# 标准二阶矩在环境质量评价中的应用

张 益 智
(吉林省环境保护研究所)

本文依照统计数学的基础理论，用矩法通过正态分布引出二阶原点矩公式，并与样本二阶原点矩相结合，得出标准二阶矩公式，用样本的统计特性来进行水质的环境质量评价．本公式也可以用于大气、土壤等环境质量评价．

## 一、矩的概念

在统计数学里，对点的估计有一种方法为矩法估计．在一般情况下，使用矩法估计可获得渐近正态的估计量． 期望值 $E_\xi$ 是理想的平均数，而分布的矩 $V_r$ 是具有该分布随机变量的幂的期望值．

若 $\xi$ 是离散的随机变量，其可能取值为 $X_n$ $(n=1,2,3\cdots\cdots)$，而 $\xi$ 取值的概率分布为 $p(\xi=X_n)=p_n$，则

1. $\xi$ 的 $r$ 阶原点矩定义为

$$V_r=E\xi^r=\sum_n X_n^r p_n \tag{1.1}$$

当 $r=1$ 时，(1.1)式描述了 $\xi$ 的取值中心．

2. $\xi$ 的 $r$ 阶中心矩定义为

$$u_r=E(\xi-E\xi)^r=\sum_n (X_n-E\xi)^r p_n \tag{1.2}$$

当 $r=2$ 时，(1.2)式描述了 $\xi$ 的可能取值与中心偏差的疏密程度．

3. $\xi$ 的矩母函数定义为随机变量函数 $e^{t\xi}$ 的均值

$$u(t)=\sum_n e^{tX_n}p_n \tag{1.3}$$

若把矩母函数对 $t$ 微分 $r$ 次，再令 $t=0$ 代入，就得到 $r$ 阶矩．

## 二、正态分布中的二阶矩

正态分布的矩母函数是

$$u(t)=e^{ut+\frac{\sigma^2t^2}{2}} \tag{2.1}$$

其一阶导数是

$$u'(t)=e^{ut+\frac{\sigma^2t^2}{2}}\cdot(u+\sigma^2t)$$

将 $t=0$ 代入

$$u'(\theta)=V_1=u \tag{2.2}$$

其二阶导数是

$$u''(t)=e^{ut+\frac{\sigma^2t^2}{2}}\cdot[(u+\sigma^2t)^2+\sigma^2]$$

将 $t=0$ 代入

$$u''(0)=V_2=u^2+\sigma^2 \tag{2.3}$$

在求出(2.2)式后，也可以用对任何分布都成立的公式

$$u_2=u''(0)-[u'(0)]^2 \tag{2.4}$$

来得到(2.3)式．

## 三、样本的标准二阶矩

从总体中抽取一部分样品 $X_n$ $(n=1,2,3,4\cdots\cdots m)$ 称为总体的一个样本，样本中样品的个数称为样本的容量．当样本的容量足够大时，我们可以假设样本矩和总体矩相等，这样就把样本特征数与总体特征数联系起来，这时有:

1. 样本的算术平均值

$$\overline{X}=u=\frac{1}{m}\sum_{n=1}^{m}X_n \tag{3.1}$$

2. 样本的方差

$$S^2=\frac{1}{m}\sum_{n=1}^{m}(X_n-\overline{X})^2 \tag{3.2}$$

上式中的分母是 $m$，而不用 $(m-1)$ 那种样本修正方差.

3. 样本的二阶原点矩是

$$V_2=\frac{1}{m}\sum_{n=1}^{m}X_n^2 \qquad (3.3)$$

样本的特征数与样本的分布无关，我们可以把 (3.3) 式同 (2.3) 式联系起来

$$\frac{1}{m}\sum_{n=1}^{m}X_n^2=\overline{X}^2+S^2 \qquad (3.4)$$

在 (3.3) 式中，由于 $X_n$ 是二次方，因次比原数据提高了，我们可以把 (3.3) 式开方变为

$$V=\sqrt{V_2}=\sqrt{\frac{1}{m}\sum_{n=1}^{m}X_n^2} \qquad (3.5)$$

这时，(3.4) 式将随之变为

$$V=\sqrt{\frac{1}{m}\sum_{n=1}^{m}X_n^2}=\sqrt{\overline{X}^2+S^2} \qquad (3.6)$$

在 (3.6) 式中，第一个等号后面是二阶原点矩的开方，第二个等号后面的 $S$ 是二阶中心矩，我们可以把 (3.6) 式概括称为二阶矩。所得结果的因次与原数据相同，依照方差的平方根称为标准差，我们可以把 (3.5) 和 (3.6) 式称为标准二阶矩公式.

(3.6) 式把一个样本中的两个特征数联系起来，一个是 $\overline{X}$，它表示了样本的中心位置，另一个是 $S$，它表示了样本中各数据与中心位置的离散程度，这两个特征数有着本质的差异，但它们也有共性的一面，它们的数值都是来自同一个样本. 我们把标准二阶矩放到以均值和标准差为两个轴的平面直角坐标系中. 虽然每个轴代表了不同质的量，但这两个轴单位矢量的长度相同，这就能得到用单位矢量测得的标准二阶矩的模式.

## 四、标准二阶矩在环境质量评价中的意义

在污染状况调查中，遇到的多数情况是人群的 50% 以下受到毒害. 在这种情况下，当以污染物剂量为横轴、人群的感受为纵轴，则剂量感受曲线可表示为指数大于 1 的幂函数曲线，即斜率是逐渐增加的. 在 (3.5) 式中，由于各数据采用的是 2 次方，就体现了污染物剂量与人类感受曲线是斜率逐渐增加的曲线函数.

在环境质量评价过程中，要把监测数据除以标准浓度来体现毒物与生态的统一. 在运算综合过程中，(3.5) 式则体现了这一特点.

从 (3.6) 式可以看出，标准二阶矩不仅体现了一个样本的中心位置，也体现了离散度，是一个样本中两个数字特征 $\overline{X}$ 和 $S$ 的矢量和. 从污染的意义上比较几个样本的轻重时，就不仅比较了各样本的中心位置，同时又比较了离散度. 在离散度相等的情况下，均值 $(\overline{X})$ 大的样本，污染就重一些；均值 $(\overline{X})$ 相同的样本，离散度大的样本污染就重一些.

当 $a$ 和 $b$ 两个样本的标准二阶矩相同时，如果 $a$ 样本中的 $\overline{X}_a$ 大于 $b$ 样本中的 $\overline{X}_b$，则 $b$ 样本的 $S_b$ 将相应的大于 $a$ 样本中的 $S_a$. 我们可以设想一种特殊情况，$a$ 样本中的数值完全相同，这时 $S_a=0$；$b$ 样本的数值不完全相同，这时 $S_b>0$. 而 $a$ 和 $b$ 两个样本的标准二阶矩相同，这时有

$$\overline{X}_b^2+S_b^2=\overline{X}_a^2 \qquad (4.1)$$

$$\overline{X}_b+S_b>\overline{X}_a \qquad (4.2)$$

由于 $\overline{X}_b$ 和 $S_b$ 都是 $b$ 样本中的各个数据的平均形式，而又不相等，则在 $b$ 样本中必有一个或若干个数据大于 $a$ 样本中的所有数据. 比较两个样本的算术平均值，用 $\overline{X}_b<\overline{X}_a$ 的结果得出的 $b$ 样本污染轻，就忽视了 $b$ 样本中一个或若干个大于 $a$ 样本中的所有数据的情况. 从污染角度讲，这一个或若干个较大数据造成的危害是绝不能轻视的.

## 五、标准二阶矩在水质评价中的应用

某条河流的一个截面 $(j)$，年际间污染物浓度变化不大. 从这种系统中得到的一种污染物在不同时间取得的许多数据 $C_t$ $(t=$

$1, 2, 3 \cdots\cdots h$)，把它们换算为相对污染值 $X_t = \frac{C_t}{C_0}$，式中 $C_0$ 为该污染物的标准浓度. $X_t$ 可被看成容量为 $h$ 的一个样本. 我们不知道这个样本属于什么分布，但可以研究这个样本的均值、方差等数字特征.

$j$ 截面、$i$ 污染物的标准二阶矩是:

$$V_{ij} = \frac{1}{C_0}\sqrt{\frac{1}{h}\sum_{t=1}^{h} C_t^2} \qquad (5.1)$$

我们用标准二阶矩估计 $j$ 截面的污染程度为:

$$V_j = \sqrt{\frac{1}{m}\sum_{i=1}^{m} V_{ij}^2} \qquad (5.2)$$

我们用标准二阶矩估计一种污染物在河流上的污染程度为:

$$V_i = \sqrt{\frac{1}{n}\sum_{j=1}^{n} V_{ij}^2} \qquad (5.3)$$

当我们把 $j$ 截面的各种污染物同全河流中同种污染物联系起来考虑时，用相关性来看这个截面的污染轻重时可有相关性标准二阶矩

$$V_j^g = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (V_{ij} \cdot V_i)^2} \qquad (5.4)$$

## 六、标准二阶矩与其它公式比较

某一点的同一种污染物在不同时间的浓度不同，对生态的影响也不一样，剂量感受曲线是一种斜率逐渐增加的曲线，而算术平均值只是线性函数的平均，用它来计算不同时间的平均值，既掩盖了极大值和次大值，又没有把毒物和生态有机的结合在一起. 在 (5.1) 式中，在加和之后除以项数，结果也是一种平均值. 用 (5.1) 式代替算术平均值，就能体现剂量对生态的影响，又能分辨出算术平均值相等而极大值或次大值不相同的情况.

在北京西郊等地的环境质量评价中，在把不同的污染物进行综合时，采用了简单的算术平均值模式. 这种模式的优点是计算简便、直观，缺点是把不同质的污染物作为同类项合并，并掩盖了极大值、次大值等对生态的影响. (5.2) 式继承了计算简便、直观等优点 (现在普及的小计算器中，都有 $\Sigma X^2$ 键)，又合理的扬弃了算术平均值不能体现极大值、次大值和把不同质的量作为同类项合并的缺点.

由于算术平均值不能体现较大值对生态的影响，在内梅罗公式中，把极大值单独拿出来，放在与同一个样本中算术平均值同等的地位，进行运算，等于把极大值加了个很大的权，这个权又是随样本的容量增加而加大的. 由于样本中的各个项是随机的，容量较大时，就过分夸大了极大值而掩盖了次大值、再次大值等对生态的影响.

申葆诚等提出的矢量公式，把各项污染物作为不同质的量处理. 这种公式不是一种均值形式，计算结果随项数增加而增大 (在数值大于 0 时)，只能用在样本容量相同时才能比较. (3.5) 式具有算术平均值的均值优点，又具有矢量公式处理不同质的量的优点，它可以处理同一地点、同一种污染物在不同时间取得的数据，又能处理不同的污染物或是不同地点为一个样本的数据. 在处理过程中，又把污染物同生态结合起来考虑.

## 参 考 文 献

[1] 数学手册编写组，数学手册，人民教育出版社，1979年.
[2] A. M. 穆德等，统计学导论，科学出版社，1978 年.
[3] 刘多森，环境科学，2, 71(1981).
[4] 申葆诚等，环境科学，4, 75(1981).