

研究简报

复杂记忆概率预测模型*

王 健

(中国北京 100081 国家地震局地球物理研究所)

现行的地震概率预测方法是假定相邻地震之间的时间间隔符合某一分布,根据上一次地震的发生时间,便能够计算预测时段内的发震概率,时间间隔的分布可以是多种多样的,因而形成了许多模型.对于这些模型来说,有一点是共同的,就是时间的起算点都是上一次地震的发震时刻,也就是发震概率仅与上一次地震有关.形象地讲,如果认为模型是有记忆能力的话,现行的模型只能“记住”一次地震.我们称之为简单记忆.

判定应用简单记忆模型是否合理,首先要看应用的前提条件:空间范围以及该范围内地震之间的关系.如果空间范围较小,相当于单一孕震区范围,在一定的假定下,可以认为地震之间的关系是直接相关的,应用简单记忆模型是合理的(如时间可预报和滑移可预报模型).但对单一孕震区范围来说往往缺乏实际资料.现阶段多需在地震带的范围内进行预测.地震带的范围相对较大,其内存在多个相互关联的孕震区,在该情况下假定地震之间关系是直接的,不尽合理,因而应用简单记忆模型也是不合理的.笔者认为,地震带内地震之间关系既不是直接的,也不是独立的,应当是间接的.从总体上来看待地震活动性,一次地震不仅仅与上一次地震有关,与前几次地震都应存在某种联系.本研究的目的是要提供一种具有较强“记忆能力”的模型,不仅能够“记住”上一次地震,还能“记住”前多次地震,以期提取更多信息,给出较为合理的地震预测.

1 基本思路

如果得到相邻地震之间的时间间隔分布 f_1 ,再根据上一次地震(图1中用-1表示)的发震时刻,就能够计算预测时段的发震概率 P_1 ;同样,如果我们得到相间地震的时间间隔分布 f_2 ,再根据已发生的倒数第二个地震(图1中用-2表示),又可以得到一个发震概率 P_2 ;如果资料允许的话,接下去可以根据倒数第三个地震(图1中用-3表示)和间隔两个地震的时间间隔分布 f_3 得到第三个发震概率 P_3 .理论上,可以得到 j 个概率值.

由此可以看到,对未来同一预测时间段,可以得到 j 个预测概率,过去常用方法得到的只是其中一个: P_1 .这 j 个概率值通常情况下是不相等的,它们分别是根据已发生的 j 次地震独立地给出预测时段的发震概率.总的发震概率应是这 j 个概率值的综合,而不仅仅是 P_1 .设总发震概率为 P ,则

* 国家地震局地球物理研究所论著 94A0081.

1993年8月20日收到初稿,1993年12月5日决定采用.

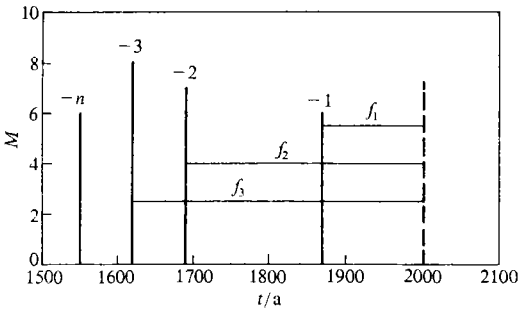


图 1 复杂记忆模型基本思路示意图

$$P = \frac{\sum_{n=1}^j w_n \cdot P_n}{j} \tag{1}$$

式中, w_n 是权系数. 权系数的合理确定有待进一步研究.

各时间间隔分布 f_j 由实际资料统计、拟合得到.

以地震带为基本空间单元. 设共有 M 个地震带、 N 个地震, 对第 k 地震带, 有 N_k 个地震, 则有 $(N_k - j)$ 个时间间隔对应于 f_j , 平均时间间隔为 $T_{av, kj}$

$$T_{av, kj} = \frac{1}{(N_k - j)} \cdot \sum_{i=1}^{N_k-j} T_{i, kj} \tag{2}$$

式中, $T_{i, kj}$ 表示第 k 地震带中第 i 个对应于 f_j 的时间间隔. $T_{i, kj}$ 除以 $T_{av, kj}$, 得到一种标准化的量 $T_{i, kj}/T_{av, kj}$. 对于 k 地震带有 $(N_k - j)$ 个 $T_{i, kj}/T_{av, kj}$. 对所有 M 个地震带求标准化变量的均值和均方差

$$T_{av, j} = \frac{1}{N - M \cdot j} \cdot \sum_{k=1}^M \sum_{i=1}^{N_k-j} \frac{T_{i, kj}}{T_{av, kj}} \tag{3}$$

$$\sigma_j = \sqrt{\frac{1}{N - M \cdot j - 1} \cdot \sum_{k=1}^M \sum_{i=1}^{N_k-j} (T_{i, kj} - T_{av, kj})^2} \tag{4}$$

对 $(N - M \cdot j)$ 个标准化的数据, 分档统计可得到其分布的直方图. 由式(3)、(4)得到的均值和均方差, 可得到一些理论概率分布, 对直方图进行拟合. 至此, 就得到了分布 f_j .

2 资料及处理

2.1 基础资料

我国东部地区历史地震资料较为丰富, 允许进行多种时间间隔的统计拟合. 本研究对资料的可靠性和完整性要求较严格. 根据黄玮琼等(1994)的研究结果, 我国东部各地区地震资料是不平衡的. 为保证各地震带资料的完整性和可靠性, 每一地震带资料的起始年代都是不同的, 详见表 1. 其中编号 08, 09 两地震带内缺乏资料. 地震带的划分采用了全国区划图(中国地震烈度区划图编委会, 1992)的

表 1 地震带基础数据表

| 地震带 编号 | 02 | 03 | 04 | 05 | 06 | 07 | 10 | 11 |
|--------------|-------|-------|------|------|-------|------|------|-------|
| 资料 起始年 | 1500 | 1300 | 1290 | 1900 | 1800 | 1800 | 1900 | 1600 |
| 地震 个数 | 12 | 18 | 19 | 5 | 4 | 9 | 4 | 10 |
| $T_{av, k1}$ | 38.7 | 39.0 | 29.1 | 12.7 | 37.3 | 17.3 | 16.2 | 35.2 |
| $T_{av, k2}$ | 82.7 | 71.6 | 55.3 | 31.2 | 98.5 | 37.8 | 37.5 | 77.9 |
| $T_{av, k3}$ | 130.9 | 107.3 | 83.2 | 46.8 | 111.8 | 57.4 | 48.6 | 132.0 |

结果.

本项研究中考虑的最小震级为 6 级. 去除了前震和余震, 包括 6 级以上强余震. 双主震型地震仅作一个地震处理, 例如 1976 年唐山地震只保留 7.8 级主震.

2.2 曲线拟合

已有研究结果显示 (Nishenko and Buland, 1987), 地震时间间隔主要服从对数正态分布和 Weibull 分布. 本项研究主要用这两种分布来拟合我国东部地区的地震资料, 以比较哪种分布更符合我国东部的实际情况.

对 $T_{i, k1}/T_{av, k1}$ 以 0.2 间隔分档累计并进行归一化, 得到相应的频率分布图 (图 2a); 对 $T_{i, k2}/T_{av, k2}$ 和 $T_{i, k3}/T_{av, k3}$ 重复上述计算, 得到相应的频率分布图 (图 2e).

分别用对数正态分布和 Weibull 分布进行拟合. 对数正态分布的密度函数为

$$f(x) = \frac{1}{\sqrt{2\pi}\zeta x} \cdot e^{-\frac{(\ln x - \lambda)^2}{2\zeta^2}} \quad 0 \leq x < \infty \quad (5)$$

式中, λ, ζ 分别是 $\ln x$ 的均值和均方差.

为了简便, 令 $T_{i, kj}/T_{av, kj}$ 的均值为 α , 均方差为 σ . α, σ 与 λ, ζ 的关系式为

$$\begin{cases} \zeta = \sqrt{\ln(1 + \frac{\sigma^2}{\alpha^2})} \\ \lambda = \ln \alpha - 0.5\zeta^2 \end{cases} \quad (6)$$

Weibull 分布的密度函数可以表示为

$$f(x) = Kx^m \cdot e^{-\frac{Kx^{m+1}}{m+1}} \quad 0 \leq x < \infty \quad (7)$$

式中, K, m 均为参数, 它们与均值和均方差的关系式为

$$\begin{cases} \alpha = \left(\frac{K}{m+1} \right)^{\frac{m+1}{m+2}} \cdot \Gamma\left(\frac{m+2}{m+1}\right) \\ \frac{\sigma}{\alpha} = \frac{\sqrt{\Gamma\left(\frac{m+3}{m+1}\right) - \Gamma^2\left(\frac{m+2}{m+1}\right)}}{\Gamma\left(\frac{m+2}{m+1}\right)} \end{cases} \quad (8)$$

现有 8 个地震带内共有符合要求地震数 81 个. 对于相邻地震时间间隔, 共有数据 73 个. 计算得到 $\alpha=1, \sigma=1.3139$, 代入式 (6) 得 $\zeta=1.0015, \lambda=-0.5015$; 代入式 (8) 得 $K=2.6264, m=1.6264$. 对数正态分布和威布尔分布的密度曲线 f_1 分别绘于图 2a, 相应的累积分布曲线绘于图 2b (图中 Weibull 分布用虚线表示).

对于相间地震时间间隔, 共有数据 65 个. 计算得到 $\alpha=1, \sigma=0.8457$, 代入式 (6) 得 $\zeta=0.7345, \lambda=-0.2698$; 代入式 (8) 得 $K=2.3644, m=1.3644$. 对数正态分布和 Weibull 分布的密度曲线 f_2 分别绘于图 2c, 相应的累积分布曲线绘于图 2d.

对于间隔两个地震的时间间隔, 共有数据 57 个. 计算得到 $\alpha=1, \sigma=0.6423$, 代入式 (6) 得 $\zeta=0.5877, \lambda=-0.1727$; 代入式 (8) 得 $K=2.0230, m=1.0230$. 对数正态分布和 Weibull 分布的密度曲线 f_3 分别绘于图 2e, 相应的累积分布曲线绘于图 2f.

2.3 有效性检验

分别对 Weibull 分布和对数正态分布进行 χ^2 检验.

对于相邻地震时间间隔, 对数正态分布的置信度为 99.5%. Weibull 分布无法通过检验 (从图 2a 中也很直观地看出).

对于相间地震时间间隔, 对数正态分布同样可通过置信度为 99.5% 的 χ^2 检验. Weibull 分布仍无法通过检验 (从图 2c 中可很直观地看出).

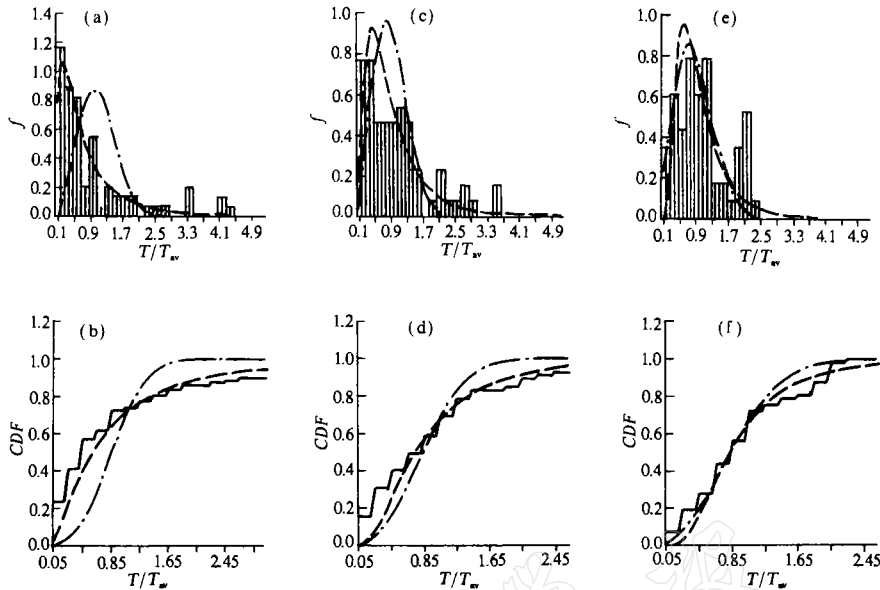


图 2 地震时间间隔的密度和分布曲线

对于间隔两个地震的时间间隔，对数正态分布无法通过置信度为 99.5% 的 χ^2 检验，但可通过置信度为 90% 的 χ^2 检验，而此时 Weibull 分布可通过置信度为 99.5% 的 χ^2 检验。

综上所述，如果取置信度为 90%，对数正态分布都可通过 χ^2 检验；而 Weibull 分布只有 f_3 可通过检验。

2.4 概率计算

从图 2a, c, e 中可以看到，实际资料统计的各种分布 f_j ，随着 j 的增大，标准化变量的最大值在减小。我们认为，这一特征用截头的对数正态分布能够较好地反映出来。

已知各历史地震的发震时刻 $t_{-1}, t_{-2}, \dots, t_{-n}$ ，令目前的时间为 t_0 ，计算未来时间段 $[t_1, t_2]$ 内，第 k 地震带、相应第 j 分布的发震概率 $P_{k,j}$ ，

$$P_{k,j} = \frac{F(\frac{t_1 - t_{-j}}{T_{av,kj}} < x \leq \frac{t_2 - t_{-j}}{T_{av,kj}})}{F(\frac{t_0 - t_{-j}}{T_{av,kj}} < x \leq \tau_j)} \tag{9}$$

式中， τ_j 为 j 分布的截点。当 $\tau_j = \infty$ 时即为对数正态分布。如何根据具体情况合理地确定截点，值得进一步研究。

对数正态的分布函数为

$$F(a < x \leq b) = \Phi(\frac{\ln b - \lambda}{\zeta}) - \Phi(\frac{\ln a - \lambda}{\zeta}) \tag{10}$$

式中， $\Phi(x)$ 为标准正态分布。

3 讨 论

(1) 从理论上可以进行 j 种地震时间间隔分布的统计、拟合，但由于资料所限，本次工作只作了 3 种。其余各种时间间隔分布的特征如何，以及何种理论曲线拟合较好，尚有待进一步研究。

(2) 计算中统一地使用了对数正态分布，也可用不同的分布，如间隔两个地震的时间间隔用

Weibull 分布. 本文阐述了复杂记忆模型的基本思路, 应用时拟合、计算中的一些具体问题可进行更细致地讨论.

本文得到了国家地震局地球物理研究所时振梁研究员、潘华博士研究生的大力帮助, 在此表示感谢.

参 考 文 献

- 黄玮琼、李文香、曹学锋, 1994. 中国大陆地震资料完整性研究之一——以华北地区为例. 地震学报, **16**, 273—280.
- 中国地震烈度区划图编委会, 1992. 中国地震烈度区划图(1990)及说明书. 中国地震, **8**, 1—10.
- Nishenko, S. P. and Buland, R., 1987. A generic recurrence interval distribution for earthquake forecasting. *Bull. Seism. Soc. Amer.*, **77**, 1382—1399.

地震学报
ACTA SEISMOLOGICA SINICA