实用水文統計法

金光炎編

水利电力出版社

实用水文統計法

金　光　炎編

*

1233S320

水利电力出版社出版（北京西郊科學路二里溝）

北京市書刊出版業營業許可證出字第105号

水利电力出版社印刷厂排印　　新华書店发行

*

850×1168 1/32开本＊3 1/4印張＊83千字

1958年12月北京第1版

1958年12月北京第1次印刷（0001—2,900册）

統一書号：15143·1018　　定价（第9类）0.43元

序　言

近来，数理統計法应用于水文計算中愈来愈多，例如频率与相关問題，儿乎在每个水文問題中都不能缺少。为了帮助初学者們了解数理統計方法的性質和計算步驟，故而編成了这本小册子。

書中，对于許多数理統計上的理論問題，一般只作了一些說明，而沒有論証它們。如果不是这样，那就会占去很多篇幅，而也会妨碍达到实用的目的。

至今，在水文統計分析中，还存在有許多爭論，書內沒有提及它們。因为，如果談論了这些問題，只会引起混乱，对初学者来說是沒有好处的。

本書采用了比較簡單而易于接受的方法，并举了許多例子来說明各种問題的意义与应用。

本書的对象是中等技术学校的学生和相当于这种程度的水利工作者。

限于編者的水平，一定会有許多缺点，希望得到同志們的批評和指正。

[redacted]日北京

14 目 录

第一章 水文統計的一般概念

1.水文現象的偶然性質

大家都知道，自然界中有这么一些現象：它們的到来是必然的、不可避免的。水文現象中很多事件都是这样，例如北京地区每年必定会下几場雨或飘几次雪；又如暴雨降落后，蒸发、滲漏及徑流等現象必然会随之而发生。

但是，北京地区每年究竟会下多少次雨；一場暴雨之后，在河川中將会出現多大的洪峯流量或者会产生多大的徑流。这些有关数量上出現多少的問題，我們很难預測。故在出現数量上，却不是必然的。

我們知道，降雨与徑流的形成，幷不是很單純的，在它們的出現过程中往往牽連着許多复杂的要素（如很多气象上和地理上的因子），甚至复杂到使我們无从考查其自始至終的变化过程。因此我們說这种錯綜复杂的結果，即在数量上的出現情況是偶然的。

这是水文現象的偶然性質，但絕不能与它本身出現的必然性相混淆。

2.水文計算中应用数理統計法的必要性

經驗告訴我們，这种偶然現象也有它一定的規律性。比喻說，經过多次观察后，就能发現河川中的历年水位：特别大的和特别小的是不大容易碰到，而一般的情况却常发生。这样我們就可根据以往的事实来对未来的情势作出某些估計。数理統計法和其有关的机率理論就是研究这种偶然規律性的数学工具。

我們知道，水利工程的設計中需要各种大小的水文数据。但目前我們还不能对未来的降雨或徑流作出長期定量的預报，而現在的水利建設又不允許等到將来預报科学极度发展后再来进行，

因此适当地运用数理統計法来估算今后可能发生的水文情况，也可以滿足工程設計的要求。

由此可見，水文計算中数理統計法的应用是必要的。但必須知道，光靠数理統計法是不可能解决复杂的水文問題的，只有与現象的物理成因分析和綜合各个自然地理特征的方法密切結合起来，才能最有效果。因为数理統計法毕竟是一种数学工具，幷不能說明或表达現象的本質，也仅在正确应用时才是强有力的助手。假如不問靑紅皂白，于实际工作中盲目地搬用数理統計法，就会犯严重的錯誤甚至歪曲事实，这一点必須引起我們特别的注意。

3.統計分析时应該注意的几項問題

应用数理統計法来分析水文現象其唯一的依据是过去的实測資料，因此在工作时，应密切注意資料的考查工作。現分别說明于后：

資料的一致性 統計法要求所有的資料是屬于同一类型幷且是在同一条件下产生的，不能选取性質不同的資料。例如：雨汛洪水和融雪洪水，由于它們的成因与类型均不相同，因此不能把它們統計在一起。又如水庫的兴建、河道的疏浚以及各种水利工程的修筑，对于工程完成前后的水文資料也是不同性質的，必須經过适当处理后才能应用。由于目前羣众性的水利运动普遍开展，对資料一致性的分析是更屬重要的。

資料的代表性 統計分析的目的是要把过去資料的变化規律推論于今后的情况，因此对資料代表性的审查必須相当細致。現在我国的水文紀录一般仅有十多年至数十年，而最長的也不超过一百年，有的地方甚至只有几年的資料。这些短期的記載，很可能都是丰水年或枯水年，因此根据它們所推得的結果，就会偏大或偏小。一般要有20至30年的資料才具有一定的可靠性。

在資料不足的情况下，应尽可能地利用鄰近地区且条件一致的長期資料来延長，或利用历史水文資料等来弥补。

資料的可靠性与独立性 資料可靠程度的檢查也是相当重要

的。例如可把前后联系起来或把有关系的現象放在一起来看，檢查它是否有矛盾或写錯伪造之处。

統計分析中还要求資料是相互独立的，因此不能把互相有关的資料統計在一起。如日流量的前后几天間都是由同一場暴雨所形成的，彼此并不独立，故我們不能用連續的日流量来进行統計計算。

第二章 机 率

1.机率的意义

我們用例子来說明机率的意义，如：一枚硬币有正反兩面，投擲一次的結果，或者是出現正面或者是出現反面，但可以断定出現正面或出現反面的机会（可能程度）是相等的。由于一枚硬币投擲一次的全部可能出現情况有兩种，而实际出現只能是这兩种情况中的一种，因此我們說出現正面或出現反面的机会都是 $\frac{1}{2}$。这个比值是代表其中一种情况（事件）出現的可能性，也就是該事件的发生率，一般称之为机率。

机率有兩种：一为事先机率，一为經驗机率。如果，某一类事件的出現与不出現的种种可能方法都非常清楚地知道，象上述擲錢問題那样，則叫做**事先机率**。假使对某一类事件，我們不能預知其出現与不出現的一切可能情况，要估計它的机率，就只能通过多次的試驗来求得，則称之为**經驗机率**。

經驗机率也叫做頻率。水文現象中如某一大小的降雨量或洪峯流量的发生率，其事先机率无法知道，这样我們仅能借助于已有的实測資料应用数理統計算法来估得它們的頻率。由机率理論中的大数定律知：当試驗次数（或实測資料）很多时，頻率就非常稳定，甚至于很接近于一个常数。意思是說当試驗次数愈多，其頻率也愈眞确。因此，水文統計分析就要求有足够的实測資料。

2.簡單的机率計算

簡單的机率計算方法可用下列公式来表示:

$$P\{A\}=\frac{m}{n}. \qquad (2\text{-}1)$$

式中 $P\{A\}$ 为事件 A 出現(或不出現)的机率,n 为所有可能的結果数,m 是出現有利(或不利)情況的可能結果数。

例 1 盒中有 4 个紅球和 6 个白球,問任意摸出一个紅球的机率若何?

解 因为全部可能結果数 $n=10$,而有利場合 $m=4$,故出一个紅球的机率为$\frac{4}{10}$或$\frac{2}{5}$。

例 2 在 52 張扑克牌中,任意抽取一張,問得 K 的机会多少?

解 由于 $n=52$ 及K共有四張,即 $m=4$,故抽得 K 的机率为$\frac{4}{52}$或$\frac{1}{13}$。

在式 (2-1) 中,如果使 $m=n$,即有利場合的可能結果数就是試驗中的全部可能情况,則

$$P\{A\}=\frac{m}{n}=1.$$

机率等于 1 的意义是表示不論在那一次試驗中我們欲求的事件总是出現的,因此我們称这种情况为**必然事件**。反之,当 $m=0$,亦即全系不利場合,則此时

$$P\{A\}=0 .$$

机率得零表示在每次試驗中均不会出現欲求的事件,故称为**不可能事件**。因此偶然事件的机率均介于 0 与 1 之間。这些情况就是机率的特性。

上述的計算方法,只能应用于一些很簡單的問題中,至于稍呈复杂的試驗,尙需下述的几个机率定理来解决。

3.机率相加定理

如果某种試驗可有n种可能不同的結果,其中有利于 A_1 事件的有 m_1 种,有利于 A_2 事件的有 m_2 种,……,有利于 A_K事件

的有 m_K 种，而这些事件 A_1、A_2、……、A_K 在一次試驗中仅能出現其中的一个（以后我們称这种情况为互斥事件），則有利于 A_1 或 A_2……或 A_K 情形的数目共有 $m_1+m_2+\cdots\cdots+m_K$ 种，故这些事件中任何一个出現的机率 $P\{A_1$ 或 A_2……或 $A_K\}$ 为

$$\frac{m_1+m_2+\cdots\cdots+m_K}{n}=\frac{m_1}{n}+\frac{m_2}{n}+\cdots\cdots+\frac{m_K}{n}.$$

由于 $\frac{m_1}{n}$ 为事件 A_1 的出現机率，或写作 $P\{A_1\}$，同样 $\frac{m_2}{n}=P\{A_2\}$，……，$\frac{m_K}{n}=P\{A_K\}$，故上式又可为

$$P\{A_1\text{ 或 }A_2\cdots\cdots\text{或 }A_3\}=P\{A_1\}+P\{A_2\}+\cdots\cdots+P\{A_K\}. \tag{2-2}$$

由此可以得出下列的机率相加定理。

定理　若干互斥事件中任何一些事件出現的机率为这些事件出現机率的总和。

例 1　盒中有紅球 5 枚、白球 10枚及藍球15 枚，問任意摸取一球得紅球或白球的机率为多少？

解　盒中共有 $n=30$ 枚球，取出一枚紅球的机率为 $\frac{5}{30}$，取一枚白球的机率为 $\frac{10}{30}$，根据机率加法定理得

$$P\{\text{紅球或白球}\}=\frac{5}{30}+\frac{10}{30}=\frac{1}{2}.$$

例 2　在上例中，問任取 2 球，得 2 白球或 2 藍球的机率为若干？

解　在30个球中任取 2 枚共有 C_2^{30}❶ 种可能情况，10 枚白球中任取 2 枚共有 C_2^{10} 种，15枚藍球中任取 2 枚共有 C_2^{15} 种，则得

$$P\{2\text{ 白球或 }2\text{ 藍球}\}=\frac{C_2^{10}+C_2^{15}}{C_2^{30}}=\frac{45+105}{435}=\frac{10}{29}.$$

例 3　投擲兩粒骰子，問擲一次而得 9 点的机率若何？

---

❶ 这是組合計算的符号。在 n 个元素中选取 r 个不計次序的組合的公式为

$$C_r^n=\frac{n!}{r!(n-r)!}.$$

惊嘆号讀作阶乘，卽連乘积的意思，如 $30!=30\times29\times28\cdots\cdots3\times2\times1$，又 $0!=1$。

解 一粒骰子共有6种可能結果，而兩粒骰子共投則有$6\times6=36$种可能結果。擲一次出現9点可有四种形式，即3与6、4与5、5与4、6与3，这四者中每种出現的机率均为$\frac{1}{36}$，故其机率总和为$\frac{1}{9}$。

4. 机率相乘定理

多种事件中，若任何一个事件的出現并不影响其他事件出現者称为独立事件。

如果甲种試驗共有n_1种結果，其中有利于A事件者有m_1种。另外又有一种与甲完全无关（即相互独立）的乙种試驗，它共有n_2种結果，其中有利于B事件的有m_2种。若甲乙共同試驗一次，则总的結果共有n_1n_2种，共同出現A和B的結果共有m_1m_2种，因此出現A也同时出現B的机率$P\{AB\}$为

$$\frac{m_1m_2}{n_1n_2}=\frac{m_1}{n_1}\times\frac{m_2}{n_2}.$$

显然$\frac{m_1}{n_2}=P\{A\}$及$\frac{m_2}{n_2}=P\{B\}$故

$$P\{AB\}=P\{A\}P\{B\}. \tag{2-3}$$

由此可导出下列的机率相乘定理。

定理 若干独立事件共同出現的机率等于这些事件出現机率的乘积。

例1 一粒骰子連續投擲三次，問均得幺点的机率若何？

解 骰子在每次投擲中得幺点的机率均为$\frac{1}{6}$，按机率相乘定理，連擲三次都得幺点的机率为$\left(\frac{1}{6}\right)^3=\frac{1}{216}$。

例2 某测站的水位，每年出現高于2公尺的机率为$\frac{1}{2}$和出現高于3公尺的机率为$\frac{1}{5}$，問在連續兩年中，头一年出現高于2公尺和第二年出現高于3公尺的机率是多少？

答 其机率为$\frac{1}{2}\times\frac{1}{5}=\frac{1}{10}$。

例3 盒中有紅球6枚和白球4枚，問一次取出5枚中有2枚紅球及3枚白球的机率若干？

解 盒中共有10枚球，取5枚共有C_5^{10}种取法，由6枚紅球中取出2枚只有C_2^6种，在4枚白球中取出3枚白球只有C_3^4种，故得2紅3白的机率为

$$\frac{C_2^6C_3^4}{C_5^{10}}=\frac{15\times4}{252}=\frac{5}{21}.$$

5. 二項机率定理

把独立事件进行多次重复試驗，每次試驗中只有两种可能結果：发生与不发生。現令发生的机率为p，不发生的机率为q，显然$p+q=1$。

現在要問在进行n次重复試驗后，恰好有m次发生的机率是多少？

在n次試驗中，有m次发生則其余的$n-m$次一定是不发生，因此在全部試驗中，有m个p及$n-m$个q。根据机率相乘定理得到它們共同出現的机率为p^mq^{n-m}。但在n次試驗过程中，p和q出現的次序排列不只有一种，也就是說，它們的出現次序可以是p、p、q、……，也可以是p、q、p……，等等，故共可有$\frac{n!}{m!(n-m)!}$或C_m^n种組合。由此可得到n次試驗中发生m次的机率为

$$\left.\begin{aligned}B(m)&=C_m^np^mq^{n-m}\\ \text{或}\quad B(m)&=C_m^np^m(1-p)^{n-m}\end{aligned}\right\}\qquad(2\text{-}4)$$

由于上式为二項式$(p+q)^n$展开式中的第$m+1$項，故又称为二項机率公式，其所引出的定理叫做二項机率定理。

如果要問在n次重复試驗后，至少有m次发生的机率（至少有m次发生的意思是：发生m次或$m+1$次……或n次都可以）。根据机率相加定理得欲求的机率P为

$$\begin{aligned}P&=C_m^np^mq^{n-m}+C_{m+1}^np^{m+1}q^{n-m-1}+\cdots\cdots+C_{n-1}^np^{n-1}q+p^n\\&=\sum_{r=m}^{n}C_r^np^rq^{n-r}.\qquad(2\text{-}5)\end{aligned}$$

当m較小时用上式来計算很麻煩，故可应用二項展开式的对称性原理簡化之，即

$$P=1-\sum_{r=0}^{m-1}C_r^n p^r q^{n-r}. \qquad (2-6)$$

例1 擲硬币10次，問出現6次正面的机率是多少？

解 在此$n=10$，$m=6$，$p=q=\frac{1}{2}$，故得欲求的机率为

$$B(6)=C_6^{10}\left(\frac{1}{2}\right)^6\left(\frac{1}{4}\right)^4=0.205.$$

例2 一粒骰子投擲五次，問至少有兩次为幺点的机率为若干？

解 至少有兩次为幺点的意思是擲得幺点的次数可以为兩次或三次或四次或五次，現分別計算每种情況的机率，且知$n=5$，$p=\frac{1}{6}$，$q=\frac{5}{6}$，故有

$$B(2)=C_2^5\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^3=\frac{1250}{7776};$$

$$B(3)=C_3^5\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^2=\frac{250}{7776};$$

$$B(4)=C_4^5\left(\frac{1}{6}\right)^4\left(\frac{5}{6}\right)=\frac{25}{7776};$$

$$B(5)=C_5^5\left(\frac{1}{6}\right)^5=\frac{1}{7776}.$$

因而至少有兩次为幺点的机率，可按式(2-5)来計算，即

$$P=B(2)+B(3)+B(4)+B(5)=\frac{1}{7776}(1250+250+25+1)=\frac{1526}{7776}=0.196.$$

这个問題同样可用式(2-6)来解，計算的步驟比上述要簡單，此时我們只要算$B(0)$及$B(1)$兩項就行了。由于

$$B(0)=\left(\frac{5}{6}\right)^5=\frac{3125}{7776};$$

$$B(1)=5\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^4=\frac{3125}{7776}.$$

故得同一的結果：

$$P=1-B(0)-B(1)=1-\frac{6250}{7776}=\frac{1526}{7776}=0.196.$$

例3 某河段，每年的最大洪峰流量超过Q_0的机率为10%，問在最近

十年中均不超过或仅超过一次的机率各为多少？

解 在此$n=10$，$p=0.1$，$q=0.9$，故均不超过的机率为

$$B(0)=(0.9)^{10}=34.7\%,$$

仅超过一次的机率为

$$B(1)=C_1^{10}(0.1)(0.9)^9=38.7\%。$$

在水利工程的规划中，类似例3的問題很多。这个例子告訴我們：如果用Q_0去設計該河段的河槽断面以排泄內澇，那么在未来十年中該河段附近均不受內澇的机率为34.7%，而要受內澇灾害（即至少有一次內澇）的机率为$1-34.7\%=65.3\%$，也就是說受灾的可能性比不受灾的可能性几乎大一倍。假如我們允許它在十年中內澇一次，且这一次的內澇已列入我們的水利計划，亦即这次澇灾虽使农业歉收，但我們已有足够的准备来弥补灾害所形成的損失，那么能实現預定計划的机率为$B(0)+B(1)=34.7\%+38.7\%=73.4\%$，而不能实現預定計划的机率为26.6%，也就是說能实現計划的把握比較大，因为它比不能实現計划的可能性几乎大兩倍。

第三章 分布系列

1.随机变数

在机率試驗中，有种种可能的結果，如果每种結果都給以一个实数值，也就是用一个具体的数目去代表某一种結果，則这个实数值就称为随机变数。

随机的意义是不帶有主观成分且系偶然的，因此随机变数是具有偶然性的变数，也可称作偶然量。例如，骰子共有六面，每面都有点数，在任何一次不帶有主观成分的試驗中，出現那一面都是偶然的，因此我們說一粒骰子有六个随机变数。又如洪峯流量，它的数值Q是試驗的結果，也是一个变数且随着偶然的情况而变动，所以說Q是一个随机变数。

随机变数可以根据所研究的任务和范圍的不同来进行組合。

如一粒骰子虽仅六个随机变数，但若要兩粒骰子共投时，此时的随机变数则成了11个，即2点、3点、……、12点；也就是說把兩粒骰子中的点数，进行了所有可能的組合(和的組合)。兩条或数条河流会合后的流量也是如此。

随机变数分为兩类：不連續型随机变数与連續型随机变数。

若相鄰兩随机变数間,并不存在有中間值,这种随机变数就称为**不連續型随机变数**。例如一粒骰子擲一次，不是出1点就是出2点或是出其他的点数，在1点与2点之間不能再有任何可能的数值。这种情况在水文資料中很少見到。

如果相鄰兩随机变数間，可以分为許多中間值，这些中間值之間的差Δx可以小到无穷小，甚至可用微分符号dx来表示，这种随机变数就称为**連續型随机变数**。水文資料中的数据大都如此。

一串随机变数所組成的單位称作**系列**。系列的范圍可以是有限的，也可以是无限的。水文統計分析中所用到的大都是有限系列。

2. 总体与样本

很多的水文現象是无限系列。例如某地的年降水量，其系列的全部情况应該是自古迄今再展延至未来极其長远岁月中的所有年降水量。这种包括整个情况的系列，我們称它为**总体**。但是，目前我們所掌握的实测資料仅仅是总体的一个很小部分，也就是說只是一个有限系列，故現在已有的水文系列就叫做这个总体的一个**样本**。

严格点說，总体是客观存在的，且是在同等基础上結合所有个别單位的集团，也就是指有一个共同联系而結合起来的許多观測單位的羣。样本則是总体中由若干單位組成的小羣。因此总体可以被分割成許多样本。

样本是可以在总体中任意抽取的，这种情况就叫做**抽样**，而这种样本有时也称作**随机样本**。

如上所述，水文現象的总体大都是无限的，因此要确切研究

总体的各种特征或某些規律性相当困难。但在水利規划中很需要知道总体的情况，比喻說为了防止洪水泛濫成灾，就必須知道今后发生各种大小洪水的可能性，这样我們就能用有效的措施（如筑水庫、修堤防等）来控制与約束洪水，把它的破坏能力减低到最小限度。

幸而在机率論和数理統計学中給出了許多由样本来估計总体的定理与方法，只要在样本系列足够大时，就能比較精确地推算出总体的近似情况，以来滿足实际分析的需要。然而总体与样本的概念絕不能因此而混淆。总体虽然能通过样本来認識，但总不能把样本認作是总体。

3.机率分布

如果把实測資料經过适当的整理后，就能发現它們具有一些特性。現以某站的年降水量为例，它共有62年实測資料，若使年降水量的数值每隔 200 公厘划分为一組，則各組的出現次数可統計如表3-1中的第二列，同时可按此算出相应的頻率、累积(出現)次数及累积頻率的值。然后我們再把年雨量与相应的出現次数或

表3-1　　某站年雨量分組統計表

組　段 (公厘)	出現次数 (年)	頻　率 (%)	累积次数 (年)	累积頻率 (%)
501～700	1	1.6	62	100.0
701～900	2	3.2	61	98.4
901～1100	15	24.2	59	95.2
1101～1300	18	29.1	44	71.0
1301～1500	13	21.0	26	41.9
1501～1700	7	11.3	13	20.9
1701～1900	3	4.8	6	9.6
1901～2100	2	3.2	3	4.8
2101～2300	1	1.6	1	1.6
总　計	62	100.0		

注：累积次数和累积頻率采用自大而小累加以与水文分析中的概念相符。

频率繪成图 3-1，把年雨量与相应的累积次数或累积频率繪于图 3-2。

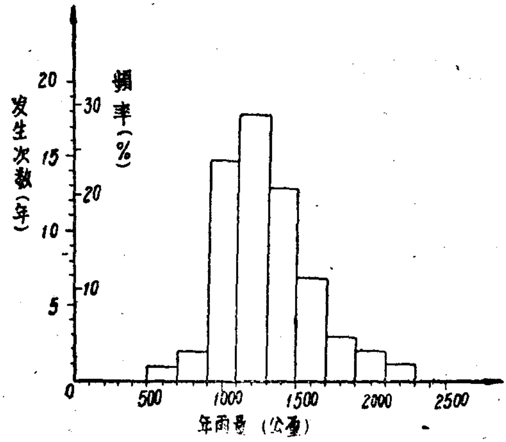

图 3-1　某站年雨量与发生次数、频率的关系图

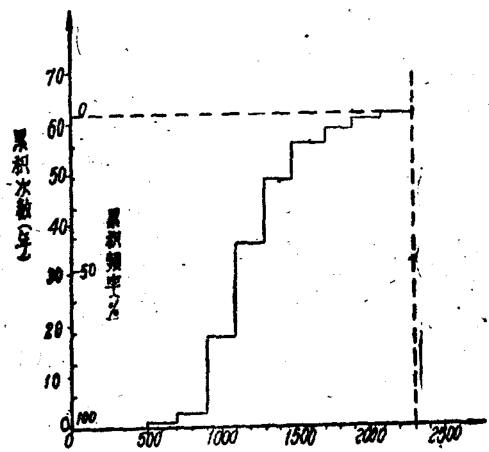

图 3-2　某站年雨量与累积次数，累积频率的关系图

經驗証明，很多的統計資料的出現次数分布形狀都与图 3-1 类同，即中間高而兩側漸低，离中間最远的地方出現次数最少。同时可以在图 3-2 中看出，累积频率的中間部分变化較緩，而