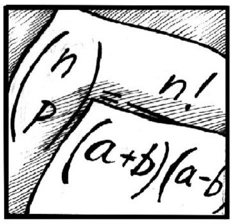

Binomijakaumasta

Markku Halmetoja
Mäntän lukio

Binomijakautuneen satunnaismuuttujan odotusarvon ja varianssin laskeminen suoraan määritelmiin tukeutuen on laskuteknisesti haastava tehtävä ja lukion oppikirjoissa ne yleensä annetaankin valmiina kaavoina. Tässä kirjoituksessa esitetään differentiaalilaskentaa soveltava oikotie kyseisten parametrien johtamiseksi. Lukijan mukavuutta ajatellen kerrataan aluksi eräitä yleisiä käsitteitä.

Lukuja

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}, \quad k = 0, \ldots, n$$

kutsutaan binomikertoimiksi, sillä jos n on positiivinen kokonaisluku, niin

$$(1+x)^n = \sum_{k=0}^{n} \binom{n}{k} x^k. \tag{1}$$

Binomikaava (1) todistuu mukavasti esimerkiksi induktiolla binomikertoimien yhteenlaskusääntöä

$$\binom{n}{k-1} + \binom{n}{k} = \binom{n+1}{k} \tag{2}$$

soveltaen. Yhtälön (1) avulla saadaan jatkossa tarvittava yleisempi binomikaava

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}. \tag{3}$$

Aktiivinen lukija suorittaa yhtälöihin (1), (2) ja (3) liittyvät todistukset yksityiskohtaisesti tai katsoo ne Pekka Alestalon kirjoituksesta [1].

Satunnaismuuttujan X todennäköisyysjakaumasta ilmenee, millä todennäköisyydellä muuttuja saa minkäkin arvon. Jos muuttujan arvojoukko on äärellinen, niin jakauma esitetään usein taulukkona

$X:$	x_1	$\ldots$	x_k	$\ldots$	x_n
$p:$	p_1	$\ldots$	p_k	$\ldots$	p_n

missä siis $\mathrm{P}(X = x_k) = p_k$ ja $p_1 + \ldots + p_n = 1$.

Todennäköisyydet p_k on mukava ajatella x-akselin pisteisiin x_k asetetuiksi massoiksi, joiden kokonaismäärä on 1. Tällöin satunnaismuuttuja odotusarvo

$$\mu = \mathbb{E}X = \sum_{k=1}^{n} x_k \cdot \mathrm{P}(X = x_k) = p_1 x_1 + \ldots + p_n x_n$$

on massan painopiste ja varianssi

$$\sigma^2 = \mathbb{D}^2 X = (x_1 - \mu)^2 p_1 + \ldots + (x_n - \mu)^2 p_n$$

sen hitausmomentti painopisteen suhteen. Varianssi on siis varsin luonnollinen mittari kertomaan todennäköisyyden keskittymisestä odotusarvon läheisyyteen. Aktiivinen lukija muuntaa varianssin kaavan jatkossa tarvittavaan muotoon

$$\sigma^2 = -\mu^2 + \sum_{k=1}^{n} p_k x_k^2. \tag{4}$$

Kerrataan lyhyesti myös binomitodennäköisyys. Siihen tullaan suorittamalla n riippumatonta toistoa kokeesta, jonka tulosmahdollisuudet ovat A ja $\bar{A}$. Jos A:n esiintymistodennäköisyys yksittäisessä toistossa on p ja satunnaismuuttuja X on A:n esiintymiskertojen lukumäärä n:ssä toistossa, niin

$$\mathrm{P}\,(X=k)=\binom{n}{k}p^k(1-p)^{n-k},\quad k=0,1,\ldots,n.$$

Tällöin sanotaan, että satunnaismuuttuja X on binomijakautunut parametrein n ja p, ja merkitään

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

Kaavaa (3) soveltaen nähdään, että binomitodennäköisyyksien summa

$$\sum_{k=0}^{n}\binom{n}{k}p^k(1-p)^{n-k}=(p+1-p)^n=1.$$

Binomijakautuneen muuttujan odotusarvo

$$\begin{aligned}\mu&=\sum_{k=0}^{n}k\cdot\mathrm{P}(X=k)\\&=\sum_{k=0}^{n}k\cdot\binom{n}{k}p^k(1-p)^{n-k},\end{aligned}\tag{5}$$

ja varianssin määrittämisessä kaavaa (4) soveltaen on laskettava summa

$$\sum_{k=0}^{n}k^2\cdot\binom{n}{k}p^k(1-p)^{n-k}.\tag{6}$$

Aktiivinen lukija tutkii millaisiin komplikaatioihin summien (5) ja (6) käsittely johtaa, sillä siten saa perspektiiviä seuraavaan yksinkertaiseen esitykseen.

Olkoon X diskreetti satunnaismuuttuja, jolla on äärellinen arvojoukko $\{x_1,\ldots,x_n\}$ ja $\mathrm{P}(X=x_k)=p_k$. Määritellään sen *generoiva funktio* g seuraavasti:

$$g(t)=p_1t^{x_1}+\ldots+p_nt^{x_n},\quad (t>0).$$

Derivoimalla saadaan

$$g'(t)=p_1x_1t^{x_1-1}+\ldots+p_nx_nt^{x_n-1},$$

joten

$$\mu=\mathbb{E}X=g'(1).$$

Varianssin laskemisessa tarvitaan g:n toinen derivaatta

$$g''(t)=p_1x_1(x_1-1)t^{x_1-2}+\ldots+p_nx_1(x_n-1)t^{x_n-2}.$$

Sen arvo pisteessä $t=1$

$$g''(1)=\sum_{k=1}^{n}p_1x_k(x_k-1)=\sum_{k=1}^{n}p_kx_k^2-g'(1),$$

joten

$$\sum_{k=1}^{n}p_kx_k^2=g'(1)+g''(1).$$

Tämän avulla varianssin kaava (4) saadaan muotoon

$$\begin{aligned}\sigma^2&=-\mu^2+\sum_{k=1}^{n}p_kx_k^2\\&=-g(1)^2+g'(1)+g''(1)\\&=g''(1)+g'(1)-g(1)^2.\end{aligned}$$

Binomijakauman generoiva funktio

$$\begin{aligned}g(t)&=\sum_{k=0}^{n}\mathrm{P}(X=k)\cdot t^k\\&=\sum_{k=0}^{n}\binom{n}{k}p^k(1-p)^{n-k}\cdot t^k\\&=\sum_{k=0}^{n}\binom{n}{k}(pt)^k(1-p)^{n-k}\\&=(pt+1-p)^n.\end{aligned}$$

Sen derivaattojen

$$\begin{aligned}g'(t)&=np(pt+1-p)^{n-1}\\g''(t)&=n(n-1)p^2(pt+1-p)^{n-2}\end{aligned}$$

avulla saadaan muuttujan X odotusarvo

$$\mu=\mathbb{E}X=g'(1)=np$$

ja varianssi

$$\begin{aligned}\sigma^2=\mathbb{D}^2X&=g''(1)+g'(1)-g(1)^2\\&=n(n-1)p^2+np-(np)^2\\&=(np)^2-np^2+np-(np)^2\\&=np-np^2\\&=np(1-p).\end{aligned}$$

Viitteet

[1] Pekka Alestalo, Binomikaava, Solmu 2/2013, `http://solmu.math.helsinki.fi/2013/2/binomi.pdf`