2.11.2 连续随机变量的联合熵、条件熵以及平均互信息量

将微分熵的概念推广到多个连续随机变量，可以得到联合微分熵和统计微分熵，它们与普通微分熵一样，都只有相对意义。

一、联合微分熵

$$h(XY) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\log f_{XY}(x,y)dxdy$$

二、条件微分熵

$$h(X|Y) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\log f_{X|Y}(x|y)dxdy$$

三、连续随机变量的平均互信息量

$$I(X;Y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\log\frac{f_{XY}(x,y)}{f_X(x)f_Y(y)}dxdy$$

$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\log\frac{f_{X|Y}(x\mid y)}{f_X(x)}dxdy$$

注意：上式可用离散化取极限的方法严格推出，是精确的，不是舍弃无穷大项而取相对值。

四、关系　1、$I(X;Y)=h(X)-h(X|Y)$

推导：$I(X;Y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\log\frac{f_{X|Y}(x|y)}{f_X(x)}dxdy$

$$=-\int_{-\infty}^{\infty}\log f_X(x)dx\int_{-\infty}^{\infty} f_{XY}(x,y)dy$$

$$+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\log f_{X|Y}(x|y)dxdy$$

$$=-\int_{-\infty}^{\infty} f_X(x)\log f_X(x)dx$$

$$+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\log f_{X|Y}(x|y)dxdy$$

$$=h(X)-h(X|Y)$$

2、恒等关系：

$$h(XY) = h(X) + h(Y|X) = h(Y) + h(X|Y)$$

$$I(X;Y) = h(X) - h(X|Y) = h(Y) - h(Y|X)$$

$$= I(Y;X)$$ （互易性）

3、不等关系：

$$h(X|Y) \le h(X); \quad h(XY) \le h(X) + h(Y)$$

$$I(X;Y) \ge 0$$ （非负性）

例2.12　XY是二维正态随机变量，$E[X]=E[Y]=0, Var[X]=\sigma_1{}^2,$ $Var[Y]=\sigma_2{}^2, \rho=\dfrac{E[XY]}{\sigma_1\sigma_2}$。求$h(X), h(X|Y), I(X:Y)$.

解：根据概率论知识，联合概率密度函数为

$$f_{XY}(x,y)$$

$$=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}e^{-(x^2/\sigma_1^2+y^2/\sigma_2^2-2\rho xy/\sigma_1\sigma_2)/2(1-\rho)}$$

ρ为相关系数。求出边缘密度和条件密度：

$$f_X(x)=\int_{-\infty}^{\infty} f_{XY}(x,y)dy=\frac{1}{\sqrt{2\pi\sigma_1^2}}e^{-x^2/2\sigma_1^2}$$

$$f_Y(y)=\int_{-\infty}^{\infty} f_{XY}(x,y)dx=\frac{1}{\sqrt{2\pi\sigma_2^2}}e^{-y^2/2\sigma_2^2}$$

$$f_{X|Y}(x\mid y)=f_{XY}(x,y)/f_Y(y)$$

$$=\frac{1}{\sqrt{2\pi\sigma_1^2(1-\rho^2)}}e^{-(x-\frac{\sigma_1}{\sigma_2}\rho y)^2/2\sigma_1^2(1-\rho^2)}$$

再利用微分熵的定义式，得：

$$h(X) = \log\sqrt{2\pi e\sigma_1^2}$$

$$h(X\mid Y) = \log\sqrt{2\pi e\sigma_1^2(1-\rho^2)}$$

$$I(X;Y) = h(X) - h(X\mid Y)$$

$$= \log\sqrt{1/(1-\rho^2)}$$