Conditional distributions and expectations

- Suppose that X and Y are discrete random variables.
  - For y such that $P(Y=y)>0$, the conditional expectation of $g(X,Y)$ given $Y=y$, denoted by $E(g(X,Y)|Y=y)$, is defined by
  $$E(g(X,Y)|Y=y)=\sum_{x:P(X=x|Y=y)>0} g(x,y)P(X=x|Y=y).$$
  - For y such that $P(Y=y)>0$, let $p_{X|Y=y}$ be a function defined by $p_{X|Y=y}(x)=P(X=x|Y=y)$ for all $x\in R$, then $p_{X|Y=y}$ is called the conditional PMF of X given $Y=y$, then
  $$E(g(X,Y)|Y=y)=\sum_{x:p_{X|Y=y}(x)>0} g(x,y)p_{X|Y=y}(x).$$
  That is, $E(g(X,Y)|Y=y)$ is the expectation with respect to the conditional PMF $p_{X|Y=y}$.
  - For y such that $P(Y=y)>0$, $Var(X|Y=y)$ is the variance with respect to the conditional PMF $p_{X|Y=y}$, that is,
  $$Var(X|Y=y)=E[(X-E(X|Y=y))^2|Y=y].$$
  - Let $h(y)=E(g(X,Y)|Y=y)$ for $y\in\{y:P(Y=y)>0\}$, then $h(Y)$ is the conditional expectation of $g(X,Y)$ given Y, denoted by $E(g(X,Y)|Y)$.
  - $Var(X|Y)=E[(X-E(X|Y))^2|Y]$.

- Example 1. Suppose that (X,Y) is a discrete random vector with joint PMF $p_{X,Y}$, where
$$p_{X,Y}(x,y)=\begin{cases}0.4 & \text{if } (x,y)=(0,2);\\ 0.4 & \text{if } (x,y)=(1,2);\\ 0.2 & \text{if } (x,y)=(4,3);\\ 0 & \text{otherwise.}\end{cases}$$
Find $E(X|Y=2)$, $E(X|Y=3)$, $Var(X|Y=2)$, $Var(X|Y=3)$, $E(X|Y)$ and $Var(X|Y)$.

  Ans. $E(X|Y=2)=0.5$, $E(X|Y=3)=4$, $Var(X|Y=2)=0.25$, $Var(X|Y=3)=0$.
$$\begin{aligned}E(X|Y) &= \begin{cases}0.5 & \text{if } Y=2;\\ 4 & \text{if } Y=3,\end{cases}\\ &= 0.5I_{\{2\}}(Y)+4I_{\{3\}}(Y).\end{aligned}$$
  and
$$Var(X|Y)=0.25I_{\{2\}}(Y).$$

- Suppose that X and Y are random vectors of dimensions k and m respectively. Suppose that (X, Y) has a joint PDF.
  - Suppose that f_Y is a PDF of Y and $R_Y = \{y \in R^m : f_Y(y) > 0\}$, and for $y \in R_Y$, h_y is a nonnegative function defined on R^k such that $\int_{R^k} h_y(x)dx = 1$, and

    $$P((X, Y) \in A \times B) = \int_{A \times B} f_Y(y)h_y(x)d(x, y)$$

    for $A \in \mathcal{B}(R^k)$, $B \in \mathcal{B}(R^m)$, then we say that $\{h_y : y \in R_Y\}$ is a version of the conditional PDF of X given Y.
  - Suppose that $f_{X,Y}$ is a PDF of (X, Y). For $y \in R^m$, define

    $$f_Y(y) = \int_{R^k} f_{X,Y}(x, y)dx.$$

    Let $R_Y = \{y \in R^m : f_Y(y) > 0\}$ and for $y \in R_Y$, define

    $$f_{X|Y=y}(x) = \frac{f_{X,Y}(x, y)}{f_Y(y)} \tag{1}$$

    for $x \in R^k$. Then $\{f_{X|Y=y} : y \in R_Y\}$ is a version of conditional PDF of X given Y.
  - Suppose that $\{f_{X|Y=y} : y \in R_Y\}$ is a version of conditional PDF of X given Y. Then

    $$E(g(X, Y)|Y) = \int_{R^k} g(x, y)f_{X|Y=y}(x)dx\bigg|_{y=Y}$$

    and $Var(X|Y) = E[(X - E(X|Y))^2|Y]$.

- Some properties of conditional expectation. Suppose that $g(X, Y)$, $g_1(X, Y)$, $g_2(X, Y)$ and $g_1(Y)g_2(X, Y)$ are integrable. Then we have
  - $E(E[g(X, Y)|Y]) = E(g(X, Y))$.
  - $E[(g_1(X, Y) + g_2(X, Y))|Y] = E[g_1(X, Y)|Y] + E[g_2(X, Y)|Y]$.
  - $E[g_1(Y)g_2(X, Y)|Y] = g_1(Y)E[g_2(X, Y)|Y]$.
  - $E[1|Y] = 1$.

- Example 2. Suppose that (X, Y) has a joint PDF $f_{X,Y}$, where

  $$f_{X,Y}(x, y) = \begin{cases} 2xy + 2(1 - x)(1 - y) & \text{if } (x, y) \in (0, 1) \times (0, 1); \\ 0 & \text{otherwise.} \end{cases}$$

  Find $E(X|Y)$.

Sol. Compute $f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx$, then

$$f_Y(y) = \begin{cases} \int_0^1 (2xy + 2(1-x)(1-y))dx = 1 & \text{if } y \in (0,1); \\ 0 & \text{if } y \notin (0,1), \end{cases}$$

and the set $\{y : f_Y(y) > 0\}$ is $(0,1)$. For $y \in (0,1)$,

$$\begin{aligned} \int_{-\infty}^{\infty} x \cdot \frac{f_{X,Y}(x,y)}{f_Y(y)} dx &= \int_{-\infty}^{\infty} x f_{X,Y}(x,y)dx \\ &= \int_0^1 x[2xy + 2(1-x)(1-y)]dx \\ &= \frac{2y}{3} + \frac{1-y}{3} = \frac{1+y}{3}, \end{aligned}$$

so

$$E(X|Y) = \left.\frac{1+y}{3}\right|_{y=Y} = \frac{1+Y}{3}.$$

- Fact 1 Suppose that $E(X^2) < \infty$. Then for $u(Y)$ such that $E[u^2(Y)] < \infty$,

  $$E[(X - E(X|Y))u(Y)] = 0$$

  and for $h(Y)$ such that $E[h^2(Y)] < \infty$,

  $$E[(X - h(Y))^2] = E[(X - E(X|Y))^2] + E[(E(X|Y) - h(Y))^2]. \tag{2}$$

  Remarks.

  - When predicting X using a function of Y, the best choice is $E(X|Y)$ in the sense that $E(X - h(Y))^2 \geq E(X - E(X|Y))^2$, which is implied by (2).
  - Take $h(Y) = E(X)$ in (2), then we have $Var(X) \geq Var(E(X|Y))$.
  - $E[(X - E(X|Y))^2] = E(Var(X|Y))$.

- Fact 2 Suppose that X and Y are random vectors of dimensions k and m respectively, and (X,Y) has a PDF. Suppose that $\{f_{X|Y=y}\colon y \in R_Y\}$ is a version of the conditional PDF of X given Y, and f_Y is a PDF of Y. Let

  $$g(x,y) = f_{X|Y=y}(x) f_Y(y) I_{R_Y}(y)$$

  for $x \in R^k$ and $y \in R^m$, then g is a PDF of (X,Y).

- Example 3. Suppose that X and Y are random variables, $f_Y(y) = 2yI_{(0,1)}(y)$ for $y \in R$ and f_Y is a PDF of Y. Suppose that for $y \in (0,1)$,

  $$f_{X|Y=y}(x) = \left(\frac{1}{2y}\right) I_{(-y,y)}(x)$$

for $x \in R$, and $\{f_{X|Y=y} : y \in (0,1)\}$ is a version of the conditional PDF of X given Y Find a PDF of X.

Sol. For $(x, y) \in R^2$, let

$$g(x, y) = f_{X|Y=y}(x) f_Y(y) I_{(0,1)}(y) = \left(\frac{1}{2y}\right) I_{(-y,y)}(x) 2y I_{(0,1)}(y),$$

then g is a PDF of (X, Y). Let

$$f_X(x) = \int_{-\infty}^{\infty} g(x, y) dy$$

for $x \in R$, then f_X is a PDF of X. The expression of $f_X(x)$ can be simplified to

$$f_X(x) = (1 - |x|) I_{(-1,1)}(x)$$

for $x \in R$.