MS-A0504 Todennäköisyyslaskennan ja tilastotieteen peruskurssi

2A Satunnaismuuttujan odotusarvo

Lasse Leskelä

Matematiikan ja systeemianalyysin laitos
Perustieteiden korkeakoulu
Aalto-yliopisto

Lukuvuosi 2018–2019
Periodi IV

Sisältö

Odotusarvo

Diskreetin lukuarvoisen satunnaismuuttujan X odotusarvo on

$$\mathrm{E}(X) \;=\; \sum_x x\,\mathrm{P}(X = x) \;=\; \sum_x x\,f(x)$$

missä summa käy yli X:n mahdollisten arvojen.

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

Esim (Noppa)

Nopanheiton tuloksen X odotusarvo on

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

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

Odotusarvon tulkinta

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

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

- Tällöin tuotto n kierrokselta on likimain

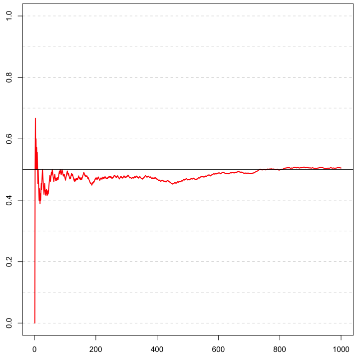
$$\sum_x x\,n\,f(x).$$

- Keskimääräinen tuotto per kierros on likimain

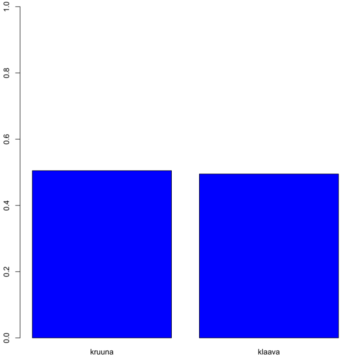
$$\frac{1}{n}\sum_x x\,n\,f(x) \;=\; \sum_x x\,f(x) \;=\; \mathrm{E}(X).$$

Mutta pitääkö oletus paikkansa?

Esimerkki: 1000 kolikkoa



Kruunan suhteellinen osuus heittojen määrän kasvaessa

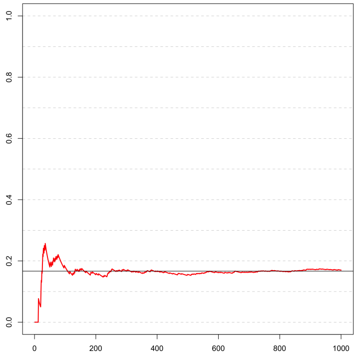


Kruunan ja klaavan suhteelliset osuudet 1000 heitossa.

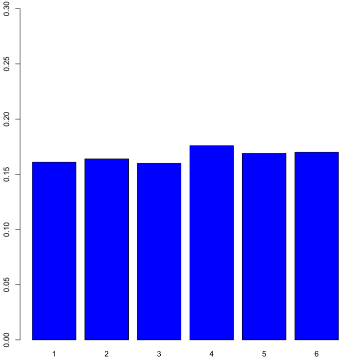
```
n <- 1000
x <- sample(c(0,1),n,replace=TRUE)
plot(cumsum(x)/(1:n),type="l")
plot(table(x))
```

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

Odotusarvo vs. keskiarvo

Lause (Suurten lukujen laki)

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

$$\frac{1}{n}\sum_{s=1}^{n} X_s \;=\; \mathrm{E}(X) \pm 0.001$$

todennäköisyys lähestyy ykköstä suurilla n arvoilla.

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

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

Päteekö tulos keskenään riippuville satunnaisluvuille?
Kyllä, jos riippuvuus on riittävän heikkoa (ergodisuus).

Todennäköisyys vs. suhteellinen esiintyvyys

Lause

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

$$\frac{\#\{s \le n : X_s \in B\}}{n} \;=\; \mathrm{P}(X \in B) \pm 0.001$$

todennäköisyydellä, joka lähestyy ykköstä suurilla n.

- Tiheysfunktiolle $f(x) = \mathrm{P}(X = x)$:

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

- Kertymäfunktiolle $F(t) = \mathrm{P}(X \le t)$:

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

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?).

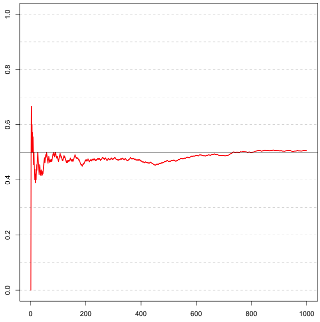
Suurten lukujen lain perusteella, kun $n \to \infty$,

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

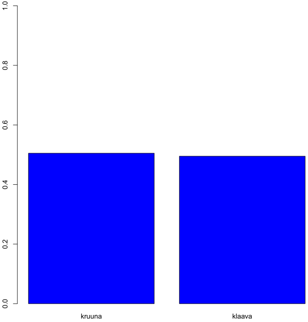
Esimerkki: 1000 kolikkoa

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

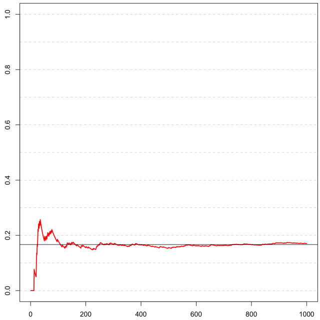


Kruunan ja klaavan suhteelliset osuudet 1000 heitossa

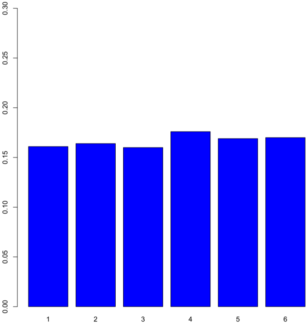
Esimerkki: 1000 noppaa

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

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



Kuutosen suhteellinen osuus heittojen lukumäärän funktiona



Silmälukujen suhteelliset osuudet 1000 heitossa

Lisää tarinoita nopista

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

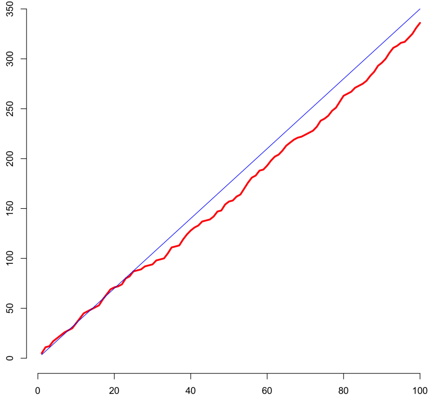
Tuomas Kukko & Risto Heikkinen:
Kimblen noppa ei ole täysin satunnainen
http://statistition.com/?p=440

Esimerkki: Noppapelin tuottokertymä

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

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

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

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

Tällöin odotusarvon $\mathrm{E}(X)$ merkitys on vähäinen?

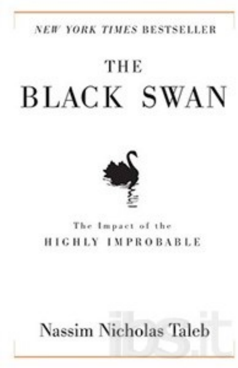
Esimerkki: “Musta joutsen”

Taulukon

k	0	1000000
$P(X = k)$	0.999999	0.000001

mukaan jakautuneen satunnaisluvun odotusarvo on

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



Odotusarvo $E(X) = 1$ 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 satunnaisluvun 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. $\lfloor 1.52793 \rfloor_3 = 1.527$).

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

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

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

Jatkuvan satunnaisluvun odotusarvo

Jatkuvan satunnaisluvun X odotusarvo määritellään kaavalla

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

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

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

Satunnaisluvun odotusarvo: Yhteenveto

Diskreetti satunnaisluku

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

$$\mathrm{P}(X \in A) = \sum_{i \in A} f(i)$$

$$\mathrm{E}(X) = \sum_{x} x\, f(x)$$

Jatkuva satunnaisluku

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

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

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

*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 $\mathrm{E}(Y)$.

Sisältö

Esimerkki: Diskreetin satunnaisluvun neliö

Tehtävä

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

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

Ratkaisu

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

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

Näin ollen

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

Esimerkki: Jatkuvan satunnaisluvun kuutio

Tehtävä

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

Ratkaisu

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

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

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

$$\begin{aligned} \mathrm{E}(X^3) &= \mathrm{E}(Y) = \int_0^{1000} t \, \frac{t^{-2/3}}{30} dt = \frac{1}{30} \int_0^{1000} t^{1/3} dt \\ &= \frac{1}{30} \Big|_0^{1000} \frac{3}{4} t^{4/3} = \frac{1000^{4/3}}{40} = 250. \end{aligned}$$

Odotusarvon muunnoskaava

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

Fakta

- *Diskreetille satunnaismuuttujalle*

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

- *Jatkuvalle satunnaismuuttujalle*

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

Esimerkki: Diskreetin satunnaisluvun neliö

Tehtävä

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

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

Ratkaisu

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

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

Esimerkki: Jatkuvan satunnaisluvun kuutio

Tehtävä

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

Ratkaisu

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

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

Siirretty ja skaalattu satunnaisluku

Fakta

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

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

(iii) $\mathrm{E}(a + bX) = a + b\,\mathrm{E}(X).$

Todistus.

Jos X on diskreetti, funktiolle $g(x) = a + bx$,

$$\begin{aligned}
\mathrm{E}(a + bX) &= \sum_x g(x) f(x) = \sum_x (a + bx) f(x) \\
&= a \sum_x f(x) + b \sum_x x f(x) \\
&= a + b\,\mathrm{E}(X).
\end{aligned}$$

$a = 0 \implies$ (ii). $\quad b = 0 \implies$ (i).

Jos X on jatkuva, sama OK vaihtamalla summat integraaleiksi. $\square$

Monen satunnaismuuttujan muunnoksen odotusarvo

Fakta

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

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

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

$$\mathrm{E}(g(X, Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y) dx dy.$$

Sisältö

Odotusarvon lineaarisuus

Fakta

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

Todistus.

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

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

$\square$

Yhteenveto

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

Diskreetti satunnaisluku

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

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

Jatkuva satunnaisluku

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

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

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

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 satunnaisluku jakaumana $f_T(k) = (1/2)^k, k = 1, 2, 3, \ldots$
Pelin odotusarvoinen tuotto on

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

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

Seuraavalla kerralla puhutaan keskihajonnasta ja korrelaatiosta...