Лабораторна робота №1
з дисципліни "комп'ютерна статистика"
Варіант №4

Горбунова Даніела Денисовича
4 курс бакалаврату
група "комп'ютерна статистика"

6 вересня 2020 р.

1 Вступ.

У даній роботі вказана інформація про отримані результати під час виконання роботи №1, а саме: побудова та перевірка якості лінійного конгруентного генератора псевдовипадкових послідовностей для різних наборів значень параметрів, генерування величин з логнормальним розподілом двома способами. Додатково була розроблена функція, що генерує дискретну послідновність методом Фібоначчі.

1.1 Теоретичні відомості.

Розглядається лінійний конгруентний генератор для побудови псевдовипадкових послідовностей з додатними цілими значеннями, що заданий рекурентною формулою

$$\forall n \geqslant 2: \ I_n = aI_{n-1} + c \mod m, \ a, c, m \in \mathbb{N} \tag{1}$$

де a - множник, c - приріст, а m - модуль генератора. Початкове значення I_1 вибирають з інтервалу $1, \ldots, m-1$. Якщо a, c, m задовольняють умовам теореми Халла-Добелла, то гаратовано, що період генератора (1) дорівнює його значенню модуля. Величини на $[0, 1]$ отримаємо діленням I_n на модуль m.

Зрозуміло, що виконання умов теореми недостатньо для того, щоб послідовність добре імітувала випадкову. Тому для перевірки на випадковість та відповідність до заданого розподілу будемо використовувати графічні методи дослідження послідовності та критерій узгодженості Колмогорова у конкретних випадках. Серед графічних наведено порівняння графіків емпіричної та теоретичної функції розподілу; побудова діаграм розсіювання елементів послідовності на площині та в просторі. Щодо критерію Колмогорова, то буде перевірена нульова гіпотеза

$$H_0 = \{P(\xi_1 < t) = F(t), \xi = (\xi_1, \ldots, \xi_n) - \text{кратна вибірка}\}$$

проти альтернативи

$$H_1 = \{P(\xi_1 < t) = G(t) \neq F(t), \xi = (\xi_1, \ldots, \xi_n) - \text{кратна вибірка}\}$$

2 Лінійні конгруентні генератори. Генератор Парка та Мілера.

Зауваження. Для усіх лінійних конгруентних генераторів, що розглядаються в роботі, в якості зернини береться число $I_1 = 2^{15}$.

Для мультиплікативного генератора числових послідовностей Парка-Мілера фіксують наступні значення параметрів:

$$m = 2^{31} - 1,\ a = 7^5,\ c = 0$$

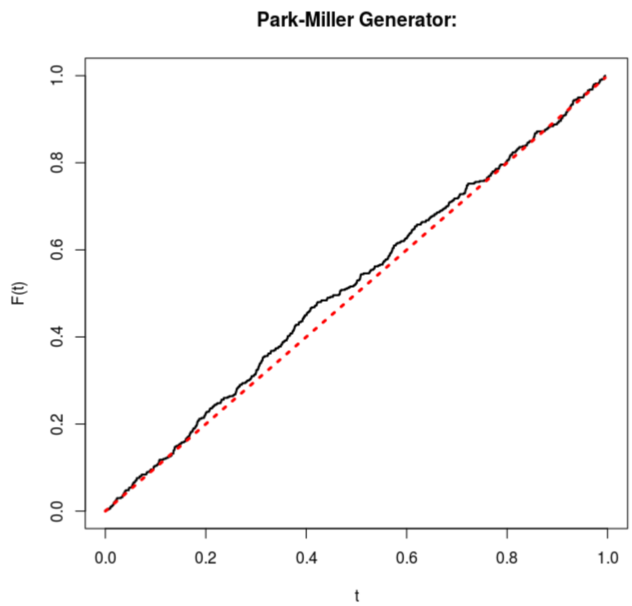

Рис. 1: Графіки емпіричної та теоретичної функції рівномірного розподілу.

Після створення масиву з 500 псевдовипадкових чисел на $[0, 1]$, можемо побудувати емпіричну функцію розподілу та порівняти її форму з теоретичною функцією рівномірного розподілу. На графіку видно, що емпірична функція розподілу подібна до теоретичної, але помітно вигинається вгору на проміжку $[0.2, 0.8]$. Такі відхилення можуть суттєво впливати на результат тесту Колмогорова. Дійсно, покажемо значення статистики та досягнутого рівня значущості:

```
One-sample Kolmogorov-Smirnov test
data:  x
D^+ = 0.057397, p-value = 0.03709
alternative hypothesis: the CDF of x lies above the null hypothesis
```

Для стандартного рівня значущості $\alpha = 0.05$ довелося б відхилити нульову гіпотезу та прийняти альтернативу. Далі розглянемо діаграми розсіювання пар та трійок отриманої послідовності.

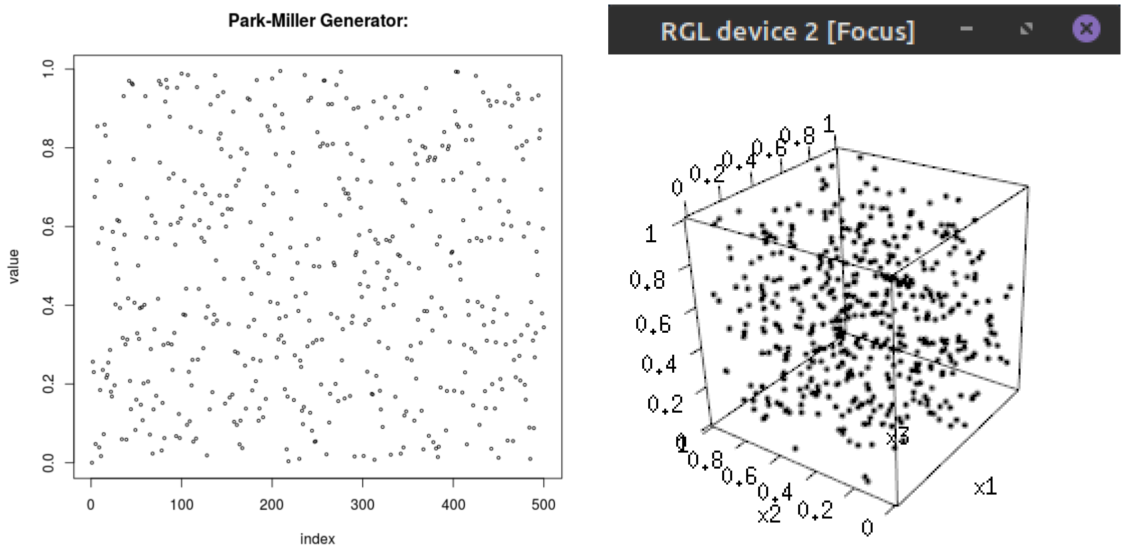

Рис. 2: Діаграми розсіювання пар та трійок елементів послідовності.

Як на площині, так і в просторі досить складно помітити закономірності між парами (трійками) елементів псевдовипадкової послідовності. Видно, що набор чисел, згенерований за допомогою генератора Парка-Мілера, добре імітує випадковість та проходить ці візуальні тести.

Розглянемо інший лінійний конгруентний генератор.

3 Лінійні конгруентні генератори. Генератор із запропонованим набором параметрів.

Для спеціального мультиплікативного генератора числових послідовностей беремо значення параметрів, які відповідають номеру варіанта:

$$m = 2^{31},\ a = 75831,\ c = 0$$

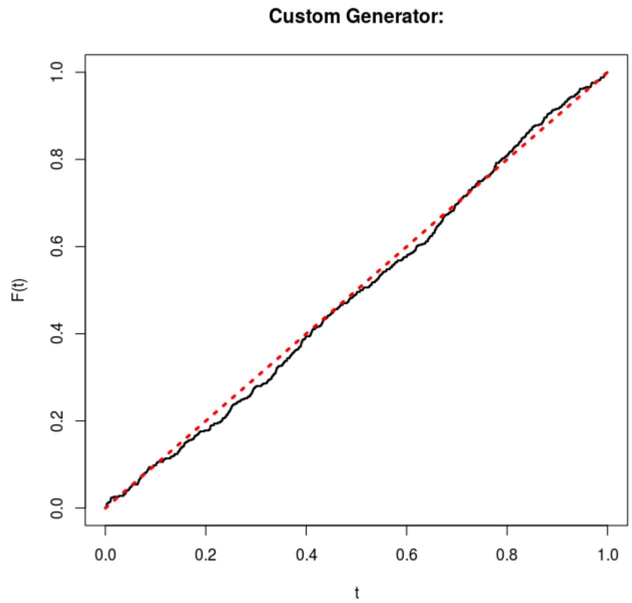

Рис. 3: Графіки емпіричної та теоретичної функції рівномірного розподілу.

На відміну від минулого конкурента, даний графік емпіричної функції розподілу розміщений ближче до кривої ймовірнісної функції розподілу. На $[0, 0.4]$ бачимо відхилення від теоретичного розподілу донизу, але не таке сильне, як бачили в минулому розділі. Критерій Колмогорова повертає нові значення, у сторону прийняття нульової гіпотези:

```
One-sample Kolmogorov-Smirnov test
data:  x
D^+ = 0.02404, p-value = 0.5611
alternative hypothesis: the CDF of x lies above the null hypothesis
```

Для стандартного рівня, зазначеного раніше, вважаємо, що істиний розподіл згенерованих спостережень подібний до неперервного рівномірного на $[0, 1]$. Тепер знову розглянемо діаграми розсіювання пар та трійок нової послідовності.

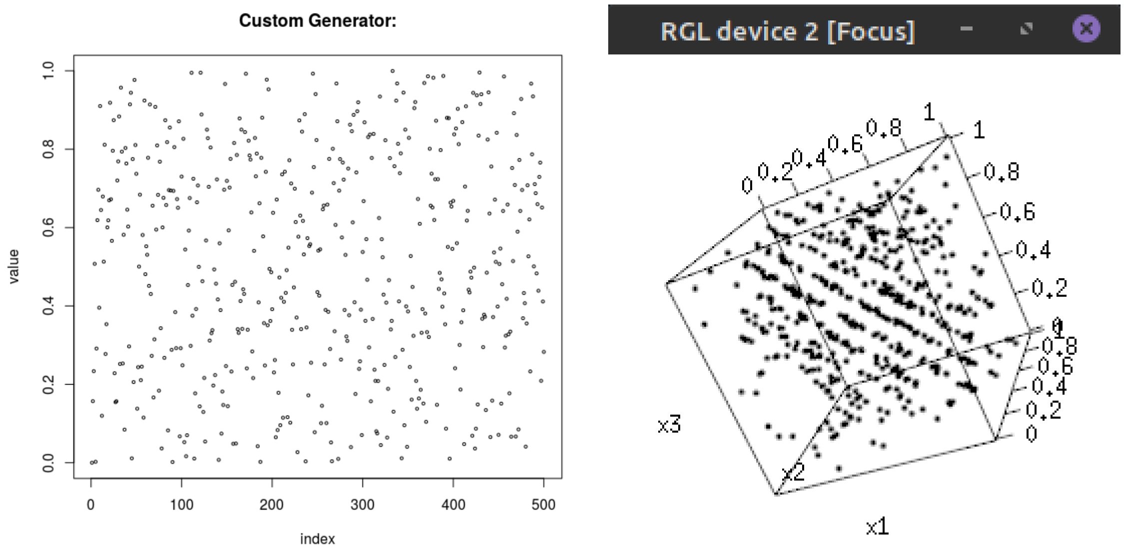

Рис. 4: Діаграми розсіювання пар та трійок елементів послідовності.

Спроби імітувати випадковість досить хороші, поки не перейти з площини у простір. Після незначної кількості поворотів можна побачити, що трійки послідовності вишукуються у паралельні в сукупності площини, що не притаманне випадковим послідовностям. На цьому етапі тест на випадковість для другого генератора закінчено з абсолютним провалом.

4 Генерування псевдовипадкових послідовностей з логнормальним розподілом. Перетворення Бокса-Мюллера.

Зауваження. У розділах 4-6 будемо користуватися лінійним конгруентним генератором Парка та Мілера, базуючись на отриманих результатах щодо перевірки на випадковість та відповідності до рівномірного розподілу.

Маючи незалежні рівномірно розподілені на $(0,1]$ величини r, φ, можемо вибрати одне з нормальних перетворень Бокса-Мюллера

$$\zeta_1 = \cos(2\pi\varphi)\sqrt{-2\ln r}$$
$$\zeta_2 = \sin(2\pi\varphi)\sqrt{-2\ln r}$$

Оскільки $\zeta_i \sim N(0,1)$, то $\exp(\mu + \sigma\zeta_i) \sim \mathrm{logN}(\mu, \sigma^2)$, де $\mu \in \mathbb{R}, \sigma > 0$.

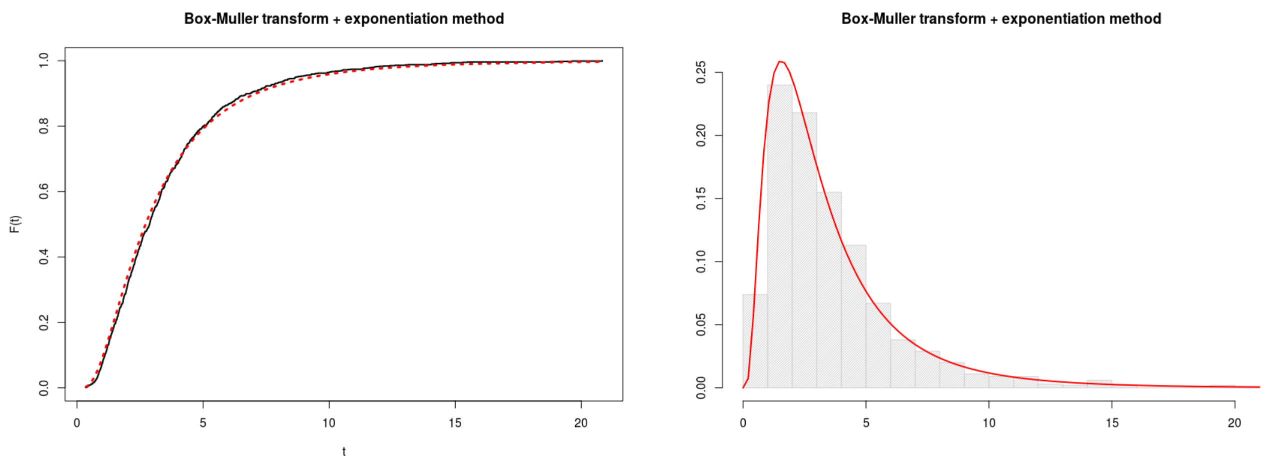

Рис. 5: Метод Бокса-Мюллера: Емпірична функція розподілу та гістограма згенерованих спостережень за $\mathrm{logN}(1, 0.75)$.

З рисунків видно, що розподіл отриманих чисел подібний до логнормального. Суттєвих відхилень від істинної функції розподілу не видно. Змальовується непогана картина на гістограмі згенерованої послідовності, яка відповідає кривій щільності логнормального розподілу.

5 Генерування псевдовипадкових послідовностей з логнормальним розподілом. Метод проріджування.

Нагадаємо функцію щільності випадкової величини за логнормальним розподілом з параметрами μ, σ^2:

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right) \mathbb{1}_{(0,\infty)}(x)$$

Далі беремо таку щільність деякого розподілу $g(x)$, для якого можна задати $C > 0$:

$$f(x) \leqslant Cg(x),\ \forall x$$

У якості $g(x)$ зафіксували щільність стандартного розподілу Коші. Взявши число $C = 4\pi$, відношення $\frac{f(x)}{Cg(x)}$ набуває вигляду

$$\frac{f(x)}{Cg(x)} = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right) \mathbb{1}_{(0,\infty)}(x) \frac{(1+x^2)}{4}\Bigg|_{\mu=1,\ \sigma=0.75}$$

Для генерування послідовності за розподілом Коші, можна скористатися відомим перетворенням з курсу теорії ймовірностей:

$$\eta \sim U[0,1] \Rightarrow \zeta = \operatorname{tg}\left(\pi\left(\eta - \frac{1}{2}\right)\right) \sim \text{Cauchy}(0,1)$$

Тоді згенеровані величини не вилучаємо з послідовності, якщо задовольняють умову:

$$u < \frac{f(\zeta)}{Cg(\zeta)};\ u, \zeta - \text{незалежні}$$

У результаті маємо послідовність величин з розподілом подібним до логнормального, але з суттєвим недоліком. Після проведених операцій функція розподілу послідовності буде зкошена від істиного теоретичного розподілу в ліву сторону. Цей приклад показує, що для побудови якісного генератора послідовностей методом проріджження необхідно приділяти увагу підбору другої функції щільності для виконання нерівності (5).

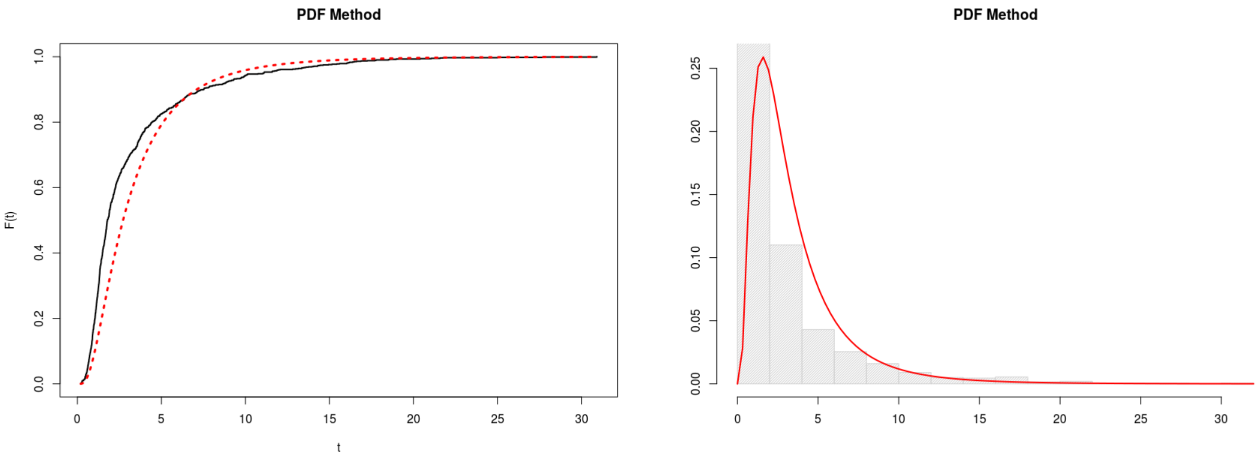

Рис. 6: Метод проріджування: Емпірична функція розподілу та гістограма згенерованих спостережень за $\log\mathrm{N}(1, 0.75)$.

6 Генерування псевдовипадкових послідовностей з логнормальним розподілом. Стандартна функція в R.

Вбудований генератор псевдовипадкових послідовностей з логнормальним розподілом, який можна викликати функцією rlnorm(n, log.mean, log.sd, . . .), використовує алгоритм квантильного перевторення, якщо користувач не буде вносити певні зміни в аргументах RNGkind або фукнції set.seed.

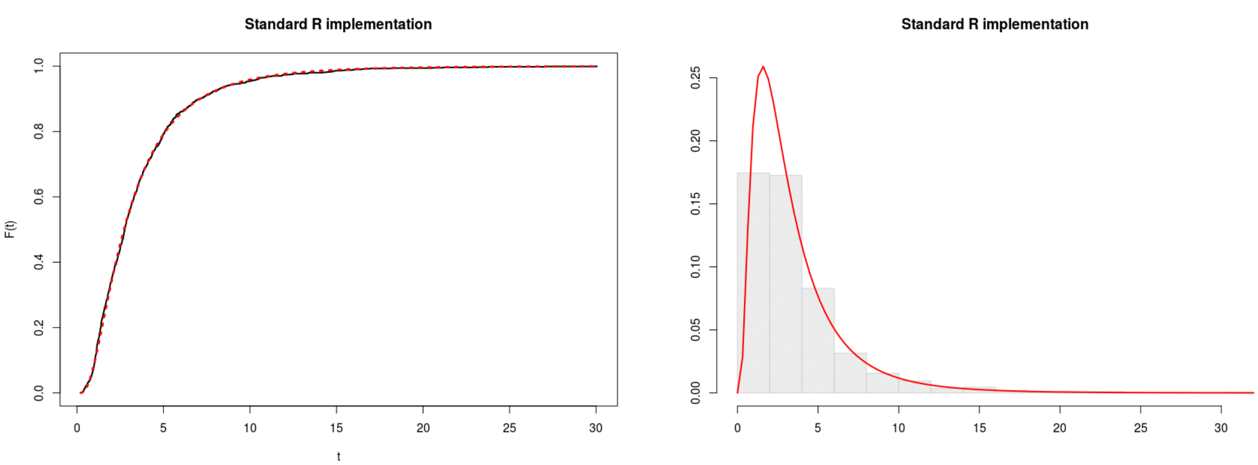

Рис. 7: Застосування стандартних функцій в R: Емпірична функція розподілу та гістограма згенерованих спостережень за $\log\mathrm{N}(1, 0.75)$.

Відхилення емпіричного розподілу від теоретичного незначне, якщо дивитися неозброєним оком. На гістограмі дивних явищ не спостерігається.

7 Деякі відомості про метод Фібоначчі.

Рекурентний генератор Фібоначчі для побудови псевдовипадкових послідовностей задається формулою

$$I_n = I_{n-k} + I_{n-l} \mod m, \tag{2}$$

де $n > l$ та $k < l$. Тобто спочатку необхідно задати перші l членів послідовності. У рамках цієї роботи поклали $l = 17, k = 5$ та перші l елементів були згенеровані за допомогою лінійного конгруентного генератора цілих чисел з параметрами, що були вказані в розділі 3. Покажемо як себе проявила модель при генерації чисел на $[0, 1]$ та за логнормальним розподілом.

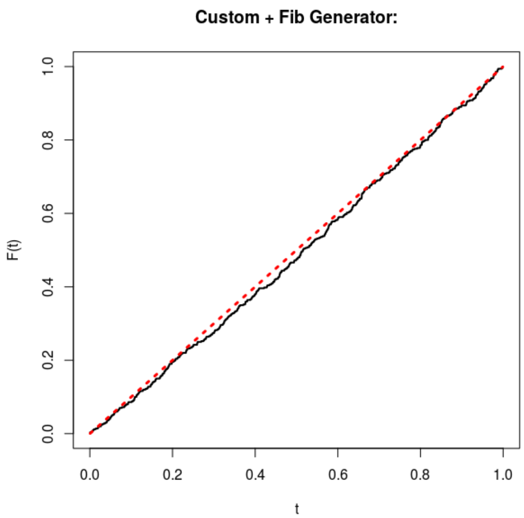

Рис. 8: Графіки емпіричної та теоретичної функції рівномірного розподілу.

Видно, що емпіричний розподіл близький до теоретичного рівномірного на одиничному відрізку. Маємо незначне відхилення на проміжку $[0.2, 0.6]$. Покажемо діаграми розсіювання.

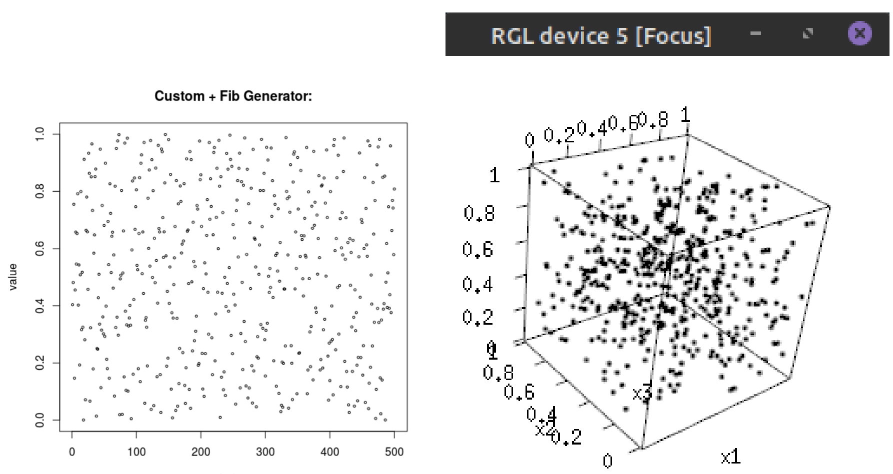

Рис. 9: Діаграми розсіювання пар та трійок елементів послідовності.

Як для діаграми пар елемнтів послідовності, так і для її трійок помітити закономірності важко. Генератор на основі методу Фібоначчі добре імітує випадковість та проходить задані візуальні тести.

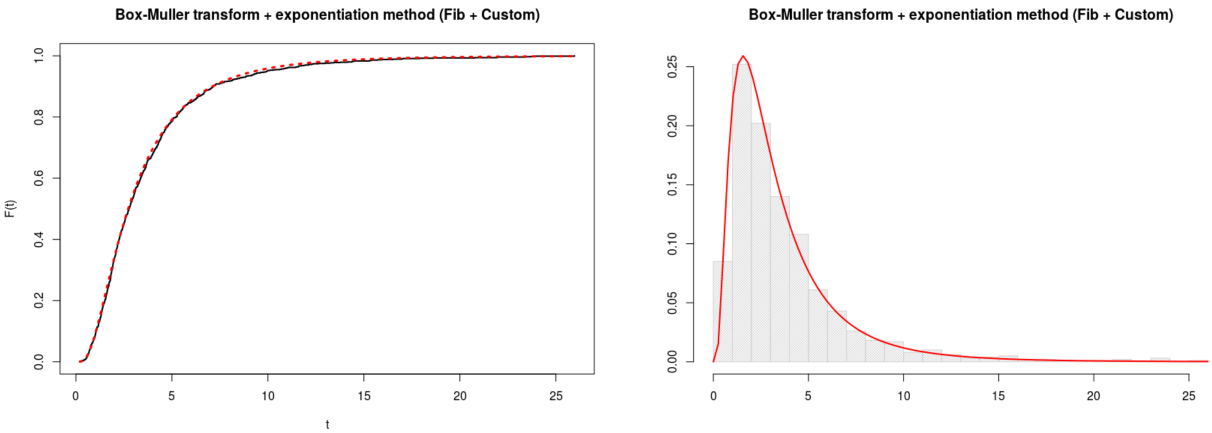

Рис. 10: Метод Бокса-Мюллера: Емпірична функція розподілу та гістограма згенерованих спостережень за $\log\mathrm{N}(1, 0.75)$.

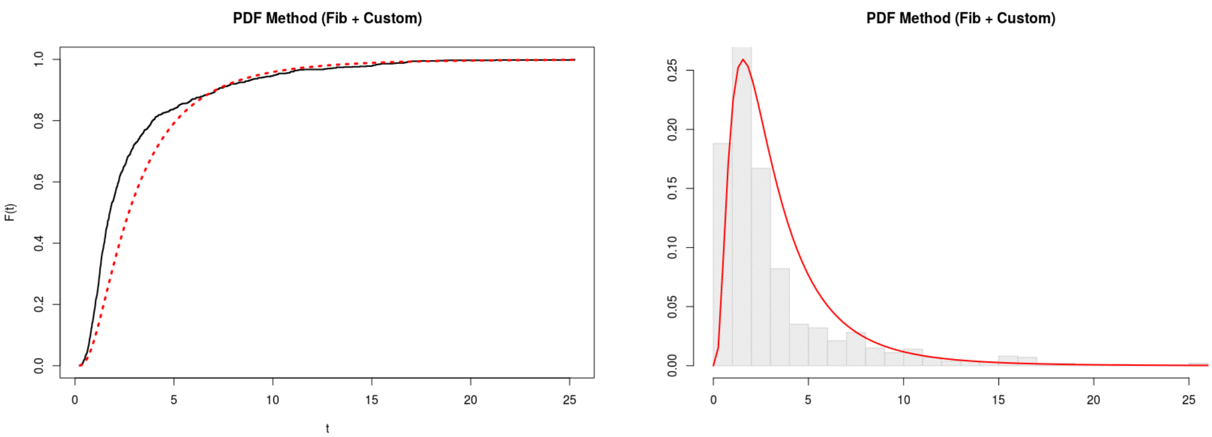

Рис. 11: Метод проріджування: Емпірична функція розподілу та гістограма згенерованих спостережень за $\log\mathrm{N}(1, 0.75)$.

Аналогічно до випадку з класичним лінійним конгруентним генератором, краща генерація чисел за розподілом отримана методом перетворення Бокса-Мюллера. Але проблему нестандартної кривизни емпіричної функції розподілу методом проріджження не вдалося вирішити зміною генератора послідовностей.

8 Висновки.

У лабораторній роботі, лінійний конгруентний генератор Парка-Мілера та генератор Фібоначчі з фіксованим запізненням показали хороші результати в тестах на випадковість. Для генерування величин з логнормальним розподілом, маємо гірші результати при застосуванні методу прорідження - це пояснюється не зовсім вдалим підбором іншої неперервної випадкової величини для виконання певних нерівностей. Метод перетворення Бокса-Мюллера по якості виходить дещо близьким до вбудованого генератора логнормальних величин в мові програмування R.