УДК 303.833.4

СТАТИСТИЧЕСКАЯ ОБРАБОТКА ДАННЫХ В СИСТЕМЕ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ С УЧЕТОМ ОШИБОК КЛАССИФИКАЦИИ И КОДИРОВАНИЯ ОБЪЕКТОВ И ПРОЦЕССОВ

Л.Н. КОСЯК, С.В. КУХТА, В.В. ТОЧИЛО, д-р техн. наук, проф. М.Л. ХЕЙФЕЦ
(Полоцкий государственный университет)

Рассматриваются процедуры формирования ограничительных перечней объектов и процессов производства, создания регламентов технологических операций. Показана их зависимость от специфики конкретного производства. Для предприятий, выпускающих большую номенклатуру продукции и оказывающих широкий спектр услуг, предложены различные методы формирования ограничительных перечней. Выбор ограниченного числа позиций для унификации и последующей регламентации рекомендовано производить в определенной последовательности по частоте применения или продолжительности использования объектов и процессов производства. Наивысшим приоритетом обладают наиболее часто встречающиеся объекты производства, а также объекты, длительность использования которых в производственных процессах максимальна. В первую очередь сокращается номенклатура объектов и устраняются процессы, которые редко используются в производстве и которые можно заменить наиболее эффективными.

Введение. Определение рациональной выборки из генеральной совокупности объектов и процессов особенно актуально в условиях периодической смены выпускаемой продукции, сопровождающейся переналадкой производства. Наиболее целесообразен путь постепенного формирования базы данных и знаний системы поддержки принятия решений при периодической статистической обработке информации с целью получения ограничительных перечней. Процесс пересмотра информации может быть закончен, когда последующий получаемый ограничительный перечень совпадает с предыдущим. Процесс может быть существенно сокращен, если формирование и изменение базы данных и знаний осуществлять в соответствии с определенными приоритетами, такими как серийность производства, величина партий деталей, степень сложности изделий, периодичность, стабильность заказов и др. [1].

Для статистической обработки информации необходимо рассматривать распределение по таким параметрам унификации, как частота встречи и продолжительность использования объектов и процессов производства. При этом для наглядности строятся гистограммы распределения или полигоны частот определенных параметров в зависимости от технологических факторов. При формировании ограничительных перечней рациональными для унификации являются позиции, в окрестностях которых распределения частот носят случайный характер, то есть вблизи которых не происходит наложение каких-либо закономерных особенностей конструктивного или технологического характера. Поэтому целесообразно при формировании ограничительного перечня проводить проверку соответствия распределений частот в окрестности принятого значения одному из законов распределения случайных величин [2].

В каждом случае имеется свой механизм возникновения суммарной погрешности, отклонения и, следовательно, свое соотношение между погрешностями различных видов. Поэтому обычно закон распределения в окрестности каждой позиции весьма индивидуален. Однако во многих случаях и особенно тогда, когда погрешности модели невелики, удается высказать соображения в пользу того или иного закона распределения, параметры которого определяются из опыта. Эти соображения связываются, прежде всего, с предельными теоремами теории вероятностей и рядом теорем, опирающихся на представления теории информации [3, 4].

Центральная предельная теорема. Среди большого числа предельных теорем важнейшей для теории ошибок является так называемая центральная предельная теорема, или теорема Ляпунова, утверждающая, что сумма независимых случайных величин, таких, что удельный вес каждого отдельного слагаемого стремится к нулю при неограниченном увеличении числа слагаемых, в пределе распределена по нормальному закону. Теореме Ляпунова можно дать и другую, более четкую формулировку [5].

Пусть $x_1, x_2,..., x_n$ – последовательность независимых случайных величин, $m_1, m_2,..., m_n$ – их математические ожидания, а $\sigma^2_1, \sigma^2_2,..., \sigma^2_n$ – дисперсии.

Рассмотрим сумму $y_n = \sum_{i=1}^{n} x_i$. Математическим ожиданием этой случайной величины будет $m_{y_n} = \sum_{i=1}^{n} m_i$, а дисперсией – $S^2_{y_n} = \sum_{i=1}^{n} \sigma^2_i$.

Обозначим:

$$z_n = \frac{y_n - m_{y_n}}{S_{y_n}};$$

$$\Phi(\alpha) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\alpha} e^{-t^2/2} dt.$$

Теорема Ляпунова для плотности вероятности P утверждает, что

$$\lim_{n\to\infty} P(\alpha \le z \le \beta) = \Phi(\beta) - \Phi(\alpha)$$

тогда и только тогда, когда соблюдается условие:

$$\lim_{n\to\infty} \frac{1}{S_{y_n}^2} \sum_{i=1}^{n} \int_{|x-m_i| \ge \varepsilon S_{y_n}} (x-m_i)^2 P_i(x) dx = 0,$$

которое называется условием Линдеберга. Оно гарантирует, что $\lim_{n\to\infty} \frac{\sigma_i}{S_{y_n}} = 0$ для всех i, т.е. малость каждого слагаемого в рассматриваемой сумме $\sum_{i=1}^{n} x_i$ случайных величин по сравнению со всей суммой.

Пусть, например, имеется сумма одинаково распределенных случайных величин с конечной дисперсией σ^2. В этом случае $S_{y_n}^2 = n\sigma^2$ и

$$\frac{1}{S_{y_n}^2} \sum_{i=1}^{n} \int_{|x-m_i| \ge \varepsilon S_{y_n}} (x-m_i)^2 P_i(x) dx = \frac{1}{n\sigma^2} n \int_{|x-m_i| \ge \varepsilon \sqrt{n\sigma^2}} (x-m)^2 P(x) dx =$$
$$= \frac{1}{\sigma^2} \int_{|x-m_i| \ge \varepsilon \sqrt{n\sigma^2}} (x-m)^2 P(x) dx = \frac{2}{\sigma^2} \int_{\varepsilon n\sigma^2}^{\infty} (x-m)^2 P(x) dx. \quad (1)$$

Если только несобственный интеграл, входящий в выражение (1), существует, он с ростом n стремится к нулю при любом ε, и поэтому все выражение также стремится к нулю.

Поскольку условие Линдеберга соблюдается, то утверждение теоремы Ляпунова применительно к рассматриваемой сумме верно. Доказанное утверждение [5] называют теоремой Линдеберга – Леви.

Так как в процессе измерений происходит суммирование большого количества погрешностей разнообразного происхождения (инструментальных, личных, внешних и др.), то, учитывая центральную предельную теорему, есть основания надеяться, что суммарная погрешность будет распределена по нормальному закону.

К предположению о нормальном характере закона распределения погрешностей можно прийти и на основании несколько иных соображений [1, 5].

Допустим, что произведено некоторое измерение, из-за погрешности которого результат будет сдвинут относительно истинного значения измеряемой величины на случайную величину Δ. Выясним, какова вероятность того, что истинное значение измеряемой величины лежит в интервале $[y, y + \Delta y]$, если в результате измерения получено значение x.

На основании формулы полной вероятности можно записать

$$P(x)P(y|x) = P(y)P(x|y), \quad (2)$$

где $P(y)$ – априорная плотность вероятности измеряемой величины; $P(x|y)$ – плотность вероятности результата измерения при условии, что истинным значением измеряемой величины является y; $P(x)$ – априорная плотность вероятности появления результата x при измерении; $P(y|x)$ – плотность вероятности того, что измеряемая величина принимает значение y, в то время как результатом измерения является x.

Из (2) имеем

$$P(x|y) = [P(x)|P(y)]P(y|x).$$

Учитывая, что априорная информация об измеряемой величине отсутствует, будем считать, что

$$P(y) = \begin{cases} 1/(2z), \text{ если } |y| \le z \\ 0, \text{ если } |y| > z, \end{cases}$$

где z – произвольно большое число, т.е. на любом интересующем нас участке $P(y) = \text{const}$.

Кроме того, априорная вероятность появления того или иного результата измерения также должна быть принята постоянной, если нет оснований полагать, что для появления некоторых результатов измерения имеются особо благоприятные или особо неблагоприятные условия. Отсюда следует, что если погрешность измерения имеет плотность вероятности $P(\Delta) = P(x - y)$, то по такому же закону распределено неизвестное значение y при условии, что измеренное значение равно x.

Нормальный закон распределения. Определим, при каком законе распределения погрешностей измерения $P = f(x)$ значение измеряемой величины в максимальной степени неопределенно. Степень неопределенности системы можно характеризовать ее близостью к равновесному состоянию, которая тем больше, чем больше величина энтропии

$$H = -\int_{-\infty}^{\infty} f(x) \ln f(x) dx. \tag{3}$$

При дальнейшем применении полученной информации о состоянии рассматриваемой системы наиболее рациональным будет использование гипотезы о том, что в пределах предположения о несмещенности и известной точности проделанного измерения в остальном состояние системы в максимальной степени неопределенно. Это прямо связано с предположением о виде плотности распределения погрешности измерения. Требуется найти плотность распределения $P = f(x)$ такую, что величина энтропии H достигает максимума при следующих ограничениях, наложенных на функцию $f(x)$:

$$f(x) \ge 0; \qquad -\int_{-\infty}^{\infty} f(x) dx = 1; \qquad -\int_{-\infty}^{\infty} x f(x) dx = y; \qquad -\int_{-\infty}^{\infty} (x-y)^2 f(x) dx = \sigma_0^2 = \text{const}. \tag{4}$$

Интегралы (4) при больших a и целом M можно сколько угодно точно представить в виде сумм:

$$H = -\int_{-\infty}^{\infty} f(x) \ln f(x) dx \approx -\int_{-a}^{a} f(x) \ln f(x) dx \approx$$

$$\approx -\frac{a}{M} \sum_{k=-M}^{M} f(x_k) \ln f(x_k) = -\frac{a}{M} \sum_{k=-M}^{M} z_k \ln z_k ;$$

$$\int_{-\infty}^{\infty} f(x) dx \approx \int_{-a}^{a} f(x) dx \approx \frac{a}{M} \sum_{k=-M}^{M} z_k ; \tag{5}$$

$$\int_{-\infty}^{\infty} x f(x) dx \approx \int_{-a}^{a} x f(x) dx \approx \frac{a}{M} \sum_{k=-M}^{M} x_k z_k ;$$

$$\int_{-\infty}^{\infty} (x-y)^2 f(x) dx \approx \int_{-a}^{a} (x-y)^2 f(x) dx \approx \frac{a}{M} \sum_{k=-M}^{M} (x_k - y)^2 z_k = \frac{a}{M} \sum_{k=-M}^{M} (x_k^2 z_k - y^2 z_k),$$

где $x_k = ak/M$; $z_k = f(x_k)$.

Для поиска экстремума воспользуемся методом неопределенных множителей Лагранжа [6]. Будем искать стационарную точку функции

$$L = -\int_{-\infty}^{\infty} f(x) \ln f(x) dx + \mu_1 \left[\int_{-\infty}^{\infty} f(x) dx - 1 \right] + \mu_2 \left[\int_{-\infty}^{\infty} x f(x) dx - y \right] + \mu_3 \left[\int_{-\infty}^{\infty} x^2 f(x) dx - y^2 - \sigma_0^2 \right] \tag{6}$$

или бесконечно близкого к ней по значению выражения:

$$L = \frac{a}{M} \left(-\sum_{k=-M}^{M} z_k \ln z_k + \mu_1 \sum_{k=-M}^{M} x_k z_k + \mu_2 \sum_{k=-M}^{M} x_k^2 z_k + \mu_3 \sum_{k=-M}^{M} x_k^2 z_k \right) - [\mu_1 + \mu_2 y + \mu_3 (y^2 + \sigma^2{}_0)],$$

введя множители Лагранжа, учитывающие ограничения (5).

Необходимым условием попадания в стационарную точку является обращение в нуль частных производных от функции Лагранжа:

$$\partial L / \partial z_1 = \partial L / \partial z_2 = ... = \partial L / \partial z_M = 0,$$

откуда с учетом (6) получаем

$$\partial L / \partial z_k = -\ln z_k - 1 + \mu_1 + \mu_2 x_k + \mu_3 x^2_k = 0. \quad (7)$$

Уравнение (7) связывает значения искомой функции и аргумента, относящиеся к одному и тому же k. Вид выражения (7) не зависит от a и M, т.е. от способа разбиения интервала интегрирования на отрезки. Поэтому уравнение для нахождения функции $f(x)$ можно записать как

$$-\ln f(x) - 1 + \mu_1 + \mu_2 x + \mu_3 x^2 = 0. \quad (8)$$

Вид уравнения (8) остается неизменным на всем интервале интегрирования. Дифференцируя левую часть (7), убедимся, что находимся в точке максимума по z_k.

Таким образом, функция, удовлетворяющая условиям (4) и максимизирующая H (3), может быть найдена из уравнения (8). В силу этого

$$f(x) = \exp(\mu_1 + \mu_2 x + \mu_3 x^2 - 1), \quad (9)$$

т.е. условие $f(x) \geq 0$ удовлетворяется автоматически.

Подставляя выражение (9) вместо $f(x)$ в (4), получаем:

$$-1 + \mu = y^2 / (2\sigma_0^2) + \ln(\sigma_0 \sqrt{2\pi}); \;\; \mu_2 = y / \sigma_0^2; \;\; \mu_3 = 1 / (2\sigma_0^2),$$

откуда

$$f(x) = 1/(\sigma_0 \sqrt{2\pi}) \exp\left[-\frac{(x-y)^2}{2\sigma_0^2} \right].$$

Уравнение (9) можно получить формальным способом, объединив под знаком интеграла все слагаемые в (6) и затем продифференцировав подынтегральное выражение по переменной f [5].

Следовательно, наибольшей неопределенностью при выполнении условий (4) будут обладать измерения, имеющие нормальное распределение.

Распределение по закону Лапласа. Обычно предполагается, что точность измерения параметра известна и на всем протяжении эксперимента не меняется. Но во многих реальных ситуациях, особенно тогда, когда велики погрешности модели и присутствуют ошибки классификации и кодирования, это предположение выглядит необоснованным [1].

Поэтому естественно высказать более осторожное суждение, состоящее в том, что основная точностная характеристика измерения – среднеквадратическое отклонение σ – известна лишь в среднем. Это же иногда целесообразно предполагать и в тех случаях, когда погрешности модели невелики, а ошибки классификации и кодирования отсутствуют, но точность параметров определена недостаточно надежно.

Следуя той же концепции о максимальной неопределенности системы в рамках сделанных предположений и учитывая неотрицательность $f(x)$, найдем плотность распределения $f^*(x)$, обращающую в максимум энтропию (3). При этом отметим, что измерения имеют нормальную плотность распределения, но их среднеквадратическое отклонение является случайной величиной с математическим ожиданием σ_0, соответствующим некоторым «средним» окружающим условиям.

В этом случае имеет место следующая система ограничений:

$$f^*(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{1}{\sigma} \exp\left[-\frac{(x-y)^2}{2\sigma^2} \right] P(\sigma) d\sigma = \int_0^{\infty} n(x,\sigma) P(\sigma) d\sigma;$$

$$\int_0^{\infty} P(\sigma) d\sigma = 1; \;\; \int_0^{\infty} \sigma P(\sigma) d\sigma = \sigma_0; \;\; 0 < \sigma < \infty, \quad (10)$$

где $n(x, \sigma) = 1/(\sigma \sqrt{2\pi}) \exp\left[-\frac{(x-y)^2}{2\sigma^2} \right]$.

Для того чтобы определить функцию $f^*(x)$, нужно предварительно найти функцию $P(\sigma)$. Для строгого решения поставленной задачи можно воспользоваться представлениями типа (5), заменив все интегралы суммами, но гораздо проще (хотя и менее строго) произвести формальное дифференцирование соответствующего выражения по функции P, считая ее независимой переменной [5]. Тогда с учетом ограничений (10) имеем

$$L = -\int_{-\infty}^{\infty}\left\{\int_0^{\infty} n(x,\sigma)P(\sigma)d\sigma \ln\left[\int_0^{\infty} n(x,\sigma)P(\sigma)d\sigma\right]\right\}dx + \mu_1\left\{\int_0^{\infty} P(\sigma)d\sigma - 1\right\} + \mu_2\left\{\int_0^{\infty}\sigma P(\sigma)d\sigma - \sigma_0\right\}.$$

Дифференцируя это выражение по P, получаем:

$$\frac{\partial L}{\partial P} = -\int_{-\infty}^{\infty}\left\{\int_0^{\infty} n\ x,\sigma\ d\sigma \ln\left[\int_0^{\infty} n\ x,\sigma\ P\ \sigma\ d\sigma\right]\right\}dx - \int_{-\infty}^{\infty}\int_0^{\infty} n\ x,\sigma\ d\sigma\, dx + \mu_1\int_0^{\infty} d\sigma + \mu_2\int_0^{\infty}\sigma d\sigma =$$

$$= \int_0^{\infty}\left\{-\int_{-\infty}^{\infty} n\ x,\sigma\ \ln\left[\int_0^{\infty} n\ x,\sigma\ P\ \sigma\ d\sigma\right]dx - \int_{-\infty}^{\infty} n\ x,\sigma\ dx + \mu_1 + \mu_2\sigma\right\}d\sigma. \tag{11}$$

Необходимое условие экстремума имеет вид:

$$\int_0^{\infty} n(x,\sigma)\ln\left[\int_0^{\infty} n(x,\sigma)P(\sigma)d\sigma\right]dx + 1 - \mu_1 - \mu_2\sigma = 0.$$

Покажем, что этому условию удовлетворяет функция

$$P(\sigma) = \frac{2}{\lambda^2}\sigma e^{-\sigma^2/\lambda^2}, \tag{12}$$

где $\lambda = 2\sigma_0/\sqrt{\pi}$.

В самом деле, во-первых, выполняются условия:

$$P(\sigma) \ge 0 \text{ при } 0 \le \sigma \le \infty;$$

$$\int_0^{\infty} P(\sigma)d\sigma = \frac{2}{\lambda^2}\int_0^{\infty}\sigma e^{-\sigma^2/\lambda^2}\ d\sigma = \frac{2}{\lambda^2}\frac{1}{2}\lambda^2 = 1;$$

$$\int_0^{\infty}\sigma P(\sigma)d\sigma = \frac{2}{\lambda^2}\int_0^{\infty}\sigma e^{-\sigma^2/\lambda^2}\ d\sigma = \frac{2}{\lambda^2}\frac{\sqrt{\pi}}{4}\lambda^3 = \frac{\sqrt{\pi}}{2}\lambda = \sigma_0.$$

Во-вторых, подстановка (12) в (11) дает

$$\int_{-\infty}^{\infty}\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{(x-y)^2}{2\sigma^2}\right]\cdot\ln\left[\frac{1}{\sqrt{2\pi}}\int_0^{\infty}\frac{1}{\sigma}\exp\left[-\frac{(x-y)^2}{2\sigma^2}\right]\sigma\frac{2}{\lambda^2}\exp(-\sigma^2/\lambda^2)d\sigma\right]dx - 1 + \mu_1 + \mu_2\sigma = 0. \tag{13}$$

Так как в [7]

$$f^*(x) = \int_0^{\infty} n(x,\sigma)P(\sigma)d\sigma = \frac{1}{\sqrt{2\pi}}\frac{2}{\lambda^2}\int_0^{\infty}\exp\left[-\frac{(x-y)^2}{2\sigma^2} - \frac{\sigma^2}{\lambda^2}\right]d\sigma = \frac{1}{2\lambda/\sqrt{2}}\exp\left[-\frac{|x-y|}{\lambda/\sqrt{2}}\right] = \frac{1}{2\Lambda}\exp\left[-\frac{|x-y|}{\Lambda}\right],$$

где $\Lambda = \lambda/\sqrt{2}$,

то, подставляя последнее выражение в (13), получаем

$$\int_{-\infty}^{\infty}\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{(x-y)^2}{2\sigma^2}\right](-\ln 2\Lambda)dx - \int_{-\infty}^{\infty}\frac{1}{\sigma\sqrt{2\pi}}\frac{|x-y|}{\Lambda}\exp\left[-\frac{(x-y)^2}{2\sigma^2}\right]dx - 1 + \mu_1 + \mu_2\sigma =$$

$$= -\ln 2\Lambda - \frac{2}{\sigma\sqrt{2\pi}}\frac{1}{\Lambda}\sigma^2 - 1 + \mu_1 + \mu_2\sigma = -\ln 2\Lambda - \sqrt{\frac{2}{\pi}}\frac{\sigma}{\Lambda} - 1 + \mu_1 + \mu_2\sigma.$$

Полагая $\ln 2\Lambda = -1 + \mu_1$ и $\mu_2 = \frac{1}{\Lambda}\sqrt{\frac{2}{\pi}} = \frac{2}{\lambda\sqrt{\pi}} = \frac{1}{\sigma_0}$, удовлетворяем необходимое условие экстремума.

Таким образом, при $P(\sigma) = 2(\sigma/\lambda^2)\exp(-\sigma^2/\lambda^2)$

$$f^*(x) = \int_0^\infty n(x,\sigma)P(\sigma)d\sigma = \frac{1}{2\lambda}\exp\left[-\frac{|x-y|}{\lambda}\right].$$

При повторном дифференцировании выражения (11) по P можно убедиться в том, что найденное значение функции $P(\sigma)$ обращает энтропию в максимум. Порождаемая ею функция распределения погрешностей измерения $P = f^*(x)$ является лапласовским распределением.

Предположение о распределении погрешностей измерений по закону Лапласа является более осторожным, чем предположение о нормальном законе во всех случаях, когда точностные характеристики измерений неизвестны или нестабильны.

Ограничительные перечни и регламенты процессов. Для определения степени репрезентативности исследуемой выборки генеральной совокупности целесообразно определить частоту (продолжительность) p использования объектов или любую другую величину, характеризующую статистические данные выборки.

Частота p соответствует одному из параметров x (размеру, точности и т.д.).

По частоте p вычисляются [8]:

$x_B = \frac{\sum_{i=1}^{n} x_i p_i}{\sum_{i=1}^{n} p_i}$ – выборочная средняя;

$\Delta_B = x_i - x_B$ – отклонение от общей средней;

$\sigma^2_B = \frac{\sum_{i=1}^{n} (x_i - x_B)^2 p_i}{\sum_{i=1}^{n} p_i}$ – выборочная дисперсия;

$\sigma_B = \sqrt{\sigma_B^2}$ – выборочное среднеквадратичное отклонение.

Это позволяет рассчитать ошибку выборки: $\Delta_\alpha = t_\alpha \sqrt{\frac{\sigma_B^2}{\sum_{i=1}^{n} p_i}} \sqrt{\left(1 - \frac{\sum_{i=1}^{n} p_i}{N}\right)}$,

где t_α – критерий Стьюдента при допустимой доверительной вероятности α; N – общее число позиций генеральной совокупности; $\left(\sum_{i=1}^{n} p_i / N\right)\cdot 100$ – процент выборки из всей генеральной совокупности.

Сравнивая ошибку выборки Δ_α с отклонением от общей средней Δ_B, можно судить о степени ее репрезентативности по принятому интервалу доверительной вероятности.

Степень соответствия закону распределения частоты случайных величин рассматриваемого фактора в окрестностях значения, выбранного для унификации, рассмотрим для нормального закона распределения [6]:

$$F(x) = \frac{1}{\sigma_B\sqrt{2\pi}}\exp\left[-\frac{(x-x_B)^2}{2\sigma_B^2}\right], \tag{14}$$

и закона Лапласа с интенсивностью появления рассматриваемых объектов $\lambda^* = \sigma/\sqrt{2}$ [6, 8]:

$$F(x) = \frac{1}{\sigma_B\sqrt{2}}\exp\left[\frac{\sqrt{2}|x-x_B|}{\sigma_B}\right]. \tag{15}$$

Для этого необходимо определить: $t = \frac{x - x_B}{\sigma_B}$ – нормированное отклонение; $\Delta t = \frac{x_{i+1} - x_i}{\sigma_B}$ – приращение нормированного отклонения на рассматриваемом интервале.

В соответствии с плотностью вероятности на интервале для нормального закона:

$$f^*(t) = \frac{1}{\sqrt{2\pi}} \exp\left[-t^2/2\right] \Delta t\,, \tag{16}$$

и для закона Лапласа:

$$f^*(t) = \frac{1}{\sqrt{2}} \exp\left[-\sqrt{2}t\right] \Delta t\,, \tag{17}$$

пересчитываются частотные характеристики на каждом интервале:

$$p_i^* = f^* \sum_{i=1}^{n} p_i.$$

Это позволяет рассчитать критерий Пирсона:

$$\chi^2 = \sum_{i=1}^{n} \frac{(p_i - p_i^*)}{p_i^{*2}}\,,$$

по которому можно определить степень соответствия распределения принятому закону.

Судить о степени соответствия распределения случайных величин принятому закону позволяет соотношение Романовского [8, 9]:

$$Q = \frac{\chi^2 - f}{\sqrt{2f}}\,, \tag{18}$$

где f – число степеней свободы, то есть число групп в изучаемом ряду рассчитанных статистических характеристик и используемых при вычислении теоретического распределения.

Для принятых законов распределения (14) и (15) существует две характеристики – это x_B и σ_B. Тогда число степеней свободы в (18) $f = N - 2$, где N – число интервалов разбиения рассматриваемого участка распределения случайных величин.

При отношении Романовского $Q \leq 3$ можно говорить о соответствии рассматриваемого участка принятому закону распределения. Статистическая обработка данных об объектах производства и их параметрах позволяет судить о степени достоверности описания в ограничительных перечнях тому многообразию случайных факторов, которые необходимо рассматривать при унификации. При достаточной репрезентативности выборки и соответствии предложенных для унификации позиций максимумам – модам в законах распределения случайных величин – можно приступать к формированию ограничительных перечней, объектов производства и регламентов технологических процессов.

Заключение. Процедуры формирования ограничительных перечней объектов и процессов производства, а также создания регламентов технологических операций зависят от специфики конкретного производства. Для предприятия, выпускающего большую номенклатуру продукции и оказывающего широкий спектр услуг по запросам заказчиков, рекомендуются различные методы формирования ограничительных перечней:

1) при обширной информации об унифицируемых объектах (обычно для большой номенклатуры изделий) целесообразно анализировать все моды – локальные максимумы на кривых распределения или полигонах вероятностных характеристик. Изучение рассеяния параметров в окрестностях моды необходимо проводить в сопоставлении с законом распределения случайных величин Лапласа (15), (17) или нормальным законом распределения (14), (16);

2) согласно соотношению Романовского (18), при удовлетворительном соответствии полученного распределения теоретическому, моду на изученном интервале можно принять в качестве позиции для ограничительного перечня, а все остальные позиции в ее окрестности по возможности убрать;

3) в случае, когда при описании объектов имеются сведения о предпочтительных интервалах использования параметров, эти интервалы на полигоне распределения или гистограмме разбиваются на участки размером меньше, чем другие. На каждом участке определяются моды распределения вероятностных характеристик, причем шаг разбиения участка для поиска абсолютного максимума также может быть уменьшен;

4) когда отсутствует достаточная информация об унифицируемых объектах, рационально использовать наиболее простой метод разбиения статистического распределения параметров на равные интервалы гистограммы в соответствии с числом позиций, которые целесообразно оставить после унификации. Последующий поиск абсолютных максимумов распределения на всех интервалах разбиения позволяет предложить позиции для начального варианта ограничительного перечня.

Выбор ограниченного числа позиций для унификации и последующей регламентации может производиться в определенной последовательности по частоте применения или продолжительности использования объектов и процессов производства. Наивысшим приоритетом обладают наиболее часто встречающиеся объекты производства, а также объекты, длительность использования которых в производственных процессах максимальна. В первую очередь сокращается номенклатура объектов и устраняются процессы, редко используемые в производстве, которым может быть достаточно эффективно предоставлена замена.

ЛИТЕРАТУРА

1. Статистический анализ конструктивных элементов и технологических параметров деталей машин / М.Л. Хейфец [и др.]. – Новополоцк: ПГУ, 2001. – 112 с.
2. Сигорский, В.П. Математический аппарат инженера / В.П. Сигорский. – Киев: Техника, 1977. – 788 с.
3. Вудворт, Ф.М. Теория вероятности и теория информации с применениями в радиолокации / Ф.М. Вудворт. – М.: Сов. радио, 1955. – 128 с.
4. Ершов, А.А. Робастный фильтр Кальмана в дискретном времени / А.А. Ершов, Р.Ш. Липцер // Автоматика и телемеханика. – 1978. – № 3. – С. 60 – 70.
5. Мудров, В.И. Методы обработки измерений. Квазиправдоподобные оценки / В.И. Мудров, В.Л. Кушко. – М.: Радио и связь, 1983. – 304 с.
6. Бронштейн, И.Н. Справочник по математике для инженеров и учащихся вузов / И.Н. Бронштейн, К.А. Семендяев. – М.: Наука, 1981. – 720 с.
7. Двайт, Г.Б. Таблицы интегралов и другие математические формулы / Г.Б. Двайт. – М.: Наука, 1969. – 228 с.
8. Кокс, Д. Прикладная статистика, принципы и примеры / Д. Кокс, Э. Снелл. – М.: Мир, 1984. – 200 с.
9. Пасхвер, И.С. Закон больших чисел и статистические закономерности / И.С. Пасхвер. – М.: Статистика, 1974. – 151 с.

Поступила 22.01.2010

STATISTICAL DATA PROCESSING IN DECISION MAKING SUPPORT SYSTEM WITH REGARD TO MISCLASSIFICATION AND CODING OF OBJECTS AND PROCESSES

L. KOSYAK, S. KUKHTA, V. TOCHILO, M. KHEIFETZ

It is shown, that forming procedures of limitative enumerations of plants and production processes, creation procedures of technological operations regulations depend on specificity of concrete manufacturing. For the firms exhausting the big nomenclature of products and rendering the wide spectrum of services, different methods of forming of limitative enumerations are offered. Choice of restricted number of positions for unification and consequent regulation is recommended for producing in the defined sequence according to frequency of application or a duration of usage of objects and production processes. Most frequently meeting objects of production and also objects, which duration of usage in productions it is maximum have the highest priority. The highest priority meeting plants of production, and also plants which duration of usage in productions is maximum have most frequently. First of all the nomenclature of plants is devided out, processes seldom used in production are eliminated, if their substitution can be effectively enough produced.