Метод двойного бутстрапа для оценивания степенного индекса с помощью экспектилей

Андрей Евгеньевич Лукьянов, ММФ НГУ, Новосибирск

ПРАВИЛЬНО МЕНЯЮЩИЙСЯ КЛАСС РАСПРЕДЕЛЕНИЙ И СТЕПЕННОЙ ИНДЕКС

Для изучения экстремальных событий и их моделирования в теории экстремальных значений вводится правильно меняющийся класс распределений.

Если $\overline{F}(x) = 1 - F(x)$ - хвост функции распределения, то для каждого правильного меняющегося распределения с некоторым коэффициентом α, являющимся степенным индексом верно: $\overline{F}(x) \sim L(x)x^{-\alpha}$, где $L(x)$ - медленно меняющаяся функция ($\lim\limits_{t\to\infty}\frac{L(tx)}{L(x)} = 1$).

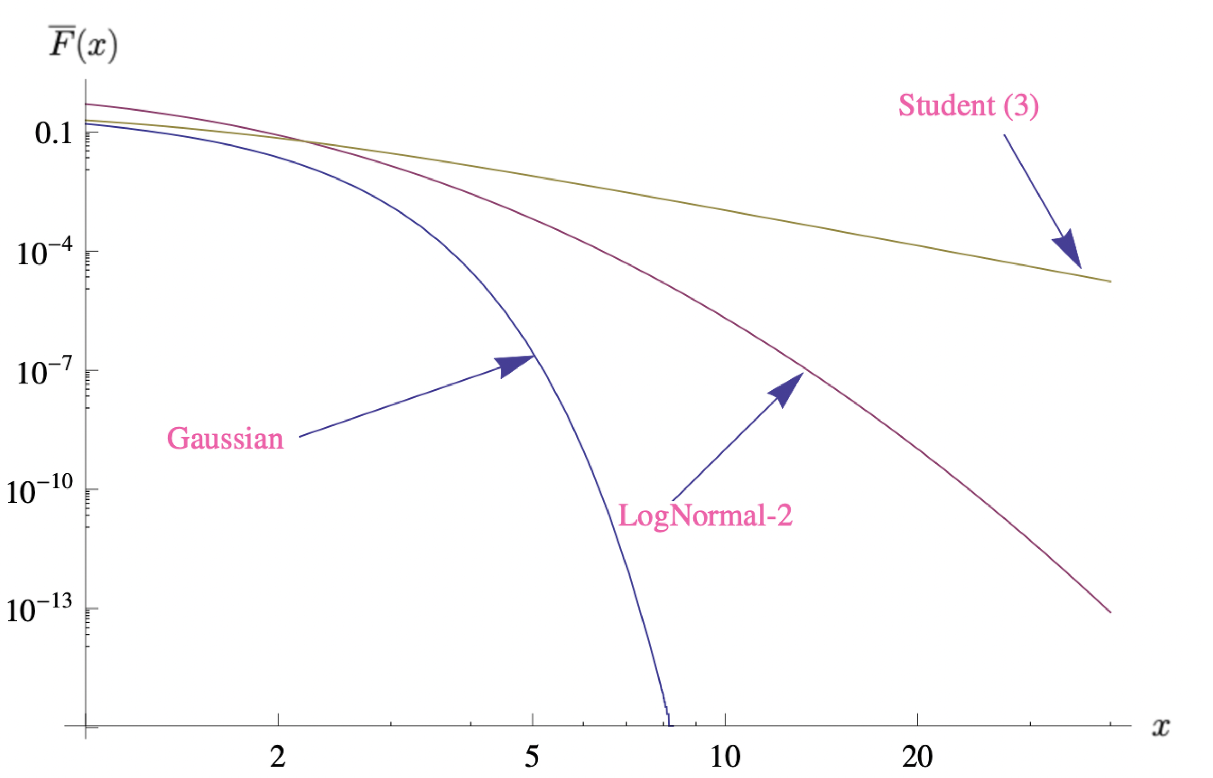

Рис.: log-log график хвоста функции распределения. Правильно меняющиеся распределения (вроде Стьюдента-3) асимптотически представляют из себя прямую линию.
[N. Taleb, Statistical Consequences of Fat Tails, 2020]

Правильно меняющиеся распределения часто возникают в экономических и естественных науках:

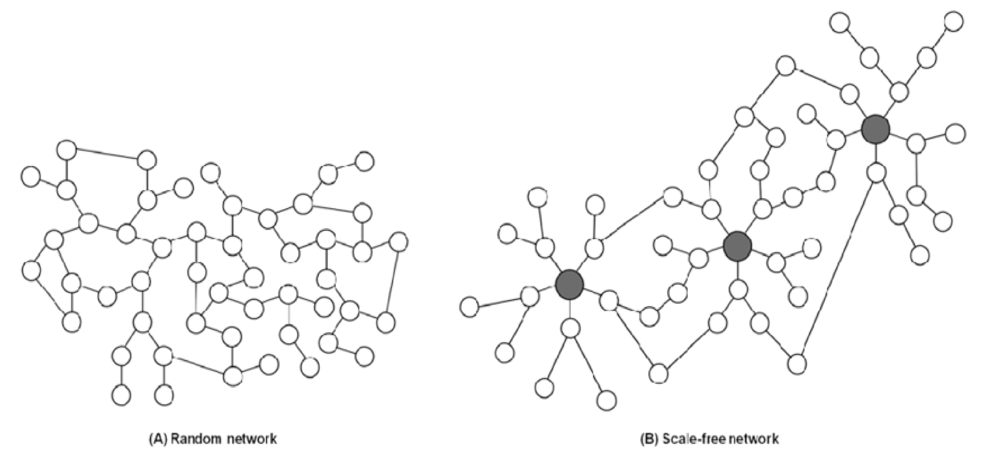

Рис.: Безмасштабные сети

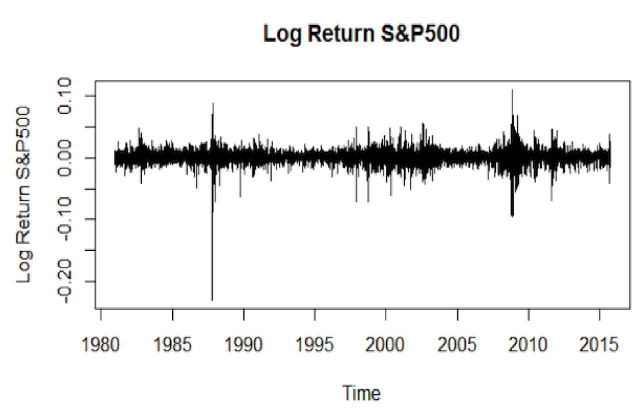

Рис.: Log-return индекса S&P500

ОЦЕНКИ ТИПА ХИЛЛА И МЕТОД ДВОЙНОГО БУТСТРАПА ДЛЯ НИХ

В теории экстремальных значений оценивается параметр $\gamma = \frac{1}{\alpha}$.

Положим $M^{(p)} = \frac{1}{k}\sum\limits_{i=0}^{k-1}\left(\log X_{(n-i,n)} - \log X_{(n-k,n)}\right)^p$, где

$X_{(i,n)}$ - i-ая порядковая статистика выборки размера n.

- Оценка Хилла : $\hat{\gamma}_H = M^{(1)}$,
- Оценка моментов : $\hat{\gamma}_M = M^{(1)} + 1 - \frac{1}{2}\left(1 - \frac{M^{(1)}}{M^{(2)}}\right)^{-1}$,
- Ядерная оценка : $\hat{\gamma}_K = \Psi\left(\frac{i}{n}, \phi(\frac{i}{n})\right) - 1 + \frac{\Psi\left((\frac{i}{n})^\lambda, \phi(\frac{i}{n})\right)}{\Psi\left(\frac{\partial(u^{\lambda+1}\phi(u))}{\partial u}(i/n), 1\right)}$, для некоторого $\lambda > 0$ и ядра

  $\phi : (0,1) \to [0,\infty)$, где $\Psi(a,b) = \sum\limits_{i=0}^{n-1} ab\log\frac{x_{(i)}}{x_{(i+1)}}$,

Для каждой из оценок доказаны состоятельность и асимптотическая нормальность [Хилл, 1975; Деккерс, 1989; Грюнбум, 2003]. Однако каждый из методов зависит от числа используемых последних порядковых статистик k, такого что:

- $k = k(n) \to \infty$ *при* $n \to \infty$
- $\frac{k(n)}{n} \to 0$

При этом выбор k существенно влияет на поведение оценок:

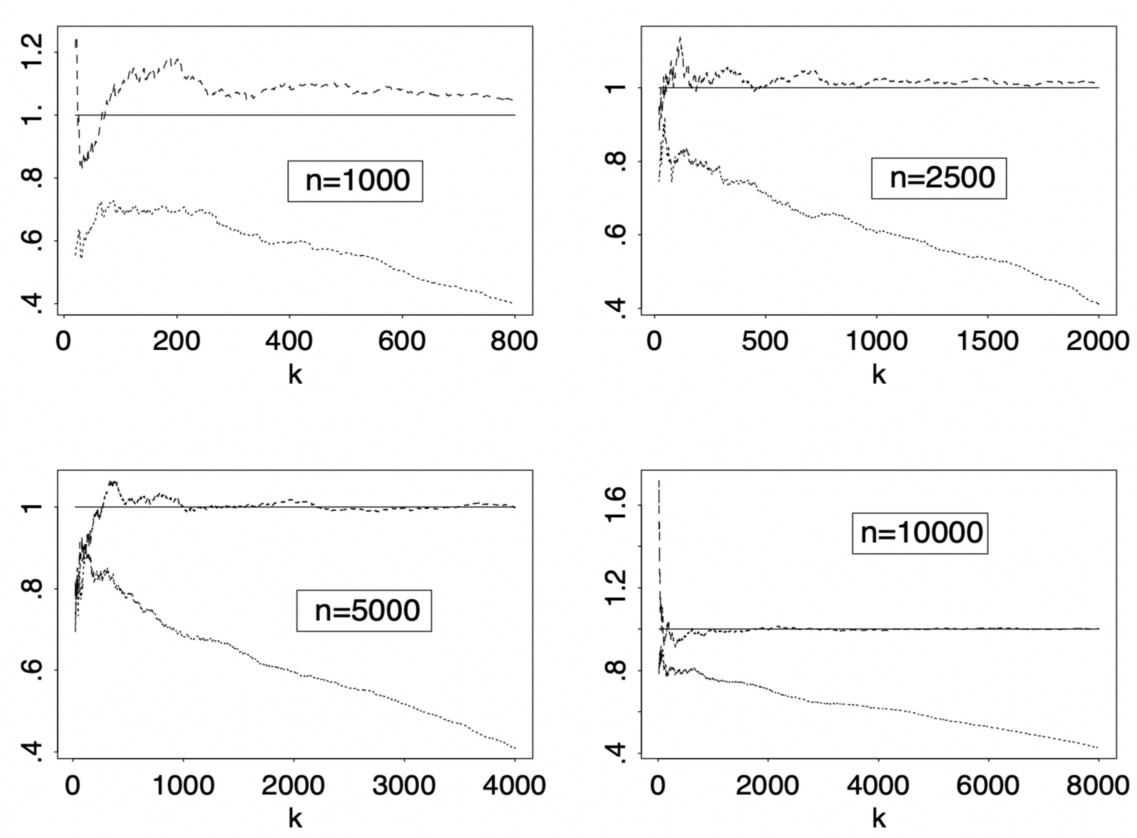

Рис.: На графиках показаны результаты оценивания α при помощи оценки Хилла в зависимости от количества используемых порядковых статистик: $\overline{F_1}(x) = \frac{1}{x}$ (длинный пунктир) и $\overline{F_2}(x) = \frac{1}{x\ln x}$ (короткий пунктир), $(x > 1.7)$, $(\alpha = 1$, прямая линия).

Для хвоста $\overline{F_2}(x)$ с логарифмической компонентой результаты намного хуже.
[Embrechts et al., Modelling Extremal Events, 1997]

Метод двойного бутстрапа для выбора оптимального k(n):

Для выбора k(n) был представлен метод двойного бутстрапа [Драйзма и др., 1999]:

- Сгенерировать с одинаковой вероятностью из выборки r новых выборок размера $n_1 = \left[\sqrt{\frac{n}{2}}\right]$.
- Выбрать две состоятельные оценки степенного индекса $\gamma^{(1)}_{\kappa_1,j}$ и $\gamma^{(2)}_{\kappa_1,j}$ посчитанные по κ_1 последним порядковым статистикам выборки j, $\kappa_1 = 1,\ldots,n_1$, $j = 1,\ldots,r$.
- Найти оптимальное $\kappa_1^* = \arg\min\limits_{\kappa_1}\frac{1}{r}\sum\limits_{j=1}^{r}\left(\gamma^{(1)}_{\kappa_1,j} - \gamma^{(2)}_{\kappa_1,j}\right)^2$.
- Повторить вышеописанную процедуру для $n_2 = [\frac{n}{2}]$ и найти оптимальное κ_2^*.
- Определить оптимальное κ^* для изначальной выборки: $\kappa^* = A(\kappa_1^*, n_1, n)\frac{(\kappa_1^*)^2}{\kappa_2^*}$, где функция A дополнительно зависит от выбранных оценок.

$\left(\frac{(\log k_1^*(n_1))^2}{(2\log n_1 - \log k_1^*(n_1))^2}\right)^{(\log n_1 - \log k_1^*(n_1))/\log n_1}$ - вид функции $A(\kappa_1^*, n_1, n)$ для оценки моментов

ЭКСПЕКТИЛИ И ИХ СВОЙСТВА

Экспектиль представляет из себя решение следующей оптимизационной задачи:
$e_\tau = \arg\min\limits_{\mu}\tau\int_\mu^\infty(x-\mu)^2dF(x) + (1-\tau)\int_{-\infty}^{\mu}(x-\mu)^2dF(x)$ для заданного уровня $\tau \in [0,1]$ и функции распределения $F(x)$.

Экспектиль является аналогом квантили, так как квантиль является решением похожей оптимизационной задачи: $q_\tau = \arg\min\limits_{\mu}\tau\int_\mu^\infty|x-\mu|dF(x) + (1-\tau)\int_{-\infty}^{\mu}|x-\mu|dF(x)$.

При этом, если $q_{0.5} = med(X)$, то $e_{0.5} = \mathbb{E}X$.

Свойства экспектилей:

- Однозначно определяют функцию распределения случайной величины: $e_{\tau,X} = e_{\tau,Y}$, $\forall\tau \in (0,1) \iff X \stackrel{d}{=} Y$.
- Устойчивы к аффинным преобразованиям: $X = aY + b \Rightarrow e_{\tau,X} = ae_{\tau,Y} + b$.
- Субаддитивны: $\forall\tau \leqslant \frac{1}{2}$, $e_{\tau,Y+Z} \geqslant e_{\tau,Y} + e_{\tau,Z}$;
  $\forall\tau \geqslant \frac{1}{2}$, $e_{\tau,Y+Z} \leqslant e_{\tau,Y} + e_{\tau,Z}$.
- Экспектильная функция $F : \tau \to e_\tau$ является непрерывной монотонно возрастающей функцией на отрезке [0,1].
- Экспектили уровня τ случайной величины Y являются квантилями уровня τ случайной величины $\tilde{Y}$ со следующей функцией распределения: $H(y) = \frac{\mathbb{E}(|Y-y|\mathbb{I}\{Y\leqslant y\})}{\mathbb{E}|Y-y|}$.

При этом экспектили позволяют решать некоторые задачи страхования при помощи экспектильной регрессии, которая является обобщением квантильной регрессии. Экспектильная регрессия, в отличие от квантильной, более гладко оценивает экстремальные события.

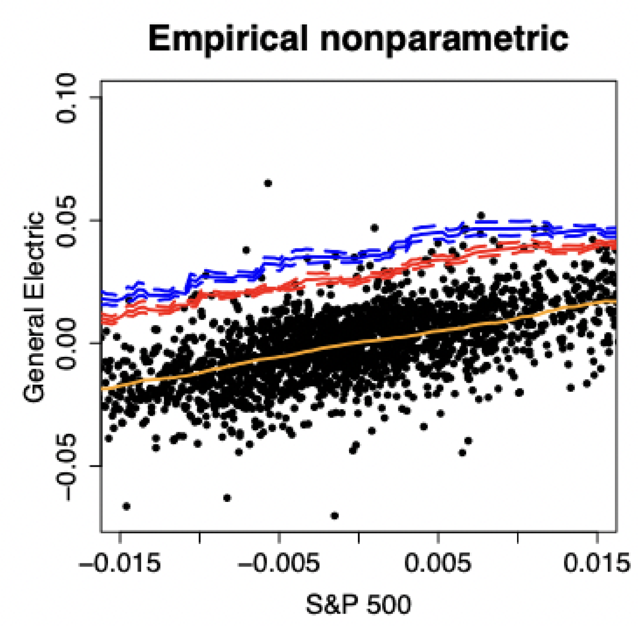

Рис.: Зависимость log-returns SP500 от log-returns General Electric.
Оранжевая линия: непараметрическая регрессия Надарая-Ватсона.
Красная линия: экспектильная регрессия, пунктирные линии образуют доверительный интервал уровня 0.95.
Синяя линия: квантильная регрессия, пунктирные линии образуют доверительный интервал уровня 0.95.
[Daouia et al., Inference for extremal regression with dependent heavy-tailed data, 2022]

ЭКСПЕКТИЛЬНЫЕ ОЦЕНКИ СТЕПЕННОГО ИНДЕКСА И ИХ ПРИМЕНЕНИЕ

Экспектили так же позволяют оценить степенной индекс [Дауя и др., 2018].
Следующие оценки являются состоятельными и асимптотически нормальными; каждая из них зависит от количества доли взятых элементов выборки τ_n:

- $\gamma_{exp*} = (1 + \frac{\hat{\overline{F}}_n(\hat{e}_{\tau_n})}{1-\tau_n})^{-1}$,
- $\gamma_{exp1} = \int_0^1\log\left(\frac{\hat{e}_{1-(1-\tau_n)s}}{\hat{e}_{\tau_n}}\right)ds$,
- $\gamma_{exp2} = \frac{1}{l}\sum\limits_{i=1}^{l}\log\frac{\hat{e}_{1-(1-\tau_n)(i-1)/l}}{\hat{e}_{\tau_n}}$,
- $\gamma_{exp3} = \frac{1}{\lfloor n(1-\tau_n)\rfloor}\sum\limits_{i=1}^{\lfloor n(1-\tau_n)\rfloor}\log\frac{\hat{e}_{1-(i-1)/n}}{\hat{e}_{1-\lfloor n(1-\tau_n)\rfloor/n}}$,

  $\hat{e}_{\tau_n} = \arg\min\limits_{u\in\mathbb{R}}\sum\limits_{i=1}^{n}|\tau_n - \mathbb{I}(Y_i - u \leqslant 0)|(Y_i - u)^2$ - эмпирическая экспектиль уровня τ_n,

  $\overline{F}^*_n(u) = n^{-1}\sum\limits_{i=1}^{n}\mathbb{I}(Y_i > u)$ - эмпирический хвост функции распределения в точке u.

Особый интерес представляет собой последняя оценка, поскольку имеет форму оценки Хилла.
Полученные оценки используются для оценивания различных мер риска, например, ES(Expected Shortfall), [Дауя, Жирар, 2018].

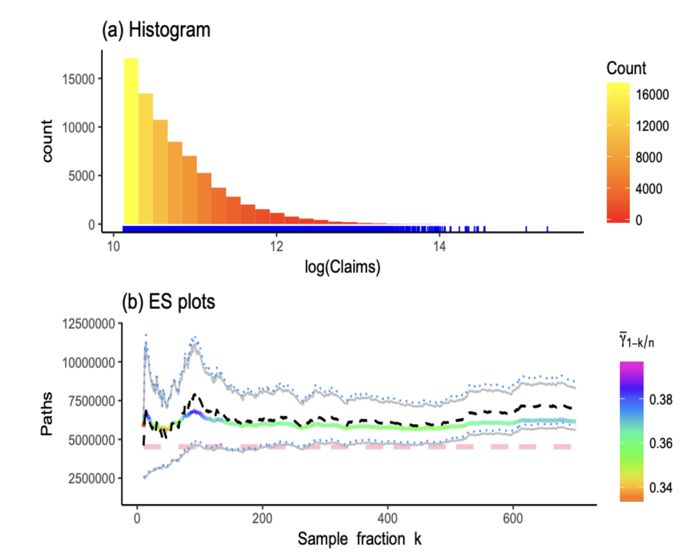

Рис.: Верхний график: Гистограмма log-claims(логарифмов выплат) страховой компании.
Нижний график: Expected Shortfall(ES) оценки:
ES(оцененный экспектилями) - линия радужного цвета, пунктирные линии голубого цвета: доверительные интервалы уровня 0.95.
ES(оцененный квантилями) - пунктирная линия черного цвета, линии серого цвета: доверительные интервалы уровня 0.95.
$Y_{(n,n)}$ - пунктирная линия розового цвета.
[Daouia et al., ExpectHill estimation, extreme risk and heavy tails, 2018]

ПОСТАНОВКА ЗАДАЧИ

В рамках данной задачи планируется создание метода двойного бутстрапа, поиска оптимального уровня доли используемой выборки для оценок степенного индекса по экспектилям, а также исследование их статистических свойств. Также планируется реализовать искомый метод на языке Python и сравнить его с уже созданным методом двойного бутстрапа для оценок типа Хилла.
Работа выполняется под научным руководством к.ф.-м.н. Е. И. Прокопенко.

ЛИТЕРАТУРА

1. A. Daouia, S. Girard and G. Stupffer: ExpectHill estimation, extreme risk and heavy tails (2018).
2. I. Voitalov, P. van der Hoorn, R. van der Hofstad, D. Krioukov: Scale-free networks well done (2018).
3. G. Draisma, L. de Haan, L. Peng, and T. T. Pereira: A Bootstrap-based Method to Achieve Optimality in Estimating the Extreme-value Index, Extremes 2(1999), 367.
4. A. Daouia, S. Girard and G. Stupffer: ExpectHill estimation, extreme risk and heavy tails, TSE-953(2018).

Novosibirsk State University
*THE REAL SCIENCE
mathematics & mechanics