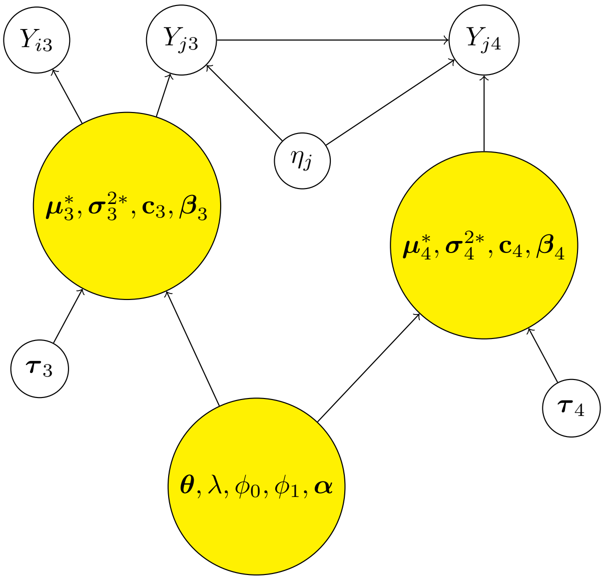
Missing Data Full Conditional

In the following, we prove conditional independence using Bayesian networks. For more information, see "Pattern Recognition and Machine Learning", Chapter 8.1, by Bishop. Let us suppose Y_{i3} is the missing datum. For the Gibbs sampler, we have to compute the full conditional of Y_{i3}, i.e.,

$$\mathcal{L}(Y_{i3} \mid \mathbf{Y}\backslash\{Y_{i3}\}, \boldsymbol{\mu}^*, \boldsymbol{\sigma}^{2*}, \mathbf{c}, \boldsymbol{\tau}, \boldsymbol{\eta}, \boldsymbol{\beta}, \boldsymbol{\theta}, \phi_0, \phi_1, \lambda, \boldsymbol{\alpha})$$

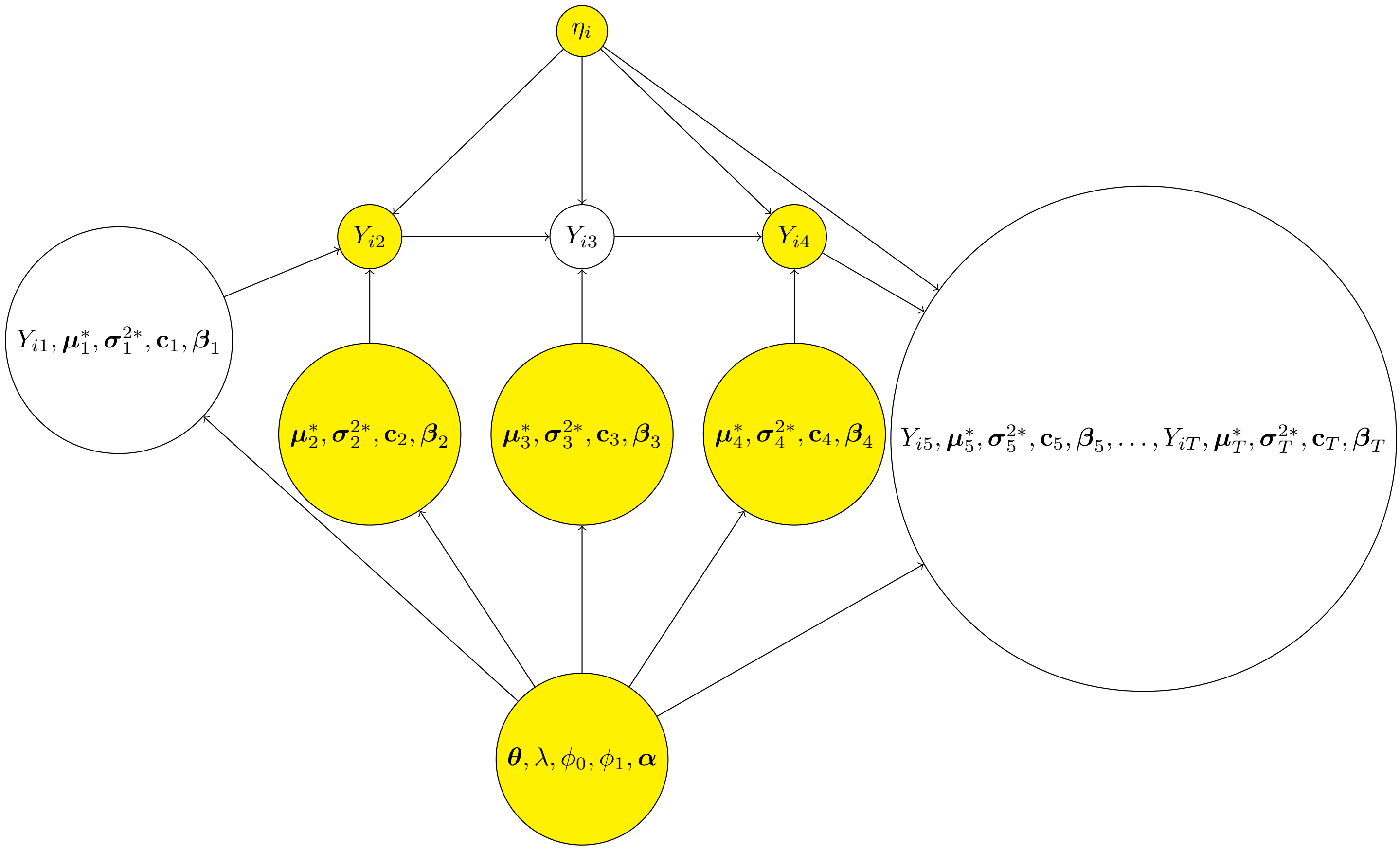
We represent part of the model's Bayesian network:



We can see that, conditioning on $(\boldsymbol{\mu}_3^*, \boldsymbol{\sigma}_3^{2*}, \mathbf{c}_3, \boldsymbol{\beta}_3, \boldsymbol{\mu}_4^*, \boldsymbol{\sigma}_4^{2*}, \mathbf{c}_4, \boldsymbol{\beta}_4, \boldsymbol{\theta}, \lambda, \phi_0, \phi_1, \boldsymbol{\alpha})$ (i.e., the colored nodes), Y_{i3} is independent of τ_3, τ_4, η_j, Y_{j3} and Y_{j4}, for $j \neq i$.
Extending the graph, we can prove in the same way that, conditioning on $(\boldsymbol{\mu}^*, \boldsymbol{\sigma}^{2*}, \mathbf{c}, \boldsymbol{\beta}, \boldsymbol{\theta}, \phi_0, \phi_1, \lambda, \boldsymbol{\alpha})$ Y_{i3} is independent of $\boldsymbol{\tau}, \boldsymbol{\eta}_j, \mathbf{Y}_j$ for $j \neq i$.
Thus, we can remove these variables from the full conditional, i.e.

$$\mathcal{L}(Y_{i3} \mid \mathbf{Y}\backslash\{Y_{i3}\}, \boldsymbol{\mu}^*, \boldsymbol{\sigma}^{2*}, \mathbf{c}, \boldsymbol{\tau}, \boldsymbol{\eta}, \boldsymbol{\beta}, \boldsymbol{\theta}, \phi_0, \phi_1, \lambda, \boldsymbol{\alpha}) = \mathcal{L}(Y_{i3} \mid \mathbf{Y}_i\backslash\{Y_{i3}\}, \boldsymbol{\mu}^*, \boldsymbol{\sigma}^{2*}, \mathbf{c}, \eta_i, \boldsymbol{\beta}, \boldsymbol{\theta}, \phi_0, \phi_1, \lambda, \boldsymbol{\alpha})$$

We represent part of the remaining Bayesian network:



We can see that, conditioning on $(Y_{i2}, \boldsymbol{\mu}_2^*, \boldsymbol{\sigma}_2^{2*}, \mathbf{c}_2, \boldsymbol{\beta}_2, \boldsymbol{\mu}_3^*, \boldsymbol{\sigma}_3^{2*}, \mathbf{c}_3, \boldsymbol{\beta}_3, Y_{i4}, \boldsymbol{\mu}_4^*, \boldsymbol{\sigma}_4^{2*}, \mathbf{c}_4, \boldsymbol{\beta}_4, \eta_i, \boldsymbol{\theta}, \lambda, \phi_0, \phi_1, \boldsymbol{\alpha})$ (i.e., the colored nodes), Y_{i3} is independent of $Y_{it}, \boldsymbol{\mu}_t^*, \boldsymbol{\sigma}_t^{2*}, \mathbf{c}_t, \boldsymbol{\beta}_t$, for $t = 1, 5, 6, \ldots, T$.
Thus, we can remove these variables from the full conditional, i.e.

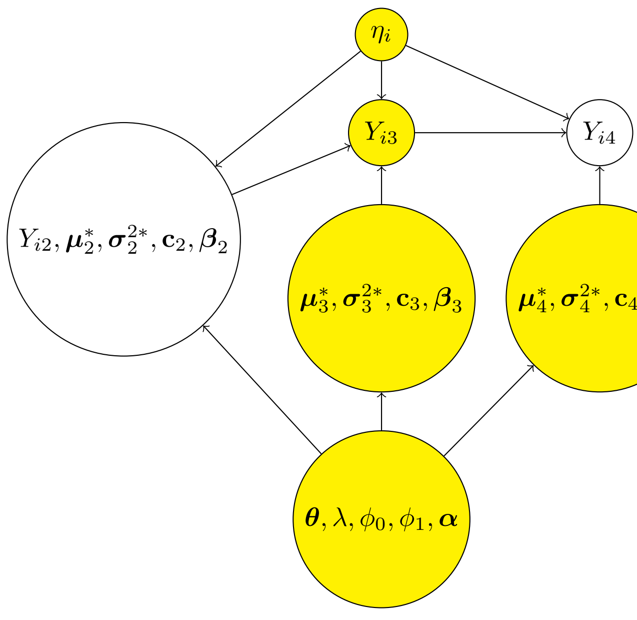
$$\mathcal{L}(Y_{i3} \mid \mathbf{Y}_i \backslash \{Y_{i3}\}, \boldsymbol{\mu}^*, \boldsymbol{\sigma}^{2*}, \mathbf{c}, \eta_i, \boldsymbol{\beta}, \boldsymbol{\theta}, \phi_0, \phi_1, \lambda, \boldsymbol{\alpha}) = \mathcal{L}(Y_{i3} \mid Y_{i2}, Y_{i4}, \boldsymbol{\mu}_2^*, \boldsymbol{\mu}_4^*, \boldsymbol{\sigma}_2^{2*}, \boldsymbol{\sigma}_4^{2*}, \mathbf{c}_2, \mathbf{c}_4, \eta_i, \boldsymbol{\beta}_2, \boldsymbol{\beta}_4, \boldsymbol{\theta}, \phi_0, \phi_1, \lambda, \boldsymbol{\alpha})$$

We use Bayes' rules:

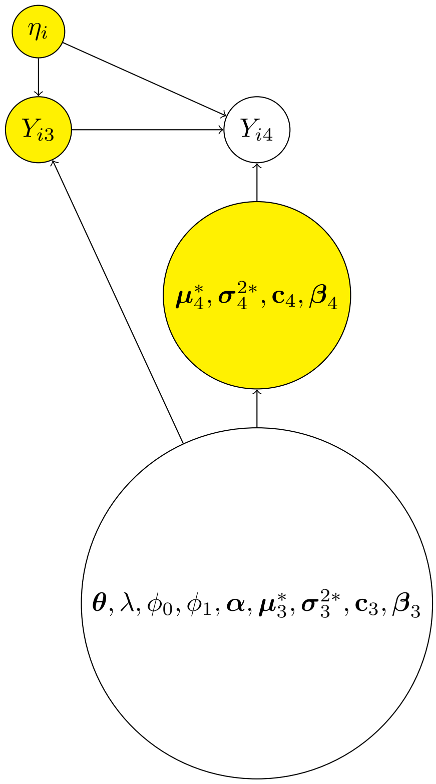
$$\begin{aligned}
&\propto \mathcal{L}(Y_{i4} \mid Y_{i2}, Y_{i3}, \boldsymbol{\mu}_2^*, \boldsymbol{\mu}_4^*, \boldsymbol{\sigma}_2^{2*}, \boldsymbol{\sigma}_4^{2*}, \mathbf{c}_2, \mathbf{c}_4, \eta_i, \boldsymbol{\beta}_2, \boldsymbol{\beta}_4, \boldsymbol{\theta}, \phi_0, \phi_1, \lambda, \boldsymbol{\alpha}) \\
&\quad \times \mathcal{L}(Y_{i3} \mid Y_{i2}, \boldsymbol{\mu}_2^*, \boldsymbol{\mu}_4^*, \boldsymbol{\sigma}_2^{2*}, \boldsymbol{\sigma}_4^{2*}, \mathbf{c}_2, \mathbf{c}_4, \eta_i, \boldsymbol{\beta}_2, \boldsymbol{\beta}_4, \boldsymbol{\theta}, \phi_0, \phi_1, \lambda, \boldsymbol{\alpha})
\end{aligned}$$

0.1 First term

Let us take the first term. We represent the remaining Bayesian network:



We can see that, conditioning on $(Y_{i3}, \boldsymbol{\mu}_3^*, \boldsymbol{\sigma}_3^{2*}, \mathbf{c}_3, \boldsymbol{\beta}_3, \boldsymbol{\mu}_4^*, \boldsymbol{\sigma}_4^{2*}, \mathbf{c}_4, \boldsymbol{\beta}_4, \eta_i, \boldsymbol{\theta}, \lambda, \phi_0, \phi_1, \boldsymbol{\alpha})$ (i.e., the colored nodes), Y_{i4} is independent of $Y_{i2}, \boldsymbol{\mu}_2^*, \boldsymbol{\sigma}_2^{2*}, \mathbf{c}_2, \boldsymbol{\beta}_2$.
We represent the new Bayesian network:



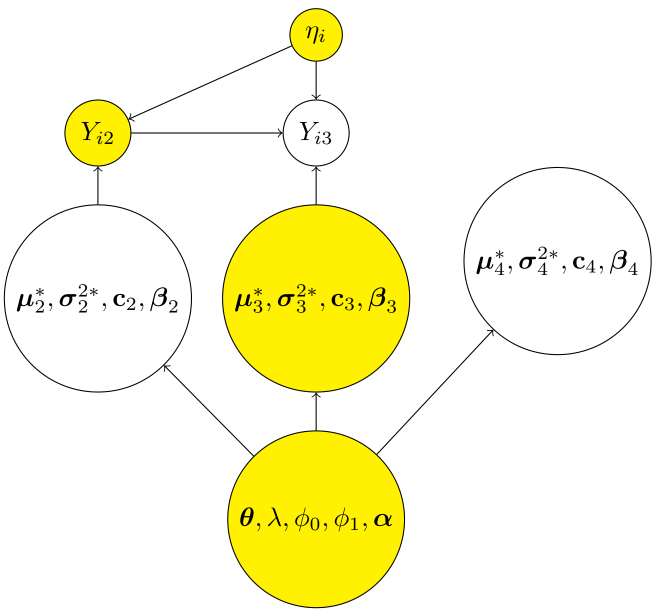
We can see that, conditioning on $(Y_{i3}, \boldsymbol{\mu}_4^*, \boldsymbol{\sigma}_4^{2*}, \mathbf{c}_4, \boldsymbol{\beta}_4, \eta_i)$ (i.e., the colored nodes), Y_{i4} is independent of $\boldsymbol{\mu}_3^*, \boldsymbol{\sigma}_3^{2*}, \mathbf{c}_3, \boldsymbol{\beta}_3, \boldsymbol{\theta}, \lambda, \phi_0, \phi_1, \boldsymbol{\alpha}$.
Thus, the first term becomes:

$$\mathcal{L}(Y_{i4} \mid Y_{i3}, \boldsymbol{\mu}_4^*, \boldsymbol{\sigma}_4^{2*}, \mathbf{c}_4, \eta_i, \boldsymbol{\beta}_4)$$

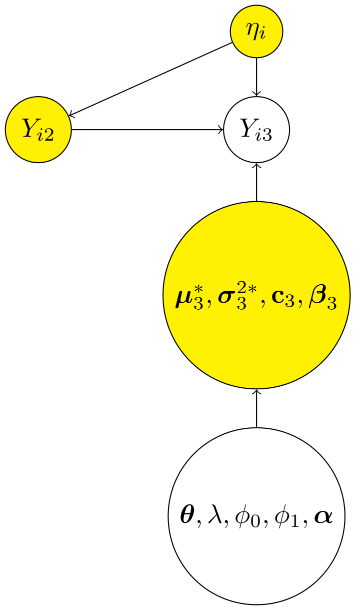
Which is the likelihood of Y_{i4}.

0.2 Second Term

Let us take the second term. We represent the remaining Bayesian network:



We can see that, conditioning on $(Y_{i2}, \boldsymbol{\mu}_3^*, \boldsymbol{\sigma}_3^{2*}, \mathbf{c}_3, \boldsymbol{\beta}_3, \eta_i, \boldsymbol{\theta}, \lambda, \phi_0, \phi_1, \boldsymbol{\alpha})$ (i.e., the colored nodes), Y_{i3} is independent of $\boldsymbol{\mu}_2^*, \boldsymbol{\sigma}_2^{2*}, \mathbf{c}_2, \boldsymbol{\beta}_2, \boldsymbol{\mu}_4^*, \boldsymbol{\sigma}_4^{2*}, \mathbf{c}_4, \boldsymbol{\beta}_4,$.
We represent the remaining Bayesian network:



We can see that, conditioning on $(Y_{i2}, \boldsymbol{\mu}_3^*, \boldsymbol{\sigma}_3^{2*}, \mathbf{c}_3, \boldsymbol{\beta}_3, \eta_i)$ (i.e., the colored nodes), Y_{i3} is independent of $\boldsymbol{\theta}, \lambda, \phi_0, \phi_1, \boldsymbol{\alpha}$.

Thus, the second term becomes:

$$\mathcal{L}(Y_{i3} \mid Y_{i2}, \boldsymbol{\mu}_3^*, \boldsymbol{\sigma}_3^{2*}, \mathbf{c}_3, \eta_i, \boldsymbol{\beta}_3)$$

Which is the likelihood of Y_{i3}.

0.3 Conclusion

In this document, we proved that the full conditional of an observation Y_{it} $(i = 1, \dots, n, \quad t = 1, \dots, T-1)$ is the product of the likelihood of Y_{it} and $Y_{i,t-1}$.
In similar fashion, it is possible to prove that the full conditional of Y_{iT} $(i = 1, \dots, n)$ is the likelihood of Y_{iT}.