**Laboratorium SJ.**

**Lab. 2 - Elementy statystyki**

**Zadania do wykonania w środowisku Matlab**

1) Założenie: prawdopodobieństwo, że wylosowana do próbki jednostka jest niezgodna wynosi p. Napisać funkcję obliczającą prawdopodobieństwo, że w próbce o liczności N wystąpi co najmniej K elementów niezgodnych oraz wykreślić zależność wartości tej funkcji od K (K=1..N) dla zadanych wartości p i N (np. p=0.05, N=50). Wejściem funkcji są zmienne p, N i K ( $0<p<1$, $K \le N$). Wzór pomocniczy: prawdopodobieństwo wystąpienia dokładnie K elementów niezgodnych wynosi:

$$P(K) = \binom{N}{K} \cdot p^K \cdot (1-p)^{N-K}$$

2) X jest zmienną losową o rozkładzie normalnym ze średnią μ i odchyleniem standardowym σ. Napisać fukcję wyznaczającą prawdopodobieństwo P zdarzenia, że $\left|\frac{X-\mu}{\sigma}\right| > k$, gdzie *k* jest zadaną liczbą. Napisać funkcję, która dla zadanego prawdopodobieństwa P zdarzenia, że $\left|\frac{X-\mu}{\sigma}\right| > k$ oblicza wartość *k*.

3) Sprawdzenie normalności rozkładu danych (lub zgodności danych z dowolnym typem rozkładu).
Dla dowolnej ustalonej liczby p, $p \in (0,1)$, kwantylem rzędu p rozkładu prawdopodobieństwa o dystrybuancie F(x) nazywamy taką wartość $x_p$, że $F(x_p) = p$. Jeżeli dokonamy ustawienia losowych wyników pomiarów $X_1$ …. $X_N$ w porządku niemalejącym: $X_{(1)} \le X_{(2)} \le X_{(r)} \le \ldots \le X_{(N)}$ oraz dokonamy standaryzacji zmiennej losowej X, tak aby wartość średnia wynosiła 0, a odchylenie standardowe 1, to wartość oczekiwana statystyki pozycyjnej $X_{(r)}$ może być przybliżona przez kwantyl $K_{(r-0.5)/N}$ rzędu (r-0.5)/N standardowej zmiennej losowej Z o rozkładzie normalnym N(0,1). Zatem jeśli przebieg wykresu składa się z punktów { $X_{(r)}$ , $K_{(r-0.5)/N}$}, to jeżeli próbka pochodzi z rozkładu normalnego, przebieg wykresu powinien być zbliżony do linii prostej. Odstępstwo od linii prostej sugeruje, że zmierzona próbka nie pochodzi z rozkładu normalnego. Podobnie testuje się zgodność rozkładu próbki dla innych typów rozkładów.
Napisać funkcję, która dla zadanej próbki losowej o liczności N wykona wykresy sprawdzające, czy próbka pochodzi z rozkładu normalnego. Funkcję przetestować dla próbek symulowanych z rozkładu normalnego (randn) i równomiernego (rand) dla różnych liczności próbek N (np. N = 100, $10^4$, $10^6$).

4) Estymatory odchylenia standardowego

Na podstawie N-próbek K-elementowych można wyznaczyć następujące estymatory nieobciążone odchylenia standardowego:

$\sigma_1 = a(K) \cdot \bar{s}$, gdzie $\bar{s}$ - średnie odchylenie standardowe wyznaczone z N-odchyleń standardowych pojedynczych próbek, a(K) – współczynnik zależny od liczności próbki K.

$\sigma_2 = b(K) \cdot \bar{R}$, gdzie $\bar{R}$ jest rozstępem średnim z N próbek K-elementowych, b(K) – współczynnik zależny od liczności próbki K.

$\sigma_3 = c(K) \cdot med(s)$, gdzie *med(s)* jest medianą odchyleń standardowych z N-próbek K-elementowych, *c(K)* – współczynnik zależny od liczności próbek K.

Współczynniki a(K), b(K) i c(K) zawarto w poniższej tabeli:

| **Liczność próbki K** | **a(K)** | **b(K)** | **c(K)** |
|---|---|---|---|
| 2 | 1.253 | 0.8865 | 1.1829 |
| 3 | 1.1284 | 0.5907 | 1.0646 |
| 4 | 1.0854 | 0.4857 | 1.0374 |
| 5 | 1.0638 | 0.4300 | 1.0260 |
| 6 | 1.0510 | 0.3946 | 1.0201 |
| 7 | 1.0423 | 0.3698 | 1.0161 |
| 8 | 1.0363 | 0.3512 | 1.0136 |
| 9 | 1.0317 | 0.3367 | 1.0116 |
| 10 | 1.0281 | 0.3249 | 1.0103 |
| 15 | 1.0180 | 0.2880 | 1.0063 |
| 20 | 1.0133 | 0.2677 | 1.0046 |

Napisać funkcję wyznaczającą opisane estymatory odchylenia standardowego dla N-próbek K-elementowych dla K=2..20. Wartości współczynników dla brakujących w tabeli liczności próbek wyznaczyć z zastosowaniem interpolacji.
Dla wybranej liczności próbki K przetestować symulacyjnie, który spośród trzech estymatorów odchylenia standardowego jest najefektywniejszy i wyrazić efektywność pozostałych dwóch względem niego.

5) Wygenerować wektor liczb losowych z rozkładu normalnego o wartości oczekiwanej $\mu$ i odchyleniu standardowym $\sigma$ (wartości parametrów rozkładu przyjąć arbitralnie). Na jego podstawie wykreślić histogram (doświadczalną funkcję gęstości) oraz dystrybuantę. Do wykonania wykresów użyć funkcji *bar()* i wyników odpowiednio wywołanej funkcji hist() lub histcounts() – które wyznaczają położenia binów i liczby zliczeń w binach. Znormalizować histogram (suma powierzchni słupków histogramu znormalizowanego powinna być równa jedności) i przedstawić go na wykresie razem z teoretyczną funkcją gęstości (normpdf). Na podstawie znormalizowanego histogramu wykreślić dystrybuantę. Porównać na drugim wykresie dystrybuantę doświadczalną z dystrybuantą teoretyczną (normcdf). Histogramy i dystrybuanty doświadczalne wykreślić dla kilku różnych długości wektora liczb losowych.