Metody probabilistyczne
Rozwiązania zadań

13. Statystyka 2

16.01.2018

Zadanie 1. *Wyznacz estymator największej wiarogodności parametru λ dla rozkład Poissona.*

Odpowiedź: Rozkład Poissona zdefiniowany jest jako:

$$p_\lambda(x) \;=\; \frac{\lambda^x}{x!}e^{-\lambda}, \qquad \text{dla} \;\; x = 0, 1, 2, \ldots,$$

gdzie przez $p_\lambda(x)$ oznaczamy skrótowo $P_\lambda(X = x)$. Załóżmy teraz, że zaobserwowaliśmy dane $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$. Funkcja wiarogodności (łączne prawdopodobieństwo zaobserwowanych danych) ma postać:

$$L(\boldsymbol{x};\lambda) \;=\; \prod_{i=1}^{n} p_\lambda(x_i) \;=\; \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!}e^{-\lambda} \;=\; \frac{1}{x_1!x_2! \cdot \ldots \cdot x_n!}\lambda^{x_1+x_2+\ldots+x_n}e^{-n\lambda}.$$

Oznaczając jak zwykle $\overline{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$, możemy przepisać:

$$L(\boldsymbol{x};\lambda) \;=\; \frac{1}{x_1!x_2! \cdot \ldots \cdot x_n!}\lambda^{n\overline{x}_n}e^{-n\lambda},$$

co po zlogarytmowaniu i przemnożeniu przez minus jeden daje:

$$-\ln L(\boldsymbol{x};\lambda) \;=\; \ln(x_1! \cdot \ldots \cdot x_n!) \;+\; n\overline{x}_n \ln\lambda \;-\; n\lambda.$$

Pierwszy człon jest niezależne od λ, więc zniknie przy braniu pochodnej. Różniczkując po λ dostajemy:

$$\frac{\partial}{\partial\lambda} - \ln L(\boldsymbol{x};\lambda) \;=\; n\overline{x}_n\frac{1}{\lambda} - n.$$

Przyrównanie pochodnej do zera daje:

$$\lambda \;=\; \overline{x}_n,$$

a więc estymatorem największej wiarogonodności parametru λ jest po prostu średnia arytmetyczna z próby: $\widehat{\lambda} = \overline{X}_n$.

Zadanie 2*. *Skonstruuj przedział ufności dla parametru p w rozkładzie dwupunktowym $B(p)$. Użyj przybliżenia rozkładem normalnym.*

Odpowiedź: Niech $X \sim B(p)$ ma rozkład dwupunktowy. Znajdziemy przedział ufności dla parametru p. Estymator punktowy parametru p na podstawie próby $X_1, \ldots, X_n$ ma postać:

$$\widehat{p} \;=\; \overline{X}_n \;=\; \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Ponieważ $X_i \in \{0, 1\}$, łatwo zauważyć, że $\widehat{p}$ jest równe liczbie sukcesów podzielonej przez n, czyli empirycznej (wyznaczonej na próbie) częstości sukcesów. Ponieważ $\widehat{p}$ jest również średnią arytmetyczną z próby, mamy:

$$E(\widehat{p}) \;=\; E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \;=\; \frac{1}{n}\sum_{i=1}^{n} \underbrace{E(X_i)}_{p} \;=\; p,$$

$$D^2(\widehat{p}) \;=\; D^2\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \;=\; \frac{1}{n^2}\sum_{i=1}^{n} \underbrace{D^2(X_i)}_{p(1-p)} \;=\; \frac{p(1-p)}{n},$$

gdzie we wzorze na wariancję wykorzystaliśmy niezależność zmiennych $X_1, \dots, X_n$. Z Centralnego Twierdzenia Granicznego wiemy, że ciąg ustandaryzowanych średnich arytmetycznych zbiega (według dystrybuant) do zmiennej o rozkładzie normalnym standardowym $N(0,1)$, tzn. że:

$$\frac{\widehat{p} - E(\widehat{p})}{D(\widehat{p})} \;=\; \frac{\widehat{p} - p}{\sqrt{p(1-p)}}\sqrt{n} \;\xrightarrow{D}\; Z \;\sim\; N(0,1) \tag{1}$$

Z kolei z Prawa Wielkich Liczb wiemy, że średnia arytmetyczna zbiega do wartości oczekiwanej:

$$\widehat{p} \xrightarrow{P} E(X) \;=\; p.$$

Tym samym, możemy[1] w mianowniku (1) zastąpić p przez $\widehat{p}$ i nadal otrzymamy zbieżność według dystrybuant do $Z \sim N(0,1)$:

$$\frac{\widehat{p} - p}{\sqrt{\widehat{p}(1-\widehat{p})}}\sqrt{n} \;\xrightarrow{D}\; Z \;\sim\; N(0,1).$$

Założymy, że n jest wystarczająco duże i użyjemy powyższej własności do przybliżenia rozkładem normalnym, tzn. przybliżymy:

$$\frac{\widehat{p} - p}{\sqrt{\widehat{p}(1-\widehat{p})}}\sqrt{n} \;\simeq\; Z \;\sim\; N(0,1) \tag{2}$$

Konstruujemy przedział ufności na poziomie ufności $1-\alpha$:

$$\begin{aligned}
1-\alpha \;&=\; P\left(-\Delta \;\leqslant\; \widehat{p} - p \;\leqslant\; \Delta\right) \\
&=\; P\Bigg(-\frac{\Delta\sqrt{n}}{\sqrt{\widehat{p}(1-\widehat{p})}} \;\leqslant\; \underbrace{\frac{\widehat{p}-p}{\sqrt{\widehat{p}(1-\widehat{p})}}\sqrt{n}}_{\simeq Z \text{ (używamy (2))}} \;\leqslant\; \frac{\Delta\sqrt{n}}{\sqrt{\widehat{p}(1-\widehat{p})}}\Bigg) \\
&=\; P\Bigg(-\frac{\Delta\sqrt{n}}{\sqrt{\widehat{p}(1-\widehat{p})}} \;\leqslant\; Z \;\leqslant\; \frac{\Delta\sqrt{n}}{\sqrt{\widehat{p}(1-\widehat{p})}}\Bigg) \\
&\simeq\; \Phi\left(\frac{\Delta\sqrt{n}}{\sqrt{\widehat{p}(1-\widehat{p})}}\right) - \Phi\left(-\frac{\Delta\sqrt{n}}{\sqrt{\widehat{p}(1-\widehat{p})}}\right) = 2\Phi\left(\frac{\Delta\sqrt{n}}{\sqrt{\widehat{p}(1-\widehat{p})}}\right) - 1,
\end{aligned}$$

gdzie w ostatniej linii wykorzystaliśmy własność dystrybuanty rozkładu normalnego standardowego $\Phi(-x) = 1 - \Phi(x)$ dla dowolnych x. Otrzymaliśmy więc równanie:

$$1-\alpha \;=\; 2\Phi\left(\frac{\Delta\sqrt{n}}{\sqrt{\widehat{p}(1-\widehat{p})}}\right) - 1,$$

co po dodaniu stronami 1 i podzieleniu przez 2 daje:

$$\Phi\left(\frac{\Delta\sqrt{n}}{\sqrt{\widehat{p}(1-\widehat{p})}}\right) \;=\; 1 - \frac{\alpha}{2}.$$

Ponieważ Φ jest odwracalna, otrzymujemy:

$$\frac{\Delta\sqrt{n}}{\sqrt{\widehat{p}(1-\widehat{p})}} \;=\; \underbrace{\Phi^{-1}\left(1-\frac{\alpha}{2}\right)}_{z_{1-\alpha/2}} \qquad \Longrightarrow \qquad \Delta \;=\; z_{1-\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}.$$

Zauważmy, że $z_{1-\alpha/2}$ jest kwantylem rzędu $1-\frac{\alpha}{2}$ rozkładu normalnego standardowego. Tym samym wyznaczyliśmy przybliżony przedział ufności na poziomie ufności $1-\alpha$:

$$[\widehat{p} - \Delta, \widehat{p} + \Delta], \qquad \text{gdzie } \Delta \;=\; z_{1-\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}.$$

[1]Uwaga: stwierdzenie to jest nietrywialne. Wymaga użycia dwóch twierdzeń: (1) jeśli $Y_n \xrightarrow{P} c$ to dla dowolnej funkcji ciągłej f mamy $f(Y_n) \xrightarrow{P} f(c)$; (2) jeśli $X_n \xrightarrow{D} X$ i $Z_n \xrightarrow{P} a$ to $X_n Z_n \xrightarrow{D} aX$. Używamy powyższych twierdzeń biorąc $Y_n = \overline{X}_n = \widehat{p} \xrightarrow{P} p$, $f(x) = \frac{\sqrt{p(1-p)}}{\sqrt{x(1-x)}}$, oraz $Z_n = f(Y_n)$, tym samym $Z_n = \frac{\sqrt{p(1-p)}}{\sqrt{\widehat{p}(1-\widehat{p})}} \xrightarrow{P} f(p) = 1$. Na koniec bierzemy $X_n = \frac{\widehat{p}-p}{\sqrt{p(1-p)}}\sqrt{n} \xrightarrow{D} N(0,1)$, co po wykorzystaniu drugiego twierdzenia daje: $\frac{\widehat{p}-p}{\sqrt{\widehat{p}(1-\widehat{p})}}\sqrt{n} = X_n Z_n \xrightarrow{D} N(0,1)$.