Wprowadzenie do ilościowej oceny ryzyka

Łukasz Bocian
Zakład Epidemiologii i Oceny Ryzyka
PIWet-PIB

Puławy, 12-13.03.2018

Ocena ryzyka

jakościowa

(wynik w skali typu: *b. wysokie, wysokie, umiarkowane, niskie* itd..)

ilościowa

(wynik przedstawiony liczbowo)

podejście deterministyczne

- model matematyczny jednoznacznie na wejściu przypisuje danemu zdarzeniu konkretny stan.
- Opis modelu nie zawiera żadnego elementu losowości.
- Wszystkie parametry (elementy) modelu są znane lub założone

podejście stochastyczne (probabilistyczne)

- obejmuje przypisanie funkcji gęstości prawdopodobieństwa różnym składnikom mającym wpływ na ryzyko, a następnie wykonanie symulacji Monte Carlo lub innych obliczeń w celu oceny prawdopodobieństwa, że zaistnieje pewne zdarzenie. Wskazuje na możliwe zakresy wartości zmiennych w postaci rozkładów prawdopodobieństwa.

Modelowanie matematyczne

Przykładowe zastosowania modeli:

- **Przewidywanie bezpieczeństwa** - modele pozwalają np. na ocenę ryzyka wzrostu lub przeżycia patogenów w określonym czasie przechowywania;
- **Kontrola jakości** - mogą wskazać warunki w jakich możliwy jest rozwój lub przeżywanie patogenów, mogą pomagać w ustalaniu kryteriów dla krytycznych punktów kontrolnych oraz pozwalają na ocenę konsekwencji mikrobiologicznych przekroczenia kryteriów;
- **Opracowanie nowych produktów** - możliwe jest szybkie sprawdzenie konsekwencji zmiany składu produktu;
- **Edukacja** - modele ułatwiają wyjaśnienie działania wielu układów, zachowania np. patogenów osobom, które nie znają tej tematki;
- **Analiza danych i planowanie badań laboratoryjnych** - modelowanie staje się rutynowym narzędziem opisu i analizy danych, pozwala wskazać właściwe miejsca i okresy pobierania próbek co oszczędza czas i pieniądze;
- **Ocena ryzyka** – np. określenie prawdopodobieństwa wywołania zachorowania przez dany produkt spożywczy;

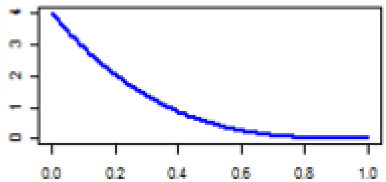
Modelowanie matematyczne

Podstawowe pojęcia:

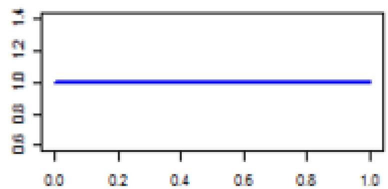
- Rozkład prawdopodobieństwa
- Iteracja
- Model matematyczny
- Metoda Monte Carlo
- Symulacja

Podejście stochastyczne w ilościowej ocenie ryzyka

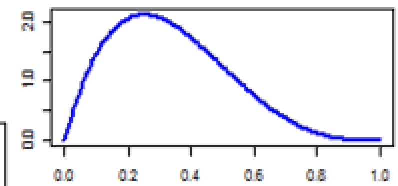
Dane wejściowe – matematyczny opis zagrożeń i innych czynników składowych



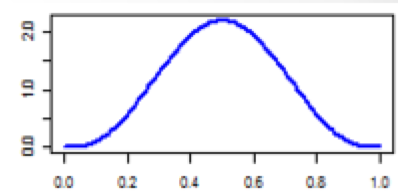
Rozkład



Rozkład



Rozkład



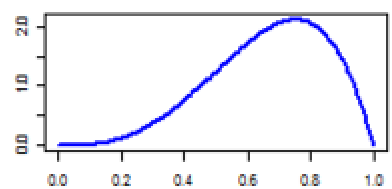
Rozkład

Matematyczny model - kompilacja i implementacja wzorów matematycznych

Model matematyczny opisujący zależności i szacujący prawdopodobieństwo

Symulacja

Wynik ilościowej oceny ryzyka



Wynik w postaci rozkładu prawdopodobieństwa (ryzyka)

Interpretacja

Wnioski

Rozkład prawdopodobieństwa

– rozpisane prawdopodobieństwa poszczególnych zdarzeń elementarnych, które razem tworzą przestrzeń zdarzeń elementarnych i których suma prawdopodobieństw wynosi 1.

Przykład

Dla rzutu kostką, przestrzeń zdarzeń elementarnych składa się z sześciu zdarzeń elementarnych:

$$\Omega = \{1,2,3,4,5,6\}$$

$$\overline{\overline{\Omega}} = 6$$

Zakładając, że kostka jest idealnie symetryczna, prawdopodobieństwo każdego zdarzenia elementarnego jest takie samo i wynosi:

$$P(\omega) = \frac{1}{\overline{\overline{\Omega}}} = \frac{1}{6}$$

Rozkład prawdopodobieństwa możemy zapisać w tabeli:

ω	1	2	3	4	5	6
$P(\omega)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Rozkład prawdopodobieństwa – cd.

Przykład – cd.

Suma prawdopodobieństw wszystkich zdarzeń elementarnych musi wynosić 1:

$$P(\omega_1) + P(\omega_2) + P(\omega_3) + P(\omega_4) + P(\omega_5) + P(\omega_6) = 1$$

$$\frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = 1$$

$$\frac{6}{6} = 1$$

SCHEMAT KLASYCZNY – rozkład prawdopodobieństwa, w którym prawdopodobieństwo każdego zdarzenia jest takie samo (**zdarzenia są jednakowo prawdopodobne**).

Rozkład prawdopodobieństwa – cd.

Przykład

Dany jest rozkład prawdopodobieństwa dla rzutu **niesymetryczną** kostką:

ω	1	2	3	4	5	6
$P(\omega)$	$\frac{1}{3}$	$\frac{1}{12}$	$\frac{1}{4}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{6}$

PRAWDOPODOBIEŃSTWO ZDARZENIA LOSOWEGO jest sumą prawdopodobieństw wszystkich zdarzeń elementarnych, które je spełniają.

Rozkład prawdopodobieństwa – cd.

ω	1	2	3	4	5	6
$P(\omega)$	$\frac{1}{3}$	$\frac{1}{12}$	$\frac{1}{4}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{6}$

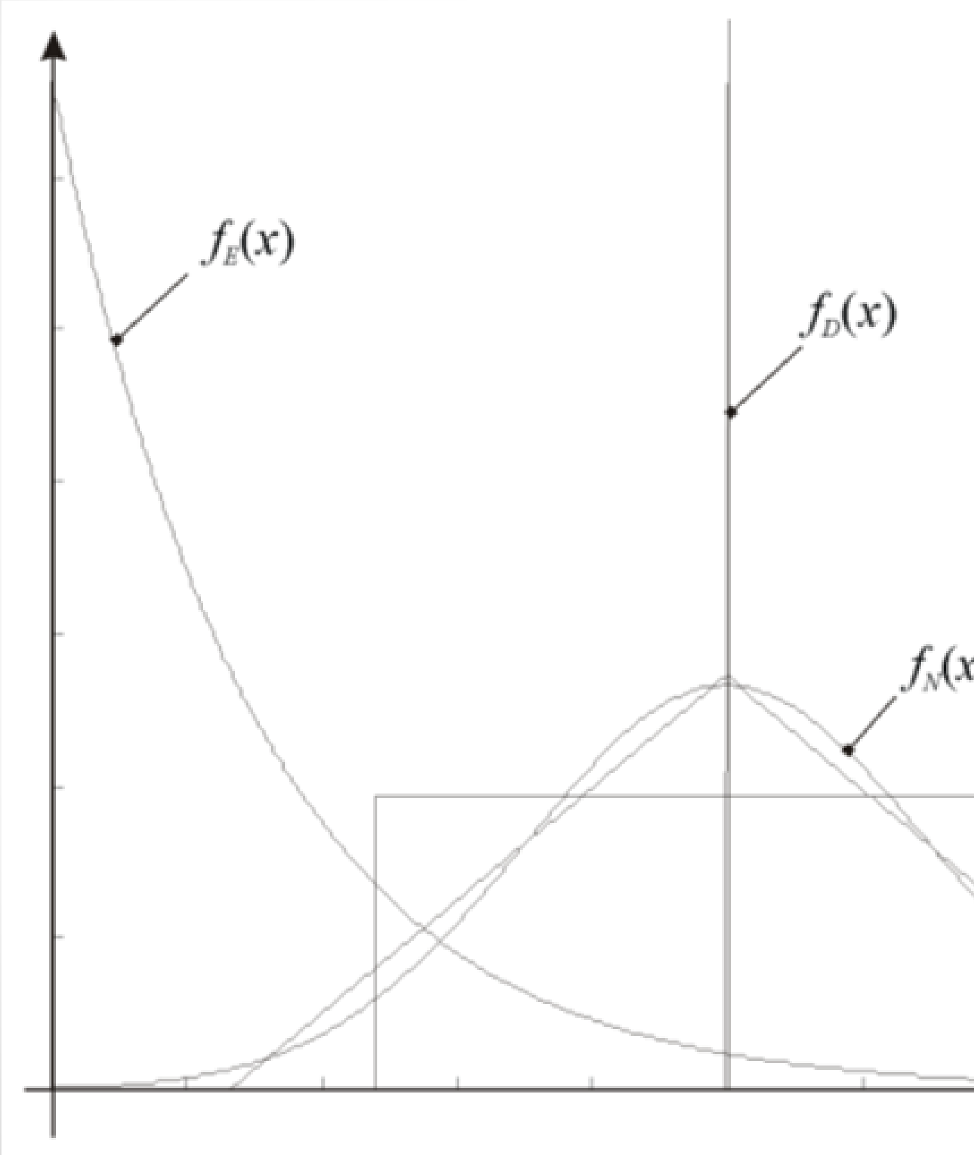
Przykład

Dla podanego rozkładu prawdopodobieństwa obliczymy prawdopodobieństwo zdarzenia losowego A, polegającego na wyrzuceniu **nieparzystej liczby oczek**.

$$P(A) = P(\omega_1) + P(\omega_3) + P(\omega_5)$$

$$P(A) = \frac{1}{3} + \frac{1}{4} + \frac{1}{12} = \frac{4}{12} + \frac{3}{12} + \frac{1}{12} = \frac{8}{12} = \frac{2}{3}$$

Rozkład prawdopodobieństwa – cd.



Wybrane rozkłady gęstości prawdopodobieństwa:

$f_N(x)$ – rozkład normalny,
$f_E(x)$ – rozkład wykładniczy,
$f_R(x)$ – rozkład jednostajny,
$f_T(x)$ – rozkład trójkątny,
$f_D(x)$– rozkład delty Diraca dla zmiennej pewnej.

Oszacowanie funkcji **rozkładu prawdopodobieństwa**

dla wyniku

- **20** rzutów **symetryczną** monetą (orzeł = sukces)
- zbadania **20** losowo pobranych próbek z populacji dzików, w której **co drugi** jest zakażony ASF (próbka dodatnia = sukces)

za pomocą **symulacji** o różnej liczbie **iteracji**

z użyciem **metody Monte Carlo**

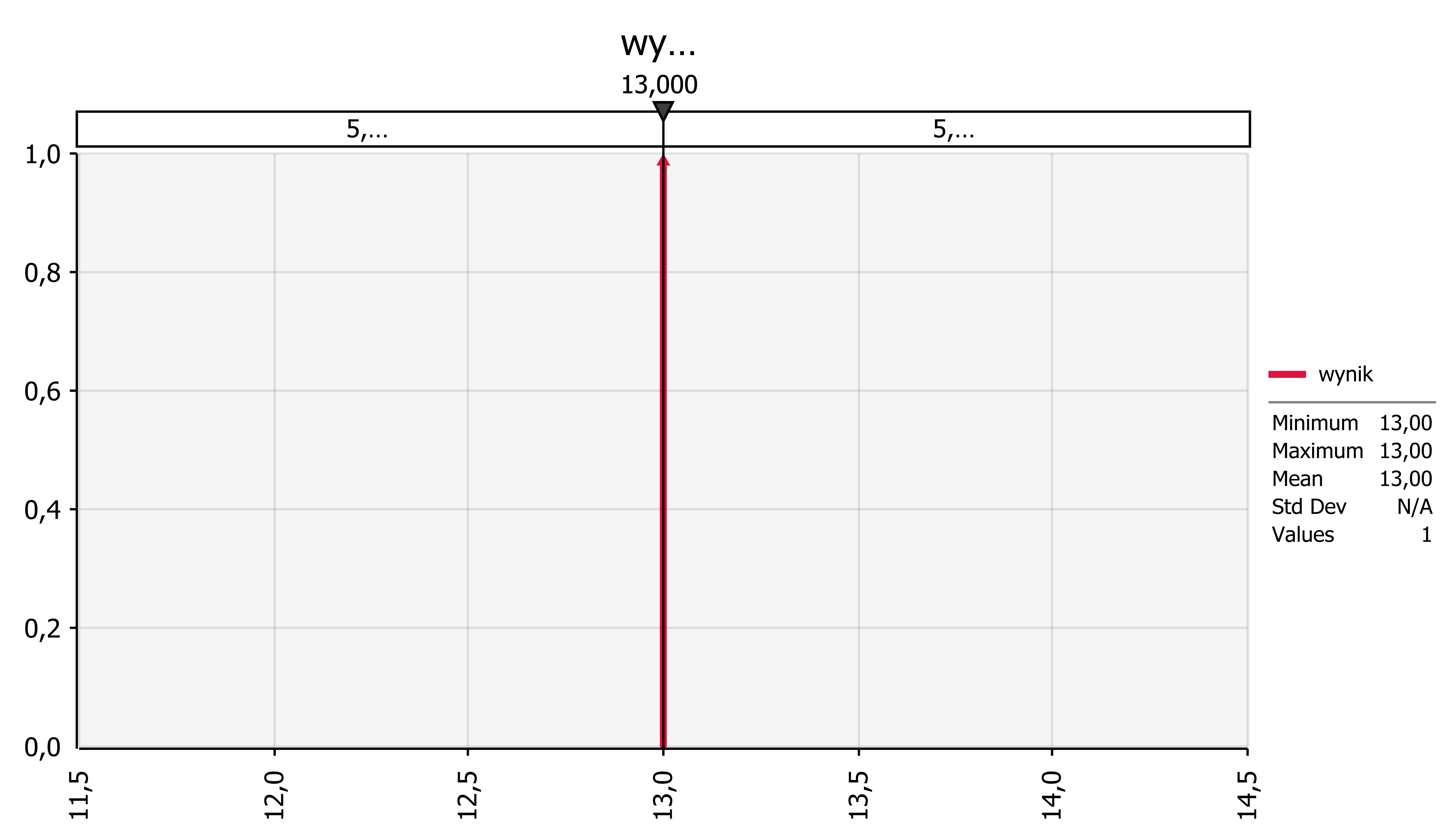
(oprogramowanie **@RISK**)

Rozkład dwumianowy

Parametry rozkładu:

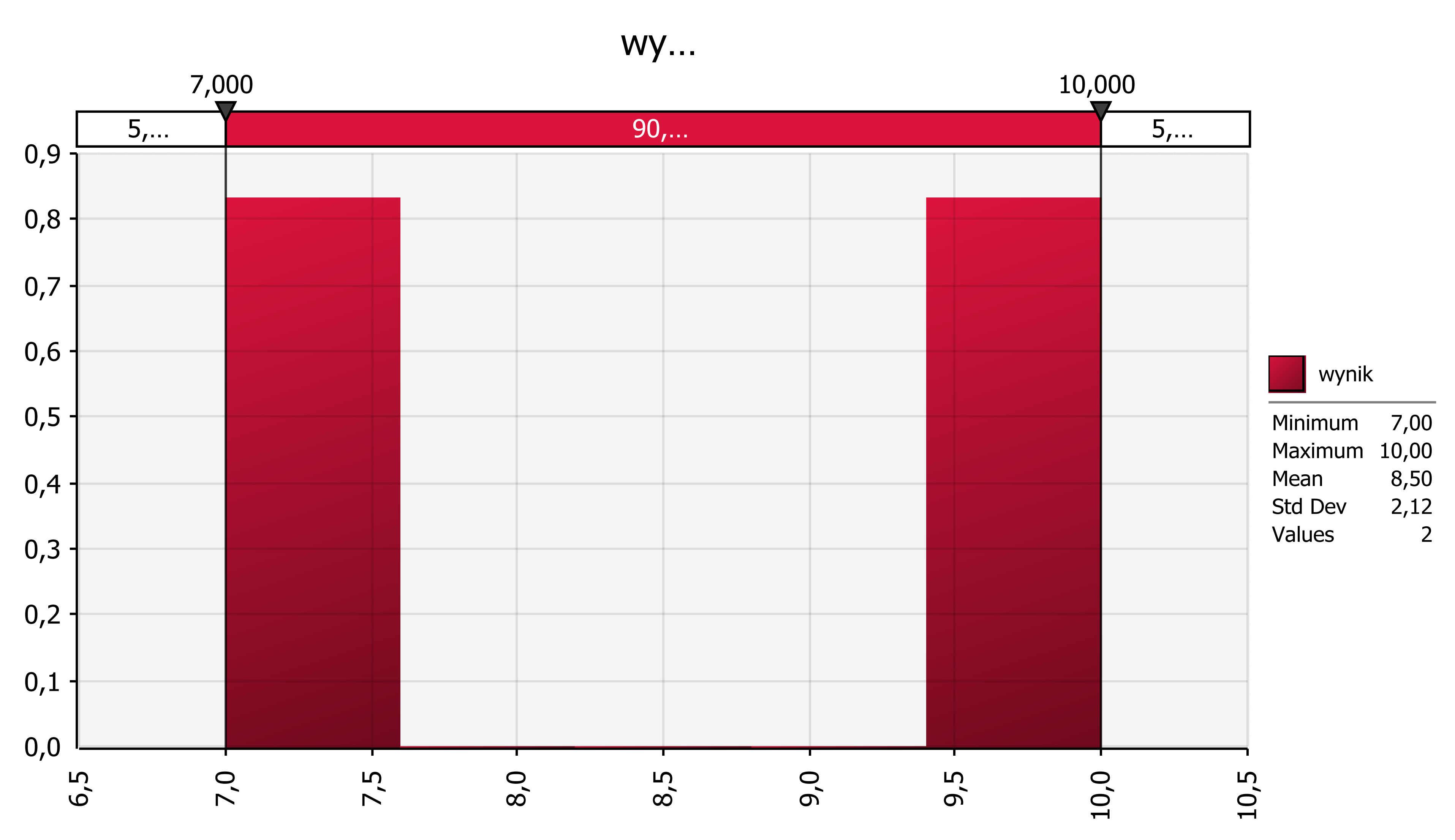
- prawdopodobieństwo „sukcesu”: **0,5**
- liczba niezależnych prób: **20**

- Oszacowanie z użyciem 1 iteracji:



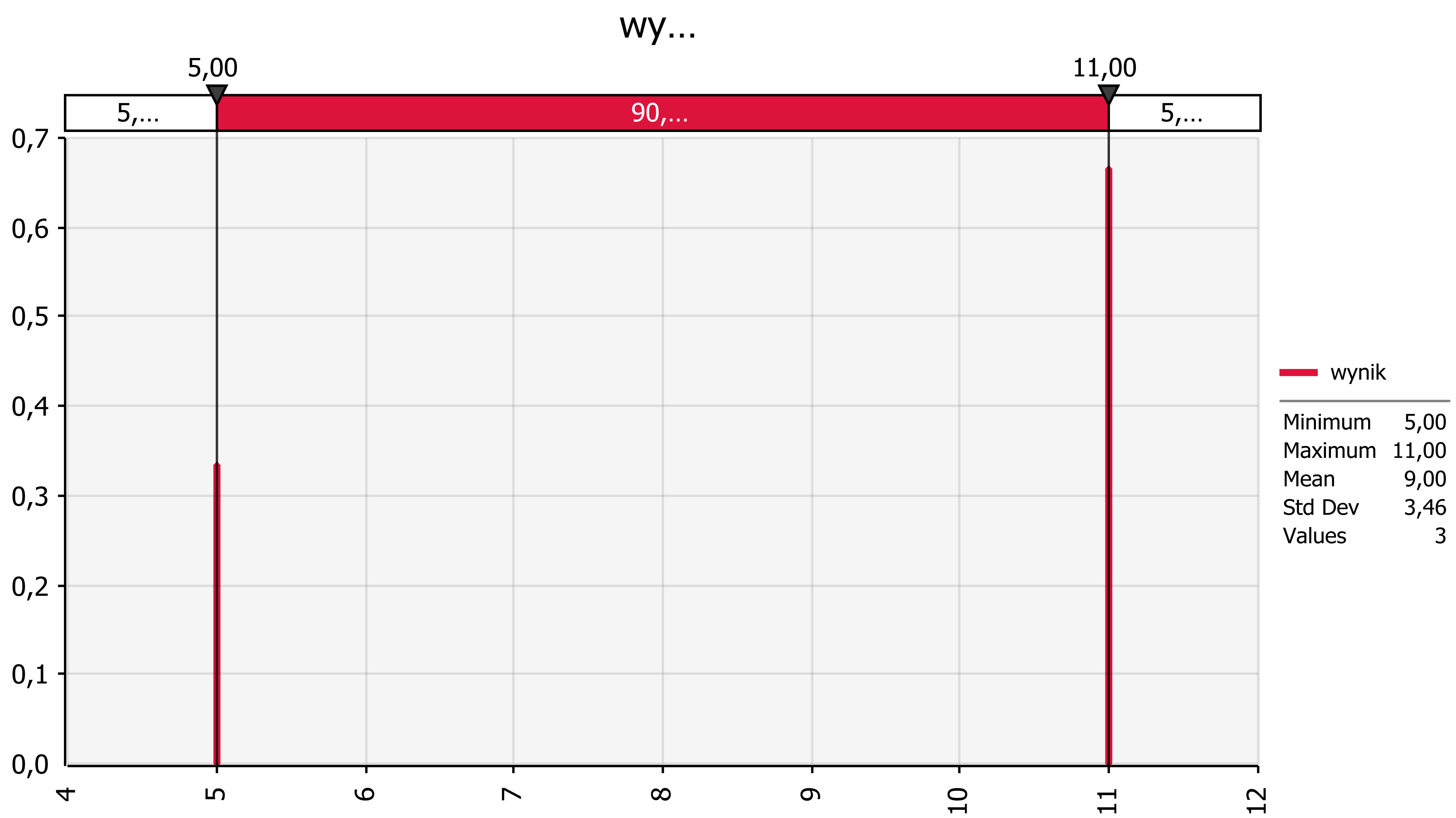
Otrzymano 13 sukcesów (**średnia: 13**).

- Oszacowanie z użyciem 2 iteracji:



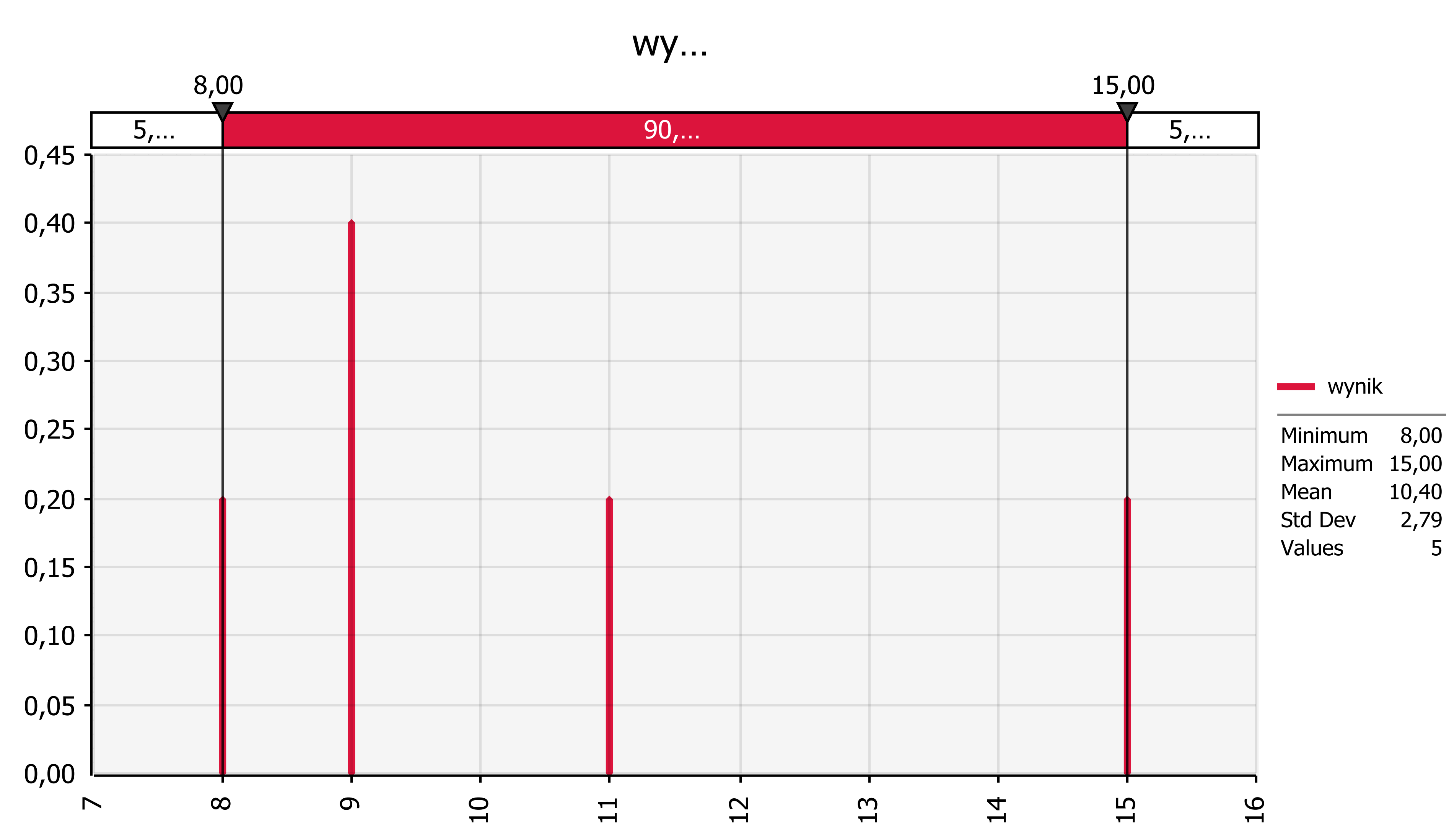
Otrzymano raz 10 a raz 7 sukcesów (**średnia: 8,5**).

• Oszacowanie z użyciem 3 iteracji:



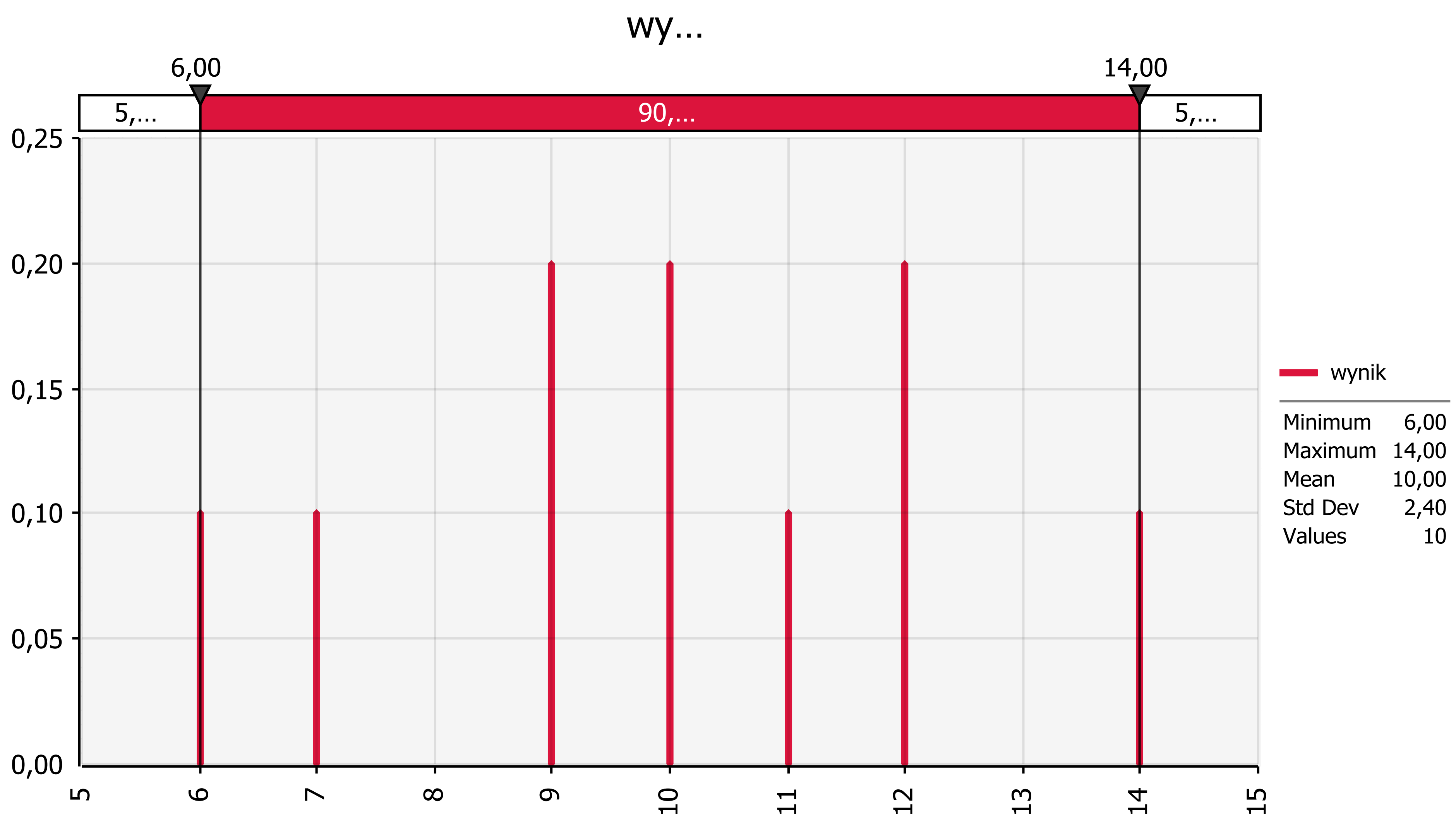
Otrzymano dwa razy 11 i raz 5 sukcesów (**średnia: 9**).

- Oszacowanie z użyciem 5 iteracji:



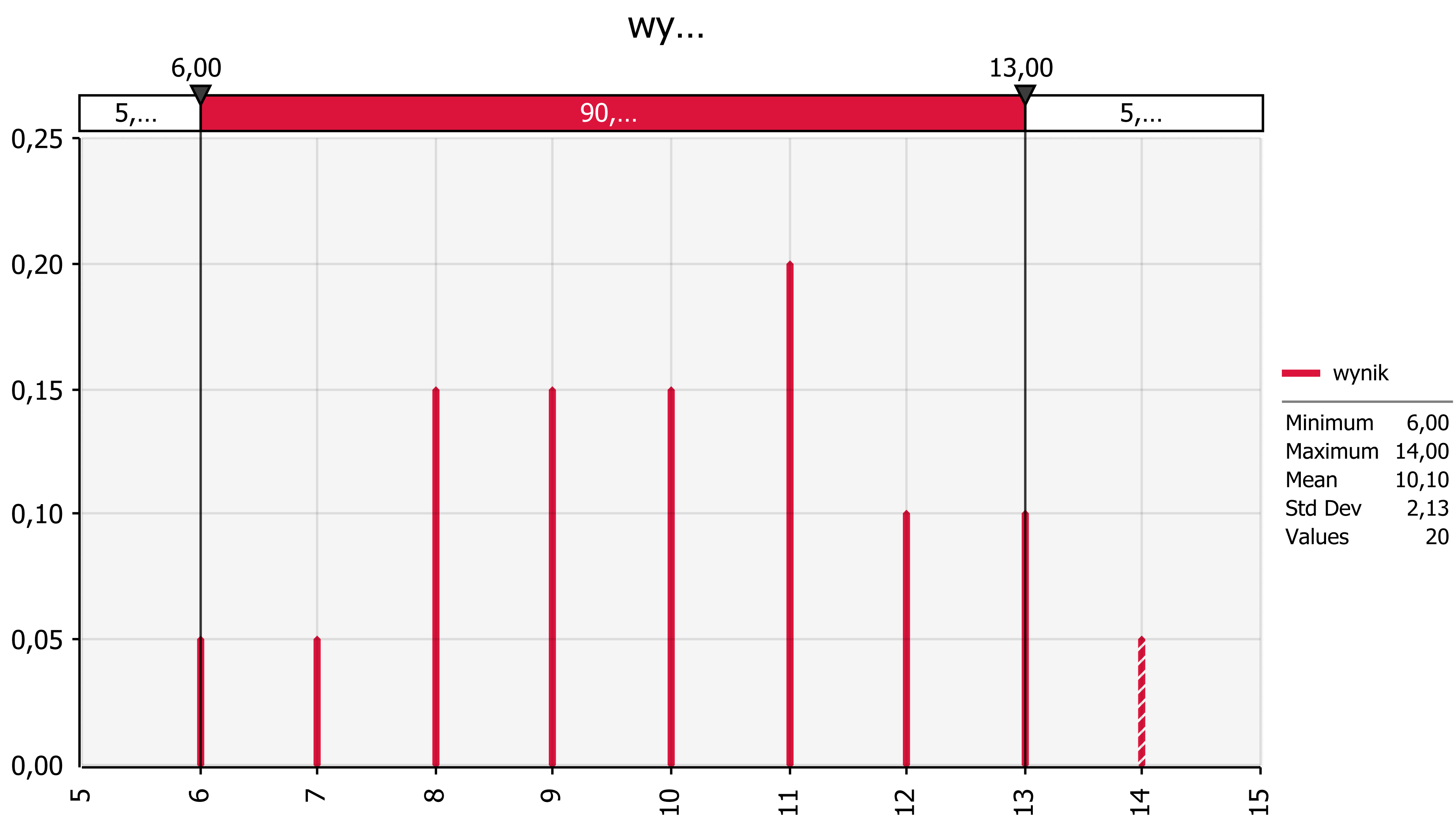
Otrzymano raz 8, raz 11, raz 15 i dwa razy 9 sukcesów (**średnia: 10,4**).

•Oszacowanie z użyciem 10 iteracji:



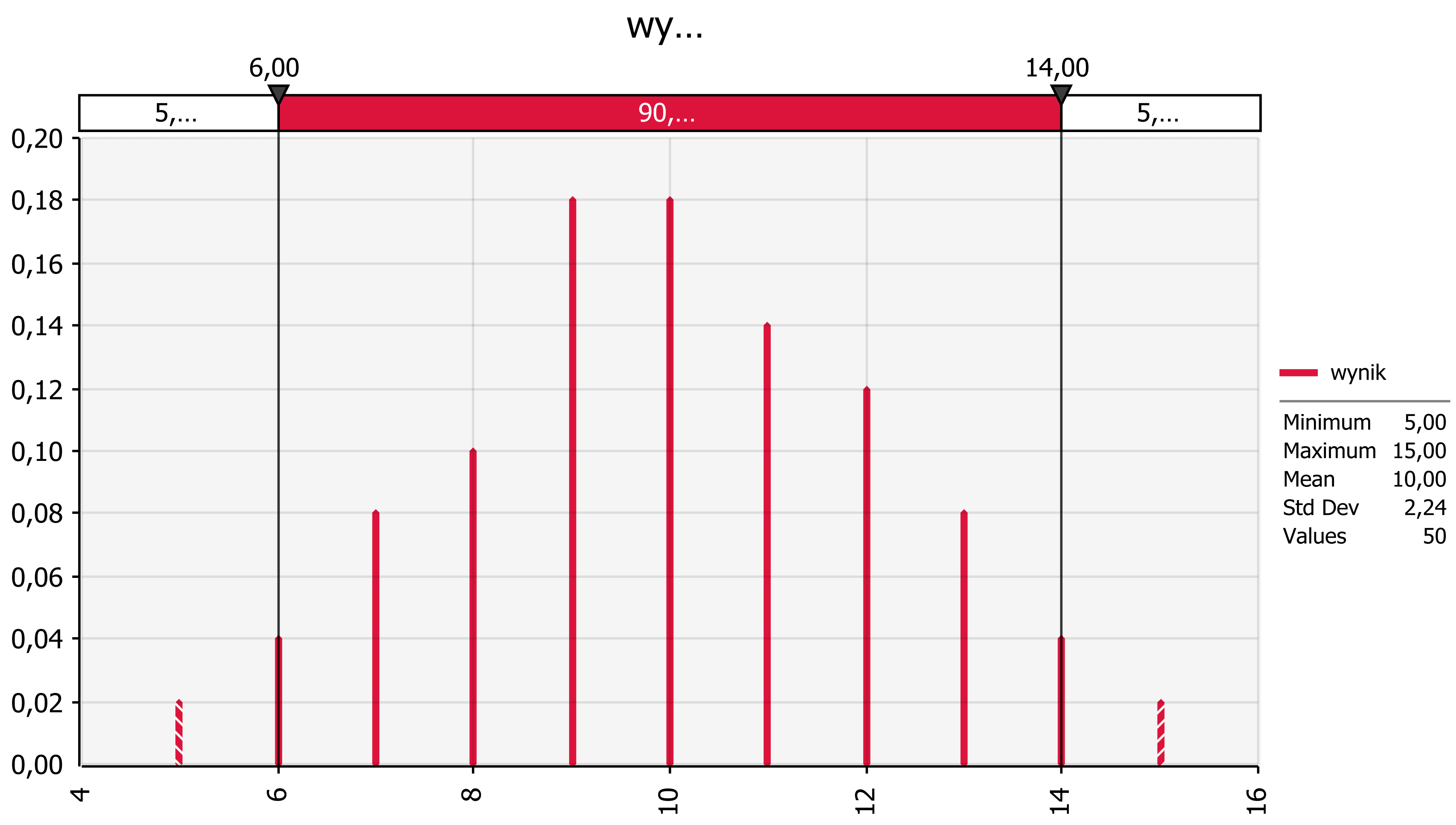
Otrzymano jeden raz 6, 7, 11 i 14 oraz dwa razy 9, 10 i 12 sukcesów (**średnia: 10,0**).

• Oszacowanie z użyciem 20 iteracji:



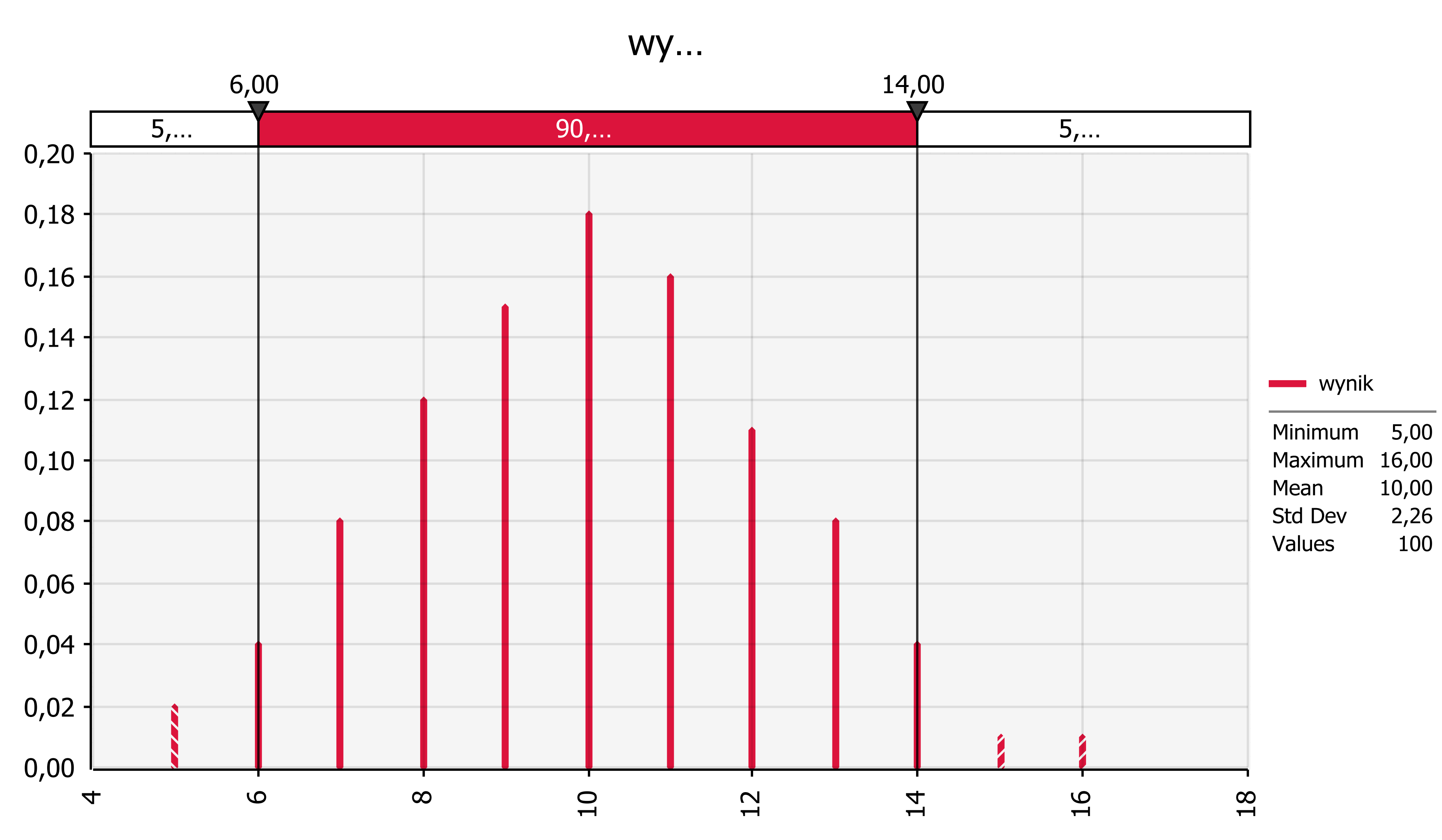
Otrzymano jeden raz 6, 7 i 14, dwa razy 12 i 13, trzy razy 8, 9, 10 oraz cztery razy 11 sukcesów (**średnia: 10,1**).

- Oszacowanie z użyciem 50 iteracji:



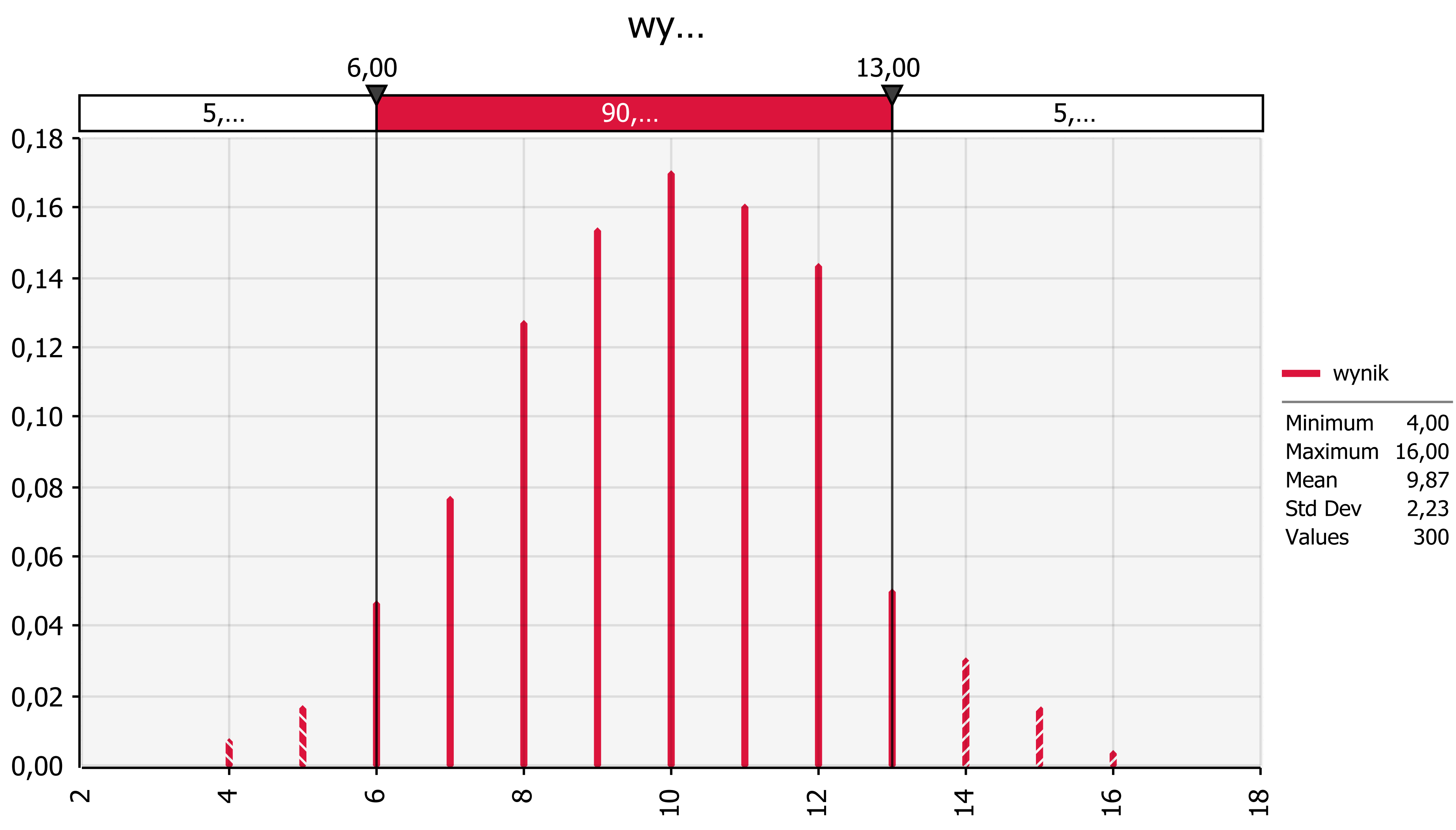
Otrzymano jeden raz 5 i 15, dwa razy 6 i 14, cztery razy 7 i 13, pięć razy 8, sześć razy 12, siedem razy 11 oraz dziewięć razy 9 i 10 sukcesów (**średnia: 10,0**).

- Oszacowanie z użyciem 100 iteracji:



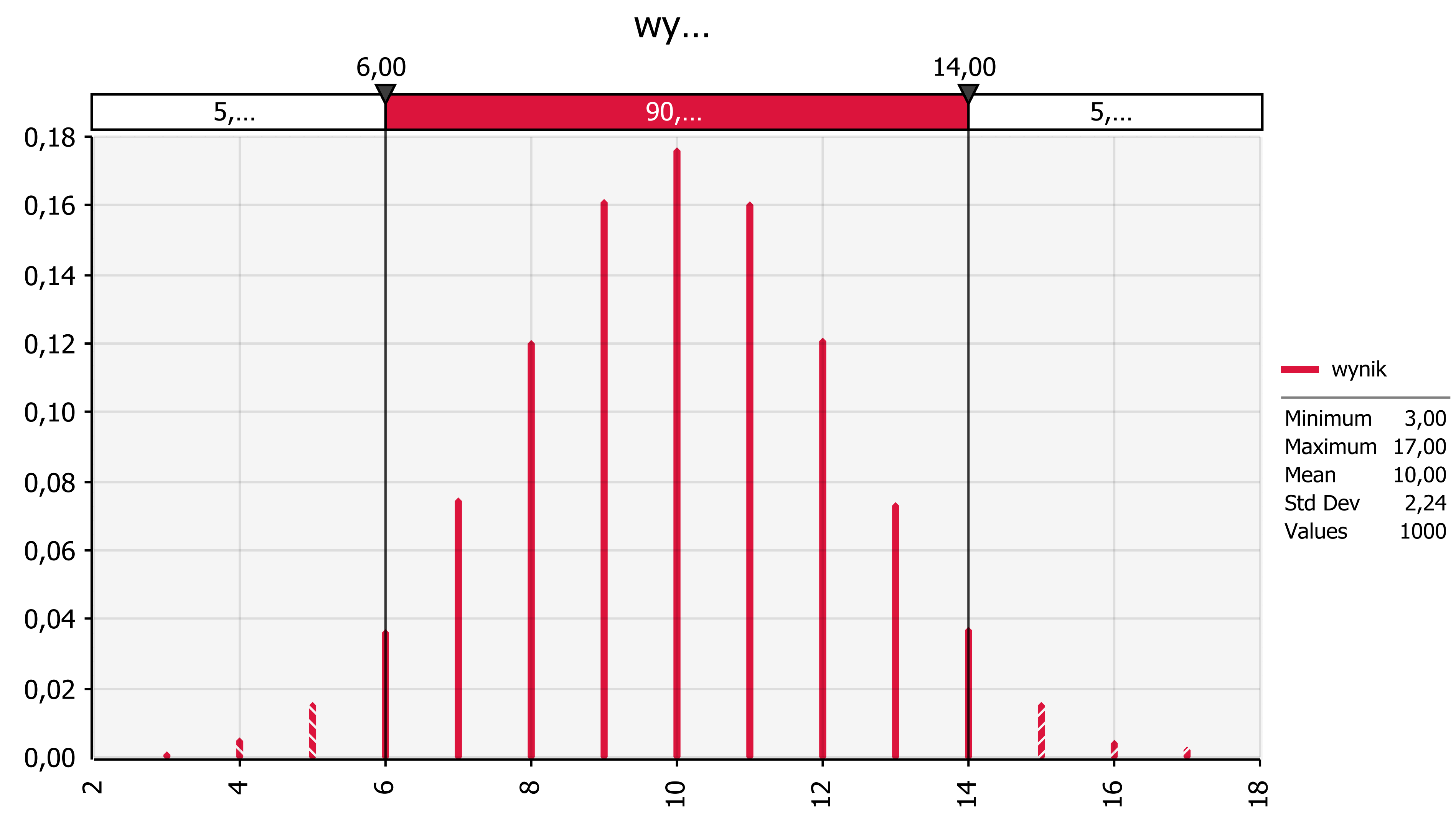
Średnia otrzymanych sukcesów: **10,0**.

- Oszacowanie z użyciem 300 iteracji:



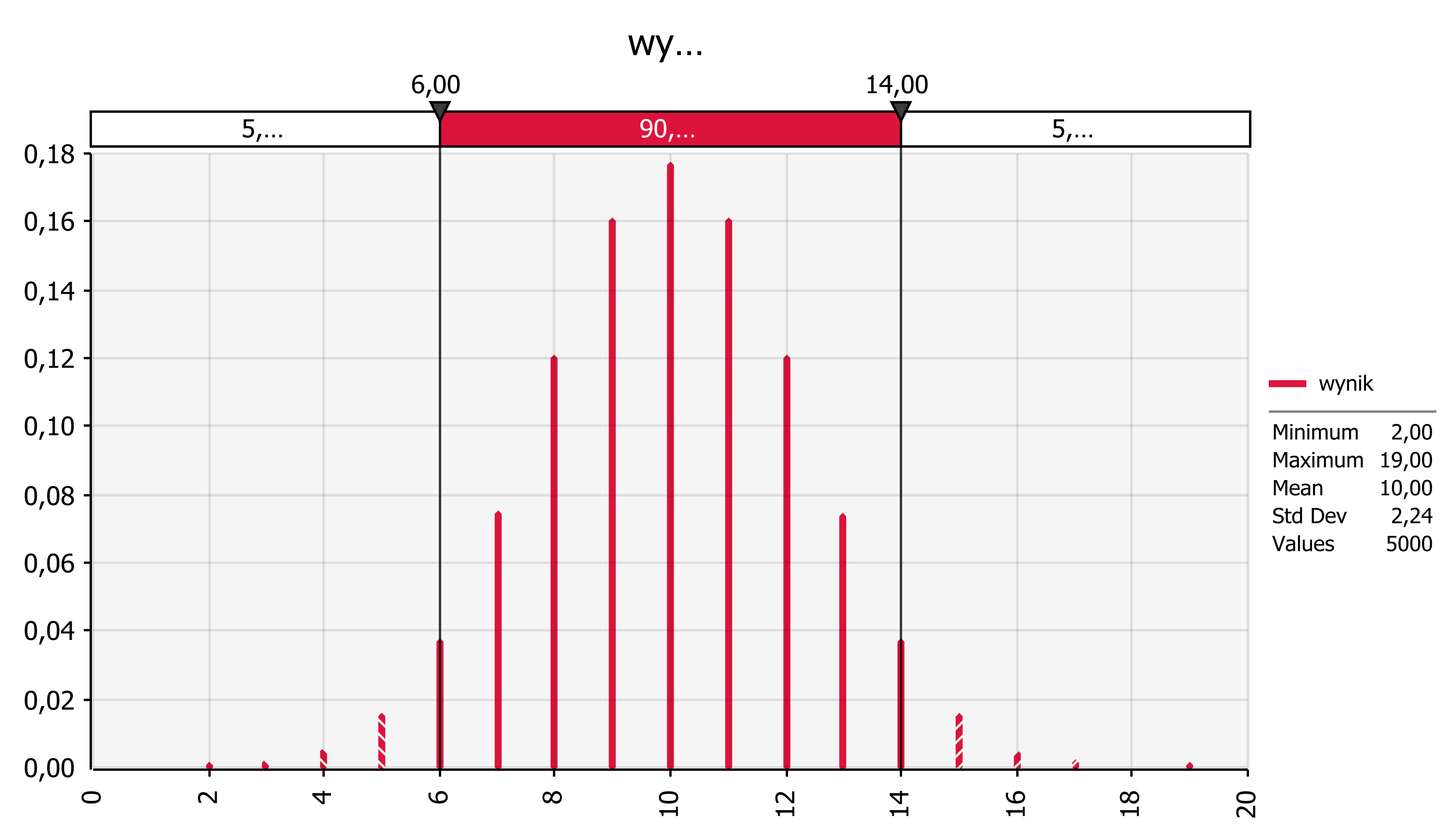
Średnia otrzymanych sukcesów: **9,9**.

• Oszacowanie z użyciem 1000 iteracji:



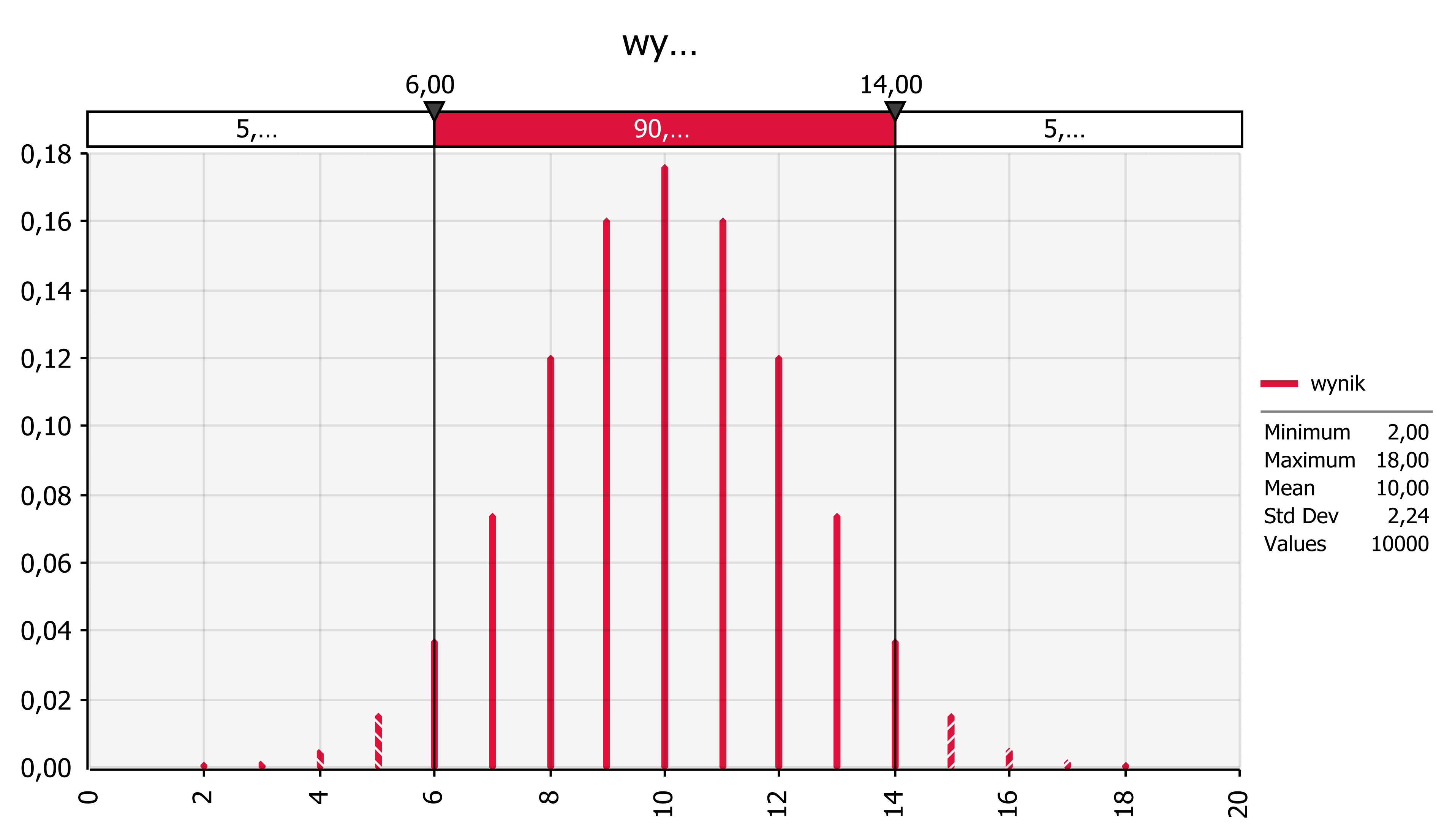
Średnia otrzymanych sukcesów: **10,0**.

- Oszacowanie z użyciem 5000 iteracji:



Średnia otrzymanych sukcesów: **10,0**.

- Oszacowanie z użyciem 10 000 iteracji:



Średnia otrzymanych sukcesów: **10,0**.

Symulacja

- przybliżone odtwarzanie zjawiska lub zachowania danego obiektu za pomocą jego **modelu**.

Szczególnym rodzajem modelu jest **model matematyczny**, często zapisany w postaci **programu komputerowego**.

Zastosowanie symulacji:

- każda dziedzina nauki i techniki
- cele wojskowe
- rozrywka, np. w grach komputerowych

Modelowanie matematyczne

- użycie języka matematyki do opisania zachowania jakiegoś układu (np. biologicznego, ekonomicznego, termodynamicznego).

Model matematyczny opisuje dany układ za pomocą **zmiennych**. Wartości zmiennych mogą należeć do różnych zbiorów: liczb rzeczywistych, całkowitych, wartości logicznych, ciągów znakowych itp.

Zmienne reprezentują pewne właściwości układu, np. zmierzone wartości wyjść układu, wartości liczników, wystąpienia zdarzeń (tak/nie) itp.

Właściwy model to grupa **funkcji** wiążących ze sobą różne zmienne i w ten sposób opisujących powiązania między wielkościami w układzie.

Np.: ***przebyta droga [km] / czas [h] = średnia prędkość [km/h]***

Modelowanie matematyczne – cd.

- **Wiedza a priori**

  "czarne skrzynki" (ang. *black-box*) i "białe skrzynki" (ang. *white-box*)

- **Złożoność**

  "białoskrzynkowy" model układu a koszt obliczeniowy i margines błędu całego modelu

- **Ocena modelu**

  Jak określić, czy model dobrze opisuje rzeczywisty układ?

  Jak określić, czy użyty zestaw pomiarów jest reprezentatywny dla wszystkich możliwych sytuacji?

  *Czy model dobrze opisuje własności układu dla danych pomiędzy punktami pomiarowymi (**interpolacja**)?*

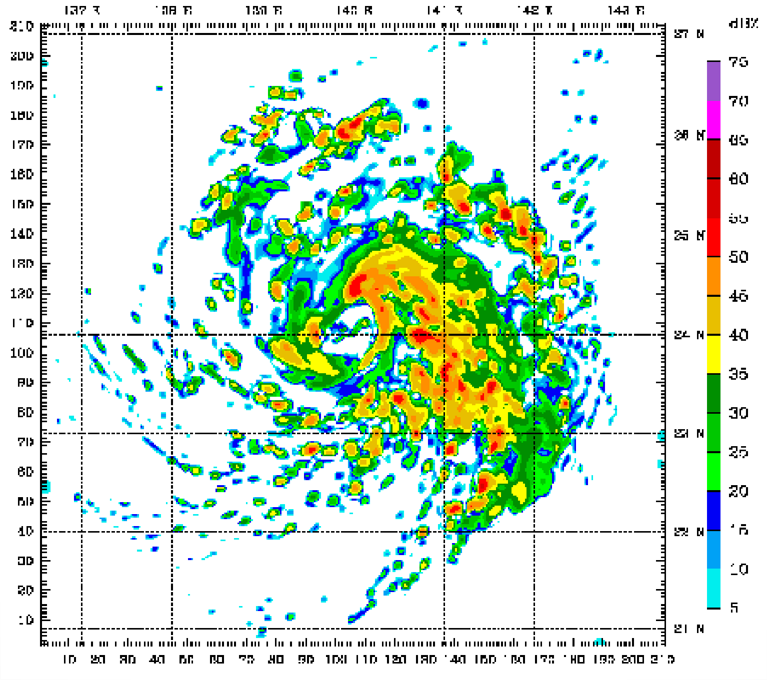
  *Czy model dobrze opisuje zdarzenia spoza przedziału pomiarów (**ekstrapolacja**)?*

Symulacja komputerowa

– symulacja z wykorzystaniem **modelu matematycznego**, zapisanego w postaci **programu komputerowego**.

- Przydatność: m.in. systemy złożone i trudności w wyznaczeniu analitycznego rozwiązania

- Przykład:



Symulacja 48 godzin tajfunu Mawar (2005)

Symulacja komputerowa – cd.

Rodzaje

– ze względu na:

- **przewidywalność zdarzeń**
    - o **stochastyczne** – korzystają z generatora liczb pseudolosowych lub (bardzo rzadko) losowych (szczególnie popularna jest metoda Monte Carlo)
    - o **deterministyczne** – wynik jest powtarzalny i zależy tylko od danych wejściowych i ewentualnych interakcji ze światem zewnętrznym
- **sposób upływu czasu**
    - o **z czasem ciągłym** – czas zwiększa się stałymi przyrostami, jak w symulacji z czasem dyskretnym, lecz wartości próbek sygnałów są interpolowane dla chwil pośrednich pomiędzy momentami odczytu
    - o **z czasem dyskretnym** – czas zwiększa się stałymi przyrostami, a krok czasowy dobiera się optymalnie ze względu na „zasobożerność" systemu, jego wydajność i charakter symulowanego obiektu i/lub zjawiska (np. mikrosekundy w obwodach elektrycznych i miliony lat przy symulacji ewolucji gwiazd)
    - o **symulacja zdarzeń dyskretnych** – czas zwiększa się skokowo, ale jego przyrosty są zmienne (ważniejsza jest tu sekwencja zdarzeń niż rzeczywisty lub wirtualny upływ czasu)
- **formę danych wyjściowych**
    - o **statyczne** – wynikiem jest zbiór danych, statyczny obraz itp.
    - o **dynamiczne** – wynikiem jest proces przebiegający w czasie np. animacja
        - interaktywne – reagują na sygnały ze świata zewnętrznego np. operatora
        - nieinteraktywne
- **liczbę użytych komputerów**
    - o **lokalne** – przetwarzanie odbywa się na pojedynczym komputerze
    - o **rozproszone** – przetwarzanie odbywa się w wielu komputerach połączonych w sieci (lokalnej lub zewnętrznej)

Symulacja komputerowa – cd.

Rodzaje

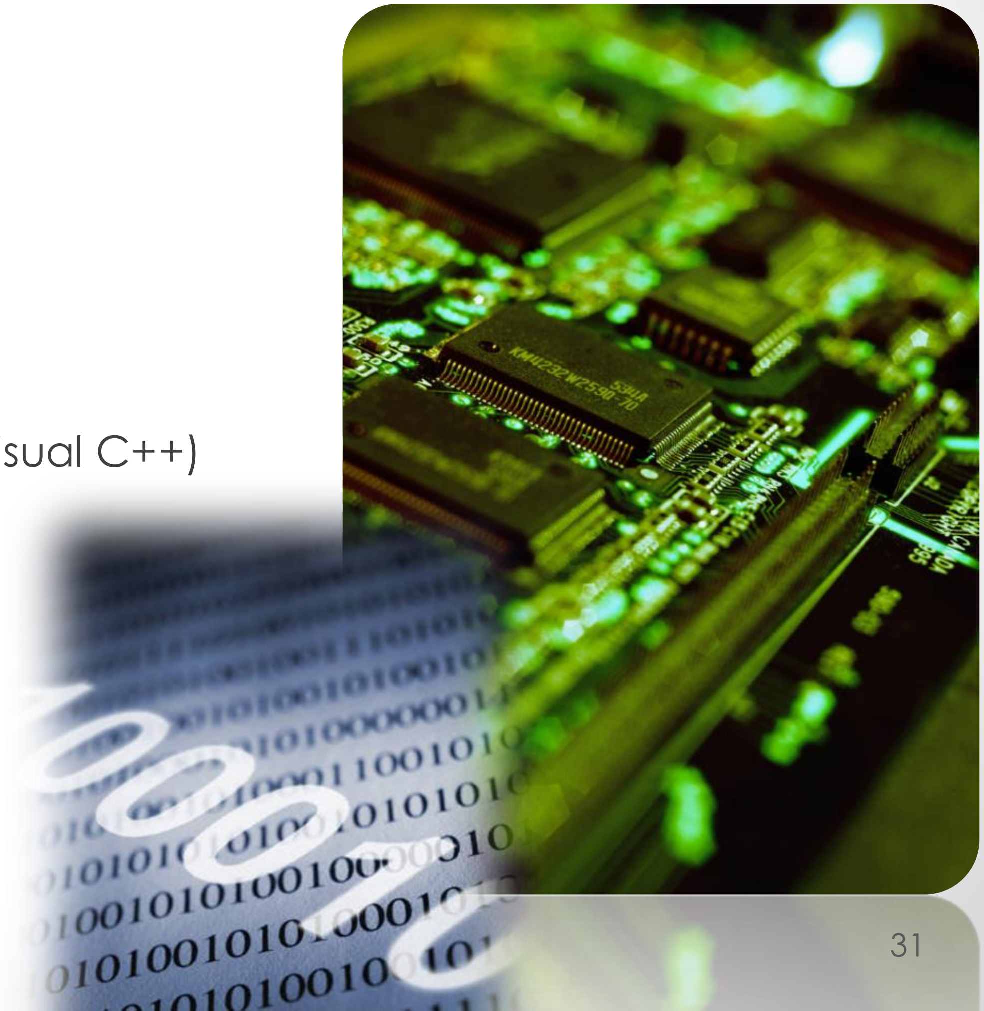
– ze względu na:

- **przewidywalność zdarzeń**
  - **stochastyczne** – korzystają z generatora liczb pseudolosowych lub (bardzo rzadko) losowych (szczególnie popularna jest metoda **Monte Carlo**)
  - **deterministyczne** – wynik jest powtarzalny i zależy tylko od danych wejściowych i ewentualnych interakcji ze światem zewnętrznym
- **sposób upływu czasu**
  - **z czasem ciągłym** – czas zwiększa się stałymi przyrostami, jak w symulacji z czasem dyskretnym, lecz wartości próbek sygnałów są interpolowane dla chwil pośrednich pomiędzy momentami odczytu
  - **z czasem dyskretnym** – czas zwiększa się stałymi przyrostami, a krok czasowy dobiera się optymalnie ze względu na „zasobożerność" systemu, jego wydajność i charakter symulowanego obiektu i/lub zjawiska (np. mikrosekundy w obwodach elektrycznych i miliony lat przy symulacji ewolucji gwiazd)
  - **symulacja zdarzeń dyskretnych** – czas zwiększa się skokowo, ale jego przyrosty są zmienne (ważniejsza jest tu sekwencja zdarzeń niż rzeczywisty lub wirtualny upływ czasu)
- **formę danych wyjściowych**
  - **statyczne** – wynikiem jest zbiór danych, statyczny obraz itp.
  - **dynamiczne** – wynikiem jest proces przebiegający w czasie np. animacja
    - interaktywne – reagują na sygnały ze świata zewnętrznego np. operatora
    - nieinteraktywne
- **liczbę użytych komputerów**
  - **lokalne** – przetwarzanie odbywa się na pojedynczym komputerze
  - **rozproszone** – przetwarzanie odbywa się w wielu komputerach połączonych w sieci (lokalnej lub zewnętrznej)

Symulacja komputerowa – cd.

Narzędzia

- język programowania GPSS
- Crystal Ball
- @Risk
- Arena
- SciLab
- ModSim (oparty na Microsoft Visual C++)
- R
- Python
- inne języki programowania
- arkusz kalkulacyjny



Symulacja komputerowa – cd.

Wybrane zastosowania

- **symulatory statków powietrznych, okrętów podwodnych, czołgów itp.**
- **w ekonomii i biznesie**
  - systemy kolejkowe
  - zarządzanie zapasami
  - wycena instrumentów pochodnych
  - ocena projektów inwestycyjnych
- **w naukach społecznych**
  - Dynamiczna teoria wpływu społecznego Nowaka-Latane
  - prognozowanie podziału miejsc w parlamencie
  - dynamika populacji
- **nauki przyrodnicze**
  - meteorologia – prognozy pogody
  - ocena ryzyka wprowadzenia i analiza rozprzestrzeniania się wirusów itp.
- **w naukach inżynieryjnych**
  - budownictwo – wytrzymałość konstrukcji
  - lotnictwo – wytrzymałość konstrukcji
  - elektronika – analiza obwodów elektrycznych
- **matematyka**
  - numeryczne wyznaczanie rozwiązań równań różniczkowych
  - symulacyjne wyznaczanie dystrybuant funkcji, które nie dają się całkować (np. rozkładu normalnego)
- **komputerowe gry symulacyjne**

Iteracja

– (łac. *iteratio* – powtarzanie) – czynność powtarzania (najczęściej wielokrotnego) tej samej instrukcji (albo wielu instrukcji) w pętli.

– powtarzanie pewnej wzorcowej czynności lub procesu **celem otrzymania kolejnego przybliżenia** (wielokrotne stosowanie jakiegoś procesu numerycznego po to, aby stopniowo ulepszać wcześniejsze wyniki)

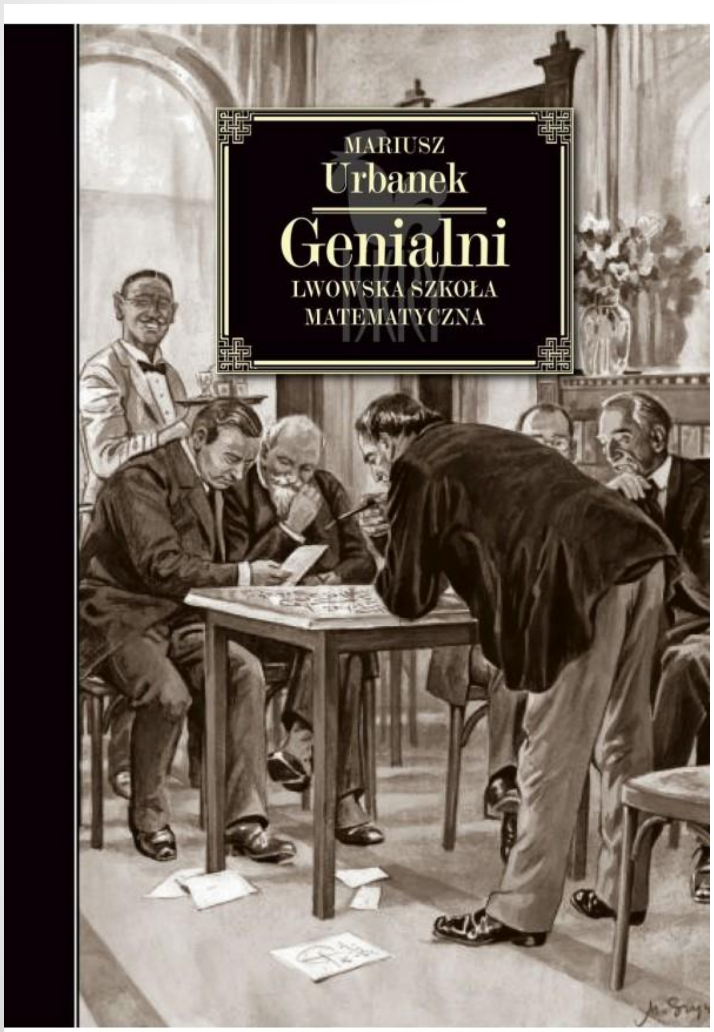
Metoda Monte Carlo (MC)

– jest stosowana do **modelowania matematycznego** procesów zbyt złożonych (obliczania całek, łańcuchów procesów statystycznych), aby można było przewidzieć ich wyniki za pomocą podejścia analitycznego.

Istotną rolę w metodzie MC odgrywa **losowanie** wielkości charakteryzujących proces, przy czym losowanie dokonywane jest **zgodnie z rozkładem**, który musi być znany.

Stanisław Ulam

Metoda Monte Carlo (MC)



188

Mariusz Urbanek

Kiedy przyjaciel zaproponował Ulamowi partię szachów, uczony bał się, że nie będzie pamiętał zasad gry. A kiedy wygrał, zaczął podejrzewać, że może partner specjalnie zagrał poniżej możliwości, chcąc dać mu wygrać. Kiedy amerykańskie Towarzystwo Matematyczne zamówiło u niego wspomnienie o Banachu, miał wątpliwości, czy tekst będzie wystarczająco dobry. „Pisanie o czyjejś śmierci zaraz po tym, jak samemu ledwie uszło się z życiem, wydawało mi się makabryczne" – wspominał. Powoli przekonywał się jednak, że choroba nie dokonała trwałych zmian w jego psychice. A układając w trakcie rekonwalescencji pasjanse, wpadł na tzw. metodę Monte Carlo, pozwalającą uzyskiwać prawidłowe wyniki obliczeń na podstawie wielu próbek losowych. Metoda okazała się niezastąpiona w sytuacjach, gdy nie ma czasu na precyzyjne obliczenia, a szybkość otrzymania wiarygodnego, choć tylko przybliżonego wyniku, jest ważniejsza od stuprocentowej dokładności.

– Nazwałem ją tak, bo chodziło o przypadek – jak wygrana w kasynie Monte Carlo – tłumaczył Ulam.

W kwietniu 1946 roku dostał zaproszenie na konferencje nauko-

Metoda Monte Carlo (MC)

Dokładność i poprawność metody Monte Carlo

- Dokładność wyniku uzyskanego tą metodą jest zależna od liczby sprawdzeń i jakości użytego generatora liczb pseudolosowych. Zwiększanie liczby prób nie zawsze zwiększa dokładność wyniku.

- Poprawność metody Monte Carlo w przypadku liczenia pól lub całek można udowodnić stosując twierdzenie Picka (lub jego wielowymiarowe uogólnienia).

Metoda Monte Carlo (MC) – cd.

Przykład

Metodą Monte Carlo można obliczyć pole figury zdefiniowanej nierównością:

$$x^2 + y^2 \leq R^2$$

czyli koła o promieniu R i środku w punkcie (0,0).

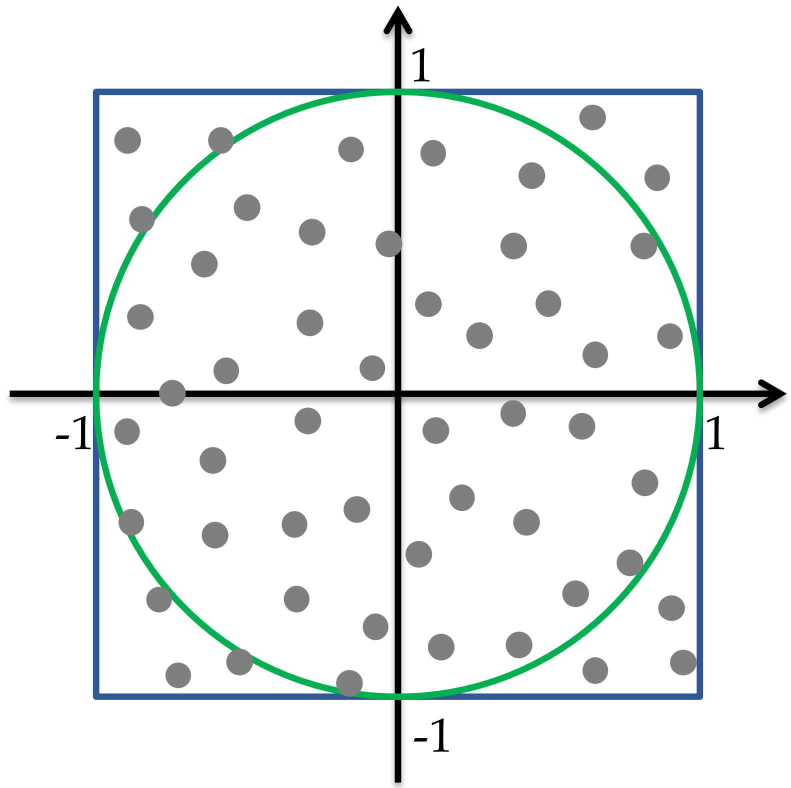
1. Losuje się n punktów z opisanego na tym kole kwadratu – dla koła o $R = 1$ współrzędne wierzchołków: (-1,-1), (-1,1), (1,1), (1,-1).
2. Po wylosowaniu każdego z tych punktów trzeba sprawdzić, czy jego współrzędne spełniają powyższą nierówność (tj. czy punkt należy do koła).

Wynikiem losowania jest informacja, że z n wszystkich prób k było trafionych, zatem pole koła wynosi

$$P_{koła} = P_{kwadratu} \cdot \frac{k}{n}$$

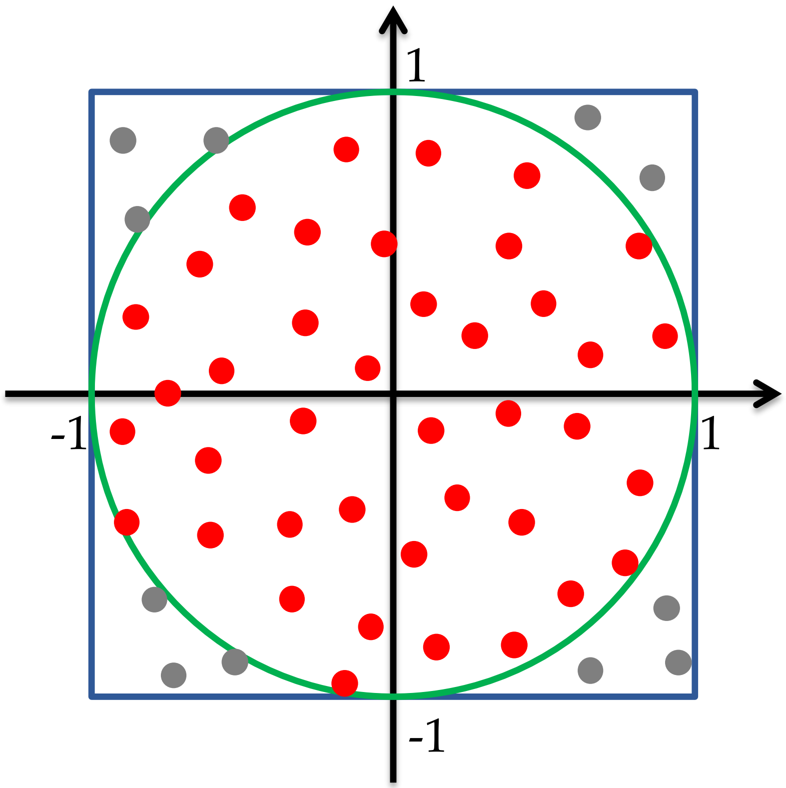
Metoda Monte Carlo (MC) – cd.

Przykład – cd.



Metoda Monte Carlo (MC) – cd.

Przykład – cd.



$$P_{koła} = P_{kwadratu} \cdot \frac{k}{n}$$

Ocena ryzyka na podstawie symulacji komputerowej

Wady:

- konieczność użycia specjalistycznego oprogramowania
- wysoki stopień trudności przeprowadzenia takich analiz

Zalety:

- otrzymanie bardziej wiarygodnych wyników analizy
- możliwość przeprowadzenia oceny ryzyka przy małej ilości lub braku danych wejściowych
- możliwość oceny zjawisk, których symulacji nie można przeprowadzić w rzeczywistości (*Co by było gdyby…*)
- symulację tzw. scenariuszy (różne wersje wydarzeń)
- wykorzystanie metaanalizy (danych opublikowanych przez inne instytucje, w czasopismach, inne programy badań, roczniki statystyczne, itd.).

Wnioski

- Zastosowanie nowoczesnych technik oceny mimo swojego skomplikowania jest nieodzowne. Tylko w taki sposób można ustalić wiarygodność uzyskanych wyników badań.
- Mając nawet niepełne dane można przeprowadzić wstępną ocenę z ustaleniem błędu tej oceny, co może służyć planowaniu dalszych badań.

Źródła

- K. Góralczyk, G. Kostka, J. K. Ludwicki, P. Struciński; Bezpieczeństwo chemiczne, bezpieczeństwo żywności - leksykon terminów; Narodowy Instytut Zdrowia Publicznego - Państwowy Zakład Higieny, Warszawa 2008
- Risk assessment of Listeria monocytogenes in ready-to-eat foods: Interpretative Summary, 2004
- Hazard characterization for pathogens in food and water: Guidelines, 2003
- http://www.matematykam.pl/
- http://pl.wikipedia.org/
- http://oen.dydaktyka.agh.edu.pl/dydaktyka/matematyka/c_metody_numeryczne/wyklad/it.htm