岩土工程反分析的最大熵原理

陈 斌[1],刘 宁[2],卓家寿[2]

(1. 南京大学地球科学系,江苏 南京 210093;2. 河海大学土木工程学院,江苏 南京 210098)

摘要:以概率与数理统计理论和现代信息理论为基础,针对目前岩土工程随机反演中常用的极大似然法和贝叶斯法,从随机过程入手,应用最大熵原理推导了这两种方法的准则函数. 它的理论意义在于,从数据信息的角度对极大似然反分析法和贝叶斯反分析法进行了重新认识,比过去仅从数据和算法本身研究反演问题多了一个参照系,使得对传统意义上反分析法中的不确定性研究有了一个新的模式.

关键词:反分析;随机;最大熵原理;极大似然法;贝叶斯法;岩土工程

中图分类号:TU43 **文献标识码**:A **文章编号**:1000-1980(2002)06-0052-04

熵(entropy)是一个历史颇长的概念,19 世纪中叶 Clausius 首先把熵引进热力学. 在 Boltzmann 方程:$S = k\log_a W$ 里,熵是用来描述热力学系统无序程度的量. 作为一个近代的概念,1948 年 Shannon 在创立信息论时,将熵引入作为量度信源不确定性的惟一量. 这个量与热力学和统计力学中的熵的数学形式和物理意义都相近,不过为了明确起见,信息论中的熵有时也称为'信息熵'或'仙农(Shannon)熵'[1].

在岩土工程反问题中,由于受建模目的和先验知识水平的影响,以及对系统的分解方式的差异,系统辨识建模问题是没有惟一解的. 即使是在确定了分解水平的描述模型后,由于状态集合的多样性,也使得同一模型结构存在多解性,用数学语言描述岩土工程反问题属于不适定问题. 对于解决此类问题,最大熵原理是最有效的方法之一,特别是当数据或条件不够的情况时. 最大熵原理就是说,在所有可行解中,满足一定约束条件时,应该选择熵最大的一个. 从熵作为不确定程度的度量来看,此时的解包含的主观成分最少,因而是最客观的.

1 最大熵原理

信息蕴涵于不确定性中,而不确定性在概率论中是用随机事件或随机变量来描述的,因此信息的定量表征必然联系着不确定性的量度[2]. 为了定量测度随机事件的不确定性,Shannon 给出了下面的定义:

设 X 是取有限个值的随机变量

$$p_i = \{X = x_i\} \qquad i = 1,2,\cdots,n \tag{1}$$

即 X 取 x_i 的先验概率为 p_i.

则 X 的熵定义为

$$H(X) = H(p_1,p_2,\cdots,p_n) = -k\sum_{i=1}^{n}(p_i \cdot \log_a p_i) \tag{2}$$

$H(X)$称为信息熵,它度量了随机变量 X 的不确定程度,可以将信息熵理解为物质系统内状态的丰富程度或复杂程度,是一种态函数. 式(2)中常数 k 决定于所选用的单位,选取不同的对数底,熵就有不同的单位. 一般为了数学运算方便,常取 e 为底,单位为 nat(奈特),此时 $k=1$,且除非特别标明,在信息熵引用时,有 $\log = \log_e = \ln$;当对数以 2 为底时,信息单位为 bit(比特);以 10 为底时,信息单位为 Hartley(哈特莱). 而对于连续信源,当 X 的分布用概率密度 $p(x)$(连续函数)来描述时,可用离散信源来逼近,即在式(2)中令 $p_i =$

收稿日期:2001-09-13

基金项目:国家自然科学基金资助项目(59809003);教育部高等学校骨干教师资助项目;教育部博士点专项科研基金

作者简介:陈斌(1964—),男,四川安岳人,博士后,工程力学专业,主要从事岩土工程反分析研究.

$p(x_i)\mathrm{d}x$. 如果我们将式(2)定义的离散信源的信息熵看作绝对值量,则式(2)可写为

$$H_{绝} = \lim_{\mathrm{d}x\to 0}\left\{-\sum_{i=1}^{\infty} p(x_i)\mathrm{d}x \cdot \ln(p(x_i)\mathrm{d}x)\right\} = -\int p(x)\ln(p(x))\mathrm{d}x - \lim_{\mathrm{d}x\to 0}\sum_{i=1}^{\infty}\ln(\mathrm{d}x)p(x_i)\mathrm{d}x \tag{3}$$

考察式(3)最后一项,当 $\mathrm{d}x\to 0$ 时,有 $\sum_{i=1}^{\infty} p(x_i)\mathrm{d}x\to 1$,而 $\ln(\mathrm{d}x)\to -\infty$. 所以式(3)最后一项趋于 ∞. 在此,定义式(3)的倒数第二项为连续信源的熵(相对量)

$$H(X) = -\int p(x)\ln(p(x))\mathrm{d}x \tag{4}$$

由式(3)(4)的定义可知,连续信源的熵并不是它实际输出的绝对熵 $H_{绝}$,而是等于绝对熵减去一个无穷大项的相对值,但这个无穷大项在计算熵的变化问题时,将出现两次,一次为正,一次为负,其大小不变,相互抵消[1]. 因此,在任何包含有熵的差的问题中,使用连续信源熵的定义及式(4),都会得到正确的结论. 而实际上,本文将涉及的都是熵的差. 当然,仍需强调的是连续信源定义的熵是一个比无穷大(∞)大多少的相对量,不是绝对量,而离散信源定义的熵是一个绝对熵,二者是不同的.

使得熵得以广泛应用的一个重要概念是 1957 年 Jaynes 在统计力学中正式提出的最大熵原理(POME, principle of maximum entropy). 热力学第二定律告诉我们,一孤立的物质系统的熵永远不会减少,并将自动加大到它所能达到最大值. 当人们将这种熵增加原理用于非热力学领域时便称之为最大熵原理(或最大熵方法). 从这个意义上来讲,最大熵原理对解所作的选择是' 合乎自然 '的[1,2]. 这里,X 可以是概率场,$p(\cdot)$ 或 p_i 也可以是频率或比例,或所谓的置信度(degree of belief). 但对于连续分布情况下的式(4),由于不像离散分布情况下的式(2)在变量之间是一一对应的变换,所以在求最大熵时,不一定用熵的原始定义,而通过引入测度 $m(x)$,用熵的导出表达式

$$H(p(x)\mid m(x)) = -\int_{-\infty}^{+\infty} p(x)\ln\left(\frac{p(x)}{m(x)}\right)\mathrm{d}x \tag{5}$$

以克服这一困难. $m(x)$ 只要依 $p(x)$ 变化的同一方向变化,就可以保持 H 的值不变.

2　最大熵原理背景下的极大似然函数

设随机变量 $\boldsymbol{u}$ 为观测数据,$p(\boldsymbol{u})$ 为其真实的概率密度函数. 所谓极大似然原理的物理意义是:对一系列参数 $\boldsymbol{x}_k$ 及相应的概率密度函数 $p(\boldsymbol{u}\mid\boldsymbol{x}_k)$ 的可能模型中,设法找到参数估计值 $\hat{\boldsymbol{x}}_{ML}$,使得 $\boldsymbol{u}$ 在 $\hat{\boldsymbol{x}}_{ML}$ 条件下的概率密度函数最大可能地逼近 $\boldsymbol{u}$ 在 $\boldsymbol{x}_0$(真值)条件下的概率密度函数,即应有

$$p(\boldsymbol{u}\mid\hat{\boldsymbol{x}}_{ML}) \xrightarrow{\max} p(\boldsymbol{u}\mid\boldsymbol{x}_0) \tag{6}$$

为此,引入熵的表达式(5),有

$$\begin{aligned} H(p(\boldsymbol{u})\mid p(\boldsymbol{u}\mid\boldsymbol{x}_k)) &= -\int p(\boldsymbol{u})\ln\left(\frac{p(\boldsymbol{u})}{p(\boldsymbol{u}\mid\boldsymbol{x}_k)}\right)\cdot\mathrm{d}\boldsymbol{u} = \\ &\int p(\boldsymbol{u})\ln(p(\boldsymbol{u}\mid\boldsymbol{x}_k))\mathrm{d}\boldsymbol{u} - \int p(\boldsymbol{u})\ln(p(\boldsymbol{u}))\mathrm{d}\boldsymbol{u} \end{aligned} \tag{7}$$

显然,式(7)最后一项与 $\boldsymbol{x}_k$ 无关,为常数. 因此,根据最大熵原理,最合适的模型便是使

$$\int p(\boldsymbol{u})\ln(p(\boldsymbol{u}\mid\boldsymbol{x}_k))\mathrm{d}\boldsymbol{u} = \max \tag{8}$$

的那一个. 又

$$E(\ln(p(\boldsymbol{u}\mid\boldsymbol{x}_k))) = \int p(\boldsymbol{u})\ln(p(\boldsymbol{u}\mid\boldsymbol{x}_k))\mathrm{d}\boldsymbol{u} \tag{9}$$

由式(9)可知,式(8)等价于

$$\bar{l}(\boldsymbol{u}\mid\boldsymbol{x}_k) = E(\ln(p(\boldsymbol{u}\mid\boldsymbol{x}_k))) = \max \tag{10}$$

其中 $\bar{l}(\cdot)$ 为平均对数似然函数.

由于 $E(\ln(p(\boldsymbol{u}\mid\boldsymbol{x}_k)))$ 与 $p(\boldsymbol{u}\mid\boldsymbol{x}_k)$ 或 $\ln(p(\boldsymbol{u}\mid\boldsymbol{x}_k))$ 之间存在单调的函数关系,所以式(8)及(10)等价于

$$L(\boldsymbol{u}\mid\boldsymbol{x}_k) = p(\boldsymbol{u}\mid\boldsymbol{x}_k) = \max \tag{11}$$

$$l(\boldsymbol{u}\mid\boldsymbol{x}_k) = \ln(p(\boldsymbol{u}\mid\boldsymbol{x}_k)) = \max \tag{12}$$

其中 $L(\cdot)$ 及 $l(\cdot)$ 分别为所谓的似然函数及对数似然函数. 式(11)及(12)即为极大似然原理的数学表达式,或称之为满足式(11)或(12)的模型即为最‘好’的模型. 由上述推论,本文为极大似然函数法给出了一个最大熵原理背景下的新的解释,同时也为这一方法的应用赋予了新的含义,开拓了一个新的空间.

在岩土工程中,极大似然法是用于量测误差分布已知的情形下的参数估计,其前提条件为:误差是可加的并具有零均值和已知方差的正态分布随机变量;参数 x 是确定性的量,而不是随机变量. 即设

$$\boldsymbol{u}^* = \boldsymbol{u}(\boldsymbol{x}) + \boldsymbol{\varepsilon}_u \tag{13}$$

$$\boldsymbol{\varepsilon}_u \sim N(\boldsymbol{0}, \mathrm{cov}(\boldsymbol{\varepsilon}_u)) \tag{14}$$

式中:$\boldsymbol{\varepsilon}_u$——量测误差列阵;$\mathrm{cov}(\cdot)$——协方差阵;$\boldsymbol{u}^*$,$\boldsymbol{u}$——观测数据及计算值列阵.

则由式(11)及(12)似然函数为

$$L(\boldsymbol{u}|\boldsymbol{x}) = p(\boldsymbol{u}|\boldsymbol{x}) = (2\pi)^{-\frac{m}{2}}|\mathrm{cov}(\boldsymbol{\varepsilon}_u)|^{-\frac{1}{2}} \cdot \exp\left(-\frac{1}{2}(\boldsymbol{u}^* - \boldsymbol{u})^{\mathrm{T}}\mathrm{cov}^{-1}(\boldsymbol{\varepsilon}_u)(\boldsymbol{u}^* - \boldsymbol{u})\right) \tag{15}$$

式中 m 为测点总数.

对数似然函数为

$$l(\boldsymbol{u}|\boldsymbol{x}) = \ln(p(\boldsymbol{u}|\boldsymbol{x})) = -\frac{1}{2}[m \cdot \ln(2\pi) + \ln(|\mathrm{cov}(\boldsymbol{\varepsilon}_u)|) + (\boldsymbol{u}^* - \boldsymbol{u})^{\mathrm{T}} \cdot \mathrm{cov}^{-1}(\boldsymbol{\varepsilon}_u) \cdot (\boldsymbol{u}^* - \boldsymbol{u})] \tag{16}$$

观察(15)(16)式,由于只有 $u(x)$ 与参数 x 有关,因此使式(15)(16)极大似然的 $\hat{x}$ 等价于使式中的 $(\boldsymbol{u}^* - \boldsymbol{u})^{\mathrm{T}} \cdot \mathrm{cov}^{-1}(\boldsymbol{\varepsilon}_u)(\boldsymbol{u}^* - \boldsymbol{u})$ 极小,所以工程中常将

$$J = (\boldsymbol{u}^* - \boldsymbol{u})^{\mathrm{T}} \cdot \mathrm{cov}^{-1}(\boldsymbol{\varepsilon}_u)(\boldsymbol{u}^* - \boldsymbol{u}) \tag{17}$$

作为极大似然法的准则函数.

3　最大熵原理背景下的贝叶斯模型

贝叶斯方法的原理是,把参数 x 看作具有某种验前概率密度 $p(x)$ 的随机变量,然后通过观测与该函数有关联的其他变量,设法从输入输出数据中提取关于参数 x 的信息. 而后者可以归结为由参数 x 的验后概率密度函数 $p(x|\boldsymbol{u})$ 的计算,来推断这个参数. 根据推断参数的方法,贝叶斯参数估计又分为极大验后参数估计方法和条件期望参数估计方法.

由概率论的基本概念有

$$p(\boldsymbol{x}|\boldsymbol{u}) = \frac{p(\boldsymbol{u}|\boldsymbol{x})p(\boldsymbol{x})}{p(\boldsymbol{u})} \tag{18}$$

极大验后参数估计方法是以使验后概率密度函数 $p(x|\boldsymbol{u})$ 达到极大值为估计准则,将所对应的参数值作为估计值,记作 $\hat{x}_{MP}$. 这就意味着在数据 u 的条件下,参数 x 落在邻域内的概率比落在其他邻域内的概率要大. 条件期望参数估计方法则是直接以参数 x 的条件数学期望作为估计值,即用随机变量的均值作为它的估计值. 不过由于条件数学期望

$$\hat{\boldsymbol{x}} = E(\boldsymbol{x}|\boldsymbol{u}) = \int_{-\infty}^{+\infty} \boldsymbol{x} \cdot p(\boldsymbol{x}|\boldsymbol{u})\mathrm{d}\boldsymbol{x} \tag{19}$$

在计算上难度较大,因此工程上多采用极大验后参数估计方法. 本文所讨论的贝叶斯方法也是指极大验后参数估计方法. 由贝叶斯方法的基本思想,引入熵表达式(5),有

$$H[p(\boldsymbol{x})|p(\boldsymbol{x}|\boldsymbol{u})] = -\int p(\boldsymbol{x})\ln\left(\frac{p(\boldsymbol{x})}{p(\boldsymbol{x}|\boldsymbol{u})}\right)\mathrm{d}\boldsymbol{x} = \int p(\boldsymbol{x})\ln(p(\boldsymbol{x}|\boldsymbol{u}))\mathrm{d}\boldsymbol{x} - \int p(\boldsymbol{x})\ln(p(\boldsymbol{x}))\mathrm{d}\boldsymbol{x} \tag{20}$$

由于验前概率密度 $p(x)$ 已知,且在求 $p(x|\boldsymbol{u})$ 时,x 是作为常量的. 所以式(20)最后一项为常数. 因此,根据最大熵原理,最合适的模型应是使

$$\int p(\boldsymbol{x})\ln(p(\boldsymbol{x}|\boldsymbol{u}))\mathrm{d}\boldsymbol{x} = \max \tag{21}$$

的那一个. 同理可知式(21)等价于

$$p(\boldsymbol{x}|\boldsymbol{u}) = \max \tag{22}$$

自然,由极大验后参数估计的思想,相应于式(21)(22)成立的参数 $\hat{x}_{MP}$ 即是待辨识参数的最‘好’估计值. 换句话说,通过最大熵原理,极大验后参数估计方法在此得到了进一步的论证.

取前提条件与式(13)(14)相同的前提条件,且增加参数 x 为随机变量.即设

$$\boldsymbol{x} \sim N(\boldsymbol{\mu}_x, \mathrm{cov}(\boldsymbol{x})) \tag{23}$$

式中 $\boldsymbol{\mu}_x$ 为参数 x 的均值.同时假设

$$\mathrm{cov}(\boldsymbol{x}, \boldsymbol{\varepsilon}_u) = \boldsymbol{0} \tag{24}$$

式(24)表明 x 与 $\boldsymbol{\varepsilon}_u$ 不相关.这里 $\mathrm{cov}(\boldsymbol{\varepsilon}_u)$,$\boldsymbol{\mu}_x$,$\mathrm{cov}(\boldsymbol{x})$均为已知矩阵.则有

$$p(\boldsymbol{u}|\boldsymbol{x}) = (2\pi)^{-\frac{m}{2}} |\mathrm{cov}(\boldsymbol{\varepsilon}_u)|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\boldsymbol{u}^* - \boldsymbol{u})^{\mathrm{T}} \mathrm{cov}^{-1}(\boldsymbol{\varepsilon}_u)(\boldsymbol{u}^* - \boldsymbol{u})\right) \tag{25}$$

$$p(\boldsymbol{x}) = (2\pi)^{-\frac{n}{2}} |\mathrm{cov}(\boldsymbol{x})|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\boldsymbol{x}^* - \boldsymbol{\mu}_x)^{\mathrm{T}} \mathrm{cov}^{-1}(\boldsymbol{x})(\boldsymbol{x}^* - \boldsymbol{\mu}_x)\right) \tag{26}$$

式中 n 为参数 x 的维数.由于 $p(u)$与参数 x 无关,故将式(25)(26)代入式(18)有

$$p(\boldsymbol{x}|\boldsymbol{u}) = \frac{p(\boldsymbol{u}|\boldsymbol{x}) \cdot p(\boldsymbol{x})}{p(\boldsymbol{u})} = \mathrm{const} \cdot \exp\left\{-\frac{1}{2}\left[(\boldsymbol{u}^* - \boldsymbol{u})^{\mathrm{T}} \mathrm{cov}^{-1}(\boldsymbol{\varepsilon}_u)(\boldsymbol{u}^* - \boldsymbol{u}) + (\boldsymbol{x} - \boldsymbol{\mu}_x)^{\mathrm{T}} \mathrm{cov}^{-1}(\boldsymbol{x})(\boldsymbol{x} - \boldsymbol{\mu}_x)\right]\right\} \tag{27}$$

式中 const 为与参数 x 的变化无关的常量.

求式(27)的极大值,等价于求

$$J = (\boldsymbol{u}^* - \boldsymbol{u})^{\mathrm{T}} \mathrm{cov}^{-1}(\boldsymbol{\varepsilon}_u)(\boldsymbol{u}^* - \boldsymbol{u}) + (\boldsymbol{x} - \boldsymbol{\mu}_x)^{\mathrm{T}} \mathrm{cov}^{-1}(\boldsymbol{x})(\boldsymbol{x} - \boldsymbol{\mu}_x) \tag{28}$$

的极小值.式(28)即为岩土工程中常用的贝叶斯反分析法的准则函数.

4　小　　结

随着人们对岩土工程研究的深入,反分析方法已成为目前解决岩土问题的主要方法之一.特别是进入20 世纪 90 年代以来,运用随机系统论的方法进行岩土工程反演分析,建立非确定性的模型研究已日益引起国内外学者的关注[3].笔者基于概率与数理统计理论和现代信息理论,应用最大熵原理推导了岩土工程随机反分析中常用的极大似然反分析法和贝叶斯反分析法的准则函数,它的理论意义在于,从数据信息的角度认识了极大似然反分析和贝叶斯反分析的过程,比过去仅从数据和算法本身研究反演问题多了一个参照系,使得对传统意义上反分析法中的不确定性研究有了一个新的模式.

参考文献:

[1] 吴乃龙,袁素云.最大熵方法[M].长沙:湖南科学技术出版社,1991.5~158.
[2] 孟庆生.信息论[M].西安:西安交通大学出版社,1986.20~80.
[3] 陈斌,卓家寿,刘宁.岩土工程反分析的非确定性模型研究与发展[J].水利水电科技进展,2001,21(5):5~8.

Principle of Maximum Entropy for back analysis in geotechnical engineering

CHEN Bin[1], LIU Ning[2], ZHUO Jia-shou[2]
(1. *Department of Earth Sciences*, *Nanjing Univ.*, *Nanjing* 210093, *China*;
2. *College of Civil Engineering*, *Hohai Univ.*, *Nanjing* 210098, *China*)

Abstract: Based on the modern information theory, two objective functions of the maximum likelihood method and Bayes' Method are deduced with random process and the Principle of Maximum Entropy. The theoretical significance of the deduction is that the maximum likelihood method and Bayes' Method are reconsidered according to the data information, and that one more reference is added to the back analysis which was performed only based on data and algorithms, thus, providing a new method for research of indeterminateness of back analysis.

Key words: back analysis; random; Principle of Maximum Entropy; maximum likelihood method; Bayes' Method; geotechnical engineering