Ch 5.4 多维随机变量函数的分布

回顾前一次课

多维随机变量函数$\max(X,Y)$ 和 $\min(X,Y)$

$$f_{X+Y}(z)=\int_{-\infty}^{+\infty}f(x,z-x)dx$$

$$f_{XY}(z)=\int_{-\infty}^{+\infty}\frac{1}{|x|}f\left(x,\frac{z}{x}\right)dx$$

$$f_{Y/X}(z)=\int_{-\infty}^{+\infty}|x|f(x,xz)dx$$

$X\sim B(n_1,p)$和$Y\sim B(n_2,p)$独立, $X+Y\sim B(n_1+n_2,p)$

$X\sim P(\lambda_1)$ 和$Y\sim P(\lambda_2)$ 独立, 则$X+Y\sim P(\lambda_1+\lambda_2)$

$X\sim N(\mu_1,\sigma_1^2)$和$Y\sim N(\mu_2,\sigma_2^2)$ 独立, 则

$$X+Y\sim \mathrm{N}(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)$$

随机变量的联合分布函数

问题：

随机变量(X, Y)的联合概率密度$f(x, y)$, 设(X, Y)的函数

$$U = u(X, Y) \qquad V = v(X, Y)$$

如何求(U, V)的联合分布？

随机变量的联合分布函数

$u = u(x, y)$和$v = v(x, y)$有连续偏导，反函数$x = x(u, v)$和$y = y(u, v)$，$U = u(X, Y)$和$V = v(X, Y)$的联合密度为

$$f_{UV}(u, v) = f_{XY}\big(x(u, v), y(u, v)\big)|J|$$

其中J为变换的雅可比行列式，即

$$J = \left|\frac{\partial(x, y)}{\partial(u, v)}\right| = \left|\frac{\partial(u, v)}{\partial(x, y)}\right|^{-1} = \begin{vmatrix} \frac{\partial \mathrm{u}}{\partial \mathrm{x}} & \frac{\partial \mathrm{u}}{\partial y} \\ \frac{\partial \mathrm{v}}{\partial \mathrm{x}} & \frac{\partial \mathrm{v}}{\partial \mathrm{y}} \end{vmatrix}^{-1}$$

上述结论可推广到一般的n维随机变量

Ch 5.5 多维随机变量的数字特征

期望

离散随机变量(X,Y)的分布列为$p_{ij}=P\left(X=x_i,Y=y_j\right)$，则随机变量$\boldsymbol{Z}=\boldsymbol{g}(\boldsymbol{X},\boldsymbol{Y})$**的期望**为

$$\boldsymbol{E}[\boldsymbol{Z}]=\boldsymbol{E}[\boldsymbol{g}(\boldsymbol{X},\boldsymbol{Y})]=\sum_{\boldsymbol{i},\boldsymbol{j}}\boldsymbol{g}\left(\boldsymbol{x}_{\boldsymbol{i}},\boldsymbol{y}_{\boldsymbol{j}}\right)\boldsymbol{p}_{\boldsymbol{ij}}$$

连续随机变量(X,Y)的概率密度为$f(x,y)$，则随机变量$\boldsymbol{Z}=\boldsymbol{g}(\boldsymbol{X},\boldsymbol{Y})$**的期望**为

$$\boldsymbol{E}[\boldsymbol{Z}]=\boldsymbol{E}[\boldsymbol{g}(\boldsymbol{X},\boldsymbol{Y})]=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\boldsymbol{g}(\boldsymbol{x},\boldsymbol{y})\boldsymbol{f}(\boldsymbol{x},\boldsymbol{y})\boldsymbol{dxdy}$$

例题

设随机变量 $X \sim N(0,1)$ 和 $Y \sim N(0,1)$ 相互独立，求 $E[\max(X,Y)]$

在长度为1米的线段上任取两点X, Y, 求$E[|X-Y|]$

性质

对任意随机变量X,Y和常数a,b有

$$\boldsymbol{E[aX+bY]=aE[X]+bE[Y]}$$

对独立随机变量X和Y, 以及任意函数$h(x)$和$g(y)$, 有

$$\boldsymbol{E[XY]=E[X]E[Y]} \text{ 和 } \boldsymbol{E[h(X)g(Y)]=E[h(X)]E[g(Y)]}$$

对任意随机变量X和Y, 有Cauchy-Schwartz不等式

$$\boldsymbol{E[XY]\le\sqrt{E[X^2]E[Y^2]}}$$

协方差

定理： 对任意随机变量X与Y有

$$\mathbf{Var}(\boldsymbol{X}+\boldsymbol{Y})=\mathbf{Var}(\boldsymbol{X})+\mathbf{Var}(\boldsymbol{Y})+\mathbf{2}\boldsymbol{E}[(\boldsymbol{X}-\boldsymbol{E}(\boldsymbol{X}))(\boldsymbol{Y}-\boldsymbol{E}(\boldsymbol{Y}))]$$

特别地, 当X与Y独立时有

$$\mathbf{Var}(\boldsymbol{X}+\boldsymbol{Y})=\mathbf{Var}(\boldsymbol{X})+\mathbf{Var}(\boldsymbol{Y})$$

定义： 随机变量X和Y的协方差为

$$\mathbf{Cov}(\boldsymbol{X},\boldsymbol{Y})=\boldsymbol{E}[(\boldsymbol{X}-\boldsymbol{E}(\boldsymbol{X}))(\boldsymbol{Y}-\boldsymbol{E}(\boldsymbol{Y}))]=\boldsymbol{E}(\boldsymbol{XY})-\boldsymbol{E}(\boldsymbol{X})\boldsymbol{E}(\boldsymbol{Y})$$

协方差的性质

对任意随机变量X和常数c, 有$\mathrm{Cov}(X, c) = 0$.

交换律：$\mathrm{Cov}(X,Y)=\mathrm{Cov}(Y,X)$

对任意常数a和b, 随机变量X和Y, 有

$$\mathrm{Cov}(aX, bY) = ab\mathrm{Cov}(X, Y)$$

$$\mathrm{Cov}(X + a, Y + b) = \mathrm{Cov}(X, Y)$$

对任意常数X_1, X_2和Y, 有

$$\mathrm{Cov}(X_1 + X_2, Y) = \mathrm{Cov}(X_1, Y) + \mathrm{Cov}(X_2, Y)$$

协方差的性质

对随机变量$X_1, X_2, \dots, X_n$和$Y_1, Y_2, \dots, Y_m$, 有

$$\mathbf{Cov}\left(\sum_{\mathrm{i=1}}^{n} X_i, \sum_{j} Y_j\right) = \sum_{\mathrm{i=1}}^{n}\sum_{\mathrm{j=1}}^{m} \mathbf{Cov}(X_i, Y_j)$$

以及进一步有

$$\mathbf{Var}\left(\sum_{i}^{n} X_i\right) = \sum_{i}^{n} \mathbf{Var}(X_i) + 2\sum_{\mathrm{i}<j} \mathbf{Cov}(X_i, X_j)$$

协方差的性质

若随机变量X与Y独立, 则$\mathrm{Cov}(X,Y)=0$, 反之不成立.

对任意随机变量X与Y有

$$\left(\mathbf{Cov}(\boldsymbol{X},\boldsymbol{Y})\right)^{2}\le \mathbf{Var}(\boldsymbol{X})\mathbf{Var}(\boldsymbol{Y})$$

等号成立的充要条件是$\boldsymbol{Y}=\boldsymbol{aX}+\boldsymbol{b}$, 即$X$与$Y$之间存在线性关系

例题

随机变量X与Y独立，且$\mathrm{Var}(X)=6$和$\mathrm{Var}(Y)=3$，求$\mathrm{Var}(2X\pm Y)$.

随机变量$X\sim P(2)$和$Y\sim N(-2,4)$，且X与Y独立，则$E[(X-Y)^2]$.

相关性

根据前面的性质可知

$$-1 \le \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} \le 1$$

等号成立的充要条件是X与Y存在线性相关

上式一定程度上反应了随机变量X和Y的线性相关程度，由此引入一个新概念——相关系数

相关系数的定义

随机变量X和Y的方差$\mathrm{Var}(X)$,$\mathrm{Var}(Y)$存在且不为0, 称

$$\rho_{XY}=\frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$$

为***X与Y的相关系数***, 简记ρ

使用相关系数而不是$\mathrm{Cov}(X,Y)$, 主要是规范$|\rho_{XY}|\leq 1$, 而$\mathrm{Cov}(X,Y)$受数值大小影响

相关系数

相关系数$|\rho_{XY}| \leq 1$:　若$\rho > 0, X$与Y正相关;

若$\rho < 0$, X与Y负相关;

$|\rho_{XY}| = 1$的充要条件为X与Y有线性关系$Y = aX + b$

本质上$\boldsymbol{\rho_{XY}}$刻画了$\boldsymbol{X}$与$\boldsymbol{Y}$的线性相关程度, 又称为 "线性相关系数"

若相关系数$|\rho_{XY}| = 0$，则随机变量***X*和*Y*不相关**

X和Y不相关　　X和Y独立