УДК 519.21; 519.61

Распределения числа состояний в двоичных марковских стохастических моделях*

Л.Я. Савельев[1,2]

[1]Новосибирский государственный университет, ул. Пирогова, 2, Новосибирск, 630090

[2]Институт математики им. С.Л. Соболева Сибирского отделения Российской академии наук, просп. Акад. В.А. Коптюга, 4, Новосибирск, 630090

E-mail: savelev@math.nsc.ru

Савельев Л.Я. Распределения числа состояний в двоичных марковских стохастических моделях // Сиб. журн. вычисл. математики / РАН. Сиб. отд-ние. — Новосибирск, 2015. — Т. 18, № 2. — С. 191–200.

В статье выводятся точные и приближенные формулы для распределения, средних значений и дисперсий числа единиц на отрезках двоичных марковских последовательностей. Предлагаются различные способы вычислений по этим формулам. Даются оценки погрешностей. Приводится пример вычислений для двоичной марковской модели процесса выпадения осадков.

DOI: 10.15372/SJNM20150207

Ключевые слова: *стохастическая модель, двоичная марковская цепь, распределение, производящая функция, среднее значение, дисперсия.*

Saveliev L.Ya. Calculation of the number of states in binary Markov stochastic models // Siberian J. Num. Math. / Sib. Branch of Russ. Acad. of Sci. — Novosibirsk, 2015. — Vol. 18, № 2. — P. 191–200.

This paper derives exact and approximate formulas for the distribution, average values and variances of the number of units on the segments of binary Markov sequences. Various ways to calculate these formulas are proposed. Estimates of the errors are given. An example of the calculation for a binary Markov model of the precipitation process is presented.

Keywords: *stochastic model, binary Markov chain, distribution, generating function, mean, variance.*

1. Конечные марковские цепи

В теории вероятностей и математической статистике разработаны марковские модели с любым числом состояний. Но использование моделей с большим числом состояний связано с большими объемами данных и вычислений. Даже применение современных компьютеров и программ не всегда позволяет с ними справиться. Главными недостатками моделей с большим числом состояний является трудность оценки вероятностей переходов из одних состояний в другие. Это существенно снижает адекватность моделей с большим числом состояний. Часто для надежной интерпретации результатов приходится группировать состояния, получая те же результаты, что и при изначальном использовании более надежной модели с небольшим числом состояний.

Термины *состояние* и *момент* в теории имеют очень общий смысл, и поэтому марковская модель может описывать широкий круг явлений и процессов любой природы. Этим объясняется широкое применение стохастических марковских моделей в самых разных областях. Исключительно важным свойством этих моделей является возможность

*Работа выполнена при поддержке РФФИ (проект № 13-01-00275), гранта Минобрнауки № 2379.

© Савельев Л.Я., 2015

достаточно адекватно прогнозировать поведение рассматриваемого процесса. Марковские прогнозы часто хорошо согласуются с реальностью. При необходимости учитывать изменение условий используются модели с изменяющимися по выбранному закону переходными вероятностями.

Рассмотрим последовательность ξ случайных переменных $\xi(t)$, $t = 1, 2, \ldots$, с конечным множеством значений $C = \{1, 2, \ldots, m\}$. Последовательность ξ называется *марковской цепью*, если она обладает *марковским свойством*: при известном настоящем $\xi(t)$ прошлое $\{\xi(s),\ s < t\}$ и будущее $\{\xi(u),\ u > t\}$ стохастически независимы. Формально такая марковская цепь определяется строкой $A = \{a_i\}(i = 1, \ldots, m)$ *начальных* вероятностей $a_i = \Pr\left[\xi(0) = i\right]$ и матрицы $Q = \{q_{ij}\}$ *переходных* вероятностей $q_{ij} = \Pr\left[\xi(t+1)j \;\middle|\; \xi(t) = i\right]$. Распределение случайной величины $\xi(t)$ выражает строка $P^t{=}\{p_i^t\}$, составленная из вероятностей $p_i^t = \Pr\left[\xi(t) = i\right]$ того, что в момент t значением случайной величины $\xi(t)$ является номер i. Верны равенства:

$$P^{t+1} = P^t Q, \qquad p_j^{t+1} = \sum_{i=1}^{m} p_i^t\, q_{ij} \quad (j = 1, \ldots, m).$$

Переход через k моментов от t к $t{+}k$ описывается матричным равенством $P^{t+k}{=}P^t Q^k$. В частности, распределение случайной величины $\xi(u)$ выражается равенством $P^u{=}AQ^u$. При $m = 2, 3$ эти распределения легко вычисляются в общем виде [5]. Получены также общие формулы для их важных характеристик, связанных со структурой серий в реализациях таких марковских последовательностей. Подробно конечные марковские последовательности описаны в [2].

2. Двоичная марковская последовательность

Так называется последовательность ξ случайных величин $\xi(t)$, $t \ge 0$, с множеством значений $C = \{1, 0\}$, начальной $1{\times}2$ матрицей A и переходной $2{\times}2$ матрицей Q:

$$A = \{a_1, a_0\} = \{a, 1-a\}, \qquad Q = \begin{pmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{pmatrix} = \begin{pmatrix} p & 1-p \\ 1-q & q \end{pmatrix}.$$

Здесь $a_i = \Pr\left[\xi(0) = i\right]$ — вероятность появления в начальный момент состояния i, а $q_{ij} = \Pr[\xi(t+1)j|\xi(t) = i]$ — вероятность появления в момент $t+1$ состояния j при условии, что в момент t появилось состояние $i\,(0 \le i,\ j \le 1)$. Будем использовать более простые обозначения a, p, q и $d = p + q - 1$, $b = (1-q)/(1-d)$.

По индукции нетрудно доказать, что верно следующее выражение для t-й степени Q^t матрицы Q:

$$Q^t = \begin{pmatrix} b + (1-b)\, d^t & 1 - b - (1-b)\, d^t \\ b - b\, d^t & 1 - b + b\, d^t \end{pmatrix}.$$

Из матричного равенства $p^u = AQ^u$ следуют формулы

$$p_1^t = \Pr\left[\xi(t) = 1\right] = b + (a-b)\, d^t, \qquad p_0^t = \Pr\left[\xi(t) = 0\right] = 1 - b + (a-b)\, d^t$$

для вероятностей появления в момент t состояний $j = 1, 0$. Если $|d| < 1$, то при достаточно больших значениях t верны равенства $p_1^t \approx b$, $p_0^t = 1 - b$.

Постепенно последовательность $\xi(t)$ стабилизируется. Распределение случайной величины $\xi(t)$ все ближе к $\{b, 1-b\}$ и все меньше зависит от начального распределения $\{a, 1-a\}$.

Распределение $\{b, 1-b\}$ *стационарно*: $\{b, 1-b\}Q = \{b, 1-b\}$. Если в некоторый момент t распределение случайной величины $\xi(t)$ оказывается равным $\{b, 1-b\}$, то это распределение имеют и все следующие случайные величины $\xi(u)$, $u \geq 0$. Процесс переходит в стационарный режим. Чем меньше $|d| < 1$, тем быстрее убывает $|d|^t$ и быстрее достигается стационарный режим.

Впервые двоичные марковские последовательности были описаны в пионерской работе А.А. Маркова [3].

3. Распределение числа единиц

Формулы для этого распределения можно вывести из общих формул для различных совместных распределений [4, 5]. Но проще получить нужные результаты и в нужной форме непосредственно.

3.1. Рассмотрим случайную величину

$$x[n] = \sum_{k=0}^{n} \xi(k).$$

Выведем рекуррентные уравнения для вероятностей значений $x[n]$. Пусть

$$p_1(i,n)=Pr\big(x[n]=i, \xi(n)=1\big); \qquad p_0(i,n)=Pr\big(x[n]=i, \xi(n)=0\big); \qquad p(i,n)=Pr\big(x[n]=i\big).$$

Используя формулу полной вероятности и марковское свойство, получаем систему уравнений:

$$\begin{aligned} p_1(i,n) &= p_1(i-1,n-1)p + p_0(i-1,n-1)(1-q) \quad && (i>0, n>0), \\ p_0(i,n) &= p_1(i,n-1)(1-p) + p_0(i,n-1)q && (i>=0, n>0). \end{aligned} \tag{1}$$

Кроме того (граничные условия):

$$\begin{aligned} &p_1(1,0) = a, && p_0(0,0) = 1-a, \\ &p_1(i,0) = 0 \ (i \neq 1), && p_0(i,0) = 0 \ (i \neq 0), \\ &p_1(i,n) = 0 \ (i > n+1), && p_0(0,n) = (1-a)q^n. \end{aligned}$$

3.2. Введем производящие функции:

$$f_1(s,v)=\sum_{i=0}^{\infty}\sum_{n=0}^{\infty} p_1(i,n)s^i v^n, \quad f_0(s,v)=\sum_{i=0}^{\infty}\sum_{n=0}^{\infty} p_0(i,n)s^i v^n, \quad f=f_1+f_0, \quad F=(f_1, f_0).$$

Используя систему (1), получаем

$$(1-psv)f_1 - (1-q)svf_0 = as, \qquad -(1-p)vf_1 + (1-qv)f_0 = 1-a. \tag{2}$$

Решая систему (2) и складывая результаты, находим

$$f_1(s,v) = \frac{sv(-a-q+1)+as}{dsv^2 - v(ps+q)+1}, \tag{3}$$

$$f_0(s,v) = \frac{1-a+(a-p)sv}{dsv^2 - v(ps+q)+1}, \tag{4}$$

$$f(s,v) = \frac{1-a+as-dsv}{dsv^2 - v(ps+q)+1}. \tag{5}$$

3.3. Разложим правую часть равенства (3) в ряд и вычислим коэффициенты при степенях v. Используя формулу для суммы геометрической прогрессии и дважды применяя биномиальную формулу, получаем

$$\sum_{m=0}^{\infty}\sum_{l=0}^{m}\sum_{k=0}^{l}\binom{l}{k}\binom{m}{l}\Big(S_1(k,l,m)v^{2m-l}+S_2(k,l,m)v^{-l+2m+1}\Big),$$

где

$$S_1(k,l,m)=(-1)^{m-l}aq^kd^{m-l}p^{l-k}s^{-k+m+1},$$
$$S_2(k,l,m)=(-1)^{m-l}(-a-q+1)q^kd^{m-l}p^{l-k}s^{-k+m+1}.$$

Так как при $l>m$ или $k>l$ произведения биномиальных коэффициентов в суммах равны 0, то можно заменить конечные суммы бесконечными и переставлять их, когда нужно. Переменную v можно считать достаточно малой, чтобы обеспечить суммируемость рассматриваемых рядов. Произведя замену $n=2m-l$ в первой сумме и $n=2m-l+1$ во второй, получаем

$$f_1(s,v)=\sum_{n=0}^{\infty}g_1(s,n)v^n,$$

$$g_1(s,n)=\sum_{m=0}^{\infty}\sum_{k=0}^{\infty}\big(T_1(k,m,n)+T_2(k,m,n)\big)s^{-k+m+1},$$

$$T_1(k,m,n)=\binom{m}{2m-n}\binom{2m-n}{k}aq^k(-d)^{n-m}p^{-k+2m-n},$$

$$T_2(k,m,n)=\binom{m}{2m-n+1}\binom{2m-n+1}{k}(-a-q+1)q^k(-d)^{-m+n-1}p^{-k+2m-n+1}.$$

Сделав замену $i=m-k+1$ и вычислив коэффициенты при степенях s, находим

$$p_1(i,n)=\sum_{m=0}^{\infty}U_1(i,m,n)(-d)^{-m+n-1}q^{-i+m+1}p^{i+m-n-1}, \tag{6}$$

$$U_1(i,m,n)=\binom{m}{2m-n+1}\binom{2m-n+1}{-i+m+1}p(-a-q+1)-\binom{m}{2m-n}\binom{2m-n}{-i+m+1}ad.$$

По аналогии из равенства (4) получаем

$$p_0(i,n)=\sum_{m=0}^{\infty}U_0(i,m,n)(-d)^{-m+n-1}q^{-i+m-1}p^{i+m-n}, \tag{7}$$

$$U_0(i,m,n)=\binom{m}{2m-n+1}\binom{2m-n+1}{-i+m+1}(a-p)q-\binom{m}{2m-n}\binom{2m-n}{-i+m}(1-a)d.$$

Складывая (6) и (7), находим

$$p(i,n)=\sum_{m=0}^{\infty}U(i,m,n)(-d)^{-m+n}q^{-i+m}p^{i+m-n-1}, \tag{8}$$

$$U(i,m,n)=\binom{m}{2m-n+1}\binom{2m-n+1}{-i+m+1}pq+\binom{m}{2m-n}\binom{2m-n}{-i+m+1}aq+$$
$$\binom{m}{2m-n}\binom{2m-n}{-i+m}(1-a)p.$$

Замечание 1. Можно выразить найденные вероятности значений $x(n)$ через гипергеометрические функции:

$$p(i,n) = \frac{{}_3\tilde{F}_2\big(1,1,-n;2-i,i-n;pq/d\big)}{\Gamma(n+1)}\,\frac{aq}{p} + \frac{{}_3\tilde{F}_2\big(1,1,1-n;2-i,i-n+1;pq/d\big)}{\Gamma(n)}\,q + \frac{{}_3\tilde{F}_2\big(1,1,-n;1-i,i-n+1;pq/d\big)}{\Gamma(n+1)}\,(1-a).$$

Эти равенства позволяют рассматривать аналитические продолжения рассматриваемых вероятностей и интегральные представления для них [6].

3.4. Приведем несколько формул, удобных для вычисления вероятностей $p(i,n) = Pr\big(x(n) = i\big)$. В этих формулах ряды заменены конечными суммами. Заметим, что $2m - n = m - (m-n) > m$ при $m > n$ и поэтому соответствующие произведения биномиальных коэффициентов в слагаемых суммы (8) равны 0. То же самое и при $m < i-1$ или $m < n-i$: тогда $m - i + 1 < 0$ или $2m - n - (m-i) = m - (n-i) < 0$. Поэтому равенство (8) эквивалентно равенству

$$p(i,n) = \sum_{m=\max(i-1,n-1)}^{n} U(i,m,n)(-d)^{-m+n}q^{-i+m}p^{i+m-n-1}. \tag{9}$$

Замена $n - m = k$ и использование правила симметрии для биномиальных коэффициентов приводит к

$$p(i,n) = \sum_{k=0}^{\min(i,n-i+1)} V(i,k,)(-d)^k p^{i-k-1}q^{n-k-i}, \tag{10}$$

$$V(i,k,n) = \binom{n-k}{k-1}\binom{n-2k+1}{i-k}pq + \binom{n-k}{k}\binom{n-2k}{i-k-1}aq + \binom{n-k}{k}\binom{n-2k}{i-k}(1-a)p.$$

В полученных формулах произведения биномиальных коэффициентов можно заменить полиномиальными и вынести общие множители. Пусть

$$M(j,k,l) = \frac{(j+k+l)!}{(j!\,k!\,l!)},$$

$$K(m,i,n) = \frac{a(i+m-n)}{(m+1)p} + \frac{(1-a)(-i+m+1)}{(m+1)q} + \frac{n-m}{m+1}.$$

Тогда (9) и (10) превращается в равенства:

$$p(i,n) = \sum_{m=\max(i-1,n-1)}^{n} K(m,i,n)M(n-m,m-n+i,m-i+1)(-d)^{-m+n}q^{-i+m+1}p^{i+m-n}, \tag{11}$$

$$p(i,n) = \sum_{k=0}^{\min(i,n-i+1)} K(n-k,i,n)M(k,i-k,n-i-k+1)(-d)^k p^{i-k}q^{n-k-i+1}. \tag{12}$$

Благодаря мультиномиальному множителю равенства (11) и (12) могут быть полезны при исследовании асимптотики распределения $p(\cdot,n)$. Подчеркнем, что $0 < p,\ q < 1$, $-1 < d < 1$ и $p+q-d = 1$. Если $0 < d < 1$, то d, p, q составляют некоторое распределение вероятностей.

4. Среднее и дисперсия числа единиц

Легко получить простую формулу для среднего числа единиц в отрезках двоичной марковской последовательности. Формула для дисперсии сложнее.

4.1. Благодаря двоичным значениям переменных $\xi(k)$ среднее значение $M(n)$ случайной величины $x(n)$ равно сумме вероятностей $p_k = Pr(\xi(k) = 1)$:

$$M(n) = \sum_{k=0}^{n} p_k = \sum_{k=0}^{n} \big(b + (a - b)d^k\big).$$

Используя формулу для суммы геометрической прогрессии и введенные обозначения, находим

$$M(n) = a + nb + cd(1 - d^n). \tag{13}$$

Слагаемое $a + nb$ является биномиальной частью формулы (13). Если $a = p$ и $d = 0$, то $c = 0$, $b = 1 - q = p$, марковская последовательность становится бернулевской и $M(n) = (n + 1)p$. Слагаемое $cd(1 - d^n)$ является аддитивной поправкой на нестационарность распределения случайных переменных $\xi(k)$. В стационарном случае $(a = p,\ c = 0)$ тоже верно равенство $M(n) = (n + 1)p$.

Те же самые равенства для средних можно получить, дифференцируя производящую функцию $f[s, v]$ по s и разлагая производную в точке 1 в ряд по v. Так как по предположению $|d| < 1$, то $d^n n^m \to 0$ $(n \to \infty)$ и

$$M(n) = a + nb + cd + O\big(d^n\big) = a + nb + cd + o\big(n^{-m}\big) \quad (n \to \infty).$$

Для погрешностей асимптотических равенств легко получить простые оценки:

$$\begin{aligned}
&|M(n) - (a + nb + cd)| <= d(1 - d)^{-1}d^n,\\
&|M(n) - (a + nb)| <= d(1 - d)^{-1}(1 + d^n),\\
&|M(n) - nb| <= 1 + d(1 - d)^{-1}(1 + d^n).
\end{aligned}$$

Приближенные значения для средних $M(n)$ можно также получить, используя центральную предельную теорему [3].

4.2. Для вычисления дисперсии случайной переменной $x[n]$ удобно использовать найденную производящую функцию. Последовательное дифференцирование равенства (5) дает

$$f'(s) = \frac{dv^2(a + q - 1) - v(ad + a + q - 1) + a}{\big(dsv^2 - v(ps + q) + 1\big)^2},$$

$$f''(s) = f'(s)\frac{2v(p - dv)}{dsv^2 - v(ps + q) + 1}.$$

При $s = 1$, разлагая числитель и знаменатель на множители и сокращая общие множители, получаем

$$f'(1) = \frac{v(-a - q + 1) + a}{(1 - v)^2(1 - dv)}, \qquad f''(1) = f'(1)\frac{2v(p - dv)}{(1 - v)(1 - dv)}.$$

Используя биномиальные ряды, после некоторых преобразований получаем

$$f'(1) = \sum_{n=0}^{\infty} M(n)v^n = \sum_{n=0}^{\infty} v^n\big(a + bn + cd(1 - d^n)\big), \quad f''(1) = 2f'(1)\Big(pv + \sum_{n=0}^{\infty} v^n\big((1 - b)d^n + b\big)\Big).$$

Откуда, вычисляя коэффициенты $c'(n)$, $c''(n)$ при v^n для $f'(1)$, $f''(1)$ в полученных рядах, находим по известной формуле для дисперсии

$$\begin{aligned} V(n) &= c''(n)+c'(n)-c'(n)^2 \\ &= 2\Big(p\big(bn+c(1-d^n)\big)+\sum_{m=2}^{n}\big(b(1-d^m)+d^m\big)\Big(b(-m+n+1)+c\big(1-d^{-m+n+1}\big)\Big)\Big)+ \\ &\quad \big(b(n+1)-cd^{n+1}+c\big)-\big(b(n+1)-cd^{n+1}+c\big)^2. \end{aligned}$$

Суммируя и выделяя главную часть, получаем следующую формулу для дисперсии:

$$V(n) = a(1-a) + nb(1-b)\gamma + A - R(n) \tag{14}$$

с остатком $R(n) = B(nd^n) + Cd^n + Dd^{2n}$, где $\gamma = (1+d)/(1-d)$. При $\rho = d/(1-d)^2$ верны равенства:

$$\begin{aligned} &A = \rho\big(a(1-a)(2-d) + (a-2b)(1-2b) - b\big(1+(1-b)d\big)\big), \quad B = 2\rho(a-b)(q-p), \\ &C = \rho\big(a\big(2(1-a)+(1-2b)(1-d)\big) + \big(1-d-4(1-b)\big)b\big), \quad D = \rho(a-b)^2 d. \end{aligned}$$

4.3. Для чисел единиц в длинных отрезках двоичной марковской последовательности целесообразно использовать асимптотические формулы дисперсии. Так как по предположению $|d| < 1$ и $d^n n^m \to 0$ при $n \to \infty$ для любого $m > 0$, то

$$V(n) = a(1-a) + nb(1-b)\gamma + A + O(nd^n) = nb(1-b)\gamma + A + o(n^{-m}) \quad (n \to \infty).$$

Слагаемые $a(1-a) + nb(1-b)\gamma$ составляют биномиальную часть формулы (14) с коэффициентом искажения $\gamma = (1+d)/(1-d)$, учитывающим зависимость между случайными переменными $\xi(k)$. Слагаемые $A + B(nd^{n+1}) + Cd^{n+1} + D\big(d^{2(n+1)}\big)$ являются аддитивной поправкой на их зависимость и нестационарность.

Целесообразно выделить стационарную последовательность. Для нее $a = b$, $c = 0$ и формулы (13), (14) дают

$$M(n) = (n+1)b, \qquad V(n) = (n+1)b(1-b)\gamma - 2(1-b)bd(1-d)^{-2}\left(1-d^{n+1}\right).$$

Замечание 2. Главные члены nb и $nb(1-b)\gamma$ формул для средних и дисперсий были выведены А.А. Марковым в [3] и использованы им в предельной теореме. Другие члены формул могут играть существенную роль при значениях d, близких к 1. Кроме того, они позволяют оценить точность указанного в [3] нормального приближения и скорость сходимости к нему.

Используя полученные равенства и заменяя $a = Pr(\xi(0) = 1)$ на $Pr(\xi(k) = 1)$, легко вывести из найденных аналогичные формулы средних $M(k,n)$ и дисперсий $V(k,n)$ для числа единиц на отрезке $(k, k+n)$. А заменяя a на b, можно получить приближенные формулы для $M(k,n)$ и $V(k,n)$ при больших значениях k.

4.4. Оценим погрешность приближения дисперсии $V(n)$ ее членами: $a(1-a) + nb(1-b)\gamma + A$, $a(1-a) + nb(1-b)\gamma$, $nb(1-b)\gamma$. Как нетрудно проверить, при $|d| < 1$ верны неравенства:

$$|A| < \frac{5}{(1-d)^2}, \quad |B| < \frac{2}{(1-d)^2}, \quad |C| < \frac{4}{(1-d)^2}, \quad |D| < \frac{1}{(1-d)^2}.$$

Откуда

$$\big|V(n) - \big(a(1-a) + nb(1-b)\gamma + A\big)\big| < 3n^{-1}(1-d)^{-2},$$

$$\big|V(n) - \big(a(1-a) + nb(1-b)\gamma\big)\big| < (5 + 3n^{-1})(1-d)^{-2},$$

$$\big|V(n) - nb(1-b)\gamma\big| < (5 + 3n^{-1})(1-d)^{-2}.$$

Последние неравенства верны, когда $n > 4 + 4(\ln|d|)^{-2}$. Неравенство $n > 4$ обеспечивает $2 + 4/n + 1/n^2 < 3$, а неравенство $n > 4(\ln|d|)^{-2}$ вместе с неравенством $\ln(n) \le \sqrt{n}$ влекут $|nd^n| < n^{-1}$ при $0 \le |d| \le 1$.

Из полученных для абсолютных погрешностей оценок выводятся оценки для относительных погрешностей. Они имеют порядок n^{-1}. При больших значениях n указанные приближения можно эффективно использовать для вычислений. Оценки и примеры показывают, что $m(n) = a + nb + cd$ и $v(n) = a(1-a) + nb(1-b)\gamma + A$ хорошо приближают $M(n)$ и $V(n)$ даже при сравнительно небольших значениях n.

Замечание 3. Вместо $x(n)$ при $d \ne 1$ можно рассматривать случайную величину $\bar{x}(n) = (1-d)x(n)$. Множитель $1-d = p+q$ оценивает сохранение состояния в следующий момент и учитывает зависимость между переменными $\xi(k)$. Распределение, среднее и дисперсия для $\bar{x}(n)$ сразу выводятся из соответствующих формул для $x(n)$. Умножение на $(1-d)$ и $(1-d)^2$ позволяет освободиться от знаменателей.

4.5. Приведем пример точных и приближенных вычислений моментов числа единиц в отрезках двоичной марковской последовательности. Пусть

$$M = M(n), \quad m = a + nb + cd, \quad \bar{m} = a + nb, \quad \hat{m} = nb,$$
$$\Delta = M - m, \quad \bar{\Delta} = M - \bar{m}, \quad \hat{\Delta} = M - \hat{m},$$
$$\mu = d(1-d)^{-1}d^n, \quad \bar{\mu} = d(1-d)^{-1}(1 + d^n), \quad \hat{\mu} = 1 + (1-d)^{-1} + d(1-d)^{-1})d^n,$$
$$V{=}V(n), \quad v = a(1-a) + nb(1-b)\gamma + A, \quad \bar{v} = a(1-a) + nb(1-b)\gamma, \quad \hat{v} = nb(1-b)\gamma,$$
$$\delta = V - v, \quad \bar{\delta} = V - \bar{v}, \quad \hat{\delta} = V - \hat{v},$$
$$\sigma = 3(1-d)^{-2}n^{-1}, \quad \bar{\sigma} = \hat{\sigma} = 5(1-d)^{-2} + 3(1-d)^{-2}n^{-1}, \quad \nu = 4\big(2 + (\ln|d|)^{-2}\big).$$

Рассмотрим три распределения с параметрами:

$$(n, a, p, q) = \big((24, 0.5, 0.5, 0.5), (24, 0.5, 0.25, 0.5), (24, 0.5, 0.75, 0.75)\big).$$

Значения точных и приближенных средних $(M, m, \bar{m}, \hat{m})$, оценки погрешностей $(\Delta, \bar{\Delta}, \hat{\Delta})$ и $(\mu, \bar{\mu}, \hat{\mu})$ для этих распределений выражаются соответствующими тройками строк значений:

$$(12.5, 12.5, 12.5, 12), \qquad (0, 0, 0.5), \qquad (0, 0, 1);$$
$$(10.08, 10.08, 10.1, 9.6), \quad (0, -0.02, 0.48), \quad (0, -0.2, 0.8);$$
$$(12.5, 12.5, 12.5, 12), \qquad (0, 0, 0.5), \qquad (0, 1, 2).$$

Значения точных и приближенных дисперсий $(V, v, \bar{v}, \hat{v})$, оценки погрешностей $(\delta, \bar{\delta}, \hat{\delta})$ и $(\sigma, \bar{\sigma}, \hat{\sigma})$ для этих распределений выражаются соответствующими тройками строк значений:

$$\begin{array}{lll}(6.25, 6.25, 6.25, 6), & (0, 0, 0.25), & (0.125, 5.125, 5.125);\\ (3.68, 3.68, 3.706, 3.456), & (0, -0.026, 0.224), & (0.08, 3.28, 3.28);\\ (17.75, 17.75, 18.25, 18), & (0, -0.5, -0.25), & (0.5, 20.5, 20.5).\end{array}$$

Длины отрезков и начальные распределения для всех трех распределений одинаковые. Таблицы показывают изменение точности приближений при изменении переходных вероятностей.

5. Приложения

Опишем некоторые приложения двоичной марковской модели к вычислительной математике и геофизике.

5.1. Рассмотрим классическую задачу вычислительной математики и теории программирования — задачу о числе переносов разряда при последовательном сложении случайных двоичных чисел. Соответствующая марковская модель подробно описана в [5]. При последовательном суммировании после группировки определяющих перенос двоичных троек дело сводится к двоичной марковской последовательности с начальной и переходной матрицами:

$$A = (1/4, 3/4), \qquad Q = \begin{pmatrix} 3/4 & 1/4 \\ 1/4 & 3/4 \end{pmatrix}.$$

Среднее число переносов

$$M(n) = \frac{1}{2}\left(n + \left(\frac{1}{2}\right)^{n+1}\right).$$

Складывая, например, 32-значные двоичные числа ($n = 31$), можно ожидать появления в среднем примерно 15 переносов разряда.

Дисперсия числа переносов равна

$$V(n) = \frac{1}{4}\left(3n - 2 + 11\left(\frac{1}{2}\right)^{n+1} - \left(\frac{1}{4}\right)^{n+1}\right).$$

При $n = 31$ стандартное отклонение числа переносов от среднего равно примерно 5. Грубо можно считать, что при сложении случайных 32-разрядных чисел в трех случаях из четырех число переносов будет между 10 и 20.

5.2. Полученные результаты можно применить для подсчета различных характеристик серий жидких осадков. Соответствующая марковская модель описана в [5]. Использовались приведенные в [1] данные 15-летних (1969–1983 гг.) наблюдений за суточным слоем осадков в теплое время года (май–сентябрь) в 47 пунктах равнинной части Новосибирской области. Статистический анализ показал, что рассматриваемая модель удовлетворительно описывает наблюдавшийся реальный процесс.

Приведем некоторые результаты предлагаемого марковского моделирования. Параметры модели собраны в табл. 1. Эмпирические и теоретические средние $m(n)$, $M(n)$ приведены в табл. 2. Разность между теоретическими и эмпирическими средними не превышает четырех сотых.

Таблица 1

Параметры	Май	Июнь	Июль	Август	Сентябрь
p	0.455	0.470	0.491	0.483	0.495
q	0.747	0.781	0.744	0.728	0.786
a	0.220	0.281	0.332	0.428	0.403
d	0.202	0.251	0.235	0.212	0.281
b	0.317	0.293	0.335	0.344	0.297

Таблица 2

Средние	Май	Июнь	Июль	Август	Сентябрь
$m(n)$	9.68	8.74	10.34	10.78	9.03
$M(n)$	9.7	8.76	10.38	10.78	9.07

Литература

1. **Дробышев А.Д., Марченко А.С., Огородников В.А., Чижиков В.Д.** Статистическая структура временных рядов суточных сумм жидких осадков в равнинной части Новосибирской области // Тр. ЗапСиб НИИ Госкомгидромета, 1989. — Вып. 86. — С. 44–74. (Drobyshev A.D., Marchenko A.S., Ogorodnikov V.A., CHizhikov V.D. Statisticheskaya struktura vremennykh ryadov sutochnykh summ zhidkikh osadkov v ravninnoj chasti Novosibirskoj oblasti // Tr. ZapSib NII Goskomgidrometa, 1989. — Vyp. 86. — S. 44–74.)
2. **Кемени Дж., Снелл Дж.** Конечные цепи Маркова. — М.: Наука, 1970. (Kemeni Dzh., Snell Dzh. Konechnye tsepi Markova. — M.: Nauka, 1970.)
3. **Марков А.А.** Исчисление вероятностей. — СПб.: Типография императорской академии наук, 1913. (Markov A.A. Ischislenie veroyatnostej. — SPb.: Tipografiya imperatorskoj akademii nauk, 1913.)
4. **Савельев Л.Я.** Случайные соответствия, двоичные матрицы и серии // Дискретная математика. — 1999. — № 4. — С. 3–26. (Savel'ev L.YA. Sluchajnye sootvetstviya, dvoichnye matritsy i serii // Diskretnaya matematika. — 1999. — № 4. — S. 3–26.)
5. **Савельев Л., Балакин С.** Конечные марковские цепи и серии (теория и приложения). — LAP LAMBERT Academic Publishing, 2012. (Savel'ev L., Balakin S. Konechnye markovskie tsepi i serii (teoriya i prilozheniya). — LAP LAMBERT Academic Publishing, 2012.)
6. **Уиттекер Э.Т., Ватсон Дж.Н.** Курс современного анализа. В 2-х томах. — М.: Гос. изд-во физ.-мат. литературы, 1963. (Uitteker E.T., Vatson Dzh.N. Kurs sovremennogo analiza. V 2-kh tomakh. — M.: Gos. izd-vo fiz.-mat. literatury, 1963.)

Поступила в редакцию 26 декабря 2013 г.,
в окончательном варианте 28 февраля 2014 г.