take the drug. Further, patients having the syndrome are more likely to avoid the drug, with probabilities $q_1 = P(x_1|z_0)$ and $q_2 = p(x_1|z_1)$.

(a) *Based on this model, compute the joint distributions $P(x, y, z), P(x, y), P(x, z)$ and $P(y, z)$ for all values of $x, y,$ and z, in terms of the parameters $(r, p_1, p_2, p_3, p_4, q_1, q_2)$.*

   Hint: Use the product decomposition of Section 1.5.2.

(b) *Calculate the difference $P(y_1|x_1) - P(y_1|x_o)$ for three populations: (1) those carrying the syndrome, (2) those not carrying the syndrome and (3) the population as a whole.*

(c) *Using your results for (b), find a combination of parameters that exhibits Simpson's reversal.*

Study question 1.5.3
Consider a graph $X_1 \to X_2 \to X_3 \to X_4$ of binary random variables, and assume that the conditional probabilities between any two consecutive variables are given by

$$P(X_i = 1|X_{i-1} = 1) = p$$
$$P(X_i = 1|X_{i-1} = 0) = q$$
$$P(X_1 = 1) = p_0$$

Compute the following probabilities

$$P(X_1 = 1, X_2 = 0, X_3 = 1, X_4 = 0)$$
$$P(X_4 = 1|X_1 = 1)$$
$$P(X_1 = 1|X_4 = 1)$$
$$P(X_3 = 1|X_1 = 0, X_4 = 1)$$

Study question 1.5.4
Define the structural model that corresponds to the Monty Hall problem, and use it to describe the joint distribution of all variables.

Bibliographical Notes for Chapter 1

An extensive account of the history of Simpson's paradox is given in Pearl (2009, pp. 174–182), including many attempts by statisticians to resolve it without invoking causation. A more recent account, geared for statistics instructors is given in (Pearl 2014c). Among the many texts that provide basic introductions to probability theory, Lindley (2014) and Pearl (1988, Chapters 1–2) are the closest in spirit to the Bayesian perspective used in Chapter 1. The textbooks by Selvin (2004) and Moore et al. (2014) provide excellent introductions to classical methods of statistics, including parameter estimation, hypothesis testing and regression analysis.

The Monty Hall problem, discussed in Section 1.3, appears in many introductory books on probability theory (e.g., Grinstead and Snell 1998, p. 136; Lindley 2014, p. 201) and

is mathematically equivalent to the "Three Prisoners Dilemma" discussed in (Pearl 1988, pp. 58–62). Friendly introductions to graphical models are given in Elwert (2013); Glymour and Greenland (2008), and the more advanced texts of Pearl (1988, Chapter 3), Lauritzen (1996) and Koller and Friedman (2009). The product decomposition rule of Section 1.5.2 was used in Howard and Matheson (1981) and Kiiveri et al. (1984) and became the semantic basis of *Bayesian Networks* (Pearl 1985)—directed acyclic graphs that represent probabilistic knowledge, not necessarily causal. For inference and applications of Bayesian networks, see Darwiche (2009), Fenton and Neil (2013), and Conrady and Jouffe (2015). The validity of the product decomposition rule for structural causal models was shown in Pearl and Verma (1991).