

多维随机变量及其分布

多维随机变量

定义：二维：样本空间上有两个随机变量
(0,1)，且单调不减
有趋于负无穷时概率为0，双趋于正无穷，则概率为1

性质：
右连续
 $P\{x_1 < X \leq x_2, y_1 < Y \leq y_2\} = F(x_2, y_2) - F(x_2, y_1) + F(x_1, y_2) - F(x_1, y_1)$
联合分布可确定X、Y的边缘分布，但反过来边缘分布不能确定联合分布，相互独立才行

主要类型：离散型、连续型

相互独立性

概念： $F(x, y) = F_X(x)F_Y(y)$

充要条件：
离散型： $P\{X=x_i, Y=y_j\} = P\{X=x_i\}P\{Y=y_j\}$
连续型： $f(x, y) = f_X(x)f_Y(y)$

性质：
系列随机变量相互独立，则其中任意k个随机变量也相互独立
随机变量相互独立，则它们的函数也相互独立
若两个随机变量相互独立，则一个相对于另一个的条件分布就是其无条件分布

离散：条件分布=边缘分布
连续：条件密度=边缘密度

分布函数

联合分布函数
边缘分布函数

离散型分布

概念：有限个或无穷可列

性质： $P_{ij} \geq 0, \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} P_{ij} = 1$

边缘分布：
 $P_{i\cdot} = P\{X=x_i\} = \sum_{j=1}^{\infty} P\{X=x_i, Y=y_j\} = \sum_{j=1}^{\infty} P_{ij}$
 $P_{\cdot j} = P\{Y=y_j\} = \sum_{i=1}^{\infty} P\{X=x_i, Y=y_j\} = \sum_{i=1}^{\infty} P_{ij}$

概率分布：
 $P\{X=x_i | Y=y_j\} = \frac{P\{X=x_i, Y=y_j\}}{P\{Y=y_j\}} = \frac{P_{ij}}{P_{\cdot j}}$
条件概率分布： $P\{Y=y_j | X=x_i\} = \frac{P\{X=x_i, Y=y_j\}}{P\{X=x_i\}} = \frac{P_{ij}}{P_{i\cdot}}$

概率分布

连续型分布

概念： $F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) du dv$ 非负可积
 $f(x, y) \geq 0$
 $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dx dy = F(+\infty, +\infty) = 1$

性质：
 $P\{(x, y) \in G\} = \iint_G f(x, y) dx dy$
 $\frac{\partial^2 F(x, y)}{\partial x \partial y} = f(x, y)$

边缘概率密度： $f_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy, f_Y(y) = \int_{-\infty}^{+\infty} f(x, y) dx$

概率分布：
 $F_{X|Y}(x|y) = \int_{-\infty}^x \frac{f(t, y)}{f_Y(y)} dt, F_{Y|X}(y|x) = \int_{-\infty}^y \frac{f(x, t)}{f_X(x)} dt$

条件概率密度： $f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)}, f_{Y|X}(y|x) = \frac{f(x, y)}{f_X(x)}$

密度乘法公式： $f(x, y) = f_X(x)f_{Y|X}(y|x) = f_Y(y)f_{X|Y}(x|y)$

两个常见的二维连续型随机变量的分布

二维均匀分布：
概念： $f(x, y) = \begin{cases} \frac{1}{A}, & (x, y) \in G \\ 0, & \text{其他} \end{cases}$
性质：
概率只与子区域的面积大小有关，与之位置和形状无关
二维均匀分布的边缘分布，条件分布以及数字特征都与区域G的形状密切相关

二维正态分布：
概念： $u_i = EX, u_2 = EY, \sigma_1^2 = DX, \sigma_2^2 = DY, \rho = \frac{Cov(X, Y)}{\sigma_1 \sigma_2}$
性质：
联合分布是正态分布，X、Y为正态分布
反之不成立，即使二者相关系数为0，其联合分布也不一定正态分布，相互独立才行
X和Y相互独立的充分必要条件是相关系数=0
两个正态分布随机变量X与Y的任意线性组合仍服从正态分布： $aX+bY \sim N(a\mu_1+b\mu_2, a^2\sigma_1^2+b^2\sigma_2^2+2ab\rho\sigma_1\sigma_2)$

两个随机变量函数的分布

分布函数法

公式法

求法：
离散型：直接计算或通过边缘分布于联合分布关系，或条件分布间接计算
连续型：例连续，先求分布函数，再求其密度函数
一连续一离散，对离散型随机变量的各种可能取值用全概率公式展开，再与连续型各点概率求和

几个常见分布的求法

$Z = X+Y$ ：
 $f_Z(z) = \int_{-\infty}^{+\infty} f(z-y, y) dy$
 $f_Z(z) = \int_{-\infty}^{+\infty} f_X(z-y)f_Y(y) dy$
 $f_Z(z) = \int_{-\infty}^{+\infty} f_X(x-z)f_Y(z-x) dx$ 独立，卷积公式

$Z = X-Y, Z=Y-X$ ：
 $f_Z(z) = \int_{-\infty}^{+\infty} f(x, x-z) dx = \int_{-\infty}^{+\infty} f(x+z, z) dz$
 $Y = X+Z, X = Y-Z$

$Z = XY$ ：
 $f_Z(z) = \int_{-\infty}^{+\infty} \frac{1}{|x|} f\left(x, \frac{z}{x}\right) dx \xrightarrow{\text{换元}} f_Z(z) = \int_{-\infty}^{+\infty} \frac{1}{|x|} f_X(x) f_Y\left(\frac{z}{x}\right) dx$

$Z = \max(X, Y), Z = \min(X, Y)$ ：
 $F_{\max}(z) = P\{\max(X, Y) \leq z\} = P\{X \leq z, Y \leq z\} = F(z, z) \xrightarrow{\text{独立}} F_X(z)F_Y(z)$
 $F_{\min}(z) = F_X(z) + F_Y(z) - F(z, z) \xrightarrow{\text{独立}} 1 - [1 - F_X(z)][1 - F_Y(z)]$

分布的可加性：
 $X \sim B(n, p), Y \sim B(m, p) \rightarrow X+Y \sim B(n+m, p)$
 $X \sim P(\lambda_1), Y \sim P(\lambda_2) \rightarrow X+Y \sim P(\lambda_1+\lambda_2)$
 $X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2) \rightarrow X+Y \sim N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$
 $X \sim \chi^2(n), Y \sim \chi^2(m) \rightarrow X+Y \sim \chi^2(m+n)$