

BIOINFORMATYKA

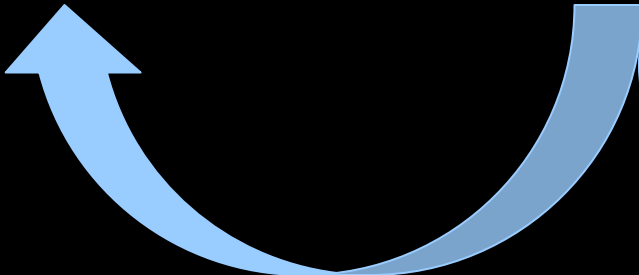
1. Wykład wstępny
2. Struktury danych w badaniach bioinformatycznych
3. Bazy danych: projektowanie i struktura
4. Bazy danych: projektowanie i struktura
5. Zastosowanie przykładowych programów do analizy danych
6. Zastosowanie przykładowych programów do analizy danych
7. Powiązania pomiędzy genami: równ. Hardyego-Weinberga, wsp. rekombinacji
8. Analiza sprzężeń - teoria
9. Analiza sprzężeń – przykłady programów
10. Analiza asocjacyjna
11. Analiza asocjacyjna - przykłady programów
- 12. Symulacje komputerowe, jackknife, bootstrap**
13. Monte Carlo Markov Models
14. Metody klasyfikacyjne
15. Wykład podsumowujący

1. Symulacje komputerowe

- Zasady wykonania analizy asocjacyjnej
- Struktura plików wsadowych
- Interpretacja wyników

2. Metody próbkowania

- Jackknife
- Bootstrap

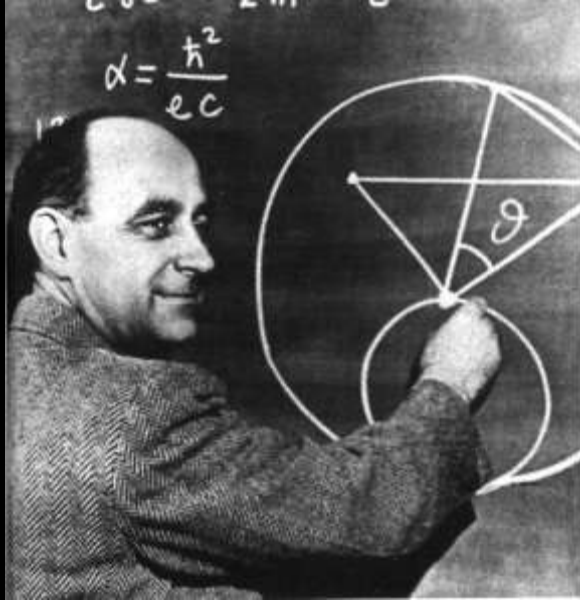


techniki często
stosowane w
bioinformatyce
→ złożoność
struktury danych

SYMULACJE MONTE CARLO

SYMULACJE MONTE CARLO - historia

Enrico Fermi



Stanisław Ulam



John vonNeumann



Instytut Los Alamos, USA



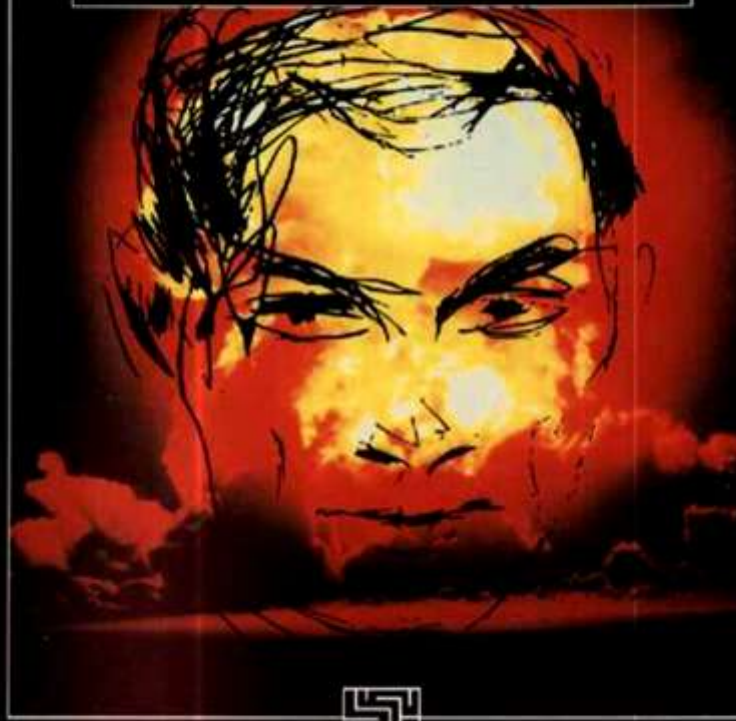
Monte Carlo



SYMULACJE MONTE CARLO - historia

STANISŁAW M. ULAM

PRZYGODY MATEMATYKA



NA SCHODKACH
NAUKI

Prószyński i S-ka

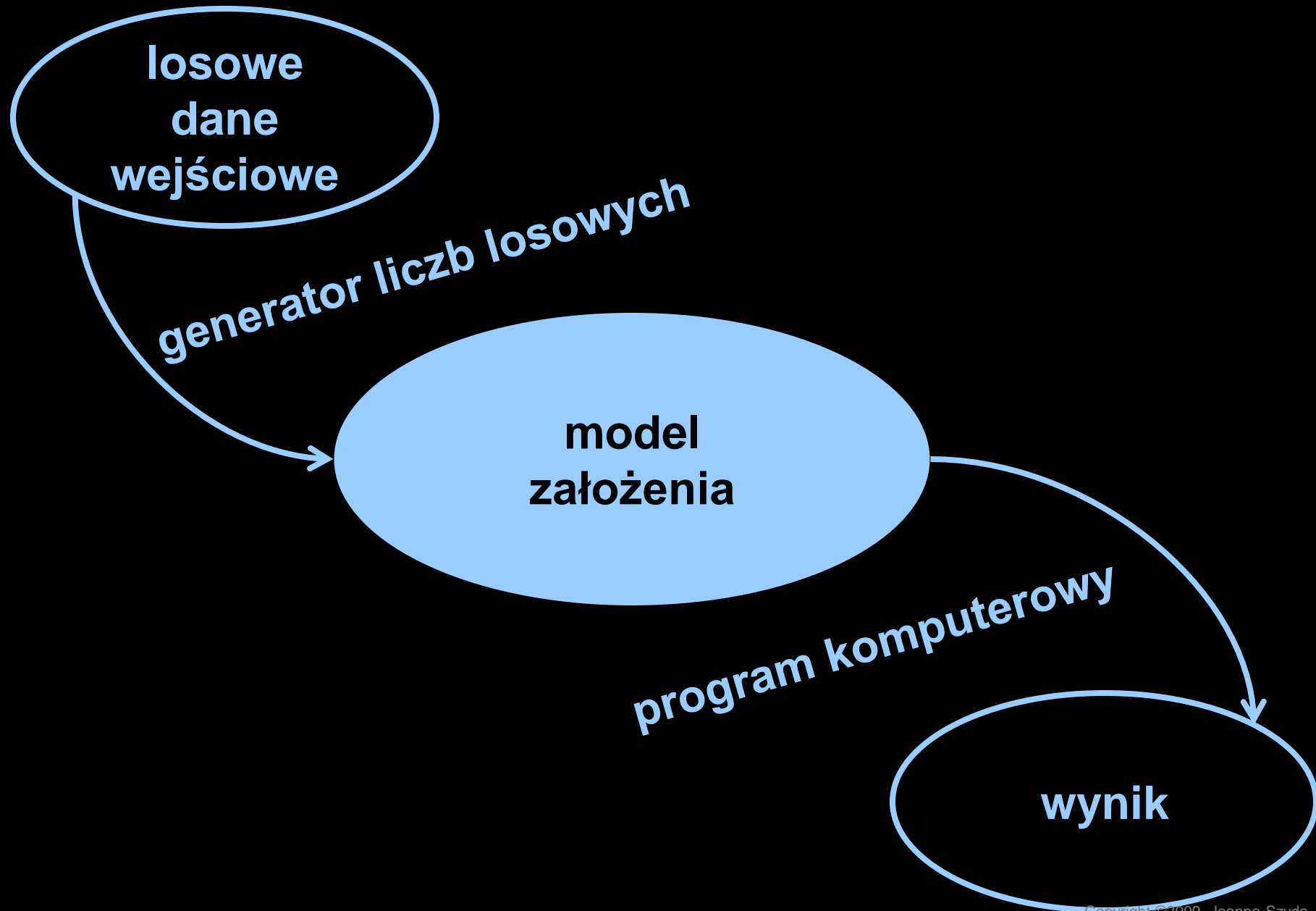
ANALIZA SYSTEMÓW ZŁOŻONYCH, WYKORZYSTUJĄCA:

- **prawdopodobieństwo**
- **losowość**
- **komputer + oprogramowanie**

SYMULACJE MONTE CARLO - zastosowanie

- 1. wnioskowanie o systemach zbyt skomplikowanych do opisu deterministycznego**
- 2. modelowanie systemów złożonych w różnych warunkach**
- 3. np.**
 - przewidywanie pogody**
 - finanse**
 - genetyka - modelowanie przepływu genów, częstość rekombinacji, ...**
 - statystyka - testowanie hipotez, określanie dokładności estymatorów, określanie wielkości błędów I-go i II-go rodzaju**
 - ...**

SYMULACJE MONTE CARLO - przebieg symulacji



SYMULACJE MONTE CARLO - przykład

ANALIZA STRUKTURY GENETYCZNEJ POPULACJI

- liczby losowe

generator liczb losowych

- frekwencje alleli
- przeżywalność
- płodność
- podatność na choroby
- liczba pokoleń

program komputerowy

- struktura populacji w pokoleniu N

SYMULACJE MONTE CARLO - przykład

GENEROWANIE GENOTYPÓW I WARTOŚCI CECH

osobnik	liczba losowa1	liczba losowa2	allel 1	allel 2	średni wzrost	odchylenie standardowe	obserwowany wzrost
1	0.84	0.201	2	2	10	-6.035247679	164
2	0.89	0.641	1	1	0	11.04124294	171
3	0.60	0.051	2	2	10	-10.89329089	159
4	0.22	0.472	1	1	10	-4.573496426	165
5	0.98	0.291	2	2	10	8.376377991	178
6	0.58	0.551	1	1	0	3.677837412	164
7	0.71	0.401	1	1	0	5.223046173	165
8	0.56	0.691	1	1	0	-15.26705091	145
9	0.28	0.442	1	1	10	19.64781404	190
10	0.22	0.112	2	2	20	14.71644282	195
frekwencja allelu 1		0.3		efekt allelu 2 na wzrost			10
frekwencja allelu 2		0.7		średni wzrost w populacji			160
				obserwowany średni wzrost			170

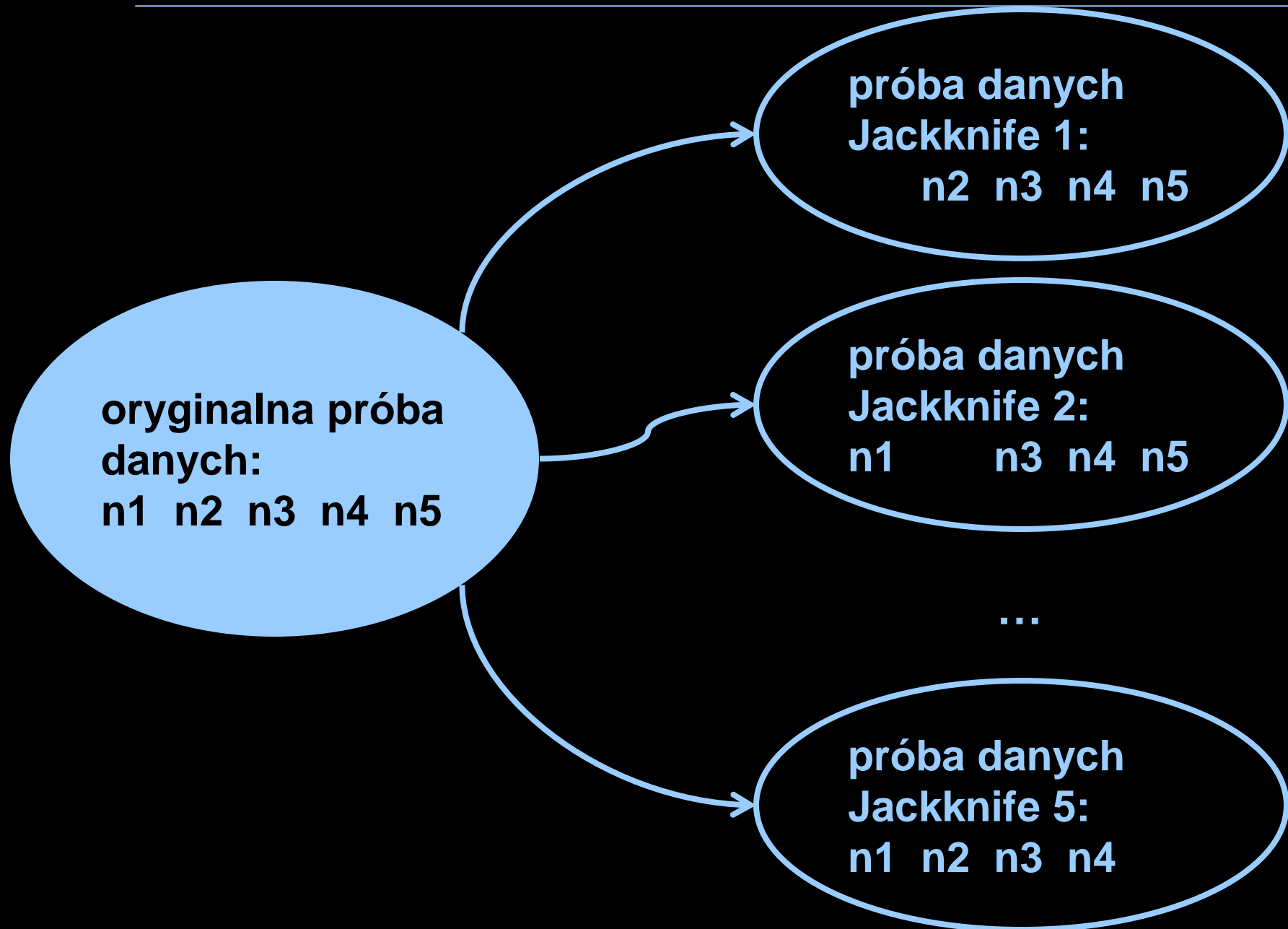


METODY PRÓBKOWANIA

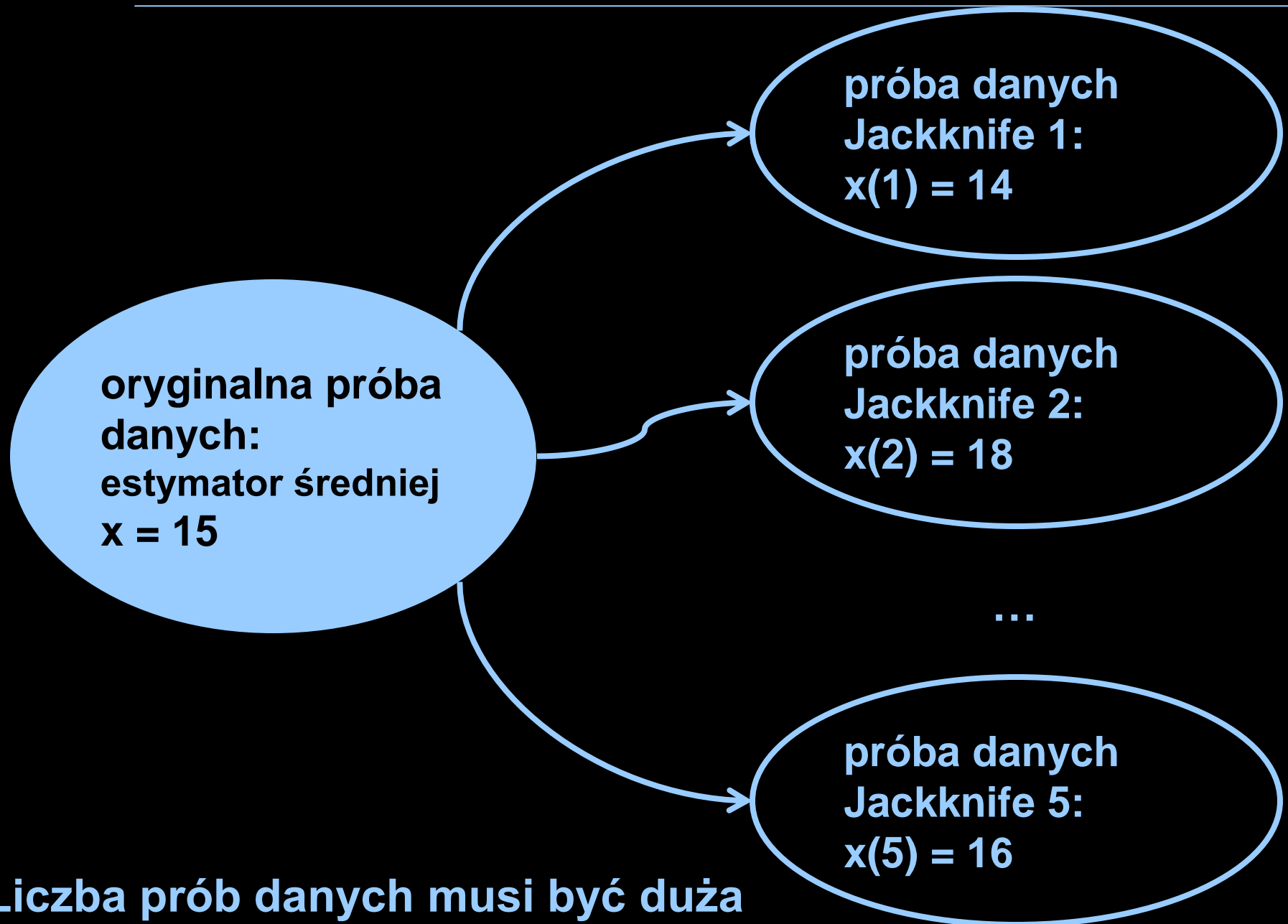
METODY PRÓBKOWANIA

- **Określenie cech estymatora – średnia, wariancja, obciążenie**
- **Alternatywa dla podejścia numerycznego – wyprowadzenie numeryczne**
- **Alternatywa dla podejścia empirycznego – replikacje prób danych**
- **Wykorzystuje pojedynczą, dostępną próbę danych**
- **Jackknife**
- **Bootstrap**

JACKKNIFE



JACKKNIFE

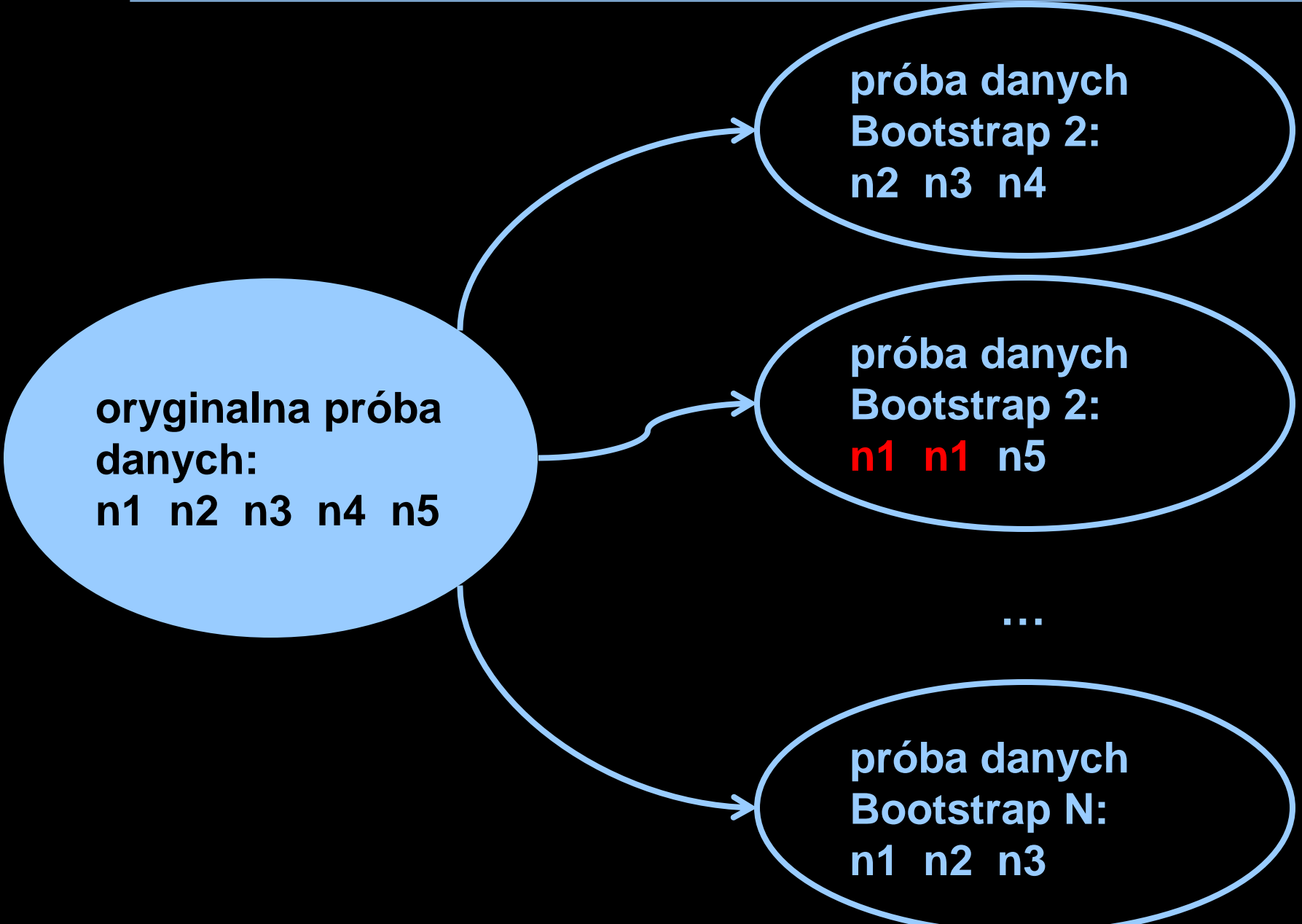


średnia $\bar{x}_J = \frac{1}{5} \sum_{i=1}^5 x_i$

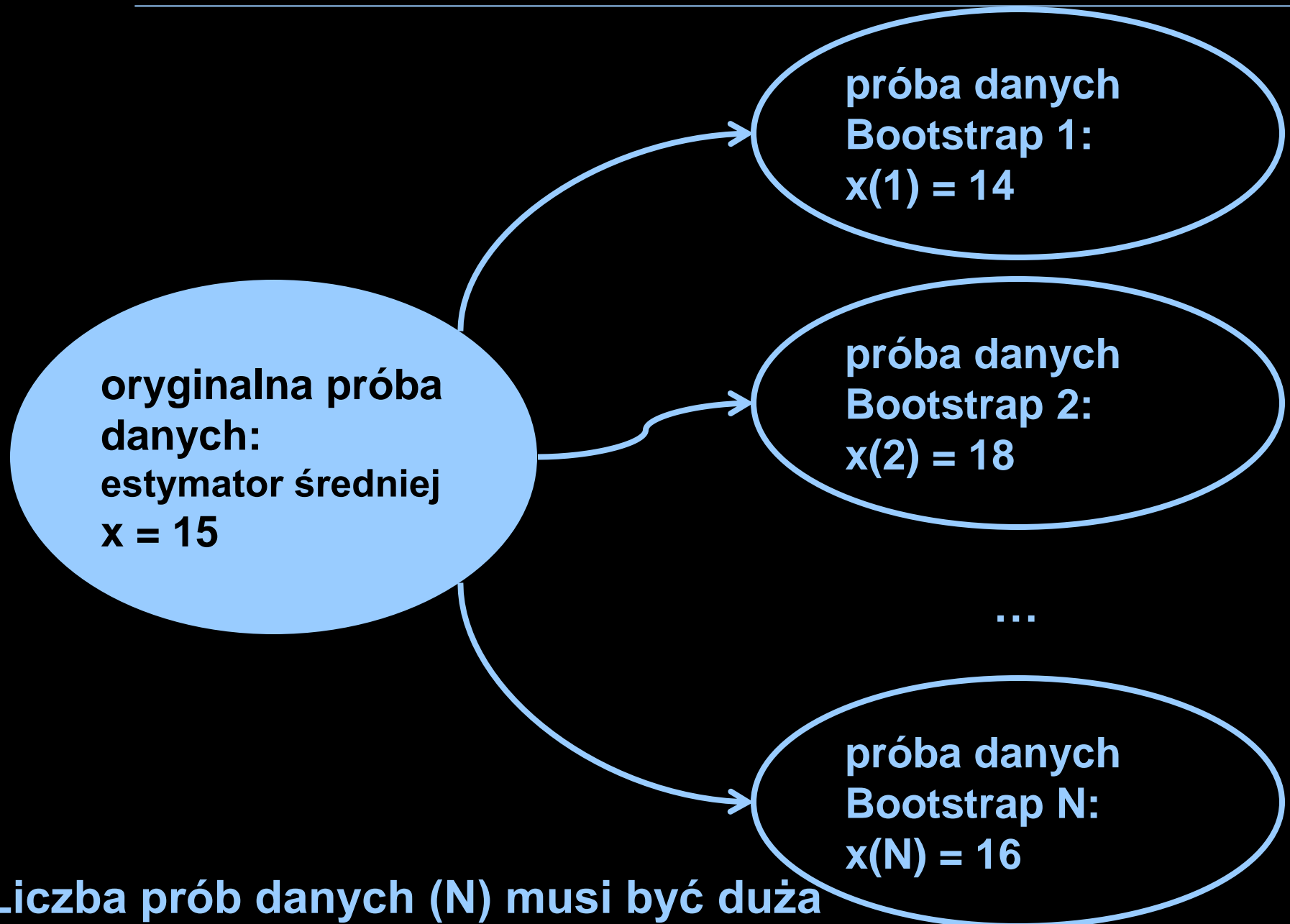
wariancja $\hat{\sigma}_J^2(\bar{x}) = \frac{5-1}{5} \sum_{i=1}^5 (\bar{x}_J - x_i)^2$

obciążenie $(\bar{x}) = (5-1)(\bar{x}_J - \bar{x})$

BOOTSTRAP



BOOTSTRAP



średnia $\bar{x}_B = \frac{1}{N} \sum_{i=1}^N \bar{x}_i$

wariancja $\hat{\sigma}_J^2(\bar{x}) = \frac{1}{N-1} \sum_{i=1}^N (\bar{x}_J - \bar{x}_i)^2$

obciążenie $(\bar{x}) = (\bar{x}_B - \bar{x})$

JACKKNIFE vs BOOTSTRAP

- **Jackknife jest łatwiejszy w implementacji**
- **Dla małych prób danych liczba możliwych "pod"próbek Jackknife jest zbyt niska**
- **Dla dużych prób danych obie metody dają bardzo zbliżone wyniki**

Przykład R

1. Symulacje komputerowe

- **Zasady wykonania analizy asocjacyjnej**
- **Struktura plików wsadowych**
- **Interpretacja wyników**

2. Metody próbkowania

- **Jackknife**
- **Bootstrap**