Statystyka Matematyczna

Ćwiczenia

Zadanie 7.6

Mamy próbkę $X_1, ..., X_n$ z rozkładu danego gęstością:

$$f_\theta(x) = \frac{3x^2}{\theta^3} I(0, \theta)(x)$$

Badamy następującą parę hipotez:

$$\begin{cases} H : \theta = \frac{1}{3} \\ K : \theta = \frac{2}{3} \end{cases}$$

Można to zadanie rozwiązać na dwa sposoby.
Pierwszy sposób to skorzystanie z Lematu Neymana-Pearsona. Postać testu:

$$\phi(\mathbb{X}) = \begin{cases} 1 \text{ gdy } f_{\frac{2}{3}}(x_1, ..., x_n) > k f_{\frac{1}{3}}(x_1, ..., x_n) \\ 0 \text{ w.p.p.} \end{cases}$$

Zadanie 7.6

Przekształcamy gęstości:

$$f_{\frac{2}{3}}(x_1, ..., x_n) > kf_{\frac{1}{3}}(x_1, ..., x_n)$$

$$= \frac{3^n \prod_{i=1}^n x_i^2}{\left(\frac{2}{3}\right)^{3n}} I(0, \frac{2}{3})(x_{n:n}) > k \frac{3^n \prod_{i=1}^n x_i^2}{\left(\frac{1}{3}\right)^{3n}} I(0, \frac{1}{3})(x_{n:n})$$

$$\iff \frac{1}{8^n} I(0, \frac{2}{3})(x_{n:n}) > kI(0, \frac{1}{3})(x_{n:n})$$

$$\iff I(0, \frac{2}{3})(x_{n:n}) > k' I(0, \frac{1}{3})(x_{n:n})$$

Zatem dla $\frac{1}{3} < x_{n:n} < \frac{2}{3}$ zawsze odrzucamy, podczas gdy dla $0 < x_{n:n} < \frac{1}{3}$ zawsze odrzucamy jeśli $k' < 1$ lub zawsze przyjmujemy gdy $k' > 1$. Zatem nie możemy znaleźć k' takiego, że:

$$\mathrm{E}_{\frac{1}{3}} \phi(\mathbb{X}) = \mathbb{P}_{\frac{1}{3}}(\phi(\mathbb{X}) = 1) = \mathbb{P}_{\frac{1}{3}}(I(0, \frac{2}{3})(x_{n:n}) > k' I(0, \frac{1}{3})(x_{n:n}))$$

$$= \mathbb{P}_{\frac{1}{3}}(I(0, \frac{1}{3})(x_{n:n}) > k' I(0, \frac{1}{3})(x_{n:n})) = \alpha$$

Zadanie 7.6

Jedyne wyjście w tym przypadku - randomizacja. Przyjmujemy, że $k' > 1$ np. $k' = 1.8$ i rozpatrujemy następujący test:

$$\phi(\mathbb{X}) = \begin{cases} \alpha \text{ gdy } 0 < X_{n:n} < \frac{1}{3} \\ 1 \text{ gdy } \frac{1}{3} < X_{n:n} < \frac{2}{3} \end{cases}$$

Wówczas:

$$\mathrm{E}_{\frac{1}{3}}\,\phi(\mathbb{X}) = \alpha\mathbb{P}_{\frac{1}{3}}(\phi(\mathbb{X}) = \alpha) = \alpha\mathbb{P}_{\frac{1}{3}}(0 < X_{n:n} < \frac{1}{3}) + 1\mathbb{P}_{\frac{1}{3}}(\frac{1}{3} < X_{n:n} < \frac{2}{3})$$
$$= \alpha$$

Policzmy moc tego testu:

$$M = \mathrm{E}_{\frac{2}{3}}\,\phi(\mathbb{X}) = \alpha\mathbb{P}_{\frac{2}{3}}(0 < X_{n:n} < \frac{1}{3}) + 1\mathbb{P}_{\frac{2}{3}}(\frac{1}{3} < X_{n:n} < \frac{2}{3})$$

Widzimy, że:

$$\mathbb{P}_{\frac{2}{3}}(0 < X_{n:n} < \frac{1}{3}) = F_X(\frac{1}{3})^n = \left(\frac{\frac{1}{3}}{\frac{2}{3}}\right)^{3n} = \frac{1}{8^n}$$

Zadanie 7.6

A także:

$$\mathbb{P}_{\frac{2}{3}}(\frac{1}{3} < X_{n:n} < \frac{2}{3}) = F_X(\frac{2}{3})^n - F_X(\frac{1}{3})^n = 1 - \frac{1}{8^n}$$

Zatem:

$$M = \alpha\frac{1}{8^n} + 1 - \frac{1}{8^n} = 1 - \frac{1-\alpha}{8^n} = 1 - \frac{0.99}{8^n}$$

Drugi sposób. Nasza przestrzeń parametrów jest równa $\Theta = \{\frac{1}{3}, \frac{2}{3}\}$ zatem hipotezę:

$$\begin{cases} H : \theta = \frac{1}{3} \\ K : \theta = \frac{2}{3} \end{cases}$$

mogę przedstawić jako:

$$\begin{cases} H : \theta \leqslant \frac{1}{3} \\ K : \theta > \frac{1}{3} \end{cases}$$

Zadanie 7.6

I zastosować tw. Karlina-Rubina. Wpierw wykażmy, że nasza rodzina rozkładów ma monotoniczny iloraz wiarogodności, Rozpatrzmy $\theta_1 > \theta_0$:

$$\frac{p_{\theta_0}(x_1,..,x_n)}{p_{\theta_1}(x_1,..,x_n)} = \frac{\frac{3^n \prod_{i=1}^n x_i^2}{(\theta_0)^{3n}} I(0,\theta_0)(x_{n:n})}{\frac{3^n \prod_{i=1}^n x_i^2}{(\theta_1)^{3n}} I(0,\theta_1)(x_{n:n})}$$

$$= \frac{\theta_1^{3n}}{\theta_0^{3n}} \frac{I(0,\theta_0)(x_{n:n})}{I(0,\theta_1)(x_{n:n})}$$

Widzimy, że jest to funkcja nierosnąca funkcją $x_{n:n}$. Zatem test ma postać:

$$\phi(\mathbb{X}) = \begin{cases} 1 \text{ gdy } X_{n:n} \geqslant c \\ 0 \text{ w.p.p.} \end{cases}$$

Zadanie 7.6

Znajdujemy c:

$$\mathbb{P}_{\frac{1}{3}}(X_{n:n} \geqslant c) = 1 - F_{X_{n:n}}(c) = 1 - (F_X(c))^n = \alpha$$

$$(1-\alpha)^{\frac{1}{n}} = F_X(c) = (3c)^3$$

$$c = \frac{(1-\alpha)^{\frac{1}{3n}}}{3}$$

P.d. policzyć moc.

Zadanie 7.7

Mamy próbkę $X_1, ..., X_{n_1}$ z rozkładu $Bern(\theta_1)$ i $Y_1, ..., Y_{n_2}$ z rozkładu $Bern(\theta_2)$. Mamy zweryfikować następującą parę hipotez:

$$\begin{cases} H : \theta_1 = \theta_2 \\ K : \theta_1 \neq \theta_2 \end{cases}$$

Rozpatrujemy następującą statystykę:

$$\bar{X} - \bar{Y}$$

Przy założeniu hipotezy zerowej wiemy, że:

$$\mathrm{E}\, X_i = \mathrm{E}\, Y_i = \theta_1$$

$$\mathrm{Var}\, X_i = \mathrm{Var}\, Y_i = \theta_1(1 - \theta_1)$$

Z CTG wiemy, że:

$$\sqrt{n_1}(\bar{X} - \theta_1) \xrightarrow{d} N\left(0, \sqrt{\theta_1(1 - \theta_1)}\right) \text{ i } \sqrt{n_2}(\bar{Y} - \theta_1) \xrightarrow{d} \sim N\left(0, \sqrt{\theta_1(1 - \theta_1)}\right)$$

Zadanie 7.7

Zatem dla dostatecznie dużej próby:

$$\bar{X} \sim N\left(\theta_1, \sqrt{\frac{\theta_1(1-\theta_1)}{n_1}}\right) \text{ i } \bar{Y} \sim N\left(\theta_1, \sqrt{\frac{\theta_1(1-\theta_1)}{n_2}}\right)$$

Zauważmy, że $\bar{X}$ i $\bar{Y}$ są niezależne, zatem:

$$\bar{X} - \bar{Y} \sim N\left(0, \sqrt{\theta_1(1-\theta_1)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}\right)$$

Wariancja może być zapisana jako:

$$\theta_1(1-\theta_1)\left(\frac{1}{n_1} + \frac{1}{n_2}\right) = \frac{\theta_1(1-\theta_1)}{\frac{n_1 n_2}{n_1+n_2}}$$

Zatem:

$$\frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\theta_1(1-\theta_1)}{\frac{n_1 n_2}{n_1+n_2}}}} \sim N(0,1)$$

Zadanie 7.7

Zauważmy, że:

$$p^* = \frac{\sum_{i=1}^{n_1} X_i + \sum_{i=1}^{n_2} Y_i}{n_1 + n_2}$$

przy założeniu hipotezy zerowej zbiega do θ_1. Zatem dla odpowiednio dużej próby

$$\sqrt{p^*(1-p^*)} \approx \sqrt{\theta_1(1-\theta_1)}$$

Więc ostatecznie statystyka testowa:

$$\frac{\bar{X} - \bar{Y}}{\sqrt{\frac{p^*(1-p^*)}{\frac{n_1 n_2}{n_1+n_2}}}} \sim N(0,1)$$

Zadanie 7.12

Mamy dwie próbki z rozkładu Poissona $X_1, ..., X_{100} \sim Poiss(\lambda_1)$ i $Y_1, ..., Y_{100} \sim Poiss(\lambda_2)$. Wiemy, że $\bar{X} = 20$ i $\bar{Y} = 22$.
Para hipotez:

$$\begin{cases} H : \lambda_1 = \lambda_0 \\ K : \lambda_1 \neq \lambda_0 \end{cases}$$

Mamy zbadać parę hipotez testem ilorazu wiarogodności. Nasza przestrzeń parametrów $\Theta = \{\lambda_1, \lambda_2\}$. Z kolei $\Theta_0 = \{\lambda = \lambda_1 = \lambda_2\}$. Policzmy:

$$\sup_{\lambda_1, \lambda_2 \in \Theta_0} L(\mathbb{X}, \lambda_1, \lambda_2) = \sup_{\lambda} L(\mathbb{X}, \lambda, \lambda)$$

Szukamy argumentu, który zmaksymalizuje nam wartość $L(\mathbb{X}, \lambda_1, \lambda_2)$ przy założeniu, że $\lambda = \lambda_1 = \lambda_2$. Zatem:

$$l(\mathbb{X}, \lambda, \lambda) = \ln\left(\prod_{i=1}^{n}\left(\frac{\lambda^{x_i} e^{-\lambda}}{x_i!}\right)\prod_{i=1}^{n}\left(\frac{\lambda^{y_i} e^{-\lambda}}{y_i!}\right)\right)$$

Zadanie 7.12

$$= \ln\left(\frac{\lambda^{\sum_{i=1}^n x_i + \sum_{i=1}^n y_i} e^{-2\lambda n}}{\prod_{i=1}^n (x_i! y_i!)}\right)$$

$$= (\sum_{i=1}^n x_i + \sum_{i=1}^n y_i) \ln\lambda - 2\lambda n - \ln(\prod_{i=1}^n (x_i! y_i!))$$

$$\frac{d}{d\lambda} l(\mathbb{X}, \lambda, \lambda) = (\sum_{i=1}^n x_i + \sum_{i=1}^n y_i)\frac{1}{\lambda} - 2n = 0$$

$$\hat{\lambda} = \frac{\sum_{i=1}^n X_i + \sum_{i=1}^n Y_i}{2n}$$

Druga pochodna:

$$\frac{d^2}{d\lambda^2} l(\mathbb{X}, \lambda, \lambda) = -(\sum_{i=1}^n x_i + \sum_{i=1}^n y_i)\frac{1}{\lambda^2} < 0$$

Zadanie 7.12

Teraz poszukamy

$$\sup_{\lambda_1,\lambda_2\in\Theta} L(\mathbb{X},\lambda_1,\lambda_2) = \sup_{\lambda_1,\lambda_2} L(\mathbb{X},\lambda_1,\lambda_2)$$

Zatem:

$$l(\mathbb{X},\lambda_1,\lambda_2) = \ln\left(\prod_{i=1}^{n}\left(\frac{\lambda_1^{x_i}e^{-\lambda_1}}{x_i!}\right)\prod_{i=1}^{n}\left(\frac{\lambda_2^{y_i}e^{-\lambda_2}}{y_i!}\right)\right)$$

$$= \ln\left(\frac{\lambda_1^{\sum_{i=1}^{n} x_i}\lambda_2^{\sum_{i=1}^{n} y_i}e^{-(\lambda_1+\lambda_2)n}}{\prod_{i=1}^{n}(x_i!y_i!)}\right)$$

$$= \sum_{i=1}^{n} x_i \ln\lambda_1 + \sum_{i=1}^{n} y_i \ln\lambda_2 - (\lambda_1+\lambda_2)n - \ln(\prod_{i=1}^{n}(x_i!y_i!))$$

Zadanie 7.12

$$\frac{d}{d\lambda_1} l(\mathbb{X}, \lambda_1, \lambda_2) = \sum_{i=1}^{n} x_i \frac{1}{\lambda_1} - n = 0$$

$$\hat{\lambda}_1 = \frac{\sum_{i=1}^{n} X_i}{n}$$

Druga pochodna:

$$\frac{d^2}{d\lambda_1^2} l(\mathbb{X}, \lambda_1, \lambda_2) = -\sum_{i=1}^{n} x_i \frac{1}{\lambda_1^2} < 0$$

Analogicznie dla λ_2:

$$\frac{d}{d\lambda_2} l(\mathbb{X}, \lambda_1, \lambda_2) = \sum_{i=1}^{n} y_i \frac{1}{\lambda_2} - n = 0$$

$$\hat{\lambda}_2 = \frac{\sum_{i=1}^{n} Y_i}{n}$$

$$\frac{d^2}{d\lambda_2^2} l(\mathbb{X}, \lambda_1, \lambda_2) = -\sum_{i=1}^{n} y_i \frac{1}{\lambda_2^2} < 0$$

Zadanie 7.12

Więc:

$$\lambda(\mathbb{X}) = \frac{\sup_\lambda L(\mathbb{X}, \lambda, \lambda)}{\sup_\lambda L(\mathbb{X}, \lambda_1, \lambda_2)} = \frac{\frac{\hat{\lambda}^{\sum_{i=1}^n x_i + \sum_{i=1}^n y_i} e^{-2\hat{\lambda}n}}{\prod_{i=1}^n (x_i! y_i!)}}{\frac{\hat{\lambda}_1^{\sum_{i=1}^n x_i} \hat{\lambda}_2^{\sum_{i=1}^n y_i} e^{-(\hat{\lambda}_1 + \hat{\lambda}_2)n}}{\prod_{i=1}^n (x_i! y_i!)}}$$

$$= \frac{\hat{\lambda}^{\sum_{i=1}^n x_i + \sum_{i=1}^n y_i} e^{-2\hat{\lambda}n}}{\hat{\lambda}_1^{\sum_{i=1}^n x_i} \hat{\lambda}_2^{\sum_{i=1}^n y_i} e^{-(\hat{\lambda}_1 + \hat{\lambda}_2)n}}$$

$$= \frac{\left(\frac{\sum_{i=1}^n x_i + \sum_{i=1}^n y_i}{2n}\right)^{\sum_{i=1}^n x_i + \sum_{i=1}^n y_i} e^{-(\sum_{i=1}^n x_i + \sum_{i=1}^n y_i)}}{\left(\frac{\sum_{i=1}^n x_i}{n}\right)^{\sum_{i=1}^n x_i} \left(\frac{\sum_{i=1}^n y_i}{n}\right)^{\sum_{i=1}^n y_i} e^{-(\sum_{i=1}^n x_i + \sum_{i=1}^n y_i)}}$$

$$= \frac{\left(\frac{\sum_{i=1}^n x_i + \sum_{i=1}^n y_i}{2n}\right)^{\sum_{i=1}^n x_i + \sum_{i=1}^n y_i}}{\left(\frac{\sum_{i=1}^n x_i}{n}\right)^{\sum_{i=1}^n x_i} \left(\frac{\sum_{i=1}^n y_i}{n}\right)^{\sum_{i=1}^n y_i}}$$

Zadanie 7.12

Rozpatrujemy statystykę $-2\ln\lambda(\mathbb{X})$.

$$-2\ln\lambda(\mathbb{X}) = -2\left((\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i)\ln\left(\frac{\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i}{2n}\right)\right.$$

$$\left. - \sum_{i=1}^{n} x_i \ln(\frac{\sum_{i=1}^{n} x_i}{n}) - \sum_{i=1}^{n} y_i \ln(\frac{\sum_{i=1}^{n} y_i}{n})\right)$$

$$-2(4200\ln(21) - 2000\ln(20) - 2200\ln(22)) \approx 9.53$$

Rozkład statystyki przy założeniu H_0 jest χ^2_1. Przedział krytyczny:

$$(-\infty, \chi^2_{1,0.005}] \cup [\chi^2_{1,0.995}, \infty) = (-\infty, 0.00004] \cup [7.88, \infty)$$

Zatem nasza statystyka znajduje się w przedziale krytycznym, zatem przeciętne wartości są istotnie różne.

Zadanie 8.2

Chcemy sprawdzić, czy wygenerowana próba pochodzi z rozkładu $Bin(3, 0.5)$:

Wygenerowana liczba losowa	0	1	2	3
Liczba uzyskanych parametrów	12	37	38	13

Liczba parametrów szacowanych z próby jest równa 1 (pierwszy parametr jest ustalony). Liczymy statystykę testową:

$$T = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i} = \frac{(12 - 100 * 0.125)^2}{100 * 0.125} + \frac{(37 - 100 * 0.375)^2}{100 * 0.375}$$

$$+ \frac{(38 - 100 * 0.375)^2}{100 * 0.375} + \frac{(13 - 100 * 0.125)^2}{100 * 0.125}$$

$$= 2\frac{0.25}{12.5} + 2\frac{0.25}{37.5} = 0.04 + 0.013 = 0.053$$

Przedział krytyczny:

$$\kappa_\alpha = [\chi^2_{1-\alpha,k-1-r}, \infty) = [\chi^2_{0.95,2}, \infty) = [5.991, \infty)$$

Jesteśmy poza przedziałem krytycznym zatem nie ma podstaw do odrzucenia

Zadanie 8.8

Weryfikacja hipotez o niezależności cech X i Y	
Hipoteza zerowa	H : cechy X i Y są niezależne
Hipoteza alternatywna	K : cechy X i Y są zależne
Statystyka testowa	$T = \sum_{i=1}^{r\cdot c} \frac{(O_i - E_i)^2}{E_i}$

gdzie O_i - liczba obserwacji w i-tej komórce, $E_i = \frac{R_i \cdot C_i}{n}$,
R_i - suma obserwacji w wierszu, w którym jest położona i-ta komórka,
C_i - suma obserwacji w kolumnie, do której należy i-ta komórka,
r - liczba wierszy, c - liczba kolumn w tabeli kontyngencji,
n - liczba wszystkich obserwacji,

Obszar krytyczny	$\left\langle \chi^2_{1-\alpha,(r-1)(c-1)}; +\infty \right)$

Chcemy zbadać zależność między
spędzania wakacji. Tabela:

	góry	jeziora	morze
kobiety	32	32	41
mężczyźni	39	33	23

$\Sigma r_1 = 105$

$\Sigma r_2 = 95$ // 200

Statystyka testowa: $\frac{\Sigma \text{ wiersz} \cdot \Sigma \text{ kolumna}}{\Sigma \text{ wszystko}}$

$$T = \sum_{i=1}^{rc} \frac{(O_i - E_i)^2}{E_i} = \frac{\left(32 - \frac{105*71}{200}\right)^2}{\frac{105*71}{200}} + \frac{\left(32 - \frac{105*65}{200}\right)^2}{\frac{105*65}{200}} + \frac{\left(41 - \frac{105*64}{200}\right)^2}{\frac{105*64}{200}}$$

$$+ \frac{\left(39 - \frac{95*71}{200}\right)^2}{\frac{95*71}{200}} + \frac{\left(33 - \frac{95*65}{200}\right)^2}{\frac{95*65}{200}} + \frac{\left(23 - \frac{95*64}{200}\right)^2}{\frac{95*64}{200}}$$

i-ta komórka tabeli

Zadanie 8.8

$$T = \sum_{i=1}^{rc} \frac{(O_i - E_i)^2}{E_i} = \frac{(32 - 37.275)^2}{37.275} + \frac{(32 - 34.125)^2}{34.125} + \frac{(41 - 33.6)^2}{33.6}$$

$$+ \frac{(39 - 33.725)^2}{33.725} + \frac{(33 - 30.875)^2}{30.875} + \frac{(23 - 30.4)^2}{30.4}$$
$$= 0.746 + 0.132 + 1.63 + 0.825 + 0.146 + 1.801 = 5.28$$

Przedział krytyczny:

$$\kappa_\alpha = [\chi^2_{1-\alpha,(r-1)(c-1)}, \infty) = [\chi^2_{0.95,2}, \infty) = [5.991, \infty)$$

Jesteśmy poza przedziałem krytycznym zatem nie ma podstaw do odrzucenia hipotezy.