Václav Fabian

Poznámka k Halperinove metodě prokládání přímky, jsou-li obě proměnné pozorovány s chybou

Aplikace matematiky, Vol. 8 (1963), No. 3, 197–(198a),199–200

Persistent URL: http://dml.cz/dmlcz/102852

POZNÁMKA K HALPERINOVĚ METODĚ PROKLÁDÁNÍ PŘÍMKY, JSOU-LI OBĚ PROMĚNNÉ POZOROVÁNY S CHYBOU

VÁCLAV FABIAN

(Došlo dne 16. března 1962.)

Z Halperinových výsledků je odvozen simultánní intervalový odhad hodnot lineární funkce, jejíž parametry jsou odhadovány v experimentu, v němž pozorování funkčních hodnot i argumentů jsou zatížena náhodnými chybami.

1. Úvod a shrnutí. Tento článek obsahuje menší doplněk k práci [1]. Matematicky jde o velmi jednoduchý výsledek, který autor uveřejňuje jen pro jeho praktickou užitečnost a proto, že přes svou jednoduchost by patrně tvrzení i jeho důkaz nebyly samy o sobě zřejmé každému experimentátoru používajícímu Halperinovy metody.

V článku [1] je studován problém odhadu neznámých parametrů α, β lineární funkce $r(u) = \alpha + \beta u$ na základě posloupnosti dvojic $[x_1, y_1], \ldots, [x_n, y_n]$, kde každé y_i je odhadem funkční hodnoty $r(u_i) = \alpha + \beta u_i$ v jistém bodě u_i. Číslo u_i je odhadováno hodnotou x_i.

Poznamenáme, že teorie regrese se obvykle zabývá případem, v němž je $u_i = x_i$, tj. v němž známe hodnoty nezávisle proměnné přesně. Obecnější problém (tzv. strukturální relace), do něhož spadá práce [1] a v němž se připouští chyba i v odhadu hodnot nezávislé proměnné, je v praxi častější, ale bohužel i obtížnější k řešení.

Halperinova metoda je založena ještě na další posloupnosti $z_1, z_2, \ldots, z_n$ odhadů hodnot $u_1, u_2, \ldots, u_n$, nezávislých na $[x_1, y_1], \ldots, [x_n, y_n]$.

Předpoklady lze nyní shrnout takto: x_i, y_i, z_i jsou náhodné proměnné, jejichž očekávané hodnoty splňují vztah $Ey_i = r(Ex_i)$. Označíme-li $\eta_i = y_i - Ey_i$, $\xi_i = x_i - Ex_i$, jsou $[\xi_1, \eta_1], \ldots, [\xi_n, \eta_n]$ vzájemně nezávislé identicky distribuované náhodné vektory s dvojrozměrným normálním rozložením pravděpodobností a jsou nezávislé na $z_1, z_2, \ldots, z_n$.

Poznamenáme, že ve formálním vyjádření předpokladů se nevyskytuje požadavek, aby z_i byly dobrými odhady hodnot Ex_i. Čím je však lépe tento požadavek splněn, tím užitečnější je odhad, který metoda poskytne; jsou-li z_i špatnými odhady čísel Ex_i (Ex_i jsou přesné hodnoty nezávisle proměnné), může být odhad, který metoda dává, neužitečný (podrobněji viz [1]).

Symbolem $F_\alpha(f_1, f_2)$ označme nyní $(1 - \alpha)$-kvantil F-rozložení s f_1 a f_2 stupni volnosti (tabelované např. v [2], tabulka 12) a položme $R = 2F_\gamma(2, n - 2)/(n - 2)$. Pišme formálně[1])

$$\bar{a} = \frac{1}{n}\sum_{i=1}^{n} a_i\,,$$

$$S_{a,b} = \sum_{i=1}^{n}(a_i - \bar{a})(b_i - \bar{b}) = \frac{n\sum_{i=1}^{n} a_i b_i - (\sum_{i=1}^{n} a_i)(\sum_{i=1}^{n} b_i)}{n}\,,$$

$$\Phi_{a,b} = S_{az}S_{bz} + R(S_{az}S_{bz} - S_{ab})\,;$$

dosazujíce za písmena a, b v předešlých dvou vztazích písmena x, y, z, dostaneme vztahy určující Φ_{xx}, Φ_{xy}, Φ_{yy}, $\bar{x}$, $\bar{y}$, $\bar{z}$.

Halperin ukázal, že s pravděpodobností $1 - \gamma$ je

$$(1.1) \qquad n(\bar{y} - \alpha - \beta\bar{x})^2 + \Phi_{xx}(\beta^* - \beta)^2 \leqq \beta^{*2}\Phi_{xx} - \Phi_{yy}\,,$$

kde

$$(1.2) \qquad \beta^* = \Phi_{xy}/\Phi_{xx}\,.$$

Tento výsledek umožňuje simultánní odhad neznámých parametrů α a β. Jak ukážeme, umožňuje též simultánní intervalový odhad všech hodnot $r(u)$ funkce r.

Označme symbolem $r^*(u)$ bodový odhad $\bar{y} + \beta^*(u - \bar{x})$ funkční hodnoty $r(u) = \alpha + \beta u$ a položme

$$(1.3) \qquad d(u) = \begin{cases} \left[\dfrac{\Phi_{xy}^2 - \Phi_{xx}\Phi_{yy}}{n\Phi_{xx}^2}(\Phi_{xx} + n(u - \bar{x})^2)\right]^{1/2}, & \text{je-li } \Phi_{xx} > 0\,, \\ +\infty & \text{jinak}\,. \end{cases}$$

Pak vztah

$$(1.4) \qquad |r(u) - r^*(u)| \leqq d(u) \text{ pro všechna } u\,,$$

plyne z (1.1) a je tedy splněn s pravděpodobností aspoň $1 - \gamma$.

Intervalovým odhadem je tedy pás mezi oběma funkcemi $r^*(u) - d(u)$ a $r^*(u) + d(u)$.

2. Poznámka. Výsledek metody je neužitečný, je-li $\Phi_{xx} \leqq 0$. Pravděpodobnost tohoto nežádoucího jevu je, zhruba řečeno, tím menší, čím jsou z_i lepšími odhady čísel Ex_i (ve skutečnosti záleží jen na chování posloupností $l_i = (z_i - a)/b$, kde a a b jsou volena tak, aby $\Sigma l_i = 0$, $\Sigma l_i^2 = 1$; viz [1]).

Bylo by ovšem prakticky velmi cenné, kdyby se podařilo metodu modifikovat (např. v sekvenční) tak, aby přinesla užitečné rozhodnutí vždy.

[1]) Vztahy jsou poněkud upraveny, aby skýtaly výhodnější návod pro výpočet.

Vztah pro Φ_{ab} na str. 198 má být

$$\Phi_{ab} = S_{az}S_{bz}/S_{zz} + R(S_{az}S_{bz}/S_{zz} - S_{ab});$$

3. Důkaz.

3.1. Toto pomocné tvrzení lze odvodit z prvé derivace pravé strany: Pro $b > 0$, $a \geqq 0$ je

$$\max_{t \in \langle 0, (a/b)^{1/2} \rangle} \left[t + (a - bt^2)^{1/2} \right] = \sqrt{\left(\frac{a(1 + b)}{b} \right)} . \tag{3.1.1}$$

3.2. Označme $h(u) = r^*(u) - r(u)$, $t = |\alpha + \beta\bar{x} - \bar{y}|$, $A = (\beta^{*2}\Phi_{xx} - \Phi_{yy})/\Phi_{xx}$, $B = n/\Phi_{xx}$. Z (1.1) plyne

$$(\beta^* - \beta)^2 \leqq A - Bt^2 , \tag{3.2.1}$$

aspoň je-li $\Phi_{xx} > 0$, což můžeme zřejmě předpokládat, neboť případ $\Phi_{xx} \leqq 0$ je triviální. Z (1.1) plyne také

$$0 \leqq t \leqq (A/B)^{1/2} . \tag{3.2.2}$$

Označíme-li $A(u) = (u - \bar{x})^2 A$ a $B(u) = (u - \bar{x})^2 B$, plyne z (3.2.1)

$$|\beta^* - \beta| \, |u - \bar{x}| \leqq (A(u) - B(u)\, t^2)^{1/2} .$$

Protože je $|h(u)| \leqq |\beta^* - \beta| \, |u - \bar{x}| + t$, je dále

$$|h(u)| \leqq t + (A(u) - B(u)\, t^2)^{1/2} ;$$

odtud a z (3.1.1) a (3.2.2) dostáváme

$$h^2(u) \leqq \frac{A(u)\,(1 + B(u))}{B(u)} = \frac{A}{B}(1 + B(u)) .$$

Odtud, z definic čísel A, B, $B(u)$ a ze vztahu (1.2) plyne (1.4) jednoduchou algebraickou úpravou.

Literatura

[1] *Max Halperin*: Fitting of straight lines and prediction when both variables are subject to error. Journal Amer. Statist. Ass., *56* (1961), 657—669.

[2] *Jaroslav Janko*: Statistické tabulky. NČSAV, Praha 1958.

Резюме

ЗАМЕЧАНИЕ К МЕТОДУ ГАЛПЕРИНА ПРОКЛАДЫВАНИЯ ПРЯМОЙ В СЛУЧАЕ, КОГДА ОБЕ ПЕРЕМЕННЫЕ НАБЛЮДАЮТСЯ С ПОГРЕШНОСТЬЮ

ВАЦЛАВ ФАБИАН (Václav Fabian)

Из результата Галперина (см. [1]) выведен симультанный доверительный интервал функциональных значений $r(u)$ (для всех u) изучаемой линейной функ-

ции r. Оценка дана соотношением (1.4), причем $d(u)$ определено в (1.3) и $r^*(u) = \bar{y} + \beta^*(u - \bar{x})$. Остальные символы имеют то же значение, как в статье [1], вероятность события (1.4) равна по крайней мере $1 - \gamma$.

Summary

NOTE ON HALPERIN'S METHOD OF FITTING STRAIGHT LINES WHEN BOTH VARIABLES ARE SUBJECT TO ERROR

VÁCLAV FABIAN

From Halperin's result (*cf.* [1]) there is derived a simultaneous interval estimate of the function values $r(u)$ (for all u) of the considered linear function r. This estimate is given in (1.4) with $d(u)$ defined by (1.3) and with $r^*(u) = \bar{y} + \beta^*(u - \bar{x})$, other notation remaining as in [1]; the probability that the event (1.4) will occur is at least $1 - \gamma$.

Adresa autora: Dr. *Václav Fabian* C. Sc., Výzkumný ústav matematických strojů, Loretánské nám. 3, Praha 1 — Hradčany.