本及它们的类别标号，要设法根据这些样本求出这个映射或者对这个映射的一个较好的估计。一个显然的条件是，对于任何一个实际问题来说，已知的样本数只能是有限的。

让我们回顾一下概率论中的一个核心命题——大数法则，其内容包括三个方面：

(1)（伯努利定理）随机事件 A 在 n 次独立试验中的频率 $\frac{\nu}{n}$ 依概率收敛于事件 A 的概率 p，即对任意 $\epsilon>0$，有

$$\lim_{n\to\infty}P\left\{\left|\frac{\nu}{n}-p\right|<\epsilon\right\}=1。$$

(2) 互相独立的随机变量 $x_1,x_2,\cdots$，如果①存在均值和方差（分别记为 μ 和 σ^2），或者②具有相同分布，且有有限均值 μ，那么，$\frac{1}{n}\sum_{i=1}^{n}x_i$ 依概率收敛于随机变量的均值，即对任意 $\epsilon>0$，有

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

(3) 如果互相独立且具有相同分布的随机变量 $x_1,x_2,\cdots$ 的均值和方差都存在（分别记为 μ 和 σ^2），那么，$\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\frac{1}{n}\sum_{j=1}^{n}x_j\right)^2$ 依概率收敛于随机变量的方差，即对任意 $\epsilon>0$，有

$$\lim_{n\to\infty}P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\frac{1}{n}\sum_{j=1}^{n}x_j\right)^2-\sigma^2\right|<\epsilon\right\}=1。$$

回顾本书第 2 章到第 6 章的内容，其中很多方法和结论都是建立在大数法则的基础上的。也就是说，我们一直在使用一个隐含的命题，就是在样本数趋向无穷的假设下得到的结论，当样本数有限时仍是有效的，或者至少是一种不错的近似。这一命题似乎是符合直观认识的，因此很长时间里没有被多数人注意。而本章的内容就是对这一命题的真实性进行研究，指出这个命题的基本思想是用经验风险最小化代替实际中无法实现的平均风险最小化（或称为"期望风险最小化"）。而在经验风险最小化准则下设计的分类器，其性能与理论上使平均风险最小化的最优分类器的性能是有差距的，进而对这一差距的界进行了研究，并得出了关于为使某一分类器达到一定的性能而必需具有的训练样本数的初步结论，同时提出了更合理的有序风险最小化的思想。可以说，在一定意义上，这些问题都是模式识别中的根本问题，本章介绍的内容主要是在 20 世纪 80 年代以前在这方面的研究成果。近十几年来，这些成果又得到了很大的发展，逐渐成为一个更完善的理论体系，这就是第 13 章将专门介绍的统计学习理论。虽然在统计学习理论中，本章介绍的部分结论已经得到了更新，但是为了让读者对这一理论的发展历程有一个更全面的了解，我们编写本书（第二版）时仍保留了第一版中本章的内容。

习　　题

7.1　试证明

$$\Phi(D,N)=\begin{cases}2\sum_{i=0}^{D-1}\binom{N-1}{i}, & 若 D\leqslant N\\ 2^N, & 若 D>N\end{cases}$$

是递归方程

$$\Phi(D,N) = \Phi(D,N-1) + \Phi(D-1,N-1)$$

的解且 $\Phi(D,1)=2$，$\Phi(1,N)=2$。

7.2 对于线性判决规则类，若特征空间的维数等于 5，当用经验风险最小化进行分类器设计，希望分类器的误识率小于 0.1 的概率不小于 90%，问最少需要多少个学习样本？当特征数减少到 2 时，所需的样本数可以减少多少？当特征数为 2，但用二次多项式作为判决函数类时，最少需要多少样本？

7.3 讨论用两段线性函数类作为判决规则类时，判决规则品质 ξ，样本数 N，维数 D 和可靠性参数 η 之间的关系。

7.4 讨论用有序风险最小化选取线性函数判决规则类的最优特征组合的方法。