

广义极值分布在地震危险性分析中的应用*

钱小仕, 王福昌, 曹桂荣, 任晴晴

(防灾科技学院, 河北 三河 065201)

摘要: 利用广义极值分布研究地震最大发震震级的规律, 给出了广义极值分布下地震危险性分析的相关公式与方法, 并对台湾地区历史地震资料进行极值统计分析, 发现最大震级超过7级的地震理论发震次数与实际发震次数完全一致, 在此基础上预测了未来几年发震的危险性。

关键词: 极值理论; 广义极值分布; 地震危险性

中图分类号: P315.7

文献标识码: A

文章编号: 1000-0666(2012)01-0073-06

0 引言

极值理论是概率论中的一个重要分支, 也是统计预报地震的重要方法之一。Nordquist (1945) 最早将 Gumbel 极值分布应用于地震统计, 其后, 国内外许多学者应用和发展了 Gumbel 理论, 使其在地震预报、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领域都有着重要应用 (Epstein, Lomnitz, 1966; Yegulalp, Kuo, 1974; Huillet, Raynaud, 1999; 高孟潭, 贾素娟, 1988; 贾素娟, 鄢家全, 1996)。考虑到震级应有上限, 陈培善和林邦慧 (1973) 给出了极值分布的修正式, 并对我国 14 个地震带作了内符检验和预测。陈虹等 (1995, 1996) 通过将极值分布线性组合成混合极值分布研究了大陆和东南沿海各震区的地震危险性, 一定程度上弥补了在震级较大区间内 I 型极值相对观测值偏高, 而 III 型相对观测值偏低的缺点。由于地震活动的复杂性, 各地区地震震级分布存在很大的差异, 在我国各多震区, 各种极值类型分布有着不同的适用性 (秦卫平, 1983; 张卫东等, 2005)。因此在实际应用过程中, 对模型的选择就显得非常重要, 如果模型分布选择不当, 可能会带来严重偏差。广义极值分布可以统一 3 种类型极值渐近分布 (De Haan, Ferreira, 2006), 本文基于广义极值分布讨论地震危险性分析, 给出了几种与地震预报有关的公式

和预报方法, 并以台湾地区震级资料为基础, 分析了未来不同年份地震危险性。

1 极值理论与方法

设 X_1, X_2, \dots, X_n 是独立同分布的随机变量列, 有共同的分布函数 $F(x)$, 记为 $M_n = \max(X_1, X_2, \dots, X_n)$, 若存在 $\{a_n > 0, b_n \in R\}$ 和非退化分布函数 $G(x)$, 使

$$\frac{M_n - b_n}{a_n} \xrightarrow{b} G(n \rightarrow \infty). \quad (1)$$

则称 G 为极值分布, 并称 F 为属于 G 的吸引场, 记作 $F \in D(G)$ 。Fisher 和 Tippett (1928) 证明了极值分布的 3 大类型定理, 指出 G 必属于下列 3 种类型之一:

$$\begin{aligned} \text{I. Gumbel } \Lambda(x) &= e^{-e^{-\frac{x-\mu}{\sigma}}}, x \in R, \sigma > 0; \\ \text{II. Frechet } \Phi_\alpha(x) &= \begin{cases} 0, & x \leq \mu, \\ e^{-\left(\frac{x-\mu}{\sigma}\right)^{-\alpha}}, & x > \mu, \end{cases} \sigma > 0, \alpha < 0; \\ \text{III. Weibull } \Psi_\alpha(x) &= \begin{cases} e^{-\left(\frac{x-\mu}{\sigma}\right)^{-\alpha}}, & x > \mu, \\ 1, & x \leq \mu, \end{cases} \sigma < 0, \alpha < 0. \end{aligned} \quad (2)$$

其中, μ 为位置参数, σ 为尺度参数, α 为极值指标。

如果知道地震震级分布的类型, 并且能验证其是否满足极值分布吸引场条件 (史道济,

* 收稿日期: 2011-04-25.

基金项目: 中央高校基本科研业务费专项资金 (ZY20120205) 和河北省自然科学基金 (A2011408006) 联合资助。

2006), 就可以确定相关问题的极值分布属于什么类型, 但是通常只能获得观测数据, 而其具有何种分布是难以确定的。另外, 即使知道分布类型, 也很难验证其是否满足最大值吸引场条件。对于我国各个多震地区, 以往应用极值理论来分析地震危险性时往往单独采用 3 种分布中的一种, 而忽视满足其分布的条件。

广义极值分布包含了以上 3 种分布, 是极值分布的统一形式, 其分布函数为

$$G_{\xi}(x)=\exp\left\{-\left(1+\xi\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\xi}}\right\},\tag{3}$$

其中 $1+\xi\frac{x-\mu}{\sigma}>0$, $\sigma>0$, $\xi\in\mathbb{R}$, $\mu\in\mathbb{R}$ 。 μ 为位置参数, σ 为尺度参数, ξ 为形状参数。当 $\xi>0$ 时, 令 $\xi=\frac{1}{\alpha}$ 对应极值 II 型分布, 位置参数为 $(\mu-\alpha\sigma)$, 尺度参数为 $\alpha\sigma$; 当 $\xi<0$ 时, $\xi=-\frac{1}{\alpha}$ 对应极值 III 型分布, 位置参数为 $(\mu+\alpha\sigma)$, 尺度参数为 $\alpha\sigma$; 当 $\xi=0$ 时对应极值 I 型分布, 因为 $\lim_{\xi\rightarrow 0}\exp\left\{-\left(1+\xi\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\xi}}\right\}=e^{-e^{-\frac{x-\mu}{\sigma}}}$ 。

国内应用极值方法进行地震活动分析时, 通常作如下 2 个假定 (陈培善, 林邦慧, 1973; 夏成明等, 2002):

(1) 地震震级 x 和相应的发震频度 $n(x)$ 遵守古登堡-里克特 (G-R) 经验公式:

$$\ln n(x)=a-bx.\tag{4}$$

(2) 单位时间内发生大于某个震级的地震次数 η 服从泊松分布:

$$P(\eta=k)=\frac{\lambda^k}{k!}e^{-\lambda},(\lambda>0,k=0,1,2,\cdots).\tag{5}$$

在这 2 个假定下可以推导出地震震级分布为指数分布:

$$F(x)=1-e^{-\beta x},x>0,\beta=b\ln 10,\tag{6}$$

单位时间内最大地震震级分布为

$$G(x)=e^{-e^{-\beta(x-\mu)}},\mu=\ln\lambda/\beta.\tag{7}$$

这实际上与极值统计中得到的指数分布属于 Gumbel 分布吸引场的结果是一致的。而事实上除指数分布外还有很多分布也属于 Gumbel 分布吸引场。最大值的渐进分布若存在, 则也属于广义极值分布, 这从另一个侧面说明在应用极值理论作统计分析时, 直接从广义极值分布模型出发将具有更广的适用性。

2 地震活动危险性分析中的极值方法

目前在地震危险性分析与安全性评价中, 应用较多的是在前述两个假定下, 基于 Gumbel 型极值分布与泊松分布推导出相应的地震预报方法。实际上可由极限理论推导出很多与地震预报相关的公式 (张卫东等, 2005)。本文基于广义极值分布推导出若干个与地震危险性分析与预报有关的公式与方法, 以拓展极值理论在地震应用中的范围。

设 X_1,X_2,\cdots,X_n 为来自于最大震级分布——广义极值分布的样本, 记 $K=\sum_{i\leq n}I_{(x_i\geq x)}$, 其中 $I_{(x_i\geq x)}$ 为示性函数, 当 $X_i\geq x$ 时为 1, 否则为 0, K 表示数据集 $\{X_i|i=1,\cdots,n\}$ 中超过 x 的个数, 则 K 服从参数为 n , $p=1-G_{\xi}(x)$ 的二项分布。

2.1 平均复发周期

设最大震级超过 x 发生一次的平均复发周期为 $T(x)$, 即要求 $T(x)$ 年内最大震级超过 x 的平均次数为 1, 也就是 $E\sum_{i\leq T}I_{(X_i\geq x)}=1$ 即 $T(x)(1-G_{\xi}(x))=1$, 得

$$T(X)=1/(1-G_{\xi}(x)).\tag{8}$$

反过来也可以推得年内最大地震震级重现水平:

$$M(T)=G_{\xi}^{-1}(1-\frac{1}{T}).\tag{9}$$

2.2 地震危险性

未来 T 年内发生最大震级超过 x 的概率

$$P\{\min(m;X_m\geq x)\leq T\}=1-[G_{\xi}(x)]^T.\tag{10}$$

或者年内发生地震震级不超过 x 的概率

$$p=P(X_1\leq x,\cdots,X_T\leq x)=[G_{\xi}(x)]^T.\tag{11}$$

据此可以推知, 年内以概率的把握推测发生地震的

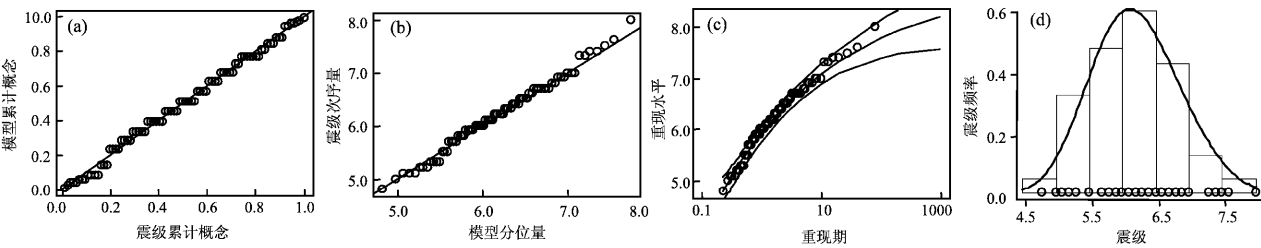


图2 广义极值分布拟合诊断图

(a) 概率 $P-P$ 图; (b) 分位数 $Q-Q$ 图; (c) 重现水平图; (d) 密度曲线估计与直方图

Fig. 2 Fitting diagnostic graphs of generatized extreme value distribution

(a) Probability graph of $P-P$; (b) Quantile graph of $Q-Q$; (c) Recrcence level graph;
(d) Density curve estimation and histogram

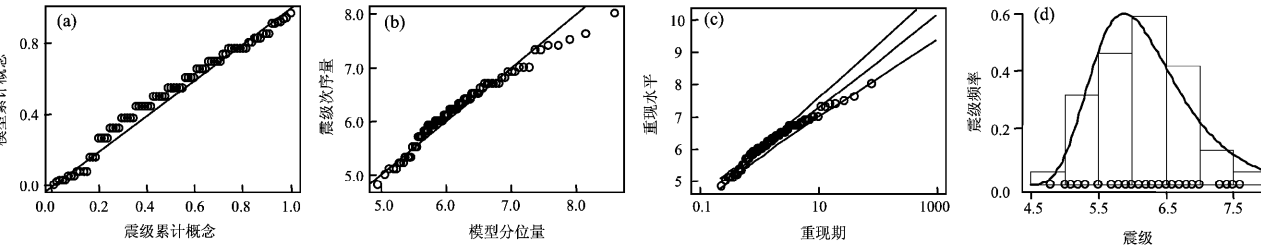


图3 Gumbel 分布拟合诊断图

(a) 概率 $P-P$ 图; (b) 分位数 $Q-Q$ 图; (c) 重现水平图; (d) 密度曲线估计与直方图

Fig. 3 Fitting diagnostic graphs of Gumbel distribution

(a) Probability graph of $P-P$; (b) Quantile graph of $Q-Q$; (c) Recurrence level graph;
(d) Density curve estimation and histogram

从图 2a 和图 2b 可看出,所有点都几乎在一条直线上,因此不能否认所拟合的模型。由于 ξ 的估计值为负,因此相应分布存在有限上界,故重现水平曲线应渐近地趋于某个有限值,且重现水平图为 1 条凸曲线(图 2c)。密度曲线的估计与直方图也显示模型拟合较优(图 2d)。因此 4 个诊断图都支持拟合的广义极值分布模型。图 3

则是用 Gumbel 分布模型拟合台湾地区震级数据的模型诊断图,与图 2 对比可见用该模型拟合效果较差。

利用第 2 节中给出的公式,以 6 个月为时间步长计算 1970~2010 年台湾地区最大地震震级的平均复发周期及理论发震次数,并预测未来 0.5 年、1 年和 5 年内发震的概率,计算结果见表 2 和表 3。

表2 台湾地区地震复发周期计算表(1970~2010 年)

Tab. 2 Table of earthquake recurrence periods in Taiwan region from 1970 to 2010

震级 M_s	3.5	4.0	4.5	5.0	5.5	6.0	6.5	7.0	7.5	8.0	8.5	9.0
复发周期 T	1.00	1.00	1.00	1.03	1.17	1.65	3.11	8.15	32.73	273.22	18 909.54	∞
理论发震数	82	82	82	80	71	50	27	10	3	1	0	0
实际发震数	82	82	82	81	70	54	29	10	3	1	0	0

表3 台湾地区未来 T 年发震危险性

Tab. 3 Seismic risks in the next T years in Taiwan region

震级 ($M_s \geq$)	3.5	4.0	4.5	5.0	5.5	6.0	6.5	7.0	7.5	8.0	8.5	9.0
0.5 年内发震概率	0.99	0.99	0.99	0.97	0.85	0.60	0.32	0.12	0.03	0	0	0
1 年内发震概率	1.0	0.99	0.99	0.99	0.98	0.84	0.54	0.23	0.06	0.007	0	0
5 年内发震概率	1.0	1.0	1.0	1.0	0.99	0.99	0.98	0.73	0.27	0.04	0	0

4 结论与讨论

(1) 针对台湾地区最大震级分布的极值建模, 应用 Gumbel 极值分布拟合效果较差, 从分析结果可以看出震级分布服从具有有限上界的 Weibull 分布。基于广义极值分布的地震危险性分析模型, 一方面理论上是合理的, 比传统方法更可靠, 另一方面也能够避免以往单独采用某一分布的不足。

(2) 根据广义极值分布计算出的地震理论发震数与实际发震数相比较, 震级超过 3.5 ~ 4.5 级的地震次数理论与实际一致, 事实上震级数据中最大震级几乎都超过 4.5 级。由于极值理论主要特点在分析分布尾部特征, 因此比较大的震级发生才是我们所关心的, 表 2 显示 $M_s \geq 7.0$ 地震次数符合实际, 由此可见极值理论是强震预报中一种较有效的方法之一。震级超过 5.0 ~ 6.5 级的地震次数实际发生数总是大于理论发震数, 这可能是因为是在所讨论的时间窗口中未能将余震信息删除。余震和震群被认为是非“独立”的事件, 是由于较大地震发生后, 在其震源区及其附近应力调整而产生的“相关”事件 (傅征祥, 1997), 这在一定程度上削弱了模型应用的变量独立性条件。

(3) 从地震发震危险性分析来看, 台湾地区半年内发震震级几乎都超过 5.0 级, 发生超过 6.0 级地震的概率达到 0.6; 1 年内发生超过 6.5 级地震的可能性也达到了 0.54。从更长时间来看, 未来 5 年台湾地区发生超过 6.5 级地震的概率为 0.98, 超过 7.0 级地震的概率高达 0.73, 危险性非常高, 应引起重视。需要说明的是, 以上针对台湾地区 41 年来的地震基础资料进行的极值分析, 仅是从有限的样本中得到的统计特征, 随着统计数据样本量的增加, 结果可能会更可靠。另外, 由于地震活动呈现出地震活跃—平静期交替变化

的周期特征和时间空间分布的丛集特点, 针对这种呈趋势性的震级序列怎样进行极值统计分析, 还值得进一步的研究和探讨。

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.
- 陈虹. 1996. 应用混合极值理论及最大似然法估计东南沿海各地震区的地震危险性[J]. 华南地震, 16(1): 10 - 15.
- 陈培善, 林邦慧. 1973. 极值理论在中长期地震预报的应用[J]. 地球物理学报, 16(1): 6 - 24.
- 傅征祥. 1997. 中国大陆地震活动性力学研究[M]. 北京: 地震出版社.
- 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.
- 贾素娟, 鄢家全. 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.
- 秦卫平. 1983. 定时段最大地震震级分布问题[J]. 地震研究, 6(4): 591 - 602.
- 史道济. 2006. 实用极值统计方法[M]. 天津: 天津科学技术出版社.
- 夏成明, 姜先畴, 张寅生, 等. 2002. 应用极值理论研究台湾和华东地区中强地震活动的相关性[J]. 地震学刊, 22(4): 40 - 45.
- 张卫东, 李茂林, 张秀梅, 等. 2005. 极值理论在地震危险性分析中的应用与研究[J]. 东北地震研究, 21(1): 24 - 30.
- De Haan L, Ferreira A. 2006. Extreme value theory[M]. New York: Springer, 3 - 33.
- 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.
- Huillet T, Raynaud H F. 1999. Rare events in a log-Weibull Scenario-Application to earthquake magnitude data[J]. Eur Phys J B, 12(3): 457 - 469.
- Nordquist J N. 1945. Theory of largest values applied to earthquake magnitudes[J]. Trans Am Geoph Union, 26: 29 - 31.
- Yegulalp T M, Kuo J A. 1974. Statistical prediction of the occurrence of maximum magnitude earthquakes[J]. BSSA, 64(2): 393 - 414.

Application of the Generalized Extreme Value Distribution in Seismic Hazard Analysis

QIAN Xiao-shi, WANG Fu-chang, CAO Gui-rong, REN Qing-qing

(*Institute of Disaster Prevention, Sanhe 065201, Hebei, China*)

Abstract

We study the maximum magnitude laws of earthquake by use of the generalized extreme value distribution, and give related methods and formula of seismic risk analysis with the distribution of generalized extreme value. From the statistical analysis of extreme values on historical earthquake data in Taiwan, we find out that the number of earthquake occurrence calculated by generalized extreme value distribution is completely consistent with that of actual earthquake occurrence when the maximum magnitude is greater than $M7$. On the basis of it, we predict the seismic risk in Taiwan in the coming years.

Key words: extreme value theory; generalized extreme value distribution; seismic risk