Metody probabilistyczne i statystyka - wykład dziesiąty[1] Podstawy teorii estymacji: estymacja punktowa.

dr Jarosław Kotowicz

Instytut Informatyki Uniwersytet w Białymstoku

30 kwietnia 2020r.

Spis treści

Pojęcie estymacji. I

Definicja 1

Estymacja to zbiór metod pozwalających na wnioskowanie o wartości parametrów lub postaci rozkładu cechy populacji generalnej na podstawie rozkładu empirycznego uzyskanego z próby losowej z tej populacji tj. na podstawie rozkładu z próby.

Estymacje dzielimy na

1. parametryczną, która polega na szacowaniu wartości parametrów rozkładu populacji generalnej,
2. nieparametryczną, która polega na szacowaniu również postaci funkcyjnej rozkładu populacji generalnej (być może z jej parametrami).

Estymację parametryczną dzielimy na

Pojęcie estymacji. II

1. punktową, w której za ocenę wartości parametru (parametrów) przyjmujemy jedną konkretną wartość (jeden konkretny wektor liczbowy) otrzymaną na podstawie wyników próby, który na podstawie przyjętych kryteriów dokładności będzie stanowił najlepsze przybliżenie nieznanego parametru (parametrów) w populacji generalnej;
2. przedziałową, w której wyznaczamy pewien liczbowy przedział w którym z określonym prawdopodobieństwem zawiera się wartość szacowanego parametru.

Założenia. I

- Parametr rozkładu może być jednowymiarowy, wielowymiarowy lub nieskończenie wymiarowy.
- Parametr rozkładu cechy jest stały (nielosowy).
- Wyniki z próby mają charakter losowy.
- Będziemy przyjmować, że rozkład cechy w populacji generalnej jest zadany za pomocą dystrybuant $F(\cdot;\theta)^2$, gdzie $\theta \in \Theta$ jest parametrem od którego zależy dystrybuanta. Zakładamy, że Θ jest zbiorem niepustym.
- Możemy równie dobrze założyć, że cecha w populacji generalnej jest zadana za pomocą funkcji rozkładu prawdopodobieństwa $p(\cdot;\theta)$ dla rozkładu dyskretnego, albo funkcji gęstości $f(\cdot;\theta)$ dla rozkładu ciągłego.

Założenia. II

Rozważamy model statystyczny tj. obserwację X i rodzinę $\{P_\theta : \theta \in \Theta\}$ rozkładów prawdopodobieństwa określoną na zbiorze $\mathcal{X}$ będącym zbiorem możliwych wartości obserwacji. Chcemy estymować parametr θ w przypadku parametru jednowymiarowego lub liczbę $\gamma(\theta)$, gdzie $\gamma\colon \Theta \to \mathbb{R}$ jest znaną funkcją, nie znamy tylko parametru θ. Zauważmy, że bez straty ogólności zawsze możemy jednak myśleć, że estymujemy wielkość $\gamma(\theta)$ kładą przy estymacji θ funkcję $\gamma \equiv \mathrm{Id}$.

Uwaga 1

Dlaczego funkcję? Popatrzmy na następujące przykłady.

Założenia. III

Przykład 1

Rozważmy $\{Exp(\theta) : \theta \in \mathbb{R}_+\}$. *Wtedy moglibyśmy rozważyć funkcje*

1. $\gamma_1 : \mathbb{R}_+ \to \mathbb{R}_+$ *daną wzorem* $\gamma_1(\theta) = \theta$,
2. $\gamma_2 : \mathbb{R}_+ \to \mathbb{R}_+$ *daną wzorem* $\gamma_2(\theta) = \frac{1}{\theta}$,
3. $\gamma_3 : \mathbb{R}_+ \to \mathbb{R}_+$ *daną wzorem* $\gamma_3(\theta) = \frac{1}{\theta^2}$.

Uwaga 2

Zauważmy, że w przypadku pierwszej funkcji estymujemy sam parametr θ, *drugiej chodzi o estymację wartości oczekiwanej, w trzeciej wariancji rozkładu wykładniczego.*

Założenia. IV

Przykład 2

Rozważmy $\{\mathcal{N}(m,\sigma) : m \in \mathbb{R} \wedge \sigma \in \mathbb{R}_+\}$. *Wtedy moglibyśmy rozważyć funkcje*

1. $\gamma_1 : \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}$ *daną wzorem* $\gamma_1(m,\sigma) = m$,
2. $\gamma_2 : \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}_+$ *daną wzorem* $\gamma_2(m,\sigma) = \sigma^2$,
3. $\gamma_3 : \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}_+$ *daną wzorem* $\gamma_3(m,\sigma) = \sigma$.

Uwaga 3

W tym przypadku chodzi nam o estymację odpowiednio wartości oczekiwanej, wariancji i odchylenia standardowego rozkładu normalnego.

Założenia. V

Przykład 3

Niech $\mathcal{F}$ będzie rodziną wszystkich dystrybuant ciągłych rozkładów prawdopodobieństwa na $\mathbb{R}$. Rozważmy $\{P_F : F \in \mathcal{F}\}$. Wtedy moglibyśmy rozważyć funkcje

1. $\gamma\colon \mathcal{F} \to \mathbb{R}$ *daną wzorem* $\gamma(F) = 2$.

Uwaga 4

W naszych dalszych rozważaniach ograniczymy się wyłącznie do estymacji parametrycznej. Ponadto zakładamy, że $\Theta \subset \mathbb{R}^d$ dla pewnej liczby naturalnej d.

Założenia. VI

Uwaga 5

1. *Parametr θ lub liczbę $\gamma(\theta)$, gdzie $\gamma\colon \Theta \to \mathbb{R}$ jest ustalą i znaną funkcją, nazywany jest parametrem estymowanym.*
2. *W dalszym ciągu naszych rozważań będziemy pisać o szacowaniu $\gamma(\theta)$.*
3. *Będziemy szacować nieznaną wartość parametru na podstawie n elementowej próby $X_1, \ldots, X_n$.*
4. *W niektórych wypadkach liczebność próby będzie ustalona, a w niektórych konstruować będziemy estymator według jednego schematu dla wielu prób o różnych liczebnościach.*

[2] Oznaczenie · oznacza argument dystrybuanty F.

Pojęcie estymatora i oceny parametru. I

Definicja 2

Niech $X_1, \ldots, X_n \sim_{iid} F(\cdot;\theta)$.
Estymatorem T_n *liczby* $\gamma(\theta)$ *nazywamy statystykę z próby* $T_n = T(X_1, \ldots, X_n)$, *która służy do oszacowania wartości* $\gamma(\theta)$[a].
Wartość liczbową t_n *estymatora* T_n *liczby* $\gamma(\theta)$ *otrzymana z konkretnej realizacji próby* $(x_1, x_2, \ldots, x_n)$ *tzn.* $t_n = T(x_1, x_2, \ldots, x_n)$ *nazywa się oceną, oceną punktową lub estymatą zmiennej losowej* T_n .

[a] Estymator, to statystyka $T_n \colon (\mathcal{X}, \mathcal{H}) \to (\Theta, \mathcal{M})$. Zakładamy, że mamy określone σ-podciała zbiorów borelowskich $\mathcal{H} = \mathcal{B}(\mathcal{X})$ i $\mathcal{M} = \mathcal{B}(\Theta)$.

Pojęcie estymatora i oceny parametru. II

Uwaga 6

1. *Jeżeli liczebność próby nie będzie istotna, to będziemy opuszczać indeks dolny przy estymatorze.*
2. *Jeżeli będziemy estymować wielkość $\gamma(\theta)$, to estymator będziemy również oznaczać $\widehat{\gamma(\theta)}$.*
3. *Ocenę punktową wielkości $\gamma(\theta)$ dla realizacji $x = (x_1, x_2, \ldots, x_n)$ próby losowej $X_1, \ldots, X_n$ oznaczamy $\widehat{\gamma}(\theta, x)$.*

Pojęcie estymatora i oceny parametru. III

Uwaga 7

1. *Ocena parametru będzie prawie zawsze różnić się od oryginalnej wartości parametru $\gamma(\theta)$.*
2. *Szacując pewien parametr populacji za pomocą określonego estymatora na podstawie wyników próby losowej, otrzymuje się różne wartości ocen. Wynika z tego, że oceny parametrów obliczane z prób obarczone są błędami. Różnica między estymatorem wielkości $\gamma(\theta)$ a tą wartością nazywana jest błędem szacunku.*
3. *Aby uzyskać mały błąd (duża precyzja szacunku), należy zadbać o prawidłowe losowanie próby, a także wybór najlepszego[a] estymatora spośród wielu dostępnych statystyk z próby.*

[a] Pozostaje pytanie co oznacza estymator najlepszy.

Pojęcie estymatora i oceny parametru. IV

Definicja 3

Błędem szacunku oceny punktowej estymowanej wielkości $\gamma(\theta)$ *odpowiadającemu realizacji* $x = (x_1, x_2, \dots, x_n)$ *nazywamy liczbę*

$$e(\gamma(\theta), x) := \widehat{\gamma}(\theta, x) - \gamma(\theta).$$

Natomiast błędem szacunku oceny punktowej estymowanej wielkości $\gamma(\theta)$ *nazywamy zmienną losową*

$$e(\gamma(\theta)) := \widehat{\gamma(\theta)} - \gamma(\theta).$$

Uwaga 8

Zauważmy, że błąd szacunku $e(\gamma(\theta), x)$ *zależy nie tylko od estymatora (formuły wykorzystanej do szacowania estymowanej wielkości), ale także od próby.*

Pojęcie estymatora i oceny parametru. V

Uwaga 9

Można też wprowadzić inną miarę błędu

Definicja 4

Niech $\widehat{\gamma(\theta)}$ *będzie estymatorem wielkości* $\gamma(\theta)$ *na podstawie próby n elementowej, a* $x = (x_1, x_2, \ldots, x_n)$ *będzie realizacją tej próby. Odchyleniem próbkowym estymatora* $\widehat{\gamma(\theta)}$ *odpowiadającemu realizacji* $x = (x_1, x_2, \ldots, x_n)$ *nazywamy liczbę równą*

$$d(\gamma(\theta), x) := \widehat{\gamma}(\theta, x) - \mathbb{E}_\theta(\widehat{\gamma(\theta)}).$$

Pojęcie estymatora i oceny parametru. VI

Uwaga 10

Zauważmy, że odchylenie próbkowe $d(\gamma(\theta), x)$, podobnie jak błąd szacunku oceny punkowej $e(\gamma(\theta), x)$, zależy nie tylko od estymatora, ale także od próby.

1. Jeżeli chcemy szacować błąd estymatora niezależnie od realizacji należy wprowadzić inną miarę błędu szacunku.
2. Zwykle miarą błędu szacunku estymatora T wielkości $\gamma(\theta)$ jest błąd średniokwadratowy (ang. MSE)

$$\forall_{\theta\in\Theta} \quad MSE_\theta(T) := \mathbb{E}_\theta((T-\gamma(\theta))^2).$$

3. Estymator to zmienna losowa.

Pojęcie estymatora i oceny parametru. VII

4. Ponieważ każda ze zmiennych X_i dla dowolnego $i \in \overline{1, n}$ ma taki sam rozkład, jak cecha X w populacji generalnej, więc mają też taką samą dystrybuantę (funkcję prawdopodobieństwa lub funkcję gęstości).

Uwaga 11

Zauważmy, że błąd średniokwadratowy ta tak naprawdę funkcja

$$\Theta \ni \theta \mapsto MSE_\theta(T).$$

Estymatory nieobciążone i obciążone. I

Definicja 5

Niech $X_1, \ldots, X_n \sim_{iid} F(\cdot;\theta)$ *i* T *będzie estymatorem wielkości* $\gamma(\theta)$. *Obciążeniem estymatora* T *nazywamy liczbę równą*

$$\forall_{\theta\in\Theta}\ b(T) := \mathbb{E}_\theta(T) - \gamma(\theta).$$

Mówimy, że estymator T *jest nieobciążony wtedy i tylko wtedy, gdy* $b(T) = 0$. *W przeciwnym wypadku estymator nazywamy obciążonym.*

Estymatory nieobciążone i obciążone. II

Wniosek 1

Niech T będzie estymatorem wielkości $\gamma(\theta)$.

1. *Estymator T jest nieobciążony wtedy i tylko wtedy $\mathbb{E}_\theta(T) = \gamma(\theta)$ dla dowolnego $\theta \in \Theta$.*
2. $MSE_\theta(T) = Var_\theta(T) + (b(T))^2$ *dla dowolnego* $\theta \in \Theta$.
3. *Jeżeli T jest nieobciążony, to $MSE_\theta(T) = Var_\theta(T)$ dla dowolnego $\theta \in \Theta$.*

Estymatory nieobciążone i obciążone. III

Uwaga 12

1. *Jeżeli estymator jest nieobciążony, to otrzymane przy jego stosowaniu oceny parametrów nie są obciążone błędami systematycznymi.*
2. *Dla estymatora nieobciążonego odchylenie wartości estymatora od szacowanego parametru obliczone na podstawie jednej próby mogą być duże, ale w wyniku wielokrotnego losowanie próby odchylenia przy obliczeniu średniej arytmetycznej znoszą się.*

Estymatory nieobciążone i obciążone. IV

Definicja 6

Niech estymator T wielkości $\gamma(\theta)$ będzie nieobciążony.

1. *Średnim błędem szacunku estymatora T nazywamy odchylenie standardowe T (tzn. $\mathbb{D}_\theta(T)$).*
2. *Względnym średnim błędem szacunku estymatora T nazywamy liczbę równą*
$$\frac{\mathbb{D}_\theta(T)}{\gamma(\theta)}.$$
3. *Precyzją szacunku estymatora T nazywamy liczbę równą*
$$\frac{1}{Var_\theta(T)}.$$

Estymatory nieobciążone i obciążone. V

Przykład 4

Niech $X_1, \ldots, X_n \sim_{iid} F(\cdot;\theta)$, gdzie rozkłady dla każdej wartości parametru θ posiadają wartość oczekiwaną i szacujemy wielkość $\gamma(\theta) = m$ i m jest wartością oczekiwaną rozkładu o dystrybuancie $F(\cdot;\theta)$. Nieobciążonym estymatorem wartości oczekiwanej jest średnia z próby $\overline{X}$.

UZASADNIENIE. Mamy

$$\mathbb{E}_\theta(\overline{X}) = \mathbb{E}_\theta\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_\theta\left(X_i\right) = \frac{1}{n}\cdot m \cdot n = m.$$

□

Estymatory nieobciążone i obciążone. VI

Uwaga 13

W przykładach 5, 6, 9 parametr jest dwuwymiarowy (m, σ) *i dlatego nie używamy estymatora parametru, a używamy funkcji* γ *określonej dla dowolnych* m, σ *wzorem* $\gamma(m, \sigma) := \sigma^2$.

Przykład 5

Niech $X_1, \ldots, X_n \sim_{iid} \mathcal{N}(m, \sigma)$. *Dla prostoty przyjmijmy oznaczenie* $\theta = (m, \sigma)$.
Obciążonym estymatorem wariancji jest wariancja z próby

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

Estymatory nieobciążone i obciążone. VII

UZASADNIENIE. Mamy

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

oraz

$$\begin{aligned}\mathbb{E}_\theta(S^2) &= \left(\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_\theta(X_i^2)\right) - \mathbb{E}_\theta(\overline{X}^2)\\ &= \mathbb{E}_\theta(X^2) - \mathbb{E}_\theta(\overline{X}^2).\end{aligned}$$

Estymatory nieobciążone i obciążone. VIII

Ponadto

$$\begin{aligned}
\mathbb{E}_\theta(\overline{X}^2) &= \mathbb{E}_\theta\left(\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2\right) = \frac{1}{n^2}\mathbb{E}_\theta\left(\sum_{i=1}^{n} X_i^2 + \sum_{\substack{j,k\in\overline{1,n}\\ j\neq k}} X_k X_j\right) \\
&= \frac{1}{n^2}\left(\sum_{i=1}^{n}\mathbb{E}_\theta(X_i^2) + \sum_{\substack{j,k\in\overline{1,n}\\ j\neq k}}\mathbb{E}_\theta(X_k)\mathbb{E}_\theta(X_j)\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{E}_\theta(X_i^2) + \frac{1}{n^2}\cdot m^2\cdot n\cdot(n-1) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{E}_\theta(X_i^2) + \frac{n-1}{n}\cdot m^2 = \frac{1}{n}E(X^2) + \frac{n-1}{n}\cdot m^2.
\end{aligned}$$

Estymatory nieobciążone i obciążone. IX

Stąd

$$\begin{aligned}\mathbb{E}_\theta(S_n^2) &= \mathbb{E}_\theta(X^2) - \frac{1}{n}\mathbb{E}_\theta(X^2) - \frac{n-1}{n}\cdot m^2 \\ &= \frac{n-1}{n}\mathbb{E}_\theta(X^2) - \frac{n-1}{n}\cdot m^2 = \frac{n-1}{n}\sigma^2.\end{aligned}$$

Wynika stąd, że obciążenie estymatora S^2 wynosi $b(S^2) = -\frac{1}{n}\sigma^2$. □

Przykład 6

Niech $X_1, \ldots, X_n \sim_{iid} \mathcal{N}(m, \sigma)$. Podobnie, jak poprzednio oznaczmy $\theta = (m, \sigma)$.
Nieobciążonym estymatorem wariancji jest wariancji z próby postaci $\tilde{S}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$.

Estymatory nieobciążone i obciążone. X

UZASADNIENIE. Ponieważ $S^2 = \frac{n-1}{n}\tilde{S}^2$, więc

$$\begin{aligned}\mathbb{E}_\theta(\tilde{S}^2) &= \mathbb{E}_\theta\left(\frac{n}{n-1}S^2\right) = \frac{n}{n-1}\mathbb{E}_\theta(S^2)\\ &= \frac{n}{n-1}\frac{n-1}{n}\sigma^2 = \sigma^2.\end{aligned}$$

□

Estymator najefektywniejszy. I

Definicja 7

Dany jest zbiór wszystkich nieobciążonych estymatorów $T^1, \ldots, T^r$ *wielkości* $\gamma(\theta)$ *dla próby o ustalonej liczebności n.*
Estymator $T^\star$ *z tego zbioru o własności* $Var_\theta(T^\star) \leqslant Var_\theta(T^i)$ *dla* $i \in \overline{1, r}$ *i* $\theta \in \Theta$ *nazywamy najefektywniejszym estymatorem liczby* $\gamma(\theta)$*. Natomiast wielkość*

$$e(T^i) = \frac{Var_\theta(T^\star)}{Var_\theta(T^i)} \quad dla\ i \in \overline{1, r} \tag{1}$$

efektywnością estymatora T^i *wielkości* $\gamma(\theta)$*.*

Estymator najefektywniejszy. II

Uwaga 14 (Problem wyznaczania estymatora najefektywniejszego)

1. *Efektywność najefektywniejszego estymatora jest równa jeden, w pozostałych zaś przypadkach należy do przedziału* $]0,1[$.
2. *Aby określić efektywność estymatora, należy znać wszystkie nieobciążone estymatory danego parametru i ich wariancję lub znać wartość wariancji estymatora najefektywniejszego.*
3. *Przy wyznaczaniu wariancji estymatora najefektywniejszego wykorzystuje się twierdzenie zwane nierównością Rao-Cramera.*
4. *Warunkiem koniecznym stosowalności nierówność Rao-Cramera jest znajomość postaci funkcyjnej funkcji prawdopodobieństwa* $p(\cdot;\theta)$ *bądź funkcji gęstości* $f(\cdot;\theta)$.

Nierówność Rao-Cramera. I

Uwaga 15

W przypadku nierówności Rao-Cramera sformułujemy ją tylko przy następującym założeniu: rozważać będziemy jednowymiarowy parametr θ *tzn.* $\Theta \subset \mathbb{R}$.
Estymować będziemy wielkość $\gamma(\theta)$.

Nierówność Rao-Cramera. II

Definicja 8 (Warunki regularności Rao-Cramera)

Niech dana będzie rodzina gęstości prawdopodobieństwa (funkcji prawdopodobieństwa) zależnych od parametru $\theta \in \Theta \subset \mathbb{R}$ *tzn* $\{f(\cdot;\theta) : \theta \in \Theta\}$.

Mówimy, że rodzina $\{f(\cdot;\theta) : \theta \in \Theta\}$ *spełnia warunek regularności wtedy i tylko wtedy, gdy*

1. *funkcje gęstości (odpowiednio funkcje prawdopodobieństwa) dla dowolnego* x *spełniają warunki*
   1. *posiada ciągłe pochodne cząstkowe rzędu drugiego względem* θ,
   2. *zbiór* $\{x : f(x;\theta) > 0\}$ *jest niezależny od* θ,
   3. $\frac{\partial}{\partial\theta} \ln f(x;\theta)$ *istnieje oraz* $\mathbb{E}_\theta\left(\left(\frac{\partial}{\partial\theta} \ln f(X;\theta)\right)^2\right) \in]0,+\infty[$,
2. *operacje całkowania względem* x *i różniczkowania ze względu na* θ *są przemienne tzn.* $\frac{\partial}{\partial\theta}\left[\int \ldots \cdot f(x;\theta)\, dx\right] = \int \ldots \cdot \left[\frac{\partial}{\partial\theta} f(x;\theta)\right] dx$ *wszędzie gdzie prawa strona jest skończona.*

Nierówność Rao-Cramera. III

Twierdzenie 1

Niech $X_1, \ldots, X_n \sim_{iid} F(\cdot; \theta)$ *i* T *będzie nieobciążonym estymatorem wielkości* $\gamma(\theta)$. *Oznaczmy* $X = (X_1, \ldots, X_n)$.
Jeżeli funkcje gęstości spełniają warunki regularności, to

$$Var_\theta(T) \geqslant \frac{\left(\frac{\partial}{\partial\theta}\gamma(\theta)\right)^2}{\mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\ln \mathbf{f}(X;\theta)\right)^2\right]} \tag{2}$$

lub

$$Var_\theta(T) \geqslant \frac{\left(\frac{\partial}{\partial\theta}\gamma(\theta)\right)^2}{\mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\ln \mathbf{p}(X;\theta)\right)^2\right]}. \tag{3}$$

Nierówność Rao-Cramera. IV

Uwaga 16

W twierdzeniu 1 mamy: **f** *(odpowiednio* **p***) we wzorze* (2) *(odpowiednio we wzorze* (3)*) oznacza funkcję gęstości (odpowiednio funkcję prawdopodobieństwa) n-wymiarowej zmiennej losowej* X *będącej iloczynem gęstości (funkcji prawdopodobieństwa) zmiennych* $X_1, \ldots, X_n$*.*

Nierówność Rao-Cramera. V

Uwaga 17

Ze względu na niezależność zmiennych losowych $X_1, \ldots, X_n$ mamy

$$\mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\ln \mathbf{f}(X;\theta)\right)^2\right] = n\mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\ln f(X_1;\theta)\right)^2\right]$$

$$\mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\ln \mathbf{p}(X;\theta)\right)^2\right] = n\mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\ln p(X_1;\theta)\right)^2\right].$$

co więcej

$$\mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\ln f(X_1;\theta)\right)^2\right] = \int_{\mathbb{R}}\left(\frac{\partial}{\partial\theta}\ln f(x;\theta)\right)^2 f(x;\theta)dx$$

$$\mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\ln p(X;\theta)\right)^2\right] = \sum_x\left(\frac{\partial}{\partial\theta}\ln p(x;\theta)\right)^2 p(x;\theta)$$

Nierówność Rao-Cramera – przykłady. I

Przykład 7

Niech $X_1, \ldots, X_n \sim_{iid} Exp(\theta)$. *Estymujemy wielkość* $\gamma(\theta) = \frac{1}{\theta}$, *czyli wartość oczekiwaną. Wtedy* $\overline{X}$ *jest estymatorem najefektywniejszym, gdyż*

$$
\begin{aligned}
Var_\theta(X_1) &= \frac{1}{\theta^2} \\
\left(\frac{\partial}{\partial\theta}\gamma(\theta)\right)^2 &= \frac{1}{\theta^4} \\
\mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\ln \mathbf{f}(X;\theta)\right)^2\right] &= nVar(X_1) \\
&= \frac{n}{\theta^2}.
\end{aligned}
$$

Nierówność Rao-Cramera – przykłady. II

Uwaga 18

Z nierówności Rao-Cramera nie wynika, że wariancja estymatora najefektywniejszego, o ile oczywiście estymator najefektywniejszy istnieje, wynosi

$$Var_\theta(T_n^\star) = \frac{\left(\frac{\partial}{\partial\theta}\gamma(\theta)\right)^2}{n\mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\ln f(x;\theta)\right)^2\right]}. \tag{4}$$

Nierówność Rao-Cramera – przykłady. III

Przykład 8

Niech $X_1, \ldots, X_n \sim_{iid} Exp(\theta)$. *Estymujemy wielkość* $\gamma(\theta) = \theta$. *Estymatorem najefektywniejszym parametru* θ *jest* $\frac{n-1}{n\overline{X}}$ *i*

$$Var_\theta\left(\frac{n-1}{n\overline{X}}\right) > \frac{1}{\mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\ln \mathbf{f}(X;\theta)\right)^2\right]}.$$

UZASADNIENIE: Estymator $n\overline{X}$ jest estymatorem obciążanym, gdyż

$$\begin{aligned} n\overline{X} &\sim Gamma(n,\theta) \\ \mathbb{E}_\theta\left(\frac{1}{n\overline{X}}\right) &= \frac{\theta}{n-1}. \end{aligned}$$

Nierówność Rao-Cramera – przykłady. IV

Estymatorem nieobciążonym jest więc statystyka

$$\mathbb{E}_\theta\left(\frac{n-1}{n\overline{X}}\right).$$

Ponieważ

$$\mathbb{E}_\theta\left(\frac{1}{(n\overline{X})^2}\right)=\frac{\theta^2}{(n-1)(n-2)},$$

więc

$$\begin{aligned}
\mathbb{E}_\theta\left(\frac{(n-1)^2}{(n\overline{X})^2}\right) &= \frac{(n-1)\theta^2}{n-2}\\
Var_\theta\left(\frac{n-1}{n\overline{X}}\right) &= \frac{\theta^2}{n-2}\\
&> \frac{\theta^2}{n}=\frac{1}{\mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\ln\mathbf{f}(X;\theta)\right)^2\right]}.
\end{aligned}$$

Informacja Fishera.

Uwaga 19

W nierówności Rao-Cramera pojawia się pewne pojęci.

Definicja 9

Niech X będzie zmienną losową o gęstości $f(\cdot;\theta)$, zależnej od jednowymiarowego parametru $\theta \in \Theta \subset \mathbb{R}$. Funkcję

$$I(\theta) := \mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\ln f(X;\theta)\right)^2\right]$$

nazywamy informacją Fishera zawartą w zmiennej (obserwacji) X.

Estymatory asymptotycznie nieobciążone. I

Uwaga 20

W przypadku dwóch kolejnych definicji konstruujemy ciąg estymatorów zgodnie z jedną ustaloną regułą, ale dla prób coraz liczniejszych.

Definicja 10

Niech n będzie dowolną liczbą naturalną, $X_1, \ldots, X_n \sim_{iid} F(\cdot; \theta)$ i T_n będzie estymatorem liczby $\gamma(\theta)$.
Mówimy, że estymator T_n jest asymptotycznie nieobciążony wtedy i tylko wtedy, gdy

$$\lim_{n \to \infty} b(T_n) = 0.$$

Estymatory asymptotycznie nieobciążone. II

Przykład 9

Niech $X_1, \ldots, X_n \sim_{iid} \mathcal{N}(m, \sigma)$.
Estymator S_n^2 wariancji jest asymptotycznie nieobciążony.

UZASADNIENIE. Wynika to z faktu

$$\lim_{n\to\infty} b(S_n^2) = \lim_{n\to\infty} \left(-\frac{\sigma^2}{n}\right) = 0.$$

□

Estymatory zgodne. I

Definicja 11

Niech $X_1, \ldots, X_n \sim_{iid} F(\cdot;\theta)$ *i* T_n *będzie estymatorem* $\gamma(\theta)$*.*
Mówimy, że estymator T_n *liczby* $\gamma(\theta)$ *jest zgodny wtedy i tylko wtedy, gdy jest zbieżny według prawdopodobieństwa do* $\gamma(\theta)$*.*

Uwaga 21

Przypomnijmy warunek zbieżności według prawdopodobieństwa ciągu (X_n) *zmiennych losowych do zmiennej* X *oznacza*

$$\forall_{\varepsilon>0} \lim_{n\to\infty} P(\{|X_n - X| < \varepsilon\}) = 1.$$

Estymatory zgodne. II

Uwaga 22

Jeżeli estymator jest zgodny, to w miarę wzrostu liczebności próby jest coraz większe prawdopodobieństwo tego, że estymator będzie przyjmować wartości coraz bliższe wartości parametru w populacji. Zmniejsza się zatem ryzyko popełnienia dużego błędu.

Estymatory zgodne, a nieobciążone (asymptotycznie nieobciążone). I

Twierdzenie 2

Jeżeli estymator wielkości $\gamma(\theta)$ jest zgodny, to jest asymptotycznie nieobciążony.

Twierdzenie 3

Jeżeli estymator T_n liczby $\gamma(\theta)$ jest nieobciążony lub asymptotycznie nieobciążony oraz $\lim_{n\to\infty} Var_\theta(T_n) = 0$ dla dowolnego θ, to T_n jest zgodny.

Dowód. Gdy jest estymator T_n jest nieobciążony wystarczy zastosować nierówność Czebyszewa - Bienaymé. Natomiast dla asymptotycznie nieobciążonego trzeba zastosować nierówność Markowa ze zmienną $T_n - \gamma(\theta)$ i $p = 2$. □

Estymatory zgodne, a nieobciążone (asymptotycznie nieobciążone). II

Przykład 10

Niech $X_1, \ldots, X_n \sim_{iid} \mathcal{N}(m, \sigma)$. *I znowu niech* $\theta = (m, \sigma)$.
Estymatorem zgodnym wartości oczekiwanej jest średnia z próby $\overline{X}_n$.

UZASADNIENIE. Mamy

$$Var_\theta(\overline{X}_n) = \frac{1}{n^2} Var_\theta \left(\sum_{i=1}^{n} X_i \right) = \frac{1}{n^2} \sum_{i=1}^{n} Var_{(m,\sigma^2)}(X_i) = \frac{\sigma^2}{n}.$$

Co uzasadnia zgodność estymatora. □

Zadanie 1 (Praca domowa)

Niech $X_1, \ldots, X_n \sim_{iid} \mathcal{N}(m, \sigma)$. *Udowodnić, że estymator* S_n^2 *jest estymatorem zgodnym wariancji* σ^2.

Estymatory zgodne, a nieobciążone (asymptotycznie nieobciążone). III

WSKAZÓWKA. Należy skorzystać ze statystyki $\frac{nS_n^2}{\sigma^2}$ i jej własności (jaki rozkład ma ta statystyka – twierdzenie Fishera).

Estymatory asymptotycznie najefektywniejsze. I

Definicja 12

Niech $X_1, \ldots, X_n \sim_{iid} F(\cdot;\theta)$ i T_n będzie nieobciążonym lub asymptotycznie nieobciążonym estymatorem wielkości $\gamma(\theta)$.
Mówimy, że estymator T_n jest asymptotycznie najefektowniejszy wtedy i tylko wtedy, gdy zachodzi równość

$$\lim_{n\to\infty} e(T_n) = 1.$$

Uwaga 23

Dla estymatora asymptotycznie najefektywniejszego w miarę wzrostu liczebności próby do nieskończoności wariancja $Var_\theta(T_n)$ estymatora T_n przyjmuje wartości coraz bliższe wariancji $T_n^\star$ estymatora najefektywniejszego.

Estymatory asymptotycznie najefektywniejsze. II

Przykład 11

Niech $X_1, \ldots, X_n \sim_{iid} \mathcal{N}(m, \sigma)$. Udowodnimy, że średnia arytmetyczna z próby jest estymatorem najefektywniejszym wartości oczekiwanej.

UZASADNIENIE. Mamy

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

Wtedy

$$\ln f(x; m, \sigma) = -\ln(\sqrt{2\pi}\sigma) - \frac{(x-m)^2}{2\sigma^2}.$$

Stąd

$$\frac{\partial}{\partial m} \ln f(x; m, \sigma) = \frac{x-m}{\sigma^2},$$

Estymatory asymptotycznie najefektywniejsze. III

a więc

$$\mathbb{E}_{(m,\sigma)}\left[\left(\frac{\partial}{\partial m}\ln f(x;m,\sigma)\right)^2\right]=\frac{1}{\sigma^2}.$$

Ostatecznie

$$Var_{(m,\sigma)}(T_n^\star)=\frac{1}{n\mathbb{E}_{(m,\sigma)}\left[\left(\frac{\partial}{\partial m}\ln f(x;m,\sigma)\right)^2\right]}=\frac{\sigma^2}{n}.$$

Ponieważ $Var_{(m,\sigma)}(\overline{X}_n)=\frac{\sigma^2}{n}$, więc dowód jest zakończony. □

Spis treści

Metody estymacji punktowej

Uwaga 24

Estymacja punktowa polega na wyborze estymatora T danego parametru θ populacji lub funkcji tego parametru i wyznaczeniu jego wartości (oceny parametru) na podstawie próby i podaniu odchylenia standardowego oceny parametru $\mathbb{D}_\theta(T)$ ($\theta = T \pm \mathbb{D}_\theta(T)$).

Istnieją następujące metody estymacji punktowej

1. metodę momentów (**MM**),
2. metodę największej wiarogodności (**MNW**),
3. metodę najmniejszych kwadratów (**MNK**).

Uwaga 25

Wiarogodności jest też w literaturze nazywana wiarygodnością

Przypomnienie – momenty zwykłe i centralne

Niech będzie ustalona przestrzeń probabilistyczna (Ω, Σ, P) oraz zmienna losowa X całkowalna z k-tą potęgą tzn. taka, że $\int\limits_{\Omega} |X|^k dP < +\infty$, gdzie k jest dowolną, lecz ustaloną liczba naturalną.

Definicja 13

Momentem zwykłym rzędu k nazywamy wartość oczekiwaną zmiennej losowej X^k, a momentem centralnym rzędu k nazywamy wartość oczekiwaną zmiennej losowej $(X - \mathbb{E}(X))^k$. Pierwszy z nich oznaczamy m_k, zaś drugi μ_k.

Metoda momentów. I

Momenty zwykłe i centralne można przedstawić jako pewne funkcje parametrów rozpatrywanego układu. Rozważmy liczbę takich równań równą liczbie nieznanych parametrów. Otrzymujemy wtedy układ równań

$$\begin{cases} \eta_1 = g_1(\theta_1, \ldots, \theta_r) \\ \eta_2 = g_2(\theta_1, \ldots, \theta_r) \\ \qquad \ldots \\ \eta_r = g_r(\theta_1, \ldots, \theta_r) \end{cases}, \tag{5}$$

gdzie η_i dla $i = 1, \ldots, r$ są momentami zwykłymi lub centralnymi tak dobranymi, aby układ ten miał jednoznaczne rozwiązanie względem parametrów $\theta_1, \ldots, \theta_r$. Ponadto za η_i podstawiamy momenty z próby.

Metoda momentów. II

Przykład 12

Niech populacja generalna ma rozkład wykładniczy $f(x;\theta) = \theta e^{-\theta x}\mathbb{I}_{]0,+\infty[}(x)$ dla $\theta > 0$. Wyznaczymy metodą momentów estymator parametru θ.

UZASADNIENIE. Rozważmy próbę losową $X_1, \ldots, X_n \sim_{iid} Exp(\theta)$. Ponieważ mamy tylko jeden parametr, więc bierzemy dokładnie jedno równanie. Ponieważ $m_1 = \mathbb{E}_\theta(X) = \frac{1}{\theta}$. Wtedy $m_1^P = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$, a stąd

$$\hat{\theta} = \frac{1}{\overline{X}}.$$

$\square$

Metoda momentów. III

Uwaga 26

W kolejnym przykładzie poszukujemy nie parametrów, ale funkcji od tych parametrów.

Przykład 13

Niech cecha w populacji generalnej ma rozkład normalny $\mathcal{N}(m, \sigma)$. Wyznaczymy metodą momentów estymatory parametrów m i σ^2.

UZASADNIENIE. Rozważmy próbę losową $X_1, \ldots, X_n \sim_{iid} \mathcal{N}(m, \sigma)$. Zauważmy, że $m_1 = \mathbb{E}_{(m,\sigma)}(X)$ oraz $m_2 = \mu_2 + m_1^2$, gdzie μ_2 jest momentem centralnym rzędu 2. Układamy układ równań.

$$\begin{cases} m_1^P = \overline{X} \\ m_2^P = \frac{1}{n} \sum\limits_{i=1}^{n} X_i^2 \end{cases}$$

Metoda momentów. IV

Wtedy

$$\begin{aligned} \widehat{m} &= m_1^P = \overline{X} \\ \widehat{\sigma^2} &= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - (\overline{X})^2 \equiv S^2. \end{aligned}$$

$\square$

Uwaga 27

Wadą tej metody jest niemożliwość określenia własności estymatorów.

Metoda największej wiarogodności. I

Niech $X_1, \dots, X_n \sim_{iid} F(\cdot; \theta_1, \dots, \theta_r)$. Wtedy oczywiście funkcja prawdopodobieństwa lub funkcja gęstości zależy od nieznanych parametrów rozkładu $\theta_1, \dots, \theta_r$. Mamy $p(x_i; \theta_1, \dots, \theta_r)$ i $f(x_i; \theta_1, \dots, \theta_r)$ dla $i \in \overline{1, n}$.

Przyjmiemy konwencję $\mathbf{x} = (x_1, \dots, x_n)$.

Metoda największej wiarogodności. II

Definicja 14

Funkcją wiarogodności obserwacji $(X_1, \ldots, X_n)$ *zmiennej dyskretnej nazywamy funkcję postaci*

$$L(\theta_1, \ldots, \theta_r; \mathbf{x}) = \prod_{i=1}^{n} p(x_i; \theta_1, \ldots, \theta_r). \tag{6}$$

Funkcją wiarogodności obserwacji $(X_1, \ldots, X_n)$ *zmiennej ciągłej nazywamy funkcję postaci*

$$L(\theta_1, \ldots, \theta_r; \mathbf{x}) = \prod_{i=1}^{n} f(x_i; \theta_1, \ldots, \theta_r). \tag{7}$$

Etapy konstruowania estymatorów:

Metoda największej wiarogodności. III

1. Określenie funkcji wiarogodności.
2. Wyznaczenie logarytmu funkcji wiarogodności.
3. Obliczanie pochodnych cząstkowych $\frac{\partial}{\partial \theta_i} \ln L(\theta_1, \ldots, \theta_r; \mathbf{x})$ dla $i = 1, \ldots, n$.
4. Rozwiązanie układu równań

$$\begin{cases} \frac{\partial}{\partial \theta_1} \ln L(\theta_1, \ldots, \theta_r; \mathbf{x}) = 0 \\ \frac{\partial}{\partial \theta_2} \ln L(\theta_1, \ldots, \theta_r; \mathbf{x}) = 0 \\ \ldots \\ \frac{\partial}{\partial \theta_n} \ln L(\theta_1, \ldots, \theta_r; \mathbf{x}) = 0 \end{cases}.$$

5. Poszukiwane estymatory, to rozwiązania ostatniego układu.

Metoda największej wiarogodności. IV

Uwaga 28

Logarytmujemy funkcję wiarogodności, gdyż ich pochodne mają te same miejsca zerowe jak sama funkcja wiarogodności.

Przykład 14

Niech $X_1, \ldots, X_n \sim_{iid} Bin(m, p)$. Stosując metodę największej wiarogodności wyznaczymy estymator parametru p.

Uzasadnienie. Mamy

$$L(p; \mathbf{x}) = \prod_{i=1}^{n} \binom{m}{x_i} p^{x_i} (1-p)^{m-x_i}.$$

Metoda największej wiarogodności. V

Liczymy logarytm funkcji wiarogodności

$$\ln L(p;\mathbf{x}) = \sum_{i=1}^{n} \ln \binom{m}{x_i} + \sum_{i=1}^{n} x_i \ln p + \sum_{i=1}^{n} (m - x_i) \ln(1-p).$$

Następnie liczymy pochodną cząstkową względem p

$$\frac{\partial}{\partial p} \ln L(p;\mathbf{x}) = \frac{1}{p} \sum_{i=1}^{n} x_i - \frac{1}{1-p} \sum_{i=1}^{n} (m - x_i).$$

Rozwiązujemy równanie

$$\frac{\partial}{\partial p} \ln L(p;\mathbf{x}) = 0.$$

Otrzymujemy ostatecznie

$$\hat{p} = \frac{\overline{X}}{m}. \tag{8}$$

Metoda największej wiarogodności. VI

$\square$

Przykład 15

Niech $X_1, \ldots, X_n \sim_{iid} \mathcal{N}(m, \sigma)$. Metodą największej wiarogodności wyznaczymy estymatory parametrów m i σ^2. Zauważmy, że mamy tutaj funkcję od parametru σ tj. $\gamma(\sigma) = \sigma^2$.

Uzasadnienie. Mamy

$$L(m, \sigma^2; \mathbf{x}) = \frac{1}{(\sigma\sqrt{2\pi})^n} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - m)^2\right).$$

Liczymy logarytm funkcji wiarogodności

$$\ln L(m, \sigma^2; \mathbf{x}) = -n \ln \sqrt{\sigma^2} - n \ln \sqrt{2\pi} - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - m)^2.$$

Metoda największej wiarogodności. VII

Liczymy pochodne cząstkowe względem wartości oczekiwanej i wariancji

$$\begin{cases} \frac{\partial}{\partial m} \ln L(m, \sigma^2; \mathbf{x}) = \frac{1}{\sigma^2} \sum_{i=1}^{n} x_i - \frac{n \cdot m}{\sigma^2} \\ \frac{\partial}{\partial \sigma^2} \ln L(m, \sigma^2; \mathbf{x}) = \frac{-n}{2\sigma^2} + \frac{\sum_{i=1}^{n} (x_i - m)^2}{2(\sigma^2)^2}. \end{cases}$$

Rozwiązując układ równań otrzymujemy

$$\begin{aligned} \widehat{m} &= \frac{1}{n} \sum_{i=1}^{n} X_i = \overline{X}_n \\ \widehat{\sigma}^2 &= \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2. \end{aligned}$$

□

Metoda największej wiarogodności. VIII

Uwaga 29 (Własności estymatorów otrzymanych metodą największej wiarogodności)

1. *Są zgodne.*
2. *Mają asymptotyczny rozkład normalny o wartości oczekiwanej θ i wariancji $\frac{1}{n\mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\ln f(x;\theta)\right)^2\right]}$, gdzie n jest liczebnością próby.*
3. *Są co najmniej asymptotycznie nieobciążone.*
4. *Jeżeli istnieje estymator najefektywniejszy, to jest otrzymywany tą metodą.*
5. *Jeżeli $\hat{\theta}$ jest estymatorem otrzymanym tą metodą, to stosując metodę największej wiarogodności dla parametru $g(\theta)$ otrzymamy estymator $g(\hat{\theta})$.*

Estymatory dostateczne. I

Uwaga 30

Wykorzystując funkcję wiarogodności możemy wprowadzić pojęcie estymatora dostatecznego. Pojęcie to wprowadził A. Fisher.

Definicja 15

Niech $X_1, \ldots, X_n \sim_{iid} F(\cdot;\theta)$ *i niech* T *będzie estymatorem parametru* θ *na podstawie tej próby.*
Powiemy, że estymator T *jest estymatorem dostatecznym (wystarczającym) parametru* θ *wtedy i tylko wtedy, gdy skupia wszystkie informacje o parametrze* θ *w próbie prostej tzn. dla danej próby i danego parametru nie ma innego estymatora, który zawierałby więcej informacji o szacowanym parametrze.*

Estymatory dostateczne. II

Twierdzenie 4

Niech T będzie estymatorem parametru θ.
Jeżeli iloraz funkcji wiarogodności i funkcji gęstości (lub funkcji rozkładu prawdopodobieństwa) badanego estymatora jest funkcją niezależną od parametru θ, to estymator T jest dostateczny.

Uwaga 31

1. *Jeżeli estymator jest najefektywnieszy, to jest dostateczny.*
2. *Jeżeli istnieje estymator dostateczny parametru θ, to najlepszy w sensie* ***MNW*** *estymator tego parametru jest równy estymatorowi dostatecznemu lub jest pewną funkcją estymatora dostatecznego.*

Metoda najmniejszych kwadratów

Szacując wartość średnią na podstawie próby możemy zapisać

$$X_i = m + \varepsilon_i, \quad dla \quad i \in \overline{1, n}.$$

Jako estymator średniej bierzemy taką wartość $\widehat{m}$, dla której wyrażenie

$$\sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (X_i - m)^2 \tag{9}$$

jest najmniejsze. Wtedy obliczając pochodną względem m otrzymujemy

$$\widehat{m} = \frac{1}{n} \sum_{i=1}^{n} X_i \equiv \overline{X}.$$