DZIAŁ NAUKOWY

MATEMATYKA STOSOWANA 8, 2007

Teresa Ledwina (Wrocław)

O asymptotycznej efektywności estymatorów

Streszczenie. W pracy przedstawiamy i dyskutujemy pojęcie asymptotycznej efektywności estymatorów w ujęciu Hájeka i Le Cama. Podajemy też ogólną konstrukcję pewnej klasy asymptotycznie optymalnych estymatorów dla parametrów z przestrzeni euklidesowej. Pokrótce szkicujemy uogólnienia dyskutowanych idei na przypadek semiparametryczny i pokazujemy, że techniczne wyniki uzyskane w teorii asymptotycznie efektywnej estymacji mogą być z powodzeniem wykorzystane w asymptotycznej teorii testowania.

Wybór materiału jest wysoce subiektywny i tylko w niewielkim stopniu oddaje złożoność rozpatrywanych współcześnie zagadnień oraz ogrom wyników, jakie uzyskano w tej tematyce. Tekst jest skróconą wersją wykładu przygotowanego na zaproszenie Organizatorów Konferencji ze Statystyki Matematycznej – Wisła 2005. Głównym celem prezentacji jest pokazanie, że klasyczne podejście do definiowania asymptotycznej efektywności nie sprawdziło się i przedyskutowanie tego jak, dla pewnej klasy zagadnień, w naturalny i elegancki sposób został ten problem rozwiązany.

Słowa kluczowe: asymptotyczna efektywność, asymptotyczna optymalność, funkcja wpływu, superefektywność, test wynikowy.

1. Klasyczne podejście do asymptotycznej efektywności. Pierwsze, niezbyt formalne, próby definiowania i udowadniania asymptotycznej optymalności estymatorów pochodziły od Edgewortha (1908). Fisher (1922, 1925) zrobił istotny krok poprzez dużo bardziej formalne rozważania dla ogólnej jednoparametrowej rodziny rozkładów. Choć jego wywody nie były całkiem ścisłe, prace te były bardzo istotne. Wielu autorów (np. Doob 1934) formalizowało wywody Fishera. Większość tych formalizacji była zbliżona do klasycznego dziś podejścia Craméra (1946). Dla kompletności prezentacji przedstawiamy poniżej wariant takiego rozwiązania. Dla prostoty ograniczymy się do przypadku, gdy estymujemy parametr z prostej.

Rozważamy model $\mathcal{P} = \{P_\theta : \theta \in \Theta \subset R\}$. Zakładamy, że rozkłady P_θ posiadają gęstości p_θ względem pewnej σ-skończonej miary dominującej μ i informacja Fishera

$$I_\theta = \int_R \left[\frac{\partial}{\partial\theta}\log p_\theta(x)\right]^2 p_\theta(x)\mu(dx)$$

istnieje oraz spełnia $I_\theta \in (0,\infty)$.

Dla odróżnienia klasycznego rozwiązania od rozwiązań współczesnych będziemy używać nazwy v-efektywność na ujęcie klasyczne. Taką nazwę wprowadził Rao (1963). Tradycyjnie rozważania ogranicza się do klasy $\{T_n\}$ zgodnych i asymptotycznie normalnych estymatorów parametru θ, to znaczy takich, że

$$\sqrt{n}(T_n - \theta) \xrightarrow{\mathcal{D}} N(0, v(\theta)). \tag{1}$$

DEFINICJA 1. Ciąg estymatorów $\{T_n\}$ spełniający (1) z $v(\theta) = 1/I_\theta$ nazywamy v-efektywnym.

Sztandarowymi przykładami estymatorów v-efektywnych były estymatory największej wiarogodności i estymatory jednokrokowe. Omówimy pokrótce oba te przykłady.

1.1. *Klasyczne założenia regularności o* $\mathcal{P} = \{P_\theta : \theta \in \Theta \subset R,\ P_\theta \prec\prec \mu\}$. Rozważmy założenia:

(i) Θ jest zbiorem otwartym.

(ii) Rozkłady P_θ mają wspólny nośnik A, który nie zależy od θ.

(iii) Dla każdego $x \in A$ gęstość p_θ jest trzykrotnie ciągle różniczkowalna względem θ.

(iv) Funkcja $\int p_\theta(x)\mu(dx)$ jest dwukrotnie różniczkowalna po θ pod znakiem całki.

(v) $I_\theta \in (0, \infty)$.

(vi) Dla każdego $\theta_0 \in \Theta$ istnieją dodatnia liczba c i funkcja $M(x)$ (być może obie zależne od θ_0) takie, że

$$\left| \frac{\partial^3 \log p_\theta(x)}{\partial \theta^3} \right| \le M(x), \quad \forall x \in A, \quad \forall \theta \in (\theta_0 - c, \theta_0 + c)$$

oraz $\int M(x) p_{\theta_0}(x)\mu(dx) < \infty$.

1.2. *Funkcja wiarogodności i estymatory największej wiarogodności.* Dla wyników $x_1, ..., x_n$ niezależnych obserwacji o rozkładzie P_θ oznaczmy przez

$$L(\theta) = \sum_{i=1}^{n} \log p_\theta(x_i)$$

logarytm funkcji wiarogodności.

Niech L' oznacza pochodną L względem θ i niech estymator $\widehat{\theta}_n$ będzie rozwiązaniem równania

$$L'(\widehat{\theta}_n) = 0. \tag{2}$$

TWIERDZENIE 1. *Zakładamy, że $\mathcal{P}$ spełnia (i)–(vi). Jeśli $\widehat{\theta}_n$, będące rozwiązaniem (2), jest zgodnym estymatorem θ, to*

$$\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{\mathcal{D}} N(0, 1/I_\theta).$$

Kwestia zgodności rozwiązania równania (2) jest problemem nietrywialnym. Warunki zgodności badali między innymi Le Cam (1953, 1970), Kiefer i Wolfowitz (1956) oraz Zacks (1971). Wiadomo, że są sytuacje, gdy zgodności nie ma. Dla uniknięcia powyższych kłopotów zaproponowano następujące przybliżone rozwiązanie.

1.3. *Estymatory jednokrokowe.* Niech $\widehat{\theta}_n$ będzie rozwiązaniem równania (2) i niech $\widetilde{\theta}_n$ będzie jakimś innym estymatorem θ. Przy założeniu (iii) funkcja wiarogodności L jest trzykrotnie różniczkowalna. Z wzoru Taylora dla L' mamy

$$0 = L'(\widehat{\theta}_n) = L'(\widetilde{\theta}_n) + (\widehat{\theta}_n - \widetilde{\theta}_n)L''(\widetilde{\theta}_n) + R_n,$$

gdzie

$$R_n = (\widehat{\theta}_n - \widetilde{\theta}_n)^2 L'''(\widetilde{\theta}_n^*)/2,$$

a $\widetilde{\theta}_n^*$ jest punktem pośrednim między $\widehat{\theta}_n$ i $\widetilde{\theta}_n$. Zdefiniujmy δ_n poprzez relację

$$0 = L'(\widetilde{\theta}_n) + (\delta_n - \widetilde{\theta}_n)L''(\widetilde{\theta}_n).$$

Rozwiązanie δ_n nazywamy jednokrokowym estymatorem opartym na $\widetilde{\theta}_n$. Oczywiście

$$\delta_n = \widetilde{\theta}_n - \frac{L'(\widetilde{\theta}_n)}{L''(\widetilde{\theta}_n)}.$$

Użyteczność tej konstrukcji wynika z poniższego twierdzenia i wniosku. Przed ich sformułowaniem przypomnijmy, że ciąg estymatorów $\{T_n\}$ parametru θ jest $\sqrt{n}$-zgodny, jeśli ciąg zmiennych losowych $\{\sqrt{n}(T_n - \theta)\}$ jest ograniczony według prawdopodobieństwa P_θ.

Twierdzenie 2. *Niech $\widetilde{\theta}_n$ będzie jakimś $\sqrt{n}$-zgodnym estymatorem θ. Przy założeniach (i)–(vi) estymator δ_n jest v-efektywny.*

Wniosek 1. *Jeśli I_θ jest ciągłą funkcją θ to, przy założeniach twierdzenia 2, estymator*

$$\delta_n^0 = \widetilde{\theta}_n + \frac{L'(\widetilde{\theta}_n)}{nI_{\widetilde{\theta}_n}} = \widetilde{\theta}_n + \frac{1}{n}\sum_{i=1}^{n} I_{\widetilde{\theta}_n}^{-1}\dot{\ell}_{\widetilde{\theta}_n}(X_i), \tag{3}$$

gdzie

$$\dot{\ell}_\theta(x) = \frac{\partial}{\partial\theta}\log p_\theta(x),$$

jest v-efektywny.

Funkcję $\dot{\ell}_\theta(x)$ będziemy nazywać funkcją wynikową.

1.4. *Superefektywność i problemy pochodne.* W 1953 r. Hodges podał przykład, który zachwiał bezkrytyczną wiarą w użyteczność i sensowność definicji v-efektywności. Mianowicie, Hodges zdefiniował ciąg estymatorów $\{S_n\}$, dla którego zachodzi

$$\sqrt{n}(S_n - \theta) \xrightarrow{\mathcal{D}} N(0, v(\theta)), \quad v(\theta) \le 1/I(\theta) \quad \forall\theta, \tag{4}$$

z ostrą nierównością dla pewnego θ.

Własność (4) nazwano superefektywnością.

Przykład Hodgesa. Niech $X_1, ..., X_n$ będą niezależnymi zmiennymi losowymi o rozkładzie $P_\theta = N(\theta, 1)$ dla każdej zmiennej. Zdefiniujmy

$$S_n = \overline{X} I\{|\overline{X}| \ge n^{-1/4}\} + a\, \overline{X} I\{|\overline{X}| < n^{-1/4}\},$$

gdzie $I\{A\}$ oznacza indykator zdarzenia A, a $\overline{X} = \sum_{i=1}^n X_i/n$. Łatwo pokazać, że $\sqrt{n}(S_n - \theta) \xrightarrow{\mathcal{D}} N(0, v(\theta))$, gdzie $v(\theta) = I\{\theta \ne 0\} + a^2 I\{\theta = 0\}$. Oczywiście, biorąc odpowiednio małe a, możemy uczynić $v(\theta)$ dowolnie małym w punkcie $\theta = 0$. Jest to jednak dość iluzoryczny zysk. Aby to zrozumieć, rozważmy znormalizowane ryzyko estymatora S_n dane wzorem

$$R_n(\theta) = nE_\theta(S_n - \theta)^2,$$

gdzie $E_\theta(\bullet)$ oznacza wartość oczekiwaną zmiennej $\bullet$ liczoną przy rozkładzie P_θ. Zanotujmy, że znormalizowane ryzyko estymatora $\overline{X}$ wynosi 1 dla każdego θ. Z postaci $R_n(\theta)$ (por. Lehmann 1983, s. 408) wynika, że $R_n(\theta) \to 1$, jeśli $\theta \ne 0$, oraz $R_n(\theta) \to a^2$, gdy $\theta = 0$. Ponadto, dla $\theta_n = n^{-1/4}$ zachodzi $R_n(\theta_n) \to \infty$, co implikuje $\sup_\theta R_n(\theta) \to \infty$. Dla $a = 0$ i kilku wybranych n rysunek 8.1 w książce van der Vaarta (2000) ilustruje, jak mocno oscyluje znormalizowane ryzyko $R_n(\theta)$ estymatora S_n w pobliżu punktu $\theta = 0$. Tak więc mniejsza wariancja S_n w $\theta = 0$ istotnie „rozregulowuje” zachowanie ryzyka w otoczeniu $\theta = 0$.

W ogólnej sytuacji Le Cam (1953) i Huber (1966) pokazali, że dla $\theta \in R$ oraz $\theta \in R^2$ superefektywność w pewnym punkcie θ_0 powoduje niepożądane własności ryzyka w otoczeniu θ_0.

Warto odnotować, że dla $\theta \in R^k$, $k \ge 3$, superefektywność nie musi mieć tak groźnego wpływu na ryzyko, a superefektywne estymatory mogą mieć dobre własności. Kilka uwag na ten temat zamieszczamy poniżej.

Dla zwartości prezentacji zauważmy, że, przy dodatkowym założeniu o jednostajnej całkowalności $n(T_n - \theta)^2$, v-efektywność estymatora T_n parametru $\theta \in R$ implikuje relację

$$\lim_{n\to\infty} nE_\theta|T_n - \theta|^2 = 1/I_\theta. \tag{5}$$

Naturalnym analogonem (5) dla $\theta \in R^k$ jest warunek

$$\lim_{n\to\infty} nE_\theta||T_n - \theta||^2 = \mathrm{tr}\{I_\theta^{-1}\}, \tag{6}$$

gdzie $|| \ \ ||$ oznacza normę euklidesową w R^k a $\mathrm{tr}\{\bullet\}$ ślad macierzy $\bullet$. Dla ilustracji rozważmy teraz k wymiarowe wektory X_i o rozkładzie $N(0, I)$, I macierz identycznościowa wymiaru $k \times k$ i następujący estymator Jamesa-Steina (1961)

$$T_n = \overline{X} - (k-2)\overline{X}/(n||\overline{X}||)^2, \tag{7}$$

gdzie, jak poprzednio, $\overline{X} = \sum_{i=1}^n X_i/n$. Znormalizowane ryzyko tego estymatora ma postać (por. Lehmann 1983, str. 306 i 294)

$$nE_\theta||T_n - \theta||^2 = k\left[1 - \frac{(k-2)^2}{k} E_\theta||\sqrt{n}\overline{X}||^{-2}\right].$$

Zauważmy, że dla $\theta \neq 0$ zachodzi

$$\lim_{n\to\infty} E_\theta||\sqrt{n}\overline{X}||^{-2} = 0 \quad \text{i} \quad \lim_{n\to\infty} nE_\theta||T_n - \theta||^2 = k = \mathrm{tr}\{I\}.$$

Ponadto, dla $X_1 \sim N(\theta, I)$ oraz $k \geq 3$, mamy

$$\frac{1}{k - 2 + ||\theta||^2} \leq E_\theta \frac{1}{||X_1||^2} \leq \frac{1}{k-2}\left(\frac{k}{k + ||\theta||^2}\right)$$

(por. Casella i Hwang 1982).

Wobec tego, dla $\theta = 0$ dostajemy $\lim_{n\to\infty} nE_\theta||T_n - \theta||^2 = 2$. To pokazuje, że T_n dany wzorem (7) jest superefektywny w sensie definicji (6). Z drugiej strony przy $k \geq 3$ zachodzi relacja

$$nE_\theta||T_n - \theta||^2 < k = nE_\theta||\overline{X} - \theta||^2, \quad \forall\theta,\ \forall n.$$

Tak więc, przy wymiarze $k \geq 3$ oraz przy dowolnych ustalonych n i k, superefektywność zredukowała ryzyko v-efektywnego estymatora $\overline{X}$ w całej przestrzeni parametrów. Tego typu zjawisko powoduje, że superefektywne estymatory są do dziś obiektem badań i okazują się użyteczne w wielu sytuacjach.

W literaturze w latach 60. i 70. ubiegłego wieku można zaobserwować rozmaite reakcje na superefektywność. Wymienimy tu cztery nurty badań.

1. Wykazywanie, że dla asymptotycznie normalnych estymatorów miara Lebesgue'a zbioru punktów θ, w których ma miejsce superefektywność, wynosi 0 (Le Cam 1953, Bahadur 1964). Były to oczywiście interesujące wyniki, ale przykład estymatora Jamesa-Steina pokazuje, że zbiór miary 0 ma znaczenie w praktyce.
2. Próby formułowania warunków, przy których $\sqrt{n}(T_n - \theta) \to L_\theta$, jednostajnie względem θ wziętej ze zbiorów zwartych, gdzie L_θ jest pewną zmienną losową (Rao 1963, Wolfowitz 1965). Wyniki te były na tyle wąskie, że nie rodziły nadziei na uzyskanie wniosków istotnych dla praktyki.
3. Ograniczenie rozważań do modeli i estymatorów T_n, dla których

   $$\sqrt{n}(T_n - \theta) \to L_\theta$$

   lokalnie jednostajnie.

4. Ograniczenie rozważań do gładkich modeli i dowodzenie lokalnych twierdzeń minimaksowych.

Dwa ostatnie punkty wiążą się ściśle z wynikami dowodzonymi przez Le Cama począwszy od 1953 r. i rozwiniętymi przez Hájeka w latach 1967–1972. Ich teoria rozwiązała pewną klasę problemów. Głównym przesłaniem wynikającym z tej teorii i problemów związanych z superefektywnością, których ona nie objęła, jest uwypuklenie wagi badań pewnych jednostajności przy porównywaniu estymatorów.

2. Asymptotyczna optymalność estymatorów w ujęciu Hájeka–Le Cama. Przypadek estymacji $\theta \in \Theta \subset R^k$.

2.1. *Nowe warunki regularności.* Prezentację podejścia Hájeka–Le Cama zaczniemy od przedstawienia wprowadzonego przez nich ujęcia warunków regularności. Ma ono dwie podstawowe zalety: pozwala na osłabienie klasycznych założeń oraz umożliwia naturalne rozszerzenie teorii modeli parametrycznych na modele semiparametryczne i nieparametryczne.

Rozważmy model $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$, gdzie P_θ są określone na pewnej przestrzeni $(\mathcal{X}, \mathcal{A})$, Θ jest zbiorem otwartym w R^k, a P_θ są absolutnie ciągłe względem pewnej σ-skończonej miary μ. Oznaczmy $p_\theta = dP_\theta/d\mu$ i wprowadźmy pomocniczą definicję.

Definicja 2 . Mówimy, że funkcja $\theta \to \sqrt{p_\theta}$ jest różniczkowalna w sensie średniokwadratowym jeśli istnieje wektor $\dot{\ell}_\theta = (\dot{\ell}_{\theta,1}, ..., \dot{\ell}_{\theta,k})^T$ funkcji mierzalnych taki, że $E_\theta \dot{\ell}_\theta(X) = 0$ i $E_\theta ||\dot{\ell}_\theta(X)||^2 < \infty$ oraz

$$\int_{\mathcal{X}} \left[\sqrt{p_{\theta+h}} - \sqrt{p_\theta} - \frac{1}{2} h^T \dot{\ell}_\theta \sqrt{p_\theta} \right]^2 d\mu = o(||h||^2), \quad h \to 0. \tag{8}$$

Tu i poniżej wszystkie wektory są kolumnowe, a górny wskaźnik ^{T} oznacza transpozycję.

Następna definicja sprowadza regularność modelu $\mathcal{P}$ do średniokwadratowej różniczkowalności funkcji $\theta \mapsto \sqrt{p_\theta}$.

Definicja 3 . Jeśli dla gęstości p_θ miar P_θ, $\theta \in \Theta$, zachodzi warunek (8), to mówimy, że model $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ jest różniczkowalny w sensie średniokwadratowym w punkcie θ. Taki model nazywać będziemy regularnym. Funkcję $\dot{\ell}_\theta$ nazywa się funkcją wynikową lub pochodną Hellingera. Natomiast macierz $I_\theta = E_\theta \dot{\ell}_\theta(X)[\dot{\ell}_\theta(X)]^T$ nosi nazwę macierzy informacji.

Przypomnijmy, że jeśli dla każdego x z nośnika gęstość $p_\theta(x)$ była różniczkowalna względem θ, to przy $k = 1$ funkcja wynikowa była zdefiniowana jako $\dot{\ell}_\theta(x) = \frac{\partial}{\partial\theta} \log p_\theta(x)$. Zauważmy, że dla $s_\theta(x) = \sqrt{p_\theta(x)}$ zachodzi

$$\dot{s}_\theta(x) = \frac{\partial}{\partial\theta} s_\theta(x) = \frac{1}{2} \dot{\ell}_\theta(x) \sqrt{p_\theta(x)}.$$

Jest oczywiste, że nakładając dalsze warunki gładkości na p_θ możnaby uzyskać dla każdego x z nośnika P_θ następującą relację

$$\sqrt{p_{\theta+h}(x)} - \sqrt{p_\theta(x)} - \frac{1}{2} h^T \dot{\ell}_\theta(x) \sqrt{p_\theta(x)} = o(||h||). \tag{9}$$

Widać więc, że zamiast postulowania kolejnych założeń o gładkości p_θ (por. rozdział 1.1), Hájek i Le Cam zaproponowali słabszy warunek stanowiący o małości lewej strony (9) w sensie średnim.

Poniższy lemat podaje proste warunki analityczne wystarczające dla zachodzenia (8).

LEMAT 1. *Niech* $\mathcal{P} = \{p_\theta : \theta \in \Theta \subset R^k\}$ *będzie rodziną gęstości względem miary* μ *na* $(\mathcal{X}, \mathcal{A})$. *Załóżmy, że*

1. Θ *jest zbiorem otwartym w* R^k.
2. *Odwzorowanie* $\theta \to s_\theta = \sqrt{p_\theta}$ *jest ciągle różniczkowalne po* θ, *dla każdego* $x \in \mathcal{X}$.
3. *Elementy macierzy*

$$I_\theta = \int\limits_{\mathcal{X}} \left(\frac{\dot{p}_\theta}{p_\theta}\right) \left(\frac{\dot{p}_\theta}{p_\theta}\right)^T p_\theta \, d\mu, \quad \textit{gdzie} \quad \dot{p}_\theta = \frac{\partial}{\partial \theta} p_\theta$$

są dobrze określone i ciągłe po θ.

Wówczas odwzorowanie $\theta \mapsto \sqrt{p_\theta}$ *jest różniczkowalne w sensie średniokwadratowym w punkcie* θ *i spełnia (8) z* $\dot{\ell}_\theta = \dot{p}_\theta / p_\theta$.

UWAGA 1. Jeśli $k = 1$, to założenie 2 w lemacie 1 można zastąpić słabszym warunkiem: w otoczeniu θ, dla wszystkich x, $p_\theta(x)$ jest absolutnie ciągła względem θ. W szczególności, dla modelu $p_\theta(x) = f(x - \theta)$, z f będącą absolutnie ciągłą funkcją swego argumentu i taką, że $\int (f'/f)^2 f d\mu < \infty$, uzyskujemy średniokwadratową różniczkowalność z

$$\dot{\ell}_\theta(x) = -\frac{f'(x-\theta)}{f(x-\theta)}.$$

Przykładem funkcji f spełniającej założenia uwagi 1, a nie spełniającej założenia 2 z lematu 1, jest funkcja $f(x) = 0.5 \exp\{-|x|\}$.

2.2. *Regularność estymatorów.* Przykład Hodgesa pokazał, że warto myśleć o zagwarantowaniu stabilnego zachowania estymatorów w otoczeniu estymowanej wartości θ. W metodologii zaproponowanej przez Hájeka i Le Cama zrealizowano ten postulat poprzez kontrolę zachowania się estymatora w ściągających się wraz z n otoczeniach θ.

Mając na względzie rozmaite zastosowania, warto rozważyć od razu ogólniejszy problem estymacji składowych funkcji $\Psi : \Theta \to R^m$, $m \le k$, $\Theta \subset R^k$. Przypuśćmy, że dla pewnego estymatora T_n parametru $\Psi(\theta)$, przy każdym ustalonym $h \in R^k$, dla obserwacji o rozkładzie $P_{\theta+h/\sqrt{n}}$, zachodzi

$$\sqrt{n}\left[T_n - \Psi\left(\theta + \frac{h}{\sqrt{n}}\right)\right] \xrightarrow{\mathcal{D}} L_{\theta,h}.$$

W dalszej części niniejszego opracowania ograniczymy rozważania do estymatorów T_n, dla których $L_{\theta,h} = L_\theta$ (według rozkładu) dla każdego h. Własność ta oznacza, że T_n stabilizuje się w sposób lokalnie jednostajny. Takie estymatory nazwiemy regularnymi. Formalizuje to następująca definicja.

Definicja 4. Ciąg estymatorów $\{T_n\}$ parametru $\Psi(\theta)$ nazywamy regularnym w punkcie θ, jeśli dla każdego $h \in R^k$, przy $P_{\theta+h/\sqrt{n}}$, zachodzi

$$\sqrt{n}\left[T_n - \Psi\left(\theta + \frac{h}{\sqrt{n}}\right)\right] \xrightarrow{\mathcal{D}} L_\theta, \quad L_\theta \sim G_\theta,$$

gdzie G_θ jest jakimś rozkładem, który zależy od θ, a nie zależy od h.

Wiele estymatorów spełnia powyższy wymóg regularności. Istnieją też dobre i złe estymatory, które regularne nie są. Jako przykład mogą służyć estymatory Jamesa–Steina i Hodgesa. Oba nie są regularne w 0 z odpowiednio wymiarowej przestrzeni.

2.3. *Formalizacja pojęcia asymptotycznej optymalności.* Zacznijmy od wprowadzenia dodatkowego oznaczenia. Niech $\Psi = (\Psi_1, ..., \Psi_m)$ będzie różniczkowalną funkcją odwzorowującą $\Theta \subset R^k$ w R^m, $m \le k$. Symbolem $\dot{\Psi}_\theta$ będziemy oznaczać macierz $m \times k$ o elementach postaci $\partial\Psi_i/\partial\theta_j$, $1 \le i \le m$, $1 \le j \le k$.

Wprowadzenie pojęcia asymptotycznej optymalności poprzedzimy kluczowym twierdzeniem Hájeka (1970) o splocie i wnioskami z niego wypływającymi.

Twierdzenie 3. *Załóżmy, że Θ jest zbiorem otwartym w R^k, a model $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ jest różniczkowalny w sensie średniokwadratowym w punkcie θ. Ponadto załóżmy, że macierz I_θ jest nieosobliwa, a funkcja $\Psi : \Theta \to R^m$, $m \le k$, jest różniczkowalna w punkcie θ. Niech $\{T_n\}$ będzie ciągiem estymatorów parametrów $\Psi(\theta)$ regularnym w θ, z rozkładem granicznym G_θ.*

Wówczas istnieje miara probabilistyczna M_θ taka, że

$$G_\theta = N\left(0, \dot{\Psi}_\theta I_\theta^{-1} \dot{\Psi}_\theta^T\right) * M_\theta. \tag{10}$$

W szczególności, jeśli G_θ ma macierz kowariancji Σ_θ, to macierz $\Sigma_\theta - \dot{\Psi}_\theta I_\theta^{-1} \dot{\Psi}_\theta^T$ jest nieujemnie określona.

Warunek (10) można zinterpretować następująco. Przy $P_{\theta+h/\sqrt{n}}$ zachodzi $\sqrt{n}\,[T_n - \Psi(\theta + h/\sqrt{n})] \xrightarrow{\mathcal{D}} Z_\theta + S_\theta$, gdzie Z_θ i S_θ są niezależne, $Z_\theta \sim N\left(0, \dot{\Psi}_\theta I_\theta^{-1} \dot{\Psi}_\theta^T\right)$, a $S_\theta \sim M_\theta$. Stąd wynika, że regularny ciąg estymatorów w modelu regularnym jest (asymptotycznie) najmniej rozproszony, jeśli ma asymptotyczny rozkład $N\left(0, \dot{\Psi}_\theta I_\theta^{-1} \dot{\Psi}_\theta^T\right)$.

Podobną interpretację twierdzenia 3 uzyskuje się poprzez zastosowanie lematu Andersona (1955). Lemat poprzedzamy pomocniczą definicją.

Definicja 5. Funkcję $l : R^m \to R_+$ nazywamy czaszokształtną, jeśli $l(x) = l(-x)$ oraz dla każdego $c \geq 0$ zbiór $\{x : l(x) \leq c\}$ jest wypukły.

Lemat 2. *Dla każdej czaszokształtnej funkcji l na R^m, każdej miary probabilistycznej M i każdej macierzy kowariancji Σ wymiaru $m \times m$ zachodzi*

$$\int_{R^m} l \, d[N(0,\Sigma) * M] \geq \int_{R^m} l \, dN(0,\Sigma).$$

Z twierdzenia o splocie, lematu Andersona i własności słabej zbieżności wynika następujące spostrzeżenie.

Wniosek 2. *Niech $Z_\theta \sim N\left(0, \dot{\Psi}_\theta I_\theta^{-1} \dot{\Psi}_\theta^T\right)$. Przy założeniach twierdzenia 3, dla każdej czaszokształtnej funkcji l na R^m zachodzi*

$$\liminf_{n\to\infty} E_{\theta+h/\sqrt{n}} \, l\left(\sqrt{n}\left[T_n - \Psi\left(\theta + \frac{h}{\sqrt{n}}\right)\right]\right) \geq El(Z_\theta).$$

W szczególności

$$\liminf_{n\to\infty} \; E_\theta \, l\left(\sqrt{n}\,[T_n - \Psi(\theta)]\right) \geq El(Z_\theta).$$

Twierdzenie 3 i wniosek 2 pokazują, że estymator o rozkładzie granicznym $N\left(0, \dot{\Psi}_\theta I_\theta^{-1} \dot{\Psi}_\theta^T\right)$ jest najbardziej skoncentrowany wokół $\Psi(\theta)$. Powyższe wyniki motywują następującą definicję.

Definicja 6. Rozważamy średniokwadratowo różniczkowalny model $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ z nieosobliwą macierzą informacji I_θ i problem estymacji różniczkowalnej funkcji $\Psi(\theta)$. Mówimy, że ciąg regularnych estymatorów $\{T_n\}$ parametru $\Psi(\theta)$ jest asymptotycznie optymalny dla estymacji $\Psi(\theta)$ w punkcie θ, jeśli przy P_θ zachodzi

$$\sqrt{n}[T_n - \Psi(\theta)] \xrightarrow{\mathcal{D}} N\left(0, \dot{\Psi}_\theta I_\theta^{-1} \dot{\Psi}_\theta^T\right).$$

Innego argumentu na to, że w regularnych przypadkach rozkład $N\left(0, \dot{\Psi}_\theta I_\theta^{-1} \dot{\Psi}_\theta^T\right)$ jest najlepszym osiągalnym wynikiem, dostarcza odpowiednie twierdzenie minimaksowe. Pierwszy ogólny rezultat tego typu podał Hájek w 1972 r. Dla ilustracji tego podejścia przytaczamy w miarę prosty wariant twierdzenia udowodniony przez van der Vaarta (2000).

Twierdzenie 4. *Załóżmy, że model $\mathcal{P} = \{P_\theta : \theta \in \Theta \subset R^k\}$ jest średniokwadratowo różniczkowalny w punkcie θ z nieosobliwą macierzą informacji*

I_θ. Niech Ψ będzie różniczkowalna w θ i niech $\{T_n\}$ będzie dowolnym ciągiem estymatorów. Wówczas dla dowolnej czaszokształtnej funkcji straty l zachodzi

$$\sup_{\mathcal{S}} \liminf_{n\to\infty} \sup_{h\in\mathcal{S}} E_{\theta+h/\sqrt{n}}\, l\left(\sqrt{n}\left[T_n - \Psi(\theta + \frac{h}{\sqrt{n}})\right]\right) \geq \int l\, dN\left(0, \dot{\Psi}_\theta I_\theta^{-1} \dot{\Psi}_\theta^T\right),$$

gdzie pierwsze supremum jest brane po wszystkich skończonych podzbiorach $\mathcal{S}$ z R^k.

2.4. *Charakteryzacja estymatorów asymptotycznie optymalnych.* Poniższe twierdzenie gra kluczową rolę w konstrukcji estymatorów asymptotycznie optymalnych. Jego sformułowanie pochodzi z ksiażki van der Vaarta (2000).

TWIERDZENIE 5. *Załóżmy, że Θ jest zbiorem otwartym, a model $\mathcal{P} = \{P_\theta : \theta \in \Theta \subset R^k\}$ jest średniokwadratowo różniczkowalny w punkcie θ z nieosobliwą macierzą informacji I_θ. Niech Ψ będzie różniczkowalna w θ i niech $\{T_n\}$ będzie ciągiem estymatorów $\Psi(\theta)$ takim, że*

$$\sqrt{n}[T_n - \Psi(\theta)] = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \dot{\Psi}_\theta I_\theta^{-1} \dot{\ell}_\theta(X_i) + o_{P_\theta}(1). \tag{11}$$

Wówczas $\{T_n\}$ jest regularny i asymptotycznie optymalny dla estymacji $\Psi(\theta)$ w punkcie θ. Ponadto, każdy asymptotycznie optymalny i regularny ciąg estymatorów parametru $\Psi(\theta)$ ma reprezentację (11).

Dla innego wysłowienia warunku (11) przypomnimy dwa standardowe pojęcia.

Mówimy, że estymator T_n parametru $\Psi(\theta)$ w modelu $\mathcal{P}$ jest asymptotycznie liniowy z funkcją wpływu h_θ, $\int h_\theta dP_\theta = 0$, $\int h_\theta^2 dP_\theta < \infty$, jeśli

$$\sqrt{n}[T_n - \Psi(\theta)] = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} h_\theta(X_i) + o_{P_\theta}(1). \tag{12}$$

Funkcję $\dot{\Psi}_\theta I_\theta^{-1} \dot{\ell}_\theta(\bullet)$ nazywa się efektywną funkcją wpływu dla parametru $\Psi(\theta)$.

Tezę twierdzenia 5 można więc streścić następująco: każdy asymptotycznie optymalny i regularny ciąg estymatorów musi być asymptotycznie liniowy z efektywną funkcją wpływu.

2.5. *Konstrukcja klasy estymatorów asymptotycznie optymalnych wektora θ.* Generalnie, charakteryzacja klasy estymatorów asymptotycznie optymalnych wydaje się być problemem łatwiejszym niż konstruowanie takich estymatorów. Oczywiście, bardzo dużo w tej tematyce już zrobiono. Niestety, czasami wiedza ta nie jest łatwo dostępna. Poniżej podajemy pewną konstrukcję klasy estymatorów asymptotycznie optymalnych dla problemu estymacji wektora θ. Poprawność takiej konstrukcji jest naszkicowana w rozdz. 2.5 książki Bickela i innych (1993). Praca Schicka (2001) dostarcza precyzyjnych narzędzi pozwalających udowodnić poniższe twierdzenie w sposób samodzielny. Podana konstrukcja estymatorów asymptotycznie optymalnych naśladuje estymatory jednokrokowe (por. (3)) i wykorzystuje trick Le Cama (1956), zwany dyskretyzacją. Opis konstrukcji zaczniemy od przypomnienia na czym polega dyskretyzacja.

Rozważamy model $\mathcal{P} = \{P_\theta : \theta \in \Theta \subset R^k\}$, niezależne zmienne losowe $X_1, ..., X_n$ o wartościach w $(\mathcal{X}, \mathcal{A})$ i rozkładzie P_θ. Dla danej realizacji próby $x_1, ..., x_n$ kładziemy $\mathbf{x} = (x_1, ..., x_n)$. Niech $\widetilde{\theta}_n$ będzie jakimś $\sqrt{n}$-zgodnym estymatorem θ. Podzielmy Θ na kostki o boku $c/\sqrt{n}$, gdzie c jest dowolnym ustalonym wektorem w R^k. Zdyskretyzowana wersja θ_n^* estymatora $\widetilde{\theta}_n$ jest zdefiniowana następująco: dla danej realizacji $\mathbf{x}$, $\theta_n^* = \theta_n^*(\mathbf{x})$ jest środkiem kostki, do której należy $\widetilde{\theta}_n(\mathbf{x})$. Dla wartości $\widetilde{\theta}_n(\mathbf{x})$ leżących na brzegach kostek przyjmuje się jakąś dodatkową regułę określenia θ_n^*.

Estymator θ_n^* ma dwie istotne własności: jest $\sqrt{n}$-zgodny oraz dla każdego $M > 0$ na zbiorze $\{\mathbf{x} : \sqrt{n}||\theta_n^* - \theta|| \le M\}$ przyjmuje skończoną liczbę wartości, która zależy od c i M, ale nie zależy od n.

Przypomnijmy teraz, że przy klasycznych założeniach regularności (i)-(vi) i ciągłości I_θ estymator δ_n^0 postaci

$$\delta_n^0 = \widetilde{\theta}_n + \frac{1}{n}\sum_{i=1}^n I_{\widetilde{\theta}_n}^{-1}\dot{\ell}_{\widetilde{\theta}_n}(X_i),$$

był v-efektywny dla estymacji θ, o ile $\widetilde{\theta}_n$ był $\sqrt{n}$-zgodny. Okazuje się, że przy odpowiednich założeniach regularności, estymator postaci

$$\delta_n^* = \theta_n^* + \frac{1}{n}\sum_{i=1}^n I_{\theta_n^*}^{-1}\dot{\ell}_{\theta_n^*}(X_i), \tag{13}$$

gdzie θ_n^* jest zdyskretyzowaną wersją $\sqrt{n}$-zgodnego estymatora parametru θ, jest asymptotycznie optymalny. Stosowne warunki regularności podaje poniższe twierdzenie.

TWIERDZENIE 6. *Załóżmy, że Θ jest zbiorem otwartym w R^k, a rodzina $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$, dominowana przez σ-skończoną miarę μ, jest różniczkowalna w sensie średniokwadratowym w otoczeniu θ z pochodną $\dot{\ell}_\theta$ w punkcie θ. Załóżmy, że I_θ jest nieosobliwa, a $\dot{\ell}_\theta$ jest ciągła w sensie Hellingera w punkcie θ, to znaczy*

$$\lim_{\tau \to \theta} \int_{\mathcal{X}} ||\dot{\ell}_\tau\sqrt{p_\tau} - \dot{\ell}_\theta\sqrt{p_\theta}||d\mu = 0,$$

gdzie $p_\theta = dP_\theta/d\mu$. Niech $\widetilde{\theta}_n$ będzie $\sqrt{n}$-zgodnym estymatorem θ, a θ_n^* jego dyskretną wersją. Wówczas estymator δ_n^*, dany wzorem (13), jest asymptotycznie optymalny.

Oczywiście, klasa estymatorów asymptotycznie optymalnych jest dużo bogatsza niż (13). Przy odpowiednich założeniach, L-, M-, R-estymatory, estymatory bayesowskie i estymatory minimalnej odległości są asymptotycznie optymalne. Przykłady takich wyników można znaleźć, np. w monografiach Bickela i innych (1993), Hubera (1981) oraz Ibragimowa i Hasminskiego (1981). Ponadto, warto zanotować, że przedstawiona teoria nie obejmuje wielu dobrych estymatorów. Powody są dwa: albo estymatory nie są regularne (jak np. estymator Jamesa-Steina) albo są regularne, ale nie są asymptotycznie normalne (jak np. estymator środka symetrii zaproponowany przez Bickela i Hodgesa 1967).

Pisząc w wielkim skrócie, większość dowodów optymalności sprowadza się do sprawdzenia czy estymator spełnia warunek (11). Do analizy tego warunku wrócimy w następnym rozdziale, w którym rozważamy ważny przypadek $\Psi(\theta_1, ..., \theta_k) = (\theta_1, ..., \theta_m)$, $m < k$. Inaczej mówiąc, jest to przypadek estymacji $(\theta_1, ..., \theta_m)$ przy parametrach zakłócających $(\theta_{m+1}, ..., \theta_k)$. Ten przypadek jest istotny sam w sobie, ale również stanowi wzorzec dla rozwiązań bardziej złożonych problemów semiparametrycznych i nieparametrycznych.

3. Asymptotyczna optymalność estymatorów wektora parametrów $\theta \in R^m$ przy parametrach zakłócających $\eta \in R^{k-m}$, $m < k$. Twierdzenie 5 podaje opis asymptotycznie optymalnych estymatorów dla estymacji składowych dowolnej funkcji $\Psi : \Theta \subset R^k \to R^m$, $m \le k$. Teraz rozważymy wyżej wspomniany szczególny przypadek funkcji Ψ. Zachowanie symbolu θ dla parametru estymowanego wymaga wprowadzenia nowego oznaczenia. Niech $\gamma = (\theta^T, \eta^T)^T$, gdzie $\theta = (\theta_1, ..., \theta_m)$, $\eta = (\theta_{m+1}, ..., \theta_k)$, $m < k$, $\Gamma = \Theta \times H$, $\Theta \subset R^m$, $H \subset R^{k-m}$ i rozważmy funkcję $\Psi : \Gamma \to R^m$ daną wzorem

$$\Psi(\gamma) = \theta. \tag{14}$$

Przepisując tezę twierdzenia 5 dla funkcji $\Psi(\gamma)$, otrzymujemy warunek

$$\sqrt{n}[T_n - \Psi(\gamma)] = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \dot{\Psi}_\gamma I_\gamma^{-1} \dot{\ell}_\gamma(X_i) + o_{P_\gamma}(1).$$

Używając terminologii z rozdziału 2 i koncentrując uwagę na funkcji (14), można powiedzieć, że $\dot{\Psi}_\gamma I_\gamma^{-1} \dot{\ell}_\gamma(\bullet)$ jest efektywną funkcją wpływu dla estymowanego parametru θ w obecności parametru zakłócającego η. Krótko będziemy tę funkcję oznaczać w następujący sposób

$$\widetilde{\ell}_\theta(\bullet) = \widetilde{\ell}_\theta(\bullet; \eta) = \dot{\Psi}_\gamma I_\gamma^{-1} \dot{\ell}_\gamma(\bullet). \tag{15}$$

Dla uzyskania jawnego wzoru na $\widetilde{\ell}_\theta(\bullet)$ dla funkcji (14) wprowadzimy pomocnicze oznaczenia.

$$\dot{\ell}_\gamma = \begin{pmatrix} \dot{\ell}_\theta \\ \dot{\ell}_\eta \end{pmatrix}, \quad \widetilde{\ell}_\gamma = \begin{pmatrix} \widetilde{\ell}_\theta \\ \widetilde{\ell}_\eta \end{pmatrix}, \quad I_\gamma = \begin{pmatrix} I_{\theta\theta} & I_{\theta\eta} \\ I_{\eta\theta} & I_{\eta\eta} \end{pmatrix}, \quad I_\gamma^{-1} = \begin{pmatrix} I^{\theta\theta} & I^{\theta\eta} \\ I^{\eta\theta} & I^{\eta\eta} \end{pmatrix}.$$

Po elementarnych rachunkach otrzymujemy

$$\widetilde{\ell}_\theta(\bullet) = \dot{\Psi}_\gamma I_\gamma^{-1} \dot{\ell}_\gamma(\bullet) = (I_\theta^*)^{-1} \ell_\theta^*(\bullet), \tag{16}$$

gdzie

$$\begin{aligned} \ell_\theta^*(\bullet) &= \dot{\ell}_\theta(\bullet) - I_{\theta\eta} I_{\eta\eta}^{-1} \dot{\ell}_\eta(\bullet), \\ I_\theta^* &= \left(I^{\theta\theta}\right)^{-1} = I_{\theta\theta} - I_{\theta\eta} I_{\eta\eta}^{-1} I_{\eta\theta} = E_\gamma \ell_\theta^* (\ell_\theta^*)^T. \end{aligned} \tag{17}$$

Powyższe rozważania motywują następującą definicję.

Definicja 7 . Funkcję ℓ_θ^* określoną wzorem (17) nazywamy efektywną funkcją wynikową dla θ w modelu $\mathcal{P} = \{P_{\theta,\eta} : \theta \in \Theta, \eta \in H\}$. Macierz I_θ^* zdefiniowaną także w (17) nazywamy macierzą informacji dla parametru θ.

Warto zanotować prostą i użyteczną interpretację geometryczną efektywnej funkcji wynikowej ℓ_θ^* jako residuum rzutu pierwszej cześci funkcji wynikowej $\dot{\ell}_\gamma$ (odpowiadającej szacowanemu parametrowi θ) na przestrzeń liniową rozpiętą przez składowe drugiej części $\dot{\ell}_\gamma$ (odpowiadającej parametrom zakłócającym η, por. (17)).

Wykorzystując powyższy wniosek z twierdzenia 5, postać efektywnej funkcji wpływu (16) oraz wyniki Schicka (2001), udowodniono następujący analogon twierdzenia 6.

Twierdzenie 7. *Załóżmy, że model* $\mathcal{P} = \{P_\gamma : \gamma \in \Theta \times H \subset R^m \times R^{k-m}\}$, Θ *i* H *otwarte, jest średniokwadratowo różniczkowalny z pochodną* $\dot{\ell}_\gamma$ *i nieosobliwą macierzą informacji* $I_\gamma = E_\gamma \dot{\ell}_\gamma (\dot{\ell}_\gamma)^T$.

Niech $\ell_\theta^*(x) = \ell_\theta^*(x;\eta)$ *będzie efektywną funkcją wynikową dla* θ, *a* $I_\theta^* = I_\theta^*(\eta)$ *niech oznacza odpowiadającą jej macierz informacji. Zakładamy, że* $\ell_\theta^*(x;\eta)$ *jest ciągła w sensie Hellingera względem obu zmiennych* θ *i* η.

Niech $\widetilde{\theta}_n$ *i* $\widetilde{\eta}_n$ *będą* $\sqrt{n}$*-zgodnymi (przy* P_γ*) estymatorami* θ *i* η *i niech* θ_n^* *oraz* η_n^* *oznaczają ich dyskretne wersje.*

Przy powyższych założeniach, estymator

$$\delta_n^* = \theta_n^* + \frac{1}{n}\Big[I_{\theta_n^*}(\eta_n^*)\Big]^{-1} \sum_{i=1}^n \ell_{\theta_n^*}^*(X_i;\eta_n^*)$$

jest asymptotycznie optymalnym estymatorem θ *w modelu* $\mathcal{P}$.

Twierdzenie 5 ilustruje rolę efektywnej funkcji wpływu w optymalnej estymacji. Zanotujmy, że efektywna funkcja wpływu w problemie estymacji z parametrami zakłócającymi pojawiła się po raz pierwszy w pracy Bartletta

(1953). Neyman (1954, 1959) odkrył kluczową rolę efektywnej funkcji wynikowej w problemach testowania. Obaj autorzy użyli $\widetilde{\ell}_\theta(\bullet)$ do wyeliminowania wpływu parametrów zakłócających na rozkład asymptotyczny estymatorów i statystyk testowych. W szczególności praca Neymana (1959) dotyczyła testowania hipotez o jednowymiarowym parametrze θ przy nieznanym wektorze parametrów zakłócających η. Bühler i Puri (1966) klasycznymi metodami uogólnili wyniki Neymana na ogólny przypadek $\theta \in R^m$, $\eta \in R^{m-k}$, używając oznaczeń z tego rozdziału naszego artykułu. Poniższe twierdzenie 8 pokazuje, że wyniki Schicka mogą być z powodzeniem wykorzystane w testowaniu i można łatwo uzyskać elegancki analogon wyniku Bühlera i Puriego (1966).

4. Testowanie w modelu $\mathcal{P} = \{P_{\theta,\eta} : \theta \in \Theta \subset R^m,\ \eta \in H \subset R^{k-m}\}$. Rozważmy problem testowania hipotezy

$$H_0 : \theta = \theta_0,\ \eta$$

przeciwko alternatywie

$$A : \theta \neq \theta_0,\ \eta$$

i zmienną losową

$$W_n(\theta_0, \eta) = \left[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \ell^*_{\theta_0}(X_i;\eta)\right]^T \left[I^*_{\theta_0}(\eta)\right]^{-1} \left[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \ell^*_{\theta_0}(X_i;\eta)\right], \tag{18}$$

gdzie ℓ^*_θ i I^*_θ są zdefiniowane wzorem (17). Łatwo sprawdzić, że przy prawdziwości H_0, $W_n(\theta_0,\eta) \xrightarrow{\mathcal{D}} \chi^2_m$, gdzie χ^2_m oznacza zmienną losową o centralnym rozkładzie chi-kwardat z m stopniami swobody.

Zmienna (18) jest prototypem (efektywnej) statystyki wynikowej. Bühler i Puri (1966), przy szeregu analitycznych założeń typu Craméra (rozbudowany wariant warunków (i)-(vi) z rozdz. 1) udowodnili, że jeśli w (18) zastąpimy nieznany parametr η jakimś $\sqrt{n}$-zgodnym estymatorem $\widehat{\eta}$, to rozkład graniczny $W_n(\theta_0,\widehat{\eta})$ będzie taki sam jak rozkład graniczny zmiennej (18). Poniższe twierdzenie pokazuje, że, przy dużo słabszych założeniach, można uzyskać taki sam efekt.

TWIERDZENIE 8. *Załóżmy, że model $\mathcal{P} = \{P_{\theta,\eta} : \theta \in \Theta \subset R^m,\ \eta \in H \subset R^{k-m}\}$, Θ i H otwarte, jest średniokwadratowo różniczkowalny z pochodną $\dot{\ell}_\gamma$ i nieosobliwą macierzą informacji $I_{\theta_0,\eta}$ oraz efektywną funkcją wynikową $\ell^*_{\theta_0}(x;\eta)$ ciągłą względem η w sensie Hellingera.*

Niech $\widehat{\eta}$ będzie $\sqrt{n}$-zgodnym (przy $P_{\theta_0,\eta}$) estymatorem η, a η^*_n niech będzie jego dyskretną wersją. Niech $\widehat{I^*_{\theta_0}}$ będzie jakimś dodatnio określonym i zgodnym (przy $P_{\theta_0,\eta}$) estymatorem $I^*_{\theta_0}(\eta)$.

Wówczas, przy prawdziwości hipotezy $H_0 : \theta = \theta_0$, *zachodzi*

$$\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\ell^*_{\theta_0}(X_i;\eta^*_n)\right]^T \left[\widehat{I^*_{\theta_0}}\right]^{-1}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\ell^*_{\theta_0}(X_i;\eta^*_n)\right] \xrightarrow{\mathcal{D}} \chi^2_m.$$

5. Uwagi

5.1. *Estymacja i testowanie.* Wyniki z rozdziałów 2–3 można przenieść na bardzo ogólne modele semiparametryczne i nieparametryczne. Monografie Ibragimowa i Hasminskiego (1981) oraz Bickela i innych (1993) zawierają bardzo obszerny przegląd takich uogólnień. Praca Stone'a (1975) może być rekomendowana jako ilustracja rozwiązania jednego z najprostszych problemów semiparametrycznej estymacji. Z kolei praca Choi i inni (1996) wskazała na możliwość stosownych uogólnień konstrukcji Neymana (1954, 1959). Ostatnie lata przyniosły dalszy postęp w omawianej dziedzinie. W szczególności, sporo wysiłku poświęcono modelom regresji (por. np. Schick 1997 oraz Klassen i Putter 2005). Na przykładzie pewnego problemu testowania o modelu regresji, Inglot i Ledwina (2006a,b) zilustrowali potencjał tkwiący w wykorzystaniu efektywnych funkcji wynikowych.

Przejście od przypadku parametrycznego, który pokrótce przedstawiono w rozdziałach 2-4, do bardziej złożonych modeli wiąże się z oczywistą zmianą przestrzeni parametrów z parametrów liczbowych na funkcyjne. To z kolei powoduje konieczność zastosowania adekwatnych metod różniczkowania i rzutowania. Wprowadzenie przez Hájeka i Le Cama różniczkowalności średniokwadratowej było bardzo pomocne w naturalnym rozwiązaniu tego problemu. Rozdział 6 pracy Inglota i Ledwiny (2006b) zawiera prostą i czytelną interpretację takiego rozszerzenia.

Ostatni rozdział niniejszej pracy poświęcimy kilku uwagom o związku regularności modeli i estymatorów z prawidłowym działaniem metody bootstrap. Jest to jeszcze jedna ilustracja konstatacji, że prawidłowe działanie statystycznych procedur wymaga pewnych stabilności w otoczeniu „modelowej" sytuacji.

5.2. *Bootstrap i regularność.* Twierdzenie 3 jest wariantem wyniku Hájeka, udowodnionego przy niemal minimalnych założeniach, potrzebnych do uzyskania tezy. Le Cam zamiast estymatorów regularnych w sensie definicji 4 rozpatrywał nieco węższą klasę estymatorów ekwiwariantnych, która ma lepsze statystyczne umocowanie. Poniżej podajemy definicję takiego estymatora. Tak jak w rozdziale 2.2 rozważamy problem estymacji składowych funkcji $\Psi : \Theta \to R^m$, $m \le k$, $\Theta \subset R^k$.

DEFINICJA 8. Ciąg estymatorów $\{T_n\}$ parametu $\Psi(\theta)$ nazywamy lokalnie asymptotycznie ekwiwariantnym w punkcie θ, jeśli dla każdego $h \in R^k$

i każdego $\{h_n\} \in R^k$ takiego, że $h_n \to h$, przy $P_{\theta+h_n/\sqrt{n}}$, zachodzi

$$\sqrt{n}\left[T_n - \Psi\left(\theta + \frac{h_n}{\sqrt{n}}\right)\right] \xrightarrow{\mathcal{D}} L_\theta, \quad L_\theta \sim G_\theta,$$

gdzie G_θ jest jakimś rozkładem, który zależy od θ, a nie zależy od h.

Bezpośrednio z definicji widać, że asymptotyczna lokalna ekwiwariantność może być nazwana lokalną asymptotyczną odpornością.

Beran (1997) (patrz również Beran 2003, rozdział 5) pokazał, że istnieje głęboki związek między lokalną asymptotyczną ekwiwariantnością, twierdzeniem o splocie (w adekwatnej wersji) i prawidłowym działaniem parametrycznego bootstrapu. Praktyczną implikacją jego rezultatów jest konkluzja, że parametryczny bootstrap nie może działać w punktach θ, w których replikowany estymator nie jest lokalnie asymptotycznie ekwiwariantny. W szczególności, punkty θ, w których ma miejsce superefektywność estymatora Hodgesa, Jamesa–Steina i innych tego typu estymatorów, wykluczają poprawne działanie tamże parametrycznego bootstrapu. Inną konkluzją Berana (1997) jest propozycja praktycznej metody diagnozowania poprawności działania metody bootstrap. Beran (1997) rozważa również zagadnienie prawidłowego działania nieparametrycznego bootstrapu. Praca Bednarskiego i Florczaka (1999) zawiera także podobne wyniki w tym ostatnim przypadku.

Dziękuję doktorowi Waldemarowi Wołyńskiemu za zaproszenie do wygłoszenia tego wykładu, profesorowi Ryszardowi Zielińskiemu za zachętę do przygotowania jego pisemnej wersji, doktorowi habilitowanemu Janowi Mielniczukowi za konstruktywne uwagi oraz recenzentowi za uważną lekturę tekstu.

Literatura

[1] T. W. Anderson, *The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities*, Proc. Amer. Math. Soc. 6 (1955), 170–176.

[2] R. R. Bahadur, *On Fisher's bound for asymptotic variances*, Ann. Math. Statist. 35 (1964), 1545–1552.

[3] M. S. Bartlett, *Approximate confidence intervals II. More than one unknown parameter*, Biometrika 40 (1953), 306–317.

[4] T. Bednarski, W. Florczak, *On local uniform bootstrap validity*, Statist. Neerl. 53 (1999), 111–121.

[5] R. Beran, *Diagnosing bootstrap success*, Ann. Inst. Statist. Math. 49 (1997), 1–24.

[6] R. Beran, *The impact of the bootstrap on statistical algorithms and theory*, Statist. Sci. 18 (2003), 175–184.

[7] P. J. Bickel, J. H. Hodges, *The asymptotic theory of Galton's test and a related simple estimate of location*, Ann. Math. Statist. 38 (1967), 73–89.

[8] P. J. Bickel, C. A. J. Klaassen, Y. Ritov, J. A. Wellner, *Efficient and Adaptive Estimation for Semiparametric Models*, Johns Hopkins University Press, Baltimore, 1993.

[9] W. J. Bühler, P. S. Puri, *On optimal asymptotic tests of composite hypotheses with several constraints*, Z. Wahrsch. verw. Gebiete 5 (1966), 71–88.
[10] G. Casella, T. J. Hwang, *Limit expression for the risk of James-Stein estimators*, Canad. J. Statist. 10 (1982), 305–309.
[11] S. Choi, W. J. Hall, A. Schick, *Asymptotically uniformly most powerful tests in parametric and semiparametric models*, Ann. Statist. 24 (1996), 841–861.
[12] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, Princeton, 1946 (przekład polski: H. Cramér, *Metody matematyczne w statystyce*, PWN, Warszawa 1958).
[13] J. L. Doob, *Probability and statistics*, Trans. Amer. Math. Soc. 36 (1934), 759–772.
[14] F. Y. Edgeworth, *On the probable errors of frequency constants*, J. Roy. Statist. Soc. 71 (1908), 381–397.
[15] R. A. Fisher, *On the mathematical foundations of theoretical statistics*, Philos. Trans. Roy. Soc. A 222 (1922), 309–365.
[16] R. A. Fisher, *Theory of statistical estimation*, Proc. Camb. Phil. Soc. 22 (1925), 700–725.
[17] J. Hájek, *A characterization of limiting distributions of regular estimates*, Z. Wahrsch. verw. Gebiete 14 (1970), 323–330.
[18] P. J. Huber, *Strict efficiency excludes superefficiency*, Ann. Math. Statist. 37 (1966), 1425.
[19] P. J. Huber, *Robust Statistics*, Wiley, New York, 1981.
[20] I. A. Ibragimow, R. Z. Hasminski, *Statistical Estimation: Asymptotic Theory*, Springer, New York, 1981.
[21] T. Inglot, T. Ledwina, *Data driven score tests for homoscedastic linear regression model: the construction and simulations*, w: Prague Stochastics 2006, M. Hušková, M. Janžura (red.), Matfyzpress, Prague, 2006a, 124–137.
[22] T. Inglot, T. Ledwina, *Data driven score tests for homoscedastic linear regression model: asymptotic results*, Probab. Math. Statist. 26 (the issue dedicated to the memory of K. Urbanik) (2006b), 41–61.
[23] W. James, C. Stein, *Estimation with quadratic loss*, w: Proc. Fourth Berkeley Symp. Math. Statist. Prob., J. Neyman (red.), Univ. California Press, Berkeley 1961, 361–380.
[24] J. Kiefer, J. Wolfowitz, *Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters*, Ann. Math. Statist. 27 (1956), 887–906.
[25] C. A. J. Klaassen, H. Putter, *Efficient estimation of Banach parameters in semiparametric models*, Ann. Statist. 33 (2005), 307–346.
[26] L. Le Cam, *On some asymptotic properties of maximum likelihood estimates and related Bayes estimates*, Univ. California Publ. Statist. 1 (1953), 277–330.
[27] L. Le Cam, *On the asymptotic theory of estimation and testing hypotheses*, w: Proc. Third Berkeley Symp. Math. Statist. Prob., J. Neyman (red.), Univ. California Press, Berkeley 1956, 129–156.
[28] L. Le Cam, *On the assumptions used to prove asymptotic normality of maximum likelihood estimates*, Ann. Math. Statist. 41 (1970), 802–828.
[29] E. L. Lehmann, *Theory of Point Estimation*, Wiley, New York, 1983 (przekład polski: E. Lehmann, *Teoria estymacji punktowej*, PWN, Warszawa, 1991).
[31] J. Neyman, *Sur une famille de tests asymptotiques des hypothèses statistiqes composées*, Trabajos de Estadistica 5 (1954), 161–168.
[32] J. Neyman, *Optimal asymptotic tests of composite statistical hypotheses*, w: The Harald Cramér Volume, U. Grenander (red.), Wiley, New York, 1959, 213–234.

[33] C. R. Rao, *Criteria of estimation in large samples*, Sankhyā 25 (1963), 189–206.
[34] A. Schick, *Efficient estimates in linear and nonlinear regression with heteroscedastic error*, J. Statist. Plann. Inference 58 (1997), 371–387.
[35] A. Schick, *On asymptotic differentiability of averages*, Statist. Probab. Lett. 51 (2001), 15–23.
[36] C. Stone, *Adaptive maximum likelihood estimators of a location parameter*, Ann. Statist. 3 (1975), 267–284.
[37] A. W. van der Vaart, *Asymptotic Statistics*, Cambridge Univ. Press, Cambridge, 2000.
[38] J. Wolfowitz, *Asymptotic efficiency of the maximum likelihood estimator*, Theory Probab. Appl. 10 (1965), 247–260.
[39] S. Zacks, *The Theory of Statistical Inference*, Wiley, New York, 1971.

Instytut Matematyczny PAN
Oddział Wrocław
ul. Kopernika 18, 51-617 Wrocław
E-mail: ledwina@impan.pan.wroc.pl

On asymptotic efficiency of estimators

Abstract. We present and discuss the notion of asymptotic efficiency of estimators as introduced by Hájek and Le Cam. We give also some general construction of a class of asymptotically efficient estimators of Euclidean parameters. Moreover, we briefly indicate some generalizations of the discussed ideas to the case of semiparametric models. We show also that technical results obtained in the asymptotic theory of efficient estimation can be successfully used in asymptotic theory of testing.

The selection of the material is highly subjective and to a little extent reflects complexity of several problems and range of results available in present-day literature. The paper is a shortened version of invited series of lectures presented at the Conference on Mathematical Statistics WISŁA 2005. Its main purpose is to show that classic approach to define efficiency was not satisfactory and to discuss how, for some class of problems, this question was solved in a natural and elegant way.

Key words: asymptotic efficiency, asymptotic optimality, influence function, superefficiency, score test.

(*wpłynęło 10 listopada 2006 r.*)