# 概率论与数理统计

## 第三章: 联合分布

李立颖 (lily@sustech.edu.cn)

南方科技大学数学系

2023 秋季

## 多维随机变量的背景

**例**

- 人的身高 $H$ 与体重 $W$
- 某地区的气温 $X$, 气压 $Y$ 与湿度 $Z$
- 射击中落点横向偏差与纵向偏差 $Y$

**问**

能不能将上述随机变量单独分别进行研究?

**分析**

- 一般人的身高 $H \sim \mathcal{N}(\cdot,\cdot)$, $W \sim \mathcal{N}(\cdot,\cdot)$. 但身高与体重之间有一定关系.
- 气象指标中的气温、气压与湿度也是相关联的
- 导弹射程误差与落点的横向偏差及纵向偏差都有关

由于同一对象的不同指标之间往往是有一定联系的, 所以应该把它们作为一个从中整体来看待.

## 二维随机变量的概念

### 定义

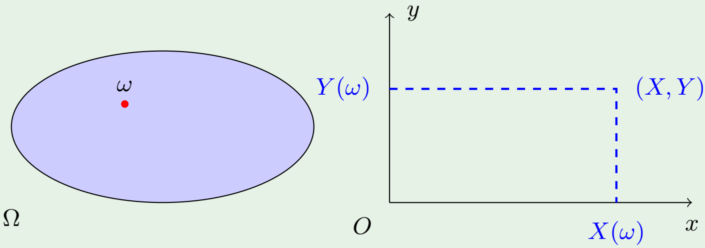
设 $\Omega$ 为样本空间, $X = X(\omega)$, $Y = Y(\omega)$, $\omega \in \Omega$, 是定义在 $\Omega$ 上的两个随机变量. 记

$$(X,Y) := \big(X(\omega), Y(\omega)\big), \quad \omega \in \Omega.$$

称 $(X,Y)$ 为**二维随机变量 (向量)**.

### 注

一个试验产生的二维随机变量可视为向二维平面 "投掷" 一个 "随机点".



# 累积分布函数

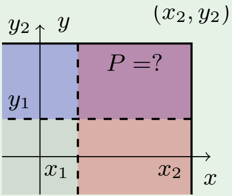
## 定义

设 $(X,Y)$ 为二维随机变量. 对任意 $x,y\in\mathbb{R}$, 定义

$$F(x,y) := \mathsf{P}\Big(\{X\le x\}\cap\{Y\le y\}\Big) = \mathsf{P}(X\le x,\ Y\le y).$$

称 $F(x,y)$ 为二维随机变量 $(X,Y)$ 的**累积分布函数**, 或称为 $X$ 与 $Y$ 的**联合累积分布函数**.

## 计算 $\mathsf{P}(X\in(x_1,x_2],\ Y\in(y_1,y_2])$




$$\mathsf{P}(X\in(x_1,x_2],\ Y\in(y_1,y_2]) = F(x_2,y_2)-F(x_1,y_2)-F(x_2,y_1)+F(x_1,y_1)\ge 0.$$

# 分布函数 $F(x,y)$ 的本质特征

## 单调性

对任意固定 $x_0$, $F(x_0,y)$ 是 $y$ 的单调不减函数; 对任意固定 $y_0$, $F(x,y_0)$ 是 $x$ 的单调不减函数.

## 无穷处极限

$0 \le F(x,y) \le 1$ 且

$$F(\infty,\infty)=1,\quad F(-\infty,y)=F(x,-\infty)=F(-\infty,-\infty)=0,\quad \forall x,y.$$

## 右连续性

对任意 $x,y$, $F(x,\cdot)$, $F(\cdot,y)$ 是右连续的.

## 二维单调性

$$F(x_2,y_2)-F(x_1,y_2)-F(x_2,y_1)+F(x_1,y_1)=\mathsf{P}(X\in(x_1,x_2],\ Y\in(y_1,y_2])\ge 0.$$

# 二维单调性

## 注

最后一条性质不能由前三条性质推出.

## 反例

$$F(x,y)=\begin{cases}1, & x+y\geq -1,\\ 0, & x+y<-1.\end{cases}$$

则 $F(1,1)-F(-1,1)-F(1,-1)+F(-1,-1)=1-1-1+0<0.$

# $n$ 维随机变量

## 定义

设 $X_1, X_2, \ldots, X_n$ 是定义在样本空间 $\Omega$ 上的 $n$ 个随机变量. 称

$$(X_1, X_2, \ldots, X_n)$$

为 $n$ **维随机变量** 或 $n$ **维随机向量**.

## 分布函数

称 $n$ 元函数

$$F(x_1, x_2, \ldots, x_n) = \mathsf{P}(X_1 \le x_1,\ X_2 \le x_2, \ldots, X_n \le x_n)$$

为 $n$ 维随机向量 $(X_1, X_2, \ldots, X_n)$ 的**分布函数** 或 $X_1, X_2, \ldots, X_n$ 的**联合分布 (函数)**.

## 二维随机变量的基本分类

- 二维离散型随机变量
- 二维连续型随机变量

# 边缘分布 (marginal distribution)

如果 $(X,Y)$ 是一个二维随机变量, 则它的分量 $X$ (或者 $Y$ ) 是一维随机变量. 因此 $X$ 或 $Y$ 也有分布.

## 定义

称 $X$ 或 $Y$ 的分布为 $X$ 或 $Y$ 关于二维随机变量 $(X,Y)$ 的**边缘分布 (边际分布)**.

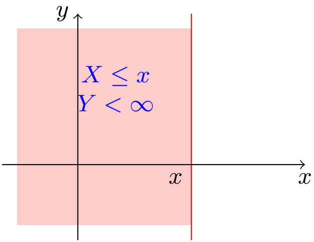
## 边际分布函数

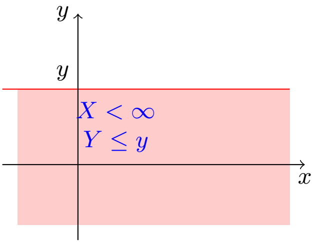
- 二维随机变量的整体概率特性: $(X,Y) \sim F(x,y)$.
- 两个一维随机变量的概率特性: $X \sim F_X(x)$, $Y \sim F_Y(y)$.

### 定义

称 $F_X(x)$、$F_Y(y)$ 为 $(X,Y)$ 关于 $X$、$Y$ 的**边际分布 (函数)**.

# 联合分布与边际分布的关系




$$F_X(x) = F(x, \infty)$$

$$F_Y(y) = F(\infty, y)$$

## 性质

随机变量的边际分布完全由它们的联合分布决定. **反之不然.**

## 例题

设随机变量 $(X,Y)$ 的联合分布函数为

$$F(x,y)=A\left(B+\arctan\frac{x}{2}\right)\left(C+\arctan\frac{y}{2}\right),\quad -\infty<x,y<\infty,$$

其中 $A,B,C$ 为常数.

1. 确定 $A,B,C$.
2. 求 $X$ 和 $Y$ 的边缘分布函数.
3. 求 $\mathrm{P}(X>2)$ .

## 解

$$\begin{aligned}F(\infty,\infty)&=A(B+\pi/2)(C+\pi/2)=1\\F(-\infty,\infty)&=A(B-\pi/2)(C+\pi/2)=0\\F(\infty,-\infty)&=A(B+\pi/2)(C-\pi/2)=0.\end{aligned}$$

因此 $B=C=\pi/2,\ A=\dfrac{1}{\pi^2}$.

**解**

$$F_X(x) = F(x, \infty) = \frac{1}{2} + \frac{1}{\pi}\arctan\frac{x}{2}, \quad -\infty < x < \infty,$$
$$F_Y(y) = F(\infty, y) = \frac{1}{2} + \frac{1}{\pi}\arctan\frac{y}{2}, \quad -\infty < y < \infty,$$

$$\mathsf{P}(X > 2) = 1 - \mathsf{P}(X \le 2) = 1 - F_X(2) = 1 - \left(\frac{1}{2} + \frac{1}{\pi}\arctan\frac{2}{2}\right) = \frac{1}{4}.$$

# 二维离散型随机变量

## 回顾: 二维随机变量分类

- 二维离散型随机变量
- 二维连续型随机变量

## 定义

设随机变量 $(X, Y)$ 的所有可能的取值为 $(x_i, y_j)$, $i, j = 1, 2, \ldots$, 取值的概率为

$$\mathsf{P}(X = x_i,\ Y = y_j) = p(x_i, y_j) := p_{i,j}, \quad i, j = 1, 2, \ldots$$

则称 $(X, Y)$ 为**离散型随机变量**, 称 $p_{i,j}$ 为它的**(联合) 频率函数 (joint frequency function)**.

# 频率函数的基本性质

## 本质特征

设随机变量 $(X,Y)$ 的频率函数为

$$p_{ij} = \mathsf{P}(X = x_i,\ Y = y_j), \quad i,j = 1,2,\ldots$$

则

- $p_{ij} \geq 0,\ i,j = 1,2,\ldots$
- $\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} p_{ij} = 1.$

## 频率函数的表格表示法

| $Y \backslash X$ | $x_1$ | $x_2$ | $\cdots$ | $x_i$ | $\cdots$ |
|---|---|---|---|---|---|
| $y_1$ | $p_{11}$ | $p_{21}$ | $\cdots$ | $p_{i1}$ | $\cdots$ |
| $y_2$ | $p_{12}$ | $p_{22}$ | $\cdots$ | $p_{i2}$ | $\cdots$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| $y_j$ | $p_{1j}$ | $p_{2j}$ | $\cdots$ | $p_{ij}$ | $\cdots$ |

## 例题

袋中装有 2 只白球及 3 只黑球. 现进行无放回的摸球, 定义随机变量如下:

$$X = \begin{cases} 1, & \text{第一次摸出白球}, \\ 0, & \text{第一次摸出黑球}, \end{cases} \quad Y = \begin{cases} 1, & \text{第二次摸出白球}, \\ 0, & \text{第二次摸出黑球}. \end{cases}$$

求 $(X, Y)$ 的频率函数.

## 解

$$\mathsf{P}(X=0,\ Y=0) = \mathsf{P}(Y=0 \mid X=0) \cdot \mathsf{P}(X=0) = \frac{2}{4} \cdot \frac{3}{5} = \frac{6}{20},$$
$$\mathsf{P}(X=0,\ Y=1) = \mathsf{P}(Y=1 \mid X=0) \cdot \mathsf{P}(X=0) = \frac{2}{4} \cdot \frac{3}{5} = \frac{6}{20},$$
$$\mathsf{P}(X=1,\ Y=0) = \mathsf{P}(Y=0 \mid X=1) \cdot \mathsf{P}(X=1) = \frac{3}{4} \cdot \frac{2}{5} = \frac{6}{20},$$
$$\mathsf{P}(X=1,\ Y=1) = \mathsf{P}(Y=1 \mid X=1) \cdot \mathsf{P}(X=1) = \frac{1}{4} \cdot \frac{2}{5} = \frac{2}{20}.$$

## 例题

有一个射击游戏, 参加游戏的人先掷一下骰子, 若出现点数为 $X$, 则射击 $X$ 次. 设某人击中目标概率为 $p = 0.9$. 记击中目标的次数为 $Y$. 求 $(X, Y)$ 的频率函数.

## 分析

$X$ 的取值为 $1, 2, \ldots, 6$, $Y$ 的取值为 $0, 1, \ldots, X$. 当 $X = i$ 时, $Y \sim \mathrm{Bin}(i, p)$, $i = 1, 2, \ldots, 6$. 我们应当用乘法公式计算概率!

## 解

由乘法公式及如上分析,

$$\begin{aligned} \mathsf{P}(X = i,\ Y = j) &= \mathsf{P}(Y = j \mid X = i)\mathsf{P}(X = i) \\ &= \frac{1}{6}\binom{i}{j} p^j (1-p)^{i-j}, \quad 0 \le j \le i,\ 1 \le i \le 6. \end{aligned}$$

## 边际频率函数

设 $(X,Y)$ 的频率函数为 $\mathsf{P}(X=x_i,\ Y=y_j)=p_{ij}$, $i,j=1,2,\ldots$. 则随机变量 $X$ 的频率函数是

$$\mathsf{P}(X=x_i)=\sum_{j=1}^{\infty}p_{ij}:=p_{i\cdot},\quad i=1,2,\ldots$$

同理 $Y$ 的频率函数是

$$\mathsf{P}(Y=y_j)=\sum_{i=1}^{\infty}p_{ij}:=p_{\cdot j},\quad j=1,2,\ldots$$

### 定义

称数列 $\{p_{i\cdot}\}$ 为 $(X,Y)$ 关于 $X$ 的**边际频率函数**, 称数列 $\{p_{\cdot j}\}$ 为 $(X,Y)$ 关于 $Y$ 的**边际频率函数 (marginal frequency function)**.

### 边际频率函数的性质

- 它是一维随机变量的频率函数
- 它可通过二维随机变量的频率函数计算得到.

## 例子

设随机变量 $X$ 从 1, 2, 3, 4 中等可能取值, 又设随机变量 $Y$ 从 $1, 2, \ldots, X$ 中等可能取值. 求 $X, Y$ 的联合频率函数以及边际频率函数.

## 解

当 $X = i$ 时, $Y$ 以 $\frac{1}{i}$ 的概率等可能取到 1 至 $i$. 因此由乘法公式,

$$\mathsf{P}(X = i,\ Y = j) = \mathsf{P}(Y = j \mid X = i) \cdot \mathsf{P}(X = i) = \frac{1}{i} \cdot \frac{1}{4}, \quad 1 \le j \le i.$$

因此 $X, Y$ 的频率函数及边际频率函数为

| $Y \backslash X$ | 1 | 2 | 3 | 4 | $p_{\cdot j} = \sum_{i=1}^{4} p_{ij}$ |
|---|---|---|---|---|---|
| 1 | 1/4 | 1/8 | 1/12 | 1/16 | 25/48 |
| 2 | 0 | 1/8 | 1/12 | 1/16 | 13/48 |
| 3 | 0 | 0 | 1/12 | 1/16 | 7/48 |
| 4 | 0 | 0 | 0 | 1/16 | 3/48 |
| $p_{i\cdot} = \sum_{j=1}^{4} p_{ij}$ | 1/4 | 1/4 | 1/4 | 1/4 | |

# $n$ 维离散型随机变量的边际频率函数

设 $X_1, \ldots, X_n$ 的联合频率函数为

$$\mathsf{P}(X_1 = x_1, \ldots, X_n = x_n) = p(x_1, \ldots, x_n).$$

## 边际频率函数

随机变量 $X_1$ 的**边际频率函数**是

$$\mathsf{P}_{X_1}(x_1) = \sum_{x_2, \ldots, x_n} p(x_1, x_2, \ldots, x_n).$$

随机变量 $X_1, X_2$ 的**二维边际频率函数** 是

$$p_{X_1 X_2}(x_1, x_2) = \sum_{x_3, \ldots, x_n} p(x_1, x_2, \ldots, x_n).$$

# 多项分布 (multinomial distribution): 二项分布的推广

## 定义

假设进行 $n$ 次独立试验，每次试验有 $r$ 种可能的结果，各自出现的概率是 $p_1, \dots, p_r$. 令 $N_i$ 是 $n$ 次试验中出现第 $i$ 种试验结果的所有次数，其中 $i = 1, \dots, r$. 则 $N_1, \dots, N_r$ 的联合频率函数是

$$p(n_1, \dots, n_r) = \binom{n}{n_1 \cdots n_r} p_1^{n_1} p_2^{n_2} \cdots p_r^{n_r},$$

称为**多项分布**.

## 两种计算 $N_i$ 边际频率函数的方法

- 将联合频率函数关于其它的 $n_j$ 求和;
- $N_i$ 可解释为 $n$ 次试验中**成功的次数**，故 $N_i \sim \mathrm{Bin}(n, p_i)$，因此

$$p_{N_i}(n_i) = \binom{n}{n_i} p_i^{n_i} (1 - p_i)^{n - n_i}.$$

## 例子

箱子里装有 4 只白球和 2 只黑球. 在其中随机地取两次, 每次取一只. 分别考虑 "有放回抽样" 与 "无放回抽样". 定义

$$X=\begin{cases}0, & \text{第一次为黑球},\\ 1, & \text{第一次为白球}.\end{cases}\quad Y=\begin{cases}0, & \text{第二次为黑球},\\ 1, & \text{第二次为白球}.\end{cases}$$

求 $X, Y$ 的联合分布律和边缘分布律.

## 解

有放回抽样:

| Y \ X | 0 | 1 | $p_{i\cdot}$ |
|---|---|---|---|
| 0 | 1/9 | 2/9 | 1/3 |
| 1 | 2/9 | 4/9 | 2/3 |
| $p_{\cdot j}$ | 1/3 | 2/3 | 1 |

无放回抽样:

| Y \ X | 0 | 1 | $p_{i\cdot}$ |
|---|---|---|---|
| 0 | 1/15 | 4/15 | 1/3 |
| 1 | 4/15 | 6/15 | 2/3 |
| $p_{\cdot j}$ | 1/3 | 2/3 | 1 |

## 例题

袋中有 1 个红球, 2 个黑球, 3 个白球. 现有放回地取两次, 每次取一球, 以 $X$, $Y$, $Z$ 分别表示取球所得红、黑、白球个数. 求: $\mathsf{P}(X=1 \mid Z=0)$, $\mathsf{P}(X=1,\ Z=0)$, 以及 $(X,Y)$ 的分布.

## 解

$$\mathsf{P}(X=1 \mid Z=0) = \mathsf{P}(\mathrm{Bin}(2,1/3)=1) = \frac{1}{3}\times\frac{2}{3}+\frac{2}{3}\times\frac{1}{3}=\frac{4}{9}$$
$$\mathsf{P}(X=1,\ Z=0) = \frac{1}{6}\times\frac{2}{6}+\frac{2}{6}\times\frac{1}{6}=\frac{1}{9}.$$

$X,Y$ 的取值均为 0, 1, 2. 我们只计算几个例子:

$$\mathsf{P}(X=0,\ Y=0) = \frac{3}{6}\times\frac{3}{6}, \qquad \text{两个均为白球}$$
$$\mathsf{P}(X=0,\ Y=1) = \frac{2}{6}\times\frac{3}{6}\times 2=\frac{1}{3}, \qquad \text{黑白或白黑}$$
$$\mathsf{P}(X=1, Y=2) = 0, \qquad \text{总数超 2 只, 不可能}$$
$$\mathsf{P}(X=2, Y=0) = \frac{1}{6}\times\frac{1}{6}=\frac{1}{36}, \qquad \text{两只均为红球.}$$

# 作业

- P76: 3
- 补充题
  1. 把一枚均匀硬币抛掷三次, 设 $X$ 为三次抛掷中正面出现的次数, 而 $Y$ 为正面出现次数与反面出现次数之差的绝对值. 求 $(X,Y)$ 的频率函数.
  2. 设 $X$ 的分布为 $\mathsf{P}(X=-1)=\mathsf{P}(X=0)=\mathsf{P}(X=1)=\frac{1}{3}$. 令 $Y=X^2$. 求 $(X,Y)$ 的联合频率函数及边缘频率函数.
  3. 设随机变量 $Y$ 服从参数为 1 的指数分布, 随机变量
  $$X_k=\begin{cases}0, & Y\le k,\\ 1, & Y>k,\end{cases}\qquad k=1,2.$$
  求二维随机变量 $(X_1,X_2)$ 的联合频率函数及边缘频率函数.

## 二维连续型随机变量

### 定义

设随机变量 $(X,Y)$ 的联合分布函数 (joint cdf) 为

$$F(x,y) = \mathrm{P}(X \le x,\ Y \le y).$$

若存在非负可积函数 $f(x,y) \ge 0$ 使得

$$F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u,v)\,dudv, \quad (x,y) \in \mathbb{R}^2,$$

则称 $(X,Y)$ 为**二维连续型随机变量**, 称 $f(x,y)$ 为**概率密度函数 (密度函数、密度、联合密度函数)**.

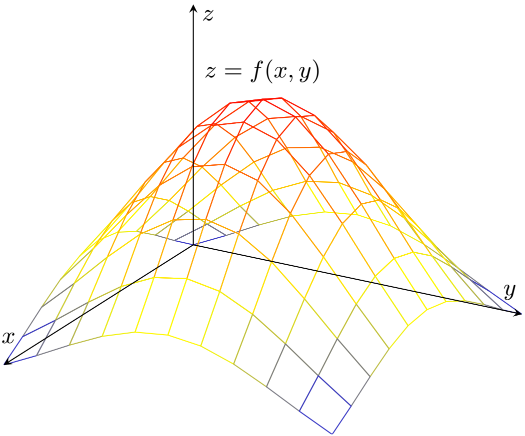
### 注

由微积分知识, $F(x,y)$ 是连续函数.

# 密度函数的基本性质

## 本质特征

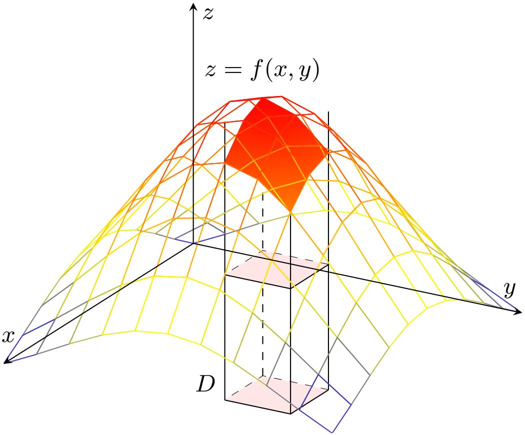
- $f(x,y) \geq 0,\ \forall (x,y) \in \mathbb{R}^2$,
- $\displaystyle\int_{\mathbb{R}}\int_{\mathbb{R}} f(u,v)\,dudv = 1$. **几何意义：曲面 $z = f(x,y)$ 与 $xOy$ 平面围成的体积为 1**.



## 区域概率

对任意 (由逐段光滑曲线围成的) 区域 $D \subset \mathbb{R}^2$,

$$\mathsf{P}\big((X,Y) \in D\big) = \iint_D f(x,y)\,dxdy.$$



# 分布函数与密度函数

## 性质

在 $f(x,y)$ 的连续点处，有

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

## 注

在 $f(x,y)$ 的连续点处，由导数的定义有

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y} = \lim_{\Delta x \to 0^+,\ \Delta y \to 0^+} \frac{\mathsf{P}(x < X \le x + \Delta x,\ y < Y \le y + \Delta y)}{\Delta x \Delta y}.$$

## 密度函数的意义

$$\mathsf{P}(x < x \le x + \Delta x,\ y < Y \le y + \Delta y) \approx f(x,y)\Delta x \Delta y.$$

## 例题

设随机变量 $(X,Y)$ 的概率密度为

$$f(x,y) = \mathbb{1}_{\{x>0,\ y>0\}} k e^{-(2x+y)}.$$

1. 确定常数 $k$;
2. 求分布函数 $F(x,y)$;
3. 计算概率 $\mathrm{P}(Y \le X)$.

## 解

1. 由归一性,

$$1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(u,v)\,du dv = k\int_0^{\infty}\int_0^{\infty} e^{-(2x+y)}\,dx dy$$
$$= k\left[\int_0^{\infty} e^{-2x}\,dx\right]\cdot\left[\int_0^{\infty} e^{-y}\,dy\right] = \frac{k}{2}.$$

因此 $k=2$.

$$f(x,y) = \mathbb{1}_{\{x>0,\ y>0\}} k e^{-(2x+y)}.$$

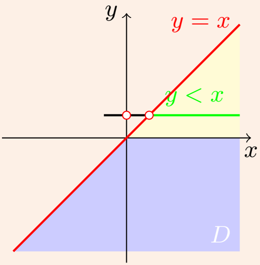
## 解

2. 显然，分布函数只有在 $x>0$, $y>0$ 时才非零. 我们有

$$\begin{aligned} F(x,y) &= \mathbb{1}_{\{x>0,\ y>0\}} \int_0^x \int_0^y 2e^{-(2u+v)}\,du dv \\ &= \mathbb{1}_{\{x>0,\ y>0\}} (1-e^{-2x})(1-e^{-y}). \end{aligned}$$

3. 记 $D = \{(x,y) \mid y \le x,\ x,y \in (-\infty,\infty)\}$. 于是

$$\begin{aligned} \mathsf{P}(Y \le X) &= \mathsf{P}\big((X,Y)\in D\big) \\ &= \iint_D f(x,y)\,dxdy \\ &= \iint_{D\cap\{x,y>0\}} 2e^{-(2x+y)}\,dxdy \\ &= \int_0^\infty dy \int_y^\infty 2e^{-(2x+y)}\,dx = \frac{1}{3}. \end{aligned}$$



## 例题

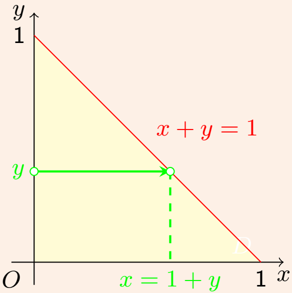
设随机变量 $(X, Y)$ 的密度函数为

$$f(x,y) = \mathbb{1}_{\{x>0,\ y>0\}} 2e^{-(2x+y)}.$$

计算概率 $\mathsf{P}(X+Y \le 1)$.

## 解

$$\begin{aligned}
\mathsf{P}(X+Y \le 1) &= \iint_{\{x+y\le 1\}} f(x,y)\,dxdy \\
&= \iint_{\{x+y\le 1, x>0, y>0\}} 2e^{-(2x+y)}\,dxdy \\
&= \int_0^1 dy \int_0^{1-y} 2e^{-(2x+y)}\,dx \\
&= 1 - 2e^{-1} + e^{-2}.
\end{aligned}$$



# 约定

## 记号

$$(X, Y) \sim F(x, y)$$

表示二维随机变量 $(X, Y)$ 的分布函数为 $F(x, y)$.

$$(X, Y) \sim f(x, y)$$

表示二维随机变量 $(X, Y)$ 的概率函数为 $f(x, y)$, 即

- 离散型随机变量 $f(x, y)$ 表示频率函数 (PMF);
- 连续型随机变量 $f(x, y)$ 表示密度函数 (PDF).

## $n$ 维随机变量的记号

$$(X_1, X_2, \ldots, X_n) \sim F(x_1, x_2, \ldots, x_n), \quad (X_1, X_2, \ldots, X_n) \sim f(x_1, x_2, \ldots, x_n).$$

设 $(X,Y)$ 的分布函数和密度函数分别为 $F(x,y)$, $f(x,y)$.

## $X$ 的边际分布

随机变量 $X$ 的分布函数为

$$F_X(x) = \mathsf{P}(X \le x) = \mathsf{P}(X \le x,\ Y < \infty) = \int_{-\infty}^{x} \left[ \int_{-\infty}^{\infty} f(u,y)\,dy \right] du$$

因此随机变量 $X$ 的密度函数为

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy, \quad x \in \mathbb{R}.$$

## $Y$ 的边际分布

$$F_Y(y) = \int_{-\infty}^{y} \left[ \int_{-\infty}^{\infty} f(x,v)\,dx \right] dv, \quad f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx.$$

## 定义

称 $f_X(x)$ $(f_Y(y))$ 为 $(X,Y)$ 关于 $X$ $(Y)$ 的**边际密度 (函数)**.

## 注

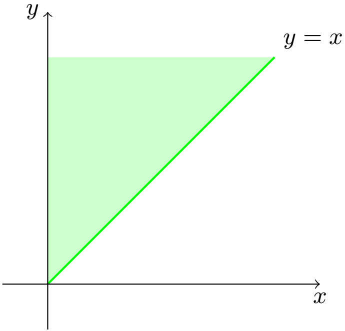
已知联合密度可以求得边际密度.

## 例子

设 $(X, Y)$ 的概率密度是

$$f(x,y) = \begin{cases} e^{-y}, & x > 0, y > x, \\ 0, & \text{其它}. \end{cases}$$

求 $(X, Y)$ 关于 $X$ 和 $Y$ 的边际密度函数.



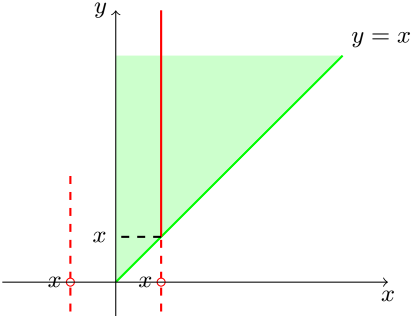
# $X$ 的边际分布

**解**

$f_X(x)=\int_{-\infty}^{\infty} f(x,y)\,dy$. 当 $x\le 0$ 时，显然 $f_X(x)=0$. 当 $x>0$ 时，由图知

$$f_X(x)=\int_x^{\infty} e^{-y}\,dy=e^{-y}|_{\infty}^{x}=e^{-x},\quad x\ge 0.$$

因此 $f_X(x)=\mathbb{1}_{\{x\ge 0\}}e^{-x}$ (指数分布).



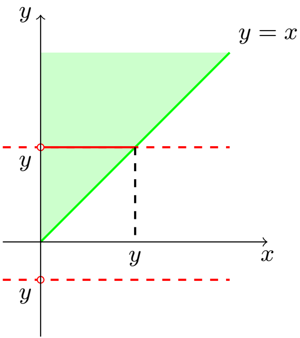
# $Y$ 的边际分布

**解**

当 $y \leq 0$ 时，$f_Y(y) = 0$. 当 $y > 0$ 时，

$$f_Y(y) = \int_0^y e^{-y}\,dx = ye^{-y}.$$

故 $f_Y(y) = \mathbb{1}_{\{y>0\}} ye^{-y}$.



## 例题

设 $(X, Y)$ 的联合密度为

$$f(x,y)=\begin{cases}6, & x^2\le y\le x,\\ 0, & \text{其它}.\end{cases}$$

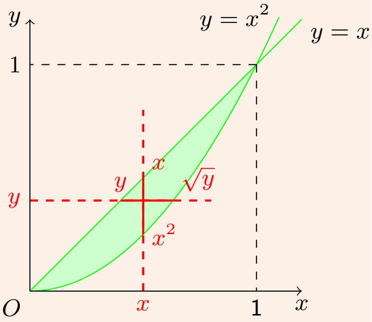
求边际密度 $f_X(x)$, $f_Y(y)$.

## 解

由图知 $0\le X,Y\le 1$. 因此

$$f_X(x)=\mathbb{1}_{(0,1)}(x)\int_{x^2}^{x}6\,dy=\mathbb{1}_{(0,1)}(x)6x(1-x),$$

$$f_Y(y)=\mathbb{1}_{(0,1)}(y)\int_{y}^{\sqrt{y}}6\,dx=6(\sqrt{y}-y).$$



# $n$ 维连续型变量的边际密度

**定义**

设 $X, Y, Z$ 的联合密度函数为 $f(x, y, z)$, 则随机变量 $X$ 的一维边际密度函数是

$$f_X(x) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y, z)\, dydz,$$

随机变量 $X$ 和 $Y$ 的二维边际密度函数是

$$f_{XY}(x, y) = \int_{-\infty}^{\infty} f(x, y, z)\, dz.$$

# 联合分布的非唯一性

## Farlie–Morgenstern 族

设 $F(x)$ 和 $G(y)$ 都是一维连续型分布函数. 则对任意的 $\alpha$, $|\alpha| \leq 1$,

$$H(x,y) = F(x)G(y)\Big[1+\alpha\Big(1-F(x)\Big)\Big(1-G(y)\Big)\Big]$$

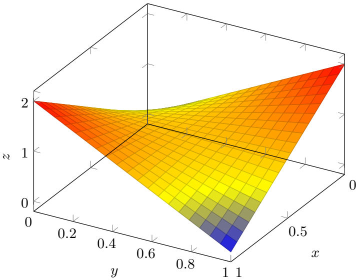
是二维连续型分布函数, 且边际分布为

$$H(x,\infty) = F(x), \quad H(\infty,y) = G(y).$$

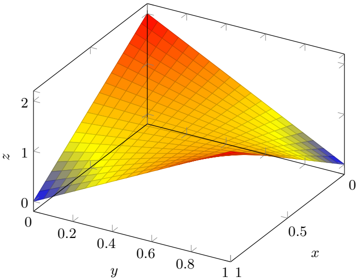
## 注

按这种方式, 可以构造给定边际分布的无数个不同的二维联合分布.

# $F = G = \mathrm{U}(0,1)$ 时 Farlie–Morgensten 族




(a) $\alpha = -1$: $h(x,y) = 2x + 2y - 4xy$




(b) $\alpha = 1$: $h(x,y) = 2 - 2x - 2y + 4xy$

# 连接函数 (copula)

## 定义

称那些使得边际分布为均匀分布的联合累积分布函数为**连接函数**, 记为 $C(u,v)$.

## 性质

1. $C(u,v)$ 关于每个变量都是非降的
2. $\mathrm{P}(U \le u) = C(u,1) = u$, $C(1,v) = v$.
3. 限定连续型连接函数, 即 $C(u,v)$ 具有密度:

$$c(u,v) = \frac{\partial^2}{\partial u \partial v} C(u,v) \ge 0.$$

# 由连接函数构造两个分布的耦合 (coupling)

## 引理

- 假设 $X$ 和 $Y$ 是分布函数分别为 $F_X(x)$ 和 $F_Y(y)$ 的连续型随机变量，则 $U = F_X(X)$ 和 $V = F_Y(Y)$ 是均匀分布 $\mathrm{U}(0,1)$.
- 对于连接函数 $C(u,v)$，考虑定义联合分布

$$F_{XY}(x,y) = C\big(F_X(x), F_Y(y)\big),$$

则其边际分布为 $F_X(x)$ 和 $F_Y(y)$，且密度为

$$f_{XY}(x,y) = c\big(F_X(x), F_Y(y)\big) f_X(x) f_Y(y).$$

## 注

由两个边际分布和任意连接函数，可以构造出相同边际分的联合分布，即：**边际函数不能决定联合分布**，两个变量的相依性由连接函数控制.

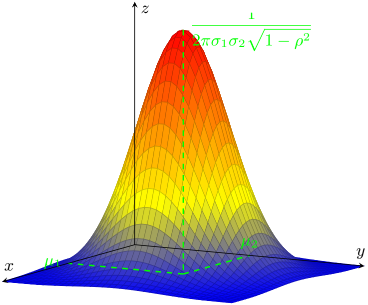
# 二维正态分布

## 定义

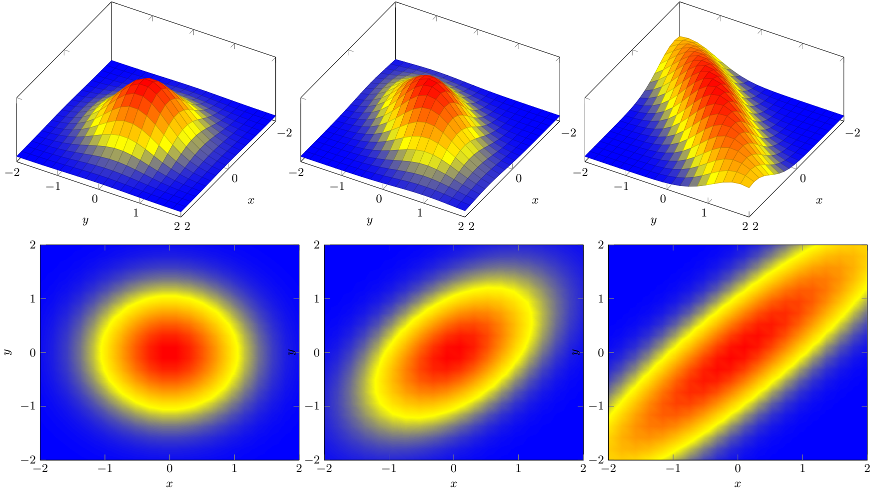
若 $X, Y$ 的联合密度为

$$f(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}e^{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2}-2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}\right]}$$

则称 $(X,Y)$ 服从参数为 $(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$ 的**二维正态分布**，记为 $(X,Y)\sim\mathcal{N}(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$. 其中 $\mu_1,\mu_2\in\mathbb{R}$, $\sigma_1,\sigma_2>0$, $|\rho|<1$.

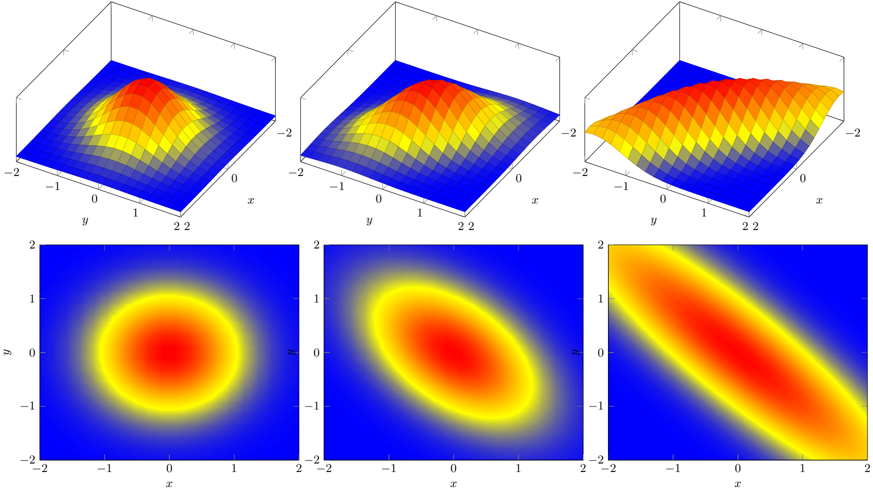


# $\mu_1 = \mu_2 = 0,\ \sigma_1 = \sigma_2 = 1$ 的二维正态分布




(a) $\rho = 0$　　(b) $\rho = 0.5$　　(c) $\rho = 0.9$

# $\mu_1 = \mu_2 = 0,\ \sigma_1 = \sigma_2 = 1$ 的二维正态分布

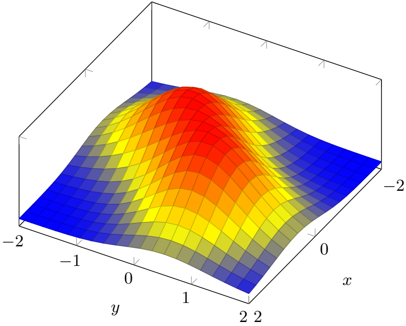


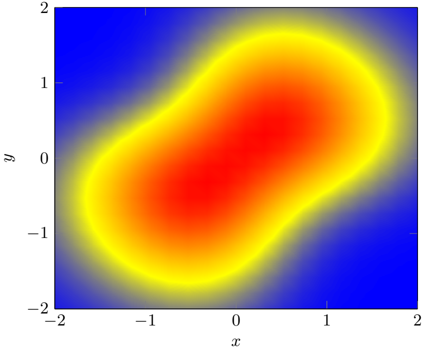

(a) $\rho = 0$ (b) $\rho = -0.5$ (c) $\rho = -0.9$

取边际密度都为 $\mathcal{N}(0,1)$ 的正态分布，利用密度为 $c(u,v)=2-2u-2v+4uv$ 的连接函数，则二维联合密度为

$$f(x,y)=\Big[2-2\Phi(x)-2\Phi(y)+4\Phi(x)\Phi(y)\Big]\cdot\varphi(x)\varphi(y).$$

**此函数具有边际正态，但不是二维正态密度.**



# 均匀分布

## 定义

设有界区域 $G$ 的面积为 $A$. 若随机变量 $(X,Y)$ 的联合密度函数为

$$f(x,y)=\begin{cases}\dfrac{1}{A}, & (x,y)\in G,\\ 0, & \text{其它}.\end{cases}$$

则称 $(X,Y)$ 服从区域 $G$ 上的**均匀分布**, 记作 $(X,Y)\sim \mathrm{U}(G)$.

## 性质

若 $(X,Y)\sim \mathrm{U}(G)$, 则对任意 $G_1\subset G$,

$$\mathsf{P}\big((X,Y)\in G_1\big)=\frac{|G_1|}{|G|}.$$

这就是二维的几何概型!

## 例题

设 $(X,Y)\sim \mathrm{U}(G)$, 其中

$$G=\{(x,y)\mid 0\le y\le x,\ 0\le x\le 1\}.$$

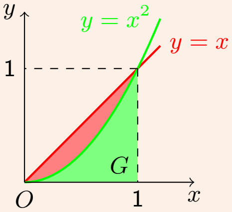
求

1. $f(x,y)$;
2. $\mathsf{P}(Y>X^2)$;
3. $(X,Y)$ 在平面上的落点到 $y$ 轴距离小于 0.3 的概率.

## 解

1. 显然 $|G|=\int_0^1 x\,dx=\frac{1}{2}$, 故 $f(x,y)=\mathbb{1}_G(x,y)\cdot 2$.

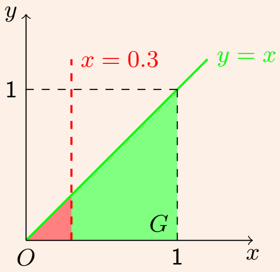
$$\mathsf{P}(Y>X^2)=\int_0^1 dx\int_{x^2}^{x} 2\,dy=\frac{1}{3}.$$




2. 

## 解

3. $\mathsf{P}(|X| < 0.3) = \mathsf{P}(-0.3 < X < 0.3) = \int_0^{0.3} dx \int_0^x 2\,dy = 0.09.$



## 例题

设随机变量 $(X,Y)$ 服从区域 $D$ 上的均匀分布, 其中
$D=\{(x,y)\mid x\geq 0, y\geq 0, x+\frac{y}{2}\leq 1\}$. 试求随机变量 $(X,Y)$ 的边际密度函数.

## 解

区域 $D$ 的面积为 $A=1$, 所以联合密度函数为 $f(x,y)=\mathbb{1}_D(x,y)$.
当 $x<0$ 或 $x>1$ 时, $f_X(x)=0$. 当 $0\leq x\leq 1$ 时,

$$f_X(x)=\int_{-\infty}^{\infty} f(x,y)\,dy=\int_0^{2(1-x)} 1\,dy=2(1-x).$$

因此, $f_X(x)=\mathbb{1}_{(0,1)}(x)\cdot 2(1-x)$.
同理, 随机变量 $Y$ 的边际密度函数为

$$f_Y(y)=\mathbb{1}_{(0,2)}(y)\int_0^{1-\frac{y}{2}} 1\,dx=\mathbb{1}_{(0,2)}(y)\cdot(1-y/2).$$

- P76: 5, 6, 7, 8
- 补充题:
  1. 设二维随机变量 $(X,Y)$ 的联合分布函数为
  $$F(x,y)=\begin{cases}k(1-e^{-x})(1-e^{-y}), & x>0,\ y>0,\\ 0, & \text{其它}.\end{cases}$$
  求边际密度函数以及 $\mathrm{P}(1<X<3,\ 1<Y<2)$.
  2. 设二维随机变量 $(X,Y)$ 的密度函数为
  $$f(x,y)=\begin{cases}x+y, & 0<x,y<1,\\ 0, & \text{其它}.\end{cases}$$
     (1) 求边际密度函数;
     (2) 求 $\mathrm{P}(X>Y)$;
     (3) 求 $\mathrm{P}(X<0.5)$.

# 概念的引入

## 事件的独立性

$$A, B \text{ 相互独立} \Leftrightarrow \ A \text{ 和 } B \text{ 之间没有任何关系}$$
$$\Leftrightarrow \mathsf{P}(AB) = \mathsf{P}(A)\mathsf{P}(B).$$

## 问

怎样定义随机变量 $X$、$Y$ 之间的独立性?

## 分析

若 $X$ 和 $Y$ 相互 "独立", 从直观上看, $X$ 和 $Y$ 取任何值之间应是没有任何关系的, 即 $\forall x, y \in \mathbb{R}^1$, 两个事件

$$\{X \le x\}, \quad \{Y \le y\}$$

应该相互独立, 即

$$\mathsf{P}(X \le x,\ Y \le y) = \mathsf{P}(X \le x) \cdot \mathsf{P}(Y \le y) \Leftrightarrow F(x, y) = F_X(x) \cdot F_Y(y).$$

# 随机变量的独立性

## 定义

设 $(X,Y) \sim F(x,y)$, $X \sim F_X(x)$, $Y \sim F_Y(y)$, 若对任意的 $x, y \in (-\infty, \infty)$, 有

$$\mathsf{P}(X \le x,\ Y \le y) = \mathsf{P}(X \le x) \cdot \mathsf{P}(Y \le y),$$

即 $F(x,y) = F_X(x) \cdot F_Y(y)$. 称随机变量 $X, Y$ **相互独立**.

## 注

两个随机变量相互独立, 联合分布函数等于两个边缘分布函数的乘积.

## 性质

若 $X, Y$ 相互独立. 对任意的 $x_1 < x_2, y_1 < y_2$, 有

$$\mathrm{P}(x_1 < X \le x_2,\ y_1 < Y \le y_2) = \mathrm{P}(x_1 < X \le x_2) \cdot \mathrm{P}(y_1 < Y \le y_2),$$

即 $\{x_1 < X \le x_2\}$ 与 $\{y_1 < Y \le y_2\}$ 相互独立.

## 证明

$$\begin{aligned}
&\mathrm{P}(x_1 < X \le x_2,\ y_1 < Y \le y_2) \\
&= F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1) \\
&= F_X(x_2)F_Y(y_2) - F_X(x_1)F_Y(y_2) - F_X(x_2)F_Y(y_1) + F_X(x_1)F_Y(y_1) \\
&= \Big[F_X(x_2) - F_X(x_1)\Big] \cdot \Big[F_Y(y_2) - F_Y(y_1)\Big] \\
&= \mathrm{P}(x_1 < X \le x_2) \cdot \mathrm{P}(y_1 < Y \le y_2).
\end{aligned}$$

## $X$ 与 $Y$ 独立的直观意义

$X$ 的取值与 $Y$ 的取值是相互独立、互不相干的.

# 独立性的定义

## 定义

设 $(X,Y)$ 的频率函数为

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

则 $X,Y$ 相互独立等价于对任意 $i,j=1,2,\ldots$ 有

$$\mathsf{P}(X=x_i,\ Y=y_j)=\mathsf{P}(X=x_i)\cdot\mathsf{P}(Y=y_j)$$

即

$$X\text{ 与 }Y\text{ 独立}\Leftrightarrow p_{ij}=p_{i\cdot}\cdot p_{\cdot j}.$$

### 例题

甲袋中有 3 个红球, 2 个白球; 乙袋中有 4 个红球, 5 个白球. 从甲、乙两袋中各任取两球, 记 $X$, $Y$ 分别表示取到白球的个数, 问 $X$, $Y$ 是否独立?

### 分析

由于从两袋中取球是相互独立的过程, 所以 $X$, $Y$ 的取值是相互独立、互不相干的, 故 $X$, $Y$ 相互独立.

### 判断随机变量的独立性的方法

- 按定义判断
- 从直观背景判断

## 例题

设随机变量 $X$ 从 1, 2, 3, 4 四个数中等可能取值. 又设随机变量 $Y$ 从 1 到 $X$ 中等可能取值. 问 $X, Y$ 是否独立?

## 解

$(X,Y)$ 的频率函数及边际频率函数为

| $Y \backslash X$ | 1 | 2 | 3 | 4 | $p_{\cdot j}=\sum_{i=1}^{4} p_{ij}$ |
|---|---|---|---|---|---|
| 1 | 1/4 | 1/8 | 1/12 | 1/16 | 25/48 |
| 2 | 0 | 1/8 | 1/12 | 1/16 | 13/48 |
| 3 | 0 | 0 | 1/12 | 1/16 | 7/48 |
| 4 | 0 | 0 | 0 | 1/16 | 3/48 |
| $p_{i\cdot}=\sum_{j=1}^{4} p_{ij}$ | 1/4 | 1/4 | 1/4 | 1/4 | |

可以直接验证, 对任意 $i,j=1,2,3,4$, $\mathrm{P}(X=i,\ Y=j)\neq \mathrm{P}(X=i)\cdot\mathrm{P}(Y=j)$. 因此 $X,Y$ 独立. **从直观上看, $X$ 和 $Y$ 也不独立.**

## 例子

设 $(X, Y)$ 的频率函数为

| $X \backslash Y$ | 1 | 2 | 3 |
|---|---|---|---|
| 1 | 1/8 | $a$ | 1/24 |
| 2 | b | 1/4 | 1/8 |

问:

1. $a, b$ 应满足什么条件?
2. 若 $X, Y$ 独立, 求 $a, b$.

## 分析

第一问是考察归一化条件, 第二问是考察独立性条件.

## 解

1. 由归一化条件, $\sum_{i,j} p_{ij} = 1 \Rightarrow a + b = 1 - (1/8 + 1/24 + 1/4 + 1/8) = \frac{11}{24}$, $a \geq 0$, $b \geq 0$.

## 解

2. 由 $X, Y$ 的独立性

$$a = \mathsf{P}(X=2,\ Y=1) = \mathsf{P}(X=2)\cdot\mathsf{P}(Y=1) = (a+1/4)(1/8+a+1/24)$$
$$b = \mathsf{P}(X=1,\ Y=2) = \mathsf{P}(X=1)\cdot\mathsf{P}(Y=2) = (b+1/8)(b+1/4+1/8).$$

求解这两个二次方程, 得到 $a = 1/12$ 或 $a = 1/2$, $b = 1/8$ 或 $b = 3/8$. 再利用归一化条件:

$$a+b = \frac{11}{24} \Rightarrow a = \frac{1}{12},\ b = \frac{3}{8}.$$

# 独立性的条件

## 定义

设 $(X,Y)$ 为连续型随机变量，且

$$(X,Y) \sim f(x,y), \quad X \sim f_X(x), \quad Y \sim f_Y(y).$$

则 $(X,Y)$ 相互独立当且仅当在 $f(x,y), f_X(x), f_Y(y)$ 的连续点处有

$$f(x,y) = f_X(x) \cdot f_Y(y).$$

## 分析

$(X,Y)$ 相互独立表明对任意的 $x, y$,

$$\mathsf{P}(X \le x,\ Y \le y) = \mathsf{P}(X \le x)\mathsf{P}(Y \le y)$$
$$\Leftrightarrow \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)\,dudv = \int_{-\infty}^{x}\int_{-\infty}^{y} f_X(u)f_Y(v)\,dudv.$$

# 独立性的推论

**定义**

$X$ 和 $Y$ 独立 $\Leftrightarrow$ 对任意 $x$ 和 $y$ 有

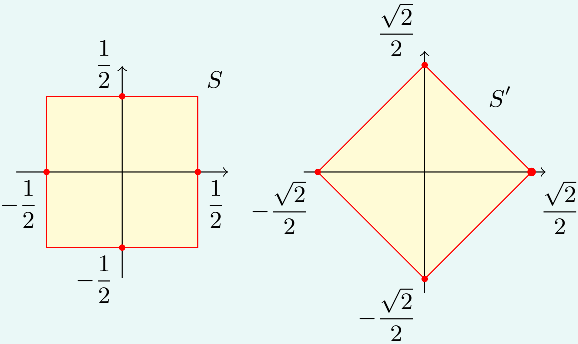
$$f(x,y)=f_X(x)\cdot f_Y(y).$$

**推论**

二维随机变量 $(X,Y)$ 相互独立, 则边缘分布完全确定联合分布.

## 例子

设 $(X,Y)$ 服从正方形 $S=\{(x,y)\mid -1/2\le x,y\le 1/2\}$ 上的均匀分布. 又设 $S'$ 为 $S$ 旋转 $45°$, $(X',Y')$ 为 $S'$ 上的均匀分布.




## 独立性验证

- 易见 $f_X(x)=\mathbb{1}_{(-1/2,1/2)}(x)$, $f_Y(y)=\mathbb{1}_{(-1/2,1/2)}(y)$, 因此 $f_{XY}(x,y)=f_X(x)f_Y(y)\Rightarrow X,Y$ 独立.
- $f_{X'}(0.5)>0$, $f_{Y'}(0.5)>0$, 但 $f_{X'Y'}(0.5,0.5)=0$, 因此 $X',Y'$ 不独立.

## 例题

设 $(X,Y)$ 的密度函数为

$$f(x,y)=\begin{cases}2e^{-(2x+y)}, & x>0,\ y>0\\ 0, & \text{其它}.\end{cases}$$

问 $X,Y$ 是否独立?

## 分析

先利用边际密度的公式计算出 $f_X(x)$, $f_Y(y)$, 再验证独立性条件.

## 解

$X,Y$ 的边际密度函数分别为

$$f_X(x)=\mathbb{1}_{(0,\infty)}(x)\int_{-\infty}^{\infty}2e^{-(2x+y)}\,dy=\mathbb{1}_{(0,\infty)}(x)2e^{-2x},$$

$$f_Y(y)=\mathbb{1}_{(0,\infty)}(y)=\int_{-\infty}^{\infty}2e^{-(2x+y)}\,dx=\mathbb{1}_{(0,\infty)}(y)e^{-y}.$$

因此 $f_{XY}(x,y)=f_X(x)f_Y(y)$, 故 $X,Y$ 相互独立.