Course 10

Joint Probablity Distribution

- **Discrete RV**
  - Joint pmf: $p(x,y)=P(X=x \text{ and } Y=y)$
  - $p(x,y)\geq 0$
  - $\sum\limits_x \sum\limits_y p(x,y)=1$
  - **Marginal pmf**:
    - $p_X(x)=\sum\limits_{y:p(x,y)>0} p(x,y)$
    - $p_Y(y)=\sum\limits_{x:p(x,y)>0} p(x,y)$
- **Continuous RV**
  - Joint pmf $f(x,y)=\displaystyle\int_A\int f(x,y)\mathrm{d}x\mathrm{d}y$
  - $f(x,y)\geq 0$
  - $\displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\mathrm{d}x\mathrm{d}y=1$
  - **Marginal pmf**:
    - $f_X(x)=\displaystyle\int_{-\infty}^{\infty} f(x,y)\mathrm{d}y,\ -\infty<x<\infty$
    - $f_Y(y)=\displaystyle\int_{-\infty}^{\infty} f(x,y)\mathrm{d}x,\ -\infty<y<\infty$
- **Independent RV**
  - $p(x,y)=p_X(x)\cdot p_Y(y)$, when X and Y are discrete
  - $f(x,y)=f_X(x)\cdot f_Y(y)$, when X and Y are continuous

More than two variable

- For discrete RV, $p(x_1,x_2,\ldots,x_n)=P(X_1=x_1,X_2=x_2,\ldots,X_n=x_n)$
- For continuous RV, $P(a_1\leq X_1\leq b_1,\ldots,a_n\leq X_n\leq b_n)=\displaystyle\int_{a_1}^{b_1}\cdots\int_{a_n}^{b^n} f(x_1\ldots,x_n)\mathrm{d}x_n\ldots\mathrm{d}x_1$

Conditional Distributions

Let X and Y be two continuous rv's with joint pdf $f(x,y)$ and marginal X pdf $f_X(x)$. Then for any X value x for which $f_X(x)>0$, the **conditional probability density function of Y given that $X=x$** is:

$$f_{Y|X}(y|x)=\frac{f(x,y)}{f_X(x)}\quad -\infty<y<\infty$$

If X and Y are discrete, then:

$$f_{Y|X}(y|x)=\frac{p(x,y)}{p_X(x)}\quad -\infty<y<\infty$$

is the conditional pmf of Y when $X=x$

Expected Value

Let X and Y be jointly distributed rv's with pmf $p(x,y)$ or pdf $f(x,y)$ according to whether the variables are discrete or continuous. Then the expected value of a function $h(X,Y)$, denoted by $E[h(X,Y)]$ or $\mu_{h(X,Y)}$, is given by:

$$E[h(X,Y)]=\begin{cases}\sum_X\sum_Y h(x,y)\cdot p(x,y) & \text{if X and Y are discrete}\\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x,y)\cdot f(x,y)\mathrm{d}x\mathrm{d}y & \text{if X and Y are continuous}\end{cases}$$

Covariance协方差

When two random variables X and Y are **not independent**, it is frequently of interest to assess **how strongly they are related to one another**.

The covariance between two rv's X and Y is:

$$\mathrm{Cov}(X,Y)=E[(X-\mu_X)(Y-\mu_X)]=\begin{cases}\sum_x\sum_y(x-\mu_X)(y-\mu_Y)p(x,y) & X,Y\text{ discrete}\\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_X)(y-\mu_Y)f(x,y)\mathrm{d}x\mathrm{d}y & X,Y\text{ continuous}\end{cases}$$

$$\mu_X = \begin{cases} \sum x p_X(x) & X \text{ is discrete} \\ \int_{\infty}^{\infty} x p_X(x) \mathrm{d}x & X \text{ is continuous} \end{cases}$$

$$\mu_Y = \begin{cases} \sum y p_Y(y) & Y \text{ is discrete} \\ \int_{\infty}^{\infty} y p_Y(y) \mathrm{d}y & Y \text{ is continuous} \end{cases}$$

- $\mathrm{Cov}(X, Y) = E(XY) - \mu_X \cdot \mu_Y$

Correlation标准化的协方差/相关性

The correlation coefficient of X and Y, denoted by $\mathrm{Corr}(X, Y)$, $\rho_{X,Y}$, or just ρ, is defined by

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y}$$

- It is a normalization of $\mathrm{Cov}(X, Y)$ (协方差的标准化)
- If a and c are either both positive or both negative, $\mathrm{Corr}(aX + b, cY + d) = \mathrm{Corr}(X, Y)$
- For any two rv's X and Y, $-1 \leq \mathrm{Corr}(X, Y) \leq 1$.

Homework

Section 5.1 9, 12, 18, 19

Section 5.2 24, 26, 33, 35