5

Possibility Theory

The English economist George Lennox Sharman Shackle [256] proposed that uncertainty is related to surprise. For example, when Jane says that an event can happen, then she will not be surprised if this event will actually happen. In general, when we say that an event can "happen", we actually mean that it is a "possible" event. Since there are degrees of surprise, it is absolutely reasonable to talk about degrees of possibility. Thus, in a way, Shackle laid the theoretical basis for a theory of possibility. Indeed, Zadeh [314] used fuzzy sets to define a mathematical *theory of possibility*. Later on, others tried to describe possibilities using different tools that culminated to what is now known as possibility theory.

5.1 Fuzzy Restrictions and Possibility Theory

Although Zadeh was quite aware of the measure-theoretic axiomatization of probability[1] (see Ref. [257] for a discussion of Kolmogorov's work), still he opted to define the notion of *possibility* in terms of *fuzzy restrictions* [309]. In Zadeh's own words, a fuzzy restriction is "a fuzzy relation which acts as an elastic constraint on the values that may be assigned to a variable." Here the term "elastic" does not have any special meaning, and it just means that the constraint is not rigid. Let us first give an informal description of fuzzy restrictions.

Consider the fuzzy proposition *Emma is young* and let Age(Emma) denote the value of variable Emma. The value is a whole number that belongs to the interval [0,100]. Using this interval as a universe, we define a fuzzy set Young for the attribute *young*. For example, it makes sense to have Young(28) = 0.7. The assertion *Emma is young* says something about Emma's age and thus is a restriction on the possible values of Emma's age. However, it says nothing about the group of

1 In fact, Zadeh used measure theory to provide "[a]dditional insight into the distinction between probability and possibility" [314, p. 9].

A Modern Introduction to Fuzzy Mathematics, First Edition.
Apostolos Syropoulos and Theophanes Grammenos.

people to which Emma possible belongs. The following equation

$$R(\text{Age}(\text{Emma})) = \text{Young}$$

denotes the *restriction* on the age of Emma. And here the restriction $R(\text{Age}(\text{Emma}))$ is a fuzzy one, since Young is a fuzzy set.

Assume that F denotes a fuzzy set. Then, the following sentences

Eila is brunette.
Avery is tall.

are instances of the proposition:

$$p \stackrel{\text{def}}{=} X \text{ is } F.$$

In general, the interpretation of "X is F" will be characterized by a *relational assignment equation*. More specifically, we have

Definition 5.1.1 The meaning of the proposition

$$p \stackrel{\text{def}}{=} X \text{ is } F,$$

where X is a name of an object and F is a label of a fuzzy set of a universe of U, is expressed by the relational assignment equation

$$R(A(X)) = F,$$

where A is an attribute which is implied by X and F, and R denotes a fuzzy restriction on $A(X)$ to which the value F is assigned by this equation.

Example 5.1.1 In the sentence "Eila is brunette," the implied attribute is color(Hair), and the relational assignment equation takes the form

$$R(\text{Color}(\text{Hair}(\text{Eila}))) = \text{brunette}.$$

Example 5.1.2 Suppose that the fuzzy set "young" is defined by

$$Y(u) = 1 - S(u; 20, 30, 40),$$

where

$$S(u; \alpha, \beta, \gamma) = \begin{cases} 0, & \text{when } u \le \alpha, \\ 2\left(\dfrac{u-\alpha}{\gamma-\alpha}\right)^2, & \text{when } \alpha \le u \le \beta, \\ 1 - 2\left(\dfrac{u-\alpha}{\gamma-\alpha}\right)^2, & \text{when } \beta \le u \le \gamma, \\ 1, & \text{when } u \ge \gamma. \end{cases}$$

Note that the semicolon separates arguments from parameters, thus, u is the argument and α, β, and γ are the parameters. Assume that Emma is 28 years old. Then, $Y(28) \approx 0.7$. Then, we interpret 0.7 as the degree of *compatibility* of 28 with the concept labeled young. In addition, one may argue that the proposition "Emma is young" transforms the meaning of 0.7 from a degree of compatibility to the degree of possibility that Emma is 28. Thus, the possibility that Emma is young is 0.7. More generally, the compatibility of a value of u with young is transformed into the possibility of that value of u given "Emma is young."

The previous example is used to describe the general notion of possibility.

Definition 5.1.2 Suppose that F is a fuzzy subset of a universe U, where $F(u)$ is assumed to be the compatibility of u with the concept that F corresponds to. Suppose that X is a variable whose values are drawn from U. Also, we view F as a fuzzy restriction, $R(X)$, associated with X. Then, the proposition "X is F," which is represented by

$$R(X) = F$$

associates a *possibility distribution* Π_X with X which we propose to be equal to $R(X)$, that is,

$$\Pi_X = R(X).$$

Correspondingly, the *possibility distribution function associated with* X is written as π_X, and its values are the corresponding membership degrees of F, that is,

$$\pi_X(u) \stackrel{\text{def}}{=} F(u).$$

Therefore, $\pi_X(u)$, the possibility that $X = u$, is proposed to be equal to $F(u)$.

5.2 Possibility and Necessity Measures

In order to introduce possibility measures and necessity measures, we need to know what a *measure* is. All the definitions that are presented in this section are borrowed from [294]. Assume that X is a nonempty set, that $\mathbf{C}$ is a nonempty class of subsets of X, and that $\mu : \mathbf{C} \to [0, \infty]$ is a nonnegative, extended real valued set function defined on $\mathbf{C}$.

Definition 5.2.1 A set E in $\mathbf{C}$ is called the *null set* (with respect to μ) if and only if $\mu(E) = 0$.

Definition 5.2.2 μ is *additive* if and only if

$$\mu(E \cup F) = \mu(E) + \mu(F),$$

where $E, F, E \cup F \in \mathbf{C}$ and $E \cap F = \emptyset$.

Definition 5.2.3 μ is *finitely additive* if and only if

$$\mu\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} \mu(E_i)$$

for any finite, disjoint class $\{E_1, E_2, \dots, E_n\}$ of sets in $\mathbf{C}$ whose union is also in $\mathbf{C}$.

Definition 5.2.4 μ is *countably additive* if and only if

$$\mu\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mu(E_i)$$

for any disjoint class $\{E_1, E_2, \dots, E_n\}$ of sets in $\mathbf{C}$ whose union is also in $\mathbf{C}$.

Definition 5.2.5 μ is *subtractive* if and only if

$$\mu(F - E) = \mu(F) - \mu(E)$$

for $E, F, E \subset F, F \setminus E \in \mathbf{C}$ and $\mu(E) < \infty$.

Theorem 5.2.1 *If μ is additive, then it is subtractive.*

Definition 5.2.6 μ is called a *measure* on $\mathbf{C}$ if and only if it is countably additive, and there exists $E \in \mathbf{C}$ such that $\mu(E) < \infty$.

Definition 5.2.7 Assume that μ is a measure on $\mathbf{C}$. Then, μ is *complete* if and only if $E \in \mathbf{C}$, $F \subset E$, and $\mu(E) = 0$ imply that $F \in \mathbf{C}$.

Definition 5.2.8 Suppose that μ is a measure on $\mathbf{C}$. Then, μ is *monotone* if and only if

(i) $\mu(\emptyset) = 0$ when $\emptyset \in \mathbf{C}$;
(ii) $\mu(E) \le \mu(F)$ when $E, F \in \mathbf{C}$ and $E \subset F$.

Definition 5.2.9 A monotone measure μ is called *maxitive* on $\mathbf{C}$ if and only if

$$\mu\left(\bigcup_{t \in T} E_t\right) = \bigvee_{t \in T} \mu(E_t)$$

for any subclass $\{E_t | t \in T\}$ of $\mathbf{C}$ whose union is in $\mathbf{C}$, where T is an arbitrary index set.

If $\mathbf{C}$ is a finite class, then the previous requirement is replaced by

$$\mu(E_1 \cup E_2) = \max[\mu(E_1), \mu(E_2)],$$

where $E_1, E_2 \in \mathbf{C}$ and $E_1 \cup E_2 \in \mathbf{C}$.

Definition 5.2.10 A monotone measure μ is called a *generalized possibility measure* on $\mathbf{C}$ if and only if it is maxitive on $\mathbf{C}$, and there exists $E \in \mathbf{C}$ such that $\mu(E) < \infty$.

In what follows, the letter π will be used to denote a generalized possibility measure.

Definition 5.2.11 A monotone measure μ on $(X, \mathbf{C})$, where $\mathbf{C} \subseteq \mathscr{P}(X)$, is normalized if $X \in \mathbf{C}$ and $\mu(X) = 1$.

Definition 5.2.12 If a generalized possibility measure π defined on $\mathscr{P}(X)$ is normalized, it is called a possibility measure.

Definition 5.2.13 If π is a possibility measure on $\mathscr{P}(X)$, then the set function ν defined as

$$\nu(E) = 1 - \pi(\overline{E}) \text{ for any } E \in \mathscr{P}(X)$$

is called a necessity measure on $\mathscr{P}(X)$.

5.3 Possibility Theory

A theory that is based on possibility and necessity measures is usually called a possibility theory [111]. Roughly, possibility theory is the reinterpretation of results about possibility and necessity measures in terms of events, and the degree to which they surprise us. Thus, we start with a set of events that are supposed to be subsets of a reference set Ω, which will be called the “event that does not surprise us at all” (practically, an event that will happen). The empty set is identified with the “most surprising event” (practically, even that is impossible). Each event $A \subseteq \Omega$ is associated with a real number $g(A)$. This number is not random, but it is computed and/or estimated by someone who happens to have knowledge of the context in which the event occurs. The number $g(A)$ is a measure of the confidence one has that this particular event will happen. Typically, $g(A)$

increases as confidence increases. In addition, if A is an absolutely possible event, then $g(A) = 1$, but if A is an impossible event, then $g(A) = 0$. In particular,

$$g(\emptyset) = 0 \quad \text{and} \quad g(\Omega) = 1.$$

However, $g(A) = 1$ or $g(A) = 0$ does not necessarily mean that A is an absolutely possible or impossible event. The following axiom is necessary in order to ensure that function g is coherent:

$$A \subseteq B \Rightarrow g(A) \leq g(B)$$

This axiom means that if the event A implies the event B, we should be sure that our confidence that B will happen should be at least as much as our confidence that A will happen. When Ω is an infinite reference set, it is possible to introduce axioms of continuity as follows: For every nested sequence $(A_n)_n$ of sets $A_0 \subseteq A_1 \subseteq \cdots A_n \subseteq \cdots$, or $A_0 \supseteq A_1 \supseteq \cdots A_n \supseteq \cdots$, we have

$$\lim_{n \to +\infty} g(A_n) = g\left(\lim_{n \to +\infty} A_n\right).$$

Given two events $A, B \subseteq \Omega$, for the conjunction and disjunction of events the following inequalities hold:

$$g(A \text{ and } B) = P(A \cap B) \leq \min(g(A), g(B)),$$
$$g(A \text{ or } B) = P(A \cup B) \geq \max(g(A), g(B)).$$

A possibility measure Π is a limiting case of confidence measures for which:

$$\Pi(A \cup B) = \max(\Pi(A), \Pi(B)), \quad \forall A, B.$$

Also, if A and $A^{\complement}$ are two contradictory events ($A^{\complement}$ is the complement of A in Ω), then

$$\max(\Pi(A), \Pi(A^{\complement})) = 1,$$

which means that of two contradictory events, at least one is completely possible. In addition, if an event is considered to be possible, this does not mean that the contrary event is completely impossible.

If the set Ω is finite, then any possibility measure Π can be defined in terms of its values on the singletons of Ω:

$$\Pi(A) = \bigvee \{\pi(\omega) | \omega \in A\}, \quad \forall A,$$

where $\pi(\omega) = \Pi(\{\omega\})$ and $\pi : \Omega \to [0, 1]$ is called *possibility distribution*. Mapping π is *normalized*, meaning that there is an ω' such that

$$\pi(\omega') = 1.$$

Of course, this happens because $\Pi(\Omega) = 1$.

Necessity measures, which are denoted by N, are another form of a limiting case of confidence measures:

$$N(A \cap B) = \min(N(A), N(B)), \quad \forall A, B.$$

In addition, we demand that

$$\Pi(A) = 1 - N(A^{\complement}), \quad \forall A.$$

A direct consequence of this requirement is that

$$N(A) = \bigwedge \{1 - \pi(\omega) | \omega \notin A\}.$$

Necessity measures satisfy the relation

$$\min(N(A), N(A^{\complement})) = 0,$$

which means that two contrary events cannot be necessary at all at the same time. Finally,

$$N(\Omega) = 1 \quad \text{and} \quad N(\varnothing) = 0.$$

Conditional possibilities and necessities are something special, and a comprehensive review of conditional possibilities is presented in [76]. Roughly, these kinds of possibilities and necessities try to answer the question: Now that the event A has happened, what is the chance of event B? In general, conditional possibilities and necessities are mostly presented as a derived notion of the unconditional ones. For instance, given a possibility Π on a Boolean algebra $\mathbf{B}$ (see Section 2.9) and a t-norm $*$, for every $H \in \mathbf{B} \setminus \varnothing$, a t-conditional possibility $\Pi(\cdot|H)$ on E is defined as any solution of the equation

$$\Pi(E \wedge H) = x * \Pi(H).$$

Depending on the choice of the t-norm the pairs $(\Pi(E \wedge H), \Pi(H))$ may not lead to a unique solution. For instance, if the t-norm is the min function, the possible solutions of the previous equation are

$$\begin{aligned} \Pi(A|B) = \Pi(A \wedge B) \quad & \text{if} \ \ \Pi(A \wedge B) < \Pi(B), \\ \Pi(A \wedge B) \leq \Pi(A|B) \leq 1 \quad & \text{if} \ \ \Pi(A \wedge B) = \Pi(B). \end{aligned}$$

In general, we do not demand that an arbitrary solution corresponds to a normalized possibility. However, the following definition avoids this problem:

$$\Pi(A|B) = \begin{cases} \Pi(A \wedge B), & \text{if } \Pi(A \wedge B) < \Pi(B), \\ 1, & \text{if } \Pi(A \wedge B) = \Pi(B). \end{cases}$$

One major problem with this definition is that, for incompatible A and B (i.e. $A \wedge B = 0$) and $\Pi(B) = 0$, we have that $\Pi(A|B) = 1$, while it is natural to expect that $\Pi(A|B) = 0$! Interestingly, if we force $\Pi(A|B) = \Pi(A \wedge B|B) = \Pi(\varnothing|B) = 0$, then $\Pi(\cdot|B)$ is not a possibility. The definition that follows solves this problem [37]:

Definition 5.3.1 Assume that $\mathbf{E} = \mathbf{B} \times \mathbf{H}$ is a set of conditional events $E|H$ such that $\mathbf{B}$ is a Boolean algebra and $\mathbf{H}$ is an additive set (i.e. closed with respect to $\wedge$), with $\mathbf{H} \subset \mathbf{B}$ and $\emptyset \notin \mathbf{H}$. Also, assume that $*$ is a t-norm. Then, a function $\Pi : \mathbf{E} \to [0, 1]$ is a $*$-conditional possibility if it satisfies the following properties:

(i) $\Pi(E|H) = \Pi(E \wedge H|H)$, for all $E \in \mathbf{B}$ and $H \in \mathbf{H}$;
(ii) $\Pi(\cdot|H)$ is a possibility measure, for all $H \in \mathbf{H}$; and
(iii) for all $H, E \wedge H \in \mathbf{H}$, and $E, F \in \mathbf{B}$

$$\Pi(E \wedge F|H) = \Pi(E|H) * \Pi(F|E \wedge H).$$

Recall that a necessity function is the dual function $N : \mathbf{B} \to [0, 1]$ of a possibility Π that satisfies

$$N(E_j \wedge E_j) = \min\{N(E_i), N(E_j)\}.$$

Conditional necessities for any $E|H \in \mathbf{B} \times \mathbf{H}$ have been introduced in [37]:

$$N(E|H) = 1 - \Pi(E^{\complement}|H).$$

This definition induces the following one:

Definition 5.3.2 Assume that $\mathbf{E} = \mathbf{B} \times \mathbf{H}$ is a set of conditional events $E|H$ such that $\mathbf{B}$ is a Boolean algebra and $\mathbf{H}$ is an additive set, with $\mathbf{H} \subset \mathbf{B}$ and $\emptyset \notin \mathbf{H}$. Then, a function $N : \mathbf{E} \to [0, 1]$ is a conditional necessity if it satisfies the following properties:

(i) $N(E|H) = N(E \wedge H|H)$, for all $E \in \mathbf{B}$ and $H \in \mathbf{H}$;
(ii) $N(\cdot|H)$ is a necessity measure, for all $H \in \mathbf{H}$; and
(iii) for all $H, E^{\complement} \wedge H \in \mathbf{H}$, and $E, F \in \mathbf{B}$

$$N(E \vee F|H) = \max\{N(E|H), N(F|E^{\complement} \wedge H)\}.$$

5.4 Possibility Theory and Probability Theory

Roughly speaking, probability is the measure of the likelihood that an event will occur in a random experiment. And this is why people often confuse randomness with probability theory. Typically, a probability is a number that takes values between 0 and 1. Here 0 indicates impossibility and 1 indicates certainty.

Obviously, two events A and B have probabilities p_A and p_B, respectively, and $p_A < p_B$, then A is less likely to occur than B. This brief description explains why possibility theory and probability theory look so similar. Also, this explains why some people superficially assume that these theories are identical. However, there are three different views of the relationship between probability theory and possibility theory. The first view sees the two theories as independent theories that are used to describe different facets of uncertainty. According to the second view, both theories are identical, and one can easily translate possibilities to probabilities and vice versa. And according to the third view, probability theory is just a special case of possibility theory.

The first view has been advocated by Zadeh [317] who, in addition, argued that probability theory is not adequate to deal with uncertainty and imprecision and that the two theories are complementary rather than competitive. In particular, the reasons for the inadequacy of probability theory to deal with uncertainty and imprecision are given below.

(i) Probability theory does not support the concept of a fuzzy event, where examples of such events are a cold day, a strong earthquake, the near future, etc.

(ii) It is not possible to deal with fuzzy quantifiers like many, most, several, and few in probability theory.

(iii) In probability theory, one performs computations with numbers, and it is not possible to perform computations with *fuzzy probabilities* such as likely, unlikely, not very likely, and so forth. Fuzzy probabilities have been introduced by Zadeh [316]:

Assume that A is a fuzzy set defined as follows:

$$A = a_1/u_1 + a_2/u_2 + \cdots + a_n/u_n,$$

where $U = \{u_1, \ldots, u_n\}$ is a finite universe. Also, assume that X is a variable that takes the values $u_1, \ldots, u_n$ with a uniform probability $(1/n)$. If $A \subset U$, then the nonfuzzy probability of the fuzzy proposition or, equivalently, of the fuzzy event

$$p \stackrel{\text{def}}{=} X \text{ is } A$$

is given by

$$P(p) = \frac{\text{card}(A)}{n}$$

and the fuzzy probability is the fuzzy number

$$FP(p) = \frac{\sum_\alpha \alpha/\text{card}({}^\alpha A)}{n}.$$

(iv) It is not possible to give an estimation of fuzzy probabilities in probability theory. This very simply means that there is no answer to questions like "What is the probability that my car may be stolen?"
(v) Probability theory cannot be used as a meaning-representation language. This means that the sentence "It is not likely that there will be a sharp increase in the price of oil in the near future" has no meaning in probability theory.
(vi) The expressive limits of probability theory is a burden to the analysis of problems in which the data are described in fuzzy terms. For example, the following problem borrowed from [317] explains this point.

> A variable X can take the values small, medium, and large with respective probabilities low, high, and low. What is the expected value of X? What is the probability that X is not large?

Of course, it is an undeniable fact that probability theory has been and continues to be used successfully in areas where the systems are mechanistic, while human reasoning, perceptions, and emotions do not play a significant role. For example, statistical mechanics, quantum mechanics, communication systems, evolutionary programming, are all areas where probability theory is used successfully. However, there are areas such as economics, pattern recognition, group decision analysis, speech and handwriting recognition, expert systems, weather and earthquake forecasting, where probability theory is not that successful. The main reason for this is that dependencies between variables are not well defined, the knowledge of probabilities is imprecise and/or incomplete, the systems are not mechanistic, and human reasoning, perceptions, and emotion do play an important role.

According to George Jiří Klir and Behzad Parviz [177]

> probability theory is a natural tool for formalizing uncertainty in situations where class frequencies are known or where evidence is based on outcomes of large series of independent random experiments. Possibility theory, on the other hand, is a natural tool for formalizing uncertainty that results from information that is both imprecise and fuzzy. When information of both kinds is available, it is useful to have the capability of transforming probabilities to possibilities or vice versa.

A very systematic approach to this "problem" was presented by Dubois et al. [113]. In particular, their solution is based on the fact that a possibility measure Π on a set X (i.e. a maxitive measure) is equivalent to the family $\mathscr{P}(\Pi)$ of probability measures (i.e. measures that have the property described in Definition 5.2.4) such that

$$\mathscr{P}(\Pi) = \{P \mid \forall A \subseteq X \text{ and } P(A) \leq \Pi(A)\}.$$

Mathematics is an abstract language that has its own rules and principles. What makes this language important is the use of mathematics to describe the world. Thus, probability and possibility theories are mathematical tools that are used to describe the world. However, just because one can use the mathematical properties of these two theories to transform one to the other, it does not mean that it makes sense to make any of these transformations. The following text by Philip Klöcking, which appeared in Philosophy Stack Exchange, explains the difference between possibility and probability and indirectly explains why we cannot transform possibilities to probabilities.

> Possibility means being able to be thought without contradiction at the same time (!). In the sense of probabilities, it means a state (the total number of states = total number of possibilities in Laplace's sense) that can be thought as an outcome without contradiction.
> Probability of an event then simply means a certain number of states (i) thought as causally invoked by or (ii) conceptually thought within an event divided by the total number of states being able to be thought without contradiction.
> This also means: no probability without possibility, but possibility without probability. Therefore, it is (strictly speaking, see below) perfectly possible to become President of the United States for every American (= no contradiction), although it is thought improbable for nearly all of them.
> The problem of thought arises when situations with a probability near or (mathematically) equal 0 are named impossible, although they can be thought without contradiction. Like saying "My neighbor becoming the president is impossible." It often occurs when we are judging heuristically, which we are doing because we are simply not able to say what the probability of an outcome really is (otherwise we would have holy wisdom). It is an equivocation, a slight move of sense ruining the scientific usability of a language. Becoming rid or at least revealing equivocations is basically the main task of philosophy, Husserl would have said.

In addition, one could say that fuzzy set theory is a mathematical theory that describes our vague world. Of course, that our world is vague does not rule out the use of probabilities in the description of certain phenomena. Nevertheless, probabilities are not that fundamental as are possibilities, since the later are very closely related to vagueness.

In conclusion, the idea that probabilities and possibilities describe different facets of uncertainty is weird since we do not know exactly where are the boundaries that separate these facets. But if we find a way to exactly specify these boundaries, then there is no uncertainty! Of course, the idea that possibility and probability theories are interchangeable is dead wrong. Thus, it only makes sense to assume that possibility theory is the most general theory and probability theory is a special case of possibility theory.

5.5 An Unexpected Application of Possibility Theory

As expected, possibility theory has been applied in many areas (e.g. data analysis, database query, diagnosis, argumentations, see Ref. [112] for more details). However, the replacement of probability theory by possibility theory, as suggested by Michael Smithson [263], is particularly interesting. Of course, the reason is that practitioners of psychology have relied almost exclusively on statistical models and methods known as the Neyman–Pearson–Fisher (NPF) framework for the quantitative analysis of human behavior. Although psychologists still use NPF (see Ref. [241] for a very recent overview of the framework), this framework was originally used in military strategic and decision analysis, industrial quality control, and agricultural experimentation. Thus, the NPF was designed for the solution of problems one encounters in these fields and not the human sciences. However, for some reasons, it was fully adopted by psychologists for the solution of their problems. Smithson used the *general linear model* (GLM) in order to assess whether the exclusive reliance on the NPF obstructs a systematic investigation of key theoretical concepts in psychology.

The GLM splits variation in human behavior into two parts: one that is predicted by the instrumental variable(s), which are necessarily one-to-one, and one that is not predicted because there are unobserved variables or random processes. Thus, psychological predictions become an injective "mapping" from states of values of the independent variables to the output variable. Consequently, human agents operate within a specific range, thus making impossible to have any kind of behavior. And the random processes do not make human behavior… random, as they have to be in agreement with the first part. Therefore, the GLM produces a stochastically deterministic view of human behavior, something that does not agree with everyday observations and with other theories. For example, cognitivism requires concepts such as intentionality, choice, or decision, that have no place in a one-to-one predictive model, where randomness is the only source of uncertainty. On the other hand, there are theories in social psychology that have been deeply affected by GLM. For example, cognitive dissonance theory originally assumed that people are mindful actors whose choices are only partially constrained and not even determined by motivations. Subsequently, the theory assumed that all subjects under the same experimental conditions should respond identically. In conclusion, Smithson notes that

> [u]nfortunately, the GLM cannot tell the clinician whether the treatment was sufficient but not necessary, necessary but not sufficient, or weakly contributing to improvement in the clients. A "statistical significant" difference between the means of treatment and control groups could arise from any of those outcomes.

The NPF and all other "statistical paradigms," as Smithson calls them, including the Bayesian approach, rely on probability theory to handle two kinds of uncertainty. The first one is directly related to the model of behavior and is about the uncertainty one encounters when studying what actors will do. The second kind is about the knowledge of what the actors may do. The first kind of uncertainty is best exemplified by the sentence "I can ride my bicycle to work," but, of course, this says nothing about what I will actually do. And it is a fact that nonprobabilistic types of uncertainty play a central role in many psychological theories.

Probability theory has one more drawback: It cannot satisfactorily capture incomplete knowledge. For example, in the simplest case of a binary outcome setup (A and B), probability theory cannot distinguish uncertainty about whether A will happen from ignorance about whether A will happen. Thus, the statement $P(A) = 0.5$ means that either A and B are equally likely or that we have no idea what is the likelihood of A or B.

In general, one can say that whether we are going to use statistical models for the analysis of quantitative data depends on our interpretive and theoretical basis instead of logical or empirical reasons. For example, Figure 5.1 is a scatter plot and Table 5.1 contains information that has an almost identical pattern. In particular, one may notice that high IEPV or animal fat intake is sufficient but not necessary to produce high APV or age-adjusted death rate, respectively. The real question is whether this says something. According to most medical models, animal fat intake and risk of cancer are causal, while some social scientists may not agree that the approval of violence is causally related to the intention to engage in it. Although it might make sense to most social scientists to conclude that people who approve violence may choose to engage in it or not, most medical researchers would not like to claim that countries in which people do not consume much animal fat may opt to have high death-rates from breast cancer. The data in Table 5.1 and the scatter plot say the same thing, logically and empirically. However, our interpretive or theoretical perspective is responsible for assigning causality or intentionality to one or the other. Therefore, Smithson concluded that statistical models cannot be replaced in all circumstances. Of course, this last statement is not entirely true

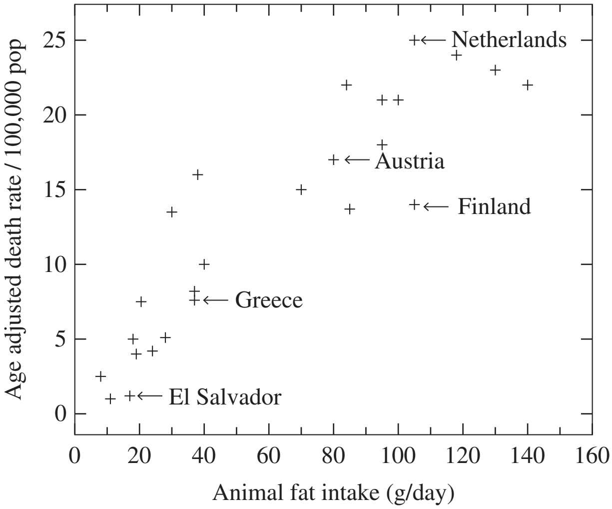

Figure 5.1 A scatter plot. Source: Data from Carroll 1975 [58].

Table 5.1 Intention to engage in political violence (IEPV) and approval of political violence (APV).

IEPV	**APV**					
	0	**1**	**2**	**3**	**4**	**5**
0	97	89	43	9	2	2
1	5	75	45	17	4	2
2	0	5	39	8	11	2
3	0	0	3	13	8	4
4	0	0	1	1	3	4
5	0	0	0	0	2	5

Source: Data from Muller 1972 [226].

today, as fuzzy statistics is gaining approval in the scientific community (see Ref. [78] for an overview of the fuzzy approach to statistical analysis).

Smithson proposed the use of possibility theory for the working psychologist. His approach is based on the “fact” that “possibility is not a kind of probability, nor can probability represent possibilistic uncertainty” [263, p. 11]. The following inequality is very important in the development of his toolkit:

$$\mathrm{ne}_i \leq pr_i \leq \mathrm{po}_i,$$

where ne_i is the necessity of the ith option, pr_i is probability, and po_i is possibility. This inequality can be used to measure the *freedom of action* and the *relative* *preference*. In general, when an action is possible, but not necessary, then one is free to choose whether to perform this action or not. Formally, this is expressed as follows:

$$F_i = \mathrm{po}_i - \mathrm{ne}_i. \tag{5.1}$$

Many surveys measure *preference* by calculating the percentage of people who select a specific option. The problem with this approach is that, for all options, the possibility equals to one and the necessity equals to zero. A better approach is the associated preference with the extent to which people freely select an option, then a valid relative preference measure can be defined as follows:

$$S_i = \frac{\mathrm{pr}_i - \mathrm{ne}_i}{\mathrm{po}_i - \mathrm{ne}_i}.$$

Assume that we have a population and that the portions of people for whom the ith and jth options are possible are po_i and po_j, respectively. Then, the question is: What is the joint possibility $\mathrm{po}(i \text{ and } j)$, that is, the portion of people for whom both options are possible? The range of values for joint possibility is

$$\max(0, \mathrm{po}_i + \mathrm{po}_j - 1) \le \mathrm{po}(i \text{ and } j) \le \min(\mathrm{po}_i, \mathrm{po}_j).$$

When there is no empirical information about $\mathrm{po}(i \text{ and } j)$, there are two definitions one could adopt. According to the first, $\mathrm{po}(i \text{ and } j) = \min(\mathrm{po}_i, \mathrm{po}_j)$. The second approach uses the inequality above as the definition of the joint possibility.

The relevant interval for $\mathrm{po}(i \text{ and } j)$ is

$$\max(\mathrm{po}_i, \mathrm{po}_j) \le \mathrm{po}(i \text{ and } j) \le \min(\mathrm{po}_i + \mathrm{po}_j, 1).$$

If we have adopted that $\mathrm{po}(i \text{ and } j) = \min(\mathrm{po}_i, \mathrm{po}_j)$, then we are forced to deduce that $\mathrm{po}(i \text{ or } j) = \max(\mathrm{po}_i, \mathrm{po}_j)$. Otherwise, the previous inequality becomes the definition. In addition, if we happen to know the value $\mathrm{po}(i \text{ and } j)$, then

$$\mathrm{po}(i \text{ or } j) = \mathrm{po}_i + \mathrm{po}_j) - \mathrm{po}(i \text{ and } j).$$

The possibility that the portion of people for whom option i is possible and have access to option j is expressed by $\mathrm{po}(i/j)$. An alternative interpretation of $\mathrm{po}(i/j)$ is that it is a measure of the degree to which option i is included by option j. This possibility is defined as follows:

$$\mathrm{po}(i/j) = \frac{\mathrm{po}(i \text{ and } j)}{\mathrm{po}_i}.$$

The concepts presented so far make up Smithson's basic calculus of possibility theory.

When we measure uncertainty with some form of possibility theory, then we need to know how relatively free are the agents of a system to make choices. Naturally, this relative freedom should take under consideration all the constraints of a given system. Eq. (5.1) could be used as a starting point for the definition of relative freedom. Smithson proposed that the *relative amount of freedom enjoyed by the ith individual* is

$$F_i = \sum_{j=1}^{r} \frac{\mathrm{po}_{ij} - \mathrm{ne}_{ij}}{r},$$

where r is the number of conceivable options, po_{ij} and ne_{ij} are the possibility and necessity values for the ith individual on the jth option, respectively. Note that the number of individuals is assumed to be N and that po_{ij} and ne_{ij} can assume only two values: 0 and 1. The marginal possibility and necessity distributions over the r options may be recovered by N:

$$\mathrm{po}_j = \sum_{i=1}^{N} \frac{\mathrm{po}_{ij}}{N} \quad \text{and} \quad \mathrm{ne}_j = \sum_{i=1}^{N} \frac{\mathrm{ne}_{ij}}{N}.$$

The following can be used to measure the relative freedom for part of the population that is free to choose the jth option:

$$F_j = \mathrm{po}_j - \mathrm{ne}_j.$$

What if we want to measure the freedom available to the entire population of N individuals? A straightforward solution is to measure the average F_i or F_j, but this solution does not take under consideration both partitions and permutations. If we take under consideration permutations, then the system freedom F^* is defined as follows:

$$\begin{aligned} F^* &= \frac{\prod_i rF_i}{rN} \\ &= 0 \text{ only if all } F_i = 0, \end{aligned} \tag{5.2}$$

where only nonzero F_i are multiplied together. The third approach ignores permutations and considers only partitions that are equipossible. Without going into the various details, when there are two options, then the relative group freedom (FG) can be computed by the following formula:

$$\mathrm{FG} = (1 - \mathrm{ne}_1 - \mathrm{ne}_2)\max(0, 1 - \mathrm{po}_1 - \mathrm{ne}_2) - \max(0, 1 - \mathrm{po}_2 - \mathrm{ne}_1).$$

When we have three options, then

$$\begin{aligned} \mathrm{FG} = {} & (1 - \mathrm{ne}_1 - \mathrm{ne}_2 - \mathrm{ne}_3)^2 - \max(0, 1 - \mathrm{po}_1 - \mathrm{ne}_2 - \mathrm{ne}_3)^2 \\ & - \max(0, 1 - \mathrm{po}_2 - \mathrm{ne}_1 - \mathrm{ne}_3)^2 - \max(0, 1 - \mathrm{po}_3 - \mathrm{ne}_1 - \mathrm{ne}_2)^2 \\ & + \max(0, 1 - \mathrm{po}_1 - \mathrm{po}_2 - \mathrm{ne}_3)^2 + \max(0, 1 - \mathrm{po}_1 - \mathrm{po}_3 - \mathrm{ne}_2)^2 \\ & + \max(0, 1 - \mathrm{po}_2 - \mathrm{po}_3 - \mathrm{ne}_1)^2. \end{aligned}$$

The general formula for the case of r options follows:

$$\begin{aligned}\mathrm{FG} &= \left(1 - \sum_{i=1}^{r} \mathrm{ne}_i\right)^{r-1} - \sum_{i=1}^{r} \max\left(0, 1 - \mathrm{po}_i - \sum_{j\neq i} \mathrm{ne}_j\right)^{r-1} \\ &\quad + \sum_{i=1}^{r} \sum_{j>i} \max\left(0, 1 - \mathrm{po}_i - \mathrm{po}_j - \sum_{k\neq i,j} \mathrm{ne}_k\right)^{r-1} - \cdots \\ &\quad + (-1)^{r-1} \sum_{i=1}^{r} \max\left(0, 1 - \mathrm{ne}_i - \sum_{j\neq i} \mathrm{po}_j\right)^{r-1}.\end{aligned}$$

Let us briefly say what happens when the N individuals are assigned possibilities over a range of values on a continuous variable rather than over a discrete option. Initially, we assume that there is a variable X that can take any value from the closed interval $[d, u]$. Also, we assume that X is the ith individual that is assigned the possibility $\mathrm{po}_i(x) = 1$ for all x in some subrange of $[a_i, b_i]$ and 0 for all y outside this subrange. Then, F_i is defined as

$$F_i = \frac{b_i - a_i}{u - d}.$$

The average F_i and the system freedom F^* are defined as above by replacing r with $u - d$ in (5.2). More generally, we assume that $\mathrm{po}_i(x)$ and $\mathrm{ne}_i(x)$ take values from the closed interval $[0, 1]$. Then, the previous equation is generalized to

$$F_i = \int_d^u \frac{\mathrm{po}_i(x) - \mathrm{ne}_i(x)}{u - d}\, dx. \tag{5.3}$$

If each state x of X is assigned a possibility and/or a necessity value po)(x) and ne(x), respectively, then these values do not refer to individuals. The relative freedom for each individual F_i is identical for all i and is defined by (5.3) without the use of the subscripts in the right-hand terms.

There is one more class of situations where the density functions constrained by possibility and necessity do not have to accumulate to 1. Instead, 1 is considered as a limit on their accumulation value. Typically, we consider M shareholders that may use a resource up to some limit L. Each shareholder's usage of the resource is also limited by her share of the holdings, h_i. Assume that H is the sum of the holdings h_i, w_i is the amount actually used by the ith shareholder, $b_i = w_i/H$, and B is the sum of the b_i. Then,

$$\mathrm{po}(b_i) = \min\left(\frac{h_i}{H}, \frac{L}{H}\right) \quad \text{and} \quad \mathrm{po}(B) = \min(L/H, 1).$$

The value of FG can be calculated by the following formula:

$$\mathrm{FG} = \frac{1}{r!\prod_{i=1}^{r}\mathrm{po}(b_i)}\left[(L/H)^r - \sum_{i=1}^{r}\max(0, L/H - \mathrm{po}(b_i))^r\right.$$
$$+\sum_{i=1}^{r}\sum_{j>i}\max(0, L/H - \mathrm{po}(b_j) - \mathrm{po}(b_j))^r - \cdots$$
$$\left. + (-1)^{r-1}\sum_{i=1}^{r}\max(0, L/H - \sum_{j\neq j}\mathrm{po}(b_j))^r\right].$$

We stop here the presentation of Smithson's work. He discussed also how this calculus can be used in psychology, but we feel the discussion is far too specialized.

Exercises

5.1 Write down the truth tables of the following expressions: $((A \Rightarrow B) \wedge (\neg B))$, $((A \Rightarrow (B \Rightarrow C)) \Leftrightarrow (A \Rightarrow B))$, and $(A \vee (\neg A))$.

5.2 Show that $(A \Rightarrow B)$ is logically equivalent to $(\neg(A \wedge (\neg B)))$.

5.3 Prove in BL the following:
(a) $A \to (B \to (A \,\&\, B))$;
(b) $(A \to B) \to ((A \,\&\, C) \to (B \,\&\, C))$; and
(c) $((A \to B) \wedge (A \to C)) \to (A \to (B \wedge C))$.

5.4 Show that $\mathbf{0}$ is equivalent to $A \odot \mathbf{0}$ in BL.