Dodatek E

Funkcja tworząca momenty (transformata Laplace'a)

§ E.1. Definicja i przykłady

Zamieszczamy tu podstawowe informacje o funkcjach tworzących momenty, które stosuje się w wielu zagadnieniach praktycznych, szczególnie tam, gdzie występują nieujemne zmienne losowe. Przykłady takich zastosowań, m.in. do teorii odnowienia, teorii ryzyka, zagadnienia ruiny w złożonym procesie Poissona można znaleźć w [FELL], t. II, roz. XIII–XIV. My pokażemy dwa zastosowania teoretyczne: do dowodu centralnego twierdzenia granicznego (tw. 12) i do oszacowania szybkości zbieżności w mocnym prawie wielkich liczb (§ E.2).

Jeśli X jest zmienną losową, a μ_X jej rozkładem prawdopodobieństwa, to definiujemy *funkcję tworzącą momenty* zmiennej losowej X (rozkładu μ_X) jako

$$M_X(t) = \mathcal{E}\mathrm{e}^{tX} = \int_{-\infty}^{\infty} \mathrm{e}^{ts}\mu_X(ds). \qquad (1)$$

Funkcja tworząca momenty jest określona dla tych t, dla których istnieje powyższa wartość oczekiwana (czyli całka po prawej stronie).

Jeśli X przyjmuje wartości całkowite nieujemne, funkcja M_X jest związana z opisaną w dodatku B funkcją tworzącą g_X w następujący sposób:

$$g_X(s) = \mathcal{E}s^X = \mathcal{E}\left[(\mathrm{e}^{\log s})^X\right] = M_X(\log s), \qquad s \in (0,1].$$

Jak kto woli, $M_X(t) = g_X(\mathrm{e}^t)$.

Funkcja tworząca momenty jest określona na przedziale zawierającym zero, który poza tym może być dowolny, jak pokazują przykłady 2 i 5.

Jeśli przedział ten zawiera otoczenie zera o dodatniej długości, to funkcja M_X wyznacza momenty zmiennej losowej X. Dokładniej, mamy

Twierdzenie 1. *Jeśli funkcja tworząca momenty $M_X(t)$ zmiennej losowej X jest określona dla $t \in (-t_0, t_0)$, $t_0 > 0$, to*

(i) *Istnieją wszystkie momenty zmiennej losowej* X, *czyli* $\mathcal{E}|X|^k < \infty$ *dla* $k = 1, 2, \ldots$.

(ii) $M_X(t) = \sum_{k=0}^{\infty} \frac{t^k \mathcal{E} X^k}{k!}, \quad |t| < t_0.$

(iii) $M_X^{(k)}(0) = \mathcal{E} X^k, \quad k = 1, 2, \ldots$.

D o w ó d. (i) Z nierówności $\mathrm{e}^{|t|} \leqslant \mathrm{e}^t + \mathrm{e}^{-t}$ wynika, że $\mathcal{E}\mathrm{e}^{t|X|} < \infty$ dla $t < t_0$. Oczywiste oszacowanie:

$$\sum_{k=0}^{n} \frac{t^k |X|^k}{k!} \leqslant \mathrm{e}^{t|X|}, \quad 0 < t < t_0, \tag{2}$$

gwarantuje teraz istnienie wszystkich momentów.

(ii) Dzięki oszacowaniu (2) można zastosować twierdzenie Lebesgue'a o zbieżności monotonicznej:

$$\mathcal{E}\left(\sum_{k=0}^{\infty} \frac{t^k |X|^k}{k!}\right) = \sum_{k=0}^{\infty} \frac{t^k \mathcal{E}|X|^k}{k!} = \mathcal{E}\mathrm{e}^{t|X|} < \infty, \quad 0 < t < t_0. \tag{3}$$

W efekcie twierdzenie Lebesgue'a o zbieżności zmajoryzowanej pozwala na usprawiedliwienie poniższych przejść granicznych:

$$M_X(t) = \mathcal{E}\left(\sum_{k=0}^{\infty} \frac{t^k X^k}{k!}\right) = \sum_{k=0}^{\infty} \frac{t^k \mathcal{E} X^k}{k!}, \quad |t| < t_0. \tag{4}$$

(iii) Wystarczy obliczyć pochodne szeregu potęgowego po prawej stronie (4). ■

Przykład 2. Zmienna losowa o gęstości $C\mathrm{e}^{-|x|^{1/2}}$ ma wszystkie momenty, ale jej funkcja tworząca momenty jest określona tylko w zerze. ■

Uwaga 3. Czasami w literaturze można się spotkać z nieco innym nazewnictwem, mianowicie funkcję tworzącą momenty nazywa się (dwustronną) transformatą Laplace'a. My nazwaliśmy transformatą Laplace'a (por. zad. 13.4.2) funkcję

$$L(\lambda) = \mathcal{E}\mathrm{e}^{-\lambda X}, \text{ gdzie } X \geqslant 0, \lambda \geqslant 0. \ ■$$

Uwaga 4. Twierdzenie Kaca (zad. 9.4.1), zawierające kryterium niezależności ograniczonych zmiennych losowych, da się nieco wzmocnić, a mianowicie: jeśli zmienne losowe X i Y mają funkcje tworzące momenty określone w pewnym otoczeniu zera, oraz $\mathcal{E}X^kY^l = \mathcal{E}X^k\mathcal{E}Y^l$ dla $k, l \in \mathbf{N}$, to X i Y są niezależne. Dowód wymaga oczywistej adaptacji argumentu podanego w rozwiązaniu zadania. ■

Przykład 5. Oto funkcje tworzące momenty dla kilku rozkładów.
1. Rozkład $\mathcal{N}(0,1)$:

$$M_X(t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{tx - x^2/2} = e^{t^2/2}, \quad t \in (-\infty, \infty),$$

a stąd otrzymujemy hurtem parzyste momenty, bowiem

$$e^{t^2/2} = \sum_{k=0}^{\infty} \frac{1}{k!} \left(\frac{t^2}{2}\right)^k = \sum_{k=0}^{\infty} \frac{1 \cdot 3 \cdot \ldots \cdot (2k-1)}{(2k)!} t^{2k},$$

wobec tego $\mathcal{E}X^{2k} = 1 \cdot 3 \cdot \ldots \cdot (2k-1) = (2k-1)!!$.
2. Rozkład wykładniczy:

$$M_X(t) = \int_0^{\infty} e^{tx} \lambda e^{-\lambda x}\, dx = \frac{\lambda}{\lambda - t} = \sum_{k=0}^{\infty} \frac{t^k}{\lambda^k}, \quad t < \lambda.$$

zatem $\mathcal{E}X^k = \frac{k!}{\lambda^k}$.
3. Rozkład Poissona:

$$M_X(t) = \sum_{k=0}^{\infty} \frac{e^{kt}\lambda^k}{k!} e^{-\lambda} = \sum_{k=0}^{\infty} \frac{(\lambda e^t)^k}{k!} e^{-\lambda} = e^{\lambda(e^t - 1)}, \quad t \in (-\infty, \infty).$$

Stąd już nieco trudniej wyznaczać momenty. Wygodniej jest obliczyć bezpośrednio $\mathcal{E}X(X-1) \cdot \ldots \cdot (X-k+1)$.

Za pomocą gęstości postaci $c\mathbf{1}_{[1,\infty)}x^{-2}$, $c\mathbf{1}_{[1,\infty)}x^{-2}e^{-ax}$ oraz wykładniczych i normalnych można skonstruować przykłady funkcji tworzących określonych w dowolnym przedziale (lewo-/prawostronnie domkniętym/otwartym) zawierającym zero. ■

Przy wykonywaniu działań na zmiennych losowych funkcje tworzące momenty zachowują się analogicznie do funkcji charakterystycznych. Zachodzi równość

$$M_{aX+b}(t) = e^{bt} M_X(at), \quad a, b \in \mathbf{R}.$$

Prawdziwe jest także podstawowe

Twierdzenie 6 (O mnożeniu). *Jeśli zmienne losowe X i Y są niezależne, a ich funkcje tworzące momenty M_X i M_Y są określone w pewnym otoczeniu zera $(-t_0, t_0)$, $t_0 > 0$, to w tym samym otoczeniu zera istnieje funkcja tworząca momenty sumy, M_{X+Y}, oraz*

$$M_{X+Y}(t) = M_X(t) M_Y(t), \quad t \in (-t_0, t_0), \quad t_0 > 0.$$

D o w ó d. Jeśli $t \in (-t_0, t_0)$, to zmienna losowa $\mathrm{e}^{t(X+Y)} = \mathrm{e}^{tX}\mathrm{e}^{tY}$ jest całkowalna jako iloczyn niezależnych i całkowalnych zmiennych losowych (uwaga c po tw. 5.8.15); z samego tw. 5.8.15 wynika, że

$$M_{X+Y}(t) = \mathcal{E}\mathrm{e}^{t(X+Y)} = \mathcal{E}\mathrm{e}^{tX}\mathcal{E}\mathrm{e}^{tY} = M_X(t)M_Y(t). \blacksquare$$

Twierdzenie 7. *Jeśli funkcje tworzące momenty zmiennych losowych X i Y są równe na przedziale $(-t_0, t_0)$, $t_0 > 0$, to X i Y mają ten sam rozkład.*

D o w ó d. Na mocy tw. 1 obie zmienne losowe mają te same momenty. Wykażemy, że ich funkcje charakterystyczne są równe.

Z tw. A.1.1 wynika, że

$$\left| \mathrm{e}^{itx} \left(\mathrm{e}^{ihx} - \sum_{k=0}^{n} \frac{(ihx)^k}{k!} \right) \right| \leqslant \frac{|hx|^{n+1}}{(n+1)!}, \quad t, h \in \mathbf{R}.$$

Stąd

$$\left| \varphi_X(t+h) - \sum_{k=0}^{n} \frac{(ih)^k}{k!} \mathcal{E}\left(X^k \mathrm{e}^{itX}\right) \right| \leqslant \frac{|h|^{n+1}}{(n+1)!} \mathcal{E}|X|^{n+1}, \quad t, h \in \mathbf{R},$$

gdzie prawa strona zmierza do zera zgodnie z (3). Ze wzoru na pochodne funkcji charakterystycznej (tw. 9.1.7) otrzymujemy:

$$\varphi_X(t+h) = \sum_{k=0}^{\infty} \frac{(ih)^k}{k!} \varphi_X^{(k)}(t), \quad |h| < t_0.$$

Tę samą równość spełnia φ_Y.

Niech teraz $t = 0$. Wtedy $\varphi_X^{(k)}(0) = i^k \mathcal{E}X^k = i^k \mathcal{E}Y^k = \varphi_Y^{(k)}(0)$ dla $k = 0, 1, 2\ldots$, zatem obie funkcje charakterystyczne pokrywają się na przedziale $(-t_0, t_0)$.

Powtarzając to rozumowanie z punktami startowymi $t = \pm t_0/2$ otrzymujemy równość $\varphi_X = \varphi_Y$ na przedziale $(-\frac{3}{2}t_0, \frac{3}{2}t_0)$, etc. — a więc wszędzie. Stąd X i Y muszą mieć ten sam rozkład. $\blacksquare$

Uwaga 8. W powyższym dowodzie nie odwoływaliśmy się do teorii funkcji analitycznych. Można jednak po prostu zauważyć, że funkcje

$$M_X(z) = \mathcal{E}\mathrm{e}^{zX}, \quad M_Y(z) = \mathcal{E}\mathrm{e}^{zY},$$

są holomorficzne w pasie $-t_0 < \operatorname{Re} z < t_0$ i są równe na odcinku $(-t_0, t_0)$, czyli na mocy tw. D.3.2 — wszędzie, co daje równość funkcji charakterystycznych, bowiem $M_X(it) = \varphi_X(t)$, $t \in \mathbf{R}$.

Uwaga 9. Jeśli rozkład prawdopodobieństwa μ ma wszystkie momenty, a jego funkcja tworząca momenty istnieje w pewnym otoczeniu zera o dodatniej długości, to nie istnieje inny rozkład prawdopodobieństwa o tych samych momentach (wynika to z (3) i powyższego twierdzenia).

Mówi się wtedy, że rozkład jest wyznaczony przez swoje momenty. Podamy teraz przykład sytuacji przeciwnej — dwóch różnych rozkładów o tych samych momentach.

Oznaczmy przez f gęstość rozkładu logarytmicznie-normalnego:

$$f(x) = \frac{1}{\sqrt{2\pi}} \frac{1}{x} \mathrm{e}^{-(\log x)^2/2} \mathbf{1}_{(0,\infty)}(x).$$

Wtedy

$$\int_0^\infty x^k f(x) \sin(2\pi \log x)\, dx = 0, \quad k = 0, 1, 2, \ldots,$$

gdyż podstawienie $\log x = s + k$ przekształca tę całkę do postaci

$$\frac{1}{\sqrt{2\pi}} \mathrm{e}^{k^2/2} \int_{-\infty}^{\infty} \mathrm{e}^{-s^2/2} \sin 2\pi s\, ds,$$

która jest równa zeru, jako że funkcja podcałkowa jest nieparzysta. W takim razie funkcje $f(x)$ oraz $f(x)[1+\sin(2\pi \log x)]$ są parą różnych gęstości o tych samych momentach.

Twierdzenie 10. *Niech $X, X_1, X_2, \ldots$ będzie ciągiem zmiennych losowych, których funkcje tworzące momenty, M_X i M_{X_k}, $k = 1, 2, \ldots$ są określone na przedziale $(-t_0, t_0)$, $t_0 > 0$.*

Jeśli $M_{X_n}(t) \xrightarrow[n\to\infty]{} M_X(t)$ dla $t \in (-t_0, t_0)$, to $X_n \xrightarrow{D} X$.

D o w ó d. Jędrność ciągu rozkładów zmiennych losowych X_n można otrzymać z wykładniczej nierówności Czebyszewa 5.7.6(c) i ze zbieżności ciągu funkcji tworzących momenty:

$$P(|X_n| > a) \leqslant \mathcal{E}\left(\mathrm{e}^{t|X_n|}\right) \mathrm{e}^{-at} \xrightarrow[n\to\infty]{} M_X(t)\mathrm{e}^{-at}, \quad |t| < t_0.$$

Dalsze rozumowanie przebiega tak, jak w dowodzie twierdzenia Lévy'ego-Craméra o ciągłości — po wyborze podciągu zbieżnego według rozkładu dowodzimy, korzystając z poprzedniego twierdzenia o jednoznaczności, że rozkładem granicznym jest μ_X i że cały ciąg jest słabo zbieżny do tego rozkładu prawdopodobieństwa. ■

Uwaga 11. Przy założeniach tw. 10 nietrudno otrzymać zbieżność momentów:

$$\mathcal{E}X_n^k \xrightarrow[n\to\infty]{} \mathcal{E}X^k, \quad k = 0, 1, 2, \ldots.$$

Istotnie, wynika to z kryterium z zad. 8.3.9, bowiem dla każdego $k = 1, 2, \ldots$:

$$\sup_n \mathcal{E}|X_n|^k \leqslant \frac{k!}{a^k} \sup_n \mathcal{E}\mathrm{e}^{a|X_n|} \leqslant \frac{k!}{a^k} \sup_n [M_{X_n}(a) + M_{X_n}(-a)], \quad |a| < t_0,$$

a wyrażenie w nawiasie kwadratowym po prawej stronie jest ograniczone, ponieważ zawiera wyrazy ciągu zbieżnego.

Odwrotnie, w zad. 8.3.8 udowodniliśmy, że jeśli rozkład X jest określony przez swoje momenty i zachodzi zbieżność momentów, to $X_n \xrightarrow{D} X$. Dlatego też mówi się o dowodzie CTG metodą momentów.

Udowodnimy teraz najprostszą wersję centralnego twierdzenia granicznego:

Twierdzenie 12. *Niech $X, X_1, X_2, \ldots$ będą niezależnymi zmiennymi losowymi o tym samym rozkładzie, $\mathcal{E}X = 0$, $\mathcal{D}^2X = 1$, i niech $M_X(t)$ będzie określona na przedziale $I = (-t_0, t_0)$ o dodatniej długości. Wtedy*

$$\frac{X_1 + \ldots + X_n}{\sqrt{n}} \xrightarrow{D} \mathcal{N}(0,1).$$

D o w ó d. Wystarczy udowodnić zbieżność funkcji tworzących momenty do funkcji $\mathrm{e}^{t^2/2}$ co najmniej na przedziale I. Ze wzoru (4) otrzymujemy

$$M_X(t) = 1 + \frac{t^2}{2} + o(t^2).$$

Z twierdzenia o mnożeniu wynika, że funkcja tworząca momenty po lewej stronie jest określona co najmniej na przedziale $(-\sqrt{n}t_0, \sqrt{n}t_0)$, oraz:

$$M_{\frac{X_1+\ldots+X_n}{\sqrt{n}}}(t) = \left[1 + \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right]^n \xrightarrow[n\to\infty]{} \mathrm{e}^{t^2/2}.$$

Zbieżność ma miejsce faktycznie dla każdego $t \in \mathbf{R}$. ■

Przypomina to co prawda wyścigi w workach.

Powyższy dowód stosuje się w szczególności do ograniczonych zmiennych losowych. Można otrzymać z niego dowód twierdzenia Lindeberga–Lévy'ego 10.2.1 metodą ucinania (której zastosowanie widzieliśmy w dowodzie mocnego prawa wielkich liczb).

§ E.2. Transformata Craméra. Oszacowanie szybkości zbieżności w mocnym prawie wielkich liczb

Oszacowanie odchyleń od średniej w nierówności Bernsteina (7.4.2) zawdzięczamy temu, że funkcja tworząca momenty liczby sukcesów S_n w schemacie Bernoulliego istnieje w przedziale o dodatniej długości. Rozwiniemy tę ideę i otrzymamy nierówności maksymalne, dające oszacowanie szybkości zbieżności w prawie wielkich liczb.

Załóżmy, że zmienna losowa X spełnia *warunek Craméra*: istnieje takie $\lambda > 0$, że $\mathcal{E}\mathrm{e}^{\lambda|X|} < \infty$. Wtedy oczywiście jej funkcja tworząca momenty, M_X, jest określona co najmniej na przedziale $[-\lambda, \lambda]$.

Warunek Craméra jest równoważny z tym, że ogon rozkładu zmiennej losowej X maleje wykładniczo: $P(|X| > t) \leqslant \mathrm{e}^{-at}$ dla pewnego $a > 0$ i wszystkich $t > 0$. Jest on spełniony przez wszystkie ograniczone zmienne losowe, a także przez zmienne losowe o rozkładach: geometrycznym, wykładniczym, Poissona i normalnym.

Niech teraz
$$\Lambda = \{t \in \mathbf{R}\colon M_X(t) < \infty\}.$$

Zbiór Λ zawiera otoczenie zera. Definiujemy funkcję
$$\psi(t) = \log M_X(t), \quad t \in \Lambda. \tag{1}$$

Funkcja ψ jest wypukła (i ściśle wypukła, jeśli X nie jest z prawdopodobieństwem 1 stała). Istotnie, dla $0 < a < 1$ otrzymujemy z nierówności Höldera z wykładnikami $p = 1/a$ i $q = 1/(1-a)$:

$$\begin{aligned}\psi(at + (1-a)s) &= \log \mathcal{E}\left(\mathrm{e}^{atX} \cdot \mathrm{e}^{(1-a)sX}\right) \\ &\leqslant \log \left(\mathcal{E}\mathrm{e}^{atX\cdot(1/a)}\right)^a \left(\mathcal{E}\mathrm{e}^{(1-a)sX\cdot(1/(1-a))}\right)^{1-a} = \\ &= a \cdot \log \mathcal{E}\mathrm{e}^{tX} + (1-a)\cdot \log \mathcal{E}\mathrm{e}^{sX} = a\psi(t) + (1-a)\psi(s).\end{aligned}$$

Mamy $\psi(0) = 0$, $\psi'(0) = \mathcal{E}X = m$,$\psi''(t) \geqslant 0$, jako że funkcja ψ jest wypukła i dwukrotnie różniczkowalna.

Przedłużymy ψ na cały zbiór liczb rzeczywistych, kładąc $\psi(t) = \infty$ dla $t \notin \Lambda$.

Jesteśmy teraz gotowi do zdefiniowania *transformaty Craméra*:

Definicja 1. *Transformatą Craméra rozkładu μ_X zmiennej losowej X nazywamy funkcję*
$$H(a) = \sup_{t\in\mathbf{R}} [at - \psi(t)],$$
gdzie funkcja ψ została zdefiniowana równością (1).

Funkcja H jest wypukła. Istotnie, niech $\lambda + \mu = 1$, $t, s \geqslant 0$. Wtedy

$$\begin{aligned}H(\lambda a + \mu b) &= \sup_{t\in\mathbf{R}}[(\lambda a + \mu b)t - \lambda\psi(t) - \mu\psi(t)] \leqslant \\ &\leqslant \lambda \sup_{t\in\mathbf{R}}[at - \psi(t)] + \mu \sup_{t\in\mathbf{R}}[bt - \psi(t)] = \lambda H(a) + \mu H(b).\end{aligned}$$

Ponadto H jest nieujemna, bowiem $\psi(0) = 0$, i wreszcie $H(a) = 0$ wtedy i tylko wtedy, gdy $a = \psi'(0) = m$.

Istotnie, $\psi(0) = 0$, z wypukłości wynika, że prosta o równaniu $y = ax$ przecina wykres funkcji w zerze i jeszcze jednym punkcie. Dla $a \neq m$ zawsze istnieje takie x, że $ax > \psi(x)$, a jeśli $a = \psi'(0)$, to prosta $y = ax$ jest styczna do wykresu ψ w punkcie 0, więc leży cała pod wykresem ψ i $H(m) = 0$.

$$a > \psi'(0) \Rightarrow H(a) = \sup_{t>0}[at - \psi(t)],$$

$$a < \psi'(0) \Rightarrow H(a) = \sup_{t<0}[at - \psi(t)],$$

Oszacujemy teraz szybkość zbieżności w MPWL.

Twierdzenie 2. *Niech X, X_1, X_2, ... będą niezależnymi zmiennymi losowymi o tym samym rozkładzie, spełniającymi warunek Craméra. Wtedy*

$$P\left(\sup_{k\geqslant n}\left|\frac{S_k}{k} - m\right| > \varepsilon\right) \leqslant 2\mathrm{e}^{-n\min(H(m+\varepsilon),H(m-\varepsilon))},$$

gdzie $m = \mathcal{E}X$, a H jest transformatą Craméra zmiennej losowej X.

D o w ó d. Niech a, λ spełniają warunek

$$\lambda a - \log M(\lambda) \geqslant 0. \tag{2}$$

Ustalmy $n \geqslant 1$. Niech $\tau = \inf\{k \geqslant n \colon S_k/k > a\}$ (zwróćmy uwagę, że $\tau = \infty$, gdy zawsze $S_k/k \leqslant a$). Mamy

$$\begin{aligned}
P\left(\sup_{k\geqslant n}\frac{S_k}{k} > a\right) &= P\left(\frac{S_\tau}{\tau} > a, \tau < \infty\right) = P(\mathrm{e}^{\lambda S_\tau} > \mathrm{e}^{\lambda a\tau}, \tau < \infty) = \\
&= P(\mathrm{e}^{\lambda S_\tau - \tau\log M(\lambda)} > \mathrm{e}^{\lambda a\tau - \tau\log M(\lambda)}, \tau < \infty) \leqslant \\
&\leqslant P(\mathrm{e}^{\lambda S_\tau - \tau\log M(\lambda)} > \mathrm{e}^{n(\lambda a - \log M(\lambda))}, \tau < \infty) \leqslant \\
&\leqslant P(\sup_{k\geqslant n}\mathrm{e}^{\lambda S_k - k\log M(\lambda)} > \mathrm{e}^{n(\lambda a - \log M(\lambda))}, \tau < \infty).
\end{aligned}$$

Oznaczmy $Z_k = \mathrm{e}^{\lambda S_k - k\log M(\lambda)}$ i niech $\mathcal{F}_k = \sigma(X_1, \ldots, X_k)$. Wtedy ciąg $(Z_k, \mathcal{F}_k)$ jest nieujemnym martyngałem, co nietrudno sprawdzić. Ponieważ

$$Z_k = Z_{k-1}\mathrm{e}^{\lambda X_k - \log M(\lambda)},$$

mamy

$$\begin{aligned}
\mathcal{E}\left(Z_k \mid \mathcal{F}_{k-1}\right) &= Z_{k-1}\mathcal{E}\left(\mathrm{e}^{\lambda X_k - \log M(\lambda)} \,\middle|\, \mathcal{F}_{k-1}\right) = Z_{k-1}\mathcal{E}\mathrm{e}^{\lambda X_k - \log M(\lambda)} = \\
&= Z_{k-1}\cdot M(\lambda)\cdot \mathrm{e}^{-\log M(\lambda)} = Z_{k-1}.
\end{aligned}$$

Ponadto $\mathcal{E}Z_k = 1$, $k = 1, 2, \ldots$.

Z nierówności maksymalnej 11.4.1 otrzymujemy, przy spełnieniu (2),

$$P\left(\sup_{k\geqslant n}\frac{S_k}{k} > a\right) \leqslant P\left(\sup_{k\geqslant n} \mathrm{e}^{\lambda S_k - k\log M(\lambda)} > \mathrm{e}^{n(\lambda a - \log M(\lambda))}\right) \leqslant$$
$$\leqslant \mathrm{e}^{-n(\lambda a - \log M(\lambda))}.$$

Jeśli $a > m$, funkcja $f(\lambda) = \lambda a - \log M(\lambda)$ spełnia następujące warunki: $f(0) = 0$, $f'(0) = a - m > 0$, zatem istnieje takie $\lambda > 0$, że spełniony jest warunek (2). Zatem dla $a > m$ mamy

$$P\left(\sup_{k\geqslant n}\frac{S_k}{k} > a\right) \leqslant \mathrm{e}^{-n\sup_{\lambda>0}[\lambda a - \log M(\lambda)]} = \mathrm{e}^{-nH(a)}.$$

Zupełnie analogicznie dla $a < m$ mamy

$$P\left(\inf_{k\geqslant n}\frac{S_k}{k} < a\right) \leqslant \mathrm{e}^{-n\sup_{\lambda<0}[\lambda a - \log M(\lambda)]} = \mathrm{e}^{-nH(a)}.$$

Ostatecznie

$$P\left(\sup_{k\geqslant n}\left|\frac{S_k}{k} - m\right| > \varepsilon\right) \leqslant P\left(\sup_{k\geqslant n}\frac{S_k}{k} > m+\varepsilon\right) + P\left(\inf_{k\geqslant n}\frac{S_k}{k} < m-\varepsilon\right) \leqslant$$
$$\leqslant 2\mathrm{e}^{-n\min(H(m+\varepsilon), H(m-\varepsilon))}.$$

Przykład 3. Jeśli $X, X_1, X_2, \ldots$ jest ciągiem Bernoulliego, to $M_X(t) = \cosh t$ i nietrudno obliczyć, że funkcja $at - \log\cosh t$ przyjmuje maksimum dla $t = \operatorname{ar\,tgh} a = \frac{1}{2}\log\frac{1+a}{1-a}$, o ile $|a| < 1$, a jeśli $|a| \geqslant 1$, jej kresem górnym jest ∞. W takim razie

$$H(a) = \begin{cases} \frac{1}{2}[(1+a)\log(1+a) + (1-a)\log(1-a)] & \text{dla } |a| < 1 \\ \infty & \text{w p.p.} \end{cases}$$

Rozwijając logarytm w szereg potęgowy widzimy, że $H(a) \geqslant \frac{1}{2}a^2$. Wobec tego

$$P\left(\sup_{k\geqslant n}\left|\frac{S_k}{k}\right| > \varepsilon\right) \leqslant 2\mathrm{e}^{-nH(\varepsilon)} \leqslant \begin{cases} 2\mathrm{e}^{-\frac{1}{2}n\varepsilon^2} & \text{dla } 0 \leqslant \varepsilon < 1 \\ 0 & \text{dla } \varepsilon \geqslant 1. \ \blacksquare \end{cases}$$