Комбинированное применение аналитико-статистических методов моделирования

О. И. Кутузов[1], Т. М. Татарникова[2]
Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина)
[1]Oleg-kutuzov@mail.ru mail, [2]tm-tatarn@tandex.ru

***Аннотация*. Обсуждается задача комплексного применения расслоенной выборки и равновзвешенного моделирования при оценке редких событий в решении подобных задач позволяет существенно сократить число испытаний без потери точности анализируемых характеристик системы.**

Ключевые слова: метод Монте-Карло; равновзвешенное расслоенное моделирование; оценка редких событий; число экспериментов; эффективность по числу экспериментов

I. Введение

Метод Монте-Карло, лежащий в основе вероятностного моделирования, является универсальным методом. Однако он отличается медленной сходимостью. Требуемое число опытов может быть столь велико, что ЭВМ не в состоянии их выполнить в приемлемые сроки. Этот недостаток метода Монте-Карло особенно проявляется при анализе сложных систем и оценивании редких событий [1], [2].

Проблема точности моделирования стохастических систем является центральной в имитационном моделировании. Общей стратегией снижения цены точности является ускорение сходимости вычисляемых оценок.

В настоящее время сложился ряд подходов, позволяющих ускорить статистического моделирования, не снижая точности получаемых результатов. К ним относятся: регенеративный метод моделирования [3]; методы понижения дисперсии [4]; комбинированные и аналитико-статистические методы [5].

Наибольшего эффекта удается достичь тогда, когда методы ускорения учитывают специфику моделируемых объектов, решаемых задач и алгоритмов их решения. Так, например, сочетание расслоенной выборки [6] и равновзвешенного моделирования [5] при анализе сетевых моделей дает весьма ощутимый эффект.

II. Метод взвешивания

Суть метода взвешивания может быть продемонстрирована на статистическом эксперименте с одномерной входной случайной величиной (СВ) x (рис. 1).

Формально искомое математическое ожидание $M(y)$ определяется выражением:

$$M(y) = \int y(x)\, f(x)\, dx. \qquad (1)$$

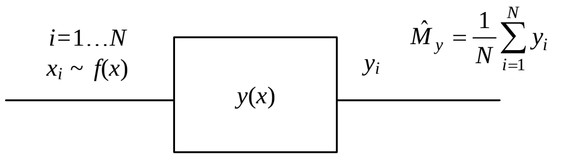

Рис. 1. Исходная схема эксперимента

Метод взвешивания основан на аналитическом преобразовании выражения (1) и на соответствующем изменении схемы эксперимента.

Разделим и умножим подынтегральное выражение в (1) на произвольную плотность распределения вероятностей $p(x)$, не равную нулю в пределах интегрирования. В результате мы увидим, что

$$(y) = \int y(x)\frac{f(x)}{p(x)}\, p(x)\, dx = M\left[y(x)\frac{f(x)}{p(x)}\right] = M(y'), \qquad (2)$$

где $y' = y(x)\, f(x)/p(x)$, а $x \sim p(x)$. Следовательно, вместо расчета оценки $\hat{M}_y$ можно выполнять расчет оценки $\hat{M}_{y'}$. Схема расчета $\hat{M}_{y'}$ представлена на рис. 2. Такой переход от исходной схемы рис. 1 к схеме рис. 2 называется методом взвешивания. Обе схемы эквивалентны с точки зрения математического ожидания выходной СВ, но различны с точки зрения ее дисперсии $D(y)$.

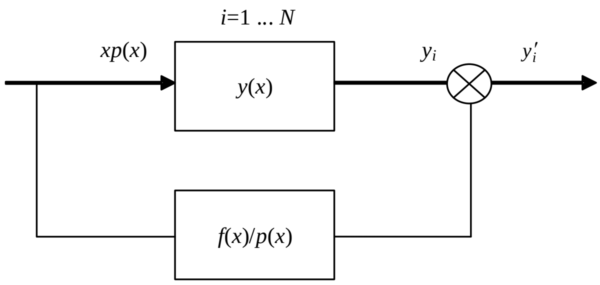

Рис. 2. Преобразованная схема

Описанный одномерный вариант метода легко распространяется и на случай многомерной СВ. Для этого достаточно в (2) вместо скаляра X записать вектор $\vec{x}=\left(x_1,...,x_n\right)$. Обобщение на случай дискретной СВ X осуществляется путем замены интеграла на сумму, а плотностей – на вероятности.

Искусство применения метода взвешивания сводится к подбору такой функции p, чтобы $D(y') << D(y)$. Тогда получится $N' << N$. Можно, например, подбирать плотность распределения вероятностей p так, чтобы СВ $y' = yf/p$ была по возможности ближе к константе, хотя бы в каком-нибудь интуитивном смысле, т.к. при $y' = \text{const}$ должно быть $D(y') = 0$.

Основное достоинство метода взвешивания – в простоте преобразования схемы эксперимента. Программа вычисления функции $y(x)$, то есть модель исследуемой системы не меняется. Добавляется только расчет множителя $f\ /\ p$ (веса), да генератор входной СВ с распределением f заменяется на генератор с распределением p.

III. Метод расслоения

Другим более распространенным методом ускоренного моделирования является метод расслоенной (стратифицированной) выборки (рис. 3).

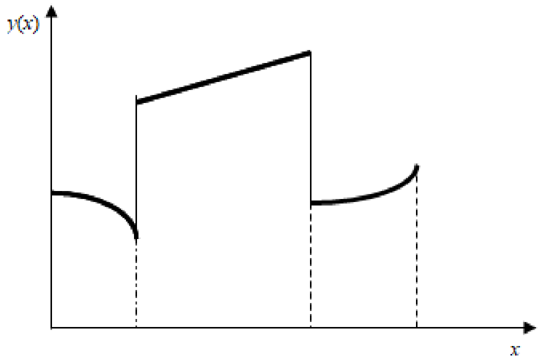

Рис. 3. Кусочно-непрерывная функция y(x) случайного аргумента x

Область значений случайного аргумента x может быть разбита на участки так, что изменение функции в пределах отдельного участка значительно меньше, чем в пределах всей области. В таком случае целесообразно сначала усреднить функцию на каждом участке отдельно, а затем из частных средних получить общее среднее с учетом вероятностей участков. Предполагается, что вероятности слоев $p_r,\ r=\overline{1,R}$ легко вычисляются через заданное распределение $p(x)$:

$$p_r = p\left\{x \in B_r\right\} = \int\limits_{B_r} p\left(x\right)dx,\ r=\overline{1,R},$$

где $B_r,\ r=\overline{1,R}$ – подмножества множества $\{x\}$ значений аргумента x, называемые слоями (стратами).

Искомый результат Q находится по формуле полной вероятности через условные математические ожидания

$$Q=\sum_{r=1}^{R}Q_r=\sum_{r=1}^{R}p_r\int\limits_{B_r}y\left(x\right)\frac{p\left(x\right)}{p_r}dx,\ r=\overline{1,R},$$

Опыты проводятся, как в обычной схеме при прямом моделировании, но по слоям. Для каждого слоя вычисляются оценки условных математических ожиданий (частные оценки)

$$\hat{Q}_r=\frac{1}{N_r}\sum_{i=1}^{N_r}y\left(x_i\right),\ x_i \sim \begin{cases}\dfrac{p(x)}{p_r}, x\in B_r\\ 0,\quad x\notin B_r\end{cases},$$

где N_r – количество опытов в r-м слое.

В итоге вычисляется искомая оценка

$$\hat{Q}=\sum_{r=1}^{R}p_r\hat{Q}.$$

Для контроля за точностью эксперимента можно в каждом слое вычислять условную дисперсию выходной СВ:

$$\hat{D}_r=\frac{1}{N_r}\sum_{j=1}^{N_r}y_j^2-\left(\hat{Q}_r\right)^2,\ x\in B_r \tag{3}$$

и дисперсию внутрислойной оценки

$$\hat{D}\left(\hat{Q}_r\right)=\frac{\hat{D}_r}{N_r}.$$

Дисперсия итоговой оценки может быть вычислена по формуле:

$$\hat{D}\left(\hat{Q}\right)=\sum_{r=1}^{R}p_r^2\hat{D}\left(\hat{Q}_r\right).$$

Если нужно знать ошибку итоговой оценки в форме коэффициента вариации, то она вычисляется как обычно:

$$\nu'^2=\frac{\hat{D}\left(\hat{Q}\right)}{\hat{Q}^2}.$$

Теоретически показано [4]: чтобы оценка $\hat{Q}$ имела меньшую дисперсию, чем обычная оценка (без расслоения), достаточно, чтобы объем выборки распределялся по слоям пропорционально их вероятностям и чтобы условные средние по слом не были одинаковы. При разбиении по слоям следует стремиться к тому, чтобы средние по слоям как можно больше отличались друг от друга.

Общее число опытов N по слоям: $N=N_1+\ldots+N_R$ распределять рационально в соответствии с формулой

$$N_r = p_r N\ ,\ \overline{1,R} \tag{4}$$

или оптимально – по формуле

$$N_r = \frac{\sigma_r\, p_r}{\sum_r \sigma_r\, p_r} N, \tag{5}$$

где $\sigma = \sqrt{D_r}$ – условное среднеквадратическое отклонение СВ y в слое B_r.

Обычно σ_r априорно неизвестны и их заменяют на оценки

$$\hat{\sigma}_r = \sqrt{\hat{D}_r},\ r = \overline{1,R},$$

где $\hat{D}_r$ находят по (3) путем выполнения пробного прогона программы с не очень большими значениями N_r.

При использовании распределения опытов (5) ошибка v' расслоенного эксперимента будет минимальной для данного значения N. Распределение (4) приводит к худшей (во всяком случае не лучшей) точности, чем (5). Тем не менее, распределение опытов (5) используется чаще, т.к. оно проще и все же гарантирует (при любом расслоении эксперимента), что ошибка получится меньше (в крайнем случае – не больше), чем в исходной схеме [4].

Таким образом, рациональное, а тем более, и оптимальное распределения опытов гарантируют, что даже при самом неудачном расслоении мы не ухудшим точности расчетов.

IV. Тестовый пример

Оценим этот возможный выигрыш при моделировании маловероятных событий на примере расчета надежности системы, структурная схема которой изображена в виде случайного графа на рис. 4 [7].

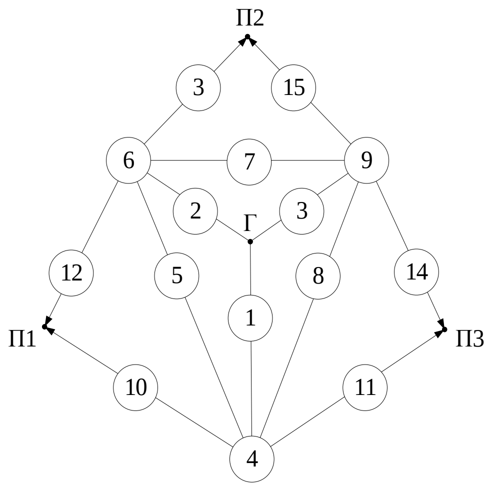

Рис. 4. Структурная схема рассматриваемой системы

Вершины (элементы системы) 1, ..., 15 имеются в графе с вероятностями $q_1, \ldots, q_{15}$ соответственно, дуги абсолютно надежны. Система работоспособна, если из полюса Г есть пути во все полюса П1, П2, П3 (движение против ориентированных дуг запрещено). Требуется найти вероятность отказа системы при следующих вариантах:

I) $q_1=q_2=q_3=10^{-5}$,

II) $q_4=q_6=q_9=2\cdot10^{-5}$,

III) $q_{10}=\ldots=q_{15}=5\cdot10^{-5}$.

Для вариантов I и II известны точные решения [6] (табл. 1, графа **A**). В [5] приведены результаты расчетов методом равновзвешенного моделирования без расслоения, содержащие значения оценок $\hat{Q}_\xi$ для $M\xi$ и оценки $\hat{\vartheta}$ для коэффициентов вариации СВ $\hat{Q}_\xi$, (табл. 1, графа **B**). Видно, что при хорошей точности оценок $\hat{Q}_\xi$ в целом, при появлении некоторого разброса в исходных данных (вариант III) значение $\hat{\vartheta}$ заметно возрастает.

ТАБЛИЦА I

Вариант	A	B		C	
		N=1000		*N=1020*	
	$M\xi\cdot10^8$	$\hat{Q}_\xi\cdot10^8$	$\hat{\vartheta}$	$\hat{m}_0\cdot10^8$	$\hat{\vartheta}$
I	12,00	12,40	0,086	11,82	0,069
II	0,120	0,1271	0,085	0,1183	0,069
III	–	1,549	0,354	1,505	0,068

Эта же задача решена методом равновзвешенного моделирования с расслоением. Система, изображенная на рис. 1, имеет связность $\gamma=2$. Анализ показывает, что для данных значений отказов элементов при расслоенном моделировании достаточно рассмотреть слои с кратностью отказов $i = \overline{2,3}$.

Разделение элементов системы на группы состояло в выделении для i=2 и i=3 одних и тех же 4 групп элементов с номерами 1,2,3; 4,6,9; 5,7,8 и 10,...,15 соответственно. При этом получены результаты, приведенные в табл. 1, графа **C**. Можно видеть, что для варианта III показатель $\hat{\vartheta}$ оказался примерно таким же, как и для первых двух вариантов.

Для рассматриваемого примера p^+=2,7·10^{-4} и $\rho_\gamma^2 = 0,303$. Согласно (2) выигрыш в числе испытаний при расслоении и равновзвешенном моделировании по сравнению с прямым моделированием по методу Монте-Карло составит t_1=2,84·10^3 раз.

V. Заключение

Комбинированное применение аналитико-статистических методов моделирования позволяет ускорить численный расчет и анализ характеристик исследуемой системы методом машинной имитации. Выполнена оценка эффективности совместного применения аналитико-статистических методов моделирования в сравнении с непосредственным моделированием по числу испытаний.

Список литературы

[1] Кутузов О.И., Татарникова Т.М. Инфокоммуникационные сети. моделирование и оценка вероятностно-временных характеристик. СПб: ГУАП, 2015. 382 с.

[2] Кутузов О.И., Татарникова Т.М. К анализу парадигм имитационного моделирования // Научно-технический вестник информационных технологий, механики и оптики. 2017. Т. 17. № 3. С. 552–558. doi: 10.17586/2226-1494-2017-17-3-552-558.

[3] Крейн М., Лемуан О. Введение в регенеративный метод анализа моделей. М.: Наука, 1983, 104 с.

[4] Клейнен Дж. Статистические методы в имитационном моделировании: Пер с англ. / Под ред. Ю.П. Адлера и В.Н. Варыгина. М. : Статистика, 1978. 556 с.

[5] Плакс Б.И. Расчет надежности систем со сложной структурой ускоренным методом Монте-Карло // Изв. АН СССР. Техническая кибернетика, 1983. № 6. С. 158-162.

[6] Kleijnen J.P.C. Design and analysis of simulation experiments. New York; London: Springer, 2008. 216 p.

[7] Рябинин И.А., Черкесов Г.Н. Логико-вероятностные методы исследования структурно-сложных систем. М.: Радио и связь,1981. 264 с.