А. А. Василев*

ЧИСЛЕНИЯТ КРИТЕРИЙ ЗА ОПТИМАЛНОСТ ПРИ ИЗБОРА НА КРИТЕРИЙ ЗА ПРОВЕРКА НА СТАТИСТИЧЕСКИ ХИПОТЕЗИ ПРИ ЗАДАЧИТЕ ЗА ОБРАБОТКА НА ЧИСЛОВИ ИКОНОМИЧЕСКИ ДАННИ

При обосновката на много икономически решения се използват статистически процедури за обработка на цифрови икономически данни, в частност критерии за проверка на статистически хипотези. При това изборът на оптимална статистическа процедура на такава обработка за решаване на конкретна практическа задача се отнася към напълно нерешимите задачи на приложната статистика.

Проверката на две (за улеснение) алтернативни статистически хипотези за значението на който и да е признак се прилага при следните условия, характерни при задачите за обработка на цифрови икономически данни: обемът при избора на значение на признака n > 30; нулевата хипотеза H_0 е проста; алтернативната хипотеза H_1 е проста или сложна; сведения за разпределението на вероятностите на изследвания признак при нулева и алтернативна хипотези отсъстват; за проверка на статистическите хипотези може да се използва множеството $K=\{k_1, k_2, \dots, k_l, \dots, k_r\}$ параметрични, непараметрични и статистически критерии с номинално равнище на значимост при $\alpha_{ном}$.

Необходимо е да се избере оптимален критерий за проверка на статистическите хипотези. Класическият аналитичен критерий за оптималност на Бейес не може да бъде използван за решаване на поставената задача, даже и с подмножествата на параметричните статистически критерии на множеството K, тъй като матрицата на глобите на статистическите критерии $k_l \in K$, вероятностите на хипотезите и намирането на плътността за разпределение на вероятностите $f(x|H_0)$ и $f(x|H_1)$ са неизвестни.

Класическият аналитичен критерий за оптималност на Нейман-Пирсон може да се използва само за приблизително решаване на поставената задача в предположенията за $f(x|H_0)$ и $f(x|H_1)$. При това положение решаването на неравенството $\alpha_{факт}(k_l) \le \alpha_{ном}$, където $\alpha_{факт}(k_l)$ е равнището за значимост на критерия за проверка на статистически хипотези $k_l \in K$, не е гарантирано.

---

* Русия, гр. Твер, Тверски държавен университет.

Редица предположения, приети при разработването на аналитични критерии за оптималност на Хюер[1] и Хампел,[2] също ограничават възможностите за тяхното коректно използване при избора на оптимален статистически критерий за проверка на статистическите хипотези в разглежданата задача. Към тях се отнасят: *първо,* необходимостта от наличие на достатъчно точни данни за функционирането при разпределението на вероятностите от случайна величина, тъй като изборът на оптимален статистически критерий е възможен само в безкрайно малък параметър на разглежданото параметрично разпределение; *второ,* допускането за симетрия на плътността при разпределението на вероятностите; *трето,* допускането за асимптоматична нормалност за разпределението на статистическите вероятности при статистическия критерий; *четвърто,* необходимостта от наличието на априорни сведения за дела на наблюденията в избраните данни.

Условията за проверка на статистическите хипотези се отразяват по-адекватно при обработката на цифровите икономически данни, числените критерии за оптималност въз основа на модела на Монте-Карло. Един от тези критерии за оптималност се разглежда като сбор от праговите точки, равнището на значимост и мощност на статистическия критерий, усреднено спрямо множеството от комбинации, зададени от апроксимации на условните разпределения за вероятностите със случайна величина при нулевата и алтернативната хипотези.[3] Този числен критерий се формулира по следния начин:

$$k^0 = \arg\max_{k_l \in K} \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} \left[\varepsilon^*_{\alpha_i}\left(k_l, \alpha_{доп}, a_i\right) \times \varepsilon^*_{\beta_{ij}}\left(k_l, \alpha_{доп}, \beta_{доп}, a_i, b_j\right) \right],$$

където: *m* е количеството на апроксимацията за разпределение на вероятности $f\left(x \middle| H_0\right)$ при нулева хипотеза; *n* – количеството на апроксимацията за разпределение на вероятности $f\left(x \middle| H_1\right)$ при алтернативна хипотеза; $\alpha_{доп}$ - пределно допустимо значение на равнището за значимост на критерия; $\beta_{доп}$ - пределно допустимото значение на вероятността на грешките от втори вид; a_i, $i = \overline{1, m}$, i - i-е - значението на параметъра за положение на *a* апроксимация за разпределение на вероятностите $f\left(x \middle| H_0\right)$ при нулева хипотеза от района на

---

[1] *Хьюбер, Дж. П.* Робастность в статистике (пер. с англ.). Москва: Мир, 1984, 304 с.

[2] *Хампель, Ф., Э. Рончетти, П. Рауссеу, В. Штаэль.* Робастность в статистике. Подход на основе функций влияния (пер. с англ.). Москва: Мир, 1989, 512 с.

[3] *Васильев, А. А.* Критерии оптимальности критериев проверки статистических гипотез в задачах обработки числовых экономических данных. - Факторы развития экономики России: Материалы Международной научно-практической конференции, 19-20 апреля 2006 г., ч. II,. Тверь: Тверской государственный университет, 2006, с. 283-288.

което и да било параметрично разпределение на вероятностите; b_j - $j=\overline{1,n}$, j - j-е - значението на параметъра b, който определя разстоянието между центровете на апроксимация за разпределението $f\left(x\middle|H_0\right)$ и $f\left(x\middle|H_1\right)$, описващи нулевата и алтернативна хипотези; $\varepsilon^*_{\alpha_i}\left(k_l,\alpha_{доп},a_i\right)$ - оценяваното чрез метода на Монте-Карло значение на праговите точки и равнището на значимост на критерия за проверка на статистическите хипотези $k_l \in K$ с номинално равнище на значимост $\alpha_{ном}$ за i-и апроксимация за разпределение на вероятностите $f\left(x\middle|H_0\right)$ при нулева хипотеза от посоченото параметрично разпределение на вероятностите при пределно допустимо значение на нивото на значимостта $\alpha_{доп}$; $\varepsilon^*_{\beta_{ij}}\left(k_l,\alpha_{доп},\beta_{доп},a_i,b_j\right)$ - оценяваното по метода на Монте Карло, значение на праговата точка на мощността на критерия за проверка на статистическите хипотези $k_l \in K$ с номинално равнище на значимост $\alpha_{ном}$ за i-и апроксимация за разпределение на вероятностите $f\left(x\middle|H_0\right)$ при нулева хипотеза от посоченото параметрично разпределение на вероятностите и за j-о значение на разстоянието между центъра на апроксимация за разпределение $f\left(x\middle|H_0\right)$ и $f\left(x\middle|H_1\right)$, описващи нулевата и алтернативната хипотези.

Под прагова точка за равнище на значимостта разбираме най-малкото значение на частта с аномалните наблюдения, при които фактическото значение на равнището на значимостта на статистическия критерий като функция от частта на аномалните наблюдения в който и да било модел за грешки превишава допустимите значения. Прагова точка за мощност е най-малкото значение на частта с аномалните наблюдения, при които фактическото значение на вероятността на грешките от втори ред за статистически критерий като функция от частта на аномалните наблюдения в който и да било модел за грешки превишава допустимите значения. При това допустимото значение на равнището за значимост и вероятността на грешките от втори ред се задават статистически, изхождайки от неговата представа за оптимална устойчивост на статистическия критерий по дадени характеристики в конкретна задача за обработка на цифрови икономически данни.

Разгледаният цифров критерий за оптималност позволява да се реши двойнокритерийната задача за избор на окончателно оптимален статистически критерий в задачите за обработка на цифрови икономически данни с практически всякакви разпределения на вероятностите.