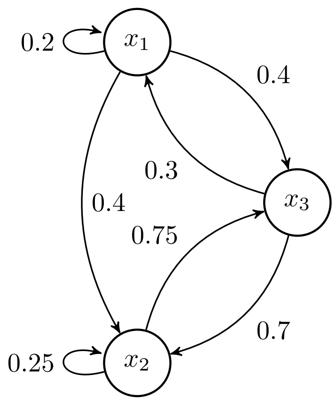

Figure 1: Markov chain 1

1 Sampling methods

Exercise 1.1 *Forward Sampling. One strategy for obtaining an estimate of the conditional probability $P(\boldsymbol{y}|\boldsymbol{e})$ is by using forward sampling to estimate $P(\boldsymbol{y}, \boldsymbol{e})$ and $P(\boldsymbol{e})$ separately, and then computing the ratio. We can use the Hoeffding Bound to obtain a bound on both the numerator and the denominator. Assuming that M is large, when does the resulting bound provide meaningful guarantees? Recall that we need $M \leq \frac{\log(2/\delta)}{2\epsilon^2}$ to get an additive error bound ϵ that holds with probability $(1-\delta)$ for our estimate.*

a) It always provides meaningful guarantees.

b) It never provides a meaningful guarantee.

c) It provides a meaningful guarantee, but only when δ is small relative to $P(\boldsymbol{y}, \boldsymbol{e})$ and $P(\boldsymbol{e})$.

d) It provides a meaningful guarantee, but only when ϵ is small relative to $P(\boldsymbol{y}, \boldsymbol{e})$ and $P(\boldsymbol{e})$.

Exercise 1.2 *Rejecting Samples. Consider the process of rejection sampling to generate samples from the posterior distribution $P(\boldsymbol{Y}|\boldsymbol{e})$. If we want to obtain M samples, what is the expected number of samples that would need to be drawn from $P(\boldsymbol{X})$?*

a) $M \cdot P(\boldsymbol{e})$

b) $M \cdot P(\boldsymbol{Y}|\boldsymbol{e})$

c) $M \cdot (1 - P(\boldsymbol{e}))$

d) $M \cdot (1 - P(\boldsymbol{Y}|\boldsymbol{e}))$

e) $M/P(\boldsymbol{e})$

f) $M/(1 - P(\boldsymbol{e}))$

Exercise 1.3 *Stationary Distributions. Consider the simple Markov chain of Figure 1. By definition, a stationary distribution ϕ for this chain must satisfy which of the following properties? You may select 1 or more options, or none of them.*

a) $\phi(x_3) = 0.4\pi(x_1) + 0.5\phi(x_2)$

b) $\phi(x_1) = 0.2\pi(x_1) + 0.3\phi(x_3)$

c) $\phi(x_1) = 0.2\pi(x_1) + 0.4\phi(x_2) + 0.4\phi(x_3)$

d) $\phi(x_1) = \pi(x_2) = \phi(x_3)$

e) $\phi(x_1) + \pi(x_2) + \phi(x_3) = 1$

f) $\phi(x_3) = 0.3\pi(x_1) + 0.7\phi(x_3)$

Exercise 1.4 *Gibbs Sampling in a Bayesian Network. Suppose that we have the Bayesian network of Figure 2. If we are sampling the variable X_{23} as a substep of Gibbs sampling, which is the closed form for the distribution that we use to sample the value x'_{23}?*

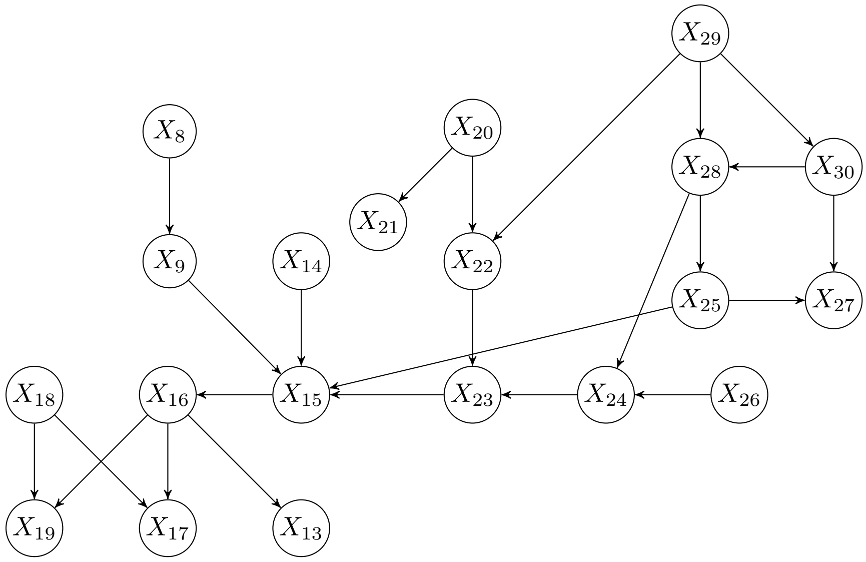

Figure 2: Large Bayesian network 1

a) $P(x_{23}'|x_{22}, x_{24})$

b) $P(x_{23}'|x_{22}, x_{24})P(x_{15}|x_{23}', x_{14}, x_9, x_{25})$

c) $\Big(P(x_{23}'|x_{22}, x_{24})P(x_{15}|x_{23}', x_{14}, x_9, x_{25})\Big)/\Big(\sum_{x_{22}'',x_{24}'',x_{15}'',x_{14}'',x_9'',x_{25}''} P(x_{23}'|x_{22}'', x_{24}'')P(x_{15}''|x_{23}', x_{14}'', x_9'', x_{25}'')\Big)$

d) $P(x_{23}'|\boldsymbol{x}_{-23})$ *where* $\boldsymbol{x}_{-23}$ *is the tuple of values for all the rest of variables but* X_{23}

e) $\Big(P(x_{23}'|x_{22}, x_{24})P(x_{15}|x_{23}', x_{14}, x_9, x_{25})\Big)/\Big(\sum_{x_{23}''} P(x_{23}''|x_{22}, x_{24})P(x_{15}|x_{23}'', x_{14}, x_9, x_{25})\Big)$

f) None, as these are all either incorrect or not in closed form

Exercise 1.5 *Gibbs Sampling. Suppose we are running the Gibbs sampling algorithm on the chain Bayesian network* $X \to Y \to Z$. *If the current sample is* (x_0, y_0, z_0) *and we sample* Y *as the first substep of the Gibbs sampling process, which is the probability that the next sample is* (x_0, y_1, z_0)*?*

a) $P(x_0, y_1, z_0)$

b) $P(x_0, z_0|y_1)$

c) $P(y_1|x_0, z_0)$

d) $P(y_1|x_0)$

Exercise 1.6 *Collecting samples from Markov chains. Assume we run a Markov chain for a sufficient burn-in time, and now we wish to collect samples so that we can use them to estimate* $P(X_i = 1)$. *Can we collect and use every sample from the Markov chain after the burn-in?*

a) No, once we collect one sample, we have to continue running the chain in order to re-mix it before we get another sample.

b) Yes, and if we collect m consecutive samples, we can use the Hoeffding bound to provide (high-probability) bounds on the error in our estimated probability.

c) No, Markov chains are only good for one sample; we have to restart the chain and burn-in it before we can collect another sample.

d) *Yes, that would give a correct estimate of the probability. However, we cannot apply the Hoeffding bound to estimate the error in our estimate.*

Exercise 1.7 *Markov Chain mixing. Which of the following classes of chains would you expect to have the shortest mixing time in general?*

a) *Markov chains where state spaces are well connected and transitions between states have high probabilities.*

b) *Markov chains for networks with nearly deterministic potentials.*

c) *Markov chains with many distinct and peaked probability modes.*

d) *Markov chains with distinct regions in the state space that are connected by low probability transitions.*

Answers

Ex. 1.1: d

Ex. 1.2: e

Ex. 1.3: b,e

Ex. 1.4: e

Ex. 1.5: c

Ex. 1.6: d

Ex. 1.7: a