第一章随机变量及其信息度量

第一节离散随机变量及其分布

陈兴同

中国矿业大学 数学学院

2021 年 8 月

内容提要

内容提要

内容提要

内容提要

内容提要

什么是分布律

离散型随机变量是指取值空间为可数离散集合（无限集合或有限集合）的随机变量。设离散型随机变量 X 取值空间为 $\mathcal{X}=\{x_1,x_2,x_3,\cdots\}$，它的**分布律**是指每个取值的概率

$$p_i=p(x_i)=P\{X=x_i\},i=1,2,\cdots, \tag{1.1}$$

并且符合要求

$$p_i\geq 0,i=1,2,\cdots,\sum_i p_i=1. \tag{1.2}$$

一维分布矩阵

如果 $\mathcal{X}$ 是一个有限集 $\mathcal{X} = \{x_1, x_2, x_3, \cdots, x_N\}$，则它的分布律也可以用两行表格或矩阵来表示，第一行表示取值，第二行表示取每个值的概率。

$$\begin{array}{c|cccc} X & x_1 & x_2 & \cdots & x_N \\ \hline p_i & p_1 & p_2 & \cdots & p_N \end{array} \text{或} X \sim \begin{pmatrix} x_1 & x_2 & \cdots & x_N \\ p_1 & p_2 & \cdots & p_N \end{pmatrix}. \quad (1.3)$$

另外，分布律（1.1）也定义了一元离散函数

$$p(x) = P\{X = x\}, x \in \mathcal{X},$$

它类似于连续型随机变量概率密度函数，不仿称为离散型随机变量 X 的**概率函数**，记作：

$$X \sim p(x), x \in \mathcal{X}. \quad (1.4)$$

一维分布函数

多个随机变量的概率函数将用下标来区分，比如 $p_X(x), p_Y(x)$。描述离散随机变量分布情况也可以用**分布函数**，它的定义为

$$F(x) = P\{X \leq x\} = \sum_{x_i \leq x} p_i, x \in \mathbb{R}, \tag{1.5}$$

其实就是离散随机变量 X 取值不大于 x 的概率，它是一元阶梯形函数，分断点 x_i 处的跃度就是概率 p_i。
最后，如果行向量 $(p_1, p_2, \cdots)$ 满足条件（1.2），就称它是一个概率分布向量。

练习

试求离散型随机变量 X 的分布函数并画图。

$$X \sim \begin{pmatrix} 0 & 1 & 2 & 3 \\ 3/8 & 1/4 & 1/4 & 1/8 \end{pmatrix}.$$

什么是联合分布

设离散型随机变量 X, Y 的取值空间分别为 $\mathcal{X} = \{x_1, x_2, x_3, \cdots\}$，$\mathcal{Y} = \{y_1, y_2, y_3, \cdots\}$，它们的**联合分布律**是

$$p_{ij} = p(x_i, y_j) = P\{X = x_i, Y = y_j\}, x_i \in \mathcal{X}, y_j \in \mathcal{Y}. \qquad \overset{(1.6)}{\cancel{(2.1)}}$$

联合分布矩阵

如果 X, Y 的取值空间 $\mathcal{X}, \mathcal{Y}$ 是有限集，则联合分布律也可以用二维表或矩阵

$X\backslash Y$	y_1	y_2	$\cdots$	y_M
x_1	p_{11}	p_{12}	$\cdots$	p_{1M}
x_2	p_{21}	p_{22}	$\cdots$	p_{2M}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
x_N	p_{N1}	p_{N2}	$\cdots$	p_{NM}

或 $(X, Y) \sim \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1M} \\ p_{21} & p_{22} & \cdots & p_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ p_{N1} & p_{N2} & \cdots & p_{NM} \end{pmatrix}$

~~(2.2)~~
(1.7)

表示。

联合概率函数

另外，分布律（2.1）也定义了一个二元离散函数

$$p(x,y)=P\{X=x,Y=y\},(x,y)\in\mathcal{X}\times\mathcal{Y},$$

不仿称为随机向量 (X,Y) 的**联合概率函数**，记作

$$(X,Y)\sim p(x,y),x\in\mathcal{X},y\in\mathcal{Y}. \qquad (2.3)$$

当有多个联合分布时，用下标来区分，比如：$p_{X,Y}(x,y),p_{U,V}(x,y)$。

二维或联合分布函数

描述联合分布也可以用二元分布函数

$$F(x,y)=P\{X\le x, Y\le y\}=\sum_{x_i\le x, y_j\le y} p_{ij}, (x,y)\in\mathbb{R}^2,$$

它是一个二元阶梯形函数，每个阶梯处的跃度就是概率 p_{ij}。

n 维分布

对于 n 维离散随机向量 $(X_1, X_2, \cdots, X_n)$ 的联合分布，也可以用 n 维联合分布律或 n 元分布函数来描述。

$$p(x_1, x_2, \cdots, x_n) = P\left\{X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n\right\}, x_i \in \mathcal{X}_i, i = 1$$

$$F(x_1, x_2, \cdots, x_n) = P\left\{X_1 \le x_1, X_2 \le x_2, \cdots, X_n \le x_n\right\}, x_i \in \mathbb{R}, i = 1$$

其中 $\mathcal{X}_i$ 是第 i 个随机变量的取值空间，当然对 n 个有限离散随机变量的联合分布律也可以用 n 维数组来描述。

什么是边缘分布

设两个离散型随机变量 X, Y 的联合分布律为（2.3），则可以求出 X, Y 各自的分布律 $p_X(x), p_Y(y)$，称为**边缘分布律**，它们的求法如下：

$$p_X(x) = \sum_y p(x,y), \quad p_Y(y) = \sum_x p(x,y). \tag{3.1}$$

多维边缘分布

对于 n 维离散随机向量 $(X_1, X_2, \cdots, X_n)$，它的每一个分量 X_i，任意二个分量 X_i, X_j，任意三个分量 X_i, X_j, X_k，任意 $n-1$ 个分量 $X_{i_1}, X_{i_2}, \cdots, X_{i_{n-1}}$ 也分别是随机变量或随机向量，也都有确定的分布律，称为这个 n 维随机向量的边缘分布。它们可以从 n 维联合分布律获得：

$$\begin{aligned}
p_{X_i}(x_i) &= \sum_{x_j \in \mathcal{X}_j, j \neq i} p(x_1, x_2, \cdots, x_n), \\
p_{X_i, X_j}(x_i, x_j) &= \sum_{\substack{x_l \in \mathcal{X}_l, l \neq i \\ x_m \in \mathcal{X}_m, m \neq j}} p(x_1, x_2, \cdots, x_n), \\
F_{X_1}(x_1) &= F(x_1, +\infty, \cdots, +\infty), \\
F_{X_i, X_j}(x_i, x_j) &= F(+\infty, \cdots, +\infty, x_i, +\infty, \cdots, +\infty, x_j, +\infty, \cdots, +\infty
\end{aligned}$$

练习

（1） 任意多个随机变量的边缘分布函数怎样写出？

（2） 如果已知边缘分布，能否求得联合分布？这可以用二维正态分布的边缘分布来说明

什么是条件分布

若两个离散型随机变量 X, Y 有联合分布律（2.3），并且边缘分布 $p_X(x) > 0$，则可以定义条件概率

$$p(y|x) = P\{Y = y|X = x\} = \frac{P\{X = x, Y = y\}}{P\{X = x\}} = \frac{p(x,y)}{p_X(x)},$$

如果取 $X = x_i, Y = y_j$，并将条件概率记为 $p(y_j|x_i)$ 或$p_{j|i}$，则每给随机变量 X 的一个取值 x_i 都能得到一个关于 Y 的条件分布律

$$p(y_1|x_i), p(y_2|x_i), p(y_3|x_i), \cdots . \qquad (\cancel{4.1})\ 1.10$$

条件分布矩阵

如果 X, Y 的取值空间 $\mathcal{X}, \mathcal{Y}$ 是有限集，就可以将每个条件分布律作为行构成条件分布矩阵。

4.1

$$
\begin{pmatrix}
p(y_1|x_1) & p(y_2|x_1) & \cdots & p(y_M|x_1) \\
p(y_1|x_2) & p(y_2|x_2) & \cdots & p(y_M|x_2) \\
\vdots & \vdots & & \vdots \\
p(y_1|x_N) & p(y_2|x_N) & \cdots & p(y_M|x_N)
\end{pmatrix} \tag{$\cancel{4.2}$}
$$

续：条件分布矩阵

如果 X, Y 的取值空间 $\mathcal{X}, \mathcal{Y}$ 是有限集，就可以将每个条件分布律作为行构成矩阵

1.11

$$\begin{pmatrix} p_{1|1} & p_{2|1} & \cdots & p_{M|1} \\ p_{1|2} & p_{2|2} & \cdots & p_{M|2} \\ \vdots & \vdots & \ddots & \vdots \\ p_{1|N} & p_{2|N} & \cdots & p_{M|N} \end{pmatrix} \tag{$\cancel{4.3}$}$$

称为**条件分布矩阵**，它可以作为信道传输特性的概率模型，同时也给出一个在集合 $\mathcal{X} \times \mathcal{Y}$ 上的二元离散函数 $p(y|x)$，称为**条件概率函数**。

多维条件分布

也可以定义用随机向量作为条件的条件概率

$$
\begin{aligned}
p(y_j|x_1x_2\cdots x_n) &= P\{Y=y_j|X_1=x_1,X_2=x_2,\cdots,X_n=x_n\} \\
&= \frac{P\{X_1=x_1,\cdots,X_n=x_n,Y=y_j\}}{P\{X_1=x_1,\cdots,X_n=x_n\}} \\
&= \frac{p(x_1,x_2,\cdots,x_n,y_j)}{p(x_1,x_2,\cdots,x_n)},
\end{aligned}
$$

其中 $y_j\in\mathcal{Y}, x_i\in\mathcal{X}_i, i=1,2,\cdots,n$。

更一般条件概率

另外还可以定义更一般的条件概率，比如：

$$
\begin{aligned}
& p(y_1 y_2 \cdots y_m | x_1 x_2 \cdots x_n) \\
= \; & P\{Y_1 = y_1, \cdots, Y_m = y_m | X_1 = x_1, \cdots, X_n = x_n\} \\
= \; & \frac{P\{Y_1 = y_1, \cdots, Y_m = y_m, X_1 = x_1, \cdots, X_n = x_n\}}{P\{X_1 = x_1, \cdots, X_n = x_n\}} \\
= \; & \frac{p(x_1, \cdots, x_n, y_1, \cdots, y_m)}{p(x_1, x_2, \cdots, x_n)},
\end{aligned}
$$

n　m

其中 $y_j \in \mathcal{Y}_j, x_i \in \mathcal{X}_i, i = 1, 2, \cdots, N, j = 1, 2, \cdots, M$。

什么是数学期望

离散型随机变量的数学期望定义为所取值的概率平均值。这里同时给出随机变量函数 $Y = f(X)$ 及 $Z = f(X_1, X_2, \cdots, X_n)$ 的数学期望公式。

$$
\begin{aligned}
E(X) &= \sum_x xp(x), \\
E(Y) &= E\left[f(X)\right] = \sum_x f(x)p(x), \\
E(Z) &= E\left[f(X_1, X_2, \cdots, X_n)\right], \\
&= \sum_{x_1 \in \mathcal{X}_1} \cdots \sum_{x_n \in \mathcal{X}_n} f(x_1, \cdots, x_n)p(x_1, \cdots, x_n).
\end{aligned}
$$

条件数学期望

条件数学期望是指条件分布的数学期望。在事件 $\{X = x_i\}$ 发生的条件下随机变量 Y 具有条件分布（4.1），当然可以求数学期望

$$E(Y|x_i) = E(Y|X = x_i) = \sum_{y \in \mathcal{Y}} yp(y|x_i), x_i \in \mathcal{X}.$$

由于条件分布矩阵（??）每一行都是一个条件分布，故每一行都可以求数学期望，可以求到一系列条件期望值 $E(Y|x_1), E(Y|x_2), \cdots$，因此条件数学期望 $E(Y|x_i)$ 可以看作是在取值空间 $\mathcal{X}$ 上的函数 $E(Y|x), x \in \mathcal{X}$。
类似地可以定义其它条件数学期望，比如在更多条件下的条件数学期望：

$$\begin{aligned} E(Y|x_1 x_2 \cdots x_n) &= E(Y|X_1 = x_1, \cdots, X_n = x_n), \\ &= \sum_{y \in \mathcal{Y}} yp(y|x_1 x_2 \cdots x_n) \end{aligned}$$

练习

证明：

$$E(Y) = \sum_x p_X(x) E(Y|x)$$