Возможные темы НИР

Родионов И.В.

Моя научная деятельность прежде всего связана со стохастической теорией экстремумов. Данная теория изучает поведение максимумов случайных величин. В основе данной теории лежит известная теорема Гнеденко, которая является неким аналогом центральной предельной теоремы, но для максимумов случайных величин. Теорема утверждает, что если при некоторых $a_n > 0$ b_n

$$P(\max(X_1, \dots, X_n) \le a_n x + b_n) = F^n(a_n x + b_n) \to G(x), \tag{1}$$

где $(X_1, \dots, X_n)$ – выборка с функцией распределения F, а G – функция распределения, принимающая более 2 значений, то G может быть одной из трёх типов (в отличие от ЦПТ, где постулируется сходимость к нормальному распределению), т.е. существуют $a > 0$ и b, что $G(ax + b) = G_\gamma(x)$ для некоторого $\gamma \in \mathbb{R}$, где

1. $G_\gamma(x) = \exp(-x^{-1/\gamma}) I(x > 0)$ при $\gamma > 0$ – класс распределений Фреше;
2. $G_0(x) = \exp(-e^{-x})$ – распределение Гумбеля;
3. $G_\gamma(x) = \exp(-(-x)^{1/\gamma}) I(x \le 0) + I(x > 0)$ для $\gamma < 0$ – класс распределений Вейбулла.

Если функция распределения F удовлетворяет (1) для некоторого γ, то будем говорить, что F принадлежит области притяжения закона G_γ, $F \in MDA(G_\gamma)$. Таким образом, различают области максимального притяжения Фреше, Вейбулла и Гумбеля.

Изучение распределения максимума независимых одинаково распределенных случайных величин основано по большей части на исследовании параметра γ, возникающего в теореме Гнеденко, который называется *индексом экстремального значения* (см. подробнее монографию [1]). Для его оценивания существует множество методов, можно считать, что исследование его поведения завершено. Области максимального притяжения Фреше и Вейбулла хорошо изучены по причине того, что довольно точно описываются индексом экстремального значения. С другой стороны, область притяжения Гумбеля, наиболее обширная из трех областей максимального притяжения, остается практически не изученной, поскольку индекс экстремального значения для всех распределений из этой области равняется 0. Кроме того, существует большой класс распределений, который не удовлетворяет условиям теоремы Гнеденко и для исследования которого классические методы статистики экстремумов не годятся.

В своей научной работе я предлагаю метод, который не зависит от принадлежности распределений какой-либо области максимального притяжения. По сути, стоит задача построения полноценной статистической теории для хвостов распределений. Некоторые сдвиги в этом направлении уже имеются:

1. Получен критерий различения двух классов хвостов распределений (аналогичный классическим критериям согласия), [2].

2. Получен метод оценки параметров хвостов распределений, доказана состоятельность и асимптотическая нормальность полученной оценки [3].

3. Получен метод оценки параметра формы хвостов распределений, не зависящий от параметра сдвига, [3].

4. Получены критерии различения распределений из области максмального притяжения Гумбеля, [4, 5].

5. Получены оценки параметров сдвига и масштаба для хвостов распределений из области максимального притяжения Гумбеля, [6].

1 Задачи

1 Обобщение предельной теоремы Гнеденко на случай зависимых случайных величин.

Пусть $\xi_k,\ k = 1, 2, \ldots$ – независимые одинаково распределённые случайные величины с функцией распределения F, η – не зависящая от них случайная величина, $f(x, y)$ – некая функция. Вывести предельную теорему для $M_n := \max(f(\eta, \xi_1), f(\eta, \xi_2), \ldots, f(\eta, \xi_n))$ при $n \to \infty$. То есть, найти последовательности $a_n > 0$, b_n такие, что предел

$$\lim_{n\to\infty} P(a_n^{-1}(M_n - b_n) < x) = G(x)$$

существует и является невырожденной функцией распределения.

а) Решить задачу, если почти наверное $f(\eta, \xi_1)$ принадлежит какой-либо из областей максимального притяжения. Указать в таком случае условия на f и η, при которых предельное распределение будет невырожденным.

б) Решить задачу, если предполагать, что у предельного распределения не может быть атомов.

в) Возможно ли решить задачу в произвольном случае?

2 Аналог метода максимального правдоподобия для k верхних членов вариационного ряда.

Пусть $X_1, \ldots, X_n$ – выборка из распределения с абсолютно непрерывной функцией распределения F, а $X_{(1)} \le \ldots \le X_{(n)}$ – вариационный ряд этой выборки. Пусть известно только k верхних (наибольших) членов вариационного ряда выборки. Предположим также, что F принадлежит некому параметрическому семейству функций распределения, $F \in \{F_a(x), a \in A\}$, где A – некое множество параметров. Определим $f(x_n, \ldots, x_{n-k}; a)$ – многомерную плотность, соответствующую функции распределения $F(x; a) = P(Y_{(n-k)} \le x_{n-k}, \ldots, Y_{(n)} \le x_n)$ (где выборка $\{Y_i\}_{i=1}^n$ сделана из распределения с функцией распределения $F_a(x)$), и $f_{k,n}(X; a) = f(X_{(n)}, \ldots, X_{(n-k)}; a)$.

а) Верно ли, что оценка методом максимального правдоподобия $\widehat{a} = \arg\max_{a\in A} f_{k,n}(X; a)$ будет состоятельной оценкой параметра a при $k \to \infty,\ n \to \infty,\ k/n \to 0$?

б) Если нет, то для какой функции $f(x_n, \ldots, x_{n-k}; a)$ это свойство будет выполнено? Будет ли для этой функции выполнено экстремальное свойство правдоподобия, т.е.

$$P_a\left(f(X_{(n)}, \ldots, X_{(n-k)}; a) > f(X_{(n)}, \ldots, X_{(n-k)}; b)\right) \to 1 \text{ при } n \to \infty, a \ne b.$$

3 Критерий различения классов хвостов распределений.

Пусть $X_1, \dots, X_n$ – выборка из распределения с непрерывной функцией распределения F, а $X_{(1)} \le \dots \le X_{(n)}$ – вариационный ряд этой выборки. Рассмотрим простую гипотезу $H_0 : F = F_0$ и определим статистику

$$R_{k,n} = \ln(1 - F_0(X_{(n-k)})) - \frac{1}{k} \sum_{i=n-k+1}^{n} \ln(1 - F_0(X_{(i)})).$$

Известно, что в условиях верности H_0

$$\sqrt{k}(R_{k,n} - 1) \xrightarrow{d} N(0,1)$$

при $k < n \to \infty$. Какие условия нужно наложить на функцию распределения F, чтобы результат оставался верным без предположения о непрерывности F? Будет ли он верен, если F – дискретна?

4 Непараметрический (ранговый) критерий проверки гипотезы сдвига о хвосте распределении.

Пусть имеются две независимые выборки $X = \{X_1, \dots, X_n\}$ и $Y = \{Y_1, \dots, Y_m\}$ с функциями распределения F и G соответственно. Предположим, что нам известны только k и l максимальных членов выборок X и Y соответственно. Рассмотрим две простые гипотезы $H_0 : F = G$ и $H_1 : F(x) = G(x - \theta)$.

а) Построить ранговый критерий проверки гипотезы H_0 против альтернативы H_1 и доказать его состоятельность (в качестве прототипа рассмотреть известный критерий Манна-Уитни-Уилкоксона).

б) Найти состоятельную оценку параметра θ по имеющимся данным, используя непараметрический подход.

5 Критерий различения областей максимального притяжения.

Пусть $X_1, \dots, X_n$ – выборка из распределения с непрерывной функцией распределения F, а $X_{(1)} \le \dots \le X_{(n)}$ – вариационный ряд этой выборки. Рассмотрим простую гипотезу $H_0 : F = F_0$ и определим статистику

$$R_{k,n} = \ln(1 - F_0(X_{(n-k)})) - \frac{1}{k} \sum_{i=n-k+1}^{n} \ln(1 - F_0(X_{(i)})).$$

Пусть $\{G_\gamma, \gamma \in \mathbb{R}\}$ – семейство экстремальных распределений из теоремы Гнеденко. Возможно ли, взяв в качестве F_0 распределение G_{γ_0}, построить состоятельный критерий для проверки гипотезы $H_0 : \gamma > \gamma_0$ против альтернативы $H_1 : \gamma < \gamma_0$ с помощью статистики $R_{k,n}$, где γ - индекс экстремального значения функции распределения F.

6 Критерии согласия для хвостов распределений.

Пусть $X_1, \dots, X_n$ – выборка из распределения с непрерывной функцией распределения F, а $X_{(1)} \le \dots \le X_{(n)}$ – вариационный ряд этой выборки. Рассмотрим простую гипотезу $H_0 : F(x) = F_0(x)$ при $x \to x_F$ и альтернативу $H_1 : F \ne F_0$ при $x \to x_F$, где $x_F = \sup\{x : F_0(x) < 1\}$. Предложить аналоги классических критериев согласия (Андерсона-Дарлинга, Колмогорова, Смирнова-фон Мизеса) для проверки данных гипотез.

Список литературы

[1] Ferreira, A., Haan, L. de. Extreme value theory. An introduction. Springer, Springer Series in Operations Research and Financial Engineering, New York, 2006.

[2] Rodionov, I. V., On discrimination between classes of distribution tails, Problems of Information Transmission, (2018). V. 54, N. 2, 124-138.

[3] Rodionov, I. V., On estimation of the parameter of the distribution tail. Extremes, in print.

[4] Родионов, И. В., Различение близких гипотез о хвостовых распределениях по максимальным членам вариационного ряда. Теория вероятностей и ее применения (2018), Т. 63, В. 3, С. 447-467.

[5] Родионов, И. В., Критерий различения для хвостов распределений вейбулловского типа. Теория вероятностей и ее применения (2018), Т. 63, В. 2, С. 402-413.

[6] Ахтямов, П. И., Родионов, И. В., Об оценке параметров сдвига и масштаба хвостов распределений. Фундаментальная и прикладная математика (2018), Т. 22, В. 3.