ANALOGICAL PREDICTIVE PROBABILITIES

SIMON M. HUTTEGGER

Abstract. How should analogical considerations enter inductive reasoning? This question was raised after Carnap developed his early systems of inductive logic and various inductive rules have been introduced since. Most of these proposals do not have an axiomatic foundation along the lines of W. E. Johnson’s and Carnap’s work. Thus, it is at least to some extent unclear to which inductive problems they are supposed to apply. By taking clues from de Finetti’s ideas about analogy, I present a new analogical inductive logic that is based on a rigorous foundation. The axioms of the new theory extend the axioms of Johnson and Carnap in fairly minimal ways, and they allow us to discuss at a precise level the merits and limitations of the resulting system of inductive inference.

1. Introduction

Inductive reasoning by simple enumeration is an integral part of Bayesian inference. In particular, a generalized form of Laplace’s rule of succession can be derived from de Finetti’s theorem for exchangeable sequences of observations (see e.g. Zabell, 1989). There is an alternative approach that also leads to generalized rules of succession by imposing constraints on conditional probabilities. This approach was pioneered by the Cambridge logician and philosopher W. E. Johnson (Johnson, 1924, 1932) and independently developed by Rudolf Carnap in his monumental endeavor to develop a general inductive system for the logic of science (Carnap, 1950, 1952, 1971, 1980). The significant advantage of the Johnson-Carnap approach over the first one is that it is axiomatic: it provides a set of axioms for sequences of observations from which an inductive rule can be derived without appealing to priors over model parameters. This makes the Johnson-Carnap approach very appealing from a philosophical point of view since the plausibility of the inductive system can be judged quite straightforwardly from the axioms.

As emphasized by de Finetti, inductive inferences are often based on analogy (de Finetti, 1938, 1959). He and his successors studied various notions of partial exchangeability that correspond to different forms of analogy. Carnap also clearly recognized the importance of analogy for inductive inference, but his own contributions did not lead to any definite solution of the analogy problem (Carnap and Stegmüller, 1959; Carnap, 1980). The chief obstacle here is that a central assumption of the Johnson-Carnap approach—the so-called ‘sufficientness postulate’ (Good, 1965)—implies that the predictive probability of an outcome does not depend on the number of times similar outcomes are observed.

Date: July 2014.

1See Zabell (2011) for an in-depth discussion of this issue.
2See also Diaconis and Freedman (1980, 1984) and Diaconis (1988).
My main goal is to develop an analogical inductive logic that both lifts the restriction of the sufficientness postulate and makes use of the idea of partial exchangeability. After briefly reviewing exchangeability, the relevant notions of partial exchangeability will be introduced in §2. In §3 our main result is stated and proved, along with some straightforward consequences. We compare our new system of inductive logic to some alternative ones in §4, discuss its merits and its limits and the type of analogical reasoning it captures.

2. Exchangeability and Partial Exchangeability

2.1. Exchangeability. The notion of an exchangeable sequence is fundamental for Johnson’s, de Finetti’s and Carnap’s approaches to inductive inference. Observations are exchangeable if their probability does not depend on the order in which they are made. Mathematically, a finite number of observations are reported in terms of a finite sequence of random variables $X_1, \ldots, X_n$ taking values in the finite set of outcomes $\{1, \ldots, s\}$. The sequence $X_1, \ldots, X_n$ is exchangeable if

$$P[X_1 = x_1, \ldots, X_n = x_n] = P[X_1 = x_{\sigma(1)}, \ldots, X_n = x_{\sigma(n)}]$$

for every permutation $\sigma$ of $\{1, \ldots, n\}$. An infinite sequence $X_1, X_2, \ldots$ is exchangeable if every finite initial sequence is exchangeable.

de Finetti proved that, whenever $X_1, X_2, \ldots$ is an infinite exchangeable sequence, the probability measure $P$ is represented by a unique mixture of i.i.d. multinomial trials (de Finetti, 1937). This means that one can think of observations as generated by fixed probabilities $p_i$ for each outcome $i$ such that one is uncertain concerning the values $p_i$. The uncertainty is expressed by a probability distribution—the mixing measure—over the possible values of $p_i$.

One important consequence of this result is a simple formula for calculating predictive probabilities. If the mixing measure in the de Finetti representation is a Dirichlet distribution, then the conditional probability of observing outcome $i$ given past observations of outcomes is

$$P[X_{n+1} = i|X_1, \ldots, X_n] = \frac{n_i + \alpha_i}{n + \sum \alpha_i},$$

where $n_i$ is the number of times $i$ occurs among $X_1, \ldots, X_n$ and the $\alpha_i$ are the parameters of the Dirichlet distribution. The formula (1) is a generalization of Laplace’s rule of succession (Laplace, 1774), which arises as a special case where the mixing measure is the uniform distribution—i.e. all values $p_i$ are equally probable. In general, Dirichlet distributions can express many kinds of uncertainty that manifest themselves in the values of $\alpha_i$.

Johnson (1924, 1932) and Carnap (1950, 1952, 1971, 1980) develop an approach to arrive at the predictive probabilities (1) that does not appeal to de Finetti’s theorem. Besides exchangeability, their central assumption is Johnson’s sufficientness postulate. Johnson’s sufficientness postulate says that the predictive probability for outcome $i$ only depends on $i$, $n_i$, and the total number of observations $n$:

$$P[X_{n+1} = i|X_1, \ldots, X_n] = f_i(n_i, n)$$

There are many extensions of this result to more general probability spaces—see Aldous (1985) for an overview.
Together with a regularity condition, the predictive probabilities (1) can be derived from exchangeability and Johnson’s sufficientness postulate.\footnote{See Zabell \citeyear{Zabell1982} and Kuipers \citeyear{Kuipers1978} for detailed accounts of the mathematical aspects of, respectively, Johnson’s and Carnap’s arguments.}

Does the Johnson-Carnap approach offer any advantages? First, it seems to be more appropriate for a subjective Bayesian point of view such as de Finetti’s. According to this view, the $p_i$ in the de Finetti representation are just artifacts of exchangeable probabilities. So any appeal to them, as in requiring the mixing measure to be Dirichlet, is philosophically problematic.\footnote{As argued in \cite{deFinetti1938}.}

Furthermore, except in very simple cases, it might be hard for you to quantify your mixing measure in the de Finetti representation, or to determine whether it is a Dirichlet distribution to begin with. Johnson’s sufficientness postulate, on the other hand, only refers to the sequence of observations, just like exchangeability. It is therefore quite straightforward to judge whether or not these two assumptions hold in a particular situation. In case they do, the parameters $\alpha_i$ can be readily estimated from one’s probabilities of outcomes before any observations are made.

Both exchangeability and Johnson’s sufficientness postulate can, in a certain sense, be thought of as judgements of analogy influences. In the case of exchangeability this was observed by de Finetti \citeyear{deFinetti1938}. For de Finetti, each observation or each trial of an experiment is, strictly speaking, unique since circumstances may vary in an enormous number of aspects—the same kind of observation can be, and usually is, made under completely different circumstances \cite{deFinetti1974}. Saying that a sequence of outcomes is exchangeable is then tantamount to asserting that different circumstances are irrelevant. In other words, trials are judged to be completely analogous in all relevant respects.

Johnson’s sufficientness postulate expresses a somewhat different kind of analogy judgement. Saying that the predictive probability of outcome $i$ only depends on $i$, $n_i$ and the total number of trials implies that the number of trials $n_k$ of outcomes $k$ similar to $i$ is judged irrelevant. Thus, there can be no analogy influence on predictive probabilities stemming from analogical relations between outcomes.

As Achinstein \citeyear{Achinstein1963} observed, this poses a problem for Carnapian inductive logic. In Achinstein’s main example different metals are observed as to whether they conduct heat. Achinstein argues that there should be an analogy influence between different types of metal. To be more specific, the number of cases where platinum or osmium was observed to conduct heat should have an influence on the predictive probability that a piece of rhodium conducts heat. The analogy here stems from the fact that platinum and osmium are in the same chemical family as rhodium. However, Johnson’s sufficientness postulate excludes such an analogy influence, and so it is clear that Carnap’s inductive logic cannot successfully treat these cases of analogical inference.

2.2. \textbf{Partial exchangeability.} In order to deal more systematically with the first type of analogy judgment—the influence of different circumstances on trials—de Finetti introduced the concept of \textit{partial exchangeability} \cite{deFinetti1938, deFinetti1959}. While exchangeability is complete analogy, partial exchangeability is located in the middle ground between complete analogy and no analogy.

de Finetti’s basic scheme of partial exchangeability involves \textit{outcomes} and \textit{types of outcomes}. Suppose there are $t < \infty$ types. Then the observation of $N$ outcomes
is given by an array with \( X_{1j}, \ldots, X_{N_j,j} \) as the \( j \)th column; here, \( X_{nj} \) is the \( n \)th outcome of type \( j \), \( N_j \) is the number of type \( j \) outcomes, and \( \sum_j N_j = N \). Thus, each outcome is indexed by its type.

As an example, consider flipping two coins. The outcomes register whether a coin flip yields heads or tails. In addition, it is noted whether an outcome is obtained with the first or the second coin—these are the types. Achinstein’s example can also be regarded as an instance of this scheme, where types are different kinds of metal that can assume two outcomes (conduct heat, doesn’t conduct heat).

The array \( X_{1j}, \ldots, X_{N_j,j}, 1 \leq j \leq t \) is partially exchangeable if it is exchangeable within each type. More precisely, let \( n_{ij} \) be the number of times outcome \( i \) of type \( j \) was observed in the first \( N \) trials. If the array is partially exchangeable, then the counts \( n_{ij} \) are a sufficient statistics for the prior—knowledge of \( n_{ij} \) is sufficient to determine the probability distribution over outcomes.

Exchangeability is a special case of partial exchangeability. If outcomes can always be permuted across types without changing the probability assignment, then the array can be thought of as an exchangeable sequence of outcomes. This is a precise formulation of the analogy judgement underlying exchangeability—that the circumstances of an observation can be ignored.

There is a representation theorem for partially exchangeable trials that generalizes de Finetti’s theorem for exchangeable trials. It states that the probability of an infinite partially exchangeable sequence is a mixture of independent but not necessarily identically distributed multinomial trials \cite{Diaconis1980, Link1980}. In de Finetti’s coin example, the representation theorem allows one to think of partially exchangeable coin flips by first choosing the bias of the first coin and the bias of the second coin according to a joint distribution over biases, and then make independent coin flips with these two coins. The resulting sequence will be partially exchangeable.

If the joint distribution makes types identically distributed, then the sequence of outcomes is exchangeable and outcomes are completely analogous. If the joint distribution is a product distribution for types, then types are independent and there are no analogy influences between them. In general, the types may be dependent. The posterior becomes peaked around the observed relative frequencies as data accumulate. The predictive probabilities for outcomes conditional on past observations coincide with the Bayes estimates (the posterior expectation of chances). Hence, predictive probabilities converge to the limiting relative frequencies almost surely.

Analogy influences between types are thus transient. The similarity between types determines the initial dependence between types but vanishes with increasing information. For instance, the similarity between two coins does have no effect on

\[ P\left[X_{11} = x_{11}, \ldots, X_{N_{1},1} = x_{N_{1},1}, \ldots; X_{1t} = x_{1t}, \ldots, X_{N_{t},t} = x_{N_{t},t}\right] = \int_{\Delta^t} \prod_{j=1}^{t} p_{1j}^{n_{1j}} \cdots p_{sj}^{n_{sj}} \, d\mu(p_1, \ldots, p_t). \]

The integral ranges over the \( t \)-fold product of the \( s-1 \)-dimensional unit simplex \( \Delta \), and \( \mu \) is the mixing measure on the probability vectors \( p_j = (p_{1j}, \ldots, p_{sj}) \in \Delta, 1 \leq j \leq t \).
predictive probabilities in the long run. This rather weak kind of analogy influence results from the conditional independence (given knowledge of chances) of types.

2.3. Generalizing partial exchangeability. A generalization of the second type of analogy judgment—given by Johnson’s sufficientness postulate—allows counts of similar outcomes to have an effect on the predictive probability of another outcome. In the context of partial exchangeability, this means that the predictive probability of a type \( j \) outcome \( i \) does not just depend on \( n_{ij} \) but also on \( n_{im} \) for \( m \neq j \). More precisely,

\[
P[X_{N+1} = i|X_1, \ldots, X_N, Y_{N+1} = j] = f_{ij}(n_{i1}, \ldots, n_{it}, N_1, \ldots, N_t)
\]

Here, \( n_{im} \) is the number of times outcome \( i \) of type \( m \) is recorded in the first \( N \) trials, and \( N_m = \sum n_{im} \) is the number of times type \( m \) is observed\footnote{In order to simplify notation, we don’t state \( Y_1, \ldots, Y_N \) explicitly but assume that a particular sequence of types is given.}. Thus the probability of observing an outcome \( i \) of type \( j \) only depends on \( i, j, n_{im}, N_m, m = 1, \ldots, t \).

Condition \( \text{(4)} \) is a generalization of Johnson’s sufficientness postulate that provides a precise formulation of the idea that analogy influences between types may affect predictive probabilities of outcomes. As examples, consider again tossing two coins or Achinstein’s example.

This idea may be in conflict with partial exchangeability. To see why we consider the random variables \( Y_n \) more explicitly. We say in the present context that \( X_1, \ldots, X_N \) is partially exchangeable with respect to \( Y_1, \ldots, Y_N \) if the probability of \( X_1 = x_1, \ldots, X_{N+1} = x_{N+1}, Y_1 = y_1, \ldots, Y_N = y_N \) only depends on the counts \( n_{ij} \). Now, partial exchangeability implies that

\[
P[X_{N+1} = i, X_{N+2} = l, X_{N+3} = k|X_1, \ldots, X_N, Y_{N+1} = j, Y_{N+2} = m, Y_{N+3} = j]
= P[X_{N+1} = k, X_{N+2} = l, X_{N+3} = i|X_1, \ldots, X_N, Y_{N+1} = j, Y_{N+2} = m, Y_{N+3} = j]
\]

for all types \( j, m \) and for all outcomes \( i, l, k \). This poses a significant constraint on how counts of outcomes of type \( m \) can influence predictive probabilities for outcomes of type \( j \). Suppose that \( l = k \). Then the left side of the equation is a conditional joint probability where a \( k \)-outcome of type \( m \) happens before a \( k \)-outcome of type \( j \). On the right side it is the other way round. Thus, if counts of type \( m \) influence counts of type \( j \), this must happen in a very particular way in order for the left side and the right side to always be equal. For this case it is shown below (under some technical assumptions) that the generalized sufficientness postulate \( \text{(4)} \) and partial exchangeability imply that types \( m \) and \( j \) are either completely analogous or there is no analogy influence between the two types (see Corollary \( \text{[2]} \)).

In order for a generalization of Johnson’s sufficientness postulate to be effective, a broader concept of partial exchangeability should be considered. More specifically, we only require that, for all types \( j, m \),

\[
P[X_{N+1} = i, X_{N+2} = l, X_{N+3} = k|X_1, \ldots, X_N, Y_{N+1} = j, Y_{N+2} = m, Y_{N+3} = j]
= P[X_{N+1} = k, X_{N+2} = l, X_{N+3} = i|X_1, \ldots, X_N, Y_{N+1} = j, Y_{N+2} = m, Y_{N+3} = j].
\]

for all outcomes \( i, l, k \) such that \( i, k \neq l \). If e.g. \( k = l \), then equality in \( \text{(5a)} \) does not need to hold. This allows for the possibility that counts of type \( m \) might genuinely
influence counts of type \( j \). Of course, we also suppose that

\[
P[X_{N+1} = i, X_{N+2} = k | X_1, \ldots, X_N, Y_{N+1} = j, Y_{N+2} = j]
= P[X_{N+1} = k, X_{N+2} = i | X_1, \ldots, X_N, Y_{N+1} = j, Y_{N+2} = j].
\]

for all types \( j \) and all outcomes \( i, k \).

Whether or not (5a) also holds in its unrestricted form depends on how one judges the analogy influence between types. If it only holds in its restricted form, then analogy effects are, as we shall see, significantly less constrained. I discuss this issue more thoroughly in \( \S 4 \).

3. An Analogical Inductive Logic

In this section we derive a simple rule for analogical predictive probabilities along the lines of Johnson and Carnap and analyze some of its properties. In our new analogical inductive logic an observation of a certain outcome does not just influence the probability of that kind of outcome but also the probability of similar outcomes. It is based on the generalized concept of partial exchangeability as expressed by (5) and the generalized sufficientness postulate (4) together with two further assumptions that are to be introduced next.

3.1. A representation theorem for analogical predictive probabilities. We suppose that the following regularity condition holds:

\[
P[X_1 = x_1, \ldots, X_{N+1} = x_{N+1}, Y_1 = y_1, \ldots, Y_{N+3} = y_{N+3}] > 0
\]

for all combinations of outcomes \( x_1, \ldots, x_{N+1} \) and for all combinations of types \( y_1, \ldots, y_{N+3} \). The regularity postulate (6) guarantees that the conditional probabilities (4) are well defined.

Finally, in order to connect the two symmetry postulates (4) and (5), we assume that

\[
P[X_{N+1} = i | X_1, \ldots, X_N, Y_{N+1} = j]
= P[X_{N+1} = i | X_1, \ldots, X_N, Y_{N+1} = j, Y_{N+2} = k]
= P[X_{N+1} = i | X_1, \ldots, X_N, Y_{N+1} = j, Y_{N+2} = k, Y_{N+3} = l]
\]

for all outcomes \( i \) and all types \( j, k, l \). This postulate says that \( X_{N+1} \) and \( Y_{N+2} \) are conditionally independent given \( X_1, \ldots, X_N \) and \( Y_1, \ldots, Y_{N+1} \), and that \( X_{N+1} \) and \( Y_{N+3} \) are conditionally independent given \( X_1, \ldots, X_N \) and \( Y_1, \ldots, Y_{N+2} \). If one thinks of types as different kinds of experiments, then the experiments planned for trials \( N+1 \) and \( N+2 \) are independent of the outcome of the \( N+1 \)st trial.

Postulate (7) is not a trivial assumption. Suppose, for example, that the experiment is a certain medical treatment; if it leads to a success tomorrow, it might become more likely that this treatment is used again. Thus, the choice of future treatments is not independent of observing a success tomorrow. In general, the experiment one chooses to perform the day after tomorrow might depend on the outcome of tomorrow’s experiment.

Our main theorem provides a representation of predictive probabilities meeting our four constraints in terms of a simple family of inductive learning rules.

**Theorem 1.** Let \( X_1, \ldots, X_N \), \( (N^* \geq 2) \) be a sequence of random variables that take on values in \( \{1, \ldots, s\} \), \( 3 \leq s \leq \infty \). Let \( Y_1, \ldots, Y_{N^*+2} \) be the corresponding sequence of types taking values in \( \{1, \ldots, t\} \), \( 1 \leq t < \infty \). Suppose that for every \( N < N^* \) the
four postulates (4), (5), (6), and (7) hold. Let \( \sum_i n_{ij} = N_j \) and \( \sum_j N_j = N \). Suppose, moreover, that for each type \( j \) outcomes of that type are not independent of each other. Then there exist nonzero constants \( \kappa_{ij} \) (for each \( j \), either all \( \kappa_{ij} \) are positive or all are negative) and constants \( \beta_{jm}, 1 \leq i \leq s, 1 \leq j, m, \leq t, m \neq j \) such that

\[
N_j + \sum_{m \neq j} \beta_{jm} N_m + K_j \neq 0 \quad \text{(where } K_j = \sum_i \kappa_{ij} \text{) and}
\]

\[
\mathbb{P}[X_{N+1} = i|X_1, \ldots, X_N, Y_{N+1} = j] = \frac{n_{ij} + \sum_{m \neq j} \beta_{jm} n_{im} + \kappa_{ij}}{N_j + \sum_{m \neq j} \beta_{jm} N_m + K_j}
\]

for all \( N < N^* \) and all \( 0 \leq n_{ij} \leq N_j, 1 \leq i \leq s, 1 \leq j \leq t \).

**Remark 1.** For Carnapian inductive logic it is well known that if the number of outcomes \( s = 2 \), then Johnson’s sufficientness postulate is trivially satisfied. The same is true here for the analogous requirement (4). In order for the conclusions of Theorem 1 and its corollary to be valid if \( s = 2 \), it needs to be assumed that predictive probabilities are linear in the \( n_{ij} \) (see Lemma 1 below). An alternative approach could probably be developed by using relevance quotients, as in Costantini (1979).

Before proving Theorem 1, it is useful to note two immediate consequences (which are also proved below). First, the constraints on the parameters are more stringent if \( X_1, X_2, \ldots \) is an infinite sequence.

**Corollary 1.** Let \( X_1, X_2, \ldots \) be an infinite sequence of outcomes such that the assumptions of Theorem 1 hold. Then \( \kappa_{ij} > 0 \) and \( \beta_{jm} \geq 0 \), \( 1 \leq i \leq s, 1 \leq j, m \leq t, m \neq j \).

The second corollary deals with the case where (4) holds unrestrictedly. It states that types must then either be independent or outcomes must be exchangeable across types.

**Corollary 2.** Let \( X_1, X_2, \ldots X_{N^*} \) (\( N^* \geq 2 \)) be a sequence of outcomes such that the assumptions of Theorem 1 hold. Then (5a) holds for \( l = k \) if and only if either \( \beta_{jm} = \beta_{mj} = 0 \) or \( \beta_{jm} = \beta_{mj} = 1 \).

If (5a) holds for all outcomes, then there are no persistent analogy influences and postulate 1 essentially reduces to a form of Johnson’s sufficientness postulate.

### 3.2. Proofs of Theorem 1 and Corollaries 1 and 2

The proof of Theorem 1 is an adaptation of Johnson’s argument (Johnson 1932; Zabell 1982) to the present setting. In fact, if the number of types \( t = 1 \), the argument reduces to this proof and postulate (7) is dispensable. Thus we assume that \( t \geq 2 \).

We start by showing that our analogue of the sufficientness postulate implies that predictive probabilities are linear in \( n_{i1}, \ldots, n_{it} \).

**Lemma 1.** If (4) and (6) hold, then for every outcome \( i \) and every type \( j \) there exist constants \( a_{ij} > 0 \) and \( b_{j1}, \ldots, b_{jt} \) such that

\[
f_{ij}(n_{i1}, \ldots, n_{it}, N_1, \ldots, N_t) = a_{ij} + b_{j1}n_{i1} + \cdots + b_{jt}n_{it},
\]

where the constants only depend on \( N_1, \ldots, N_t \).

**Proof.** Fix an outcome \( i \) and a type \( j \). We start with considering the influence of type 1 outcomes. Suppose that \( N_1 \geq 2 \), and that \( 0 < n_{i1}, n_{k1} \) and \( n_{i1}, n_{i1} < N_1 \),
where \( i, k, l \) are distinct outcomes. Let
\[
N_1 = (n_{11}, \ldots, n_{i1}, \ldots, n_{k1}, \ldots, n_{lt}, \ldots, n_{s1})
\]
\[
N_2 = (n_{11}, \ldots, n_{i1} + 1, \ldots, n_{k1} - 1, \ldots, n_{lt}, \ldots, n_{s1})
\]
\[
N_3 = (n_{11}, \ldots, n_{i1}, n_{k1} - 1, \ldots, n_{lt} + 1, \ldots, n_{s1})
\]
\[
N_4 = (n_{11}, \ldots, n_{i1} - 1, \ldots, n_{k1}, \ldots, n_{lt} + 1, \ldots, n_{s1})
\]
The equality
\[
\sum_r f_{rj}(n_{r1}, \ldots, n_{rt}) = 1,
\]
holds for each \( N_1, N_2, N_3, N_4 \), with fixed values for the \( n_{rm}, m \neq 1 \), the only constraint being \( \sum_r n_{rm} = N_m \) (here we suppress the counts \( N_1, \ldots, N_t \) in order to simplify notation). This results in four equations, which together imply, as in the first part of the proof of Lemma 2.1 in Zabell (1982), that
\[
f_{ij}(n_{i1} + 1, \ldots, n_{it}) - f_{ij}(n_{i1}, \ldots, n_{it})
= f_{kj}(n_{k1}, \ldots, n_{kt}) - f_{kj}(n_{k1} - 1, \ldots, n_{kt})
\]
\[
= f_{ij}(n_{i1} + 1, \ldots, n_{it}) - f_{ij}(n_{i1}, \ldots, n_{it})
\]
\[
= f_{ij}(n_{i1}, \ldots, n_{it}) - f_{ij}(n_{i1} - 1, \ldots, n_{it}).
\]
Hence
\[
f_{ij}(n_{i1}, \ldots, n_{it}, N_1, \ldots, N_t) = f_{ij}(0, n_{i2}, \ldots, n_{it}, N_1, \ldots, N_t) + b_{j1}n_{i1},
\]
where
\[
b_{j1} = f_{ij}(n_{i1}, \ldots, n_{it}, N_1, \ldots, N_t) - f_{ij}(n_{i1} - 1, \ldots, n_{it}, N_1, \ldots, N_t).
\]
The second part of Zabell’s proof of Lemma 2.1 shows that this also holds if \( N_1 = 1 \).

In the above calculations it is essential to fix the counts \( N_1, \ldots, N_t \). We now show that \( n_{i2}, \ldots, n_{it} \) may vary. In order to prove that \( b_{j1} \) does not depend on \( n_{i2} \), it is sufficient to show that
\[
f_{ij}(n_{i1}, n_{i2}, \ldots, n_{it}) - f_{ij}(n_{i1} - 1, n_{i2}, \ldots, n_{it})
= f_{ij}(n_{i1}, n_{i2} + 1, \ldots, n_{it}) - f_{ij}(n_{i1} - 1, n_{i2} + 1, \ldots, n_{it}).
\]
for \( 0 \leq n_{i2} \leq N_2 - 1 \). If \( N_2 = 0 \), nothing needs to be shown. If \( N_2 > 0 \), suppose that \( n_{k2} > 0 \). Consider the two equations
\[
f_{ij}(n_{i1}, n_{i2} + 1, \ldots, n_{it}) + f_{kj}(n_{k1}, n_{k2} - 1, \ldots, n_{kt})
+ f_{ij}(n_{i1}, \ldots, n_{it}) + \sum_{r \neq i, k, l} f_{rj}(n_{r1}, \ldots, n_{rt}) = 1
\]
and
\[
f_{ij}(n_{i1} - 1, n_{i2} + 1, \ldots, n_{it}) + f_{kj}(n_{k1}, n_{k2} - 1, \ldots, n_{kt})
+ f_{ij}(n_{i1} + 1, \ldots, n_{it}) + \sum_{r \neq i, k, l} f_{rj}(n_{r1}, \ldots, n_{rt}) = 1.
\]
Subtracting the first from the second equation yields
\[
f_{ij}(n_{i1} + 1, \ldots, n_{it}) - f_{ij}(n_{i1}, \ldots, n_{it})
= f_{ij}(n_{i1}, n_{i2} + 1, \ldots, n_{it}) - f_{ij}(n_{i1} - 1, n_{i2} + 1, \ldots, n_{it}).
\]
Equation (11) now follows from (10). Thus \( b_{j1} \) does not depend on \( n_{i2} \). The same type of argument can be applied to \( n_{i3}, \ldots, n_{it} \). Hence, \( b_{j1} \) may depend on \( N_1, \ldots, N_t \), but it does not depend on \( i \) (because of (10)) nor on \( n_{i1}, \ldots, n_{it} \).
The same arguments can now be used for \( n_{i2} \), with the result that
\[
f_{ij}(0, n_{i2}, \ldots, n_{it}, N_1, \ldots, N_t) = f_{ij}(0, 0, n_{i3}, \ldots, n_{it}, N_1, \ldots, N_t) + b_{j2} n_{i2}
\]
for some constant \( b_{i2} \) depending on \( N_1, \ldots, N_t \). Since \( t < \infty \), repeated applications lead to
\[
f_{ij}(n_{i1}, \ldots, n_{it}, N_1, \ldots, N_t) = f_{ij}(0, \ldots, 0, N_1, \ldots, N_s) + b_{j1} n_{i1} + \cdots + b_{jt} n_{it}.
\]
If we set
\[
a_{ij} = f_{ij}(0, \ldots, 0, N_1, \ldots, N_t),
\]
then \( \text{(9)} \) is established. That \( a_{ij} > 0 \) follows from \( \text{(6)} \). □

Lemma 1 establishes the basic representation of our predictive probabilities. In order to bring them into the form as given in \( \text{(8)} \), let \( A_j = \sum_r a_{rj} \). Then
\[
A_j + b_{j1} N_1 + \ldots + b_{jt} N_t = 1.
\]
Suppose that \( b_{jj} \neq 0 \) and let \( \kappa_{ij} = a_{ij}/b_{jj}, \beta_{jm} = b_{jm}/b_{jj}, k \neq j \). Then \( K_j = A_j/b_{jj} \) and
\[
\frac{1}{b_{jj}} = N_j + \sum_{m \neq j} \beta_{jm} N_m + K_j.
\]
It follows from Lemma 1 that
\[
f_{ij}(n_{i1}, \ldots, n_{it}, N_1, \ldots, N_t) = \frac{n_{ij} + \sum_{m \neq j} \beta_{jm} n_{im} + \kappa_{ij}}{N_j + \sum_{m \neq j} \beta_{jm} N_m + K_j}.
\]
Note that \( a_{ij} \) and \( b_{jm} \) depend in general on \( N_1, \ldots, N_t \). Thus it remains to show that the \( \kappa_{ij}, \beta_{jm} \) do not depend on \( N_1, \ldots, N_t \). This is achieved in Lemma 3 for the case where observations within a type are not independent. The case of independence is considered in the next lemma.

**Lemma 2.** Let \( X_1, \ldots, X_{N+1}, X_{N+2}, X_{N+3}, N \geq 1 \) be a sequence of random variables such that \( \text{(4)}, \text{(5)}, \text{(6)}, \) and \( \text{(7)} \) hold. Then for all types \( j \) and \( m \), if
\[
b_{jj}(N_1, \ldots, N_m, \ldots, N_t) : b_{jj}(N_1, \ldots, N_m + 1, \ldots, N_t) = 0,
\]
then
\[
b_{jj}(N_1, \ldots, N_m, \ldots, N_t) = b_{jj}(N_1, \ldots, N_m + 1, \ldots, N_t) = 0,
\]
provided that \( N_j > 0 \).

**Proof.** Let \( i \neq k \). Consider first the case \( j = m \). Postulate \( \text{(5b)} \) says that
\[
\begin{array}{rcl}
\mathbb{P}[X_{N+1} = i, X_{N+2} = k | X_1, \ldots, X_N, Y_{N+1} = j, Y_{N+2} = j] & = & \mathbb{P}[X_{N+1} = k, X_{N+2} = i | X_1, \ldots, X_N, Y_{N+1} = j, Y_{N+2} = j].
\end{array}
\]
Let \( a_{ij} = a_{ij}(N_1, \ldots, N_j, \ldots, N_t), a'_{ij} = a_{ij}(N_1, \ldots, N_j + 1, \ldots, N_t), b_{jj} = b_{jj}(N_1, \ldots, N_j, \ldots, N_t), b'_{jj} = b'_{jj}(N_1, \ldots, N_j + 1, \ldots, N_t) \), etc. Let \( \sum_r = \sum_{m \neq j} b_{jm} n_{rm} \) and \( \sum'_r = \sum_{m \neq j} b'_{jm} n_{rm} \). Then, by using \( \text{(7)} \) and \( \text{(9)}, \text{(13)} \) implies
\[
\begin{array}{rcl}
\left( a_{ij} + b_{jj} n_{ij} + \sum_i \right) \left( a'_{kj} + b'_{jj} n_{kj} + \sum_k' \right) & = & \left( a_{kj} + b_{jj} n_{kj} + \sum_k \right) \left( a'_{ij} + b'_{jj} n_{ij} + \sum_i' \right).
\end{array}
\]
Suppose that \( b_{ij} = 0 \) and \( n_{im} = n_{km} = 0, m \neq j \) (this is possible without changing the counts \( N_m \) since \( s \geq 3 \)). By first setting \( n_{kj} = N_j \) and then setting \( n_{ij} = N_j \) we get the following two equations:

\[
\begin{align*}
    a_{ij} (a_{kj} + b'_{jj} N_j) &= a_{kj} a'_{ij} \quad a_{ij} a'_{kj} = a_{kj} (a_{ij} + b'_{jj} N_j) \\
\end{align*}
\]

Subtracting the second from the first equation yields \( a_{ij} b'_{jj} N_j = -a_{kj} b'_{jj} N_j \). If \( N_j > 0 \), then \( b'_{jj} = 0 \) since \( a_{ij}, a_{kj} > 0 \). An analogous argument shows that assuming \( b'_{jj} = 0 \) implies \( b_{jj} = 0 \).

Consider now the case \( j \neq m \). Postulate (5a) states that

\[
P[X_{N+1} = i, X_{N+2} = l, X_{N+3} = k]X_{1, \ldots, X_N}, Y_{N+1} = j, Y_{N+2} = m, Y_{N+3} = j]
= P[X_{N+1} = k, X_{N+2} = l, X_{N+3} = i]X_{1, \ldots, X_N}, Y_{N+1} = j, Y_{N+2} = m, Y_{N+3} = j],
\]

where we assume that \( i, l, k \) are distinct. By again using postulate (7) it follows that

\[
\begin{align*}
    f_{ij}(n_{i1}, \ldots, n_{it}, N_1, \ldots, N_j, \ldots, N_t) \\
    \cdot f_{kj}(n_{k1}, \ldots, n_{kt}, N_1, \ldots, N_m + 1, \ldots, N_j + 1, \ldots, N_t)
    &= f_{kj}(n_{k1}, \ldots, n_{kt}, N_1, \ldots, N_j, \ldots, N_t) \\
    \cdot f_{ij}(n_{i1}, \ldots, n_{it}, N_1, \ldots, N_m + 1, \ldots, N_j + 1, \ldots, N_t)
\end{align*}
\]

Since the second term is the same on each side, this implies

\[
(a_{ij} + b_{jj} n_{kj} + \sum_k) (a''_{ij} + b''_{jj} n_{ij} + \sum_{ij})
= (a_{kj} + b_{jj} n_{kj} + \sum_k) (a''_{kj} + b''_{jj} n_{ij} + \sum_{ij})
\]

where now \( a''_{ij} = a_{ij}(N_1, \ldots, N_m + 1, \ldots, N_j + 1, \ldots, N_t) \) and \( b''_{jj} = b_{jj}(N_1, \ldots, N_m + 1, \ldots, N_j + 1, \ldots, N_t) \), etc. The same argument following equation (14) shows that \( b''_{jj} = 0 \) whenever \( b_{jj} = 0 \) and vice versa (assuming that \( N_j > 0 \)). From the first part of the proof it follows that \( b'_{jj} = b_{jj}(N_1, \ldots, N_m + 1, \ldots, N_j, \ldots, N_t) \), and the conclusion of the lemma follows.

Lemma (2) shows that if \( b_{jj}(0, \ldots, N_j, \ldots, 0) = 0 \) for \( N_j = 1 \), then we have \( b_{jj}(N_1, \ldots, N_m, \ldots, N_t) = 0 \) for all \((N_1, \ldots, N_t)\) with \( N_j > 0 \). If \( b_{jj}(0, \ldots, N_j, \ldots, 0) = 0 \) for \( N_j = 1 \), then an observation of an outcome of type \( j \) is probabilistically irrelevant for the predictive probability of another outcome of type \( j \) since in this case

\[
f_{ij}(0, \ldots, n_{ij}, \ldots, 0, \ldots, N_j, \ldots, 0) = a_{ij}(0, \ldots, N_j, \ldots, 0),
\]

where \( n_{ij} \) may be zero or one.

The next lemma considers the case of \( b_{jj}(0, \ldots, N_j, \ldots, 0) \neq 0 \) for \( N_j = 1 \). By Lemma (2) this implies \( b_{jj}(N_1, \ldots, N_m, \ldots, N_t) \neq 0 \) for all \((N_1, \ldots, N_t)\) with \( N_j > 0 \).

**Lemma 3.** Let \( X_1, X_{N+1}, X_{N+2}, X_{N+3}, N \geq 1 \) be a sequence of random variables such that (1), (2), (3), and (4) hold. Suppose that \( N_j > 0 \), and that for all \( 1 \leq m \leq t, \ b_{jj}(N_1, \ldots, N_m, \ldots, N_t) \neq 0 \). Then the following three statements are true for any type \( m \):

(1) \( b_{jj}(N_1, \ldots, N_m, \ldots, N_t) b_{jj}(N_1, \ldots, N_m + 1, \ldots, N_t) > 0 \);
(2) \( n_{ij}(N_1, \ldots, N_m, \ldots, N_t) = \kappa_{ij}(N_1, \ldots, N_m + 1, \ldots, N_t) \); and
(3) for \( u \neq j \), if \( N_u > 0 \), then \( \beta_{ju}(N_1, \ldots, N_m, \ldots, N_t) = \beta_{ju}(N_1, \ldots, N_m + 1, \ldots, N_t) \).
Proof. (1) and (2): Suppose that $m = j$. The relevant predictive probabilities can be given by (12) since $n_{ij}(N_1, \ldots, N_m, \ldots, N_l) \cdot b_{ij}(N_1, \ldots, N_m + 1, \ldots, N_l) \neq 0$. Setting $n_{il} = n_{kl} = 0$ for all $l \neq i$ in equation (14) implies

$$\frac{n_{ij} + \kappa_{ij}}{N_j + \sum_{j} + K_j} \left( \frac{n_{kj} + \kappa'_{kj}}{N_j + 1 + \sum_{k} + K_j} \right) = \frac{n_{ij} + \kappa'_{ij}}{N_j + \sum_{j} + K_j} \left( \frac{n_{ij} + \kappa''_{ij}}{N_j + 1 + \sum_{j} + K_j} \right)$$

where $\kappa_{ij} = \kappa_{ij}(N_1, \ldots, N_m, \ldots, N_s)$, $\kappa'_{ij} = \kappa_{ij}(N_1, \ldots, N_j + 1, \ldots, N_s)$, $\beta_{ij} = \beta_{jm}(N_1, \ldots, N_j, \ldots, N_s)$, $\beta'_{jm} = \beta_{jm}(N_1, \ldots, N_j + 1, \ldots, N_s)$, $\sum_j = \sum_{m \neq j} \beta_{jm} N_m$, $\sum'_{j} = \sum_{m \neq j} \beta'_{jm} N_m$, etc.

Hence

$$\kappa_{ij} n_{ij} + \kappa_{kj} n_{ij} + \kappa_{ij} \kappa'_{kj} = \kappa_{kj} n_{ij} + \kappa_{ij} \kappa'_{kj}.$$  

First setting $n_{kj} = N_j$ and then $n_{ij} = N_j$, and subtracting yields $\kappa_{ij} + \kappa_{kj} = \kappa'_{ij} + \kappa'_{kj}$. Since $s \geq 3$, it follows that $\kappa_{ij} = \kappa'_{ij}$. Since $a_{ij}, a'_{ij} > 0$, we must have $b_{ij}(N_1, \ldots, N_j, \ldots, N_l) \cdot b_{ij}(N_1, \ldots, N_j + 1, \ldots, N_l) > 0$.

Suppose now that $m \neq j$. Then as in (15)

$$\frac{n_{ij} + \sum_{ij} + \kappa_{ij}}{N_j + \sum_{j} + K_j} \left( \frac{n_{kj} + \sum'_{kj} + \kappa'_{kj}}{N_j + 1 + \sum'_{j} + K_j} \right) = \frac{n_{ij} + \sum'_{ij} + \kappa'_{ij}}{N_j + \sum_{j} + K_j} \left( \frac{n_{ij} + \sum''_{ij} + \kappa''_{ij}}{N_j + 1 + \sum''_{j} + K_j} \right)$$

where $\kappa''_{ij} = \kappa_{ij}(N_1, \ldots, N_m + 1, \ldots, N_j + 1, \ldots, N_s)$, $\beta''_{ij} = \beta_{ij}(N_1, \ldots, N_m + 1, \ldots, N_j + 1, \ldots, N_s)$, $\sum''_j = \sum_{r \neq j} \beta_{jr} N_r$, $\sum'_{rj} = \sum_{r \neq j,m} \beta'_{jr} N_r$, etc. Now the argument following (17) can be applied to show that $\kappa_{ij} = \kappa''_{ij}$, and hence, that $b_{ij}(N_1, \ldots, N_m, \ldots, N_j, \ldots, N_s) \cdot (N_1, \ldots, N_m + 1, \ldots, N_j, \ldots, N_s) > 0$.

(3) Let $m = j$. Under the assumptions of the lemma,

$$f_{ij}(n_{i1}, \ldots, n_{it}, N_1, \ldots, N_j, \ldots, N_l) \cdot f_{kj}(n_{kt}, \ldots, n_{kt}, N_1, \ldots, N_j + 1, \ldots, N_l) = f_{kj}(n_{kt}, \ldots, n_{kt}, N_1, \ldots, N_j, \ldots, N_l) \cdot f_{ij}(n_{i1}, \ldots, n_{it}, N_1, \ldots, N_j + 1, \ldots, N_l)$$

implies that

$$\frac{\beta_{ju} n_{ku} + \kappa_{ij}}{N_j + \sum_{j} + K_j} = \frac{\beta''_{ju} n_{ku} + \kappa'_{kj}}{N_j + 1 + \sum_{k} + K_j}$$

provided that $n_{il} = n_{kl} = 0$ for all $l \neq u$. Since, by the first part of the proof, $\kappa_{ij} = \kappa'_{ij}$ for all outcomes $i$, it follows that

$$\kappa_{ij} n_{ku} (\beta''_{ju} - \beta_{ju}) = \kappa_{kj} n_{iu} (\beta''_{ju} - \beta_{ju}).$$
If \((\beta'_{jn} - \beta_{ju}) \neq 0\), then \(\kappa_{ij}n_{ku} = \kappa_{kj}n_{lu}\), which can only hold for all \(0 \leq n_{iu}, n_{ku} \leq N_u\) if \(N_u = 0\) (since \(\kappa_{iu} \kappa_{ku} \neq 0\)). Hence \(\beta_{ju}(N_1, \ldots, N_j, \ldots, N_l) = \beta_{ju}(N_1, \ldots, N_j + 1, \ldots, N_l)\) whenever \(N_u > 0\).

Consider now the case \(m \neq j\). Equation (18) implies that

\[
(n_{ij} + \sum_{ij} + \kappa_{ij})(n_{kj} + n'_{kj}) = (n_{kj} + \sum_{kj} + \kappa_{kj})(n_{ij} + n''_{ij} + \kappa_{ij}).
\]

Hence, if \(n_{ij} = n_{kj} = 0\),

\[
\sum_{ij} + \kappa_{ij} \sum_{kj} + \kappa_{kj} = \sum_{ij} + \kappa_{ij} \sum_{kj} + \kappa_{kj} + \kappa_{ij} \sum_{ij} + \kappa_{ij} \sum_{kj} + \kappa_{kj} \sum_{ij}''
\]

Suppose also that \(n_{ir} = n_{kr} = 0, r \neq u\). Then (17) reduces to

\[
\beta_{ju}n_{iu} \cdot \beta'_{j'u}n_{ku} + \kappa_{ij}\beta_{ju}n_{iu} + \kappa_{ij}\beta'_{j'u}n_{ku} = \beta_{ju}n_{ku} + \kappa_{ij}\beta_{ju}n_{ku} + \kappa_{ij}\beta'_{j'u}n_{iu}.
\]

By the same argument as in the case \(m = j\) it follows that \(\beta_{ju} = \beta'_{j'u}\) provided that \(N_u > 0\). The conclusion now follows since \(\beta'_{j'u} = \beta_{ju}(N_1, \ldots, N_m + 1, \ldots, N_j, \ldots, N_l)\).

Together, the three lemmata imply Theorem 1. The assumption that for each type the random variables \(X_i\) of that type are not independent excludes case (1) of Lemma 3; since then, for any \(j, b_{ij}(0, \ldots, N_j, \ldots, 0) \neq 0\) for \(N_j = 1\). By continuity, the representation of predictive probabilities (8) also holds as \(N_1, \ldots, N_l\) go to zero.

In order to prove Corollary 1 suppose that \(b_{ij}(0, \ldots, N_j, \ldots, 0) < 0\) for \(N_j = 1\). It follows from Lemma 3(1) that \(b_{j1}(N_1, \ldots, N_j, \ldots, N_l) < 0\) for all \(N_j \geq 1\). But then

\[
N_j + \sum_{m \neq j} \beta_{jm} N_m + K_j = \frac{1}{b_{jj}} < 0.
\]

To see that this cannot be the case for arbitrarily large \(N\) just set \(N_m = 0\) for \(m \neq j\) and let \(N_j\) be sufficiently large. Hence \(b_{jj}(N_1, \ldots, N_j, \ldots, N_l) > 0\), and thus also \(\kappa_{ij} > 0\) for all \(i\). Moreover, \(N_j + \sum_{m \neq j} \beta_{jm} N_m + K_j > 0\). Setting \(N_m = 0\) for all \(m \neq j, l\) and \(N_l = 1\), and letting \(N_l\) be sufficiently large shows that \(\beta_{jl} \geq 0\).

For proving Corollary 2 suppose that (5a) holds if \(l = k\). It follows from Theorem (1) that

\[
\left(\frac{n_{ij} + \sum_{r \neq j} \beta_{jr} n_{ir} + \kappa_{ij}}{N_j + \sum_{r \neq j} \beta_{jr} N_r + K_j}\right) \left(\frac{n_{km} + \sum_{r \neq m} \beta_{mr} n_{kr} + \kappa_{km}}{N_m + \beta_{mj}(N_j + 1) + \sum_{r \neq j, m} \beta_{mr} N_r + K_m}\right)
\]

\[
\left(\frac{n_{kj} + \beta_{jm}(n_{km} + 1)}{N_j + \sum_{r \neq j} \beta_{jr} N_r + K_j}\right) \left(\frac{n_{mj} + \beta_{jm}(n_{kj} + 1)}{N_m + \beta_{mj}(N_j + 1) + \sum_{r \neq j, m} \beta_{mr} N_r + K_m}\right)
\]

\[
\left(\frac{n_{ij} + \sum_{r \neq j} \beta_{jr} n_{ir} + \kappa_{ij}}{N_j + \sum_{r \neq j} \beta_{jr} N_r + K_j}\right).
\]

Set \(n_{kr} = 0\) for all \(r \neq j, m\). Then the above equation reduces to

\[
\beta_{jm}(n_{km} + \beta_{mj} n_{kj} + \kappa_{km}) = \beta_{mj}(n_{kj} + \beta_{jm} n_{km} + \kappa_{kj}).
\]

Suppose that \(n_{kj} = 0\). If \(n_{km} = 0\), then \(\beta_{jm} \kappa_{km} = \beta_{mj} \kappa_{kj}\). Thus, if \(n_{km} > 0\), \(\beta_{jm} n_{km} = \beta_{jm} \beta_{mj} n_{km}\). Hence, if \(\beta_{jm} \neq 0\), then \(\beta_{mj} = 1\). A similar argument shows that, if \(n_{kj}\) is set to zero and \(n_{km}\) is allowed to vary, then \(\beta_{mj} \neq 0\) implies
\[ \beta_{jm} = 1. \] It follows that either \( \beta_{jm} = \beta_{mj} = 0 \) or \( \beta_{jm} = \beta_{mj} = 1. \) If \( \beta_{jm} = \beta_{mj} = 0 \) or \( \beta_{jm} = \beta_{mj} = 1, \) then it is clear from \( 3 \) that \( 3a \) holds for \( l = k. \)

### 3.3. Urn models.

The predictive probabilities of the Johnson-Carnap continuum of inductive methods are mathematically equivalent to a Polya urn process.\(^8\) Analogical predictive probabilities can also be generated by urn processes.

Consider the case \( N^* = \infty. \) First, choose an infinite sequence of types \((y_1, y_2, \ldots)\) according to some probability distribution that assigns positive probability to every finite initial sequence of types. For each type \( j \) there is an urn containing \( \kappa_{ij} \) balls of color \( i \) (one color for each outcome). The process starts with choosing a ball at random from urn \( y_1, \) observing the color of the ball and replacing it with two balls of the same color. In addition, \( \beta_{jm} \geq 0 \) balls of that color are put in every urn \( m. \) This procedure is repeated for urns \( y_2, y_1, \ldots. \) The same urn model can be used if \( N^* < \infty \) as long as all parameters are nonnegative.

Consider now the case where all \( \kappa_{ij} \) are negative and all \( \beta_{jm} \) are such that \( |\beta_{jm}| \leq 1. \) Suppose that \( N^* < \min\{-\kappa_{ij}\}. \) First, choose a finite sequence of types of length \( N^* \) from a distribution that assigns every such sequence a positive probability. For each type \( j \) there is an urn containing \( -\kappa_{ij} \) balls of color \( i. \) Following the sequence of types, balls are chosen without replacement from urns. If a ball of color \( i \) is chosen from an urn of type \( j, \) then it is not replaced and \( \beta_{jm} \) balls of color \( i \) are subtracted from urn \( m \) (or added, in case \( \beta_{jm} \) is negative).

### 3.4. Special cases.

There are several important reductions of Theorem \( 11 \) to simpler cases. If the number of types \( t = 1, \) then the theorem reduces to the Johnson-Carnap continuum of inductive methods. The same happens if all \( \beta_{jm} = 1. \) In this case the analogy influence is the same among all types and as strong as the influence from an outcome’s own type. Another interesting case is given by \( \beta_{jm} = \beta \neq 1 \) for all \( j, m. \) The analogy influence among all types on \( j \) other than \( j \) itself is the same. This leads to a simplification of the inductive rule \( 8: \)

\[
P[X_{N+1} = i|X_1, \ldots, X_N, Y_{N+1} = j] = \frac{n_{ij} + \beta(N^i - n_{ij}) + \kappa_{ij}}{N_j + \beta(N - N_j) + K_j},
\]

where \( N^i = \sum_m n_{im} \) is the number of outcomes \( i \) regardless of type.

A further plausible requirement is that analogical influences are symmetric, that is \( \beta_{jm} = \beta_{mj} \) for all \( 1 \leq j, m \leq t. \) Finally, there is the case of no analogy influence: \( \beta_{jm} = 0 \) for all \( j \neq i. \) Here it is judged that only observations of type \( j \) are relevant for predictive probabilities regarding type \( j. \) This judgment again reduces the new inductive rule to the Johnson-Carnap continuum since the inductive probabilities of a type can be analyzed without regard of other types. By Corollary \( 2, \) this is the case when outcomes are partially exchangeable.

### 3.5. Some consequences.

The \( \beta_{jk}, k \neq j \) in \( 8 \) can be thought of as analogy parameters. To see this, note first that \( \beta_{jk} \) depends both on the priors \( \kappa_{ij} \) and on \( p_{ij,ik} = P[X_2 = i|X_1 = i, Y_1 = k, Y_2 = j]. \) Since

\[
p_{ij,ik} = \frac{\beta_{jk} + \kappa_{ij}}{\beta_{jk} + K_j},
\]

\(^8\)In the Polya urn process one starts with an urn containing a finite number of balls of finitely many different colors. One then chooses a ball at random from the urn, observes its colors and returns two balls of the same color to the urn.
Because $p_{ij,ik} < 1$, $\beta_{jk}$ is positive (negative, zero) whenever $p_{ij,ik} > P[X_1 = i|Y_1 = j]$ ($p_{ij,ik} < P[X_1 = i|Y_1 = j]$, $p_{ij,ik} = P[X_1 = i|Y_1 = j]$). Thus, the sign of $\beta_{jk}$ depends on whether observing an outcome-type pair $(i, j)$ beyond the prior probability of making that observation. The absolute value of $\beta_{jk}$ increases as the probability $p_{ij,ik}$ gets closer to one, i.e., if the observation of an $(i, k)$ event renders the observation of an $(i, j)$ event very probable.

Another reason to think of the $\beta_{jk}$ as analogy parameters is given in the following proposition, which describes the effect of these parameters on predictive probabilities.

**Proposition 1.** Suppose that all assumptions of Theorem 1 hold and that for type $j$ all $\kappa_{ij}$ are positive. If $\beta_{jk} > \beta_{jl}$, then for all $N < N^*$

$$P[X_{N+1} = i|X_1, \ldots, X_{N-1}, X_N = i, Y_N = k, Y_{N+1} = j] \geq P[X_{N+1} = i|X_1, \ldots, X_{N-1}, Y_N = l, Y_{N+1} = j].$$

If $\beta_{jk} > \beta_{jl}$, then the inequality is strict.

**Proof.** By Theorem 1 the inequality (21) is equivalent to

$$\frac{n_{ij} + \beta_{jk}(n_{ik} + 1) + \sum_{m \neq j,k} \beta_{jm} n_{im} + \kappa_{ij}}{N_j + \beta_{jk}(N_k + 1) + \sum_{m \neq j,k} \beta_{jm} N_m + K_j} \geq \frac{n_{ij} + \beta_{jl}(n_{il} + 1) + \sum_{m \neq j,l} \beta_{jm} n_{im} + \kappa_{ij}}{N_j + \beta_{jk}(N_l + 1) + \sum_{m \neq j,l} \beta_{jm} N_m + K_j},$$

where the $n_{ij}, N_j$ etc. denote counts up to period $N - 1$. Rearranging and simplifying shows that this inequality is equivalent to

$$(\beta_{jk} - \beta_{jl}) \left( K_j - \kappa_{ij} + N_j + \sum_{m \neq j} \beta_{jm} N_m - \left( n_{ij} + \sum_{m \neq j} \beta_{jm} n_{im} \right) \right) \geq 0.$$ 

Since

$$N_j + \sum_{m \neq j} \beta_{jm} N_m \geq n_{ij} + \sum_{m \neq j} \beta_{jm} n_{im}$$

and because $\beta_{jk} \geq \beta_{jl}$, the inequality (22) holds if

$$K_j - \kappa_{ij} \geq 0.$$ 

If the $\kappa_{ij}$ are assumed to be positive, the assertion of the proposition follows. □

Thus, the larger $\beta_{jk}$ the more influence do observations of type $k$ have on the predictive probabilities for type $j$. In fact, Theorem 1 does not exclude the possibility that $\beta_{jk} > 1$; in this case, observations of type $k$ have a higher influence on the predictive probability for outcomes of type $j$ than observations of that type. This does not need to be unreasonable, for example in situations where observations of type $k$ are considered to be more reliable. However, analogy influences are often thought of in terms of similarity, where each type is considered to be most similar to itself. The idea that the influence from other types should not surpass the influence from the focal type can be easily incorporated into the present framework by requiring, in addition to the axioms assumed by Theorem 1, that for all $k \neq j$

$$P[X_2 = i|X_1 = i, Y_1 = j, Y_2 = j] \geq P[X_2 = i|X_1 = i, Y_1 = k, Y_2 = j].$$
It is then easy to show that every type of observations has the maximal analogy influence on itself.

**Proposition 2.** Suppose that all assumptions of Theorem (1) hold, and that for type \( j \) all \( \kappa_{ij} \) are positive. If (23) is true for all \( k \neq j \), then \( \beta_{jk} \leq 1, k \neq j \).

**Proof.** Suppose that (23) holds. It follows from Theorem 1 that
\[
\frac{1 + \kappa_{ij}}{1 + K_j} \geq \frac{\beta_{jk} + \kappa_{ij}}{\beta_{jk} + K_j}.
\]
This implies
\[
(1 - \beta_{jk})(K_j - \kappa_{ij}) \geq 0.
\]
Since all \( \kappa_{ij} \) are positive, it follows that \( \beta_{jk} \leq 1 \).

Under the assumptions of the previous proposition, it immediately follows from Proposition 1 that
\[
P[X_{N+1} = i|X_1, \ldots, X_N = i, Y_N = j, Y_{N+1} = j] \geq P[X_{N+1} = i|X_1, \ldots, X_N = i, Y_N = k, Y_{N+1} = j]
\]
for all \( k \neq j \).

In the light of the previous two propositions, Achinstein’s example fits the present setting very well. Recall that Achinstein (1963) argued that the number of times a piece of platinum or osmium is observed to conduct heat should influence the predictive probability that the next piece of rhodium to be observed conducts heat. Rhodium, osmium and platinum are different types, and the outcomes are whether the piece of metal under consideration conducts heat.\(^9\) Now, given that the next observation is a piece of rhodium, whether it conducts heat depends on the other types of metal that were observed, the extent of the dependance being given by the analogy influence.

A deeper understanding of the process described by the analogical predictive probabilities (8) can be gained by studying the limit of (8) for infinite sequences \( X_1, X_2, \ldots \). Denote the limit of \( N_j/N_k \) by \( \rho_{jk} \) (provided that it exists). Let \( A_{jk} \) be given by
\[
A_{jk} = \frac{\beta_{jk} N_k}{N_j + \sum_{m \neq j} \beta_{jm} N_m + K_j},
\]
and let \( B_j \) be given by
\[
B_j = \frac{K_j}{N_j + \sum_{m \neq j} \beta_{jm} N_m + K_j}
\]
Then
\[
P[X_{N+1} = i|X_1, \ldots, X_N, Y_{N+1} = j] = \left( 1 - \sum_{k \neq j} A_{jk} \right) \frac{n_{ij}}{N_j} + \sum_{k \neq j} A_{jk} \frac{n_{ik}}{N_k} + B_j \kappa_{ij}.
\]

\(^9\)There is a technical caveat here since any application of Theorem 1 requires there to be at least three outcomes. See Remark 1 for a discussion. In Achinstein’s example, one could instead think of three outcomes: clearly conducts heat, seems to conduct heat, and clearly doesn’t conduct heat.
Suppose that each type occurs infinitely often, that \( \rho_{jk} \) exists for all types \( j \neq k \), and that \( \lim n_{ij}/N_j = \eta_{ij} \) exists for all outcomes \( i \) and all types \( j \). Then the predictive probability converges to

\[
(1 - \sum_{k \neq j} \beta_{jk} \rho_{jk} + \sum_{m \neq j} \beta_{jm} \rho_{mk}) \eta_{ij} + \sum_{k \neq j} \beta_{jk} \rho_{jk} \eta_{ik}.
\]

It is clear that, in general, the predictive probabilities \( P[X_{N+1} = i | X_1, \ldots, X_N, Y_{N+1} = j] \) will not converge to the relative frequencies \( \eta_{ij} \) but to a convex combination of the relative frequencies \( \eta_{ik}, 1 \leq k \leq t \). Under the plausible assumption that \( \rho_{jk} = 1, 1 \leq j, k \leq t, k \neq j \), this mixture is given by

\[
\frac{1}{1 + \sum_{k \neq j} \beta_{jk}} \left( \eta_{ij} + \sum_{k \neq j} \beta_{jk} \eta_{ik} \right).
\]

This reflects an assumption built into the basic symmetry postulate (5): if (5) does not hold unrestrictedly, then analogy influences between types may be persistent. Thus, sample information from types other than \( j \) can have a non-vanishing effect on predictive probabilities for outcomes of type \( j \).

4. Discussion

4.1. Related literature. It is well known that Carnap’s original systems of inductive logic (Carnap, 1950, 1952) can account for limited forms of analogical inference (Carnap and Stegmüller, 1959; Niiniluoto, 1981; Romeijn, 2006). But there are various examples of valid inductive analogical inferences that are not captured by Carnapian predictive probabilities, as well as a number of attempts at solving this problem (Achinstein, 1963; Hesse, 1964; Carnap, 1980; Niiniluoto, 1981; Spohn, 1981; Costantini, 1983; Kuipers, 1984; Skyrms, 1993a; di Maio, 1995; Pesta, 1997; Maher, 2000, 2001; Romeijn, 2006; Hill and Paris, 2013).

This is not the place for a detailed review of alternative analogical inductive logics. I will make a few general remarks, though, before discussing in more detail an aspect that seems to be particularly relevant for our new inductive logic.

As mentioned in §2, there are two main sources for why Carnapian inductive logic cannot capture relevant types of analogical inference. One reason is that exchangeability assumes “too much” analogy since trials are assumed to be perfectly similar. The other reason is that Johnson’s sufficientness postulate (equivalently, the assumption of a Dirichlet prior) makes it impossible that observations of one type make observations of a similar type more probable.

Many of the above mentioned analogical inference rules (e.g., Skyrms, 1993a; Festa, 1997; Romeijn, 2006; Hill and Paris, 2013) preserve exchangeability. This generally leads to predictive probabilities where the influence of analogy wears off rather quickly as more observations are made. In this respect these proposals are similar to de Finetti’s understanding of analogy based on partial exchangeability. As we have seen in our system, there is a certain tension between partial exchangeability and the generalized sufficientness postulate (4), which might explain why the analogy influence vanishes quite rapidly in these systems.

Several authors (Niiniluoto, 1981; Spohn, 1981; Costantini, 1983) consider, as we do, the possibility of seriously incorporating the counts of similar types to influence predictive probabilities. Their proposals are not without merits, since they derive
certain desirable consequences of their inductive rules. The same is true for Kuipers (1984). However, these approaches generally lack a foundation in terms of a set of axioms from which an inductive rule or a family of rules can be derived. (A notable exception is Maher (2000), whose inductive logic is different in several other respects, though.) As was stressed earlier, such an axiomatic foundation is the gold standard for judging to what extent an inductive logic is adequate and when it should be used. Regarding our system as given by (8) I shall discuss this issue now in more detail.

4.2. Reichenbach’s axiom. The property of our analogical inductive logic mentioned at the end of §3—that analogy factors have a non-vanishing influence—might be thought of as a liability of the present approach. In order to understand why, recall that Reichenbach’s axiom (also called the axiom of convergence) requires that predictive probabilities converge to the limiting relative frequencies, provided that they exist (Carnap, 1980). This is the case for the inductive rule (1): If the limit of \(n_i/n\) exists, then the ratio on the right side of (1) clearly converges to that limit.

Something like this is not in general the case for our analogical inductive logic. In (24) we see that the predictive probability for an outcome of one type may not just depend on its limiting relative frequency but also on the limiting relative frequency of that outcome for other types. Thus Reichenbach’s axiom appears to fail. The same is true in the systems of Niiniluoto (1981) and Costantini (1983) (who, on the other hand, don’t use de Finetti’s framework of partial exchangeability).

There is a strong tendency to take Reichenbach’s axiom in fact as axiomatic for any inductive logic. This view is articulated in Spohn (1981) and in many other works on analogical inductive logic. Reichenbach’s axiom is one way to express the idea that analogy influences should vanish over time as increasing evidence makes obsolete any information one might have from analogical considerations.

Since the present system violates these features, I now try to give a plausible defense of it along the lines suggested by Johnson himself. Discussing the sufficientness postulate he states that

“the postulate adopted in a controversial kind of theorem cannot be generalised to cover all sorts of working problems; so that it is the logician’s business, having once formulated a specific postulate, to indicate very carefully the factual or epistemic conditions under which it has practical value.”

I thus suggest that there are some epistemic situations where the axioms of Theorem 1 and the resulting inductive logic are appropriate, without claiming that this is the case whenever analogy is involved. Quite to the contrary—I think there are different types of sound analogical inductive inferences, some of which have been introduced in the literature cited above. However—and here I refer to Johnson’s idea—since most of the alternative systems lack a rigorous axiomatic foundation, it is much more difficult to determine the “factual or epistemic conditions” which would make them appropriate.

4.3. Weak and strong analogy. One feature of our inductive logic is that it also applies if the total number of observations \(N^*\) is finite. In this case the problem with Reichenbach’s axiom does not arise.
This, however, does not tell us yet about the situations where the inductive rule (8) is adequate. In a certain sense, this question is answered by simply pointing to the assumptions of Theorem 1. Whenever they hold, you are bound by consistency to apply rule (8). From this it follows that, whenever the assumptions of Theorem 1 hold for an infinite sequence of observations, Reichenbach’s axiom may fail. Are there any examples that would make such a situation plausible?

Let’s return to the example of flipping two coins and Achinstein’s example of the conductivity of different types of metal. The first one was used by de Finetti to illustrate his idea of partial exchangeability. It is clear that he thought of analogy influences in this and other examples as transient; the initial similarity between the coins is superficial, and since coin flips are independent given the biases, the superficial influence between them cannot, and should not, have a persistent influence on predictive probabilities since they only provide initial information about the two coins.

Achinstein’s example might be construed differently. Whether or not rhodium, osmium or platinum conduct heat may be due to some deep and perhaps not fully known chemical similarity between the three elements. In a manner of speaking one could refer to a “common cause” that determines certain physical properties of these metals. The deep similarity between the three types may arguably have a non-vanishing influence on predictive probabilities. Even after you have observed many pieces of rhodium, osmium and platinum, you may judge the three to be sufficiently similar so that further observations of one type always provide you with information about the other two.

The same can be said of an example of partially exchangeable medical trials discussed by Diaconis and Freedman (1980). Subjects are categorized into four groups: treated males, untreated males, treated females, and untreated females. If the male and female subjects judged, with respect to a particular disease, to be sufficiently similar so that observations about, say, treating a female subject always provides significant information about treating male subjects, then an inductive logic like ours may be adequate.

In general, the conditions for which the inductive rule (8) is adequate are ones where the covariate between types is supposed to have a sustained influence on the outcomes that are being observed throughout the process. In this case, (5a) is a plausible inductive assumption. It should be noted that one instance where this scenario does not conflict with Reichenbach’s axiom is when the limiting relative frequencies of outcomes of different types are the same. This can be taken as one expression of judging types to be similar in a strong sense.

The issues discussed so far suggest a rough distinction between two types of analogical inductive inference into weak and strong analogies. Weak analogies are treated in inductive logics where analogy influences vanish more or less quickly. Some of the theories mentioned earlier—such as Kuipers (1984),Skyrms (1993a), Festa (1997), Romeijn (2006), or Hill and Paris (2013)—are of this type. Our inductive logic, as well as the theories proposed by Niiniluoto (1981) and Costantini (1983), are of the strong type. Whether the strong type is appropriate—especially with regard to violations of Reichenbach’s axiom—needs to be discussed within the context of a particular inductive situation where various issues will play a role (e.g. the total number of observations and background information concerning the kind of similarity between types).
It would be desirable to develop a more general model than ours where analogy factors may or may not decrease over time or can be revised by evidence. At this point it is not clear how this goal could be achieved in a straightforward way. Note that our generalized sufficientness postulate (4) together with (6) implies the rather strong conclusion that the analogy parameters $b_{jm}$ in (9) only depend on $N_1, \ldots, N_t$. This already constrains the possible forms of predictive probabilities considerably. But if analogy influences should be decreasing in $N_1, \ldots, N_t$, then the appropriate symmetry conditions will be more complex than those given by (5). Suppose, for example, that the analogy parameters are given by $\frac{1}{N}$. Then the conditional probability of observing outcomes $i$ and $k$ (both of type $j$) at trials $N + 1, N + 2$ is

\[
\frac{ni_{ij}}{N_j + \frac{N-N_1}{N} + K_j} \frac{nk_{kj}}{N_j + 1 + \frac{N-N_1}{N} + K_j}
\]

It is not difficult to see that the outcomes $j, k$ cannot in general be interchanged without also changing the resulting conditional probability, which would violate (5b).

A more promising alternative is to embed our inductive logic in a hierarchical model with a prior probability measure over the parameters $\beta_{jk}$. With increasing information the posterior may assign higher probability to certain analogy parameters. Thus the analogy parameters become revisable, and for particular priors the analogy influence may vanish in the limit. Of course, in a hierarchical model some of the Carnapian spirit of an axiomatic approach to inductive inference might get lost.

5. Concluding Remarks

There are several coherent ways in which analogical reasoning can become a part of inductive logic. Here we have worked out the axiomatic foundation for one that is suitable for strong analogy influences and discussed some its limitations and advantages. Let me end with two remarks—one on the kind of symmetry we have used, and the other one on a possible extension of the present theory to continuous random variables.

Johnson’s and Carnap’s approach to inductive inference provides, in a certain sense, an alternative to de Finetti’s theorem on exchangeable sequences, since Johnson’s sufficientess postulate and the requirement that for all $i, j$

\[
P[X_{n+1} = i, X_{n+2} = j|X_1, \ldots, X_n] = P[X_{n+1} = j, X_{n+2} = i|X_1, \ldots, X_n]
\]

implies exchangeability (Kuipers 1978; Fortini et al. 2000). Condition (25) is similar to (5). Both requirements can be thought of as "local symmetries", which place restrictions on conditional probabilities given the evidence. In contrast, exchangeability is a constraint on the unconditional prior probability. In our case, using local symmetries as in (5) considerably simplifies the formulation of the symmetry relevant for analogy.

Another feature of our inductive logic is that it only applies to discrete probability spaces. It would be desirable to extend it to more general situations, for instance to random variables that can take on a continuum of values. In such probability spaces there often is a natural sense of analogy based on a metric between outcomes. Zabell (2011) discusses analogical inference for this setting based on
well known symmetry assumptions. For instance, in the case of an exchangeable prior the posterior expresses the belief that an observation similar to the observed relative frequency will be observed. The inductive logic introduced by [Skyrms (1993b)], which is based on the Blackwell-Mac Queen urn process [Blackwell and MacQueen (1973)], could be utilized for an analysis of analogical inference in more general probability spaces.

ACKNOWLEDGEMENTS

I would like to thank Hannes Leitgeb, Jeff Paris, Brian Skyrms, Marta Sznaider, Greg Wheeler, and Sandy Zabell for helpful comments.

REFERENCES


DEPARTMENT OF LOGIC & PHILOSOPHY OF SCIENCE, UC IRVINE

E-mail address: shuttegg@uci.edu