Poznań Studies in the Philosophy of the Sciences and the Humanities
Vol. 42, pp. 285–297

Uwe Konerding

PROBABILITY AS AN IDEALIZATION OF RELATIVE FREQUENCY: A CASE STUDY BY MEANS OF THE BTL-MODEL

Introduction

It is common scientific experience that deterministic empirical laws are seldom exactly valid. In nearly every non-trivial empirical application there are some data which are not completely in correspondence with the law. Mainly two approaches are applied for handling this problem. The first of these approaches has a long and successful tradition in the natural sciences. It consists of interpreting the empirical law only as an idealization. This means that the law is not expected to hold exactly; instead it is only expected to approximate data well enough to serve the practical purposes for which the respective empirical application is performed. This point of view presupposed, small deviations from predicted data can easily be tolerated. The second approach is mainly applied in the social sciences. It consists of combining the law with assumptions concerning random deviations from ideal data. As a result formulations arise which refer to probabilities.

In philosophy of science and consequently also within the structuralist considerations [Sneed 1971, Balzer, Moulines and Sneed 1987] both strategies have received quite different attention. The whole conceptual framework of structuralism has been developed with regard to deterministically formulated theories. Consequently, it is primarily appropriate for analyzing and reconstructing scientific activities which are examples of the first approach. The second approach however, i.e., the formulation and empirical application of probabilistic models, has hardly ever been treated within the structuralist framework. In order to counter this deficit, the attempt is made within this article to apply the structuralist conception for reconstructing a probabilistic model together with common strategies of its empirical application. For this purpose the special version of the structuralist framework provided by Balzer, Moulines and Sneed [1987] is selected. The probabilistic model which is to be reconstructed is the so-called BTL-model, which is of great importance in psychology [Luce 1959].

Two alternative reconstructions are presented. The first reconstruction is nearer to usual expositions of the model in textbooks of mathematical psychology [Colonius 1984]. It will, however, be argued that this reconstruction provides no adequate conceptual basis for reconstructing the empirical application of the model. In contrast, the second reconstruction is less similar to usual expositions; but on the other hand it is adequate for reconstructing how the model is empirically applied. The core idea of the latter reconstruction consists of interpreting probability as an idealized relative frequency.

I
Informal Description of the BTL-Model

The BTL-model is meant to describe and explain choice-behaviour. It refers to situations in which subjects are confronted with *n*-tuples of objects from which they have to select the most dominant object with respect to an a-priori given attribute. The objects may be marbles, politicians, types of beer or different types of crimes; the attribute may be physical weight, competence, good taste or severeness of the offense. There is practically no limitation concerning the kind of objects or the kind of the attribute which can be considered by means of the BTL-model.

In its most generalized form the model can be applied to *n*-tuples of different size. In this article only the restricted version, which refers to pairs of objects, is considered. In the most common intended application of the model, each pair of objects which can be constituted out of a given sample of objects is presented at least once to the subject. The model can be applied for predicting either individual or group behaviour. In the first case, the model refers to repeated independent choices performed by the same person with respect to the same pair of objects; in the second case, it refers to choices performed by different persons with respect to the same pair of objects. In both cases the choice probabilities are reduced to a latent variable which refers to the presented objects. When individual choice behaviour is considered, this variable can be interpreted as the subjective sensation of the attribute in question; in the case of group behaviour it can be interpreted as the modal sensation of the group.

From this description it follows that there are some essential differences between the BTL-model and its applications on the one hand and a prototypical physical theory-element, as for example the basic element of classical particle mechanics (CPM), on the other. One difference concerns the set of intended applications. In contrast to CPM there is, in the case of the BTL-model, no clearly defined set of intended applications for which the model is generally expected to be empirically valid. Consequently, the whole set of intended applications of the BTL-model is by far more amorphous than the set of intended applications of CPM. A second difference concerns the nature of the terms involved. In the case of CPM the term 'mass' can easily be considered as the same regardless of whether CPM is applied to systems of planets or to atoms. In the case of the BTL-model it may, however, be questionable whether the indivual sensation of weight is the same concept as the modal group sensation of political competence.

In view of these differences it may even be questionable whether the BTL-model and its intended applications can be reconstructed as one single theory-element at all. For the purpose of this article, i.e., a special consideration of those problems which arise when a probabilistic model is reconstructed, the answer to this question is rather irrelevant. For this purpose it may be assumed that a reconstruction of the BTL-model and its intended applications in one single theory-element is feasible.

First Reconstruction

To give an overview of all involved terms, the reconstructions presented here start with the definition of the potential models. In the first possible reconstruction of the BTL-model this is

D1 x is a *potential BTL-model* ($x \in M_p(BTL)$) iff there exist A, p, τ, so that

(1) $x = \langle A, [0, 1] \text{ subset of } \mathbb{R}, \mathbb{R}^+, p, \tau \rangle$;
(2) A is a finite set with at least two elements;
(3) $p: A \times A \rightarrow [0, 1]$ subset of $\mathbb{R}$;
(4) $\tau: A \rightarrow \mathbb{R}^+$.

In this reconstruction A is the sample of objects out of which the presented pairs of objects are constituted; this set is the only non-auxiliary base-set in this definition. The function p is the probability that the first object of the respective pair is selected when it is presented together with the second object of this pair; this probability will also be refered to as probability of preference. The function τ is the subjective sensation of the attribute in question.

In order to discriminate non-theoretical and theoretical terms, a very pragmatic criterion of theoreticity is applied. According to this criterion a term is theoretical with respect to a model iff it is usually estimated from data under presupposition of this model. The use of such a vague term as 'usually' becomes necessary because of the vast set of very different intended applications of the BTL-model. But again, with respect to the purpose of this article the application of such a vague criterion will do no harm. Anyway, according to this criterion the discrimination between non-theoretical and theoretical terms can easily be performed. The values for the sensation are usually estimated from data under presupposition of the model; therefore, this term is theoretical with respect to the model. In contrast, the probability of preference is usually not determined under presupposition of the model; therefore, this term is non-theoretical with respect to the model.

Consequently, the definition for the partial potential models is

D2 x is a *partial potential BTL-model* ($x \in M_{pp}(BTL)$) iff there exist A, p, so that

(1) $x = \langle A, [0, 1] \text{ subset of } \mathbb{R}, p \rangle$;

(2) A is a finite set with at least two elements;

(3) $p: A \times A \rightarrow [0, 1]$ subset of $\mathbb{R}$;

The definition of the actual models is

D3 x is an *actual BTL-model* ($x \in M(BTL)$) iff

(1) $x \in M_p(BTL)$;

(2) $p(a,b) = \dfrac{\tau(a)}{\tau(a) + \tau(b)}$ for all $a,b \in A$.

Here the second axiom describes the already mentioned relationship between the probability of preference and the subjective sensations of the two objects which are presented together.

Up to this point the reconstruction may seem easy. Difficulties arise, however, when the conceptual apparatus developed thus far is applied for reconstructing how the model is empirically applied. The reason for these difficulties is the concept of probability. Social scientists usually interpret probability as the limiting value of relative frequency when sample size tends to infinity. Data sets, however, are always finite. This means that the term 'probability of preference' is not suitable for describing data and, furthermore, that the BTL-model in the reconstruction just presented contains no function at all, which is suitable for describing data. As a consequence the empirical claim cannot be reconstructed. Taken at a whole this means that the reconstruction just presented turns out to be a conceptual dead end.

Second Reconstruction

In the second possible reconstruction the definition of the potential models is

D4 x is a *potential BTL-model* ($x \in M_p(BTL)$) iff there exist A, n, k, τ, so that

(1) $x = \langle A, \mathbb{N}_0, \mathbb{R}^+, n, k, \tau \rangle$;

(2) A is a finite set with at least two elements;

(3) $n: A \times A \rightarrow \mathbb{N}_0$;

(4) $k: A \times A \rightarrow \mathbb{N}_0$;

(5) $\tau: A \rightarrow \mathbb{R}^+$.

As far as the non-auxiliary base-sets are concerned, the second reconstruction is the same as the first. As in the first reconstruction, there is only one set of this kind, namely the sample of objects which is again denoted by A. Both reconstructions differ, however, with respect to their functional terms. The term 'probability of preference' is omitted in the second reconstruction. Instead two new terms are introduced. The first new term, the function n, is the number of times the respective pair of objects has been presented; the second new term, the function k, is the number of times the first object has been selected. The last term, the function τ, is, exactly as in the first reconstruction, the sensation of the attribute in question.

According to the theoreticity-criterion which has been proposed in the preceding part, the terms 'number of presentations' and 'number of times the first object has been selected' can easily be identified as non-theoretical with respect to the model. The term sensation again is theoretical. In contrast to the first reconstruction, all non-theoretical terms in the second reconstruction can be directly applied for describing data.

The definition for the partial potential models is

D5 x is a *partial potential BTL-model* ($x \in M_{pp}(BTL)$) iff there exist A, n, k, so that

(1) $x = \langle A, \mathbb{N}_0, n, k \rangle$;

(2) A is a finite set with at least two elements;

(3) $n: A \times A \rightarrow \mathbb{N}_0$;

(4) $k: A \times A \rightarrow \mathbb{N}_0$;

The definition of the actual models is

D6 x is an *actual BTL-model* ($x \in M(BTL)$) iff
(1) $x \in M_p(BTL)$;
(2) for all $a,b \in A$: if $n(a,b) > 0$ then
$$\frac{k(a,b)}{n(a,b)} =_{\text{IDEALIZED}} \frac{\tau(a)}{[\tau(a) + \tau(b)]}.$$

This last definition is quite analogous to the corresponding definition in the first reconstruction. Again the second axiom describes the 'fundamental law' of the model. In contrast to the first reconstruction, however, there is no relationship stated between sensation and probability but instead a relationship between sensation and relative frequency. This relationship can, of course, only be understood as an idealization. Especially for small data sets it will usually be impossible to find theoretical terms so that this relationship holds exactly. The general idea of probability as the limiting case of relative frequency can now be taken as a kind of metaprinciple which determines how this idealization has to be understood. In other words, if the right hand side of the equation is interpreted as probability, then the derivation of strategies for empirically applying the model can be based upon all that is known from mathematical statistics about the relationship between probability and relative frequencies.

II
Reconstruction of the Empirical Application Procedure

The next step is to elaborate how the actually performed procedure of empirically applying the BTL-model can be reconstructed within the structuralist framework. This reconstruction will be performed in two stages: firstly, the conception which will serve as a reconstructional frame will be discussed; secondly, the procedures which are actually applied when the BTL-model is tested empirically will be subsumed under this frame.

Reconstructional Frame

The reconstructional frame will be produced partly by selecting relevant concepts from the structuralist conception and partly by adequately modifying these selected concepts. The part of the structuralist considerations which is most promising in terms of providing relevant concepts is the part concerned with empirical claims. Here, the attempt is made to find a general characterization of the claim which can be made by means of the mathematical structure of a theory element. Most of these considerations refer to exact empirical claims made by means of an idealized theory element. Balzer, Moulines and Sneed, however, also discuss so-called approximative claims, i.e. claims in which a certain degree of inaccuracy is incorporated. They consider characterizations of such claims as more realistic representations of the manner in which idealized theory elements are actually applied. Therefore, this part of their considerations will be discussed here more thoroughly.

For the exact characterization of approximative empirical claims Balzer, Moulines and Sneed define some special concepts. The most elementary of these concepts is that of a blur, which is simply a set of pairs of models. By means of this concept two more complicated concepts are defined. The first, the uniformity on a set of potential models, is a set of blurs constituted out of the respective potential models so that every blur includes all possible pairs of identical potential models. The second, the class of admissable blurs in a uniformity on a set of potential models, is a special kind of subset of the given uniformity.

The respective subset of the uniformity has to meet four requirements for being a class of admissable blurs: firstly, it has to be non-empty; secondly, membership in this class has to be invariant when the order of elements within the pairs of the blurs is changed; i.e. if a certain blur belongs to the class, then the blur in which the order of models within the pairs is reversed must also belong to the class; thirdly, blurs with equal corresponding blurs on the non-theoretical level must not be separated; i.e. if two blurs of potential models are transformed into the same blur of partial potential models by removing the theoretical functions, then both blurs of potential models have to be either both inside or both outside the class of admissable blurs; fourthly, the class of admissable blurs has to be bounded; i.e. there has to be a set of 'greatest' blurs, so that each admissable blur is a subset of one of these 'greatest' blurs [cf. Balzer, Moulines and Sneed 1987, p. 348]. In structuralist language this set of greatest blurs is refered to as the bound of the class of admissable blurs.

Balzer, Moulines and Sneed apply the concept of a class of admissable blurs to represent those pairs of potential models which can be accepted as approximations of each other. All their various characterizations of approximative empirical claims rely on this concept. In doing this, they adopt a notion of approximation which may be unnecessarily restrictive. To be specific, it is the second requirement which is criticized here. Application of this requirement implies that a model x can only be approximated by a model x' if this relationship also holds the other way round. Balzer, Moulines and Sneed formulate this demand without further extensive investigations of the empirical testing procedures which are actually performed. It will, however, be shown further below that in the case of the BTL-model the actually performed procedures do not meet this requirement. Therefore, the original concept proposed by Balzer, Moulines and Sneed will be modified here by removing the second requirement. In the resulting concept a different epistemological function is attributed to both models of the pair. That is to say, the first model is considered as the approximated and the second as the approximating model.

Balzer, Moulines and Sneed discuss three possible characterizations of an approximative empirical claim. All these three versions result from combining the idea of approximation with the exact empirical claim made by means of an idealized theory element. In the first version the approximation pertains to the empirical content of the theory element, in the second version to the data, and in the third version to both sides. Balzer, Moulines and Sneed [1987] show that the first version implies the second and that the second implies the third. From this it follows that the third version is the empirically least restrictive one. Balzer, Moulines and Sneed argue that this empirically least restrictive version is the most appropriate for reconstructing the empirical testing procedures which are actually performed in empirical sciences. In the case of the BTL-model even the second version will do.

The original structuralist considerations are concerned with empirical claims which refer to all intended application of a theory element. For the purpose of this article only the application to one single intended application is of interest. Therefore, the just selected version of an empirical claim will be correspondingly simplified in order to construct an adequate frame for reconstruction.

Now let i be a single intended application, A a class of admissable blurs for potential models (in the liberalized sense proposed above) and $B(A)$ the corresponding class of blurs for the partial potential models. Let $x \approx_{B(A)} x'$ denote that there is a $v \in B(A)$ so that $<x,x'> \in v$; this relation will be refered to as the empirical approximation relation. Furthermore let $cn_{nt}(T)$ denote the set of all partial potential models of a theory element which can be extended to a potential model so that an actual model results; this set will be refered to as the isolated empirical content. If now the second version of an approximative empirical claim is simplified so that it refers only to single intended applications the resulting formulation is[1]:

$$\text{There is an } x \in M_{pp}(T) \text{ so that } i \approx_{B(A)} x \text{ and } x \in cn_{nt}(T).$$

Actual Reconstruction

With this conception as an interpretational pattern two aspects have to be discussed. Firstly, it must be determined whether the criteria which are usually applied to decide whether the BTL-model empirically holds for a given set of data can also be applied to define a class of admissable blurs. Secondly, it must be determined whether the actual empirical application procedure can be interpreted as an item of the version presented here of an approximative empirical claim.

The procedure which is actually performed for testing the BTL-model consists of two separate steps: firstly, those sensation values are estimated which fit best to the data under presupposition of the model; secondly, a check is made as to whether the probabilities predicted by the model and the estimated sensation values are sufficiently consistent with the data. In order to avoid unneccessary complications, only the case will be discussed where the respective set of data is large[2]. In this case parameters are usually estimated by means of a maximum-likelihood-procedure [Van Putten, 1982] and the final test is performed by means of the Pearson-test-function [cf. Bosch, 1976, Chap. 6]. This test-function and its application is treated here as the best candidate for a conceptual basis for defining the required class of admissable blurs.

Suppose now that the set of objects is ordered according to any arbitrarily given order. Furthermore let p_{ij} be the prefence-probability which refers to choice pair (a_i,a_j) and which is predicted by the model and the corre-

---

1 For readers who want to compare the simplified version with the original more general version of Balzer, Moulines and Sneed [1987, p. 355, (ß)] it should be noted that there seems to be an erratum in the original formulation. As far as can be judged by a reader it should be $X \in$ Cn(K) instead of **X**$\in$**Cn(K)**.

2 In this context a large set of data can be characterized by at least four different choice-objects and by about twenty independent data for each pair of choice objects. For the exact requirements which data have to meet so that the discussed procedures can be applied see van Putten [1982] in the case of parameter estimation and Bosch [1976] in the case of the statistical test.

sponding estimated parameters. The Pearson-test-function applied to the BTL-model is then

$$f = \sum_{i=1}^{|A|} \sum_{\substack{j=1 \\ i \neq j}}^{|A|} \frac{[k(a_i, a_j) - n(a_i, a_j) \times p_{ij}]^2}{[n(a_i, a_j) \times p_{ij}]} .$$

Under the given presuppositions and under validity of the model this function is approximately chi-square-distributed with $|A| \times (|A|-3)$ degrees of freedom[3]. As a prerequisite for the empirical decision the acceptable probability for falsely rejecting the model is chosen and a test-function-value is determined so that the corresponding distribution function value is equal to one minus the given probability. If the empirically determined function value exceeds this critical value, the model is rejected, otherwise it is retained.

In the usual statistical jargon the p_{ij} are often refered to as theoretical and the $k(a_i,a_j)$ and $n(a_i,a_j)$ as empirical values. Hence it may be tempting to interpret the Pearson-test-function as a measure of similarity between non-theoretical and theoretical expressions which are both constructed out of function values of the same potential model. With respect to reconstructing the empirical testing procedure by means of the concept of a class of admissable blurs this interpretation, however, leads in the wrong direction. To be specific the concept of admissable blurs is concerned with similarities between two different models and not with similarities between two expressions constructed from different terms out of the same model. Therefore, it is more promising to conceive of the k- and n-values as values of one model and the p-values as values of a different model.

This interpretation, however, is still confronted with a fundamental difficulty, namely, that the third requirement in the original definition of a class of admissable blurs of potential models is violated. According to this requirement two blurs of potential models which are transformed into the same blur of partial potential models by removing the theoretical functions have to be either both inside or both outside the class of admissable blurs. If, however, – as it has just been discussed – similarity measures defined by means of theoretical functions are applied for determining the extension of a class of admissable blurs, this reqirement will always be violated. Therefore, only such similarity functions should be applied which are defined by means of non-theoretical terms. The Pearson-test-function can be transformed into such a non-theoretical similarity measure, when the area of possible arguments of the function is restricted to pairs of models with the same number of choice-objects and identical n-values and when then the expression $n(a_i,a_j) \times p_{ij}$ is interpreted as an estimation of the $k(a_i,a_j)$ in a different model for which the BTL-model holds exactly.

Now let α be a given accepted probability for falsely rejecting the model and let $f(\alpha, |A|)$ be the critical Pearson-test-function-value which has been determined by the procedure described above. An admissable blur of potential models can then be defined as

$$u_\alpha = \{<x,x'> \mid x,x' \in M_p(BTL) \text{ and } |A| = |A'|$$

and the objects in both models can be ordered so that for all $a_i, a_j \in A$ and for all $a_i', a_j' \in A$: $n(a_i,a_j) = n(a_i',a_j')$ and

$$\sum_{i=1}^{|A|} \sum_{\substack{j=1 \\ i \neq j}}^{|A|} \frac{[k(a_i, a_j) - k(a_i', a_j')]^2}{k(a_i', a_j')} < f(\alpha, |A|) \}.$$

The definition for the corresponding blur for the partial potential models can be easily generated by replacing $x,x' \in M_p(BTL)$ with $x,x' \in M_{pp}(BTL)$.

As already hinted above application of the Pearson-test-function for defining a class of blurs results in a mathematical entity which does not meet the second requirement of the original definition of a class of admissable blurs. Thus if the bound of the class of admissable blurs is defined by means of a fixed α then pairs of models can be constructed so that the calculated test-function will exceed the critical value for one order of the models and so that it will not exceed this value for the revised order. If the less restrictive conception which has been proposed here is applied this aspect will cause no problems for the further reconstruction.

The process of parameter estimation for the BTL-model can now be interpreted as determining the theoretical values of that actual BTL-model which is – with respect to a theoretical approximation relation defined by means of u_α (see above) – most similar to the potential models which can be constructed by extending data with theoretical function values. Under presupposition of this interpretation the n-values of the most similar actual model are also uniquely defined; according to the approximation relation

3 The calculation of the degrees of freedom is based upon the assumption that $k(a_i,a_j)$ and $k(a_j,a_i)$ result from the same sample of presentations; i.e. that

$$k(a_i,a_j) = n(a_i,a_j) - k(a_j,a_i) = n(a_j,a_i) - k(a_j,a_i).$$

For the principles for calculating degrees of freedom for the Pearson-test-function see Bosch [1976] or Fisz [1971].

they have to be equal to the n-values of data. Consequently, the k-values of this most similar actual model are likewise uniquely defined. They can be calculated by means of the n- and p-values and by means of the fundamental law. In total this means that the process of parameter estimation can also be interpreted as indirectly determining that element of the isolated empirical content of the BTL-model which is – with respect to the corresponding empirical approximation relation – most similar to the data. The process of empirical testing can then be conceived as checking whether this similarity is strong enough for the empirical approximation relation to hold.

Concluding Remarks

Within this article one possible structuralist reconstruction of a special probabilistic model has been presented. The core idea of this reconstruction consists of conceiving probability as idealized relative frequency. Consequently, investigations have been made into whether the strategies which are actually applied for testing the empirical validity of the model can be reconstructed by means of the structuralist concept of an approximative realistic empirical claim. With the exception of one aspect this concept proved to be adequate. A possible improvement of this concept with respect to this one problematic aspect has been proposed.

The reconstruction presented here only pertains to one selected model. The applied principle of reconstructing, however, could also be applied to the reconstruction of different probabilistic models. This may be performed quite straightforwardly for all other choice models discussed in psychology [cf. Colonius 1984]. For different probabilistic models further, more fundamental considerations might be required.

Institut für Psychologie
RWTH Aachen
Jägerstrasse 17/19
52066 Aachen, Germany

ACKNOWLEDGEMENTS

I would like to thank Mrs. Ute Schmid for a critical review of an earlier version of the manuscript and Mr. Peter Bereza for revising my English.

REFERENCES

Balzer, W., Moulines, C.U. and Sneed, J.D. [1987]. *An Architectonic for Science*. Dordrecht: Reidel.

Bosch, K. [1976]. *Angewandte mathematische Statistik*. Wiesbaden: Vieweg.

Colonius, H. [1984]. Stochastische Theorien individuellen Wahlverhaltens. In: Albert, D., Pawlik, K., Stapf, K.-H. and Stroebe, W. (Eds.). *Lehr- und Forschungstexte Psychologie*. Berlin, Heidelberg: Springer.

Fisz, M. [1971]. *Wahrscheinichkeitsrechnung und mathematische Statistik*. Berlin: Deutscher Verlag der Wissenschaften.

Luce, R.D. [1959]. *Individual Choice Behavior*. New York: Wiley.

Sneed, J.D. [1971]. *The Logical Structure of Mathematical Physics*. Dordrecht.

Van Putten, W.L.J. [1982]. Maximum Likelihood Estimation for Luce's Choice Model. *Journal of Mathematical Psychology*, 25, 163–174.