Ch 5.4 随机变量的独立性

回顾前一次课

二维连续性随机向量、联合密度函数、边缘密度函数、性质

均匀分布 = (平面)几何概型

二维正太分布$(X,Y)\sim N(\mu_x,\mu_y,\sigma_x^2,\sigma_y^2,\rho)$及其边缘分布

$$\frac{1}{2\pi\sqrt{1-\rho^2}\sigma_x\sigma_y}\exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_x)^2}{\sigma_x^2}+\frac{(y-\mu_y)^2}{\sigma_y^2}-\frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right]\right)$$

随机变量X与Y相互独立: $F(x,y)=F_X(x)\,F_Y(y)$

- 离散随机变量: 独立性等价于$p_{ij}=p_{i\cdot}p_{\cdot j}$
- 连续随机变量：独立性等价于$f(x,y)=f_X(x)f_Y(y)$

性质

性质： 设随机变量X与Y相互独立，则$f(X)$与$g(Y)$相互独立（其中$f(x)$和$g(y)$是连续或分段连续函数）

性质： 如果存在函数$h(x)$ 和$g(y)$，使得随机变量X和Y的联合密度函数$f(x,y)$对任意实数x和y都有

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

则随机变量X和Y相互独立

正太分布的独立性

设二维随机向量$(X, Y) \sim \boldsymbol{N}\left(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho\right)$**, 则**

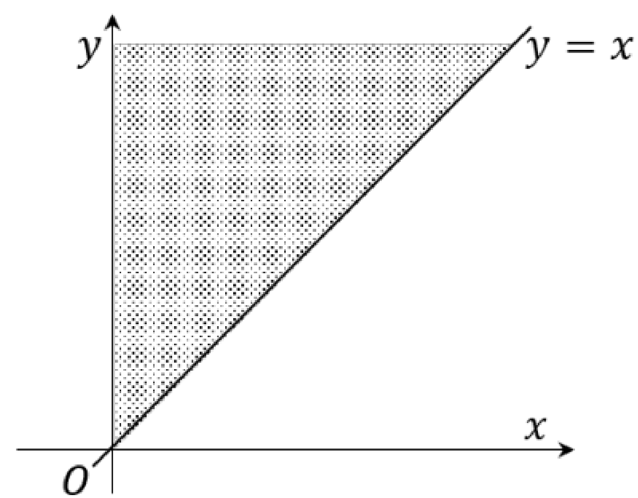
$$X\text{与}Y\text{独立的充要条件是}\rho = 0$$

例题

设二维随机向量(X,Y)的密度函数

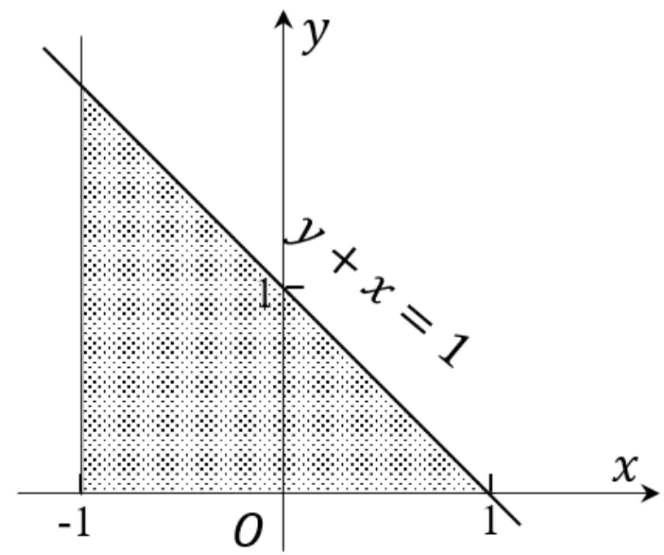
$$f(x,y)=\begin{cases} cxe^{-y} & 0<x<y<+\infty \\ 0 & \text{其它}, \end{cases}$$

问随机变量X与Y是否相互独立?



例题

设随机变量X与Y相互独立, 且X服从$[-1,1]$ 均匀分布, Y服从参数为$\lambda = 2$的指数分布, 求$P(X + Y \leq 1)$



Ch 5.5 条件分布

条件分布

前面学过随机事件的条件概率，即在事件B发生的条件下事件A发生的条件概率

$$P(A|B) = P(AB)/P(B)$$

推广到随机变量：给定随机变量Y取值的条件下，求随机变量X的概率分布, 即条件分布

离散型随机变量的条件分布

设二维离散型随机变量(X,Y)的分布列为$p_{ij} = P\left(X = x_i, Y = y_j\right)$,
若Y的边缘分布$P\left(Y = y_j\right) = p_{\cdot j} > 0$, 称

$$P\left(X = x_i \middle| Y = y_j\right) = \frac{P\left(X = x_i, Y = y_j\right)}{P\left(Y = y_j\right)} = \frac{p_{ij}}{p_{\cdot j}}$$

在$Y = y_j$条件下随机变量X的**条件分布列** (conditional probability distribution)

类似定义在$X = x_i$条件下随机变量Y的条件分布列

离散型随机变量的条件分布

- 非负性: $P\left(X=x_i \middle| Y=y_j\right) \geq 0$
- 规范性: $\sum_{i=1}^{\infty} P(X=x_i|Y=y_j)=1$
- 若离散随机变量X和Y相互独立, 则有

$$P\left(X=x_i \middle| Y=y_j\right)=P(X=x_i) \quad P(Y=y_j|X=x_i)=P(Y=y_j)$$

若出现条件概率$P(X=x_i|Y=y_j)$, 一般都默认$P(Y=y_j)>0$
条件分布列也可以通过下面的表格给出:

X	x_1	x_2	$\cdots$	x_n	$\cdots$
$P(X=x_i\|Y=y_j)$	$p_{1j}/p_{\cdot j}$	$p_{2j}/p_{\cdot j}$	$\cdots$	$p_{nj}/p_{\cdot j}$	$\cdots$

例题

一选手随机进行射击训练，击中目标的概率为p，射击进行到击中两次目标为止，用X表示首次击中目标的射击次数，用Y表示第二次射中目标的射击次数, 求X和Y的联合分布和条件分布.

连续随机变量的条件概率

对连续随机变量(X,Y)，对任意$x,y\in(-\infty,+\infty)$，有$P(X=x)=0$和$P(Y=y)=0$成立，因此不能用离散随机变量的条件概率推导连续随机变量的条件分布

以$f_{X|Y}(x|y)=f(x,y)/f_Y(y)$为例, 分布函数

$$F_{X|Y}(x|y)=\lim_{\epsilon\to 0^+}P[X\le x|\,y\le Y\le y+\epsilon]$$

$$=\lim_{\epsilon\to 0^+}\frac{P[X\le x,y\le Y\le y+\epsilon]}{P[y\le Y\le y+\epsilon]}$$

根据积分中值定理展开

连续随机变量的条件概率

定义： 设连续随机变量(X,Y)的联合概率密度为$f(x,y)$，以及Y的边缘概率密度为$f_Y(y)$, 对任意$f_Y(y)>0$, 称

$$f_{X|Y}(x|y)=f(x,y)/f_Y(y)$$

在$\boldsymbol{Y=y}$**条件下随机变量**$\boldsymbol{X}$**的条件概率密度**

$$F_{X|Y}(x|y)=P[X\le x|Y=y]=\int_{-\infty}^{x}f_{X|Y}(u|y)du$$

为$\boldsymbol{Y=y}$**条件下**$\boldsymbol{X}$**的条件分布函数**

条件密度函数的性质

条件密度函数本质上是密度函数, 具有以下性质:

- 非负性：对任意实数x, y有$f_{Y|X}(y|x) \geq 0$
- 规范性：对任意实数y: $f_Y(y) > 0$ 有

$$\int_{-\infty}^{+\infty} f_{X|Y}(x|y)dx = \int_{-\infty}^{+\infty} \frac{f(x,y)}{f_Y(y)} dx = \frac{1}{f_Y(y)} dx \int_{-\infty}^{+\infty} f(x,y)dx = 1$$

- 乘法公式:

$$f(x,y) = f_X(x) f_{Y|X}(y|x) \qquad f(x,y) = f_Y(y) f_{X|Y}(x|y)$$

- 对独立的随机变量X和Y, 有

$$f_{Y|X}(y|x) = f_Y(y) \qquad f_{X|Y}(x|y) = f_X(x)$$

条件密度函数的性质

根据条件概率的乘法公式有

$$f(x|y)=\frac{f(x,y)}{f_Y(y)}=\frac{f_{Y|X}(y|x)f_X(x)}{f_Y(y)}=\frac{f_{Y|X}(y|x)f_X(x)}{\int_{-\infty}^{+\infty}f(x,y)dx}=\frac{f_{Y|X}(y|x)f_X(x)}{\int_{-\infty}^{+\infty}f(y|x)f_X(x)dx}$$

可以将其看作**密度函数的贝叶斯公式**

如何构造二维随机向量的联合分布函数

- 根据实际问题或实际数据归纳为$f(x,y)$
- 根据随机变量的独立性有$f(x,y)=f_X(x)f_Y(y)$
- 根据乘法公式$f(x,y)=f_X(x)f_{Y|X}(y|x)$

二维正太分布的条件分布

设随机向量$(X,Y)\sim N\left(\mu_x,\mu_y,\sigma_x^2,\sigma_y^2,\rho\right)$，则随机变量$X$和$Y$的条件分布分别是正太分布

$$X\Big|_{Y=y}\sim N(\mu_x-\rho\sigma_x\left(y-\mu_y\right)/\sigma_y,\left(1-\rho^2\right)\sigma_x^2)$$

$$Y\Big|_{X=x}\sim N(\mu_y-\rho\sigma_y(x-\mu_x)/\sigma_x,\left(1-\rho^2\right)\sigma_y^2)$$

例题

设二维随机变量(X,Y)的概率密度

$$f(x,y)=\begin{cases}\dfrac{e^{-\frac{x}{y}}e^{-y}}{y} & 0<x<+\infty, 0<y<+\infty \\ 0 & \text{其它}\end{cases}$$

求$P(X>1|Y=y)$.

例题

已知随机变量$X \sim U(0,1)$，当观察到$X = x$的条件下，随机变量$Y \sim U(x, 1)$, 求Y的概率密度.