

Evidence Evaluation: a Study of Likelihoods and Independence

Silja Renooij

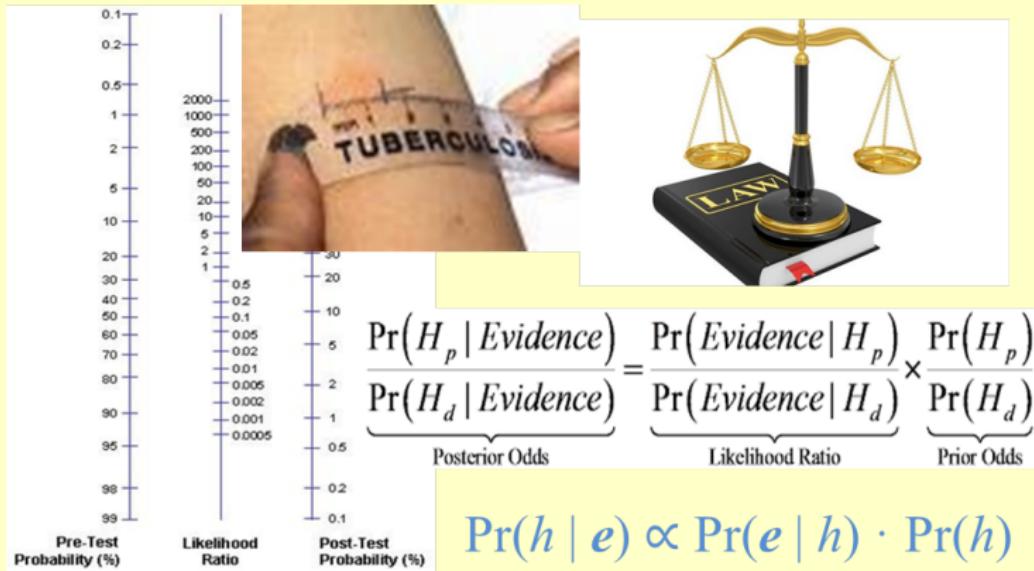
Department of Information and Computing Sciences
Utrecht University, The Netherlands



Universiteit Utrecht

PGM 2016, Lugano

Motivation



- Who can and may judge what?
- How are multiple pieces of evidence combined?

Multiple pieces of evidence

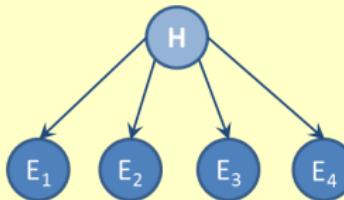
$$\Pr(e_1 \dots e_n \mid h) = \\ \Pr(e_n \mid e_1 \dots e_{n-1} h) \cdot \dots \cdot \Pr(e_2 \mid e_1 h) \cdot \Pr(e_1 \mid h)$$

Multiple pieces of evidence

$$\Pr(e_1 \dots e_n \mid h) =$$

correct: $\Pr(e_n \mid e_1 \dots e_{n-1} h) \cdot \dots \cdot \Pr(e_2 \mid e_1 h) \cdot \Pr(e_1 \mid h)$

naive: $\Pr(e_n \mid h) \cdot \dots \cdot \Pr(e_2 \mid h) \cdot \Pr(e_1 \mid h)$

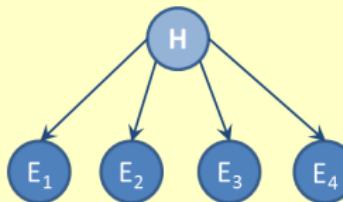


Multiple pieces of evidence

$$\Pr(e_1 \dots e_n \mid h) =$$

correct: $\Pr(e_n \mid e_1 \dots e_{n-1}, h) \cdot \dots \cdot \Pr(e_2 \mid e_1, h) \cdot \Pr(e_1 \mid h)$

naive: $\Pr(e_n \mid h) \cdot \dots \cdot \Pr(e_2 \mid h) \cdot \Pr(e_1 \mid h)$



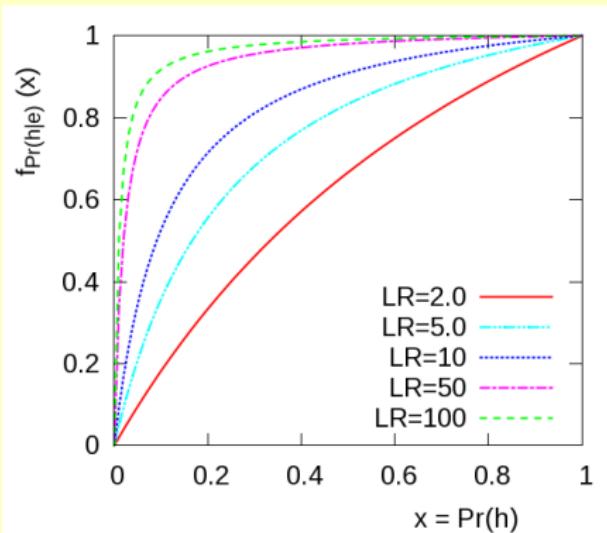
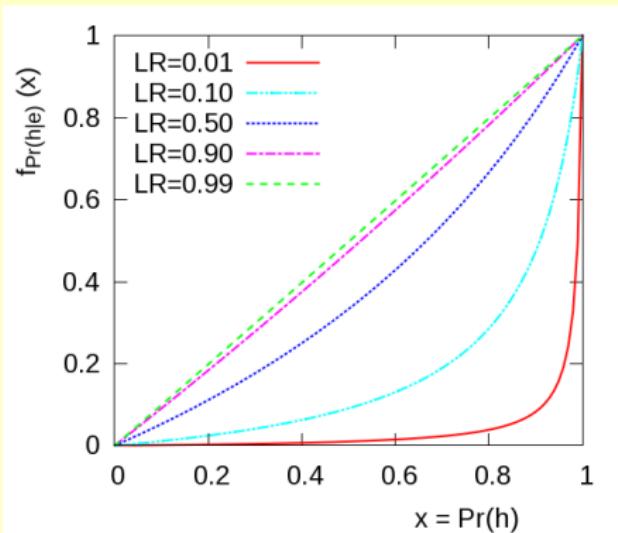
What is the effect of incorrectly assuming independence?

- Previous work¹: effect on Naive Bayes classification
- This paper: effect on likelihood(ratio)

¹a.o. Domingos & Pazzani (1997); Hand & Yu (2001); Rish, Hellerstein & Thathachar (2001); Zhang (2004); Kuncheva & Hoare (2008)

Importance of likelihood-ratio LR

Sensitivity functions as alternative nomograms:



$$\Pr(h \mid e) = \frac{LR \cdot \Pr(h)}{(LR - 1) \cdot \Pr(h) + 1}$$

$$\left(LR = \Pr(e \mid h) / \Pr(e \mid \bar{h}) \right)$$

Dependencies through parameter changes

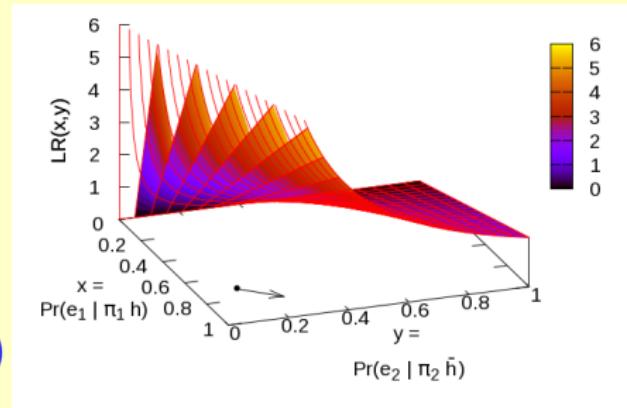
Recall: likelihood-ratio $LR = \Pr(e \mid h) / \Pr(e \mid \bar{h})$

Let $x = \Pr(e_i \mid \pi_i h)$ and
 $y = \Pr(e_j \mid \pi_j \bar{h})$

Then

$$LR(x, y) = \frac{c_1 \cdot x}{c_2 \cdot y}, \quad c_i > 0$$

$$\frac{\partial}{\partial y} LR(x, y) = -\frac{x}{y} \cdot \frac{\partial}{\partial x} LR(x, y)$$



E.g. derivative in $(0.70, 0.13)$ in direction $(1, 1)$: -36.16 .

Dependencies through parameter changes

Recall: likelihood-ratio $LR = \Pr(e \mid h) / \Pr(e \mid \bar{h})$

Let $x = \Pr(e_i \mid \pi_i h)$ and
 $y = \Pr(e_j \mid \pi_j \bar{h})$

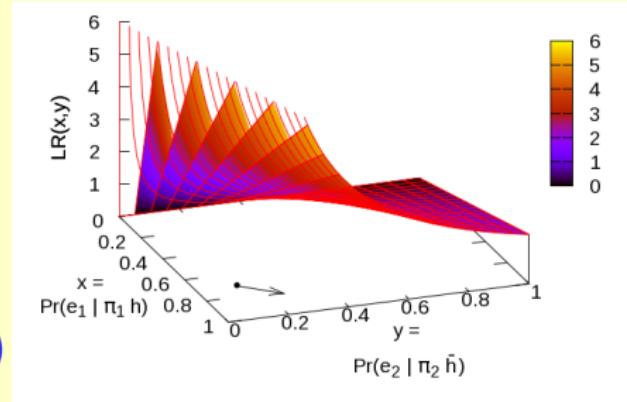
Then

$$LR(x, y) = \frac{c_1 \cdot x}{c_2 \cdot y}, c_i > 0$$

$$\frac{\partial}{\partial y} LR(x, y) = -\frac{x}{y} \cdot \frac{\partial}{\partial x} LR(x, y)$$

E.g. derivative in $(0.70, 0.13)$ in direction $(1, 1)$: -36.16 .

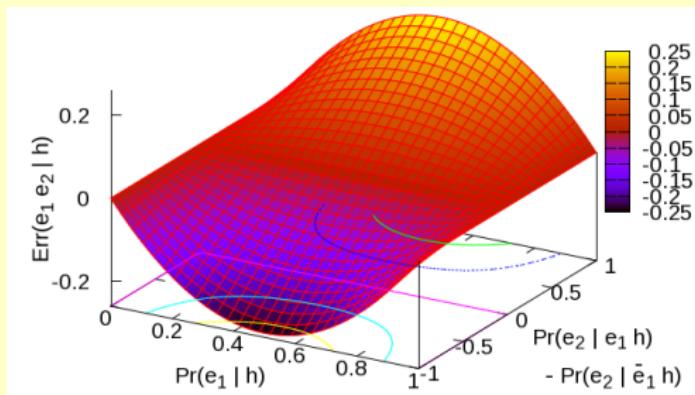
Neglecting one dependency can be compensated by neglecting another.



Error in likelihood: theoretical bound

$$\text{Err}(\mathbf{e}|h) = \Pr(\mathbf{e} | h) - \prod_{i=1}^n \Pr(e_i | h)$$

Error due to neglecting a single dependency between (binary) e_1 and e_2 , when $n \geq 2$ pieces of evidence:

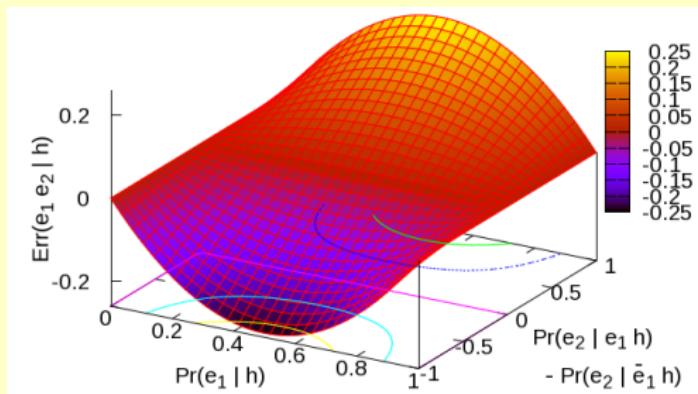


$$n = 2 \rightarrow |\text{Err}(\mathbf{e}|h)| \leq \frac{1}{4}$$
$$n > 2 \rightarrow |\text{Err}(\mathbf{e}|h)| \downarrow$$

Error in likelihood: theoretical bound

$$\text{Err}(\mathbf{e}|h) = \Pr(\mathbf{e} | h) - \prod_{i=1}^n \Pr(e_i | h)$$

Error due to neglecting a single dependency between (binary) e_1 and e_2 , when $n \geq 2$ pieces of evidence:

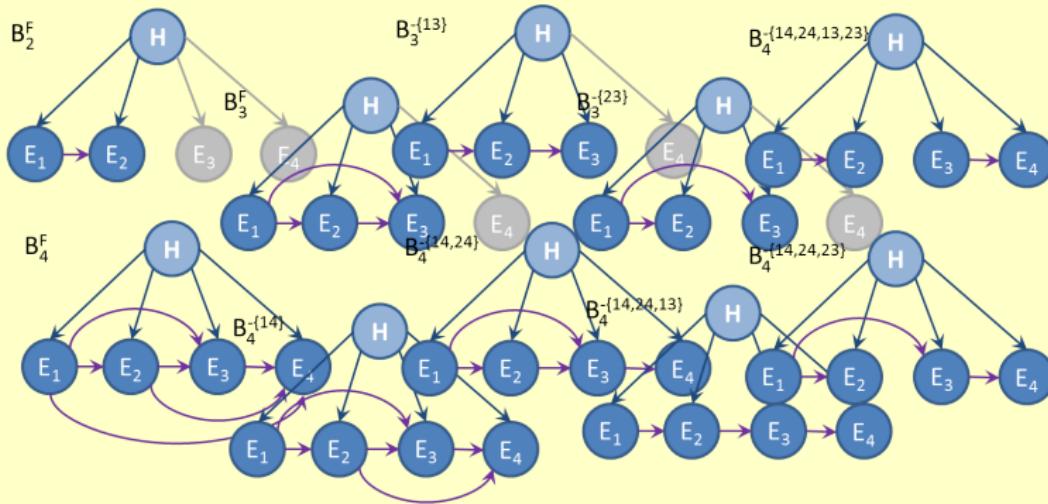


$$n = 2 \rightarrow |\text{Err}(\mathbf{e}|h)| \leq \frac{1}{4}$$

$$n > 2 \rightarrow |\text{Err}(\mathbf{e}|h)| \downarrow$$

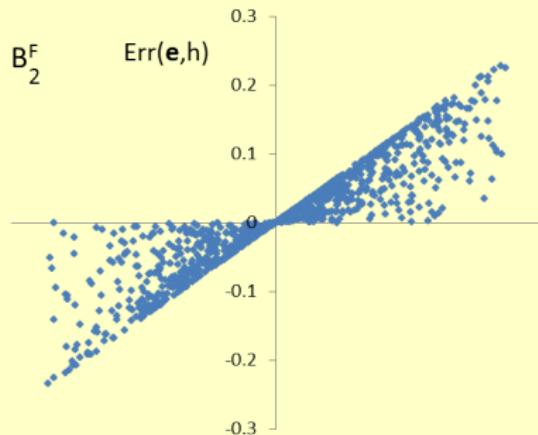
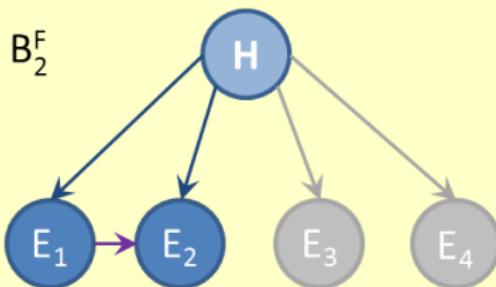
← dependency e_1 and e_2

Error in likelihood: experimental study



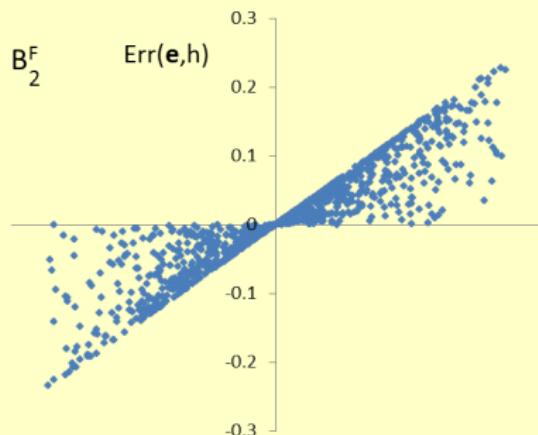
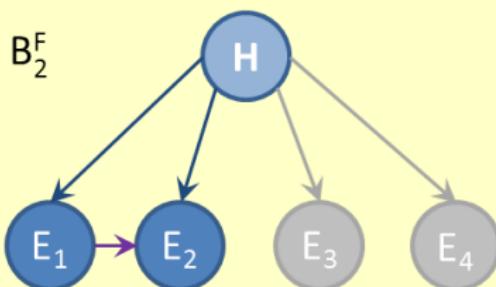
- 10 structures with up to 4 pieces of evidence
- 1000 distributions $\Pr(E_1 \dots E_4 | h)$ per structure
- arc-removal: $\Pr(e_j | e_i \mathbf{z}) = \Pr(e_j | \overline{e_i} \mathbf{z})$

Results: verify the theoretical bound



- $\max_{i=1}^{1000} |\text{Err}(e_1 e_2 \mid h)| = 0.234 < 0.25$
- $\text{avg}_{i=1}^{1000} |\text{Err}(e_1 e_2 \mid h)| = 0.057$ = less than 6%

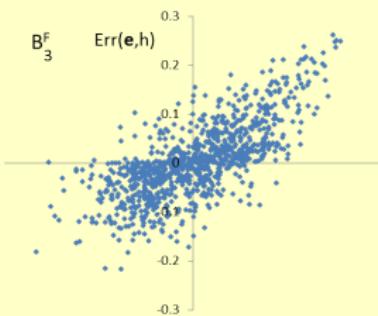
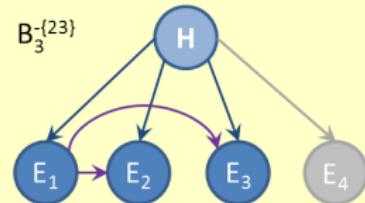
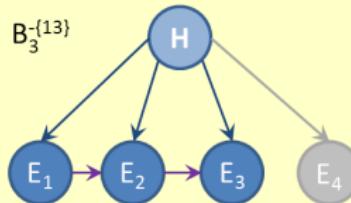
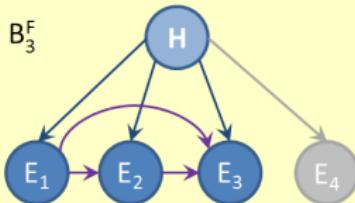
Results: verify the theoretical bound



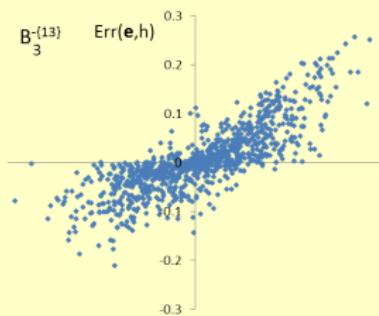
- $\max_{i=1}^{1000} |\text{Err}(e_1 e_2 | h)| = 0.234 < 0.25$
- $\text{avg}_{i=1}^{1000} |\text{Err}(e_1 e_2 | h)| = 0.057$ = less than 6%

x-axis? We'll get to that later...

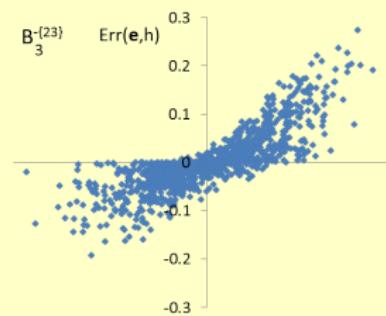
Results (max, avg) for $|\text{Err}(e_1 e_2 e_3 \mid h)|$



(0.261, 0.056)

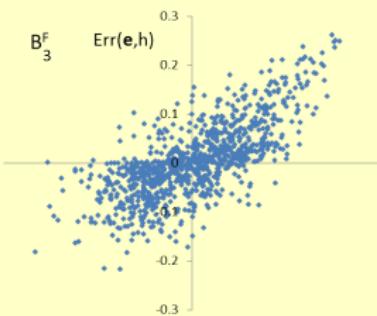
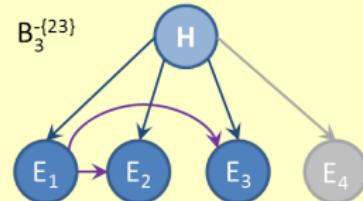
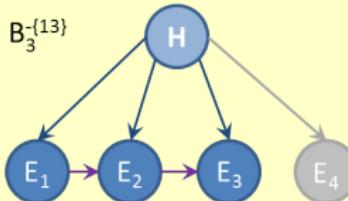
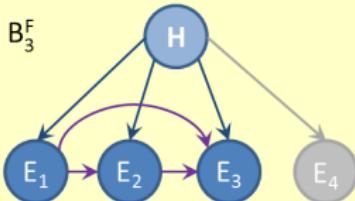


(0.256, 0.048)

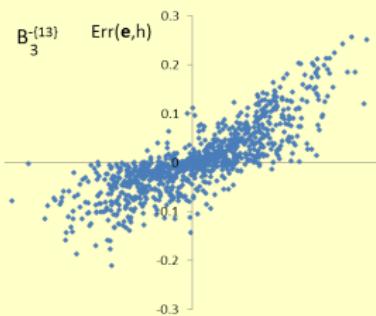


(0.273, 0.043)

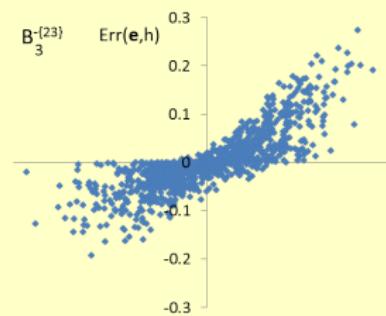
Results (max, avg) for $|\text{Err}(e_1 e_2 e_3 \mid h)|$



(0.261, 0.056)



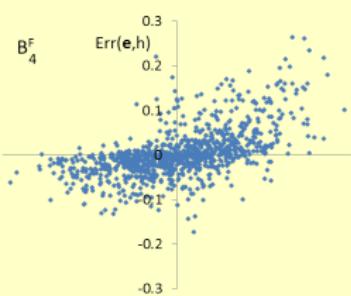
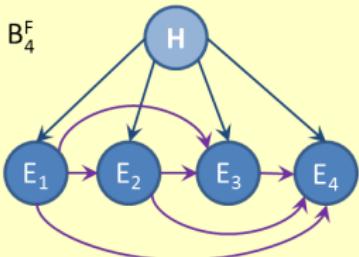
(0.256, 0.048)



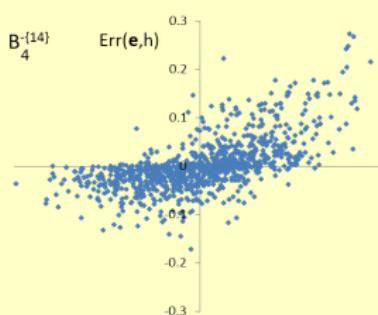
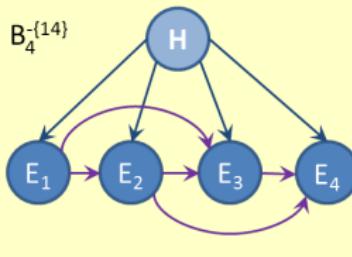
(0.273, 0.043)

Average decreases with fewer independence violations
(though max doesn't).

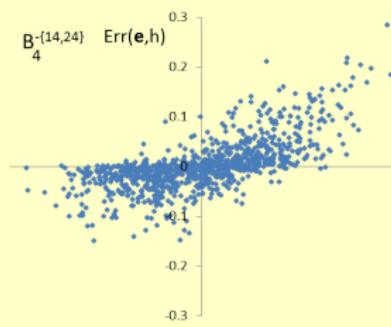
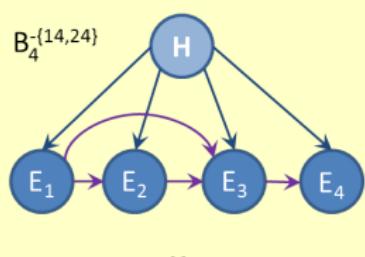
Results (max, avg) for $|\text{Err}(e_1 e_2 e_3 e_4 \mid h)|$



(0.263, 0.037)



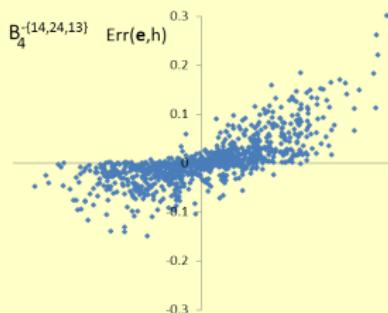
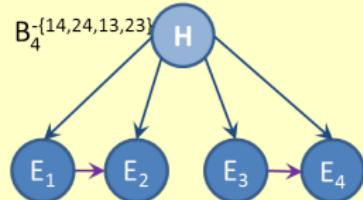
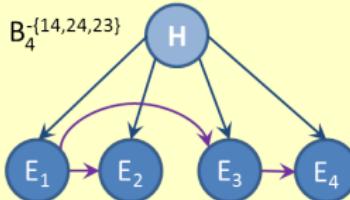
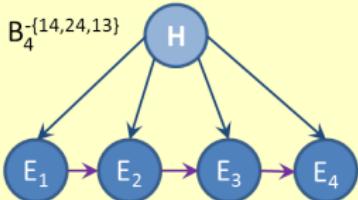
(0.273, 0.036)



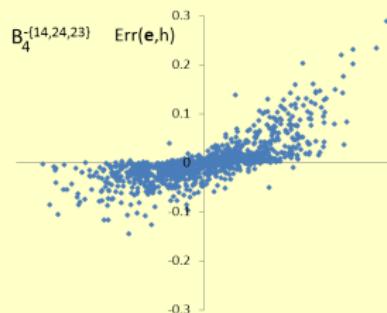
(0.282, 0.033)

Average decreases with fewer independence violations (though max increases).

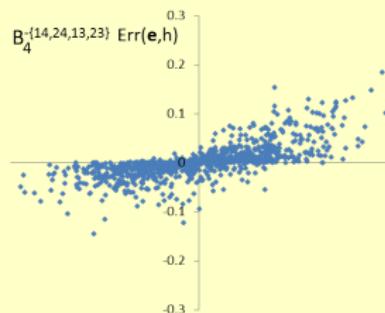
Results (max, avg) for $|\text{Err}(e_1 e_2 e_3 e_4 \mid h)| \parallel$



(0.300, 0.030)



(0.288, 0.028)



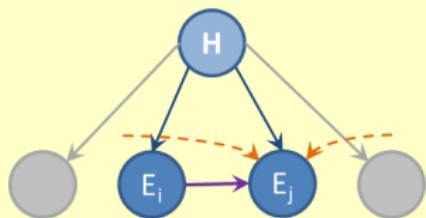
(0.185, 0.022)

Average decreases with fewer independence violations
(and so does max).

Organising the data points

x -axis represents tailored dependency measure:

$$R_{\text{tot}}^{\text{avg}}(\mathbf{e} \mid h) = \sum_{ij} \text{avg}_k \left(\Pr(e_j \mid e_i \mathbf{z}_k h) - \Pr(e_j \mid \bar{e}_i \mathbf{z}_k h) \right)$$



R correlates better with $\text{Err}(\mathbf{e}|h)$ than

- mutual information $I(E_i, E_j \mid h)$
- Yule's Q statistic

Useful measure for establishing a theoretical bound?

Conclusions

- Sensitivity functions give insight in relation between posterior, LR and prior
- Effects of dependencies can be simulated with parameter changes
- Theoretical result: $\text{Err}(e_1 e_2 | h)$ is at most 1/4
- Empirical results (preliminary!):
 - $\text{Err}(e|h) \leq 0.3$
 - avg $|\text{Err}(e|h)|$ decreases with more evidence and fewer independence violations
 - $R_{\text{tot}}^{\text{avg}}$ correlates (somewhat) with $\text{Err}(e|h)$
- Results are relevant for Naive Bayes as well

Future research

- Does upper bound on error hold for larger networks?
- What happens to upper bound under extreme distributions?
- Do monotonicity properties affect upper bound?
- Can we formulate a theoretical bound for more than 2 pieces of evidence, perhaps using the R measure?
- (How does $\text{Err}(e|h)$ affect posterior and/or classification?)

ADDITIONAL SLIDES

Correlations

B_n^c	correlation with Err (for R , Q) or $ Err $ (for I)			
	$R_{\text{tot}}^{\text{avg}}$	$R_{\text{tot}}^{\text{sum}}$	Q_{tot}	I_{tot}
B_2^F	0.917	0.917	0.878	0.932
B_3^F	0.726	0.712	0.759	0.471
B_4^F	0.580	0.510	0.400	0.180
$B_3^{-\{13\}}$	0.816	0.796	0.813	0.554
$B_3^{-\{23\}}$	0.826	0.789	0.816	0.643
$B_4^{-\{14\}}$	0.625	0.557	0.429	0.232
$B_4^{-\{14,24\}}$	0.675	0.614	0.151	0.275
$B_4^{-\{14,24,13\}}$	0.726	0.659	0.517	0.320
$B_4^{-\{14,24,23\}}$	0.738	0.658	0.575	0.454
$B_4^{-\{14,24,13,23\}}$	0.709	0.609	0.678	0.469

Dependency measures

- $I_{ij} = I(E_i, E_j | h) = \sum_{e_i^* \in \text{va}(E_i)} \sum_{e_j^* \in \text{va}(E_j)} \Pr(e_i^* e_j^* | h) \cdot \log \frac{\Pr(e_i^* e_j^* | h)}{\Pr(e_i^* | h) \cdot \Pr(e_j^* | h)}$
- $Q_{ij} = \frac{\Pr(e_i e_j | h) \cdot \Pr(\overline{e_i} \overline{e_j} | h) - \Pr(\overline{e_i} e_j | h) \cdot \Pr(e_i \overline{e_j} | h)}{\Pr(e_i e_j | h) \cdot \Pr(\overline{e_i} \overline{e_j} | h) + \Pr(\overline{e_i} e_j | h) \cdot \Pr(e_i \overline{e_j} | h)}$

Both are summed over all arcs.