Centrális határeloszlástétel

Tétel (Centrális határeloszlástétel)

Legyenek $X_1, X_2, \ldots$ **független azonos eloszlású** *valószínűségi változók, melyekre* $\mathbb{E}(X_1) = m$ *és* $D(X_1) = \sigma < \infty$*, azaz* **szórásuk véges***. Ekkor tetszőleges* t *valós számra*

$$\mathbb{P}\left(\frac{X_1 + X_2 + \ldots + X_n - n \cdot m}{\sigma\sqrt{n}} \le t\right) \to \mathbb{P}(Z \le t) \qquad (n \to \infty),$$

ahol Z *standard normális eloszlású, azaz*

$$\mathbb{P}(Z \le t) = \Phi(t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx.$$

Ezt úgy is fogalmazhatjuk, hogy

$$\frac{X_1 + X_2 + \ldots + X_n - n \cdot m}{\sigma\sqrt{n}} \to N(0,1)$$

teljesül $n \to \infty$ esetén eloszlásban. Azonos eloszlású: $\mathbb{P}(X_i \le t) = P(X_j \le t)$ minden i, j párra és t valós számra

Centrális határeloszlástétel

Legyenek $X_1, X_2, \ldots$ független azonos eloszlású valószínűségi változók, melyekre $\mathbb{E}(X_1) = m$ és $D(X_1) = \sigma < \infty$. Ekkor

$$\lim_{n\to\infty} \mathbb{P}\left(a \leq \frac{X_1 + X_2 + \ldots + X_n - n \cdot m}{\sigma\sqrt{n}} < b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx.$$

A határértéket $\Phi(b) - \Phi(a) = \mathbb{P}(a \leq Y \leq b)$ alakban is írhatjuk, ahol $Y \sim N(0,1)$.

Centrális határeloszlástétel

Legyenek $X_1, X_2, \ldots$ független azonos eloszlású valószínűségi változók, melyekre $\mathbb{E}(X_1) = m$ és $D(X_1) = \sigma < \infty$. Ekkor

$$\lim_{n\to\infty} \mathbb{P}\left(a \leq \frac{X_1 + X_2 + \ldots + X_n - n \cdot m}{\sigma\sqrt{n}} < b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx.$$

A határértéket $\Phi(b) - \Phi(a) = \mathbb{P}(a \leq Y \leq b)$ alakban is írhatjuk, ahol $Y \sim N(0,1)$.

Így is átfogalmazható a tétel állítása:

$$\mathbb{P}(nm + a\sigma\sqrt{n} \leq X_1 + X_2 + \ldots + X_n < nm + b\sigma\sqrt{n}) \to \Phi(b) - \Phi(a).$$

Ez azt jelenti, hogy az $\overline{X}_n$ átlag eloszlása közel van egy m várható értékű, $\sigma/\sqrt{n}$ szórású normális eloszláshoz.

Centrális határeloszlástétel

Legyenek $X_1, X_2, \ldots$ **független** azonos eloszlású valószínűségi változók, melyekre $\mathbb{E}(\mathrm{X}_1) = \mathrm{m}$ és $D(X_1) = \sigma < \infty$, azaz **szórásuk véges**. Ekkor

$$\lim_{n\to\infty} \mathbb{P}\left(a \leq \frac{X_1 + X_2 + \ldots + X_n - n \cdot m}{\sigma\sqrt{n}} < b \right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx =$$

$$= \Phi(b) - \Phi(a) = \mathbb{P}(a \leq Z \leq b),$$

ahol $Z \sim N(0,1)$ **standard normális eloszlású**. Tovább alakítva:

Centrális határeloszlástétel

Legyenek $X_1, X_2, \ldots$ **független** azonos eloszlású valószínűségi változók, melyekre $\mathbb{E}(X_1) = \mathrm{m}$ és $D(X_1) = \sigma < \infty$, azaz **szórásuk véges**. Ekkor

$$\lim_{n\to\infty} \mathbb{P}\left(a \leq \frac{X_1 + X_2 + \ldots + X_n - n \cdot m}{\sigma\sqrt{n}} < b \right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx =$$

$$= \Phi(b) - \Phi(a) = \mathbb{P}(a \leq Z \leq b),$$

ahol $Z \sim N(0,1)$ **standard normális eloszlású**. Tovább alakítva:

$$\mathbb{P}(nm + a\sigma\sqrt{n} \leq X_1 + X_2 + \ldots + X_n < nm + b\sigma\sqrt{n}) \to \mathbb{P}(a \leq Z \leq b).$$

Ha n-nel osztunk, hogy az átlag jelenjen meg:

$$\mathbb{P}\left(\mathrm{m} + \mathrm{a}\frac{\sigma}{\sqrt{n}} \leq \frac{X_1 + X_2 + \ldots + X_n}{n} < \mathrm{m} + \mathrm{b}\frac{\sigma}{\sqrt{n}} \right) \to \mathbb{P}(\mathrm{a} \leq Z \leq \mathrm{b}).$$

Vagyis az **átlag eloszlása** „közel van” egy m várható értékű, $\frac{\sigma}{\sqrt{n}}$ szórású **normális eloszláshoz**.

Centrális határeloszlástétel: példa

Legyenek $X_1, X_2, \ldots$ független, 2 várható értékű, exponenciális eloszlású valószínűségi változók. Mi a limesze a $\mathbb{P}(X_1 + \ldots + X_n - 2n < 2\sqrt{n})$ mennyiségnek $n \to \infty$ esetén?

Centrális határeloszlástétel: példa

Legyenek $X_1, X_2, \ldots$ független, 2 várható értékű, exponenciális eloszlású valószínűségi változók. Mi a limesze a $\mathbb{P}(X_1 + \ldots + X_n - 2n < 2\sqrt{n})$ mennyiségnek $n \to \infty$ esetén?

Mivel a valószínűségi változók **függetlenek**, **azonos eloszlásúak és véges szórásúak**, teljesülnek a centrális határeloszlástétel feltételei. Ezért

$$\mathbb{P}(X_1 + \ldots + X_n - 2n < 2\sqrt{n}) = \mathbb{P}\left(\frac{X_1 + \ldots + X_n - 2n}{2\sqrt{n}} < 1\right) \to \Phi(1),$$

ha $n \to \infty$, hiszen $\mathrm{m} = 2$ a várható érték, és mivel az eloszlás exponenciális, a várható érték egyenlő a szórással, így $\sigma = 2$ a szórás.

Házi feladat december 10., kedd, 8:15-ig

Legyenek $X_1, X_2, \ldots, X_n$, $Y_1, Y_2, \ldots, Y_n$ független exponenciális eloszlású valószínűségi változók. Az X-ek paramétere legyen $1/10$, az Y-oké pedig 1. Mennyi az alábbi határérték?

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{X_1 + \ldots + X_n + Y_1 + \ldots + Y_n - 11n}{\sqrt{n}} < 10\right)?$$

Konvergenciafajták

Definíció

A $Z_1, Z_2, \ldots,$ *valószínűségi változókból álló sorozat* **eloszlásban konvergál** *az* Z *valószínűségi változóhoz, ha minden olyan* t *számra, melyre* Z *eloszlásfüggvénye folytonos* t*-ben, teljesül, hogy*

$$\mathbb{P}(Z_n \leq t) \to \mathbb{P}(Z \leq t) \qquad (n \to \infty).$$

Konvergenciafajták

Definíció

A $Z_1, Z_2, \ldots,$ *valószínűségi változókból álló sorozat* **eloszlásban konvergál** *az* Z *valószínűségi változóhoz, ha minden olyan* t *számra, melyre* Z *eloszlásfüggvénye folytonos* t*-ben, teljesül, hogy*

$$\mathbb{P}(Z_n \leq t) \to \mathbb{P}(Z \leq t) \qquad (n \to \infty).$$

Ha $Z_n \to Z$ teljesül 1 valószínűséggel, akkor $Z_n \to Z$ sztochasztikusan és eloszlásban is.

Lehetséges, hogy $Z_n \to Z$ eloszlásban, de Z_n nem tart Z-hez sztochasztikusan (és ezért 1 valószínűséggel sem).

A centrális határeloszlástétel alkalmazása

Tegyük fel, hogy egy véletlenszerűen választott ember p valószínűséggel szavaz egy adott pártra. Legalább hány embert kell megkérdeznünk (feltételezve, hogy mindenki a többiektől függetlenül válaszol és igazat mond), hogy annak valószínűsége, hogy a pártra szavazók aránya legfeljebb $0,01$-gyel tér el p-től, tetszőleges p esetén legalább 95% legyen?

A centrális határeloszlástétel alkalmazása

Tegyük fel, hogy egy véletlenszerűen választott ember p valószínűséggel szavaz egy adott pártra. Legalább hány embert kell megkérdeznünk (feltételezve, hogy mindenki a többiektől függetlenül válaszol és igazat mond), hogy annak valószínűsége, hogy a pártra szavazók aránya legfeljebb $0,01$-gyel tér el p-től, tetszőleges p esetén legalább 95% legyen?

n megkérdezett, mindenki p valószínűséggel támogatja a pártot

X: a pártot támogatók száma a megkérdezettek között

Kell:

$$\mathbb{P}\left(\left|\frac{X}{n} - p\right| \leq 0,01\right) \geq 0,95$$

teljesüljön minden $0 \leq p \leq 1$-re.

A centrális határeloszlástétel alkalmazása

n megkérdezett, mindenki p valószínűséggel támogatja a pártot

X: a pártot támogatók száma a megkérdezettek között,

Kell:

$$\mathbb{P}\left(\left|\frac{X}{n} - p\right| \leq 0,01\right) \geq 0,95$$

teljesüljön minden $0 \leq p \leq 1$-re, ahol $X = \sum_{j=1}^{n} X_j$, az X_j-k függetlenek, és

$$\mathbb{P}(X_j = 1) = 1 - \mathbb{P}(X_j = 0) = p; \qquad \mathbb{E}(X_j) = p; \qquad D(X_j) = \sqrt{p(1-p)}.$$

$$\mathbb{P}\left(\left|\frac{X}{n} - p\right| \leq 0,01\right) = \mathbb{P}\left(\left|\frac{\sum_{j=1}^{n} X_j - np}{\sqrt{np(1-p)}}\right| \leq \frac{0,01\sqrt{n}}{\sqrt{p(1-p)}}\right)$$
$$\approx 2\Phi\left(\frac{0,01\sqrt{n}}{\sqrt{p(1-p)}}\right) - 1.$$

A centrális határeloszlástétel alkalmazása

n megkérdezett, mindenki p valószínűséggel támogatja a pártot

Kell:

$$\mathbb{P}\left(\left|\frac{X}{n} - p\right| \leq 0,01\right) \approx 2\Phi\left(\frac{0,01\sqrt{n}}{\sqrt{p(1-p)}}\right) - 1 \geq 0,95$$

teljesüljön minden $0 \leq p \leq 1$-re. Vagyis mivel $p(1-p) \leq 1/4$:

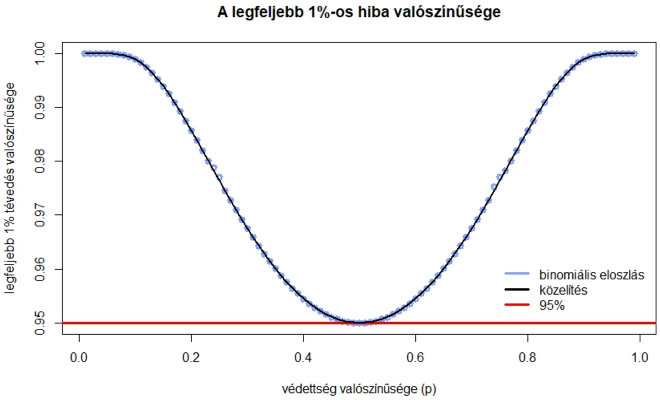
$$\Phi\left(\frac{0,01\sqrt{n}}{\sqrt{p(1-p)}}\right) \geq 0,975;$$

$$\frac{0,01\sqrt{n}}{\sqrt{p(1-p)}} \geq \Phi^{-1}(0,975) = \mathtt{qnorm}(0,975) = 1,96;$$

$$n \geq p(1-p) \cdot 1,96^2 \cdot \frac{1}{0,01^2};$$

$$n \geq \frac{1}{4} \cdot 1,96^2 \cdot \frac{1}{0,01^2} = 9607.$$

A hiba valószínűsége



A hiba valószínűsége a p függvényében

A centrális határeloszlástétel alkalmazása

Tegyük fel, hogy egy véletlenszerűen választott ember p valószínűséggel szavaz egy adott pártra. Legalább hány embert kell megkérdeznünk (feltételezve, hogy mindenki a többiektől függetlenül válaszol és igazat mond), hogy annak valószínűsége, hogy a pártra szavazók aránya legfeljebb $0,01$-gyel tér el p-től, tetszőleges p esetén legalább 95% legyen?

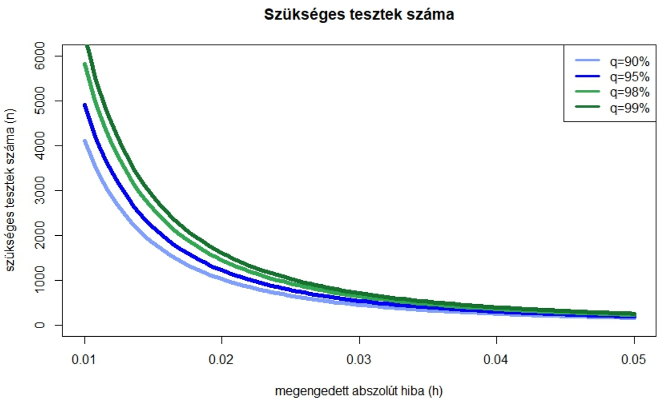
- Csebisev-egyenlőtlenséggel: $n \geq 50000$ biztosan elég
- centrális határeloszlástétellel közelítve: $n \geq 9607$ elég

A centrális határeloszlástétel alkalmazása

Tegyük fel, hogy egy véletlenszerűen választott ember p valószínűséggel szavaz egy adott pártra. Legalább hány embert kell megkérdeznünk (feltételezve, hogy mindenki a többiektől függetlenül válaszol és igazat mond), hogy annak valószínűsége, hogy a pártra szavazók aránya legfeljebb $0,01$-gyel tér el p-től, tetszőleges p esetén legalább 95% legyen?

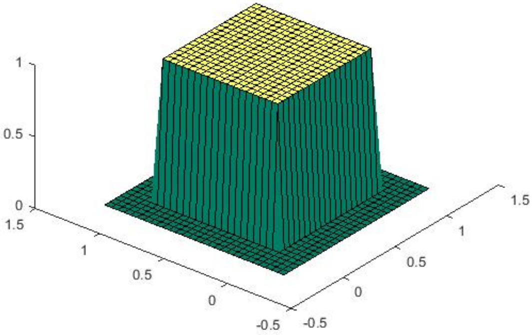
- Csebisev-egyenlőtlenséggel: $n \geq 50000$ biztosan elég
- centrális határeloszlástétellel közelítve: $n \geq 9607$ elég
- valójában: $n = 9607, p = 1/2$ esetén $0,94987$ adódik a $0,95$ helyett
- valójában $n \geq 9650$ kell (pontos számolással)

Szükséges mintaelemszám



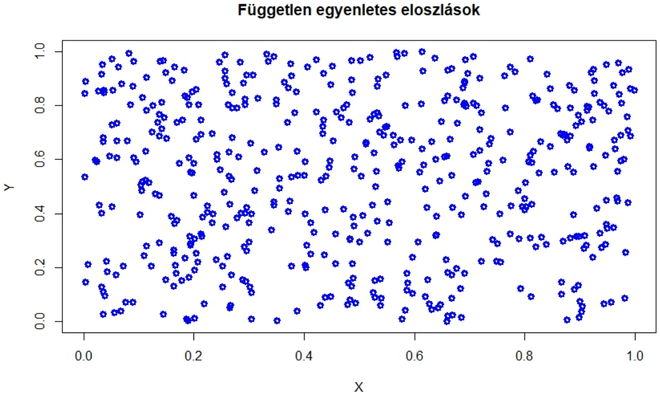
A szükséges mintaelemszám a megengedett hibák függvényében

Kétdimenziós egyenletes eloszlás



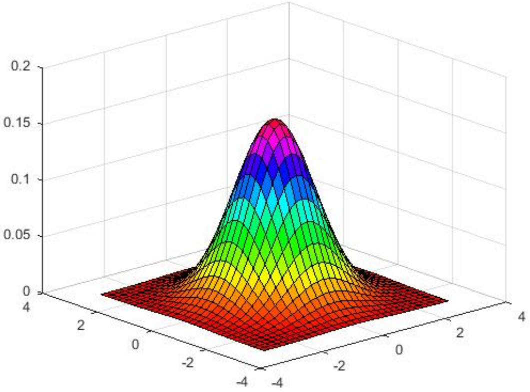
(X, Y) együttes sűrűségfüggvénye, ahol X és Y függetlenek és a $[0, 1]$ intervallumon egyenletes eloszlásúak

Kétdimenziós egyenletes eloszlás



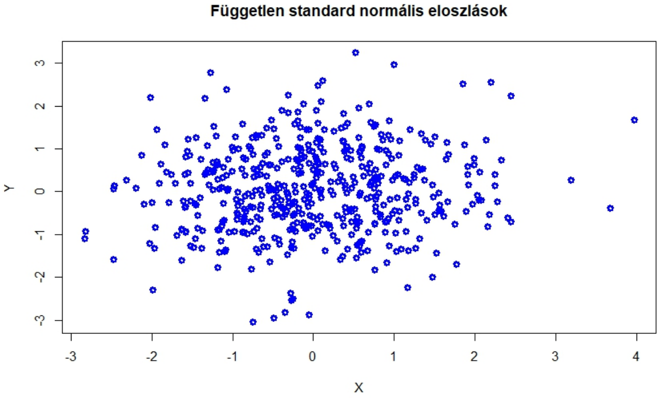
500 darab véletlen pont a síkon, melyek koordinátái függetlenek és a $[0, 1]$ intervallumon egyenletes eloszlásúak (az együttes sűrűségfüggvény az előző ábrán látható).

Kétdimenziós normális eloszlás



Két független standard normális eloszlás együttes sűrűségfüggvénye Azaz: (X, Z) együttes sűrűségfüggvénye, ahol X, Z függetlenek, $N(0, 1)$ eloszlásúak

Kétdimenziós normális eloszlás



500 darab véletlen pont a síkon, melyek koordinátái független standard nomális eloszlásúak. Ahol nagyobb az együttes sűrűségfüggvény (előző ábra), oda több pont esik.

Együttes sűrűségfüggvény

Definíció

Az $\underline{X} = (X_1, \ldots, X_n)$ *valószínűségi vektorváltozó abszolút folytonos, ha van olyan* $f : \mathbb{R}^n \to \mathbb{R}$ *függvény, melyre*

$$F(t_1, \ldots, t_n) = \int_{-\infty}^{t_1} \ldots \int_{-\infty}^{t_n} f(s_1, \ldots, s_n) ds_1 \ldots ds_n.$$

teljesül minden $(t_1, \ldots, t_n) \in \mathbb{R}^n$ *esetén. Ilyenkor az* f *függvényt az* $(X_1, X_2, \ldots, X_n)$ **együttes sűrűségfüggvényének** *nevezzük.*

Együttes sűrűségfüggvény

Definíció

Az $\underline{X} = (X_1, \ldots, X_n)$ *valószínűségi vektorváltozó abszolút folytonos, ha van olyan* $f : \mathbb{R}^n \to \mathbb{R}$ *függvény, melyre*

$$F(t_1, \ldots, t_n) = \int_{-\infty}^{t_1} \ldots \int_{-\infty}^{t_n} f(s_1, \ldots, s_n) ds_1 \ldots ds_n.$$

teljesül minden $(t_1, \ldots, t_n) \in \mathbb{R}^n$ *esetén. Ilyenkor az* f *függvényt az* $(X_1, X_2, \ldots, X_n)$ **együttes sűrűségfüggvényének** *nevezzük.*

Tegyük fel, hogy az $(X_1, X_2, \ldots, X_n)$ valószínűségi vektorváltozó együttes sűrűségfüggvénye f. Ekkor egy $A \subseteq \mathbb{R}^n$ halmazra

$$\mathbb{P}((X_1, X_2, \ldots, X_n) \in A) = \int_A f(t_1, t_2, \ldots, t_n)\, dt_1\, dt_2 \ldots dt_n.$$

Következmény:

$$\int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} f(t_1, t_2, \ldots, t_n)\, dt_1\, dt_2 \ldots dt_n = 1.$$

Peremeloszlások sűrűségfüggvénye

Tegyük fel, hogy az $(X_1, X_2, \ldots, X_n)$ valószínűségi vektorváltozó **együttes sűrűségfüggvénye** f. Hogyan kapható meg például az **első peremeloszlás, azaz X_1 sűrűségfüggvénye**?

Peremeloszlások sűrűségfüggvénye

Tegyük fel, hogy az $(X_1, X_2, \ldots, X_n)$ valószínűségi vektorváltozó **együttes sűrűségfüggvénye** f. Hogyan kapható meg például az **első peremeloszlás, azaz X_1 sűrűségfüggvénye**?

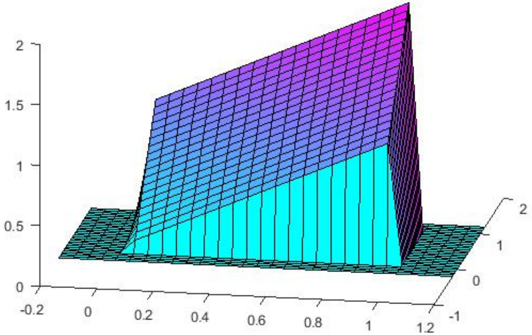
Az X_j valószínűségi változó sűrűségfüggvénye (melyet f_j-vel jelölünk), azaz a j. peremsűrűségfüggvény így kapható meg f-ből:

$$f_j(t) = \int_{\mathbb{R}^{n-1}} f(s_1, \ldots, s_{j-1}, t, s_{j+1}, \ldots, s_n) ds_1 \ldots ds_{j-1} ds_{j+1} \ldots ds_n.$$

Speciálisan $n = 2$-re:

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y) dy; \qquad f_2(y) = \int_{-\infty}^{\infty} f(x, y) dx.$$

Együttes sűrűségfüggvény: példa



A $[0,1] \times [0,1]$ négyzeten $x + y$ alakú együttes sűrűségfüggvény

Együttes sűrűségfüggvény: példa

Tegyük fel, hogy az (X, Y) valószínűségi vektorváltozó együttes sűrűségfüggvénye

$$f(x,y) = \begin{cases} x + y, & \text{ha } 0 \leq x \leq 1 \text{ és } 0 \leq y \leq 1; \\ 0 & \text{különben.} \end{cases}$$

Számítsuk ki X és Y korrelációs együtthatóját:

$$R(X, Y) = \frac{\operatorname{cov}(X, Y)}{D(X)D(Y)} = \frac{\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)}{D(X)D(Y)}.$$

Együttes sűrűségfüggvény: példa

$$f(x,y) = \begin{cases} x+y, & \text{ha } 0 \le x \le 1 \text{ és } 0 \le y \le 1; \\ 0 & \text{különben.} \end{cases}$$

Állítás

Legyen az (X, Y) valószínűségi vektorváltozó együttes sűrűségfüggvénye f. Ekkor

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

Együttes sűrűségfüggvény: példa

$$f(x,y) = \begin{cases} x+y, & \text{ha } 0 \le x \le 1 \text{ és } 0 \le y \le 1; \\ 0 & \text{különben.} \end{cases}$$

Állítás

Legyen az (X,Y) valószínűségi vektorváltozó együttes sűrűségfüggvénye f. Ekkor

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

$$\mathbb{E}(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy \cdot f(x,y)\,dy\,dx = \int_0^1\int_0^1 xy(x+y)dydx =$$

$$= \int_0^1\int_0^1 x^2ydydx + \int_0^1\int_0^1 xy^2dydx = \int_0^1 \frac{x^2}{2}dx + \int_0^1 \frac{x}{3}dx = \frac{1}{6} +$$

felhasználva, hogy $\int_0^1 x^k dx = [x^{k+1}/(k+1)]_{x=0}^1 = 1/(k+1)$.

Együttes sűrűségfüggvény: példa

Az (X, Y) együttes sűrűségfüggvénye:

$$f(x,y) = \begin{cases} x+y, & \text{ha } 0 \le x \le 1 \text{ és } 0 \le y \le 1; \\ 0 & \text{különben.} \end{cases}$$

Az X valószínűségi változó sűrűségfüggvénye:

$$f_1(x) = \int_{-\infty}^{\infty} f(x,y)dy = \int_0^1 x + y\, dy = x + \frac{1}{2},$$

ha $0 \le x \le 1$, és 0 különben. Ezért

Együttes sűrűségfüggvény: példa

Az (X, Y) együttes sűrűségfüggvénye:

$$f(x,y) = \begin{cases} x+y, & \text{ha } 0 \leq x \leq 1 \text{ és } 0 \leq y \leq 1; \\ 0 & \text{különben.} \end{cases}$$

Az X valószínűségi változó sűrűségfüggvénye:

$$f_1(x) = \int_{-\infty}^{\infty} f(x,y)dy = \int_0^1 x + y\, dy = x + \frac{1}{2},$$

ha $0 \leq x \leq 1$, és 0 különben. Ezért

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f_1(x) dx = \int_0^1 x \cdot \left(x + \frac{1}{2}\right) dx = \int_0^1 x^2 + \frac{x}{2} dx = \frac{1}{3} + \frac{1}{4} = \frac{7}{12}.$$

Együttes sűrűségfüggvény: példa

Az X valószínűségi változó sűrűségfüggvénye:

$$f_1(x) = \int_{-\infty}^{\infty} f(x,y)dy = \int_0^1 x + ydy = x + \frac{1}{2},$$

ha $0 \leq x \leq 1$, és 0 különben. Ezért

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} xf_1(x)dx = \int_0^1 x \cdot \left(x + \frac{1}{2}\right) dx = \int_0^1 x^2 + \frac{x}{2} dx = \frac{1}{3} + \frac{1}{4} = \frac{7}{12}.$$

$$\mathbb{E}(X^2) = \int_{-\infty}^{\infty} x^2 f_1(x)dx = \int_0^1 x^2 \cdot \left(x + \frac{1}{2}\right) dx = \int_0^1 x^3 + \frac{x^2}{2} dx = \frac{1}{4} + \frac{1}{6} = \frac{5}{12}.$$

Együttes sűrűségfüggvény: példa

Az X valószínűségi változó sűrűségfüggvénye:

$$f_1(x) = \int_{-\infty}^{\infty} f(x,y)dy = \int_0^1 x + ydy = x + \frac{1}{2},$$

ha $0 \leq x \leq 1$, és 0 különben. Ezért

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} xf_1(x)dx = \int_0^1 x \cdot \left(x + \frac{1}{2}\right) dx = \int_0^1 x^2 + \frac{x}{2} dx = \frac{1}{3} + \frac{1}{4} = \frac{7}{12}.$$

$$\mathbb{E}(X^2) = \int_{-\infty}^{\infty} x^2 f_1(x)dx = \int_0^1 x^2 \cdot \left(x + \frac{1}{2}\right) dx = \int_0^1 x^3 + \frac{x^2}{2} dx = \frac{1}{4} + \frac{1}{6} = \frac{5}{12}.$$

A szimmetria miatt hasonlóképpen:

$$\mathbb{E}(Y) = \frac{7}{12}; \qquad \mathbb{E}(Y^2) = \frac{5}{12}.$$

Együttes sűrűségfüggvény: példa

$$f(x,y) = \begin{cases} x+y, & \text{ha } 0 \le x \le 1 \text{ és } 0 \le y \le 1; \\ 0 & \text{különben.} \end{cases}$$

$$\mathbb{E}(XY) = \frac{1}{3}; \qquad \mathbb{E}(X) = \mathbb{E}(Y) = \frac{7}{12}; \qquad \mathbb{E}(X^2) = \mathbb{E}(Y^2) = \frac{5}{12}.$$

Az X és Y korrelációs együtthatója:

$$\begin{aligned} R(X,Y) &= \frac{\operatorname{cov}(X,Y)}{D(X)D(Y)} = \frac{\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)}{D(X)D(Y)} = \\ &= \frac{1/3 - (7/12)^2}{\sqrt{5/12 - (7/12)^2} \cdot \sqrt{5/12 - (7/12)^2}} = \frac{1/3 - (7/12)^2}{5/12 - (7/12)^2} = \\ &= -0,091. \end{aligned}$$

Nagyon gyenge negatív korreláció van a két valószínűségi változó között.

Feltételes eloszlás

Legyen X diszkrét valószínűségi változó, lehetséges értékei $x_1, x_2, \ldots$, a hozzájuk tartozó valószínűségek: $\mathbb{P}(X = x_k) = p_k$. Legyen A pozitív valószínűségű esemény. Ekkor az X-nek az A eseményre vonatkozó feltételes eloszlása:

$$q_k = \mathbb{P}(X = x_k | A) = \frac{\mathbb{P}(\{X = x_k\} \cap A)}{\mathbb{P}(A)}.$$

Ekkor

$$\sum_{k=1}^{\infty} q_k = \sum_{k=1}^{\infty} \frac{\mathbb{P}(\{X = x_k\} \cap A)}{\mathbb{P}(A)} = 1,$$

azaz a (q_k) sorozat is valószínűségeloszlás.

Másrészt az $\{X = x_k\}$, $k = 1, 2, \ldots$ események teljes eseményrendszert alkotnak.

Feltételes eloszlás

Legyen X diszkrét valószínűségi változó, lehetséges értékei $x_1, x_2, \ldots$, a hozzájuk tartozó valószínűségek: $\mathbb{P}(X = x_k) = p_k$. Legyen A pozitív valószínűségű esemény. Ekkor az X-nek az A eseményre vonatkozó feltételes eloszlása:

$$q_k = \mathbb{P}(X = x_k | A) = \frac{\mathbb{P}(\{X = x_k\} \cap A)}{\mathbb{P}(A)}.$$

Ekkor

$$\sum_{k=1}^{\infty} q_k = \sum_{k=1}^{\infty} \frac{\mathbb{P}(\{X = x_k\} \cap A)}{\mathbb{P}(A)} = 1,$$

azaz a (q_k) sorozat is valószínűségeloszlás.

Másrészt az $\{X = x_k\}$, $k = 1, 2, \ldots$ események teljes eseményrendszert alkotnak.

Feltételes várható érték:

$$\mathbb{E}(X|A) = \sum_{k=1}^{\infty} k \cdot q_k = \sum_{k=1}^{\infty} k \cdot \mathbb{P}(X = x_k | A).$$

Együttes eloszlás: példa

Kétszer dobunk szabályos kockával. Legyen X az első dobás, Y pedig a dobott számok közül a nagyobb. Ekkor az (X, Y) valószínűségi vektorváltozó együttes eloszlása:

X/Y	1	2	3	4	5	6	összesen
1	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/6$
2	0	$1/18$	$1/36$	$1/36$	$1/36$	$1/36$	$1/6$
3	0	0	$1/12$	$1/36$	$1/36$	$1/36$	$1/6$
4	0	0	0	$1/9$	$1/36$	$1/36$	$1/6$
5	0	0	0	0	$5/36$	$1/36$	$1/6$
6	0	0	0	0	0	$1/6$	$1/6$
összesen	$1/36$	$1/12$	$5/36$	$7/36$	$1/4$	$11/36$	1

Legyen $A = \{Y = 3\}$ a feltétel.

Ekkor X feltételes eloszlása:

$$q_1 = \mathbb{P}(X = 1 | Y = 3) = \frac{1/36}{5/36} = \frac{1}{5} = q_2; \qquad q_3 = \frac{1/12}{5/36} = \frac{3}{5}.$$

Együttes eloszlás: példa

Kétszer dobunk szabályos kockával. Legyen X az első dobás, Y pedig a dobott számok közül a nagyobb. Ekkor az (X, Y) valószínűségi vektorváltozó együttes eloszlása:

X/Y	1	2	3	4	5	6	összesen
1	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/6$
2	0	$1/18$	$1/36$	$1/36$	$1/36$	$1/36$	$1/6$
3	0	0	$1/12$	$1/36$	$1/36$	$1/36$	$1/6$
4	0	0	0	$1/9$	$1/36$	$1/36$	$1/6$
5	0	0	0	0	$5/36$	$1/36$	$1/6$
6	0	0	0	0	0	$1/6$	$1/6$
összesen	$1/36$	$1/12$	$5/36$	$7/36$	$1/4$	$11/36$	1

Legyen $A = \{Y = 3\}$ a feltétel. Ekkor X feltételes várható értéke:

$$\mathbb{E}(X|Y = 3) = \sum_{k=1}^{3} \mathbb{P}(X = k|Y = 3) \cdot k = \frac{1}{5} \cdot 1 + \frac{1}{5} \cdot 2 + \frac{3}{5} \cdot 3 = \frac{12}{5} = 2,4.$$

Feltételes sűrűségfüggvény és várható érték

Legyen az (X, Y) valószínűségi vektorváltozó együttes sűrűségfüggvénye a f. Az X, illetve Y sűrűségfüggvénye:

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)dy; \qquad f_Y(y) = \int_{-\infty}^{\infty} f(x, y)dx.$$

Definíció

Az X valószínűségi változónak az $Y = y$ feltételre vonatkozó **feltételes sűrűségfüggvénye** *adott y valós számra:*

$$f_{X|Y=y}(x) = \frac{f(x, y)}{f_Y(y).}$$

A $g(X)$ mennyiség **feltételes várható értéke** *az $Y = y$ feltétel mellett:*

$$\mathbb{E}(g(X)|Y = y) = \int_{-\infty}^{\infty} g(x)f_{X|Y=y}(x)dx.$$

Feltételes várható érték: példa

Legyenek X és Z független standard normális eloszlású valószínűségi változók. Ekkor az $(X, X+Z)$ együttes sűrűségfüggvénye:

$$f(x,y) = \frac{1}{2\pi}\exp\left(-\frac{2x^2 - 2xy + y^2}{2}\right).$$

Az $X+Z$ valószínűségi változó normális eloszlású $m = 0$ várható értékkel és $\sigma = 2$ szórásnégyzettel, így

$$f_{X+Z}(y) = \frac{1}{2\sqrt{\pi}}\exp\left(-\frac{y^2}{4}\right).$$

Mennyi lehet az $\mathbb{E}(X|X+Z=y)$ feltételes várható érték?

Feltételes várható érték: példa

Legyenek X és Z független standard normális eloszlású valószínűségi változók. Ekkor az $(X, X+Z)$ együttes sűrűségfüggvénye:

$$f(x,y) = \frac{1}{2\pi} \exp\left(-\frac{2x^2 - 2xy + y^2}{2} \right).$$

Mennyi lehet az $\mathbb{E}(X|X+Z=y)$ feltételes várható érték?

Feltételes várható érték: példa

Legyenek X és Z független standard normális eloszlású valószínűségi változók. Ekkor az $(X, X+Z)$ együttes sűrűségfüggvénye:

$$f(x,y) = \frac{1}{2\pi} \exp\left(-\frac{2x^2 - 2xy + y^2}{2} \right).$$

Mennyi lehet az $\mathbb{E}(X|X+Z=y)$ feltételes várható érték?

$$\begin{aligned}
\mathbb{E}(X|X+Z=y) &= \int_{-\infty}^{\infty} x \frac{f(x,y)}{f_{X+Z}(y)} dx = \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi}} x \exp\left(-\frac{2x^2 - 2xy + y^2/2}{2} \right) dx = \\
&= \int_{-\infty}^{\infty} x \cdot \frac{1}{\sqrt{\pi}} \exp\left(-\frac{(x-y/2)^2}{2 \cdot 1/2} \right) dx = \frac{y}{2}.
\end{aligned}$$