Lecture 7 - Multivariate Random Variables

💡 We will consider multiple random variables defined over the same probability space Ω

Example

- Rolling two dice
- X - outcome on the first die
- Y - outcome on the second die

$$P(X = 1) = \frac{1}{6}$$
$$P(2 \leq Y \leq 3) = \frac{2}{6}$$
$$P(X \geq 4) = \frac{3}{6}$$
$$P(X = 2 \wedge Y = 3) = \frac{1}{36}$$
$$P(X = 2 | Y = 3) = \frac{1}{6}$$
$$P(X + Y = 10) = \frac{3}{36}$$

Joint distribution

- Let $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$ be discrete random variables defined over the same probability space.
- A joint distribution of X and Y specifies probabilities of simultaneously taking certain values by X and Y:
- Notation: $P(X = x, Y = y)$; this is the same as $P(\{X = x \wedge Y = y\})$
- When we want to calculate an event A that is a subspace of $\mathbb{R}^2$ we can use the joint distribution:

$$P((X, Y) \in A) = \sum_{(x,y) \in A} P(X = x, Y = y)$$

Marginal distribution

- Given a joint distribution of random variables $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$ we define marginal distributions:
  - With respect to X:

$$P(X = x) = \sum_{y \in \mathcal{Y}} P(X = x, Y = y)$$

  - With respect to Y:

$$P(Y = y) = \sum_{x \in \mathcal{X}} P(X = x, Y = y)$$

- Marginal distribution specifies probabilities associated with only one of the variables.
- Example: ($\sum$ symbol represents the marginal distribution)

	y_1	y_2	Σ
x_1	$P(X = x_1, Y = y_1)$	$P(X = x_1, Y = y_2)$	$P(X = x_1)$
x_2	$P(X = x_2, Y = y_1)$	$P(X = x_2, Y = y_2)$	$P(X = x_2)$
Σ	$P(Y = y_1)$	$P(Y = y_2)$	1

Conditional distribution

- Conditional distribution of a random variable X given $Y \in B$, where $P(Y \in B) > 0$, is a probability distribution defined for any $A \subseteq \mathbb{R}$ as:

$$P(X \in A | Y \in B) = \frac{P(X \in A, Y \in B)}{P(Y \in B)}$$

- If A and B contain only one number then the conditional distribution takes the form:

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

The law of total probability

- For discrete random variables $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$, where $P(Y = y) > 0$ for all $y \in \mathcal{Y}$ it holds:

$$P(X \in A) = \sum_{y \in \mathcal{Y}} P(X \in A|Y = y)P(Y = y)$$

Conditional expectation

- The quantity:

$$E(X|Y = y) = \sum_{x \in \mathcal{X}} xP(X = x|Y = y)$$

- is called conditional expected value (expectation) of a random variable X given $Y = y$.
- It is the mean value of X calculated using the conditional distribution $P(X = x|Y = y)$
- The conditional expected value has the same properties of the usual expected value, e.g.:
  - $E(g(X)|Y = y) = \sum_{x \in \mathcal{X}} g(x)P(X = x|Y = y)$
  - $E(aX + b|Y = y) = aE(X|Y = y) + b$
- Since $E(X|Y = n) = g(n)$ is a function of the value n taken by Y, we can treat is as a random variable which is a function of Y:

$$E(X|Y) = g(Y)$$

- From this we can also prove that:

$$E(E(X|Y)) = EX$$

More random variables

- Consider n random variables $X_1, X_2, \ldots, X_n$ defined over the same probability space.
- We can treat those variables as a multivariate random vector: $\mathbf{X} : \Omega \to \mathbb{R}^n$ where $\mathbf{X} = (X_1, X_2, \ldots, X_n)$
- In case of discrete random variables the joint distribution is fully specified by probabilities of the form:

$$P(\mathbf{X} = \mathbf{x}) = P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$$

- The marginal distribution is defined as:

$$P(X_i = a) = \sum_{\mathbf{x}:\ x_i = a} P(\mathbf{X} = \mathbf{x})$$

- One can also define marginal distributions with respect to a subset of random variables, e.g.:

$$P(X_i = a, X_j = b) = \sum_{\mathbf{x}:\ x_i=a, x_j=b} P(\mathbf{X} = \mathbf{x})$$

- The conditional distribution of a random vector $\mathbf{X}$ given $X_i = x_i$ is defined as:

$$P(\ldots, X_{i-1} = x_{i-1}, X_{i+1} = x_{i+1}, \ldots | X_i = x_i) = \frac{P(X_1 = x_1, \ldots, X_i = x_i, \ldots, X_n = x_n)}{P(X_i = x_i)}$$

- One can also define conditional distributions conditioning on a subset of random variables, e.g.:

$$P(X_1 = x_1, X_2 = x_2 | X_3 = x_3, X_4 = x_4) = \frac{P(X_1 = x_1, X_2 = x_2, X_3 = x_3, X_4 = x_4)}{P(X_3 = x_3, X_4 = x_4)}$$

- Analogously one can generalize to n variables:
  - The law of the total probability:

$$P(X_1 \in A) = \sum_{x_2, x_3} P(X_1 \in A | X_2 = x_2, X_3 = x_3) P(X_2 = x_2, X_3 = x_3)$$

  - Conditional expectation, e.g.:

$$E(X_1 | X_2 = x_2, X_3 = x_3) = \sum_{x_1} x_1 p(X_1 = x_1 | X_2 = x_2, X_3 = x_3)$$