

# 随机化检验方法及其应用\*

陈汉尧 胡聿贤

(中国北京 100081 国家地震局地球物理研究所)

## 摘要

首先系统叙述了无分布检验中的一种方法——随机化检验。此方法具有不需作任何分布假定及相应参数估计,对小样本及样本来自随机及非随机均适用的优点。作为实际应用,利用此方法,对鲜水河断裂带上地震的迁移规律及华北地区大地震的减震作用进行了随机化检验。检验结果表明,鲜水河断裂带上 6 级以上地震来回迁移规律具有 98.7% 以上的置信度;大华北地区 8 级以上地震对 7 级以上地震具有比较明显的减震作用。所得结果及方法本身在实际中具有一定意义。

**关键词** 无分布检验; 随机化检验; 强震迁移; 减震作用

## 1 前 言

地震中有多种复杂的现象,如大地震的迁移、大地震的减震作用及地震的填空性等等。对它们的解释一般有两种方法:一种是基于对物理过程的理解,利用已有的记录数据加以验证分析;另一种则是基于纯统计的数学手段,然后辅以一定的物理解释。无论用哪一种方法,均有一个对所研究的地震现象是否存在及存在的可能性有多大的问题。众所周知,地震数据短缺及影响地震发生因素多而复杂,这二者使得一些地震中似是“规律”的现象实际上是具有不确定性的,而具有不确定性的结论若没有给出其相应的置信度,则难以让人信服。因而,若能利用统计学中置信度的概念,对所得结论给出一定的置信度,就使人比较容易作出合理的拒绝或接受的决定。

本文的目的是针对地震数据少,且由此组成的样本并非完全随机的情形,引进一种无分布检验方法——随机化检验。同时,作为实际应用的例子,利用它对鲜水河断裂带上 6 级以上地震来回迁移规律及大华北地区大地震的减震作用进行了随机化检验。

## 2 随机化检验方法

统计学中,检验方法可分成分布检验及无分布检验两大类。常用的检验方法属于分布检验。无分布检验包括排序、Bootstrap 方法 (Erfon, 1979) 及随机化检验方法 (Edging-

\* 国家地震局地球物理研究所论著 93A0089  
1992 年 5 月 8 日收到本文初稿,1992 年 11 月 16 日决定采用。

ton, 1987)等,这些方法虽然互有差异,但它们有一共同特点,就是不需对样本及待估量作任何分布假定,直接由观察数据计算置信度.除此之外,随机化检验方法还适用小样本及样本是非随机的情形,而地震研究中地震数据少,同时由地震数据组成的样本并非随机,这些使得随机化检验方法在地震研究中具有较为广阔的应用前景.为此,下面就对其基本思想及特点作一简要介绍.

## 2.1 随机化检验方法的基本思想

现设有两个(或两组)变量  $X, Y$ ,其观察值为  $(x_i, y_i) (i, \dots, n)$ . 假设根据这些观察值得出变量  $X, Y$  间存在某种特定的关系,这种特定的关系可由统计量  $T(X, Y)$  加以度量.为了求解这种特定的关系的置信度,随机化检验方法采用零假设  $H_0$ ,为

$H_0$ : 变量  $X, Y$  间无任何相互关系.

基于  $H_0$ ,保证了变量  $X, Y$  的观察值可以予以随机化.当变量  $X, Y$  的观察值随机化以后,对于每一组由随机化得到的值  $(x_i^*, y_i^*)$ ,可以相应求得其统计量  $T_i^*(X, Y)$ . 假定随机化的次数为  $N$  次,另外设由实际观察数据计算得到的统计量为  $T_0(X, Y)$ . 由此,根据  $T_i^*(X, Y) (i=1, \dots, N)$  及  $T_0(X, Y)$  即可得到变量  $X, Y$  间存在这种特定关系的置信度.若随机化以后很容易得到与观察值相应的统计量  $T_0(X, Y)$ ,则说明  $X, Y$  存在由  $T(X, Y)$  表述的特定关系的置信度较低,反之则较高.

以研究地震在一条断层带内来回迁移问题为例.假设地震只能在断层上的  $A, B$  两点(对应于变量  $X$ )发生,实际上有两个地震分别在这两点发生了,它们的时间为 1900 与 1950 年(对应于变量  $Y$ )(如图 1),而统计量  $T(X, Y)$  为地震发生在  $A, B$  两点来回迁移.现在,根据零假设  $H_0$ ,在这里为地震的发生地点与发生时间无关.由此,可以将地震的发生地点予以随机化,可以得到以下 4 种可能结果:

| 随机化结果  | 1   | 2   | 3   | 4   |
|--------|-----|-----|-----|-----|
| 1900 年 | $A$ | $A$ | $B$ | $B$ |
| 1950 年 | $A$ | $B$ | $A$ | $B$ |

将这 4 组随机化后得到的结果与实际观察到的结果相比,可以发现仅有一组结果与实际观察到的结果相同,因而其置信度为  $3/4$ . 假设历史上有 3 次地震发生,即再加上一次在 1990 年发生于  $C$  点,则随机化后可以得到以下 27 种可能:

| 随机化结果  | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10  | 11  | 12  | 13  | 14  |
|--------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 1900 年 | $A$ | $A$ | $A$ | $B$ | $A$ | $B$ | $B$ | $B$ | $A$ | $A$ | $C$ | $C$ | $A$ | $C$ |
| 随机化结果  | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10  | 11  | 12  | 13  | 14  |
| 1900 年 | $A$ | $A$ | $A$ | $B$ | $A$ | $B$ | $B$ | $B$ | $A$ | $A$ | $C$ | $C$ | $A$ | $C$ |
| 1950 年 | $A$ | $A$ | $B$ | $A$ | $B$ | $A$ | $B$ | $B$ | $A$ | $C$ | $A$ | $A$ | $C$ | $C$ |
| 1990 年 | $A$ | $B$ | $A$ | $A$ | $B$ | $B$ | $A$ | $B$ | $C$ | $A$ | $C$ | $A$ | $C$ | $A$ |
|        | 15  | 16  | 17  | 18  | 19  | 20  | 21  | 22  | 23  | 24  | 25  | 26  | 27  |     |
| 1990 年 | $C$ | $B$ | $B$ | $C$ | $B$ | $C$ | $C$ | $A$ | $B$ | $C$ | $A$ | $C$ | $B$ | $B$ |
| 1950 年 | $C$ | $B$ | $C$ | $B$ | $C$ | $B$ | $C$ | $B$ | $C$ | $A$ | $C$ | $B$ | $A$ |     |
| 1990 年 | $C$ | $C$ | $B$ | $B$ | $C$ | $C$ | $B$ | $C$ | $A$ | $B$ | $B$ | $A$ | $C$ |     |

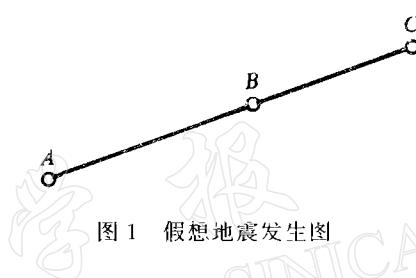


图 1 假想地震发生图

这时,从 27 个组合可以看出,中间只有一个(第 22) 随时间先后从 A 向 C 方向顺序迁移,因而其置信度为 26/27.

由以上例子可以看出,用随机化检验方法求取置信度有 3 个显著的特点:第一,零假设  $H_0$  非常简单,上例中为地震发生地点与地震发生时间无关;第二,对于地震发生未作任何分布假定;最后,就是直接由观察数据求取置信度,而对于观察数据是否是来自一总体的随机样本则无关紧要.

## 2.2 精确随机化方法与近似随机化方法

精确随机化方法与近似随机化方法没有本质的区别.若将所有各种随机组合的可能一一列出用以计算置信度时,就称其为精确随机化方法,上述地震迁移的例子即属于此类;当仅利用所有可能组合的某一随机样本计算置信度时,则称其为近似随机化方法.实际上,由于上述例子比较简单,因而可以将其各种可能一一列出,当问题比较复杂,观察数据又较大时,即使使用现代计算机,要将各种可能一一列出以求取置信度也几乎是不可能的.所以,在实际使用中,更多的还是利用近似随机化方法.已经证明(Edgington, 1969; Dwass, 1971),当随机化次数足够大时,由近似随机化方法得到的结果趋近于精确随机化方法得到的结果.因而,为简单计,下面就将近似随机化方法简称为随机化方法.

## 2.3 随机化检验方法的实施步骤

根据以上所述,随机化检验方法的实施步骤可归纳如下:

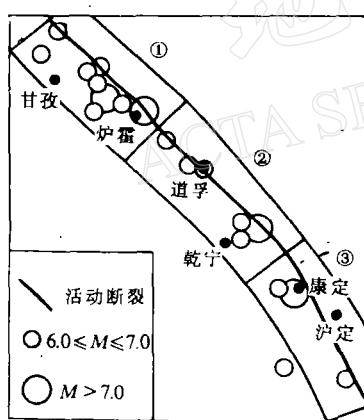


图 2 鲜水河地震带震中分布图

西南地区是我国地震活动强烈的地区之一,该区的地震活动特别是强震主要发生在几条断裂带上,其中鲜水河断裂带是地震活动最强烈的地震带,曾引起不少研究者的注意(邓天刚、龙德雄, 1986; 林邦慧等, 1986; 王阜, 1987).该带 6 级以上强震震中分布如图 2 所示.对鲜水河断裂带,许多学者认为其上地震发生有往返迁移的规律.为此,本文利用随机化检验方法,基于现有观察数据,计算其上地震发生具有往返迁移规律的置信度.为简便起见,根据邓天刚和龙德雄(1986)的研究,粗略地将鲜水河断裂带分成 3 段(如图 2 所示).

将发震地点作为纵坐标轴,发震时间作为横坐标轴,则可得出如图 3 所示的震级大于等于 6 级地震的强震迁移图(林邦慧等, 1986).由图可见,实际观察到的地震在鲜水河断裂带上表现出来回迁移,即若在①发生一次地震,下一次地震或相继朝前,或返回朝②、③

(1) 首先由观察数据计算待检验的统计量(或统计分布)  $T_0$ :

(2) 按检验目的的要求,将观察数据予以随机化,同时对应每一随机化后得到的新样本,计算相应的统计量  $T_i^*$ ,  $i=1, 2, \dots, N-1$ , 其中  $N$  为随机化次数;

(3) 由  $T_0$  及  $T_i^*$ ,  $i=1, 2, \dots, N-1$ , 即可以求得统计量  $T_0$  的置信度.

## 3 实际应用

### 3.1 在强震迁移规律研究中的应用

西南地区是我国地震活动强烈的地区之一,

该区的地震活动特别是强震主要发生在几条断裂

段移动. 在所有已有地震数据中, 没有在一段内有两个连续拐弯点, 同时在②段内未曾产生过拐弯. 因此, 在作随机化检验时, 作如下定义:

- (1) 在同段内不能连续发生两个拐弯点;
- (2) 在②段内不能产生拐点, 即若②段发生一个地震后, 其后一个地震必然顺前一地震与此地震的连线方向前移, 不产生拐弯.

满足以上两个条件的迁移现象即为顺序迁移.

据以上定义的顺序迁移现象作了随机化检验, 结果得置信度为 98.7%. 若放宽顺序迁移的要求, 即若在来回迁移过程中, 有一步不符合以上定义仍为正确, 此时置信度为 89%. 从以上的随机化检验结果可以看出, 鲜水河断裂带上 6 级以上强震来回迁移规律具有 98.7% 以上的置信度.

### 3.2 在大地震减震作用研究中的应用

大地震的减震作用有时也称作免疫性, 意指大地震发生以后, 在相当长的时间内, 周邻地区很少发生 6 级以上的大地震. 这个问题曾引起国内不少研究者的注意<sup>(1)(2)(3)</sup>, 大家的结论基本上均认为有减震作用存在, 但对于存在减震作用的置信度如何则未作过研究. 国外, Veniziano 等<sup>(4)</sup>曾利用华北地区的资料对大地震的减震作用进行了检验, 结果认为大地震的减震作用不明显. 然而在他们的研究中, 存在以下几个问题:

- (1) 震级具有拒绝性, 即认为不能将  $M+0.5$  以上的地震放在  $M$  位置上, 这无疑给随机化检验方法一人为约束, 增强了空区, 与随机化检验方法概念不符;
- (2) 去余震时, 仅考虑不同震级大地震后发生余震时间上的差异, 没有考虑空间范围的差异;
- (3) 随机化检验时, 地震位置的随机化是在潜在震源区间进行的, 没有考虑距离的影响. 为此, 一方面为了引进随机化检验方法, 同时也对大地震的减震作用作一研究, 这里就用随机化检验方法对其进行具体的分析.

与鄂家全等<sup>(1)</sup>相同, 选用资料的范围为东经 106 度至 125 度、北纬 30 度至 42 度. 时间范围自 1500 年至今, 参与随机化检验分析的震级下限为 5.0. 去余震时, 按以下标准删除:

- 1) 时间上与 Veniziano 等<sup>(5)</sup>一样, 即当震级在 5.0—6.9 时,  $\Delta T = 1a$ ; 当震级在 7.0—

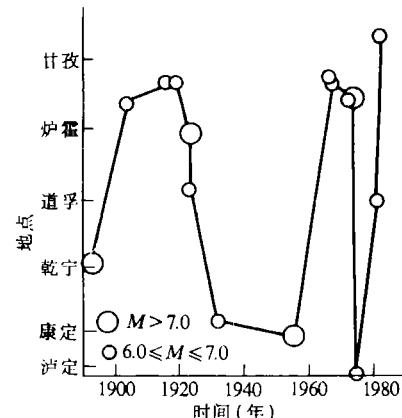


图 3 强震迁移图

<sup>(1)</sup> 张国民等, 1985. 首钢迁安厂址地震危险性评定工作报告.

<sup>(2)</sup> 鄂家全等, 1988. 华北地区大地震的减震作用, 地震区域划分和方法的研究报告.

<sup>(3)</sup> Veniziano *et al.*, 1990. Seismic hazard assessment: spatial and temporal dependent earthquake process, analysis of seismicity in North China, presentation of preliminary results. Risk Engineering, Inc.

<sup>(4)</sup> 鄂家全等, 1988. 华北地区大地震的减震作用, 地震区域划分和方法的研究报告.

<sup>(5)</sup> Veniziano *et al.*, 1990. Seismic hazard assessment: spatial and temporal dependent earthquake process, analysis of seismicity in North China, presentation of preliminary results. Risk Engineering, Inc.

7.9 时,  $\Delta T = 5a$ ; 当震级大于 8 时,  $\Delta T = 10a$ .

2) 空间上, 考虑震源体的大小与  $M$  的关系, 为(邢军、时振梁, 1985)

$$M = 3.46 + 1.84 \lg L$$

检验时, 假定考虑  $M_2$  以上地震(类型 I)对  $M_2$  级以上地震(类型 II)的减震作用, 具体按以下步骤进行:

- 以任意一个类型 I 的地震为圆心, 以  $R$  为半径作一个圆. 若在这个圆内有几个类型 II 的地震, 则取离此类型 I 地震时间为最近的一个 II 型地震; 若无地震, 则取 1990 年减去类型 I 地震的时间;
- 按 a 计算相对于每一个类型 I 的类型 II 地震时间, 同时计算随时间变化的比例;
- 按 a,b 先对观察值进行计算;
- 保持地震时间、空间位置不变, 将地震震级随机化, 同时对于得到的随机化结果, 同样按 a,b 进行计算;
- 计算所有随机化后总体中的 80%、50% 及 20% 分位数.

计算时, 据以上步骤, 对 8 级以上地震对 7 级以上地震及 8 级以上地震对 6.5 级以上地震的减震作用进行了随机化检验. 随机化次数取 200 次, 分别按  $R = 100, 200$  及  $300 \text{ km}$  计算, 所得结果如图 4a,b,c 及图 5a,b,c 所示.

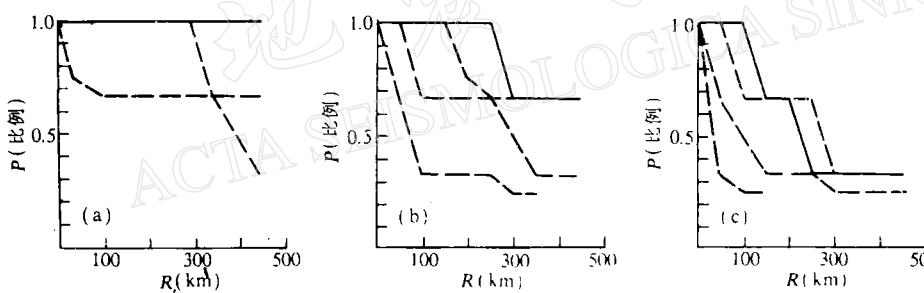


图 4 8 级以上地震对 7 级以上地震减震作用随机化检验结果

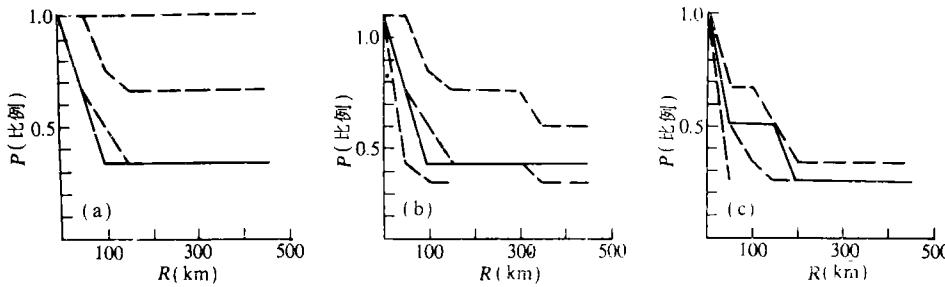


图 5 8 级以上地震对 6.5 级以上地震减震作用随机化检验结果

由图中可以看出, 在 80% 置信度下, 8 级以上地震对 7 级以上地震有比较明显的减震作用, 而对 6.5 级以上地震的减震作用不明显.

## 4 结语

以上利用随机化检验方法对鲜水河断裂带上地震的往返迁移规律及大华北地区大地震的减震作用进行了随机化检验。检验结果表明,鲜水河断裂带上6级以上地震具有较强的往返迁移规律;大华北地区8级以上地震对7级以上地震具有较明显的减震作用,而对6.5级以上地震的减震作用则不显著。

在本研究过程中,曾与时振梁、金严、鄢家全、黄玮琼及高孟潭等作过有益的讨论,在此表示衷心的感谢。

## 参 考 文 献

- 邓天刚、龙德雄,1986. 鲜水河断裂带的基本结构与地震. 地震研究,9, 1, 81—87.
- 林邦慧、陈天长、蒲晓红、刘万琴、彭美宣、张卫平,1986. 鲜水河断裂带强震的破裂过程与地震活动. 地震学报,8, 1—20.
- 王阜,1987. 地震发生的概率模型及其在地震危险性分析中的应用. 国家地震局工程力学所博士论文, 1—59.
- 邢军、时振梁,1985. 震源体与地震影响场,重大工程抗震问题学术讨论会论文集,1—9. 哈尔滨.
- Dwass, M. , 1971. Modified randomization tests for nonparametric hypothesis. *Ann. Math. Stat.* , 28, 1, 181—187.
- Edgington, E. S. , 1987. *Randomization Test* , 1—341. Marcel Dekker, Inc. , New York.
- Erfon, B. , 1979. Bootstrap methods: another look at the Jackknife. *Ann. Stat.* , 7, 1, 1—26.
- Edgington, E. S. , 1969. Approximate randomization tests. *Jour. Psych.* , 72, 1, 143—149.