Figure 14.5 A simple network with discrete variables (Subsidy and Buys) and continuous variables (Harvest and Cost).

possible values into a fixed set of intervals. For example, temperatures could be divided into (<0°C), (0°C-100°C), and (>100°C) Discretization is sometimes an adequate solution, but often results in a considerable loss of accuracy and very large CPTs. The most common solution is to define standard families of probability density functions (see Appendix A) that are specified by a finite number of parameters. For example, a Gaussian (or normal) distribution $N(\mu, \sigma^2)$ has the mean $\mu$ and the variance $\sigma^2$ as parameters. Yet another solution—sometimes called a nonparametric representation—is to define the conditional distribution implicitly with a collection of instances, each containing specific values of the parent and child variables. We explore this approach further in Chapter 18.

A network with both discrete and continuous variables is called a hybrid Bayesian network. To specify a hybrid network, we have to specify two new kinds of distributions: the conditional distribution for a continuous variable given discrete or continuous parents; and the conditional distribution for a discrete variable given continuous parents. Consider the simple example in Figure 14.5, in which a customer buys some fruit depending on its cost, which depends in turn on the size of the harvest and whether the government’s subsidy scheme is operating. The variable Cost is continuous and has continuous and discrete parents; the variable Buys is discrete and has continuous and discrete parents.

For the Cost variable, we need to specify $P(Cost \mid Harvest, Subsidy)$. The discrete parent is handled by enumeration—that is, by specifying both $P(Cost \mid Harvest, subsidy)$ and $P(Cost \mid Harvest, \neg subsidy)$. To handle Harvest, we specify how the distribution over the cost $c$ depends on the continuous value $h$ of Harvest. In other words, we specify the parameters of the cost distribution as a function of $h$. The most common choice is the linear Gaussian distribution, in which the child has a Gaussian distribution whose mean $\mu$ varies linearly with the value of the parent and whose standard deviation $\sigma$ is fixed. We need two distributions, one for subsidy and one for $\neg subsidy$, with different parameters:

$$P(c \mid h, subsidy) = N(\mu, \sigma^2)$$

$$P(c \mid h, \neg subsidy) = \frac{1}{2\pi \sigma_f^2} e^{-\frac{(c-fh)^2}{2\sigma_f^2}}$$

For this example, then, the conditional distribution for Cost is specified by naming the Linear Gaussian distribution and providing the parameters $a$, $b$, $\sigma_f$, and $\sigma_f$. Figures 14.6(a)
Figure 14.6 The graphs in (a) and (b) show the probability distribution over Cost as a function of Harvest size, with Subsidy true and false, respectively. Graph (c) shows the distribution $P(Cost | Harvest)$, obtained by summing over the two subsidy cases.

and (b) show these two relationships. Notice that in each case the slope is negative, because cost decreases as supply increases. (Of course, the assumption of linearity implies that the cost becomes negative at some point; the linear model is reasonable only if the harvest size is limited to a narrow range.) Figure 14.6(c) shows the distribution $P(Cost | h)$, averaging over the two possible values of Subsidy and assuming that each has prior probability 0.5. This shows that even with very simple models, quite interesting distributions can be represented.

The linear Gaussian conditional distribution has some special properties. A network containing only continuous variables with linear Gaussian distributions has a joint distribution that is a multivariate Gaussian distribution (see Appendix A) over all the variables (Exercise 14.9). Furthermore, the posterior distribution given any evidence also has this property. When discrete variables are added as parents (not as children) of continuous variables, the network defines a conditional Gaussian, or CG, distribution: given any assignment to the discrete variables, the distribution over the continuous variables is a multivariate Gaussian.

Now we turn to the distributions for discrete variables with continuous parents. Consider, for example, the Buys node in Figure 14.5. It seems reasonable to assume that the customer will buy if the cost is low and will not buy if it is high and that the probability of buying varies smoothly in some intermediate region. In other words, the conditional distribution is like a "soft" threshold function. One way to make soft thresholds is to use the integral of the standard normal distribution:

$$\Phi(x) = \int_{-\infty}^{x} N(0,1)(t)dt$$

Then the probability of Buys given Cost might be

$$P(buys | Cost = c) = \Phi\left(\frac{c - \mu}{\sigma}\right)$$

which means that the cost threshold occurs around $\mu$, the width of the threshold region is proportional to $\sigma$, and the probability of buying decreases as cost increases. This probit distribution leads to fast inference in linear Gaussian networks.

It follows that inference in linear Gaussian networks takes only $O(n^3)$ time in the worst case, regardless of the network topology. In Section 14.4, we see that inference for networks of discrete variables is NP-hard.