

Přednáška V.

Úvod do teorie odhadu

- ➔ Pojmy a principy teorie odhadu
- ➔ Nestranné odhady
- ➔ Metoda maximální věrohodnosti
- ➔ Průměr vs. medián



evropský
sociální
fond v ČR



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání
pro konkurenceschopnost



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Opakování – střední hodnota

→ Uvažujme diskrétní náhodnou veličinu

$$\rightarrow X = \{x_1, \dots, x_k\}$$

$$\rightarrow P(X=x_1) = p_1, \dots, P(X=x_k) = p_k$$

→ Pak střední hodnota má tvar:

$$E(X) = \mu = \sum_{i=1}^k x_i p(x_i)$$

→ **Jaká je její interpretace?**



Opakování – pravidlo ± 3 sigma

➔ Co to znamená?

➔ K čemu to může být dobré?

1. Pojmy a principy teorie odhadu

Jak se vlastně přišlo na použití průměru?

→ Použití průměru jako sumarizace n pozorovaných hodnot se učí už na základní škole, nicméně zmínka o jeho používání je až z konce 17. století.

→ Byl navržen bez ohledu na jakoukoliv souvislost s teorií pravděpodobnosti jako hodnota, označme ji a , která má následující vlastnosti:

1. Hodnota a minimalizuje reziduální součet čtverců, tedy součet čtverců rozdílů pozorovaných hodnot a hodnoty a :

$$\sum_{i=1}^n (x_i - a)^2 = \sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - a)^2$$

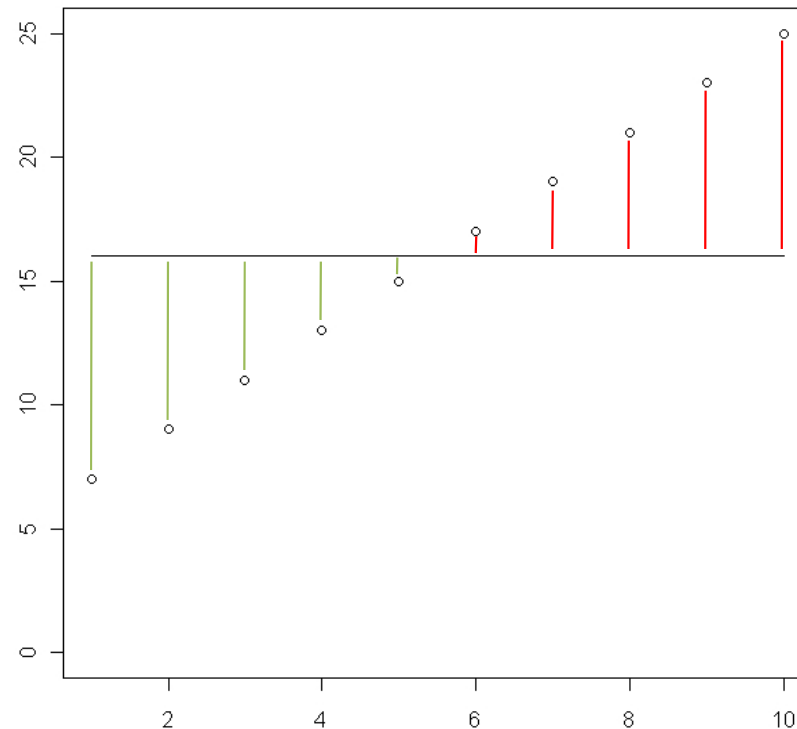
2. Součet reziduí vzhledem k hodnotě a je nula, tedy kladná i záporná rezidua jsou v rovnováze:

$$\sum_{i=1}^n (x_i - a) = 0$$

→ Tyto dvě kritéria zohledňují pouze pozorovaná data, vůbec se nezabývají jakýmkoliv rozdělením pravděpodobnosti a jeho parametry.

Příklad – průměr pozorovaných hodnot

- ➔ V případě, že osa x nepředstavuje žádnou informaci, je použití průměru v pořádku (kladná i záporná rezidua jsou v rovnováze).



- ➔ Co když osa x ponese nějakou informaci?

Cíl snažení v teorii odhadu

- ➔ Na základě reálných pozorování náhodné veličiny X chceme získat informaci o parametrech rozdělení pravděpodobnosti této veličiny.
- ➔ Teorie odhadu se snaží sestavit statistiku, která by na základě pozorovaných dat poskytla nejlepší možný odhad neznámého parametru / parametrů.
- ➔ Teorie odhadu předpokládá, že pozorované hodnoty nesou informaci o neznámém parametru.
- ➔ Někdy je třeba pozorované hodnoty před použitím statistiky „značně“ upravit
→ normalizace dat z DNA mikročipů.



Základní pojmy

- ➔ **Náhodná veličina** X – číselné ohodnocení výsledku experimentu, zajímá nás její pravděpodobnostní chování – popisuje ho **rozdělení pravděpodobnosti** náhodné veličiny X .
- ➔ **Parametr** rozdělení pravděpodobnosti – neznámá hodnota, θ , na které závisí předpis rozdělení pravděpodobnosti
- ➔ **Parametrická funkce** – reálná funkce parametru θ .

- ➔ **Realizace náhodné veličiny** (n realizací) – představují je pozorované hodnoty: $\mathbf{x} = x_1, x_2, \dots, x_n$. Předpokládám jejich vzájemnou nezávislost.
- ➔ **Odhad parametru** θ – reálná funkce $\mathbf{x} = d(\mathbf{x}) = \hat{\theta}$.
- ➔ **Odhad parametrické funkce** $g(\theta)$ – reálná funkce $\mathbf{x} = d(\mathbf{x}) = g(\hat{\theta})$.

Klasifikace odhadů

- ➔ **Parametrické odhady** – vycházejí z předpokladu znalosti rozdělení pravděpodobnosti, kterým se náhodná veličina řídí. Případně předpokládají i znalost rozdělení pravděpodobnosti sledovaného parametru (tedy náhodné veličiny) – Bayesovské odhady.
- ➔ **Neparametrické odhady** – v tomto případě nejsou uvažovány žádné předpoklady o pravděpodobnostním chování dat. Výsledkem jsou robustní odhady se širokým použitím, u kterých ale nelze hodnotit optimálnost vzhledem k pravděpodobnostnímu modelu.

Klíčové otázky v teorii odhadu

- ➔ Jak najít bodový odhad?
- ➔ Jak hodnotit kvalitu odhadu?

Jak najít bodový odhad?

- Existuje řada postupů k nalezení bodového odhadu neznámého parametru – liší se jak filozofií (např. Bayesovské odhady) tak definicí kritéria optimálních vlastností odhadu. Zaměříme se pouze na vybrané pojmy a postupy.
- **Metoda založená na Rao-Blackwellově větě** – slouží k nalezení nestranného odhadu s nejmenší variabilitou (ne vždy to však lze spočítat).
- **Metoda maximální věrohodnosti** – slouží k nalezení odhadu (hodnoty), který je ve smyslu pozorovaných dat nejvíce pravděpodobný. Respektive lze říci, že při „platnosti“ této hodnoty jsou data nejvíce věrohodná.
- **Bayesovské metody** – nehledají jednu hodnotu parametru, ale celé rozdělení pravděpodobnosti (parametr je zde vlastně náhodná veličina).
- ...

2. Nestranné odhady

Střední kvadratická chyba odhadu

- Významnou rizikovou funkcí ve statistice je tzv. **střední kvadratická chyba odhadu** („mean squared error“) definovaná jako

$$MSE(\theta, \hat{\theta}) = E_{\theta}((\hat{\theta} - \theta)^2)$$

- Výraz pro MSE , respektive MSE odhadu, se dá rozdělit na dvě komponenty – **vychýlení** (jeho druhou mocninu) a **variabilitu**:

$$MSE(\theta, \hat{\theta}) = E_{\theta}((\hat{\theta} - \theta + E(\hat{\theta}) - E(\hat{\theta}))^2) = (\theta - E(\hat{\theta}))^2 - E((\hat{\theta} - E(\hat{\theta}))^2)$$

$$MSE(\theta, \hat{\theta}) = \text{bias}^2(\hat{\theta}) + \text{var}(\hat{\theta})$$

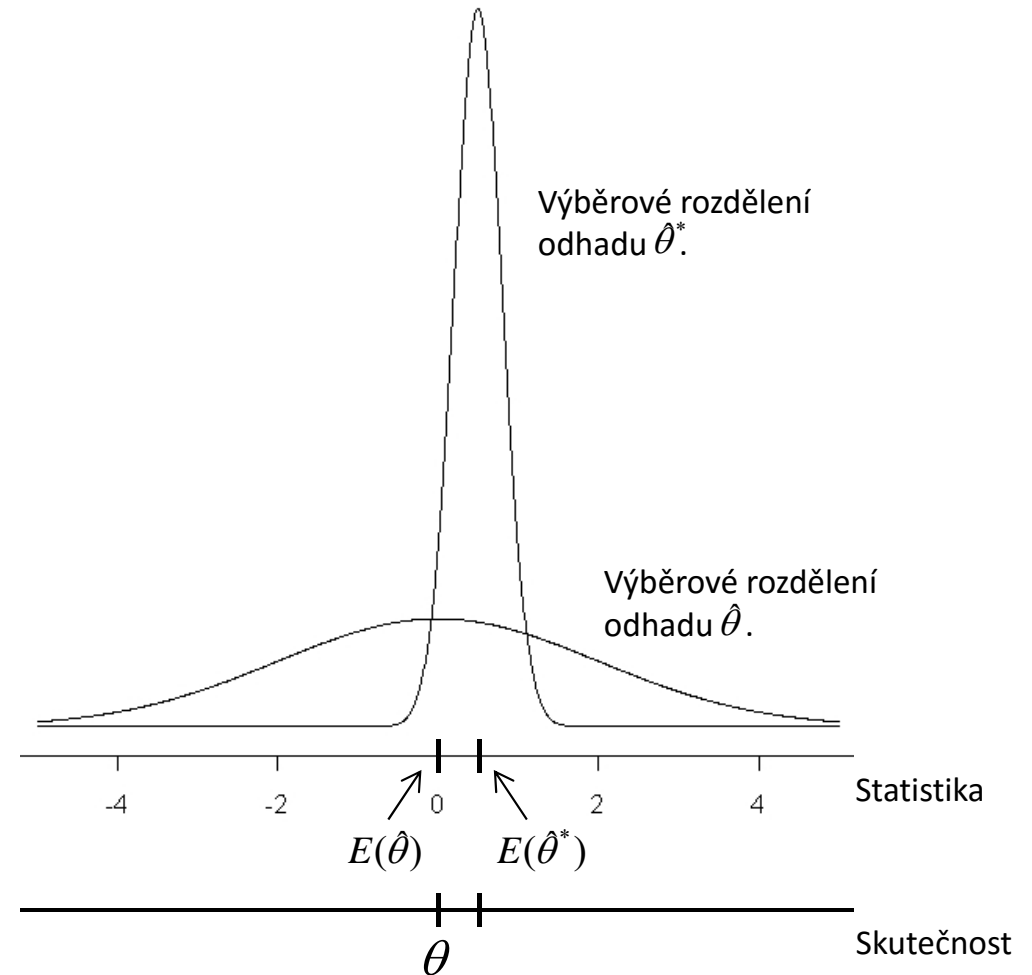


vychýlení² + variabilita

„bias²“ + „precision“

Příklad

- ➔ Máme dva odhady neznámého parametru θ .
- ➔ Jeden je vychýlený s malou variabilitou.
- ➔ Druhý je nevychýlený s větší variabilitou.
- ➔ Ne vždy musí být lepším odhadem ten, který je nevychýlený!



Nestrannost

- Celkem **logickým omezením odhadů**, které nás zajímají, je jejich **nestrannost**.
- Odhad $d(\mathbf{x})$ parametru θ je nestranný když

$$E_{\theta}(d(X)) = \theta \quad \text{pro každé } \theta \in \Theta$$

- Platí tedy:

$$E_{\theta}(d(X) - \theta) = 0 \quad \text{pro každé } \theta \in \Theta$$

- V množině nestranných odhadů se poté **snažíme najít odhad s nejmenší variabilitou** – abychom měli i minimální *MSE*.
- V úvodní přednášce jsme mluvili o zkreslení výsledků („biased results“) – nestrannost je ve své podstatě to samé.



Průměr – nestranný odhad?

→ Normální rozdělení pravděpodobnosti:

$$X_i \sim N(\mu, \sigma^2)$$

$$E(\bar{X}) = E\left(\frac{1}{n} \sum X_i\right) = \frac{1}{n} \sum EX_i = \mu \text{ pro každé } \mu \in R$$

→ Poissonovo rozdělení pravděpodobnosti:

$$X_i \sim Po(\lambda)$$

$$E(\bar{X}) = E\left(\frac{1}{n} \sum X_i\right) = \frac{1}{n} \sum EX_i = \lambda \text{ pro každé } \lambda \in R$$

→ Použití průměru pro tato rozdělení má smysl, ale je třeba si ověřit dané rozdělení pravděpodobnosti.

Nestranný odhad – příklad

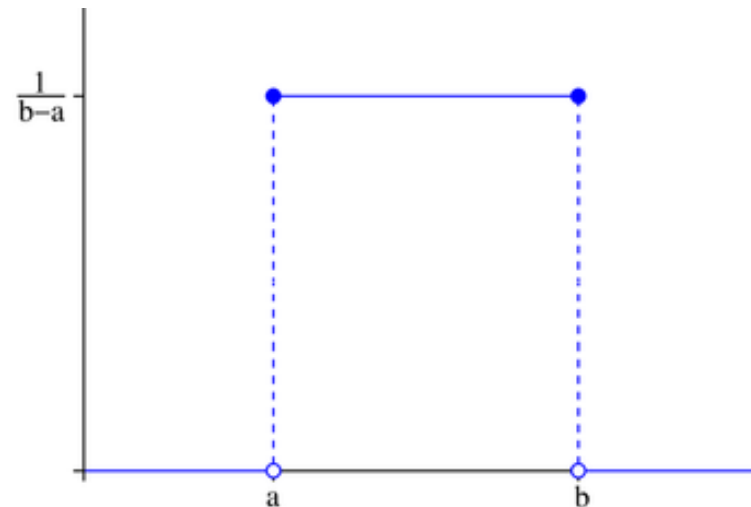
- ➔ Měříme čas, který trvá lékaři určitá činnost (např. ambulantní ošetření).
Chceme najít odhad maxima tohoto času, tedy jak maximálně dlouho mu daná činnost může trvat.
- ➔ Uvažujme rovnoměrně spojité rozdělení pravděpodobnosti na intervalu $[0, \theta]$:

$$X \sim R_s(0, \theta)$$

$$\rightarrow f(x) = 1/\theta \quad \text{pro každé } x \in (0, \theta)$$

$$\rightarrow f(x) = 0 \quad \text{pro každé } x \notin (0, \theta)$$

- ➔ Jak můžeme hodnotu θ odhadnout?



Nestranný odhad – příklad

→ Máme tedy náhodný výběr X_1, X_2, \dots, X_n i.i.d. z rozdělení $Rs[0, \theta]$, které ještě seřadíme podle velikosti: $X_{(1)}, X_{(2)}, \dots, X_{(n)}$.

$$E(X_i) = \theta \quad D(X_i) = \frac{1}{12} \theta^2$$

→ Máme dvě zajímavé hodnoty:

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

$$X_{(n)} = \max X_i$$

→ Uvažujeme dva odhady:

$$\left. \begin{aligned} T_1 &= 2\bar{X} = \frac{2}{n} \sum_{i=1}^n X_i \\ T_2 &= \frac{n+1}{n} X_{(n)} = \frac{n+1}{n} \max X_i \end{aligned} \right\} \text{Který je lepší?}$$

Nestranný odhad – příklad

→ Máme tedy X_1, X_2, \dots, X_n , které seřadíme podle velikosti: $X_{(1)}, X_{(2)}, \dots, X_{(n)}$.

→ Máme dvě zajímavé hodnoty:

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

$$E\bar{X} = \frac{1}{n} \sum_{i=1}^n EX_i = \theta / 2$$

$$D(\bar{X}) = \frac{1}{12n} \theta^2$$

$$X_{(n)} = \max X_i$$

$$EX_{(n)} = E(\max X_i) = \frac{n}{n+1} \theta$$

$$D(X_{(n)}) = \frac{n\theta^2}{(n+1)^2(n+2)}$$

→ Uvažujeme dva odhady:

$$T_1 = 2\bar{X} = \frac{2}{n} \sum_{i=1}^n X_i$$

$$ET_1 = E(2\bar{X}) = 2\left(\frac{\theta}{2}\right) = \theta$$

$$D(T_1) = \frac{1}{3n} \theta^2$$

$$T_2 = \frac{n+1}{n} X_{(n)} = \frac{n+1}{n} \max X_i$$

$$ET_2 = E\left(\frac{n+1}{n} X_{(n)}\right) = \frac{n+1}{n} \frac{n}{n+1} \theta = \theta$$

$$D(T_2) = \frac{\theta^2}{n(n+2)}$$

Který je lepší?

Nestranný odhad – příklad

→ Máme tedy X_1, X_2, \dots, X_n , které seřadíme podle velikosti: $X_{(1)}, X_{(2)}, \dots, X_{(n)}$.

→ Máme dvě zajímavé hodnoty:

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

$$E\bar{X} = \frac{1}{n} \sum_{i=1}^n EX_i = \theta / 2$$

$$D(\bar{X}) = \frac{1}{12n} \theta^2$$

$$X_{(n)} = \max X_i$$

$$EX_{(n)} = E(\max X_i) = \frac{n}{n+1} \theta$$

$$D(X_{(n)}) = \frac{n\theta^2}{(n+1)^2(n+2)}$$

→ Uvažujeme dva odhady:

$$T_1 = 2\bar{X} = \frac{2}{n} \sum_{i=1}^n X_i$$

$$ET_1 = E(2\bar{X}) = 2\left(\frac{\theta}{2}\right) = \theta$$

$$D(T_1) = \frac{1}{3n} \theta^2$$

$$T_2 = \frac{n+1}{n} X_{(n)} = \frac{n+1}{n} \max X_i$$

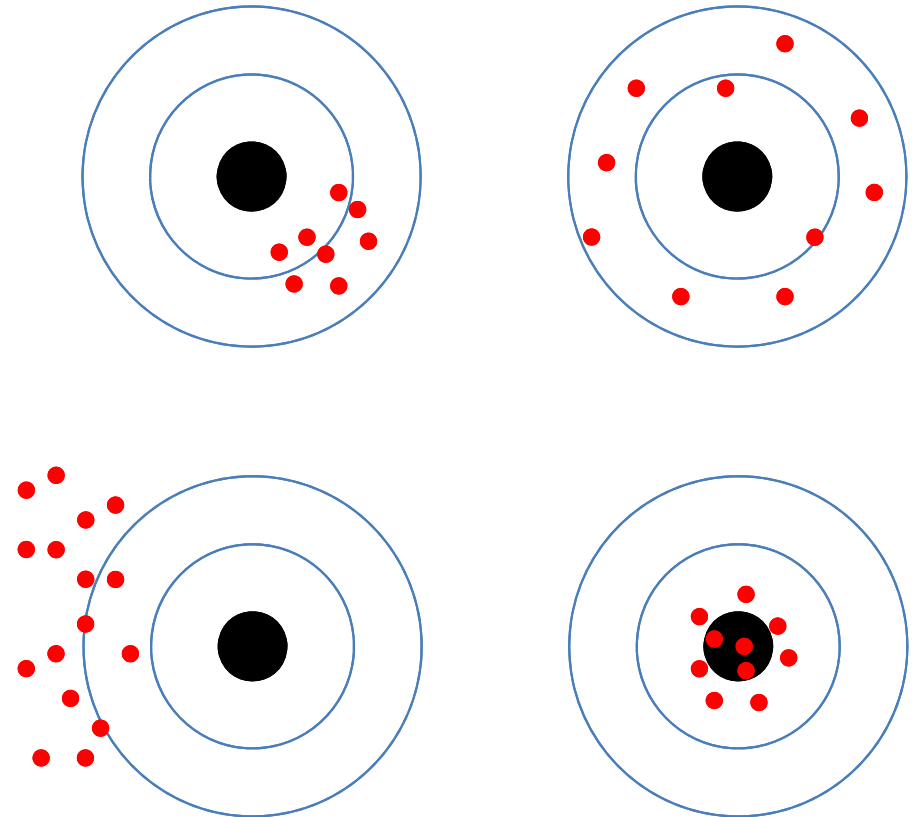
$$ET_2 = E\left(\frac{n+1}{n} X_{(n)}\right) = \frac{n+1}{n} \frac{n}{n+1} \theta = \theta$$

$$D(T_2) = \frac{\theta^2}{n(n+2)}$$

Vítězem se stal odhad T_2 , jeho variabilita s rostoucím n rychleji klesá k 0.

Vztah vychýlení a variability odhadu

- ➔ Odhady můžeme kombinací vychýlení a variability rozdělit (hypoteticky) do čtyř skupin.
- ➔ Význam není až tak v jednoduchých sumarizacích dat, ale spíš ve stochastickém modelování.



- Skutečná hodnota neznámého parametru
- Odhad neznámého parametru

Poznámka o stochastickém modelování

- ➔ Modely, které jsou příliš jednoduché (mají málo vysvětlujících proměnných) mohou být nepřesné kvůli velkému vychýlení, protože nejsou dostatečně flexibilní vzhledem k pozorovaným datům.
- ➔ Modely, které jsou příliš složité (mají mnoho vysvětlujících proměnných) mohou být nepřesné kvůli velké variabilitě, protože se příliš přizpůsobují pozorovaným datům (tzv. „overfitting“).
- ➔ Tomuto fenoménu se říká „**bias-variance tradeoff**“.
- ➔ Identifikovat správný model není jednoduché, je třeba najít správný počet vysvětlujících proměnných („model complexity“).

3. Metoda maximální věrohodnosti

Metoda maximální věrohodnosti

- ➔ Autorem je R. A. Fisher (1922). Anglicky „maximum likelihood estimation“.
- ➔ Máme n nezávislých stejně rozdělených pozorování (i.i.d.) z rozdělení s hustotou $f(x; \theta)$.
- ➔ **Sdružená hustota** odpovídající n pozorovaným hodnotám x_1, x_2, \dots, x_n je:

Jaká? A proč?

Metoda maximální věrohodnosti

- Autorem je R. A. Fisher (1922). Anglicky „maximum likelihood estimation“.
- Máme n nezávislých stejně rozdělených pozorování (i.i.d.) z rozdělení s hustotou $f(x; \theta)$.
- **Sdružená hustota** odpovídající n pozorovaným hodnotám x_1, x_2, \dots, x_n je:

$$f(x_1, \dots, x_n | \theta) = \prod_{i=1}^n f(x_i; \theta)$$

- Sdružená hustota vyjadřuje (za předpokladu, že známe θ), jak moc je pravděpodobné, že pozorované hodnoty pochází z rozdělení s hustotou $f(x; \theta)$
- **Pointa metody maximální věrohodnosti:** Dívat se na sdruženou hustotu jako na funkci θ a vybrat θ takové, aby výraz $f(x_1, \dots, x_n | \theta) = \prod_{i=1}^n f(x_i; \theta)$ byl co největší (maximum).

Věrohodnostní funkce

→ Zavádíme tzv. **věrohodnostní funkci** („likelihood function“):

$$L(\theta | x_1, \dots, x_n) = f(x_1, \dots, x_n | \theta)$$

→ Maximálně věrohodný odhad, značíme ho $\hat{\theta}_{MLE}$, je číslo, které maximalizuje věrohodnostní funkci, tedy

$$\hat{\theta}_{MLE} = \arg \max_{\theta \in \Theta} L(\theta | x_1, \dots, x_n)$$

→ Výpočetně se jedná o řešení rovnice (rovníc):

$$dL(\theta | x_1, \dots, x_n) / d\theta = 0$$

→ Musíme si ještě ověřit, že se jedná o maximum – např. pomocí druhých derivací.



Logaritmus věrohodnostní funkce

- ➔ Často je výhodnější (hlavně výpočetně jednodušší) maximalizovat logaritmus věrohodnostní funkce:

$$l(\theta | x_1, \dots, x_n) = \ln L(\theta | x_1, \dots, x_n) = \ln \prod_{i=1}^n f(x_i; \theta) = \sum_{i=1}^n \ln f(x_i; \theta)$$

- ➔ Bude maximum pro věrohodnostní funkci i logaritmus věrohodnostní funkce stejné? Pokud ano, tak proč?

ML odhad parametru λ Poissonova rozdělení

→ Máme n i.i.d. pozorování z Poissonova rozdělení: x_1, x_2, \dots, x_n .

→ Sdružená hustota má tvar:

$$f(x_1, \dots, x_n | \lambda) = \prod_{i=1}^n \frac{e^{-\lambda} \lambda^{x_i}}{x_i!}$$

→ Věrohodnostní funkce má tvar:

$$L(\lambda | x_1, \dots, x_n) = f(x_1, \dots, x_n | \lambda) = e^{-n\lambda} \lambda^{\sum_i x_i} / \prod_i x_i!$$

→ Logaritmus věrohodnostní funkce má tvar:

$$\ln L(\lambda | x_1, \dots, x_n) = \sum_i x_i \ln \lambda - n\lambda - \ln(\prod_i x_i!)$$

→ Jak vypadá $\hat{\theta}_{MLE}$?

ML odhad parametru λ Poissonova rozdělení

→ Derivace logaritmu věrohodnostní funkce má tvar:

$$\frac{d \ln L}{d\lambda} = \sum_i x_i / \lambda - n = 0$$

→ Výsledkem je průměr:

$$\hat{\lambda} = \frac{\sum_i x_i}{n}$$

→ Je to maximum?

$$\frac{d^2 \ln L}{d\lambda^2} = -\sum_i x_i / \lambda^2 < 0$$

ML odhad parametru μ normálního rozdělení

→ Máme n i.i.d. pozorování z normálního rozdělení: x_1, x_2, \dots, x_n .

→ Sdružená hustota má tvar:

$$f(x_1, \dots, x_n | \mu, \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_i - \mu)^2 / 2\sigma^2}$$

→ Logaritmus věrohodnostní funkce má tvar:

$$\ln L(\lambda | x_1, \dots, x_n) = -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2$$

→ Parciální derivace logaritmu věrohodnostní funkce mají tvar:

$$\partial \ln L / \partial \mu = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) = 0$$

$$\partial \ln L / \partial \sigma^2 = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2 = 0$$

ML odhad parametru μ normálního rozdělení

➔ Výsledkem jsou následující odhady:

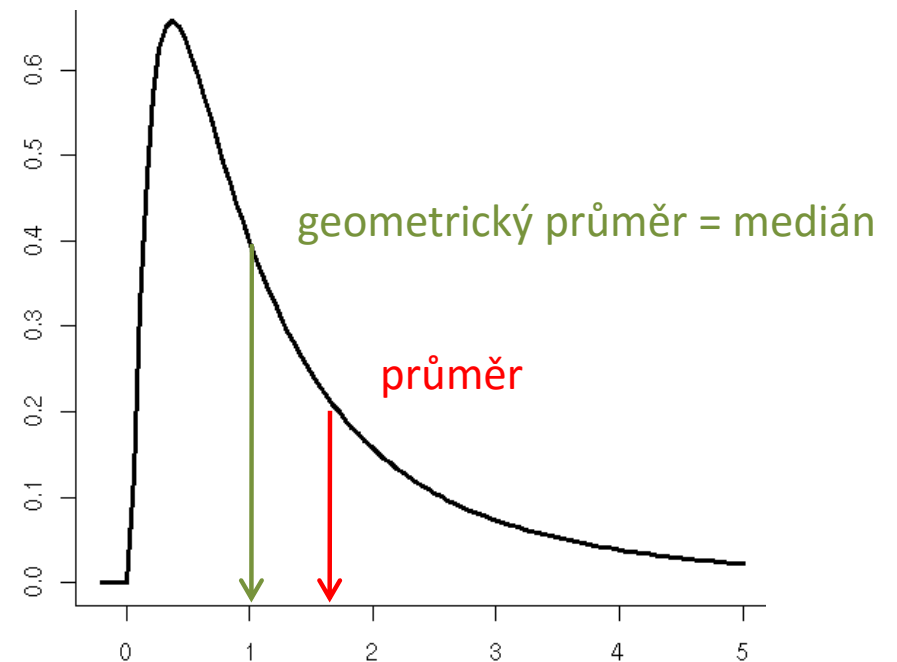
$$\hat{\mu}_{MLE} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}$$

$$\hat{\sigma}_{MLE}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

4. Srovnání průměru a mediánu

Nesmyslné použití průměru u asymetrických dat

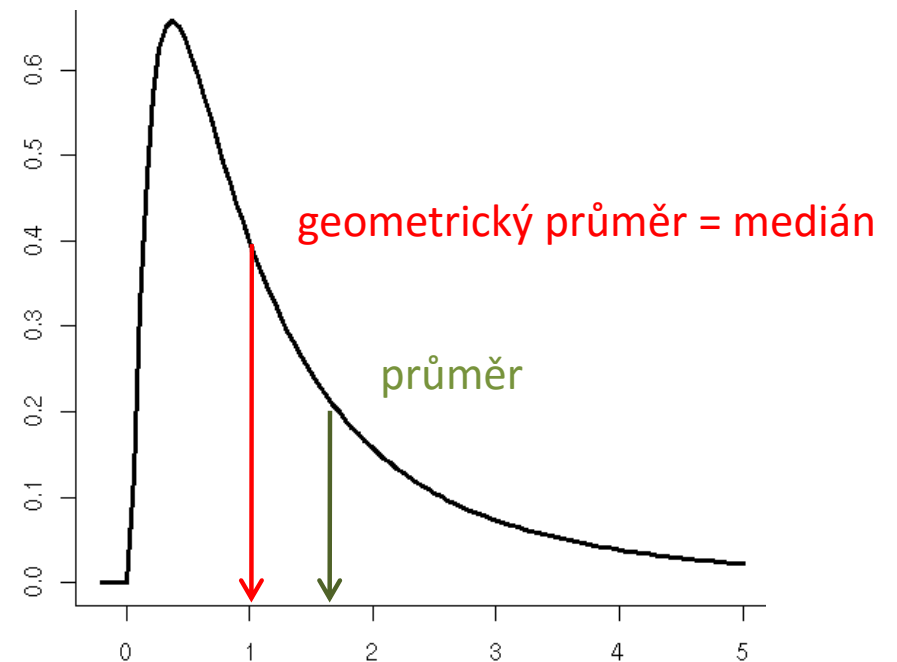
- ➔ Chceme-li charakterizovat log-normální rozdělení z hlediska střední hodnoty, je použití průměru nesmyslné. Není totiž splněn model, pro který byl jako optimální odhad odvozen!
- ➔ Vhodnějším odhadem je **medián** a **geometrický průměr** (jsou teoreticky ekvivalentní pro log-normální data)
- ➔ Geometrický průměr je průměr spočítaný na normálních datech, tedy po transformaci $y = \ln(x)$.
- ➔ **Příklad:** počty bílých krvinek.



Smysluplné použití průměru u asymetrických dat

➔ Chceme-li charakterizovat log-normální rozdělení z hlediska celkového součtu pozorovaných hodnot, je použití průměru smysluplné. Jedná-li se totiž např. o spotřebu nějakého materiálu, alkoholu nebo peněz, průměr popisuje z hlediska celkového součtu spotřebu lépe.

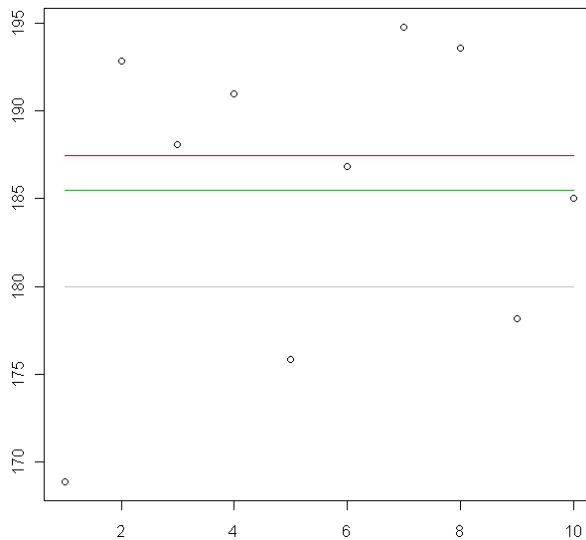
➔ **Příklad:** plánování celkové spotřeby nějakého materiálu, alkoholu nebo peněz do budoucna.



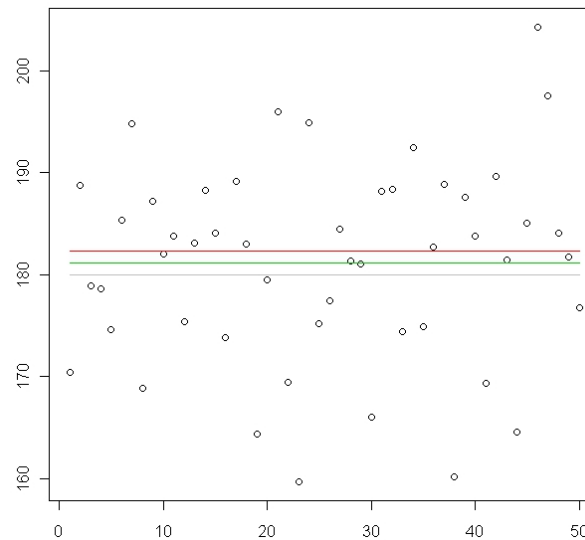
Smysluplné použití průměru u symetrických dat

- ➔ Pokud je splněn pravděpodobnostní model, tedy zejména normalita dat, je použití průměru na místě.
- ➔ **Průměr je konzistentní odhad** – pro $n \rightarrow \infty$ konverguje k θ podle pravděpodobnosti. Pro rostoucí n máme zaručeno, že se průměr přibližuje k θ .

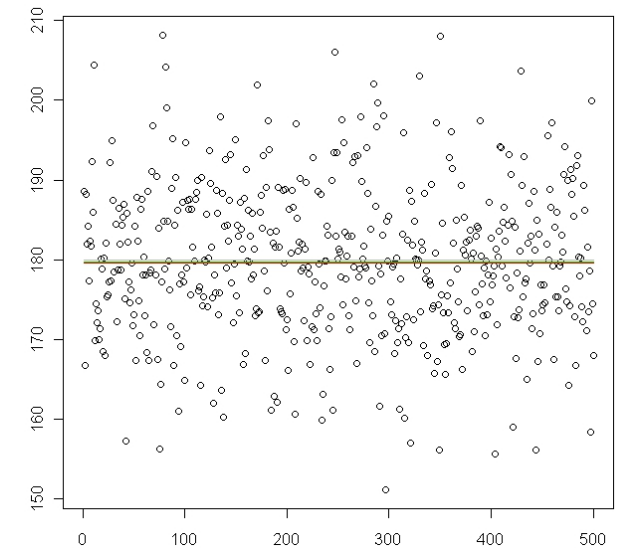
$n = 10$



$n = 50$



$n = 500$



skutečná hodnota

průměr

medián

Shrnutí – průměr vs. medián

	Výhody	Nevýhody
Průměr	Využívá informace celého souboru dat	Citlivý na odlehlá pozorování
	Jednoduché rozdělení pravděpodobnosti	Omezené použití u asymetrických dat
Medián	Není citlivý na odlehlá pozorování	Využívá informaci pouze jednoho pozorování
	Použití pro všechny typy dat	Komplikované rozdělení pravděpodobnosti

Shrnutí

- ➔ Používejte průměr!
 - ➔ Ale vždy si ověřte předpoklad normality (nebo alespoň symetrie), případně Poissonova rozdělení dat! A taky se nezapomeňte podívat na odlehlé hodnoty!
 - ➔ Pokud si něčím nejste jistí, použijte i medián.
-
- ➔ **Useknutý průměr** – odhad, který je svými vlastnostmi mezi průměrem a mediánem, spočítáme ho tak, že „odsekne“ m nebo $m\%$ minimálních a maximálních hodnot a ze zbytku spočítáme průměr.

Poděkování...

Rozvoj studijního oboru „Matematická biologie“ PŘF MU Brno je finančně podporován prostředky projektu ESF č. CZ.1.07/2.2.00/07.0318 „Víceoborová inovace studia Matematické biologie“ a státním rozpočtem České republiky



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ