Part III-B: Probability Theory and Mathematical Statistics

Lecture by 李漫漫

Note by THF

2024 年 10 月 16 日

目录

Lecture 7

两个随机变量的独立性

证明. 当 A, B 独立时：$P(AB) = P(A)P(B)$

$$\forall i, j : P\{X = x_i, Y = y_j\} = P\{X = x_i\} P(Y = y_j).$$

□

Notation. 离散随机变量独立的情况下：$P_{ij} = P_{i\cdot} \cdot P_{\cdot j}$

如何证明 X, Y 独立：

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

由联合分布律得边缘分布律:

$$F_X(x) = \lim_{y\to+\infty} F(x,y) = \lim_{y_0\to+\infty} P\{X \le x, Y \le y_0\}.$$

Notation. 连续性随机变量的区间 D 概率:

$$F(x,y) = P\{X \le x, Y \le y\} = \iint_{u=x,v=y} f(u,v)\,\mathrm{d}x\mathrm{d}y.$$

概率函数

Notation. 二维均匀分布:

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

Example.

$$f(x,y) = \begin{cases} Axy, & x\in(0,1), y\in(0,1) \\ 0, & \text{Others} \end{cases}.$$

1. 求 A

$$\iint_{x\in[0,y],y\in[0,1]} Axy\mathrm{d}x\mathrm{d}y = 1.$$

$A = 8$

2. $P\{X+Y\ge 1\}$

$$\begin{aligned} P\{X+Y\ge 1\} &= \iint_{x+y\ge 1} f(x,y)\,\mathrm{d}x\mathrm{d}y \\ &= \iint_{x+y\ge 1} 8xy\mathrm{d}x\mathrm{d}y \\ &= \int_{0.5}^{1}\mathrm{d}y\int_{1-y}^{y} 8xy\mathrm{d}x \\ &= \frac{5}{6} \end{aligned}$$

图 1: f(x,y)

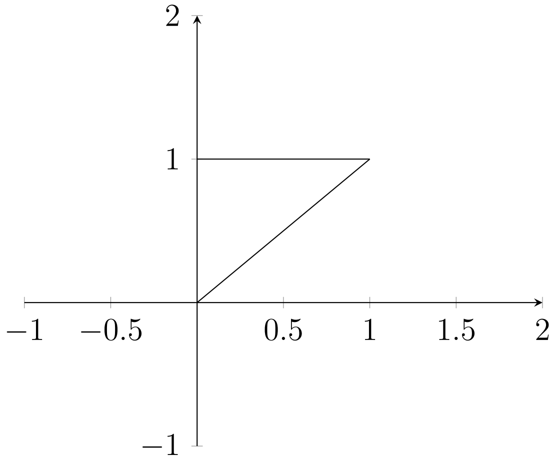
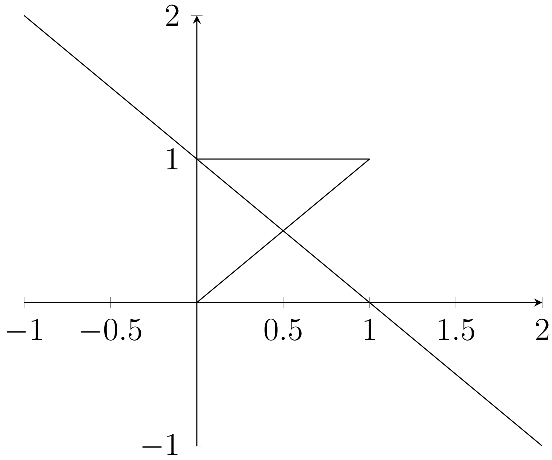


图 2: P(x+y>= 1)



3. X, Y 的分布函数

3.1 $x \in (-\infty, 0), y \in (-\infty, 0)$

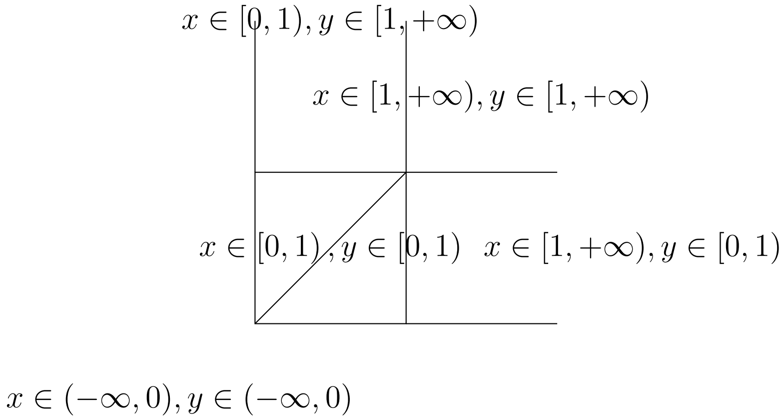
3.2 $x \in [0, 1), y \in [0, x)$

3.3 $x \in [0, 1), y \in [x, 1)$

3.4 $x \in [0, 1), y \in [1, +\infty)$

3.5 $x \in [1, +\infty), y \in [0, 1)$

3.6 $x \in [1, +\infty), y \in [1, +\infty)$



分段对 Axy 积分:

$$F(x, y) = \iint_{x \le u, y \le v} f(u, v) \, \mathrm{d}x \mathrm{d}y$$

0.1 多维随机变量及分布

Definition. 二维推广至多维:

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

称 F 为 n 维随机变量的联合分布函数

Definition. 多维联合分布律:

$$P\{X_1 = a_{1k_1}, X_2 = a_{2k_2}, \cdots, X_n = a_{nk_n}\}.$$

联合分布函数和联合分布律的关系:

$$F\left(x_1, x_2, \cdots, x_n\right).$$

$$=\sum_{a_{1k_1}\leq x_1}\sum_{a_{2k_2}\leq x_2}\cdots\sum_{a_{nk_n}\leq x_n} P\left\{X_1=a_{1k_1}, X_2=a_{2k_2}, \cdots, X_n=a_{nk_n}\right\}.$$

Notation. 二项分布推广多项分布:

$A_1, A_2, \cdots, A_r$ 是 E 的完备事件组，$P\left(A_i\right)=p_i, i=1,2,\cdots,r$，对 E 进行 n 次独立重复试验，X_i 表示 A_i 发生的次数，则:

$$P\left\{X_1=k_1, X_2=k_2, \cdots, X_r=k_r\right\}=\frac{n!}{k_1!k_2!\cdots k_r!}\prod_{i=1}^{r} p_i^{k_i}.$$

其中 $k_i \geq 0, \sum_{i=1}^{r} k_i=n$，当 $n=2$ 时为二项分布

0.1.1 多维随机变量的独立性

Definition.

$$F\left(x_1, x_2, \cdots, x_n\right)=\prod_{i=1}^{n} F_{X_i}\left(x_i\right).$$

称随机变量相互独立

等价于:

$$f\left(x_1, x_2, \cdots, x_n\right)=\prod_{i=1}^{n} f_{X_i}\left(x_i\right).$$

Lecture 8

10.15

0.1.2 条件分布

Notation. 离散型条件分布:

$$P\left\{X=a_i \mid Y=b_j\right\}=\frac{P\left\{X=a_i, Y=b_j\right\}}{P\left\{Y=b_j\right\}}.$$

$$P\left\{Y=b_j\right\}=\sum_{i=1}^{+\infty} P\left\{Y=b_j, X=a_i\right\}.$$

性质:

1. $P\left\{X=a_i|Y=b_j\right\} \geq 0$
2. $\sum\limits_{i=1}^{+\infty} P=1$

Notation. 连续型条件分布：

$$P\left\{X=a|Y=b\right\}=0.$$

（无穷多个点）

通过微元法：

$$\begin{aligned}
P\left\{X \leq x|Y=y\right\} &= \lim_{\varepsilon \to 0^+} P\left\{X \leq x|y-\varepsilon < Y \leq y\right\} \\
&= \lim_{\varepsilon \to 0^+} \frac{P\left\{X \leq x, Y \in (y-\varepsilon, y]\right\}}{P\left\{Y \in (y-\varepsilon, y]\right\}} \\
&= \lim_{\varepsilon \to 0^+} \frac{F\left(x,y\right)-F\left(x,y-\varepsilon\right)}{F_Y\left(y\right)-F_Y\left(y-\varepsilon\right)} \\
&= \frac{\dfrac{\partial F\left(x,y\right)}{\partial y}}{\dfrac{\mathrm{d}F_Y\left(y\right)}{\mathrm{d}y}} \\
&= \frac{\displaystyle\int_{-\infty}^{x} f\left(u,y\right)\mathrm{d}u}{f_Y\left(y\right)} \\
&= \int_{-\infty}^{x} \frac{f\left(u,y\right)}{f_Y\left(y\right)}\mathrm{d}u
\end{aligned}$$

.

0.2 二维随机变量函数的分布

$$\begin{cases} aX+bY+c \\ \max\{X,Y\} \\ \min\{X,Y\} \end{cases}.$$

重点公式：3.4.5 ~ 3.4.8

假设随机变量 Z=aX+bY+c，有分布函数 $(X,Y) \to F\left(x,y\right)$

$$
\begin{aligned}
F_Z(z) &= P\{Z \le z\} \\
&= P\{aX + bY + c \le z\} \\
&= \begin{cases} \displaystyle\sum_{ax_i+by_j+c\le z}\sum P\{X = x_i, Y = y_j\}, & (X,Y)\text{ 离散} \\ \displaystyle\iint_{ax+by+c\le z} P\{X = x, Y = y\}\,\mathrm{d}x\mathrm{d}y, & (X,Y)\text{ 连续} \end{cases}
\end{aligned}
$$

.

Notation. 二项分布可加性：有两个相互独立的试验 $X \sim B(m,p), Y \sim B(n,p)$，相当于一个试验 $Z \sim B(m+n,p)$

泊松分布可加性：相互独立的两个随机变量 $X \sim P(\lambda_1), Y \sim P(\lambda_2)$，相当于一个分布 $Z \sim P(\lambda_1 + \lambda_2)$

Notation. 极值公式：

两个随机变量（连续）X, Y 相互独立，求 $Z_1 = \max(X,Y), Z_2 = \min(X,Y)$ 的分布函数和密度函数

1. $F_{Z_1}(z)$：最大的不超过 z 等价于每一个都不超过 z

$$
\begin{aligned}
F_{Z_1}(z) &= P\{X \le z, Y \le z\} \\
&= P\{X \le z\} \cdot P\{Y \le z\} \\
&= F_X(z) \cdot F_Y(z)
\end{aligned}
$$

.

2. $F_{Z_2}(z)$：最小的不超过 z 不等价于每一个都不超过 z，但最小的超过 z 等价与每一个都超过了 z

$$
\begin{aligned}
F_{Z_2}(z) &= 1 - P\{X > z, Y > z\} \\
&= 1 - [1 - F_X(z)] \cdot [1 - F_Y(z)]
\end{aligned}
$$

.

Notation. 独立同分布：变量相互独立且分布律相同

对极值公式扩展：$(X_1, X_2, \ldots, X_n)$ 独立同分布：

$$F_{Z_1}(z) = \prod_{i=1}^{n} F_{X_i}(z).$$

由于同分布，因此 $F_{X_i}(z) = F_X(z)$

$$F_{Z_1}(z) = (F_X(z))^n.$$

$$f_{Z_1}(z) = (F_{Z_1}(z))' = n(F_X(z))^{n-1}.$$

Notation. 3.4.5:

$$\begin{aligned}
F_Z(z) &= P\{X + Y \le z\} \\
&= \iint_D f(x, y)\,\mathrm{d}x\mathrm{d}y \\
&= \int_{-\infty}^{+\infty} \mathrm{d}x \int_{-\infty}^{z-x} f(x, y)\,\mathrm{d}y \\
&= \int_{-\infty}^{z} \mathrm{d}t \int_{-\infty}^{+\infty} f(x, t-x)\,\mathrm{d}x \\
f_Z(z) &= F_Z'(z) \\
&= \int_{-\infty}^{+\infty} f(x, z-x)\,\mathrm{d}x
\end{aligned}$$

.

同理：

$$f_Z(z) = \int_{-\infty}^{+\infty} f(z-y, y)\,\mathrm{d}y.$$

Notation. 正态分布的可加性：

$X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2)$ ，且 X, Y 相互独立，则：

$$X + Y = Z \sim N\left(\mu_1 + \mu_2, \sigma_1^2, \sigma_2^2\right).$$

多元：

$$Z = \sum_{i=1}^{n} a_i X_i \sim N\left(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right).$$

0.3 二元正态分布

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\{-\frac{1}{2(1-\rho^2)}$$

$$\times \left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]\}.$$