中山大学学报（自然科学版）
第32卷 第1期 1993年
ACTA SCIENTIARUM NATURALIUM UNIVERSITATIS SUNYATSENI
Vol.32 No.1 1993

洪水风险分析的簇生过程模型*

邓永录 徐宗学
（中山大学数学系） （清华大学水利水电工程系）

摘 要 对较小的洪水作风险分析时，由流量过程线截取得到的点过程现实呈现成丛性．为了适应这一特点，试图利用簇生点过程模型进行洪水风险分析．在本文中选用了 Neyman-Scott 过程模型，并把它用于长江宜昌水文站1963—1984年的（阈值$Q_b=45000\mathrm{m}^3/\mathrm{s}$）较小洪水风险分析．

关键词 洪水风险，簇生过程，Neyman-Scott过程

传统的洪水风险分析模型都是基于频率分析方法，一般只能以最大值法选样，难以利用无确定量值的洪水历史资料．为了克服这些缺陷并进一步把洪水风险分析的研究推向深入，近几年来我们把随机点过程理论和统计分析方法应用于洪水风险分析进行了探讨和研究，在文[1]，[2]中分别对洪水风险分析的泊松过程模型和更新过程模型作了较详细的讨论，并取得较为满意的结果．

点过程模型的应用范围是很广的．例如，文[3，4]中利用点过程模型对地震的发生进行了有益的研究，使用的方法和思想原则上也适用于其它领域的应用．

考虑到在水文过程中每一次洪峰出现在时间轴上也有局部化的特点(图1)，我们结合位于长江中上游的宜昌水文站的历年洪水观测数据对随机点过程模型在洪水风险分析的应用开展了研究．在[2]中首先尝试用一类简单的点过程——齐次泊松过程作模型，以50000m³/s作阈值，从宜昌站100 多年的实测资料中选取超过阈值的洪峰作洪水风险分

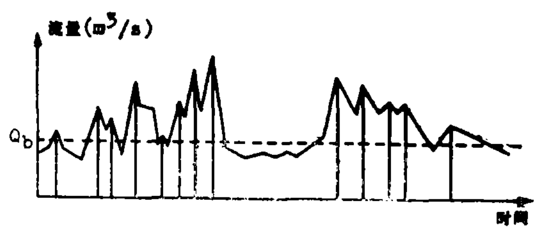

图1 一个典型的流量过程线及对应的洪峰点过程现实
Fig.1 A typical continuous streamflow hydrograph and associated realization of flood peak point process

本文1991年6月26日收到
* 高等学校博士学科点专项科研基金资助

析，并得到较满意的结果。根据水文学家的看法，特大洪水的发生一般不是无后效的。因此，以无后效性为重要特征的泊松过程模型对于特大洪水的研究就不十分合适。于是我们接着在〔1〕中进一步利用更新过程模型对以71100m³/s 为阈值的宜昌站千余年的洪水历史资料作统计分析。由于更新过程比齐次泊松过程更一般。利用这种模型从理论上应能得到更满意的结果。

各次洪水的发生相互影响除了表现为后效性外，还可以其它形式出现，簇生现象(或者说成丛性)就是一种较常见的形式。因为洪水是由降水产生的地面迳流引起的。天气情况的(长期或中期的)变化常常呈现降水较多、较集中的时期和降水较少的时期交替出现的模式(这可以从宏观的水量平衡得到合理的解释)。所以，当我们取较小的洪峰阈值Q_b时，就会出现有较多洪峰密集出现的时段和洪峰稀少的时段交替的现象。例如，我们对宜昌站从1963～1984年的实测洪水资料以45000m³/s 为阈值截取到的一个点过程的现实如图2所示。从中可看出存在某种程度的成丛性。为了更好反映这种特性，本文将利用簇生过程模型进行分析研究。

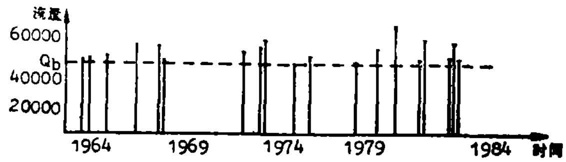

图2 宜昌站1963—1984年的洪峰点过程的现实(阈值$Q_b=45000\mathrm{m}^3/\mathrm{s}$)

Fig.2 The realization of the flood peak point process at Yichang station from 1963—1984 ($Q_b=45000\mathrm{m}^3/\mathrm{s}$)

一个点过程称做簇生过程，如果它由如下方式构成：①有一个簇生中心过程作为基本(第一级)过程；②每一簇生中心产生一个从属(第二级)过程。这样，所有簇(即从属过程)的叠合就是簇生过程。为了便于研究，一般假设各簇的点分布相互独立，而且与簇生中心过程独立。

Neyman和Scott[5]在研究宇宙星体分布时提出了一类簇生过程，人们称之为Neyman-Scott过程。这类过程进一步满足以下条件：①簇生中心过程是齐次泊松过程，设其强度是μ。②簇的大小(即点数)U有由概率母函数$H(z)=\sum_k \pi_k z^k$确定的离散分布。③当给定簇的大小时，簇中各点和中心的距离是相互独立同分布的随机变量，其分布函数是$A(t)$。

簇生过程由两级过程构成，其统计结构是复杂的。研究这类过程的一种重要工具是概率母泛函。点过程$\{N(t)\}$的概率母泛函定义为

$$G[\xi]=E\{\exp[\int \log\xi(t)\mathrm{d}N(t)]\}=E\{\prod_i \xi(t_i)\} \quad (1)$$

其中$\xi(t)$属于某一适当的函数类，例如，满足$0\leqslant\xi(t)\leqslant 1$且在某一有限区间外恒等于1的有界变差函数类。$\{t_i\}$是点过程$\{N(t)\}$的点发生时间序列。我们约定当$\xi(t)$在某集合$A$上恒等于零时，若$N(A)$不为零则令(1)式中的指数函数等于零，若$N(A)$为零则令其等于1。应当指出，概率母泛函可以看作是有限维随机向量的概率母函数概念在无穷

多个随机变量集合的情形的自然推广，它保有概率母函数的某些重要性质（如卷积定理）。还要指出，若令

$$\xi(t)=\begin{cases}1 & t\notin(a,b]\\ z & t\in(a,b]\end{cases} \tag{2}$$

其中z在[0，1]中取值，则(1)式简约为随机变量$N(a,b]$——点过程$\{N(t)\}$在区间$(a,b]$中的点数的概率母函数，特别地，若在(2)式中取$z=0$则有

$$G[\xi]=P(N(a,b]=0) \tag{3}$$

令$G_1[\xi]$是簇生过程$\{N(t)\}$的簇生中心过程的概率母泛函，$G_2[\xi|y]$是当给定簇生中心位于y时由这中心产生的从属过程的概率母泛函。这时，合成过程$\{N(t)\}$的概率母泛函由

$$G\{\xi\}=G_1[G_2[\xi|y]] \tag{4}$$

给出。对于Neyman-Scott过程，

$$G_1[\xi]=\exp\{\mu\int[\xi(t)-1]\,dt\} \tag{5}$$

和

$$G_2[\xi|y]=H\{\int\xi(x)d_xA(x-y)\}$$

从而有

$$G[\xi]=\exp\{\mu\int[H(\int\xi(x)d_xA(x-y))-1]dy\} \tag{6}$$

故由(2)，(3)式知

$$\begin{aligned}P(N(0,t]=0)&=\exp\left\{\mu\int[H\left(\int_{-\infty}^{0}d_xA(x-y)+\int_t^{\infty}d_xA(x-y)-1]\right)dy\right\}\\&=\exp\left\{\mu\int[H\left(1-\int_0^t d_xA(x-y)\right)-1]dy\right\}\end{aligned} \tag{9}$$

这给出从观测原点算起长为t的时间区间内没有超过阈值Q_b的洪水发生的概率。于是，在这时间区间内至少有一次超过阈值Q_b的洪水发生的概率，亦即在这区间的洪水风险率

$$\begin{aligned}R(t)&=1-P(N(0,t]=0)\\&=1-\exp\left\{\mu\int[H\left(1-\int_0^t d_xA(x-y)\right)-1]dy\right\}\end{aligned} \tag{8}$$

如果再假设：(i)簇的大小U遵从参数为θ的泊松分布，即

$$\pi_k=P(U=k)=\frac{\theta^k}{k!}e^{-\theta}\qquad k=0,1,2,\cdots.$$

于是有

$$H(z)=\exp\{\theta(z-1)\} \tag{9}$$

(ii)任一点到它所在簇中心的距离遵从双边指数分布，即有密度

$$\lambda(\gamma)=\frac{1}{2}\eta\exp(-\eta|\gamma|),\qquad -\infty<\gamma<\infty \tag{10}$$

将(9)、(10)式代入(8)式得

$$R(t)=1-\exp\left\{\mu\int\left\{\exp\left[-\frac{\theta}{2}\left(\int_0^{\min(t,y)}\eta e^{-\eta(x-y)}\mathrm{d}x+\int_{\min(t,y)}^t\eta e^{-\eta(x-y)}\mathrm{d}x\right)\right]-1\right\}\mathrm{d}y\right\} \tag{11}$$

在实际应用中要求出洪水风险率(11)，必须解决以下两个问题：第一，利用观测数据估计未知参数μ，θ和η。第二，对所得的估计作拟合优度检验。仍然由于簇生过程结构的复杂性，我们需要先作如下的简化假设：每一簇的所有点都集中在簇中心，于是就能先对μ和θ作出估计。注意这时实际上是把Neyman-Scott过程看作是广义齐次泊松过程，其基本过程是强度为μ的齐次泊松过程，而在由基本过程确定的每一个点发生时刻的点数则遵从参数为θ的泊松分布。对于这样的过程，在$(0,t]$区间中点数$N(0,t]$的概率母函数是

$$K(z)=\exp\{\mu t\{\exp[\theta(z-1)]-1\}\} \tag{12}$$

展成幂级数得

$$K(z)=e^{-\mu t}\exp\{\mu te^{-\theta}e^{\theta z}\}$$

$$=e^{-\mu t}\sum_{k=0}^{\infty}\{[(\mu t)^k\exp(-\theta k)]/k!\}\sum_{j=0}^{\infty}[(k\theta z)^j/j!] \tag{13}$$

将(13)与$K(z)=\sum_{j=0}^{\infty}p_jz^j$比较即得

$$p_j=P(N(0,t]=j)=e^{-\mu t}\frac{\theta^j}{j!}\sum_{k=0}^{\infty}\frac{(\mu t)^k\exp(-\theta k)k^j}{k!},\quad j=0,1,\cdots \tag{14}$$

将(12)式对z求导得

$$K'(z)=\mu t\theta\exp\{\mu t[\exp[\theta(z-1)]-1]+\theta)z-1)\}$$

$$K''(z)=\mu t\theta^2\exp\{\mu t[\exp[\theta(z-1)]-1]+\theta(z-1)\}\{\mu t[\exp[\theta(x-1)]+1\}$$

于是

$$EN(0,t]=K'(1)=\mu t\theta \tag{15}$$

$$VarN(0,t]=E[N(0,t]]^2-[EN(0,t]]^2$$

$$=K''(1)+K'(1)-[K'(1)]^2=\mu t\theta(\theta+1) \tag{16}$$

联合(15)，(16)式可解出

$$\theta=\frac{VarN(0,t]}{EN(0,t]}-1 \tag{17}$$

$$\mu=\frac{EN(0,t]}{\theta t} \tag{18}$$

基于(17)，(18)式，我们可以根据观测数据估计θ和μ，具体做法如下：将观测期分为K个等长为t的小区间，这里K是一个适当的正整数。在每个簇的所有点都发生在簇中心的假设下，对每个长为t的小区间中的点计数，并记第i个这样的区间中的点数为n_i，$i=1,\cdots,K$。令

$$\bar{n}=\sum_{i=1}^{K}\frac{n_i}{K} \tag{19}$$

和

$$S^2=\frac{1}{K-1}\sum_{i=1}^{K}(n_i-\bar{n})^2 \tag{20}$$

分别是$N(0,t]$的样本均值和样本方差。当分别取$\bar{n}$和S^2作$EN(0,t]$和$\mathrm{Var}N(0,t]$的估计时，由(17)、(18)式即可得到参数θ和μ的估计

$$\tilde{\theta}=\frac{S^2}{\bar{n}}-1 \tag{21}$$

$$\tilde{\mu}=\frac{\bar{n}}{\tilde{\theta}t} \tag{22}$$

利用χ^2检验作参数θ和μ的拟合优度检验。令m_j表示恰含j个点的区间数目($j=0,1,\cdots,K=\sum_{j=1}^{\infty}m_j$)，则当假设$\theta=\tilde{\theta}$和$\mu=\tilde{\mu}$真确时，统计量

$$\chi^2=\sum_{j=0}^{J}\frac{(m_j-p_jK)^2}{p_jK}+\frac{\left[\left(1-\sum_{j=0}^{J}p_j\right)K\right]^2}{\left(1-\sum_{j=0}^{J}p_j\right)K} \tag{23}$$

渐近地有自由度为J的χ^2分布，这里p_j是利用(14)式并令$\theta=\tilde{\theta}$和$\mu=\tilde{\mu}$算出，J是满足$m_J>0$和$m_{J+k}=0$（对所有$k\geqslant1$）的整数。若取显著性水平为α，对应的临界值为$\chi^2(\alpha,J)$，则当$\chi^2\leqslant\chi^2(\alpha,J)$时接受假设$H_0$：观测数据来自具有分布参数$\tilde{\theta}$和$\tilde{\eta}$的总体。

余下要估计点和簇中心距离的分布参数η。首先按递增的次序把观测到的距离数据（按绝对值）重排为$r_1\leqslant r_2\leqslant\cdots\leqslant r_L$并计算

$$u(i)=\log\left(\frac{L+1}{L+1-i}\right),\quad i=1,2,\cdots,L \tag{24}$$

这里L是观测数据的数目。于是，η的一个基于最小二乘法原理的估计由

$$\tilde{\eta}=\left[\sum_{i=1}^{L}r_iu(i)\right]\Big/\sum_{i=1}^{L}r_i^2 \tag{25}$$

给出。

最后，我们利用Anderson-Darling的W^2-检验对估计$\tilde{\eta}$作拟合优度检验如下：检验的统计量是

$$W^2=-L-\frac{1}{L}\sum_{i=1}^{L}\{(2i-1)\log z(i)+(2L+1-2i)\log[1-z(i)]\} \tag{26}$$

在这里$z(i)$由下式

$$z(i)=1-\exp\{-\tilde{\eta}r_i\} \tag{27}$$

算出。若选定显著性水平α，对应的临界值是$W^{2*}(\alpha)$，则当由观测数据算出的$W^2\leqslant W^{2*}(\alpha)$时接受假设$H_0$：观测数据来自具有参数$\tilde{\eta}$的总体。

有了参数θ，μ和η的估计，我们就可利用(11)式计算洪水风险率$R(t)$。

利用上面的Neyman-Scott模型对宜昌站1963～1984年超过阈值$Q_b=45000\mathrm{m^3/s}$的

洪峰点过程(图2)进行研究和分析。在分别用(21)、(22)和(25)式求得参数估计$\tilde{\theta}$，$\tilde{\mu}$和$\tilde{\eta}$之后，由于观测数据序列还不够长，我们没有作χ^2检验而只作了W^2检验。取显著性水平$\alpha=5\%$，对应的临界值$W^{2*}(\alpha)=2.492$，而由(26)，(27)式算出的$W^2=1.224\leqslant W^{2*}(\alpha)$，故可认为观测数据来自具有参数$\eta=\tilde{\eta}$的总体。最后，利用(11)式算出的对应于不同的$t$值的洪水风险率如下：

t(年)	0	0.2	0.3	0.8	1	1.5	2	2.5	3
风险率$R(t)$	0	0.385	0.7032	0.8566	0.9116	0.9735	0.9921	0.9976	1

应当指出，我们在前面的讨论中对簇生过程模型作了一些简化的假设，这样做显然会对模型的适用范围有所限制，在理论上也带来一些缺陷。为了克服这些不足，今后需要作进一步的研究。但是，对于存在成丛现象的情形，本文考虑的簇生过程模型无疑是比泊松过程模型和更新过程模型进了一步，所做的统计检验也表明采用这种模型是比较合适的。

参考文献

1 李景玉，徐宗学。洪水风险率Poisson模型应用分析。数理统计与应用概率，1988(3)：392～401

2 邓永录，徐宗学。洪水风险率分析的更新过程模型及其应用。水电能源科学，1989(7)：226～232

3 Vere-Jones D. Stochastic models for earthquake occurrence. J R Stat Soc, 1970, B32: 1～62

4 Vere-Jones D, Deng Yonglu. A point process analysis of historical earthquakes from North China, Earthquake Research in China 1989(2):165～181

5 Neyman J E, Scott E L. A statistical approach to problems of cosmology. J R Stat Soc, 1958, B20: 1～43

Cluster Process Model For Flood Risk Analysis

Deng Yonglu Xu Zhongxue*

Abstract In practice we also have to consider the case of small floods which exceed a low critical value of flood Q_b. In that case there are many more flood peaks exceeding Q_b, and the peak exceedence counts may then present some degree of clustering which would necessitate a model other than those in [1],[2]. There is a variety of cluster models in the theory of stochastic point processes, in this paper we discuss only Neyman-Scott process model and use it to analyse the small flood risk at Yichang station on the Yangtze river from 1963 to 1984.

Keywords flood risk, cluster process, Neyman-Scott process

* Department of Mathematics, Zhongshan University