

第三章 联合分布

- § 1 引言：联合累积分布函数
- § 2 (二维)离散随机变量
- § 3 (二维)连续随机变量
- § 4 独立随机变量
- § 5 条件分布
- § 6 联合分布随机变量函数
- § 7 极值和顺序统计量

回顾事件的独立性

A, B 相互独立 $\iff A, B$ 之间没有任何关系

$$\iff P(AB) = P(A)P(B)$$

 怎样定义 r.v X, Y 之间的独立性?

分析 若 X, Y 相互“独立”，从直观上看， X, Y 取任何值之间应是没有任何关系的，即 $\forall x, y \in R^1$

$$\{X \leq x\}, \{Y \leq y\}$$

应相互独立，即

$$P\{X \leq x, Y \leq y\} = P\{X \leq x\} \cdot P\{Y \leq y\}$$



$$F(x, y) = F_X(x) \cdot F_Y(y)$$

定义 设

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

若 $\forall x, y \in (-\infty, \infty)$ 有

$$P\{X \leq x, Y \leq y\} = P\{X \leq x\} \cdot P\{Y \leq y\}$$

即

$$F(x, y) = F_X(x) \cdot F_Y(y)$$

则称 r.v X, Y **相互独立**.

设 X, Y 相互独立, $\forall x_1 < x_2, y_1 < y_2$ 有

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

$\therefore \{x_1 < X \leq x_2\}, \{y_1 < Y \leq y_2\}$ 相互独立.

r. v独立性的直观意义

X 的取值与 Y 的取值是相互独立、互不相干的.

(一) 二维离散型 r.v 的独立性

设 (X, Y) 的频率函数为

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

则 X, Y 相互独立等价于 $\forall i, j = 1, 2, \dots$ 有

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

例 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 r.v 的独立性的方法

- ① 按定义判断
- ② 从直观背景判断

例 设 r.v X 从 1, 2, 3, 4 四个数中等可能取值, 又设 r.v Y 从 $1 \sim X$ 中等可能取值. 问 X, Y 是否独立?

解 由 § 2 例, (X, Y) 的频率函数及边际频率函数为

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

$$\because P\{X = i, Y = j\} \neq P\{X = i\} \cdot P\{Y = j\}, \quad i, j = 1, 2, 3, 4$$

$\therefore X, Y$ 不独立.

从直观上看, X, Y 也不独立

例 设 (X, Y) 的频率函数为

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

① a, b 应满足什么条件? ② 若 X, Y 独立, 求 a, b .

解 ① $\because \sum_{i,j} p_{ij} = 1$

$$\therefore a + b = 1 - \left(\frac{1}{8} + \frac{1}{24} + \frac{1}{4} + \frac{1}{8}\right) = \frac{11}{24}, \quad a \geq 0, b \geq 0$$

② 若 X, Y 相互独立, 则

$$a = P\{X=2, Y=1\} = P\{X=2\} \cdot P\{Y=1\} = (a + \frac{1}{4}) \left(\frac{1}{8} + a + \frac{1}{24}\right)$$

$$b = P\{X=1, Y=2\} = P\{X=1\} \cdot P\{Y=2\} = (b + \frac{1}{8}) \left(b + \frac{1}{4} + \frac{1}{8}\right)$$

解得 $a = \frac{1}{12}$ 或 $a = \frac{1}{2}$; $b = \frac{1}{8}$ 或 $b = \frac{3}{8}$.

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

更简单的做法?

(二) 二维连续型 r.v 的独立性

设 (X, Y) 为连续型 r. v, 且

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

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

若 X, Y 相互独立, 则

$$\begin{aligned} \int_{-\infty}^x \int_{-\infty}^y f(u, v) du dv &= P\{X \leq x, Y \leq y\} \\ &= P\{X \leq x\} P\{Y \leq y\} \\ &= \int_{-\infty}^x f_X(u) du \cdot \int_{-\infty}^y f_Y(v) dv \\ &= \int_{-\infty}^x \int_{-\infty}^y f_X(u) f_Y(v) du dv \end{aligned}$$

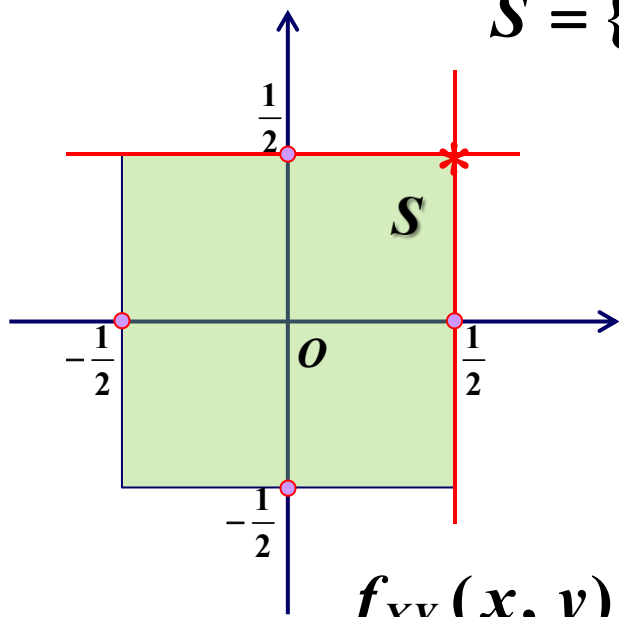
从而在 $f(x, y), f_X(x), f_Y(y)$ 的连续点处有

$$f(x, y) = f_X(x) f_Y(y)$$

即有 X, Y 相互独立 $\iff f(x, y) \stackrel{a.e.}{=} f_X(x) f_Y(y)$

例 设 (X, Y) 服从正方形 S 上的均匀分布, 其中

$$S = \{(x, y) \mid -\frac{1}{2} \leq x \leq \frac{1}{2}, -\frac{1}{2} \leq y \leq \frac{1}{2}\}$$



$$\text{则 } f_{XY}(x, y) = \begin{cases} 1, & (x, y) \in S \\ 0, & (x, y) \notin S \end{cases}$$

$$\text{易知 } f_X(x) = 1, \quad -1/2 \leq x \leq 1/2$$

$$f_Y(y) = 1, \quad -1/2 \leq y \leq 1/2$$

$$f_{XY}(x, y) = f_X(x)f_Y(y), \quad (X, Y) \text{ 独立.}$$

现考虑将 S 旋转 45° , 正方形 \rightarrow 菱形.

则 $f_X(x), f_Y(y)$ 不再是均匀分布密度.

(X, Y) 也不再独立. (如: $f_X(0.5) > 0, f_Y(0.5) > 0$,
 $f_{XY}(0.5, 0.5) = 0$. 且这样的点很多)

例 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?

解 X, Y 的边际密度分别为

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy = \begin{cases} 2e^{-2x}, & x > 0, \\ 0, & x \leq 0, \end{cases}$$
$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx = \begin{cases} e^{-y}, & y > 0, \\ 0, & y \leq 0, \end{cases}$$

$$\because f(x, y) = f_X(x) f_Y(y), \quad -\infty < x, y < \infty$$

$\therefore X, Y$ 相互独立.

是否为密度?

问 若 (X, Y) 的密度函数能分解为

$$f(x, y) = g(x)h(y)$$

其中 $g(x) \geq 0, h(y) \geq 0$, 问 X, Y 是否独立?



补充例子 若 (X, Y) 的概率密度为

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

情况又怎样?

密度函数形式可分离,
但支撑区域不可分离.

解

$$f_X(x) = \int_x^1 2dy = 2(1-x), \quad 0 < x < 1$$

$$f_Y(y) = \int_0^y 2dx = 2y, \quad 0 < y < 1$$

由于存在面积不为0的区域,

$$f(x, y) \neq f_X(x)f_Y(y)$$

故 X 和 Y 不独立.

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 $|\alpha|\leq 1$, 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.

可见, 仅当 $\alpha = 0$ 时, X 和 Y 才是独立的, 此时

$$H(x, y) = F(x)G(y)$$

分解成了边际分布 $F(x)$ 和 $G(y)$ 的乘积.

例 在某一分钟内，信号进入收信机是等可能的. 若收到两个互相独立的信号的时间间隔小于0.5秒，则信号将相互产生干扰，求两信号相互干扰的概率.

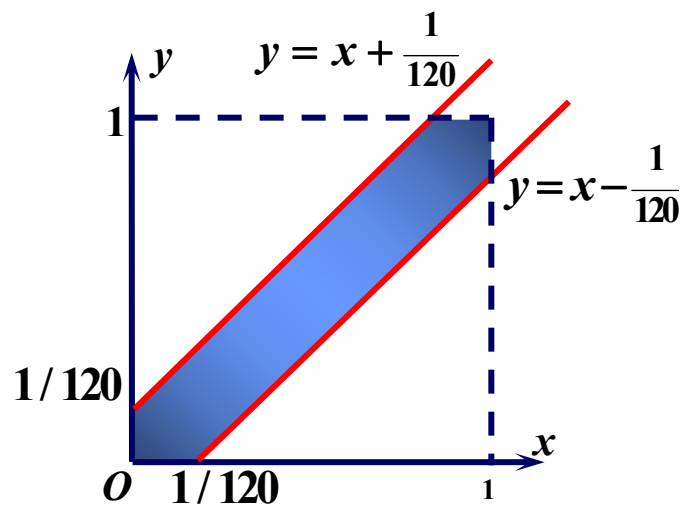
解 设两信号进入收信机时间分别为 X, Y (分钟), 则
$$X \sim U(0,1), Y \sim U(0,1)$$

$\therefore X, Y$ 独立, 故联合密度为

$$f(x, y) = f_X(x)f_Y(y) = \begin{cases} 1, & 0 < x < 1, 0 < y < 1 \\ 0, & \text{其它} \end{cases}$$

故两信号互相干扰的概率为

$$\begin{aligned} & P\{ |X - Y| < \frac{1}{120} \} \\ &= \iint_{|x-y| < 1/120} f(x, y) dx dy = \iint_{\substack{|x-y| < 1/120 \\ 0 < x < 1, 0 < y < 1}} dx dy \\ &= 1 - (1 - \frac{1}{120})^2 \approx 0.016 \end{aligned}$$



回忆：二维正态分布

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

密度函数为

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

其中各参数满足

$$-\infty < \mu_1 < \infty, -\infty < \mu_2 < \infty, \sigma_1 > 0, \sigma_2 > 0, |\rho| < 1$$

重要结论

$$(X, Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho) \rightarrow \begin{cases} X \sim N(\mu_1, \sigma_1^2) \\ Y \sim N(\mu_2, \sigma_2^2) \end{cases}$$



参数 ρ 与独立性有什么关系？

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

X, Y 相互独立 $\iff \rho = 0$

证 \Rightarrow 若 (X, Y) 相互独立, 则 $f(x, y) = f_X(x)f_Y(y)$, 即

$$\begin{aligned} & \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right]\right\} \\ &= \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left\{-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right\} \cdot \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left\{-\frac{(y-\mu_2)^2}{2\sigma_2^2}\right\} \end{aligned}$$

令 $x = \mu_1, y = \mu_2$, 则有

$$\sqrt{1-\rho^2} = 1$$

$$\therefore \rho = 0$$

\Leftarrow 若 $\rho = 0$, 显然

$$f(x, y) \equiv f_X(x)f_Y(y)$$

即 X, Y 相互独立.

(三) n 维 r.v 的边缘分布、独立性

设 n 维 r.v (X_1, X_2, \dots, X_n) 的分布函数

$$F(x_1, x_2, \dots, x_n) = P\{X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n\}$$

一维边缘分布

X_i 的分布函数

$$\begin{aligned} F_{X_i}(x_i) &= P\{X_1 < \infty, \dots, X_{i-1} < \infty, X_i \leq x_i, X_{i+1} < \infty, \dots, X_n < \infty\} \\ &= F(\infty, \dots, \infty, x_i, \infty, \dots, \infty) \quad (i = 1, 2, \dots, n) \end{aligned}$$

二维边缘分布

X_1, X_2 的联合分布是

$$\begin{aligned} F_{X_1 X_2}(x_1, x_2) &= P\{X_1 \leq x_1, X_2 \leq x_2, X_3 < \infty, \dots, X_n < \infty\} \\ &= F(x_1, x_2, \infty, \dots, \infty) \end{aligned}$$

类似可定义三维、四维等
高维边缘分布

X_2, X_3 的联合分布是

$$\begin{aligned} F_{X_2 X_3}(x_2, x_3) &= P\{X_1 < \infty, X_2 \leq x_2, X_3 \leq x_3, X_4 < \infty, \dots, X_n < \infty\} \\ &= F(\infty, x_2, x_3, \infty, \dots, \infty) \end{aligned}$$

随机向量的独立性

$\forall x_1, x_2, \dots, x_n \in R^1$ 若有

$$F(x_1, x_2, \dots, x_n) = F_{X_1}(x_1)F_{X_2}(x_2) \cdots F_{X_n}(x_n)$$

则称 X_1, X_2, \dots, X_n 相互独立.

设

$$(X_1, X_2, \dots, X_m) \sim F_1(x_1, x_2, \dots, x_m)$$

$$(Y_1, Y_2, \dots, Y_n) \sim F_2(y_1, y_2, \dots, y_n)$$

$$(X_1, X_2, \dots, X_m; Y_1, Y_2, \dots, Y_n) \sim F(x_1, x_2, \dots, x_m; y_1, y_2, \dots, y_n)$$

$\forall x_1, x_2, \dots, x_m, y_1, y_2, \dots, y_n \in R^1$ 若有

$$F(x_1, x_2, \dots, x_m; y_1, y_2, \dots, y_n) = F_1(x_1, x_2, \dots, x_m) \cdot F_2(y_1, y_2, \dots, y_n)$$

则称 $(X_1, X_2, \dots, X_m), (Y_1, Y_2, \dots, Y_n)$ 相互独立.

从直观上看: 随机向量的独立性是指各随机向量的取值是相互独立、互不相干的

定理 设 $(X_1, X_2, \dots, X_m), (Y_1, Y_2, \dots, Y_n)$ 相互独立, 则

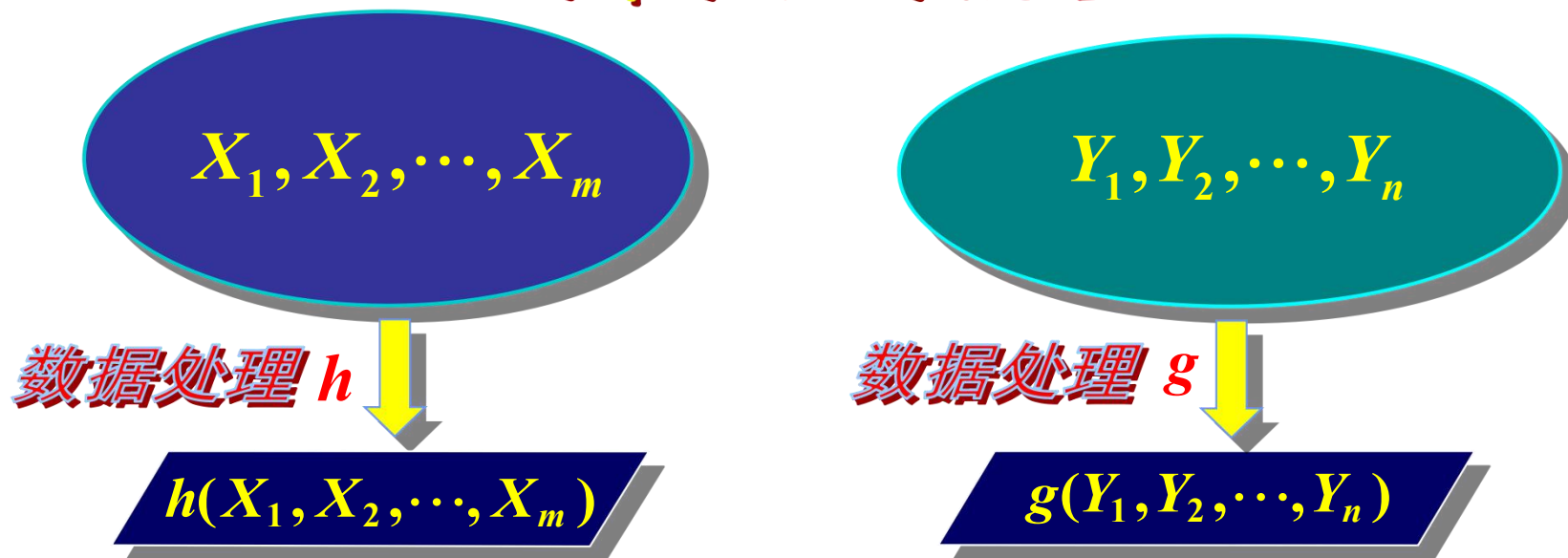
① X_i, Y_j 相互独立 ($i = 1, 2, \dots, m; j = 1, 2, \dots, n$)

② 设 h, g 分别是 m 元和 n 元的连续函数, 则

$$h(X_1, X_2, \dots, X_m), g(Y_1, Y_2, \dots, Y_n)$$

相互独立.

两堆独立的数据



处理后的数据仍是独立的



课后作业

P77: 19、补充题

补充题 设在 $\triangle ABC$ 内部任取一点 P , 在底边 BC 上任取一点 Q , 求直线 PQ 与线段 AB 相交的概率.

一个袋中有 5 个球, 其中 2 个白球 3 个黑球,

(1) 先后有放回的任取一球,

(2) 先后无放回的任取一球,

取到的白球个数分别为 X 和 Y , 求 (X,Y) 的联合频率函数及边缘频率函数, 讨论独立性。

在一个以原点为圆心半径为 R 的圆内随机选取一点, 令 (X,Y) 表示这一点的分布, 则 (X,Y) 服从

$$f(x,y) = \begin{cases} c, & x^2 + y^2 \leq R^2, \\ 0, & \text{其他,} \end{cases}$$

(1) 求 c ; (2) 求边缘密度函数;

(3) 讨论 X 和 Y 独立性。

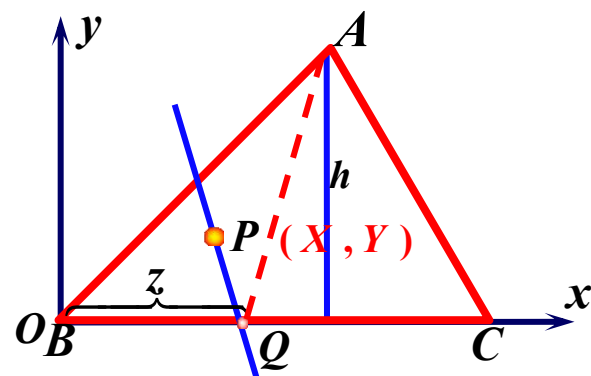
例 设在 $\triangle ABC$ 内部任取一点 P , 在底边 BC 上任取一点 Q , 求直线 PQ 与线段 AB 相交的概率.

解 建立坐标系如图.

依题意, 点 P 服从 $\triangle ABC$ 上的均匀分布, 点 Q 服从区间 $(0, BC)$ 上的均匀分布, 其概率密度分别为

$$f_{XY}(x, y) = \begin{cases} \frac{2}{h \cdot BC}, & (x, y) \in \triangle ABC \\ 0, & \text{其它} \end{cases}$$

$$f_Z(z) = \begin{cases} \frac{1}{BC}, & z \in (0, BC), \\ 0, & \text{其它} \end{cases}$$



因为点 P 与点 Q 相互独立, 故联合概率密度为

$$f(x, y, z) = f_{XY}(x, y) f_Z(z)$$

问 直线 PQ 与线段 AB 相交等价于什么?

例 设在 $\triangle ABC$ 内部任取一点 P , 在底边 BC 上任取一点 Q , 求直线 PQ 与线段 AB 相交的概率.

解 建立坐标系如图.

直线 PQ 与线段 AB 相交的概率为

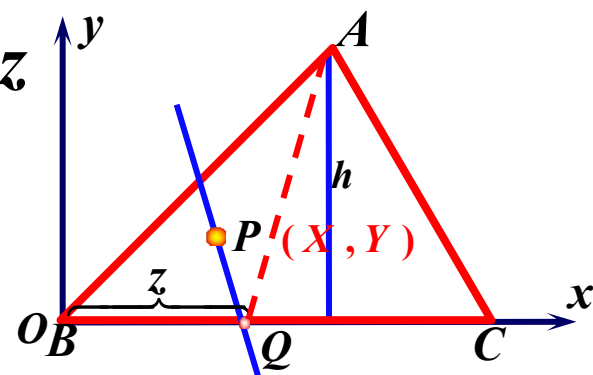
$$P\{(X, Y) \in \triangle ABQ, 0 < Z < BC\}$$

$$= \iiint_{\left\{ \begin{array}{l} (x, y) \in \triangle ABQ, \\ 0 \leq z \leq BC \end{array} \right\}} f_{XY}(x, y) f_Z(z) dx dy dz$$

$$= \int_0^{BC} f_Z(z) dz \iint_{(x, y) \in \triangle ABQ} f_{XY}(x, y) dx dy$$

$$= \frac{2}{h \cdot BC} \cdot \frac{1}{BC} \int_0^{BC} \frac{1}{2} h \cdot z dz$$

$$= \frac{1}{BC^2} \frac{BC^2}{2} = \frac{1}{2}$$



$$\iint_{(x, y) \in \triangle ABQ} dx dy = \triangle ABQ \text{ 面积}$$