水利水电工程试验数据计算分析方法

张龙 郭林涛 洪镝 编著

中国水利水电出版社
知识产权出版社

水利水电工程试验数据计算分析方法

●张龙　郭林涛　洪镝　编著

中国水利水电出版社
知识产权出版社

内容提要

本书主要针对水利水电工程质量控制和检测工作的需要，介绍了水利水电工程中常用混凝土试验、原材料试验、土工试验数据的计算方法。书中还根据试验数值处理需要，介绍了部分 Excel 函数的应用方法和技巧，供读者在实际工作中参考。本书内容深入浅出、紧密贴合工程实际，实用性强。

本书可供水利水电工程现场试验人员、施工人员和监理人员以及高等院校相关专业学生学习参考。

责任编辑： 阳　淼　张宝林　E-mail：yangsanshui@vip.sina.com；z-baolin@263.net
文字编辑： 彭天赦

图书在版编目（CIP）数据

水利水电工程试验数据计算分析方法 / 张龙，郭林涛，洪镝编著. -- 北京 : 中国水利水电出版社 : 知识产权出版社，2009.9
ISBN 978-7-5084-6840-2

Ⅰ. ①水… Ⅱ. ①张… ②郭… ③洪… Ⅲ. ①水利工程一实验数据一计算方法②水利工程一实验数据一分析方法 Ⅳ. ①TV-33

中国版本图书馆CIP数据核字(2009)第171981号

水利水电工程试验数据计算分析方法
张龙　郭林涛　洪镝　编著
中国水利水电出版社 / 知识产权出版社　出版发行（北京市海淀区玉渊潭南路1号D座；电话：010-68367658 / 北京市海淀区马甸南村1号；电话：010-82005070）
北京科水图书销售中心零售（电话：010-88383994、63202643）
全国各地新华书店和相关出版物销售网点经售
中国水利水电出版社微机排版中心排版
北京瑞斯通印务发展有限公司印刷
184mm×260mm　16 开本　16 印张　379 千字
2009 年 9 月第 1 版　2009 年 9 月第 1 次印刷
印数：0001—4000 册
定价：**31.00** 元

前　　言

水利水电工程直接关系到国计民生，而水利水电工程质量控制和检测则是水利水电工程建设中极为重要的一项工作，其试验方法是否科学、试验结果是否准确关系到整个工程的质量，如何做好质量控制和检测工作，提高水利水电工程现场试验检测人员的技能水平，是一直以来工程建设管理人员工作的重点。本书作者长期从事水利水电工程施工现场试验和质量控制管理工作，在实践过程中总结了大量的现场试验计算方法和分析方法，编写本书，旨在为从事水利水电工程试验人员提供交流学习的平台，提高现场试验人员的试验水平和计算分析能力，使试验人员能够更好地掌握试验技能，适应水利水电工程建设快速高效发展的需要。

本书依据行业或国家有关标准、规程、规范和设计文件及有关水利水电工程中混凝土、原材料、土工的试验要求，归纳了水利水电工程试验中常用的试验方法和计算方法。本书分为5章，其中，第1章介绍数据计算方法，包括数据的简单分析，随机变量的基本概念、数学特征、基本定理，行列式和矩阵，回归分析，非线性回归，多元回归，方差分析，正交设计，插值，以及数据处理；第2章介绍部分混凝土试验数据计算方法；第3章介绍部分原材料试验数据计算方法；第4章介绍部分土工试验数据计算方法；第5章介绍混凝土质量统计分析方法。第2、3、4章主要通过现场试验的例证，使试验人员掌握不同的计算方法，得到准确的试验结果。第5章主要以混凝土抗压强度试验结果为例，通过大量的试验数据进行统计分析以及利用直方图、管理图、正态分布图对试验结果进行统计分析，并对记录数据中的混凝土拌和物的温度、坍落度、含气量三个常用指标进行统计分析。

为了计算方便，本书试验数据选用 Microsoft Excel 函数进行计算。Microsoft Excel 是一款电子表格软件，也是一种数据库软件，它能够完成许多复杂的数据运算，进行数据统计，并具有制作图表的功能。其操作原理简单，操作方法易学，为大家所熟悉。用 Excel 函数处理一些实际问题的计算方法很多，本书根据水利水电工程试验数据处理方法，介绍了部分 Excel 函数的应用、程序编写的方法和作图的方法，可供读者参考。

本书由西北农林科技大学水利与建筑工程学院（水利部西北水利科学研究所）的张龙、郭林涛同志和黄河上游水电开发有限责任公司建设工程分公

司的洪镝同志共同编写。本书可供水利水电工程现场试验人员、施工人员和监理人员以及高等院校相关专业学生学习参考。

由于作者水平所限，书中疏漏和不足之处在所难免，敬请各位同行批评指正。

作　者

2009年6月

目　录

第1章　数 据 计 算 方 法

1.1　数据的简单分析

1.1.1　算术平均数

算术平均数简称平均数或均值。它是反映总体中某一项特征集中位置的统计特征值，也就是将所测得的质量特征性能数据的代数和除以数据的总和（样本总数），即

$$\overline{x}=\frac{1}{n}(x_1+x_2+\cdots+x_n)=\frac{1}{n}\sum_{i=1}^{n}x_i \tag{1.1.1}$$

式中　$\overline{x}$——算术平均数；

x_i——数据中的第 i 个数；

n——数据的总数。

1.1.2　加权平均数

若给定一组数据 x_1、x_2、…、x_n 和满足条件 $\sum\limits_{i=1}^{n}f_i=1$ 的一组正数 f_1、f_2、…、f_n，则

$$\sum_{i=1}^{n}f_ix_i=f_1x_1+f_2x_2+\cdots+f_nx_n \tag{1.1.2}$$

就称为 x_1、x_2、…、x_n 的加权平均数，f_1、f_2、…、f_n 称为 x_1、x_2、…、x_n 相应的权（$i=1$，2，…，n）。很明显，当权 $f_i=f_1=f_2=\cdots=f_n=1/n$ 时，加权平均数就是算术平均数。“权”就是衡量数据 x_i 在平均时的重视程度。可用下式表示加权平均数：

$$\overline{x}=\frac{f_1x_1+f_2x_2+\cdots+f_nx_n}{f_1+f_2+\cdots+f_n}=\frac{\sum\limits_{i=1}^{n}f_ix_i}{\sum\limits_{i=1}^{n}f_i} \tag{1.1.3}$$

1.1.3　方差、标准差

对给定的一组数据 x_1、x_2、…、x_n 来说，$\overline{x}$ 是与这一组 n 个数最接近的数，即 $\sum\limits_{i=1}^{n}(x_i-\overline{x})$ 在 $\overline{x}$ 变化时达到最小值。$(x_i-\overline{x})^2$ 是与偏差的平方，这 n 个数与 $\overline{x}$ 的平方的偏差的平方就是

$$\frac{1}{n}[(x_1-\bar{x})^2+(x_2-\bar{x})^2+\cdots+(x_n-\bar{x})^2]=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2 \tag{1.1.4}$$

它就为数据 x_1、x_2、…、x_n 的方差（或均方差），它的算术平方根称为标准差，记作 σ，即

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

$$\sigma=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2} \tag{1.1.5}$$

可以看出，σ 越大，这组数据就越分散，也就是说这组数据的变异性就越大；σ 越小，这组变异性就越小，数据就越相对集中，当 $\sigma=0$，$x_1=x_2=\cdots=x_n=\bar{x}$。由于 σ 的大小代表数据之间的差异大小，也就是表示工程质量的波动程度，所以 σ 是衡量质量好坏的一个重要指标。

对一组数据 x_1、x_2、…、x_n 来分析，$\bar{x}$ 与 σ（或 σ^2）是两个常用的量，前者是反映总体数据中统计的特征值，后者是描述数据的变异性的值。

1.1.4 极值、极差、变异系数

当描述数据之间的差异程度（变异性），也可以用其他的指标。

1. 极值

一组数据 x_1、x_2、…、x_n 按大小顺序排列后，处于首位和末位的最大和最小的两个数 x_{max}、x_{min} 就是极值。

2. 极差

一组数据 x_1、x_2、…、x_n 中最大值减去最小值，即 $x_{max}-x_{min}$ 称为 x_1、x_2、…、x_n 的极差，它反映了一组数据之间最大的差距。

3. 变异系数

x_1、x_2、…、x_n 的标准差 σ 都是有单位的量，单位不同时不易比较。为了消除单位的影响，考虑相对的变异性，采用变异系数 C_v：

$$C_v=\frac{\sigma}{\bar{x}} \tag{1.1.6}$$

因此，标准差 σ 是表示数据的绝对波动大小，变异系数 C_v 是表示数据相对波动的大小。

1.1.5 中位数、众数

1. 中位数

从 $\bar{x}$ 的定义可以看到，x_1、x_2、…、x_n 中个别数字很大或者个别数字很小，都会对 $\bar{x}$ 产生影响，有时这些数据并不反映真实情况，但是不能随便剔除它们，这时往往采用中位数。将数据 x_1、x_2、…、x_n 按大小排列，当 n 是奇数时，居中间的一个数就是中位数，当 n 为偶数时，居中间的两个数取平均值就是中位数。

2. 众数

众数是一组数据 x_1、x_2、…、x_n 最集中的那个值附近的一个数，从直方图中很容易反映众数。

1.2 随机变量的基本概念

1.2.1 事件和随机事件

(1) 必然事件。在一定条件下必然发生的事件称为必然事件。例如混凝土抗压强度试验结果为正值，是必然事件。

(2) 不可能事件。在一定条件下不可能发生的事件成为不可能事件。例如混凝土抗压强度为负值即为不可能事件。

(3) 随机事件。在一定条件下可能发生或者不发生的事件。例如混凝土抗压强度试验结果出现在 21.6～28.1MPa 之间，是一个随机事件。随机事件是随机现象的某种结果。

1.2.2 随机变量及其分布

1. 随机变量及其分布函数的概念

如果对于任一实数 x，变量 ξ 在区间 $(-\infty, x)$ 内取值的概率是 x 的给定函数 $f(x)$，则称 ξ 是给定分布函数 $f(x)$ 取值的随机变量，并称 $f(x)$ 是随机变量 ξ 的分布函数。

例如，混凝土抗压强度试验结果在 21.6～28.1MPa 区间取值的概率为随机事件，这样的量称为随机变量。

2. 随机变量的分类

随机变量根据其取值的特征可以分为两类：

(1) 连续型随机变量。如果随机变量 ξ 的分布函数 $f(x)$ 可表示为：$f(x)=\int_{-\infty}^{x} f(t)\mathrm{d}t$，则称 ξ 是连续随机变量，并称 $f(x)=\mathrm{d}f/\mathrm{d}x$ 为 ξ 的分布函数。也就是说 x 的取值可以在坐标轴上某一区间取任一数值，若取值布满整个区间或整个实数轴，则称 x 为连续型随机变量。

(2) 离散型随机变量。如果随机变量 ξ 只取有限个值 x_1、x_2、…、x_k、…，并且取各个值的概率 P_1、P_2、…、P_k、…，则称 ξ 是离散型随机变量，这时分布函数为 $f(x)=\sum_{x} P(\xi=x_i)$。也就是说随机变量 x 的取值可以离散地排列为 x_1、x_2、…、x_k、…，而且 x 以各种确定的概率取这些不同的值，即只取有限个实数值，则称 x 为离散型随机变量。

例如，在取有效数字的位数时，数字的舍入误差属于离散型随机变量。

1.2.3 事件与概率

1. 概率的概念

随机事件的特点是在一定条件下可能发生也可能不发生，但是在大量的试验中呈现统

计规律，用大写字母 A、B、C 表示。

概率随机事件的概率 $P(A)$ 是用于度量随机事件 A 发生的可能性大小的数值。必然事件的概率为1，不可能事件的概率为0，随机事件的概率 $P(A)$ 满足 $0 \leqslant P(A) \leqslant 1$。所以必然事件和不可能事件是随机事件的特例。

例如，混凝土抗压强度 n 次试验中，混凝土合格事件发生 m 次，m 成为事件的频数，m/n 则成为事件的相对频数或频率。当 n 极大时，频率 m/n 稳定地趋于某一个常数 P，此常数 P 称为事件 A 的概率。

2. 概率的几个基本性质

(1) 包含：$A \subset B$，表示 A 发生时 B 一定发生，所以有

$$A \subset B \Rightarrow P(A) \leqslant P(B)$$

(2) 相等：$A \subset B$ 且 $B \subset A$，则称 A、B 相等，记为 $A=B$，所以有

$$A=B \Rightarrow P(A)=P(B)$$

(3) 和：$A+B$（或 $A \cup B$），表示 A、B 中至少有一个发生。但是一般有

$$P(A+B) \neq P(A)+P(B)$$

(4) 积：AB（或 $A \cap B$），表示 A、B 均发生。但是一般有

$$P(AB) \neq P(A)P(B)$$

(5) 必然事件与不可能事件：在一定条件下发生的事件称为必然事件，用 $\overline{U}$ 表示，显然 $P(\overline{U})=1$。在一定条件下，不发生的事件称为不可能事件，用 ϕ 表示，显然 $P(\phi)=0$。

(6) 不相容（互斥）事件：若 $AB=\phi$，则称 A 与 B 互斥。互斥事件成立的加法公式为

$$A \cap B=\phi \Rightarrow P(A+B)=P(A)+P(B)$$

(7) 对立事件（逆事件）：A 的对立（或逆）事件用 $\overline{A}$ 表示，A 是 $\overline{A}$ 的逆，$\overline{A}$ 是 A 的逆，而且 A 与 $\overline{A}$ 不能一起发生。即有 $A \cap \overline{A}=\phi$，又有 $A \cup \overline{A}=\overline{U}$，故得

$$P(A)+P(\overline{A})=P(\overline{U})=1 \text{ 或 } P(\overline{A})=1-P(A)$$

(8) 减：事件 $(A-B)$ 表示 A 与 $\overline{B}$ 的积，即 A 发生而 B 不发生。注意的是当 A 与 B 相减时，并不要求 $B \subset A$ 或 $A \supset B$，并且一般有

$$P(A-B) \neq P(A)-P(B)$$

1.2.4 几种常用分布

1. 二点分布

随机变量 ξ 只取0、1两个值，即 $P(\xi=0)=P$，$P(\xi=1)=1-P$，$(0 \leqslant P \leqslant 1)$。二点分布描述某一特定事件是否发生。二点分布属于离散型随机变量。

2. 二项分布

若随机变量 ξ 的取值范围 $k=0$、1、2、…、n，其分布为

$$P(\xi=k)=\binom{n}{k}P^k(1-P)^{n-k} \qquad (0<P<1, q=1-P)$$

记作 $\xi \sim B(n, P)$，称 ξ 服从于 n、P 的二项分布。

二项分布是 n 重贝努利试验中试件 A 恰好发生 k 次的频率的分布。

3. 正态分布

随机变量 ξ 的密度函数为

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

通常用 $N(\mu,\ \sigma^2)$ 表示正态分布，当 $\mu=0$，$\sigma^2=1$ 时，相应的 $N(0,\ 1)$ 称为标准正态分布。正态分布是水利水电工程试验中常用的统计分析方法之一。

4. 均匀分布

随机变量 ξ 的密度函数为

$$f(x)=\begin{cases}\dfrac{1}{b-a}\\ 0\end{cases}$$

通常用 $U(a,\ b)$ 表示，称为在 $(a,\ b)$ 区间上的均匀分布，一般数据的舍入误差遵从均匀分布 $U(0,\ 1)$。

5. 指数分布

随机变量 ξ 的密度函数为

$$F(t)=1-e^{-t}(t>0)$$

通常称为负指数分布，或称为指数分布。指数分布有很多重要特征，它在寿命试验的数据分析中有特殊位置。

1.2.5　样品、样本和统计量

(1) 总体：统计要研究的问题，是一类事物的某个量的指标分布，如混凝土抗压强度分布，这是客观存在的，这个客观存在的分布函数 $F(x)$ 就称为总体。

(2) 样品：所考察对象中任意抽取一个，研究的那个数量指标就是一个随机变量 X，显然 $X\sim F(x)$ 称为来自总体 X 的一个样品。

(3) 样本：有若干个样品 X_1、X_2、…、X_n 组成的一组独立同分布 $F(x)$ 的随机变量，称为来自 $F(x)$ 的一个样本。

(4) 统计量：样本的函数就称为统计量。它是数理统计中要着重研究的随机变量。

1.3　随机变量的数学特征

利用分布函数或分布密度函数可以完全确定一个随机变量，但在实际问题中求分布函数或分布密度函数十分困难。例如，测量零件的长度得到一系列观测值，往往只需要知道零件长度这个随机变量的一些特征量，比如长度的平均值及用标准差描述的随机变量的分散程度。用一些数字来描述随机变量的主要特征，在概率论和数理统计中称它们为随机变量的数学特征。这些特征量有数学期望、方差和矩等。

1.3.1　数学期望

随机变量 ξ 的数学期望值记为 $E(x)$，简记为 μ_x，用它可以表示随机变量本身的大

小，说明随机变量ξ的取值中心在数轴上的位置，也称为期望值。数学期望表征随机变量分布的中心位置，数学期望的估计值为一系列观测值的算术平均值。

1. 离散型随机变量的数学期望

若离散型随机变量ξ的取值范围为x_1、x_2、…的概率为$P_i(i=1,2,\cdots)$，则

$$E(\xi)=\sum_i x_i P(\xi=x_i=\sum_i x_i P_i)$$

称为ξ的期望值。

2. 连续型随机变量的数学期望

设ξ的分布函数为$f(x)$（连续随机变量），则ξ的期望值为

$$E(\xi)=\int_{-\infty}^{\infty} xf(x)\mathrm{d}x$$

1.3.2 方差

ξ的期望值$E(\xi)$相当一组数据的平均值，则相应于该组数据的方差就是随机变量ξ的方差。

1. 离散型随机变量的方差

若离散型随机变量取值为x_1、x_2、…、x_n、…，其概率为P_1、P_2、…、P_n、…，则它的方差为

$$D(\xi)=\sum_i[x_i-E(\xi)]^2P(\xi=x_i)$$
$$=\sum[x_i-E(\xi)]^2P_i$$

2. 连续型随机变量的方差

对连续的、密度为$f(x)$的随机变量ξ，它的方差为

$$D(\xi)=\int_{-\infty}^{\infty}[x-E(\xi)]^2f(x)\mathrm{d}x$$

从定义可以看出，随机变量ξ的期望值$E(\xi)$与方差$D(\xi)$完全是由它的分布决定的，因而也称它们是分布的期望值与方差。

3. 期望值和方差

在水利水电工程质量检测中，遵从各种分布的随机变量，现将几种常用分布列入表1.3.1。

表 1.3.1 几种常用分布的期望值和方差

名　称	概　率　分　布	期望值$E(\xi)$	方差$D(\xi)$	参数范围
二点分布	$P(\xi=x)=P^xq^{1-x}$ $(x=0,1)$	P	Pq	$0<P<1$ $q=1-P$
二项分布	$P(\xi=k)=\binom{n}{k}P^k(1-P)^{n-k}$ $(k=0,1,2,\cdots,n)$	nP	nPq	$0<P<1$ $q=1-P$

续表

名　称	概　率　分　布	期望值 $E(\xi)$	方差 $D(\xi)$	参数范围
均匀分布	$\begin{cases}\frac{1}{a-b} & x\in(a,b)\\ 0 & \text{其他}\end{cases}\quad (a<b)$	$(a+b)/2$	$(b-a)^2/12$	$b>a$
正态分布	$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$ $x\in(-\infty,\infty)$	μ	σ^2	μ（任意） $\sigma>0$
指数分布	$F(t)=1-e^{-t}$ $(t>0)$	1	1	

1.3.3　随机变量的线性变换

随机变量 ξ 大都与计量单位有关，例如 ξ 表示温度，温度有摄氏、华氏、绝对温度等几个单位来表示，相应的值也不一样。华氏 F 与摄氏 C 之间有下列关系：

$$C=\frac{5}{9}(F-32)$$

从上式可以看出，一个数是另一个数的线性函数。因此计量单位的改变往往需要考虑线性变换。

又如，用标准正态分布表查出一般正态分布函数值。

设标准正态分布函数为 $N(0,1)$，考虑它的线性函数 $t=a\xi+b$，其中 $a>0$，b 为常数，求分布密度 t。

$$P(t<y)=P(a\xi+b<y)=P\left(\xi<\frac{y-b}{a}\right)=\int_{-\infty}^{\frac{y-b}{a}}\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}du$$

两边对 y 求微商，就得到分布密度 t 为

$$t=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2a^2}(y-b)^2}$$

它就是正态分布 $N(b,a^2)$，即 $t\sim N(b,a^2)$。可见任何一个给定的分布 $N(b,a^2)$，都可以看成由标准正态分布 $N(0,1)$经过线性变换之后导出来的。即有 $t\sim N(b,a^2)$，则有 $\xi\sim N(0,1)$使

$$t=\sigma\xi+\mu\quad\text{或}\quad\xi=\frac{t-\mu}{\sigma}$$

因此，只要有一张标准正态分布表，就可以查出一般正态分布函数值。通常用$\Phi(x)$表示标准正态分布函数，$\varphi(x)$表示 $N(0,1)$的密度，即

$$\Phi(x)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}du$$

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$$

根据正态分布 $\Phi(x)$表，反查 x 的值。已知 $t\sim N(3,0.25)$，求 $P(t>3.11)$。

$$P(t>3.11)=P\left(\frac{t-3}{0.5}<\frac{3.11-3}{0.5}\right)=P(\xi<0.22)=\Phi(0.22)=0.5871 \text{（查表）}$$

由于当 $\xi\sim N(0,1)$ 时，有 $P(|\xi|<1)=0.6826$，$P(|\xi|<2)=0.9544$，$P(|\xi|<3)=0.9974$。对于一般正态分布 $N(b,a^2)$，当 $t\sim N(b,a^2)$ 时，就有

$$P(|t-\mu|<\sigma)=0.6826$$
$$P(|t-\mu|<2\sigma)=0.9544$$
$$P(|t-\mu|<3\sigma)=0.9974$$

可见，$|t-\mu|\geqslant 2\sigma$ 和 $|t-\mu|\geqslant 3\sigma$ 的概率是很小的，如果出现这种情况，则认为这个数据 x 很难接受。因此，当确认这个数据 x 来自 $N(\mu,\sigma^2)$ 这个总体时，认为这个数据 x 必须满足下列不等式：

$$|x-\mu|\leqslant 2\sigma \quad \text{或} \quad |x-\mu|\leqslant 3\sigma$$

否则就不予承认（即剔除这个数据），这就是通常所说的 2σ 原则（或 3σ 原则）。

例如，检测一组混凝土抗压强度数值遵从正态分布 $N(27.8, 9.0)$，发现试验值中有一数据为 37.2 可疑，是否应该剔除？

由 3σ 原则知道，全部检测数据应在 $27.8-3.0\times3$ 与 $27.8+3.0\times3$ 之间，即在 18.8 与 36.8 之间，而 $37.2>36.8$，故这个数据应予以剔除。

1.4 随机变量的基本定理

1.4.1 大数定理

1. 切比雪夫不等式

设 x 为随机变量，且 $E(x)=\mu$，$D(x)=\sigma^2$ 存在，则对任给 $\varepsilon>0$，有

$$P(|x-\mu|\geqslant\varepsilon)\leqslant\frac{\sigma^2}{\varepsilon^2}$$

由上式可以看出期望值 μ 与方差 σ^2 的意义，期望值 μ 是表明随机变量取值的平均数，方差 σ^2 表明随机变量的分散程度。可以看出，当 σ^2 很小时，x 与期望值 μ 之差的绝对值大于 ε 的概率很小，即 x 取值在 μ 附近。

2. 大数定理

设 x_1、x_2、…、x_n 独立同分布且 $E(x_i)=\mu$，$D(x_i)=\sigma^2(i=1,2,\cdots,n)$，则有

$$\lim_{n\to\infty}P\left(\left|\frac{1}{n}\sum_{i=1}^{n}x_i-\mu\right|\geqslant\varepsilon\right)=0$$

称为弱大数定理。

上式等价于

$$\lim_{n\to\infty}P\left(\left|\frac{1}{n}\sum_{i=1}^{n}x_i-\mu\right|<\varepsilon\right)=1$$

它表明事件 $\left(\left|\frac{1}{n}\sum_{i=1}^{n}x_i-\mu\right|<\varepsilon\right)$，当 $n\to\infty$ 时几乎总是成立的，或者当 n 充分大时，

对任意 $\varepsilon>0$，不等式 $\left|\frac{1}{n}\sum_{i=1}^{n}x_i-\mu\right|<\varepsilon$ 几乎都成立，n 很大时，平均值 $\bar{x}$ 与期望值 μ 之差的绝对值可任意小，所以可以用平均值 $\bar{x}$ 代替期望值 μ，而做试验的次数越多其代表性就越好。

$$P(\lim_{n\to\infty}\bar{x}_n=\mu)=1$$

称为强大数定理。

3. 贝努利定理

在 n 次独立试验中，事件 A 的出现次数为 m，则当 n 无限增大时，频率 m/n 收敛于它的概率 p，即对任意的 $\varepsilon>0$，则有

$$\lim_{n\to\infty}P\left(\left|\frac{m}{n}-p\right|<\varepsilon\right)=1$$

称为大数定理，也称为贝努利定理。它表明在试验的条件稳定不变时，如果 n 充分大，则可用频率代替概率，此时频率具有很高的稳定性。

1.4.2　中心极限定理

设 x_1、x_2、…、x_n 是一组独立同分布的随机变量，且 $E(x_i)=\mu$，$D(x_i)=\sigma^2(i=1,2,\cdots,n)$，则有

$$\lim_{n\to\infty}P\left(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}<x\right)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{n}e^{\frac{\mu^2}{2}}du$$

称为中心极限定理。大数定理知识粗略地描述了随机变量取值的极限情况，中心极限定理确切地说明 x_n 的极限分布是正态分布，也就是说，无论 x_1、x_2、…、x_n 服从什么分布，只要是独立同分布，则 $\bar{x}_n$ 趋向正态分布。由此可以看出正态分布在概率统计中的地位，在数理统计中常常设样本 x_1、x_2、…、x_n 服从正态分布（也可能本来不是正态分布）。只要 n 较大，中心极限定理也能保证相当好地接近正态分布。

1.5　行列式和矩阵

1.5.1　行列式

1. 二元线性方程组和二阶行列式

我们知道，一次方程又称为线性方程，一次方程组又称为线性方程组。含有两个未知数、两个一次方程称为二元线性方程组。二元线性方程组一般形式为

$$\begin{cases}a_{11}x+a_{12}y=b_1 & \text{(a)}\\ a_{21}x+a_{22}y=b_2 & \text{(b)}\end{cases}$$

下面用消元法解这个方程组。

由 (a) $\times a_{22}-a_{12}\times$ (b)，得

$$(a_{11}a_{22}-a_{12}a_{21})x=b_1a_{22}-a_{12}b_2 \qquad \text{(c)}$$

由 $a_{11}\times$ (b) − (a) $\times a_{21}$，得

$$(a_{11}a_{22}-a_{12}a_{21})y=b_2a_{11}-a_{21}b_2 \tag{d}$$

如果 $a_{11}a_{22}-a_{12}a_{21}b_2\neq 0$，则得到方程组的求解公式：

$$\begin{cases} x=\dfrac{b_1a_{12}-a_{12}b_2}{a_{11}a_{22}-a_{12}a_{21}} \\ y=\dfrac{b_2a_{11}-a_{21}b_1}{a_{11}a_{22}-a_{12}a_{21}} \end{cases} \tag{1.5.1}$$

将任意一个二元线性方程组化程一般形式后，只要 $a_{11}a_{22}-a_{12}a_{21}b_2\neq 0$。就可以利用式（1.5.1）求解。

为了便于应用，我们引入二阶行列式的概念。

我们把方程组（1.5.1）中的 x、y 的系数，按照它们在方程组中原来的位置排列起来，并在两旁各加一条竖线，就得出下列形式：

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \tag{1.5.2}$$

用它来表示左上角和右下角的两个数的积 $a_{11}a_{22}$减去右上角和左下角的两个数的积 $a_{12}a_{21}$之差，即

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}=a_{11}a_{22}-a_{12}a_{21} \tag{1.5.3}$$

式（1.5.2）有二行二列，称为二阶行列式，其中 a_{11}、a_{12}、a_{21}、a_{22}称为二阶行列式的元素。式（1.5.3）右边的式子称为该二阶行列式的展开式。实线上两个元素的积取正号，虚线上两个元素的积取负号。

利用行列式，式（1.5.1）就可以写成

$$\begin{cases} x=\dfrac{\begin{vmatrix} b_1 & a_{12} \\ b_2 & a_{22} \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}} \\ y=\dfrac{\begin{vmatrix} a_{12} & b_1 \\ a_{21} & b_2 \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}} \end{cases} \quad \left(\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}\neq 0\right) \tag{1.5.4}$$

为了简便起见，通常用 D、D_x、D_y 分别表示式（1.5.4）中作为分母与分子的行列式：

$$D=\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix},\ D_x=\dfrac{\begin{vmatrix} b_1 & a_{12} \\ b_2 & a_{22} \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}},\ D_y=\dfrac{\begin{vmatrix} a_{12} & b_1 \\ a_{21} & b_2 \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}}$$

此时，方程组（1.5.1）的解可以写为