Bootcamp Data Science

Statystyka

Przemysław Spurek

Estymatory

Definicja

Prostą próbą losową (lub krócej próbą losową) o liczności n nazywamy ciąg niezależnych zmiennych losowych $X_1, X_2, \ldots, X_n$ określonych na przestrzeni zdarzeń elementarnych Ω i takich, że każda ze zmiennych ma taki sam rozkład.

Definicja

Statystyką nazywamy każdą zmienną losową będącą ustaloną funkcją próby losowej $X_1, X_2, \ldots, X_n$.

Definicja

Każdą statystykę, którą przyjmujemy jako oszacowanie (przybliżenie) nieznanego parametru rozkładu będziemy nazywać *estymatorem*.

Estymatory

Uwaga

Statystyką jest więc, na przykład, najmniejsza, największa wartość w próbie, iloczyn lub suma kwadratów wszystkich wartości. Oczywiście, wybór konkretnej statystyki związany jest z nieznaną wielkością (parametrem) charakteryzującą populację, którą chcemy szacować. Statystykę

$$\bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

nazywamy średnią z próby losowej $X_1, X_2, \ldots, X_n$.

Estymatory

Uwaga

Statystyką jest więc, na przykład, najmniejsza, największa wartość w próbie, iloczyn lub suma kwadratów wszystkich wartości. Oczywiście, wybór konkretnej statystyki związany jest z nieznaną wielkością (parametrem) charakteryzującą populację, którą chcemy szacować. Statystykę

$$\bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

nazywamy średnią z próby losowej $X_1, X_2, \ldots, X_n$.

Definicja

Każdą statystykę, którą przyjmujemy jako oszacowanie (przybliżenie) nieznanego parametru rozkładu nazywamy *estymatorem*.

Estymatory

Jednym z zadań statystyki jest znajdowanie estymatorów (a więc statystyk), które w jakimś sensie mówią nam o rozkładzie zmiennej losowej X, z której pochodzi dana próbka. Na przykład, wydaje się, że znajomość średniej arytmetycznej:

$$\bar{X} = \frac{X_1 + \cdots + X_n}{n},$$

daje nam pewne informacje o nadziei matematycznej $\mathbb{E}(X)$.

Estymatory

Zauważmy jednak, że istnieją inne estymatory, które także dają pewne informacje o wartości średniej - przykładowo:

$$T(x_1, \ldots, x_n) = \frac{x_1 + x_n}{2}$$

lub

$$T(x_1, \ldots, x_n) = \frac{\min_{1 \leq i \leq n}\{x_i\} + \max_{1 \leq i \leq n}\{x_i\}}{2}.$$

Oczywiście, można wskazać jeszcze inne, dość "rozsądne" estymatory nadziei matematycznej.

Estymatory

Zauważmy jednak, że istnieją inne estymatory, które także dają pewne informacje o wartości średniej - przykładowo:

$$T(x_1, \ldots, x_n) = \frac{x_1 + x_n}{2}$$

lub

$$T(x_1, \ldots, x_n) = \frac{\min_{1 \leq i \leq n}\{x_i\} + \max_{1 \leq i \leq n}\{x_i\}}{2}.$$

Oczywiście, można wskazać jeszcze inne, dość "rozsądne" estymatory nadziei matematycznej.

Powstaje więc problem, jaki estymator należy stosować w konkretnej sytuacji?

Estymatory

Zauważmy jednak, że istnieją inne estymatory, które także dają pewne informacje o wartości średniej - przykładowo:

$$T(x_1, \ldots, x_n) = \frac{x_1 + x_n}{2}$$

lub

$$T(x_1, \ldots, x_n) = \frac{\min_{1 \le i \le n}\{x_i\} + \max_{1 \le i \le n}\{x_i\}}{2}.$$

Oczywiście, można wskazać jeszcze inne, dość "rozsądne" estymatory nadziei matematycznej.

Powstaje więc problem, jaki estymator należy stosować w konkretnej sytuacji?

Rozwiązuje się go w ten sposób, że wprowadza się kilka kryteriów, które powinien spełniać "dobry" estymator, a następnie bada się, czy rozpatrywany przez nas estymator spełnia te kryteria.

Estymatory

Definicja

Estymator $\hat{\Theta}_n$ parametru Θ będziemy nazywać **nieobciążonym** jeżeli (dla wszystkich n)

$$E(\hat{\Theta}_n) = \Theta.$$

Zadanie

Czy średnia z próby $X_1, \ldots, X_n$ jest nieobciążonym estymatorem parametru średniej $E(X)$?

Oznaczmy

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Zadanie

Czy średnia z próby $X_1, \ldots, X_n$ jest nieobciążonym estymatorem parametru średniej $E(X)$?

Oznaczmy

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} x_i.$$

Mamy pokazać, że

$$E(\bar{X}) = E(X).$$

Estymatory

Zadanie

Czy średnia z próby $X_1, \ldots, X_n$ jest nieobciążonym estymatorem parametru średniej $E(X)$?

Oznaczmy

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Mamy pokazać, że

$$E(\bar{X}) = E(X).$$

Czyli

$$E(\bar{X}) = E\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{1}{n}E(X_1 + \cdots + X_n) = \frac{1}{n}nE(X) = E(X).$$

Estymatory

Zadanie

Niech $X_1, \ldots, X_n$ będzie próbą prostą pobraną z populacji, w której cecha X ma skończoną i różną od zera wartość σ^2, gdy wartość oczekiwana jest znana $E(X) = m$. Pokaż, że wariancja empiryczna $S_1^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - m)^2$. jest estymatorem nieobciążonym nieznanej wariancji σ^2.

Estymatory

Zadanie

Niech $X_1, \ldots, X_n$ będzie próbą prostą pobraną z populacji, w której cecha X ma skończoną i różną od zera wartość σ^2, gdy wartość oczekiwana jest znana $E(X) = m$. Pokaż, że wariancja empiryczna $S_1^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - m)^2$. jest estymatorem nieobciążonym nieznanej wariancji σ^2.

Mamy pokazać, że:

$$E(S_1^2(X_1, \ldots, X_n)) = D^2(X).$$

Zadanie

Niech $X_1, \ldots, X_n$ będzie próbą prostą pobraną z populacji, w której cecha X ma skończoną i różną od zera wartość σ^2, gdy wartość oczekiwana jest znana $E(X) = m$. Pokaż, że wariancja empiryczna $S_1^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - m)^2$. jest estymatorem nieobciążonym nieznanej wariancji σ^2.

Mamy pokazać, że:

$$E(S_1^2(X_1, \ldots, X_n)) = D^2(X).$$

Czyli

$$E(S_1^2(X_1, \ldots, X_n)) = E\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - m)^2\right) =$$
$$= \frac{1}{n}\sum_{i=1}^{n} E((X_i - m)^2) = \frac{1}{n} n D^2(X) = D^2(X).$$

Zadanie*

W przypadku, gdy nie znamy nadziei matematycznej m, możemy także estymować wariancję - definiujemy wtedy S^2 następująco:

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2, \text{ gdzie } \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Pokazać, że jest to estymator obciążony oraz zachodzi wzór:

$$E(S^2(X_1, \dots, X_n)) = \frac{n-1}{n} D^2(X).$$

Estymatory

Zadanie*

Wykorzystując wynik z powyższego zadania pokaż, że:

$$S^{2*} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2, \text{ gdzie } \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

jet estymatorem nieobciążonym nieznanej wariancji σ^2.

Estymatory

Zrozumiałym jest też, że wymagamy, aby ze wzrostem liczności próbki wzrastała dokładność oszacowanego parametru Θ.

Zrozumiałym jest też, że wymagamy, aby ze wzrostem liczności próbki wzrastała dokładność oszacowanego parametru Θ.
Wymaganie to prowadzi do spełnienia (dla każdej liczby $\epsilon > 0$) warunku:

$$\lim_{n\to\infty} P(|\hat{\Theta}_n - \Theta| < \epsilon) = 1.$$

oznaczającego zbieżność w sensie prawdopodobieństwa (zbieżność stochastyczną) Θ_n do Θ.

Zrozumiałym jest też, że wymagamy, aby ze wzrostem liczności próbki wzrastała dokładność oszacowanego parametru Θ.
Wymaganie to prowadzi do spełnienia (dla każdej liczby $\epsilon > 0$) warunku:

$$\lim_{n\to\infty} P(|\hat{\Theta}_n - \Theta| < \epsilon) = 1.$$

oznaczającego zbieżność w sensie prawdopodobieństwa (zbieżność stochastyczną) Θ_n do Θ.

Definicja

Estymator Θ_n spełniający powyższy warunek nazywamy **estymatorem zgodnym** parametru Θ.

Estymacja Punktowa (Metoda największej wiarygodności)

Niech rozkład badanej cechy X zależy od k nieznanych parametrów

$$\theta_1, \ldots, \theta_1,$$

które chcemy oszacować na podstawie próbki $X_1, \ldots, X_n$.

Estymacja Punktowa (Metoda największej wiarygodności)

Etap 1.
Najpierw wyznaczamy funkcję wiarygodności próby zgodnie ze wzorami:

$$L(x_1, ..., x_n; \theta_1, ..., \theta_r) = \prod_{i=1}^{n} f(x_i; \theta_1, ..., \theta_r) \text{ dla rozkładów ciągłych}$$

$$L(x_1, ..., x_n; \theta_1, ..., \theta_r) = \prod_{i=1}^{n} p(x_i; \theta_1, ..., \theta_r) \text{ dla rozkładów skokowych}$$

gdzie f oznacza funkcję gęstości rozkładu, zaś p funkcję prawdopodobieństwa.

Estymacja Punktowa (Metoda największej wiarygodności)

Etap 2.
Wyznaczamy

$$\ln\left(L(x_1, ..., x_n; \theta_1, ..., \theta_r)\right)$$

Etap 2.
Wyznaczamy

$$\ln\left(L(x_1, ..., x_n; \theta_1, ..., \theta_r)\right)$$

Etap 3.
Wyznaczamy pochodne cząstkowe:

$$\frac{\partial \ln L}{\partial \theta_i} \text{ dla } i = 1, ..., r.$$

Gdy $L(\theta)$ jest dyskretna nie możemy różniczkować, wyliczamy $\frac{L(n+1)}{L(n)}$.
Wiarygodność wtedy jest maksymalizowana przez najmniejsze n, przy którym ten stosunek jest ≤ 1.
(Funkcje $\ln L(\theta)$ i $L(\theta)$ osiągają maksimum dla tej samej wartości, a często zamiast $L(\theta)$ wygodniej jest używać logarytmu funkcji wiarygodności.)

Etap 2.
Wyznaczamy:

$$\ln\left(L(x_1, ..., x_n; \theta_1, ..., \theta_r)\right)$$

Etap 3.
Wyznaczamy pochodne cząstkowe:

$$\frac{\partial \ln L}{\partial \theta_i} \text{ dla } i = 1, ..., r.$$

Etap 4.
Rozwiązujemy układ równań:

$$\frac{\partial \ln L}{\partial \theta_i} = 0$$

względem θ_i.

Estymacja Punktowa (Metoda największej wiarygodności)

Etap 2.
Wyznaczamy:

$$\ln\left(L(x_1, ..., x_n; \theta_1, ..., \theta_r)\right)$$

Etap 3.
Wyznaczamy pochodne cząstkowe:

$$\frac{\partial \ln L}{\partial \theta_i} \text{ dla } i = 1, ..., r.$$

Etap 4.
Rozwiązujemy układ równań:

$$\frac{\partial \ln L}{\partial \theta_i} = 0$$

względem θ_i.

Rozwiązania układu stanowią estymatory szukanych parametrów.

Estymacja Punktowa (Metoda największej wiarygodności)

Zadanie

Znaleźć estymatory największej wiarygodności parametrów m oraz σ rozkładu normalnego.

Estymacja Punktowa (Metoda największej wiarygodności)

Zadanie

Znaleźć estymatory największej wiarygodności parametrów m oraz σ rozkładu normalnego.

Wyznaczamy funkcję wiarygodności:

$$L(X, m, \sigma) = \prod_{i=1}^{n} f_{m,\sigma}(x_i).$$

Estymacja Punktowa (Metoda największej wiarygodności)

Zadanie

Znaleźć estymatory największej wiarygodności parametrów m oraz σ rozkładu normalnego.

Wyznaczamy funkcję wiarygodności:

$$L(X, m, \sigma) = \prod_{i=1}^{n} f_{m,\sigma}(x_i).$$

Wyznaczamy logarytmiczną funkcję wiarygodności:

$$l = \ln\left(L(X, m, \sigma)\right) = \ln\left(\prod_{i=1}^{n} f_{m,\sigma}(x_i)\right) = \sum_{i=1}^{n} \ln(f_{m,\sigma}(x_i))$$

Mamy

$$l(X) = \sum_{i=1}^{n} \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}\right) = \sum_{i=1}^{n} \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) + \sum_{i=1}^{n} \ln\left(e^{-\frac{(x-m)^2}{2\sigma^2}}\right) =$$

Estymacja Punktowa (Metoda największej wiarygodności)

Mamy

$$l(X) = \sum_{i=1}^{n} \ln(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}) = \sum_{i=1}^{n} \ln(\frac{1}{\sqrt{2\pi\sigma^2}}) + \sum_{i=1}^{n} \ln(e^{-\frac{(x-m)^2}{2\sigma^2}}) =$$

$$= \sum_{i=1}^{n} \ln(\frac{1}{\sqrt{2\pi\sigma^2}}) - \sum_{i=1}^{n} \frac{(x-m)^2}{2\sigma^2} =$$

Estymacja Punktowa (Metoda największej wiarygodności)

Mamy

$$
\begin{aligned}
l(X) = \sum_{i=1}^{n} \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}\right) = \sum_{i=1}^{n} \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) + \sum_{i=1}^{n} \ln\left(e^{-\frac{(x-m)^2}{2\sigma^2}}\right) = \\
= \sum_{i=1}^{n} \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \sum_{i=1}^{n} \frac{(x-m)^2}{2\sigma^2} = \\
= \sum_{i=1}^{n} \ln\left(\frac{1}{\sqrt{2\pi}}\right) + \sum_{i=1}^{n} \ln\left(\frac{1}{\sigma}\right) - \sum_{i=1}^{n} \frac{(x-m)^2}{2\sigma^2} =
\end{aligned}
$$

Mamy

$$
\begin{aligned}
l(X) = \sum_{i=1}^{n} \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}\right) &= \sum_{i=1}^{n} \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) + \sum_{i=1}^{n} \ln\left(e^{-\frac{(x-m)^2}{2\sigma^2}}\right) = \\
&= \sum_{i=1}^{n} \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \sum_{i=1}^{n} \frac{(x-m)^2}{2\sigma^2} = \\
&= \sum_{i=1}^{n} \ln\left(\frac{1}{\sqrt{2\pi}}\right) + \sum_{i=1}^{n} \ln\left(\frac{1}{\sigma}\right) - \sum_{i=1}^{n} \frac{(x-m)^2}{2\sigma^2} = \\
&= \sum_{i=1}^{n} \ln\left(\frac{1}{\sqrt{2\pi}}\right) - n\ln(\sigma) - \sum_{i=1}^{n} \frac{(x-m)^2}{2\sigma^2}
\end{aligned}
$$

Policzmy pochodną funkcji wiarygodności po parametrze m i przyrównajmy ją do zera:

$$\frac{\partial}{\partial m} L(X, m_i, \sigma) = \frac{\partial}{\partial m}\left(\sum_{i=1}^{n} \ln(\frac{1}{\sqrt{2\pi}}) - n\ln(\sigma) - \sum_{i=1}^{n} \frac{(x-m)^2}{2\sigma^2} \right) =$$

Estymacja Punktowa (Metoda największej wiarygodności)

Policzmy pochodną funkcji wiarygodności po parametrze m i przyrównajmy ją do zera:

$$\frac{\partial}{\partial m} L(X, m_i, \sigma) = \frac{\partial}{\partial m}\left(\sum_{i=1}^{n} \ln(\tfrac{1}{\sqrt{2\pi}}) - n\ln(\sigma) - \sum_{i=1}^{n} \frac{(x-m)^2}{2\sigma^2}\right) =$$
$$= \sum_{i=1}^{n} \frac{\partial}{\partial m}\left(-\frac{(x-m)^2}{2\sigma^2}\right) =$$

Estymacja Punktowa (Metoda największej wiarygodności)

Policzmy pochodną funkcji wiarygodności po parametrze m i przyrównajmy ją do zera:

$$\frac{\partial}{\partial m} L(X, m_i, \sigma) = \frac{\partial}{\partial m}\left(\sum_{i=1}^{n} \ln(\frac{1}{\sqrt{2\pi}}) - n\ln(\sigma) - \sum_{i=1}^{n} \frac{(x-m)^2}{2\sigma^2}\right) =$$

$$= \sum_{i=1}^{n} \frac{\partial}{\partial m}\left(-\frac{(x-m)^2}{2\sigma^2}\right) =$$

$$= \sum_{i=1}^{n} \frac{2(x_i-m)}{2\sigma^2} = \frac{1}{\sigma^2}(\sum_{i=1}^{n} x_i - nm) = 0$$

Estymacja Punktowa (Metoda największej wiarygodności)

Policzmy pochodną funkcji wiarygodności po parametrze m i przyrównajmy ją do zera:

$$\frac{\partial}{\partial m} L(X, m_i, \sigma) = \frac{\partial}{\partial m}\left(\sum_{i=1}^{n} \ln\left(\frac{1}{\sqrt{2\pi}}\right) - n\ln(\sigma) - \sum_{i=1}^{n} \frac{(x-m)^2}{2\sigma^2}\right) =$$

$$= \sum_{i=1}^{n} \frac{\partial}{\partial m}\left(-\frac{(x-m)^2}{2\sigma^2}\right) =$$

$$= \sum_{i=1}^{n} \frac{2(x_i - m)}{2\sigma^2} = \frac{1}{\sigma^2}\left(\sum_{i=1}^{n} x_i - nm\right) = 0$$

Otrzymujemy równanie:

$$\sum_{i=1}^{n} x_i - nm = 0,$$

$$\sum_{i=1}^{n} x_i = nm,$$

$$m = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Estymacja Punktowa (Metoda największej wiarygodności)

Policzmy pochodną funkcji wiarygodności po parametrze σ i przyrównajmy ją do zera:

$$\frac{\partial}{\partial\sigma} L(X, m_i, \sigma) = \frac{\partial}{\partial\sigma} \left(\sum_{i=1}^{n} \ln(\frac{1}{\sqrt{2\pi}}) - n\ln(\sigma) - \sum_{i=1}^{n} \frac{(x-m)^2}{2\sigma^2} \right) =$$

Estymacja Punktowa (Metoda największej wiarygodności)

Policzmy pochodną funkcji wiarygodności po parametrze σ i przyrównajmy ją do zera:

$$\frac{\partial}{\partial\sigma}L(X, m_i, \sigma) = \frac{\partial}{\partial\sigma}\left(\sum_{i=1}^{n}\ln(\tfrac{1}{\sqrt{2\pi}}) - n\ln(\sigma) - \sum_{i=1}^{n}\frac{(x-m)^2}{2\sigma^2}\right) =$$

$$= \frac{\partial}{\partial\sigma}\left(-n\ln(\sigma) - \sum_{i=1}^{n}\frac{(x-m)^2}{2\sigma^2}\right) =$$

Estymacja Punktowa (Metoda największej wiarygodności)

Policzmy pochodną funkcji wiarygodności po parametrze σ i przyrównajmy ją do zera:

$$\frac{\partial}{\partial\sigma}L(X,m_i,\sigma)=\frac{\partial}{\partial\sigma}\left(\sum_{i=1}^{n}\ln\left(\frac{1}{\sqrt{2\pi}}\right)-n\ln(\sigma)-\sum_{i=1}^{n}\frac{(x-m)^2}{2\sigma^2}\right)=$$

$$=\frac{\partial}{\partial\sigma}\left(-n\ln(\sigma)-\sum_{i=1}^{n}\frac{(x-m)^2}{2\sigma^2}\right)=$$

$$=\frac{-n}{\sigma}-\frac{1}{2\sigma^3}(-2)\sum_{i=1}^{n}(x_i-\bar{x})^2=-\frac{n}{\sigma}+\frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i-\bar{x})^2=0.$$

Policzmy pochodną funkcji wiarygodności po parametrze σ i przyrównajmy ją do zera:

$$\frac{\partial}{\partial\sigma}L(X, m_i, \sigma) = \frac{\partial}{\partial\sigma}\left(\sum_{i=1}^{n}\ln(\tfrac{1}{\sqrt{2\pi}}) - n\ln(\sigma) - \sum_{i=1}^{n}\frac{(x-m)^2}{2\sigma^2}\right) =$$
$$= \frac{\partial}{\partial\sigma}\left(-n\ln(\sigma) - \sum_{i=1}^{n}\frac{(x-m)^2}{2\sigma^2}\right) =$$
$$= \frac{-n}{\sigma} - \frac{1}{2\sigma^3}(-2)\sum_{i=1}^{n}(x_i - \bar{x})^2 = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i - \bar{x})^2 = 0.$$

Otrzymujemy równanie:

$$-\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i - \bar{x})^2 = 0,$$
$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \bar{x})^2 = n,$$
$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

Zadanie:
https://github.com/przem85/statistic_4/blob/master/D04_Z01_maximum_likelihood_estimation.ipynb