

III型极值分布在年最大地震 震级分析中的应用

徐人平

(云南工业大学,昆明 650051)

傅惠民

(北京航空航天大学,北京 100083)

摘要 在探讨年最大地震震级分布服从III型极值分布的基础上,给出了震级分布参数估计方法,讨论了地震趋势预报中的点估计和区间估计问题,指出了区间估计是一种较为有效的统计推断形式。它能够提高对地震预测预报的准确程度,从而为中长期地震预报提供一种新的方法。

关键词: 趋势预报 震级 震级分布 III型极值分布

1 引言

关于极值理论在地震长期预报中的应用问题,已有很多作者进行了研究。文献 [1-4] 对此进行了报导。但以往的研究主要集中在年最大地震震级的分布类型上。本文深入讨论了III型极值分布的性质,建立了大地震统计分布模型,并对年最大地震震级和重现周期进行了有效的预测预报。

2 年最大地震震级分布

对同一地区而言,在一段时期中地震震级分布规律具有相对的稳定性,这是因为在相当长的一段时期内,该地区的外界环境基本变化不大。另外,考虑该地区发生的地震是彼此独立的,每年发生的地震震级受到各种因素的影响,可将其看作是随机变量,用 d 表示。

设 1 年中地震震级 d 的分布函数为 $F(x)$, $d_1, d_2, d_3, \dots, d_n$ 为其子样,子样中的最大项 $d_n = \max\{d_1, d_2, \dots, d_n\}$ 的分布函数为 $F_n(u)$, 则有:

$$\begin{aligned} F_n(u) &= P\{d_n < u\} \\ &= P\{d_1 < u, d_2 < u, \dots, d_n < u\} \\ &= [F(u)]^n \end{aligned} \quad (1)$$

其分布密度函数为

$$f_n(u) = nf(u)[F(u)]^{n-1} \quad (2)$$

$F_n(u)$ 和 $f_n(u)$ 分别称为最大极值分布函数和最大极值分布密度函数。它们都是严格的函数,往往不能用于计算。这是因为一方面没有对震级分布函数 $F(x)$ 作任何确定,另一方面即使能给出严格的 $f_n(u)$ 分布函数表达式,有时也因计算复杂而难以实现。但只要 n 充分大, $F_n(u)$ 将趋向一个极限的分布函数——渐近分布。这种渐近分布的形式往往是简单的。

根据极值分布理论,这种极限分布存在并且退化,其极限分布只可能是下述 3种形式:

I 型渐近型式: Gumbel 双重指数型

$$F_{x_1}(x) = \exp\{-e^{-a_n(x-u_n)}\} \quad -\infty < x < +\infty \quad (3)$$

II 型渐近型式: Gumbel 指数型

$$\begin{cases} 0 & x \leq w \\ F_{x_2}(x) = \exp\left\{-\left(\frac{v_n}{x}\right)^k\right\} & x > 0, k > 0 \end{cases} \quad (4)$$

III型渐近型式: 威布尔指数型

$$\begin{cases} 0 & x > w \\ F_{x_3}(x) = \exp\left\{-\left(\frac{w-x}{w-w_n}\right)^k\right\} & x \leq w, k > 0 \end{cases} \quad (5)$$

当 $n \rightarrow \infty$ 时,极值可收敛于一种或几种极值分布的渐近型式。在某些情况下,初始分布的尾部情况可能是明显的,因而很容易确定其分属的渐近型式。但是,在许多情况下,其尾部的情况不明确,这就会涉及到如何确定其极值分布的极限型式问题。为此,可利用数学准则作为指南。

在这里采用冯米泽斯的准则概念来确定年最大地震震级的分布。

$$\text{若} \quad \lim_{x \rightarrow \infty} \frac{d}{dx} \left[\frac{1}{h_n(x)} \right] = 0 \quad (6)$$

则对于大的 n 值, x 的极值将收敛于I 型渐近分布型式。

$$\text{式中} \quad h_n(x) = \frac{f_n(x)}{1 - F_n(x)}$$

$$\text{若} \quad \lim_{x \rightarrow \infty} h_n(x) = k \quad k > 0 \text{且为常数} \quad (7)$$

则取自初始变量 X 的极值,最大值将收敛(分布)于II 型渐近型式。

$$\text{若} \quad F_X(w') = 1.0 \quad (\text{对于上限 } w') \quad (8)$$

$$\text{及} \quad \lim_{x \rightarrow w} (w' - x)h_n(x) = k \quad (k > 0 \text{且为常数})$$

则来自初始变量 X 的最大值(分布)将收敛于III型渐近型式。

每年地震震级可以认为是在 0与最大值 w 之间有均匀分布的初始变量 X ,其概率密度函数

$$f_X(x) = \frac{1}{w - 0} = \frac{1}{w} \quad 0 \leq X \leq w$$

及分布函数

$$F_X(x) = \frac{x - 0}{w - 0} = \frac{x}{w}$$

在这里 $w' = w$ 因此,最大值可由方程(8)得出:

$$F_X(w') = 1$$

$$\lim(w' - x) \frac{\frac{1}{w}}{1 - \frac{x}{w}} = 1$$

于是,可以认为每年地震震级的最大值将逐渐收敛于Ⅲ型渐近型式(威布尔分布)。

3 年最大地震震级分布的参数估计

在(5)式中,系数 w 为极值上限,即年最大地震震级理论极限;参数 w_n 为特征最大值,亦等于众数;参数 k 为形状参数,特征极值,而 $1/k$ 可以进行离散性量度。上述 3 个参数可以由下面诸式计算:

$$k = L_{xy} / L_{xx} \quad (9)$$

$$W_n = w - \exp(\bar{x} - \bar{y} L_{xx} / L_{xy}) \quad (10)$$

$$E(w) = L_{x0} / L_{xx} - L_{y0} / L_{xy} \quad (11)$$

$$X_i = \ln(w - w_n), \quad Y_i = \ln \ln\left(\frac{1}{p_i}\right) \quad (p_i \text{ 为平均秩}) \quad (12)$$

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad (13)$$

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i \quad (14)$$

$$L_{xx} = \sum_{i=1}^n X_i^2 - \frac{1}{n} \left(\sum_{i=1}^n X_i \right)^2 \quad (15)$$

$$L_{yy} = \sum_{i=1}^n Y_i^2 - \frac{1}{n} \left(\sum_{i=1}^n Y_i \right)^2 \quad (16)$$

$$L_{xy} = \sum_{i=1}^n X_i Y_i - \frac{1}{n} \left(\sum_{i=1}^n X_i \right) \left(\sum_{i=1}^n Y_i \right) \quad (17)$$

$$L_{x0} = \sum_{i=1}^n \frac{X_i}{w - x_i} - \frac{1}{n} \left(\sum_{i=1}^n X_i \right) \left(\sum_{i=1}^n \frac{1}{w - x_i} \right) \quad (18)$$

$$L_{y0} = \sum_{i=1}^n \frac{Y_i}{w - x_i} - \frac{1}{n} \left(\sum_{i=1}^n Y_i \right) \left(\sum_{i=1}^n \frac{1}{w - x_i} \right) \quad (19)$$

首先通过(11)式求出 w , 然后通过(9)式和(10)式求得 k 和 w_n 。(11)式较复杂,需用迭代方法求得。由于 $E(w)$ 具有如下性质: 设 $\hat{w}$ 为 w 的预估值, 则当 $\hat{w} < w$ 时, $E(\hat{w}) > 0$ 而当 $\hat{w} > w$ 时, $E(\hat{w}) < 0$ 。又因为 w 必位于区间 (x_n, ∞) 内, x_n 为诸最大地震震级 $x_i (i=1, 2, \dots, n)$ 中的最大者, 所以可以容易地求得 w 。

其线性相关系数 V 由下式给出:

$$V = \frac{L_{xy}}{L_{xx} L_{yy}}$$

4 地震预报的点估计和区间估计

由于对同一地区来讲,在相当长的一段时期内,可以认为基本外界环境变化不大,这样就可以把未来发生的地震与过去发生的地震看作具有相同的特性,因而在确定了年最大地震震

级的分布类型后,可以根据历史数据预测未来若干年可能发生的年最大地震震级。

对某个较大的震级 x

$$P = p(X \geq x) = 1 - p(X < x) = 1 - F(x)$$

p 表示年最大地震震级大于等于 x 的概率,或者说,平均 $1/p$ 年可能发生一次震级大于等于 x 的地震。

$$T(x) = \frac{1}{p} = \frac{1}{1 - F(x)}$$

称为重现期。反过来,给定重现期,也可求出在此期间内,平均最大震级 x 可能是什么值,即

$$F(x) = 1 - \frac{1}{T}$$

$$x(T) = F^{-1}(1 - \frac{1}{T})$$

以上是根据历年年最大地震震级随机变量 X 进行研究的。取得样本并依靠样本根据III型极值去估计参数,根据年最大地震震级预测重现期或根据重现期分析年最大地震震级,由此而得到的估计为点估计。

点估计虽然在某种平均的意义上反映出年最大地震震级的预测,但从区间估计的角度来看,仅凭点估计并不能给出带普遍性的结论。这种解法只是近似的而非确切的,并且无法看出它的可信程度(精度)有多大。

实际问题往往是要求根据带有随机因素的观测数据建立数学模型,并以此预测大地震发生的震级或重现期,预测点应该落入拟合曲线两侧的“置信带”内,这事实上就是一种区间估计,即

$$P(x_1 \leq x \leq x_2) \geq 1 - T$$

用这个随机区间作为年最大地震震级 x 的估计,则称 (x_1, x_2) 是 x 的置信水平为 $1 - T$ 的区间估计, $1 - T$ 为置信水平, x_2 和 x_1 分别为上、下置信限,即

$$P(x_1 \leq x) \geq 1 - T$$

$$P(x \leq x_2) \geq 1 - T$$

文献 [5] 给出了III型极值分布置信限曲线的计算方法,其单侧置信上、下限曲线可以分别表示为:

$$P_u = \exp\left[-\left(\frac{w_u - x}{w_u - w_{nu}}\right)^{K_u}\right]$$

$$P_L = \exp\left[-\left(\frac{w_L - x}{w_L - w_{nL}}\right)^{K_L}\right]$$

式中 w_u, w_{nu}, K_u 和 w_L, w_{nL}, K_L 均为置信参数,分别由下面方程组计算:

$$\frac{\partial Q}{\partial w_n} = 0 \quad \frac{\partial Q}{\partial w_{nu}} = 0 \quad \frac{\partial Q}{\partial w_{Ku}} = 0$$

式中

$$Q = \sum_{i=1}^n \{ [F_{nu}(x_i, w_u, w_{nu}, w_{Ku})] - H(P_{ui}) \}^2$$

$$= \sum_{i=1}^n \{ \ln \left[\ln \frac{1}{F_{nu}(x_i, w_n, w_{nu}, w_{Ku})} \right] - \ln \left(\ln \frac{1}{1 - P_{ui}} \right) \}^2$$

$$= \sum_{i=1}^n \left[K_u \ln(w_u - x_i) - K_u \ln(w_h - w_{nu}) - \ln \left(\ln \frac{1}{1 - P_{ui}} \right) \right]^2$$

和
$$\frac{\partial \mathcal{Q}}{\partial w_L} = 0 \qquad \frac{\partial \mathcal{Q}}{\partial w_{hL}} = 0 \qquad \frac{\partial \mathcal{Q}}{\partial w_{kL}} = 0$$

式中
$$\mathcal{Q} = \sum_{i=1}^n \left[K_L \ln(w_L - x_i) - K_L \ln(w_L - w_{nL}) - \ln \left(\ln \frac{1}{1 - P_{Li}} \right) \right]^2$$

P_{ui} 和 P_{Li} 分别为对应于母体百分率 $P= T$ 和 $P= 1- T$ 的百分位秩

$$P_{ui} = \frac{\lceil i/(n-i+1) \rceil F_{1-T, 2i, 2(n-i+1)}}{1+ \lceil i/(n-i+1) \rceil F_{1-T, 2i, (n-i+1)}}$$
$$P_{Li} = \frac{i/(n-i+1)}{i/(n-i+1) + F_{1-T, 2(n-i+1), 2i}}$$

5 应用举例

文献 [1]提供了中国大陆的最大地震震级数据 (表 1)。根据表 1数据和根据平均秩求出的点估计为:

$$F(x,T,U,X) = \exp \Big[- \big(\frac{9.9851-x}{3.3429} \big)^{5.0440} \Big]$$

并且可以求出置信水平为 90% 的区间估计的上、下界:

$$F(x,T,U,X) = \exp \Big[- \big(\frac{11.8499-x}{5.0221} \big)^{7.6202} \Big]$$
$$F(x,T,U,X) = \exp \Big[- \big(\frac{8.9974-x}{2.5232} \big)^{3.9124} \Big]$$

为了便于分析对照,用 1913- 1970年的 58个数数据预测了未来 10年中国大陆年最大地震震级和重现期,并与 1971- 1980年内实际发生的最大地震情况进行了比较 (表 2和表 3)。

表 1 中国大陆年最大地震震级数据表

序数	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
年份	1913	1914	1915	1916	1917	1918	1919	1920	1921	1922	1923	1924	1925	1926	1927	1928	1929
震级	6.5	7.5	7.0	5.0	6.5	6.0	6.5	8.5	6.5	6.5	7.25	7.25	7.0	6.25	8.0	6.0	6.5
序数	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
年份	1930	1931	1932	1933	1934	1935	1936	1937	1938	1939	1940	1941	1942	1943	1944	1945	1946
震级	6.5	8.0	7.5	7.5	7.0	6.5	6.75	7.5	6.25	6.0	7.25	7.0	6.75	5.75	7.25	6.25	7.25
序数	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51
年份	1947	1948	1949	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960	1961	1962	1963
震级	7.75	7.25	7.25	8.5	8.0	7.5	6.25	7.25	7.5	5.75	7.0	6.5	6.75	6.75	6.8	6.8	7.0
序数	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68
年份	1964	1965	1966	1967	1968	1969	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980
震级	6.5	6.6	7.2	6.8	5.7	7.4	7.7	6.8	6.2	7.9	7.3	7.3	7.8	6.6	6.4	6.8	6.6

表 2 根据 1913— 1970年数据预测 1971— 1980年地震重现期 (年)

最大震级	点估计	区间估计	实际发生
$x= 7$	2.3	2.3— 3.0	2.5
$x= 7.5$	4.7	4.7— 8.0	5

表 3 根据 1913— 1970年数据预测 1971— 1980年年最大地震震级

重现期 (年)	点估计	区间估计	实际发生
$T= 5$	7.5	7.8	7.8
$T= 10$	7.8	8.0	7.9

表 2和表 3中用到的III型极值分布数学表达式分别为:
点估计:

$$F(x,T,U,X)=\exp\Big[-\Big(\frac{10.1352-x}{3.5131}\Big)^{5.134}\Big]$$

区间估计:

$$F(x,T_L,U_L,X)=\exp\Big[-\Big(\frac{9.0477-x}{2.6122}\Big)^{3.9580}\Big]$$

$$(T=0.1) \qquad F(x,T_u,U_u,X)=\exp\Big[-\Big(\frac{12.2702-x}{5.4400}\Big)^{7.9801}\Big]$$

从表 2和表 3可以看出 ,点估计只是近似的而非确切的 ,只有区间估计才能保证至少 (1- α)% 是正确的。

6 结论

- (1)III型极值分布 (威尔分布)适用于分析年最大地震震级分布。
- (2)采用本文推荐的方法可以准确、方便、迅速地估计年最大地震震级分布的参数 ,从而对最大地震震级或重现期进行预测。
- (3)区间估计是一种很重要的统计推断形式 ,区间估计的宗旨就是充分利用样本提供的信息 ,以作出尽可能可靠和精确的估计 ,从而在大地震预测中有着特殊重要的意义。
- (4)应用举例说明 ,区间估计能够准确地进行中长期预报。

参考文献

1 史道济.年最大地震震级分布的统计检验.数理统计与应用概率.1987,2(3):257-275.
2 秦卫平.定时段最大地震的震级分布问题.地震研究,1983,6(4):592-602
3 徐云波.极值分布理论在地震趋势预报中的作用.东北师范大学学报(自然科学版).1985,(1):31-34.
4 陈培善,林邦慧.极值理论在中长期地震预报中的应用.地球物理学报,1973,(5):7-24.
5 傅惠民,高镇同,徐人平.三参数威布尔分布的置信限.航空学报,1992,13(3):A153-162.

TYPE-III EXTREME-VALUE DISTRIBUTION AND ITS APPLICATION TO ANALYSIS OF ANNUALLY LARGEST MAGNITUDES OF EARTHQUAKES

Xu Renping

(Yunnan Polytechnic University, Kunming 650051)

Fu Huimin

(Beijing University of Aeronautics and Astronautics, Beijing 100083)

Abstract

In this paper, based on Type-III extreme-value distribution, the method of parameter estimation is given and point estimation and interval estimation of earthquake prediction are discussed. The prediction accuracy will greatly be improved and the theoretical basis for medium term and long term earthquake prediction is provided.

Key words Trend prediction, Magnitude, Magnitude distribution, Type-III extreme-value distribution

(上接第 47 页)

EFFECT OF FUTURE EARTHQUAKES ON THE DUNHUANG MOGAO GROTTOES AND THEIR SUBSIDIARY BUILDINGS

Shi Yucheng

(Earthquake Research Institute of Lanzhou, SSB, Lanzhou 730000)

Abstract

From the angle of earthquake disaster mitigation of the Mogao Grottoes, this paper expounds their seismogeology background and effects of historical earthquakes on them, gives the factors influencing the seismic stability of Mogao caves and their subsidiary buildings, and theoretically puts forward the preliminary prediction method for earthquake damages of the caves and their subsidiary buildings.

Key words Earthquake damage, Seismic damage prediction, Mogao Grottoes, Subsidiary building, Aseismic property