Санкт-Петербургский государственный университет

На правах рукописи

Сагателян Ваагн Каренович

Предельные теоремы и характеризационные соотношения для упорядоченных случайных величин

01.01.05 – теория вероятностей и математическая статистика

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
кандидата физико-математических наук

Санкт-Петербург
2008

Работа выполнена на кафедре теории вероятностей и математической статистики математико-механического факультета Санкт-Петербургского государственного университета

Научный руководитель: доктор физико-математических наук, профессор Невзоров Валерий Борисович

Официальные оппоненты: доктор физико-математических наук, профессор Розовский Леонид Викторович

кандидат физико-математических наук Запорожец Дмитрий Николаевич

Ведущая организация: Санкт-Петербургский государственный электротехнический университет (ЛЭТИ)

Защита состоится "___" _________ 2009г. в ___ часов на заседании диссертационного совета Д 002.202.01 в Санкт-Петербургском отделении Математического института им. В.А. Стеклова РАН по адресу 191023 наб. р. Фонтанки, д. 27, к. 311.

С диссертацией можно ознакомиться в библиотеке Санкт-Петербургского отделения Математического института им. В.А. Стеклова РАН по адресу 191023 наб. р. Фонтанки, д. 27.

Автореферат разослан "___" _________ 200 г.

Ученый секретарь
диссертационного совета Д 002.202.01
доктор физико-математических наук Зайцев А.Ю.

Общая характеристика работы

Актуальность проблемы. Диссертация посвящена изучению новой модели математических рекордов и проблеме характеризации распределений различными свойствами порядковых статистик.

Определенная и исследуемая в диссертации новая модель рекордных величин, названная ”рекордами с подтверждением”, является обобщением обыкновенных математических рекордов. Теория рекордов является одной из наиболее динамично развивающихся областей современной теории вероятностей. Это объясняется неугасающим интересом Человечества ко всевозможным рекордам, которые фиксируются в самых различных областях жизни. Сама классическая рекордная модель и ее различные обобщения, такие как слабые рекорды, k-е рекорды, верхние-k-списки (top-k-lists) и т.д., являются предметом активного исследования. Систематическое изложение ”рекордных” результатов можно найти, например, в работах Невзорова [2], Ахсануллаха [3, 4], Арнольда и соавторов [6].

Задачам характеризаций распределений различными свойствами порядковых статистик посвящено немало работ. В частности схемы случайного сжатия рассмотрены, например, у Весоловского и Ахсануллаха [12], Наварро [9], Бейтнера и Кампса [7]. Эти схемы позволяют при определенных условиях выражать одни порядковые статистики через другие (чаще всего соседние). Отметим также проблему оценивания математического ожидания для определенных сумм порядковых статистик и рекордных величин, рассмотренных у Ахундова, Берреда и Невзорова [5], Нагараджи [8] и некоторые результаты из монографии Дэйвида [1]. Все эти статистики достаточно широко используются для изучения всевозможных выборочных характеристик.

Цель работы. Целью диссертации является решение следующих задач.

1. Построить и исследовать математическую модель рекордов с подтверждением.

2. Получить характеризации распределений с помощью равенств порядковых статистик, включая схемы случайного сжатия/расширения.

3. Для ряда величин получить оценки, обобщающие известные результаты. На основе этих оценок получить характеризации максимизирующих распределений.

Методы исследований. При изучении рекордов с подтверждением используются классические результаты теории рекордов. В частности, применяются известные представления для экспоненциальных и равномерных рекордов, вероятностное преобразование Смирнова [11] и метод, предложеный в работе Резника [10] для доказательства предельных теорем.

При изучении проблемы характеризации распределений различными свойствами порядковых статистик применяются классические результаты и методы из различных областей математики.

Основные результаты. При исследовании новой модели рекордных величин были получены следующие результаты.

1. Мотивирована и построена новая модель рекордных величин,

названная "рекордами с подтверждением".

2. Получены представления для экспоненциальных рекордов с подтверждением через суммы независимых одинаково распределенных случайных величин, имеющих стандартное экспоненциальное распределение. Похожее представление получено и для равномерных рекорда с подтверждением.

3. Описан класс всех возможных предельных функций распределения для произвольных рекордов с подтверждением.

При исследовании проблемы характеризации распределений различными свойствами упорядоченных случайных величин получены следующие результаты.

1. Рассмотрена новая схема случайного одностороннего сжатия/расширения на основе экспоненциальных случайных величин для выборочных минимумов. Получены характеризации соответствующих распределений. Кроме того, аналогичная задача решена для двухсторонней схемы.

2. Рассмотрено равенство двух произвольных порядковых статистик и описан класс распределений, при которых это равенство выполняется.

3. Получены оценки математического ожидания для сумм рекордных величин и выборочных максимумов. Кроме того получены характеризации максимизирующих исходных распределений.

4. Введено понятие обобщенного выборочного размаха, получена оценка математического ожидания этой величины и характе-

ризация максимизирующего исходного распределения. При исследовании асимптотического поведения описана предельная функция распределения, к которой сходится последовательность соответствующих максимизирующих исходных функций распределения.

Научная новизна. Все основные результаты диссертации являются новыми.

Теоретическая и практическая ценность. Диссертация носит теоретический характер. Построенная в работе новая модель рекордов может быть существенно использована при работе с т.н. "грязными" выборками, кроме того может иметь значение при дальнейшем исследовании рекордных моделей. Полученные характеризации распределений и оценки будут полезны при дальнейшем развитии этого направления.

Апробация работы. Результаты были доложены дважды на Санкт-Петербургском городском семинаре по теории вероятностей и математической статистике ПОМИ РАН (в 2008г.).

Публикации. По теме диссетации опубликовано четыре работы [13]–[16].

Структура и объем диссертации. Диссертация состоит из введения, двух глав, разбитых на параграфы и пункты, заключения и списка литературы, включающего 53 названия. Общий объем диссертации – 98 страниц.

Содержание работы

Во **введении** приводится краткое описание существующих результатов, формулируются поставленные задачи и кратко излагаются полученные результаты.

В первой главе строится и исследуется новая модель рекордных величин.

Параграф 1.1 посвящен краткому обзору существующих рекордных моделей и известных результатов, использованных в дальнейшей работе.

В **параграфе 1.2** строится модель рекордов с подтверждением. Пусть $k = 1, 2, \ldots$ – некоторое произвольное фиксированное число. При $k = 1$ рекорды с подтверждением совпадают с обычными рекордами. Рассмотрим последовательность независимых случайных величин $X_1, X_2, ...$, имеющих общую непрерывную функцию распределения $F(x)$. Для $k > 1$ возьмем случайный вектор $(X_1, X_2, ..., X_k)$ и упорядочим его компоненты в порядке возрастания: $(X_{1,k}, X_{2,k}, ..., X_{k,k})$. Первым рекордным моментом с подтверждением считаем $L^{(k)}(1) = k$, а соответствующей рекордной величиной объявляем $X_{k,k}$. Вместе с этими величинами будем рассматривать также первый рекордный вектор $(X^{(1)}_{k(1)}, X^{(1)}_{k(2)}, ..., X^{(1)}_{k(k)})$, совпадающий с вектором $(X_{1,k}, X_{2,k}, ..., X_{k,k})$. Далее нужно получить k случайных величин, превышающих $X_{k,k}$, упорядочив которые, получим второй рекордный вектор. Обозначим его $(X^{(2)}_{k(1)}, X^{(2)}_{k(2)}, ..., X^{(2)}_{k(k)})$. Остальные рекордные моменты, величины и векторы получаются последовательно таким же образом.

Определение. Рекордные моменты $L^{(k)}(n)$, соответствующие им рекордные величины $X^{(n)}_{k(k)}$ и рекордные векторы $(X^{(n)}_{k(1)}, X^{(n)}_{k(2)}, ..., X^{(n)}_{k(k)})$, построенные по последовательности $X_1, X_2, ...$, определяются следующим образом:
$L^{(k)}(1) = k,$
$X^{(1)}_{k(k)} = X_{k,k},$
$L^{(k)}(n) = \min\{j : X_{j-k+1,j} > X^{(n-1)}_{k(k)}\},$
$X^{(n)}_{k(k)} = \max\{X_1, X_2, ..., X_{L^{(k)}(n)}\} = X_{L^{(k)}(n),L^{(k)}(n)}, n = 2, 3, ...,$
$(X^{(n)}_{k(1)}, X^{(n)}_{k(2)}, ..., X^{(n)}_{k(k)}) = (X_{L^{(k)}(n)-k+1,L^{(k)}(n)}, ..., X_{L^{(k)}(n),L^{(k)}(n)}), n = 1, 2, ...$

Параграф 1.3 посвящен изучению рекордов с подтверждением в двух важных частных случаях: равномерном и экспоненциальном. Для этих случаев получены представления, сформулированные в следующих двух теоремах.

Теорема 1.4. *Пусть* $Z^{(1)}_{k(k)}, Z^{(2)}_{k(k)}, ...$ *- обозначают рекорды с подтверждением, построенные по последовательности* $Z_1, Z_2, ...$*– независимых случайных величин, с общей функцией распределения* $H(x) = \max\{0, 1 - e^{-x}\}$*, а* $(Z^{(n)}_{k(1)}, Z^{(n)}_{k(2)}, ..., Z^{(n)}_{k(k)})$*,* $n = 1, 2, ...$ *- соответствующие им рекордные векторы. Тогда для любого* $n = 1, 2, ...$ *имеют место соотношения:*

$$(Z^{(n)}_{k(1)}, Z^{(n)}_{k(2)}, ..., Z^{(n)}_{k(k)}) \stackrel{d}{=} \left(\frac{1}{k}\sum_{i=0}^{n-1}\eta_{ik+1} + \frac{1}{k-1}\sum_{i=0}^{n-2}\eta_{ik+2} + ... + \sum_{i=1}^{n-1}\eta_{ik},\right.$$

$$\left.\frac{1}{k}\sum_{i=0}^{n-1}\eta_{ik+1} + \frac{1}{k-1}\sum_{i=0}^{n-1}\eta_{ik+2} + ... + \sum_{i=1}^{n-1}\eta_{ik}, ..., \frac{1}{k}\sum_{i=0}^{n-1}\eta_{ik+1} + ... + \sum_{i=1}^{n}\eta_{ik}\right),$$

$(Z^{(n)}_{k(1)}, ..., Z^{(n)}_{k(k)}) \stackrel{d}{=} (Z(n, k) + Z(n-1, k-1) + ... + Z(n-1, 1), Z(n, k)+$

$+Z(n,k-1)+...+Z(n-1,1),...,Z(n,k)+Z(n,k-1)+...+Z(n,1))$,

где $\eta_1, \eta_2, ...$ *- последовательность независимых случайных величин, имеющих стандартное экспоненциальное распределение, а* $Z(n,k)$, $n=1,2,...$ *– k-е рекордные величины, построенные по той же последовательности* $Z_1, Z_2, \ldots$.

Теорема 1.5. *Пусть* $U_{k(k)}^{(1)}, U_{k(k)}^{(2)}, ...$ *- рекорды с подтверждением, построенные по последовательности* $U_1, U_2, ...$ *. Тогда*

$$U_{k(k)}^{(n)} \stackrel{d}{=} 1-(u_1^{(1)})^{\frac{1}{k}}(u_2^{(1)})^{\frac{1}{k}}\cdot\ldots\cdot(u_n^{(1)})^{\frac{1}{k}}(u_1^{(2)})^{\frac{1}{k-1}}\cdot\ldots\cdot(u_n^{(2)})^{\frac{1}{k-1}}\cdot\ldots\cdot(u_1^{(k)})\cdot\ldots\cdot(u_n^{(k)}),$$

где $u_i^{(j)}, i=1,...,n, j=1,..,k$ *- независимые случайные величины с равномерным на отрезке* $[0,1]$ *распределением.* Результат теоремы 1.4 позволил в **параграфе 1.4** показать асимптотическую нормальность экспоненциальных рекордов с подтверждением. Более того, удалось описать класс всех возможных предельных распределений для рекордов с подтверждением из произвольной совокупности. Этот результат сформулирован в том же параграфе в виде следующих двух теорем.

Теорема 1.7. *Для того, чтобы при некотором выборе констант* $A(n)$ *и* $B(n)$ *существовала невырожденная предельная функция распределения*

$$Q(x)=\lim_{n\to\infty}\mathbf{P}\left\{\frac{X_{k(k)}^{(n)}-B(n)}{A(n)}<x\right\},$$

необходимо и достаточно, чтобы существовала предельная функция

$$g(x)=\lim_{n\to\infty}\frac{T(xA(n)+B(n))-\tau_n}{\sigma_n},$$

где $\tau_n = n\sum_{i=0}^{k-1}\frac{1}{k-i}$, $\sigma_n^2 = n\sum_{i=0}^{k-1}\frac{1}{(k-i)^2}$, *причем функции* $Q(x)$ *и* $g(x)$ *связаны соотношением*

$$Q(x) = \Phi\left(g(x)\right),$$

где $\Phi(x)$ *– функция распределения стандартного нормального закона*

Теорема 1.8. *Предельная функция* $g(x)$ *может принадлежать только одному из следующих трех типов:*

1. $g_1(x) = -\infty$, *если* $x \leq 0$ *и* $g_1(x) = \frac{1+\sqrt{a_k}}{\sqrt{b_k}}\alpha\ln x$, $\alpha > 0$, *если* $x > 0$;

2.$g_2(x) = -\frac{1+\sqrt{a_k}}{\sqrt{b_k}}\alpha\ln(-x)$, *если* $x \leq 0$ *и* $g_2(x) = \infty$, *если* $x > 0$;

3.$g_3(x) = \frac{1+\sqrt{a_k}}{\sqrt{b_k}}x$, $-\infty < x < \infty$,

где $a_k = \sum_{i=0}^{k-1}\frac{1}{k-i}$, $b_k = \sum_{i=0}^{k-1}\frac{1}{(k-i)^2}$.

Глава II посвящена проблеме характеризации распределений различными свойствами упорядоченных случайных величин.

Первый параграф этой главы – **параграф 2.1** посвящен краткому изложению основных определений и результатов классической теории порядковых статистик, которые были использованы нами в работе. В следующем параграфе проводится обзор существующих результатов по тем конкретным направлениям, по которым мы и проводили исследования. В частности, речь идет о схеме одностороннего случайного сжатия/расширения, которое представляет собой равенство по распределению

$$X \stackrel{d}{=} YU,$$

связывающее три случайные величины, где U имеет известное распределение (от которого по сути зависит, будет это схемой сжатия или схемой расширения), а произведение YU – называется случайным сжатием/расширением величины Y. В последние годы авторы рассматривали схемы такого типа и для порядковых статистик, например Весоловски и Ахсануллах [12], Бейтнер и Кампс [7], Наварро [9].

Речь также идет о результатах работ Ахундова, Берреда и Невзорова [5], Нагараджи [8] и классических результатах, приведеных в монографии Дэйвида [1]. Последние относятся к проблеме оценивания моментов ряда известных статистик и характеризации соответствующих исходных распределений, при которых эти моменты достигают своего максимума.

В **параграфе 2.3** вводятся в рассмотрение новые схемы одностороннего случайного сжатия/расширения для выборочных минимумов.

$$Y_{1,k} \stackrel{d}{=} Y_{1,n} + W, \tag{1}$$

$$Y_{1,k} - W \stackrel{d}{=} Y_{1,n}, \tag{2}$$

где W имеет экспоненциальное распределение, а случайные величины $Y_1, Y_2, \ldots$ полагаем независимыми с общей непрерывной функцией распределения F. В следующих двух теоремах дается вид исходных распределений, обеспечивающих такие равенства.

Теорема 2.1. *Равенство по распределению (1) при некоторых фиксированных $1 \le k < n$ выполняется тогда и только тогда, когда*

$$F(x) = 1 - \left[\left(e^{C-x}\right)^{\frac{k-n}{k}} + 1\right]^{\frac{1}{k-n}}, -\infty < x < \infty,$$

где C - произвольная константа.

Теорема 2.2. *Равенство по распределению (2) при некоторых фиксированных $1 \le k < n$ выполняется тогда и только тогда, когда функция распределения $F(x)$ имеет вид*

$$F(x) = 1 - \left[1 - \left(e^{x-C}\right)^{\frac{n-k}{n}}\right]^{\frac{1}{n-k}}, -\infty < x < C,$$

где C- произвольная константа.

Эти теоремы позволяют получить аналогичные характеризации и для максимальных порядковых статистик, тем самым давая ответ на интересующие нас вопросы, касающиеся данной схемы сжатия/расширения для выборочных экстремумов. В том же параграфе рассмотрена двухсторонняя схема сжатия/расширения, для которой получена характеризация, сформулированная в нижеследующей теореме.

Теорема 2.4. *Равенство по распределению*

$$Y_{1,k} - W_\alpha \stackrel{d}{=} Y_{1,n} + W_\beta,$$

при некоторых фиксированных $1 \le k < n$ и $\alpha > 0, \beta > 0$ выполняется тогда и только тогда, когда функция распределения F задается с помощью своей обратной функции G следующим образом:

$$G(x) = \ln \frac{\left(1 - (1-x)^{n-k}\right)^{\frac{n\beta + k\alpha}{\alpha\beta(n-k)}} C_1}{(1-x)^{\frac{k}{\beta}}},$$

где C_1– произвольная положительная постоянная.

В **параграфе 2.4** рассматривается довольно общая проблема равенства по распределению двух произвольных порядковых статистик:

$$X_{k,n} \stackrel{d}{=} X_{l,m}.$$

Изучается вопрос существования таких исходных функций распределения и таких индексов n, m, k, l при которых это равенство имеет место. Доказана теорема, отвечающая на этот вопрос.

Теорема 2.5. *Если выполняется равенство* $X_{k,n} \stackrel{d}{=} X_{l,m}, m \neq n$, *то исходная функция распределения является либо вырожденной, либо имеет лишь две точки роста.*

Более того, оказалось, что функция распределения будет двухточечной в том случае, когда индексы n, m, k, l удовлетворяют условиям $n - m > k - l > 0$ или $m - n > l - k > 0$.

Параграф 2.5 посвящен изучению сумм рекордных величин $V_{n,m} = X(N(m)+1) + ... + X(N(n))$ и сумм последовательных максимумов $T_{n,m} = X_{m,m} + X_{m+1,m+1} + ... + X_{n,n}$. Отметим, что исходные величины $X_1, X_2, \ldots$ предполагаются независимыми, с общей непрерывной функцией распределения $F(x)$, нулевым математическим ожиданием и единичной дисперсией. При исследовании свойств этих сумм были получены следующие результаты.

Теорема 2.6. *Для любых* $n, m = 1, 2, ..., 0 \leq m < n$ *имеет место неравенство*

$$EV_{n,m} \leq \gamma_{n,m}, \tag{3}$$

где

$$\gamma_{n,m} = \left(\int_0^1 \left(\frac{x^m - x^n}{1-x} \right)^2 dx - \left(\sum_{k=m+1}^{n} \frac{1}{k} \right)^2 \right)^{\frac{1}{2}},$$

причем равенство в (3) достигается тогда и только тогда, когда

$$G(x) = \frac{\frac{x^m - x^n}{1-x} - \sum\limits_{k=m+1}^{n} \frac{1}{k}}{\gamma_{n,m}}, 0 < x < 1,$$

где G – функция, обратная функции распределения F.

Теорема 2.7. *Справедливо следующее неравенство*

$$ET_{n,m} \leq \sigma_{n,m}, \tag{4}$$

где

$$\sigma_{n,m} = \left(\int\limits_0^1 \left(\sum_{k=m}^{n} kx^{k-1} \right)^2 dx - (n-m+1)^2 \right)^{\frac{1}{2}},$$

причем равенство в (4) достигается тогда и только тогда, когда

$$G(x) = F^{-1}(x) = \frac{\sum\limits_{k=m}^{n} kx^{k-1} - (n-m+1)}{\sigma_{n,m}}.$$

В **параграфе 2.6** вводится понятие обобщенного выборочного размаха

$$W_{k,n} = (X_{n,n} + ... + X_{n-k+1,n}) - (X_{1,n} + ... + X_{k,n}).$$

Для этой статистики также получен результат, связанный с оцениванием математического ожидания и характеризацией максимизирующего распределения.

Теорема 2.8. *Для любых $k = 1, 2, \ldots,$ $n = 2, 3, \ldots,$ $k \leq n$ имеет* *место неравенство*

$$EW_{k,n} \le knC_{n-1}^{k} \left(2 \int\limits_{\frac{1}{2}}^{1} \left(\int\limits_{1-u}^{u} v^{n-k-1}(1-v)^{k-1} dv \right)^{2} du \right)^{\frac{1}{2}},$$

равенство в котором достигается, когда $G = F^{-1}$ *имеет вид:*

$$G(u) = \frac{1}{\mu} \left(knC_{n-1}^{k} \int\limits_{1-u}^{u} z^{n-k-1}(1-z)^{k-1} dz \right), \quad \frac{1}{2} \le u \le 1.$$

Далее было исследовано предельное поведение $EW_{k,n}$, при $n \to \infty$.

Теорема 2.9 *При* $n \longrightarrow \infty$ *и* $0 < \alpha < \frac{1}{2}$ *максимальное значение величины* $E\left(\frac{W_{\alpha n,n}}{n}\right)$ *стремится к* $\sqrt{2\alpha}$, *а последовательность* $F_1, F_2, \ldots$ *соответствующих максимизирующих функций распределения стремится к функции трехточечного симметричного распределения, сосредоточенного в точках* $-\frac{1}{\sqrt{2\alpha}}$, 0 *и* $\frac{1}{\sqrt{2\alpha}}$, *с соответствующими весами* α, $1-2\alpha$ *и* α.

В конце проводится сравнение обобщенного выборочного размаха с известной статистикой – выборочным дифференциалом (Selection Differential), который был определен в указанной выше работе Нагараджи [8].

В **заключении** подводится итог проделанной работы.

Автор пользуется случаем поблагодарить своего научного руководителя Валерия Борисовича Невзорова за постановку интересных задач, постоянное внимание к работе и многочисленные советы, данные в течение всего времени руководства.

Список литературы

[1] Дэйвид Г. Порядковые статистики. – М.: Наука, 1979.

[2] Невзоров В.Б. Рекорды. Математическая теория. – М.: ФАЗИС, 2000.

[3] Ahsanullah M. Record statistics. – New York: Nova Science Publishers, 1995.

[4] Ahsanullah M. Record Values – Theory and Applications. – Lanham: University Press of America, 2004.

[5] Akhundov I.S., Berred A., Nevzorov V.B., On the influence of record terms in the addition of independent random variables // Communications in Statistics: Theory and Methods, 2007, v.36, n. 7, 1291-1303.

[6] Arnold B.C., Balakrishnan N., Nagaraja H.N. Records. – New York: Wiley, 1998.

[7] Beutner E., Kamps U. Random contraction and random dilation of generalized order statistics // Commun. Statist. Theory Meth., 2008, 37, pp. 2185-2201.

[8] Nagaraja H.N. Some finite sample results for the selection differential // Ann. Inst. Statist. Math., 1981, 33, pp. 437-448.

[9] Navarro J. Characterizations by power contractions of order statistics // Commun. Statist. Theory Meth., 2008, 37, pp. 987-997.

[10] Resnick S.I. Limit laws for record values // Stohast. Processes Appl., 1(1), 1973, pp. 67-82.

[11] Smirnoff N.V. Sur la distribution de ω^2 // C. R. Acad. Sci. Paris, 1936, 202, 449.

[12] Wesolowsky J., Ahsanullah M. Switching order statistics through random power contractions // Austral. NZ J. Statist., 2004, 46, pp. 297-303.

Публикации автора по теме диссертации

[13] Сагателян В.К. Характеризация распределений равенством порядковых статистик // Записки научных семинаров ПОМИ, 2007, т. 341, с. 168-173.

[14] Сагателян В.К. Характеризации распределений моментными свойствами сумм рекордов и последовательных максимумов // Записки научных семинаров ПОМИ, 2008, т. 361, с. 138-144.

[15] Сагателян В.К. Об одной новой модели рекордных величин // Вестн. С.-Петербург. ун-та. Сер. 1. 2008. Вып.(3) с. 144-147.

[16] Nevzorov V.B., Sagateljan V.K. Characterizations of distributions by equalities of order statistics // Journal of Stat. Theory and Appl., 2007, vol. 6, 3, pp. 264-271.