岩土参数概率分布统计意义上的优化分析

李小勇[1] 白晓红[2] 谢康和[1]
(1 浙江大学土木工程系, 杭州 310027)(2 太原理工大学土木系, 太原 030024)

【摘 要】 通过实验数据的可靠性检验、概率模型有限比较法拟合优度检验和分布参数的推广贝叶斯法估计, 来实现岩土参数概率分布在统计意义上的优化分析。

【关键词】 岩土参数;概率分布;有限比较法;推广贝叶斯法

【中图分类号】 TU521.3

【Abstract】 The paper, which is based on the reliability test of experiment data, the optimum fit of probabilistic model by finite contrast method and the estimation of distribution parameters by extended Bayesian method, brings about the statistically optimal analysis of probabilistic dictribution about geotechnical parameters.

【Key words】 geotechnical parameters ;probabilistic distribution ;finite contrast method ;extended Bayesian method

0 引 言

岩土参数具有显著的不确定性, 在概率分析中常作为随机变量来对待。在评价岩土工程的可靠性、估计失效概率时, 需要了解岩土参数的概率分布特征。岩土参数概率模型和分布参数对岩土工程可靠性分析结果和精度产生直接的影响, 因此岩土参数概率分布的统计分析必须反映土性的实际情况。在岩土工程可靠性设计过程中, 常会遇到由于土性参数测试数据少而无法用传统的统计学理论确定其概率分布的问题, 因此研究小子样概率分布分析方法, 具有重要的现实意义。本文通过实验数据可靠性检验、概率模型有限比较法拟合优度检验和分布参数推广贝叶斯法估计, 来实现岩土参数概率分布在统计意义上的优化分析。文后给出了应用实例。

1 土工实验数据的可靠性检验

土工实验数据的变异性较大, 其变异性包括土性因有的变异性和实验误差。实验误差按其性质分为三类: 随机误差、系统误差、过失误差。随着实验次数的增加, 随机误差的算术平均值将越来越小, 并逐渐趋近于零。系统误差可以通过一定的方法识别和消除。过失误差一般是由于实验观测系统测错、记错等不正常的原因造成的, 必须消除。

根据抽样理论, 要使一组样本中得到的实验有意义, 必须满足条件: 从母体中取出的样本必须具有代表性以及样本数必须充分。《结构可靠性总原则》(ISO/DIS2394 :1996)规定: 随机变量的概率模型应以有效数据的统计分析为依据, 在可能条件下所有数据应进行校核, 以消除量测误差。因此, 实验数据的可靠性检验内容应包括: 异常实验数据的舍弃、实验数据的相关性检验、实验数据中最小样本数检验、量测误差的消除。

2 概率模型有限比较法拟合优度检验

在数理统计理论中常用的拟合优度检验方法有: 概率纸检验法, χ^2 检验法及 K-S、

作者简介: 李小勇, 1970 年生, 男, 汉族, 山西翼城人, 岩土工程硕士毕业, 现为浙江大学博士生 . 研究方向是岩土工程可行性、地基处理与基础工程。

C－M、A－D 等多种“EDF”检验法。概率纸检验法只适用于正态分布的情况, 检验结果较粗, 在可靠性分析中较少采用。χ^2 检验法是根据皮尔逊定理, 利用当子样充分大时统计量总是渐近地服从 χ^2 分布的特点, 来进行拟合优度检验。各种“EDF”检验法是通过实测子样, 构造一个经验分布函数, 然后计算经验分布函数与假设分布函数的检验偏差来确定是否接受假设。

在拟合优度检验中, 可能犯“弃真”和“取伪”两类错误。在进行小子样检验时, 经常会遇到同时接受多种不同假设的情况, 这必然犯了“取伪”错误。从理论上讲, 这是由于被检样本较少, 检验统计量的判别值不准确, 而犯“弃真”错误率固定, 必然加大“取伪”错误率, 所以会出现同时接受多种不同假设的“取伪”现象。要解决这一问题, 可采取加大子样数, 而大多数岩土参数加大子样数常常是不可能的, 因此只能寻找新的检验方法。

结合实践经验和已有的研究成果, 某一岩土参数只可能是有限的几种概型。虽然在小子样的情况下, 进行拟合优度检验时, 会出现同时接受几种不同假设的情况, 但根据子样与母体的关系, 应该是符合原分布的假设拟合优度最优。因此, 只需对有限的几种可能的分布假设进行统计检验, 并加以分析对比, 就可以确定小子样的最佳概型。对于分布假设 H_{0i} 接受水平 K_{Hi} 定义为检验统计量与标准临界值之比, 取 $K_{opt}=\min\{K_{H0}\}$ 对应的分布假设为最优分布概型。这种以常规检验法为基础, 并能唯一地确定检验结果的拟合优度检验方法, 称为有限比较法[1]。以 χ^2 检验法为基础的有限比较法称 χ^2 比较法。

有限比较法的选用原则[1]: ①当样本数 $n>150$ 时, 可由传统的检验方法唯一地确定最优概型; ②当样本数 $8<n<50$ 时, 可用各种“EDF”比较法来确定最优概型; ③当样本数 $50<n<150$ 时, 可用 χ^2 比较法来确定最优概型。

3 分布参数的推广贝叶斯法估计

贝叶斯(Bayes)定理是随机变量验后概率 $f''(\theta)$ 是验前概率 $f'(\theta)$、似然函数 $L(\theta)$、正规化常数 K 乘积[2]。推广的 Bayes 法是根据 Bayes 统计理论引申而来的, 利用先验分布的信息对分布概型进行初估, 再根据后验信息对后验分布参数重新加以估计, 能较好地避免由于实验数据不足引起的统计参数的偏差, 并确定岩土参数的概率分布参数。

不同类型的概率分布其参数各异, 但是它们往往可以通过测值的均值和方差求得。因此研究测值的均值和方差的估计特别重要。根据 Bayes 公式有:

$$f''(\theta_i)\Delta\theta = KL(\theta_i)f'(\theta_i)\Delta\theta \qquad (1)$$

假定变量后验分布概型与先验分布概型一致, 则分布的均值和方差可近似确定:

$$\mu=\sum_{i=1}^{h}\overline{\theta}_i f''(\theta_i)\Delta\theta \qquad (2)$$

$$\sigma^2=\sum_{i=1}^{h}(\overline{\theta}_i-\mu)^2 f''(\theta_i)\Delta\theta \qquad (3)$$

式中: $\overline{\theta}_i$ 为分段区间 $[\theta_i, \theta_i+\Delta\theta]$ $(i=1, 2, \cdots, h)$ 上 0 的中值; $\Delta\theta$ 为分段间路; h 为区间数。

4 应用实例

本文以太原粉质粘土的压缩指数 C_c 为例, 结合“太原日报社编辑业务楼”工程来进行其概率分布统计上的优化分析。资料来源: 第一套资料是从太原地区现有 20 个建筑工程场地《岩土工程勘察报告》中收集了 58 个数据, 进行可靠性检验舍弃了 3 个指标, 所余的 55 个指标满足统计分析的要求, 列于表 1; 第二套资料是由 C_c-e 间的经验关系式[3] $C_c=-0.084+0.348e$(剩余方差 $\sigma^2{}_c=0.239^2$)间接求得 10 个数据(见表 2); 第三套资料是直接测定的 5 个数据, 0.102, 0.135, 0.166, 0.206, 0.312。

1)概率分布拟合

利用 χ^2 比较法对压缩指数 C_c 进行概型拟合优度检验, 确定其最优概型。根据实践经

表 1 压缩指数 C_c 的统计数据

编号	压缩指数 C_c	编号	压缩指数 C_c
1	0.075	29	0.246
2	0.082	30	0.251
3	0.093	31	0.259
4	0.105	32	0.262
5	0.112	33	0.265
6	0.121	34	0.266
7	0.125	35	0.275
8	0.136	36	0.278
9	0.139	37	0.282
10	0.147	38	0.285
11	0.152	39	0.296
12	0.156	40	0.301
13	0.171	41	0.307
14	0.172	42	0.312
15	0.175	43	0.318
16	0.182	44	0.326
17	0.186	45	0.329
18	0.192	46	0.331
19	0.195	47	0.335
20	0.205	48	0.342
21	0.216	49	0.348
22	0.218	50	0.351
23	0.226	51	0.363
24	0.227	52	0.372
25	0.232	53	0.385
26	0.235	54	0.397
27	0.239	55	0.420
28	0.241		

验和前人研究成果，假设压缩指数 C_c 符合以下分布，H_{01}：正态分布，H_{02}：对数正态分布，H_{03}：β 分布。由 χ^2 检验法的要求，将测值从小到大重新排列，划分成 7 个区段 D_h（h =1，2，…7），并计算各 D_h 区间上的试验频数 N_h 和不同假设下的理论频数 $M_h = nf_h\triangle$ 计算统计量：$\chi^2_{H0i}=\sum_{h=1}^{7}(N_h-M_h)^2/M_h=\sum_{h=1}^{7}d_h(i=1,2,3)$。检验统计量：$\chi^2_{0.95}(4)=9.488$。计算过程及检验结果见表 3 表 4。

根据拟合优度检验的结果，如按常规的 χ^2 检验法，C_c 的概率分布可以接受正态分布、对数正态分布、β 分布三种假设。造成这种现象的原因，是由于样本容量不足，而在实践中样本容量总是有限的。而按 χ^2 比较法只接受 K_{min} 为最优拟合分布，因此压缩指数 C_c 服从正态分布 $N(0.241, 0.088)$。在进行参数估计时可用此结果作为先验分布。

2）分布参数估计

如果随机变量 θ 验前分布 $f'(\theta)$ 为正态分布 $N[\mu', \sigma']$，那么似然函数 $L(\theta)$ 亦为正态分布 $N[\mu^*, \sigma^*]$，其中 $\mu^*=\mu$，$\sigma^*=\sigma/\sqrt{n}$，则验后分布 $f''(\theta)$ 是二个正态分布乘积，它也是一个正态分布 $N[\mu'', \sigma'']$，其 μ'' 和 σ'' 已有坦格（Tang 1984）推导如下：

$$\mu''=[\mu^*(\sigma')^2+\mu'(\sigma^*)^2]/[(\sigma')^2+(\sigma^*)^2] \quad (4)$$

表 2 孔隙比 e 测定值和压缩指数 C_c 的计算值（1～10 号）

	1	2	3	4	5	6	7	8	9	10
孔隙比 e	0.678	0.733	0.968	0.920	0.856	0.710	0.589	0.830	0.767	0.891
压缩指数 C_c	0.152	0.171	0.253	0.236	0.214	0.163	0.121	0.205	0.183	0.226

$$\sigma''=(\sigma'\sigma^*)/[\sigma_x')2+(\sigma_x^*)^2]^{1/2} \quad (5)$$

①以式（4）为验前分布对第二套资料进行估计，已知 $\mu'=0.241$、$\sigma'=0.088$，由表 2 可得 $\mu^*=0.189$，$\sigma^{*2}=0.04^2$。似然函数：$\mu^*=0.189$，$\sigma^{*2}=(1/10)(\sigma^2+\sigma^2_c)=0.077^2$，根据公式（4）和（5）可以计算验后分布参数均值和标准差：$\mu''=0.212$，$\sigma''=0.058$，则验后分布密度为 $N(0.212, 0.058)$，由此可见，验后方差比验前方差和似然方差都要小。

②以 $N(0.212, 0.058)$ 为验前分布，对第三套资料进行估计，已知 $\mu'=0.212$、$\sigma'=0.058$。根据 5 组测值可得：$\mu=0.184$，$\sigma^2=0.081^2$，似然函数均值和方差为：$\mu^*=0.184$，$\sigma^{*2}=0.036^2$。根据公式（4）和（5）可以计算验

表 3　压缩指数 C_c 的概率分布 χ^2 比较法检验

区间号 h	样本区间 D_h	实验频数 N_h/次	H_{01}:正态分布 $\mu_x=0.241,\sigma_x=0.088$ 理论频数 M_h/次	d_h	H_{02}:对数正态分布 $\mu_{\ln x}=-1.502,\sigma_{\ln x}=0.424$ 理论频数 M_h/次	d_h	H_{03}:β 分布 $\alpha=0.415,\beta=0.516$ 理论频数 M_h/次	d_h
1	0.07~0.12	5	4.620	0.031	3.960	0.273	5.720	0.091
2	0.12~0.17	7	6.875	0.002	10.56	1.200	8.580	0.291
3	0.17~0.22	10	10.78	0.056	12.93	0.664	9.185	0.072
4	0.22~0.27	12	12.32	0.008	9.625	0.586	9.955	0.420
5	0.27~0.32	9	10.29	0.161	7.040	0.546	8.855	0.002
6	0.32~0.37	8	6.215	0.513	4.675	2.365	7.865	0.002
7	0.37~0.42	4	3.905	0.002	6.215	0.789	4.840	0.146

表 4　压缩指数 C_c 的概率分布 χ^2 比较法检验结果

项	目	H_{01}:正态分布	H_{02}:对数正态分布	H_{03}:β 分布
计算结果	χ^2_{H0i}	0.773	6.421	1.028
	K_{H0i}	0.081	0.677	0.108
检验结果		压缩指数 c_c 的总体符合正态分布		

后分布参数均值和标准差：$\mu''=0.192$，$\sigma''=0.031$，则验后分布密度函数为：$N(0.192, 0.031)$

计算结果表明该工程粉质粘土地基压缩指数 Cc 服从均值 0.192、标准差 0.031 正态分布，应该认为 Bayes 法估计值能更好的接近总体分布的参数，而工程上 5 组测值的统计值与总体分布参数有一定的偏差，在进行可靠性分析时宜采用前者。

5　结　论

土工实验数据可靠性检验是实现概率分布统计优化的最基本的要求。Bayes 原理为利用地区经验和同类工程经验提供了一条途径，有限比较法应用的关键是要保证所假设的几种可能分布中必须包含有母体分布概型，而岩土参数一般具有地域性特点，因此研究和总结岩土参数地区性的概率分布，具有重要现实意义和指导意义。

通过土工实验数据的可靠性检验、概率模型有限比较法拟合优度检验和分布参数推广贝叶斯法估计可以实现岩土参数概型分布在统计意义上的优化分析。这一分析方法与岩土参数统计的实际过程相一致，因而具有较高的实用价值。

参　考　文　献

1　张博庭．用有限比较法进行拟合优度检验．岩土工程学报，1991，13(6)：84~91

2　包承纲．贝叶斯定理及其在三峡工程中的应用．人民长江，1985(1)：46~52.

3　李小勇．太原粉质粘土工程性质指标的概率统计特征：[学位论文]．太原：太原理工大学，1999

收稿日期：2000-02-21