

ЗАКОН РАСПРЕДЕЛЕНИЯ ДИСКРЕТНОЙ СЛУЧАЙНОЙ ВЕЛИЧИНЫ НА ВЫХОДЕ КОМБИНАЦИОННОГО ГЕНЕРАТОРА

Эмиль Фауре

Черкасский государственный технологический университет, Украина



ФАУРЕ Эмиль Витальевич, к.т.н.

Год и место рождения: 1983 год, г. Черкассы, Украина.

Образование: Черкасский государственный технологический университет, 2005 год.

Должность: доцент кафедры компьютерных систем.

Научные интересы: модели, методы и средства формирования псевдослучайных последовательностей чисел; кодовые и некодовые методы повышения достоверности передаваемых данных.

Публикации: более 40 научных публикаций, учебно-методические работы.

E-mail: faureemil@gmail.com

Аннотация. В статье рассматриваются статистические свойства дискретной случайной величины на выходе комбинационного генератора, выполняющего операцию суммирования по некоторому модулю слов, полученных от двух первичных генераторов равномерно распределенных случайных чисел. Определен закон распределения дискретной случайной величины на выходе комбинационного генератора. Определены условия, при которых этот закон распределения является строго равномерным. В качестве исходных первичных последовательностей случайных чисел рассмотрены последовательности истинно случайных чисел как с неограниченными, так и с ограниченными периодами, а также последовательности, представляющие собой циклически повторяющиеся подстановки. Полученные результаты позволяют расширить теоретическую базу проектирования комбинационных генераторов случайных чисел и создают основу для дальнейшего анализа, разработки и практической реализации подобного рода генераторов.

Ключевые слова: дискретная случайная величина, последовательность случайных чисел, комбинационный генератор.

Введение

Задача формирования случайных (псевдослучайных) последовательностей чисел является одной из наиболее сложных и актуальных задач из области моделирования сложных процессов, систем и объектов, а также из области криптографической защиты информации.

Сформированная последовательность должна обладать необходимыми свойствами, определяющимися особенностями ее применения. Такими свойствами могут, например, являться: равномерный закон распределения дискретной случайной величины, отсутствие корреляции слов (чисел) последовательности, максимально возможный период повторения, непредсказуемость, простота реализации и воспроизводимость последовательности и т.п.

Наиболее широко используются последовательности случайных чисел с равномерным законом распределения, в том числе и по той причине, что на их основе формируются случайные последовательности с другими законами распределения. Последнее и определяет актуальность разработки и исследования новых высокоэффективных методов и технических решений формирования последовательностей равномерно распределенных случайных чисел.

Постановка проблемы

Недостатком многих существующих методов формирования последовательностей равномерно распределенных псевдослучайных чисел (например, последовательностей, образованных регистром сдвига с линейными обратными связями – РСЛОС (LFSR), линейным конгруэнтным генератором – ЛКГ (LCG), вихрем Мерсенна (Mersenne twister, MT) и пр.) является малая линейная сложность и, как следствие, простота вскрытия закона образования последовательности по ее отрезку. Кроме того, ряд методов (РСЛОС, ЛКГ и др.) имеют слишком малый период повторения последовательности и не позволяют получить равномерное распределение дискретной случайной величины на множестве значений, включающем значение нуля. В таком случае следует прибегать к операции рандомизации, что усложняет процедуру формирования случайных чисел. Ограниченный период также может являться недостатком в случае использования предварительно сформированной и записанной последовательности истинно случайных чисел.

Для устранения указанных недостатков в ряде известных работ ([1-3] и др.), подробно освещенных и систематизированных в [4], предлагается использование комбинации нескольких случайных процессов.

Генератор, содержащий в своем составе несколько первичных автономных генераторов (псевдо) случайных чисел и обеспечивающий их комбинирование, будем называть комбинационным генератором.

В работе [5] уделяется внимание определению длины периода комбинационного генератора, комбинирующей функцией которого является операция суммирования по некоторому модулю M двух последовательностей целых чисел, распределенных на отрезке $[0, M - 1]$. Кроме того, в этой работе представлен анализ применения t -мерного критерия серий к сформированной рассматриваемым комбинационным генератором последовательности.

Несмотря на то, что некоторые вопросы, касающиеся исследования комбинационных генераторов, находят решения в указанных работах, многие задачи все еще остаются нерешенными.

Целью настоящего исследования является определение закона распределения дискретной случайной величины на выходе комбинационного генератора, комбинирующей функцией которого является операция суммирования по некоторому модулю M , а также определение условий, при которых этот закон распределения будет строго равномерным. Ошибка воспроизведения закона распределения дискретной случайной величины определяется по методике, предложенной в работе [6], как число символов ошибочного потока, приходящееся на единицу объема выборки.

Постановка задачи

В общем случае количество (первичных) генераторов случайных чисел в комбинационном генераторе равняется n . В настоящей работе рассматривается случай $n=2$. Таким образом, имеется два независимых первичных генератора равномерно распределенных случайных величин X и Y с мощностями алфавита M_x и M_y . Области определения этих дискретных случайных величин – $X \in [0, (M_x - 1)]$ и $Y \in [0, (M_y - 1)]$. Так как комбинирующей функцией генератора является суммирование по некоторому модулю M , то рассматриваемый комбинационный генератор выполняет операцию $Z = |X + Y|_M$ (здесь $|A|_B$ обозначает вычет (остаток) числа A по модулю B).

Задача исследования сводится к нахождению вероятностей $P(Z = z_i)$ для $\forall z_i \in [0, M - 1]$, а также определению условий, при которых полученный закон распределения дискретной случайной величины Z будет строго равномерным, т.е. $P(Z = z_i) = \frac{1}{M}$ для $\forall z_i \in [0, M - 1]$.

Решение задачи

Рассмотрим сначала функцию $Z' = X + Y$ с множеством значений $Z' \in [0, M_x + M_y - 2]$.

Определим функцию распределения $F(z')$ и закон распределения $f(z')$ случайной величины Z' .

Для этого найдем функцию распределения $G(x, y)$ и закон распределения $g(x, y)$ системы случайных величин (X, Y) . Воспользуемся рис. 1.

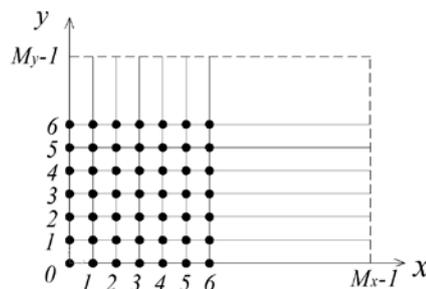


Рис. 1 – Размещение системы случайных величин X и Y на числовой оси

Отметим, что функцией распределения системы двух случайных величин (X, Y) называется функция двух аргументов $G(x, y)$, равная вероятности совместного выполнения двух неравенств $X < x, Y < y$.

В силу равномерного распределения дискретных случайных величин X и Y , а также их независимости $G(0,0) = \frac{0}{M_x} \cdot \frac{0}{M_y} = 0$, $G(0,1) = \frac{0}{M_x} \cdot \frac{1}{M_y} = 0$,

$$G(1,0) = \frac{1}{M_x} \cdot \frac{0}{M_y} = 0, \quad G(1,1) = \frac{1}{M_x} \cdot \frac{1}{M_y} = \frac{1}{M_x M_y},$$

$$G(1,2) = \frac{1}{M_x} \cdot \frac{2}{M_y} = \frac{2}{M_x M_y} \text{ и т.д. В общем виде}$$

$$G(x, y) = P(X < x, Y < y) = \frac{xy}{M_x M_y}, \quad x \in [0, M_x], \quad y \in [0, M_y].$$

Закон распределения системы дискретных случайных величин (X, Y)

$$g(x, y) = G(x+1, y+1) - G(x+1, y) - G(x, y+1) + G(x, y) = \frac{(x+1)(y+1)}{M_x M_y} - \frac{(x+1)y}{M_x M_y} - \frac{x(y+1)}{M_x M_y} + \frac{xy}{M_x M_y}$$

или

$$g(x, y) = \frac{1}{M_x M_y}. \quad (1)$$

Закон распределения системы дискретных случайных величин (X, Y) $g(x, y) = \frac{1}{M_x M_y}$ означает, что вероятность события, соответствующего любой из точек на рис. 1, одинакова и равна $\frac{1}{M_x M_y}$.

Для нахождения закона распределения дискретной случайной величины $Z' = X + Y$ построим на плоскости xOy прямую $z' = x + y$ (рис. 2).

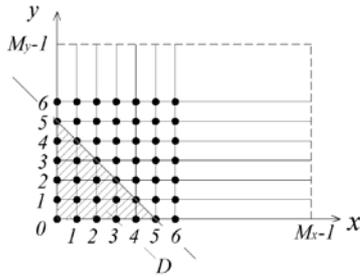


Рис. 2 - Область D

Обозначим D область, для которой высота поверхности $z' = x + y$ над плоскостью xOy меньше z' . Чтобы выполнялось неравенство $F(z') = P(Z' < z')$, случайная точка (x, y) должна попасть в область D . Следовательно, функция распределения величины $Z' = X + Y$

$$F(z') = P((X', Y') \subset D) = \sum_{(x', y') \in D} g(x, y) \text{ или}$$

$$F(z') = \begin{cases} \frac{1}{M_x M_y} \left(\frac{z'(z'+1)}{2} \right) & \text{при } z' \in [0, \min(M_x, M_y)]; \\ F(\min(M_x, M_y)) + \frac{1}{M_x M_y} (z' - \min(M_x, M_y)) \cdot \min(M_x, M_y) & \text{при } z' \in [\min(M_x, M_y), \max(M_x, M_y)]; \\ 1 - \frac{1}{M_x M_y} \cdot \frac{(M_x + M_y - z' - 1)(M_x + M_y - z')}{2} & \text{при } z' \in [\max(M_x, M_y), M_x + M_y - 1]. \end{cases}$$

Закон распределения величины $Z' = X + Y$ $P(Z' = z'_i) = F(z'_i + 1) - F(z'_i)$ или

$$P(Z' = z'_i) = \begin{cases} \frac{z'+1}{M_x M_y} & \text{при } z' \in [0, \min(M_x, M_y) - 1]; \\ \frac{1}{\max(M_x, M_y)} & \text{при } z' \in [\min(M_x, M_y) - 1, \max(M_x, M_y) - 1]; \\ \frac{M_x + M_y - z' - 1}{M_x M_y} & \text{при } z' \in [\max(M_x, M_y) - 1, M_x + M_y - 2]. \end{cases} \quad (2)$$

Графическое представление закона распределения случайной величины $Z' = X + Y$ показано на рис. 3.

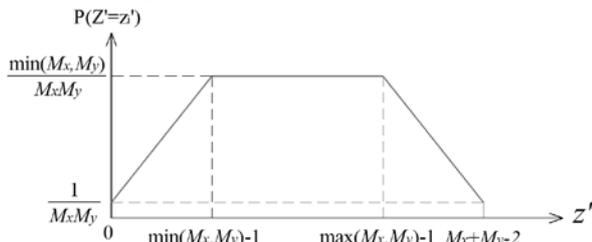


Рис. 3 - Закон распределения случайной величины $Z' = X + Y$

При равенстве M_x и M_y закон распределения случайной величины $Z' = X + Y$ вырождается в закон Симпсона, а график закона распределения, соответственно, - в треугольник Симпсона.

Рассмотрим теперь функцию $Z = |X + Y|_M = |Z'|_M$ с множеством значений $Z \in [0, M - 1]$. Закон распределения функции Z имеет вид:

$$P(Z = z_i) = \sum_{k=0}^{K_i} P(Z' = kM + z_i),$$

где K_i - максимальное число, при котором $K_i M + z_i \leq M_x + M_y - 2$.

Таким образом,

$$P(Z = z_i) = \sum_{k=0}^{\lfloor \frac{M_x + M_y - 2 - z_i}{M} \rfloor} P(Z' = kM + z_i),$$

где $\lfloor A \rfloor$ обозначена целая часть (функция «пол») от числа A .

Определим условия, при которых $P(Z = z_i) = \frac{1}{M}$ для $\forall z_i \in [0, M - 1]$:

1) очевидно, что при $M > \max(M_x, M_y)$ закон распределения случайной величины Z не является равномерным;

2) при $M = \max(M_x, M_y)$ для всех $z_i \in [0, \min(M_x, M_y) - 1]$,

$$P(Z = z_i) = \frac{z_i + 1}{M_x M_y} + \frac{M_x + M_y - \max(M_x, M_y) - z_i - 1}{M_x M_y} = \frac{1}{\max(M_x, M_y)}$$

и закон распределения случайной величины Z является равномерным с $P(Z = z_i) = \frac{1}{M}$ для всех

$z_i \in [0, M - 1]$;

3) при $M = \frac{\max(M_x, M_y)}{n}$ закон распределения случайной величины Z также является равномерным

$$\text{с } P(Z = z_i) = \frac{n}{\max(M_x, M_y)} = \frac{1}{M} \text{ для всех } z_i \in [0, M - 1];$$

4) для $M = \min(M_x, M_y)$ воспользуемся рис. 4.

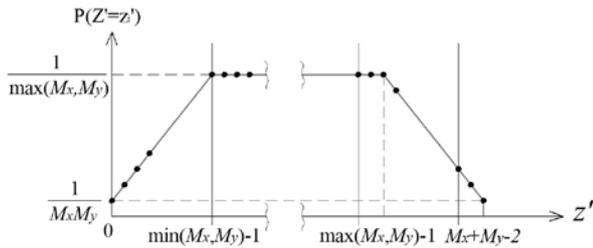


Рис. 4 - Закон распределения случайной величины $Z' = X + Y$ для $M = \min(M_x, M_y)$

При $m = \left\lfloor \frac{\max(M_x, M_y) - \min(M_x, M_y)}{M} \right\rfloor$ для всех

$$z_i \in [0, \lfloor M_x + M_y - 2 \rfloor_M + 1]$$

$$P(Z = z_i) = \frac{z_i + 1}{M_x M_y} + m \frac{1}{\max(M_x, M_y)} + \frac{M_x + M_y - (m+1)M - z_i - 1}{M_x M_y} = \frac{1}{\min(M_x, M_y)}$$

Для всех $z_i \in [\lfloor M_x + M_y - 2 \rfloor_M + 2, M - 1]$

$$P(Z = z_i) = \frac{z_i + 1}{M_x M_y} + (m-1) \frac{1}{\max(M_x, M_y)} + \frac{M_x + M_y - mM - z_i - 1}{M_x M_y} = \frac{1}{\min(M_x, M_y)}$$

Таким образом, при $M = \min(M_x, M_y)$ для всех

$$z_i \in [0, M - 1] \quad P(Z = z_i) = \frac{1}{\min(M_x, M_y)} = \frac{1}{M}$$

и закон распределения случайной величины Z является равномерным;

5) при $M = \frac{\min(M_x, M_y)}{n}$ закон распределения

случайной величины Z также является равномерным с $P(Z = z_i) = \frac{n}{\min(M_x, M_y)} = \frac{1}{M}$ для всех $z_i \in [0, M - 1]$.

Таким образом, для дискретной случайной величины $Z = |X + Y|_M$ можно сформулировать следующее утверждение.

Утверждение 1. Для равномерного распределения на множестве целых чисел мощности M дискретной случайной величины, полученной в результате суммирования по модулю M двух независимых первичных случайных величин, равномерно распределенных на множествах целых чисел из диапазонов $[0, M_x - 1]$ и $[0, M_y - 1]$, достаточно, чтобы хотя бы одно из значений M_x или M_y было кратно M .

Заметим, что доказанное утверждение справедливо для истинно случайных исходных величин, являющихся реализациями естественного (природного) «белого» шума, с бесконечными периодами повторения. Если же первичные последовательности равномерно распределенных случайных чисел X и Y периодичны с периодами T_x и T_y , то период повторения последовательности

случайных чисел на выходе комбинационного генератора зависит от периодов повторения первичных последовательностей случайных чисел и равен

$$T = \text{НОК}(T_x, T_y).$$

В таком случае, достаточным условием справедливости утверждения 1 для первичных случайных величин X и Y с периодами T_x и T_y

$(|T_x|_{M_x} = 0, |T_y|_{M_y} = 0)$ является условие взаимной

простоты периодов T_x и T_y ($\text{НОД}(T_x, T_y) = 1$). Это

объясняется тем, что при таком условии события, соответствующие любой из точек на рис. 1, встречаются одинаковое количество

$\left(n = \frac{\text{НОК}(T_x, T_y)}{M_x \cdot M_y} \in Z \right)$ раз на всем периоде

последовательности и, соответственно, закон распределения системы дискретных случайных

величин (X, Y) $g(x, y) = \frac{1}{M_x M_y}$.

Если же $|T_x|_{M_x} \neq 0, |T_y|_{M_y} \neq 0$ или $\text{НОД}(T_x, T_y) \neq 1$

то требуются дополнительные исследования с учетом принципов формирования первичных последовательностей случайных чисел X и Y .

В частном случае в качестве первичных генераторов могут использоваться генераторы подстановок, каждый из которых циклически формирует некоторую подстановку, или таблицы подстановок, реализованные с помощью циклических сдвиговых регистров, содержащих предварительно сформированные различные последовательности подстановок. В процессе работы комбинационного генератора исходные значения, подлежащие композиции, формируются циклически генераторами подстановок или считываются с выходов циклических сдвиговых регистров. Отсюда $T_x = M_x$, а $T_y = M_y$.

В таком случае закон распределения системы дискретных случайных величин (X, Y)

$g(x, y) = \frac{1}{M_x M_y}$, а закон распределения системы

дискретной случайной величины $Z' = X + Y$ определяется выражением, определенным в (2), при условии, что M_x и M_y взаимно просты. Это объясняется тем, что условие взаимной простоты значений M_x и M_y является необходимым и достаточным для того, чтобы период формируемой композиции был максимальным и равным $T_{\max} = M_x \cdot M_y$, а события, соответствующие любой из точек на рис. 1, встречались ровно по одному разу на всем периоде последовательности. В иных случаях период последовательности будет меньшим

$T_{\max} = M_x \cdot M_y$ и, как следствие, $g(x, y) \neq \frac{1}{M_x M_y}$.

Если же $\text{НОД}(T_x, T_y) \neq 1$, то требуются дополнительные исследования с учетом принципов

формирования первичных последовательностей подстановок X и Y . В любом случае период $T = НОК(T_x, T_y) = НОК(M_x, M_y)$ должен быть кратен M .

В силу сказанного, сформулируем следующее утверждение.

Утверждение 2. Для равномерного распределения дискретной случайной величины на множестве целых чисел мощности M на выходе комбинационного генератора, выполняющего операцию суммирования по модулю M слов от двух первичных генераторов, каждый из которых циклически формирует некоторую подстановку на множествах целых чисел из диапазонов $[0, M_x - 1]$ и $[0, M_y - 1]$ для первого и второго генератора, соответственно, достаточно, чтобы M_x и M_y были взаимно просты и одно из значений M_x или M_y было кратно M .

Экспериментальные исследования показывают, что если M_x и M_y взаимно просты, то для равномерного распределения дискретной случайной величины на множестве целых чисел мощности M на выходе комбинационного генератора условие кратности значению M одного из значений M_x или M_y является обязательным (случаев равномерного распределения дискретной случайной величины на выходе генератора, когда $НОД(M_x, M_y) = 1$, $|M_x|_M \neq 0$, $|M_y|_M \neq 0$, а $|НОК(M_x, M_y)|_M = 0$ (например, $M_x = 9$, $M_y = 16$, $M = 6$), не наблюдалось).

В общем случае, когда не требуется строгого соответствия закона распределения случайной величины равномерному закону распределения (не требуется строго одинакового количества всех значений случайной величины Z из области ее определения на всем периоде композиции), однако требуется выполнение статистической гипотезы о ее равномерном распределении на всем периоде генерируемой последовательности, можно, например, воспользоваться критерием Пирсона. Для этого следует вычислить значение

$$\chi^2 = V \sum_{z_i=0}^{M-1} \frac{(P(Z = z_i) - p_0(z))^2}{p_0(z)} =$$

$$= V \sum_{z_i=0}^{M-1} \frac{\left(\sum_{k=0}^{\left\lfloor \frac{M_x + M_y - 2 - z_i}{M} \right\rfloor} P(Z' = kM + z_i) - p_0(z) \right)^2}{p_0(z)},$$

где $p_0(z) = \frac{1}{M}$ соответствует гипотетическому (теоретическому) закону распределения равномерно распределенной случайной величины Z в области ее определения ($Z \in [0, M - 1]$);

$V = T = НОК(M_x, M_y)$ - объем выборки, равный периоду последовательности.

Далее рассчитанное значение требуется сравнить с квантилем закона распределения χ^2 заданного уровня значимости и сделать вывод о соответствии или несоответствии рассматриваемого закона распределения равномерному закону.

Для определения величины ошибки воспроизведения закона распределения дискретной случайной величины как числа символов ошибочного потока, приходящимся на единицу объема выборки, воспользуемся формулой из работы [[6]], где ошибка воспроизведения закона распределения определяется как

$$\xi = \frac{1}{2} \sum_{z_i=0}^{M-1} |P(Z = z_i) - p_0(z)|.$$

Если первичные последовательности равномерно распределенных случайных величин X и Y периодичны с периодами T_x и T_y , $|T_x|_{M_x} = 0$, $|T_y|_{M_y} = 0$, $НОД(T_x, T_y) = 1$ и хотя бы одно из значений M_x или M_y кратно M , то $g(x, y) = \frac{1}{M_x M_y}$,

$|T|_M = 0$, а $P(Z = z_i) = \frac{1}{M}$. Это свидетельствует о нулевой ошибке воспроизведения равномерного закона распределения символов на выходе комбинационного генератора:

$$\xi = \frac{1}{2} \sum_{z_i=0}^{M-1} |P(Z = z_i) - p_0(z)| = 0.$$

Заметим, что для любой подстановки на некотором множестве мощности M ошибка воспроизведения равномерного закона распределения также равна нулю: $\xi = 0$.

Выводы

Проведенное исследование позволило получить следующие результаты:

- определен закон распределения дискретной случайной величины на выходе комбинационного генератора, выполняющего операцию суммирования по некоторому модулю M слов, полученных от двух первичных генераторов равномерно распределенных случайных чисел;

- определены условия, при которых закон распределения дискретной случайной величины на выходе комбинационного генератора будет строго равномерным, т.е. ошибка воспроизведения равномерного закона распределения будет равна нулю. В качестве исходных первичных последовательностей случайных чисел рассмотрены последовательности истинно случайных чисел как с неограниченными, так и с ограниченными периодами, а также последовательности, представляющие собой циклически повторяющиеся подстановки.

Полученные результаты позволяют расширить теоретическую базу проектирования комбинационных генераторов случайных чисел и создают основу для дальнейшего анализа, разработки и практической реализации подобного рода генераторов.

Литература

[1] Giffe P.R. How to Protect Data With Ciphers That are Really Hard to Break / P.R. Giffe // Electronics. -1973. - V. 46. - N. 1. - PP. 99-101.

[2] Both T., Piper F.C. The Stop-and-Go Generator, Advances in Cryptology: Proceedings of EUROCRYPT 84, Springer-Verlag, 1984, pp. 88-92.

[3] D. Coppersmith, H. Krawczyk, Y. Mansour. The Shrinking Generator // Advances in Cryptology-CRYPTO '93 Proceedings, Springer-Verlag. - 1994. - pp. 22-39.

[4] Шнайер Б. Прикладная криптография. Протоколы, алгоритмы, исходные тексты на языке

Си; [пер. с англ. под ред. Семьянова П.В.]. - [2-е изд.]. - М.: Триумф, 2002. - 816 с.

[5] Кнут Д.Э. Искусство программирования: В 7 т.; [пер. с англ. В. Тертышный] . - [3-е изд.]. - М.: «Вильямс», 2007. - Т.2: Получисленные алгоритмы. - 832 с.

[6] Фауре Э.В., Береза А.С., Ярославская Е.А. Оценка точности воспроизведения закона распределения дискретной случайной величины при ее преобразовании // Вестник Хмельницкого национального университета. - 2012. - №5. - С. 176-182.

УДК 004.421.5 (045)

Фауре Е.В. Закон розподілу дискретної випадкової величини на виході комбінаційного генератора

Анотація. У статті розглядаються статистичні властивості дискретної випадкової величини на виході комбінаційного генератора, що виконує операцію підсумовування за деяким модулем слів, отриманих від двох первинних генераторів рівномірно розподілених випадкових чисел. Визначено закон розподілу дискретної випадкової величини на виході комбінаційного генератора. Визначено умови, за яких цей закон розподілу є строго рівномірним. У якості вихідних первинних послідовностей випадкових чисел розглянуто послідовності істинно випадкових чисел як з необмеженими, так і з обмеженими періодами, а також послідовності, що представляють собою циклічно повторювані підстановки. Отримані результати дозволяють розширити теоретичну базу проектування комбінаційних генераторів випадкових чисел і створюють основу для подальшого аналізу, розробки та практичної реалізації подібного роду генераторів.

Ключові слова: дискретна випадкова величина, послідовність випадкових чисел, комбінаційний генератор.

Faure E. Distribution law of discrete random variable in the combination generator output

Abstract. In the article the statistical properties of discrete random variable in the output sequence of the combination generator are reviewed. Combination generator performs the operation of summing by some modulo of words from two primary generators of uniformly distributed random numbers. The distribution law of discrete random variable in the generator output is defined. The conditions under which this distribution will be strictly uniform are defined. As the initial primary random numbers sequences are reviewed truly random numbers sequences with both limited and unlimited periods, as well as sequences which are cyclically repeated permutations. The obtained results allow us to expand the theoretical basis of design of combination of random number generators and provide a basis for further analysis, development and implementation of such generators.

Key words: discrete random variable, random numbers sequence, combination generator.

Отримано 15 квітня 2014 року, затверджено редколегією 20 травня 2014 року