

负二项式极值模式在地震活动中、长期趋势分析中的初步应用

张 德 齐

(江苏省地震局)

负二项式分布远比简单波松分布更符合于地震次数的观测分布。所以用负二项式分布代替简单波松分布以建立极值模式, 将使统计效果得到改善。

一、问题的提出与负二项式分布检验

在地震活动中、长期趋势分析中, 常用到的极值模式

$$G(x_i) = e^{-a} \left(\frac{e^{-\beta x_1} - e^{-\beta x_2}}{1 - e^{-\beta x_2}} \right) \quad (1)$$

是在简单波松分布的假定下推导出来的[1]。在简单波松分布的假定下, 任何彼此不重叠的时间间隔内所发生的地震事件数都视为相互独立的其期望值与方差都是不变的。事实上由于地质构造之间的相互联系以及震源物理过程的复杂性等等原因, 在实际观测中经常会遇到许多地震事件相互之间并非独立而是表现出某种相关性; 单位时间里发生的地震次数的均值亦并非恒定不变。对于这样的资料使用极值模式(1)去进行统计, 往往会出现较大的误差。为了寻找合理的分布, 设想有一个随机变数 ξ 。它服从的分布与一个参数 a 有关, 而 a 也是一个随机变数, 它服从一个确定的分布。即 ξ 服从于一个复合分布。负二项式分布[2]就是属于这样一种类型。其概率函数式为:

$$p(\xi = k) = (-1)^k \binom{-v}{k} p^k q^v \\ = g^v p^k v(v+1)(v+2) \cdots (v+k-1)/k! \quad (2)$$

为了与波松分布作比较, 采用柯莫哥洛夫——斯米尔洛夫最优拟合检验方法[3]对有关地震资料进行负二项式分布与波松分布检验。其要点如下:

- 1) 令 $F(x)$ 为完满地满足零假设 H_0 下的理论累计分布函数。
- 2) 令 $S_n(x)$ 为基于 n 次观测的样本分布函数, 对于任何观测值 x , $S_n(x) = k/n$; 这里 k 是小于或等于 x 的观测值的数目。
- 3) 最大偏差 D 定义为 $D = \max |F(x) - S_n(x)|$ 。

4)若对于所选定的显著水平 α ,其D的观测值大于或等于相应的临界值时,则拒绝接受 H_0 。

$$\text{波松分布的 } F(x) = \sum_{k=0}^x e^{-\lambda} \lambda^k / k! \quad (3)$$

$$\text{负二项式分布的 } F(x) = \sum_{k=0}^x (-1)^k \binom{-v}{k} p^k q^v \quad (4)$$

其中, x 及 $k = 0, 1, 2, \dots, n$ 。 $v > 0$; $0 < p < 1$; $0 < q < 1$; $p + q = 1$ 。计算观测值的数学期望及方差,然后利用负二项式分布的一、二阶原点矩及二阶中心矩公式便可计算出 v 、 p 的值,进而计算出 $F(x)$ 。

举出几个检验结果综合列表如下:

统计 区域	资 料 年 限	震 级 范 围	观 测 次 数	D = max	F(x) - S _n (x)	检 验 结 果	
				波松分布	负二项式分布	波松分布	负二项式分布
扬州~铜陵带	1505~1974	$M_s \geq 4^{3/4}$	47(年)	0.0680	0.0300	$\alpha = 0.20$ 接受 H_0	$\alpha = 0.20$ 接受 H_0
下扬子北东带	1970.1~1980.6	$M_s \geq 2.0$	126(月)	0.1054	0.0338	$\alpha = 0.15$ 拒绝 H_0	$\alpha = 0.20$ 接受 H_0
华北地区	999~1976	$M_s \geq 4^{3/4}$	978(年)	0.0550	0.0184	$\alpha = 0.01$ 拒绝 H_0	$\alpha = 0.20$ 接受 H_0
全 国	1900~1976	$M_s \geq 6$	77(年)	0.1745	0.0610	$\alpha = 0.05$ 拒绝 H_0	$\alpha = 0.20$ 接受 H_0
全 球	1904.1 1972.12	$M_s \geq 7$	828(月)	0.0454	0.0167	$\alpha = 0.10$ 拒绝 H_0	$\alpha = 0.20$ 接受 H_0

由表可见地震事件发生的负二项式规律与观测分布之间的最大偏差总是显著小于按波松规律时所产生的最大偏差。且多数情况下可在一定的 α 下拒绝接受波松分布。而负二项式分布却可以全部被接受。

二、负二项式极值概率统计模型

经过分布检验之后,以负二项式分布代替波松分布并采用与文献[1]所介绍的模式(1)的推导过程相同的步骤,推导出一种新的极值分布函数式为:

$$G(x) = \left[1 + Q \left(\frac{e^{-\beta x} - e^{-\beta x_2}}{1 - e^{-\beta x_2}} \right) \right]^{-v} \quad (5)$$

它仍然是负二项式形式,故称为负二项式极值分布函数。其中 $Q = P/g$; x_2 、 β 、 b 的物理意义均与模式(1)的相同。

由以下演算可以找到“负二项式极值分布函数”与“简单波松极值分布函数”之间的关系。

在(5)式中,令其对于任何 V 值均有 $p v / q = a$,于是 $p = a / (v + a)$ 。而后,令 $V \rightarrow \infty$ 即有

$$\begin{aligned} G(x) &= \lim_{V \rightarrow \infty} \sum_{k=0}^{\infty} g^v p^k \left(\frac{1 - e^{-\beta x}}{1 - e^{-\beta x_2}} \right)^k \frac{v(v+1)(v+2)\dots(v+k-1)}{k!} \\ &= \lim_{V \rightarrow \infty} \sum_{k=0}^{\infty} \left(\frac{1 - e^{-\beta x}}{1 - e^{-\beta x_2}} \right)^k \frac{a^k v(v+1)(v+2)\dots(v+k-1)}{k!} \left(1 - \frac{a}{v+a} \right)^v \end{aligned}$$

因为 $\lim_{v \rightarrow \infty} \frac{v(v+1)(v+2)\cdots(v+k-1)}{k!} = 1$ 及

$$\lim_{v \rightarrow \infty} \left(1 - \frac{a}{v+a}\right)^v = e^{-a}$$

故有 $G(x) = e^{-a} \sum_{k=0}^{\infty} \frac{\left(\frac{1 - e^{-\beta x}}{1 - e^{-\beta x_2}}\right)^k}{k!} a^k = e^{-a} \left(\frac{e^{-\beta x} - e^{-\beta x_2}}{1 - e^{-\beta x_2}}\right)$

得到了简单波松分布假定下的极值分布函数式。即在单位时间里 0 级以上地震平均发震次数恒定不变及 $v \rightarrow \infty$ 的极端情况假定下, 很容易由极值分布函数式(5)得到极值分布函数式(1)。可见后者是前者的特例, 因而负二项式极值模式更具有普遍意义, 也更有利于反映地震发生的复杂性。

三、新极值分布函数中的参数的确定

(5)式是个复杂的超越函数式, 不能直接采用最小二乘法逼近的方法求解参数, 经大量运算结果来看, 它的积分很可能是不收敛的。因此也无法应用矩估计的方法。为了以 n 个每单位时间中观测到的最大震级值 $x_j = x_1, x_2, \dots, x_n$ 来估计 v, β, Q, x_2 的值, 使用了一种三因素优选法[4]与最小二乘法混合并用的方法。其思路是, 考虑到在

$$G(x_j) = j/(n+1) = \left[1 + Q \left(\frac{e^{-\beta x_j} - e^{-\beta x_2}}{1 - e^{-\beta x_2}}\right)\right]^{-v} \quad (6)$$

式中 $j/(n+1)$ 是不含未知数的顺序统计量。因而一旦 Q, β, x_2 的值给定之后, (6)式就由超越方程变为一般的指数方程形式。很容易通过取对数的方法进行线性化处理, 而后直接用最小二乘法逼近相应的 v 值。其大体步骤是:

$$\ln h_j = \left[1 - Q \left(\frac{e^{-\beta x_j} - e^{-\beta x_2}}{1 - e^{-\beta x_2}}\right)\right]^{-v} \quad (7)$$

据三因素优选法原理在一定区间内挑选 Q, β, x_2 的值代入(6)式, 并对其等式两端取对数而后按最小二乘法公式得到

$$v = \sum_{j=1}^n \left[\ln G(x_j) \ln h_j \right] / \sum_{j=1}^n (\ln h_j)^2 \quad (8)$$

便可算得相应的 v 值。三因素优选法的挑选原则是使得挑选出来的 Q, β, x_2 及按(8)式所得到的相应的 v 值代入(6)式后, 等式两端数值之差的平方和越小越好。按预定的精度最终得的 Q, β, x_2 和 v 值即为所求的参数值。然后按以下公式及方法计算平均地震次数、复周期、理论频数和发震概率:

单位时间里发生 $M \geq x$ 的地震平均次数

$$N(x) = vQ(e^{-\beta x} - e^{-\beta x_2}) / (1 - e^{-\beta x_2}) \quad (9)$$

$M \geq x$ 的地震平均复发周期即为

$$T(x) = 1 / N(x) \quad (10)$$

$$m+1 \text{ 个单位时间里发生 } M \geq x \text{ 的地震的理论次数 } N_p = (m+1)N(x) \quad (11)$$

将 N_p 与 m 个单位时间里已发生的 $M \geq x$ 的地震次数 N 相比较, 并利用(2)式便可计算出

第 $m+1$ 个单位时间里再发生 $M \geq x$ 的地震的概率。

四、初步检验

1974年底以来,应用负二项式极值模型并结合其他地震活动性方法对扬州~铜陵地震带地震活动中、长趋势进行概率预测,经实际验证说明作为背景分析是有一定意义的。当时的预测值与5年来的实况列表如下:

表一、 扬州——铜陵地震带

资料: 1906~1970, $t = 5$ 年, $M_s \geq 4.3/4$, $m = 13$, $Q = 0.482$, $v = 45.27$, $\beta = 0.573$, $x_2 = 7.25$

时 间	震 级 x	4.5	5.0	5.5	6.0	6.5	7.0	备 注
1970~1975	理论频数 Np	17.3	11.9	7.8	4.7	2.4	0.7	内 部 检 验
	1906~1970已发震次数 Ns	12	9	5	4	2	0	
	预测概率 P	0.83	0.70	0.69	0.48	0.41	0.44	
	预测期内发震次数 N	1	1	1	0	0	0	
1976~1980	理论频数 Np	18.6	12.8	8.4	5.1	2.6	0.8	预 测
	1906~1975已发震次数 Ns	15	11	6	4	2	0	
	预测概率 p	0.71	0.59	0.71	0.54	0.46	0.48	
	预测期内发震次数 N	1	0	(1)*	1	0	0	
1906~19080.10实际频数		17(18)	12(13)	7(8)	5	2	0	

* 圆括号内为记入1976年北黄海5.5级地震后的数据。

据1910~1973年的历史资料以一年为时间单位所计算的结果与表一结果很吻合。为了与用模式(1)所计算的有关结果相比较,顺便列出一组用模式(1)对该地震带1971~1980年间的地震活动所进行的概率预测结果(表二)。按表二的结果,则该地震带在1971~1980年间不会有发生5级以上地震的危险。即便按原表所列的数值进行分析亦不会有发生6级以上地

表二 扬州——铜陵地震带

资料: 1901~1970, $t = 10$ 年, $m = 7$, $\beta = 1.13$, $\mu = 4.88$, $x_2 = 8$

震 级 x	4.5	5.0	5.5	6.0	6.5	备 注
复发周期 $T(x)$	0.65 (0.66)	1.15 (1.2)	2.02 (2.14)	3.54 (3.96)	6.23 (7.6)	从1970年外推到 1980年的理论发震 次数
理论值 $m/T(x)$	10.8 (10.6)	6.1 (5.8)	3.5 (3.3)	2.0 (1.8)	1.1 (0.9)	
外推值 $(m+1)/T(x)$	12.3 (12.1)	7.0 (6.7)	4.0 (3.7)	2.3 (2.1)	1.3 (1.1)	
* 实际值 $Ns(x)$	5 (12)	4 (9)	3 (5)	2 (4)	1 (2)	
预测期内发震次数 N	2	1	1(2)	1	0	1901年~1970年 内已发震次数
						1971年~1980年 发震次数

* 该栏未加[]号者是原表中的数字。加[]号者, $M_s \geq 4\frac{3}{4}$ 地震数目是按本局1974年编印的扬州~铜陵地震带地震自录剔除余震而得到的。

震的危险。这一结果与近十年来的实况出入较明显。其误差较大的原因，除与所用的资料残缺有关之外还与所用模式有关。

负二项式极值模式初步应用验证的结果虽比模式(1)的效果有改进，但还未对外地区特别是多震区的资料进行过验证；在震级-频度关系、顺序统计量等方面也仍有值得进一步改进之处。不过，可能会使得计算量变得更加庞大。

(本文1981年11月6日收到)

参 考 资 料

- [1] 地球物理所，极值理论在中长期地震预报中的应用，地震技术参考资料，地震战线编辑部，1973年 第2期、
- [2] [波兰] M费史，概率论及数理统计，王福保译，第五章，上海科学技术出版社，1957.
- [3] Bernard Ostle, Statistics in research, 471—472, U, S, A, 1963.
- [4] 北京市优选法应用推广小组，优选法及其应用，24，北京人民出版社，1972.

THE INITIAL APPLICATION OF THE NEGATIVE BINOMIAL MODEL OF EXTREME VALUE IN ANALYSIS OF MEDIUM-TERM AND LONG-TERM SEISMICITY TENDENCY

Zhang De-qi