2.10 Independence of Random Variables

Consider two events $\{X = x\}$ and $\{Y = y\}$.

Definition 8 *Two random variables X and Y are independent if*

$$P(\{X = x\} \cap \{Y = y\}) = P(\{X = x\})P(\{Y = y\})$$

for all x, y.

Intuitively speaking, knowledge on Y conveys no information on X, and vice versa.

Independence of two random variables conditioned on an event A: $p_{X,Y|A}(x, y) = p_{X|A}(x)p_{Y|A}(y)$ for all x, y.

When X and Y are independent, $E[XY] = E[X]E[Y]$.

If X and Y are independent, so are $g(X)$ and $h(Y)$.

The independence definition given above can be extended to multiple random variables in a straightforward way. For example, three random variables X, Y, Z are independent if:

2.10.1 Variance of the Sum of Independent Random Variables

Let us calculate the variance of the sum $X + Y$ of two independent random variables X, Y.

If one repetitively uses the above result, the general formula for the sum of independent random variables is obtained:

Ex: During April in Ankara, it rains with probability p each day, independently of every other day. Compute the variance of the number of rainy days in the month. Consider how the variance changes with p.

Ex: Show that, when $E(XY) = E(X)E(Y)$ is satisfied, then the variance of the sum $X + Y$ is equal to the sum of the variances, that is:

$$E(XY) = E(X)E(Y) \rightarrow \text{var}(X + Y) = \text{var}X + \text{var}Y$$

- Note that $E(XY) = E(X)E(Y)$ always holds when X and Y are independent. In general, when $E(XY) = E(X)E(Y)$ holds, the random variables are said to be "uncorrelated", they are not necessarily independent.

- Also note that in contrast, expectation is always linear, expectation of the sum is equal to the sum of expectations:

$$E[X+Y] = E[X] + E[Y]$$

  This is true whether the random variables are dependent or not.

Ex: The number of e-mail messages I get every day is Poisson distributed with mean 10. Let L be the total number of e-mail messages I receive in a week. Compute the mean and variance of R.

Ex: (Mean and variance of the sample mean) An opinion poll is conducted to determine the average public opinion on an issue. It is modelled that a person randomly selected from the society will vote in favour of the issue with probability p, and against it with probability $1 - p$, independently of everyone else. The goal of the survey is to estimate p. To keep the cost of the poll at a minimum, we are interested in surveying the smallest number of people such that the variance of the result is below 0.001. (Hint: Note that an upperbound on the variance of a Bernoulli random variable is $1/4$.)