Fourth International Conference on Construction in the 21st Century (CITC-IV)
"Accelerating Innovation in Engineering, Management and Technology"
July 11-13, 2007, Gold Coast, Australia

Applying Bayesian Belief Networks to Construction Management

Zhong Tang
(Lecturer, Department of Civil Engineering, Curtin University of Technology ,Perth, WA, Australia)

Brenda McCabe
(Associate Professor, Department of Civil Engineering, University of Toronto, Toronto, ON, Canada)

Abstract
Bayesian belief networks (BBN) is a probabilistic technique in artificial intelligence. BBN make complex system models operational by showing relationships explicitly - the structure and degree of dependence, which will facilitate the research and practice of decision making. Construction projects become more and more complex and risky in rapidly changing modern societies. Risk analysis and decision-making under uncertainty are real challenges in construction management. BBN provide a potential method to estimate risks and facilitate decision-making in construction projects. In this paper, the basic concept of BBN is presented and BBN establishment and procedure are discussed. One critical step applying BBN to construction management is knowledge elicitation from construction experts. The concerns in knowledge elicitation are examined. Finally, an example of BBN established upon the input from experts is demonstrated.

Keywords
Bayesian belief networks, probabilistic models, knowledge-based systems, knowledge elicitation.

1. Introduction

Risk and decision-making accompany construction projects all the time. Project managers are often challenged to choose the optimal plan to maximize the project's success. Although they benefit from the traditional apprenticeship to gain knowledge and skills from older generations, project managers strongly feel the need to develop new techniques to facilitate decision-making under uncertainty. In recent years, researchers introduced formalized methods for decision analysis to construction management. Bayesian Belief Networks (BBN) is a useful tool with potential applications in the construction industry.

Scientists in the 1980's developed a theoretical framework of BBN for decision analysis (Pearl 1988). The applications of BBN can be found in many fields such as Microsoft's products (Heckerman et al. 1995), interplanetary probes and deep space explorations (Stutz et al. 1998) and bioinformatics and medical informatics (Husmeier et al. 2005; Di Bacco et al. 2004). BBN were first introduced into the area of construction engineering and management research in the late 1990s and applied to construction performance diagnostics, which shed light on new techniques to forecasting and decision support in construction engineering and management (McCabe et al. 1998). Since then, several applications in construction have been developed, such as BBN models to facilitate the determination of the schedule and cost risks (Nasir et al. 2003; Eyers 2001; Tang 2005; Tang et al. 2007).

In this paper, the basic concept of BBN is presented and BBN establishment and procedure are discussed. One critical step applying BBN to construction management is knowledge elicitation from construction experts. The concerns in knowledge elicitation are examined. Finally, an example of BBN established upon the input from experts is demonstrated.

2. Bayesian Belief Networks

Bayesian belief networks are directed acyclic graphs (DAGs) in which the nodes represent variables, and the arcs signify the existence of dependence between the linked variables. The strength of these influences is expressed by conditional probabilities (Pearl 1988). BBN provide a knowledge base to encode the structure of relevancies as well as probabilistic relationships.

Figure 1 shows a simple BBN describing the influences and strengths between events. Abbreviations are used to represent activities: G for favourable geological condition, T for available excavator, E for excavation on schedule, R for soil removal on schedule, F for foundation on schedule. True denotes the positive state of the activity while False denotes the negative state.

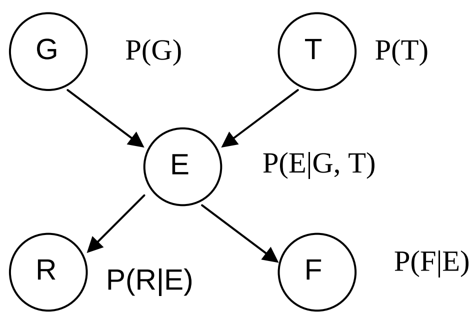

Figure 1 An example of Bayesian belief networks

Generally, both geological condition and availability of an excavator have a direct effect on the progress of excavation, but they don't affect the soil removal and foundation progress directly. (Note that the terms 'cause' and 'effect' are used casually and do not have any statistical meaning.) G and T are called the parents of E, and G and T are orphans in that they have no parents. Formally, the parents of X_i are those variables judged to have a direct influence on X_i. The degree of influence is quantified via conditional probability tables (CPT) embedded in the networks. The number of probabilities required for each node is shown in Eq. (1):

$$NP = (m-1)\prod_{i=1}^{k} n_i \tag{1}$$

where NP is the number of probabilities, *m* is the number of states of the child node, and *n* is the number of states of parent node *i*. As *i* goes from 1 to *k* representing the number of parents, the case of no parents (e.g. variable G) means that the product term drops from the equation (it does not take the value of zero) and only the prior probabilities (*m*-1) are needed. For example, node E has two states, t=True and f=False, its parent nodes G and T also have two states each. The number of probabilities needed for node E is (2-1) x 2 x 2 = 4. These probabilities are P(E=t | G=t, T=t), P(E=t | G=t, T=f), P(E=t | G=f, T=t), P(E=t | G=f, T=f). The probabilities of P(E=f | G, T) can be calculated based on the four probabilities above as they are complementary.

In the BBN in Figure 1, the state of F is affected by the state of E. When F is concerned, a phrase "…given that what is known about E" is always associated with the statement. Syntactically, this is denoted by placing E behind the conditioning bar (|) in a statement such as P(F|E)=*p*. This statement combines the notions of knowledge and belief by attributing to F a degree of belief *p*, given the knowledge E. E is also called the context of the belief in F, and the notation P(F|E) is called Bayes conditionalization. Due to the contribution of Thomas Bayes (1702-1761), the statement can be expressed in a ratio formula in Eq. (2)

$$P(F \mid E) = \frac{P(F,E)}{P(E)} \ , \tag{2}$$

which is called Bayes' Theorem and has become a definition of conditional probabilities.

In the Bayesian formalism, belief measures obey the three basic axioms of probability theory:

$$0 \le P(A) \le 1 \tag{3}$$

$$P(Sure \quad proposition) = 1 \tag{4}$$

$$P(A \quad or \quad B) = P(A) + P(B) \text{ if A and B are mutually exclusive.} \tag{5}$$

For any event A, the complement of A, denoted ¬A, consists of all outcomes in the sample space that are not in A (Ross 1987). A and ¬A must be mutually exclusive and collectively exhaustive. A proposition and its complement must be assigned a total belief of unity,

$$P(A) + P(\neg A) = 1 \ , \tag{6}$$

because one of the two statements is certain to be true.

Since any event A can be written as the union of the joint events (A and B) and (A and ¬B), their associated probabilities are given by

$$P(A) = P(A,B) + P(A,\neg B) \ , \tag{7}$$

where P(A, B) is short for P(A and B). More generally, if B_i, *i*=1, 2, …, *n*, is a set of exhaustive and mutually exclusive propositions, then P(A) can be computed from P(A, B_i), *i*=1, 2, …, *n*, using the sum

$$P(A) = \sum_i P(A, B_i) \ . \tag{8}$$

The basic expressions in the Bayesian formalism are statements about conditional probabilities. If P(A|B)=P(A), then A and B are independent. If P(A|B, C)=P(A|C), then A and B are conditionally independent given C.

Bayes' Theorem can also be written in a product form

$$P(A,B) = P(A \mid B) P(B) \ , \tag{9}$$

which is also called product rule. If the product rule is applied repeatedly, the chain rule formula can be derived. It states that if there is a set of *n* events, E_1, E_2, …, E_n, then the probability of the joint event (E_1, E_2, …, E_n) can be written as a product of *n* conditional probabilities:

$$P(E_1, E_2 \ldots, E_n) = P(E_n \mid E_{n-1}, \ldots, E_2, E_1) \ldots P(E_2 \mid E_1) P(E_1) \ . \tag{10}$$

The heart of Bayesian techniques lies in the celebrated inversion formula,

$$P(H \mid e) = \frac{P(e \mid H) P(H)}{P(e)}, \tag{11}$$

which states that the belief of a hypothesis (H) accorded upon obtaining evidence (*e*) can be computed by multiplying the belief P(H) by the likelihood P(*e*|H) that *e* will materialize if H is true. P(H|*e*) is called the posterior probability or simply posterior, and P(H) is called the prior probability or prior. Applying Eq. (6) to require that P(H|*e*) and P(¬H|*e*) sum to unity, the denominator P(*e*) in Eq. (11) can be derived and computed by $P(e) = P(e \mid H) P(H) + P(e \mid \neg H) P(\neg H)$.

The importance of the inversion formula is that it expresses a quantity P(H|*e*) – that people often find hard to assess – in terms of quantities that often can be drawn directly from their experiential knowledge. For example, if a construction site engineer is asked "What is the likelihood that it is raining when you experience low productivity on site?", s/he might have difficulties in estimating the probability, because low productivity is not a cause of rain. But if s/he is asked "What is the chance of experiencing low productivity on site when it rains?", s/he can estimate the probability because the rain may lead to low productivity and therefore is of keen interest to the site engineer.

3. Establishing Bayesian Belief Networks for Construction Management

BBN make complex system models operational by showing relationships explicitly **-** the structure and degree of dependence, which will facilitate the research and practice of decision making. Mathematically, BBN can be constructed based on the following procedure.

Consider a finite set **X** = {X_1, …, X_n} of discrete random variables where each variable X_i may take on values from a finite set, denoted by *Val*(X_i). A BBN consists of (1) a network structure *S* that encodes a set of conditional independence assertion about variables in **X**, and (2) a set *P* of local probability distributions associated with each variable. Formally, a BBN can be denoted as *B* = (S, Θ). *S* is a directed acyclic graph whose vertices correspond to the random variables X_1, …, X_n. Each variable X_i is independent of its non-descendants given its parents in *S*. Θ represents the set of parameters that quantifies the network. It contains a parameter $\theta_{X_i|\mathbf{Pa}(X_i)} = P(x_i|\mathbf{pa}(X_i))$ for each possible value x_i of X_i, and **pa**(X_i) of **Pa**(X_i). Here **Pa**(X_i) denotes the set of parents of X_i in *S* and **pa**(X_i) is a particular instantiation of the parents. A BBN *B* specifies a unique joint probability distribution over **X** given by Eq. (12).

$$P_B(X_1, \ldots, X_n) = \prod_{i=1}^{n} P_B(X_i \mid Pa(X_i)) \tag{12}$$

The probabilities encoded by a BBN may be Bayesian (based on opinion) or physical (based on data). When building BBN from prior knowledge alone, the probabilities will be Bayesian. When learning these networks from data, the probabilities will be physical although their values may be uncertain.

The evaluation of a BBN is based upon probability calculus, including Bayes theorem. Suppose the background knowledge ξ is known, the uncertainty about Θ can be expressed using the probability density function $p(\theta|\xi)$. Given D as the set of the observations, the probability distribution for Θ can be obtained as shown in Eq. (13). The probability distributions $p(\theta|\xi)$ and $p(\theta|D,\xi)$ are commonly referred to as the *prior* and *posterior* for Θ respectively.

$$p(\theta \mid D, \xi) = \frac{p(\theta \mid \xi)\, p(D \mid \theta, \xi)}{p(D \mid \xi)}, \tag{13}$$

where $p(D \mid \xi) = \int p(D \mid \theta, \xi)\, p(\theta \mid \xi)\, d\theta$

According to Heckerman (1995), the Bayesian and classical approaches are fundamentally different methods for learning from data. In the classical approach, θ is fixed though unknown, all data sets of size N that may be generated by sampling from a certain distribution determined by θ are imagined. Each data set D will occur with some probability $p(D|\theta)$ and will produce an estimate $\theta^*(D)$. To evaluate an estimator, the expectation and variance of the estimate with respect to all such data sets are computed using Eq. (14) and Eq. (15). Then an estimator is chosen to balance the bias (θ-$E_{p(D|\theta)}(\theta^*)$) and variance of these estimates over the possible values for θ. Finally, this estimator is applied to the data set actually observed. A commonly-used estimator is the maximum-likelihood (ML) estimator, which selects the value of θ that maximizes the likelihood $p(D|\theta)$.

$$E_{p(D|\theta)}(\theta^*) = \sum_D p(D \mid \theta)\theta^*(D) \tag{14}$$

$$Var_{p(D|\theta)}(\theta^*) = \sum_D p(D \mid \theta)(\theta^*(D) - E_{p(D|\theta)}(\theta^*))^2 \tag{15}$$

In contrast, in the Bayesian approach, D is fixed, and all possible values of θ from which this data set could have been generated are imagined. Given θ, the estimate of the physical probability is just θ itself. Nonetheless, people are uncertain about θ, and so the final estimate, shown in Eq. (16), is the expectation of θ with respect to people's posterior beliefs about its value. The expectations in $E_{p(D|\theta)}(\theta^*)$ and $E_{p(\theta|D,\xi)}(\theta)$ are different and lead to different estimates, which indicate different approaches in learning BBN from data.

$$E_{p(\theta|D,\xi)}(\theta) = \int \theta p(\theta \mid D, \xi)\, d\theta \tag{16}$$

Generally, there are two groups of approaches to learn BBN from data: constraint-based search and Bayesian methods (Cheng & Greiner 1999; Friedman et al. 1999; Dash& Druzdzel 1999). In the constraint-based approach, an effort is made to estimate properties of conditional independence among the attributes in the data. Usually this is done using a statistical hypothesis test. Then, a network that exhibits the observed dependencies and independencies is built. In the Bayesian approach, an effort is made to define a statistically motivated score that describes the fitness of each possible structure to the observed data. These scores include Bayesian scores and minimum description length (MDL) scores, which can be computed by various algorithms (Lam & Segre 2002). Then, a structure that maximizes the score is identified. In general, this is an NP-hard problem and thus heuristic methods are needed for score maximization.

The probabilistic information is available from various sources, which includes statistical data, literature and human experts (Druzdzel and Van der Gaag 2000). In many areas of the construction industry, historical and systematic data are not available. Therefore, elicitation of the knowledge from domain experts for constructing Bayesian belief network structures and quantifying the strength between the nodes in the networks becomes the preamble and provides the foundation for the learning of Bayesian belief networks.

4. Knowledge Elicitation for Bayesian Belief Networks in Construction Management

Expert knowledge is an informed opinion based on the expert's training and experience. In BBN construction, a domain expert's knowledge and opinion can be valuable in establishing the structure of the network, assessing the probabilities required, fine tuning probabilities from other sources and verifying probability tables in the network. The knowledge of experts can be of greater potential value and used more efficiently and accurately if it is expressed in a probabilistic form. However, a domain expert may not have the expertise in statistics and probability. Therefore, the goal of elicitation is to make it as easy as possible for domain experts to tell researchers what they believe, in probabilistic terms, while reducing how much they need to know about probability theory to do so.

Based on the knowledge elicited from experts of airport development, the authors developed a BBN focused on the decision of whether to build a new airport terminal building or renovate the existing terminal given certain characteristics of an operating airport. Because of a shortage of data from which a model could be developed, reliance was placed on expert knowledge for both model structure and probabilistic quantification. Personal interview was selected to elicit the knowledge from the airport development experts.

In an existing airport, the airport authority must decide whether to renovate the existing building or to construct a new one. Through examination with 7 airport experts, referred to as E1, E2, …, and E7, five critical factors influencing the decision for a passenger building (PB) were identified and conditional probabilities were elicited, as shown in Table 1. Based on the relationships between the factors provided by the domain experts, a BBN was established as shown in Figure 2.

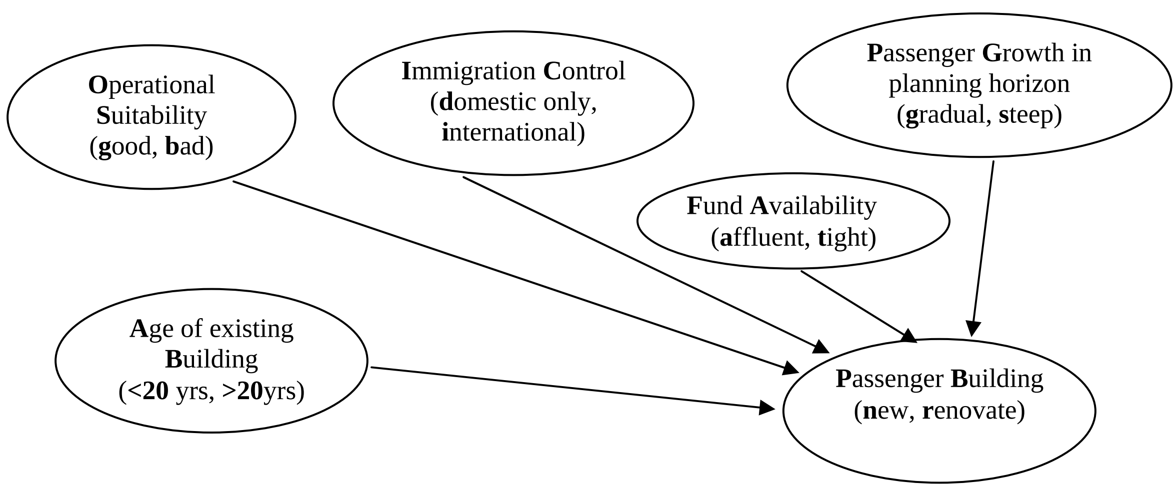

Figure 2 Bayesian belief network for passenger building construction

Having determined the structure and the associated conditional probability table of the BBN, the authors completed one cycle of a BBN construction. The constructed BBN will be tested and calibrated against real projects. Once the model is accepted, decision makers can use it to help determine the influence of various factors on the decision under different scenarios. The model can provide a guidance to identify critical factors by comparing probabilities in different scenarios so that efforts can be put in the most cost-effective area.

Table 1 Conditional probabilities (%) elicited from the experts

Item	E1	E2	E3	E4	E5	E6	E7
Pr(PB=n | OS=b, PG=s, IC=i, FA=a, AB=>20)	90	90	90	90	80	99	99
Pr(PB=n | OS=b, PG=s, IC=i, FA=a, AB=<20)	80	70	90	80	60	99	90
Pr(PB=n | OS=b, PG=s, IC=i, FA=t, AB=>20)	50	70	90	80	60	80	80
Pr(PB=n | OS=b, PG=s, IC=i, FA=t, AB=<20)	40	50	80	70	40	70	70
Pr(PB=n | OS=b, PG=s, IC=d, FA=a, AB=>20)	70	70	80	80	50	90	60
Pr(PB=n | OS=b, PG=s, IC=d, FA=a, AB=<20)	60	50	70	80	30	80	50
Pr(PB=n | OS=b, PG=g, IC=i, FA=a, AB=>20)	30	60	90	80	70	70	70
Pr(PB=n | OS=b, PG=g, IC=i, FA=a, AB=<20)	20	40	70	70	40	70	60
Pr(PB=n | OS=g, PG=s, IC=i, FA=a, AB=>20)	30	90	90	80	60	90	80
Pr(PB=n | OS=g, PG=s, IC=i, FA=a, AB=<20)	20	80	90	70	50	80	70
Pr(PB=n | OS=b, PG=s, IC=d, FA=t, AB=>20)	20	60	80	70	60	80	70
Pr(PB=n | OS=b, PG=s, IC=d, FA=t, AB=<20)	10	40	70	60	50	70	50
Pr(PB=n | OS=b, PG=g, IC=i, FA=t, AB=>20)	20	50	80	70	60	60	60
Pr(PB=n | OS=b, PG=g, IC=i, FA=t, AB=<20)	10	20	70	60	40	50	50
Pr(PB=n | OS=g, PG=s, IC=i, FA=t, AB=>20)	30	90	80	70	50	70	70
Pr(PB=n | OS=g, PG=s, IC=i, FA=t, AB=<20)	20	70	70	60	30	60	50
Pr(PB=n | OS=b, PG=g, IC=d, FA=a, AB=>20)	50	40	30	80	60	80	50
Pr(PB=n | OS=b, PG=g, IC=d, FA=a, AB=<20)	40	20	20	70	50	70	40
Pr(PB=n | OS=b, PG=g, IC=d, FA=t, AB=>20)	20	50	30	70	50	30	30
Pr(PB=n | OS=b, PG=g, IC=d, FA=t, AB=<20)	10	30	20	60	40	20	20
Pr(PB=n | OS=g, PG=s, IC=d, FA=a, AB=>20)	20	90	70	80	60	80	50
Pr(PB=n | OS=g, PG=s, IC=d, FA=a, AB=<20)	10	70	60	70	40	70	40
Pr(PB=n | OS=g, PG=s, IC=d, FA=t, AB=>20)	10	60	60	70	50	70	30
Pr(PB=n | OS=g, PG=s, IC=d, FA=t, AB=<20)	1	40	60	60	20	60	20
Pr(PB=n | OS=g, PG=g, IC=i, FA=a, AB=>20)	10	80	80	70	60	70	40
Pr(PB=n | OS=g, PG=g, IC=i, FA=a, AB=<20)	1	60	70	70	30	60	30
Pr(PB=n | OS=g, PG=g, IC=i, FA=t, AB=>20)	10	70	60	70	50	60	30
Pr(PB=n | OS=g, PG=g, IC=i, FA=t, AB=<20)	1	50	50	60	20	50	20
Pr(PB=n | OS=g, PG=g, IC=d, FA=a, AB=>20)	1	50	60	70	40	60	20
Pr(PB=n | OS=g, PG=g, IC=d, FA=a, AB=<20)	1	70	30	60	20	50	20
Pr(PB=n | OS=g, PG=g, IC=d, FA=t, AB=>20)	1	50	40	60	50	20	10
Pr(PB=n | OS=g, PG=g, IC=d, FA=t, AB=<20)	1	30	30	40	10	10	1

5. Conclusion

Bayesian belief networks provide a means to graphically expose the interrelationship and probabilistically quantify the strength in a system. The theory of BBN construction is solid and knowledge elicitation from domain experts is practical. The expertise from domain experts captured in BBN can be used in similar projects in the future. The example has demonstrated that the potential of BBN applied to construction management is promising.

6. References

Cheng, J & Greiner,R. (1999). Comparing Bayesian Network Classifiers, Proceedings of the Fifteenth Annual Conference on Uncertainty in Artificial Intelligence (UAI-99), 101-108.

Dash, D. & Druzdzel, M.J. (1999). A Hybrid Anytime Algorithm for the Construction of Causal Models from Sparse Data, Proceedings of the Fifteenth Annual Conference on Uncertainty in Artificial Intelligence (UAI-99), pp142-149.

Di Bacco, M., D'Amore, G. & Scalfari, F. (2004). Applied Bayesian Statistical Studies in Biology and Medicine. Kluwer Academic Publishers, Boston.

Druzdzel, M. J. & Van der Gaag, L. C. (2000). Building probabilistic networks: "Where do the numbers come from?" Guest Editors' introduction, IEEE Transactions on Knowledge and Data Engineering, 12(4), 481-486.

Eyers, K. (2001), Belief network analysis of direct cost risk in building construction, MASc. thesis, University of Toronto.

Friedman, N., Nachman, I. & Peer, D. (1999). Learning Bayesian Network Structure from Massive Datasets: The "Sparse Candidate" Algorithm, Proceedings of the Fifteenth Annual Conference on Uncertainty in Artificial Intelligence (UAI-99), pp206-215.

Heckerman, D. & Geiger, D. (1995). Learning Bayesian Networks, Technical Report MSR-TR-95-02, Microsoft Research.

Heckerman, D. (1995). A Tutorial on Learning with Bayesian Networks, Technical Report MSR-TR-95-06, Microsoft Research.

Husmeier, D., Dybowski, R. & Roberts, S. (Eds.) (2005). Probabilistic Modeling in Bioinformatics and Medical Informatics, Springer-Verlag, London.

Lam, W. & Segre, A. M. (2002). A distributed learning algorithm for Bayesian inference networks, IEEE Transactions on Knowledge and Data Engineering, 14(1), 93-105.

McCabe, B., AbouRizk, S. M. and Goebel, R. (1998). Belief networks for construction performance diagnostics. J. Computing in Civil Engrg., ASCE, 12(2), 93-100.

Nasir, D., McCabe, B. and Hartono, L. (2003), Evaluating risk in construction-schedule model (ERIC-S): construction schedule risk model, J. Constr. Eng. Manage., ASCE, 129(5), 518-527.

Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems: networks of plausible inference, Morgan Kaufmann.

Ross, S. M. (1987) Introduction to probability and statistics for engineers and scientists, John Wiley & Sons.

Stutz, J., Taylor ,W. & Cheeseman, P. (1998). AutoClass C - General Information, NASA, Ames Research Center.

Tang, Z. (2005). "Developing Complete Conditional Probability Tables From Fractional Data For Bayesian Belief Networks In Engineering Decision Making ", A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy, Department of Civil Engineering, University of Toronto.

Tang, Z. & McCabe, B. (2007). "Developing Complete Conditional Probability Tables from Fractional Data for Bayesian Belief Networks" J. Computing in Civil Eng., ASCE, 21(4).