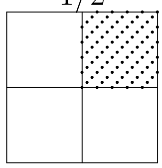
Probability Theory Notes

Neşe Yıldız



What is the probability that six will appear when a fair die is rolled? What is the probability that a point chosen at random on the unit square above falls in the dotted region? What is the probability that a point chosen randomly from [0,1] will be irrational number?

To answer these questions you had to assign a measure to the event in question. Assigning such a measure in the first question is the easiest. One can list all the possibilities and assign each of them equal measure since the die is fair. In the second question, assigning the right measure is slightly harder but still quite easy. The area of the shaded area is $1/4^{th}$ of the total area. Since the process to choose points from this set is random, if two subsets, A and B, have equal areas, the probability of the point chosen will fall into A and the probability that it will fall into B should be equal. This reasoning leads us to 1/4 as the answer to the second question. The third question is harder to answer intuitively. When the sets are simple or easy to imagine assigning them measures is easy, but when the sets are not "nice" in this sense we need mathematical tools to get to the answer.

Kolmogorov was the first person to provide axiomatic foundation for probability theory. The concept of "measure" was already developed by Borel & Lebesgue. Kolmogorov noticed analogies between measure of a set and the probability of an event, and between integral of a function and mathematical expectation of a random variable. Another analogy is between independent random variables and orthogonal functions.

While we need more sophisticated tools to talk about probability more generally, the more sophisticated tools should give us the same answer as when we use our intuition. In particular, our intuition tells us the probability of an event A cannot be negative; the probability of the set of all possible outcomes (sample space) has to be 1; if A and B are disjoint events the probability of their union should be the sum of the probabilities of these events; and if we know the probability of event A then we can figure out the probability of the event not A or A^C simply by subtracting the probability of A from 1. We will build a structure which will obey these intuitive rules.

Let S be a set. It will denote the set of all possible outcomes or "set of states", or "sample space"; it captures the whole uncertainty.

Definition 1 *A collection Σ_0 of subsets of S is called an* <u>*algebra*</u> *on S if*

1. $S \in \Sigma_0$,

2. $F \in \Sigma_0 \Rightarrow F^C := S \setminus F \in \Sigma_0$,

3. $F, G \in \Sigma_0 \Rightarrow F \cup G \in \Sigma_0$.

Remark 1 *1.* $\emptyset \in \Sigma_0$*, because* $\emptyset = S^C$*.*

2. $F, G \in \Sigma_0 \Rightarrow F \cap G \in \Sigma_0$*, because* $F \cap G = F^C \cup G^C$*.*

Now we could define a function $P : \Sigma_0 \to [0, 1]$ and hope that this would work, but we would run into difficulties if the sample space has infinitely many subsets.

Definition 2 *A collection* Σ *of subsets of* S *is called a* <u>*σ-algebra*</u> *on* S *if* Σ *is an algebra on* S *such that whenever* $\{F_n\}_{n\in\mathbb{N}} \subseteq \Sigma$, $\cup_{n\in\mathbb{N}} F_n \in \Sigma$*.*

Remark 2 *If* Σ *is a* σ*-algebra on* S *and* $\{F_n\}_{n\in\mathbb{N}} \subseteq \Sigma$*, then* $\cap_{n\in\mathbb{N}} F_n = (\cup_{n\in\mathbb{N}} F_n^C)^C \in \Sigma$*.*

Note that a σ*-algebra on* S *is the collection of distinguishable subsets of* S*. If two sets belong to a* σ*-algebra then potentially we are able to distinguish them from one another. The concept of a* σ*-algebra captures in a sense "what is knowable" in a given uncertain situation.*

Terminology: A pair (S, Σ), where is a set and Σ is a σ-algebra on S is called a measurable space. An element of Σ is called a Σ-measurable subset of S. In probability theory an element of Σ is called an <u>event</u>.

Example 1 *Let* $S \neq \emptyset$, $\Sigma = \{\emptyset, S\}$*.*

Example 2 *Let* S *be any set and* Σ *be the power set of* S*.*

Example 3 $S = \{1, 2, 3\}$, $\Sigma = \{\emptyset, S, \{1\}, \{2, 3\}\}$*.*

Definition 3 *Let* $\mathcal{C}$ *be a class of subsets of* S*. Then* $\sigma(\mathcal{C})$*, the* σ*-algebra generated by* $\mathcal{C}$ *is the smallest* σ*-algebra* Σ *on* S *containing* $\mathcal{C}$*.*

Example 4 $S = \{1, 2, 3\}$, $\mathcal{C} = \{\{1\}\}$*. Then* $\sigma(\mathcal{C}) = \{\emptyset, S, \{1\}, \{2, 3\}\}$*.*

Note that in the last example power set of S would be another σ-algebra containing $\{1\}$ but it would not be the smallest one.

Example 5 *(Borel* σ*-algebra) Let* S *be any subset of a topological space. The Borel* σ*-algebra on* S*,* $\mathcal{B}(S)$*, is the* σ*-algebra generated by the family of open subsets of* S*. This is the most important* σ*-algebra for us.*

Definition 4 *Let* (S, Σ) *be a measurable space. A function* $\mu : \Sigma \to [0, \infty]$ *is called a measure if it satisfies*

1. $\mu(\emptyset) = 0$,

2. Whenever $\{F_n : n \in \mathbb{N}_+\}$ is a sequence of disjoint sets in Σ, $\mu(\cup_{n=1}^{\infty} F_n) = \sum_{n=1}^{\infty} \mu(F_n)$.

The triple (S, Σ, μ) is called a measure space.

Definition 5 *A measure μ is called a probability measure if $\mu(S) = 1$ and (S, Σ, μ) is then called a probability triple.*

Examples to measure spaces:

Example 6 *$S \neq \emptyset$, $\Sigma =$ Power set of S, $\mu(A)$ equals number of elements in A if A is finite, and equals ∞ otherwise.*

Example 7 *$S = \mathbb{R}$, $\Sigma = \mathcal{B}(\mathbb{R})$, $\mu(A) =$ length of A.*

Example 8 *$S = \mathbb{R}$, $\Sigma = \{A \subseteq \mathbb{R} :$ either A is countable A^C is countable$\}$, and $\mu(A) = 0$ if A is countable and $\mu(A) = 1$ if A is uncountable.*

Example 9 *In Example 6, restrict S to be a finite set so that $\mu(S) < \infty$. Then we can construct a probability measure that would correspond to a frequency:*

$$P(A) := \frac{\mu(A)}{\mu(S)}.$$

1 Independence

From your undergraduate statistics courses you might remember that two events A and B are called independent if $P(A \cap B) = P(A)P(B)$.[1] To be able to talk about independence of random variables we will need a more general concept of independence.

Definition 6 *Let $\mathcal{F}$ be a σ-algebra on S. $\mathcal{G}$ is called a sub σ-algebra of $\mathcal{F}$ if $\mathcal{G}$ itself is a σ-algebra and $\mathcal{G} \subseteq \mathcal{F}$.*

Definition 7 *Let $(\Omega, \mathcal{F}, P)$ be a probability triple. Let $\mathcal{G}_1, \mathcal{G}_2, ...$ be all sub-σ-algebras of $\mathcal{F}$. They are called <u>independent</u> if, whenever $G_n \in \mathcal{G}_n$ and $i_1, ..., i_n$ are distinct indices,*

$$P(G_{i_1} \cap G_{i_2} \cap ... G_{i_n}) = \Pi_{k=1}^{n} P(G_{i_k}).$$

Note that if a collection of σ-algebras are pairwise independent it does not necessarily mean that they will be mutually independent. To see this, consider the sample space representing the possible outcomes of two independent coin tosses, so that $S = \{TT, TH, HT, HH\}$. Let Σ be the power set of S. Let $P(E)$ denote the number of elements in E divided by 4. Let A denote the event that the outcome of the first coin toss is H. Let B denote the event that the outcome of the second toss is H. Finally, let C denote the event that either both tosses yield H or both yield T. So $A = \{HH, HT\}$, $B = \{TH, HH\}$, and $C = \{HH, TT\}$. Now $A \cap B = A \cap C = B \cap C = \{HH\}$. It is thus easy to see the pairwise independence of the σ-algebras generated by the sets A, B and C. Next, note that $A \cap B \cap C = \{HH\}$ and $P(A \cap B \cap C) = 1/4 \neq 1/8 = P(A)P(B)P(C)$. Thus, $\sigma(A)$, $\sigma(B)$, $\sigma(C)$ are not (mutually) independent.

[1] Note that Amemiya (1994) calls this "pairwise independence" and the definition below "mutual independence".

References

[1] Amemiya, T., 1994, *Introduction to Statistics and Econometrics* (Harvard University Press: London, England).

[2] Amemiya, T., (1985), *Advanced Econometrics* (Harvard University Press: London, England).

[3] Pollard, D., 2002, *A User's Guide to Measure Theoretic Probability* (Cambridge University Press: Cambridge, England).

[4] Williams, D., 1991, *Probability with Martingales* (Cambridge University Press: Cambridge, England).