Computers in Biology and Medicine 108 (2019) 382–399

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Bidirectional General Graphs for inference. Principles and implications for medicine

Barry Robson*

Ingine Inc. Viginia, USA
The Dirac Foundation, OxfordShire, UK

ARTICLE INFO

Keywords:
Inference net
Directed acyclic graph
Bidirectional general graph
Bayes' rule
Bayes net
Hyperbolic Dirac net
Clinical decision support

SYNOPSIS

Probabilistic inference methods require a more general and realistic description of the world as a Bidirectional General Graph (BGG). While in its original form the Bayes Net (BN) has been promoted as a predictive tool, it is more immediately a way of testing a hypothesis or model about interactions in a system usually considered on a causal basis. Once established, the model can be used in a predictive way, but the problem here is that for a traditional BN the hypotheses or models that can be formed are limited to the Directed Acyclic Graph (DAG) by definition. Three interrelated features are highlighted that represent deficiencies of the DAG which are corrected by conversion to a method based on a BGG: (i) lack of intrinsic representation of coherence by Bayes' rule, (ii) relatedly the need to consider interdependence in parent nodes, and (iii) the need for management of a property called recurrence. These deficiencies can represent large errors in absolute estimates of probabilities, and while relative and renormalized probabilities ameliorate that, they can often make much of a net superfluous through cancelations by division. The Hyperbolic Dirac Net (HDN) based on Dirac's quantum mechanics is a solution that led naturally to avoiding these deficiencies. It encodes bidirectional probabilities in an $\boldsymbol{h}$-complex value rediscovered by Dirac, i.e. with the imaginary number $\boldsymbol{h}$ such that $\boldsymbol{hh} = +1$. Properties of the HDN described previously are reviewed (though emphasis is on descriptions in familiar probability terms), the issue of recurrence is introduced, methods of construction are simplified, and the severity of the quantitative differences between BNs and analogous HDNs are exemplified. There is also discussion of how results compare with other approaches in practice.

1. Introduction and review

1.1. Background

The amount of medical information in digital form worldwide began its explosion into thousands of petabytes around 2009; it promised to increase the potential for data analytics and automated reasoning to support the industries of clinical practice, insured healthcare, and pharmaceutical development in the face of uncertainty and risk [1]. By "uncertainty and risk" is meant both in regard to (a) probabilities associated with statements about the world, reflecting the fact that we cannot always be certain about what they say, or that their scope of applicability is limited, and (b) *probabilistic inference*, i.e. the probabilistic predictions and assessments of risk made from collections of those statements [1,2]. Pearl's Bayes Net (BN) provides a well-known example using conditional probabilities [3]. The purpose of the present paper is to improve probabilistic inference and decision support in medicine by overcoming certain deficiencies in this kind of approach, drawing on principles and lessons from other disciplines as diverse as the financial industry, bioinformatics and theoretical physics.

1.2. Probability and risk

Many industries think of probabilistic inference as primarily concerned with risk and *risk management*. Not all of medicine is viewed that way, but it is has been presented as an important justification for medical information technology and it is particularly evident in the disciplines of evidence based medicine and clinical decision support [1]. Much of a book by Rebonato [2] on financial stress management is of interest to inference researchers in disciplines outside finance, including medicine. Rebonato's arguments are of particular interest here as he promotes a property of consistency between probabilities called *coherence* to combine with BN use that is, however, not *intrinsic* to the BN, although it is intrinsic to the related more general approach

*.
E-mail address: barryrobson@ingine.com.

https://doi.org/10.1016/j.compbiomed.2019.04.005
Received 18 February 2019; Received in revised form 3 April 2019; Accepted 4 April 2019

promoted below. As discussed in this paper, coherence is largely a matter of observing Bayes' rule (see below and Eqn. (2) in Theory Section 2.3), although a much less well known related issue called *recurrence* is also discussed below. Throughout, it is also useful to keep in mind the distinction between *predicted risk*, essentially an estimated probability that specified things are the case or will happen in the future, and *experiential risk* as relating to the actual proportion of undesirable results obtained in the past, from which probabilities can be evaluated empirically. Using the latter to improve the former requires better *estimates* of complicated probabilities (involving many factors), as discussed in the following Sections.

1.3. Inadequacies of current tools

Unfortunately, there appear to be serious limitations in the state of the art, or at least in the practice of it, for managing risk. Rebonato has argued that neglect of coherence, addressed in the present paper, may in part explain the 2007–2009 global market crash [2]. "Understanding the reason for this renewed interest is simple: the financial crisis of 2007–2008-2009 has shown with painful clarity the limitations of the purely statistical techniques... that were supposed to provide the cornerstones of the financial edifice" [2]. This is despite a probabilistic, analytical and guideline toolkit that has been available to risk managers for decades [2], including access to the BN that Rebonato favored but wished to make safer by emphasizing coherence, discussed below. The earlier promise of medical Big Data to reduce substantially waste and error in healthcare [1] has not yet been realized (e.g. Ref. [4]), and the experiential risks of developing new drugs and FDA rejection rates remain at such a discouraging level that innovation in the global pharmaceutical industry is declining [5]. Of course, challenges to do with estimating probability discussed in the present paper are not necessarily the only ones that must be overcome in order that the promise of medical Big Data can be realized, but many known ones relate to them, including the need for interoperability between data sources and algorithms in the face of uncertainty and inconsistent ontologies, and the need to process diverse, heterogeneous and high-dimensional data [1]. These are difficult problems, and in particular the latter leads (with mathematical inevitability) to a combinatorial explosion that requires huge amounts of machine memory and processing time [1]. These challenges are being tackled by many other workers and of great interest to the present author as part of our larger Q-UEL project discussed briefly below. The following, nonetheless, remains the chief focus of the present paper.

1.4. Simple considerations in deficiencies relating to probability estimation

Not all challenges that research must address in this arena are formidable. The issue of coherence discussed by Rebonato can be addressed by considering fairly simple probability theory, and other issues raised below are of similar flavor. Improving probability estimation by these considerations will be relatively straightforward if some of the data needed to provide more information are available, which is usually likely to be the case, but just not as yet incorporated. The challenge is one of seeing more clearly where it is needed. It is possibly less obvious because all probabilistic inference nets are at their most useful as estimates of a complicated probability when there is insufficient data to calculate it directly. If there *were* sufficient data, then by definition that means that one could directly obtain a joint probability such as P(A,B,C,D,E) or a conditional probability such as P(A|B,C,D,E) (See Eqn. (1) in Theory Section 2) by observing and counting; there may be statistical aspects to consider, but no inference net is required. Here and throughout this paper, A, B etc. stand for example for states, events, attributes, observations or measurements. Of particular interest are clinical and demographic factors in medicine. One may set up an inference net that looks like an *estimation formula*, e.g. when one algebraically *expands* a joint probability into component parts, but if such nets represent exact expansions, they are merely of educational value. The simplest possible example is P(A, B). If it cannot be calculated by counting, then it may hopefully be estimated by P(A) x P(B), which is the estimation formula in this case. The assumption there is that A and B are independent, which is not generally true, as the example of estimating the probability of patients that are male and pregnant shows. What has happened is that information about the very relevant *interaction* between A and B was missed. A more practical example is the simple BN P(A|B,C)P(B|D)P(D)P(C|E)P(E) as the estimation formula for joint probability P(A,B,C,D,E). It is not an exact algebraic expansion: to be so it would require further corrective terms in the formula to do with missing interactions, and so it is only a *partial expansion*. There is no direct interaction modelled between A and D, between B and E, and so on. In particular, that B and C come together in P(A|B,C) and hence have an interaction as implied in the joint event (A,B,C), but they are treated as independent in joint events (B,D) and (C,E), was early noted as a problem for the BN [6]. In the above example, half of the possible interactions are implied in the estimation formula, but the number of possible formulae grows rapidly as the number of factors A,B,C,D,E,F,G, …. is progressively increased. This problem is one of lack of symmetry that is closely related to, and breaches, Bayes' rule. It should be kept in mind that the name "Bayes Net" was chosen to honor Bayes, but does not incorporate his rule in the net itself.

1.5. Deficiencies as relating to graph structure

The present paper is to a large extent concerned with "graph problems" but they have nothing to do with any kind of deficiencies in graphs considered abstractly. A graph is a general structure from the discipline of discrete mathematics, and has multiple applications. One of these applications is that a specific graph structure can specify a probabilistic inference model like the BN that is based on multiplication of conditional and self-probabilities. Adhering to a particular *class* of graph structure can help with choosing valid estimation formulae. In context, the specific graph defines the specific estimation formula. The General Graph (GG) is as the name suggests the most general, or almost the most general, type of graph as considered in the field of graph theory in discrete mathematics, i.e. a set of nodes (or vertices), and a set of connections (or edges, or arc) between them. The factors A, B etc. above are now seen as the nodes. By "almost the most general" was meant that the connections may have a direction associated with them, and any two nodes may be associated conceptually by two such connections pointing in opposite directions, so to ensure that this final generalization is understood, the term *Bidirectional General Graph* (BGG) is used below. It is the graph of interest here. It could be a simple graph with relatively few connections between nodes as interactions, but it potentially allows for interactions between every node, and in both directions. The BN is based on use of a more restricted graph model called the *Directed Cyclic Graph* (DAG), traditionally and by definition [3]. As a point of caution, a *Directed Graph* usually means a graph that has connections that are directed and they can be in two directions between two nodes as in a BGG, so the term DAG is rather misleading, and stating "unidirectional" above rather than just "directional" is important. The DAG is discussed in some depth in this paper but briefly, by being originally based on a particular causal model, the connections are unidirectional and all in the same direction and cyclic paths through the connections are not allowed. By confining the choice of all possible valid estimation formulae to this class of graph, assembly is relatively easy, but has also led to criticisms of the BN. Notably, branch points in the DAG corresponding to such as construction P(A|B,C, …)P(B, …)P(C, ….) have been criticized by authors, e.g. Ref. [6]. The problem is stated as "*variables starting from the same parent are treated as conditionally independent*" [6]. One should perhaps say that is true in *one direction of conditionality*, since they are not assumed independent in P(A|B,C, …). Criticism could be avoided if the assumption was that B, C etc are always independent, but that would limit the usefulness of the DAG, and

is usually not intended.

1.6. Historical context and information-theoretic perspectives

In the present author's opinion, use of the DAG in inference is not fully consistent with information theory if the DAG is used outside its proper, but limited, context. Similar issues arose many years ago in a project by the present author for inference in molecular biology and bioinformatics using a Bayesian information-theoretic network approach (e.g. Ref. [7]). These earlier efforts led to specific methods called GOR methods that continued to be developed into the late 1990s [8] and, implemented on a large number of servers, still show significant levels of use. At least some of the approaches would be seen today as naïve Bayes classifiers in logarithmic form [7], and GOR was described by one reviewer as based on “conditional probabilities” and “therefore essentially Bayesian in its analysis” [8], although the approaches predated Pearl's BN [3] by several years. These early models saw the graph model used as representing a *physical information flow* [7]. The important point here is that these models [7–10] were used in bioinformatics to follow the *Central Dogma* of molecular biology [10], i.e. that there is a *one way* flow of information from DNA to protein structure and function [10]. For that purpose, the DAG, including simpler tree graphs, would be appropriate, but it was in part the concern about the general applicability outside the context of unidirectional flow that confined application of these early methods to bioinformatics and protein science [10]. A later development extended it well with data mining for discovery purposes [9], but for discovery of relationships, rather than using them to assemble an inference net, DAG criticisms are not an issue. These early models were explicitly based on Bayesian thinking, on which was built in turn the Theory of Expected Information [7]. The key feature was that one can ultimately only hold probability-like degrees of belief, called say Be, about the world, but one can at least choose to believe in a classical probability "out there", say Be(P(A) | data). Solving integrals for expected information obtained from such Bayesian posterior distributions resulted in expressions with terms that were the partially summated harmonic series $1 + 1/2 + 1/3 + \ldots + 1/n$ that expressed the information content in n observations, optionally including virtual frequencies of observations as prior belief [7,9]. A primary reason for its use was to handle the sparse data prominent in bioinformatics at that time and avoid a data sufficiency problem in “classical” statistics, i.e. to recognize that a large amount of weak evidence, expressed as information, can add up to outweigh a judgement made without it, as in a court of law. However, in modern more general data mining, the same problem arises, because high dimensionality inevitably leads the researcher to encounter sparse data as the natural low occurrence of joint events containing many factors [9]. For example, calculation of P(A,B,C,E,F,G) requires a lot more data than P(A,B,C), and the joint event (A,B,C,E,F,G) may not even be seen at all. There is, of course, no reason to believe its true probability value is zero, but there are common cases when the causal DAG model does suggest zero probabilities in conflict with information theory, as follows.

A concern is that Pearl's causality model with P(A|B) (represented graphically as A ← B) appears to imply that B is causal of A with the probability of that happening as the value of P(A|B), and that it is *impossible* that A is causal of B. This would signify that P(B|A) = 0. By Bayes' Rule (Theory Section 2.3, Eqn. (2)) it would also imply that P(A|B) = 0. It is now fairly widely recognized that a causal model is not essential to the BN. In a population, P(A, B) can be imagined as existing, and so can P(A), so P(B|A) > 0 can be defined (Eqn. (1)). Even if the causal interpretation were enforced, there are other interpretations that avoid P(B|A) = 0. Notably, one could have a degree of belief in the mechanism that A causes B and a degree of belief in the mechanism that B causes A, such that P(B|A) > 0, but then, a DAG approach cannot represent it, Bayes' rule is required, and a BGG approach would be needed. If there is conversion of a DAG to a BGG, and one wishes to preserve what the original author of the BN wanted to convey, it will not usually include the insistence that P(B|A) = 0. If it did, the above Theory of Expected information [7] (and its later extension as zeta theory [9]) avoids or reduces the problems. It dictates that a probability based on observations and counting, or more precisely our expectations of such a probability, cannot be meant to be precisely zero because that would require an infinite amount of information. It can be no closer to it than implied by the fact that there were, for example, 50 observations of B so far, in which A did not show up; it could show up on the 51st. Roughly stated, the *expected probability* P(A) cannot be less than approximately 1/n for n refuting observations, unless there is prior knowledge, in which case n includes a “virtual frequency” of observations to represent it [9]. More precisely, it cannot be much less than $e^{-\zeta(s,\ n)}$ where ζ( ) is the Riemann zeta function with s ≥ 1, partially summated, up to n. In the case of present interest, s = 1, and ζ( ) is the above harmonic series [9]. This could be usefully applied to a BN, but then again arguably, it suggests that the user should be using a BGG, not a DAG. In any event, all workers would presumably agree that constant presence of a probability of zero (or even an unrealistically small one) in a any probabilistic inference net that is *purely multiplicative* is clearly a bad idea, and that is avoided in the above approach.

1.7. Use of the Bidirectional General Graph for inference perceived as difficult

It appears to be a problem for adoption of the BGG in inference that it looks likely to be difficult. Pearl's insight and causal interpretation [3] allows the DAG to serve as the basis of guidance for probabilistic inference net construction, but for the BGG, guidance of that kind is necessarily lost, simply because it is the most general kind of graph possible, and consequently many more valid estimation formulae are possible. Separating the valid from the invalid is harder. But ultimately, that may be seen as questionable. A BN's DAG in general can be more complicated than a tree graph and general recipes of BN construction are still non-trivial. Rather, the recipes described here as solutions of the BGG approach are of a different flavor. It is emphasized here that a BN is repairable by conversion to a BGG in steps that use simple familiar probability-theoretic principles. Nonetheless, it remains that they represent an extra layer of conceptual difficulty, and the collective application of these steps, and the mathematical machinery required to develop them into a much larger self-consistent approach to probabilistic semantics, has arguably not been helped by drawing on the mathematics of theoretical physics as discussed in Section 1.14. A defense is that the BGG model emerged naturally from this mathematics, not *vice versa*, and BN to BGG conversion is just a small part of a theoretical system with a far broader class of applications (Section 1.14). The emphasis in this paper is therefore to simplify description, but without bypassing the original mathematical framework when this helps or adds clarification or insight.

1.8. Specific aims of this paper

The main specific aim of this paper is to demonstrate how a departure from traditional DAG methods by the use of the BGG allows consideration of bidirectional interactions by means of what is essentially a simple recipe, and not one involving any iteration, i.e. by just one evaluation a single formula that can be seen as the probabilistic inference net. Step by step conversion of a fairly simple but practical BN to BGG form by hand is decribed, after assigning arbitrary probabilities, and then goes on to discuss semi-automated methods of assembly with a summary of their performance. The BGG enables a much larger set of less restricted models and not every one of them can reasonably be said to have a BN counterpart, but this focus on manual BN to BGG conversion is useful (a) because it illustrates the recipe for repair that is automatable, (b) because it indicates how much of the work in identifying and assigning values to probabilities in specific BNs can still be

used, and (c) because it highlights the deficiencies of the DAG not only in general but also in the cases of specific BNs as models. The following Sections 1.9–1.13 describe problems that have been addressed in the course of this work. They are considered as adequately solved for present purposes by the approach in this paper, but in each case there are still issues arising as possible opportunities for further research, and these will be briefly indicated.

1.9. Research problems addressed: (i) realism

It seems self-evident that the real everyday world is better modelled by relationships described by a BGG rather than a DAG, because it is best seen as a network of interactions in the everyday sense. Notably, interactions are typically bidirectional. In addition, one can commonly trace cyclic paths through the relationships. Examples include road, rail and subway maps, biological neural networks and metabolic pathways. Note that a bidirectional graph does not necessarily imply cycles, but conversely the bidirectional graph is required if cycles are to be represented in a way that can be immediately be solved without iteration. In many cases, a DAG would imply a *discontinuity* as more realistic detail is added. For modeling flow more physically, as in traffic, blood flow, electrical circuits and chemical reactions, it should be possible to convert the graph to physical probabilities and their processing rules, and in turn convert these using formulae for physical laws known for such systems, so adding a first layer of physics of immense practical value. A good basic graph model can also evolve to a much richer kind of graph description. For example, the information in the early inference nets used in bioinformatics to predict protein structure [7–9] was quickly extended into “protein folding” simulations, in which the nodes became atoms interacting by covalent bonds and non-covalent energy interactions or forces [10]. Here note that the interactions are bidirectional and cyclic structure abounds. Because the DAG excludes the bidirectional and cyclic interactions that are typical of real systems, the BN would seem hard to justify as a general predictive model. There are more assumed independencies. By the objections in Ref. [6], use of the DAG would seem to imply the assumption that two or more events must always occur independently to give rise to their stated quantitative final effect. On consideration, this seems a startling deficiency because if belief in this independence is genuinely held by the user, it should apply in both directions of conditionality and would seem to eliminate the need for branch points that are the quintessential features of any practical network. In order to overcome this, use of hidden nodes was proposed, and to calibrate the conditional probability matrices for the hidden node, the use of a gradient descent method [6]. Here the objective function to be minimized was the squared-error between the measured and computed values of the instantiated nodes using both forward and backward propagation to compute the node probabilities. As with any optimization approach, this may sometimes suffer from entrapment in local optima. Implementation is therefore fairly complicated and still not widely used in BNs. It is also important that such approaches do not appear to be able to account for all the limitations of the DAG. Notably, it is still the case that the DAG forbids a cyclic path in the network, as with A ↔ B ↔ C ↔ A, but it does so even if the connections are treated as unidirectional, as in A → B →C → A. The hidden variable approach is however still of some interest because any detection of lack of coherence suggests that there could be missing probabilities in the model for which new variables can substitute.

1.10. Research problems addressed: (ii) coherence

The coherency principle in Bayesian decision theory was fairly well known before Rebonato's work but has been promoted by him [2]. It states that whether probabilities are subjective and personal or based on observations and counting, all possible relevant conditional probabilities and the self-probabilities of what they are conditional upon must come to the same quantitative conclusion so that consistent decisions can be obtained. It reduces to depending on (a) the very simple equation called Bayes' rule (or law, theorem, or equation - see Eqns. (2) and (3) and associated discussion in Theory section 2) combined with (b) even simpler notions of normalization, i.e. how probabilities add up to 1 or to marginal probabilities. Rebonato favors the BN, but only when ensuring that probabilities are made coherent prior to their introduction into the BN by use of contingency tables in which all the relevant data and possible probabilities from them are considered, even if only a few of them are used in the BN. By not paying attention to coherence, as Rebanato shows, risk analysts may be lead to false or misleading conclusions [2]. Bayes' rule is not intrinsic to the DAG, because Bayes' rule is a matter of bi-directionality. Bayes' rule itself might be considered as described by the simplest useful BGG when a joint probability is being estimated in two directions (see Methods Section 3), but only one side of the equation could be considered a BN. An initial perceived research challenge was to develop an approach in which not all probabilities were required and that use could be made of conditional probabilities and other kinds of probabilistic knowledge in a repository called a knowledge representation store (KRS). It became important that coherence by Bayes' rule was inherent whenever possible in each entry in that store. It is important here to distinguish the many examples of BNs that perform well because they are exact solutions (Section 1.4). Provided that values are correctly assigned, the network represented will be coherent. It such cases it would be incorrect to claim that the BN tests the hypothesis that a specific DAG represents. That is simply because any exact expansion into parts yields the same result, even if common sense dictates that the particular model is illogical.

1.11. Research problems addressed: (iii) proportion and prevalence

Proportions differ somewhat from probabilities, albeit largely in a conceptual way, because they are intended to represent or predict an actual fraction of the population that would be obtained by exhaustive counting, in the manner of a census. In epidemiology, it is essentially the same as *prevalence*, e.g. the fraction of the population with influenza. Such probabilities are absolute values and so harder to calculate correctly in comparison to applications that can make use of a further step to generate relative values, including renormalized distributions (see below), and so reduce errors. Consider the epidemiologists' chain rule (a simple linear inference net, and also a DAG). If there is 0.1 probability of contact with an infected person, 0.1 probability of being infected, 0.1 probability of symptoms if infected, 0.1 probability of complications if symptoms, and 0.1 of dying if complications, one's chance of dying is reassuringly only 0.00001. But that is also the predicted mortality rate due to the disease which gives the number of people dying in the population as 0.001%. If that were out by a factor of 10, it could be a serious miscalculation for practical public health purposes. Here a DAG as a linear graph works, although an epidemiologist would almost certainly be interested in more elaborate networks with more influencing interactions, if there were greater confidence in their performance. A problem in departing from the chain rule is that ref [6] essentially points out that Bayes' rule generally fails almost every time there is a branch in a BN. A student epidemiologist might be frustrated that the probability of a particular etiology or cause cannot be established by a BN set up to predict outcomes, because as a DAG it cannot necessarily be used correctly backwards, but at the same time take confidence in the idea that the problem of independency in parent nodes [6] can be avoided as long as one does not attempt read the BN backwards. Unfortunately this latter view is not correct when calculating probabilities as absolute values and the two issues are entangled (see Results Section 4.1 below). Nonetheless a valid defense for use of the DAG is that such difficulties can be circumvented when a user is more interested in comparing values, or when normalizing a distribution of outputs so that probabilities of options add to 1. However, the particular model chosen cannot then properly be tested because the

use of relative values or normalization causes the contributions of invariant interactions not relevant to a prediction to become superfluous, by cancellation due to division. Sometimes this is almost all of the inference net. For example, consider distribution of A = 1, A = 2, A = 3, etc. using say Net(A = 1) for P(A = 1|B,C)P(B|D)P(D)P(C|E)P(E) divided by Net(A = 1) + Net(A = 2) + Net (A = 3) + ….. Here P(B|D)P(D)P(C|E)P(E) cancels by division, leaving only the normalized probability P(A = 1|B,C). Odds such as Net(A = 1)/Net(A = 2) also result in the same cancellations. This is correct but loses information about absolute values and limits any use for testing a causality hypothesis. Were further interactions included such as A-E, they would not cancel, but for typical DAGs most of the rest of the inference net, perhaps assembled and assigned values with significant effort, would in effect disappear.

1.12. Research problems addressed: (iv) recurrence and incidence

In conversion of a BN to a BGG-based inference method, the following is an aspect of which to be wary, though the potential for further research is that it could allow prediction of *incidence* rather than prevalence, i.e. the occurrence of medical events such as onset or first reporting of a disease in a patient, in principle even to the point of predicting that a patient may have two heart attacks in the next five years, or three attacks, and so on. Consider a branch point in an inference net P(A|C)P(A|C′)P(C,C′|D) and then compare the more common P(A|C)P(B|C)P(C|D). Is it not P(A|C) P(B|C)P(C,C |D) that is really meant? It would almost certainly not be so in a typical BN, but P(C,C) is perfectly legal if we mean that C and C are *distinguishable by recurrence*, as in counting males in a population. Sampling and counting theory go further and require *independence in recurrence*, i.e. that P(C) = P(C|C) = P(C|C,C) and so on, and P(C,C) = P(C)P(C), P(C,C,C) = P(C)P(C)P(C) and so on. Compare P(C,C) = P(C) etc. that means that C is *not distinguishable by recurrence*, and intermediate cases where movement of “heads” being counted and noise does not guarantee that we avoid the same subject being counted twice or more. In the other extreme, an argument could be made for P(C, C) = 0 if recurrence is impossible, or to construct a reasonable distinguishability theory to allow it. In a BN such things are not generally meant, but in using a BGG the difference is unambiguous. In the present paper, conversion of a BN to an inference net based on a BGG requires correction to ensure that things occur only once, because this is the usual BN assumption, by implication.

1.13. Research problems addressed: (v) limitations of any simple causal model

Rebonato [2] favorably critiques the BN in terms of Pearl's discussion of wet lawns, garden sprinklers and rain, to illustrate the value of common sense use of causality. One objection is that this particular BN is an example of an exact expansion of a joint probability that must get the right answer, and so its predictive performance cannot test any particular causal hypothesis. Even if were not exact, medical examples are often more complicated and do not adhere to such simple causal models. A purely causal model implies the assumption that any influences from a node, say A, on other nodes cannot directly or indirectly influence node A. One has in mind here that in the causal interpretation, nodes also represent separate events in time, so that one would have to implement time-travel to influence something that happened earlier. However, there is an argument that BN cannot in any case be a test of causality *per se* because the same kind of DAG structure can be used as all or part of a net with non-causal interpretation. For example, P(A|B) can certainly work as P(“B causes A”) in a DAG, but also as P(“All B are A”), P(“A sampled from B”), P(“A is greater than B”), amongst others, and very generally by P(“A if B”) without much commitment to interpretation. When a researcher investigates a population, interactions between individuals and exchanges and influences between groups of people are frequently occurring in all directions, but this has never prohibited sampling in classical statistics and for deducing probabilities in general. All that is required is that a system such a system is sampled on a much shorter timescale than that on which the overall system evolves. In that, we should not preempt a causal interpretation. For example, in industrial nations some 80–90% of adults with type 2 diabetes are overweight, while 30–60% of overweight patients have type 2 diabetes, begging the question of which mostly causes which. A simple one-way causal model removes the ability to test and enforce coherence by Bayes' rule in any kind of net, and the problem with lack of access to that rule is that what it dictates is frequently not intuitive, as many popular probability puzzles illustrate. Finally, causality is not a clear cut concept in practical contexts. Design involves causal steps, and may be the reverse of the description of a natural process. For example, recall that very early inference methods (Section 1.6) were essentially DAGs, but that seemed a clean-cut case: the Central Dogma of molecular biology was that information can only flow from DNA to amino acid sequence to protein structure and function, not in the reverse direction [10]. However, the subsequent development of laboratory DNA synthesis and growth of protein engineering [10] highlighted that one can to a useful extent design modified proteins by starting from required modified conformations and functions and then synthesizing novel DNA required to produce them [10]. Analysis of this is beyond present scope, though simple considerations lead to the notion of a design cycle with opportunities for inference and causality researchers.

1.14. Previous work relevant to the Hyperbolic Dirac Net as a Bidirectional General Graph

There is some case for stating that any use of a BGG in probabilistic inference must be essentially use of what the author calls a *Hyperbolic Dirac Net* (HDN). Certainly, starting from the principles of quantum mechanics (QM) [11], has led to use of the BGG in a way that solves the above problems naturally. In physics, QM does not have quite the same issues concerning what causes what, while uncertainty, and bidirectionality including oscillations between states as waves, evidently abound [11,12]. Moreover, cyclic relationships and paths are fundamental to Feynman's diagrams of interactions and computations of probabilistic inference known as “path integrals” [13]. These features are highlighted in the *Dirac notation*, or *braket notation*, and associated algebra developed by Dirac [12] which forms the basis of the approach described here. Its use for representing uncertainties in observations and for inferring probabilities and expected values from it has been a standard for quantum mechanics (QM) in physics and chemical physics (and later biophysics) since the 1930s [12]. Although in his Nobel Prize speech and elsewhere Dirac indicated that his methods should be more generally applicable to almost anything except poetry and current economics (which he appeared to distrust), he did not give details. A likely key point to address is the quantity called the probability amplitude in QM because its description in terms of complex algebra allows it to encode the probability dual such as {P(A|B), P(B|A)}. The problem for present purposes it that the imaginary number needed is usually seen as $\boldsymbol{i}$ such that $\boldsymbol{ii} = -1$. Exponentials of $\boldsymbol{i}$-complex expressions lead to waves responsible for the weirdness that is notoriously associated with QM's predictions for everyday life. These should be avoided for most practical medical applications. Following initial speculation [14], the author developed the mathematical principles of the HDN [15–20] as a kind of inference net based on Dirac's work. There the emphasis is on the part of the Dirac algebra with imaginary part based on the split complex number or real tessarines, also called the *hyperbolic number* (here called $\boldsymbol{h}$) because the property $\boldsymbol{hh} = +1$ relates to hyperbolic functions. It has been rediscovered independently by many other workers, including by Dirac, although in many different guises. In Dirac's system flavors of an $\boldsymbol{i}$ part plus an $\boldsymbol{h}$ part coexist, but it is “getting rid of $\boldsymbol{i}$” by its conversion (Lorentz rotation) to $\boldsymbol{h}$ in wave

mechanics that turns out to be important for allowing classical behavior [15]. A previous paper has described HDN construction based on ***h***-complex probability amplitudes [19], but that description was rather theoretical and illustrative, exploring various approaches to construction, and lacked the treatment of recurrence (Section 1.12). This work on the theoretical basis of the HDN was followed by the associated Q-UEL (Quantum Universal exchange Language) based on the Dirac notation [21–24]. Q-UEL also takes inspiration from the Semantic Web and other efforts discussed and compared with the HDN and Q-UEL approach that are fairly extensively reviewed in Ref. [22]. The recent industrial implementation of the HDN and Q-UEL is as a high performance system called the BioIngine, which includes many new algorithms and applications for HDN use [25–31].

1.15. Related efforts by other workers

The approaches to inference described in the present paper as well as in those of many other workers have historical roots in earlier work on thought and reasoning, such as the original paper by Bayes [32], and the work by philosopher Karl Popper [33]. The Theory of Expected Information [7] and several Q-UEL conventions and HDN algorithms are seen as consistent with Popper's principles [34]. Considerations of these in conjunction with Pearl's work [3] and ideas concerning deductive, induction and abductive inference influenced design of the BioIngine [27–31] (see Theory Section 2.1) and the simplified Q-UEL-based AI language POPPER [34], as well as current studies (mentioned briefly in the present paper) of HDNs in relation to hidden variables or layers in BNs [6,37] and neural networks [38–40]. Related efforts by other workers have been reviewed in some detail [22–31] and especially in Ref. [22], but none of those other efforts could reasonably be considered as relating closely to the HDN or representing an interoperability language based on Dirac notation. The efforts of the present author and collaborators were influenced by early Expert Systems, and especially the book [41] on the Stanford MYCIN project (1974–1984), usually described as the earliest major attempt at a medical decision support system. Expert Systems were a popular approach when Big Data was in the future and information was easier to obtain from human experts. The basic principles and limitations of the DAG are seen in MYCIN's IF-THEN rules (which are unidirectional and relate to conditional probabilities in a BN), and in its method for combining evidence. The latter is related to the form P(A) + P(B) − P(A)P(B) which the formula for logical OR when A and B are not mutually exclusive but *independent*. This is thus subject to the same kind of criticism that was raised for the DAG [6]. However, some use was made of "self-referencing rules", which have some of the character of a cyclic path. Consequently MYCIN foresaw the BN approach to some extent, but its founders concluded that "causal Information is useful but not sufficient for problem solving in most medical domains" [41].

Modern efforts reflecting industry trends might reasonably be summarized as follows. Concerns over the huge cost of the inability to detect fraudulent or erroneous healthcare insurance claims [42] have encouraged many workers to look for approaches based on data mining and inference [43] as well as broader practical [44] solutions, but there is a strong element of detecting the anomalous as associations that are abnormal, and so the work is often of different flavor. However, diagnosis and risk as prediction of disease remains an important and popular topic promoting the use of inference nets for clinical decision support (e.g. Refs. [45–51]) and for recovering relevant medical information (e.g. Refs. [52,53]). These involve BNs or are BN-like. Such studies are increasingly aided by continuous progress, but also increasing diversity, in how to construct and assign values to inference nets automatically (e.g. Refs. [52–64]). A significant amount of work continues that critiques the BN. Such publications often also describe how the deficiencies might be repaired, though they at are in the main not too easy to use and are of nature of the solution in Ref. [6]. Related to the concerns addressed in Ref. [6], it has been noted that there are conditional independence properties that hold at the parameter level, but often there are conditional independence relations that hold only for specific values of a parent or child [65,66]. Again, workers have found ways to fix these problems within the Bayesian network framework (e.g. Ref. [67]]) and related to that, there is also research into the so-called *"noisy OR"* where one can quantify the causal power of the parent on the child independently for each node [65]. This has been address by several workers (e.g. Ref. [68]) but the matter may not be fully resolved and not yet fully exploited [69]. These kinds of concerns do not appear to relate to the more general solution discussed in the present paper, since they ae more concerned with the causal aspect of the BN model. Some earlier criticisms of the BN as a causal model have been vigorous. For example: "… serious words of caution must be raised on account of several claims made regarding evidential principles which are supposed to validate inferential applications of such proclaimed 'causal networks' in any practical instance" [70]. More recent authors appear to share similar views (e.g. Refs. [71,72]). Pearl has defended the value of the causal model as the basis of the BN, using arguments in increasingly refined form [73–75]. These are interesting and insightful and can stimulate further thought and research, and certainly remain relevant, but of greater concern to the present author is the neglect of interactions and the loss of symmetry in the BN, both of which affect quantitative calculations, as follows.

2. Theory and review of relevant theoretical principles

2.1. Inference

There are several theories of inference, some of which go back hundreds of years. For example, Roger Bacon (1214–1284) is considered the first to have formally promoted inductive reasoning as part of the scientific method. The philosophical and logical definitions do not, however, always map too crisply to computational science, and there seems some confusion in the computer science literature as to the distinction between deductive and inductive reasoning. This is perhaps in part because a practical approach may contain mixed aspects of both and even be a cycle through both, and in part because it can depend on whether one is speaking about generating the required graph model, or assigning and using the probability values. Fig. 1 describes the approach of the present author and collaborators that also illustrates this.

The main functions of our algorithms are to transform information from observations in the "real world", by data mining, into probabilistic elements of knowledge in the canonical form called Q-UEL, and to use

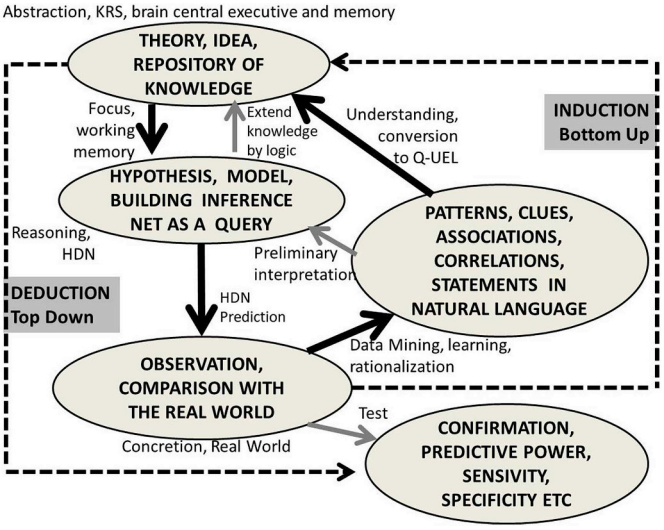

Fig. 1. Basic transformation principles of deduction and induction in the BioIngine.

these in automated inference via the Hyperbolic Dirac Net (HDN) to provide insight and support decision making. Fig. 1 uses the definitions for deduction and induction that currently tend to be in syllabuses at undergraduate level (e.g. Ref. [35]). Insomuch that deductive learning starts with generalizations or rules as canonical representations of knowledge and inductive learning results in such generalizations, the assignments would seem appropriate in this case. While the BN is often considered deductive inference, methods of constructing it can be considered *abstractive* [36] or can involve training of hidden variables [37]. For brevity in Fig. 1, matters to do with *abstraction* meaning the internal world of mind or computation are pooled into the top ellipse on the left hand side, and those to do with *concretion*, i.e. the external world of reality, are pooled into the bottom ellipse on the left hand side.

2.2. BGG generation as part of inference: summary of steps

In the above approach, top down inference using inference nets is seen a process that includes gathering the relevant probabilities together in a way consistent with the basic laws of probability theory so that the resulting inference net represents a valid (i.e. "legal"), if not necessarily best, estimate. The inference net is in this case specifically a Hyperbolic Dirac Net (HDN - reviewed in section 1.14) that is not confined to a DAG and may conform to a BGG. When one can start from a familiar BN, the steps in overview are as follows.

(i) Construct a BN according to the DAG principle in the usual way, or use an existing one from a project or the literature that conforms to that principle.
(ii) Replace the conditional and self-probabilities of the BN (see below) by analogous entities called *brakets* (see below). When required to be a joint probability (recommended), terms representing self or prior probabilities will be need to be included to represent "input" in both forward and reverse directions.
(iii) Note the brakets that occur at branch points in the structure and act on them with an @ operator, actually a procedure that applies two types of correction:
   a. a correction to render variables starting from the same parent are treated as conditionally interdependent, rather than as independent as they are in the BN, and
   b. a correction to ensure that any probabilities of factors as states or events etc. that were intended in the BN to be the probabilities of them occurring only once are still interpreted that way, and not as multiple occurrences.

When starting from a BN it is conceptually easiest to start from one that already estimates a joint probability, and if not, convert it to one that does.

2.3. Probabilities and basic notation used

The basic building blocks of a Bayes Net (BN) is conditional probability is usually expressed abstractly as

$$P(A \mid B) = P(A, B) / P(B) \quad (1)$$

See next Section 2.4 for the nature of A and B. Eqn. (1) might reasonably be called "de Moivre's rule" because Bayes' contemporary de Moivre clarified conditional probabilities. In standard probability theory '|' is usually read as "conditional upon", "given", "if", or "from the sample of". Recall Bayes' rule, theorem or law [32] that builds on the de Moivre rule.

$$P(A \mid B)P(B) = P(B \mid A)P(A) \quad (2)$$

In estimating joint probability P(A,B), it may be said that P(A|B)P(B) from de Moivre's rule is the simplest example of a BN, and that the *probability dual* {P(A|B)P(B), P(B|A) P(A)} is the simplest example of an HDN. Bayes' paper focused on the form P(A|B) = P(B|A) P(A)/ P(B), with P(A) as prior belief, P(B|A) as likelihood (the effect of the data) and P(A|B) as posterior belief. Often it is particularly expressed as a probability distribution or element of such, for example

$$P(Class_j \mid x) = P(x \mid Class_j)\, P(Class_j) / P(x) \quad (3)$$

where $P(Class_j \mid x)$ is the *posterior probability* of Class j given predictor x, $P(x \mid Class_j)$ is the likelihood, the probability of predictor x given class j, and $P(Class_j)$ is the *prior probability* of Class j before seeing x. $P(x)$ is typically determined by normalization, or said to be invariant. $P(Class_j)$ is thought of as a fixed *prior degree of belief* (though it could be probability derived from older data) and the likelihood is considered as the new information, as the *update operator*, such that the posterior probability becomes more specifically the *posterior degree of belief*. However, we have merely substituted A by $Class_j$ and B by x in P(A|B)P(B) = P(B|A)P(A) and rearranged. In the context of BNs the more customary representation (Eqn. (3)) is possibly confusing because it is often the prior probabilities that are usually intended to be routinely changed or updated while the likelihood based on the available data stays constant as a core feature of the network, at least for a period.

2.4. Attributes: the meaning of A, B, C etc

Since in this paper brakets map to probabilities, the same theoretical basis is used here in both and a practical example of what can be actually meant by P(A |B, C) is

P('systolic BP(mmHg)': = 145 | male 'type 2 diabetes' age(years): = 45, 'Fasting glucose(mg/dl): = 'ge180′)

Often in the present paper, A B, C, etc. are called *attributes* in Q-UEL's XML-like tags, by analogy with attributes in XML, or *factors* when having in mind demographic and clinical factors on a patient record. They usually correspond to what are meant by *states, events, observations*, *preparations, measurements, parameters, variables, nodes* etc. in different fields and contexts. The key feature is that they, and also expressions containing them, are *in principle countable*, or that a similar subjective notion may be held. A same attribute can occur twice or more, even with the same value, representing recurrence (Section 1.12). Commas or blanks between attributes are defaults for logic **and.** The bar '|' can be read as logical **if**, appearing in Q-UEL as | **if** |. Importantly, HDN construction usually requires the use of a special "universal" attribute '?'. It is needed in order to form brakets representing prior probabilities P(A) and is typically intended to represent the act of preparation or observation of a state analogous to that process in QM, an act that can be performed with certainty even though what is measured is uncertain prior to the measurement..

$$P(?) = 1 \quad (4)$$

$$P(A, ?) = P(?, A) = P(A) \quad (5)$$

$$P(?|A) = 1 \quad (6)$$

$$P(A|?) = P(A) \quad (7)$$

2.5. Algebraic joint probability estimate (AJPE) and symbolic manipulation

Joint probability estimates by an HDN are important for establishing coherence by Bayes' rule. An example of an expression for *exact calculation* of a joint probability is

$$P(A, B, C) = P(A \mid B, C)\, P(B \mid C)\, P(C) \quad (8)$$

It is also a simple example of a BN that is exact, not an estimate. Note that the joint probability can always be made conditional, which is often the more interesting value, by division by an appropriate probability. For example, dividing the above by P(C) gives P(A, B | C). Real

practical value arises when it is necessary to *estimate* a more complicated joint probability because the knowledge or data available is inadequate for doing so, e.g. the count n(A, B, C) and hence P(A, B, C) is not yet available.

$$P(A, B, C, D, E) \approx P(A \mid B, C)\, P(B \mid D) P(D)\, P(C \mid E) P(E) \quad (9)$$

Here the approximation sign ≈ means "an estimate of", because Eqn. (3) assumes direct independencies between certain attributes A, D and E. Recall that it does not follow that it is always a good estimate. The right hand side of Eqn. (9) is the a *joint probability estimate* JPE. Many different JPEs with different values may be valid solutions, in effect eigensolutions. Some are better than others, and they will usually be solutions with fewest independencies showing in the *algebraic joint probability estimate* (AJPE), i.e. a JPR in algebraic form re-expressed to contain only joint and self-probabilities. Reduced to use of non-conditional probabilities, the estimate on the right hand side of Eqn. (9) above becomes

$$\frac{P(A, B, C)\, P(B, D)\, P(C, E)}{P(B, C)}$$

Though not strictly speaking part of HDN construction other than as a last recommended verification step, using similar algebraic reductions at each step of BGG construction not only provides one-off mathematical proof but is also a valuable routine aid to manual HDN construction and in development of automatic methods. Symbolic manipulation of the AJPE from the outset, using the instruction steps as algebraic manipulations, is a valid alternative viewpoint and construction strategy. In an AJPE, the joint events in brackets specify the interdependencies. Attributes not appearing in the same brackets are assumed to be independent, i.e. randomly associated, expected on such a chance basis. A simple rule is that the number of appearances of each attribute A, B etc. should be one less in the denominator (divisor) than in the numerator. That is, providing that each attribute is occurring just once in observations or in such a way that subsequent observations are indistinguishable from the first observed. This is the assumption in BN to HDN conversion as discussed above. By dividing the above by relevant joint or self-probabilities one obtains an *algebraic conditional probability* estimate (ACPE). The simple or self-probabilities such as P(A) and P(B) can be considered *prior probabilities* in the BN so that multiplying the above EJP by the *prior probability ratio* $P(A_{new})P(B_{new})/P(A_{old})P(B_{old})$ provides the update, or transfer to a new relevant context. The above assumes A and B are independent, and $P(A_{new}, B_{new})/P(A_{old}, B_{old})$ is the better choice. However, it is convenient to work with separate inputs. A check for adequate coherence should however be done in each case.

2.6. Dirac's original braket notation and algebra

Probabilities in a BN become brakets in an HDN. A typical example of a braket or bracket from Dirac Notation [12] as originally used in physics is as follows.

< momentum(eV-sec): = 0.2 | position(Angstrom): = 6.3 >

It can be seen the basic element from which vectors, and matrices as operators, (all these are also valid Q-UEL) and also expressions representing inference nets such as path integrals can be constructed. Only < A|B > and < A |**operator**|B > and these expressions are important in this paper. Their *probability amplitude* is scalar, occasionally real but traditionally an ***i***-complex scalar quantity. It is Dirac's recipe for converting the braket to observable conditional probabilities P(A|B) and P(B|A) [11] that implies that < A|B > encodes the *probability dual* {P(A|B), P(B|A)} by applying ket normalization to < A|B > and < A|B > *. Here the asterisk * indicates the operation of complex conjugation, i.e. it changes the sign of the imaginary part. Compare P(A|B) = P(A|B)* because it is purely real, and Bayes' rule makes it clear that in general, P(A|B) is not equal in value to P(B|A), but it *is* true that

$$< A|B > \; = \; < B|A >^* \quad (10)$$

$$\{P(A|B),\, P(B|A)\} \; = \; \{P(B|A),\, P(A|B)\}^* \quad (11)$$

Almost everything from Dirac's QM is transferable to the HDN and Q-UEL, except that multiplications of brakets to conform to classical probability laws requires the Lorentz rotation $\boldsymbol{i} \rightarrow \boldsymbol{h}$ of QM as wave mechanics, or equivalently a perspective as follows.

2.7. The hyperbolic complex interpretation and Dirac dualization

Real probability values in a BN become ***h***-complex probability amplitudes in am HDN. For a probabilistic semantics to describe the more familiar everyday world and medicine, we are not usually interested in reasoning about particles as waves, but are more concerned with examples such as follows.

< 'systolic BP(mmHg)': = 145 | Glucose(mg/dl): = 180 >

Dirac rediscovered ***h*** such that $\boldsymbol{hh} = +1$. Although most often thought of as γ matrices that form the basis of quantum field and particle theory, it appears very early in Dirac's book [11] as the entity σ that can split or *dualize* < A|B > into two additive parts, ½(1 + σ) < A|B > + ½(1-σ) < A|B > (Dirac used the ket $|P>$ in place of braket < A|B >, but the above follows directly from bra and ket multiplication rules). A hyperbolic imaginary component is thus seen as a valid aspect of a probability amplitude (the value of < A|B >). A purely hyperbolic complex probability amplitude can be expressed as a the *Hermitian commutator* of empirical adjoint conditional probabilities, and also as an equivalent *spinor form*:

$$\begin{aligned}\langle A|B\rangle &= \tfrac{1}{2}[P(A|B) + P(B|A)] + \boldsymbol{h}\tfrac{1}{2}[P(A|B) - P(B|A)] \\ &= \iota P(A|B) + \iota^* P(B|A)\end{aligned} \quad (12)$$

The probabilities P( ) are empirical, usually obtained from observing and counting. Otherwise, all statements in Section 2.6 and above still apply. Physicist's *spinor projectors* expressed using Q-UEL's ι or *iota* notation are $\iota = ½ (1 + \boldsymbol{h})$ and $\iota^* = ½ (1\text{-}\boldsymbol{h})$ (in physics ***h*** typically appears in spinor projectors as γ5 with the same hyperbolic property γ5γ5 = +1). Indeed it is important for consistency with QM that the form ι < A|B > + ι* < A|B >, where < A|B > is dualized to ½(1 + σ) < A|B > + ½(1-σ) < A|B > as above but with σ = ***h***, turns out as equal to ιP(A|B) + ι*P(B|A), a possibly counterintuitive result that is easily shown from the following. The spinor projectors have the following important properties that make proofs and calculations easy: the idempotent property ιι = ι, ι*ι* = ι*, the annihilation property ιι* = ι*ι = 0, and the normalization property ι+ι* = ι*+ι = 1. It is common to write ιPfwd and ι*Pbwd as shorthand for the "probability forward" and "probability backward" for any braket or bra-operator-ket or expression using them. The terms Pfwd and Pbwd come from the names for Q-UELs' tag value attributes which must carry the values of the tag up front when it is transmitted or stored as a tag on the Internet, for example,

< 'Public Safety': = '60–69′ Pfwd: = 0.5795 | **if** | 'Population Health': = '80–89′ **and** 'Environment': = '60–69′ Pbwd: = 0.0993 >

The author and collaborators commonly use $\iota HDN_{fwd} + \iota^* HDN_{bwd}$ when writing about an inference network that is composed of many such tags or their simplified forms. The steps of HDN construction in this paper were declared in section 2.2. These steps are greatly facilitated by first estimating a *joint probability* in each direction; we hope to find, or must adjust the probability terms to find, the following equalities.

$$HDN = \iota HDN_{fwd} + \iota^* HDN_{bwd} = HDN_{fwd} = HDN_{bwd} \quad (13)$$

A significant imaginary part (See Eqn. (12)) means that the constructor has not correctly built a net that is a joint probability, or a cyclic path correctly, or represented certain types of network correctly where symmetry or equal reciprocal relations or equivalence is required, e.g. probabilities that A is equivalent to B and *vice versa*, with probability 1 in each case. From the above perspective, converting a BN to an HDN usually involves considering the BN first converted to estimate of a joint probability, then as HDN_{fwd}, and then finding the HDN_{bwd} that balances it, meaning that it causes Eqn. (13) to hold when a joint probability is first estimated. If the original BN was conditional or the HDN is required to be conditional, a coherent joint probability, once established, is divided by a simpler coherent joint probability (with less factors or attributes) to be obtain it. The latter may simply be a self-probability such as P(B).

2.8. Semantic interpretation of conditional probabilities

Sometimes in BN to HDN conversion, it is useful to look at an author's description of a BN to ensure what was intended. In construction of an HDN *de novo* from Q-UEL tags, the wording used on the tags for the relationship (the operator) is important. Irrespective of Pearl's causality approach [3,73–75], there are other kinds of relationship that can follow or use the same graph rules even in a DAG: simple conditional or implicative, categorical, relative such as "larger than", and so on. A summary of some of the points in this Section and related ideas of importance in the present paper, including coherence, is shown in Fig. 2.

Verbs of action etc. based on Dirac operators like relator **R** in < A | **R** | B > are not needed in the present paper, but as in a BN, conditional relationships can usually reasonably be interpreted in a verbal way, and it becomes important as to how we think about and describe P(B|A) when encountering P(A|B) in conversion of a BN to an HDN. The *categorical interpretation* of < A|B > is e.g. < B | **is** |A > and the *causal interpretation* is as < B | **causes** |A > (note the inversions of A and B). When Pearl speaks of A being causally dependent on B in P(A|B) [3], this maps to < A | **'is caused by'** |B > as a component of the overall BN as a hypothesis. In the categorical interpretation, < A| **'is equivalent to'**| B > means < B| **'is equivalent to'**| A > = < A|B > = < B|A > = 1. If it were less than 1, it would imply < A| **some** | B > = < B| **some**| A > = < A|B > = < B|A > . Note that for interpreting verbal, prepositional and most other relationships (as for the important cases in QM), the operator **R** is *Hermitian*, meaning that we can write

$$< A|\ \mathbf{R}\ |B> \ = \ < B\ |\ \mathbf{R}^*\ |A> \ = \ < B|\ \mathbf{R}\ |\ A>^* \qquad (14)$$

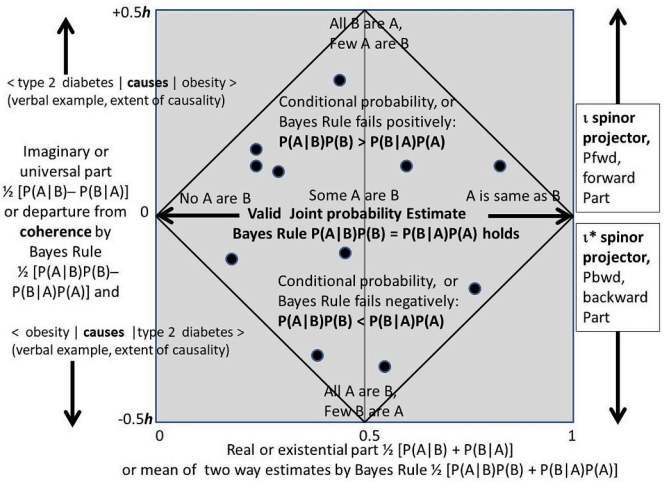

Fig. 2. Diagram for quantitative categorical logic and coherence by Bayes' rule.

Note also that **'is equivalent to'** and **some** satisfy Eqn. (14) in full, but are said to be *trivially Hermitian* because it implies that the bra-operator-ket is purely real. More generally the complex conjugation of **R** as **R*** gives a distinct active-passive inversion of the verb, e.g. if **R** = **is** then we can imagine an converse such as **R*** = **'is exemplified by'.** If **R** = **causes**, then **R*** = **'is caused by'**.

2.9. Significance of probability 1 and 0 in the HDN

Within the current system there are some formal considerations concerning use of extreme probability values. Checking that this still gives adequate *coherence* in an HDN is important, so major Q-UEL software applications always report the imaginary component, which should be of a significantly smaller order than the probability being estimated for a joint probability or other symmetric relationship. Giving a braket or other Q-UEL tag a value of 1 (in both directions) in an HDN is equivalent to removing it (the same is true for a probability in a BN). Conversely, not knowing of some probability value, or perhaps not knowing the need for the probability at all, is as if the probability, or in this case braket, were included with probability 1. In accordance with Popper's principle [33], all statements are initially seen as assertions awaiting refutation by contrary evidence, and initially take the value 1 which is also used to indicate ignorance or absurdity. It is the default value for probabilities, odds and association constants on Q-UEL tags and in HDNs. This is all consistent with the fact that information I = -logP approaches 0 as probability P as e^1 approaches 1. In the Theory of Expected Information [7,9] the probability approaches 1 (or a specified value if prior subjective knowledge is available) when amount of data vanishes toward zero [25]. Also, a low probability should never be considered as zero: consistent with discussion in Section 1.6, $0 = e^{-\infty}$ would imply infinite amount of information refuting an assertion, but in addition it would prohibit use of Bayes rule, and arguably cause theoretical problems in use of spinor projectors.

2.10. Braket representation of prior probabilities

To establish a joint probability as a method of BGG construction and as a check on coherence requires that probabilities are introduced in a manner analogous to use of P(B) in P(A, B) = P(A|B)P(B) and of P(A) in P(A, B) = P(B|A)P(A). Recall that '?' is the state or event of preparation or observation that certainly occurred, so that P(?) = 1. The analogues of self-probabilities are tags of the following general forms and values.

$$< A\ |\ ?\ > \ = \iota P(A) \ + \ \iota^* \qquad (15)$$

$$< ?\ |B> \ = \iota \ + \ \iota^*\ P(B) \qquad (16)$$

Note that < A|? > < ?|B > = < A|B > when A are B are independent. By some analogy with Dirac notation, the operator **?** = |? > < ?| defines < A|**?**|B > = < A|? > < ?|B > and means "is independent of". Similar concepts normally expressed as < A|B > ' = ιP(A|B) + ι* and ' < B|A > = ι + ι* P(B|A) also play a deep role analogous to Dirac's recipe for obtaining observable probabilities [11], but this is beyond present scope.

3. Methods of HDN construction

3.1. Basic points regarding conversion of a BN to and HDN

Recall that BNs and simple HDNs are *not connectionist* in the manner of a neural net NN [38] and each represents an expression in conditional probabilities that are *commutatively multiplied* together. The steps described here for HDN construction represent restoration of interdependencies by including further probabilities. This can imply cyclic paths that commonly arise in BGG construction even if one starts with an inference net based on a DAG. The DAG chosen here illustrates these points and is a realistic example problem in that it uses an typical sized

BN which is an actual successful model from the literature, although in Results Section 4.1, probabilities are selected arbitrarily (using 0.1, 0.2, 0.3 etc.) in order to demonstrate how a coherent result may be obtained from such an arbitrary assignment. The choice of a small BN allows the manual approach that illustrates the construction method outlined in Theory Section 2.2. For BGG construction in general, without a BN as starting point step 1 would then be replaced by a proposed estimation formula, i.e. a specific preliminary BGG. Having established a coherent joint probability, one may divide by required probabilities to obtain a conditional form (if that is wanted). In the above process, it should be kept in mind that an HDN as a JPE may be balanced *algebraically*, and hence the graph is correct, but the values will still not in general be coherent if the probability *values* assigned to the algebraic variables are not a consistent set. To check this, the imaginary part ½ [HDN_{fwd} - HDN_{bwd}] of an HDN vanishes for a JPE. The above manual construction is, however, a toy problem in the sense that than one would normally have constructed an HDN by semi-automatic methods for real computations of practical use [25,31]. These are used in some comparative studies. Briefly, in the first method [25], the user makes an initial proposal as to the structure of the net, and the system then provides the probability values and suggests estimates for any probabilities that are not available. In the second [31], the user provides a simple specification of interdependent and independent factors, each in a short list. The latter approach is facilitated by working with odds (probability ratios) and allows the equivalent of millions of probabilities in the inference net.

3.2. Step 1 of HDN construction. Replace probabilities in a Bayes Net by brakets

The example BN chosen from the literature [37] for conversion is shown in Fig. 3.

For brevity this figure covers both the original BN and the result of the first step in conversion, which is simply to replace prior and conditional probabilities by the corresponding brakets. This BN is described by its authors as a Dynamic Bayes Net (BN), with the interesting feature of a hidden variable (meaning here attribute, node etc.) that can have values calibrated by the methods described [37]. It was considered as having A and B as discrete observed values, C as a hidden variable, and D, E, F, and G as observed continuous variables. The observed variables are not thought of that way in the present paper and this net is treated as static. The reason for choosing this BN is partly that a cyclic path is generated simply because of that conversion to the bidirectional case: C → E → G → C and C → G → E → C. It is immediately evident that there are direct interactions here that are not included in the original BN. It also shows how probabilities can be assigned to a hidden node in order that coherence is satisfied. If P(C) were unknown valid values could be deduced by the need to be consistent with other available probabilities.

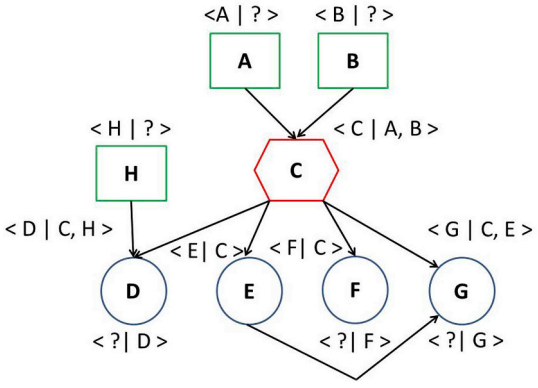

Fig. 3. A Bayes Net with conditional and prior probabilities replaced by brakets. See Ref. [37] for original source as a Bayes Net.

A useful adjunct to the first step involves showing that we have a valid Algebraic Joint Probability Estimate AJPE formed by the probabilities implied by HDN_{fwd}. The first sub-step is to write the list of probabilities corresponding to Pfwd.

P(A) = P(A|?)

P(B) = P(B|?)

P(H) = P(H|?)

P(E|C) = P(C, E) / P(C)

P(F|C) = P(C, F) / P(C)

P(C | A, B) = P(A, B, C) / P(A, B)

P(D |C, H) = P(D, C, H) / P(C, H)

P(G |C, E) = P(C, E, G) / P(C, E)

This gives

$$HDN = \iota \frac{P(A, B, C)\, P(C, E, G)\, P(C, D, H)\, P(C, F)\, P(A)P(B)P(H)}{P(A, B),\, P(C, H)\, P(C)^2} \quad (17)$$

There is a strong argument for adding ι* to the above, but since at this point the equation only serves a conceptual purpose, and in any event (e.g. whether Pbwd is 0 or 1) it is clear that Pfwd overall does not in general match Pbwd overall, that discussion can be ignored here. The step of creating a meaningful ι* component is trivial: switch x and y in p ($x|y$) to create P($y|x$). The resulting *adjoint forms* are

P(?|A) = 1

P(?|B) = 1

P(?|H) = 1

P(C|E) = P(C, E) / P(E)

P(C|F) = P(C, F)/P(F)

P(A, B | C) = P(A, B, C) / P(C)

P(C, H | D) = P(D, C, H) / P(D)

P(C, E | G) = P(C, E, G) / P(G)

From this point on it should be clear how an algebraic approach can be automated. There are now two lists called the Pfwd or iota (ι) and the Pbwd iota-star (ι*) lists to be manipulated by the following "rules of thumb". Equivalently, and it soon becomes surprisingly easy manually, one can manipulate the brakets. Conceptually, one usually works directly with brakets, and that produces the effects described in the developing AJPEs, but one can work directly with the latter following the "rules of thumb". Briefly, these rules reflect that outputs, that will be seen as inputs when the net is read in the reverse direction, may need to be added to the Pbwd list, and then that lists need to be manipulated to give symmetry and a valid joint probability. At this point we have

$$HDN = \iota \frac{P(A, B, C)\, P(C, E, G)\, P(C, D, H)\, P(C,F)P(A)P(B)P(H)}{P(A, B),\, P(C, H)\, P(C)^2} + \iota^* \frac{P(A, B, C)\, P(C, E, G)\, P(C, D, H)\, P(C,E)\, P(F,C)}{P(C)\, P(D)\, P(E)\, P(F)\, P(G)} \quad (18)$$

Evidently the two terms still do not balance. In a BN, it is not always obvious as to what nodes are to be considered as output, i.e. which nodes are normally, or options, to be interrogated as to their values. The simple rule-of-thumb in BN to HDN conversion is that $< x \mid ? >$ is used when there are no nodes coming into x, in the net as seen in the forward direction of the net and no nodes going out in the backward direction of the net. $< ? \mid x >$ is used when there are no nodes going out from x, in the forward direction of the net (and hence in the BN), and no nodes

coming in, in the backward direction of the net. We do not see $< D \mid$ *something* $>$, $< F \mid$ *something* $>$ or $< G \mid$ *something* $>$. This here affects the Pbwd and ket $\mid >$ part. The effect of this rule of thumb here is to create the brakets $< ?|D >$, $< ?|F >$ and $< ?|G >$ at the bottom Fig. 2. These have the value 1 in the forward direction so there is no effect on effect on the first (i.e. ι) term, but now we have

$$HDN = \iota \frac{P(A, B, C)\, P(C, E, G)\, P(C, D, H)\, P(C, F) P(A) P(B) P(H)}{P(A, B),\, P(C, H)\, P(C)^2} + \iota^* \frac{P(A, B, C)\, P(C, E, G)\, P(C, D, H)\, P(C, E)\, P(F, C)\, P(D)\, P(F)\, P(G)}{P(C)\, P(D)\, P(E)\, P(F)\, P(G)}$$

Simplifying,

$$HDN = \iota \frac{P(A, B, C)\, P(C, E, G)\, P(C, D, H)\, P(C,F) P(A) P(B) P(H)}{P(A, B),\, P(C, H)\, P(C)^2} + \iota^* \frac{P(A, B, C)\, P(C, E, G)\, P(C, D, H)\, P(C, E)\, P(F, C)}{P(C)\, P(E)} \quad (19)$$

Evidently the two terms still do not balance.

3.3. Step 2 of HDN construction. Append the $^{@}$ association operator to brakets at branch points in order to give symmetry by Bayes' rule

The problem to be resolved here is that the BN usually has an unwanted broken symmetry at branch points. In this step, the issue early seen in Ref. [6] is addressed. Symmetry can be restored by making x and y *inter*dependent in both directions of conditionality or *in*dependent in both. Unless we wish to loose information, the first is required and the solution when starting from a BN is to correct the effects of the attributes Dirac *bras* $< |$, not to correct for the contents of the *kets* $| >$.It corresponds to a BN case where a joint probability such as B, C connects with B and C separately as in P(A|B,C) P(B|D)P(C|E). In general for a HDN, if corresponding to the inverse case such as P(D|B)PE|C) P(B, C | A), the contents of bras and kets in the corrective procedure below would need to be interchanged. One must apply a branch point correction by an ***h***-complex association constant for forward-backward symmetry as to interdependence. It is interpreted as the action of the $^{@}$ operator:

$$< a \mid b, c >^{@} = < a \mid b, c > K(b; c \mid ?) = < a \mid b, c > < b, c \mid ? > / < b| ? > < c| ? > \quad (20)$$

In general, $< a \mid b, c, d, \ldots >^{@} = < a \mid b, c, \ldots. > K(b; c;d; \ldots \mid ?)$

$$= < a \mid b, c, d \ldots > < b, c, d \ldots \mid ? > / < b| ? > < c| ? > < d|? > \ldots \quad (21)$$

That is, the association constant K for the interaction between the attributes in the ket is being used to overcome the assumption that they are interdependent the condition of the probability of the probability to the left but independent in the bra to the right. The mutual information between e.g. x and y that is likely to be available because P(x, y) is likely to be known. It is to be implemented as in Fig. 4.

The situation after applying the $^{@}$ operator as above is described in the layout in Fig. 4. The effect of this is to multiply the first (ι) term by P (A, B)/P(A) P(B) and by P(C, H)/ P(C)P(H) and by P(C, E)/ P(C)P(E). *In practice, one can data mine for these correction values as association constants* such as K(A; B) = P(A, B)/P(A) P(B). See Results Section 4.3. The

Fig. 4. Terms to multiply arranged in a standard layout.

above "correction" leads to

$$HDN = \iota \frac{P(A, B, C)\, P(C, E, G)\, P(C, D, H)\, P(C,F) P(A) P(B) P(H)\, P(A, B),\, P(C, E)\, P(C, H)}{P(A)\, P(B)\, P(A, B),\, P(C, H)\, P(C)^4\ \ P(E)\, P(H)} + \iota^* \frac{P(A, B, C)\, P(C, E, G)\, P(C, D, H)\, P(C,F) P(A) P(B) P(H)\, P(A, B),\, P(C, E)\, P(C, H)}{P(C)^4\, P(E)}$$

Simplifying

$$HDN = \iota \frac{P(A, B, C)\, P(C, E, G)\, P(C, D, H)\, P(C, E)\, P(C, F)}{P(C)^4 P(E)} + \iota^* \frac{P(A, B, C)\, P(C, E, G)\, P(C, D,H)\, P(C,E)\, P(F,C)}{P(C)\, P(E)} \quad (22)$$

3.4. Step 3 of HDN construction. Modify the action of the @ operator extension for a fork where several same attributes meet

This can be considered as part of the @ operation, with null affect if the following situation does not arise. In starting from a BN, this affects the attributes in Dirac *kets* $| >$. It corresponds to a BN case where several conditional probabilities are converging to the same node such as C in P (A|C)P(B|C)P(C|D) where there are too many C's to the left of the branch point. Uncorrected, the simple example $< A|C > < B|C > < C|D >$ would suggest that what we really mean is

$$< A|C > \quad < C, C \mid D > . \quad < B|C >$$

The problem is that usually we do *not* mean that an attribute such as C can occur more than once when starting from a BN. A division by $< ?|C >$ is required for each *extra* C on the left side of the branch in the ket parts $|C, \ldots >$. In the case of Fig. 3, we see 4 such kets, so that conceptually we divide by 3 $< ?|C >$.

$$HDN = \iota \frac{P(A, B, C)\, P(C, E, G)\, P(C, D, H)\, P(C, E)\, P(C, F)}{P(C)^4\, P(E)} + \iota^* \frac{P(A, B, C)\, P(C, E, G)\, P(C, D, H)\, P(C, E)\, P(C, F)}{P(C)^4\, P(E)} = \frac{P(A, B, C)\, P(C, E, G)\, P(C, D, H)\, P(C, E)\, P(C, F)}{P(C)^4\, P(E)} \quad (23)$$

In this case, balance is achieved. The result is a real value: recall $\iota + \iota^* = 1$, so $\iota x + \iota^* x = x$.

3.5. Step 4 of HDN construction. Check the HDN constructed by calculating the value of the so-called algebraic joint probability estimate AJPE

In summary the three step algorithm for BN to HDN conversion is to consider estimation of a joint probability, replace conditional probabilities by corresponding brakets and ensure that all brakets corresponding to prior probabilities are included for both directions, and to apply corrections to branch points both to prevent information loss in the reverse direction and to ensure that attributes are not counted more than once. After that, it should be verified that the evaluated imaginary part ½ [Pfwd – Pbwd] for the whole net is negligible in comparison to the magnitudes of Pfwd and Pbwd for the net. In addition, if the AJPE is evaluated separately at each step, one might say that it is "triple entry bookkeeping". Providing the HDN is correctly constructed, the three values of the net Pfwd, Pbwd and final AJPE will be close but may differ slightly as a consequence precision in manual calculations due to rounding. See Section 4.1.

3.6. Step 5. generating conditional forms if required

Conditional probabilities can be set up by judicious removal of brakets from the HDN or by division by brakets. For example, for the

above HDN to be made into the estimate of a probability conditional on, say, all of A and B, the following is required;

$$\frac{P(A, B, C)\, P(C, E, G)\, P(C, D, H)\, P(C,E)\, P(C, F)}{P(A)P(B)\, P(C)^4\, P(E)}$$

Except by coincidence of values, the imaginary part is no longer real for a conditional form. Note that if there is sufficient departure from coherence, a new P(A, B, C) value may need to be obtained to correct it, and that this really applies also to a BN, since in P(A | B, C) removes the contribution from P(B, C) but still reflects sampling from that population with a B−C interdependency. Irrespective of that, recall that updating by e.g. $P(A_{new}, B_{new})/ P(A_{old}, B_{old})$ is the better choice with less independency assumptions, but is typically less convenient for use. If it is required only to consider output D, F, G conditional on everything else, one can take advantage of the known joint probabilities and take out any joint probabilities that depend on more than one of A, B, C, and H, then make up the remainder with self-probabilities. This requires some discussion and will be described elsewhere, because more commonly one predicts one factor at a time and typically one is then considering an odds HDN already described elsewhere [31].

3.7. Practical calculations and comparisons with other work

Comparison with published efforts of other workers, in order to see how the method might be improved, is more effective if all methods are brought into exact alignment leaving as distinct only what are arbitrarily judged as the distinguishing essential features, but this is well known to be extremely difficult. Fortunately, for systematic review of what is available in the literature, methods in their entirety can at least be treated as a "black boxes". In principle, their merits then stand or fall on how each performs. Unfortunately there still remains the significant problem that the data are very different, or are used differently, even if the work described is intended to perform essentially the same task. For example, predictions, not diagnoses, are of interest in the present study, but they are typically harder to do than diagnoses because risk concerns the future, and diagnostic data is supposed to provide powerful clues. Distinctions among lifestyle, etiologies, triggers, biomarkers and symptoms, are often blurred. For example, some studies compared that predicted heart disease included chest pain. In the present case, an existing systematic review [44] was used which represented a great deal of work already done in reconciling the approaches described in it, concerning the application of Naive Bayesian Networks to disease prediction. Primary sources were however examined. Two relevant studies on kidney disease of 673 patients [45] and 169 patients [46] respectively were selected for comparison because adequate data was available to the present author (of the order of 700,000 patient records [31]), while data for the particular cancers and glaucoma considered in the systematic review were not available for comparison.

Computer "experiments" as studies for best methods are described in Results Section 4 but the general approach was as follows. Preliminary runs using DiracMiner [25] were performed then using nets manually constructed by DiracBuilder [25] that sought to emulate as closely as possible what demographic and clinical factors were covered in the studies compared in the above systematic review, building preliminary HDNs to establish coherence (Results Section 4.4). DiracBuilder then also allows studies in which the discrepancies between a DAG and BGG could be directly computed (Results Section 4.2). As discussed in Results 4.3, DiracMiner also produces tags that include association constants that can measure the discrepancies that arise in the use of the BN as a DAG. The factors found effective was used to set up the so-called hitlist and wishlist (or shortlist) used to hold what are considered as interdependent and independent factors respectively, for use by SMASH [31], which automatically adds unsupervised data mining in hunt for factors relevant to these. In practice, the interdependent factors were considered as the demographic factors including major classifications such as male, age range and diabetes, while the independent factors were clinical evidence. SMASH automatically provides a two-state distribution for whatever is chosen as the prediction target by using odds, in which the HDN that the target predicted such as renal failure will occur is in effect divided by the corresponding HDN that it will not occur. As probability ratios, these can be seen in the overall HDN as

$$\text{Odds-HDN} = \imath \text{ predictive-odds} + \imath^* \text{ likelihood-ratio} \quad (24)$$

These probability ratios play the same role as prediction forward and prediction backward. While inclusion of negative evidence is a desirable feature, and the brakets and the probabilities they represent are tested, there are cancelations by division of those that do not contain the target. It is therefore important to emphasize that there is not only still bidirectional and $\hbar$-complex character, but also that coherence is still an issue and can be tested, as follows. A joint probability ratio is the same as the predictive odds because of calculation of the conditional probability by division, and the following equation then follows from Bayes Rule (Eqn. 1.2), where T is the target such a renal failure, and C is a set of conditions.

$$P(T, C) / P(\textbf{not } T, C) = \imath\, P(T|C) / P(\textbf{not } T|C) + \imath^* (P(C|T) / P(C|\textbf{not } T) \times P(T)/P(\textbf{not } T)) \quad (25)$$

As for the basic HDN as a joint probability, if the $\imath$ and $\imath^*$ parts are not equal, there is a significant imaginary part indicating lack of coherence.

4. Results

4.1. Worked example by hand calculation

See Table 1 a number of computer experiments and comparative studies have been performed regarding the relative ability of a BN based on the DAG and HDN based on the BGG to estimate proportion or prevalence in a population (Section 1.11), as opposed to distributions and normalized values. One aim in this first example is to show how discrepancies can arise in estimation if interactions are omitted because a DAG rather than a BGG is used. The conversion procedure should be applicable to any initial BN values. For that reason, in this worked example, which is for the inference net in Fig. 3 the self-probabilities were *arbitrarily* assigned to the BN of choice [37] for conversion: as 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8. Joint probabilities were then initially arbitrarily selected using only the rules that P(a, b, …) was very roughly of the magnitude P(a)P(b) … and, P(a, b, …P(a), P(a, b) ≤ P(b), and so on, and P(a, b, c, …) ≤ P(a, b) and P(a, c) and P(b, c) … and so on. This tests the ability to produce a coherent HDN from an arbitrary choice. A good tactic is to examine the probabilities to see if anything is unaccounted for only the steps in Sections 3.2 to 3.5, assuming that they have been correctly applied. In the first column showing self and joint probabilities, there are five joint probabilities underlined that appear in the AJPE for the HDN, being P(C, E), P(C, F), P(A, B, C), P(C, D, H) and P(C,E, G), so unless C is to be allowed to appear several times such that P(C, C, C, C, C) is meaningful and not equal to P(C), four must be removed, i.e. by multiplication by self-probability $P(C)^{-4}$. In all the following we can ignore terms like P(C, E | ?)/ P(C|?)P(E|?) where the number of C is not changed, and in Pfwd we can ignore the C counting issue entirely because it uses P(?|C) and P(?|x) = 1 for any x. In the Pbwd Column, there are 4 Cs to the left of the conditional bar and 1 to the right, requiring correction by multiplication by $P(C|?)^{-3}$. The real part of this HDN is 0.00002213 and the Imaginary part of this HDN is 0.00000012. The Joint Probability Estimate is small and is said to be *sufficiently coherent*. Since a reasonable effort was made to ensure that the probabilities were coherent, the small imaginary part is primarily due to precision in manual assignments. If further refinement is required the imaginary part tells how much and in which direction. Having addressed the first step of BN to BGG conversion by including

Table 1
"Triple Entry Bookkeeping" to Show Adequate Coherence by Bayes' Rule and Correct HDN Construction..

Probability	AJPE value (multiply underlined values and use the power)	Pfwd for HDN	Pfwd value	Pbwd for HDN	Pbwd value
P(A)	0.1	P(A | ?)	0.1	P(? | A)	1
P(B)	0.2	P(B | ?)	0.2	P(? | B)	1
P(C)	$\underline{0.3}^{-4}$	P(? | C)$^{-3}$	1	P(C | ?)$^{-3}$	37.037
P(D)	0.4	P(?|D)	1	P(D | ?)	0.4
P(E)	$\underline{0.5}^{-1}$				
P(F)	0.6	P(? | F)	1	P(F | ?)	0.6
P(G)	0.7	P(? | G)	1	P(G | ?)	0.7
P(H)	0.8	P(H | ?)	0.8	P(?| H)	1
P(A,B)	0.06	P(A, B | ?)/ P(A|?)P(B|?)	3	P(? |A, B) P(?|A)P(?|B)	1
P(C, E)	$\underline{0.03}$	P(E | C)	0.1	P(C | E)	0.06
		P(C, E | ?)/ P(C|?)P(E|?)	0.2	P(? | C,E)/ P(?|C)P(?|E)	1
P(C,H)	0.2	P(C, H | ?)/ P(C|?))P(H|?)	0.833	P(? | C, H)/ P(?|C)P(?|H)	1
P(C, F)	$\underline{0.2}$	P(F | C)	0.67	P(C | F)	0.333
P(A, B, C)	$\underline{0.05}$	P(C | A, B)	0.83	P(A, B | C)	0.167
P(C, D, H)	$\underline{0.03}$	P(D | C, H)	0.15	P(C, H | D)	0.075
P(C, E, G)	$\underline{0.01}$	P(G | C, E)	0.33	P(C, E | G)	0.01429
Algebraic	0.00002222	HDN Forward	0.00002201	HDN Backward	0.00002225

the reverse direction of conditionality, one can ask what difference was made by converting to a BGG compared with the value obtain when using the same probabilities in Eqn. (17), i.e. P(A, B, C) P(C, E, G) P(C, D, H) P(C, F) P(A)P(B)P(H)/ P(A, B), P(C, H) P(C)^2. This evaluates as 0.05×0.01 x $0.03 \times 0.2 \times 0.1 \times 0.2$ x $0.8/ 0.06 \times 0.2$ x $0.3^2 = 0.00004444$. it is to be compared with P(A, B, C) P(C, E, G) P(C, D, H) P(C, E) P(C, F)/ P(C)^4 P(E) which evaluates as 0.05×0.01 x 0.03×0.03 x $0.2/ 0.3^4 \times 0.5 = 0.00002222$ as shown in the table. Recall that it is readily shown that the real part of this HDN is 0.00002213 and the imaginary part of this HDN is -0.00000012, and that since a reasonable effort was made to ensure that the probabilities were coherent, the small imaginary part is primarily due to precision in manual assignments. The discrepancy between the DAG and the BGG is much larger than this imaginary part. The discrepancy is a factor of 2. The distinction in that discrepancy is in the changes in interactions revealed by the ratio of the above AJPEs, i.e.

$$P(A)P(B)\,P(C)^2\,P(E)P(H) \,/\, P(A,B)\,P(C,E)\,P(C,H)$$
$$= 0.1 \times 0.2 \times 0.3^2 \times 0.5 \times 0.8 \,/\, 0.06 \times 0.03 \times 0.2 = 2$$

That is, each of the three joint events (A, B), (C, E), (C, H) were initially assumed to have independent parameters, (A)(B), (C)(E) and (C)(H), in one direction of conditionality which relates to the objection raised that was early raised regarding the BN, i.e. the need to treat variables starting from the same parent as conditionally interdependent [6], and the BN assumption that this can be ignored is only possible if coherence by Bayes' rule is ignored. Focusing on these discrepancies P(A,B)/P(A)P(B) = 0.3333, P(C,E)/P(C)P(E) = 5.0000, and P(C,H)/P(C)P(H) = 1.2000, by the above choice of convenient arbitrary numbers as input. Their product is 2.0.

4.2. Similar practical calculations

To explore how large the above kind of discrepancy can be in practice, some studies were carried out with real data as described in Methods Section 3.7 but limited to BNs built with DiracBuilder [25] for predictions. These used probabilities and associations on Q-UEL tags generated by DiracMiner [25] and other Q-UEL data mining applications, including those of SMASH data mining but without proceeding on to use of SMASH in predictive mode. DiracBuilder allowed an initial BN entry to be compared with a final HDN as BGG, using the same data. An example of an initial net created by the user, plausible as a BN, is as follows.

```
<'History of high cholesterol':='0' |'BMI':='30-39' and 'HDL(mg/dl)':='50-59'>
<'BMI':='30-39' |  'Systolic BP(mmHg)':='130-139'>
<'HDL(mg/dl)':='50-59' | 'Taking diabetes medication':='1'> = 25.11, 14.92
<'Taking diabetes medication':='1' | 'Male':='0' and 'Age(years)':='50-59'> = 11.69, 22.91
<'Male':='0' and 'Age(years)':='50-59' | 'Non-HDL(mg/dl)':='120-129'> = 0.0, 0.0
<'History of high cholesterol':='0' |'BMI':='30-39' and 'HDL(mg/dl)':='50-59'>=0.7087,0.0665
<'BMI':='30-39' | 'Systolic BP(mmHg)':='130-139'>=0.4644,0.2851
<'HDL(mg/dl)':='50-59' | 'Taking diabetes medication':='1'>=0.2511,0.1492
<'Taking diabetes medication':='1' | 'Male':='0' and 'Age(years)':='50-59'>=0.1169,0.2291
<'Male':='0' and 'Age(years)':='50-59' | 'Non-HDL(mg/dl)':='120-129'>=0.2081,0.0809
#   NET = {0.07816%,0.01349%} = iota(0.07816%) + iota*(0.01349%)
#   where real part = existential joint probability component = 0.045825%
#   where imaginary part = universal purely conditional component = h*0.032335%
```

Note that simple multiplication between the tags is indicated unless stated otherwise. The nets can be operationally generated by text input which is processed using various trigger word like "given", "when" and "for" and various synonyms, but while this is currently under more advanced development to facilitate use by physicians, it is not yet recommended for complicated research studies or construction of an inference net depository by a human expert. Note the assignment convention using the notation $< A|B > = P(A|B), P(B|A)$, where the (here percentage) probabilities were automatically assigned by DiracBuilder by hunting for and if necessary indirectly deducing appropriate tags. Manual conversion of an initial net, BN in the present studies, proceeds as in Section 4.1. For example:

```
<'History of high cholesterol':='0' |'BMI':='30-39' and 'HDL(mg/dl)':='50-59'>
<'BMI':='30-39' |  'Systolic BP(mmHg)':='130-139'>
<'HDL(mg/dl)':='50-59' | 'Taking diabetes medication':='1'> = 25.11, 14.92
```

This needs correction by a association constant of form P(A,B)/P(A)P(B) on the bra side of the tags, which is introducing the following tags.

```
< 'BMI':='30-39' and 'HDL(mg/dl)':='50-59' | ?>
/< 'BMI':='30-39' | ?>
/< 'HDL(mg/dl)':='50-59' | ?>
```

Note the division sign '/' for the divisors. A general finding from such studies is that predicted abundances (proportions, prevalence) in a population can be significantly in error typically by a factor of 5 times for a net represented by a graph of 6–10 nodes, though sometimes with much more agreement, and sometimes much less. There is a great deal of variation that will be explored elsewhere, but briefly, as discussed below, the discrepancy between a DAG and a BGG tends to be greater in medical contexts when some factors are supposed to provide strong information, e.g. clinical results are intended to facilitate a clear diagnosis.

4.3. Direct computation of Bayes Net branch point discrepancies and hence correction factors as association constants

Although there was found a great deal of variation in HDN-BN discrepancies that will be explored elsewhere, in practice a more helpful and more readily extensible summary, expressed on a "per branch point" basis, is possible. This is because of the observation made at the end of Section 4.1 concerning the dominating and often sole importance of the need to treat variables starting from the same parent as conditionally interdependent [6]. Consequently, we simply need to know the association constant for a joint event such as 'BMI': = '30–39′ and 'HDL(mg/dl)': = '50–59'. This also provides the required correction factor. It is an idea which extends naturally to three or more factors at that time: then we need to consider the *atomic* association constant, i.e. P(A,B,C)/P(A)P(B)P(C) rather than, say, P(A,B,C)/ P(A)P(B,C). To illustrate the above and also diversity of the computer "experiments" performed, the following is an example Q-UEL tag from DiracMiner for a geographic socioeconomic public health study. Note the *standard: = 1.8125* and *classical: = 1.8125: = 22*500/119*51* branch of the assoc attribute, which carries the association constant, that represent "standard" and "atomic" association constant are, of course, the same in the case of just a pairwise association between two factors. It is the estimates Kzeta based on the Theory of Expected Information and zeta function [7,9] that are usually of greater interest, but the level of data is sufficient that these are similar values in most cases.

<Q-UEL-DIRACMINER-KMETHOD-2-FACTOR-CHF-SURVEY:=(application:='Perl version v5.16.3':=DiracMiner158.txt, input:=AetnaOverallClass.csv, patient#:=all:=0-500, samplesize:=0, incidences:=22, prior=0, tagtime(gmt):='Tue Jan 8 02:48:17 2019')

'Equity':='40-49' Pfwd:=(Pfzeta:=0.1922:=exp[3.1387-4.7881], classical:=0.1849:=22/119)

| **if**:='do all':=(assoc:=(standard:=1.8125, atomic:=(strength(nats):=3.1524:=3.1387--0.0137, Kzeta:=1.8450:=exp[3.1387+6.2146-4.7881-3.9527], classical:=1.8125:=22*500/119*51)), Pjoint:=0.0440:=22/500), events:=(P[Population Health:=80-89]=0.2402~=119/500), P[Equity:=40-49]=0.1042~=51/500)) |

'Population Health':='80-89' Pbwd:=(Pbzeta:=0.4431:=exp[3.1387-3.9527], classical:=0.4314:=22/51)

Q-UEL-DIRACMINER-KMETHOD-2-FACTOR-CHF-SURVEY>

The above shows an association of 1.8125 between somewhat poor equity and population health, a somewhat controversial social finding for US data confirmed by classical Pearson's correlation between equity and population health which was significantly negative at −0.324. The important point for present purposes, however, is that most associations constants in this kind of study are close to 1 but the above association does approximate the largest discrepancy of BN to BGG of 2 that seems to be obtained in at least in public health studies of this kind. That is, most BNs tend to agree with the HDN, except when including factors like the above which approximates the discrepancy obtained in Section 4.1.

This relatively small kind of discrepancy between BN and HDN was not the case, however for public health and evidence based medicine studies incorporating clinical data, especially those involving factors that are intended to carry strong information to the physician, e.g. have diagnostic power. The choice of probabilities used in Section 4.1 may be considered relatively kind to the BN considering what three association constants could have been multiplied: using the same medical data association constants for associations for pairs and triplets ranged from 0.0337 to 11.1340, exemplified by the pairwise association of high blood urea nitrogen and high creatinine at 9.21, and pairwise associations with one as congestive heart failure from 0.0829 to 8.0849, exemplified by congestive heart failure with pulmonary circulation disorder at 6.5535. The situation tends to get worse when there are several factors involved, because several interdependent interactions are then implied. That is to say it tends to worsen discrepancies between DAG and BGG based inference nets that are based on high dimensional data mining. This exemplified by the association constant of 14 in the following SMASH data mining tag in simplified format (the default in SMASH).

< 'Renal failure':='eqY' Pfwd:=0.1049 Ofwd:=0.1127

| **if**:=(assoc:=14.0182, count:=2, factors:=6) |

'Year of birth':='le1953' 'Diabetes uncomplicated':='Y' 'Obesity':='Y' 'HCTLast HEMATOCRIT%':='low' 'Tracheoscopy and laryngoscopy with biopsy':='N' Pbwd:=0.000611374283454888 Obwd:=14.9412 >

This also illustrates that significance as information strength is not necessarily dependent on the so-called "support" for an association "rule". In this case the 6-factor concurrence from high dimensional data mining is "supported" by just 2 observations that were nonetheless deemed by the SMASH (and DiracMiner) method as significant because an occurrence of just 2 was still a lot more than would be expected from the product of the self-probabilities for the individual factors, in the same data.

4.4. Practical calculations and comparisons with other work

These studies focus on predictive performance and represent comparison with a systematic review [44]: see Methods Section 3.7. SMASH [31] used all records data mined to obtain the brakets for odds HDN predictions. They all contained *interdependent* variables age, sex, diabetes Y/N, and obesity Y/N, and the probabilities implied by the brakets used in HDN construction contained at least one of the *independent* variables, creatinine, blood urea nitrogen, and urea protein. Initially, all the latter included "slightly above normal range", i.e. not necessarily severe, and were first test only. This gave accuracy 91%, specificity 62%, sensitivity 92%, but dropping to 82%, 40% 83% respectively, when obesity and diabetes were excluded. Studies in the present project to improve predictions showed that this reflected a high probability of minor renal disease or of recovery, leading to studies discussed below. The results seem nonetheless already reasonable in the current state of the art. Accuracy of 90% obtained by the HDN for kidney disease in diabetics compare reasonably well with the accuracies which were the measures available from the study cited by Leung et al. [45] in the systematic review [44]. It specifically considered diabetic kidney disease (their counterpart of the HDN interdependent variables would be age, age at diagnosis, diabetes, ethnicity Chinese). "Diabetic kidney disease" was classed as an eGFR less than 60 ml/min/1.73 m^2, a range which includes mild and reversible forms. It achieved 84% and 71% accuracy for Naïve Bayes' Net and least squares regression respectively (sensitivity and specificity seem not to be clearly specified in the primary source and were not cited in the systematic review [44]). The study used 17 clinical factors selected reduced to 8, systolic BP, triglycerides, white blood cell count, cholesterol total, LDL, and HDL,

waist to hip ratio, alcohol intake (and originally 10 genomic markers reduced to best 5 known to be relevant). In the HDN studies including BP, triglycerides, white blood cell count, cholesterol total, LDL, and HDL, and excessive alcohol intake (Y/N) as independent factors lead to accuracy 90%, sensitivity 62%, specificity 92% with diabetes as an interdependent factor. This reduced to 85%, 55% 87% without diabetes as a factor.

Accuracy of 89% obtained by HDN methods also compared well with the study by Huang et al. [46] also included in the above systematic review, also interesting because Naïve Bayes Net, Support Vector Machine, Random Forest and Decision Tree methods were compared for a target of severe kidney disease needing analysis. The four methods obtained 61–63% accuracy, 41–64% specificity, and 67–85% specificity for variously 3–7 clinical factors. Including up to 5 genetic biomarkers changed these values to 61–65%, 56–63%, and 51–72% respectively. The primary clinical factors in the latter study were clinical were creatinine, blood urea nitrogen, urea protein, but apparently persistent, rather than just first readings as in the analogous odds HDN study. Predictions of severe forms are often (though not always) harder to predict than mild forms (or just a general specification as “disease) largely because milder forms may or may not progress to more severe forms, so benefiting in principle from time series analysis [47]. Consequently, odds HDN predictions using first, last, median and average values were explored. With obesity and diabetes still *excluded*, this improved the results to give the above accuracy of 89%, with sensitivity 60%, and specificty 91%, i.e. comparable with including obesity and diabetes though less than the improvement expected. The above systematically reviewed studies [45,46] also introduced genomic biomarkers; these have not yet been included in the above HDN studies but will be described elsewhere. However, it seems reasonable to say that in the above original studies [45,46] their inclusion did not have dramatic effect.

4.5. Congestive heart disease

SMASH has previously obtained fairly high performance scores for predicting congestive heart disease (which was not covered in the above systematic review), given fairly obvious test cases for comorbidities used as *independent* variables, where just one was needed to occur to have impact. These were pulmonary circulation disorder, severe kidney disease, valvular disease, and diabetes, which gave accuracy 91%, sensitivity 67%, and specificity 93%. Obviously, high scores would be expected here. But also in a public health study using the same data, an application to public health obesity, hypertension, depression and alcohol abuse as independent variables it scored moderately well too, at 86%, 56%, 88%, boosting to 86%, 78%, 73% with diabetes added to the list of interdependent variables. The gold standard for predicting cardiovascular risk was for some years the Framingham Risk Score which was extended in a study by Brindle et al. [48] who used an extended form taking account of age, blood pressure, HDL, smoking, electrocardiographic left ventricular hypertrophy, and diabetes as factors. At a 15% threshold for predicting an event, they obtained accuracy 65%, sensitivity 75%, specificity 55%. Based on that finding, inclusion of blood pressure, HDL, and smoking as independent factors in the odds HDN obtained 86%, 78%, 74%, which is, somewhat surprisingly, not a significant increase. In odds HDNs including fasting blood sugar as an independent factor had no significant effect with diabetes Y/N already included. Anooj [50] obtained results 47–62%, 32–77%, and 45–75% over three data sets. Some data included fasting blood sugar and exercise-induced angina, but for comparison purposes this was considered too close to symptoms limiting the diagnosis, and to outcome. The SMASH odds approach also has a somewhat unusual feature regarding what governs a true positive and true negative and false positive and false negative (e.g. see Table 2, ref [31]). However, it remains that a blind prediction is made for what is in each record, e.g. kidney disease or not kidney disease, that emerges as 100% correct or 100% incorrect, so the approach is fairly on the same basis as that of other researchers. Nonetheless in comparisons, the above treatment by SMASH [31] should be considered as part of the method as a whole.

4.6. Preliminary studies comparing other methods using identical data

Because many differences between studies by different authors can make comparisons between the approaches difficult, preliminary comparisons with neural networks (NN) and similar-looking HDNs with hidden variables were also investigated within the present study, using the same data in the same conditions and established NN [38] and hidden node [6] BN methods, as well as exploring some new developments. For example the techniques of ref [6] were extended to HDNs. For the NN these in-house modifications included particular ways of using prior probabilities and of starting with expected important features ranked toward the beginning of the record as input, and then progressively lengthening the binary representation from the beginning. Because neither the “traditional” or modified approaches performed quite as well as the approaches described above, they will be described elsewhere. Briefly, however, for NNs and hidden layer BNs and above, 10 predictions of for each congestive heart disease and kidney disease obtained 65–75% accuracy, sensitivity and specificity, discarding runs that showed poor convergence due to local entrapment, while corresponding runs for HDNs under identical conditions showed 70–85% accuracy, 40–70% sensitivity and 75–90% specificity. Also while HDN-based methods and some approaches built on related techniques such as ALERT and BILL [31] now put 90% of true positives in the top 10% of records ranked by predicted anomalous character, ranking of results NN and experimental hidden layer approaches have so far put only 50% there. It has been disappointing that HDNs with hidden nodes made no significant change to the results of the basic HDN. However, this might be a promising indication: see discussion below.

5. Discussion

5.1. Summary: restoring missing interactions

The requirement for coherence by Bayes rule, interdependence in parent nodes satisfying association constants, and the need to handle recurrence, form a triad of interrelated considerations that provides strong constraints and guidelines for the probabilities in an inference net. The repair of a BN to meet those requirements and its conversion to a Bidirectional General Graph as a basic HDN are essentially the same thing. Recall that the values obtained in all inference nets of this general kind are in general probability estimates that imply at least one assumption of independency, otherwise an inference net is not required and a probability can be calculated directly from observations and counts. The issue is one of whether important interactions have been included. A golden rule should be that all information about relationships between things that can be considered in making an important judgment should be considered, to represent our best efforts based on the best available knowledge. In fairness to Pearl's BN, it must constantly be kept in mind that while its original form has been promoted by many as a predictive tool, it is more immediately a way of testing a hypothesis or model about interactions in a system. Nonetheless, the DAG can be criticized particularly because it implies somewhat strong assumptions that represent inconsistency between two directions of conditionality and yet do not necessarily have to be made. Notably, if say $P(A|B,C) = P(A,B,C)/ P(B,C)$ can be evaluated, it is extremely likely that at least P(B, C) etc. can be evaluated in order to ensure that independence of B and C does not have to be assumed in the parent nodes. Association constants such as $P(B, C)/ P(B)P(C)$ representing the B—C interaction emerge as the quantities that are deficient in the Bayes Net and key to its repair by its conversion to a Bidirectional General Graph. Of course, the HDN based on the BGG also has its inevitable limitations

because, as a tool for estimation, by definition it misses some interactions, and one must wonder if all these actually need be missed. For the graph of Fig. 3, one may in a simple thought experiment using an example from Section 1.4, imagine that B and G are male and pregnant respectively, in Fig. 3. Then P(B, G) = 0, because they are in effect mutually exclusive, and the overall BN *and* the HDN derived from it must then the value 0. Similarly, any P(A, B) and P(B, C) may be large but P(A, B, C) can be vanishingly small. For example, elderly males associate with anemia, and anemia associates with pregnant patients, but few if any pregnant elderly male patients with anemia are seen. Fortunately, such strongly negative associations can often be detected by data mining, although as "unicorn events" that are not seen at all, there are some challenges [76,77].

5.2. Improvements and future work

Here comments are confined to the topic of inference nets and neural nets that are more readily seen as mutually competitive in the same arena, because there are more elaborate HDNs for probabilistic semantics that have an even stronger AI flavor and about which work is also ongoing [26,27,34]. Preliminary efforts discussed in Results Section 4.6 are consistent with the superiority of the basic HDN but in all fairness to other competing techniques, the results of these studies are as yet inconclusive. The common problem for many kinds of hidden nodes technique, with automatic calibration of weights by local gradient methods based on fitting predictions to data when weights are permutable, is that the fitting function may have many minima, maxima, saddle points (and potentially other topological features in higher dimensional cases). This leads to entrapment in local solutions rather than the in the best global solution, and attempts to overcome this with random restarts gives results that fluctuate over many runs. Many more runs with random value restarts are needed before any conclusions can be drawn as to their future potential, and these studies are still in progress. These too will be described elsewhere but it is worth noting that, with the above modifications, the NN required only 1001 hidden layer neurons which is small in the current state of the art, in order to achieve the above plausible results in reasonable time without evidence as yet for persistent entrapment in local minima.

Recently, there have been proposals that gradient methods must lead to a global minimum in neural nets (notably ref [78]) but this appears to be referring to local gradient methods and the present author is doubtful. As a minimization problem it appears more like that encountered in the "folding problem", here meaning prediction of the three dimensional structure of a protein that has been an archetypal case study for the global minimization problem. The Globex method developed by the present author showed promise by establishing many deep local minima and then using these results to find a coarser, more general trend in the function surface [10], but this failed for very complicated functions that can easily imply astronomic numbers of local minima. The quest for further solutions thus continues [79], and some of these may extensible to more general applications. Certainly the challenge is difficult. Optimization of more advanced semantic HDNs has been explored in which not only probability values but graph structure evolves under syllogistic and similar rules, and it was suspected that a result may be unreachable by Gödel's theorem [34]. Recalling that any graph that changes could for purely multiplicative inference nets be considered as the always same graph in which a probability value takes on the value 1 and *vice versa*, and recalling also the deep relationship between conditional and more elaborate semantic (e.g. verbal) relationships from a Dirac notation perspective, such a suspected conclusion seems likely to be more general.

The following needs some investigation. As declared in Results Section 4.6, it was at first disappointing that HDNs with hidden nodes made no significant improvement to the basic HDN. However, it appears that there may be a readily accessible solution very close to what would be assigned to the HDN in order to ensure coherence, interdependence in parent nodes, and proper treatment of recurrence and so that, in effect, "there is not much room for improvement". As noted above, requiring coherence, interdependence in parent nodes, and proper treatment of recurrence when also applied are strong constraints and guidelines for any calibration method such as minimization. Starting with an estimate of a joint probability and a single hidden node and all other probabilities considered reliable and fixed, it is evident that there is a single solution that at least removes the imaginary part of the value of the HDN, irrespective of, and in effect overriding the need for, any minimization. Extending that to multiple hidden nodes allows a smooth continuum of solutions of multiplied alternative weights. This is rather different to the possible permutations of weights in a neural net that represent local minima. The behavior of an HDN with hidden variables and minimization seems closer to HDN's in which all or most probabilities evolve in time [19]. Although the iteration function differs, it still implies a relationship that may benefit from further thought and research.

6. Conclusions

More realistic modeling of the world requires consideration of the Bidirectional General Graph as a basis for probabilistic inference, and a way to achieve this is conveniently expressed in the language of ***h***-complex quantum mechanics. While the origin of such quantum mechanics lies with Dirac, the notion of an extensively hyperbolic quantum mechanics for physics is fairly recent [80,81] and its possible relation to mental information has not gone unnoticed [82]. However, such efforts seek to paint the grander physical picture of the nature of reality and have not led to an HDN-like proposal or addressed relatively "mundane" issues such as DAG to BGG conversion for inference. Indeed the present author believes that the main potential in the HDN and Q-UEL lies in the application of Dirac's quantum mechanics to probabilistic semantics as a basis of language and thought [14–31,34]. Despite that, the intent in this paper has been to promote broader consideration of some simpler issues in probabilistic inference, and the reason is, as pointed out by Rebonato [2], that their neglect may have had had profound impact on mission critical industries such as finance and medicine. Their re-appreciation could bring about immense benefit. In the present author's view, the reason for neglect may be part the popularity of the BN, which based on the DAG, makes inclusion of these aspects more difficult, and the direct representation of bidirectional character and cyclic paths impossible. In the present paper, the less familiar mathematics persist in that the "triad" or requirements associated with coherence by Bayes rule, interdependence in parent nodes, and the need to handle recurrence, can be seen as important operations in the conversion of scalar real probabilities to scalar ***h***-complex probability amplitudes. This again provides convenience, and a pleasing consistency with quantum mechanics. "Convenience", is, of course, a matter of taste, so it is fortunate that the "fixes" (repairs) described in the paper can be expressed alongside in terms of familiar probability theory.

Be that all as it may, complex graphs for probabilistic knowledge presentation have become a popular theme increasingly appreciated in industry [84], and consideration of how to render these probabilistic could pose a challenge for the traditional approaches, and motivate continued research into broader methods.

Conflicts of interest

This paper is provided to the community to promote the more general applications of the thinking of Professor Paul A. M. Dirac in human and animal medicine in accordance with the charter of The Dirac Foundation, to emphasize the advantages and simplicity of the basic form of the Hyperbolic Dirac Net, to encourage its use, and to propose at least some of the principles of the associated Q-UEL, a universal exchange language for medicine, as a basis for a standard for

interoperability. The basic idea of the HDN is very easy to implement and a simple worked example done by hand is given to illustrate this as a central feature of the paper. These mathematical and engineering principles are used, amongst many others in an integrated way, in the algorithms and internal architectural features of the BioIngine.com, a distributed system developed by Ingine Inc. VA for the mining of, and inference from, Very Big Data for commercial purposes.

References

[1] B. Robson, O.K. Baek, The Engines of Hippocrates. From the Dawn of Medicine to Medical and Pharmaceutical Informatics, Wiley, 2009.
[2] R. Rebanato, Coherent Stress Testing. A Bayesian Approach to the Analysis of Financial Stress, John Wiley & Sons, 2010.
[3] J. Pearl, Probabilistic Reasoning in Intelligent Systems, Morgan Kaufmann, San Mateo, 1988.
[4] http://www.thefiscaltimes.com/2015/10/06/Audit-Uncovers-1247-B-Overpayments-and-Fraud-Medicare-and-Medicaid.
[5] https://www.washingtonpost.com/news/theworldpost/wp/2018/10/17/pharmaceutical/?noredirect=on&utm_term=.8b34ea6af97a.
[6] C.K. Kwoh, D.F. Gillies, Using hidden nodes in Bayesian networks, Artif. Intell. 88 (1996) 1–38.
[7] B. Robson, Analysis of the code relating sequence to conformation in globular proteins: theory and application of expected information, Biochem. J. 141 (1974) 853–867 (1974).
[8] The GOR Method, Wikipedia, https://en.wikipedia.org/wiki/GOR_method , Accessed date: 21 December 2018.
[9] B. Robson, Clinical and pharmacogenomic data mining: 3. Zeta theory as a general tactic for clinical bioinformatics, J. Proteome Res. 4 (2) (2005) 445–455 (2005.
[10] B. Robson, J. Garnier, Introduction to Proteins and Protein Engineering, second ed., Elsevier Press, 1998.
[11] P.A.M. Dirac, The Principles of QM, fourth ed., Oxford University Press, 1958.
[12] P.A.M. Dirac, A new notation for QM, Math. Proc. Camb. Phil. Soc. 35 (3) (1939) 416–418.
[13] R.P. Feynman, A.R. Hibbs, Quantum Mechanics and Path Integrals, McGraw Hill, 1965.
[14] B. Robson, The new physician as unwitting quantum mechanic: is adapting Dirac's inference system best practice for personalized medicine, genomics and proteomics? J. Proteome Res. 6 (No. 8) (2007) 3114–3126.
[15] B. Robson, Towards Intelligent Internet-Roaming Agents for Mining and Inference from Medical Data, Future of Health Technology Congress, Technology and Informatics vol. 149, IOS Press, 2009, pp. 157–177.
[16] B. Robson, Links between Quantum Physics and Thought (A. I. Applications in Medicine), Future of Health Technology Congress, Technology and Informatics vol. 149, IOS Press, 2009, pp. 236–248.
[17] B. Robson, Towards automated reasoning for drug discovery and pharmaceutical business intelligence, Pharmaceutical Technology and Drug Research 1 (3) (2012).
[18] B. Robson, Towards new tools for pharmacoepidemiology, Adv. Pharmacoepidemiol. Drug Saf. 1 (2013) 6 https://doi.org/10.4172/2167-1052.100012.
[19] B. Robson, Hyperbolic Dirac nets for medical decision support. Theory, methods, and comparison with Bayes nets, Comput. Biol. Med. 51 (2014) 183–197.
[20] S. Deckelman, B. Robson, B. Split-complex numbers and Dirac bra-kets, Commun. Inf. Syst. 14 (3) (2015) 135–149.
[21] B. Robson, U.G.J. Balis, U. G. J, T.P. Caruso, Considerations , for a universal exchange Language for healthcare, Proceedings of 2011 IEEE 13th International Conference on e-Health Networking, Applications and Services (Healthcom 2011), IEEE, Columbus, MO, 2011, pp. 173–176.
[22] B. Robson, U.G.J. Balis, U. G. J, T.P. Caruso, Suggestions for a web based universal exchange and inference language for medicine, Comput. Biol. Med. 1 (2013) 2297–2310 43(12).
[23] B. Robson, T.P. Caruso, A universal exchange Language for healthcare MedInfo, in: C.U. Lehmann, E. Ammenwerth, C. Nohr (Eds.), 13: Proceedings of the 14th World Congress on Medical and Health Informatics, Copenhagen, Denmark, IOS Press, Washington, DC, USA, 2013.
[24] B. Robson, T.P. Caruso, U.G.J. Balis, Suggestions for a web based universal exchange and inference language for medicine. Continuity of patient care with PCAST disaggregation, Comput. Biol. Med. 56 (2014) 51–66.
[25] B. Robson, S. Boray, Implementation of a web based universal exchange and inference language for medicine. Sparse data, probabilities and inference in data mining of clinical data repositories, Comput. Biol. Med. 66 (2015) 82–102.
[26] B. Robson, S. Boray, Interesting things for computer systems to do: keeping and data mining millions of patient records, guiding patients and physicians, and passing medical licensing exams, Bioinformatics and Biomedicine (BIBM), Proceedings 2015 IEEE International Conference, IEEE, 2015, pp. 1397–1404.
[27] B. Robson, S. Boray, Data-mining to build a knowledge representation store for clinical decision support. Studies on curation and validation based on machine performance in multiple choice medical licensing examinations, Comput. Biol. Med. 73 (2015) 71–93.
[28] B. Robson, Studies in using a universal exchange and inference language for evidence based medicine. Semi-automated learning and reasoning for PICO methodology, systematic review, and environmental epidemiology, Comput. Biol. Med. 79 (2016) 299–323.
[29] B. Robson, S. Boray, Studies of the role of a smart web for precision medicine supported by biobanking, personalized medicine, FTG, Pers. Med. 13 (2016) 4.
[30] B. Robson, S. Boray, Methods and Systems of a Hyperbolic-Dirac-Net-Based Bioingine Platform and Ensemble of Applications, (2017) US20170185729A1.
[31] B. Robson, S. Boray, Studies in the extensively automatic construction of large odds-based inference networks from structured data. Examples from medical, bioinformatics, and health insurance claims data, Comput. Biol. Med. 95 (2018) 147–166.
[32] T. Bayes, An Essay towards Solving a Problem in the Doctrine of Chances Philosophical Transactions of the Royal Society of London vol. 53, (1763), p. 370.
[33] K. Popper, The Logic of Scientific Discovery, Springer, 1934 (as Logik der Forschung, English translation 1959).
[34] B. Robson, POPPER, a simple programming language for probabilistic semantic inference in medicine, Comput. Biol. Med. 56 (2014) 107–123.
[35] Deduction and Induction, Web Center for Social Research Methods, https://socialresearchmethods.net/kb/dedind.php.
[36] S. Raghaven, R. Mooney, Bayesian abductive logic programs, Proceedings of the 6th AAAI Conference on Statistical Relational Artificial Intelligence 82-87, AAAI Press, 2010.
[37] D. Smith, G. Timms, P. De Souza, C. D'Este, A Bayesian framework for the automated online assessment of sensor data quality, Sensors 12 (7) (2012) 9476–9501.
[38] C.M. Bishop, Neural Networks for Pattern Recognition, Springer, 2006.
[39] R.M. Neal, Bayesian Learning for Neural Networks vol. 118, Springer Science & Business Media, 2012.
[40] M.S. Hung, M.Y. Hu, M.S. Shanker, B.E. Patuwo, Estimating posterior probabilities in classification problems with neural networks, International journal of Computational intelligence and Organizations 1 (1) (1996) 49–60.
[41] B.G. Buchanan, E.H. Shortliffe, Rule-Based Exert Systems. The MYCIN Experiments of the Stanford Heuristic Programming Project, Addison-Wesley, 1984.
[42] Healthcare IT News, (2011) June 20 https://www.healthcareitnews.com/news/ama-blasts-insurers-193-percent-claims-error-rate.
[43] M. Kumar, R. Ghani, Z.S. Mei, Data mining to predict and prevent errors in health insurance claims processing, Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2010, pp. 65–74.
[44] P. Viechnicki, W.D. Eggers, B. Lorenze, M. Greene, J. Guszcza, D. Olson, Shutting Down Fraud, Waste, and Abuse. Moving from Rhetoric to Real Solutions in Government Benefit Programs, Deloitte University Press, 2017, https://www2.deloitte.com/content/dam/insights/us/articles/fraud-waste-and-abuse-in-entitlement-programs-benefits-fraud/DUP_3043_WasteFraudError_FINAL.pdf.
[45] M. Langarizadeh, F. Moghbeli, Applying naive bayesian networks to disease prediction: a systematic review, Acta Inf. Med. 24 (5) (2016) 364–369.
[46] R.K. Leung, Y. Wang Y, R.C.W. Ma, A.O.Y. Luk, V. Lam, M. Ng M, W.Y. So, S.K.W. Sui, J.C.N. Chan, Using a multi-staged strategy based on machine learning and mathematical modeling to predict genotype-phenotype risk patterns in diabetic kidney disease: a prospective case-control cohort analysis, BMC Nephroogyl 14 (2013) 162.
[47] G.-M. Huang, K.-Y. Huang, T-Yi Lee, J.T.-Y. Weng, An interpretable rule-based diagnostic classification of diabetic nephropathy among type 2 diabetes patients, BMC Bioinf. 16 (Suppl 1) (2015) S5.
[48] T. Verplancke, S. Van Looy, K. Steurbaut, D. Benoit D, F. De Turck, G. De Moor G, J. Decruyenaere, A novel time series analysis approach for prediction of dialysis in critically ill patients using echo-state networks, BMC Med. Inf. Decis. Mak. 10 (2010) 4.
[49] P. Brindle, J. Emberson, F. Lampe, M. Walker, P. Whincup, T. Fahey, S. Ebrahim, Predictive accuracy of the Framingham coronary risk score in British men: prospective cohort study, Br. Med. J. 327 (7426) (2003) 1267.
[50] P.K. Anooj, Clinical decision support system: risk level prediction of heart disease using weighted fuzzy rules, Journal of King Saud University - Computer and Information Sciences 24 (1) (2012) 27–40 https://www.sciencedirect.com/science/journal/13191578.
[51] G. Yadav, Y. Kumar, G. Sahoo, Predication of Parkinson's disease using data mining methods: a comparative analysis of tree statistical and support vector machine classifiers, National Conference on Computing and Communication Systems (NCCCS) 1–8 (2012).
[52] C.-Y. Lin, J.-X. Yin, L.-H. Ma, J.-Y. Chen, An intelligent model based on fuzzy bayesian networks to predict astrocytoma malignant degree, 2006 IEEE Conference on Cybernetics and Intelligent Systems, 2006, pp. 1–5.
[53] H.Yazid, K. Kalti, N. Ben, A. Essoukri, F. Elouni, and K. Tlili, A probabilistic network based similarity measure for cerebral tumors MRI cases retrieval", 2011 IEEE Third International Workshop on Computational Intelligence in Medical Imaging (CIMI), pp. 1-7 (2011).
[54] A. Rodríguez-González, M. A. Mayer, G. Alor-Hernandez, J. M. Gomez-Berbis, G. Cortes-Robles and A. L. Lemos, Using ontologies and probabilistic networks to develop a preventive stroke diagnosis system (PSDS)", 2010 IEEE 23rd International Symposium on, Computer-Based Medical Systems (CBMS) 370-377 (2010).
[55] D. Nikovski, Constructing Bayesian networks for medical diagnosis from incomplete and partially correct statistics, IEEE Trans. Knowl. Data Eng. 12 (4) (2000) 509–516.
[56] J. Huang, H. Pan, Y. Wan, Cooperative learning of Bayesian network structure based on PG algorithms, The 8th International Conference on, Computer Supported Cooperative Work in Design 2004. Proceedings, vol. 2, 2004, pp. 708–711 2.
[57] S.S. Liao, H.Q. Wang, Q.D. Li, W.Y. Liu, A functional-dependencies-based Bayesian networks learning method and its application in a mobile commerce system, Systems Man and Cybernetics Part B: Cybernetics 36 (3) (2005) 660–671.
[58] X.-W. Chen, G. Anantha, X. Lin, Improving bayesian network structure learning with mutual information-based node ordering in the K2 algorithm, Engineering IEEE Transactions on, Knowledge and Data 20 (5) (2008) 628–640.

[59] Y. Zhang, Q. Hu, W. Zhang, J. Liu, A novel Bayesian network structure learning algorithm based on Maximal Information Coefficient, 2012 IEEE Fifth International Conference on Advanced Computational Intelligence (ICACI), 2012, pp. 862–867.
[60] P. Suchánek, F. Marecki, R. Bucki, Self-learning bayesian networks in diagnosis, Procedia Computer Science 35 (2014) 1426–1435.
[61] Y. Tang, Z. Xu, A score based approach towards improving bayesian network structure learning, 2014 Second International Conference Advanced Cloud and Big Data (CBD), 2014, pp. 39–44.
[62] B. Ojeme, A. Mbogho, T. Meyer, Probabilistic Expert systems for reasoning in clinical depressive disorders, 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA), 2016, pp. 599–604.
[63] N. Fortier, J. Sheppard, S. Strasser, Learning Bayesian classifiers using overlapping swarm intelligence, 2014 IEEE Symposium on Swarm Intelligence (SIS), 2014, pp. 1–8.
[64] P. Laitila, K. Virtanen, Improving construction of conditional probability tables for ranked nodes in bayesian networks, Transactions on, Knowledge and Data Engineering IEEE 28 (7) (2016) 1691–1705.
[65] B. D'Ambrosio, Infrence in bayesian networks, AI magazine, AI Mag. 20 (2) (1999).
[66] B. D'Ambrosio, Local expression languages for probabilistic dependence, Proceedings of the Seventh Annual Conference on Uncertainty in Artificial Intelligence, vol. 66, Morgan Kaufmann, San Francisco, 1991, pp. 95–102.
[67] D. Poole, Probabilistic partial evaluation: exploiting rule structure in probabilistic inference, Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence, Menlo Park, California, International Joint Conferences on Artificial Intelligence, 1997, pp. 1284–1291.
[68] N. Zhang, Intercausal independence and heterogeneous factorization, Proceedings of the Tenth Conference on Uncertainty in Artificial Intelligence, Morgan Kaufmann, San Francisco, 1994, pp. 606–614.
[69] M. Takikawa, Representation and Inference in Bayesian Networks, Ph.D. thesis Department of Computer Science, Oregon State University, 1998 (1998).
[70] F. Lad, Assessing the foundation for Bayesian networks: a challenge to the principles and the practice', Soft Computing 3 (3) (1999) 174–180.
[71] C. Glymore, What is right with 'Bayes net methods' and what is wrong with 'hunting causes and using them'? Br. J. Philos. Sci. 61 (1) (2010) 161–211.
[72] N. Cartwright, What Is Wrong with Bayes Nets? MONIST, April 1, (2001) http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.821.4897&rep=rep1&type=pdf.
[73] J. Pearl, R. Dechter, Identifying independencies in causal graphs with feedback, Proceedings of the 12th Conference of Uncertainty in Artificial Intelligence, Morgan Kaufmann, Portland OR, 1996.
[74] J. Pearl, Causality. Models Reasoning and Inference, Cambridge University Press, 2009.
[75] J. Pearl, D. Mackenzie, The Book of Why: the New Science of Cause and Effect, Allen Lane, 2018.
[76] B. Robson, The dragon on the gold: myths and realities for data mining in biotechnology using digital and molecular libraries, J. Proteome Res. 3 (6) (2004) 1113–1119.
[77] B. Robson, Clinical and pharmacogenomic data mining: 4. The FANO program and command set as an example of tools for biomedical discovery and evidence based medicine, J. Proteome Res. 7 (9) (2008) 3922–3947.
[78] S.S. Du, J.D. Lee, H. Li, L. Wang, X. Zhai, Gradient Descent Finds Global Minima of Deep Neural Networks, Cornell University, 2019 arXiv:1811.03804.
[79] B. Robson, A. Vaithilingham, Protein folding revisited" pp 161-202 in progess in molecular biology and translational science, Molecular Biology of Protein Folding, vol. 84, Elsevier Press/Academic Press, 2008.
[80] A.Y. Khrennikov, Hyperbolic quantum mechanics, Adv. Appl. Clifford Algebras 13 (2003) 1.
[81] A.Y. Khrennikov, Contextual Approach to Quantum Formalism, Springer, 2009.
[82] A.Y. Khrennikov, On quantum-like probabilistic structure of mental information, Open Syst. Inf. Dynam. 11 (3) (2004) 267–275.