Paulina Busz
Sylwia Dzierżawa

Projekt zaliczeniowy

Wnioskowanie Bayesowskie w ekonomii empirycznej

Cel projektu i opis danych:

Niniejszy projekt zawiera opis procesu wnioskowania bayesowskiego dla zagadnienia dotyczącego brutalnych przestępstw popełnionych w Stanach Zjednoczonych Ameryki Północnej w latach 1994 – 2013. Dane zostały pobrane z oficjalnej strony Federalnego Biura Śledczego (FBI). Według klasyfikacji FBI do brutalnych przestępstw zaliczane są: przestępstwa polegające na zabójstwie lub nieumyślnym zabójstwie, gwałt, rozbój a także agresywny atak. Dane, które Federalne Biuro Śledcze umieszcza każdego roku na swojej oficjalnej stronie, zbierane są za pomocą *„ Uniform Crime Reporting Program”*. Jest to program zrzeszający ograny ścigania w celu pobierania od nich danych statystycznych.

Dane 1994 – 2013 : https://ucr.fbi.gov/crime-in-the-u.s/2013/crime-in-the-u.s.-2013/tables/1tabledatadecoverviewpdf/table_1_crime_in_the_united_states_by_volume_and_rate_per_100000_inhabitants_1994-2013.xls

UCR: https://www.ucrdatatool.gov/

Proces wnioskowania bayesowskiego:

a) Model próbkowy:

$$p(y_t|\mu) = f_N(y_t|\mu) = (2\pi)^{-\frac{T}{2}} \sigma^{-T} \exp\left\{-\frac{1}{2}\sigma^{-2}\sum_{t=1}^{T}(y_t - \mu)^2\right\}.$$

b) Standaryzowana funkcja wiarygodności:

$$\widetilde{Ly}(\mu) = \frac{Ly(\mu)}{\int_{-\infty}^{\infty} Ly(\mu)} \,\alpha\, Ly(\mu),$$

$$\widetilde{Ly}(\mu) \,\alpha\, Ly(\mu) \,\alpha\, \exp\left\{-\frac{1}{2}\sigma^{-2}T(\bar{y} - \mu)^2\right\}.$$

Parametry:

$$\begin{cases} \bar{\mu} = \bar{y} = 1\,439\,004{,}3 \\ \bar{v} = \sqrt{\dfrac{1}{T\sigma^{-2}}} = 41\,580{,}5 \end{cases}.$$

c) Rozkład a priori:

Sprzężonym do rozkładu próbkowego (zapisanego w pkt. a)) rozkładem a priori jest rozkład normalny o następującej funkcji gęstości:

$$p(\mu) = f_N(\mu) = \frac{1}{v_0\sqrt{2\pi}}\, e^{\frac{(\mu - m_0)^2}{2v_0^2}},$$

gdzie v_0 oraz m_0 są hiperparametrami.

Aby wyznaczyć hipreparamtery rozkładu a priori posłużono się danymi dotyczącymi brutalnych przestępstw dla 5 państw Unii Europejskiej w latach 1998 – 2007 (stąd w zbiorze danych zawiera się Anglia), których suma powierzchni w znaczącym stopniu odpowiada powierzchni Stanów Zjednoczonych. Są to:

- Niemcy,
- Hiszpania,
- Francja,
- Włochy,
- Wielka Brytania (Anglia oraz Walia, Szkocja, Irlandia Północna).

Z danych wyznaczono średnią oraz odchylenie standardowe z próby i przyjęto je jako wartości hiperparametrów:

$$m_0 = 2\,073\,956{,}3,$$

$$v_0 = 214\,406{,}5.$$

d) Model bayesowski:

$$p(y,\mu) = p(y|\mu)p(\mu) = (2\pi)^{-\frac{T-1}{2}}\,\sigma^{-T} v_0^{-1} \exp\left\{-\frac{1}{2}\left(\sigma^{-2}\sum_{t=1}^{T}(y_t - \mu)^2 + v_0^{-2}(\mu - m_0)^2\right)\right\}.$$

e) Rozkład a posteriori:

$$p(\mu|y)\;\alpha\;p(y,\mu)\;\alpha\;\exp\left\{-\frac{1}{2}(T\sigma^{-2} + v_0^{-2})\left(\mu - \frac{T\sigma^{-2}\bar{y} + v_0^{-2}m_0}{T\sigma^{-2} + v_0^{-2}}\right)^2\right\},$$

gdzie:

$$\bar{m}_0 = \frac{T\sigma^{-2}\bar{y} + v_0^{-2}m_0}{T\sigma^{-2} + v_0^{-2}} = 1\,462\,019{,}3,$$

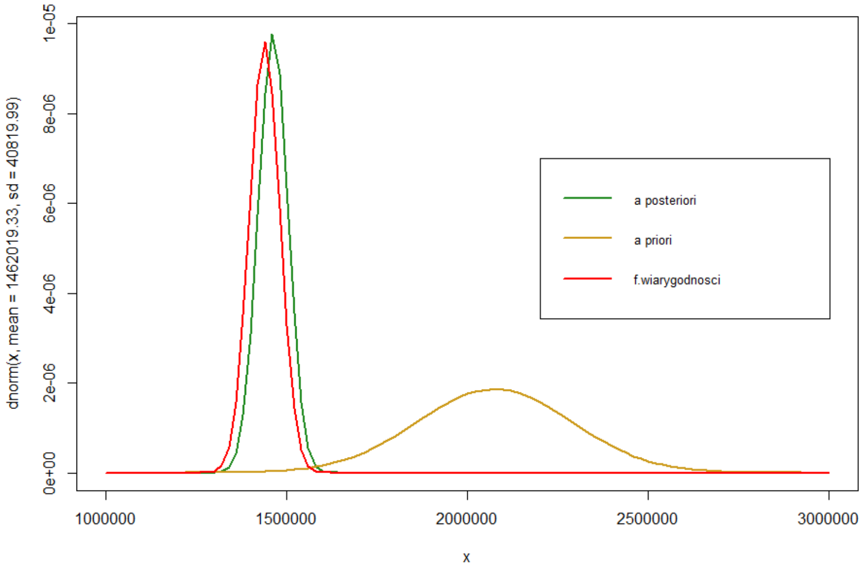
$$\bar{v}_0 = \sqrt{\frac{1}{T\sigma^{-2} + v_0^{-2}}} = 40\,820.$$

Ostatecznie:

$$p(\mu|y) = \frac{1}{\bar{v}_0\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\bar{v}_0^{-2}(\mu - \bar{m}_0)^2\right\}.$$

f) Wyniki estymacji:

- Funkcje gęstości:



- Wartości oczekiwane:

$$E(\mu) = 2\,073\,956{,}3$$

$$E(\mu|y) = 1\,462\,019{,}3.$$

Oczekiwana przed wglądem w dane liczba brutalnych przestępstw w Stanach Zjednoczonych wynosi 2073956,3.

Oczekiwana po wglądzie w dane liczba brutalnych przestępstw w Stanach Zjednoczonych wynosi 1462019,3.

- Modalne:

$$Mo(\mu) = 2\,073\,956{,}3$$

$$Mo(\mu|y) = 1\,462\,019{,}3$$

Największe zagęszczenie masy prawdopodobieństwa a priori ma miejsce wokół wartości 2073956,3 przestępstw.

Największe zagęszczenie masy prawdopodobieństwa a posteriori ma miejsce wokół wartości 1462019,3 przestępstw.

- Mediany:

$$Me(\mu) = 2\,073\,956{,}3$$

$$Me(\mu|y) = 1\,462\,019{,}3$$

Na lewo i na prawo od 2073956,3 znajduje się ½ masy prawdopodobieństwa a priori.

Na lewo i na prawo od 1462019,3 znajduje się ½ masy prawdopodobieństwa a posteriori.

- Odchylenia standardowe:

$$D(\mu) = 214\,406{,}5,$$

$$D(\mu|y) = 40\,820.$$

Przeciętna a posteriori różnica pomiędzy liczbą brutalnych przestępstw w Stanach Zjednoczonych a ich wartością oczekiwaną wynosi 40820.

Ponieważ odchylenie standardowe a posteriori jest mniejsze od odchylenia standardowego a priori możemy stwierdzić, że informacja zawarta w danych zredukowała niepewność.

- Kwantyle:

$$Q_{0,25}(\mu) = 1\,929\,341{,}28$$

$$Q_{0,75}(\mu) = 2\,218\,571{,}32$$

25% masy prawdopodobieństwa a priori znajduje się na lewo od wartości 1929341,28.
75% masy prawdopodobieństwa a priori znajduje się na lewo od wartości 2218571,32.

$$Q_{0,25}(\mu|y) = 1\,434\,486{,}66$$

$$Q_{0,75}(\mu|y) = 1\,489\,552{,}99$$

25% masy prawdopodobieństwa a posteriori znajduje się na lewo od wartości 1434486,66.
75% masy prawdopodobieństwa a posteriori znajduje się na lewo od wartości 1489552,99.

- 90 % przedział wiarygodności:

$$\left(Q_{0,025}(\mu)\,;\ Q_{0,975}(\mu)\right) = (1\,653\,727\,;\ 2\,494\,185)$$

A priori przedział $(1\,653\,727\,;\ 2\,494\,185)$ zawiera nieznaną wartość parametru μ z wiarygodnością 90%.

$$\left(Q_{0,025}(\mu|y)\,;\ Q_{0,975}(\mu|y)\right) = (808\,870{,}5\,;\ 2\,037\,523)$$

A posteriori przedział $(808\,870{,}5\ ;\ 2\,037\,523)$ zawiera nieznaną wartość parametru μ z wiarygodnością 90%.

g) Brzegowa gęstość wektora obserwacji:

$$p(y) = \frac{p(\mu, y)}{p(\mu|y)} = \frac{(2\pi)^{-\frac{T}{2}}\sigma^{-T}v_0^{-1}}{\bar{v}_0{}^{-1}} = \ 8{,}12E - 115.$$

h) Nieinformacyjny rozkład a priori (reguła Jeffreysa):

- Właściwość rozkładu:

$$p(\mu) = \ T^{\frac{1}{2}}\sigma^{-2},$$

rozkład nie jest rozkładem sprzężonym.

$$\int_{-\infty}^{\infty} T^{\frac{1}{2}}\sigma^{-1} d\mu\,,$$

jest to całka niewłaściwa.

- Rozkład a posteriori:

$$p(y|\mu)p(\mu)\ \alpha \exp\left\{-\frac{1}{2}T\sigma^{-2}(\bar{y} - \ \mu)^2\right\}\ \alpha\ f_N(\mu|\bar{\mu}, \bar{v}),$$

$$p(\mu|y) = \ (2\pi)^{-\frac{1}{2}}\bar{v}^{-1} \exp\left\{-\frac{1}{2}\bar{v}^{-2}(\bar{\mu} - \mu)^2\right\},$$

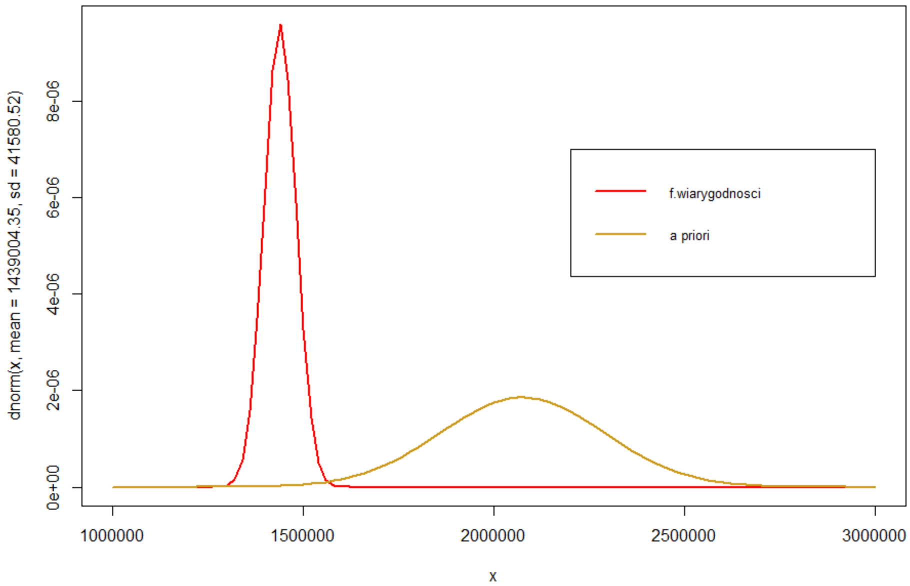
rozkład a posteriori jest rozkładem właściwym.

$$\bar{\mu} = \bar{y} = \ 1\,439\,004{,}35$$

$$\bar{v} = \sqrt{\frac{1}{T\sigma^{-2}}} = \ 41\,580{,}5$$

i) Wyniki estymacji:

- *Funkcje gęstości:*



- Wartości oczekiwane:

$$E(\mu) = \ 2\,073\,956{,}3$$

$$E(\mu|y) = \ 1\,439\,004{,}35$$

Oczekiwana przed wglądem w dane liczba brutalnych przestępstw w Stanach Zjednoczonych wynosi 2073956,3.

Oczekiwana po wglądzie w dane liczba brutalnych przestępstw w Stanach Zjednoczonych wynosi 1439004,35.

- Modalne:

$$Mo(\mu) = 2\,073\,956{,}3$$

$$Mo(\mu|y) = 1\,439\,004{,}35$$

Największe zagęszczenie masy prawdopodobieństwa a priori ma miejsce wokół wartości 2073956,3 przestępstw.

Największe zagęszczenie masy prawdopodobieństwa a posteriori ma miejsce wokół wartości 1439004,35 przestępstw.

- Mediany:

$$Me(\mu) = \ 2\,073\,956{,}3$$

$$Me(\mu|y) = \ 1\,439\,004{,}35$$

Na lewo i na prawo od 2073956,3 znajduje się ½ masy prawdopodobieństwa a priori.

Na lewo i na prawo od 1439004,35 znajduje się ½ masy prawdopodobieństwa a posteriori.

- Odchylenia standardowe:

$$D(\mu) = \ 214\ 406{,}5,$$

$$D(\mu|y) = 41\ 580{,}5.$$

Przeciętna a posteriori różnica pomiędzy liczbą brutalnych przestępstw w Stanach Zjednoczonych a ich wartością oczekiwaną wynosi 41580,5.

Ponieważ odchylenie standardowe a posteriori jest mniejsze od odchylenia standardowego a priori możemy stwierdzić, że informacja zawarta w danych zredukowała niepewność.

- Kwantyle:

$$Q_{0,25}(\mu) = 1\ 929\ 341$$

$$Q_{0,25}(\mu|y) = 2\ 228\ 571$$

25% masy prawdopodobieństwa a priori znajduje się na lewo od wartości 1929341,28.
75% masy prawdopodobieństwa a priori znajduje się na lewo od wartości 2218571,32.

$$Q_{0,75}(\mu) = 1\ 410\ 959$$

$$Q_{0,75}(\mu|y) = 1\ 467\ 050$$

25% masy prawdopodobieństwa a posteriori znajduje się na lewo od wartości 1434486,66.
75% masy prawdopodobieństwa a posteriori znajduje się na lewo od wartości 1489552,99.

- 90 % przedział wiarygodności:

$$\left(Q_{0,025}(\mu)\ ;\ Q_{0,975}(\mu)\right) = \ (1\ 653\ 727\ ;\ 2\ 494\ 185)$$

A priori przedział $\left(1\ 653\ 727\ ;\ 2\ 494\ 185\right)$ zawiera nieznaną wartość parametru μ z wiarygodnością 90%.

$$\left(Q_{0,025}(\mu|y)\ ;\ Q_{0,975}(\mu|y)\right) = \ (1\ 357\ 508\ ;\ 1\ 520\ 501)$$

A posteriori przedział $\left(1\ 357\ 508\ ;\ 1\ 520\ 501\right)$ zawiera nieznaną wartość parametru μ z wiarygodnością 90%.