

基于 Bootstrap 方法确定岩土参数的标准值^①

王晓明^{1,2}, 曹正波¹

(1. 河北省交通规划设计院, 河北 石家庄 050011; 2. 河北工业大学 土木工程学院, 天津 300401)

摘要:提出一种基于 Bootstrap 重抽样技术的岩土参数标准值确定方法, 将置信概率为 95% 对应的重抽样均值作为标准值, 并通过一组实测土体的抗剪强度参数阐述标准值的确定方法和过程。对两个实例进行研究, 分别探讨 Bootstrap 重抽样次数(即 Bootstrap 样本集大小)和原始数据的变异性对标准值精度和收敛性的影响。结果表明, 基于 Bootstrap 方法确定的岩土参数标准值具有较好的收敛性和精度, 当重抽样次数不小于 100 时, Bootstrap 重抽样次数和原始样本的变异性对标准值的收敛性和精度不会产生显著影响。该方法能够有效地处理岩土参数的不确定性问题。

关键词: 岩土参数; Bootstrap 方法; 标准值; Bootstrap 样本; 置信概率

中图分类号:

文献标志码: A

文章编号: 1000-0844(2015)增刊 1-0007-05

DOI: 10.3969/j.issn.1000-0844.2015.02.0563

Determination of the Standard Value of Geotechnical Parameters Using a Bootstrap Method

WANG Xiao-ming^{1,2}, CAO Zheng-bo¹

(1. Hebei Provincial Communications Planning and Design Institute, Shijiazhuang, Hebei 050011, China;

2. School of Civil Engineering, Hebei University of Technology, Tianjin 300401, China)

Abstract: Scientific and rational determination of the representative values of geotechnical parameters constitutes a premise for the design of a safe and economic project. In this paper, an approach based on a Bootstrap resampling technique is proposed to determine the standard values of geotechnical parameters that have a confidence probability of 95% over the Bootstrap sample mean. First, the Bootstrap method is briefly introduced and a set of shear strength parameters is utilized to elaborate the procedures. Two examples were analyzed to assess the influence of the number of Bootstrap simulations (i. e., the size of bootstrap samples) and the variability of original data on the convergence and accuracy of standard values. It was shown that the standard values determined using the proposed Bootstrap method have good convergence and accuracy. When the number of Bootstrap simulations was at least 100, the number of bootstrap simulations and the variability of original data had a negligible impact on the convergence and accuracy of standard values. Therefore, this method is capable of assessing the uncertainty in geotechnical parameters.

Key words: geotechnical parameter; Bootstrap method; the standard value; Bootstrap sample; confidence probability

0 引言

岩土体是经过漫长、复杂的地质历史过程形成的地质材料, 其物理力学性质存在很大程度的不确

定性, 不仅随时间和空间变化, 而且还具有模糊性、知识不完备性和结构特性^[1]。由于岩土体的空间变异性、抽样数量以及试验结果受经费、场地等制约因

① 收稿日期: 2015-04-01

作者简介: 王晓明(1984-)男, 博士, 工程师, 主要从事岩土工程可靠性方面的研究. E-mail: wangxiaoming1984@126.com

素的影响,获得岩土体参数的“真值”是非常困难的。科学合理确定岩土体参数的代表值是岩土工程勘察的重要内容,也是确保工程设计经济安全的前提条件。为此,不少学者针对岩土体参数的取值开展研究,并取得了不少成果^[2-6]。在岩土勘察和设计实践中,应用最多的是《岩土工程勘察规范》(GB 50021-2011)(2009年版)和《建筑地基基础设计》(GB 50007-2011)推荐的标准值计算公式^[7-8],本文简称为规范法。该方法确定的参数标准值是在统计学区间估计理论上得到的关于参数母体平均值置信区间的单侧置信界限值^[7]。为推广该方法在岩土工程实践中的应用,规范采用了标准差折减系数和学生氏函数界限值的近似公式,使得规范中的简化方式与理论公式计算结果存在一定的误差^[8]。朱红霞等^[6]将随机场理论引入抗剪强度指标的统计方法,并考虑了土性指标的相关性,但基于随机场理论的参数确定方法不易被工程技术人员理解和应用,且尚未到达完全实用化的程度。

本文提出一种基于 Bootstrap 重抽样技术的岩土参数标准值确定方法,该方法通过对岩土参数样本进行多次 Bootstrap 重抽样,确定重抽样样本的均值及其分布特征,取置信概率 α 为 95% 对应的重抽样均值作为标准值。通过两个实例对该方法确定的参数标准值与随机实现次数(即 Bootstrap 样本集大小)及原始数据变异性的关系进行探讨。

1 Bootstrap 方法及岩土参数标准值的确定

1.1 Bootstrap 方法及程序实现

Bootstrap 方法是 Efron 于 1979 年提出的一种统计推断方法^[9]。该方法的基本思想是在原始数据的范围内进行可重复的再抽样,重抽样样本容量与原始数据数量保持一致,所得样本称为 Bootstrap 样本。通过多次抽样获得多组 Bootstrap 样本,对 Bootstrap 样本的均值、方差等参数进行统计分析进而推断总体的参数特征。已知某一给定参数的原始样本 $X = \{x_1, x_2, \dots, x_n\}$, 对其进行重新抽样,每一个参数值 x_i 被抽取的概率相等且可以重复抽取,经过 n 次抽样,得到一个 Bootstrap 样本 $Y_j = \{y_{1,j}, y_{2,j}, \dots, y_{n,j}\}$ 。根据获得的 Bootstrap 样本 Y_j , 可以通过式(1)和式(2)计算该样本的平均值 $\mu(Y_j)$ 和方差 $S^2(Y_j)$:

$$\mu(Y_j) = \frac{1}{n} \sum_{i=1}^n y_{i,j} \quad (1)$$

$$S^2(Y_j) = \frac{1}{n-1} \sum_{i=1}^n [y_{i,j} - \mu(Y_j)]^2 \quad (2)$$

将上述过程进行 m 次随机实现,得到包含 m 个 Bootstrap 样本的样本集 $B = \{Y_1, Y_2, \dots, Y_m\}$, 进而可计算出 m 个 Bootstrap 样本的平均值和方差。当选取的 m 值较大时,可以获得 Bootstrap 样本平均值和方差的分布特征,进而对总体均值和方差的置信区间进行分析和确定。Bootstrap 方法特别适用于小样本($n < 30$)参数的估计,且无需考虑参数总体的分布情况^[10-11]。

作者根据 Bootstrap 方法的核心思想,采用 Lisp 语言编制了计算机程序,该程序的基本流程如图 1 所示。从图中可以看出,Bootstrap 方法的计算机实现较为简单,计算机程序仅 20 余行代码。

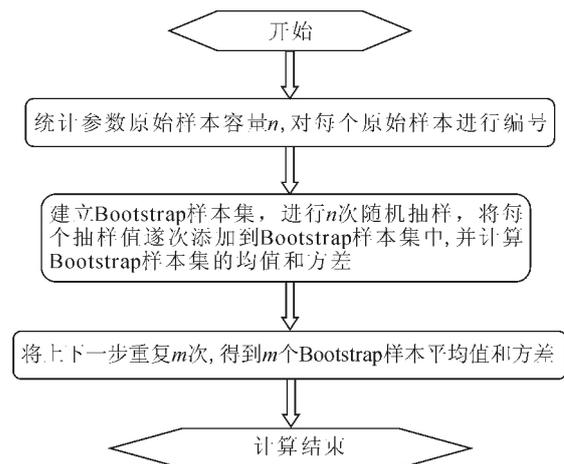


图 1 Bootstrap 程序的基本流程

Fig. 1 Basic procedure of Bootstrap code

2.2 岩土参数标准值的确定

通过一组实测参数样本说明岩土参数标准值的确定方法。土体样本为京沪高速公路某工点经钻探采取的原状粉土,有效样本 16 个。对其进行快剪试验,获得其抗剪强度指标(表 1)。经计算,黏聚力 c 的平均值 μ_c 为 13.06 kPa,变异系数 δ_c 为 0.10,规范法确定的标准值 c_k 为 12.48 kPa;内摩擦角 φ 的平均值 μ_φ 为 23.10°,变异系数 δ_φ 为 0.05,标准值 φ_k 为 22.63°。在参数统计过程中,如遇到离散性较大的数据,可采用“三倍标准差原则”进行剔除。

利用 Bootstrap 程序,对原始样本进行随机抽样,每组 Bootstrap 样本容量为 16,与原始样本容量相等。共进行 50 000 次随机实现,得到 50 000 组 Bootstrap 样本,其 c 、 φ 平均值的分布情况见图 2。Bootstrap 样本集估计值可能受随机抽样次数(即 Bootstrap 样本集容量)的影响,50 000 次的随机实现是比较稳健的做法,有关内容将在第 3 节进行讨论。从图 2 中可以看出,Bootstrap 样本均值近似服

表 1 土体抗剪强度参数

Table 1 Shear strength parameters of soil samples

样本编号	抗剪强度参数		样本编号	抗剪强度参数	
	c/kPa	$\varphi/(\circ)$		c/kPa	$\varphi/(\circ)$
1	14.8	21.7	9	14.0	24.1
2	12.4	23.0	10	13.5	24.4
3	14.0	22.0	11	10.5	22.5
4	11.7	24.1	12	12.1	21.5
5	13.6	23.3	13	12.6	24.2
6	12.1	24.2	14	15.4	24.1
7	12.3	22.2	15	12.6	23.1
8	12.6	23.7	16	14.7	21.5

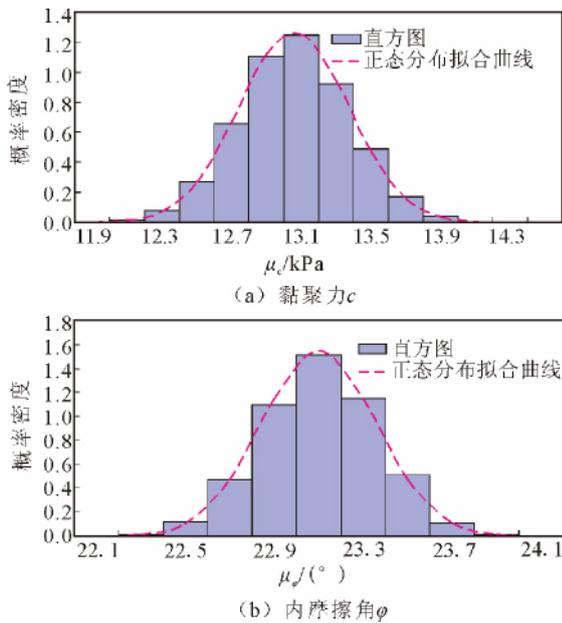


图 2 Bootstrap 样本参数均值概率分布直方图

Fig. 2 Probability distribution of the mean values of Bootstrap sample parameters

从正态分布,采用正态分布曲线对其进行拟合,拟合效果良好。取置信概率 α 为 95% 对应的 Bootstrap 样本均值作为标准值,即取其概率分布的 0.05 分位数,这样就保证 Bootstrap 样本的均值对标准值的超越概率 P 为 95%。对于土体任意参数 Φ ,其标准值 Φ_k 的概率学定义为

$$P(\mu_\Phi \geq \Phi_k) = 95\% \quad (3)$$

式中, μ_Φ 为参数 Φ 的 Bootstrap 样本均值。根据本文提出的标准值确定方法,得到土体抗剪强度标准值 c_k 和 φ_k 分别为 12.54 kPa 和 22.67°,与规范法确定的标准值基本相等,说明本文的方法是切实可行的。本文提出的基于 Bootstrap 方法的参数标准值采用参数平均值置信区间的单侧置信界限值,但与规范法相比,无需考虑岩土参数点标准差和空间均值标准差的相互关系,而且直接根据 Bootstrap 重

抽样均值的概率分布确定参数标准值,确定过程简单直接、容易理解。此外,该方法还可以用于估计岩土参数空间均值标准差,确定其置信区间,为随机场研究和岩土工程可靠性研究提供帮助。确定空间均值标准差及其置信区间的方法与确定标准值的方法类似,本文不再进行详细说明。

3 敏感性分析

3.1 Bootstrap 样本集容量对标准值的影响

根据 Bootstrap 方法确定的岩土参数标准值,收敛性在一定程度上受样本集容量 m 的影响。本节对标准值与 Bootstrap 样本集容量 m 的关系展开研究,以合理确定样本集的容量。对 2.2 节中的土体抗剪强度指标进行多次 Bootstrap 重抽样,获得多个 Bootstrap 样本集,样本集容量 m 分别设为 100, 200, ..., 1 000, 2 000, ..., 10 000, 20 000, ..., 50 000, 标准值大小与 m 的关系见图 3。图 3 反应了标准值随 Bootstrap 样本集容量 m 的变化情况,当容量 m 较小时(如 $m < 1 000$),标准值会呈现一定程度的上下波动,随着抽样次数的增加,样本集容量 m 增大,标准值逐渐趋于稳定。当 $m \geq 8 000$ 时,黏聚力标准值 c_k 不再发生变化,稳定后的 c_k 为 12.538;对于内摩擦角标准值 φ_k ,当 $m \geq 6 000$ 时变为常数,此时 φ_k 为 22.669。然而,应当注意到无论是 c_k 还是 φ_k ,其变化区间非常狭窄, c_k 为 12.50 ~ 12.58, φ_k 为 22.60 ~ 22.68。从满足工程要求的角度, $m = 100$ 也是可以接受的,但考虑数学精度,样本集容量 m 宜大于或等于 8 000。

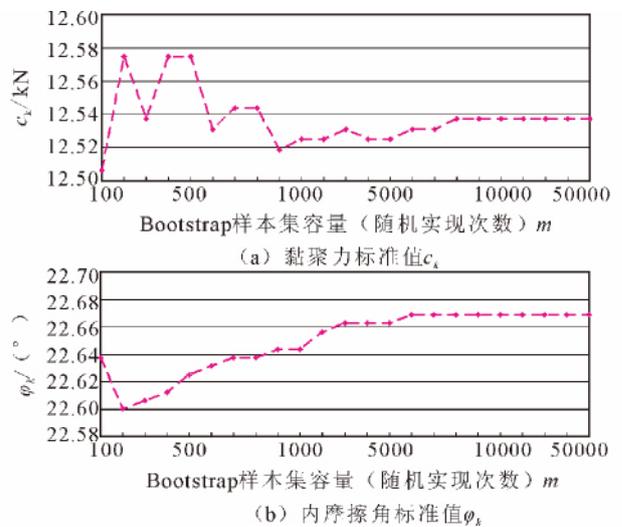


图 3 参数标准值与 Bootstrap 样本集容量关系

Fig. 3 Relationship between the standard value of parameters and Bootstrap sample-set size

3.2 原始样本离散性对标准值的影响

为研究原始样本参数的离散性对确定标准值的影响,对三组岩样饱和单轴抗压强度 R_c 进行 Bootstrap 重抽样,分析变异系数(COV)对确定标准值可能产生的影响。三组岩样取自河北省张涿高速分水岭隧道围岩,每组包含 15 个岩样,根据点荷载试验确定每个岩样的 R_c ,三组岩样 R_c 值的变异系数分别为 0.26, 0.37 和 0.49,具体统计参数见表 2。根据 Duncan^[12]、Baecher 与 Christian^[13] 总结的岩土参数变异系数可知,多数岩土参数的变异系数小于 0.5,所以本例给出的变异系数最大值能涵盖多数岩土参数的变异系数。

表 2 三组岩样 R_c 统计结果

Table 2 Statistical results of R_c for rock samples in three groups

统计参数	第 1 组	第 2 组	第 3 组
平均值/MPa	73.37	71.09	40.72
标准差/MPa	18.74	26.62	20.12
变异系数	0.26	0.37	0.49
最大值/MPa	44.44	26.75	17.00
最小值/MPa	99.76	111.74	89.28
Bootstrap 标准值/MPa	65.70	60.14	32.94
规范法标准值/MPa	64.74	58.83	31.45

对每组岩样的 R_c 值进行 Bootstrap 重抽样,确定每组岩样在不同随机实现次数时的标准值 $R_{c,k}$ (图 4)。由图 4 可知,具有不同变异系数的三组岩样,其饱和单轴抗压强度标准值 $R_{c,k}$ 的变化曲线均近似水平直线,说明当随机实现次数 $m \geq 100$ 时,变异系数并不会显著影响标准值的精度。当变异系数为最大值 0.49 时,对应组别的 $R_{c,k}$ 最小为 32.331,最大为 32.967,两者较为接近。因此当 $m \geq 100$ 时,确定的标准值能够满足工程要求,这与 3.1 节中得到的结论一致。所以原始数据的离散性不会对标准值的精度和收敛性产生显著影响。将该方法与规范法确定的标准值进行比较发现,二者非常接近,相对误差 $< 5\%$,因此 Bootstrap 标准值完全满足工程精度要求。至于说哪种方法确定的标准值更接近真值,还有待进一步研究。

4 结语

本文提出了一种确定岩土参数标准值的新方法。该方法基于 Bootstrap 重抽样技术,通过对岩土参数的原始样本进行多次随机抽样,建立 Bootstrap 样本均值的概率分布模型,将置信概率为 95% 对应的 Bootstrap 样本均值作为标准值。Boot-

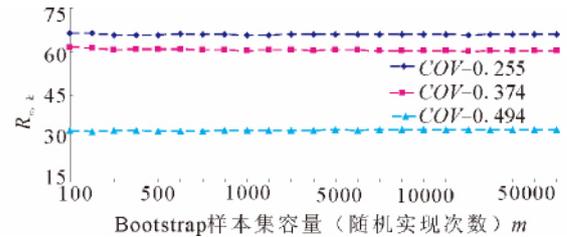


图 4 不同随机实现次数下三组岩样的 $R_{c,k}$

Fig. 4 $R_{c,k}$ for rock samples in three groups with different COV versus different number of simulations

strap 方法能有效确定样本估计量(均值、方差等)的变化区间,尤其适用于小样本参数的估计,且无需掌握参数总体的分布特征。通过对两个实例的分析表明,基于 Bootstrap 方法确定的岩土参数标准值具有较好的收敛性和精度。当随机实现次数不小于 100 时,随机实现次数(即 Bootstrap 样本集容量)和原始样本的变异性对标准值的收敛性和精度不会产生显著影响。从工程角度而言,随机实现次数 $m \geq 100$ 时就能够满足工程需求,但为了保证标准值具有足够的精度以满足其它要求,建议随机实现次数 m 取 5 000~10 000 之间。本文提出的标准值确定方法简单、直接,容易理解,易于程序实现,与规范推荐的方法相比,无需考虑标准差折减系数,具有一定的研究与应用前景。

与其他参数统计方法一样,Bootstrap 方法也是根据已知样本数据推断总体的参数及置信区间,因此,样本数据的代表性对确定岩土参数标准值是至关重要的。原始样本越具有代表性,Bootstrap 估计值的精度和可信度越高。

参考文献 (References)

- [1] 高谦,吴顺川,万海林,等. 土木工程可靠性理论及其应用[M]. 北京:中国建材工业出版社,2007.
- [2] 高大钊. 土的抗剪强度指标的统计方法[J]. 工程勘察,1986,(4):1-5.
- [3] 黄志全,王思敬,李华晔,等. 岩体力学参数取值的置信度及其可靠性[J]. 岩石力学与工程学报,1999,18(1):33-35.
- [4] 熊文林,李胡生. 岩石样本力学参数值的随机-模糊处理方法[J]. 岩土工程学报,1992,14(6):101-108.
- [5] 张润明,郑文棠. 相关型岩土参数分析和选用[J]. 岩土力学,2013,34(7):1995-1999.
- [6] 朱红霞,闫澎旺. 随机场理论在抗剪强度指标中的应用[J]. 岩石力学与工程学报,2008,27(增1):3040-3046.
- [7] 中华人民共和国建设部. 岩土工程勘察规范(GB 50021-2001)(2009年版)[S]. 北京:中国建筑工业出版社,2009.
- [8] 中华人民共和国住房和城乡建设部. 建筑地基基础设计规范(GB 50007-2011)[S]. 北京:中国建筑工业出版社,2011.
- [9] 高大钊. 关于岩土设计参数标准值计算公式的讨论[J]. 工程勘

- 察,1996,(3):5-8,16.
- [10] Efron B. Bootstrap Methods; Another Look at the Jack Nife [J]. The Annals of Statistics,1979,(1):1-26.
- [11] LUO Z, Atamturktur S, Juang C H. Bootstrapping for Characterizing the Effect of Uncertainty in Sample statistics for Braced Excavations[J]. Journal of Geotechnical and Geoenvironmental Engineering,2013,139(1):13-23.
- [12] Johnson R W. An Introduction to the Bootstrap[J]. Teaching Statistics,2001,23(2):49-54.
- [13] Duncan J M. Factors of Safety and Reliability in Geotechnical Engineering[J]. Journal of Geotechnical and Geoenvironmental Engineering,2000,126(4):307-316.
- [14] Baecher G B, Christian J T. Reliability and Statistics in Geotechnical Engineering[M]. New York: John Wiley & Sons,2003.