

# 随机变量及其分布

by zijeff

## 2.1 随机变量

- 随机变量的引入

我们高中就接触过**古典概型**中的**样本空间** $S$ ，我们需要找出样本空间所有的**样本点** $e$ ，进而计算一系列问题的概率。在实际问题中，样本点 $e$ 既有可能是简单的数字，也有可能是文字上的描述。为了更好的进行数学分析，同时在**具体概率分析**的时候我们也并不关心结果的**具体文字描述**，所以基于以上两点原因，我们希望找到一个**对应关系**。这个**对应关系**使得随机试验的每个结果都被赋予一个数值，在样本空间 $S$ 和实数值建立一种对应关系（类似于函数中从定义域到值域的一种**映射**）。综上所述，我们需要引入某个东西来帮助我们分析，这个引入的东西就是**随机变量**了。

- 随机变量的概念

定义：设随机试验的样本空间 $S = \{e\}$ ， $X = X(e)$ 是定义在样本空间 $S$ 上的单值实值函数。称 $X = X(e)$ 为随机变量。

### 💡 Tip

- 随机变量通常用**大写字母**表示或**希腊字母**等表示
- 随机变量所取到的值通常用**小写字母**表示

如定义所言，随机变量是**单值实值函数**，那它与我们之前学的函数又存在什么**区别**呢？

### 📌 Important

- 普通函数是定义在实数轴上的，而随机变量是定义在样本空间上的（样本空间的元素不一定是实数）。
- 随机变量随着试验的结果不同而取不同的值，由于试验的各个结果的出现具有一定的概率，因此随机变量的取值也有一定的概率规律。

## 2.2 离散型随机变量及其分布律

- 离散型随机变量的分布律

设离散型随机变量 $X$ 所有的可能取值为：

$$x_k \quad (k = 1, 2, \dots)$$

$X$ 取各个可能值的概率，即事件 $\{X = x_k\}$ 的概率，为：

$$P(X = x_k) = p_k \quad (k = 1, 2, \dots)$$

称此为离散型随机变量 $X$ 的分布律。

### 📌 Note

显然由**概率的定义**，易得如下性质：

- $p_k \geq 0, k = 1, 2, \dots$
- $\sum_{k=1}^{\infty} p_k = 1$

同时，分布律我们也可以用**表格**来表示。

$X$	$x_1$	$x_2$	$\dots$	$x_n$	$\dots$
$p_k$	$p_1$	$p_2$	$\dots$	$p_n$	$\dots$

表格直观地表示了随机变量取各个值的概率的规律。 $X$ 取各个值各占一些概率，这些概率之和为1。

- 常见离散型随机变量的概率分布

1. 常数随机变量（退化分布）：最简单的随机变量是常数，它在整个样本空间上仅取一个值。

$$X \equiv c$$
$$P(X \equiv c) = 1$$

2. 0-1分布（两点分布，伯努利分布）：

设随机变量 $X$ 只能取到0与1这两个值，其分布为：

$$P(X = k) = p^k(1-p)^{1-k} \quad k = 0, 1 \quad p \in (0, 1)$$

则称随机变量 $X$ 服从以 $p$ 为参数的0-1分布或者两点分布，分布律表格如下：

$X$	0	1
$p_k$	$1-p$	$p$

3. 二项分布（ $n$ 重伯努利试验）：

设试验 $E$ 只有两个可能的结果： $A$ 和 $\bar{A}$ 。那我们称这种试验为伯努利试验，将 $E$ 独立重复进行 $n$ 次，则称这一系列的试验为 $n$ 重伯努利试验。

我们设单次伯努利试验中 $P(A) = p$ ，易得 $P(\bar{A}) = 1-p$ ，其中 $p \in (0, 1)$ ，用随机变量 $X$ 表示 $n$ 重伯努利试验中事件 $A$ 的发生次数，显然 $X$ 的所有可能取值为 $\{1, 2, \dots, n\}$ 。

接下来，我们推导二项分布的分布律。不失一般性地，也就是求当 $X = k$ ， $k \in \{1, 2, \dots, n\}$ 时的概率。由高中学习的组合知识，显然我们只要在 $n$ 次试验中选取 $k$ 次事件 $A$ 发生即可，从而得到如下分布律公式。

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} = C_n^k p^k (1-p)^{n-k}$$

符合上述分布律的分布称为二项分布，记作 $X \sim b(n, p)$ 。易证，当 $n = 1$ 时，即为两点分布。

因此二项分布的分布律表格也就不难画出了，这里省略，读者可以自画。

4. 泊松分布：

设随机变量 $X$ 的所有可能取值为 $\{1, 2, \dots, n\}$ ，而取到各个值得概率为：

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, k = 0, 1, 2, \dots,$$

其中 $\lambda > 0$ 是常数，则称 $X$ 服从以 $\lambda$ 为参数的泊松分布，记作 $X \sim \pi(\lambda)$ 。

**Important**

泊松分布的引入与其分布律的推导过程：

5. 几何分布：

考虑一个随机试验 $E$ ，它只有两个结果，比如成功和失败，设两个结果的概率分别为 $p$ 和 $1-p$ 。现在我们将这个随机试验独立重复进行，直到其成功为止。用 $X$ 表明所需要的试验次数，显然 $X \in N_+$ ，同时易得 $X$ 取每个值的时候的概率为：

$$P(X = k) = p(1-p)^{k-1}, k \in N_+$$

那么，我们称 $X$ 服从于几何分布，其中 $p$ 为参数，记作 $X \sim G(p)$ 。

**Note**

几何分布具有一个比较重要的性质，即无记忆性。

若前 $m$ 次试验均已失败，之后的试验次数 $X$ 仍为几何分布，即如下式子成立：

$$P(X = m + k | X > m) = p(1 - p)^{k-1}$$

## 6. 负二项分布：

给定正整数 $r$ ，用 $X_r$ 表示首次取得 $r$ 次成功的试验次数。显然， $X_r \in \{r, r+1, r+2, \dots\}$ ，并且：

$$P(X_r = k) = \binom{k-1}{r-1} p^r (1-p)^{k-r} = C_{k-1}^{r-1} p^r (1-p)^{k-r}$$

## • 二项分布的泊松逼近

当试验次数 $n$ 很大的时候，计算二项分布的精确概率变得十分麻烦。因此，我们需要寻找近似计算二项分布的方法。这里我们先介绍泊松近似，后面我们将介绍二项分布的正态近似。

**泊松定理：**设 $\lambda > 0$ 是一个常数， $n$ 是任意正整数，设 $np_n = \lambda$ ，则对于任一固定的非负整数 $k$ ，有如下式子成立

$$\lim_{n \rightarrow \infty} \binom{n}{k} p_n^k (1-p_n)^{n-k} = \frac{\lambda^k e^{-\lambda}}{k!}$$

该定理的**证明如下**：由已知条件可知， $p_n = \frac{\lambda}{n}$ ，将其带回等式左边得到

$$\binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left[1 - \left(\frac{\lambda}{n}\right)\right]^{n-k} = \frac{n(n-1) \cdots (n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

继续变形后得到

$$\frac{\lambda^k}{k!} \left[1 \cdot \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right)\right] \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k}$$

值得注意的是，我们始终关心的是上述式子在两个特定整数 $\lambda, k$ 下随着 $n \rightarrow \infty$ 时的**极限值**，又因为上述式子的项数是**有限可列**的，所以我们可以大胆地使用**极限的运算法则**。

直接将上面这一长串式子，拆分成几个部分：

$$\begin{aligned} \lim_{n \rightarrow \infty} \left(1 - \frac{k-1}{n}\right) &= 1 \\ \lim_{n \rightarrow \infty} \left[1 \cdot \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right)\right] &= 1 \\ \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n &= e^{-\lambda} \\ \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^{-k} &= 1 \end{aligned}$$

综上所述，将各个极限值带回待证等式，即可得到原式成立，证毕。

## 2.3 随机变量的分布函数

### • 分布函数的概念

- 概念的引入：对于随机变量 $X$ ，我们不仅要知道 $X$ 取哪些值，还要知道 $X$ 取这些值的概率；而且在实际问题中，更重要的是想知道 $X$ 在任意有限区间 $(a, b)$ 内取值的概率。
- 分布函数的定义：设 $X$ 是一个随机变量， $x$ 是任意实数，函数

$$F(x) = P(X \leq x) \quad (-\infty < x < +\infty)$$

函数 $F(x)$ 称为 $X$ 的分布函数。

现在我们可以思考一个问题，为什么我们要定义分布函数  $F(x) = P(X \leq x)$ ，这背后有什么说法吗？

首先，我们知道分布函数的提出与定义是为了求解  $X$  在任意有限区间  $(a, b)$  内取值的概率。所以，现在不妨考虑求解随机变量  $X$  落在区间  $(x_1, x_2]$  内的概率。显然，求解过程如下：

$$P(x_1 < X \leq x_2) = P(X \leq x_2) - P(X \leq x_1)$$

所以我们可以令  $P(X \leq x_2) = F(x_2)$ ,  $P(X \leq x_1) = F(x_1)$ ，因此也就给出了分布函数的定义。

#### • 分布函数的性质

1.  $F(x)$  是一个不减函数。
2.  $0 \leq F(x) \leq 1$ ，且  $F(x)$  的极限满足如下等式：

$$\lim_{x \rightarrow -\infty} F(x) = 0$$

$$\lim_{x \rightarrow \infty} F(x) = 1$$

3.  $F(x+0) = F(x)$ ，即  $F(x)$  是右连续的。

#### 📌 Important

由分布函数得一些计算上的重要公式：

1.  $P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a)$
2.  $P(X > x) = P(X < \infty) - P(X \leq x) = 1 - F(x)$
3.  $P(X = x) = F(x) - F(x-0)$
4.  $P(X < x) = F(x-0)$
5.  $P(a \leq X \leq b) = F(b) - F(a-0)$
6.  $P(a \leq X < b) = F(b-0) - F(a-0)$
7.  $P(a < X < b) = F(b-0) - F(a)$

一般地，设离散型随机变量  $X$  的分布律：

$$P(X = x_k) = p_k \quad (k = 1, 2, \dots)$$

由概率的可列可加性可得  $X$  的分布函数为：

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} P(X = x_k)$$

即，化简后分布函数为：

$$F(x) = \sum_{x_i \leq x} p_k$$

这里和式是对所有满足  $x_k \leq x$  的  $k$  求和的。所以分布函数  $F(x)$  在  $x = x_k$  ( $k = 1, 2, \dots$ ) 处有跳跃，其跳跃值为  $p_k$ 。

#### ④ Note

不同的随机变量，它们的分布函数可能相同。

## 2.4 连续型随机变量及其概率密度

连续型随机变量  $X$  所有可能取值充满一个区间，对这种类型的随机变量，不能像离散型随机变量那样，以指定它取每个值概率的方式，去给出其概率分布，而是通过给出概率密度函数的方式进行计算。

#### • 概率密度的概念与性质

1. 概率密度函数的定义：如果对于随机变量 $X$ 的分布函数 $F(x)$ ，存在非负可积函数 $f(x)$ ，使对于任意实数 $x$ 有下式成立：

$$F(x) = \int_{-\infty}^x f(t)dt$$

则称 $X$ 为连续型随机变量，其中函数 $f(x)$ 称为 $X$ 的**概率密度函数**，简称**概率密度**。

**Note**

连续型随机变量的分布函数是**连续函数**。

2. 概率密度函数的性质：

- $f(x) \geq 0$
- $\int_{-\infty}^{\infty} f(x)dx = 1$ ，也就意味着**概率密度函数**与坐标横轴围成的**面积为定值1**
- 对于任意的实数 $x_1, x_2, (x_1 < x_2)$ ， $P(x_1 < X \leq x_2) = \int_{x_1}^{x_2} f(x)dx$
- 若 $f(x)$ 在点 $x$ 处连续，则有 $F'(x) = f(x)$ 成立

**Note**

第一条和第二条性质是判定一个函数 $f(x)$ 是否为某个**连续型随机变量 $X$** 的概率密度函数的**充要条件**。

3. 利用上述性质得以下计算公式：

- $P(X \leq a) = F(a) = \int_{-\infty}^a f(x)dx$
- $P(X > a) = 1 - P(X \leq a) = 1 - F(a) = \int_{-\infty}^{\infty} f(x)dx - \int_{-\infty}^a f(x)dx = \int_a^{\infty} f(x)dx$

**Important**

**连续型随机变量**取任一指定实数值 $a$ 的概率均为0。即 $P(X = a) = 0$

证明如下：

从随机变量本身性质出发，我们可以得到如下不等关系：

$$0 \leq P(X = a) \leq P(a - \delta < X \leq a) = F(a) - F(a - \delta)$$

当 $\delta \rightarrow 0^+$ 时，得到 $P(X = a) = 0$ ，因此我们可以得到如下两条重要性质：

- 由 $P(A) = 0$ ，我们不能推出 $A = \emptyset$
- 对于连续型随机变量而言， $P(a < X < b) = P(a \leq X < b) = P(a < X \leq b) = P(a \leq X \leq b)$

也就是说：**连续型随机变量取值落在某区间的概率与端点无关**。

• **常见连续型随机变量及其概率分布**

1. 均匀分布：若连续型随机变量 $X$ 具有概率密度

$$f(x) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & \text{其他} \end{cases}$$

则称 $X$ 在 $(a, b)$ 上服从均匀分布，记为 $X \sim U(a, b)$

均匀分布的意义在于：在区间 $(a, b)$ 上服从均匀分布的随机变量 $X$ 落在区间 $(a, b)$ 中任意等长度的子区间内的可能性是相同的。

对概率密度函数积分后可以得到**分布函数**，如下：

$$F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x < b \\ 1 & x \geq b \end{cases}$$

2. 指数分布：若连续型随机变量 $X$ 具有概率密度

$$f(x) = \begin{cases} \frac{1}{\theta} e^{-\frac{x}{\theta}} & x > 0 \\ 0 & \text{其他} \end{cases}$$

其中,  $\theta > 0$  是常数, 则称  $X$  服从参数为  $\theta$  的指数分布。易知  $f(x) \geq 0$ , 且  $\int_{-\infty}^{\infty} f(x) dx = 1$  对概率密度函数积分后可以得到分布函数, 如下:

$$F(x) = \begin{cases} 1 - e^{-\frac{x}{\theta}} & x > 0 \\ 0 & \text{其他} \end{cases}$$

### Important

指数分布也具有无记忆性的重要性质。

也就是说, 对于任意的  $s, t > 0$ , 有下式成立:

$$P(X > s + t \mid X > s) = P(X > t)$$

3. 正态分布: 若连续型随机变量  $X$  具有概率密度

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

其中  $\mu, \sigma (\sigma > 0)$  为常数, 则称  $X$  服从参数为  $\mu, \sigma$  的正态分布或高斯分布, 记为  $X \sim N(\mu, \sigma^2)$ 。

正态分布具有如下重要性质:

- 曲线关于  $\mu = x$  对称, 这表明对于任意的  $h > 0$ , 有  $P(\mu - h < X \leq \mu) = P(\mu < X \leq \mu + h)$
- 当  $x = \mu$  时取到最(极)大值,  $f(\mu) = \frac{1}{\sqrt{2\pi}\sigma}$
- 在  $x = \mu \pm \sigma$  处曲线有拐点
- 曲线以  $x$  轴为渐近线
- $\mu$  称为正态分布概率密度曲线的位置参数, 意味着曲线的位置将完全由参数  $\mu$  所决定。
- 改变  $\sigma$  时, 曲线对称轴不改变, 而形状在改变。 $\sigma$  越小, 图形越高越瘦,  $\sigma$  越大, 图形越矮越胖。

## 2.5 随机变量的函数分布

从实际出发, 生活中存在各式各样的随机变量, 而不同的随机变量之间往往不是独立存在的, 之间存在着某种关系