Metody probabilistyczne

12. Elementy statystki matematycznej I

Wojciech Kotłowski

Instytut Informatyki PP
http://www.cs.put.poznan.pl/wkotlowski/

9.01.2018

Zagadnienia statystki

Przeprowadzamy pewien eksperyment losowy, którego wynikiem są dane.

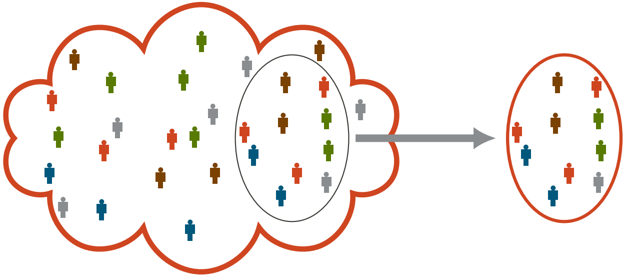
Celem jest wyciągnięcie ogólnych wniosków na temat zbiorowości / zjawiska / procesu, z którego dane pochodzą.

Wnioskowanie statystyczne jest podstawową metodologią badań w większości nauk (np. w medycynie, psychologii, ekonomii, fizyce eksperymentalnej), jak również stanowi bazę nowoczesnych metod sztucznej inteligencji (uczenie maszynowe) i analizy danych

Zagadnienia statystyki – przykłady

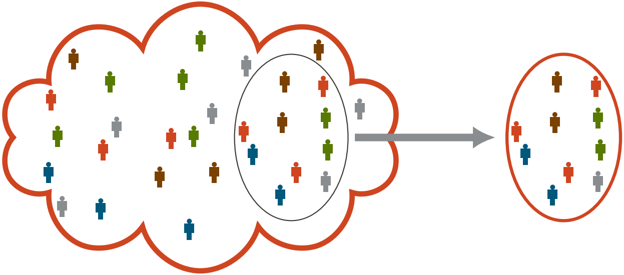
- Sondaż wyborczy na próbie losowo wybranych osób celem oszacowania prawdziwego poparcia dla partii w całej populacji kraju
- Wnioskowanie o zmianach temperatury na podstawie pomiarów ze stacji pogodowych
- Sprawdzenie, czy istnieje zależność między danymi zjawiskami (np. palenie papierosów a zachorowalność na nowotworowy)
- Testowanie skuteczności nowego leku poprzez porównanie wyników próby badawczej (podajemy lek) i kontrolnej (podajemy placebo)
- Szacowanie czasu przejazdu po drogach na podstawie danych z GPS urządzeń mobilnych
- Badania marketingowe celem kierowania do konsumentów personalizowanych reklam

Próba i populacja



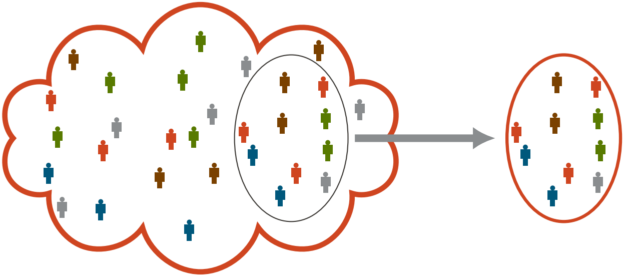
- Populacja (generalna) – cała rozważana zbiorowość, np. wszyscy potencjalni wyborcy, wartości temperatur na całym globie, wszyscy możliwi pacjenci, itp.
- Próba – zespół elementów wylosowany z populacji („dane”), np. wyborcy wybrani do sondażu, wartości temperatur w stacjach pomiarowych, pacjenci wybrani do testowania leku, itp.

Próba i populacja



Celem wnioskowania statystycznego jest wyciągnięcie wniosków o całej populacji na podstawie losowo wybranej próby

Próba i populacja

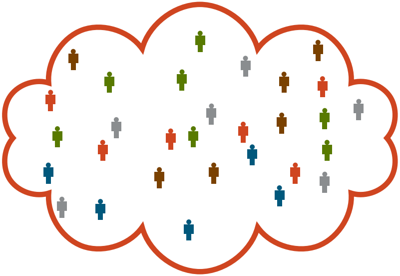


Celem wnioskowania statystycznego jest wyciągnięcie wniosków o całej populacji na podstawie losowo wybranej próby

Uwaga: Zakładamy, że próba jest reprezentatywna (każdy element populacji ma równe szanse na wejście do próby) i prosta (każdy element próby losowany niezależnie od pozostałych).

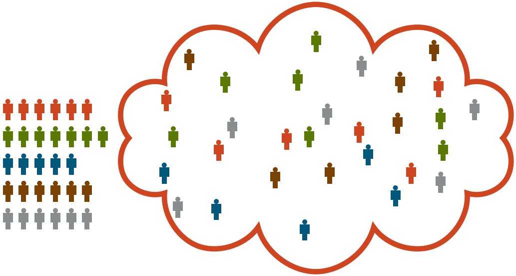
Populacja a rozkład prawdopodobieństwa

Zwykle interesuje nas tylko konkretna cecha populacja, np. preferencje wyborcze, wysokość zarobków, temperatura, itp.



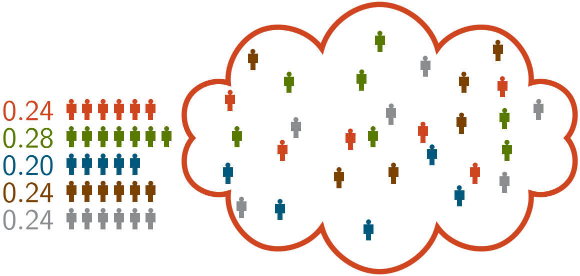
Populacja a rozkład prawdopodobieństwa

Zwykle interesuje nas tylko konkretna cecha populacja, np. preferencje wyborcze, wysokość zarobków, temperatura, itp.



Populacja a rozkład prawdopodobieństwa

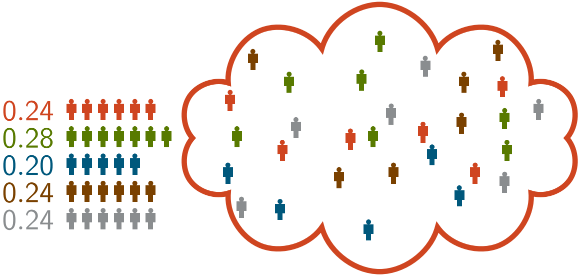
Zwykle interesuje nas tylko konkretna cecha populacja, np. preferencje wyborcze, wysokość zarobków, temperatura, itp.



Zliczając częstości wystąpienia danej wartości cechy populację możemy traktować abstrakcyjnie jako rozkład prawdopodobieństwa cechy

Populacja a rozkład prawdopodobieństwa

Zwykle interesuje nas tylko konkretna cecha populacja, np. preferencje wyborcze, wysokość zarobków, temperatura, itp.



Zliczając częstości wystąpienia danej wartości cechy populację możemy traktować abstrakcyjnie jako rozkład prawdopodobieństwa cechy

Dla ciągłych cech modelujemy populację gęstością prawdopodobieństwa

Statystyka matematyczna

- Badaną cechę X modelujemy jako zmienną losową o pewnym rozkładzie P (na populacji)

Statystyka matematyczna

- Badaną cechę X modelujemy jako zmienną losową o pewnym rozkładzie P (na populacji)
- Próba o liczności n może być modelowana jako realizacje (wartości) n niezależnych zmiennych losowych

$$X_1, X_2, \ldots, X_n$$

  o tym samym rozkładzie P

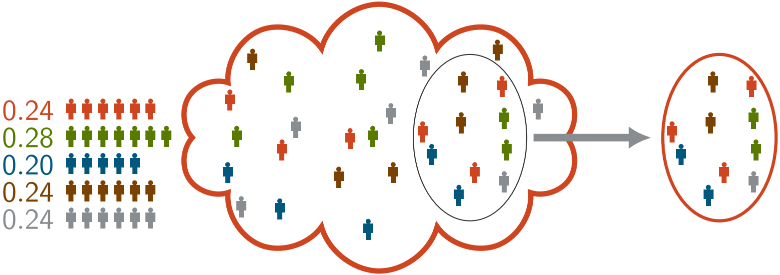
Statystyka matematyczna

- Badaną cechę X modelujemy jako zmienną losową o pewnym rozkładzie P (na populacji)
- Próba o liczności n może być modelowana jako realizacje (wartości) n niezależnych zmiennych losowych

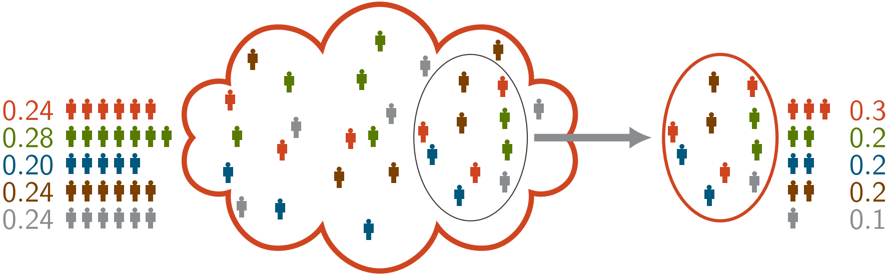
$$X_1, X_2, \ldots, X_n$$

  o tym samym rozkładzie P
- Obserwując próbę pochodzącą z nieznanego rozkładu P, wnioskujemy o interesujących nas własnościach tego rozkładu.

Rozkład empiryczny



Rozkład empiryczny



Rozkład empiryczny powstaje poprzez zliczenie częstości wystąpień poszczególnych wartości cechy na próbie.

Przykład: preferencje wyborcze

- Rozważmy drugą turę wyborów prezydenckich z kandydatami A i B. W populacji kandydat A ma poparcie 60%, a B – 40%

Przykład: preferencje wyborcze

- Rozważmy drugą turę wyborów prezydenckich z kandydatami A i B. W populacji kandydat A ma poparcie 60%, a B – 40%
- Niech $X \in \{0, 1\}$ koduje poparcie kandydata A ($X = 1$) lub B ($X = 0$)

Przykład: preferencje wyborcze

- Rozważmy drugą turę wyborów prezydenckich z kandydatami A i B. W populacji kandydat A ma poparcie 60%, a B – 40%
- Niech $X \in \{0, 1\}$ koduje poparcie kandydata A ($X = 1$) lub B ($X = 0$)
- Cecha X ma więc rozkład dwupunktowy z parametrem $p = 0.6$

Przykład: preferencje wyborcze

- Rozważmy drugą turę wyborów prezydenckich z kandydatami A i B. W populacji kandydat A ma poparcie 60%, a B – 40%
- Niech $X \in \{0, 1\}$ koduje poparcie kandydata A ($X = 1$) lub B ($X = 0$)
- Cecha X ma więc rozkład dwupunktowy z parametrem $p = 0.6$
- Pobraliśmy próbę $n = 10$ elementów otrzymując wartości:

$$0, 1, 0, 1, 1, 1, 0, 0, 1, 0$$

Przykład: preferencje wyborcze

- Rozważmy drugą turę wyborów prezydenckich z kandydatami A i B. W populacji kandydat A ma poparcie 60%, a B – 40%
- Niech $X \in \{0, 1\}$ koduje poparcie kandydata A ($X = 1$) lub B ($X = 0$)
- Cecha X ma więc rozkład dwupunktowy z parametrem $p = 0.6$
- Pobraliśmy próbę $n = 10$ elementów otrzymując wartości:

$$0, 1, 0, 1, 1, 1, 0, 0, 1, 0$$

- Rozkład empiryczny jest rozkładem dwupunktowym z parametrem $\widehat{p} = \frac{5}{10} = 0.5$

Teoria estymacji

W ramach tego wykładu zajmiemy się tylko zadaniem estymacji parametrów rozkładu

Teoria estymacji zajmuje się szacowaniem interesujących parametrów rozkładu na populacji na podstawie próby

Teoria estymacji

W ramach tego wykładu zajmiemy się tylko zadaniem estymacji parametrów rozkładu

Teoria estymacji zajmuje się szacowaniem interesujących parametrów rozkładu na populacji na podstawie próby

Przykład: Cecha w populacji ma rozkład normalny $N(\mu, \sigma^2)$, gdzie μ i σ^2 są nieznane. Szacujemy te wartości na podstawie próby.

Statystyka i estymator

Próba: Niezależne zmienne losowe $X_1, X_2, \ldots, X_n$ o rozkładzie P.

Statystyka i estymator

Próba: Niezależne zmienne losowe $X_1, X_2, \ldots, X_n$ o rozkładzie P.

Statystka

Statystyką $T = T(X_1, X_2, \ldots, X_n)$ nazywamy dowolną funkcję próby

Uwaga: statystyka jest zmienną losową, ponieważ jest funkcją zmiennych losowych!

Statystyka i estymator

Próba: Niezależne zmienne losowe $X_1, X_2, \ldots, X_n$ o rozkładzie P.

Statystka

Statystyką $T = T(X_1, X_2, \ldots, X_n)$ nazywamy dowolną funkcję próby

Uwaga: statystyka jest zmienną losową, ponieważ jest funkcją zmiennych losowych!

Estymator

Estymatorem $\widehat{\theta} = \widehat{\theta}(X_1, X_2, \ldots, X_n)$ nazywamy statystykę szacującą nieznany parametr θ rozkładu P

Uwaga: estymator jest zmienną losową, więc ma swój rozkład, wartość oczekiwaną, wariancję, itp.

Przykład: szacowanie wartości oczekiwanej

Załóżmy, że chcemy oszacować wartość oczekiwaną $\mu = EX$ rozkładu zmiennej losowej X.

Przykład: szacowanie wartości oczekiwanej

Załóżmy, że chcemy oszacować wartość oczekiwaną $\mu = EX$ rozkładu zmiennej losowej X.

Rozważmy estymator parametru μ będący średnią arytmetyczną próby

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

Przykład: szacowanie wartości oczekiwanej

Załóżmy, że chcemy oszacować wartość oczekiwaną $\mu = EX$ rozkładu zmiennej losowej X.

Rozważmy estymator parametru μ będący średnią arytmetyczną próby

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

Zgodnie z prawem wielkich liczb:

$$\overline{X}_n \overset{\text{z pr. 1}}{\longrightarrow} \mu,$$

a więc zwiększając wielkość próby n do nieskończoności, estymator $\widehat{\mu}$ zbiega do prawdziwej wartości parametru μ.

Zgodność estymatora

Estymator $\widehat{\theta}$ parametru θ nazywamy silnie zgodnym jeśli:

$$\widehat{\theta} \xrightarrow{\text{z pr. 1}} \theta$$

Estymator nazywamy słabo zgodnym jeśli

$$\widehat{\theta} \xrightarrow{P} \theta$$

Zgodność estymatora oznacza, że zbiega on do prawdziwej wartości parametru, gdy rozmiar próby n rośnie do nieskończoności

Zgodność estymatora

Estymator $\widehat{\theta}$ parametru θ nazywamy silnie zgodnym jeśli:

$$\widehat{\theta} \stackrel{\text{z pr. } 1}{\longrightarrow} \theta$$

Estymator nazywamy słabo zgodnym jeśli

$$\widehat{\theta} \stackrel{P}{\longrightarrow} \theta$$

Zgodność estymatora oznacza, że zbiega on do prawdziwej wartości parametru, gdy rozmiar próby n rośnie do nieskończoności

Wniosek: estymator $\widehat{\mu} = \overline{X}_n$ jest silnie zgodnym estymatorem wartości oczekiwanej $\mu = EX$.

Przykład: preferencje wyborcze

Rozważmy drugą turę wyborów prezydenckich (kandydaci $A \equiv 1$ i $B \equiv 0$). Zmienna $X \in \{0, 1\}$ (poparcie A) ma rozkład $B(p)$, gdzie $p = EX$ jest poparciem kandydata w populacji.

Przykład: preferencje wyborcze

Rozważmy drugą turę wyborów prezydenckich (kandydaci $A \equiv 1$ i $B \equiv 0$). Zmienna $X \in \{0, 1\}$ (poparcie A) ma rozkład $B(p)$, gdzie $p = EX$ jest poparciem kandydata w populacji.

Estymator wartości oczekiwanej p:

$$\widehat{p} \;=\; \frac{1}{n}\sum_{i=1}^{n} X_i \;=\; \frac{\#\text{sukcesów w próbie}}{n}$$

jest częstością sukcesów w próbie.

Przykład: preferencje wyborcze

Rozważmy drugą turę wyborów prezydenckich (kandydaci $A \equiv 1$ i $B \equiv 0$). Zmienna $X \in \{0, 1\}$ (poparcie A) ma rozkład $B(p)$, gdzie $p = EX$ jest poparciem kandydata w populacji.

Estymator wartości oczekiwanej p:

$$\widehat{p} \;=\; \frac{1}{n}\sum_{i=1}^{n} X_i \;=\; \frac{\#\text{sukcesów w próbie}}{n}$$

jest częstością sukcesów w próbie.

$\widehat{p}$ jest silnie zgodnym estymatorem parametru p.

Metoda momentów

Estymator wartości oczekiwanej $\overline{X}_n$ jest przykładem zastosowania tzw. metody momentów:

Chcąc oszacować moment rozkładu $m_k = E(X^k)$, użyjmy estymatora $\widehat{m}_k$ będącego odpowiednim momentem rozkładu empirycznego:

$$\widehat{m}_k = \frac{1}{n} \sum_{i=1}^{n} X_i^k$$

(zamiast wartości oczekiwanej bierzemy średnią po próbie)

Przykład: estymator wariancji

Chcemy oszacować wariancję $\sigma^2 = D^2(X)$ rozkładu zmiennej losowej X.

Przykład: estymator wariancji

Chcemy oszacować wariancję $\sigma^2 = D^2(X)$ rozkładu zmiennej losowej X.

Ponieważ $D^2(X) = E(X^2) - (EX)^2 = m_2 - m_1^2$, bierzemy estymator wariancji:

$$\widehat{\sigma}^2 \;=\; \widehat{m}_2 - \widehat{m}_1^2 \;=\; \frac{1}{n}\sum_{i=1}^{n} X_i^2 - (\overline{X}_n)^2$$

Przykład: estymator wariancji

Chcemy oszacować wariancję $\sigma^2 = D^2(X)$ rozkładu zmiennej losowej X.

Ponieważ $D^2(X) = E(X^2) - (EX)^2 = m_2 - m_1^2$, bierzemy estymator wariancji:

$$\widehat{\sigma}^2 \;=\; \widehat{m}_2 - \widehat{m}_1^2 \;=\; \frac{1}{n}\sum_{i=1}^{n} X_i^2 - (\overline{X}_n)^2 \;=\; \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

Przykład: estymator wariancji

Chcemy oszacować wariancję $\sigma^2 = D^2(X)$ rozkładu zmiennej losowej X.

Ponieważ $D^2(X) = E(X^2) - (EX)^2 = m_2 - m_1^2$, bierzemy estymator wariancji:

$$\widehat{\sigma}^2 \;=\; \widehat{m}_2 - \widehat{m}_1^2 \;=\; \frac{1}{n}\sum_{i=1}^{n} X_i^2 - (\overline{X}_n)^2 \;=\; \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

Zadanie 1

Uzasadnij, ostatnią równość, tzn. pokaż wzór skróconego mnożenia:

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 \;=\; \frac{1}{n}\sum_{i=1}^{n} X_i^2 - (\overline{X}_n)^2$$

Zwróć uwagę na analogię do $E\left((X - EX)^2\right) = E(X^2) - (EX)^2$, tyle że zamiast uśredniać po rozkładzie, uśredniamy po próbie.

Przykład: estymator wariancji

Chcemy oszacować wariancję $\sigma^2 = D^2(X)$ rozkładu zmiennej losowej X.

Ponieważ $D^2(X) = E(X^2) - (EX)^2 = m_2 - m_1^2$, bierzemy estymator wariancji:

$$\widehat{\sigma}^2 \; = \; \widehat{m}_2 - \widehat{m}_1^2 \; = \; \frac{1}{n}\sum_{i=1}^{n} X_i^2 - (\overline{X}_n)^2 \; = \; \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

Zadanie 2

Pokaż, że $\widehat{\sigma}^2$ jest silnie zgodny poprzez dwukrotne zastosowania prawa wielkich liczb do $\frac{1}{n}\sum_{i=1}^{n} X_i^2$ oraz do $\overline{X}_n$.

Estymator nieobciążony

Zgodność estymatora jest własnością asymptotyczną: mówi o zachowaniu estymatora tylko dla bardzo dużych prób ($n \longrightarrow \infty$)

Estymator nieobciążony

Zgodność estymatora jest własnością asymptotyczną: mówi o zachowaniu estymatora tylko dla bardzo dużych prób ($n \to \infty$)

Obciążenie estymatora

Obciążeniem estymatora $\widehat{\theta}$ parametru θ nazywamy różnicę między wartością oczekiwaną estymatora a nieznaną wartością parametru:

$$E(\widehat{\theta}) - \theta$$

Estymator nieobciążony

Zgodność estymatora jest własnością asymptotyczną: mówi o zachowaniu estymatora tylko dla bardzo dużych prób ($n \to \infty$)

Obciążenie estymatora

Obciążeniem estymatora $\widehat{\theta}$ parametru θ nazywamy różnicę między wartością oczekiwaną estymatora a nieznaną wartością parametru:

$$E(\widehat{\theta}) - \theta$$

Estymator nieobciążony

Estymator nazywamy nieobciążonym jeśli jego obciążenie jest równe zero, tzn:

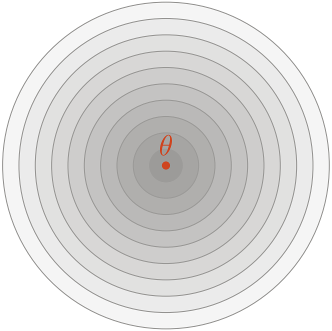
$$E(\widehat{\theta}) = \theta$$

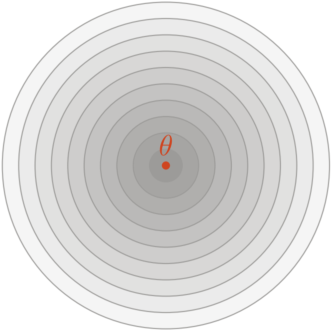
Estymator „średnio” trafia w wybrany parametr.

Uwaga: to ma sens, bo estymator jest zmienną losową, więc ma swoją wartość oczekiwaną!

Estymator nieobciążony

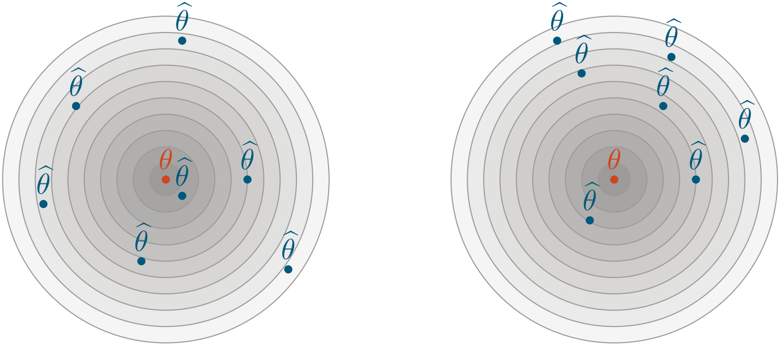
Zgodnie z prawem wielkich liczb, wartość oczekiwaną estymatora $E(\widehat{\theta})$ można interpretować jako średnią wartość estymatora przy wielokrotnym losowaniu próby





Estymator nieobciążony

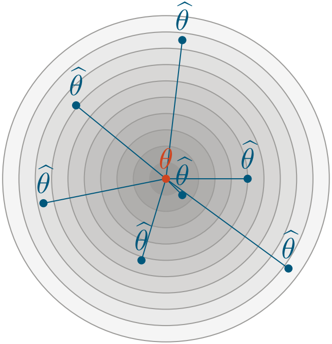
Zgodnie z prawem wielkich liczb, wartość oczekiwaną estymatora $E(\widehat{\theta})$ można interpretować jako średnią wartość estymatora przy wielokrotnym losowaniu próby



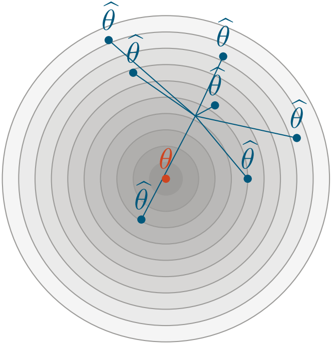
Estymator nieobciążony

Zgodnie z prawem wielkich liczb, wartość oczekiwaną estymatora $E(\widehat{\theta})$ można interpretować jako średnią wartość estymatora przy wielokrotnym losowaniu próby

Estymator nieobciążony



Estymator obciążony



Estymator wartości oczekiwanej

Estymator wartości oczekiwanej $\mu = EX$ postaci:

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

jest nieobciążony

Estymator wartości oczekiwanej

Estymator wartości oczekiwanej $\mu = EX$ postaci:

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

jest nieobciążony

Dowód:

$$E(\widehat{\mu}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)$$

Estymator wartości oczekiwanej

Estymator wartości oczekiwanej $\mu = EX$ postaci:

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

jest nieobciążony

Dowód:

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

Estymator wartości oczekiwanej

Estymator wartości oczekiwanej $\mu = EX$ postaci:

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

jest nieobciążony

Dowód:

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

Estymator wartości oczekiwanej

Estymator wartości oczekiwanej $\mu = EX$ postaci:

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

jest nieobciążony

Dowód:

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

(jest to przypomnienie z wykładu o twierdzeniach granicznych, gdzie już udowodniliśmy, że $E\overline{X}_n = \mu$)

Estymator wariancji

Wyznacz obciążenie estymatora wariancji $\sigma^2 = D^2(X)$ postaci:

$$\widehat{\sigma}^2 \;=\; \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 \;=\; \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}_n^2$$

Estymator wariancji

Wyznacz obciążenie estymatora wariancji $\sigma^2 = D^2(X)$ postaci:

$$\widehat{\sigma}^2 \;=\; \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 \;=\; \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}_n^2$$

Odpowiedź:

$$E(\widehat{\sigma}^2) \;=\; \frac{1}{n}\sum_{i=1}^{n} E(X_i^2) - E(\overline{X}_n^2) \;=\; E(X^2) - E(\overline{X}_n^2)$$

Estymator wariancji

Wyznacz obciążenie estymatora wariancji $\sigma^2 = D^2(X)$ postaci:

$$\widehat{\sigma}^2 \;=\; \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 \;=\; \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}_n^2$$

Odpowiedź:

$$E(\widehat{\sigma}^2) \;=\; \frac{1}{n}\sum_{i=1}^{n} E(X_i^2) - E(\overline{X}_n^2) \;=\; E(X^2) - E(\overline{X}_n^2)$$

Ze wzoru skróconego mnożenia mamy:

$$D^2(X) = E(X^2) - (\underbrace{EX}_{=\mu})^2, \qquad \text{oraz} \quad D^2(\overline{X}_n) = E(\overline{X}_n^2) - (\underbrace{E\overline{X}_n}_{=\mu})^2,$$

Estymator wariancji

Wyznacz obciążenie estymatora wariancji $\sigma^2 = D^2(X)$ postaci:

$$\widehat{\sigma}^2 \;=\; \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 \;=\; \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}_n^2$$

Odpowiedź:

$$E(\widehat{\sigma}^2) \;=\; \frac{1}{n}\sum_{i=1}^{n} E(X_i^2) - E(\overline{X}_n^2) \;=\; E(X^2) - E(\overline{X}_n^2)$$

Ze wzoru skróconego mnożenia mamy:

$$D^2(X) = E(X^2) - (\underbrace{EX}_{=\mu})^2, \qquad \text{oraz} \quad D^2(\overline{X}_n) = E(\overline{X}_n^2) - (\underbrace{E\overline{X}_n}_{=\mu})^2,$$

stąd:

$$E(\widehat{\sigma}^2) \;=\; D^2(X) + \mu^2 - \left(D^2(\overline{X}_n) + \mu^2\right) \;=\; D^2(X) - D^2(\overline{X}_n)$$

Estymator wariancji

Wyznacz obciążenie estymatora wariancji $\sigma^2 = D^2(X)$ postaci:

$$\widehat{\sigma}^2 \;=\; \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 \;=\; \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}_n^2$$

Odpowiedź:

$$E(\widehat{\sigma}^2) \;=\; D^2(X) - D^2(\overline{X}_n)$$

Estymator wariancji

Wyznacz obciążenie estymatora wariancji $\sigma^2 = D^2(X)$ postaci:

$$\widehat{\sigma}^2 \;=\; \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 \;=\; \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}_n^2$$

Odpowiedź:

$$E(\widehat{\sigma}^2) \;=\; D^2(X) - D^2(\overline{X}_n)$$

Zauważmy, że:

$$D^2(X) \;=\; \sigma^2$$

Estymator wariancji

Wyznacz obciążenie estymatora wariancji $\sigma^2 = D^2(X)$ postaci:

$$\widehat{\sigma}^2 \;=\; \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 \;=\; \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}_n^2$$

Odpowiedź:

$$E(\widehat{\sigma}^2) \;=\; D^2(X) - D^2(\overline{X}_n)$$

Zauważmy, że:

$$\begin{aligned} D^2(X) &= \sigma^2 \\ D^2(\overline{X}_n) &= D^2\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \end{aligned}$$

Estymator wariancji

Wyznacz obciążenie estymatora wariancji $\sigma^2 = D^2(X)$ postaci:

$$\widehat{\sigma}^2 \; = \; \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 \; = \; \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}_n^2$$

Odpowiedź:

$$E(\widehat{\sigma}^2) \; = \; D^2(X) - D^2(\overline{X}_n)$$

Zauważmy, że:

$$D^2(X) \; = \; \sigma^2$$

$$D^2(\overline{X}_n) \; = \; D^2\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \; = \; \frac{1}{n^2}\sum_{i=1}^{n}\underbrace{D^2(X_i)}_{\sigma^2} \; = \; \frac{\sigma^2}{n}$$

Estymator wariancji

Wyznacz obciążenie estymatora wariancji $\sigma^2 = D^2(X)$ postaci:

$$\widehat{\sigma}^2 \;=\; \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 \;=\; \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}_n^2$$

Odpowiedź:

$$E(\widehat{\sigma}^2) \;=\; D^2(X) - D^2(\overline{X}_n)$$

Zauważmy, że:

$$D^2(X) \;=\; \sigma^2$$

$$D^2(\overline{X}_n) \;=\; D^2\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \;=\; \frac{1}{n^2}\sum_{i=1}^{n}\underbrace{D^2(X_i)}_{\sigma^2} \;=\; \frac{\sigma^2}{n}$$

Stąd:

$$E(\widehat{\sigma}^2) \;=\; \sigma^2 - \frac{\sigma^2}{n} \;=\; \frac{n-1}{n}\sigma^2 \;\neq\; \sigma^2$$

Estymator wariancji $\widehat{\sigma}^2$ jest obciążony (niedoszacowuje!)

Ratujemy estymator wariancji

Pokazaliśmy, że:

$$E(\widehat{\sigma}^2) = \frac{n-1}{n}\sigma^2, \qquad \text{gdzie} \;\; \widehat{\sigma}^2 \;=\; \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

Ratujemy estymator wariancji

Pokazaliśmy, że:

$$E(\widehat{\sigma}^2) = \frac{n-1}{n}\sigma^2, \qquad \text{gdzie} \;\; \widehat{\sigma}^2 \;=\; \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

Modyfikujemy estymator wariancji:

$$\widehat{\sigma}^2_* \;=\; \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 \;=\; \frac{n}{n-1}\widehat{\sigma}^2$$

Ratujemy estymator wariancji

Pokazaliśmy, że:

$$E(\widehat{\sigma}^2) = \frac{n-1}{n}\sigma^2, \qquad \text{gdzie } \widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

Modyfikujemy estymator wariancji:

$$\widehat{\sigma}^2_* = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 = \frac{n}{n-1}\widehat{\sigma}^2$$

Nowy estymator jest nieobciążony, ponieważ:

$$E(\widehat{\sigma}^2_*) = E\left(\frac{n}{n-1}\widehat{\sigma}^2\right) = \frac{n}{n-1}E(\widehat{\sigma}^2) = \frac{\cancel{n}}{\cancel{n-1}}\frac{\cancel{n-1}}{\cancel{n}}\sigma^2 = \sigma^2$$

Podsumowanie

Nieobciążony estymator wartości średniej $\mu = EX$:

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

Nieobciążony estymator wariancji $\sigma^2 = D^2(X)$:

$$\widehat{\sigma}^2_* \;=\; \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

Brak obciążenia nie wystarcza

Łatwo uzyskać estymator nieobciążony, który jest trywialny

Brak obciążenia nie wystarcza

Łatwo uzyskać estymator nieobciążony, który jest trywialny

Rozważmy estymator wartości oczekiwanej postaci $\widehat{\mu}_1 = X_1$

Brak obciążenia nie wystarcza

Łatwo uzyskać estymator nieobciążony, który jest trywialny

Rozważmy estymator wartości oczekiwanej postaci $\widehat{\mu}_1 = X_1$

Mamy: $E(\widehat{\mu}_1) = EX_1 = \mu$ (estymator nieobciążony!)

Brak obciążenia nie wystarcza

Łatwo uzyskać estymator nieobciążony, który jest trywialny

Rozważmy estymator wartości oczekiwanej postaci $\widehat{\mu}_1 = X_1$

Mamy: $E(\widehat{\mu}_1) = EX_1 = \mu$ (estymator nieobciążony!)

Ale ten estymator odrzuca wszystkie elementy próby poza X_1!

Brak obciążenia nie wystarcza

Łatwo uzyskać estymator nieobciążony, który jest trywialny

Rozważmy estymator wartości oczekiwanej postaci $\widehat{\mu}_1 = X_1$

$$\text{Mamy:} \quad E(\widehat{\mu}_1) = EX_1 = \mu \qquad \text{(estymator nieobciążony!)}$$

Ale ten estymator odrzuca wszystkie elementy próby poza X_1!

Np. przy szacowaniu poparcia kandydata A w wyborach, $\widehat{\mu}_1$ zwraca wartość równą pierwszemu elementowi próby, czyli 0 lub 1!

Brak obciążenia nie wystarcza

Łatwo uzyskać estymator nieobciążony, który jest trywialny

Rozważmy estymator wartości oczekiwanej postaci $\widehat{\mu}_1 = X_1$

Mamy: $E(\widehat{\mu}_1) = EX_1 = \mu$ (estymator nieobciążony!)

Ale ten estymator odrzuca wszystkie elementy próby poza X_1!

Np. przy szacowaniu poparcia kandydata A w wyborach, $\widehat{\mu}_1$ zwraca wartość równą pierwszemu elementowi próby, czyli 0 lub 1!

Średnio jest to poprawne (równe prawdziwemu poparciu), ale jest to raczej kiepskie oszacowanie poparcia.

Brak obciążenia nie wystarcza

Łatwo uzyskać estymator nieobciążony, który jest trywialny

Rozważmy estymator wartości oczekiwanej postaci $\widehat{\mu}_1 = X_1$

$$\text{Mamy: } E(\widehat{\mu}_1) = EX_1 = \mu \qquad \text{(estymator nieobciążony!)}$$

Ale ten estymator odrzuca wszystkie elementy próby poza X_1!

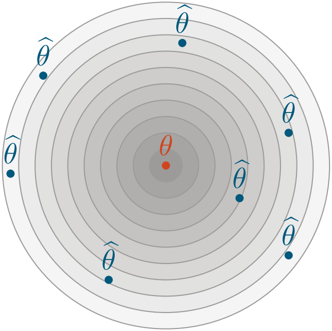
Np. przy szacowaniu poparcia kandydata A w wyborach, $\widehat{\mu}_1$ zwraca wartość równą pierwszemu elementowi próby, czyli 0 lub 1!

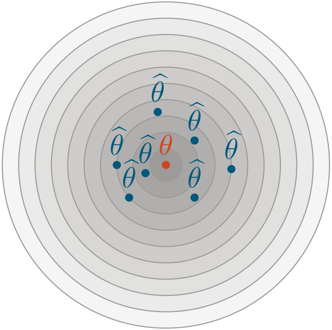
Średnio jest to poprawne (równe prawdziwemu poparciu), ale jest to raczej kiepskie oszacowanie poparcia.

Potrzebujemy innych kryteriów oceny estymatorów

Wariancja estymatora

Chcemy, aby nieobciążony estymator miał jak najmniejszą wariancję $D^2(\widehat{\theta})$

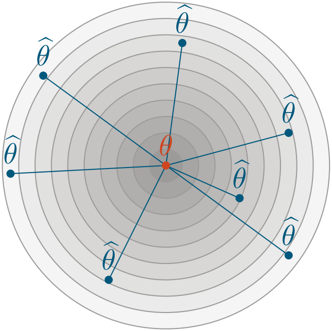




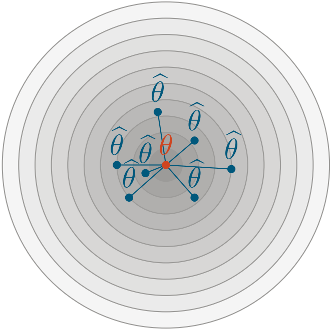
Wariancja estymatora

Chcemy, aby nieobciążony estymator miał jak najmniejszą wariancję $D^2(\widehat{\theta})$

Estymator nieobciążony



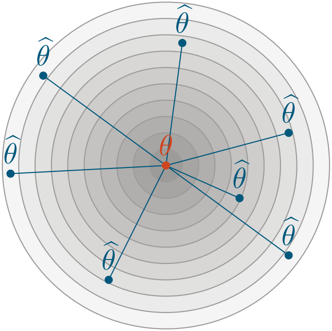
Estymator nieobciążony



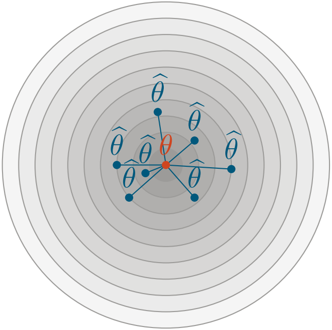
Wariancja estymatora

Chcemy, aby nieobciążony estymator miał jak najmniejszą wariancję $D^2(\widehat{\theta})$

Estymator nieobciążony
Duża wariancja



Estymator nieobciążony
Mała wariancja



Relacja efektywności

Rozważmy dwa estymatory nieobciążone $\widehat{\theta}_1$ i $\widehat{\theta}_2$ parametru θ. Mówimy, że $\widehat{\theta}_1$ jest efektywniejszy od estymatora $\widehat{\theta}_2$, jeśli

$$D^2(\widehat{\theta}_1) \leqslant D^2(\widehat{\theta}_2)$$

dla każdej wartości parametru θ.

Relacja efektywności

Rozważmy dwa estymatory nieobciążone $\widehat{\theta}_1$ i $\widehat{\theta}_2$ parametru θ. Mówimy, że $\widehat{\theta}_1$ jest efektywniejszy od estymatora $\widehat{\theta}_2$, jeśli

$$D^2(\widehat{\theta}_1) \leqslant D^2(\widehat{\theta}_2)$$

dla każdej wartości parametru θ.

- Ściślej powinno być: „efektywniejszy lub tak samo efektywny jak”

Relacja efektywności

Rozważmy dwa estymatory nieobciążone $\widehat{\theta}_1$ i $\widehat{\theta}_2$ parametru θ. Mówimy, że $\widehat{\theta}_1$ jest efektywniejszy od estymatora $\widehat{\theta}_2$, jeśli

$$D^2(\widehat{\theta}_1) \leqslant D^2(\widehat{\theta}_2)$$

dla każdej wartości parametru θ.

- Ściślej powinno być: „efektywniejszy lub tak samo efektywny jak"
- Warunek „dla każdego θ" jest potrzebny, ponieważ wariancje estymatorów mogą zależeć od parametru rozkładu

Relacja efektywności

Rozważmy dwa estymatory nieobciążone $\widehat{\theta}_1$ i $\widehat{\theta}_2$ parametru θ. Mówimy, że $\widehat{\theta}_1$ jest efektywniejszy od estymatora $\widehat{\theta}_2$, jeśli

$$D^2(\widehat{\theta}_1) \leqslant D^2(\widehat{\theta}_2)$$

dla każdej wartości parametru θ.

- Ściślej powinno być: „efektywniejszy lub tak samo efektywny jak"
- Warunek „dla każdego θ" jest potrzebny, ponieważ wariancje estymatorów mogą zależeć od parametru rozkładu
- Relacja efektywności dotyczy tylko estymatorów nieobciążonych

Relacja efektywności

Rozważmy dwa estymatory nieobciążone $\widehat{\theta}_1$ i $\widehat{\theta}_2$ parametru θ. Mówimy, że $\widehat{\theta}_1$ jest efektywniejszy od estymatora $\widehat{\theta}_2$, jeśli

$$D^2(\widehat{\theta}_1) \leqslant D^2(\widehat{\theta}_2)$$

dla każdej wartości parametru θ.

- Ściślej powinno być: „efektywniejszy lub tak samo efektywny jak”
- Warunek „dla każdego θ” jest potrzebny, ponieważ wariancje estymatorów mogą zależeć od parametru rozkładu
- Relacja efektywności dotyczy tylko estymatorów nieobciążonych
- Relacja efektywności jest porządkiem częściowym, tzn. dwa estymatory mogą być nieporównywalne w sensie efektywności

Relacja efektywności

Rozważmy dwa estymatory nieobciążone $\widehat{\theta}_1$ i $\widehat{\theta}_2$ parametru θ. Mówimy, że $\widehat{\theta}_1$ jest efektywniejszy od estymatora $\widehat{\theta}_2$, jeśli

$$D^2(\widehat{\theta}_1) \leqslant D^2(\widehat{\theta}_2)$$

dla każdej wartości parametru θ.

- Ściślej powinno być: „efektywniejszy lub tak samo efektywny jak"
- Warunek „dla każdego θ" jest potrzebny, ponieważ wariancje estymatorów mogą zależeć od parametru rozkładu
- Relacja efektywności dotyczy tylko estymatorów nieobciążonych
- Relacja efektywności jest porządkiem częściowym, tzn. dwa estymatory mogą być nieporównywalne w sensie efektywności

Estymator nieobciążony efektywniejszy od wszystkich innych estymatorów nieobciążonych (jeśli istnieje) nazywamy estymatorem efektywnym

Przykład

Dla próby losowej $X_1, X_2, \ldots, X_n$ rozważmy rodzinę estymatorów wartości oczekiwanej μ postaci:

$$\widehat{\mu}_k = \frac{1}{k}\sum_{i=1}^{k} X_i, \qquad k = 1, \ldots, n$$

(czyli $\widehat{\mu}_k$ bierze średnią z k pierwszych elementów próby)

Przykład

Dla próby losowej $X_1, X_2, \ldots, X_n$ rozważmy rodzinę estymatorów wartości oczekiwanej μ postaci:

$$\widehat{\mu}_k = \frac{1}{k}\sum_{i=1}^{k} X_i, \qquad k = 1, \ldots, n$$

(czyli $\widehat{\mu}_k$ bierze średnią z k pierwszych elementów próby)

$$E(\widehat{\mu}_k) = \frac{1}{k}\sum_{i=1}^{k} \underbrace{EX_i}_{\mu} = \mu \qquad \text{(nieobciążone)}$$

Przykład

Dla próby losowej $X_1, X_2, \ldots, X_n$ rozważmy rodzinę estymatorów wartości oczekiwanej μ postaci:

$$\widehat{\mu}_k = \frac{1}{k}\sum_{i=1}^{k} X_i, \qquad k = 1, \ldots, n$$

(czyli $\widehat{\mu}_k$ bierze średnią z k pierwszych elementów próby)

$$E(\widehat{\mu}_k) = \frac{1}{k}\sum_{i=1}^{k} \underbrace{EX_i}_{\mu} = \mu \qquad \text{(nieobciążone)}$$

$$D^2(\widehat{\mu}_k) = \frac{1}{k^2}\sum_{i=1}^{k} \underbrace{D^2(X_i)}_{\sigma^2} = \frac{\sigma^2}{k}$$

Przykład

Dla próby losowej $X_1, X_2, \ldots, X_n$ rozważmy rodzinę estymatorów wartości oczekiwanej μ postaci:

$$\widehat{\mu}_k = \frac{1}{k}\sum_{i=1}^{k} X_i, \qquad k = 1, \ldots, n$$

(czyli $\widehat{\mu}_k$ bierze średnią z k pierwszych elementów próby)

$$E(\widehat{\mu}_k) = \frac{1}{k}\sum_{i=1}^{k} \underbrace{EX_i}_{\mu} = \mu \qquad (\text{nieobciążone})$$

$$D^2(\widehat{\mu}_k) = \frac{1}{k^2}\sum_{i=1}^{k} \underbrace{D^2(X_i)}_{\sigma^2} = \frac{\sigma^2}{k}$$

Estymator $\widehat{\mu}_n = \overline{X}_n$ jest najefektywniejszy, a $\widehat{\mu}_1 = X_1$ najmniej efektywny

Czy estymator może mieć dowolnie małą wariancję?

Czy estymator może mieć dowolnie małą wariancję?

Nierówność Craméra-Rao

Niech $\widehat{\theta}$ będzie nieobciążonym estymatorem parametru θ wyznaczonym z próby $X_1, X_2, \ldots, X_n$. Zdefiniujmy funkcję informacji Fishera $I(\theta)$ jako:

- Dla rozkładów dyskretnych:

$$I(\theta) = E\left(\left(\frac{\partial \ln p(X)}{\partial \theta}\right)^2\right), \qquad \text{gdzie } p(x) \text{ jest prawd. wartości } x$$

- Dla rozkładów ciągłych:

$$I(\theta) = E\left(\left(\frac{\partial \ln f(X)}{\partial \theta}\right)^2\right), \qquad \text{gdzie } f(x) \text{ jest gęstością}$$

Wtedy:

$$D^2(\widehat{\theta}) \geqslant \frac{1}{nI(\theta)}$$

Czy estymator może mieć dowolnie małą wariancję?

Nierówność Craméra-Rao

Niech $\widehat{\theta}$ będzie nieobciążonym estymatorem parametru θ wyznaczonym z próby $X_1, X_2, \ldots, X_n$. Zdefiniujmy funkcję informacji Fishera $I(\theta)$ jako:

- Dla rozkładów dyskretnych:

$$I(\theta) = E\left(\left(\frac{\partial \ln p(X)}{\partial \theta}\right)^2\right), \qquad \text{gdzie } p(x) \text{ jest prawd. wartości } x$$

- Dla rozkładów ciągłych:

$$I(\theta) = E\left(\left(\frac{\partial \ln f(X)}{\partial \theta}\right)^2\right), \qquad \text{gdzie } f(x) \text{ jest gęstością}$$

Wtedy:

$$D^2(\widehat{\theta}) \geqslant \frac{1}{nI(\theta)}$$

Zadanie 3

Udowodnij tę nierówność (a raczej zobacz jak wygląda dowód)

Przykład

Załóżmy że X ma rozkład normalny $N(\mu, \sigma^2)$ i chcemy estymować μ

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

Przykład

Załóżmy że X ma rozkład normalny $N(\mu, \sigma^2)$ i chcemy estymować μ

$$f(x) \;=\; \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

Mamy: $\ln f(x) \;=\; \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{(x-\mu)^2}{2\sigma^2}$

Przykład

Załóżmy że X ma rozkład normalny $N(\mu, \sigma^2)$ i chcemy estymować μ

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

Mamy:
$$\begin{aligned} \ln f(x) &= \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{(x-\mu)^2}{2\sigma^2} \\ \frac{\partial \ln f(x)}{\partial \mu} &= -\frac{\partial}{\partial \mu}\left(\frac{(x-\mu)^2}{2\sigma^2}\right) = \end{aligned}$$

Przykład

Załóżmy że X ma rozkład normalny $N(\mu, \sigma^2)$ i chcemy estymować μ

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

$$\text{Mamy:} \quad \begin{aligned} \ln f(x) &= \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{(x-\mu)^2}{2\sigma^2} \\ \frac{\partial \ln f(x)}{\partial \mu} &= -\frac{\partial}{\partial \mu}\left(\frac{(x-\mu)^2}{2\sigma^2}\right) = \frac{x-\mu}{\sigma^2} \end{aligned}$$

Przykład

Załóżmy że X ma rozkład normalny $N(\mu, \sigma^2)$ i chcemy estymować μ

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

Mamy:
$$\begin{aligned}
\ln f(x) &= \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{(x-\mu)^2}{2\sigma^2} \\
\frac{\partial \ln f(x)}{\partial \mu} &= -\frac{\partial}{\partial \mu}\left(\frac{(x-\mu)^2}{2\sigma^2}\right) = \frac{x-\mu}{\sigma^2} \\
I(\mu) &= E\left(\left(\frac{X-\mu}{\sigma^2}\right)^2\right) =
\end{aligned}$$

Przykład

Załóżmy że X ma rozkład normalny $N(\mu, \sigma^2)$ i chcemy estymować μ

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

Mamy:
$$\begin{aligned}
\ln f(x) &= \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{(x-\mu)^2}{2\sigma^2} \\
\frac{\partial \ln f(x)}{\partial \mu} &= -\frac{\partial}{\partial \mu}\left(\frac{(x-\mu)^2}{2\sigma^2}\right) = \frac{x-\mu}{\sigma^2} \\
I(\mu) &= E\left(\left(\frac{X-\mu}{\sigma^2}\right)^2\right) = \frac{1}{\sigma^4}\underbrace{E\left((X-\mu)^2\right)}_{\sigma^2} = \frac{1}{\sigma^2}
\end{aligned}$$

Przykład – c.d.

$$X \sim N(\mu, \sigma^2), \qquad I(\mu) \;=\; \frac{1}{\sigma^2}$$

Z nierówności Craméra-Rao wynika więc, że dla dowolnego nieobciążonego estymatora $\widehat{\mu}$ wartości oczekiwanej μ:

$$D^2(\widehat{\mu}) \;\geqslant\; \frac{1}{nI(\mu)} \;=\; \frac{1}{n1/\sigma^2} = \frac{\sigma^2}{n}.$$

Przykład – c.d.

$$X \sim N(\mu, \sigma^2), \qquad I(\mu) \ = \ \frac{1}{\sigma^2}$$

Z nierówności Craméra-Rao wynika więc, że dla dowolnego nieobciążonego estymatora $\widehat{\mu}$ wartości oczekiwanej μ:

$$D^2(\widehat{\mu}) \ \geqslant \ \frac{1}{nI(\mu)} \ = \ \frac{1}{n1/\sigma^2} = \frac{\sigma^2}{n}.$$

Ale wiemy, że estymator $\overline{X}_n$ jest nieobciążony i ma wariancję $\frac{\sigma^2}{n}$!

Przykład – c.d.

$$X \sim N(\mu, \sigma^2), \qquad I(\mu) = \frac{1}{\sigma^2}$$

Z nierówności Craméra-Rao wynika więc, że dla dowolnego nieobciążonego estymatora $\widehat{\mu}$ wartości oczekiwanej μ:

$$D^2(\widehat{\mu}) \geqslant \frac{1}{nI(\mu)} = \frac{1}{n1/\sigma^2} = \frac{\sigma^2}{n}.$$

Ale wiemy, że estymator $\overline{X}_n$ jest nieobciążony i ma wariancję $\frac{\sigma^2}{n}$!

Wniosek: Estymator $\widehat{\mu} = \overline{X}_n$ wartości oczekiwanej μ jest efektywny jeśli X ma rozkład normalny.

Zadanie

Zadanie 4

Pokaż, że funkcja informacji Fishera $I(p)$ dla rozkładu dwupunktowego $B(p)$ ma postać:

$$I(p) = \frac{1}{p(1-p)}$$

Następnie pokaż, że estymator $\overline{X}_n$ wartości oczekiwanej p jest efektywny dla tego rozkładu