MS-A0504 Todennäköisyyslaskennan ja tilastotieteen peruskurssi

2A Odotusarvo ja muunnokset

Pekka Pere

matematiikan ja systeemianalyysin laitos
perustieteiden korkeakoulu
Aalto-yliopisto

lukuvuosi 2023–2024
periodi I

Sisältö

Odotusarvo

Diskreetin lukuarvoisen satunnaismuuttujan X odotusarvo on

$$\mathbb{E}(X) = \sum_{x \in \mathcal{X}} x \mathbb{P}(X = x) = \sum_{x \in \mathcal{X}} x\, f(x)$$

missä $\mathcal{X}$ tarkoittaa X:n mahdollisten arvojen joukkoa.

Odotusarvo on X:n mahdollisten arvojen todennäköisyyksillä $f(x) = \mathbb{P}(X = x)$ painotettu summa.

Esim (Noppa)

Nopanheiton tuloksen X odotusarvo on

$$\mathbb{E}(X) = (1 \times \frac{1}{6}) + (2 \times \frac{1}{6}) \cdots + (6 \times \frac{1}{6}) = 3.5.$$

Mitä odotusarvo kertoo satunnaismuuttujasta X?
(Ei ainakaan "odotettua arvoa", koska noppa ei koskaan saa arvoa 3.5.)

Odotusarvon tulkinta: Pitkän sarjan keskiarvo

Pelataan n kierrosta peliä, jossa yhden kierroksen tuotto on X.
Tiheysfunktio $f(x) = \mathbb{P}(X = x)$.

Oletus: Tulos x esiintyy pelissä likimain $n\,f(x)$ kertaa.

- Tällöin tuotto n kierrokselta on likimain

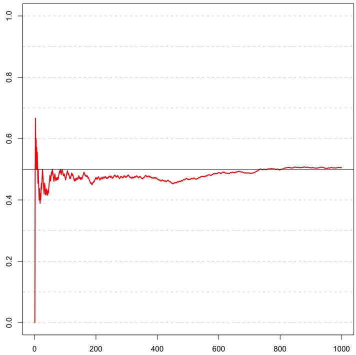
$$\sum_{x \in \mathcal{X}} x\, n\, f(x).$$

- Keskimääräinen tuotto per kierros on likimain

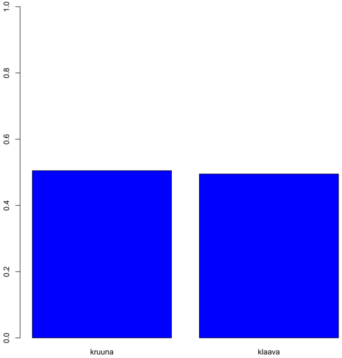
$$\frac{1}{n} \sum_{x \in \mathcal{X}} x\, n\, f(x) \;=\; \sum_{x \in \mathcal{X}} x\, f(x) \;=\; \mathbb{E}(X).$$

Mutta pitääkö oletus paikkansa?

Esimerkki: 1000 kolikkoa



Kruunan suhteellinen osuus heittojen määrän kasvaessa

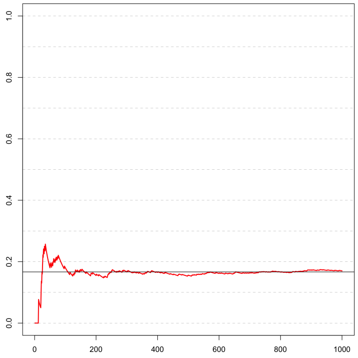


Kruunan ja klaavan suhteelliset osuudet 1000 heitossa.

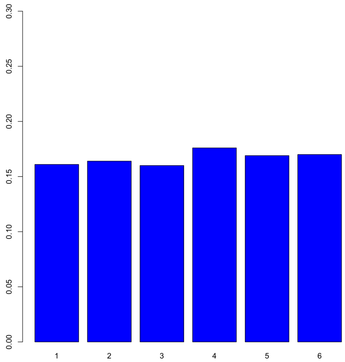
```
n <- 1000
x <- sample(c(0,1),n,replace=TRUE)
plot(cumsum(x)/(1:n),type="l")
plot(table(x))
```

http://www.r-project.org/
http://www.random.org/

Esimerkki: 1000 noppaa



Kuutosen suhteellinen osuus heittojen lukumäärän funktiona



Silmälukujen suhteelliset osuudet 1000 heitossa

```
n <- 1000
x <- sample(1:6,n,replace=TRUE)
plot(cumsum(x==6)/(1:n),type="l")
plot(table(x))
```

http://www.r-project.org/
http://www.random.org/

Sisältö

Satunnaismuuttujan odotusarvo vs. pitkän ajan keskiarvo

Lause (Suurten lukujen laki)

Jos $X_1, X_2, X_3, \ldots$ ovat keskenään riippumattomia X:n tavoin jakautuneita satunnaismuuttujia, niin tapahtuman

$$\frac{1}{n}\sum_{i=1}^{n} X_i \;=\; \mathbb{E}(X) \pm 0.001$$

todennäköisyys lähestyy ykköstä suurilla n arvoilla.

Tämä on stokastiikan tärkein teoreema, jonka mukaan keskiarvon satunnaisuus katoaa suurilla n:

- Keskiarvo $\frac{1}{n}\sum_{i=1}^{n} X_i$ on satunnaismuuttuja
- Odotusarvo $\mathbb{E}(X)$ on deterministinen luku
- 0.001 voidaan korvata millä tahansa $\epsilon > 0$.

Päteekö tulos keskenään riippuville satunnaismuuttujille?

Kyllä, jos riippuvuus on riittävän heikkoa (ergodisuus).

Lukumäärän laskenta = indikaattorien summaus

Olkoon meillä keskenään samoin jakautuneita ja riippumattomia satunnaismuuttujia $X_1, X_2, X_3, \ldots$, esim. nopanheittoja.

Haluamme tutkia *montako kertaa* tulos kuului tiettyyn joukkoon, esim. joukkoon $B = \{2, 4, 6\}$.

Työkaluna voimme käyttää indikaattorimuuttujia

$$I_i = \begin{cases} 1, & \text{jos } X_i \in B \\ 0, & \text{jos } X_i \notin B \end{cases}$$

Nyt lukumäärän laskenta onkin yhteenlaskua:

i	1	2	3	4	5	6	7	8	9	10	
X_i	3	6	2	2	1	2	6	4	6	1	$\#\{i : X_i \in B\} = 7$
I_i	0	1	1	1	0	1	1	1	1	0	$\sum I_i = 7$

Tapahtuman todennäköisyys vs. suhteellinen esiintyvyys

Lause

Jos $X_1, X_2, \ldots$ ovat riippumattomia X:n tavoin jakautuneita satunnaismuuttujia, niin minkä tahansa arvojoukon B suhteellinen esiintyvyys listassa $(X_1, \ldots, X_n)$ toteuttaa

$$\frac{\#\{s \in \{1, 2, \ldots, n\} \ : \ X_s \in B\}}{n} \ = \ \mathbb{P}(X \in B) \pm 0.001$$

todennäköisyydellä, joka lähestyy ykköstä suurilla n.

- Diskreetille tiheysfunktiolle $f(x) = \mathbb{P}(X = x)$:

$$\frac{\#\{s \ : \ X_s = x\}}{n} \ \approx \ f(x)$$

- Kertymäfunktiolle $F(t) = \mathbb{P}(X \leq t)$:

$$\frac{\#\{s \ : \ X_s \leq t\}}{n} \ \approx \ F(t)$$

Todennäköisyys vs. suhteellinen esiintyvyys: Todistus

Arvojoukon B suhteellinen esiintyvyys voidaan kirjoittaa muodossa

$$\frac{1}{n}\sum_{s=1}^{n} I_s, \quad \text{missä} \quad I_s = \begin{cases} 1, & \text{jos } X_s \in B, \\ 0, & \text{muuten,} \end{cases}$$

on tapahtuman $\{X_s \in B\}$ indikaattorimuuttuja.

Satunnaismuuttujat $I_1, I_2, \ldots$ ovat toisistaan riippumattomia ja tapahtuman $\{X \in B\}$ indikaattorimuuttujan I kanssa samoin jakautuneita (miksi?).

Suurten lukujen lain perusteella, kun $n \to \infty$,

$$\frac{1}{n}\sum_{s=1}^{n} I_s \approx \mathbb{E}(I) = 0 \times \mathbb{P}(I = 0) + 1 \times \mathbb{P}(I = 1) = \mathbb{P}(X \in B).$$

Esimerkki: Nopan tutkiminen empiirisesti

Yritetään kokeilemalla tutkia todennäköisyyttä $\mathbb{P}(X \leq 2)$, missä X on nopanheiton (tietokoneella simuloitu) tulos.

n	suht. esiintyvyys	käytetty laskenta-aika
100	0.38000000	0.00 s
10000	0.33260000	0.00 s
1e+06	0.33351000	0.02 s
1e+08	0.33332494	1.55 s
1e+10	0.33333081	159.33 s

Tässä tiedämme oikean todennäköisyyden, joten näemme kuinka oikeita desimaaleja kertyy (virhe pienenee).

Oikeissa sovelluksissa emme yleensä tiedä todennäköisyyttä, vaan yritämme arvioida sitä. Olisi myös kiva pystyä arvioimaan virheen suuruutta, vaikka oikeaa arvoa ei tiedetä. Tähän saadaan myöhemmin työkaluja.

Todennäköisyys vs. suhteellinen esiintyvyys: Käyttö

Saimme keinon kokeellisesti arvioida jonkin tapahtuman todennäköisyyttä, kunhan vain voimme toistaa sitä monta kertaa.

Onko tässä tn-laskennan Graalin malja? Hankalien kaavojen sijaan toistetaan koetta ja katsotaan, miten usein tapahtuma tapahtui?

Tavallaan, mutta

- tarvitsemme menetelmän toistaa koetta monta kertaa (oikeasti tai simuloidusti)
- oikea kokeileminen voi olla vaikeaa, kallista, vaarallista
- simuloitu koe ei ehkä vastaa todellisuutta
- suuri tarkkuus edellyttää paljon toistoja: tn-arvion virhe on karkeasti suhteessa lukuun $1/\sqrt{n}$, joten yksi oikea lisädesimaali vaatii 100 kertaa enemmän toistoja!

Todennäköisyys vs. suhteellinen esiintyvyys: Käyttö

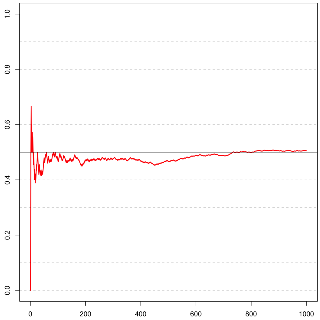
Kutenkin se, että esiintyvyys pitkässä sarjassa on hyvä estimaatti todennäköisyydelle, on pohjana suurelle osalle modernia tilastotiedettä.

- otanta: poimimme väestöstä n ihmistä, heistä k:lla on diabetes, arvioimme osuuden k/n pätevän koko väestössä
- lääketieteellinen koe: kokeilemme hoitoa n kertaa, se onnistuu k kertaa, arvioimme jatkossakin sen onnistuvan tn:llä k/n
- jakauman estimointi empiirisellä histogrammilla
- Monte Carlo -menetelmät fysiikassa, biologiassa jne. pohjautuvat siihen, että koetta simuloidaan tietokoneella monta (miljoonaa) kertaa ja katsotaan esiintyvyys. (Vaikeaa on suunnitella se simulaatio.)
- Monte Carlo -integroinnissa heitetään tikkaa ja katsotaan, miten usein osutaan tiettyyn joukkoon $\to$ arvio ko. joukon pinta-alalle

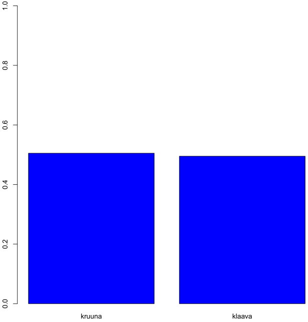
Esimerkki: 1000 kolikkoa

Suurten lukujen lain perusteella kruunan suhteellinen esiintyvyys satunnaisjonossa $(X_1, \ldots, X_n)$ on

$$\frac{\#\{s \leq n : X_s = \text{“kruuna”}\}}{n} \approx \frac{1}{2}$$



Kruunan suhteellinen osuus heittojen määrän kasvaessa

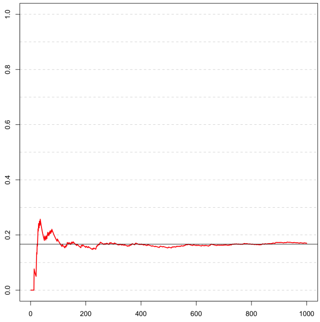


Kruunan ja klaavan suhteelliset osuudet 1000 heitossa

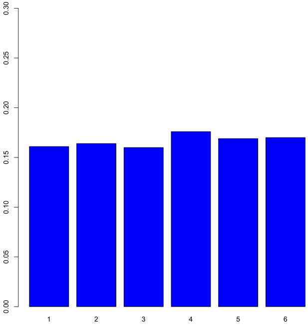
Esimerkki: 1000 noppaa

Suurten lukujen lain perusteella 6:n suhteellinen esiintyvyys satunnaisjonossa $(X_1, \ldots, X_n)$ on

$$\frac{\#\{s \leq n : X_s = 6\}}{n} \approx \frac{1}{6}$$



Kuutosen suhteellinen osuus heittojen lukumäärän funktiona



Silmälukujen suhteelliset osuudet 1000 heitossa

Lisää tarinoita nopista

Zacariach Labby: Weldon's dice, automated
https://www.youtube.com/watch?v=95EErdouO2w
https://link.springer.com/article/10.1007/s00144-009-0036-8

Prof. Samuli Siltanen:
Samun tiedepläjäys: arpakuutio ja todennäköisyyden olemus
https://www.youtube.com/watch?v=rkJv4BveY4g

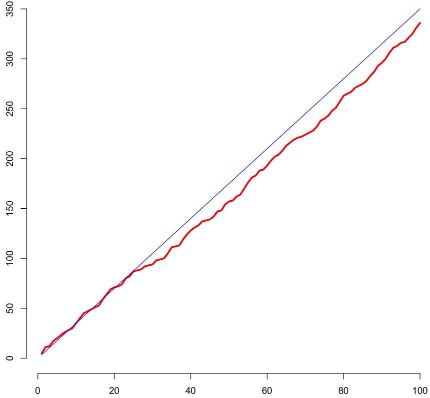
Tuomas Kukko & Risto Heikkinen:
Kimblen noppa ei ole täysin satunnainen
http://statistition.com/?p=440

Esimerkki: Noppapelin tuottokertymä

Noppapelissä voittaa kierroksella i silmäluvun X_i verran euroja. Yhden kierroksen tuoton odotusarvo on $\mathbb{E}(X_i) = 3.5$ EUR.

Tuotto suurelta määrältä n kierroksia on suurten lukujen lain mukaan likimain

$$\sum_{i=1}^{n} X_i = \left(\frac{1}{n} \sum_{i=1}^{n} X_i \right) n \approx 3.5n.$$



Odotusarvo vs. keskiarvo: Yhteenveto

Satunnaismuuttujan X odotusarvolle ja todennäköisyyksille on saatu tulkinnat

$$\mathbb{E}(X) \approx \frac{1}{n} \sum_{s=1}^{n} X_s,$$

$$\mathbb{P}(X = x) \approx \frac{\#\{s \le n : X_s = x\}}{n},$$

missä $X_1, X_2, \ldots$ ovat riippumattomia ja samoin jakautuneita.

Entä jos riippumattomia toistoja ei ole saatavilla?

- $X =$ startup-yhtiön seuraavan vuoden liikevaihto
- $X =$ taloyhtiön materiaalivahingot tulipaloista

Tällöinkin odotusarvolla on jokin merkitys, mutta “pitkän ajan keskiarvo” ei kovin mielekäs.

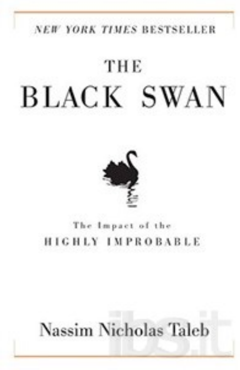
Esimerkki: "Musta joutsen"

Taulukon

k	0	1000000
$\mathbb{P}(X = k)$	0.999999	0.000001

mukaan jakautuneen satunnaismuuttujan odotusarvo on

$$\begin{aligned} \mathbb{E}(X) &= 0 \times 0.999999 + 1000000 \times 0.000001 \\ &= 1. \end{aligned}$$



Odotusarvo $\mathbb{E}(X) = 1$ kertoo hyvin vähän satunnaisilmiöstä?

Jos X:n tavoin jakautuneita satunnaislukuja generoidaan toisistaan riippumattomasti, niin 10000 ensimmäistä satunnaislukua ovat kaikki nollia todennäköisyydellä $0.999999^{10000} \approx 99\%$.

http://www.fooledbyrandomness.com/

Sisältö

Jatkuvan satunnaismuuttujan diskretointi

Diskreetti satunnaisluku $\lfloor X \rfloor_k = \frac{\lfloor 10^k X \rfloor}{10^k}$ on X:n arvo pyöristettynä alaspäin k desimaalin tarkkuuteen (esim. $\lfloor 1.52793 \rfloor_3 = 1.527$).

$$\begin{aligned}
\mathbb{E}(\lfloor X \rfloor_k) &= \sum_{i=-\infty}^{\infty} \frac{i}{10^k} \mathbb{P}\left(\lfloor X \rfloor_k = \frac{i}{10^k} \right) \\
&= \sum_{i=-\infty}^{\infty} \frac{i}{10^k} \mathbb{P}\left(\frac{i}{10^k} \le X < \frac{i+1}{10^k} \right) \\
&= \sum_{i=-\infty}^{\infty} \frac{i}{10^k} \int_{\frac{i}{10^k}}^{\frac{i+1}{10^k}} f(x)dx \;=\; \int_{-\infty}^{\infty} \lfloor x \rfloor_k f(x)dx.
\end{aligned}$$

Koska $\lfloor X \rfloor_k \to X$ kun tarkkuus $k \to \infty$, määritellään

$$\mathbb{E}(X) = \lim_{k\to\infty} \mathbb{E}(\lfloor X \rfloor_k) \;=\; \lim_{k\to\infty} \int_{-\infty}^{\infty} \lfloor x \rfloor_k f(x)dx \;=\; \int_{-\infty}^{\infty} x\, f(x)dx.$$

Jatkuvan satunnaismuuttujan odotusarvo

Jatkuvan satunnaismuuttujan X odotusarvo määritellään kaavalla

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

Kyseessä on jatkuva versio ideasta "mahdollisten arvojen tiheyspainotettu keskiarvo".

Esim (Metro)

Jos seuraavan metron saapumiseen kuluva aika X noudattaa välin $[0, 10]$ tasajakaumaa tiheysfunktiona

$$f(x) = \begin{cases} \frac{1}{10}, & x \in (0, 10), \\ 0, & \text{muuten}, \end{cases}$$

niin

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x\, f(x) dx = \int_{0}^{10} x\, \frac{1}{10} dx = 5.$$

Jatkuva odotusarvo — Lisäesimerkkejä

Esim (Polynomi tiheysfunktiona)

Printterin korjausaika X (tunteina) on jatkuvasti jakautunut tiheysfunktiolla $f(x) = 2x$, kun $0 < x < 1$.

Korjausaika on siis aina välillä 0—1 tuntia, mutta todennäköisemmin lähellä välin yläpäätä (tiheys on siellä suurempi).

Korjausajan odotusarvo on

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x\, f(x) dx = \int_0^1 x\; 2x\; dx = \int_0^1 2x^2 dx = 2/3.$$

Jatkuva odotusarvo — Lisäesimerkkejä

Esim (Eksponenttijakauma)

Hyönteisiä osuu tuulilasiin siten, että osumien väliajat X ovat eksponenttijakautuneet taajuusparametrilla $\lambda = 1$ kpl minuutissa.
Tiheysfunktio on

$$f(x) = e^{-x}$$

kun $x > 0$, ja nolla muualla.
Nyt väliajan odotusarvo on

$$\mathbb{E}(X) \;=\; \int_{-\infty}^{\infty} x\ f(x)dx \;=\; \int_{0}^{\infty} x\ e^{-x}dx \;=\; 1.$$

(Integraalin laskemiseen tarvittiin osittaisintegrointia.)

Tulkinta pitkän ajan keskiarvona: Jos odotetaan, että hyönteisiä on kertynyt 50 kpl, niin keskimääräinen väliaika on SLL:n nojalla *noin* 1 minuutti, ja aikaa on siis yhteensä kulunut *noin* 50 minuuttia.

Satunnaismuuttujan odotusarvo: Yhteenveto

Diskreetti satunnaismuuttuja

- Esim. joukon $\{1, \ldots, 6\}$ tasajakauma, binomijakauma

$$\mathbb{P}(X \in A) = \sum_{x \in A} f(x)$$

$$\mathbb{E}(X) = \sum_{x \in \mathcal{X}} x\, f(x)$$

Jatkuva satunnaismuuttuja

- Esim. välin $[0, 10]$ tasajakauma, eksp. jakauma

$$\mathbb{P}(X \in A) = \int_A f(x)\, dx$$

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

Sisältö

Esimerkki: Diskreetin satunnaismuuttujan neliö

Tehtävä (Vrt. viime luennon neliölaatat)

Laske $\mathbb{E}(X^2)$, kun X:n jakauma on

k	0	1	2
$\mathbb{P}(X = k)$	0.2	0.5	0.3

Ratkaisu

$Y = X^2$ on diskreetti satunnaismuuttuja arvojoukkona $\{0, 1, 4\}$ ja jakaumana

k	0	1	4
$\mathbb{P}(Y = k)$	0.2	0.5	0.3

Näin ollen

$$\mathbb{E}(X^2) = \mathbb{E}(Y) = 0 \times 0.2 + 1 \times 0.5 + 4 \times 0.3 = 1.7.$$

Esimerkki: Jatkuvan satunnaismuuttujan kuutio

Kone tekee kuutioita, joiden sivu on tasaisesti jakautunut 0 ja 10 cm välillä.

Tehtävä

Laske $\mathbb{E}(X^3)$, kun X noudattaa välin $[0, 10]$ tasajakaumaa.

Ratkaisu

Satunnaismuuttujan $Y = X^3$ arvojoukon pisteissä $t \in [0, 1000]$,

$$F_Y(t) \;=\; \mathbb{P}(Y \le t) \;=\; \mathbb{P}(X^3 \le t) \;=\; \mathbb{P}(X \le t^{1/3}) \;=\; \frac{t^{1/3}}{10}.$$

$$\text{Tiheysfunktio} \quad f_Y(t) \;=\; \begin{cases} \frac{t^{-2/3}}{30}, & 0 < t < 1000, \\ 0, & \text{muuten.} \end{cases}$$

$$\mathbb{E}(X^3) \;=\; \mathbb{E}(Y) \;=\; \int_0^{1000} t\,\frac{t^{-2/3}}{30}dt \;=\; \frac{1}{30}\int_0^{1000} t^{1/3}dt$$

$$=\; \frac{1}{30}\Big|_0^{1000}\frac{3}{4}t^{4/3} \;=\; \frac{1000^{4/3}}{40} \;=\; 250.$$

Odotusarvon muunnoskaava

Jos g on funktio satunnaismuuttujan X arvojoukosta reaaliluvuille, niin $Y = g(X)$ on satunnaismuuttuja, joka liittää satunnaisilmiön toteumaan s luvun $g(X(s))$.

Fakta

- *Diskreetille satunnaismuuttujalle*

$$\mathbb{E}(g(X)) \; = \; \sum_x g(x)\, f(x).$$

- *Jatkuvalle satunnaismuuttujalle*

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

Esimerkki: Diskreetin satunnaismuuttujan neliö

Tehtävä

Laske $\mathbb{E}(X^2)$, kun X:n jakauma on

k	0	1	2
$\mathbb{P}(X = k)$	0.2	0.5	0.3

Ratkaisu

Soveltamalla odotusarvon muunnoskaavaa funktioon $g(k) = k^2$,

$$\mathbb{E}(X^2) \;=\; \sum_k k^2 f(k) \;=\; 0^2 \times 0.2 + 1^2 \times 0.5 + 2^2 \times 0.3 \;=\; 1.7.$$

Esimerkki: Jatkuvan satunnaismuuttujan kuutio

Tehtävä

Laske $\mathbb{E}(X^3)$, kun X noudattaa välin $[0, 10]$ tasajakaumaa.

Ratkaisu

Soveltamalla odotusarvon muunnoskaavaa funktioon $g(t) = t^3$,

$$\mathbb{E}(X^3) \;=\; \int_{-\infty}^{\infty} t^3 \, f(t) dt \;=\; \int_0^{10} t^3 \, \frac{1}{10} dt \;=\; \frac{1}{10} \Big|_0^{10} \frac{1}{4} t^4 \;=\; 250.$$

Huom. Sama tulos kuin aiemmin, mutta muunnoskaavan avulla paljon helpommin.

Siirretty ja skaalattu satunnaismuuttuja

Tässä pienet kirjaimet ovat vakioita ja isot kirjaimet satunnaismuuttujia.

Fakta

(i) $\mathbb{E}(a) = a$.

(ii) $\mathbb{E}(bX) = b\mathbb{E}(X)$.

(iii) $\mathbb{E}(X + a) = \mathbb{E}(X) + a$.

Todistus.

Jos X on diskreetti, (ii) saadaan käyttämällä muunnoskaavaa funktioon $g(x) = bx$:

$$\mathbb{E}(bX) = \sum_x (bx) f(x) = b \sum_x x f(x) = b\mathbb{E}(X).$$

Samoin (iii) käyttämällä funktiota $g(x) = x + a$ (taululla).
(i) saadaan yhdistämällä (ii) ja (iii) käyttäen arvoa $b = 0$.
Jos X on jatkuva, sama OK vaihtamalla summat integraaleiksi. $\square$

Monen satunnaismuuttujan muunnoksen odotusarvo

Fakta

- *Diskreeteille satunnaismuuttujille X ja Y, joiden yhteisjakaumalla on tiheysfunktio $f(x,y)$,*

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

- *Jatkuville satunnaismuuttujille X ja Y, joiden yhteisjakaumalla on tiheysfunktio $f(x,y)$,*

$$\mathbb{E}(g(X,Y)) \; = \; \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y)\, f(x,y) dxdy.$$

Monen satunnaismuuttujan muunnoksen odotusarvo

Esimerkki (Laatikko, jossa kaksi diskreettiä mittaa)

Kone tekee laatikoita, joiden pohja on X-sivuinen neliö ja korkeus on H. Laatikon tilavuus on siis $g(X, H) = X^2 H$.

Pohjan sivu on 10 tai 20, ja korkeus on 3 tai 5, yhteisjakaumalla

	$H = 3$	$H = 5$
$X = 10$	0.4	0.3
$X = 20$	0.2	0.1

Tilavuuden odotusarvo on

$$\begin{aligned}
\mathbb{E}(g(X, H)) &= g(10, 3)f(10, 3) + g(10, 5)f(10, 5) \\
&\quad + g(20, 3)f(20, 3) + g(20, 5)f(20, 5) \\
&= (300 \times 0.4) + (500 \times 0.3) + (1200 \times 0.2) + (2000 \times 0.1) \\
&= 710.
\end{aligned}$$

Sisältö

Summman odotusarvo

Fakta

$\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$.

Todistus.

Jos X ja Y ovat diskreettejä (jatkuva tapaus samaan tapaan), soveltamalla odotusarvon muunnoskaavaa funktioon $g(x, y) = x + y$,

$$
\begin{aligned}
\mathbb{E}(X+Y) &= \sum_x \sum_y (x+y)\, f(x,y) \\
&= \sum_x \sum_y x\, f(x,y) + \sum_x \sum_y y\, f(x,y) \\
&= \sum_x x \left(\sum_y f(x,y) \right) + \sum_y y \left(\sum_x f(x,y) \right) \\
&= \sum_x x\, f_X(x) + \sum_y y\, f_Y(y) \\
&= \mathbb{E}(X) + \mathbb{E}(Y).
\end{aligned}
$$

$\square$

Usean sm:n summa

Pidemmän summan odotusarvo saadaan käyttämällä kaavaa $\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$ monta kertaa.

Kolmen sm:n summa $X + Y + Z$. Annetaan sm:lle $X + Y$ lyhennysnimi U.

$$\begin{aligned}\mathbb{E}(X+Y+Z) &= \mathbb{E}(U+Z)\\ &= \mathbb{E}(U) + \mathbb{E}(Z)\\ &= \mathbb{E}(X+Y) + \mathbb{E}(Z)\\ &= \mathbb{E}(X) + \mathbb{E}(Y) + \mathbb{E}(Z).\end{aligned}$$

Jatkamalla samaan tapaan saadaan n-pituiselle summalle

$$\mathbb{E}(X_1 + \ldots + X_n) = \mathbb{E}(X_1) + \ldots + \mathbb{E}(X_n).$$

Toisin sanoen: odotusarvon voi ottaa termeittäin. Tätä (ja skaalausta vakiokertoimella) kutsutaan odotusarvon lineaarisuudeksi.

Esimerkki: Binomijakauma

Olkoon n riippumatonta indikaattorimuuttujaa $I_1, \ldots, I_n$, kukin indikoi onnistumista (1) tai epäonnistumista (0), onnistumisen tn $P(I_i = 1) = p$ ja epäonnistumisen $q = 1 - p$.
Silloin $X = \sum_{i=1}^{n} I_i$, eli onnistumisten lukumäärä, on binomijakautunut.
Miten laskea $\mathbb{E}(X)$? Voisit käyttää kaavaa $\sum_x xf(x)$, mutta se on työlästä. Helpompi tapa: odotusarvot indikaattorimuuttujista termeittäin.

$$\begin{aligned}
\mathbb{E}(X) &= \mathbb{E}(I_1 + I_2 + \ldots + I_n) \\
&= \mathbb{E}(I_1) + \mathbb{E}(I_2) + \ldots + E(I_n) \\
&= p + p + \ldots p \\
&= np.
\end{aligned}$$

Esim. $n = 100$ yritystä, $p = 0.20$ onnistumistn $\implies$ onnistumisten lukumäärän odotusarvo on $np = 20$.

Odotusarvolle ei voi tehdä mitä tahansa

Edellä nähtiin, että joitakin "lineaarisia" operaatioita voi "siirtää" odotusarvon sisältä ulos ja päinvastoin: esim. $\mathbb{E}(bX) = b\mathbb{E}(X)$ ja $E(X+Y) = E(X) + E(Y)$.

Mitä tahansa operaatioita ei suinkaan voi näin siirrellä!

Esimerkki

Aiempi kuutioita tekevä kone, sivu X tasaisesti jakautunut 0 ja 10 cm välillä. Edellä laskimme, että $\mathbb{E}(X^3) = 250$.

Sen sijaan $(E(X))^3 = 5^3 = 125 \neq E(X^3)$.

(Odotusarvion kuutio ei ole kuution odotusarvo.)

Yhteenveto

Satunnaismuuttujan odotusarvo $\mathbb{E}(X)$ antaa likiarvon keskiarvolle, joka lasketaan suuresta määrästä X:n kanssa samoin jakautuneita riippumattomia satunnaismuuttujia.

Diskreetti satunnaismuuttuja

$$\mathbb{E}(X) = \sum_x x\, f(x)$$

$$\mathbb{E}(g(X)) = \sum_x g(x)\, f(x)$$

Jatkuva satunnaismuuttuja

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

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

$$\mathbb{E}\left(a + \sum_{i=1}^{n} b_i X_i\right) = a + \sum_{i=1}^{n} b_i \mathbb{E}(X_i)$$

Sisältö

Pietarin paradoksi

Kasinolla on tarjolla uhkapeli, jossa kolikkoa heitetään kunnes saadaan klaava. Pelin tuotto on

- 2 EUR, jos ensimmäinen klaava ilmestyy 1. heitolla
- 4 EUR, jos ensimmäinen klaava ilmestyy 2. heitolla
- 8 EUR, jos ensimmäinen klaava ilmestyy 3. heitolla
- ...

Paljonko olisit valmis maksamaan oikeudesta osallistua peliin?

Pelin tuotto on $g(T) = 2^T$, missä pelin kesto T on diskreetti satunnaismuuttuja jakaumana $f_T(k) = (1/2)^k, k = 1, 2, 3, \ldots$ Pelin odotusarvoinen tuotto on

$$\mathbb{E}[g(T)] \;=\; 2^1(1/2)^1 + 2^2(1/2)^2 + 2^3(1/2)^3 + \cdots \;=\; \infty.$$

Vaikka tuotto on varmasti äärellinen, sen odotusarvo on ääretön.

https://en.wikipedia.org/wiki/St._Petersburg_paradox

*Lisätehtävä (yli kurssialueen)

Y = odotusaika (min) asemalla, jonne metroja saapuu 10 min välein, ja jossa metrot pysähtyvät 1 min ajan.

Viime luennolla johdettiin Y:n kertymäfunktio

$$F_Y(t) = \begin{cases} 0, & t < 0, \\ \frac{1}{10} + \frac{t}{10}, & 0 \le t \le 9, \\ 1, & t > 9. \end{cases}$$

ja todettiin, että Y:n jakauma ei ole diskreetti eikä jatkuva vaan niiden sekoitus.

Tehtävä

Kehitä luonteva odotusarvon määritelmä diskreetin ja jatkuvan jakauman sekoituksille ja laske $\mathbb{E}(Y)$.

Seuraavalla kerralla puhutaan keskihajonnasta ja korrelaatiosta…