УДК 519.2, 519.6, 658.562.0121.7

ПРО ПОБУДОВУ ОПЕРАТИВНОЇ ХАРАКТЕРИСТИКИ СТАТИСТИЧНОГО КОНТРОЛЮ ЯКОСТІ ПРОМИСЛОВОЇ ПРОДУКЦІЇ НА БАЗІ БІНОМІАЛЬНОГО ПОСЛІДОВНОГО КРИТЕРІЮ ВІДНОШЕННЯ ЙМОВІРНОСТЕЙ (ПКВІ)

С.М. КРАСНИТСЬКИЙ, М.С. КРАСНИТСЬКИЙ, Є.В. ЗАРЖИЦЬКИЙ, Л.В. ХОМЕНКО

Київський національний університет технологій та дизайну

У роботі запропоновано методику обчислення значень оперативної характеристики і очікуваного обсягу вибірки поширеної процедури послідовного біноміального статистичного контролю якості виробничої продукції, що є більш ефективною і є нескладною для розробки відповідних комп'ютерних програм, ніж методики пропоновані в літературі [1,2]

Важливою характеристикою процедури статистичного контролю якості продукції є так звана *оперативна характеристика* (**ОХ**) [2 – 4]. За означенням, це функція $P(\theta)$, $\theta \in [0,1]$, яка дорівнює ймовірності того, що при використанні цього алгоритму вибіркового контролю партія виробів буде визнана придатною, якщо рівень браку у партії виробів (або ймовірність створення бракованого виробу у технологічному процесі) дорівнює θ. Практичне значення має можливість ефективного обчислення значень оперативної характеристики і побудова її графічного зображення. Важливим є і знання очікуваного потрібного обсягу вибірки одиниць виробничої продукції, що дозволяє у першому наближенні планувати витрати на проведення обстежень її якості.

Об'єкти та методи дослідження.

Об'єктом дослідження у роботі є процес обчислення значень оперативної характеристики процедури статистичного контролю якості промислової продукції, що ґрунтується на біноміальному послідовному критерії відношення ймовірностей.

Результати та їх обговорення.

Як відомо, усі статистичні методи вибіркової перевірки якості промислової продукції можна розділити на два класи: ті, що використовують вибірки *фіксованого* обсягу, і ті, що використовують схему *послідовного* відбору. У першому випадку контролю підлягає вся вибірка, що складається з заданої кількості виробів. При *послідовному* відборі процес контролю припиняється, якщо обстежена *частина* вибірки вже містить достатньо інформації, щоб прийняти або відхилити гіпотезу H_0 = «партія виробів задовольняє прийнятним вимогам якості». Якщо інформації не вистачає, то контроль триває. Таким чином, при послідовному контролі якості продукції контролер на кожному кроці може прийняти одне з трьох рішень: «H_0 приймається», «H_0 відхиляється», «потрібно продовжувати процедуру контролю». Підстави, на яких приймаються належні рішення, ґрунтуються на відповідних положеннях теорії перевірки статистичних гіпотез [1,2, 5,6].

Нехай $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$ — параметрична сім'я розподілів ймовірностей на вимірному просторі (X, $\mathcal{X}$) [2,3]. Тут θ — параметр, Θ — параметрична множина. (Мається на увазі, що виконується деякий стохастичний експеримент, результати якого представляються елементами множини X; останні з'являються відповідно до розподілу ймовірностей P_θ, що визначені на множині підмножин X, яка вище позначена як X. Параметр θ невідомий.) Нехай θ_0, θ_1— елементи Θ, $\theta_0 \neq \theta_1$. Припустимо, йдеться про перевірку гіпотези

$$H_0 = \{\theta = \theta_0\} \tag{1}$$

проти альтернативи

$$H_1 = \{\theta = \theta_1\} \tag{2}$$

(проста гіпотеза і проста альтернатива [1,6]). Серед різних можливих схем перевірки H_0 проти H_1, виділимо відому схему, що ґрунтується на відношенні правдоподібності [1,2,5,6] (нижче наводяться необхідні означення). Для вибірок фіксованого обсягу оптимальні властивості таких схем базуються на фундаментальній лемі Неймана — Пірсона [1,63]; використання відношення правдоподібності у послідовному відборі започаткував Вальд [5].

У загальних рисах згадана схема при послідовному відборі полягає у такому. Припустимо, що з множини X відповідно до розподілу P_θ послідовно обираються елементи $x_1, x_2, \ldots$. Нехай $f(x_i;\theta)$ дорівнює ймовірності одержати значення $x_i \in X$ як i- го вибіркового значення, $i = 1,2,\ldots$, якщо розподіли з сім'ї $\mathcal{P}$ є дискретними або щільності ймовірності в точці x_i, якщо остання існує — все за умови, що значення параметра розподілу дорівнює θ. Складемо так зване відношення правдоподібності λ_m:

$$\lambda_m = \prod_{i=1}^{m} \frac{f(x_i;\theta_1)}{f(x_i;\theta_0)},\ m \geq 1. \tag{3}$$

Зафіксуємо значення ймовірності помилки I роду α і ймовірності помилки II роду β [1,2,6]. Далі розшукують дві такі константи A та B, що після m спостережень подальші дії визначаються правилами:

1) якщо $\lambda_m \leq B$, то припинити випробування і прийняти H_0;
2) якщо $\lambda_m \geq A$, то припинити випробування і прийняти H_1;
3) якщо $B < \lambda_m < A$, то продовжити випробування.

Далі подану вище схему перевірки статистичних гіпотез будемо позначати для скорочення ПКВІ (послідовний критерій відношення ймовірностей).

Доведено [1,2], що при заданих α, β сталі A та B повинні задовольняти нерівності

$$A \leq \frac{(1-\beta)}{\alpha}, \tag{4}$$

$$B \geq \frac{\beta}{(1-\alpha)} \tag{5}$$

Якщо покласти, що A, B рівним правим частинам нерівностей (4), (5) відповідно, то хоча ймовірності помилок першого і другого роду не будуть дорівнювати α та β, але для досить малих α і β відмінностями між цими величинами і реальними ймовірностями помилок першого і другого роду можна знехтувати [1,2].

У цій роботі нас цікавитиме випадок, коли $\Theta = \{\theta: 0 \leq \theta \leq 1\}$, $\theta_0 < \theta_1$, кожне x_i дорівнює 0 або 1, $f(1;\theta) = \theta$, $f(0;\theta) = 1 - \theta$. Відповідну процедуру ПКВІ будемо називати біноміальною.

Наведені співвідношення можна інтерпретувати так. Є партія виробів, частка бракованих серед яких дорівнює θ. Бажано перевірити гіпотезу $\{\theta \leq \theta_0\}$ проти альтернативи $\{\theta > \theta_0\}$, де θ_0 — деяке задане число.

Хоча такі гіпотези і не є простими, але практично [1,2 2, 7] відповідну перевірку можна звести до ситуації простої гіпотези з простою альтернативою, задавши число θ_1 таке, що

$$\theta_0 < \theta_1, \tag{6}$$

і перевіряючи H_0 проти H_1 (останні задаються рівностями (1) і (2)) таким чином, щоб імовірність прийняти H_0 за умови, що $\theta \le \theta_0$, була б не менша за $1 - \alpha$, а імовірність прийняти цю гіпотезу за умови, що $\theta \ge \theta_1$, була б не більша за β. (При цьому вважається, що прийняття H_0 свідчить на користь гіпотези $\{\theta \le \theta_0\}$[1,2].)

Важливою характеристикою процедури статистичного контролю якості продукції, що дає можливість , зокрема, обчислити щойно вказані ймовірності, є так звана *оперативна характеристика* (**ОХ**) [2–47]. За означенням, це функція $P(\theta)$, $\theta \in [0,1]$, яка дорівнює ймовірності того, що при використанні даного алгоритму вибіркового контролю партія виробів буде визнана придатною, якщо рівень браку у партії виробів (або ймовірність створення бракованого виробу у технологічному процесі) дорівнює θ. Такою характеристикою завжди повинна супроводжуватися процедура вибіркового контролю, що дістає офіційний статус [3,4]. Типовий графік функції $P(\theta)$ представлено на рис.1.

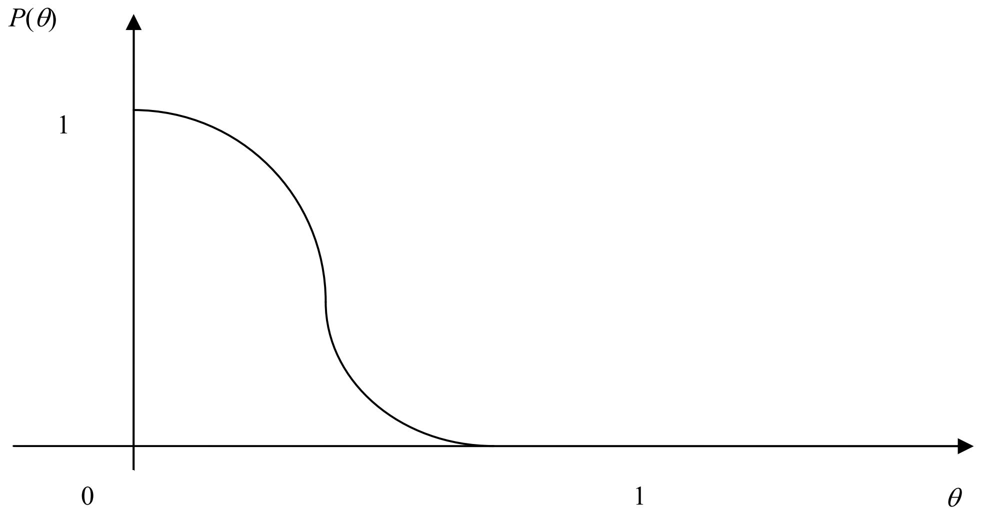

Рис. 1. **Графічне зображення типової оперативної характеристики процедури вибіркового контролю**

У цій роботі запропоновано алгоритм обчислення значень оперативної характеристики біноміального ПКВІ, що є більш ефективним, ніж ті, що звичайно пропонуються у літературі [1,2] і є нескладним для розробки відповідних комп'ютерних програм.

Відомо [1,2], що ОХ біноміального ПКВІ з практично достатньою точністю задається таким виразом:

$$P(\theta) = \frac{\left(\frac{1-\beta}{\alpha}\right)^h - 1}{\left(\frac{1-\beta}{\alpha}\right)^h - \left(\frac{\beta}{1-\alpha}\right)^h}, \tag{7}$$

де $h = h(\theta)$ – корінь рівняння

$$\frac{1-[(1-\theta_1)/(1-\theta_0)]^h}{(\theta_1/\theta_0)^h-[(1-\theta_1)/(1-\theta_0)]^h}=\theta. \qquad (8)$$

У літературі [1,2] є рекомендації з приводу обчислення значень *P(θ)на зразок* того, що «змінюючи значення h, з рівності (8) можна отримувати відповідні пари значень (θ, $h(\theta)$). Далі слід виконати підстановки значень $h(\theta)$ при всіх розглядуваних θ у рівність (7)». Недостатня практичність цієї рекомендації полягає саме в незручності дій щодо перебору значень h, наприклад, через необмеженість множини, у якій слід виконувати перебір. При цьому, апріорно є невідомими і крок зміни параметра h, потрібний для досягнення належного рівня дискретності параметра θ, і кількість таких кроків. У даній роботі пропонується такий (альтернативний) спосіб дій. Виконується перебір не по змінній $h \in(-\infty, +\infty)$, а по достатньо щільній сітці значень $\theta \in[0,1]$, і при кожному такому значенні розв'язується рівняння (8). Для забезпечення вказаного процесу при кожному θ спочатку виконується відокремлення кореня цього рівняння на підставі міркувань, що наводяться нижче. Після згаданого відокремлення для розв'язку (8) можна застосувати будь-який з числових методів розв'язання нелінійних рівнянь [8,9]. (Зокрема, у комп'ютерній програмі, що створена авторами спеціально для числового та графічного визначення ОХ, застосовується класичний метод хорд (січних) [8,9] як такий, що потребує мінімуму обчислень. Виграш у швидкості, який можна мати при застосуванні інших методів, у цій ситуації не має практичного значення). Розглянемо цей процес детальніше.

Введемо функцію $g(h)$, що залежить ще від двох параметрів a, b, за допомогою рівності

$$g(h)=\begin{cases}\dfrac{1-b^h}{a^h-b^h}, & h\neq 0\\[2mm] -\dfrac{\ln b}{\ln a-\ln b}, & h=0\end{cases}, \qquad (9)$$

вважаючи параметри a, b такими, що

$$a > 1,\ 0 < b < 1 \qquad (10)$$

(функція $g(h)$ для $h = 0$ визначена як межах при $h \to 0$ виразу $g(h)$, $h \neq 0$).

Нижче буде описано алгоритм розв'язку рівняння

$$g(h) = \theta,\quad 0 < \theta < 1, \qquad (11)$$

яке, зважаючи на нерівність (6), є узагальненням рівняння (8).

Введемо функції $g_1(h)$, $h \in(-\infty, 0)$ та $g_2(h)$, $h \in(0, +\infty)$ за допомогою рівностей

$$g_1(h)=\frac{b^h-1}{b^h},\ h\in(-\infty, 0), \qquad (12)$$

$$g_2(h)=\frac{1}{a^h},\ h\in(0, +\infty). \qquad (13)$$

Неважко побачити, що мають місце нерівності

$$g(h) > g_1(h),\ h\in(-\infty, 0), \qquad (14)$$

$$g(h) < g_2(h),\ h\in(0, +\infty). \qquad (15)$$

(Наприклад, маємо $g(h) - g_1(h) = a^h(1-b^h)/[b^h(a^h-b^h)] > 0$ при $h \in(-\infty, 0)$).

З іншого боку, корені рівнянь

$$g_1(h) = \theta \tag{16}$$

і

$$g_2(h) = \theta \tag{17}$$

легко знаходяться: у першому випадку, очевидно,

$$h = h(\theta) = -\frac{ln(1-\theta)}{ln b} \quad , \tag{18}$$

а в другому

$$h = h(\theta) = -\frac{ln \theta}{ln a} \quad . \tag{19}$$

Графічне зображення співвідношень між функціями g, g_1, g_2 та коренями рівнянь (11), (16) і (11) дається на рис.2.

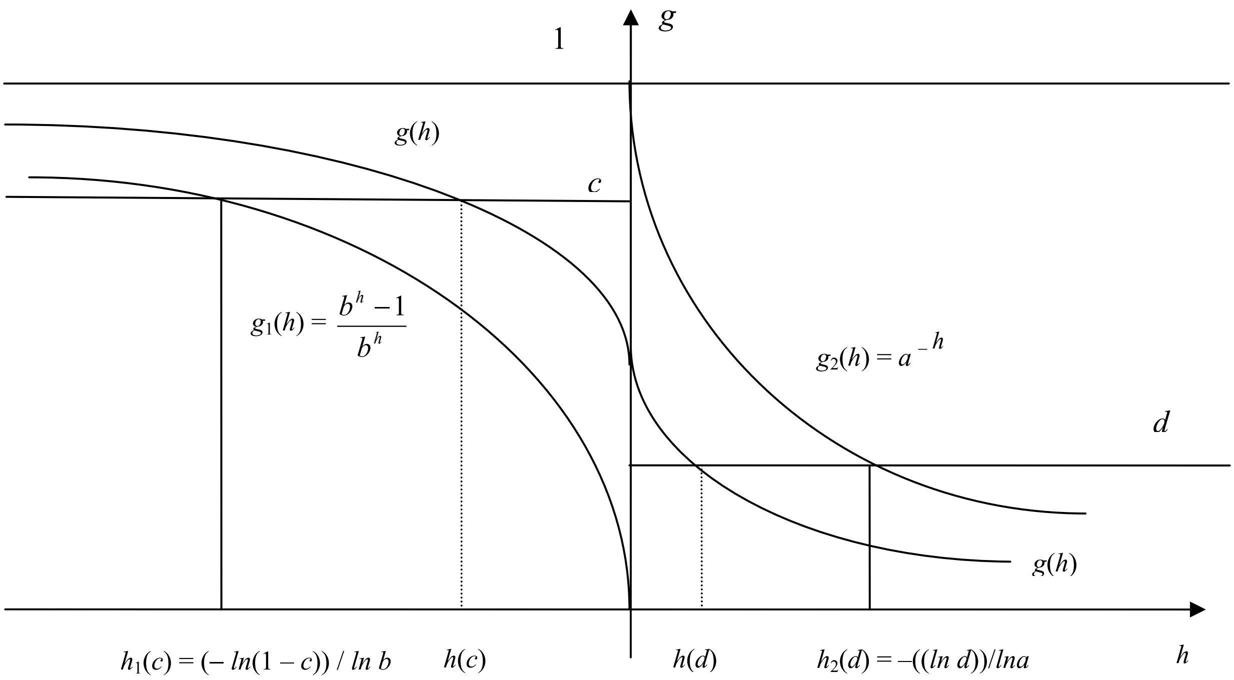

Рис. 2. **Графічна ілюстрація розв'язку рівняння (11)**

За означенням, ордината точки перетину вісі g з кривою $g = g(h)$ є

$$-\frac{ln b}{(ln a - ln b)} \; .$$

Отже, з рис. 2 бачимо:

при $\theta = -\frac{ln b}{ln a - ln b}$ корінь рівняння (11) дорівнює 0;

при $\theta = c > -\frac{ln b}{ln a - ln b}$ корінь рівняння (11) міститься у проміжку

$$(-\frac{ln(1-c)}{ln b},\ 0);$$

при $\theta = d < -\frac{ln b}{ln a - ln b}$ корінь рівняння (11) міститься у проміжку

$$(0,\ -\frac{ln d}{ln a}).$$

Позначимо $f(h) = g(h) - \theta$.

Наступний алгоритм знаходження (єдиного) кореня $h(\theta)$ рівняння (11) спирається на наведені вище відомості. Для запису алгоритму використано алгоритмічну мову, схожу за синтаксисом на Паскаль.

Вхід: $a, b, \varepsilon, \theta$.

Вихід: h.

If $\theta = -\frac{ln b}{ln a - ln b}$, then $h := 0$, end;

if $-\frac{ln b}{ln a - ln b} < \theta$, then $h_0 := -\frac{ln(1-\theta)}{ln b}$

else $h_0 := -\frac{ln\theta}{ln a}$;

if $|f(h_0)| < \varepsilon$, then $h := h_0$, end

else repeat $h_0 := h_0 - (f(h_0) - f(0))^{-1} f(h_0)\, h_0$,

until $|f(h_0)| < \varepsilon$,

$h := h_0$, end.

Зауважимо, що в цьому алгоритмі ітераційний процес уточнення значень $h(\theta)$ (команда repeat until) відповідає класичному методу хорд [8], хоча, очевидно, в цьому місці можна застосувати і будь-який інший числовий метод, для роботи якого достатнім є початкове відділення коренів.

Розроблена авторами роботи комп'ютерна програма виконує достатньо щільне розбиття множини значень параметра θ (щільність розбиття може варіюватися) і для кожної точки цього розбиття знаходить розв'язок рівняння (11). (Параметри a та b вказаного рівняння визначаються по заданих значеннях θ_0 і θ_1, таким що виконується нерівність (6), за допомогою рівностей

$$a = -\frac{\theta_1}{\theta_0},\ b = -\frac{(1-\theta_1)}{(1-\theta_0)}.$$

Після цього для тих самих значень θ за допомогою рівності (7) обчислюються значення ОХ. Передбачено графічне зображення цієї функції.

Крім сказаного вище, комп'ютерне розв'язання рівняння (8) дає змогу миттєво обчислювати і деякі інші важливі статистичні характеристики біноміального ПКВІ. До таких характеристик належить, зокрема, очікуваний обсяг вибірки $M(N|\theta)$ (останній є функцією $\theta \in \Theta$). Так називається математичне сподівання числа вибіркових значень, які потрібно одержати для прийняття рішення щодо прийняття чи відхилення гіпотези H_0 за умови, що значення параметра розподілу $\mathcal{P} \in \theta$. Іншими словами, це – середня кількість одиниць виробничої продукції, яку треба обстежити для прийняття рішення щодо якості всієї партії виробів за умови, що рівень браку в обстежуваній партії дорівнює θ.

У разі біноміального ПКВІ наближений вираз для $M(N|\theta)$ дається таким співвідношенням [1,2]:

$$M(N|\theta) \cong [P(\theta)\ ln(\frac{\beta}{1-\alpha}) + (1-P(\theta))\ ln(\frac{1-\beta}{\alpha})] / u(\theta).$$

де $P(\theta)$ — оперативна характеристика ПКВІ,

$$u(\theta) = \theta \cdot ln(\frac{\theta_1}{\theta_0}) + (1-\theta) \cdot ln\left[\frac{1-\theta_1}{1-\theta_0}\right]$$

Комп'ютерна програма, про яку йшлося мова вище, виконує, зокрема, обчислення $M(N|\theta)$ при заданих θ_0, θ_1, α, β для довільної (вказаної) множини значень θ, видаючи таблицю потрібних значень на зразок наведеної в праці [2].

Висновки

1. Однією з поширених процедур перевірки якості виробничої продукції є статистичний контроль на базі біноміального послідовного відношення ймовірностей. Важливими характеристиками зазначеної процедури є оперативна характеристика і очікуваний обсяг вибірки. Пропоновані у відомій літературі [1,2] методи обчислення значень цих характеристик є не досить ефективними для реальних розрахунків і незручними для складання комп'ютерних програм.
2. У роботі запропоновано альтернативну методику вказаних розрахунків і наведено простий алгоритм, що може служити для реалізації зазначеної методики.

ЛІТЕРАТУРА

1. Леман Э. Проверка статистических гипотез. – М.: Наука, 1964. – 498 с.
2. Ллойд Э., Ледерман У. Справочник по прикладной статистике, т.2. – М.: Финансы и статистика, 1990. – 526 с.
3. Гельфанд С.Ю., Дьяконова Э.В. Статистические методы контроля качества продукции в консервной и пищеконцентратной промышленности. – М.: Лёгкая и пищевая пром-сть, 1984. – 160 с.
4. Богатырёв А.А., Филиппов Ю.Д. Стандартизация статистических методов управления качеством. – М.: Изд-во стандартов, 1989. – 120 с.
5. Вальд А. Последовательный анализ. – М.: ФМ, 1960. – 328 с
6. Барра Ж.-Р. Основные понятия математической статистики. – М.: Мир, 1974. – 275 с.
7. Ширяев А.Н. Статистический последовательный анализ: оптимальные правила остановки. – М.: Наука, 1976. – 272 с.
8. Демидович Б.П., Марон И.А. Основы вычислительной математики. – М.: Наука, 1970. – 548 с.
9. Бахвалов Н.С., Жидков Н.П., Кобельков Г.М. Численные методы. – М.: Наука, 1987. – 598 с.

Надійшла 02.09.2009