Wstęp do rachunku prawdopodobieństwa
11. Prawa wielkich liczb. Centralne twierdzenie graniczne

Zadanie 1. *Średnica sekwoi wzrasta od 1 do 2 cm rocznie, przy czym zakładamy, że przyrost ma rozkład jednostajny na odcinku $[1,2]$, a przyrosty w kolejnych latach nie zależą od siebie. Oszacuj prawdopodobieństwo, że w ciągu 300 lat średnica zwiększy się o co najmniej 460 cm, stosując:*

(a) *nierówność Markowa,*

Zacznijmy od wprowadzenia zmiennej losowej S_{300} oznaczającej przyrost średnicy sekwoi w ciągu 300 lat. Możemy wyrazić tę zmienną losową za pomocą sumy pomocniczych zmiennych losowych X_i, $i = 1, 2, \ldots, 300$, oznaczających przyrosty w kolejnych latach, czyli

$$S_{300} = X_1 + X_2 + \ldots + X_{300}.$$

Ponieważ w ciągu roku każdy przyrost pomiędzy 1 i 2 cm jest równoprawdopodobny, każda ze zmiennych losowych X_i ma ten sam rozkład, mianowicie rozkład jednostajny na odcinku $[1,2]$. Jak pamiętamy z poprzednich zajęć, wartość oczekiwana takiego rozkładu jest równa dokładnie środkowi przedziału, na którym ten rozkład jest określony, a zatem dla każdego $i = 1, 2, \ldots, 300$ mamy

$$\mathbb{E}X_i = \frac{1+2}{2} = \frac{3}{2},$$

skąd otrzymujemy

$$\mathbb{E}S_{300} = \mathbb{E}X_1 + \mathbb{E}X_2 + \ldots + \mathbb{E}X_{300} = 300 \cdot \frac{3}{2} = 450.$$

Stosując nierówność Markowa do zmiennej losowej $S_{300} = |S_{300}|$ otrzymujemy:

$$\mathbb{P}\left(S_{300} \geqslant 460\right) \leqslant \frac{\mathbb{E}S_{300}}{460} = \frac{450}{460} \approx 0{,}978.$$

(b) *nierówność Czebyszewa,*

Aby móc zastosować nierówność Czebyszewa musimy najpierw wyznaczyć wariancję zmiennej losowej S_{300}. W tym celu przyjmiemy założenie, że przyrosty sekwoi w kolejnych latach nie zależą od siebie, a zatem zmienne losowe $X_1, X_2, \ldots, X_{300}$ są niezależne. Dla pomocniczej zmiennej losowej X_i o rozkładzie jednostajnym na odcinku $[1,2]$ mamy:

$$\mathrm{Var}X_i = \frac{(2-1)^2}{12} = \frac{1}{12}.$$

W związku z tym, przyjmując założenie o niezależności zmiennych losowych $X_1, X_2, \ldots, X_{300}$, otrzymujemy:

$$\mathrm{Var}S_{300} = \mathrm{Var}X_1 + \mathrm{Var}X_2 + \ldots + \mathrm{Var}X_{300} = 300 \cdot \frac{1}{12} = 25.$$

Z nierówności Czebyszewa mamy następujące oszacowanie:

$$\mathbb{P}\left(S_{300} \geqslant 460\right) = \mathbb{P}\left(S_{300} - \mathbb{E}S_{300} \geqslant 460 - 450\right) \leqslant \mathbb{P}\left(|S_{300} - \mathbb{E}S_{300}| \geqslant 10\right) \leqslant \frac{\mathrm{Var}S_{300}}{10^2} = \frac{25}{100} = \frac{1}{4}.$$

(c) *Centralne Twierdzenie Graniczne.*

Znając już wartość oczekiwaną oraz wariancję zmiennej losowej S_{300} możemy od razu przejść do szacowania prawdopodobieństwa interesującego nas zdarzenia:

$$\begin{aligned}\mathbb{P}\left(S_{300} \geqslant 460\right) &= \mathbb{P}\left(\frac{S_{300} - \mathbb{E}S_{300}}{\sqrt{\mathrm{Var}S_{300}}} \geqslant \frac{460-450}{\sqrt{25}}\right) = \mathbb{P}\left(\frac{S_{300} - \mathbb{E}S_{300}}{\sqrt{\mathrm{Var}S_{300}}} \geqslant 2\right) \\ &= 1 - \mathbb{P}\left(\frac{S_{300} - \mathbb{E}S_{300}}{\sqrt{\mathrm{Var}S_{300}}} < 2\right) \approx 1 - \Phi(2) \approx 1 - 0{,}977 = 0{,}023.\end{aligned}$$

Zadanie 2. *Rzucamy symetryczną kostką 100 razy. Oszacuj prawdopodobieństwo, że wyrzucimy szóstkę mniej niż 10 razy, stosując:*

(a) *twierdzenie Czebyszewa,*

Niech X będzie liczbą szóstek wyrzuconych w 100 rzutach kostką. Wówczas zmienna losowa X ma rozkład dwumianowy $Bin(100, 1/6)$ z parametrami

$$\mathbb{E}X = 100 \cdot \frac{1}{6} = \frac{50}{3} \quad \text{oraz} \quad \mathrm{Var}X = 100 \cdot \frac{1}{6} \cdot \frac{5}{6} = \frac{125}{9}.$$

Stosując nierówność Czebyszewa do zmiennej losowej $X = |X|$ mamy:

$$\mathbb{P}\left(X < 10\right) = \mathbb{P}\left(X - \mathbb{E}X < 10 - \frac{50}{3}\right) \leqslant \mathbb{P}\left(|X - \mathbb{E}X| > \frac{20}{3}\right) \leqslant \mathbb{P}\left(|X - \mathbb{E}X| \geqslant \frac{20}{3}\right)$$
$$\leqslant \frac{\mathrm{Var}X}{\left(\frac{20}{3}\right)^2} = \frac{125}{9} \cdot \frac{9}{400} = 0{,}3125.$$

(b) *CTG bez poprawki 1/2,*

$$\mathbb{P}\left(X < 10\right) = \mathbb{P}\left(\frac{X - \mathbb{E}X}{\sqrt{\mathrm{Var}X}} < \frac{10 - \frac{50}{3}}{\sqrt{\frac{125}{9}}}\right) = \mathbb{P}\left(\frac{X - \mathbb{E}X}{\sqrt{\mathrm{Var}X}} < -\frac{20}{3} \cdot \frac{3}{5\sqrt{5}}\right) = \mathbb{P}\left(\frac{X - \mathbb{E}X}{\sqrt{\mathrm{Var}X}} < -\frac{4\sqrt{5}}{5}\right)$$
$$\approx \Phi\left(-\frac{4\sqrt{5}}{5}\right) \approx \Phi(-1{,}79) = 1 - \Phi(1{,}79) \approx 1 - 0{,}96327 = 0{,}03673$$

(c) *CTG z poprawką 1/2.*

Ponieważ zmienna losowa X przyjmuje tylko wartości całkowite, mamy:

$$\mathbb{P}\left(X < 10\right) = \mathbb{P}\left(X < 9{,}5\right) = \mathbb{P}\left(\frac{X - \mathbb{E}X}{\sqrt{\mathrm{Var}X}} < \frac{\frac{19}{2} - \frac{50}{3}}{\sqrt{\frac{125}{9}}}\right) = \mathbb{P}\left(\frac{X - \mathbb{E}X}{\sqrt{\mathrm{Var}X}} < -\frac{43}{6} \cdot \frac{3}{5\sqrt{5}}\right) = \mathbb{P}\left(\frac{X - \mathbb{E}X}{\sqrt{\mathrm{Var}X}} < -\frac{43\sqrt{5}}{50}\right)$$
$$\approx \Phi\left(-\frac{43\sqrt{5}}{50}\right) \approx \Phi(-1{,}92) = 1 - \Phi(1{,}92) \approx 1 - 0{,}97257 = 0{,}02743$$

Czy możemy oszacować to prawdopodobieństwo, stosując twierdzenie Markowa?

Spróbujmy oszacować to prawdopodobieństwo stosując nierówność Markowa do zmiennej losowej $X = |X|$:

$$\mathbb{P}\left(X < 10\right) = 1 - \mathbb{P}\left(X \geqslant 10\right) \geqslant 1 - \frac{\mathbb{E}X}{10} = 1 - \frac{5}{3} = -\frac{2}{3}.$$

A zatem widzimy, że dostajemy oszacowanie wynikające z trywialnego oszacowania $\mathbb{P}\left(X < 10\right) \geqslant 0$.

Stosując CTG, oszacuj prawdopodobieństwo wyrzucenia dokładnie 10 szóstek i porównaj je z rzeczywistą wartością tego prawdopodobieństwa.

Ponieważ zmienna losowa X przyjmuje tylko wartości całkowite, do oszacowania $\mathbb{P}\left(X = 10\right)$ możemy zastosować poprawkę 1/2:

$$\mathbb{P}\left(X = 10\right) = \mathbb{P}\left(9{,}5 < X \leqslant 10{,}5\right) = \mathbb{P}\left(\frac{\frac{19}{2} - \frac{50}{3}}{\sqrt{\frac{125}{9}}} < \frac{X - \mathbb{E}X}{\sqrt{\mathrm{Var}X}} \leqslant \frac{\frac{21}{2} - \frac{50}{3}}{\sqrt{\frac{125}{9}}}\right) = \mathbb{P}\left(-\frac{43\sqrt{5}}{50} < \frac{X - \mathbb{E}X}{\sqrt{\mathrm{Var}X}} \leqslant -\frac{37\sqrt{5}}{50}\right)$$
$$\approx \Phi\left(-\frac{37\sqrt{5}}{50}\right) - \Phi\left(-\frac{43\sqrt{5}}{50}\right) \approx \Phi(-1{,}65) - \Phi(-1{,}92)$$
$$= \Phi(1{,}92) - \Phi(1{,}65) \approx 0{,}97257 - 0{,}95053 = 0{,}02204.$$

Natomiast rzeczywista wartość tego prawdopodobieństwa z dokładnością do pięciu miejsc po przecinku wynosi:

$$\binom{100}{10}\left(\frac{1}{6}\right)^{10}\left(\frac{5}{6}\right)^{90} \approx 0{,}02140.$$

A zatem widzimy, że oszacowanie z CTG dało nam wynik, który różni się od rzeczywistej wartości o około 0,001.

Zadanie 3. *Kierowniczka urzędu pocztowego twierdzi, że w jej placówce klient czeka w kolejce do okienka krócej niż 15 minut. Dociekliwa ekipa telewizyjna chce zweryfikować jej deklarację, mierząc średni czas obsługi 50 klientów. Jakie jest prawdopodobieństwo, że potwierdzi ona deklarację kierowniczki urzędu jeśli wiadomo, że rzeczywisty czas T stania w kolejce jest taką ciągłą zmienną losową, że $\mathbb{E}T = 14$ i $VarT = 23$? Czy należy stosować w tym przypadku „poprawkę 1/2”?*

W tym zadaniu interesuje nas średni czas oczekiwania w kolejce na podstawie pomiaru czasów oczekiwania 50 klientów. A zatem wprowadźmy zmienną losową S_{50}, która będzie liczyła łączny czas oczekiwania 50 klientów. Wówczas średni czas wyraża się wzorem $S_{50}/50$. Dodatkowo wprowadźmy pomocnicze zmienne losowe $T_1, T_2, \ldots, T_{50}$, które będą zwracały czasy oczekiwania poszczególnych klientów, czyli

$$S_{50} = T_1 + T_2 + \ldots + T_{50}.$$

Załóżmy idealistycznie, że zmienne te są niezależne (co oznacza, że nie powinniśmy mierzyć czasów oczekiwania kolejnych klientów, tylko dokonywać pomiaru co jakiś czas, na przykład co godzinę).

Każda ze zmiennych losowych T_i ma rozkład taki sam jak zmienna losowa T, czyli:

$$\mathbb{E}T_i = 14 \ \text{ oraz } \ \mathrm{Var}T_i = 23, \ \ i = 1, 2, \ldots, 50.$$

Przy założeniu niezależności zmiennych losowych $T_1, T_2, \ldots, T_{50}$ otrzymujemy również:

$$\mathbb{E}S_{50} = \mathbb{E}T_1 + \mathbb{E}T_2 + \ldots + \mathbb{E}T_{50} = 50 \cdot 14 = 700,$$

$$\mathrm{Var}S_{50} = \mathrm{Var}T_1 + \mathrm{Var}T_2 + \ldots + \mathrm{Var}T_{50} = 50 \cdot 23 = 1150.$$

Oszacujmy teraz z CTG prawdopodobieństwo tego, że średni czas obsługi jest krótszy niż 15 minut. Nie będziemy stosować w tym przypadku poprawki 1/2, ponieważ mamy do czynienia z czasem, który jest ciągły. A zatem otrzymujemy następujące oszacowanie:

$$\mathbb{P}\left(\frac{S_{50}}{50} < 15\right) = \mathbb{P}\left(S_{50} < 750\right) = \mathbb{P}\left(\frac{S_{50} - \mathbb{E}S_{50}}{\sqrt{\mathrm{Var}S_{50}}} < \frac{750 - 700}{\sqrt{1150}}\right) \approx \Phi\left(\frac{750 - 700}{\sqrt{1150}}\right)$$
$$\approx \Phi(1{,}47) \approx 0{,}92922.$$

Przypuśćmy, że ekipa telewizyjna zapytała uprzednio kierowniczkę, czas ilu kolejnych klientów ma zmierzyć. Co powinna ona odpowiedzieć, by mieć 99% pewności, że wyliczenia ekipy nie obalą jej twierdzenia?

Jak widzimy z poprzedniego podpunktu, przy pomiarze czasów 50 klientów mamy pewność potwierdzenia tezy na poziomie około 93%. Aby mieć jeszcze większą pewność, powinniśmy przebadać większą grupę klientów. Niech zatem n będzie szukaną liczbą klientów. Postępując podobnie jak w poprzednim podpunkcie, korzystając z CTG oszacujemy średni czas ich obsługi. W tym celu znów wprowadzamy zmienną losową S_n – łączny czas obsługi n klientów, oraz zmienne pomocnicze $T_1, T_2, \ldots, T_n$ – czasy obsługi poszczególnych klientów. Mamy zatem:

$$\mathbb{E}S_n = \mathbb{E}T_1 + \mathbb{E}T_2 + \ldots + \mathbb{E}T_n = n \cdot 14,$$

$$\mathrm{Var}S_n = \mathrm{Var}T_1 + \mathrm{Var}T_2 + \ldots + \mathrm{Var}T_n = n \cdot 23.$$

Korzystając z CTG szacujemy następnie prawdopodobieństwo tego, że średni czas obsługi jest krótszy niż 15 minut:

$$\mathbb{P}\left(\frac{S_n}{n} < 15\right) = \mathbb{P}\left(S_n < 15n\right) = \mathbb{P}\left(\frac{S_n - \mathbb{E}S_n}{\sqrt{\mathrm{Var}S_n}} < \frac{15n - 14n}{\sqrt{23n}}\right) = \mathbb{P}\left(\frac{S_n - \mathbb{E}S_n}{\sqrt{\mathrm{Var}S_n}} < \frac{\sqrt{23n}}{23}\right) \approx \Phi\left(\frac{\sqrt{23n}}{23}\right).$$

Naszym celem jest tak dobrać n, aby powyższe prawdopodobieństwo było równe co najmniej 0,99. Z tablic dystrybuanty standardowego rozkładu normalnego otrzymujemy, że $\Phi(2{,}33) > 0{,}99$ i $\Phi(2{,}32) < 0{,}99$. A zatem jeżeli weźmiemy n spełniające:

$$\frac{\sqrt{23n}}{23} \geqslant 2{,}33,$$

to otrzymamy

$$\mathbb{P}\left(\frac{S_n}{n} < 15\right) \approx \Phi\left(\frac{\sqrt{23n}}{23}\right) \geqslant \Phi(2{,}33) > 0{,}99.$$

W związku z tym wystarczy wziąć n spełniające poniższą nierówność:

$$(2{,}33 \cdot \sqrt{23})^2 \approx 124{,}86 < 125 \leqslant n.$$

TABLICA 1. Dystrybuanta $\Phi(x)$ standardowego rozkładu normalnego $\mathcal{N}(0,1)$

x	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0	0,5	0,50399	0,50798	0,51197	0,51595	0,51994	0,52392	0,5279	0,53188	0,53586
0,1	0,53983	0,5438	0,54776	0,55172	0,55567	0,55962	0,56356	0,56749	0,57142	0,57535
0,2	0,57926	0,58317	0,58706	0,59095	0,59483	0,59871	0,60257	0,60642	0,61026	0,61409
0,3	0,61791	0,62172	0,62552	0,6293	0,63307	0,63683	0,64058	0,64431	0,64803	0,65173
0,4	0,65542	0,6591	0,66276	0,6664	0,67003	0,67364	0,67724	0,68082	0,68439	0,68793
0,5	0,69146	0,69497	0,69847	0,70194	0,7054	0,70884	0,71226	0,71566	0,71904	0,7224
0,6	0,72575	0,72907	0,73237	0,73565	0,73891	0,74215	0,74537	0,74857	0,75175	0,7549
0,7	0,75804	0,76115	0,76424	0,7673	0,77035	0,77337	0,77637	0,77935	0,7823	0,78524
0,8	0,78814	0,79103	0,79389	0,79673	0,79955	0,80234	0,80511	0,80785	0,81057	0,81327
0,9	0,81594	0,81859	0,82121	0,82381	0,82639	0,82894	0,83147	0,83398	0,83646	0,83891
1	0,84134	0,84375	0,84614	0,84849	0,85083	0,85314	0,85543	0,85769	0,85993	0,86214
1,1	0,86433	0,8665	0,86864	0,87076	0,87286	0,87493	0,87698	0,879	0,881	0,88298
1,2	0,88493	0,88686	0,88877	0,89065	0,89251	0,89435	0,89617	0,89796	0,89973	0,90147
1,3	0,9032	0,9049	0,90658	0,90824	0,90988	0,91149	0,91309	0,91466	0,91621	0,91774
1,4	0,91924	0,92073	0,9222	0,92364	0,92507	0,92647	0,92785	0,92922	0,93056	0,93189
1,5	0,93319	0,93448	0,93574	0,93699	0,93822	0,93943	0,94062	0,94179	0,94295	0,94408
1,6	0,9452	0,9463	0,94738	0,94845	0,9495	0,95053	0,95154	0,95254	0,95352	0,95449
1,7	0,95543	0,95637	0,95728	0,95818	0,95907	0,95994	0,9608	0,96164	0,96246	0,96327
1,8	0,96407	0,96485	0,96562	0,96638	0,96712	0,96784	0,96856	0,96926	0,96995	0,97062
1,9	0,97128	0,97193	0,97257	0,9732	0,97381	0,97441	0,975	0,97558	0,97615	0,9767
2	0,97725	0,97778	0,97831	0,97882	0,97932	0,97982	0,9803	0,98077	0,98124	0,98169
2,1	0,98214	0,98257	0,983	0,98341	0,98382	0,98422	0,98461	0,985	0,98537	0,98574
2,2	0,9861	0,98645	0,98679	0,98713	0,98745	0,98778	0,98809	0,9884	0,9887	0,98899
2,3	0,98928	0,98956	0,98983	0,9901	0,99036	0,99061	0,99086	0,99111	0,99134	0,99158
2,4	0,9918	0,99202	0,99224	0,99245	0,99266	0,99286	0,99305	0,99324	0,99343	0,99361
2,5	0,99379	0,99396	0,99413	0,9943	0,99446	0,99461	0,99477	0,99492	0,99506	0,9952
2,6	0,99534	0,99547	0,9956	0,99573	0,99585	0,99598	0,99609	0,99621	0,99632	0,99643
2,7	0,99653	0,99664	0,99674	0,99683	0,99693	0,99702	0,99711	0,9972	0,99728	0,99736
2,8	0,99744	0,99752	0,9976	0,99767	0,99774	0,99781	0,99788	0,99795	0,99801	0,99807
2,9	0,99813	0,99819	0,99825	0,99831	0,99836	0,99841	0,99846	0,99851	0,99856	0,99861
3	0,99865	0,99869	0,99874	0,99878	0,99882	0,99886	0,99889	0,99893	0,99896	0,999
3,1	0,99903	0,99906	0,9991	0,99913	0,99916	0,99918	0,99921	0,99924	0,99926	0,99929
3,2	0,99931	0,99934	0,99936	0,99938	0,9994	0,99942	0,99944	0,99946	0,99948	0,9995
3,3	0,99952	0,99953	0,99955	0,99957	0,99958	0,9996	0,99961	0,99962	0,99964	0,99965
3,4	0,99966	0,99968	0,99969	0,9997	0,99971	0,99972	0,99973	0,99974	0,99975	0,99976
3,5	0,99977	0,99978	0,99978	0,99979	0,9998	0,99981	0,99981	0,99982	0,99983	0,99983
3,6	0,99984	0,99985	0,99985	0,99986	0,99986	0,99987	0,99987	0,99988	0,99988	0,99989
3,7	0,99989	0,9999	0,9999	0,9999	0,99991	0,99991	0,99992	0,99992	0,99992	0,99992
3,8	0,99993	0,99993	0,99993	0,99994	0,99994	0,99994	0,99994	0,99995	0,99995	0,99995
3,9	0,99995	0,99995	0,99996	0,99996	0,99996	0,99996	0,99996	0,99996	0,99997	0,99997
4	0,99997	0,99997	0,99997	0,99997	0,99997	0,99997	0,99998	0,99998	0,99998	0,99998