

ГЕНЕРАЦИЯ СЛУЧАЙНЫХ СОБЫТИЙ МЕТОДОМ ПСЕВДОНИМОВ

*М.Е. Семёнов, к.т.н., доц.,
Е.Е. Кудряшова, студент гр. 0ВМ01
Томский политехнический университет
E-mail: eek19@tpu.ru*

Введение

Существуют различные подходы для генерации случайных чисел с настраиваемым распределением. Выбор зависит от многих условий, в том числе от размера выборки, закона распределения, а также меняется ли этот закон со временем. Целью данной работы является изучение алгоритма псевдонимов и сравнение его с другими методами.

Описание алгоритма

Метод псевдонима состоит в следующем:

1. Каждая заданная вероятность увеличивается в n раз, где n - количество различных значений, для которых задано распределение. Полученные значения - "нормированными вероятностями".
2. Нормированные вероятности разделяются на 2 группы. В первой группе $P_{norm} \geq 1$, во второй группе $P_{norm} < 1$.
3. Строится таблица вероятностей `prob` и таблица псевдонимов `alias`, для этого:
 - a. случайным образом берется вероятность P_{2i} из второй группы и перемещается в итоговую таблицу вероятностей `prob` (в ячейку, соответствующую данному значению);
 - b. из первой группы также выбирается случайная вероятность P_{1i} , и из нее вычитается $(1 - P_{2i})$;
 - c. если в итоге P_{1i} становится меньше 1, то она перемещается во вторую группу, иначе - остается в первой;
 - d. значение, соответствующее вероятности P_{1i} , становится псевдонимом для P_{2i} (это значение добавляется в таблицу псевдонимов `alias` в соответствующую ячейку);
 - e. вышеуказанные шаги повторяются до тех пор, пока вторая группа не пуста;
 - f. последним шагом в итоговую таблицу добавляются оставшиеся вероятности из первой группы (все они будут равны 1). Псевдонимы для них не требуются.

Сравнение алгоритма с другими методами

Усредненное сравнение скорости работы алгоритма при обработке различного набора кадров видеофайла на одинаковом фрагменте поиска представлено в таблице 1.

Для проверки правильности написанного алгоритма проведена генерация 1000 выборок с заданным распределением вероятности. Далее с целью статистического сравнения проведена проверка гипотезы H_0 : "полученное распределение статистически не отличается от целевого распределения", с уровнем значимости $\alpha=0,95$. Такая гипотеза выдвинута в отношении каждого из 1000 полученных распределений. Далее рассмотрим пример проверки для равномерного, нормального и биномиального распределений. Для сравнения гистограмм были использованы следующие методы: метрика Вассерштейна, критерий хи-квадрат и критерий Колмогорова-Смирнова.

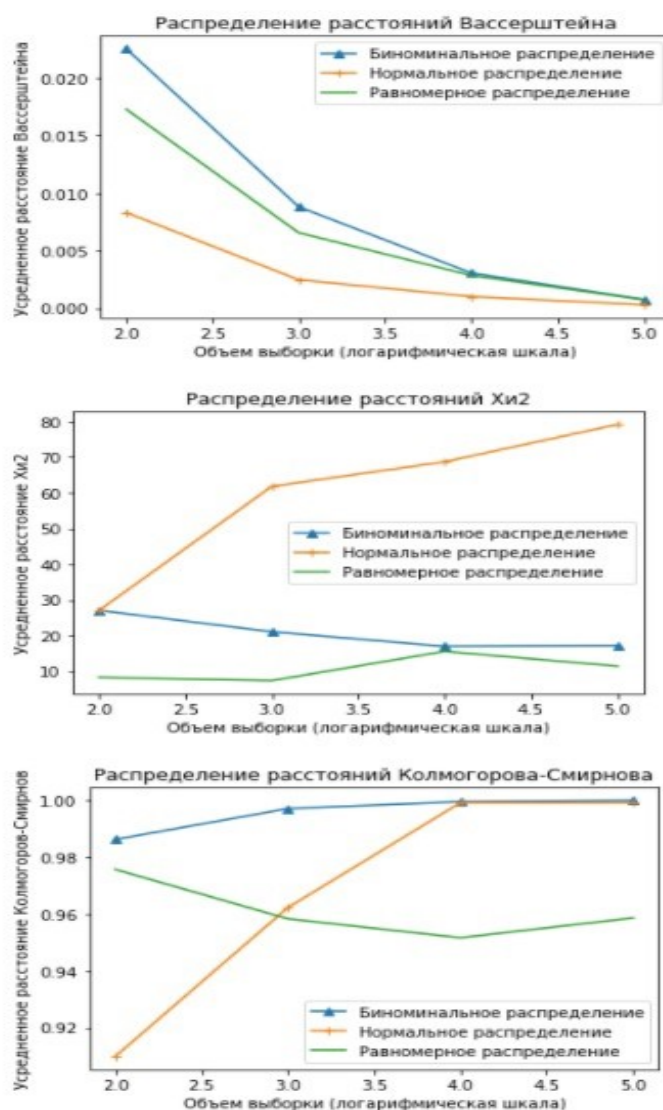


Рис. 1. График кривых для расстояний Вассерштейна, хи-квадрат и Колмогорова-Смирнова.

Заключение

В ходе проделанной работы был реализован алгоритм генерации случайных величин с заданным равномерным, нормальным и биномиальным распределениями вероятностей с использованием метода псевдонима. В результате исследования было установлено, что *alias method* позволяет достаточно эффективно генерировать случайные числа с заданным распределением, а именно - за время $O(1)$, в то время как на генерацию стандартными методами затрачивается время $O(\log n)$ или даже $O(n)$.

При этом, как и в стандартных методах, подготовка данных происходит за время $O(n)$, и используется $O(n)$ памяти. То есть, по указанным параметрам *alias method* также не уступает остальным.

Список использованных источников

1. Ермаков С. М., Михайлов Г. А. Статистическое моделирование. М.: Наука, 1982.
2. Ермаков С. М. Метод Монте-Карло в вычислительной математике: Вводный курс. С.-Пб.: Невский диалект; М.: БИНОМ. Лаборатория знаний, 2011.
3. Walker A. J. New fast method for generating discrete random numbers with arbitrary frequency distributions // *Electronic Letters*. 1974. Vol. 10. P. 127—128.
4. Walker A. J. An efficient method for generating discrete random variables with general distributions // *ACM Trans. Math. Software*. 1977. Vol. 3, No 3. P. 253—256.
5. Brown F. B., Martin W. R., Calaham D. A. A discrete sampling method for vectorized Monte Carlo calculations // *Trans. Am. Nucl. Soc.* 1981. Vol. 38. P. 354—355.