ROCZNIKI POLSKIEGO TOWARZYSTWA MATEMATYCZNEGO
Seria III: MATEMATYKA STOSOWANA XXI (1982)

M. BIEŃKIEWICZ (Łódź)

O metodzie „jackknife"

(Praca wpłynęła do Redakcji 14.04.1980)

0. Wstęp. Jackknife to w dosłownym tłumaczeniu scyzoryk, a więc ostre i praktyczne narzędzie. W statystyce nazwa jackknife określa metodę redukcji obciążenia estymatorów, zaproponowaną przez M. H. Quenouille'a w 1956 r. [6]. Zagadnieniami związanymi z jacknifem od szeregu już lat zajmują się liczni matematycy, m.in. A. F. Bissel, R. A. Ferguson, H. L. Gray, L. A. Jaeckel, R. G. Miller, W. R. Schucany i T. A. Watkins, ale sprawa możliwości zastosowań tej techniki jest nadal otwarta. Ponieważ w polskiej literaturze matematycznej brak jest omówień tej problematyki, poniższa praca ma charakter informacyjny, stanowiąc zarazem pewną próbę przedstawienia teorii dotyczącej jackknife'u.

Nazwę jackknife wprowadził J. W. Tukey, chociaż — jak wyżej wspomniano — nie on był twórcą metody. Zaproponowana przez Tukey'a nazwa sugeruje, że narzędzie to ma być „ostre i szybkie w użyciu". Mam pewne opory z przeniesieniem na nasz grunt dokładnego tłumaczenia tej nazwy, gdyż żadne z określeń: nóż, scyzoryk, czy po prostu ostrze nie brzmi dobrze ani jako nazwa estymatora, ani jako nazwa metody (nóż? nożowanie? estymator nożowy?). Poczekam więc chętnie w tej kwestii na sugestie innych, a sama, na użytek chociażby tej pracy, posługiwać się będę nazwami: jackknife i jackknifing.

Jackknife jest metodą lub techniką — aby nie użyć często spotykanego w literaturze określenia — matematyczną sztuczką, która pozwala zredukować obciążenie większości estymatorów i estymować ich wariancję (bez względu na rozkład próby). Uwalnia ona więc statystyków od często kwestionowanych założeń o rozkładzie, a przez to ma bezpośredni związek z zagadnieniem odporności. Oczywiście nieobciążoność estymatorów jest tylko jednym z aspektów dobrej estymacji i dlatego jackknife nie zawsze będzie dawał najlepsze rozwiązania, ale i w swoim założeniu nie ma to być narzędzie uniwersalne.

1. Jackknife — definicja ogólna. Niech $\{\mathscr{X}, \mathscr{A}, (P_\theta; \theta \in \Theta)\}$ będzie modelem statystycznym i niech zadanie polega na estymacji parametru θ indeksującego rodzinę rozkładów prawdopodobieństwa na $\mathscr{A}$. Interesować nas przy tym będą estymatory nieobciążone, a więc takie estymatory T, że $E_\theta T = \theta$. Estymatory parametru θ oznaczać będziemy przez $\hat{\theta}$ lub ewentualnie $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_i, \ldots$

DEFINICJA 1. Niech $\hat{\theta}_1$, $\hat{\theta}_2$ będą ustalonymi estymatorami parametru θ. *Uogólnionym jackknife'm* estymatorów $\hat{\theta}_1$, $\hat{\theta}_2$ nazywamy estymator $G(\hat{\theta}_1, \hat{\theta}_2)$ określony wzorem:

$$G(\hat{\theta}_1, \hat{\theta}_2) = \frac{\hat{\theta}_1 - R\hat{\theta}_2}{1-R}, \tag{1}$$

gdzie $R \neq 1$ jest dowolną, ale ustaloną liczbą rzeczywistą.

Zauważmy, że jeżeli wartości oczekiwane estymatorów $\hat{\theta}_1, \hat{\theta}_2$ wynoszą odpowiednio:

$$E\hat{\theta}_1 = \theta + b_1(n, \theta),$$

$$E\hat{\theta}_2 = \theta + b_2(n, \theta), \quad b_2(n, \theta) \neq 0,$$

to dla $R = b_1(n, \theta)/b_2(n, \theta)$ estymator $G(\hat{\theta}_1, \hat{\theta}_2)$ jest nieobciążony.

Często obciążenia rozważanych estymatorów mają pewną szczególną postać, np

$$E\hat{\theta} - \theta = \frac{1}{n} g_1(\theta) + \frac{1}{n} g_2(\theta) + \ldots \tag{2}$$

Ogólniej, załóżmy, że

$$E\hat{\theta}_1 = \theta + \sum_{i=1}^{\infty} b_{1i}(n, \theta),$$

$$E\hat{\theta}_2 = \theta + \sum_{i=1}^{\infty} b_{2i}(n, \theta).$$

Jeżeli teraz $b_{11}(n, \theta)/b_{12}(n, \theta) \neq 1$, a we wzorze (1) przyjmiemy $R = b_{11}(n, \theta)/$ $/b_{12}(n, \theta)$, to

$$EG(\hat{\theta}_1, \hat{\theta}_2) = \theta + \Big(\sum_{i=2}^{\infty} b_{1i}(n, \theta) - R \sum_{i=2}^{\infty} b_{2i}(n, \theta)\Big) / (1-R),$$

a więc obciążenie $G(\hat{\theta}_1, \hat{\theta}_2)$ jest mniejszego rzędu niż obciążenia każdego z estymatorów $\hat{\theta}_1, \hat{\theta}_2$.

2. Opis metody Quenouille'a. Definiując uogólniony jackknife $G(\hat{\theta}_1, \hat{\theta}_2)$ zakładaliśmy, że dane są estymatory (obciążone) $\hat{\theta}_1$ oraz $\hat{\theta}_2$. Obecnie zajmiemy się zagadnieniem wyboru tych dwóch estymatorów, przedstawiając pewną konstrukcję podaną przez Quenouille'a [6].

Niech $X_1, \ldots, X_n$ będzie ciągiem niezależnych zmiennych losowych o jednakowym rozkładzie („próbą losową") i niech $\hat{\theta} = \hat{\theta}_n = \hat{\theta}_n(X_1, \ldots, X_n)$ będzie estymatorem parametru θ, opartym na tej próbie losowej. Załóżmy przy tym, że estymator $\hat{\theta}_n$ jest określony dla każdego $n = 1, 2, \ldots$

Daną próbę dzielimy w sposób losowy na N rozłącznych podzbiorów o rozmiarze M każdy. Oczywiście, $n = N \cdot M$. Bez zmniejszania ogólności rozważań można przyjąć, że dokonany podział ma postać:

$$(X_1, \dots, X_M), (X_{M+1}, \dots, X_{2M}), \dots, (X_{(N-1)M+1}, \dots, X_{NM}).$$

Tworzymy nową $(N-1)M$-elementową próbę losową pomijając i-ty podzbiór i określamy estymator:

$$\hat{\theta}^i = \hat{\theta}_{n-M} = \hat{\theta}(X_1, \dots, X_{(i-1)M}, X_{iM+1}, \dots, X_n); \quad i = 1, \dots, N. \tag{3}$$

DEFINICJA 2. *Pseudowartościami jackknife* nazywamy estymatory

$$J_i(\hat{\theta}) = N\hat{\theta} - (N-1)\hat{\theta}^i \tag{4}$$

a *jackknife'm* $J(\hat{\theta})$ ich średnią arytmetyczną, tzn.

$$J(\hat{\theta}) = \frac{1}{N}\sum_{i=1}^{N} J_i(\hat{\theta}) = N\hat{\theta} - (N-1)\hat{\hat{\theta}}, \tag{5}$$

gdzie

$$\hat{\hat{\theta}} = \frac{1}{N}\sum_{i=1}^{N} \hat{\theta}^i.$$

Zauważmy, że tak skonstruowany estymator $J(\hat{\theta})$ jest szczególnym przypadkiem uogólnionego jackknife'u postaci (1). Istotnie, przyjmując $\hat{\theta}_1 = \hat{\theta}, \hat{\theta}_2 = \hat{\hat{\theta}}$ oraz $R = R(N) = (N-1)/N$, mamy $G(\hat{\theta}_1, \hat{\theta}_2) = J(\hat{\theta})$. W dalszym ciągu rozważań wygodnie będzie utożsamiać estymator $G(\hat{\theta}_1, \hat{\theta}_2)$ o parametrze $R = (N-1)/N$ z estymatorem $J(\hat{\theta})$.

Aby wyznaczyć obciążenie tak określonego jackknife'u załóżmy, że estymator $\hat{\theta}$ jest obciążony i że wartość oczekiwana tego estymatora ma postać

$$E\hat{\theta} = \theta + \sum_{i=1}^{\infty} \frac{a_i}{n^i} = \theta + \sum_{i=1}^{\infty} \frac{a_i}{N^i M^i},$$

gdzie a_i mogą być funkcjami parametru $\hat{\theta}$, ale nie liczności próby n. Okazuje się, że sytuacja taka ma miejsce w przypadku większości praktycznie używanych estymatorów. Otrzymujemy

$$E\hat{\hat{\theta}} = \theta + \sum_{i=1}^{\infty} \frac{a_i}{(n-M)^i} = \theta + \sum_{i=1}^{\infty} \frac{a_i}{(N-1)^i M^i},$$

a stąd

$$EJ(\hat{\theta}) = \theta - \frac{a_2}{M^2 N(N-1)} - \frac{a_3}{M^3 N^2 (N-1)^2} - \dots =$$

$$= \theta - \frac{a_2}{n(n-M)} - \frac{a_3}{n^2(n-M)^2} - \dots$$

Obciążenie estymatora $J(\hat{\theta})$ jest więc rzędu n^{-2}, podczas gdy obciążenie $\hat{\theta}$, przy $a_1 \neq 0$, było rzędu n^{-1}. Ponadto, gdy $a_2 = a_3 = \ldots = 0$, jackknife $J(\hat{\theta})$ jest nieobciążony.

3. Przykłady.

PRZYKŁAD 1. *Schemat Bernoullie'go.* Wykonujemy n niezależnych prób, z których każda z prawdopodobieństwem p kończy się sukcesem lub z prawdopodobieństwem $1-p$ — porażką. Niech X_i będzie zmienną losową przyjmującą wartość 1, gdy i-ta próba kończy się sukcesem i wartość 0, gdy i-ta próba kończy się porażką. Łączna liczba sukcesów

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

określa zmienną losową, która ma rozkład dwumianowy:

$$f(x) = \binom{n}{x} p^x(1-p)^{n-x}, \qquad x = 0, 1, \ldots, n.$$

Niech zadanie polega na estymacji p^2. Ponieważ X/n jest nieobciążonym estymatorem parametru p, to rozważmy

$$\hat{\theta} = (X/n)^2.$$

Jest to estymator obciążony, gdyż

$$E\hat{\theta} = p^2 + \frac{1}{n}(p-p^2).$$

Dla estymatorów $\hat{\theta}^i$ (wzór (3)) otrzymujemy wartości:

$$((X-1)/(n-1))^2 \quad \text{lub} \quad (X/(n-1))^2,$$

w zależności od tego czy pominięta próba była sukcesem czy porażką. Średnia $\hat{\hat{\theta}}$ jest równa:

$$\hat{\hat{\theta}} = \frac{1}{n}\left[X\left(\frac{X-1}{n-1}\right)^2 + (n-X)\left(\frac{X}{n-1}\right)\right]^2 = \frac{(n-2)X^2+X}{n(n-1)^2},$$

a stąd, na mocy wzoru (5)

$$J(\hat{\theta}) = \frac{X(X-1)}{n(n-1)}.$$

Otrzymana modyfikacja $J(\hat{\theta})$ estymatora $\hat{\theta}$ jest estymatorem nieobciążonym, gdyż $EJ(\hat{\theta}) = p^2$.

PRZYKŁAD 2. *Rozkład równomierny.* Niech $X_1, \ldots, X_n$ będzie próbą losową z rozkładu równomiernego na przedziale $(0, \theta)$ i niech $X_{(1)}, \ldots, X_{(n)}$ będą statystykami pozycyjnymi z tej próby. Rozpatrzmy następujący estymator parametru θ:

$$\hat{\theta} = \max(X_1, \ldots, X_n) = X_{(n)}.$$

Jest to estymator obciążony, gdyż $E\hat{\theta} = \theta - \theta/(n+1)$. Jest oczywiste, że $(n+1)/n\hat{\theta}$ byłoby estymatorem nieobciążonym, ale warto również i w tym przypadku prześledzić konstrukcję jackknife'u opisaną w punkcie 2.

Zauważmy, że

$$\hat{\theta}^i = \begin{cases} X_{(n)}, & \text{gdy} \quad X_i \neq X_{(n)}, \\ X_{(n-1)}, & \text{gdy} \quad X_i = X_{(n)}, \end{cases}$$

a stąd

$$\hat{\hat{\theta}} = \frac{1}{n}\left((n-1)X_{(n)} + X_{(n-1)}\right).$$

Zgodnie z konstrukcją Quenouille'a mamy:

$$J(\hat{\theta}) = X_{(n)} + \frac{n-1}{n}(X_{(n)} - X_{(n-1)}).$$

Jest to wprawdzie estymator obciążony, ale obciążenie jego jest mniejsze od obciążenia estymatora $\hat{\theta}$, gdyż

$$EJ(\hat{\theta}) = \theta - \theta/\left(n(n+1)\right).$$

Skonstruujmy z kolei uogólniony estymator $G(\hat{\theta}_1, \hat{\theta}_2)$, który jak wiemy będzie nieobciążony, jeżeli tylko właściwie wybierzemy parametr R. Przyjmijmy

$$R = \frac{E(\hat{\theta} - \theta)}{E(\hat{\hat{\theta}} - \theta)} = \frac{n}{n+1}.$$

Wtedy

$$G(\hat{\theta}_1, \hat{\theta}_2) = (\hat{\theta} - R\hat{\hat{\theta}})/(1-R) = 2X_{(n)} - X_{(n-1)}$$

i istotnie $G(\hat{\theta}_1, \hat{\theta}_2)$ jest estymatorem nieobciążonym, gdyż $EG(\hat{\theta}_1, \hat{\theta}_2) = 2EX_{(n)} - EX_{(n-1)} = \theta$.

Pokażemy teraz, że redukcja obciążenia estymatora nie musi prowadzić do pogorszenia innych jego własności, np. błędu średniokwadratowego (MSE). W rozważanym przypadku mamy:

$$\mathrm{MSE}(\hat{\theta}) = 2\theta^2/\left((n+1)(n+2)\right),$$

$$\mathrm{MSE}\,G(\hat{\theta}_1, \hat{\theta}_2) = 2\theta^2/\left((n+1)(n+2)\right),$$

oraz

$$\mathrm{MSE}\,J(\hat{\theta}) = 2\theta^2(n^2 - n + 1)/\left(n^2(n+1)(n+2)\right).$$

Okazuje się więc, że obciążenia estymatorów $J(\hat{\theta})$ oraz $G(\hat{\theta}_1, \hat{\theta}_2)$ są mniejsze niż obciążenie estymatora $\hat{\theta}$, przy czym $\mathrm{MSE}\,J(\hat{\theta}) < \mathrm{MSE}\hat{\theta}$ dla $n \geqslant 2$ oraz $\mathrm{MSE}\,G(\hat{\theta}_1, \hat{\theta}_2) = \mathrm{MSE}\hat{\theta}$.

4. Własności estymatorów typu jackknife. Niech $\hat{\theta}_1$, $\hat{\theta}_2$ będą estymatorami parametru θ określonymi na próbie o liczności n i niech

$$b_1(n, \theta) = E(\hat{\theta}_1 - \theta) \neq 0,$$
$$b_2(n, \theta) = E(\hat{\theta}_2 - \theta) \neq 0.$$

Jeżeli $|\lim_{n\to\infty} b_1(n, \theta)/b_2(n, \theta)| = 1$, to mówimy, że $\hat{\theta}_1$ jest *estymatorem o tym samym rzędzie obciążenia* co $\hat{\theta}_2$ i piszemy $\hat{\theta}_1 \sim \hat{\theta}_2$.

Jeżeli $0 < |\lim_{n\to\infty} b_1(n, \theta)/b_2(n, \theta)| < 1$, to mówimy, że θ_1 jest *lepszym estymatorem o tym samym rzędzie obciążenia* co $\hat{\theta}_2$, i piszemy $\hat{\theta}_1 \precsim \hat{\theta}_2$.

Jeżeli $\lim_{n\to\infty} b_1(n, \theta)/b_2(n, \theta) = 0$ (bez założenia, że $b_1(n, \theta) \neq 0$), to mówimy, że $\hat{\theta}_1$ jest *estymatorem o niższym rzędzie obciążenia* niż $\hat{\theta}_2$ i piszemy $\hat{\theta}_1 \prec \hat{\theta}_2$.

LEMAT 1. *Jeżeli $\hat{\theta}$ jest estymatorem parametru θ, określonym na próbie losowej $X_1, \ldots, X_n$, oraz*

$$E(\hat{\theta} - \theta) = b(n, \theta),$$

to

$$EJ(\hat{\theta}) = \theta + b(n, \theta) + (n-1)\Delta b(n, \theta), \tag{6}$$

gdzie

$$\Delta b(n, \theta) = b(n, \theta) - b(n-1, \theta).$$

Dla dowodu wystarczy zauważyć, że

$$J(\hat{\theta}) = \hat{\theta} + (n-1)\big((\hat{\theta} - \theta) - (\hat{\hat{\theta}} - \theta)\big)$$

oraz

$$E(\hat{\hat{\theta}} - \theta) = b(n-1, \theta),$$

co wynika bezpośrednio z założenia i definicji estymatora $\hat{\hat{\theta}}$.

TWIERDZENIE 1. *Jeżeli istnieje taka liczba $p > 0$, że spełnione są warunki*

$\lim_{n\to\infty} n^p b(n, \theta) = C(\theta)$, *przy czym $C(\theta)$ jest skończone i różne od zera,*

$\lim_{n\to\infty} n^{p+1} \Delta b(n, \theta)$ *istnieje,*

to

(i) $p = 1 \Rightarrow J(\hat{\theta}) \prec \hat{\theta}$,

(ii) $p < 2, p \neq 1 \Rightarrow J(\hat{\theta}) \precsim \hat{\theta}$,

(iii) $p = 2 \Rightarrow J(\hat{\theta}) \sim \hat{\theta}$,

(iv) $p > 2 \Rightarrow \hat{\theta} \precsim J(\hat{\theta})$.

Podobne twierdzenie można sformułować dla uogólnionego jackknife'u $G(\hat{\theta}) = G(\hat{\theta}_1, \hat{\theta}_2)$.

TWIERDZENIE 2. *Niech $b(n,\theta) \neq 0$ i niech $p > 0$ będzie liczbą spełniającą założenia twierdzenia* 1. *Jeżeli*

$$R(n) = [(n-1)/n]^m, \quad m > 0,$$

to

(i) $p = m \Rightarrow G(\hat{\theta}) \prec \hat{\theta}$,

(ii) $0 < p < 2m, p \neq m \Rightarrow G(\hat{\theta}) \precsim \hat{\theta}$,

(iii) $p = 2m \Rightarrow G(\hat{\theta}) \sim \theta$,

(iv) $p > 2m \Rightarrow \hat{\theta} \precsim G(\hat{\theta})$,

(v) $p = m \neq 1 \Rightarrow G(\hat{\theta}) \prec J(\hat{\theta})$,

(vi) $\left.\begin{matrix} p > 1, 1 < m < p/(2-p) \\ p < 1, p/(2-p) < m < 1 \end{matrix}\right\} \Rightarrow G(\hat{\theta}) \precsim J(\hat{\theta})$,

(vii) $p = m = 1 \Rightarrow G(\hat{\theta}) \sim J(\hat{\theta})$,

(viii) $\left.\begin{matrix} p > 1, m < 1 \\ p < 1, m > 1 \end{matrix}\right\} \Rightarrow J(\hat{\theta}) \precsim G(\hat{\theta})$,

(ix) $p = 1, m \neq 1 \Rightarrow J(\hat{\theta}) \prec G(\hat{\theta})$.

Aby udowodnić twierdzenie 1, wystarczy posłużyć się lematem 1 i pewnymi własnościami ciągów monotonicznych. Otrzymuje się wtedy

$$\left| \lim_{n\to\infty} \frac{E\big(J(\hat{\theta})-\theta\big)}{E(\hat{\theta}-\theta)} \right| = \left| 1 + \lim_{n\to\infty} \frac{(n-1)\Delta b(n,\theta)}{b(n,\theta)} \right| = |1-p|,$$

skąd już w sposób oczywisty wynikają kolejne tezy twierdzenia 1.

Dla dowodu twierdzenia 2 zauważmy, że dla $R(n) = \big((n-1)/n\big)^m$ mamy

$$\lim_{n\to\infty} \frac{1}{n-1} \frac{R(n)}{1-R(n)} = \frac{1}{m}$$

a z dowodu twierdzenia 1

$$\lim_{n\to\infty} (n-1) \frac{\Delta b(n,\theta)}{b(n,\theta)} = -p.$$

Stąd

$$\left| \lim_{n\to\infty} \frac{E\big(G(\hat{\theta})-\theta\big)}{E(\hat{\theta}-\theta)} \right| = \left| \lim_{n\to\infty} \frac{\big(b(n,\theta)-R(n)b(n-1,\theta)\big)/\big(1-R(n)\big)}{b(n,\theta)} \right| =$$

$$= \left| \lim_{n\to\infty} \left(1 + \frac{R(n)}{1-R(n)} \frac{\Delta b(n,\theta)}{b(n,\theta)} \right) \right| =$$

$$= \left| \lim_{n\to\infty} \left(1 + \frac{R(n)}{1-R(n)} \frac{1}{1-n} \cdot (n-1) \frac{\Delta b(n,\theta)}{b(n,\theta)} \right) \right| =$$

$$= \left| 1 - \frac{p}{m} \right|;$$

więc tezy (i)–(iv) są natychmiastowe. Aby wykazać (vii) zauważmy, że dla $p = m = 1$ $J(\hat{\theta}) = G(\hat{\theta})$ więc i $J(\hat{\theta}) \sim G(\hat{\theta})$. Ponieważ

$$\left|\lim_{n\to\infty} \frac{E\big(G(\hat{\theta})-\theta\big)}{E\big(J(\hat{\theta})-\theta\big)}\right| = \left|\lim_{n\to\infty} \frac{E\big(G(\hat{\theta})-\theta\big)/E(\hat{\theta}-\theta)}{E\big(J(\hat{\theta})-\theta\big)/E(\hat{\theta}-\theta)}\right| = \left|\frac{1-p/m}{1-p}\right|,$$

to na mocy przyjętych określeń mamy (v)–(ix), co kończy dowód.

Aby zilustrować przedstawione twierdzenia zauważmy, że w omawianym wcześniej przykładzie 2 mieliśmy

$$\hat{\theta} = X_{(n)} \quad \text{i} \quad b(n, \theta) = -\theta/(n+1).$$

Stąd dla $p = 1$

$$\lim_{n\to\infty} n^p b(n, \theta) = -\theta = C(\theta)$$

i

$$\lim_{n\to\infty} n^{p+1} \Delta b(n, \theta) = \lim_{n\to\infty} n\theta/(n+1) = C,$$

więc zgodnie z tezą (i) twierdzenia 1 mamy $J(\hat{\theta}) \prec \hat{\theta}$, co rzeczywiście miało miejsce, gdyż

$$\lim_{n\to\infty} \frac{E\big(J(\hat{\theta})-\theta\big)}{E(\hat{\theta}-\theta)} = \lim_{n\to\infty} \frac{-\theta/\big(n(n+1)\big)}{-\theta/(n+1)} = 0.$$

5. Jackknife wyższego rzędu. Rozważmy ponownie estymator $G(\hat{\theta}_1, \hat{\theta}_2)$, który dla $R = a_{11}/a_{12}$ można zapisać w postaci:

$$G(\hat{\theta}_1, \hat{\theta}_2) = \frac{\begin{vmatrix} \hat{\theta}_1 & \hat{\theta}_2 \\ a_{11} & a_{12} \end{vmatrix}}{\begin{vmatrix} 1 & 1 \\ a_{11} & a_{12} \end{vmatrix}}. \tag{7}$$

Wzór ten sugeruje następujące uogólnienie:

Definicja 3. Niech $\hat{\theta}_1, \ldots, \hat{\theta}_{k+1}$ będą estymatorami parametru θ, określonymi na próbie losowej $X_1, \ldots, X_n$ i niech a_{ij}, $i = 1, \ldots, k$, $j = 1, \ldots, k+1$ będą takimi liczbami rzeczywistymi, że wyznacznik

$$\begin{vmatrix} 1 & 1 & \ldots 1 \\ a_{11} & a_{12} & \ldots a_{1k+1} \\ \cdot & \cdot & \cdot\;\cdot\;\cdot \\ a_{k1} & a_{k2} & \ldots a_{kk+1} \end{vmatrix} \tag{8}$$

jest różny od zera. *Uogólnionym jackknife'm k-tego rzędu* nazywać będziemy estymator $G(\hat{\theta}_1, \ldots, \hat{\theta}_{k+1})$, określony wzorem

$$(9) \qquad G(\hat{\theta}_1, \ldots, \hat{\theta}_{k+1}) = \frac{\begin{vmatrix} \hat{\theta}_1 & \hat{\theta}_2 & \ldots & \hat{\theta}_{k+1} \\ a_{11} & a_{12} & \ldots & a_{1k+1} \\ . & . & . & . \\ a_{k1} & a_{k2} & \ldots & a_{kk+1} \end{vmatrix}}{\begin{vmatrix} 1 & 1 & \ldots & 1 \\ a_{11} & a_{12} & \ldots & a_{1k+1} \\ . & . & . & . \\ a_{k1} & a_{k2} & \ldots & a_{kk+1} \end{vmatrix}}.$$

Jeżeli obciążenia estymatorów $\hat{\theta}_1, \ldots, \hat{\theta}_{k+1}$ dadzą się zapisać jako iloczyny pewnej funkcji liczności próby n i funkcji parametru θ, to prawdziwe jest następujące twierdzenie:

TWIERDZENIE 3. *Jeżeli*

$$E\hat{\theta}_j - \theta = \sum_{i=1}^{\infty} f_{ij}(n) b_i(\theta), \qquad j = 1, \ldots, k+1$$

oraz wyznacznik (8) *dla* $a_{ij} = f_{ij}(n)$ *jest różny od zera, to*

$$EG(\hat{\theta}_1, \ldots, \hat{\theta}_{k+1}) = \theta + b_G(n, \theta),$$

gdzie

$$b_G(n, \theta) = \frac{\begin{vmatrix} b_1^* & b_2^* & \ldots & b_{k+1}^* \\ f_{11} & f_{12} & \ldots & f_{1k+1} \\ . & . & . & . \\ f_{k1} & f_{k2} & \ldots & f_{kk+1} \end{vmatrix}}{\begin{vmatrix} 1 & 1 & \ldots & 1 \\ f_{11} & f_{12} & \ldots & f_{1k+1} \\ . & . & . & . \\ f_{k1} & f_{k2} & \ldots & f_{kk+1} \end{vmatrix}}$$

oraz

$$b_j^* = \sum_{i=k+1}^{\infty} f_{ij}(n) b_i(\theta), \qquad j = 1, \ldots, k+1,$$

Dowód wynika wprost z definicji 3 i własności wyznacznika.

Natychmiastowym wnioskiem jest stwierdzenie, że jeżeli

$$E\hat{\theta}_j - \theta = \sum_{i=1}^{k} f_{ij}(n) b_i(\theta), \qquad j = 1, \ldots, k+1,$$

to

$$EG(\hat{\theta}_1, \ldots, \hat{\theta}_{k+1}) = 0,$$

a więc jackknife k-tego rzędu jest wtedy nieobciążony.

Ograniczmy teraz nasze rozważania do szczególnej postaci uogólnionego jackknife'u, a mianowicie do przypadku, kiedy funkcje f_{ij} mają postać:

$$f_{ij}(n) = (n-j-1)^{-i}.$$

Zachodzi wtedy następujące twierdzenie:

TWIERDZENIE 4. *Jeżeli*

$$E\hat{\theta}_j - \theta = \sum_{i=1}^{\infty} f_{ij}(n) b_i(\theta), \quad j = 1, \ldots, k+1$$

oraz

$$f_{ij}(n) = (n-j+1)^{-i},$$

to

$$EG(\hat{\theta}_1, \ldots, \hat{\theta}_{k+1}) - \theta = O(n^{-k-1}).$$

Z twierdzenia tego wynika, że gdy estymatory $\hat{\theta}_1, \ldots, \hat{\theta}_{k+1}$ mają obciążenia rzędu n^{-1}, estymator $G(\hat{\theta}_1, \ldots, \hat{\theta}_{k+1})$ ma obciążenie rzędu tylko n^{-k-1}. Ponadto, jeżeli

$$E\hat{\theta}_j - \theta = \sum_{i=1}^{k} \frac{b_i(\theta)}{(bn-j+1)^i},$$

to estymator $G(\hat{\theta}_1, \ldots, \hat{\theta}_{k+1})$ jest nieobciążony.

Dowód powyższego twierdzenia znaleźć można np. w [2].

6. Rozszerzona metoda Quenouille'a. Jak widać z definicji 3, konstrukcja jackknife'u wyższego rzędu dla estymatora $\hat{\theta}_1$ wymaga utworzenia k nowych estymatorów $\hat{\theta}_2, \ldots, \hat{\theta}_{k+1}$. Podobnie jak poprzednio, i tutaj można się posłużyć odpowiednio rozszerzoną metodą Quenouille'a.

Niech $\hat{\theta}_1 = \hat{\theta}_1(X_1, \ldots, X_n)$ będzie estymatorem określonym dla każdego $n = 1, 2, \ldots$, a $\hat{\theta}^{i_1, \ldots, i_j}$ — estymatorem $\hat{\theta}_1$ na próbie $(X_1, \ldots, X_n)$, z której odrzucono obserwacje $X_{i_1}, \ldots, X_{i_j}$, przy czym wskaźniki $i_1, \ldots, i_j$ wybiera się losowo. Niech $k < n$. Weźmy pod uwagę ciąg estymatorów $\hat{\theta}_1^{i_2}, \ldots, \hat{\theta}_1^{i_2, \ldots, i_{k+1}}$ i zdefiniujmy estymatory $\hat{\theta}_j$ wzorem

$$\hat{\theta}_j = \hat{\bar{\theta}}{}_1^{i_2, \ldots, i_j} = \frac{1}{\binom{n}{j-1}} \sum \hat{\theta}_1^{i_2, \ldots, i_j} \tag{10}$$

dla $j = 2, \ldots, k+1$, przy czym sumowanie rozciąga się na wszystkie permutacje liczb $i_2, \ldots, i_j$.

Na przykład, w rozważanym już wcześniej przykładzie 2 mieliśmy

$$\hat{\theta}_1 = \hat{\theta} = X_{(n)},$$

$$\hat{\theta}_2 = \hat{\hat{\theta}} = \frac{1}{n}[(n-1)X_{(n)} + X_{(n-1)}].$$

Aby wyznaczyć $\hat{\theta}_3$ wyznaczmy estymatory $\hat{\theta}^{ij}$, które mogą przyjmować następujące wartości:

$$\begin{array}{lll} X_{(n)} & \text{gdy} & X_i, X_j \neq X_{(n)}, \\ X_{(n-1)} & \text{gdy} & (X_i = X_{(n)}, X_j \neq X_{(n-1)}) \quad \text{lub} \quad (X_j = X_{(n)}, X_i \neq X_{(n-1)}), \\ X_{(n-2)} & \text{gdy} & (X_i = X_{(n)}, X_j = X_{(n-1)}) \quad \text{lub} \quad (X_j = X_{(n)}, X_i = X_{(n-1)}). \end{array}$$

Stąd średnia $\binom{n}{2}$ możliwych wartości $\hat{\theta}^{ij}$, przy uwzględnieniu częstości pojawiania się każdej z wartości $X_{(n)}$, $X_{(n-1)}$, $X_{(n-2)}$, wynosić będzie

$$\hat{\theta}_3 = \frac{2}{n(n-1)}\left[\frac{(n-1)(n-2)}{2}X_{(n)} + (n-2)X_{(n-1)} + X_{(n-2)}\right].$$

Zauważmy, że jeżeli $E\hat{\theta}_1 - \theta = b(n, \theta)$, to przy omawianym sposobie konstrukcji $\hat{\theta}_j$ mamy $E\hat{\theta}_j - \theta = b(n-j+1, \theta)$ a stąd, jeżeli tylko obciążenie estymatora $\hat{\theta}_1$ można rozwinąć względem potęg wyrażenia $1/n$, w postaci

$$E\hat{\theta}_1 - \theta = \frac{a_1}{n} + \frac{a_2}{n^2} + \ldots,$$

to

$$E\hat{\theta}_j - \theta = \frac{a_1}{n-j+1} + \frac{a_2}{(n-j+1)^2} + \ldots$$

i, zgodnie z definicją,

$$G(\hat{\theta}_1, \ldots, \hat{\theta}_{k+1}) = \frac{\begin{vmatrix} \hat{\theta}_1 & \hat{\theta}_2 & \ldots & \hat{\theta}_{k+1} \\ \frac{1}{n} & \frac{1}{n-1} & \ldots & \frac{1}{n-k} \\ \ldots & \ldots & \ldots & \ldots \\ \frac{1}{n^k} & \frac{1}{(n-1)^k} & \ldots & \frac{1}{(n-k)^k} \end{vmatrix}}{\begin{vmatrix} 1 & 1 & \ldots & 1 \\ \frac{1}{n} & \frac{1}{n-1} & \ldots & \frac{1}{n-k} \\ \ldots & \ldots & \ldots & \ldots \\ \frac{1}{n^k} & \frac{1}{(n-1)^k} & \ldots & \frac{1}{(n-k)^k} \end{vmatrix}} \tag{11}$$

Na mocy twierdzenia 4 mamy teraz, że

$$EG(\hat{\theta}_1, \ldots, \hat{\theta}_{k+1}) - \theta = O(n^{-k-1}),$$

a dla $a_{k+1} = a_{k+2} = \ldots = 0$ estymator $G(\hat{\theta}_1, \ldots, \hat{\theta}_{k+1})$ jest nieobciążony.

Skonstruowany estymator postaci (11) nazywać też będziemy *zwykłym jackknife'm k-tego rzędu* i oznaczać będziemy symbolem $J^{(k)}(\hat{\theta})$. Ponadto, wygodnie będzie stosować uproszczony zapis:

$$G^{(k)}(\hat{\theta}) = G(\hat{\theta}_1, \ldots, \hat{\theta}_{k+1}).$$

PRZYKŁAD 3. Wyznaczmy postać estymatora $J^{(2)}(\hat{\theta})$. Na mocy przyjętych określeń mamy

$$J^{(2)}(\hat{\theta}) = G^{(2)}(\hat{\theta})$$

dla $f_{ij}(n) = 1/(n-j+1)^i$, $j = 1, 2, 3$; $i = 1, 2$; czyli

$$J^{(2)}(\hat{\theta}) = \frac{\begin{vmatrix} \hat{\theta}_1 & \hat{\theta}_2 & \hat{\theta}_3 \\ \frac{1}{n} & \frac{1}{n-1} & \frac{1}{n-2} \\ \frac{1}{n^2} & \frac{1}{(n-1)^2} & \frac{1}{(n-2)^2} \end{vmatrix}}{\begin{vmatrix} 1 & 1 & 1 \\ \frac{1}{n} & \frac{1}{n-1} & \frac{1}{n-2} \\ \frac{1}{n^2} & \frac{1}{(n-1)^2} & \frac{1}{(n-2)^2} \end{vmatrix}}$$

$$= \frac{1}{2}\left(n^2\hat{\theta}_1 - 2(n-1)^2\hat{\theta}_2 + (n-2)^2\hat{\theta}_3\right).$$

Istnieje wiele możliwości zarówno modyfikacji samej metody jak i wykorzystania estymatorów $J(\hat{\theta})$. Jedną z takich propozycji jest pomysł T. Sharota [7] „wyostrzenia'' jackknife'u (sharpening the jackknife). Sugeruje on mianowicie utworzenie nowej rodziny estymatorów typu jackknife, składającej się z liniowych kombinacji estymatorów $\hat{\theta}, \hat{\hat{\theta}}, \hat{\theta}^{ij}$;

$$J(p) = pn\hat{\theta} + (1-2p)(n-1)\hat{\hat{\theta}} + (p-1)(n-2)\hat{\theta}^{ij},$$

przy czym za p wybieramy taką wartość, która minimalizuje wariancję $J(p)$. Oczywiście $J(1) = J(\hat{\theta})$, a $J(n/2) = J^{(2)}(\hat{\theta})$.

7. Pewne zastosowania metody jackknife. Początkowo jedyną motywacją dla jackknife'u była redukcja obciążenia estymatorów lub wręcz konstruowanie estymatorów nieobciążonych. Dopiero późniejsze badania asymptotycznych własności uczyniły z jackknife'u szczególnie użyteczne narzędzie w analizie danych.

Już w 1958 r. J. W. Tukey [8] wskazał, że pseudowartości $J_i(\hat{\theta})$ mogą być w przybliżeniu traktowane jako niezależne zmienne losowe o jednakowym rozkładzie — a stąd jackknifing może być użyteczny przy konstrukcji przybliżonych przedziałów ufności. Procedura ta, stosowana z zadawalającymi rezultatami, ma obecnie szeroką podbudowę teoretyczną w postaci twierdzeń asymptotycznych, z których wiele znaleźć można w pracy [1]. Pozwalają one traktować pseudowartości

$$J_i(\hat{\theta}) = n\hat{\theta} - (n-1)\hat{\theta}^i$$

jako zmienne losowe, asymptotycznie niezależne i o jednakowym rozkładzie. Zmienne te w pewnym sensie reprezentują wpływ każdej pojedyńczej obserwacji na $\hat{\theta}$, więc estymatorem wariancji $\hat{\theta}$ może być wyrażenie

$$\operatorname{var}(\hat{\theta}) = \frac{1}{n(n-1)} \sum_{i=1}^{n} (J_i(\hat{\theta}) - J(\hat{\theta}))^2.$$

Statystyka

$$\frac{(J(\hat{\theta}) - \theta)\sqrt{n(n-1)}}{\sqrt{\sum (J_i(\hat{\theta}) - J(\hat{\theta}))^2}}$$

ma wtedy asymptotycznie rozkład t-Studenta z $(n-1)$ stopniami swobody.

Najcenniejszym zastosowaniem techniki jackknife'u, zwłaszcza z punktu widzenia odporności metod statystycznych, jest testowanie hipotez o wariancji. Wiadomo, że w tym problemie niewygodnie jest opierać test na rozkładzie χ^2 lub F, gdyż taki test jest bardzo czuły na odchylenia od normalności. Z drugiej strony, niektóre z wersji klasycznych metod przewidywały podział próby losowej na grupy, co sugerowało możliwość zastosowania jackknife'u. Wiele miejsca temu zagadnieniu poświęcił w swoich pracach R. G. Miller [3].

Na przykład, niech zadanie polega na testowaniu hipotezy H_0: $\sigma_y^2 = \sigma_x^2$, gdzie

$$X_1, \ldots, X_n \text{ — próba losowa z populacji} \sim F\left(\frac{x-\mu_x}{\sigma_x}\right),$$

$$Y_1, \ldots, Y_m \text{ — próba losowa z populacji} \sim F\left(\frac{y-\mu_y}{\sigma_y}\right).$$

Parametry $(\mu_x, \mu_y, \sigma_x, \sigma_y)$ są nieznane, tak jak i postać rozkładu F (poza założeniem, że F ma skończone momenty do czwartego rzędu włącznie). Do konstrukcji testu może służyć statystyka

$$t = \frac{(J(\hat{\theta}_x) - \theta_x) - (J(\hat{\theta}_y) - \theta_y)}{\sqrt{\frac{1}{n(n-1)} \sum_{i=1}^{n} (J_i(\hat{\theta}_x) - J(\hat{\theta}_x))^2 + \frac{1}{m(m-1)} \sum_{i=1}^{m} (J_i(\hat{\theta}_y) - J(\hat{\theta}_y))^2}},$$

gdzie

$$\theta_x = \log \sigma_x^2, \qquad \theta_y = \log \sigma_y^2.$$

Statystyka ta ma rozkład t-Studenta z $(n+m-2)$ stopniami swobody.

Niektórzy autorzy zwracają jednak uwagę na fakt, że niekiedy natura estymatora $\hat{\theta}$ jest taka, że jego pseudowartości nie tworzą n różnych wartości. Sugerują oni, aby w takiej sytuacji ilość stopni swobody rozważanej statystyki obniżyć do liczby równej ilości tych różnych wartości, pomniejszonej o jeden. Na przykład, w omawianym przykładzie 2 mieliśmy tylko dwie różne wartości, ponieważ $\hat{\theta}^i$ było równe $X_{(n)}$ lub $X_{(n-1)}$ — więc zgodnie z tym rozkład t należałoby traktować jako rozkład z jednym stopniem swobody.

8. Uwagi końcowe. Jak już wcześniej zauważono „jackknife — to dostosowanie podstaw teoretycznych do tego, co każdy dobry i doświadczony statystyk już dawno robił w sposób nieformalny". Jest jeszcze wprawdzie wiele niejasności przy formułowaniu zagadnień związanych z samą metodą, jak i wiele trudności natury technicznej przy liczności podgrup $M > 1$, ale z drugiej strony jackknife jest jeszcze jedną, jeżeli nie jedyną, możliwością rozwiązania postawionego problemu.

Jak podkreśla R. G. Miller [4], [5] wynik zastosowania jackknife'u w dużym stopniu zależy od „rozsądnego" podziału danych na grupy; występują bowiem sytuacje, gdy przyjęte założenia o asymptotycznej niezależności $J_i(\hat{\theta})$ nie są spełnione. Nie umniejsza to jednak znaczenia samej techniki, która w wielu przypadkach daje szybkie i proste rozwiązania — jak przy cięciu nożem — czyli jackknifing.

Prace cytowane

[1] H. L. Grey, W. L. Schucany, *The generalized jackknife statistic*, Marcel Dekker, New York 1972.

[2] —, —, *On bias reduction in estimation*, J. Amer. Statist. Assoc. 66 (1971), str. 524–533.

[3] R. G. Miller, *Jackknifing variances*, Ann. of Math. Statist. 39 (1968), str. 567–582.

[4] —, *Jackknifing — a review*, Biometrika 61 (1974a), str. 1–15.

[5] —, *A trustworthy jackknife*, Ann. of Math. Statist. 35 (1964), str. 1594–1605.

[6] M. H. Quenouille, *Notes on bias in estimation*, Biometrika 43 (1956), str. 353–360.

[7] T. Sharot, *Sharpening the jackknife*, ibid. 63 (1976), str. 315–321.

[8] J. W. Tukey, *Bias and confidence in not-quite large samples*, Ann. of Math. Statist. 29 (1958) (Abstract), str. 614.