УДК 519.2

АСИМПТОТИЧЕСКОЕ ПОВЕДЕНИЕ ЭКСТРЕМУМОВ СЛУЧАЙНЫХ ПРИЗНАКОВ ЧАСТИЦ В ВЕТВЯЩИХСЯ ПРОЦЕССАХ С НАСЛЕДСТВЕННОСТЬЮ[1]

А.В. Лебедев[2]

Рассматриваются надкритические ветвящиеся процессы без вырождения, в которых каждая частица обладает некоторым случайным признаком. Предполагается, что признаки частиц зависимы, причем эта зависимость обусловлена общей наследственностью и определяется степенью родства. Нас интересует асимптотическое поведение максимумов признака в популяции. Рассмотрены случаи, когда признаки имеют стандартное нормальное распределение и когда они имеют распределения с "тяжелыми хвостами". Получены невырожденные предельные законы для максимумов при линейной нормировке.

Ключевые слова: ветвящийся процесс, максимум, экстремумы, нормальное распределение, тяжелые хвосты, корреляция, случайный признак, наследственность, родство, невырожденный предельный закон

Keywords: branching process, maximum, extreme values, normal distribution, heavy tails, correlation, random score, heredity, kindred, nondegenerate limit law

1. Введение. В работах [1, 2], а также автором в [3] изучались максимумы на ветвящихся процессах с дискретным временем. А именно, рассматривался классический процесс Гальтона-Ватсона [4], в котором каждая частица обладает некоторым случайным признаком, и изучалось поведение максимумов признака в популяции. Если в качестве случайного признака частицы берется число ее потомков, получаем задачу о максимальном размере семейства [5]. В [6] изучались процессы с двумя типами частиц. Некоторые итоги зарубежных исследований подведены в [7]. Автором в [8] изучались также процессы с непрерывным временем. Всюду предполагалось, что признаки частиц независимы и одинаково распределены.

Однако для реальных биологических популяций предположение о независимости признаков является довольно грубом. На самом деле между ними существует зависимость, обусловленная общностью происхождения организмов, их наследственностью. До сих пор этот фактор при изучении максимумов никак не учитывался.

Следует упомянуть, что вопрос о зависимости количественных признаков детей и родителей (например, роста) изучался еще Ф.Гальтоном в конце XIX века [9, гл. 2, §6]. Как раз для описания этого явления он ввел модель линейной регрессии, получившую впоследствии широкое распространение. Мы также будем полагать зависимость

[1]Исправленный вариант работы, опубликованной в журнале "Ярославский педагогический вестник", Серия "Физико-математические и естественные науки", 2010, № 1, с. 7–14. Работа выполнена при поддержке по грантам РФФИ № 07-01-00077, № 07-01-00373.

[2]Лебедев Алексей Викторович, avlebed@yandex.ru, кафедра теории вероятностей, механико-математический факультет МГУ имени М.В. Ломоносова, Москва, 119991.

между признаками родственных частиц линейной. Учет наследования признаков "через поколение" приводит к авторегрессии второго порядка и т.д.

Далее сначала рассмотрим случай, когда признаки частиц имеют стандартное нормальное распределение, а корреляции между ними зависят от степени родства частиц. Оказывается, что при условии экспоненциального убывания корреляций максимум признака в популяции асимптотически ведет себя так же, как в случае независимых величин.

Потом мы рассмотрим случай тяжелых (правильно меняющихся) хвостов. Такая модель может описывать возможность больших скачков (мутаций), которые крайне маловероятны при нормальном распределении. Оказывается, что в этом случае наследственность оказывает большое влияние на асимптотическое поведение максимума.

Основные результаты для модели авторегрессии первого порядка были кратко изложены в [10], однако в полном объеме до сих пор не были опубликованы.

Введем необходимые объекты и обозначения.

Пусть Z_n — число частиц в n-ом поколении; $Z_0 = 1$. Будем рассматривать надкритические процессы без вырождения: число непосредственных потомков не менее одного, имеет конечное среднее μ и конечный второй момент. Напомним, что тогда имеет место сходимость почти наверное [4, гл. 1, §8]:

$$\frac{Z_n}{\mu^n} \to W, \quad n \to \infty. \tag{1}$$

Случайная величина W в нашем случае оказывается положительной почти наверное. Ее преобразование Лапласа-Стилтьеса, однозначно определенное теоремой 8.2 [4, гл. 1, §8], обозначим через $\varphi(t)$.

Пусть $\xi_{m,n}$ — признак m-ой частицы в n-ом поколении. Нас интересует асимптотическое поведение величины

$$M_n = \bigvee_{m=1}^{Z_n} \xi_{m,n}, \quad n \to \infty.$$

Рассматривая частицы текущего поколения, будем говорить, что они находятся в k-родстве, если они имеют ближайшего общего предка k поколений назад. В генеалогической терминологии (например, по женской линии), частицы 1-родства имеют общую мать и являются сестрами, 2-родства — имеют только общую бабушку и являются двоюродными сестрами и т.д. Кроме того, k-родственной группой назовем совокупность частиц, находящихся не более чем в k-родстве между собой. Вся популяция в любой момент времени разбивается на такие группы однозначно.

2. Гауссовский случай. Предположим, что признаки всех частиц имеют стандартное нормальное распределение Φ и зависимость между ними определяется только степенью родства. А именно, пусть признаки частиц, находящихся в k-родстве, имеют коэффициент корреляции ρ_k. Так, в модели Гальтона получаем $\rho_k = \rho^{2k}$, где ρ — коэффициент корреляции между признаками частицы и ее непосредственного предка (в силу того, что путь на генеалогическом дереве, соединяющий частицы, находящиеся в k-родстве, имеет длину $2k$). В общем случае мы не определяем корреляции между признаками частиц разных поколений, поскольку для наших целей это не нужно.

Введем следующее условие на величины ρ_k, $k \geq 1$. Пусть существует число $r \in (0,1)$ такое, что
$$|\rho_k| \leq r^k, \quad \forall k \geq 1. \tag{2}$$
Это условие заведомо выполняется для любой модели наследственности типа стационарной авторегрессии конечного порядка [11], поскольку в этом случае вклад общих предков в признаки родственных частиц убывает экспоненциально, причем все $\rho_k \geq 0$.

Если говорить о теоретической возможности отрицательной зависимости признаков, то она вряд ли может быть обусловлена наследственностью, однако может возникнуть, например, когда на признак сильно влияют условия жизни (питание или жизненное пространство) в условиях конкуренции частиц за ресурсы.

Напомним, что стандартное нормальное распределение принадлежит области притяжения максимум-устойчивого закона Гумбеля $\Lambda(x) = \exp\{-e^{-x}\}$ с известными нормирующими константами. С помощью теоремы 1.5.3 [12] получаем

$$\Phi^s(a(s)x + b(s)) \to \Lambda(x), \quad s \to \infty,$$
$$a(s) = (2\ln s)^{-1/2}, \quad b(s) = (2\ln s)^{1/2} - \tfrac{1}{2}(2\ln s)^{-1/2}(\ln\ln s + \ln 4\pi),$$

где $\Phi(x)$ — функция стандартного нормального распределения. Тогда, согласно теореме 1 [3], в случае независимых признаков частиц имеет место асимптотика

$$\mathsf{P}(M_n \leq a(\mu^n)x + b(\mu^n)) \to \varphi(e^{-x}), \quad n \to \infty, \tag{3}$$

где $\varphi(t)$ — преобразование Лапласа-Стилтьеса случайной величины W в предельном соотношении (1). Отсюда, в частности, следует, что $M_n \sim (2n\ln\mu)^{1/2}$, $n \to \infty$ (по вероятности). Кроме того, в ряде случаев предельное распределение можно выписать явно. Например, если число непосредственных потомков частицы распределено геометрически, то $\varphi(t) = 1/(1+t)$, откуда $\varphi(e^{-x}) = 1/(1+e^{-x})$, т.е. предельный закон для максимумов оказывается логистическим.

Теорема 1. *При условии (2) верно (3).*

Доказательство. Далее будем пользоваться следующей оценкой для стандартных нормальных величин $X_1, \ldots, X_N$ с корреляциями r_{ij}, удовлетворяющими условию $|r_{ij}| \leq \delta < 1$. Согласно следствию 4.2.4 [12] для них выполнено неравенство

$$\left| \mathsf{P}\left(\bigvee_{1 \leq i \leq N} X_i \leq u \right) - \Phi^N(u) \right| \leq K(\delta) \sum_{1 \leq i < j \leq N} |r_{ij}| \exp\left\{ -\frac{u^2}{1 + |r_{ij}|} \right\}. \tag{4}$$

Обозначим $u_n = a(\mu^n)x + b(\mu^n)$. Заметим, что имеет место асимптотика

$$\exp\left\{ -\frac{u_n^2}{2} \right\} \sim \sqrt{2\pi} e^{-x} \frac{(2n\ln\mu)^{1/2}}{\mu^n}, \quad n \to \infty. \tag{5}$$

Найдем среднее число пар частиц в k-родстве (в текущем, n-ом поколении). Пусть ζ — случайное число непосредственных потомков частицы, тогда среднее число всевозможных пар в 1-родственной группе составляет $\mathsf{M}\zeta(\zeta - 1)/2$. Обозначим это число через ν. Число 1-родственных групп в поколении совпадает с числом частиц в предыдущем поколении, со средним μ^{n-1}. Таким образом, среднее число пар частиц в 1-родстве составляет $\nu\mu^{n-1}$. Рассуждая далее, получаем, что среднее число пар частиц в k-родстве равно $\nu\mu^{n+k-2}$, $1 \le k \le n$. Таким образом, из (2) и (4) следует

$$|\mathsf{P}(M_n \le u_n) - \mathsf{M}\Phi^{Z_n}(u_n)| \le \frac{K(r)\nu}{\mu^2}\mu^n \sum_{k=1}^{n} (\mu r)^k \exp\left\{-\frac{u_n^2}{1+r^k}\right\}. \qquad (6)$$

Поскольку $r^k \to 0$, $k \to \infty$, можно выбрать число $l = l(r, \mu)$ такое, что

$$1 + \frac{\ln(\mu r \vee 1)}{\ln \mu} - \frac{2}{1+r^l} < 0.$$

Обозначим

$$A_n = \mu^n \sum_{k=1}^{l} (\mu r)^k \exp\left\{-\frac{u_n^2}{1+r^k}\right\}, B_n = \mu^n \sum_{k=l+1}^{n} (\mu r)^k \exp\left\{-\frac{u_n^2}{1+r^k}\right\},$$

тогда правая часть (6) может быть записана в виде $C(A_n + B_n)$. Используя асимптотику (5), получаем

$$A_n \le l(\mu r \vee 1)^l \mu^n \exp\left\{-\frac{u_n^2}{1+r}\right\} \sim C_1 n^{1/(1+r)} \mu^{(1-2/(1+r))n} \to 0, \quad n \to \infty.$$

Аналогично,

$$B_n \le n((\mu r \vee 1)\mu)^n \exp\left\{-\frac{u_n^2}{1+r^l}\right\} \sim C_2 n^{1+1/(1+r^l)} \mu^{\alpha n}, \quad n \to \infty,$$

где в силу выбора l имеем

$$\alpha = 1 + \frac{\ln(\mu r \vee 1)}{\ln \mu} - \frac{2}{1+r^l} < 0,$$

а значит и $B_n \to 0$, $n \to \infty$. С учетом (1) и (6) получаем

$$\lim_{n\to\infty} \mathsf{P}(M_n \le u_n) = \lim_{n\to\infty} \mathsf{M}\Phi^{Z_n}(u_n) = \mathsf{M}\Lambda^W(x) = \varphi(e^{-x}).$$

Заметим, что наш результат имеет асимптотический характер. На практике наследственность может приводить к стохастическому уменьшению максимума (за счет сокращения разброса признака в популяции), что в гауссовском случае также следует из нормальной леммы сравнения (следствие 4.2.3 [12]) при $\rho_k \ge 0$, $k \ge 1$.

3. Случай тяжелых хвостов. Пусть наследственность описывается линейной авторегрессией первого порядка:

$$\xi_{m,n} = a\xi_{\kappa(m,n),n-1} + b\xi^*_{m,n}, \quad a \in (0,1), b > 0, \qquad (7)$$

где $\xi_{m,n}$ — признак m-ой частицы в n-ом поколении, $\kappa(m,n)$ — номер предка этой частицы в предыдущем поколении, а случайные инновации $\xi^*_{m,n}$, $m \ge 1$, $n \ge 1$, независимы и имеют одинаковое распределение F, которое удовлетворяет условиям:

$$\bar{F}(x) \sim x^{-\gamma} L(x), \quad F(-x)/\bar{F}(x) \to r \ge 0, \quad x \to \infty, \quad \gamma > 0, \qquad (8)$$

где $L(x)$ — медленно меняющаяся функция [13, гл. 8, §8]. Этим условиям, в частности, удовлетворяют устойчивые распределения с показателями $\gamma \in (0, 2)$, распределения Парето, лог-гамма и др.

Определим также неотрицательную функцию $u(s)$, $s > 0$, такую, что $s\bar{F}(u(s)) \to 1$, $s \to \infty$. Заметим, что $u(s)$ определена корректно и правильно меняется с показателем $1/\gamma$, т.е. $u(s) \sim s^{1/\gamma} L_2(s)$, $s \to \infty$, где $L_2(s)$ — медленно меняющаяся функция [14, §1.5].

В модели (8) существует и единственно стационарное распределение Ψ, совпадающее с распределением случайного ряда

$$X = \sum_{k=0}^{\infty} b a^k Z_k,$$

где Z_k независимы и имеют распределение F. По лемме A3.26 [15, с. 583] получаем асимптотические соотношения:

$$\bar{\Psi}(x) \sim \frac{b^\gamma}{1 - a^\gamma} \bar{F}(x), \quad \Psi(-x)/\bar{\Psi}(x) \to r, \quad x \to \infty.$$

Далее полагаем, что все признаки частиц имеют стационарное распределение.

Для выявления роли наследственности "в чистом виде" желательно обеспечить независимость распределения признаков от коэффициентов авторегрессии (как это имело место в гауссовском случае). Здесь можно добиться этого только для строго устойчивых распределений, полагая

$$a^\gamma + b^\gamma = 1. \tag{9}$$

Для произвольных F, удовлетворяющих (8), условие (9) обеспечивает асимптотическую эквивалентность хвостов: $\bar{\Psi}(x) \sim \bar{F}(x)$, $x \to \infty$. Будем предполагать, что это условие выполнено.

Обозначим для краткости $C = b^\gamma/(1 - a^\gamma/\mu)$.

Теорема 2. *При условиях (7) и (8) верно*

$$\frac{M_n}{u(\mu^n)} \xrightarrow{d} (CW)^{1/\gamma} \eta, \quad n \to \infty, \tag{10}$$

где W определено в (1), η имеет распределение Фреше $\Phi_\gamma(x) = \exp\{-x^{-\gamma}\}$, $x > 0$, и не зависит от W.

Доказательство. Утверждение (10) можно представить в виде

$$\mathsf{P}(M_n \le x u(\mu^n)) \to \mathsf{M} \exp\left\{-CWx^{-\gamma}\right\}, \quad n \to \infty, \quad x > 0.$$

Обозначим через $\tilde{M}_k$ максимум признака в k-родственной группе ($k < n$) при условии, что ее основатель (общий предок в k-ом поколении) произошел от непосредственного предка с признаком, равным нулю. Тогда легко получить рекуррентную формулу

$$\tilde{M}_k \stackrel{d}{=} \bigvee_{i=1}^{\zeta} \tilde{M}_{k-1}^{(i)} + b a^k \xi, \quad \tilde{M}_0 = b\xi,$$

где $\tilde{M}_k^{(i)}$, ξ, ζ независимы, $\tilde{M}_k^{(i)} \stackrel{d}{=} \tilde{M}_k$, $\xi \stackrel{d}{=} \xi_{1,1}$, ζ имеет распределение числа непосредственных потомков.

Воспользовавшись известными свойствами распределений с правильно меняющимися хвостами [13, гл. 8, §8], получаем

$$\mathsf{P}(\tilde{M}_k > x) \sim c_k \bar{F}(x), \quad x \to \infty; \quad c_k = \mu c_{k-1} + (ba^k)^\gamma, \quad c_0 = b^\gamma,$$

откуда следует асимптотика

$$c_k \sim C\mu^k, \quad k \to \infty. \tag{11}$$

Для стохастического оценивания M_n снизу разобъем популяцию на k-родственные группы, порожденные частицами $(n-k)$-го поколения. Рассмотрев два случая: когда все признаки частиц $(n-k-1)$-го поколения не менее $-u(\mu^n)$ и когда хотя бы один из них меньше, получаем оценку

$$\begin{aligned}&\mathsf{P}(M_n \le xu(\mu^n)) \le \\ &\le \mu^{n-(k+1)}\Psi(-u(\mu^n)) + \mathsf{M}\mathsf{P}^{Z_{n-k}}(\tilde{M}_k \le (x + a^{k+1})u(\mu^n)) \to \\ &\to \mu^{-(k+1)} r b^\gamma/(1-a^\gamma) + \mathsf{M}\exp\{-c_k\mu^{-k}W(x+a^{k+1})^{-\gamma}\}, \quad n \to \infty.\end{aligned}$$

Переходя к пределу по $k \to \infty$ с учетом (11), получаем

$$\limsup_{n\to\infty} \mathsf{P}(M_n \le xu(\mu^n)) \le \mathsf{M}\exp\left\{-CWx^{-\gamma}\right\},$$

Для получения стохастической оценки M_n сверху заметим, что максимумы признаков по отдельным k-родственным группам положительно зависимы (ассоциированы) между собой как возрастающие функции от независимых величин $\{\xi_{m,n}\}$ [16, 17]. Поэтому их общий максимум (по всей популяции) стохастически оценивается сверху максимумом случайных величин с теми же распределениями, но независимыми. Таким образом, получаем

$$\begin{aligned}&\mathsf{P}(M_n \le xu(\mu^n)) \ge \mathsf{M}\mathsf{P}^{Z_{n-k}}(\tilde{M}_k + a^{k+1}\xi \le xu(\mu^n)) \to \\ &\to \mathsf{M}\exp\left\{-(c_k + a^{\gamma(k+1)})\mu^{-k}Wx^{-\gamma}\right\}, \quad n \to \infty.\end{aligned}$$

Переходя к пределу по $k \to \infty$ с учетом (11), получаем

$$\liminf_{n\to\infty} \mathsf{P}(M_n \le xu(\mu^n)) \ge \mathsf{M}\exp\left\{-CWx^{-\gamma}\right\},$$

Совпадение верхнего и нижнего пределов доказывает теорему.

При условии (9) получаем $C = (1-a^\gamma)/(1-a^\gamma/\mu)$. Таким образом, наследственность приводит к появлению в асимптотике максимумов дополнительного множителя (от 0 до 1) по сравнению с максимумами независимых величин, причем этот множитель убывает с ростом коэффициента a. Подобный эффект выглядит вполне естественно: чем сильнее наследственность, тем меньше разброс признака в поколении и, соответственно, меньше максимум.

В теории экстремумов стационарных случайных последовательностей известно понятие экстремального индекса θ [15, §8.1]. А именно, во многих случаях максимум n элементов последовательности асимптотически растет как максимум θn независимых случайных величин с тем же распределением. Таким образом, экстремальный индекс описывает влияние зависимости на максимум. Распространяя это понятие с последовательностей на поколения частиц, приходим к выводу, что величина C здесь как раз и играет роль экстремального индекса.

Для последовательностей значение $\theta \in (0, 1)$ означает, что превышения высокого уровня происходят не по одиночке, а группами (кластерами) средней величины $1/\theta$. В нашем случае также можно предположить образование подобных кластеров. Очевидно, речь должна идти о родственных группах частиц, имеющих общего предка с аномальным признаком, и унаследовавшим эту мутацию.

Было проведено компьютерное моделирование. Для простоты рассмотрен ветвящийся процесс, в котором каждая частица имеет ровно двух непосредственных потомков. Инновации имели распределение Стьюдента с двумя степенями свободы (так что $\gamma = 2$); параметр $a = 0,9$. В этом случае получаем $C \approx 0,32$ и ожидаемый средний размер кластера $1/C \approx 3,13$.

Результат для популяции из 1024 частиц (в 10-м поколении) представлен на рис. 1. Видно, что основная масса (более 80%) сосредоточена в диапазоне признаков $[-2, 2]$, но есть и значительные отклонения. Частицы пронумерованы в лексикографическом порядке (по своему положению на двоичном дереве), поэтому родственники находятся рядом (обратное необязательно). Группы родственных частиц с аномальными признаками выглядят как вертикально вытянутые группы точек.

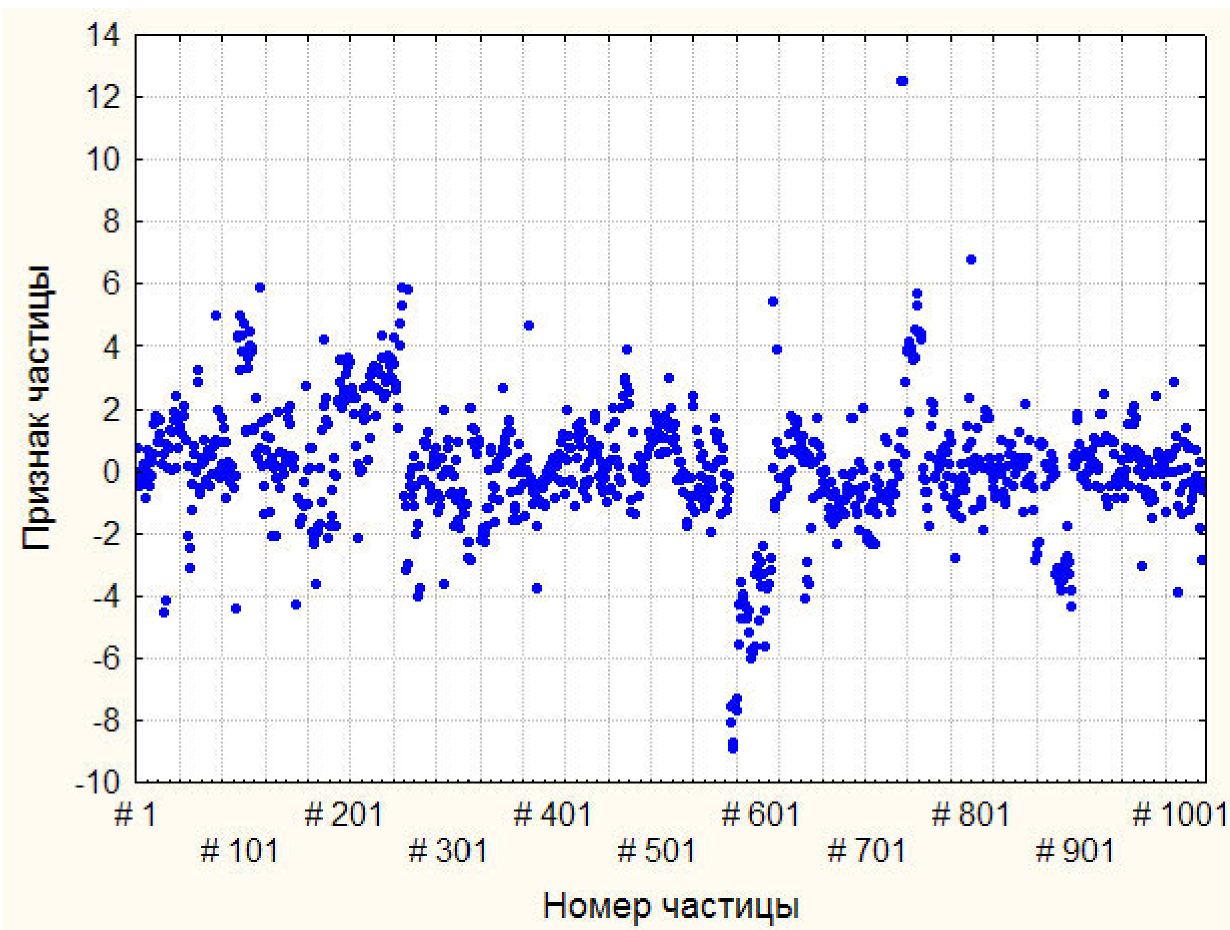

Рис. 1:

В заключение отметим, что популяция (в каждом поколении) может быть описана случайным полем в некотором (случайном) метрическом пространстве, точки которого соответствуют отдельным частицам, а расстояния между ними — степени их родства. Коэффициенты ρ_k, $k \geq 1$, либо модели авторегрессии определяют структуру зависимости поля, а задача сводится к изучению его глобального максимума.

Одной из особенностей введенного пространства в том, что из любых трех расстояний между точками (частицами) либо равны все три расстояния, либо два равны, а третье — меньше. Пространство, обладающее таким свойством, называют ультраметрическим. Подобные пространства возникают в различных математической физики и др. В качестве обзора на эту тему можно указать [18]. Таким образом, мы изучаем экстремумы случайного поля в ультраметрическом пространстве.

Автор благодарен Л.Г.Афанасьевой, М.В.Козлову и Е.Б.Яровой за внимание к работе и полезное обсуждение.

Список литературы

[1] *Arnold B.C., Villasenor J.A.* The tallest man in the world / Statistical theory and applications. Papers in honor of H.A.David. Springer. 1996. P. 81–88.

[2] *Pakes A.G.* Extreme order statistics on Galton-Watson trees // Metrika. 1998. V. 47. P. 95–117.

[3] *Лебедев А.В.* Максимумы случайных признаков частиц в надкритических ветвящихся процессах. // Вестник МГУ. Сер.1. Математика. Механика. 2008. № 5. С. 3–6.

[4] *Харрис Т.* Теория ветвящихся случайных процессов. М.: Мир, 1966.

[5] *Rahimov L., Yanev G.P.* On maximum family size in branching processes // J. Appl. Probab. 1999. V. 36. P. 632–643.

[6] *Mitov K.V., Yanev G.P.* Maximum individual score in critical two-type branching processes // C. R. Acad. Bulg. Sci. 2002. V. 55. № 11. P.17–22.

[7] *Yanev G.P.* Revisiting offspring maxima in branching processes // Pliska Studia Mathematica Bulgarica. 2007. V. 18. P.401–426.

[8] *Lebedev A.V.* Maxima of random particles scores in Markov branching processes with continuous time // Extremes. 2008. V. 11. № 2. P. 203–216.

[9] *Секей Г.* Парадоксы в теории вероятностей и математической статистике. М.: Институт компьютерных исследований, 2003.

[10] *Лебедев А.В.* Максимумы наследуемых признаков частиц в ветвящихся процессах // Труды V Колмогоровских чтений. Ярославль, 22–25 мая 2007. Изд-во ЯГПУ, 2007. С. 62–66.

[11] *Бокс Дж., Дженкинс Г.* Анализ временных рядов. М.: Мир, 1974.

[12] *Лидбеттер М., Линдгрен Г., Ротсен Х.* Экстремумы случайных последовательностей и процессов. М.: Мир, 1989.

[13] *Феллер В.* Введение в теорию вероятностей и ее приложения. Т.2. М.: Мир, 1984.

[14] *Сенета Е.* Правильно меняющиеся функции. М.: Наука, 1985.

[15] *Embrechts P., Klüppelberg C., Mikosh T.* Modelling Extremal Events for Insurance and Finance. Springer, 1997.

[16] *Esary J., Prochan F., Walkup D.* Association of random variables with applications // Ann. Math. Stat. 1967. V. 38. № 5. P. 1466–1474.

[17] *Булинский А.В., Шашкин А.П.* Предельные теоремы для ассоциированных случайных полей и родственных систем. М.: Физматлит, 2008.

[18] *Rammal R., Toulose G., Virasoro M.A.* Ultrametricity for physicists // Rev. Mod. Phys. 1986. V. 58. P. 765–788.