Čemu se (ne)máme divit

Mirko Navara
Centrum strojového vnímání, katedra kybernetiky
elektrotechnická fakulta ČVUT, Praha
http://cmp.felk.cvut.cz/~navara

27. 4. 2024

Motto

Kniha přírody je psána jazykem matematiky.

Galileo Galilei

Motto

Kniha přírody je psána jazykem matematiky.

Galileo Galilei

Ale dnes řekneme něco skoro bez vzorců.

Chaos v obálkách

Posílám 4 obálky, v každé je na začátku 10 lístků.
Kdykoli k vám dojdou, vyberte náhodně 2 obálky
a z jedné do druhé přesuňte 1 lístek (pokud tam je).

Náhodné losování

Máme vybrat jednoho za dvou lidí.
Střihněte si! (Kámen, nůžky, papír.)

Náhodné losování

Máme vybrat jednoho za dvou lidí.
Střihněte si! (Kámen, nůžky, papír.)
Jak dlouho potrvá, než se rozhodne?

Náhodné losování

Máme vybrat jednoho za dvou lidí.
Střihněte si! (Kámen, nůžky, papír.)
Jak dlouho potrvá, než se rozhodne?

počet kol n	pravděpodobnost konce v kole n	pravděpodobnost konce do kola n
1	$\frac{2}{3}$	$\frac{2}{3} \doteq 0.67$

Náhodné losování

Máme vybrat jednoho za dvou lidí.
Střihněte si! (Kámen, nůžky, papír.)
Jak dlouho potrvá, než se rozhodne?

počet kol n	pravděpodobnost konce v kole n	pravděpodobnost konce do kola n
1	$\frac{2}{3}$	$\frac{2}{3} \doteq 0.67$
2	$\frac{1}{3} \cdot \frac{2}{3}$	$\frac{8}{9} \doteq 0.89$

Náhodné losování

Máme vybrat jednoho za dvou lidí.
Střihněte si! (Kámen, nůžky, papír.)
Jak dlouho potrvá, než se rozhodne?

počet kol n	pravděpodobnost konce v kole n	pravděpodobnost konce do kola n
1	$\frac{2}{3}$	$\frac{2}{3} \doteq 0.67$
2	$\frac{1}{3} \cdot \frac{2}{3}$	$\frac{8}{9} \doteq 0.89$
3	$\frac{1}{3} \cdot \frac{1}{3} \cdot \frac{2}{3}$	$\frac{26}{27} \doteq 0.96$

Náhodné losování

Máme vybrat jednoho za dvou lidí.
Střihněte si! (Kámen, nůžky, papír.)
Jak dlouho potrvá, než se rozhodne?

počet kol n	pravděpodobnost konce v kole n	pravděpodobnost konce do kola n
1	$\frac{2}{3}$	$\frac{2}{3} \doteq 0.67$
2	$\frac{1}{3} \cdot \frac{2}{3}$	$\frac{8}{9} \doteq 0.89$
3	$\frac{1}{3} \cdot \frac{1}{3} \cdot \frac{2}{3}$	$\frac{26}{27} \doteq 0.96$
4	$\frac{1}{3} \cdot \frac{1}{3} \cdot \frac{1}{3} \cdot \frac{2}{3}$	$\frac{80}{81} \doteq 0.99$
...	...	...

Náhodné losování

Máme vybrat jednoho za dvou lidí.
Střihněte si! (Kámen, nůžky, papír.)
Jak dlouho potrvá, než se rozhodne?

počet kol n	pravděpodobnost konce v kole n	pravděpodobnost konce do kola n
1	$\frac{2}{3}$	$\frac{2}{3} \doteq 0.67$
2	$\frac{1}{3} \cdot \frac{2}{3}$	$\frac{8}{9} \doteq 0.89$
3	$\frac{1}{3} \cdot \frac{1}{3} \cdot \frac{2}{3}$	$\frac{26}{27} \doteq 0.96$
4	$\frac{1}{3} \cdot \frac{1}{3} \cdot \frac{1}{3} \cdot \frac{2}{3}$	$\frac{80}{81} \doteq 0.99$
...	...	...

Na 99 % budeme po 4 kolech hotovi,

Náhodné losování

Máme vybrat jednoho za dvou lidí.
Střihněte si! (Kámen, nůžky, papír.)
Jak dlouho potrvá, než se rozhodne?

počet kol n	pravděpodobnost konce v kole n	pravděpodobnost konce do kola n
1	$\frac{2}{3}$	$\frac{2}{3} \doteq 0.67$
2	$\frac{1}{3} \cdot \frac{2}{3}$	$\frac{8}{9} \doteq 0.89$
3	$\frac{1}{3} \cdot \frac{1}{3} \cdot \frac{2}{3}$	$\frac{26}{27} \doteq 0.96$
4	$\frac{1}{3} \cdot \frac{1}{3} \cdot \frac{1}{3} \cdot \frac{2}{3}$	$\frac{80}{81} \doteq 0.99$
...	...	...

Na 99 % budeme po 4 kolech hotovi, ale kdy to bude určitě?

Náhodné losování

Máme vybrat jednoho za dvou lidí.
Střihněte si! (Kámen, nůžky, papír.)
Jak dlouho potrvá, než se rozhodne?

počet kol n	pravděpodobnost konce v kole n	pravděpodobnost konce do kola n
1	$\frac{2}{3}$	$\frac{2}{3} \doteq 0.67$
2	$\frac{1}{3} \cdot \frac{2}{3}$	$\frac{8}{9} \doteq 0.89$
3	$\frac{1}{3} \cdot \frac{1}{3} \cdot \frac{2}{3}$	$\frac{26}{27} \doteq 0.96$
4	$\frac{1}{3} \cdot \frac{1}{3} \cdot \frac{1}{3} \cdot \frac{2}{3}$	$\frac{80}{81} \doteq 0.99$
...	...	...

Na 99 % budeme po 4 kolech hotovi, ale kdy to bude určitě?
Nikdy!

Chytřejší náhodné losování

Nápad: Změňme pravidla tak, aby bylo možno rozhodnout po konečném počtu kroků.

Chytřejší náhodné losování

Nápad: Změňme pravidla tak, aby bylo možno rozhodnout po konečném počtu kroků.
S volbou mezi třemi výsledky to pro dva hráče nejde, pro tři ano.

Chytřejší náhodné losování

Nápad: Změňme pravidla tak, aby bylo možno rozhodnout po konečném počtu kroků.
S volbou mezi třemi výsledky to pro dva hráče nejde, pro tři ano.
Ale dva hráči se mohou dohodnout např., že budou používat jen dva znaky (nůžky, papír);
jeden hráč vyhrává, dají-li stejné znaky, druhý při různých znacích.

Chytřejší náhodné losování

Nápad: Změňme pravidla tak, aby bylo možno rozhodnout po konečném počtu kroků.
S volbou mezi třemi výsledky to pro dva hráče nejde, pro tři ano.
Ale dva hráči se mohou dohodnout např., že budou používat jen dva znaky (nůžky, papír);
jeden hráč vyhrává, dají-li stejné znaky, druhý při různých znacích.
Je to spravedlivé a rozhodne se hned v prvním kole!

Házení mincí

Alice a Bob házejí mincí (společnou).
Bob vyhraje, padne-li **líc dvakrát po sobě**.
Alice vyhraje, padne-li **rub dvakrát po sobě**.
Je to spravedlivé?

Házení mincí

Alice a Bob házejí mincí (společnou).
Bob vyhraje, padne-li **líc dvakrát po sobě**.
Alice vyhraje, padne-li **rub dvakrát po sobě**.
Je to spravedlivé?
Ano. Líc a rub jsou rovnocenné.

Házení mincí

Alice a Bob házejí mincí (společnou).
Bob vyhraje, padne-li **líc dvakrát po sobě**.
Alice vyhraje, padne-li **líc a po něm rub**.
Je to spravedlivé?

Házení mincí

Alice a Bob házejí mincí (společnou).
Bob vyhraje, padne-li **líc dvakrát po sobě**.
Alice vyhraje, padne-li **líc a po něm rub**.
Je to spravedlivé?
Ano. Až padne poprvé líc, následující hod rozhodne.

Házení mincí

Alice a Bob házejí mincí (společnou).
Bob vyhraje, padne-li **líc dvakrát po sobě**.
Alice vyhraje, padne-li **rub a po něm líc**.
Je to spravedlivé?

Házení mincí

Alice a Bob házejí mincí (společnou).
Bob vyhraje, padne-li **líc dvakrát po sobě**.
Alice vyhraje, padne-li **rub a po něm líc**.
Je to spravedlivé?
Ne!

Házení mincí

Alice a Bob házejí mincí (společnou).
Bob vyhraje, padne-li **líc dvakrát po sobě**.
Alice vyhraje, padne-li **rub a po něm líc**.
Je to spravedlivé?
Ne!
Padne-li napoprvé líc, Bob čeká, jestli padne ještě jednou.

Házení mincí

Alice a Bob házejí mincí (společnou).
Bob vyhraje, padne-li **líc dvakrát po sobě**.
Alice vyhraje, padne-li **rub a po něm líc**.
Je to spravedlivé?
Ne!
Padne-li napoprvé líc, Bob čeká, jestli padne ještě jednou.
Padne-li napoprvé rub, Alice už jen čeká na svou jistou výhru.

Házení mincí

Alice a Bob házejí mincí (společnou).
Bob vyhraje, padne-li **líc dvakrát po sobě**.
Alice vyhraje, padne-li **rub a po něm líc**.
Je to spravedlivé?
Ne!
Padne-li napoprvé líc, Bob čeká, jestli padne ještě jednou.
Padne-li napoprvé rub, Alice už jen čeká na svou jistou výhru.
A pokud padl líc a rub, Alice opět očekává jistou výhru.
Sázím 3:1 na Alici.

Házení mincí

Alice a Bob házejí mincí (společnou).
Bob vyhraje, padne-li **líc dvakrát po sobě**.
Alice vyhraje, padne-li **rub a po něm líc**.
Je to spravedlivé?
Ne!
Padne-li napoprvé líc, Bob čeká, jestli padne ještě jednou.
Padne-li napoprvé rub, Alice už jen čeká na svou jistou výhru.
A pokud padl líc a rub, Alice opět očekává jistou výhru.
Sázím 3:1 na Alici. **Jistou** výhru?

Házení mincí

Alice a Bob házejí mincí (společnou).
Bob vyhraje, padne-li **líc dvakrát po sobě**.
Alice vyhraje, padne-li **rub a po něm líc**.
Je to spravedlivé?
Ne!
Padne-li napoprvé líc, Bob čeká, jestli padne ještě jednou.
Padne-li napoprvé rub, Alice už jen čeká na svou jistou výhru.
A pokud padl líc a rub, Alice opět očekává jistou výhru.
Sázím 3:1 na Alici. **Jistou** výhru?
Přesněji: s pravděpodobností 1.

Házení mincí

Alice a Bob házejí mincí (společnou).
Bob získává bod, padne-li **líc dvakrát po sobě** (a **hází se dál**).
Alice získává bod, padne-li **rub a po něm líc**.
Je to spravedlivé?

Házení mincí

Alice a Bob házejí mincí (společnou).
Bob získává bod, padne-li **líc dvakrát po sobě** (a **hází se dál**).
Alice získává bod, padne-li **rub a po něm líc**.
Je to spravedlivé?
Bob může bodovat třeba pořád, Alice nanejvýš v každém druhém hodu. Rozdělení pravděpodobností počtu bodů není stejné.

Házení mincí

Alice a Bob házejí mincí (společnou).
Bob získává bod, padne-li **líc dvakrát po sobě** (a **hází se dál**).
Alice získává bod, padne-li **rub a po něm líc**.
Je to spravedlivé?
Bob může bodovat třeba pořád, Alice nanejvýš v každém druhém hodu. Rozdělení pravděpodobností počtu bodů není stejné.
Ale je to spravedlivé! (Pokud hrají dlouho.)

Házení mincí

Alice a Bob házejí mincí (společnou).
Bob získává bod, padne-li **líc dvakrát po sobě** (a **hází se dál**).
Alice získává bod, padne-li **rub a po něm líc**.
Je to spravedlivé?
Bob může bodovat třeba pořád, Alice nanejvýš v každém druhém hodu. Rozdělení pravděpodobností počtu bodů není stejné.
Ale je to spravedlivé! (Pokud hrají dlouho.)
Kdybychom hru sledovali pozpátku, po líci by v následujícím kroku jeden bodoval: líc Bob, rub Alice.

Házení mincí

Alice a Bob házejí mincí (**každý svou**).
Bob čeká, až padne **líc dvakrát po sobě**.
Alice čeká, až padne **líc a po něm rub**.
Vyhrává ten, kdo dosáhne cíle dříve (po menším počtu hodů).
Je to spravedlivé?

Házení mincí

Alice a Bob házejí mincí (**každý svou**).
Bob čeká, až padne **líc dvakrát po sobě**.
Alice čeká, až padne **líc a po něm rub**.
Vyhrává ten, kdo dosáhne cíle dříve (po menším počtu hodů).
Je to spravedlivé?
Ne!
Bob potřebuje v průměru 6 hodů, Alice 4.

Buffonova úloha

Házíme jehlu na linkovaný papír.
Délka jehly je stejná jako vzdálenost linek.
Jaká je pravděpodobnost, že jehla protne některou z linek?

Buffonova úloha

Házíme jehlu na linkovaný papír.
Délka jehly je stejná jako vzdálenost linek.
Jaká je pravděpodobnost, že jehla protne některou z linek?

$$\frac{2}{\pi}$$

Monty Hall problem

Troje dveře, za jedněmi je výhra.
Moderátor ví, za kterými.

Monty Hall problem

Troje dveře, za jedněmi je výhra.
Moderátor ví, za kterými.
Vybereme jedny dveře.
Moderátor nám ukáže jedny z dveří, které jsme nevybrali, a za nimi nic není.
Nyní smíme změnit svoji volbu; máme to udělat?

Monty Hall problem

Troje dveře, za jedněmi je výhra.
Moderátor ví, za kterými.
Vybereme jedny dveře.
Moderátor nám ukáže jedny z dveří, které jsme nevybrali, a za nimi nic není.
Nyní smíme změnit svoji volbu; máme to udělat?

Názor sedláka: Je to jedno, máme šanci $1/3$.

Monty Hall problem

Troje dveře, za jedněmi je výhra.
Moderátor ví, za kterými.
Vybereme jedny dveře.
Moderátor nám ukáže jedny z dveří, které jsme nevybrali, a za nimi nic není.
Nyní smíme změnit svoji volbu; máme to udělat?

Názor sedláka: Je to jedno, máme šanci $1/3$.

Názor praktika: Znovu náhodně vybereme.

Monty Hall problem

Troje dveře, za jedněmi je výhra.
Moderátor ví, za kterými.
Vybereme jedny dveře.
Moderátor nám ukáže jedny z dveří, které jsme nevybrali, a za nimi nic není.
Nyní smíme změnit svoji volbu; máme to udělat?

Názor sedláka: Je to jedno, máme šanci $1/3$.

Názor praktika: Znovu náhodně vybereme.
Teď totiž vybíráme jedny ze dvojích dveří, máme šanci $1/2$.

Monty Hall problem

Troje dveře, za jedněmi je výhra.
Moderátor ví, za kterými.
Vybereme jedny dveře.
Moderátor nám ukáže jedny z dveří, které jsme nevybrali, a za nimi nic není.
Nyní smíme změnit svoji volbu; máme to udělat?

Názor sedláka: Je to jedno, máme šanci $1/3$.

Názor praktika: Znovu náhodně vybereme.
Teď totiž vybíráme jedny ze dvojích dveří, máme šanci $1/2$.

Názor matematika: Vždy měníme.

Monty Hall problem

Troje dveře, za jedněmi je výhra.
Moderátor ví, za kterými.
Vybereme jedny dveře.
Moderátor nám ukáže jedny z dveří, které jsme nevybrali, a za nimi nic není.
Nyní smíme změnit svoji volbu; máme to udělat?

Názor sedláka: Je to jedno, máme šanci $1/3$.

Názor praktika: Znovu náhodně vybereme.
Teď totiž vybíráme jedny ze dvojích dveří, máme šanci $1/2$.

Názor matematika: Vždy měníme. Máme šanci $2/3$!

Problém dvou obálek

Vylosovali jsme jednu ze dvou obálek a víme, že v jedné je $2\times$ víc než ve druhé.

Problém dvou obálek

Vylosovali jsme jednu ze dvou obálek a víme, že v jedné je $2\times$ víc než ve druhé. V obálce najdeme 2000. Nyní smíme obálky vyměnit. Vyplatí se to?

Problém dvou obálek

Vylosovali jsme jednu ze dvou obálek a víme, že v jedné je $2\times$ víc než ve druhé. V obálce najdeme 2000. Nyní smíme obálky vyměnit. Vyplatí se to?

Názor sedláka: Ne. Je to jedno.

Problém dvou obálek

Vylosovali jsme jednu ze dvou obálek a víme, že v jedné je $2\times$ víc než ve druhé. V obálce najdeme 2000. Nyní smíme obálky vyměnit. Vyplatí se to?

Názor sedláka: Ne. Je to jedno.

Názor praktika:
Je-li to ta cennější obálka, přijdu o 1000.
Je-li to ta méně cenná obálka, získám 2000.
Měním! V průměru vydělám 500.

Problém dvou obálek

Vylosovali jsme jednu ze dvou obálek a víme, že v jedné je $2\times$ víc než ve druhé. V obálce najdeme 2000. Nyní smíme obálky vyměnit. Vyplatí se to?

Názor sedláka: Ne. Je to jedno.

Názor praktika:
Je-li to ta cennější obálka, přijdu o 1000.
Je-li to ta méně cenná obálka, získám 2000.
Měním! V průměru vydělám 500.

Názor matematika:

Problém dvou obálek

Vylosovali jsme jednu ze dvou obálek a víme, že v jedné je $2\times$ víc než ve druhé. V obálce najdeme 2000. Nyní smíme obálky vyměnit. Vyplatí se to?

Názor sedláka: Ne. Je to jedno.

Názor praktika:
Je-li to ta cennější obálka, přijdu o 1000.
Je-li to ta méně cenná obálka, získám 2000.
Měním! V průměru vydělám 500.

Názor matematika: Je to jedno.
„Velká“ čísla jsou méně pravděpodobná než „malá“.

Problém dvou obálek

Vylosovali jsme jednu ze dvou obálek a víme, že v jedné je $2\times$ víc než ve druhé. V obálce najdeme 2000. Nyní smíme obálky vyměnit. Vyplatí se to?

Názor sedláka: Ne. Je to jedno.

Názor praktika:
Je-li to ta cennější obálka, přijdu o 1000.
Je-li to ta méně cenná obálka, získám 2000.
Měním! V průměru vydělám 500.

Názor matematika: Je to jedno.
„Velká“ čísla jsou méně pravděpodobná než „malá“. Ale neprodělám a mohu vydělat, pokud se budu rozhodovat náhodně a menší hodnoty měním s větší pravděpodobností než větší.

Nekonečný balíček karet

[Schrödinger]: Na nekonečný balíček karet natiskneme výhry;
na 1 výhru 1$,
na 10 výhru 10$,
na 100 výhru 100$...

Nekonečný balíček karet

[Schrödinger]: Na nekonečný balíček karet natiskneme výhry;
na 1 výhru 1$,
na 10 výhru 10$,
na 100 výhru 100$...
Každý hráč si vybere jednu kartu. Aniž by se na ně podíval, nalepí si ji na čelo, takže druhý ji vidí.
Chce si ji ponechat, nebo vyměnit se soupeřem?

Nekonečný balíček karet

[Schrödinger]: Na nekonečný balíček karet natiskneme výhry;
na 1 výhru 1$,
na 10 výhru 10$,
na 100 výhru 100$...
Každý hráč si vybere jednu kartu. Aniž by se na ně podíval, nalepí si ji na čelo, takže druhý ji vidí.
Chce si ji ponechat, nebo vyměnit se soupeřem?

Nekonečný balíček karet

[Schrödinger]: Na nekonečný balíček karet natiskneme výhry;
na 1 výhru 1$,
na 10 výhru 10$,
na 100 výhru 100$...
Každý hráč si vybere jednu kartu. Aniž by se na ně podíval, nalepí si ji na čelo, takže druhý ji vidí.
Chce si ji ponechat, nebo vyměnit se soupeřem?
Kde je chyba?

Nekonečný balíček karet

[Schrödinger]: Na nekonečný balíček karet natiskneme výhry;
na 1 výhru 1$,
na 10 výhru 10$,
na 100 výhru 100$...
Každý hráč si vybere jednu kartu. Aniž by se na ně podíval, nalepí si ji na čelo, takže druhý ji vidí.
Chce si ji ponechat, nebo vyměnit se soupeřem?
Kde je chyba?
Lze si snad představit nekonečný balíček karet, nebo aspoň stroj, který ho simuluje tím, že rozdá (konečně mnoho) karet z takového balíčku.

Nekonečný balíček karet

[Schrödinger]: Na nekonečný balíček karet natiskneme výhry;
na 1 výhru 1$,
na 10 výhru 10$,
na 100 výhru 100$...
Každý hráč si vybere jednu kartu. Aniž by se na ně podíval, nalepí si ji na čelo, takže druhý ji vidí.
Chce si ji ponechat, nebo vyměnit se soupeřem?
Kde je chyba?
Lze si snad představit nekonečný balíček karet, nebo aspoň stroj, který ho simuluje tím, že rozdá (konečně mnoho) karet z takového balíčku.
Jenže ten balíček nelze „spravedlivě zamíchat.“
Neexistuje rovnoměrné rozdělení na všech přirozených číslech.

Ruinování rulety

V ruletě si můžeme vsadit na barvu červenou nebo černou; pokud trefíme, dostaneme dvojnásobek vkladu.
Pravděpodobnost výhry je blízká 1/2.

Ruinování rulety

V ruletě si můžeme vsadit na barvu červenou nebo černou; pokud trefíme, dostaneme dvojnásobek vkladu.
Pravděpodobnost výhry je blízká 1/2.
Taktika:

1. Vsadíme 1 na jakoukoli barvu.
2. Pokud vyhrajeme, končíme se ziskem 1.
3. Pokud prohrajeme, zdvojnásobíme vklad a pokračujeme od bodu 2.

Ruinování rulety

V ruletě si můžeme vsadit na barvu červenou nebo černou; pokud trefíme, dostaneme dvojnásobek vkladu.
Pravděpodobnost výhry je blízká 1/2.
Taktika:

1. Vsadíme 1 na jakoukoli barvu.
2. Pokud vyhrajeme, končíme se ziskem 1.
3. Pokud prohrajeme, zdvojnásobíme vklad a pokračujeme od bodu 2.

Při konci po n kolech jsme vsadili $1 + 2 + 4 + \ldots + 2^{n-1} = 2^n - 1$ a vyhráli 2^n.
Končíme se ziskem 1, a to
s pravděpodobností > 99 % do 7 kol,
s pravděpodobností 1, je-li hra nekonečná.

Ruinování rulety

V ruletě si můžeme vsadit na barvu červenou nebo černou; pokud trefíme, dostaneme dvojnásobek vkladu.
Pravděpodobnost výhry je blízká $1/2$.
Taktika:

1. Vsadíme 1 na jakoukoli barvu.
2. Pokud vyhrajeme, končíme se ziskem 1.
3. Pokud prohrajeme, zdvojnásobíme vklad a pokračujeme od bodu 2.

Při konci po n kolech jsme vsadili $1 + 2 + 4 + \ldots + 2^{n-1} = 2^n - 1$ a vyhráli 2^n.
Končíme se ziskem 1, a to
s pravděpodobností > 99 % do 7 kol,
s pravděpodobností 1, je-li hra nekonečná.
Nedělejte to!

Chcete být milionářem? (zjednodušená verze bez záchytných bodů)

Zaplatíme 1000 za vstup do hry. Odpovídáme na otázky, pravděpodobnost správné odpovědi je c.
Po správné odpovědi můžeme skončit s výhrou $1000, 2000, 4000, 8000, \ldots$, obecně $1000 \cdot 2^{n-1}$, kde n je počet správně zodpovězených otázek.
Po chybné odpovědi končíme s nulou.

Chcete být milionářem? (zjednodušená verze bez záchytných bodů)

Zaplatíme 1000 za vstup do hry. Odpovídáme na otázky, pravděpodobnost správné odpovědi je c.
Po správné odpovědi můžeme skončit s výhrou $1000, 2000, 4000, 8000, \ldots$, obecně $1000 \cdot 2^{n-1}$, kde n je počet správně zodpovězených otázek.
Po chybné odpovědi končíme s nulou.

- Jak dlouho hrát, abychom maximalizovali zisk?

Chcete být milionářem? (zjednodušená verze bez záchytných bodů)

Zaplatíme 1000 za vstup do hry. Odpovídáme na otázky, pravděpodobnost správné odpovědi je c.
Po správné odpovědi můžeme skončit s výhrou $1000, 2000, 4000, 8000, \ldots$, obecně $1000 \cdot 2^{n-1}$, kde n je počet správně zodpovězených otázek.
Po chybné odpovědi končíme s nulou.

- Jak dlouho hrát, abychom maximalizovali zisk?
- Jaká je adekvátní cena za vstup do hry?

Chcete být milionářem?

Zaplatili jsme za vstup do hry, určitě máme hrát 1. kolo.
Před n-tým kolem se rozhodujeme mezi jistou výhrou 2^{n-1} a další otázkou, která může vést k výhře buď 0, nebo 2^n
(s pravděpodobností c).

Chcete být milionářem?

Zaplatili jsme za vstup do hry, určitě máme hrát 1. kolo.
Před n-tým kolem se rozhodujeme mezi jistou výhrou 2^{n-1} a další otázkou, která může vést k výhře buď 0, nebo 2^n
(s pravděpodobností c).
Optimální rozhodnutí bude vždy stejné (závislé na c, nikoli na n).
Pro $c < 1/2$ končíme po 1. kole.
Pro $c > 1/2$ je optimální hrát co nejdéle.

Chcete být milionářem?

Zaplatili jsme za vstup do hry, určitě máme hrát 1. kolo.
Před n-tým kolem se rozhodujeme mezi jistou výhrou 2^{n-1} a další otázkou, která může vést k výhře buď 0, nebo 2^n
(s pravděpodobností c).
Optimální rozhodnutí bude vždy stejné (závislé na c, nikoli na n).
Pro $c < 1/2$ končíme po 1. kole.
Pro $c > 1/2$ je optimální hrát co nejdéle.
Pravděpodobnost výhry je nulová!

Chcete být milionářem?

Ve 12. kole bychom vyhráli 4 096 000,
ale i při $c = 2/3$ se nám to s pravděpodobností > 99 % nepodaří.

Chcete být milionářem?

Ve 12. kole bychom vyhráli 4 096 000,
ale i při $c = 2/3$ se nám to s pravděpodobností $> 99\ \%$ nepodaří.
Nicméně průměrná výhra po 12. kole je

$$4\,096\,000 \cdot \left(\frac{2}{3}\right)^{12} \doteq 31\,569\,,$$

po n. kole

$$\left(\frac{2}{3}\right)^{n} 2^{n-1} = \frac{1}{2}\left(\frac{4}{3}\right)^{n} \to \infty \qquad \text{pro } n \to \infty\,.$$

Chcete být milionářem?

Ve 12. kole bychom vyhráli 4 096 000,
ale i při $c = 2/3$ se nám to s pravděpodobností > 99 % nepodaří.
Nicméně průměrná výhra po 12. kole je

$$4\,096\,000 \cdot \left(\frac{2}{3}\right)^{12} \doteq 31\,569\,,$$

po n. kole

$$\left(\frac{2}{3}\right)^{n} 2^{n-1} = \frac{1}{2}\left(\frac{4}{3}\right)^{n} \to \infty \qquad \text{pro } n \to \infty\,.$$

Adekvátní cena za vstup do hry je nekonečná!

Nejlepší hráč golfu

Občas se někomu podaří z odpalu přímo trefit jamku.

Nejlepší hráč golfu

Občas se někomu podaří z odpalu přímo trefit jamku.
Ale nedaří se mu to zopakovat.

Nejlepší hráč golfu

Občas se někomu podaří z odpalu přímo trefit jamku.
Ale nedaří se mu to zopakovat.
Jednomu se to podařilo.
Je to nejlepší hráč golfu?

Nejlepší hráč golfu

Občas se někomu podaří z odpalu přímo trefit jamku.
Ale nedaří se mu to zopakovat.
Jednomu se to podařilo.
Je to nejlepší hráč golfu? Není.
Ne proto, že to nedokázal zopakovat.

Nejlepší hráč golfu

Občas se někomu podaří z odpalu přímo trefit jamku.
Ale nedaří se mu to zopakovat.
Jednomu se to podařilo.
Je to nejlepší hráč golfu? Není.
Ne proto, že to nedokázal zopakovat.
Podrobnější analýza říká, že za rok je na celém světě asi 100 000 000 odpalů, takže pravděpodobnost, že se to někomu z tolika hráčů podaří, není zas tak malá.

Úmrtí po očkování

Jeden člověk v ČR po očkování proti covidu zemřel ještě dřív, než odešel z očkovacího centra.
Máme se bát?

Úmrtí po očkování

Jeden člověk v ČR po očkování proti covidu zemřel ještě dřív, než odešel z očkovacího centra.
Máme se bát? Ne!
Z 10 000 000 lidí zemře asi 400 za den, tj. v průměru 1 za 4 minuty.

Úmrtí po očkování

Jeden člověk v ČR po očkování proti covidu zemřel ještě dřív, než odešel z očkovacího centra.
Máme se bát? Ne!
Z 10 000 000 lidí zemře asi 400 za den, tj. v průměru 1 za 4 minuty.
Stejně nebezpečná tedy vyjde návštěva jakéhokoli jiného místa, např. restaurace, muzea, úřadu...

Úmrtí po očkování

Jeden člověk v ČR po očkování proti covidu zemřel ještě dřív, než odešel z očkovacího centra.
Máme se bát? Ne!
Z 10 000 000 lidí zemře asi 400 za den, tj. v průměru 1 za 4 minuty.
Stejně nebezpečná tedy vyjde návštěva jakéhokoli jiného místa, např. restaurace, muzea, úřadu...

Nejnebezpečnější činnost je ležet v posteli. Tam umřelo nejvíc lidí.

Syndrom náhlého úmrtí kojence

(Sally Clark): Syndrom náhlého úmrtí kojenců se vyskytuje s pravděpodobností $1/8500$.
Jaká je pravděpodobnost dvou úmrtí po sobě v jedné rodině?

Syndrom náhlého úmrtí kojence

(Sally Clark): Syndrom náhlého úmrtí kojenců se vyskytuje s pravděpodobností $1/8500$.
Jaká je pravděpodobnost dvou úmrtí po sobě v jedné rodině?
Soud uvěřil žalobci, že $1/8500^2 = 1/72\,250\,000$, a matku odsoudil.

Syndrom náhlého úmrtí kojence

(Sally Clark): Syndrom náhlého úmrtí kojenců se vyskytuje s pravděpodobností $1/8500$.
Jaká je pravděpodobnost dvou úmrtí po sobě v jedné rodině?
Soud uvěřil žalobci, že $1/8500^2 = 1/72\,250\,000$, a matku odsoudil.
Později osvobozena ona i 3 další ženy.

Syndrom náhlého úmrtí kojence

(Sally Clark): Syndrom náhlého úmrtí kojenců se vyskytuje s pravděpodobností $1/8500$.
Jaká je pravděpodobnost dvou úmrtí po sobě v jedné rodině?
Soud uvěřil žalobci, že $1/8500^2 = 1/72\,250\,000$, a matku odsoudil.
Později osvobozena ona i 3 další ženy.
I kdyby platila nezávislost, velmi pravděpodobně by byl někdo neprávem odsouzen.

Syndrom náhlého úmrtí kojence

(Sally Clark): Syndrom náhlého úmrtí kojenců se vyskytuje s pravděpodobností $1/8500$.
Jaká je pravděpodobnost dvou úmrtí po sobě v jedné rodině?
Soud uvěřil žalobci, že $1/8500^2 = 1/72\,250\,000$, a matku odsoudil.
Později osvobozena ona i 3 další ženy.
I kdyby platila nezávislost, velmi pravděpodobně by byl někdo neprávem odsouzen.
Není dostatek dat pro kvalifikovaný odhad.

Syndrom náhlého úmrtí kojence

(Sally Clark): Syndrom náhlého úmrtí kojenců se vyskytuje s pravděpodobností $1/8500$.
Jaká je pravděpodobnost dvou úmrtí po sobě v jedné rodině?
Soud uvěřil žalobci, že $1/8500^2 = 1/72\,250\,000$, a matku odsoudil.
Později osvobozena ona i 3 další ženy.
I kdyby platila nezávislost, velmi pravděpodobně by byl někdo neprávem odsouzen.
Není dostatek dat pro kvalifikovaný odhad.
Alespoň jsou známy vlivy, které (podmíněné) riziko snižují (*nekuřácká rodina, stabilní, dobře zajištěná*).

Syndrom náhlého úmrtí kojence

(Sally Clark): Syndrom náhlého úmrtí kojenců se vyskytuje s pravděpodobností $1/8500$.
Jaká je pravděpodobnost dvou úmrtí po sobě v jedné rodině?
Soud uvěřil žalobci, že $1/8500^2 = 1/72\,250\,000$, a matku odsoudil.
Později osvobozena ona i 3 další ženy.
I kdyby platila nezávislost, velmi pravděpodobně by byl někdo neprávem odsouzen.
Není dostatek dat pro kvalifikovaný odhad.
Alespoň jsou známy vlivy, které (podmíněné) riziko snižují (*nekuřácká rodina, stabilní, dobře zajištěná*).
Tytéž vlivy snižují i riziko vraždy novorozence, takže **podmíněná** pravděpodobnost, že šlo o trestný čin, zůstává nízká.

Higgsův boson

Detektor v CERNu zachytí *událost* (současný vznik Higgsova bosonu a páru top kvarků) s účinností 0.01;
falešně pozitivní je s pravděpodobností 0.000 05.
Hledaná událost se vyskytne v 6 případech z 10^{12}. Jaká je pravděpodobnost, že detekovaná událost skutečně nastala? [1]

[1] André Sopczak: Recognition and categorization of events resulting from proton collisions at CERN. Přednáška na katedře kybernetiky FEL ČVUT, 22. 1. 2019

Higgsův boson

Detektor v CERNu zachytí *událost* (současný vznik Higgsova bosonu a páru top kvarků) s účinností 0.01;
falešně pozitivní je s pravděpodobností 0.000 05.
Hledaná událost se vyskytne v 6 případech z 10^{12}. Jaká je pravděpodobnost, že detekovaná událost skutečně nastala? [1] H ... událost nastala, $P(H) = 6 \cdot 10^{-12}$,
D ... detektor hlásí událost, $P(D|H) = 0.01$,
$P(D|\bar{H}) = 5 \cdot 10^{-5}$.

[1] André Sopczak: Recognition and categorization of events resulting from proton collisions at CERN. Přednáška na katedře kybernetiky FEL ČVUT, 22. 1. 2019

Higgsův boson

Detektor v CERNu zachytí *událost* (současný vznik Higgsova bosonu a páru top kvarků) s účinností 0.01;
falešně pozitivní je s pravděpodobností 0.000 05.
Hledaná událost se vyskytne v 6 případech z 10^{12}. Jaká je pravděpodobnost, že detekovaná událost skutečně nastala? [1] H … událost nastala, $P(H) = 6 \cdot 10^{-12}$,
D … detektor hlásí událost, $P(D|H) = 0.01$,
$P(D|\bar{H}) = 5 \cdot 10^{-5}$.

$$P(D) = P(H) \cdot P(D|H) + P(\bar{H}) \cdot P(D|\bar{H}) =$$
$$= 6 \cdot 10^{-12} \cdot 0.01 + (1 - 6 \cdot 10^{-12}) \cdot 5 \cdot 10^{-5} = 5.000000006 \cdot 10^{-5}$$

$$P(H|D) = \frac{P(H) \cdot P(D|H)}{P(D)} \doteq \frac{6 \cdot 10^{-14}}{5 \cdot 10^{-5}} = 1.2 \cdot 10^{-9}\,.$$

[1] André Sopczak: Recognition and categorization of events resulting from proton collisions at CERN. Přednáška na katedře kybernetiky FEL ČVUT, 22. 1. 2019

První statistické průzkumy

Astrologové, zejména Johannes Kepler.

První statistické průzkumy

Astrologové, zejména Johannes Kepler.

- Položili počátky statistiky i astrologii.

První statistické průzkumy

Astrologové, zejména Johannes Kepler.

- Položili počátky statistiky i astrologii.
- Astrologická znamení mj. už dávno nejsou, co bývala.

První statistické průzkumy

Astrologové, zejména Johannes Kepler.

- Položili počátky statistiky i astrologii.
- Astrologická znamení mj. už dávno nejsou, co bývala.
- Přesto na datu narození závisí např. úspěšnost ve sportu (fotbal, šach).

První statistické průzkumy

Astrologové, zejména Johannes Kepler.

- Položili počátky statistiky i astrologii.
- Astrologická znamení mj. už dávno nejsou, co bývala.
- Přesto na datu narození závisí např. úspěšnost ve sportu (fotbal, šach).
- Jenže kvůli věkovým rozdílům při srovnávání.

První statistické průzkumy

Astrologové, zejména Johannes Kepler.

- Položili počátky statistiky i astrologii.
- Astrologická znamení mj. už dávno nejsou, co bývala.
- Přesto na datu narození závisí např. úspěšnost ve sportu (fotbal, šach).
- Jenže kvůli věkovým rozdílům při srovnávání.
- Nicméně byly zjištěny drobné odchylky mezi stejně starými dětmi narozenými v různých ročních obdobích, vysvětlují se sezónními rozdíly ve výživě matek.

První statistické průzkumy

Astrologové, zejména Johannes Kepler.

- Položili počátky statistiky i astrologii.
- Astrologická znamení mj. už dávno nejsou, co bývala.
- Přesto na datu narození závisí např. úspěšnost ve sportu (fotbal, šach).
- Jenže kvůli věkovým rozdílům při srovnávání.
- Nicméně byly zjištěny drobné odchylky mezi stejně starými dětmi narozenými v různých ročních obdobích, vysvětlují se sezónními rozdíly ve výživě matek.
- Rozhodně je mnoho podstatnějších vlivů, např. výživa dětí.

Čínské horoskopy

- Podle roku narození předpovídají možný rok úmrtí.

Čínské horoskopy

- Podle roku narození předpovídají možný rok úmrtí.
- Zjištěna závislost u populace Číňanů v USA.

Čínské horoskopy

- Podle roku narození předpovídají možný rok úmrtí.
- Zjištěna závislost u populace Číňanů v USA.
- Jenže funguje jen u těch, kteří tomu věří.

Čínské horoskopy

- Podle roku narození předpovídají možný rok úmrtí.
- Zjištěna závislost u populace Číňanů v USA.
- Jenže funguje jen u těch, kteří tomu věří.
- A ti umírají v průměru o 2 roky dříve!

Výhoda astrologie

Výhoda astrologie

- I kdybych zamíchal horoskopy jako karty, nehnal by mě nikdo k zodpovědnosti.

Výhoda astrologie

- I kdybych zamíchal horoskopy jako karty, nehnal by mě nikdo k zodpovědnosti.
- Pokud někdo chce poradit, jak zhodnotit své úspory nebo vyhrát soud s podvodníkem, potřebuje solidnější rady, i když ani ty nebudou zaručené.

Každý může být někomu dobrým astrologem (i finančním poradcem)

Obešleme 128 lidí.
Polovině předpovíme růst akcií, polovině pokles.

Každý může být někomu dobrým astrologem (i finančním poradcem)

Obešleme 128 lidí.
Polovině předpovíme růst akcií, polovině pokles.
V každé skupině polovině předpovíme růst jiných akcií, polovině pokles.

Každý může být někomu dobrým astrologem (i finančním poradcem)

Obešleme 128 lidí.
Polovině předpovíme růst akcií, polovině pokles.
V každé skupině polovině předpovíme růst jiných akcií, polovině pokles.
...
Po 7 kolech zbyde jeden, jemuž jsme vše předpověděli správně.
Ten si řekne: *„To nemůže být náhoda! Jinak by se na víc než* 99 % *někdy zmýlil. Tomu věřím!“*

Typické otázky

- Kde se déle žije, víc pije, víc vydělává?
- Co nám škodí/prospívá?
- Jak vyhodnotit úspěšnost předpovědí počasí?
- Proč mohou selhat volební předpovědi?

Zbohatlíkov [Swoboda]

Průměrný příjem rodiny je 82 320.
Příjem více než poloviny rodin je $\geq 29\,000$.

Zbohatlíkov [Swoboda]

Průměrný příjem rodiny je 82 320.
Příjem více než poloviny rodin je $\geq$ 29 000.
Nejčetnější příjem rodiny je 18 000.
Většina lidí nemá ani 7 500.
88 % lidí má méně než 25 000.

Zbohatlíkov [Swoboda]

Příjmy 25 rodin (v závorce počet členů rodiny):

1 200 000	(3)	60 000	(1)	45000	(2)
150 000	(5)	51 000	(3)	42 000	(2)
86 000	(4)	49 000	(4)	38 000	(4)
37 000	(3)	20 000	(7)	14 000	(1)
35 000	(5)	18 000	(3)	13 000	(4)
32 000	(3)	18 000	(8)	11 000	(1)
29 000	(3)	18 000	(4)	10 000	(2)
26 000	(4)	16 000	(3)		
24 000	(4)	16 000	(2)		

Zbohatlíkov [Swoboda]

Příjmy 25 rodin (v závorce počet členů rodiny):

1 200 000	(3)	60 000	(1)	45000	(2)
150 000	(5)	51 000	(3)	42 000	(2)
86 000	(4)	49 000	(4)	38 000	(4)
37 000	(3)	20 000	(7)	14 000	(1)
35 000	(5)	18 000	(3)	13 000	(4)
32 000	(3)	18 000	(8)	11 000	(1)
29 000	(3)	18 000	(4)	10 000	(2)
26 000	(4)	16 000	(3)		
24 000	(4)	16 000	(2)		

„Pan milionář si žije skvěle, obklopen několika tucty chudých otroků.“

Průměr nebo medián?

[Gonick, Smith] V r. 1984 University of Virginia oznámila, že průměrný nástupní plat absolventů katedry rétoriky a komunikací byl (tehdy impozantních) 55 000 $.

Průměr nebo medián?

[Gonick, Smith] V r. 1984 University of Virginia oznámila, že průměrný nástupní plat absolventů katedry rétoriky a komunikací byl (tehdy impozantních) 55 000 \$. Příčinou nebyly tam vyučované dovednosti, ale jeden absolvent – basketbalový profesionál Ralph Sampson.

Průměr nebo medián?

[Gonick, Smith] V r. 1984 University of Virginia oznámila, že průměrný nástupní plat absolventů katedry rétoriky a komunikací byl (tehdy impozantních) 55 000 \$. Příčinou nebyly tam vyučované dovednosti, ale jeden absolvent – basketbalový profesionál Ralph Sampson.

Bývalý premiér: „Co je to za stát, kde 2/3 obyvatel má podprůměrný příjem?“

Průměr nebo medián?

[Gonick, Smith] V r. 1984 University of Virginia oznámila, že průměrný nástupní plat absolventů katedry rétoriky a komunikací byl (tehdy impozantních) 55 000 $. Příčinou nebyly tam vyučované dovednosti, ale jeden absolvent – basketbalový profesionál Ralph Sampson.

Bývalý premiér: „Co je to za stát, kde 2/3 obyvatel má podprůměrný příjem?“
A co má být? Co navrhujete?

Průměr nebo medián?

[Gonick, Smith] V r. 1984 University of Virginia oznámila, že průměrný nástupní plat absolventů katedry rétoriky a komunikací byl (tehdy impozantních) 55 000 \$. Příčinou nebyly tam vyučované dovednosti, ale jeden absolvent – basketbalový profesionál Ralph Sampson.

Bývalý premiér: „Co je to za stát, kde 2/3 obyvatel má podprůměrný příjem?"
A co má být? Co navrhujete?

Medián pravděpodobnostního rozdělení je taková hodnota, že
s pravděpodobností 1/2 dostáváme hodnoty **větší**,
s pravděpodobností 1/2 dostáváme hodnoty **menší**.

„Statistika je věda o přesném součtu nepřesných čísel“

„Statistika je věda o přesném součtu nepřesných čísel“

Není!

„Statistika je věda o přesném součtu nepřesných čísel"

Není!

- **Absolutní** chyba součtu roste, jen **relativní** chyba klesá (a zůstává nenulová).

„Statistika je věda o přesném součtu nepřesných čísel"

Není!

- **Absolutní** chyba součtu roste, jen **relativní** chyba klesá (a zůstává nenulová).
- Lépe by bylo hovořit o „přesném průměru nepřesných čísel".

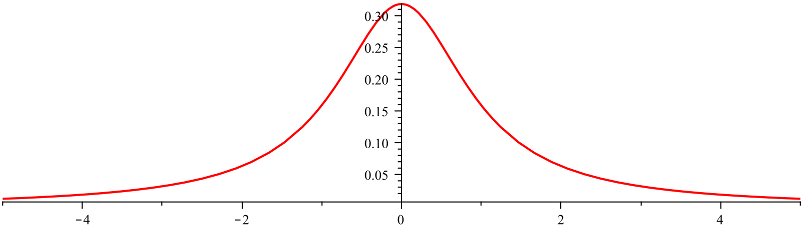
„Statistika je věda o přesném součtu nepřesných čísel"

Není!

- **Absolutní** chyba součtu roste, jen **relativní** chyba klesá (a zůstává nenulová).
- Lépe by bylo hovořit o „přesném průměru nepřesných čísel".
- To vše však jen za určitých předpokladů (např. nezávislost).

Protipříklad

Cauchyho rozdělení (=rozdělení podílu dvou nezávislých náhodných veličin s Gaussovým normálním rozdělením).



Pro něj má průměr dokonce stejné rozdělení jako původní! Nemá střední hodnotu.

Role statistiky ve vědě

[Janko: Jak vytváří statistika obrazy života a světa]

Není potřeba, pokud jde o zjevné zákonitosti, ale odhaluje nám řadu jemnějších vztahů.

Příklady závislostí:

- Jedovatost muchomůrky hlíznaté – nesporná

Příklady závislostí:

- Jedovatost muchomůrky hlíznaté – nesporná
- Jedovatost zmije – nezpochybňovaná

Příklady závislostí:

- Jedovatost muchomůrky hlíznaté – nesporná
- Jedovatost zmije – nezpochybňovaná
- Škodlivost dalších vlivů – diskutovaná
    - Hliníkové nádobí
    - Chronický únavový syndrom
    - Syndrom války v Zálivu
    - Letní čas

Příklady závislostí:

- Jedovatost muchomůrky hlíznaté – nesporná
- Jedovatost zmije – nezpochybňovaná
- Škodlivost dalších vlivů – diskutovaná
  - Hliníkové nádobí
  - Chronický únavový syndrom
  - Syndrom války v Zálivu
  - Letní čas
- Placebo efekt apod.

Příklady závislostí:

- Jedovatost muchomůrky hlíznaté – nesporná
- Jedovatost zmije – nezpochybňovaná
- Škodlivost dalších vlivů – diskutovaná
  - Hliníkové nádobí
  - Chronický únavový syndrom
  - Syndrom války v Zálivu
  - Letní čas
- Placebo efekt apod.
- Nepodložená tvrzení
  - Geopatogenní zóny
  - Účinky homeopatik

Chceme-li si udělat jasno, **neobejdeme se bez statistiky**.

K čemu je tedy statistika?

Statistika dovoluje rozpoznat **společné vlastnosti velkých souborů**.

K čemu je tedy statistika?

Statistika dovoluje rozpoznat **společné vlastnosti velkých souborů**.

Nabízí pouze problematický kompromis, nic víc ani očekávat nelze:

- Závěry nejsou zaručené, pouze hodně pravděpodobné.

K čemu je tedy statistika?

Statistika dovoluje rozpoznat **společné vlastnosti velkých souborů**.

Nabízí pouze problematický kompromis, nic víc ani očekávat nelze:

- Závěry nejsou zaručené, pouze hodně pravděpodobné.
- Nelze je přímo uplatnit na jednotlivce.

K čemu je tedy statistika?

Statistika dovoluje rozpoznat **společné vlastnosti velkých souborů**.
Nabízí pouze problematický kompromis, nic víc ani očekávat nelze:

- Závěry nejsou zaručené, pouze hodně pravděpodobné.
- Nelze je přímo uplatnit na jednotlivce.
- Hypotézy lze pouze **zamítat** jako nepravděpodobné, **nikoli potvrdit**, jeví-li se pravděpodobné.

Dvojitě zaslepený experiment

Jak statistik očkuje dvojčata.

Dvojitě zaslepený experiment

Jak statistik očkuje dvojčata.

Ideál: Experimentátor ani účastník experimentu nemá vědět, zda je ve výběrovém nebo kontrolním souboru.

Dvojitě zaslepený experiment

Jak statistik očkuje dvojčata.

Ideál: Experimentátor ani účastník experimentu nemá vědět, zda je ve výběrovém nebo kontrolním souboru.
Někdy je těžké to dodržet.

Časté omyly v chápání statistiky

Proč máme statistiku respektovat i s těmito nedostatky:

- Kdy máme očekávat stoletou vodu?
- Kdy máme očekávat výhru v loterii?
- Kdo je na titulní stránce, s tím to půjde z kopce.

Časté omyly v chápání statistiky

Proč máme statistiku respektovat i s těmito nedostatky:

- Kdy máme očekávat stoletou vodu?
- Kdy máme očekávat výhru v loterii?
- Kdo je na titulní stránce, s tím to půjde z kopce.

Stejný princip platí i pro akcie a investiční fondy!

Kde všude je/má být statistika

$\Longrightarrow$ je třeba vyučovat statistiku a pravděpodobnost už od základní školy, na odpovídající úrovni (kvalitativní, nikoli kvantitativní).

Kde je místo statistiky?

„V dotazníku byli respondenti dotázáni následovně: Absolvovaná fakulta ... jako celek podle Vás patřila ve srovnání s jinými fakultami (případně školami) mezi:
1 – velmi špatné,
2 – velmi podprůměrné,
3 – podprůměrné,
4 – průměrné,
5 – nadprůměrné,
6 – vysoce nadprůměrné,
7 – vynikající.
Pomocí škály vyplývající z nabízených odpovědí (od 1 do 7) byl pro každou školu a fakultu vypočítán průměr

[2] Zásady financování VŠ, 3.2011, Příloha 5

Kde je místo statistiky?

„V dotazníku byli respondenti dotázáni následovně: Absolvovaná fakulta ... jako celek podle Vás patřila ve srovnání s jinými fakultami (případně školami) mezi:
1 – velmi špatné,
2 – velmi podprůměrné,
3 – podprůměrné,
4 – průměrné,
5 – nadprůměrné,
6 – vysoce nadprůměrné,
7 – vynikající.
Pomocí škály vyplývající z nabízených odpovědí (od 1 do 7) byl pro každou školu a fakultu vypočítán průměr, přičemž hodnota průměru za celý soubor činí **4,94**.“[2]

[2]Zásady financování VŠ, 3.2011, Příloha 5

Kde je místo statistiky?

„V dotazníku byli respondenti dotázáni následovně: Absolvovaná fakulta ... jako celek podle Vás patřila ve srovnání s jinými fakultami (případně školami) mezi:
1 – velmi špatné,
2 – velmi podprůměrné,
3 – podprůměrné,
4 – průměrné,
5 – nadprůměrné,
6 – vysoce nadprůměrné,
7 – vynikající.
Pomocí škály vyplývající z nabízených odpovědí (od 1 do 7) byl pro každou školu a fakultu vypočítán průměr, přičemž hodnota průměru za celý soubor činí **4,94**.“[2]
Tedy v průměru byli z nadprůměrných škol.

[2] Zásady financování VŠ, 3.2011, Příloha 5

Kde je místo statistiky?

„V dotazníku byli respondenti dotázáni následovně: Absolvovaná fakulta … jako celek podle Vás patřila ve srovnání s jinými fakultami (případně školami) mezi:
1 – velmi špatné,
2 – velmi podprůměrné,
3 – podprůměrné,
4 – průměrné,
5 – nadprůměrné,
6 – vysoce nadprůměrné,
7 – vynikající.
Pomocí škály vyplývající z nabízených odpovědí (od 1 do 7) byl pro každou školu a fakultu vypočítán průměr, přičemž hodnota průměru za celý soubor činí **4,94**."[2]

Tedy v průměru byli z nadprůměrných škol.

Je možné, že odpovídali správně?

[2] Zásady financování VŠ, 3.2011, Příloha 5

Jaký práh pro zamítnutí hypotézy?

Test na nemoc, škodlivost léku apod.

Jaký práh pro zamítnutí hypotézy?

Test na nemoc, škodlivost léku apod.
V podstatě obrácený „boží soud“, respektujeme presumpci neviny.

Jaký práh pro zamítnutí hypotézy?

Test na nemoc, škodlivost léku apod.
V podstatě obrácený „boží soud", respektujeme presumpci neviny.
Můžeme nastavit práh citlivosti, a tím měnit pravděpodobnost, že obviníme nevinného, ale současně také pravděpodobnost, že neodhalíme viníka.

Test na drogy

Motivační příklad (test na drogy): Policií užívaný test DrugWipe 5S je falešně pozitivní u 5 % nezdrogovaných a falešně negativní u 3 % zdrogovaných. Jaká je (podmíněná) pravděpodobnost, že řidič s pozitivním testem je pod vlivem drog?

Test na drogy

Motivační příklad (test na drogy): Policií užívaný test DrugWipe 5S je falešně pozitivní u 5 % nezdrogovaných a falešně negativní u 3 % zdrogovaných. Jaká je (podmíněná) pravděpodobnost, že řidič s pozitivním testem je pod vlivem drog?

H ... zdrogovaný, D ... pozitivní test,

$P(D|\bar{H}) = 0.05,$ $P(D|H) = 1 - P(\bar{D}|H) = 1 - 0.03 = 0.97.$

Test na drogy

Motivační příklad (test na drogy): Policií užívaný test DrugWipe 5S je falešně pozitivní u 5 % nezdrogovaných a falešně negativní u 3 % zdrogovaných. Jaká je (podmíněná) pravděpodobnost, že řidič s pozitivním testem je pod vlivem drog?

H ... zdrogovaný, D ... pozitivní test,

$P(D|\bar{H}) = 0.05,$ $P(D|H) = 1 - P(\bar{D}|H) = 1 - 0.03 = 0.97.$

Potřebujeme ještě znát podíl zdrogovaných řidičů na silnicích $P(H)$.

Test na drogy

Motivační příklad (test na drogy): Policií užívaný test DrugWipe 5S je falešně pozitivní u 5 % nezdrogovaných a falešně negativní u 3 % zdrogovaných. Jaká je (podmíněná) pravděpodobnost, že řidič s pozitivním testem je pod vlivem drog?

H … zdrogovaný, D … pozitivní test,

$P(D|\bar{H}) = 0.05,$ $P(D|H) = 1 - P(\bar{D}|H) = 1 - 0.03 = 0.97.$

Potřebujeme ještě znát podíl zdrogovaných řidičů na silnicích $P(H)$. Policejní ani statistická data nemáme, tedy jen zkusíme, co bychom dostali pro

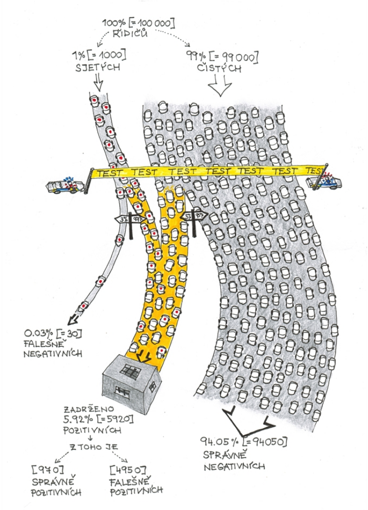
$P(H) = 0.01,$ $P(\bar{H}) = 0.99$:

Test na drogy

Motivační příklad (test na drogy): Policií užívaný test DrugWipe 5S je falešně pozitivní u 5 % nezdrogovaných a falešně negativní u 3 % zdrogovaných. Jaká je (podmíněná) pravděpodobnost, že řidič s pozitivním testem je pod vlivem drog?
H ... zdrogovaný, D ... pozitivní test,
$P(D|\bar{H}) = 0.05$, $P(D|H) = 1 - P(\bar{D}|H) = 1 - 0.03 = 0.97$.
Potřebujeme ještě znát podíl zdrogovaných řidičů na silnicích $P(H)$. Policejní ani statistická data nemáme, tedy jen zkusíme, co bychom dostali pro
$P(H) = 0.01$, $P(\bar{H}) = 0.99$:

$$P(H|D) = \frac{P(H) \cdot P(D|H)}{P(H) \cdot P(D|H) + P(\bar{H}) \cdot P(D|\bar{H})} = \frac{0.0097}{0.0592} \doteq 0.164\,.$$

Test na drogy



Zdroj: https://finmag.penize.cz/kaleidoskop/407932-velka

Jaký práh pro zamítnutí hypotézy?

Obvykle nastavujeme podle pravděpodobnosti, že obviníme nevinného.

Jaký práh pro zamítnutí hypotézy?

Obvykle nastavujeme podle pravděpodobnosti, že obviníme nevinného.
To můžeme, aniž bychom věděli parametry druhé skupiny (nemocných, padělků atd.).

Jaký práh pro zamítnutí hypotézy?

Obvykle nastavujeme podle pravděpodobnosti, že obviníme nevinného.
To můžeme, aniž bychom věděli parametry druhé skupiny (nemocných, padělků atd.).
Někdy minimalizujeme ztrátu při chybném rozhodnutí.

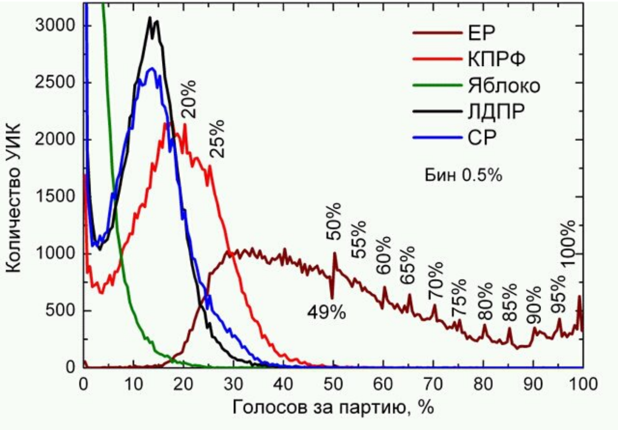
Jaký práh pro zamítnutí hypotézy?

Obvykle nastavujeme podle pravděpodobnosti, že obviníme nevinného.
To můžeme, aniž bychom věděli parametry druhé skupiny (nemocných, padělků atd.).
Někdy minimalizujeme ztrátu při chybném rozhodnutí.
Ta bude jiná u spam filtru a jiná u pyrotechnika.

Jaký práh pro zamítnutí hypotézy?

Obvykle nastavujeme podle pravděpodobnosti, že obviníme nevinného.
To můžeme, aniž bychom věděli parametry druhé skupiny (nemocných, padělků atd.).
Někdy minimalizujeme ztrátu při chybném rozhodnutí.
Ta bude jiná u spam filtru a jiná u pyrotechnika.
U gravitestu by se hodily dvě varianty, ale na to si lidé těžko zvykají.

Volby v Rusku do Státní dumy 2011 – výsledky stran po okrscích[3]



[3] http://www.pirati.cz/tiskove-zpravy/rusko-falsovani-voleb-vypinani-internetu-pirati-v-ulicich

Newcombův-Benfordův zákon

Vezměte

- poslední 3 nenulové cifry vašeho (nebo jiného) telefonního čísla,
- číslo domu,
- číslo bot,

Newcombův-Benfordův zákon

Vezměte

- poslední 3 nenulové cifry vašeho (nebo jiného) telefonního čísla,
- číslo domu,
- číslo bot,

vše vynásobte.

Newcombův-Benfordův zákon

Vezměte

- poslední 3 nenulové cifry vašeho (nebo jiného) telefonního čísla,
- číslo domu,
- číslo bot,

vše vynásobte.
Komu výsledek začíná číslicí 1?

Newcombův-Benfordův zákon

Vezměte

- poslední 3 nenulové cifry vašeho (nebo jiného) telefonního čísla,
- číslo domu,
- číslo bot,

vše vynásobte.
Komu výsledek začíná číslicí 1?
Na podobném principu se vyhledávají podvodná data u finančních transakcí, daňových dokladů apod.

Statistické odhady

Jsou zatíženy chybou; chceme, aby nebyla systematická:

Statistické odhady

Jsou zatíženy chybou; chceme, aby nebyla systematická:
Nestranný odhad je takový, že jeho střední hodnota je stejná jako střední hodnota odhadovaného rozdělení.

Statistické odhady

Jsou zatíženy chybou; chceme, aby nebyla systematická:
Nestranný odhad je takový, že jeho střední hodnota je stejná jako střední hodnota odhadovaného rozdělení.
Ve stavebnictví se zjevně nepoužívá. ☹

Statistické odhady

Jsou zatíženy chybou; chceme, aby nebyla systematická:

Nestranný odhad je takový, že jeho střední hodnota je stejná jako střední hodnota odhadovaného rozdělení.

Ve stavebnictví se zjevně nepoužívá. ☹

Někdy jsou jiná kritéria pro volbu odhadu, např. minimalizace rizik. (Odhad věku, IQ apod.)

Co lze říci o přesnosti volebních preferencí?

Ve vzorku 1000 z 10 000 000 voličů se 200 respondentů vyjádřilo, že budou volit určitou stranu.
Odhad preferencí je $200/1000 = 20\,\%$.
Jaká je garantovaná přesnost tohoto odhadu?

Co lze říci o přesnosti volebních preferencí?

Ve vzorku 1000 z 10 000 000 voličů se 200 respondentů vyjádřilo, že budou volit určitou stranu.
Odhad preferencí je $200/1000 = 20\,\%$.
Jaká je garantovaná přesnost tohoto odhadu?

S **jistotou** lze říci pouze, že preference jsou 0.000 02 **až** 0.999 92, protože zaručíme jen to, že aspoň 200 voličů ji bude volit a aspoň 800 ji volit nebude (a to ještě nesmí změnit názor).

Co lze říci o přesnosti volebních preferencí?

Ve vzorku 1000 z 10 000 000 voličů se 200 respondentů vyjádřilo, že budou volit určitou stranu.
Odhad preferencí je $200/1000 = 20\,\%$.
Jaká je garantovaná přesnost tohoto odhadu?

S **jistotou** lze říci pouze, že preference jsou 0.000 02 **až** 0.999 92, protože zaručíme jen to, že aspoň 200 voličů ji bude volit a aspoň 800 ji volit nebude (a to ještě nesmí změnit názor).

Více nám řekne informace, že preference jsou 16.7 **až** 23.3 % **s pravděpodobností** 99 %. (K tomu je potřebný správný výběr respondentů.)

Co lze říci o přesnosti volebních preferencí?

Ve vzorku 1000 z 10 000 000 voličů se 200 respondentů vyjádřilo, že budou volit určitou stranu.
Odhad preferencí je $200/1000 = 20\,\%$.
Jaká je garantovaná přesnost tohoto odhadu?

S **jistotou** lze říci pouze, že preference jsou 0.000 02 **až** 0.999 92, protože zaručíme jen to, že aspoň 200 voličů ji bude volit a aspoň 800 ji volit nebude (a to ještě nesmí změnit názor).

Více nám řekne informace, že preference jsou 16.7 **až** 23.3 % **s pravděpodobností** 99 %. (K tomu je potřebný správný výběr respondentů.)

Podobně: Kdy mám vyjet, abych byl na místě určitě včas?

D'Hondtova metoda v praxi

Tabulka udává rozdělení (podmíněné) pravděpodobnosti, že volič strany zastoupené v parlamentu volil danou stranu. Posuďte, zda tomu odpovídá rozdělení poslaneckých křesel.[4]

relativní preference	0.376	0.344	0.136	0.077	0.067	1
počet poslanců	81	74	26	13	6	200
teor. četnost	75.2	68.8	27.2	15.4	13.4	200

Hypotézu nezamítáme, takový (nebo ještě nevyváženější) výsledek má pravděpodobnost 7.5 % (poněkud překvapivý závěr vzhledem k tomu, že poslední dvě strany mají téměř stejnou podporu u voličů, ale poslední má více než 2× méně poslanců).

[4] Údaje z parlamentních voleb v ČR 2006.

Jak (ne)vyšly předpovědi výsledků voleb v ČR 2013

Vstupní data[5]

Rozsah	**Volby**	Factum 1000	Median 1000	STEM 1000	CVVM 1000	Sanep 7000	Médea 1000	Aisa 1000
ČSSD	**20.5**	**22.8**	25.5	25.9	26	23.8	**22.2**	**23**
ANO	**18.7**	12.1	13	16.1	**16.5**	11.6	**16.9**	16
KSČM	**14.9**	**17.1**	**16**	**13.3**	18	16.9	11.8	**14**
TOP 09	**12**	**13.2**	**13**	**11.5**	9	**11.9**	9.6	**10.5**
ODS	**7.7**	**7.2**	**8**	**8.6**	**6.5**	**7.5**	5.5	**7**
Úsvit	**6.9**	3.7	4	**5.9**	5	5.3	**8.2**	**6**
KDU	**6.8**	**5.9**	**6**	4.5	5	5.7	**6.2**	**6**
SZ	**3.2**	**4.3**	**3**	**2.6**	2	**3.1**	**2.9**	**3**
SPOZ	**1.5**	4.7	5	2.6	3.5	5.2	3.7	4
ostatní	**7.8**	**9**	**6.5**	**9**	**8.5**	**9**	13	10.5
Účast	**59.5**	54	**60**	67	63	**59.3**	71	78
χ^2-test		≈ 0	≈ 0	$5 \cdot 10^{-6}$	$2 \cdot 10^{-12}$	≈ 0	≈ 0	$3 \cdot 10^{-10}$

Tučně jsou vyznačeny předpovědi, které se vejdou do 95% symetrického intervalového odhadu.

[5] http://zpravy.idnes.cz/jak-se-trefily-pruzkumy-0wa-/domaci.aspx?c=A131027_120534_domaci_jw. Rozsah výběru je pouze přibližný.

Proč volební předpovědi nevycházejí?

1. Lidé nejsou losy, své názory mění a nesdělují je vždy správně.

Proč volební předpovědi nevycházejí?

1. Lidé nejsou losy, své názory mění a nesdělují je vždy správně.
2. Výsledky jsou ovlivněny volební neúčastí.

Proč volební předpovědi nevycházejí?

1. Lidé nejsou losy, své názory mění a nesdělují je vždy správně.
2. Výsledky jsou ovlivněny volební neúčastí.
3. Průzkumy jsou zdlouhavé, tudíž provedené v delším časovém období před volbami.

Proč volební předpovědi nevycházejí?

1. Lidé nejsou losy, své názory mění a nesdělují je vždy správně.
2. Výsledky jsou ovlivněny volební neúčastí.
3. Průzkumy jsou zdlouhavé, tudíž provedené v delším časovém období před volbami.
4. Výběrový soubor nemá zcela odpovídající složení, což se kompenzuje váhováním údajů; tím však vzroste rozptyl.

Proč volební předpovědi nevycházejí?

1. Lidé nejsou losy, své názory mění a nesdělují je vždy správně.
2. Výsledky jsou ovlivněny volební neúčastí.
3. Průzkumy jsou zdlouhavé, tudíž provedené v delším časovém období před volbami.
4. Výběrový soubor nemá zcela odpovídající složení, což se kompenzuje váhováním údajů; tím však vzroste rozptyl.
5. **Celý model je špatně, protože měření ovlivňuje zkoumaný systém (stejně jako v kvantové fyzice).**

„Kvantové“ jevy ve statistice

Omezené možnosti opakovaného nebo současného testování jevů
$\Longrightarrow$ paradoxy nevysvětlitelné klasickou teorií.
Nejen ve fyzice mikročástic, ale např. v sociologii a psychologii.

- Můžeme najednou provést jen některé z možných testů.
- Původní stav systému nelze znovu nastavit.

Důsledky ignorování kvantových jevů ve statistice

- Použití klasické statistiky může vést ke zcela chybným závěrům kvůli ignorování nesprávných předpokladů.
- Nejedná se o **statistickou** chybu, ale o chybu v popisu, kterou klasická teorie ani nemůže vyjádřit.

Robustní statistika

- Klasická statistika pracovala především s chybami, které měly přibližně normální rozdělení, vznikaly **součem mnoha malých nezávislých chyb**.

Robustní statistika

- Klasická statistika pracovala především s chybami, které měly přibližně normální rozdělení, vznikaly **součem mnoha malých nezávislých chyb**.
- S digitalizací dat dochází častěji k **hrubým chybám** (chyby přenosu dat, špatně umístěná desetinná čárka apod.); postupy je potřeba jim přizpůsobit.

Robustní statistika

- Klasická statistika pracovala především s chybami, které měly přibližně normální rozdělení, vznikaly **součem mnoha malých nezávislých chyb**.
- S digitalizací dat dochází častěji k **hrubým chybám** (chyby přenosu dat, špatně umístěná desetinná čárka apod.); postupy je potřeba jim přizpůsobit.
- V klasickém modelu by takové chyby byly považovány za vysoce nepravděpodobné.

Fuzzifikace statistiky

- Statistika se z principu zabývá jevy, jejichž výskyt lze postihnout dvěma pravdivostními hodnotami.

Fuzzifikace statistiky

- Statistika se z principu zabývá jevy, jejichž výskyt lze postihnout dvěma pravdivostními hodnotami.
- Okolní svět však často popisujeme v termínech, které nejsou dvouhodnotové („černobílé"), např. „zítra očekáváme místy přízemní mrazíky a četné přeháňky, ojediněle sněhové".
- To jsou užitečné informace, avšak takové, že na ně nelze např. uzavírat sázky.
- Jsou svou povahou vícehodnotové, **fuzzy**.

Fuzzifikace statistiky

- Statistika se z principu zabývá jevy, jejichž výskyt lze postihnout dvěma pravdivostními hodnotami.
- Okolní svět však často popisujeme v termínech, které nejsou dvouhodnotové („černobílé"), např. „zítra očekáváme místy přízemní mrazíky a četné přeháňky, ojediněle sněhové".
- To jsou užitečné informace, avšak takové, že na ně nelze např. uzavírat sázky.
- Jsou svou povahou vícehodnotové, **fuzzy**.
- Statistika se takovým jevům dosud vyhýbá, což může mít negativní důsledky.
- Učící se klasifikátor se hodně zdrží, když jeho klasifikace sporného případu bude ohodnocena jako zcela špatná.

Paradox hromady

- „Odborníci Fondu OSN pro otázky populace tvrdí, že se roku 2050 nemusíme bát. Svět podle nich začne být lidstvu skutečně těsný až ve chvíli, kdy počet obyvatel planety dosáhne čtrnácti miliard.“[6]

[6] Sandra Krištofová: Rok 2050: bude nás devět miliard. Planeta začne být lidstvu těsná. MF Dnes, 22. 3. 2001

Paradox hromady

- „Odborníci Fondu OSN pro otázky populace tvrdí, že se roku 2050 nemusíme bát. Svět podle nich začne být lidstvu skutečně těsný až ve chvíli, kdy počet obyvatel planety dosáhne čtrnácti miliard.“[6]
- … a televize bude při tom, až se narodí 14 000 000 000-tý občan …

[6] Sandra Krištofová: Rok 2050: bude nás devět miliard. Planeta začne být lidstvu těsná. MF Dnes, 22. 3. 2001

Paradox hromady

- „Odborníci Fondu OSN pro otázky populace tvrdí, že se roku 2050 nemusíme bát. Svět podle nich začne být lidstvu skutečně těsný až ve chvíli, kdy počet obyvatel planety dosáhne čtrnácti miliard.“[6]
- … a televize bude při tom, až se narodí 14 000 000 000-tý občan …
- Řešení paradoxu hromady ve fuzzy logice: S každým narozeným člověkem roste pravdivost výroku „Země je přelidněná“.

[6] Sandra Krištofová: Rok 2050: bude nás devět miliard. Planeta začne být lidstvu těsná. MF Dnes, 22. 3. 2001

Proč ještě nemáme kvantovou fuzzy statistiku?

- Jakékoli zobecnění znamená, že model má více parametrů.
- Je snadné kritizovat, že klasické výsledky jsou chybné, ale mnohem těžší navrhnout **funkční** alternativu.
- Nicméně je potřebné ji hledat.

Obvyklé aplikace statistiky

- V reklamě: Průzkumy trhu, zaměření reklamy na cílovou skupinu.

Obvyklé aplikace statistiky

- V reklamě: Průzkumy trhu, zaměření reklamy na cílovou skupinu.
- V politice: Totéž.

Obvyklé aplikace statistiky

- V reklamě: Průzkumy trhu, zaměření reklamy na cílovou skupinu.
- V politice: Totéž.
- V podnikání: Průzkumy trhu, odhad adekvátního rozsahu výroby a odbytu.

Obvyklé aplikace statistiky

- V reklamě: Průzkumy trhu, zaměření reklamy na cílovou skupinu.
- V politice: Totéž.
- V podnikání: Průzkumy trhu, odhad adekvátního rozsahu výroby a odbytu.

Obvyklé aplikace statistiky

Umí naznačit, co by spolu mohlo souviset.

Obvyklé aplikace statistiky

Umí naznačit, co by spolu mohlo souviset.
Problém je určit, zda je příčinná souvislost a kterým směrem.

Obvyklé aplikace statistiky

Umí naznačit, co by spolu mohlo souviset.
Problém je určit, zda je příčinná souvislost a kterým směrem.
Bohatší lidé se dožívají vyššího věku.
Ženy matky mají vyšší IQ než bezdětné.
...

Neobvyklé aplikace statistiky a pravděpodobnosti

- V bankovnictví a kriminalistice: Vytipování podezřelých finančních operací.

Neobvyklé aplikace statistiky a pravděpodobnosti

- V bankovnictví a kriminalistice: Vytipování podezřelých finančních operací.
- V daňovém systému: Vytipování zfalšovaných dat.

Neobvyklé aplikace statistiky a pravděpodobnosti

- V bankovnictví a kriminalistice: Vytipování podezřelých finančních operací.
- V daňovém systému: Vytipování zfalšovaných dat.
- Ve strategických hrách (šachy, go).

Neobvyklé aplikace statistiky a pravděpodobnosti

- V bankovnictví a kriminalistice: Vytipování podezřelých finančních operací.
- V daňovém systému: Vytipování zfalšovaných dat.
- Ve strategických hrách (šachy, go).
- V komunikaci (ethernet): Pokud si dva počítače „skáčí do řeči“ (=začaly vysílat oba současně), odmlčí se oba na náhodně zvolenou dobu, než pokus o vysílání zopakují.

Podezřelé losování

- V r. 2006 v ČR proběhlo výběrové řízení: Z 16 firem bylo pro 4 zakázky vylosováno do užšího výběru vždy 5 firem, jejichž nabídky byly dále posuzovány. Jedna z firem se zúčastnila prostřednictvím čtyř filiálek, tj. podala 4 z 16 přihlášek.
- S pravděpodobností $> 99\,\%$ by firma uspěla už v některém ze 3 kol.

[7] Informace laskavě poskytl Prof. RNDr. J. Štěpán, DrSc., který se spolupracovníky vypracoval znalecký posudek a referoval o něm v televizi.

Podezřelé losování

- V r. 2006 v ČR proběhlo výběrové řízení: Z 16 firem bylo pro 4 zakázky vylosováno do užšího výběru vždy 5 firem, jejichž nabídky byly dále posuzovány. Jedna z firem se zúčastnila prostřednictvím čtyř filiálek, tj. podala 4 z 16 přihlášek.
- S pravděpodobností $> 99\,\%$ by firma uspěla už v některém ze 3 kol.
- Nebyla vybrána v žádném ze 4 kol.
  Firma zažalovala zadavatele, neboť považuje losování za zmanipulované.[7]

[7] Informace laskavě poskytl Prof. RNDr. J. Štěpán, DrSc., který se spolupracovníky vypracoval znalecký posudek a referoval o něm v televizi.

Podezřelé losování

- V r. 2006 v ČR proběhlo výběrové řízení: Z 16 firem bylo pro 4 zakázky vylosováno do užšího výběru vždy 5 firem, jejichž nabídky byly dále posuzovány. Jedna z firem se zúčastnila prostřednictvím čtyř filiálek, tj. podala 4 z 16 přihlášek.
- S pravděpodobností $> 99\,\%$ by firma uspěla už v některém ze 3 kol.
- Nebyla vybrána v žádném ze 4 kol.
  Firma zažalovala zadavatele, neboť považuje losování za zmanipulované.[7]
- Stejně neúspěšných bylo celkem 9 firem z 16. Kdyby se tak losovalo každý rok, tak nepravděpodobný výsledek by vyšel **v průměru jednou za 333 000 let**.

[7] Informace laskavě poskytl Prof. RNDr. J. Štěpán, DrSc., který se spolupracovníky vypracoval znalecký posudek a referoval o něm v televizi.

Neobvyklé aplikace statistiky a pravděpodobnosti[8]

[8] Mahdian, B.: Blind Digital Image Authentication. PhD. Thesis, FJFI ČVUT, Praha, 2008

Neobvyklé aplikace statistiky a pravděpodobnosti

Figure 1.7: An example of famous image manipulations of recent years.

Figure 1.8: An example of famous image manipulations of recent years.

Neobvyklé aplikace statistiky a pravděpodobnosti

- V kódování: Generování klíče pro šifrování obvykle začíná: „vezmeme dvě velká prvočísla p, q a vypočteme $N = p\,q$“.[9]

[9] Demlová, M.: Střípky z diskrétní matematiky. Inspirativní setkání s matematikou, FEL ČVUT, 8. března 2013

Neobvyklé aplikace statistiky a pravděpodobnosti

- V kódování: Generování klíče pro šifrování obvykle začíná: „vezmeme dvě velká prvočísla p, q a vypočteme $N = p\,q$“.[9]
- Ve skutečnosti se vezmou dvě náhodně vybraná čísla a provede se řada **nezávislých** testů, zda nejsou něčím dělitelná.

[9] Demlová, M.: Střípky z diskrétní matematiky. Inspirativní setkání s matematikou, FEL ČVUT, 8. března 2013

Neobvyklé aplikace statistiky a pravděpodobnosti

- V kódování: Generování klíče pro šifrování obvykle začíná: „vezmeme dvě velká prvočísla p, q a vypočteme $N = p\,q$“.[9]
- Ve skutečnosti se vezmou dvě náhodně vybraná čísla a provede se řada **nezávislých** testů, zda nejsou něčím dělitelná.
- Připouštíme, že to nemusí být prvočísla a vzniklá šifra RSA bude vadná, pokud riziko je $\leq 2^{-100} \doteq 7.9 \cdot 10^{-31}$.

[9] Demlová, M.: Střípky z diskrétní matematiky. Inspirativní setkání s matematikou, FEL ČVUT, 8. března 2013

Neobvyklé aplikace statistiky a pravděpodobnosti

- Buňky v mozku ptáků rozpoznávající směr (*head direction cells*) mají přesnost cca 45 stupňů. Průměruje se jich $> 200\ 000$. Tím se dosahuje vysoká přesnost.[10]

---

[10] Pavel Němec: Neurální substrát magnetické kompasové orientace ptáků, přednáška, CTS, Praha 20.10.2011

[11] Matas, J.: Strojové učení a jak poznat ve fotografiích kluky od holek a jak z mnoha slabých studentů vytvořit jednoho experta. Inspirativní setkání s matematikou, FEL ČVUT, 13. prosince 2013

Neobvyklé aplikace statistiky a pravděpodobnosti

- Buňky v mozku ptáků rozpoznávající směr (*head direction cells*) mají přesnost cca 45 stupňů. Průměruje se jich $> 200\,000$. Tím se dosahuje vysoká přesnost.[10]
- Statistické rozpoznávání na základě „hlasování" mnoha nedokonalých klasifikátorů.[11]

[10] Pavel Němec: Neurální substrát magnetické kompasové orientace ptáků, přednáška, CTS, Praha 20.10.2011

[11] Matas, J.: Strojové učení a jak poznat ve fotografiích kluky od holek a jak z mnoha slabých studentů vytvořit jednoho experta. Inspirativní setkání s matematikou, FEL ČVUT, 13. prosince 2013