Mathematical foundations of Econometrics

G.Gioldasis, UniFe & prof. A.Musolesi, UniFe

March 19, 2016

The axiomatic definition of Probability

Definition

A function $\mathrm{P} : \mathscr{F} \to [0,1]$ from a *σ-algebra* $\mathscr{F}$ of events of a set Ω into the unit interval is a probability measure on $\{\Omega, \mathscr{F}\}$ if it satisfies the following three conditions (axioms):

1. For all $\mathrm{A} \in \mathscr{F}$, $P(\mathrm{A}) \geq 0$
2. $\mathrm{P}(\Omega) = 1$
3. For disjoint sets $A_j \in \mathscr{F}$, $\mathrm{P}\left(\bigcup_{j=1}^{\infty} \mathrm{A}_j\right) = \sum_{j=1}^{\infty} \mathrm{P}(\mathrm{A}_j)$

   (σ - *additivity*)

Note on Probability

The axiomatic definition of probability does not impose a particular probability measure for an experiment. Therefore, any function that satisfies the aforementioned conditions can be a probability. Consequently, the researcher must choice on the probability measure that is most appropriate for the problem-experiment.

Example: In the **"fair" coin tossing experiment**, suppose that the researcher thinks that:

$$P(A) = \begin{cases} 1 \text{ if} & A = \{H, T\} \\ 1 \text{ if} & A = \{H\} \\ 0 \text{ if} & A = \{T\} \end{cases}$$

According to the definition, such a mapping can be considered as a probability measure! The researcher must "choose" the appropriate probability which best describes the experiment, in particular:

$$P(A) = \begin{cases} 1 \text{ if} & A = \{H, T\} \\ 1/2 \text{ if} & A = \{H\} \\ 1/2 \text{ if} & A = \{T\} \end{cases}$$

Probability space

The triplet $\{\Omega, \mathscr{F}, \mathrm{P}\}$ is called a **probability space**, while $\{\Omega, \mathscr{F}\}$ is a **measurable space**.

Note: A probability rule is defined only inside a probability space. So, a measurable space $\{\Omega, \mathscr{F}\}$ must be defined first.

Example: Lets remember the example of asking ten people about their employment status. Assume that the unemployment rate is p.

Again, what is the sample space Ω and what can be a "suitable" σ-*algebra* for the experiment?
After defining the above (*in practice they are implicit...*), we can then answer questions as the following one:

What is the probability of the event

$$\{\text{At most one person surveyed is unemployeed}\}$$

The answer should be $P(\{0,1\}) = (1-p)^{10} + 10p(1-p)^{9}$ (*Exercise: Why?*)

Question

Question
Why is the domain of P a σ-*algebra*? *Why don't we just take all subsets of* Ω *as the domain of* P?

Answer
The notion of "information" is introduced in our problem through the definition of a **suitably selected** σ-*algebra*, not restrictively through just the powerset $\mathscr{P}(\Omega)$ of Ω.

Question: Why the domain of P is a σ-*algebra* and not just an *algebra*?

Properties of probabilities

Theorem

Let $\{\Omega, \mathscr{F}, \mathrm{P}\}$ *be a probability space. The following hold for sets in* $\mathscr{F}$*:*

- $\mathrm{P}(\emptyset) = 0$
- $\mathrm{P}(\tilde{\mathrm{A}}) = 1 - \mathrm{P}(\mathrm{A})$
- $\mathrm{A} \subset \mathrm{B}$ *implies* $\mathrm{P}(\mathrm{A}) \leq \mathrm{P}(\mathrm{B})$
- $\mathrm{P}(\mathrm{A} \cup \mathrm{B}) + \mathrm{P}(\mathrm{A} \cap \mathrm{B}) = \mathrm{P}(\mathrm{A}) + \mathrm{P}(\mathrm{B})$
- *If* $\mathrm{A}_n \subset \mathrm{A}_{n+1}$ *for* $n = 1, 2, \ldots,$ *then* $\mathrm{P}(\mathrm{A}_n) \uparrow \mathrm{P}(\cup_{n=1}^{\infty} \mathrm{A}_n)$
- *If* $\mathrm{A}_n \supset \mathrm{A}_{n+1}$ *for* $n = 1, 2, \ldots,$ *then* $\mathrm{P}(\mathrm{A}_n) \downarrow \mathrm{P}(\cap_{n=1}^{\infty} \mathrm{A}_n)$
- $\mathrm{P}(\cup_{n=1}^{\infty} \mathrm{A}_n) \leq \sum_{n=1}^{\infty} \mathrm{P}(\mathrm{A}_n)$

Exercise: Can we derive the first four properties?

Conditional probabilities

Consider the game of rolling a dice once.

- The sample space of the experiment is $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- We choose *σ-algebra* $\mathscr{F}$ to be the powerset $\mathscr{P}(\Omega)$ (i.e. all subsets of Ω) and
- We choose the probability rule to be $P(\{\omega\}) = 1/6$ for all $\omega = 1, 2, 3, 4, 5, 6$.

Let B be the event that the outcome is even, i.e. $B = \{2, 4, 6\}$. We would like to find the probability of $A = \{1, 2, 3\}$ given that B is true.

Common sense: If we know that the outcome is even, we know that the outcomes $\{1, 3\}$ in A did not occur. Therefore, the probability that A occurs given that B occured is $P(A \cap B)/P(B) = 1/3$.

Definition

Let A, B be events in $\{\Omega, \mathscr{F}, P\}$ and $P(B) > 0$. The conditional probability of A given B is defined as:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Conditional probability as a probability measure

Question: Is the conditional probability $P(\cdot|B) : \mathscr{F} \to [0,1]$ a probability according to Kolmogorov's axiomatic definition?

Answer: Yes, because all 3 axioms are met*.

Therefore, the arrival of a "new" information B updates the probability rule from $P(\cdot)$ to $P(\cdot|B)$, reflecting the fact that our expectation for an event to happen or not changes according to the information we have. Hence, the probability space from $\{\Omega, \mathscr{F}, P\}$ becomes $\{\Omega, \mathscr{F}, P(\cdot|B)\}$.

*Exercise: Can we prove that $P(\cdot|B) : \mathscr{F} \to [0,1]$ is a probability in $\{\Omega, \mathscr{F}\}$?

Simple exercise

Select two persons at random. What is the probability of both being female given that at least one is female? Is it 1/2 ?

Answer

The initial sample space is $\Omega = \{MM, FF, FM, MF\}$. We condition with respect to $B = \{FF, FM, MF\}$. If $A = \{FF\}$, then the probability of interest is:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1}{3}$$

So, instead of the initial probability space $\{\Omega, \mathscr{P}(\Omega), P\}$ we used another one $\{\Omega, \mathscr{P}(\Omega), P(\cdot|B)\}$.

General question

Question: Apart from the probability space $\{\Omega, \mathscr{F}, P(\cdot|B)\}$, can we define another probability space where the sample space, the σ- algebra and the probability are different?

Answer: Given an "information" B, we can also create another sample space $\Omega' = B$ and another respective σ-*algebra* $\mathscr{F}' \subset \mathscr{F}$ are **also** as follows:

$$\mathscr{F}' = \mathscr{F} \cap B \stackrel{def.}{=} \{all\ A \cap B \text{ so that } A \in \mathscr{F}\}$$

As already been told, also the probability measure becomes $P' : \mathscr{F}' \rightarrow [0, 1]$ so that

$$P'(A) = P(A|B) = P(A \cap B)/P(B) \text{ for every } A \in \mathscr{F}'$$

Therefore, we can think that conditioning on $B \in \mathscr{F}$, not only $\{\Omega, \mathscr{F}, P(\cdot|B)\}$ but also another probability space $\{B, \mathscr{F} \cap B, P(\cdot|B)\}$ is created.

Exercise continued

We recall that in the experiment of selecting two persons, the sample space is $\Omega = \{MM, FF, FM, MF\}$. Moreover, we chose for σ-algebra the powerset $\mathscr{F} = \mathscr{P}(\Omega) = \{\text{all events in } \Omega\} =$
$\{\Omega, \emptyset, \{MM\}, \{FM\}, \{MF\}, \{FF\}, \{MM, FM\}, \{MM, MF\}, \{MM, FF\},$
$\{FM, MF\}, \{FM, FF\}, \{MF, FF\}, \{MM, FM, MF\}, \{MM, FM, FF\},$
$\{FM, MF, FF\}, \{MF, FF, MM\}\}$

After the "arrival" of the information $B = \{FF, FM, MF\}$, we can also define a different experiment, namely:
"We peak a pair of persons from which at least one is a female"

In this new experiment, the sample space is B and our information is described by:

$$\mathscr{F}' = \mathscr{F} \cap B = \{B, \emptyset, \{FM\}, \{MF\}, \{FF\}, \{FM, MF\}, \{FM, FF\}, \{MF, FF\}$$

Obviously $\mathscr{F}' \subset \mathscr{F}$
The probability defined in $\{B, \mathscr{F} \cap B\}$ is: $P(\cdot|B) = P(\cdot \cap B)/P(B)$

Conclusion: After the arrival of information B, apart from $\{\Omega, \mathscr{F}, P(\cdot|B)\}$ we can also consider another probability space $\{B, \mathscr{F} \cap B, P(\cdot|B)\}$.

Properties

All properties of probability measures described before carry to conditional probabilities. So, for example:

- $\mathrm{P}(\tilde{\mathrm{A}}|B) = 1 - \mathrm{P}(\mathrm{A}|B)$
- $\mathrm{A} \subset \mathrm{B}$ implies $\mathrm{P}(\mathrm{A}|C) \leq \mathrm{P}(\mathrm{B}|C)$
- $\mathrm{P}(\mathrm{A} \cup \mathrm{B}|C) + \mathrm{P}(\mathrm{A} \cap \mathrm{B}|C) = \mathrm{P}(\mathrm{A}|C) + \mathrm{P}(\mathrm{B}|C)$
- If $\mathrm{A}_n \subset \mathrm{A}_{n+1}$ for $n = 1, 2, \ldots,$ then $\mathrm{P}(\mathrm{A}_n|B) \uparrow \mathrm{P}(\cup_{n=1}^{\infty} \mathrm{A}_n|B)$
- If $\mathrm{A}_n \supset \mathrm{A}_{n+1}$ for $n = 1, 2, \ldots,$ then $\mathrm{P}(\mathrm{A}_n|B) \downarrow \mathrm{P}(\cap_{n=1}^{\infty} \mathrm{A}_n|B)$
- $\mathrm{P}(\cup_{n=1}^{\infty} \mathrm{A}_n|B) \leq \sum_{n=1}^{\infty} \mathrm{P}(\mathrm{A}_n|B)$

Law of total probability

Suppose a probabilty space $\{\Omega, \mathscr{F}, P\}$. Also, assume a **finite or countably infinite partition of** Ω, i.e. $\{A_n : n = 1, 2, \ldots\}$ with $A_n \in \mathscr{F}$. Then, according to the *Law of total probability*,for an event $B \in \mathscr{F}$, it holds that:

$$P(B) = \sum_n P(B|A_n)P(A_n)$$

The above is depicted in the following graph.

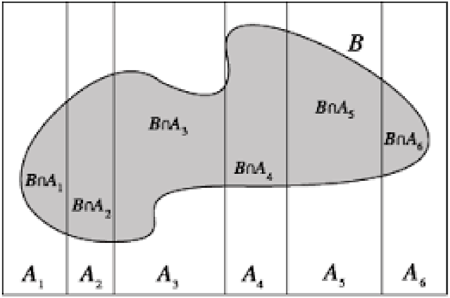

Figure 1: Law of total probability

Bayes' rule

Let A and B be sets in $\mathscr{F}$. According to the definition of conditional probabilities:

$$P(A|B) = P(A \cap B)/P(B) \tag{1}$$

Moreover, A and $\tilde{A}$ form a partition of the Ω sample space. Therefore $B \cap A$, $B \cap \tilde{A}$ are disjoint and, moreover, $B = (B \cap A) \cup (B \cap \tilde{A})$. Therefore

$$P(B) = P(B \cap A) + P(B \cap \tilde{A}) \tag{2}$$

Substituting (2) in (1) and applying the definition of conditional probability again:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\tilde{A})P(\tilde{A})}$$

More generally,

Theorem (Bayes' rule)

If $A_j, j = 1, 2, \ldots, n$ is a partition of the sample space Ω and $A_j, B \in \mathscr{F}$, then:

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{n} P(B|A_j)P(A_j)}$$

Example on Bayes' rule

Consider the previous example of selecting two persons at random. Obviously, a partition of the sample space can be $\{A_1, A_2\}$, where:
$A_1 = \{FF\} = \{\text{both persons are females}\}$
$A_2 = \{FM, MF, MM\} = \{\text{not all persons are females}\}$
Conditioning on the new information

$$B = \{FF, FM, MF\} = \{\text{at least one is female}\}$$

Bayes' rule dictates that:

$$P(A_1|B) = \frac{P(B|A_1)P(A_1)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2)} = \frac{1 \cdot 1/4}{1 \cdot 1/4 + 2/3 \cdot 3/4} = \frac{1}{3}$$

Note: Bayes rule updates the probability of an event from $P(A)$ to $P(A|B)$ according to the new information B that is provided to us.

Pairwise independence

Definition (Pairwise independence)

Sets A and B in $\mathscr{F}$ are (pairwise) independent if $P(A \cap B) = P(A)P(B)$. If $P(B) > 0$ then A and B are independent if $P(A|B) = P(A)$.

Note: The transitive property does not hold, i.e. if A and B are independent and B and C are independent, this does not mean that A and C are independent.

For example, consider the case where A and B are independent, $C = \tilde{A}$ and $0 < P(A) < 1$. Then, also B and C are independent (why?) but A and $\tilde{A}$ are not!

Exercise

Show that the following hold:

- If $P(A) = 0$ or $P(B) = 0$ then A and B are independent
- If A and B are independent, then so are $\tilde{A}$ and B, $\tilde{A}$ and $\tilde{B}$, A and $\tilde{B}$.

Independence

Definition

A sequence A_j of sets in $\mathscr{F}$ is independent if for any subsequence A_{j_i}, $i = 1, 2, \ldots, n$, $P(\bigcap_{i=1}^{n} A_{j_i}) = \prod_{i=1}^{n} P(A_{j_i})$.

For example, if 3 set A_1, A_2 and A_3 are independent, then:

- $P(A_1 \cap A_2) = P(A_1)P(A_2)$
- $P(A_1 \cap A_3) = P(A_1)P(A_3)$
- $P(A_2 \cap A_3) = P(A_2)P(A_3)$ and
- $P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3)$

Therefore independence implies pairwise independence. Nevertheless, the inverse does not hold; a group of pairwise independent sets may not be independent.

Example of independent events

Consider the example of tossing a fair coin twice. The probability space can be $\{\Omega, \mathscr{F}, p\}$, with $\Omega = \{HH, HT, TH, TT\}$, $\mathscr{F} = \mathscr{P}(\Omega)$ and the usual probability measure P.
Let us consider also the following events:

$A = \{\text{head appears in the first toss}\} = \{HH, HT\}$
$B = \{\text{head appears in the second toss}\} = \{TH, HH\}$ and
$C = \{\text{both tosses yield heads or both tosses yield tails}\} = \{HH, TT\}$
Then obviously $P(A \cap B) = P(A)P(B)$, $P(A \cap C) = P(A)P(C)$ and $P(B \cap C) = P(B)P(C)$ but $P(A \cap B \cap C) \neq P(A)P(B)P(C)$.

Therefore, the 3 sets are pairwise independent but **not** independent.

Conditional independence

Definition (Conditional independence)

Sets A and B in $\mathscr{F}$ are conditionally independent given $C \in \mathscr{F}$, if $P(A \cap B|C) = P(A|C)P(B|C)$.

Note: Conditional independence does not imply independence. Also, independence does not imply conditional independence. This holds because conditioning alters the sample space. See for instance the following graph.

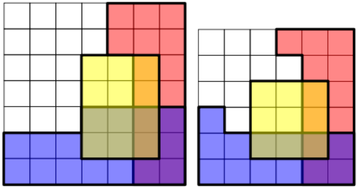

Figure 2: Conditional but not unconditional independence

Obviously, $P(B \cap R|Y) = P(B|Y)P(R|Y)$, but $P(B \cap R) \neq P(B)P(R)$.