# 水文地质手册

SHUIWENDIZHI SHOUCE

# 水文地质手册

主编：刘正峰

第四卷



银声音像出版社

# 第九章　水文统计

## 第一节　水文统计的基本概念

水文资料是水文现象的一种物质表现形式，通过对水文资料的进一步分析，可以揭示水文现象所隐藏的客观规律。

### 一、水文现象的统计规律

水文现象是一种自然现象，它的产生、发展和演变过程，既有必然性的一面，又有偶然性的一面。例如一条河流每年必有一个最大的洪峰流量，它是由水文循环所决定的，因此是必然的；而各年的最大洪峰流量数量的大小，何时发生，这又是未知的，带有一定的偶然性（数学上称为随机性），因此可以说水文现象也是一种随机现象。对于随机现象，从表现上看似乎是无规律的，但分析大量实测水文资料后便知，它是遵循一定规律的。如河流某断面年径流量的多年平均值是一个比较稳定的数值，在长期过程中，河流的径流量接近于多年平均值的年份出现较多，而特大或特小值出现的年份则比较少。随机现象的这种规律性，只有通过大量的观测、统计、分析后才能发现，故称之为统计规律。而数理统计法正是揭示这种规律的一种有力工具，所以也就成为水文分析计算的主要方法，在水文学中称之为水文统计。

### 二、随机变量、总体和样本

在日常生活中我们会遇到各种各样的试验，如科学种田试验，导弹发射试验等。可以说试验具有广泛的意义。在概率论中提出这样一种试验：①可以在相同的条件下重复进行；②每次试验的可能结果不止一个，并且事先知道试验所有可能出现的结果或范围；③每次试验之前无法确定究竟哪种结果会出现。如掷硬币、掷骰子、摸扑克牌等等均是如此。具有这种特性的试验称为随机试验。随机试验的结果叫事件，事件可以分为两大类：

在一定条件的组合下，必然会发生的事件，称为必然事件；肯定不发生的事件，称为不可能事件。它们发生与否在试验前就可确定，因此属于确定性事件。另一类是在一定条件的组合下，可能发生也可能不发生的事件称为随机事件，简称事件。例如在一次掷骰子中，设事件为“2点出现”，掷骰子时该事件有可能发生，也有可能不发生，这种事先不能确定的事件就是随机事件。水文观测也可以看成是随机试验，观测到的结果是随机事件。必然事件和不可能事件本来没有随机性，但为了研究的方便，把它们看成是随机事件的特例。

为了便于分析事件的统计特性，通常用变量来表示事件，即随机试验的结果。每次试验后事件发生与否就相当于给这个变量赋值，这个变量就叫随机变量。

通常用大写字母表示随机变量，其取值用相应的小写字母表示。如某随机变量为 $X$，其取值可以记为 $x_i$，$i=1,2、\cdots、n$。在上面掷骰子例子中，掷骰子的结果用变量 $X$ 表示，$X=2$ 代表“2点出现’的事件发生。一般将 $x_1、x_2、\cdots、x_n$ 称为随机系列，简称系列。

在数理统计中，把随机变量所有取值的全体称为总体。从总体中任意抽取的一部分称为样本，样本的项数称为样本容量。掷骰子试验中。随机变量总体的取值共有6种情况。而水文变量的总体都是无限的，实际上无法获得。例如某地的年降水量是随机变量，其总体是指自古到今以至未来无限年代的所有降水量，人类是不可能全部观测得到的。目前设站所观测到的几十年甚至是上百年降水量资料，只不过是总体中的一小部分，是一个很有限的样本。因此，我们通过各种途径收集来的水文资料都是样本资料。

总体和样本之间既有区别又有联系。由于样本是总体中的一部分，因而样本的特征在一定程度上（或部分地）反映了总体的特征，故总体的规律可以借助样本来逐步地认识。这就是我们目前用已有水文资料来推断总体或预估未来水文情势的依据。但样本毕竟只是总体中的一部分，当然不能完全代表总体的情况，其中存在着一定的差别。这种差别就是我们后面将要介绍的抽样误差。可以相信，随着科学技术的发展，随着资料和经验的积累，总体的规律将愈来愈为人们所认识。

### 三、选样

水文过程是随时间变化的，具有周期性、趋势性和持续性的随机过程，目前人们还不能构造成一个复杂的随机模型来准确地模拟水文变量的变化过程。现行水利水电设计中的水文分析计算方法，基本上都是把水文过程看成是不含周期性、趋势性和持续性的纯随机过程。为此，首先以“年”为时段把实际连续的水文过程离散化，以消除水文变量在年内随季节波动的影响。为了进一步消除相邻年际间河川径流的相关关系，在年径流计算时，又改用水文年（以水文现象的循环规律来划分，即从每年汛期开始起到下一年的汛期开始止）或水利年（以水库蓄泄水的周期来划分，即从每年水库蓄水之日起到下一年蓄水之日止）作为划分时段。使逐年出现的水文过程尽量与次序无关，即随机独立。

其次，由于每年的水文过程也是一次复杂的随机事件，目前只能简化成用水文特征值（水文变量）来从不同角度刻画其主要特征，如时段径流量、洪峰流量等。并按成因相同，条件一致的原则从中选出水文特征值组成年样本系列，作为统计分析的对象，这个过程称为选样。

组成年样本系列的水文特征值的大小与统计时段有关。统计时段分设计时段和控制时段两种。能够完整地反映设计流域水文特征的最长统计时段为设计时段，在设计时段之内划定的，用以细致刻画和控制设计时段内水文变化过程的较短时段为控制时段。设计时段和控制时段应根据工程要求和流域水文特性而定。如在分析设计流域河川来水量时，通常以年作为设计时段，根据灌溉或发电要求，选灌溉期或枯水期为控制时段；在分析暴雨时，如根据历年暴雨资料选定某小流域暴雨的设计时段为 24h，则可取 1h、3h、6h 等为控制时段。控制时段的多少以能控制设计时段内的水文过程为原则。

为了使样本资料满足随机独立性的要求，并考虑最不利情况，一些水文特征值的选样通常按"年极值"来挑选，如年最大洪峰流量；年最大 1h、年最大 6h、年最大 24h 雨量；年最小 1 个月、年最小 3 个月枯季径流量等。每年独立选取水文特征值，组成各自的年样本系列，作为水文统计计算的基础，水文变量的选样如图 12－9－1 所示。

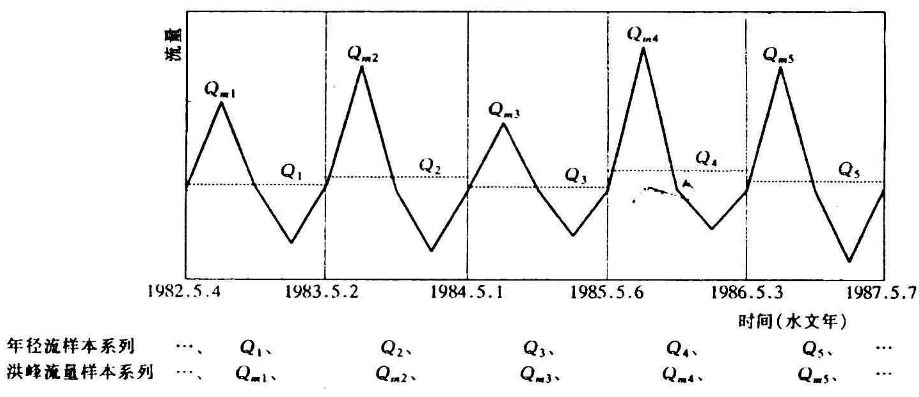


图 12－9－1　水文选样示意图

水文统计的任务，就是通过数理统计的途径和方法，对上述观测所取得的样本系列进行统计分析，寻求其统计规律。其实质就是由水文样本特征估计水文总体的特征，由此推测同属水文总体一部分的工程未来运用期的水文情势。

## 第二节　样本的统计参数和抽样误差

由总体和样本的关系可知，样本是总体的一部分，样本特征对总体特征有一定的代表

性。一般从样本系列的数字特征和频率特征两个侧面揭示总体的特征，并用抽样误差来衡量样本对总体代表性的优劣程度。

## 一、样本的统计参数

水文样本系列是水文随机变量的一组取值。水文随机变量取值应有一定的规律，这种规律的数字特征叫统计参数。样本系列常用的统计参数有：算术平均数、均方差、变差系数和偏差系数。

1. 算术平均数($\bar{x}$)

设随机变量的样本系列为 $x_1$、$x_2$、…、$x_n$，则其算术平均数可用下式计算：

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad (9-1)$$

算术平均数简称均值，它表示样本系列的平均情况，反映系列总体水平的高低。例如，甲乙两条河流的多年平均流量分别为 1000m³/s 和 100m³/s，就说明甲河流域的水资源比乙河流域的丰富得多。

2. 均方差与变差系数

(1)均方差($\sigma$)。均方差 $\sigma$ 表示系列各个取值相对于均值的离散程度。计算公式为

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x}^2)}{n-1}} \qquad (9-2)$$

例如有甲乙两个系列，其值为

甲系列：5，10，15

乙系列：2，10，18

两个系列的均值都是 10，但两个系列中各个取值相对于均值的离散程度显然是不同的，两系列的均方差可按(9—2)式计算：

$$\sigma_{甲} = \sqrt{\frac{(5-10)^2 + (10-10)^2 + (15-10)^2}{3-1}} = 5$$

$$\sigma_{乙} = \sqrt{\frac{(2-10)^2 + (10-10)^2 + (18-10)^2}{3-1}} = 8$$

可见 $\sigma$ 甲<$\sigma$ 乙，说明甲系列的离散程度小，乙系列的离散程度大。

(2)变差系数($C_v$)。均方差的大小只能衡量等均值系列的离散程度，而对于均值不等的系列，为了消除均值的影响，用均方差与均值的比值，即变差系数 $C_v$ 来比较。样本系列变差系数的计算公式为

$$C_v = \frac{\sigma}{\bar{x}} = \frac{1}{\bar{x}}\sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} \qquad (9-3)$$

例如有丙、丁两个系列，其值为

丙系列：5，10，15

丁系列：995，1000，1005

两系列的均值分别为 $\bar{x}_{丙}=10$，$\bar{x}_{丁}=1000$，均方差分别为 $\sigma_{丙}=5$、$\sigma_{丁}=5$，用式(9－3)求得两系列的变差系数分别为 $C_{v丙}=0.5$、$C_{v丁}=0.005$，说明丙系列的离散程度远比丁系列的离散程度大。

在水文学中，计算出某样本系列的变差系数后，通常需要分析其在地区上的变化规律，和邻近流域的变差系数进行比较，所以经常使用变差系数来反映样本系列的离散程度。

3. 偏差系数($C_s$)

偏差系数 $C_s$ 也称偏态系数，它是反映系列中各值在均值两侧对称程度的一个参数。

计算公式为

$$C_s=\frac{\sum_{i=1}^{n}(x_i-\bar{x})^3}{(n-3)\bar{x}^3C_v^3} \qquad (9-4)$$

当样本系列中各值在均值两侧对称分布时，$C_s=0$，称为正态分布。若分布不对称时，$C_s\neq 0$，称为偏态分布。其中 $C_s>0$，它表示随机变量大于均值的取值比小于均值的取值机会少，称为正偏分布。反之，$C_s<0$，称为负偏分布。水文现象大多属正偏分布。

样本的统计参数 $\bar{x}$、$\sigma$、$C_s$ 可由电子表格 Excel 软件的内置函数 AVERAGE、SIDEV 和 SKEW 直接算出。

## 二、抽样误差

用式(9－1)～式(9－4)计算出来的都是样本的统计参数，它们与总体统计参数是有区别的。

对于水文现象而言，几乎所有的水文变量的总体都是无限的，目前掌握的资料仅仅是一个容量十分有限的样本，这样要由样本的统计参数去估计总体的统计参数，总会存在一定的误差，这种误差是由随机抽样而引起的，故称之为抽样误差。

根据数理统计的理论和实践经验，样本统计参数的抽样误差一般随样本的均方差 $\sigma$、变差系数 $C_v$ 及偏差系数 $C_s$ 的增大而增大；随样本容量 $n$ 的增大而减小。因此一般来讲，样本系列越长，抽样误差将愈小，样本对总体的代表性也就愈好。反之，样本系列愈短，抽样误差愈大，样本对总体的代表性也就越差。所以在水文分析过程中，一般要求样本容量要有足够长度。

用式(9－1)～式(9－4)估算参数的方法称为矩法估计。经验表明，矩法估算参数，除

了有上述抽样误差外，还有一定的系统误差。因此，在实际的水文分析计算中，通常不直接使用矩法估计的参数，而是以矩法公式计算的参数作为初始参数值，然后经过频率适线法来确定。这种方法是我国水文界目前广泛使用的一种方法，将在后面详细介绍。

# 第三节　水文样本的频率分析

水文样本系列，不仅可以用样本的统计参数来描述随机变量取值的数字特征，还可以用随机变量的频率分布来进一步刻画。

## 一、概率、频率和重现期

### (一)概率

随机事件在试验结果中可能出现也可能不出现，但其出现(或不出现)的可能性的大小是不同的。为了比较这种可能性的大小，必须赋予一种数量指标，这个数量指标就是事件的概率。

简单随机事件 $A$ 发生的概率，计算如下：

$$(A) = \frac{m}{n} \tag{9-5}$$

式中　$P(A)$——一定条件下随机事件 A 发生的概率；

$m$——属于随机事件 $A$ 的结果数；

$n$——试验中所有可能出现的结果数。

例如掷骰子试验，所有可能出现的结果数 $n=6$，即可能出现 1、2、3、4、5、6 点。设事件 $A$ 表示为“3 点出现”，则所有可能出现的六种结果中属于 3 点的结果数为 $m=l$，因此“3 点出现”的概率为 $P(A)=\frac{m}{n}=\frac{1}{6}$；若事件 $A$ 表示“大于 3 点的点数出现”，则属于事件 $A$ 的可能出现的结果数为 m=3(即 4、5、6 点出现)，因此 $P(A)=\frac{5}{6}$。假若将骰子的六个面全部刻成 3 点，则 3 点出现的结果数为 $m=n=6$，此时，$P(A)=1$，该事件为必然事件；除 3 点以外其它各点数出现的结果数为 0，也就是 $m=0$，$P(A)=0$，说明其它各点数的出现是不可能事件。由此可以得出随机事件 $A$ 的概率介于 0 和 1 之间，即 $0\leqslant P(A)\leqslant 1$。

式(9－5)只适用于所谓的“古典概型事件”，即试验的所有可能结果都是等可能的，且试验中所有可能出现的结果总数是有限的简单随机事件。而对于水文上的复杂随机事

件而言，试验所有可能结果数 $n$ 是无限的，无法知道。因此，其出现的可能性大小无法用古典概型的概率来描述，这便引出了频率的概念。

（二）频率

设随机事件 $A$ 在 $n$ 次随机试验中，实际出现了 $m$ 次，则

$$P(A)=\frac{m}{n} \tag{9-6}$$

为事件 $A$ 在 $n$ 次试验中出现的频率，简称为事件 $A$ 的频率。

实践证明，当试验次数 $n$ 较少时，事件的频率很不稳定，有时大、有时小；但当试验次数 $n$ 无限增多时，事件的频率就逐渐趋近于一个稳定值，这个稳定值便是事件发生的概率。如掷硬币试验，从理论上讲，正面（或反面）出现的概率为 0.5（即 $n=2, m=1, p(A)=\frac{1}{2}$），以前曾有人做过掷硬币试验 4040 次、12000 次和 24000 次，分别统计正面出现的次数为 2048 次、6019 次和 12012 次，相应频率为 0.5080、0.5016 和 0.5005。可见，随着试验次数的增多，频率越来越接近于事件的概率 0.5。

随机事件频率的这种特性，不仅为大量实践所证实，而且在概率论中也有严格的理论证明。因此对于复杂的随机事件，可以用频率近似地代替概率来描述其发生的可能性，从这个意义上讲也要求试验次数 $n$ 要足够大。

水文现象中的各种随机变量如降水、径流、洪水等的变化均属于复杂事件，生产上的水文频率通常是指水文随机变量大于或等于某一固定值这一随机事件发生的可能性。例如，河流某断面洪水的频率 $p=5\%$，表示该断面的洪水大于或等于某一给定值这一事件发生的可能性是 5%。

（三）重现期

由于频率是概率论中的一个概念，比较抽象，因此水文分析计算中常用“重现期”来代替频率表示随机事件（或随机变量取值）出现的机会。

所谓重现期，是指某随机事件在长期过程中平均多少年出现一次，即“多少年一遇”，用 $T$ 表示。例如，某水文随机变量大于或等于某值的频率 $P=5\%$，表示该随机事件平均 100 年可以出现 5 次，或平均 20 年出现 1 次，亦即重现期 $T=20$ 年，称为“20 年一遇”。

频率与重现期的关系，在不同的情况下有不同的表示方法。

在研究暴雨、洪水时，水文变量大于或等于某值事件发生的频率 $p<50\%$，该事件的重现期为

$$T=\frac{1}{p}(\text{年}) \tag{9-7}$$

例如，某防洪工程设计依据洪水频率 $p=5\%$，则重现期 $T=\frac{1}{0.05}=20$ 年，即洪水或该工

程的防洪标准是 20 年一遇。表示长期过程中，平均 20 年出现 1 次大于或等于该级别的洪水。

在灌溉、发电、供水工程规划设计时，需要研究枯水问题，水文变量大于或等于某值的频率 $p>50\%$，则水文变量小于某值，即枯水事件发生的频率为 $1-p$，枯水事件的重现期为

$$T=\frac{1}{1-p}(\text{年}) \tag{9-8}$$

例如，为保证灌区的用水需求，某灌区的设计依据为径流大于或等于某一值的频率 $p=95\%$，则径流小于该值，灌区用水遭到破坏，即枯水事件发生的频率为 1－95%，那么枯水的重现期 $T=\frac{1}{1-0.95}=20$ 年，表示平均在 20 年中有 1 年供水不足，其余 19 年用水可以得到保证。鉴于此，在灌溉、发电、供水工程规划设计时，常把所依据的径流频率（P＞50%）称为保证率，可理解为兴利用水得到保证的概率。

必须指出，因为水文现象一般并无固定的周期，所谓“多少年一遇”，是指长期过程中的平均情况。如百年一遇的洪水，表示大于或等于某一级别的洪水平均是 100 年出现一次，但并不意味着每隔 100 年就必须会遇上 1 次。此处着重强调的是长期过程中的平均情况，对于某个具体的 100 年来说，大于或等于这个级别的洪水可能出现几次，也有可能一次都未出现。

## 二、随机变量的频率分布

水文变量基本上都是随机变量，如年径流量、降水量、最大洪峰流量等，各随机变量可以取不同的值，而且每一个取值都对应一定的频率，我们将随机变量的各个取值与其频率之间的对应关系称为随机变量的频率分布。它可以用表格和图形来表示。

【例 9－1】 已知某站 1938～2001 年共 64 年的年降水量（见表 12－9－1），试分析该样本系列的频率分布规律。

解：(1)为了使研究问题简单而实用，将年降水量分组并统计各组出现次数和累积次数，拟定分组的组距 $\Delta x=100$ mm，统计结果列于表 12－1－1 中的①、②、③、④栏。

表 12－1－1 某站年降水量表 单位：mm

| 年份 | 年降水量 | 年份 | 年降水量 | 年份 | 年降水量 | 年份 | 年降水量 | 年份 | 年降水量 |
|---|---|---|---|---|---|---|---|---|---|
| 1938 | 476 | 1951 | 285 | 1964 | 549 | 1977 | 841 | 1990 | 556 |
| 1939 | 486 | 1952 | 528 | 1965 | 702 | 1978 | 386 | 1991 | 526 |
| 1940 | 905 | 1953 | 583 | 1966 | 563 | 1979 | 565 | 1992 | 548 |
| 1941 | 207 | 1954 | 618 | 1967 | 612 | 1980 | 623 | 1993 | 627 |

| 年份 | 年降水量 | 年份 | 年降水量 | 年份 | 年降水量 | 年份 | 年降水量 | 年份 | 年降水量 |
|---|---|---|---|---|---|---|---|---|---|
| 1942 | 472 | 1955 | 388 | 1968 | 760 | 1981 | 558 | 1994 | 672 |
| 1943 | 513 | 1956 | 609 | 1969 | 658 | 1982 | 585 | 1995 | 514 |
| 1944 | 598 | 1957 | 817 | 1970 | 528 | 1983 | 784 | 1996 | 346 |
| 1945 | 580 | 1958 | 464 | 1971 | 802 | 1984 | 561 | 1997 | 530 |
| 1946 | 436 | 1959 | 626 | 1972 | 554 | 1985 | 488 | 1998 | 491 |
| 1947 | 229 | 1960 | 446 | 1973 | 643 | 1986 | 543 | 1999 | 512 |
| 1948 | 328 | 1961 | 457 | 1974 | 592 | 1987 | 629 | 2000 | 545 |
| 1949 | 331 | 1962 | 641 | 1975 | 586 | 1998 | 410 | 2001 | 545 |
| 1950 | 430 | 1963 | 481 | 1976 | 754 | 1999 | 663 | | |

表 12－9－2 某站年降水量分组统计表

| 序 号 | 年降水量分组<br>组距 Δx=100<br>(mm) | 各组出现次数<br>(次) | 累计出现次数<br>(次) | 各组出现频率<br>p(x)<br>(%) | 累计频率 p<br>(%) |
|---|---|---|---|---|---|
| ① | ② | ③ | ④ | ⑤ | ⑥ |
| 1 | 900～999 | 1 | 1 | 1.6 | 1.6 |
| 2 | 800～899 | 3 | 4 | 4.7 | 6.3 |
| 3 | 700～799 | 5 | 9 | 7.8 | 14.1 |
| 4 | 600～699 | 12 | 21 | 18.7 | 32.8 |
| 5 | 500～599 | 23 | 44 | 36 | 68.8 |
| 6 | 400～499 | 12 | 56 | 18.7 | 87.5 |
| 7 | 300～399 | 5 | 61 | 7.8 | 95.3 |
| 8 | 200～299 | 3 | 64 | 4.7 | 100 |
| 总 计 | | 64 | | 100 | |

(2)计算各组出现的频率和累积频率。各组出现的频率均按 $p(A)=\dfrac{m}{n}$ 计算，并表示成百分数。如第一组(900～999 mm)出现的频率为

$$P(900 \leqslant x \leqslant 999) = \frac{1}{64}$$

=1.6%

第二组(800～899mm)出现的频率为

$$P(800 \leqslant x \leqslant 899) = \frac{3}{64} = 4.7\%$$

……

将计算结果填入表中⑤栏。

累积次数实际含义是降水量大于或等于某个数值出现的次数。累积频率就是降水量大于或等于某个数值出现的频率，可表示为 $p(x \geqslant x_i)=\frac{m}{n}$，式中 $m$ 为累积次数，按此式计算累积频率填入表中第⑥栏。

(3)绘图。由表 12－9－2 中②栏和⑤栏绘成年降水量频率分布直方图，如图 12－9－2 所示；②栏和⑤栏绘成累积频率阶梯图，如图 12－9－3 所示。

从图 12－9－2 中可以看出，直方图呈中间高两边渐低的形状，表示 400～499 ㎜的降水量出现机会最大，而最大组和最小组降水量出现的机会均很小；从图 12－9－3 中可以看出，累积频率分布图为一阶梯状图，每一台阶的宽度也就反映了该组降水量出现的频率，其分布的规律和直方围规律是一致的。因此可以说频率分布直方图和累积频率分布图的作用是相同的，它们都反映了随机变量取值与其频率之间对应关系的分布规律，只是表现的形式不同而已。如果资料再加多，分组值 $\Delta x$ 再取小，则图 12－9－2 直方图的外包线可以接近于比较光滑的铃形曲线(图中虚线)，称为频率密度曲线；图 12－9－3 中的阶梯状图，台阶的高度也将逐渐变小，其外包线也就近似于一条 $S$ 形曲线(图中虚线)，称为累积频率分布曲线。

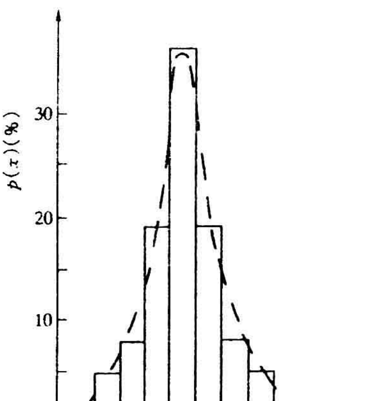


**图 12－9－2　某站降水量频率分布直方图**

在水文分析计算中，一般不绘制随机变量的频率密度曲线，而是习惯用累积频率分布曲线，简称频率曲线。

## 三、经验频率曲线

一般情况下，现有的水文资料系列较短，仅几十年，不便像上述那样分组统计，而是将

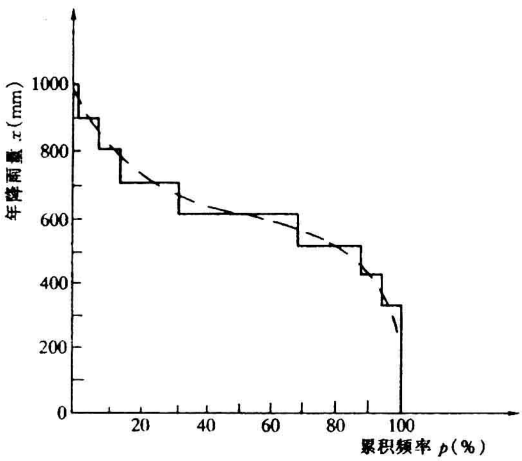


图 12—9—3　某站年降水量累积频率分布图

样本中水文变量的各个取值按由大到小的顺序排列为 $x_1$、$x_2$、…、$x_m$、…、$x_n$，其中某一取值 $x_m$ 的序号 $m$。也正是在 $n$ 次观测中，水文变量大于或等于该值的累积次数。这就可以用频率的定义式(9—5)计算水文变量大于或等于某一取值的累积频率，即经验频率。但当 $m=n$ 时，则 $P=100\%$，即表示随机变量取大于或等于样本中最小值的事件是必然事件，或者说样本的最小值一定是总体的最小值。这显然是不合乎实际的，因此用式 9—5 计算经验频率对总体来说是合适的，但对于样本系列则要进行修正。生产上计算样本的经验频率改用频率的数学期望公式计算，即

$$P=\frac{m}{n+1}\times 100\% \tag{9-9}$$

式中　$p$——随机变量取值大于或等于某值的经验频率；

$m$——系列按由大到小排序时，各值对应的序号；

$n$——观测资料的总项数，即样本系列的容量。

计算出样本各数值 $x_i$；对应的经验频率 $x_i$ 后，点绘相应的坐标点($p_i$，$x_i$)，这些经验频率点据简称经验点据，目估过点群中心绘制一条光滑的累积频率曲线，在水文计算中把该曲线称为经验频率曲线。

【例 9—2】　选取表 9—9 中有代表性的 1970～2001 年降水量资料，计算并绘制该站年降水的经验频率曲线。

解：按式(12—9—1)计算该站年降水量的经验频率(见表 12—9—3)，并将其点绘在普通坐标纸上，目估点群中心绘制的经验频率曲线，如图 12—9—4 所示。

表 12—9—3 某站年降水量经验频率计算表

| 年份 | $x_i$ (mm) | 序号 $m$ | $x_i$ (mm) | P (%) | 年份 | $x_i$ (mm) | 序号 $m$ | $x_i$ (mm) | P (%) |
|---|---|---|---|---|---|---|---|---|---|
| ① | ② | ③ | ④ | ⑤ | ① | ② | ③ | ④ | ⑤ |
| 1970 | 528 | 1 | 841 | 3.0 | 1986 | 543 | 17 | 558 | 51.5 |
| 1971 | 802 | 2 | 802 | 6.1 | 1987 | 629 | 18 | 556 | 54.5 |
| 1972 | 554 | 3 | 784 | 9.1 | 1988 | 410 | 19 | 554 | 57.6 |
| 1973 | 643 | 4 | 745 | 12.1 | 1989 | 663 | 20 | 548 | 60.6 |
| 1974 | 592 | 5 | 726 | 15.2 | 1990 | 556 | 21 | 545 | 63.6 |
| 1975 | 586 | 6 | 672 | 18.2 | 1991 | 526 | 22 | 543 | 66.7 |
| 1976 | 745 | 7 | 663 | 21.2 | 1992 | 548 | 23 | 530 | 69.7 |
| 1977 | 841 | 8 | 643 | 24.2 | 1993 | 627 | 24 | 528 | 72.7 |
| 1978 | 386 | 9 | 629 | 27.3 | 1994 | 672 | 25 | 526 | 75.8 |
| 1979 | 565 | 10 | 627 | 30.3 | 1995 | 514 | 26 | 514 | 78.8 |
| 1980 | 623 | 11 | 623 | 33.3 | 1996 | 346 | 27 | 512 | 81.8 |
| 1981 | 558 | 12 | 592 | 36.4 | 1997 | 530 | 28 | 491 | 84.8 |
| 1982 | 585 | 13 | 586 | 39.4 | 1998 | 491 | 29 | 488 | 87.9 |
| 1983 | 784 | 14 | 585 | 42.4 | 1999 | 512 | 30 | 410 | 90.9 |
| 1984 | 561 | 15 | 565 | 45.5 | 2000 | 726 | 31 | 386 | 93.9 |
| 1985 | 488 | 16 | 561 | 48.5 | 2001 | 545 | 32 | 346 | 97.0 |

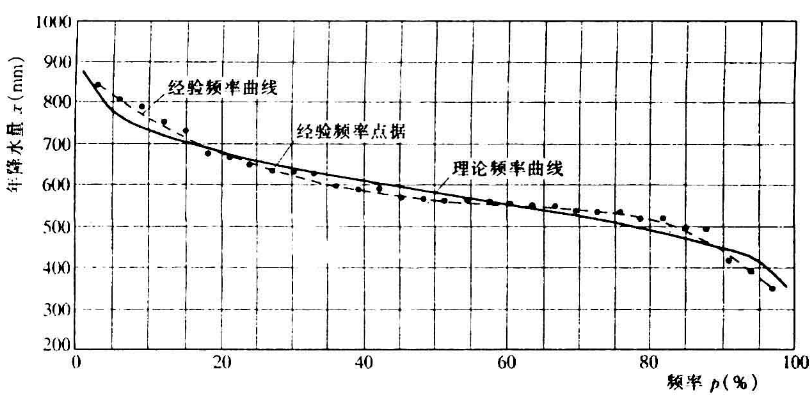


表 12—9—4 某站年降水量频率曲线图

## 四、理论频率曲线

上述经验频率曲线是过点群中心目估绘制的，曲线的形状会因人而异，尤其在经验点据分布较散时更是如此。另外，由于样本系列长度有限，据此点绘的经验频率点据往往集中在常遇频率的范围内，反映不出极小和极大频率的分布情况，若将曲线延长，则因无点子控制任意性更大，为了克服经验频率曲线的上述缺点，使样本的频率分析标准统一，便于综合比较，实际工作中常采用数学方程式来拟合经验频率点据，这种由数学方程式表达的频率曲线习惯上称为理论频率曲线。

数理统计学中有很多类型的分布曲线，究竟哪一条可以应用到水文上来，主要是看该曲线的形状与水文变量的分布规律是否吻合，其中皮尔逊Ⅲ型频率曲线应用的最为广泛，公认它和水文频率点据的配合较好。有的单位偶尔也采用前苏联的克里茨基—闵凯里频率曲线。由此可见，所谓的理论频率曲线，并不是从水文现象的物理成因方面推导出来的，它不能从根本上揭示水文现象的总体分布规律，只是作为一种数学工具，以达到规范和延长经验频率曲线的目的。

皮尔逊Ⅲ型频率曲线，简称 $P$—Ⅲ型曲线。它是英国生物学家皮尔逊于 1895 年在分析大量实际资料的基础上建立起来的曲线簇中的一种线型，方程中除频率 $p$ 和相应的变量 $x_p$ 外，还包含三个统计参数，即均值 $\overline{x}$、变差系数 $C_v$ 和偏差系数 $C_s$，其方程为

$$p = p(x \geqslant x_p) = F(x) = \int_{x\ p}^{\infty} \frac{\left(\frac{2}{C_v C_s \overline{x}}\right)^{\frac{4}{C_s^2}}}{\Gamma\left(\frac{4}{C_s^2}\right)} - \left(x - \overline{x} + \frac{2C_v}{C_s}\overline{x}\right)^{\frac{4}{C_s^2}-1} e^{\frac{2}{C_v C_s \overline{x}}\left(x-\overline{x}+\frac{2C_v}{C_s}\overline{x}\right)} \mathrm{d}x$$

$$= f(x_p; \overline{x}, C_v, C_s \qquad (9-10)$$

式中　$\Gamma\left(\frac{4}{C_s^2}\right)$——$\frac{4}{C_s^2}$的伽玛函数。

由于该方程比较复杂，直接求解是相当困难的。为了简化计算，经过数学推导，得出如下公式：

$$X_p = (1 + C_v \Phi_p)x = k_p x \qquad (9-11)$$

式中　$\Phi_p$——离均系数，与 p 和 $C_s$ 有关；

$k_p$——模比系数，与 $P$、$C_v$ 和 $C_s$ 有关。

现已把 $\Phi_p$(或 $k_p$)与影响参数之间的关系制成了专用的查算表，以方便使用。选用 $\Phi_P$ 和 $\Phi_P$ 的计算结果相同，只是虽然用 $K_p$ 计算 $x_p$ 的公式简单，但查算表中的 $C_s$ 是按 $C_v$ 的倍比给出的，有时会遇到 $C_v$ 和 $C_s/C_v$ 两个参数的内插，比较麻烦。这时用 $\Phi_p$ 计算就简单多了。

可见，正如方程式 $y=ax+b$ 中的参数 $a$ 和 $b$ 决定直线位置和坡度一样，$P$—Ⅲ型曲

线是由三个统计参数决定的。只要知道三个统计参数 $\bar{x}$、$C_v$、$C_s$，就可以通过查表得到不同频率 $P$ 相应的 $\Phi_p$(或 $k_p$)，然后由式(9—11)计算出相应的变量值 $x_p$，有了 $p$ 和相应的变量值 $x_p$，就可绘出一条 P—Ⅲ型理论频率曲线。

【例 9—3】 用表 12—9—3 年降水资料计算并绘制该站年降水量的 P—Ⅲ型理论频率曲线。

解：由表 12—9—3 年降水资料计算得到，$\bar{x}=584$mm，$C_v=0.19$，$C_s=0.35$。选取不同频率 $P$，由式(9—11)计算得出相应的年降水量 $x_p$，如频率 $P=1\%$时，由 $C_s=0.35$，查附表 1 得 $\Phi_p=2.58$，则 $x_{1\%}=(1+C_v\Phi_p)\bar{x}=(1+0.19\times2.58)\times584=870$(mm)。具体见表 12—9—4。依此绘制的 P—Ⅲ型理论频率曲线如图 12—9—4 所示。

**表 12—9—4　某站年降水且理论频率曲线计算表**

| $p$(%) | ① | 1 | 5 | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 95 | 99 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\Phi_p$ | ② | 2.58 | 1.74 | 1.31 | 0.82 | 0.48 | 0.20 | −0.00 | −0.31 | −0.56 | −0.85 | −1.24 | −1.54 | −2.07 |
| $x_p$(mm) | ③ | 870 | 777 | 729 | 675 | 637 | 606 | 577 | 550 | 522 | 490 | 446 | 413 | 354 |

频率曲线点绘在等分格的普通坐标纸上，两端陡峭，曲度较大，难以外延，而生产上又经常使用稀遇频率或大频率。如在图 12—9—4 上很难查到 0.1%对应的年降水量 $x_{0.1\%}$。为了克服外延的这种困难，通常把等分格的横坐标改为中间密两边疏的不均匀分格，表示累积频率，这种坐标纸叫频率格纸，如图 12—9—5 所示。在频率格纸上绘制频率曲线，两端的坡度比在普通坐标纸上绘制大大变缓，对曲线的外延是较为有利的。

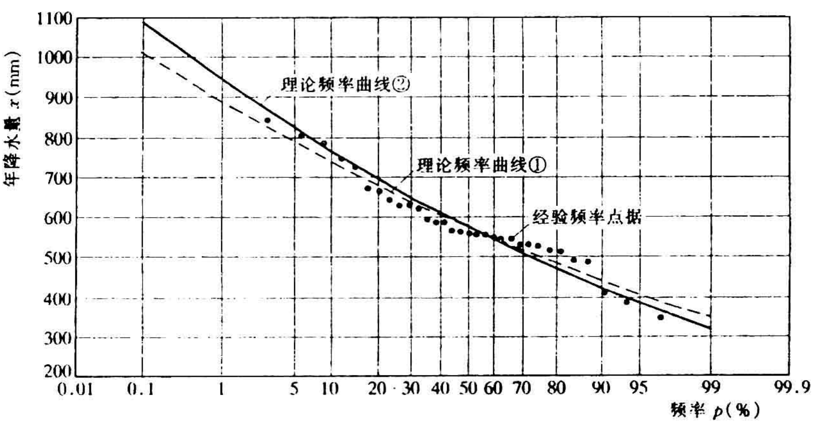


图 12—9—5　某站年降水量适线法频率计算图

# 第四节　适线法频率计算

## 一、适线法频率计算及其步骤

由上述可知，水文样本系列可以用作本的统计参数反映随机变量取值的数字特征，并可由统计参数惟一确定的P—Ⅲ型理论频率曲线表示随机变量取值的频率特征。但是理论和经验表明，由矩法公式(9—1)～式(9—4)计算样本的统计参数抽样误差较大，算得的统计参数及其相应的P—Ⅲ型理论频率曲线不能很好地反映总体的数字特征及总体的概率分布，生产上通常采用调整样本的统计参数及相应的P—Ⅲ型理论频率曲线来拟合样本的经验点据，将与经验点据配合最好的理论频率曲线近似地作为总体的概率分布，相应的统计参数作为总体数字特征的最佳估计。这种通过调整理论频率曲线逼近经验频率点据来确定统计参数的方法在水文上称为适点配线法，简称适线法。

适线法的计算步骤如下：

(1)点绘经验点据。将审查后的水文样本资料按由大到小排队，按式(9—9)计算各值的经验频率，然后将($p_i$,$x_i$)点绘在特制的频率格纸上。

(2)估算统计参数初值。根据样本资料系列，用矩法公式(9—1)和式(9—3)计算均值$\overline{x}$和变差系数$C_v$；作为适线的初值。至于偏差系数$C_s$，用矩法公式计算的抽样误差很大，一般不用公式9—11计算，而是根据经验选定$C_s$与$C_v$的倍比值($C_s/C_v$)。

(3)适线。即由统计参数初值$\overline{x}$,$C_v$,$C_s$，查附表1或附表2，按公式计算并绘制P—Ⅲ型理论频率曲线，判断该曲线与经验点配合情况，若配合良好，则表明该线就是所求频率曲线。若配合不好，调整统计参数，再次适线，直至曲线与经验点配合最佳为止。因为矩法计算的$\overline{x}$误差相对较小，主要是调整$C_u$和$C_s/C_v$。把这条配合最佳的频率曲线作为采用曲线，相应于该曲线的统计参数便可看成是水文总体统计参数的最佳估计值。

由此可见，适线法层次清楚，图像明显，方法灵活，容易操作，所以在水文计算中广泛采用。这一方法不仅是由样本的统计参数估计总体的数字特征的一种有效方法。同时也是由样本的频率分布估计总体的概率分布的基本途径。

在实际工作中，有两种适线方法：一种是目估经验适线，通过主观判断，决定适线的优劣，此法虽然简单灵活，但适线因人而异，任意性大；另一种是计算机编程适线，把理论频率曲线与经验点据的拟合误差作为适线准则，优选统计参数，求得最佳的拟合曲线，但是这种方法不便处理样本资料中的特大或特小水文数据，很难反映设计人员的经验。