

钱小仕, 王福昌, 盛书中. 2013. 基于广义帕累托分布的地震震级分布尾部特征分析. 地震学报, 35(3): 341—350.  
Qian Xiaoshi, Wang Fuchang, Sheng Shuzhong. 2013. Characterization of tail distribution of earthquake magnitudes via generalized Pareto distribution. *Acta Seismologica Sinica*, 35(3): 341—350.

# 基于广义帕累托分布的地震 震级分布尾部特征分析<sup>\*</sup>

钱小仕<sup>\*</sup> 王福昌 盛书中

(中国河北三河 065201 防灾科技学院)

**摘要** 极值理论在地震危险性分析中有着重要应用, 发震震级超过某一阈值的超出量分布可以近似为广义帕累托分布. 基于广义帕累托分布给出了若干地震活动性参数的估计公式, 包括强震震级分布、地震复发周期和重现水平、期望重现震级、地震危险性概率和潜在震级上限等; 以云南地区震级资料为基础数据, 讨论了阈值选取、模型拟合诊断和参数估计; 在此基础上计算了该地区的地震活动性参数. 结果表明, 广义帕累托分布较好地刻画了强震震级分布, 通过超阈值(POT)模型计算的复发周期与实际复发间隔统计基本一致, 高分位数估计在一定阈值范围内表现稳定, 为工程抗震中潜在震级上限的确定提供了一种途径.

**关键词** 广义帕累托分布 POT 模型 期望重现震级 云南地区

doi:10.3969/j.issn.0253-3782.2013.03.006 中图分类号: P315.7 文献标志码: A

## Characterization of tail distribution of earthquake magnitudes via generalized Pareto distribution

Qian Xiaoshi<sup>\*</sup> Wang Fuchang Sheng Shuzhong

(Institute of Disaster Prevention, Sanhe City, Hebei Province 065201, China)

**Abstract:** Extreme value theory in seismic risk analysis has important application, and the excess of earthquake magnitude distribution over a threshold can be approximated by generalized Pareto distribution. Based on the generalized Pareto distribution, we developed several estimation formulas of seismic activity parameters, including strong earthquake magnitude distribution, earthquake recurrence period and return level, expected recurrence magnitude, probability of seismic risk and maximum earthquake magnitude; then, based on historical seismic data in Yunnan region, we discussed how to choose the threshold value, model fitting diagnosis and parameter estimation. Finally we calculated seismicity parameters in the region. The results show that generalized Pareto distribution characterized strong earthquake magnitude distribution quite satisfactorily, the recurrence periods by POT (peaks over threshold method) model agree with

<sup>\*</sup> 基金项目 中央高校基本科研业务费专项资金(ZY20120205, ZY20110101)资助.

收稿日期 2012-02-23 收到初稿, 2012-06-06 决定采用修改稿.

<sup>\*</sup> 通讯作者 e-mail: qxs\_prob@qq.com

the actual recurrence interval statistics, and the high quantile is stable within a specific threshold range. So the generalized Pareto distribution is a possible approach to determining the potential upper limit earthquake magnitude in engineering seismology.

**Key words:** generalized Pareto distribution; peaks over threshold method; expected recurrence magnitude; Yunnan region

## 引言

极值统计在地震预报、地震危险性分析、工程地震安全性评价及地震区划等诸多地震学领域有着广泛应用(Epstein, Lomnitz, 1966; Yegulalp, Kuo, 1974; 陈培善, 林邦慧, 1973; 高孟潭, 贾素娟, 1988; 陈虹, 黄忠贤, 1995; 贾素娟, 鄢家全, 1996). 极值分布有 3 种类型: Gumbel(I 型)、Frechet(II 型)、Weibull(III 型), 它们可以统一为广义极值分布. 钱小仕等(2012)基于广义极值分布给出了若干地震预报分析的公式与方法, 并以台湾地区为震例进行了研究, 拓展了极值理论的应用范围. 但是利用极值分布作统计分析的不足是仅能利用所研究时间窗中单个数据点信息. 例如在估计最大地震震级分布时, 仅使用了年或给定时间区间上最大震级数据, 严重制约着获得稳健分布参数估计的数据规模, 导致参数估计的统计稳定性降低, 概率地震危险性预测的不确定性增大. Balkema 和 de Haan (1974)以及 Pickands (1975)指出, 对于充分大的阈值, 随机变量超过阈值的超出量的极限分布为广义帕累托分布(generalized Pareto distribution, 简称为 GPD), 为研究随机变量尾部特征提供了统计理论依据. 近年来, 国外已有学者将广义帕累托分布引入地震研究, Pisarenko 和 Sornette (2003)利用广义帕累托分布分析了哈佛地震目录中 18 个地震区的浅层地震地震矩分布; Huyse 等(2010)利用太平洋地震工程研究的地面峰值加速度(PGA)数据和基于 Abrahamson-Silva 衰减模型的 PGA 残差数据, 比较了对数正态分布与广义帕累托分布的拟合优度.

本文基于广义帕累托分布给出了若干地震活动性参数的估计方法, 包括强震震级分布、地震平均复发周期、重现水平和期望重现震级、地震危险性概率和潜在震级上限估计等; 利用超阈值(peaks over threshold, 简称为 POT)模型对云南地区历史地震资料进行分析, 估计该地区强震震级分布并与 G-R 经验关系作比较; 在此基础上计算了各震级档复发时间和未来几年发震危险性.

## 1 广义帕累托分布与 POT 模型

设  $X_1, X_2, \dots, X_n$  为地震震级随机变量列, 假设它们相互独立且服从同一分布  $F(x)$ , 记其最大震级为  $M_n = \max(X_1, X_2, \dots, X_n)$ , 若存在  $\{a_n > 0, b_n \in R\}$  和非退化分布函数  $H(x)$ , 使

$$\lim_{n \rightarrow \infty} P\left(\frac{M_n - b_n}{a_n} \leq x\right) = H(x) \quad (1)$$

则称  $H(x)$  为极值分布, 而称分布  $F(x)$  属于极值分布  $H(x)$  的最大值吸引场. Fisher 和 Tippet(1928)获得了极值分布的 3 种形式, 它们也可以统一为广义极值分布

$$H_{\xi}(x; \mu, \sigma, \xi) = \exp\left\{-\left(1 + \xi \frac{x - \mu}{\sigma}\right)^{-\frac{1}{\xi}}\right\} \quad (2)$$

式中,  $\mu$  为位置参数,  $\sigma$  为尺度参数 ( $\sigma > 0$ ),  $\xi$  为形状参数,  $x$  满足  $1 + \xi \frac{x - \mu}{\sigma} \geq 0$ . 当  $\xi > 0$  时对应极值 II 型 Frechet 分布; 当  $\xi < 0$  时对应极值 III 型 Weibull 分布; 当  $\xi = 0$  时对应极值 I 型 Gumbel 分布. 若记  $x^+ = \sup\{x : 0 < F(x) < 1\}$  为分布函数的右端点, 则称

$$F_u(x) = P(X - u \leq x | X > u) = \frac{F(x + u) - F(u)}{1 - F(u)} \quad x \geq 0 \quad (3)$$

为地震震级  $X$  的超过阈值  $u$  的超出量的分布函数, 简称超出量分布. Balkema 和 de Haan (1974) 以及 Pickands (1975) 指出, 当震级分布  $F$  属于极值分布  $H$  的最大值吸引场时有

$$\lim_{u \rightarrow x^+} \sup_{0 \leq x \leq x^+ - u} |F_u(x) - G(x; \bar{\sigma}, \xi)| = 0 \quad (4)$$

式中,  $G(x; \bar{\sigma}, \xi) = 1 - \left(1 + \xi \frac{x}{\bar{\sigma}}\right)^{-1/\xi}$ ,  $x \geq 0$ ,  $1 + \xi \frac{x}{\bar{\sigma}} \geq 0$ ,  $\bar{\sigma} = \sigma + \xi(u - \mu)$  为两参数广义帕累托分布, 即若最大震级  $M_n$  近似服从广义极值分布  $H(x; \mu, \sigma, \xi)$ , 则对充分大的震级阈值  $u$ , 震级超出量  $X - u$  近似服从广义帕累托分布  $G(x; \bar{\sigma}, \xi)$ , 且两者具有相同的形状参数  $\xi$ .  $\xi$  也称为极值指数, 其值反映了分布尾部的收敛性质,  $\xi$  越大则尾部越厚, 尾分布收敛速度越缓慢. 基于广义帕累托分布对超过某一充分大的阈值的所有观测数据进行极值统计建模, 渐近地刻画分布的尾部特征模型称为 POT 模型.

## 2 POT 模型的应用

钱小仕等 (2012) 基于广义极值分布推导出若干地震危险性预测预报的公式与方法, 本节基于广义帕累托分布给出一些相关的地震活动性参数的估计. 记  $\bar{F}(x) = 1 - F(x)$  表示震级分布  $F(x)$  的尾分布;  $N_u = \sum_{i=1}^n I_{\{X_i > u\}}$  表示  $X_1, X_2, \dots, X_n$  中超过阈值  $u$  的次数, 其中  $I_{\{X_i > u\}}$  为示性函数, 当  $X_i > u$  时为 1, 否则为 0;  $\Delta_n(u) = \{i : X_i > u\}$  表示超过  $u$  的观测值的下标集;  $e(u) = E(X - u | X > u)$  为当震级  $X$  超过  $u$  时的平均超出量函数, 则易计算出基于广义帕累托分布  $G(x; \bar{\sigma}, \xi)$  的平均超出量函数  $e(u) = \frac{\bar{\sigma}}{1 - \xi} + \frac{\xi}{1 - \xi} u$  是  $u$  的线性函数;  $e_n(u) = \frac{1}{N_u} \sum_{i \in \Delta_n(u)} (X_i - u)$ ,  $u > 0$  为样本平均超出量函数.

### 2.1 强震震级分布估计

由式 (3) 可得  $\bar{F}(u + x) = \bar{F}_u(x) \bar{F}(u)$ . 其中  $\bar{F}_u(x)$  为超出量分布的尾, 可用广义帕累托分布近似;  $\bar{F}(u)$  为震级超过阈值  $u$  的概率, 用样本量  $N_u/n$  估计, 于是得分布尾部估计

$$\hat{\bar{F}}(u + x) = \hat{\bar{F}}_u(x) \hat{\bar{F}}(u) = \frac{N_u}{n} \left(1 + \hat{\xi} \frac{x}{\hat{\sigma}}\right)^{-1/\hat{\xi}} \quad (5)$$

或者分布函数的估计

$$\hat{F}(y) = 1 - \frac{N_u}{N} \left(1 + \hat{\xi} \frac{y - u}{\hat{\sigma}'}\right)^{-1/\hat{\xi}} = 1 - \left(1 + \hat{\xi} \frac{y - h}{\hat{\sigma}'}\right)^{-1/\hat{\xi}} \quad y \geq u \quad (6)$$

为参数  $h, \xi, \sigma'$  的广义帕累托分布, 其中  $h = u + \frac{\hat{\sigma}}{\hat{\xi}} \left[\left(\frac{N_u}{n}\right)^{\hat{\xi}} - 1\right]$ ,  $\sigma' = \hat{\sigma} \left(\frac{N_u}{n}\right)^{\hat{\xi}}$ .

根据式(3),若超过某一震级阈值  $u$  的震级分布为广义帕累托分布  $F(x|u) = 1 - \left(1 + \xi \frac{x-u}{\bar{\sigma}}\right)^{-1/\xi}$ , 则对另一阈值  $u_1 > u$ , 在震级超过新阈值  $u_1$  的条件下震级分布

$$F(x|u_1) = \frac{F(x|u) - F(u_1|u)}{1 - F(u_1|u)} = 1 - \left(1 + \xi \frac{x-u_1}{\sigma'}\right)^{-1/\xi} \quad \sigma' = \bar{\sigma} + \xi(u_1 - u) \quad (7)$$

仍为广义帕累托分布,且形状参数保持不变.因此,基于超阈值震级分布的 POT 模型具有理论上的阈值分布稳定性.

## 2.2 平均复发周期估计

假设地震数据为日观测数据,一年按 365 天计算,若震级分布为广义帕累托分布  $G(x; u, \bar{\sigma}, \xi)$ , 则发生震级为  $x$  的地震平均复发周期为

$$T(x) = 1/365[1 - G(x; u, \bar{\sigma}, \xi)] \quad (8)$$

## 2.3 重现水平估计

对给定概率  $p$ , 根据式(5)不超过概率  $p$  的发震震级(即  $p$ -分位数)为

$$\hat{x}_p = u + \frac{\hat{\bar{\sigma}}}{\hat{\xi}} \left\{ \left[ \frac{n}{N_u} (1-p) \right]^{-\hat{\xi}} - 1 \right\} \quad (9)$$

若假设地震数据为日观测数据,则  $T$  年重现期的重现水平相当于  $p = 1 - \frac{1}{365T}$  的分位数  $\hat{x}_p$ .

重现水平估计只是估计了发震超过某一震级的可能性,不能确定当超过该震级地震发生时可能的地震震级大小.为解决这一问题,我们提出期望重现震级概念,以期发震震级预测提供一种新的途径.称  $E(X|X > x_p)$  为期望重现震级,表示当发震震级超过特定水平  $x_p$  下该震级平均大小,记为  $\bar{x}_p$ , 则

$$\bar{x}_p = E(X|X > x_p) = \frac{1}{F(x_p)} \int_{x_p}^{x^+} x dF(x) = x_p + E(X - x_p|X > x_p) = x_p + e(x_p) \quad (10)$$

为震级分布的  $p$ -分位数与平均超出量函数之和.当震级分布为广义帕累托分布  $G(x; u, \bar{\sigma}, \xi)$  时可以获得期望重现震级的估计

$$\hat{\bar{x}}_p = \hat{x}_p + \frac{\hat{\bar{\sigma}}}{1 - \hat{\xi}} + \frac{\hat{\xi}}{1 - \hat{\xi}} (\hat{x}_p - u) = \frac{\hat{x}_p}{1 - \hat{\xi}} + \frac{\hat{\bar{\sigma}} - \hat{\xi}u}{1 - \hat{\xi}} \quad (11)$$

## 2.4 地震危险性概率

未来  $T$  年发震震级超过  $x$  的概率

$$P(\min\{m: X_m \geq x\} \leq 365T) = 1 - [G(x; u, \bar{\sigma}, \xi)]^{365T} \quad (12)$$

## 2.5 潜在震级上限估计

当形状参数  $\xi < 0$  时,对应分布有有限右端点,即震级分布有上限

$$\hat{x}^+ = u - \frac{\hat{\bar{\sigma}}}{\hat{\xi}} \quad (13)$$

# 3 实例分析

## 3.1 资料选取

2005 年 10 月 31 日前的数据引自国家自然科学基金委员会“中国西部环境与生态科学

数据中心”(http://westdc.westgis.ac.cn)<sup>①</sup>. 该目录包括公元前 2300 年—公元 2005 年间发生在全国各地的各级地震, 共 33 万余条目录, 每条包括发震时间、震中经纬度、震源深度和震级. 2005 年 11 月 1 日—2010 年 12 月 31 日数据引自中国地震信息网(http://www.csi.ac.cn)提供的测震目录<sup>②</sup>. 地震震级采用面波震级  $M_s$ , 在  $M_s$  缺失部分利用面波震级  $M_s$  与近震震级  $M_L$  之间的转换关系式(汪素云等, 2010):  $M_s = 1.13M_L - 1.08$ , 将近震震级转换为  $M_s$ . 从中选取云南地区( $21^\circ$ — $29^\circ\text{N}$ ,  $97^\circ$ — $106^\circ\text{E}$ )1900—2010 年地震目录(包含四川省与中缅交界地区的部分地震), 并进行余震删除. 余震删除方法采用 C-S 余震时空窗(陈凌等, 1998), 经余震删除后共有 74159 个震级数据, 其中  $M_s \geq 5.0$  地震共 539 次,  $M_s \geq 6.0$  地震 122 次,  $M_s \geq 7.0$  地震 17 次.

由于 POT 模型主要针对震级分布的尾部, 即强震震级特征的分析, 因此需考虑强震目录的完整性. 图 1 给出了 1900—2010 年云南地区  $M_s \geq 5.0$  地震的震中分布和  $M-t$  图. 由图 1b 可见, 大致在 1930 年前后, 地震记载开始丰富起来. 相关研究表明, 20 世纪 30 年代以后该地区 5 级以上地震记录才较为完整(黄玮琼等, 1994; 皇甫岗, 李中华, 2010; 苏有锦, 李中华, 2011). 结合本文阈值的选择, 我们选取 1930 年作为 5 级以上地震目录基本

完整的起始年代. 共获得 74102 个震级数据, 其中  $M_s \geq 5.0$  地震共 487 次,  $M_s \geq 6.0$  地震 107 次,  $M_s \geq 7.0$  地震 15 次. 为了比较广义帕累托分布与基于 G-R 关系的指数分布的拟合效果, 我们选择震级数据较好地符合 G-R 关系式的最小震级作为起始震级. 图 2 给出了 1930—2010 年的 G-R 关系拟合. 从

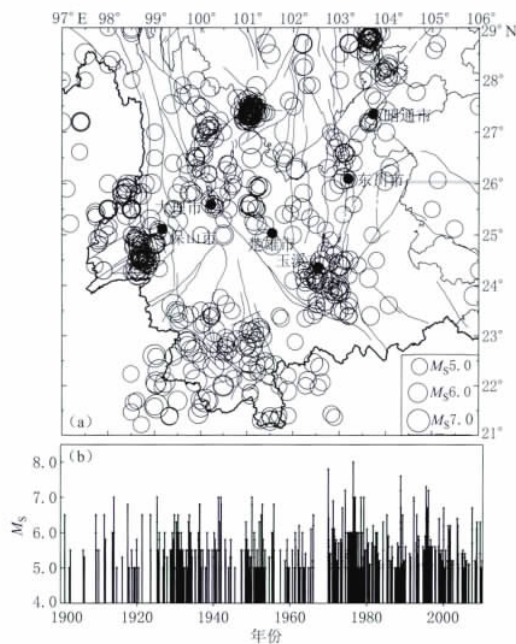


图 1 1900—2010 年云南地区  $M_s \geq 5.0$  地震震中分布图(a)和  $M-t$  图(b)

Fig. 1 Epicenters (a) and  $M-t$  relation (b) for  $M_s \geq 5.0$  earthquakes in Yunnan region from 1900 to 2010

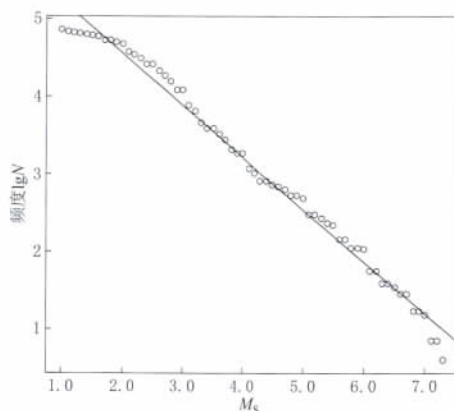


图 2 1930—2010 年云南地区地震的震级频度关系

Fig. 2 Magnitude-frequency relation of earthquakes in Yunnan region from 1930 to 2010

① 马瑾, 等. 2005. 中国地震目录. 中国地震局地质研究所. [2011-11-09]http://westdc.westgis.ac.cn/data/236a607a-245f-4444-ba4c-f7ccfd53271.

② 中国地震台网中心. 2012. 中国地震台网(CSN)地震目录. 中国地震台网中心. [2011-11-15] http://www.csnedmc.ac.cn/newweb/catalog\_direct\_link.htm.

图中可以看出, 云南地区  $M_s \geq 2.0$  的地震较好地符合 G-R 关系式,  $M_s < 2.0$  地震明显偏离上述关系, 因此本文震例计算中, 震级下限取为  $M_s \geq 2.0$ .

### 3.2 阈值选取与模型诊断

用广义帕累托分布拟合超出量时, 首要的问题是阈值  $u$  的选取.  $u$  值太大将只有少数几个超出量, 数据量太少会导致估计量方差过大;  $u$  值太小将使得超出量分布与广义帕累托分布出现较大偏差, 估计量成为有偏估计, 误差增大. 因此在  $u$  的选取中需要综合考虑

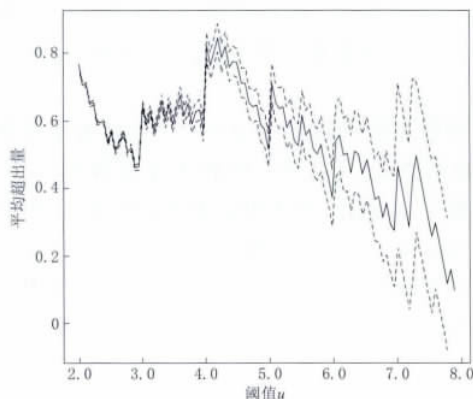


图3 云南地区样本平均超出量函数图

Fig. 3 Sample mean excess function for Yunnan region

图4给出了广义帕累托分布参数估计在不同阈值选取下的变化图, 其中修正的尺度参数为  $\sigma^* = \bar{\sigma} - \xi u$ . 对于适当选定的阈值, 相应的超出量服从广义帕累托分布, 则形状参数和修正的尺度参数的估计应保持不变. 从图4可以看出, 当阈值约在区间(4.0, 5.0)时, 参数估计比较稳定. 根据极值理论, 为保证极值模型的准确性, 应在参数估计值与所选阈值附近呈平稳性的基础上尽量选取较大的阈值, 所以我们取云南地区震级尾部阈值为  $u=5.0$ . 此时, 超阈值数据达320

个, 约占总量(48456)的0.66%, 不超过1%, 因此可以较好地满足尾部极值建模的条件.

阈值选定后, 需对广义帕累托分布拟合状况进行诊断. 图5a为P-P图, 图5b为Q-Q图, 图5c为重现水平图, 图5d为尾部密度曲线的估计和直方图. P-P图是根据变量的累积概率与指定分布的累积概率之间的关系绘制的图形. Q-Q图是用变量数据分布的分位数与指定分布的分位数之间的关系绘制的图形. 如所检验的数据符合指定的分布, 则P-P图和Q-Q图上的点应近似在一条直线上. 重现水平图为重现期的对数与重现水平之间关系绘制

偏和方差. 实际操作中, 阈值一般依据样本平均超出量函数是否呈直线形来选取. 对某个阈值  $u_0$  之后的  $u$ , 当样本平均超出量函数在一条直线附近波动时, 点  $u_0$  通常就可以作为阈值. 图3给出了云南地区震级数据的样本平均超出量函数图及相应的95%置信区间. 由图3可见, 阈值约在区间(4.0, 5.0)以上的部分近似呈直线形, 可在此区间内选取阈值, 但是选取仍存在一定程度的任意性. 为解决图3中阈值选取的问题, 可以利用不同阈值选取下参数  $\xi$  和  $\bar{\sigma}$  估计值变化的稳定状况进行阈值的进一步判断.

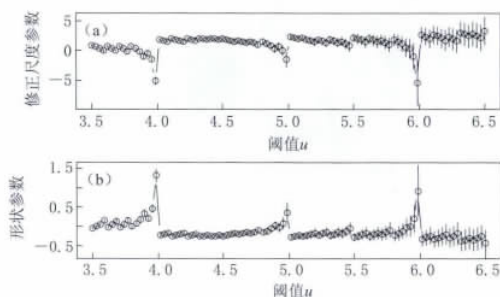


图4 广义帕累托分布参数估计随阈值变化图

(a) 修正的尺度参数; (b) 形状参数

Fig. 4 Parameter estimation of generalized Pareto distribution changing with threshold

(a) Modified scale parameters;  
(b) Shape parameters

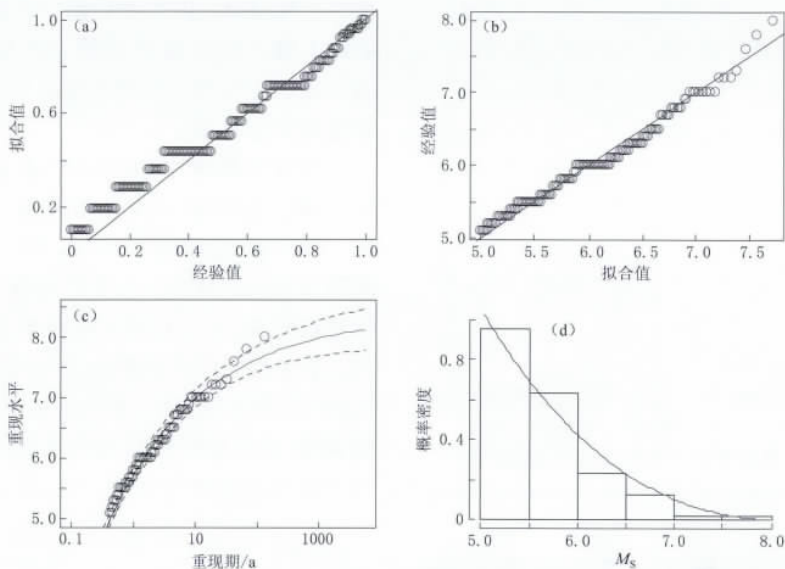


图 5 广义帕累托分布拟合诊断图

(a) P-P 图; (b) Q-Q 图; (c) 重现水平图; (d) 直方图与密度曲线估计图

Fig. 5 Diagnostic graphs of generalized Pareto distribution fitting

(a) P-P graph; (b) Q-Q graph; (c) Recurrence level graph;

(d) Histogram and density curve estimation

的图形, 如所检验数据符合广义帕累托分布, 则样本数据应落在给定分布分位数估计置信区间内. 从图 5 中可以看出, 各散点数据基本紧密围绕各参考线分布, 表明拟合状态良好, 利用 POT 模型分析云南地区震级分布特征较适宜.

### 3.3 地震活动性参数估计

在取定阈值后, 广义帕累托分布的参数可用矩估计法、极大似然法、概率矩法、 $L$ 矩法等进行估计(史道济, 2006). 本文利用 R 软件(<http://www.r-project.org/>)中 ismev 工具包, 采用极大似然法估计广义帕累托分布的参数, 得  $(\hat{\xi}, \hat{\sigma}) = (-0.287, 0.955)$ , 相应的对数似然函数值为 212.95, 参数估计的协方差矩阵近似为  $\begin{pmatrix} 0.001 & -0.002 \\ -0.002 & 0.004 \end{pmatrix}$ ,  $\hat{\xi}$  和  $\hat{\sigma}$  的标准误差分别为 0.0329 和 0.0603, 于是可得  $\xi$  的 95% 置信区间为  $[-0.351, -0.222]$ . 从  $\xi$  的 95% 置信区间可以看出,  $\xi$  为负值, 所以广义帕累托分布有有限的右端点, 即云南地区震级有上限. 进一步, 通过式(13)获得震级上限的估计  $M_{\max} = 8.3$ . 通过  $\hat{\xi}$  和  $\hat{\sigma}$  的置信区间, 计算出最大震级的区间估计为  $[7.4, 9.8]$ , 对比云南地区所给地震目录资料范围内的历史最大震级(8.0 级), 说明广义帕累托分布不仅可以作为确定潜在震源区最大震级上限的一种有效方法, 同时还给出了震级上限不确定性的度量. 此外, 在工程抗震中, 潜在震级上限的确定总是与特定建筑结构的使用年限相联系, 即给定年限内可能出现的最大地震. 根据前节所述, 这相当于要估计分布的高分位数. 图 6 给出了百年重现期所对应的 99.997% 分位数随阈值变化的估计. 可以看出, 阈值选择在  $M_s 4.0-6.0$  左右重现水平估计表现出较高的稳定性, 基本在  $M_s 7.6$  左右波动, 表明高分位数估计在一定范围内受阈值选取的影响较小, 而且在实际应用中还可以根据建筑结构的抗震要求选择相应的高分位数



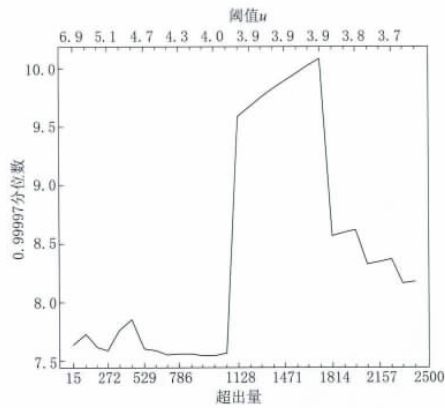


图 6 百年重现水平随阈值变化的估计

Fig. 6 Estimation of recurrence magnitudes in 100 years under threshold change

信区间为 $[5.82, 6.01]$ . 类似地, 表 1 给出了各重现期的地震复发震级估计和 95% 的置信区间以及期望重现震级的估计. 表 2 给出了云南地区各强震复发周期预测和未来 1 a, 5 a 的发震概率. 从表 1 的点估计和表 2 的复发周期可以看出, 基于广义帕累托分布估计的复发震级和复发周期, 与苏有锦和李中华(2011)的结果基本一致, 各震级复发周期相对于苏有锦和李中华(2011)平均复发间隔稍高, 但均处于  $0.5\sigma$  方差线内. 表 2 的概率估计表明, 云南地区未来一段时间内 5—6 级地震发生危险性较高.

作为上限估计. 因此利用高分位数估计潜在震级上限不仅具备使用的灵活性而且还具有一定的稳定性, 同时还可以给出估计不确定性的定量表示.

苏有锦和李中华(2011)研究了云南地区 6 级以上强震时间分布特征, 统计得到  $M_s \geq 6.0$  平均间隔 1.27 a,  $M_s \geq 6.5$  平均间隔 2.75 a,  $M_s \geq 7.0$  平均间隔 8.21 a, 实际地震有 60% 以上均发生在平均间隔时间内. 本文利用广义帕累托分布估计相应复发时间间隔的震级水平. 图 7 是重现期为 1.27 a 的震级重现水平轮廓似然函数图, 复发震级的最大似然估计为  $M_s = 5.91$ , 95% 的置

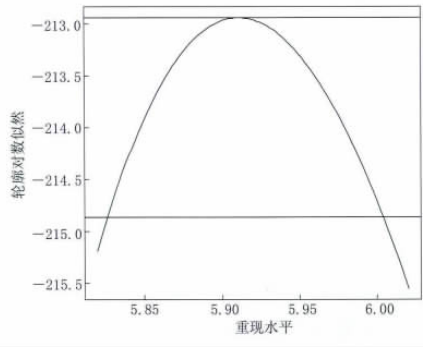


图 7 重现期为 1.27 a 的重现水平轮廓对数似然估计

Fig. 7 The profile log-likelihood for recurrence magnitude at 1.27 year recurrence period

表 1 云南地区发震震级重现水平估计

Table1 Estimation of recurrence magnitude for Yunnan region

重现期/a	1.27	2.75	8.21	30	50	100
点估计	5.91	6.39	6.91	7.35	7.49	7.64
区间估计	[5.82, 6.01]	[6.28, 6.51]	[6.79, 7.06]	[7.21, 7.57]	[7.35, 7.73]	[7.49, 7.93]
期望重现震级	6.45	6.83	7.23	7.57	7.69	7.79

表 2 云南地区地震危险性分析

Table 2 Seismic risks in Yunnan region

震级 $M_s$	5.0	5.5	6.0	6.5	7.0	8.0
复发周期 $T/a$	0.41	0.73	1.44	3.35	10.19	1317.77
1 a 发震概率	0.91	0.75	0.50	0.26	0.09	0.00
5 a 发震概率	1.0	1.0	0.97	0.78	0.39	0.00



在参数估计的基础上, 由式(6)可得强震震级分布函数的估计为

$$F(x) = 1 - 0.0066[1 - 0.3(x - 5)]^{3.484} \quad x \geq 5 \quad (14)$$

Gutenberg 和 Richter (1956) 提出的震级-频度经验公式, 已成为地震学中最基础的定律之一, 在地震活动性研究与地震危险性预测的实践中被广泛引用. 在地震活动平稳性假设下, 相当于地震震级分布为指数分布(陈培善, 林邦慧, 1973). 图 8 给出了云南地区  $M_s \geq 5.0$  地震资料的经验分布(圆圈)、基于广义帕累托分布的震级分布拟合(实线)和基于 G-R 关系式的指数分布拟合(虚线)比较. 从图 8 可以看出, 在尾部区域对给定的尾概率水平  $1 - F(x)$ , G-R 经验关系在 5—6 级段低估而在 7 级以上高估发震震级或危险性, 而广义帕累托

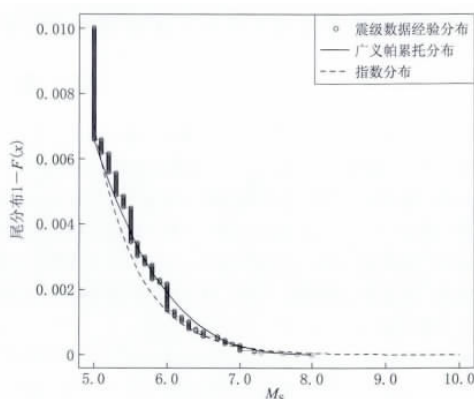


图 8 云南地区强震震级分布的广义帕累托分布拟合与指数分布拟合比较

Fig. 8 Comparison between generalized Pareto distribution fitting and exponential distribution fitting for strong earthquake magnitudes in Yunnan region

分布的估计总体上与实际数据拟合良好, 而且给出了震级应有上限的一种自然描述, 无需求像指数分布那样通过人为的右截断(Cosentino *et al.*, 1977)来描述震级分布.

## 4 讨论与结论

一般说来, 基于独立同分布随机变量的极值建模有两种基本统计模型: 一种是经典的极值分析区组最大值法——仅利用研究时间窗中最大值数据信息; 另一种是超阈值分析 POT 方法——利用超过给定阈值的所有数据信息. 本文基于广义帕累托分布给出了地震活动性分析的几种公式, 包括强震震级分布的估计、重现期与重现水平的预测、给定时间内地震危险性概率和震级上限的确定等, 利用 POT 模型对云南地区历史地震资料进行分析. 结果表明, 云南地区震级资料用广义帕累托分布拟合较好, 在参数估计的基础上, 获得云南地区强震震级分布为  $F(x) = 1 - 0.0066[1 - 0.3(x - 5)]^{3.484}$ ,  $x \geq 5$ . 通过与 G-R 关系式的对比发现, 基于 G-R 关系的指数分布在强震级端可能或低或高地估计了地震危险性, 而基于 POT 模型的广义帕累托分布则与实际资料吻合良好, 而且 POT 模型具备严格的数学理论依据.

最大地震震级是描述区域地震活动性的一个基本参数, 地震学家们已发展了多种最大地震震级的确定方法(陈颢, 陈凌, 1999), 对各种震级上限确定方法的不确定性也有研究(冉洪流, 2009). 本文指出若震级分布可以用广义帕累托分布拟合, 则最大震级及其不确定性可以通过参数估计方法得到, 为确定震级上限增添了一种新的途径.

本文基于广义帕累托分布的云南强震复发时间的估计, 与苏有锦和李中华(2011)通过对实际地震资料发震间隔统计得到的结果基本一致, 表明利用 POT 模型估计地震复发时间具有一定的实际意义. 对云南地区未来 1 a, 5 a 强震发生概率的估计可以发现, 未来该

地区中强震发生的危险性偏高,应引起重视.

重现水平估计只是给出了超过某一震级档地震发生的可能,不能确定实际可能发生的地震大小.本文提出了期望重现震级概念,计算在超过某震级地震发生时发震震级的平均大小.关于期望重现震级在实际应用中的效果,还有待进一步研究.

### 参 考 文 献

- 陈虹,黄忠贤. 1995. 应用混合极值理论及最大似然法估计中国大陆地震危险性[J]. 地震学报, 17(2): 264–269.
- 陈凌,刘杰,陈颢,陈龙生. 1998. 地震活动性分析中余震的删除[J]. 地球物理学报, 41(增刊): 244–252.
- 陈培善,林邦慧. 1973. 极值理论在中长期地震预报中的应用[J]. 地球物理学报, 16(1): 6–24.
- 陈颢,陈凌. 1999. 地震危险性分析中最大地震震级的确定[J]. 地球物理学报, 42(3): 351–357.
- 高孟潭,贾素娟. 1988. 极值理论在工程地震中的应用[J]. 地震学报, 10(3): 317–326.
- 黄玮琼,李文香,曹学峰. 1994. 中国大陆地震资料完整性研究之二: 分区地震资料基本完整的起始年分布图[J]. 地震学报, 16(4): 423–432.
- 皇甫岗,李中华. 2010. 20 世纪云南地区地震记录完全性评价[J]. 地震研究, 33(1): 1–6.
- 贾素娟,郗家全. 1996. 利用历史地震影响烈度的统计特性进行地震区划[J]. 地震研究, 19(3): 277–285.
- 钱小仕,王福昌,曹桂荣,任晴晴. 2012. 广义极值分布在地震危险性分析中的应用[J]. 地震研究, 35(1): 73–78.
- 冉洪流. 2009. 潜在震源区震级上限不确定性研究[J]. 地震学报, 31(4): 396–402.
- 史道济. 2006. 实用极值统计方法[M]. 天津: 天津科学技术出版社: 83–85.
- 苏有锦,李中华. 2011. 云南地区 6 级以上强震时间分布特征及其概率预测模型研究[J]. 地震研究, 34(1): 1–7.
- 汪素云,高阿甲,冯义钧,和锐. 2010. 中国地震目录间的对比及标准化[J]. 地震, 30(2): 38–45.
- Balkema A A, de Haan L. 1974. Residual life time at great age[J]. *Ann Probab*, 2(5): 792–804.
- Cosentino P, Ficara V, Luzio D. 1977. Truncated exponential frequency-magnitude relationship in the earthquake statistics[J]. *Bull Seism Soc Amer*, 67(6): 1615–1623.
- Epstein B, Lomnitz C. 1966. A model for the occurrence of the largest earthquakes[J]. *Nature*, 211(5052): 954–956.
- Fisher R, Tippett L H. 1928. Limiting forms of the frequency distributions of the largest or smallest member of a sample[J]. *Proc Camb Phil Soc*, 24(2): 180–190.
- Gutenberg B, Richter C. 1956. Earthquake magnitude, intensity, energy, and acceleration, part II [J]. *Bull Seism Soc Amer*, 46(2): 105–145.
- Huyse L, Chen R, Stamatakis J A. 2010. Application of generalized pareto distribution to constrain uncertainty in peak ground accelerations[J]. *Bull Seism Soc Amer*, 100(1): 87–101.
- Pickands J. 1975. Statistical inference using extreme order statistics[J]. *Ann Stat*, 3(1): 119–131.
- Pisarenko V F, Sornette D. 2003. Characterization of frequency of extreme earthquake events by the generalized pareto distribution[J]. *Pure Appl Geophys*, 160(12): 2343–2364.
- Yegulalp T M, Kuo J A. 1974. Statistical prediction of the occurrence of maximum magnitude earthquakes[J]. *Bull Seism Soc Amer*, 64(2): 393–414.

## 作者简介



工作. 甘肃省地震学会会员.

杨兴悦 甘肃省地震局工程师. 2012 年中国地震局兰州地震研究所固体地球物理专业毕业, 获理学硕士学位. 曾从事台站地震监测和青藏高原强震数值模拟方面的研究. 现从事地震监测预测研究方面的工作. 甘肃省地震学会会员.



理论及其应用研究.

钱小仕 防灾科技学院讲师. 2003 年淮北煤炭师范学院数学系数学与应用数学专业毕业, 获理学学士学位; 2009 年南京师范大学数学科学学院概率论与数理统计专业毕业, 获理学硕士学位. 现从事极值



方 炜 陕西省地震局高级工程师. 2002 年兰州地震研究所固体地球物理专业毕业, 获理学硕士学位. 主要从事地球内部结构、地震预测与地震电磁学方面的研究工作. 中国地球物理学会会员.



会会员.

谢俊举 中国地震局地球物理研究所副研究员. 2010 年中国地震局地球物理研究所毕业, 获理学博士学位. 主要从事强震地面运动、近断层地震动对结构的影响以及近场强震动模拟研究. 中国地震学



事数字地震学和综合预报方法的研究工作.

唐兰兰 新疆维吾尔自治区地震局工程师. 2003 年新疆大学建筑工程学院测绘专业毕业, 获工学学士学位; 2007 年新疆大学资源与环境工程学院地球探测与信息技术专业毕业, 获工学硕士学位. 主要从事



李 琳 哈尔滨学院工学院土木系讲师. 中国地震局工程力学研究所在读博士研究生. 2007 年沈阳建筑大学土木学院固体力学专业毕业, 获工学硕士学位. 现主要从事地震地面运动记录的工程选取工作.