

第 2 讲*

随机变量，概率质量/密度函数，以及累积分布函数(单变量模型)

麻省理工学院 14.30 2006 年春季

Herman Bennett

3 随机变量

3.1 直观定义

一个**随机变量**是一个未知具体数值的变量，它的取值可能取到或代表样本空间中任何可能的元素。样本空间中的元素有一定的概率分布(概率函数)。因此，我们说**随机变量**是以一定的概率取到不同的可能值。

存在两种随机变量：(1) 离散型随机变量，即 X 只能取到几个有限 (或者有限可列) 数值，(2) 连续型随机变量，即 X 可以在有界或无界区间内取到所有的实数值。

3.2 数学定义

一个**随机变量** $X(s)$ 是从样本空间 S (域) 到实数域的函数，其中 $s \in S$ 。

为了说明随机变量，必须对样本空间进行重新界定。例如，在抛三次硬币的试验中，我们可以把“出现正面的总次数”定义为一个随机变量。

注意：这些讲义不一定是自封的。它们只是对讲座的一种补充而不是替代。

例 3.1. 试验：抛掷均匀的硬币 3 次，

	s	HHH	HHT	HTH	THH	TTH	THT	HTT	TTT
X(s)	“结果”								
X(s)	“正面 H 的次数”								

说明：随机变量通常用大写字母来表示，而变量的实际数值（或者它的范围）用相应的小写字母来表示。因此，随机变量 X 的取值是 x 。

- 表格：

在样本空间上定义的概率函数决定随机变量的（概率）分布。累积分布函数或概率密度函数可以完全反映离散（连续）随机变量的分布特征（病态的情况除外）。

4 离散模型

4.1 概率质量函数

离散型随机变量的概率质量函数 (pmf) 记作 $f_X(x)$, 指

$$f_X(x) = P(X = x), \text{ 对于任意 } x; \quad (4)$$

对于任意 x , 满足如下性质:

$$(1) \quad f(x) \geq 0$$

$$(2) \quad \sum_{i=1}^{\infty} f(x_i) = 1$$

注意: 如果有必要强调 $f(x)$ 是随机变量 X 的概率质量函数, 那么就用 $f_X(x)$ 表示。

如果从上下文中已经明确它是概率质量函数, 那么下标数字是可以省略的: 表示为 $f(x)$ 。

这种方法同样适用于我们后面将遇到的函数。

例 4.1. 接例 3.1, 请列出随机变量“出现正面 H 的次数”的所有可能数值, 写出它的概率质量函数, 验证它的性质, 并画出它的图形。

例 4.2 均匀分布：对于 $x=1,2,\dots,k$, $f(x)=1/k$, 否则为 0。画出 $f(x)$ 的图形。
设 $k=10$, 计算 $P(X=3)$ 和 $P(2 \leq X)$ 。

例 4.3. 二项分布。进行 n 个独立相同试验, 每个试验只有两种可能的结果: “成功”或“失败”, 被称作二项试验。描述连续试验中“成功”次数的随机变量属于二项分布, 性质如下:

$$f(x) = P(X = x) = p^x(1-p)^{n-x} \binom{n}{x} \quad \text{for } x = 0, 1, \dots, n \quad (5)$$

$$x \sim b(n, p)$$

其中 n 为试验次数, p 表示每次试验成功的概率。如果 $n=1$, 此分布叫做贝努利分布。
应用: 掷一枚骰子, 随机变量 X 为 4 次投掷过程中获 6 点的次数, 计算两次 (仅有两次) 为 6 点的概率。

4.2 累积分布函数

离散型随机变量的累积分布函数记作 $F_X(x)$ ，定义为：对于任意 x ，

$$F_X(x) = P(X \leq x) = \sum_{i=1}^x f_X(i), \quad \text{对于任意 } x; \quad (6)$$

满足如下性质：

$$(1) \lim_{x \rightarrow 0} F(x) = 0 \text{ 且 } \lim_{x \rightarrow \infty} F(x) = 1$$

(2) $F(x)$ 是 x 的单调不递减函数

(3) $F(x)$ 右连续，且对于任意 x_0 , $\lim_{x|x_0 \rightarrow \infty} F(x) = F(x_0)$

- $P(X > x) = 1 - P(X \leq x) = 1 - F(x)$

- 若 $x_1 < x_2$, 则 $P(x_1 < X < x_2) = F(x_1) - F(x_2)$

- 若 $x_1 \sim F_X(x)$, $Y \sim F_Y(y)$, 则对于所有 x , $X \sim Y \leftrightarrow F_X(x) = F_Y(y)$

- 为什么是右连续？(参考例 4.4)

记住：在**本节(4.2)**中关于离散型随机变量的累积分布函数同样适用于连续型随机变量（后面将会详细介绍）。

例 4.4. 接例 4.1，写出累积分布函数，画出它的图形，验证它的性质。

5 连续模型

5.1 概率密度函数

连续型随机变量的概率密度函数 (pdf) 记为 $f_X(x)$, 定义为

$$\int_{x_1}^{x_2} f_X(x) dx = P(x_1 \leq X \leq x_2), \quad \text{对于任意 } x_1 \text{ 和 } x_2; \quad (7)$$

它满足如下性质:

(1) 对于任意 x , $f(x) \geq 0$;

$$(2) \quad \int_{-\infty}^{\infty} f(x) dx = 1$$

- 如果 X 是连续型随机变量, 那么 $P(X=x_0)=?$
- $P(a < X < b) = P(a \leq X < b) = P(a < X \leq b) = P(a \leq X \leq b)$
- $\frac{P(x_0 - \varepsilon < X < x_0 + \varepsilon)}{P(x_1 - \varepsilon < X < x_1 + \varepsilon)} \rightarrow \frac{f_X(x_0)}{f_X(x_1)}, \text{ 当 } \varepsilon \rightarrow 0$

5.2 累积分布函数

正如在 4.2 节中提到, 离散型随机变量的有关累积分布函数的所有定义和结论同样也适用于连续型随机变量的累积分布函数。连续型随机变量的累积分布函数具体公式为:

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f_X(u) du \quad \text{对于任意 } x; \quad (8)$$

- 混合分布
- 假定 F' 存在, 且 X 是一个连续随机变量, 则 $\frac{\partial F_X(x)}{\partial x} = f_X(x)$

例 5.1. 均匀分布: $f(x) = \frac{1}{b-a}, \rightarrow a \leq x \leq b$, 否则为 0。画出概率密度函数的图形, 并计算累积分布函数。

例 5.2. X 是一个连续型随机变量, 它的概率密度函数为

$$f(x) = \begin{cases} ax^2, & 0 < x < 3 \\ 0 & \text{否则} \end{cases}$$

计算 a 和 $P(1 < x < 2)$ 的值。

例 5.3. (作业) X 是一个连续型随机变量, 它的累积分布函数为:

$$F(x) = \frac{1}{1 + e^{-x}} \quad \text{对于任意 } x \in \mathbb{R}.$$

计算 $f(x)$ 并验证它的性质。分别运用概率密度函数和累积分布函数计算

$P(-1 < x < 2)$, 并验证两种方法获得的结果相同。