Probability Calculus 2021/2022
Lecture 2

1. An Additional Example and Some More Properties of Probability

We will start with another example of a probabilistic model, this time infinite – and continuous.

(6) A large (and useful) class of experiments deal with situations where a point is drawn from an area within the plane, or the section of the line, a cube etc. In this case, contrary to the countable case, we can not assign a non-negative probability to the points within the area in question – there are too much (uncountably many) of them, and thus a sum of positive probabilities could not converge to 1. We must therefore revise our thinking about what results we should (and may) really be interested in, and move from specific points to whole sets of points; only then, we will be able to assign non-negative probability (connected to the measure of the studied area).

In the case of areas, it is natural to assume that the measure used as probability should be the measure dedicated to the specific dimension of the sample space; if we wish to study experiments where a point is drawn from a section of the line, we will use a single-dimensional measure, i.e. length; if we wish to study experiments where a point is drawn from an area of the plane, we will be using area; in a three dimensional space – volume, etc. The most natural way to define probability is then assuming that it is proportional to the (relative) measure of the area, i.e.

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

where $|\cdot|$ is the measure of a set. We must be careful, however, because not all sets are measurable, which means that the above definition of probability will not be suitable for unmeasurable sets. This is not a serious problem, however – it will suffice to constrain our σ-algebra (i.e. the set of events of interest to us) to (sub)sets which are measurable. The most commonly used σ-algebra in this context is the σ-algebra of Borel sets (over Ω), denoted by $\mathcal{B}(\Omega)$ – a σ-algebra generated by all open subsets of Ω (i.e. including those sets and all their possible complements, intersections, unions, the complements of those intersections and unions, etc.).

A useful property of probability in both the countable and uncountable infinite cases is the behavior of the probability function in the case of sequences of events. We will specifically describe two types of sequences of events:

Definition 1. *Assume $A_1, A_2, \ldots$ is a sequence of events. We will call this sequence* **expanding** *if $A_1 \subseteq A_2 \subseteq A_3 \subseteq \ldots$, and* **contracting** *if $A_1 \supseteq A_2 \supseteq A_3 \supseteq \ldots$.*

We are now in a position to formulate the following theorem:

Theorem 1 (Rule of Continuity)**.** *Assume that $(A_n)_{n=1}^{\infty}$ is a sequence of events.*
(i) If the series is expanding, then $\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right)$.
(ii) If the series is contracting, then $\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right)$.

This theorem permits us for example to calculate the probability of winning (or losing) in a game which is potentially infinite – such as gambling in a casino, where a strategy might be “bet \$1 until winning \$1000000 or going bankrupt” (with an initial capital of, say, \$500). We will also use it later on this semester.

2. Conditional Probability

In many real-life situations, the probability that an event will occur is not “given” unconditionally; rather, it depends on some additional factors that may be taken into consideration. For example, the probability that a consumer will buy a specific product may depend on the sex, or age of this consumer; the probability that it will snow on a given day depends on the time of year, etc. In such cases, to be able to describe the probability accurately we will need some additional information about the "conditions" under which the experiment is conducted.

This is not the only reason why we may be interested in conditional probability.

(1) A record company may be interested, for example, in whether or not it may be possible to target a campaign of one product to a target group of a different product successfully. If we know something about the group of classical music lovers within a population, we may be interested in how many out of this group are also jazz music lovers (will advertising aimed at a group we have already "caught" for a different reason be efficient?). If by C we denote the set of classical music fans among the population (Ω), and by J we denote the set of jazz fans, we may be interested in the fraction

$$\frac{|J \cap C|}{|C|} = \frac{|J \cap C|/|\Omega|}{|C|/|\Omega|}.$$

Note that in the fraction on the right hand side, both the numerator and denominator are also fractions, denoting probabilities in a classical scheme.

In many cases, commonly in economic sciences, we may also be interested in answering questions of the "what may have led to this specific outcome /situation?" type. We will visualize the problem with a simple example.

(2) Assume we have rolled two dice, but we were only told that the sum of points on the dice was equal to four. What is the chance that we have obtained a "2" in the first roll?

The information we have may be formally described by a set $A = \{(1,3),(2,2),(3,1)\}$, subset of $\Omega = \{(a,b)\colon a,b \in \{1,2,3,4,5,6\}\}$. What we wish to determine is: what is the probability, that given A, event $B = \{(a,b)\colon a = 2\}$ will occur? Intuitively, we can see that the three possibilities in A are equally probable; we should therefore expect that the probability of having actually drawn the sole one where 2 appears will be equal to $\frac{1}{3}$. Formally, the experiment may be described by a classical probability scheme, and

$$\frac{1}{3} = \frac{1/36}{3/36} = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}.$$

These considerations motivate the following definition.

Definition 2. *Let X and Y be events, such that $\mathbb{P}(Y) > 0$. By* **conditional probability** *of event X under the condition Y we will understand*

$$\mathbb{P}(X|Y) = \frac{\mathbb{P}(X \cap Y)}{\mathbb{P}(Y)}.$$

A useful observation is that conditional probability is also probability (i.e., for a given condition Y, $P(A|Y)$ satisfies the definition of probability). This means that this function has all the properties of a "normal" probability function, which simplifies many calculations (and problem-solving). For example, we may have

$$\mathbb{P}(A|Y) = 1 - \mathbb{P}(A'|Y),$$

or

$$\mathbb{P}(A \cup B|Y) = \mathbb{P}(A|Y) + \mathbb{P}(B|Y) - \mathbb{P}(A \cap B|Y).$$

On the other hand, the notion of conditional probability widens the use of probability in general. For example, we may formulate the following

Theorem 2 (Chain Rule)**.** *For any sequence of events $A_1, \ldots, A_n$, such that*

$$\mathbb{P}(A_1 \cap A_2 \cap \ldots \cap A_{n-1}) > 0,$$

we have

$$\mathbb{P}(A_1 \cap A_2 \cap \ldots \cap A_n) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2|A_1) \cdot \mathbb{P}(A_3|A_1 \cap A_2) \cdots \mathbb{P}(A_n|A_1 \cap A_2 \cap \ldots \cap A_{n-1}).$$

The chain rule theorem validates a widely-used method of calculating probabilities with the use of "trees".

We will illustrate this theorem with a very simple example. Assume we draw, one by one, three cards from a deck of 52. What is the probability that all three are aces? Now let us denote by A_1 – the event that the first card will be an ace. We know that $\mathbb{P}(A_1) = \frac{4}{52}$. If by A_2 we denote the event that the second card will be an ace, we have, obviously, $\mathbb{P}(A_2|A_1) = \frac{3}{51}$. Similarly, we have $\mathbb{P}(A_3|A_2 \cap A_1) = \frac{2}{50}$ (if A_3 denotes the event of drawing an ace in the third draw). Using the chain rule, we have

$$P(A_1 \cap A_2 \cap A_3) = \frac{4}{52} \cdot \frac{3}{51} \cdot \frac{2}{50}.$$

At the beginning of this section, we have hinted that the probability of an event may depend on the "state of nature", or a set of conditions (whose probabilities we may be able to calculate easily). In this case, it is often useful to look at "normal" probability through the conditional perspective, and dissect the probability space into a suitable group of subsets, for which the initial probability may be calculated easily.

Definition 3. *Any family of events $\{H_i\}_{i \in I}$, such that $H_i \cap H_j = \emptyset$ for $i \neq j$ and $\bigcup_{i \in I} H_i = \Omega$ is called a* **partition** *of the sample space Ω.*

A partition does not have to be finite (i.e. with a finite indexing set I); we may also look at infinite countable partitions.

Theorem 3 (Law of total probability)**.** *For any finite partition $\{H_1, H_2, \ldots, H_n\}$ of the sample space Ω, such that all H_i have positive probability, and for any event A, we have*

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A|H_i) \cdot \mathbb{P}(H_i).$$

For a countable infinite partition (consisting of sets of positive probability), the situation is analogous.

Examples.

(1) A car manufacturer has two plants. 5% of vehicles produced in the first plant (A) are defective. For the second facility (B), this fraction amounts to 10%. $\frac{2}{3}$ of the production takes place in the first plant. What is the probability that a randomly selected car of this manufacturer will be defective? Our event of interest (D) is randomly choosing a defective car. We know that $\mathbb{P}(D|A) = 0.05$ and $\mathbb{P}(D|B) = 0.1$, while $\mathbb{P}(A) = \frac{2}{3}$ and $\mathbb{P}(B) = \frac{1}{3}$. We may therefore write

$$\mathbb{P}(D) = \mathbb{P}(D|A) \cdot \mathbb{P}(A) + \mathbb{P}(D|B) \cdot \mathbb{P}(B) = 0.05 \cdot \frac{2}{3} + 0.1 \cdot \frac{1}{3} = \frac{2}{30} \approx 0.067$$

(2) We have w white balls and b black balls in a box. We randomly draw a ball, and throw it away without looking at it. What is the probability of drawing a white ball in the second draw? Let W_i denote drawing a white ball, and B_i denote a black ball, in draw number i for $i = 1, 2$.

$$\begin{aligned} \mathbb{P}(W_2) &= \mathbb{P}(W_2|B_1)\mathbb{P}(B_1) + \mathbb{P}(W_2|W_1)\mathbb{P}(W_1) = \\ &= \frac{w}{w+b-1} \cdot \frac{b}{w+b} + \frac{w-1}{w+b-1} \cdot \frac{w}{w+b} = \\ &= \frac{wb + (w-1)w}{(w+b)(w+b-1)} = \frac{w}{w+b} = \mathbb{P}(W_1). \end{aligned}$$

This example is an illustration of the fact that the "material" occurrence of an event does not imply a change in the perceived probability of another event; it is our *knowledge* of the fact that changes the perceived probability. Since we did not know what was the color of the ball drawn first, this could not have had an impact on our assessment of the probability of drawing a ball of a specific color in the second draw.

In many cases, we only know the final outcome, and we would like to assess the probability of the events that have led to this result. In the second example, we may ask – after having drawn a white ball in the second draw – what was the probability that a white ball was drawn first? In the first example, we may ask what is the probability that a car was manufactured in the second plant, given that it was found to be defective?

The rule that permits to answer these types of questions is called the

Theorem 4 (Bayes' Rule)**.** *Let $\{H_i\}_{i\in I}$ be a countable (finite or infinite) partition of Ω into sets of positive probability. For any event A of positive probability, we have*

$$\mathbb{P}(H_j|A) = \frac{\mathbb{P}(A|H_j)\mathbb{P}(H_j)}{\sum_{i\in I}\mathbb{P}(A|H_i)\mathbb{P}(H_i)}.$$

In statistics and econometrics, the probabilities $\mathbb{P}(H_i)$ are often called *a priori* (before the experiment), while the probabilities $\mathbb{P}(H_i|A)$ – *a posteriori* (after the experiment).

Reverting to our two examples:

(1) The probability that a car was manufactured in the second plant, given that it was found to be defective, would be

$$\mathbb{P}(B|D) = \frac{\mathbb{P}(D|B)\mathbb{P}(B)}{\mathbb{P}(D|B)\mathbb{P}(B)+\mathbb{P}(D|A)\mathbb{P}(A)} = \frac{0.1\cdot\frac{1}{3}}{0.1\cdot\frac{1}{3}+0.05\cdot\frac{2}{3}} = \frac{1/30}{2/30} = \frac{1}{2}.$$

(2) The probability that a white ball was drawn in the first try, given that a white ball was drawn in the second try, would be

$$\mathbb{P}(W_1|W_2) = \frac{\mathbb{P}(W_2|W_1)\mathbb{P}(W_1)}{\mathbb{P}(W_2|W_1)\mathbb{P}(W_1)+\mathbb{P}(W_2|B_1)\mathbb{P}(B_1)} = \frac{\frac{w(w-1)}{(w+b)(w+b-1)}}{\frac{w}{w+b}} = \frac{w-1}{w+b-1}.$$