3.4 边缘分布

上节将二维随机变量视为一个整体，讨论$F(x, y)$，

但X、Y也是一个随机变量，他们各自的分布函数如何？

定义

X、Y的分布函数记为$F_X(x)$和$F_Y(y)$称为**边缘分布函数**

$$\begin{cases} F_X(x) = F(x, +\infty) \\ F_Y(y) = F(+\infty, y) \end{cases}$$

因为 $F_X(x) = P(X \le x) = P(X \le x, Y \le +\infty)$

令联合分布函数 $F(x, y)$中的 $y \to +\infty$ 得到$F_X(x)$

二维离散型随机变量

(X、Y) 联合分布律为 $P(X=x_i, Y=y_j)=p_{ij},\ i,j=1,2,\cdots$

X、Y 的边缘分布函数为

$$F_X(x)=F(x,+\infty)=\sum_{x_i\le x}\sum_{j=1}^{\infty}p_{ij}$$

$$F_Y(y)=F(+\infty,y)=\sum_{y_j\le y}\sum_{i=1}^{\infty}p_{ij}$$

因为X的分布律 $P(X=x_i)=P(X=x_i,Y<+\infty)=\sum_j p_{ij}\overset{\text{记为}}{=\!=}p_{i\bullet}\quad i=1,2,\cdots$

因为Y的分布律 $P(Y=y_j)=P(X<+\infty,Y=y_j)=\sum_i p_{ij}\overset{\text{记为}}{=\!=}p_{\bullet j}\quad j=1,2,\cdots$

由联合分布律的表格写出边缘分布律

$X \backslash Y$	y_1	y_2	…	y_j	…	$P(X=x_i)$
x_1	p_{11}	p_{12}	…	p_{1j}	…	$p_{1\bullet}$
x_2	p_{21}	p_{22}	…	p_{2j}	…	$p_{2\bullet}$
⋮	⋮	⋮		⋮		⋮
x_i	p_{i1}	p_{i2}	…	p_{ij}	…	$p_{i\bullet}$
⋮	⋮	⋮		⋮		⋮
$P(Y=y_j)$	$p_{\bullet 1}$	$p_{\bullet 2}$	…	$p_{\bullet j}$	…	1

记号$p_{i\cdot}$中·表示$p_{i\cdot}$是由p_{ij}关于j求和后得到的；同样$p_{\cdot j}$是由p_{ij}关于i求和后得到的

例 对一群体的吸烟及健康状况进行调查，引入随机变量X和Y如下：

$$X=\begin{cases}0, & \text{健康}\\ 1, & \text{一般}\\ 2, & \text{不健康}\end{cases}\qquad Y=\begin{cases}0, & \text{不吸烟}\\ 10, & \text{一天吸烟不多于15支}\\ 20, & \text{一天吸烟多于15支}\end{cases}$$

根据调查结果，得(X, Y)的如下的联合分布律：

X \ Y	0	10	20
0	0.35	0.04	0.025
1	0.025	0.15	0.04
2	0.020	0.10	0.25

(i) 关于X和Y的边缘分布律；

(ii) 求$P(X=2|Y=20)$

解 (i)

X	0	1	2
P	0.415	0.215	0.370

Y	0	10	20
P	0.395	0.290	0.315

(ii) $$P(X=2|Y=20)=\frac{P(X=2, Y=20)}{P(Y=20)}=\frac{0.25}{0.315}=0.794$$

例 已知(X, Y)的联合分布律为

$X \backslash Y$	-1	0	1
1	0.1	a	0.2
2	0.1	0.2	b

已知$P(Y=1|X=1)=0.5$，求:

(i) a、b的值；

(ii) X、Y的边缘分布律；(iii) 求$P(X=1|Y=1)$。

解 (i) 由分布律性质知 $a+b+0.6=1$ 即$a+b=0.4$

$$P(Y=1|X=1)=\frac{0.2}{0.3+a} \Rightarrow \frac{0.2}{0.3+a}=\frac{1}{2} \Rightarrow a=0.1,\ b=0.3$$

(ii)

X	1	2
$p_{i\cdot}$	0.4	0.6

Y	-1	0	1
$p_{\cdot j}$	0.2	0.3	0.5

(iii) $$P(X=1|Y=1)=\frac{P(X=1,Y=1)}{P(Y=1)}=\frac{0.2}{0.5}=0.4$$

二维连续型随机变量

$(X、Y)$ 联合概率密度为$f(x,y)$，联合分布函数为 $F(x,y)=\int_{-\infty}^{y}\int_{-\infty}^{x} f(u,v)dudv$

X，Y 的边缘概率密度为

$$f_X(x)=\int_{-\infty}^{+\infty} f(x,y)dy \qquad f_Y(y)=\int_{-\infty}^{+\infty} f(x,y)dx$$

因为$F_X(x)=F(x,+\infty)=\int_{-\infty}^{x}\left[\int_{-\infty}^{+\infty} f(t,y)dy\right]dt=\int_{-\infty}^{x} f_X(t)dt$

因为$F_Y(y)=F(+\infty,y)=\int_{-\infty}^{y}\left[\int_{-\infty}^{+\infty} f(x,t)dx\right]dt=\int_{-\infty}^{y} f_Y(t)dt$

$f_X(x)$和$f_Y(y)$是(X,Y)关于X和关于Y的边缘概率密度

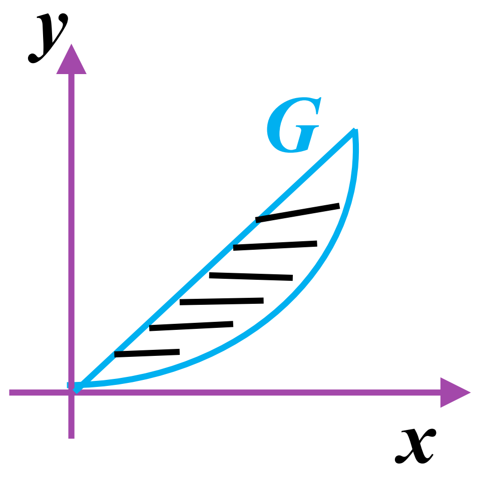
例 设G是平面上的有界区域，其面积为A，若二维随机变量(X, Y)具有联合概率密度

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

则称(X, Y)在G上服从均匀分布。现设(X, Y)在有界区域$x^2\le y\le x$上服从均匀分布，其联合概率密度为

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

求边缘概率密度$f_X(x)$和$f_Y(y)$



解

$$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)dy=\begin{cases}\int_{x^2}^{x}6dy=6(x-x^2), & 0\le x\le 1\\ 0, & \text{其他}\end{cases}$$

$$f_Y(y)=\int_{-\infty}^{+\infty}f(x,y)dx=\begin{cases}\int_{y}^{\sqrt{y}}6dx=6(\sqrt{y}-y), & 0\le y\le 1\\ 0, & \text{其他}\end{cases}$$

二维均匀分布$(X, Y)\sim U(G)$：
$(X, Y)\sim$在G的任一子区域取值的概率等价于平面区域G的几何概率

例 若二维随机变量(X, Y)具有联合概率密度

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

其中μ_1、μ_2、σ_1、σ_2都是常数，且$\sigma_1>0$、$\sigma_2>0$、$-1<\rho<1$。称(X, Y)服从参数为μ_1、μ_2、σ_1、σ_2、ρ的二维正态分布，记为$(X,Y)\sim N(\mu_1,\mu_2;\sigma_1^2,\sigma_2^2;\rho)$，求二维正态分布的边缘概率密度。

解

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

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

$$=\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x-\mu_1)^2}{2\sigma_1{}^2}}\qquad-\infty<x<+\infty$$

同理 $$f_Y(y)=\frac{1}{\sqrt{2\pi}\sigma_2}e^{-\frac{(y-\mu_2)^2}{2\sigma_2{}^2}}\qquad-\infty<y<+\infty$$

即二维正态分布的两个边缘分布都是一维正态分布，且都不依赖于参数ρ；反推不成立

如果只知道关于X和关于Y的边缘概率分布，一般不能推导出X和Y的联合概率分布

本节回顾

□ 边缘分布函数

X、Y 的分布函数记为 $F_X(x)$ 和 $F_Y(y)$ 称为**边缘分布函数**
$$\begin{cases} F_X(x) = F(x, +\infty) \\ F_Y(y) = F(+\infty, y) \end{cases}$$

□ 二维离散型随机变量的边缘分布律

$(X$、$Y)$ 联合分布律为 $P(X = x_i, Y = y_j) = p_{ij},\ i, j = 1, 2, \cdots$

$$P(X = x_i) = P(X = x_i, Y < +\infty) = \sum_j p_{ij} \overset{\text{记为}}{=\!=} p_{i\bullet} \qquad i = 1, 2, \cdots$$

$$P(Y = y_j) = P(X < +\infty, Y = y_j) = \sum_i p_{ij} \overset{\text{记为}}{=\!=} p_{\bullet j} \qquad j = 1, 2, \cdots$$

□ 二维连续型随机变量的边缘概率密度

$(X$、$Y)$ 联合概率密度为 $f(x, y)$，联合分布函数为 $F(x, y) = \int_{-\infty}^{y}\int_{-\infty}^{x} f(u, v)\,du\,dv$

X、Y 的边缘概率密度为 $f_X(x) = \int_{-\infty}^{+\infty} f(x, y)\,dy$ 　 $f_Y(y) = \int_{-\infty}^{+\infty} f(x, y)\,dx$