УДК 519.24

А. В. Гурьянов

ГЕНЕРАЦИЯ РАВНОМЕРНО РАСПРЕДЕЛЕННЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ В КОМПЬЮТЕРНОМ СТОХАСТИЧЕСКОМ ЭКСПЕРИМЕНТЕ ПО МЕТОДУ МОНТЕ-КАРЛО

UNIFORMLY DISTRIBUTED SEQUENCES GENERATION IN MONTE CARLO STOCHASTIC COMPUTER SIMULATION EXPERIMENT

Ивановский государственный университет,
кафедра вычислительной и прикладной математики
153025 Иваново, ул. Ермака, д. 39

Роль имитационного компьютерного моделирования в изучении жидкокристаллической фазы и самоорганизующихся систем в целом неуклонно возрастает. В связи с этим рассматривается ряд вопросов, возникающих при планировании и оптимизации как компьютерного, так и натурного эксперимента. Обсуждаются некоторые аспекты применения в вычислительном эксперименте по методу Монте-Карло равномерно распределенных $ЛП_\tau$ последовательностей.

Computer simulation becomes very importent in liquid crystals and self-organized systems study. Some questions of uniformly distributed LPτ sequences usage in computer and full-scale Monte Carlo experiments are discussed below.

Ключевые слова: *планирование эксперимента, метод Монте-Карло, имитационное моделирование, псевдослучайная последовательность, равномерно распределенная последовательность, $ЛП_\tau$ последовательность.*

Keywords: *experiment planning, Monte Carlo method, computer simulation, pseudorandom sequence, uniform distributed sequence, LPτ sequence.*

Автоматизированная обработка информации получила широкое распространение во многих сферах научно-исследовательской и практической деятельности, начиная от обработки экспериментальных данных и заканчивая комплексными системами автоматизации научных исследований. В основу таких систем обычно закладываются компьютерные модели исследуемых процессов и явлений, которые и служат предметом вычислительного эксперимента. Модели, используемые в структурном анализе, в большинстве своем носят стохастический характер и основаны на методе Монте-Карло, точность и достоверность результатов которого существенно зависит от качества программы-генератора псевдослучайных числовых последовательностей. Рассмотрим некоторые аспекты применения методик планирования эксперимента в имитационном моделировании с целью повышения качества и достоверности его результатов. Очевидно, что имитационное моделирование следует отнести к количественному эксперименту. С этой точки зрения главное достоинство имитационных моделей состоит в том, что все факторы моделируемой системы являются контролируемыми и управляемыми. С другой стороны, очевидна высокая ресурсоемкость компьютерного моделирования стохастических систем (нехватка оперативной памяти, длительное время счета), что часто приводит к «огрублению» модели и, естественно – к снижению достоверности и точности результатов моделирования. В этом случае необходимо использовать методику активного эксперимента, которая достаточно эффективна в многофакторных моделях, например при подборе оптимального состава смеси из нескольких компонентов. Эта методика предусматривает контроль погрешности на всех этапах эксперимента и позволяет построить многомерное линейное регрессионное уравнение для отклика системы с заданной точностью, а при итерационном повторении процедуры – найти оптимальное значение отклика. Рассмотрим некоторые аспекты применения этой методики к стохастической имитационной модели в $\boldsymbol{n}$-мерном случае.

Пусть состояние системы $\boldsymbol{\Sigma}$ зависит от набора из $\boldsymbol{n}$ факторов (вектора) $\boldsymbol{X \in R^n}$, и характеризуется откликом $\boldsymbol{Y=Y(X),\ Y \in R}$. Будем считать, что факторы определены (имеют смысл) в некоторой ограниченной области факторного пространства $\boldsymbol{X \in \Omega \subset R^n}$. Под поверхностью отклика системы $\boldsymbol{\Sigma}$ будем понимать множество точек $\boldsymbol{\Psi = \Psi(X,Y(X)) \subset R^{n+1}}$, $\boldsymbol{X \in \Omega}$**.**

Область $\boldsymbol{\Omega}$ представляет собой n-мерный параллелепипед, то есть для факторов должны быть определены минимальные и максимальные допустимые значения. Затем производится нормирование значений факторов, с тем, чтобы они принадлежали отрезку $[-1,\ 1]$, таким образом, область $\boldsymbol{\Omega}$ приобретает вид $\boldsymbol{n}$-мерного куба со стороной в две единицы и центром в начале координат. Эксперимент проводится со значениями параметров, соответствующими вершинам этого куба. В каждой вершине эксперимент повторяется несколько раз для обеспечения статистической достоверности его результатов. Затем по полученным данным строится многомерное линейное уравнение регрессии. Таким образом, общее количество испытаний составит $\boldsymbol{2^n k}$, где $\boldsymbol{k}$ – количество повторов эксперимента, определяемое по классической методике. Кроме того, ресурсоемкость стохастического эксперимента существенно возрастает из-за необходимости заполнения некоторой области факторного пространства точками в соответствии с выбранным законом распределения. Основой для получения любых псевдослучайных последовательностей в компьютерном моделировании является генератор равномерно распределенных псевдослучайных чисел. Отдельный вопрос – определение длины последовательности псевдослучайных чисел, достаточной для обеспечения заданной точности результата. Из самых общих соображений можно сказать, что для того, чтобы погрешность моделирования в методе Монте-Карло не превышала некоторого $\boldsymbol{\varepsilon > 0}$, необходимо сгенерировать $\boldsymbol{m > 1/\varepsilon}$ псевдослучайных чисел. (Заметим, что в [1] приводится еще более жесткая оценка: $\boldsymbol{m > 1/\varepsilon^2}$.) То есть, для обеспечения хотя бы двух верных цифр в мантиссе, необходимо не менее 100 чисел в последовательности. В случае нескольких факторов случайные последовательности для каждого из них независимы. Следовательно, общее количество обращений к генератору случайных чисел должно превышать $\boldsymbol{1/\varepsilon^n}$. Например, для обеспечения двух верных цифр в мантиссе при пяти случайных факторах необходимо более 10^{10} обращений к генератору псевдослучайных чисел.

Очевидно, что резко снизить затраты вычислительных ресурсов при моделировании возможно только изменив саму методику проведения стохастического эксперимента. Одним из подходов к решению этой проблемы является применение в методе Монте-Карло так называемых «ЛПτ последовательностей», использование которых позволяет равномерно заполнить $\boldsymbol{n}$-мерный параллелепипед $\boldsymbol{2^k}$ точками, где $\boldsymbol{k}$ – необходимое количество верных знаков в двоичной мантиссе [1]. Следует особо отметить, что число точек в этом случае зависит только от необходимой точности решения и не зависит от количества учитываемых в модели факторов. Например, для обеспечения не менее двух верных знаков в десятичной мантиссе необходимо всего $2^7 = 128$ точек при любой размерности модели. Еще одним преимуществом этого подхода является ослабление требований к форме области $\boldsymbol{\Omega}$: достаточно, чтобы ее можно было заключить в параллелепипед. Последовательность генерируется внутри этого параллелепипеда, а затем точки, не принадлежащие $\boldsymbol{\Omega}$ **,** исключаются из дальнейшей обработки [1]. Боле того, результатам такого эксперимента является поверхность отклика системы $\boldsymbol{\Psi}$, которую можно подвергнуть дальнейшему анализу по любой классической методике, а не только линейному регрессионному анализу.

К сожалению, методика генерации ЛПτ последовательностей основана на применении «таблицы числителей», разработанной для устаревших ЭВМ М-20 [1, 2], которая ограничивает факторное пространство 51 измерением, а период последовательности – 2^{20}, что снижает максимально достижимую точность примерно до 10^{-4}. В рамках НИР и НИРС, проведенных в ИвГУ, удалось адаптировать данную таблицу к возможностям современной вычислительной техники и систем программирования. В частности, период последовательности удалось увеличить до 2^{64}, что обеспечивает максимально возможную для современных персональных компьютеров точность вычислений порядка 10^{-19}. Кроме того, был обнаружен и устранен ряд неточностей в исходной таблице, опубликованной в [1]. Также был проведен сравнительный вычислительный эксперимент по решению двух наиболее типичных для метода Монте-Карло задач – вычислению многомерного объема (кратного определенного интеграла) и поиску экстремума функции многих переменных (оптимального состава смеси). В обоих случаях генератор ЛПτ последовательностей оказался предпочтительнее традиционного «датчика псевдослучайных чисел», что позволило сделать вывод о целесообразности дальнейших исследований в этой области.

В настоящее время построена 128-битовая «таблица числителей» для компьютеров с процессором AMD64, что обеспечивает период последовательности равный 2^{128} и теоретическую точность порядка 10^{-38}, достижимую только при использовании формата вещественных чисел с разрядностью мантиссы не менее 128 бит. Разработка программ ведется на языке FORTRAN 2003 с использованием 64-битовой версии компилятора GCC версии 4.3.2 [4] в IDE Eclipse [5]. В таблице приведены первые двенадцать столбцов числителей в десятичной системе счисления и последний (128-й) – в шестнадцатеричной системе счисления. Эти данные позволяют построить ЛПτ последовательность с периодом 4096 в факторном пространстве с размерностью до 51, что вполне достаточно для планирования и обработки результатов «натурного» эксперимента (например, по подбору оптимального состава смеси). В имитационном компьютерном эксперименте должно использоваться $\boldsymbol{k > log_2 N}$ столбцов таблицы, где $\boldsymbol{N}$ – необходимое для моделирования число точек. Генерация последовательности по таблице числителей может быть выполнена программой, опубликованной в [1] или по алгоритму, описанному в [2]. В таблице также приведены в шестнадцатеричной системе счисления коды моноциклических операторов в поле ***GF(2)*** с операцией сложения по модулю 2 [1, 2], которые использованы при генерации строк таблицы. Первая цифра кода – порядок оператора, остальные, после перевода их в двоичную систему счисления, задают коэффициенты моноциклического полинома соответствующего порядка. Следует отметить, что такое сопоставление кодов полиномов и соответствующих им строк таблицы в [1, 2] не приводится.

Фрагмент 128-битовой таблицы числителей

№	Код	1	2	3	4	5	6	7	8	9	10	11	12	128(Hex)
0	0000	1	1	1	1	1	1	1	1	1	1	1	1	00000000000000000000000000000001
1	1003	1	3	5	15	17	51	85	255	257	771	1285	3855	FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF
2	2007	1	1	7	11	13	61	67	79	465	721	823	4091	4F00FFFF00000000FFFFFFFFFFFFFFFF
3	300D	1	3	7	5	7	43	49	147	439	1013	727	987	C30500000000C3050000000050050093
4	300B	1	1	5	3	15	51	125	141	177	759	267	1839	42F300000000BD0CFFFFFFFF300CFF8D
5	4019	1	3	1	1	9	59	25	89	321	835	833	4033	59000000999999999999999999999999
6	4013	1	1	3	7	31	47	109	173	181	949	471	2515	AD00000021210000DD00DD0011111111
7	5029	1	3	3	9	9	57	43	43	225	113	1601	579	9000038531A76318C6319F300900002B
8	5037	1	3	7	13	3	35	89	9	235	929	1341	3863	D000028700A46B5AD6B54C3C5851B509
9	503D	1	1	5	11	27	53	69	25	103	615	913	977	B000080FF7A4EF7BDEF7640A73D10019
10	502F	1	3	5	1	15	19	113	115	411	157	1725	3463	10000184213755110000D0321A332173
11	503B	1	1	7	3	29	51	47	97	233	39	2021	2909	30000CC1636100000000B5D0028C6361
12	5025	1	3	7	7	21	61	55	19	59	761	1905	3379	7000070000570D070000B0300A330013
13	6061	1	1	1	9	23	37	97	97	353	169	375	1349	40014575514500000000000035500061
14	6043	1	3	3	5	19	33	3	197	329	983	893	3739	C0090005090035C8C78971C7C875D795
15	606D	1	1	3	13	11	7	37	101	463	657	1599	347	40025049B2CB40025049B2CB09B09B65
16	605B	1	1	7	13	25	5	83	255	385	647	415	387	400EDF7DF7DF7DF7DF7DF7DF900EDF2F
17	6073	1	3	5	11	7	11	103	29	111	581	605	2881	C00700DD07008C4255F85555B95145AD
18	6067	1	1	1	3	13	39	27	203	475	505	819	2821	4002210000009C0351FB00007E4670FB
19	70C1	1	3	1	15	17	63	13	65	451	833	975	1873	8103000F000000000000000000003003
20	7083	1	1	5	5	1	27	33	195	263	139	915	1959	815C4976C00F0089EFCE7972CB97142D
21	7091	1	3	3	3	25	17	115	177	19	147	1715	1929	81000003000000B10000000000007103
22	7089	1	1	3	15	29	15	41	105	249	203	1223	2389	805076153468D157930B0000000072C1
23	70F1	1	3	1	7	3	23	79	17	275	81	1367	3251	818D8FFE271700110000000000000FD03
24	708F	1	3	7	9	31	29	17	47	369	337	663	1149	817CF9F9000000117CF9F3E7CF9FD17F
25	70B9	1	1	5	13	11	3	29	169	393	829	629	243	80715550A142858A657D000000002A41
26	709D	1	3	1	9	5	21	119	109	167	989	525	3609	81AD500AD038BB5D3EDBB7190000BBD3
27	70E5	1	1	3	1	23	13	75	149	333	375	469	1131	80ADA7031AE6F7ECC0D0056B000022B1
28	70A7	1	3	3	11	27	31	73	15	473	365	981	1701	8113910000000076FC4EBF7EFDFBA0AF
29	70D5	1	1	7	7	19	25	105	213	469	131	1667	143	81009005400900E93113078F0000EDF1
30	70AB	1	3	5	5	21	9	7	135	101	215	1587	1339	81C1EDD42733DD7E05D006FC6EDDEE07
31	70FD	1	1	1	15	5	49	59	253	21	733	1251	3497	80DC3A7E74E10054D32A72A90000E4D1
32	70BF	1	1	1	1	1	33	65	191	451	451	451	2499	81FD0006EFDFBF7EFDFBF7EFDFBF2EFD
33	70D3	1	3	5	15	17	19	21	155	229	447	481	1571	81F0393D07FEA323C0EA3BFBD1A303BF
34	70CB	1	1	7	11	13	29	3	175	247	177	721	983	81C71A1AD87100FCB5000D0F50A14CB5
35	70F7	1	3	7	5	7	11	113	63	297	57	483	4021	81FF75006AD5AB63BFDFB56AD5AB67EF
36	70EF	1	1	5	3	15	19	61	47	403	471	1209	1625	802F0DFFC2CF00907EFDFBF7EFDFDC7D
37	8171	1	3	1	1	9	27	89	7	497	979	1457	3217	85E7EEE78CF70000F7F7F7F7F7F7F7F7
38	811D	1	1	3	7	31	15	45	23	61	197	415	1163	0FDCD3A38D5384A34FDC5700A3A3A3A3
39	81A9	1	3	3	9	9	25	107	39	361	251	1435	2977	3DDB69106684688F9AC08F8F8F8F8F8F
40	812B	1	3	7	13	3	3	25	55	215	517	725	3391	B289734D00000000D50D1000CE34FDFD
41	8169	1	1	5	11	27	21	5	71	393	137	861	675	0304787F0304787FC6407F7F7F7F7F7F
42	812D	1	3	5	1	15	51	49	87	125	567	41	3093	14434343434343439700970043434343
43	8165	1	1	7	3	29	19	111	103	285	1021	1619	1495	DD07A6FB9777570050FC57008B8B8B8B
44	814D	1	3	7	7	21	29	119	119	501	167	1579	3443	794FA4300F5BAC6BD224B7006B6B6B6B
45	81F5	1	1	1	9	23	5	33	135	277	877	1701	557	56E7E5D03437170013D41700E3E3E3E3
46	815F	1	3	3	5	19	1	67	153	199	929	869	675	41630D00C90000000A2C0D00085C1313
47	818D	1	1	3	13	11	39	101	169	301	269	1151	1489	77D151101110D4EDDA3CE900EDEDEDED
48	8163	1	1	7	13	25	37	19	185	19	327	1897	2303	DE30123F563F3F3F39EC2D0037DC3F3F
49	81C3	1	3	5	11	7	43	39	201	83	997	1679	3925	4894C7C78194C7C75F4319108194C7C7
50	8187	1	1	1	3	13	7	91	217	351	91	1355	3705	61D4DD50F3F3F3F3FB9083F3EE14F3F3

Это косвенно свидетельствует о том, что кроме теоретического обоснования свойств ЛПτ последовательностей и алгоритмов их построения авторы руководствовались и другими критериями, которые не были опубликованы. Особенно острой является проблема выбора «направляющих чисел» моноциклических операторов, которые задают начальные точки последовательности. В частности, авторы при построении таблиц не следуют собственной рекомендации использовать в качестве направляющих чисел δ_{ij} [1,2]. Хотя этот факт никак не влияет на практическую ценность ЛПτ последовательностей для компьютерного имитационного моделирования, он серьезно затрудняет дальнейшие теоретические разработки в этом направлении. Особый интерес представляет разработка алгоритма генерации равномерно распределенных ЛПτ последовательностей точек в $\boldsymbol{R}^n$ для произвольного $\boldsymbol{n}$ и с периодом, ограниченным только разрядностью целочисленной арифметики компьютера. В [1] отмечается, что если ЛПτ последовательность построена в $\boldsymbol{R}^n$, то ее проекция в $\boldsymbol{R}^{n-1}$ тоже является ЛПτ последовательностью. Тем не менее, авторы [1, 2] не ставят перед собой задачу построения *n*-мерной пространственной структуры, из которой можно было бы получить ЛПτ последовательности меньшей размерности.

Исходя из самых общих соображений, можно сказать, что подобная структура должна обладать свойством самоподобия, то есть должна быть фракталом. Известны фракталы-дендриты, основанные на позиционных системах счисления [3]. Следует отметить, что единственный моноциклический оператор первого порядка (табл.) генерирует последовательность числителей, которую можно получить из широко известного «треугольника Паскаля» по модулю 2. В результате получается числовой фрактал, визуально подобный геометрическому фракталу, известному как «треугольник Серпинского» [3]. Таким образом, задачу генерации *n*-мерной равномерно распределенной последовательности можно свести к построению *n*-мерного числового фрактала с последующим вычислением его проекций.

Список литературы

1. *Соболь И. М., Статников Р. Б.* Выбор оптимальных параметров в задачах со многими критериями. М.: Наука, 1981. 108 с.
2. *Соболь И. М.* Многомерные квадратурные формулы и функции Хаара. М.: Наука, 1969. 288 с.
3. *Морозов А. Д.* Введение в теорию фракталов. М.: Институт компьютерных исследований, 2002. 160 с.
4. http://gcc.gnu.org/
5. http://www.eclipse.org/

Поступила в редакцию 7.11.2008 г.