

Задачи к занятиям 10 марта

1 (Стандартный способ моделирования с.в. – метод обратной функции). Покажите, что если с.в. η равномерно распределена на отрезке $[0,1]$, то с.в. $\xi = F^{-1}(\eta)$ имеет функцию распределения $F(x)$. Предполагается, что $F(x)$ непрерывна и строго монотонна. Как выглядит формула для моделирования с.в. из показательного распределения с функцией распределения $F(x) = (1 - e^{-\lambda x})I\{x > 0\}$?

2 (Нелинейное преобразование). Пусть с.в. η_1, η_2 имеют равномерное распределение на отрезке $[0,1]$. Докажите, что с.в. X и Y : $X = \sqrt{-2 \ln \eta_1} \cos(2\pi\eta_2)$, $Y = \sqrt{-2 \ln \eta_1} \sin(2\pi\eta_2)$ – независимые и одинаково распределенные: стандартно нормально $\mathcal{N}(0,1)$.

Указание. По-сути нужно показать, что

$$f_{XY}(x, y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}.$$

Перейдите к полярным координатам (не забыв о якобиане замены переменных). Воспользуйтесь задачей 1.

Метод суперпозиции.

1) Разыгрывается значение дискретной с.в., принимающей значения $k = 0, 1, 2, \dots$ с вероятностями p_k .

2) Моделируется с.в. с функцией распределения $F_k(x)$ (например, методом обратной функции).

3 (Метод суперпозиции). Пусть ξ распределена на $[0,1]$ с плотностью $f_\xi(x)$, представимой в виде степенного ряда $\sum_{k=0}^{\infty} a_k x^k$ с $a_k \geq 0$. Положим $p_k = a_k / (k+1)$. Тогда

$$f_\xi(x) = \sum_{k=0}^{\infty} p_k \cdot (k+1)x^k. \text{ Примените метод суперпозиции для моделирования с.в. } \xi.$$

4 (теорема Бернштейна). а) С помощью неравенства Чебышёва установите следующий результат из анализа: $\forall f \in C[0,1] \rightarrow \|f_n - f\|_{C[0,1]} \xrightarrow{n \rightarrow \infty} 0$, $f_n(x) = \sum_{k=0}^n f\left(\frac{k}{n}\right) C_n^k x^k (1-x)^{n-k}$.

б) Исходя из предыдущей задачи и п. а) предложите способ генерирования распределения с.в. ξ , имеющей плотность $f_\xi(x)$ с финитным носителем, для определенности, пусть носителем будет отрезок $[0,1]$.

5. Как с помощью с.в. ξ , равномерно распределенной на отрезке $[0,1]$ ($\xi \in R[0;1]$), и симметричной монетки построить с.в. X , имеющую плотность распределения $f_X(x) = \frac{1}{4} \left(\frac{1}{\sqrt{x}} + \frac{1}{\sqrt{1-x}} \right)$, $x \in [0,1]$?

6 (Метод фон Неймана). Пусть с.в. ξ распределена на отрезке $[a,b]$, причем ее плотность распределения ограничена: $\max_{x \in [a;b]} f_\xi(x) = C < \infty$. Пусть с.в. η_1, η_2, \dots – независимы и рав-

номерно распределены на $[0,1]$, $X_i = a + (b-a)\eta_{2i-1}$, $Y_i = C\eta_{2i}$, $i = 1, 2, \dots$, т.е. пары (X_i, Y_i) независимы и равномерно распределены в прямоугольнике $[a,b] \times [0,C]$. Обозначим через ν номер первой точки с координатами (X_i, Y_i) , попавшей под график плотности $f_\xi(x)$, т.е. $\nu = \min\{i: Y_i \leq f_\xi(X_i)\}$. Положим $X_\nu = \sum_{n=1}^{\infty} X_n I\{\nu = n\}$.

а) Покажите, что с.в. X_ν распределена также как ξ .

б) Сколько в среднем точек (X_i, Y_i) потребуется «вбросить» в прямоугольник $[a,b] \times [0,C]$ для получения одного значения ξ ?

7 (метод Монте Карло для вычисления значения интеграла). а) Требуется вычислить с заданной точностью ε и с заданной доверительной вероятностью γ абсолютно сходящийся интеграл $J = \int_{[0,1]^m} f(\vec{x}) d\vec{x}$. Считайте, что $\forall \vec{x} \in [0,1]^m \rightarrow |f(\vec{x})| \leq 1$.

Пояснение. Введем случайный m -вектор $\vec{X} \in R([0,1]^m)$ и с.в. $\xi = f(\vec{X})$. Тогда

$M\xi = \int_{[0,1]^m} f(\vec{x}) d\vec{x} = J$. Поэтому получаем оценку интеграла $\bar{J}_n = \frac{1}{n} \sum_{k=1}^n f(\vec{x}^k)$, где \vec{x}^k ,

$k = 1, \dots, n$ – повторная выборка значений случайного вектора \vec{X} (т.е. все \vec{x}^k , $k = 1, \dots, n$ – независимы и одинаково распределены: также как и вектор \vec{X}). В задаче требуется оценить сверху число n ($n \gg m$), начиная с которого $P(|J - \bar{J}_n| \leq \varepsilon) \geq \gamma$.

б) Решите задачу из п. а) при дополнительном предположении липшецевости функции $f(\vec{x})$, разбив единичный куб на $n = N^m$ одинаковых кубиков со стороной $1/N$, и используя оценку $\bar{J}_n = \frac{1}{n} \sum_{k=1}^n f(\vec{x}^k)$, где \vec{x}^k – имеет равномерное распределение в k -м кубике.

в) **(метод выделения главной части, метод замены меры (метод существенной выборки), метод включения особенности в плотность)** Решите задачу п. а) не предполагая, что $f(\vec{x})$ – ограниченная функция на единичном кубе. Предложите способы уменьшения дисперсии полученной оценки интеграла. Как можно использовать информацию об особенностях функции $f(\vec{x})$?

8 (Алгоритм Кнута–Яо). С помощью бросаний симметричной монетки требуется сгенерировать распределение заданной дискретной с.в., принимающей конечное число значений. Обобщите описанную ниже схему на общий случай. Предположим, что нам нужно сгенерировать распределение с.в., принимающей три значения 1, 2, 3 с равными вероятностями $1/3$. Действуем таким образом. Два раза кидаем монетку: если выпало 00, то считаем что выпало значение 1, если 01, то 2, если 11, то 3. Если 10, то еще два раза кидаем монетку и повторяем рассуждения. Покажите, что можно сгенерировать распределение дискретной с.в., принимающей, вообще говоря, с разными вероятностями n значений в среднем с помощью не более чем $\log_2(n-1) + 2$ подбрасываний симметричной монетки.

9 (Markov Chain Monte Carlo (MCMC)). Чтобы построить однородный дискретный марковский процесс с конечным числом состояний, имеющий заданную инвариантную (стационарную) меру π , переходные вероятности ищутся в следующем виде: $p_{ij} = p_{ij}^0 a_{ij}$,

$i \neq j$; $p_{ii} = 1 - \sum_{j: j \neq i} p_{ij}$, где p_{ij}^0 – некоторая затравочная матрица, которую будем далее предполагать симметричной. Покажите, что матрица p_{ij} имеет инвариантную (стационарную) меру π , если

$$\frac{a_{ij}}{a_{ji}} = \frac{\pi_j p_{ji}^0}{\pi_i p_{ij}^0} = \frac{\pi_j}{\pi_i}.$$

Чтобы найти a_{ij} достаточно найти функцию $F: \mathbb{R}_+ \rightarrow [0, 1]$ такую, что

$$\frac{F(z)}{F(1/z)} = z \text{ и положить } a_{ij} = F\left(\frac{\pi_j p_{ji}^0}{\pi_i p_{ij}^0}\right) = F\left(\frac{\pi_j}{\pi_i}\right).$$

Пожалуй, самый известный пример такой функции $F(z) = \min(z, 1)$ определяет алгоритм Метрополиса.

10 (задача о наилучшем приближении). Предположим, что с.в. $X \in L_2$, это означает $MX^2 < \infty$. Докажите, что

$$\|X - M(X | Y_1, \dots, Y_n)\|_{L_2} = \min_{\varphi \in H} \|X - \varphi(Y_1, \dots, Y_n)\|_{L_2}, \quad (*)$$

где H – подпространство пространства L_2 всевозможных борелевских функций $\varphi(Y_1, \dots, Y_n) \in L_2$; $M(X | Y_1, \dots, Y_n)$ – условное математическое ожидание с.в. X относительно σ -алгебры порожденной с.в. Y_1, \dots, Y_n , часто говорят просто относительно с.в. Y_1, \dots, Y_n ;

$$\|X\|_{L_2} = \sqrt{(X, X)_{L_2}} = \sqrt{M(X \cdot X)} = \sqrt{M(X^2)}.$$

Пояснение. Утверждение задачи означает, что в гильбертовом пространстве $L_2(\Omega, \Xi, \mu)$ (со скалярным произведением $(X, Y) = M(X \cdot Y)$) всевозможных квадратично интегрируемых с.в. над заданным вероятностным пространством (Ω, Ξ, μ) , ограниченный линейный оператор условного математического ожидания $M(\cdot | Y_1, \dots, Y_n): L_2 \rightarrow L_2$ ($X \rightarrow E(M | Y_1, \dots, Y_n)$) есть ни что иное, как проекция $X \in L_2$ на подпространство H .

11 (задача о линейной регрессии). Докажите, что если в условиях предыдущей задачи $(X, Y_1, \dots, Y_n)^T$ – является нормальным случайным вектором (без ограничения общности можно также считать, что $(Y_1, \dots, Y_n)^T$ – невырожденный нормальный случайный вектор), то в качестве H можно взять подпространство всевозможных линейных комбинаций с.в. Y_1, \dots, Y_n . Т.е. мы можем более конкретно сказать на каком именно классе борелевских функций достигается минимум в (*).

Указание. Будем искать $M(X | Y_1, \dots, Y_n)$ в виде

$$M(X | Y_1, \dots, Y_n) = c_1 Y_1 + \dots + c_n Y_n.$$

Поскольку $(X, Y_1, \dots, Y_n)^T$ – нормальный случайный вектор, то из условий, которые как не трудно проверить (проверьте), однозначно определяют вектор коэффициентов $(c_1, \dots, c_n)^T$,

$$\forall k = 1, \dots, n \rightarrow (X - c_1 Y_1 - \dots - c_n Y_n, Y_k)_{L_2} = M((X - c_1 Y_1 - \dots - c_n Y_n) \cdot Y_k) = 0$$

следует (объясните почему), что

$$X - M(X | Y_1, \dots, Y_n) \text{ и } \varphi(Y_1, \dots, Y_n) \text{ – независимые с.в..}$$

Функция φ – произвольная борелевская функция из L_2 . Так как (поясните почему)

$$E(X - M(X|Y_1, \dots, Y_n)) = 0, \quad M((X - M(X|Y_1, \dots, Y_n)) \cdot \varphi(Y_1, \dots, Y_n)) = 0,$$

то $X - M(X|Y_1, \dots, Y_n)$ ортогонален подпространству Π пространства L_2 всевозможных борелевских функций $\varphi(Y_1, \dots, Y_n) \in L_2$. Следовательно,

$$M(X|Y_1, \dots, Y_n) = c_1 Y_1 + \dots + c_n Y_n$$

удовлетворяет условию (*).

12 (статистическая эргодическая теорема фон Неймана для динамических систем).

Пусть X – измеримое пространство, $T: X \rightarrow X: \forall B \in \sigma(X) \rightarrow \mu(T^{-1}(B)) = \mu(B)$. Тогда

$\forall f \in L_2(X) \rightarrow \frac{1}{n} \sum_{k=1}^n f(T^k(x)) \xrightarrow[n \rightarrow \infty]{L_2} M(f|\Theta)$, где Θ порождено функциями (случайными величинами) $\left\{ f: f(x) \stackrel{L_2}{=} f(T(x)) \right\}$.

Если $B \in \sigma(X): T^{-1}(B) \stackrel{L_2}{=} B \Rightarrow B = \{\emptyset\} \vee X$, то $M(f|\Theta) = M(f) = \int_X f(x) d\mu(x)$.

Не ограничивая общности, будем считать, что $\mu(X) = 1$. Построим стационарный в широком смысле случайный процесс $Y(k) = f(T^k x)$, где x – случайная величина, распределенная согласно мере μ . Верно ли, что

$$M(f|\Theta) = M(f) \Leftrightarrow \frac{1}{n^2} \sum_{i=1, j=1}^{n, n} R_Y(i, j) = \frac{1}{n^2} \sum_{i=1, j=1}^{n, n} M\left[\overset{\circ}{Y}(i) \overset{\circ}{Y}(j) \right] \xrightarrow[n \rightarrow \infty]{} 0?$$

13 (Г. Вейль). Пусть α – какое-то иррациональное число. Последовательность $x_k = \{\alpha k\}$.

а) С помощью предыдущей задачи, объясните почему для любой $f \in C[0, 1]$ выполняется:

$$\frac{1}{n} \sum_{k=1}^n f(x_k) \xrightarrow[n \rightarrow \infty]{} \int_0^1 f(x) dx.$$

б) Пусть функция, определенная в единичном квадрате на плоскости, имеет вид:

$$f(x, y) = \begin{cases} 1, & y \leq x \wedge x + y \leq 1 \\ 0, & \text{иначе} \end{cases}.$$

Верно ли, что

$$\frac{1}{n} \sum_{k=1}^n f(x_{2k-1}, x_{2k}) \xrightarrow[n \rightarrow \infty]{} \int_0^1 \int_0^1 f(x, y) dx dy?$$

14. Пусть $\vec{X} \in N(\vec{0}_n, I_n)$. Покажите, что $\vec{\xi} = \vec{X} / \|\vec{X}\| \in R(S^n(1))$, т.е. $\vec{\xi}$ имеет равномерное распределение на сфере в \mathbb{R}^n . Верно ли, что если мы рассматриваем группу $SO(n)$ поворотов единичной сферы в \mathbb{R}^n (всевозможных движений сферы, переводящих её в себя), то единственной инвариантной мерой μ ($\forall S \in SO(n), B \in \sigma(S^n(1)) \rightarrow \mu(S^{-1}B) = \mu(B)$) будет равномерная (лебегова) мера на сфере?

15. Решить методом Монте Карло линейное интегральное уравнение вида:

$$\varphi(x) = \int_G K(x, x') \varphi(x') dx' + f(x) = K\varphi(x) + f(x),$$

в предположении, что спектральный радиус линейного интегрального оператора K меньше 1.

Указание. В условиях задачи единственное решение интегрального уравнения представляется рядом Неймана, в котором можно ограничиться конечным числом слагаемых. Количество слагаемых определяется точностью, которую мы хотим получить. В фиксированной точке x_0 , имеем

$$\varphi(x_0) \approx f(x_0) + Kf(x_0) + \dots + K^l f(x_0). \quad (i)$$

Таким образом, задача сводится к приближенному вычислению слагаемых этого ряда, т.е. кратных интегралов типа $K^j f(x_0)$.

Выберем в G такую плотность $p(x) > 0$ и условную плотность $p(x|x') > 0$, что $\int_G p(x) dx = \int_G p(x|x') dx = 1$. Определим в G случайную траекторию $T_k = \{x_0 \rightarrow x_1 \rightarrow \dots \rightarrow x_k\}$, где x_0 – реализация с.в. ξ_0 с плотностью $p(x)$, x_j – реализация с.в. ξ_j с плотностью $p(x_j|x_{j-1})$, $j = 1, \dots, k$. Положим, $w_0 = 1$, $w_j = w_{j-1} K(x_{j-1}, x_j) / p(x_j|x_{j-1})$, $j = 1, \dots, k$; $v_j = w_j \frac{\varphi(x_0)}{p(x_0)} f(x_j)$, $j = 1, \dots, k$. Покажите, что $Mv_j = \langle \varphi, K^j f \rangle$, где $\langle \cdot, \cdot \rangle$ – означает скалярное произведение в гильбертовом пространстве $L_2(G)$.

Реализовав n случайных траекторий $(T_{k1}, T_{k2}, \dots, T_{kn})$, получаем оценку \tilde{y}_j скалярных произведений $\tilde{y}_j = \langle \varphi, K^j f \rangle$ согласно формулам

$$\tilde{y}_j = \frac{1}{n} \sum_{i=1}^n v_{ij}, \quad (ii)$$

где v_{ij} , $i = 1, \dots, n$ – реализация с.в. v_j на i -й траектории T_{ji} .

Положим $\varphi(x) = \delta(x - x_0)$. Тогда $\langle \delta(x - x_0), K^j f(x) \rangle = K^j f(x_0)$. Возьмем в качестве $p(x) = \delta(x - x_0)$, т.е. все траектории T_{ji} “стартуют” из фиксированной точки x_0 и $v_{j0} = w_{j0} f(x_j)$. Имеем: $Mv_{j0} = K^j f(x_0)$.

Обоснуйте, почему слагаемые в (i) можно оценивать формулой (ii), где v_{ij} , $i = 1, \dots, n$ – реализация с.в. v_j на i -й траектории T_{ji} с фиксированной точкой “старта” $x = x_0$ для всех траекторий.

16 (Hit and Run <http://cde.ru/node/320>, http://dame.mipt.ru/studyandscience/a_5g3xhb.html)*. Ряд задач, в которых используется метод Монте Карло, предполагает возможность случайно равномерно набрасывать точки в некоторое наперед заданное множество (не обязательно выпуклое и связное). Например, при вычислении интеграла с помощью метода Монте Карло или при численном решении задач оптимизации, в которых нужно уметь приближенно находить центр тяжести множества. Исходя из МСМС подхода обоснуйте, аккуратно оговорив детали, следующий способ генерации точек: Берем любую точку внутри множества и проводим случайно направление через эту точку, далее с помощью граничного оракула случайно генерируем (с помощью равномерного распределения) на этом направлении внутреннюю точку рассматриваемого множества. Через эту точку снова проводим случайное направление и т.д. Хорошо ли будет работать Hit and Run для вытянутых множеств или для множеств, имеющих достаточно острые углы? Предложите мо-

дификацию (например, с помощью эллипсоидов Дикина) алгоритма Hit and Run для таких “плохих” множеств.

Предложите другие способы случайно равномерно набрасывать точки в некоторое наперед заданное множество (например, Shake and Bake). Хорошо ли будет работать метода Shake and Bake для множеств в пространствах большой размерности?

Как следует действовать, если рассматриваемое множество имеет простую структуру: n -мерный куб, n -мерный шар, n -мерный симплекс, многогранник.

Ниже приводится исследовательская задача, решение которой будет подразумевать получение автомата.

17 (Markov Chain Monte Carlo Revolution)*.** В руки опытных криптографов попало закодированное письмо (10 000 символов). Чтобы это письмо прочитать нужно его декодировать. Для этого берется стохастическая матрица переходных вероятностей $P = \|p_{ij}\|$ (линейный размер которой определяется числом возможных символов (букв, знаков препинания и т.п.) в языке на котором до шифрования было написано письмо – этот язык известен и далее будет называться базовым), в которой p_{ij} – отвечает за вероятность появления символа с номером j сразу после символа под номером i . Такая матрица может быть идентифицирована с помощью статистического анализ какого-нибудь большого текста, скажем, Войны и мира Л.Н. Толстого.

Пускай способ (де)шифрования определяется некоторой, неизвестной, функцией \bar{f} – преобразование (перестановка) множества кодовых букв во множество символов базового языка.

В качестве, “начального приближения” выбирается какая-то функция f , например, полученная исходя из легко осуществимого частотного анализа. Далее рассчитывается вероятность выпадения полученного закодированного текста \vec{x} , сгенерированного при заданной функции f (функция правдоподобия):

$$L(\vec{x}; f) = \prod_k p_{f(x_k), f(x_{k+1})}.$$

Случайно выбираются два аргумента у функции f и значения функции при этих аргументах меняются местами. Если в результате получилась такая f^* , что $L(\vec{x}; f^*) \geq L(\vec{x}; f)$, то $f := f^*$, иначе независимо бросается монетка с вероятностью выпадения орла $p = L(\vec{x}; f^*) / L(\vec{x}; f)$, и если выпадает орёл, то $f := f^*$, иначе $f := f$. Далее процедура повторяется.

Объясните, почему предложенный алгоритм “сходится” именно к \bar{f} ? Почему сходимость оказывается такой быстрой (0.01 сек. на современном РС)?

Замечание. Приведенный алгоритм является частным случаем более общего алгоритма Метрополиса: для того чтобы сгенерировать требуемое дискретное распределение вероятностей (нормировка распределения на единицу – не существенна), строится неразложимая марковская цепь, с сильно разреженной матрицей переходных вероятностей, (единственным) стационарным распределением которой будет требуемое распределение. По эргодической теореме для марковских цепей, это означает, что после некоторого числа шагов построенной марковской цепи “получится” (в независимости от начального распределения) распределение вероятностей близкое к стационарному, т.е. требуемому. Поскольку матрица переходных вероятностей сильно разрежена, то эволюция, согласно описанной марковской динамике, вычислительна малозатратна. Осталось заметить, что при довольно естественных предположениях марковская цепь может крайне быстро “выходить” на свое стационарное распределение. Собственно, одним из революционных направлений последних десятилетий при построении эффективных алгоритмов стало (Mag-

kov Chain Monte Carlo Revolution) использование только что отмеченного факта (возможности потенциально быстрой сходимости в эргодической теореме для марковских процессов). В качестве другого нетривиального и интересного примера укажем вероятностные алгоритмы (работающие быстрее известных детерминированных) поиска центра тяжести выпуклого множества и его объема (одна из работ в этом направлении была удостоена премии Фалкерсона – аналога Нобелевской премии в области Computer Science).

Обратим также внимание, что школьники могли познакомиться с этой темой на примере задачи тасования карт (сколько раз надо “перемешивать” колоду, чтобы все возможные перестановки были практически равновероятны) из выступления Андрея Окунькова в ЛШСМ-2010.

Для лучшего понимания происходящего в условиях задачи, отметим, что одним из самых универсальных способов получения асимптотически наилучших оценок неизвестных параметров по выборке является метод наибольшего правдоподобия. Напомним вкратце в чем он заключается. Пусть имеется выборка из распределения, зависящего от неизвестного параметра – в нашем случае выборкой \vec{x} из 10 000 элементов будет письмо, а неизвестным “параметром” будет функция f (хотя наш параметр может принимать много значений, но все-таки конечное число). Далее считается вероятность (или плотность вероятности в случае непрерывных распределений) $L(\vec{x}; f)$ того что выпадет данный \vec{x} при условии, что значения параметра f . Если посмотреть на распределение $L(\vec{x}; f)$, как на распределение в пространстве параметров (\vec{x} – зафиксирован), то при большом объеме выборки (размерности \vec{x}) при естественных условиях это распределение концентрируется в малой окрестности наиболее вероятного значения

$$f(\vec{x}) = \arg \max_f L(\vec{x}; f),$$

которое “асимптотически” совпадает с неизвестным истинным значением \bar{f} .

ПРИЛОЖЕНИЕ

ПРЕДВАРИТЕЛЬНЫЕ СВЕДЕНИЯ ИЗ ТЕОРИИ МАРКОВСКИХ ПРОЦЕССОВ

В настоящем приложении мы приводим необходимую в дальнейшем информацию о том, что такое *марковский случайный процесс* и какими свойствами он обладает [1, 2]. При этом нужно отметить колоссальную практическую значимость таких процессов. Приложения теории марковских процессов далеко не исчерпываются рассмотренными в данной статье. Из известного продвинутым школьникам, например, можно напомнить, что с помощью марковских процессов эффективно численно решаются разнообразные уравнения [3] и задачи типа “задачи о разборчивой невесте” [4]. Из не очень известного можно, например, отметить вероятностные алгоритмы (часто базирующиеся на специально подобранных по данной задаче марковских процессах) [5], которые намного эффективнее детерминированных решают разнообразные задачи (скажем, поиск центра тяжести выпуклого множества и его объема). Однако за эту эффективность приходится платить тем, что с маленькой, но ненулевой вероятностью, алгоритм может дать неверный ответ.

Итак, перейдем к описанию простейшего марковского процесса: конечной (число состояний конечно) однородной (по времени) марковской цепи (время дискретно, если время непрерывно, то вместо слова “цепь” говорят “процесс”).

Пусть некоторая система может находиться в одном из n состояний. В начальный момент времени положение системы характеризуется вектором распределения вероятностей:

$$\vec{p}^T(0) = (p_1(0), \dots, p_k(0), \dots, p_n(0)), \quad \sum_{k=1}^n p_k(0) = 1,$$

где $p_k(0) \geq 0$ – вероятность того, что система находилась в момент времени $t=0$ в состоянии с номером k . Пусть система эволюционирует во времени согласно *матрице переходных вероятностей* $P = \|p_{ij}\|_{i,j=1,1}^{n,n}$. При этом $0 \leq p_{ij} \leq 1$ – вероятность того, что система

перейдет за один шаг из состояния i в состояние j . Очевидно, что $\sum_{j=1}^n p_{ij} = 1, i = 1, \dots, n$.

Такие матрицы также называют *стохастическими матрицами*. Таким образом, распределение вероятностей в момент времени $t=1$ можно найти по формулам (*полной вероятности*):

$$p_k(1) = \sum_{i=1}^n p_i(0) p_{ik}, \quad k = 1, \dots, n.$$

Аналогично на любом шаге $(m+1) \in N$:

$$p_k(m+1) = \sum_{i=1}^n p_i(m) p_{ik}, \quad k = 1, \dots, n.$$

Эти соотношения можно переписать более компактно:

$$\vec{p}^T(m+1) = \vec{p}^T(m)P \quad (\text{уравнение Колмогорова–Чэмпена}).$$

Осуществим предельный переход к непрерывному времени (скейлинг): шаг по времени $\Delta t \rightarrow 0+$ и $p_{ij} := \lambda_{ij} \Delta t, i \neq j$. Тогда уравнение Колмогорова–Чэмпена перейдет в уравнение:

$$\frac{d\vec{p}^T(t)}{dt} = \vec{p}^T(t)\Lambda \quad (\text{уравнение Колмогорова–Феллера}),$$

где $[\Lambda]_{ij} = \lambda_{ij}$ при $i \neq j$ (интенсивности переходов) и $[\Lambda]_{ii} = -\sum_{j:j \neq i} \lambda_{ij}, i = 1, \dots, n$. Матрицу Λ

называют инфинитезимальной матрицей.

Таким образом для того чтобы задать конечный однородный марковский процесс (цепь) нужно задать начальное распределение вероятностей и инфинитезимальную мат-

рицу (или матрицу переходных вероятностей в случае дискретного времени). В следующем пункте будут рассмотрены конкретные примеры. А сейчас поставим основной вопрос: что будет происходить с марковским процессом на больших временах? Ответ подсказывают выписанные уравнения динамики: $\lim_{m \rightarrow \infty} \vec{p}(m) = \vec{p}^*$, где $\vec{p}^{*T} = \vec{p}^{*T} P$ (в дискретном времени) и $\lim_{t \rightarrow \infty} \vec{p}(t) = \vec{p}^*$, где $\vec{p}^{*T} \Lambda = \vec{0}^T$ (в непрерывном времени). Так определяемый вектор распределения вероятностей \vec{p}^* называется *стационарным (инвариантным) распределением*. Предельное распределение также называют *финальным распределением*. Выше лишь сказано, что финальные распределения следует искать среди стационарных.

Но, во-первых, где гарантии, что предел существует? Возьмем марковскую цепь с матрицей переходных вероятностей:

$$P = \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} \text{ и начальным состоянием } \vec{p}^T(0) = (1 \ 0)^T.$$

Очевидно, что $\vec{p}^T(m) = 0.5 \cdot (1 + (-1)^m, 1 + (-1)^{m+1})^T$ предела в обычном смысле не имеет.

Легко при этом проверить, что стационарное распределение $\vec{p}^{*T} = (0.5, 0.5)^T$ (заметим, что стационарное распределение всегда существует) будет пределом $\vec{p}^T(m)$ по Чезаро (см. ниже).

Во-вторых, при каких условиях стационарное распределение единственно? Ведь если взять, например, в качестве матрицы переходных вероятностей единичную матрицу, то линейно независимых стационарных состояний будет столько сколько состояний у цепи.

Чтобы ответить на поставленные вопросы напомним следующий простой факт: монотонное преобразование отрезка в себя имеет неподвижную точку. Причем это утверждение можно доказать с помощью *принципа сжимающих отображений*. Марковская динамика – это монотонная динамика в том смысле, что если $\vec{x} \geq \vec{y} \geq \vec{0}$ (покомпонентно, следовательно, “ \geq ” задает лишь частичный порядок), то $\vec{x}^T P \geq \vec{y}^T P$ (это следует из того, что все элементы матрицы P неотрицательные). Поэтому возникает желание воспользоваться одним из основных инструментов прикладного функционального анализа “принципом сжимающих отображений”, чтобы исследовать на устойчивость стационарные распределения (неподвижные точки марковской динамики). Далее осуществляется этот план.

Определение 1. Множество X называется метрическим пространством, если на парах его элементов задано расстояние (метрика) $\rho(x, y)$, удовлетворяющая условиям (аксиомам):

1. $\rho(x, y) = 0 \Leftrightarrow x = y$;
2. $\forall x, y \in X \rightarrow \rho(x, y) = \rho(y, x)$;
3. $\forall x, y, z \in X \rightarrow \rho(x, z) \leq \rho(x, y) + \rho(y, z)$.

Определение 2. Последовательность $\{x_n\}_{n=1}^{\infty} \in X$ называется фундаментальной или последовательностью Коши, если

$$\forall \varepsilon > 0 \exists N_\varepsilon > 0: \forall n, m > N_\varepsilon \rightarrow \rho(x_n, x_m) < \varepsilon.$$

Простым упражнением является показать, что любая сходящаяся последовательность фундаментальна. Обратное верно не всегда. Например, $X = [0, 1)$, $\rho(x, y) = |x - y|$. Последовательность $x_n = 1 - 1/n$, $n \in \mathbb{N}$ фундаментальная, но не сходится в X . Точнее говоря, она сходится в обычном смысле к 1, но $1 \notin X$.

Определение 3. Метрическое пространство называется полным, если любая фундаментальная последовательность в нем сходится.

В действительности, любое метрическое пространство может быть “пополнено”, подобно тому, как можно пополнить $X = [0, 1)$, $\rho(x, y) = |x - y|$, добавив 1.

Теорема (принцип сжимающих отображений, монотонные операторы, эргодическая теорема для конечных однородных марковских цепей [6]).

а) (принцип сжимающих отображений) Если оператор (вообще говоря, нелинейный) A действует в полном метрическом пространстве X и

$$\exists k \in \mathbb{N}: \forall x, y \in X \rightarrow \rho(A^k(x), A^k(y)) \leq \theta \rho(x, y), \theta \in (0, 1), \text{ то}$$

$$\exists! x^* \in X: A(x^*) = x^* \text{ и } \forall x \in X \rightarrow \rho(A^n(x), x^*) = O(\theta^{n/k}).$$

б) (монотонные операторы) Пусть $X = \mathbb{P}\mathbb{R}_{++}^n$ – множество лучей пространства \mathbb{R}^n , лежащих во внутренней неотрицательного ортанта, на котором введена метрика Биркгофа:

$$\rho(x, y) = \ln \min \left\{ \frac{\beta}{\alpha} : \alpha x \leq y \leq \beta x \right\} = \ln \min_{i,j=1,\dots,n} \frac{x_i y_j}{x_j y_i}.$$

Здесь под элементами x и y в левой части равенства понимаются лучи, а вот в правой части уже какие-то векторы, лежащие на соответствующих лучах. Какие именно векторы – не важно.

Тогда X – полное метрическое пространство. И если линейный оператор $A: X \rightarrow X$

($A = \|a_{ij}\|_{i,j=1}^{n,n}$ – матрица $n \times n$) положительный, т.е. $\forall i, j = 1, \dots, n \rightarrow a_{ij} > 0$, то

$$\exists \theta \in (0, 1): \forall x, y \in X \rightarrow \rho(Ax, Ay) \leq \theta \rho(x, y).$$

в) (стохастический вариант теоремы Фробениуса-Перрона, или эргодическая теорема для конечных однородных марковских цепей) Следующие условия для стохастической матрицы $n \times n$ P равносильны:

1) $\exists m_0 \in \mathbb{N}: P^{m_0} = \|p_{ij}(m_0)\|_{i,j=1}^{n,n} > 0$, т.е. $\forall i, j = 1, \dots, n \rightarrow p_{ij}(m_0) > 0$;

2) $i)$ У системы

$$\vec{p}^{*T} = \vec{p}^{*T} P, \sum_{k=1}^n p_k^* = 1. \quad (S)$$

существует притом единственно решение \vec{p}^* . $ii)$ Причем $\vec{p}^* > \vec{0}$.

$$iii) \forall \vec{p}(0) \geq \vec{0} \left(\sum_{k=1}^n p_k(0) = 1 \right) \rightarrow \lim_{m \rightarrow \infty} \vec{p}^T(m) = \lim_{m \rightarrow \infty} \vec{p}^T(0) P^m = \vec{p}^{*T},$$

другими словами,

$$\lim_{m \rightarrow \infty} P^m = \underbrace{\left[\vec{p}^*, \vec{p}^*, \dots, \vec{p}^* \right]}_n^T.$$

Заметим также, что условия 1), 2) пункта в) теоремы равносильны следующим требованиям: конечная однородная марковская цепь с матрицей переходных вероятностей P – неразложимая = неприводимая (т.е. из произвольного состояния «можно прийти» в любое, наперед заданное) и непериодическая

$$\text{Н.О.Д.} \left\{ k : \left[P^k \right]_{11} > 0 \right\} = 1.$$

Например, для цепи с матрицей переходных вероятностей

$$P = \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix}$$

период равняется 2. Если убрать условие неперериодичности, то для неразложимой цепи с периодом d будет иметь место сходимость средних временных (сходимость по Чезаро):

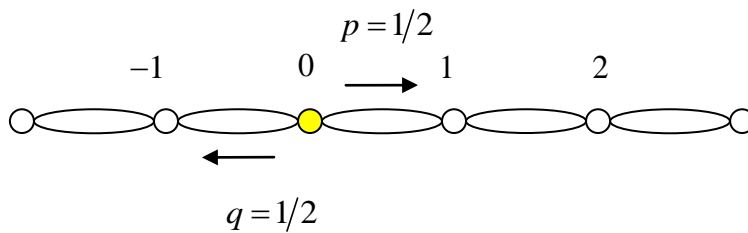
$$\exists! \bar{p}^* > \vec{0} \ (\bar{p}^* \in S): \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{m=1}^N P^m = \lim_{m \rightarrow \infty} \frac{1}{d} \sum_{k=1}^d P^{m+k-1} = \underbrace{[\bar{p}^*, \bar{p}^*, \dots, \bar{p}^*]}_n^T.$$

Если перейти к непрерывному времени, осуществляя соответствующий скейлинг, то легко показать, что необходимость в условии неперериодичности исчезает.

Если же цепь разложима, то система (S), вообще говоря, уже будет разрешима не единственным образом (аналогично и в непрерывном времени). Финальное распределение существует, но уже может зависеть от того, с какого распределения стартуем. Кроме того, не обязательно будет верно условие

$$\bar{p}^* > \vec{0}.$$

В заключение заметим, что эргодическая теорема переносится и на марковские процессы со счетным числом состояний с одним лишь нюансом: финальное распределение = стационарному распределению уже не обязательно будет распределением вероятностей. Хорошо это можно продемонстрировать на примере симметричного случайного блуждания на бесконечной прямой. Эта марковская цепь имеет единственное финальное распределение (хотя цепь и периодическая, зато неразложимая), которое совпадает с единственным стационарным распределением, в котором все вероятности равны нулю, т.е. не являющимся распределением вероятностей.



Кстати говоря, симметричное случайное блуждание на прямой обладает интересным свойством: стартуя из любого состояния, мы с вероятностью 1 вернемся в него за конечное число шагов (аналогичным свойством обладает случайное блуждание на плоскости, но вот уже в пространстве с отличной от нуля вероятностью, случайно симметрично гуляя, мы уже никогда не вернемся в точку старта), при этом математическое ожидание (т.е. среднее значение) времени возвращения равно бесконечности.

Упражнение. Дана однородная марковская цепь, которая может находиться в 3-ех состояниях. Вероятность перейти за один шаг из состояния 1 в состояние 2 равна $2/3$, а в состояние 3 – 0. Вероятность перейти за один шаг из состояния 2 в состояние 1 равна $1/2$, а в состояние 3 – $1/2$. Вероятность перейти за один шаг из состояния 3 в состояние 1 равна $2/3$, а в состояние 2 – $1/3$.

1. Выписать матрицу P переходных вероятностей за один шаг.
2. Определите стационарное распределение, финальное распределение и элементы матрицы $\lim_{m \rightarrow \infty} P^m$.
3. Найдите математическое ожидание относительного времени нахождения цепи в состояниях: а) 1; б) 2; в) 3, считая, что $\bar{p}^T(0) = (1/4, 1/2, 1/4)$. Зависит ли ответ от $\bar{p}^T(0)$?
4. Найдите математическое ожидание времени возвращения в состояния: а) 1; б) 2; в) 3.

Литература

1. *Кельберт М. Я., Сухов Ю. М.* Вероятность и статистика в примерах и задачах. Т. 2. М.: МЦНМО, 2010.
2. *Ширяев А.Н.* Вероятность 1, 2. М.: МЦНМО, 2007.
3. *Сосинский А. Б.* Мыльные плёнки и случайные блуждания. М.: МЦНМО, Библиотека “Математ. просвещение”, вып. 6. 2000.
4. *Гусейн-Заде С. М.* Разборчивая невеста. М.: МЦНМО, Библиотека “Математ. просвещение”, вып. 25. 2003.
5. *Motwani R., Raghavan P.* Randomized algorithms. Cambridge Univ. Press, 1995.
6. *Красносельский М. А., Лифшиц Е. А., Соболев А. В.* Позитивные линейные системы. Метод положительных операторов. М.: Наука, 1985.