Доверительное оценивание параметров статистик экстремальных значений

Старков Артём, гр. 622

Санкт-Петербургский государственный университет
Математико-механический факультет
Статистическое моделирование

Научный руководитель: д.ф.-м.н., профессор Ермаков М.С.
Рецензент: д.ф.-м.н., профессор Шевляков Г.Л.

Санкт-Петербург
2018

Предсказание редких событий

Задача предсказания редких событий возникает во многих отраслях человеческой деятельности. Часто редкие события имеют распределения с тяжелыми хвостами.
Хвост распределения $F(x)$ при степенном убывании:

$$\bar{F}(x) = 1 - F(x) = Cx^{-\alpha}, \tag{1}$$

где α – параметр формы.
Оценка Хилла (*Hill, 1975*):

$$\hat{\alpha}_H^{-1} = \frac{1}{n-m+1} \sum_{i=m}^{n} \ln X_{i:n} - \ln X_{m:n}. \tag{2}$$

Метод существенной выборки

Пусть P_0 – теоретическое, $\hat{P}_n$ – эмпирическое распределение, $T(P)$ – оцениваемый функционал. Задача оценивания:

$$\omega = P(T(\hat{P}_n) - T(P_0) > b), b \in R. \qquad (3)$$

Поиск оптимальной меры Q. Решение задачи оценивания:

$$\hat{\omega}_n = \frac{1}{k}\sum_{j=1}^{k} \chi\left(T(\hat{Q}_n^{(j)}) - T(P_0) > b_n\right) \prod_{i=1}^{n} \frac{dP_0}{dQ_n}(Y_i^{(j)}). \qquad (4)$$

$$\mathrm{E}\,(\hat{\omega}_n) = \omega_n, \quad \mathrm{Var}(\hat{\omega}_n) = U_n - \omega_n^2, \qquad (5)$$

$$U_n = \mathrm{E}_{Q_n}\left[\frac{1}{k}\sum_{j=1}^{k} \chi\left(T(\hat{Q}_n^{(j)}) - T(P_0) > b_n\right) \prod_{i=1}^{n} q_n^{-2}(Y_i^{(j)})\right] \qquad (6)$$

Критерий оптимальности Q: асимптотическая эффективность

$$\overline{\lim_{n\to\infty}} \frac{\log U_n}{\log \omega_n^2} = 1. \qquad (7)$$

Асимптотически эффективная мера (Ermakov, 2007)

При соблюдении некоторых ограничений (Ermakov, 2007) процедура существенной выборки, основанная на вероятностной мере

$$q_n(x) = \lambda_n + b_n h(x) \chi \left(h(x) > -\frac{\delta}{b_n} \right), \qquad (8)$$

где $\lambda_n \in R, \delta \in [0, 1]$ – константы нормализации; $h(x)$ – функция влияния функционала $T(P)$,

$$\mathrm{E} h(x) = 0, \mathrm{E}|h(x)| < \infty, \mathrm{E} h^2(x) < \infty; \qquad (9)$$

$$b_n = \frac{N_{0,1}^{-1}(\gamma)}{\sqrt{n}\sigma(F)}, \gamma \in [0, 1], \qquad (10)$$

будет асимптотически эффективна.

В работе оптимальная мера $q_n(x) = g(x)$ используется в виде

$$g(x) = p(x)(1 + \mu b_n h(x)), p(x) - \text{базовое распределение.} \qquad (11)$$

Цель и задачи

Цель: определение границ доверительных интервалов для оценки Хилла на основе статистического моделирования методом существенной выборки.

Задачи:

- определить функцию влияния $h(x)$ и дисперсию статистики Хилла;
- разработать алгоритм моделирования случайных величин оптимальной меры $g(x)$;
- осуществить моделирование и найти оценки малых вероятностей;
- исследовать объем моделирования в зависимости от коэффициента масштаба μ;
- оценить вычислительную эффективность данной оценки по сравнению с прямым методом.

Оцениваемый функционал

Оцениваемый функционал $T(F)$:

$$T(F) = \hat{\alpha}_H^{-1} = \frac{1}{n-m+1} \sum_{i=m}^{n} Y_{i:n} - Y_{m:n}, \qquad (12)$$

где $Y = \ln X$, $Y \sim exp(\alpha)$ при $X \sim Cx^{-\alpha}$ (Embrechts, 1997).

Обозначим:

- $x_1 = F^{-1}(\beta)$, $\beta = \frac{m}{n}$ - доля выборки, не участвующая в оценке;
- $x_2 = x_1 + \alpha^{-1} : g(x_2) = exp(x_2)$.

Функция влияния $T(F)$:

$$h(x) = \frac{x - x_2}{1-\beta} \chi(x > x_1), \qquad (13)$$

Асимптотическая дисперсия $T(F)$:

$$\sigma^2(F) = \frac{1+\beta}{\alpha^2(1-\beta)}. \qquad (14)$$

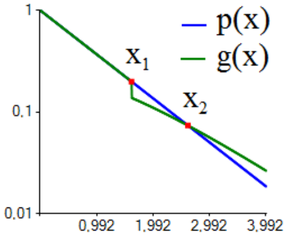
Моделирование оптимальной меры, $b_n > 0$

Оптимальная мера:

$$g(x) = \alpha e^{-\alpha x}\left(1 + \mu b_n \frac{x - x_2}{1-\beta}\chi(x > x_1)\right). \tag{15}$$

Моделирование, $b_n > 0$.

- $x \in [0, x_1)$: $g(x) = exp(x; \alpha)$.
- $x \in [x_1, x_2)$: Метод мажорант. Мажорирующая функция $g_m(x) = exp(x; \alpha)$.
- $x \in [x_2, \infty)$: Метод композиции.



$$g(x) = exp(x; \alpha) + \frac{b_n}{e\alpha}\gamma(x; 2, \alpha^{-1}), \tag{16}$$

где $\gamma(x; k, \theta)$ – гамма-распределение:

$$\gamma(x; k, \theta) = \frac{1}{\theta^k \Gamma(k)} x^{k-1} e^{\frac{-x}{\theta}}, \tag{17}$$

Моделирование оптимальной меры, $b_n < 0$

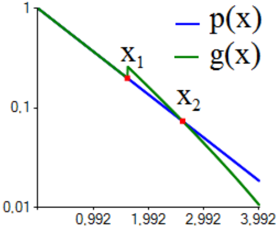
Оптимальная мера:

$$g(x) = \alpha e^{-\alpha x}\left(1 - \mu b_n \frac{x - x_2}{1-\beta}\chi(x > x_1)\right). \qquad (18)$$

Моделирование, $b_n < 0$

- $x \in [0, x_1)$: $g(x) = exp(x;\alpha)$.
- $x \in [x_1, x_2)$: Метод мажорант. Мажорирующая функция:

$$g_m(x) = \alpha e^{-\alpha x}\left(1 + b_n \frac{x_2 - x_1}{1-\beta}\right). \quad (19)$$



- $x \in [x_2, \infty)$: Метод мажорант. Мажорирующая функция $g_m(x) = exp(x;\alpha)$.

Построение доверительных интервалов

Оценка вероятности уклонения:

$$\hat{\omega}_n = \frac{1}{k} \sum_{j=1}^{k} \chi\big(T(\hat{Q}_n^{(j)}) - T(P_0) > b_n\big) \prod_{i=1}^{n} \frac{1}{1 + \mu b_n h(Y_i^{(j)})}. \quad (20)$$

Доверительный интервал уровня ω:

$$\bar{\alpha}^{-1} = \hat{\alpha}_H^{-1} \pm b_n(\gamma) \frac{\sqrt{n}}{\sigma}. \quad (21)$$

Нормальная асимптотика:

$$x = \frac{b_n \sqrt{n}}{\sigma} = \frac{N^{-1}(\gamma)}{\sigma^2}. \quad (22)$$

Построение доверительных интервалов

Пример построения доверительных интервалов для различных n.

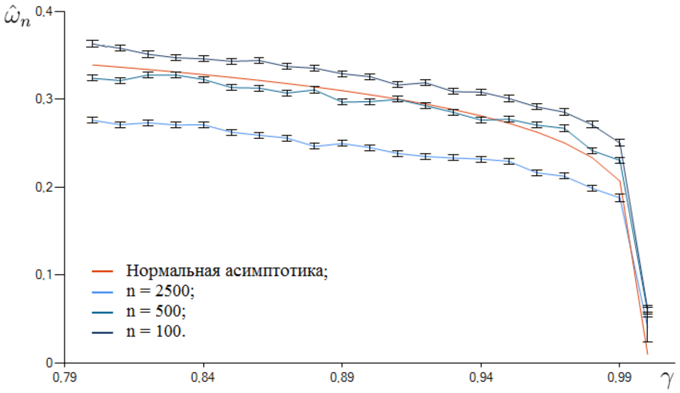


Рис.: Вероятность уклонения от $\gamma; \alpha = 2, \beta = 0.90, n = 100, 500, 2500.$

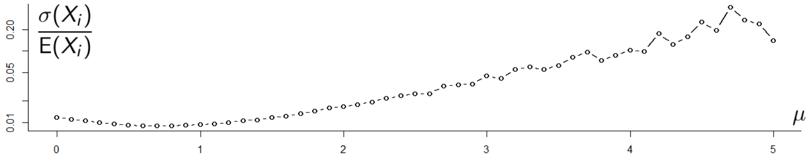
Анализ формы оптимальной меры

Плотность оптимальной меры:

$$g(x) = p(x)(1 + \mu b_n h(x)), \tag{23}$$

$$\mu_{min} = \arg\min_{\mu} \frac{\sigma(X_i)}{\mathrm{E}(X_i)}, X_i \sim g(x;\mu). \tag{24}$$

При больших μ оценка ведет себя нестабильно. Пример зависимости нормализованной дисперсии от μ:



Анализ формы оптимальной меры

Таблица: Результат поиска оптимального μ для $\alpha = 1$

	Beta							
	μ_{min}				Normalized sigma ratio			
γ	0.80	0.85	0.90	0.95	0.80	0.85	0.90	0.95
0.90	0.9	0.9	0.6	0.7	1.03	1.03	1.03	1.05
0.92	0.7	0.8	0.6	0.7	1.06	1.06	1.09	1.06
0.94	0.6	0.5	0.6	0.6	1.07	1.07	1.13	1.14
0.96	0.7	0.5	0.6	0.5	1.13	1.15	1.18	1.21
0.98	0.5	0.6	0.4	0.5	1.26	1.29	1.33	1.41

Сравнение эффективности

Прямая оценка вероятности уклонения:

$$\hat{\omega}_D = \frac{1}{k}\sum_{i=1}^{k}\chi\big(T(\hat{P}_n^{(i)}) - T(P_0) > b_n\big), \qquad (25)$$

где $P = exp(\alpha)$. Дисперсия оценки:

$$D\hat{\omega}_D = \sqrt{\frac{\hat{\omega}_D(1-\hat{\omega}_D)}{k}}. \qquad (26)$$

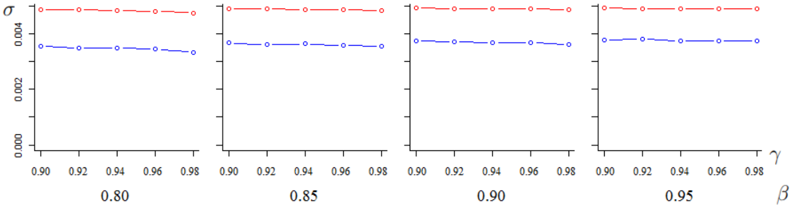


Рис.: σ прямого метода (красный) и метода сущ. выборки (синий)

Заключение

В работе исследована задача построения доверительного интервала для вероятности уклонения параметра формы, полученного при помощи оценки Хилла, на основе асимптотически эффективной процедуры существенной выборки.

- Найден явный вид асимптотически эффективной меры моделирования. Предложен и осуществлен алгоритм моделирования случайных величин, по этой мере.
- Проанализирована эффективность метода в зависимости от параметров оценки. Найдены оптимальные параметры, с их учетом осуществлено моделирование для построения доверительных интервалов.
- Получены результаты сравнения дисперсий данного метода и прямой оценки. Представлены примеры построения доверительных интервалов.