Formula Sheet

1. Probability

- Conditional Probability

$$P(A|B) = \frac{P(A,B)}{P(B)}$$
$$P(A,B) = P(A|B)P(B)$$

- Law of total probability

$$P(A) = \sum_{B \in \mathcal{B}} (P(A,B))$$
$$P(A) = \sum_{B \in \mathcal{B}} (P(A|B)P(B))$$
$$P(A|C) = \sum_{B \in \mathcal{B}} (P(A|B,C)P(B|C))$$

- Bayes' Rule

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

2. Probabilistic Representation

- Chain Rule for Bayesian Networks

$$P(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} (P(x_i| \text{ Parents of } x_i))$$

- Independence and Conditional Independence

$$A \perp B \Leftrightarrow P(A,B) = P(A)P(B) \text{ or } P(A) = P(A|B)$$
$$(A \perp B)|C \Leftrightarrow P(A,B|C) = P(A|C)P(B|C) \text{ or } P(A|C) = P(A|B,C)$$

- Conditional Independence in Bayesian Networks
  $(A \perp B|\mathcal{C})$ or $\mathcal{C}$ d-separates A and B if for all paths from A to B, one of the following is true
  (a) The path contains a chain $X \to Y \to Z$ for $Y \in \mathcal{C}$.
  (b) The path contains a fork $X \leftarrow Y \to Z$, for $Y \in \mathcal{C}$.
  (c) The path contains an inverted fork or v-structure $X \to Y \leftarrow Z$, for $Y \notin \mathcal{C}$ and the children/descendents of Y not in $\mathcal{C}$.

3. Inference

- Exact Inference
  Use Law of total probability and sum out hidden variables. Also need to use chain rule. Can eliminate one variable at a time using tables.
- Approximate Inference
  - Topological Sort - Ordered list of nodes such that parents come before children.
  - Direct Sampling
    Sample in order of topological sort, and estimate probability.

$$P(c|o_{1:n}) = \frac{\text{Number of times } c \text{ and } o_{1:n} \text{ is observed in sample}}{\text{Number of times } o_{1:n} \text{ is observed}}$$

  - Likelihood Weighted Sampling
    Sample in order of topological sort, enforce given conditions but assign weight from known conditional distribution. For x_i given/observed

$$w = w \times P(x_i|\text{parents of } x_i)$$
$$P(c|o_{1:n} = \frac{\text{Sum of weights were } c \text{ observed}}{\text{Sum of weights}}$$

  - Gibbs Sampling
    Take random initial sample, with given/observed variables. Iteratively update variables in order using probability given all other current sample. Update all unknown variables with probability $P(X_i|x'_{1:n/i})$.

4. Parameter Learning

- Maximum Likelihood
  - Binary Discrete Variable - C

$$P(C=1) \approx \hat{\theta} = \frac{m}{n} = \frac{\text{number of observations of 1}}{\text{total observations}}$$

  - Discrete variable with k options

$$P(C=i) \approx \hat{\theta}_i = \frac{m_i}{\sum_{j=1}^{k}(m_j)} = \frac{\text{number of observations of i}}{\text{total observations}}$$

  - Continuous variable - Normal Distribution

$$\hat{\mu} = \frac{\sum_i (v_i)}{n}$$
$$\hat{\sigma}^2 = \frac{\sum_i \left((v_i - \hat{mu})^2\right)}{n}$$

- Bayesian Parameter Learning
  - Binary Discrete Variable, uniform prior $\alpha = \beta = 1$

$$p(\theta|o_i) = \text{Beta}(\alpha + m, \beta + n - m)$$

  - Discrete Variable with k options

$$p(\theta_{1:n}|\alpha_{1:n}, m_{1:n}) = \text{Dir}(\theta_{1:n}|\alpha_1 m_1, \ldots, \alpha_n + m_n)$$

5. Structure Learning

- Bayesian Score

$$m_{ijk} = (\#X_i = k|\pi_{ij}) \qquad \alpha_{ij0} = \sum_{k=1}^{r_i}(\alpha_{ijk}) \qquad m_{ij0} = \sum_{k=1}^{r_i}(m_{ijk})$$

$$\ln(P(G|D)) = \ln(P(G)) + \sum_{i=1}^{n}\sum_{j=1}^{q_i}\left(\ln\left(\frac{\Gamma(\alpha_{ij0})}{\Gamma(\alpha_{ij0} + m_{ij0}}\right) + \sum_{k=1}^{r_i}\ln\left(\frac{\Gamma(\alpha_{ijk} + m_{ijk})}{\Gamma(\alpha_{ijk}}\right)\right)$$

- Graph Searches
  - K2 search - Greedily add parents to nodes to maximize score
  - Local Search - Search through local neighbors for max score
  - Genetic Algorithm
  - Memetic Algorithm - Genetic Algorithm with local search
- Markov equivalence Classes
  Two graphs are Markov equivalent if they have the same edges without regard to direction and the same v-structures. A Bayesian score is Markov equivalent is one were $\sum_j\left(\sum_k(\alpha_{ijk})\right)$ is constant.
- Partially directed graph search
  A partially directed graph can encode a Markov Equivalence Class. The local operations on a partially directed graph are
  - If an edge between A and B doesn't exist, add either $A - B$ or $A \to B$.
  - If $A - B$ $A \to B$, then remove the edge.
  - If $A \to B$, then reverse direction $A \leftarrow B$.
  - If $A - B - C$, then add $A \to B \leftarrow C$.