На правах рукописи

БАКЛАНОВ Евгений Анатольевич

ВЕРОЯТНОСТНЫЕ НЕРАВЕНСТВА И ПРЕДЕЛЬНЫЕ ТЕОРЕМЫ ДЛЯ ОБОБЩЕННЫХ L-СТАТИСТИК

01.01.05 – теория вероятностей и математическая статистика

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата физико-математических наук

Новосибирск – 2002

Работа выполнена на кафедре теории вероятностей и математической статистики Новосибирского государственного университета

Научный руководитель:	д.ф.-м.н., профессор И. С. Борисов
Официальные оппоненты:	д.ф.-м.н., профессор А. А. Могульский д.ф.-м.н., профессор А. И. Саханенко
Ведущая организация:	Омский государственный университет

Защита состоится 17 апреля 2002 г. в 17 часов на заседании диссертационного совета Д 003.015.01 в Институте математики им. С. Л. Соболева СО РАН по адресу: 630090, Новосибирск, пр. Коптюга, 4, к. 417.

С диссертацией можно ознакомиться в библиотеке Института математики им. С. Л. Соболева СО РАН.

Автореферат разослан 15 марта 2002 г.

Ученый секретарь
диссертационного совета
д.ф.-м.н. Ю. В. Шамардин

Общая характеристика работы

Актуальность темы. Пусть $X_1, X_2, \ldots$ – независимые одинаково распределенные случайные величины. Обозначим через $X_{n:1} \leq \ldots \leq X_{n:n}$ – порядковые статистики, построенные по выборке $\{X_i, i \leq n\}$. Рассмотрим линейную комбинацию порядковых статистик

$$L_n = \sum_{i=1}^{n} c_{ni} X_{n:i},$$

называемую (классической) *L-статистикой.* L-статистики находят широкое применение, в частности, в теории оценивания. Они используются, например, при оценке параметров сдвига и масштаба (см. [5, 9]). Для некоторых параметрических семейств коэффициенты c_{ni} могут быть подобраны так, чтобы L-статистики были в известном смысле эквивалентны оценкам максимального правдоподобия (их дисперсии асимптотически эквивалентны), которые, как правило, являются оптимальными.

Наиболее общей формой L-статистик, встречающейся в литературе, являются аддитивные функционалы от порядковых статистик следующего вида:

$$\Phi_n = \sum_{i=1}^{n} h_{ni}(X_{n:i}), \tag{GL}$$

где $h_{ni} : \mathbb{R} \to \mathbb{R}$, $i = 1, \ldots, n$, – некоторые измеримые функции. В частности, если $h_{ni}(y) = c_{ni} h(y)$ и функция h монотонна, то Φ_n – классическая L-статистика.

Функционалы вида (GL) в рассматриваемой общности естественно называть *обобщенными L-статистиками.* Они впервые были введены в [3, 4], где были получены асимптотические разложения для распределений этих статистик в некоторых частных случаях. Анализ Фурье распределений Φ_n содержится в [6]. Отметим, что интегральные статистики (интегральные функционалы от эмпирических функций распределения, например, статистики

Крамера – Андерсона – Дарлинга) представимы в виде (GL), но не в виде классических L-статистик (подробнее см. [3, 6]).

Цель работы. Основной целью работы является получение моментных неравенств и показательных оценок для хвостов распределения обобщенных L-статистик, а также доказательство асимптотической нормальности и исследование асимптотики вероятностей больших уклонений указанных статистик.

Методика исследований основана на общих методах теории вероятностей (в частности, на использовании вероятностных неравенств для сумм независимых банаховозначных случайных элементов).

Научная новизна. В работе предложены подходы, позволяющие сводить анализ обобщенных L-статистик к изучению сумм независимых случайных элементов в некотором функциональном банаховом пространстве, что дало возможность воспользоваться известными вероятностными неравенствами для сумм независимых банаховозначных случайных элементов. Полученные неравенства используются при изучении вероятностей больших уклонений рассматриваемых статистик.

В большинстве работ, посвященных асимптотическому анализу L-статистик, чаще всего рассматривался случай так называемых *регулярных* коэффициентов c_{ni}, когда

$$c_{ni} = n^{-1}J(i/(n+1)) \qquad \text{или} \qquad c_{ni} = \int_{(i-1)/n}^{i/n} J(t)dt,$$

где J – некоторая достаточно гладкая функция, или *асимптотически регулярных* коэффициентов, когда c_{ni} вычисляются по вышеприведенным формулам с точностью до $o(1/n)$ равномерно по всем i (см., например, [1, 2, 10, 11]). Предложенный в настоящей диссертации метод позволяет отказаться от вышеупомянутых условий регулярности коэффициентов c_{ni} при доказательстве асимптотической нормальности L-статистик и выводе вероятностных неравенств для них.

Практическая ценность. Работа носит теоретический характер.

Апробация работы. Результаты диссертации докладывались на объединенном семинаре кафедры теории вероятностей и математической статистики НГУ и лаборатории теории вероятностей и математической статистики Института математики СО РАН под руководством академика А. А. Боровкова. Результаты работы также докладывались на XXXV и XXXVI международных научных студенческих конференциях (Новосибирск, 1997 и 1998 гг.).

Публикации. По теме диссертации опубликованы 4 работы, которые приведены в конце автореферата.

Структура и объем диссертации. Объем диссертации 55 страниц. Она состоит из введения, двух глав и списка литературы, содержащего 36 наименований.

Используемая в автореферате нумерация теорем и формул не совпадает с соответствующей нумерацией в диссертации.

Содержание работы

В § 1. 1. изучаются статистики вида (GL), построенные по выборке из показательного распределения. Получены экспоненциальные оценки для хвостов распределения указанных статистик, а также моментные неравенства для них. В § 1. 2. получены аналогичные оценки для статистик вида (GL), построенных по выборке из равномерного на отрезке $[0, 1]$ распределения. Отметим, что с помощью квантильных преобразований (т.е. несколько видоизменив ядра обощенной L-статистики) мы можем свести задачу к анализу статистик, построенных по выборке с наперед заданным непрерывным распределением – например, с показательным или равномерным. Традиционный в научной литературе выбор этих двух распределений объясняется удобной для анализа структурой порядковых статистик.

Аналогичные оценки получены также и для линейных комбинаций функций от порядковых статистик (так называемых L-статистик с расщепляющимися ядрами), причем на распределение выборки не налагается дополнительных ограничений, кроме моментных, и не требуется монотонности ядра и регулярного представления коэффициентов c_{ni}.

Предлагаемый в настоящей работе подход при выводе вышеупомянутых результатов иллюстрирует возможности бесконечномерного анализа: рассматриваемые задачи сводятся к аналогичным проблемам для сумм независимых случайных элементов со значениями в некотором функциональном банаховом пространстве. Так, предложенный метод позволяет получать значительно более точные по сравнению с [1] неравенства для хвостов распределения L-статистик и отказаться от регулярного представления весов.

Во второй главе доказана предельная теорема для одного класса обобщенных L-статистик, построенных по набору центрированных порядковых статистик, когда в качестве предельного распределения выступает интегральный функционал от некоторого гауссовского случайного процесса. Кроме того, в этой же главе изучается асимптотическая нормальность обобщенных L-статистик, построенных по выборке из равномерного на отрезке $[0, 1]$ распределения, и L-статистик с расщепляющимися ядрами. В частности, для асимптотической нормальности последних существенно ослаблены ограничения на веса, накладываемые в [11]. Эти ограничения по существу означают, что коэффициенты c_{ni} асимптотически регулярны, в то время как предложенные в настоящей диссертации условия на c_{ni} могут и не удовлетворять этому требованию. В третьем параграфе второй главы исследуется асимптотика вероятностей больших уклонений (в зоне нормальных уклонений) статистик вида (GL), построенных по выборке из показательного распределения. Аналогичные результаты получены в [7]. В настоящей работе по сравнению с [7] расширяется класс L-статистик, допускающий степенные зоны уклонений в соответствующем асимптотическом представлении хвостов распределения.

Рассмотрим центрированную обобщенную L-статистику

$$\overline{A}_n = \sum_{i=1}^{n} h_{ni}(U_{n:i}) - \sum_{i=1}^{n} h_{ni}(\mathbf{E}U_{n:i}), \tag{1}$$

построенную по выборке из равномерного на отрезке $[0,1]$ распределения.

Теорема 1. *Пусть функции* $\{h_{ni};\, i \le n\}$ *в* (1) *удовлетворяют на* $[0,1]$ *условию Липшица с константами* b_{ni} *соответственно. Тогда*

$$\mathbf{P}\{\overline{A}_n \ge y\} \le 4 \exp\left\{ -\frac{y^2 - 8\beta_1 y}{8(B_1^2 + H_1 y)} \right\}, \tag{2}$$

где

$$\begin{aligned} \beta_1 &= \frac{1}{n+1}\sum_{i=1}^{n} i^{1/2} b_{ni}, \quad B_1^2 = \frac{2}{(n+1)^2}\sum_{i=1}^{n}\left(\sum_{j=i}^{n} b_{nj}\right)^2, \\ H_1 &= \frac{1}{n+1}\sum_{i=1}^{n} b_{ni}. \end{aligned}$$

Рассмотрим обобщенные L-статистики, построенные по выборке из показательного с параметром 1 распределения:

$$\overline{\Phi}_n = \sum_{i=1}^{n} h_{ni}(X_{n:i}) - \sum_{i=1}^{n} h_{ni}(\mathbf{E}X_{n:i}). \tag{3}$$

Теорема 2. *Пусть функции* $\{h_{ni};\, i \le n\}$ *в* (3) *непрерывно дифференцируемы на* $[0,\infty)$ *и при всех* $x \ge 0$

$$|h'_{ni}(x)| \le \alpha_{ni} + \beta_{ni} x^p \quad \textit{для некоторого} \quad p \ge 0, \tag{4}$$

где α_{ni} *и* β_{ni} *– положительные постоянные, зависящие только от* i *и* n*. Тогда для всех* $r \ge 2$

$$\begin{aligned} \mathbf{E}|\overline{\Phi}_n|^r &\le 8^{r-1}\beta_1^r + (K_1 r)^r \left\{ \Gamma(r+1)\widetilde{B}_{n,r} + \widetilde{B}_{n,2}^{r/2} \right\} + (4\beta_2)^{r(p+1)} \\ &+ (K_2 r p)^{r(p+1)} \left\{ \Gamma(r(p+1)+1)\overline{B}_{n,r} + \overline{B}_{n,2/(p+1)}^{r(p+1)/2} \right\}, \end{aligned} \tag{5}$$

где β_k, $k=1,2$, $\overline{B}_{n,r}$, $\widetilde{B}_{n,r}$ выражаются через коэффициенты a_{ni} и b_{ni}, в диссертации приведен их явный вид, K_1 и K_2 – положительные абсолютные постоянные. В диссертации приводится пример, показывающий, что неравенства (2) и (5) достаточно точные.

Рассмотрим L-статистики следующего вида:

$$L_n = \sum_{i=1}^{n} c_{ni} h(X_{n:i}), \tag{6}$$

где c_{ni}, $i=1,\ldots,n$, – некоторые постоянные, h – произвольная измеримая (не обязательно монотонная) функция, а X_1 имеет произвольную функцию распределения F.

Без ограничения общности можно считать, что $\sum\limits_{i=1}^{n} c_{ni} = 0$, поскольку статистика L_n представима в следующем виде:

$$L_n = \sum_{i=1}^{n} \tilde{c}_{ni} h(X_{n:i}) + \tilde{c}_n \sum_{i=1}^{n} h(X_i), \tag{7}$$

где $\tilde{c}_{ni} = c_{ni} - \tilde{c}_n$, $\tilde{c}_n = n^{-1} \sum\limits_{i=1}^{n} c_{ni}$, а второе слагаемое в правой части (7) представляет собой сумму независимых одинаково распределенных случайных величин, для которых хорошо известны как моментные неравенства, так и оценки для хвостов распределения.

Введем в рассмотрение функцию $\varphi_n(x)$:

$$\varphi_n(x) = nc_{nk}x + \sum_{i=1}^{k} c_{ni} - kc_{nk}, \quad \frac{k-1}{n} < x < \frac{k}{n}, \quad k=1,\ldots,n.$$

Далее мы считаем, что функция h имеет ограниченную вариацию на каждом конечном отрезке и непрерывна слева. Введем следующие обозначения:

$$\gamma_n = \int_{\mathbb{R}} \varphi_n(F(t))dh(t), \quad H_F = \int_{\mathbb{R}} \Big(F(t)(1-F(t))\Big)^{1/2} |dh(t)|,$$

$$
\begin{aligned}
\alpha_k &\equiv \alpha(k,F,h) = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} g(t,z)|dh(t)| \right)^k dF(z), \\
g(t,z) &= \begin{cases} F(t) & \text{при} \quad t \le z, \\ 1 - F(t) & \text{при} \quad t > z, \end{cases} \\
c_n &= \max_{1 \le k \le n} |c_{nk}|,
\end{aligned}
$$

где $|dh(t)|$ – мера полной вариации, порожденная функцией h. Всюду в дальнейшем существование интеграла $\int f dh$ подразумевает конечность $\int |f||dh|$.

Теорема 3. *Пусть $H_F < \infty$ и $\mathbf{E}|h(X_1)|^k \le k!B^2H^{k-2}/2$ для некоторых постоянных B^2 и $H > 0$ и всех целых $k \ge 2$. Тогда*

$$
\mathbf{P}\{L_n + \gamma_n \ge y\} \le \exp\left\{ -\frac{y^2 - 2H_F c_n \sqrt{n} y}{4c_n(2nc_nB^2 + yH)} \right\}. \tag{8}
$$

З а м е ч а н и е. Зависимость от y в показателе экспоненты последнего неравенства по существу такая же, что и в классическом неравенстве С. Н. Бернштейна для сумм независимых случайных величин. Неравенство (8) существенно уточняет аналогичное неравенство из [1] в зоне уклонений $y \gg n^{1/6}$.

Теорема 4. *Пусть $H_F < \infty$ и $\alpha_k < \infty$ для некоторого $k \ge 2$. Тогда*

$$
\mathbf{E}|L_n + \gamma_n|^k \le 2^{k-1} c_n^k \Big((Ck)^k \alpha_k + H_F^k \Big) n^{k/2}, \tag{9}
$$

где C – абсолютная положительная постоянная.

Рассмотрим обобщенные L-статистики

$$
A_n = \sum_{i=1}^n h_{ni}(U_{n:i} - i/(n+1)), \tag{10}
$$

построенные по выборке из равномерного на отрезке $[0,1]$ распределения. Рассмотрим также гауссовский случайный процесс $w^0(t)$, $0 \le t \le 1$, с нулевым средним и корреляционной функцией $\min\{s,t\} - st$, $0 \le s,t \le 1$. Символом $\to^d$ будем обозначать слабую сходимость распределений.

Теорема 5. *Пусть существует функция $\varphi(t,x)$, непрерывная по паре переменных (t,x), $0 \le t \le 1$, $x \in \mathbb{R}$, такая, что для всех $x \in \mathbb{R}$*

$$\max_{1\le i\le n} \sup_{(i-1)/n<t\le i/n} |nh_{ni}(x/\sqrt{n}) - \varphi(t,x)| \le \varepsilon_n \psi(x), \tag{11}$$

где $\varepsilon_n \to 0$ при $n \to \infty$, $\psi(x) \ge 0$ – произвольная непрерывная функция. Тогда при $n \to \infty$

$$A_n \to^d \int_0^1 \varphi(t, w^0(t))dt. \tag{12}$$

Рассмотрим центрированную обобщенную L-статистику

$$\overline{A}_n = \sum_{i=1}^n h_{ni}(U_{n:i}) - \sum_{i=1}^n h_{ni}(\mathbf{E}U_{n:i}), \tag{13}$$

также построенную по выборке из равномерного на отрезке $[0,1]$ распределения и имеющую гладкие ядра.

Для всех $t \in ((i-1)/n, i/n]$, $i = 1,\ldots,n$, положим

$$\alpha_n(t) = n^{1/2} h'_{ni}(i/(n+1)).$$

Обозначим

$$\sigma_n^2 = \int_0^1 \int_0^1 \alpha_n(x)\alpha_n(y)(\min\{x,y\} - xy)dxdy.$$

Так как σ_n^2 – второй момент некоторой гауссовской случайной величины, то $\sigma_n^2 \ge 0$. Отметим, что если $h'_{ni}(i/(n+1)) \ne 0$ хотя бы для одного i, то $\sigma_n^2 > 0$. Обозначим также $\mathbf{N}(0,1)$ случайную величину, имеющую стандартное нормальное распределение.

Теорема 6. *Пусть функции $\{h_{ni}; i \le n\}$ в (13) непрерывно дифференцируемы на $[0,1]$ и удовлетворяют следующим условиям:*

$$|h'_{ni}(x) - h'_{ni}(y)| \le b_{ni}|x-y|^{\alpha}, \quad 0 < \alpha \le 1, \tag{14}$$

$$\sum_{i=1}^{n} b_{ni} = o\left(n^{(\alpha+1)/2}\sigma_n\right), \tag{15}$$

$$\sum_{i=1}^{n} \left|h'_{ni}(i/(n+1))\right| = o(n\sigma_n(\ln n)^{-1}). \tag{16}$$

Тогда при $n \to \infty$

$$\sigma_n^{-1}\overline{A}_n \to^d \mathbf{N}(0,1). \tag{17}$$

Рассмотрим L-статистики вида (6), где ядро h имеет ограниченную вариацию на каждом конечном отрезке и непрерывно слева. Функцию $c_n(t)$ определим следующим образом:

$$c_n(t) = nc_{ni}, \quad t \in ((i-1)/n, i/n], \quad 1 \le i \le n, \quad c_n(0) = nc_{n1}.$$

Введем следующие обозначения (при условии существования соответствующих интегралов):

$$\begin{aligned}
\mu_n &= \int_{\mathbb{R}} h(t)d\varphi_n(F(t)),\\
\sigma_n^2 &= \int_{\mathbb{R}}\int_{\mathbb{R}} c_n(F(x))c_n(F(y))\Big(F(\min\{x,y\}) - F(x)F(y)\Big)dh(x)dh(y),\\
\tilde{c}_n &= \max_{1\le i\le n-1} |c_{n,i+1} - c_{ni}|,\\
Y_{n1} &= \int_{\mathbb{R}} c_n(F(t))\Big(F(t) - \mathbf{I}\{X_1 < t\}\Big)dh(t).
\end{aligned}$$

Теорема 7. *Пусть* $0 < \sigma_n < \infty$ *при всех* n,

$$\mathbf{E}Y_{n1}^2\mathbf{I}\left\{|Y_{n1}| \ge \varepsilon\sigma_n\sqrt{n}\right\} = o(\sigma_n^2) \quad \text{для любого} \quad \varepsilon > 0, \tag{18}$$

$$\tilde{c}_n = o(n^{-3/2}\sigma_n), \tag{19}$$

$$\int_{\mathbb{R}} F(t)(1 - F(t))|dh(t)| < \infty. \tag{20}$$

Тогда при $n \to \infty$

$$\frac{\sqrt{n}(L_n - \mu_n)}{\sigma_n} \to^d \mathbf{N}(0,1). \tag{21}$$

З а м е ч а н и е. В теореме 7 не требуется равномерного притяжения ступенчатой функции $c_n(t)$ к некоторой ограниченной функции, тогда как, например, в [11] такое условие является существенным. Отметим также, что в теореме 7 ослаблены условия гладкости на ядро h по сравнению с [8].

Рассмотрим обобщенные L-статистики, построенные по выборке из показательного распределения и ядрам h_{ni}, производные которых удовлетворяют условию Гельдера с показателем α:

$$\Phi_n = \sum_{i=1}^{n} h_{ni}(X_{n:i}). \tag{22}$$

Во второй главе приведены некоторые достаточные условия для справедливости соотношения

$$\mathbf{P}\{(\Phi_n - \mu_n)/\sigma_n \geq x\} = (1 - \Phi(x))(1 + o(1)), \tag{23}$$

где μ_n и σ_n – некоторые постоянные, $\Phi(x)$ – функция распределения стандартного нормального закона. В частности, при выполнении условий работы [7] допустимая зона уклонений в (22) $x = o(n^{\alpha/(4+2\alpha)})$. Предложенный в настоящей работе метод позволяет, в отличие от [7], получать степенную зону применимости соотношения (23) для статистик более общего вида.

Литература

[1] *Алешкявичене А.К.* О больших уклонениях для линейных комбинаций порядковых статистик // *Литов. мат. сб.* 1989. Т. 29, № 2. С. 212–222.

[2] *Алешкявичене А.К.* Большие и умеренные уклонения для L-статистик // *Литов. мат. сб.* 1991. Т. 31, № 2. С. 227–241.

[3] *Зитикис Р.* О гладкости функции распределения $\mathcal{FL}$-статистики. **I** // *Литов. мат. сб.* 1990. Т. 30, № 2. С. 233–246.

[4] *Зитикис Р.* О гладкости функции распределения $\mathcal{FL}$-статистики. **II** // *Литов. мат. сб.* 1990. Т. 30, № 3. С. 499–512.

[5] *Blom G.* Statistical estimates and transformed beta-variables. N.-Y., John Wiley and Sons, 1958.

[6] *Borisov I.S.* Bounds for characteristic functions of additive functionals of order statistics // *Siberian Adv. Math.* 1995. V. 5, N 4. P. 1–15.
[7] *Callaert H., Vandemaele M., Veraverbeke N .* A Cramér type large deviation theorem for trimmed linear combinations of order statistics // *Commun. Statist.-Theor. Meth.* 1982. 11(23), P. 2689–2698.
[8] *Chernoff H., Gastwirth J.L., and Johns M.V., Jr.* Asymptotic distribution of linear combinations of order statistics, with applications to estimation // *Ann. Math. Statist.* 1967. V. 38. P. 52–72.
[9] *Lloyd E.H.* Least-squares estimation of location and scale parameters using order statistics // *Biometrica.* 1952. V. 39. P. 88–95.
[10] *Mason D.M., Shorack G.R.* Necessary and sufficient conditions for asymptotic normality of L-statistics // *Ann. Probab.* 1992. V. 20, N 4. P. 1779–1803.
[11] *Norvaiša R.* The central limit theorem for L-statistics // *Studia Scien. Math. Hung.* 1997. V. 33. P. 209–238.

Публикации по теме диссертации

[12] *Борисов И.С., Бакланов Е.А.* Моментные неравенства для обобщенных L-статистик // *Сиб. мат. журн.* 1998. Т. 39, № 3. С. 483–489.
[13] *Борисов И.С., Бакланов Е.А.* Вероятностные неравенства для обобщенных L-статистик // *Сиб. мат. журн.* 2001. Т. 42, № 2. С. 258–274.
[14] *Бакланов Е.А.* Моментные неравенства для обобщенных L-статистик // *Материалы XXXV международной научной студенческой конференции «Студент и научно-технический прогресс». Математика.* Новосибирск. 1997. С. 11–12.
[15] *Бакланов Е.А.* Вероятностные неравенства для обобщенных L-статистик // *Материалы XXXVI международной научной студенческой конференции «Студент и научно-технический прогресс». Математика.* Новосибирск. 1998. С. 10-11.

Подписано в печать 5 марта 2002 г. Уч.-изд. л. 1
Офсетная печать. Формат 60x84 1/16.
Заказ № Тираж 100 экз.

Лицензия ЛР № 021285 от 6 мая 1998г.

Издательский центр НГУ;
630090, Новосибирск, ул. Пирогова, 2.