1 Laplacian Mixture Model (LMM)

1.1 The maximum likelihood of the model

Let $\log(L(\lambda; X, Z))$ be the log-ML of the model where $X = \{x_1, x_2, \ldots, x_N\}$ are data points in R^D, $Z_i = k$ is the event of x_i being associated with mixture component k and λ are the parameters of the model.

$$\begin{aligned}
\log(L(\lambda; X, Z)) &= \sum_{i=1}^{N} \log(\sum_{k=1}^{K} I(z_i = k)\tau_k \cdot f(x_i; m_k, s_k)) \\
&= \sum_{i=1}^{N}\sum_{k=1}^{K} I(z_i = k)\left[log(\tau_k) + log(f(x_i; m_k, s_k)\right] \\
&= \sum_{i=1}^{N}\sum_{k=1}^{K} I(z_i = k)\left[log(\tau_k) + log\left(\prod_{d=1}^{D}\left(\frac{1}{2s_{k,d}}exp\left(-\frac{|x_{i,d} - m_{k,d}|}{s_{k,d}}\right)\right)\right)\right] \\
&= \sum_{i=1}^{N}\sum_{k=1}^{K} I(z_i = k)\left[log(\tau_k) + \left(\sum_{d=1}^{D}\left(-\log\left(2s_{k,d}\right) - \frac{|x_{i,d} - m_{k,d}|}{s_{k,d}}\right)\right)\right]
\end{aligned}$$

1.2 Expectation step (E step)

Let $T_{k,i}^{(t)}$ be the conditional probability of sample x_i being associated with component k, given the sample x_i and the current estimate of the parameters $\boldsymbol{\lambda}^{(t)}$

$$\begin{aligned}
T_{k,i}^{(t)} &= P\left(Z_i = k | X_i = x_i; \lambda\left(t\right)\right) \\
&= \frac{P\left(X_i = x_i | Z_i = k; \lambda\left(t\right)\right) P\left(Z_i = k\right)}{\sum_{r=1}^{K} P\left(X_i = x_i | Z_i = r; \lambda\left(t\right)\right) P\left(Z_i = r\right)} \\
&= \frac{\tau_k \cdot f\left(x_i; m_k, s_k\right)}{\sum_{r=1}^{K} \tau_r \cdot f\left(x_i; m_r, s_r\right)}
\end{aligned}$$

Let $Q\left(\lambda|\lambda^{(t)}\right)$ be the expected value of the log likelihood function, with respect to the conditional distribution of $\mathbf{Z}$ given $\mathbf{X}$ under the current estimate of the parameters $\boldsymbol{\lambda}^{(t)}$

$$\begin{aligned}
Q\left(\lambda|\lambda^{(t)}\right) &= \mathrm{E}_{Z|X,\lambda^{(t)}}\left[\log\left(L\left(\lambda; X, Z\right)\right)\right] \\
&= \sum_{i=1}^{N}\sum_{k=1}^{K} T_{k,i}^{(t)}\left[log(\tau_k) + \left(\sum_{d=1}^{D}\left(-\log\left(2s_{k,d}\right) - \frac{|x_{i,d} - m_{k,d}|}{s_{k,d}}\right)\right)\right]
\end{aligned}$$

978-1-4673-6964-0/15/$31.00 ©2015 IEEE

1.3 Maximization step (M step)

Compute $\boldsymbol{\lambda}^{(t+1)} = \arg\max_{\boldsymbol{\lambda}} Q(\boldsymbol{\lambda}|\boldsymbol{\lambda}^{(t)})$

1.3.1 Derive τ_k

Solve $\frac{\partial Q\left(\lambda|\lambda^{(t)}\right)}{\partial \tau_k} = 0$ under the restriction $\sum_{r=1}^{K} \tau_r = 1$. Using lagrange multipliers:

$$Q\left(\lambda|\lambda^{(t)}\right) + \beta \cdot \left(\sum_{r=1}^{K} \tau_r - 1\right) = 0$$

Derive by τ_k:

$$\frac{1}{\tau_k} \cdot \sum_{i=1}^{N} T_{k,i}^{(t)} + \beta = 0$$
$$\tau_k = -\frac{1}{\beta} \cdot \sum_{i=1}^{N} T_{k,i}^{(t)}$$

Derive by β:

$$\sum_{r=1}^{K} \tau_r = 1$$

Therefore:

$$\sum_{r=1}^{K} -\frac{1}{\beta} \cdot \sum_{i=1}^{N} T_{r,i}^{(t)} = 1$$
$$\beta = -\sum_{r=1}^{K} \sum_{i=1}^{N} T_{r,i}^{(t)}$$
$$\tau_k = \frac{\sum_{i=1}^{N} T_{k,i}^{(t)}}{\sum_{r=1}^{K} \sum_{i=1}^{N} T_{r,i}^{(t)}}$$

1.3.2 Derive $m_{k,d}$

Solve $\frac{\partial Q\left(\lambda|\lambda^{(t)}\right)}{\partial mk,d} = 0$

$$\sum_{i=1}^{N} T_{k,i}^{(t)} \cdot (-1) \frac{\partial}{\partial m_{k,d}} \begin{cases} m_{k,d} - x_{i,d} & \text{if } m_{k,d} > x_{i,d} \\ x_{i,d} - m_{k,d} & \text{o.w.} \end{cases} = 0$$

$$\sum_{i=1}^{N} T_{k,i}^{(t)} \begin{cases} -1 & \text{if } m_{k,d} > x_{i,d} \\ 1 & \text{o.w.} \end{cases} = 0$$

$$\sum_{m_{k,d} \le x_{i,d}} T_{k,i}^{(t)} = \sum_{m_{k,d} > x_{i,d}} T_{k,i}^{(t)}$$

1.3.3 Derive $s_{k,d}$

Solve $\frac{\partial Q\left(\lambda|\lambda^{(t)}\right)}{\partial sk,d} = 0$

$$\sum_{i=1}^{N} T_{k,i}^{(t)} (-1) \cdot \left(\frac{1}{s_{k,d}} - \frac{|x_{i,d} - m_{k,d}|}{s_{k,d}^2} \right) = 0$$

$$\sum_{i=1}^{N} T_{k,i}^{(t)} |x_{i,d} - m_{k,d}| = s_{k,d} \cdot \sum_{i=1}^{N} T_{k,i}^{(t)}$$

$$s_{k,d} = \frac{\sum_{i=1}^{N} T_{k,i}^{(t)} |x_{i,d} - m_{k,d}|}{\sum_{i=1}^{N} T_{k,i}^{(t)}}$$

1.4 Fisher Vector

Let $X = x_1, x_2, \ldots, x_N$ be new samples and let $\lambda = (m, s)$ be the parameters of the LMM that were found by running the EM on the training samples X_{Trn}. Let $\mathcal{L}(\lambda|X)$ be the log-likelihood of the new samples given λ.

$$\mathcal{L}(\lambda|X) = \sum_{i=1}^{N} \log\left(\sum_{k=1}^{K} \tau_k \cdot f(x_i; m_k, s_k)\right)$$

The fisher vector entries for $x_1, x_2, \ldots, x_N$ are the gradients of $\mathcal{L}(\lambda|X)$ with respect to λ. For convenience, let $T_{k,i} = \frac{\tau_k \cdot f(x_i; m_k, s_k)}{\sum_{r=1}^{K} \tau_r \cdot f(x_i; m_r, s_r)}$

1.4.1 $m_{k,d}$ entries

$$\frac{\partial}{\partial m_{k,d}} \mathcal{L}(\lambda|X) = \sum_{i=1}^{N} \frac{\frac{\partial}{\partial m_{k,d}} \sum_{r=1}^{K} \tau_r \cdot f(x_i; m_r, s_r)}{\sum_{r=1}^{K} \tau_r \cdot f(x_i; m_r, s_r)}$$

$$= \sum_{i=1}^{N} \frac{\tau_k \cdot \frac{\partial}{\partial m_{k,d}} f(x_i; m_k, s_k)}{\sum_{r=1}^{K} \tau_r \cdot f(x_i; m_r, s_r)}$$

$$= \sum_{i=1}^{N} \frac{\tau_k \cdot \frac{\partial}{\partial m_{k,d}} \prod_{u=1}^{D} \left(\frac{1}{2s_{k,u}} \exp\left(-\frac{|x_{i,u} - m_{k,u}|}{s_{k,u}}\right)\right)}{\sum_{r=1}^{K} \tau_r \cdot f(x_i; m_r, s_r)}$$

$$= \sum_{i=1}^{N} \frac{\tau_k \cdot \prod_{u \neq d}^{D} \left[\left(\frac{1}{2s_{k,u}} \exp\left(-\frac{|x_{i,u} - m_{k,u}|}{s_{k,u}}\right)\right)\right]}{\sum_{r=1}^{K} \tau_r f(x_i; m_r, s_r)} \cdot \frac{\partial}{\partial m_{k,d}} \left(\frac{1}{2s_{k,d}} \exp\left(-\frac{|x_{i,d} - m_{k,d}|}{s_{k,d}}\right)\right)$$

$$= \sum_{i=1}^{N} \frac{\tau_k \cdot f(x_i; m_k, s_k)}{\sum_{r=1}^{K} \tau_r \cdot f(x_i; m_r, s_r)} \cdot \frac{\partial}{\partial m_{k,d}} \left(-\frac{|x_{i,d} - m_{k,d}|}{s_{k,d}}\right)$$

$$= \sum_{i=1}^{N} \frac{T_{k,i}}{s_{k,d}} \cdot \begin{cases} 1 & \text{if } x_{i,d} > m_{k,d} \\ -1 & \text{o.w.} \end{cases}$$

1.4.2 $s_{k,d}$ entries

$$\frac{\partial}{\partial m_{k,d}}\mathcal{L}\left(\lambda|X\right)=\sum_{i=1}^{N}\frac{\frac{\partial}{\partial s_{k,d}}\sum_{r=1}^{K}\tau_r\cdot f\left(x_i;m_r,s_r\right)}{\sum_{r=1}^{K}\tau_r\cdot f\left(x_i;m_r,s_r\right)}$$

$$=\sum_{i=1}^{N}\frac{\tau_k\cdot\frac{\partial}{\partial m_{k,d}}f\left(x_i;m_k,s_k\right)}{\sum_{r=1}^{K}\tau_r\cdot f\left(x_i;m_r,s_r\right)}$$

$$=\sum_{i=1}^{N}\frac{\tau_k\cdot\frac{\partial}{\partial s_{k,d}}\prod_{u=1}^{D}\left(\frac{1}{2s_{k,u}}\exp\left(-\frac{|x_{i,u}-m_{k,u}|}{s_{k,u}}\right)\right)}{\sum_{r=1}^{K}\tau_r\cdot f\left(x_i;m_r,s_r\right)}$$

$$=\sum_{i=1}^{N}\frac{\tau_k\cdot\prod_{u\neq d}^{D}\left[\left(\frac{1}{2s_{k,u}}\exp\left(-\frac{|x_{i,u}-m_{k,u}|}{s_{k,ud}}\right)\right)\right]}{\sum_{r=1}^{K}\tau_r f\left(x_i;m_r,s_r\right)}\cdot\frac{\partial}{\partial s_{k,d}}\left(\frac{1}{2s_{k,d}}\exp\left(-\frac{|x_{i,d}-m_{k,d}|}{s_{k,d}}\right)\right)$$

$$=\sum_{i=1}^{N}\frac{\tau_k\cdot\prod_{u\neq d}^{D}\left[\left(\frac{1}{2s_{k,u}}\exp\left(-\frac{|x_{i,u}-m_{k,u}|}{s_{k,ud}}\right)\right)\right]}{\sum_{r=1}^{K}\tau_r f\left(x_i;m_r,s_r\right)}\cdot$$

$$\left[\left(-\frac{1}{2s_{k,d}^2}\exp\left(-\frac{|x_{i,d}-m_{k,d}|}{s_{k,d}}\right)\right)+\left(\frac{1}{2s_{k,d}}\exp\left(-\frac{|x_{i,d}-m_{k,d}|}{s_{k,d}}\right)\cdot\left(\frac{|x_{i,d}-m_{k,d}|}{s_{k,d}^2}\right)\right)\right]$$

$$=\sum_{i=1}^{N}T_{k,i}\left(\frac{|x_{i,d}-m_{k,d}|}{s_{k,d}^2}-\frac{1}{s_{k,d}}\right)$$

1.5 Derivation of the Fisher Information Matrix

See 2.5.3 and 2.5.4 as the derivations are almost identical.

2 Hybrid Gaussian-Laplacian Mixture Model ($HGLMM$)

2.1 The maximum likelihood of the model

Let $\log(L(\lambda; X, Z))$ be the log-ML of the model where $X = \{x_1, x_2, \ldots, x_N\}$ are data points in R^D, $Z_i = k$ is the event of x_i being associated with mixture component k and λ are the parameters of the model.

$$
\begin{aligned}
\log(L(\lambda; X, Z)) &= \sum_{i=1}^{N} \log(\sum_{k=1}^{K} I(z_i = k)\tau_k \cdot f(x_i; \mu_k, \sigma_k, m_k, s_k, b_k)) \\
&= \sum_{i=1}^{N} \sum_{k=1}^{K} I(z_i = k) \left[log(\tau_k) + log(f(x_i; \mu_k, \sigma_k, m_k, s_k, b_k)\right] \\
&= \sum_{i=1}^{N} \sum_{k=1}^{K} I(z_i = k) \left[log(\tau_k) + log\left(\prod_{d=1}^{D}\left(\left(\frac{1}{2s_{k,d}} exp\left(-\frac{|x_{i,d} - m_{k,d}|}{s_{k,d}}\right)\right)^{b_{k,d}}\right.\right.\right. \\
&\cdot \left.\left.\left.\left(\frac{1}{\sqrt{2\pi}\sigma_{k,d}} exp\left(-\frac{(x_{i,d} - \mu_{k,d})^2}{2\sigma_{k,d}^2}\right)\right)^{(1-b_{k,d})}\right)\right)\right] \\
&= \sum_{i=1}^{N} \sum_{k=1}^{K} I(z_i = k) \left[log(\tau_k) + \left(\sum_{d=1}^{D} (b_{k,d}) \left(-\log(2s_{k,d}) - \frac{|x_{i,d} - m_{k,d}|}{s_{k,d}}\right)\right.\right. \\
&+ \left.\left. (1 - b_{k,d}) \cdot \left(-\log\left(\sqrt{2\pi}\sigma_{k,d}\right) - \frac{(x_{i,d} - \mu_{k,d})^2}{2\sigma_{k,d}^2}\right)\right)\right]
\end{aligned}
$$

2.2 Expectation step (E step)

Let $T_{k,i}^{(t)}$ be the conditional probability of sample x_i being associated with component k, given the sample x_i and the current estimate of the parameters $\boldsymbol{\lambda}^{(t)}$

$$
\begin{aligned}
T_{k,i}^{(t)} &= P\left(Z_i = k | X_i = x_i; \lambda(t)\right) \\
&= \frac{P\left(X_i = x_i | Z_i = k; \lambda(t)\right) P\left(Z_i = k\right)}{\sum_{r=1}^{K} P\left(X_i = x_i | Z_i = r; \lambda(t)\right) P\left(Z_i = r\right)} \\
&= \frac{\tau_k \cdot f\left(x_i; \mu_k, \sigma_k, m_k, s_k, b_k\right)}{\sum_{r=1}^{K} \tau_r \cdot f\left(x_i; \mu_r, \sigma_r, m_r, s_r, b_r\right)}
\end{aligned}
$$

Let $Q\left(\lambda|\lambda^{(t)}\right)$ be the expected value of the log likelihood function, with respect to the conditional distribution of $\mathbf{Z}$ given $\mathbf{X}$ under the current estimate of the parameters $\boldsymbol{\lambda}^{(t)}$

$$
\begin{aligned}
Q\left(\lambda|\lambda^{(t)}\right) &= \mathrm{E}_{Z|X,\lambda^{(t)}}\left[\log\left(L\left(\lambda;X,Z\right)\right)\right] \\
&= \sum_{i=1}^{N}\sum_{k=1}^{K} T_{k,i}^{(t)}\left[log(\tau_k) + \left(\sum_{d=1}^{D}\left(b_{k,d}\right)\left(-\log\left(2s_{k,d}\right) - \frac{|x_{i,d}-m_{k,d}|}{s_{k,d}}\right)\right.\right. \\
&+ \left.\left.\left(1-b_{k,d}\right)\cdot\left(-\log\left(\sqrt{2\pi}\sigma_{k,d}\right) - \frac{\left(x_{i,d}-\mu_{k,d}\right)^2}{2\sigma_{k,d}^2}\right)\right)\right]
\end{aligned}
$$

2.3 Maximization step (M step)

Compute $\boldsymbol{\lambda}^{(t+1)} = \arg\max_{\boldsymbol{\lambda}} Q(\boldsymbol{\lambda}|\boldsymbol{\lambda}^{(t)})$

2.3.1 Derive τ_k

Solve $\frac{\partial Q\left(\lambda|\lambda^{(t)}\right)}{\partial \tau k} = 0$ under the restriction $\sum_{r=1}^{K} \tau_r = 1$. Using lagrange multipliers:

$$Q\left(\lambda|\lambda^{(t)}\right) + \beta \cdot \left(\sum_{r=1}^{K} \tau_r - 1\right) = 0$$

Derive by τ_k:

$$\frac{1}{\tau_k} \cdot \sum_{i=1}^{N} T_{k,i}^{(t)} + \beta = 0$$

$$\tau_k = -\frac{1}{\beta} \cdot \sum_{i=1}^{N} T_{k,i}^{(t)}$$

Derive by β:

$$\sum_{r=1}^{K} \tau_r = 1$$

Therefore:

$$\sum_{r=1}^{K} -\frac{1}{\beta} \cdot \sum_{i=1}^{N} T_{r,i}^{(t)} = 1$$

$$\beta = -\sum_{r=1}^{K} \sum_{i=1}^{N} T_{r,i}^{(t)}$$

$$\tau_k = \frac{\sum_{i=1}^{N} T_{k,i}^{(t)}}{\sum_{r=1}^{K} \sum_{i=1}^{N} T_{r,i}^{(t)}}$$

2.3.2 Derive $\mu_{k,d}$

Solve $\frac{\partial Q\left(\lambda|\lambda^{(t)}\right)}{\partial \mu_{k,d}} = 0$

$$\sum_{i=1}^{N} T_{k,i}^{(t)} \left(1 - b_{k,d}\right) \cdot \left(\frac{\mu_{k,d} - x_{i,d}}{\sigma_{k,d}^2}\right) = 0$$

$$\sum_{i=1}^{N} T_{k,i}^{(t)} \left(\mu_{k,d} - x_{i,d}\right) = 0$$

$$\mu_{k,d} = \frac{\sum_{i=1}^{N} T_{k,i}^{(t)} \cdot x_{i,d}}{\sum_{i=1}^{N} T_{k,i}^{(t)}}$$

2.3.3 Derive $\sigma_{k,d}$

Solve $\frac{\partial Q\left(\lambda|\lambda^{(t)}\right)}{\partial \sigma_{k,d}} = 0$

$$\begin{aligned}
&\sum_{i=1}^{N} T_{k,i}^{(t)}\left(1-b_{k,d}\right)\cdot\left(-\frac{1}{\sigma_{k,d}}+\frac{\left(x_{i,d}-\mu_{k,d}\right)^{2}}{\sigma_{k,d}^{3}}\right)=0\\
&\sum_{i=1}^{N} T_{k,i}^{(t)}\left(x_{i,d}-\mu_{k,d}\right)^{2}=\sigma_{k,d}^{2}\cdot\sum_{i=1}^{N} T_{k,i}^{(t)}\\
&\sigma_{k,d}^{2}=\frac{\sum_{i=1}^{N} T_{k,i}^{(t)}\left(x_{i,d}-\mu_{k,d}\right)^{2}}{\sum_{i=1}^{N} T_{k,i}^{(t)}}
\end{aligned}$$

2.3.4 Derive $m_{k,d}$

Solve $\frac{\partial Q\left(\lambda|\lambda^{(t)}\right)}{\partial mk,d} = 0$

$$\begin{aligned}
&\sum_{i=1}^{N} T_{k,i}^{(t)}\cdot\left(-b_{k,d}\right)\frac{\partial}{\partial m_{k,d}}\begin{cases} m_{k,d}-x_{i,d} & \text{if } m_{k,d}>x_{i,d}\\ x_{i,d}-m_{k,d} & \text{o.w.}\end{cases}=0\\
&\sum_{i=1}^{N} T_{k,i}^{(t)}\begin{cases} -1 & \text{if } m_{k,d}>x_{i,d}\\ 1 & \text{o.w.}\end{cases}=0\\
&\sum_{m_{k,d}\leq x_{i,d}} T_{k,i}^{(t)}=\sum_{m_{k,d}>x_{i,d}} T_{k,i}^{(t)}
\end{aligned}$$

2.3.5 Derive $s_{k,d}$

Solve $\frac{\partial Q\left(\lambda|\lambda^{(t)}\right)}{\partial sk,d} = 0$

$$\begin{aligned}
&\sum_{i=1}^{N} T_{k,i}^{(t)}\left(-b_{k,d}\right)\cdot\left(\frac{1}{s_{k,d}}-\frac{\left|x_{i,d}-m_{k,d}\right|}{s_{k,d}^{2}}\right)=0\\
&\sum_{i=1}^{N} T_{k,i}^{(t)}\left|x_{i,d}-m_{k,d}\right|=s_{k,d}\cdot\sum_{i=1}^{N} T_{k,i}^{(t)}\\
&s_{k,d}=\frac{\sum_{i=1}^{N} T_{k,i}^{(t)}\left|x_{i,d}-m_{k,d}\right|}{\sum_{i=1}^{N} T_{k,i}^{(t)}}
\end{aligned}$$

2.3.6 Derive $b_{k,d}$

The contribution of $b_{k,d}$ to the log maximum likelihood is

$$\sum_{i=1}^{N} T_{k,i}^{(t)} \left[b_{k,d} \left(-\log\left(2s_{k,d}\right) - \frac{|x_{i,d} - m_{k,d}|}{s_{k,d}} \right) + (1 - b_{k,d}) \cdot \left(-\log\left(\sqrt{2\pi}\sigma_{k,d}\right) - \frac{\left(x_{i,d} - \mu_{k,d}\right)^2}{2\sigma_{k,d}^2} \right) \right] =$$
$$b_{k,d} \cdot \sum_{i=1}^{N} T_{k,i}^{(t)} \left(-\log\left(2s_{k,d}\right) - \frac{|x_{i,d} - m_{k,d}|}{s_{k,d}} \right) + (1 - b_{k,d}) \cdot \sum_{i=1}^{N} T_{k,i}^{(t)} \left(-\log\left(\sqrt{2\pi}\sigma_{k,d}\right) - \frac{\left(x_{i,d} - \mu_{k,d}\right)^2}{2\sigma_{k,d}^2} \right)$$

Let

$$L_{b_{k,d}} = \sum_{i=1}^{N} T_{k,i}^{(t)} \left(-\log\left(2s_{k,d}\right) - \frac{|x_{i,d} - m_{k,d}|}{s_{k,d}} \right)$$
$$G_{b_{k,d}} = \sum_{i=1}^{N} T_{k,i}^{(t)} \left(-\log\left(\sqrt{2\pi}\sigma_{k,d}\right) - \frac{\left(x_{i,d} - \mu_{k,d}\right)^2}{2\sigma_{k,d}^2} \right)$$

Therefore under the restriction that $0 \leq b_{k,d} \leq 1$ we get that

$$b_{k,d} = \begin{cases} 1 & \text{if } L_{b_{k,d}} > G_{b_{k,d}} \\ 0 & \text{o.w.} \end{cases}$$

2.4 Fisher Vector

Let $X = x_1, x_2, \ldots, x_N$ be new samples and let $\lambda = (\mu, \sigma, m, s, b)$ be the parameters of the $HGLMM$ that were found by running the EM on the training samples X_{Trn}. Let $\mathcal{L}(\lambda|X)$ be the log-likelihood of the new samples given λ.

$$\mathcal{L}(\lambda|X) = \sum_{i=1}^{N} \log \left(\sum_{k=1}^{K} \tau_k \cdot f(x_i; \mu_k, \sigma_k, m_k, s_k, b_k) \right)$$

The fisher vector entries for $x_1, x_2, \ldots, x_N$ are the gradients of $\mathcal{L}(\lambda|X)$ with respect to λ. For convenience, let $T_{k,i} = \frac{\tau_k \cdot f(x_i;\mu_k,\sigma_k,m_k,s_k,b_k)}{\sum_{r=1}^{K} \tau_r \cdot f(x_i;\mu_r,\sigma_r,m_r,s_r,b_r)}$

2.4.1 $\mu_{k,d}$ entries

$\frac{\partial}{\partial \mu_{k,d}} \mathcal{L}(\lambda|X) = 0$ if $b_{k,d} = 1$. Otherwise (i.e. $b_{k,d} = 0$):

$$\frac{\partial}{\partial \mu_{k,d}} \mathcal{L}(\lambda|X) = \sum_{i=1}^{N} \frac{\frac{\partial}{\partial \mu_{k,d}} \sum_{r=1}^{K} \tau_r \cdot f(x_i; \mu_r, \sigma_r, m_r, s_r, b_r)}{\sum_{r=1}^{K} \tau_r \cdot f(x_i; \mu_r, \sigma_r, m_r, s_r, b_r)}$$

$$= \sum_{i=1}^{N} \frac{\tau_k \cdot \frac{\partial}{\partial \mu_{k,d}} f(x_i; \mu_k, \sigma_k, m_k, s_k, b_k)}{\sum_{r=1}^{K} \tau_r \cdot f(x_i; \mu_r, \sigma_r, m_r, s_r, b_r)}$$

$$= \sum_{i=1}^{N} \frac{\tau_k \cdot \frac{\partial}{\partial \mu_{k,d}} \prod_{u=1}^{D} \left[\left(\frac{1}{2s_{k,u}} \exp\left(-\frac{|x_{i,u} - m_{k,u}|}{s_{k,u}} \right) \right)^{b_{k,u}} \cdot \left(\frac{1}{\sigma_{k,u}\sqrt{2\pi}} \exp\left(-\frac{(x_{i,u} - \mu_{k,u})^2}{2\sigma_{k,u}^2} \right) \right)^{(1-b_{k,u})} \right]}{\sum_{r=1}^{K} \tau_r \cdot f(x_i; \mu_r, \sigma_r, m_r, s_r, b_r)}$$

$$= \sum_{i=1}^{N} \frac{\tau_k \cdot \prod_{u \neq d}^{D} \left[\left(\frac{1}{2s_{k,u}} \exp\left(-\frac{|x_{i,u} - m_{k,u}|}{s_{k,u}} \right) \right)^{b_{k,u}} \cdot \left(\frac{1}{\sigma_{k,u}\sqrt{2\pi}} \exp\left(-\frac{(x_{i,u} - \mu_{k,u})^2}{2\sigma_{k,u}^2} \right) \right)^{(1-b_{k,u})} \right]}{\sum_{r=1}^{K} \tau_r \cdot f(x_i; \mu_r, \sigma_r, m_r, s_r, b_r)}$$

$$\cdot \left(\frac{1}{2s_{k,u}} \exp\left(-\frac{|x_{i,d} - m_{k,d}|}{s_{k,d}} \right) \right)^{b_{k,d}} \frac{\partial}{\partial \mu_{k,d}} \left[\left(\frac{1}{\sigma_{k,d}\sqrt{2\pi}} \exp\left(-\frac{(x_{i,d} - \mu_{k,d})^2}{2\sigma_{k,d}^2} \right) \right)^{(1-b_{k,d})} \right]$$

$$= \sum_{i=1}^{N} \frac{\tau_k \cdot f(x_i; \mu_k, \sigma_k, m_k, s_k, b_k)}{\sum_{r=1}^{K} \tau_r \cdot f(x_i; \mu_r, \sigma_r, m_r, s_r, b_r)} \cdot \frac{\partial}{\partial \mu_{k,d}} \left(-\frac{(x_{i,d} - \mu_{k,d})^2}{2\sigma_{k,d}^2} \right)$$

$$= \sum_{i=1}^{N} T_{k,i} \cdot \frac{x_{i,d} - \mu_{k,d}}{\sigma_{k,d}^2}$$

2.4.2 $\sigma_{k,d}$ entries

$\frac{\partial}{\partial \sigma_{k,d}} \mathcal{L}(\lambda|X) = 0$ if $b_{k,d} = 1$. Otherwise (i.e. $b_{k,d} = 0$):

$$\frac{\partial}{\partial \sigma_{k,d}} \mathcal{L}(\lambda|X) = \sum_{i=1}^{N} \frac{\frac{\partial}{\partial \sigma_{k,d}} \sum_{r=1}^{K} \tau_r \cdot f(x_i; \mu_r, \sigma_r, m_r, s_r, b_r)}{\sum_{r=1}^{K} \tau_r \cdot f(x_i; \mu_r, \sigma_r, m_r, s_r, b_r)}$$

$$= \sum_{i=1}^{N} \frac{\tau_k \cdot \frac{\partial}{\partial \sigma_{k,d}} f(x_i; \mu_k, \sigma_k, m_k, s_k, b_k)}{\sum_{r=1}^{K} \tau_r \cdot f(x_i; \mu_r, \sigma_r, m_r, s_r, b_r)}$$

$$= \sum_{i=1}^{N} \frac{\tau_k \cdot \frac{\partial}{\partial \sigma_{k,d}} \prod_{u=1}^{D} \left[\left(\frac{1}{2s_{k,u}} \exp\left(-\frac{|x_{i,u} - m_{k,u}|}{s_{k,u}} \right) \right)^{b_{k,u}} \cdot \left(\frac{1}{\sigma_{k,u}\sqrt{2\pi}} \exp\left(-\frac{(x_{i,u} - \mu_{k,u})^2}{2\sigma_{k,u}^2} \right) \right)^{(1-b_{k,u})} \right]}{\sum_{r=1}^{K} \tau_r \cdot f(x_i; \mu_r, \sigma_r, m_r, s_r, b_r)}$$

$$= \sum_{i=1}^{N} \frac{\tau_k \cdot \prod_{u \neq d}^{D} \left[\left(\frac{1}{2s_{k,u}} \exp\left(-\frac{|x_{i,u} - m_{k,u}|}{s_{k,u}} \right) \right)^{b_{k,u}} \cdot \left(\frac{1}{\sigma_{k,u}\sqrt{2\pi}} \exp\left(-\frac{(x_{i,u} - \mu_{k,u})^2}{2\sigma_{k,u}^2} \right) \right)^{(1-b_{k,u})} \right]}{\sum_{r=1}^{K} \tau_r \cdot f(x_i; \mu_r, \sigma_r, m_r, s_r, b_r)}$$

$$\cdot \left(\frac{1}{2s_{k,u}} \exp\left(-\frac{|x_{i,d} - m_{k,d}|}{s_{k,d}} \right) \right)^{b_{k,d}} \frac{\partial}{\partial \sigma_{k,d}} \left[\left(\frac{1}{\sigma_{k,d}\sqrt{2\pi}} \exp\left(-\frac{(x_{i,d} - \mu_{k,d})^2}{2\sigma_{k,d}^2} \right) \right)^{(1-b_{k,d})} \right]$$

$$= \sum_{i=1}^{N} \frac{\tau_k \cdot \prod_{u \neq d}^{D} \left[\left(\frac{1}{2s_{k,u}} \exp\left(-\frac{|x_{i,u} - m_{k,u}|}{s_{k,u}} \right) \right)^{b_{k,u}} \cdot \left(\frac{1}{\sigma_{k,u}\sqrt{2\pi}} \exp\left(-\frac{(x_{i,u} - \mu_{k,u})^2}{2\sigma_{k,u}^2} \right) \right)^{(1-b_{k,u})} \right]}{\sum_{r=1}^{K} \tau_r \cdot f(x_i; \mu_r, \sigma_r, m_r, s_r, b_r)}$$

$$\cdot \left(\frac{1}{2s_{k,u}} \exp\left(-\frac{|x_{i,d} - m_{k,d}|}{s_{k,d}} \right) \right)^{b_{k,d}} \left[\frac{1}{\sigma_{k,d}\sqrt{2\pi}} \exp\left(-\frac{(x_{i,d} - \mu_{k,d})^2}{2\sigma_{k,d}^2} \right) \left(\frac{(x_{i,d} - \mu_{k,d})^2}{\sigma_{k,d}^3} \right) \right.$$

$$\left. - \frac{1}{\sigma_{k,d}^2\sqrt{2\pi}} \exp\left(-\frac{(x_{i,d} - \mu_{k,d})^2}{2\sigma_{k,d}^2} \right) \right]$$

$$= \sum_{i=1}^{N} \frac{\tau_k \cdot f(x_i; \mu_k, \sigma_k, m_k, s_k, b_k)}{\sum_{r=1}^{K} \tau_r \cdot f(x_i; \mu_r, \sigma_r, m_r, s_r, b_r)} \left(\frac{(x_{i,d} - \mu_{k,d})^2}{\sigma_{k,d}^3} - \frac{1}{\sigma_{k,d}} \right)$$

$$= \sum_{i=1}^{N} T_{k,i} \left(\frac{(x_{i,d} - \mu_{k,d})^2}{\sigma_{k,d}^3} - \frac{1}{\sigma_{k,d}} \right)$$

2.4.3 $m_{k,d}$ entries

$\frac{\partial}{\partial m_{k,d}}\mathcal{L}\left(\lambda|X\right)=0$ if $b_{k,d}=0$. Otherwise (i.e. $b_{k,d}=1$):

$$
\begin{aligned}
&\frac{\partial}{\partial m_{k,d}}\mathcal{L}\left(\lambda|X\right)=\sum_{i=1}^{N}\frac{\frac{\partial}{\partial m_{k,d}}\sum_{r=1}^{K}\tau_r\cdot f\left(x_i;\mu_r,\sigma_r,m_r,s_r,b_r\right)}{\sum_{r=1}^{K}\tau_r\cdot f\left(x_i;\mu_r,\sigma_r,m_r,s_r,b_r\right)}\\
&=\sum_{i=1}^{N}\frac{\tau_k\cdot\frac{\partial}{\partial m_{k,d}}f\left(x_i;\mu_k,\sigma_k,m_k,s_k,b_k\right)}{\sum_{r=1}^{K}\tau_r\cdot f\left(x_i;\mu_r,\sigma_r,m_r,s_r,b_r\right)}\\
&=\sum_{i=1}^{N}\frac{\tau_k\cdot\frac{\partial}{\partial m_{k,d}}\prod_{u=1}^{D}\left[\left(\frac{1}{2s_{k,u}}\exp\left(-\frac{\left|x_{i,u}-m_{k,u}\right|}{s_{k,u}}\right)\right)^{b_{k,u}}\cdot\left(\frac{1}{\sigma_{k,u}\sqrt{2\pi}}\exp\left(-\frac{\left(x_{i,u}-m_{k,u}\right)^2}{2\sigma_{k,u}^2}\right)\right)^{\left(1-b_{k,u}\right)}\right]}{\sum_{r=1}^{K}\tau_r\cdot f\left(x_i;\mu_r,\sigma_r,m_r,s_r,b_r\right)}\\
&=\sum_{i=1}^{N}\frac{\tau_k\cdot\prod_{u\neq d}^{D}\left[\left(\frac{1}{2s_{k,u}}\exp\left(-\frac{\left|x_{i,u}-m_{k,u}\right|}{s_{k,u}}\right)\right)^{b_{k,u}}\cdot\left(\frac{1}{\sigma_{k,u}\sqrt{2\pi}}\exp\left(-\frac{\left(x_{i,u}-\mu_{k,u}\right)^2}{2\sigma_{k,u}^2}\right)\right)^{\left(1-b_{k,u}\right)}\right]}{\sum_{r=1}^{K}\tau_r\cdot f\left(x_i;\mu_r,\sigma_r,m_r,s_r,b_r\right)}\\
&\cdot\left(\frac{1}{\sigma_{k,d}\sqrt{2\pi}}\exp\left(-\frac{\left(x_{i,d}-\mu_{k,d}\right)^2}{2\sigma_{k,d}^2}\right)\right)^{\left(1-b_{k,d}\right)}\frac{\partial}{\partial m_{k,d}}\left[\left(\frac{1}{2s_{k,u}}\exp\left(-\frac{\left|x_{i,d}-m_{k,d}\right|}{s_{k,d}}\right)\right)^{b_{k,d}}\right]\\
&=\sum_{i=1}^{N}\frac{\tau_k\cdot f\left(x_i;\mu_k,\sigma_k,m_k,s_k,b_k\right)}{\sum_{r=1}^{K}\tau_r\cdot f\left(x_i;\mu_r,\sigma_r,m_r,s_r,b_r\right)}\cdot\frac{\partial}{\partial m_{k,d}}\left(-\frac{\left|x_{i,d}-m_{k,d}\right|}{s_{k,d}}\right)\\
&=\sum_{i=1}^{N}\frac{T_{k,i}}{s_{k,d}}\cdot\begin{cases}1 & \text{if } x_{i,d}>m_{k,d}\\ -1 & \text{o.w.}\end{cases}
\end{aligned}
$$

2.4.4 $s_{k,d}$ entries

$\frac{\partial}{\partial s_{k,d}}\mathcal{L}\left(\lambda|X\right)=0$ if $b_{k,d}=0$. Otherwise (i.e. $b_{k,d}=1$):

$$\frac{\partial}{\partial s_{k,d}}\mathcal{L}\left(\lambda|X\right)=\sum_{i=1}^{N}\frac{\frac{\partial}{\partial s_{k,d}}\sum_{r=1}^{K}\tau_r\cdot f\left(x_i;\mu_r,\sigma_r,m_r,s_r,b_r\right)}{\sum_{r=1}^{K}\tau_r\cdot f\left(x_i;\mu_r,\sigma_r,m_r,s_r,b_r\right)}$$

$$=\sum_{i=1}^{N}\frac{\tau_k\cdot\frac{\partial}{\partial s_{k,d}}f\left(x_i;\mu_k,\sigma_k,m_k,s_k,b_k\right)}{\sum_{r=1}^{K}\tau_r\cdot f\left(x_i;\mu_r,\sigma_r,m_r,s_r,b_r\right)}$$

$$=\sum_{i=1}^{N}\frac{\tau_k\cdot\frac{\partial}{\partial s_{k,d}}\prod_{u=1}^{D}\left[\left(\frac{1}{2s_{k,u}}\exp\left(-\frac{\left|x_{i,u}-m_{k,u}\right|}{s_{k,u}}\right)\right)^{b_{k,u}}\cdot\left(\frac{1}{\sigma_{k,u}\sqrt{2\pi}}\exp\left(-\frac{\left(x_{i,u}-\mu_{k,u}\right)^2}{2\sigma_{k,u}^2}\right)\right)^{\left(1-b_{k,u}\right)}\right]}{\sum_{r=1}^{K}\tau_r\cdot f\left(x_i;\mu_r,\sigma_r,m_r,s_r,b_r\right)}$$

$$=\sum_{i=1}^{N}\frac{\tau_k\cdot\prod_{u\neq d}^{D}\left[\left(\frac{1}{2s_{k,u}}\exp\left(-\frac{\left|x_{i,u}-m_{k,u}\right|}{s_{k,u}}\right)\right)^{b_{k,u}}\cdot\left(\frac{1}{\sigma_{k,u}\sqrt{2\pi}}\exp\left(-\frac{\left(x_{i,u}-\mu_{k,u}\right)^2}{2\sigma_{k,u}^2}\right)\right)^{\left(1-b_{k,u}\right)}\right]}{\sum_{r=1}^{K}\tau_r\cdot f\left(x_i;\mu_r,\sigma_r,m_r,s_r,b_r\right)}$$

$$\cdot\left(\frac{1}{\sigma_{k,d}\sqrt{2\pi}}\exp\left(-\frac{\left(x_{i,d}-\mu_{k,d}\right)^2}{2\sigma_{k,d}^2}\right)\right)^{\left(1-b_{k,d}\right)}\frac{\partial}{\partial s_{k,d}}\left[\left(\frac{1}{2s_{k,u}}\exp\left(-\frac{\left|x_{i,d}-m_{k,d}\right|}{s_{k,d}}\right)\right)^{b_{k,d}}\right]$$

$$=\sum_{i=1}^{N}\frac{\tau_k\cdot\prod_{u\neq d}^{D}\left[\left(\frac{1}{2s_{k,u}}\exp\left(-\frac{\left|x_{i,u}-m_{k,u}\right|}{s_{k,u}}\right)\right)^{b_{k,u}}\cdot\left(\frac{1}{\sigma_{k,u}\sqrt{2\pi}}\exp\left(-\frac{\left(x_{i,u}-\mu_{k,u}\right)^2}{2\sigma_{k,u}^2}\right)\right)^{\left(1-b_{k,u}\right)}\right]}{\sum_{r=1}^{K}\tau_r\cdot f\left(x_i;\mu_r,\sigma_r,m_r,s_r,b_r\right)}$$

$$\cdot\left(\frac{1}{\sigma_{k,d}\sqrt{2\pi}}\exp\left(-\frac{\left(x_{i,d}-\mu_{k,d}\right)^2}{2\sigma_{k,d}^2}\right)\right)^{\left(1-b_{k,d}\right)}\left[\frac{1}{2s_{k,d}}\exp\left(-\frac{\left|x_{i,d}-m_{k,d}\right|}{s_{k,d}}\right)\left(\frac{\left|x_{i,d}-m_{k,d}\right|}{s_{k,d}^2}\right)\right.$$

$$\left.-\frac{1}{2s_{k,d}^2}\exp\left(-\frac{\left|x_{i,d}-m_{k,d}\right|}{s_{k,d}}\right)\right]$$

$$=\sum_{i=1}^{N}\frac{\tau_k\cdot f\left(x_i;\mu_k,\sigma_k,m_k,s_k,b_k\right)}{\sum_{r=1}^{K}\tau_r\cdot f\left(x_i;\mu_r,\sigma_r,m_r,s_r,b_r\right)}\left(\frac{\left|x_{i,d}-m_{k,d}\right|}{s_{k,d}^2}-\frac{1}{s_{k,d}}\right)$$

$$=\sum_{i=1}^{N}T_{k,i}\left(\frac{\left|x_{i,d}-m_{k,d}\right|}{s_{k,d}^2}-\frac{1}{s_{k,d}}\right)$$

2.5 Derivation of the Fisher Information Matrix

Similarly to [1], our derivations are based on two assumptions. The first one is that the number of low-level features x_i extracted from each sample is constant and equal to N. The second one is that for each observation x_i, the distribution of the occupancy probability $T_{k,i}$ is sharply peaked. This means that there is one index k such that $T_{k,i} \approx 1$ and that $\forall r \neq k, T_{r,i} = 0$. This second property is based on empirical observation. Therefore, $T_{k,i}^2 \approx T_{k,i}$.

For convenience let $H_{b_{k,d}}(x) = \begin{cases} G(x|\mu_{k,d}, \sigma_{k,d}) & \text{if } b_{k,d} = 1 \\ L(x|m_{k,d}, s_{k,d}) & \text{o.w.} \end{cases}$
Where G is the univariate Gaussian distribution and L is the univariate Laplacian distribution.

Let us denote by $f_{\mu_{k,d}}$, $f_{\sigma_{k,d}}$, $f_{m_{k,d}}$ and $f_{s_{k,d}}$ the terms on the diagonal of the Fisher information matrix F which correspond respectively to $\frac{\mathcal{L}(X|\lambda)}{\partial \mu_{k,d}}$, $\frac{\mathcal{L}(X|\lambda)}{\partial \sigma_{k,d}}$, $\frac{\mathcal{L}(X|\lambda)}{\partial m_{k,d}}$ and $\frac{\mathcal{L}(X|\lambda)}{\partial s_{k,d}}$.

Our derivations follow Appendix A in [1].

2.5.1 $F_{\mu_{k,d}}$

The normalization is only interesting if $b_{k,d} = 0$
Using the definition of the fisher kernel and the value of $\frac{\partial \mathcal{L}(Y|\lambda)}{\partial \mu_{k,d}}$ we get:

$$f_{\mu_{k,d}} = \int_Y p\left(Y|\lambda\right) \left[\sum_{i=1}^N \frac{\partial \mathcal{L}(y_i|\lambda)}{\partial \mu_{k,d}}\right]^2 dY$$

Let $A_i\left(k,d\right) = \frac{\partial \mathcal{L}(y_i|\lambda)}{\partial \mu_{k,d}} = T_{k,d}\left(\frac{y_{i,d}-\mu_{k,d}}{\sigma_{k,d}^2}\right)$ then:

$$f_{\mu_{k,d}} = \sum_{\substack{t=1\ldots N\\u=1\ldots N\\t\neq u}} \int_{y_t,y_u} A_t(k,d)A_u(k,d)p\left(y_t,y_u|\lambda\right)dy_t dy_u + \sum_{t=1}^N \int_{y_t} A_t(k,d)^2 p\left(y_t|\lambda\right)dy_t$$

For $t \neq u$, using independence of y_t and y_u, we get:
$\int_{y_t,y_u} A_t(k,d)A_u(k,d)p\left(y_t,y_u|\lambda\right)dy_t dy_u = \int_{y_t} A_t(k,d)p\left(y_t|\lambda\right)dy_t \cdot \int_{y_u} A_u(k,d)p\left(y_u|\lambda\right)dy_u$

Using the definition of $T_{k,i}$ we get:
$A_t(k,d)p\left(y_t|\lambda\right) = T_{k,t}\left(\frac{y_{t,d}-\mu_{k,d}}{\sigma_{k,d}^2}\right)p\left(y_t|\lambda\right) = \frac{\tau_k p_k(y_t|\lambda)}{p(y_t|\lambda)}\left(\frac{y_{t,d}-\mu_{k,d}}{\sigma_{k,d}^2}\right)p\left(y_t|\lambda\right) = \tau_k p_k(y_t|\lambda)\left(\frac{y_{t,d}-\mu_{k,d}}{\sigma_{k,d}^2}\right)$

Therefore:

$$\begin{aligned}
\int_{y_t} A_t(k,d)p\left(y_t|\lambda\right)dy_t &= \int_{y_t} \tau_k p_k(y_t|\lambda)\left(\frac{y_{t,d}-\mu_{k,d}}{\sigma_{k,d}^2}\right)dy_t \\
&= \tau_k \cdot \left[\prod_{c\neq d}\int_{y_{t,c}} p_{k,d}(y_{t,c}|\lambda)dy_{t,c}\right]\int_{y_{t,d}} p_{k,d}(y_{t,d}|\lambda)\left(\frac{y_{t,d}-\mu_{k,d}}{\sigma_{k,d}^2}\right)dy_{t,d} \\
&= \tau_k \cdot \left[\prod_{c\neq d}\int_{y_{t,c}} H_{b_{k,d}}\left(y_{t,c}\right)dy_{t,c}\right]\int_{y_{t,d}} p_{k,d}(y_{t,d}|\lambda)\left(\frac{y_{t,d}-\mu_{k,d}}{\sigma_{k,d}^2}\right)dy_{t,d} \\
&= \tau_k \cdot \int_{y_{t,d}} p_{k,d}(y_{t,d}|\lambda)\left(\frac{y_{t,d}-\mu_{k,d}}{\sigma_{k,d}^2}\right)dy_{t,d} \\
&= \frac{\tau_k\sqrt{2}}{\sigma_{k,d}^2\sqrt{2\pi}} \cdot \int_{y_{t,d}} \exp\left(-\left(\frac{y_{t,d}-\mu_{k,d}}{\sqrt{2}\sigma_{k,d}}\right)^2\right)\left(\frac{y_{t,d}-\mu_{k,d}}{\sqrt{2}\sigma_{k,d}}\right)dy_{t,d} \\
&= \frac{\tau_k\sqrt{2}\sigma_{k,d}}{\sigma_{k,d}^2}\sqrt{\pi} \cdot \int_v \exp\left(-v^2\right)v\,dv \\
&= 0
\end{aligned}$$

$A_t(k,d)^2 p\left(y_t|\lambda\right) = T_{k,t}^2 \left(\frac{y_{t,d}-\mu_{k,d}}{\sigma_{k,d}^2}\right)^2 p\left(y_t|\lambda\right) \approx T_{k,t} \left(\frac{y_{t,d}-\mu_{k,d}}{\sigma_{k,d}^2}\right)^2 p\left(y_t|\lambda\right) = \tau_k p_k(y_t|\lambda) \left(\frac{y_{t,d}-\mu_{k,d}}{\sigma_{k,d}^2}\right)^2$

Therefore:

$$
\begin{aligned}
\int_{y_t} A_t(k,d)^2 p\left(y_t|\lambda\right) dy_t &= \int_{y_t} \tau_k p_k(y_t|\lambda) \left(\frac{y_{t,d}-\mu_{k,d}}{\sigma_{k,d}^2}\right)^2 dy_t \\
&= \tau_k \cdot \left[\prod_{c\neq d} \int_{y_{t,c}} p_{k,d}(y_{t,c}|\lambda) dy_{t,c}\right] \int_{y_{t,d}} p_{k,d}(y_{t,d}|\lambda) \left(\frac{y_{t,d}-\mu_{k,d}}{\sigma_{k,d}^2}\right)^2 dy_{t,d} \\
&= \tau_k \cdot \left[\prod_{c\neq d} \int_{y_{t,c}} H_{b_{k,d}}\left(y_{t,c}\right) dy_{t,c}\right] \int_{y_{t,d}} p_{k,d}(y_{t,d}|\lambda) \left(\frac{y_{t,d}-\mu_{k,d}}{\sigma_{k,d}^2}\right)^2 dy_{t,d} \\
&= \tau_k \cdot \int_{y_{t,d}} p_{k,d}(y_{t,d}|\lambda) \left(\frac{y_{t,d}-\mu_{k,d}}{\sigma_{k,d}^2}\right)^2 dy_{t,d} \\
&= \frac{2\tau_k}{\sigma_{k,d}^3\sqrt{2\pi}} \cdot \int_{y_{t,d}} \exp\left(-\left(\frac{y_{t,d}-\mu_{k,d}}{\sqrt{2}\sigma_{k,d}}\right)^2\right) \left(\frac{y_{t,d}-\mu_{k,d}}{\sqrt{2}\sigma_{k,d}}\right)^2 dy_{t,d} \\
&= \frac{2\sqrt{2}\tau_k\sigma_{k,d}}{\sigma_{k,d}^3\sqrt{2\pi}} \cdot \int_v \exp\left(-v^2\right) v^2 dv \\
&= \frac{2\tau_k}{\sigma_{k,d}^2\sqrt{\pi}} \frac{\sqrt{\pi}}{2} \\
&= \frac{\tau_k}{\sigma_{k,d}^2}
\end{aligned}
$$

Finally: $f_{\mu_{k,d}} = 0 + N \cdot \frac{\tau_k}{\sigma_{k,d}^2} = \frac{M\tau_k}{\sigma_{k,d}^2}$

2.5.2 $F_{\sigma_{k,d}}$

The normalization is only interesting if $b_{k,d} = 0$
Using the definition of the fisher kernel and the value of $\frac{\partial \mathcal{L}(Y|\lambda)}{\partial \sigma_{k,d}}$ we get:

$$f_{\sigma_{k,d}} = \int_Y p\left(Y|\lambda\right) \left[\sum_{i=1}^N \frac{\partial \mathcal{L}(y_i|\lambda)}{\partial \sigma_{k,d}}\right]^2 dY$$

Let $A_i\left(k,d\right) = \frac{\partial \mathcal{L}(y_i|\lambda)}{\partial \sigma_{k,d}} = T_{k,d}\left(\frac{\left(y_{i,d}-\mu_{k,d}\right)^2}{\sigma_{k,d}^3} - \frac{1}{\sigma_{k,d}}\right)$ then:

$$f_{\sigma_{k,d}} = \sum_{\substack{t=1\ldots N\\u=1\ldots N\\t\neq u}} \int_{y_t,y_u} A_t(k,d)A_u(k,d)p\left(y_t,y_u|\lambda\right)dy_t dy_u + \sum_{t=1}^N \int_{y_t} A_t(k,d)^2 p\left(y_t|\lambda\right)dy_t$$

For $t \neq u$, using independence of y_t and y_u, we get:
$\int_{y_t,y_u} A_t(k,d)A_u(k,d)p\left(y_t,y_u|\lambda\right)dy_t dy_u = \int_{y_t} A_t(k,d)p\left(y_t|\lambda\right)dy_t \cdot \int_{y_u} A_u(k,d)p\left(y_u|\lambda\right)dy_u$

Using the definition of $T_{k,i}$ we get:
$A_t(k,d)p\left(y_t|\lambda\right) = T_{k,t}\left(\frac{\left(y_{i,d}-\mu_{k,d}\right)^2}{\sigma_{k,d}^3} - \frac{1}{\sigma_{k,d}}\right)p\left(y_t|\lambda\right) = \frac{\tau_k p_k(y_t|\lambda)}{p(y_t|\lambda)}\left(\frac{\left(y_{i,d}-\mu_{k,d}\right)^2}{\sigma_{k,d}^3} - \frac{1}{\sigma_{k,d}}\right)p\left(y_t|\lambda\right) = \tau_k p_k(y_t|\lambda)\left(\frac{\left(y_{i,d}-\mu_{k,d}\right)^2}{\sigma_{k,d}^3} - \frac{1}{\sigma_{k,d}}\right)$

Therefore:

$$\begin{aligned}
\int_{y_t} A_t(k,d)p\left(y_t|\lambda\right)dy_t &= \int_{y_t} \tau_k p_k(y_t|\lambda)\left(\frac{\left(y_{i,d}-\mu_{k,d}\right)^2}{\sigma_{k,d}^3} - \frac{1}{\sigma_{k,d}}\right)dy_t \\
&= \tau_k \cdot \left[\prod_{c\neq d}\int_{y_{t,c}} p_{k,d}(y_{t,c}|\lambda)dy_{t,c}\right]\int_{y_{t,d}} p_{k,d}(y_{t,d}|\lambda)\left(\frac{\left(y_{i,d}-\mu_{k,d}\right)^2}{\sigma_{k,d}^3} - \frac{1}{\sigma_{k,d}}\right)dy_{t,d} \\
&= \tau_k \cdot \left[\prod_{c\neq d}\int_{y_{t,c}} H_{b_{k,d}}\left(y_{t,c}\right)dy_{t,c}\right]\int_{y_{t,d}} p_{k,d}(y_{t,d}|\lambda)\left(\frac{\left(y_{i,d}-\mu_{k,d}\right)^2}{\sigma_{k,d}^3} - \frac{1}{\sigma_{k,d}}\right)dy_{t,d} \\
&= \tau_k \cdot \int_{y_{t,d}} p_{k,d}(y_{t,d}|\lambda)\left(\frac{\left(y_{i,d}-\mu_{k,d}\right)^2}{\sigma_{k,d}^3} - \frac{1}{\sigma_{k,d}}\right)dy_{t,d} \\
&= \frac{2\tau_k}{\sigma_{k,d}} \cdot \int_{y_{t,d}} p_{k,d}(y_{t,d}|\lambda)\left(\frac{y_{i,d}-\mu_{k,d}}{\sqrt{2}\sigma_{k,d}}\right)^2 dy_{t,d} - \frac{\tau_k}{\sigma_{k,d}} \\
&= \frac{2\tau_k}{\sigma_{k,d}^2\sqrt{2\pi}} \cdot \int_{y_{t,d}} \exp\left(-\left(\frac{y_{t,d}-\mu_{k,d}}{\sqrt{2}\sigma_{k,d}}\right)^2\right)\left(\frac{y_{t,d}-\mu_{k,d}}{\sqrt{2}\sigma_{k,d}}\right)dy_{t,d} - \frac{\tau_k}{\sigma_{k,d}} \\
&= \frac{2\sqrt{2}\tau_k\sigma_{k,d}}{\sigma_{k,d}^2\sqrt{2\pi}} \cdot \int_v \exp\left(-v^2\right)v dv - \frac{\tau_k}{\sigma_{k,d}} \\
&= \frac{2\sqrt{2}\tau_k\sigma_{k,d}}{\sigma_{k,d}^2\sqrt{2\pi}} \cdot \frac{\sqrt{\pi}}{2} - \frac{\tau_k}{\sigma_{k,d}} \\
&= 0
\end{aligned}$$

$A_t(k,d)^2 p\left(y_t|\lambda\right) = T_{k,t}^2 \left(\frac{\left(y_{i,d}-\mu_{k,d}\right)^2}{\sigma_{k,d}^3} - \frac{1}{\sigma_{k,d}}\right)^2 p\left(y_t|\lambda\right) \approx T_{k,t} \left(\frac{\left(y_{i,d}-\mu_{k,d}\right)^2}{\sigma_{k,d}^3} - \frac{1}{\sigma_{k,d}}\right)^2 p\left(y_t|\lambda\right) =$
$\tau_k p_k(y_t|\lambda) \left(\frac{\left(y_{i,d}-\mu_{k,d}\right)^4}{\sigma_{k,d}^6} - 2\frac{\left(y_{i,d}-\mu_{k,d}\right)^2}{\sigma_{k,d}^4} + \frac{1}{\sigma_{k,d}^2}\right)$
Therefore:

$$\begin{aligned}
\int_{y_t} A_t(k,d)^2 p\left(y_t|\lambda\right) dy_t &= \int_{y_t} \tau_k p_k(y_t|\lambda) A_t(k,d)^2 dy_t \\
&= \tau_k \cdot \left[\prod_{c\neq d} \int_{y_{t,c}} p_{k,d}(y_{t,c}|\lambda) dy_{t,c}\right] \int_{y_{t,d}} p_{k,d}(y_{t,d}|\lambda) A_t(k,d)^2 dy_{t,d} \\
&= \tau_k \cdot \left[\prod_{c\neq d} \int_{y_{t,c}} H_{b_{k,d}}\left(y_{t,c}\right) dy_{t,c}\right] \int_{y_{t,d}} p_{k,d}(y_{t,d}|\lambda) A_t(k,d)^2 dy_{t,d} \\
&= \tau_k \cdot \int_{y_{t,d}} p_{k,d}(y_{t,d}|\lambda) \left(\frac{(y_{i,d}-\mu_{k,d})^4}{\sigma_{k,d}^6} - 2\frac{(y_{i,d}-\mu_{k,d})^2}{\sigma_{k,d}^4} + \frac{1}{\sigma_{k,d}^2}\right) dy_{t,d}
\end{aligned}$$

We'll split this expression to three parts:
1)

$$\begin{aligned}
&= \tau_k \cdot \int_{y_{t,d}} p_{k,d}(y_{t,d}|\lambda) \frac{1}{\sigma_{k,d}^2} dy_{t,d} \\
&= \frac{\tau_k}{\sigma_{k,d}^2}
\end{aligned}$$

2)

$$
\begin{aligned}
&= \tau_k \cdot \int_{y_{t,d}} p_{k,d}(y_{t,d}|\lambda) \frac{(y_{i,d} - \mu_{k,d})^4}{\sigma_{k,d}^6} dy_{t,d} \\
&= \frac{\tau_k}{\sigma_{k,d}\sqrt{2\pi}} \cdot \int_{y_{t,d}} \exp\left(-\left(\frac{y_{t,d} - \mu_{k,d}}{\sqrt{2}\sigma_{k,d}}\right)^2\right) \frac{(y_{i,d} - \mu_{k,d})^4}{\sigma_{k,d}^6} dy_{t,d} \\
&= \frac{4\tau_k}{\sigma_{k,d}^3\sqrt{2\pi}} \cdot \int_{y_{t,d}} \exp\left(-\left(\frac{y_{t,d} - \mu_{k,d}}{\sqrt{2}\sigma_{k,d}}\right)^2\right) \left(\frac{y_{i,d} - \mu_{k,d}}{\sqrt{2}\sigma_{k,d}}\right)^4 dy_{t,d} \\
&= \frac{4\sqrt{2}\tau_k}{\sigma_{k,d}^2\sqrt{2\pi}} \int_v \exp\left(-v^2\right) v^4 dv \\
&= \frac{4\sqrt{2}\tau_k}{\sigma_{k,d}^2\sqrt{2\pi}} \frac{3\sqrt{\pi}}{4} \\
&= \frac{3\tau_k}{\sigma_{k,d}^2}
\end{aligned}
$$

3)

$$
\begin{aligned}
&= -2\tau_k \cdot \int_{y_{t,d}} p_{k,d}(y_{t,d}|\lambda) \frac{(y_{i,d} - \mu_{k,d})^2}{\sigma_{k,d}^4} dy_{t,d} \\
&= \frac{-2\tau_k}{\sigma_{k,d}\sqrt{2\pi}} \cdot \int_{y_{t,d}} \exp\left(-\left(\frac{y_{t,d} - \mu_{k,d}}{\sqrt{2}\sigma_{k,d}}\right)^2\right) \frac{(y_{i,d} - \mu_{k,d})^2}{\sigma_{k,d}^4} dy_{t,d} \\
&= \frac{-4\tau_k}{\sigma_{k,d}^3\sqrt{2\pi}} \cdot \int_{y_{t,d}} \exp\left(-\left(\frac{y_{t,d} - \mu_{k,d}}{\sqrt{2}\sigma_{k,d}}\right)^2\right) \left(\frac{y_{i,d} - \mu_{k,d}}{\sqrt{2}\sigma_{k,d}}\right)^2 dy_{t,d} \\
&= \frac{-4\sqrt{2}\tau_k}{\sigma_{k,d}^2\sqrt{2\pi}} \int_v \exp\left(-v^2\right) v^2 dv \\
&= \frac{-4\sqrt{2}\tau_k}{\sigma_{k,d}^2\sqrt{2\pi}} \frac{\sqrt{\pi}}{2} \\
&= \frac{-2\tau_k}{\sigma_{k,d}^2}
\end{aligned}
$$

Finally: $f_{\sigma_{k,d}} = 0 + N \cdot \left(\frac{\tau_k}{\sigma_{k,d}^2} - \frac{2\tau_k}{\sigma_{k,d}^2} + \frac{3\tau_k}{\sigma_{k,d}^2}\right) = \frac{2N\tau_k}{\sigma_{k,d}^2}$

2.5.3 $F_{m_{k,d}}$

The normalization is only interesting if $b_{k,d} = 1$

Using the definition of the fisher kernel and the value of $\frac{\partial \mathcal{L}(Y|\lambda)}{\partial m_{k,d}}$ we get:

$$f_{m_{k,d}} = \int_Y p\left(Y|\lambda\right)\left[\sum_{i=1}^N \frac{\partial \mathcal{L}(y_i|\lambda)}{\partial m_{k,d}}\right]^2 dY$$

Let $A_i\left(k,d\right) = \frac{\partial \mathcal{L}(y_i|\lambda)}{\partial m_{k,d}} = \frac{T_{k,d}}{s_{k,d}} \cdot \begin{cases} 1 & \text{if } y_{t,d} > m_{k,d} \\ -1 & \text{o.w.} \end{cases}$ then:

$$f_{m_{k,d}} = \sum_{\substack{t=1\ldots N \\ u=1\ldots N \\ t\neq u}} \int_{y_t,y_u} A_t(k,d)A_u(k,d)p\left(y_t,y_u|\lambda\right)dy_t dy_u + \sum_{t=1}^N \int_{y_t} A_t(k,d)^2 p\left(y_t|\lambda\right)dy_t$$

For $t \neq u$, using independence of y_t and y_u, we get:

$$\int_{y_t,y_u} A_t(k,d)A_u(k,d)p\left(y_t,y_u|\lambda\right)dy_t dy_u = \int_{y_t} A_t(k,d)p\left(y_t|\lambda\right)dy_t \cdot \int_{y_u} A_u(k,d)p\left(y_u|\lambda\right)dy_u$$

Using the definition of $T_{k,i}$ we get:

$A_t(k,d)p\left(y_t|\lambda\right) = \frac{T_{k,t}}{s_{k,d}} \cdot \begin{cases} 1 & \text{if } y_{t,d} > m_{k,d} \\ -1 & \text{o.w.} \end{cases} p\left(y_t|\lambda\right) = \frac{\tau_k p_k(y_t|\lambda)}{p(y_t|\lambda)s_{k,d}} \begin{cases} 1 & \text{if } y_{t,d} > m_{k,d} \\ -1 & \text{o.w.} \end{cases} p\left(y_t|\lambda\right) = \tau_k p_k(y_t|\lambda) \begin{cases} 1 & \text{if } y_{t,d} > m_{k,d} \\ -1 & \text{o.w.} \end{cases}$

Therefore:

$$\begin{aligned}
\int_{y_t} A_t(k,d)p\left(y_t|\lambda\right)dy_t &= \int_{y_t} \frac{\tau_k}{s_{k,d}} p_k(y_t|\lambda) \begin{cases} 1 & \text{if } y_{t,d} > m_{k,d} \\ -1 & \text{o.w.} \end{cases} dy_t \\
&= \frac{\tau_k}{s_{k,d}} \cdot \left[\prod_{c\neq d} \int_{y_{t,c}} p_{k,d}(y_{t,c}|\lambda)dy_{t,c}\right] \int_{y_{t,d}} p_{k,d}(y_{t,d}|\lambda) \begin{cases} 1 & \text{if } y_{t,d} > m_{k,d} \\ -1 & \text{o.w.} \end{cases} dy_{t,d} \\
&= \frac{\tau_k}{s_{k,d}} \cdot \left[\prod_{c\neq d} \int_{y_{t,c}} H_{b_{k,d}}\left(y_{t,c}\right)dy_{t,c}\right] \int_{y_{t,d}} p_{k,d}(y_{t,d}|\lambda) \begin{cases} 1 & \text{if } y_{t,d} > m_{k,d} \\ -1 & \text{o.w.} \end{cases} dy_{t,d} \\
&= \frac{\tau_k}{s_{k,d}} \cdot \int_{y_{t,d}} p_{k,d}(y_{t,d}|\lambda) \begin{cases} 1 & \text{if } y_{t,d} > m_{k,d} \\ -1 & \text{o.w.} \end{cases} dy_{t,d} \\
&= \frac{\tau_k}{2s_{k,d}^2} \cdot \int_{y_{t,d}} \exp\left(-\frac{|y_{t,d} - m_{k,d}|}{s_{k,d}}\right) \begin{cases} 1 & \text{if } y_{t,d} > m_{k,d} \\ -1 & \text{o.w.} \end{cases} dy_{t,d} \\
&= \frac{\tau_k}{2s_{k,d}^2} \left[-\int_{y_{t,d}<m_{k,d}} \exp\left(-\frac{|y_{t,d} - m_{k,d}|}{s_{k,d}}\right) + \int_{y_{t,d}>m_{k,d}} \exp\left(-\frac{|y_{t,d} - m_{k,d}|}{s_{k,d}}\right)\right] \\
&= 0
\end{aligned}$$

$A_t(k,d)^2 p\left(y_t|\lambda\right) = \frac{T_{k,t}^2}{s_{k,d}^2}\left[\begin{cases} 1 & \text{if } y_{t,d} > m_{k,d} \\ -1 & \text{o.w.} \end{cases}\right]^2 p\left(y_t|\lambda\right) \approx \frac{T_{k,t}}{s_{k,d}^2} p\left(y_t|\lambda\right) = \frac{\tau_k}{s_{k,d}^2} p_k(y_t|\lambda)$

Therefore:

$$\begin{aligned}
\int_{y_t} A_t(k,d)^2 p\,(y_t|\lambda)\,dy_t &= \int_{y_t} \frac{\tau_k}{s_{k,d}^2} p_k(y_t|\lambda) dy_t \\
&= \frac{\tau_k}{s_{k,d}^2} \cdot \left[\prod_{c \neq d} \int_{y_{t,c}} p_{k,d}(y_{t,c}|\lambda) dy_{t,c}\right] \int_{y_{t,d}} p_{k,d}(y_{t,d}|\lambda) dy_{t,d} \\
&= \frac{\tau_k}{s_{k,d}^2}
\end{aligned}$$

Finally: $f_{m_{k,d}} = 0 + N \cdot \frac{\tau_k}{s_{k,d}^2} = \frac{N\tau_k}{s_{k,d}^2}$

2.5.4 $F_{s_{k,d}}$

The normalization is only interesting if $b_{k,d} = 1$

Using the definition of the fisher kernel and the value of $\frac{\partial \mathcal{L}(Y|\lambda)}{\partial s_{k,d}}$ we get:

$$f_{s_{k,d}} = \int_Y p\left(Y|\lambda\right) \left[\sum_{i=1}^{N} \frac{\partial \mathcal{L}(y_i|\lambda)}{\partial s_{k,d}}\right]^2 dY$$

Let $A_i\left(k,d\right) = \frac{\partial \mathcal{L}(y_i|\lambda)}{\partial s_{k,d}} = T_{k,d}\left(\frac{\left|y_{i,d}-m_{k,d}\right|}{s_{k,d}^2} - \frac{1}{s_{k,d}}\right)$ then:

$$f_{s_{k,d}} = \sum_{\substack{t=1\ldots N\\ u=1\ldots N\\ t\neq u}} \int_{y_t,y_u} A_t(k,d)A_u(k,d)p\left(y_t,y_u|\lambda\right) dy_t dy_u + \sum_{t=1}^{N} \int_{y_t} A_t(k,d)^2 p\left(y_t|\lambda\right) dy_t$$

For $t \neq u$, using independence of y_t and y_u, we get:

$$\int_{y_t,y_u} A_t(k,d)A_u(k,d)p\left(y_t,y_u|\lambda\right) dy_t dy_u = \int_{y_t} A_t(k,d)p\left(y_t|\lambda\right) dy_t \cdot \int_{y_u} A_u(k,d)p\left(y_u|\lambda\right) dy_u$$

Using the definition of $T_{k,i}$ we get:

$$A_t(k,d)p\left(y_t|\lambda\right) = T_{k,t}\left(\frac{\left|y_{t,d}-m_{k,d}\right|}{s_{k,d}^2} - \frac{1}{s_{k,d}}\right) p\left(y_t|\lambda\right) = \frac{\tau_k p_k(y_t|\lambda)}{p(y_t|\lambda)}\left(\frac{\left|y_{t,d}-m_{k,d}\right|}{s_{k,d}^2} - \frac{1}{s_{k,d}}\right) p\left(y_t|\lambda\right) = \tau_k p_k(y_t|\lambda)\left(\frac{\left|y_{t,d}-m_{k,d}\right|}{s_{k,d}^2} - \frac{1}{s_{k,d}}\right)$$

Therefore:

$$
\begin{aligned}
\int_{y_t} A_t(k,d)p\left(y_t|\lambda\right)dy_t &= \int_{y_t} \tau_k p_k(y_t|\lambda)\left(\frac{|y_{t,d}-m_{k,d}|}{s_{k,d}^2}-\frac{1}{s_{k,d}}\right)dy_t \\
&= \tau_k \cdot \left[\prod_{c\neq d}\int_{y_{t,c}} p_{k,d}(y_{t,c}|\lambda)dy_{t,c}\right]\int_{y_{t,d}} p_{k,d}(y_{t,d}|\lambda)\left(\frac{|y_{t,d}-m_{k,d}|}{s_{k,d}^2}-\frac{1}{s_{k,d}}\right)dy_{t,d} \\
&= \tau_k \cdot \left[\prod_{c\neq d}\int_{y_{t,c}} H_{b_{k,d}}\left(y_{t,c}\right)dy_{t,c}\right]\int_{y_{t,d}} p_{k,d}(y_{t,d}|\lambda)\left(\frac{|y_{t,d}-m_{k,d}|}{s_{k,d}^2}-\frac{1}{s_{k,d}}\right)dy_{t,d} \\
&= \tau_k \cdot \int_{y_{t,d}} p_{k,d}(y_{t,d}|\lambda)\left(\frac{|y_{t,d}-m_{k,d}|}{s_{k,d}^2}-\frac{1}{s_{k,d}}\right)dy_{t,d} \\
&= \frac{\tau_k}{s_{k,d}} \cdot \int_{y_{t,d}} p_{k,d}(y_{t,d}|\lambda)\left(\frac{|y_{t,d}-m_{k,d}|}{s_{k,d}}\right)dy_{t,d} - \frac{\tau_k}{s_{k,d}} \\
&= \frac{\tau_k}{2s_{k,d}^2} \cdot \int_{y_{t,d}} \exp\left(-\frac{|y_{t,d}-m_{k,d}|}{s_{k,d}}\right)\left(\frac{|y_{t,d}-m_{k,d}|}{s_{k,d}}\right)dy_{t,d} - \frac{\tau_k}{s_{k,d}} \\
&= \frac{\tau_k}{2s_{k,d}^2} \cdot \left[\int_{-\infty}^{y_{t,d}<m_{k,d}} \exp\left(-\frac{m_{k,d}-y_{t,d}}{s_{k,d}}\right)\left(\frac{m_{k,d}-y_{t,d}}{s_{k,d}}\right)dy_{t,d}\right. \\
&+ \left.\int_{y_{t,d}>m_{k,d}}^{\infty} \exp\left(-\frac{y_{t,d}-m_{k,d}}{s_{k,d}}\right)\left(\frac{y_{t,d}-m_{k,d}}{s_{k,d}}\right)dy_{t,d}\right] - \frac{\tau_k}{s_{k,d}} \\
&= \frac{\tau_k}{2s_{k,d}} \cdot \left[-\int_{\infty}^{v>0} \exp\left(-v\right)\left(v\right)dv + \int_{v>0}^{\infty} \exp\left(-v\right)\left(v\right)dv\right] - \frac{\tau_k}{s_{k,d}} \\
&= \frac{\tau_k}{s_{k,d}} \cdot \int_{v>0}^{\infty} \exp\left(-v\right)\left(v\right)dv - \frac{\tau_k}{s_{k,d}} \\
&= 0
\end{aligned}
$$

$A_t(k,d)^2 p\left(y_t|\lambda\right) = T_{k,t}^2\left(\frac{|y_{t,d}-m_{k,d}|}{s_{k,d}^2}-\frac{1}{s_{k,d}}\right)^2 p\left(y_t|\lambda\right) \approx T_{k,t}\left(\frac{|y_{t,d}-m_{k,d}|}{s_{k,d}^2}-\frac{1}{s_{k,d}}\right)^2 p\left(y_t|\lambda\right) =$
$\tau_k p_k(y_t|\lambda)\left(\frac{|y_{t,d}-m_{k,d}|^2}{s_{k,d}^4}-2\frac{|y_{t,d}-m_{k,d}|}{s_{k,d}^3}+\frac{1}{s_{k,d}^2}\right)$
Therefore:

$$
\begin{aligned}
\int_{y_t} A_t(k,d)^2 p\left(y_t|\lambda\right) dy_t &= \int_{y_t} \tau_k p_k(y_t|\lambda) A_t(k,d)^2 dy_t \\
&= \tau_k \cdot \left[\prod_{c \neq d} \int_{y_{t,c}} p_{k,d}(y_{t,c}|\lambda) dy_{t,c} \right] \int_{y_{t,d}} p_{k,d}(y_{t,d}|\lambda) A_t(k,d)^2 dy_{t,d} \\
&= \tau_k \cdot \left[\prod_{c \neq d} \int_{y_{t,c}} H_{b_{k,d}}\left(y_{t,c}\right) dy_{t,c} \right] \int_{y_{t,d}} p_{k,d}(y_{t,d}|\lambda) A_t(k,d)^2 dy_{t,d} \\
&= \tau_k \cdot \int_{y_{t,d}} p_{k,d}(y_{t,d}|\lambda) \left(\frac{\left|y_{t,d} - m_{k,d}\right|^2}{s_{k,d}^4} - 2\frac{\left|y_{t,d} - m_{k,d}\right|}{s_{k,d}^3} + \frac{1}{s_{k,d}^2} \right) dy_{t,d}
\end{aligned}
$$

We'll split this expression to three parts:
1)

$$
\begin{aligned}
&= \tau_k \cdot \int_{y_{t,d}} p_{k,d}(y_{t,d}|\lambda) \frac{1}{s_{k,d}^2} dy_{t,d} \\
&= \frac{\tau_k}{s_{k,d}^2}
\end{aligned}
$$

2)

$$
\begin{aligned}
&= \tau_k \cdot \int_{y_{t,d}} p_{k,d}(y_{t,d}|\lambda) \frac{|y_{t,d}-m_{k,d}|^2}{s_{k,d}^4} dy_{t,d} \\
&= \frac{\tau_k}{2s_{k,d}} \cdot \int_{y_{t,d}} \exp\left(-\frac{|y_{t,d}-m_{k,d}|}{s_{k,d}}\right) \frac{|y_{t,d}-m_{k,d}|^2}{s_{k,d}^4} dy_{t,d} \\
&= \frac{\tau_k}{2s_{k,d}^3} \cdot \int_{y_{t,d}} \exp\left(-\frac{|y_{t,d}-m_{k,d}|}{s_{k,d}}\right) \frac{|y_{t,d}-m_{k,d}|^2}{s_{k,d}^2} dy_{t,d} \\
&= \frac{\tau_k}{2s_{k,d}^3} \cdot \int_{y_{t,d}} \exp\left(-\frac{|y_{t,d}-m_{k,d}|}{s_{k,d}}\right) \frac{|y_{t,d}-m_{k,d}|^2}{s_{k,d}^2} dy_{t,d} \\
&= \frac{\tau_k}{2s_{k,d}^3} \cdot \left[\int_{-\infty}^{y_{t,d}<m_{k,d}} \exp\left(-\frac{m_{k,d}-y_{t,d}}{s_{k,d}}\right) \left(\frac{m_{k,d}-y_{t,d}}{s_{k,d}}\right)^2 dy_{t,d} \right. \\
&+ \left. \int_{y_{t,d}>m_{k,d}}^{\infty} \exp\left(-\frac{y_{t,d}-m_{k,d}}{s_{k,d}}\right) \left(\frac{y_{t,d}-m_{k,d}}{s_{k,d}}\right)^2 dy_{t,d} \right] \\
&= \frac{\tau_k}{2s_{k,d}^2} \cdot \left[-\int_{\infty}^{v>0} \exp(-v)\left(v^2\right) dv + \int_{v>0}^{\infty} \exp(-v)\left(v^2\right) dv\right] \\
&= \frac{\tau_k}{s_{k,d}^2} \cdot \int_{v>0}^{\infty} \exp(-v)\left(v^2\right) dv \\
&= \frac{2\tau_k}{s_{k,d}^2}
\end{aligned}
$$

3)

$$
\begin{aligned}
&= -2\tau_k \cdot \int_{y_{t,d}} p_{k,d}(y_{t,d}|\lambda) \frac{|y_{t,d}-m_{k,d}|}{s_{k,d}^3} dy_{t,d} \\
&= \frac{-\tau_k}{s_{k,d}^3} \cdot \int_{y_{t,d}} \exp\left(-\frac{|y_{t,d}-m_{k,d}|}{s_{k,d}}\right) \frac{|y_{t,d}-m_{k,d}|}{s_{k,d}} dy_{t,d} \\
&= \frac{-2\tau_k}{s_{k,d}^2}
\end{aligned}
$$

Finally: $f_{s_{k,d}} = 0 + N \cdot \left(\frac{2\tau_k}{s_{k,d}^2} - \frac{2\tau_k}{s_{k,d}^2} + \frac{\tau_k}{s_{k,d}^2}\right) = \frac{N\tau_k}{s_{k,d}^2}$

References

[1] F. Perronnin and C. Dance. Fisher kernels on visual vocabularies for image categorization. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pages 1–8. IEEE, 2007.