§3.2　边际分布

一、　边际分布函数

定义1 设$X_1,X_2,\cdots,X_n$是定义在$(\Omega,\mathbb{F},P)$上的n个随机变量，它们之中任意k个$X_{i_1},X_{i_2},\cdots,X_{i_k}$随机变量也构成$k$随机向量，$k=1,2,\cdots,n-1$. 我们把这样的$k$维随机向量称为$n$维随机向量$(X_1,X_2,\cdots,X_n)$的$k$维边际随机向量，共有$C_n^k$个.每个$k$维边际随机向量都有自己的分布函数，这种分布函数称为k维边际分布，特别$k=1$时，有n个不同的一维边际分布.

本节主要讨论二维随机变量的边缘分布问题.

二、 二维随机变量的边缘分布函数

已知(X,Y)的分布函数，如何求X,Y的分布函数？

$$F(x,y)=P\{X\le x,\ Y\le y\}\xrightarrow{?\ ?}\begin{array}{l}F_X(x)=P\{X\le x\}\\ F_Y(x)=P\{Y\le y\},\end{array}$$

$$P\{X\le x\}=P\{X\le x,\ Y<\infty\}=F(x,\infty)=F_X(x).$$

$$\Updownarrow$$

(X,Y)关于X的边缘分布函数.

$$P\{Y\le y\}=P\{X<\infty,\ Y\le y\}=F(\infty,y)=F_Y(y).$$

$$\Updownarrow$$

(X,Y)关于Y的边缘分布函数..

三、 二维离散型随机变量的边缘分布

设二维离散型随机变量(X,Y)的概率分布为

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

则X,Y也是离散型随机变量.

X的取值为 $x_1,\ x_2,\ \cdots\cdots,\ x_m,\cdots\cdots$

相应的概率为

$$P(X=x_i)=\sum_j p_{ij}\triangleq p_{i\cdot},\quad i=1,2,\cdots$$

Y的取值为 $y_1,\ y_2,\ \cdots\cdots,\ y_n,\cdots\cdots$

相应的概率为

$$P(Y=y_j)=\sum_i p_{ij}\triangleq p_{\cdot j},\quad j=1,2,\cdots$$

分别称以上两式为二维离散型随机变量(X,Y)的关于X和Y的边缘概率分布.

联合分布与边际分布表

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

【例1】袋中装有2只白球和3只黑球．现在进行有放回和不放回的两种摸球，每次摸一球，设X、Y分别表示第一、二次摸出的白球数，求(X,Y)的联合分布和边缘分布.

解：可能取值为 (0,0)，(0,1)，(1,0)，(1,1).

对有放回摸球，可求得

$$P\{X=0,Y=0\}=\frac{3}{5}\cdot\frac{3}{5},$$

$$P\{X=1,Y=0\}=\frac{2}{5}\cdot\frac{3}{5}.\quad \text{等}$$

有放回摸球的概率分布

$X \backslash Y$	0	1	$p_{i\cdot}$
0	$\frac{3}{5}\cdot\frac{3}{5}$	$\frac{2}{5}\cdot\frac{3}{5}$	$3/5$
1	$\frac{3}{5}\cdot\frac{2}{5}$	$\frac{2}{5}\cdot\frac{2}{5}$	$2/5$
$p_{\cdot j}$	$3/5$	$2/5$	

不放回摸球的概率分布

$X \backslash Y$	0	1	$p_{i\cdot}$
0	$\frac{3}{5}\cdot\frac{2}{4}$	$\frac{2}{5}\cdot\frac{3}{4}$	$3/5$
1	$\frac{3}{5}\cdot\frac{2}{4}$	$\frac{2}{5}\cdot\frac{1}{4}$	$2/5$
$p_{\cdot j}$	$3/5$	$2/5$	

注意

联合分布 $\underset{\not\longleftarrow}{\longrightarrow}$ **边缘分布**

四、 二维连续型随机变量的边缘分布

定义2 设二维连续型随机变量(X,Y)的分布函数与密度函数分别为$F(x,y)$与$f(x,y)$，由于

$$F_X(x)=F(x,\infty)=\int_{-\infty}^{x}\left[\int_{-\infty}^{\infty}f(u,y)\,\mathrm{d}y\right]\mathrm{d}u,$$

$$F_Y(y)=F(\infty,y)=\int_{-\infty}^{y}\left[\int_{-\infty}^{\infty}f(x,v)\,\mathrm{d}x\right]\mathrm{d}v,$$

因此，X和Y也是连续型随机变量. 记

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

$$f_Y(y)=\int_{-\infty}^{\infty}f(x,y)\,\mathrm{d}x,$$

分别称$f_X(x)$, $f_Y(y)$为二维连续型随机变量(X,Y)关于X和Y的边缘概率密度函数, 简称边际密度函数.

【例2】 设随机变量X和Y具有联合概率密度

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

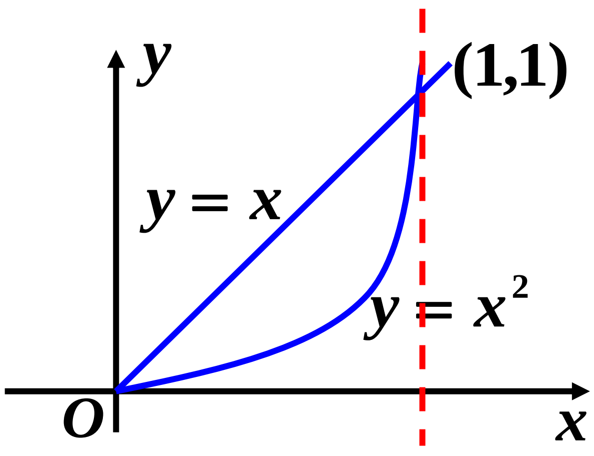
求边缘概率密度$f_X(x), f_Y(y)$.

解: $f_X(x)=\int_{-\infty}^{+\infty} f(x,y)\,\mathrm{d}y.$

当 $0\le x\le 1$ 时,

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

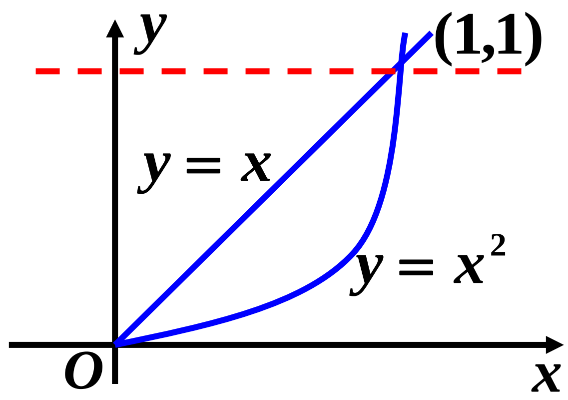
$$=\int_{x^2}^{x} 6\,\mathrm{d}y=6(x-x^2).$$



当 $x<0$ 或 $x>1$时,

$$f_X(x)=\int_{-\infty}^{+\infty} f(x,y)\,\mathrm{d}y=0.$$

得 $f_X(x)=\begin{cases}6(x-x^2), & 0\le x\le 1,\\ 0, & \text{其他}.\end{cases}$



当 $0\le y\le 1$ 时,

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

$$=\int_{y}^{\sqrt{y}} 6\,\mathrm{d}x=6(\sqrt{y}-y).$$

当 $y<0$ 或 $y>1$时,

$$f_Y(y)=\int_{-\infty}^{+\infty} f(x,y)\,\mathrm{d}x=0.$$

得 $f_Y(y)=\begin{cases}6(\sqrt{y}-y), & 0\le y\le 1,\\ 0, & \text{其他}.\end{cases}$

【例3】设二维随机变量(X,Y)的概率密度为

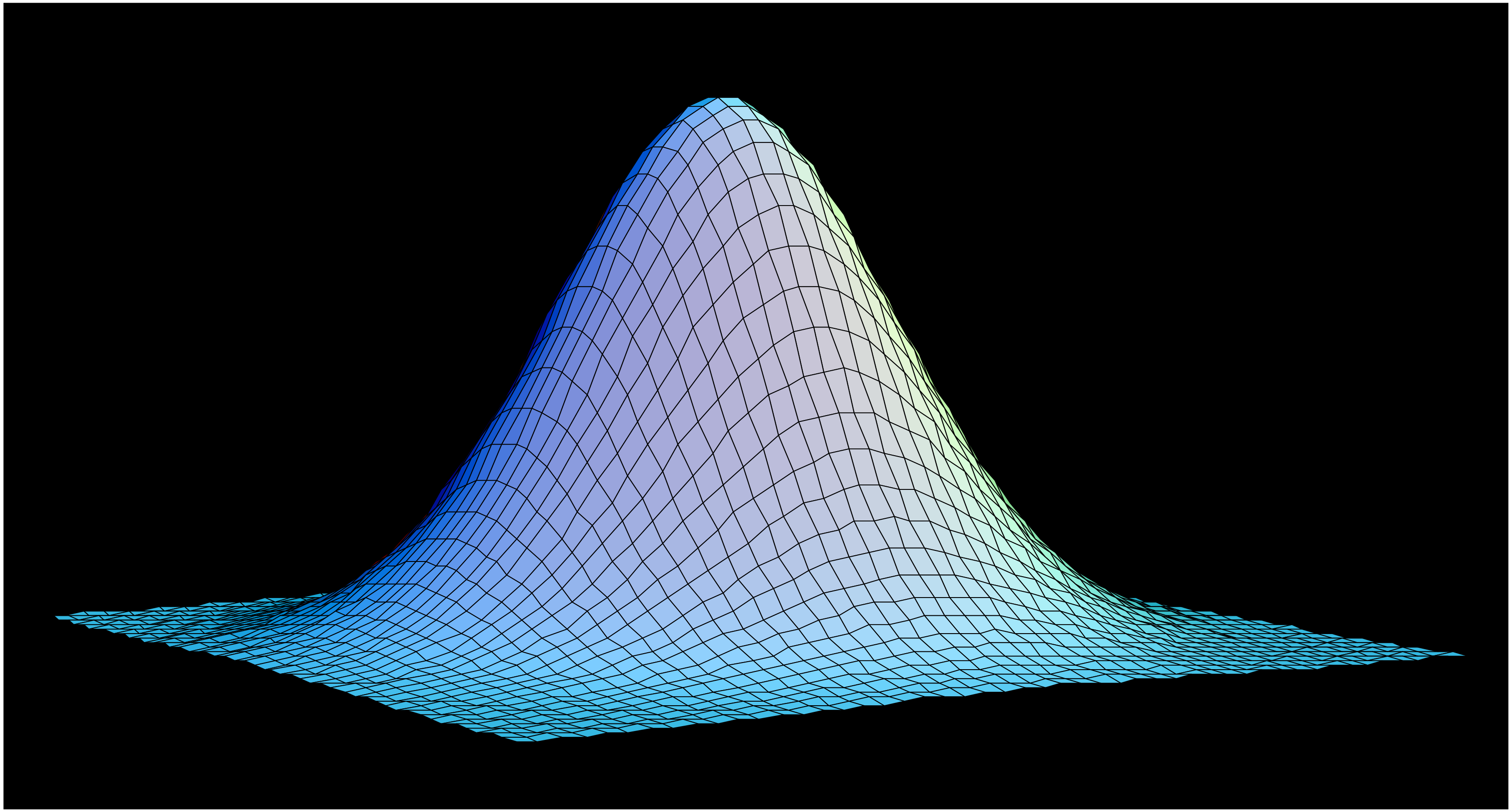
$$f(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}$$

$$\cdot\exp\left\{\frac{-1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2}-2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}\right]\right\}$$

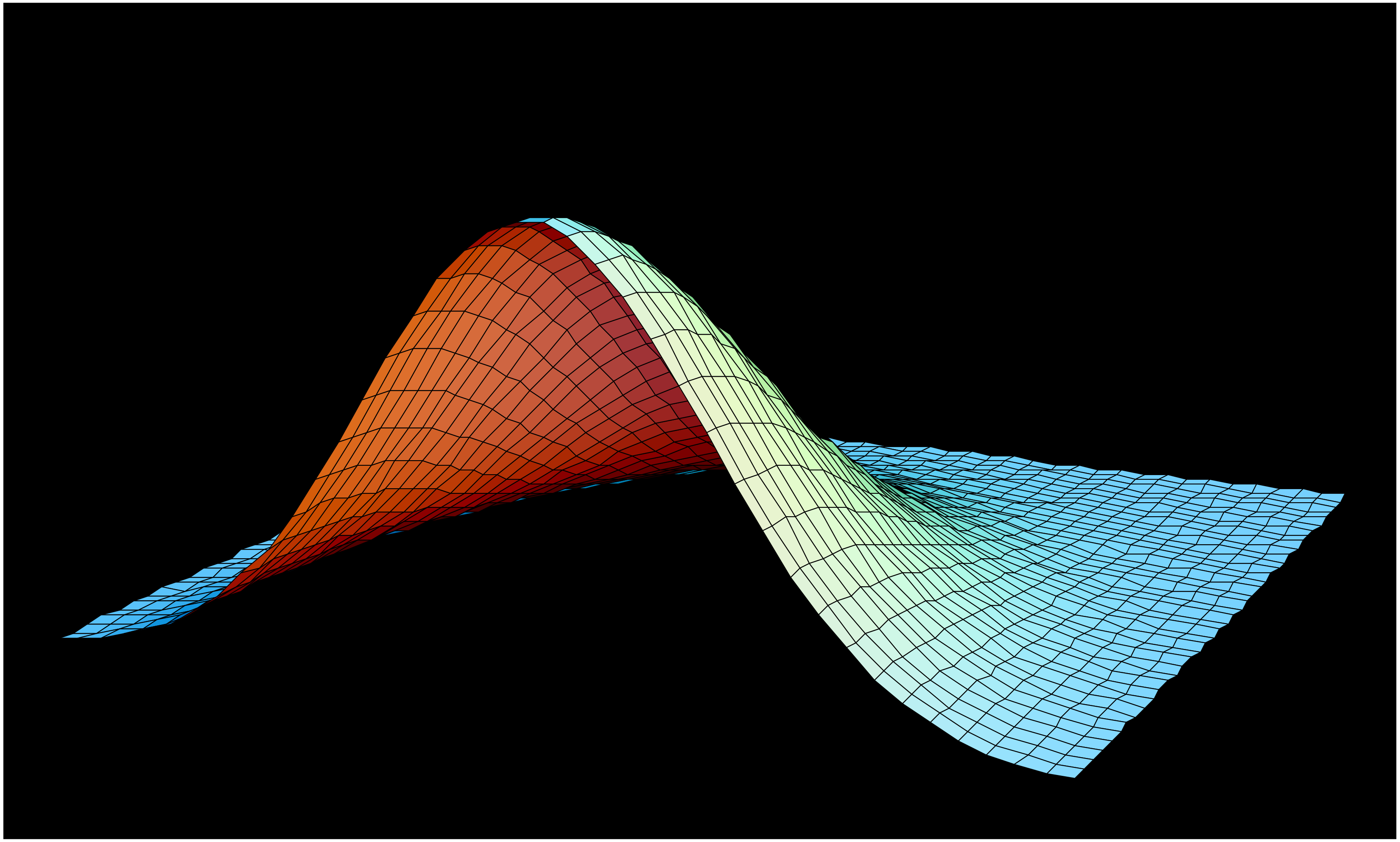
$$-\infty<x<+\infty,\quad -\infty<y<+\infty,$$

其中 $\mu_1,\mu_2,\sigma_1,\sigma_2,\rho$ 都是常数,且$\sigma_1>0,\ \sigma_2>0,$ $-1<\rho<1.$

试求二维正态随机变量的边缘概率密度.



二元正态分布图



二元正态分布剖面图

解： 二元正态分布的密度函数有下列两种分解式

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

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

利用上述结论不难求得二元正态分布的边际分布

$$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)dy=\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}},$$

$$f_Y(y)=\int_{-\infty}^{+\infty}f(x,y)dx=\frac{1}{\sqrt{2\pi}\sigma_2}e^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}}.$$

因此，二维正态分布的两个边缘分布都是一维正态分布，并且不依赖于参数ρ.

注:(1)n维正态分布的k维边缘分布也是正态分布.

(2)边缘分布均为正态分布的随机变量，其联合分布不一定是二维正态分布.

例如：(X,Y)的联合密度函数为

$$f(x,y)=\frac{1}{2\pi}\mathrm{e}^{-\frac{x^2+y^2}{2}}(1+\sin x\sin y),$$

显然(X,Y)不服从正态分布，但是

$$f_X(x)=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}},\quad f_Y(y)=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{y^2}{2}}.$$

因此边缘分布均为正态分布的连续型随机向量，其联合分布不一定是二维正态分布.

作业题 P82

习题8，9，10，11