Katedra pravděpodobnosti a matematické statistiky

MATEMATICKO-FYZIKÁLNÍ FAKULTA
Univerzita Karlova

Statistický seminář

Waldovy rovnosti

Anna Michálková

Motivace

- $X_1, X_2, \dots, X_n$ iid
  $\implies \mathbb{E}[\sum_{i=1}^n X_i] = n \cdot \mathbb{E}\, X_1$, var $[\sum_{i=1}^n X_i] = n \cdot \mathbb{E}\, X_1$,
- pokud počet sčítaných náhodných veličin je náhodný, bude platit něco podobného
- **sekvenční analýza** (připomenutí):
  - H_0 přijmeme, pokud $\prod_{i=1}^n \frac{f_1(X_i)}{f_0(X_i)} \leq B$,
    H_0 zamítneme, pokud $\prod_{i=1}^n \frac{f_1(X_i)}{f_0(X_i)} \geq A \implies$
  - počet pozorování je náhodná veličina:

$$N := \min\{n \in \mathbb{N} : \prod_{i=1}^n \frac{f_1(X_i)}{f_0(X_i)} \notin (B, A)\}$$

  - Wald: jaký je střední počet pozorování, která budeme potřebovat?

Terminologie z teorie pravděpodobnosti

Podmíněná střední hodnota

- předpokládáme: $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ reálná, $\mathcal{F} \in \mathcal{A}$,
- **podm. stř. hodnota** = zobrazení $\mathbb{E}[X|\mathcal{F}] \in L^1(\Omega, \mathcal{F}, \mathbb{P}\restriction_{\mathcal{F}})$ splňující:

$$\int_F X \mathrm{d}\,\mathbb{P} = \int_F \mathbb{E}[X|\mathcal{F}] \mathrm{d}\,\mathbb{P}\restriction_{\mathcal{F}}, \quad \forall F \in \mathcal{F}.$$

- vlastnosti: jednoznačnos sj., linearita, monotonie, „$X\mathcal{F}$-měřitelná $\implies \mathbb{E}[X|\mathcal{F}] \overset{\text{sj}}{=} X$“, vztah s nezávislostí...

Terminologie z teorie pravděpodobnosti (pokr.)

Markovský čas

- předpokládáme: **filtrace** $\{\mathcal{F}_n\}_{n=1}^{\infty}$ (tj. $\mathcal{F}_n \subseteq \mathcal{F}_{n+1} \subseteq \cdots \subseteq \mathcal{A}$),
- **markovský čas** = zobrazení $T : \Omega \to \mathbb{N} \cup \{\infty\}$ splňující $[T \leq n] \in \mathcal{F}_n, n \in \mathbb{N}$.

Martingal

- předpokládáme: filtrace $\{\mathcal{F}_n\}_{n=1}^{\infty}$,
- $\mathcal{F}_n$**-martingal** = náh. posloupnost $X_n \in L^1(\Omega, \mathcal{A}, \mathbb{P})$, která je $\mathcal{F}_n$**-adaptovaná** (tj. $\sigma(X_1, \ldots, X_n) \subseteq \mathcal{F}_n$) a $\mathbb{E}[X_{n+1}|\mathcal{F}_n] \stackrel{\text{sj}}{=} X_n, n \in \mathbb{N}$.

Náhodná procházka

$$S_n := \sum_{i=1}^n X_i, \text{ kde } X_1, \dots, X_n \text{ iid}$$

Vlastnosti:

- 4 možné scénáře:
  1. $\mathbb{P}(S_n = 0) = 1, \forall n \in \mathbb{N}$,
  2. $S_n \xrightarrow{\text{sj}} \infty, n \to \infty$,
  3. $S_n \xrightarrow{\text{sj}} -\infty, n \to \infty$,
  4. $\limsup_{n\to\infty} S_n = \infty$ a $\liminf_{n\to\infty} S_n = -\infty$.
- netriviální (tj. $\mathbb{P}(X_i \neq 0) > 0$) $\implies$ $T_B := \min\{n \in \mathbb{N} : S_n \notin B \text{ omez. borel.}\}$ je konečný sj.
- netriviální $\implies \forall r \in \mathbb{N} : \mathbb{E}\, T_B^r < \infty$.

Příklad (náh. procházka, markovský čas)

- SSRW: $X_i \sim R\{-1; 1\}$,
- $T_a := \min\{n \in \mathbb{N} : S_n = a\}, a \in \mathbb{Z}$.

Potom $T_a < \infty$ sj.

Náhodná procházka v sekvenční analýze

$$N := \min\{n \in \mathbb{N} : \prod_{i=1}^{n} \frac{f_1(X_i)}{f_0(X_i)} \notin (B, A)\}$$

- $Y_i := \log \frac{f_1(X_i)}{f_0(X_i)}$ jsou iid $\implies$ $S_n := \sum_{i=1}^{n} Y_i$ je náhodná procházka,
- $N := \min\{n \in \mathbb{N} : \sum_{i=1}^{n} Y_i \notin (\log B, \log A)\}$ je markovský čas (vzhledem k filtraci $\sigma(Y_1, \ldots, Y_n)$),
- $N < \infty$ sj., dokonce $\mathbb{E}\, N^r < \infty, r \in \mathbb{N}$.
- $\mathbb{E}\, N =???$ $\implies$ Waldovy rovnosti!

„Optional sampling“

- pro $\{X_n\}_{n=1}^{\infty}$ náh. posl. a T markovský čas:

$$X_T(\omega) := \begin{cases} X_{T(\omega)}(\omega), & \text{pokud } T(\omega) < \infty \\ 0, & \text{pokud } T(\omega) = \infty, \end{cases}$$

- $S \leq T < \infty$ sj. markovské časy, $\{X_n\}_{n=1}^{\infty}$ martingal $\implies$ $\mathbb{E}\, X_T = \mathbb{E}\, X_S$?
  Platí, pokud:
  - $X_T \in L^1$,
  - $\int_{[T>n]} |X_n| \mathrm{d}\,\mathbb{P} \to 0, n \to \infty$.

  Postačující podmínka:
  - $T \in L^1$,
  - $\exists c \in (0, \infty) : [T > n] \implies \mathbb{E}[|X_{n+1} - X_n| | \mathcal{F}_n] \overset{\text{sj}}{\leq} c$.

Waldovy rovnosti - základní verze

Věta

Předpokládáme: $\{X_n\}_{n=1}^{\infty}$ iid, $T \in L^1$ s hodnotami v $\mathbb{N}$ **nezávislá** na $\{X_n\}_{n=1}^{\infty}$.

(1) Pokud $X_1 \in L^1$, pak $\mathbb{E}\, S_T = \mathbb{E}\, X_1 \cdot \mathbb{E}\, T$.

(2) Pokud $\mathbb{E}\, X_1 = 0, X_1 \in L^2$, pak $\operatorname{var}\ S_T = \mathbb{E}\, X_1^2 \cdot \mathbb{E}\, T$.

(3) Pokud $\mathbb{E} \exp(\alpha X_1) < \infty$ pro nějaké α, pak $\mathbb{E}[\frac{\exp(\alpha S_T)}{(\mathbb{E} \exp(\alpha X_1))^T}] = 1$.

Waldovy rovnosti - obecná verze

Věta

Předpokládáme: $\{X_n\}_{n=1}^{\infty}$ iid, $T \in L^1$ **markovský čas** (vzhledem k $\sigma(X_1, \dots, X_n)$).

(1) **Waldova rovnost I**: Pokud $X_1 \in L^1$, pak $\mathbb{E}\, S_T = \mathbb{E}\, X_1 \cdot \mathbb{E}\, T$.

(2) **Waldova rovnost II**: Pokud $\mathbb{E}\, X_1 = 0, X_1 \in L^2$ a $\exists c \in (0, \infty) : [T > n] \implies |S_n| \overset{\text{sj}}{\leq} c$, pak $\operatorname{var} S_T = \mathbb{E}\, X_1^2 \cdot \mathbb{E}\, T$.

(3) **Waldova fundamentální identita**: Pokud $1 \leq \mathbb{E} \exp(\alpha X_1) < \infty$ pro nějaké α a $\exists c \in (0, \infty) : [T > n] \implies |S_n| \overset{\text{sj}}{\leq} c$, pak $\mathbb{E}[\frac{\exp(\alpha S_T)}{(\mathbb{E} \exp(\alpha X_1))^T}] = 1$.

Příklad (Waldovy rovnosti)

- $X_i \sim R\{-1;1\}$ iid,
- $T := \min\{n \in \mathbb{N} : S_n \notin (b,a)\}, b < 0 < a.$

Potom:

- $\mathbb{E}\, S_T = \mathbb{E}\, T \cdot \mathbb{E}\, X_1 = 0,$
- $\mathbb{E}\, T = \frac{\mathbb{E}\, S_T^2}{\mathbb{E}\, X_1^2},$
- $\mathbb{E}[\frac{\exp(\alpha S_T)}{(\mathbb{E}\exp(\alpha X_1))^T}] = 1.$

Využití v sekvenční analýze

$$N := \min\{n \in \mathbb{N} : \textstyle\sum_{i=1}^{n} \log \frac{f_1(X_i)}{f_0(X_i)} \notin (\log B, \log A)\}$$

- odhad $\mathbb{E}\, N$:
  - pokud $\mathbb{E}\, Y_1 \neq 0 \to \mathbb{E}\, N \approx \widehat{\frac{\mathbb{E}\, S_N}{\mathbb{E}\, Y_1}}$,
  - pokud $\mathbb{E}\, Y_1 = 0 \to \mathbb{E}\, N \approx \widehat{\frac{\mathbb{E}\, S_N^2}{\mathbb{E}\, Y_1^2}}$,
- aproximace chyb 1. a 2. druhu: výpočet $\mathbb{E} \exp(\alpha S_T) \to$ kdy je $\mathbb{E}\, e^{\alpha X_1} = 1$?

Příklad (Waldovy rovnosti) 2

- $X_i \sim N(\mu, \sigma^2)$ iid,
- $T \in L^1$ markovský čas $\{X_n\}_{n=1}^{\infty}$.

Potom:

$$\mathbb{E}\, e^{tX} = \exp\big(\frac{t^2\sigma^2}{2} + t\mu\big) = 1 \iff \big(t = 0 \text{ nebo } t = \frac{-2\mu}{\sigma^2}\big).$$

Vlastnosti charakteristických funkcí

Věta

Nechť $\varphi(t) = \mathbb{E}\, e^{tX} < \infty$ pro $t = \pm b$, $b > 0$. Potom:

(1) $\forall t \in [-b, b] : \varphi(t) < \infty$,

(2) φ je spojitá na $[-b, b]$ a $\mathcal{C}^{\infty}(-b, b)$,

(3) φ je konvexní na $[-b, b]$, $\varphi^{(r)}$ je konvexní na $(-b, b)$ $\forall r \in \mathbb{N}$,

(4) $\mathbb{E}\,|X|^r < \infty, r \in \mathbb{N}$ a $\mathbb{E}\, X^r = \varphi^{(r)}(0)$.

Vlastnosti charakteristických funkcí 2

Věta

Nechť $\varphi(t) = \mathbb{E}\, e^{tX} < \infty, t \in \mathbb{R}$; $\mathbb{P}(X > 0) > 0$ & $\mathbb{P}(X < 0) > 0$. Potom platí:

(1) pokud $\mathbb{E}\, X \neq 0$, pak $\exists! t_0 \neq 0 : \varphi(t_0) = 1$,

(2) pokud $\mathbb{E}\, X = 0$, pak $\varphi(t) = 1$ pouze pro $t = 0$.