

STATYSTYKA

DLA STUDENTÓW III ROKU

MATEMATYKI EKONOMICZNEJ, MATEMATYKI Z INFORMATYKĄ I MATEMATYKI NAUCZYCIELSKIEJ

Lista 2

1. (Wzór Hardy'ego-Weinberga)¹ Najprostszy model genetyczny zakłada, że cechy dziedziczne występują w genach parami. Geny mogą przybierać dwie formy, nazywane allelami. Oznacza się je literami **A** i **a**. Wobec tego są trzy różne genotypy: **AA**, **Aa**² i **aa**.³ Załóżmy, że pary kojarzą się w sposób losowy i każdy przekazuje potomkowi losowo jeden ze swoich alleli. Oznaczmy przez θ prawdopodobieństwo wystąpienia w pokoleniu 0 w populacji allele **A**. Prawdopodobieństwo wystąpienia allele **a** w pokoleniu 0 wyniesie więc $1 - \theta$.

- (a) Pokaż, że prawdopodobieństwo wystąpienia genotypów **AA**, **Aa** i **aa** jest stałe we wszystkich pokoleniach, począwszy od pierwszego i wynosi

$$p_1 = \theta^2, p_2 = 2\theta(1 - \theta), p_3 = (1 - \theta)^2$$

- (b) Pokaż, że prawdopodobieństwo wystąpienia alleli **A** i **a** jest również stałe we wszystkich pokoleniach i wynosi odpowiednio θ i $1 - \theta$.
 - (c) Oszacuj parametr θ metodą podstawienia częstości na podstawie próby n -elementowej, w której genotyp **AA** pojawił się n_1 razy, genotyp **Aa** pojawił się n_2 razy, genotyp **aa** - n_3 razy. Pokaż, dobierając odpowiednio zdarzenia, że jego estymatorami mogą być

$$\hat{\theta}_1 = \frac{n_1}{n}, \hat{\theta}_2 = 1 - \frac{n_3}{n}, \hat{\theta}_3 = \frac{n_1}{n} + \frac{n_2}{2n}$$

- (d) Przypuśćmy, że parametrem modelu statystycznego jest stosunek ϑ prawdopodobieństwa wystąpienia alleli **A** do prawdopodobieństwa wystąpienia alleli **a**, równy

$$\vartheta = \frac{\theta}{1 - \theta}$$

Znajdź metodą podstawienia częstości estymator ϑ .

2. Niech obserwacje $X^T = [x_1, x_2, \dots, x_n]$ będą próbą prostą z rozkładu $\mathcal{N}(0, \sigma)$ z nieznanym parametrem σ .
 - (a) Znajdź estymator parametru σ^2 metodą momentów, opierając się na drugim momencie,
 - (b) Znajdź estymator parametru σ na podstawie estymatora z punktu (a) i relacji $\sigma = \sqrt{\sigma^2}$,

¹Więcej na ten temat w W. Feller, Wstęp do rachunku prawdopodobieństwa, t. I Rozdział V, §5

²genotypy **Aa** i **aA** są nierozróżnialne

³Na przykład, kolor kwiatu rośliny może zawierać dwa allele: **A** - czerwony i **a** - biały. Genotypy, które odpowiadają tym allelom są: **AA** - czerwony, **Aa** - różowy, **aa** - biały.

- (c) Znajdź, metodą momentów estymator parametru σ , korzystając z tożsamości $E|X| = \sigma \sqrt{\frac{2}{\pi}}$, gdzie X ma rozkład $\mathcal{N}(0, \sigma)$. Udowodnij tę tożsamość.
3. Rzucamy monetą tak długo, aż wypadnie orzeł. Orzeł pojawił się po raz pierwszy po n rzutach. Znajdź estymator największej wiarygodności dla prawdopodobieństwa wypadnięcia orła w jednym rzucie. Wykonaj analogiczne obliczenia, gdybyśmy powtarzali taki eksperyment k -krotnie i orzeł pojawiłby się po raz pierwszy w tych powtórzeniach po $n_1, n_2, \dots, n_k$ rzutach.
4. Niech obserwacje $X^T = [x_1, x_2, \dots, x_n]$ będą próbą prostą z rozkładu jednostajnego na odcinku $[\theta - \frac{1}{2}, \theta + \frac{1}{2}]$. Pokaż, że dowolny estymator $\hat{\theta}(X)$, spełniający warunek

$$x_{(n)} - \frac{1}{2} < \hat{\theta}(X) < x_{(1)} + \frac{1}{2}$$

jest estymatorem największej wiarygodności parametru θ .

5. Obserwacje $X^T = [x_1, x_2, \dots, x_n]$ są realizacjami niezależnych zmiennych losowych X_i o rozkładzie $\mathcal{N}(\mu_i, 1)$
- (a) Znajdź, metodą największej wiarygodności estymatory parametrów μ_i ,
- (b) Wykonaj zadanie (a) przy założeniu, że $n = 2$ i $\mu_1 \leq \mu_2$.
6. Rzucamy monetą aż do uzyskania orła, ale nie więcej, niż M razy. Wynik x takiego doświadczenia będziemy zapisywać w następujący sposób:
- $$x = \begin{cases} j & \text{gdy orzeł wypadł po raz pierwszy za } j\text{-tym razem} \\ M + 1 & \text{gdy orzeł nie wypadł ani razu.} \end{cases}$$

Znajdź estymator największej wiarygodności dla prawdopodobieństwa wypadnięcia orła w jednym rzucie. Znajdź przykłady praktyczne, w których taki model ma zastosowanie.

Andrzej Dąbrowski, 20.10.99