Исследование качества генераторов псевдослучайных чисел при помощи батарей статистических тестов

Яковлева Ольга Валентиновна, гр.522

Санкт-Петербургский государственный университет
Математико-механический факультет
Кафедра статистического моделирования

Научный руководитель: к.ф.-м.н., доц. Коробейников А.И.
Рецензент: к.ф.-м.н., доц. Некруткин В.В.

Санкт-Петербург
2015г.

Постановка задачи

Генераторы вырабатывают последовательность целых чисел на решётке $\{0, \ldots, 2^d - 1\}$.

1. Моделирование последовательности вещественных чисел

$$\xi_1, \xi_2, \ldots, \xi_n.$$

$\mathcal{H}_0^U$: $\xi_1, \xi_2, \ldots, \xi_n$ — независимые, о.р., $U(0, 1)$.

2. Моделирование последовательности единиц и нулей (бит)

$$\beta_1, \beta_2, \ldots, \beta_n.$$

$\mathcal{H}_0^B$: $\beta_1, \beta_2, \ldots, \beta_n$ — независимые, о.р., $Ber(0.5)$.

Постановка задачи

Генераторы вырабатывают последовательность целых чисел на решётке $\{0, \ldots, 2^d - 1\}$.

1. Моделирование последовательности вещественных чисел

$$\xi_1, \xi_2, \ldots, \xi_n.$$

$\mathcal{H}_0^U$: $\xi_1, \xi_2, \ldots, \xi_n$ — независимые, о.р., $U(0,1)$.

2. Моделирование последовательности единиц и нулей (бит)

$$\beta_1, \beta_2, \ldots, \beta_n.$$

$\mathcal{H}_0^B$: $\beta_1, \beta_2, \ldots, \beta_n$ — независимые, о.р., $Ber(0.5)$.

Проверка гипотезы: построение теста

1. $\tau = f(\beta_1, \beta_2, \ldots, \beta_n)$ — статистика теста
2. $\mathcal{F}$ — известная функция распределения τ, если верна $\mathcal{H}_0^B$
3. $\rho = \mathcal{F}(\tau)$

   Если $\mathcal{F}$ непрерывна и гипотеза $\mathcal{H}_0^B$ верна, то $\rho \sim U(0,1)$
4. $p\text{-}value$ теста — реализация случайной величины ρ

Процесс тестирования

1. Вычисление k реплик статистики τ:

$$\tau_1 = f(\beta_1^{(1)}, \beta_2^{(1)}, \ldots, \beta_n^{(1)})$$
$$\vdots$$
$$\tau_k = f(\beta_1^{(k)}, \beta_2^{(k)}, \ldots, \beta_n^{(k)})$$

2. $\rho_1 = \mathcal{F}(\tau_1), \ldots, \rho_k = \mathcal{F}(\tau_k)$
3. Проверка гипотезы $\mathcal{H}_0$:
   $\rho_1, \rho_2, \ldots, \rho_k$ — независимые, о.р., $U(0,1)$
   с помощью критерия Колмогорова-Смирнова
4. $p\text{-}value^{Fin}$

Задачи

1. Проверить известные факты о качестве линейных конгруэнтных генераторов
2. Исследовать качество широко используемых генераторов с помощью битовых тестов
3. Исследовать способ улучшения линейных конгруэнтных генераторов с помощью метода замещения значений траектории

Процесс тестирования

- Система тестов TestU01:
  Alphabit, 17 тестов
  Rabbit, 33 теста
- Параметры тестирования

$$\tau_1 = f(\beta_1^{(1)}, \beta_2^{(1)}, \ldots, \beta_n^{(1)})$$
$$\vdots$$
$$\tau_k = f(\beta_1^{(k)}, \beta_2^{(k)}, \ldots, \beta_n^{(k)})$$

$n \approx 2^{23}$, $k = 100$

- Тест **пройден**, если $p\text{-}value^{Fin} > 0.05$

Извлечение последовательности бит:

- Последовательность целых можно рассматривать как последовательность бит:

$$\underbrace{010111\ldots0111}_{u_1},\underbrace{110101\ldots1010}_{u_2},\ldots$$

- Можно использовать не все двоичные разряды целого числа:

$$\underbrace{010111\ldots0\_\_\_}_{u_1},\underbrace{110101\ldots1\_\_\_}_{u_2},\ldots$$

Пример теста: RandomWalk, τ_C

- Входная последовательность бит $b_1, b_2, \ldots, b_n$ разбивается на N непересекающихся блоков длины L:

$$[b_1, b_2, \ldots, b_L], \ldots, [b_{L(N-1)+1}, b_{L(N-1)+2}, \ldots, b_{LN}]$$

- Для каждого блока вычисляются следующие величины:

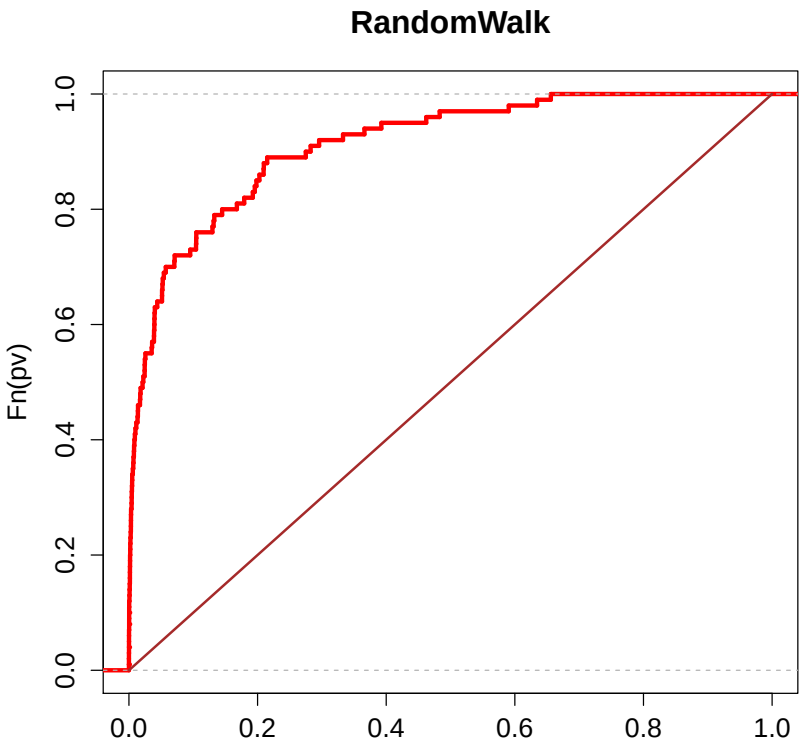
$$\Gamma_k = \sum_{j=1}^{k} c_j, \quad \text{где} \quad c_j = \begin{cases} 1, & \text{если } b_j = 1 \\ -1, & \text{если } b_j = 0. \end{cases}$$

$$\tau_C = \sum_{k=3}^{L} \mathbb{I}(\Gamma_{k-2}\Gamma_k < 0)$$

- τ_C — число перемен знака траектории процесса $\{\Gamma_k\}$
- После вычисления N реализаций случайной величины τ_C полученные частоты сравниваются с ожидаемыми с помощью критерия χ^2

Результат тестирования. LCG

$$u_{i+1} = (69069u_i + 1) \mod 2^{32}$$

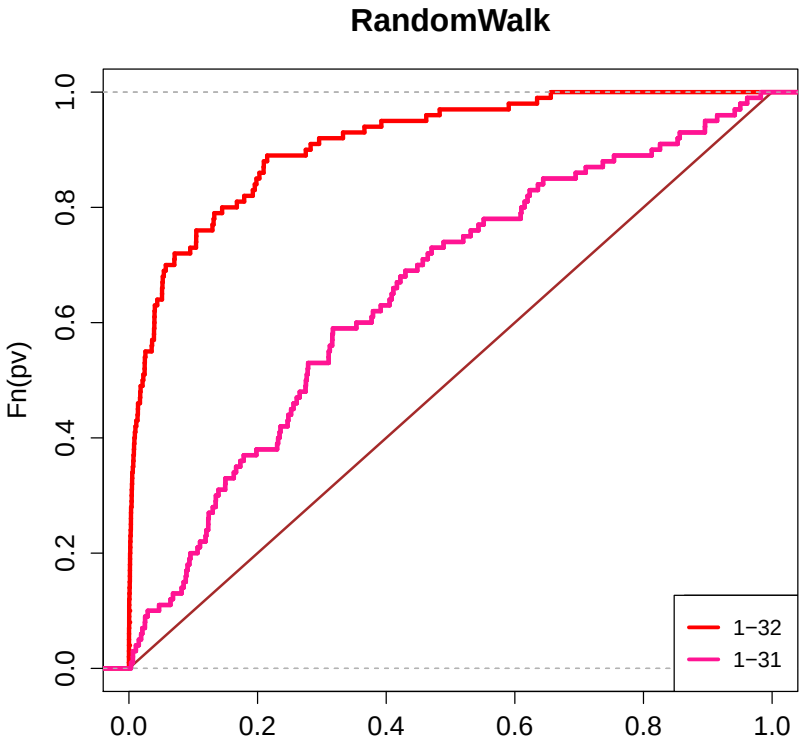


Количество пройденных тестов

Alphabit	Rabbit
17	33
1	7

Младшие двоичные разряды u_i плохого качества

Результат тестирования. LCG

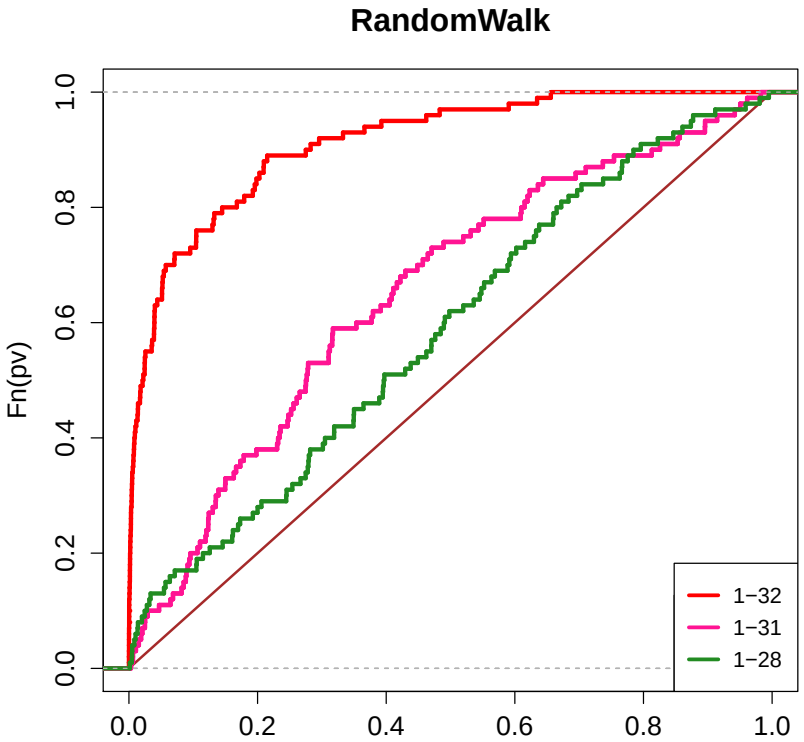


Количество пройденных тестов

Исп. биты	Alphabit 17	Rabbit 33
1-32	1	7
1-31	3	14

Исп. биты	1-32	1-31
$p\text{-}value^{Fin}$	<0.01	<0.01

Результат тестирования. LCG

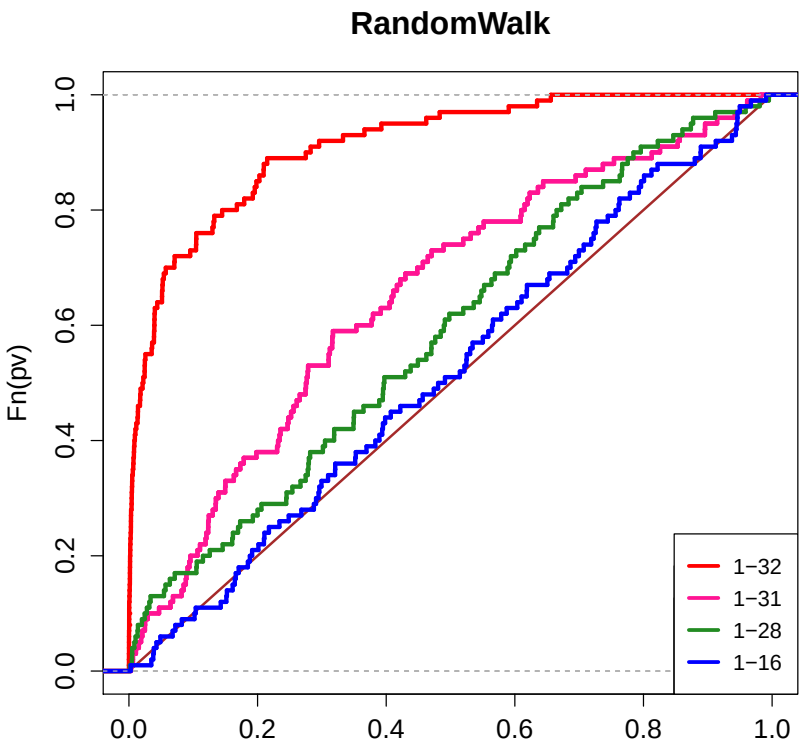


Количество пройденных тестов

Исп. биты	Alphabit 17	Rabbit 33
1-32	1	7
1-31	3	14
1-28	4	16

Исп. биты	1-32	1-31	1-28
$p\text{-}value^{Fin}$	<0.01	<0.01	0.04

Результат тестирования. LCG

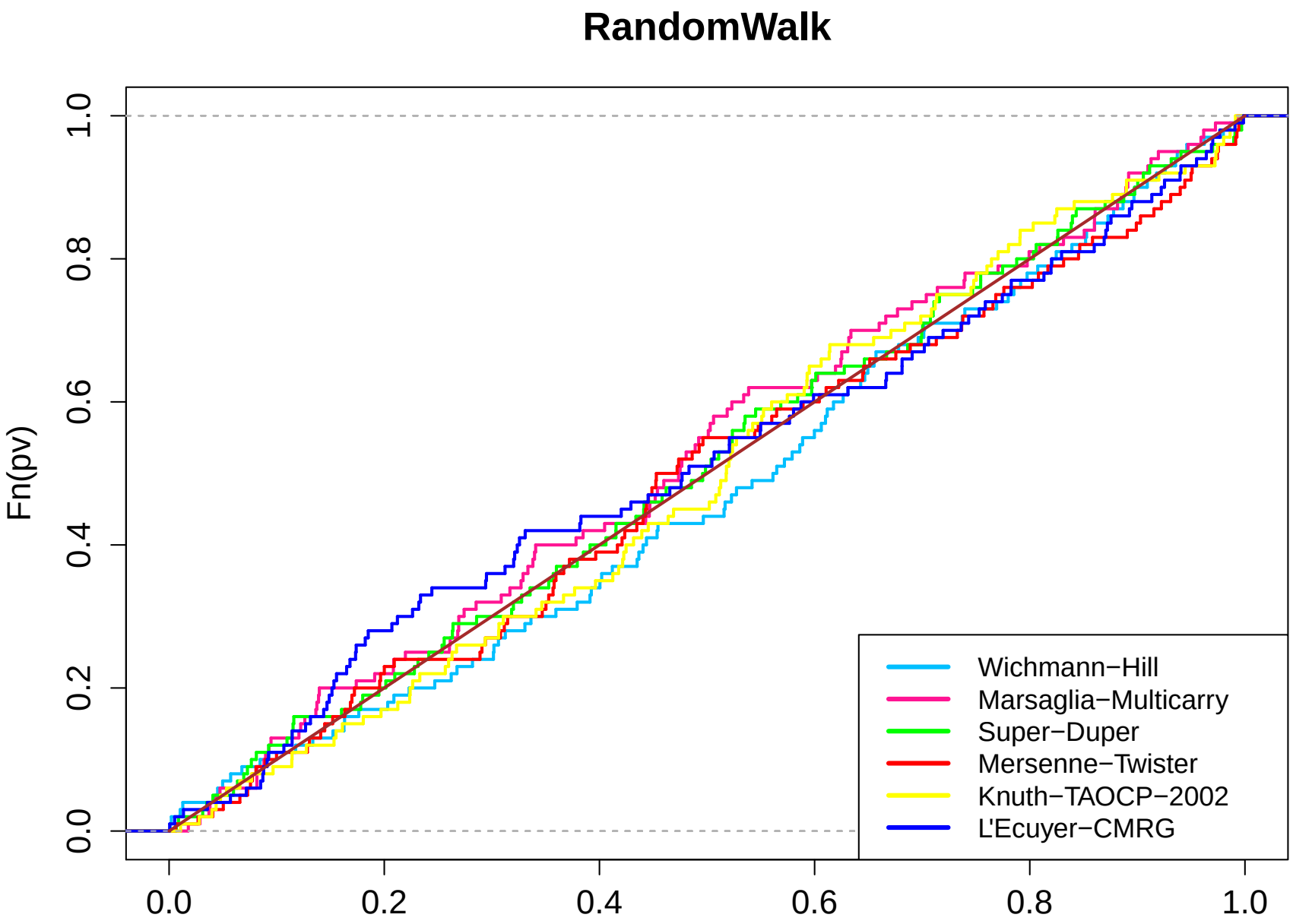


Количество пройденных тестов

Исп. биты	Alphabit 17	Rabbit 33
1-32	1	7
1-31	3	14
1-28	4	16
1-16	16	28

Исп. биты	1-32	1-31	1-28	1-16
$p\text{-}value^{Fin}$	<0.01	<0.01	0.04	0.88

Встроенные генераторы пакета R



Встроенные генераторы пакета R

Генератор	Alphabit (17 тестов)	Rabbit (33 теста)
Wichmann-Hill	16	30
Marsaglia-Multicarry	17	32
Super-Duper	17	29
Mersenne-Twister	16	30
Knuth-TAOCP-2002	17	28
L'Ecuyer-CMRG	16	29

Все генераторы проходят большинство тестов

Комбинация генераторов

Исходный генератор: $u_{i+1} = (69069u_i + 1) \mod 2^{32}$

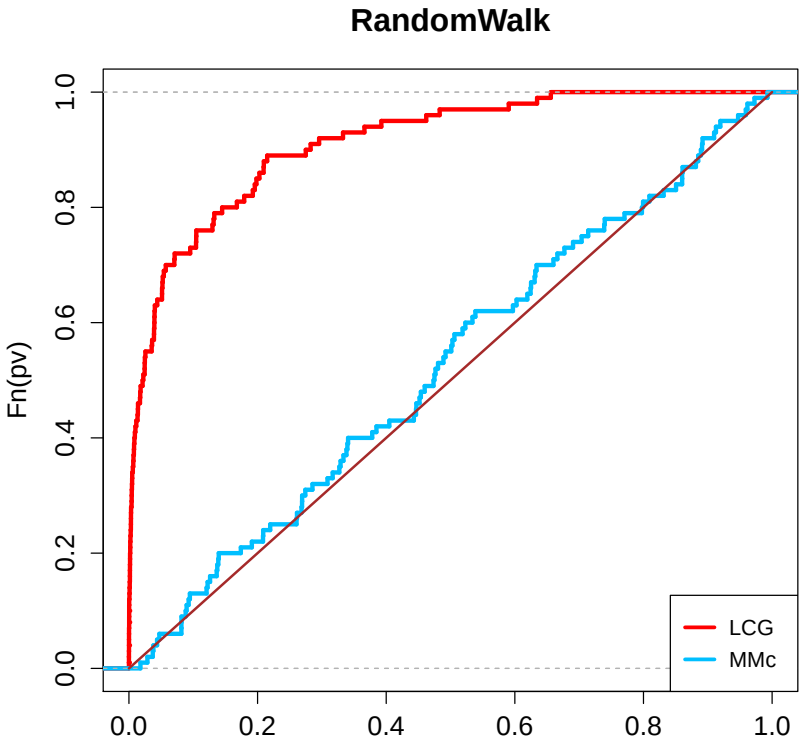
- $g_1, g_2, \ldots$ — последовательность целых, выработанных надёжным генератором
- Через каждые r значений траектория начинается с нового значения:

  $u_1^{(0)}, u_2^{(0)}, \ldots, u_{r-1}^{(0)}, \quad g_1, u_1^{(1)}, u_2^{(1)}, \ldots, u_{r-1}^{(1)}, \quad g_2, \ldots$

Предполагаемая выгода:

- Улучшение свойств
- Увеличение периода
- Сохранение высокой скорости работы

Комбинация генераторов

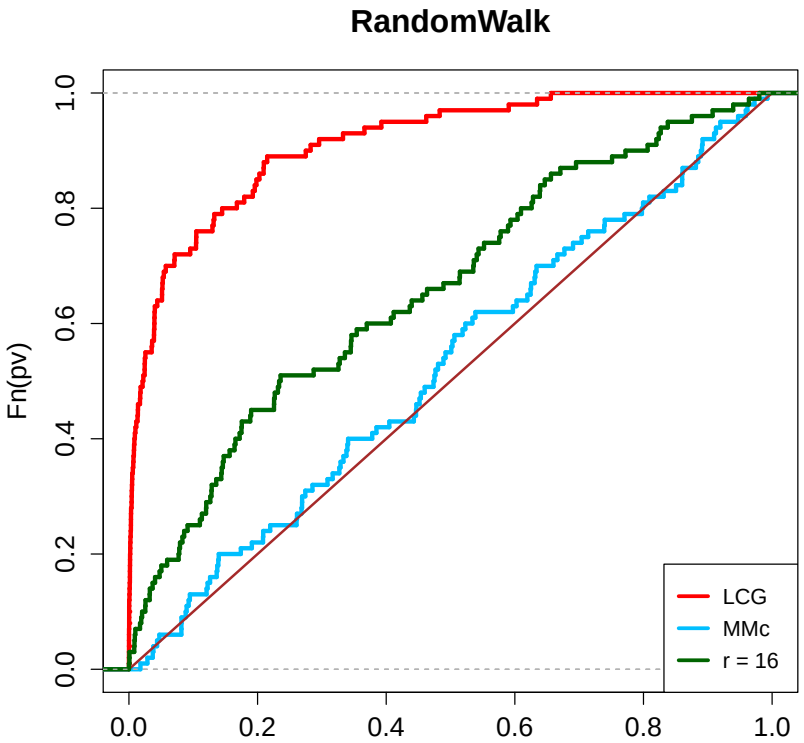


Количество пройденных тестов

	Alphabit 17	Rabbit 33
LCG	1	7
MMc (R)	17	32

	LCG	MMc (R)
$p\text{-}value^{Fin}$	<0.01	0.52

Комбинация генераторов

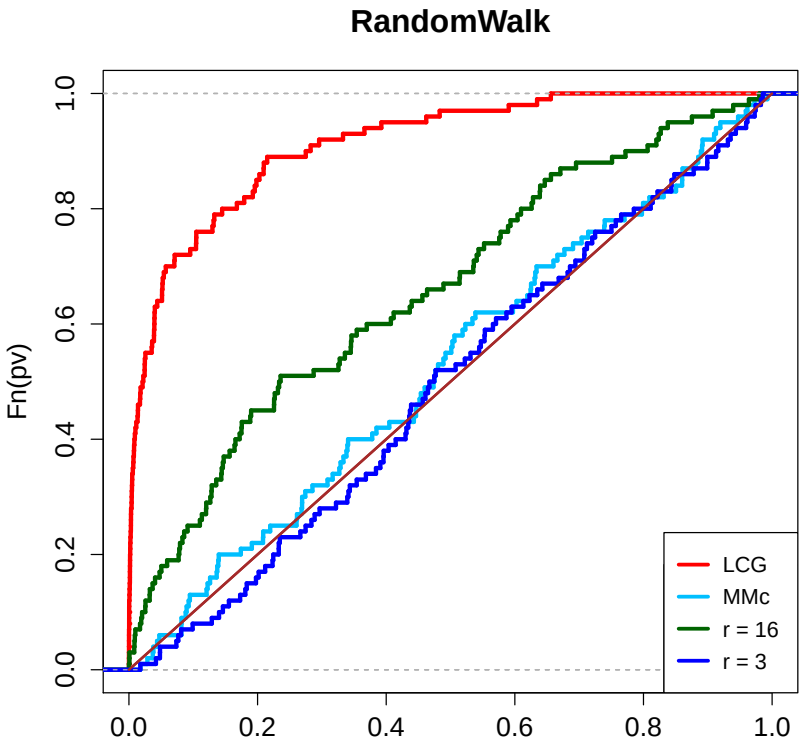


Количество пройденных тестов

	Alphabit 17	Rabbit 33
LCG	1	7
MMc (R)	17	32
Комбинация $r = 16$	7	14

	LCG	MMc (R)	$r = 16$
$p\text{-}value^{Fin}$	<0.01	0.52	<0.01

Комбинация генераторов



Количество пройденных тестов

	Alphabit 17	Rabbit 33
LCG	1	7
MMc (R)	17	32
Комбинация		
$r = 16$	7	14
$r = 3$	11	27

	LCG	MMc (R)	$r = 16$	$r = 3$
$p\text{-}value^{Fin}$	<0.01	0.52	<0.01	0.94

Результаты

1. С помощью батарей битовых тестов Alphabit и Rabbit проведено тестирование широко распространённых генераторов
2. Исследован способ комбинации генераторов с помощью метода замещения значений траектории
3. Показано, что предложенный метод позволяет улучшить статистические свойства исходного генератора