ehdollinen jakauma* on eräs beetajakauma.
  $\longrightarrow$ Kevään 2015 kurssi Bayes-päättely
- Toinen beetajakauman sovellus on liittyy järjestettyihin otoksiin. Jos $X_1, \ldots, X_n \sim U(0,1)$ ovat riippumattomia, ja P on otoksen pienin ja S suurin luku, niillä on kummallakin eräs beetajakauma (vrt. Tuominen: Todennäköisyyslaskenta, kappale 3.8). Yleisemmin, otoksen k:nneksi suurimmalla luvulla on eräs beetajakauma.

5.3.6 Normaalijakauma

- Standardinormaalijakaumaa $N(0,1)$ noudattavan sm:n $Z \sim N(0,1)$ tf on

$$f_Z(z) = \phi(z) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2} z^2), \qquad z \in \mathbb{R}\,. \tag{4}$$

- Standardinormaalijakuman tiheysfunktiolle käytetään yleisesti merkintää ϕ ja sen kertymäfunktiolle merkintää Φ.
- Tämä on tiheysfunktio sen takia, että funktio on ei-negatiivinen (peräti aidosti positiivinen) ja sen integraali koko reaaliakselin yli on yksi. (Normalisointivakio laskettiin edellä gamma- ja beetafunktioiden avulla.)

Standardinormaalijakauman $N(0, 1)$ momenttiemäfunktio

$$M_Z(t) = E\mathrm{e}^{tZ} = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}z^2 + tz)\,\mathrm{d}z.$$

- Täydennetään eksponenttifunktion argumentti neliöksi,

$$-\frac{1}{2}z^2 + tz = -\frac{1}{2}(z-t)^2 + \frac{1}{2}t^2,$$

minkä jälkeen (sijoituksella $u = z - t$) nähdään, että

$$M_Z(t) = \exp(\frac{1}{2}t^2) \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right) \mathrm{d}u = \exp(\frac{1}{2}t^2).$$

- Laskemalla kumulanttiemäfunktion $K_Z(t) = \frac{1}{2}t^2$ derivaatat nähdään, että

$$EZ = 0, \qquad \operatorname{var} Z = 1.$$

Normaalijakauma $N(\mu, \sigma^2)$

- Sm X noudattaa normaalijakaumaa parametrein μ, σ^2, eli $X \sim N(\mu, \sigma^2)$, jos se voidaan esittää muodossa

$$X = \mu + \sigma Z, \qquad Z \sim N(0, 1). \tag{5}$$

- Tällöin

$$EX = \mu, \qquad \operatorname{var} X = \sigma^2 \operatorname{var} Z = \sigma^2,$$

joten μ on $N(\mu, \sigma^2)$-jakauman odotusarvo ja σ^2 sen varianssi.

- Jos $\sigma > 0$, niin jakauman tf on

$$f_X(x) = \frac{1}{\sigma} \phi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sigma} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}\right), \qquad x \in \mathbb{R}. \tag{6}$$

Normaalijakauman ominaisuuksia

- $N(\mu, \sigma^2)$-jakauman momenttiemäfunktio on

$$M_X(t) = M_{\mu+\sigma Z}(t) = E \exp((\mu + \sigma Z)t) = \exp(\mu t)\, M_Z(\sigma t) \\ = \exp(\mu t + \frac{1}{2}\sigma^2 t^2). \tag{7}$$

- Yhteenlaskuominaisuus. Jos $X \sim N(\mu_1, \sigma_1^2)$ ja $Y \sim N(\mu_2, \sigma_2^2)$, ja ne ovat riippumattomia, niin

$$M_{X+Y}(t) = M_X(t)\, M_Y(t) = \exp((\mu_1 + \mu_2)t + \frac{1}{2}(\sigma_1^2 + \sigma_2^2)t^2),$$

joten $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

5.3.7 Normaalijakaumasta johdettuja jakaumia

Määritelmä (Khiin neliön jakauma)

Jos $X_1, \ldots, X_n \sim N(0,1)$ ja ne ovat riippumattomia, niin niiden neliöiden summalla sanotaan olevan χ^2_n-jakauma eli khiin neliön jakauma n:llä vapausasteella (tai vapausasteluvulla n).

Seuraavaksi nähdään, että khiin neliön jakauma on erikoistapaus gammajakaumasta.

χ^2 yhdellä vapausasteella on $\mathrm{Gam}(\frac{1}{2}, \frac{1}{2})$

- Olkoon $X \sim N(0,1)$, ja $Y = X^2$. Tietysti $Y \geq 0$.
- Jos $y \geq 0$, niin

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= \Phi(\sqrt{y}) - \Phi(-\sqrt{y}). \end{aligned}$$

- $F_Y(y) = 0$, jos $y < 0$. Kf F_Y on jatkuva pisteessä $y = 0$ ja jatkuvasti derivoituva, kun $y \neq 0$. Jakauma on jatkuva (ks. lause 2.7), joten tf:n saa laskea derivoimalla kf:n.
- Jos $y > 0$, niin

$$\begin{aligned} f_Y(y) &= [\phi(\sqrt{y}) + \phi(-\sqrt{y})] \, \frac{1}{2\sqrt{y}} \\ &= \frac{(\frac{1}{2})^{1/2}}{\Gamma(\frac{1}{2})} \, y^{\frac{1}{2}-1} \, \mathrm{e}^{-\frac{1}{2} y} \end{aligned}$$

χ^2_n on $\mathrm{Gam}(\frac{n}{2}, \frac{1}{2})$

- Olk. $X_1, \ldots, X_n \sim N(0, 1)$ riippumattomia.
- Gammajakauman yhteenlaskuominaisuuden perusteella $X_1^2 + X_2^2 \sim \mathrm{Gam}(\frac{1}{2} + \frac{1}{2}, \frac{1}{2})$,
- $(X_1^2 + X_2^2) + X_3^2 \sim \mathrm{Gam}(\frac{2}{2} + \frac{1}{2}, \frac{1}{2})$,
- samaa päättelyä jatkamalla

$$X_1^2 + \cdots + X_n^2 \sim \mathrm{Gam}(\frac{n}{2}, \frac{1}{2}).$$

- Ts. χ^2_n on sama kuin $\mathrm{Gam}(\frac{n}{2}, \frac{1}{2})$.
- Yleistys. Jos $\nu > 0$ ei ole kokonaisluku, niin määrittelemme, että χ^2_ν-jakauma on gammajakauma $\mathrm{Gam}(\frac{\nu}{2}, \frac{1}{2})$.

t-jakauma

Määritelmä (t-jakauma)

Jos $Z \sim N(0,1)$ ja $V \sim \chi^2_\nu$ (jollekin $\nu > 0$) ja $Z \perp\!\!\!\perp V$, niin satunnaismuuttujalla

$$T = \frac{Z}{\sqrt{V/\nu}}$$

on jakauma, jota kutsutaan (Studentin) t-jakaumaksi ν:llä vapausasteella tai vapausasteluvulla ν (engl. *degrees of freedom, df*), eli $T \sim t_\nu$.

F-jakauma

Määritelmä (F-jakauma.)

Jos $U \sim \chi_k^2$ ja $V \sim \chi_m^2$ ja $U \perp\!\!\!\perp V$, niin sm:lla

$$Y = \frac{U/k}{V/m}$$

on jakauma, jota kutsutaan F-jakaumaksi parametreilla k (osoittajan vapausasteluku) ja m (nimittäjän vapausasteluku), eli $Y \sim F_{k,m}$.