文章编号：1000-8608(2010)01-0086-07

小样本问题的概率分析与工程应用

贡金鑫*， 李荣庆， 王利欢

（大连理工大学 海岸和近海工程国家重点实验室，辽宁 大连 116024）

摘要： 统计分析是工程中数据处理的基本手段，也是可靠度分析的基础.由于条件和试验费用的限制，有时只能获得容量极小的样本，如工程中荷载试验桩的数量一般不多于4根.在这种情况下，采用的数据统计分析方法至关重要.研究表明，线性最小偏差方法对于极小样本的统计分析具有良好的适用性.针对线性最小偏差参数估计方法，给出了相关统计量的概率密度函数，以此为基础可分析所估计参数的概率特性.以工程中按试桩确定桩极限承载力的标准值为例讨论了所得公式的应用性，证明可用于工程中小样本数据的分析和处理.

关键词： 统计；小样本；可靠度；桩

中图分类号： TU311.2 **文献标志码：** A

0 引 言

样本的概率分析是数理统计研究的基本问题，主要包括参数估计、假设检验等，其中 χ^2 分布、F 分布和 t 分布是经典的三大抽样分布[1]. 在工程中，样本分布可用于小样本的参数估计和推断.关于小样本，没有明确的样本容量规定，事实上，用样本估计总体参数的准确与否还与总体的变异性有关.有些情况下，即使样本容量不大，但总体变异性很小，如变异系数为0.01，不需太大的样本容量即能比较准确地估计出总体的参数，这种情况并不是小样本问题.但在工程中，有时样本的容量非常小，一般只有2～3个数据，可称为极小样本.如工程中按桩静荷载试验确定桩的承载力就是如此[2~6]，其判断的准确性对工程的安全性和经济性至关重要.早在1954年，文献[7]就提出了小样本问题参数的线性最小偏差估计方法，并指出这种方法与传统分析方法相比所具有的优越性.之后，其他一些研究者也做了很多工作[8]，但迄今只给出了参数估计的方法，没有给出样本参数的分布.为此，本文研究线性最小偏差估计的样本分布，并探讨样本分布在确定桩基承载力标准值中的应用.

1 参数的线性最小偏差估计

如前所述，参数估计是数理统计研究的基本问题之一.传统的参数估计方法有矩法、最大似然法、贝叶斯法等.对于总体中变量概率分布形式为 $F[(x-\lambda)/\alpha]$ 的问题，根据文献[7、8]，参数 λ 和 α 可用下式估计：

$$\hat{\lambda} = \sum_{i=1}^{n} a_i \cdot x_{(i)} \tag{1}$$

$$\hat{\alpha} = \sum_{i=1}^{n} b_i \cdot x_{(i)} \tag{2}$$

式中：a_i 和 b_i 为系数，与样本容量 n 有关，文献[8]的附表提供了不同概率分布时的值；$x_{(1)} \leqslant x_{(2)} \leqslant \cdots \leqslant x_{(n)}$ 为样本 $x_1, x_2, \cdots, x_n$ 自小到大排列的值.

当样本的母体为正态分布时，分布参数为平均值 μ 和标准差 σ，样本的平均值和标准差可用下式计算[8]：

$$\bar{x} = \frac{1}{n}(x_{(1)} + x_{(2)} + \cdots + x_{(n)}) = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{3}$$

$$s = b_1 x_{(1)} + b_2 x_{(2)} + \cdots + b_n x_{(n)} = \sum_{i=1}^{n} b_i x_{(i)} \tag{4}$$

其中 $b_1, b_2, \cdots, b_n$ 为系数，按表1取用[8].

收稿日期： 2008-01-10； **修回日期：** 2009-09-18.
基金项目： “十一五”国家科技支撑计划资助项目(2006BAJ01B6-06).
作者简介： 贡金鑫*(1964-)，男，教授，博士生导师，E-mail：gong_jx.vip@eyou.com.

表 1　样本母体为正态分布时计算样本标准差的系数

Tab. 1　Coefficient for calculation of standard variance of samples taking from normal distribution

n	b_1	b_2	b_3	b_4	b_5	b_6	b_7
2	−0.886 2	0.886 2					
3	−0.590 8	0	0.590 8				
4	−0.453 9	−0.110 2	0.110 2	0.453 9			
5	−0.372 4	−0.135 2	0	0.135 2	0.372 4		
6	−0.317 5	−0.138 6	−0.043 2	0.043 2	0.138 6	0.317 5	
7	−0.277 8	−0.135 1	−0.062 5	0	0.062 5	0.135 1	0.277 8

2　统计量的概率分布

2.1　统计量

统计量是不包括未知参数的函数，是由样本构成的随机变量，其分布可用来研究所要估计参数的概率特性，式(3) 和式(4)$\bar{x}$ 和 s 均为统计量. 不难证明[1]，$\overline{X} \sim N(\mu, \sigma^2/n)$，这样只需研究$(\overline{X}-\mu)/S$ 的分布.

设有 $X_{(1)} \leqslant X_{(2)} \leqslant \cdots \leqslant X_{(n)}$ 为来自正态母体的一个顺序统计量，如表 1 所示，$b_i(i = 1,2,\cdots,n)$ 的值是对称的，所以

$$V = \frac{S}{\sigma} = \frac{\sum_{i=1}^{n} b_i X_{(i)}}{\sigma} = \sum_{i=1}^{n} b_i \frac{X_{(i)}-\mu}{\sigma} = \sum_{i=1}^{n} b_i Y_{(i)} \tag{5}$$

$$Z = \frac{\overline{X}-\mu}{S} = \frac{\frac{1}{n}\sum_{i=1}^{n}\frac{X_{(i)}-\mu}{\sigma}}{\frac{S}{\sigma}} = \frac{\frac{1}{n}\sum_{i=1}^{n} Y_{(i)}}{V} \tag{6}$$

其中 $Y_{(i)} = \dfrac{X_{(i)}-\mu}{\sigma}(i = 1,2,\cdots,n)$. 显然 $Y_{(1)} \leqslant Y_{(2)} \leqslant \cdots \leqslant Y_{(n)}$ 也是一个顺序统计量.

2.2　顺序统计量的变换

将顺序统计量 $Y_{(1)} \leqslant Y_{(2)} \leqslant \cdots \leqslant Y_{(n)}$ 变换为随机变量 $U_1, U_2, \cdots, U_n$，即

$$U_i = a_{ij} Y_j (i,j = 1,2,\cdots,n) \tag{7}$$

其中

$$a_{ij} = \begin{cases} \dfrac{1}{\sqrt{(i+1)i}}; & i \geqslant j \\ -\dfrac{i-1}{\sqrt{(i+1)i}}; & i < j \\ \dfrac{1}{\sqrt{n}}; & i = n \end{cases}$$

式中$(a_{ij})_{n\times n}$ 是一个正交矩阵[1].

由式(7) 可解得

$$Y_{(i)} = c_{ij} U_j (i,j = 1,2,\cdots,n) \tag{8}$$

其中 c_{ij} 为矩阵$(a_{ij})_{n\times n}$ 逆矩阵(或转置矩阵) 的元素，且 $c_{nn} = 0$.

将式(8) 代入式(5) 和(6) 得

$$V = \sum_{i=1}^{n}\sum_{j=1}^{n} b_i c_{ij} U_j = \sum_{j=1}^{n-1} d_j U_j \tag{9}$$

$$Z = \frac{\frac{1}{n}\sqrt{n}U_n}{V} = \frac{U_n}{\sqrt{n}V} \tag{10}$$

其中 $d_j = \sum_{i=1}^{n-1} b_i c_{ij} (j = 1,2,\cdots,n-1)$，按表 1 中的 $b_i(i = 1,2,\cdots,n)$ 计算得 $d_n = 0$. 这样，S 与 U_n 是相互独立的.

将行向量$(d_1 \quad d_2 \quad \cdots \quad d_{n-1})$ 除以 $d_0 = \sqrt{\sum_{j=1}^{n-1} d_j^2}$ 得到一个单位向量. 根据线性代数原理，可以以该单位向量作为矩阵最后一行的行向量，利用 Gram-Schmidt 标准正交化方法产生正交矩阵 $\boldsymbol{E} = (e_{ij})_{(n-1)\times(n-1)}$. 对于表 1 中的 $b_i(i = 1,2,\cdots,n)$，计算表明 $\boldsymbol{E}$ 是一个具有如下形式的三角矩阵：

$$\boldsymbol{E} = \begin{pmatrix} e_{11} & e_{12} & 0 & 0 & \cdots & 0 \\ e_{21} & e_{22} & e_{23} & 0 & \cdots & 0 \\ e_{31} & e_{32} & e_{33} & e_{34} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ e_{n-2,1} & e_{n-2,2} & e_{n-2,3} & e_{n-2,4} & \cdots & e_{n-2,n-1} \\ e_{n-1,1} & e_{n-1,2} & e_{n-1,3} & e_{n-1,4} & \cdots & e_{n-1,n-1} \end{pmatrix}$$

利用正交矩阵 $\boldsymbol{E}$，进一步将随机变量 $U_1, U_2, \cdots, U_{n-1}$ 变换为随机变量 $V_1, V_2, \cdots, V_{n-1}$，即

$$V_i = \sum_{j=1}^{i+1} e_{ij} U_j (i = 1,2,\cdots,n-1) \tag{11}$$

$$V_{n-1}=\sum_{j=1}^{n-1}e_{n-1,j}=\sum_{j=1}^{n-1}\frac{d_j}{d_0}U_j \tag{12}$$

这样式(9)表示为

$$V=d_0V_{n-1} \tag{13}$$

对于本文研究的问题,计算表明,如果将式(11)和式(12)的第 $1,2,\cdots,n-1$ 行各分别除 $|e_{11}|,|e_{21}|,\cdots,|e_{n-1,1}|$,可得到下面形式的表达式(这是由表1中的 $b_i(i=1,2,\cdots,n)$ 决定的):

$$\frac{V_i}{|e_{i,1}|}=-\sum_{j=1}^{i}h_jU_j+\frac{e_{i,i+1}}{|e_{i,1}|}U_{i+1}$$
$$(i=1,2,\cdots,n-2) \tag{14}$$

$$\frac{V_{n-1}}{|e_{n-1,1}|}=-\sum_{j=1}^{n-1}h_jU_j \tag{15}$$

其中 $h_1,h_2,\cdots,h_{n-2}$ 均大于 0.

2.3 各变量的上下限

如前所述,$X_{(1)}\leqslant X_{(2)}\leqslant\cdots\leqslant X_{(n)}$ 是一个顺序统计量,经变换后 $Y_{(1)}\leqslant Y_{(2)}\leqslant\cdots\leqslant Y_{(n)}$ 也是一个顺序统计量,相应地 $U_1,U_2,\cdots,U_n$ 及 $V_1,V_2,\cdots,V_n$ 也存在一定的关系,在确定 S/σ 的概率密度函数时,需要知道这些随机变量间的关系.

由式(7)得

$$\frac{U_{i-1}}{\sqrt{i+1}}-\frac{U_i}{\sqrt{i-1}}=\frac{i}{\sqrt{(i+1)i(i-1)}}\times$$
$$(Y_{(i+1)}-Y_{(i)})\geqslant 0 \tag{16}$$

所以

$$U_i\leqslant\sqrt{\frac{i-1}{i+1}}U_{i-1}\leqslant\sqrt{\frac{i-1}{i+1}}\sqrt{\frac{i-2}{i}}U_{i-2}\leqslant\cdots\leqslant$$
$$\sqrt{\frac{i-1}{i+1}}\sqrt{\frac{i-2}{i}}\cdots\sqrt{\frac{2}{4}}U_2\leqslant$$
$$\sqrt{\frac{i-1}{i+1}}\sqrt{\frac{i-2}{i}}\cdots\sqrt{\frac{1}{3}}U_1\leqslant 0$$
$$(i=2,3,\cdots,n) \tag{17}$$

即

$$U_i\leqslant\sqrt{\frac{(j+1)j}{(i+1)i}}U_j(i\geqslant j) \tag{18}$$

或

$$U_j\geqslant\sqrt{\frac{(i+1)i}{(j+1)j}}U_i(j\leqslant i) \tag{19}$$

由于 $h_1>0,h_2>0,\cdots,h_{n-1}>0$,由式(15)和(7)得

$$\frac{V_{n-1}}{|e_{n-1,1}|}\geqslant-\sum_{j=i}^{n-1}h_jU_j\geqslant-\sum_{i=j}^{n-1}\sqrt{\frac{(j+1)j}{(i+1)i}}h_iU_j$$

所以有

$$U_j\geqslant u_{t,j}=-\frac{1}{\sum_{i=j}^{n-1}\sqrt{\frac{(j+1)j}{(i+1)i}}h_i}\frac{v_{n-1}}{|e_{n-1,1}|}$$
$$(j=1,2,\cdots,n-1) \tag{20}$$

由式(15)和(20)得

$$\frac{V_i}{|e_{i,1}|}\geqslant\frac{e_{i,i+1}}{|e_{i,1}|}U_{i+1}\geqslant\frac{e_{i,i+1}}{|e_{i,1}|}u_{t,i+1}$$

所以有

$$V_i\geqslant e_{i,i+1}u_{t,i+1} \tag{21}$$

进一步由式(15)和(20)得

$$\frac{V_{n-1}}{|e_{n-1,1}|}\leqslant-\sum_{i=1}^{n-1}\sqrt{\frac{n(n-1)}{(i+1)i}}h_iU_{n-1}$$

所以有

$$U_{n-1}\leqslant u_{s,n-1}=-\frac{1}{\sum_{i=1}^{n-1}\sqrt{\frac{n(n-1)}{(i+1)i}}h_i}\frac{v_{n-1}}{|e_{n-1,1}|} \tag{22}$$

进一步由式(19)得

$$-U_j\leqslant-u_{s,j}=\frac{\sqrt{\frac{n(n-1)}{(j+1)j}}}{\sum_{i=1}^{n-1}\sqrt{\frac{n(n-1)}{(i+1)i}}h_i}\frac{v_{n-1}}{|e_{n-1,1}|}$$
$$(j=1,2,\cdots,n-1) \tag{23}$$

由式(14)和(23)得

$$\frac{V_i}{|e_{i,1}|}\leqslant-\sum_{j=1}^{i-1}u_{s,j}h_j$$

所以有

$$V_i\leqslant-|e_{i,1}|\sum_{j=1}^{i-1}u_{s,j}h_j \tag{24}$$

进一步由式(14)和(20)得

$$\frac{V_i}{|e_{i,1}|}-\frac{V_{i+1}}{|e_{i+1,1}|}=\left(\frac{e_{i,i+1}}{|e_{i,1}|}+h_{i+1}\right)U_{i+1}-$$
$$\frac{e_{i+1,i+2}}{|e_{i+1,1}|}U_{i+2}\geqslant$$
$$\left(\frac{e_{i,i+1}}{|e_{i,1}|}+h_{i+1}\right)U_{i+1}-$$
$$\frac{e_{i+1,i+2}}{|e_{i+1,1}|}\sqrt{\frac{i+1}{i+3}}U_{i+1}\geqslant$$
$$\left[\left(\frac{e_{i,i+1}}{|e_{i,1}|}+h_{i+1}\right)-\right.$$
$$\left.\frac{e_{i+1,i+2}}{|e_{i+1,1}|}\sqrt{\frac{i+1}{i+3}}\right]u_{t,i+1}$$

所以有

$$V_i \geqslant \frac{|e_{i,1}|}{|e_{i+1,1}|}V_{i+1} + |e_{i,1}| \left[\left(\frac{e_{i,i+1}}{|e_{i,1}|} + h_{i+1}\right) - \frac{e_{i+1,i+2}}{|e_{i+1,1}|}\sqrt{\frac{i+1}{i+3}}\right]u_{t,i+1}$$

$$(i = 1,2,\cdots,n-3) \tag{25}$$

由式(14)和(23)得

$$\frac{V_i}{|e_{i,1}|} - \frac{V_{i+1}}{|e_{i+1,1}|} \leqslant -\frac{e_{i+1,i+2}}{|e_{i+1,1}|}U_{i+2} \leqslant -\frac{e_{i+1,i+2}}{|e_{i+1,1}|}u_{s,i+2}$$

$$(i = 1,2,\cdots,n-3)$$

所以有

$$V_i \leqslant \frac{|e_{i,1}|}{|e_{i+1,1}|}V_{i+1} - \frac{|e_{i,1}|}{|e_{i+1,1}|}e_{i+1,i+2}u_{s,i+2}$$

$$(i = 1,2,\cdots,n-3) \tag{26}$$

结合式(21)和(25)得 $V_i(i = 1,2,\cdots,n-2)$ 的上限为

$$t_i = \max\left\{e_{i,i+1}u_{t,i+1}, \frac{|e_{i,1}|}{|e_{i+1,1}|}v_{i+1} + |e_{i,1}| \times \left[\left(\frac{e_{i,i+1}}{|e_{i,1}|} + h_{i+1}\right) - \frac{e_{i+1,i+2}}{|e_{i+1,1}|}\sqrt{\frac{i+1}{i+3}}\right]u_{t,i+1}\right\} \tag{27}$$

结合式(24)和(26)得 $V_i(i = 1,2,\cdots,n-2)$ 的下限为

$$s_i = \min\left\{-|e_{i,1}| \sum_{j=1}^{i-1}u_{s,j}h_j, \frac{|e_{i,1}|}{|e_{i+1,1}|}v_{i+1} - \frac{|e_{i,1}|}{|e_{i+1,1}|}e_{i+1,i+2}u_{s,i+2}\right\} \tag{28}$$

2.4　概率密度函数

将顺序统计量 $X_{(1)} \leqslant X_{(2)} \leqslant \cdots \leqslant X_{(n)}$ 经过 $Y_{(1)} \leqslant Y_{(2)} \leqslant \cdots \leqslant Y_{(n)}$ 和 $U_1,U_2,\cdots,U_n$ 变换到随机变量 $V_1,V_2,\cdots,V_n$ 后，其联合概率密度函数也相应地变化. 下面依次给出变量联合概率密度函数的表达式.

来自正态母体的顺序统计量 $X_{(1)} \leqslant X_{(2)} \leqslant \cdots \leqslant X_{(n)}$ 的联合概率密度函数为[1]

$$f_X(x_{(1)},x_{(2)},\cdots,x_{(n)}) = \frac{n!}{(\sqrt{2\pi}\sigma)^n} \times \exp\left[-\frac{1}{2}\sum_{i=1}^{n}\frac{(x_{(i)}-\mu)^2}{\sigma^2}\right] \tag{29}$$

将 $X_{(1)} \leqslant X_{(2)} \leqslant \cdots \leqslant X_{(n)}$ 变换为顺序统计量 $Y_{(1)} \leqslant Y_{(2)} \leqslant \cdots \leqslant Y_{(n)}$ 后，联合概率密度函数为

$$f_Y(y_{(1)},y_{(2)},\cdots,y_{(n)}) = \frac{n!}{(\sqrt{2\pi})^n}\exp\left(-\frac{1}{2}\sum_{i=1}^{n}y_{(i)}^2\right) \tag{30}$$

由于 $Y_{(1)} \leqslant Y_{(2)} \leqslant \cdots \leqslant Y_{(n)}$ 到 $U_1,U_2,\cdots,U_n$ 的变换是正交变换，根据矩阵理论，有

$$|\boldsymbol{J}| = \left|\frac{\partial Y_i}{\partial X_j}\right|_{n\times n} = 1, \sum_{i=1}^{n}y_{(i)}^2 = \sum_{i=1}^{n}u_i^2$$

从而得到 $U_1,U_2,\cdots,U_n$ 的联合概率密度函数为

$$f_U(u_1,u_2,\cdots,u_n) = f_Y(y_{(1)},y_{(2)},\cdots,y_{(n)})|\boldsymbol{J}| = \frac{C_1}{(\sqrt{2\pi})^n}\exp\left(-\frac{1}{2}\sum_{i=1}^{n}u_i^2\right) \tag{31}$$

其中，C_1 由下式确定：

$$C_1 = \frac{(\sqrt{2\pi})^n}{\int\int\cdots\int \exp\left(-\frac{1}{2}\sum_{i=1}^{n}u_i^2\right)\mathrm{d}u_1\mathrm{d}u_2\cdots\mathrm{d}u_n}$$

由式(7)知，U_n 服从标准正态分布，即 $U_n \sim N(0,1)$，其概率密度函数为

$$f_{U_n}(u_n) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}u_n^2\right) \tag{32}$$

$U_1,U_2,\cdots,U_{n-1}$ 的联合概率密度函数为

$$f_U(u_1,u_2,\cdots,u_{n-1}) = \frac{C_1}{(\sqrt{2\pi})^{n-1}}\exp\left(-\frac{1}{2}\sum_{i=1}^{n-1}u_i^2\right) \tag{33}$$

同样由于 $\boldsymbol{E}$ 为正交矩阵，$|\boldsymbol{J}| = 1, \sum_{i=1}^{n-1}u_i^2 = \sum_{i=1}^{n-1}v_i^2$，$V_1,V_2,\cdots,V_{n-1}$ 的联合概率密度函数为

$$f_V(v_1,v_2,\cdots,v_{n-1}) = f_U(u_1,u_2,\cdots,u_{n-1})|\boldsymbol{J}| = \frac{C_2}{(\sqrt{2\pi})^{n-1}}\exp\left(-\frac{1}{2}\sum_{i=1}^{n-1}v_i^2\right) \tag{34}$$

其中，C_2 由下式确定：

$$C_2 = \frac{(\sqrt{2\pi})^{n-1}}{\int\int\cdots\int \exp\left(-\frac{1}{2}\sum_{i=1}^{n-1}v_i^2\right)\mathrm{d}v_1\mathrm{d}v_2\cdots\mathrm{d}v_{n-1}}$$

所以，V_{n-1} 的概率密度函数为

$$f_{V_{n-1}}(v_{n-1})=\frac{C_2}{(\sqrt{2\pi})^{n-1}}\int_{t_{n-2}}^{s_{n-2}}\int_{t_{n-3}}^{s_{n-3}}\cdots\int_{t_1}^{s_1}\exp\left(-\frac{1}{2}\times\sum_{i=1}^{n-1}v_i^2\right)\mathrm{d}v_1\mathrm{d}v_2\cdots\mathrm{d}v_{n-2} \quad (35)$$

根据式(15),$V=S/\sigma$ 的概率密度函数为

$$f_V(v)=\frac{C_2}{d_0(\sqrt{2\pi})^{n-1}}\exp\left[\left(-\frac{v}{d_0}\right)^2\right]\times\int_{t_{n-2}}^{s_{n-2}}\int_{t_{n-3}}^{s_{n-3}}\cdots\int_{t_1}^{s_1}\exp\left(-\frac{1}{2}\sum_{i=1}^{n-2}v_i^2\right)\times\mathrm{d}v_1\mathrm{d}v_2\cdots\mathrm{d}v_{n-2} \quad (36)$$

根据式(13),$Z=(\overline{X}-\mu)/S$ 的概率密度函数由下式计算:

$$f_Z(z)=\int_0^{+\infty}\frac{\sqrt{n}d_0C_2v_{n-1}}{\sqrt{2\pi}}\cdot\exp\left[-\frac{1}{2}n\times(d_0zv_{n-1})^2\right]f_{V_{n-1}}(v_{n-1})\mathrm{d}v_{n-1} \quad (37)$$

2.5 计算结果与模拟结果的比较

为检验本文推导的 S/σ 和$(\overline{X}-\mu)/S$ 概率密度函数的正确性,进行了 Monte-Carlo 模拟.模拟的方法是,首先假定正态总体变量的平均值 μ 和标准差σ,然后产生容量为 n 的样本$x_1,x_2,\cdots,x_n$,将样本值由小到大排序,得到 $x_{(1)}\leqslant x_{(2)}\leqslant\cdots\leqslant x_{(n)}$,由式(3)和(4)计算$\overline{x}$和$s$,进而计算$s/\sigma$和$(\overline{x}-\mu)/s$.重复上述过程多次(不少于 10 000 次).对模拟结果按划分的区段进行统计处理,得到模拟的 S/σ 和$(\overline{X}-\mu)/S$ 的频数.图 1 给出了 $n=2$ 和 5 时由式(36)计算的概率密度函数与模拟结果的比较,图 2 和图 3 给出了 $n=2$ 和 5 时由式(37)计算的概率密度函数与模拟结果的比较.由这些图可以看出,计算结果与模拟结果符合得很好,说明本文推导的计算公式是正确的.图 4 为 $n=2\sim7$ 时由式(37)计算的概率密度曲线.

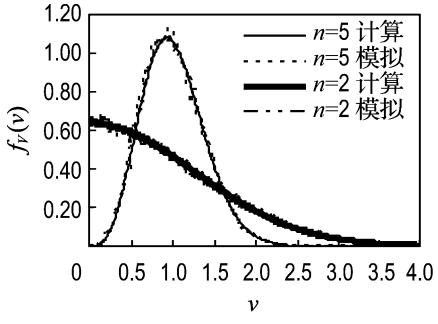

图 1 计算和模拟的 S/σ 概率密度函数的比较

Fig. 1 Comparison of probability density value of S/σ between calculated and simulated

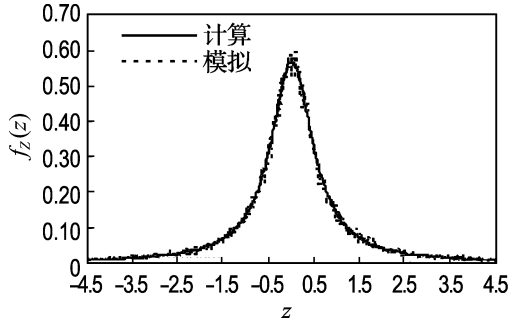

图 2 $n=2$ 时计算和模拟的$(\overline{X}-\mu)/S$ 概率密度函数的比较

Fig. 2 Comparison of probability density value of $(\overline{X}-\mu)/S$ between calculated and simulated for $n=2$

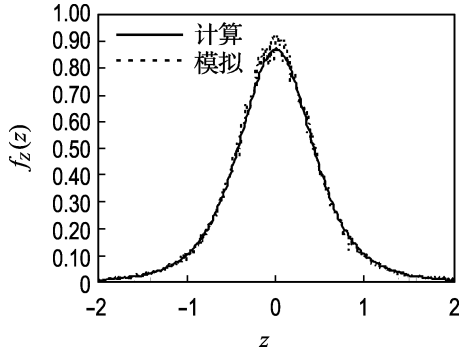

图 3 $n=5$ 时计算和模拟的$(\overline{X}-\mu)/S$ 概率密度函数的比较

Fig. 3 Comparison of probability density value of $(\overline{X}-\mu)/S$ between calculated and simulated for $n=5$

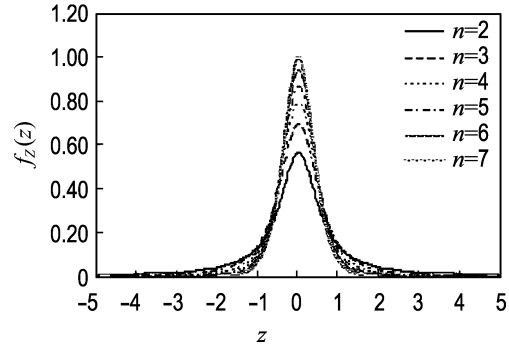

图 4 $n=2\sim7$ 时$(\overline{X}-\mu)/S$ 的概率密度函数

Fig. 4 Probability density value of $(\overline{X}-\mu)/S$ for $n=2\sim7$

3　在桩极限承载力分析中的应用

在工程设计中，材料和土工参数的代表值是用标准值定义的[9~11]. 混凝土和钢材标准值的保证率为 95%，而土工参数标准值一般取平均值. 由于样本容量很小时桩承载力的平均值是不确定的，估计桩的承载力时应考虑由小样本产生的不确定性. 假定桩的承载力服从正态分布，则可将桩承载力的标准值定义为

$$Q_k = \mu = \bar{x} - z_\alpha(n-1) \cdot s \tag{38}$$

式中：$\bar{x}$ 和 s 分别由式(3)和式(4)计算；$z_\alpha(n-1)$ 为 Z 的 α- 分位点值，由下式计算：

$$P(Q > Q_k) = P[Z < z_\alpha(n-1)] = \int_0^{z_\alpha(n-1)} f_Z(z)\mathrm{d}z = 1-\alpha \tag{39}$$

式(39)表示用式(38)确定的桩极限承载力的保证率为 $1-\alpha$. 表 2 给出了 $n = 2 \sim 7$，$1-\alpha = 0.75$、0.90 和 0.95 时的 $z_\alpha(n-1)$ 值，这 3 个 $1-\alpha$ 值是国际上常用的概率值[10,11].

表 2　不同保证率 $1-\alpha$ 和不同样本容量时的 $z_\alpha(n-1)$

Tab. 2　Values of $z_\alpha(n-1)$ in various size and accepted probability $1-\alpha$

n	$1-\alpha$		
	0.75	0.90	0.95
2	0.56	1.75	3.61
3	0.41	0.96	1.50
4	0.35	0.76	1.09
5	0.31	0.66	0.92
6	0.29	0.59	0.81
7	0.27	0.55	0.74

如果桩的静载承载力服从对数正态分布，这样，静荷载试验结果按下式计算：

$$\overline{\ln x} = \frac{1}{n}(\ln x_{(1)} + \ln x_{(2)} + \cdots + \ln x_{(n)}) = \frac{1}{n}\sum_{i=1}^{n} \ln x_i \tag{40}$$

$$\ln s = b_1 \ln x_{(1)} + b_2 \ln x_{(2)} + \cdots + b_n \ln x_{(n)} = \sum_{i=1}^{n} b_i \ln x_{(i)} \tag{41}$$

类似于式(38)，可确定保证率为 $1-\alpha$ 时桩静载承载力对数的标准值：

$$\ln Q_k = \overline{\ln x} - z_\alpha(n-1) \cdot \ln s \tag{42}$$

保证率为 $1-\alpha$ 时桩静载承载力的标准值为

$$Q_k = \exp(\overline{\ln x} - z_\alpha(n-1) \cdot \ln s) \tag{43}$$

4　算　例

某码头水工结构采用预应力混凝土桩，设计桩长 41.9 m，桩径 600 mm. 进行了两根桩的静荷载试验. 按照我国现行《港口工程桩基规范》(JTJ 254—98)[4]的规定进行判断，两根试验桩的极限承载力分别为 4 382 和 3 987 kN. 按 75%和 95%的保证率确定桩的极限承载力标准值.

根据试验结果进行排列：$x_{(1)} = 3\,987$ kN，$x_{(2)} = 4\,382$ kN. 按式(40)和(41)进行计算：

$$\overline{\ln x} = \frac{1}{2}(\ln x_{(1)} + \ln x_{(2)}) = \frac{1}{2} \times (\ln 3\,987 + \ln 4\,382) = 8.338\,0$$

$$\ln s = b_1 \ln x_{(1)} + b_2 \ln x_{(2)} = -0.886\,2 \times \ln 3\,987 + 0.886\,2 \times \ln 4\,382 = 0.083\,7$$

(1)保证率为 75%的情况

由 $n = 2$，$1-\alpha = 0.75$，查表 2 得 $z_\alpha(n-1) = 0.56$. 根据式(43)：

$$Q_k = \exp(\overline{\ln x} - z_\alpha(n-1) \cdot \ln s) = \exp(8.338\,0 - 0.56 \times 0.083\,7) = 3\,988.330\,5 \text{ kN}$$

(2)保证率为 95%的情况

由 $n = 2$，$1-\alpha = 0.95$，查表 2 得 $z_\alpha(n-1) = 3.61$. 根据式(43)：

$$Q_k = \exp(\overline{\ln x} - z_\alpha(n-1) \cdot \ln s) = \exp(8.338\,0 - 3.61 \times 0.083\,7) = 3\,089.742\,4 \text{ kN}$$

5　结　语

数据的小样本是工程设计和评估中经常遇到的问题，目前有多种不同的分析方法. 本文针对文献[1]提出的参数线性最小偏差估计方法，给出了样本分布的概率密度函数，可用于工程中小样本数据的分析和处理.

参考文献：

[1] 茆诗松，程依明，濮晓龙. 概率论与数理统计教程[M]. 北京：高等教育出版社，2004

[2] 中华人民共和国建设部. GB 50007—2002 建筑地基基础设计规范[S]. 北京：中国建筑工业出版社，2002

[3] 路桥集团第一公路工程局. JTJ 041—2000 公路桥涵施工技术规范[S]. 北京：人民交通出版社，2000

[4] 交通部第三航务工程勘察设计院. JTJ 254—98 港口工程桩基规范[S]. 北京：人民交通出版社，1998

[5] 中国建筑科学研究院. JGJ 106—2003 建筑基桩检测技术规范[S]. 北京：中国建筑工业出版社，2003

[6] CEN. EN 1997—1：2004 Eurocode 7：Geotechnical design—Part 1：General Rules [S]. London：BSI，2004

[7] LIEBLEIN J. NACATN 3053 A new method of analyzing extreme-value data [S]. MA：NACA，1954

[8] 日本腐蚀防蚀协会. 装置材料的寿命预测入门[M]. 陈家福，周 卫，翁 端，译. 大连：大连理工大学出版社，1991

[9] 中华人民共和国住房和城乡建设部. GB 50153—2008 工程结构可靠性设计统一标准[S]. 北京：中国建筑工业出版社，2008

[10] International Organization for Standardization. ISO 2394：1998 General Principles on Reliability for Structures [S]. New York：ANSI，1997

[11] International Organization for Standardization. ISO 12491：1997 Statistical Methods for Quality Control of Building Materials and Components [S]. New York：ANSI，1997

Probabilistic analysis of less size samples problem and its application to engineering

GONG Jin-xin*， LI Rong-qing， WANG Li-huan

（State Key Laboratory of Coastal and Offshore Engineering，Dalian University of Technology，Dalian 116024，China）

Abstract：Statistical analysis is the basic technique for processing data and the basis of reliability analysis. In some instances，only least size samples could be collected due to limited test conditions and expenses，for example，the number of tested piles for loading is no more than four. So it is important to find a way to deal with the data. In view of the efficiency of parameter evaluation method of linear least variance for less size samples problem，the probabilistic distributions of the evaluated parameters were derived and its application to determining the characteristic limit bearing capacity of tested piles in foundation engineering was discussed. The results show that it is feasible for analyzing and processing data of less size samples in engineering.

Key words：statistics；less size samples；reliability；piles