Definition: Longer Markov Chains

We can extend the definition of Markov chains to more than three variables:

Definition: Markov chain (of length n)

The random variables $X_1, X_2, \dots, X_n$ form a Markov chain (notation: $X_1 \to X_2 \to \dots \to X_n$) if for all $3 \leq i \leq n$,

$$P_{X_i|X_1 \cdots X_{i-1}} = P_{X_i|X_{i-1}} .$$

Markov chains of length n exhibit similar properties to the properties we have seen for Markov chains of length 3. In particular, the reverse chain is also a Markov chain ($X_n \to \dots \to X_2 \to X_1$), and a more general form of the data-processing inequality holds in the sense that the further apart two variables are in the chain, the less correlated they are.

Proposition

If $X_1 \to X_2 \to X_3 \to X_4$ is a Markov chain, the following are Markov chains as well:

a. $X_1 \to X_2 \to X_3$
b. $X_2 \to X_3 \to X_4$
c. $X_1 \to X_2X_3 \to X_4$
d. $X_4 \to X_3 \to X_2 \to X_1$

Proof

left as exercise

In the exercises, we will prove that if $X_1, X_2, \dots, X_n$ forms a Markov chain, then for all $1 \leq i \leq j \leq k \leq n$: $I(X_i, X_j) \geq I(X_i, X_k)$. This is a generalized form of the **data-processing inequality**.