ZADANIA Z METOD STATYSTYCZNYCH

Zestaw V - na 2.12.2009

1. Niezależne zmienne losowe $x_1, \ldots, x_n$ mają rozkład jednorodny w przedziale $(0, b)$. Podaj estymator najwięszej wiarygodności parametru b.

2. W stawie żyje nieznana liczba (N) ryb, z których K wyławiamy, oznaczamy, po czym wpuszczamy ponownie do stawu. Po pewnym czasie wyławiamy n ryb, wśród których rozpoznajemy k zaznaczonych. Podaj estymator największej wiarygodności dla N.

3. Współrzędne punktów na płaszczyźnie $(x_1^{(j)}, x_2^{(j)})$ mają dwuwymiarowy rozkład normalny o gęstości prawdopodobieństwa

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)} \times \left\{ \frac{(x_1-a_1)^2}{\sigma_1} + \frac{(x_2-a_2)^2}{\sigma_2^2} - 2\rho\frac{(x_1-a_1)(x_2-a_2)}{\sigma_1\sigma_2} \right\}\right].$$

   Znaleźć estymatory największej wiarygodności parametrów rozkładu a_1, a_2, σ_1, σ_2, ρ.

4. Rzucamy n razy monetą niesymetryczną, dla której prawdopodobieństwo wyrzucenia orła wynosi p. Wyniki zapisujemy jako $k_j = 0$ gdy wypadł orzeł lub 1 gdy wypadła reszka ($j = 1, \ldots, n$).
   a) Pokaż, że statystyką wystarczającą dla parametru $\theta = (1-p)^2$ jest $T(k) = \sum_j k_j$, oraz że estymator największej wiarygodności dla θ to $(T/n)^2$.
   b) Estymator największej wiarygodności jest estymatorem obciążonym. Znajdź estymator nieobciążony bedący funkcją statystyki wystarczającej T.

5. Prawdopodobieństwo, że komputer na którym zainstalowano w chwili $t = 0$ pewien popularny system operacyjny ulegnie awarii w przedziale czasu pomiędzy t i $t + dt$, opisane jest wzorem

$$f(t)dt = \lambda \exp(-t/\lambda)dt.$$

   Załóżmy, że zaobserwowano awarie n komputerów w chwilach $t_1, \ldots, t_n$.
   a) Znajdź statystykę wystarczająca dla λ i jej rozkład prawdopodobieństwa.
   b) Znajdź estymator największej wiarygodności dla λ. Pokaż, że jest on obciążony i znajdź wielokrotność tego estymatora będącą estymatorem nieobciążonym.
   c) Znajdź estymator będący wielokrotnością estymatora największej wiarygodności o najmniejszym średnim kwadracie błędu.

   Wskazówka: Rozkład $f(T)$ zmiennej losowej $T = \sum_j t_j$ można wyliczyć jako

$$f(T) = \int dt_1 \ldots dt_n \, \delta\left(T - \sum_j t_j\right) f(t_1) \ldots f(t_n),$$

   gdzie kładziemy $\delta(z) = \int_{-\infty}^{\infty} (dk/2\pi) e^{ikz}$ i najpierw wykonujemy całkowania po $t_1 \ldots t_n$ a potem po k. Wynikiem jest $f(T) = \frac{\lambda^n}{(n-1)!} T^{n-1} e^{-\lambda T}$.