20 Истинно-истинно случайные числа

Задача. Трент подарил Бобу истинно-истинно случайный 6-гранный кубик, результаты бросков которого независимы и равновероятны. Бросая кубик, Боб генерирует истинно-истинно случайные буквы русского алфавита. Верно ли, что для генерации отдельной буквы (одной из 33) Боб может использовать менее $\sqrt{5}$ бросков кубика в среднем?
Дополнительный вопрос $(*)$: верно ли, что для генерации отдельной буквы английского алфавита (одной из 26) потребуется больше$\sqrt{5}$ бросков в среднем?
Дополнительный вопрос $(\star)$: изменятся ли ответы на предыдущие вопросы при замене $\sqrt{5}$ на $21163/9721$?

Решение. Бросив кубик дважды, Боб получает случайные числа $r_0, r_1 \in \{0, 1, \ldots, 5\}$ ($\boxdot = 0, \boxdot = 1, \ldots, \boxplus = 5$). По ним Боб может построить случайное число $r = r_0 + 6r_1$ с равномерным распределением на $\{0, 1, \ldots, 35\}$. Если $r < 33$, то Боб преобразует r в истинно-истинно случайный символ русского алфавита ($\texttt{А} = 0, \texttt{Б} = 1, \ldots, \texttt{Я} = 32$). Если же $r \in \{33, 34, 35\}$, то Боб снова бросает кубик дважды, снова формирует r и так далее, до тех пор, пока условие $r < 33$ не будет выполнено.

Среднее число бросков для генерации одного символа:

$$\sum_{t\geq 1} 2t\mathsf{P}\left\{\text{потребуется } 2t \text{ бросков}\right\} = 2\sum_{t\geq 0} \mathsf{P}\left\{\text{потребуется более } 2t \text{ бросков}\right\} =$$
$$= 2\sum_{t\geq 0}\left(\frac{3}{36}\right)^t = 2\cdot\frac{36}{33} < \sqrt{5},$$

и ответ на первый вопрос положительный.

При переходе к английскому алфавиту Бобу после двух бросков потребуется сделать еще по крайней мере один с вероятностью $\frac{36-26}{36}$. Поэтому среднее число бросков не меньше

$$2 + \frac{10}{36} > \sqrt{5},$$

и ответ на второй вопрос также положительный.

Описанный способ генерации случайных чисел соответствует методу исключения фон Неймана. Является ли способ оптимальным? Оказывается, что нет — существует алгоритм, в котором среднее число бросков меньше. Основная идея алгоритма — не игнорировать случайное число r, даже если оно не попадает в допустимое множество.

Опишем алгоритм для общего случая. Пусть имеются истинно-истинно случайные цифры от 0 до $a-1$. Покажем как, используя эти цифры, построить истинно-истинно случайное число от 0 до $n-1$. На шагах алгоритма будем пересчитывать порог R и формировать случайное число r с равномерным распределением на $\{0, 1, \ldots, R-1\}$. Будем ожидать выполнения условия $r < n$. Конкретнее, алгоритм имеет следующий вид:

1. Установить $r \leftarrow 0$, $R \leftarrow 1$.
2. Найти $d \leftarrow \lceil \log_a n/R \rceil$.
3. Получить d очередных случайных цифр $r_0, \ldots, r_{d-1}$.
4. Установить $r \leftarrow r_0 + ar_1 + \ldots + a^{d-1}r_{d-1} + a^d r$, $R \leftarrow R \cdot a^d$ (r снова имеет равномерное распределение на $\{0, 1, \ldots, R-1\}$).
5. Найти $e \leftarrow R \bmod n$.
6. Если $r \in \{0, 1, \ldots, R-e-1\}$, т. е. $r < n$, то возвратить r.

7. Здесь r имеет равномерное распределение на $\{R-e,\dots,R-1\}$. Учтем этот факт следующими образом: $r \leftarrow r-(R-e)$, $R \leftarrow e$.

8. Вернуться к шагу 2.

Оценим сложность алгоритма. Пусть $f_{a,n}(R)$ — среднее число цифр, которое потребуется использовать при наличии случайного числа r с равномерным распределением на $\{0,1,\dots,R-1\}$. Тогда

$$f_{a,n}(R) = d + \frac{Ra^d \bmod n}{Ra^d} f_{a,n}(Ra^d \bmod n), \quad d = \lceil \log_a n/R \rceil.$$

Например, в интересующем нас случае $a=6$, $n=33$ получаем систему уравнений:

$$\begin{aligned}
f_{6,33}(1) &= 2 + \frac{1}{12} f_{6,33}(3),\\
f_{6,33}(3) &= 2 + \frac{1}{12} f_{6,33}(9),\\
f_{6,33}(9) &= 1 + \frac{7}{18} f_{6,33}(21),\\
f_{6,33}(21) &= 1 + \frac{3}{14} f_{6,33}(27),\\
f_{6,33}(27) &= 1 + \frac{5}{27} f_{6,33}(30),\\
f_{6,33}(30) &= 1 + \frac{1}{12} f_{6,33}(15),\\
f_{6,33}(15) &= 1 + \frac{4}{15} f_{6,33}(24),\\
f_{6,33}(24) &= 1 + \frac{1}{12} f_{6,33}(12),\\
f_{6,33}(12) &= 1 + \frac{1}{12} f_{6,33}(6),\\
f_{6,33}(6) &= 1 + \frac{1}{12} f_{6,33}(3).
\end{aligned}$$

Решая эту систему, определяем интересующее нас среднее число бросков

$$f_{6,33}(1) = \frac{84653}{38885} < \frac{84653-1}{38885-1} = \frac{21163}{9721} \approx 2.177.$$

Для сравнения,

$$f_{6,26}(1) = \frac{544111}{233285} \approx 2.332.$$