第三章 联合分布

二维连续型随机变量

设 r.v (X,Y)的联合分布函数(joint CDF)为

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

若存在非负可积函数 $f(x, y) \ge 0$使得

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

则称(X,Y)为二维连续型r.v，$f(x, y)$称为**概率密度函数（密度函数、密度）**，或称为 X,Y的**联合概率密度(joint pdf)**.

由高等数学知：

$F(x, y)$ 是连续函数

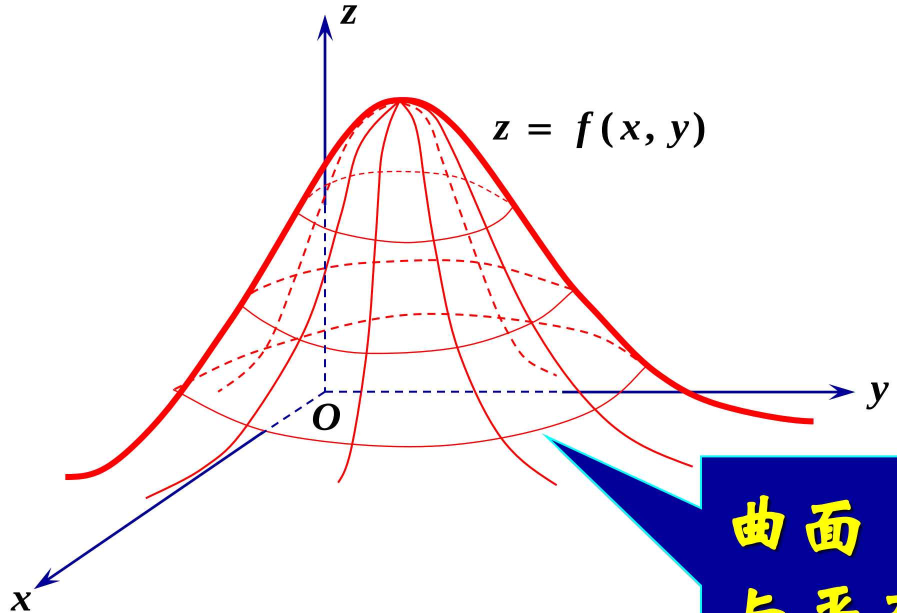
密度函数的基本性质

① $f(x,y)\geq 0 \quad (\forall (x,y)\in R^2)$

② $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(u,v)dudv = 1$

密度函数的本质特征

几何意义



曲面 $z=f(x,y)$

与平面 xOy 围成的

“山丘”的体积为 1

密度函数的基本性质

① $f(x,y)\geq 0 \quad (\forall (x,y)\in R^2)$

② $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(u,v)dudv = 1$

密度函数的本质特征

③ $\forall\ D \subset R^2$

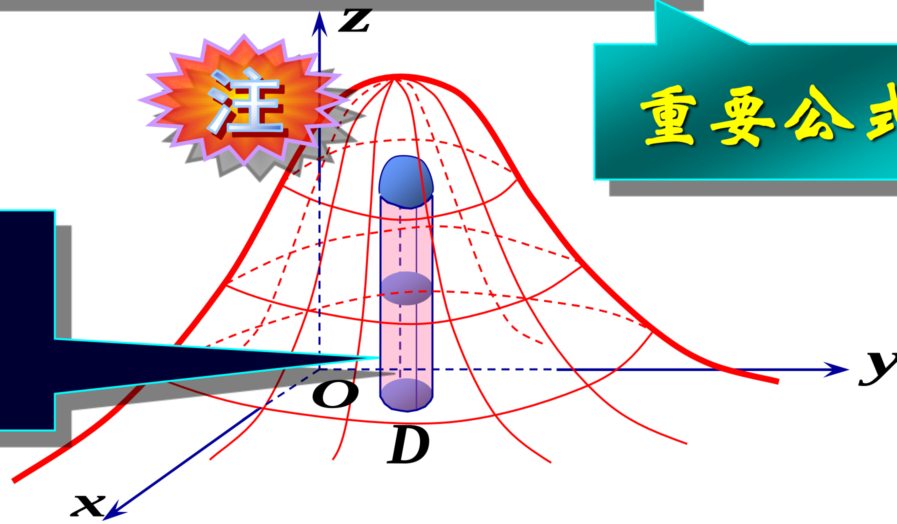
由逐段光滑曲线围成的平面区域

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

重要公式

注

$P\{(X,Y)\in D\} =$ 曲顶柱体体积



密度函数的基本性质

① $f(x, y) \geq 0 \quad (\forall (x, y) \in R^2)$

② $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(u,v)dudv = 1$

密度函数的本质特征

③ $\forall\ D \subset R^2$

$$P\{(X,Y) \in D\} = \iint\limits_{D} f(x, y)dxdy$$

④ 在 $f(x, y)$ 的连续点处，有

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

由性质(4)，在 $f(x,y)$ 的连续点处，有

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

$$=\lim_{\substack{\Delta x\to 0^+\\ \Delta y\to 0^+}} \frac{F(x+\Delta x,y+\Delta y)-F(x+\Delta x,y)-F(x,y+\Delta y)+F(x,y)}{\Delta x\,\Delta y}$$

$$=\lim_{\substack{\Delta x\to 0^+\\ \Delta y\to 0^+}} \frac{P\{x<X\le x+\Delta x,y<Y\le y+\Delta y\}}{\Delta x\,\Delta y}$$

故当 $\Delta x\Delta y$充分小时，有

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

例 设 r.v (X,Y)的概率密度为

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

① 确定常数 k;

② 求分布函数 $F(x,y)$;

③ 计算概率 $P\{Y \le X\}$.

解 ① $\because\ 1=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(u,v)dudv$

$$=k\int_0^{+\infty}\int_0^{+\infty} e^{-(2x+y)}dxdy$$

$$=k\int_0^{+\infty} e^{-2x}dx\cdot\int_0^{+\infty} e^{-y}dy$$

$$=\frac{k}{2}$$

$\therefore\ k=2$

例 设 r.v (X,Y)的概率密度为

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

① 确定常数 k;

② 求分布函数 $F(x,y)$;

③ 计算概率 $P\{Y \le X\}$.

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

$$=\begin{cases} \int_0^x\int_0^y 2e^{-(2u+v)}dudv, & x>0, y>0 \\ 0, & \text{其它} \end{cases}$$

$$=\begin{cases} 2\int_0^x e^{-2u}du\cdot\int_0^y e^{-v}dv, & x>0, y>0 \\ 0, & \text{其它} \end{cases}$$

$$=\begin{cases} (1-e^{-2x})(1-e^{-y}), & x>0, y>0 \\ 0, & \text{其它} \end{cases}$$

例 设 r.v (X,Y) 的概率密度为

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

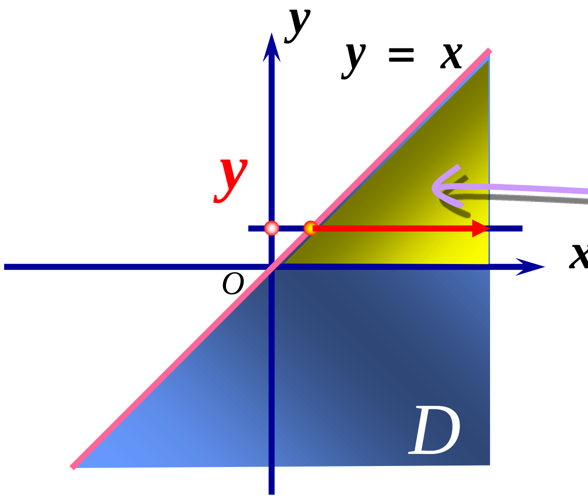
① 确定常数 k;

② 求分布函数 $F(x,y)$;

③ 计算概率 $P\{Y\le X\}$.

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

$$\therefore \quad P\{Y\le X\}=P\{(X,Y)\in D\}$$



$$=\iint_D f(x,y)dxdy$$

$$=\iint_{D\cap\{x>0,\ y>0\}} 2e^{-(2x+y)}dxdy$$

$$=\int_0^{+\infty}dy\int_y^{+\infty}2e^{-(2x+y)}dx$$

$$=\frac{1}{3}$$

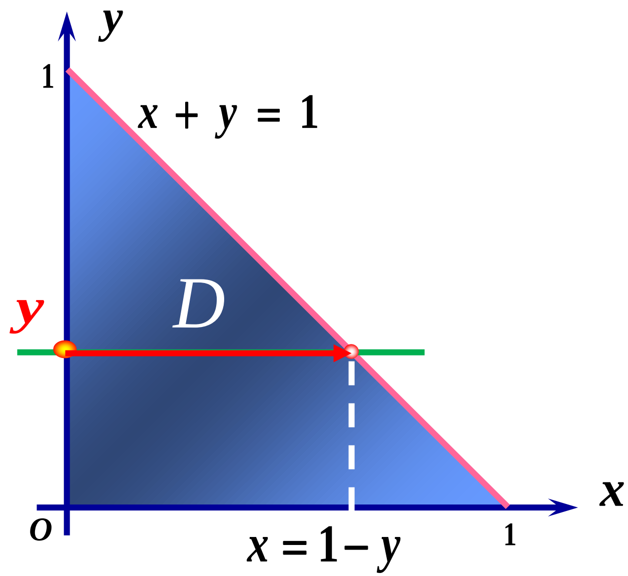
例 设 r.v (X,Y) 的概率密度为

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

计算概率 $P\{X+Y\le 1\}$.

解

$$P\{X+Y\le 1\}=\iint\limits_{x+y\le 1} f(x,y)\,dxdy$$



$$=\iint\limits_{D:\begin{cases}x+y\le 1\\ x>0,y>0\end{cases}} 2e^{-(2x+y)}dxdy$$

$$=\int_0^1 dy\int_0^{1-y} 2e^{-(2x+y)}dx$$

$$=1-2e^{-1}+e^{-2}$$

几个约定：

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

表示二维 r.v (X,Y) 的分布函数为 $F(x,y)$

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

表示二维 r.v (X,Y) 的概率函数为 $f(x,y)$

对于离散型 r.v 它表示频率函数(PMF),
对于连续型 r.v 它表示密度函数(PDF).

对于 n 维r.v也采用相应的记号：

$$(X_1, X_2, \cdots, X_n) \sim F(x_1, x_2, \cdots, x_n)$$

$$(X_1, X_2, \cdots, X_n) \sim f(x_1, x_2, \cdots, x_n)$$

二维连续型随机变量的边际分布密度

设 (X,Y) 的分布函数和密度函数分别为

$$F(x,y),\ f(x,y)$$

则 r.v X 的分布函数为

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

$$=\int_{-\infty}^{x}\left(\int_{-\infty}^{+\infty} f(u,y)dy\right)du$$

故 r.v X 的密度函数为

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

同理 Y 的分布函数为

$$F_Y(y)=\int_{-\infty}^{y}\left(\int_{-\infty}^{+\infty} f(x,v)dx\right)dv$$

Y 的密度函数为

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

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

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

二维连续型随机变量的边缘分布

$$F_X(x)=\int_{-\infty}^{x}\int_{-\infty}^{+\infty}f(u,v)dvdu$$

$$F_Y(y)=\int_{-\infty}^{y}\int_{-\infty}^{+\infty}f(u,v)dudv$$

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

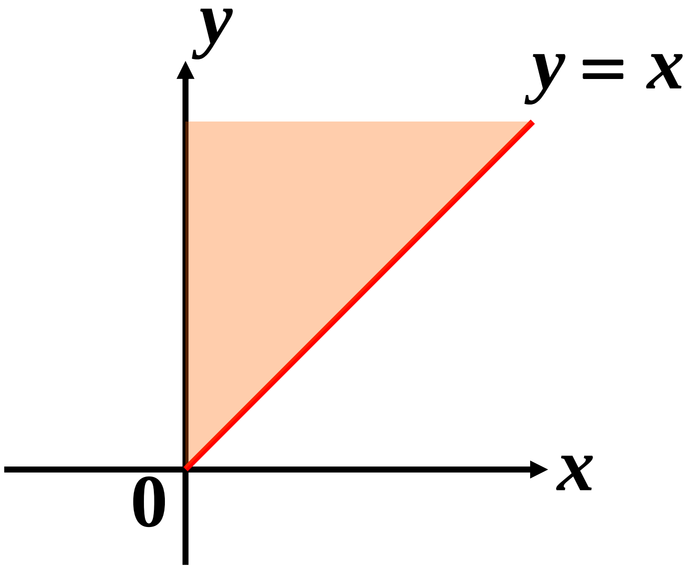
$$f_Y(y)=\int_{-\infty}^{+\infty}f(x,y)dx$$

已知联合密度可以求得边缘密度

例 设 (X,Y) 的概率密度是

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

求(X,Y)关于 X 和 Y 的边际概率密度.



x暂时固定

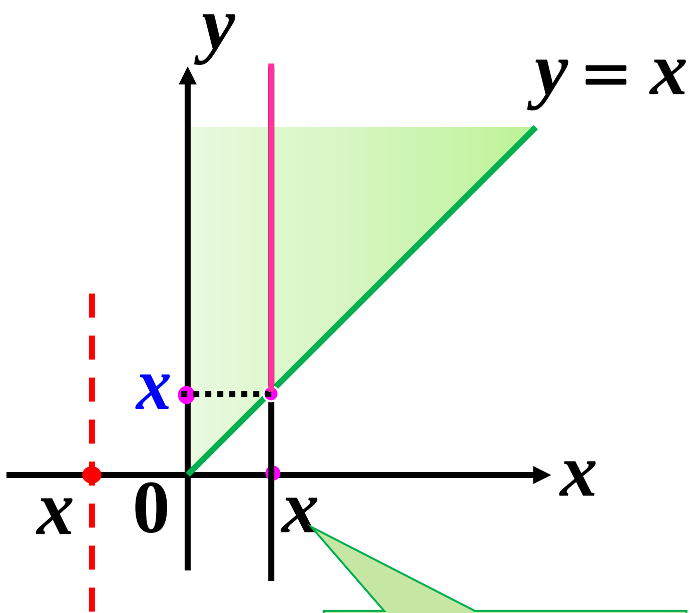
解 $f_X(x)=\int_{-\infty}^{+\infty} f(x,y)\,dy$

当 $x\leq 0$ 时，$f_X(x)=\int_{-\infty}^{+\infty} 0\,dy=0$

当 $x>0$ 时，

$$f_X(x)=\int_{x}^{+\infty} e^{-y}\,dy$$

$$=e^{-y}\Big|_{+\infty}^{x}=e^{-x}$$



故

$$f_X(x)=\begin{cases} e^{-x}, & x>0, \\ 0, & x\leq 0. \end{cases}$$

y暂时固定

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

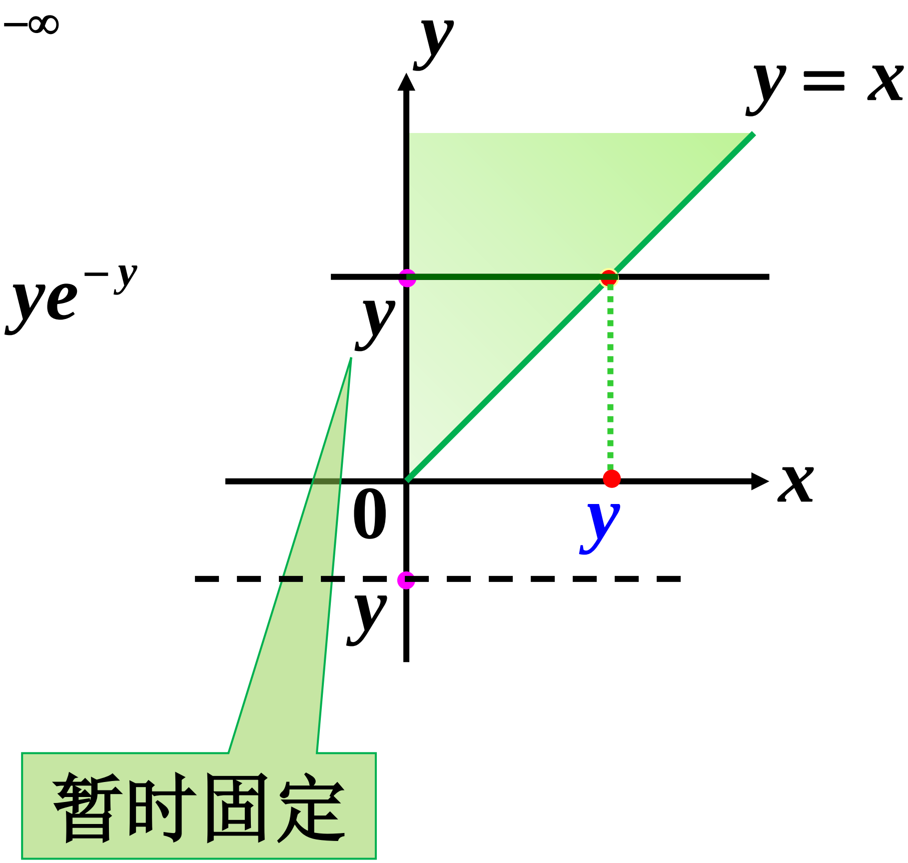
当 $y\le 0$ 时，$f_Y(y)=\int_{-\infty}^{+\infty} 0\,dx=0$

当 $y>0$ 时，

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

故

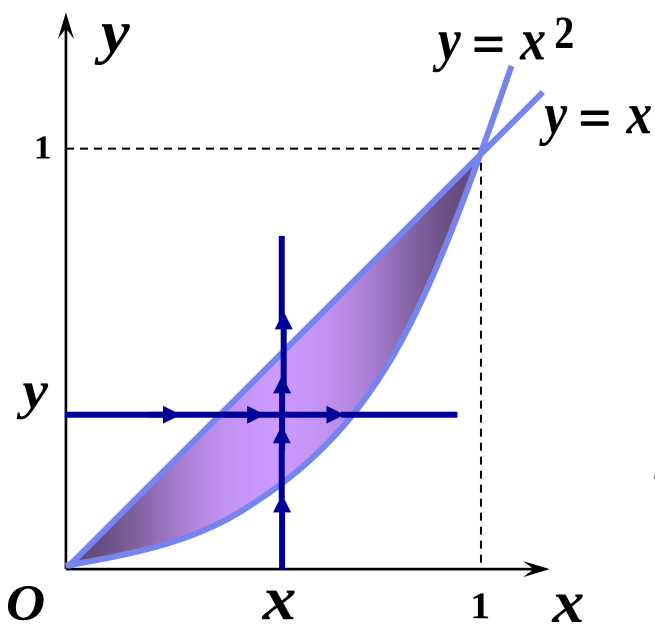
$$f_Y(y)=\begin{cases} ye^{-y}, & y>0, \\ 0, & y\le 0. \end{cases}$$



例 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 (如图)



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

$$=\begin{cases}\int_{x^2}^{x} 6dy, & 0\le x\le 1\\ 0, & \text{其它}\end{cases}$$

$$=\begin{cases}6x(1-x), & 0\le x\le 1\\ 0, & \text{其它}\end{cases}$$

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

$$=\begin{cases}\int_{y}^{\sqrt{y}} 6dx=6(\sqrt{y}-y), & 0\le y\le 1\\ 0, & \text{其它}\end{cases}$$

n维连续型随机变量的边际密度函数

设 X, Y, Z 的联合密度函数为

$$f(x, y, z)$$

则r.v X 的一维边际密度函数是

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

r.v X 和 Y 的二维边际密度函数是

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

例 (Farlie-Morgenstern族)

设 $F(x)$和$G(y)$都是一维连续型分布函数, 可以证明, 对于任意的 α, 只要满足 $|\alpha|\le 1$, 就有

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

是二元连续型分布函数.

其边际分布为

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

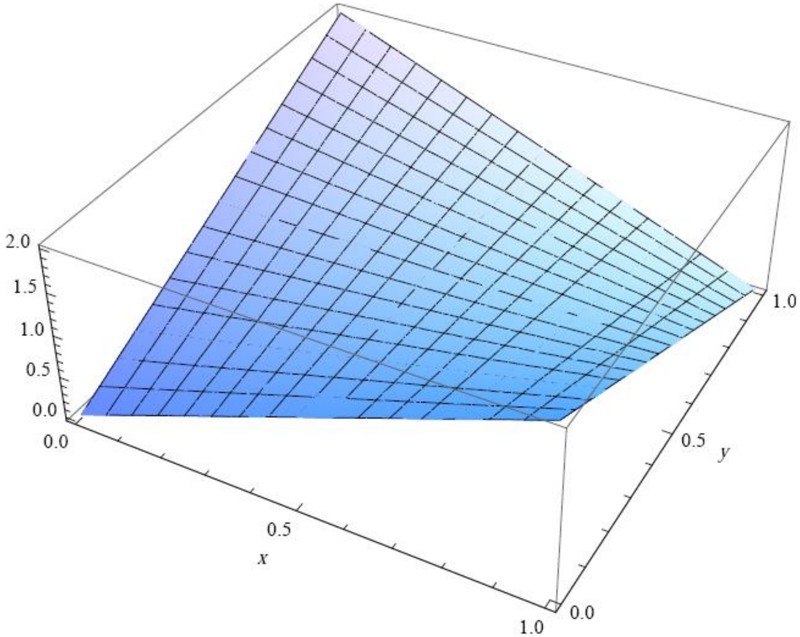
按这种方式, 可以构造给定边际分布的无数个不同的二维联合分布.

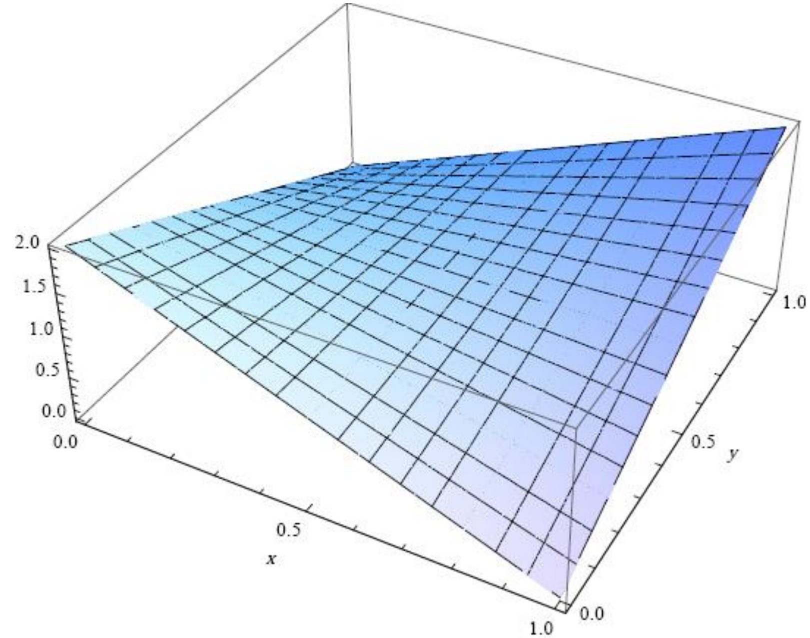
$$H(x,y)=F(x)G(y)\{1+\alpha[1-F(x)][1-G(y)]\}$$

例如, 取$F(x)$和$G(y)$都是[0,1]上的均匀分布.

此时, $F(x) = x,\ \ 0\le x \le 1;\ \ G(y) = y,\ 0\le y \le 1.$

取不同的 α:





$\alpha = -1$

$$h(x,y)=2x+2y-4xy$$

$\alpha = 1$

$$h(x,y)=2-2x-2y+4xy$$

$0\le x \le 1;\ 0\le y \le 1.$

连接函数(copula)：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 $C(u, v)$.

具有性质：

1、$C(u, v)$ 关于每个变量都是非降的;

2、 $P\{U \le u\} = C(u,1) = u, \quad C(1,v) = v.$

3、讨论具有密度函数的连接函数, 此时

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

连接函数(copula)：边际分布为均匀分布的联合累积分布函数, 记为 $C(u, v)$.

具有性质:

4、假设 X 和 Y 是分布函数分别为$F_X(x)$ 和 $F_Y(y)$的连续r.v. , 则 $U = F_X(X)$ 和 $V = F_Y(Y)$ 是均匀分布r.v.

对于连接函数$C(u, v)$, 考虑定义联合分布

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

则其边际分布为$F_X(x)$ 和 $F_Y(y)$,

相应的密度为 $f_{XY}(x, y) = c(F_X(x), F_Y(y)) f_X(x) f_Y(y)$

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

二维正态分布

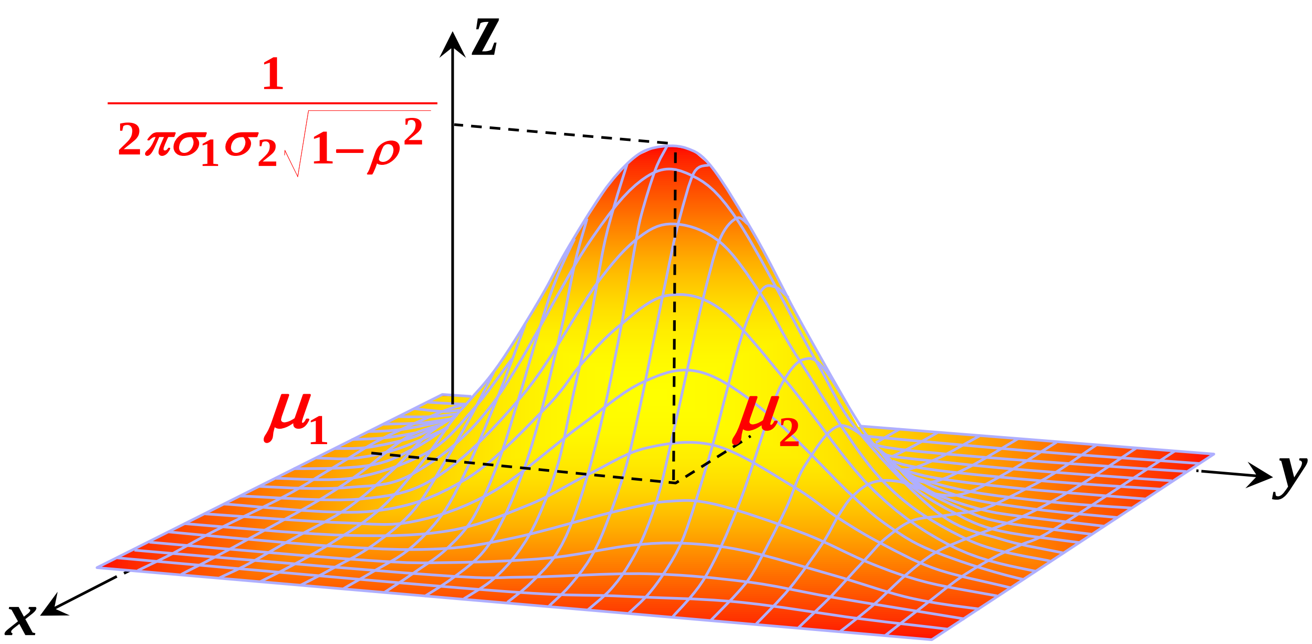
若 X,Y的联合密度为

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

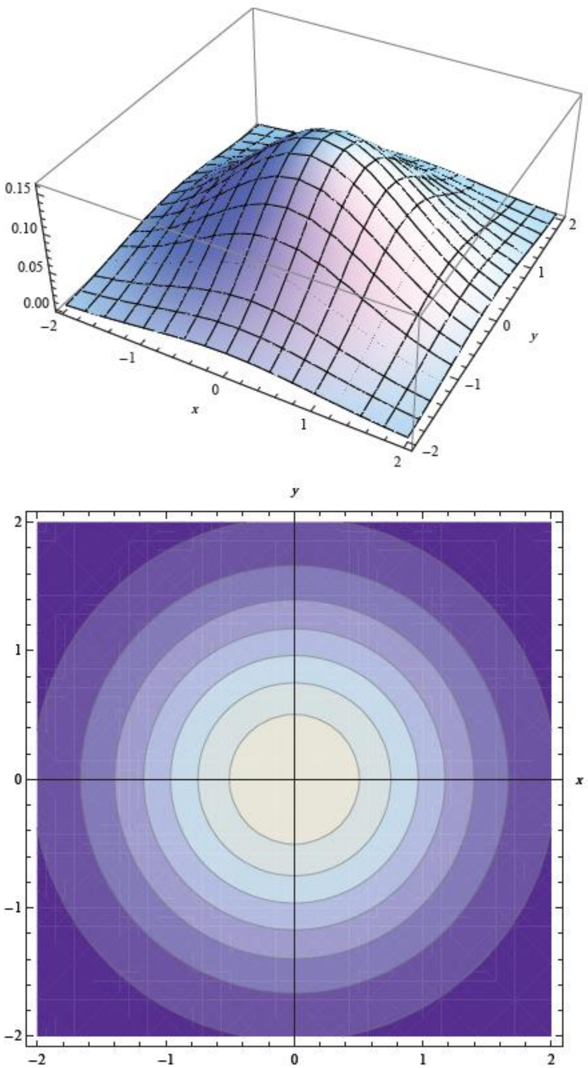
则称(X,Y)服从参数为$(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$的**二维正态分布**,记为

$$(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$$

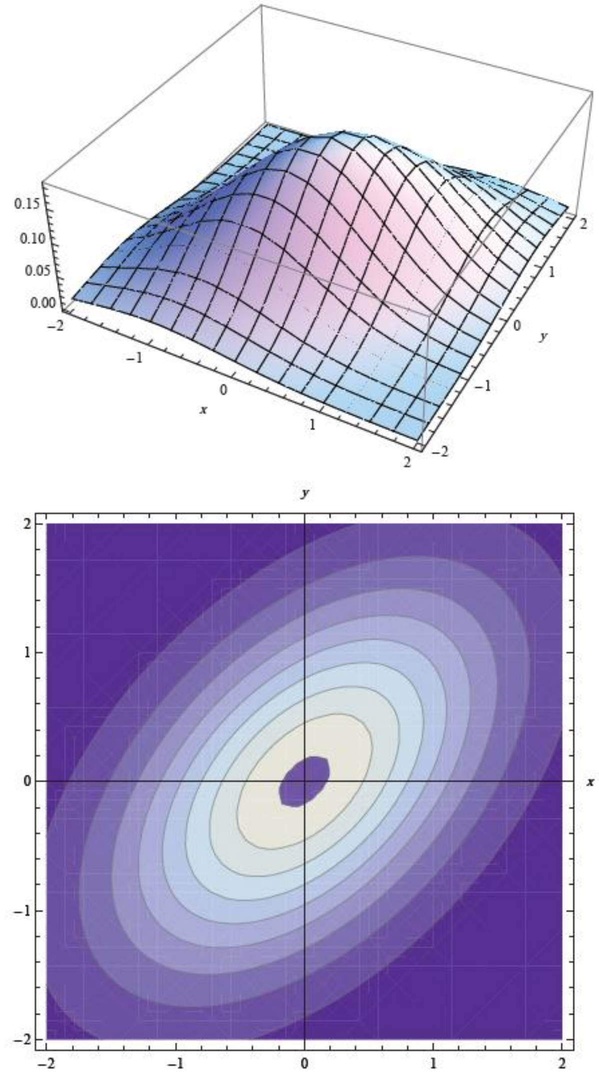
其中各参数满足 $-\infty<\mu_1<+\infty,-\infty<\mu_2<+\infty,\sigma_1>0,\sigma_2>0,|\rho|<1.$



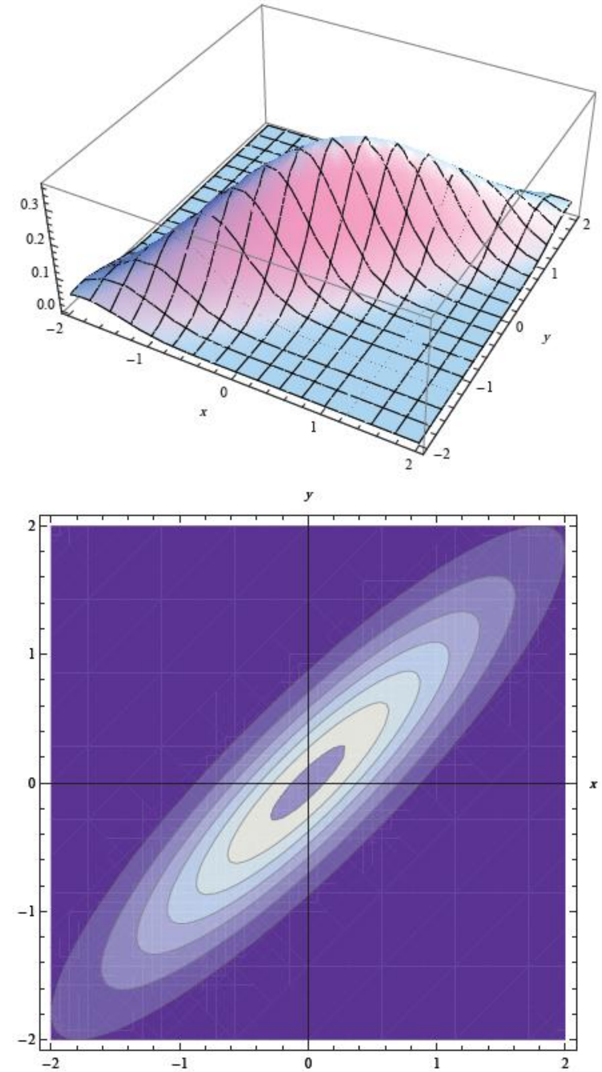
$\mu_1=\mu_2=0,\quad \sigma_1=\sigma_2=1$ 的二维正态密度



$\rho=0$

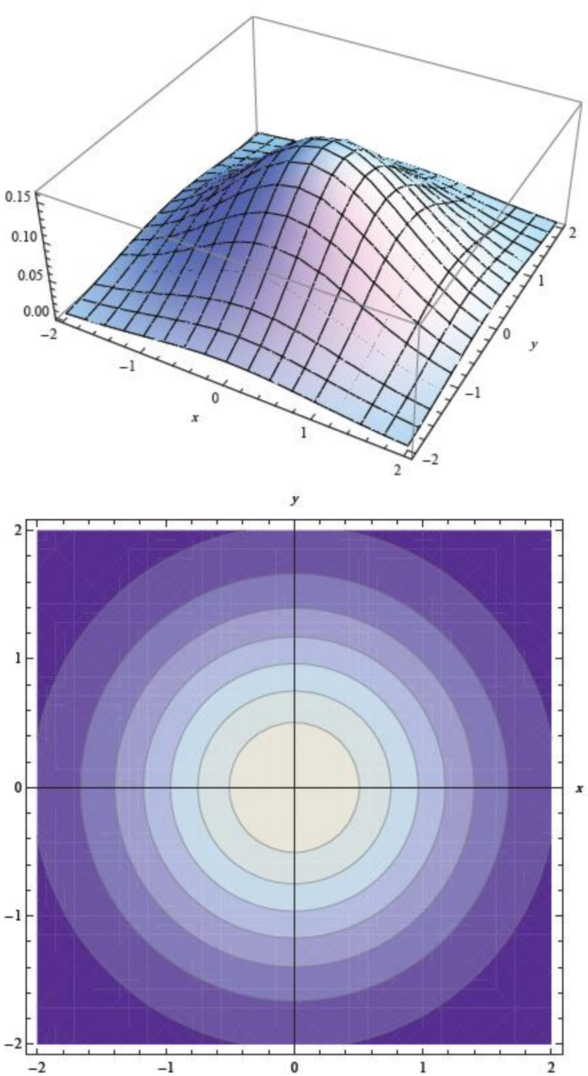


$\rho=0.5$

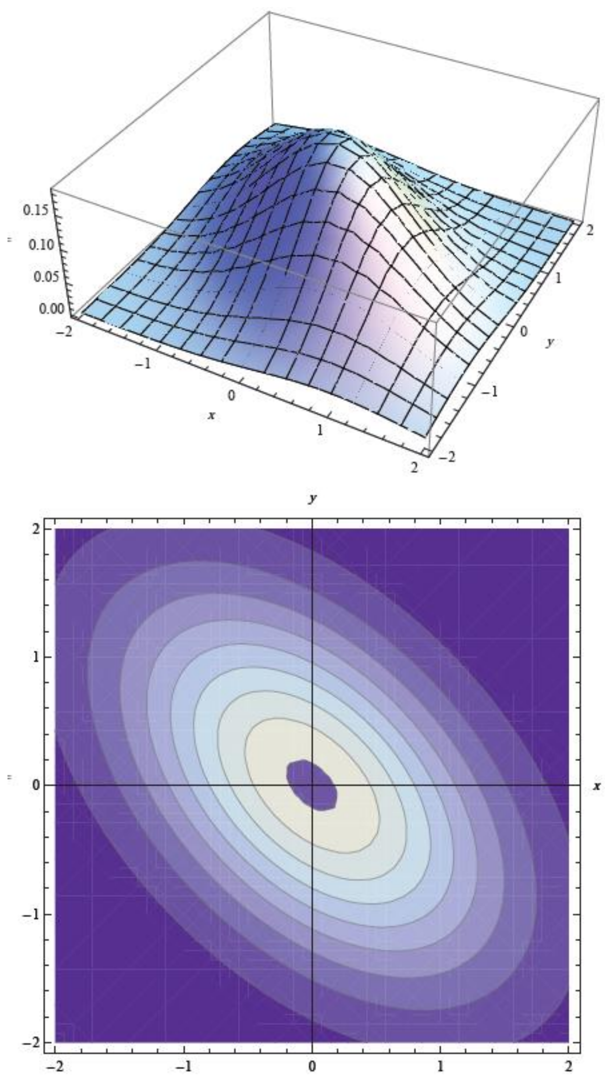


$\rho=0.9$

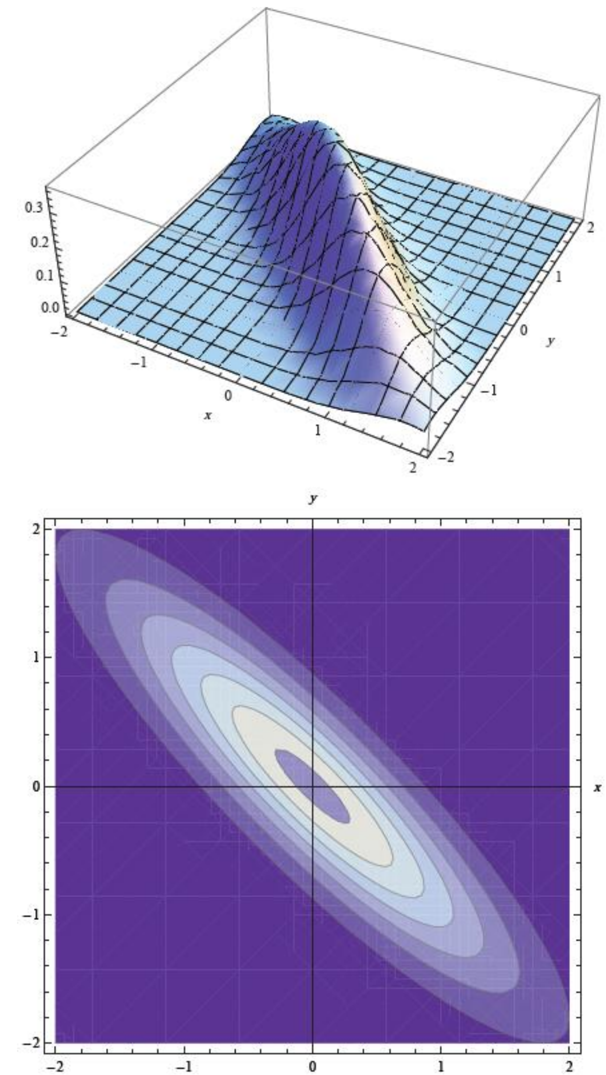
$\mu_1=\mu_2=0,\quad \sigma_1=\sigma_2=1$ 的二维正态密度



$\rho=0$



$\rho=-0.5$



$\rho=-0.9$

定理 若 $(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，则

$$\underline{X\sim N(\mu_1,\sigma_1^2)}\ ,\quad \underline{Y\sim N(\mu_2,\sigma_2^2)}$$

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

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

$$\times[\frac{(x-\mu_1)^2}{\sigma_1^2}-2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}]\}$$

$$=(\frac{y-\mu_2}{\sigma_2}-\rho\frac{x-\mu_1}{\sigma_1})^2-\rho^2\frac{(x-\mu_1)^2}{\sigma_1^2}$$

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

定理 若$(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$, 则

$$X\sim N(\mu_1,\sigma_1^2)\ ,\quad Y\sim N(\mu_2,\sigma_2^2)$$

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

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

令 $\frac{1}{\sqrt{1-\rho^2}}\left(\frac{y-\mu_2}{\sigma_2}-\rho\frac{x-\mu_1}{\sigma_1}\right)=t$

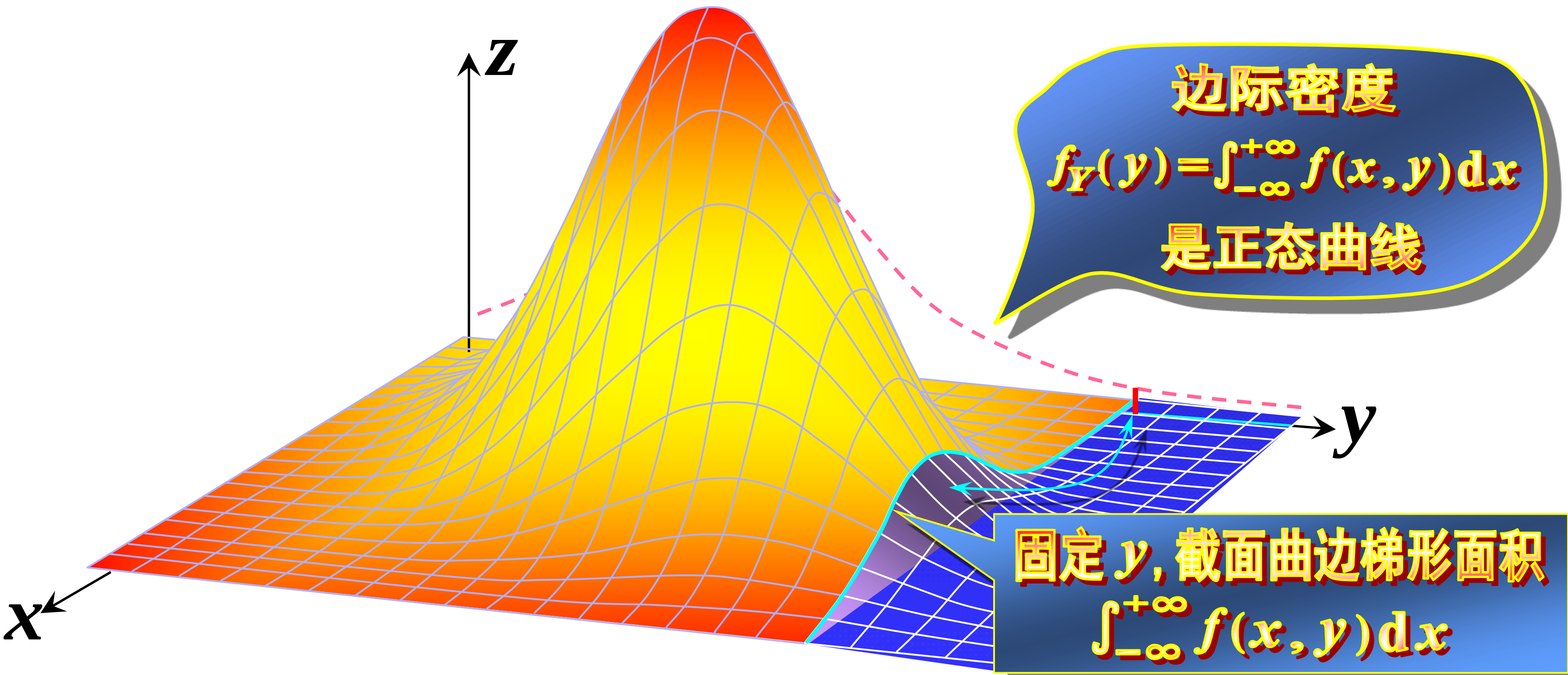
$$=\frac{1}{2\pi\sigma_1}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}\int_{-\infty}^{\infty}e^{-\frac{t^2}{2}}dt$$

$$=\frac{1}{2\pi\sigma_1}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}\cdot\sqrt{2\pi}$$

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

$\therefore\ X\sim N(\mu_1,\sigma_1^2)$ 同理可证 $Y\sim N(\mu_2,\sigma_2^2)$

正态密度的图形及边际密度的几何意义



问 由 X,Y 的边际分布能否确定联合分布？

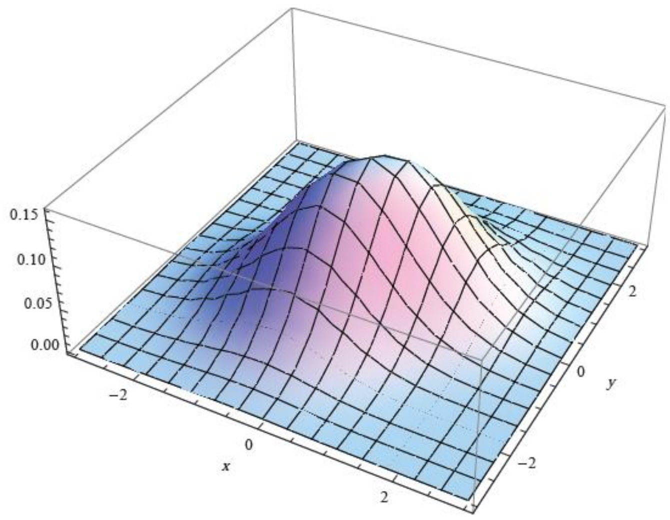
取边际密度都为 N(0,1) 的正态分布

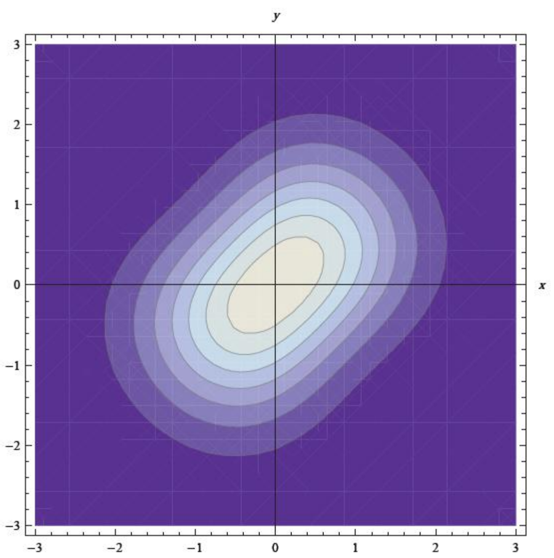
利用密度为 $c(u,v)=2-2u-2v+4uv$ 的连接函数

则二维联合密度函数为

$$f(x,y)=(2-2\Phi(x)-2\Phi(y)+4\Phi(x)\Phi(y))\varphi(x)\varphi(y)$$

具有边际正态，但不是二维正态密度.





均匀分布

- **区域G上的均匀分布，记作$U(G)$**

G是平面上的有界区域，面积为A

若r.v.(X,Y)的联合d.f.为

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

则称(X,Y)服从区域G上的均匀分布

若(X,Y)服从区域G上的均匀分布，

则 $\forall G_1 \subseteq G$，设G_1的面积为A_1，

$$P\left((X,Y) \in G_1\right) = \frac{A_1}{A}$$

例 设$(X,Y)\sim G$上的均匀分布,

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

求

(1) $f(x,y)$;

(2) $P(Y>X^2)$;

(3) (X,Y) **在平面上的落点到** y **轴距离小于0.3的概率.**

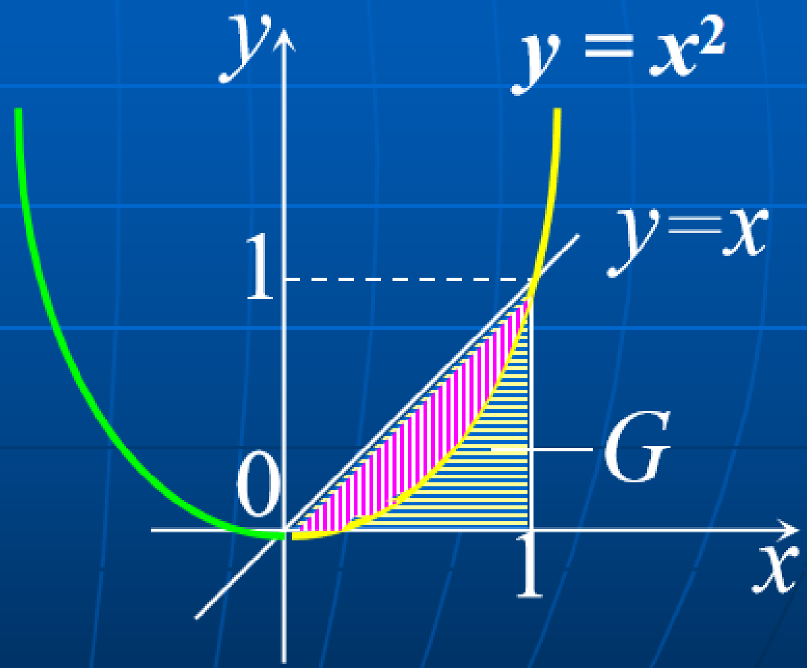
解 (1)

$$f(x,y)=\begin{cases}2, & 0\le y\le x, 0\le x\le 1\\ 0, & \text{其他}\end{cases}$$

(2) $P(Y>X^2)$

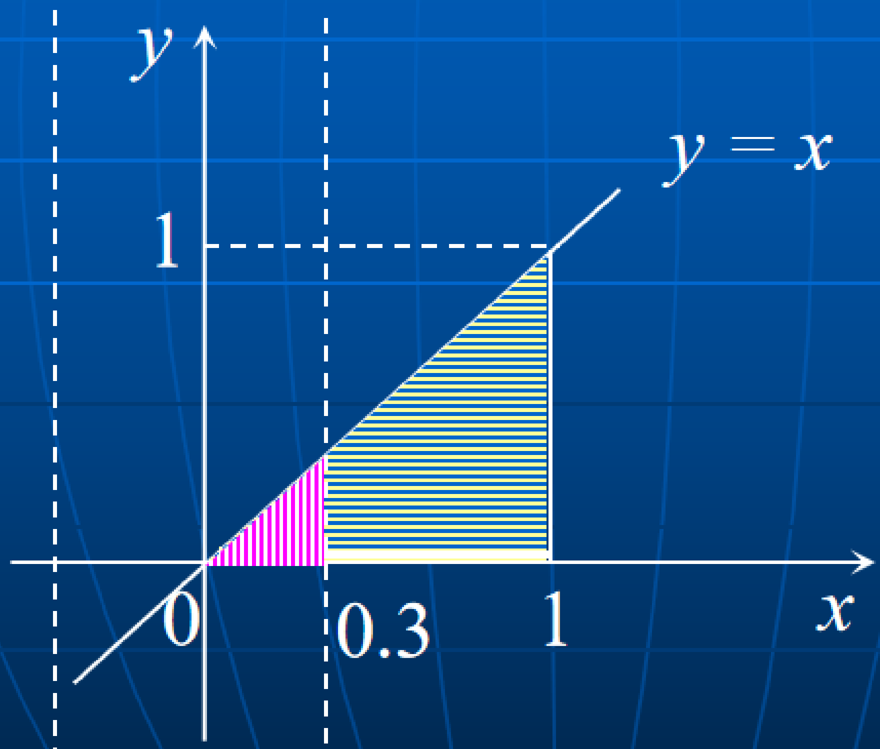
$$=\int_0^1 dx\int_{x^2}^{x} 2dy$$

$$=1/3.$$



(3) $P(|X|<0.3)=P(-0.3<X<0.3)$

$$=2\cdot\frac{1}{2}\cdot(0.3)^2=0.09$$



例 设随机变量$(X, \quad Y)$服从区域D上的均匀分布.

其中$D=\{(x,y) \mid x\ge 0, y\ge 0, x+\dfrac{y}{2}\le 1\}$,

试求随机变量$(X, \quad Y)$的边缘密度函数.

解: 区域D的面积为 $A=1$

$(X, \quad Y)$的联合密度函数为

$$f(x, \quad y)=\begin{cases} 1, & (x, \quad y)\in D \\ 0, & (x, \quad y)\notin D \end{cases}$$

当$0 \le x \le 1$时，

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

$$= 2(1-x)$$

当$x < 0$或$x > 1$时，$f_X(x) = 0$

随机变量X的边缘密度函数为

$$f_X(x) = \begin{cases} 2(1-x) & 0 \le x \le 1 \\ 0 & \text{其它} \end{cases}$$

同理，随机变量Y的边缘密度函数为

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

$$f_Y(y)=\begin{cases}1-\dfrac{y}{2}, & 0\le y\le 2\\ 0, & \text{其它}\end{cases}$$

课后作业 P76：5、6、7、8

补充题

1. 设二维连续随机变量(X,Y)的联合分布函数为

$$F(x,y)=\begin{cases}k(1-e^{-x})(1-e^{-y}), & x>0,y>0,\\ 0, & 其他,\end{cases}$$

求边缘密度函数及 P(1<X<3, 1<Y<2)。

2. 设二维连续随机变量(X,Y)的概率密度为

$$f(x,y)=\begin{cases}x+y, & 0<x,y<1,\\ 0, & 其他,\end{cases}$$

(1) 求边缘密度函数; (2)求 P(X>Y); (3)求 P(X <0.5)