# 第1章

# 随机变量基础

概率论与随机变量是随机信号分析与处理的理论基础，本章简要介绍随机变量的基本理论，更为详细的内容请大家参考有关教材。

## 1.1 概率论的基本术语

### 1. 随机试验

满足下列3个条件的试验称为随机试验：

（1）在相同条件下可重复进行；

（2）试验的结果不止一个，所有可能的结果能事先明确；

（3）每次试验前不能确定会出现哪一个结果。

随机试验通常用$E$表示，比如投掷硬币，就是一个随机试验，它满足以上3个条件。首先，投掷硬币是可以重复进行的；其次试验的结果可能是正面，也可能是反面，即有两种可能的结果，而且只有这两种结果，事先可以明确，但具体到某次试验，试验前是不能预知出现哪种结果的。

### 2. 随机事件

在随机试验中，对试验中可能出现也可能不出现、而在大量重复试验中却具有某种规律性的事情，称为随机事件，简称为事件，如投掷硬币出现正面就是一个随机事件。

### 3. 基本事件

随机试验中最简单的随机事件称为基本事件，如投掷骰子出现1，2，…，6点是基本事件，出现偶数点是随机事件，但不是基本事件。

#### 4. 样本空间

随机试验 $E$ 的所有基本事件组成的集合称为样本空间，记为 $S$，如投掷骰子的样本空间为$\{1,2,3,4,5,6\}$。

#### 5. 频数和频率

在相同条件下的 $n$ 次重复试验中，事件 $A$ 发生的次数 $n_A$ 称为事件 $A$ 的频数，比值$\frac{n_A}{n}$称为事件 $A$ 发生的频率。频率反映了事件 $A$ 发生的频繁程度，若事件 $A$ 发生的可能性大，那么相应的频率也大，反之则较小。

#### 6. 概率

概率是事件发生的可能性大小的度量。事件的频率可以刻画事件发生的可能性大小，但是频率具有随机波动性，对于相同的试验次数 $n$，事件 $A$ 发生的频率可能不同，$n$ 越小，这种波动越大，$n$ 越大，波动越小，当 $n$ 趋于无穷时，频率趋于一个稳定的值，可以把这个稳定的值定义为事件 $A$ 发生的概率，记为 $P(A)$，即

$$P(A)=\lim_{n\to\infty}\frac{n_A}{n} \tag{1.1.1}$$

这一定义称为概率的统计定义。概率的统计定义不仅提供了事件 $A$ 发生的可能性大小的度量方法，而且还提供了估计概率的方法，只要重复试验的次数 $n$ 足够大，就可以用下式来估计概率：

$$\hat{P}(A)=\frac{n_A}{n} \tag{1.1.2}$$

概率还有多种定义方式，如古典概型的古典定义，几何概型的几何定义，以及更一般的概率的公理化定义，这些定义大家可以参阅概率论的有关书籍。

## 1.2 随机变量的定义

在随机试验中，试验的结果不止一个，如投掷骰子可能出现的点数，打靶命中的环数及一批产品中的次品数等。另一些随机试验尽管其可能结果与数值间没有直接的联系，如投掷硬币出现正面或反面、雷达探测发现“有目标”或“无目标”等，但可以规定一些数值来表示试验的可能结果。如对于投掷硬币，用“1”表示“正面”，“0”表示“反面”，对雷达探测用“1”表示“有目标”，“0”表示“无目标”。为了表示这些试验的结果，我们定义一个变量，变量的取值反映试验的各种可能结果，由于试验前无法确知试验结果，所以变量的值在试验前是无法确知的，即变量的值具有随机性，称这个变量为随机变量。下面给出详细的定义。

**定义** 设随机试验 $E$ 的样本空间为 $S=\{e\}$，如果对于每一个 $e\in S$，有一个实数 $X(e)$与之对应，这样就得到一个定义在 $S$ 上的单值函数 $X(e)$，称 $X(e)$为随机变量，简记为 $X$。

从以上的定义可以看出，随机变量是定义在样本空间 $S$ 上的一个单值函数。对应于不同的样本 $e$，$X(e)$的取值不同，$X(e)$的随机性在样本 $e$ 中体现出来，因为在试验前究竟出现哪个样本事先无法确知，只有试验后才知道。$X$ 的取值可以是连续的，也可以是离散的，所以，根据 $X$ 取值的不同，可以分为连续型随机变量和离散型随机变量。

所谓离散型随机变量是指它的全部可能取值为有限个或可列无穷个。离散型随机变量的概率特性通常用概率分布律来描述。

设离散型随机变量 $X$ 的所有可能取值为 $x_k(k=1,2,\cdots,n)$，其概率为

$$P(X=x_k)=p_k\quad(k=1,2,\cdots,n)\tag{1.2.1}$$

称上式为 $X$ 的概率分布或分布律。通常如表 1.1 所示。

**表 1.1　$X$ 的概率分布**

| $X$ | $x_1$ | $x_2$ | $\cdots$ | $x_n$ |
|---|---|---|---|---|
| $p_k$ | $p_1$ | $p_2$ | $\cdots$ | $p_n$ |

其中

$$\sum_{k=1}^{n}p_k=1\tag{1.2.2}$$

下面介绍几种典型的离散随机变量的概率分布。

### 1.（0,1）分布

设随机变量 $X$ 的可能取值为 0 和 1 两个值，其概率分布为

$$P(X=1)=p,\quad P(X=0)=1-p\quad(0<p<1)\tag{1.2.3}$$

称 $X$ 服从(0,1)分布。如投掷硬币的试验，假定出现正面用 1 表示，出现反面用 0 表示，用 $X$ 表示试验结果，那么 $X$ 的可能取值为 0、1，$X$ 是一个离散型随机变量，且服从(0,1)分布，

$$P(X=1)=P(X=0)=0.5$$

### 2. 二项式分布

设随机试验 $E$ 只有两种可能的结果 $A$ 及 $\overline{A}$，且 $P(A)=p$，$P(\overline{A})=1-p=q$，将 $E$ 独立地重复 $n$ 次，这样的试验称为贝努里(Bernoulli)试验，那么在 $n$ 次试验中事件 $A$ 发生 $m$ 次的概率为

$$P_n(m)=C_n^m p^m q^{n-m}\quad(0\leqslant m\leqslant n)\tag{1.2.4}$$

(1.2.4)式刚好是 $(p+q)^n$ 展开式的第 $m+1$ 项，故称为二项式分布。

### 3. 泊松(Poisson)分布

设随机变量 $X$ 的可能取值为 $0,1,2,\cdots$，且概率分布为

$$P(X=k)=\frac{\lambda^k \mathrm{e}^{-\lambda}}{k!},\quad k=0,1,\cdots,\quad \lambda>0\tag{1.2.5}$$

则称 $X$ 服从参数为 $\lambda$ 的泊松分布。

## 1.3 随机变量的分布函数与概率密度

设 $X$ 为随机变量，$x$ 为任意实数，定义

$$F(x)=P(X\leqslant x) \tag{1.3.1}$$

为 $X$ 的概率分布函数或简称为分布函数。

分布函数具有如下性质：

(1) 它是 $x$ 的不减函数，即

$$F(x_2)-F(x_1)\geqslant 0 \quad x_2>x_1 \tag{1.3.2}$$

(2) $0\leqslant F(x)\leqslant 1$。 (1.3.3)

(3) $F(-\infty)=0$，　$F(\infty)=1$。 (1.3.4)

(4) 若 $F(x_0)=0$，则对任何 $x<x_0$，有 $F(x_0)=0$。

(5) $P(X>x)=1-F(x)$。 (1.3.5)

(6) 函数 $F(x)$是右连续的，即

$$F(x^+)=F(x) \tag{1.3.6}$$

(7) 对于任意实数 $x_1$，$x_2(x_1<x_2)$，有

$$P(x_1<X\leqslant x_2)=F(x_2)-F(x_1) \tag{1.3.7}$$

(8) $P(X=x)=F(x)-F(x^-)$。 (1.3.8)

(9) $P(x_1\leqslant X\leqslant x_2)=F(x_2)-F(x_1^-)$。 (1.3.9)

对于连续型随机变量，其分布函数是连续的，在这种情况下，$F(x)=F(x^-)$，所以对于任意 $x$ 都有

$$P(X=x)=0 \tag{1.3.10}$$

对离散型随机变量，分布函数是阶梯型的。设 $x_i$ 表示 $F(x)$的不连续点，则

$$F(x_i)-F(x_i^-)=P(X=x_i)=p_i \tag{1.3.11}$$

这时 $X$ 的统计特性由它的取值 $x_i$ 及取值的概率 $p_i$ 确定，也即由概率分布律确定。分布函数可表示为

$$F(x)=\sum_i p_i U(x-x_i) \tag{1.3.12}$$

其中 $p_i=P(X=x_i)$，$U(\cdot)$为单位阶跃函数。由(1.3.12)式可以看出，离散型随机变量的分布函数是阶梯型函数，阶梯的跳变点位于随机变量的取值点，跳变的高度等于随机变量取该值的概率。

比如(0,1)分布的随机变量 $X$(见图 1.1)，其分布函数为

$$F(x)=P(X\leqslant x)=\begin{cases}0 & x<0\\ 1/2 & 0\leqslant x<1\\ 1 & x\geqslant 1\end{cases}$$

如果 $X$ 的概率分布既不是连续的，也不是离散的，那么称 $X$ 为混合型随机变量。

随机变量 $X$ 的分布函数的导数定义为它的概率分布密度，简称为概率密度或分布密度，记为 $f(x)$，即

$$f(x)=\frac{\mathrm{d}F(x)}{\mathrm{d}x} \tag{1.3.13}$$

由概率密度的定义及分布函数的性质,可以得出概率密度的如下性质:

(1) $f(x)\geqslant 0$,也即概率密度是非负的函数。

(2) $\int_{-\infty}^{+\infty}f(x)\mathrm{d}x=1$,即概率密度函数与横轴 $x$ 所围成的面积为1。

(3) $P(x_1<X\leqslant x_2)=F(x_2)-F(x_1)=\int_{x_1}^{x_2}f(x)\mathrm{d}x$,这说明随机变量 $X$ 落在区间 $(x_1,x_2]$ 上的概率等于图1.2中阴影区的面积。从这条性质也可以看出,对于连续型随机变量,有 $P(X=x)=0$。

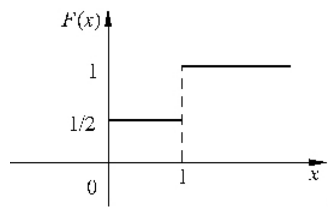


图 1.1 (0,1)分布的分布函数

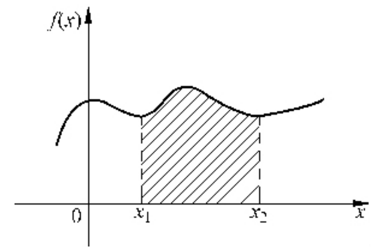


图 1.2 随机变量 X 落在区间$(x_1,x_2)$上的概率

对于离散型随机变量,由于它的概率分布函数是阶梯型,那么它的概率密度函数是一串 $\delta$ 函数之和,$\delta$ 函数出现在随机变量的取值点,强度为取该值的概率,即

$$f(x)=\sum_i p_i\delta(x-x_i) \tag{1.3.14}$$

其中 $x_i$ 为离散型随机变量 $X$ 的取值,$p_i=P(X=x_i)$。

下面介绍常见的连续型随机变量分布。

### 1. 正态分布

若随机变量 $X$ 的概率密度为

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x-m)^2}{2\sigma^2}\right] \tag{1.3.15}$$

其中 $m$、$\sigma$ 为常数,则称 $X$ 服从正态分布,正态分布通常也简记为 $N(m,\sigma^2)$。均值为0,方差为1的正态分布 $N(0,1)$称为标准正态分布。正态分布随机变量的概率密度是一个高斯曲线,所以又称为高斯随机变量,概率密度曲线如图1.3(a)所示。

正态分布函数为

$$F(x)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x-m)^2}{2\sigma^2}\right]\mathrm{d}x \tag{1.3.16}$$

标准正态分布函数通常用 $\Phi(x)$表示,即

$$\Phi(x)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)\mathrm{d}x \tag{1.3.17}$$

### 2. 均匀分布

如果随机变量 $X$ 的概率密度函数为

$$f(x)=\begin{cases}\dfrac{1}{b-a} & a<x<b\\ 0 & \text{其他}\end{cases} \tag{1.3.18}$$

则称 $X$ 在区间 $(a,b)$ 上服从均匀分布，概率密度曲线如图 1.3(b)所示。

### 3. 瑞利分布

如果随机变量 $X$ 的概率密度函数为

$$f(x)=\frac{x}{\sigma^2}\exp\left(-\frac{x^2}{2\sigma^2}\right)\quad x\geqslant 0 \tag{1.3.19}$$

其中 $\sigma$ 为常数，则称 $X$ 服从瑞利分布，概率密度曲线如图 1.3(c)所示。

### 4. 指数分布

如果随机变量 $X$ 的概率密度函数为

$$f(x)=\frac{1}{\mu}\exp\left(-\frac{x}{\mu}\right)\quad x>0 \tag{1.3.20}$$

其中 $\mu$ 为常数，则称 $X$ 服从指数分布，概率密度曲线如图 1.3(d)所示。

### 5. 韦伯分布

如果随机变量 $X$ 的概率密度为

$$f(x)=\begin{cases}\dfrac{b}{a}\left(\dfrac{x}{a}\right)^{b-1}\exp\left[-\left(\dfrac{x}{a}\right)^{b}\right] & x\geqslant 0\\ 0 & x<0\end{cases}$$

其中 $a$、$b$ 为常数，则称 $X$ 服从韦伯分布，参数 $a$ 称为尺度参数，$b$ 称为形状参数，雷达的地杂波的幅度特性通常可以用韦伯分布来描述，概率密度曲线如图 1.3(e)所示。

### 6. 对数正态分布

如果随机变量 $X$ 的概率密度为

$$f(x)=\begin{cases}\dfrac{1}{x\sqrt{2\pi}\sigma}\exp\left[-\dfrac{\ln^2(x/m)}{2\sigma^2}\right] & x\geqslant 0\\ 0 & x<0\end{cases}$$

其中 $m$,$\sigma$ 均为非负的常数，则称 $X$ 服从对数正态分布，雷达的海杂波的幅度特性通常可以用对数正态分布来描述，概率密度曲线如图 1.3(f)所示。

### 7. K 分布

如果随机变量 $X$ 的概率密度函数为

$$f(x)=\frac{\sqrt{2\nu}}{\sqrt{\mu}\,2^{\nu-1}\,\Gamma(\nu)}\left(\sqrt{\frac{2\nu}{\mu}}x\right)^{\nu}\mathrm{K}_{\nu-1}\left(\sqrt{\frac{2\nu}{\mu}}x\right)\quad x\geqslant 0 \tag{1.3.21}$$

则称 $X$ 服从 K 分布，其中 $\mu>0$ 为比例参数，$\nu>0$ 为形状参数，$\Gamma(\cdot)$ 为伽马函数，$\mathrm{K}_{\nu-1}(\cdot)$ 为第二类 $\nu-1$ 阶修正贝塞尔函数。K 分布是描述现代高分辨率雷达杂波的一种统计模型，

概率密度曲线如图 1.3(g)所示。

### 8. 拉普拉斯分布

如果随机变量 $X$ 的概率密度函数为

$$f(x) = \frac{1}{2}c\exp(-c|x-m|) \tag{1.3.22}$$

其中 $c,m$ 均为常数，且 $c>0$，则称 $X$ 服从拉普拉斯分布。拉普拉斯分布被广泛应用于语音信号和图像灰级的统计建模，概率密度曲线如图 1.3(h)所示。

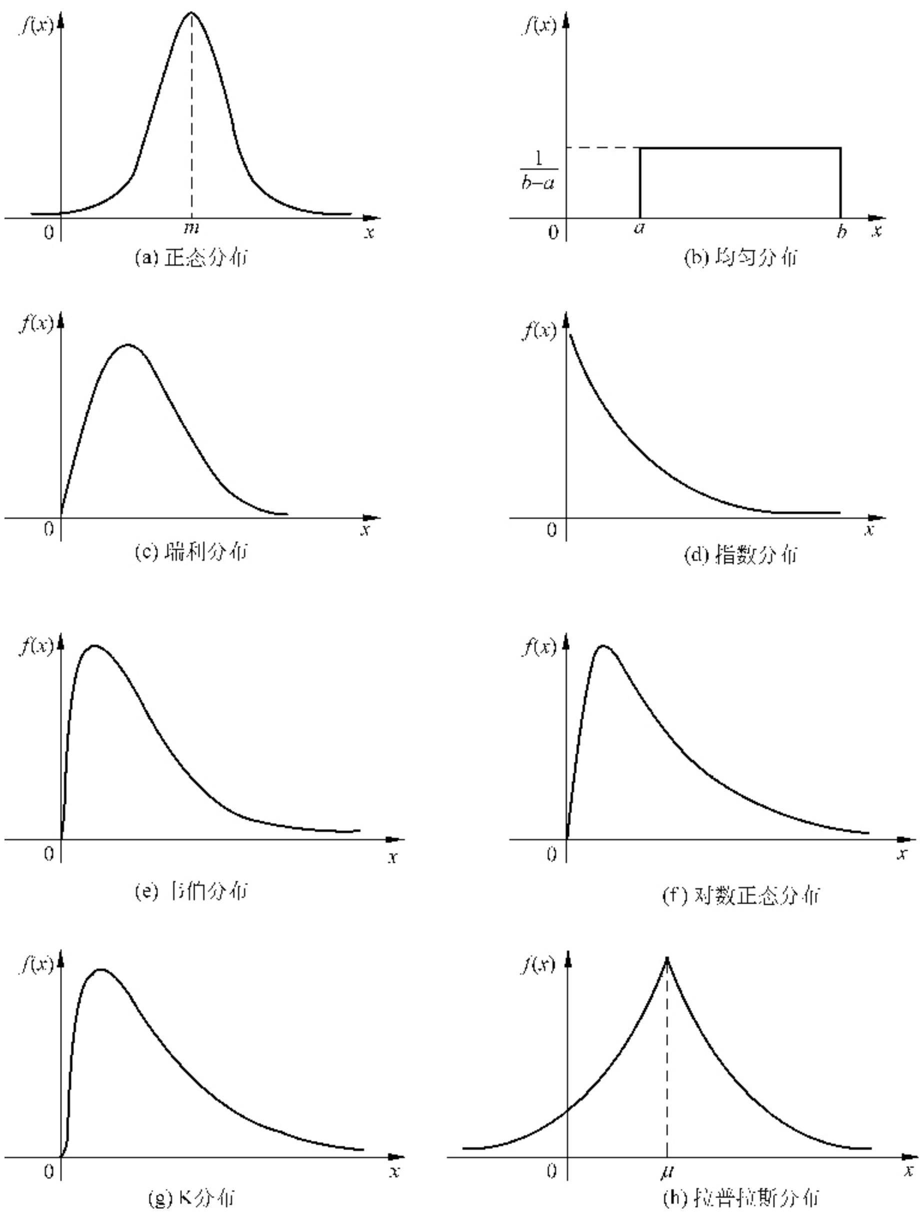


**图 1.3 常见概率密度分布**

# 1.4 多维随机变量及分布

在实际中，实验结果通常需要用多个随机变量才能加以描述，例如回波信号的幅度和相位需要两个不同的随机变量来描述。由多个随机变量构成的矢量称为多维随机变量或随机矢量。

## 1.4.1 二维随机变量

设随机试验 $E$ 的样本空间 $S=\{e\}$，$X=X(e)$ 和 $Y=Y(e)$ 是定义在样本空间 $S$ 上的两个随机变量，由 $X$ 和 $Y$ 构成的矢量 $(X,Y)$ 称为二维随机变量或二维随机矢量。

### 1. 二维分布函数

设 $x$、$y$ 为任意实数，那么二维随机变量 $(X,Y)$ 的分布函数定义为

$$F(x,y) = P(X \leqslant x, Y \leqslant y) \tag{1.4.1}$$

二维随机变量 $(X,Y)$ 的取值 $(x,y)$ 可以看作为平面上的一个点，那么二维分布函数就是二维随机变量 $(X,Y)$ 的取值落在图 1.4 所示的阴影区域的概率。

二维随机变量的分布函数具有下列性质：

(1) $0 \leqslant F(x,y) \leqslant 1$。

(2) 分布函数满足 $F(-\infty,y)=0$，$F(x,-\infty)=0$，$F(-\infty,-\infty)=0$，$F(\infty,\infty)=1$。

(3) $F(x,\infty)=F_X(x)$，$F(\infty,y)=F_Y(y)$，$F_X(x)$ 和 $F_Y(y)$ 称为边缘分布，即随机变量 $X$ 和 $Y$ 的分布，由二维分布函数可以求出一维分布函数。

(4) 对于任意的 $(x_1,y_1)$ 和 $(x_2,y_2)$，且 $x_2>x_1$，$y_2>y_1$，则

$$P(x_1 < X \leqslant x_2;\ y_1 < Y \leqslant y_2) = F(x_2,y_2) - F(x_2,y_1) - F(x_1,y_2) + F(x_1,y_1) \tag{1.4.2}$$

(1.4.2)式给出了利用二维分布函数计算二维随机变量落在某一区域的概率的方法，如图 1.5 所示。

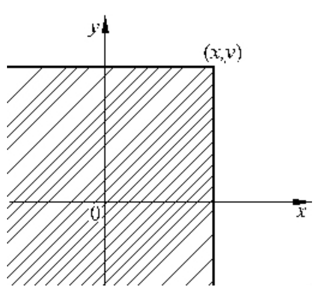


**图 1.4　二维分布函数图**

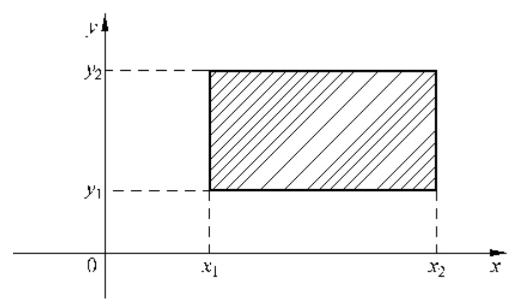


**图 1.5　二维随机变量落在某一区域的概率**

如果二维随机变量$(X,Y)$的可能取值为有限个或可列无穷个，则称$(X,Y)$为离散型随机变量。设

$$P(X=x_i,Y=y_i)=p_{ij} \quad i,j=1,2,\cdots$$

那么

$$\sum_i \sum_j p_{ij}=1 \tag{1.4.3}$$

$P(X=x_i,Y=y_i)=p_{ij}$称为$(X,Y)$的联合概率分布列或简称为分布列。

### 2. 二维概率密度

二维分布函数$F(x,y)$的二阶偏导数：

$$f(x,y)=\frac{\partial^2 F(x,y)}{\partial x \partial y} \tag{1.4.4}$$

定义为$(X,Y)$的二维联合概率密度，简称为二维概率密度。

二维概率密度具有以下性质：

(1) $f(x,y)\geqslant 0$，即概率密度是非负的函数。

(2) $F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y} f(x,y)\mathrm{d}x\mathrm{d}y$，$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\mathrm{d}x\mathrm{d}y=F(\infty,\infty)=1$。 (1.4.5)

(3) 边缘概率密度可由二维概率密度求得

$$f_X(x)=\int_{-\infty}^{+\infty} f(x,y)\mathrm{d}y, \quad f_Y(y)=\int_{-\infty}^{+\infty} f(x,y)\mathrm{d}x \tag{1.4.6}$$

(4) 设$G$是$x-y$平面上的一个区域，则二维随机变量$(X,Y)$的取值落在该区域的概率为

$$P\{(X,Y)\in G\}=\iint_G f(x,y)\mathrm{d}x\mathrm{d}y \tag{1.4.7}$$

## 1.4.2 条件分布

设$X$为一随机变量，$A$是一随机事件，定义

$$F(x \mid A)=P\{X\leqslant x \mid A\} \tag{1.4.8}$$

为随机变量$X$在事件$A$发生时的条件分布函数，对应的条件概率密度定义为条件分布函数的导数，即

$$f(x \mid A)=\frac{\mathrm{d}F(x \mid A)}{\mathrm{d}x} \tag{1.4.9}$$

由概率的特性，(1.4.8)式可以写成

$$F(x \mid A)=\frac{P\{X\leqslant x,A\}}{P\{A\}} \tag{1.4.10}$$

设有二维随机变量$(X,Y)$，令$A=\{X=x\}$，定义

$$F_{Y|X}(y \mid x)=P\{Y\leqslant y \mid X=x\} \tag{1.4.11}$$

为随机变量$Y$在$X=x$时的条件分布函数，或称为$Y$对$X$的条件分布函数。相应地定义

$$f_{Y|X}(y \mid x)=\frac{\partial F_{Y|X}(y \mid x)}{\partial y} \tag{1.4.12}$$

为随机变量 $Y$ 在 $X=x$ 时的条件概率密度，或称为 $Y$ 对 $X$ 的条件概率密度。

可以证明（证明留作习题，参见习题 1.1）

$$f_{Y|X}(y \mid x)=\frac{f(x,y)}{f_X(x)},\quad f_{X|Y}(x \mid y)=\frac{f(x,y)}{f_Y(y)} \tag{1.4.13}$$

于是有

$$f(x,y)=f_{X|Y}(x \mid y)f_Y(y)=f_{Y|X}(y \mid x)f_X(x) \tag{1.4.14}$$

如果

$$f(x,y)=f_X(x)f_Y(y) \tag{1.4.15}$$

则称随机变量 $X$ 和 $Y$ 是相互独立的。

### 1.4.3 多维分布

下面将二维分布的一些结论直接推广到多维的情况。

#### 1. 多维分布函数

设有 $n$ 维随机变量 $(X_1,X_2,\cdots,X_n)$，定义

$$F(x_1,x_2,\cdots,x_n)=P(X_1\leqslant x_1,X_2\leqslant x_2,\cdots,X_n\leqslant x_n) \tag{1.4.16}$$

为 $n$ 维随机变量的 $n$ 维分布函数。$n$ 维分布函数具有下列性质：

(1) $F(x_1,x_2,\cdots,-\infty,\cdots,x_n)=0$，其中 $x_i=-\infty(i=1,2,\cdots,n)$。

(2) $F(\infty,\infty,\cdots,\infty)=1$。

(3) $F(\infty,\infty,\cdots,\infty,x_m,x_{m+1},\cdots,x_n)=F(x_m,x_{m+1},\cdots,x_n)$。

#### 2. 多维概率密度

若 $n$ 维分布函数的 $n$ 阶混合偏导数存在，那么定义

$$f(x_1,x_2,\cdots,x_n)=\frac{\partial^n F(x_1,x_2,\cdots,x_n)}{\partial x_1\partial x_2\cdots\partial x_n} \tag{1.4.17}$$

为 $n$ 维随机变量的 $n$ 维概率密度。显然

$$F(x_1,x_2,\cdots,x_n)=\int_{-\infty}^{x_1}\int_{-\infty}^{x_2}\cdots\int_{-\infty}^{x_n}f(x_1,x_2,\cdots,x_n)\mathrm{d}x_1\mathrm{d}x_2\cdots\mathrm{d}x_n \tag{1.4.18}$$

对于 $n$ 维随机变量，其取值落在区域 $G$ 内的概率可表示为

$$P\{(X_1,X_2,\cdots,X_n)\in G\}=\iint\limits_G f(x_1,x_2,\cdots,x_n)\mathrm{d}x_1\mathrm{d}x_2\cdots\mathrm{d}x_n \tag{1.4.19}$$

#### 3. 多维条件概率密度

对于 $n$ 维随机变量 $(X_1,X_2,\cdots,X_n)$，在 $X_{k+1},X_{k+2},\cdots,X_n$ 的取值为 $x_{k+1},x_{k+2},\cdots,x_n$ 的条件下，$X_1,X_2,\cdots,X_k$ 的条件概率密度为

$$f(x_1,x_2,\cdots,x_k \mid x_{k+1},x_{k+2},\cdots,x_n)=\frac{f(x_1,x_2,\cdots,x_n)}{f(x_{k+1},x_{k+2},\cdots,x_n)} \tag{1.4.20}$$

显然，$n$ 维概率密度与条件概率密度之间有如下关系：

$$f(x_1,x_2,\cdots,x_n)=f(x_1)f(x_2 \mid x_1)f(x_3 \mid x_1,x_2)\cdots f(x_n \mid x_1,x_2,\cdots,x_{n-1}) \tag{1.4.21}$$