

УДК 622.012

МЕТОДОЛОГИЯ И АЛГОРИТМИЧЕСКАЯ ОСНОВА ПОСТРОЕНИЯ ШКАЛЫ С ДИНАМИЧЕСКИМИ ОГРАНИЧЕНИЯМИ

С. В. ПРОКОПЧИНА

*профессор, доктор технических наук, Финансовый университет
при Правительстве Российской Федерации,
Москва, Россия*

E-mail: svprokopchina@mail.ru

В статье рассмотрены методология и методика создания шкал с динамическими ограничениями, ориентированных на применение в условиях неопределенности. Приведены формульные зависимости для определения границ классов реперов шкалы. Определены и формализованы условия различимости классов реперов шкалы. Приведен пример применения такого типа шкал для измерения коэффициентов асимметрии и эксцесса в алгоритме аппроксимации законов распределения случайных величин и процессов типовыми распределениями из системы Пирсона. Рассмотрен пример группирования выборок.

Ключевые слова: методика создания шкал с динамическими ограничениями, формульные зависимости для определения границ классов реперов шкалы, группирования выборок.

Investment, financial and management analysis

METHODOLOGY AND ALGORITHMIC BASIS OF SCALE CONSTRUCTION WITH DYNAMIC CONSTRAINTS

S. V. PROKOPCHINA

*Professor, Doctor of Technical Sciences, Financial University
under the Government of the Russian Federation,
Moscow, Russia*

E-mail: svprokopchina@mail.ru

The article describes the methodology and technique of creating scales with dynamic constraints, focused on the application of uncertainty. Formula dependences for determination of limits of classes of reference points of a scale are given. Defined and formalized the terms of the distinctiveness of the classes of reference points of the scale. An example of the application of this type of scales for measuring the coefficients of asymmetry and kurtosis in the algorithm of approximation of the distribution laws of random variables and processes of typical distributions from the Pearson system. An example of the grouping of the samples.

Keywords: *method of creating scales with dynamic constraints, formula dependences for determining the boundaries of the scale reference classes, grouping of samples.*

В работах автора при создании теоретических основ и практических методик измерения и управления сложными системами в условиях неопределенности [1–3] были предложены концепция и принципы построения шкалы динамическими ограничениями (ШДО).

В основе принципа построения шкалы с динамическими ограничениями положено байесовское решающее правило. Это правило, обеспечивающее выбор модели плотности вероятности измеряемого объекта, основывается на сравнении определенной меры близости R исследуемого распределения H_i с распределением каждого репера [ШДО]. Выдвигается гипотеза H_i в том случае, если мера близости R с H_i -м классом превышает меру близости с любым дру-

совское решающее правило. Это правило, обеспечивающее выбор модели плотности вероятности измеряемого объекта, основывается на сравнении определенной меры близости R исследуемого распределения H_i с распределением каждого репера [ШДО]. Выдвигается гипотеза H_i в том случае, если мера близости R с H_i -м классом превышает меру близости с любым дру-

гим j -м эталонным классом $j = 1, \dots, m, j \neq i$ или аналогично [3]:

$$R(H_1 H_i) = \text{extr}_i R(H_1 H_i), i = 1, 2, \dots, m, \quad (1)$$

где m – количество классов эталонных гипотез.

БРП минимизирует средний риск статистических решений, который применительно к рассматриваемому алгоритму может быть записан в виде:

$$\rho = \sum_{j=1}^m \sum_{i=1}^m P(H_i) \prod_{ji} \int f_i(a_1, \dots, a_n) da_1 \dots da_n, \quad (2)$$

где $P(H_i)$ – априорная вероятность появления эталонной гипотезы H_i ;

Π_{ji} – элементы платежной матрицы, которые отражают цену выбора H_i эталонной гипотезы, если в действительности исследуемое распределение принадлежит -му эталонному классу ($j, i = 1, \dots, m$);

$f_i(a_1, \dots, a_n)$ – условная плотность вероятности значений признаков $a_1, \dots, a_n$ в i -м классе;

L_j – область пространства признаков, соответствующая H_j эталонной гипотезе.

Элементы платежной матрицы представляют собой цены правильных и неправильных решений о принадлежности исследуемого распределения определенной эталонной гипотезе. Если установить различия в плате за отдельные решения представляет значительные трудности, то, как правило, цены правильных решений устанавливаются равными нулю, а в целом матрица потерь берется симметричной. В соответствии с этим, дифференцируя (2) по признакам $a_1, a_2, \dots, a_n$ и приравнявая производную к нулю, получают универсальное уравнение, из которого можно определить значения разделяющих границ n -мерного пространства ШДО [3]:

$$\frac{f_i(a_1, \dots, a_n)}{f_j(a_1, \dots, a_n)} = \frac{P(H_j) \Pi_{ij}}{P(H_i) \Pi_{ji}}. \quad (3)$$

Реперы ШДО можно рассматривать как признаки параметров сложных систем. Приведем пример построения двумерной для аппроксимации распределений унимодальными типовыми распределениями, подробно рассмотренный в [1]. Признаками в таком алгоритме являются коэффициенты асимметрии и эксцесса, сочетания которых единственным образом определяют все эталонные классы системы аппроксимирующих распределений, которая представляется в

виде многомерной ШДО. Двумерное совместное распределение признаков вследствие их коррелированности, в общем виде может быть записано как плотность вероятности совместного распределения зависимых величин:

$$f(\tilde{A}_i, \tilde{E}_i) = f(\tilde{A}_i) \cdot f\left(\frac{\tilde{E}_i}{\tilde{A}_i}\right), \quad (4)$$

где $f(\tilde{A}_i)$ – плотность распределения выборочного коэффициента асимметрии $\tilde{A}$, с математическим ожиданием, равным A_i , и среднеквадратическим отклонением σ_{A_i} ;

$f(\tilde{E}_i/A_i)$ – условная плотность распределения выборочного коэффициента эксцесса $\tilde{E}$, вычисленная при условии, что значение коэффициента асимметрии равно A_i .

Ввиду асимптотической нормальности распределений выборочных значений коэффициентов асимметрии и эксцесса, а также симметричности матрицы потерь, формулы для вычисления значений разделяющих границ двумерного пространства признаков могут быть записаны аналогично [1] следующим образом: ($\sigma A_i \neq \sigma A_j$)

$$x_i = \frac{A_j \sigma_{A_i}^2 - A_i \sigma_{A_j}^2 - \sigma_{A_j} \sigma_{A_i} \sqrt{(A_j - A_i)^2 + 2(\sigma_{A_i}^2 - \sigma_{A_j}^2) \ln \frac{P(A_j) \sigma_{A_i}}{P(A_i) \sigma_{A_j}}}}{\sigma_{A_i}^2 - \sigma_{A_j}^2}, \quad (5)$$

при условии $A = A_i$;

$$y_i = \frac{E_j \sigma_{E_i}^2 - E_i \sigma_{E_j}^2 - \sigma_{E_j} \sigma_{E_i} \sqrt{(E_j - E_i)^2 + 2(\sigma_{E_i}^2 - \sigma_{E_j}^2) \ln \frac{\sigma_{E_i} P[(E_j)/A_i]}{\sigma_{E_j} P[(E_i)/A_i]}}}{\sigma_{E_i}^2 - \sigma_{E_j}^2}. \quad (6)$$

где $i, j = 1, \dots, n_1$

n_1 – для (5) – число эталонных распределений с различными значениями коэффициентов асимметрии;

n_1 – для (6) – число эталонных распределений с одинаковым значением коэффициента асимметрии A ;

A_i, E_i – теоретические значения коэффициентов асимметрии и эксцесса, соответствующие H_i эталонному распределению и являющиеся математическими ожиданиями их выборочных значений;

$\sigma_{A_i}, \sigma_{E_i}$ – среднеквадратические отклонения соответствующих распределений выборочных коэффициентов $\tilde{A}$ и $\tilde{E}$;

$P(A_i), P(E_i/A_i)$ – априорные вероятности появления теоретических значений коэффициентов A_i, E_i .

Преодоление априорной неопределенности, может быть проведено несколькими путями. Один из них состоит в тщательном изучении ис-

следуемых процессов, их физической природы, реальных условий их протекания и т.п., которое может быть проведено как при обработке экспериментальных данных, полученных ранее, так и с помощью накопленных теоретических сведений о процессе. Однако, возможны случаи, когда непосредственно изучить априорную информацию об исследуемом процессе представляет значительные трудности или невозможно. В литературе имеются рекомендации по использованию для конструирования априорного распределения «метода экспертов» и других специальных методов. Но и такие методы часто невозможно применить. В таком случае, обычно принимают допущение о равновероятном априорном распределении. В системе выбранных аппроксимирующих распределений возможны случаи, когда эталонным гипотезам $H_i, H_j, \dots H_{n_1}$ будет соответствовать одинаковое значение коэффициента асимметрии A_i . С учетом этого замечания априорная вероятность появления $P(A_i)$ признака асимметрии может быть вычислена по следующей формуле

$$P(A_i) = \sum_{j=1}^{n_1} P(H_j), \quad (7)$$

где n_1 – число эталонных гипотез с теоретическими значениями коэффициента асимметрии, равным A_i .

В таком случае, априорная вероятность появления теоретического значения коэффициента эксцесса, соответствующего H_i гипотезе с коэффициентом асимметрии A_i аналогично (2) равна:

$$P(E_i/A_i) = \frac{P(H_i)}{P(A_i)}. \quad (8)$$

Дисперсии выборочных значений коэффициентов асимметрии и эксцесса $D_{\hat{A}_i}, D_{\hat{E}_i/A_i}$ вычисляются в соответствии с формулами (10), приведенными в [1].

Полностью алгоритм программы построения разделяющих границ приведен в [1]. В программе предусмотрено группирование значений коэффициентов асимметрии с шагом LD , который выбирается в долях $L1$ дисперсии выборочного коэффициента асимметрии нормального распределения (как одной из самых малых для эталонных распределений).

$$LD = L1 \sqrt{\frac{6}{N}}. \quad (9)$$

Такое группирование снижает требования к величине объема выборки, необходимого для по-

строения разделяющих границ по коэффициенту асимметрии. Однако нужно иметь в виду что такое группирование приводит к большей степени сгруппированности по коэффициенту эксцесса для одного значения A_i , что приводит к значительному увеличению объема выборочных данных, необходимых при построении разделяющих границ по этому признаку. В связи с чем значения $L1$ имеет смысл выбирать меньшими единицы. Результатом работы программы является организация массивов разделяющих границ X, Y по признакам асимметрии и эксцесса соответственно, которые очерчивают области эталонных гипотез на плоскости моментов.

Условие различимости реперов ШДО определяется теоремой, доказанной автором, приведенной в [2]. Существования разделяющих границ по каждому из признаков для каждого соседних 1-го и $i+1$ -го эталонных классов, определяется из условия неотрицательности подкоренного выражения в (5), (6), которое в общем виде может быть записано следующим образом:

$$(m_i - m_{i+1})^2 \geq -2(\sigma_i^2 - \sigma_{i+1}^2) \ln \frac{P_{i+1}\sigma_i}{P_i\sigma_{i+1}}, \quad (10)$$

где m_i, σ_i^2 – теоретическое значение и дисперсия оценки признака, соответствующих H_i -му эталонному аппроксимирующему распределению;

P_i – априорная вероятность появления i -й эталонной гипотезы.

Как видно из (10) границы существуют всегда, когда дисперсии оценок признаков одинаковы или когда выполняется условие:

$$P_{i+1}\sigma_i = P_i\sigma_{i+1} \quad (11)$$

Если же $\sigma_{i+1} \neq \sigma_i$ и условие (11) не выполняется, возможны два случая, когда $\sigma_i < \sigma_{i+1}$ и $\sigma_i > \sigma_{i+1}$. В первом случае, т.е. когда дисперсия оценки признака, соответствующего i -му классу мала, при выполнении условия:

$$P_{i+1} > \frac{\sigma_{i+1}}{\sigma_i} P_i. \quad (12)$$

Когда априорная вероятность появления $i+1$ -го класса достаточно велика, построение разделяющих границ возможно для любого сколь угодно малого расстояния между теоретическими значениями признаков 1-го и $i+1$ -го классов. Во втором случае, построение разделяющих границ для таких расстояний m_i и m_{i+1} возможно при условии:

$$P_{i+1} < \frac{\sigma_{i+1} P_i}{\sigma_i}. \quad (13)$$

Если же для рассматриваемых случаев соответствующие условия (12), (13) не выполняются, то для существования разделяющих границ между i -м и $i+1$ -м эталонными классами расстояние между ними не может быть сколь угодно малым из (10) при фиксированных значениях P_i , σ_i . В связи с тем что априорные вероятности появления эталонных гипотез и классы распределений с соответствующими им значениями m_i и m_{i+1} выбираются заранее, до построения разделяющих границ, то очевидно из (10), что возможность существования разделяющих границ будет определяться соотношением дисперсий оценок признаков σ_i^2 , σ_{i+1}^2 . В соответствии с (2') дисперсии оценок коэффициентов асимметрии и эксцесса обратно пропорциональны объему выборочных данных N . Поэтому возникает вопрос выбора необходимого объема выборочных данных, при котором разделяющие границы по всем признакам для всех классов существуют. Формульные зависимости для определения необходимого объема выборочных данных получены автором и впервые приведены в [1].

Список литературы

1. Прокопчина С. В. Диссертация на соискание ученой степени кандидата технических наук «Разработка метода аппроксимации плотности вероятности случайной величины и случайного процесса типовыми распределениями», Л., ЛЭТИ, 1976.
2. Прокопчина С. В. Диссертация на соискание ученой степени доктора технических наук. СПб, 1996.

3. Недосекин Д. Д., Прокопчина С. В., Чернявский Е. А. Интеллектуализация информационных технологий измерительных процессов. Л., Энергоатомиздат, 1996, 187 с.
3. Прокопчина С. В. Создание развивающихся информационных технологий на основе регуляризирующего байесовского подхода. Сборник докладов IX Международной конференции «Мягкие вычисления и измерения», СПб, ЛЭТИ, 2007 г
5. Прокопчина С. В. Концепция байесовской интеллектуализации измерений в задачах мониторинга сложных объектов // Новости искусственного интеллекта. 1997. № 3. С. 7–56.

References

1. Prokopchina S. V. (1976) Thesis for the degree of candidate of technical Sciences "Development of the method of approximation of probability density of a random variable and a random process by typical distributions", L., LETI.
2. Prokopchina S. V. (1996) Thesis for the degree of doctor of technical Sciences. St. Petersburg.
3. Nedosekin D. D., Prokopchina S. V., Chernyavsky E. A. (1996) Intellectualization of information technologies of measuring processes. L., Energoatomizdat. 187 p.
3. Prokopchina S. V. (2007) Development of emerging information technologies on the basis of regularizing Bayesian approach. Collection of reports of the IX international conference "Soft computing and measurement", St. Petersburg, LETI.
5. Prokopchina S. V. (1997) Concept of Bayesian intellectualization of measurements in the problems of monitoring of complex objects. *News of artificial intelligence*. No. 3. Pp. 7–56.