than training one from scratch. Furthermore, it only applies
to DNs with tree CPDs.

4 CONVERTING CONSISTENT DEPENDENCY NETWORKS

We now discuss how to construct a joint distribution from a set of positive conditional distributions. To begin with, we assume that the conditional distributions, \( P(X_j | X_{-j}) \), are the conditionals of some unknown joint probability distribution \( P \). Later, we will relax this assumption and discuss how to find the most effective approximation.

Consider two instances, \( x \) and \( x' \), that only differ in the state of one variable, \( X_j \). In other words, \( x_i = x'_i \) for all \( i \neq j \), so \( x_{-j} = x'_{-j} \). We can express the ratio of their probabilities as follows:

\[
\frac{P(x)}{P(x')} = \frac{P(x_j, x_{-j})}{P(x'_j, x'_{-j})} = \frac{P(x_j | x_{-j})P(x_{-j})}{P(x'_j | x'_{-j})P(x'_{-j})} = \frac{P(x_j | x_{-j})}{P(x'_j | x'_{-j})} \tag{3}
\]

Note that the final expression only involves a conditional distribution, \( P(X_j | X_{-j}) \), not the full joint. The conditional distribution must be positive in order to avoid division by zero.

If \( x \) and \( x' \) differ in multiple variables, then we can express their probability ratio as the product of multiple single-variable transitions. We construct a sequence of instances \( \{x^{(0)}, x^{(1)}, \ldots, x^{(n)}\} \), each instance differing in at most one variable from the previous instance in the sequence. Let order \( o \in S_n \) be a permutation of the numbers 1 to \( n \) and let \( o[i] \) refer to the \( i \)th number in the order. We define \( x^{(i)} \) inductively:

\[
x^{(0)} = x \\
x^{(i)} = (x'_{o[i]}, x'_{-o[i]})
\]

In other words, the \( i \)th instance \( x^{(i)} \) simply changes the \( o[i] \)th variable from \( x_{o[i]} \) to \( x'_{o[i]} \) and is otherwise identical to the previous element, \( x^{(i-1)} \). Thus, in \( x^{(i)} \) the first \( i \) variables in the order are set to their values in \( x' \) and the latter \( n - i \) are set to their values in \( x \). Note that \( x^{(n)} = x' \), since all \( n \) variables have been changed to their values in \( x' \).

We can use these intermediate instances to express the ratio \( P(x)/P(x') \) as a product:

\[
\frac{P(x)}{P(x')} = \frac{P(x^{(0)})}{P(x^{(n)})} = \frac{P(x^{(0)})}{P(x^{(1)})} \times \cdots \times \frac{P(x^{(n-1)})}{P(x^{(n)})} = \prod_{i=1}^{n} \frac{P(x^{(i-1)}/x^{(i)})}{P(x^{(i)})} = \prod_{i=1}^{n} \frac{P(x_{o[i]} | x_{-o[i]})}{P(x'_{o[i]} | x'_{-o[i]})} \tag{4}
\]

The last equality follows from substituting (3), since \( x^{(i-1)} \) and \( x^{(i)} \) only differ in the \( o[i] \)th variable, \( X_{o[i]} \).

Letting \( \phi_i(X) = P(x_{o[i]} | x_{-o[i]})/P(x'_{o[i]} | x'_{-o[i]}) \) and \( Z = 1/P(x') \):

\[
\frac{1}{Z} \sum_i \phi_i(x) = P(x') \frac{P(x)}{P(x')} = P(x)
\]

Therefore, a Markov network with the factors \( \{\phi_i\} \) exactly represents the probability distribution \( P(X) \). Since the factors are defined only in terms of conditional distributions of one variable given evidence, any consistent dependency network can be converted to a Markov network representing the exact same distribution. This holds for any ordering \( o \) and base instance \( x' \).

Note that, unlike \( x' \), \( x \) is not a fixed vector of values but can be set to be any instance in the state space. This is necessary for \( \phi_i \) to be a function \( x \). The following subsection will make this clearer with an example.

4.1 EXAMPLE

We now show how this idea can be applied to a simple, consistent DN. Consider the following conditional distributions over binary variables \( X_1 \) and \( X_2 \):

\[
P(X_1 = T | X_2 = T) = 4/5 \\
P(X_1 = T | X_2 = F) = 2/5 \\
P(X_2 = T | X_1 = T) = 2/3 \\
P(X_2 = T | X_1 = F) = 1/4
\]

Let \( x' = [T, T] \) and \( o = [1, 2] \). Following the earlier construction:

\[
\phi_1(x_1, x_2) = \frac{P(x_1 | x_1')}{P(x_1 | x_1')} = \frac{P(x_1 | x_2)}{P(X_1 = T | x_2)} \\
\phi_2(x_2) = \frac{P(x_2 | x_2')}{P(x_2 | x_2')} = \frac{P(x_2 | x_1 = T)}{P(X_2 = T | x_1 = T)}
\]

Note that \( \phi_2 \) is not a function of \( x_1 \). This is because \( x_1^{(2)} = x_2' \), so the value of \( X_1 \) in the evidence is defined by the base