

# Problem Set #2: Learning Forests from Data

Joe Suzuki

July 25, 2017

## Problems (Corrected)

1. Compute the correlation coefficient in page five.
2. Generate a forest using  $J_n$  from  
 $x^5 = (1, 0, 1, 0, 1)$ ,  $y^5 = (1, 1, 1, 1, 1)$ ,  $z^5 = (1, 1, 1, 0, 0)$ ,  $w^5 = (0, 1, 0, 1, 0)$ .
3. Generate a Bayesian network from the data  $x^5, y^5, z^5$  ( $p = 3$ ) in the previous problem using  $J_n$ .
4. Install R language and BNSL package to generate a forest for insurance data using  $J_n$  and  $I_n$ .  
`mi_matrix(insurance)` instead of `mi_matrix(asia)`

## Hints

1.  $E[X] = E[Y] = E[XY] = 0$  means

$$\text{cov}(X, Y) = E[(X - E[X])(Y - E[Y])] = E[XY] = 0$$

- 2.

$$Q(x^5) = Q(z^5) = Q(w^5) = \left(\frac{1}{2}\right)^2 \cdot \left(\frac{3}{2}\right)^2 \cdot \frac{5}{2} \cdot \frac{1}{5!} = \frac{3}{2^8}$$

$$Q(y^5) = \frac{1}{2} \cdot \frac{3}{2} \cdot \frac{5}{2} \cdot \frac{7}{2} \cdot \frac{9}{2} \cdot \frac{1}{5!} = \frac{63}{2^8}$$

$$Q(x^5, y^5) = \left(\frac{1}{2}\right)^3 \cdot \left(\frac{3}{2}\right)^2 \cdot \frac{1}{6!} = \frac{9}{2^5 6!}$$

$$Q(x^5, z^5) = Q(z^5, w^5) = \left(\frac{1}{2}\right)^4 \cdot \frac{3}{2} \cdot \frac{1}{6!} = \frac{3}{2^5 6!}$$

$$Q(x^5, w^5) = Q(y^5, z^5) = Q(y^5, w^5) = \left(\frac{1}{2}\right)^2 \cdot \left(\frac{3}{2}\right)^2 \cdot \frac{5}{2} \cdot \frac{1}{6!} = \frac{45}{2^5 6!}$$

$$\frac{Q(x^5, y^5)}{Q(x^5)Q(y^5)} = \frac{9}{2^5 6!} / \left\{ \frac{3}{2^8} \cdot \frac{63}{2^8} \right\} = \frac{1}{21} \cdot \frac{2^6}{6!}$$

$$\frac{Q(x^5, z^5)}{Q(x^5)Q(z^5)} = \frac{3}{2^5 6!} / \left\{ \frac{3}{2^8} \cdot \frac{3}{2^8} \right\} = \frac{1}{3} \cdot \frac{2^6}{6!}$$

$$\frac{Q(x^5, w^5)}{Q(x^5)Q(w^5)} = \frac{45}{2^5 6!} / \left\{ \frac{3}{2^8} \cdot \frac{3}{2^8} \right\} = 5 \cdot \frac{2^6}{6!}$$

$$\frac{Q(y^5, z^5)}{Q(y^5)Q(z^5)} = \frac{45}{2^5 6!} / \left\{ \frac{63}{2^8} \cdot \frac{3}{2^8} \right\} = \frac{5}{21} \cdot \frac{2^6}{6!}$$

$$\frac{Q(y^5, w^5)}{Q(y^5)Q(w^5)} = \frac{45}{2^5 6!} / \left\{ \frac{63}{2^8} \cdot \frac{3}{2^8} \right\} = \frac{5}{21} \cdot \frac{2^6}{6!}$$

$$\frac{Q(z^5, w^5)}{Q(z^5)Q(w^5)} = \frac{3}{2^5 6!} / \left\{ \frac{3}{2^8} \cdot \frac{3}{2^8} \right\} = \frac{1}{3} \cdot \frac{2^6}{6!}$$

If  $J_n(i, j) = J_n(i', j')$ , you may choose any of them. In this case, there are many solutions. Only one edge set is sufficient.

3. Simply copy the following lines after install the CRAN BNSL package.

```
library(BNSL)
mm=mi_matrix(insurance, proc=9) # I_n is used
edge.list=kruskal(mm)
g=graph_from_edgelist(edge.list, directed=FALSE)
plot(g)
mm=mi_matrix(insurance) # J_n is used
edge.list=kruskal(mm)
g=graph_from_edgelist(edge.list, directed=FALSE)
plot(g)
```