**Analiza statyczna promieniowania, testowanie hipotez statystycznych**
Michał Urbański

**Ćwiczenie składa się z trzech zadań:**

(z1) Weryfikacja hipotezy, że dla krótkiego czasu zliczania (gdy liczba jedynek i dwójek jest dużo mniejsza niż liczba zer) liczby zliczeń opisuje rozkład Poissona.

(z2) Weryfikacja hipotezy, że dla dużych liczby zliczeń rozkład może być opisany rozkładem Gaussa. Należy zrobić weryfikację hipotezy zarówno dla rozkładu Gaussa jak i Poissona

(z3) Sprawdzenie testem chikwadrat, że statystyka $\chi^2$ ma rozkład $\chi^2$. W tym celu wykonujemy pomiar liczby zliczeń dla krótkiej bramki (7ms) 100 serii po 500 zliczeń.

**Wykonanie ćwiczenia. Ćwiczenie ma dwie części:**

1. pomiar liczby zliczeń dla krótkiej bramki, $L = 100$ razy po $N = 500$ zliczeń. Zanotować liczbę zer, jedynek, dwójek itd. dla każdej serii.

2. dla długiej bramki $L = 2$ razy wykonać pomiary liczby zliczeń dla $N = 1000$ powtórzeń w każdej serii, zanotować liczbę zer, jedynek itd.

Pojedyńczym zdarzeniem losowym (próbą losową lub pojedyńczym pomiarem) jest liczba zliczeń w czasie $\tau$, który nazywamy czasem bramki. Komputer przeprowadza takich pomiarów $N$ (np. $N$ =500) i bezpośrednim wynikiem takiej obserwacji jest ciąg liczb $(k_1, k_2, \cdots, k_i, \cdots, k_N)$, gdzie $k_i$ jest liczbą rozpadów zarejestrowaną w $i$–tym pomiarze (w czasie $\tau$). Dla małej bramki i $N$ =10 przykładem jest następujący ciąg obserwacji: (0,0,0,1,2,0,0,1,0,0). Komputer nie podaje całego ciągu liczb ale wyświetla jedynie liczbę $n_k$ wystąpień $k$ zliczeń. Dla przykładu (0,0,0,1,2,0,0,1,0,0) mamy: $n_0 = 7$, $n_1 = 2$, $n_2 = 1$, $n_3 = 0$ (liczba trójek jest zero), oraz $\sum_k n_k = n_0+n_1+n_2+n_3 = 10$ . Wynikiem każdej serii obserwacji jest więc ciąg liczb $n_k$ opisujący liczbę zer, jedynek, dwójek itd.($k = 0, 1, \cdots .K$, $K$–maksymalna liczba zliczeń). Oczywiście $\sum_k n_k = N$. Całkowita liczba zdarzeń w $L$ seriach pomiarowych wynosi więc $L \bullet N$

Liczba zliczeń $k$ w pojedyńczym pomiarze opisywana jest rozkładem Poissona: $P(k) = \frac{m^k}{k!}e^{-m}$, gdzie $P(k)$ –prawdopodobieństwo uzyskania $k$ zliczeń, $m$–wartość oczekiwana (pierwszy moment rozkładu).

Dla dużych $k$ rozkład ten w granicy zbliża się do rozkładu normalnego.

**Opracowanie wyników pomiarów:**

**1.** Dla pomiarów liczby zliczeń dla krótkiej bramki:

(a) wyznaczyć statystykę $\chi^2$ dla każdej serii pomiarowej (N=100 wartości $\chi^2$):

$$\chi^2 = \sum_{l=0}^{B} \frac{(n_l - Np_l)^2}{Np_l} \tag{1}$$

gdzie $l$ numeruje biny (koszyki), $B$–liczba binów.

Ponieważ warunkiem przy którym statystyka $\chi^2$ na rozkład $\chi^2$ jest to aby $n_k$ miało rozkład normalny a to wymaga aby $n_k > 10$. Wobec tego podzielimy wyniki na dwie grupy: $k = 0$ i $k > 0$, jeśli przypadek $k = 0$ oznaczymy $l = 0$ i przypadek $k > 0$ jako $l = a$ to: $p_0 = P(0) = e^{-m}$ i $p_a = P(k > 0) = 1 - p_0$. Wtedy w równaniu (1) są dwa człony. Wyrażenie (1) należy wyliczyć $N$ razy, uzyskamy więc $N$ wartości statystyki $\chi^2$, zmienną tą dla każdej $i$ tej serii oznaczymy $\chi_i^2$. Należy sprawdzić $N$ razy czy mamy powód odrzucić hipotezę na poziomie istotności $p = 0,95$. Sprawdzić czy rzeczywiści liczba odrzuceń będzie wynosić 5% całkowitej liczby serii pomiarowych.

**Uwaga**: Wartość średnią $m$ wyliczyć raz dla wszystkich $L \bullet N$ danych: $m = \frac{\sum_{k=0}^{K} kn_k}{\sum_{k=0}^{K} n_k}$ gdzie: $K$ jest maksymalna wartościa zliczeń, $k = 0, 1, \cdots, K$. Tutaj $n_k = n_{k,i}$, gdzie $n_{k,i}$–liczna zliczeń w $k$–tym pomiarze $i$–tej serii.

(b) Dla każdej serii pomiarowej $\chi_i^2$ jest inne bowiem wielkość $\chi^2$ (statystyka $\chi^2$) wyliczona z danych doświadczalnych (wzór 1) jest zmienną losową i należy testem $\chi^2$ sprawdzić czy statystyka $\chi^2$ ma rozkład $\chi^2$. W tym celu należy wyliczyć wartość $\chi^2$ ze wzoru (1) dla każdej serii danych pamiętając o tym, że:

(1) trzeba dobrać koszyki tak aby każdym było przynajmniej 10 zliczeń. Należy to zrobić tak, że dane $x_i^2$ (w naszym przypadku $\chi_i^2$.) porządkuje się (w arkuszu kalkulacyjnym można porządkować jedynie liczby więc należy w osobnym miejscu wkleić obliczone wartości $\chi_i^2$ jako liczby). Uporządkowany ciąg $x_i^2$ danych dzieli się na koszyki po 10 w każdym (czyli 10 binów po 10 serii liczbowych), granice binów oznaczamy $a_l$, gdzie $l = 0, \cdots, L$

(2) prawdopodobieństwo $p_l$ dla każdego koszyka opisane jest całką:

$$p_l = \int_{a_{l-1}}^{a_l} f(x)dx = F(a_l) - F(a_{l-1}) \tag{2}$$

$f(x)$ jest gęstością rozkładu prawdopodobieństwa (hipoteza) a $F(x)$ jego dystrybuantą, dystrybuanty dane są w programach typu arkusz kalkulacyjny (lub innych).

(c) Dla serii danych $\chi_i^2$ narysowac histogram i porównac na rysunku z gęstością prawdopodobieństwa rozkładu $\chi^2$, w tym celu należy dane podzielić na ok. 10 równych przedziałów.

**2.** Dla długiej bramki $L = 2$ razy sprawdzić czy otrzymane dane mogą być opisane rozkładem Gaussa czy tez trzeba je opisać rozkładem Poissona. W tym celu należy:

(a) Wykreślić histogram uzyskanych danych, czy zależność $n_k$ od $k$ razem z wykresem teoretycznym zarówno rozkładu Pissona jak i rozkładu Gaussa. Przy wyznaczaniu wartości $P(k)$ rozkładu Gaussa należy pamiętać, że jest to rozkład ciągły a badany rozkład jest rozkładem dyskretnym. (b) Wyliczyć wartość statystyki $\chi^2$ (dla $L = 1$ i 2) ze wzoru (1) pamiętając o tym co jest napisane w punkcie 1b) i sprawdzić czy są powody aby hipotezę o rozkładzie normalnym (i Poissona) odrzucić.