Rozdział 2

Rachunek prawdopodobieństwa

2.1 Przestrzeń probabilistyczna, zdarzenia losowe i prawdopodobieństwo

Mówimy o **doświadczeniu losowym** jeśli nie możemy przewidzieć wyniku, ale możemy oszacować *prawdopodobieństwo* wystąpienia poszczególnych wyników doświadczenia.

Przykład. Wyniki 100 rzutów kostką były takie:

R R R 0 0 0 0 R 0 R 0 0 0 R R 0 0 0 0 0 0 R 0 R R R R R 0 R 0 0 0 R R 0 R
R 0 R R R R R R 0 0 R R R R R 0 0 R R R R 0 R R R 0 0 0 R 0 R 0 R 0 0 0
R 0 0 0 0 R 0 0 0 R R 0 R R 0 R R R 0 R R 0 R 0 0 R 0

Nie możemy *przewidzieć*, co wyjdzie w kolejnym, 101 rzucie ale oceniamy prawdopodobieństwo otrzymania reszki na 0.5. Co to właściwie znaczy, że prawdopodobieństwo reszki jest 0.5? W 100 rzutach zdarzyły się akurat 52 reszki. Gdy powtórzono doświadczenie jeszcze dodatkowo 400 razy, w kolejnych setkach rzutów pojawiły się 53, 47, 42 i 41 reszki. W dużej liczbie rzutów udział ,,reszek" zbliża się do 50%. Obrazuje to nast'epująca tabelka, która opisuje stosunek liczby reszek do liczby rzutów dla doświadczenia, kontynuowanego aż do 100000 rzutów.

liczba rzutów	liczba rezek	liczba reszek/liczba rzutów
100	52	0.52000
200	105	0.52500
300	152	0.50667
400	194	0.48500
500	235	0.47000
1000	485	0.48500
10000	5025	0.50250
100000	50194	0.50194

□

Matematycznym modelem doświadczenia losowego jest przestrzeń probabilistyczna, wyposażona w rozkład prawdopodobieństwa. Opiszemy teraz i wyjaśnimy te pojęcia.

DEFINICJA. Przestrzeń probabilistyczna jest to zbiór wszystkich możliwych wyników doświadczenia losowego. Będziemy oznaczali ten zbiór literą

$$\Omega.$$

Elementy tego zbioru, czyli pojedyncze wyniki oznaczamy zwykle

$$\omega \in \Omega.$$

Przykład. (Rzut kostką) $\Omega = \{1, 2, 3, 4, 5, 6\}$.

Zauważmy, że pojęcie ,,doświadczenia losowego" rozumiemy bardzo szeroko: może to być obserwacja każdego zjawiska, w którym pewną rolę odgrywa przypadek. Dla rozmaitych doświadczeń losowych nie jest trudno dobrać odpowiednie przestrzenie probabilistyczne. Przyjrzyjmy sie kilku przykładom:

- Obserwujemy liczbę wypadków drogowych w powiecie toruńskim w dniu 20 października 2020 r. Wynikiem tego ,,doświadczenia" jest pewna liczba całkowita, dodatnia lub zero. Możemy zatem przyjąć, że $\Omega = \{0, 1, 2, \ldots\}$.
- Obserwujemy czas życia p. Ildefonsa Kowalskiego, który właśnie obchodzi trzydzieste piąte urodziny. Jasne, że mamy do czynienia z doświadczeniem losowym. (powiedzmy, w milimetrach). Wynikiem jest pewna liczba dodatnia lub zero, niekoniecznie całkowita (powiedzmy, że czas wyrażamy w latach). Przyjmijmy, że $\Omega = [35, \infty)$. Jasne jest, że raczej nie spodziewamy się czasu życia rzędu tysiąca lat. Wybraliśmy przedział nieograniczony aby uniknąc jałowych dyskusji o tym, jaka największy czas życia jest jeszcze możliwy!

- W losowaniu TOTO-LOTKA, wynikiem doświadczenia (przeprowadzanego przez specjalną maszynę przed oczami telewidzów) jest zbiór sześciu liczb wybranych spośród 49. Ω jest więc zbiorem wszystkich podzbiorów sześcioelementowych zbioru $\{1, 2, \ldots, 49\}$. Nauczymy się później, jak policzyć liczbę takich podzbiorów. Okazuje się, że $\#\Omega = 13983816$.

DEFINICJA. Rozkład prawdopodobieństwa na przestrzeni probabilistycznej

$$\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$$

jest opisany tabelką:

wynik	ω_1	ω_2	$\ldots$	ω_n
prawdopodobieństwo	p_1	p_2	$\ldots$	p_n

gdzie $p_i \geqslant 0$ i $\sum_{i=1}^{n} p_i = 1$.

Przykład. (Rzut kostką) Rozkład prawdopodobieństwa:

wynik	1	2	3	4	5	6
prawdopodobieństwo	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

DEFINICJA. Zdarzenie losowe jest to podzbiór przestrzeni probabilistycznej:

$$A \subset \Omega.$$

Prawdopodobieństwo zdarzenia A jest to liczba

$$\mathbb{P}(A) = \sum_{\omega_i \in A} p_i$$

Przykład. (Rzut kostką) Zdarzenie ,,wypadła parzysta liczba oczek" jest to zbiór $A = \{2, 4, 6\}$. Prawdopodobieństwo tego zdarzenia jest równe $\mathbb{P}(A) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$.

,,Klasyczna definicja prawdopodobieństwa:" Wszystkie możliwe wyniki są *jednakowo prawdopodobne.*

wynik	ω_1	ω_2	$\ldots$	ω_n
prawdopodobieństwo	$\frac{1}{n}$	$\frac{1}{n}$	$\ldots$	$\frac{1}{n}$

Wtedy

$$\mathbb{P}(A) = \frac{\#A}{\#\Omega},$$

gdzie $\#A$ oznacza *liczbę elementów* zbioru A.

Uwaga. **Nie zawsze** można zakładać, że tak jest!

Przykład. (2 rzuty monetą)

$$\Omega = \{OO, OR, RO, RR\}.$$

Rozkład prawdopodobieństwa:

wynik	OO	OR	RO	RR
prawdopodobieństwo	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

Przykład. (Rzut dwiema jednakowymi monetami)

$$\Omega = \{2O, OR, 2R\}$$

(nie odróżniamy OR od RO). Rozkład prawdopodobieństwa:

wynik	$2O$	OR	$2R$
prawdopodobieństwo	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

Prawdopodobieństwo geometryczne

Niekiedy za przestrzeń probabilistyczną Ω można przyjąć pewien zbiór na płaszczyźnie, zaś prawdopodobieństwo zdarzenia $A \subset \Omega$ określić wzorem

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|},$$

gdzie $|A|$ oznacza *pole* ,,figury" A.

Przykład. (Rzucanie strzałką do tarczy). Wynikiem doświadczenia jest punkt trafienia. Załóżmy, że *zawsze* trafiamy w tarczę i trafienie w każdy punkt tarczy jest ,,jednakowo prawdopodobne". Możemy wtedy przyjąć, że Ω jest tarczą i szansa trafienia w zbiór $A \subset \Omega$ jest równa $|A|/|\Omega|$.

Jeśli tarcza jest kołem o promieniu b, zaś ,,dziesiątka" – małym kółkiem o promieniu $b/10$, to szansa trafienia w ,,dziesiątkę" jest równa

$$\frac{\pi(b/10)^2}{\pi b^2} = \frac{1}{100}.$$

Uwaga. Mamy tu *nieskończony* zbiór Ω. Prawdopodobieństwo otrzymania każdego pojedynczego wyniku, czyli trafienia *dokładnie* w punkt $\omega \in \Omega$ (np. w środek tarczy) jest równe 0.

Działania na zdarzeniach:

$A \cap B$	jest zdarzeniem	„zaszło A **i** B"
$A \cup B$	jest zdarzeniem	„zaszło A **lub** B"
$A' = \Omega \backslash A$	jest zdarzeniem	„**nie** zaszło A"
Ω	jest zdarzeniem	**pewnym**
$\emptyset$	jest zdarzeniem	**niemożliwym**

$A \cap B = \emptyset$: zdarzenia A i B się **wykluczają**;
$A \subset B$: jeśli zajdzie A, to **musi** zajść B

Podstawowe własności prawdopodobieństwa:

- Jeśli $A \cap B = \emptyset$ to $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.
- $\mathbb{P}(\Omega) = 1$, $\mathbb{P}(\emptyset) = 0$.
- $\mathbb{P}(A') = 1 - \mathbb{P}(A)$.
- (Wzór „włączeń i wyłączeń"): dla *dowolnych* zdarzeń A i B,

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

- (Wzór „włączeń i wyłączeń"): dla *dowolnych* zdarzeń A, B i C,

$$\begin{aligned}\mathbb{P}(A \cup B \cup C) = {} & \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) \\ & - \mathbb{P}(A \cap B) - \mathbb{P}(B \cap C) - \mathbb{P}(A \cap C) \\ & + \mathbb{P}(A \cap B \cap C).\end{aligned}$$

2.2 Kombinatoryka

DEFINICJA. Permutacja zbioru $\{a_1, a_2, \ldots, a_n\}$ jest to *uporządkowanie* elementów tego zbioru.

Przykład. Permutacje zbioru $\{1, 2, 3\}$:

123	213	312
132	231	321

TWIERDZENIE. *Liczba permutacji zbioru n-elementowego jest równa*

$$n! = 1 \cdot 2 \cdot 3 \cdots n.$$

DEFINICJA. Wariacja z powtórzeniami k-wyrazowa ze zbioru $\{a_1, a_2, \ldots, a_n\}$ jest to *ciąg* (uporządkowany) k elementów tego zbioru (niekoniecznie różnych).

Przykład. Wariacje 2-wyrazowe z powtórzeniami ze zbioru $\{1, 2, 3\}$:

11	12	13
21	22	23
31	32	33

TWIERDZENIE. *Liczba k-wyrazowych wariacji z powtórzeniami ze zbioru n-elementowego jest równa*

$$n^k.$$

DEFINICJA. Wariacja bez powtórzeń k-wyrazowa ze zbioru $\{a_1, a_2, \ldots, a_n\}$ jest to *ciąg* (uporządkowany) k *różnych* elementów tego zbioru.

Przykład. Wariacje 2-wyrazowe bez powtórzeń ze zbioru $\{1, 2, 3\}$:

12	13
21	23
31	32

TWIERDZENIE. *Liczba k-wyrazowych wariacji bez powtórzeń ze zbioru n-elementowego jest równa*

$$n \cdot (n-1) \cdot (n-2) \cdot \cdots \cdot (n-k+1) = \frac{n!}{(n-k)!}.$$

DEFINICJA. Kombinacja k-elementowa ze zbioru $\{a_1, a_2, \ldots, a_n\}$ jest to *podzbiór* (nieuporządkowany) złożony z k różnych elementów tego zbioru.

Uwaga. Mówimy tu tylko o kombinacjach *bez powtórzeń.*

Przykład. Kombinacje 2-elementowe ze zbioru $\{1, 2, 3\}$:

$$\{1,2\} \quad \{1,3\} \quad \{2,3\}.$$

TWIERDZENIE. *Liczba kombinacji k-elementowych ze zbioru n-elementowego jest równa*

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

$$= \frac{n \cdot (n-1) \cdot (n-2) \cdots (n-k+1)}{k!}$$

2.3 Prawdopodobieństwo warunkowe i niezależność

Oceniamy szanse zajścia zdarzenia losowego A *wiedząc, że zaszło* zdarzenie losowe B. Mówimy wtedy o *prawdopodobieństwie warunkowym* i używamy oznaczenia $\mathbb{P}(A|B)$.

DEFINICJA. Prawdopodobieństwo warunkowe zdarzenia A pod warunkiem zajścia zdarzenia B jest określone wzorem

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Zakładamy przy tym, że $\mathbb{P}(B) > 0$.

Dodatkowa informacja, którą mamy, ,,zawęża przestrzeń probabilistyczną do zbioru B".

Przykład. (Losowe rendez-vous) Dwie osoby umówiły się na spotkanie. Każda z nich przychodzi "w chwili losowo wybranej pomiędzy godziną 0:00 i 1:00, niezależnie od drugiej osoby". Wynikiem doświadczenia jest para

$$\omega = (x, y),$$

gdzie

x – czas przyjścia Pani;

y – czas przyjścia Pana.

Przestrzeń probabilistyczna: kwadrat

$$\Omega = \{(x, y) : 0 < x < 1, 0 < y < 1]\},$$

Równomierny rozkład prawdopodobieństwa:

$$\mathbb{P}(A) = |A|/|\Omega| = |A|,$$

gdzie $|A|$ oznacza pole zbioru płaskiego A.

- A – zdarzenie „Pan czekał na Panią”: $A = \{(x, y) :\ x > y\}$.
- B – zdarzenie „ Pani przyszła po 0:30”: $B = \{(x, y) :\ x > 1/2\}$.

Mamy $\mathbb{P}(A) = |A|/|\Omega| = 1/2$
ale $\mathbb{P}(A|B) = |A \cap B|/|B| = (3/8)/(1/2) = 3/4$.

TWIERDZENIE. (Wzór łańcuchowy) *Z definicji wynika, że*

$$\mathbb{P}(A \cap B) = \mathbb{P}(B)\mathbb{P}(A|B).$$

Dla trzech zdarzeń, powiedzmy :

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_1)\mathbb{P}(A_2 \mid A_1)\mathbb{P}(A_3 \mid A_1 \cap A_2).$$

Posługiwanie się wzorem łańcuchowym w życiu codziennym:

- prawdopodobieństwo, że pojadę w tym roku do Australii oceniam na 0.5;
- jeśli będę w Australii, to z prawdopodobieństwem 0.001 mogę zostać zaatakowany przez rekina;
- jeśli zaatakuje mnie rekin, to mnie zje z prawdopodobieństwem 0.8.

⋆ Zatem prawdopodobieństwo, że *w Australii zje mnie rekin*

$$= 0.5 \times 0.001 \times 0.8 = 0.0004.$$

TWIERDZENIE. (Wzór na prawdopodobieństwo całkowite) *Załóżmy, że zdarzenia* $C_1, C_2, \ldots, C_k$ *spełniają następujące warunki:* $C_i \cap C_j = \emptyset$, *gdy* $i \neq j$, $C_1 \cup C_2 \cup \cdots \cup C_k = \Omega$ *i* $\mathbb{P}(C_i) > 0$ *dla* $i = 1, 2, \ldots, k$. *Wtedy dla dowolnego zdarzenia* B,

$$\mathbb{P}(B) = \sum_{i=1}^{k} \mathbb{P}(B \mid C_i)\mathbb{P}(C_i).$$

Dowód. Mamy

$$\begin{aligned}
\mathbb{P}(B) &= \mathbb{P}((B \cap C_1) \cup (B \cap C_2) \cup \cdots (B \cap C_k) \\
&= \sum_i \mathbb{P}(B \cap C_i) \\
&= \sum_i \mathbb{P}(B \mid C_i)\mathbb{P}(C_i),
\end{aligned}$$

co było do okazania. ■

TWIERDZENIE. (Wzór Bayesa) *Przy założeniach wzoru na prawdopodobieństwo całkowite, jeśli ponadto* $\mathbb{P}(B) > 0$, *to*

$$\mathbb{P}(C_i \mid B) = \frac{\mathbb{P}(C_i \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B \mid C_i)(\mathbb{P}(C_i)}{\sum_k \mathbb{P}(B \mid C_k)\mathbb{P}(C_k)}.$$

Dowód. Z definicji,

$$\mathbb{P}(C_i \mid B) = \frac{\mathbb{P}(C_i \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|C_i)\mathbb{P}(C_i)}{\mathbb{P}(B)}.$$

Mianownik obliczyć można ze wzoru na prawdopodobieństwo całkowite. ■

Przykład. (3 Fabryki) Trzy fabryki produkują ten sam produkt. Każda z nich wypuszcza z pewnym prawdopodobieństwem wyroby wadliwe:

1. fabryka – $100p_1\%$ wadliwych wyrobów,
2. fabryka – $100p_2\%$,
3. fabryka – $100p_3\%$.

W partii towaru znajduje się

1. n_1 sztuk towaru z fabryki 1,
2. n_2 sztuk towaru z fabryki 2,
3. n_3 sztuk towaru z fabryki 3.

Losowo wybieramy jedną sztukę towaru. Jakie jest prawdopodobieństwo, że wylosujemy produkt wadliwy? Przy oczywistych oznaczeniach mamy $\mathbb{P}(W|F_i) = p_i$ i $\mathbb{P}(F_i) = n_i/n$ dla $i = 1, 2, 3$ i $n = n_1+n_2+n_3$. Wzór na prawdopodobieństwo całkowite daje

$$\mathbb{P}(W) = p_1 n_1/n + p_2 n_2/n + p_3 n_3/n.$$

Wybraliśmy losowo sztukę towaru i widzimy, że jest wadliwa. Jakie jest prawdopodobieństwo, że pochodzi z pierwszej fabryki? Ze wzoru Bayesa mamy

$$\mathbb{P}(F_1|W) = \frac{\mathbb{P}(W|F_1)\mathbb{P}(F_1)}{\mathbb{P}(W)} = \frac{p_1 n_1}{p_1 n_1 + p_2 n_2 + p_3 n_3}.$$

Typowe zastosowanie wzoru Bayesa dotyczy sytuacji, gdy zajście (lub niezajście) zdarzenia B jest ,,widoczne", podczas gdy zdarzenia C_i ,,mają wpływ" na zdarzenie A, ale same są trudne lub niemożliwe do zaobserwowania.

DEFINICJA. Zdarzenia A i B są **niezależne**, jeśli

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

Zauważmy, że definicja ta traktuje A i B w sposób symetryczny. Jeśli $\mathbb{P}(B) > 0$ to zdarzenia A i B są niezależne wtedy i tylko wtedy, gdy

$$\mathbb{P}(A \mid B) = \mathbb{P}(A),$$

czyli informacja o zajściu zdarzenia B *nie wpływa na ocenę szansy zajścia zdarzenia* A.

Przykład. (Kolor i wysokość karty) Losujemy jedną kartę z talii. Przyjmijmy, że każda karta, czyli element zbioru $\Omega = \{1, 2, ..., 52\}$ jest jednakowo prawdopodobna (bardzo dobrze potasowaliśmy talię). Niech F oznacza, że wyciągniemy figurę a T, że wyciągniemy trefla. Te zdarzenia są niezależne. Istotnie, $\mathbb{P}(T) = 13/52$, $\mathbb{P}(F) = 16/52$, $\mathbb{P}(T \cap F) = 4/52$, więc $\mathbb{P}(T)\mathbb{P}(F) = \mathbb{P}(T \cap F)$. Zauważmy, że $\mathbb{P}(F|T) = 4/13 = 16/52 = \mathbb{P}(F)$.

Wyobraźmy teraz sobie, że talia zawiera tylko 51 kart, bo zgubiliśmy dwójkę karo. Czy zdarzenia T (trefl) i F (figura) są niezależne? Nie, bo teraz $\mathbb{P}(F|T) = 4/13 < 16/51 = \mathbb{P}(F)$. Informacja o tym, że wyciągnięta karta jest treflem *zmniejsza* szansę tego, że jest ona figurą.

Jeżeli zdarzenia $A_1, A_2, \dots, A_k$ są niezależne, to

$$\mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_k) = \mathbb{P}(A_1)\mathbb{P}(A_2)\dots\mathbb{P}(A_k).$$

i

$$\mathbb{P}(A_1 \cup A_2 \cup \dots \cup A_k) = 1 - \big(1 - \mathbb{P}(A_1)\big)\big(1 - \mathbb{P}(A_2)\big)\dots\big(1 - \mathbb{P}(A_k)\big).$$

Schemat Bernoulli'ego.

Rozważmy n *niezależnych powtórzeń* takiego samego doświadczenia losowego, które ma dwa możliwe wyniki. Przyjęło się mówić, że mamy

n ,,prób", z których każda kończy się

- ,,sukcesem" – z prawdopodobieństwem p

  lub

- ,,porażką" – z prawdopodobieństwem $1 - p = q$.

Przykład. (30 rzutów kością) Przyjmijmy, że ,,sukcesem" jest wyrzucenie ,,szóstki", zaś otrzymanie każdej innej liczby oczek jest ,,porażką". Mamy $n = 30$ prób Bernoulli'ego z prawdopodobieństwem sukcesu $p = 1/6$ i prawdopodobieństwem porażki $q = 5/6$.

TWIERDZENIE. *Rozważmy n prób Bernoulli'ego z prawdopodobieństwem sukcesu p w pojedynczej próbie. Prawdopodobieństwo tego, że uzyskamy* dokładnie *k sukcesów (a zatem $n - k$ porażek) jest równe*

$$\binom{n}{k} p^k (1-p)^{n-k}.$$

Dowód. (Dla 5 prób Bernoulli'ego) Za wynik doświadczenia, powtórzonego 5 razy, możemy uznać ciąg 5 liter B (sukces) lub C (porażka). Prawdopodobieństwo otrzymania sekwencji

$$BBBCC$$

jest równe

$$pppqq = p^3 q^2,$$

bo próby są niezależne. Takie samo jest prawdopodobieństwo otrzymania innej sekwencji, zawierającej 3 sukcesy i 2 porażki, np. $BCBCB$. Wszystkich takich sekwencji jest

$$\binom{5}{3} = 10.$$

Istotnie, można ustawić 3 litery B wybierając 3 miejsca spośród 5 miejsc w ciągu, a na pozostałych miejscach ustawić litery C. Oto wszystkie takie sekwencje:

$$\begin{array}{ccccc} BBBCC & BBCBC & BBCCB & BCBBC & BCBCB \\ BCCBB & CBBBC & CBBCB & CBCBB & CCBBB \end{array}$$

Zatem prawdopodobieństwo uzyskania 3 sukcesów i 2 porażek jest równe

$$\binom{5}{3} p^3 q^2.$$

W przypadku ogólnym rozumowanie jest identyczne. ■

Przykład. (30 rzutów kostką) Prawdopodobieństwo uzyskania w 30 rzutach równo 5 razy ,,szóstki" jest równe

$$\binom{30}{5}\left(\frac{1}{6}\right)^5\left(\frac{5}{6}\right)^{25} \approx 0.1921$$

($n = 30$, $k = 5$, $p = 1/6$, $q = 5/6$).

Losowanie próbki ze zwracaniem

Wybieramy losowo n-krotnie po 1 elemencie z populacji liczącej r elementów, przy czym możemy wielokrotnie wylosować ten sam element (losujemy n kul z urny zawierającej r kul, zwracając za każdym razem wylosowaną kulę do urny). Przyjmijmy, że wynikiem jest *uporządkowany* układ n kul, czyli *wariacja z powtórzeniami* n z r. Przestrzeń probabilistyczna Ω zawiera

$$r^n$$

jednakowo prawdopodobnych wyników.

Zauważmy, że w losowaniu *ze zwracaniem* powtarzamy n-krotnie, niezależnie to samo doświadczenie losowe: wyciągnięcie 1 kuli z urny. Jest tak, bo poprzednie losowania nie zmieniają składu urny.

Powiedzmy, że wyróżniamy pewne elementy populacji. Wyobraźmy sobie wyróżnione elementy jako kule ,,białe", a pozostałe jako kule ,,czarne". Jeśli wyciągnięcie kuli białej uznamy za sukces, a kuli czarnej – za porażkę, to mamy *schemat Bernoulli'ego.* Prawdopodobieństwo sukcesu jest tu frakcją kul białych w urnie: jeśli urna zawiera m kul białych i $r - m$ czarnych, to $p = m/r$.

TWIERDZENIE. *Jeśli z urny zawierającej m kul białych i $r - m$ czarnych losujemy* ze zwracaniem *n razy po jednej kuli, to prawdopodobieństwo wylosowania dokładnie k razy kuli białej (i $n - k$ razy czarnej) jest równe*

$$\binom{n}{k}\frac{m^k(r-m)^{n-k}}{r^n}.$$

Losowanie próbki bez zwracania

Wybieramy losowo n różnych elementów z populacji liczącej r elementów (losujemy n kul z urny zawierającej r kul, nie zwracając uprzednio wylosowanych kul do urny).

Gdy interesuje nas kolejność wyciąganych kul, to za wynik przyjmiemy *uporządkowany* układ n kul, czyli *wariację bez powtórzeń* n z r. Przestrzeń probabilistyczna Ω zawiera

$$r(r-1)\cdots(r-n+1)$$

jednakowo prawdopodobnych wyników.

Gdy nie interesuje nas kolejność wyciąganych kul, to za wynik doświadczenia można równie dobrze przyjąć *nieuporządkowany* układ n kul, czyli *kombinację* n z r. Mamy wtedy przestrzeń probabilistyczną Ω, która zawiera

$$\binom{r}{n}$$

jednakowo prawdopodobnych wyników.

Podobnie jak poprzednio wyróżnimy pewne elementy populacji. Wyobraźmy sobie wyróżnione elementy jako kule ,,białe", a pozostałe jako kule ,,czarne".

Przykład. (losowanie 2 kul) Urna zawiera m kul białych i $r - m$-czarnych.

Losujemy bez zwracania 2 kule, notując ich kolejność. Rozważmy zdarzenia

- B_1 – ,,kula biała w pierwszym ciągnieniu";
- B_2 – ,,kula biała w drugim ciągnieniu";
- C_1 – ,,kula czarna w pierwszym ciągnieniu";
- C_2 – ,,kula czarna w drugim ciągnieniu".

Mamy ze wzoru łańcuchowego

$$\begin{aligned}\mathbb{P}(B_1 \cap B_2) &= \mathbb{P}(B_1)\mathbb{P}(B_2|B_1)\\ &= \frac{m}{r} \cdot \frac{m-1}{r-1}.\end{aligned}$$

Ze wzoru na prawdopodobieństwo całkowite

$$\begin{aligned}\mathbb{P}(B_2) &= \mathbb{P}(B_1)\mathbb{P}(B_2|B_1) + \mathbb{P}(C_1)\mathbb{P}(B_2|C_1)\\ &= \frac{m}{r} \cdot \frac{m-1}{r-1} + \frac{m-r}{r} \cdot \frac{m}{r-1}\\ &= \frac{m}{r}.\end{aligned}$$

Zauważmy, że $\mathbb{P}(B_2) = \mathbb{P}(B_1) = m/r$, ale zdarzenia B_1 i B_2 *nie są niezależne*, bo

$$\mathbb{P}(B_2|B_1) < \mathbb{P}(B_2).$$

Dla losowania ze zwracaniem mamy również $\mathbb{P}(B_2) = \mathbb{P}(B_1) = m/r$, ale przy tym

$$\begin{aligned}\mathbb{P}(B_1 \cap B_2) &= \mathbb{P}(B_1)\mathbb{P}(B_2)\\ &= \left(\frac{m}{r}\right)^2,\end{aligned}$$

czyli zdarzenia B_1 i B_2 są niezależne.

Losowanie bez zwracania *nie jest* schematem Bernoulli'ego, bo kolejne ciągnienia nie są niezależne.

TWIERDZENIE. *Jeśli z urny zawierającej m kul białych i $r-m$ czarnych losujemy* bez zwracania *n kul, to prawdopodobieństwo wylosowania dokładnie k kul białych (i $n-k$ czarnych) jest równe*

$$\frac{\binom{m}{k}\binom{r-m}{n-k}}{\binom{r}{n}}$$

(zakładając, że $0 \leqslant k \leqslant n$, $k \leqslant r$ i $n-k \leqslant r-m$).

Przykład. (MULTI-LOTEK) Gra polega na wytypowaniu 10 numerów spośród 80. Organizator gry losuje potem 20 numerów spośród 80. Prawdopodobieństwo tego, że wśród wylosowanych 20 numerów znajdą się wszystkie 10 wytypowane przez nas jest równe

$$\frac{\binom{10}{10}\binom{70}{10}}{\binom{80}{20}} \approx \frac{1}{8911711.2}$$

($r=80$, $m=10$, $n=20$, $k=10$). Trafienie „dziesiątki" zdarza się raz na 9 milionów gier. Organizator wypłaca wtedy w nagrodę 100000-krotność stawki.

2.4 Zmienne losowe

DEFINICJA. Zmienną losową nazywamy funkcję o wartościach rzeczywistych, określoną na przestrzeni probabilistycznej:

$$X : \Omega \to \mathbb{R}.$$

X interpretujemy jako „zmienną, której wartość zależy od wyniku doświadczenia losowego":

$$\text{wynikowi } \omega \in \Omega \text{ odpowiada liczba } X(\omega) \in \mathbb{R}.$$

Zdarzenia losowe "dotyczące zmiennej losowej X" jak zwykle, utożsamiamy z podzbiorami przestrzeni probabilistycznej. Np. jeśli a jest liczbą, to

zdarzenie „$X = a$" jest zbiorem $\{\omega \in \Omega : X(\omega) = a\}$,

zdarzenie „$X \leqslant a$" jest zbiorem $\{\omega \in \Omega : X(\omega) \leqslant a\}$,

DEFINICJA. Załóżmy, że zmienna losowa X przyjmuje wartości ze zbioru $\{x_1, x_2, \ldots, x_n\} \subset \mathbb{R}$. **Rozkład prawdopodobieństwa zmiennej losowej** X możemy przedstawić przy pomocy tabelki

wartość	x_1	x_2	$\ldots$	x_n
prawdopodobieństwo	f_1	f_2	$\ldots$	f_n

gdzie

$$f_i = \mathbb{P}(X = x_i).$$

Przykład. (Liczba orłów w 2 rzutach monetą) Oznaczmy liczbę orłów przez S. Wartości zmiennej losowej S i rozkład prawdopodobieństwa *na przestrzeni probabilistycznej* przedstawia tabelka:

wynik	OO	OR	RO	RR
prawdopodobieństwo	1/4	1/4	1/4	1/4
wartość	2	1	1	0

Rozkład prawdopodobieństwa *zmiennej losowej* S przedstawia tabelka:

wartość	0	1	2
prawdopodobieństwo	1/4	1/2	1/4

Na przykład

$$\mathbb{P}(S = 1) = \mathbb{P}(\{OR, RO\}) = 1/2.$$

DEFINICJA. Mówimy, że zmienna S ma rozkład **dwumianowy** (inaczej rozkład *Bernoulli'ego*) z parametrami n i p, jeśli

$$\mathbb{P}(S = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Będziemy wtedy pisali $S \sim \mathrm{Bin}(n, p)$.

Oczywiście, liczba sukcesów w schemacie Bernoulli'ego jest zmienną losową o rozkładzie dwumianowym.

DEFINICJA. Dystrybuanta rozkładu zmiennej losowej X jest to funkcja $F : \mathbb{R} \to [0, 1]$ określona wzorem

$$F(x) = \mathbb{P}(X \leqslant x).$$

Jeśli wartości zmiennej losowej X jest skończenie wiele i są one uporządkowane:

$$x_1 < x_2 < \cdots < x_n,$$

to możemy umieścić w tabelce także ,,skumulowane prawdopodobieństwa:"

wartość	x_1	x_2	$\ldots$	x_n
prawdopodobieństwo	f_1	f_2	$\ldots$	f_n
pr-stwo skumulowane	c_1	c_2	$\ldots$	c_n

gdzie $c_i = \mathbb{P}(X \leqslant x_i)$ i $f_i = \mathbb{P}(X = x_i)$. Mamy zatem

$$\begin{aligned} &c_1 = f_1; \\ &c_2 = f_1 + f_2; \\ &\ldots \\ &c_n = f_1 + f_2 + \cdots + f_n. \end{aligned}$$

Dystrybuanta jest wygodnym graficznym przedstawieniem tabelki ,,skumulowanych" prawdopodobieństw.

Przykład. (Kontynuacja) Niech S oznacza liczbę orłów w 2 rzutach monetą. Zestawmy razem tabelkę rozkładu prawdopodobieństwa i tabelkę prawdopodobieństwskumulowanych:

wartość	0	1	2
prawdopodobieństwo	1/4	1/2	1/4
pr-stwo skumulowane	1/4	3/4	1

Dystrybuanta jest dana wzorem

$$F(s) = \mathbb{P}(S \leqslant s) = \begin{cases} 0 & \text{dla } s < 0; \\ 1/4 & \text{dla } 0 \leqslant s < 1; \\ 3/4 & \text{dla } 1 \leqslant s < 2; \\ 1 & \text{dla } 2 \leqslant s. \end{cases}$$

Wykres tej funkcji ma postać ,,schodków".

Własności dystrybuanty.

Jeśli F jest dystrybuantą zmiennej losowej X, to

F jest funkcją niemalejącą;

F jest funkcją prawostronnie ciągłą;

$\lim_{x\to-\infty} F(x) = 0$ i $\lim_{x\to+\infty} F(x) = 1$.

Mamy przy tym:

$$\mathbb{P}(a < X \leqslant b) = F(b) - F(a);$$
$$\mathbb{P}(X > b) = 1 - F(b);$$

Dystrybuanta F zmiennej losowej przyjmującej wartości $x_1 < x_2 < \cdots < x_n$ jest funkcją ,,schodkową", to znaczy jest stała na każdym przedziale $[x_k, x_{k+1})$. W punktach $x_1, x_2, \ldots, x_n$ jest nieciągła i $\mathbb{P}(X = x_{k+1}) = F(x_{k+1}) - F(x_k)$. Dla $x < x_1$ mamy $F(x) = 0$, zaś dla $x \geqslant x_n$ mamy $F(x) = 1$.

Zmienne losowe ciągłe.

Dystrybuanta zmiennej losowej może być funkcją ciągłą.

<u>*Przykład*</u>. (Rzucanie do tarczy) Rozważaliśmy przestrzeń probabilistyczną Ω, która jest kołem o promieniu b i przyjęliśmy, że $\mathbb{P}(A) = |A|/|\Omega|$. Niech R będzie odległością punktu trafienia od środka tarczy. Dystrybuanta zmiennej losowej R jest dana wzorem

$$F(r) = \mathbb{P}(R \leqslant r) = \begin{cases} 0 & \text{dla } r < 0; \\ r^2/b^2 & \text{dla } 0 \leqslant r \leqslant b; \\ 1 & \text{dla } r > b. \end{cases}$$

Istotnie, $\mathbb{P}(R \leqslant r) = (\pi r^2)/(\pi b^2) = r^2/b^2$, bo zdarzenie ,,$R \leqslant r$" jest kołem o promieniu r (dla $0 \leqslant r \leqslant b$).

Dwa podstawowe typy zmiennych losowych:

- Jeśli zmienna losowa X przyjmuje skończoną liczbę wartości $\{x_1, x_2, \ldots, x_n\}$ lub nieskończenie wiele wartości, które można ustawić w ciąg $\{x_1, x_2, \ldots\}$, to mówimy, że jest to zmienna typu **dyskretnego**.

Dla zmiennej *dyskretnej*, rozkład prawdopodobieństwa możemy opisać przy pomocy tabelki lub dystrybuanty.

- Jeśli zmienna losowa X może przyjmować każdą z wartości należących do pewnego przedziału liczbowego i jej dystrybuanta jest funkcją ciągłą, to mówimy, że jest to zmienna typu **ciągłego**.

Dla zmiennej *ciągłej*, rozkładu prawdopodobieństwa *nie możemy* opisać przy pomocy tabelki. Możemy opisać rozkład używając dystrybuanty. Zwykle rozkład zmiennej ciągłej można opisać podając *gęstość* prawdopodobieństwa.

DEFINICJA. Funkcja $f : \mathbb{R} \to \mathbb{R}$ nazywa się **gęstością rozkładu prawdopodobieństwa zmiennej losowej** X, jeśli dla dowolnych $a < b$ mamy

$$\mathbb{P}(a < X < b) = \mathbb{P}(a \leqslant X \leqslant b) = \int_a^b f(x)\mathrm{d}x.$$

Interpretacja: jeśli gęstość jest ciągła w punkcie x, to

$$f(x) = \lim_{h \to 0} \frac{\mathbb{P}(x \leqslant X \leqslant x+h)}{h}.$$

Jeśli f jest gęstością, to dystrybuanta jest funkcją ciągłą i jest dana wzorem

$$F(a) = \int_{-\infty}^a f(x)\mathrm{d}x.$$

Odwrotnie,

$$f(x) = F'(x) = \frac{\mathrm{d}}{\mathrm{d}x}F(x),$$

(przynajmniej w punktach x, w których f jest ciągła).

Własności gęstości prawdopodobieństwa.

Jeśli funkcja f jest gęstością rozkładu prawdopodobieństwa, to

$$f(x) \geqslant 0; \qquad \int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1.$$

Przykład. (Rzucanie do tarczy) Zmienna losowa R ma gęstość

$$f(r) = \begin{cases} 0 & \text{dla } r < 0; \\ 2r/b^2 & \text{dla } 0 \leqslant r \leqslant b; \\ 0 & \text{dla } r > b. \end{cases}$$

Istotnie, dla $0 < r < b$ mamy $f(r) = F'(r) = (r^2/b^2)' = 2r/b^2$.

DEFINICJA. Zmienna losowa X ma **rozkład jednostajny** na przedziale $[a, b]$, jeśli ma gęstość

$$f(x) = \begin{cases} 1/(b-a) & \text{dla } a \leqslant x \leqslant b; \\ 0 & \text{w przeciwnym przypadku.} \end{cases}$$

Obrazowo mówiąc, wartość zmiennej losowej X jest "punktem losowo wybranym z odcinka $[a, b]$". Gęstość prawdopodobieństwa rozkładu jednostajnego jest funkcją stałą na tym przedziale. Zauważmy, że dla $a \leqslant x \leqslant y \leqslant b$ mamy

$$\mathbb{P}(x < X < y) = \frac{y - x}{b - a}.$$

Oznaczenie: $X \sim \mathrm{U}(a, b)$.

DEFINICJA. Zmienna losowa X ma **rozkład normalny** (inaczej: *rozkład Gaussa*) z parametrami μ i σ^2, jeśli ma gęstość

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right].$$

Oznaczenie: $X \sim \mathrm{N}(\mu, \sigma^2)$.

W szczególnym przypadku, jeśli $Z \sim \mathrm{N}(0, 1)$ to mówimy, że Z ma *standardowy* rozkład normalny.

Gęstość rozkładu $\mathrm{N}(0, 1)$:

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{z^2}{2}\right].$$

Dystrybuanta rozkładu $\mathrm{N}(0, 1)$: jest funkcja

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{t^2}{2}\right] \mathrm{d}t.$$

Wartości dystrybuanty Φ standardowego rozkładu normalnego $N(0, 1)$ podaje następująca tablica:

z	0.0	0.5	1.0	1.5	2.0	2.33
$\Phi(z)$	0.50	0.69	0.84	0.93	0.98	0.99

Symetria funkcji Φ:

$$\Phi(-z) = 1 - \Phi(z).$$

Przykład.

$$\Phi(-1.5) = 1 - \Phi(1.5) = 1 - 0.93 = 0.07.$$

Standaryzacja zmiennej o rozkładzie normalnym:

- jeśli $X \sim \mathrm{N}(\mu, \sigma^2)$ to $Z = \frac{X-\mu}{\sigma} \sim \mathrm{N}(0, 1)$;
- jeśli $Z \sim \mathrm{N}(0, 1)$ to $X = \sigma Z + \mu \sim \mathrm{N}(\mu, \sigma^2)$;

Korzystając z tablic funkcji Φ możemy obliczać prawdopodobieństwa zdarzeń dotyczących zmiennej losowej o *dowolnym* rozkładzie normalnym $\mathrm{N}(\mu, \sigma^2)$.

Przykład. Jeśli $X \sim \mathrm{N}(\mu, \sigma^2)$, to

$$
\begin{aligned}
& \mathbb{P}(\mu - 1.5\sigma \leqslant X \leqslant \mu + 1.5\sigma) \\
= \; & \mathbb{P}(-1.5 \leqslant \frac{X-\mu}{\sigma} \leqslant 1.5) \\
= \; & \mathbb{P}(-1.5 \leqslant Z \leqslant 1.5) \\
= \; & \Phi(1.5) - \Phi(-1.5) \\
= \; & 2\Phi(1.5) - 1 = 0.86.
\end{aligned}
$$

DEFINICJA. Niech X będzie dyskretną zmienną losową o wartościach $x_1, x_2, \ldots, x_n$. **Wartością przeciętną** zmiennej losowej X nazywamy liczbę

$$\mathbb{E}(X) = \sum_{i=1}^{n} x_i f_i,$$

gdzie $f_i = \mathbb{P}(X = x_i)$.

Inaczej nazywamy $\mathbb{E}(X)$ *wartością oczekiwaną* lub po prostu *średnią*.

Przykład. Jeśli rozkład zmiennej losowej X jest dany tabelką

x_i	-5	2	5	10
f_i	0.2	0.5	0.2	0.1

to

$$\mathbb{E}(X) = (-5) \times 0.2 + 2 \times 0.5 + 5 \times 0.2 + 10 \times 0.1 = 2.$$

DEFINICJA. Jeśli X jest zmienną losową o wartości przeciętnej $\mathbb{E}(X) = \mu$, to **wariancją** tej zmiennej losowej nazywamy liczbę

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

Odchyleniem standardowym nazywamy liczbę

$$\mathrm{D}(X) = \sqrt{\mathrm{Var}(X)}.$$

Interpretacja: wariancja i odchylenie standardowe są miarami *rozrzutu* zmiennej losowej wokół jej wartości przeciętnej.

Czasem oznacza się wariancję innym symbolem:

$$\mathrm{Var}(X) = \mathrm{D}^2(X).$$

Inny wzór na obliczanie wariancji:

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - \mu^2.$$

Jeśli X jest dyskretną zmienną losową, która przyjmuje wartości $x_1, x_2, \ldots, x_n$, to

$$\begin{aligned}\mathrm{Var}(X) &= \sum_{i=1}^{n} (x_i - \mu)^2 f_i \\ &= \sum_{i=1}^{n} x_i^2 f_i - \mu^2,\end{aligned}$$

gdzie $f_i = \mathbb{P}(X = x_i)$ i $\mu = \mathbb{E}(X)$.

Przykład. (Kontynuacja) Dla zmiennej losowej X z poprzedniego przykładu, $\mu = 2$. Mamy więc

$$\begin{aligned}\mathrm{Var}(X) =& (-5-2)^2 \times 0.2 + (2-2)^2 \times 0.5 \\ &+ (5-2)^2 \times 0.2 + (10-2)^2 \times 0.1 = 18.\end{aligned}$$

Inaczej,

$$\begin{aligned}\mathrm{Var}(X) =& (-5)^2 \times 0.2 + 2^2 \times 0.5 \\ &+ 5^2 \times 0.2 + 10^2 \times 0.1 - 2^2 = 22 - 4 = 18.\end{aligned}$$

Oczywiście, $\mathrm{D}(X) = \sqrt{18} = 4.243$.

Uwaga. Wartość przeciętną i wariancję można określić również dla zmiennej losowej o rozkładzie ciągłym. Jeśli F jest *gęstością* rozkładu zmiennej losowej X, to

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} x f(x) \mathrm{d}x.$$

Własności wartości przeciętnej i wariancji.

Niech a będzie liczbą, zaś X i Y – zmiennymi losowymi.

- $\mathbb{E}(X + a) = \mathbb{E}(X) + a$;
- $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$;
- $\mathbb{E}(aX) = a\mathbb{E}(X)$;
- $\mathrm{Var}(X + a) = \mathrm{Var}(X - a) = \mathrm{Var}(X)$;
- $\mathrm{Var}(aX) = a^2\mathrm{Var}(X)$;
- $\mathrm{D}(aX) = |a|\mathrm{D}(X)$;
- jeśli zmienne X i Y są *niezależne* to
$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y).$$

Wartości przeciętne i wariancje typowych rozkładów prawdopodobieństwa.

- **Rozkład dwumianowy.** Jeśli $S \sim \mathrm{Bin}(n, p)$, to
$$\mathbb{E}(S) = np \text{ i } \mathrm{Var}(S) = np(1 - p).$$
- **Rozkład jednostajny.** Jeśli $X \sim \mathrm{U}(a, b)$, to
$$\mathbb{E}(X) = (a + b)/2 \text{ i } \mathrm{Var}(X) = (b - a)^2/12.$$
- **Rozkład normalny.** Jeśli $X \sim \mathrm{N}(\mu, \sigma^2)$, to
$$\mathbb{E}(X) = \mu \text{ i } \mathrm{Var}(X) = \sigma^2.$$

  Zmienna o rozkładzie normalnym $\mathrm{N}(\mu, \sigma^2)$ ma *wartość przeciętną μ, wariancję σ^2 i odchylenie standardowe σ.*

Przykład. (Liczba orłów w wielu rzutach monetą) Niech S_{100} oznacza liczbę orłów w 100 rzutach monetą. Mamy tu do czynienia ze schematem Bernoulli'ego (za sukces uznajmy wyrzucenie orła, $p = 1/2$, $n = 100$), a więc S_{100} ma rozkład dwumianowy: $S_{100} \sim \mathrm{Bin}(100, 1/2)$. Zatem

$$\begin{aligned}
&\mathbb{E}(S_{100}) = 100 \cdot \frac{1}{2} = 50,\\
&\mathrm{Var}(S_{100}) = 100 \cdot \frac{1}{2} \cdot \frac{1}{2} = 25,\\
&\mathrm{D}(S_{100}) = \sqrt{25} = 5.
\end{aligned}$$

Uwaga. Można pokazać, że jeśli zmienne losowe mają rozkłady normalne *i są niezależne*, to ich suma też ma rozkład normalny. Jeśli $X_1 \sim \mathrm{N}(\mu_1, \sigma_1^2)$ i $X_2 \sim \mathrm{N}(\mu_2, \sigma_2^2)$, to

$$X_1 + X_2 = \mathrm{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

Przykład. (Średnia niezależnych pomiarów) Wyniki pomiarów są zwykle obciążone błędami, które można uznać za losowe. Często przyjmuje się, że wynik pomiaru wielkości μ jest zmienną losową X o rozkładzie normalnym:

$$X \sim \mathrm{N}(\mu, \sigma^2).$$

Odchylenie standardowe σ interpretujemy jako ,,dokładność pomiaru" (średni błąd). Jeśli przeprowadzimy n niezależnych pomiarów tej samej wielkości z tą samą dokładnością, to mamy niezależne zmienne losowe $X_1, X_2, \ldots, X_n$ o jednakowym rozkładzie normalnym. Za oszacowanie mierzonej wielkości μ można przyjąć *średnią* pojedynczych pomiarów, czyli

$$\bar{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n).$$

Mamy

$$\begin{aligned}
\mathbb{E}(\bar{X}) &= \mathbb{E}\left[\frac{1}{n}(X_1 + X_2 + \cdots + X_n)\right] \\
&= \frac{1}{n}[\mathbb{E}(X_1) + \mathbb{E}(X_2) + \cdots + \mathbb{E}(X_n)] \\
&= \frac{1}{n}[\underbrace{\mu + \mu + \cdots + \mu}_{n}] \\
&= \frac{1}{n}n\mu = \mu,
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}(\bar{X}) &= \mathrm{Var}\left[\frac{1}{n}(X_1 + X_2 + \cdots + (X_n)\right] \\
&= \frac{1}{n^2}\mathrm{Var}\left[X_1 + X_2 + \cdots + X_n\right] \\
&= \frac{1}{n^2}[\mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \cdots + \mathrm{Var}(X_n)] \\
&= \frac{1}{n^2}[\underbrace{\sigma^2 + \sigma^2 + \cdots + \sigma^2}_{n}] \\
&= \frac{1}{n^2}n\sigma^2 = \frac{1}{n}\sigma^2.
\end{aligned}$$

Zatem

$$\bar{X} \sim \mathrm{N}(\mu, \sigma^2/n).$$

Uśredniony wynik n pomiarów ma odchylenie standardowe

$$\mathrm{D}\bar{X} = \frac{\sigma}{\sqrt{n}}.$$

Jeśli uśrednimy, powiedzmy 100 niezależnych pomiarów, to dokładność wyniku zwiększy się $\sqrt{100} = 10$ razy w porównaniu z dokładnością pojedynczego pomiaru.

2.5 Centralne Twierdzenie Graniczne

TWIERDZENIE. (Centralne Twierdzenie Graniczne) *Jeżeli $X_1, \ldots, X_n$ są niezależnymi zmiennymi losowymi o jednakowym rozkładzie prawdopodobieństwa o wartości oczekiwanej $\mathbb{E}X_i = \mu$ i wariancji $\mathrm{Var}X_i = \sigma^2$ oraz*

$$S_n = X_1 + \cdots + X_n,$$

to dla każdej liczby a,

$$\lim_{n\to\infty} \mathbb{P}\Big(\frac{S_n - n\mu}{\sqrt{n}\sigma} \leqslant a\Big) = \Phi(a),$$

gdzie Φ jest dystrybuantą standardowego rozkładu normalnego $\mathrm{N}(0,1)$.

Interpretacja: W praktyce najczęściej stosuje się Centralne Twierdzenie Graniczne (CTG) w następujący sposób.

> *Sumę dużej liczby niezależnych zmiennych losowych można w przybliżeniu traktować jak zmienną losową o rozkładzie normalnym.*

Jeśli n jest ,,dostatecznie duże" to,

$$\frac{S_n - n\mu}{\sqrt{n}\sigma} \approx Z \sim \mathrm{N}(0,1),$$

czyli

$$S_n \approx n\mu + \sqrt{n}\sigma Z \sim \mathrm{N}(n\mu, n\sigma^2),$$

czyli *w przybliżeniu*

$$S_n \sim \mathrm{N}\left(\mathbb{E}(S_n), \mathrm{Var}(S_n)\right).$$

Przykład. (Sumaryczna wysokość szkód) Przybliżenie normalne sumarycznej wysokości szkód jest powszechnie stosowane w matematyce ubezpieczeniowej. Niech S_n będzie sumaryczną wysokością szkód dla n polis, w pewnym ustalonym okresie czasu. Założymy, że $S = X_1 + \cdots + X_n$, gdzie każda ze zmiennych X_i opisuje wysokość szkód dla *pojedynczej* polisy (oczywiście, zmienne X_i mogą przyjmować wartość 0 ze sporym prawdopodobieństwem). Na przykład, załóżmy, że wysokość szkód dla pojedynczej polisy jest zmienną losową X o następującym rozkładzie prawdopodobieństwa:

wysokość szkody x (w tys. PLN)	0	1	4
prawdopodobieństwo $\mathbb{P}(X = x)$	0.75	0.20	0.05

Mamy

$$\begin{aligned}\mathbb{E}(X) &= 0.4,\\ \mathrm{Var}(X) &= 0.84\end{aligned}$$

Rozpatrzmy teraz portfel złożony z 2100 takich samych polis i załóżmy, że szkody dla poszczególnych polis są niezależnymi zmiennymi losowymi $X_1, X_2, \ldots, X_{2100}$, z których każda ma rozkład prawdopodobieństwa opisany powyżej. Zmienna losowa

$$S = X_1 + X_2 + \cdots + X_{2100}$$

jest sumaryczną wysokością szkód. Mamy

$$\begin{aligned}\mathbb{E}(S) &= 2100 \times 0.4 = 840,\\ \mathrm{Var}(S) &= 2100 \times 0.84 = 1764\\ \mathrm{D}(S) &= \sqrt{1764} = 42.\end{aligned}$$

Stosując przybliżenie rozkładem normalnym, czyli CTG, wnioskujemy, że

$$\begin{aligned}\mathbb{P}(S \leqslant 924) &= \mathbb{P}\left(\frac{S-840}{42} \leqslant \frac{924-840}{42}\right)\\ &= \mathbb{P}\left(\frac{S-840}{42} \leqslant 2\right)\\ &\approx \mathbb{P}\left(Z \leqslant 2\right)\\ &= \Phi(2) = 0.98\end{aligned}$$

(Z oznacza zmienną losową o rozkładzie $\mathrm{N}(0,1)$). Zatem z prawdopodobieństwem ok. 0.98 sumaryczne szkody nie przekroczą 924. Podkreślmy: suma S wielu niezależnych szkód ma w przybliżeniu rozkład normalny, nawet jeśli rozkład *pojedynczego* składnika wcale nie przypomina rozkładu normalnego.

W szczególności, CTG stosuje się do schematu Bernoulli'ego. Liczba sukcesów w n próbach jest sumą niezależnych zmiennych losowych:

$$S_n = X_1 + X_2 + \cdots + X_n,$$

gdzie

$$X_i = \begin{cases} 1 & \text{jeśli w } i\text{-tej próbie otrzymamy sukces;}\\ 0 & \text{jeśli w } i\text{-tej próbie otrzymamy porażkę.}\end{cases}$$

Oczywiście $\mathbb{P}(X_i = 1) = p$, $\mathbb{P}(X_i = 0) = 1 - p$,

$$\begin{aligned}
\mathbb{E}(X_i) &= 0 \cdot (1-p) + 1 \cdot p = p, \\
\mathrm{Var}(X_i) &= (0-p)^2(1-p) + (1-p)^2 p = p(1-p), \\
\mathbb{E}(S_n) &= \mathbb{E}(X_1) + \mathbb{E}(X_2) + \cdots + \mathbb{E}(X_n) \\
&= \underbrace{p + p + \cdots + p}_{n} = np, \\
\mathrm{Var}(S_n) &= \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \cdots + \mathrm{Var}(X_n) \\
&= \underbrace{p(1-p) + p(1-p) + \cdots + p(1-p)}_{n} = np(1-p).
\end{aligned}$$

TWIERDZENIE. (Twierdzenie de Moivre'a – Laplace'a). *Jeśli S_n jest liczbą sukcesów w n próbach Bernoulliego z prawdopodobieństwem pojedynczego sukcesu równym p, to dla każdej liczby a,*

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{S_n - np}{\sqrt{np(1-p)}} \leqslant a \right) = \Phi(a).$$

Przykład. (Fluktuacje losowe w rzutach monetą) Wykonujemy 1000 rzutów monetą. Jakie jest prawdopodobieństwo, że liczba orłów będzie zawarta pomiędzy 475 a 525? Wiemy, z grubsza, że liczba orłów jest ,,z dużym prawdopodobieństwem bliska 500". CTG pozwala zbadać rzecz dokładniej. Niech $S = S_{1000}$ będzie liczbą orłów. Mamy $\mathrm{D}S = \sqrt{np(1-p)} = \sqrt{250}$. Przybliżenie rozkładem normalnym jest uzasadnione ze względu na to, że $n = 1000$ jest duże.

$$\begin{aligned}
\mathbb{P}(475 \leqslant S \leqslant 525) &= \mathbb{P}(474 < S < 526) = \mathbb{P}(474.5 \leqslant S \leqslant 525.5) \\
&= \mathbb{P}(|S - 500| \leqslant 25.5) = \mathbb{P}\left(\frac{|S-500|}{\sqrt{250}} \leqslant \frac{25.5}{\sqrt{250}} \right) \\
&\approx \mathbb{P}(|Z| \leqslant 1.61) \\
&= 2\Phi(1.61) - 1 = 0.892.
\end{aligned}$$

Zastąpienie liczb 475 i 525 przez 474.5 i 525.5 jest tak zwaną ,,poprawką na ciągłość". Wobec tego, że aproksymujemy rozkład dyskretny (dwumianowy) rozkładem ciągłym (normalnym), ta poprawka poprawia nieco jakość przybliżenia.

Zauważmy, że *dokładne* wyrażenie

$$\mathbb{P}(475 \leqslant S \leqslant 525) = \sum_{i=475}^{525} \binom{1000}{i} \frac{1}{2^{1000}}$$

jest łatwe do napisania ale, oględnie mówiąc, mało wygodne. Obliczenia wykorzystujące CTG dają *przybliżone* wyniki, ale są szalenie proste.