Санкт-Петербургский государственный университет

Прикладная математика и информатика

Статистическое моделирование

Старков Артём Константинович

Доверительное оценивание параметров предельных распределений статистик экстремальных значений

Отчет о научно-исследовательской работе

Научный руководитель:

д. ф.-м. н., профессор М. С. Ермаков

Санкт-Петербург

2017

Оглавление

Введение

В вопросах оценивания вероятностей больших уклонений весьма важную роль играет вопрос о распределении максимумов. Например, пусть дана последовательность независимых случайных величин $(X_i)_{i=1}^n$. Тогда M_n — их выборочный максимум:

$$M_n = \max(X_1, ..., X_n).$$

Одним из результатов изучения распределения M_n является теория предельных распределений. Важнейший результат этой теории заключен в теореме, представленной ниже [1, стр. 132].

Теорема 1 (Фишера-Типпета-Гнеденко)**.** *Пусть* $(X_i)_{i=1}^n$ *— последовательность независимых случайных величин,* M_n *— их выборочный максимум* $M_n = \max(X_1, ..., X_n)$*. Если существуют такие* $c_n > 0, d_n \in \mathbb{R}$ *и некоторая плотность распределения* H*, что выполняется*

$$\frac{(M_n - d_n)}{c_n} \xrightarrow{p} H,$$

тогда H *принадлежит одному из семейств предельных распределений:*

Фреше: $\Phi_\alpha(x) = e^{-x^{-\alpha}}, \quad x > 0, \alpha > 0;$

Вейбулл: $\Psi_\alpha(x) = e^{-(-x)^\alpha}, \quad x \leq 0, \alpha > 0;$

Гумбель: $\Lambda(x) = e^{-e^{-x}}, \quad x \in \mathbb{R}.$

Очевидно, что вид предельного распределения напрямую зависит от исходного распределения X_i. Будем говорить, что если некоторая последовательность $(X_i)_{i=1}^n$ с функцией распределения $F(x)$ имеет предельное распределение H, то она принадлежит его области максимального притяжения (maximum domain attraction), и записывать:

$$F \in \mathsf{MDA}(H).$$

В книге Гумбеля [2, стр. 194] детально изложен вывод предельных распределений из теоремы 1, данный Фишером и Типпетом. Там же продемонстрирован общий вид распределений, предложенный Дженкинсоном:

$$\Phi(x) = \exp\left[-\left(1 - \frac{x}{a}\right)^{\frac{1}{k}}\right], ak > 0. \tag{1}$$

В данной работе рассматривается распределение Фреше. Оно соответствует семействам распределений Парето, Коши, Булла и др. (для случая $k < 0$ из (1)).

Распределение Фреше имеет параметры shape α, scale s и location m; $\alpha > 0, s > 0, x > m$:

$$F(x) = e^{-(\frac{x-m}{s})^{-\alpha}},$$
$$p(x) = \frac{\alpha}{s}\left(\frac{x-m}{s}\right)^{-1-\alpha} e^{-(\frac{x-m}{s})^{-\alpha}}.$$

Для $\mathsf{MDA}(\Phi_\alpha)$ существует следующая теорема [1, стр. 142]:

Теорема 2. *Функция распределения $F(x)$ принадлежит области максимального притяжения распределения Фреше $F \in \mathsf{MDA}(\Phi_\alpha)$ для $\alpha > 0$, если и только если $\bar{F}(x) = x^{-\alpha}L(x)$, для некоторой медленно меняющейся функции L, где $\bar{F}(x) = 1 - F(x)$ — распределение хвоста $F(x)$.*

На основании этого факта Хиллом [3] была получена оценка для α, совпадающая с оценкой максимального правдоподобия:

$$\hat{\alpha}_H = \left(\frac{1}{k}\sum_{i=1}^{k}\ln X_{i:n} - \ln X_{k:n}\right)^{-1}, \tag{2}$$

где $k = k(n) \to \infty$. Это оценка т.н. «нижнего хвоста» (*lower tail estimate*), в той же статье делается оценка и «верхнего хвоста» (*upper tail estimate*), т. е. для распределения хвоста на интервале $[\beta, 1)$. Она идентична (2) и делается путем замены $X = Y^{-1}$; соответственно с учетом смены направления упорядочивания вариационного ряда получаем

$$Y_{i:n} = X_{n-i+1:n},$$
$$\left(\frac{1}{n-m+1}\sum_{i=m}^{n}\ln X_{i:n} - \ln X_{m:n}\right)^{-1}, \tag{3}$$

где $m = n - k + 1 : m = [\beta n]$.

Для многих прикладных задач параметр α имеет большое значение. Например, при $\alpha < 2$ дисперсия бесконечна: $\mathsf{E}X^2 = \infty$. Этот случай часто наблюдается на практике, и его нужно отслеживать при моделировании. В связи с этим возникает задача определения качества полученной оценки. Так как оценка строится в условиях $k = k(n) \to \infty$, то любая попытка определения качества оценки затруднительна ввиду необходимости получения большого количества реализаций $n \to \infty$ при том, что действительно полезных из них будет относительно немного (в зависимости от параметра β). В решении этой проблемы в задачах статистического моделирования может помочь асимптотическая эффективность метода существенной выборки.

1. Метод существенной выборки

Пусть поставлена задача вычисления следующей вероятности:

$$\omega = P(T(\hat{P}_n) - T(P_0) \in b_n\Omega), \tag{4}$$

где P_0 – теоретическое распределение; $\hat{P}_n$ – эмпирическая функция распределения; $T(P)$ – некоторый функционал. Согласно методу, необходимо выбрать меру Q_n такую, что $Q_n \ll P_0$. Затем моделируются k независимых выборок с распределением Q_n

$$Y_1^{(i)}, Y_2^{(i)}, ..., Y_n^{(i)}, 1 \le i \le k.$$

Оценка вероятности:

$$\hat{\omega}_n = \frac{1}{k}\sum_{i=1}^{k} \chi\big(T(\hat{Q}_n^{(i)}) - T(P_0) > b_n\big) \prod_{j=1}^{n} q_n^{-1}(Y_j^{(i)}), \tag{5}$$

где $\hat{Q}_n^{(i)}$ – эмпирическое распределение выборки $Y_j^{(i)}$, $q_n = \frac{dQ_n}{dP_0}$.

Введем величину

$$U_n = \mathsf{E}_{Q_n}\left[\chi\big(T(\hat{Q}_n^{(1)}) - T(P) > b_n\big) \prod_{j=1}^{n} q_n^{-2}(Y_j^{(1)})\right],$$

тогда дисперсия оценки (5) имеет вид

$$V(Q_n) = \mathsf{Var}[\hat{\omega}_n] = U_n - \omega_n^2,$$

где ω_n – математическое ожидание (5). Потому естественно рассмотреть асимптотическую эффективность процедуры существенной выборки. Будем определять ее в терминах теории вероятности больших уклонений. В силу неотрицательности дисперсии ясно, что

$$U_n \le \omega_n^2.$$

Определение 1. *Процедура называется асимптотически эффективной (в смысле логарифмической асимптотики), если*

$$\overline{\lim_{n\to\infty}} \frac{\log U_n}{2\log \omega_n} = 1.$$

Определение 2. *Процедура называется эффективной, если*

$$\overline{\lim_{n\to\infty}} \frac{U_n}{2\omega_n} = 1.$$

Так как задачей является оценка малых вероятностей, их математические ожидания и дисперсии также будут малыми величинами. Тогда даже малые ошибки в вычислениях могут привести к значительным ошибкам в значении эффективности или даже невозможности ее вычислить. На асимптотическую эффективность малые ошибки влияют незначительно; поэтому для определения качества процедуры будет использоваться асимптотическая эффективность.

Основной трудностью при использовании метода существенной выборки является выбор меры Q_n при условии достижения процедурой асимптотической эффективности.

Предпосылкой для решения задачи о выборе меры Q_n стала работа математика И. В. Санова. Им была исследована задача вычисления вероятности

$$P(\hat{P}_n \in \Omega), \tag{6}$$

где $\hat{P}_n$ – эмпирическая функция распределения, построенная по выборке $X_i \sim P_0, 1 \leq i \leq n$; Ω принадлежит пространству мер. Он исследовал асимптотическое поведение вероятности (6) и получил следующие результаты:

$$\begin{aligned} \overline{\lim_{n\to\infty}} \tfrac{1}{n} \ln P(\hat{P}_n \in \Omega) &\leq -\inf_{H \in \mathfrak{int}(\Omega)} K(H, P_0), \\ \underline{\lim_{n\to\infty}} \tfrac{1}{n} \ln P(\hat{P}_n \in \Omega) &\geq -\inf_{H \in \mathfrak{cl}(\Omega)} K(H, P_0), \end{aligned} \tag{7}$$

где $\mathfrak{int}(\Omega)$ и $\mathfrak{cl}(\Omega)$ – соответственно замыкание и внутренность Ω в специальной топологии со слабой сходимостью,

$$K(H, P_0) = \begin{cases} \int \ln \frac{dH}{dP_0} dH, & \text{если } H \ll P_0; \\ \infty, & \text{иначе.} \end{cases} \tag{8}$$

На основе этих результатов в дальнейшем была развита теория вероятности больших уклонений, окончательно формализованная С. Р. Варадханом. В рамках этой теории неравенства типа (7) стали называться принципом больших уклонений, а функционалы типа (8) – функционалами действия. Они являются в некотором смысле мерой близости распределений H и P_0, схожей с расстоянием Кульбака-Лейблера.

Рассмотрим задачу (4) в зоне больших уклонений:

$$P(T(\hat{P}_n) - T(P_0) \in \Omega), \Omega \in \mathbb{R}.$$

На практике для такой задачи нахождение H часто является неразрешимой задачей. В зоне умеренных уклонений и действия ЦПТ для задачи в виде

$$P(T(\hat{P}_n) - T(P_0) \in b_n\Omega), \Omega \in \mathbb{R}, \tag{9}$$

где $nb_n^2 \to \infty, b_n \to \infty$, действует нормальная асимптотика, но для логарифма распределения. Здесь задача нахождения H разрешима для очень широкого класса функционалов T.

М. Джонсоном была построена асимптотически эффективная процедура метода существенной выборки в зоне нормальной аппроксимации. В зону умеренных уклонений его результаты были перенесены М. С. Ермаковым в статье [4]. В статье рассматривается задача (9) в случае, когда статистика $T(\hat{P}_n)$ имеет функцию влияния g и верны предположения, рассматриваемые ниже.

Введем следующие множества:

1. множество функций Φ такое, что для любой $f \in \Phi$ $\mathsf{E}f(x) = 0$ и
$$\lim_{n\to\infty} \frac{1}{nb_n^2} \log\big(nP_0(|f(X)| > nb_n)\big) = -\infty;$$
2. множество Λ_Φ всех мер $Q \in \Lambda$ таких, что для любой $f \in \Phi$ $\int_\Omega |f| dQ < \infty$;
3. множество $\Lambda_{0\Phi}$ всех зарядов $G = P - R; P, R \in \Lambda$.

Тогда предположим, что $g \in \Phi$ и существует полунорма $N \in \Lambda_{0\Phi}$, такая, что для любого $Q \in \Lambda_\Phi$

$$|T(Q) - T(P_0) - \int_\Omega g dQ| \le \omega\big(N(Q - P_0), \int_\Omega g dQ, T(Q) - T(P_0)\big)$$

с функцией $\omega\colon \mathbb{R}^3 \to \mathbb{R}_+$, такой, что

$$\lim_{t_1,t_2,t_3\to 0} \frac{\omega(t_1,t_2,t_3)}{t_1+t_2+t_3} = 0.$$

М. С. Ермаковым была доказана следующая теорема.

Теорема 3. *В предположениях, изложенных выше, рассмотрим процедуру существенной выборки, основанную на вероятностной мере Q_n с плотностью*

$$q_{1n}(x) = \lambda_n + b_n h(x) \chi\left(h(x) > -\frac{\delta}{b_n}\right)$$

или

$$q_{2n}(x) = c_n e^{b_n h(x)} \chi\left(h(x) < \frac{\delta}{b_n}\right),$$

где λ_n, c_n – константы нормализации, $\delta \in [0,1]$ и $\mathsf{E}h(x) = 0, \mathsf{E}|h(x)| < \infty, \mathsf{E}h^2(x) < \infty$. Тогда процедура существенной выборки асимптотически эффективна, если $h = \frac{g}{\sigma_g^2}$

2. Постановка задачи

Пусть $(X_i)_{i=1}^n$ — независимые случайные величины с функцией распределения $F \in \mathsf{MDA}(\Phi_\alpha)$. Упорядочим выборку по возрастанию, получим вариационный ряд:

$$X_{1:n} \leq X_{2:n} \leq \ldots \leq X_{n:n}.$$

Хвост распределения $\bar{F}(x) = 1 - F(x)$ принадлежит семейству предельного распределения Фреше Φ_α, в котором неопределенным остается параметр формы (shape) α. Имеем для него оценку Хилла:

$$\hat{\alpha}_H = \left(\frac{1}{n-m+1} \sum_{i=m}^{n} \ln X_{i:n} - \ln X_{m:n} \right)^{-1}, \tag{10}$$

где $m = [\beta n]$.

Рассмотрим задачу оценки вероятности уклонения оцененного значения от истинного, представленной в формуле (4), где в качестве функционала $T(F)$ будет выступать оценка $\hat{\alpha}_H$. Так как сама оценка строится в условиях $m = m(n) \to \infty$, то $\beta \to 1$, и обычные оценки становятся неэффективными. Для решения данной проблемы будет использована асимптотическая эффективность метода существенной выборки.

По теореме 3 будем использовать вероятностную меру Q_n вида

$$g(x) = p(x)(1 + b_n h(x)), \tag{11}$$

где $h(x)$ — функция влияния оцениваемого функционала, b_n – константа для определения ширины доверительного интервала,

$$b_n = \pm \frac{N^{-1}_{(0,1)}(\gamma)}{\sqrt{n}\sigma}, \tag{12}$$

$$\sigma^2(F) = \int_{-} \infty^\infty \int_{-} \infty^\infty J(F(x))J(F(y))\big(F(x \wedge y) - F(x)F(y)\big) \tag{13}$$

— асимптотическая дисперсия функционала $T(F)$.

Определим вид оцениваемого функционала и его функцию влияния.

Для удобства будем оценивать значение $\hat{\alpha}_H^{-1}$; таким образом избавляемся от показателя степени -1 в правой части. Обозначим $Y_j = \ln X_j$, тогда $F_y(\ln t) = F(t)$, F_y – функция распределения Y. Получим следующий функционал:

$$T(F_y) = \hat{\alpha}_H^{-1} = \frac{1}{n-m+1} \sum_{i=m}^{n} Y_{i:n} - Y_{m:n} \tag{14}$$

Общий вид таких функционалов для L-оценок [5, стр. 263]:

$$T(F) = \int_0^1 F^{-1}(t)J(t)dt + \sum_{j=1}^{l} a_j F^{-1}(p_j), \qquad (15)$$

где $J(t)$ — некоторая интегрируемая весовая функция, p_j — уровни квантилей $F^{-1}(p_j)$, a_j – соответствующие им веса. Там же предложен общий вид функции влияния:

$$IC(x;T,F) = -\int_{-\infty}^{\infty} [\gamma(x \le y) - F(y)]J(F(y))dy + \sum_{j=1}^{l} a_j \frac{p_j - \gamma(x \le F^{-1}(p_j))}{f(F^{-1}(p_j))}.$$

Перепишем (14) с учетом формы (15):

$$T(F_y) = \frac{1}{n-m+1}\sum_{i=m}^{n} Y_{i:n} - Y_{m:n} = \frac{1}{n-m+1}\sum_{i=m}^{n} F_y^{-1}(\frac{i}{n}) - F_Y^{-1}(\beta) =$$

$$= \frac{1}{1-\beta}\int_{\beta}^{1} F_y^{-1}(t)dt - F_y^{-1}(\beta), \qquad (16)$$

$$J(t) = \frac{\gamma(t > \beta)}{1-\beta}.$$

Соответственно функция влияния принимает вид:

$$h(x) = IC(x;T,F_y) = \frac{1}{\beta - 1}\int_{F_y^{-1}(\beta)}^{\infty} (\gamma(x \le t) - F_y(t))dt - \frac{\beta - \gamma(x \le F_y^{-1}(\beta))}{f_y(F_y^{-1}(\beta))};$$

Обозначив $F_y^{-1}(\beta) = x_1$, получаем:

$$h(x) = \begin{cases} \frac{1}{\beta-1}\int_{x_1}^{\infty}(1 - F_y(t))dt - \frac{\beta-1}{f_y(x_1)}, & \text{если } x \le x_1; \\ \frac{1}{1-\beta}\int_{x_1}^{x} F_y(t)dt + \frac{1}{\beta-1}\int_{x}^{\infty}(1 - F_y(t))dt - \frac{\beta}{f_y(x_1)}, & \text{иначе.} \end{cases} \qquad (17)$$

Моделирование производится для хвоста распределения; по теореме 2 искомое распределение $\bar{F}(x) = Cx^{-\alpha}$. Выразим f_y:

$$Cx^{-\alpha} = Ce^{-\alpha \ln x}|_{t=\ln x} = Ce^{-\alpha t} = \alpha e^{-\alpha t} = f_y(t),$$

или экспоненциальное распределение с параметром $\lambda = \alpha$.

Целью данной работы является построение доверительного интервала вероятности вида (4) для функционала оценки Хилла (16). Для ее достижения, исходя из описанного выше, необходимо:

1. промоделировать вспомогательное распределение Q_n (11);

2. на основе модельных данных построить доверительный интервал для вероятности (5).

3. Моделирование

Из (17) имеем функцию влияния:

$$h(x) = \begin{cases} \frac{1}{\alpha(\beta-1)}\left(e^{-\alpha x_1} + \beta - 1\right), & \text{если } x \le x_1; \\ \frac{1}{\alpha(1-\beta)}\left(\alpha(x - x_1) - e^{-\alpha x_1} - \beta\right), & \text{иначе.} \end{cases}$$

Заметим, что

$$e^{-\alpha x_1} = e^{-\alpha F^{-1}(\beta)} = e^{-\alpha\left(-\frac{1}{\alpha}\ln(1-\beta)\right)} = e^{\ln(1-\beta)} = 1 - \beta.$$

С учетом этого получаем

$$h(x) = \frac{x - x_1 - \frac{1}{\alpha}}{1-\beta}\gamma(x > x_1).$$

Введем дополнительное обозначение, нужное при моделировании:

$$x_2 = h^{-1}(0) = \frac{1}{\alpha} + x_1.$$

Для удобства записи используем это обозначение в функции влияния:

$$h(x) = \frac{x - x_2}{1-\beta}\gamma(x > x_1).$$

По (13) для $F(x) = 1 - e^{-\alpha x}$ имеем:

$$\sigma^2(F) = \frac{3-\beta}{\alpha^2(1-\beta)}.$$

Интеграл функции $g(x)$ для $x > x_1$:

$$\int\limits_a^b g(x)dx = \int\limits_a^b \alpha e^{-\alpha x}(1 + b_n\frac{x - x_2}{1-\beta})dx =$$

$$= (e^{-\alpha a} - e^{-\alpha b})(1 - \frac{b_n x_2}{1-\beta}) - \frac{b_n}{\alpha(1-\beta)}\left(e^{-\alpha b}(\alpha b + 1) - e^{-\alpha a}(\alpha a + 1)\right). \qquad (18)$$

Разобьем область моделирования на 3 части:

$$0 \le A < x_1 \le B < x_2 \le C < \infty$$

. В каждой области функция имеет свои особенности. Внешний вид плотности распределения показан на рисунке 1. Зелеными линиями отмечены точки x_1 и x_2.

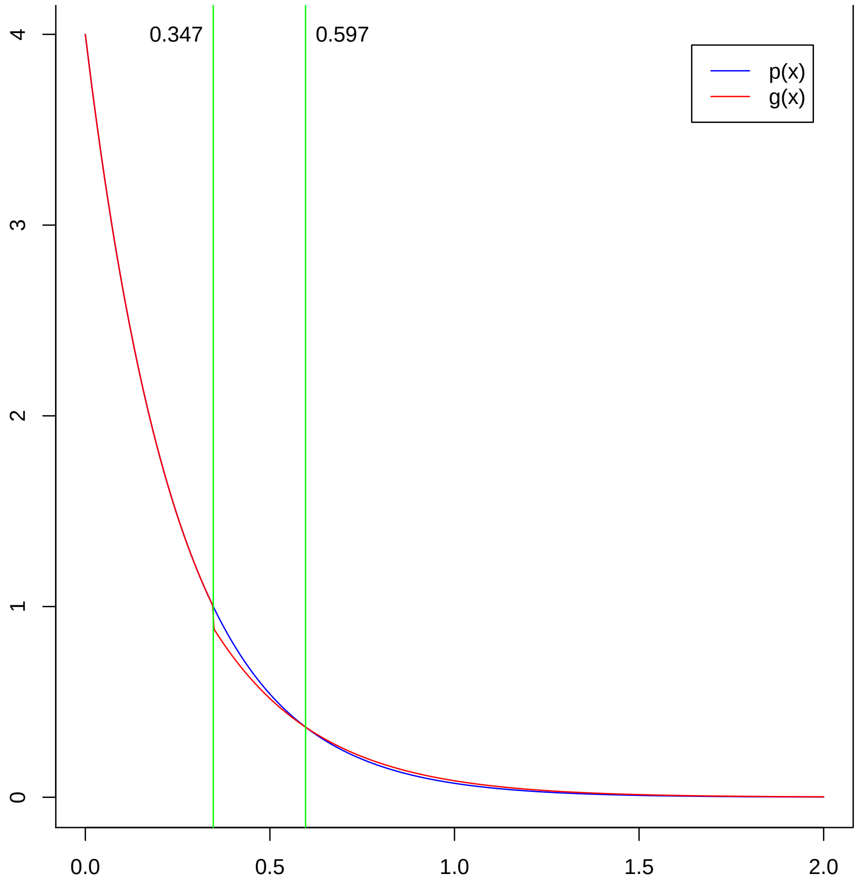

Рис. 1. Распределение $g(x), p(x)$

3.1. Область $A = [0, x_1)$

Область A не отличается от исходного показательного распределения. Ее моделирование не представляет никакой сложности. Площадь данного участка выражается через функцию экспоненциального распределения вероятности и равна $S_A = \beta$.

3.2. Область $B = [x_1, x_2)$

В области B применим метод мажорант. Моделирование проходит по следующему алгоритму [6]:

Алгоритм 1 (Метод мажорант)**.** *Пусть*

$$D_S = \frac{\int_S p(x)dx}{\int_S g(x)dx},$$

где $g(x)$ — моделируемая плотность, $p(x)$ — некоторая другая плотность, такая, что $\forall x \in S \; p(x) \geq g(x)$. Тогда:

1. *получим реализацию $\xi \sim \frac{p(X)}{|D|}$;*

2. *получим реализацию* η_{s+1} $U(0, p(\xi))$*;*

3. *если* $\eta_{s+1} > g(\xi)$*, переходим к пункту 1; иначе* ξ *— искомая случайная величина,* $\xi \sim g(x)$*.*

Эффективность метода равна $\frac{1}{|D_B|}$*.*

Отношение интегралов функций в данной области успешно выражается при помощи формулы (18) и следующих замечаний:

$$e^{-\alpha x_1} = 1 - \beta, e^{-\alpha x_2} = \frac{1-\beta}{e}. \tag{19}$$

С учетом этого площадь $g(x), x \in B$ и отношение площадей $\frac{p(x)}{g(x)}, x \in B$ принимает вид:

$$S_B = \int\limits_B g(x)dx = \frac{e-1}{e}(1 - \beta - b_n x_2 + \frac{b_n}{\alpha}) + b_n(x_1 - \frac{x_2}{e});$$

$$D_B = \frac{\int_B p(x)dx}{\int_B g(x)dx} = \frac{1-\beta}{1 - \beta + b_n(\frac{1}{\alpha} + \frac{e}{e-1}(x_1 - x_2))}.$$

Результат моделирования распределения в виде гистограммы с наложенными функциями $g(x), p(x)$ представлен на рисунке 2.

3.3. Область $C = [x_2, \infty)$

В области C используем метод композиции [6].

Алгоритм 2 (Метод композиции)**.** *Пусть искомая плотность* $g(x)$ *представлена суммой плотностей* $p_i(x)$ *с некоторыми неотрицательными весами* a_i*:*

$$g(x) = \sum_{i=1}^{M} a_i p_i(x). \tag{20}$$

Тогда метод выражается следующим алгоритмом:

1. *получаем реализацию* $\xi \sim U(0, \sum_i a_i)$*;*

2. *выбираем индекс* $s = \arg\max(a_i : a_i < \xi)$;

3. *искомая случайная величина —* $\eta \sim p_s(x)$*.*

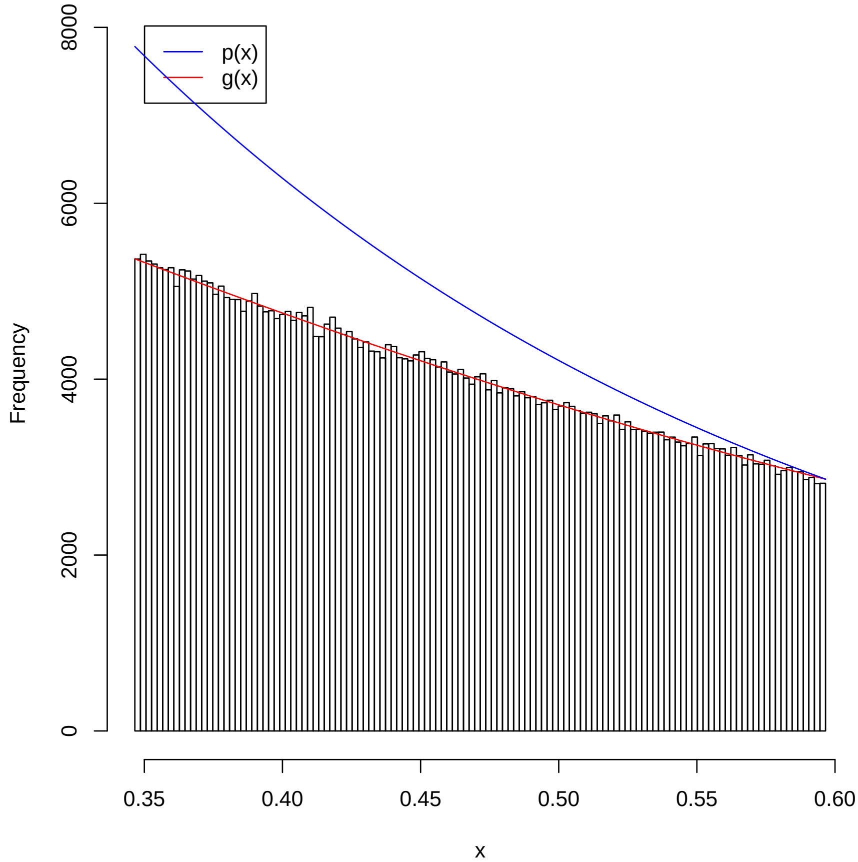

Рис. 2. Распределение $g(x), p(x)$ при $x \in B$

Представим функцию в виде:

$$g(x) = p(x)(1 + b_n h(x)) = \alpha e^{-\alpha x} + \frac{b_n(x - x_2)}{1 - \beta}\alpha e^{-\alpha x} = p(x) + p_1(x).$$

Заменим во второй части $y = x - x_2$:

$$p_1(y) = \frac{b_n y}{1 - \beta}\alpha e^{-\alpha(y + x_2)} = \frac{b_n}{\alpha(1 - \beta)}\alpha^2 y e^{-\alpha y} e^{-\alpha x_2} = \frac{b_n e}{\alpha}\alpha^2 y e^{-\alpha y}.$$

Теперь $p_1(y)$ выражается гамма-распределением с параметрами $shape = 2, scale = \frac{1}{\alpha}$. Получаем форму (20) с весами $a_1 = 1, a_2 = \frac{b_n e}{\alpha}$ и распределениями

$$p_1(x) = p(x) = exp(x, \alpha),$$
$$p_2(x) = \gamma(x, 2, \frac{1}{\alpha}) + x_2.$$

Используя формулу (18) и замечания (19), получаем формулу для площади $g(x) : x \in C$:

$$\int\limits_C g(x)dx = \frac{\alpha - \alpha\beta + b_n}{\alpha e}.$$

Результат моделирования распределения в виде гистограммы с наложенными функциями $g(x), p(x)$ представлен на рисунке 3.

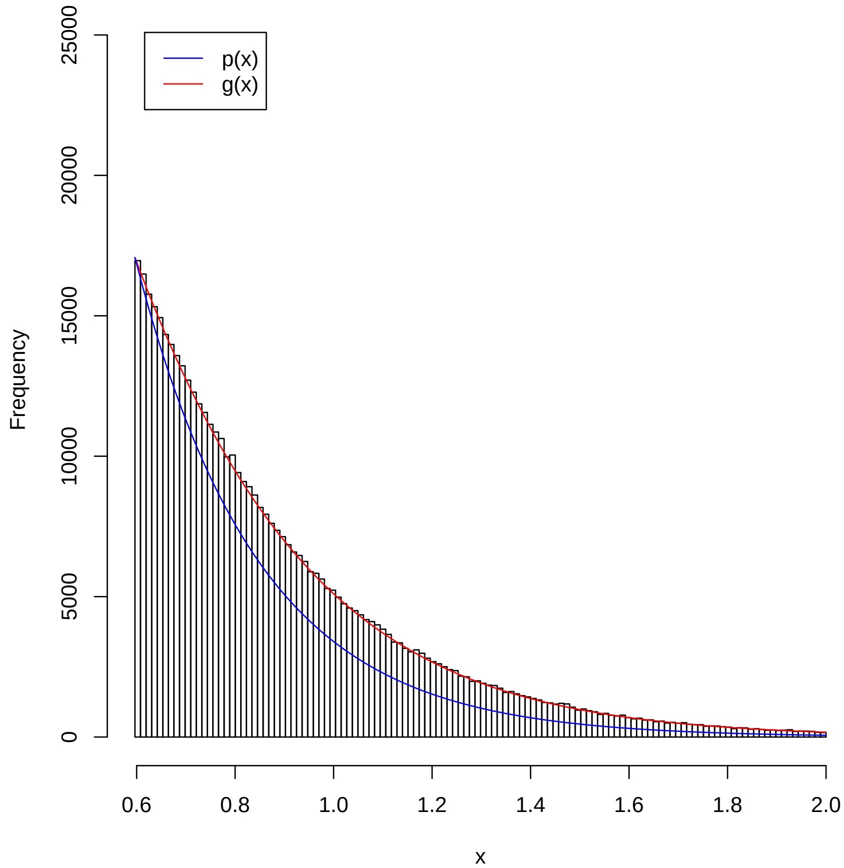

Рис. 3. Распределение $g(x), p(x)$ при $x \in C$

3.4. Моделирование общего распределения $g(x)$

Моделирование проводится при помощи алгоритма 2. В качестве весов берутся площади плотности $g(x)$ в соответствующих областях. Результат моделирования распределения в виде гистограммы с наложенными функциями $g(x), p(x)$ представлен на рисунке 4.

3.5. Вычисление доверительного интервала

Строим k выборок по n реализаций, $Y_n^{(k)} \sim g(x)$. Выберем n достаточно небольшим. По (5) имеем оценку вероятности:

$$\hat{\omega}_n = \frac{1}{k}\sum_{i=1}^{k} \chi\big(T(\hat{Q}_n^{(i)}) - T(P_0) > b_n\big) \prod_{j=1}^{n} \frac{1}{1 + b_n h(Y_j^{(i)})};$$

$$U_n = \frac{1}{k}\sum_{i=1}^{k} \chi\big(T(\hat{Q}_n^{(1)}) - T(P) > b_n\big) \prod_{j=1}^{n} \frac{1}{(1 + b_n h(Y_j^{(i)}))^2}.$$

Дисперсия оценки вероятности $V = U_n - \hat{\omega}_n^2$.

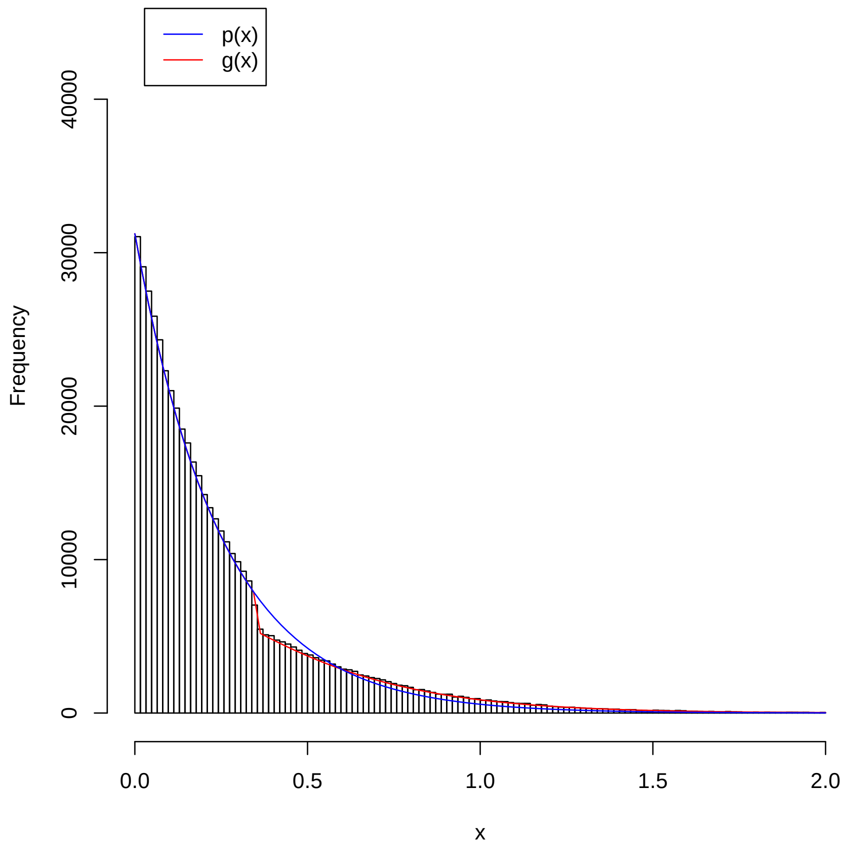

Рис. 4. Распределение $g(x), p(x)$

Для демонстрации работы алгоритма были построены доверительные интервалы ω для следующих параметров: $\alpha = 4, \beta = 0.9, \gamma \in [0.8, 0.99] n = 500, k = 1000$ и $k = 2000$.

Выбор α обусловлен лишь ограничением на конечность дисперсии: $\alpha > 2$. Выбор β сделан исходя из стандартной практики в данной области, ее значение выбирают в зависимости от объема выборки в диапазоне от 0.8 до 0.9; γ это уровень доверия вероятности ω. Следует учесть, что при моделировании необходимо перестраивать плотность $g(x)$ каждый раз при изменении γ, так как от нее зависит b_n (12). Итоговые доверительные интервалы показаны на рисунках 5 и 6.

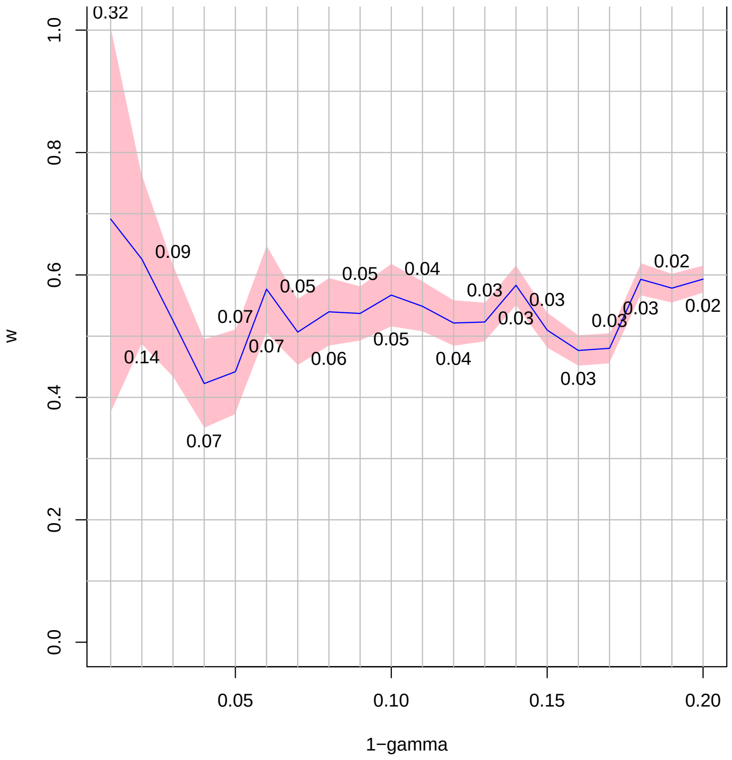

Рис. 5. Оценка ω и доверительный интервал при $k = 1000$.

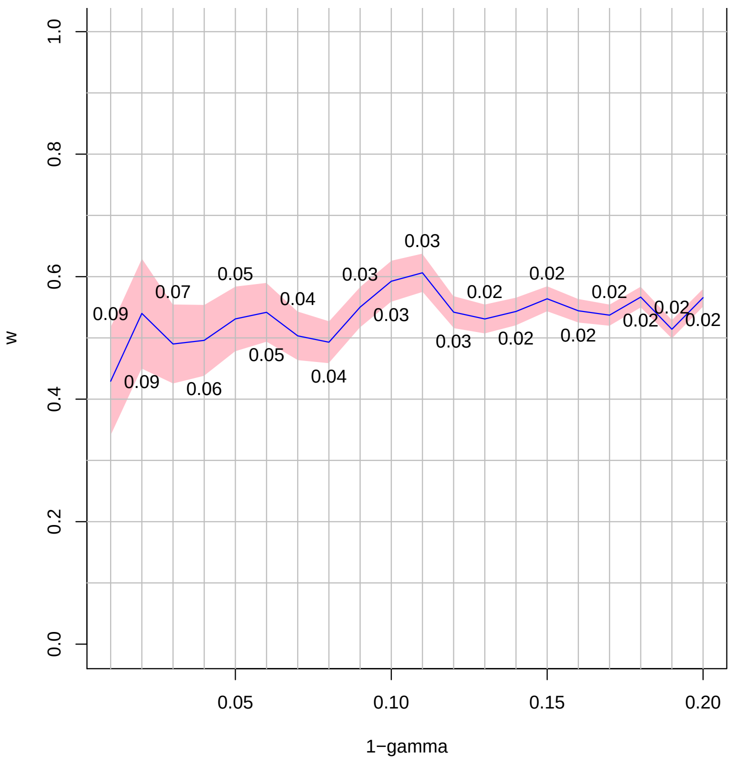

Рис. 6. Оценка ω и доверительный интервал при $k = 2000$.

Заключение

В работе исследована задача построения доверительного интервала для вероятности уклонения параметра формы (*shape*), полученного при помощи оценки Хилла. Задача решалась на основе асимптотической эффективности метода существенной выборки для $b_n > 0$.

В дальнейшем планируется рассмотреть способы оценки полученных результатов. Необходимо исследовать эффективность метода по сравнению с обычным доверительным оцениванием. Также планируется рассмотреть построение доверительных интервалов других оценок параметров асимптотических распределений.

Список литературы

1. Embrects P. Modelling Extremal Events for Insurance and Finance. New York: Springer, 1997. 655 p.
2. Гумбель Э. Статистика экстремальных значений. М.: Мир, 1965. 451 с.
3. Hill B. M. A simple general approach to inference about the tail of a distribution // Annals of Statistics, 13. 1975. P. 331–341.
4. Ermakov M. S. Importance sampling for simulatins of moderate deviation probailities of statistics // Statistics and Decisions, 25. 2007. P. 265–284.
5. Serfling R. J. Approximation Theorems of Mathematical Statistics. New York: Wiley, 1980. 371 p.
6. Ермаков С. М. Метод Монте-Карло в вычислительной математике. Санкт-Петербург, 2009. 192 с.