# A sztochasztika alapjai
# MBNXK262

## 6. előadás: Nevezetes folytonos eloszlások; várható érték

Kevei Péter

2023/24 tavasz

# Ismétlés

$\xi$ *folytonos eloszlású*, ha

$$F(x) = \mathbf{P}(\xi < x) = \int_{-\infty}^{x} f(y)\mathrm{d}y, \quad x \in \mathbb{R}.$$

$f(x)$ a $\xi$ sűrűségfüggvénye.
Várható érték

$$\mathbf{E}(\xi) = \int_{-\infty}^{\infty} yf(y)\mathrm{d}y,$$

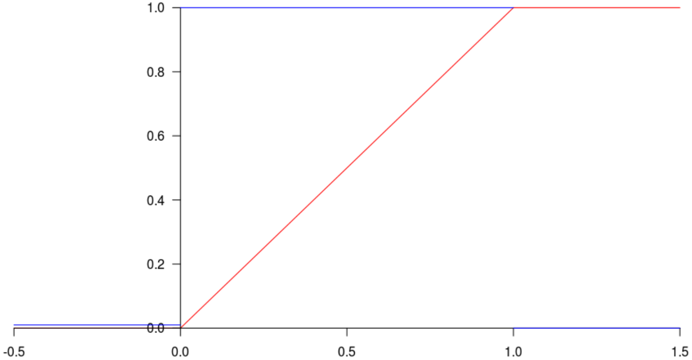
szórásnégyzet

$$\mathbf{D}^2(\xi) = \mathbf{E}[(\xi - \mathbf{E}(\xi))^2] = \mathbf{E}(\xi^2) - (\mathbf{E}(\xi))^2$$

# Egyenletes eloszlás

$\xi$ *egyenletes eloszlású az* $(a, b)$ *intervallumon*, $\xi \sim \text{Egy}(a, b)$, $-\infty < a < b < \infty$, ha sűrűségfüggvénye

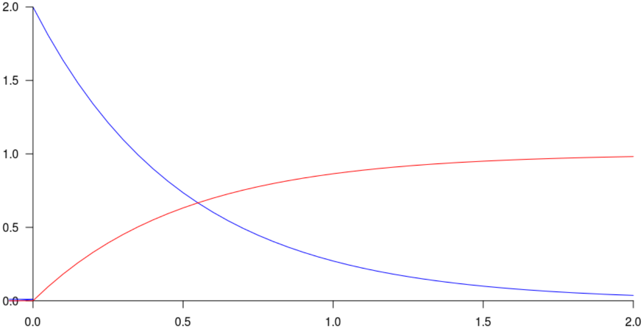
$$f(y) = \begin{cases} \frac{1}{b-a}, & \text{ha } y \in (a, b), \\ 0, & \text{különben.} \end{cases}$$


1.0
0.8
0.6
0.4
0.2
0.0
-0.5
0.0
0.5
1.0
1.5

# Exponenciális eloszlás

$\xi$ $\lambda$*-paraméterű exponenciális eloszlású*, $\xi \sim \text{Exp}(\lambda)$, $\lambda > 0$, ha sűrűségfüggvénye

$$f(y) = \begin{cases} \lambda e^{-\lambda y}, & y \geq 0, \\ 0, & y < 0. \end{cases}$$

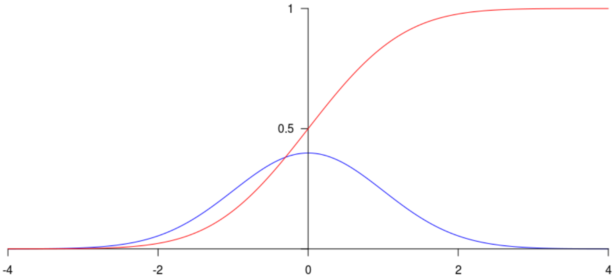

2.0
1.5
1.0
0.5
0.0
0.0
0.5
1.0
1.5
2.0

# Momentumok

# Normális eloszlás

$\xi$ *normális eloszlású* $\mu$ *és* $\sigma^2$ *paraméterekkel*, $X \sim \mathrm{N}(\mu, \sigma^2)$, $\mu \in \mathbb{R}, \sigma^2 > 0$, ha sűrűségfüggvénye

$$f_{\mu,\sigma}(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}}.$$


1
0.5
-4
-2
0
2
4

# Standardizálás

## Állítás

*Ha* $\xi \sim N(\mu, \sigma^2)$*, akkor* $(\xi - \mu)/\sigma \sim N(0, 1)$*.*

| $x$ | 0,00 | 0,01 | 0,02 | 0,03 | 0,04 | 0,05 | 0,06 | 0,07 | 0,08 | 0,09 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0,0 | 0,5000 | 0,5040 | 0,5080 | 0,5120 | 0,5160 | 0,5199 | 0,5239 | 0,5279 | 0,5319 | 0,5359 |
| 0,1 | 0,5398 | 0,5438 | 0,5478 | 0,5517 | 0,5557 | 0,5596 | 0,5636 | 0,5675 | 0,5714 | 0,5753 |
| 0,2 | 0,5793 | 0,5832 | 0,5871 | 0,5910 | 0,5948 | 0,5987 | 0,6026 | 0,6064 | 0,6103 | 0,6141 |
| 0,3 | 0,6179 | 0,6217 | 0,6255 | 0,6293 | 0,6331 | 0,6368 | 0,6406 | 0,6443 | 0,6480 | 0,6517 |
| 0,4 | 0,6554 | 0,6591 | 0,6628 | 0,6664 | 0,6700 | 0,6736 | 0,6772 | 0,6808 | 0,6844 | 0,6879 |
| 0,5 | 0,6915 | 0,6950 | 0,6985 | 0,7019 | 0,7054 | 0,7088 | 0,7123 | 0,7157 | 0,7190 | 0,7224 |
| 0,6 | 0,7257 | 0,7291 | 0,7324 | 0,7357 | 0,7389 | 0,7422 | 0,7454 | 0,7486 | 0,7517 | 0,7549 |
| 0,7 | 0,7580 | 0,7611 | 0,7642 | 0,7673 | 0,7704 | 0,7734 | 0,7764 | 0,7794 | 0,7823 | 0,7852 |
| 0,8 | 0,7881 | 0,7910 | 0,7939 | 0,7967 | 0,7995 | 0,8023 | 0,8051 | 0,8078 | 0,8106 | 0,8133 |
| 0,9 | 0,8159 | 0,8186 | 0,8212 | 0,8238 | 0,8264 | 0,8289 | 0,8315 | 0,8340 | 0,8365 | 0,8389 |
| 1,0 | 0,8413 | 0,8438 | 0,8461 | 0,8485 | 0,8508 | 0,8531 | 0,8554 | 0,8577 | 0,8599 | 0,8621 |
| 1,1 | 0,8643 | 0,8665 | 0,8686 | 0,8708 | 0,8729 | 0,8749 | 0,8770 | 0,8790 | 0,8810 | 0,8830 |
| 1,2 | 0,8849 | 0,8869 | 0,8888 | 0,8907 | 0,8925 | 0,8944 | 0,8962 | 0,8980 | 0,8997 | 0,9015 |
| 1,3 | 0,9032 | 0,9049 | 0,9066 | 0,9082 | 0,9099 | 0,9115 | 0,9131 | 0,9147 | 0,9162 | 0,9177 |
| 1,4 | 0,9192 | 0,9207 | 0,9222 | 0,9236 | 0,9251 | 0,9265 | 0,9279 | 0,9292 | 0,9306 | 0,9319 |
| 1,5 | 0,9332 | 0,9345 | 0,9357 | 0,9370 | 0,9382 | 0,9394 | 0,9406 | 0,9418 | 0,9429 | 0,9441 |
| 1,6 | 0,9452 | 0,9463 | 0,9474 | 0,9484 | 0,9495 | 0,9505 | 0,9515 | 0,9525 | 0,9535 | 0,9545 |
| 1,7 | 0,9554 | 0,9564 | 0,9573 | 0,9582 | 0,9591 | 0,9599 | 0,9608 | 0,9616 | 0,9625 | 0,9633 |
| 1,8 | 0,9641 | 0,9649 | 0,9656 | 0,9664 | 0,9671 | 0,9678 | 0,9686 | 0,9693 | 0,9699 | 0,9706 |
| 1,9 | 0,9713 | 0,9719 | 0,9726 | 0,9732 | 0,9738 | 0,9744 | 0,9750 | 0,9756 | 0,9761 | 0,9767 |
| 2,0 | 0,9772 | 0,9778 | 0,9783 | 0,9788 | 0,9793 | 0,9798 | 0,9803 | 0,9808 | 0,9812 | 0,9817 |
| 2,1 | 0,9821 | 0,9826 | 0,9830 | 0,9834 | 0,9838 | 0,9842 | 0,9846 | 0,9850 | 0,9854 | 0,9857 |
| 2,2 | 0,9861 | 0,9864 | 0,9868 | 0,9871 | 0,9875 | 0,9878 | 0,9881 | 0,9884 | 0,9887 | 0,9890 |
| 2,3 | 0,9893 | 0,9896 | 0,9898 | 0,9901 | 0,9904 | 0,9906 | 0,9909 | 0,9911 | 0,9913 | 0,9916 |
| 2,4 | 0,9918 | 0,9920 | 0,9922 | 0,9925 | 0,9927 | 0,9929 | 0,9931 | 0,9932 | 0,9934 | 0,9936 |
| 2,5 | 0,9938 | 0,9940 | 0,9941 | 0,9943 | 0,9945 | 0,9946 | 0,9948 | 0,9949 | 0,9951 | 0,9952 |
| 2,6 | 0,9953 | 0,9955 | 0,9956 | 0,9957 | 0,9959 | 0,9960 | 0,9961 | 0,9962 | 0,9963 | 0,9964 |
| 2,7 | 0,9965 | 0,9966 | 0,9967 | 0,9968 | 0,9969 | 0,9970 | 0,9971 | 0,9972 | 0,9973 | 0,9974 |
| 2,8 | 0,9974 | 0,9975 | 0,9976 | 0,9977 | 0,9977 | 0,9978 | 0,9979 | 0,9979 | 0,9980 | 0,9981 |
| 2,9 | 0,9981 | 0,9982 | 0,9982 | 0,9983 | 0,9984 | 0,9984 | 0,9985 | 0,9985 | 0,9986 | 0,9986 |
| 3,0 | 0,9987 | 0,9987 | 0,9987 | 0,9988 | 0,9988 | 0,9989 | 0,9989 | 0,9989 | 0,9990 | 0,9990 |
| 3,1 | 0,9990 | 0,9991 | 0,9991 | 0,9991 | 0,9992 | 0,9992 | 0,9992 | 0,9992 | 0,9993 | 0,9993 |
| 3,2 | 0,9993 | 0,9993 | 0,9994 | 0,9994 | 0,9994 | 0,9994 | 0,9994 | 0,9995 | 0,9995 | 0,9995 |
| 3,3 | 0,9995 | 0,9995 | 0,9995 | 0,9996 | 0,9996 | 0,9996 | 0,9996 | 0,9996 | 0,9996 | 0,9997 |
| 3,4 | 0,9997 | 0,9997 | 0,9997 | 0,9997 | 0,9997 | 0,9997 | 0,9997 | 0,9997 | 0,9997 | 0,9998 |

$\xi \sim$ N$(\mu, \sigma^2)$ és $Z \sim$ N$(0,1)$, akkor

$$\begin{aligned}
\mathbf{P}(|\xi - \mu| \leq \lambda\sigma) &= \mathbf{P}(|Z| \leq \lambda) \\
&= \Phi(\lambda) - \Phi(-\lambda) \\
&= 2\Phi(\lambda) - 1 \\
&= \begin{cases} 0,6827, & \lambda = 1, \\ 0,9545, & \lambda = 2, \\ 0,9973, & \lambda = 3, \\ 0,9999, & \lambda = 4. \end{cases}
\end{aligned}$$

# Várható érték

## Definíció

$\xi$ diszkrét $x_1, x_2, \ldots$ lehetséges értékekkel. $\xi$ *várható értéke*

$$\mathbf{E}(\xi) = \sum_i x_i \mathbf{P}(\xi = x_i),$$

ha $\sum_i |x_i| \mathbf{P}(\xi = x_i) < \infty$.
Ha $\xi$ folytonos $f(x)$ sűrűségfüggvénnyel, akkor $\xi$ várható értéke

$$\mathbf{E}(\xi) = \int_{-\infty}^{\infty} y f(y) \mathrm{d}y,$$

ha $\int_{-\infty}^{\infty} |y| f(y) \mathrm{d}y < \infty$.

# Várható érték tulajdonságai

## Állítás

*$\xi, \eta$ véletlen változók, $g : \mathbb{R} \to \mathbb{R}$, $h : \mathbb{R}^2 \to \mathbb{R}$. Ekkor*

$$\mathbf{E}(g(\xi)) = \sum_{i=1}^{\infty} g(x_i)\mathbf{P}(\xi = x_i), \text{ ill. } \mathbf{E}(g(\xi)) = \int_{-\infty}^{\infty} g(y)f(y)\mathrm{d}y,$$

*ahol $f(x)$ az $\xi$ sűrűségfüggvénye, és*

$$\mathbf{E}(h(\xi, \eta)) = \sum_i \sum_j h(x_i, y_j)\mathbf{P}(\xi = x_i, \eta = y_j).$$

# Várható érték tulajdonságai

## Állítás

*A várható érték lineáris, azaz tetszőleges $a, b \in \mathbb{R}$ állandókra*

$$\mathbf{E}(a\xi + b) = a\mathbf{E}(\xi) + b.$$

# Várható érték tulajdonságai

### Állítás

*Ha* $a \leq \xi \leq b$*, akkor* $a \leq \mathbf{E}(\xi) \leq b$ *tetszőleges* $a, b \in \mathbb{R}$ *számok esetén.*

# Várható érték tulajdonságai

### Állítás

$\mathbf{E}(\xi + \eta) = \mathbf{E}(\xi) + \mathbf{E}(\eta)$.

# Várható érték tulajdonságai

## Állítás

*Ha $\xi_1, \xi_2, \ldots, \xi_n$ véletlen változók, akkor*

$$\mathbf{E}\left(\sum_{i=1}^{n} \xi_i\right) = \sum_{i=1}^{n} \mathbf{E}(\xi_i).$$

### Példa

Csodaország munka törvénykönyve szerint egy cég minden munkása fizetett szabadságot kap azokon a napokon, amikor legalább az egyiküknek születésnapja van. Ezen napok kivételével azonban az év minden napján mindenkinek dolgoznia kell. Minden munkás 1 TV-készüléket készít egy nap alatt. Hány alkalmazottat vegyen fel a cégtulajdonos, ha azt akarja, hogy a gyártott TV-készülékek számának a várható értéke maximális legyen?

# Csodaország

# Csodaország

# Huffman-kód

$\mathcal{X} = \{x_1, x_2, \ldots, x_n\}$ véges halmaz, a *forrásábécé*.
Kód $f : \mathcal{X} \to \{$véges 0-1 sorozatok $\}$
Az $f$-hez tartozó lehetséges kódszavak $f(x_1), f(x_2), \ldots, f(x_n)$.

# Huffman-kód

$\mathcal{X} = \{x_1, x_2, \ldots, x_n\}$ véges halmaz, a *forrásábécé*.
Kód $f : \mathcal{X} \to \{$véges 0-1 sorozatok $\}$
Az $f$-hez tartozó lehetséges kódszavak $f(x_1), f(x_2), \ldots, f(x_n)$.
Az $f$ kód *prefix*, ha a lehetséges kódszavak közül egyik sem folytatása a másiknak. Jelölje $x \in \mathcal{X}$ esetén $|f(x)|$ a kódszó hosszát.

### Példa

Legyen $\mathcal{X} = \{a, b, c\}$, és legyen $f_1(a) = 0$, $f_1(b) = 01$, $f_1(c) = 011$. Ekkor $f_1$ nem prefix kód, de könnyen látható, hogy egyértelműen dekódolható. Az $f_2(a) = 01$, $f_2(b) = 00$, $f_2(c) = 1$, kód prefix.

Legyen $X$ egy véletlen betű, és eloszlása $\mathbf{P}(X = x_k) = p_k$, $k = 1, 2, \ldots, n$. Tehát $p_k$ a $x_k$ betű gyakorisága az adott nyelvben.

Adott $f$ kód esetén egy hosszú szövegben az egy karakterre eső átlagos kódszóhossz:

Feltehető, hogy $p_1 \geq p_2 \geq \ldots \geq p_n$. Ha az $f$ prefix kód optimális, akkor feltehető, hogy teljesülnek a következők:

(i) Hosszabb kódhoz ritkább betűk tartoznak, azaz

$$|f(x_1)| \leq |f(x_2)| \leq \ldots \leq |f(x_n)|.$$

Feltehető, hogy $p_1 \geq p_2 \geq \ldots \geq p_n$. Ha az $f$ prefix kód optimális, akkor feltehető, hogy teljesülnek a következők:

(i) Hosszabb kódhoz ritkább betűk tartoznak, azaz

$$|f(x_1)| \leq |f(x_2)| \leq \ldots \leq |f(x_n)|.$$

(ii) A két legkisebb valószínűséghez tartozó kód hossza egyenlő.

Feltehető, hogy $p_1 \geq p_2 \geq \ldots \geq p_n$. Ha az $f$ prefix kód optimális, akkor feltehető, hogy teljesülnek a következők:

(i) Hosszabb kódhoz ritkább betűk tartoznak, azaz

$$|f(x_1)| \leq |f(x_2)| \leq \ldots \leq |f(x_n)|.$$

(ii) A két legkisebb valószínűséghez tartozó kód hossza egyenlő.

(iii) $f(x_{n-1})$ és $f(x_n)$ csak az utolsó bitben térnek el.

## Tétel

*Tegyük fel, hogy az*

$$\mathcal{X}' = \{x_1, \ldots, x_{n-2}, y_{n-1}\}$$

$(n-1)$ *elemű forrásábécé és* $p_1, \ldots, p_{n-2}, p_{n-1} + p_n$ *eloszlás esetén* $g$ *egy optimális prefix kód. Ekkor az eredeti problémához tartozó optimális prefix kódot kapunk, ha az* $x_{n-1}$*, ill.* $x_n$ *kódját úgy választjuk, hogy a* $g(y_{n-1})$ *kódszót kiegészítjük 0-val, ill. 1-gyel, a többi kódszót változatlanul hagyjuk.*

## Példa

Legyen $n = 6$, $\mathcal{X} = \{x_1, \ldots, x_6\}$, és $p_1 = 0.132$, $p_2 = 0.329$, $p_3 = 0.329$, $p_4 = 0.165$, $p_5 = 0.041$, $p_6 = 0.004$.

(i) $x_5, x_6 \to x_{56}$, $p_{56} = 0.045$;

## Példa

Legyen $n = 6$, $\mathcal{X} = \{x_1, \ldots, x_6\}$, és $p_1 = 0.132$, $p_2 = 0.329$, $p_3 = 0.329$, $p_4 = 0.165$, $p_5 = 0.041$, $p_6 = 0.004$.

(i) $x_5, x_6 \to x_{56}$, $p_{56} = 0.045$;

(ii) $x_1, x_{56}$, $p_{156} = 0.177$;

## Példa

Legyen $n = 6$, $\mathcal{X} = \{x_1, \ldots, x_6\}$, és $p_1 = 0.132$, $p_2 = 0.329$, $p_3 = 0.329$, $p_4 = 0.165$, $p_5 = 0.041$, $p_6 = 0.004$.

(i) $x_5, x_6 \to x_{56}$, $p_{56} = 0.045$;

(ii) $x_1, x_{56}$, $p_{156} = 0.177$;

(iii) $x_{156}, x_4$, $p_{1564} = 0.342$;

## Példa

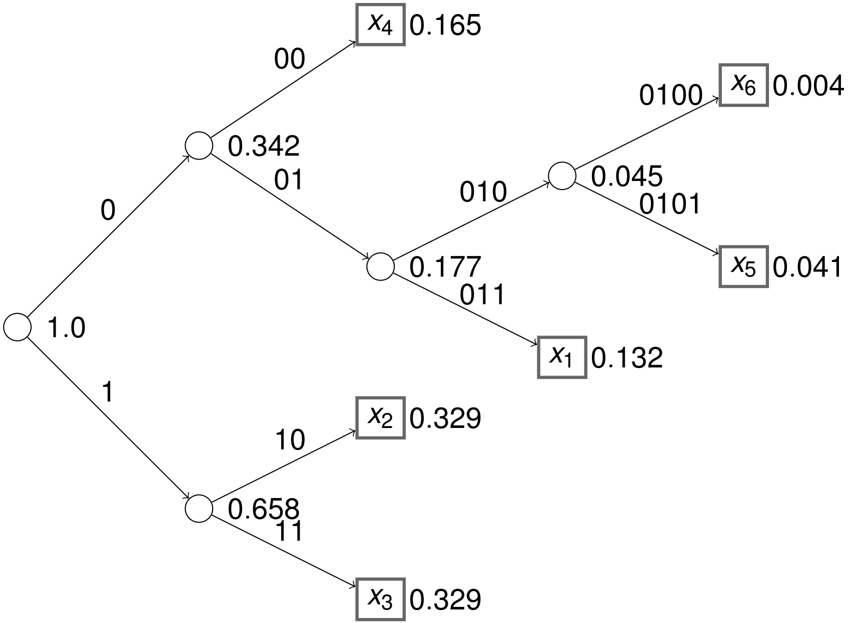
Legyen $n = 6$, $\mathcal{X} = \{x_1, \ldots, x_6\}$, és $p_1 = 0.132$, $p_2 = 0.329$, $p_3 = 0.329$, $p_4 = 0.165$, $p_5 = 0.041$, $p_6 = 0.004$.

(i) $x_5, x_6 \to x_{56}$, $p_{56} = 0.045$;

(ii) $x_1, x_{56}$, $p_{156} = 0.177$;

(iii) $x_{156}, x_4$, $p_{1564} = 0.342$;

(iv) $x_2, x_3$, $p_{23} = 0.658$.



$x_4$ 0.165
00
$x_6$ 0.004
0100
0.342
01
0.045
0101
0
010
$x_5$ 0.041
0.177
011
1.0
$x_1$ 0.132
1
$x_2$ 0.329
10
0.658
11
$x_3$ 0.329

Így az optimális kód:

| $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ |
|---|---|---|---|---|---|
| 011 | 10 | 11 | 00 | 0101 | 0100 |
| 0.132 | 0.329 | 0.329 | 0.165 | 0.041 | 0.004 |

A várható érték:

$$\begin{aligned}\mathbf{E}(f(X)) &= 0.132 \cdot 3 + 0.329 \cdot 2 + 0.329 \cdot 2 \\ &\quad + 0.165 \cdot 2 + 0.041 \cdot 4 + 0.004 \cdot 4 = 2.22\end{aligned}$$

# Entrópia

Az optimális várható kódhosszra teljesül, hogy

$$\sum_{k=1}^{n} p_k \log_2 \frac{1}{p_k} \leq \mathbf{E}(|f(\xi)|) < \sum_{k=1}^{n} p_k \log_2 \frac{1}{p_k} + 1.$$

Pl.: JPEG, MP3.