CZEN PIN (Warszawa)

O MINIMAKSOWYM ESTYMATORZE LICZNOŚCI POPULACJI

Wstęp. W ekologii często spotykamy się z zagadnieniem estymacji liczności pewnej populacji. Przykładem może być estymacja liczności ryb określonego gatunku i wielkości w jeziorze, estymacja liczności pewnego gatunku ptaków w Tatrach itp. Zwykle robi się to *metodą łowienia i znakowania*. Metoda ta wygląda następująco: Łapiemy t ptaków, znakujemy ptaki złapane i puszczamy je. Po pewnym czasie, gdy ptaki znakowane i nieznakowane są już dobrze pomieszane, łapiemy n ptaków i notujemy liczbę ptaków znakowanych. Niech ich będzie s wśród n . Z tych trzech liczb t, s, n budujemy estymator liczności ptaków w tej populacji.

Przy metodzie łowienia i znakowania można też postępować inaczej, mianowicie tak zwaną *metodą odwrotnych próbek* (inverse sample censuses): Najpierw łapiemy t ptaków i znakujemy je. Po znakowaniu puszczamy je, tak jak w sposobie poprzednim, a następnie łapiemy po jednym, ze zwracaniem lub bez zwracania, dotąd, aż złapiemy z góry zadaną liczbę s ptaków. Niech n oznacza liczbę łowień w drugiej serii. Z liczb n, s, t budujemy estymator liczności ptaków w tej populacji.

W niniejszej pracy zajmiemy się bliżej postępowaniem drugim. Niech N oznacza liczbę ptaków w populacji. D. G. Chapman [1] znalazł nieobciążone estymatory $\hat{N}_c^1, \hat{N}_c^2$ parametru N metodą maksimum wiarygodności.

a. Dla próbek ze zwracaniem

$$\hat{N}_c^1 = \frac{nt}{s},$$

$$(1) \quad E(n) = \frac{s}{t} N,$$

$$E(\hat{N}_c^1 - N)^2 = \frac{N(N-t)}{s}.$$

b. Dla próbek bez zwracania

$$\hat{N}_c^2 = \frac{(t+1)n}{s} - 1,$$

$$(2) \quad E(n) = \frac{s}{t+1} (N+1),$$

$$E(\hat{N}_c^2 - N)^2 = \frac{t-s+1}{s(t+2)} (N+1)(N-t).$$

W niniejszej pracy znajdziemy estymator minimaksowy parametru N i porównamy go z estymatorami Chapmana.

DEFINICJA. $M(n)$ nazywamy *minimaksowym estymatorem parametru* N dla $N \in \Omega$, przy funkcji straty $R(f, N) = \frac{1}{N^2} (f - N)^2$, jeżeli minimizuje on $\sup_{N \in \Omega} E[R(f, N)]$.

1. Pobieranie próbek ze zwracaniem. $n = k$ oznacza tu, że w pierwszych $k-1$ łowieniach otrzymaliśmy $s-1$ ptaków znakowanych oraz że w k -tym łowieniu złapaliśmy ptaka znakowanego. Prawdopodobieństwo tego zdarzenia jest równe

$$(3) \quad \begin{aligned} P[n = k | N] &= \binom{k-1}{s-1} \left(\frac{t}{N}\right)^{s-1} \left(1 - \frac{t}{N}\right)^{k-s} \frac{t}{N} = \\ &= \binom{k-1}{s-1} \left(\frac{t}{N}\right)^s \left(1 - \frac{t}{N}\right)^{k-s} \quad (k = s, s+1, \dots). \end{aligned}$$

Łatwo zauważyć, że funkcja

$$\binom{k-1}{s-1} \left(\frac{t}{N}\right)^s \left(1 - \frac{t}{N}\right)^{k-s} \quad (k = s, s+1, \dots)$$

jest określona dla dowolnych liczb rzeczywistych $N \geq t$, $t > 0$ i dowolnej liczby naturalnej s . Postawmy nasze zagadnienie następująco: Niech zmienna losowa n ma rozkład z nieznanym parametrem N

$$p_N(k) = \binom{k-1}{s-1} \left(\frac{t}{N}\right)^s \left(1 - \frac{t}{N}\right)^{k-s} \quad (k = s, s+1, \dots).$$

Oznaczamy przez Ω zbiór wartości parametru N , gdzie $\Omega \subseteq [t, \infty)$. Zagadnienie polega na znalezieniu minimaksowego estymatora parametru N dla $N \in \Omega$.

Łatwo obliczyć, że

$$(4) \quad \begin{aligned} E_N(n) &= \frac{s}{t} N, \\ E_N\left(n - \frac{s}{t} N\right)^2 &= \frac{s}{t^2} N(N-t), \\ E_N\left(\frac{\partial}{\partial N} \log p_N(n)\right)^2 &= \frac{s}{N(N-t)}. \end{aligned}$$

Rozpatrując estymator parametru N , mający postać $an+b$, gdzie a, b nie zależą od n , lecz mogą zależeć od s i t , mamy

$$\begin{aligned} \frac{1}{N^2} E_N(an+b-N)^2 &= \frac{1}{N^2} E_N\left[a\left(n - \frac{sN}{t}\right) + b - N\left(1 - \frac{as}{t}\right)\right]^2 = \\ &= a^2 \frac{s(N-t)}{t^2 N} + \frac{1}{N^2} \left[b - N\left(1 - \frac{as}{t}\right)\right]^2 = \\ &= \left(1 - \frac{as}{t}\right)^2 + \frac{a^2 s}{t^2} - \left[\frac{a^2 s}{t} + 2\left(1 - \frac{as}{t}\right)b\right] \frac{1}{N} + \frac{b^2}{N^2}. \end{aligned}$$

Bierzemy takie a , żeby $\left(1 - \frac{as}{t}\right)^2 + \frac{a^2 s}{t^2}$ osiągnęło minimum. Zachodzi to dla $a = \frac{t}{s+1}$. Mamy zatem

$$(5) \quad \frac{1}{N^2} E_N\left(\frac{t}{s+1} n + b - N\right)^2 = \frac{1}{s+1} - \left[\frac{ts}{(s+1)^2} + \frac{2b}{s+1}\right] \frac{1}{N} + \frac{b^2}{N^2}.$$

Łatwo zauważyć, że dla

$$\frac{N_0}{s+1} \left(1 - \sqrt{1 + \frac{st}{N_0}}\right) \leq b \leq \frac{N_0}{s+1} \left(1 + \sqrt{1 + \frac{st}{N_0}}\right)$$

zachodzi równość

$$(6) \quad \sup_{N \geq N_0 \geq t} \frac{1}{N^2} E_N\left(\frac{t}{s+1} n + b - N\right)^2 = \frac{1}{s+1}.$$

TWIERDZENIE 1. Jeżeli

(i) zbiór Ω zawiera rosnący ciąg $\{N_i\}$ taki, że $\lim_{i \rightarrow \infty} N_i = \infty$, $N_{i+1} - N_i \leq K$ (const) dla $i = 1, 2, \dots$,

$$(ii) \quad \frac{N_0}{s+1} \left(1 - \sqrt{1 + \frac{st}{N_0}}\right) < b < \frac{N_0}{s+1} \left(1 + \sqrt{1 + \frac{st}{N_0}}\right),$$

gdzie $N_0 = \inf_{N \in \Omega} N$,

to $\frac{t}{s+1} n + b$ jest minimaksowym estymatorem parametru N dla $N \in \Omega$.

Dowód. Skorzystamy tu z nierówności Rao-Craméra [2], [3]. Łatwo zauważyć, że dla $t < \underline{N} \leq N \leq \bar{N}$

$$\begin{aligned} \left| \frac{\partial}{\partial N} p_N(k) \right| &= \\ &= \left| \frac{kt - \bar{N}s}{N(N-t)} \right| p_N(k) \leq \frac{kt + \bar{N}s}{N(N-t)} \binom{k-1}{s-1} \left(\frac{t}{N} \right)^s \left(1 - \frac{t}{N} \right)^{k-s} = \\ &= \frac{kt + \bar{N}s}{N(N-t)} \left(\frac{\bar{N}}{N} \right)^s p_{\bar{N}}(k) \quad (k = s, s+1, \dots) \end{aligned}$$

oraz

$$\sum_{k=0}^{\infty} \frac{kt + \bar{N}t}{N(N-t)} \left(\frac{\bar{N}}{N} \right)^s p_{\bar{N}}(k) = \frac{1}{N-t} \left(\frac{\bar{N}}{N} \right)^s \left(s + \frac{\bar{N}}{N} t \right) < \infty.$$

Stąd widzimy, że warunki potrzebne do tego, aby zachodziła nierówność Rao-Craméra, są spełnione, a więc

$$\sigma_f^2 = E_N[f(n) - E_N(f(n))]^2 \geq \frac{(1 + b'(N))^2}{E_N \left(\frac{\partial}{\partial N} \log p_N(n) \right)^2},$$

gdy $t < \underline{N} \leq N \leq \bar{N} < \infty$ i $E_{\bar{N}}|f(n)n| < \infty$, gdzie $b(N) = E_N(f(n)) - \underline{N}$.

Ponieważ $\underline{N}, \bar{N}$ są dowolne, więc dla każdego estymatora $f(n)$ takiego, że $E_N|f(n)n| < \infty$, zachodzi nierówność

$$(7) \quad \sigma_f^2 \geq \frac{(1 + b'(N))^2}{E_N \left(\frac{\partial}{\partial N} \log p_N(n) \right)^2},$$

jeżeli tylko $t < N$. Gdy $t = N$, mamy

$$(8) \quad E_N(f(n) - N)^2 = (f(s) - N)^2 = (E_N(f(n)) - N)^2 = b^2(N).$$

Z wzorów (7) i (8) wynika, że dla każdego estymatora $f(n)$ spełniającego $E_N|f(n)n| < \infty$

$$(9) \quad \frac{1}{N^2} E_N(f(n) - N)^2 = \frac{1}{N^2} (\sigma_f^2 + b^2(N)) \geq \frac{b^2(N)}{N^2} + \frac{N-t}{sN} (1 + b'(N))^2,$$

jeżeli tylko $N \geq t$.

Przystępujemy do dowodu twierdzenia. Gdyby $\frac{t}{s+1}n + b$ nie był minimaxowym estymatorem parametru N dla $N \in \Omega$, to istniałoby takie $\varepsilon > 0$ i estymator $f_0(n)$, że zachodziłaby nierówność

$$\begin{aligned}
(10) \quad & \sup_i \frac{1}{N_i^2} E_{N_i} (f_0(n) - N)^2 \leq \\
& \leq \sup_{N \in \Omega} \frac{1}{N^2} E_N (f_0(n) - N)^2 \leq \sup_{N \in \Omega} \frac{1}{N^2} E_N \left(\frac{t}{s+1} n + b - N \right)^2 - \varepsilon \leq \\
& \leq \sup_{N \geq N_0} \frac{1}{N^2} E_N \left(\frac{t}{s+1} n + b - N \right)^2 - \varepsilon = \frac{1}{s+1} - \varepsilon.
\end{aligned}$$

A więc wzór (9) zachodzi dla $N \in \Omega$, co daje

$$(11) \quad b^2(N_i) \leq E_{N_i} (f_0(n) - N_i)^2 \leq N_i^2 \left(\frac{1}{s+1} - \varepsilon \right).$$

Z warunku (i) i wzorów (10), (11) wynika, że dla $N_{i-1} \leq N \leq N_i$

$$\begin{aligned}
(12) \quad & E_N (f_0(n) - N)^2 = E_{N_i} (f_0(n) - N_i)^2 + E_N (f_0(n) - N)^2 - E_{N_i} (f_0(n) - N_i)^2 = \\
& = E_{N_i} (f_0(n) - N_i)^2 + \\
& + \sum_{k=s}^{\infty} [(f_0(k) - N)^2 - (f_0(k) - N_i)^2] p_{N_i}(k) + \sum_{k=s}^{\infty} (f_0(k) - N)^2 (p_N(k) - p_{N_i}(k)) = \\
& = E_{N_i} (f_0(n) - N_i)^2 + 2(N_i - N)b(N_i) + (N_i - N)^2 + \\
& + \sum_{k=s}^{\infty} (f_0(n) - N_i)^2 \binom{k-1}{s-1} \left\{ \left[\left(\frac{t}{N} \right)^s - \left(\frac{t}{N_i} \right)^s \right] \left(1 - \frac{t}{N_i} \right)^{k-s} + \right. \\
& \quad \left. + \left(\frac{t}{N} \right)^s \left[\left(1 - \frac{t}{N} \right)^{k-s} - \left(1 - \frac{t}{N_i} \right)^{k-s} \right] \right\} \leq \\
& \leq E_{N_i} (f_0(n) - N_i)^2 + 2K|b(N_i)| + K^2 + \sum_{k=s}^{\infty} (f_0(k) - N)^2 p_{N_i}(k) \frac{K}{N} \sum_{v=0}^{s-1} \left(\frac{N_i}{N} \right)^v \leq \\
& \leq E_{N_i} (f_0(n) - N_i)^2 + 2K|b(N_i)| + K^2 + \frac{K(1+K)^s}{N_i - 1} E_{N_i} (f_0(n) - N)^2 \leq \\
& \leq E_{N_i} (f_0(n) - N_i)^2 + 2K|b(N_i)| + K^2 + \\
& \quad + \frac{K(1+K)^s}{N_i - 1} [E_{N_i} (f_0(n) - N_i)^2 + 2K|b(N_i)| + K^2] \leq \\
& \leq E_{N_i} (f_0(n) - N_i)^2 + 2K^2(1+K)^s + 4K(1+K)^s N_i \sqrt{\frac{1}{1+s}} - \varepsilon + \\
& \quad + \frac{K(1+K)^s}{N_i - 1} N_i^2 \left(\frac{1}{s+1} - \varepsilon \right) \leq E_{N_i} (f_0(n) - N_i)^2 + MN_i, \quad i = 1, 2, \dots,
\end{aligned}$$

gdzie M jest stałą niezależną od N .

Z wzorów (10), (12) i warunku (i) wynika, że

$$\begin{aligned} \frac{1}{N^2} E_N(f_0(n) - N)^2 &\leq \frac{1}{N^2} E_{N_i}(f_0(n) - N_i)^2 + \frac{N_i}{N^2} M \leq \\ &\leq \left(\frac{N_i}{N}\right)^2 \left(\frac{1}{s+1} - \varepsilon\right) + \frac{1}{N} \left(1 + \frac{K}{N}\right) M \leq \\ &\leq \frac{1}{s+1} - \varepsilon + \frac{2K}{N} \left(\frac{1}{s+1} - \varepsilon + 1\right) + \frac{K}{N^2} (K + M) \end{aligned}$$

dla $N_{i-1} \leq N \leq N_i$, $i = 1, 2, \dots$

Ponieważ $\lim_{i \rightarrow \infty} N_i = \infty$, więc istnieje liczba N' taka, że dla $N \geq N'$ zachodzi

$$\frac{1}{N^2} E_N(f_0(n) - N)^2 \leq \frac{1}{s+1} - \frac{\varepsilon}{2}.$$

Korzystając więc z wzoru (9) otrzymujemy

$$\frac{b^2(N)}{N^2} + \frac{N-t}{sN} (1+b'(N))^2 \leq \frac{1}{s+1} - \frac{\varepsilon}{2}.$$

Zatem dla $N \geq N' > t$ otrzymujemy

$$\begin{aligned} (1+b'(N))^2 &\leq \frac{sN}{N-t} \left(\frac{1}{s+1} - \frac{\varepsilon}{2}\right) < \\ &< s(t+1) \left(\frac{1}{s+1} - \frac{\varepsilon}{2}\right). \end{aligned}$$

Istnieje więc N'' takie, że dla $N \geq N''$ zachodzi nierówność

$$\begin{aligned} \frac{b^2(N)}{N^2} + \frac{(1+b'(N))^2}{s} &\leq \frac{1}{s+1} - \frac{\varepsilon}{2} + \frac{t}{sN} (1+b'(N))^2 \leq \\ &\leq \frac{1}{s+1} - \frac{\varepsilon}{4}. \end{aligned}$$

Niech

$$C(N) = b(N) + \frac{N}{s+1}.$$

Mamy zatem

$$\frac{C^2(N)}{N^2} + \frac{2}{s+1} \left(C'(N) - \frac{C(N)}{N}\right) + \frac{[C'(N)]^2}{s} \leq -\frac{\varepsilon}{4}.$$

Przyjmując $g(N) = \frac{C(N)}{N}$ otrzymujemy

$$g^2(N) + \frac{2N}{s+1} g'(N) \leq -\frac{\varepsilon}{4},$$

czyli

$$\frac{1}{N} \leq -\frac{2}{s+1} \cdot \frac{g'(N)}{g^2(N) + \frac{1}{4}\varepsilon}.$$

Całkując obie strony tej nierówności od N'' do N otrzymujemy

$$\begin{aligned} \log N - \log N'' &= \int_{N''}^N \frac{dx}{x} \leq -\frac{2}{s+1} \int_{N''}^N \frac{g'(x)}{g^2(x) + \frac{1}{4}\varepsilon} dx = \\ &= -\frac{4}{s+1} \cdot \frac{1}{\sqrt{\varepsilon}} \operatorname{arctg} \frac{2g(x)}{\sqrt{\varepsilon}} \Big|_{N''}^N. \end{aligned}$$

Lewa strona tej nierówności dąży do nieskończoności, gdy N dąży do nieskończoności, lecz prawa strona tej nierówności zawsze jest ograniczona. Wobec tego nie istnieje estymator spełniający warunek (10), c.b.d.o.

WNIOSEK. Niech Ω_{N_0} oznacza zbiór wszystkich naturalnych liczb $N \geq N_0$. Jeżeli

$$\frac{N_0}{s+1} \left(1 - \sqrt{1 + \frac{st}{N_0}} \right) \leq b \leq \frac{N_0}{s+1} \left(1 + \sqrt{1 + \frac{st}{N_0}} \right),$$

to $\frac{t}{s+1}n + b$ jest minimaksowym estymatorem parametru N dla $\Omega = \Omega_{N_0}$.

Z wniosku tego widzimy, że otrzymaliśmy całą rodzinę (z parametrem b) estymatorów minimaksowych dla $\Omega = \Omega_{N_0}$. Parametr b możemy dobrać w zależności od konkretnego zagadnienia.

(i) $b = \frac{t}{s+1}$, otrzymamy wówczas estymator

$$(13) \quad m_1(n) = \frac{t}{s+1} (n+1)$$

i wzór

$$E_N(m_1(n) - N)^2 = \frac{N(N-t)}{s+1} \left(1 - \frac{t}{N(s+1)} \right).$$

(ii) Jeżeli wiemy, że $N \geq \bar{N} > t\sqrt{s}$ i N jest bliskie $\bar{N}$, to wybieramy b tak, żeby nasz estymator osiągał minimum straty przy $N = \bar{N}$, tj.

przyjmujemy $b = \frac{\bar{N}}{s+1}$. Mamy zatem estymator

$$(14) \quad m_2(n) = \frac{1}{s+1} (tn + \bar{N})$$

i wzór

$$E_N(m_2(n) - N)^2 = \frac{N^2}{s+1} - \frac{N}{(s+1)^2} \left(ts + 2\bar{N} - \frac{\bar{N}^2}{N} \right).$$

Porównując $m_1(n)$ i $m_2(n)$ z wzorami Chapmana otrzymamy

$$E_N(m_1(n) - N)^2 < E_N(\hat{N}_c^1 - N)^2, \quad \text{gdy } N > t,$$

$$E_N(m_2(n) - N)^2 < E_N(\hat{N}_c^1 - N)^2, \quad \text{gdy } N > \bar{N}.$$

2. Pobieranie próbek bez zwracania. $n = k$ oznacza tu, że w pierwszych $k-1$ łowieniach otrzymaliśmy $s-1$ ptaków znakowanych, i w k -tym łowieniu złapaliśmy ptaka znakowanego. Prawdopodobieństwo tego zdarzenia jest równe

$$(15) \quad P[n = k | N] = \frac{\binom{t}{s-1} \binom{N-t}{k-s}}{\binom{N}{k-1}} \cdot \frac{t-s+1}{N-k+1} = \frac{\binom{k-1}{s-1} \binom{N-k}{t-s}}{\binom{N}{t}}.$$

($k = s, s+1, \dots, k+s$).

Łatwo obliczyć, że

$$E_N(n) = \frac{s(N+1)}{t+1},$$

$$(16) \quad E_N(n^2) = \frac{s(s+1)(N+1)(N+2)}{(t+1)(t+2)} - \frac{s(N+1)}{t+1},$$

$$E_N(n^v) = \frac{s(s+1)\dots(s+v-1)}{(t+1)\dots(t+v)} N^v + O(N^{v-1}) \quad (v = 1, 2, \dots).$$

Rozpatrujemy estymator parametru N mający postać $an+b$, gdzie a, b nie zależą od n , lecz mogą zależeć od s i t . Mamy

$$\begin{aligned} & \frac{1}{N^2} E_N(an+b-N)^2 = \\ &= \frac{a^2s(s+1)}{(t+1)(t+2)} - \frac{2as}{t+1} + 1 + \frac{1}{N} \left[\frac{3a^2s(s+1)}{(t+1)(t+2)} + \frac{2abs-2as-a^2s}{t+1} - 2b \right] + \\ & \quad + \frac{1}{N^2} \left[b^2 + \frac{2a^2s(s+1)}{(t+1)(t+2)} - \frac{as(a-2b)}{t+1} \right]. \end{aligned}$$

Dobieramy a tak, żeby $\frac{a^2 s(s+1)}{(t+1)(t+2)} - \frac{2as}{t+1} + 1$ osiągało minimum.

W tym celu trzeba przyjąć $a = \frac{t+2}{s+1}$. Mamy więc

$$(17) \quad \frac{1}{N^2} E_N \left(\frac{t+2}{s+1} n + b - N \right)^2 =$$

$$= \frac{t-s+1}{(t+1)(s+1)} \left[1 - \frac{1}{N} \left(\frac{s(t+2)}{s+1} + 2b \right) \right] +$$

$$+ \frac{1}{N^2} \left[b^2 + \frac{(t+2)s}{(s+1)(t+1)} \left(2b + 2 - \frac{t+2}{s+1} \right) \right].$$

Oznaczmy przez Ω_{N_0} zbiór wszystkich liczb naturalnych niemniejszych od liczby N_0 . Łatwo obliczyć, że gdy

$$\frac{N_0(t-s+1)}{(s+1)(t+1)} \left(1 - \sqrt{\frac{N_0+st-2}{N_0}} \right) \leq b+1 \leq$$

$$\leq \frac{N_0(t-s+1)}{(s+1)(t+1)} \left(1 + \sqrt{\frac{N_0+st-2}{N_0}} \right),$$

to zachodzi równość

$$(18) \quad \sup_{N \in \Omega_{N_0}} \frac{1}{N^2} E_N \left(\frac{t+2}{s+1} n + b - N \right)^2 = \frac{t-s+1}{(s+1)(t+1)}.$$

Istnieje hipoteza, że $\frac{t+2}{s+1} n + b$ jest minimaksowym estymatorem parametru N dla $\Omega = \Omega_{N_0}$. Tutaj udowodnimy tylko twierdzenie następujące.

Oznaczmy przez E zbiór wszystkich estymatorów, które mają formę wymierną $\frac{a_0 n^k + a_1 n^{k-1} + \dots + a_k}{b_0 n^l + b_1 n^{l-1} + \dots + b_l}$, gdzie mianownik nie równa się zeru dla $n \in \Omega_s$.

Twierdzenie 2. Jeżeli zachodzi nierówność

$$\frac{N_0(t-s+1)}{(s+1)(t+1)} \left(1 - \sqrt{\frac{N_0+st-2}{N_0}} \right) \leq b+1 \leq$$

$$\leq \frac{N_0(t-s+1)}{(s+1)(t+1)} \left(1 + \sqrt{\frac{N_0+st-2}{N_0}} \right),$$

to estymator $\frac{t+2}{s+1}n+b$ spełnia warunek

$$(19) \quad \inf_{f \in E} \sup_{N \in \Omega_{N_0}} \frac{1}{N^2} E_N(f(n)-N)^2 = \sup_{N \in \Omega_{N_0}} \frac{1}{N^2} E_N\left(\frac{t+2}{s+1}n+b-N\right)^2 = \\ = \frac{t-s+1}{(s+1)(t+1)}.$$

Dowód. Niech

$$M(n) = \frac{a_0 n^k + \dots + a_k}{b_0 n^l + \dots + b_l} = f(n) + g(n),$$

gdzie $f(n)$ jest wielomianem, a $g(n)$ jest funkcją wymierną, której stopień mianownika jest większy od stopnia licznika. Ponieważ mianownik funkcji $g(n)$ nie równa się zero, dla $n \in \Omega_s$, więc istnieje stała c taka, że $|g(n)| \leq c$ dla $n \in \Omega_s$. Mamy zatem

$$|E_N[g(n)(f(n)-N)]| \leq c E_N|f(n)-N| \leq c E_N^{1/2}(f(n)-N)^2.$$

Stąd otrzymujemy

$$(20) \quad \sup_{N \in \Omega_{N_0}} \frac{1}{N^2} E_N(M(n)-N)^2 = \sup_{N \in \Omega_{N_0}} \frac{1}{N^2} E_N(f(n)-N+g(n))^2 \geq \\ \geq \lim_{N \rightarrow \infty} \frac{1}{N^2} E_N(f(n)-N+g(n))^2 = \lim_{N \rightarrow \infty} \frac{1}{N^2} E_N(f(n)-N)^2.$$

Jeżeli stopień wielomianu $f(n)$ jest większy niż jeden, to z wzorów (16), (20) otrzymujemy

$$\lim_{N \rightarrow \infty} \frac{1}{N^2} E_N(f(n)-N)^2 = \infty.$$

Jeżeli natomiast $f(n) = an+b$, to

$$(21) \quad \sup_{N \in \Omega_{N_0}} \frac{1}{N^2} E_N(M(n)-N)^2 \geq \lim_{N \rightarrow \infty} \frac{1}{N^2} E_N(an+b-N)^2 = \\ = \frac{a^2 s(s+1)}{(t+1)(t+2)} - \frac{2as}{t+1} + 1 \geq \frac{t-s+1}{(t+1)(s+1)}.$$

Ponadto z wzoru (18) otrzymujemy

$$(22) \quad \sup_{N \in \Omega_{N_0}} \frac{1}{N^2} E_N(M(n)-N)^2 \geq \sup_{N \in \Omega_{N_0}} \frac{1}{N^2} E_N\left(\frac{t+2}{s+1}n+b-N\right)^2 = \\ = \frac{t-s+1}{(s+1)(t+1)}.$$

Wzory (21), (22) implikują więc wzór (19), c.b.d.o.

Widzimy zatem, że istnieje cała rodzina estymatorów spełniających wzór (19). W praktyce możemy wybierać b na podstawie dodatkowych informacji.

(i) $b = \frac{t-2s}{s+1}$. Mamy wtedy

$$(23) \quad \hat{m}_1(n) = \frac{t+2}{s+1}n + \frac{t-2s}{s+1},$$

$$E_N(\hat{m}_1(n) - N)^2 = \frac{t-s+1}{(s+1)(t+1)}(N+1)(N-t) \left[1 - \frac{t-s+1}{(s+1)(N+1)} \right].$$

(ii) Gdy wiemy, że na pewno $N \geq \bar{N} > t\sqrt{s}$ i N jest bliskie $\bar{N}$, to bierzemy $b = \frac{(t-s+1)(\bar{N}+1)}{(s+1)(t+1)} - 1$; otrzymujemy wtedy estymator

$$(24) \quad \hat{m}_2(n) = \frac{t+2}{s+1}n + \frac{(t-s+1)(\bar{N}+1)}{(s+1)(t+1)} - 1$$

i wzór

$$E_N(\hat{m}_2(n) - N)^2 = \frac{t-s+1}{(s+1)(t+1)} \left[(N+1)^2 - \frac{s(t+2)}{s+1}(N+1) - \frac{[(t-s+1)(\bar{N}+1)]}{(s+1)(t+1)}(2N - \bar{N} + 1) \right].$$

Porównując $\hat{m}_1(n)$, $\hat{m}_2(n)$ z wzorem (2), mamy

$$E_N(\hat{m}_1(n) - N)^2 < E_N(\hat{N}_c^2 - N)^2, \quad \text{gdy } N > t,$$

$$E_N(\hat{m}_2(n) - N)^2 < E_N(\hat{N}_c^2 - N)^2, \quad \text{gdy } N > \bar{N}.$$

Prace cytowane

- [1] D. G. Chapman, *Inverse, multiple and sequential sample censuses*, *Biometrics* 8 (1952), str. 286-306.
- [2] H. Cramer, *Mathematical methods of statistics*, Princeton University 1946, str. 294.
- [3] C. R. Rao, *Information and the accuracy attainable in the estimation of statistical parameters*, *Bull. Calcutta Math. Soc.* 37 (1945), str. 81-91.
- [4] David Blackwell and M. A. Girshik, *Theory of games and statistical decisions*, New York 1954, str. 294.

INSTYTUT MATEMATYCZNY POLSKIEJ AKADEMII NAUK

Praca wpłynęła 26. 1. 1960

ЧЕН ПИН (Варшава)

О МИНИМАКСОВОЙ ОЦЕНКЕ ЧИСЛЕННОСТИ СОВОКУПНОСТИ

■ РЕЗЮМЕ

Объектом настоящей работы является вопрос об оценке численности совокупности. Д. Чэпмен [1] предложил для этой цели метод обратных выборок и нашел соответствующие оценки. В настоящей статье найдены минимаксовые оценки численности совокупности для функции потери $R(f, N) = \frac{1}{N^2} (f - N)^2$ и исследованы их оптимальные свойства (теоремы 1 и 2). Затем эти оценки сравнены с формулами Чэпмена.

CZEN PIN (Warszawa)

ON THE MINIMAX ESTIMATE OF THE POPULATION SIZE

SUMMARY

The paper deals with the problem of estimating the population size. D. G. Chapman [1] has employed the method of inverse sample censuses and gave some estimates for the population size. This paper gives the minimax estimates for the population size under the loss function $R(f, N) = \frac{1}{N^2} (f - N)^2$. Some of the optimal properties of these estimates are investigated (theorems 1 and 2); later they are compared with the Chapman's estimates.