Machine Learning II Exercise 2

Matthias Tangemann, Cornelia Vogel and Paul Schmidt-Barbo

June 27, 2018

1 Bayesian Networks

E1

The path $(X_1, X_4, X_2, X_6, X_7, X_5)$ is not blocked: In nodes X_4 and X_7 the edges of the path meet head-to-head, but both nodes are in the set of observed variables C. In the remaining intermediate nodes X_2 and X_6, the edges meet tail-to-tail and head-to-tail, respectively. But neither of the nodes is in C.

E2

If $C_1 := \{x_2, x_3\}$ are given, both paths from x_1 to x_4 are blocked: In the intermediate nodes x_2 and x_3 of the paths (x_1, x_2, x_4) and (x_1, x_3, x_4), the edges meet head-to-tail and both nodes are in C_1. So $x_1 \perp x_4 | (x_2, x_3)$.

If $C_2 := \{x_1, x_4\}$ are given, the path (x_2, x_4, x_3) is not blocked, because in the only intermediate node x_4, the edges meet head-to-head, but $x_4 \in C_2$. So $x_2 \perp x_3 | (x_1, x_4)$ does not hold.

E3

Let $C_2 = \{x_1, x_4\}$ be given. Then both paths from x_2 to x_3, e.g. (x_2, x_1, x_3) and (x_2, x_4, x_3), are blocked: In both intermediate nodes x_1 and x_4 the edges meet tail-to-tail, and both nodes are in C_2. So C_1. So $x_1 \perp x_4 | (x_2, x_3)$.

If $C_1 = \{x_2, x_3\}$ are given, the path (x_1, x_2, x_4) is not blocked, because in the only intermediate node x_2, the edges meet head-to-head, but $x_2 \in C_1$. So $x_1 \perp x_4 | (x_2, x_3)$ does not hold.

2 Hidden Markov Models

E4

The given Bayesian network represent the following factorization of the combined probability:

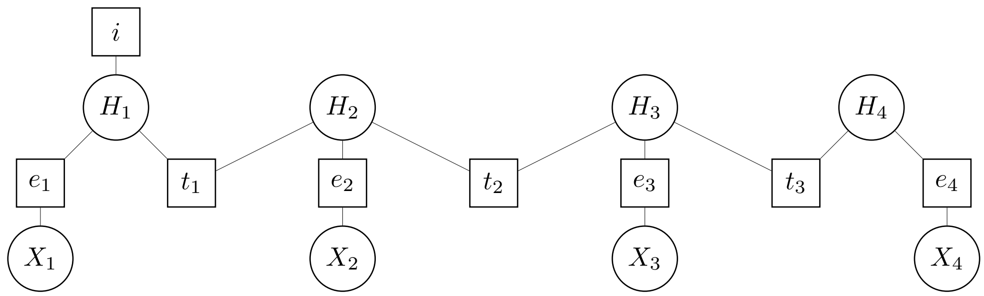

Figure 1: The factor graph corresponding to the hidden markov model

$$P(H_1, H_2, H_3, H_4, X_1, X_2, X_3, X_4) =$$
$$P(H_1)P(X_1|H_1)P(H_2|H_1)P(X_2|H_2)P(H_3|H_2)P(X_3|H_3)P(H_4|H_3)P(X_4|H_4)$$

So using the factors $i(H_1) := P(H_1)$, $e_i(X_i, H_i) := P(X_i|H_i)$ and $t_i(H_{i+1}, H_i) := P(H_{i+1}|H_i)$, the network can be transformed into the factor graph shown in figure 1.

The sum-product algorithm can be applied, because the graph is singly connected: Between each pair of nodes, there is exactly one path.

E5

The marginal distributions $P(X_i)$ can be computed using the sum-product algorithm. The variable node $P(X_1)$ is arbitrarily chosen as root. In the resulting tree, the factor node i and variable nodes X_2, X_3, X_4 are leaves. Starting from this leaves, messages are iteratively passed along the tree, until the root is reached:

$$
\begin{aligned}
\mu_{i\to H_1}(H_1) &= i(H_1) \\
\mu_{X_2\to e_2}(X_2) &= 1 \\
\mu_{e_2\to H_2}(H_2) &= \sum_{X_2} e_2(X_2, H_2) \\
\mu_{X_3\to e_3}(X_3) &= 1 \\
\mu_{e_3\to H_3}(H_3) &= \sum_{X_3} e_3(X_3, H_3) \\
\mu_{X_4\to e_4}(X_4) &= 1 \\
\mu_{e_4\to H_4}(X_4) &= \sum_{X_4} e_4(X_4, H_4) \\
\mu_{H_4\to t_3}(H_4) &= \mu_{e_4\to H_4}(X_4) \\
\mu_{t_3\to H_3}(H_3) &= \sum_{H_4} t_3(H_4, H_3)\mu_{H_4\to t_3}(H_4) \\
\mu_{H_3\to t_2}(H_3) &= \mu_{e_3\to H_3}(H_3)\mu_{t_3\to H_3}(H_3) \\
\mu_{t_2\to H_2}(H_2) &= \sum_{H_3} t_2(H_3, H_2)\mu_{H_3\to t_2}(H_3) \\
\mu_{H_2\to t_1}(H_2) &= \mu_{e_2\to H_2}(H_2)\mu_{t_2\to H_2}(H_2) \\
\mu_{t_1\to H_1}(H_1) &= \sum_{H_2} t_1(H_2, H_1)\mu_{H_2\to t_1}(H_2) \\
\mu_{H_1\to e_1}(H_1) &= i(H_1)\mu_{t_1\to H_1}(H_1) \\
\mu_{e_1\to X_1}(X_1) &= \sum_{H_1} e_1(X_1, H_1)\mu_{H_1\to e_1}(H_1)
\end{aligned}
$$

The remaining messages can now be calculated by iteratively traversing the tree from the root to the leaves:

$$
\begin{aligned}
\mu_{X_1\to e_1}(X_1) &= 1 \\
\mu_{e_1\to H_1}(H_1) &= \sum_{X_1} e_1(X_1, H_1) \\
\mu_{H_1\to i}(H_1) &= \mu_{e_1\to H_1}(H_1)\mu_{t_1\to H_1}(H_1) \\
\mu_{H_1\to t_1}(H_1) &= \mu_{i\to H_1}(H_1)\mu_{e_1\to H_1}(H_1) \\
\mu_{t_1\to H_2}(H_2) &= \sum_{H_1} t_1(H_2, H_1)\mu_{H_1\to t_1}(H_1) \\
\mu_{H_2\to e_2}(H_2) &= \mu_{t_1\to H_2}(H_2)\mu_{t_2\to H_2}(H_2) \\
\mu_{e_2\to X_2}(X_2) &= \sum_{H_2} e_2(X_2, H_2)\mu_{H_2\to e_2}(H_2) \\
\mu_{H_2\to t_2}(H_2) &= \mu_{t_1\to H_2}(H_2)\mu_{e_2\to H_2}(H_2) \\
\mu_{t_2\to H_3}(H_3) &= \sum_{H_2} t_2(H_3, H_2)\mu_{H_2\to t_2}(H_2) \\
\mu_{H_3\to e_3}(H_3) &= \mu_{t_2\to H_3}(H_3)\mu_{t_3\to H_3}(H_3) \\
\mu_{e_3\to X_3}(X_3) &= \sum_{H_3} e_3(X_3, H_3)\mu_{H_3\to e_3}(H_3) \\
\mu_{H_3\to t_3}(H_3) &= \mu_{t_2\to H_3}(H_3)\mu_{e_3\to H_3}(H_3) \\
\mu_{t_3\to H_4}(H_4) &= \sum_{H_3} t_3(H_4, H_3)\mu_{H_3\to t_3}(H_3) \\
\mu_{H_4\to e_4}(H_4) &= \mu_{t_3\to H_4}(H_4) \\
\mu_{e_4\to X_4}(X_4) &= \sum_{H_4} e_4(X_4, H_4)\mu_{H_4\to e_4}(H_4)
\end{aligned}
$$

Now one message has passed in each direction across each link, and only values available at the respective step were used for the computation. The marginals can now be evaluated as $P(X_i) = \mu_{e_i\to X_i}(X_i)$. As the initial condition and the transition probabilites are symmetric with respect to values of the state variables, and due to the symmetric emission probabilities it can be easiliy seen that the marginal distributions evalutate to $P(X_i = T) = P(X_i = F) = 0.5$ for $i = 1, 2, 3, 4$.

3 Solution to E6-E9 with HMM notation from Cornelia Vogel

Exercise 6

In the following we are going to use the notations from lecture 6.

We have

$$\begin{aligned}
\alpha(H_1) &= P(X_1 = T|H_1)P(H_1),\\
\alpha(H_1 = T) &= 0.8 \cdot 0.5 = 0.4,\\
\alpha(H_1 = F) &= 0.2 \cdot 0.5 = 0.1
\end{aligned}$$

Using the forward algorithm, we obtain

$$\begin{aligned}
\alpha(H_2) &= P(X_2 = F|H_2) \sum_{H_1} P(H_2|H_1)\, \alpha(H_1),\\
\alpha(H_2 = T) &= P(X_2 = F|\ H_2 = T)\, (P(H_2 = T|H_1 = T)\, \alpha(H_1 = T) + P(H_2 = T|H_1 = F)\, \alpha(H_1 = F))\\
&= 0.2 \cdot (0.7 \cdot 0.4 + 0.3 \cdot 0.1)\\
&= 0.062,\\
\alpha(H_2 = F) &= P(X_2 = F|\ H_2 = F)\, (P(H_2 = F|H_1 = T)\, \alpha(H_1 = T) + P(H_2 = F|H_1 = F)\, \alpha(H_1 = F))\\
&= 0.8 \cdot (0.3 \cdot 0.4 + 0.7 \cdot 0.1)\\
&= 0.152
\end{aligned}$$

$$\begin{aligned}
\alpha(H_3) &= P(X_3 = F|H_3) \sum_{H_2} P(H_3|H_2)\alpha(H_2),\\
\alpha(H_3 = T) &= P(X_3 = F|\ H_3 = T)\, (P(H_3 = T|H_2 = T)\, \alpha(H_2 = T) + P(H_3 = T|H_2 = F)\, \alpha(H_2 = F))\\
&= 0.2 \cdot (0.7 \cdot 0.062 + 0.3 \cdot 0.152)\\
&= 1.78 \cdot 10^{-2},\\
\alpha(H_3 = F) &= P(X_3 = F|\ H_3 = F)\, (P(H_3 = F|H_2 = T)\, \alpha(H_2 = T) + P(H_3 = F|H_2 = F)\, \alpha(H_2 = F))\\
&= 0.8 \cdot (0.3 \cdot 0.062 + 0.7 \cdot 0.152)\\
&= 0.1,
\end{aligned}$$

$$\begin{aligned}
\alpha(H_4) &= P(X_4 = T|H_4) \sum_{H_3} P(H_4|H_3)\alpha(H_3),\\
\alpha(H_4 = T) &= P(X_4 = T|H_4 = T)\, (P(H_4 = T|H_3 = T)\, \alpha(H_3 = T) + P(H_4 = T|H_3 = F)\, \alpha(H_3 = F))\\
&= 0.8 \cdot \big(0.7 \cdot 1.78 \cdot 10^{-2} + 0.3 \cdot 0.1\big)\\
&= 3.40 \cdot 10^{-2},\\
\alpha(H_4 = F) &= P(X_4 = T|H_4 = F)\, (P(H_4 = F|H_3 = T)\, \alpha(H_3 = T) + P(H_4 = F|H_3 = F)\, \alpha(H_3 = F))\\
&= 0.2 \cdot \big(0.3 \cdot 1.78 \cdot 10^{-2} + 0.7 \cdot 0.1\big)\\
&= 1.51 \cdot 10^{-2}.
\end{aligned}$$

We have

$$
\begin{aligned}
P(X_1 = T, X_2 = F, X_3 = F, X_4 = T) &= \sum_{H_4} P(X_1 = T, X_2 = F, X_3 = F, X_4 = T, H_4) \\
&= \sum_{H_4} \alpha(H_4) = \alpha(H_4 = T) + \alpha(H_4 = F) \\
&= 3.40 \cdot 10^{-2} + 1.51 \cdot 10^{-2} \\
&= 4.91 \cdot 10^{-2}.
\end{aligned}
$$

Finally, we obtain

$$
\begin{aligned}
P(H_4 = T | X_1 = T, X_2 = F, X_3 = F, X_4 = T) &= \frac{P(H_4 = T, X_1 = T, X_2 = F, X_3 = F, X_4 = T)}{P(X_1 = T, X_2 = F, X_3 = F, X_4 = T)} \\
&= \frac{3.40 \cdot 10^{-2}}{4.91 \cdot 10^{-2}} \\
&= 0.692, \\
P(H_4 = F | X_1 = T, X_2 = F, X_3 = F, X_4 = T) &= \frac{1.51 \cdot 10^{-2}}{4.91 \cdot 10^{-2}} \\
&= 0.308
\end{aligned}
$$

Exercise 7

Applying the backward algorithm, we obtain

$$
\begin{aligned}
\beta(H_4) &= 1,\\
\\
\beta(H_3) &= \sum_{H_4} P(X_4 = T|H_4)P(H_4|H_3)\beta(H_4),\\
\beta(H_3 = T) &= P(X_4 = T|\ H_4 = T)P(H_4 = T|\ H_3 = T)\beta(H_4 = T)\\
&\quad + P(X_4 = T|\ H_4 = F)P(H_4 = F|\ H_3 = T)\beta(H_4 = F)\\
&= 0.8 \cdot 0.7 + 0.2 \cdot 0.3\\
&= 0.62,\\
\beta(H_3 = F) &= P(X_4 = T|\ H_4 = T)P(H_4 = T|\ H_3 = F)\beta(H_4 = T)\\
&\quad + P(X_4 = T|\ H_4 = F)P(H_4 = F|\ H_3 = F)\beta(H_4 = F)\\
&= 0.8 \cdot 0.3 + 0.2 \cdot 0.7\\
&= 0.38,\\
\\
\beta(H_2) &= \sum_{H_3} P(X_3 = F|\ H_3)P(H_3|H_2)\beta(H_3),\\
\beta(H_2 = T) &= P(X_3 = F|H_3 = T)P(H_3 = T|\ H_2 = T)\beta(H_3 = T)\\
&\quad + P(X_3 = F|\ H_3 = F)P(H_3 = F|\ H_2 = T)\beta(H_3 = F)\\
&= 0.2 \cdot 0.7 \cdot 0.62 + 0.8 \cdot 0.3 \cdot 0.38\\
&= 0.178,\\
\beta(H_2 = F) &= P(X_3 = F|H_3 = T)P(H_3 = T|\ H_2 = F)\beta(H_3 = T)\\
&\quad + P(X_3 = F|\ H_3 = F)P(H_3 = F|\ H_2 = F)\beta(H_3 = F)\\
&= 0.2 \cdot 0.3 \cdot 0.62 + 0.8 \cdot 0.7 \cdot 0.38\\
&= 0.25,\\
\\
\beta(H_1) &= \sum_{H_2} P(X_2 = F|\ H_2)P(H_2|H_1)\beta(H_2),\\
\beta(H_1 = T) &= P(X_2 = F|H_2 = T)P(H_2 = T|H_1 = T)\beta(H_2 = T)\\
&\quad + P(X_2 = F|\ H_2 = F)P(H_2 = F|\ H_1 = T)\beta(H_2 = F)\\
&= 0.2 \cdot 0.7 \cdot 0.178 + 0.8 \cdot 0.3 \cdot 0.25\\
&= 8.49 \cdot 10^{-2},\\
\beta(H_1 = F) &= P(X_2 = F|H_2 = T)P(H_2 = T|H_1 = F)\beta(H_2 = T)\\
&\quad + P(X_2 = F|\ H_2 = F)P(H_2 = F|\ H_1 = F)\beta(H_2 = F)\\
&= 0.2 \cdot 0.3 \cdot 0.178 + 0.8 \cdot 0.7 \cdot 0.25\\
&= 0.151
\end{aligned}
$$

Now we obtain

$$
\begin{aligned}
P(H_1 = T \mid X_1 = T, X_2 = F, X_3 = F, X_4 = T) &= \frac{P(H_1 = T, X_1 = T, X_2 = F, X_3 = F, X_4 = T)}{P(X_1 = T, X_2 = F, X_3 = F, X_4 = T)} \\
&= \frac{\alpha(H_1 = T)\beta(H_1 = T)}{P(X_1 = T, X_2 = F, X_3 = F, X_4 = T)} \\
&= \frac{0.4 \cdot 8.49 \cdot 10^{-2}}{4.91 \cdot 10^{-2}} \\
&= 0.692, \\
P(H_1 = F \mid X_1 = T, X_2 = F, X_3 = F, X_4 = T) &= \frac{\alpha(H_1 = F)\beta(H_1 = F)}{P(X_1 = T, X_2 = F, X_3 = F, X_4 = T)} \\
&= \frac{0.1 \cdot 0.151}{4.91 \cdot 10^{-2}} \\
&= 0.308
\end{aligned}
$$

Exercise 8

We have

$$\mu(H_4) = 1,$$

$$\begin{aligned}
\mu(H_3) &= \max_{H_4} P(X_4 = T|H_4)P(H_4|\ H_3)\mu(H_4), \\
P(X_4 = T|H_4 = T)P(H_4 = T|H_3 = T) &= 0.8 \cdot 0.7 = 0.56, \\
P(X_4 = T|H_4 = F)P(H_4 = F|H_3 = T) &= 0.2 \cdot 0.3 = 0.06, \\
\Rightarrow \mu(H_3 = T) &= 0.56, \\
P(X_4 = T|\ H_4 = T)P(H_4 = T|\ H_3 = F) &= 0.8 \cdot 0.3 = 0.24, \\
P(X_4 = T|\ H_4 = F)P(H_4 = F|\ H_3 = F) &= 0.2 \cdot 0.7 = 0.14 \\
\Rightarrow \mu(H_3 = F) &= 0.24,
\end{aligned}$$

$$\begin{aligned}
\mu(H_2) &= \max_{H_3} P(X_3 = F|H_3)P(H_3|H_2)\mu(H_3), \\
P(X_3 = F|H_3 = T)P(H_3 = T|\ H_2 = T)\mu(H_3 = T) &= 0.2 \cdot 0.7 \cdot 0.56 = 7.84 \cdot 10^{-2}, \\
P(X_3 = F|H_3 = F)P(H_3 = F|\ H_2 = T)\mu(H_3 = F) &= 0.8 \cdot 0.3 \cdot 0.24 = 5.76 \cdot 10^{-2}, \\
\Rightarrow \mu(H_2 = T) &= 7.84 \cdot 10^{-2}, \\
P(X_3 = F|\ H_3 = T)P(H_3 = T|\ H_2 = F)\mu(H_3 = T) &= 0.2 \cdot 0.3 \cdot 0.56 = 3.36 \cdot 10^{-2}, \\
P(X_3 = F|H_3 = F)P(H_3 = F|\ H_2 = F)\mu(H_3 = F) &= 0.8 \cdot 0.7 \cdot 0.24 = 0.134, \\
\Rightarrow \mu(H_2 = F) &= 0.134,
\end{aligned}$$

$$\begin{aligned}
\mu(H_1) &= \max_{H_2} P(X_2 = F|H_2)P(H_2|\ H_1)\mu(H_2), \\
P(X_2 = F|\ H_2 = T)P(H_2 = T|\ H_1 = T)\mu(H_2 = T) &= 0.2 \cdot 0.7 \cdot 7.84 \cdot 10^{-2} = 1.10 \cdot 10^{-2}, \\
P(X_2 = F|\ H_2 = F)P(H_2 = F|\ H_1 = T)\mu(H_2 = F) &= 0.8 \cdot 0.3 \cdot 0.134 = 3.22 \cdot 10^{-2}, \\
\Rightarrow \mu(H_1 = T) &= 3.22 \cdot 10^{-2}, \\
P(X_2 = F|\ H_2 = T)P(H_2 = T|\ H_1 = F)\mu(H_2 = T) &= 0.2 \cdot 0.3 \cdot 7.84 \cdot 10^{-2} = 4.70 \cdot 10^{-3}, \\
P(X_2 = F|\ H_2 = F)P(H_2 = F|\ H_1 = F)\mu(H_2 = F) &= 0.8 \cdot 0.7 \cdot 0.134 = 7.50 \cdot 10^{-2}, \\
\Rightarrow \mu(H_1 = F) &= 7.50 \cdot 10^{-2}.
\end{aligned}$$

We compute

$$\begin{aligned}
P(X_1 = T|H_1 = T)P(H_1 = T)\mu(H_1 = T) &= 0.8 \cdot 0.5 \cdot 3.22 \cdot 10^{-2} = 1.29 \cdot 10^{-2}, \\
P(X_1 = T|H_1 = F)P(H_1 = F)\mu(H_1 = F) &= 0.2 \cdot 0.5 \cdot 7.50 \cdot 10^{-2} = 7.50 \cdot 10^{-3}.
\end{aligned}$$

Thus the most likely state for H_1 is given by

$$H_1^* = \arg\max_{H_1} P(X_1 = T|H_1)P(H_1)\mu(H_1) = T.$$

For the other states we obtain with the formula

$$H_t^* = \arg\max_{H_t} P(X_t|H_t)P(H_t|H_{t-1}^*)\mu(H_t)$$

the following optimal states

$$\begin{aligned}
P(X_2 = F|H_2 = T)P(H_2 = T|H_1^* = T)\mu(H_2 = T) &= 0.2 \cdot 0.7 \cdot 7.84 \cdot 10^{-2} = 1.10 \cdot 10^{-2}, \\
P(X_2 = F|H_2 = F)P(H_2 = F|H_1^* = T)\mu(H_2 = F) &= 0.8 \cdot 0.3 \cdot 0.134 = 3.22 \cdot 10^{-2} \\
&\Rightarrow H_2^* = F,
\end{aligned}$$

$$\begin{aligned}
P(X_3 = F|\ H_3 = T)P(H_3 = T|H_2^* = F)\mu(H_3 = T) &= 0.2 \cdot 0.3 \cdot 0.56 = 3.36 \cdot 10^{-2}, \\
P(X_3 = F|\ H_3 = F)P(H_3 = F|H_2^* = F)\mu(H_3 = F) &= 0.8 \cdot 0.7 \cdot 0.24 = 0.134, \\
&\Rightarrow H_3^* = F,
\end{aligned}$$

$$\begin{aligned}
P(X_4 = T|\ H_4 = T)P(H_4 = T|\ H_3^* = F)\mu(H_4 = T) &= 0.8 \cdot 0.3 = 0.24, \\
P(X_4 = T|\ H_4 = F)P(H_4 = F|\ H_3^* = F)\mu(H_4 = F) &= 0.2 \cdot 0.7 = 0.14, \\
&\Rightarrow H_4^* = T.
\end{aligned}$$

Hence the most probable hidden state sequence is T, F, F, T.

Exercise 9

We know from Exercise 6:

$$\begin{aligned}
\alpha(H_3 = T) &= P(H_3 = T, X_1 = T, X_2 = F, X_3 = F) = 1.78 \cdot 10^{-2}, \\
\alpha(H_3 = F) &= P(H_3 = F, X_1 = T, X_2 = F, X_3 = F) = 0.1, \\
P(X_1 = T, X_2 = F, X_3 = F, X_4 = T) &= 4.91 \cdot 10^{-2}.
\end{aligned}$$

It follows

$$\begin{aligned}
P(X_1 = T, X_2 = F, X_3 = F) &= \sum_{H_3} P(H_3, X_1 = T, X_2 = F, X_3 = F) \\
&= 1.78 \cdot 10^{-2} + 0.1 \\
&= 0.118
\end{aligned}$$

Hence we obtain

$$\begin{aligned}
P(X_4 = T|X_1 = T, X_2 = F, X_3 = F) &= \frac{P(X_4 = T, X_1 = T, X_2 = F, X_3 = F)}{P(X_1 = T, X_2 = F, X_3 = F)} \\
&= \frac{4.91 \cdot 10^{-2}}{0.118} \\
&= 0.416, \\
P(X_4 = F|X_1 = T, X_2 = F, X_3 = F) &= 1 - P(X_4 = T|X_1 = T, X_2 = F, X_3 = F) \\
&= 1 - 0.416 \\
&= 0.584
\end{aligned}$$

E10

The independence of variables can be obtained by applying d-Independence on the original Bayesian network:

If no variable is observed, e.g. $C_1 = \emptyset$, then X_1 and X_4 are not independent: In the Bayesian network, the path $(X_1, H_1, H_2, H_3, H_4, X_4)$ is not blocked, because in all intermediate nodes H_i the edges meet tail-to-tail oder head-to-tail, but none of them is in C_1.

If $C_2 = H_3$ is observed, the same path is blocked: The edges meet head-to-tail in H_3, and $H_3 \in C_2$. So in this case, X_1 and X_4 are independent.