Bibliographical Notes for Chapter 2

The distinction between chains and forks in causal models was made by Simon (1953) and Reichenbach (1956) while the treatment of colliders (or common effect) can be traced back to the English economist Pigou (1911) (see Stigler 1999, pp. 36–41). In epidemiology, colliders came to be associated with “Selection bias” or “Berkson paradox” (Berkson 1946) while in artificial intelligence it came to be known as the “explaining away effect” (Kim and Pearl 1983). The rule of $d$-separation for determining conditional independence by graphs (Definition 2.4.1) was introduced in Pearl (1986) and formally proved in Verma and Pearl (1988) using the theory of graphoids (Pearl and Paz 1987). Gentle introductions to $d$-separation are available in Hayduk et al. (2003), Glymour and Greenland (2008), and Pearl (2009, pp. 335–337). Algorithms and software for detecting $d$-separation, as well as finding minimal separating sets are described in Tian et al. (1998), Kyono (2010), and Textor et al. (2011). The advantages of local over global model testing, are discussed in Pearl (2009, pp. 144–145) and further elaborated in Chen and Pearl (2014). Recent applications of $d$-separation include extrapolation across populations (Pearl and Bareinboim 2014) and handling missing data (Mohan et al. 2013). In extrapolation problems, also called “external validity,” “transportability,” or “meta analysis,” graphical methods enable statisticians to take experimental findings from one population and apply them to another population, differing from the first in several statistical and causal features. In missing data applications, graphical models tell analysts how to compute bias-free estimates from data in which the values of some variables are missing.