1.2. МОДУЛЬНЫЙ АНАЛИЗ СЛОЖНЫХ СИСТЕМ И СЛОЖНЫЕ СОСТАВНЫЕ СУММЫ

Малиновский В.К., ЦЭМИ РАН, Москва, Россия

Статья посвящена модульному анализу сложных составных сумм случайных величин, возникающих при исследовании различных сложных систем в экономике, технике, и проч. Составные суммы представляют значительный теоретический интерес, например, как вспомогательный инструмент при анализе марковских зависимостей. Они представляют значительный практический интерес, поскольку к таким суммам сводится ряд известных динамических моделей, начиная со стандартных моделей теории восстановления. Сложные составные суммы лежат в основе ряда многопериодических, с несколькими периодами, обычно от трех до семи, динамических моделей. В таких моделях, представляющих интерес для анализа различных количественных методов регулирования, модели, относящиеся к одному периоду, обычно статичны и многокомпонентны.

1. Модульность: способ управления сложностью.

К анализу природных, социальных, экономических и технических сложных систем (см., например, [Anscombe,1952], [Callebaut et al.,2005], [Holling, 1973], [Ito, 2008], [Langlois,1999]) применяется модульный подход. Под модульностью (см. [Baldwin et al., 2000]) обычно понимается то, что компоненты сложной системы сперва разделяются на модули, а потом объединяются специальным образом для достижения управляемости системой или для повышения гибкости управления системой.

Модульный подход имеет очевидное практическое значение. Примером является (см. [Brooks,1975]) разработка операционной системы OS/360 для линейки компьютеров IBM 360, которая сначала была организована неразложимым образом. Изначально считалось, что каждый программист должен быть знаком со всей текущей работой, т.е. должен иметь в своем распоряжении копию общей рабочей документации. Но это быстро привело к нарастанию проблемам в управлении проектом, в частности из-за того, что даже ведение такой документации стало отнимать значительное время из-за постоянного роста ее объема.

Возник вопрос, нужно ли программистам, разрабатывающим один конкретный модуль, знать о деталях всего проекта. Утверждалось, что, особенно в крупных проектах, ответ отрицательный: руководители должны стремиться к минимизации несущественных взаимозависимостей (см. [Parnas, 1972]). Информация, относящаяся только к отдельному модулю, если она не может повлиять на другие составляющие проекта, не должна передаваться прочим исполнителям. Эта идея была сформулирована как «сокрытие информации», ключевое понятие в современной объектно-ориентированном подходе к программированию.

3. Многокомпонентные модели экологического регулирования.

Модульный подход, обычно приводящий к появлению составных сумм, применим ко многим техническим и экономическим проблемам. Приведем два примера многокомпонентных моделей экологического контроля, когда промышленность действует в рамках государственного регулирования.

Пример 1 (Простая модель). Пусть отрасль состоит из ряда производственных цепочек, т. е. групп предприятий, находящихся в зависимости друг от друга. Эти производственные цепочки называются единицами или модулями. Если зависимость между цепочками намного меньше, чем внутри цепочек, то при использовании модульного подхода можно делать предположения о независимости между модулями. Пусть t – общий объем промышленных отходов, например, сточных вод или оксидов углерода, выбрасываемых в атмосферу, который регулирующий орган считает допустимым. Пусть T_i – отходы, выбрасываемые i-м модулем, а X_i – общий продукт, произведенный $i-m$ модулем. В некоторых случаях естественно предположить, что это случайные величины. Тогда $N_t := \max\{k \geqslant 1 : \sum_{i=1}^{k} T_i \leqslant t\}$, или 0, если $T_1 > t$ – это количество групп предприятий, или модулей, которые исчерпывают лимит промышленных отходов, а составная сумма $S_{N_t} := \sum_{i=1}^{N_t} X_i$ – это выход, или общий продукт, выпускаемый этими группами.

С одной стороны, эта модель излишне проста. Так, в ней не учитывается динамика по годам, изменение ситуации от года к году, зависимость между группами предприятий, которые считаются независимыми модулями. Однако эта модель лежит в основе построения многошаговой модели по нескольким годам, включающей все перечисленное. Такая модель аналогична многошаговой модели долгосрочного страхового регулирования (см. [Малиновский, 2020]) и будет рассматриваться отдельно.

Модель примера 1 может способствовать выработке гибких и обоснованных требований к производствам, являющимся источником загрязнений. Не запрещая такие производства, органы экологического контроля предъявляют количественные требования на основе анализа изменяющейся по годам (обычно, с горизонтом планирования не выше пяти-семи лет) ситуации с экологической безопасностью. С точки зрения постановки и методов исследования, к модели примера 1 близки модели, возникающие при рассмотрении мер риска в страховании (см. [Malinovskii, 2021a] и [Malinovskii,2021b].

Заметим, что ни приведенная выше модель, ни ее многошаговое расширение, не являются динамическими моделями в смысле теории восстановления, широко известными в теории массового обслуживания и теории надежности технических восстанавливаемых систем.

Пример 2 (Усложненная модель). Предположим, что в модели примера 1 каждая производственная цепочка снабжена очистным сооружением, которое частично, полностью или даже избыточно перерабатывает сточные воды или оксиды углерода, т. е. совершает действие, противоположное загрязнению: очистку, рекультивацию или мелиорацию окружающей среды. Тогда T_i могут принимать отрицательные значения. В этом случае общий промышленный выход моделируется сложной составной суммой S_{N_t}, базис которой образован независимыми случайными векторами $(T_i, X_i), i = 1,2, ...$, где T_i принимают не только положительные, но и отрицательные значения.

С точки зрения приложений модель примера 1 мало отличается от модели примера 2, но с позиций математического исследования разница существенная: если $T_i, i = 1,2, ...$, не положительны, то суммы $V_n := \sum_{i=1}^{n} T_i$ не возрастают монотонно с ростом n и множество

$$\{N_t = n\} = \{V_1 \leqslant t, \dots, V_n \leqslant t < V_{n+1}\}$$

не совпадает с множеством

$$\{N_t = n\} = \{V_n \leqslant t < V_{n+1}\}.$$

3. Имитационное моделирование и аналитические методы.

Для широкого круга пользователей имитационное моделирование характеристик составной суммы $S_{N_t} := \sum_{i=1}^{N_t} X_i$ представляются естественным и часто единственным

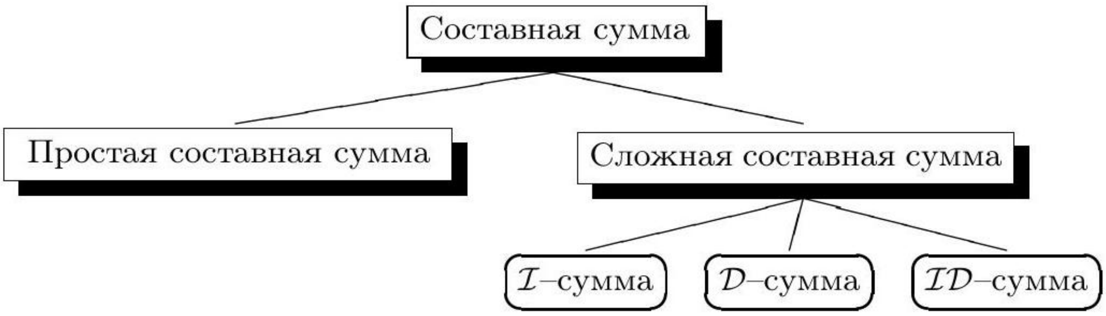

Рис. 1. Классификация составных сумм.

методом реализации модульного анализа. К сожалению, имитационное моделирование имеет ряд врожденных недостатков. Одним из них является относительно низкая точность при заданном объеме вычислений. По существу, имитационное моделирование похоже на определение вероятности выпадения орла и решки с помощью многократного подбрасывания монеты. Известно, что для определения этой вероятности с заданной точностью необходимо произвести большое количество бросков; количественный ответ на вопрос, сколько именно их нужно, дает теорема Муавра-Лапласа. Другой недостаток заключается в том, что имитационное моделирование дает только частные, численные результаты, т. е. охватывает только ситуацию, относящуюся к выбранному набору параметров модели.

Наоборот, если имеется аналитическое решение, то оно дает общее видение для ситуаций, относящихся ко всему набору параметров модели, хотя и в тех условиях, при которых такое решение было получено. Поэтому аналитический подход, в отличие от имитационного моделирования, лучше подходит для выработки общих рекомендаций. Аналитические приближения, даже если не всегда подходят для практических вычислений, много способствуют как качеству, так и эффективности процедур имитационного моделирования.

Простые и сложные составные суммы. В моделях примеров 1 и 2 центральным объектом является составная сумма

$$S_{N_t} := \sum_{i=1}^{N_t} X_i,$$

базисом которой являются двумерные случайные векторы $(T_i, X_i), i = 1,2, ...$

По-видимому, впервые составная сумма S_{N_t} впервые появилась (см. [Dœblin,1938]) в 1938 г. при доказательстве центральной предельной теоремы для дискретной цепи Маркова, где случайные индексы $0 < \Upsilon_0 < \Upsilon_1 <$ являются моментами последовательных возвращений в некоторое фиксированное состояние. Эти случайные индексы задают модульную структуру, где независимые и одинаково распределенные (н.о.р.) блочные слагаемые $X_i := \sum_{j=\Upsilon_{i-1}+1}^{\gamma_i} \zeta_j$ складываются друг с другом до тех пор, пока соответствующая сумма блочных слагаемых $T_i := \Upsilon_i - \Upsilon_{i-1}$, положительных по определению, не превысит n. Таким образом, зависимость парадоксальным образом перетекает в независимость. Эта процедура в настоящее время известна как разложение Деблина. Дальнейшие примеры составных сумм S_{N_t} связаны с именами Ф.Дж. Анскомба (см. [Anscombe, 1952]) и Д. Блэквелла. (см. [Blackwell, 1953]) и относятся к началу 1950-х гг. С тех пор такие суммы стали основным инструментом последовательного анализа, теории надежности, теории страховых рисков, теории массового обслуживания и многих других областей прикладной теории вероятностей.

Мы классифицируем (см. Рис. 1) составные суммы следующим образом. Когда случайные векторы $(T_i, X_i) := (X, T), i = 1,2, \ldots,$ н.о.р. и их компоненты $T_i,\ i = 1,2, \ldots,$ положительны, соответствующая составная сумма S_{N_t} называется простой и ее базис называется простым.

Когда случайные векторы $(T_i, X_i), i = 1,2, \ldots,$ н.о.р. и их компоненты $T_i,\ i = 1,2, \ldots,$ не обязательно положительны, соответствующая составная сумма S_{N_t} является сложной и называется $\mathcal{I}$-суммой. Ее базис называется $\mathcal{I}$-базисом.

Когда случайные векторы $(T_i, X_i), i = 1,2, \ldots,$ являются зависимыми (следовательно, метод характеристических функций, который переводит свертки в произведения, не работает в полной мере) а случайные величины $T_i, i = 1,2, \ldots,$ положительны, соответствующая сложная составная сумма S_{N_t} называется $\mathcal{D}$ суммой. Ее базис называется $\mathcal{D}$-базисом.

Когда случайные векторы $(T_i, X_i), i = 1,2, \ldots,$ являются зависимыми, а случайные величины $T_i, i = 1,2, \ldots,$ не обязательно положительны, соответствующая сложная составная сумма называется $\mathcal{ID}$-суммой. Ее базис называется $\mathcal{ID}$-базисом.

4. Аппроксимация распределения простой составной суммы.

Обозначим плотность и функцию распределения стандартного нормального распределения через

$$\varphi_{(0,1)}(x) := \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}},\ \Phi_{(0,1)}(x) := \int_{-\infty}^{x} \varphi_{(0,1)}(z)dz.$$

Для краткости положим $\mu_T := \mathrm{E}T, \mu_X := \mathrm{E}X, \sigma_T^2 := \mathrm{D}T, \sigma_X^2 := \mathrm{D}X$, и $\varkappa_{XT} := \mathrm{E}\big((X-\mu_X)(T-\mu_T)\big)$. Для $i,j > 0, i+j > 2$ положим $\varkappa_{XT}^{(i,j)} := \mathrm{E}\big((X-\mathrm{E}X)^i(T-\mathrm{E}T)^j\big)$. Очевидно, что $\varkappa_{TX}^{(i,j)} = \varkappa_{TX}^{(j,i)}$. Положим также:

$$\begin{aligned}
M_S &:= \frac{\mu_X}{\mu_T},\ D_S^2 := \frac{\mu_T^2\sigma_X^2 - 2\mu_T\mu_X\varkappa_{XT} + \mu_X^2\sigma_T^2}{\mu_T^3} \\
A_S &:= \Big(\Big(\varkappa_{XT}^{(3,0)} - 3\sigma_X^2\varkappa_{XT}\mu_T^{-1} + 6\sigma_X^2\sigma_T^2\mu_X\mu_T^{-2} - 6\sigma_T^2\varkappa_{XT}\mu_X^2\mu_T^{-3}\Big)\mu_T^{-1} \\
&\quad -3\Big(\varkappa_{XT}^{(2,1)} - 2\varkappa_{XT}^2\mu_T^{-1} + \sigma_X^2\sigma_T^2\mu_T^{-1}\Big)\mu_X\mu_T^{-2} \\
&\quad +3\Big(\varkappa_{XT}^{(1,2)} - \sigma_T^2\varkappa_{XT}\mu_T^{-1}\Big)\mu_X^2\mu_T^{-3} - \Big(\varkappa_{XT}^{(0,3)} - 3\sigma_T^4\mu_T^{-1}\Big)\mu_X^3\mu_T^{-4}\Big)D_S^{-3} \\
B_S &:= \mu_X(\sigma_T^2\mu_T^{-2} + 1)D_S^{-1}
\end{aligned}$$

Приведем пример аналитического результата, относящегося к распределению простой составной суммы.

ТЕОРЕМА 1. Пусть функция плотности случайного вектора (T, X) ограничена сверху конечной константой, $D_S^2 > 0, \mathrm{E}(T^4) < \infty, \mathrm{E}(X^4) < \infty$. Тогда

$$\begin{aligned}
\sup_{x\in R} \mid\ & P\{S_{N_t} - M_S t \leqslant x D_S t^{1/2}\} - \Phi_{(0,1)}(x) \\
&+ \frac{1}{6}(A_S(x^2-1) + 3B_S)\varphi_{(0,1)}(x)t^{-1/2} \mid = O(t^{-1}),\ t \to \infty.
\end{aligned}$$

Этот результат, называемый асимптотическим разложением в предельной теореме для распределения составной суммы S_{N_t}, позволяет уменьшить величину t, при котором аппроксимация такого распределения удовлетворительна. В этом результате учитывается корреляция между компонентами случайного вектора (T, X) и третьи моменты этих компонент.

Для использования приближения (1) достаточно вычислить входящие в M_S, D_S^2, A_S, B_S моменты. Это можно делать аналитически, вычисляя интегралы, которыми выражаются моменты $\mu_T, \mu_X, \sigma_T^2, \sigma_X^2$ и проч., и численно, методом имитационного моделирования. Эта численная процедура намного проще той, которая потребовалась бы для расчета распределения сложной суммы S_{N_t}.

5. Метод доказательства Теоремы 1.

Воспользовавшись формулой полной вероятности, имеем

$$\begin{aligned}
\mathrm{P}\{S_{N_t} \leqslant x\} &= \sum_{n=1}^{\infty} \mathrm{P}\{S_{N_t} \leqslant x, N_t = n\} \\
&= \mathrm{P}\left\{\sum_{i=1}^{n} X_i \leqslant x, \sum_{i=1}^{n} T_i \leqslant t < \sum_{i=1}^{n+1} T_i\right\} \\
&= \sum_{n=1}^{\infty} \int_0^t \mathrm{P}\left\{\sum_{i=1}^{n} X_i \leqslant x \mid \sum_{i=1}^{n} T_i = t - z\right\} f_T^{*n}(t-z) \\
&\quad \times \mathrm{P}\{T_{n+1} > z\}dz.
\end{aligned}$$

Подынтегральные выражения подходят для применения интегро-локальных центральных предельных теорем с неравномерным остаточным членом для н.о.р. двумерных случайных векторов.

Обобщения. Теорема 1 является примером применения общего подхода к исследованию составных сумм. Впервые этот подход был предложен в [Малиновский, 1993]. Существенно более общие результаты получены для трех типов сложных составных сумм (см. Рис. 1): $\mathcal{J}$-сумм, $\mathcal{D}$-сумм и $\mathcal{JD}$-сумм. Эти результаты будут изложены отдельно.

Литература

1. Anscombe, F.J. (1952) Large–sample theory of sequential estimation, Proc. Cambridge Phil. Soc., 48, 600–607.
2. Baldwin, C.Y., and Clark, K.B. (2000) Design Rules. Vol. 1: The Power of Modularity. MIT Press, Cambridge, MA.
3. Blackwell, D. (1953) Extension of a renewal theorem, Pacific J. Math., 3, 315–320.
4. Brooks, F.P. (1975) The Mythical Man–Month: Essays on Software Engineering. Addison–Wesley, Reading.
5. Callebaut,W, and Rasskin–Gutman, D. (2005) Modularity: Understanding the Development and Evolution of Natural Complex Systems. The Vienna Series in Theoretical Biology. MIT Press, Cambridge, MA.
6. Dœblin,W. (1938) Sur deux probl`emes de M. Kolmogoroff concernant les chaˆines d´enombrables, Bull. Soc. Math. de France, 66, 210–220.
7. Holling, C.S. (1973) Resilience and stability of ecological systems, Annual Review of Ecology and Systematics, 4, 1–23.
8. Ito, Y. (2008) Modular Design for Machine Tools. McGraw–Hill, New York.
9. Langlois, R.N. (1999) Modularity in technology, organization, and society, University of Connecticut Working Paper 1999–05.
10. [Малиновский, В.К. (1993) Предельные теоремы для остановленных случайных последовательностей I. Теор. вер. применен., 38, 4, 890–826.
11. Малиновский, В.К. (2020) Модели долгосрочного страхового планирования. М.: Янус–К.
12. Malinovskii, V.K. (2021) Level–Crossing Problems and Inverse Gaussian Distributions. Chapman and Hall/CRC, Boca Raton.
13. Malinovskii, V.K. (2021) Risk Measures and Insurance Solvency Benchmarks. Chapman and Hall/CRC, Boca Raton,
14. Parnas, D.L. (1972) On the criteria to be used in decomposing systems into modules, Communications of the ACM, 15, 1053–1058.

Всеволод Константинович Малиновский, д.ф.-м.н.,
Центральный экономико-математический институт (ЦЭМИ РАН), 117418,
Нахимовский просп., 47, Москва, Россия
E-mail address: admin@actlab.ru, Vsevolod.Malinovskii@mail.ru
URL: http://www . actlab.ru

Ключевые слова

Модульный анализ, количественные методы регулирования, суммы случайных величин, сложные составные суммы, предельные теоремы, имитационное моделирование.

Vsevolod Malinovsky, Modular analysis of complex systems and complex composite sums

Keywords

Modular analysis, quantitative approach to regulation, sums of random variables, complex compound sums, limit theorems, simulation.

DOI: 10.34706/DE-2023-01-02

JEL classification C02 – Математические методы; M15 Управление информационными технологиями.

Abstract

he paper deals with modular analysis of complex compound sums of random variables that arise in the study of various complex systems in economics, technology, etc. The compound sums are of a considerable theoretical interest, e.g., as an auxiliary tool in the analysis of Markov dependencies, and of a considerable practical interest, since a number of well-known dynamical models, starting from standard models of renewal theory, are reduced to them. The complex compound sums underlie a number of multi-periodic, usually comprising from three to seven periods, dynamic models. In such models, which are of interest for the analysis of various quantitative approaches to regulation, the models related to one period are usually static and multicomponent.