

β 分布的参数确定及其在岩土工程中的应用

刘 勇 郑俊杰 郭 嘉

(华中科技大学岩土与地下工程研究所, 湖北武汉 430074)

【摘 要】 对岩土工程中随机变量的空间概率特征进行了统计分析, 介绍了确定 β 分布各参数的迭代方法, 提出了经验公式简化迭代过程, 有效地处理了随机变量分布范围的估计问题, 并阐明了 β 分布在可靠度领域中的适用性。作为对比, 对同一样本采用了其他分布进行拟合。分析结果表明, β 分布拟合精度高于其他分布。

【关键词】 随机变量; 分布拟合; β 分布; 经验公式; 可靠性分析

【中图分类号】 TU 44

Determination of Parameters for Beta Distribution and Its Application in Geotechnical Engineering

Liu Yong Zheng Junjie Guo Jia

(Institute of geotechnical and underground engineering,
Huazhong University of Science and Technology, Wuhan Hubei 430074 China)

【Abstract】 This paper statistically analyses the spatial probabilistic characteristics of random variables in geotechnical problems. Both the iteration process to determine the parameters for a beta probability density function and the empirical formula to simplify this iteration are presented. Therefore, the estimation of the bounds of random variables is handled efficiently and the applicability of the beta distribution in reliability analysis is illustrated. Certain other types of distribution are chosen to fit the same random sample for comparison and the results prove that the beta distribution is superior to others.

【Key Words】 random variable; distribution fitting; beta distribution; empirical formula; reliability analysis

0 引 言

进行岩土工程中的可靠度分析时, 随机变量概率分布类型的选取是一项关键的工作。一些学者对参数的分布类型进行了研究和探讨: 谭忠盛^[1]认为隧道围岩强度指标 c 、 $\tan \varphi$ 服从正态分布; 陈立宏^[2]认为土体抗剪强度指标服从对数正态分布; 张小敏^[3]、罗书学^[4]对桩基承载力试计比(单桩承载力的试验值与按规范经验公式的计算值两者的比值)选用对数正态分布。可靠性分析中, 一般允许的失效概率是在 $10^{-3} \sim 10^{-4}$ 数量级。这时, 计算分布尾部小概率的误差就成为估计失效概率的关键问题^[5]。岩土工程问题中, 绝大多数随机变量的分布区域有限, 选用正态、对数正态、极值型等无界分布, 会使估计的可靠性指标在一定程度上的失真, 甚至使计算结果失去实际意义。因此, 对相应的随机样本选用具有边界的分布显得尤为重要。文献[5-7]

引入具有上、下边界的 β 分布拟合土体抗剪强度指标, 取得了理想的拟合效果, 然而这些文献将样本的最大(小)值确定为分布的上(下)边界, 将已发生事件(样本最大值、最小值)做零概率处理, 对可能出现但未出现的小概率事件未做“空间预留”。本文在此基础上, 收集国内外相关资料就如何合理确定 β 分布上、下界及其形参数进行了探讨并通过实例说明了该分布的拟合效果。

1 β 分布

称随机变量 X 服从 β 分布, 如果 X 的概率密度函数为

$$f_X(x) = \begin{cases} \frac{1}{(b-a)B(\gamma, \eta)} \left(\frac{x-a}{b-a} \right)^{\gamma-1} \times \\ \left[1 - \frac{x-a}{b-a} \right]^{\eta-1}, & (a \leq x \leq b) \\ 0, & \text{其他} \end{cases} \quad (1)$$

式中: $\gamma > 0, \eta > 0, B(\gamma, \eta)$ 为 Beta 函数。称 a 和 b 分别为分布的下限和上限; γ, η 为分布的形参数。

形参数 γ, η 的表达式为^[8]

$$\left. \begin{aligned} \gamma &= \frac{(\mu_x - a)^2(b - \mu_x) - \sigma_x^2(\mu_x - a)}{\sigma_x^2(b - a)} \\ \eta &= \frac{(\mu_x - a)(b - \mu_x)^2 - \sigma_x^2(b - \mu_x)}{\sigma_x^2(b - a)} \end{aligned} \right\} \quad (2)$$

式中: μ_x 和 σ_x 分别为 X 的均值和标准差。

取归一化变量:

$$Z = \frac{X - a}{b - a} \quad (3)$$

则 Z 的概率密度函数为

$$f_z(z) = \begin{cases} \frac{1}{B(\gamma', \eta')} z^{\gamma'-1} (1-z)^{\eta'-1}, & (0 \leq z \leq 1) \\ 0, & \text{其他} \end{cases} \quad (4)$$

式中: $\gamma' > 0, \eta' > 0$, 此时分布为标准 β 分布, γ', η' 为标准 β 分布的形参数。记为: $Z \sim BE(\gamma', \eta')$ 。

β 分布概率密度函数的形状由形参数 γ, η 来控制, 从均匀分布到近似正态分布, 从对称到不对称, 而且分布范围有限。形参数对分布形状的影响见图 1。

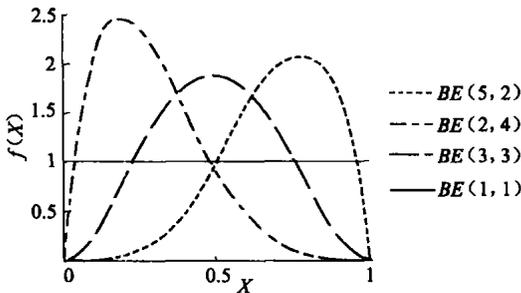


图 1 标准 β 分布概率密度函数

2 β 分布参数的确定

2.1 直接确定法

文献[5-7]采用了一套确定 β 分布四个参数的方法:

1) 将样本的最小值定为分布下限 a ; 将样本最大值确定为分布上限 b 。

2) 给出分布形参数 γ, η 与归一化变量 Z 的期望 $E(Z)$ 和方差 $V(Z)$ 的一个关系:

$$\left\{ \begin{aligned} E(Z) &= \frac{\gamma + 1}{\gamma + \eta + 2} \\ V(Z) &= \frac{(\gamma + 1)(\eta + 1)}{(\gamma + \eta + 2)^2(\gamma + \eta + 3)} \end{aligned} \right. \quad (5)$$

基于此关系, 反求出 γ, η 的值。

这样处理使得计算简便。然而在由样本最小值和最大值确定的区间以外, 还会存在一些小概率事

件, 这些事件在特定一次样本值中没有出现。由失效概率的几何意义可知, 这样处理不利于该分布在可靠度领域的应用。

2.2 传统的迭代方法

文献[8]考虑了样本值的特性, 给出确定 β 分布(本文引用样本容量大于 20 的情况)参数的迭代过程, 步骤如下:

① 计算样本均值和方差

$$X = \sum_{i=1}^n \frac{X_i}{n}; \quad S_x^2 = \sum_{i=1}^n \frac{(X_i - X)^2}{n-1}$$

② 将样本 X_i 排序, 令

$$Y_1 = X_{\min}, \quad Y_n = X_{\max}, \quad Y_1 \leq Y_2 \leq \dots \leq Y_n$$

③ 计算边界 a, b 的估计值 \hat{a}, \hat{b}

$$\hat{a} = Y_1 + F_A(i, n); \quad \hat{b} = Y_n + F_B(i, n)$$

$$\text{式中: } F_A(i, n) = Y_1 - (e-1) \sum_{i=1}^{21} e^{-i} Y_i$$

$$F_B(i, n) = Y_n - (1-e^{-1}) \sum_{i=0}^{20} e^{-i} Y_{n-i}$$

④ 分别用估计值 $X, S_x^2, \hat{a}, \hat{b}$ 替代 μ_x, σ_x^2, a, b ,

由式(2)计算 γ, η 的估计值 γ, η

$$\gamma = \frac{(X - \hat{a})^2(b - X) - S_x^2(X - \hat{a})}{S_x^2(b - \hat{a})}$$

$$\eta = \frac{(X - \hat{a})(b - X)^2 - S_x^2(b - X)}{S_x^2(b - \hat{a})}$$

⑤ 计算 b 的新估计值 \hat{b}

$$\hat{b} = Y_n + C_2(\nu) F_B(i, n)$$

式中: $C_2(\nu) = a_1(\nu) / a_2(\nu)$

$$a_1(\nu) = \Gamma(\nu + 1)(1 - e^{-1}) \sum_{i=0}^{21} G_A(i, \nu) - \Gamma(2\nu + 1)$$

$$a_2(\nu) = \Gamma(2\nu + 1) \{ 1 + (1 - e^{-1})^2 (1 - e^{-2})^{-2\nu-1} \} -$$

$$2\Gamma(\nu + 1)(1 - e^{-1}) \sum_{i=0}^{21} G_A(i, \nu) +$$

$$2(1 - e^{-1})^2 \sum_{i=1}^{21} G_A(i, \nu) \sum_{j=0}^{i-1} G_B(j, \nu)$$

$$G_A(i, \nu) = e^{-i} \frac{\Gamma(2\nu + i + 1)}{\Gamma(\nu + i + 1)}$$

$$G_B(j, \nu) = e^{-j} \frac{\Gamma(\nu + j + 1)}{\Gamma(j + 1)}$$

此处 ν 值取为 $1/\eta$; $\Gamma(\cdot)$ 为伽马函数。

⑥ 按第 4 步由 \hat{a}, \hat{b} 当前值计算 γ, η 的新估计值。

⑦ 计算 a 的新估计值 \hat{a}

$$\hat{a} = Y_1 + C_2(\nu) F_A(i, n)$$

这一步 ν 取值改为: $1/\gamma$ 。

⑧ 再按第 4 步由 \hat{a}, \hat{b} 当前值计算 γ, η 的新估计值。

⑨ 重复步骤⑤~⑧, 直到 \hat{a}, \hat{b} 的当前值与前一

步的估计值之差均等于或小于预定精度 ϵ_0 为止。

2.3 简化迭代方法

在 2.2 节中的迭代步骤⑤~⑧, 需要反复计算函数 $C_2(\nu)$, 计算过程烦琐。作者通过反复计算、曲线拟合, 得出 $C_2(\nu)$ 与 ν 之间的双曲线经验关系:

$$C_2(\nu) = -0.130 + \frac{1.31}{\nu} \quad (6)$$

将式(6)代入 2.2 节中的迭代步骤⑤~⑧可以将传统迭代方法大为简化, 方便计算机编程实现迭代过程。实践表明该经验式在迭代过程中不会影响收敛效果, 且其与文献[9]提供的节点估计数据误差平方和为零。

3 算例分析

3.1 打入桩承载力试计比统计分析

进行桩基承载力可靠度分析时, 桩的承载力受桩径、桩长、地区、施工方法等诸多因素影响, 很难得到数量足够多的可认为条件相同的样本来统计。利用量纲一的量随机变量试计比, 将试桩极限承载力归一化, 从而以使各种试桩处于同一分析水平上, 解决样本容量不够大的问题。桩基承载力试计比对可靠度指标影响最大, 占 90% 以上^[3], 因此对其分布类型的确定有着重要的意义。

文献[4]收集了 128 根打入桩试桩资料, 利用各根桩静载试桩承载力以及《铁路桥涵设计规范》经验公式计算得到极限承载力的标准值, 算出它们的试计比(见表 1), 同时分别采用正态、对数正态、极值 I 型分布对数组拟合, 通过 K-S 检验法, 统计分析表明对数正态分布更符合该数组。

表 1 打入桩承载力试计比统计数据

0.535	0.612	0.710	0.728	0.776	0.827	0.829	0.832
0.871	0.871	0.872	0.887	0.887	0.899	0.910	0.915
0.934	0.940	0.954	0.955	0.973	0.978	0.983	1.023
1.032	1.036	1.042	1.048	1.048	1.062	1.068	1.070
1.074	1.080	1.082	1.085	1.086	1.091	1.091	1.092
1.092	1.097	1.098	1.098	1.099	1.104	1.110	1.110
1.115	1.121	1.122	1.124	1.127	1.128	1.130	1.132
1.132	1.134	1.141	1.143	1.147	1.154	1.155	1.155
1.156	1.159	1.160	1.184	1.187	1.187	1.193	1.208
1.210	1.211	1.213	1.215	1.227	1.228	1.231	1.231
1.233	1.233	1.234	1.237	1.238	1.240	1.245	1.246
1.255	1.263	1.271	1.273	1.278	1.283	1.287	1.289
1.290	1.302	1.304	1.310	1.311	1.312	1.323	1.325
1.325	1.343	1.348	1.353	1.354	1.360	1.381	1.385
1.404	1.406	1.408	1.414	1.430	1.438	1.443	1.465
1.473	1.491	1.492	1.512	1.537	1.565	1.592	1.599

在此基础上, 本文引入 β 分布对表 1 的样本进行统计分析。从试计比的物理意义及其频率直方图(见图 2)可知分布是左偏的, 而正态分布、对数正态分布图形分别是对称、右偏的, 从而用这两者来拟合会带来一定偏差。

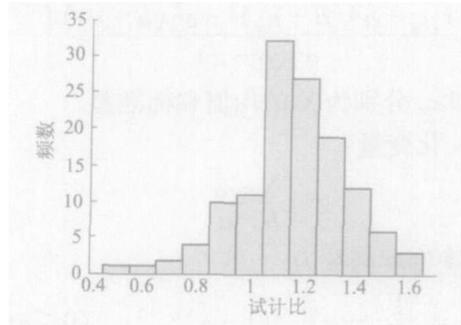


图 2 打入桩试计比频率直方图

取显著性水平 $\alpha=0.05$, 对样本的以下三种分布假设用 K-S 法检验:

H_0 为正态分布(见文献[4]);

H_0 为对数正态分布(见文献[4]);

H_0 为本文推荐的 β 分布($a=0.0305$, $b=1.6381$, $\gamma=8.9051$, $\eta=3.7192$)。

检验结果见表 2。接受水平定义为检验统计量与检验临界值的比值。文献[10]认为接受水平(小于 1 的情况下)越接近于 0, 则拟合效果越好, 由此确定出拟合度最优的分布概型, 并称之为有限比较检验法。表 2 显示三种假设都能接受, 本文接受有限比较检验法的概念可得出: 选择假设 H_0 优于假设 H_0 、 H_0 。

表 2 打入桩承载力试计比的 K-S 检验结果

	正态分布 H_0	对数正态分布 H_0	β 分布 H_0
检验统计量 D_n	0.076 43	0.074 25	0.065 96
检验临界值 $D_{n, \alpha}$	0.078 31	0.078 31	0.1186 6
接受否	接受	接受	接受
接收水平	0.975 99	0.948 15	0.555 89

3.2 粉质粘土压缩指数统计分析

土是在漫长的地质年代里形成的, 是自然历史的产物, 故土性指标可以视为随机变量。在岩土工程可靠性分析时, 需要了解土性指标这些随机变量的概率分布特征。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^[11]。

文献[11]收集了太原地区粉质粘土的压缩指数 C_c 的相关资料,给出了压缩指数 C_c 的一个容量为 55 的样本(见表 3),并由有限比较检验法得出本例中正态分布拟合优于按 2.1 节方法确定的 β 分布、对数正态分布。本文结合该样本,对粉质粘土压缩指数 C_c 的分布类型进行分析。下面用本文推荐的方法确定 β 分布来拟合该数组。按常规的 χ^2 检验方法对以下两种分布假设做检验(显著性水平 $\alpha=0.05$):

H_{01} 为正态分布 $(N(0.241, 0.088^2))^{[11]}$

H_{02} 为本文推荐的 β 分布 ($a=0.0647, b=0.4475, \gamma=1.7157, \eta=2.0087$)。

表 3 粉质粘土压缩指数 C_c 的统计数据

0.075	0.082	0.093	0.105	0.112	0.121	0.125
0.136	0.139	0.147	0.152	0.156	0.171	0.172
0.175	0.182	0.186	0.192	0.195	0.205	0.216
0.218	0.226	0.227	0.232	0.235	0.239	0.241
0.246	0.251	0.259	0.262	0.265	0.266	0.275
0.278	0.282	0.285	0.296	0.301	0.307	0.312
0.318	0.326	0.329	0.331	0.335	0.342	0.348
0.351	0.363	0.372	0.385	0.397	0.420	

检验结果见表 4。

表 4 粉质粘土压缩指数 C_c 的 χ^2 检验结果

分组数	临界值	正态分布		β 分布	
		统计量	接收水平	统计量	接收水平
7	9.488	0.774	0.082	0.808	0.085
8	11.070	1.573	0.142	0.749	0.068
9	12.592	2.348	0.186	1.533	0.122

表 4 中显示将样本分为 7 组时,两假设下的接受水平基本持平;样本被分别分为 8、9 组时, H_{02} 优于 H_{01} 。由有限比较检验法综合比较可得出:选择假设更合理。本例也间接比较了假设 H_{02} 与按 2.1 节的公式确定的 β 分布,前者拟合效果更理想。

4 结 论

β 分布本身具有分布形状适应性强、分布区域有界等性质,用它来描述非对称并具有一定分布范

围的统计量有优越性。将样本最大(小)值确定为 β 分布的上(下)限,未对随机变量的小概率事件做“空间预留”。正态、对数正态、极值型分布等的分布域为无穷,它们对样本尾部预留过多。这两者运算上较方便,但在精度要求较高的可靠度分析中存在缺陷。本文提供的迭代算法有效地解决了这一不足,经验公式的提出使得迭代过程变得简便,便于计算机编程。同时为岩土工程可靠度分析提供了一个偏度变化灵活、具有上下边界、拟合精度高的分布类型。值得注意的是,本文结论是在给定样本的容量大于 20 的情况下得出的。

参 考 文 献

- [1] 谭忠剩,高波,关宝树. 隧道围岩抗剪强度指标 c 、 $\tan\varphi$ 的概率特征[J]. 岩土工程学报, 1999, 21(6): 760-762.
- [2] 陈立宏,陈祖煜,刘金梅. 土体抗剪强度指标的概率分布类型研究[J]. 岩土力学, 2005, 26(1): 37-40.
- [3] 张小敏,郑俊杰. CFG 桩复合地基承载力可靠度分析[J]. 岩土力学, 2002, 23(6): 810-812.
- [4] 罗书学. 桩基概率极限状态法研究及其工程应用[D]. 成都: 西南交通大学, 2004.
- [5] 高大钊. 土力学可靠性原理[M]. 北京: 中国建筑工业出版社, 1989.
- [6] Lumb P. Safety factors and the probability distribution of soil strength[J]. Canadian Geotechnical Journal, 1970, 7(3): 225-241.
- [7] 宁宝宽,杨军. 沈阳粉质粘土工程指标的概率统计及分布模型[J]. 沈阳工业大学学报, 2004, 26(4): 461-463.
- [8] He J R. Estimating the distribution of manufactured dimensions with the beta probability density function[J]. Mach. Tools Manufacture, 1991, 31(3): 383-396.
- [9] Cooke P. Statistical inference for bounds of random variables[J]. Biometrika, 1979, 66(2): 367-374.
- [10] 张博庭. 用有限比较法进行拟合优度检验[J]. 岩土工程学报, 1991, 13(6): 84-91.
- [11] 李小勇,白晓红,谢康和. 岩土参数概率分布统计意义上的优化分析[J]. 岩土工程技术, 2000(3): 130-133.

收稿日期: 2006-08-22