3 多为随机变量及其分布

3.1 多维随机变量及其分布

3.1.1 多维随机变量定义

如果$X_1(\omega)$，$X_2(\omega)$，$\cdots$，$X_n(\omega)$是定义在同一个样本空间$\Omega=\{\omega\}$上的n个随机变量，则称：

$$X(\omega)=(X_1(\omega),X_2(\omega),\cdots,X_n(\omega))$$

为n维随机变量或随机向量。

3.1.2 联合概率质量函数

如果二维随机向量(X,Y)所有可能的取值为$(x_i,y_j),i,j=1,2,\cdots$，这两个随机变量同时发生的概率可以用函数表示如下：

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

且此函数满足如下性质（即概率的三大公理）：

- **非负性**：

$$p_{ij}\geq 0$$

- **规范性和可加性**

$$\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}p_{ij}=1$$

则称此函数为(X,Y)的联合概率质量函数（Joint Probability Mass Function），或者称为联合分布列，此定义可以推广到多维离散随机变量上去。

3.1.2 联合概率密度函数

对于某二维随机向量(X,Y)存在二元函数$p(x,y)$满足：

- **非负性**：

$$p(x,y)\geq 0$$

- **规范性和可加性**（连续的都通过积分来相加）：

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}p(x,y)\mathrm{d}x\mathrm{d}y=1$$

则称此函数为(X,Y)的联合概率密度函数（Joint Probability Density Function），或者称为联合分布列，此定义可以推广到多维连续随机变量上去。

3.1.3 联合累积分布函数

3.1.3.1 定义

设(X,Y)是二维随机变量，对于任意实数x、y，可以定义一个二元函数来表示两个事件同时发生的概率：

$$F(x,y)=P\Big(\{X\le x\}\ \text{且}\ \{Y\le y\}\Big)=P(X\le x,Y\le y)$$

称为二维随机变量(X,Y)的联合累积分布函数（Joint Cumulative Distribution Function），如果混合偏导存在的话，那么：

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

得到$p(x,y)$就是此分布的概率密度函数。此定义和性质可以推广到多维随机变量。

3.1.3.2 性质

1 单调性

$F(x,y)$对于x或y是单调非减的，即：

$$\begin{aligned}F(x_1,y)&\le F(x_2,y),\quad x_1<x_2\\F(x,y_1)&\le F(x,y_2),\quad y_1<y_2\end{aligned}$$

2 有界性

对任意的x和y有：

$$0\le F(x,y)\le 1$$

且：

$$\begin{aligned}F(-\infty,y)&=\lim_{x\to-\infty}F(x,y)=0\\F(x,-\infty)&=\lim_{y\to-\infty}F(x,y)=0\\F(+\infty,+\infty)&=\lim_{x,y\to+\infty}F(x,y)=1\end{aligned}$$

3 右连续性

对于单独的每一个变量都是右连续的，即：

$$\begin{aligned}F(x+0,y)&=\lim_{h\to0^+}F(x+h,y)=F(x,y)\\F(x,y+0)&=\lim_{h\to0^+}F(x,y+h)=F(x,y)\end{aligned}$$

4 非负性

对于任意$a<b$、$c<d$有：

$$\begin{aligned}&P(a<X\leq b,c<Y\leq d)\\=&F(b,d)-F(a,d)-F(b,c)+F(a,c)\geq 0\end{aligned}$$

3.1.4 多维均匀分布

设D为R^n中的一个有界区域，其度量（直线为长度，平面为面积，空间为体积等）为S_D，如果多维随机变量$(X_1,X_2,\cdots,X_n)$的联合概率密度函数为：

$$p(x_1,x_2,\cdots,x_n)=\begin{cases}\frac{1}{S_D}, & (x_1,x_2,\cdots,x_n)\in D\\ 0, & \text{其它}\end{cases}$$

则称$(X_1,X_2,\cdots,X_n)$服从D上的多维均匀分布，记作：

$$(X_1,X_2,\cdots,X_n)\sim U(D)$$

3.2 边缘分布与随机变量的独立性

3.2.1 边缘概率质量函数

如果二维离散随机变量(X,Y)的联合概率质量函数为：

$$P(X=x_i,Y=Y_j),i,j=1,2,\cdots$$

对j求和所得的函数：

$$\sum_{j=1}^{\infty}P(X=x_i,Y=y_j)=P(X=x_i)$$

称为X的边缘概率质量函数（Marginal Probability Mass Function），或者称为边缘分布列。类似的对i求和所得的函数：

$$\sum_{i=1}^{\infty}P(X=x_i,Y=y_j)=P(Y=y_j)$$

称为Y的边缘概率质量函数。

3.2.2 边缘概率密度函数

如果二维连续随机变量(X,Y)的联合概率密度函数为$p(x,y)$，则：

$$p_X(x)=\int_{-\infty}^{+\infty}p(x,y)\mathrm{d}y$$

称为X的边缘概率密度函数（Marginal Probability Density Function）。类似的：

$$p_Y(y)=\int_{-\infty}^{+\infty}p(x,y)\mathrm{d}x$$

称为Y的边缘概率密度函数。

3.2.3 边缘累积分布函数

如果二维连续随机变量(X,Y)的联合累积分布函数为$F(x,y)$，如下可以得到X的累积分布函数：

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

称为X的 边缘累积分布函数（Marginal Cumulative Distribution Function）。可记作：

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

同理可以得到Y的边缘累积分布函数：

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

3.2.4 随机变量的独立性

随机变量本身就代表了一系列事件，所以随机变量的独立性定义也和事件独立的定义类似：

- 从概率分布函数出发的定义：

设n维随机变量$(X_1,X_2,\cdots,X_n)$的联合累积分布函数为$F(x_1,x_2,\cdots,x_n)$，$F_i(x_i)$为X_i的边缘累积分布函数，如果对任意n个实数$x_1,x_2,\cdots,x_n$满足：

$$F(x_1,x_2,\cdots,x_n)=\prod_{i=1}^{n}F_i(x_i)$$

则称$X_1,X_2,\cdots,X_n$相互独立。

- 从概率质量函数出发的定义：

在离散随机变量场合，如果对任意n个实数$x_1,x_2,\cdots,x_n$满足：

$$P(X=x_1,X=x_2,\cdots,X=x_n)=\prod_{i=1}^{n}P(X=x_i)$$

则称$X_1,X_2,\cdots,X_n$相互独立。

- 从概率密度函数出发的定义：

在连续随机变量场合，如果对任意n个实数$x_1,x_2,\cdots,x_n$满足：

$$p(x_1,x_2,\cdots,x_n)=\prod_{i=1}^{n}p_i(x_i)$$

则称 $X_1,X_2,\cdots,X_n$相互独立。

3.3 多维条件分布

3.3.1 离散的条件分布

设(X,Y)是二维离散型随机变量，对于固定的j，若$P(Y=y_j)\geq 0$，则称：

$$P\left(X=x_i|Y=y_j\right)=\frac{P\left(X=x_i,Y=y_j\right)}{P\left(Y=y_j\right)},i=1,2,\cdots$$

为$Y=y_j$条件下的随机变量X的条件概率质量函数。同样的对于固定的i，若$P(X=x_i)\geq 0$，则称：

$$P\left(Y=y_j|X=x_i\right)=\frac{P\left(X=x_i,Y=y_j\right)}{P\left(X=x_i\right)},i=1,2,\cdots$$

为$X=x_i$条件下的随机变量Y的条件概率质量函数。

3.3.2 泊松分布的条件分布

问　设一段时间内进入某商店的顾客人数X服从泊松分布$P(\lambda)$：每个顾客购买某种物品的概率为p，并且各个顾客是否购买该种物品相互独立，求进入商店的顾客购买这种物品的人数Y的PMF。

答　可以凭直觉来思考下这个问题，$X\sim P(\lambda)$，也就是说一段时间内平均有λ个客人进入商店；购买某物品的概率为p，那么就是说一段时间内平均有λp个客人购买此物品，这就意味着$Y\sim P(\lambda p)$，我们来证明看看这个推论对吗？

根据题意有：

$$P(X=m)=\frac{\lambda^m}{m!}\mathrm{e}^{-\lambda},\quad m=0,1,2,\cdots$$

在$X=m$的条件下，且各个顾客购买该种物品相互独立，所以购买此物品的人数Y的条件分布为二项分布$b(m,p)$：

$$P(Y=k|X=m)=\binom{m}{k}p^k(1-p)^{m-k},\quad k=0,1,2,\cdots,m$$

由全概率公式可得：

$$
P(Y=k)=\sum_{m=k}^{\infty}P(X=m)P(Y=k|X=m)
$$

$$
\begin{aligned}
&=\sum_{m=k}^{\infty}\frac{\lambda^m}{m!}\mathrm{e}^{-\lambda}\cdot\frac{m!}{k!(m-k)!}p^k(1-p)^{m-k}\\
&=\mathrm{e}^{-\lambda}\sum_{m=k}^{\infty}\frac{\lambda^m}{k!(m-k)!}p^k(1-p)^{m-k}
\end{aligned}
$$

$$
=e^{-\lambda}\frac{(\lambda p)^k}{k!}\sum_{m=k}^{\infty}\frac{[(1-p)\lambda]^{m-k}}{(m-k)!}
$$

$$
=\frac{(\lambda p)^k}{k!}e^{-\lambda}e^{\lambda(1-p)}
$$

$$
=\frac{(\lambda p)^k}{k!}\mathrm{e}^{-\lambda p},\quad k=0,1,2,\cdots
$$

所以确实有：

$$
Y\sim P(\lambda p)
$$

3.3.3 连续的条件分布

设二维连续型随机变量(X,Y)的概率密度函数为$p(x,y)$，若对于固定的y有边缘概率密度函数$p_Y(y)>0$，则：

$$
p_{X|Y}(x\mid y)=\frac{p(x,y)}{p_Y(y)}
$$

为$Y=y$条件下的随机变量X的条件概率密度函数。对应的条件累积分布函数为：

$$
F_{X|Y}(x\mid y)=\int_{-\infty}^{x}\frac{p(u,y)}{p_Y(y)}\mathrm{d}u
$$

同样的道理，以$X=x$为条件有：

$$
p_{Y|X}(y\mid x)=\frac{p(x,y)}{p_X(x)}
$$

$$
F_{Y|X}(y\mid x)=\int_{-\infty}^{y}\frac{p(x,u)}{p_X(x)}\mathrm{d}u
$$

3.3.4 连续的全概率和贝叶斯

- 全概率

全概率公式的概率密度函数形式：

$$p_Y(y) = \int_{-\infty}^{+\infty} p(y|x)p_X(x)\mathrm{d}x$$

$$p_X(x) = \int_{-\infty}^{+\infty} p(x|y)p_Y(y)\mathrm{d}y$$

- 贝叶斯

贝叶斯公式的概率密度函数形式：

$$\begin{aligned} p(x|y) &= \frac{p(y|x)p_X(x)}{p_Y(y)} \\ &= \frac{p(y|x)p_X(x)}{\int_{-\infty}^{+\infty} p(y|x)p_X(x)\mathrm{d}x} \end{aligned}$$

3.4 多维随机变量函数的分布

3.4.1 随机变量的和

3.4.1.1 卷积

假设有两个函数f、g，称$(f*g)(n)$为两者的卷积，其定义为：

- 连续的情况：假设f、g的定义域为实数，有：

$$(f*g)(n) = \int_{-\infty}^{\infty} f(\tau)g(n-\tau)d\tau$$

- 离散的情况：假设f、g的定义域为离散的，有：

$$(f*g)(n) = \sum_{\tau=-\infty}^{\infty} f(\tau)g(n-\tau)$$

这两种定义有一个共同的特征：

$$(f*g)(n)=\int_{-\infty}^{\infty}f(\tau)g(n-\tau)d\tau$$

$$n=\tau+(n-\tau)$$

$$(f*g)(n)=\sum_{\tau=-\infty}^{\infty}f(\tau)g(n-\tau)$$

3.4.1.2 随机变量加和公式

- 离散

设X、Y为两个相互独立的离散随机变量，取值范围为0，1，2，$\cdots$，则其和的概率质量函数为：

$$P(X+Y=k)=\sum_{i=0}^{k}P(X=i)P(Y=k-i)$$

- 连续

设(X,Y)为二维连续型随机变量，概率密度函数为$p(x,y)$，则$Z=X+Y$仍为连续型随机变量，其概率密度为：

$$p_{X+Y}(z)=\int_{-\infty}^{+\infty}p(z-y,y)\mathrm{d}y=\int_{-\infty}^{+\infty}p(x,z-x)\mathrm{d}x$$

若X、Y为相互独立，其边缘密度函数分别为$p_X(x)$和$p_Y(y)$，则其和$Z=X+Y$的概率密度函数为：

$$p_Z(z)=\int_{-\infty}^{+\infty}p_X(z-y)p_Y(y)\mathrm{d}y=\int_{-\infty}^{+\infty}p_X(x)p_Y(z-x)\mathrm{d}x$$

3.4.2 常见分布的和

- 伯努利分布

假设相互独立的随机变量$X_1,X_2,\cdots,X_n$都服从伯努利分布：

$$P(X_i = x) = p(x) = \begin{cases} p, & x = 1 \\ 1-p, & x = 0 \end{cases}, \quad i = 1, 2, \cdots n$$

则它们的和服从二项分布：

$$Y = X_1 + X_2 + \cdots + X_n \sim b(n, p)$$

- 二项分布

假设相互独立的随机变量X, Y都服从二项分布：

$$X \sim b(n, p), \quad Y \sim b(m, p)$$

则它们的和也服从二项分布：

$$X + Y \sim b(n + m, p)$$

- 几何分布

假设相互独立的随机变量$X_1, X_2, \cdots, X_r$都服从几何分布$Ge(p)$，则它们的和服从负二项分布：

$$Y = X_1 + X_2 + \cdots + X_r \sim Nb(r, p)$$

- 正态分布

假设相互独立的随机变量X, Y都服从正态分布：

$$X \sim N\left(\mu_1, \sigma_1^2\right), \quad Y \sim N\left(\mu_2, \sigma_2^2\right))$$

则它们的和也服从正态分布：

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

小结

随机变量的和把之前很多常见分布都串起来了，从扔一枚硬币到扔多枚硬币：

$$\text{伯努利分布} + \text{伯努利分布} = \text{二项分布}$$

继续扔更多的硬币：

$$\text{二项分布} + \text{二项分布} = \text{二项分布}$$

硬币实在扔得太多的时候可以用正态分布来近似：

$$\text{二项分布} \xrightarrow{n\text{很大}} \text{正态分布}$$

最后不论怎么扔都始终是正态分布了：

$$\text{正态分布} + \text{正态分布} = \text{正态分布}$$

3.4.3 随机变量的商和积

设(X, Y)为二维连续型随机变量，概率密度函数为$p(x, y)$，则$Z = \dfrac{Y}{X}$、$Z = XY$仍为连续型随机变量，其概率密度分别为：

$$p_{Y/X}(z) = \int_{-\infty}^{+\infty} |x| p(x, xz) \mathrm{d}x$$

$$p_{XY}(z) = \int_{-\infty}^{+\infty} \frac{1}{|x|} p(x, \frac{z}{x}) \mathrm{d}x$$