Insinöörimatematiikka: Todennäköisyyslaskenta

Mika Hirvensalo
mikhirve@utu.fi

Matematiikan ja tilastotieteen laitos
Turun yliopisto

2024

Normaalijakauma

Määritelmä

Jatkuvalla satunnaismuuttujalla on *standardoitu* normaalijakauma, jos sen tiheysfunktio on muotoa

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}.$$

Tällöin kertymäfunktio on muotoa

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}t^2}\, dt.$$

Normaalijakaumaa kutsutaan myös *Gaussin jakaumaksi* ja sen tiheysfunktion kuvaajaa *Gaussin kellokäyräksi*.

Huomautus

Integraali

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2}t^2}\, dt = \sqrt{2\pi}$$

on hankala laskea.
Vielä hankalampi laskettava on

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}t^2}\, dt.$$

Aiemmin taulukot, nykyisin ohjelmistot.

Normaalijakauma

Lause

Standardoidulle normaalijakaumalle

$$\mathbb{E}(X) = 0 \text{ ja } \mathrm{Var}(X) = 1 \text{ ja } \Phi(-x) = 1 - \Phi(x).$$

Esimerkki

Jos jatkuvalla satunnaismuuttujalla on standardoitu normaalijakauma, on

- $\mathbb{P}(X \leq 1,28) = \Phi(1,28) = 0,8997$
- $\mathbb{P}(X > 1,28) = 1 - \Phi(1,28) = 0,1003$
- $\mathbb{P}(X < -1,28) = 1 - \Phi(1,28) = 0,1003$
- $\mathbb{P}(X \geq -1,28) = 1 - \Phi(-1,28) = 0,8997$

Normaalijakauma

Yleinen normaalijakauma

Jakaumalla $N(\mu, \sigma^2)$ tarkoitetaan jakaumaa, jonka tiheysfunktio on

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

ja tästä johdettava kertymäfunktio

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right).$$

Jakaumalle $N(\mu, \sigma^2)$ on $\mathbb{E}(X) = \mu$ ja $\mathrm{Var}(X) = \sigma^2$.

Esimerkki

Jos satunnaismuuttujalla X on jakauma $N(\mu, \sigma^2)$, voidaan todennäköisyyksiä määrittää edellisen huomautuksen perusteella. Jos esim. $\mu = 6$ ja $\sigma^2 = 5$, on

- $\mathbb{P}(X \leq 7,54) = \Phi(\frac{7,54-6}{\sqrt{5}}) = \Phi(0,69) = 0,7549$
- $\mathbb{P}(X > 7,54) = 1 - \Phi(\frac{7,54-6}{\sqrt{5}}) = 0,2451$
- $\mathbb{P}(X \leq 4,46) = \Phi(\frac{4,46-6}{\sqrt{5}}) = \Phi(-0,69) = 0,2451$
- $\mathbb{P}(X > 4,46) = 1 - \Phi(\frac{4,46-6}{\sqrt{5}}) = 1 - \Phi(-0,69) = \Phi(0,69) = 0,7549$

Esimerkki

Led-lampun kestoikä T noudattaa normaalijakaumaa, jossa odotusarvo $\mu = 25000$ h ja keskihajonta $\sigma = 3000$ h. Määritetään seuraavat todennäköisyydet:

- $\mathbb{P}(T < 20000)$
- $\mathbb{P}(T > 50000)$
- $\mathbb{P}(22000 \leq T \leq 28000)$

Satunnaismuuttujien summat

Markovin epäyhtälö

Jos $\mathbb{P}(X < 0) = 0$ ja $a > 0$, on

$$\mathbb{P}(X > a) \leq \frac{\mathbb{E}(X)}{a}.$$

Chebyschevin epäyhtälö

$$\mathbb{P}(|X - \mathbb{E}(X)| > a) \leq \frac{\mathrm{Var}(X)}{a^2}$$

Satunnaismuuttujien yhteisjakauma

Yhteisjakauma

Olkoon $p(x, y)$ satunnaismuuttujien X ja Y yhteinen jakauma ja $f(X, Y)$ jokin satunnasmuuttujien X ja Y funktio. Tällöin

$$\mathbb{E}(f(X, Y)) = \sum_x \sum_y f(x, y)p(x, y).$$

Riippumattomuus

Satunnaismuuttujat X ja Y ovat riippumattomat, jos $p(x, y) = p_x(x)p_y(y)$ ja lisäksi samoin jakautuneet, jos $p_x = p_y$. Jos satunnaismuuttujat ovat riippumattomat, on

- $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$
- $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$
- $\mathrm{Var}(aX + bY) = a^2 \mathrm{Var}(X) + b^2 \mathrm{Var}(Y)$.

Otoskeskiarvo, sen varianssi

Jos $X_1, \ldots, X_n$ ovat samoin jakautuneita $((\mu, \sigma^2))$, riippumattomia satunnaismuuttujia ja

$$\overline{X}_n = \frac{1}{n}(X_1 + \ldots + X_n)$$

näiden aritmeettinen keskiarvo (ns. otoskeskiarvo), on

- $\mathbb{E}(\overline{X}_n) = n \cdot \frac{1}{n}\mathbb{E}(X_1) = \mu.$
- $\mathrm{Var}(\frac{1}{n}(X_1 + \ldots + X_n)) = n \cdot \frac{1}{n^2}\sigma^2 = \frac{\sigma^2}{n}.$

Lause (Suurten lukujen laki)

Jos $X_1, \ldots, X_n$ ovat samoin jakautuneita $((\mu, \sigma^2))$, riippumattomia satunnaismuuttujia ja

$$\overline{X}_n = \frac{1}{n}(X_1 + \ldots + X_n)$$

näiden aritmeettinen keskiarvo (ns. otoskeskiarvo), on jokaista $\epsilon > 0$ kohti

$$\mathbb{P}(\left|\overline{X}_n - \mu\right| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2} \xrightarrow{n\to\infty} 0.$$

Esimerkki

Heitetään noppaa n kertaa ja olkoon X_i heitolla i saatu silmäluku. Merkitään

$$\overline{X}_n = \frac{1}{n}(X_1 + X_2 + \ldots + X_n).$$

Kuinka suuri pitää heittojen määrän n olla, että varmasti

$$\mathbb{P}(\left|\overline{X}_n - \mu\right| > 0,1) \leq 0,01?$$

Keskeinen raja-arvolause

Olkoot X_i ja $\overline{X}_n$ kuten edellisessä lauseessa. Suurilla n:n arvoilla $\overline{X}_n$ jakautuu likimain normaalijakauman $N(\mu, \frac{\sigma^2}{n})$ mukaisesti. Toisin sanoen,

$$\mathbb{P}(\overline{X}_n \leq x) \approx \Phi(\frac{x-\mu}{\sigma/\sqrt{n}})$$

kun n on suuri.

Esimerkki

Alueelle suunnitellaan asuntoja tuhannelle perheelle. Perheissä on 0, 1, 2 tai 3 lasta todennäköisyyksillä 0.35, 0.45, 0.15 ja 0.05. Mikä on todennäköisyys sille, että 1000 koulupaikkaa ei riitä alueen lapsille?

Olkoon X_i lapsiluku perheessä i. Tällöin $\mathbb{E}(X_i) = 0.9$ (laske) ja $\mathrm{Var}(X_i) = 0.69$ (laske). Jos merkitään $L = X_1 + \ldots + X_{1000}$, on L (kokonaislapsiluku) satunnaismuuttuja, jonka odotusarvo on $0,9 \cdot 1000 = 900$ ja varianssi 690. Tällöin $\sigma = \sqrt{690} \approx 26.3$. Näin ollen

$$
\begin{aligned}
\mathbb{P}(L > 1000) &= 1 - \mathbb{P}(L \leq 1000) \approx 1 - \Phi\left(\frac{1000 - 900}{\sqrt{690}}\right) \\
&= 0,0000703499\ldots
\end{aligned}
$$

Esimerkki

Eräässä kaupungissa huomattiin että 344 kaikista 747:stä kuolemantapauksesta oli sattunut korkeintaan 3 kk syntymäpäivän jälkeen. Mikä on tällaisen todennäköisyys jos kuolemantapaukset jakautuvat tasaisesti?

Jos kuolemat jakautuisivat tasaisesti, pitäisi näitä olla 3 kk syntymäpäivän jälkeen noin $747 \cdot \frac{1}{4} = 186,75$. 344 on kuitenkin noin $1,8$-kertainen tähän nähden.

Määritetään todennäköisyys $\mathbb{P}(X \geq 344)$ sillä otaksumalla, että $p = 0,25$ on todennäköisyys kuolla minä hyvänsä 3 kuukauden ajanjaksona.

Keskeinen raja-arvolause

Binomijakauma

Binomijakauman perustella voidaan laskea

$$\mathbb{P}(X \geq 344) = \sum_{k=344}^{747} \binom{747}{k} 0,25^k \cdot 0,75^{747-k}.$$

Suorassa laskussa voi tulla paljon pöyristysvirheitä. Varianssi $np(1-p) = 747 \cdot 0.25 \cdot 0.75 = 140,06$, joten keskeinen raja-arvolause antaa

$$\begin{aligned} \mathbb{P}(X \geq 344) &= 1 - \mathbb{P}(0 \leq X \leq 343) \\ &\approx 1 - \left(\Phi\left(\frac{343,5 - 186,75}{\sqrt{140,06}}\right) - \Phi\left(\frac{-0,5 - 186,75}{\sqrt{140,06}}\right)\right) \\ &= 1,09491 \cdot 10^{-56} = 0,00000000\ldots000109491\ldots \end{aligned}$$

Vrt. Suoraan binomijakauman perusteella laskettu arvo on $1,19478 \cdot 10^{-35}$.