

改进 Bootstrap 法在海上风电场岩土参数区间估计中的应用

李亚杰

中国电建集团福建省电力勘测设计院有限公司 福建 福州 350000

【摘要】：海上风电场地质勘察因成本高、钻孔间距大，岩土试验数据样本量往往较小，传统 t 分布法区间估计易偏宽、结果保守，影响工程经济性。为此，提出一种基于样本空间扩展的改进 Bootstrap 区间估计方法，通过构建扩展样本空间提高小样本自助抽样的离散程度，减少重抽样对原始样本的过度依赖，从而提升区间估计的稳定性与可靠性。以福建某海上风电场预可研勘察资料为例，对压缩模量、压缩系数、黏聚力及内摩擦角分别采用两种方法进行区间估计对比。结果表明：相同置信水平下，改进 Bootstrap 法所得区间更集中、稳定，有效降低了传统方法因样本不足带来的保守性；当样本容量 $n \geq 8$ 、扩展参数 $h \geq 3$ 时，估计区间趋于稳定。研究成果可为海上风电场预可研阶段小样本岩土参数统计分析提供参考。

【关键词】：改进 Bootstrap 方法；海上风电；岩土参数；区间估计；小样本

DOI:10.12417/3083-5526.25.11.018

1 引言

岩土参数的分析与选定是岩土工程勘察的基础性工作，其合理性直接关系工程的安全性与经济性。参数取值过于保守，将显著增加工程造价，造成资源浪费；取值失准，则可能危及工程安全。海上风电场地质勘察中，场地范围广、钻孔间距大，可获取的岩土参数样本容量通常较小；且远离陆地，缺乏区域性经验取值依据。因此，如何从小样本中充分提取有效信息、准确确定参数估计值，对保障海上风电场工程安全性与经济性具有重要意义。

尽管 Bootstrap 方法已广泛应用于岩土参数统计分析，但面对海上风电场预可研阶段的极小样本数据，传统 Bootstrap 法仍存在局限：样本容量较小时，自助抽样易产生大量与原始样本高度相似的重抽样样本，导致估计结果对原始样本过度依赖，影响区间估计的稳定性与可靠性。针对该问题，本文改进传统 Bootstrap 抽样过程并应用于海上风电场岩土参数区间估计，通过构建扩展样本空间提高重抽样样本的离散程度，结合工程案例分析其适用性，并探讨样本容量 n 及扩展参数 h 对估计结果的影响，提出适用的参数取值建议，为小样本条件下岩土参数合理选取提供参考。

2 岩土参数估计区间计算方法

2.1 自助抽样法

假设总体分布 F 未知，给定来自该总体的容量为 n 的样本数据 $X = (X_1, X_2, \dots, X_n)$ 。通过有放回抽样从该样本中抽取一个容量仍为 n 的样本，称为自助样本（Bootstrap 样本）。重复上述过程 B 次，即可获得 B 个自助样本。基于这 B 个自助样本对总体分布 F 进行统计推断的数理统计方法，即称为自助抽样（Bootstrap）法。

应用自助抽样法估计置信水平下的置信区间时，通常需借助样本数据的经验分布函数。一般地，可按式（1）计算。

$$F_n(X) = \begin{cases} 0 & (X < X_1) \\ \frac{i}{n} & (X_i \leq X < X_{i+1}, i = 1, 2, \dots, n) \\ 1 & (X \geq X_n) \end{cases} \quad (1)$$

直接使用自助抽样法的问题在于，当样本容量很低的时候，取出与原样本具有很高相似度的自助样本，使得参数的概率分布集中于少量几个点上，导致最终的计算结果偏离实际的分布情况。针对这种情况，有两种改进方法：一种是对经验分布函数的改进，改进出更为合理的经验分布函数，另一种是对小样本进行改造，调整抽样方法，增大样本的容量。对于海上风电场的岩土体，其参数分布情况受到形成条件、埋藏深度、地下水活动等多重条件的制约，难以给出合适经验函数，因此对小样本岩土参数改造扩充是可以改进的方向。

2.2 改进的自助抽样法

本文使用一种方便易操作的样本空间扩展方法，首先将容量为 n 岩土参数样本按照从小到大的顺序排列，将得到的结果记为 $x(x_1, x_2, \dots, x_n)$ ，然后对每个值 x_i 作如下处理：

$$U_1 = [x_1 - \frac{x_2 - x_1}{h}, x_1 + \frac{x_2 - x_1}{h}] \quad (2)$$

$$U_i = [x_i - \frac{x_i - x_{i-1}}{h}, x_i + \frac{x_{i+1} - x_i}{h}] \quad i = 2, \dots, n-1 \quad (3)$$

$$U_n = [x_n - \frac{x_n - x_{n-1}}{h}, x_n + \frac{x_n - x_{n-1}}{h}] \quad (4)$$

参数 h 通常取大于 2 的值以保证扩展样本的稳定性，因此本文初步取 $h=3$ 进行工程算例分析，并在后续章节进一步讨论 h 取值对估计结果的影响。这里取 $h=3$ ，这样就得到一个样本空间 $U(U_1, U_2, \dots, U_n)$ ，将 $U(U_1, U_2, \dots, U_n)$ 与原有样本 $x(x_1, x_2, \dots, x_n)$ 合并并按照从小到大的顺序排序可获得一个扩

展了的岩土参数样本空间，样本容量为3N，记为X(X1,X2,...,X3N)。新的样本空间在尽可能保持原样本分布特性的条件下，扩大了样本空间容量，大大降低自助样本与原样本的相似性。得到了新的样本空间后就可以使用自助抽样法进行计算。

3 工程实例计算分析

3.1 工程实例

以福建海域某海上风电场预可研勘测设计阶段为例，根据钻孔揭露情况，选取具有代表性的地层③1、⑤3、⑦1，各地层岩性特征如下：

③1—粉砂：灰色，饱和，中密。颗粒级配差，主要矿物成分为石英、长石。含贝壳碎片。偶见黏性土团块，局部夹10-12cm的粉质黏土夹层。该层砂土层顶埋深8.10-14.20m，厚度1.70-10.10m。

⑤3—粉质黏土：灰色，可塑。含少量贝壳碎片，混有粉细砂，局部夹2-3cm的粉砂薄层。该层土层层顶埋深43.00-61.50m，厚度4.40-24.10m。

⑦1—粉质黏土：灰色，局部见褐黄色，可塑。夹8-15mm粉细砂薄层，个别达到25-30mm，局部夹粉细砂层。该层土层层顶埋深61.10-72.80m，厚度3.7-19.15m。

以③1的压缩模量、⑤3的压缩系数、⑦1层的固结快剪聚力和内摩擦角为例，进行对比计算，原始土工数据如表1、表2、表3所示：

表1 ③1—粉砂层压缩模量土工数据

层号	岩土参数	①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩
③1	压缩模量E(MPa)	12.35	10.29	8.27	9.34	10.21	9.76	8.74	13.34	11.25	10.82

表2 ⑤3—粉质黏土层压缩系数土工数据

层号	岩土参数	①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩
⑤3	压缩系数a(MPa ⁻¹)	0.34	0.29	0.26	0.44	0.49	0.36	0.35	0.32	0.33	0.3
	岩土参数	⑪	⑫	⑬	⑭	⑮	⑯	⑰	⑱	⑲	/
	压缩系数a(MPa ⁻¹)	0.35	0.27	0.33	0.27	0.25	0.18	0.28	0.37	0.17	/

表3 ⑦1—粉质黏土层固结快剪聚力和内摩擦角土工数据

层号	岩土参数	①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩	⑪	⑫
⑦1	粘聚力C _{sk} (kpa)	29	26	18	14	21	32	25	29	27	26	20	16
	内摩擦角φ _{sk} (°)	16.5	18.2	16	15.1	16	15.9	17.5	17.5	18	18.3	22.7	18.3
	岩土参数	⑬	⑭	⑮	⑯	⑰	⑱	⑲	⑳	㉑	㉒	㉓	㉔

表4 t分布法与改进 Bootstrap 法区间估计结果

土层	岩土参数	估计方法	下限值	上限值
③1	压缩模量E(MPa)	t分布法	8.65	12.86
		改进自助法	9.76	11.93
		t分布法	0.28	0.34
⑤3	压缩系数a(MPa ⁻¹)	改进自助法	0.29	0.33
		t分布法	20.03	33.13
		改进自助法	23.15	31.58
⑦1	粘聚力C _{sk} (kpa)	t分布法	16.46	18.54
		改进自助法	16.93	18.26
	内摩擦角φ _{sk} (°)	t分布法	16.46	18.54
		改进自助法	16.93	18.26

图1 原始土工试验数据

由表4可知，在相同置信水平条件下，改进 Bootstrap 法计算得到的各项岩土参数置信区间均位于传统 t 分布法估计区间范围内，且区间宽度整体较小。说明在本工程算例条件下，改进 Bootstrap 法能够在保持统计可靠性的基础上获得更加集中的区间估计结果。

从工程应用角度来看，海上风电场岩土参数取值通常直接影响基础设计及工程投资。当参数估计区间过宽时，往往会导

致设计参数偏于保守，增加基础尺寸和工程造价。本文提出的改进 Bootstrap 法能够充分利用有限试验数据中的统计信息，为岩土参数选取提供更加合理的参考依据。

3.2 样本容量 n 对改进自助取样法区间估计的影响

对来自⑤3—粉质黏土层压缩系数 a 土工数据总体随机抽取数量为6、7、8、9、10、11、12、13、14、15、16、17、18、19的子样本，计算 t 分布法双侧置信水平α=90%的置信区间和改进自助抽样法双侧置信水平α=90%的置信区间，其中自助样本数 B=10 000，参数 h=3。则计算出置信区间随样本容量的变化情况如图2所示。

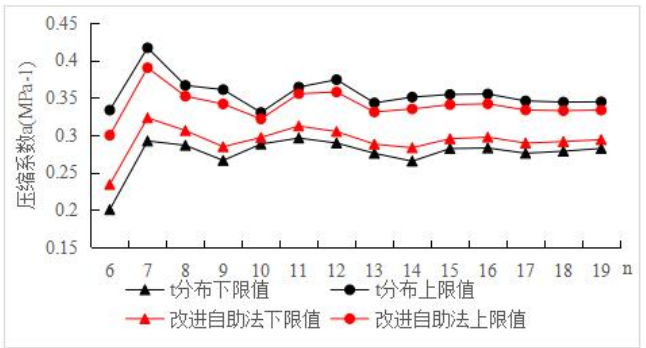


图2 不同样本容量下压缩系数估计区间变化曲线

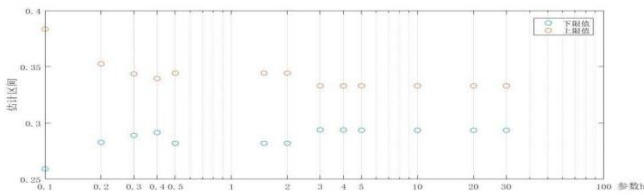
图2反映了样本容量对区间估计结果的影响规律。随着样本容量n增加，两种方法计算得到的区间上下限均逐渐收敛，说明样本容量增加有助于提高估计结果稳定性。当n=6~7时，区间边界波动明显，估计结果受个别样本影响较大；当n≥8时，区间变化幅度显著减小，估计结果趋于稳定。

此外，在不同样本容量条件下，改进 Bootstrap 法计算得到的区间均较 t 分布法更加集中，说明扩展样本空间能够有效缓解小样本条件下重抽样样本过度重复的问题。对于海上风电场预可研阶段常见的小样本数据，建议样本容量不低于8组，以提高区间估计结果的可靠性。

3.3 参数 h 取值对改进自助取样法区间估计的影响

随着 h 逐渐增大，区间上下限逐步收敛，估计结果趋于稳定。当 h≤2 时，区间边界波动较明显，表明扩展样本空间尚未充分形成；当 h≥3 时，区间变化幅度显著减小，不同 h 值对应的估计结果差异较小。

从工程应用角度考虑，过大的 h 虽然能够进一步扩展样本空间，但计算量也随之增加，而估计结果改善有限。因此综合估计精度与计算效率，建议实际工程应用中优先采用 h≥3 的参数设置。


 图3 参数 h 对压缩系数估计区间的影响

3.4 方法适用性讨论

本文所采用的改进 Bootstrap 法不依赖总体分布假设,适用于海上风电场勘察阶段常见的小样本岩土参数分析。与传统 t 分布法相比,其能够充分利用有限样本信息,减小由于样本容量不足导致的区间估计偏保守现象。

需要指出的是,本文研究基于实际工程案例展开,所得 $n \geq 8$ 和 $h \geq 3$ 的建议主要适用于本文所涉及岩土参数及样本规模条件。对于不同地区、不同地层类型以及更复杂参数分布情况,仍需结合更多工程案例和数值模拟进行验证。

此外,本文主要采用区间宽度和估计稳定性作为评价指标,尚未进一步分析区间覆盖率等统计性能。后续可结合 Monte

Carlo 模拟方法对改进 Bootstrap 法进行更全面评价。

4 结语

(1) 针对海上风电场岩土工程勘察中样本容量较小、传统 t 分布法区间估计偏于保守的问题,提出了一种基于样本空间扩展的改进 Bootstrap 区间估计方法。

(2) 工程算例表明,在相同置信水平条件下,改进 Bootstrap 法所得置信区间整体较传统 t 分布法更加集中,并表现出较好的估计稳定性,能够有效缓解小样本条件下参数估计过于保守的问题。

(3) 通过对不同样本容量及扩展参数 h 的敏感性分析发现,随着样本容量增加,区间估计结果逐渐趋于稳定;对于本文算例,当 $n \geq 8$ 时可获得较稳定的估计结果。

(4) 参数 h 对区间估计结果具有一定影响。当 $h \geq 3$ 时,估计区间变化趋于平稳,建议工程应用中优先取 $h \geq 3$ 。

(5) 本文研究主要基于实际工程案例开展分析,后续可进一步结合 Monte Carlo 模拟试验及覆盖率评价指标,对改进 Bootstrap 方法的统计性能进行深入研究。

参考文献:

- [1] 骆飞.小样本岩土参数统计特征估计及边坡稳定可靠性分析[D].西南交通大学,2017.
- [2] 罗强,骆飞,蒋良淮,张文生.基于 Bootstrap 法的小样本岩土参数标准值计算[J].华中科技大学学报(自然科学版),2017,45(08):93-98.
- [3] 赵宇飞.岩土强度参数选取方法研究及工程应用[D].中国水利水电科学研究院,2005.
- [4] 谢桂华.岩土参数随机性分析与边坡稳定可靠度研究[D].中南大学,2009.
- [5] 唐小松,李典庆,曹子君,周创兵,方国光.有限数据条件下边坡可靠度分析的 Bootstrap 方法[J].岩土力学,2016,37(03):893-901.
- [6] 骆飞,罗强,蒋良淮,张文生,李昂.小样本岩土参数的 Bootstrap 估计及边坡稳定分析[J].岩石力学与工程学报,2017.
- [7] 王栋.基于自助抽样法与少量试验数据的边坡可靠性分析[J].铁道建筑,2020.
- [8] 盛骤,谢式千,潘承毅.概率论与数理统计.4版[M].高等教育出版社,2008.
- [9] Efron B. Bootstrap Methods: Another Look at the Jackknife[J].Annals of Statistics,1979
- [10] Efron B, Tibshirani R J. An Introduction to the Bootstrap[M].New York: Chapman & Hall,1993.