УДК 65.012; 519,24

Р.М. Трищ, С.В. Романов

Украинская инженерно-педагогическая академия, Харьков, Украина

МЕТОДЫ ПОЛУЧЕНИЯ ОЦЕНОК ПАРАМЕТРОВ СТАТИСТИЧЕСКИХ МОДЕЛЕЙ РАСПРЕДЕЛЕНИЯ СЛУЧАЙНЫХ ВЕЛИЧИН ПОКАЗАТЕЛЕЙ КАЧЕСТВА

Рассмотрены существующие методы получения оценок параметров статистических моделей рассеивания показателей качества и проведен их сравнительный анализ при условии ограниченности статистической информации. Предложен метод получения оценок, основанный на применении модальных значений порядковых статистик, который позволяет получить эффективные оценки параметров.

Ключевые слова*: эффективные оценки параметров, порядковые статистики, модель распределения.*

Введение

Эффективность и экономичность управления качеством определяется достоверностью определения статистических характеристик процесса рассеивания показателей качества изделий по выборочным данным небольшого объема. Для высокой достоверности решения задач по применению статистических методов в процессе управления качеством необходимо нахождение эффективных статистических оценок параметров статистических моделей распределения.

Нахождение эффективных статистических оценок параметров моделей распределения случайных величин, которым соответствуют процессы рассеивания показателей качества, по опытным данным, является одним из условий управления качеством статистическими методами по малому количеству информации. Чем точнее известна функция плотности распределения, и чем эффективнее ее статистические оценки, тем меньше необходимый объем контрольной выборки.

Точность решения прикладных задач с применением статистических методов, в условиях ограниченного объема информации, зависит от близости полученных статистических оценок к их истинным значениям по выборке небольшого объема. При описании процессов, у которых параметры распределены по нормальному закону, точность их решения зависит от оценки параметра среднего квадратического отклонения σ.

Цель статьи – рассмотрение существующих методов получения оценок параметров статистических моделей распределения показателей качества и предложить новый метод, позволяющий получать эффективные оценки параметров.

Любой статистический метод дает некоторую погрешность в определении оценки параметров распределения. Необходимо максимально точно знать эту ошибку для учета ее при решении практических задач.

Остановимся кратко на некоторых понятиях, связанных со статистическими методами оценки качества. Если описание процесса известно, то вероятность получения результатов наблюдения за исключением того или иного его параметра выражается некоторой функцией распределения. Обычно ее вид можно определить на основании теоретических соображений или предшествующего опыта наблюдений. Таким образом, задача нахождения вероятностных характеристик качества параметров изделия или процесса сводится к отысканию значений неизвестных параметров моделей распределения. Следует отметить, что выбор определённой функции распределения является в каждом случае гипотезой, требующей последующей проверки.

Никакой ограниченный статистический материал не позволяет точно определить вероятностные характеристики качества. Вместо этого с помощью такого материала можно найти статистические характеристики (оценки параметров), т.е. функции от наблюдаемых величин, которые должны давать достаточно хорошие (в смысле заданного критерия) приближения истинного параметра. Существуют различные критерии, позволяющие судить о качестве оценки.

Как известно, оценка a* параметра a основанная на n независимых наблюдениях, называется состоятельной, если она стремится по вероятности к параметру a при $n \to \infty$, т. е. если для её функции распределения $\Phi_n(a, a^*)$ имеет место равенство

$$\lim_{n\to\infty} \Phi_n(a, a^*) = \delta(a^* - a),$$

где $\delta(x)$ – дельта-функция.

При нахождении вероятностных характеристик важно знать точность их определения, т. е. степень возможных отклонений оценок от соответствующих им точным значениям параметров. В качестве меры точности принимаются две величины: отклонение Δ математического ожидания оценки от истинного значения соответствующего параметра и дисперсия $D\{a^*\}$, или среднее квадратичное отклонение $\sqrt{D\{a^*\}}$ оценки а*.

Оценка а* является несмещенной, если для всех а

$$M\{a^*\} = a, \tag{1}$$

где $M\{a^*\}$ – математическое ожидание оценки а*.

Очевидно, что для несмещённой оценки $\Delta = 0$.

Наряду с величинами Δ и $D\{a^*\}$, или вместо них, за меру точности принимается также средняя квадратичная ошибка S, характеризующая разброс оценок параметров относительно истинных значений параметров.

$$S = \sqrt{M\{a^* - a\}}\,. \tag{2}$$

Нетрудно показать, что

$$S = \sqrt{\Delta^2 + D\{a^*\}}\,.$$

Известно, что для несмещённой оценки S^2 и $D\{a^*\}$ совпадают при $\Delta = 0$.

Для сравнения точности статистических характеристик разных параметров пользуются относительными величинами. В качестве меры относительной точности оценки параметра а, принимается величина:

$$\eta = \frac{S}{a}\,. \tag{3}$$

В математической статистике доказывается, что существует нижняя граница для дисперсии несмещённой оценки

$$D\{a\} \ge \frac{1}{I(a)}\,, \tag{4}$$

где I(a) – величина, характеризующая количество информации содержащейся в материалах испытаний относительно параметра а. Если при определении статистической характеристики используется вся эта информация, то такую оценку называют достаточной. Для несмещённой оценки а*, отношение минимально возможной дисперсии к фактической дисперсии этой оценки называется эффективностью оценки e(a*).

Если e(a*) = 1, то оценка а* называется эффективной.

Известно, что коэффициент корреляции между эффективной оценкой и несмещённой оценкой а* того же параметра есть

$$K = \sqrt{e(a^*)}\,. \tag{5}$$

Если $e(a^*)<1$ при конечных n, а $e(a^*) \to 1$ при $n \to \infty$, то оценка называется асимптотически эффективной [2].

Из неравенства (4) следует важный вывод: всякое уменьшение объёма испытаний, сопровождается уменьшением полученной из испытания информации, что в конечном итоге приводит к понижению точности оценки. Поэтому чтобы уменьшить объём испытаний, сохранив при этом необходимую точность оценки, следует потерянную информацию восполнить дополнительными сведениями, которые иногда требуют построения более адекватной модели или получение несмещённых, эффективных оценок.

На основе анализа испытаний и теоретических соображений делается предположение о характере распределения исследуемого показателя качества. Данные предположения проверяются с использованием методов математической статистики, которые называются критериями согласия. Но для того, чтобы решить данную задачу нужно уметь находить оценки неизвестных параметров моделей распределения показателя качества. Приведем некоторые методы, которые будут применяться для решения этих задач.

Для получения статистических оценок параметров моделей распределения применяют метод максимального правдоподобия и метод моментов, так как с их применением, для некоторых моделей распределения, можно получить эффективные оценки. На практике метод максимального правдоподобия приводит иногда к довольно сложным системам уравнений, которые только в настоящий момент могут быть решены с помощью современных компьютерных программ. Получение оценок методом моментов содержит в себе элемент произвола, так как, кроме моментов, можно рассматривать и другие числовые характеристики распределений, например, медиану, моду, дисперсию, коэффициент асимметрии, эксцесс и т.п.

Оценки, полученные методом максимального правдоподобия имеют такие существенные характеристики:

1. Они являются состоятельными.
2. Их распределение при увеличении n приближается к нормальному.
3. Они являются достаточными, если только достаточная статистика существует.
4. Они являются асимптотически эффективными, хотя для некоторых распределений являются смещенными.

Метод получения оценок, состоящий в приравнивании теоретических моментов распределения к эмпирическим, называется методом моментов.

Преимуществом метода моментов является сравнительная простота его применения.

Нет оснований, предполагать, что с помощью метода моментов можно получить наилучшие в каком-либо смысле оценки. Можно показать, что оценки, полученные по методу моментов, вообще говоря, не всегда имеют наименьший, возможный разброс и не являются асимптотически эффективными для некоторых распределений. Однако в силу простоты метод моментов используется на практике.

Классические методы статистического оценивания – метод максимального правдоподобия и метод моментов, ориентированы на применение к выборкам большого объема n. Получаемые при этом оценки, обладают наилучшими свойствами только при $n >> 1$, однако на практике их часто применяют при любых объемах выборки, в том числе и малых, что приводит к значительным ошибкам при решении прикладных задач.

Метод получения оценок среднего значения параметра *a* и его рассеивания σ по обобщенному способу наименьших квадратов, для распределений вида $F[(X - a)/\sigma$, предложил Э. Ллойд. Если удается получить математические ожидания $e_{i:n}$ нормированных порядковых статистик $(X_{(i)} - a)/\sigma$, их дисперсии и ковариации V_{ij} для любой выборки объема n, то соответствующие характеристики упорядоченных наблюдений $X_{(i)}$ будут линейными функциями искомых параметров *a* и σ с известными коэффициентами. То есть математическое ожидание i-й порядковой статистики $M[X_{(i)}]$ и ковариация между i-й и j-й порядковой статистиками $Cov[X_{(i)}, X_{(j)}]$ выражается формулами [3]:

$$M[X_{(i)}] = a + \sigma e_{i:n} ; \qquad (7)$$

$$Cov[X_{(i)}, X_{(j)}] = \sigma^2 V_{ij}, \ (i \le j). \qquad (8)$$

Тогда можно по обобщенному методу наименьших квадратов получить несмещенные и имеющие наименьшую дисперсию среди всех этих линейных оценок оценки a^* и σ^* в виде линейных функций упорядоченных результатов наблюдений $X_{(i)}$:

$$a^* = \sum_{i=1}^{n} \alpha_i X_{(i)} \quad \text{и} \quad \sigma^* = \sum_{i=1}^{n} \beta_i X_{(i)} ,$$

где α_i и β_i – весовые коэффициенты порядковых статистик.

Для многих распределений найти математическое ожидание порядковых статистик и их ковариацию очень сложно. Поэтому необходимы методы нахождения оценок параметров распределений, которые не требуют их вычислений, и могут дать несмещенные и близкие к оптимальным оценки.

Предлагается метод нахождения оценок параметров любых распределений, который требует вычисления только математического ожидания порядковых статистик объема выборки n, совпадающего с числом оцениваемых параметров. Метод основан на полученном рекуррентном соотношении функции плотности i-й порядковой статистики объема n $\psi_n(x_{(i)})$ с функцией плотности любой *l*-й порядковой статистики выборки объема, меньшего n.

Так как не для всех моделей распределения можно найти математические ожидания порядковых статистик, то, для них, предложим новый метод получения оценок его параметров, используя модальные значения первой и последней порядковых статистик.

Так как функция распределения наибольшего значения (n-ой порядковой статистики) выборки объема n описывается зависимостью

$$F(r_{(n)}) = \left[F(r)\right]^n ,$$

где F(r) – функция распределения генеральной совокупности.

Тогда функция плотности распределения n-й порядковой статистики

$$f(r_{(n)}) = F'(r_{(n)}) = n\left[F(r)\right]^{n-1} f(r) ,$$

где f(r) – плотность распределения генеральной совокупности $f(r) = F'(r)$.

Для нахождения моды n-й порядковой статистики необходимо определить производную от ее плотности распределения и решить дифференциальное уравнение второго порядка:

$$(n-1)\left[F'(r)\right]^2 + F(r) \cdot F''(r) = 0 .$$

Как видно, это уравнение содержит только функцию F(r) и ее производные, но не содержит независимой переменной r. Поэтому, согласно [4], оно допускает понижение порядка с помощью замены

$$F'(r) = z ; \qquad F''(r) = z\frac{dz}{dF} .$$

Отсюда уравнение имеет вид:

$$z\frac{dz}{dF} F = (1-n)z^2 .$$

Решим его способом, разделения переменных, имеем:

$$\frac{dz}{z} = (1-n)\frac{dF}{F} ; \quad \ln|z| = (1-n)\ln|F| + \ln C_1 ;$$

$$z = C_1 F^{1-n}, \text{ где } C_1 \text{ – константа.}$$

Подставляя вместо z – $F'(r)$ имеем уравнение с разделяющими переменными:

$$\frac{dF}{dr} = C_1\left[F(r)\right]^{1-n} ; \quad F^{n-1}dF = C_1 dr ,$$

получим:

$$F^n(r) = (C_1 r + C_2)n .$$

Зная числовые характеристики изучаемой модели, можем получить уравнение, позволяющее найти моду n-ой порядковой статистики для разных моделей.

В дальнейших исследованиях необходимо определить наилучшие оценки для моделей распределения, которым подчиняются рассеивание нормируемых показателей качества. Это позволит применять методы сбора и обработки статистической информации с целью управления качеством продукции при меньшем объеме контрольных выборок.

Вывод

Предложен метод получения оценок параметров моделей распределений случайных величин, которые можно применять для оценки параметров моделей, применяемых при изучении вопросов управления качеством. Новый метод основан на теории порядковых статистик.

Список литературы

1. Воробьев Л.Н. Технология машиностроения и ремонт машин / Л.Н. Воробьев. – М.: Высш. Шк., 1981. – 334 с.

2. Дейвид Г. Порядковые статистики / Г. Дейвид. – М.: Наука. Главная редакция физико-математической литературы, 1979. – 336 с.

3. Введение в теорию порядковых статистик / Под ред. А.Я. Боярского. – М.: Статистика, 1970. – 416 с.

4. Введение в теорию порядковых статистик / Под ред. А.Я. Боярского. – М.: Статистика, 1970. – 416 с.

Поступила в редколлегию 13.04.2010

Рецензент: д-р техн. наук, проф. И.П. Захаров, Харьковский национальный университет радиоэлектроники, Харьков.

МЕТОДИ ОДЕРЖАННЯ ОЦІНОК ПАРАМЕТРІВ СТАТИСТИЧНИХ МОДЕЛЕЙ РОЗПОДІЛУ ВИПАДКОВИХ ВЕЛИЧИН ПОКАЗНИКІВ ЯКОСТІ

Р.М. Тріщ, С.В. Романов

Розглянуто існуючі методи одержання оцінок параметрів статистичних моделей розсіювання показників якості й проведений їхній порівняльний аналіз за умови обмеженості статистичної інформації. Запропоновано метод одержання оцінок, заснований на застосуванні модальних значень порядкових статистик, що дозволяє одержати ефективні оцінки параметрів.

Ключові слова: *ефективні оцінки параметрів, порядкові статистики, модель розподілу.*

METHODS OF RECEPTION OF PARAMETERS ESTIMATIONS OF DISTRIBUTION STATISTICAL MODELS OF RANDOM VARIABLES OF QUALITY PARAMETERS

R.M. Trishch, S.V. Romanov

Existing methods of reception of estimations of parameters of statistical models of dispersion of parameters of quality are considered and their comparative analysis under condition of limitation of the statistical information is carried out. The method of reception of the estimations, based on application of modal values serial statistics which allows to receive effective estimations of parameters is offered.

Key words*: effective estimations of parameters, serial statistics, model of distribution.*