# 理解概率密度函数

概率密度函数是概率论中的核心概念之一，用于描述连续型随机变量所服从的概率分布。在机器学习中，我们经常对样本向量 x 的概率分布进行建模，往往是连续型随机变量。很多同学对于概率论中学习的这一抽象概念是模糊的。在今天的文章中，SIGAI 将直观的解释概率密度函数的概念，帮你更深刻的理解它。

## 从随机事件说起

回忆我们在学习概率论时的经历，随机事件是第一个核心的概念，它定义为可能发生也可能不发生的事件，因此是否发生具有随机性。例如，抛一枚硬币，可能正面朝上，也可能反面朝上，正面朝上或者反面朝上都是随机事件。掷骰子，1 到 6 这 6 种点数都可能朝上，每种点数朝上，都是随机事件。

与每个随机事件 a 关联的有一个概率值，它表示该事件发生的可能性：

$$p(a)$$

这个概率值必须在 0 到 1 之间，22 即满足下面的不等式约束：

$$0 \le p(a) \le 1$$

另外，对于一次实验中所有可能出现的结果，即所有可能的随机事件 $a_1, a_2, ..., a_n$，它们的概率之和必须为 1：

$$\sum_{i=1}^{n} p(a_i) = 1$$

这些随机事件不会同时发生，但必须有一件会发生。例如，对于抛硬币，不是正面朝上就是反面朝上，不会出现其他情况（这里假设硬币抛出去后不会立着），因此有：

$$p(\text{正面朝上}) + p(\text{反面朝上}) = 1$$

很多时候，我们假设这些基本的随机事件发生的概率都是相等的，因此，如果有 n 个基本的随机事件，要使得它们发生的概率之和为 1，则它们各自发生的概率都为：

$$\frac{1}{n}$$

对于抛硬币，正面朝上和反面朝上的概率各为 1/2，对于掷骰子，每个点朝上的概率各为 1/6。对于这种只有有限种可能的情况，我们通过枚举各种可能的情况，可以算出每个事件发生的概率。例如，如果我们要计算掷骰子出现 1 点或者 2 点的概率，只需要将这两点至少有一点出现的情况数，比上所有可能的情况数，就得到概率值：

$$\frac{\text{1点朝上的情况数+2点朝上的情况数}}{\text{所有可能的情况数}}=\frac{2}{6}$$

上面的例子中，随机事件所有可能的情况只有有限种，而且可以用整数对这些随机事件进行编号，如 $a_1, a_2, a_3 \ldots$。

然而，有有限就有无限，对于可能有无限种情况的随机事件，我们该如何计算它发生的概率？考虑一个简单的问题，有一个长度和高度都为 1 的正方形，如果我们随机的扔一个点到这个正方形里，这个点落在右上方也就是红色区域里的概率是多少？

你可能已经想到了，直接用红色三角形的面积，比上整个正方形的面积，应该就是这个概率

$$\frac{\text{红色三角形的面积}}{\text{正方形的面积}}=\frac{1}{2}$$

在这里，随机点所落的位置坐标(x, y)的分量 x 和 y 都是[0, 1]区间内的实数，这有无限多种情况，不能再像之前那样把所有的情况全部列出来，统计出这些情况的数量，然后和总情况数相除得到概率值。而是使用了“面积”这一指标来计算。看来，对这种类型的随机事件，我们得借助于“长度”，“面积”，“体积”这样的积分值来计算。

如果用集合来描述这些随机事件的话，第一种情况是有限集，我们可以给集合里的每个元素编号。第二种情况是无限集，元素的个数多到无法用整数下标来编号。

## 整数集与实数集

高中时我们学过集合的概念，并且知道整数集是 z，实数集是 R。对于有限集，可以统

计集合中元素的数量即集合的基数（cardinal number，也称为集合的势 cardinality）。对于无限集，元素的个数显然是无穷大，但是，都是无穷大，能不能分个三六九等呢？

回忆微积分中的极限，对于下面的极限：

$$\lim_{x\to+\infty}\frac{x}{e^x}=0$$

虽然当 x 趋向于正无穷的时候，x 和 exp(x)都是无穷大，但它们是有级别的，在 exp(x)面前，x 是小巫见老巫。

同样的，对于整数集和实数集，也是有级别大小的。任意两个整数之间，如 1 与 2 之间，都密密麻麻的分布着无穷多个实数，而且，只要两个实数不相等，不管它们之间有多靠近，如 0.0000001 和 0.0000002，在它们之间还有无穷多个实数。在数轴上，整数是离散的，而实数则是连续的，密密麻麻的布满整个数轴。因此，实数集的元素个数显然比整数要高一个级别。

## 随机变量

变量是我们再熟悉不过的概念，它是指一个变化的量，可以取各种不同的值。随机变量可以看做是关联了概率值的变量，即变量取每个值有一定的概率。例如，你买彩票，最后的中奖金额 x 就是一个随机变量，它的取值有 3 种情况，以 0.9 的概率中 0 元，0.09 的概率中 100 元，0.01 的概率中 1000 元。变量的取值来自一个集合，可以是有限集，也可以是无限集。对于无限集，可以是离散的，也可以是连续的，前者对应于整数集，后者对应于实数集。

## 离散型随机变量

随机变量是取值有多种可能并且取每个值都有一个概率的变量。它分为离散型和连续型两种，离散型随机变量的取值为有限个或者无限可列个（整数集是典型的无限可列），连续型随机变量的取值为无限不可列个（实数集是典型的无限不可列）。

描述离散型随机变量的概率分布的工具是概率分布表，它由随机变量取每个值的概率 $p\left(x=x_i\right)=p_i$ 依次排列组成。它满足：

$$p_i \geq 0$$
$$\sum p_i = 1$$

下面是一个概率分布表的例子：

**表 2.2　一个随机变量的概率分布表**

| $x$ | **概率值** |
|---|---|
| 1 | 0.1 |
| 2 | 05 |
| 3 | 0.2 |
| 4 | 0.2 |

如果我们把前面例子中掷骰子的点数 x 看做是随机变量，则其取值为 1-6 之间的整数，取每个值的概率为 1/6，这是典型的离散型随机变量。

## 连续型随机变量

把分布表推广到无限情况，就可以得到连续型随机变量的概率密度函数。此时，随机变量取每个具体的值的概率为 0，但在落在每一点处的概率是有相对大小的，描述这个概念的，

就是概率密度函数。你可以把这个想象成一个实心物体，在每一点处质量为 0，但是有密度，即有相对质量大小。

以上面在正方形内随机扔一个点的问题为例，此时，落点的坐标(x, y)就是连续型随机变量，落到任意一点(x, y)的概率值为 0。因为这一个点的数量为 1，而整个正方形内的点数为无穷大，二者之比值为 0：

$$\frac{1}{[0,1]\text{内的正方形的点数}}=0$$

这实际上是均匀分布，即落在任何一点处的概率值相等。对于有些问题，落在各个不同的点处的概率是不相等的，就像一个实心物体，有些点处的密度大，有些点处的密度小，由此引入了概率密度函数的概念。

一个函数如果满足如下条件，则可以称为概率密度函数：

$$f(x)\geq 0$$

$$\int_{-\infty}^{+\infty} f(x)dx=1$$

这可以看做是离散型随机变量的推广，积分值为 1 对应于取各个值的概率之和为 1。分布函数是概率密度函数的变上限积分，它定义为：

$$F(y)=p(x\leq y)=\int_{-\infty}^{y} f(x)dx$$

显然这个函数是增函数，而且其最大值为 1。分布函数的意义是随机变量 $x\leq y$ 的概率。注意，连续型随机变量取某一个值的概率为 0，但是其取值落在某一个区间的值可以不为 0：

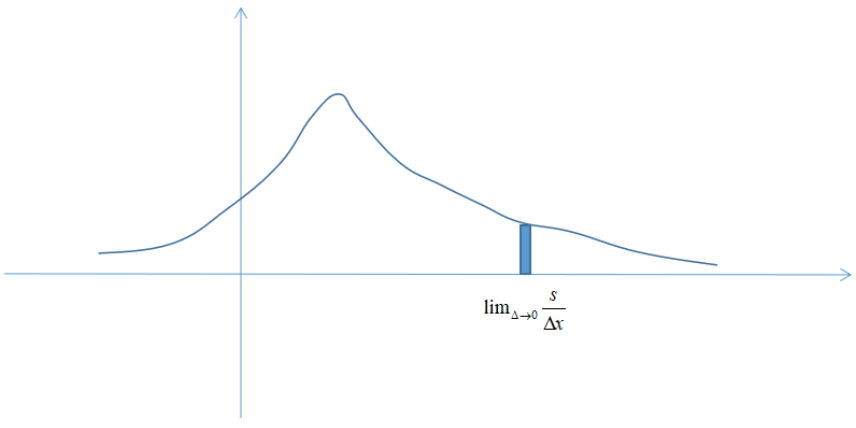
$$p(x_1<x<x_2)=\int_{x_1}^{x_2} f(x)dx=F(x_2)-F(x_1)$$

虽然连续型随机变量取一个值的概率为 0，但取各个不通过的值的概率还是有相对大小的，这个相对大小就是概率密度函数。这就好比一个物体，在任意一点处的质量为 0，但在这一点有密度值，密度值衡量了在各点处的质量的相对大小。

从这个角度，我们可以将概率密度函数解释为随机变量落在一个区间内的概率与这个区间大小的比值在区间大小趋向于 0 时的极限：

$$\lim_{\text{区间大小}\to 0}\frac{\text{落在这个区间内的概率}}{\text{区间大小}}$$

这个过程如下图所示：



还是以上面的正方形为例，如果要计算随机点(x, y)都落在区间[0, 0.5]内的概率，可以这样计算：

$$\frac{[0,0.5]\text{这个区间的正方形的面积}}{[0,1]\text{这个区间的正方形的面积}}=\frac{0.5\times 0.5}{1\times 1}=0.25$$

这个面积，就是积分值，对应于分布函数。最常见的连续型概率分布是正态分布，也称为高斯分布。它的概率密度函数为：

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

其中 $\mu$ 和 $\sigma^2$ 分别为均值和方差。现实世界中的很多数据，例如人的身高、体重、寿命等都近似服从正态分布。另外一种常用的分布是均匀分布，如果随机变量 $x$ 服从区间 $\left[a,b\right]$ 内的均匀分布，则其概率密度函数为：

$$f\left(x\right)=\begin{cases}\dfrac{1}{b-a} & a\le x\le b\\ 0 & x<a,x>b\end{cases}$$

在程序设计和机器学习中，这两种分布是最为常见的。