DISKREETIT JAKAUMAT

Generoiva funktio (z-muunnos)

Määritelmä

Olkoon X diskreetti sm, jonka arvot ovat ei-negatiivisia kokonaislukuja, $X \in \{0, 1, 2, \ldots\}$.

Merkitään tämän pistetodennäköisyyksiä p_i:llä

$$p_i = \mathrm{P}\{X = i\}$$

X:n generoiva funktio $\mathcal{G}(z)$ (tai $\mathcal{G}_X(z)$; myös $X(z)$ tai $\hat{X}(z)$)

$$\boxed{\mathcal{G}(z) = \sum_{i=0}^{\infty} p_i z^i = \mathrm{E}[z^X]}$$

- Kokoaa kätevästi yhteen arvot $\{p_0, p_1, \ldots\}$; z on 'kirjanpitomuuttuja'
- Monissa tapauksissa $\mathcal{G}(z)$ kyetään laskemaan (yksinkertainen analyyttinen lauseke)
- Kun $\mathcal{G}(z)$ on annettu, siitä kyetään takaperin päättelemään arvot $\{p_0, p_1, \ldots\}$
- Eräät jakaumille suoritettavat operaatiot vastaavat paljon yksinkertaisempia operaatioita generoiville funktioille
- Usein yksinkertaistaa rekursiivisten yhtälöiden ratkaisemista

Käänteismuunnos

Tehtävänä on päätellä todennäköiyydet p_i, kun $\mathcal{G}(z)$ on annettu.

Kolme eri keinoa

1. Kehittämällä $\mathcal{G}(z)$ potenssisarjaksi, josta p_i identifioidaan potenssin z^i kertoimena. Kerroin voidaan laskea myös mekaanisesti
$$p_i = \frac{1}{i!}\frac{d^i\mathcal{G}(z)}{dz^i}\Big|_{z=0} = \frac{1}{i!}\mathcal{G}^{(i)}(0)$$
2. Tarkastelemalla: hajoitetaan $\mathcal{G}(z)$ osiin, joiden käänteismuunnokset tunnetaan; esim. osamurtokehitelmä
3. Kompleksitason (polku)integraalina
$$p_i = \frac{1}{2\pi i}\oint \frac{\mathcal{G}(z)}{z^{i+1}}\,dz$$
polku origon ympäri (valitaan siten, että $\mathcal{G}(z)$:n navat jäävät polun ulkopuolelle)

Esim. 1

$$\mathcal{G}(z) = \frac{1}{1-z^2} = 1 + z^2 + z^4 + \cdots$$

$$\Rightarrow \quad p_i = \begin{cases} 1 & \text{kun } i \text{ parillinen} \\ 0 & \text{kun } i \text{ pariton} \end{cases}$$

Esim. 2

$$\mathcal{G}(z) = \frac{2}{(1-z)(2-z)} = \frac{2}{1-z} - \frac{2}{2-z} = \frac{2}{1-z} - \frac{1}{1-z/2}$$

Koska $\dfrac{A}{1-az}$ vastaa jonoa $A \cdot a^i$ päätellään

$p_i = 2 \cdot (1)^i - 1 \cdot (\frac{1}{2})^i = 2 - (\frac{1}{2})^i$

Jakauman momenttien laskeminen $\mathcal{G}(z)$:n avulla

Koska p_i:t edustavat todennäköisyysjakaumaa, niiden summa on 1 ja

$$\mathcal{G}(1) = \mathcal{G}^{(0)}(1) = \sum_{i=1}^{\infty} p_i \cdot 1^i = 1$$

Derivoimalla todetaan

$$\begin{aligned} \mathcal{G}^{(1)}(z) &= \tfrac{d}{dz}\mathrm{E}[z^X] \\ &= \mathrm{E}[X z^{X-1}] \\[1ex] \mathcal{G}^{(1)}(1) &= \mathrm{E}[X] \end{aligned}$$

Jatkamalla samoin saadaan

$$\boxed{\mathcal{G}^{(i)}(1) = \mathrm{E}[X(X-1)\cdots(X-i+1)] = F_i}$$

missä F_i on i:s kertomamomentti.

Kertomamomenttien ja tavallisten (origo)momenttien suhde

Kertomamomenttien $F_i = \mathrm{E}[X(X-1)\cdots(X-i+1)]$ ja tavallisten (origo)momenttien $M_i = \mathrm{E}[X^i]$ ja välillä vallitsevat lineaariset relaatiot:

$$\boxed{\left\{\begin{array}{rcl} F_1 &=& M_1 \\ F_2 &=& M_2 - M_1 \\ F_3 &=& M_3 - 3M_2 + 2M_1 \\ \vdots && \end{array}\right. \qquad \left\{\begin{array}{rcl} M_1 &=& F_1 \\ M_2 &=& F_2 + F_1 \\ M_3 &=& F_3 + 3F_2 + F_1 \\ \vdots && \end{array}\right.}$$

Esimerkiksi

$$F_2 = \mathcal{G}^{(2)}(1) = \mathrm{E}[X(X-1)] = \mathrm{E}[X^2] - \mathrm{E}[X]$$

$$\Rightarrow \quad M_2 = \mathrm{E}[X^2] = F_2 + F_1 = \mathcal{G}^{(2)}(1) + \mathcal{G}^{(1)}(1)$$

$$\Rightarrow \quad \mathrm{V}[X] = M_2 - M_1^2 = \mathcal{G}^{(2)}(1) + \mathcal{G}^{(1)}(1) - (\mathcal{G}^{(1)}(1))^2 = \mathcal{G}^{(2)}(1) + \mathcal{G}^{(1)}(1)(1 - \mathcal{G}^{(1)}(1))$$

Momenttien laskeminen suoraan

Momentit voidaan laskea generoivasta funktiosta myös suoraan (ilman kertomamomentteja) seuraavasti:

$$\begin{aligned} \frac{d}{dz}\mathcal{G}(z)\Big|_{z=1} &= \mathrm{E}[X z^{X-1}]_{z=1} &= \mathrm{E}[X] \\ \frac{d}{dz} z \frac{d}{dz}\mathcal{G}(z)\Big|_{z=1} &= \mathrm{E}[X^2 z^{X-1}]_{z=1} &= \mathrm{E}[X^2] \end{aligned}$$

Yleisesti

$$\boxed{\mathrm{E}[X^i] = \frac{d}{dz}(z\,\frac{d}{dz})^{i-1}\mathcal{G}(z)\Big|_{z=1} = (z\,\frac{d}{dz})^{i}\mathcal{G}(z)\Big|_{z=1}}$$

Riippumattomien sm:ien summan generoiva funktio

Olkoot X ja Y riippumattomia satunnaismuuttujia. Tällöin

$$\begin{aligned} \mathcal{G}_{X+Y}(z) &= \mathrm{E}[z^{X+Y}] = \mathrm{E}[z^X z^Y] \\ &= \mathrm{E}[z^X]\mathrm{E}[z^Y] \qquad \text{riippumattomuus} \\ &= \mathcal{G}_X(z)\mathcal{G}_Y(z) \end{aligned}$$

$$\boxed{\mathcal{G}_{X+Y}(z) = \mathcal{G}_X(z)\mathcal{G}_Y(z)}$$

Alkuperäisillä diskreeteillä jakaumilla

$$\begin{cases} p_i = \mathrm{P}\{X = i\} \\ q_j = \mathrm{P}\{Y = j\} \end{cases}$$

ilmaistuna summaa vastaa konvoluutio $p \otimes q$

$$\mathrm{P}\{X + Y = k\} = (p \otimes q)_k = \sum_{i=0}^{k} p_i q_{k-i}$$

Kahden jakauman konvoluutiona saadun jakauman generoiva funktio on siten alkuperäisten jakaumien generoivien funktioiden tulo.

Yhdistetty jakauma (compound distribution) ja sen generoiva funktio

Olkoon Y riippumattomien, samoin jakautuneiden ($i.i.d.$) satunnaismuuttujien X_i summa

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

missä N itse on ei-negatiivinen kokonaislukuarvoiden satunnaismuuttuja.

Merkitään

$$\begin{cases} \mathcal{G}_X(z) & \text{muuttujien } X_i \text{ yhteinen generoiva funktio} \\ \mathcal{G}_N(z) & N\text{:n generoiva funktio} \end{cases}$$

Halutaan laskea $\mathcal{G}_Y(z)$

$$\begin{aligned} \mathcal{G}_Y(z) &= \mathrm{E}[z^Y] \\ &= \mathrm{E}[\mathrm{E}\left[z^Y \,|\, N\right]] \\ &= \mathrm{E}[\mathrm{E}\left[z^{X_1+\cdots X_N} \,|\, N\right]] \\ &= \mathrm{E}[\mathrm{E}\left[z^{X_1} \cdots z^{X_N} \,|\, N\right]] \\ &= \mathrm{E}[\mathcal{G}_X(z)^N] \\ &= \mathcal{G}_N(\mathcal{G}_X(z)) \end{aligned}$$

$$\boxed{\mathcal{G}_Y(z) = \mathcal{G}_N(\mathcal{G}_X(z))}$$

Bernoulli-jakauma $X \sim \mathrm{Bernoulli}(p)$

Yksinkertainen koe, jossa on kaksi mahdollista lopputulosta: 'onnistuminen' ja 'epäonnistuminen'.

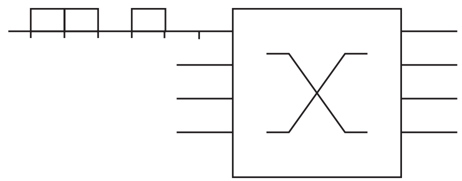
Liitetään näihin sm X seuraavasti

$$X = \begin{cases} 1 & \text{kun koe onnistuu; todennäköisyys } p \\ 0 & \text{kun koe epäonnistuu; todennäköisyys } q = 1 - p \end{cases}$$

Esim. 1. X kuvaa bittivirtaa liikennelähteestä, joka on joko päällä tai pois päältä.
Generoiva funktio

$$\begin{aligned} \mathcal{G}(z) &= p_0 z^0 + p_1 z^1 = q + pz \\ \mathrm{E}[X] &= \mathcal{G}^{(1)}(1) = p \\ \mathrm{V}[X] &= = \mathcal{G}^{(2)}(1) + \mathcal{G}^{(1)}(1 - \mathcal{G}^{(1)}) = p(1-p) = pq \end{aligned}$$

Esim. 2. ATM-kytkimen tuloporttiin saapuva soluvirta: aikalovessa (soluväli) on solu todennäköisyydellä p tai se on tyhjä todennäköisyydellä q.



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

X on onnistumisten lukumäärä n kertaa toistetussa Bernoulli-kokeessa.

$$X = \sum_{i=1}^{n} Y_i \qquad \text{missä } Y_i \sim \text{Bernoulli}(p) \text{ ja } Y_i\text{:t riippumattomia } (i = 1, \ldots, n)$$

Generoiva funktio saadaan suoraan Bernoulli-muuttujan generoivasta funktiosta $q + pz$

$$\mathcal{G}(z) = (q + pz)^n = \sum_{i=1}^{n} \binom{n}{i} p^i (1-p)^{n-i} z^i$$

Identifioimalla z^i:n kerroin

$$\boxed{p_i = \mathrm{P}\{X = i\} = \binom{n}{i} p^i (1-p)^{n-i}}$$

$$\begin{cases} \mathrm{E}[X] & = & n\mathrm{E}[Y_i] = np \\ \mathrm{V}[X] & = & n\mathrm{V}[Y_i] = np(1-p) \end{cases}$$

Rajamuoto kun $\lambda = \mathrm{E}[X] = np$ on annettu ja $n \to \infty$:

$$\mathcal{G}(z) = (1 - (1-z)p)^n = (1 - (1-z)\lambda/n)^n \to e^{(1-z)\lambda}$$

joka on Poisson-jakautuneen satunnaismuuttujan generoiva funktio.

Binomijakautuneiden sm:ien summan jakauma

Olkoot X_i:t $(i = 1, \ldots, k)$ binomijakautuneita samalla parametrilla p. Tällöin niiden summan jakauma on

$$X_1 + \cdots + X_k \sim \text{Bin}(n_1 + \cdots + n_k, p)$$

koska summa edustaa onnistumisten lukumäärää toistetussa Bernoulli-kokeessa, jossa toistojen lukumäärä on $n_1 + \cdots + n_k$.

Multinomijakauma

Tarkastellaan n-kertaista toistokoetta, mutta nyt kukin koe voi päättyä k:hon ($k \geq 2$) erilaiseen tulokseen. Näiden todennäköisyydet yksittäisessä kokeessa ovat $p_1, p_2, \ldots, p_k$ ($\sum_{i=1}^k p_i = 1$).

Olkoon tuloksen i esiintymien lkm toistokokeessa N_i. Halutaan laskea yhteistapahtuman $\{N_1 = n_1, \ldots, N_k = n_k\}$ todennäköisyys $p(n_1, \ldots, n_k) = \mathrm{P}\{N_1 = n_1, \ldots, N_k = n_k\}$.

Määritellään monen muuttujan $N_1, \ldots, N_k$ yhteisjakauman generoiva funktio

$$\mathcal{G}(z_1, \ldots, z_k) = \mathrm{E}[z_1^{N_1} \cdots z_k^{N_k}] = \sum_{n_1=0}^{\infty} \cdots \sum_{n_k=0}^{\infty} p(n_1, \ldots, n_k) z_1^{n_1} \cdots z_k^{n_k}$$

Yhden kokeen jälkeen yksi esiintymälukumääristä N_i on 1 ja muut ovat nollia. Yhtä koetta vastaava generoiva funktio on siten $p_1 z_1 + \cdots + p_k z_k$.

Riippumattomien kokeiden tapauksessa n-kertaisen toistokokeen tuottamien esiintymälukumäärien gf on yksittäisten kokeiden generoivien funktioiden tulo eli $(p_1 z_1 + \cdots + p_k z_k)^n$.

z_i-muuttujien potenssien kertoimista identifioidaan

$$\boxed{p(n_1, \ldots, n_k) = \frac{n!}{n_1! \cdots n_k!} p_1^{n_1} \cdots p_k^{n_k}} \qquad \begin{array}{l} \text{kun } n_1 + \ldots + n_k = n, \\ 0 \text{ muulloin} \end{array}$$

Geometrinen jakauma $X \sim \text{Geom}(p)$

X on toistetussa Bernoulli-kokeessa (onnistumistodennäköisyys p) ensimmäiseen onnistumiseen tarvittavien toistojen lukumäärä

$$\boxed{p_i = \text{P}\{X = i\} = (1-p)^{i-1}p}$$

$i = 1, 2, \ldots$
Huom. joskus määritellään $X - 1$ geometriseksi jakaumaksi (nollasta alkava geometrinen jakauma)

Generoiva funktio

$$\mathcal{G}(z) = p\sum_{i=1}^{\infty}(1-p)^{i-1}z^i = \frac{pz}{1-(1-p)z}$$

Tämän avulla lasketaan odotusarvo ja varianssi:

$$\boxed{\begin{aligned}
\text{E}[X] &= \mathcal{G}'(1) = \frac{p(1-(1-p)z) - p(1-p)z}{(1-(1-p)z)^2}\Big|_{z=1} = \frac{1}{p} \\
\text{E}[X^2] &= \mathcal{G}'(1) + \mathcal{G}''(1) = \frac{1}{p} + \frac{2(1-p)}{p^2} \\
\text{V}[X] &= \text{E}[X^2] - \text{E}[X]^2 = \frac{1-p}{p^2}
\end{aligned}}$$

Geometrinen jakauma (jatkoa)

Todennäköisyys että ensimmäiseen onnistumiseen tarvitaan yli n toistoa

$$\mathrm{P}\{X > n\} = \sum_{i=n+1}^{\infty} p_i = (1-p)^n$$

Geometrisen jakauman muistittomuus

$$\begin{aligned}\mathrm{P}\{X > i+j \,|\, X > i\} &= \frac{\mathrm{P}\{X > i+j \cap X > i\}}{\mathrm{P}\{X > i\}} = \frac{\mathrm{P}\{X > i+j\}}{\mathrm{P}\{X > i\}} \\ &= \frac{(1-p)^{i+j}}{(1-p)^i} = \mathrm{P}\{X > j\}\end{aligned}$$

Jos epäonnistumisia on ollut jo i kpl, todennäköisyys sille, että ensimmäiseen onnistumiseen tarvitaan vielä yli j koetta lisää on sama kuin kokonaan uudessa toistokokeessa todennäköisyys sille, että ensimmäiseen onnistumiseen tarvitaan enemmän kuin j koetta.

Näin tulee tietenkin ollakin, koska jo suoritetut kokeet eivät vaikuta jatkokokeisiin, jotka kaikki ovat riippumattomia.

Negatiivinen binomijakauma $X \sim \mathrm{NBin}(n, p)$

X on toistetussa Bernoulli-kokeessa tarvittavien toistojen lukumäärä, jotta saadaan n onnistumista.

Jos $X = i$, niin $(i-1)$:n ensimmäisen kokeen joukossa on täytynyt olla $n-1$ onnistumista ja kokeen i täytyy johtaa onnistumiseen:

$$\boxed{p_i = \mathrm{P}\{X = i\} = \binom{i-1}{n-1} p^{n-1}(1-p)^{i-n} \cdot p = \binom{i-1}{n-1} p^n (1-p)^{i-n}} \qquad \begin{array}{l} \text{kun } i \ge n \\ 0 \text{ muulloin} \end{array}$$

Ensimmäiseen onnistumiseen tarvittavien toistojen lkm $\sim \mathrm{Geom}(p)$. Samoin tästä eteenpäin toiseen onnistumiseen jne, joten

$$X = X_1 + \cdots + X_n \qquad \text{missä } X_i \sim \mathrm{Geom}(p) \quad (i.i.d.)$$

Siten jakauman generoiva funktio on

$$\mathcal{G}(z) = \left(\frac{pz}{1-(1-p)z}\right)^n$$

Ylläolevat pistetodennäköisyydet voidaan päätellä myös tästä

Keskiarvo ja varianssi ovat n-kertaiset geometrisen jakauman vastaaviin nähden

$$\boxed{\mathrm{E}[X] = \frac{n}{p}} \qquad \boxed{\mathrm{V}[X] = n\,\frac{1-p}{p}}$$

Poisson-jakauma $X \sim \text{Poisson}(a)$

X on ei-negatiivinen kokonaislukuarvoinen sm, jonka pistetodennäköisyydet ovat

$$\boxed{p_i = \text{P}\{X = i\} = \frac{a^i}{i!}\, e^{-a}} \qquad i = 0, 1, \ldots$$

Generoiva funktio

$$\mathcal{G}(z) = \sum_{i=0}^{\infty} p_i z^i = e^{-a} \sum_{i=0}^{\infty} \frac{(za)^i}{i!} = e^{-a} e^{za}$$

$$\boxed{\mathcal{G}(z) = e^{(z-1)a}}$$

Kuten aikaisemmin nähtiin, tämä generoiva funktio saadaan $\text{Bin}(n, p)$-jakautuneen sm:n generoivasta funktiosta rajamuotona, kun keskimääräinen onnistumisten lkm pidetään vakiona, $np = a$, ja n:n annetaan rajatta kasvaa.

Vastaavasti $X \sim \text{Poisson}(\lambda t)$ edustaa tapahtumien lukumäärää t:n pituisella välillä Poisson-prosessista, jonka intensiteetti on λ:

- pienessä välissä dt tapahtuman todennäköisyys on λdt ('koe onnistuu')
- kahden yhtäaikaisen tapahtuman todennäköisyys on $\mathcal{O}(\lambda dt)$
- tapahtumien lukumäärät yhteispisteettömissä väleissä ovat riippumattomia

Poisson-jakauma (jatkoa)

Poisson-jakaumaa noudattavat esimerkiksi

- Saapuvien puheluiden lkm tiettynä aikavälinä
- Käynnissä olevien puheluiden lukumäärä suuressa (estottomassa) johtoryhmässä

Keskiarvo ja varianssi

$$\begin{cases} \mathrm{E}[X] = \mathcal{G}'(1) = \frac{d}{dz} e^{(z-1)a} \Big|_{z=1} = a \\ \mathrm{E}[X^2] = \mathcal{G}''(1) + \mathcal{G}'(1) = a^2 + a \end{cases} \quad \Rightarrow \quad \mathrm{V}[X] = \mathrm{E}[X^2] - \mathrm{E}[X]^2 = a^2 + a - a^2 = a$$

$$\boxed{\mathrm{E}[X] = a} \qquad \boxed{\mathrm{V}[X] = a}$$

Poisson-jakauman ominaisuuksia

1. Poisson-jakautuneiden satunnaismuuttujien summa on Poisson-jakautunut.

$$X = X_1 + X_2, \quad \text{missä } X_1 \sim \text{Poisson}(a_1),\ X_2 \sim \text{Poisson}(a_2)$$

$$\Rightarrow \quad X \sim \text{Poisson}(a_1 + a_2)$$

Todistus:

$$\mathcal{G}_{X_1}(z) = e^{(z-1)a_1},\ \mathcal{G}_{X_2}(z) = e^{(z-1)a_2}$$

$$\mathcal{G}_X(z) = \mathcal{G}_{X_1}(z)\mathcal{G}_{X_2}(z) = e^{(z-1)a_1}e^{(z-1)a_2} = e^{(z-1)(a_1+a_2)}$$

2. Jos joukon koko N on Poisson-jakautunut, $N \sim \text{Poisson}(a)$, ja siitä suoritetaan satunnaispoiminta todennäköisyydellä p (kukin alkio poimitaan tällä tn:llä), niin saadun joukon koko $K \sim \text{Poisson}(pa)$.

Todistus: K noudattaa yhdistettyä jakaumaa

$$K = X_1 + \cdots + X_N, \quad \text{missä } N \sim \text{Poisson}(a) \text{ ja } X_i \sim \text{Bernoulli}(p)$$

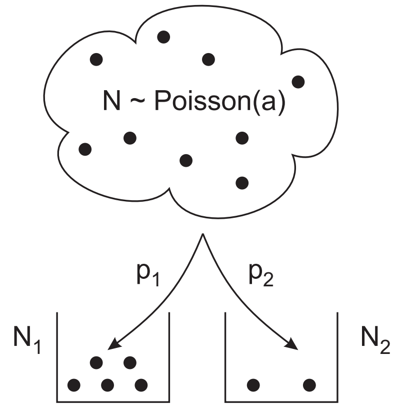
$$\mathcal{G}_X(z) = (1-p) + pz, \quad \mathcal{G}_N(z) = e^{(z-1)a}$$

$$\mathcal{G}_K(z) = \mathcal{G}_N(\mathcal{G}_X(z)) = e^{(\mathcal{G}_X(z)-1)a} = e^{[(1-p)+pz-1]a} = e^{(z-1)pa}$$

Poisson-jakauman ominaisuuksia (jatkoa)

3. Jos joukon $N \sim \text{Poisson}(a)$ elementit lajitellaan satunnaisesti kahteen joukkoon 1 ja 2 todennäköisyyksillä p_1 ja $p_2 = 1 - p_1$, niin joukkojen 1 ja 2 koot N_1 ja N_2 ovat riippumattomia ja jakautuneet kuten

   $N_1 \sim \text{Poisson}(p_1 a), \quad N_2 \sim \text{Poisson}(p_2 a)$



Todistus: Kokonaistodennäköisyyden perusteella

$$\begin{aligned}
\mathrm{P}\{N_1 = n_1, N_2 = n_2\} &= \sum_{n=0}^{\infty} \underbrace{\mathrm{P}\{N_1 = n_1, N_2 = n_2 \,|\, N = n\}}_{\text{multinomijakauma}} \underbrace{\mathrm{P}\{N = n\}}_{\text{Poisson-jak.}} \\
&= \frac{n!}{n_1! n_2!} p_1^{n_1} p_2^{n_2} \cdot \frac{a^n}{n!} e^{-a}\Big|_{n=n_1+n_2} = \frac{p_1^{n_1} p_2^{n_2}}{n_1! n_2!} \cdot a^{n_1+n_2} e^{-a \overbrace{(p_1+p_2)}^{1}} \\
&= \frac{(p_1 a)^{n_1}}{n_1!} e^{-p_1 a} \cdot \frac{(p_2 a)^{n_2}}{n_2!} e^{-p_2 a} = \mathrm{P}\{N_1 = n_1\} \cdot \mathrm{P}\{N_2 = n_2\}
\end{aligned}$$

Yhteisjakauma on tulomuotoinen. N_1 ja N_2 ovat siten riippumattomia. Tulon tekijät ovat Poisson($p_1 a$)- ja Poisson($p_2 a$)-jakaumien pistetodennäköisyyksiä.

Huom. Tulos yleistyy luonnollisella tavalla mille tahansa lukumäärälle lajittelujoukkoja.

Kollektiivisten merkkien menetelmä (Dantzig)

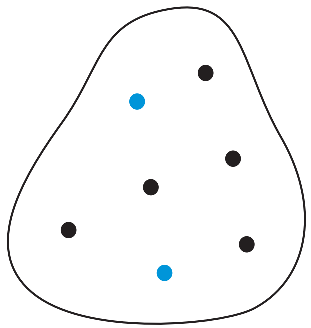
Edellä generoivan funktion muuttujaa z on käsitelty vain teknisenä apumuuttujana ('kirjanpitomuuttuja').

Ns. kollektiivisten merkkien menetelmässä muuttujalle z annetaan todennäköisyystulkinta. Tämä mahdollistaa joidenkin tulosten elegantin johtamisen yksinkertaisella päättelyllä.

Olkoon $N = 0, 1, 2, \ldots$ ei-negatiivinen kokonaislukuarvoinen sm. ja $\mathcal{G}_N(z)$ sen generoiva funktio:

$$\mathcal{G}_N(z) = \sum_{n=0}^{\infty} p_n z^n, \qquad p_n = \mathrm{P}\{N = n\}$$

Tulkinta: Ajatellaan, että N edustaa jonkin joukon kokoa. Merkitään joukon kukin alkio muista riippumatta todennäköisyydellä $1 - z$ ja siis jätetään ilman merkkiä todennäköisyydellä z. *Tällöin $\mathcal{G}_N(z)$ on tn sille, että koko joukossa ei ole merkkiä.*

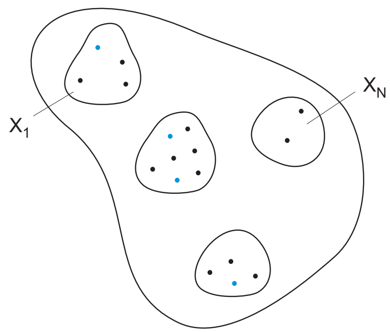


Kollektiivisten merkkien menetelmä (jatkoa)

Esimerkki: Yhdistetyn jakauman generoiva funktio

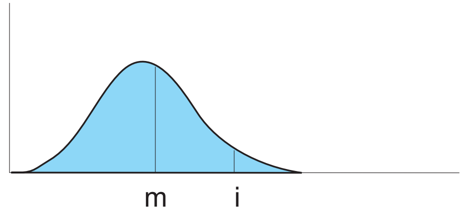
$$Y = X_1 + \cdots + X_N \ , \quad \text{missä}$$

$$\begin{cases} X_1 \sim X_2 \sim \cdots \sim X_N, \text{ yhteinen gen. funktio } \mathcal{G}_X(z) \\ N \text{ on sm., jonka generoiva funktio on } \mathcal{G}_N(z) \end{cases}$$



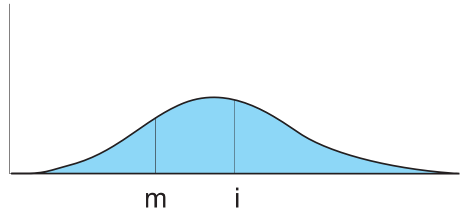
$$\begin{aligned} \mathcal{G}_Y(z) &= \text{P}\{\text{joukon } Y \text{ mikään elementti ei ole merkitty}\} \\ &= \underbrace{\mathcal{G}_N(\underbrace{\mathcal{G}_X(z)}_{\substack{\text{tn. että yksittäi-} \\ \text{nen aliryhmä ei} \\ \text{ole merkitty}}})}_{\substack{\text{tn. että mikään aliryhmä} \\ \text{ei ole merkitty}}} \end{aligned}$$

Todennäköisyyden vinoutus: pistetn:ien likim. laskenta



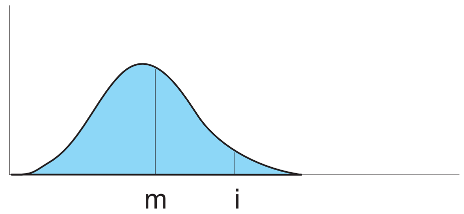
Monia jakaumia voidaan kohtuullisella tarkkuudella approksimoida normaalijakaumalla, kun keskiarvoparametri on suuri.

Esim. Poisson(a) $\approx$ N(a, a), kun $a \gg 1$



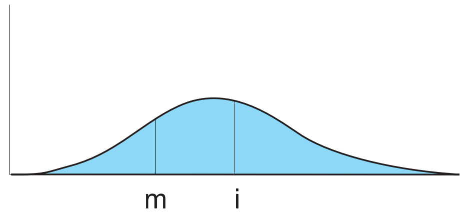
- Approksimaatio on yleensä tarkka lähellä jakauman keskiarvoa, mutta kaukana tästä suhteellinen virhe voi olla huomattava.
- Approksimaatiota voidaan merkittävästi parantaa todennäköisyyden vinoutusmenetelmällä (probability shift/tilt method).
- Tämä tarjoaa keinon laskea sellaisen jakauman pistetodennäköisyys (jakauman hännässä), jonka generoiva funktio tunnetaan.

Todennäköisyyden vinoutus (jatkoa)



Halutaan laskea sm:n X pistetodennäköisyys

$$p_i = \mathrm{P}\{X = i\}\ , \quad \text{kun } i \gg \mathrm{E}[X]\ (= m)$$



Suoritetaan jakauman vinoutus ja tarkastellaan satunnaismuuttujaa X', jonka pistetodennäköisyydet ovat

$$p'_i = \frac{p_i z^i}{\mathcal{G}(z)}$$

Tämä on normitettu jakauma, koska $\mathcal{G}(z) = \sum_i p_i z^i$.

Vinoutetun jakauman momentit

$$\begin{cases} m'(z) = \mathrm{E}[X'] = \dfrac{1}{\mathcal{G}(z)} z \dfrac{d}{dz} \mathcal{G}(z) \\ \qquad\qquad \mathrm{E}[X'^2] = \dfrac{1}{\mathcal{G}(z)} (z \dfrac{d}{dz})^2 \mathcal{G}(z) \\ \sigma'^2(z) = \mathrm{V}[X'] = \mathrm{E}[X'^2] - \mathrm{E}[X']^2 \end{cases}$$

Todennäköisyyden vinoutus (jatkoa)

Valitaan erityisesti $z = z^*$ siten, että $m'(z^*) = i$ eli että vinoutetun jakauman keskiarvo sijaitsee kiinnostavassa pisteessä i. Soveltamalla nyt normaaliapproksimaatiota vinoutettuun jakaumaan saadaan

$$p'_i \approx \frac{1}{\sqrt{2\pi\sigma'^2}}$$

Ratkaisemalla kääntäen p_i aikaisemmasta relaatiosta saadaan haluttu approksimaatio

$$\boxed{p_i \approx \frac{\mathcal{G}(z^*)}{(z^*)^i\sqrt{2\pi\sigma'^2(z^*)}}}$$

missä z^* on yhtälön $m'(z^*) = i$ ratkaisu

Lausekkeen laskemiseksi tarvitsee ainoastaan tuntea X:n generoiva funktio.

Menetelmä on erityisen käyttökelpoinen silloin, kun X on usean eri tavoin jakautuneen riippumattoman satunnaismuuttujan summa, joiden kunkin jakauma (ja generoiva funktio) tunnetaan.

X:n jakauma on tällöin monimutkainen, mutta koska sen generoiva funktio tunnetaan (generoivien funktioiden tulo) ylläolevaa menetelmää voidaan soveltaa.

Todennäköisyyden vinoutus (jatkoa)

Esimerkki (sinänsä triviaali/epämielekäs)

Poisson-jakauma

$$p_i = \frac{a^i}{i!}e^{-a}, \quad \mathcal{G}(z) = e^{(z-1)a}$$

$$p'_i = \frac{p_i z^i}{\mathcal{G}(z)} = \frac{(az)^i}{i!}e^{-az} \qquad \text{Poisson}(za)\text{-jakauma, joten momentit saadaan välittömästi}$$

$$\Rightarrow \quad m'(z) = az, \quad \sigma'^2(z) = az$$

Yhtälön $m'(z^*) = i$ ratkaisu on $z^* = \dfrac{i}{a}$

$$p_i \approx \frac{e^{(i/a-1)a}}{(i/a)^i\sqrt{2\pi i}} = \frac{a^i}{\sqrt{2\pi i}e^{-i}i^i}e^{-a}$$

Havaitaan, että approksimaatio antaa miltei oikean Poisson-todennäköisyyden, mutta nimittäjässä $i!$ on korvautunut tunnetulla Stirlingin approksimaatiollaan $i! \approx \sqrt{2\pi i}e^{-i}i^i$.