Házíme mincí

Chceme rozhodnout, zda je daná mince spravedlivá (orel padá s pravděpodobností 1/2) nebo falešná (orel padá s jinou pravděpodobností).

Ve 4 hodech padne 4x orel. Je mince spravedlivá?

Ve 40 hodech padne 40x orel. Je mince spravedlivá?

Nulová hypotéza H_0: orel padá s pravděpodobností $= 1/2$

Alternativní hypotéza H_1: orel padá s pravděpodobností $\neq 1/2$

Házíme mincí

Pro spravedlivou minci (za nulové hypotézy):

Jaký je očekávaný počet orlů ve 4, resp. ve 40 hodech mincí?

Jaká je pravděpodobnost, že ve 4, resp. ve 40 hodech mincí padnou samí orli?

Jaké je rozdělení počtu orlů ve 4, resp. ve 40 hodech mincí?

Házíme mincí

Spravedlivá mince:

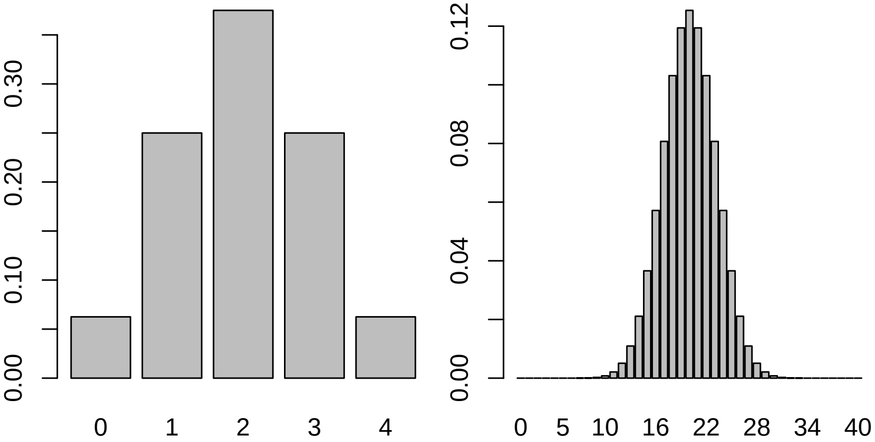
$X_4 \sim Bi(4, 1/2)$
$\mathbb{E}X_4 = 2$
$\mathbb{P}(X_4 = 4) = \left(\frac{1}{2}\right)^4 = 0{,}0625$

$X_{40} \sim Bi(40, 1/2)$
$\mathbb{E}X_{40} = 20$
$\mathbb{P}(X_{40} = 40) = \left(\frac{1}{2}\right)^{40} \doteq 10^{-12}$

Falešná mince:

???
(vlastnosti falešné mince neznáme přesně, je mnoho způsobů, jak se může lišit od spravedlivé)

Spravedlivá mince



K diskusi

Jaké hodnoty X_{40} hovoří ve prospěch nulové hypotézy (mince je spravedlivá) a jaké hodnoty hovoří proti ní?

Obvyklý postup

Proti nulové hypotéze hovoří velmi velké i velmi malé hodnoty X_{40}. Jaká hodnota už je *moc velká*, resp. *moc malá*?

Ponecháme si určitý prostor pro chybu, pravděpodobnost α (*hladina testu*), často volíme $\alpha = 0{,}05$.

Pak zvolíme *kritický obor W* hodnot veličiny X_{40}, pro které budeme zamítat nulovou hypotézu, tak, aby platilo $\mathbb{P}_{H_0}(X_{40} \in W) \leq \alpha$.

Všimněme si:
$\mathbb{P}(\text{zamítnu nulovou hypotézu, přestože platí}) = \mathbb{P}_{H_0}(X_{40} \in W) \leq \alpha$.

Obvyklý postup

V našem případě je vhodné volit $W = \{0, 1, \dots, 13, 27, 28, \dots, 40\}$, pak je $\mathbb{P}(X_{40} \in W) \doteq 0{,}0385$.

Tento kritický obor nejde rozšířit, aniž bychom překročili stanovenou hladinu 5 %. Jde ale volit třeba $W' = \{0, 1, \dots, 14, 28, 29, \dots, 40\}$, pak je $\mathbb{P}(X_{40} \in W') \doteq 0{,}0486$, ale kritický obor není symetrický.

Výsledek testu

$X_{40} \in W \Rightarrow$ zamítáme H_0 (máme dostatečně silné důkazy),
$X_{40} \notin W \Rightarrow$ nezamítáme H_0 (nemáme dostatečně silné důkazy).

Proč nejde nulovou hypotézu potvrdit? Může platit $p = 0{,}500001$, tedy nulová hypotéza neplatí, ale potřebovali bychom enormní množství dat, abychom v nich ten rozdíl poznali.

Nutno pečlivě volit, co prohlásíme za nulovou hypotézu.

H_0: nový lék má stejnou účinnost jako placebo,
H_1: nový lék má vyšší účinnost než placebo.

Zamítnutí H_0 je zde žádoucí.

Jiný pohled

Jaká je pravděpodobnost, že za platnosti H_0 napozorujeme stejně extrémní nebo ještě extrémnější data (stejně nebo silněji hovořící proti H_0)? Této pravděpodobnosti říkáme *p-hodnota*.

Předpokládejme, že jsme ve 40 hodech zaznamenali 38 orlů. Pak p-hodnota je $\mathbb{P}(X_{40} \in \{0, 1, 2, 38, 39, 40\}) \doteq 1{,}49 \cdot 10^{-9}$.

p-hodnota $\leq \alpha \Rightarrow$ zamítáme H_0 (máme dostatečně silné důkazy),
p-hodnota $> \alpha \Rightarrow$ nezamítáme H_0 (nemáme dostatečně silné důkazy).

V příkladu s televizemi a očekávanou dobou života byla p-hodnota $7 \cdot 10^{-6}$, měli jsme velmi silné důkazy hovořící proti nulové hypotéze $\rho = 0$.

Vlastnosti rozptylu

$$\mathrm{var}(aX) = \mathbb{E}(aX - \mathbb{E}aX)^2 = a^2\mathbb{E}(X - \mathbb{E}X)^2 = a^2\mathrm{var}X$$

X, Y nezávislé $\Rightarrow \mathrm{var}(X + Y) = \mathrm{var}X + \mathrm{var}\,Y$ (bez důkazu)

$X_1, \ldots, X_n$ nezávislé $\Rightarrow \mathrm{var}(X_1 + \ldots X_n) = \mathrm{var}X_1 + \ldots + \mathrm{var}X_n$

Výběrový průměr

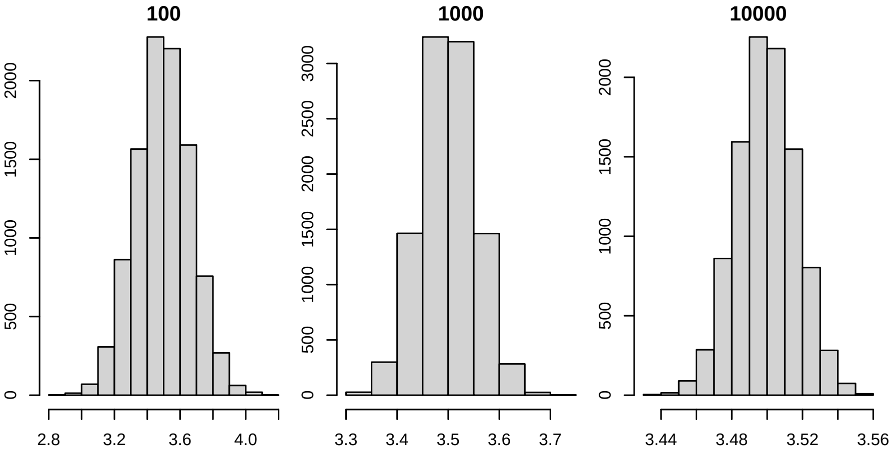
$X_1, \ldots, X_n$ nezávislé, stejně rozdělené, $\mathbb{E}X_i = \mu$, $\mathrm{var}X_i = \sigma^2 < \infty$

$\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$

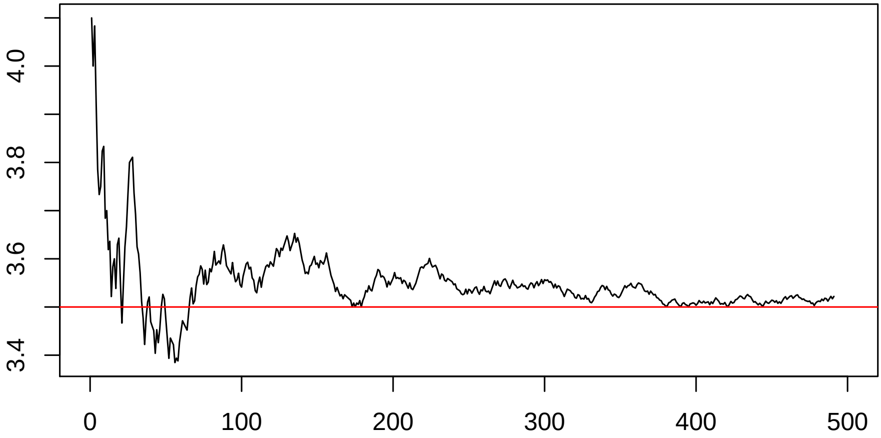
$\mathbb{E}\overline{X}_n = \frac{1}{n}\sum_{i=1}^n \mathbb{E}X_i = \frac{1}{n}\sum_{i=1}^n \mu = \mu$

$\mathrm{var}\overline{X}_n = \frac{1}{n^2}\mathrm{var}(\sum_{i=1}^n X_i) = \frac{1}{n^2}\sum_{i=1}^n \mathrm{var}X_i = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$

Průměry *k* hodů kostkou, 10 000 opakování



Průměry *k* hodů kostkou pro rostoucí *k* (prvních 10 hodnot vynecháno)



Průměry konvergují ke střední hodnotě

Viděli jsme, že $\mathbb{E}\overline{X}_n = \mu$ a $\operatorname{var}\overline{X}_n \to 0$ pro $n \to \infty$.

Rozdělení průměrů se více a více koncentruje kolem střední hodnoty.

Z toho se dá odvodit:

$$\forall \epsilon > 0 : \mathbb{P}\left(|\overline{X}_n - \mu| > \epsilon\right) \to 0, n \to \infty.$$

Tuto vlastnost popisuje *slabý zákon velkých čísel*,
píšeme $\overline{X}_n \stackrel{\mathbb{P}}{\to} \mu, n \to \infty$.

Příklad: průzkum veřejného mínění

$Y_i = \mathbb{I}(i\text{-tý respondent bude volit stranu XY})$

$\mathbb{E} Y_i = \mathbb{P}(Y_i = 1) = p$

Nezávisle vybereme n respondentů, dostaneme $Y_1, \ldots, Y_n$

$\overline{Y}_n \xrightarrow{\mathbb{P}} p, n \to \infty$