POBABILITY AND STATISTICS, Lesson 5.

- The **Joint Distribution** of $X_1, \dots, X_n$ is given by the collection of probabilities $\mathbb{P}(\mathbf{X} \in B)$ ($B \in \mathcal{B}^n$), where $\mathbf{X} = (X_1, \dots, X_n)$ is random vector and $\mathcal{B}^n$ denotes the set of Borel-sets of $\mathbb{R}^n$. The rv's $X_1, \dots, X_n$ are *independent*, if

$$\mathbb{P}(X_1 \in B_1, \dots, X_n \in B_n) = \prod_{i=1}^{n} \mathbb{P}(X_i \in B_i), \qquad \forall B_1, \dots, B_n \in \mathcal{B}.$$

- *Special types of random vectors (X, Y) (the $n = 2$ case):*

  1. **Discrete joint distributions**: X takes on values $x_1, x_2, \dots$ and Y takes on values $y_1, y_2, \dots$. The distribution of (X, Y) is given by the joint p.m.f.

  $$p_{ij} = \mathbb{P}(X = x_i, Y = y_j), \qquad i = 1, 2, \dots; \quad j = 1, 2, \dots,$$

  where $\sum_i \sum_j p_{ij} = 1$.
  The *marginal distribution* of X is $p_{i.} = \sum_j p_{ij}, \quad i = 1, 2, \dots$.
  The *marginal distribution* of Y is $p_{.j} = \sum_i p_{ij}, \quad j = 1, 2, \dots$
  X and Y are *independent* if and only if $p_{ij} = p_{i.}p_{.j}$, $\forall i, j$.
  The *mode* of (X, Y): the value(s) taken with the largest probability.

  2. **Absolutely continuous joint distributions**: The range of (X, Y) is not countable and for any $(x, y) \in \mathbb{R}^2$: $\mathbb{P}((X, Y) = (x, y)) = 0$. Still, there is an $f : \mathbb{R}^2 \to \mathbb{R}$ nonnegative, integrable function such that

  $$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dx\, dy = 1 \quad \text{and} \quad \iint_B f(x, y)\, dx\, dy = \mathbb{P}((X, Y) \in B), \quad \forall B \in \mathcal{B}^2.$$

  f is called *joint p.d.f.* of (X, Y).
  The *marginal distribution* of X is given by the p.d.f. $f_1(x) = \int_{-\infty}^{\infty} f(x, y)\, dy$.
  The *marginal distribution* of Y is given by the p.d.f. $f_2(y) = \int_{-\infty}^{\infty} f(x, y)\, dx$.
  X and Y are *independent* if and only if $f(x, y) = f_1(x) f_2(y)$, $\forall (x, y) \in \mathbb{R}^2$.

- **Conditional distributions, conditional expectation**

  1. The conditional distribution of Y given $X = x_i$ is:

  $$\mathbb{P}(Y = y_j | X = x_i) = \frac{p_{ij}}{p_{i.}}, \qquad j = 1, 2, \dots$$

  and the conditional expectation of Y given $X = x_i$ is:

  $$\mathbb{E}(Y | X = x_i) = \sum_j y_j \frac{p_{ij}}{p_{i.}} = \frac{1}{p_{i.}} \sum_j y_j p_{ij}, \quad i = 1, 2, \dots$$

  2. The conditional distribution of Y given $X = x$ is given by the p.d.f.

  $$f_{2|1}(y|x) = \frac{f(x, y)}{f_1(x)}, \qquad y \in \mathbb{R}$$

  and the conditional expectation of Y given $X = x$ is:

  $$\mathbb{E}(Y | X = x) = \int_{-\infty}^{\infty} y f_{2|1}(y|x)\, dy = \frac{1}{f_1(x)} \int_{-\infty}^{\infty} y f(x, y)\, dy = g(x), \quad x \in \mathbb{R},$$

  where g is the *regression function*. Hence, $\mathbb{E}(Y|X) = g(X)$.
  Optimum property of the conditional expectation: $\mathbb{E}(Y - t(X))^2 \geq \mathbb{E}(Y - \mathbb{E}(Y|X))^2$ for any measurable $t : \mathbb{R} \to \mathbb{R}$ (least square approximation).