

2 Podstawowe własności estymatorów.

Przedmiot teorii estymacji.

Teoria estymacji jest działem statystyki matematycznej zajmującej się oceną nieznanych parametrów rozkładu interesujących nas cech w populacji generalnej na podstawie wartości tych parametrów.

Wnioskowanie statystyczne opiera się na zbieraniu i przetwarzaniu informacji. Zbieranie informacji jest niczym innym, jak przygotowaniem wyników badań wszelkiego rodzaju do określonej analizy. Wyniki prób losowych, czy też obserwacje z szeregów czasowych stanowią materiał, dzięki któremu analityk jest w stanie wnioskować o prawidłowościach zachodzących w czasie, czy przestrzeni określonej populacji generalnej.

Przykładowo badamy rozkład wzrostu ludności w Polsce – cecha X . Zakładamy, że rozkład tej cechy X w populacji jest rozkładem normalnym, zaś szukaną wielkością jest wartość oczekiwana m . Wartość m jest zatem szukanym parametrem rozkładu cechy X . W celu oszacowania tych wielkości zbieramy dane z próby losowej o liczebności n . Następnym krokiem będzie znalezienie wygodnej statystyki (funkcji) T_n z próby, która posłuży do oszacowania parametru m . Rolę takiej statystyki może spełniać np. wartość średnia z próby. Mówimy zatem, że wartość średnia z próby jest **estymatorem** wartości oczekiwanej rozkładu normalnego. Obliczoną przez nas na podstawie konkretnej próby wartość średnią nazywamy **oceną parametru - estymatorem**.

Definicje

Załóżmy, że badamy rozkład cechy X w populacji, rozkład ten jest zależny od parametru θ . Wartość parametru zostanie oszacowana na podstawie n - elementowej próby losowej.

Estymator T_n parametru θ to dowolna statystyka z próby $T_n = t(X_1, X_2, \dots, X_n)$, pozwalająca wyznaczyć wartości parametru θ .

Ponieważ każda ze zmiennych losowych X_i ma rozkład identyczny z rozkładem cechy X w populacji generalnej, a rozkład ten zależy od parametru θ , zatem T_n jest zmienną losową, mającą rozkład również zależny od parametru θ .

Oceną parametru nazwiemy każdą realizację t_n zmiennej losowej T_n . Oczywiście ocena parametru będzie prawie zawsze różnić się od oryginalnej wartości parametru θ .

Własności estymatorów.

Definicja estymatora pozostawia dużą dowolność w wybraniu danej statystyki do szacowania parametru, nie pozwalając jednocześnie na ocenę która ze statystyk jest "dobrym" estymatorem. Aby sprawdzić, czy dana statystyka jest dobrym kandydatem na estymator parametru, powinniśmy sprawdzić, czy spełnia ona zestaw własności charakteryzujących estymator.

Estymatorem parametru θ nazywamy każdą funkcję (statystykę) $\bar{\theta}_n = U(X_1, X_2, \dots, X_n)$, która ma tę własność, że prawdopodobieństwo zdarzenia $\hat{\theta}_n$ zbliża się do θ i jest tym bliższe jedności, im większa jest liczebność próbki. Wynika z tego, że estymatorem parametru θ w populacji może być każda taka funkcja $\hat{\theta}_n$ z wartości wylosowanych do próbki, że dla arbitralnie obranej, lecz niekoniecznie dowolnej małej dodatniej liczby c zachodzi relacja:

$$\lim_{n \rightarrow \infty} P\{|\hat{\theta}_n - \theta| < c\} = 1,$$

Jest rzeczą oczywistą, że estymator tym lepiej spełnia swą rolę, im dla mniejszych wartości c może on czynić zadość powyższej równości.

Jeżeli znamy formę rozkładu prawdopodobieństwa populacji, a poszukujemy tylko jego parametrów – mówimy o **ocenie parametrycznej**.

Gdy nie znamy rozkładu prawdopodobieństwa populacji, a szukamy parametrów – mówimy o **ocenie nieparametrycznej**.

Ocena parametryczna może być:

Estymacją punktową - znajdowanie takiej wartości parametru (liczby), która może być uznana za najlepszą ocenę nieznanego parametru.

Estymacją przedziałową - znajdowanie takiego losowego przedziału, do którego z zadaniem prawdopodobieństwem należy wartość nieznanego parametru.

Najczęściej stosowanymi estymatorami w badaniach statystycznych, ze względu na jedną cechę, są estymatory wartości średniej i estymatory wariancji.

Łatwo zauważyć, że istnieje wiele estymatorów danego parametru. Zachodzi zatem potrzeba ich dokładniejszej charakterystyki. Oceniając jakość estymatora bierzemy pod uwagę następujące ich cechy:

- **estymatory zgodne,**
- **estymatory nieobciążone,**
- **estymatory najefektywniejsze,**
- **estymatory asymptotycznie najefektywniejsze.**

1. Zgodność estymatora.

Mówimy, że estymator $\hat{\theta}_n$ wielkości θ jest estymatorem zgodnym tej wielkości, jeśli jest dla niego spełnione słabo prawo wielkich liczb, tzn., że zachodzi:

$$\bigwedge_{\varepsilon > 0} \lim_{n \rightarrow \infty} P\{|\hat{\theta}_n - \theta| > \varepsilon\} = 0$$

Inaczej mówiąc, **estymator dąży do estymowanego parametru z prawdopodobieństwem 1 gdy licznosc próby rośnie (dąży do nieskończoności).**

Z nierówności Czebyszewa wynika, że wystarczy spełnić warunek:

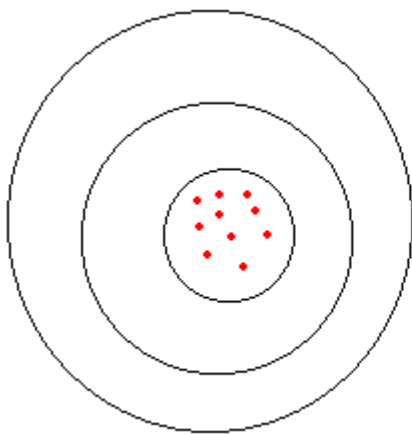
$$\lim_{n \rightarrow \infty} E(\hat{\theta}_n - \theta)^2 = 0$$

tzn. wystarcza zbieżność w sensie średniokwadratowym.

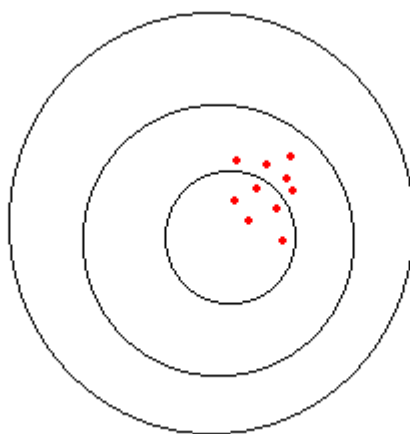
2. Obciążenie estymatora.

Przykład (Bowers USA):

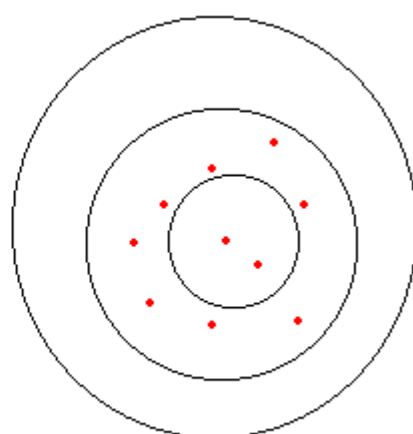
Dokonano oceny jakości 3 rewolwerów. Z każdego z nich oddaje się np. 10 próbnych strzałów do tarczy. Jeśli wyniki są takie jak na rysunku, to najlepszym rewolwerem jest rewolwer strzelający do tarczy A:



Tarcza A



Tarcza B



Tarcza C

Ocena ta jest oparta na:

- położeniu punktu wokół, którego skupiają się trafienia,
- wielkości rozrzutu.

W ocenie estymatora trzeba więc brać pod uwagę 2 jego własności:

1. zaobserwowane wartości estymatora powinny skupiać się dookoła estymowanej wielkości,
2. ich rozrzut powinien być możliwie mały.

Warunek 1 będzie spełniony, jeśli wartość średnia estymatora jest identyczna z wartością, którą chcemy estymować.

Warunek 2 będzie spełniony jeśli wariancja estymatora jest możliwie mała.

Estymator spełniający pierwszy warunek nazywamy **estymatorem nieobciążonym**. Jeśli istnieje nieobciążony estymator o najmniejszej wariancji, to nazywamy go **estymatorem nieobciążonym o minimalnej wariancji**.

Mówimy, że estymator $\hat{\theta}_n$ wielkości θ jest estymatorem nieobciążonym tej wielkości, jeśli jego wartość oczekiwana równa jest szacowanemu parametrowi θ , czyli:

$$\bigwedge_{n \in N} E[\hat{\theta}_n] = \theta$$

Jeśli $\bigwedge_{n \in N} E[\hat{\theta}_n] \neq \theta$, to estymator $\hat{\theta}_n$ nazywamy **obciążonym**.

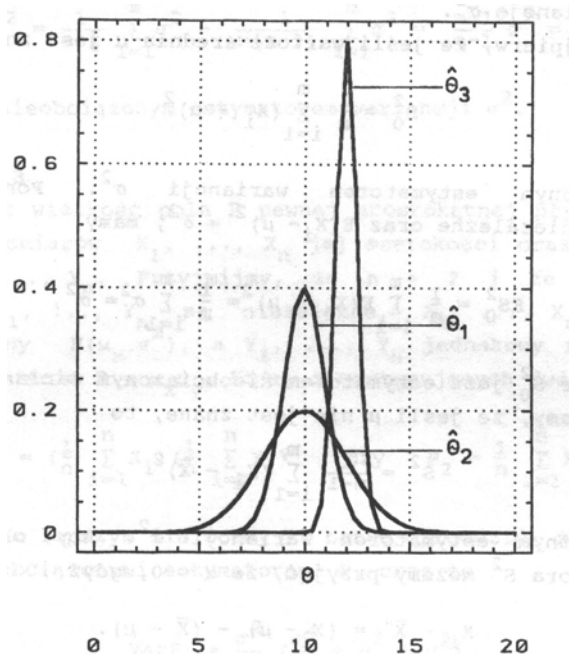
Estymator $\hat{\theta}_n$ nazywamy **asymptotycznie nieobciążonym** estymatorem parametru θ , jeśli jest spełniony warunek:

$$\bigwedge_{n \in N} \lim_{n \rightarrow \infty} E[\hat{\theta}_n] = \theta$$

Różnicę: $o(\hat{\theta})_n = E[\hat{\theta}_n] - \theta$ nazywamy **obciążeniem estymatora $\hat{\theta}_n$** .

Przykład 1:

Jeśli mamy 3 estymatory: $\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3$ parametru θ o wykresach jak na rysunku:



to $\hat{\theta}_1$ jest najlepszym estymatorem parametru θ .

Przykład 2:

Niech zmienne losowe X_1, X_2, \dots, X_m są niezależne oraz, mają jednakową wartość średnią μ i jednakową wariancję σ^2 , tzn., że:

$$EX_1 = EX_2 = \dots = EX_m = \mu \quad DX_1^2 = DX_2^2 = \dots = DX_m^2 = \sigma^2$$

Ponieważ: $E(X_1 + X_2 + \dots + X_m) = EX_1 + EX_2 + \dots + EX_m$, zatem wartość średnia:

$$\bar{X} = \frac{1}{m} \sum_{i=1}^m x_i \quad \text{jest estymatorem nieobciążonym wartości średniej } \mu,$$

ponieważ $E\bar{X} = \mu$. Estymator \bar{X} jest estymatorem nieobciążonym dla wartości średniej μ nawet wtedy, gdy zmienne losowe X_1, X_2, \dots, X_m zależne.

Przykład 3:

Weźmy pod uwagę wariancję z próby S^2 jako estymator wariancji σ^2 zmiennej losowej X . Jak łatwo policzyć:

$$S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2. \quad \text{Zatem wartość oczekiwana tego estymatora:}$$

$$E(S^2) = E\left(\frac{1}{n} \sum_{i=1}^n X_i^2\right) - E(\bar{X}^2) = m_2 - E(\bar{X}^2), \quad \text{gdzie:}$$

$$m_2 = E(X^2) = \sigma^2 + m^2$$

a ponieważ: $E(\bar{X}) = m$, $D^2(\bar{X}) = \frac{\sigma^2}{n}$, więc: $E(\bar{X}^2) = D^2(\bar{X}) + [E(\bar{X})]^2 = \frac{\sigma^2}{n} + m^2$

Czyli wartość oczekiwana tego estymatora wariancji wyniesie:

$$E(S^2) = m_2 - \frac{\sigma^2}{n} - m^2 - \sigma^2 = \frac{n-1}{n} \sigma^2$$

Tym samym wariancja z próby S^2 jest estymatorem obciążonym parametru σ^2 . Wzór ten sugeruje, aby jako estymatora wariancji σ^2 używać statystyki:

$$\hat{S}^2 = \frac{n}{n-1} S^2, \quad \text{która jest estymatorem nieobciążonym: } E(\hat{S}^2) = \sigma^2$$

3. Efektywność estymatora.

Definicja. Estymatorem najefektywniejszym parametru nazywamy ten spośród nieobciążonych estymatorów tego parametru, który ma najmniejszą wariancję, bądź też błąd średniokwadratowy estymatora jest mniejszy niż dla wszystkich innych możliwych estymatorów:

$$E[(\hat{\Theta}_1 - \Theta)^2] \leq E[(\hat{\Theta}_i - \Theta)^2],$$

gdzie – $\hat{\Theta}_1$ – nasz estymator, $\hat{\Theta}_i$ – dowolny inny estymator.

Jeśli najefektywniejszy estymator $\hat{\Theta}_n$ parametru Θ istnieje, to wybór jego umożliwia tzw.

nierówność Rao-Cramera.

Twierdzenie. Jeśli X_1, X_2, \dots, X_n jest n-wymiarową zmienną losową o gęstości:

$f_n(x, \Theta) = f_1(x_1, \Theta)f_1(x_2, \Theta) \dots f_1(x_n, \Theta)$, gdzie: f_1 – gęstość zm. los. X , spełniającą warunek:

$$\frac{\partial}{\partial \Theta} \int_{R^1} f_1(x_1, \Theta) dx_1 = \int_{R^1} \frac{\partial f_1(x_1, \Theta)}{\partial \Theta} dx_1,$$

a $\hat{\Theta}_n$ jest nieobciążonym estymatorem parametru Θ , to zachodzi:

$$V(\hat{\Theta}_n) \geq \frac{1}{E\left(\frac{\partial \ln f_n(X, \Theta)}{\partial \Theta}\right)^2} = \frac{1}{nE\left(\frac{\partial \ln f_1(X_1, \Theta)}{\partial \Theta}\right)^2}$$

Przy czym równość zachodzi wtedy i tylko wtedy, gdy z prawdopodobieństwem 1 jest:

$$\frac{\partial \ln f_n(X, \Theta)}{\partial \Theta} = c(\hat{\Theta}_n - \Theta),$$

gdzie c jest stałą, być może zależną od Θ .

Jeśli najefektywniejszy estymator parametru Θ istnieje, to dla jego wariancji zachodzi równość w powyższym wzorze.

Fakt istnienia estymatora najefektywniejszego wykorzystuje się do porównywania wariancji innych estymatorów tego samego parametru z wariancją estymatora najefektywniejszego. Niech $\hat{\Theta}$ i $\hat{\Theta}_0$ będą dwoma nieobciążonymi estymatorami tego samego parametru Θ i niech $\hat{\Theta}_0$ będzie estymatorem najefektywniejszym. Wielkość:

$$eff(\hat{\Theta}) = \frac{V(\hat{\Theta}_0)}{V(\hat{\Theta})} \quad \text{przyjmuje się za miarę efektywności}$$

estymatora $\hat{\Theta}$. Liczba $eff(\hat{\Theta})$, nazywana jest **efektywnością estymatora** $\hat{\Theta}$ i spełnia nierówność:

$$0 < eff(\hat{\Theta}) \leq 1$$

Estymator $\hat{\Theta}$, dla którego $eff(\hat{\Theta}) = 1$ jest estymatorem najefektywniejszym. Jeśli dla estymatora $\hat{\Theta}$ nie jest spełniona ta nierówność, ale jest spełniona równość:

$$\lim_{n \rightarrow \infty} eff(\hat{\Theta}) = 1, \text{ to estymator } \hat{\Theta} \text{ nazywamy } \mathbf{asymptotycznie}$$

najefektywniejszym.

Wielkość $leff(\hat{\Theta}) = \lim_{n \rightarrow \infty} eff(\hat{\Theta})$ bywa nazywana **asymptotyczną efektywnością estymatora**.

Estymatory wartości przeciętnej

Niech cecha X elementów populacji generalnej ma moment rzędu pierwszego $E[X] = m$. Przyjmujemy za estymator parametru m empiryczną wartość przeciętną:

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

Ponieważ: $E[\bar{X}_n] = E[\frac{1}{n} \sum_{i=1}^n X_i] = \frac{1}{n} E[\sum_{i=1}^n X_i] = \frac{1}{n} nm_1 = m_1$, zatem estymator \bar{X}_n jest estymatorem nieobciążonym. Błąd średniokwadratowy wartości średniej z próby \bar{X}_n , dla populacji o rozkładzie normalnym $N(m; \sigma)$, jest równy:

$$E[(\bar{X}_n - m_1)^2] = E[\frac{1}{n} \sum_{i=1}^n (X_i - m_1)^2] = \frac{1}{n^2} E[\sum_{i=1}^n (X_i - m_1)^2] = \frac{1}{n^2} n \sigma_x^2 = \frac{\sigma_x^2}{n},$$

Na podstawie prawa wielkich liczb Markowa: $\lim_{n \rightarrow \infty} E[(\bar{\phi} - \phi)^2] = 0$ można stwierdzić, że estymator \bar{X}_n jest estymatorem zgodnym. A efektywność estymatora? Można wykazać, stosując wzór Rao-Cramera, że **\bar{X}_n jest najefektywniejszym estymatorem parametru m dla populacji o rozkładzie normalnym.**

Dowód : (estymacja wartości średniej)

Chcemy estymować wartość średnią m cechy X w populacji generalnej. W tym celu pobierzemy próbę prostą n elementową (x_1, x_2, \dots, x_n) i wyznaczymy z niej

$$m_n = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

którą traktujemy jako ocenę parametru m , czyli wartość estymatora tego parametru danego wzorem

$$M_n = \frac{1}{n} \sum_{i=1}^n X_i, \text{ gdzie}$$

(X_1, X_2, \dots, X_n) jest próbą losową.

Na podstawie słabego prawa wielkich liczb *Chinczyzna* jest to estymator zgodny parametru m . Zbadajmy, czy jest to estymator nieobciążony :

$$E(M_n) = E(\frac{1}{n} \sum_{i=1}^n X_i) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{m \cdot n}{n} = m, \quad n = 1, 2, \dots$$

Żeby stwierdzić, czy jest to estymator najefektywniejszy parametru m musimy przyjąć założenie o rozkładzie cechy X w populacji generalnej. Przyjmijmy, że jest to rozkład $N(m, \sigma)$

$$f(x, g) = f(x, m) = \frac{1}{\sigma \sqrt{2\pi}} \exp[-\frac{(x-m)^2}{2\sigma^2}]$$

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

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

mamy zatem

$$\frac{1}{n E[\frac{(x-m)^2}{\sigma^4}]} = \frac{1}{n \cdot \frac{\sigma^2}{\sigma^4}} = \frac{\sigma^2}{n}$$

Wyrażmy teraz wariancję naszego estymatora : ($D^2(cX) = c^2 D^2(X)$)

$$D^2(M_n) = D^2\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n D^2(X_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

czyli

$$\sigma_{M_n} = \frac{\sigma}{\sqrt{n}}$$

Porównując oba wyniki widzimy, że nierówność *Rao - Cramera* jest spełniona, czyli widzimy, że estymator jest estymatorem najefektywniejszym.

Estymatory wariancji

Jeśli cecha X elementów populacji ma **rozkład normalny** $N(\mathbf{m}; \sigma)$, gdzie σ – jest nieznanne, to estymator wariancji S_n^2 :

$$S_n^2 = \frac{1}{n} \sum_{k=1}^n (X_k - \bar{X}_n)^2 \text{ jest zgodnym estymatorem parametru } \sigma^2.$$

Można również udowodnić, że S_n^2 jest obciążonym estymatorem tego parametru. Obciążenie tego estymatora wynosi:

$$\frac{n-1}{n} \sigma^2 - \sigma^2 = -\frac{1}{n} \sigma^2$$

Gdy $n \rightarrow \infty$, to obciążenie estymatora $S_n^2 \rightarrow 0$, tzn. że estymator ten jest asymptotycznie nieobciążony.

Można również wprowadzić nieobciążony estymator wariancji σ^2 , określony wzorem:

$$S_{nieob}^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2$$

Jeśli zbadamy efektywność estymatorów S_n^2 i S_{nieob}^2 , to okaże się, że:

$$eff(S_{nieob}^2) = \frac{n-1}{n}, \text{ czyli estymator } S_{nieob}^2 \text{ nie jest estymatorem}$$

najefektywniejszym, jest natomiast estymatorem asymptotycznie najefektywniejszym. Dla dużych wartości n oba estymatory mają w przybliżeniu równe wartości. W praktyce dla $n \leq 30$ korzysta się z estymatora S_{nieob}^2 , a dla $n > 30$ z estymatora S_n^2 .

Metody wyznaczania estymatorów.

Jeśli nie jest oczywiste jaką statystykę należy wybrać jako kandydata na estymator, z pomocą przychodzą różne metody ich wyznaczania:

1. **Metoda momentów (Pearsona).**
2. **Metoda największej wiarygodności (Fishera).**
3. **Metoda najmniejszych kwadratów.**

Metoda momentów.

Metoda ta polega na przyjmowaniu za oszacowanie nieznanych momentów cechy X elementów populacji generalnej, zaobserwowanych wartości momentów empirycznych. Estymatory uzyskane metodą momentów mają tę zaletę, że znajdowanie ich wartości jest związane na ogół z prostymi obliczeniami. Istotną ich wadą jest natomiast ich mała na ogół efektywność (za wyjątkiem, gdy cecha X ma rozkład normalny).

Przyjmijmy, że parametr θ jest jednoznacznie określony przez wartości pierwszych k momentów teoretycznych cechy. Oznacza to, że: $\theta = f(m_1, \dots, m_k)$. Estymator $\hat{\theta} = f(M_1, \dots, M_k)$, gdzie M_i są empirycznymi odpowiednikami momentów zwykłych m_i . W szczególności jeśli parametr θ jest funkcją tylko pierwszego momentu teoretycznego m , to estymator jest funkcją statystyki \bar{X} .

Metoda największej wiarygodności

Idea metody największej wiarygodności polega na oszacowaniu nieznanych parametrów tak, aby empiryczne dane były przy tym oszacowaniu najbardziej prawdopodobne. Dla znalezienia takiego estymatora konstruuje się **funkcję wiarygodności** L .

Niech rozkład zmiennej losowej dyskretnej X zależy od wektora parametrów $\theta = (\theta_1, \dots, \theta_n)$ lub w szczególności od jednego parametru θ . Niech (x_1, x_2, \dots, x_n) będzie zaobserwowaną wartością próby prostej (X_1, X_2, \dots, X_n) z populacji generalnej mającej cechę X . Dla przypadku dyskretnego funkcję wiarygodności określa się wzorem:

$$L(\theta) = L(x_1, x_2, \dots, x_n; \theta) = p(x_1; \theta) p(x_2; \theta) \dots p(x_n; \theta) \text{ gdzie } p(x_i; \theta) = P(X = x_i) .$$

Dla cechy typu ciągłego o gęstości $f(x, \theta)$, funkcja wiarygodności określona jest wzorem:

$$L(\theta) = L(x_1, x_2, \dots, x_n; \theta) = f(x_1; \theta) f(x_2; \theta) \dots f(x_n; \theta)$$

Estymatorem parametru θ jest ta jego wartość, przy której funkcja wiarygodności osiąga wartość największą. Jeśli funkcja L jest różniczkowalna, to jej maksimum można znaleźć, szukając miejsca zerowania się pochodnych cząstkowych $\partial L / \partial \theta_i$. Ponieważ funkcja L jest iloczynem funkcji, to wygodniej jest badać pochodne nie funkcji wiarygodności L , a jej logarytmu $\ln L$. Pozwala to na znaczne uproszczenie obliczeń. Można bowiem warunek konieczny wystąpienia maksimum funkcji L zapisać jako:

$$\frac{\partial \ln L}{\partial \theta_i} = 0 \text{ dla } i = 1, 2, \dots, n$$

Ta równość tworzy układ n równań o n niewiadomych i jeśli funkcja L ma maksimum, to estymatory można wyznaczyć z tego układu równań. Uzyskanie estymatorów metodą największej wiarygodności może być kłopotliwe ze względów rachunkowych.

Przykład:

Dla rozkładu wykładniczego z parametrem $\lambda > 0$: $f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{dla } x \geq 0, \\ 0 & \text{dla } x < 0, \end{cases}$

mamy: $L(\lambda) = \lambda^n e^{-\lambda(x_1 + x_2 + \dots + x_n)}$ Po logarytmowaniu otrzymamy:

$$\ln L = \ln \lambda^n - \lambda(x_1 + x_2 + \dots + x_n)$$

zaś po różniczkowaniu, otrzymuje się równanie:

$$\frac{\partial \ln L}{\partial \lambda} = \frac{n}{\lambda} - (x_1 + x_2 + \dots + x_n) = 0$$

czyli: $\frac{n}{\lambda} - (x_1 + x_2 + \dots + x_n) = 0$, a więc $\hat{\lambda} = 1/\bar{X}$

Metoda najmniejszych kwadratów

Niech X_1, X_2, \dots, X_n stanowi ciąg zmiennych losowych, których zaobserwowane wartości w próbie stanowią ciąg x_1, x_2, \dots, x_n . Rozkład zmiennych losowych X_i niech zależy od nieznanymi parametrów $\theta_1, \theta_2, \dots, \theta_k$, których wartości należy oszacować. Niech zależność funkcyjna zmiennej losowej X_i względem parametrów $\theta_1, \theta_2, \dots, \theta_k$ będzie znana i może być zapisana w formie funkcji:

$$X_i = g_i(\theta_1, \theta_2, \dots, \theta_k)$$

przy czym postać funkcji g powinna być liniowa lub za pomocą różniczek zupełnej powinna być doprowadzona do formy liniowej. Parametry θ_j oraz kształt funkcji g zależą od specyfiki rozpatrywanego zagadnienia.

Metoda najmniejszych kwadratów polega na takim dobraniu ocen parametrów θ_j , aby odchyłki (oceny) δ_j dla zmiennej losowej X_i spełniały warunek:

$$F = \sum_{i=1}^n [x_i - g_i(\theta_1, \theta_2, \dots, \theta_k)]^2 = \sum_{i=1}^n (\delta_i)^2 = \min \quad (2.1)$$

Minimum funkcji powyższej zachodzi dla takich wartości θ_j , dla których wszystkie pochodne cząstkowe będą równe zeru, czyli:

$$\frac{\partial F}{\partial \theta_j} = -2 \sum_{i=1}^n [x_i - g_i(\theta_1, \theta_2, \dots, \theta_k)] \frac{\partial g_i}{\partial \theta_j} = 0 \quad (2.2)$$

Warunek powyższy, dla tego typu funkcji, stanowi również warunek dostateczny do istnienia minimum funkcji F .

Ze względu na liniowy model funkcji g_i , pochodne cząstkowe $\frac{\partial g_i}{\partial \theta_j}$ będą zawsze odpowiadać stałym współczynnikom niezależnym od wartości parametrów θ_j , zatem powyższy układ równań $\frac{\partial F}{\partial \theta_j}$ będzie przedstawiać liniowy układ równań względem parametrów θ_j , o wyrazach wolnych x_i lub $x_i - g_{0i}$.

Rozwiązanie układu równań tego typu prowadzi do oszacowania parametrów θ_j , czyli do wartości estymatora $\hat{\theta}_j$.

Zastosowanie metody najmniejszych kwadratów do estymacji liniowych modeli stochastycznych sprowadza się do oszacowania nieznanych parametrów a_i w modelu liniowym postaci:

$$Y = \sum_{i=1}^k a_i Z_i \quad \text{gdzie: } Y, Z_i - \text{zmiennne losowe}$$

Przykład:

Estymacja wartości oczekiwanej ciągu zmiennych losowych X_1, X_2, \dots, X_n , o których wiadomo, że:

$$E(X_1) = E(X_2) = \dots = E(X_n) = \theta, \text{ oraz, że wszystkie zmienne mają tę samą wariancję.}$$

Niech x_1, x_2, \dots, x_n oznaczają zaobserwowane wartości zmiennych losowych w próbie. Warunek (2.1) zapisuje się w następującej postaci:

$$F = \sum_{i=1}^n [x_i - \theta]^2 = \min$$

Po obliczeniu pochodnej i przyrównaniu jej do zera, otrzymamy:

$$\frac{\partial F}{\partial \theta} = 2 \sum_{i=1}^n (x_i - \theta) = 0 \quad \text{co prowadzi do równania: } \sum_{i=1}^n x_i - n\theta = 0, \text{ zatem:}$$

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

Udowodniliśmy, że najlepszą oceną w sensie MNK dla wartości przeciętnej θ , jest średnia arytmetyczna wyników obserwacji, pod warunkiem, że wszystkie obserwowane w próbie zmienne losowe mają tę samą wariancję.

Zadania.

Zadanie 1.

Obliczyć prawdopodobieństwo tego, że losowo wybrany w Polsce mężczyzna ma wzrost w granicach między 157,5 cm a 162,5 cm.

Rozwiązanie:

Aby obliczyć to prawdopodobieństwo, trzeba najpierw zbudować model probabilistyczny doświadczenia polegającego na losowym wyborze mężczyzny i zmierzeniu jego wzrostu. Na podstawie licznych badań stwierdzono, że wzrost człowieka można traktować jako zmienną losową X o rozkładzie normalnym. Pozostaje więc tylko wybrać odpowiednie wartości parametrów m i σ tego rozkładu, albo inaczej

mówiąc oszacować te parametry. W tym celu mierzymy wzrost dużej liczby n losowo wybranych mężczyzn, otrzymując pewną próbę x_1, x_2, \dots, x_n zmiennej losowej X . Teraz trzeba wybrać odpowiednią metodę szacowania parametrów m i σ na podstawie próby. Przyjmujemy estymatory:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i, \quad \hat{S} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})^2}$$

Przypuśćmy, że otrzymaliśmy następujące wartości liczbowe: $\bar{X}=170$, $\hat{S}=5$. Przyjmujemy jako probabilistyczny model wzrostu mężczyzn w Polsce zmienną losową X o rozkładzie normalnym $N(170;5)$. Obliczamy szukane prawdopodobieństwo:

$$P(157,5 < X < 162,5) = P\left(\frac{157,5-170}{5} < \frac{X-170}{5} < \frac{162,5-170}{5}\right) = F(2,5) - F(1,5) = 0,061$$

Interpretacja wyniku: 6,1% mężczyzn w Polsce ma wzrost w granicach między 157,5 a 162,5 cm. Zatem producent garniturów męskich może stąd wyciągnąć wniosek, że 6,1% ogólnej liczby produkowanych garniturów powinno odpowiadać wzrostowi 160 cm (przy numeracji co 5 cm).

Zadanie 2.

Niech X_1 i X_2 będą zmiennymi losowymi niezależnymi takimi, że:

$EX_1 = 1$, $EX_2 = 3$, $D^2 X_1 = D^2 X_2 = \sigma^2$. Dla jakiej stałej c statystyka $T = cX_1^2 + (1-c)X_2^2$ jest estymatorem nieobciążonym parametru σ^2 .

Zadanie 3.

Niech X_1, X_2, X_3 będą niezależnymi zmiennymi losowymi o jednakowym rozkładzie z nieznaną wartością oczekiwaną m i skończoną wariancją σ^2 . Udowodnić, że poniższe 3 statystyki:

$$T_1 = \frac{1}{6} X_1 + \frac{1}{6} X_2 + \frac{2}{3} X_3, \quad T_2 = \frac{1}{5} X_1 + \frac{2}{5} X_2 + \frac{2}{5} X_3, \quad T_3 = \frac{1}{3} X_1 + \frac{1}{3} X_2 + \frac{1}{3} X_3$$

są nieobciążonymi estymatorami parametru m . Który z nich jest najlepszy?

Zadanie 4.

Gęstość rozkładu zmiennej losowej X jest następująca:

$$f(x) = \begin{cases} \frac{|x|}{a^2} & \text{dla } x \in [-a, a] \\ 0 & \text{dla } x \notin [-a, a] \end{cases}$$

Znaleźć estymator parametru a metodą momentów dla próby X_1, X_2, \dots, X_n o tym samym rozkładzie mającej realizację x_1, x_2, \dots, x_n .

Zadanie 5.

Wyznaczyć metodą momentów estymator parametru c dla rozkładu o gęstości:

$$f(x) = \begin{cases} \frac{c}{x^{c+1}} & \text{dla } x > 1 \\ 0 & \text{dla } x \leq 1 \end{cases}$$