

ЭМПИРИЧЕСКИЕ СТЕПЕННЫЕ РАСПРЕДЕЛЕНИЯ – 3 + УРОВНЯ ОПИСАНИЯ

Родкин М.В.

ИТПЗ РАН, rodkin@mitp.ru

Содержание доклада:

Можно полагать, что самоподобие и степенные распределения – как бы разные представления единого явления: первое более геометрическое и визуальное, второе более статистическое (в определённых пределах из первого следует второе и наоборот). При не целых коэффициентах степени мы имеем фрактальные объекты. Фрактальная концепция завораживающе красива и в её рамках выполнено огромное число работ.

Некоторые «наши»-сейсмологические результаты применения (на мой взгляд, несмотря на огромное число работ, они не столь впечатляющи):

- Обобщённый закон подобия для землетрясений (здесь его и докладывали и критиковали, Кособоков, Молчан)**
- Уменьшение фрактальной размерности в области подготовки сильного события (вполне ожидаемо)**
- Уменьшение наклона графика повторяемости в области подготовки сильного события.**

Это как бы и почти тавтология. Так уменьшение b можно переформулировать: прогнозным признаком сильного события является рост средней магнитуды землетрясений (одна из формул расчёта: $b = \lg(e)/(M_{\text{mean}} - M_0)$).

При обработке геофизических данных, мы обычно сталкиваемся со степенными распределениями, а не с анализом карт разломов или изображений. Рассмотрим, что можно сказать и какие «подводные камни» возникают при анализе степенных распределений.

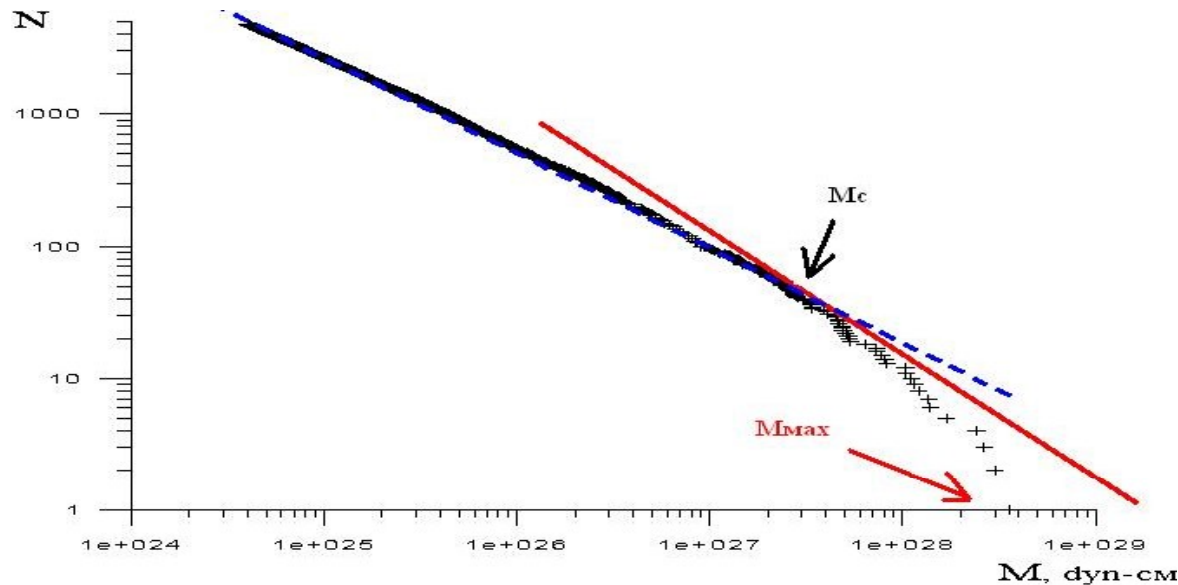
ремарка:

Большая часть материалов, на которых я сегодня буду останавливаться, были получены и опубликованы в совместных работах с В.Ф.Писаренко. Но в вопросах интерпретации мы несколько разошлись, ниже я даю свою интерпретацию.

Мне показалось, что нижеследующее стоит представить как в методическом плане (это полезно иметь в виду при работе со степенными распределениями), так и как некий пример того слона, который оказывается существенно разным, в зависимости от того, ощупывают ли его ногу, хвост, или хобот.

Классический пример - закон распределения землетрясений.

Часто эмпирические степенные распределения дают расходящийся ряд.
Известный пример - распределение обычных и сильных землетрясений

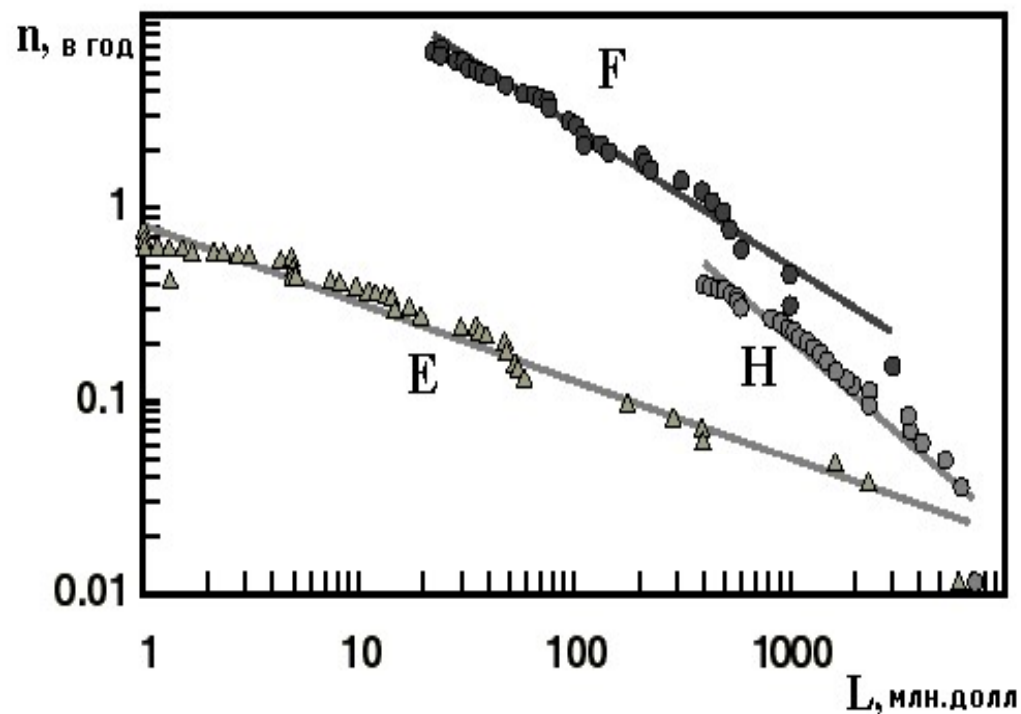


- + - данные Гарвардского каталога сейсмических моментов.
- - обычный закон Гутенберга-Рихтера, $\beta \approx 0.65$ (диапазон $\beta < 1$, бесконечное среднее, бесконечная мощность источника энергии сейсмогенеза);
- - критический наклон графика Г-Р, $\beta = 1$ (область M_c , диапазон $\beta \approx 1$).
Область M_{max} - диапазон $\beta > 1$ (неизвестное распределение с конечным средним значением и конечной мощностью источника энергии сейсмогенеза).

Распределение сильных землетрясений явным образом отличается от обычного неограниченного закона Г-Р с $\beta < 1$ и (формально) бесконечным средним значением.

Очевидно, Земля конечна, бесконечно сильных событий быть не может, безграничных распределений тоже?

Другой пример степенного распределения



Распределение величин экономических потерь от единичных событий (млн. долл. США) за год для: наводнений (F), землетрясений (E) и ураганов (H).

Данные по США за 1900-1989 гг. Линиями даны графики аппроксимирующих степенных распределений со значениями показателя степени $\beta=0.74$ (для наводнений), $\beta=0.98$ (для ураганов) и $\beta=0.41$ (для землетрясений).

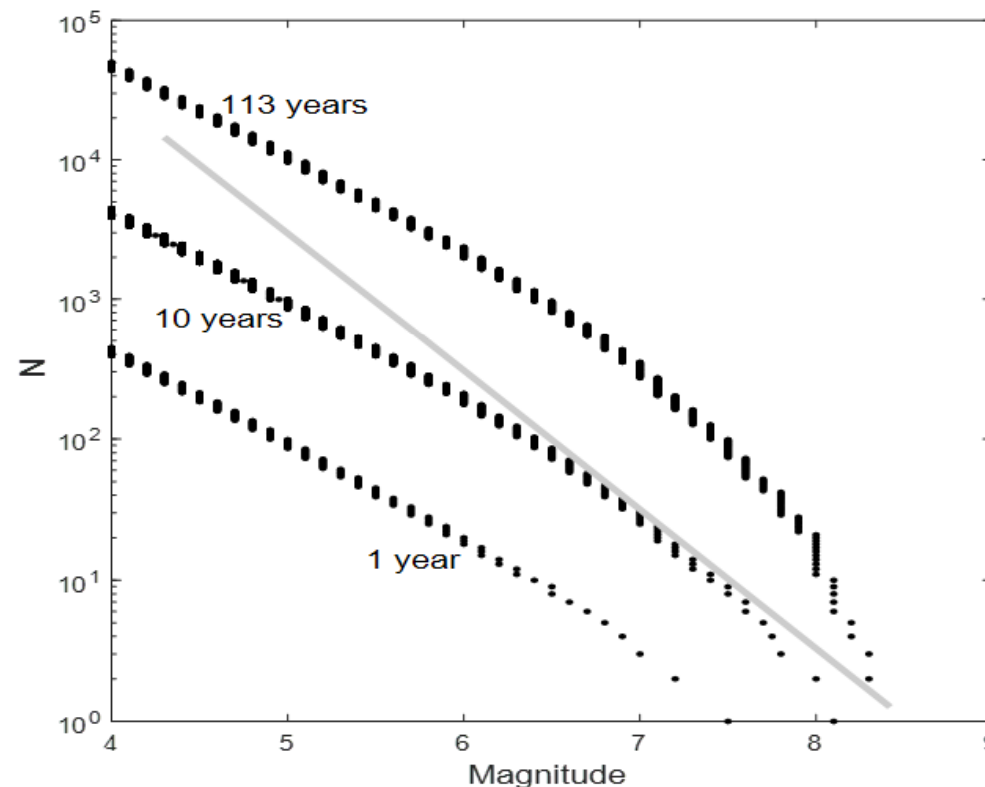
По обоим осям **x** и **y** – логарифмический масштаб.

Заметим, как бы намечается «загиб вниз»

Поведение и возможности описания на разных интервалах времени (с разной доступной статистикой) **РАЗНЫЕ**.

Можно выделить как бы 3 уровня описания

Пример – каталог Японского Метеорологического Агентства
(типичный хороший каталог)



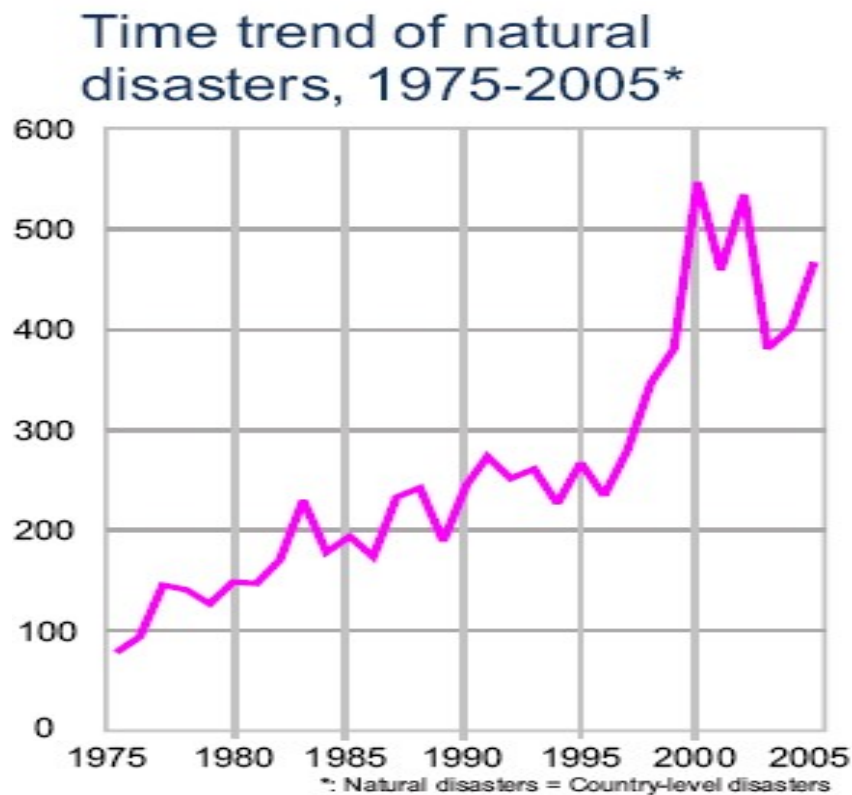
Три уровня описания степенных распределений по данным каталога землетрясений JMA. Указаны интервалы времени наблюдения – 1 год, 10 лет и весь каталог (для 1 и 10 лет – типичные распределения, полученные бутстреп-методом).

Точка касания прямой с наклоном $\beta=-1$ со средним графиком характеризует величину M_{ch} .

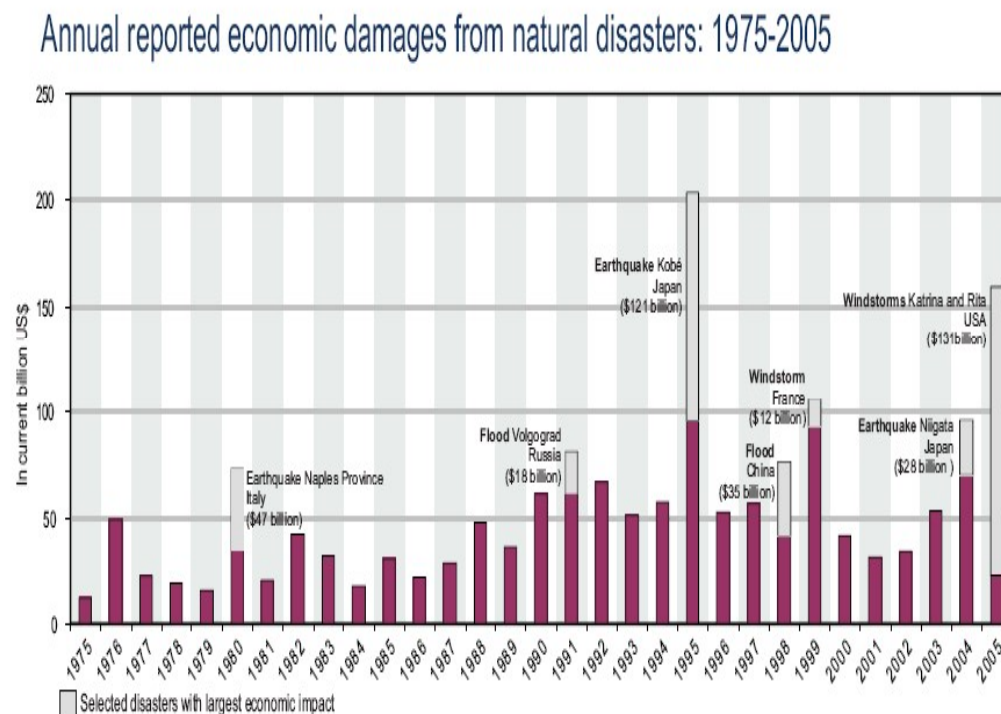
Диапазон 1

.

Как бы парадокс или опасность: мировые данные роста числа природных катастроф и величин ущерба (аналогично, как бы бесконечная сейсмическая энергия)



Изменения числа катастроф:
International Disaster Database
(www.em-dat.net).



Данные по ущербам за год: International Disaster Database (www.em-dat.net).

Свойства степенного распределения с показателем степени $\beta \leq 1$

Степенной закон Парето с функцией распределения $F(x)$:

$$F(x) = 1 - (A/x)^\beta; \quad x \geq A, \quad (1)$$

Если $\beta \leq 1$ то закон Парето (1) имеет «тяжелый хвост» “heavy tail” с бесконечным средним значением и (тем более) дисперсией. Использование выборочных средних (и дисперсии) даёт некорректный результат.

Медианы кумулятивного эффекта Σ_n и максимальное единичное значение μ_n нелинейное растут с числом случаев наблюдения (временем):

$$\Sigma_n \sim n^{1/\beta} \quad \text{и} \quad \mu_n \sim n^{1/\beta} \quad (2)$$

$$\Sigma_t \sim t^{1/\beta} \quad \text{and} \quad \mu_t \sim t^{1/\beta} \quad (2a)$$

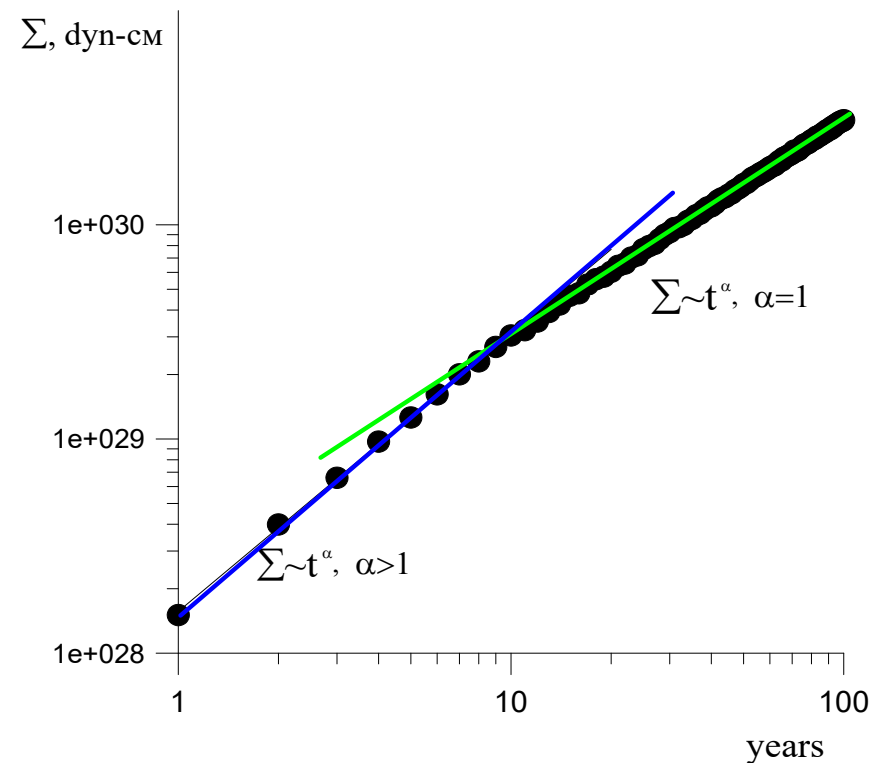
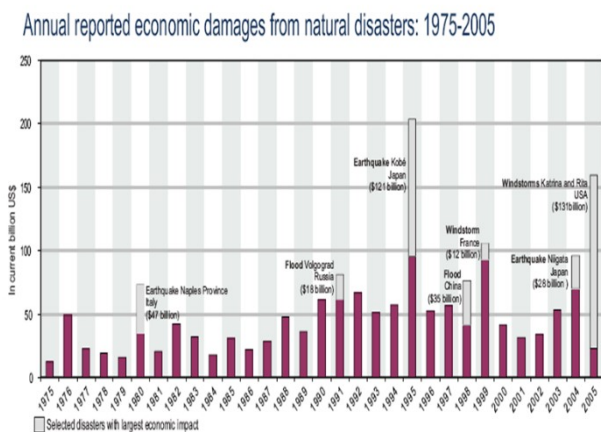
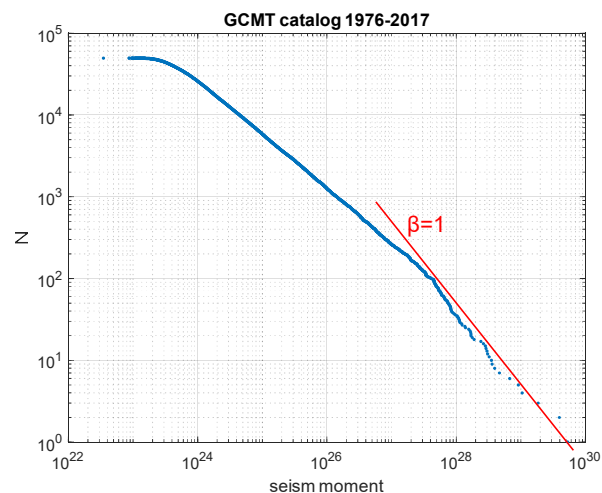
Рост (2) может быть некорректно интерпретирован как указание на нестационарность режима (в частности, быстрый рост ущербов от природных катастроф).

Диапазон 2

.

Связь графика повторяемости (слева) и кумулятивных сумм эффекта за время t (справа).

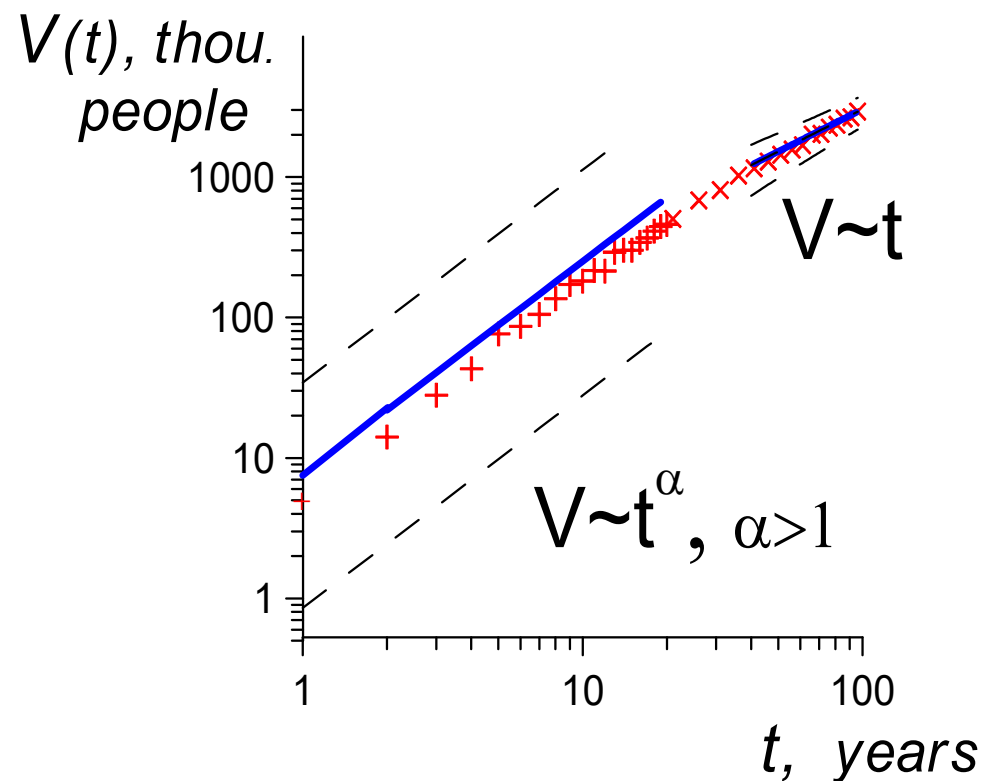
Бутстреп-метод позволяет из довольно «экзотической случайной кривульки» роста кумулятивной выделенной сейсмической энергии или момента получить весьма гладенькую зависимость с точкой изгиба



Слева – примеры графиков повторяемости каталог GCMT, В – статистически аналогичный на срок 100 лет. Использован бут-стреп boot-strap метод численного моделирования.

Результаты моделирования типичного числа жертв за t лет

$V(t)$



Можно видеть изменение режима роста характерных кумулятивных величин (выделенной сейсмической энергии (момента), величины потерь) с нелинейного на линейный режим роста : $V \sim t$

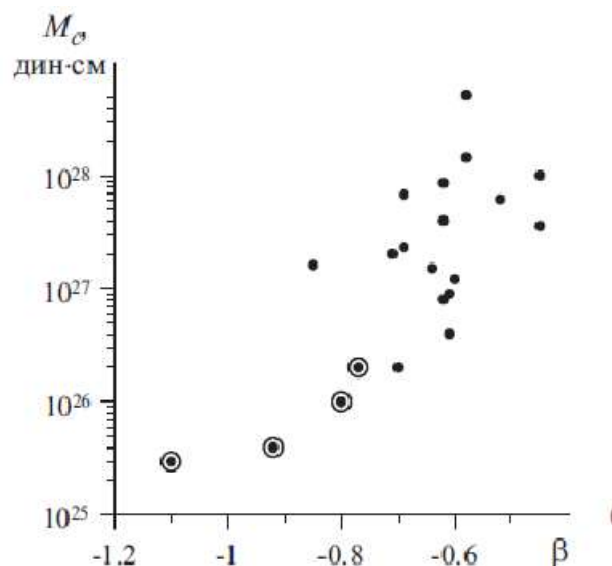
Мс – Ммах-характ - информативная устойчивая характеристика режима

Мс для сейсмического режима значительно различна для зон СОХ, зон конт коллизии и зон субдукции (кружочки – районы СОХ, точки – зоны субдукции) .

Мс более устойчивая характеристика, чем наблюдаемое макс событие.

Мс для числа жертв от природных катастроф имеет тенденцию существенно уменьшаться в развитых странах со временем.

С интервалом магнитуд $M = M_c$ связан МАКСИМУМ выделения сейсмической энергии, момента, максимум ущерба

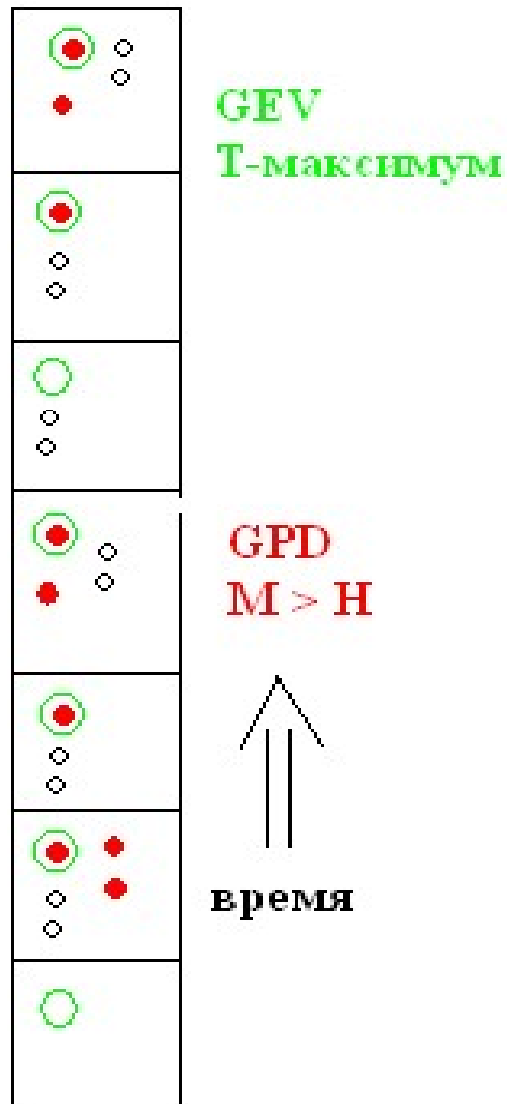


Disaster type	Region, country	Recurrence time T_p , years	Characteristic event, D_p , number of fatalities	Maximum event, number of fatalities	Data source, data base
Earth-quakes	Developed countries				Significant earthquakes, NOAA, https://www.ngdc.noaa.gov/nndc/struts/form
	1900-1959	33	95000	110000	
	1960-1999	30	24000	17000	
	Developing countries				Em-dat, The International disaster database, https://www.emdat.be
	1900-1959	40	270000	200000	
	1960-1999	60	260000	240000	
Floods	North America and European Union				
	1950-1980	15	1500	650	
	1980-2005	10	500	200	
	SE and S Asia 1984-2006	20	10000	6000	

Диапазон 3

.

Подходы на основе предельных распределений: Обобщенного распределения экстремальных значений (**GEV**) и Обобщенного распределения Парето (**GPD**) теории экстремальных значений



ПОЯСНЕНИЕ МЕТОДИКИ

2 предельных распределения
- 2 подхода к описанию хвоста
распределения (к распределению
редких сильных событий):

1. Рассматриваются события
превосходящие некоторый уровень,
 $M > H$.
Метод GPD

2. Рассматривается распределение
максимальных значений за некие
равные интервалы времени.
Метод GEV.

Обобщенное Распределение Парето, GPD

Определения:

Правая предельная точка x_F функции распределения $F(x)$: $x_F = \sup \{x: F(x) < 1\}$

Экссесс-функция $F_h(x)$ (условная функция распределения выше порога h):

$$F_h(x) = P\{X - h < x \mid X > h\}, \quad x \geq 0.$$

Теорема Гнеденко-Пикандса-Балкема-Хаана

Пусть $x_1, x_2, \dots, x_n$ - выборка n независимых случайных величин с непрерывной функцией распределения $F(x)$ и экссесс-функцией $Fh(x)$, $h > 0$. Если существуют такие нормирующие константы c_n, d_n , что нормированный максимум $(M_n - d_n)/c_n$ сходится по распределению к некоторой невырожденной случайной величине. Тогда найдется такая неотрицательная функция $s(h)$, что

$$\lim_{h \uparrow x_F} \sup_{0 \leq x \leq x_F - h} |Fh(x) - Gh(x) \xi s(h)| = 0.$$

Для любой исходной функции распределения $F(x)$ экссесс-функция $F_h(x)$ значений $(x-h)$ сходится в пределе к Обобщенному Распределению Парето (GPD):

$$GPD_h(x \mid \xi, s) = 1 - (1 + \xi(x-h)/s)^{-1/\xi}, \quad x \geq h; \quad 1 + \xi(x-h)/s \geq 0.$$

Задается параметр порога h , подлежат определению параметр масштаба s и параметр формы ξ . Параметр формы ξ характеризует вид исходного распределения $F(x)$:

Для ограниченного распределения $F(x)$ имеем $\xi < 0$.

(что, как выяснилось, и имеет обычно место – и ДОЛЖНО иметь место - в реальности)

Для экспоненциального поведения хвоста $F(x)$ имеем $\xi = 0$.

(как закон распределения по Я.Кагану)

Для степенного поведения хвоста $F(x)$ имеем $\xi > 0$, $\xi = 1/\beta$.

Параметр формы GEV γ = параметру формы GPD ξ .

Обобщенное распределение экстремальных значений (GEV)

$$GEV(x|m,s,\gamma) = \exp\{ -[1+\gamma (x-m) /s]^{-1/\gamma} \},$$

Определению подлежат 3 параметра: параметр центрирования - m ,
параметр масштаба s , и параметр формы - γ .

Параметр формы γ изменяется от $-\infty$ до $+\infty$, его знак определяет тип предельного распределения и область изменения

$\gamma > 0$; $x > m - s/\gamma$; Pareto; степенное распределение
 $\gamma < 0$; $x < m - s/\gamma$; Weibull; ограниченное распределение
 $\gamma = 0$; $|x| < \infty$; Gumbel; экспоненциальное.

Примеры GPD распределений

При уменьшении параметра формы ξ по абсолютной величине ($\xi < 0$) «носик распределения право» становится все длиннее и все тоньше.

$$\underline{M_{max} = h - s/\xi}$$

Это является математическим выражением неустойчивости оценок величины

M_{max} , $\xi < 0$ при этом часто $\xi \sim 0$.

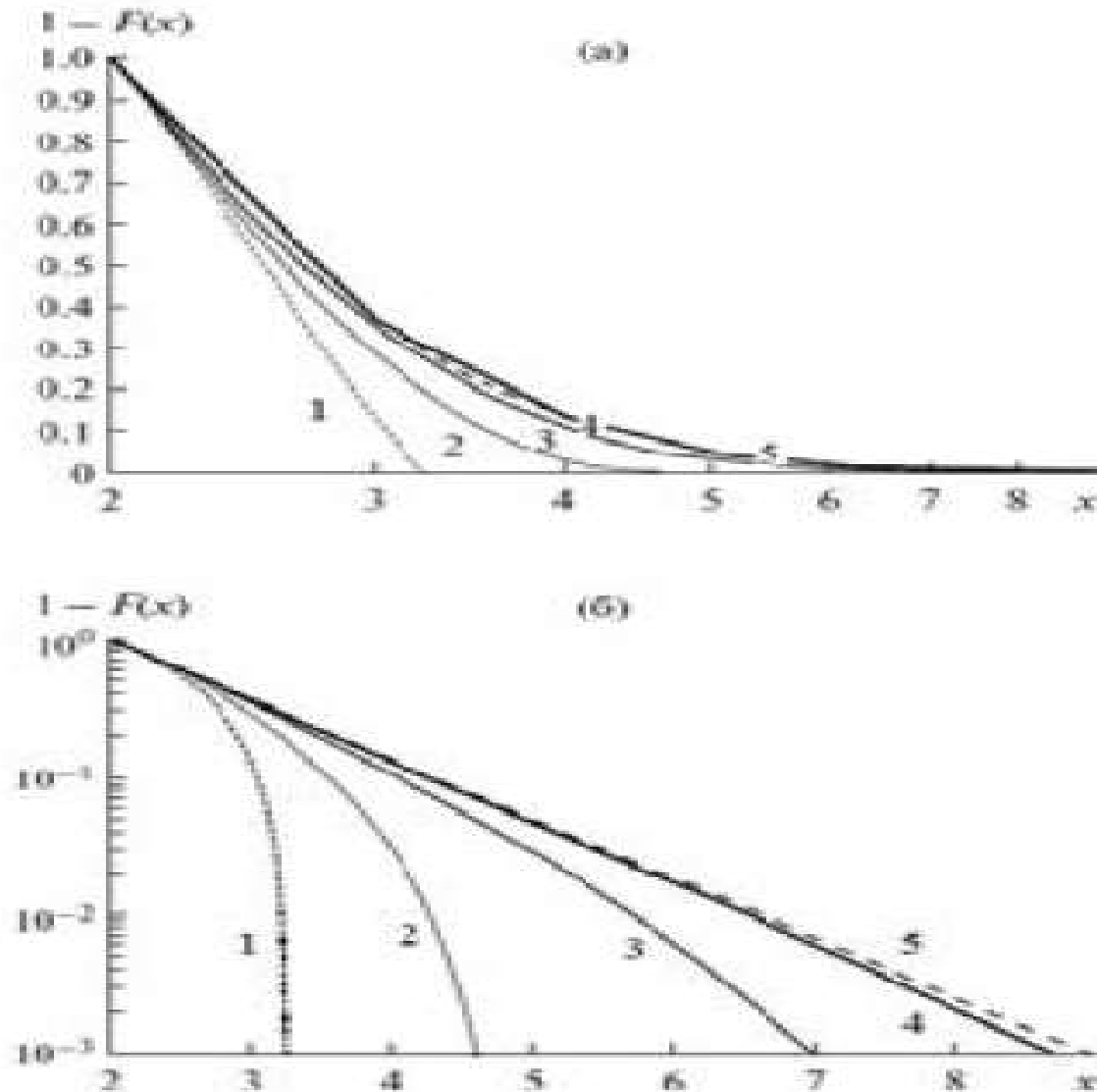


Рис. 1. Хвост GPD-распределения для разных значений параметра формы ξ : -0.8 (кривая 1); -0.35 (кривая 2); -0.1 (кривая 3); -0.01 (кривая 4) и 0 (кривая 5). Для всех графиков $h = 2$, $s = 1$. Для облегчения понимания поведения функции ось y дана в линейном (a) и логарифмическом масштабе (b).

Еще раз и подробнее про важный частный случай

Если $\xi < 0$, то $h \leq x \leq h - s/\xi$, т.е.,
$$\underline{M_{max}} = \underline{h - s/\xi} \quad (1)$$

Достаточно типична и важна ситуация:

- распределение ограничено: $h \leq x \leq h - s/\xi$
- максимально возможное событие $M_{max} = h - s/\xi$
существует,
- но оценивается с ОЧЕНЬ большой погрешностью $\xi \cong 0$.

Формула (1) «математически» поясняет неустойчивость значений M_{max} , что и отвечает необходимости регулярного пересмотра карт ОСР

Достоинства и ограничения метода $GEV-GPD$

Достоинства метода $GEV-GPD$:

1. Теоретическая статистическая обоснованность метода.
2. Анализ различных сейсмических каталогов дает разнообразие поведение хвоста распределения. Как правило, получаем оценки параметров, соответствующих конечным распределениям, что соответствует утверждению об ограниченности «всего сущего». (энергии землетрясений, любых иных реально оцениваемых величин, н.п., ущербов).

Ограничения метода $GEV-GPD$.

Использование метода предполагает получение достаточно точных оценок трех параметров GEV или GPD распределения. А для этого нужно не менее 30-50 событий в диапазоне, описывающим хвост распределения магнитуд. Это требование накладывает существенное ограничение на разрешающую способность метода.

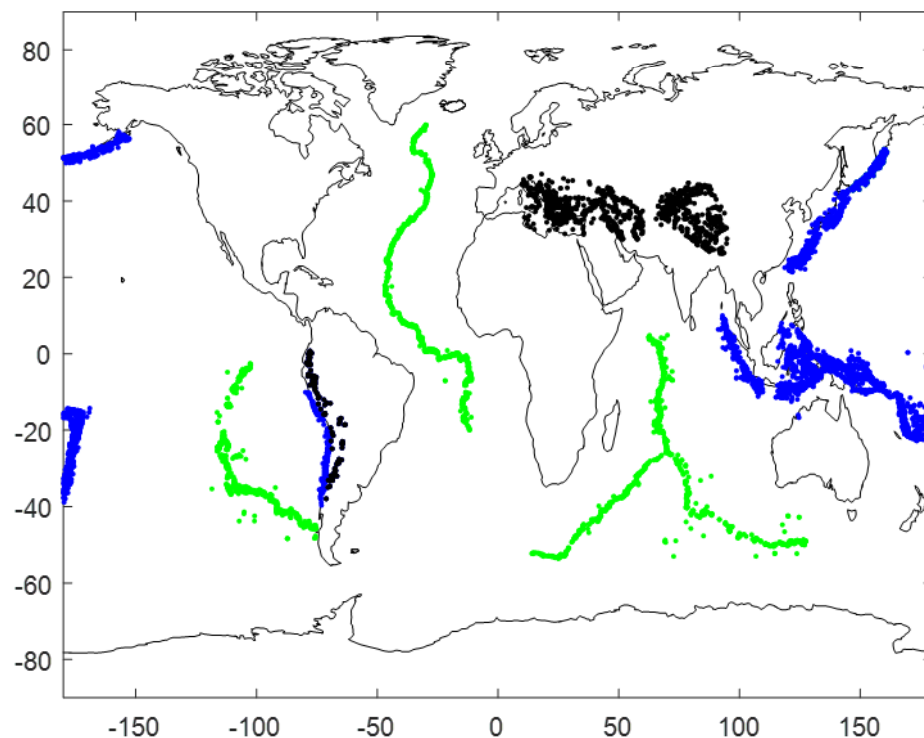
Результаты применения методов теории экстремальных значений:

1. Показано, что распределение редких сильнейших землетрясений не только (вполне ожидаемо) имеет конечное среднее значение, но и описывается конечным распределением ($\xi < 0$). Значения параметра формы обычно оказываются слабоизменчивыми - $\xi = -0.2 \pm 0.1$ (???)
2. Для Японии была показана значительная вероятность реализации землетрясений Mw9 и 9+ (Pisarenko et al., 2010), что подтвердилось возникновением мегаземлетрясения Тохоку (Mw9, 2011)
3. Значение $M_{\max}$ существует (распределение конечно), но при малых по абсолютной величине значениях ξ неустойчиво ($M_{\max} = h - s/\xi$). Это – математическое выражение необходимости регулярного пересмотра карт ОСР.
4. Вместо неустойчивых значений $M_{\max}$ лучше использовать более устойчивые значения квантилей $Q\tau(q)$ - максимальной магнитуды, реализующейся с вероятностью q (0.9, 0.95, 0.999, любое) за будущий интервал времени продолжительностью τ (10, 50, 500 лет, любое). *По сути - то же $M_{\max}$, но не абсолютное, а на заданный интервал времени.*

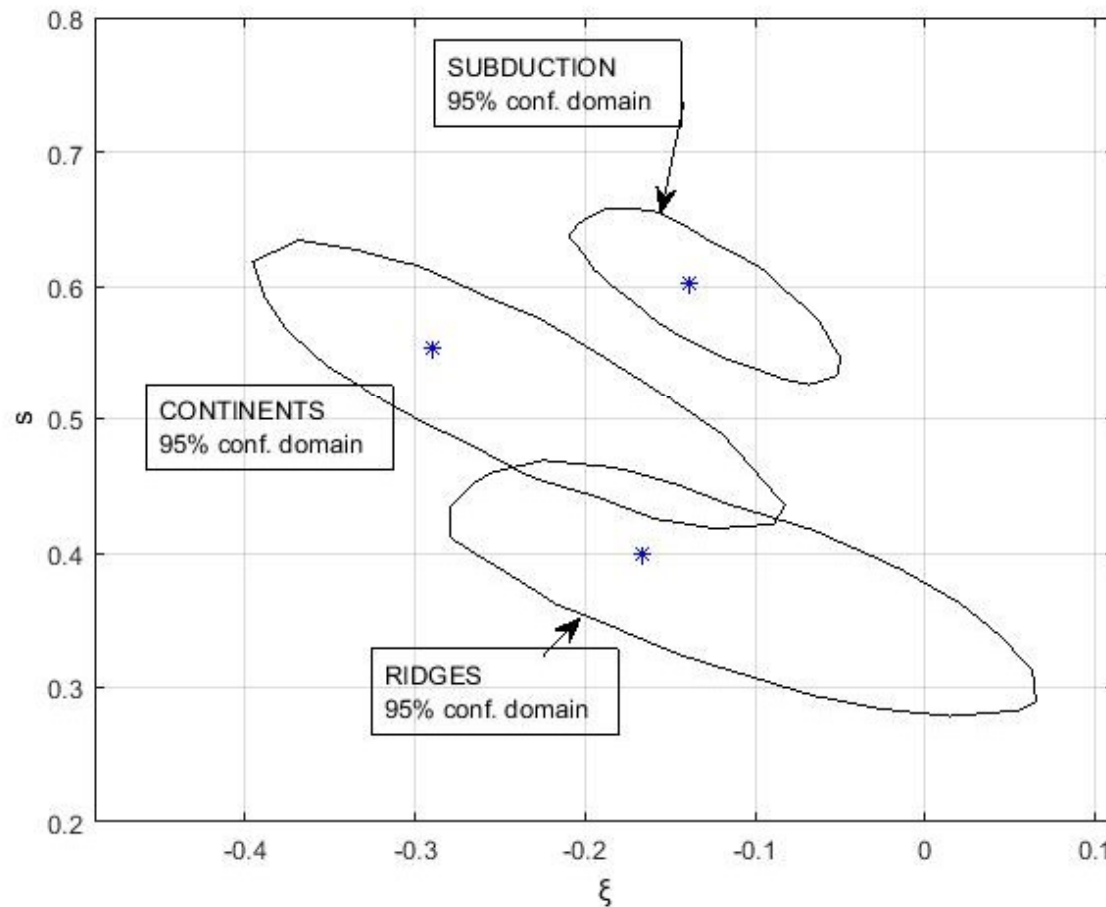
ПОПЫТКИ ПРЕОДОЛЕНИЯ ОГРАНИЧЕНИЙ

Для оценки неизвестных параметров при условии МАЛЫХ ВЫБОРОК, кроме метода максимального правдоподобия (ММП) нами был опробован метод квантилей (МК) и метод сейсмических моментов (ММ). Для GEV оказался точнее ММ, а для GPD – МК.

Для сравнения эффективности предлагаемого подхода нами были сформированы 3 виртуальные массивы данных на основе каталогов основных событий (с исключением афтершоков): **зоны субдукции,** **зоны внутриконтинентальной сейсмичности,** **зоны срединно-океанических хребтов.**



Доверительные 68%-ные области изменения (ξ , s) параметров **GPD** **распределения** для зон субдукции, внутриконтинентальной сейсмичности и срединно-океанических хребтов.



68% - области доверия для пар соответствующих параметров распределения. Оценки получены методом квантилей (МК)

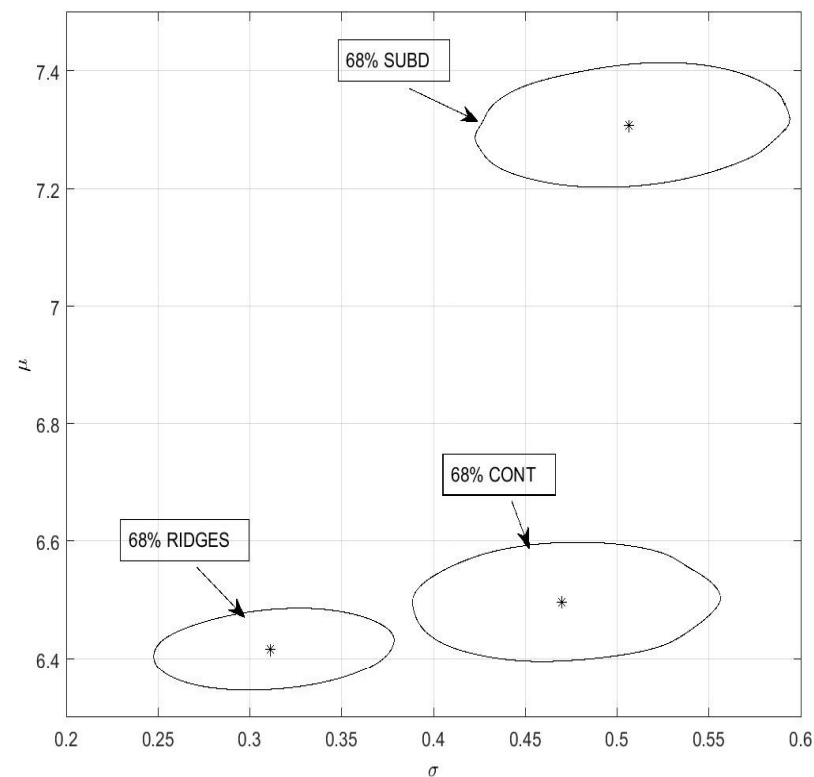
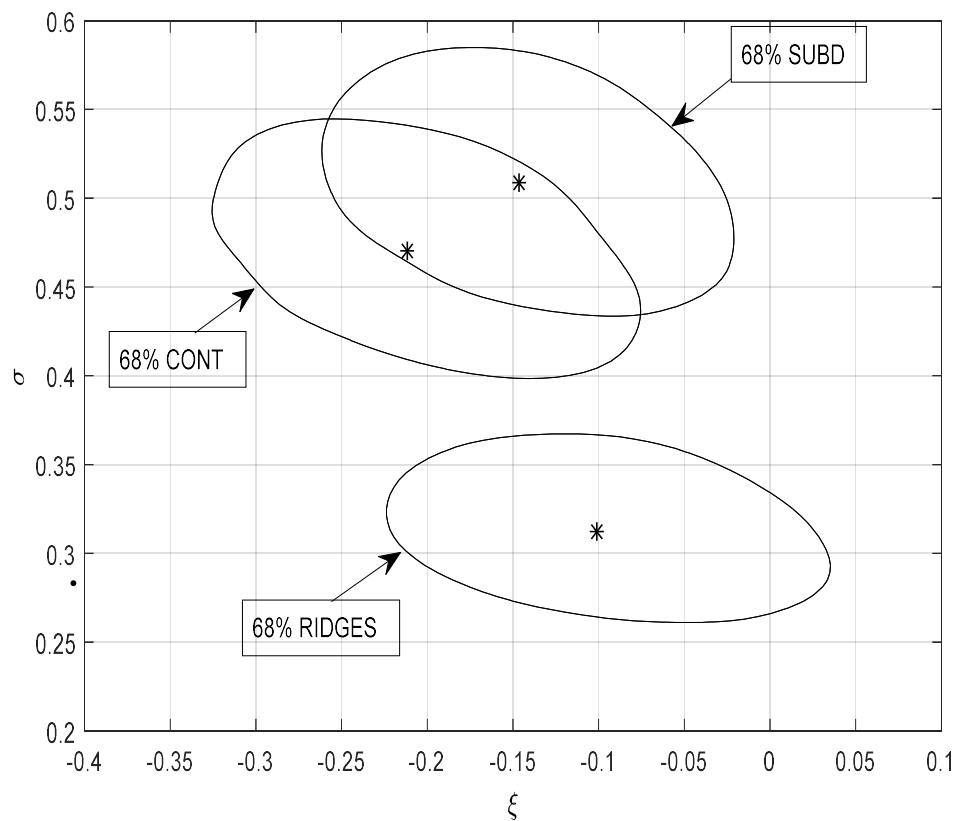
Видно различие характерных значений пар параметров для разных сеймотектонических ситуаций

Аналогично видно
различие для пар
значений параметров

ξ , σ , μ

для **GEV** распределения

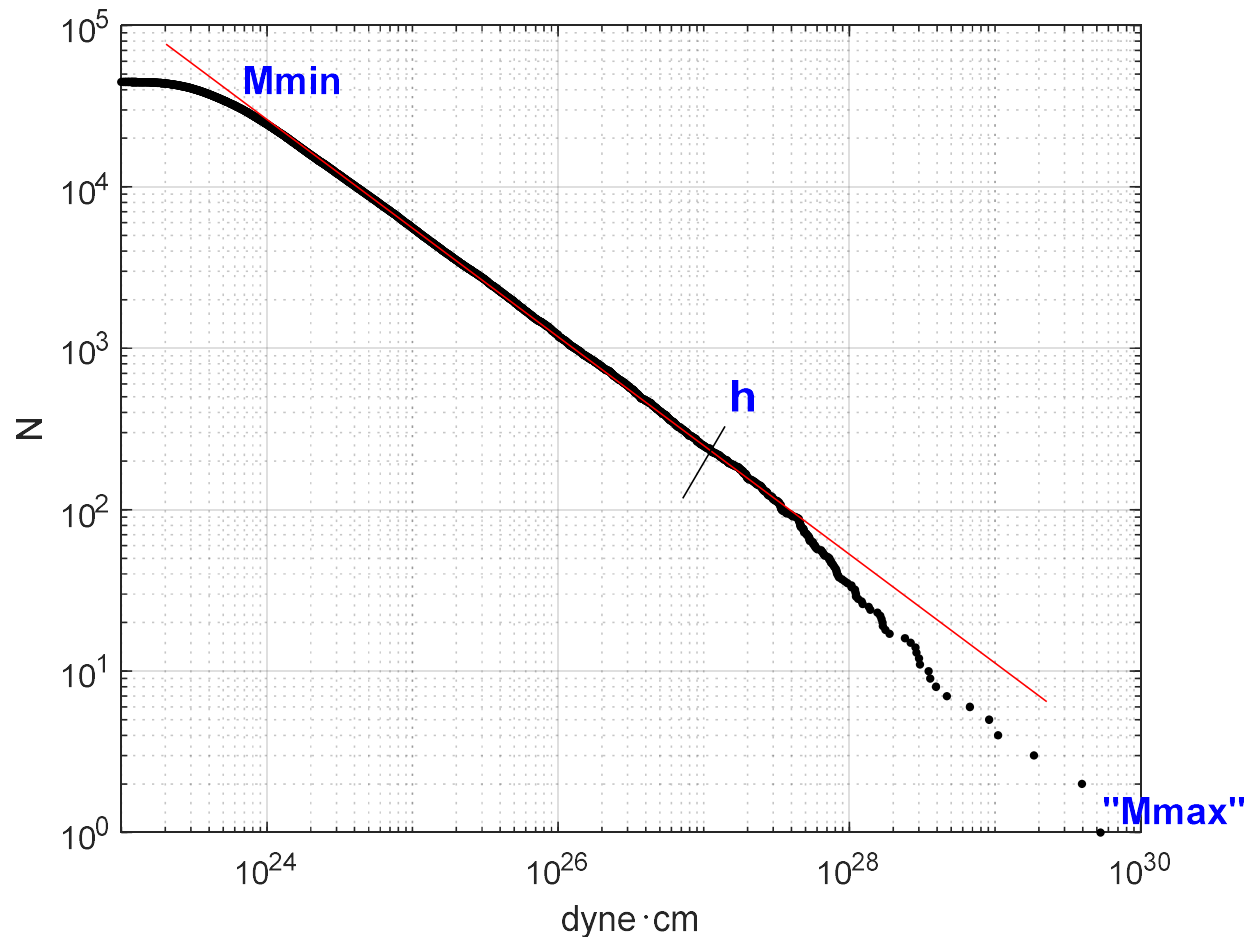
Оценки получены
методом моментов (ММ)



**Видно различие характерных
значений параметров для разных
сейсмотектонических ситуаций -**

2-х звенная модель закона повторяемости

Закон Г-Р по данным GCMT каталога за 1976 – 2016 годы.



Естественно описать такое эмпирическое распределение составной моделью: Диапазон $[M_{\min} ; h]$ стандартно описывается законом Гутенберга-Рихтера; диапазон $[h ; M_{\max}]$ – Обобщенным законом Парето (GPD). Значение параметра h не критично. В этой точке должна быть непрерывна плотность распределения и её производная.

**Формально поставим задачу следующим образом,
зададим две ветви распределения:**

$$F(m) = \begin{cases} C_1 \{1 - \exp[-b \cdot (m - m_0)]\}; & m_0 \leq m \leq h; \\ C_1 \{1 - \exp[-b \cdot (h - m_0)]\} + C_2 \{1 - [1 + (\xi/s) \cdot (m - h)]^{-1/\xi}\}; & h \leq m \leq M_{\max} = h - s/\xi, \xi < 0. \end{cases}$$

Модель содержит 5 неизвестных параметров. Порог h разделяет 2 ветви модели, Г-Р и GPD; b , m_0 , s , ξ – параметры; C_1 , C_2 – константы (они зависят от указанных параметров), которые должны обеспечить нормировку функции распределения $F(m)$ и её непрерывность:

$$C_1 = 1 / \{1 + bs \cdot \exp[-b(h - m_0)] - \exp[-b(h - m_0)]\},$$

$$C_2 = 1 - C_1 \{1 - \exp[-b(h - m_0)]\}.$$

Если обозначить

$$C_3 = C_1 [1 - \exp[-b \cdot (h - m_0)]],$$

То модель запишется в виде:

$$F(m) = \begin{cases} C_1 \{1 - \exp[-b \cdot (m - m_0)]\}; & m_0 \leq m \leq h; \\ C_3 + C_2 \{1 - [1 + (\xi/s) \cdot (m - h)]^{-1/\xi}\}; & h \leq m \leq M_{\max} = h - s/\xi, \xi < 0. \end{cases}$$

Важно, что использование данных по распределению событий умеренной силы и условие гладкого сопряжения закона Г-Р и GPD-распределения уменьшают на 2 число подлежащих оценке параметров модели

$$s = (1+\xi)/b$$

В окончательном виде модель имеет вид:

$$C_1 \{1 - \exp[-b \cdot (m - m_0)]\}; \quad m_0 \leq m \leq h;$$

$$F(m) = \begin{cases} C_3 + C_2 \{1 - [1 + (m - h) \cdot \xi \cdot b / (1 + \xi)]^{-1/\xi}\}; & h \leq m \leq M_{\max} = h - (1 + \xi) / (b \cdot \xi), \quad -1 < \xi < 0; \end{cases}$$

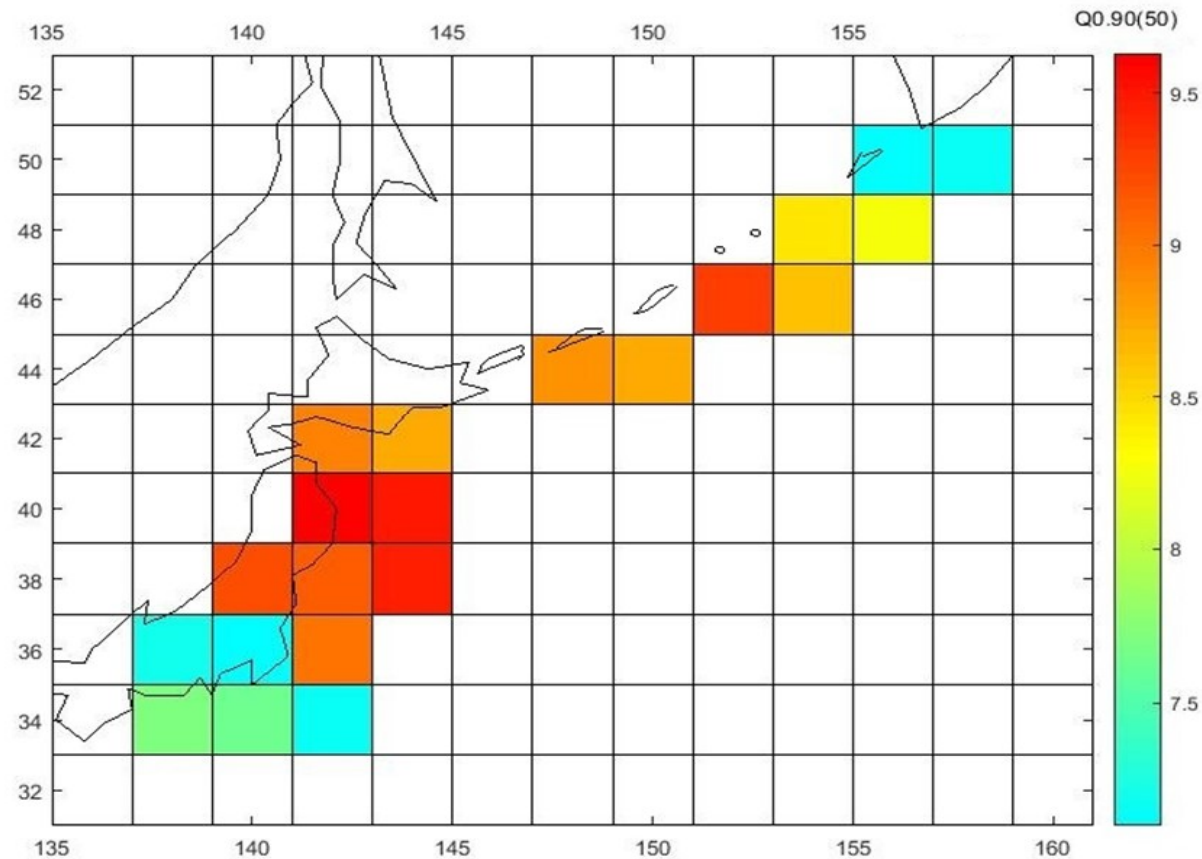
$$C_1 = 1 / (1 + \xi \exp[-b \cdot (h - m_0)]);$$

$$C_2 = (1 + \xi) \exp[-b \cdot (h - m_0)] / (1 + \xi \exp[-b \cdot (h - m_0)]);$$

$$C_3 = \{1 - \exp[-b \cdot (h - m_0)]\} / (1 + \xi \exp[-b \cdot (h - m_0)]);$$

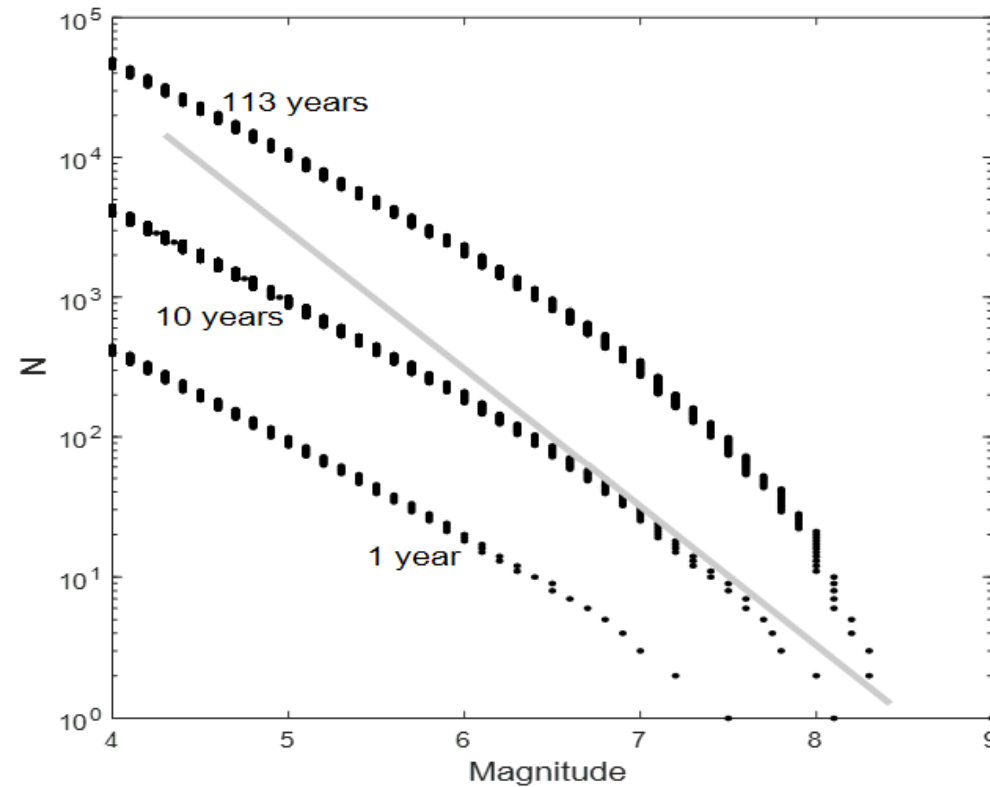
Модель содержит уже только 3 неизвестных параметра: h , b , ξ ; значение m_0 предполагается известным.

С использованием 2-х звенной модели – нового варианта закона повторяемости землетрясений – удастся в расчетах квантилей $Q_q(T)$ для $M_{\max}(T)$ выйти на уровень пространственного разрешения, сравнимого с масштабом карт ОСР



Пример: схема пространственного распределения значений квантиля $Q_{0.9}(50)$.

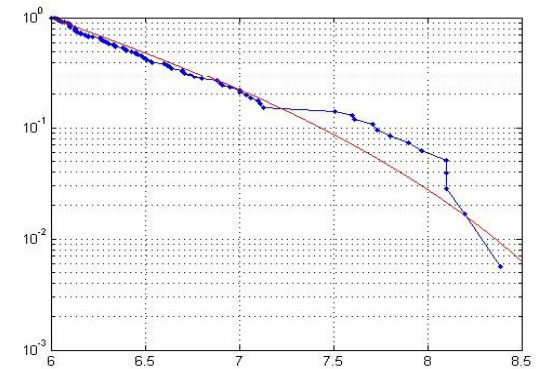
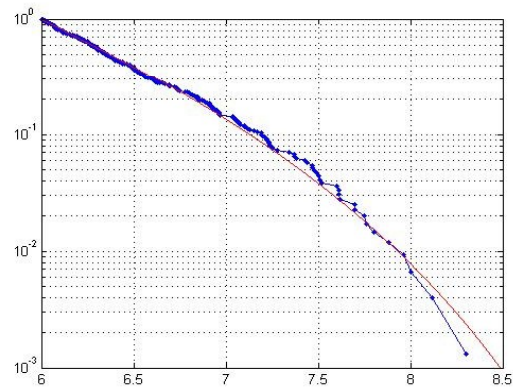
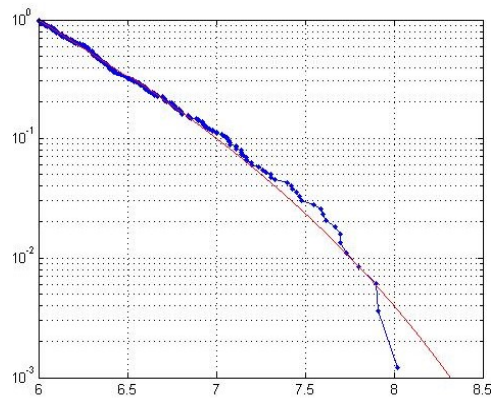
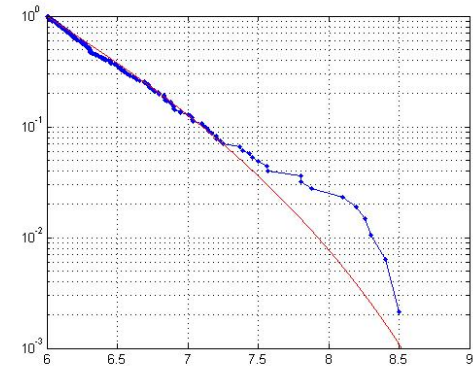
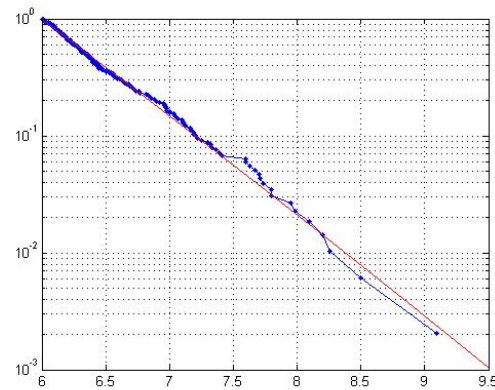
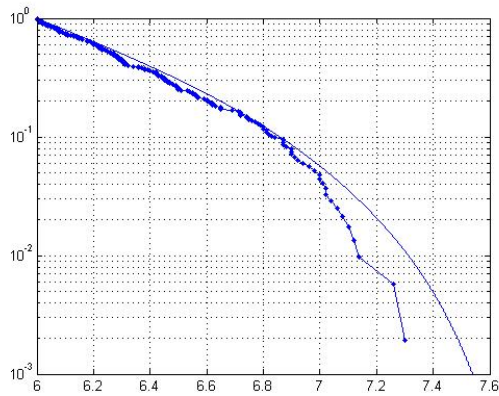
Снова 3 уровня описания.



Три уровня описания степенных распределений по данным каталога землетрясений ЛМА. Указаны интервалы времени наблюдения, точка касания прямой с наклоном $\beta = -1$ со средним графиком характеризует величину M_{ch} .

УРОВЕНЬ 3+?

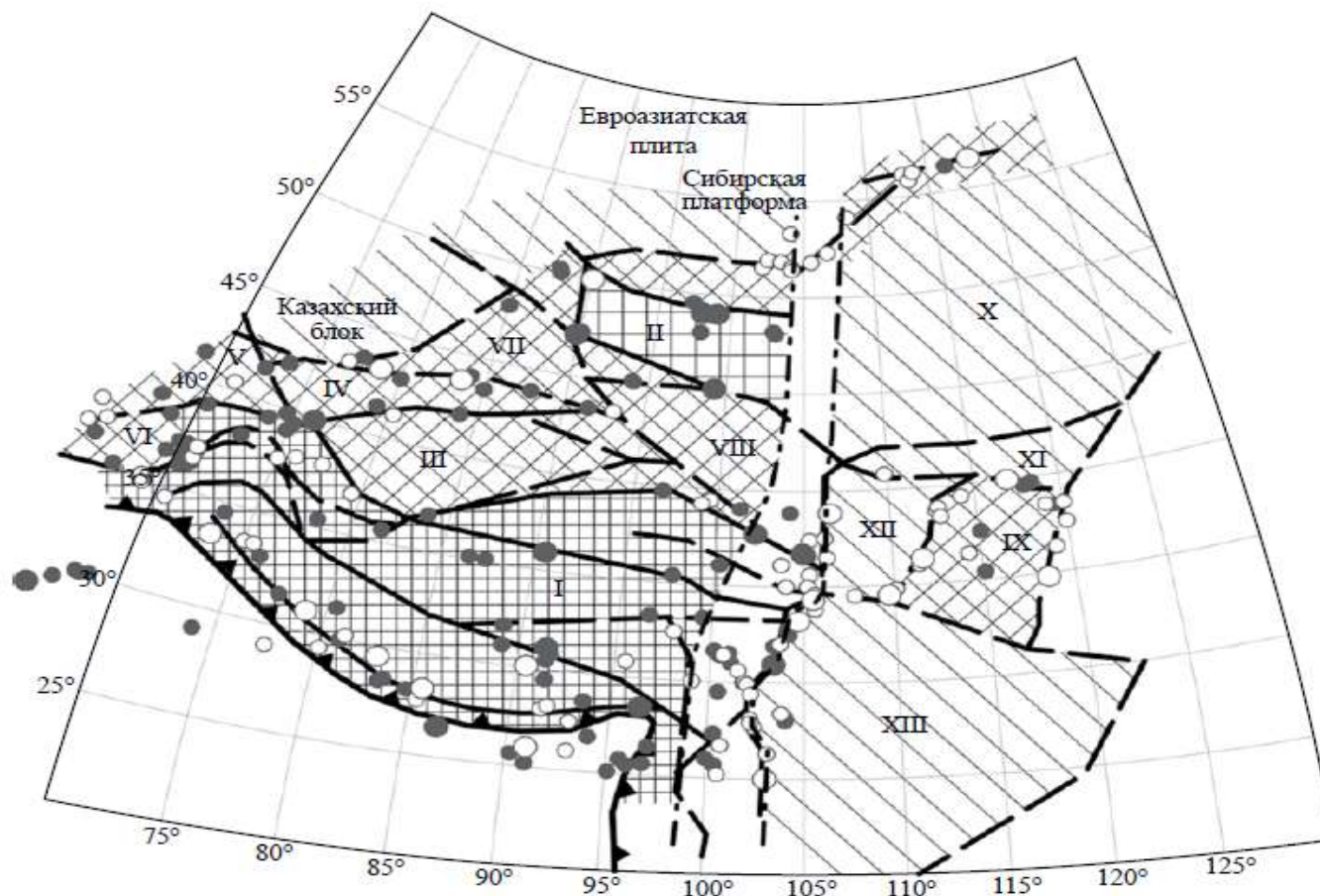
Варианты поведения хвоста



Каталог ISC-GEM - наиболее длинный и единообразный (Mw).

Районы Срединно-Атлантического хребта, Японии, Курильских островов, Новых Гебрид, Филиппин и Перу.

Некая неожиданность:



Данные по эмпирическим графикам повторяемости сильных землетрясений в окрестности региональных разломов Центральной Азии (по С.И.Шерману)

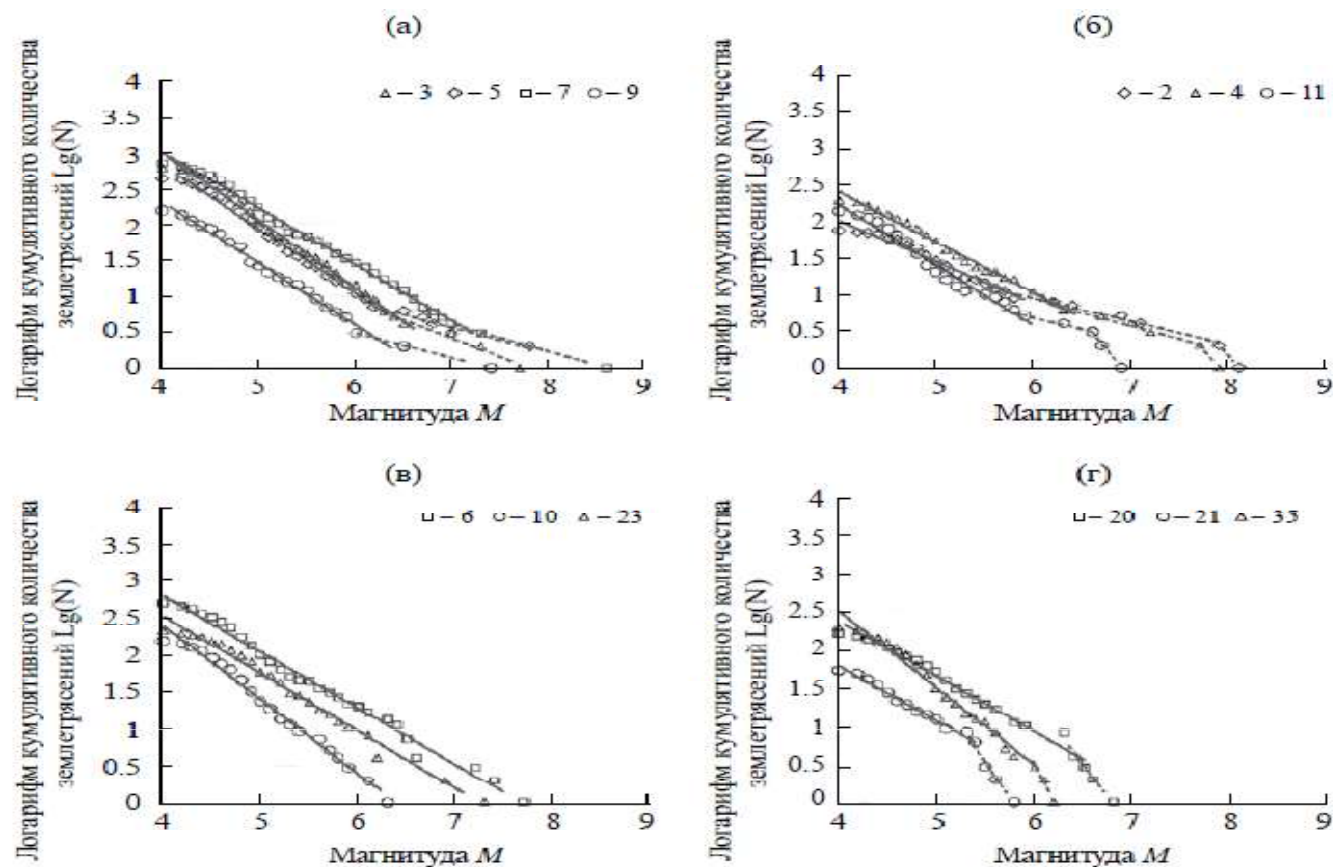


Рис. 3. Типы графиков повторяемости землетрясений по форме правых окончаний графиков для сейсмоактивных разломов Центральной Азии. а–г — типы графиков повторяемости землетрясений. Цифрами на рисунках указаны номера разломов, пронумерованных на рис. 4 (в кружках).

Таблица . **Типизация поведения хвоста графиков повторяемости землетрясений**

Тип графика повторяемости	Реализуется ли режим характеристических землетрясений	Достаточен ли столетний интервал для явного выявления загиба вниз графика повторяемости
1	да	нет
2	да	да
3	нет	нет
4	нет	да

Описание в рамках модели мультипликативного каскада

Модель сейсмического режима как последовательности большого числа эпизодов лавинообразной релаксации, случайным образом реализующихся на множестве метастабильных подсистем. Модель имитирует все надежно установленные эмпирические закономерности сейсмического режима.

Пусть начавшееся произвольное событие (землетрясение) величиной (энергией) X_i в некоторый момент времени t_i с вероятностью p может продолжить свое развитие или прекратиться с вероятностью $(1 - p)$. В случае прекращения процесса на i -ом шаге величина события полагается равной достигнутому этом шаге значению X_i . В случае продолжения процесса релаксации метастабильной подсистемы положим, что величина события X_{i+1} в следующий момент времени $t_i + 1$ возрастает до значения

$$X_{i+1} = r \times X_i,$$

$$\beta = \lg(1/p) / \lg(r),$$

$$X_{i+1} = r \times X_i + r_1 \times (X_i / A_1)^2 - r_2 \times (X_i / A_2)^3$$

Добавление нелинейных членов в кинетическое уравнение мультипликативного каскада позволяет описывать различные типы поведения хвоста распределения

Представлен ряд подходов к описанию степенных распределений, применимых в зависимости от объёма статистических данных (времени наблюдения). Каждый последующий уровень описания демонстрирует ограниченность предыдущего.

Уровень 1, нет событий близких к максимально возможным. Тенденция нелинейного роста величин эффекта со временем (длиной ряда); показана некорректность интерпретации как указания на нестационарность (например, как бы быстрый рост ущерба от природных катастроф)

Уровень 2, несколько событий близких к максимально возможным. Предложена робастная характеристика M_c – типичного максимального события и способ ее оценки. M_c характеризует самые опасные события. Устойчивая характеристика для сейсмотектонического районирования, оценок тенденций временных изменений ущербов от природных катастроф, иные применения.

Уровень 3, несколько десятков событий близких к максимальному. Точное описание поведения хвоста распределения ..

Уровень 3+ ... Ой!!!! А может это ещё и не хвост!!

Спасибо за внимание

Полученные данные для Курильских островов

широта долгота центра ячейки	число толчков $M_{\max}$	$\lg(-\xi)$ в скобках $(-\xi)$	b	h	KD	pKD	$Q_{0.5}(50)$	$Q_{0.9}(50)$
Все Курилы	307 (184;123) 8.30	-1.403 ± 0.366 (-0.0395)	1.832 ± 0.152	5.84 ± 0.62	0.83	0.16	8.35 ± 0.30	9.12 ± 0.27
44 148	100 (79;21) 8.25	-0.8814 ± 0.306 (-0.1314)	1.364 ± 0.205	6.31 ± 0.43	0.54	0.72	8.18 ± 0.38	8.84 ± 0.60
44 150	116 (96;20) 8.25	-0.8184 ± 0.294 (-0.1519)	1.417 ± 0.210	6.54 ± 0.43	0.60	0.58	8.15 ± 0.33	8.73 ± 0.55
46 152	118 (98;20) 8.30	-9.852 ± 0.034 ($1.407 \cdot 10^{-10}$)	1.814 ± 0.167	6.26 ± 0.21	0.46	0.95	8.26 ± 0.36	9.30 ± 0.36
46 154	99 (68;31) 8.30	-9.657 ± 0.030 ($2.204 \cdot 10^{-10}$)	2.125 ± 0.222	5.83 ± 0.18	0.83	0.24	7.75 ± 0.24	8.63 ± 0.34
48 154	118 (20;98) 8.30	-1.877 ± 0.459 (-0.0133)	2.198 ± 0.264	5.40 ± 0.34	0.64	0.49	7.36 ± 0.29	8.41 ± 0.48
48 156	108 (20;88) 8.30	-1.623 ± 0.419 (-0.0238)	2.175 ± 0.272	5.40 ± 0.32	0.67	0.37	7.54 ± 0.29	8.27 ± 0.46
50 156	99 (59;40) 7.00	-0.5450 ± 0.157 (-0.2851)	1.611 ± 0.335	5.83 ± 0.28	0.79	0.16	6.91 ± 0.14	7.11 ± 0.22
50 158	99 (62;37) 7.00	-0.5380 ± 0.141 (-0.2897)	1.578 ± 0.316	5.89 ± 0.29	0.78	0.15	6.96 ± 0.16	7.16 ± 0.22