ECEN 655: Advanced Channel Coding Spring 2014

Lecture 4 — 1/23/2014

Prof. Henry Pfister *Scribe: Sriharsha Madala*

1 Overview

In the last lecture, we covered marginalization and how functions that factorize allow one to minimize the complexity of marginalization. In particular, we discussed:

- how to represent such a factorization as a factor graph and
- how passing messages on that factor graph can leads to efficient marginalization.

In this lecture, we will:

- connect factor graphs with probability,
- discuss the relationship between Bayesian Networks (BNs) and FGs,
- conditional independence in FGs
- log-likelihood-ratio (LLR) messages for binary random variables, and
- discuss the use of FGs in coding.

2 Connection to Probability

Any non-negative function $f(x_1, \ldots x_n)$ can be normalized into the PMF for the RVs $X_1, \ldots X_n$ with

$$\Pr(X_1 = x_1, \ldots X_n = x_n) = \frac{1}{Z} f(x_1, \ldots x_n),$$

where $Z = \sum_{x_1^n} f(x_1, \ldots x_n)$ is called the *partition function.*

3 Bayesian Networks

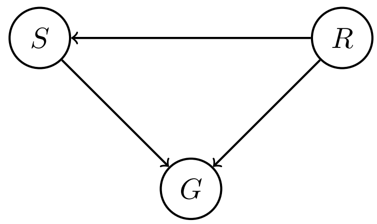
Bayesian networks (BNs) and factor graphs are two ways of representing a joint PMF. For any BN, there is a tree FG with similar graphical structure that represents the same distribution. For any tree FG, one can also define a BN with similar graphical structure. But, for FGs with cycles there is typically no equivalent BN. Thus, FGs are more general.

Let the RVs $X_1, \ldots X_n$ be the vertices of a directed acyclic graph.

$$\Pr(X_1 = x_1, \ldots X_n = x_n) = \prod_{i=1}^{n} P\left(X_i = x_i | X_{\pi(i)} = x_{\pi(i)}\right)$$

where $\pi(i)$ is the set of nodes with a directed edge to i (i.e., the parents of node X_i).

3.1 Rain-Sprinkler-Grass Example



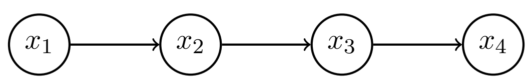
A common example used to motivate BNs is the question, "What is the probability it is raining given the grass is wet?". Since grass may be wet due to either rain or a sprinkler, there are three random variables R, S, G of interest. The main assumption is that people typically don't water their grass when it is raining. Thus, we get the BN

$$P(G, S, R) = P(R) \cdot P(S|R) \cdot P(G|R, S).$$

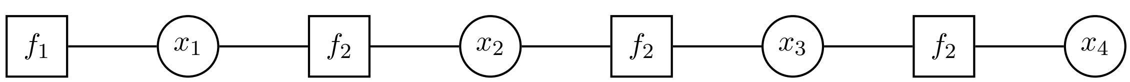
Of course, one problem with this example is that the factorization is generic and works for any three random variables.

3.2 Markov Chain

Let us consider a homogenous Markov chain. We can represent it using both a BN with



and as a factor graph with



where $f(x_1, \ldots x_4) = f_1(x_1) f_2(x_1, x_2) \cdot f_2(x_2, x_3) \cdot f_2(x_3, x_4)$ and $f_2(x, x') = \Pr(X_{i+1} = x' | X_i = x)$ and $f_1(x_1) = \Pr(X_1 = x_1)$. Using these definitions, the FG is essentially the same as the BN. It is also possible to use factor nodes where $\sum_{x'} f_2(x, x')$ is not a constant with respect to x. In this case, the resulting factorization is not easily converted into a BN.

4 Conditional Independence

Let A, B, S be disjoint subsets of VNs. We say that S *separates* A and B if any path from $u \in A$ to $v \in B$ must pass through $w \in S$.

Lemma: If S separates A and B, then

$$\Pr\left(X_A = x_A, X_B = x_B | X_S = x_S\right) = \Pr\left(X_A = x_A | X_S = x_S\right) \cdot \Pr\left(X_B = x_B | X_S = x_S\right)$$

Proof: Fixing $X_S = x_S$ allows one to remove all VNs in S and update attached factors. Since S separates A and B, the new factor graph must break into multiple two disjoint components with A and B if different components. Revisiting the generalized distributive law, we see that $\Pr\left(X_A = x_A, X_B = x_B | X_S = x_S\right)$ decomposes into the product of two terms where one term depends only on x_A and the other term depends only on x_B.

5 Binary Random Variables

Looking at the update equations for the message-passing algorithm, we see that scaling an input message by a constant results only in scaling the output message by the same constant. Therefore, if we are only interested in normalized marginals, then we can safely normalize messages at each stage and we define

$$\overline{\mu}_{i\to a}^{(t)}(x_i) = \frac{\mu_{i\to a}^{(t)}(x_i)}{\sum_{x'} \mu_{i\to a}^{(t)}(x')}$$

For binary random variables, normalized messages can be represented by scalars because $\mu(1) = 1 - \mu(0)$. It is sometimes convenient to use other representations:

- Likelihood-Ratio: $\frac{\mu(0)}{\mu(1)}$
- Log Likelihood-Ratio: $\ln \frac{\mu(0)}{\mu(1)}$
- Difference: $\mu(0) - \mu(1)$

Now, let us consider the VN update equation

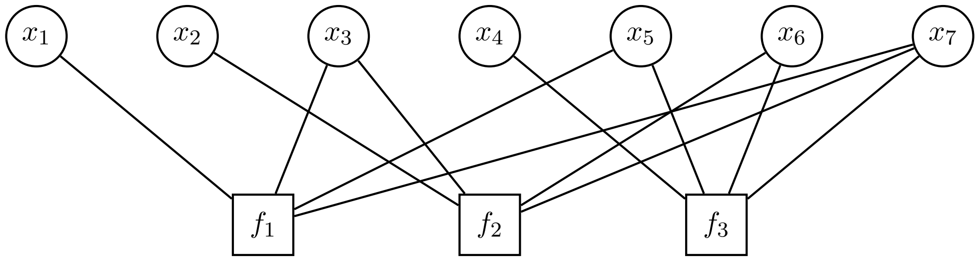
$$\mu_{i\to a}^{(t)}(x_i) = \prod_{b\in\partial a\setminus i} \widehat{\mu}_{b\to i}^{(t-1)}(x_i)$$

For binary random variables with LLR messages:

$$\begin{aligned}
L_{i\to a}^{(t)} &= \ln \frac{\mu_{i\to a}^{(t)}(0)}{\mu_{i\to a}^{(t)}(1)} \\
&= \ln \frac{\prod_{b\in\partial a\setminus i} \widehat{\mu}_{b\to i}^{(t-1)}(0)}{\prod_{b\in\partial a\setminus i} \widehat{\mu}_{b\to i}^{(t-1)}(1)} \\
&= \ln \prod_{b\in\partial a\setminus i} \frac{\widehat{\mu}_{b\to i}^{(t-1)}(0)}{\widehat{\mu}_{b\to i}^{(t-1)}(1)} \\
&= \sum_{b\in\partial a\setminus i} \widehat{L}_{b\to i}^{(t-1)}
\end{aligned}$$

Hence, the VN update becomes a sum in the LLR domain.

6 Connection to LDPC Codes



Given the $r \times n$ parity-check matrix H of a binary linear code $\mathcal{C}$, we define a variable node for each column and a factor (i.e., check) node for each row. The parity constraint enforced by a check node gives rise to the FN function

$$f_{even}(x_1^d) = I(x_1 \oplus \ldots \oplus x_d = 0).$$

The factor associated with row a is connected to the variable node associated with column i iff $H_{a,i} = 1$. The resulting factor graph is typically called the *Tanner graph* of the code.

Since the binary vector x_1^n is a codeword iff it satisfies all parity checks, we see that the indicator function of the code factors into the form

$$1_{\mathcal{C}}(x_1^n) = \prod_{a=1}^{r} f_{even}(x_{C(a)}),$$

where $C(a) = \{i \in [n] \,|\, H_{a,i} = 1\}$. For decoding, each VN also has a local factor determined by the channel observation.