第4章 概率及其分布

自然界和人类社会中存在着两类现象：

必然现象与随机现象。

必然现象就是指在一定条件下必然发生的现象。如在标准大气压下，水加热到100℃必然会沸腾；抛出的物体其初速度只要小于第一宇宙速度，必然会落回地面等，这些现象都是必然现象。

随机现象是指在一定条件下有时发生有时不发生的现象。如往地面掷一枚硬币，“出现正面”这一现象有时发生，有时不发生；篮球投篮中“投中”这一现象有时发生有时不发生，这些都是随机现象。

4.1 随机事件及其概率

一、随机事件

对随机现象进行观察，会观察到不同的结果，如观察掷硬币这一随机现象就可能看到“出现正面”或“出现反面”这两种不同的结果，“出现正面”是掷硬币这一随机现象的一种观察结果，我们称之为随机事件，同样“出现反面”也是随机事件。

随机事件：对随机现象进行观察，其观察结果叫随机事件，简称事件,用大写英文字母A、B、C等表示。

作为随机事件的特例，若某事件在每次试验中总是发生，则称该事件为**必然事件**，一般用字母Ω表示；反之若某事件在每次试验中都不发生，则称该事件为**不可能事件**，一般用字母Φ表示。

二、随机事件的概率

1.频率

在相同的条件下，进行了n次试验，在这n次试验中事件A出现了m次，则称比值m/n为事件A的频率，记为F（A）＝m/n。

显然任一事件A都有

0≤F（A）≤1

上页 下页 返回

以掷硬币为例，记正面向上为随机事件A，抛掷总次数为n，出现正面向上的次数为m，比值

F＝m/n为事件A的频率。

随着试验次数的增多，随机事件频率的波动会越来越小，且会在一个固定的常数附近作微小的波动。

上页 下页 返回

表 4-2　历史上数学家们抛掷硬币的试验数据

实验者	掷硬币次数	出现正面次数	出现正面频率
蒲丰	4040	2048	0.5069
皮尔逊	12000	6019	0.5016
皮尔逊	24000	12012	0.5005

可以用一个数来描述随机事件在一次试验中发生的可能性大小，该数就是概率。

2.概率

随机事件的概率：在n次重复试验中随机事件A发生的次数记为m，当n很大时，频率m/n会稳定地在某一数值p的附近摆动，而且随着试验次数n的增加，其摆动的幅度越来越小，称p为随机事件A的概率，记为：

P（A）＝p

例如，在投硬币的试验中，“出现正面”这一随机事件发生的频率在０.５附近摆动，且随着试验次数的增多摆动的幅度会越来越小，因此，可以认为“出现正面”这一随机事件的概率为０.５。

而对不可能事件Φ必有m=0，对必然事件Ω一定有m=n，可知它们的频率为

F（Ω）=1，F（Φ）=0

对概率类似地有

0≤P（A）≤1

以及

P（Ω）=1， P（Φ）=0

例：在电话号码簿中任取一号码，求后面4位数全不相同的概率。

解：

$$P(A)=\frac{10\times 9\times 8\times 7}{10\times 10\times 10\times 10}=0.504$$

例：投掷两枚骰子，求两枚骰子中至少有一枚出现6点且点数之和为偶数的概率？

解：有得场合为：（6，2），（6，4），（6，6），（2，6），（4，6）

则：P（A）=5/36=0.139

例如，赌徒心理：前几次赌博都输了，后面赢的希望较大；超生的孕妇，可能认为前几个孩子都是女孩，后面生男孩的希望应该较大。这些观点都是错误的，其实概率是一样的。

例子，蒲丰曾投掷硬币 4040次，得正面2048次；

皮尔逊曾投掷硬币 12000次，得正面6019次；24000次，得正面12012次。

3.小概率事件原则

一般若P（A）$\leq$0.05，则称事件A为小概率事件。

小概率事件在一次试验中可看作不可能事件，认为不可能发生，这一原则称之为小概率事件原则。小概率事件原则是统计推断的重要原则，在以后的学习中将会多次用到此原则。

4.2 随机变量及其概率分布

一、随机变量

当用一个变量的取值来表示随机试验的结果时，该变量随着试验的不同结果而取不同的值，也就是说变量的取值是随机的，称此变量为随机变量，随机变量一般用大写英文字母X、Y、Z表示，也可以用ξ、η等表示。

二、随机变量的概率分布

1.概率分布的概念

概率分布：随机变量的取值及取值的概率称为随机变量的概率分布。

2.概率分布的表示方法

⑴ 分布列法

⑵ 分布曲线法

分布列：设随机变量 X 可能取到的值为 $X_1, X_2, \ldots X_n, \ldots$，X 取到各个值的概率为 p_1，$p_2, \ldots p_n, \ldots$，则称

X 的取值	X_1	X_2	…	X_n	…
X 取对应值的概率	p_1	p_2	…	p_n	…

为随机变量 X 的概率分布列或**分布列**。

［例 4-1］ 用 X=n 表示射击中命中 n 环（n=0、6、7、8、9、10），某运动员在一段时间内的射击水平可用 X 的概率分布表示如下：

X 的取值	0	6	7	8	9	10
X 取值的概率	0.01	0.14	0.3	0.35	0.15	0.05

例如： $6 \le X \le 8$的概率（即该运动员命中6至8环的概率）为：

$P(6 \le X \le 8) = 0.14 + 0.3 + 0.35 = 0.79$

上页 下页 返回

[例4-2] 某一不透明的盒中装有10个外形一样的球，其中5个黑球，5个白球，现从中任取3球，用Y表示取到的白球数，求Y的概率分布列。

分析：求Y的概率分布列，就是求Y能取哪些值及取这些值的概率。

解: 由于取出的3个球中可能有0个白球，1个白球，2个白球，3个白球，因此Y的取值范围为0、1、2、3。

从 10 个球中任取 3 球的总取法数 $n=C_{10}^{3}=\dfrac{10\times9\times8}{3\times2\times1}=120$

其中所取的 3 个球全是黑球（此时 Y=0）的取法数 $m=C_{5}^{3}=\dfrac{5\times4\times3}{3\times2\times1}=10$

所以

$$P（Y=0）=\frac{10}{120}=0.08$$

类似求出

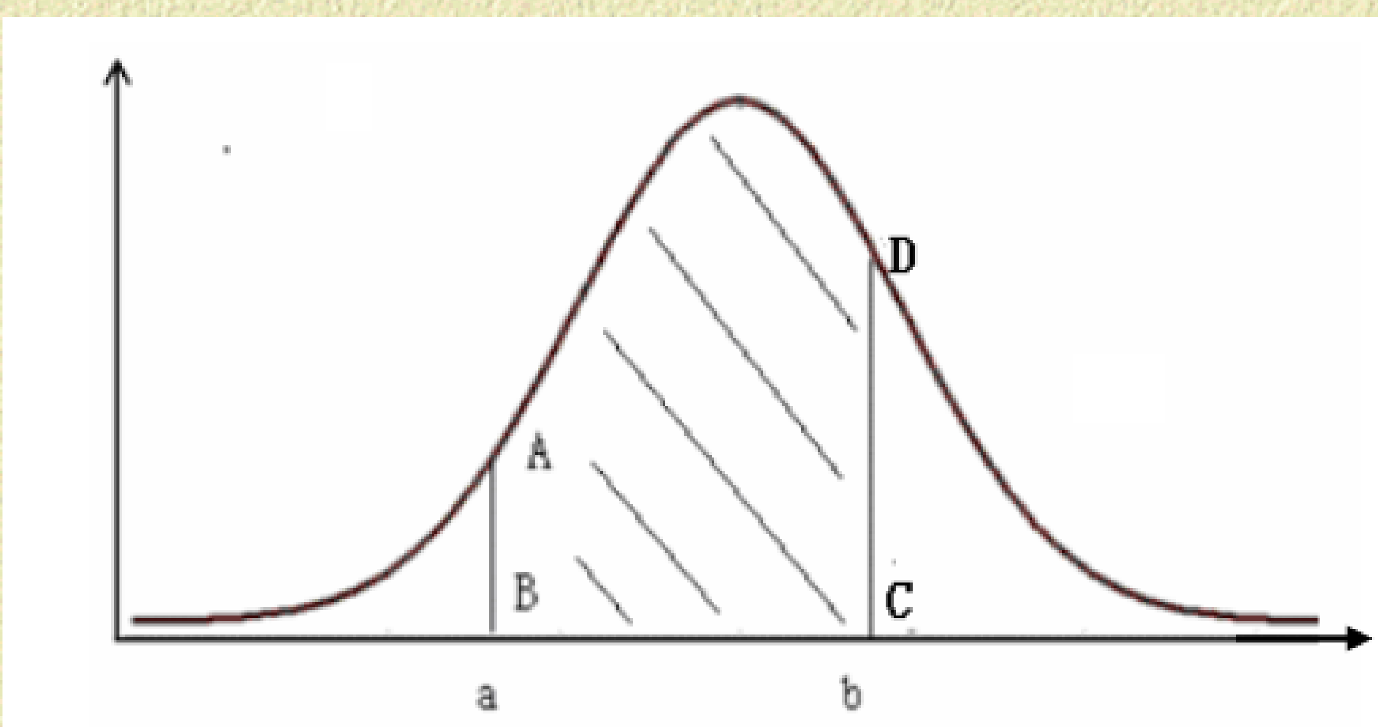
P（Y=1）=0.42，P（Y=2）=0.42，P（Y=3）=0.08

Y 的概率分布列如下：

Y 的取值	0	1	2	3
Y 取各值的概率	0.08	0.42	0.42	0.08

上页 下页 返回

概率密度曲线的含义：概率密度曲线与X轴、直线X＝a、直线X＝b所组成的曲边梯形的面积等于随机变量ξ的取值落在区间［a，b］内的概率。



上页 下页 返回

4.3 几种常用的概率分布

一、两点分布

[例 4-4] 不透明盒中装有 100 只外形一样的球，其中 **90** 个白球，**10** 个红球，从中随机取一个球，显然取到白球的概率为 0.9，取到红球的概率为 0.1，求取球结果的概率分布。

解：令 $X=\begin{cases}1, & \text{当取得白球时}\\ 0, & \text{当取得红球时}\end{cases}$

则取球结果的概率分布如下：

X 的取值	0	1
X 取各值的概率	0.1	0.9

即 X 是两点分布。

二、二项分布

贝努里试验：一般地，如果在相同条件下进行了n次相互独立的试验，每次试验只有两个可能的结果：A或 $\overline{A}$，且P（A）=p，相应地P（$\overline{A}$）=q=1-p，则称这样的n次试验为n重贝努里试验。

$$P(X=k)=C_n^k p^k (1-p)^{n-k} \quad （k=0,1,2,\ldots n）$$

例：

掷一枚质地均匀的硬币，重复地掷 5 次，记正面朝上的次数为随机变量 x，求 x=2 的概率。

解：P（x=2）$=C_5^2\cdot\left(\frac{1}{2}\right)^2\cdot\left(\frac{1}{2}\right)^3=\frac{5}{16}=0.3125$

其中：$C_5^2=\frac{5!}{(5-2)!\cdot 2!}=\frac{5!}{3!\cdot 2!}=10$

1．排列 $P_n^m=\frac{n!}{(n-m)!}=\frac{P_n}{P_{n-m}}$

2．组合 $C_n^m=\frac{n!}{m!\cdot(n-m)!}=\frac{P_n^m}{P_m}$

上页 下页 返回

[例4-6] 不透明盒中装有100只外形一样的球，其中10个白球，90个红球，采用有放回取球方式，从中任取5球，求恰好取到3个白球的概率。

解：将取一个球看作一次试验，取 5 个球相当于做了 5 重贝努力试验，用 X 表示取到的白球数，显然 X 是n=5，p=0.1的二项分布，恰好取到 3 个白球就是 X ＝ 3 ，由二项分布可知其概率为：

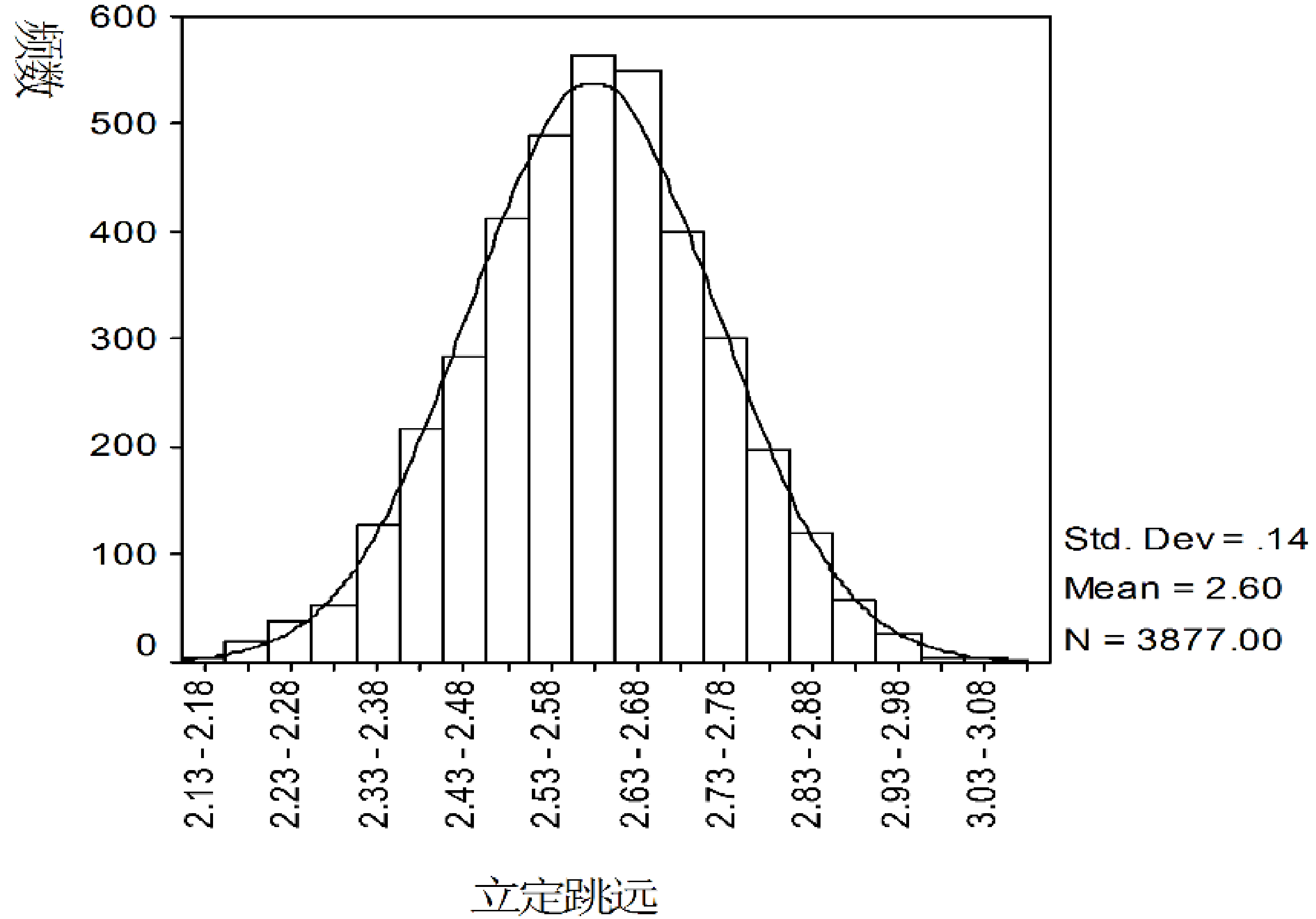
$$P(X=3)=C_5^3\,0.1^3(1-0.1)^{5-3}=0.0081$$

三、正态分布

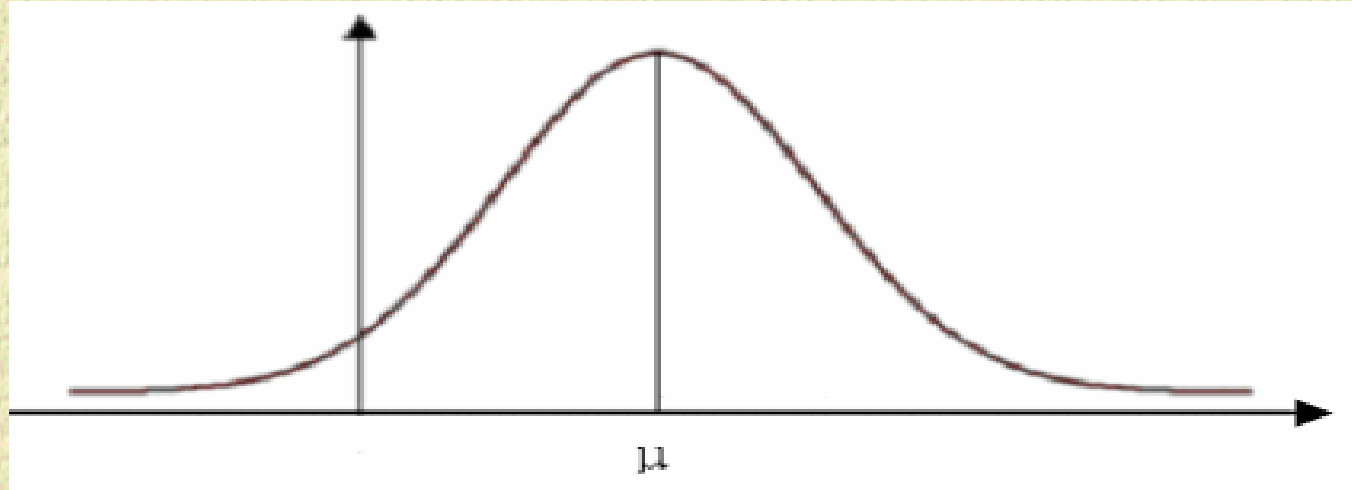
（一） 正态分布的概念

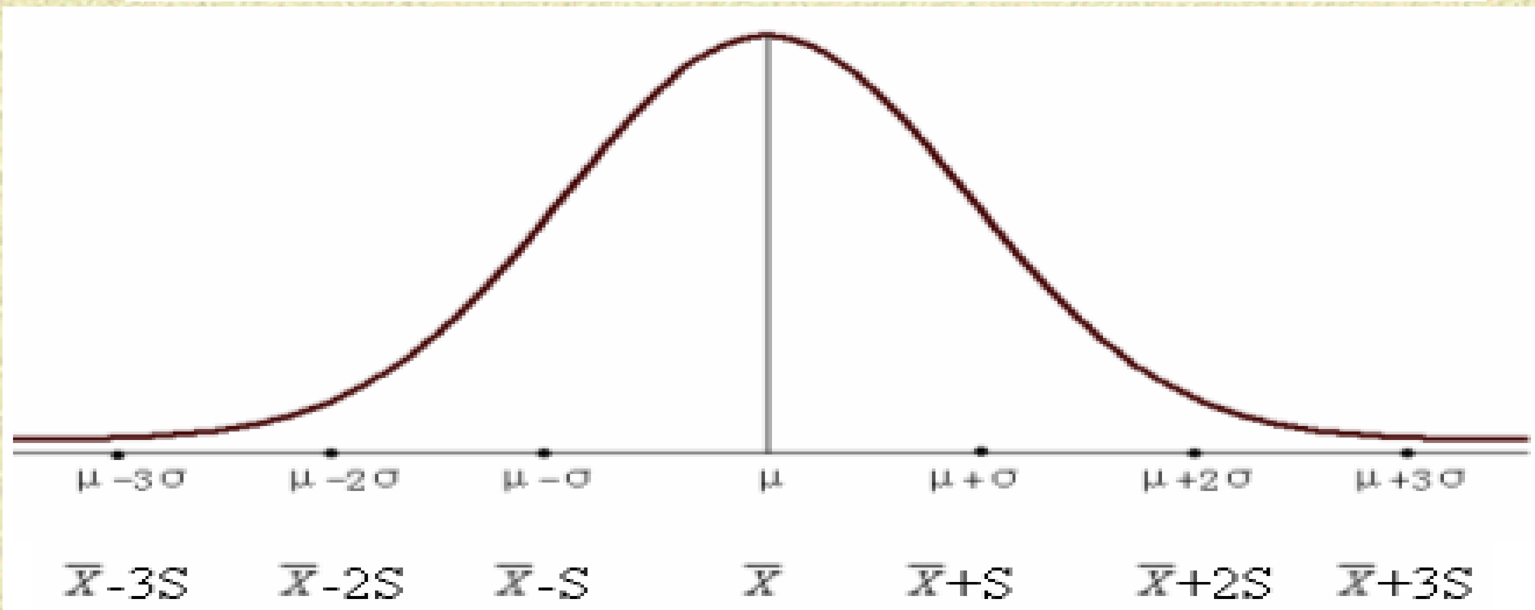
正态分布是实践中最为常见的一种分布，如同年龄、同性别人的身高、体重、运动成绩等都服从正态分布。

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


频数
600
500
400
300
200
100
0
2.13 - 2.18
2.23 - 2.28
2.33 - 2.38
2.43 - 2.48
2.53 - 2.58
2.63 - 2.68
2.73 - 2.78
2.83 - 2.88
2.93 - 2.98
3.03 - 3.08
立定跳远
Std. Dev = .14
Mean = 2.60
N = 3877.00

上页 下页 返回





记为X～N（μ，σ²），其对应的曲线叫正态曲线。