CSE 291-D: Homework 1

Hao-en Sung [A53204772] (wrangle1005@gmail.com)
Department of Computer Science
University of California, San Diego
San Diego, CA 92092

Problem 1

(a)

- A **is not** independent of A given J.
  One node cannot be independent with itself.
- B **is not** independent of A given J.
  For path A-G-D-B, all three rules cannot be applied on any node.
- C **is** independent of A given J.
  All path between A and C must pass node I, which fits the third rule.
- D **is not** independent of A given J.
  For route A-G-D, all three rules cannot be applied on any node.
- E **is not** independent of A given J.
  For route A-G-D-B-E, all three rules cannot be applied on any node.
- F **is** independent of A given J.
  All path between A and F must pass node I, which fits the third rule.
- G **is not** independent of A given J.
  A and G are connected directly.
- H **is not** independent of A given J.
  For route A-G-D-H, all three rules cannot be applied on any node.
- I **is not** independent of A given J.
  For route A-G-D-B-E-I, all three rules cannot be applied on any node.

(b)

To prove the formula, I first plan to remove most of the conditional independent variables in the form $P(X_i|X_{\neg i})$, which can results in only $P(X_i|Pa(X_i), Ch(X_i), CP(X_i))$. From Fig. 1, there are five potential cases that I need to take care of, where case-by-case proofs are listed as follows.

Case 1. *Via a parent of a parent of X, i.e. node 1 in the hint figure.*

Sol. Since the parent of X is given, it applies to type-1 d-separate. □

Case 2. *Via a parent of a parent of a child of X, i.e. node 2 in the hint figure.*

Sol. Denote node 2 as N_2, the child of N_2 as A and the co-child of A and X as B. We would like to prove $P(X, N_2|A, B) = P(X|A, B) \cdot P(N_2|A, B)$.

30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain.

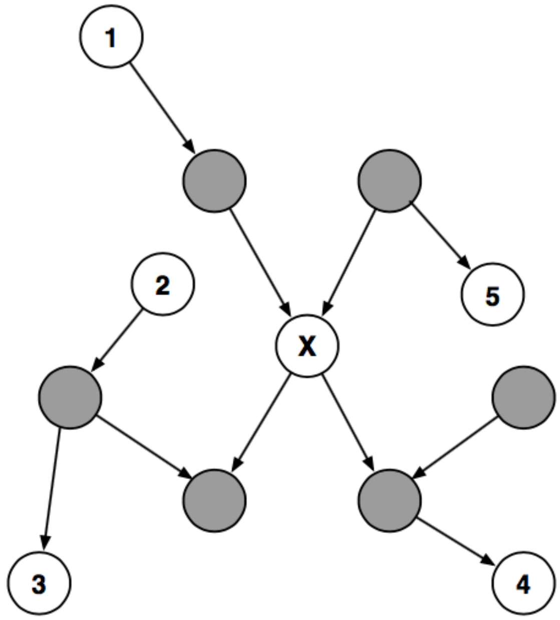

Figure 1: Markov Blanket

$$
\begin{aligned}
P(X, N_2|A, B) &= \frac{P(X, 3, A, B)}{P(A, B)} = \frac{P(X) \cdot P(N_2) \cdot P(A|N_2) \cdot P(B|A, X)}{P(A, B)} \\
&= \frac{P(X) \cdot P(N_2) \cdot P(A|N_2) \cdot P(B|A, X) \cdot \frac{P(A,X)}{P(A,X)} \cdot}{P(A, B)} \\
&= P(X|A, B) \cdot \frac{P(X) \cdot P(N_2) \cdot P(A|N_2)}{P(A) \cdot P(X)} \\
&= P(X|A, B) \cdot P(N_2|A) \\
&= P(X|A, B) \cdot P(N_2|A, B)
\end{aligned}
$$

□

Case 3. *Via a child of a parent of a child of X, i.e. node 3 in the hint figure.*

Sol. Denote node 3 as N_3, the parent of N_3 as A and the co-child of A and X as B. We would like to prove $P(X, N_3|A, B) = P(X|A, B) \cdot P(N_3|A, B)$.

$$
\begin{aligned}
P(X, N_3|A, B) &= \frac{P(X, 3, A, B)}{P(A, B)} = \frac{P(X) \cdot P(A) \cdot P(B|A, X) \cdot P(N_3|A)}{P(A, B)} \\
&= \frac{\frac{P(X) \cdot P(A)}{P(X,A)} \cdot P(X, A) \cdot P(B|A, X) \cdot P(N_3|A)}{P(A, B)} \\
&= 1 \cdot P(X|A, B) \cdot P(N_3|A) \\
&= P(X|A, B) \cdot P(N_3|A, B)
\end{aligned}
$$

□

Case 4. *Via a child of a child of X, i.e. node 4 in the hint figure.*

Sol. Since the child of X is given, it applies to type-1 d-separate. $\square$

Case 5. *Via a child of a parent of X, i.e. node 5 in the hint figure.*

Sol. Since the parent of X is given, it applies to type-2 d-separate. $\square$

After simplifying the original formula to $P(X_i|Pa(X_i), Ch(X_i), CP(X_i))$, it is then easy to derive the rest of proofs.

$$\begin{aligned}
P(X_i|Pa(X_i), Ch(X_i), CP(X_i)) &\propto P(X_i, Pa(X_i), Ch(X_i), CP(X_i)) \\
&= P(Pa(X_i)) \cdot P(X_i|Pa(X_i)) \\
&\quad \cdot P(CP(X_i)|Pa(X_i)) \cdot P(Ch(X_i)|X_i, CP(X_i)) \\
&\propto P(X_i|Pa(X_i)) \cdot P(Ch(X_i)|X_i, CP(X_i)),
\end{aligned}$$

where $Pa(X_i)$, $Ch(X_i)$, $CP(X_i)$ means the set of parent nodes, child nodes, and co-parent, respectively.

Problem 2

(a)

The derivation can be written as follows.

$$\begin{aligned}
\text{posterior} &\propto \text{likelihood} \times \text{prior} \\
&= \binom{n}{k} p^k (1-p)^{n-k} \cdot \frac{1}{B(\alpha, \beta)} p^{\alpha-1} (1-p)^{\beta-1} \\
&= \frac{\binom{n}{k}}{B(\alpha, \beta)} p^{k+\alpha-1} (1-p)^{n-k+\beta-1} \\
&\propto \binom{n+\alpha+\beta-2}{k+\alpha-1} p^{k+\alpha-1} (1-p)^{n-k+\beta-1}
\end{aligned}$$

One call tell that posterior is the form of binomial, which meets the definition of conjugate prior.

(b)

It is known that $P(\theta = u|x) = \frac{P(x|\theta=u) \cdot P(\theta=u)}{\sum_{u'} P(x|\theta=u') \cdot P(\theta=u')}$. In this special problem, we know that $P(\theta = u) = \frac{1}{5}, \forall u \in \{1, \cdots, 5\}$, which can later help me further simplify the problem.

$$
\begin{aligned}
P(\theta = 1|x) &= \frac{P(x|\theta = 1) \cdot P(\theta = 1)}{\sum_{u'} P(x|\theta = u') \cdot P(\theta = u')} \\
&= \frac{\binom{5}{3} \cdot \left(\frac{1}{5}\right)^3 \left(\frac{4}{5}\right)^2}{\sum_{u'} P(x|\theta = u')} \\
&= \frac{16}{16 + 72 + 108 + 64 + 0} = \frac{16}{260} = \frac{4}{65} \\
P(\theta = 2|x) &= \frac{P(x|\theta = 2) \cdot P(\theta = 2)}{\sum_{u'} P(x|\theta = u') \cdot P(\theta = u')} \\
&= \frac{\binom{5}{3} \cdot \left(\frac{2}{5}\right)^3 \left(\frac{3}{5}\right)^2}{\sum_{u'} P(x|\theta = u')} \\
&= \frac{72}{16 + 72 + 108 + 64 + 0} = \frac{72}{260} = \frac{18}{65} \\
P(\theta = 3|x) &= \frac{P(x|\theta = 3) \cdot P(\theta = 3)}{\sum_{u'} P(x|\theta = u') \cdot P(\theta = u')} \\
&= \frac{\binom{5}{3} \cdot \left(\frac{3}{5}\right)^3 \left(\frac{2}{5}\right)^2}{\sum_{u'} P(x|\theta = u')} \\
&= \frac{108}{16 + 72 + 108 + 64 + 0} = \frac{108}{280} = \frac{27}{65} \\
P(\theta = 4|x) &= \frac{P(x|\theta = 4) \cdot P(\theta = 4)}{\sum_{u'} P(x|\theta = u') \cdot P(\theta = u')} \\
&= \frac{\binom{5}{3} \cdot \left(\frac{4}{5}\right)^3 \left(\frac{1}{5}\right)^2}{\sum_{u'} P(x|\theta = u')} \\
&= \frac{64}{16 + 72 + 108 + 64 + 0} = \frac{64}{280} = \frac{16}{65} \\
P(\theta = 5|x) &= \frac{P(x|\theta = 5) \cdot P(\theta = 5)}{\sum_{u'} P(x|\theta = u') \cdot P(\theta = u')} \\
&= \frac{\binom{5}{3} \cdot \left(\frac{5}{5}\right)^3 \left(\frac{0}{5}\right)^2}{\sum_{u'} P(x|\theta = u')} \\
&= \frac{0}{16 + 72 + 108 + 64 + 0} = 0
\end{aligned}
$$

(c)

$$
\begin{aligned}
&P(\text{next ball is black}|\text{3 black balls out of 5 draws}) \\
&= \sum_u P(\text{next ball is black} \cap \theta = u|x) \\
&= \sum_u P(\text{next ball is black}|\theta = u) \cdot P(\theta = u|x) \\
&= \frac{1}{5} \cdot \frac{4}{65} + \frac{2}{5} \cdot \frac{18}{65} + \frac{3}{5} \cdot \frac{27}{65} + \frac{4}{5} \cdot \frac{16}{65} \\
&= \frac{185}{325} = \frac{37}{65}
\end{aligned}
$$

Problem 3

(a)

Define $L = e^{-\lambda} \cdot \frac{\lambda^x}{x!}$, then our goal is to maximize $\log L = -\lambda + x \log \lambda - \log(x!)$ in terms of λ. It can then derive as follows.

$$\frac{\partial \log L}{\partial \lambda} = -1 + \frac{x}{\lambda}$$

Thus, the maximum happens when $\lambda = x$, where $L = e^{-x} \cdot \frac{x^x}{x!}$.

(b)

The derivation can be written as follows.

$$\begin{aligned}
P(\lambda|D) &\propto P(D|\lambda) \times P(\lambda) \\
&= e^{-\lambda} \frac{\lambda^x}{x!} \cdot \frac{b^a}{\Gamma(a)} \lambda^{a-1} e^{-\lambda b} \\
&= \frac{b^a}{x! \cdot \Gamma(a)} \lambda^{x+a-1} e^{-\lambda(b+1)} \\
&\propto \frac{(b+1)^(x+a)}{\Gamma(x+a)} \lambda^{x+a-1} e^{-\lambda(b+1)}
\end{aligned}$$

The last formula is in fact Gamma$(\lambda|x+a, b+1)$, which meets our expectation.

(c)

From the conclusion of (b), we know that the mean of posterior can be represented as $\frac{x+a}{b+1}$. Thus, when $a \to 0, b \to 0$, mean of posterior $\to x$.

Problem 4

(a)

Since there maybe many machines, if assumed machines follows iid with each other, the likelihood can be expressed as

$$\begin{aligned}
L = P(X|\theta) &= \prod_x P(x|\theta) = \prod_x \theta e^{-\theta x} \\
\log L &= \sum_x (\log x - \theta x)
\end{aligned}$$

It is known that MLE happens when the derivative of it in terms of θ is 0. Then, I have

$$\begin{aligned}
0 &= \frac{\partial \log L}{\partial \theta} \\
&= \sum_x \left(\frac{1}{\theta} - x \right) \\
&= N\theta - \sum_x x
\end{aligned}$$

Thus, MLE can be achieved when $\theta = \frac{1}{\bar{x}}$, where $\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$.

(b)

The MLE of θ is $\frac{1}{\frac{5+6+4}{3}} = \frac{1}{5}$.

(c)

Since exponential distribution is a special case of gamma distribution where $a = 1, b = \lambda$, to get $E[\theta] = \frac{1}{3}$, I just need to set $\lambda = 3$.

(d)

$$\begin{aligned}
P(\theta|D,\lambda) &= \frac{P(D|\theta,\lambda)\cdot P(\theta|\lambda)}{\sum_\theta P(D|\theta,\lambda)\cdot P(\theta|\lambda)} \\
&= \frac{P(D|\theta)\cdot P(\theta|\lambda)}{\sum_\theta P(D|\theta)\cdot P(\theta|\lambda)} \\
&\propto \prod_x \left(\theta e^{-\theta x}\right)\cdot \lambda e^{-\lambda\theta} \\
&= \theta^N \cdot \lambda \cdot e^{-\theta\cdot(\sum_{i=1}^N x_i+\lambda)} \\
P(\theta|X_1=5, X_2=6, X_3=4, \lambda=3) &\propto \theta^3 e^{-18\theta}
\end{aligned}$$

(e)

From (d), I know that the posterior also follows an exponential distribution. Since prior and posterior are the same distribution, one can tell exponential prior is conjugate to the exponential likelihood.

(f)

From (d), the posterior mean can be calculated as $E[\theta|D,\lambda] = \frac{N+1}{\sum_{i=1}^N x_i+\lambda} = \frac{4}{18} = \frac{2}{9}$.

(g)

It is different from original MLE, since we add some knowledge, or said prior, into our formulation.

Problem 5

To avoid ambiguity, I replace parameter Z for Σ in the original formula.

The derivation of its exponential family form can be written as.

$$\begin{aligned}
\mathcal{N}(\boldsymbol{x}|\mu,\boldsymbol{Z}) &= \frac{1}{(2\pi)^{\frac{D}{2}}|\boldsymbol{Z}|^{\frac{1}{2}}} e^{-\frac{1}{2}(\boldsymbol{x}-\mu)^\intercal \boldsymbol{Z}^{-1}(\boldsymbol{x}-\mu)} \\
&= e^{-\frac{1}{2}\left(\boldsymbol{x}^\intercal\boldsymbol{Z}^{-1}\boldsymbol{x}-\mu^\intercal\boldsymbol{Z}^{-1}\boldsymbol{x}+\mu^\intercal\boldsymbol{Z}^{-1}\mu\right)-\log\left[(2\pi)^{\frac{D}{2}}|\boldsymbol{Z}|^{\frac{1}{2}}\right]} \\
&= e^{-\frac{1}{2}vec(\boldsymbol{Z}^{-1})\cdot vec(\boldsymbol{x}\boldsymbol{x}^\intercal)+\mu^\intercal\boldsymbol{Z}^{-1}\boldsymbol{x}-\frac{1}{2}\mu^\intercal\boldsymbol{Z}^{-1}\mu-\log\left[(2\pi)^{\frac{D}{2}}|\boldsymbol{Z}|^{\frac{1}{2}}\right]}
\end{aligned}$$

where

$$\begin{aligned}
h(\boldsymbol{x}) &= 1 \\
\boldsymbol{\theta} &= [-\frac{1}{2}vec(\boldsymbol{Z}^{-1}), \mu^\intercal\boldsymbol{Z}^{-1}] \\
\boldsymbol{x} &= [vec(\boldsymbol{x}\boldsymbol{x}^\intercal), \boldsymbol{x}] \\
A(\boldsymbol{\theta}) &= \frac{1}{2}\mu^\intercal\boldsymbol{Z}^{-1}\mu + \log\left[(2\pi)^{\frac{D}{2}}|\boldsymbol{Z}|^{\frac{1}{2}}\right]
\end{aligned}$$