14.12 Two astronomers in different parts of the world make measurements M1 and M2 of the number of stars N in some small region of the sky, using their telescopes. Normally, there is a small possibility e of error by up to one star in each direction. Each telescope can also (with a much smaller probability f) be badly out of focus (events F1 and F2), in which case the scientist will undercount by three or more stars (or if N is less than 3, fail to detect any stars at all). Consider the three networks shown in Figure 14.22.

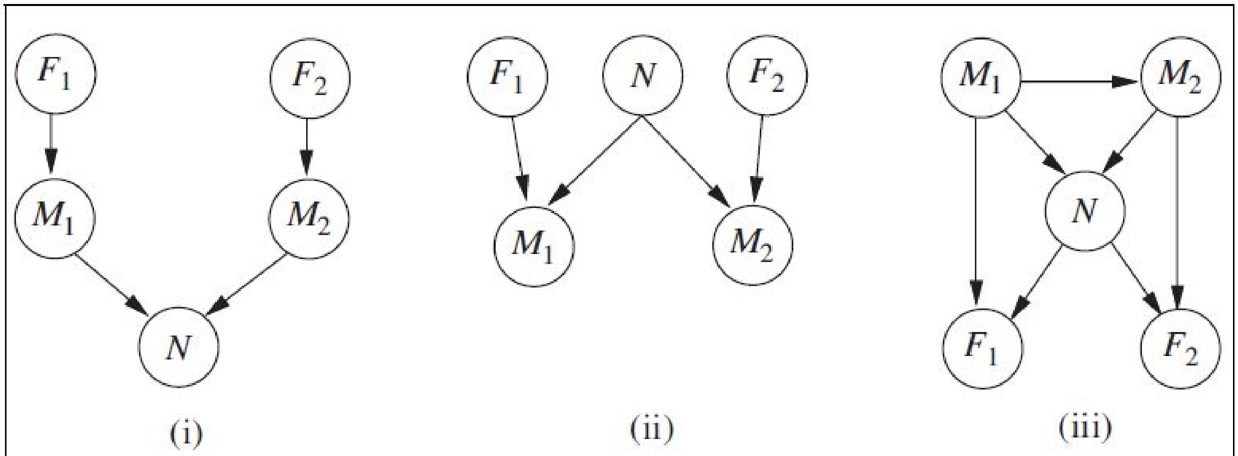

Figure 14.22 Three possible networks for the telescope problem.

a. Which of these Bayesian networks are correct (but not necessarily efficient) representations of the preceding information?
b. Which is the best network? Explain.
c. Write out a conditional distribution for P(M1 |N), for the case where N ∈{1, 2, 3} and M1 ∈{0, 1, 2, 3, 4}. Each entry in the conditional distribution should be expressed as a function of the parameters e and/or f.
d. Suppose M1 =1 and M2 =3. What are the possible numbers of stars if you assume no prior constraint on the values of N?
e. What is the most likely number of stars, given these observations? Explain how to compute this, or if it is not possible to compute, explain what additional information is needed and how it would affect the result.

a. Which of these Bayesian networks are correct (but not necessarily efficient) representations of the preceding information?

First at all and looking at the way in how astronomers do their work we can say something about these variables:

- M1 depends on F1 and N
- M2 depends on F2 and N
- F1 is independent of F2
- M1 is independent of M2
- F1 is independent of N
- F2 is independent of N

Network (i) is incorrect because it defines a CPT for variable N given M1 and M2 and it is not linked to variables F1 and F2. In a weird vision: variable N could depends on M1, M2, F1 and F2 too, but not only on M1 and M2.

Network (ii) is correct because it defines almost perfectly the variables behaviour detailed in the exercise introduction.

Network (iii) is correct but not efficient at all.

In the network (iii), M2 depends on M1. We could think that M2 will be more likely to be equal to M1 +/- 1 given M1. N will tend to be equal to M1 +/- 1 or M2 +/- 1. F1 will be more likely correct when there is no big difference between M1 and N values, and the similar will happen to F2 given M2 and N.

If we try to build a diagnostic model with links from symptoms to causes … we end up having to specify additional dependencies between otherwise independent causes (and often between separately occurring symptoms as well). If we stick to a causal model, we end up having to specify fewer numbers, and the numbers will often be easier to come up with. [AIMA 3E, "Compactness and node ordering", p.517]

b. Which is the best network? Explain.

Both networks (ii) and (iii) presents the same number and type of variables, but network (ii) presents a lower number of connections (arrows), so it will require a lower number of parameters than network (iii). Conclusion: network (ii) is the winner.

c. Write out a conditional distribution for P(M1| N), for the case where N ∈ {1, 2, 3} and M1 ∈ {0, 1, 2, 3, 4}. Each entry in the conditional distribution should be expressed as a function of the parameters e and/or f.

CPT P(M1 | N)	N=1	N=2	N=3
M1=0	CPT P(M1=0 | N=1)	CPT P(M1=0 | N=2)	CPT P(M1=0 | N=3)
M1=1	CPT P(M1=1 | N=1)	CPT P(M1=1 | N=2)	CPT P(M1=1 | N=3)
M1=2	CPT P(M1=2 | N=1)	CPT P(M1=2 | N=2)	CPT P(M1=2 | N=3)
M1=3	CPT P(M1=3 | N=1)	CPT P(M1=3 | N=2)	CPT P(M1=3 | N=3)
M4=4	CPT P(M1=4 | N=1)	CPT P(M1=4 | N=2)	CPT P(M1=4 | N=3)

I) Marginalization Rule
P(M1, N) = P(M1, N, F1) + P(M1, N, -F1)

II) Product Rule
P(M1, N, F1) = P(M1 | N, F1).P(N, F1)
P(M1, N, -F1) = P(M1 | N, -F1).P(N, -F1)

$P(M1, N) = P(M1 | N).P(N)$

III) Replace (II) in (I)
$P(M1 | N).P(N) = P(M1 | N, F1).P(N, F1) + P(M1 | N, -F1).P(N, -F1)$

IV) Product Rule of P(N, F1)
$P(M1 | N).P(N) = P(M1 | N, F1).P(F1 | N).P(N) + P(M1 | N, -F1).P(-F1 | N).P(N)$
$P(M1 | N) = P(M1 | N, F1).P(F1 | N) + P(M1 | N, -F1).P(-F1 | N)$

V) Focus does not depend on number of stars
$P(M1 | N) = P(M1 | N, F1).P(F1) + P(M1 | N, -F1).P(-F1)$

VI) If -F1 means not focused, then $P(F1) = 1 - P(-F1)$
$P(M1 | N) = P(M1 | N, F1).(1-P(-F1)) + P(M1 | N, -F1).P(-F1)$
VII) If $P(-F1) = f$ then:
$P(M1 | N) = P(M1 | N, F1).(1-f) + P(M1 | N, -F1).f$

And now we can use the above formula to view every special case.

$P(M1=0 | N=1)$ $= P(M1=0 | N=1, F1).(1-f) + P(M1=0 | N=1, -F1).f$
$P(M1=0 | N=1, F1) = e$ by hypothesis
$P(M1=0 | N=1, -F1) = 1$ by hypothesis
$P(M1=0 | N=1) = e.(1-f) + f$

$P(M1=0 | N=2)$ $= P(M1=0 | N=2, F1).(1-f) + P(M1=0 | N=2, -F1).f$
$P(M1=0 | N=2, F1) = 0$ by hypothesis
$P(M1=0 | N=2, -F1) = 1$ by hypothesis
$P(M1=0 | N=2) = 0.(1-f) + 1.f = f$

$P(M1=0 | N=3)$ $= P(M1=0 | N=3, F1).(1-f) + P(M1=0 | N=3, -F1).f$
$P(M1=0 | N=3, F1) = 0$ by hypothesis
$P(M1=0 | N=3, -F1) = 1$ by hypothesis
$P(M1=0 | N=3) = 0.(1-f) + 1.f = f$

$P(M1=1 | N=1)$ $= P(M1=1 | N=1, F1).(1-f) + P(M1=1 | N=1, -F1).f$
$P(M1=0 | N=1, F1) = P(M1=2 | N=1, F1) = e$
$P(M1=1 | N=1, F1) + P(M1=3 | N=1, F1) + P(M1=4 | N=1, F1) = 1 - 2e$
$P(M1=1 | N=1, F1) + 0 + 0 = 1 - 2e$
$P(M1=1 | N=1, F1) = 1 - 2e$
$P(M1=1 | N=1) = (1 - 2e).(1-f) +0.f = (1 - 2e).(1 - f)$

$P(M1=1 | N=2)$ $= P(M1=1 | N=2, F1).(1-f) + P(M1=0 | N=2, -F1).f$
$P(M1=1 | N=2, F1) = e$ by hypothesis
$P(M1=1 | N=2, -F1) = 0$ by hypothesis
$P(M1=1 | N=2) = e.(1-f) + 0.f = e.(1-f)$

$P(M1=1 | N=3)$ $= P(M1=1 | N=3, F1).(1-f) + P(M1=1 | N=3, -F1).f$
$P(M1=1 | N=3, F1) = 0$ by hypothesis

P(M1=1 | N=3, -F1) = 0 by hypothesis
P(M1=1 | N=3) = 0.(1-f) + 0.f = 0

P(M1=2 | N=1) = P(M1=2 | N=1, F1).(1-f) + P(M1=2 | N=1, -F1).f
P(M1=2 | N=1, F1) = e by hypothesis
P(M1=2 | N=1, -F1) = 0 by hypothesis
P(M1=2 | N=1) = e.(1-f) + 0.f = e.(1-f)

P(M1=2 | N=2) = P(M1=2 | N=2, F1).(1-f) + P(M1=2 | N=2, -F1).f
P(M1=2 | N=2, F1) = 1 - 2e (idem case P(M1=1 | N=1, F1))
P(M1=2 | N=2, -F1) = 0 by hypothesis
P(M1=2 | N=2) = (1 - 2e).(1-f) + 0.f = (1 - 2e).(1-f)

P(M1=2 | N=3) = P(M1=2 | N=3, F1).(1-f) + P(M1=2 | N=3, -F1).f
P(M1=2 | N=3, F1) = e
P(M1=2 | N=3, -F1) = 0
P(M1=2 | N=3) = e.(1-f) + 0.f = e.(1-f)

P(M1=3 | N=1) = P(M1=3 | N=1, F1).(1-f) + P(M1=3 | N=1, -F1).f
P(M1=3 | N=1, F1) = 0 by hypothesis
P(M1=3 | N=1, -F1) = 0 by hypothesis
P(M1=3 | N=1) = 0.(1-f) + 0.f = 0

P(M1=3 | N=2) = P(M1=3 | N=2, F1).(1-f) + P(M1=3 | N=2, -F1).f
P(M1=3 | N=2, F1) = e by hypothesis
P(M1=3 | N=2, -F1) = 0 by hypothesis
P(M1=3 | N=2) = e.(1-f) + 0.f = e.(1-f)

P(M1=3 | N=3) = P(M1=3 | N=3, F1).(1-f) + P(M1=3 | N=3, -F1).f
P(M1=3 | N=3, F1) = 1 - 2e (idem case P(M1=1 | N=1, F1))
P(M1=3 | N=3, -F1) = 0 by hypothesis
P(M1=3 | N=3) = (1 - 2e).(1 - f) + 0.f = (1 - 2e).(1 - f)

P(M1=4 | N=1) = P(M1=4 | N=1, F1).(1-f) + P(M1=4 | N=1, -F1).f
P(M1=4 | N=1, F1) = 0 by hypothesis
P(M1=4 | N=1, -F1) = 0 by hypothesis
P(M1=4 | N=1) = 0.(1-f) + 0.f = 0

P(M1=4 | N=2) = P(M1=4 | N=2, F1).(1-f) + P(M1=4 | N=2, -F1).f
P(M1=4 | N=2, F1) = 0 by hypothesis
P(M1=4 | N=2, -F1) = 0 by hypothesis
P(M1=4 | N=2) = 0.(1-f) + 0.f = 0

P(M1=4 | N=3) = P(M1=4 | N=3, F1).(1-f) + P(M1=4 | N=3, -F1).f
P(M1=4 | N=3, F1) = e by hypothesis
P(M1=4 | N=3, -F1) = 0 by hypothesis
P(M1=4 | N=3) = e.(1 - f) + 0.f = e.(1 - f)

CPT P(M1 | N)	N=1	N=2	N=3
M1=0	e.(1 - f) + f	f	f
M1=1	(1 - 2e).(1 - f)	e.(1 - f)	0
M1=2	e.(1 - f)	(1 - 2e).(1 - f)	e.(1 - f)
M1=3	0	e.(1 - f)	(1 - 2e).(1 - f)
M4=4	0	0	e.(1 - f)

d. Suppose M1 =1 and M2 =3. What are the possible numbers of stars if you assume no prior constraint on the values of N?

CPT of P(M2 | N) works the same as P(M1 | N). Then we can see both CPTs and the probabilities for N in the previous range:

	N=1	N=2	N=3
P(M1=1 | N)	(1 - 2e).(1 - f)	e.(1 - f)	0
P(M2=3 | N)	0	e.(1 - f)	(1 - 2e).(1 - f)

Looking at the table above the only case where it is possible that P(M1=1| N) > 0 and P(M2=3 | N) > 0 is when the number of stars is equal to 2 (two).

Now let see what happen with no prior constraint over N. Then N >= 0. I can not think of a negative number of stars over a region.

We need to look for cases of N where P(M1=1 | N=n) > 0 and P(M2=3 | N=n) > 0:

Case P(M1=1 | N) > 0

P(M1=1 | N) = P(M1=1 | N, F1).(1-f) + P(M1=1 | N, -F1).f > 0
With 0 < f < 1 then we need to check that at least P(M1=1 | N, F1) > 0 or P(M1=1 | N, -F1) > 0
P(M1=1 | N, F1) > 0 for all pairs (M1, N) where |M1-N| <= 1
P(M1=1 | N, -F1) > 0 for N=0 or N>=4 else this probability is always zero (1 <= N <=3).
Then P(M1=1 | N) > 0 ⇔ P(M1=1 | N, F1) > 0 ⇔ |M1-N| <= 1 with N > 0

Case P(M2=3 | N) > 0

P(M2=3 | N) = P(M2=3 | N, F2).(1-f) + P(M2=3 | N, -F2).f > 0
With 0 < f < 1 then we need to check that at least P(M2=3 | N, F2) > 0 or P(M2=3 | N, -F2) > 0
P(M2=3 | N, F2) > 0 for all pairs (M2, N) where |M2-N| <= 1
P(M2=3 | N, -F2) > 0 for 0 <= N <= 3 or N >= M2+3 >= 6

So, these is the general condition for valid values of N:
(|M1-N| <= 1 or N=0 or N>=4) and (|M2-N| <= 1 or 0 < N <= 3 or N>=6)

	P(M1=1 | N)	P(M2=3 | N)
N=0	>0	0
N=1	0	>0
N=2	>0	>0
N=3	0	>0
N=4	>0	>0
N=5	>0	0
N=6	>0	>0
N=7	>0	>0
N>7	>0	>0

Answer: N=2 or N=4 or N>=6

e. What is the most likely number of stars, given these observations? Explain how to compute this, or if it is not possible to compute, explain what additional information is needed and how it would affect the result.

	P(M1=1 | N)	P(M2=3 | N)
N=0	e.(1-f)+f	0
N=1	(1-2e).(1-f)	0
N=2	e.(1-f)	e.(1-f)
N=3	0	(1-2e).(1-f)
N=4	P(M1=1 | N=4, -F1).f	e.(1-f)
N=5	P(M1=1 | N=5, -F1).f	0
N=6	P(M1=1 | N=6, -F1).f	P(M2=3 | N=6, -F2).f
N=7	P(M1=1 | N=7, -F1).f	P(M2=3 | N=7, -F2).f
N=n (n >=6)	P(M1=1 | N=n, -F1).f	P(M2=3 | N=n, -F2).f

P(N=2 | M1=1, M2=3) = P(M1=1, M2=3 | N=2) * P(N=2) / P(M1=1, M2=3)
P(N=2 | M1=1, M2=3) = alpha * P(M1=1, M2=3 | N=2) * P(N=2)
P(N=2 | M1=1, M2=3) = alpha * P(M1=1 | N=2) * P(M2=3 | N=2) * P(N=2)
P(N=2 | M1=1, M2=3) = alpha * e(1-f).e(1-f) x P(N=2)

P(N=4 | M1=1, M2=3) = P(M1=1, M2=3 | N=4) * P(N=4) / P(M1=1, M2=3)
P(N=4 | M1=1, M2=3) = alpha * P(M1=1, M2=3 | N=4) * P(N=4)
P(N=4 | M1=1, M2=3) = alpha * P(M1=1| N=4) * P(M2=3| N=4) * P(N=4)
P(N=4 | M1=1, M2=3) = alpha * P(M1=1 | N=4, -F1).f.e(1-f) * P(N=4)

P(N=6 | M1=1, M2=3) = P(M1=1, M2=3 | N=6) * P(N=6) / P(M1=1, M2=3)
P(N=6 | M1=1, M2=3) = alpha * P(M1=1, M2=3 | N=6) * P(N=6)
P(N=6 | M1=1, M2=3) = alpha * P(M1=1| N=6) * P(M2=3| N=6) * P(N=6)
P(N=6 | M1=1, M2=3) = alpha * P(M1=1 | N=6, -F1).f * P(M2=3 | N=6, -F2).f * P(N=6)

Just to simplify comparison we can think of N as a uniform distributed variable. Then P(N=n) = p for all n. We need to compare:

e(1-f).e(1-f) vs P(M1=1 | N=4, -F1).f.e(1-f) vs P(M1=1 | N=6, -F1).f * P(M2=3 | N=6, -F2).f

Another observations is over term P(M1=1 | N=4, -F1). For N=4 this probability is > 0 only for M1=1 and M1=0. For M1 > 1 is zero. So P(M1=1 | N=4, -F1) could be as most close to 1. We can observe something similar on P(M1=1 | N=6, -F1) and P(M2=3 | N=6, -F2). So in the case that this probabilities were close to one, the comparison can be boiled down to:

e(1-f).e(1-f) vs f.e.(1-f) vs f.f

Case N=2 vs N=4:
N=2 will win to N=4 when e/(e+1) > f
And e > e/(e+1) Then N=2 because e > f by hypothesis.
N is more likely to be 2 that 4

Case N=2 vs N=6:
N=2 will win to N=4 when e(1-f).e(1-f) > f.f
e(1-f) > f
e > f / (1-f) > f
e > f by hypothesis (*with a much smaller probability f*)
N is more likely to be 2 that 6

The likely number of stars is 2 when M1=1 and M2=3