582481 Causal Analysis
Exercises #5 / Solutions
7.10.2008

5.1 The generating process is

$$\begin{aligned} P(A = \text{"}head''\text{)} &= P(A=1) = 0.5, & P(A=0) &= 0.5, \\ P(B = \text{"}head''\text{)} &= P(B=1) = 0.5, & P(B=0) &= 0.5, \end{aligned}$$

and

$$\begin{aligned} P(C = \text{"}ring''\text{)} = P(C=1) &= P(A=1)P(B=1) + P(A=0)P(B=0) \\ &= 0.5, \\ P(C=0) &= 0.5. \end{aligned}$$

Let us see whether there are such indepencenies in the joint distribution that are not implied by the generating process.

$P(a, b, c)$:

$C = 0$:

$A\backslash B$	0	1
0	0	0.25
1	0.25	0

$C = 1$:

$A\backslash B$	0	1
0	0.25	0
1	0	0.25

Let us find out whether A and C are dependent.

$P(a, c) = \sum_b P(a, b, c)$:

$A\backslash C$	0	1
0	0.25	0.25
1	0.25	0.25

$P(a)P(c)$:

$A\backslash C$	0	1
0	0.25	0.25
1	0.25	0.25

So we have $P(a, c) = P(a)P(c) \iff A \perp\!\!\!\perp C$ in the distribution $P(a, b, c)$, which is not true in the generating process. Similarily we get $P(b, c) = P(b)P(c)$. The distribution $P(a, b, c)$ is not stable.

5.2 Two directed acyclic graphs (DAGs) are d-separation-equivalent, if they have same the skeleton and they have the same v-structures. Here all graphs are over the same set of variables, and they all have the same skeleton, so let us examine the DAG-property and v-structures. In the model graph there is one v-structure: $C \to E \leftarrow A$.

(a) Not equivalent, because the v-structure in E is destroyed.

(b) Not a DAG, because there is a cycle $D \to A \to B \to D$.

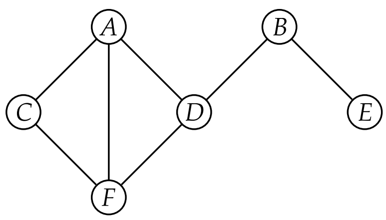
(c) Not equivalent, because there is a new v-structure $A \to B \leftarrow C$.

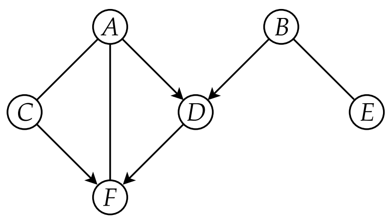
(d) Equivalent.

5.3 The following table shows all intermediate steps taken in steps 1 and 2 in the IC-algorithm:

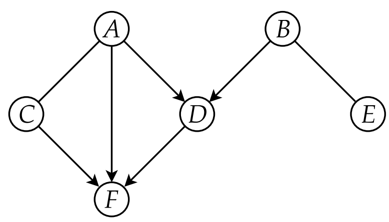
Pair	Step 1	Step 2
A, B	$S = \emptyset$	$D \notin S \Longrightarrow A \rightarrow D \leftarrow B$
A, C	$\nexists S$	—
A, D	$\nexists S$	—
A, E	$S = \emptyset$	$\nexists$ neighbours
A, F	$\nexists S$	—
B, C	$S = \emptyset$	$\nexists$ neighbours
B, D	$\nexists S$	—
B, E	$\nexists S$	—
B, F	$S = \{A, D\}$	$D \in S$
C, D	$S = \{A\}$	$A \in S, F \notin S \Longrightarrow C \rightarrow F \leftarrow D$
C, E	$S = \emptyset$	$\nexists$ neighbours
C, F	$\nexists S$	—
D, E	$S = \{B\}$	$B \in S$
D, F	$\nexists S$	—
E, F	$S = \{B\}$	$\nexists$ neighbours

In step 1, we draw an arc between the pair if no conditioning set S that makes the variables independent can be found (below left). In step 2, we add the v-structures, changing A–C–B to $A \rightarrow C \leftarrow B$, if $C \notin S$ (below right).

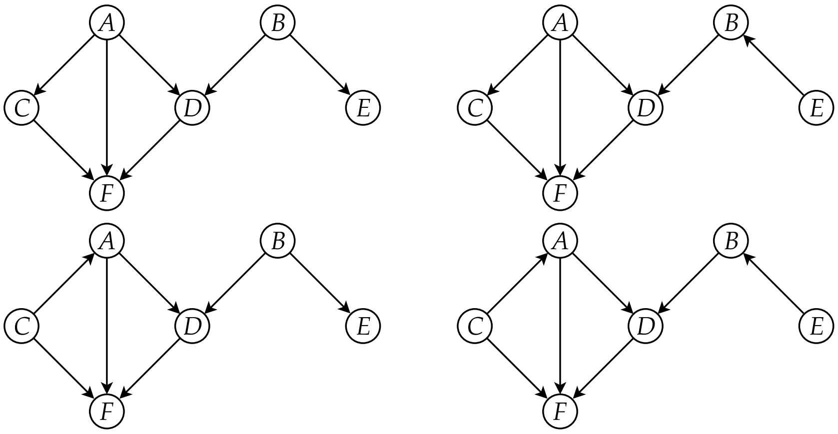




Step 3: rule R_2 directs the arc $A \rightarrow F$:



We get all the DAGs which the final pattern represents by directing the arcs A–C and B–E in every possible way:



5.4 The observed frequencies and marginal frequencies are

$Z = 0:$

$X\backslash Y$	0	1	2	Tot.
0	21	15	29	65
1	31	22	53	106
Tot.	52	37	82	171

$Z = 1:$

$X\backslash Y$	0	1	2	Tot.
0	30	54	13	97
1	13	30	9	52
Tot.	43	84	22	149

Let $n_{i\cdot k} = \sum_j o_{ijk}$ and $n_{\cdot jk} = \sum_i o_{ijk}$ be the marginal frequencies, where o_{ijk} are the observed frequencies, and $n_{\cdot\cdot k} = \sum_{ij} o_{ijk}$ the total number of observations for a given value of Z. The expected frequencies $e_{ijk} = \frac{n_{i\cdot k} n_{\cdot jk}}{n_{\cdot\cdot k}}$ are

$Z = 0:$

$X\backslash Y$	0	1	2
0	19.77	14.06	31.17
1	32.23	22.94	50.83

$Z = 1:$

$X\backslash Y$	0	1	2
0	27.99	54.68	14.32
1	15.01	29.32	7.68

The degrees of freedom is $\nu = (2-1)(3-1)2 = 4$. The test statistic is

$$
\begin{aligned}
\underline{\chi}^2 &= \sum_i \sum_j \sum_k \frac{o_{ijk}^2}{e_{ijk}} - n \\
&= \frac{21^2}{19.77} + \frac{15^2}{14.06} + \cdots + \frac{53^2}{50.83} \\
&\quad + \frac{30^2}{27.99} + \frac{54^2}{54.68} + \cdots + \frac{9^2}{7.68} - 320 \\
&\approx 1.255.
\end{aligned}
$$

Because $\underline{\chi}^2 \approx 1.255 < \chi^2_{0.05}(4) = 9.488$. we cannot reject the null hypothesis $X \perp\!\!\!\perp Y \,|\, Z$ at a significance level of 5%.