Chow et Liu dans le cas gaussien

December 19, 2017

Abstract

Approximating Gaussian distribution with dependence Trees. The likelihood of a sample of n random Gaussian variables is calculated assuming tree dependencies between variables. We consider pairs of variables in a symetric formula for the likelihood, underlying the undirectedness of trees we use. We find the Maximum Likelihood tree-structure and parameters following Chow and Liu (1968) idea for discrete variables.

1 Notations and definitions

Let $\boldsymbol{x}$ be a vector of n random continuous variables, $\boldsymbol{x} = (x_1, ..., x_n) \in \mathcal{R}^n$. Let $\mathcal{S}$ be a collection of samples of $\boldsymbol{x}$: $\mathcal{S} = \{\boldsymbol{x}^1, ..., \boldsymbol{x}^s\}$.

We consider a tree structure, each nod of the tree being a variable. All nodes i except for one have an ancestor nod, indexed a_i.The structure of a tree is recorded in the vector $\boldsymbol{\alpha} = (a_1, ..., a_n)$.

We assume Gaussian densities :

$$f(x_i|\mu_i, \sigma_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(x_i-\mu_i)^2}{2\sigma_i^2}\right)$$

$$f(x_i, x_{a_i}|\mu_i, \mu_{a_i}, \Sigma_{ia_i}) = \frac{1}{2\pi\sqrt{\det(\Sigma_{ia_i})}} \exp\left(-\frac{1}{2}(x_i-\mu_i, x_{a_i}-\mu_{a_i})^T \Sigma_{ia_i}^{-1} (x_i-\mu_i, x_{a_i}-\mu_{a_i})\right)$$

Where $\Sigma_{ia_i} = \begin{pmatrix} \sigma_i^2 & \sigma_{ia_i} \\ \sigma_{a_i i} & \sigma_{a_i}^2 \end{pmatrix}$.

The gaussian dependence-tree is parametrised by $\boldsymbol{\alpha}$:

$$\begin{aligned}
f(\boldsymbol{x}|\boldsymbol{\alpha}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) &= f(x_1|\mu_1, \sigma_1) \prod_{i=2}^{n} f(x_i|x_{a_i}, \mu_i, \mu_{a_i}, \Sigma_{ia_i}) \\
&= f(x_1|\mu_1, \sigma_1) \prod_{i=2}^{n} \frac{f(x_i, x_{a_i}|\mu_i, \mu_{a_i}, \Sigma_{ia_i})}{f(x_{a_i}|\mu_{a_i}, \sigma_{a_i})} \\
&= \prod_{i=1}^{n} f(x_i|\mu_i, \sigma_i) \prod_{i=2}^{n} \frac{f(x_i, x_{a_i}|\mu_i, \mu_{a_i}, \Sigma_{ia_i})}{f(x_{a_i}|\mu_{a_i}, \sigma_{a_i}) \times f(x_i|\mu_i, \sigma_i)}
\end{aligned}$$

2 Likelihood maximisation

2.1 Likelihood

We want to maximise the likelihood function, which writes

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \frac{1}{|\mathcal{S}|} \sum_{\boldsymbol{x} \in \mathcal{S}} \log(f(\boldsymbol{x}|\boldsymbol{\alpha}, \boldsymbol{\mu}, \boldsymbol{\Sigma})) \\
&= \frac{1}{|\mathcal{S}|} \sum_{\boldsymbol{x} \in \mathcal{S}} \left[\sum_{i=1}^{n} \log(f(x_i|\mu_i, \sigma_i)) + \sum_{i=2}^{n} \log\left(\frac{f(x_i, x_{a_i}|\mu_i, \mu_{a_i}, \Sigma_{ia_i})}{f(x_{a_i}|\mu_{a_i}, \sigma_{a_i}) \times f(x_i|\mu_i, \sigma_i)}\right)\right]
\end{aligned}$$

$$
\begin{aligned}
\mathcal{L}(\boldsymbol{\alpha},\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{|\mathcal{S}|}\sum_{\boldsymbol{x}\in\mathcal{S}} \Bigg[\sum_{i=1}^{n} \left(-\frac{1}{2}\log(2\pi\sigma_i^2) - \frac{x_i^2}{2\sigma_i^2} \right) + \sum_{i=2}^{n} \Bigg[\left(-\frac{1}{2}\log((2\pi)^2 \det(\Sigma_{ia_i})) - \frac{1}{2}(x_i, x_{a_i})^T \Sigma_{ia_i}^{-1}(x_i, x_{a_i}) \right) \\
- \left(-\frac{1}{2}\log(2\pi\sigma_i^2) - \frac{1}{2}\log(2\pi\sigma_{a_i}^2) - \frac{x_i^2}{2\sigma_i^2} - \frac{x_{a_i}^2}{2\sigma_{a_i}^2} \right) \Bigg] \Bigg] \\
= \frac{1}{|\mathcal{S}|}\sum_{\boldsymbol{x}\in\mathcal{S}} \Bigg[\sum_{i=1}^{n} -\frac{1}{2}\left(\log(2\pi\sigma_i^2) + \frac{x_i^2}{\sigma_i^2} \right) + \sum_{i=2}^{n} -\frac{1}{2}\Bigg(\log\left(1 - \frac{\sigma_{ia_i}^2}{\sigma_i^2\sigma_{a_i}^2}\right) + \frac{(x_i^2\sigma_{a_i}^2 + x_{a_i}^2\sigma_i^2 - 2\sigma_{ia_i}x_i x_{a_i})}{\det(\Sigma_{ia_i})} \\
- \left(\frac{x_i^2}{\sigma_i^2} + \frac{x_{a_i}^2}{\sigma_{a_i}^2} \right) \Bigg) \Bigg]
\end{aligned}
$$

2.2 Estimation

For any structure α of the dependance tree, the log-likelihood $\mathcal{L}(\boldsymbol{\alpha},\boldsymbol{\mu},\boldsymbol{\Sigma})$ is maximised by the MLE of the two-dimensional marginals. The maximum-likelihood estimators for the parameters μ_i, σ_i and σ_{ia_i} for Gaussian distributions are known :

$$
\begin{aligned}
\hat{\mu}_i &= \frac{1}{|\mathcal{S}|}\sum_{\boldsymbol{x}\in\mathcal{S}} x_i, \\
\hat{\sigma}_i^2 &= \frac{1}{|\mathcal{S}|}\sum_{\boldsymbol{x}\in\mathcal{S}} (x_i - \hat{\mu}_i)^2, \\
\hat{\sigma}_{ia_i}^2 &= \frac{1}{|\mathcal{S}|}\sum_{\boldsymbol{x}\in\mathcal{S}} (x_i - \hat{\mu}_i)(x_{a_i} - \hat{\mu}_{a_i}).
\end{aligned}
$$

Replacing these estimators, we finally get

$$
\mathcal{L}(\boldsymbol{\alpha},\hat{\boldsymbol{\mu}},\hat{\boldsymbol{\Sigma}}) = \sum_{i=1}^{n} -\frac{1}{2}\left(1 + \log(2\pi\hat{\sigma}_i^2)\right) + \sum_{i=2}^{n} \underbrace{-\frac{1}{2}\log\left(1 - \frac{\hat{\sigma}_{ia_i}^2}{\hat{\sigma}_i^2\hat{\sigma}_{ia_i}^2}\right)}_{\mathcal{I}\left(f(\cdot|\hat{\mu}_i,\hat{\sigma}_i),\, f(\cdot|\hat{\mu}_{a_i},\hat{\sigma}_{a_i})\right)}
$$

The first term of this quantity is independant of α, meaning it is independant of the structure of the dependance-tree. We then only need to maximise the second term to obtain an MLE estimator for the tree structure.

The last term is known as the Shannon information between two variables i and a_i. As the Shannon information is increasing with the correlation between these variables, it is a weight on each branch of the tree. The structure is then optimized by the maximum-weight spanning tree :

$$
\hat{\boldsymbol{\alpha}} = \arg\max_{\alpha} \left\{ \sum_{i=2}^{n} \mathcal{I}\left(f(\cdot|\hat{\mu}_i,\hat{\sigma}_i),\, f(\cdot|\hat{\mu}_{a_i},\hat{\sigma}_{a_i})\right) \right\}
$$