دانشگاه شهیدبهشتی
دانشکده علوم ریاضی
گروه ریاضی

بازه تغییر احتمال پسین شبکه بیزی براساس پیش‌بینی شواهدآینده

علی مشهدی علیپوری

تیر۱۳۹۸

برای ارائه در درس مدل‌های گرافی احتمالاتی –مقطع کارشناسی ارشد ریاضی کاربردی–علوم داده ها

Section 1: introducing problem

- Most applications of Bayesian networks focus on calculating posterior probabilities over variables of interest given observations of other variables.

- Because not all observations are available at the outset, one would like to know how future observations may lead to changes of these posterior probabilities.

Section 1: introducing problem

- For example, a probability of a disease in a patient with little or no symptoms or test results is close to disease prevalence in general population. This probability can go up or down, depending on the patient's specifics.

Section 1: introducing problem

- A user of a probabilistic decision support system might want to know where this probability can go as more information becomes available.

We propose to address this problem by deriving variation intervals over posterior probabilities.

Section 2: previous activities

- Most of the literature on uncertainty in results of Bayesian network inference focuses on the impact of possible imprecision in parameters of the network. Such uncertainty can be captured by means of error bars or uncertainty intervals (e.g., work by Donald and Mengersen (2014) or Van Allen et al. (2008)).

Section 2: previous activities

- If the imprecision in parameters can be expressed by intervals, it can be propagated over the model to derive uncertainty intervals over results (Cano et al., 1993; Fagiuoli and Zaalon, 1998).

Section 2: previous activities

- Uncertainty over results has also been a focus of sensitivity analysis, which amounts to studying the impact of small changes in individual model parameters on the result. For example, Laskey (1995) describes the derivation of error bars for probability assessment.

Section 2: previous activities

- Even though the question posed in this paper is useful and asked by users of probabilistic decision support systems, we have not found any literature analyzing the uncertainty intervals for posterior probabilities in anticipation of future observations.

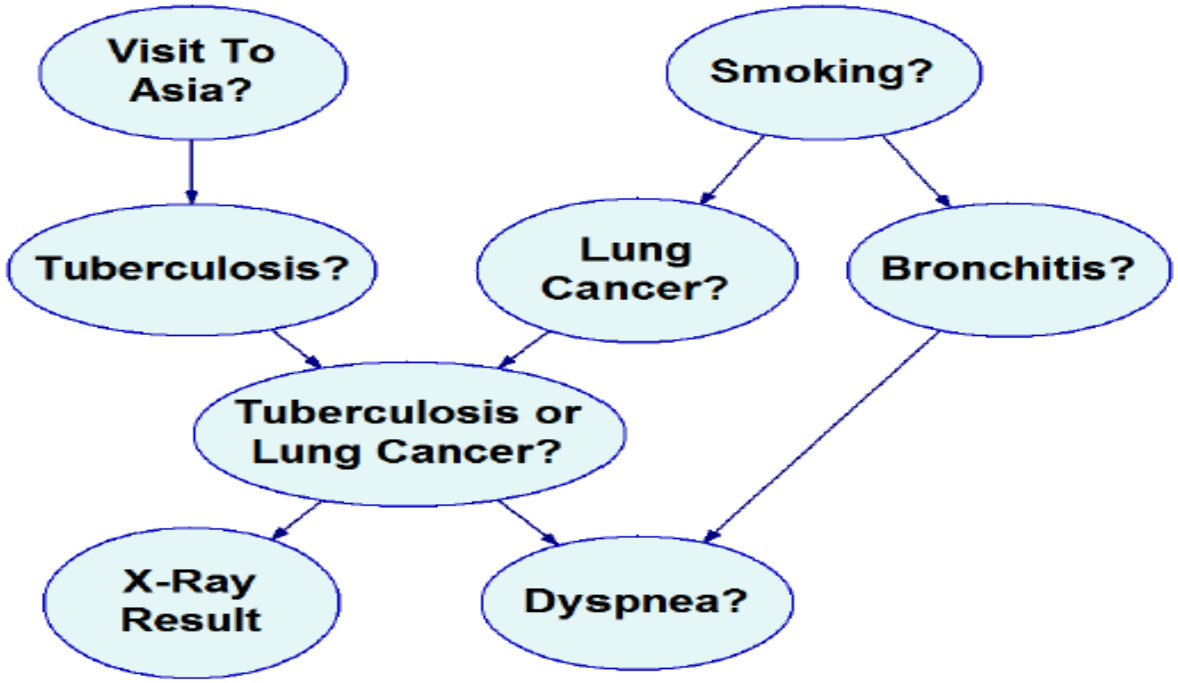


Figure 1: The ASIA Bayesian network (Lauritzen and Spiegelhalter, 1988).

Section 3: overall review of method

CI Sample All Observable

Input : BN $(\mathcal{G}, \Theta)$, target variable V_t, target assignment $v_{t,j}$, evidence $\mathbf{E}$, unobserved variables $\mathbf{S}_U$, number of samples N, confidence level $1 - \alpha$

Output: Sample H of possible probabilities $\Pr(v_{t,j}|\mathbf{E}^*)$, variation interval (p_L, p_U)

$H \leftarrow \emptyset$
for $k = 1, \ldots, N$ **do**
  $\mathbf{E}^* \leftarrow \mathbf{E}$
  foreach $V_i \in \mathbf{S}_U$ **do**
    Calculate $\Pr(V_i|\mathbf{E}^*)$
    Draw $v_{i,k} \sim \Pr(V_i|\mathbf{E}^*)$
    $\mathbf{E}^* \leftarrow \mathbf{E}^* \cup \{v_{i,k}\}$
  end
  Calculate $\Pr(V_t|\mathbf{E}^*)$
  $H \leftarrow (H, \Pr(v_{t,j}|\mathbf{E}^*))$
end
Construct $1 - \alpha$ variation interval (p_L, p_U) using sample H

Section 3: overall review of method

CI Sample All Observable algorithm explanation:

- Algorithm iterates through the set of all observable variables to assign a value to each unobserved variable (line 4).
- To draw an outcome for a variable, it calculates the posterior probability distribution over its outcomes given evidence (line 5).
- Then, it samples an outcome from the calculated posterior probability distribution (line 6) .
- Having outcomes assigned to all the observable variables, the algorithm calculates the posterior probability of the pursued outcome of the target variable, which amounts to one sample (lines 9-10).
- Based on the sample, we derive a variation interval (empirical confidence interval) over the posterior probability of the pursued outcome (line 12).

Section 3: overall review of method

CI Sample Extended Markov Blanket

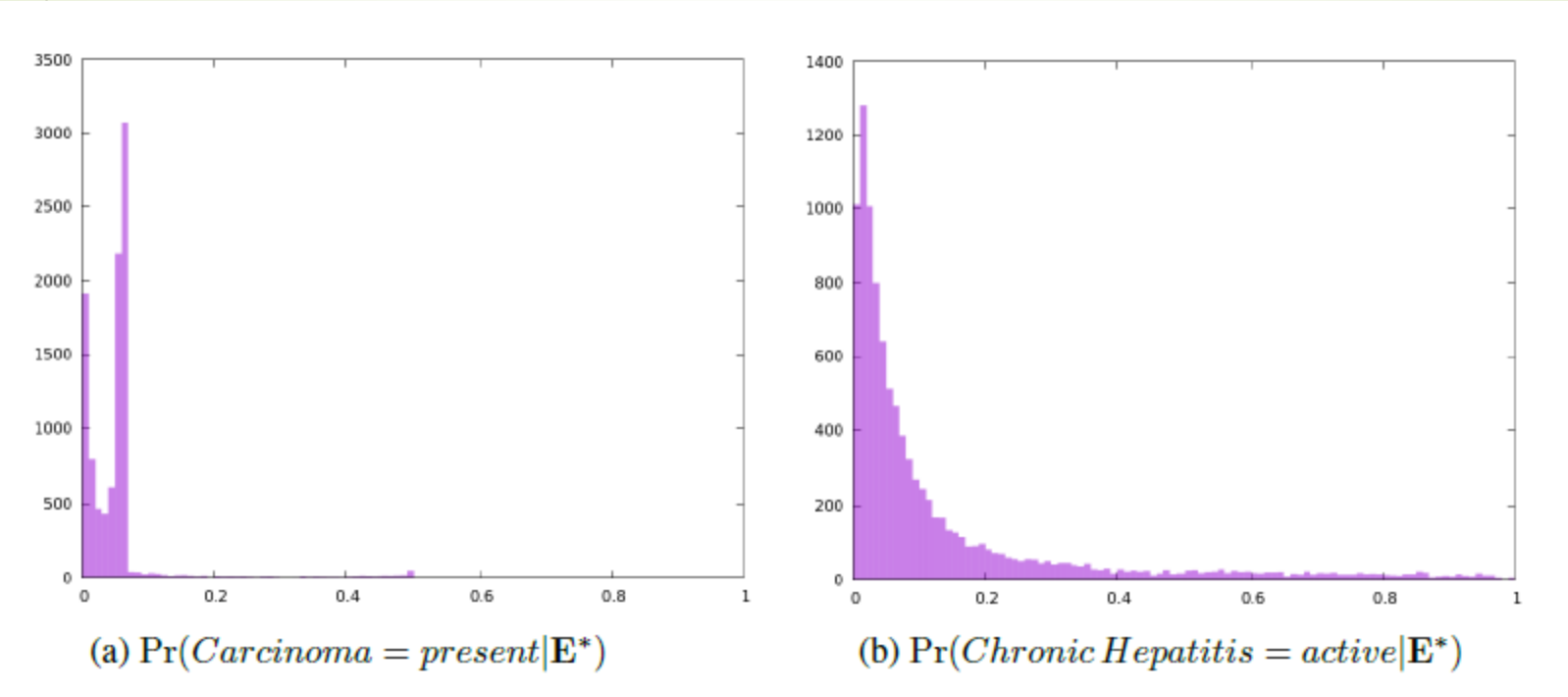
Input : BN $(\mathcal{G}, \Theta)$, target variable V_t, target assignment $v_{t,j}$, evidence $\mathbf{E}$, observable variables $\mathbf{S}$, number of samples N, confidence level $1 - \alpha$
Output: Sample H of possible probabilities $\Pr(v_{t,j}|\mathbf{E}^*)$, variation interval (p_L, p_U)

$\mathbf{M}^*(V_t) \leftarrow \mathbf{M}(V_t) \cap \mathbf{S}$
$\mathbf{A} \leftarrow \mathbf{M}(V_t) \setminus \mathbf{S}$
$\mathbf{A}_D \leftarrow \emptyset$
while $\mathbf{A} \neq \emptyset$ **do**
  pick any V_i from $\mathbf{A}$
  $\mathbf{A} \leftarrow \mathbf{A} \setminus \{V_i\}$
  $\mathbf{A}_D \leftarrow \mathbf{A}_D \cup \{V_i\}$
  $\mathbf{A} \leftarrow \mathbf{A} \cup (\mathbf{M}(V_i) \setminus (\mathbf{S} \cup \mathbf{A}_D))$
  $\mathbf{M}^*(V_t) \leftarrow \mathbf{M}^*(V_t) \cup (\mathbf{M}(V_i) \cap \mathbf{S})$
end
$H \leftarrow \emptyset$
for $k = 1, \ldots, N$ **do**
  $\mathbf{E}^* \leftarrow \mathbf{E}$
  foreach $V_i \in \mathbf{M}^*(V_t) \setminus \mathbf{S}_O$ **do**
    Calculate $\Pr(V_i|\mathbf{E}^*)$
    Draw $v_{i,k} \sim \Pr(V_i|\mathbf{E}^*)$
    $\mathbf{E}^* \leftarrow \mathbf{E}^* \cup \{v_{i,k}\}$
  end
  Calculate $\Pr(V_t|\mathbf{E}^*)$
  $H \leftarrow (H, \Pr(v_{t,j}|\mathbf{E}^*))$
end
Construct $1 - \alpha$ variation interval (p_L, p_U) using sample H

Finding extended markov blankets of target variables

CI Sample All Observable algorithm

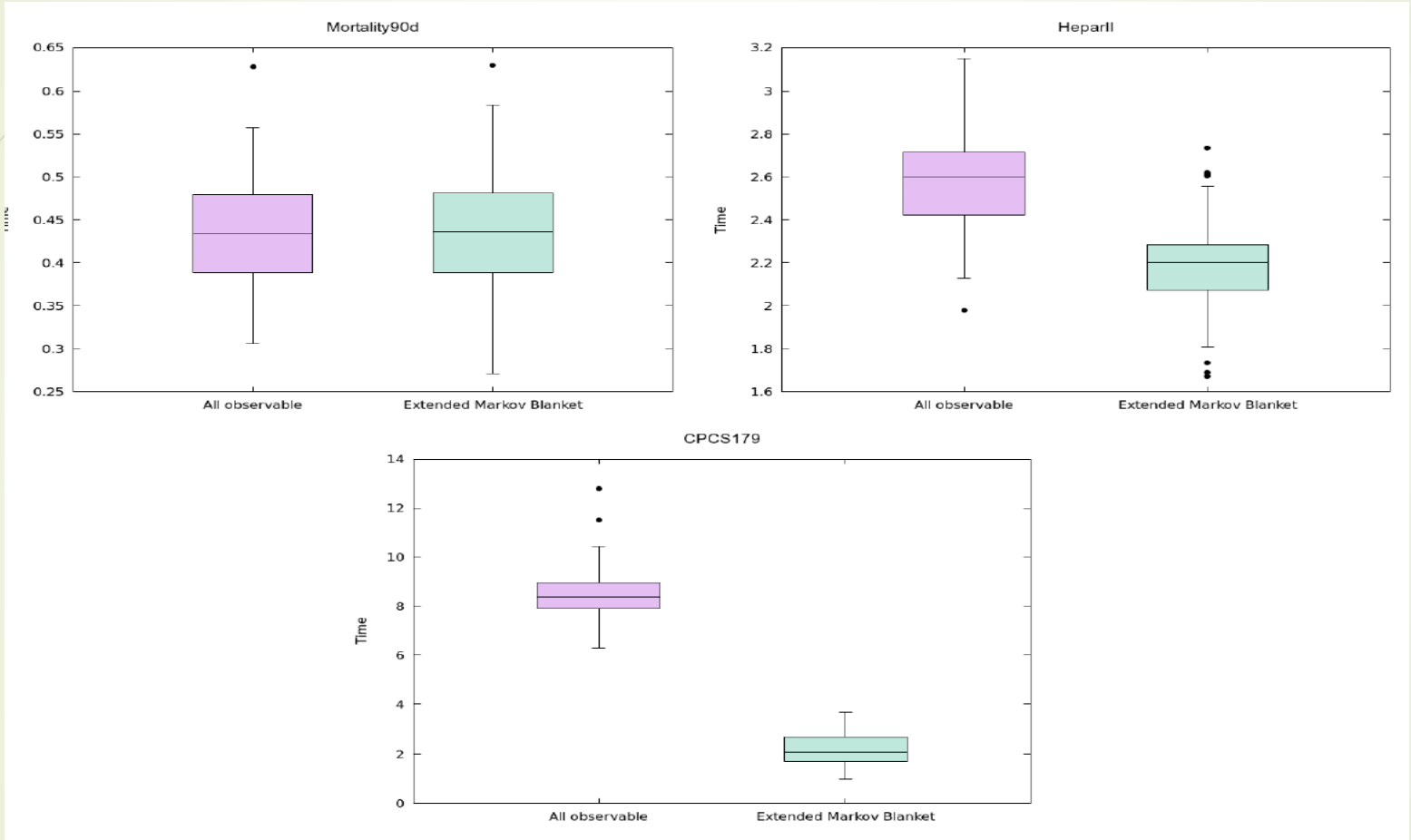
Section 3: overall review of method



(a) $\Pr(Carcinoma = present|\mathbf{E}^*)$ (b) $\Pr(Chronic\,Hepatitis = active|\mathbf{E}^*)$

- Histograms representing samples of posterior probabilities values given one assignment to a variable in HEPAR II model

Section 3: overall review of method

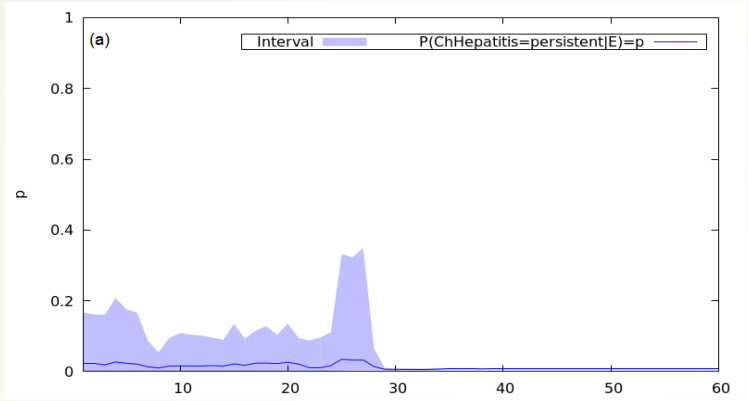


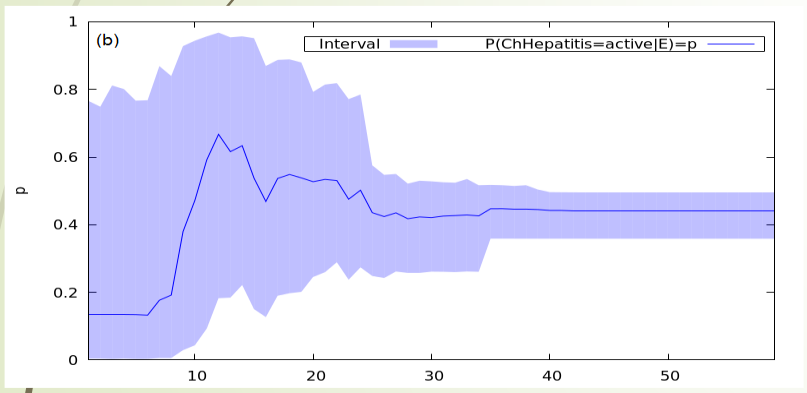
- Box plots comparing computation times of variation intervals for posterior probabilities with both versions of the algorithm (measured in seconds)

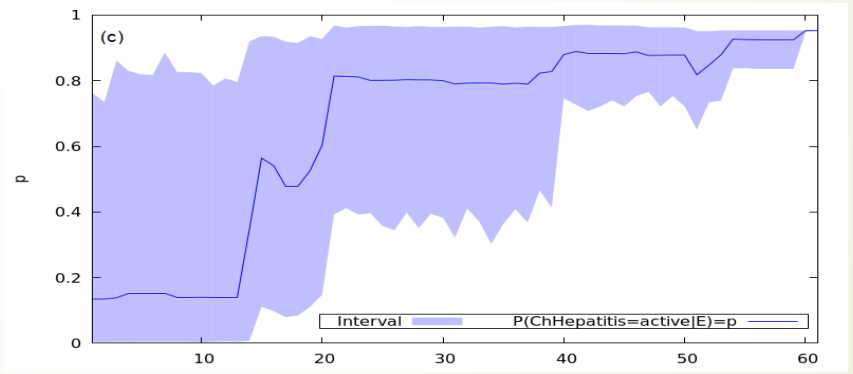
Section 4: Examples of the derived variation intervals

Figure(on next slide) shows eight examples of 95% variation intervals over the posterior probability of Chronic Hepatitis being persistent (a), Chronic Hepatitis being active for two diferent cases (b-c), PBC (primary biliary cirrhosis) (d) being present, Toxic Hepatitis being present (d), Cirrhosis being compensated for three di_erent cases (f-h). There are 61 possible observations (referring to risk factors, symptoms, and test results in the Hepar II model) for each case and they are made individually from left to right. We used a fixed number of N = 1; 000 samples in each experiment. The solid line running from left to right demonstrates the development of the probability of the target event in question as new observations are made. The area around the probability line shows the variation interval over the probability at each point in time. Please note that the variation intervals start by being very wide in the beginning, which corresponds to the situation when nothing about the patient is known. As more and more evidence is accumulated, the variation intervals narrow, to the point of becoming either a point probability (when all possible 61 observations have been made) or a fixed interval, when some of the observations have never been made in a patient's case.

Section 4: Examples of the derived variation intervals

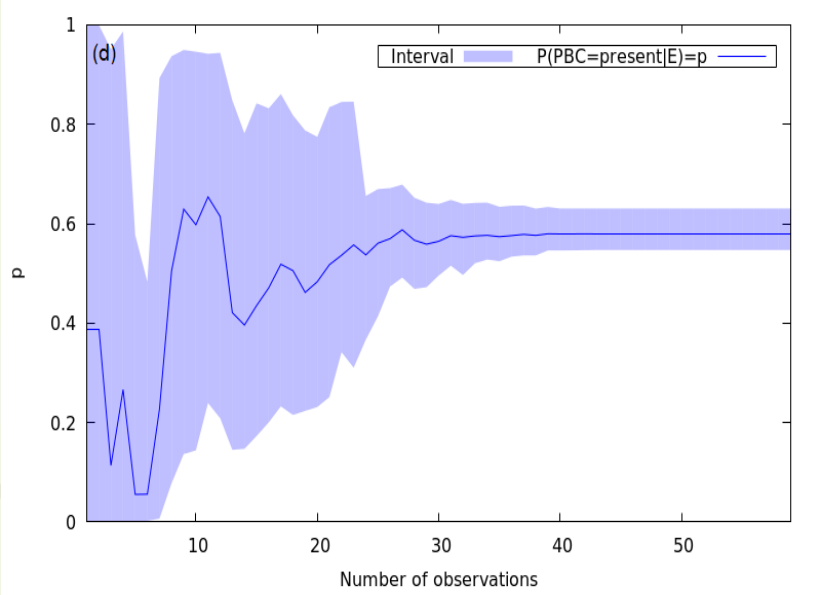


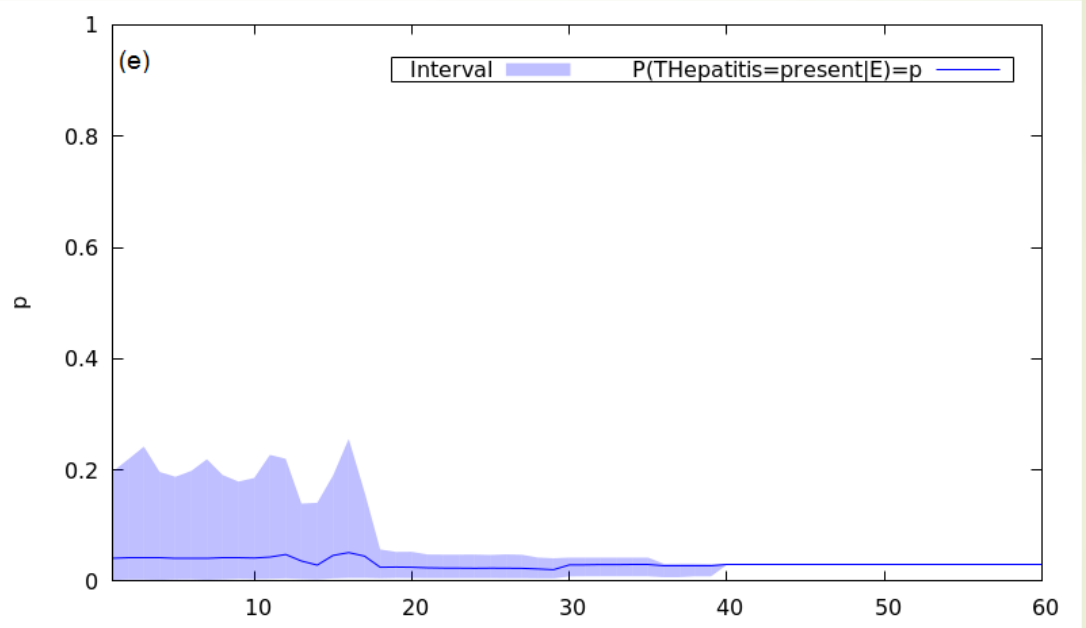




Examples of 95% variation intervals over the posterior probability of Chronic Hepatitis being persistent (a), Chronic Hepatitis being active (b-c) in Hepar II model

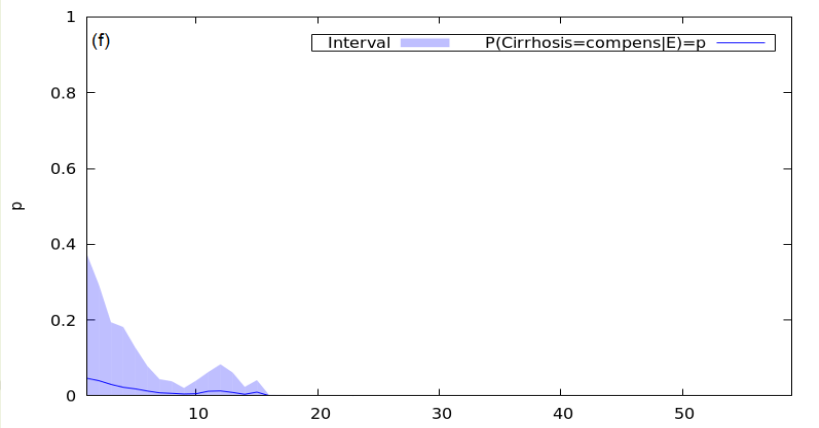
Section 4: Examples of the derived variation intervals

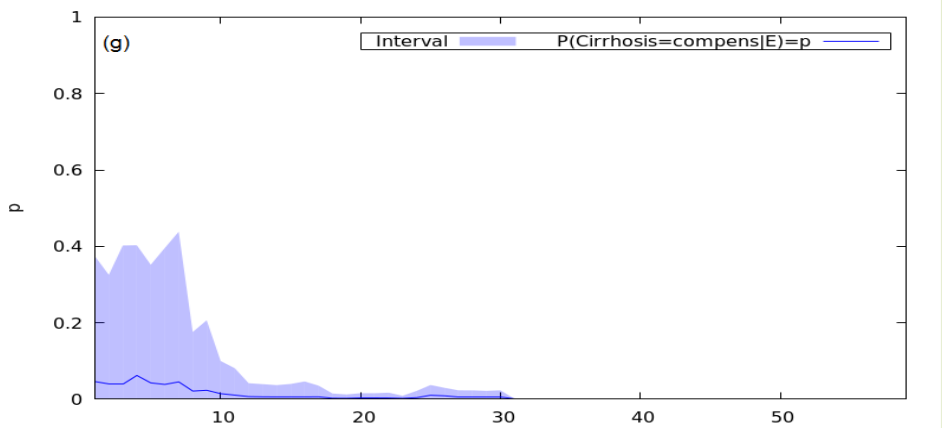


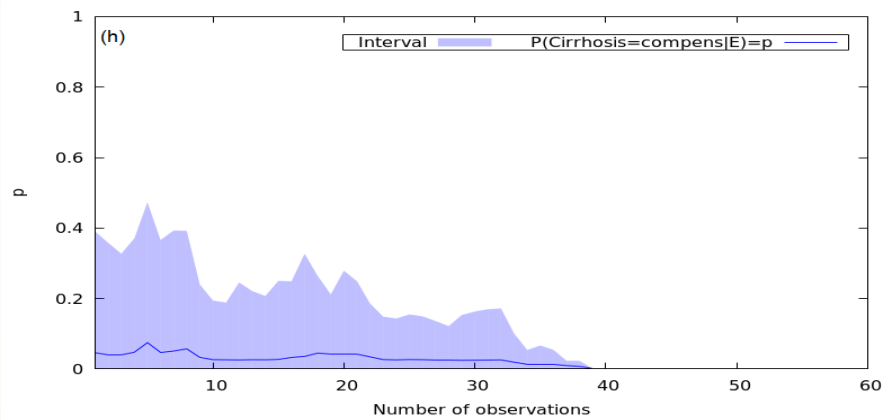


Examples of 95% variation intervals over the posterior probability of PBC (primary biliary cirrhosis) being present (d), Toxic Hepatitis being present (e) in Hepar II model

Section 4: Examples of the derived variation intervals







Examples of 95% variation intervals over the posterior probability ofCirrhosis being compensated (f-g-h) in the Hepar II model

Section 5: SIMULATION SETUP

In the simulation, I used Hepar II Bayesian model. I identified variables of interest (target variable). Models Hepar II include only one target variable. The node associated with target variable variable in the structure of this model is a predecessor for each of the other nodes as the model follows augmented naive Bayes (ANB) structure. Below Table presents model used in this simulation with some statistics and target variables identified for the purpose of the simulation.

Model	# nodes	# arcs	# targets	target variables
HEPAR II (Oniśko et al., 2000)	70	123	9	THepatitis, ChHepatitis, PBC, fibrosis, Steatosis, Cirrhosis, Hyperbilirubinemia, RHepatitis, carcinoma

Section 5: SIMULATION SETUP

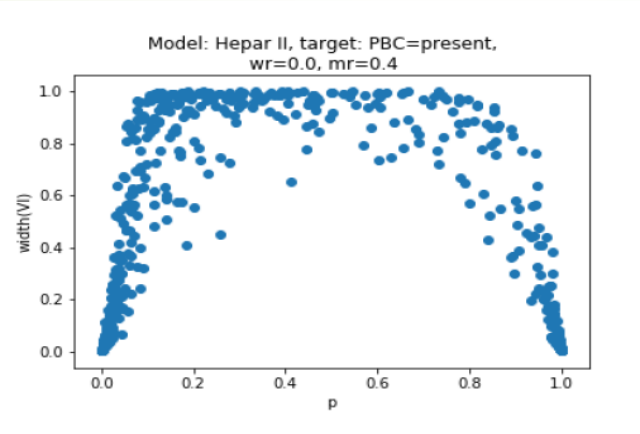
For model, I generated a set of 8,000 cases by means of probabilistic logic sampling (Henrion, 1988). I removed the information about variables at various rates m∈{0.0 ,0.2, 0.4 ,0.6 } to simulate missing values in the data. I also distorted observations randomly at various rates w ∈{0.0 ,0.2, 0.4 ,0.6 } to simulate erroneous information in the data. As a result, I obtained 500 cases for each pair of m and w values For each case E, the target variable (associated with the model) Vi and its value vij , I calculated the posterior probability p = Pr(Vi = vij |E), and the length of the variation interval over the posterior probability Pr(Vi = vij |E).

Section 6: OBSERVATIONS

Figures below shows the results of the experiment. The first observation that I made based on the plot presenting the length of intervals against the posterior probability p. variation intervals tend to get tighter as p approaches the values of zero or one. I present an example of this pattern in Figure 9. In some of the plots, I got a similar pattern, but truncated, which is a result of the fact that a posterior probability of an event may not get the value close enough to either of the ends of the interval [0; 1], which is the feature of the modeled variable. Thus, I produced scatter plots of the length of variation interval against d(p), where

$$\mathbf{d(p) = min(p; 1 - p)} ;$$

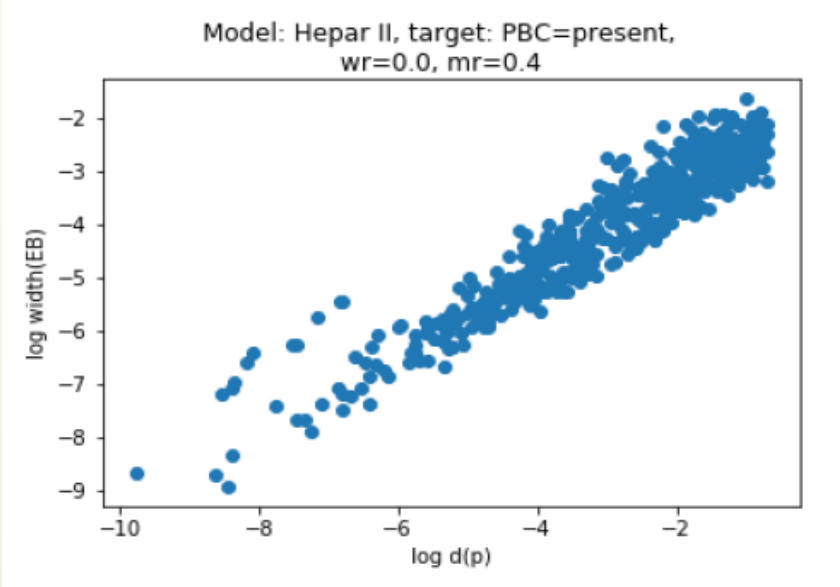
which is the distance between the value p = Pr(Vi = vij | E) and the closer end of the interval[0; 1].



An example of scatter plot showing how the interval length depends on the posteriorprobability p = Pr(PBC = present|E). The right plot shows the relationship between p and the length of the variation interval over the value of p

Section 6: OBSERVATIONS

Figure below presents several examples in logarithmic scale. Reviewing these plots, we can observe an exponential relationship between these values d(p) and the lengths of the intervals. This patterns seem to be stronger. For some target values in some models, I observed that d(p) interacts in this way with length of the variation interval.



Scatter plots showing how the variation interval depends on the posterior probability d(p) interact in log-log scale

Section 7: CONCLUSION

There seem to be an exponential relationship among lengths of variation intervals and the posterior probability (its distance from the ends of [0; 1] range). I could find plenty of examples, where lengths of variation intervals are dependent on the posterior probability value p = Pr(Vi = vij |E), although lengths of intervals are not perfectly explained by probability probability p, especially when the value of p is far from the ends of the interval [0; 1].

Section 8: references

- J. Cano, M. Delgado, and S. Moral. An axiomatic framework for propagating uncertainty in directed acyclic networks. International Journal of Approximate Reasoning, 8(4):253–280, 1993.
- G. F. Cooper. The computational complexity of probabilistic inference using Bayesian belief networks. Artificial Intelligence, 42:393–405, 1990.
- M. R. Donald and K. L. Mengersen. Methods for constructing uncertainty intervals for queries of Bayesian nets. Australian & New Zealand Journal of Statistics, 56(4):407–427, 2014. ISSN 1467-842X.
- M. J. Druzdzel. SMILE: Structural Modeling, Inference, and Learning Engine and GeNIe: A development environment for graphical decision-theoretic models. In Proceedings of the Sixteenth National Conference on Artificial Intelligence (AAAI–99) and the eleventh Innovative Applica-tions of Artificial Intelligence Conference (IAAI-99), pages 902–903, Menlo Park, CA, July 18–22 1999. AAAI Press/The MIT Press. ISBN 0–262–51106–1.
- J. Pearl. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann, 1988.
- M. Pradhan, G. Provan, B. Middleton, and M. Henrion. Knowledge engineering for large belief networks. In Proceedings of the Tenth International Conference on Uncertainty in Artificial Intelligence, pages 484–490. Morgan Kaufmann Publishers Inc., 1994.
- T. Van Allen, A. Singh, R. Greiner, and P. Hooper. Quantifying the uncertainty of a belief net response: Bayesian error-bars for belief net inference. Artificial Intelligence, 172(4-5):483–513,2008.

Section 8: references

- E. Fagiuoli and M. Zaffalon. 2U: An exact interval propagation algorithm for polytrees with binary variables. Artificial Intelligence, 106(1):77–107, 1998.
- M. Henrion. Propagation of uncertainty by probabilistic logic sampling in Bayes' networks. In Uncertainty in Artificial Intelligence, volume 2, pages 149–164, 1988.
- M. K. Kanwar, L. C. Lohmueller, R. L. Kormos, N. A. Loghmanpour, R. L. Benza, R. J. Mentz,
- S. H. Bailey, S. Murali, and J. F. Antaki. Low accuracy of the HeartMate risk score for predicting mortality using the INTERMACS registry data. ASAIO Journal, 63(3):251–256, 2017.
- K. B. Laskey. Sensitivity analysis for probability assessments in Bayesian networks. IEEE Transactions on Systems, Man, and Cybernetics, 25(6):901–909, 1995.
- S. L. Lauritzen and D. J. Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems. Journal of the Royal Statistical Society. Series B (Methodological), pages 157–224, 1988.
- A. Oni´sko, M. J. Druzdzel, and H. Wasyluk. Learning Bayesian network parameters from small data sets: Application of Noisy-OR gates. International Journal of Approximate Reasoning, 27 (2):165–182, 2001.