Zestaw 6 / Metody analizy danych (27.11.08):

1. (RN 7.2.3) Dysponujemy dwiema niezależnymi ocenami pewnego kąta θ: wartością c dla $\cos\theta$ i wartością s dla $\sin\theta$, przy czym obie oceny charakteryzują się rozkładem gaussowskim z tą samą, znaną dyspersją σ.
a) Znajdź metodą największej wiarogodności estymator kąta θ i jego niepewność.
b) Rozwiąż to samo zadanie, gdy dyspersja σ nie jest znana. Znajdź estymator dyspersji.
Odp. $\hat{\theta} = \arctan\left(\frac{s}{c}\right) \pm \frac{\sigma}{\sqrt[4]{s^2+c^2}}$.

2. (RN 7.2.14) Zakładając jednorodny rozkład gwiazd w naszym najbliższym otoczeniu, tzn. przyjmując, że liczba gwiazd w zadanej objętości V przestrzeni kosmicznej podlega rozkładowi Poissona $P_k(\lambda k)$ z nieznanym parametrem λ określającym oczekiwaną liczbę gwiazd w jednostce objętości,
a) znajdź rozkład $f_n(r)$ odległości od pierwszej, drugiej, ... , n-tej gwiazdy od Słońca.
b) Wykorzystując te rozkłady jak również dane o odległościach (w parsekach) od Słońca do najbliższych 31 gwiazd (wg. E.Rybka, Astronomia Ogólna, PWN, 1983): 1.31, 1.32, 1.32, 1.81, 2.32, 2.49, 2.65, 2.65, 2.74, 2.74, 2.9, 3.15, 3.28, 3.31, 3.32, 3.42, 3.42, 3.44, 3.48, 3.48, 3.52, 3.52, 3.55, 3.55, 3.58, 3.66, 3.76, 3.85, 3.91, 3.94, 3.94, znajdź estymatęnajwiększej wiarogodności parametru λ i jej niepewność.
c) Porównaj z wynikiem uzyskanym bezpośredniego podzielenia liczby gwiazd w kuli o objętości czterech parseków przez objętość tej kuli. Jak należy określić niepewność estymaty wielkości λ uzyskanej w tym podejściu?
Odp. b) $\hat{\lambda} = 0.113 \pm 0.005$, c) $\hat{\lambda} = 0.116 \pm 0.021$.

3. Napisz program generujący próbki o liczebności n wielkości t podlegającej rozkładowi wykładniczemu
$$f(t;\tau) = \frac{1}{\tau}e^{-t/\tau}, \quad t \geq 0$$
Estymatorem największej wiarygodności parametru τ jest $\hat{\tau} = \bar{t}$ a estymatorem jego wariancji jest $\hat{\tau}^2/n$.
a) Wygeneruj 1000 próbek z $\tau = 1$ i $n = 5, 10, 100$. Oblicz wartość estymatora $\hat{\tau}$ dla każdej próbki a wyniki dla wszystkich 1000 próbek przedstaw w postaci histogramu. Porównaj wariancje otrzymane z metody Monte Carlo z wynikami analitycznymi. Czy wartość średnia estymatorów $\hat{\tau}$ jest konsystentna z prawdziwą wartością $\tau = 1$?
b) Załóżmy, że rozkład prawdopodobieństwa został sparametryzowany poprzez parametr $\lambda = 1/\tau$
$$f(t;\lambda) = \lambda e^{-\lambda t}, \quad t \geq 0$$
Estymatorem największej wiarygodności parametru λ jest $\hat{\lambda} = 1/\bar{t}$. Zmodyfikuj program z (a) tak aby histogramował również estymatory $\hat{\lambda}$. Jak wiadomo estymator największej wiarygodności $\hat{\lambda}$ jest obciążony. Z wyników symulacji Monte Carlo oblicz wartości obciążenia $b = E[\hat{\lambda}] - \lambda$. Porównaj ze spodziewanymi wartościami dla $n = 5, 10, 100$

4. Rozważ N niezależnych zmiennych $n_1, ..., n_N$ podlegających rozkładowi Poissona o wartościach średnich równych odpowiednio $\mu_1, ..., \mu_N$. Załóżmy, że wartości średnie zależą od zmiennej x z godnie ze wzorem

$$\mu(x) = \theta a(x),$$

gdzie θ jest nieznanym parametrem a $a(x)$ jest dowolną znaną funkcją. N wartości μ_i dane są więc zależnością $\mu(x_i) = \theta a(x_i)$, gdzie wartości x_i są znane. Pokaż, że estymatorem największej wiarygodności parametru θ jest

$$\hat{\theta} = \frac{\sum_{i=1}^{N} n_i}{\sum_{i=1}^{N} a(x_i)}.$$

Pokaż, że $\hat{\theta}$ jest estymatorem nieobciążonym a jego wariancja jest równa tej otrzymanej z warunku Cramera-Rao.

5. Przykładem sytuacji opisanej w poprzednim zadaniu są dane z rozpraszania (anty)neutrin na nukleonie. W modelu kwarkowo-partonowym, przekroje czynne na procesy $\nu N \to \mu^- X$ i $\bar{\nu} N \to \mu^+ X$ mają postać

$$\sigma(\nu N \to \mu^- X) = \frac{G^2 M E}{\pi} \left(\langle q \rangle + \frac{1}{3} \langle \bar{q} \rangle \right) \equiv \theta_\nu E$$

$$\sigma(\bar{\nu} N \to \mu^+ X) = \frac{G^2 M E}{\pi} \left(\frac{1}{3} \langle q \rangle + \langle \bar{q} \rangle \right) \equiv \theta_{\bar{\nu}} E$$

gdzie E jest energią wiązki (anty)neutrin, M masą nukleonu-tarczy a G stałą Fermiego. Tutaj zmienna x odpowiada energii E. Pomiary zostały przeprowadzone dla N różnych wartości E. Dla każdej energii, spodziewana liczba przypadków jest dana przez

$$\mu_i = \sigma(E_i) \epsilon(E_i) L_i,$$

gdzie $\sigma(E_i)$ jest przekrojem czynnym dla energii E_i, L_i jest świetlnością a $\epsilon(E_i)$ jest wydajnością detektora na rejestarcję przypadku, która zależy od energii. Załóżmy dla uproszczenia, że energie E_i, świetlności L_i oraz wydajności $\epsilon(E_i)$ są znane bezbłędnie. Znajdź estymatory najwiekszej wiarygodności parametrów θ_ν i $\theta_{\bar{\nu}}$ a na ich podstawie estymatory wielkości $\langle q \rangle$ oraz $\langle \bar{q} \rangle$. W modelu kwarkowo-partonowym wielkości te odpowiadją ułamkowi pędu nukleonu unoszonemu przez odpowiednio kwarki i antykwarki tworzące nukleon. Oblicz ułamek pędu nukleonu unoszony przez partony nie bedące ani kwarkami ani antykwarkami (czyli są gluonami), $\langle g \rangle = 1 - \langle q \rangle - \langle \bar{q} \rangle$.