

# 概率论基本概念汇总

---

## 1. 联合概率与边缘概率

如果知道联合概率分布  $P(x, y)$ , 可以根据求和法则来计算边缘概率  $P(x)$ :

- 离散随机变量

$$\forall x \in \mathbf{x}, P(\mathbf{x} = x) = \sum_y P(\mathbf{x} = x, y = y)$$

- 连续随机变量

$$p(x) = \int p(x, y) dy$$

## 2. 条件概率

### 2.1 定义

给定  $\mathbf{x} = x$ ,  $y = y$  发生的条件概率记为  $P(y = y | \mathbf{x} = x)$ , 定义如下:

$$P(y = y | \mathbf{x} = x) = \frac{P(y = y, \mathbf{x} = x)}{P(\mathbf{x} = x)}$$

### 2.2 链式法则

任何多维随机变量的联合概率分布, 都可以分解成只有一个变量的条件概率相乘的形式:

$$P(\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)}) = P(\mathbf{x}^{(1)}) \prod_{i=2}^n P(\mathbf{x}^{(i)} | \mathbf{x}^{(1)}, \dots, \mathbf{x}^{(i-1)})$$

这个规则被称为概率的**链式法则** (chain rule) 或者**乘法法则** (product rule)。它可以直接从条件概率的定义中得到。例如, 使用两次定义可以得到、

$$\begin{aligned} P(a, b, c) &= P(a | b, c) P(b, c) \\ P(b, c) &= P(b | c) P(c) \\ P(a, b, c) &= P(a | b, c) P(b | c) P(c) \end{aligned}$$

## 3. PMF, PDF与CDF

### 3.1 PMF

**离散型**随机变量的概率分布使用**概率质量函数** (probability mass function, PMF) 来描述。概率质量函数是离散随机变量在各特定取值上的概率, 即

$$P(x) = \Pr(X = x)$$

如果一个函数 $P$ 是随机变量 $x$ 的PMF，则必须满足以下条件：

- $P$ 的定义域必须是 $x$ 所有可能状态的集合；
- $\forall x \in \mathbf{x}, 0 \leq P(x) \leq 1$ 。不可能发生的事件概率为 0，并且不存在比这概率更低的状态。类似的，能够确保一定发生的事件概率为 1，并且不存在比这概率更高的状态；
- $\sum_{x \in \mathbf{x}} P(x) = 1$ 。即需要满足归一化条件。

### 3.2 PDF

**连续型**随机变量的概率分布用概率密度函数（probability density function, PDF）来描述。用PDF在某一区间上的积分来刻画随机变量落在这个区间中的概率，即

$$\Pr(a \leq X \leq b) = \int_a^b p(x) dx$$

PDF并没有直接给出特定状态的概率，相对地，它给出了落在面积为 $\Delta x$ 的无限小区域内的概率为 $p(x)\Delta x$ 。

PDF需要满足以下条件：

- $p$ 的定义域必须是 $x$ 所有可能状态的集合；
- $\forall x \in \mathbf{x}, p(x) \geq 0$ 。注意，这里并不要求 $p(x) \leq 1$ ；
- $\int p(x) dx = 1$

### 3.3 CDF

累积分布函数（cumulative distribution function），用来完整地描述一个随机变量 $X$ 的概率分布。不管是离散还是连续随机变量，都可以定义它的CDF。

CDF代表的是随机变量 $X$ 的取值小于某个特定值 $x$ 的概率，即分布函数 $F(x)$ 在点 $x$ 处的值表示 $X$ 落在区间 $(-\infty, x]$ 内的概率。

对于连续型随机变量，CDF就是PDF的积分，而PDF是CDF的导数，有：

$$F_X(x) = \Pr(X \leq x) = \int_{-\infty}^x p_X(t) dt$$

对于离散型随机变量，其CDF是分段函数，比如在掷硬币中，正面向上记1，反面向上记0，那么有

$$F_X(x) = \Pr(X \leq x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{1}{2} & \text{if } 0 \leq x < 1 \\ 1 & \text{if } x \geq 1 \end{cases}$$

#### 【概念的区别】

1) PDF是连续变量特有的，PMF是离散随机变量特有的； 2) PDF的取值本身不是概率，它是一种趋势（或者变化率、密度，类似于在某一点的导数），只有对连续随机变量的取值进行积分后才是概率，也就是说对于连续值确定它在某一点的概率是没有意义的； 3) PMF的取值本身代表该值的概率。

#### 【对PDF的理解】

可以把PDF与物理中一个物体的密度作类比。

如果在区间 $[a, b]$ 上随机取点，求取在某一点处的概率，由于点的长度无限小，此概率一定为0，并没有太大意义。

这时我们就需要引入PDF，来求解所取得点落在某一段上的概率。如果这个小段足够小，那么从微分元的角度来说，概率密度函数 $p(x)$ 的值乘以微分元 $dx$ 可以得到小区间 $(x, x + dx)$ 上的概率近似值。也就是说 $P(x < X < x + dx) \approx p(x)dx$ 。

## 4. 期望

离散型随机变量的期望的定义：

$$\mathbb{E}[f] = \sum_x p(x) f(x)$$

连续型随机变量的期望的定义：

$$\mathbb{E}[f] = \int p(x) f(x) dx$$

期望可以用满足概率分布 $p(x)$ 的有限 $N$ 个点的平均来估计：

$$\mathbb{E}[f] \simeq \frac{1}{N} \sum_{n=1}^N f(x_n)$$

期望的性质：

- 期望是线性函数： $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$ ；其中 $X$ 和 $Y$ 为在同一概率空间的两个随机变量（可以独立或者非独立）， $a$ 和 $b$ 为任意实数；
- 在一般情况下，两个随机变量的积的期望值不等于这两个随机变量的期望值的积。仅当两个变量相互独立时，才有 $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ ；
- 常数随机变量的期望值仍然是它本身： $\mathbb{E}(a) = a$ 。

## 5. 方差

定义为：

$$\text{var}[f] = \mathbb{E}[(f(x) - \mathbb{E}[f(x)])^2]$$

可以进一步展开为：

$$\text{var}[f] = \mathbb{E}[f(x)^2] - \mathbb{E}[f(x)]^2$$

即平方的期望减去期望的平方。

方差的性质：

- 一个常数随机变量的方差为0，反之也成立：

$$P(X = a) = 1 \Leftrightarrow \text{Var}(X) = 0$$

- $\text{Var}(X + a) = \text{Var}(X)$
- $\text{Var}(aX) = a^2 \text{Var}(X)$
- 两个随机变量和的方差为：

$$\begin{aligned}\text{Var}(aX + bY) &= a^2 \text{Var}(X) + b^2 \text{Var}(Y) + 2ab \text{Cov}(X, Y) \\ \text{Var}(X - Y) &= \text{Var}(X) + \text{Var}(Y) - 2 \text{Cov}(X, Y)\end{aligned}$$

## 6. 协方差

协方差衡量两个随机变量之间线性相关性的强度：

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$

意义：

协方差的绝对值如果很大则意味着变量值变化很大并且它们同时距离各自的均值很远。如果协方差是正的（正相关），那么两个变量都倾向于同时取得相对较大的值。如果协方差是负的（负相关），那么其中一个变量倾向于取得相对较大的值的同时，另一个变量倾向于取得相对较小的值，反之亦然。**如果协方差为0，那么两个变量不是线性相关的（不代表独立，因为还有可能非线性相关），但是如果两个变量独立，那么它们的协方差一定为0。**

性质：

- $\text{cov}(X, X) = \text{var}(X)$
- $\text{cov}(X, Y) = \text{cov}(Y, X)$
- $\text{cov}(aX, bY) = ab \text{cov}(X, Y)$

### 6.1 协方差矩阵：

假设  $X$  是  $n$  个随机变量组成的列向量，那么它的协方差矩阵定义为：

$$\Sigma = \mathbb{E}[(\mathbf{X} - \mathbb{E}[\mathbf{X}])(\mathbf{X} - \mathbb{E}[\mathbf{X}])^T]$$

矩阵中的第  $(i, j)$  个元素是  $X_i$  与  $X_j$  的协方差。 $\Sigma$  的对角元为  $X_i$  的方差。 $\Sigma$  是对称和半正定的矩阵。