*Statistics, Probability and Game Theory*
IMS Lecture Notes – Monograph Series (1996) Volume 30

# CARNAPIAN INDUCTIVE LOGIC AND BAYESIAN STATISTICS

BRIAN SKYRMS
*University of California, Irvine*

**Abstract**

In his last work on inductive logic, Carnap listed three outstanding problems for future research: inductive logic for a value continuum, inductive logic sensitive to "analogy by proximity" and inductive logic sensitive to "analogy by similarity". The first problem is completely solved by work of Ferguson and Blackwell and MacQueen. Bayesian statistics also throws light on Carnap's two problems of analogy.

**1. Introduction.** In 1941, at the age of 50, Rudolf Carnap embarked on the project of developing inductive logic. In the early days of the Vienna Circle the logical positivists, heavily influenced by Hilbert, looked to deductive logic and the axiomatic method to analyze the scientific method. Over the years Carnap, Hempel, Reichenbach and others came to the conclusion that important aspects of scientific method were irreducibly inductive. The analysis of inductive inference thus became central to the development of scientific philosophy.

While Reichenbach focused on relative frequency and the consistency of statistical estimators, Carnap investigated probability as rational degree of belief. He started with Bayes and progressed to Laplace. By the end of his life, he had moved from a logical conception of probability inspired by Keynes and a unique inductive rule based on a flat prior for an IID process to a subjective conception of probability and a class of inductive rules corresponding to the Dirichlet priors.

Carnap hoped that after his death coworkers would carry forward the construction of inductive logic. In his last (posthumous) work (1980) he lists three major tasks for future research. One is the construction of confirmation functions for the case where the outcome can take on a continuum of possible values. The others have to do with the construction of confirmation functions sensitive to two kinds of analogy, which he calls "analogy by proximity" and "analogy by similarity".

Within three years of his death, Carnap's first problem had been completely solved in a way quite consonant with Carnapian techniques, by Ferguson(1973) and Blackwell and MacQueen (1973). Carnap's "coworkers" have turned out, in large measure, to be Bayesian statisticians rather than philosophers. In addition to the solution to the problem of a value continuum, Bayesian statistics can throw some light on Carnap's two problems of analogy.

**2. Carnap's Continua.** Suppose that we have an exhaustive family of $k$ mutually exclusive categories, and a sample of size $N$ of which $n$ are of category $F$. Carnap (1950) originally proposed the following inductive rule, C*, to give the probability that a new sampled individual, $a$, would be in $F$ on the basis of the given sample evidence, $e$.

$$pr(Fa|e) = \frac{1+n}{k+N}$$

On the basis of no sample evidence, each category gets equal probability of $1/k$. As the sample grows larger, the effect of the initial equiprobable assignment shrinks and the probability attaching to a category approaches the empirical average in the sample, $n/N$. Soon Carnap (1952) shifted from this method to a class of inductive methods of which it is a member, the $\lambda$-continuum of inductive methods.

$$pr(Fa|e) = \frac{\lambda+n}{\lambda k+N}$$

Here again, we have initial equiprobability of categories and predominance of the empirical average in the limit with the parameter, $\lambda(\lambda > 0)$, controlling the rate at which the sample evidence swamps the prior probabilities. In his posthumous (1980) paper, Carnap introduced the more general $\lambda - \gamma$ continuum:

$$pr(F_i a|e) = \frac{\lambda\gamma_i + n}{\lambda + n}$$

The new parameters, $\gamma_i > 0$, allow unequal *a priori* probabilities for different categories. For Carnap these are intended to reflect different "logical widths" of the categories. The parameter, $\lambda$, again determines how quickly the empirical average swamps the prior probability of an outcome. One could equivalently formulate Carnap's $\lambda - \gamma$ rules as follows. Take any $k$ positive numbers, $b_1, \ldots, b_k$, and let the rule be:

$$pr(F_i a|e) = \frac{b_i + n}{\sum_j b_j + N}$$

where $\lambda = \sum_j b_j$ and $\gamma_i = b_i / \lambda$.)

We can think about the problem addressed by Carnap's inductive rules in the following way. The experiment is represented by a discrete random variable, taking as possible numerical values the integers 1 through $k$, according to whether the experimental result is $F_1, \ldots, F_k$. This experimental result generates a measurable space, $S = < W, A >$, where the points in $W$

are the $k$ possible outcomes and the propositions (measurable sets) in $A$ are gotten by closing the atoms under Boolean combination.

Induction takes place when the experiment is iterated and a general analysis does not place any finite upper bound on the number of possible iterations. Thus we are led to consider an infinite sequence of such random variables, indexed by the positive integers. The relevant measurable space is a product of an infinite number of copies of the one shot probability space, $S_1 \times S_2 \times S_3 \times \ldots$ The points in the infinite product space are infinite sequences, $w_1, w_2, \ldots$, of integers from 1 through $k$. In this setting we can give a fairly general definition of an inductive rule.

An *inductive rule*, R, takes as input any finite initial sequence of results, $< w_1, \ldots, w_j >$ and any proposition about the experimental outcome the next time around, $a_{j+1}$ in $A_{j+1}$ and outputs a numerical prediction in [0,1] such that: (i) for any finite outcome sequence, $< w_1, \ldots, w_j >$, the rule gives a probability, $\mathrm{R}(< w_1, \ldots, w_j >, \cdot)$ on the space for the next moment of time, $< W_{j+1}, A_{j+1} >$ and (ii) For every proposition about the next outcome, $a_{j+1}$, $\mathrm{R}(\cdot, a_{j+1})$ is measurable on the space of histories up until then, $S_1 \times \cdots \times S_j$.

Note on (i): This includes empty sequences, so the rule must specify a prior probability on $\mathrm{S}_1$. Note on (ii): This technical regularity requirement is automatically satisfied in the finite case considered in this section, but is required for a definition which will generalize.

Then an *inductive rule* determines a unique probability on the infinite product space, $S_1 \times S_2 \times \cdots$, according to which the probabilities given by the inductive rules are conditional probabilities (Neveu, 1965, Ch. V). Thus, the properties of an inductive rule can be studied at either the "operational" level of a predictive rule or at the "metaphysical" level of a prior probability on an infinite product space.

The rules in Carnap's $\lambda - \gamma$ continuum are all inductive rules in the foregoing sense. For each of these rules the corresponding probability on the infinite product space is *exchangeable*: that is, invariant under finite permutations of outcomes. By de Finetti's theorem, any such probability has a unique representation as a mixture of probabilities which make the outcome random variables independent and identically distributed. That is to say that the Carnapian inductive logician behaves as if she were multinomial sampling. The examples of throwing a die with unknown bias or sampling with replacement from an urn of unknown composition are canonical.

**3. The Classical Bayesian Parametric View of Carnap's $\lambda - \gamma$ Continuum.** One way of thinking about Carnap's rules is to take this representation at face value. There is a fixed statistical model of the chance mechanism with subjective prior uncertainty as to the true values of the parameters of the model. There is a chance setup (the die and the throwing mechanism) generating a sequence of independent and identically distributed random variables, but the distribution of the random variables (the bias of the die) is initially unknown. Inductive inference depends on this prior uncertainty via Bayes' theorem. The goal of inductive inference is to learn the true bias of the coin. This is the parametric Bayesian conceptualization of the problem.

From this point of view, Carnap's postulate that inductive methods should satisfy the *Reichenbach Axiom* – that the probabilities given by the inductive rule converge to the empirical average as the number of trials goes to infinity – is well-motivated. By the strong law of large numbers, this is just the requirement that with chance equal to one, the inductive rule learns in the limit true bias of the die, no matter what the true bias is. That is to say that Carnap's methods correspond to priors on the bias which are Bayesian consistent with respect to the multinomial statistical model.

Not every consistent prior for multinomial sampling generates an inductive rule which is a member of Carnap's $\lambda-\gamma$ continuum. Carnap's rules correspond to the class of natural conjugate priors for multinomial sampling: the Dirichlet priors. As the predictive rules which are generated by Dirichlet priors, Carnap's $\lambda-\gamma$ methods are well-known to Bayesian statisticians, Good (1965), DeGroot (1970) Ch. 5. Carnap's (1942) $\lambda$ continuum is generated by the symmetric Dirichlet priors and his (1950) confirmation function, c*, by the uniform prior.

**4. The Subjective Bayesian View of Carnap's $\lambda-\gamma$ continuum.** For pure subjective Bayesians talk of objective chances is strictly meaningless and the multinomial statistical model discussed in the last section is only an artifact of the de Finetti representation. Exchangeability is consistent with degrees of belief which give rise to inductive rules outside Carnap's continuum. Is there any subjective characterization of the $\lambda-\gamma$ continuum?

There is. It is due to the Cambridge logician W. E. Johnson. Johnson introduced the concept of exchangeability or symmetry in 1924 (before de Finetti.) We defined exchangeability in terms of the infinite product space induced by an inductive rule, as invariance under finite permutations of trials. There is an equivalent formulation in terms of the inductive rule itself. That is that the vector of frequency counts of the possible outcomes is sufficient to determine the probabilities for the next trial: $R(<w_1,\ldots,w_j>,\cdot) = R(<w'_1,\ldots,w'_j>,\cdot)$ if the outcome sequences $<w_1,\ldots,w_j>$ and $<w'_1,\ldots,w'_j>$ have the same frequency counts. For example, in rolling a die, the outcome sequences 1234561 and 1135264 would lead to the same probabilities for the eighth trial; order is presumed not to be relevant.

Johnson was led to postulate a stronger kind of sufficiency. That is that (i) The probability of an outcome on the next trial should only depend on the frequency that *it* has occurred in the preceding trials (for each fixed number of trials) and not on the relative frequencies of the other trials, and

(ii) The dependence should be the same for all categories. In the example of the die, (i) says that the probability of a one on the eighth trial given initial outcome sequences 1213654 and 1155555 should be the same; (ii) adds that this probability should be equal to the probability of a 2 on the eighth trial given the initial sequence 1253266. Taken together (i) and (ii) are Johnson's *sufficientness postulate.*

If we have (1) exchangeability but not independence (2) sufficientness in Johnson's sense (3) the number of categories is at least three and (4) the relevant conditional probabilities are all well-defined, then we get Carnap's 1952 $\lambda$-continuum of inductive methods. If Johnson's (ii) is dropped from the foregoing, we get Carnap's $\lambda - \gamma$ continuum. (Zabell, 1982) Thus, from a purely subjective point of view, Carnap's continua correspond to strong symmetries in a predictor's degrees of belief.

**5. Blackwell-MacQueen Inductive Rules.** Carnap was interested in developing inductive logic to the point where it could make contact with mathematical physics. He saw the first important step to be to generalize his methods so that they could deal with the case where the outcome of an experiment could take on a continuum of values. We will accordingly change our canonical example from that of a die to that of a wheel of fortune with unit circumference. Repeated spins of the wheel produce as (ideal) outcomes, real numbers in the interval [0,1). We equip this outcome space with the metric corresponding to the shortest distance measured around the circumference of the circle. We will take as our problem, the specification of an inductive rule in this setting. The techniques which solve this problem, however, will apply very generally.

One could approach this problem by thinking about how a Carnapian in possession only of Carnap's $\lambda - \gamma$ continuum for finite numbers of outcomes could approximate a solution. The natural thing to try might be to partition the unit interval into a finite number of subintervals, and apply a method from the $\lambda - \gamma$ continuum taking the elements of the partition as outcomes. Finer and finer partitions might be thought of as giving better and better approximations. Recall the reformulation of Carnap's $\lambda - \gamma$ continuum of section 2. Take any $k$ positive numbers, $b_1, \ldots, b_k$, and let the rule be:

$$pr(F_i a|e) = \frac{b_i + n}{\sum_j b_j + N}$$

where $\lambda = \sum_j b_j$ and $\gamma_i = b_i/\lambda$. Thus, for a partition into $k$ subintervals, we have a class of Carnapian inductive rules whose members are specified by $k$ parameters, $b_1, \ldots, b_k$.

In the case of a continuum of possible outcomes, the appropriate parameter will be a nonnegative bounded measure, $\alpha$, defined on the Borel algebra of the unit interval. As a generalization of Carnap's condition that the $b_i$s should be positive in the finite case, we will require that the measure, $\alpha$, be absolutely continuous with respect to Lebesgue measure on the unit interval. For any Borel set, $O$, in the unit interval and evidence, $e$, consisting of $n$ points in $O$ in $N$ trials, we will take the inductive rule with parameter, $\alpha$, to give the probability of an outcome in $O$ in the next trial as:

$$pr(O|e) = \frac{\alpha(O) + n}{\alpha([0,1)) + N}$$

We will call this the Blackwell-MacQueen inductive rule, since Blackwell and MacQueen used this formula in generalizing Polya urn models for what philosophers know as Carnap's $\lambda - \gamma$ inductive rules, to more general Polya urn models for non-parametric Bayesian inference.

It can be seen that the Blackwell-MacQueen rules are consistent with Carnap's $\lambda - \gamma$ continuum in the following sense: For a Blackwell-MacQueen rule and a "coarse-graining" of outcomes according to which member of a finite partition they fall into, the induced inductive rule for the finite partition is a member of Carnap's $\lambda - \gamma$ continuum. For a simple illustration, let the parameter, $\alpha$, for the Blackwell-MacQueen rule just be Lebesgue measure. In particular, if $I$ is a subinterval of [0,1), then $\alpha(I)$ is just the length of $I$. Partition the unit interval into $k$ equal subintervals. Then for some fixed one of these subintervals, $I$, let $n$ be the number of sample points in $I$ and $N$ be the total number of sample points. Then the Blackwell-MacQueen rule gives the probability that the next point will fall into $I$ as:

$$Pr(I|e) = \frac{\frac{1}{k} + n}{1 + N}$$

which is a member of Carnap's $\lambda$-continuum. (Notice that as the partitions get finer the value of $\lambda$ gets proportionately smaller to preserve consistency with the Blackwell-MacQueen rule.) It should be clear that this class of inductive rules is the natural generalization of Carnapian rules to problems where the outcomes can be represented as real numbers in the unit interval.

Blackwell-MacQueen rules are *Inductive Rules* in the sense made precise in section 2. Thus they induce a probability measure on the infinite product space. This probability measure makes the random variables which represent the experimental outcomes exchangeable. By de Finetti's theorem, it can be represented as a mixture of probabilities which make the trials independent and identically distributed. From a classical Bayesian viewpoint, the mixing measure is the ignorance prior over the true chances governing the IID process. The IID probabilities correspond to a distribution on [0,1). The prior corresponds to a distribution on the distributions on [0,1).

**6. The Classical Bayesian Non-Parametric View: Ferguson Distributions.** One might try to generalize Carnap's inductive methods in a different way – by working at the level of the de Finetti priors rather than at the level of inductive rules. As noted in section 3, the members of Carnap's $\lambda - \gamma$ continuum are just those rules which arise from multinomial sampling with Dirichlet priors. The natural generalization of a Dirichlet prior is a Ferguson distribution (called by Ferguson, and also known as, the Dirichlet process). A *Ferguson distribution* with parameter $\alpha$ is a distribution which for every k member partition of the interval, $P_1, \ldots, P_k$, is distributed as Dirichlet with parameters, $\alpha(P_1), \ldots, \alpha(P_n)$. A Ferguson distribution is a distribution over random chance distributions for a random chance probability, $p$. The parameter, $\alpha$, of the Ferguson distribution can be any finite, non-null measure on [0,1). Thus for any finite measurable partition, $\{P_1, \ldots, P_n\}$ of [0,1), there is a corresponding vector of numbers, $< \alpha_1, \ldots, \alpha_n >$ where $\alpha_i = \alpha(P_i)$. The requirement is then that for any such partition, the random chance probability vector for members of the partition, $< p(P_1), \ldots, p(P_n) >$, has a Dirichlet distribution with parameter $< \alpha_1, \ldots, \alpha_n >$.

The good news for Carnap's program is that both roads lead to the same place. The de Finetti prior distribution corresponding to the probability on the infinite product space induced by a Blackwell-MacQueen inductive rule with parameter $\alpha$ is a Ferguson distribution with parameter $\alpha$. In fact, the paper of Blackwell and MacQueen was written to give a simple proof of the existence of Ferguson distributions.

Furthermore, just as the class of Dirichlet probabilities is closed under multinomial sampling, the class of Ferguson distributions is closed under IID sampling of points in [0,1). Consider a finite sample sequence consisting of data points $x_1, \ldots, x_n$, and let $\delta_1, \ldots, \delta_n$ be probability measures giving mass one to the points $x_1, \ldots, x_n$, respectively. Let the prior be given by a Ferguson distribution with parameter $\alpha$. Then conditioning on the data points takes one to a posterior which is a Ferguson distribution with parameter $\alpha'$, where:

$$\alpha' = \alpha + \sum_{i=1}^{n} \delta_i$$

From the classical Bayesian viewpoint, Ferguson distributions are natural conjugate priors for this non-parametric sampling problem in just the same way that Dirichlet priors are natural conjugate priors for multinomial sampling.

Blackwell-MacQueen Inductive Rules satisfy a version of *Reichenbach's axiom*. Freedman (1963), Fabius (1964), Diaconis and Freedman (1986). As evidence accumulates the probability for the next trial is a weighted average of the a priori probability and the empirical relative frequency probability, with all weight concentrating on the empirical relative frequency probability as the number of data points goes to infinity. Can we say in the classical

Bayesian setting that Blackwell-MacQueen inductive rules will with chance one - learn the true chances?

In this case "learning the true chances" is not just learning the values of a finite number of parameters as in the multinomial case, but rather learning the true chance probability on [0,1). Thus we need a sense of convergence for the space of all probability measures on [0,1). A sequence of probabilities measures, $\mu_n$, converges weak* to measure $\mu$ iff for every bounded continuous function of [0,1), its expectation with respect to the measures $\mu_n$ converges to its expectation with respect to $\mu$.

A prior (or alternatively the corresponding inductive rule) is Bayesian consistent with respect to a chance probability $\mu$, if with probability one in $\mu$, the posterior under IID (in $\mu$) sampling will converge weak* to $\mu$. A prior is Bayesian consistent if it is Bayesian consistent for all possible chance probabilities. In the multinomial case, all the rules of Carnap's $\lambda - \gamma$ continuum (alternatively, the corresponding Dirichlet priors) are Bayesian consistent.

The Blackwell-MacQueen rules as defined in section 5 are also Bayesian consistent in this sense. It should be noted that this is a consequence of a restriction that I put on the parameter, $\alpha$, of those rules; that $\alpha$ be absolutely continuous with respect to Lebesgue measure on the unit interval. I take this to be a natural generalization of Carnap's requirement of regularity (or strict coherence) in the finite case. Likewise, Bayesian consistency is a natural generalization of Reichenbach's axiom.

**7. Subjective Bayesian Analysis of Blackwell-MacQueen Inductive Rules.** In this section we return to the subjective point of view. The notion of random variables which are independent and identically distributed according to the true unknown chances gives way to the subjective symmetry of exchangeability. In the case where the random variables take on a finite number of values, an additional symmetry assumption – W.E. Johnson's sufficientness postulate – (together with a few other technical assumptions) get us Carnap's $\lambda$-$\gamma$ continuum. In the case under consideration, where our random variables can take on a continuum of values in [0,1), we have the analogous result. That is – roughly speaking – that Exchangeability + Sufficientness gives the Blackwell-MacQueen inductive rules.

Let us first review the case of random variables taking a finite number of values discussed in Section 4 in a little more detail. This is based on Zabell (1982). Suppose that we an infinite sequence of random variables, $X_1, X_2, \ldots$, each taking values in a finite set $O = \{1, ..., k\}$. And suppose that the number of possible outcome values, $k$, is at least three. We consider a number of conditions on our probabilities: (1) *Exchangeability*: This guarantees by de Finetti's theorem that our (degree of belief) probability can be uniquely represented as a mixture of probabilities which make the outcomes independent and identically distributed. (2) *Non-Independence*: If our beliefs make the outcomes independent then we will not learn from experience. (3) *Strict Coherence*: The probability of any finite outcome sequence is non-zero. This is a kind of open-mindedness condition. It guarantees that all the conditional probabilities in the next condition are well-defined. (4) *Generalized Sufficientness* Let $n_i$ be the frequency count of outcomes in category $i$ in the trials $X_1, \ldots, X_n$. Then:

$$Pr(X_{N+1} = i | X_1, ..., X_N) = f_i(n_i)$$

That is, for each category, $i$, the probability of the next outcome being in that category is a function only of the frequency count for that category in the preceding sequence of observations. (Johnson assumed that these functions would be the same for all categories, but that is not assumed here.) Zabell shows that under assumptions 1-4, the predictive conditional probabilities:

$$Pr(X_{N+1} = i | X_1, ..., X_N)$$

are just those given by the inductive rules of Carnap's $\lambda$-$\gamma$ continuum.

Now consider the case of the wheel of fortune, where an infinite sequence of random variables takes on values in the interval [0,1). We assume (1) *Exchangeability* and (2) *Nonindependence* as before. As (3*) *Regularity* we assume that for any measurable set, $B$, which has non-zero Lebesgue measure and for any finite sequence of observations, $X_1, \ldots, X_n$:

$$Pr(X_{N+1} \in B_i | X_1, ..., X_N) = f_i(n_i)$$

As (4*) *Generalized Sufficientness* we require that for any measurable set, $B_i$, and any finite sequence of observations, the probability that the next observation fall in $B_i$ is a function only of the count, $n_i$, of previous observations that have fallen in $B_i$:

$$Pr(X_{N+1} \in B_i | X_1, ..., X_N) = f_i(n_i)$$

If 1, 2, 3*, and 4* are fulfilled, then our predictive conditional probabilities are Blackwell-MacQueen inductive rules.

This is almost immediate from the finite case. Here is a sketch of an argument. If 1,2,3* and 4* are fulfilled then for any finite partition of the unit interval whose members have non-zero Lebesgue measure, 1,2,3,4 hold. Then for any such partition with three or more members, Zabell's version of W.E. Johnson's result holds. Thus the probabilities of falling in members of the partition update according to Carnap $\lambda$-$\gamma$ rules. This must also be the case for even partitions of only two members, which can be seen by subdividing them into partitions of four members. Then the de Finetti priors for all these partitions must be distributed as Dirichlet, and the de Finetti prior for [0,1) must be a Ferguson distribution. Thus the inductive rules induced by the prior must be Blackwell-MacQueen inductive rules.

Carnap would not have thought that inductive logic for a value continuum ended with the class of Blackwell-MacQueen inductive rules but rather that it started there. These rules are just right for those settings which characterize them – where 1,2,3* and 4* hold. But in some contexts they will not all hold.

**8. Analogy by Similarity.** Ferguson (1974) considers 4* as a drawback of the use of Dirichlet Processes (Ferguson distributions):

> One would like to have a prior distribution for $P$ with the property that if $X$ is a sample from $P$ and $X = x$, then the posterior guess at $P$ gives more weight to values close to $x$ than the prior guess at $P$ does. For the Dirichlet process prior, the posterior guess at $P$ gives more weight to the point $x$ itself, but it treats all other points equally. In particular, the posterior guess at $P$ actually gives less weight to points near $x$ but not equal to $x$. (Ferguson, 1974 p.622). (Note that in Ferguson's characterization 3 on that page, T1, T2, and T3 are eliminated by my nonindependence and regularity conditions.)

Carnap makes exactly the same point in his last writings on inductive logic under the heading of the problem of *analogy by similarity*. Carnap raises the question in the context of his current system where there are only a finite number of possible outcomes.

Where it is desirable that sample x gives more weight to values close to x, W. E. Johnson's Sufficientness postulate must be given up. Sufficientness is just the statement that we do not have analogy by similarity. Thus we must move outside Carnap's continuum of inductive methods. The most conservative move outside Carnap's $\lambda$-$\gamma$ continuum would be to consider finite mixtures of methods that are themselves in the $\lambda$-$\gamma$ continuum. One could think of this as putting a "hyperprior" probability on a finite number of metahypotheses as to the values of the $\lambda$ and $\gamma_i$ hyperparameters. Conditional on each metahypothesis, one calculates the predictive probabilities according to the Carnapian method specified by that metahypothesis. The probabilities of the metahypotheses are updated using Bayes' theorem. We will call these *hyperCarnapian* Methods.

The hierarchical gloss, however, is inessential. The model just described is mathematically equivalent to using a prior on the multinomial parameters which is not Dirichlet but rather a finite mixture of Dirichlet priors. It is evident that if the number of Carnapian methods in the mixture is not too great, the computational tractability of Carnapian methods is not severely compromised. Furthermore, Bayesian consistency is retained. Finite mixtures of Dirichlet priors are consistent (Diaconis and Freedman, 1986).

Furthermore, they can exhibit the kind of analogy by similarity that Carnap wished to model (Skyrms, 1993a). We can illustrate this by means of a simple example: A wheel of fortune is divided into four quadrants: N, E, S, W. There are four "metahypotheses" which are initially equiprobable. Each requires updating by a different Carnapian rule as indicated in the following table:

| | N | E | S | W |
|---|---|---|---|---|
| **H1:** | $\frac{5+n}{10+N}$ | $\frac{2+n}{10+N}$ | $\frac{1+n}{10+N}$ | $\frac{2+n}{10+N}$ |
| **H2:** | $\frac{2+n}{10+N}$ | $\frac{5+n}{10+N}$ | $\frac{2+n}{10+N}$ | $\frac{1+n}{10+N}$ |
| **H3:** | $\frac{1+n}{10+N}$ | $\frac{2+n}{10+N}$ | $\frac{5+n}{10+N}$ | $\frac{2+n}{10+N}$ |
| **H4:** | $\frac{2+n}{10+N}$ | $\frac{1+n}{10+N}$ | $\frac{2+n}{10+N}$ | $\frac{5+n}{10+N}$ |

where $n$ is the number of successes in $N$ trials.

Since the hypotheses are initially equiprobable, the possible outcomes, N, E, S, W, are also initially equiprobable. Suppose that we have one trial whose outcome is N. Then updating the probabilities of our hypotheses by Bayes' Theorem, the probabilities of H1, H2, H3, H4 respectively become .5, .2, .1, .2. Applying the Carnapian rule of each hypothesis and mixing with the new weights gives probabilities:

$$\begin{aligned} pr(\mathrm{N}) &= 44/110 \\ pr(\mathrm{E}) &= 24/110 \\ pr(\mathrm{S}) &= 18/110 \\ pr(\mathrm{W}) &= 24/110 \end{aligned}$$

The outcome, N, has affected the probabilities of the non-outcomes E, S, W differentially even though each Carnapian rule treats them the same. The outcome N has reduced the probability of the distant outcome, S, much more than that of the close outcomes, E and W, just as Carnap thought it should.

In a certain sense, this is the only solution to Carnap's problem. Carnap clearly was interested in sensitivity to analogy by similarity in the presence of exchangeability. For this problem we are, in effect, restricted to choosing a prior over IID processes. But every prior can be approximated arbitrarily well by finite mixtures of Dirichlet priors. (Diaconis and Freedman, 1986). HyperCarnapian inductive methods are the general solution to Carnap's problem of analogy by similarity.

There is an investigation of HyperCarnapian inductive methods at the level of Blackwell-MacQueen rules in Antoniak (1972). Of course, analogy by similarity may also be important in other domains where exchangeability fails to hold.

**9. Analogy by Proximity.** Carnap also discussed a different kind of analogy which his methods could not represent as the problem of *analogy by proximity*. This is the problem of taking into account temporal patterns in the data, for instance in inference about Markov chains. How should one treat Markov chains in the spirit of Carnap's original inductive logic? Carnap never addressed this question in his published work. I believe that the question would fall somewhere on Carnap's agenda for the future development of inductive logic but it was not one which he actively tried to answer. In response to an inquiry, John Kemeny replied that Carnap never discussed Markov chains with him, and in fact that when he worked as Carnap's research assistant he had not yet heard of a Markov chain. Richard Jeffrey and Haim Gaifman, who also worked with Carnap, confirm that Carnap did not actively investigate this problem.

There is, however, a natural treatment of inductive logic for finite Markov chains, which fits neatly into Carnap's program. It is put forward by Theo Kuipers in Kuipers (1988). The leading idea is this: Carnap already has an inductive logic suitable for sampling from an urn with replacement. Just apply this inductive logic to the natural urn model of a Markov chain, under the assumption that transitions originating in one state give us no information about transitions originating in a different state. Parametric Bayesians, such as Martin (1967), operating in a somewhat different tradition, have followed the same path. We have a Markov chain with finite state space and known initial state, but unknown transition probabilities and we observe the successive states. Our inductive problem is to predict future states from history. The Carnapian solution is to apply Carnapian inductive rules to transition probabilities.

From a parametric Bayesian point of view, we can raise Reichenbach's question of *consistency*. A state of a Markov chain is called *recurrent* if the probability that it is visited an infinite number of times is one. The chain is recurrent if all its states are recurrent. Carnapian inductive logic for Markov chains is consistent for recurrent Markov chains. Here is a quick sketch of a proof. Suppose that the true state of nature is a recurrent Markov chain (with a finite number of states). Then the set of sample sequences in which some state does not recur has probability zero. Delete these sequences and restrict the chance measure to get a new probability space. In this space define the random variables $f[ni]$ as having the value $j$ if the $n$th occurrence of state $i$ is followed by state $j$. For each fixed $i$, the sequence $f(1i), f(2i), \ldots$, is an infinite sequence of independent and identically distributed random variables so the strong law of large numbers applies. This means that the limiting transition relative frequencies from $i$ to $j$ equal the true transition probabilities with chance 1. Thus if the true state of nature is a recurrent Markov chain, then there are only a finite number of ways in which Carnapian inductive logic for Markov chains can fail to learn the true transition matrix, and each of these has a chance of zero.

From the subjective Bayesian point of view, the treatment is again parallel to that of Carnapian inductive methods, as long as we have a recurrent stochastic process. In dealing with Markov chains we do not have exchangeability, but rather a weaker kind of symmetry condition which can play a role with respect to Markov processes analogous to that played by exchangeability with respect to Bernoulli processes. The analysis is developed in Freedman (1962), de Finetti (1974), Diaconis and Freedman (1980). A stochastic process is *Markov exchangeable* if the vector of initial state and transition counts is a sufficient statistic for all finite sequences of given length generated by the process. That is to say that sequences of the same length having the same transition counts and the same initial state, are equiprobable. Markov exchangeability, like ordinary exchangeability, can also be given an equivalent formulation in terms of invariance (Diaconis and Freedman, 1980). A primitive block-switch transformation of a sequence takes two disjoint blocks of the sequence with the same starting and ending states and switches them. A block switch transformation is the composition of a finite number of primitive block switch transformations. A probability is then Markov exchangeable just in case it is invariant under all block switch transformations. Diaconis and Freedman (1980) show that *recurrent* stochastic processes of this type which are Markov exchangeable have a unique representation as a mixture of Markov chains.

Zabell (1995) shows that the subjective condition which guarantees that the de Finetti prior for a recurrent Markov exchangeable stochastic process is of the type that induces Carnapian inductive logic for Markov chains is again a form of W. E. Johnson's sufficientness postulate that $Pr(S_j|S_i)$ depend only on $i, j, N[S_j, S_i]$ and $\sum_m N[S_m, S_i]$, where $N[S_j, S_i]$ is the transition count. Notice that this automatically gets us the independence that Martin and Kuipers assume. If our beliefs satisfy the postulate then transition counts from one state are not taken as giving any evidence about transitions from another state. Since a recurrent Markov exchangeable process is a mixture of recurrent Markov chains, for each $i$ the embedded process $f(1i), f(2i), \ldots$ discussed above is mixture of IID processes and thus exchangeable. The sufficientness postulate for Markov chains gives the original sufficientness postulate for these embedded processes, and the application of the original sufficientness argument (Zabell, 1982) to them gives the desired result.

As before, W. E. Johnson's sufficientness postulate has no *logical* status. It merely serves to characterize cases in which degrees of belief have certain interesting and computationally tractable symmetries.

**10. Conclusion.** At the beginning of his investigations his investigations in inductive logic Carnap hoped that all of scientific inference could be based on one inductive rule. That rule would have a necessary status. The logical character of inductive logic would derive from the logical status of this rule and of the prior which led to it. He soon started down the road that leads from Keynes to de Finetti. Ever larger classes of inductive rules were seen as part of inductive logic. In the end he saw a need for further expansion, in particular to deal with the three problems discussed here, and in general to arrive at an adequate treatment of scientific inference.

In posthumous Basic System (1971), (1980), Carnap realizes somewhat reluctantly that he has become a subjective Bayesian. The logic in inductive logic is now the logic of coherence. From this point of view, it is not just that Bayesian statistics has useful things to contribute to inductive logic. From the most general convergence theorems, such as that of Blackwell and Dubins (1962), to the analysis of particular problems – Bayesian statistics *is* inductive logic.

**Acknowledgement**. I would like to thank Persi Diaconis, Haim Gaifman, Richard Jeffrey, John Kemeny, Jim MacQueen, and Sandy Zabell for valuable discussions.

## REFERENCES

ANTONIAK, C. (1974). Mixtures of Dirichlet Processes with Applications to Bayesian Nonparametric Problems. *Annals of Statistics* **2** 1152-1174.

BLACKWELL, D. AND DUBINS, L. (1962). Merging of Opinions with Increasing Information. *Annals of Mathematical Statistics*. **33** 882-886.

BLACKWELL, D. AND MACQUEEN, J. B. (1973). Ferguson distributions via Polya Urn Schemes. *Annals of Statistics* **1** 353-355.

CARNAP, R. (1950). *Logical Foundations of Probability.* Univ. Chicago Press. (second edition 1962).

CARNAP, R. (1952). *The Continuum of Inductive Methods*. Univ. Chicago Press.

CARNAP, R. (1971). A Basic System of Inductive Logic, Part 1. In *Studies in Inductive Logic and Probability: Volume I*. ed. R. Carnap and R. Jeffrey 35-165.

CARNAP, R. (1980). A Basic System of Inductive Logic, Part 2. In *Studies in Inductive Logic and Probability vol. II*. ed. R. C. Jeffrey. Berkeley: Univ. California Press.

COSTANTINI, D. (1983). Analogy by Similarity. *Erkenntnis* **20** 103-114.

DE FINETTI, B. (1974). *Probability, Induction and Statistics*. Wiley, New York.

DEGROOT, M. (1970). *Optimal Statistical Decisions*. McGraw Hill: New York.

DIACONIS, P. AND FREEDMAN, D. (1980). De Finetti's Theorem for Markov Chains. *Annals of Probability*. **8** 115-130.

DIACONIS, P. AND FREEDMAN, D. (1986). On the Consistency of Bayes Estimates. *Annals of Statistics*. **14** 1-26.

FABIUS, J. (1964). Asymptotic Behavior of Bayes Estimates. *Annals of Mathematical Statistics*. **35** 846-856.

FERGUSON, T. (1973). A Bayesian Analysis of Some Non-Parametric Problems. *Annals of Statistics* **1** 209-230.

FERGUSON, T. (1974). Prior Distributions on Spaces of Probability Measures. *Annals of Statistics* **2** 615-629.

FREEDMAN, D. (1962). Mixtures of Markov Processes. *Annals of Mathematical Statistics* **2** 615-629.

FREEDMAN, D. (1963). On the asymptotic behavior of Bayes' estimates in the discrete case. *Annals of Mathematical Statistics* **24** 1386-1403.

GOOD, I. J. (1965). *The Estimation of Probabilities: An Essay on Modern Bayesian Methods*. MIT Press, Cambridge, Massachusetts.

JOHNSON, W. E. (1924). *Logic, Part III: The Logical Foundations of Science*. Cambridge Univ. Press, Cambridge.

JOHNSON, W. E. (1932). Probability: the deductive and inductive problems. *Mind*. **49** 409-423.

KUIPERS, T. A. F. (1978). *Studies in Inductive Logic and Rational Expectation*. Reidel, Dordrecht.

KUIPERS, T. A. F. (1984a). Two Types of Inductive Analogy by Similarity.*Erkenntnis* **21** 6387.

KUIPERS, T. A. F. (1984b). Inductive Analogy in the Carnapian Spirit. In *PSA 1984* v.1 (ed. P. D. Asquith and P. Kitcher). Philosophy of Science Association, East Lansing, Michigan.

KUIPERS, T. A. F. (1988). Inductive Analogy by Similarity and Proximity. In *Analogical Reasoning* (D. H. Helman,ed.). Kluwer, Dordecht.

MARTIN, J. J. (1967). *Bayesian Decision Problems and Markov Chains*. New York: Wiley.

NEVEU, J. (1965). *Mathematical Foundations of the Theory of Probability*. tr. A. Feinstein. Holden-Day, San Francisco.

NIINILUOTO, I. (1981). Analogy and Inductive Logic. *Erkenntnis* **16** 1-34.

NIINILUOTO, I. (1988). Analogy and Similarity in Scientific Reasoning. In *Analogical Reasoning* (D. H. Helman,ed.). Reidel, Dordrecht.

PUTNAM, H. (1963). Degree of Confirmation and Inductive Logic. In *The Philosophy of Rudolf Carnap* ed. P. A. Schilpp. pp. 761-783. Open Court, LaSalle, Illinois. Reprinted in Putnam, H. (1975) *Mathematics, Matter and Method*: Philosophical Papers vol.1. Cambridge Univ. Press 270-292.

SKYRMS, B. (1991). Carnapian Inductive Logic for Markov Chains. *Erkenntnis* **35** 439-460.

SKYRMS, B. (1993a). Analogy by Similarity in HyperCarnapian Inductive Logic. In *Philosophical Problems of the Internal and External Worlds*. ed. Earman et al. Univ. Pittsburgh Press, Pittsburgh.

SKYRMS, B. (1993b). Carnapian Inductive Logic for a Value Continuum. In *The Philosophy of Science.* Midwest Studies in Philosophy, vol. 18 (H. Wettstein, ed.) 78-89. Univ. Notre Dame Press, South Bend, Indiana.

SPOHN, W. (1981). Analogy and Inductive Logic: A Note on Niiniluoto *Erkenntnis* **16** 35-52.

ZABELL, S. L. (1982). W. E. Johnson's "sufficientness" postulate. *Annals of Statistics* **10** 1091-1099.

ZABELL, S. L. (1992). Predicting the Unpredictable. *Synthese* **90** 205-232.

ZABELL, S. L. (1995). Characterizing Markov Exchangeable Sequences. *Journal of Theoretical Probability* **8** 175-178.

DEPARTMENT OF PHILOSOPHY
UNIVERSITY OF CALIFORNIA
IRVINE, CALIFORNIA 92717
bskyrms@benfranklin.hnet.uci.edu