多维随机变量及其分布函数

定义

1. n 维随机变量：
   设 Ω 为随机试验的样本空间，若：
   $$\forall\omega\in\Omega\to\exists(X_1(\omega),X_2(\omega),\dots,X_n(\omega))\in\mathbb{R}^n$$
   则称 $X=(X_1,X_2,\dots,X_n)$ 为 n 维随机变量（向量）。

2. 联合分布函数：
   对于 $x_1,x_2,\dots,x_n\in\mathbb{R}$，n元函数：
   $$F(x_1,x_2,\dots,x_n)=P(X_1\le x_1,X_2\le x_2,\dots,X_n\le x_n)$$
   称为 n 维随机变量 $X=(X_1,X_2,\dots,X_n)$ 的分布函数或者联合分布。

3. n 维随机变量 X 的任意 $k(1\le k\le n-1)$ 个分量
   $$X_{r1},X_{r2},\dots,X_{rk}\ \ (1\le r1<r2<\dots<rk\le n)$$
   可构成一个 k 维随机变量，其分布函数 $F(x_{r1},x_{r2},\dots,x_{rk})$ 称为 X 关于其对应的 k 个分量的边缘分布函数。

二维随机变量

联合分布函数性质

1. $$0\le F(x,y)\le 1$$
   $$F(+\infty,+\infty)=1,\quad F(-\infty,-\infty)=0$$
   $$F(x,-\infty)=0,\quad F(-\infty,y)=0$$

2. 对每个变量单调不减
   $$\forall y_1<y_2\ \ F(x,y_1)\le F(x,y_2)$$
   $$\forall x_1<x_2\ \ F(x_1,y)\le F(x_2,y)$$

3. 对每个变量右连续：
   $$F(x_0,y_0)=F(x_0+0,y_0)=F(x_0,y_0+0)$$

边缘分布函数

- $$\begin{aligned}F_X(x)&=P(X\le x)\\&=P(X\le x,Y<+\infty)\\&=F(x,+\infty)\end{aligned}$$
  $$\begin{aligned}F_Y(y)&=P(Y\le y)\\&=P(X<+\infty,Y\le y)\\&=F(+\infty,y)\end{aligned}$$

二维离散型随机变量

联合概率分布

- 定义：设离散型 (X, Y) 的所有可能取值为

$$(x_i, y_i),\quad i, j = 1, 2, \cdots$$

  则称 $P(X = x_i, Y = y_i) = p_{ij},\quad i, j = 1, 2, \cdots$ 为 (X, Y) 的联合概率分布或分布律。
- 求法：
  1. 直接利用古典概型等求解。
  2. 可利用乘法公式求解:

$$p_{ij} = P(X = x_i)P(Y = y_j|X = x_i)$$

边缘概率分布

- 定义：对二维离散型 (X, Y)，分量 X 和 Y 的概率分布称为：(X, Y) 关于 X 和 Y 的边缘概率分布。
- 求法：

$$\begin{aligned} p_{i\cdot} &= P(X = x_i) \\ &= \sum_j P((X, Y) = (x_i, y_j)) = \sum_j p_{ij},\quad i = 1, 2, \cdots \\ p_{\cdot j} &= \sum_i p_{ij},\quad j = 1, 2, \cdots \end{aligned}$$

关系

1. 已知联合分布函数，可求联合概率分布：

$$p_{ij} = F(x_i, y_j) - F(x_i - 0, y_j) - F(x_i, y_j - 0) + F(x_i - 0, y_j - 0)$$

2. 已知联合概率分布，可求联合分布函数：

$$F(x_i, y_j) = \sum_{i,j} p_{ij}$$

3. 已知联合概率分布，可求边缘概率分布：

$$p_{i\cdot} = \sum_j p_{ij}$$

4. 已知边缘概率分布，不一定能求出联合概率分布 **（在变量独立时可求）**
5. 已知边缘分布函数，可求出边缘概率分布：

$$p_{i\cdot} = F_X(x_i) - F_X(x_i - 0)$$

6. 已知边缘概率分布，可求出边缘分布函数：

$$F_X(x_i) = \sum_{k=1}^{i} p_{k\cdot}$$

二维连续型随机变量

联合概率密度

- 定义：设二维随机变量 (X, Y) 的联合分布函数为 $F = (x, y)$，若存在非负可积函数 $f(x, y)$，使得对于任意实数 x, y 有

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

  则称 (X, Y) 为二维连续型随机变量，称 $f(x, y)$ 为 (X, Y) 的联合概率密度函数。

边缘概率密度和边缘分布函数

首先已知边缘分布函数

$$F_X(x)=F(x,+\infty),\ \ F_Y(y)=F(+\infty,y)$$

因此

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

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

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

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

关系

- 已知联合分布函数，**且$F(x,y)$在(x,y)处连续**，可求联合概率密度：

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

- 已知联合概率密度，可求联合分布函数：

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

- 已知联合概率密度，可求边缘概率密度：

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

- 已知边缘密度分布，不一定可求联合概率密度 **（在变量独立时可求）**
- 已知边缘分布函数，可求边缘概率密度：

$$f_X(x)=\frac{\partial F_X(x)}{\partial x}$$

- 已知边缘概率密度，可求边缘分布函数：

$$F_X(x)=\int_{-\infty}^{x}f_X(u)du$$

二维随机变量的常见分布

1. 二维均匀分布：设 G 是平面上的有界区域，面积为 A，则二维连续型随机变量 (X,Y) 的联合概率密度为：

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

   记作 $U(G)$

2. 二维正态分布：

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

   记作 $(X,Y)\sim N(\mu_1,\sigma_1^2;\mu_2,\sigma_2^2;\rho)$，其中 $\sigma_1,\sigma_2>0;\ \ -1<\rho<1$

   - 二维正态分布的边缘分布仍然为正态分布：

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

$$f_Y(y)=\frac{1}{\sqrt{2\pi}\sigma_2}exp(-\frac{(y-\mu_2)^2}{2\sigma_2^2}),\quad -\infty<y<+\infty$$

随机变量的独立性

- 定义：设 (X,Y) 为二维随机变量，若对任意实数 x,y ，都有：

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

  则称随机变量 X 和 Y 相互独立。推广到 n 维随机变量，若

$$\begin{aligned}&P(X_1\le x_1,\ X_2\le x_2,\ \cdots,\ X_n\le x_n)\\ =&P(X_1\le x_1)P(X_2\le x_2)P(X_n\le x_n)\end{aligned}$$

  则称随机变量 $X_1,\ X_2,\ \cdots,\ X_n$ 相互独立。

- 性质：

$$\begin{aligned}&X,Y\ \ independent\\ \Leftrightarrow&F(x,y)=F_X(x)F_Y(y)\\ \Leftrightarrow&\forall a<b,\ c<d\\ &P(a<X\le b,c<Y\le d)\\ =&P(a<X\le b)P(c<Y\le d)\\ \Leftrightarrow&\forall a,c\in\mathbb{R}\\ &P(X>a,Y>c)=P(X>a)P(Y>c)\\ \Leftrightarrow&P(X=x_i,Y=y_j)=P(X=x_i)P(Y=y_j)\\ &namely\ \ p_{ij}=p_{i\cdot}p_{\cdot j}\\ &(for\ discrete\ random\ variable)\\ \Leftrightarrow&f(x,y)=f_X(x)f_Y(y)\\ &(for\ continuous\ random\ variable)\\ \Leftrightarrow&f_X(x)=f_{X|Y}(x|y)\ (f_Y(y)>0)\\ \Leftrightarrow&f_Y(y)=f_{Y|X}(y|x)\ (f_X(x)>0)\end{aligned}$$

- 应用
  1. 设 X,Y 为相互独立的随机变量，$u(x),\ v(y)$ 为连续函数，则 $U=u(X),\ V=v(Y)$ 也相互独立。

  2. 若 (X,Y) 为满足正态分布的二维随机变量。则 X 和 Y 独立等价于 $\rho=0$

- 判定定理：
  1. 设 $f(x,y)$ 是连续型二维随机变量(X,Y)的联合密度函数，$r(x),\ g(y)$ 为非负可积函数，且

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

     则 X 与 Y 相互独立，且

$$f_X(x)=\frac{r(x)}{\int_{-\infty}^{+\infty}r(x)dx}$$

$$f_Y(y)=\frac{g(y)}{\int_{-\infty}^{+\infty}g(y)dy}$$

  2. 设 $F(x,y)$ 是连续型二维随机变量 (X,Y) 的联合分布函数，则 X 与 Y 相互独立的充要条件为：

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

     且有

$$F_X(x) = \frac{R(x)}{R(+\infty)}$$

$$F_Y(y) = \frac{G(y)}{G(+\infty)}$$