Exercise 1: Conditional independencies in Bayes nets

In each case, state the conditions (what sets of nodes must and/or must not be known) under which the specified node sets will be conditionally independent from one another. If the node sets are always independent or can never be independent, say so.

Example:

- \{W\} and \{A\} are conditionally independent if and only if \(D\) is unknown.
- \{W\} and \{D\} are never conditionally independent.

Examples to solve:

1. A variant of the disfluency model we saw earlier:

   \[ M \quad \text{intended meaning to be conveyed} \]
   \[ W \quad \text{is the word intended to be spoken a hard word?} \]
   \[ A \quad \text{was the speaker’s attention distracted?} \]
   \[ D \quad \text{was a disfluency uttered?} \]
2. The relationship between a child’s linguistic environment, his/her true linguistic abilities/proficiency, and measures of his/her proficiency in separate spoken and written tests

- $E$: a child’s linguistic environment
- $P$: the child’s linguistic proficiency (number of words known, etc.)
- $S$: the child’s performance on a spoken language proficiency test
- $W$: the child’s performance on a written language proficiency test

(a) $\{W\}$ and $\{A\}$
(b) $\{M\}$ and $\{D\}$
(c) $\{M\}$ and $\{A\}$
(d) $\{D\}$ and $\{A\}$

(a) $\{S\}$ and $\{W\}$
(b) $\{E\}$ and $\{P\}$
(c) $\{E\}$ and $\{S\}$
(d) $\{E,P\}$ and $\{S\}$
(e) $\{E,P\}$ and $\{S,W\}$
3. Speakers’ familiarities (quantified, say, on a scale of 1 to 10) with different words

- $S_i$ the $i$-th speaker’s general vocabulary size
- $W_j$ the $j$-th word’s general difficulty/rarity
- $\Sigma_S$ the variability in vocabulary sizes across speakers
- $\Sigma_W$ the variability in difficulties/rarities across words
- $Y_{ij}$ the $i$-th speaker’s familiarity with the $j$-th word

Exercise 2: Binomial and beta-binomial predictive distributions

Three native English speakers start studying a new language together. This language has flexible word order, so that sometimes the subject of the sentence can precede the verb (SV), and sometimes it can follow the verb (VS). Of the first three utterances of the new language they are taught, one is VS and the other two are SV.

Speaker A abandons her English-language preconceptions and uses the method of maximum likelihood to estimate the probability that an utterance will be SV. Speakers B and
C carry over some preconceptions from English; they draw inferences regarding the SV/VS word order frequency in the language according to a beta-distributed prior, with $\alpha_1 = 8$ and $\alpha_2 = 1$ (here, SV word order counts as a “success”), which is then combined with the three utterances they’ve been exposed to thus far. Speaker B uses maximum a-posterior (MAP) probability to estimate the probability that an utterance will be SV. Speaker C is fully Bayesian and retains a full posterior distribution on the probability that an utterance will be SV.

It turns out that the first three utterances of the new language were uncharacteristic; of the next twenty-four utterances our speakers hear, sixteen of them are VS. Which of our three speakers was best prepared for this eventuality, as judged by the predictive distribution placed by the speaker on the word order outcomes of these twenty-four utterances? Which of our speakers was worst prepared? Why?