Section 1G. Two Random Variables

Statistics for Data Science

Victor M. Preciado, PhD MIT EECS
Dept of Electrical & Systems Engineering
University of Pennsylvania
preciado@seas.upenn.edu

Two Random Variables

Joint distributions for two r.v.'s X and Y are defined below:

- The ***joint CDF*** of two r.v.'s X and Y is defined as

$$F_{XY}(x, y) = \Pr(X \leq x, Y \leq y)$$

- The ***joint PDF*** is defined as $f_{XY}(x, y) = \partial^2 F_{XY}(x, y) / \partial x \partial y$. Hence, for a set $A \in \mathbb{R}^2$, we have that

$$\Pr((X, Y) \in A) = \iint_{(x,y) \in A} f_{XY}(x, y) \, dxdy$$

- The ***joint PMF*** of two discrete r.v.'s is defined by

$$p_{XY}(x, y) = \Pr(X = x, Y = y)$$

Two Random Variables: Marginalization

Marginal distributions are related to joint distributions as indicated below:

- The ***marginal CDF*** of X, F_X, satisfies

$$F_X(x) = \lim_{y \to \infty} F_{XY}(x, y)$$

- The ***marginal PDF*** (or marginal density) of X, f_X, satisfies

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

- The ***marginal PMF*** is given by

$$p_X(x) = \sum_{y \in \mathcal{Y}} p_{XY}(x, y)$$

  where $\mathcal{Y}$ is the set of values that Y can take.

Two Random Variables: Conditioning

The probability distribution of a r.v. Y given that you already know that X takes the exact value x is described using the conditional distribution.

- Assuming to discrete r.v.'s X and Y, the ***conditional PMF*** of Y given X is:

$$p_{Y|X}(y|x) = \frac{\Pr(X = x, Y = y)}{\Pr(X = x)} = \frac{p_{XY}(x, y)}{p_X(x)}$$

- Assuming to continuous r.v.'s X and Y, the ***conditional PDF*** of Y given X is:

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

Two Random Variables: Independence

- **Independence**: Two r.v.'s are independent if "knowing" the value of one variable does not affect the conditional probability distribution of the other variable.
  - Two discrete r.v.'s are independent if for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, we have

    $$p_{XY}(x, y) = p_X(x)\, p_Y(y)$$

    This is equivalent to $p_{X|Y}(x|y) = p_X(x)$ and $p_{Y|X}(y|x) = p_Y(y)$
  - Two continuous r.v.'s are independent if for all $x, y \in \mathbb{R}$

    $$f_{XY}(x, y) = f_X(x)\, f_Y(y)$$

    This is equivalent to $f_{X|Y}(x|y) = f_X(x)$ and $f_{Y|X}(y|x) = f_Y(y)$

Penn Engineering