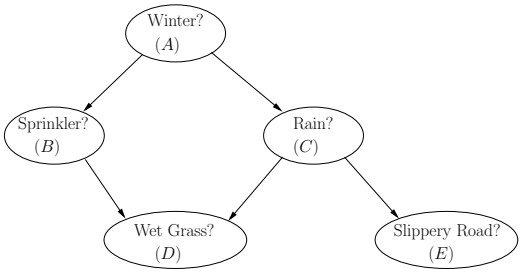
Probabilistic Models: Spring 2014
Scoring Functions Example

We are given the following Bayesian network N.



We are also given the following dataset D.

A	B	C	D	E	Count
T	F	T	T	T	20
T	F	F	F	F	15
F	T	F	T	T	10
F	F	T	T	T	15
F	F	F	F	F	5
T	T	F	T	F	2

For all of the calculations, we need the counts, n_{ijk}. They are as follows.

A	PA_A	i	j	k	n_{ijk}
T	$\emptyset$	1	1	1	37
F	$\emptyset$	1	1	2	30

B	PA_B (A)	i	j	k	n_{ijk}
T	T	2	1	1	2
F	T	2	1	2	35
T	F	2	2	1	10
F	F	2	2	2	20

C	PA_C (A)	i	j	k	n_{ijk}
T	T	3	1	1	20
F	T	3	1	2	17
T	F	3	2	1	15
F	F	3	2	2	15

E	PA_E (C)	i	j	k	n_{ijk}
T	T	5	1	1	35
F	T	5	1	2	0
T	F	5	2	1	10
F	F	5	2	2	22

D	PA_D (BC)	i	j	k	n_{ijk}
T	TT	4	1	1	0
F	TT	4	1	2	0
T	TF	4	2	1	12
F	TF	4	2	2	0
T	FT	4	3	1	35
F	FT	4	3	2	0
T	FF	4	4	1	0
F	FF	4	4	2	20

1. Calculate the MDL score for the network
2. Calculate the BDeu score for the network with ESS=0.1
3. Calculate the BDeu for the network with ESS=100

Useful Equations

Minimum description length

$$MDL(N:D) = -\sum_{i}^{n}\left\{\sum_{j}^{q_i}\sum_{k}^{r_i} N_{ijk}\log\frac{N_{ijk}}{N_{ij}}\right\} + \frac{\log_2 N}{2}\cdot(r_i-1)\cdot q_i$$

$$MDL(N:D) = -\sum_{i}^{n}\ell(X_i|PA_i) + \frac{\log_2 N}{2}\cdot(r_i-1)\cdot q_i$$

Bayesian Dirichlet with likelihood equivalence and uninformative priors

$$\begin{aligned}
P(D,N) &= P(N)P(D|\ N) && \text{Rewrite using chain rule}\\
&= P(N)\prod_{i}^{n}\prod_{j}^{q_i}\frac{\Gamma(\alpha_{ij})}{\Gamma(\alpha_{ij}+n_{ij})}\prod_{k}^{r_i}\frac{\Gamma(\alpha_{ijk}+n_{ijk})}{\Gamma(\alpha_{ijk})} && \text{Substitute probability of data}\\
&\propto \prod_{i}^{n}\prod_{j}^{q_i}\frac{\Gamma(\alpha_{ij})}{\Gamma(\alpha_{ij}+n_{ij})}\prod_{k}^{r_i}\frac{\Gamma(\alpha_{ijk}+n_{ijk})}{\Gamma(\alpha_{ijk})} && \text{Assume a uniform structure prior}\\
&\propto \prod_{i}^{n}\prod_{j}^{q_i}\frac{\Gamma(\frac{\alpha}{q_i})}{\Gamma(\frac{\alpha}{q_i}+n_{ij})}\prod_{k}^{r_i}\frac{\Gamma(\frac{\alpha}{r_i\cdot q_i}+n_{ijk})}{\Gamma(\frac{\alpha}{r_i\cdot q_i})} && \text{Replace the } \alpha\text{s}\\
BDeu(N:D,\alpha) &= \sum_{i}^{n}\sum_{j}^{q_i}\log\frac{\Gamma(\frac{\alpha}{q_i})}{\Gamma(\frac{\alpha}{q_i}+n_{ij})} + \sum_{k}^{r_i}\log\frac{\Gamma(\frac{\alpha}{r_i\cdot q_i}+n_{ijk})}{\Gamma(\frac{\alpha}{r_i\cdot q_i})} && \text{Work in log-space}\\
BDeu(N:D,\alpha) &= \sum_{i}^{n}\sum_{j}^{q_i}\log\Gamma(\frac{\alpha}{q_i}) - \log\Gamma(\frac{\alpha}{q_i}+n_{ij}) + && \text{Remove divisions}\\
&\qquad \sum_{k}^{r_i}\log\Gamma(\frac{\alpha}{r_i\cdot q_i}+n_{ijk}) - \log\Gamma(\frac{\alpha}{r_i\cdot q_i})
\end{aligned}$$