ЭКОНОМИКА И МАТЕМАТИЧЕСКИЕ МЕТОДЫ 1991 Том 27, вып. 3

ЗАМЕТКИ И ПИСЬМА

РАНДОМИЗИРОВАННЫЙ КРИТЕРИЙ СРАВНЕНИЯ КАЧЕСТВА СЛОЖНЫХ ОБЪЕКТОВ

Рожков Н.Н.

(Ленинград)

Качество многопараметрических объектов (изделий, проектов, управленческих решений и т. п.) характеризуется вектором значений признаков, каждый из которых отражает определенное свойство объекта или составляющую его качества. Если требуется оценить выходное качество объекта в целом, то итоговый показатель Q должен быть синтезирован по некоторому правилу и зачастую ищется в виде линейной комбинации

$$Q = p_1 x_1 + \ldots + p_m x_m, \tag{1}$$

где $x_1, \ldots, x_m$ – значения признаков отдельных свойств объектов; $p_1, \ldots, p_m$ – весовые коэффициенты, отражающие степень влияния признаков на результирующий показатель качества, т. е. их сравнительную важность.

Существуют различные критерии подсчета (оценки) коэффициентов p_i, причем иногда некоторый произвол в выборе такого критерия порождает критику методов, основанных на функционалах (1) с фиксированными значениями p_i. В связи с этим целесообразно использовать рандомизацию весовых коэффициентов, рассматривая их как реализацию m-мерной случайной величины $(P_1, \ldots, P_m)$, определенной на симплексе

$$S = \{(p_1, \ldots, p_m): \sum_{i=1}^{m} p_i = 1; \quad p_i \geqslant 0\}.$$

В [1, 2] дополнительно предполагается, что все P_i могут принимать лишь конечное множество возможных значений

$$P_i \in R_N = \{0, \frac{1}{N}, \ldots, \frac{N-1}{N}, 1\}, \tag{2}$$

где N – заданное натуральное число, $N > m$. Общее число L возможных реализаций m-мерного случайного вектора весов при этом также конечно: $L = \binom{N+m-1}{N}$. Распределение вектора весов будет определено, если задать распределение вероятностей непосредственно на множестве всех L возможных реализаций. Если, в частности, есть основания считать их в равной степени приемлемыми, то каждой из них сопоставляется вероятность L^{-1}.

В более общей постановке задачи можно полагать, что вектор весовых коэффициентов подчиняется распределению Дирихле (этот факт сокращенно обозначим: $(P_1, \ldots, P_m) \in D\ (\alpha_1, \ldots, \alpha_m)$), т. е. плотность распределения в любой точке множества S имеет вид

$$f(p_1, \ldots, p_m) = \Gamma(\alpha_1 + \ldots + \alpha_m) \prod_{i=1}^{m} \frac{p_i^{\alpha_i - 1}}{\Gamma(\alpha_i)} \tag{3}$$

и равна нулю вне этого множества.

Положительные константы $\alpha_1, \ldots, \alpha_m$ могут в данной задаче иметь смысл меры важности отдельных признаков объекта.

Таким образом, показатель качества (1) и, следовательно, все выводы, основанные на нем, приобретают вероятностный характер. Однако это неудобство безусловно окупается тем, что такой подход позволяет охватить весьма широкий класс наборов весовых коэффициентов.

Пусть объекты $O^{(j)}$ характеризуются значениями признаков $x_1^{(j)}, \ldots, x_m^{(j)}$. В качестве критерия для сравнения качества двух объектов $O^{(1)}$ и $O^{(2)}$ рассмотрим вероятность

$$\pi_{1,2} = P\{Q_1 > Q_2\}, \tag{4}$$

где

$$Q_j = \sum_{i=1}^{m} P_i x_i^{(j)}, \qquad j = 1, 2, \tag{5}$$

– значения рандомизированного показателя качества (1) сравниваемых объектов. Зная вероятность (4), можно принимать решения о предпочтительности объектов: $\pi_{1,2} > 0{,}5 + \gamma \Rightarrow$ объект $O^{(1)}$ предпочтительнее, $\pi_{1,2} < 0{,}5 - \gamma \Rightarrow$ объект $O^{(2)}$ предпочтительнее, $|\pi_{1,2} - 0{,}5| < \gamma \Rightarrow$ объекты значимо не различаются, где γ – заданный уровень значимости различия.

Цель данной работы – нахождение критерия (4) для двух частных случаев модели рандомизации, имеющих практический интерес.

Пусть $Z_1, \ldots, Z_m$ – независимые в совокупности неотрицательные случайные величины, такие, что Z_i имеет плотность гамма-распределения

$$f_{z_i}(x) = \frac{x^{\alpha_i - 1}}{\Gamma(\alpha_i)} \; e^{-x}, \quad x > 0, \tag{6}$$

где $\alpha_i > 0$. Тогда, как известно [3], для случайных величин

$$Y_i = \frac{Z_i}{Z_1 + \ldots + Z_m} \tag{7}$$

выполняется $(Y_1, \ldots, Y_m) \in D(\alpha_1, \ldots, \alpha_m)$.

Представив рандомизированные веса P_i в виде (7), запишем вероятность (4)

$$\pi_{1,2} = P\{\sum_{i=1}^{m} \Delta_i Z_i > 0\} = P\{H_1 > H_2\}, \tag{8}$$

где $\Delta_i = x_i^{(1)} - x_i^{(2)}$, $i = 1, \ldots, m$. Случайные величины

$$H_1 = \Sigma \;\; \Delta_i Z_i, \quad H_2 = \Sigma \;\; \tilde{\Delta}_j Z_j \tag{9}$$

построены так, что для H_1 суммирование ведется по всем i, для которых $\Delta_i > 0$, а для H_2 – по всем j, для которых $\tilde{\Delta}_j = -\Delta_j > 0$. Таким образом, обе величины в (9) – суммы, содержащие соответственно m_1 и m_2 положительных слагаемых $(m_1 + m_2 \leqslant m)$.

Согласно [2], если выполняется ограничение (2) и все реализации равновероятны, то при $N \to \infty$ для любого $k \leqslant m$ вектор рандомизированных весов $(P_1, \ldots, P_k)$ сходится (по функции распределения) к случайной величине $(Y_1, \ldots, Y_k)$

$$(P_1, \ldots, P_k) \to (Y_1, \ldots, Y_k)$$

такой, что

$$(Y_1, \ldots, Y_k, 1 - \sum_{i=1}^{k} Y_i) \in D(1, \ldots, 1, m - k). \tag{10}$$

Следовательно, предельный переход в модели с конечным числом равновероятных реализаций приводит к частному случаю $\alpha_i = 1$, $i = 1, \ldots, m$, модели, основанной на распределении Дирихле (3).

Рассмотрим подробнее этот частный случай. Поскольку случайные величины $X_i = \Delta_i Z_i$ имеют при $x > 0$ плотность

$$f_{X_i}(x) = \frac{1}{\Delta_i} \, e^{-x/\Delta_i}, \tag{11}$$

нетрудно получить, применяя индукцию по числу слагаемых, что сумма $X_1 + \ldots + X_k$ слагаемых с плотностью (11), подчиняется (при $\Delta_i \neq \Delta_l$) распределению с плотностью

$$f(x) = \sum_{i=1}^{k} \frac{\Delta_i^{k-2} \, e^{-\frac{x}{\Delta_i}}}{\prod\limits_{l=1}^{k} (\Delta_i - \Delta_l)}, \; l \neq i, \quad x > 0, \tag{12}$$

называемому обобщенным распределением Эрланга $(k = 2, \ldots)$.

Формула (12) позволяет непосредственно получить плотности распределения вероятностей H_1 и H_2, определяемых в (9), полагая $k = m_1, m_2$. Теперь, применив очевидное соотношение

$$\pi_{1,2} = \int F_{H_2}(x) \, dF_{H_1}(x),$$

где F – функция распределения указанной в индексе случайной величины, после несложных преобразований получим окончательно

$$\pi_{1,2} = \sum_{j=1}^{m_2} \sum_{i=1}^{m_1} \frac{\Delta_i^{m_1-1}}{\prod\limits_{\substack{k=1 \\ (i \neq k)}}^{m_1} (\Delta_i - \Delta_k)} \frac{\widetilde{\Delta}_j^{m_2-1}}{\prod\limits_{\substack{k=1 \\ (j \neq k)}}^{m_2} (\widetilde{\Delta}_j - \widetilde{\Delta}_k)} \frac{\Delta_i}{\Delta_i + \widetilde{\Delta}_j}. \tag{13}$$

Если среди величин Δ_i (или $\widetilde{\Delta}_j$) в (9) имеются равные между собой, например: $\Delta_1 = \Delta_2 = \ldots = \Delta_k = \Delta$, то, как известно, сумма k слагаемых с плотностью (11) здесь будет подчиняться распределению Эрланга с плотностью

$$\frac{x^{k-1}}{(k-1)!\,\Delta^k} e^{-x/\Delta}$$

при $x > 0$. Для этого случая также можно находить искомую вероятность $\pi_{1,2}$, однако рассчитываемые формулы окажутся гораздо сложнее (содержат неполную гамма-функцию).

В качестве иллюстрации рассмотрим два объекта, характеризуемые пятью признаками, значения которых таковы

	x_1	x_2	x_3	x_4	x_5
$O^{(1)}$:	0,2	0,8	0,2	0,9	0,4
$O^{(2)}$:	0,7	0,7	0,8	0,5	0,2.

(14)

Все признаки для удобства измерены в относительных единицах: значения 0 и 1 показывают соответственно наиболее низкую и наиболее высокую степени проявления того или иного признака. Предполагается, что более высокие степени соответствуют и более высокому итоговому качеству объекта. Легко видеть, что по признакам x_2, x_4 и x_5 предпочтительнее объект $O^{(1)}$, а по остальным двум – $O^{(2)}$.

Найдем $m_1 = 3$ значений Δ_i: 0,1 0,4, 0,2, а также $m_2 = 2$ значений $\widetilde{\Delta}_j$: 0,5, 0,6. Затем, применив (13), получаем $\pi_{1,2} = 0{,}35$, т. е. при любом уровне различимости γ, удовлетворяющем условию: $\gamma < 0{,}15$, объект $O^{(2)}$, согласно критерию (4), следует считать более предпочтительным.

В приведенном примере, как нетрудно подсчитать, математические ожидания рандомизированных показателей качества (5) данных двух объектов будут: $E(Q_1) = 0{,}50$, $E(Q_2) = 0{,}58$, однако сама по себе разность этих математических ожиданий еще не позволяет судить о том, следует ли данные объекты считать существенно различающимися по их качеству. Это лишний раз оправдывает использование критерия (4).

Применяя предельный переход (10) и представление (7), легко получить явный вид распределения рандомизированного показателя качества (5)

$$F_{Q_j}(x) = P\{\sum_{i=1}^{m} (x_i^{(j)} - x) Z_i < 0\},$$

и тогда, положив $\Delta_i = x_i^{(j)} - x$, можно воспользоваться (9) – (12).

Заканчивая рассмотрение первого частного случая модели (3), (5), заметим, что при производных значениях α_i явно выражение критерия (4) рассчитать много труднее. Однако этот общий случай представляет разве лишь теоретический интерес, поскольку на практике выбор конкретных значений α_i, т. е. численная оценка сравнительной важности признаков объекта, может вызвать затруднения. К тому же любой выбор и фиксация этих констант всегда уязвимы для критики почти в той же мере, что и выбор фиксированных значений самих весовых коэффициентов в нерандомизированной модели (1). Поэтому в отличие от численных оценок важности признаков более естественной для данной задачи является информация об упорядочении признаков по степени их влияния на итоговое качество объекта. Такого рода данные в каждой практической ситуации могут быть получены от экспертов: они более достоверны, устойчивы и легко представимы в виде системы отношений порядка

$$\{x_i \Leftarrow x_j\},$$

понимаемых, как "признак x_i не менее важен, чем признак x_j".

Соответствующие числовые неравенства для рандомизированных весов

$$\{P_i \geqslant P_j\}$$

должны выполняться с вероятностью единица.

Рассмотрим частный случай, когда все признаки могут быть линейно упорядочены

$$x_m \Leftarrow \ldots \Leftarrow x_2 \Leftarrow x_1.$$

Зададим на симплексе $S^* = \{(p_1, \ldots, p_m): \sum_{i=1}^{m} p_i = 1, 0 \leqslant p_1 \leqslant \ldots \leqslant p_m\}$ равномерное распределение, считая в равной мере допустимым и любые векторы рандомизированных весов $\overline{P} = (P_1, ..., P_m)$, удовлетворяющих с вероятностью единица условиям

$$\sum_{i=1}^{m} P_i = 1, \quad 0 \leqslant P_1 \leqslant ... \leqslant P_m. \tag{15}$$

Нетрудно установить, что симплекс S переводится в S^* линейным изоморфным отображением, определяемым нижней треугольной матрицей C с элементами

$$c_{ij} = \begin{cases} \dfrac{1}{m-j+1}, & i \geqslant j, \\ 0, & i < j. \end{cases}$$

Представив вектор $\overline{P} \in S^*$ в виде $\overline{P} = C\overline{T}$, где $\overline{T} = (T_1, ..., T_m) \in S$ – новый вектор весов, сведем модель с упорядоченными весами (15) к исходной модели.

Вернемся для иллюстрации к исходным данным (14). Пусть теперь в отличие от рассмотренного ранее дополнительно известно, что $x_5 \Leftarrow x_4 \Leftarrow ... \Leftarrow x_1$, т. е. наименее существенным для итогового показателя качества является первый признак, а наиболее существенным – пятый.

Итоговый показатель (5) качества объекта $O^{(j)}$ запишем в матричной форме

$$Q_j = X_j \overline{P} = X_j C\overline{T}, \tag{16}$$

где X_j – вектор-строка значений исходных признаков для объекта j.

Для данных (14) (после их умножения на C) выражения (16) примут вид

$$Q_1 = 0{,}50\, T_1 + 0{,}575\, T_2 + 0{,}50\, T_3 + 0{,}65\, T_4 + 0{,}40\, T_5,$$

$$Q_2 = 0{,}58\, T_1 + 0{,}550\, T_2 + 0{,}50\, T_3 + 0{,}35\, T_4 + 0{,}20\, T_5.$$

Теперь легко находим $m_1 = 3$ значений Δ_i: 0,025; 0,30; 0,20, а также $m_2 = 1$ значение $\widetilde{\Delta}_j$: 0,08. Применив (13), находим $\pi_{1,2} = 0{,}953$. Таким образом, почти для всех наборов весовых коэффициентов итоговый показатель качества у первого объекта оказывается выше, чем у второго, т. е. получен вывод, противоположный тому, который был сделан в случае, когда упорядочение признаков не учитывалось. Разумеется, при ином упорядочении признаков (что легко достигается их перенумерацией) результат и вывод могут быть иными.

Данный метод легко реализуется на ЭВМ и позволяет находить оценки и строить доверительные интервалы для рандомизированного показателя качества, принимать управленческие решения о предпочтимости одних объектов по сравнению с другими, а также о значимости таких предпочтений. В отличие от прямого перебора на ЭВМ всех допустимых наборов весов $P_i \in R_N$ (при конечном N), данный метод дает заметную экономию машинного времени, допуская работу с ЭВМ в диалоговом режиме.

ЛИТЕРАТУРА

1. *Хованов Н.В.* АСПИД – система квалиметрических методов оценивания в условиях дефицита информации качества сложных технических объектов // Методология и практика оценки качества продукции на ленинградских предприятиях. Л.: ЛДНТП, 1988.
2. *Корников В.В., Колесникова О.Н., Рожков Н.Н., Хованов Н.В.* Стохастические процессы с равновероятными монотонными реализациями, моделирующие дефицит информации // Вестн. ЛГУ. 1987. № 1.
3. *Уилкс С.* Математическая статистика. М.: Наука, 1967.

Поступила в редакцию
6 IV 1989