Philosophy and Methodology of Informatics

ЗАКОН БОЛЬШИХ ЧИСЕЛ ДЛЯ ГИПЕРСЛУЧАЙНОЙ ВЫБОРКИ

Игорь Горбань

***Аннотация**: Доказаны две теоремы, определяющие условия сходимости выборочного среднего гиперслучайной величины и алгоритм расчета состоятельных оценок границ начальных моментов гиперслучайных величин при быстрых изменениях статистических условий. Доказана для гиперслучайных событий теорема, определяющая предельные границы функции распределения частоты гиперслучайных событий при устремлении объема выборки к бесконечности. Доказано, что нижняя граница математического ожидания гиперслучайной величины не меньше математического ожидания верхней границы распределения, а верхняя граница математического ожидания гиперслучайной величины не больше математического ожидания нижней границы распределения.*

***Ключевые слова**: теория гиперслучайных явлений, выборка, сходимость.*

***ACM Classification Keywords**: G3 Probability and Statistics*

***Conference**: The paper is selected from XV[th] International Conference "Knowledge-Dialogue-Solution" KDS-2 2009, Kyiv, Ukraine, October, 2009.*

Введение

Недавно была опубликована монография [Горбань, 2007], посвященная новой теории гиперслучайных явлений, под которыми подразумевается множество условных случайных явлений – множество условных случайных событий, величин и функций. Гиперслучайную величину X наиболее полно характеризуют условные функции распределения $F_{x/g}(x)$, соответствующие множеству условий $g \in G$, менее полно – верхняя и нижняя границы функции распределения $F_{Sx}(x)$, $F_{Ix}(x)$. Представление о гиперслучайной величине дают начальные и центральные моменты границ распределения, в частности, математические ожидания границ m_{Sx}, m_{Ix}, дисперсии границ D_{Sx}, D_{Ix} и пр., а также границы моментов, к числу которых относятся границы математического ожидания m_{sx}, m_{ix}, границы дисперсии D_{sx}, D_{ix} и др.

В книге особое внимание уделено методам формирования соответствующих состоятельных оценок: оценок условных функций распределения $F^*_{x/g}(x)$, оценок границ функции распределения $F^*_{Sx}(x)$, $F^*_{Ix}(x)$, оценок математических ожиданий m^*_{Sx}, m^*_{Ix} этих границ, оценок дисперсий границ D^*_{Sx}, D^*_{Ix}, оценок границ математического ожидания m^*_{sx}, m^*_{ix}, оценок границ дисперсии D^*_{sx}, D^*_{ix} и пр. Исследована сходимость этих оценок к соответствующим точным характеристикам и параметрам. Доказан ряд теорем, определяющих условия сходимости.

Описанная в книге процедура формирования оценок построена по следующей схеме. Для всего множества G статистических условий g формируется выборка

$$\vec{x} = \{x_1, ..., x_N / g \in G\},$$

по ней для каждого условия g в отдельности рассчитывается оценка условных характеристик и условных параметров: оценка условной функции распределения $F^*_{x/g}(x)$, оценка условного математического ожидания $m^*_{x/g}$, оценка условной дисперсии $D^*_{x/g}$ или др. По оценкам условной функции распределения $F^*_{x/g}(x)$ вычисляются оценки границ функции распределения

$$F^*_{Sx}(x) = \sup_{g \in G} F^*_{x/g}(x), \quad F^*_{Ix}(x) = \inf_{g \in G} F^*_{x/g}(x)$$

и оценки характеристик, характеризующие эти границы: оценки математических ожиданий границ m^*_{Sx}, m^*_{Ix}, оценки дисперсий границ D^*_{Sx}, D^*_{Ix} и пр. По оценкам условных параметров определяются оценки границ соответствующих величин, например, по оценкам условных математических ожиданий $m^*_{x/g}$ – оценки границ математического ожидания $m^*_{sx} = \sup_{g \in G} m^*_{x/g}$, $m^*_{ix} = \inf_{g \in G} m^*_{x/g}$, по оценкам условных дисперсий $D^*_{x/g}$ – оценки границ дисперсии $D^*_{sx} = \sup_{g \in G} D^*_{x/g}$, $D^*_{ix} = \inf_{g \in G} D^*_{x/g}$ и т.д.

В работах [Горбань, 2007, Горбань, 2006] отмечалось, что определенные трудности можно ожидать при формировании выборки $\vec{x} = \{x_1, ..., x_N / g \in G\}$ из-за сложности обеспечения, контроля и поддержания условий $g \in G$. Однако, в ряде случаев вопрос облегчается тем, что для расчета искомых характеристик не требуются знания того, в каких именно условиях получены условные характеристики или параметры. Главное, чтобы были представлены все возможные условия g множества G и в массив данных, используемый для расчета условных характеристик и параметров, не попадали данные, соответствующие другим условиям.

Иногда можно предположить, что статистические условия меняются непрерывно и относительно медленно. Кроме того, может быть известна максимально возможная скорость изменения условий. На основе этих данных нетрудно рассчитать максимальное число последовательных элементов выборки $N_{\max}$, для которых условия остаются практически неизменными. Если объем выборки $N_{\max}$ достаточен для получения условных оценок приемлемого качества, то алгоритм нахождения искомых граничных оценок сводится к сбору данных на большом интервале наблюдения, разбиению полученного массива на фрагменты по $N_{\max}$ последовательных элементов, расчету по ним промежуточных условных оценок, а затем вычислению граничных оценок.

Описанный алгоритм достаточно прост и прозрачен. Его недостатком является то, что требуется вспомогательная априорная информации о скорости изменении статистических условий и, то, что он применим лишь при достаточно медленной смене условий.

До настоящего времени остаются не ясными целый ряд вопросов, касающихся различных оценок характеристик и параметров гиперслучайных явлений.

Целью настоящей статьи является прояснение некоторых из них на основе доказываемых ниже теорем для гиперслучайных событий и величин. Первая теорема, аналогичная известной теореме Чебышева, определяет условия сходимости выборочного среднего гиперслучайной величины, вторая теорема, базирующаяся на первой, обосновывает возможность расчета состоятельных оценок начальных моментов гиперслучайных величин при быстрых изменениях статистических условий. Третья теорема, аналогичная теореме Бернулли [Королюк, 1985, Гнеденко, 1961], определяет условия сходимости границ частоты гиперслучайного события, четвертая теорема устанавливает связь между математическими ожиданиям границ распределения и границами математического ожидания.

Теорема о сходимости границ выборочного среднего

Теорема 1. Пусть гиперслучайная величина $X = \{X / g \in G\}$ представляет собой совокупность случайных величин X / g для различных статистических условий $g \in G$ с условными математическими ожиданиями $m_{x/g}$ и ограниченными условными дисперсиями $D_{x/g}$. Нижняя и верхняя границы математического ожидания гиперслучайной величины X равны соответственно m_{ix}, m_{sx}. Из генеральной совокупности гиперслучайной величины X, полученной в неконтролируемо меняющихся статистических условиях, формируется гиперслучайная выборка $\{X_1, \ldots, X_N\}$ объемом N. По этой выборке рассчитывается гиперслучайное выборочное среднее $m_x^* = \frac{1}{N}\sum_{n=1}^{N} X_n$.

Тогда при устремлении объема выборки к бесконечности ($N \to \infty$) гиперслучайное выборочное среднее m_x^* сходится по вероятности к множеству чисел $m_x = \{m_{x/\vec{g}},\ \vec{g} = (g_1, ..., g_N),\ g_n \in G,\ n = \overline{1, N}\}$, представляющих собой множество средних $m_{x/\vec{g}} = \frac{1}{N}\sum_{n=1}^{N} m_{x_n/g_n}$ условных математических ожиданий $m_{x_1/g_1}, \ldots, m_{x_N/g_N}$ случайных величин $X_1 / g_1, \ldots, X_N / g_N$, рассчитанных для всевозможных $g_n \in G,\ n = \overline{1, N}$, а нижняя и верхняя границы этого выборочного среднего сходятся соответственно к нижней и верхней границам математического ожидания гиперслучайной величины X:

$$\begin{aligned} &\lim_{N\to\infty} P(\left|\inf m_x^* - m_{ix}\right| > \varepsilon) = 0, \\ &\lim_{N\to\infty} P(\left|\sup m_x^* - m_{sx}\right| > \varepsilon) = 0, \end{aligned} \tag{1}$$

где $P(z)$ – вероятность условия z, а ε – как угодно малое положительное число.

Для доказательства теоремы рассмотрим случайную выборку $\{X_1 / g_1, \ldots, X_N / g_N\}$, полученную при фиксированной последовательности условий $g_1, \ldots, g_N \in G$. Рассчитанное по ней выборочное среднее $m_{x/\vec{g}}^* = \frac{1}{N}\sum_{n=1}^{N} X_n / g_n$ при устремлении объема выборки N к бесконечности согласно теореме Чебышева сходится по вероятности к среднему условных математических ожиданий $m_{x/\vec{g}}$:

$$\lim_{N\to\infty} P(\left|m_{x/\vec{g}}^* - m_{x/\vec{g}}\right| > \varepsilon) = 0 .$$

Сходимость случайной величины $m_{x/\vec{g}}^*$ к $m_{x/\vec{g}}$ для всех $\vec{g} = (g_1, ..., g_N),\ g_n \in G,\ n = \overline{1, N}$ означает сходимость по вероятности гиперслучайного выборочного среднего m_x^* к множеству чисел m_x.

При любой фиксированной последовательности условий и $N \to \infty$ выборочное среднее $m^*_{x/\vec{g}}$ ограничено интервалом $[\inf m^*_x,\ \sup m^*_x]$, а среднее условных математических ожиданий $m_{x/\vec{g}}$ – интервалом $[m_{ix},\ m_{sx}]$. При этом минимальному среднему значению m_{ix} среднего математического ожидания соответствует нижняя граница $\inf m^*_x$ выборочного среднего, а максимальному среднему значению m_{sx} – верхняя граница $\sup m^*_x$ выборочного среднего, т. е. справедливо равенство (1).

Теорема о сходимости оценок границ выборочного среднего

Теорема 2. Пусть гиперслучайная величина X имеет границы математического ожидания m_{ix}, m_{sx} и конечную верхнюю границу дисперсии D_{sx}. Из генеральной совокупности гиперслучайной величины X в неконтролируемо меняющихся статистических условиях формируется L непересекающихся гиперслучайных выборок $(X_{11},...,X_{N1}),...,(X_{1L},...,X_{NL})$ объемом N каждая ($L \ge 2$). Пусть $m^*_{x_1},...,m^*_{x_L}$ – соответствующие этим выборкам гиперслучайные выборочные средние:

$$m^*_{x_1} = \frac{1}{N}\sum_{n=1}^{N} X_{n1},...,m^*_{x_L} = \frac{1}{N}\sum_{n=1}^{N} X_{nL},$$

а m^*_{ix}, m^*_{sx} – гиперслучайные оценки границ выборочных средних $m^*_{x_1},...,m^*_{x_L}$:

$$m^*_{ix} = \inf_{l=\overline{1,L}} m^*_{x_l}, \quad m^*_{sx} = \sup_{l=\overline{1,L}} m^*_{x_l}. \tag{2}$$

Тогда при неограниченном увеличении количества выборок и объема каждой выборки оценки границ выборочного среднего m^*_{ix}, m^*_{sx} сходятся по вероятности к соответствующим границам m_{ix}, m_{sx} математического ожидания гиперслучайной величины X:

$$\lim_{L\to\infty}\lim_{N\to\infty} m^*_{ix} = m_{ix}, \qquad \lim_{L\to\infty}\lim_{N\to\infty} m^*_{sx} = m_{sx}.$$

Доказательство теоремы состоит в следующем. Согласно теореме 1, для любой l-ой выборки при устремлении объема выборки к бесконечности границы гиперслучайного выборочного среднего $m^*_{x_l}$ стремятся по вероятности к границам математического ожидания m_{ix}, m_{sx} гиперслучайной величины X. Это означает, что область значений гиперслучайного выборочного среднего $m^*_{x_l}$ ограничена интервалом, стремящимся к интервалу $[m_{ix}, m_{sx}]$.

Принимая во внимание равенства (2), при $L \to \infty$ и $N \to \infty$ оценки границ выборочного среднего m^*_{ix}, m^*_{sx} сходятся по вероятности к тем же границам m_{ix}, m_{sx}.

Из теоремы следует, что состоятельные оценки границ математического ожидания гиперслучайной величины могут быть получены путем вычисления средних по множеству отсчетов для множества выборок и расчета по ним искомых границ.

Нетрудно убедиться, что таким же путем можно вычислять состоятельные оценки границ $m^*_{ix\nu} = \inf_{l=\overline{1,L}} m^*_{x_l\nu}$, $m^*_{sx\nu} = \sup_{l=\overline{1,L}} m^*_{x_l\nu}$ начальных моментов любого порядка ν, где $m^*_{x_l\nu}$ – оценка начального момента ν-го порядка, соответствующая l-ой выборке.

Следует заметить, что рассчитывать границы центральных моментов μ^*_{ixv}, μ^*_{sxv} по описанной выше схеме нельзя из-за отсутствия необходимой для этого информации об оценках условных математических ожиданий.

Теорема, аналогичная теореме Бернулли

Теорема 3. Пусть в неконтролируемо меняющихся статистических условиях проводится серия N независимых опытов, в каждом из которых может произойти некоторое событие A, рассматриваемое как гиперслучайное событие, представляемое множеством случайных событий A/g в условиях $g \in G$: $A = \{A/g \in G\}$.

Вероятность появления события A/g в фиксированных условиях $g \in G$ равняется $p_{a/g}$. Нижняя и верхняя границы вероятности события A соответственно равны P_{Ia}, P_{Sa}. Частота появления события A в рассматриваемой серии опытов описывается выражением $\Xi = \dfrac{N_a}{N}$, где N_a – число опытов, в которых произошло событие A. Эта частота – гиперслучайная величина, представляемая множеством случайных величин Ξ/g ($g \in G$): $\Xi = \{\Xi/g \in G\}$.

Тогда границы $F_{I\xi}(\xi)$, $F_{S\xi}(\xi)$ функции распределения $F_{\xi}(\xi)$ частоты Ξ при $N \to \infty$ сходятся по вероятности к функциям единичного скачка в точках P_{Ia}, P_{Sa}.

Доказательство основано на теореме 1. Гиперслучайное событие A можно рассматривать как гиперслучайную величину X, принимающую значение, равное единице, когда происходит событие A, и значение, равное нулю, когда событие не происходит. Условное математическое ожидание $m_{x/g}$ случайной величины X/g равно $p_{a/g}$, а условная дисперсия

$$D_{x/g} = (1 - p_{a/g})^2 p_{a/g} + (0 - p_{a/g})^2 (1 - p_{a/g}) = p_{a/g}(1 - p_{a/g})$$

есть величина ограниченная. Границы математического ожидания гиперслучайной величины X равны P_{Ia}, P_{Sa}. Выборочное среднее гиперслучайной величины X представляет собой частоту $\Xi = \dfrac{N_a}{N}$ появления события A. Тогда на основании теоремы 1 справедливо утверждение доказываемой теоремы.

Из теоремы следует, что при увеличении числа опытов $(N \to \infty)$ частота появления события A вне зависимости от количества условий G (если $G \neq 1$) находится в диапазоне $[P_{Ia}, P_{Sa}]$, не стремясь к какому-то конкретному значению.

Теорема о соотношении границ математического ожидания и математического ожидания границ

Теорема 4. Пусть m_{ix} и m_{sx} – соответственно нижняя и верхняя границы математического ожидания гиперслучайной величины $X = \{X/g \in G\}$ с функциями распределения $F_{ix}(x)$, $F_{sx}(x)$, а m_{Sx} и m_{Ix} – математические ожидания соответственно верхней $F_{Sx}(x)$ и нижней $F_{Ix}(x)$ границ функции распределения этой гиперслучайной величины.

Тогда $m_{ix} \geq m_{Sx}$ и $m_{sx} \leq m_{Ix}$.

Для доказательства теоремы рассмотрим функции распределения $F_{ix}(x)$ и $F_{Sx}(x)$. На одних участках эти функции могут совпадать, на других – нет. На интервалах, где они не совпадают, кривая $F_{ix}(x)$ располагается правее кривой $F_{Sx}(x)$. Поэтому $\int_0^1 x\mathrm{d}F_{ix}(x) \geq \int_0^1 x\mathrm{d}F_{Sx}(x)$, т.е. $m_{ix} \geq m_{Sx}$. Аналогично доказывается неравенство $m_{sx} \leq m_{Ix}$.

Выводы

1. На базе известной теоремы Чебышева для случайных величин доказаны для гиперслучайных величин две теоремы, определяющие условия сходимости выборочного среднего гиперслучайной величины к границам ее математического ожидания и алгоритм расчета состоятельных оценок границ начальных моментов гиперслучайных величин при быстрых изменениях статистических условий.

2. Доказана для гиперслучайных событий теорема, аналогичная теореме Бернулли для случайных событий, определяющая границы функции распределения частоты гиперслучайных событий при устремлении объема выборки к бесконечности.

3. Доказано, что нижняя граница математического ожидания гиперслучайной величины не меньше математического ожидания верхней границы распределения, а верхняя граница математического ожидания гиперслучайной величины не больше математического ожидания нижней границы распределения.

Благодарности

Работа опубликована при финансовой поддержке проекта **ITHEA XXI** Института информационных теорий и приложений FOI ITHEA Болгария www.ithea.org и Ассоциации создателей и пользователей интеллектуальных систем ADUIS Украина www.aduis.com.ua.

Литература

[Гнеденко, 1961] Гнеденко Б.В. Курс теории вероятностей. – М.: Изд-во физ.-мат. литературы, 1961. – 406 с.

[Горбань, 2006] Горбань И.И. Оценки характеристик гиперслучайных величин // Математичні машини і системи. – 2006. – № 1. – С. 40 –48.

[Горбань, 2007] Горбань И.И. Теория гиперслучайных явлений К.: Інститут проблем математичних машин и систем НАН Украины, ГП «УкрНИУЦ» Госпотребстандарта Украины, 2007. – 184 с (http://ifsc.ualr.edu/jdberleant/intprob/).

[Королюк, 1985] Королюк В.С. и др. Справочник по теории вероятностей и математическая статистика. – М.: Наука, 1985. – 637 с.

Сведения об авторе

***Игорь Горбань** – главный научный сотрудник ИПММС НАН Украины, доктор технических наук, профессор, Украина, 03187, Киев, пр. академика Глушкова. 42; e-mail: igor.gorban@yahoo.com.*