Joint and Marginal Distributions

Random variables X and Y

Joint probability density function (pdf) $f_{X,Y}(x,y)$.

Cumulative distribution function (cdf) $F_Y(y)$ of Y:

$F_Y(c) = P(Y \le c)$

$= P(-\infty < X < \infty, Y \le c)$

= volume under the graph of $f_{X,Y}(x,y)$
above the region

$$R: \begin{cases} -\infty < x < \infty \\ y \le c \end{cases}$$

Set up the integral:

$F_Y(c) =$

$$\iint_R f_{X,Y}(x,y)dxdy =$$

$$\int_{-\infty}^{c} \left(\int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx \right) dy =$$

$$\int_{-\infty}^{c} g(y)\, dy,$$

where $g(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx$.

Thus: The pdf of Y is

$$f_Y(y) = F_Y'(y) = g(y)$$

In other words: the marginal pdf of Y is

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

In words:

Similarly, the marginal pdf of X is

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

Note: When X or Y is discrete, the corresponding integral becomes a sum.

Joint and Conditional Distributions:

Case I: X and Y both discrete.

Marginal pdf's (or pmf's = probability mass functions for discrete random variables):

$$f_Y(y) = P(Y = y) \qquad f_X(x) = P(X = x).$$

Joint pdf (pmf):

$$f_{X,Y}(x,y) = P(X = x \text{ and } Y = y).$$

Conditional pdf of Y|X:

$$f_{Y|X}(y|x) = P(Y = y|X = x)$$

$$= \frac{P(X = x \text{ and } Y = y)}{P(X = x)}$$

$$= \frac{f_{X,Y}(x,y)}{f_X(x)}.$$

In words:

Case II: Continuous X and Y:

Does $f_{Y|X}(y|x) = \dfrac{f_{X,Y}(x,y)}{f_X(x)}$ in this case also?

Def: $f_{Y|X}(y|x)$ is the function with

$$\int_c^d f_{Y|X}(y \mid a)\, dy$$

$= P(c \le Y \le d \mid X = a)$ for every a, c, d

So:

Does $\displaystyle\int_c^d \frac{f_{X,Y}(a,y)}{f_X(a)}\, dy$

$= P(c \le Y \le d \mid X = a)$ for every a, c, d?

Since

$$\int_c^d \frac{f_{X,Y}(a,y)}{f_X(a)}\, dy$$

$$= \int_{-\infty}^d \frac{f_{X,Y}(a,y)}{f_X(a)}\, dy - \int_{-\infty}^c \frac{f_{X,Y}(a,y)}{f_X(a)}\, dy$$

and

$P(c \le Y \le d \mid X = a)$

$= P(Y \le d \mid X = a) - P(Y \le c \mid X = a)$,

it's enough to show that

$$\int_{-\infty}^d \frac{f_{X,Y}(a,y)}{f_X(a)}\, dy = P(Y \le d \mid X = a)$$

for every a and d.

For small Δx,

$$P(Y \le d \mid X = a)$$

$$\approx P(Y \le d \mid a \le X \le a + \Delta x)$$

$$= \frac{P(Y \le d \text{ and } a \le X \le a + \Delta x)}{P(a \le X \le a + \Delta x)}$$

$$\approx \frac{P(Y \le d \text{ and } a \le X \le a + \Delta x)}{f_X(a)\Delta x}$$

$$= \frac{\int_{-\infty}^{d}\left(\int_{a}^{a+\Delta x} f_{X,Y}(x,y)\,dx\right)dy}{f_X(a)\Delta x}$$

$$\approx \frac{\int_{-\infty}^{d} f_{X,Y}(a,y)\Delta x\,dy}{f_X(a)\Delta x}$$

$$= \frac{\int_{-\infty}^{d} f_{X,Y}(a,y)\,dy}{f_X(a)}$$

$$= \int_{-\infty}^{d} \frac{f_{X,Y}(a,y)}{f_X(a)}\,dy$$, as desired.

Summarize: The conditional distribution Y|X has pdf

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

In words:

Conditional density of Y given X =

$$\frac{\textit{joint density of } X \textit{ and } Y}{\textit{marginal density of } X}$$

(Similarly for X|Y)