# Math Formula

## Factor graphs

Shows how a function of several variables can be factored into a product of simpler functions.

> $f(x, y, z) = (x + y) \cdot (y + z) \cdot (x + z)$

Very useful for representing posteriors.

`$

$$P(x1, \ldots, x_n) = P(x_1)\Pi P(x_i|x_{i-1})$$

$`

$P(m|x1, \ldots, x_n) = P(m) \cdot \Pi P(x_i|m)$

### modeling

- What graph should I use for this data?

### Inference

- Given the graph and data, what is the mean of x
- algorithm
  - Sampling
  - Variable elimination
  - Message-passing(Expectation Propagation,

Variational Bayes)

## Cutter problem

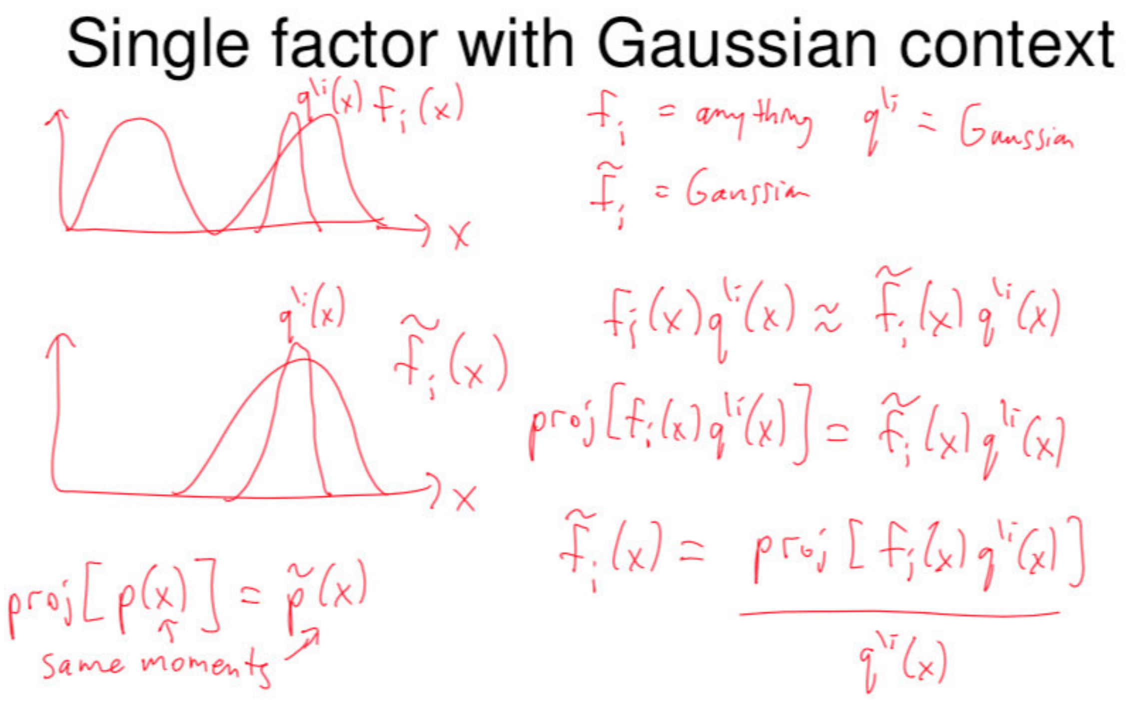
- Want to estimate x given multiple y's
- $p(x) = \mathcal{N}(x; 0, 100)$
- $p(y_i|x) = (0.5)\mathcal{N}(y_i; x, 1) + (0.5)\mathcal{N}(y_i; 0, 10)$

-> $P(x|y1, \ldots, y_n) = P(x) \cdot \Pi P(y_i|x)$

if we only have 2 points:

$P(x) \cdot P(y_1|x) \cdot P(y_2|x) \to p(y_i|x) = (0.5)\mathcal{N}(y_i; x, 1) + (0.5)\mathcal{N}(y_i; 0, 10)$

2 points have 4 Gaussians -> N points $2^N$ Gaussians



# Gaussian multiplication formula

$$\mathcal{N}(x; m_1, v_1)\mathcal{N}(x; m_2, v_2) = \mathcal{N}(m_1; m_2, v_1 + v_2)\mathcal{N}(x; m, v)$$

$$\text{where } v = \frac{1}{\frac{1}{v_1} + \frac{1}{v_2}}$$

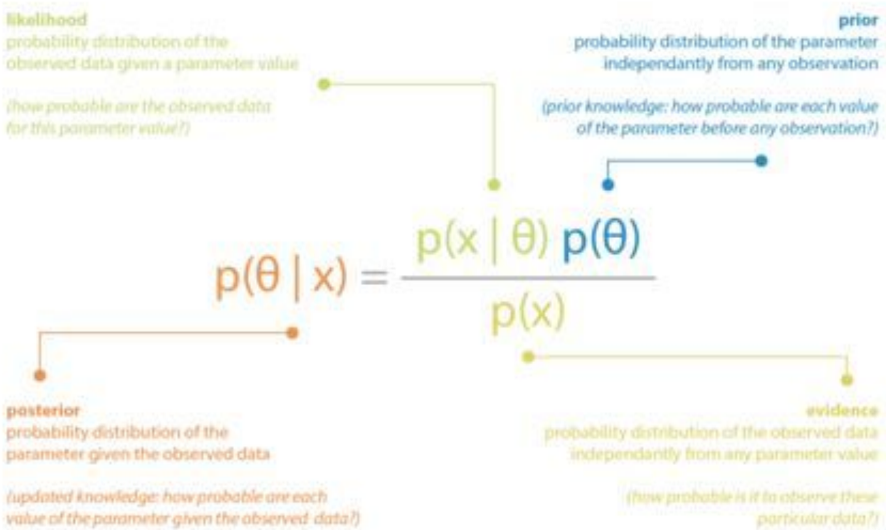
$$m = v\left(\frac{m_1}{v_1} + \frac{m_2}{v_2}\right)$$

$$\mathcal{N}(x; m_1, v_1)/\mathcal{N}(x; m_2, v_2) = \frac{v_2\mathcal{N}(x; m, v)}{(v_2 - v_1)\mathcal{N}(m_1; m_2, v_2 - v_1)}$$

$$\text{where } v = \frac{1}{\frac{1}{v_1} - \frac{1}{v_2}}$$

$$m = v\left(\frac{m_1}{v_1} - \frac{m_2}{v_2}\right)$$

20




> https://zhuanlan.zhihu.com/p/75617364

$$p(z|w) = \frac{p(w|z)p(z)}{p(w)} = \frac{p(w|z)p(z)}{\int_z p(w|z)p(z)dz}$$

Because it extends belief propagation. Belief propagation passes the entire distribution is the message. While EP will only pass onto the distribution certain expectation distribution allows you to you get a very compact message.

## Expectation Propagation

1. Fits an exponential-family approximation to the posterior.
2. Belief propagation is a special case
3. Kalman filtering is a special case
4. Does not always converge.
    1. May get stuck due to improper distributions
    2. May oscillate due to loopy graph

---

## AGM

$$p(\mathbf{X}|\Theta) = \sum_{j=1}^{M} p_j p\left(\mathbf{X}|\xi_j\right)$$

- $\xi_j$ is the set of the parameters of component j.
- $p_j$ are the mixing proptions which must be positive and sum to one.
- $\Theta = \{p_1, \ldots, p_M, \xi_1, \ldots, \xi_M\}$ is the complete set of parameters fully characterizing the mixture.
- $M \geq 1$ is number of components in the mixture.

$$p\left(X|\theta\right) = \prod_{d=1}^{D} \sqrt{\frac{2}{\pi}} \frac{1}{\left(\sigma_{l_d} + \sigma_{r_d}\right)} \times \begin{cases} \exp\left[-\frac{\left(X_d - \mu_d\right)^2}{2\sigma_{l_d}^2}\right] & \text{if } X_d < \mu_d \\ \exp\left[-\frac{\left(X_d - \mu_d\right)^2}{2\sigma_{r_d}^2}\right] & \text{if } X_d \geq \mu_d \end{cases}$$

- $\vec{\mu} = (\mu_1, \ldots, \mu_D)$ is the mean
- $\vec{\sigma}_l = (\vec{\sigma}_{l_1}, \ldots, \vec{\sigma}_{l_D})$ is the left standard deviation
- $\vec{\sigma}_r = (\vec{\sigma}_{r_1}, \ldots, \vec{\sigma}_{r_D})$ is the right standard deviation

$$\log P = \sum_{d=1}^{D} \log \sqrt{\frac{2}{\pi}} - \frac{1}{2} \log(\sqrt{v_{l_d}} + \sqrt{v_{r_d}}) - \begin{cases} \frac{(X_d - \mu_d)^2}{2 v_{l_d}} & \text{if } X_d < \mu_d \\ \frac{(X_d - \mu_d)^2}{2 v_{r_d}} & \text{if } X_d \geq \mu_d \end{cases}$$

$$\frac{\partial \log P}{\partial v_{l_d}} = -\frac{1}{4} \frac{1}{v_{l_d} + \sqrt{v_{l_d} v_{r_d}}} + \begin{cases} \frac{(X_d - \mu_d)^2}{2 v_{l_d}^2} & \text{if } X_d < \mu_d \\ 0 & \text{if } X_d \geq \mu_d \end{cases}$$

---

$$p(X, \boldsymbol{\theta}) = \prod_i f_i(\boldsymbol{\theta}) = \prod_i p(x_i | \boldsymbol{\theta})$$

Here, $p(\vec{X})$ is very intractable to calculate and we don't know $f_i(\boldsymbol{\theta})$.

Now we consider using **EP**. The approximation, $q(\theta_j)$ , of the posterior, $p\left(\theta_j | \vec{X}\right)$ , is assumed to have same functional form.

$$q(\theta_j) = \frac{1}{Z} \prod_i \tilde{f}_i(\theta_j)$$

We hope that:

$$\mathrm{KL}(p \| q) = \mathrm{KL}\left(\frac{1}{p(X)} \prod_i f_i(\boldsymbol{\theta}) \| \frac{1}{Z} \prod_i \tilde{f}_i(\boldsymbol{\theta})\right)$$

In general, this minimization will be intractable because the KL divergence involves averaging with respect to the true distribution.

But We can use EP:

1. first choose a factor $\tilde{f}_j$ to approximate.
   Begin Loop, until the following steps are convergence.
2. second compute the cavity distribution $q^{\backslash j}(\boldsymbol{\theta})$:

$$
\begin{aligned}
q^{\backslash j}(\boldsymbol{\theta}) &= \frac{q(\boldsymbol{\theta})}{\tilde{f}_j(\boldsymbol{\theta})} \\
\hat{p} &= \frac{1}{Z_j} f_j(\boldsymbol{\theta}) q^{\backslash j}(\boldsymbol{\theta})
\end{aligned}
$$

Here $q^{\backslash j}(\boldsymbol{\theta})$ is called the cavity distribution. $\hat{p}$ is defined as a product of the *exact* factor $f_i$ with the rest of the factors *approximated*, normalised to 1, and the cavity distribution needs to be computed in order to express $\hat{p}$.

> $$
> \begin{aligned}
> q^{\text{new}}(\boldsymbol{\theta}) &\propto \tilde{f}_j(\boldsymbol{\theta}) \prod_{i \neq j} \tilde{f}_i(\boldsymbol{\theta}) \\
> p \propto \hat{p} &= f_j(\boldsymbol{\theta}) \prod_{i \neq j} \tilde{f}_i(\boldsymbol{\theta})
> \end{aligned}
> $$

3. Then compute the approximative distribution $q^{new}$

$$
\arg\min \mathrm{KL}(\hat{p} \| q^{new}(\theta)) = \arg\min \mathrm{KL}\left( \frac{f_j(\boldsymbol{\theta}) q^{\backslash j}(\boldsymbol{\theta})}{Z_j} \| q^{\text{new}}(\boldsymbol{\theta}) \right)
$$

> More generally, it is straightforward to obtain the required expectations for any member of the exponential family, provided it can be normalized, because the expected statistics can be related to the derivatives of the normalization coefficient.

bishop:

$$p(\mathbf{x}|\boldsymbol{\eta}) = h(\mathbf{x})g(\boldsymbol{\eta})\exp\{\boldsymbol{\eta}^{\mathrm{T}}\mathbf{u}(\mathbf{x})\}$$
$$-\nabla \ln g(\boldsymbol{\eta}) = \mathbb{E}[\mathbf{u}(\mathbf{x})]$$

4. Update the factor

$$q^{\text{new}}(\boldsymbol{\theta}) \propto \hat{p} = \frac{1}{Z_j} f_j(\boldsymbol{\theta}) q^{\backslash j}(\boldsymbol{\theta})$$

Then we easily obtain the formula for the approximation of $f_i$:

$$f_i \approx \tilde{f}_i = Z_i \frac{q^{\text{new}}(\boldsymbol{\theta})}{q^{\backslash j}(\boldsymbol{\theta})}$$

This division of distributions is from exponetial family, so does the result $\tilde{f}_i$. Now repeat it until parameter covergence.

End Loop.

5. Evaluate the approximation to the model evidence

After the algorithm has converged to a set of factors $\{\tilde{f}_i\}$, the approximate posterior as well as the model evidence can be computed as following:

$$p(X, \boldsymbol{\theta}) \simeq \prod_i \tilde{f}_i(\boldsymbol{\theta})$$
$$p(X) \simeq \int \prod_i \tilde{f}_i(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta}$$

---

$$p(\mathbf{X}|\boldsymbol{\theta}) = (1-w)\mathcal{A}(\mathbf{X}|\boldsymbol{\theta}, \mathbf{I_l}, \mathbf{I_r}) + w\mathcal{A}(\mathbf{X}|\mathbf{0}, a\mathbf{I_l}, a\mathbf{I_r})$$

where w is the proportion of background clutter. And the prior over $\theta$(mean) is taken to be Asymmetric Gaussian.

And

$$p(\boldsymbol{\theta}) = \mathcal{A}(\mathbf{X}|\mathbf{0}, b\mathbf{I_l}, b\mathbf{I_r})$$

$$p(X, \boldsymbol{\theta}) = p(\boldsymbol{\theta}) \prod_{n=1}^{N} p\left(\mathbf{x}_n|\boldsymbol{\theta}\right)$$

## 1. initialize the approximating factors

we select an approximating distribution from the exponential family to approximate the stochastic variables $\theta$

$$q_0(\boldsymbol{\theta}) = \mathcal{A}(\boldsymbol{\theta}|\mathbf{0}, b\mathbf{I_l}, b\mathbf{I_r})$$

$$\tilde{f}_n(\boldsymbol{\theta}) = s_n \mathcal{A}\left(\boldsymbol{\theta}|\mu_n, \sigma^{\mathbf{2}}_{\mathbf{r_n}}, \sigma^{\mathbf{2}}_{\mathbf{l_n}}\right) = s_n \mathcal{A}\left(\boldsymbol{\theta}|\mu_n, \mathbf{v_{r_n}I}, \mathbf{v_{l_n}I}\right)$$

$$s_n = \prod_{d=1}^{D} \sqrt{\frac{2}{\pi}} \frac{1}{(\sigma_{l_d} + \sigma_{r_d})}$$

While $\sigma_{l_n} \to \infty, \sigma_{r_n} \to \infty$ and $\mu_n = 0$.

## 2. initialize the posterior approximation $q(\boldsymbol{\theta})$

We chooses the parameter values a = 10, b = 100 and w = 0.5 and use $v$ denote $\sigma^2$ as following, then $\mathbf{v_r} = \mathbf{v_l} = b = 100$

## 3. Until all $(\mu_n, v_{l_n}, v_{r_n}, s_n)$ converge:

$$
\begin{aligned}
q^{\setminus n}(\boldsymbol{\theta}) &= \frac{q(\boldsymbol{\theta})}{\tilde{f}_n(\boldsymbol{\theta})} = \frac{\mathcal{A}(\boldsymbol{\theta}|\mu, \mathbf{v_r I}, \mathbf{v_l I})}{s_n \mathcal{A}\left(\boldsymbol{\theta}|\mu_n, \mathbf{v_{r_n} I}, \mathbf{v_{l_n} I}\right)} \\
&\propto \begin{cases} \frac{\exp\left\{-\frac{1}{2}(\boldsymbol{X}-\mu)^T (v_l \mathbf{I})^{-1} (\boldsymbol{X}-\mu)\right\}}{\exp\left\{-\frac{1}{2}(\boldsymbol{X}-\mu_{\mathbf{n}})^T (v_{l_n} \mathbf{I})^{-1} (\boldsymbol{X}-\mu_{\mathbf{n}})\right\}} & \text{if } X < \mu \\ \frac{\exp\left\{-\frac{1}{2}(\boldsymbol{X}-\mu)^T (v_r \mathbf{I})^{-1} (\boldsymbol{X}-\mu)\right\}}{\exp\left\{-\frac{1}{2}(\boldsymbol{X}-\mu_{\mathbf{n}})^T (v_{r_n} \mathbf{I})^{-1} (\boldsymbol{X}-\mu_{\mathbf{n}})\right\}} & \text{if } X > \mu \end{cases} \\
&= \begin{cases} \exp\{-\frac{1}{2}(\boldsymbol{X}-\mu)^T (v_l \mathbf{I})^{-1} (\boldsymbol{X}-\mu) + \frac{1}{2}(\boldsymbol{X}-\mu_{\mathbf{n}})^T (v_{l_n} \mathbf{I})^{-1} (\boldsymbol{X}-\mu_{\mathbf{n}})\} & \text{if } X < \mu \\ \exp\{-\frac{1}{2}(\boldsymbol{X}-\mu)^T (v_r \mathbf{I})^{-1} (\boldsymbol{X}-\mu) + \frac{1}{2}(\boldsymbol{X}-\mu_{\mathbf{n}})^T (v_{r_n} \mathbf{I})^{-1} (\boldsymbol{X}-\mu_{\mathbf{n}})\} & \text{if } X > \mu \end{cases}
\end{aligned}
$$

- Remove the current estimate $\tilde{f}_j(\boldsymbol{\theta})$ from $q(\theta)$, then we has mean and inverse variance given by:

$$
\begin{cases} \left(v_l{}^{\setminus n}\right)^{-1} = v_l{}^{-1} - v_{l\,n}^{-1} & \text{if } X < \mu \\ \left(v_r{}^{\setminus n}\right)^{-1} = v_r{}^{-1} - v_{r\,n}^{-1} & \text{if } X > \mu \end{cases}
$$

$$
\mu^{\setminus n} = \mu + \begin{cases} v_l{}^{\setminus n} v_{l\,n}^{-1}\left(\mu - \mu_n\right) & \text{if } X < \mu \\ v_r{}^{\setminus n} v_{r\,n}^{-1}\left(\mu - \mu_n\right) & \text{if } X > \mu \end{cases}
$$

$$
----
$$

$$
\begin{aligned}
v^{\setminus n\,-1} &= v^{-1} - v_n^{-1} \\
\mu^{\setminus n} &= v^{\setminus n}(\mu v^{-1} - \mu_n v_n^{-1}) \\
&= v^{\setminus n}[\mu(v^{\setminus n\,-1} + v_n^{-1}) - \mu_n v_n^{-1}] \\
&= \mu + v^{\setminus n} v_n^{-1} \mu - v^{\setminus n} v_n^{-1} \mu_n \\
&= \mu + v^{\setminus n} v_n^{-1} (\mu - \mu_n)
\end{aligned}
$$

> Cavity Distribution:
>
> $$q^{\setminus n}(\boldsymbol{\theta}) = \frac{q(\boldsymbol{\theta})}{\tilde{f}_n(\boldsymbol{\theta})}$$

- Recompute $(\mu, v, Z)$ from $(\mu^{\setminus n}, v_l{}^{\setminus n}, v_r{}^{\setminus n})$

$$Z_n = (1-w)\mathcal{A}\left(\mathbf{x}_n|\mu^{\setminus n}, \left(v_l^{\setminus n}+1\right)\mathbf{I}, \left(v_r^{\setminus n}+1\right)\mathbf{I}\right) + w\mathcal{A}\left(\mathbf{x}_n|\mathbf{0}, a\mathbf{I_l}, a\mathbf{I_r}\right)$$

$$\begin{aligned}
Z_n &= \int q^{\setminus n}(\boldsymbol{\theta}) f_n(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta} \\
&= \int q^{\setminus n}(\boldsymbol{\theta}) P(X|\mu) \mathrm{d}\boldsymbol{\theta} \\
&= \int \mathcal{A}(\boldsymbol{\theta}|\mu^{\setminus \mathbf{n}}, v_l^{\setminus n}\mathbf{I}, v_r^{\setminus n}\mathbf{I}) \cdot \{(1-w)\mathcal{A}(\mathbf{x_n}|\boldsymbol{\mu}, \mathbf{I_l}, \mathbf{I_r}) + w\mathcal{A}(\mathbf{x_n}|\mathbf{0}, a\mathbf{I_l}, a\mathbf{I_r})\} \mathrm{d}\boldsymbol{\theta} \\
&= (1-w)\int \mathcal{A}(\boldsymbol{\theta}|\mu^{\setminus \mathbf{n}}, v_l^{\setminus n}\mathbf{I}, v_r^{\setminus n}\mathbf{I}) \mathcal{A}(\mathbf{x_n}|\boldsymbol{\mu}, \mathbf{I_l}, \mathbf{I_r}) \mathrm{d}\boldsymbol{\theta} \\
&+ w\int \mathcal{A}(\boldsymbol{\theta}|\mu^{\setminus \mathbf{n}}, v_l^{\setminus n}\mathbf{I}, v_r^{\setminus n}\mathbf{I}) \mathcal{A}(\mathbf{x_n}|\mathbf{0}, a\mathbf{I_l}, a\mathbf{I_r})\} \mathrm{d}\boldsymbol{\theta} \\
&= (1-w)\mathcal{A}\left(\mathbf{x}_n|\mu^{\setminus n}, \left(v_l^{\setminus n}+1\right)\mathbf{I}, \left(v_r^{\setminus n}+1\right)\mathbf{I}\right) + w\mathcal{A}\left(\mathbf{x}_n|\mathbf{0}, a\mathbf{I_l}, a\mathbf{I_r}\right)
\end{aligned}$$

we assumed that $f_0(\boldsymbol{\theta}) = p(\boldsymbol{\theta})$ and $f_n(\boldsymbol{\theta}) = p(\mathbf{x}_n|\boldsymbol{\theta}) = (1-w)\mathcal{A}(\mathbf{X}|\boldsymbol{\mu}, \mathbf{I_l}, \mathbf{I_r}) + w\mathcal{A}(\mathbf{X}|\mathbf{0}, a\mathbf{I_l}, a\mathbf{I_r})$, also $q(\boldsymbol{\theta}) = \mathcal{A}(\boldsymbol{\theta}|\mathbf{m}, v_l\mathbf{I}, v_r\mathbf{I})$

$$\begin{aligned}
\rho_n &= \frac{1}{Z_n}(1-w)\mathcal{A}\left(\mathbf{x}_n|\mu^{\setminus n}, \left(v_l^{\setminus n}+1\right)\mathbf{I}, \left(v_r^{\setminus n}+1\right)\mathbf{I}\right) \\
&= \frac{1}{Z_n}(1-w) \cdot \frac{Z_n - w\mathcal{A}\left(\mathbf{x}_n|\mathbf{0}, a\mathbf{I_l}, a\mathbf{I_r}\right)}{1-w} \\
&= 1 - \frac{w}{Z_n} \cdot \mathcal{A}\left(\mathbf{x}_n|\mathbf{0}, a\mathbf{I_l}, a\mathbf{I_r}\right)
\end{aligned}$$

Then our goal is to minimize:

$$\mathrm{KL}\left(\frac{f_n(\boldsymbol{\theta}) q^{\setminus n}(\boldsymbol{\theta})}{Z_n} \| q^{\text{new}}(\boldsymbol{\theta})\right)$$

Basic rule for Asymmetric Gaussian:

[https://stats.stackexchange.com/questions/27436/how-to-take-derivative-of-multivariate-normal-density](https://stats.stackexchange.com/questions/27436/how-to-take-derivative-of-multivariate-normal-density)

$$\nabla_{\boldsymbol{\mu}} \mathcal{A}(\mathbf{x}|\boldsymbol{\mu}, \mathbf{v_l}, \mathbf{v_r}) = \begin{cases} \mathcal{A}(\mathbf{x}|\boldsymbol{\mu}, \mathbf{v_l}, \mathbf{v_r}) \cdot (\mathbf{x} - \boldsymbol{\mu}) \mathbf{v_l}^{-1} & \text{if } X < \mu \\ \mathcal{A}(\mathbf{x}|\boldsymbol{\mu}, \mathbf{v_l}, \mathbf{v_r}) \cdot (\mathbf{x} - \boldsymbol{\mu}) \mathbf{v_r}^{-1} & \text{if } X > \mu \end{cases}$$

So we compute the mean and variance:

$$\begin{aligned}
\nabla_{\mu^{\backslash n}} \ln Z_n &= \frac{1}{Z_n} \cdot \nabla_{\mu^{\backslash n}} Z_n \\
&= \frac{1}{Z_n} \cdot \nabla_{\mu^{\backslash n}} \int q^{\backslash n}(\boldsymbol{\theta}) f_n(\boldsymbol{\theta}) d\boldsymbol{\theta} \\
&= \frac{1}{Z_n} \cdot \nabla_{\mu^{\backslash n}} \int q^{\backslash n}(\boldsymbol{\theta}) p\left(\mathbf{x}_n | \boldsymbol{\theta}\right) d\boldsymbol{\theta} \\
&= \frac{1}{Z_n} \cdot \int \left\{ \nabla_{\mu^{\backslash n}} q^{\backslash n}(\boldsymbol{\theta}) \right\} \cdot p\left(\mathbf{x}_n | \boldsymbol{\theta}\right) d\boldsymbol{\theta} \\
&= \frac{1}{Z_n} \cdot \int \frac{1}{v^{\backslash n}} \left( \boldsymbol{\theta} - \mu^{\backslash n} \right) \cdot q^{\backslash n}(\boldsymbol{\theta}) \cdot p\left(\mathbf{x}_n | \boldsymbol{\theta}\right) d\theta \\
&= \frac{1}{Z_n} \cdot \frac{1}{v^{\backslash n}} \cdot \left\{ \int \boldsymbol{\theta} \cdot q^{\backslash n}(\boldsymbol{\theta}) \cdot p\left(\mathbf{x}_n | \boldsymbol{\theta}\right) d\boldsymbol{\theta} - \int \mu^{\backslash n} \cdot q^{\backslash n}(\boldsymbol{\theta}) \cdot p\left(\mathbf{x}_n | \boldsymbol{\theta}\right) d\boldsymbol{\theta} \right\} \\
&= \frac{1}{v^{\backslash n}} \cdot \left\{ \mathbb{E}[\boldsymbol{\theta}] - \mu^{\backslash n} \right\} \\
&= \left\{ \mathbb{E}[\boldsymbol{\theta}] - \mu^{\backslash n} \right\} \cdot \begin{cases} \frac{1}{v_l^{\backslash n}} & \text{if } X_d < \mu_d \\ \frac{1}{v_r^{\backslash n}} & \text{if } X_d \geqslant \mu_d \end{cases}
\end{aligned}$$

According to the following :

$$q^{\backslash n}(\boldsymbol{\theta}) = \mathcal{A}(\boldsymbol{\theta}|\mu^{\backslash \mathbf{n}}, v_l^{\backslash n}\mathbf{I}, v_r^{\backslash n}\mathbf{I})$$

$$q^{\backslash n}(\boldsymbol{\theta}) \cdot p\left(\mathbf{x}_n | \boldsymbol{\theta}\right) = Z_n \cdot q^{new}(\theta)$$

$$
\begin{aligned}
\mathbb{E}[\boldsymbol{\theta}] &= \mu^{\backslash n} + v^{\backslash n} \cdot \nabla_{\mu^{\backslash n}} \ln Z_n \\
&= \mu^{\backslash n} + v^{\backslash n} \cdot \frac{1}{Z_n} \nabla_{\mu^{\backslash n}} Z_n \\
&= \mu^{\backslash n} + v^{\backslash n} \cdot \frac{1}{Z_n} \nabla_{\mu^{\backslash n}} (1-w) \mathcal{A}\left(\mathbf{x}_n | \mu^{\backslash n}, \left(v_l^{\backslash n} + 1\right) \mathbf{I}, \left(v_r^{\backslash n} + 1\right) \mathbf{I}\right) + w \mathcal{A}\left(\mathbf{x}_n | \mathbf{0}, a\mathbf{I_l}, a\mathbf{I_r}\right) \\
&= \mu^{\backslash n} + v^{\backslash n} \cdot \frac{1}{Z_n} (1-w) \nabla_{\mu^{\backslash n}} \mathcal{A}\left(\mathbf{x}_n | \mu^{\backslash n}, \left(v_l^{\backslash n} + 1\right) \mathbf{I}, \left(v_r^{\backslash n} + 1\right) \mathbf{I}\right) \\
&= \mu^{\backslash n} + v^{\backslash n} \cdot \rho_n \cdot \frac{1}{v^{\backslash n} + 1} \left(\mathbf{x}_n - \mu^{\backslash n}\right) \\
&= \mu^{\backslash n} + \rho_n \cdot \begin{cases} \frac{1}{v_l^{\backslash n}+1} \left(\mathbf{x}_n - \mu^{\backslash n}\right) \cdot v_l^{\backslash n} & \text{if } X_d < \mu_d \\ \frac{1}{v_r^{\backslash n}+1} \left(\mathbf{x}_n - \mu^{\backslash n}\right) \cdot v_l^{\backslash n} & \text{if } X_d \geqslant \mu_d \end{cases}
\end{aligned}
$$

According to: $\rho n = 1 - \frac{w}{Z_n} \cdot \mathcal{A}\left(\mathbf{x}_n | \mathbf{0}, a\mathbf{I_l}, a\mathbf{I_r}\right)$

Here we match first moment : $\mathbb{E}[\boldsymbol{\theta}] = \mu^{\mathbf{new}}$

Now we consider when:

$$\text{if } X_d < \mu_d$$

$$
\begin{aligned}
\nabla_{v_l^{\setminus n}} \ln Z_n &= \frac{1}{Z_n} \cdot \nabla_{v_l^{\setminus n}} Z_n \\
&= \frac{1}{Z_n} \cdot \nabla_{v_l^{\setminus n}} \int q^{\setminus n}(\boldsymbol{\theta}) p\left(\mathbf{x}_n | \boldsymbol{\theta}\right) d\boldsymbol{\theta} \\
&= \frac{1}{Z_n} \cdot \int \left\{ \nabla_{v_l^{\setminus n}} q^{\setminus n}(\boldsymbol{\theta}) \right\} p\left(\mathbf{x}_n | \boldsymbol{\theta}\right) d\boldsymbol{\theta} \\
&= \frac{1}{Z_n} \cdot \int \left\{ \frac{1}{2\left(v_l^{\setminus n}\right)^2} \left\| \boldsymbol{\theta} - \mu^{\setminus n} \right\|^2 - \frac{D}{4v_l^{\setminus n} + \sqrt{v_l^{\setminus n} \cdot v_r^{\setminus n}}} \right\} q^{\setminus n}(\boldsymbol{\theta}) \cdot p\left(\mathbf{x}_n | \boldsymbol{\theta}\right) d\boldsymbol{\theta} \\
&= \int q^{\text{new}}(\boldsymbol{\theta}) \cdot \left\{ \frac{1}{2\left(v_l^{\setminus n}\right)^2} \left(\mu^{\setminus n} - \boldsymbol{\theta}\right)^T \left(\mu^{\setminus n} - \boldsymbol{\theta}\right) - \frac{D}{4v_l^{\setminus n} + \sqrt{v_l^{\setminus n} \cdot v_r^{\setminus n}}} \right\} d\boldsymbol{\theta} \\
&= \frac{1}{2\left(v_l^{\setminus n}\right)^2} \left\{ \mathbb{E}\left[\boldsymbol{\theta}\boldsymbol{\theta}^T\right] - 2\mathbb{E}[\boldsymbol{\theta}]\mu^{\setminus n} + \left\| \mu^{\setminus n} \right\|^2 \right\} - \frac{D}{4v_l^{\setminus n} + \sqrt{v_l^{\setminus n} \cdot v_r^{\setminus n}}}
\end{aligned}
$$

So we rearrange the above equation:

$$
\mathbb{E}\left[\boldsymbol{\theta}\boldsymbol{\theta}^T\right] = 2\left(v_l^{\setminus n}\right)^2 \cdot \nabla_{v_l^{\setminus n}} \ln Z_n + 2\mathbb{E}[\boldsymbol{\theta}]\mathbf{m}^{\setminus n} - \left\| \mathbf{m}^{\setminus n} \right\|^2 + \frac{D \cdot \left(v_l^{\setminus n}\right)^2}{4v_l^{\setminus n} + \sqrt{v_l^{\setminus n} \cdot v_r^{\setminus n}}}
$$

Also according to:

$$
Z_n = (1-w)\mathcal{A}\left(\mathbf{x}_n | \mu^{\setminus n}, \left(v_l^{\setminus n} + 1\right)\mathbf{I}, \left(v_r^{\setminus n} + 1\right)\mathbf{I}\right) + w\mathcal{A}\left(\mathbf{x}_n | \mathbf{0}, a\mathbf{I_l}, a\mathbf{I_r}\right)
$$

$$
\begin{aligned}
\nabla_{v_l^{\backslash n}} \ln Z_n &= (1-w)\mathcal{A}\left(\mathbf{x}_n | \mu^{\backslash n}, \left(v_l^{\backslash n}+1\right)\mathbf{I}, \left(v_r^{\backslash n}+1\right)\mathbf{I}\right) \cdot \\
&\left(\frac{1}{2\left(v_l^{\backslash n}+1\right)^2}\left\|\mathbf{x_n}-\mu^{\backslash n}\right\|^2 - \frac{D}{4v_l^{\backslash n}+\sqrt{v_l^{\backslash n}\cdot v_r^{\backslash n}}}\right) \\
&= \rho_n \cdot \left(\frac{1}{2\left(v_l^{\backslash n}+1\right)^2}\left\|\mathbf{x_n}-\mu^{\backslash n}\right\|^2 - \frac{D}{4v_l^{\backslash n}+\sqrt{v_l^{\backslash n}\cdot v_r^{\backslash n}}}\right)
\end{aligned}
$$

According to the bellow formula:

$$
v\mathbf{I} = \mathbb{E}\left[\boldsymbol{\theta}\boldsymbol{\theta}^T\right] - \mathbb{E}[\boldsymbol{\theta}]\mathbb{E}\left[\boldsymbol{\theta}^T\right]
$$

$$
\begin{aligned}
v_l^{new} &= \frac{1}{D}\cdot\left\{\mathbb{E}\left[\boldsymbol{\theta}^T\boldsymbol{\theta}\right] - \mathbb{E}\left[\boldsymbol{\theta}^T\right]\mathbb{E}[\boldsymbol{\theta}]\right\} = \frac{1}{D}\cdot\left\{\mathbb{E}\left[\boldsymbol{\theta}^T\boldsymbol{\theta}\right] - \|\mathbb{E}[\boldsymbol{\theta}]\|^2\right\} \\
&= \frac{1}{D}\cdot\left\{2\left(v_l^{\backslash n}\right)^2\cdot\nabla_{v_l^{\backslash n}}\ln Z_n + 2\mathbb{E}[\boldsymbol{\theta}]\mu^{\backslash n} - \left\|\mu^{\backslash n}\right\|^2 + \frac{D\cdot\left(v_l^{\backslash n}\right)^2}{4v_l^{\backslash n}+\sqrt{v_l^{\backslash n}\cdot v_r^{\backslash n}}} - \|\mathbb{E}[\boldsymbol{\theta}]\|^2\right\} \\
&= \frac{1}{D}\cdot\left\{2\left(v_l^{\backslash n}\right)^2\cdot\nabla_{v_l^{\backslash n}}\ln Z_n - \left\|\mathbb{E}[\boldsymbol{\theta}]-\mu^{\backslash n}\right\|^2 + \frac{D\cdot\left(v_l^{\backslash n}\right)^2}{4v_l^{\backslash n}+\sqrt{v_l^{\backslash n}\cdot v_r^{\backslash n}}}\right\} \\
&= \frac{1}{D}\cdot\left\{2\left(v_l^{\backslash n}\right)^2\cdot\nabla_{v_l^{\backslash n}}\ln Z_n - \left\|\rho_n\cdot\frac{1}{v_l^{\backslash n}+1}\left(\mathbf{x}_n-\mu^{\backslash n}\right)\cdot v_l^{\backslash n}\right\|^2 + \frac{D\cdot\left(v_l^{\backslash n}\right)^2}{4v_l^{\backslash n}+\sqrt{v_l^{\backslash n}\cdot v_r^{\backslash n}}}\right\}
\end{aligned}
$$

substitute $\nabla_{v_l^{\backslash n}}\ln Z_n$:

$$v_l^{new} = \frac{1}{D}\cdot\left\{2\left(v_l^{\backslash n}\right)^2\cdot\nabla_{v_l^{\backslash n}}\ln Z_n - \left\|\rho_n\cdot\frac{1}{v_l^{\backslash n}+1}\left(\mathbf{x}_n-\mu^{\backslash n}\right)\cdot v_l^{\backslash n}\right\|^2 + \frac{D\cdot\left(v_l^{\backslash n}\right)^2}{4v_l^{\backslash n}+\sqrt{v_l^{\backslash n}\cdot v_r^{\backslash n}}}\right\}$$

$$= \frac{1}{D}\cdot\left\{2\left(v_l^{\backslash n}\right)^2\cdot\rho_n\cdot\left(\frac{1}{2\left(v_l^{\backslash n}+1\right)^2}\|\mathbf{x_n}-\mu^{\backslash n}\|^2 - \frac{D}{4v_l^{\backslash n}+\sqrt{v_l^{\backslash n}\cdot v_r^{\backslash n}}}\right) - \left\|\rho_n\cdot\frac{1}{v_l^{\backslash n}+1}\left(\mathbf{x}_n-\mu^{\backslash n}\right)\cdot v_l^{\backslash n}\right\|^2 + \frac{D\cdot\left(v_l^{\backslash n}\right)^2}{4v_l^{\backslash n}+\sqrt{v_l^{\backslash n}\cdot v_r^{\backslash n}}}\right\}$$

$$= \frac{\left(v_l^{\backslash n}\right)^2 - 2\left(v_l^{\backslash n}\right)^2\rho_n}{4v_l^{\backslash n}+\sqrt{v_l^{\backslash n}\cdot v_r^{\backslash n}}} + \rho_n\left(1-\rho_n\right)\frac{\left(v_l^{\backslash n}\right)^2\|\mathbf{x}_n-\mu^{\backslash n}\|^2}{D\left(v_l^{\backslash n}+1\right)^2}$$

So in conclusions: we have:

$$\mu^{\mathbf{new}} = \mu^{\backslash n} + \begin{cases} \rho_n\frac{v_l^{\backslash n}}{v_l^{\backslash n}+1}\left(\mathbf{x}_n-\mu^{\backslash n}\right) & \text{if } X<\mu \\ \rho_n\frac{v_r^{\backslash n}}{v_r^{\backslash n}+1}\left(\mathbf{x}_n-\mu^{\backslash n}\right) & \text{if } X>\mu \end{cases}$$

$$\begin{cases} v_l^{new} = \frac{\left(v_l^{\backslash n}\right)^2-2\left(v_l^{\backslash n}\right)^2\rho_n}{4v_l^{\backslash n}+\sqrt{v_l^{\backslash n}\cdot v_r^{\backslash n}}} + \rho_n\left(1-\rho_n\right)\frac{\left(v_l^{\backslash n}\right)^2\left\|\mathbf{x}_n-\mu^{\backslash n}\right\|^2}{D\left(v_l^{\backslash n}+1\right)^2} & \text{if } X<\mu \\ v_r^{new} = \frac{\left(v_r^{\backslash n}\right)^2-2\left(v_r^{\backslash n}\right)^2\rho_n}{4v_r^{\backslash n}+\sqrt{v_l^{\backslash n}\cdot v_r^{\backslash n}}} + \rho_n\left(1-\rho_n\right)\frac{\left(v_r^{\backslash n}\right)^2\left\|\mathbf{x}_n-\mu^{\backslash n}\right\|^2}{D\left(v_r^{\backslash n}+1\right)^2} & \text{if } X>\mu \end{cases}$$

- Evaluate and store the new factor

> Update formula for $\hat{f}_i$

$$\begin{cases} \left(v_{l_n}\right)^{-1} = \left(v_l^{new}\right)^{-1} - \left(v_l{}^{\backslash n}\right)^{-1} & \text{if } X<\mu \\ \left(v_{r_n}\right)^{-1} = \left(v_r^{new}\right)^{-1} - \left(v_r{}^{\backslash n}\right)^{-1} & \text{if } X>\mu \end{cases}$$

$$\mu_n = \mu^{\setminus n} + \begin{cases} \left(v_n + v^{\setminus n}\right)\left(v^{\setminus n}\right)^{-1}\left(\mu^{\text{new}} - \mu^{\setminus n}\right) & \text{if } X < \mu \\ \left(v_n + v^{\setminus n}\right)\left(v^{\setminus n}\right)^{-1}\left(\mu^{\text{new}} - \mu^{\setminus n}\right) & \text{if } X > \mu \end{cases}$$

$$\begin{aligned}
&\tilde{f}_n(\boldsymbol{\theta}) = Z_n \frac{q^{\text{new}}(\boldsymbol{\theta})}{q^{\setminus n}(\boldsymbol{\theta})} \\
&\Rightarrow Z_n q^{\text{new}}(\boldsymbol{\theta}) = s_n \mathcal{A}\left(\boldsymbol{\theta}|\mu_n, \mathbf{v}_{\mathbf{r_n}}\mathbf{I}, \mathbf{v}_{\mathbf{l_n}}\mathbf{I}\right) q^{\setminus n}(\boldsymbol{\theta}) = s_n \mathcal{A}\left(\boldsymbol{\theta}|\mu_n, \mathbf{v}_{\mathbf{r_n}}\mathbf{I}, \mathbf{v}_{\mathbf{l_n}}\mathbf{I}\right) \mathcal{A}\left(\boldsymbol{\theta}|\mu^{\setminus n}, \mathbf{v}_{\mathbf{r}}^{\setminus \mathbf{n}}\mathbf{I}, \mathbf{v}_{\mathbf{l}}^{\setminus \mathbf{n}}\mathbf{I}\right) \\
&\Rightarrow \int Z_n q^{\text{new}}(\boldsymbol{\theta}) d\theta = \int s_n \mathcal{A}\left(\boldsymbol{\theta}|\mu_n, \mathbf{v}_{\mathbf{r_n}}\mathbf{I}, \mathbf{v}_{\mathbf{l_n}}\mathbf{I}\right) \mathcal{A}\left(\boldsymbol{\theta}|\mu^{\setminus n}, \mathbf{v}_{\mathbf{r}}^{\setminus \mathbf{n}}\mathbf{I}, \mathbf{v}_{\mathbf{l}}^{\setminus \mathbf{n}}\mathbf{I}\right) d\theta \\
&\Rightarrow Z_n = s_n \int q^{\text{new}}(\boldsymbol{\theta}) d\theta = \int s_n \mathcal{A}\left(\mu_n - \boldsymbol{\theta}|0, \mathbf{v}_{\mathbf{r_n}}\mathbf{I}, \mathbf{v}_{\mathbf{l_n}}\mathbf{I}\right) \mathcal{A}\left(\boldsymbol{\theta}|\mu^{\setminus n}, \mathbf{v}_{\mathbf{r}}^{\setminus \mathbf{n}}\mathbf{I}, \mathbf{v}_{\mathbf{l}}^{\setminus \mathbf{n}}\mathbf{I}\right) d\theta \\
&\Rightarrow Z_n = s_n \mathcal{A}\left(\mu_n|\mu^{\setminus n}, (\mathbf{v}_{\mathbf{r}}^{\setminus \mathbf{n}} + \mathbf{v}_{\mathbf{r_n}})\mathbf{I}, (\mathbf{v}_{\mathbf{l}}^{\setminus \mathbf{n}} + \mathbf{v}_{\mathbf{r_n}})\mathbf{I}\right)
\end{aligned}$$

$$s_n = \frac{Z_n}{\mathcal{A}\left(\mu_n|\mu^{\setminus n}, (\mathbf{v}_{\mathbf{r}}^{\setminus \mathbf{n}} + \mathbf{v}_{\mathbf{r_n}})\mathbf{I}, (\mathbf{v}_{\mathbf{l}}^{\setminus \mathbf{n}} + \mathbf{v}_{\mathbf{r_n}})\mathbf{I}\right)}$$

- Evaluate the approximation to the model evidence - Posterior probability. (When $(\mu_n, v_{l\,n}, v_{r\,n}, S_n)$ unchanged )

$$p(X) \simeq q(\boldsymbol{\theta}) = \mathcal{A}(\boldsymbol{\theta}|\mu, \mathbf{v}_{\mathbf{l}}, \mathbf{v}_{\mathbf{r}}) = \prod_{n=0}^{N} \tilde{f}_n(\boldsymbol{\theta}) = f_0(\boldsymbol{\theta}) \prod_{n=1}^{N} \tilde{f}_n(\boldsymbol{\theta}) = \mathcal{A}(\boldsymbol{\theta}|\mathbf{0}, b\mathbf{I}_{\mathbf{l}}, b\mathbf{I}_{\mathbf{r}}) \cdot \prod_{i=1}^{N} \mathcal{A}\left(\boldsymbol{\mu}_i, \mathbf{v}_{\mathbf{l}\,i}, \mathbf{v}_{\mathbf{r}\,i}\right)$$