

Contents

Wzory Matematyczne - 27 Testów Statystycznych dla Analizy Pi	1
NIST Statistical Test Suite (17 testów)	1
Step 01 - Frequency Test (Test Częstotliwości)	1
Step 02 - Runs Test (Test Przejść)	2
Step 03 - Block Frequency Test (Test Częstotliwości w Blokach)	2
Step 04 - Entropy Test (Test Entropii)	2
Step 05 - Spectral FFT Test (Test Spektralny FFT)	2
Step 06 - Compression Test (Test Kompresji)	2
Step 07 - Empirical Entropy Bounds (Empiryczne Granice Entropii)	3
Step 09 - Cumulative Sums Test (Test Skumulowanych Sum)	3
Step 10 - Approximate Entropy Test (Test Przybliżonej Entropii)	3
Step 11 - Serial Test (Test Seryjny)	3
Step 12 - Linear Complexity Test (Test Złożoności Liniowej)	4
Step 13 - Random Excursions Test (Test Losowych Wycieczek)	4
Step 14 - Random Excursions Variant Test (Wariant Testu Losowych Wycieczek)	4
Step 15 - Universal Statistical Test (Test Uniwersalny Statystyczny)	4
Step 16 - Non-overlapping Template Matching Test	4
Step 17 - Overlapping Template Matching Test	5
SmallCrush Test Suite (10 testów)	5
Step 18 - Birthday Spacings Test	5
Step 19 - Collision Test	5
Step 20 - Gap Test	5
Step 21 - Simple Poker Test	5
Step 22 - Coupon Collector Test	5
Step 23 - MaxOft Test	6
Step 24 - Weight Distribution Test	6
Step 25 - Matrix Rank Test	6
Step 26 - Hamming Independence Test	6
Step 27 - Random Walk 1 Test	6
Wzory Ogólne	7
Chi-square Test	7
Z-test	7
Entropia Shannona	7
P-value Interpretation	7
Notacja	7

Wzory Matematyczne - 27 Testów Statystycznych dla Analizy Pi

NIST Statistical Test Suite (17 testów)

Step 01 - Frequency Test (Test Częstotliwości)

Wzór Chi-square: $\chi^2 = \sum_{i=0}^9 (f_i - n/10)^2 / (n/10)$

Gdzie: - f_i = częstotliwość cyfry i (0-9) - n = całkowita liczba cyfr - $E[f_i] = n/10$ = oczekiwana częstotliwość każdej cyfry

P-value: $p\text{-value} = 1 - \text{CDF}(\chi^2, \text{df} = 9)$

Step 02 - Runs Test (Test Przejść)

Oczekiwana liczba runs: $E[\text{runs}] = 2 \cdot \text{ones} \cdot \text{zeros} / n$

Wariancja: $\text{Var}[\text{runs}] = (2 \cdot \text{ones} \cdot \text{zeros} \cdot (2 \cdot \text{ones} \cdot \text{zeros} - n)) / (n^2 \cdot (n - 1))$

Z-score: $Z = (\text{runs} - E[\text{runs}]) / \sqrt{\text{Var}[\text{runs}]}$

P-value: $p\text{-value} = 2 \cdot (1 - \Phi(|Z|))$

Gdzie: - ones = liczba nieparzystych cyfr - zeros = liczba parzystych cyfr - runs = liczba przejść między parzystymi a nieparzystymi

Step 03 - Block Frequency Test (Test Częstotliwości w Blokach)

Chi-square dla bloków: $\chi^2 = \sum_j (\text{ones_per_block}_j - \text{block_size}/2)^2 / (\text{block_size}/2)$

Gdzie: - ones_per_block_j = liczba jedynek w bloku j - block_size = rozmiar bloku - $E[\text{ones}] = \text{block_size}/2$ = oczekiwana liczba jedynek w bloku

P-value: $p\text{-value} = 1 - \text{CDF}(\chi^2, \text{df} = \text{num_blocks})$

Step 04 - Entropy Test (Test Entropii)

Entropia Shannona: $H(X) = - \sum_{x=0}^9 p(x) \cdot \log_2(p(x))$

Gdzie: - $p(x) = \text{count}(x)/n$ = prawdopodobieństwo wystąpienia cyfry x - $H_{\max} = \log_2(10) \approx 3.321928$ = maksymalna entropia dla 10 cyfr

Stosunek entropii: $\text{ratio} = H(X)/H_{\max}$

Step 05 - Spectral FFT Test (Test Spektralny FFT)

Transformata Fouriera: $X[k] = \sum_{n=0}^{N-1} x[n] \cdot e^{-2\pi i k n / N}$

Widmo mocy: $P[k] = |X[k]|^2$

Entropia spektralna: $H_s = - \sum_k (P[k] / \sum P) \cdot \log_2(P[k] / \sum P + \varepsilon)$

Gdzie: - $x[n]$ = pary cyfr ($\text{digits}[i] \cdot 10 + \text{digits}[i+1]$) - $\varepsilon = 10^{-10}$ = mała stała dla stabilności numerycznej

Step 06 - Compression Test (Test Kompresji)

Współczynnik kompresji: $\text{compression_ratio} = \text{compressed_size} / \text{original_size}$

Gdzie: - original_size = rozmiar oryginalnych danych (cyfry jako string) - compressed_size = rozmiar po kompresji zlib (level 9)

Interpretacja: Niższy współczynnik = większa losowość

Step 07 - Empirical Entropy Bounds (Empiryczne Granice Entropii)

Model entropii: $H(N) = \log_2(10) \cdot (1 - c/\log(N))$

Gdzie: - N = liczba analizowanych cyfr - c = parametr dopasowania (estymowany przez curve fitting) - $H_{\max} = \log_2(10) \approx 3.321928$

Dopasowanie modelu: $c = \arg \min \sum (H_{\text{observed}}(N) - H_{\text{model}}(N, c))^2$

Confidence interval (95%): $CI = c \pm 1.96 \cdot \sigma_c$

Step 09 - Cumulative Sums Test (Test Skumulowanych Sum)

Forward cumulative sum: $S_{\text{forward}}[i] = \sum_{j=0}^i (2 \cdot \text{binary}[j] - 1)$

Backward cumulative sum: $S_{\text{backward}}[i] = \sum_{j=i}^n (2 \cdot \text{binary}[j] - 1)$

Maksymalna wartość: $\max_{\text{forward}} = \max |S_{\text{forward}}[i]|$ $\max_{\text{backward}} = \max |S_{\text{backward}}[i]|$

Oczekiwana wartość: $E[\max] \approx 0.95 \cdot \sqrt{n}$

Z-score: $Z_{\text{forward}} = \max_{\text{forward}} / \sqrt{n}$ $Z_{\text{backward}} = \max_{\text{backward}} / \sqrt{n}$

P-value: $p\text{-value} = \min(1 - \Phi(Z_{\text{forward}}), 1 - \Phi(Z_{\text{backward}}))$

Gdzie: - $\text{binary}[j] = \text{digits}[j] \bmod 2 = \text{konwersja na binarną (0 lub 1)}$

Step 10 - Approximate Entropy Test (Test Przybliżonej Entropii)

Przybliżona entropia: $\text{ApEn}(m) = \phi(m) - \phi(m+1)$

Gdzie: $\phi(m) = -\sum_{\text{pattern}} p(\text{pattern}) \cdot \log_2(p(\text{pattern}))$

P-value (uproszczony): $\chi^2 = (\text{ApEn} - 0)^2 / 0.1$ $p\text{-value} = 1 - \text{CDF}(\chi^2, \text{df} = 1)$

Gdzie: - m = długość wzorca (zwykle $m = 2$) - $p(\text{pattern})$ = częstotliwość wzorca w sekwencji

Step 11 - Serial Test (Test Seryjny)

Chi-square dla par: $\chi^2_{\text{pairs}} = \sum_{\text{pair}} (\text{count}_{\text{pair}} - \text{expected})^2 / \text{expected}$

Gdzie: - $\text{expected} = n_{\text{pairs}} / 100$ (100 możliwych par: 00-99) - $n_{\text{pairs}} = n - 1$ = liczba par

Chi-square dla trójek: $\chi^2_{\text{triplets}} = \sum_{\text{triplet}} (\text{count}_{\text{triplet}} - \text{expected})^2 / \text{expected}$

Gdzie: - $\text{expected} = n_{\text{triplets}} / 1000$ (1000 możliwych trójek: 000-999) - $n_{\text{triplets}} = n - 2$ = liczba trójek

P-value: $p\text{-value} = \min(p\text{-value}_{\text{pairs}}, p\text{-value}_{\text{triplets}})$

Step 12 - Linear Complexity Test (Test Złożoności Liniowej)

Algorytm Berlekamp-Massey: $L = \text{Berlekamp-Massey}(\text{sequence})$

Oczekiwana złożoność: $E[L] = M/2 + (9 + (-1)^{M+1})/36$

Gdzie: - M = długość bloku (zwykle $M = 500$) - L = długość najkrótszego LFSR generującego sekwencję

Chi-square test: $\chi^2 = \sum (\text{observed_complexity} - \text{expected})^2 / \text{expected}$

Step 13 - Random Excursions Test (Test Losowych Wycieczek)

Random walk: $S[i] = \sum_{j=0}^i (2 \cdot \text{binary}[j] - 1)$

Oczekiwana liczba odwiedzin stanu x: $E[\text{visits}(x)] = 1/(2|x|)$ dla $|x| \geq 1$ $E[\text{visits}(0)] = 0.5$

Prawdopodobieństwa: $p_0 = 1 - E[\text{visits}]$ $p_1 = E[\text{visits}] \cdot (1 - E[\text{visits}])$ $p_{2+} = E[\text{visits}]^2$

Chi-square: $\chi^2 = \sum (\text{observed} - \text{expected})^2 / \text{expected}$

Gdzie: - $x \in \{-4, -3, -2, -1, 1, 2, 3, 4\}$ = stany do analizy

Step 14 - Random Excursions Variant Test (Wariant Testu Losowych Wycieczek)

Oczekiwana liczba odwiedzin: $E[\text{visits}(x)] = n/(2|x|)$ dla $|x| \geq 1$ $E[\text{visits}(0)] = n/2$

Chi-square: $\chi^2 = (\text{observed} - \text{expected})^2 / \text{expected}$

Gdzie: - $x \in \{-9, -8, \dots, -1, 1, 2, \dots, 9\}$ = stany do analizy - n = liczba kroków w random walk

Step 15 - Universal Statistical Test (Test Uniwersalny Statystyczny)

Statystyka testowa: $f_n = (1/K) \cdot \sum \log_2(\text{distance})$

Gdzie: - distance = odległość między powtórzeniami wzorca - K = liczba bloków

Oczekiwana wartość (dla L=6): $E[f_n] \approx 5.2177052$ $\text{Var}[f_n] \approx 2.954$

Z-score: $Z = (f_n - E[f_n]) / \sqrt{\text{Var}[f_n]/K}$

P-value: $p\text{-value} = 2 \cdot (1 - \Phi(|Z|))$

Step 16 - Non-overlapping Template Matching Test

Oczekiwana liczba wystąpień: $E[\text{matches}] = n/(2^m + m - 1)$

Gdzie: - m = długość wzorca binarnego - n = długość sekwencji binarnej

Chi-square: $\chi^2 = (\text{matches} - E[\text{matches}])^2 / E[\text{matches}]$

P-value: $p\text{-value} = 1 - \text{CDF}(\chi^2, \text{df} = 1)$

Step 17 - Overlapping Template Matching Test

Oczekiwana liczba wystąpień: $E[\text{matches}] = (n - m + 1)/2^m$

Gdzie: - m = długość wzorca binarnego - n = długość sekwencji binarnej

Chi-square: $\chi^2 = (\text{matches} - E[\text{matches}])^2 / E[\text{matches}]$

P-value: $p\text{-value} = 1 - \text{CDF}(\chi^2, \text{df} = 1)$

SmallCrush Test Suite (10 testów)

Step 18 - Birthday Spacings Test

Rozkład odstępów: $P(\text{spacing} = k) = (1 - 1/m)^k \cdot (1/m)$

Gdzie: - m = zakres wartości (10 dla cyfr 0-9) - spacing = odległość między powtarzającymi się wartościami

Chi-square: $\chi^2 = \sum (\text{observed_spacings} - \text{expected})^2 / \text{expected}$

Step 19 - Collision Test

Oczekiwana liczba kolizji: $E[\text{collisions}] = t - m + m \cdot (1 - 1/m)^t$

Gdzie: - t = liczba próbek - m = zakres wartości (10 dla cyfr 0-9) - collisions = liczba powtórzeń wartości

Chi-square: $\chi^2 = (\text{collisions} - E[\text{collisions}])^2 / E[\text{collisions}]$

Step 20 - Gap Test

Rozkład geometryczny: $P(\text{gap} = k) = (1 - p)^k \cdot p$

Gdzie: - $p = 1/m$ = prawdopodobieństwo wystąpienia wartości docelowej - m = zakres wartości (10 dla cyfr 0-9)

Chi-square: $\chi^2 = \sum (\text{observed_gaps} - \text{expected})^2 / \text{expected}$

Step 21 - Simple Poker Test

Prawdopodobieństwo k unikalnych wartości w bloku 5: $P(k \text{ unique}) = C(5, k) \cdot P(\text{permutation}) / 10^5$

Gdzie: - $C(5, k)$ = kombinacja 5 po k - $P(\text{permutation})$ = prawdopodobieństwo permutacji

Chi-square: $\chi^2 = \sum_{k=1}^5 (\text{observed}(k) - \text{expected}(k))^2 / \text{expected}(k)$

Step 22 - Coupon Collector Test

Oczekiwana długość: $E[\text{length}] = m \cdot H_m$

Gdzie: - $H_m = \sum_{k=1}^m 1/k$ = liczba harmoniczna - $m = 10$ = liczba różnych wartości (cyfry 0-9)

Z-test: $Z = (\text{observed_mean} - E[\text{length}]) / (\text{std} / \sqrt{n_{\text{trials}}})$

P-value: $p\text{-value} = 2 \cdot (1 - \Phi(|Z|))$

Step 23 - MaxOfT Test

Rozkład maksimum: $P(\max \leq k) = (k/9)^t$

Gdzie: - t = liczba próbek w grupie (zwykle $t = 5$) - $k \in \{0, 1, 2, \dots, 9\}$ = wartości cyfr

Prawdopodobieństwo: $P(\max = k) = (k/9)^t - ((k-1)/9)^t$

Chi-square: $\chi^2 = \sum (\text{observed} - \text{expected})^2 / \text{expected}$

Step 24 - Weight Distribution Test

Oczekiwana suma: $E[\text{sum}] = \text{block_size} \cdot 4.5$

Gdzie: - block_size = rozmiar bloku (zwykle 10) - 4.5 = średnia cyfr 0-9

Z-test: $Z = (\text{observed_mean} - E[\text{sum}]) / (\text{std} / \sqrt{n_{\text{blocks}}})$

P-value: $p\text{-value} = 2 \cdot (1 - \Phi(|Z|))$

Step 25 - Matrix Rank Test

Ranga macierzy: $\text{rank} = \text{matrix_rank}(\text{binary_matrix})$

Gdzie: - binary_matrix = macierz binarna 32×32 utworzona z sekwencji binarnej

Test: Sprawdza rozkład rang macierzy binarnych (większość powinna mieć pełną rangę)

Chi-square: $\chi^2 = \sum (\text{observed_ranks} - \text{expected})^2 / \text{expected}$

Step 26 - Hamming Independence Test

Rozkład binarny: $P(\text{weight} = k) = C(\text{block_size}, k) \cdot 0.5^{\text{block_size}}$

Gdzie: - weight = liczba jedynek w bloku binarnym - block_size = rozmiar bloku (zwykle 32) - $C(n, k)$ = kombinacja n po k

Oczekiwana waga: $E[\text{weight}] = \text{block_size} / 2$

Chi-square: $\chi^2 = \sum (\text{observed_weights} - \text{expected})^2 / \text{expected}$

Step 27 - Random Walk 1 Test

Random walk: $S[i] = \sum_{j=0}^i (2 \cdot \text{binary}[j] - 1)$

Gdzie: - $\text{binary}[j] = \text{digits}[j] \bmod 2$ = konwersja na binarną - $S[i]$ = skumulowana suma (random walk)

Test: Sprawdza rozkład wartości $S[i]$ (powinien być normalny dla losowej sekwencji)

Wzory Ogólne

Chi-square Test

$$\chi^2 = \sum (\text{observed} - \text{expected})^2 / \text{expected} \quad p\text{-value} = 1 - \text{CDF}(\chi^2, \text{df})$$

Z-test

$$Z = (\text{observed} - \text{expected}) / (\text{std} / \sqrt{n}) \quad p\text{-value} = 2 \cdot (1 - \Phi(|Z|))$$

Entropia Shannona

$$H(X) = - \sum p(x) \cdot \log_2(p(x))$$

P-value Interpretation

- $p\text{-value} \geq 0.01 \rightarrow$ **PASS** (sekwencja jest losowa)
 - $p\text{-value} < 0.01 \rightarrow$ **FAIL** (sekwencja nie jest losowa)
-

Notacja

- \sum = suma
 - $E[\dots]$ = wartość oczekiwana
 - $\text{Var}[\dots]$ = wariancja
 - $C(n, k)$ = kombinacja n po k
 - $\Phi(x)$ = dystrybuenta rozkładu normalnego
 - $\text{CDF}(x, \text{df})$ = dystrybuenta rozkładu chi-square
 - $\log_2(x)$ = logarytm o podstawie 2
 - \sqrt{x} = pierwiastek kwadratowy
 - $|x|$ = wartość bezwzględna
-

Data utworzenia: 2024

Liczba testów: 27 (17 NIST + 10 SmallCrush)

Zastosowanie: Analiza statystyczna 10 miliardów cyfr liczby Pi