Lecture 3

多元随机变量与分布

ruanyl@buaa.edu.cn
September 2016, Beihang University

二元分布

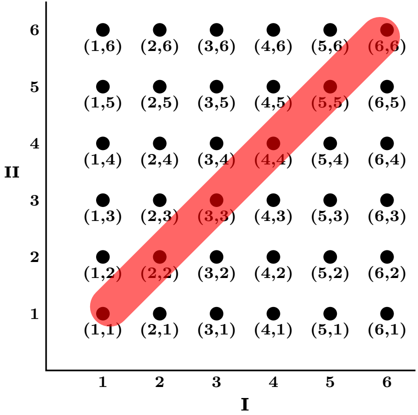
定义 假设 X, Y 是随机变量. 那么 (X, Y) 的**联合分布或者二元分布**定义为所有如下形式的概率的集合：$P((X, Y) \in C)$ 对任意 $C \subset \mathbb{R}^2$. 其中 $\{(X, Y) \in C\}$ 是 $\{\omega \in S : (X(\omega), Y(\omega)) \in C\}$ 的简写 . ω 通常表示一个样本.

二元离散分布

定义 如果联合分布 (X, Y) 的可能取值范围是有限点集或者无穷可数点集, 那么这个分布被称为**离散分布**.

定理 如果 X, Y 是离散随机变量, 那么 (X, Y) 是一个联合离散分布.

例题 投掷两个骰子, 分别用 X 和 Y 表示每个骰子出现的点数. 那么下图则是 (X, Y) 的联合分布的表示. 颜色区域表示事件 $\{X = Y\}$



与一元情形类似, 离散联合分布也有概率函数

定义 如果 (X, Y) 是一个离散联合分布, 那么下面的函数称为其联合概率函数

$$f(x, y) = P(X = x, Y = y).$$

定理 如果 (X, Y) 是一个离散联合分布, 那么其联合概率函数满足

$$\sum_{x,y} f(x,y) = 1$$

并且对任意 $C \subset \mathbb{R}^2$,

$$P\left((X,Y) \in C\right) = \sum_{(x,y)\in C} f(x,y)\,.$$

二元连续分布

定义 随机变量 X, Y 称为具有连续联合分布 , 如果存在定义在 $\mathbb{R}^2$ 上的非负函数 $f(x,y)$ 使得下面的等式成立,

$$P\left((X,Y) \in E\right) = \int\int_E f(x,y)\,dxdy$$

其中 E 是 $\mathbb{R}^2$ 上的任意有界, 无界，开或者闭的实数区域. **非负**函数 $f(x,y)$ 则称为该连续分布的**密度函数**.

需要注意的是在这个定义中我们并不假定密度函数 $f(x,y)$ 的连续性, 或是有界性.

定理 由概率公理可知, 密度函数应满足

$$\int\int_{R^2} f(x,y)\,dxdy = 1.$$

定理 假设随机变量 X, Y 具有连续联合分布. 那么

$$P\left((X,Y)\in C_0\right) = 0$$

其中 C_0 是 $\mathbb{R}^2$ 上的任何有限点集, 无穷点列或者函数曲线的图集.

例题 假设随机变量 X, Y 的联合分布具有密度函数

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

试确定常数 c, 并求出 $P(X\geqslant Y)$.

- 密度函数在全空间积分应当为 1,

$$\begin{aligned} 1 &= \int_{-1}^{1}\left(\int_{x^2}^{1} cx^2y\,dy\right)dx \\ &= c\int_{-1}^{1}\left(\frac{x^2}{2}-\frac{x^2x^4}{2}\right)dx \\ &= c\left(\frac{1}{3}-\frac{1}{7}\right)=\frac{4c}{21} \end{aligned}$$

因而 $c=21/4$.

$$P(X\geqslant Y)=\frac{21}{4}\int_{0}^{1}\left(\int_{x^2}^{x} x^2y\,dy\right)dx$$

$$
\begin{aligned}
&= \frac{21}{4} \int_0^1 \left(\frac{x^2 x^2}{2} - \frac{x^2 x^4}{2} \right) dx \\
&= \frac{21}{8} \left(\frac{1}{5} - \frac{1}{7} \right) = \frac{3}{20}
\end{aligned}
$$

二元混合分布

定义 假设 X 是离散随机变量，Y 是连续随机变量, 如果存在 $\mathbb{R}^2$ 上的非负函数 $f(x, y)$ 使得下面的积分对任何 $A, B \subset \mathbb{R}$ 都有意义并且成立

$$
P(X \in A, Y \in B) = \int_B \sum_{x \in A} f(x, y)\, dy.
$$

累积分布函数

定义 给定概率空间以及概率 P. 该空间上的随机变量 X, Y 的联合**分布函数**定义为

$$F(x, y) = P(X \leqslant x, Y \leqslant y), \forall x, y.$$

联合分布函数有如下性质

- 如果 X, Y 的分布函数分别记作 $F_X(x), F_Y(y)$, 那么成立

$$F_X(x) = P(X \leqslant x, Y < \infty) = \lim_{y \to \infty} F(x, y),$$

以及

$$F_Y(y) = P(X < \infty, Y \leqslant y) = \lim_{x \to \infty} F(x, y).$$

这里 $F_X(x)$ 和 $F_Y(y)$ 分别称为联合分布关于 X 和 Y 的**边际分布**.

$F(\cdot,-\infty)=F(-\infty,\cdot)=0, F(\infty,\infty)=1.$

下面的定理给出了分布函数与密度函数的关系.

定理 假设 X, Y 的联合分布具有密度函数 $f(x,y)$. 那么 X, Y 的联合分布函数是连续的, 并且在其二阶导数存在时, 联合分布函数

$$F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y} f(s,t)\,dsdt$$

满足以下关系

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

特别地, 如果 f 在 (x,y) 存在连续, 那么上面的求导关系式在该连续点成立.

定理 假设 X, Y 的联合分布具有密度函数 $f(x, y)$. 那么 X, Y 的联合分布函数关于 X 的分布具有密度函数

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

关于 Y 的分布具有密度函数

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

如果 X 或者 Y 或者两者同时是离散随机变量, 那么上面的相应的积分应改写为求和.

例题 本节开始投掷两个骰子的例题中我们给出了其联合分布, 试写出它的两个边际分布.

■ 对 $x, y \in \{1,2,3,4,5,6\}$,

$$f_X(x) = \sum_y f(x,y) = \sum_y P(X=x, Y=y) = \frac{1}{6}$$

$$f_Y(y) = \sum_x f(x,y) = \sum_x P(X=x, Y=y) = \frac{1}{6}$$

这个例题中的联合分布做成一个**各个元素都非负**的矩阵, 其各列和都是相等的. 乘以适当的常数后可使得各列和变为 1, 这一列和为 1 的矩阵被称为**随机矩阵**.

例题 假设 X, Y 的联合分布由下面的矩阵给出

$$\frac{1}{3}\begin{pmatrix} 5/12 & 1/4 & 1/3 \\ 5/12 & 1/4 & 1/3 \\ 1/6 & 1/2 & 1/3 \end{pmatrix}$$

试写出它的两个边际分布.

例题 假设随机变量 X, Y 的联合分布具有密度函数

$$f(x,y)=\begin{cases}\frac{21}{4}x^2y, & x^2\leqslant y\leqslant 1\\ 0, & \text{其他}\end{cases}$$

试求出其边际密度函数

- 由前面的定理

$$f_X(x)=\int_{-\infty}^{\infty}f(x,y)\,dy=\int_{x^2}^{1}\frac{21}{4}x^2ydy=\frac{21}{8}\left(x^2-x^6\right)$$

$$f_Y(y)=\int_{-\infty}^{\infty}f(x,y)\,dx=\int_{-y^{1/2}}^{y^{1/2}}\frac{21}{4}x^2ydx=\frac{7}{2}x^{\frac{5}{2}}$$

多元分布

许多二元分布的概念和相关性质都可以直接地推广到多元情形, 在此不做累述.

例题 在一个临床试验中, 有 n 个病人接受治疗. 用 X_i 表示第 i 个病人在接收治疗后的状态,

$$X_i = \begin{cases} 1, & \text{恢复} \\ 0, & \text{其他} \end{cases}$$

假设每个病人能痊愈的概率是 p, 各个病人情况是相互独立的, 那么 $\mathbf{X} = (X_1, \ldots, X_n)$ 的联合密度应该是

$$f(\mathbf{x}) = p^{x_1+\ldots+x_n}(1-p)^{n-x_1-\ldots-x_n}$$

其中 $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$.

例题 排队系统是一个由客户以及服务组成的系统, 客户排队等待并按照某流程接受服务. 一个简单的情形是所有客户都排队

等待同一个终端提供的服务, (例如客户排队等待从同一台取款机取款). 假设由个客户. 第个客户接受服务所消耗的时间是. 为了描述这个系统中客户等待时间的分布情况, 我们通常假设 $\mathbf{X} = (X_1, \ldots, X_n)$ 的联合密度是

$$f(\mathbf{x}) = \begin{cases} \frac{c}{\left(2+\sum_{i=1}^{n} x_i\right)^{n+1}}, & x_i > 0, \forall i \\ 0, & \text{其他} \end{cases}$$

确定常数 c.

- 联合密度 $f(\mathbf{x})$ 应满足

$$I = \int_0^{\infty} \ldots \int_0^{\infty} f(x_1, \ldots, x_n)\, dx_1 dx_2 \ldots dx_n = 1$$

将 $f(\mathbf{x})$ 的表达式代入并做累次积分, 首先对 x_n 积分得到

$$I = \int_0^{\infty} \ldots \int_0^{\infty} \frac{c}{n\left(2 + \sum_{i=1}^{n-1} x_i\right)^{n}} dx_1 dx_2 \ldots dx_{n-1} = 1$$

然后再依次对 $x_{n-1}, ..., x_1$ 积分得到

$$I = \frac{c}{n! \cdot 2} = 1$$

因此 $c = 2n!$

随机变量独立性

定义 随机变量 X, Y 称为是独立的, 如果:

$$P(X \in A, Y \in B) = P(X \in A) P(Y \in B)$$

对任意 $A, B \subset \mathbb{R}$. 严格地说 A, B 应该是由简单区间（开区间, 闭区间, 或半开闭区间）生成的"可测集", 这部分内容已超出本课程要求, 在此不做深入讨论. 但我们有如下定理

定理 随机变量 X, Y 是独立的当且仅当对 $\forall x, y \in \mathbb{R}$

$$P(X \leqslant x, Y \leqslant y) = P(X \leqslant x) P(Y \leqslant y).$$

定理中 $\leqslant$ 替换为 $<$ 时仍然成立. 这一定理是告诉我们随机变量 X, Y 独立的充分必要条件是联合分布函数等于边际分布函数乘积. 从而借助于分布函数与密度函数的关系, 我们还有

定理 假设 X, Y 其具有密度函数. 随机变量 X, Y 是独立的当且仅当其联合密度函数 $f(x,y)$ 满足

$$f(x,y)=f_X(x)f_Y(y),\ a.e.\ x,y.$$

如果 $f(x,y), f_X(x), f_Y(y)$ 都是连续的, 那么独立性等价于

$$f(x,y)=f_X(x)f_Y(y),\ \forall\ x,y.$$