Ran Liu

Problem	**Points**
Problem 1	
Problem 2	
Problem 3	
Problem 4	
Problem 5	
Bonus	
Total	

Problem 1.1

1. False. Since the Parent of Chills: Flu is unknown value.
2. True. $X_i \perp NonDesc(X_i) | Pa(X_i)$
3. False. Since the parent of headache: dehydration is unkown
4. True. same as 2.
5. False. N has another parent Z relates to S.
6. True. D and H control the parent of N are known.
7. False. $P(F,D) \neq P(F)P(D)$ when S is unknown.
8. False. The v-structure $F \rightarrow H \leftarrow D$ is active.
9. True. same as 2.
10. False. The trail $F \rightarrow H \rightarrow Z \rightarrow N \leftarrow D$ is active.
11. False. they share the same grandparent S.
12. False. same as 11.

Problem 1.2

1. $P(S,F,D,C,H,N,Z) = P(S) \cdot P(F|S) \cdot P(D|S) \cdot P(C|F) \cdot P(H|F,D) \cdot P(N|D,Z) \cdot P(Z|H)$.

2. $P(S,F,D,C,H,N,Z) = \frac{1}{Z}\phi_S(S)\phi_F(F)\phi_D(D)\phi_C(C)\phi_H(H)\phi_N(N)$

$\phi_Z(Z) \cdot \phi_{S,F}(S,F)\phi_{SD}(S,D)\phi_{FC}(F,C)\phi_{FH}(F,H)\phi_{DH}(D,H)\phi_{DN}(D,N)$

$\phi_{HZ}(H,Z)\phi_{NZ}(N,Z) = \frac{K}{Z}$, where $Z = \sum_{S,F,D,C,H,N,Z}(K)$.

Problem 1.3.

1. $P(F) = P(F=\text{true}|S=\text{winter})\cdot P(S=\text{winter}) + P(F=\text{true}|S=s)\cdot P(S=s)$

$= 0.4 \times 0.5 + 0.1 \times 0.5$

$\boxed{= 0.25}$

2. $P(F=\text{true}|S=\text{winter}) = \boxed{0.4}$ $\boxed{T=\text{true},\ F=\text{false}}$ $\Downarrow$

3. $P(F=\text{true}|S=\text{winter}, H=\text{true})$

$= P(F=\text{true}, D=\text{true}|S=\text{winter}, H=\text{true}) + P(F=T, D=F|S=\text{winter}, H=T)$

$$= \frac{\sum_D P(F=\text{true}, D, S=\text{winter}, H=\text{true})}{\sum_{F,D} P(S=\text{winter}, H=\text{true}, F, D)}$$

$P(F,D,S,H) = P(S)P(F|S)P(D|S)P(H|F,D)$

denominator $= \frac{1}{2}\times 0.4\times 0.1\times 0.9 + \frac{1}{2}\times 0.6\times 0.1\times 0.8 + \frac{1}{2}\times 0.4\times 0.9\times 0.8$

$+ \frac{1}{2}\times 0.6\times 0.9\times 0.3 = 0.267$

nominator $= \frac{1}{2}\times 0.4\times 0.1\times 0.9 + \frac{1}{2}\times 0.4\times 0.9\times 0.8 = 0.162$

Thus, $P(F=\text{true}|S=\text{winter}, H=\text{true}) = \frac{0.162}{0.267} = \boxed{0.6067}$

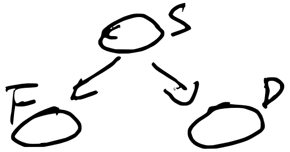
4. $P(F=\text{true}|S=\text{winter}, H=\text{true}, D=\text{true})$

$$= \frac{P(F=\text{true}, D=\text{true}, S=\text{winter}, H=\text{true})}{\sum_F P(S=\text{winter}, H=\text{true}, D=\text{true}, F)}$$

$$= \frac{\frac{1}{2}\times 0.4\times 0.1\times 0.9}{\frac{1}{2}\times\left[0.4\times 0.1\times 0.9+0.6\times 0.1\times 0.8\right]}$$

$$= \boxed{0.4286}$$

5. $\boxed{\text{Decrease}}$ if that is all your information.



is active if S is unknown.

Therefore, the infomation in D will be passed to F, which happens to decrease P(have flu) because $P(S=\text{summer} \mid D=\text{true})$ is higher, and $P(F=\text{true} \mid \text{summer})$ is low. Therefore, $\boxed{\underbrace{P(F=\text{true} \mid D=\text{true})}_{A} < \underbrace{P(F=\text{true})}_{B}}$

$$A = \frac{\sum_{S} P(F=\text{true}, D=\text{true}, S)}{\sum_{S} P(D=\text{true})} = \frac{\frac{1}{2}\times\left[0.4\times 0.1+0.1\times 0.3\right]}{\frac{1}{2}\times\left[0.1+0.3\right]} = \frac{7}{40}$$

$$B = \frac{1}{2}\times 0.4+\frac{1}{2}\times 0.1 = 0.25 = \frac{10}{40}$$

$B > A$ so, $\boxed{\text{derease}}$

Problem 1.4

1、 $\boxed{\text{No}}$ no difference.

There is no marginal independencies.

2、 $\boxed{\text{YES}}$ there is different v-structure.

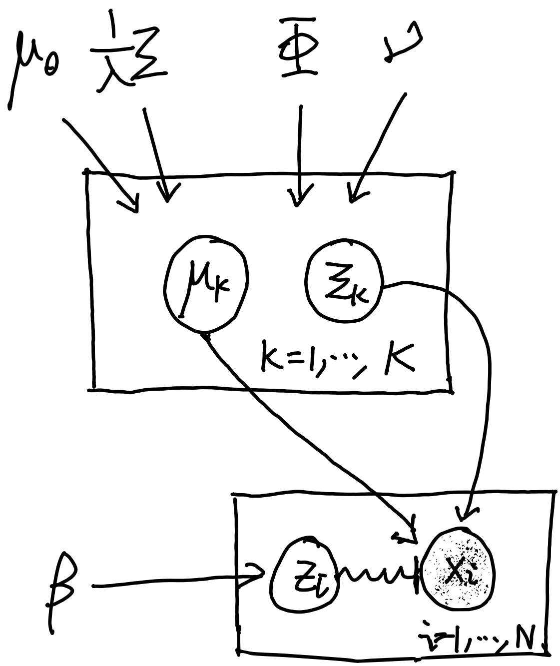
For example: $(F \perp D \mid S, H)$ in MN,

but it does not exist in BN.

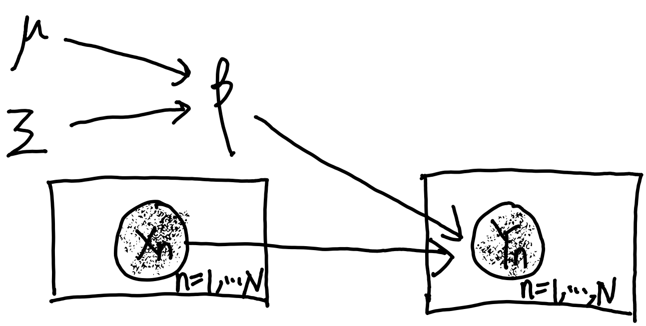
due to $F \rightarrow H \leftarrow D$

Problem 2.1

1. Gaussian Mixture Model.



2. Bayesian Logistic Regression.



Problem 2.2

Proof. If P factorizes according to G,

$$P(X_1,\dots,X_m)=\prod_{i=1}^{m}P(X_i\mid Pa(X_i))$$ can be written.

consider the last variable X_m:

$$p(X_m\mid X_1,\dots,X_{m-1})=\frac{p(X_1,\dots,X_{m-1},X_m)}{p(X_1,\dots,X_{m-1})}$$

$$=\frac{p(X_1,\dots,X_m)}{\sum_{X_m}p(X_1,\dots,X_m)}$$

$$=\frac{\prod_{i=1}^{m}p(X_i\mid Pa(X_i))}{\sum_{X_m}\prod_{i=1}^{m}p(X_i\mid Pa(X_i))}$$

[since X_m is the last]

$$=\frac{p(X_m\mid Pa(X_m))\prod_{i=1}^{m-1}p(X_i\mid Pa(X_i))}{\prod_{i=1}^{m-1}p(X_i\mid Pa(X_i))\cdot\sum_{X_m}p(X_m\mid Pa(X_m))}$$

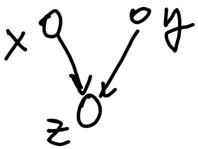
$$=p(X_m\mid Pa(X_m))$$

We can repeat this step one by one by selecting the last variable after each iteration. And thus we can get the conclusion that G is an I-map for P because $I_\ell(G)\subseteq I(P)$.

Problem 2.3

① if d-$sep_G(X;Y|Z)$, there is no active trail between $x \in X$ and $y \in Y$ given Z.

Proof: That means that, for any $x \in X$, $y \in Y$, $z \in Z$,



⇐ this v-structure does not exist.

Hence, the moralization will not introduce new link between X and Y. thus, X need to connect Z before connect Y. Thus, $sep_H(X;Y|Z)$ must hold.

② if $sep_H(X;Y|Z)$,

due to the limitation of moralization, independence from Bayesian Network $\geq$ from MN.

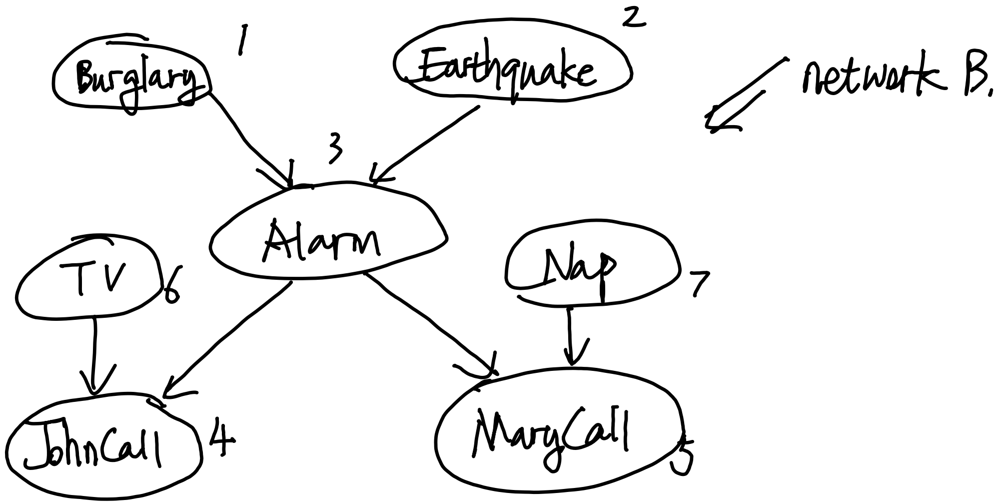
That is to say $I(H) \subseteq I(G)$

Thus, there will not be active trail between $x \in X$ and $y \in Y$. given Z.

Thus, $sep_H(X;Y|Z) \Rightarrow$ d-$sep_G(X;Y|Z)$. ∎

Problem 2.4

(a).



$$P_B(B,E,T,N,J,M) = P_B(B,E,T,N,J,M \mid A=0)\cdot P(A=0)$$
$$+ P_B(B,E,T,N,J,M \mid A=1)\cdot P(A=1)$$

[assume A is binary here. A can also have multi values.].

A is considered unobserved here.

Hence, the active trails pass A are: ⑤ $J \leftarrow A \rightarrow M$

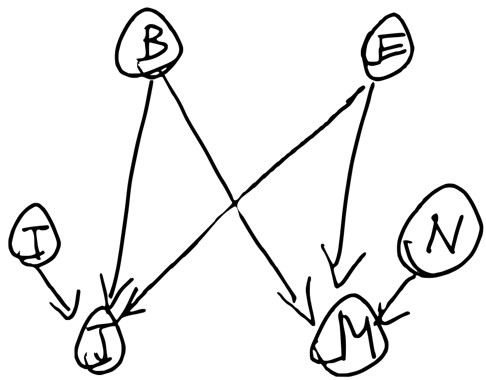
① $B \rightarrow A \rightarrow J$ ② $B \rightarrow A \rightarrow M$, ③ $E \rightarrow A \rightarrow J$ ④ $E \rightarrow A \rightarrow M$

The independencies are:

① $B \perp E \perp T \perp N$

Take the unique order B, E, T, J, N, M.

We can construct BN B' according to the above independencies and dependencies (active trails).



Problem 2.5

$I = I(\mathcal{H}) = \{(\mathbf{X} \perp \mathbf{Y} \mid \mathbf{Z}) : sep_{\mathcal{H}}(\mathbf{X}; \mathbf{Y} \mid \mathbf{Z})\}$.

$\Rightarrow$ there is no active path between $x \in \mathbf{X}$ and $y \in \mathbf{Y}$ given $\mathbf{Z}$.

$\Rightarrow$ all of the sub-paths between $x \in \mathbf{X}$ and $y \in \mathbf{Y}$ contain observed variable $\mathbf{Z}$.

$\Leftrightarrow$ for any $x \in \mathbf{X}$ and $y \in \mathbf{Y}$, their any path $x - \cdots - z - \cdots - y$ contains observed z if this path exist.

Proof of strong Union:

$(\mathbf{X} \perp \mathbf{Y} \mid \mathbf{Z}) : sep_{\mathcal{H}}(\mathbf{X}; \mathbf{Y} \mid \mathbf{Z})$

$\Rightarrow$ any sub-path contains observed variable $\mathbf{Z}$.

Adding observed variable $w \in \mathbf{W}$ will not change this fact

Thus, $(\mathbf{X} \perp \mathbf{Y} \mid \mathbf{Z}) \Rightarrow (\mathbf{X} \perp \mathbf{Y} \mid \mathbf{Z}, \mathbf{W})$

Proof of Transitivity:

consider subpath $x - \cdots - A - \cdots - y$

if $\neg(X \perp A \mid Z)$ & $\neg(A \perp Y \mid Z)$, there exist one path in

① $x - \cdots - A$ that nothing inside this path is observed, and

② $A - \cdots - y$ that nothing inside is observed.

Also, A is not observed. so $x - \cdots - y$ is active path.

Hence, $\neg(X \perp Y \mid Z)$ $\blacksquare$.

Problem 3.

(1). Ising Markov random fields

$X = \{X_1, \cdots, X_p\}$

$$p(X;\theta) \propto \exp\{\sum_{s\in V} \theta_s X_s + \sum_{(s,t)\in E} \theta_{st} X_s X_t\}.$$

For node $i \in V$

$$P(X_i = 1 \mid X_{-i};\theta) = \frac{P(X_i=1, X_{-i};\theta)}{P(X_{-i};\theta)}$$

since it is binary

$$= \frac{P(X_i=1, X_{-i};\theta)}{P(X_i=1, X_{-i};\theta) + P(X_i=0, X_{-i};\theta)}$$

$\longrightarrow$ separate P into "with X_i" part and "without X_i" part

$$= \frac{\exp\{\sum_{s\neq i}\theta_s X_s + \sum_{s\neq i}\theta_{s,t} X_s X_t\} \cdot \exp\{\theta_i + \sum_{t\in V}\theta_{i,t} X_t\}}{\exp\{\sum_{s\neq i}\theta_s X_s + \sum_{s\neq i}\theta_{s,t} X_s X_t\} \cdot [\exp\{\theta_i + \sum_{t\in V}\theta_{i,t} X_t\} + 1]}$$

$\Uparrow$

since $X_i = 0$.

$\exp(0) = 1$.

$$= \frac{\exp\{\theta_i + \sum_{t\in V}\theta_{i,t} X_t\}}{\exp\{\theta_i + \sum_{t\in V}\theta_{i,t} X_t\} + 1}$$

We know that logistic regression is in the form

$$1 - \eta_0(\vec{x}) = \frac{1}{1+\exp(-(W^T\vec{x}+b))}$$

the $\eta_0(\vec{x})$ is distribution.

Our $P(X_i=1 \mid X_{-i};\theta) = 1 - \dfrac{1}{\exp\{\theta_i + \sum_{t\in V}\theta_{i,t} X_t\}+1}$

Δ $\theta_0 + \sum_{t\in V}\theta_{i,t} X_t$ is linear.

can be written as the form $-(W^T\vec{x}+b)$

Thus, $1 - P(X_i=1 \mid X_{-i};\theta) = \dfrac{1}{1+\exp(-(W^T\vec{x}+b))}$, P is logreg

$\boxtimes$

3.2. Ising CRF.

conditional log-likelihood:

$$\mathcal{L}(X_i | W; \theta, \beta) = \log P(X_i | W; \theta, \beta)$$

$$= \sum_{s\in V} \theta_s X_s + \sum_{(s,t)\in E} \theta_{s,t} X_s X_t + \sum_{s\in V, u\in[q]} \beta_{su} X_s W_u - \log Z$$

derivative of $\mathcal{L}$ when considering a single sample:

$$\frac{\partial \mathcal{L}}{\partial \theta_s} = X_s - \frac{\partial}{\partial \theta_s} \log Z(W, \theta, \beta)$$

since $\log Z = \log \sum_{X} \exp\{\sum_{s\in V} \theta_s X_s + \sum \theta_{st} X_s X_t + \sum \beta_{su} X_s W_u\}$

$$\frac{\partial \mathcal{L}}{\partial \theta_s} = X_s - \frac{\frac{\partial Z(W, \theta, \beta)}{\partial \theta_s}}{Z(W, \theta, \beta)}$$

we only care a single sample!

$$= X_s - \frac{\sum_{X' = X_s'} \exp\{\sum \theta_s X_s' + \sum \theta_{st} X_s X_t + \sum \beta_{su} X_s W_u\} \cdot X_s'}{Z(W, \theta, \beta)}$$

$$= X_s - P(X_s' | W, \theta, \beta) \cdot X_s'$$

$$\boxed{= X_s - \mathbb{E}_{p(X_s | W, \theta, \beta)}(X_s')} = \frac{\partial \mathcal{L}}{\partial \theta_s}$$

first term　　　second term

Similarly.

$$\frac{\partial \mathcal{L}}{\partial \theta_{st}} = X_s X_t - \frac{\frac{\partial Z(W,\theta,\beta)}{\partial \theta_{st}}}{Z(W|\theta,\beta)}$$

$$= X_s X_t - \frac{\sum_{X'} X_s' X_t' \cdot \exp\{\sum \theta_s X_s + \sum \theta_{st} X_s X_t + \sum \beta_{su} X_s W_u\}}{Z(W,\theta,\beta)}$$

$$= X_s X_t - \sum_{X'} X_s' X_t' \cdot P(X'|W,\theta,\beta)$$

$$= X_s X_t - \mathbb{E}_{p(X'|W,\theta,\beta)}[X_s' X_t']$$

$$\frac{\partial \mathcal{L}}{\partial \beta_{su}} = X_s W_u - \frac{\frac{\partial Z(W,\theta,\beta)}{\partial \beta_{su}}}{Z(W,\theta,\beta)}$$

$$= X_s W_u - \frac{\sum_{X'} W_u X_s' \exp\{\sum \theta_s X_s + \sum \theta_{st} X_s X_t + \sum \beta_{su} X_s W_u\}}{Z(W,\theta,\beta)}$$

$$= X_s W_u - W_u \cdot \sum_{X'} X_s' P(X_s'|W,\theta,\beta)$$

$$= X_s W_u - W_u \cdot \mathbb{E}_{p(X_s'|W,\theta,\beta)}[X_s'].$$

Problem 4.

[See Code file].

Code, instruction, result in next page.

Result 1:

```
(base) ➜  hmm-data python eval_gene_tagger.py gene.key gene_test.p1.out
Found 2669 GENEs. Expected 642 GENEs; Correct: 424.

         precision      recall          F1-Score
GENE:    0.158861       0.660436        0.256116
(base) ➜  hmm-data
```

Result 2:

```
(base) ➜  hmm-data python eval_gene_tagger.py gene.key
Found 373 GENEs. Expected 642 GENEs; Correct: 202.

         precision      recall          F1-Score
GENE:    0.541555       0.314642        0.398030
(base) ➜  hmm-data
```

How to run my code:

Baseline

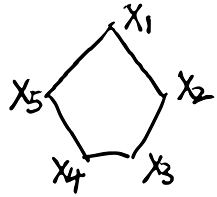
1. Use ran_hw.py function “get_final_rare_list()” to generate rare_word_list.txt
2. Use ran_hw.py function “get_rare_gene_train()” to generate rare_gene.train
3. Use ran_hw.py function “baseline_usage()” to generate gene_test.p1.out
4. Use the eval_gene_tagger.py to evaluate gene_test.p1.out

Trigram HMM

1. Use ran_hw.py function “viterbi_usage()” to generate gene_test.p2.out
2. Use the eval_gene_tagger.py to evaluate gene_test.p2.out

Note that I changed the provided original file. Please run the code inside my package.

Problem 5.



For a general MN, joint distribution $p(x_1,x_2,x_3,x_4,x_5)$ can be written as:

$$p(x_1,x_2,x_3,x_4,x_5)=\frac{\phi(x_1,x_2)\phi(x_2,x_3)\phi(x_3,x_4)\phi(x_4,x_5)\phi(x_5,x_1)}{\underbrace{\sum_{x_1,x_2,x_3,x_4,x_5}\phi(x_1,x_2)\phi(x_2,x_3)\phi(x_3,x_4)\phi(x_4,x_5)\phi(x_5,x_1)}_{\boxed{\text{Call it The Sum}}}}$$

By summing over 3, 4, we have

$$p(x_1,x_2,x_5)=\sum_{x_3,x_4}p(x_1,x_2,x_3,x_4,x_5)$$

$$=\frac{① \left[\phi(x_1,x_2)\phi(x_5,x_1)\right]\sum_{x_3,x_4}\phi(x_2,x_3)\phi(x_3,x_4)\phi(x_4,x_5)}{\sum_{x_1,x_2,x_3,x_4,x_5}\phi(x_1,x_2)\phi(x_2,x_3)\phi(x_3,x_4)\phi(x_4,x_5)\phi(x_5,x_1)}$$

Similarly,

$$p(x_2,x_4,x_5)=\frac{② \left[\phi(x_4,x_5)\right]\sum_{1,3}\phi(x_1,x_2)\phi(x_2,x_3)\phi(x_3,x_4)\phi(x_5,x_1)}{\text{The Sum}}$$

$$p(x_2,x_3,x_4)=\frac{③ \left[\phi(x_2,x_3)\phi(x_3,x_4)\right]\sum_{1,5}\phi(x_1,x_2)\phi(x_4,x_5)\phi(x_5,x_1)}{\text{The Sum}}$$

Also,

$$p(x_2,x_5)=\frac{\sum_{1,3,4}\phi(1,2)\phi(2,3)\phi(3,4)\phi(4,5)\phi(5,1)}{\text{The Sum}}$$

$$p(x_2,x_4)=\frac{\sum_{1,3,5}\phi(1,2)\phi(2,3)\phi(3,4)\phi(4,5)\phi(5,1)}{\text{The Sum}}$$

$\boxed{①②③ = P}$ ⇓

Hence, $\boxed{\text{Equation A}}$

$$\frac{P(X_1,X_2,X_5)\,P(X_2,X_4,X_5)\,P(X_2,X_3,X_4)}{\underbrace{\boxed{P}}_{=\text{①②③}}\cdot P(X_2,X_5)\,P(X_2,X_4)} = \frac{\sum_{3,4}\phi_{23}\phi_{34}\phi_{45}\cdot\sum_{1,3}\phi_{12}\phi_{23}\phi_{34}\phi_{51}\cdot\sum_{1,5}\phi_{12}\phi_{45}\phi_{51}}{\boxed{Sum}\cdot\sum_{1,3,4}\phi_{12}\phi_{23}\phi_{34}\phi_{45}\phi_{15}\cdot\sum_{1,3,5}\phi_{12}\phi_{23}\phi_{34}\phi_{45}\phi_{51}}$$

$$\sum_{1,3,4}\phi_{12}\phi_{23}\phi_{34}\phi_{45}\phi_{15} = \sum_{1}\phi_{12}\phi_{15}\cdot\sum_{3,4}\phi_{23}\phi_{34}\phi_{45}$$

$$\sum_{1,3,5}\phi_{12}\phi_{23}\phi_{34}\phi_{45}\phi_{15} = \sum_{3}\phi_{23}\phi_{34}\cdot\sum_{1,5}\phi_{12}\phi_{45}\phi_{15}$$

Thus, Equation A $= \dfrac{\sum_{1,3}\phi_{12}\phi_{23}\phi_{34}\phi_{51}\cdot\dfrac{1}{\text{①②③}}}{\sum_{1}\phi_{12}\phi_{15}\cdot\sum_{3}\phi_{23}\phi_{34}\cdot(\text{The Sum})}$

$$= \frac{\sum_{1}\phi_{12}\phi_{51}\cdot\sum_{3}\phi_{23}\phi_{34}}{\sum_{1}\phi_{12}\phi_{15}\cdot\sum_{3}\phi_{23}\phi_{34}} = \frac{1}{SUM}$$

Notice that: $P = \text{①②③} = \phi_{12}\phi_{23}\phi_{34}\phi_{45}\phi_{51}$

Thus, $\dfrac{P}{Sum} = P(X_1,X_2,X_3,X_4,X_5) = \dfrac{P(X_1,X_2,X_5)\,P(X_2,X_4,X_5)\,P(X_2,X_3,X_4)}{P(X_2,X_5)\,P(X_2,X_4)}$

∎

Marginal P tables:

I think I have written them when trying to solve the previous question. Please refer to previous pages. Thank You!