Московский государственный университет имени
М. В. Ломоносова
Факультет вычислительной математики и кибернетики
Кафедра системного анализа

Отчет по практикуму

«Стохастический анализ и моделирование»

Студент 415 группы
И. Р. Удовиченко

Руководитель практикума
к.ф.-м.н., доцент С. Н. Смирнов

Москва, 2020

Содержание

Задание 1

Пусть дана случайная величина $X \sim U[0,1]$, рассмотрим случайную величину $Y = \mathbb{1}(X < p)$. Тогда Y принимает значение 1 с вероятностью p. Это следует из свойства: $\mathbb{E}(\mathbb{1}_A) = \mathbb{P}(A)$.

Определение 1 *Биномиальным распределением случайной величины с параметрами n, p будем называть случайную величину, принмающую значения, равные числу успехов в n испытаниях Бернулли с вероятностью успеха p.*

Из определения следует способ генерирования биномиальной случайной величины:

$$X = \sum_{i=1}^{n} Y_i,$$

где случайные величины Y_i имеют распределение Бернулли с вероятностью успеха p.

Определение 2 *Геометрическим распределением с параметром p (и $q = 1 - p$) будем называть распределение числа неудач до 1-го успеха в схеме Бернулли.*

Функция вероятность геометрического распределения имеет вид: $\mathbb{P}(X = n) = q^n p,\ n \in \mathbb{N}_0$.

Пусть случайная величина $Y \sim exp(\lambda)$. Тогда

$$\mathbb{P}(n \leqslant Y < n+1) = e^{-\lambda n}\left(1 - e^{-\lambda}\right),$$

$$X = \lfloor Y \rfloor \sim geom(1 - e^{-\lambda}).$$

Тогда чтобы получить геометрическое распределение с параметром p надо взять целую часть от експоненциально распределенной случайной величины с параметром $\lambda = -\ln(1-p)$ Экспоненциальное распределение получим обращением функции распределения: если $Z \sim U[0,1]$, тогда $Y = -\frac{1}{\lambda}\ln(1 - Z) \sim exp(\lambda)$. Окончательно получаем $X = \left\lfloor \frac{\ln(1-Z)}{\ln(1-p)} \right\rfloor \sim geom(p)$.

Утверждение 1 *Для $X \sim geom(p)$ верно свойство отсутствия памяти: $\mathbb{P}(X \geqslant m + n \mid X \geqslant m) = \mathbb{P}(X \geqslant n)$.*

Доказательство.

$$\mathbb{P}(X \geqslant m+n \mid X \geqslant m) = \frac{\mathbb{P}(X \geqslant m+n)}{\mathbb{P}(X \geqslant m)} = \frac{q^{m+n}p/(1-q)}{q^m p/(1-q)} = q^n = \sum_{i=n}^{\infty} q^i p = \mathbb{P}(X \geqslant n).$$

■

Рассмотрим игру «орлянку» со случайной величиной X:

$$X_i = \begin{cases} 1 \text{ с вероятностью } p = 0{,}5, \\ -1 \text{ с вероятностью } 1-p = 0{,}5. \end{cases}$$

Пусть $S_n = \sum\limits_{i=1}^{n} X_i$ и $Y = S_i/\sqrt{n}$. Так как $\mathbb{E}S_n = 0$ и $var[S_n] = n$, то из центральной предельной теоремы следует, что $Y = S_i/\sqrt{n} \xrightarrow[n\to\infty]{d} \mathcal{N}(0,1)$.

Задание 2

Определение 3 *Канторовой лестницей нахывается функция f, которая строится следующим образом:*

1. $f(0) = 0,\ f(1) = 1.$
2. *Отрезок $[0,1]$ разбивается на 3 равных части, затем на средней части функция f полагается равной полусумме значений на концах.*
3. *для первого и третьего отрезка процедура повторяется рекурсивно.*

По определению канторова лестница обладает свойством фрактальности или самоподобия: $f(x) = 2f(x/3)$. Так как f монотонна и лежит в отрезке $[0,1]$, то существует случайная величина X (сингулярная), функция распределения которой является канторовой лестницей.

Используя свойство фрактальности данной случайной величины, найдем ее математическое ожидание и дисперсию:

$$X \stackrel{\mathrm{d}}{=} 1 - X \implies \mathbb{E}X = 0.$$

$$\mathbb{E}X^2 = \int\limits_0^1 x^2 dF(x) = \int\limits_0^{\frac{1}{3}} x^2 dF(x) + \int\limits_{\frac{2}{3}}^1 x^2 dF(x).$$

Отсюда используя замену $y = 3x$ и свойство $F(x/3) = F(x)/2$ получаем

$$\operatorname{var}[X] = \mathbb{E}X^2 = \frac{3}{8}.$$

Для проверки свойств распределения будем использовать критерии Колмогорова и Смирнова. В тесте Колмогорова проверяется гипотеза $H_0\colon F = F_0$ соответствия распределения некоторому наперед заданному распределению F_0. Для проверки гипотезы строится следующая статистика:

$$D_n = \sup_x \bigl|F_n(x) - F_0(x)\bigr|,$$

где $F_n(x)$ — эмпирическая функция распределения, потстроенная по выборке $X_1, \ldots, X_n$. Теорема Колмогорова утверждает, что $\sqrt{n}D_n \xrightarrow[n\to\infty]{d} K$, где K — распределение Колмогорова. Зафиксируем некоторый уровень доверия α и будем отклонять гипотезу, когда

$$\sqrt{n}D_n > K^{-1}(1-\alpha).$$

Для проверки гипотезы о принадлежности двух выборок размеров m и n к одному распределению используется следующая статистика:

$$D_{mn} = \sup_x \bigl|F_n(x) - G_m(x)\bigr|,$$

где F_n и G_m — эмпирические функции распределения. По теореме Смирнова

$$\sqrt{\frac{mn}{m+n}} D_{mn} \xrightarrow[n\to\infty]{d} K,$$

поэтому при больших размерах выборки $(m, n > 20)$ отклоняем гипотезу, если

$$\sqrt{\frac{mn}{m+n}} D_{mn} > K^{-1}(1-\alpha).$$

Задание 3

Определение 4 *Экспоненциальным распределением с параметром λ будем называть распределение с плотностью $p(x) = \lambda e^{-\lambda x}$.*

Утверждение 2 *Экспоненциальное распределение обладает свойством отсутствия памяти:*

$$\mathbb{P}(X \geqslant x + y \mid X \geqslant y) = \mathbb{P}(X \geqslant x).$$

Доказательство.

$$\mathbb{P}(X \geqslant x + y \mid X \geqslant y) = \frac{e^{-\lambda(x+y)}}{e^{-\lambda y}} = e^{-\lambda x} = \mathbb{P}(X \geqslant x).$$

■

Рассмотрим еще одно свойство экспоненциального распределения. Пусть $X_1, \ldots, X_n$ нещависимы и распределены экспоненциально с показателями $\lambda_1, \ldots \lambda_n$ соответственно. Пусть также $Y = \min\{X_1, \ldots, X_n\}$. Тогда

$$F_Y(x) = \mathbb{P}(Y < x) = 1 - \mathbb{P}\left(X_1 \geqslant x, \ldots, X_n \geqslant x\right) = 1 - \prod_{i=1}^{n} e^{-\lambda_i x} = 1 - e^{-x \sum\limits_{i=1}^{n} \lambda_i}.$$

Таким образом, Y распределена экспоненциально с показателем $\sum\limits_{i=1}^{n} \lambda_i$.

Рассмотрим два способа моделирования распределения Пуассона. Первый заключается в том, что промежутки между скачками пуассоновского процесса с параметром λ распределены экспоненциально, а случайная величина $X(1)$ имеет распределение Пуассона с параметром λ.

Второй способ дает

Теорема 1 (Пуассона) *Пусть в схеме серий испытаний Бернулли с вероятностями p_n выполнено:*

$$p_n \xrightarrow[n \to \infty]{} 0, \qquad np_n \xrightarrow[n \to \infty]{} \lambda > 0.$$

Тогда число успехов сходится к распределению Пуассона с параметром λ. При этом верна оценка:

$$\left| \mathbb{P}(S_n \in A) - \sum_{k \in A} e^{-\lambda} \frac{\lambda^k}{k!} \right| \leqslant \lambda p.$$

Чтобы получить вероятность, отличающуюся от распределения Пуассона не больше чем на ε, положим $p = \frac{\varepsilon}{\lambda}$ и $n = \frac{\lambda^2}{\varepsilon}$. Тогда биномиальное распределение с такими параметрами будет хорошо приближать распределение Пуассона.

Для моделирования стандартного нормального распределения рассмотрим пару случайных величин (X_1, X_2), которую будем рассматривать, как двумерный вектор. Пусть (r, φ) — запись вектора в полярных координатах. Из симметрии $\varphi \sim U[0, 2\pi]$. Из свойств нормального распределения $r^2 \sim exp(\frac{1}{2})$. Тогда моделируя угол равномерным распределением и квадрат радиуса экспоненциальным, будем получать 2 стандартных нормальных случайных величины после перехода в декартовы координаты.

Критерий χ^2 Пирсона используется для проверки гипотезы равенства распределения заданному $H_0\colon F = F_0$. Разобьем прямую на непересекающиеся части $\Delta_1, \ldots \Delta_m$ и положим $p_i = \mathbb{P}(X \in \Delta_i \mid H_0)$. Обозначим за N_i число элементов выборки, попавших в Δ_i. Тогда статистика

$$Q = \sum_{i=1}^{m} \frac{(N_i - np_i)^2}{np_i} \sim \chi^2_{m-1}.$$

Тогда отклоняем гипотезу, если $\chi^2_{m-1}(Q) < \alpha$.

Пусть дана выборка $X_1, \ldots X_n$ из нормального распределения с неизвестными параметрами. Для проверки гипотезы $H_0\colon \mathbb{E}X = \mu$ используется критерий Стьюдента. Положим

$$\hat{\sigma} = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2}.$$

Тогда статистика

$$U = \sqrt{n} \frac{\overline{X} - \mu}{\hat{\sigma}} \sim St_{n-1}.$$

Будем отклонять гипотезу, если $St_{n-1}(U) < \frac{\alpha}{2}$.

Для сравнения дисперсий двух нормально распределенных выборок используется критерий Фишера. Пусть даны две выборки $X_1, \ldots, X_n$ и $Y_1, \ldots Y_m$. Тогда

$$V = \frac{(m-1) \sum\limits_{i=1}^{n} (X_i - \overline{X})^2}{(n-1) \sum\limits_{i=1}^{m} (Y_i - \overline{Y})^2} \sim F_{m-1, n-1}.$$

Гипотеза отвергается, если $F_{m-1,n-1}(V) < \frac{\alpha}{2}$.

Задание 4

Список литературы