Set theoretic methods

Adrian Dușa

06 July 2019

Abstract

Apart from the well established and ever evolving quantitative research methods, a relatively new method begins to gain momentum. Political sciences, and more generally macro-comparative sociological studies work with a small number of cases (usually countries), and that renders standard statistics as sometimes unsuitable. This new methodology, called QCA – Qualitative Comparative Analysis (Ragin, 1987, 2000, 2008) is able to detect configurational patterns even for these small-N scenarios, but perhaps more interestingly it is fundamentally different from conventional statistics due to its asymmetric nature, inherited from its set theoretical nature. Conventional correlations are symmetric (more x, more y, or less x, less y), and the same explanatory model is responsible with both ends of the dependent variable's continuum. By contrast, set theory is inherently asymmetric: a case belongs to a set or it does not, an outcome is either produced or it is not, and more importantly the configuration of causal factors that produce the outcome is most of the times different from the configuration responsible with the absence of the outcome. Both necessity and sufficiency can be determined by exploiting the set theoretic nature of the social science concepts, where their asymmetric nature becomes obvious: what is necessary for an outcome might not be necessary for its absence, and while necessary it might not always be sufficient. The other way round, what is sufficient for an outcome might not always be necessary, a fact that signals what is called "equifinality", which posits the same outcome can be produced by multiple configurational patterns.

Introduction

There are various ways to analyze social phenomena. The traditional, qualitative and quantitative approaches involve specialized languages and seemingly incompatible methods. But such phenomena can also be framed in terms of set relations, as it is often the case in the common, everyday language. For instance, poverty research can either employ quantitative, nationally representative samples, or they can use case studies to unfold particular, exemplar life stories that are usually obscured by numbers, or it can be framed in a set theoretical perspective as recently demonstrated by Ragin and Fiss (2017).

They studied the relation between poverty and various configurational patterns that include race, class, and test scores, and found that white people are mainly characterized by multiple advantages that protect from poverty, while there are configurations of disadvantages that are mainly prevalent in black people. These disadvantages do no necessarily lead to poverty, with an important exception: black women.

Being black, and being a woman, and having children, and a configuration of disadvantages, is more than sufficient to explain poverty. This approach is less concerned about the relative effects of each independent variables included in the model, but rather about identifying membership in a particular set (in the current example, of the disadvantaged black women). It is a set relational perspective, more precisely with a focus on set intersections to explain social phenomena.

This chapter begins with a short background of set theory and the different types of sets that are used in the social sciences. It presents the most important set operations that are commonly used in the mathematical framework behind a set theoretical methodology, it shows how to formulate hypotheses using sets and exemplifies how to calculate set membership scores via the different calibration methods. Finally, it presents important concepts related to necessity and sufficiency and ends with a discussion about how to apply set theory in the Qualitative Comparative Analysis.

Short background of set theory

Formally initiated by philosopher and mathematician Georg Cantor at the end of the 19th century (Dauben, 1979), the classical set theory became part of the standard foundation of modern mathematics, well suited for the treatment of numbers (Pinter, 2014). Elementary mathematics is embedded with notions such as the set of real numbers, or the set of natural numbers, and formal demonstrations almost always employ sets and their elements as inherent, prerequisite properties of a mathematical problem.

It is nowadays called the naive set theory (Halmos, 1974), and was later extended to other versions, but the basic properties prevailed. A set can be defined as a collection of objects that share a common property. If an element x is a member of a set A, it is written as $x \in A$, and if it is not a member of that set, it is written as $x \notin A$. This is the very essence of what is called binary crisp sets, where objects are either in or out of a set.

For any object, it can be answered with "yes" if it is inside the set, and "no" if it is not. There are only two possible truth values in this version: 1 (true) and 0 (false): a country is either in, or outside the European Union, a law is either passed or not passed, an event either happens or does not happen etc. It has certain roots into Leibniz's binary mathematics from the beginning of the 18th century (Aiton, 1985), later formalized into a special system of logics and mathematics called Boolean algebra, to honor the work of George Boole from the middle of the 19th century.

In formal notation, a membership function can be defined, to attribute these two values:

$$\mu_A(x) = \begin{cases} 0 & \text{if } x \notin \text{A} \\ 1 & \text{if } x \in \text{A} \end{cases}$$

It is perhaps worth mentioning the work of all these people was influenced by the Aristotelian logic, a bivalent system based on three principles (laws of thought): the principle of identity, the principle of non-contradiction and the principle of excluded middle. A single truth value could be assigned for any proposition (either true, or false), but this was only possible for past events. No truth value could be assigned to a proposition referring to the future, since a future event has not yet happened. Future events can be treated deterministically (what is going to be, is going to be) or influenced by peoples' free will (we decide what is going to be), leading to a paradox formulated by Aristotle himself.

A solution to this problem was proposed by the Polish mathematician Łukasiewicz (1970), who created a system of logic at the beginning of the 20th century, that extends the classical bivalent philosophy. His system (denoted by $Ł_3$) presents not just two but three truth values:

$$\mu_A(\mathrm{x}) = \begin{cases} 0 & \text{false} \\ \frac{1}{2} & \text{undetermined (neither false nor true, or partially true)} \\ 1 & \text{true} \end{cases}$$

Łukasiewicz's system (using a finite number of values) was eventually generalized to multi-valent systems with $n = v - 1$ values, obtained through a uniform division of the interval [0, 1]:

$$\left\{0 = \frac{0}{n}, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n}{n} = 1\right\}$$

While some phenomena are inherently bivalent (an element is either in, or out of a set), there are situations where two values are unable to describe the whole picture. A social problem is not necessarily solved or unsolved, but can be more or less dealt with. A country is not simply rich or poor, but it can be more or less included in the set of rich countries. There is a certain degree of uncertainty regarding the truth value, that was modeled in the middle of the 20th century by another great mathematician, who laid out the foundations of the fuzzy sets (Zadeh, 1965). These types of sets have a continuous (infinite) number of membership values, in the interval bounded by 0 (completely out of the set) to 1 (completely in the set).

Set operations

Set operations are mathematical transformations that reflect the logical relations between sets, to reflect various configurations that involve intersections, unions and/or negations. The simplest way to think about these operations is an analogy to basic mathematical algebra: addition, subtraction, multiplication and division are all very simple, yet essential operations to build upon. In set theory, there are essentially three main operations that are used extensively in the set theoretical research and comparative analysis: set intersection, set union and set negation.

These operations perform differently for crisp and fuzzy sets, but the fuzzy version is more general and can be applied to crisp situations as well.

Set intersection (logical AND)

In the crisp version, the goal of this operation is to find the common elements of two sets. A truth value is involved, that is assigned a "true" value if the element is common, and "false" otherwise. Out of the four possible combinations of true/false values in Table 1 for the membership in the two sets, only one is assigned a "true" value for the intersection, where both individual values are true.

It is a called a "conjunction", meaning the logical AND expression is true only when both sets are (conjunctively) true. It is usually denoted using the "$\cap$" or multiplication "$\cdot$" signs.

The fuzzy version of the set intersection formula is obtained by calculating the minimum between two (or more) values:

$$A \cap B = min(A, B) \tag{1}$$

Table 1: Set intersection for crisp sets

A		B		
0	AND	0	=	0
0	AND	1	=	0
1	AND	0	=	0
1	AND	1	=	1

As the minimum between 0 (false) and any other truth value is always 0, this formula holds for the data from Table 1, where a minimum equal to 1 (true) is obtained only when both values are equal to 1.

Set union (logical OR)

The counterpart of the set intersection is the set union, used to form larger and larger sets by pulling together all elements from all sets. In the crisp sets version, the result of the union operation is true if the element is part of at least one of the sets. Contrary to set intersection, the only possible way to have a false truth value is the situation where an object is not an element of any of the (two) sets:

Table 2: Set union for crisp sets

A		B		
0	OR	0	=	0
0	OR	1	=	1
1	OR	0	=	1
1	OR	1	=	1

The union of two sets is called a "disjunction" and it is usually denoted with the "∪" or "+" signs, and the later should not be confused with the arithmetic addition.

The fuzzy version of this operation is exactly the opposite of the set intersection, by calculating the maximum between two (or more) values:

$$A \cup B = max(A, B) \tag{2}$$

Set negation

Set negation is a fundamental operation in set theory, consisting in finding the complement of a set A from a universe U (which is a different set of its own, formed by the elements of U that are not in A). It is many times denoted with the "∼" or "¬" signs, and sometimes (especially in programming) with the exclamation sign "!".

Negating multivalue crisp sets involves taking all elements that are not equal to a specific value. It is still a binary crisp operation, by first coercing the multivalue set into a binary crisp one, then negating the resulting values.

Table 3: Set negation for crisp sets

	A		
NOT	0	=	1
NOT	1	=	0

Negation is a unary operation, and it's fuzzy version is a simple subtraction from 1:

$$\sim A = 1 - A \tag{3}$$

The importance of set negation will be revealed later, especially when comparing the quantitative methods with the set theoretical ones, to reveal a certain asymmetry that is specific to sets, with a methodological effort to explain both the presence and the negation (its absence) of a certain phenomenon.

Formulating hypotheses using sets

There are multiple ways to conceptualize, measure and hypothesize social and political phenomena. Previous chapters from this book present several such approaches, from the quantitative types centered on variables to qualitative methods focused on cases. The quantitative approach relies on very precise statistical properties stemming from large samples, and it describes the net effect of each independent variable on the outcome (the dependent variable), controlling for all other variables in the model. It is a relatively straightforward, albeit specialized statistical language that is extensively used in quantiative research, however it is not the most common language to formulate scientific hypotheses.

Hsieh (1980), Novák (1991), Arfi (2010) and even Zadeh (1983) himself have shown how the set theory, and especially the fuzzy sets, can be related to the natural language. And contrary to most common expectations, scientific hypotheses do not usually mention the specific net effects of various independent variables, instead they seem very compatible with the language of sets, much like the natural language.

For instance, hypothesing that democratic countries do not go to war with each other (Babst, 1964) can be naturally translated into sets. The elements are countries, and there are two sets involved: the set of democratic countries, and the set of countries that do not go to war with each other. It is the type of hypothesis that can be best expressed in terms of sufficiency and subset relation, but for the moment it should suffice stating it is a concomitant membership of the two sets: those countries that are included in the set of democratic countries, are also included in the set of countries that do not go to war with each other.

The same type of language can be applied to another common type of hypothesis, in a if-then statement, for instance: "if a student passes the final exam, then he or she graduates". Here too, it is about two sets: the set of students who pass the final exam, and the set of students who graduate, and membership in the first guarantees membership in the second.

It seems natural to specify such hypotheses in terms of set language, in both fuzzy sets form (more or less democratic countries) and even binary crisp form (either graduate, or not). Scientific thinking, at least in the social and political sciences, is a constant interplay between abstractization and exact measurement: we first start by specifying the (pre)conditions that make a certain outcome possible, and only then we measure exact values for each such condition or variable.

A statement such as: "welfare is education and health" does not mention any specific values of the education, or of the health, that produce the welfare. This is but one among many possible causal recipees (in the vein of the welfare typologies contributed by Esping-Andersen, 1990) where only the ingredients (education and health) are mentioned, without specifying the exact net effects that are needed to produce welfare. It is entirely possible to assign precise mathematical numbers to sets (more exactly, to set membership scores), which is the topic of the next section, but formulating hypotheses is more a matter of specifying abstract concepts (similar to sets), and less about exact values for each.

Using a different perspective on the relation between fuzzy sets and natural language, George Lakoff rejects that natural language can be perfectly mapped over the set theory (Ramzipoor, 2014). He also criticizes Charles Ragin's approaches to assign membership scores (presented in the next section about set calibration), based on his expertise combining linguistics and cognitive science. More recently, Mendel and Korjani (in press) propose a new method using the Type-2 fuzzy sets.

The whole debate is extremely interesting, for social science concepts have a dual nature stemming from both linguistics and theoretical corpus, but it is by now evident that set theory is well established in social and political research. Conceptual thinking has long tradition in sociology, with Max Weber's ideal types being similar to set theoretic concepts that play a central role in the comparative analysis. In fact, the whole process of concept formation is embedded with the language of set theory (Goertz, 2006*b*; Mahoney, 1980; Schneider and Wagemann, 2012).

Despite the predominance of the quantitative methods in the social and political sciences, there are situations where statistical analyses are impossible (mainly due to a very small number of cases) and most importantly where the use of set theory is actually more appropriate, for both formulating and testing theories.

Set calibration

In the natural sciences, assigning membership scores to sets is a straightforward procedure. Objects have physical properties that can be measured and transformed into such membership scores. In the social and political sciences, the situation is much more complex. These sciences deal with highly complex phenomena that can only be conceptualized at a very abstract level. They don't exist in the physical reality and don't have visible properties to measure directly.

Concepts are very abstract things, and their measurement is even more complex: it depends on theory, which determines their definition, which in turn has a direct effect over their operationalization, that has an influence on constructing the research instrument and only then some measurements can be collected.

Each of these stages involve a highly specialized training, requiring years (sometimes a lifetime) of practice before mastering the activity. Theoreticians are rare, at least those who have a real impact over the research praxis of the entire academic community. Most researchers follow a handful of theories that attempt to explain the social and political reality. Each such theory should be ideally reflected into a clear definition of the abstract concept.

Based on the definition, the process of operationalization is yet another very complex step before obtaining some kind of numerical measurements about the concept. It is based on the idea that, given the impossibility of directly measuring the concept, researchers can only resort to measuring its effect over the observable reality. For instance, we cannot tell how altruistic a person is, unless we observe how the person behaves in certain situations related to altruism. There are multiple ways for a person to manifest this abstract concept, and the operationalization is a process to transform a definition into measurable indicators, usually via some other abstract dimensions and subdimensions of the concept.

Finally, obtaining numerical scores based on the indicators from the operationalization phase is yet another complex activity. There are multiple ways to measure (counting only the traditional four levels of measurement: nominal, ordinal, interval and ratio, but there are many others), and the process of constructing the research instrument, based on the chosen level of measurement for each indicator, is an art. It is especially complex as the concepts should also be equivalent in different cultures, and huge efforts are being spent to ensure the compatibility between the research instruments from different languages (translation being a very sensitive activity).

The entire process ends up with some numerical measurements for each indicator, and a final task to aggregate all these numbers to a single composite measure that should be large if the concept is strong, and small if the concept is weak. In the above example, highly altruistic people should be allocated large numbers, while unconcerned people should be allocated low numbers, on a certain numerical scale.

In set theory, calibration is the process of transforming these (raw) numbers into set membership scores, such that a completely altruistic person should receive a value of 1, while a non-altruistic person should receive a value of 0. This process is far from straightforward, even for the natural sciences.

Describing the procedure, Ragin (2008) makes a distinction between "calibration" and "measurement" processes, and exemplifies with the temperature as a directly measurable phisical property. While exact temperatures can be obtained from the absolute zero to millions of degrees, no such procedure would even be able to automatically determine what is "hot" and what is "cold". These are human interpreted concepts, and need to be associated with some subjective numerical anchors (thresholds). On the Celsius scale, 0 degrees is usually associated with cold, while 100 degrees is usually associated with very hot, and these numbers are not picked at random. They correspond to the points where the water changes states: to ice at 0 degrees and to steam at 100 degrees, when the water boils.

The choice of thresholds is very important, for it determines the point where something is completely out of a set (for instance at 0 degrees, the ice is completely out of the set of hot matter) and the point where something is completely inside the set (at 100 degrees, steam is completely inside the set of hot matter). A third threshold is also employed, called the "crossover", the point of maximum ambiguity where it is impossible to determine whether something is more in than out of a set, corresponding to the set membership score of 0.5.

The set of thresholds (exclusion, crossover and inclusion) is not universal, even for the same concept. A "tall" person means something in countries like Norway and Netherlands, with an average male height of more than 1.8 meters, and something different in countries like Indonesia and Bolivia where the average is about 1.6 meters. It is the concept that matters, not its exact measurement, therefore different thresholds need to be used in different cultural contexts, depending on the local perception.

Traditionally, there are two types of calibrations for each type of sets, crisp and fuzzy. Calibrating to crisp sets is essentially a matter of recoding the raw data, establishing a certain number of thresholds for each value of the calibrated set. When binary crisp sets are intended to be obtained, a single threshold is needed to divide the raw values in two categories: those below the threshold will be allocated a value of 0 (out of the set) and those above the threshold a value of 1 (in the set). When multi-value crisp sets are intended, there will be two thresholds to divide into three categories, an so on. The general formula for the number of thresholds is the number of values minus 1.

Even for this (crude) type of recoding, the values of the thresholds should not be mechanically determined. A statistician will likely divide the values using the median, which would many times be a mistake. It is not the number of cases that should determine the value of the threshold, but rather the meaning of the concept and the expert's intimate knowledge about which cases belong to which category.

For instance, there will be a certain value of the threshold to divide countries' GDP in the set of "developed countries", and a different value of the threshold for the set of "very developed countries". The exact value should be determined only after an inspection of the distribution of GDP values, especially if they are not clearly clustered. In such a situation, the researcher's experience should act as a guide in establishing the best threshold value that would correctly separate different countries in different categories, even if the difference is small. All of this process should be thoroghly described in the methodological part, with strong theoretical justifications for the chosen value of the threshold.

Calibrating to fuzzy sets is more challenging, and at the same time more interesting because there are multiple ways to obtain fuzzy membership scores from the same raw numerical data. The most widely used is called the "direct method", first described by Ragin (2000). It uses the logistical function to allocate membership scores, using the exclusion, cross-over and inclusion thresholds.

Table 4 below displays the two relevant columns extracted from Ragin's book, first with the national income in US dollars, and the second the degree of membership (the calibrated counterparts of the national income) into the set of developed countries.

Table 4: Per capita income (INC), calibrated to fuzzy sets membership scores (fsMS).

	INC	fsMS
Switzerland	40110	1.00
United States	34400	1.00
Netherlands	25200	0.98
Finland	24920	0.98
Australia	20060	0.95
Israel	17090	0.92
Spain	15320	0.89
New Zealand	13680	0.85
Cyprus	11720	0.79
Greece	11290	0.78
Portugal	10940	0.77
Korea, Rep.	9800	0.72

Argentina	7470	0.62
Hungary	4670	0.40
Venezuela	4100	0.25
Estonia	4070	0.25
Panama	3740	0.18
Mauritius	3690	0.17
Brazil	3590	0.16
Turkey	2980	0.08
Bolivia	1000	0.01
Cote d'Ivoire	650	0.01
Senegal	450	0.00
Burundi	110	0.00

At the top of the list, Switzerland and the United States are certainly developed country, which explain their full membership score of 1, while Senegal and Burundi, with national incomes of 450 USD and 110 USD respectively, are too poor to have any membership whatsoever in the set of developed countries.

What threshold values best describe this set, and how are the membership values calculated? A quantiative quick solution would be to calculate the ratio of every other country from the income of Switzerland, the richest country in that data.

Aside from the fact such a method is mechanical and data driven, it would immediately become obvious that for instance Netherlands (which currently has an almost full inclusion of 0.98 in the set of developed countries), would have a ratio equal to 0.628 that does not seem to accurately reflect our knowledge. Likewise, a median value of 8635 USD would leave Argentina more out of the set than more in, and the average of 11294 USD is even more misleading, leaving Greece more out than in.

Ragin started by first deciding the crossover threshold at a value of 5000 USD, which is the point of maximum ambiguity about a country being in more in than more out of the set of developed countries. He then applied some mathematical calculations based on the logistic function and the associated log odds, arriving at a full inclusion score of 20000 USD (corresponding to a membership score of at least 0.95 and the log odds of membership of at least +3) and a full exclusion score of 2500 USD (corresponding to a membership score of at most 0.05, and a log odds of membership lower than -3).

Employing the logistic function, the generated set membership scores follow the familiar increasing S shape displayed in Figure 1, but this function is only one among many other possible ones to perform calibration. Linear mathematical transformations are also possible, such as the one from the Equation (4), as extracted from Dușa (2019, p.84):

$$dm_x = \begin{cases} 0 & \text{if } x \leq e, \\ \frac{1}{2}\left(\frac{e-x}{e-c}\right)^b & \text{if } e < x \leq c, \\ 1 - \frac{1}{2}\left(\frac{i-x}{i-c}\right)^a & \text{if } c < x \leq i, \\ 1 & \text{if } x > i. \end{cases} \tag{4}$$

where:

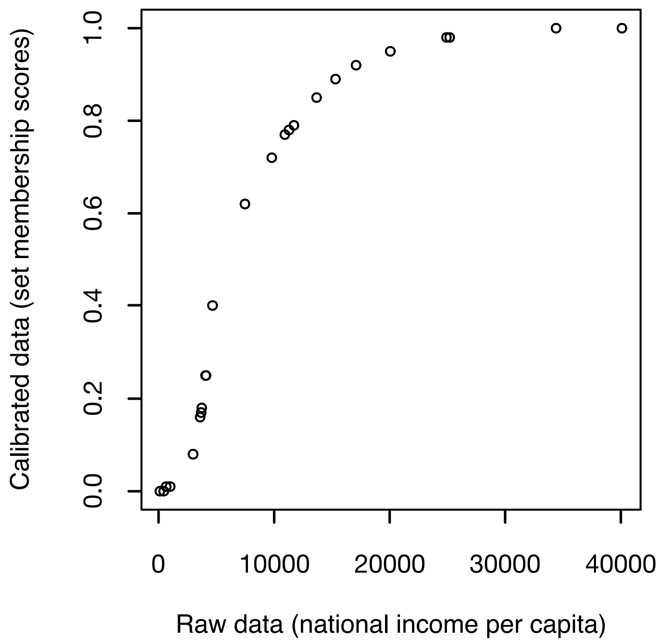


Figure 1: Calibration in the set of developed countries

- e is the threshold for full exclusion
- c is the crossover
- i is the threshold for full inclusion
- x is the raw value to be calibrated
- b determines the shape below the crossover (linear when $b = 1$ and curved when $b > 1$)
- a determines the shape above the crossover (linear when $a = 1$ and curved when $a > 1$)

The calibration functions in Figure 2 refer to the calibration of 100 randomly selected heights ranging from 150 cm to 200 cm. These values are calibrated in the set of "tall people" (the linear increasing function that could act as a replacement for the logistical S shape) as well as in the set of "average height people" (with a triangular shape, and also with a trapezoidal shape). This is an example that shows how the calibrated values depend on the conceptual meaning of the calibrated set. All shapes refer to people's heights and all use the exact same raw values, but the meaning is different for "average height" and for "tall" people.

The set of three threshold values (155 for full exclusion, 175 for the crossover and 195 for full inclusion) can be used only for the increasing linear that approximates an S shape for the set of "tall" people. The other linear functions that approximate a bell shape (for the set of "average height" people) are more challenging, and need a set of six values for the thresholds (three for the first part that increases towards the middle, and the other three for the second part that decreases from the middle towards the higher heights). There are two full exclusion thresholds, two crossover values and finally two full inclusion thresholds (that coincide for the triangular shape), the calibrated values being obtained via the mathematical transformations from Equation (5):

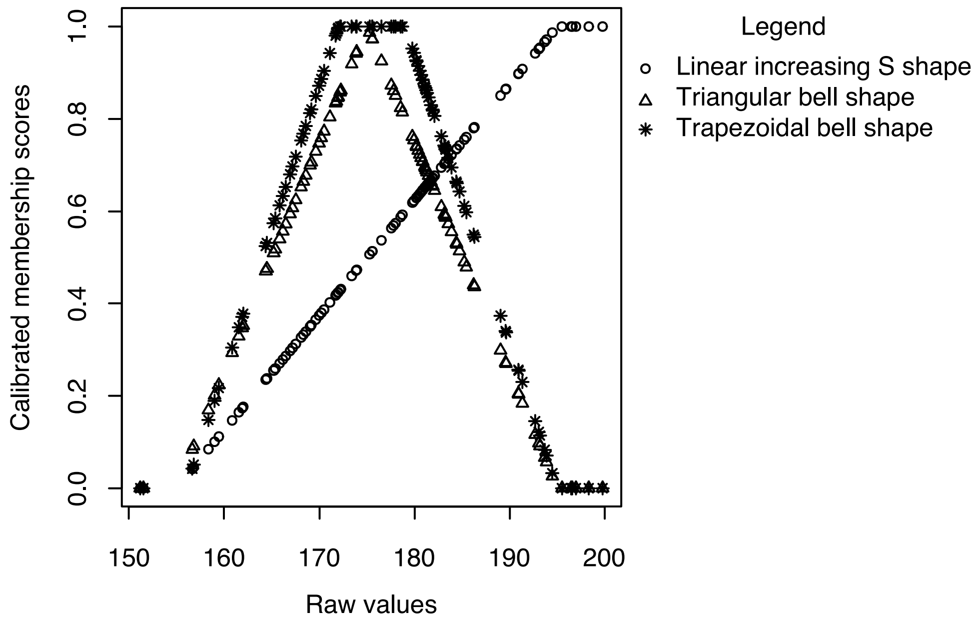


Figure 2: Other possible calibration functions

$$dm_x = \begin{cases} 0 & \text{if } x \leq e_1, \\ \frac{1}{2}\left(\frac{e_1 - x}{e_1 - c_1}\right)^b & \text{if } e_1 < x \leq c_1, \\ 1 - \frac{1}{2}\left(\frac{i_1 - x}{i_1 - c_1}\right)^a & \text{if } c_1 < x \leq i_1, \\ 1 & \text{if } i_1 < x \leq i_2, \\ 1 - \frac{1}{2}\left(\frac{i_2 - x}{i_2 - c_2}\right)^a & \text{if } i_2 < x \leq c_2, \\ \frac{1}{2}\left(\frac{e_2 - x}{e_2 - c_2}\right)^b & \text{if } c_2 < x \leq e_2, \\ 0 & \text{if } x > e_2. \end{cases} \tag{5}$$

Apart from the direct method, Ragin also presented an "indirect" one, to obtain fuzzy membership scores from interval level raw data. In this method, no qualitative anchors (thresholds) need to be specified in advance, but rather it involves creating an artificial dependent variable where each case is allocated a certain fuzzy membership category from 0 to 1 (usually six, an even number to avoid the point of maximum ambiguity 0.5), then perform a (quasi)binomial logistic regression using a fractional polynomial equation with the raw values as independent variable, against the newly formed dependent variable containing the fuzzy membership categories (for more details, see Dușa, 2019, p.92).

A different type of calibration is applied for categorical causal conditions (for instance, containing values from a response Likert type scale). It is not possible to determine any thresholds, because the variation is extremely small and data can sometimes be severely skewed, which limits the variation even more. For the same reasons, no regression equation can be applied with the indirect method, since it assumes at least the independent variable to be metric.

A possible solution to this problem is to manually allocate fuzzy membership scores for each category (the so called "direct assignment" method), introduced by Verkuilen (2005) who also criticized it for containing bias due to researcher's subjectivity. Verkuilen mentions a possibly better solution, by employing the TFR - Totally Fuzzy and Relative method (Cheli and Lemmi, 1995), making use of the empirical cumulative distribution function of the observed data E, then calculating the distance from each CDF value to the first value from the Likert scale (1), and divide that to the distance between 1 (the maximum possible fuzzy score) to the same CDF of the first value 1 in the same Likert response scale:

$$TFR = max\left(0, \frac{E(x) - E(1)}{1 - E(1)}\right) \tag{6}$$

Calibration is a very important topic in set theoretical methods, as many of the subsequent results depend on this operation. It should not be a mechanical process, but rather an informed activity where the researcher should present the methodological reasons that led to one method or another.

Set membership scores vs. probabilities

Despite this topic being discussed numerous times before (Kosko, 1990; Dubois and Prade, 1989; Zadeh, 1995; Ragin, 2008; Schneider and Wagemann, 2012), and attempts to combine set theory and statistics (Heylen and Nachtegael, 2013) set membership scores and probabilities can still be confused as both range from 0 to 1 and, at a first glance, seem very similar.

Before delving into the formal difference, consider the following example involving a potentially hot stove. If the stove has a 1% probability of being very hot, there is still a small (but real) chance to get severely burned when touching it. But if we say the stove has a 1% inclusion in the set of hot objects, the stove can be safely touched without any risk of getting burned, whatsoever.

Ragin's example with the glass of water has the same interpretation. If there is a 1% probability the glass will contain a deadly poison, there is a small but definite chance to die drinking that water. But if the glass has a 1% inclusion in the set of poisenous drinks, there is absolutely no risk to die.

Intuitive as they may seem, these two examples still don't explain the fundamental difference. At formal level, the probability has to obey the Kolmogorov axioms:

- the probability of an event that is certain is equal to 1: $P(C) = 1$
- the probability of an impossible event is equal to 0: $P(\varnothing) = 0$
- if two events do not overlap ($A \cap B = \varnothing$), then $P(A + B) = P(A) + P(B)$

The probability can essentially be interpreted as a relative frequency obtained from an infinite repetition of an experiment. It is a frequentist statistic (based on what is called a frequentist approach), where the conclusions are drawn from the relative proportions in the data.

But frequencies can only be computed for categorical variables, in this situation for events either happening or not. To calculate probabilities (relative frequencies) there are only two possible values for the event: 1 (true, happening) or 0 (false, not happening). The first section already presented the different types of sets, and this corresponds to the definition of a binary crisp set.

Therefore the meaning of probability is necessarily related to crisp sets, while membership scores are related to fuzzy sets. They simply refer to different things, given that crisp sets are only particular cases of fuzzy sets. Set membership scores refer to various degrees of membership to a set, they are related to the uncertainty about set membership that cannot be computed the same as a probability because the set itself is not crisp, but fuzzy.

When flipping a coin, there are only two possible outcomes (heads or tails) and an exact probability of occurence for each can be computed by flipping the coin numerous times. These are clear-cut categories (either heads, or tails) but not all concepts are so clear. Whether a person is "young" is a matter of uncertainty, and each person can be included (more, or less) in the set of young people. Same with "smart", "healty" etc, which cannot be determined unequivocally.

There are situations where probabilities and fuzzy sets can be combined (Singpurwalla and Booker, 2004; Demey, Kooi and Sack, 2017), especially with Bayesian probabilities (Mahoney, 2016; Fairfield and Charman, 2017; Barrenechea and Mahoney, 2017) in conjunction with process tracing, but these two concepts do not completely overlap. In the words of Zadeh (1995) himself, they are "complementary rather than competitive".

Polarity and asymmetry

There is an even deeper layer of understanding that needs to be uncovered with respect to probabilities and fuzzy sets. Describing probability, Kosko (1994, p.32) shows that it works with bivalent sets only (an event either happens or it does not happen), and another important difference refers to how a set relates to its negation.

In probability theory, $A \cap \sim A = \varnothing$, and $A \cup \sim A = 1$. For fuzzy sets, it turns out that $A \cap \sim A \neq \varnothing$, and $A \cup \sim A \neq 1$. These inequalities (especially the first one) essentially entail that objects can be part of both a set and its negation, and the union of the two sets might not always be equal to the universe.

This has deep implications over how we relate to events, their negation and the common misperception of bipolarity. Sets are unipolar, therefore a bipolar measurement scale (for instance, a Likert type response scale) cannot be easily accommodated with a single set.

In a bipolar space, "good" is the opposite of "bad". But a "not bad" thing is not precisely the same as a "good" thing: it is just not bad. Same with "ugly" vs. "beautiful": if a thing is not ugly, that does not means it is necessarily beautiful, or the other way round, something that is not beautiful is not necessarily ugly. Bauer et al. (2014) encountered similar difficulties in evaluating a bipolar scale with left-right political attitudes, analysing the vagueness of the social science concepts in applied survey research.

Things, or people, can have membership scores of more than 0.5 in both a set and its negation. A person can be both happy and unhappy at the same time, therefore translating a bipolar scale into a single set is difficult, if not impossible. There should be two sets, first for the happy persons and the second for the unhappy ones, and a person can be allocated membership scores in both, such that the sum of the two scores can exceed 1 (something impossible with probabilities).

The set negation leads to another point of misunderstanding between quantitative statistics (especially the correlation based techniques, for instance the regression analysis) and set theoretic methods. Numerous articles have been written comparing empirical results (Katz, vom Hau and Mahoney, 2005; Grendstad, 2007; Grofman and Schneider, 2009; Fujita, 2009; Woodside, 2014), pointing to the deficiencies of regression techniques (Pennings, 2003; Marx and Soares, 2015), criticizing fuzzy sets (Munck, 2016; Paine, 2015; Seawright, 2005), and revealing the advantages of fuzzy sets (Cooper and Glaesser, 2010) or more recently focusing on the integration and complementarity between the two methods (Skaaning, 2007; Mahoney, 2010; Fiss, Sharapov and Cronqvist, 2013; Radaelli and Wagemann, 2018).

The sheer amount of written publications suggest at least a couple of things. First, that set theoretic methods are increasingly used in a field traditionally dominated by the quantitative analysis, and second, there is a lot of potential for these methods to be confused (despite the obvious differences) as they both refer to explanatory causal models for a given phenomenon.

Correlation based techniques assume an ideal linear relation between the independent and dependent variables. When high values of the dependent variable (that can be interpreted as the "presence" of the outcome, in set theory) are explained by high values of the independent variable(s), then low values of the dependent ("absence" of the outcome) are necessarily explained by low values of the independent(s).

By contrast, set-theoretical methods do not assume this kind of linearity. While the presence of the outcome can be explained by a certain configuration of causal conditions, the absence of the outcome can have a very different explanation, involving different causal combinations. If welfare can be explained by the combination of education and health, it is perfectly possible for the absence of welfare to be explained by different causes.

While the correlation based analyses are symmetric with respect to the dependent variable, the set theoretic methods are characterised by an asymmetric relation between a set of causes and a certain outcome. This is a fundamental ontological difference that separates the two analysis systems, which should explain both why they are sometimes confused, as well as why their results are seemingly different.

Necessity and sufficiency

Natural language abounds with expressions containing the words "necessary" and "sufficient". In trying to identify the most relevant conditions that are associated with an outcome, theorists often ask: what are the necessary conditions for the outcome? (without which the outcome cannot happen), or what conditions are sufficient to trigger an event? (that when present, the event is guaranteed to happen).

The contrast between the correlational perspective and the set theoretic methods can be further revealed by analyzing the Figure 3. The crosstable on the left side is a typical, minimal representation of the quantitative statistical perspective, focused on the perfect correlation from the main diagonal. Everything off the main diagonal is problematic and decreases the coefficient of correlation.

The crosstable on the right side, however, tells a different story. In the language of statistics, the 45 cases in the upper left quadrant potentially ruin the correlation, but they make perfect sense from a set theoretical point of view: since there are no cases in the lower right quadrant, this crosstable tells the story of X being a perfect subset of Y. The "problematic" upper left quadrant simply says there are cases where Y is present and X is absent, in other words X does not cover (does not "explain") all of Y.

The zero cases in the lower right quadrant, combined with the 14 cases in the upper right quadrant, say there is no instance of X where Y is absent, which means that X is completely included in Y (it

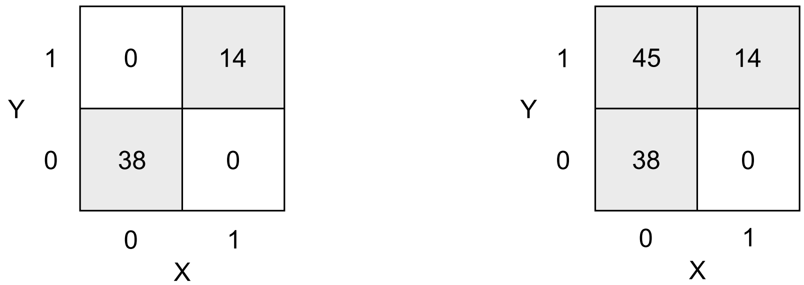


Figure 3: Correlation (left) and subset sufficiency (right)

is a subset of Y). Whenever X happens, Y happens as well, that is to say X is "sufficient" for Y ("if X, then Y").

This is a different type of language, a set theoretical one, that is foreign to the traditional quantitative analysis. Regression analysis and the sufficiency analysis have the very same purpose, to seek the relevant causal conditions for a given phenomenon. But when inspecting for sufficiency, the focus is not the main diagonal (correlation style) but rather on the right side of the crosstable where X happens (where X is equal to 1).

This is naturally a very simplified example using just two values for both X and Y. Quantitative researchers would be right to argue that, when the dependent variable is binary, a logistic regression model is more appropriate than a linear regression. But set theoretical data need not necessarily be crisp; they can also be fuzzy with a larger variation between 0 and 1 and a cross-table is not enough to represent the data.

At a closer inspection on Figure 4, the situation is identical for fuzzy sets. The left plot displays the characteristic elipse shape of the cloud, with a very positive correlation between the independent and the dependent variables. It does not really matter whether the points are located above or below the diagonal, as long as they are close.

The cloud of points from the right plot would be considered problematic. Not only the points are located far from the main diagonal (ideally, the regression line) but they also display inconstant variance (a phenomenon called heteroskedasticity). However, this is not problematic for set theory: as long as the points are located above the main diagonal (values of X always smaller than corresponding values of Y), it is a perfect representation of a fuzzy subset relation. In set theoretical language, such a subset relation is also described as perfectly "consistent".

Not all subset relations are that perfect. In fact, there can be situations where X can happen and Y to be absent, without affecting the sufficiency relation (too much). Just as there are no countries with perfect democracies (they are "more or less" democratic), situations with perfect sufficiency are also extremely rare. When perfect sufficiency happens, it is mainly the result of our calibration choice: it can happen in crisp sets, but this is almost never observed with fuzzy sets.

The concept of fuzziness teaches us that conditions can be "more or less" sufficient, just as two sets can be more or less included one into the other. The causal set should be "consistent enough" with (or "included enough" in) the outcome set, to be accepted as sufficient

The big question is how much of outcome set Y is explained by causal set X, a very common question in traditional statistics that is usually answered with the R^2 coefficient in the regression analysis. In set theory, this is a matter of coverage. There can be situations with imperfect consistency but large

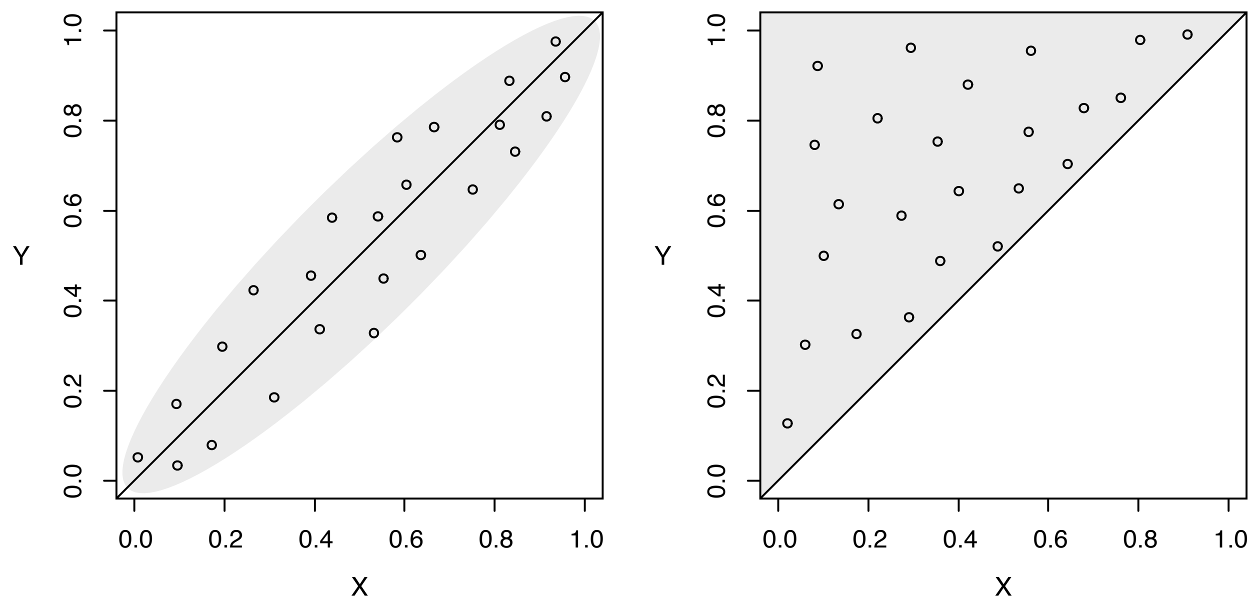


Figure 4: Fuzzy correlation (left) and fuzzy subset sufficiency (right)

coverage, and perfect consistency but low coverage.

Out of the two situations in Figure 5, the relation from the left plot is the most relevant. Despite the imperfect consistency (inclusion), the causal condition X covers a lot of the cases in the outcome Y, qualifying as a highly relevant (albeit imperfect) sufficient condition for Y.

In the plot from the right side, X is perfectly consistent with Y but it covers only a very small area, which means there are very many cases in Y that are not explained by X, suggesting we should search for more causal conditions that explain the entire diversity of the outcome's presence. In such situations, X is called sufficient but not necessary, an expression which is also described by the concept of "equifinality": the very same outcome can be produced via multiple causal paths, just as there are many roads that lead to the same city.

Inclusion and coverage can be precisely measured, with the same formula being valid for both crisp and fuzzy sets. Equation (7) calculates the consistency for sufficiency (*inclS*), while Equation (8) calculates the coverage for sufficiency (*covS*), where the sufficiency relation is denoted by the forward arrow sign $X \Rightarrow Y$:

$$inclS_{X \Rightarrow Y} = \frac{\sum min(\mathrm{X},\ \mathrm{Y})}{\sum \mathrm{X}} \tag{7}$$

$$covS_{X \Rightarrow Y} = \frac{\sum min(\mathrm{X},\ \mathrm{Y})}{\sum \mathrm{Y}} \tag{8}$$

In the regression analysis, independent variables may be collinear, meaning they explain the same part of the dependent variable's variation. This is usually detected with the contribution of each independent variable to the model's R^2 coefficient: only those variables that contribute a significant increase of the R^2 are preferred.

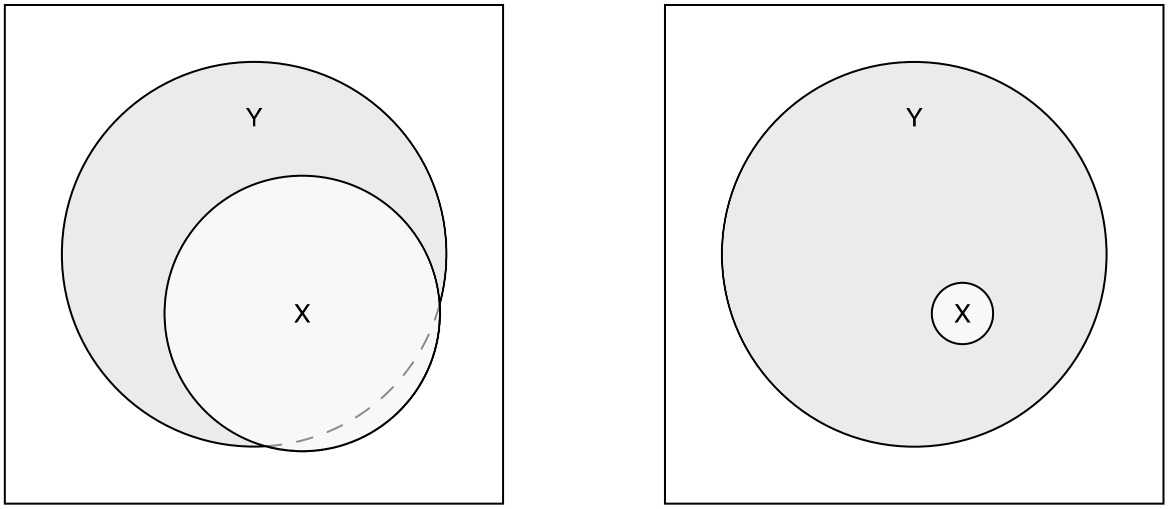


Figure 5: Incomplete inclusion / large coverage (left) and complete inclusion / low coverage (right)

Similarly, in set theory the causal conditions have a so called "raw" coverage and also a "unique" coverage. Their unique coverage (*covU*) is the area from the outcome Y which is solely covered by a certain causal condition, as shown in Equation (9) and Figure 6.

$$covU_{A \Rightarrow Y} = \frac{\sum min(\mathrm{Y, A})}{\sum \mathrm{Y}} - \frac{\sum min(\mathrm{Y, A,}\ max(\mathrm{B, C, ...}))}{\sum \mathrm{Y}} \tag{9}$$

In Figure 6, the unique coverage of condition A can be computed as the area of Y covered by A, minus the intersection of A and B (its area jointly covered by condition B). More generally, minus the intersection between A and the union of all other causal conditions that cover the same area of Y covered by A).

Necessity and sufficiency are mirrored concepts. While sufficiency is about the subset relation of the causal condition within the outcome set, necessity is the other way round: the superset relation of the causal condition over the outcome set. A causal condition is necessary *iff* it is a superset of the outcome: when Y happens, X is always present. When the outcome Y does not occur in the absence of X, it means that X is necessary.

The upper left quadrant in a 2 $\times$ 2 crosstable should be empty (where Y = 1 and X = 0), and correspondingly the area above the main diagonal in a fuzzy XY plot should also be emtpy in order to determine necessity.

Mirrored scores for the consistency of necessity (*inclN*, how much of Y is included in X), as well as for the coverage of necessity (*covN*, how much of X is covered by Y) can be calculated, as shown in Equations (10) and (11):

$$inclN_{X \Leftarrow Y} = \frac{\sum min(\mathrm{X,\ Y})}{\sum \mathrm{Y}} \tag{10}$$

$$covN_{X \Leftarrow Y} = \frac{\sum min(\mathrm{X,\ Y})}{\sum \mathrm{X}} \tag{11}$$

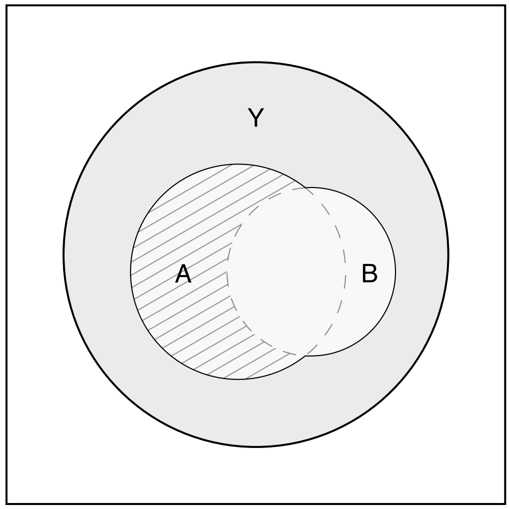


Figure 6: Unique coverage of A (hashed area)

When analysing necessity, the most important thing is to determine how relevant a necessary condition is. Oxygen is a necessary condition for a fire, but it is an irrelevant necessary condition as oxygen can be found everywhere, and in most situations where oxygen is present, a fire is not observed. A more important necessary condition would be the heat, and another necessary condition may be a sparkle. Both of these are truly necessary (hence relevant) to start a fire.

The relevance of a necessary condition is revealed by the coverage score. If the outcome Y covers only a very small area of the causal condition, it is a sign that X might be irrelevant.

Goertz (2006*a*) is a leading scholar in the analysis of necessity, further differentiating between irrelevant and trivial necessary conditions. Irrelevance and triviality are somewhat similar and they are frequently used as synonims, but there is a subtle difference between them. Triviality is a maximum of irrelevance, to the point that not only the subset outcome Y covers a very small area of the causal condition X, but the superset condition X becomes so large that it fills the entire universe. When trivial, the causal condition is omnipresent, with no empirical evidence of its absence.

In the previous example, oxygen is an irrelevant, but not exactly a trivial necessary condition for a fire, as there are many (in fact, most) places in the Universe where oxygen is absent. In the Euler/Venn diagram from Figure 7, Y is completely consistent with X but it covers a very small area. More over, it can be noticed that X occupies the entire universe represented by the rectangle: it is an omnipresent necessary condition.

The same line of reasoning can be applied on the XY plot from the right side, where the focus on necessity is the area below the main diagonal, and X is trivial since all of its points are locate on the extreme right where X is always equal to 1, and most of the points are located in the lower half of the plot where Y is more or less absent (below the crossover 0.5 point).

A condition becomes less and less trivial (hence more and more relevant) when the points move away from the extreme right where X is always equal to 1, towards the main diagonal. Goertz proposed a measure of triviality by simply measuring the distance between the fuzzy values and 1. Later, Schneider and Wagemann (2012) advanced Goertz's work and proposed a measure called Relevance of Necessity (RoN), that is the current standard to complement the coverage score for necessity:

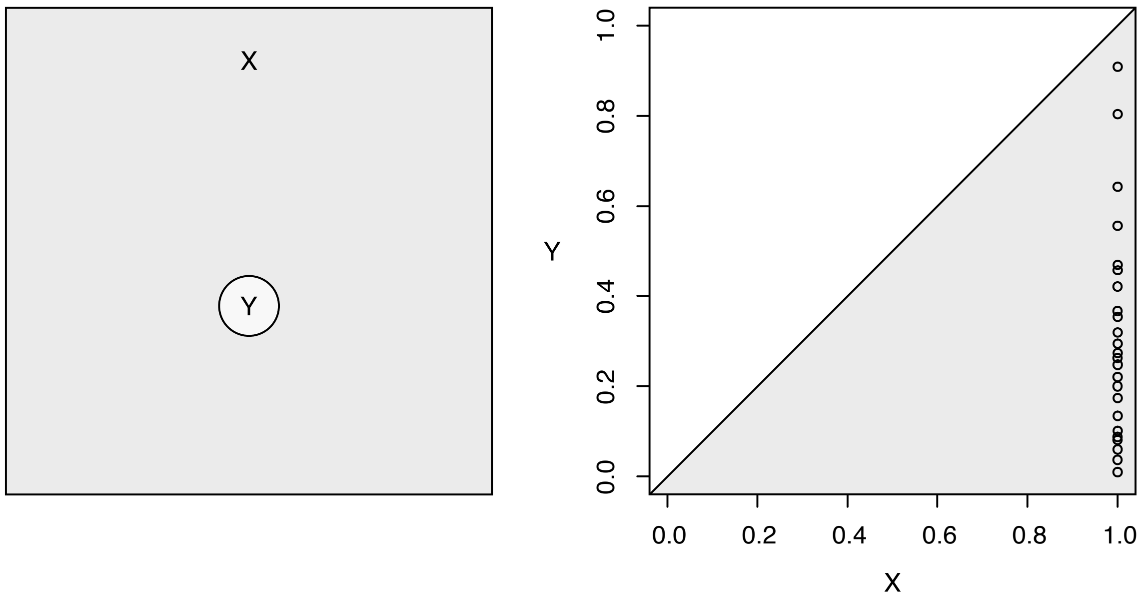


Figure 7: X as a trivial necessary condition for Y

$$RoN = \frac{\sum(1 - X)}{\sum(1 - min(X, Y))} \tag{12}$$

The XY plots and Venn/Euler diagrams have a couple of more interesting properties to discuss. If points are located all over the plot, there is no clear relationship between the cause and the outcome. We expect the points to be positioned either above the main diagonal (for sufficiency) or below (for necessity). If that happens, it means that if a cause is perfectly sufficient it is usually not necessary, and conversely if it is perfectly necessary it is usually not sufficient.

Ideally, we would like to find a causal condition that is both necessary and sufficient for and outcome, thus having the greatest explanatory power. At a first sight, that would seem impossible since a causal set X cannot be a subset and a superset of the outcome Y, at the same time. It may in fact happen when the two sets coincide: the subset X covers 100% of the set Y.

In terms of XY plots, the points are located neither above, nor below the main diagonal. When the two sets coincide, the points are located exactly along the main diagonal, which would also correspond to a (close to) perfect correlation in the statistical tradition, similar to the left plot from Figure 4.

But such a perfect correlation is difficult to obtain in practice, and it would usually mean we are not dealing with two different concepts (for the cause and for the effect) but with one and the same concept under two different measurements. No causal set is perfectly correlated with the outcome, and perhaps more importantly a single causal set is neither necesary nor sufficient by itself. It is very rare to obtain an explanatory model with a single causal condition, a typical outcome being produced by various combinations of causes.

Causal factors combine in conjunction, which is set theory are set intersections. Where a single cause might not be (sufficiently) included into an outcome set, an intersection with other condition(s) might be small enough to fit.

Table 5: Boolean minimization example

A	B	Y
1	1	1
1	0	1
1	-	1

The same thing happens for necessity, but in reverse. If a single causal condition is not big enough to qualify as a necessary superset of the outcome, disjunctions (set unions) of two or more causal conditions might eventually form a big enough superset to cover the outcome. But if conjunctions are easy to interpret (the simultaneous presence of two causal sets), disjunctions need to have theoretically valid interpretations, much like the quantitative researchers having to find a meaningful interpretation for the latent constructs resulted from the principal component analysis.

More recent and interesting developments in the analysis of necessity include the NCA - Necessary Condition Analysis by Dul (2016), while on sufficiency Schneider and Rohlfing (2016) bring important insights in the cutting edge, so called STMMR - Set Theoretic Multi-Method Research which is an entire topic on its own and deserves a separate and more extended presentation.

Set Theory and the Qualitative Comparative Analysis

Having presented the background of set theory, the stage is set to introduce a way (a third way) to tackle research problems traditionally aproached through the qualitative and quantitative methods.

The trouble with quantitative research is that it needs many cases (a large N) to make the Central Limit Theorem work, and a typical political science research compares only a handful of countries or events and does not have that many cases. There are only 28 countries in the European Union, and a comparative study on the Eastern European countries will have even less cases. When studying very rare events such as revolutions, Skocpol (1979) had only three cases to work upon: France, Russia and China.

It is difficult to argue there is an underlying, potentially infinite population of "possible" such events to draw large samples from, in order to justify the use of the quantitative analysis, even with Monte Carlo simulations for small samples. On the other side, the qualitative analysis is very much case oriented and produces perfect explanations for all individual cases. This is often useful for theory formation, but it is usually regarded as too specific to have generalizable value.

With both sides having strong arguments to defend one method or another in different situations, Ragin (1987) employed set theory and Boolean algebra to import a methodology created for electrical engineering (Quine, 1955; McCluskey, 1956), into the social and political sciences. He showed how, through a systematic comparative analysis of all possible pairs of cases, the relevant causal factors can be identified and the irrelevant ones eliminated. More importantly he showed how to identify the patterns, or the combinations of causal conditions that are sufficient to produce an outcome.

The essence of the entire procedure can be reduced to a process called Boolean minimization, which was itself imported into the electrical engineering from the canons of logical induction formulated by J.S. Mill (1843).

The two expressions in Table 5 are equivalent to AB + A$\sim$B, which can be simplified to A alone since the condition B is redundant, present in the first and absent in the second: A(B + $\sim$B) = A. In such an example, B is said to be "minimized" (or eliminated), hence the name of the Boolean minimization procedure.

Each case that is added to the analysis displays a certain combination (of presence or absence) of causal conditions, and the algorithm exhaustively compares all possible pairs cases to first identify if they differ by only one literal, then iteratively and progressively minimize until nothing else can be further simpified. The final product of this procedure is the so-called "prime implicants", which are simpler (more parsimonious) but equivalent expressions to the initial, empirically observed cases.

Since pairs of cases are compared, the process is more qualitative than quantiative, therefore the "Q" in QCA stands for the "Qualitative" Comparative Analysis. It has absolutely nothing to do with traditional statistics, yet it employs a systematical and solid mathematical algorithm such as the Boolean minimization to identify the minimal configurations of (relevant) causal conditions which are sufficient to produce an outcome.

Crisp sets are very attractive, by allowing to map the empirically observed configurations over a finite number of combinations of presence / absence for the causal conditions (equal to $\prod l_c$, where l is the number of levels for each causal condition $c = 1 \ldots n$). This finite space is called a truth table, and it contains all positive and negative observed configurations, as well as those for which there is no empirical information about (called "remainders").

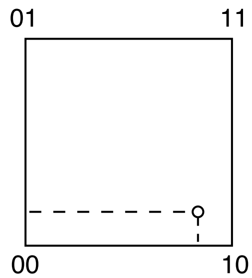
But it is precisely the "Boolean" nature of the algorithm which attracted a lot of criticism (Goldthorpe, 1997; Goldstone, 1997), since it suggests a very deterministic view of the reality (Lieberson, 1991) and as pointed many times before, most social phenomena are not simply present or absent, but somewhere in between.

The debate led to an upgrade of the Qualitative Comparative Analysis from Boolean to fuzzy sets (Ragin, 2000, 2008). Instead of crisp values, each case has a membership score for each of the causal condition sets. The challenge, that was also solved by Ragin (2004), was to translate fuzzy membership scores to truth table crisp scores, because the minimization process is Boolean.

In the fuzzy version, the truth table configurations act as the corners of a multidimensional vector space where the set membership scores play the role of fuzzy coordinates for the position of each case. Figure 8 presents the simplest possible vector space with two dimensions, and a case having two fuzzy membership scores of 0.85 on the horizontal and 0.18 on the vertical. For only two causal conditions, the truth table contains $2 \cdot 2 = 4$ rows, represented by the corners of the square, and the case is located close to the lower right corner 10 (presence of the first condition and absence of the second).

It is rather clear to which truth table configuration does the case belongs to in this example, but it would be more difficult to assess were the case be located close to the middle of the vector space. Ragin's procedure uses the fuzzy coordinates of each case to calculate consistency scores for each of the corners, and finally determine to which corner are the cases more consistent with (or closest to).

The consistency score of this case is the set intersection between the first membership score (0.85) and the negation of the second ($1 - 0.18 = 0.82$), which is the fuzzy minimum between 0.85 and 0.82, equal to 0.82. Provided there is no fuzzy membership score of exactly 0.5 (the point of maximum ambiguity), there is only one corner to which cases have a higher than 0.5 consistency.



A	B	cons.
0	0	
0	1	
1	0	0.82
1	1	

Figure 8: Bidimensional vector space (left) and the corrresponding truth table (right)

The corners of the vector space can be interpreted as genuine ideal types in the Weberian tradition, that an imperfect fuzzy configuration is most similar to. Upon determining where each case is ideally positioned in the truth table configurations, the algorithm proceeds with the same Boolean minimization procedure as in the crisp version, to identify minimally sufficient configurations that are related to the presence (or absence) of an outcome.

It is beyond the purpose of this chapter to offer a complete presentation of the QCA procedure with all its details. There are entire books written for this purpose (Ragin, 2000, 2008; Rihoux and Ragin, 2009; Schneider and Wagemann, 2012; Dușa, 2019), and the intereste reader is invited to consult the relevant literature. The main purpose was to reveal how the language of sets and the Boolean algebra can be employed for social and political research.

To conclude, set theoretic methods are rather young compared to the long established quantitative tradition, but they already compensate through a sound and precise mathematical procedure that uses set relations (subsets and supersets) to identify multiple conjunctural causation, where the outcome can be produced via several (sufficient) combinations of causal conditions.

Different from the strict statistical assumptions in the quantitative analysis, the causal conditions in QCA are not assumed to be independent of each other. What matters is how they conjunctively combine to form sufficient subsets of the outcome, and their relevance in terms of both coverage of the outcome and how well they explain the empirically observed configurations.

References

Aiton, Eric J. 1985. *Leibniz. A Biography*. Bristol and Boston: Adam Hilger Ltd.

Arfi, Badredine. 2010. *Linguistic Fuzzy Logic Methods in Social Sciences*. Vol. 253 of *Studies in Fuzziness and Soft Computing* Berlin, Heidelberg: Springer Verlag.

Babst, Dean V. 1964. "Elective Governments. A Force for Peace." *The Wisconsin Sociologist* 3(1):9–14.

Barrenechea, Rodrigo and James Mahoney. 2017. "A Set-Theoretic Approach to Bayesian Process Tracing." *Sociological Methods & Research* .

Bauer, Paul C., Pablo Barberá, Kathrin Ackermann and Aaron Venetz. 2014. "Vague Concepts in Survey Questions: A General Problem Illustrated with the Left-Right Scale." *SSRN Electronic Journal* . Available on: http://ssrn.com/abstract=2467595.

Cheli, Bruno and Achille Lemmi. 1995. "A "Totally" Fuzzy and Relative Approach to the Multidimensional Analysis of Poverty." *Economic Notes* 1:115–134.

Cooper, Barry and Judith Glaesser. 2010. "Contrasting Variable-Analytic and Case-Based Approaches to the Analysis of Survey Datasets: Exploring How Achievement Varies by Ability across Configurations of Social Class and Sex." *Methodological Innovations* 5(1):4–23.

Dauben, Joseph Warren. 1979. *Georg Cantor. His Mathematics and Philosophy of the Infinite*. New Jersey: Princeton University Press.

Demey, Lorenz, Barteld Kooi and Joshua Sack. 2017. Logic and Probability. In *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta. Summer 2017 ed. Metaphysics Research Lab, Stanford University.

Dubois, Didier and Henri Prade. 1989. "Fuzzy sets, probability and measurement." *European Journal of Operational Research* 40:135–154.

Dul, Jan. 2016. "Necessary Condition Analysis (NCA). Logic and Methodology of "Necessary but Not Sufficient" Causality." *Organizational Research Methods* 19(1):10–52.

Dușa, Adrian. 2019. Calibration. In *QCA with R. A Comprehensive Resource*, ed. Adrian Dușa. Cham, Switzerland: Springer International Publishing.

Fairfield, Tasha and Andrew E. Charman. 2017. "Explicit Bayesian Analysis for Process Tracing: Guidelines, Opportunities, and Caveats." *Political Analysis* 25(3):363–380.

Fiss, Peer C., Dmitry Sharapov and Lasse Cronqvist. 2013. "Opposites Attract? Opportunities and Challenges for Integrating Large-N QCA and Econometric Analysis." *Political Research Quarterly* 66(1):191–198.

Fujita, Taisuke. 2009. "Developed and Democratic Countries' Policy-making on Dispute Settlement in the GATT/WTO: Exploring Conjunctural and Multiple Causations by Comparing QCA and Regression Analysis." *Sociological Theory and Methods* 24:181–202.

Goertz, Gary. 2006*a*. "Assessing the Trivialness, Relevance, and Relative Importance of Necessary or Sufficient Conditions in Social Science." *Studies in Comparative International Development* 41(2):88–109.

Goertz, Gary. 2006*b*. *Social Science Concepts. A User's Guide*. Princeton and Oxford: Princeton University Press.

Goldstone, Jack A. 1997. "Methodological issues in comparative macrosociology." *Comparative Social Research* 16:121–132.

Goldthorpe, John H. 1997. "Current issues in comparative macrosociology: a debate on methodological issues." *Comparative Social Research* 16:1–26.

Grendstad, Gunnar. 2007. "Casual Complexity and Party Preference." *European Journal of Political Research* 46:121–149.

Grofman, Bernard and Carsten Q. Schneider. 2009. "An Introduction to Crisp Set QCA, with a Comparison to Binary Logistic Regression." *Political Research Quarterly* 62(4):662–672.

Halmos, Paul R. 1974. *Naive Set Theory*. New York: Springer.

Heylen, Ben and Mike Nachtegael. 2013. "The integration of fuzzy sets and statistics: toward strict falsification in the social sciences." *Quality & Quantity* 47(6):3185–3200.

Hsieh, Hsin-I. 1980. "Set theory as a meta-language for natural languages." *International Journal of Human Communication* 13(3):529–542.

Katz, Aaron, Matthias vom Hau and James Mahoney. 2005. "Explaining the Great Reversal in Spanish America: Fuzzy-Set Analysis Versus Regression Analysis." *Sociological Methods & Research* 33(4):539–573.

Kosko, Bart. 1990. "Fuzziness vs. Probability." *International Journal of General Systems* 17(1):211–240.

Kosko, Bart. 1994. "The Probability Monopoly." *IEEE Transactions on Fuzzy Systems* 2(1):32–33.

Lieberson, Stanley. 1991. "Small N's and Big Conclusions: An Examination of the Reasoning in Comparative Studies Based on a Small Number of Cases." *Social Forces* 70(2):307–320.

Mahoney, James. 1980. "Qualitative methodology and comparative politics." *Comparative Political Studies* 40(2):122–144.

Mahoney, James. 2010. "After KKV: The New Methodology of Qualitative Research." *World Politics* 62(1):120–147.

Mahoney, James. 2016. "Mechanisms, Bayesianism, and process tracing." *New Political Economy* 21(5):493–499.

Marx, Axel and Jadir Soares. 2015. "Applying new methodological tools in human rights research. The case of qualitative comparative analysis." *The International Journal of Human Rights* 20(3):365–385.

McCluskey, E. J. 1956. "Minimization of Boolean Functions." *Bell System Technical Journal* 35(6).

Mendel, Jerry M. and Mohammad M. Korjani. in press. "A New Method for Calibrating the Fuzzy Sets Used in fsQCA." *Information Sciences* .

Mill, John Stuart. 1843. *A System of Logic, Ratiocinative and Inductive*. Vol. 1 London: John W. Parker, Harrison & Co.

Munck, Gerardo L. 2016. "Assessing Set-Theoretic Comparative Methods: A Tool for Qualitative Comparativists?" *Comparative Political Studies* 49:1–6.

Novák, Vilém. 1991. Fuzzy set theory and modelling of natural language semantics. In *Interactive Fuzzy Optimization*, ed. Mario Fedrizzi, Janusz Kacprzyk and Marc Roubens. Vol. 368 of *Lecture Notes in Economics and Mathematical Systems* Berlin Heidelberg: Springer Verlag.

Paine, Jack. 2015. "Set-Theoretic Comparative Methods: Less Distinctive Than Claimed." *Comparative Political Studies* 49(6):1–39.

Pennings, Paul. 2003. "Beyond dichotomous explanations: Explaining constitutional control of the executive with fuzzy-sets." *European Journal of Political Research* 42(4):541–567.

Pinter, Charles C. 2014. *A Book of Set Theory*. Mineola, New York: Dover Publications.

Quine, W. V. 1955. "A Way to Simplify Truth Functions." *The American Mathematical Monthly* 62(9):624–631.

Radaelli, Claudio M. and Claudius Wagemann. 2018. "What did I leave out? Omitted variables in regression and qualitative comparative analysis." *European Political Science* pp. 1–16.

Ragin, Charles C. 1987. *The Comparative Method*. Berkeley, Los Angeles, London: University of California Press.

Ragin, Charles C. 2000. *Fuzzy-Set Social Science*. 1 ed. University of Chicago Press.

Ragin, Charles C. 2004. "From fuzzy sets to crisp truth tables." COMPASSS Working Paper.
URL: *http://www.compasss.org/wpseries/Ragin2004.pdf*

Ragin, Charles C. 2008. *Redesigning Social Inquiry. Fuzzy Sets And Beyond*. Chicago and London: University of Chicago Press.

Ragin, Charles C. and Peer Fiss. 2017. *Intersectional Inequality: Race, Class, Test Scores, and Poverty*. Chicago: University of Chicago Press.

Ramzipoor, Roxanna. 2014. "Set Theory and Fuzzy Sets: Their Relationship to Natural Language. An Interview with George Lakoff." *Qualitative & Multi-Method Research* 12(1):9–14.

Rihoux, Benoît and Charles C. Ragin. 2009. *Configurational Comparative Methods - Qualitative Comparative Analysis (QCA) and Related Techniques*. Vol. 51 of *Applied Social Research Methods* SAGE Publications.

Schneider, Carsten and Claudius Wagemann. 2012. *Set-Theoretic Methods for the Social Sciences. A Guide to Qualitative Comparative Analysis*. Cambridge: Cambridge University Press.

Schneider, Carsten Q. and Ingo Rohlfing. 2016. "Case Studies Nested in Fuzzy-set QCA on Sufficiency: Formalizing Case Selection and Causal Inference." *Sociological Methods and Research* 45(3):536–568.

Seawright, Jason. 2005. "Qualitative comparative analysis vis-à-vis regression." *Studies in Comparative International Development* 40(1):3–26.

Singpurwalla, Nozer D. and Jane M. Booker. 2004. "Membership functions and probability measures of fuzzy sets." *Journal of the American Statistical Association* 99(467):867–889.

Skaaning, Svend-Erik. 2007. "Explaining post-communist respect for civil liberty: A multi-methods test." *Journal of Business Research* 60(5):493–500.

Skocpol, Theda. 1979. *States and Social Revolutions: A Comparative Analysis of France, Russia and China*. Cambridge University Press.

Verkuilen, Jay. 2005. "Assigning Membership in a Fuzzy Set Analysis." *Sociological Methods and Research* 33(4):462–496.

Woodside, Arch G. 2014. "Embrace perform model: Complexity theory, contrarian case analysis, and multiple realities." *Journal of Business Research* 67(12):2495–2503.

Zadeh, Lotfi A. 1965. "Fuzzy Sets." *Information and Control* 8:338–353.

Zadeh, Lotfi A. 1983. A fuzzy-set-theoretic approach to the compositionality of meaning: propositions, dispositions and canonical forms. In *Fuzzy Sets, Fuzzy Logic, and Fuzzy Systems: Selected Papers by Lotfi A. Zadeh*, ed. George J. Klir and Bo Yuan. Vol. 6 of *Advances in Fuzzy Systems: Application and Theory* World Scientific Pub Co Inc pp. 594–613.

Zadeh, Lotfi A. 1995. "Probability theory and fuzzy logic are complementary rather than competitive." *Technometrics* 37:271–276.

Łukasiewicz, Jan. 1970. On 3-valued logic. In *Jan Łukasiewicz: Selected Works*, ed. Leo Borkowski. Warsaw: Polish Scientific Publishers.