

ПРИМЕНЕНИЕ LP_τ РАВНОМЕРНО РАСПРЕДЕЛЕННЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ ДЛЯ РЕШЕНИЯ ПРИКЛАДНЫХ ЗАДАЧ МОДЕЛИРОВАНИЯ

*Национальный технический университет Украины "Киевский политехнический институт", Киев, Украина

Анотація. Розглянуто застосування LP_τ рівномірно розподілених послідовностей для рішення прикладних задач моделювання. Сформульовані проблеми побудови багатofакторних планів експериментів на основі цих послідовностей. Показано, що застосування LP_τ послідовностей як планів експериментів без спеціального дослідження неможливо.

Ключові слова: LP_τ рівномірно розподілені послідовності, планування експерименту, кореляція.

Аннотация. Рассмотрено использование LP_τ равномерно распределенных последовательностей для решения прикладных задач моделирования. Сформулированы проблемы построения многофакторных планов экспериментов на основе этих последовательностей. Показано, что использование LP_τ последовательностей в качестве планов экспериментов без специального исследования не представляется возможным.

Ключевые слова: LP_τ равномерно распределенные последовательности, планирование эксперимента, корреляция.

Abstract. The article deals with the use of LP_τ uniformly distributed sequences designed to solve applied simulation problems. The problems of constructing of multifactor experiment designs on the base of these sequences are formulated as well. It is shown that the use of LP_τ sequences as experiment designs without special research is impossible.

Keywords: LP_τ uniformly distributed sequences, experiment designs, correlation.

1. Введение. Постановка проблемы

Научное исследование реальной действительности предполагает создание и использование математических моделей, которые формализованно описывают связь комплекса начальных условий с группой критериев качества изучаемого объекта, системы или процесса. Большинство научных исследований связано с экспериментом. Он проводится в лабораториях, на производстве, в клиниках и т.д. Планирование эксперимента входит составной частью в общее планирование исследования, представляющее один из этапов исследовательского процесса, предшествующий непосредственному проведению опытов. Суть его заключается в составлении набора экспериментальных условий с определенными комбинациями независимых и зависимых переменных.

Основой любого эксперимента является план эксперимента. Правильно выбранный план эксперимента позволяет получить многофакторные статистические модели с наилучшими возможными критериями качества.

Последовательное планирование многофакторных экспериментов с использованием планов экспериментов на основе LP_τ равномерно распределенных последовательностей – необходимое условие получения искомой структуры многофакторной статистической модели.

Последовательность точек P_1, \dots, P_i, \dots называется равномерно распределенной [3, с. 20] в n -мерном кубе K^n , если для любого параллелепипеда Π

$$\lim_{N \rightarrow \infty} \frac{S_N(\Pi)}{N} = V_{\Pi},$$

где $S_N(\Pi)$ – количество точек P_i с номерами $1 \leq i \leq N$, принадлежащих Π ;

V_{Π} – объем (n -мерный) параллелепипеда Π .

Последовательность точек $P_0, P_1, \dots, P_i, \dots, P_0$, n -мерного куба K^n называется ЛП $_{\tau}$ последовательностью, если любой ее двоичный участок, содержащий не менее чем $2^{\tau+1}$ точек, представляет собой Π_{τ} -сетку [3, с. 134–135]. Название «ЛП $_{\tau}$ последовательность» образовано как сокращение фразы «последовательность, любой двоичный участок которой представляет собой Π_{τ} -сетку».

Цель статьи

Обзор существующих областей использования ЛП $_{\tau}$ последовательностей; исследование свойств и применения ЛП $_{\tau}$ равномерно распределенных последовательностей при планировании многофакторных экспериментов для получения регрессионных моделей.

Анализ публикаций по теме исследования

ЛП $_{\tau}$ последовательности имеют широкую область применения: вычисление многомерных интегралов, случайный поиск, задачи многокритериальной оптимизации, моделирование физических и экономических процессов.

Разработанные И.М. Соболев ЛП $_{\tau}$ последовательности, предназначенные изначально для расчета многомерных интегралов, стали позже применяться и для реализации поисковых процедур. Планируемый ЛП-поиск – это метод рационального проектирования объектов искусственной природы. Он сконструирован на основе планирования ЛП $_{\tau}$ последовательностей и принадлежит семейству методов Монте-Карло [1]. Метод используется для анализа математических моделей функционирования проектируемых объектов. Структура указанных в [1, 2] ЛП $_{\tau}$ последовательностей позволяет строить сетки в k -мерном пространстве параметров исследуемых функций. Эти сетки позволяют определить:

- какие из варьируемых параметров с заданной вероятностью оказывают существенное влияние на значения функции – критерия качества;
- области концентрации наилучших значений критериев по заданной метрике между текущим значением критерия качества и его экстремальным значением;
- построить в многомерном пространстве критериев качества множество Парето.

Использование точек ЛП $_{\tau}$ последовательностей обеспечивает более высокую точность вычислений по некоторым алгоритмам Монте-Карло и более равномерный просмотр пространства параметров при решении задач многофакторной оптимизации для поиска экстремальных значений критериев качества. При решении некоторых задач методом Монте-Карло требуется, чтобы при генерации равномерного распределения обеспечивалось более равномерное покрытие области значений, частично жертвуя при этом «случайностью».

Метод исследования пространства параметров на основе ЛП $_{\tau}$ последовательностей изложен в работах [3, 4]. Он используется для постановки и решения прикладных задач

оптимизации со многими критериями качества. В основе метода лежит построение допустимого и Парето-оптимального множества решений. Приведены многочисленные результаты по теории метода. Рассматриваются такие важные классы задач: проектирование, идентификация, проектирование с управлением, поиск в базе данных. Описан программный комплекс MOVI, реализующий метод исследования пространства параметров. Комплекс позволяет находить решения многих реальных оптимизационных задач в параллельном режиме, что значительно экономит затраты машинного времени.

При экономическом моделировании для достижения высокой точности оценок производных функций необходимо провести большое количество испытаний, для этого необходимо много времени. ЛП_τ последовательности позволяют уменьшить дисперсию оценок и значительно сэкономить время проведения испытаний [5, с. 576]. Их свойства позволяют снизить величину стандартной ошибки оценки так, что она становится пропорциональна величине $1/M$, а не $1/\sqrt{M}$, где M – объем выборки [5, с. 576]. Целью их использования является извлечение репрезентативных величин для базовых переменных.

Генераторы ЛП_τ последовательностей активно используются в экономических исследованиях. В работе [6] анализируется влияние свойств равномерности A и A' [3] на характеристики генератора ЛП_τ последовательностей при решении многомерных задач. В статье [6] отмечается, что эти свойства обеспечивают добавочные гарантии равномерности для многомерных задач даже при малом количестве пробных точек. Наложение дополнительных свойств равномерности на маломерные проекции ЛП_τ последовательностей, вдобавок к свойствам равномерности d -мерных последовательностей, может увеличить эффективность ЛП_τ последовательностей. Генератор ЛП_τ последовательностей SobolSeq16384 удовлетворяет дополнительные свойства равномерности.

Более широкое описание различных методов применения ЛП_τ последовательностей в экономическом моделировании приведено в работе [7].

Проблема применения квазислучайных последовательностей в имитационном моделировании рассматривается в статье [8]. Рассмотрены некоторые статистические свойства ЛП_τ последовательностей. Показано отличие между квазислучайными и псевдослучайными последовательностями. Псевдослучайные последовательности – это числа из интервала $(0,1)$, полученные с помощью некоторого детерминированного алгоритма, но имеют все свойства случайной последовательности чисел в рамках решаемой задачи [8]. Квазислучайной называют последовательность n -мерных точек P_1, P_2, \dots в единичном кубе, отклонения которых [1, с. 152]

$$D(P_1, \dots, P_N) \leq C(n)(\ln N)^n.$$

Квазислучайные последовательности, в отличие от псевдослучайных, это равномерно распределенные последовательности, элементы которых не обладают свойством независимости [9].

В работе [9] указывается о возможности построения планов экспериментов на основе ЛП_τ последовательностей. В [10] изложено оптимальное планирование эксперимента в системе «план эксперимента – структура модели». Приведены статистические свойства планов на основе ЛП_τ равномерно распределенных последовательностей. Предложены рекомендации по использованию планов экспериментов.

Детальное описание способов построения, теоретических сведений и свойств ЛП_τ последовательностей можно найти в работах [3, 6, 11–15]. В работе [13] представлены направляющие числа (двоично-рациональные дроби в двоичной системе [1, с. 139], позво-

ляющие генерировать LPP_τ последовательности для аппроксимирования интегралов размерностью до 1111. Данные направляющие числа генерируют LPP_τ последовательности, удовлетворяющие свойству А. Однако двумерные проекции этих последовательностей могут иметь неудовлетворительные характеристики равномерности [14]. Приведенные в работе [14] LPP_τ последовательности позволяют решить эту проблему. Полученная размерность – 21201. Компания BRODA [15], занимающаяся разработкой и распространением многомерных LDS генераторов, предлагает разработанный проф. И. Соболев SobolSeq370 генератор, который генерирует последовательности для размерности – 370 [15]. Недавно компания BRODA разработала SobolSeq32000 генератор, обладающий лучшими характеристиками и эффективностью, чем предыдущий, с размерностью LPP_τ последовательностей – 32000 [15].

Нерешенные вопросы

Разработка качественных математических моделей без планирования эксперимента невозможна. Поэтому актуальным являются постановка и использование многофакторного эксперимента на основе LPP_τ равномерно распределенных последовательностей. Имеются единичные упоминания о применении их для получения математических моделей. В большинстве статей не приводятся конкретные планы экспериментов. Не проведены статистические исследования LPP_τ последовательностей для выявления коррелированности между ними и последующего их ранжирования. Использование LPP_τ последовательностей при последовательном планировании экспериментов не разработано. Статистические свойства планов, использующих квазислучайное размещение точек в многофакторном пространстве (по известным публикациям), разработаны слабо.

2. Статистические свойства LPP_τ последовательностей

Планы экспериментов должны соответствовать различным критериям качества. Критерии, позволяющие выбрать структуру математической модели, практически не используются. При выборе структуры математической модели главные эффекты и эффекты взаимодействия должны быть ортогональными или слабо коррелированными.

Величину коррелированности последовательностей будем характеризовать парным коэффициентом корреляции:

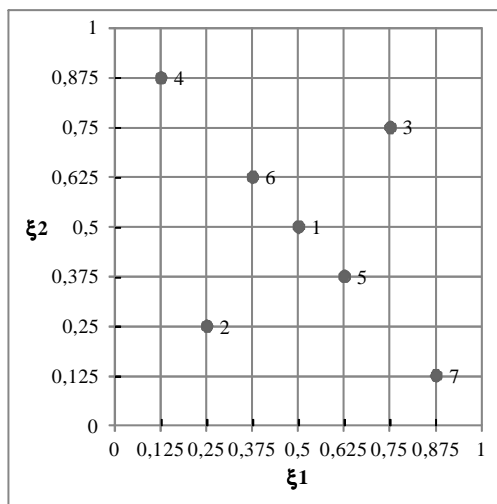
$$r_{ij}(\xi_i, \xi_j) = \frac{\sum_{u=1}^N (\xi_{iu} - \bar{\xi}_i)(\xi_{ju} - \bar{\xi}_j)}{\sqrt{\sum_{u=1}^N (\xi_{iu} - \bar{\xi}_i)^2 \sum_{u=1}^N (\xi_{ju} - \bar{\xi}_j)^2}},$$

где ξ_{iu} , ξ_{ju} – значение i -той, j -той последовательности в u -той строке таблицы последовательностей, $1 \leq i < j \leq k$, $1 \leq u \leq N$, k – общее число последовательностей, соответствующее размерности исследуемого пространства, N – общее число строк в таблице последовательностей:

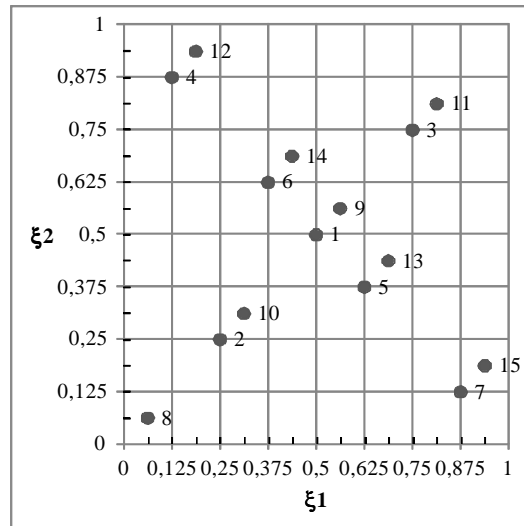
$$\bar{\xi}_i = \sum_{u=1}^N \xi_{iu} / N,$$

$$\bar{\xi}_j = \sum_{u=1}^N \xi_{ju} / N.$$

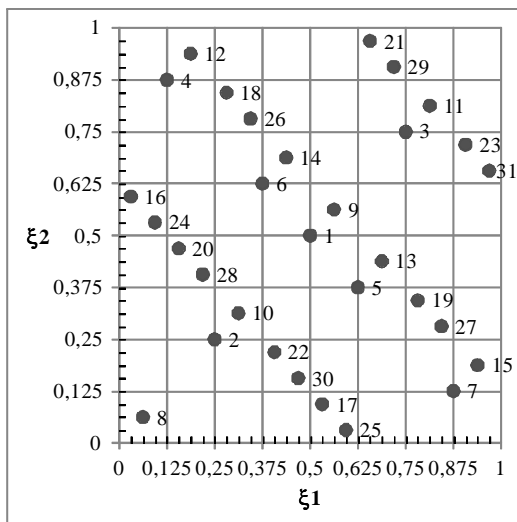
Это требование достигается путем равномерного распределения точек плана эксперимента в многофакторном пространстве. ЛП_τ последовательности в настоящее время являются наиболее равномерно распределенными последовательностями (рис. 1). Исследования показали, что последовательности точек распределены равномерно в диапазонах: $N = 1 \dots 7$; $N = 1 \dots 15$; $N = 1 \dots 31$; $N = 1 \dots 63$ и т.д.



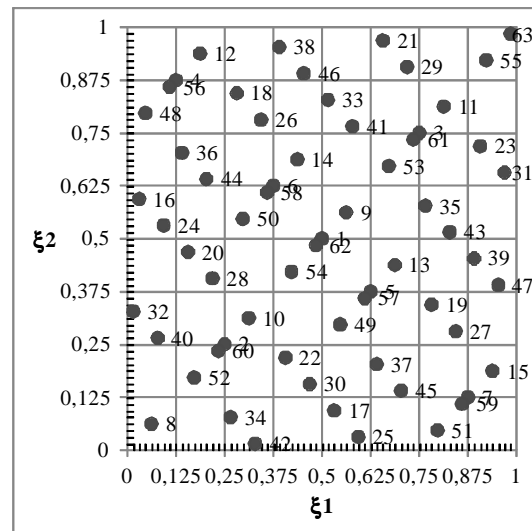
а) опытов – 7



б) опытов – 15



в) опытов – 31



г) опытов – 63

Рис. 1. Распределение точек ЛП_τ последовательностей ξ_1 и ξ_2 в двумерном пространстве

ЛП_τ последовательности И. М. Соболя обладают лучшими свойствами равномерности распределения, чем любые другие последовательности точек в многомерном единичном кубе. Свойства ЛП_τ последовательностей [1, 10]:

1. Проекция любой ЛП_τ последовательности из N точек в k -мерном пространстве на $(k - j)$ -мерную грань ($1 \leq j \leq k - 1$) многомерного единичного куба образует также равномерно распределенную последовательность из N проекций точек.

2. Выбор в качестве точек плана эксперимента в многомерном пространстве ЛП_τ равномерно распределенных последовательностей позволяет получить сравнительно слабо коррелированные главные эффекты и эффекты взаимодействий факторов ($|r_{ij}(\xi_i, \xi_j)| < 0,4$) при выборе структуры математической модели.

3. Планы экспериментов на основе ЛП_τ равномерно распределенных последовательностей позволяют, по сравнению с регулярными планами, получить расположение точек, более близкое к экстремальным значениям поверхности отклика [10, с. 104]. Число уровней s_i для непрерывных факторов равно числу опытов $N_{\text{ЛП}\tau}$, что позволяет более точно определить структуру получаемой статистической модели.

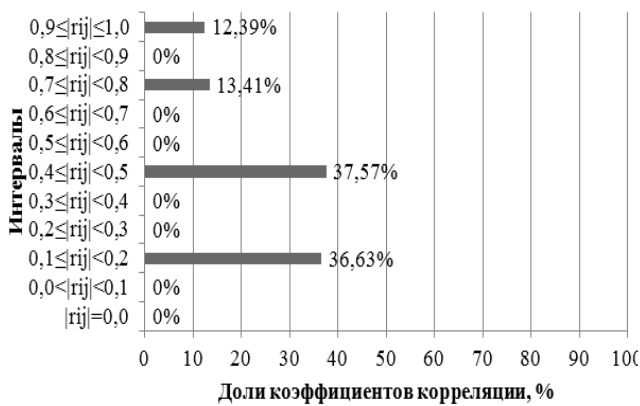
4. С увеличением числа опытов вероятность получения в плане эксперимента точек, достаточно близких к точкам экстремума и перегиба поверхности отклика, стремится к единице, а коэффициент корреляции r_{ij} между различными эффектами стремится к нулю [10, с. 104].

Математическое моделирование сложных систем, как правило, осуществляется на основе многофакторных статистических моделей (уравнений регрессии), полученных путем аппроксимации данных экспериментов, статистических испытаний, экспертных оценок, трудоемких вычислений на ЭВМ. Математические модели особенно необходимы в тех случаях, когда возможности конструирования, производства и эксплуатации, основанные на традиционных физических принципах, исчерпаны или приводят к нецелесообразно большим затратам. Планирование эксперимента на основе ЛП_τ равномерно распределенных последовательностей позволяет найти неизвестную структуру математической модели полиномиального вида, практически произвольной сложности, а также решить задачу многокритериальной оптимизации.

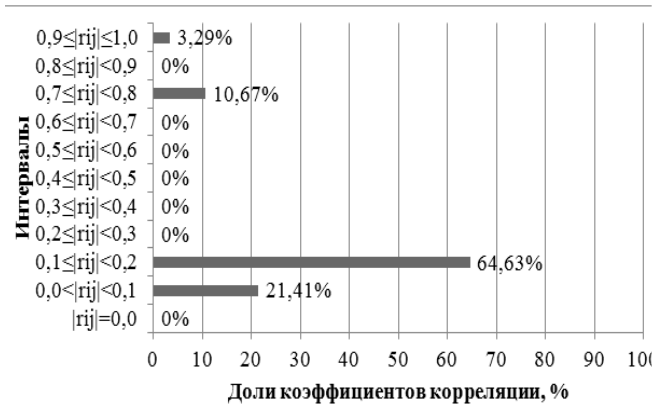
Коррелированность факторов весьма типична во множественном регрессионном анализе. При коррелированности факторов главные эффекты и взаимодействия факторов также коррелированы между собой. В таких условиях коэффициенты математической модели полиномиального вида определяются со значительными погрешностями и становятся смещенными. Найденная структура математической модели является неустойчивой, выделение истинных эффектов становится трудным, решение таких задач неустойчиво. С ростом коррелированности эффектов найденная модель теряет свою прикладную пригодность.

Исследование коррелированности последовательностей проводилось для 51 равномерно распределенной последовательности для диапазонов точек с равномерным распределением: $N = 1 \dots 7$; $N = 1 \dots 15$; $N = 1 \dots 31$; $N = 1 \dots 63$. Была выявлена значительная коррелированность некоторых последовательностей (рис. 2). Например, для семи опытов коэффициенты парной корреляции (по модулю) последовательностей ξ_1 и ξ_6 , ξ_2 и ξ_{20} равны $|r_{ij}| = 1$. Эта коррелированность убывает с ростом количества опытов.

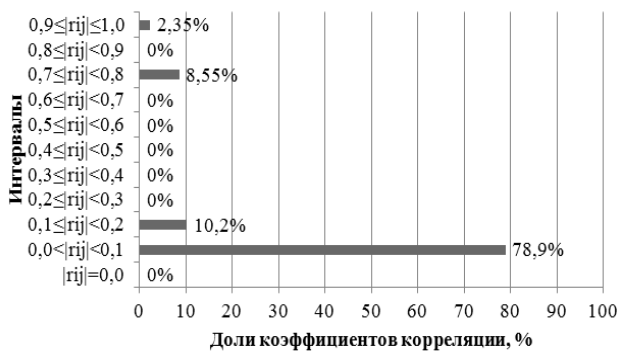
Поэтому применение ЛП_τ равномерно распределенных последовательностей в качестве планов экспериментов без специального исследования не представляется возможным. Необходимо найти последовательности, близкие к ортогональным друг относительно друга.



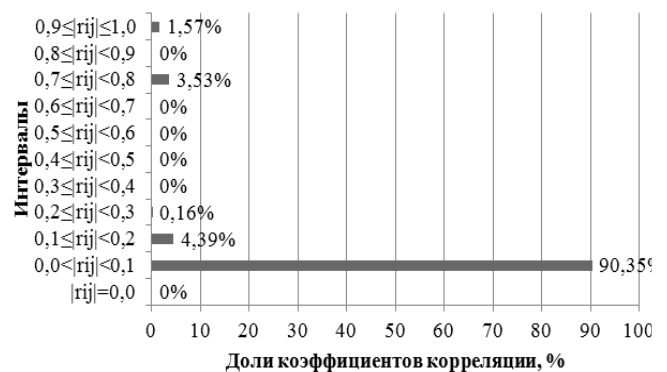
Среднее абсолютных величин коэффициентов корреляции 0,4331. Среднее квадратичное отклонение 0,2852. Опытов – 7



Среднее абсолютных величин коэффициентов корреляции 0,2350. Среднее квадратичное отклонение 0,2371. Опытов – 15



Среднее абсолютных величин коэффициентов корреляции 0,1623. Среднее квадратичное отклонение 0,2271. Опытов – 31



Среднее абсолютных величин коэффициентов корреляции 0,0844. Среднее квадратичное отклонение 0,1755. Опытов – 63

Рис. 2. Распределение коэффициентов корреляции 51 последовательности

С разработанными методами решения регрессионных задач и полученными результатами можно ознакомиться в [16].

3. Выводы

Проведенное исследование показало, что абсолютно не коррелированных последовательностей, на основе ЛП_τ последовательностей, нет. Однако присутствует значительное количество слабо коррелированных последовательностей. Поэтому построение планов экспериментов может проводиться только после определения всех слабо коррелированных последовательностей и их ранжирования. Построенные таким образом планы экспериментов будут обладать наименьшей коррелированностью эффектов, а полученные на их основе математические модели – соответствовать реальной действительности.

СПИСОК ЛИТЕРАТУРЫ

1. Соболев И.М. Многомерные квадратурные формулы и функции Хаара / Соболев И.М. – М.: Физматлит, 1969. – 288 с.
2. Соболев И.М. ЛП-поиск и задачи оптимального проектирования / И.М. Соболев, Р.Б. Статников // Проблемы случайного поиска: сб. статей. – Рига: Зинатне, 1972. – С. 117 – 135.
3. Соболев И.М. Выбор оптимальных параметров в задачах со многими критериями / И.М. Соболев, Р.Б. Статников. – [2-е изд., перераб. и доп.]. – М.: Дрофа, 2006. – 175 с.

4. Statnikov R.B. The Parameter Space Investigation Method Toolkit [with DVD] / R.B. Statnikov, A. Statnikov. – Boston/London: Artech House Publishers, 2011. – 214 p.
5. Халл К.Д. Опционы, фьючерсы и другие производные финансовые инструменты / К.Д. Халл. – [6-е изд.]; пер. с англ. – М.: ООО «И.Д. Вильямс», 2007. – 1056 с.
6. Construction and Comparison of High-Dimensional Sobol' Generators / I.M. Sobol', D. Asotsky, A. Kreinin [et al.] // Wilmott Journal. – 2012. – Vol. 2011, Is.56. – P. 64 – 79.
7. Brotherton-Ratcliffe R. Monte Carlo motoring / R. Brotherton-Ratcliffe // Risk. – 1994. – Vol. 7, N 12. – P. 53 – 57.
8. Ermakov S. On the Quasi-Random Sequence in the Random Processes Modeling Algorithms // S. Ermakov, T. Tovstik // Focus on Applied Statistics. Nova Science Publishers. – 2003. – P. 91 – 102.
9. Орлов В.А. Новое семейство квазислучайных последовательностей / В.А. Орлов, В.И. Рейзлин // Известия Томского политехнического университета. – 2012. – Т. 320, № 2. – С. 24 – 26.
10. Радченко С.Г. Методология регрессионного анализа / Радченко С.Г. – К.: Корнійчук, 2011. – 376 с.
11. Numerical recipes in C: The art of Scientific computing / W.H. Press, S.A. Teukolsky, W.T. Vetterling [et al.]. – [2nd ed.]. – Cambridge University Press, 1992. – 1018 p.
12. Антонов Я.А. Экономичный способ вычисления LTP_7 последовательностей / Я.А. Антонов, В.М. Салеев // Журн. вычисл. мат. и мат. физ. – 1979. – Т. 19, № 1. – С. 243 – 245.
13. Joe S. Remark on Algorithm 659: Implementing Sobol's quasirandom sequence generator / S. Joe // ACM Trans. Math. Softw. – 2003. – Vol. 29. – P. 49 – 57.
14. Joe S. Constructing Sobol sequences with better two-dimensional projections / S. Joe, F.Y. Kuo // SIAM J. Sci. Comput. – 2008. – Vol. 30. – P. 2635 – 2654.
15. Broda [Электронный ресурс]. – Режим доступа: <http://www.broda.co.uk>.
16. Лаборатория экспериментально-статистических методов исследований (ЛЭСМИ) [Электронный ресурс]. – Режим доступа: <http://www.n-t.org/sp/lesmi>.

Стаття надійшла до редакції 27.09.2013