Anna Janicka

Probability Calculus 2021/2022
Lecture 3

1. Independence of Events

In common language, we often say that an event influences (or does not influence) another event, which we intuitively understand as the fact that the occurrence of one event changes (or does not change) our forecast about the other event. In probabilistic terms, if we were to formalize the notion that the knowledge that one event (A) occurred does not change the perceived probability of the occurrence of a second event (B), we could write:

$$\mathbb{P}(B|A) = \mathbb{P}(B).$$

This equation may be rewritten as

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot P(B),$$

which then leads to the conclusion that (assuming that also $\mathbb{P}(B) > 0$),

$$\mathbb{P}(B|A) = \mathbb{P}(A),$$

which we understand as a constatation that the knowledge that the event B occurred does not change anything in our perception of the probability that A occurred, either. Therefore, the mathematical understanding of independence is symmetrical, and the most useful form is the middle one (which works also for events with null probability):

Definition 1. *Events A and B are* **independent***, if $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot P(B)$.*

Due to the fact that this notion is not exactly the same as the intuitive understanding of independence (which is not symmetrical),[1] events which are independent according to the above definition are sometimes referred to as *stochastically independent*.

The following examples illustrate the mathematical understanding of independence.

(1) For a throw of one die, we may define the following events: A – an even number was obtained; B – a number not greater than 4 was obtained; C – a number not greater than 5 was obtained. The probabilities of these events are: $\mathbb{P}(A) = \frac{3}{6}$, $\mathbb{P}(B) = \frac{4}{6}$, $\mathbb{P}(C) = \frac{5}{6}$. If we look at event B, we see that the proportion of odd and even outcomes from the initial sample space Ω is not upset; therefore, we expect that events A and B will be independent, and indeed: $\mathbb{P}(A \cap B) = \frac{2}{6} = \frac{1}{2} \cdot \frac{2}{3} = \mathbb{P}(A) \cdot \mathbb{P}(B)$ – these events are independent. However, If we look at event C, the proportion of odd and even outcomes is changed; we therefore expect that events A and C will not be independent, and indeed: $\mathbb{P}(A \cap C) = \frac{2}{6} \neq \frac{1}{2} \cdot \frac{5}{6} = \mathbb{P}(A) \cdot \mathbb{P}(C)$ – these events are not independent. Also, if we know that a number not greater than 4 was obtained, it is obvious that a number not greater than 5 was obtained, so events B and C should not be independent – and indeed: $\mathbb{P}(B \cap C) = \frac{4}{6} \neq \frac{4}{6} \cdot \frac{5}{6} = \mathbb{P}(B) \cdot \mathbb{P}(C)$.

(2) We draw a card from a deck of cards. Let A – a club was obtained and B – a figure was obtained. Events A and B are independent: $\mathbb{P}(A \cap B) = \frac{4}{52} = \frac{13}{52} \cdot \frac{16}{52} = \mathbb{P}(A) \cdot \mathbb{P}(B)$. But if C – a black card was drawn, then A and C are not independent, while B and C are.

We may wish to compare not only pairs of events, but also larger groups of events. In this case, the definition becomes slightly more complicated:

Definition 2. *Events $A_1, A_2, \ldots, A_n$ are* **independent***, if for all indices $1 \leqslant i_1 < i_2 < \ldots < i_k \leqslant n$, $k = 2, 3, \ldots, n$, we have*

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \cdot \mathbb{P}(A_{i_2}) \cdot \ldots \cdot \mathbb{P}(A_{i_k}).$$

[1] The common-sense intuitive understanding of dependence is usually unidirectional: we will say that whether or not we will decide to take an umbrella with us when going out depends on the weather outside; however, we will not say that the weather outside depends on our decision to take an umbrella.

That is, to check independence of a group of events we must check the independence of all subgroups consisting of these events.

(1) We randomly draw a number from 1 to 30. Let A – an even number was obtained; B – a number divisible by 3 was obtained; C – a number divisible by 5 was obtained. We have $\mathbb{P}(A) = \frac{1}{2}$, $\mathbb{P}(B) = \frac{1}{3}$ and $\mathbb{P}(C) = \frac{1}{5}$. To check the independence of these events, we now have to check if
$\mathbb{P}(A \cap B) = \frac{1}{6} = \frac{1}{2} \cdot \frac{1}{3} = \mathbb{P}(A) \cdot \mathbb{P}(B)$,
$\mathbb{P}(A \cap C) = \frac{1}{10} = \frac{1}{2} \cdot \frac{1}{5} = \mathbb{P}(A) \cdot P(C)$,
$\mathbb{P}(B \cap C) = \frac{1}{15} = \frac{1}{3} \cdot \frac{1}{5} = \mathbb{P}(B) \cdot \mathbb{P}(C)$, and
$\mathbb{P}(A \cap B \cap C) = \frac{1}{30} = \frac{1}{2} \cdot \frac{1}{3} \cdot \frac{1}{5} = \mathbb{P}(A) \cdot \mathbb{P}(B) \cdot \mathbb{P}(C)$.
Since all equalities hold, these events are independent.

It is natural to ask if the definition of independence of a group of n events needs to be so complicated. Unfortunately, it does, which may be illustrated with the following examples:

(2) There are four balls in a box: one of them is red, one of them is blue, one of them is green, and the fourth one has red, blue and green stripes. Let A – drawing a ball with red, B – drawing a ball with blue, C – drawing a ball with green. Then $\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = \frac{1}{2}$. Also,

$$\mathbb{P}(A \cap B) = \mathbb{P}(A \cap C) = \mathbb{P}(B \cap C) = \mathbb{P}(\text{ball with stripes}) = \frac{1}{4} = \mathbb{P}(A) \cdot \mathbb{P}(B) = \mathbb{P}(A) \cdot \mathbb{P}(C) = \mathbb{P}(B) \cdot \mathbb{P}(C).$$

However, $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(\text{ball with stripes}) = \frac{1}{4} \neq \frac{1}{8} = \mathbb{P}(A) \cdot \mathbb{P}(B) \cdot \mathbb{P}(C)$. Events A, B, C are not independent; they are only **pairwise independent**.

(3) If $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A) \cdot \mathbb{P}(B) \cdot \mathbb{P}(C)$, then events A, B, and C do not need to be pairwise independent (and independent): let $\Omega = \{1, 2, \ldots, 8\}$, and $A = \{1, 2, 3, 4\} = B$, while $C = \{1, 5, 6, 7\}$. Then, $\mathbb{P}(A \cap B \cap C) = \frac{1}{8} = \mathbb{P}(A) \cdot \mathbb{P}(B) \cdot \mathbb{P}(C)$, but A and B are obviously not independent.

The following theorem in many cases allows to simplify the verification of the $2^n - n - 1$ equalities when checking the independence of a group of n events.

Theorem 1. *Let $A_1, A_2, \ldots, A_n$ be a sequence of events, and denote $A_i^0 = A_i, A_i^1 = A_i'$. The following conditions are equivalent:*

(i) *events $A_1, A_2, \ldots, A_n$ are independent,*
(ii) *for any sequence $\varepsilon_1, \ldots, \varepsilon_n$, where $\varepsilon_i \in \{0, 1\}$ $(i = 1, \ldots, n)$, events $B_1 = A_1^{\varepsilon_1}, \ldots, B_n = A_n^{\varepsilon_n}$ are independent,*
(iii) *for any sequence $\varepsilon_1, \ldots, \varepsilon_n$, where $\varepsilon_i \in \{0, 1\}$ $(i = 1, \ldots, n)$, we have*

$$\mathbb{P}(A_1^{\varepsilon_1} \cap \ldots \cap A_n^{\varepsilon_n}) = \mathbb{P}(A_1^{\varepsilon_1}) \cdot \ldots \cdot \mathbb{P}(A_n^{\varepsilon_n}).$$

In particular, from the above theorem it follows that if events A and B are independent, then events A' and B' (or A and B', or A' and B) are also independent.

2. The Bernoulli Process

In many real-life situations (some more obvious than other), we deal with a sequence of repetitions of an experiment with two possible outcomes: a sequence of coin flips, an opinion poll (with two possible answers) etc. If the subsequent repetitions of the experiment are independent, we are in a position to describe this situation with the use of a Bernoulli process.

Definition 3. *A* **Bernoulli process** *is a sequence of n independent repetitions of a single experiment (referred to as a* Bernoulli trial*) with two possible outcomes: one of these outcomes is referred to as a* success *(usually denoted as 1), and occurs with probability $p \in [0, 1]$, and the other one is a failure (usually denoted as 0), and occurs with probability $q = 1 - p$.*

If we flip a symmetric coin 10 times, then we will have a Bernoulli process with 10 trials and a probability of success in a single trial equal to $\frac{1}{2}$. If we roll a die 11 times and are interested in the outcome being a six, then we have a Bernoulli process with 11 trials and a probability of success in a single trail equal to $\frac{1}{6}$, etc.

A natural question which arises when dealing with a Bernoulli proces is: how many successes were there, and what was the probability of this type of outcome? In order to be able to answer this question, we have to define a probability space, a σ-algebra and a probability function for the process. The first part is easy: we take $\Omega = \{(x_1, x_2, \ldots, x_n) : x_i \in \{0,1\}\}$. Since this set is finite, we may very well take $\mathcal{F} = 2^\Omega$. As far as probability is concerned, it is enough to assign probabilities to elementary events (as we have shown, all other may be derived easily). If the trials are to be independent, then for an $\omega = (x_1, x_2, \ldots, x_n)$, we will take $\mathbb{P}(\omega) = p^{x_1+x_2+\cdots+x_n}(1-p)^{n-x_1-x_2-\cdots-x_n} = p^k(1-p)^{n-k}$, where k is the number of successes among the n trials for this event.

A simple consequence of the above definition of probability over elementary events in a Bernoulli process is the fact that the probability of obtaining exactly k successes in the n trials will be equal to

$$\binom{n}{k} p^k (1-p)^{n-k}.$$

The definition of a Bernoulli process may easily be extended to an infinite sequence; this simple extension, however, has serious repercussions in terms of the definition of probability for the process, which is not as simple anymore (the set of elementary events is uncountable).

Examples.

(1) What is the probability of obtaining 5 heads (successes) in a sequence of 10 symmetric coin flips? We have a Bernoulli process with $n = 10$, $p = \frac{1}{2}$, and we search for $k = 5$, therefore the requested probability is $\binom{10}{5}\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^{10-5} = \frac{63}{256} \approx 0.246$

(2) What is the probability of obtaining no more than a single six in a sequence of 11 rolls of a cubic die? We have a Bernoulli process with $n = 11$, $p = \frac{1}{6}$, and we ask for $k = 0, 1$, therefore the requested probability is $\binom{11}{0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^{11} + \binom{11}{1}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^{10} \approx 0.431$.

(3) For a Bernoulli process of n trials and a probability of success p, what is the most probable number of successes? Let us denote by p_k the probability that exactly k successes will occur, $p_k = \binom{n}{k}p^k(1-p)^{n-k}$. In order to determine the maximum, we may look at the ratios $\frac{p_{k+1}}{p_k}$ for different values of k. If for some k the ratio is greater than 1, then the probability of obtaining $k+1$ successes is higher than the probability of obtaining k successes; if the ratio is smaller than 1, then this probability decreases with a unit increase of k. We have

$$\frac{p_{k+1}}{p_k} = \frac{\binom{n}{k+1}p^{k+1}(1-p)^{n-(k+1)}}{\binom{n}{k}p^k(1-p)^{n-k}} = \frac{(n-k)p}{(k+1)(1-p)},$$

and thus $\frac{p_{k+1}}{p_k} > 1$ if $k < (n+1)p - 1$, and $\frac{p_{k+1}}{p_k} < 1$ if $k > (n+1)p - 1$. In order to find the maximum, we must therefore determine if the value $(n+1)p$ is an integer or not (if there may be any equalities instead of inequalities). The final answer is thus the following: if $(n+1)p$ is an integer, then there are two values of the number of successes which are most probable, i.e. $(n+1)p - 1$ and $(n+1)p$. If $(n+1)p$ is not an integer, there is only one such value, and this value is equal to the floor (smallest integer not greater than) of $(n+1)p$, i.e. $\lfloor (n+1)p \rfloor$.

In the case of the two examples above, we have: for a sequence of 10 symmetric coin flips, $(n+1)p = \frac{11}{2}$ which is not an integer, and the most probable number of successes is $\lfloor 5.5 \rfloor = 5$. For a sequence of 11 die rolls, we have $(n+1)p = \frac{12}{6} = 2$, and there are two most probable values of the number of successes: 1 and 2.

(4) What is the probability that in an infinite sequence of coin flips, there will only be heads? In order to answer this question we should either define probability for an infinite Bernoulli process (which we won't do), or make use of some reasoning using limits. If we use the rule of continuity, the problem is easy: let A_∞ denote the event

that only heads occur in an infinite sequence of flips, and let A_n denote the event that only heads occur in a finite n-trial Bernoulli process. Then, $A_\infty = \bigcap_{n=1}^{\infty} A_n$, and using the rule of continuity, we may write $\mathbb{P}(A_\infty) = \lim_{n\to\infty} \mathbb{P}(A_n) = \lim_{n\to\infty} p^n$. Now, if only $p < 1$ (obtaining a head in a single trial is not a sure event), then this limit is equal to 0; i.e., the probability that an infinite sequence of heads will occur is 0.

3. Poisson Theorem

In case of a Bernoulli process with a large number of trials, the problem of calculating the exact value of the probability that exactly k successes will occur becomes complicated (technically, due to the factorials and powers). If the product $n \cdot p$ is *moderate*, this calculation may be simplified by using the Poisson theorem.

Theorem 2. *If $p_n \in [0,1]$, $\lim_{n\to\infty} np_n = \lambda > 0$, then for $k = 0, 1, 2, \ldots$, we have that $\lim_{n\to\infty} \binom{n}{k} p_n^k (1-p_n)^{n-k} = \frac{\lambda^k}{k!} e^{-\lambda}$.*

In common applications, this limit theorem is used as an approximation for large n. λ is interpreted as the mean number of successes in n trials. Immediately, however, the question suggests itself: how good is this approximation for a given pair of n and p (and thus λ)? The following theorem provides the answer:

Theorem 3. *Let S_n denote the number of successes in a Bernoulli process with n trials and a probability of success in a single trial equal to p, and let $\lambda = np$. For any $A \subset \{0, 1, 2, \ldots\}$, we have*

$$\left| \mathbb{P}(S_n \in A) - \sum_{k\in A} \frac{\lambda^k}{k!} e^{-\lambda} \right| \leqslant \frac{\lambda^2}{n}.$$

Examples.

(1) There are 999 white balls and one black ball in a box. We draw a ball 500 times (and always put it back). What is the probability that we will draw a black ball twice? The experiment may be described by means of a Bernoulli process with $n = 500$ trials and a probability of success in a single trial equal to $p = \frac{1}{1000}$. We therefore have the mean number of successes equal to $\lambda = 500 \cdot \frac{1}{1000} = \frac{1}{2}$. The requested probability is $\binom{500}{2} \left(\frac{1}{1000}\right)^2 \left(\frac{999}{1000}\right)^{498}$, which we may approximate by $\approx \frac{\frac{1}{2}^2}{2!} e^{-\frac{1}{2}} \approx 0.075816$. The error we make is, up to the absolute value, not greater than $\frac{\lambda^2}{n} = \frac{1}{2000} = 0.0005$, which looks good. (In truth, the value of the probability when calculated directly from the Bernoulli process formula is equal to ≈ 0.075797.)

(2) From the interval $[0, 2]$ we randomly choose 100 points. What is the probability that at least two will fall into the interval $[0, \frac{1}{4}]$? The experiment may be described by a Bernoulli process with $n = 100$ trials and $p = \frac{1}{8}$. We have $\lambda = \frac{100}{8} = 12.5$. The event that at least two successes will occur may be calculated from the probability of the complementing event that no successes will occur or one success will occur. This complementing event has the probability equal to $\left(\frac{7}{8}\right)^{100} + 100 \cdot \frac{1}{8} \cdot \left(\frac{7}{8}\right)^{99}$, which would be approximated with the use of the Poisson theorem as $\approx e^{-12.5}(1 + 12.5) \approx 0.0000503$, and thus the probability of the main event of interest would be approximated as 0.9999497. However, the analysis of possible error leads to the conclusion that it may be as large as $\frac{12.5^2}{100} = 1.5625$, and this means that the approximation is possibly useless. This results from the fact that λ was not "moderate" in comparison with n. (In truth, the approximation is not that bad, as the value calculated directly from the Bernoulli process formula is equal to ≈ 0.0000243, but note that here the first significant figure is incorrect, and if a different set A was chosen, the result could be worse.)