Aplikace matematiky

Jana Jurečková

Poznámka k volbě rozsahu prvního výběru při Steinově dvouvýběrové metodě

Aplikace matematiky, Vol. 8 (1963), No. 3, 201–205

Persistent URL: `http://dml.cz/dmlcz/102853`

POZNÁMKA K VOLBĚ ROZSAHU PRVNÍHO VÝBĚRU PŘI STEINOVĚ DVOUVÝBĚROVÉ METODĚ

JANA PŘISTOUPILOVÁ

(Došlo dne 26. dubna 1962.)

Práce se týká volby rozsahu prvního výběru při Steinově dvouvýběrové metodě pro intervalový odhad výběrového průměru a přináší řešení (včetně tabulek) pro jistý speciální případ při exponenciální nákladové funkci.

1. ÚVOD

Jde o volbu rozsahu prvního výběru při konstrukci konfidenčního intervalu pro průměr ξ normálního rozložení $N(\xi, \sigma^2)$ s neznámým rozptylem σ^2, jehož délka je shora omezena daným číslem $2d$, Steinovou dvouvýběrovou metodou. Při Steinově metodě je celkový rozsah obou výběrů dán vztahem

$$N = \max\left(m; \left\{\frac{\hat{s}_m^2 t^2}{d^2}\right\}\right), \tag{1}$$

kde m je rozsah prvního výběru; $(m - 1)\,\hat{s}_m^2/\sigma^2$ je náhodná veličina s rozložením χ^2 o $m - 1$ stupních volnosti; t je $100(1 - \frac{1}{2}\alpha)$-procentní kvantil Studentova t rozložení o $m - 1$ stupních volnosti, kde $1 - \alpha$ je požadovaný koeficient spolehlivosti konfidenčního intervalu; $\{a\}$ označuje největší celé číslo menší než a.

Volba rozsahu m předběžného výběru je posuzována z hlediska minimalizace střední hodnoty nákladové funkce. V tabulce 1 je tabelována střední hodnota exponenciální nákladové funkce pro určitý speciální případ. V tabulce 3 jsou tabelovány nejmenší celočíselné hodnoty $\bar{m}$, pro které platí $P\{N > \bar{N} \mid \bar{m}\} < q$ pro některé volby dvojice $\bar{N}$, q.

B. M. SEELBINDER [2] navrhuje volit za rozsah prvního výběru číslo, které minimalizuje střední hodnotu EN celkového počtu pozorování; příslušná nákladová funkce je tedy lineární. Lineární nákladová funkce však často nevystihuje konkrétní situaci. Proto je zde navržena a pro jistý speciální případ vyšetřena exponenciální nákladová funkce. S tím souvisí i návrh J. MOSHMANA [3], který kritizuje Seelbinderův návrh a ve snaze vyhnout se rozsáhlým výběrům navrhuje volit za m hodnotu, která minimalizuje jistou lineární kombinaci střední hodnoty EN a 95procentního kvantilu $N_{0,95}$ náhodné veličiny N.

2. VÝPOČET TABULEK

Tabulka 1. Uvažujme nákladovou funkci tvaru

$$(2) \qquad \Phi(N) = e^{cN}, \quad c > 0,$$

kde různým hodnotám koeficientu c odpovídají různé rychlosti růstu nákladů v závislosti na počtu pozorování. Střední hodnota nákladové funkce (2) vzhledem k (1) je

$$(3) \qquad E\{e^{cN}\} = e^{cm} . P\{\chi^2(m-1) < m(m-1)\lambda\} +$$
$$+ \sum_{k=m+1}^{\infty} e^{ck} . P\{(k-1)(m-1)\lambda \leqq \chi^2(m-1) < k(m-1)\lambda\} =$$
$$= e^{cm} . P\{\chi^2(m-1) < m(m-1)\lambda\} +$$
$$+ \sum_{k=m+1}^{M} e^{ck} . P\{(k-1)(m-1)\lambda \leqq \chi^2(m-1) < k(m-1)\lambda\} + R_M,$$

kde $\lambda = d^2/t^2\sigma^2$; $\chi^2(m-1)$ označuje náhodnou veličinu s χ^2 rozložením o $m-1$ stupních volnosti; $M > 0$ je celé číslo. Analogickým způsobem jako Stein počítá horní a dolní hranici pro střední hodnotu EN, lze stanovit horní a dolní hranici pro zbytek R_M. Pro $m > 2c/\lambda + 1$ platí

$$(4) \qquad R_M = K_c . e^{\Theta . c},$$

kde

$$K_c = \left[\frac{(m-1)\lambda}{(m-1)\lambda - 2c}\right]^{(m-1)/2} . P\{\chi^2(m-1) \geqq M[(m-1)\lambda - 2c]\}$$

a $0 < \Theta < 1$. Z výpočtu zároveň vyplývá, že řada, obsažená ve (4), pro $m = 2c/\lambda + 1$ diverguje k $+\infty$.

Tabulka 1. Hodnoty $E\{e^{cN}\}$ pro některá m; $\alpha = 0{,}05$; $\lambda = 0{,}1$

$c = 0{,}1$			$c = 0{,}2$			$c = 0{,}3$		
m	$E\{e^{cN}\}$	A_M	m	$E\{e^{cN}\}$	A_M	m	$E\{e^{cN}\}$	A_M
5	4,265	0,09	9	18,732	0,27	11	123,653	3,64
7	3,665	0,05	11	16,667	0,08	13	115,951	0,80
9	3,673	0,03	13	17,971	0,03	15	117,107	0,21
11	3,701	0,02	15	20,469	0,03	17	176,944	0,06
13	4,058	0,02	17	31,036	0,01			
15	4,667	0,01						

Tabulka 1 obsahuje pro speciální případ $\lambda = 0{,}1$ a pro $c = 0{,}1; 0{,}2; 0{,}3$ a některá m aproximace střední hodnoty $E\{e^{cN}\}$, které získáme tímto způsobem: Dosaďme do (3) za R_M

$$(5) \qquad R_M = K_c . \frac{e^c + 1}{2};$$

vzhledem ke (4) je tedy Θ voleno tak, aby $e^{\Theta c} = \frac{1}{2}(e^c + 1)$. Chyba hodnoty (5) a tudíž i tabelované hodnoty $E\{e^{cN}\}$ je shora omezeno číslem

$$A_M = K_c \cdot \frac{e^c - 1}{2} .$$

Volba čísla M, postupně pro $c = 0{,}1$; $0{,}2$; $0{,}3$ je $M = 20$; 30; 33. Tabulka je vypočtena pro speciální případ $\lambda = 0{,}1$, kterého se týká též Moshmanův článek.

V tabulce 2 je pro srovnání uvedeno, k jakým optimálním hodnotám m vedou ve speciálním případě $\lambda = 0{,}1$ jednotlivá kritéria: lineární nákladová funkce, Moshmanovo kritérium a exponenciální nákladová funkce.

Tabulka 2. Srovnání jednotlivých kritérií; $\lambda = 0{,}1$

	m_{opt}
lineární nákladová funkce	3
Moshmanovo kritérium	6
exponenciální nákladová funkce; $c = 0{,}1$	8–10
exponenciální nákladová funkce; $c = 0{,}2$	10–12
exponenciální nákladová funkce; $c = 0{,}3$	12–14

V tabulce 3 jsou tabelovány nejmenší celočíselné hodnoty $\overline{m}$, pro které platí

$$P\{N \geqq \overline{N} \mid \overline{m}\} < q \tag{6}$$

pro $q = 0{,}01$ a $0{,}05$ a pro některé hodnoty $\overline{N}$. Tabulka 3 je užitečná, volíme-li m např. tak, aby EN bylo minimální při splnění podmínky (6).

Tabulka 3. Hodnoty $\overline{m}$ pro některé hodnoty $\overline{N}$

a) $\alpha = 0{,}05$ a $d/\sigma = 0{,}5$; $q = 0{,}05$ a $0{,}01$

$\overline{N}$	$\overline{m}$ ($q = 0{,}05$)	$\overline{m}$ ($q = 0{,}01$)	$\overline{N}$	$\overline{m}$ ($q = 0{,}05$)	$\overline{m}$ ($q = 0{,}01$)
30	18	29	75	5	7
40	10	15	80	5	7
45	9	12	85	5	6
50	8	11	90	5	6
55	7	9	95	5	6
60	6	8	100	5	5
65	6	8	110	4	5
70	6	7	120	4	5

b) $\alpha = 0{,}05$; $d/\sigma = 0{,}01$; $q = 0{,}05$ a $0{,}01$

$\bar{N}$	$\bar{m}\ (q = 0{,}05)$	$\bar{m}\ (q = 0{,}01)$	$\bar{N}$	$\bar{m}\ (q = 0{,}05)$	$\bar{m}\ (q = 0{,}01)$
450	202	—	500	84	160
460	183	—	510	77	139
470	145	—	520	68	122
480	121	202	530	60	108
490	102	187	540	54	97

Literatura

[1] *Ch. Stein*: A two-sample test for a linear hypothesis whose power is independent of the variance. Ann. Math. Statist. 16 (1945), s. 243–249.

[2] *B. M. Seelbinder*: On Stein's two-stage sampling scheme. Ann. Math. Statist. 24 (1953), s. 640–649.

[3] *J. Moshman*: A method for selecting the size of the initial sample in Stein's two-sample procedure. Ann. Math. Statist. 29 (1958), s. 1271–1275.

[4] *E. C. Molina*: Poisson's exponential binomial limit. D. van Nostrand comp. inc. New York (1942).

Резюме

ЗАМЕЧЕНИЕ К ВЫБОРУ ОБЪЕМА ПЕРВОЙ ВЫБОРКИ ПРИ ДВУХВЫБОРОЧНОМ МЕТОДЕ СТАЙНА

ЯНА ПРЖИСТОУПИЛОВА (Jana Přistoupilová)

В статье анализируется проблема выбора объема первой выборки при двухвыборочном методе Стайна. Предполагается, что функция стоимости – показательная. Затем для частного случая даны в форме таблиц: среднее значение функции стоимости, оптимальные значения объема первой выборки, определенные при помощи трех различных критериев (для сравнения с другими методами) и значения объема первой выборки, гарантирующие, что общий объём обеих выборок превосходит данный предел только с малой вероятностью.

Summary

NOTE ON THE CHOICE OF MAGNITUDE OF THE FIRST SAMPLE IN STEIN'S TWO-SAMPLE METHOD

JANA PŘISTOUPILOVÁ

The problem treated in this paper is the choice of the first sample in Stein's two-sample method. It is assumed that the cost-function is exponential. There are given, in form of tables, the expected values of the cost-function, optimal values of the size of the first sample determined by three different criteria (for comparing with other methods), and the values of the first sample; these ensure that the total size of both samples exceeds the given boundary with only small probability.

Adresa autora: Jana Přistoupilová, Katedra vědeckého programování Vysoké školy ekonomické, nám. G. Klimenta 4, Praha 3 — Žižkov.