Základy teorie pravděpodobnosti

Zdeněk Dvořák

14. listopadu 2018

Příklad 1. *Hrajete ruletu, sázíte vždy na červenou. Před každým kolem si zvolíte výši sázky h, pokud padne červené číslo, získáte h korun, jinak h korun prohrajete. Pravděpodobnost výhry je $18/37 \approx 48.6\%$.*

Máte následující strategii: v prvním kole vsadíte 1 korunu. V dalších kolech, pokud jste v předchozím kole prohráli, zdvojnásobíte sázku, jinak vsadíte 1 korunu. Nechť V označuje vaši výhru po 5 kolech.

Rozdělení V:

v	5	4	3	2	1	0	-1	-2	-5	-6	-14	-31
$P[V = v]$	2.72	11.51	21.09	21.93	13.00	6.42	6.41	3.38	3.20	3.38	3.38	3.57

Tedy s pravděpodobností 70.75% vyhrajete (ale málo), s pravděpodobností jen 23.32% prohrajete (ale hodně).

Střední hodnota je -0.28 kč: budeme-li tuto strategii opakovat n-krát pro dost velké n, prohrajeme zhruba $0.28n$ kč.

Pro nezáporné veličiny lze alespoň pomocí střední hodnoty omezit pravděpodobnost, že veličina je příliš velká.

Lemma 1 (Markovova nerovnost)**.** *Pro náhodnou veličinu $X \geq 0$ se střední hodnotou $E[X] = m > 0$ platí pro každé reálné číslo $t \geq 1$*

$$P[X \geq tm] \leq \frac{1}{t}.$$

Důkaz.

$$\begin{aligned} P[X \geq tm] &= \sum_{z\in\Omega, X(z)\geq tm} P[z] \leq \sum_{z\in\Omega, X(z)\geq tm} P[z] \cdot \frac{X[z]}{tm} \\ &\leq \sum_{z\in\Omega} P[z] \cdot \frac{X[z]}{tm} = \frac{1}{tm}\mathrm{E}[X] = \frac{1}{t}. \end{aligned}$$

□

Příklad 2. *Mějme n bodů v rovině (v obecné poloze) a mezi každými dvěma z nich nakresleme úsečku nezávisle s pravděpodobností $10/n$. Střední hodnota počtu T trojúhelníků (tvořených třemi ze zadaných bodů) je*

$$E[T] = \binom{n}{3} \cdot \frac{1000}{n^3} \leq 1000/6.$$

Pravděpodobnost, že vznikne alespoň 1000 trojúhelníků, je tedy

$$P[T \geq 1000] \leq P[T \geq 6 \cdot E[T]] \leq 1/6.$$

Dolní odhad se pouze ze střední hodnoty zjistit nedá (a ve skutečnosti je tato pravděpodobnost mnohem nižší, než 1/6.

1 Rozptyl a Čebyševova nerovnost

Rozptyl náhodné veličiny X je definován jako

$$\mathrm{Var}[X] = \mathrm{E}[(X - \mathrm{E}[X])^2].$$

Příklad 3. *Hrajeme ruletu, vsadíme h na červenou (pravděpodobnost výhry $p = 18/37$). Výše výhry V_h má střední hodnotu*

$$E[V_h] = ph + (1-p)(-h) = (2p-1)h \approx -0.03h$$

a rozptyl

$$\begin{aligned} Var[V_h] &= E[(V_h - (2p-1)h)^2] = p(h-(2p-1)h)^2 + (1-p)(-h-(2p-1)h)^2 \\ &= \big(p(2-2p)^2 + (1-p)(-2p)^2\big)h^2 = 4p(1-p)h^2 \approx h^2. \end{aligned}$$

Pozorování 2. *Jestliže $|X| \leq s$, pak $Var[X] \leq s^2$.*

Věta 3 (Čebyševova nerovnost)**.** *Pro náhodnou veličinu $X \geq 0$ se střední hodnotou $E[X] = m$ platí pro každé reálné číslo $t \geq 1$*

$$P[|X - m| \geq t\sqrt{Var[X]}] \leq \frac{1}{t^2}.$$

Důkaz. Uvažujme náhodnou veličinu $Y = (X-m)^2$; ta je zjevně nezáporná a její střední hodnota je rovna $s = \mathrm{Var}[X]$. Z Markovovy nerovnosti máme

$$P[|X - m| \geq t\sqrt{\mathrm{Var}[X]}] = P[Y \geq t^2 s] \leq \frac{1}{t^2}.$$

□

Náhodné veličiny X a Y jsou nezávislé, jestliže pro všechna reálná čísla v_x a v_y jsou jevy $X = v_x$ a $Y = v_y$ nezávislé. Například: výše výhry v různých kolech rulety (jestliže sázky zvolíme předem).

Lemma 4. *Jestliže jsou jevy X_1 a X_2 nezávislé, pak*

$$E[X_1X_2] = E[X_1]E[X_2].$$

Jestliže jsou jevy $X_1, \ldots, X_n$ po dvou nezávislé, pak

$$Var\Big[\sum_{i=1}^{n} X_i\Big] = \sum_{i=1}^{n} Var[X_i].$$

Důkaz. Máme

$$\begin{aligned}
\mathrm{E}[X_1]\mathrm{E}[X_2] &= \Big(\sum_{v_1} P[X_1 = v_1]v_1\Big) \cdot \Big(\sum_{v_2} P[X_2 = v_2]v_1\Big) \\
&= \sum_{v_1,v_2} P[X_1 = v_1]P[X_2 = v_2]v_1v_2 = \sum_{v_1,v_2} P[X_1 = v_1, X_2 = v_2]v_1v_2 \\
&= \sum_{v} P[X_1X_2 = v]v = \mathrm{E}[X_1X_2].
\end{aligned}$$

Nechť $m_i = \mathrm{E}[X_i]$ a $m = \sum_{i=1}^{m} m_i = \mathrm{E}\big(\sum_{i=1}^{m} X_i\big)$. Označme $Y_i = X_i - m_i$, takže $\mathrm{E}[Y_i] = 0$, $\mathrm{Var}[X_i] = E[Y_i^2]$, a veličiny $Y_1, \ldots, Y_n$ jsou po dvou nezávislé. Máme

$$\begin{aligned}
\mathrm{Var}\Big[\sum_{i=1}^{n} X_i\Big] &= \mathrm{E}\Big[\Big(\Big(\sum_{i=1}^{n} X_i\Big) - m\Big)^2\Big] \\
&= \mathrm{E}\Big[\Big(\sum_{i=1}^{n}(X_i - m_i)\Big)^2\Big] = \mathrm{E}\Big[\Big(\sum_{i=1}^{n} Y_i\Big)^2\Big] \\
&= \mathrm{E}\Big[\sum_{i=1}^{n} Y_i^2 + 2\sum_{i<j} Y_iY_j\Big] = \sum_{i=1}^{n} \mathrm{E}[Y_i^2] + 2\sum_{i<j} \mathrm{E}[Y_i]\mathrm{E}[Y_j] \\
&= \sum_{i=1}^{n} \mathrm{E}[Y_i^2] = \sum_{i=1}^{n} \mathrm{Var}[X_i].
\end{aligned}$$

□

Důsledek 5 (Centrální limitní věta, slabá verze)**.** *Nechť $X_1, \ldots, X_n$ jsou nezávislé jevy a $Var[X_i] \leq s^2$ pro $i = 1, \ldots, n$. Nechť $X = \sum_{i=1}^{n} X_n$ a $h = E[X]$. Pak pro každé $t \geq 1$*

$$P(|X - h| \geq ts\sqrt{n}) \leq \frac{1}{t^2}.$$

Důkaz. Dle Lemma 4 máme $\mathrm{Var}[X] = \sum_{i=1}^{n} \mathrm{Var}[X_i] \leq ns^2$, výsledek plyne z Čebyševovy věty. □

Příklad 4. *Hodíme n-krát spravedlivou mincí. Jev I_{rub} má střední hodnotu 1/2 a rozptyl nejvýše 1 dle Pozorování 2. Pravděpodobnost, že rub padne více než $(n/2 + 10\sqrt{n})$-krát nebo méně než $(n/2 - 10\sqrt{n})$-krát je tedy nejvýše 1%.*

Příklad 5. *V kasínu je strop na sázky $s = 32$ kč a minimální sázka 1 kč. Hráč postupně sází částky $1 \leq h_1, \ldots, h_n \leq s$ na červenou (sázky jsou zvolené předem, nezávisí na výhrách v jednotlivých kolech). Označme $h = \sum_{i=1}^{n} h_i$. Nechť V_i je výhra v i-tém kole a $V = \sum_{i=1}^{n} V_i$ je celková výhra. V minulém příkladu jsme určili $E[V_i] = (2p-1)h_i$, a proto $E[V] = (2p-1)h$. Dále $Var[V_i] \leq h_i^2 \leq s^2$ dle Pozorování 2. Výhry v jednotlivých kolech jsou nezávislé, proto pravděpodobnost, že $V \geq 0$ (tedy že kasíno na hráči nevydělá) je*

$$\begin{aligned}
P[V \geq 0] &= P[V - E[V] \geq |E[V]|] \leq P[|V - E[V]| \geq |E[V]|] \\
&= P\left[|V - E[V]| \geq \frac{|E[V]|}{s\sqrt{n}} \cdot s\sqrt{n}\right] \leq \frac{s^2 n}{(E[V])^2} \\
&= \frac{s^2 n}{(1-2p)^2 h^2} \leq \frac{s^2}{(1-2p)^2 n} = \frac{1401856}{n}.
\end{aligned}$$

Tedy, budete-li hrát dost dlouho, skoro jistě prohrajete (výrazně dříve, než naznačuje tento hrubý odhad!).