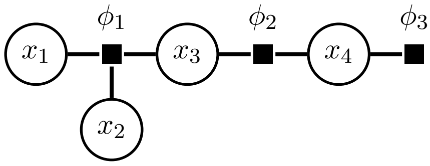
Probabilistic Modelling and Reasoning Notes (Inference)

Autumn 2025
Hodari & Gutmann

These notes summarise selected lecture concepts and are not a substitute for working through the lecture slides, tutorials, and self-study exercises. Feel free to personalise and develop them into your own summary sheet.

Factor graph — A factor graph represents an arbitrary function in terms of factors and their connections with variables. For example, a factor graph can represent a distribution written as a Gibbs distribution – $p(\mathbf{x}) = \frac{1}{Z}\prod_c \phi_c(\mathcal{X}_c)$ – where variables $x_i \in \mathbf{x}$ are represented with variable nodes (circles) and potentials ϕ_c are represented with factor nodes (squares). Edges connect each factor node ϕ_c to all its variable nodes $x_i \in \mathcal{X}_c$.

$$p(x_1, x_2, x_3, x_4) = \frac{1}{Z}\phi_1(x_1, x_2, x_3)\phi_2(x_3, x_4)\phi_3(x_4)$$



Variable elimination — Given $p(\mathcal{X}) \propto \prod_c \phi_c(\mathcal{X}_c)$, we compute the marginal $p(\mathcal{X} \setminus x^*)$ via the sum rule by exploiting the factorisation by means of the distributive law.

We sum out the variable x^* by first finding all factors $\phi_i(\mathcal{X}_i)$ such that $x^* \in \mathcal{X}_i$, and forming the compound factor $\phi^*(\mathcal{X}^*) = \prod_{i:x^* \in \mathcal{X}_i} \phi_i(\mathcal{X}_i)$, with $\mathcal{X}^* = \bigcup_{i:x^* \in \mathcal{X}_i} \mathcal{X}_i$. Summing out x^* then produces a new factor $\tilde{\phi}^*(\tilde{\mathcal{X}}^*) = \sum_{x^*} \phi^*(\mathcal{X}^*)$ that does not depend on x^*, i.e. $\tilde{\mathcal{X}}^* = \mathcal{X}^* \setminus x^*$. This is possible as products are commutative, and a sum can be distributed within a product as long as all terms depending on the variable(s) being summed come to the right of the sum.

$$p(\mathcal{X} \setminus x^*) \propto \sum_{x^*} \prod_c \phi_c(\mathcal{X}_c) \propto \left[\prod_{i:x^* \notin \mathcal{X}_i} \phi_i(\mathcal{X}_i)\right] \left[\sum_{x^*} \prod_{i:x^* \in \mathcal{X}_i} \phi_i(\mathcal{X}_i)\right] \tag{1}$$

$$\propto \left[\prod_{i:x^* \notin \mathcal{X}_i} \phi_i(\mathcal{X}_i)\right] \tilde{\phi}^*(\tilde{\mathcal{X}}^*) \tag{2}$$

When eliminating variables, order of elimination matters. However, optimal choice of elimination order is difficult. Picking variables greedily is a common heuristic, where the "best" x^* is the one that fewest factors ϕ_c depend upon.

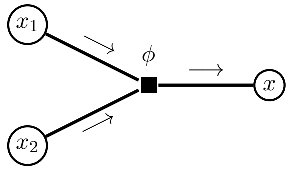
Sum-product algorithm — Variable elimination for factor trees reformulated with "messages" which allows for re-use of computations already done. See table on following page.

Max-sum algorithm — Message-passing algorithm to compute the most likely state and its probability. Obtained from sum-product by replacing $\sum$ with max, $\prod$ with $\sum$, and factors with log-factors. See table on following page.

Sum-product algorithm

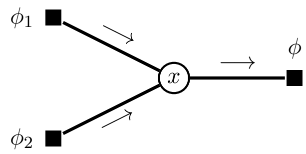
$\mu_{\phi \to x}(x)$

Factor to variable
$\mu_{\phi \to x}(x) = \sum_{x_1,\ldots,x_j} \phi(x_1, \ldots, x_j, x) \prod_{i=1}^{j} \mu_{x_i \to \phi}(x_i)$
where $\{x_1, \ldots, x_j\} = \mathrm{ne}(\phi) \setminus \{x\}$



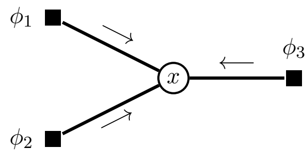
$\mu_{x \to \phi}(x)$

Variable to factor
$\mu_{x \to \phi}(x) = \prod_{i=1}^{j} \mu_{\phi_i \to x}(x)$
where $\{\phi_1, \ldots, \phi_j\} = \mathrm{ne}(x) \setminus \{\phi\}$



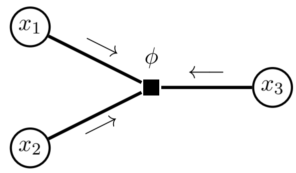
$\tilde{p}(x)$

Univariate marginals – unnormalised
$p(x) \propto \prod_{i=1}^{j} \mu_{\phi_i \to x}(x)$
where $\{\phi_1, \ldots, \phi_j\} = \mathrm{ne}(x)$



$\tilde{p}(x_1, \ldots, x_j)$

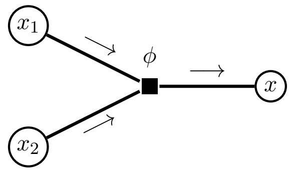
Joint marginals of variables sharing a factor– unnormalised
$p(x_1, \ldots, x_j) \propto \phi(x_1, \ldots, x_j) \prod_{i=1}^{j} \mu_{x_i \to \phi}(x_i)$
where $\{x_1, \ldots, x_j\} = \mathrm{ne}(\phi)$



Max-sum algorithm

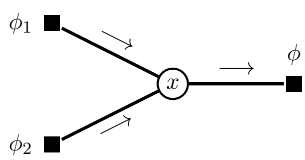
$\gamma_{\phi \to x}(x)$

Factor to variable
$\gamma_{\phi \to x}(x) = \max_{x_1,\ldots,x_j} \log \phi(x_1, \ldots, x_j, x) + \sum_{i=1}^{j} \gamma_{x_i \to \phi}(x_i)$
$\gamma^*_{\phi \to x}(x) = \mathrm{argmax}_{x_1,\ldots,x_j} \log \phi(x_1, \ldots, x_j, x) + \sum_{i=1}^{j} \gamma_{x_i \to \phi}(x_i)$
where $\{x_1, \ldots, x_j\} = \mathrm{ne}(\phi) \setminus \{x\}$



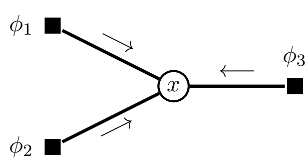
$\gamma_{x \to \phi}(x)$

Variable to factor
$\gamma_{x \to \phi}(x) = \sum_{i=1}^{j} \gamma_{\phi_i \to x}(x)$
where $\{\phi_1, \ldots, \phi_j\} = \mathrm{ne}(x) \setminus \{\phi\}$



$\log p_{\max}$

Maximum probability
$\log p_{\max} = \max_x \gamma^*(x), \quad \gamma^*(x) = -\log Z + \sum_{i=1}^{j} \gamma_{\phi_i \to x}(x)$
where $\{\phi_1, \ldots, \phi_j\} = \mathrm{ne}(x)$



$\mathrm{argmax}_{\mathbf{x}}\, \tilde{p}(\mathbf{x})$

Maximum probability states – no need for normalisation
Init: $\hat{x} = \mathrm{argmax}_x \gamma^*(x) = \mathrm{argmax}_x \sum_{i=1}^{j} \gamma_{\phi_i \to x}(x)$
Backtrack to leaves via $\gamma^*_{\phi \to x}(x)$