第 3 章补充材料

● 高斯分布参数的极大似然估计

样本集合 $D=\{\mathbf{x}_1,\cdots,\mathbf{x}_n\}$ 独立抽样自均值为 $\boldsymbol{\mu}$，协方差矩阵为 $\boldsymbol{\Sigma}$ 的高斯分布。建立对数自然函数：

$$l(\boldsymbol{\mu},\boldsymbol{\Sigma})=\sum_{i=1}^{n}\ln p(\mathbf{x}_i|\boldsymbol{\mu},\boldsymbol{\Sigma})$$

其中：

$$\begin{aligned}\ln p(\mathbf{x}_i|\boldsymbol{\mu},\boldsymbol{\Sigma})&=\ln\left[\frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}}\exp\left(-\frac{1}{2}(\mathbf{x}_i-\boldsymbol{\mu})^t\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i-\boldsymbol{\mu})\right)\right]\\&=-\frac{d}{2}\ln(2\pi)-\frac{1}{2}\ln|\boldsymbol{\Sigma}|-\frac{1}{2}(\mathbf{x}_i-\boldsymbol{\mu})^t\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i-\boldsymbol{\mu})\end{aligned}$$

因此，对数似然函数为：

$$l(\boldsymbol{\mu},\boldsymbol{\Sigma})=\sum_{i=1}^{n}\left[-\frac{d}{2}\ln(2\pi)-\frac{1}{2}\ln|\boldsymbol{\Sigma}|-\frac{1}{2}(\mathbf{x}_i-\boldsymbol{\mu})^t\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i-\boldsymbol{\mu})\right] \qquad (1)$$

首先来推导均值矢量 $\boldsymbol{\mu}$ 的最大似然估计。对数似然函数对均值矢量 $\boldsymbol{\mu}$ 求偏导数及极值：

$$\frac{\partial l(\boldsymbol{\mu},\boldsymbol{\Sigma})}{\partial\boldsymbol{\mu}}=\sum_{i=1}^{n}\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i-\boldsymbol{\mu})=\boldsymbol{\Sigma}^{-1}\left[\sum_{i=1}^{n}(\mathbf{x}_i-\boldsymbol{\mu})\right]=\mathbf{0}$$

这里利用了协方差矩阵为对称矩阵的事实，上式两边左乘 $\boldsymbol{\Sigma}$：

$$\sum_{i=1}^{n}(\mathbf{x}_i-\boldsymbol{\mu})=\sum_{i=1}^{n}\mathbf{x}_i-n\boldsymbol{\mu}=\mathbf{0}$$

这样就得到了均值矢量 $\boldsymbol{\mu}$ 的最大似然估计：

$$\boldsymbol{\mu}=\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i$$

协方差矩阵 $\boldsymbol{\Sigma}$ 的最大似然估计推导要复杂一些。首先给出需要用到的几个关于 $d\times d$ 维方阵 $\mathbf{A}$ 的基本性质：

1．逆矩阵的行列式值等于行列式值的倒数：

$$|\mathbf{A}^{-1}|=\frac{1}{|\mathbf{A}|}$$

2．令 $f(\mathbf{A})=|\mathbf{A}|$，则矩阵 $\mathbf{A}$ 的行列式值对矩阵的导数：

$$\frac{df(\mathbf{A})}{d\mathbf{A}}=\frac{d(|\mathbf{A}|)}{d\mathbf{A}}=|\mathbf{A}|\mathbf{A}^{-1}$$

3．令 $g(\mathbf{A})=\mathbf{x}^t\mathbf{A}\mathbf{y}$， $\mathbf{x}$ 和 $\mathbf{y}$ 为 d 维列矢量，函数 $g(\mathbf{A})$ 对矩阵 $\mathbf{A}$ 的导数：

$$\frac{\partial g(\mathbf{A})}{\partial\mathbf{A}}=\mathbf{x}\mathbf{y}^t$$

根据性质 1，公式（1）的对数自然函数可以写成：

$$l(\boldsymbol{\mu},\boldsymbol{\Sigma})=\sum_{i=1}^{n}\left[-\frac{d}{2}\ln(2\pi)+\frac{1}{2}\ln\left|\boldsymbol{\Sigma}^{-1}\right|-\frac{1}{2}(\mathbf{x}_i-\boldsymbol{\mu})^t\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i-\boldsymbol{\mu})\right] \tag{2}$$

$l(\boldsymbol{\mu},\boldsymbol{\Sigma})$ 对协方差矩阵的逆阵 $\boldsymbol{\Sigma}^{-1}$ 求偏导数及极值：

$$\begin{aligned}
\frac{\partial l(\boldsymbol{\mu},\boldsymbol{\Sigma})}{\partial\boldsymbol{\Sigma}^{-1}}&=\sum_{i=1}^{n}\left\{\frac{1}{2}\frac{\partial\left(\ln\left|\boldsymbol{\Sigma}^{-1}\right|\right)}{\partial\boldsymbol{\Sigma}^{-1}}-\frac{1}{2}\frac{\partial}{\partial\boldsymbol{\Sigma}^{-1}}\left[(\mathbf{x}_i-\boldsymbol{\mu})^t\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i-\boldsymbol{\mu})\right]\right\}\\
&=\sum_{i=1}^{n}\left[\frac{1}{2}\frac{1}{\left|\boldsymbol{\Sigma}^{-1}\right|}\frac{\partial\left|\boldsymbol{\Sigma}^{-1}\right|}{\partial\boldsymbol{\Sigma}^{-1}}-\frac{1}{2}(\mathbf{x}_i-\boldsymbol{\mu})(\mathbf{x}_i-\boldsymbol{\mu})^t\right]\\
&=\sum_{i=1}^{n}\left[\frac{1}{2}\frac{1}{\left|\boldsymbol{\Sigma}^{-1}\right|}\left|\boldsymbol{\Sigma}^{-1}\right|\boldsymbol{\Sigma}-\frac{1}{2}(\mathbf{x}_i-\boldsymbol{\mu})(\mathbf{x}_i-\boldsymbol{\mu})^t\right]\\
&=\frac{1}{2}\sum_{i=1}^{n}\left[\boldsymbol{\Sigma}-(\mathbf{x}_i-\boldsymbol{\mu})(\mathbf{x}_i-\boldsymbol{\mu})^t\right]\\
&=\frac{n}{2}\boldsymbol{\Sigma}-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_i-\boldsymbol{\mu})(\mathbf{x}_i-\boldsymbol{\mu})^t\\
&=\mathbf{0}
\end{aligned}$$

其中第 2 行和第 3 行分别用到了性质 3 和性质 2，由此可以得到：

$$\boldsymbol{\Sigma}=\frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i-\boldsymbol{\mu})(\mathbf{x}_i-\boldsymbol{\mu})^t$$

● 高斯混合模型 EM 算法的迭代公式推导

我们首先来推导一般混合密度模型参数估计的 EM 算法迭代公式，然后再将一般的混合密度模型具体化为高斯混合模型。

I. 混合密度模型

假设样本集 $X=\{\mathbf{x}_1,\cdots,\mathbf{x}_n\}$ 中的样本相互独立，并且按照如下的过程产生：

1. 样本是依据概率由 K 个分布中的一个产生的，分布的概率密度函数为 $p(\mathbf{x}|\boldsymbol{\theta}_k)$，$k=1,\cdots,K$，$\boldsymbol{\theta}_k$ 为分布的参数；
2. 由第 k 个分布产生样本的先验概率为 α_k；
3. 先验概率 $\boldsymbol{\alpha}=(\alpha_1,\cdots,\alpha_K)^t$，以及分布的参数 $\boldsymbol{\theta}_1,\cdots,\boldsymbol{\theta}_K$ 均未知。

我们称样本集 X 来自于一个“混合密度模型”，混合密度模型的概率密度函数为：

$$p(\mathbf{x}|\boldsymbol{\Theta})=\sum_{k=1}^{K}\alpha_k p(\mathbf{x}|\boldsymbol{\theta}_k) \tag{1}$$

其中 $\boldsymbol{\Theta}=(\boldsymbol{\alpha},\boldsymbol{\theta}_1,\cdots,\boldsymbol{\theta}_K)$ 为模型的参数，每个 $p(\mathbf{x}|\boldsymbol{\theta}_k)$ 称为一个分量密度。

II. 混合密度模型参数估计的 EM 迭代公式

混合密度模型的参数估计中，由于样本是由哪个分量密度所产生的信息 $Y=\{y_1,\cdots,y_n\}$ 是未知的，因此需要将其视作“丢失”信息，使用 EM 算法进行估计。EM 算法中 E 步和 M 步的迭代公式：

E 步：$$Q(\boldsymbol{\Theta};\boldsymbol{\Theta}^g)=E_Y\left[\ln p(X,Y|\boldsymbol{\Theta})\middle|X,\boldsymbol{\Theta}^g\right] \tag{2}$$

M 步：$$\boldsymbol{\Theta}^*=\arg\max_{\boldsymbol{\Theta}} Q(\boldsymbol{\Theta};\boldsymbol{\Theta}^g) \tag{3}$$

其中 $\boldsymbol{\Theta}^g$ 是对参数 $\boldsymbol{\Theta}$ 的一个猜测值。E 步计算的是在已知样本集 X 和参数猜测值 $\boldsymbol{\Theta}^g$ 的条件下期望对数似然函数；而 M 步则是对 $Q(\boldsymbol{\Theta};\boldsymbol{\Theta}^g)$ 的优化。更新参数的猜测值设置：$\boldsymbol{\Theta}^g=\boldsymbol{\Theta}^*$，进入下一轮迭代。

E 步期望对数似然函数 $Q(\boldsymbol{\Theta};\boldsymbol{\Theta}^g)$ 的推导：

训练样本 $\mathbf{x}_i$ 是由第 y_i 个分量密度函数产生的，$y_i=1,\cdots,K$，这两个随机事件的联合概率密度：

$$p(\mathbf{x}_i,y_i|\boldsymbol{\Theta})=\alpha_{y_i}p(\mathbf{x}_i|\boldsymbol{\theta}_{y_i})$$

因此，关于完整数据集 $D=\{X,Y\}$ 的对数似然函数为：

$$l(\boldsymbol{\Theta})=\ln p(X,Y|\boldsymbol{\Theta})=\sum_{i=1}^{n}\ln\left[\alpha_{y_i}p(\mathbf{x}_i|\boldsymbol{\theta}_{y_i})\right] \tag{4}$$

另外根据贝叶斯公式，在已知参数的一个猜测值 $\boldsymbol{\Theta}^g=(\alpha_1^g,\cdots,\alpha_K^g,\boldsymbol{\theta}_1^g,\cdots,\boldsymbol{\theta}_K^g)$ 和样本 $\mathbf{x}_i$ 的条件下，$\mathbf{x}_i$ 由第 y_i 个分量产生的概率为：

$$P\left(y_i\middle|\mathbf{x}_i,\mathbf{\Theta}^g\right)=\frac{p\left(\mathbf{x}_i,y_i\middle|\mathbf{\Theta}^g\right)}{p\left(\mathbf{x}_i\middle|\mathbf{\Theta}^g\right)}=\frac{a_{y_i}^g p\left(\mathbf{x}_i\middle|\mathbf{\theta}_{y_i}^g\right)}{\sum_{k=1}^K a_k^g p\left(\mathbf{x}_i\middle|\mathbf{\theta}_k^g\right)} \tag{5}$$

考虑到样本的独立同分布性，y_i 只与 $\mathbf{x}_i$ 有关，独立于其它 $\mathbf{x}_j$ 和 y_j，$j\neq i$，因此：

$$P\left(Y\middle|X,\mathbf{\Theta}^g\right)=P\left(y_1,\cdots,y_n\middle|\mathbf{x}_1,\cdots,\mathbf{x}_n,\mathbf{\Theta}^g\right)=\prod_{i=1}^n P\left(y_i\middle|\mathbf{x}_i,\mathbf{\Theta}^g\right) \tag{6}$$

将（4）、（6）式代入到（2）式 E 步的期望对数似然函数，同时考虑到每一个 y_i 是离散的，只取 $\{1,\cdots,K\}$ 中的某一个值，对 Y 的数学期望可以由如下的求和式计算：

$$\begin{aligned}
Q\left(\mathbf{\Theta};\mathbf{\Theta}^g\right)&=\sum_{y_1=1}^K\sum_{y_2=1}^K\cdots\sum_{y_n=1}^K \ln p\left(X,Y\middle|\mathbf{\Theta}\right)P\left(Y\middle|X,\mathbf{\Theta}^g\right)\\
&=\sum_{y_1=1}^K\sum_{y_2=1}^K\cdots\sum_{y_n=1}^K\left\{\sum_{i=1}^n\ln\left[\alpha_{y_i}p\left(\mathbf{x}_i\middle|\mathbf{\theta}_{y_i}\right)\right]\right\}\prod_{i=1}^n P\left(y_i\middle|\mathbf{x}_i,\mathbf{\Theta}^g\right)\\
&=\sum_{y_1=1}^K\sum_{y_2=1}^K\cdots\sum_{y_n=1}^K\sum_{i=1}^n\sum_{l=1}^K\left\{\delta_{l,y_i}\ln\left[\alpha_{y_i}p\left(\mathbf{x}_i\middle|\mathbf{\theta}_{y_i}\right)\right]\prod_{j=1}^n P\left(y_j\middle|\mathbf{x}_j,\mathbf{\Theta}^g\right)\right\}\\
&=\sum_{i=1}^n\sum_{l=1}^K\left\{\ln\left[\alpha_l p\left(\mathbf{x}_i\middle|\mathbf{\theta}_l\right)\right]\left\{\sum_{y_1=1}^K\sum_{y_2=1}^K\cdots\sum_{y_n=1}^K\left[\delta_{l,y_i}\prod_{j=1}^n P\left(y_j\middle|\mathbf{x}_j,\mathbf{\Theta}^g\right)\right]\right\}\right\}
\end{aligned} \tag{7}$$

其中：

$$\delta_{l,y_i}=\begin{cases}1, & l=y_i\\ 0, & l\neq y_i\end{cases}$$

由于 $\sum_{j=1}^K P\left(y_j\middle|\mathbf{x}_j,\mathbf{\Theta}^g\right)=1$，因此（7）式内层大括号中的内容可以简化为：

$$\begin{aligned}
\sum_{y_1=1}^K\sum_{y_2=1}^K\cdots\sum_{y_n=1}^K\left[\delta_{l,y_i}\prod_{j=1}^n P\left(y_j\middle|\mathbf{x}_j,\mathbf{\Theta}^g\right)\right]&=\left[\sum_{y_1=1}^K\cdots\sum_{y_{i-1}=1}^K\sum_{y_{i+1}=1}^K\cdots\sum_{y_n=1}^K\prod_{j=1,j\neq i}^n P\left(y_j\middle|\mathbf{x}_j,\mathbf{\Theta}^g\right)\right]P\left(l\middle|\mathbf{x}_i,\mathbf{\Theta}^g\right)\\
&=\prod_{j=1,j\neq i}^K\left(\sum_{y_j=1}^K P\left(y_j\middle|\mathbf{x}_j,\mathbf{\Theta}^g\right)\right)P\left(l\middle|\mathbf{x}_i,\mathbf{\Theta}^g\right)\\
&=P\left(l\middle|\mathbf{x}_i,\mathbf{\Theta}^g\right)
\end{aligned} \tag{8}$$

（8）式第 2 步过程使用的是乘法的分配率。代入（7）式可得：

$$\begin{aligned}
Q\left(\mathbf{\Theta};\mathbf{\Theta}^g\right)&=\sum_{i=1}^n\sum_{l=1}^K\left\{\ln\left[\alpha_l p\left(\mathbf{x}_i\middle|\mathbf{\theta}_l\right)\right]P\left(l\middle|\mathbf{x}_i,\mathbf{\Theta}^g\right)\right\}\\
&=\sum_{i=1}^n\sum_{l=1}^K\left[\ln\alpha_l P\left(l\middle|\mathbf{x}_i,\mathbf{\Theta}^g\right)\right]+\sum_{i=1}^n\sum_{l=1}^K\left[\ln p\left(\mathbf{x}_i\middle|\mathbf{\theta}_l\right)P\left(l\middle|\mathbf{x}_i,\mathbf{\Theta}^g\right)\right]
\end{aligned} \tag{9}$$

上式中的期望对数似然函数 $Q\left(\mathbf{\Theta};\mathbf{\Theta}^g\right)$ 只是参数 $\mathbf{\Theta}$ 的函数，而 $\mathbf{x}_1,\cdots,\mathbf{x}_n$ 以及 $\mathbf{\Theta}^g$ 均为已知。

M 步期望对数似然函数 $Q\left(\mathbf{\Theta};\mathbf{\Theta}^g\right)$ 的优化：

下面来求解公式（3）M 步的优化问题，需要注意的是参数 $\mathbf{\alpha}=\left(\alpha_1,\cdots,\alpha_K\right)^t$ 存在约束

$\sum_{k=1}^{K}\alpha_k=1$，因此构造 Lagrange 函数：

$$
\begin{aligned}
L(\boldsymbol{\Theta},\lambda) &= Q\left(\boldsymbol{\Theta};\boldsymbol{\Theta}^g\right)+\lambda\left(\sum_{k=1}^{K}\alpha_k-1\right) \\
&= \sum_{i=1}^{n}\sum_{l=1}^{K}\left[\ln\alpha_l P\left(l|\mathbf{x}_i,\boldsymbol{\Theta}^g\right)\right]+\sum_{i=1}^{n}\sum_{l=1}^{K}\left[\ln p\left(\mathbf{x}_i|\boldsymbol{\theta}_l\right)P\left(l|\mathbf{x}_i,\boldsymbol{\Theta}^g\right)\right]+\lambda\left(\sum_{k=1}^{K}\alpha_k-1\right)
\end{aligned} \quad (10)
$$

Lagrange 函数对 α_l 求偏导数及极值：

$$
\frac{\partial L(\boldsymbol{\Theta},\lambda)}{\partial\alpha_l}=\sum_{i=1}^{n}\left[\frac{1}{a_l}P\left(l|\mathbf{x}_i,\boldsymbol{\Theta}^g\right)\right]+\lambda=0
$$

因此有：

$$
a_l\lambda+\sum_{i=1}^{n}P\left(l|\mathbf{x}_i,\boldsymbol{\Theta}^g\right)=0 \quad (11)
$$

等式对 l 求和：

$$
\sum_{l=1}^{K}\left[a_l\lambda+\sum_{i=1}^{n}P\left(l|\mathbf{x}_i,\boldsymbol{\Theta}^g\right)\right]=\lambda\sum_{l=1}^{K}a_l+\sum_{i=1}^{n}\sum_{l=1}^{K}P\left(l|\mathbf{x}_i,\boldsymbol{\Theta}^g\right)=\lambda+n=0
$$

因此 Lagrange 系数 $\lambda=-n$，代入（11）式得到关于混合密度组合系数 a_l 的估计公式：

$$
a_l=\frac{1}{n}\sum_{i=1}^{n}P\left(l|\mathbf{x}_i,\boldsymbol{\Theta}^g\right) \quad (12)
$$

其中 $P\left(l|\mathbf{x}_i,\boldsymbol{\Theta}^g\right)$ 可以由（5）式计算。

III. 高斯混合模型参数估计的 EM 迭代公式

对于每一个分量密度函数参数的估计，需要考虑具体的分量密度函数形式，下面推导高斯混合模型中分量高斯的均值矢量 $\boldsymbol{\mu}_l$ 和协方差矩阵 $\boldsymbol{\Sigma}_l$ 的估计公式。

高斯混合模型中，第 l 个分量密度函数是一个高斯函数：

$$
p_l\left(\mathbf{x}|\boldsymbol{\theta}_l\right)=\frac{1}{(2\pi)^{d/2}\left|\boldsymbol{\Sigma}_l\right|^{1/2}}\exp\left[-\frac{1}{2}\left(\mathbf{x}-\boldsymbol{\mu}_l\right)^t\boldsymbol{\Sigma}_l^{-1}\left(\mathbf{x}-\boldsymbol{\mu}_l\right)\right]
$$

考虑到（10）式 Lagrange 函数中第 1 项和第 3 项与均值矢量 $\boldsymbol{\mu}_l$ 和协方差矩阵 $\boldsymbol{\Sigma}_l$ 无关，在优化时不起作用，为了书写简单可以将其省略。将高斯函数代入（10）式：

$$
\begin{aligned}
L(\boldsymbol{\Theta},\lambda) &= \sum_{l=1}^{K}\sum_{i=1}^{n}\left[\ln p\left(\mathbf{x}_i|\boldsymbol{\theta}_l\right)P\left(l|\mathbf{x}_i,\boldsymbol{\Theta}^g\right)\right] \\
&= \sum_{l=1}^{K}\sum_{i=1}^{n}\left[\left(-\frac{d}{2}\ln 2\pi-\frac{1}{2}\ln\left|\boldsymbol{\Sigma}_l\right|-\frac{1}{2}\left(\mathbf{x}_i-\boldsymbol{\mu}_l\right)^t\boldsymbol{\Sigma}_l^{-1}\left(\mathbf{x}_i-\boldsymbol{\mu}_l\right)\right)P\left(l|\mathbf{x}_i,\boldsymbol{\Theta}^g\right)\right]
\end{aligned} \quad (13)
$$

首先对 $\boldsymbol{\mu}_l$ 求偏导数及极值：

$$
\begin{aligned}
\frac{\partial L(\boldsymbol{\Theta},\lambda)}{\partial\boldsymbol{\mu}_l} &= \sum_{i=1}^{n}\left[\boldsymbol{\Sigma}_l^{-1}\left(\mathbf{x}_i-\boldsymbol{\mu}_l\right)P\left(l|\mathbf{x}_i,\boldsymbol{\Theta}^g\right)\right] \\
&= \boldsymbol{\Sigma}_l^{-1}\left[\sum_{i=1}^{n}\mathbf{x}_iP\left(l|\mathbf{x}_i,\boldsymbol{\Theta}^g\right)-\boldsymbol{\mu}_l\sum_{i=1}^{n}P\left(l|\mathbf{x}_i,\boldsymbol{\Theta}^g\right)\right] \\
&= \mathbf{0}
\end{aligned}
$$

两边左乘 $\mathbf{\Sigma}_l$，可以得到均值矢量 $\boldsymbol{\mu}_l$ 的估计公式：

$$\boldsymbol{\mu}_l = \sum_{i=1}^{n} \mathbf{x}_i P\left(l \middle| \mathbf{x}_i, \mathbf{\Theta}^g\right) \Big/ \sum_{i=1}^{n} P\left(l \middle| \mathbf{x}_i, \mathbf{\Theta}^g\right) \tag{14}$$

（13）式对 $\mathbf{\Sigma}_l^{-1}$ 求偏导数及极值（具体过程参见高斯分布参数最大似然估计的推导过程）：

$$\begin{aligned}
\frac{\partial L(\mathbf{\Theta}, \lambda)}{\partial \mathbf{\Sigma}_l^{-1}} &= \sum_{i=1}^{n} \left[\left(\frac{1}{2} \frac{1}{\left|\mathbf{\Sigma}_l^{-1}\right|} \left|\mathbf{\Sigma}_l^{-1}\right| \mathbf{\Sigma}_l - \frac{1}{2} (\mathbf{x}_i - \boldsymbol{\mu}_l)(\mathbf{x}_i - \boldsymbol{\mu}_l)^t \right) P\left(l \middle| \mathbf{x}_i, \mathbf{\Theta}^g\right) \right] \\
&= \frac{1}{2} \left\{ \mathbf{\Sigma}_l \left[\sum_{i=1}^{n} P\left(l \middle| \mathbf{x}_i, \mathbf{\Theta}^g\right) \right] - \sum_{i=1}^{n} P\left(l \middle| \mathbf{x}_i, \mathbf{\Theta}^g\right) (\mathbf{x}_i - \boldsymbol{\mu}_l)(\mathbf{x}_i - \boldsymbol{\mu}_l)^t \right\} \\
&= \mathbf{0}
\end{aligned}$$

因此得到协方差矩阵 $\mathbf{\Sigma}_l$ 的估计公式：

$$\mathbf{\Sigma}_l = \left[\sum_{i=1}^{n} P\left(l \middle| \mathbf{x}_i, \mathbf{\Theta}^g\right) (\mathbf{x}_i - \boldsymbol{\mu}_l)(\mathbf{x}_i - \boldsymbol{\mu}_l)^t \right] \Big/ \sum_{i=1}^{n} P\left(l \middle| \mathbf{x}_i, \mathbf{\Theta}^g\right) \tag{15}$$

总结（5）、（12）、（14）和（15）式，得到高斯混合模型参数估计 EM 算法第 j 轮的迭代公式：

$$P\left(l \middle| \mathbf{x}_i, \mathbf{\Theta}^{j-1}\right) = \alpha_l^{j-1} p\left(\mathbf{x}_i \middle| \boldsymbol{\theta}_l^{j-1}\right) \Big/ \sum_{k=1}^{K} \alpha_k^{j-1} p\left(\mathbf{x}_i \middle| \boldsymbol{\theta}_k^{j-1}\right)$$

其中 $p\left(\mathbf{x}_i \middle| \boldsymbol{\theta}_l^{j-1}\right)$ 为高斯函数

$$a_l^j = \frac{1}{n} \sum_{i=1}^{n} P\left(l \middle| \mathbf{x}_i, \mathbf{\Theta}^{j-1}\right)$$

$$\boldsymbol{\mu}_l^j = \sum_{i=1}^{n} \mathbf{x}_i P\left(l \middle| \mathbf{x}_i, \mathbf{\Theta}^{j-1}\right) \Big/ \sum_{i=1}^{n} P\left(l \middle| \mathbf{x}_i, \mathbf{\Theta}^{j-1}\right)$$

$$\mathbf{\Sigma}_l^j = \left[\sum_{i=1}^{n} P\left(l \middle| \mathbf{x}_i, \mathbf{\Theta}^{j-1}\right) (\mathbf{x}_i - \boldsymbol{\mu}_l^j)(\mathbf{x}_i - \boldsymbol{\mu}_l^j)^t \right] \Big/ \sum_{i=1}^{n} P\left(l \middle| \mathbf{x}_i, \mathbf{\Theta}^{j-1}\right)$$

● 103 页，例 2 的详细推导过程

二维空间中 4 个样本，其中的一个样本丢失 1 个特征：

$$D=\{\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3,\mathbf{x}_4\}=\left\{\begin{pmatrix}0\\2\end{pmatrix},\begin{pmatrix}1\\0\end{pmatrix},\begin{pmatrix}2\\2\end{pmatrix},\begin{pmatrix}*\\4\end{pmatrix}\right\}$$

假设样本满足正态分布，协方差矩阵为对角阵，EM 算法估计参数：

$$\boldsymbol{\theta}=\begin{pmatrix}\mu_1\\\mu_2\\\sigma_1\\\sigma_2\end{pmatrix},\qquad \text{初始参数：}\ \boldsymbol{\theta}_0=\begin{pmatrix}0\\0\\1\\1\end{pmatrix}$$

推导：

首先计算已知参数 $\boldsymbol{\theta}_0$ 的条件下，参数 $\boldsymbol{\theta}$ 的对数似然函数：

E 步：

$$\begin{aligned}Q(\boldsymbol{\theta};\boldsymbol{\theta}_0)&=E_{x_{41}}\left[\ln p\left(\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3,\mathbf{x}_4\middle|\boldsymbol{\theta}\right)\middle|\boldsymbol{\theta}_0,\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3,x_{42}\right]\\&=\int_{-\infty}^{+\infty}\ln p\left(\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3,\mathbf{x}_4\middle|\boldsymbol{\theta}\right)p\left(x_{41}\middle|\boldsymbol{\theta}_0,x_{42}\right)dx_{41}\end{aligned}\qquad（1）$$

其中：

$$\ln p\left(\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3,\mathbf{x}_4\middle|\boldsymbol{\theta}\right)=\sum_{i=1}^{3}\ln p\left(\mathbf{x}_i\middle|\boldsymbol{\theta}\right)+\ln p\left(x_{41},x_{42}\middle|\boldsymbol{\theta}\right)\qquad（2）$$

$$\begin{aligned}p\left(x_{41}\middle|\boldsymbol{\theta}_0,x_{42}\right)&=\frac{p\left(x_{41},x_{42}\middle|\boldsymbol{\theta}_0\right)}{p\left(x_{42}\middle|\boldsymbol{\theta}_0\right)}\\&=\frac{p\left(x_{41},x_{42}\middle|\boldsymbol{\theta}_0\right)}{\int_{-\infty}^{+\infty}p\left(x_{41},x_{42}\middle|\boldsymbol{\theta}_0\right)dx_{41}}\\&=\frac{1}{K}p\left(x_{41},x_{42}\middle|\boldsymbol{\theta}_0\right)\end{aligned}\qquad（3）$$

$$\begin{aligned}K&=\int_{-\infty}^{+\infty}p\left(x_{41},x_{42}\middle|\boldsymbol{\theta}_0\right)dx_{41}\\&=\int_{-\infty}^{+\infty}\frac{1}{2\pi\sigma_{10}\sigma_{20}}\exp\left[-\frac{\left(x_{41}-\mu_{10}\right)^2}{2\sigma_{10}^2}-\frac{\left(x_{42}-\mu_{20}\right)^2}{2\sigma_{20}^2}\right]dx_{41}\\&=\frac{1}{\sqrt{2\pi}\sigma_{20}}\exp\left[-\frac{\left(x_{42}-\mu_{20}\right)^2}{2\sigma_{20}^2}\right]\left\{\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}\sigma_{10}}\exp\left[-\frac{\left(x_{41}-\mu_{10}\right)^2}{2\sigma_{10}^2}\right]dx_{41}\right\}\\&=\frac{1}{\sqrt{2\pi}\sigma_{20}}\exp\left[-\frac{\left(x_{42}-\mu_{20}\right)^2}{2\sigma_{20}^2}\right]=\frac{e^{-8}}{\sqrt{2\pi}}\end{aligned}$$

将（2）和（3）带入（1）：

$$\begin{aligned}Q(\boldsymbol{\theta};\boldsymbol{\theta}_0)&=\int_{-\infty}^{+\infty}\left[\sum_{i=1}^{3}\ln p\left(\mathbf{x}_i\middle|\boldsymbol{\theta}\right)\right]p\left(x_{41}\middle|\boldsymbol{\theta}_0,x_{42}\right)dx_{41}+\int_{-\infty}^{+\infty}\ln p\left(x_{41},x_{42}\middle|\boldsymbol{\theta}\right)p\left(x_{41}\middle|\boldsymbol{\theta}_0,x_{42}\right)dx_{41}\\&=\left[\sum_{i=1}^{3}\ln p\left(\mathbf{x}_i\middle|\boldsymbol{\theta}\right)\right]+\frac{1}{K}\int_{-\infty}^{+\infty}\ln p\left(x_{41},x_{42}\middle|\boldsymbol{\theta}\right)p\left(x_{41}\middle|\boldsymbol{\theta}_0,x_{42}\right)dx_{41}\end{aligned}\qquad(4)$$

其中：

$$
\begin{aligned}
\ln p\left(x_{41}, x_{42} \middle| \boldsymbol{\theta}\right) &= \ln\left\{\frac{1}{2\pi\sigma_1\sigma_2}\exp\left[-\frac{\left(x_{41}-\mu_1\right)^2}{2\sigma_1^2}-\frac{\left(x_{42}-\mu_2\right)^2}{2\sigma_2^2}\right]\right\} \\
&= -\ln\left(2\pi\sigma_1\sigma_2\right)-\frac{\left(x_{41}-\mu_1\right)^2}{2\sigma_1^2}-\frac{\left(x_{42}-\mu_2\right)^2}{2\sigma_2^2}
\end{aligned}
\tag{5}
$$

将（5）带入（4）：

$$
\begin{aligned}
Q\left(\boldsymbol{\theta};\boldsymbol{\theta}_0\right) &= \sum_{i=1}^{3}\ln p\left(\mathbf{x}_i \middle| \boldsymbol{\theta}\right)-\frac{1}{K}\int_{-\infty}^{+\infty}\left[\ln\left(2\pi\sigma_1\sigma_2\right)+\frac{\left(x_{42}-\mu_2\right)^2}{2\sigma_2^2}\right]p\left(x_{41}\middle|\boldsymbol{\theta}_0, x_{42}\right)dx_{41} \\
&\quad -\frac{1}{K}\int_{-\infty}^{+\infty}\frac{\left(x_{41}-\mu_1\right)^2}{2\sigma_1^2}p\left(x_{41}\middle|\boldsymbol{\theta}_0, x_{42}\right)dx_{41} \\
&= \sum_{i=1}^{3}\ln p\left(\mathbf{x}_i \middle| \boldsymbol{\theta}\right)-\ln\left(2\pi\sigma_1\sigma_2\right)-\frac{\left(x_{42}-\mu_2\right)^2}{2\sigma_2^2}-\frac{1}{K}\int_{-\infty}^{+\infty}\frac{\left(x_{41}-\mu_1\right)^2}{2\sigma_1^2}p\left(x_{41}\middle|\boldsymbol{\theta}_0, x_{42}\right)dx_{41}
\end{aligned}
\tag{6}
$$

计算积分：

$$
\begin{aligned}
\int_{-\infty}^{+\infty}\frac{\left(x_{41}-\mu_1\right)^2}{2\sigma_1^2}p\left(x_{41}\middle|\boldsymbol{\theta}_0, x_{42}\right)dx_{41} &= \frac{1}{2\sigma_1^2\times 2\pi\sigma_{10}\sigma_{20}}\exp\left[-\frac{\left(x_{42}-\mu_{20}\right)^2}{2\sigma_{20}^2}\right]\int_{-\infty}^{+\infty}\left(x_{41}-\mu_1\right)^2\exp\left[-\frac{\left(x_{41}-\mu_{10}\right)^2}{2\sigma_{10}^2}\right]dx_{41} \\
&= \frac{e^{-8}}{4\pi\sigma_1^2}\int_{-\infty}^{+\infty}\left(x_{41}^2-2\mu_1 x_{41}+\mu_1^2\right)e^{-\frac{x_{41}^2}{2}}dx_{41}
\end{aligned}
\tag{7}
$$

积分公式：

$$\int_{-\infty}^{+\infty}x_{41}^2 e^{-\frac{x_{41}^2}{2}}dx_{41}=\sqrt{2\pi}$$

$$\int_{-\infty}^{+\infty}-2\mu_1 x_{41}e^{-\frac{x_{41}^2}{2}}dx_{41}=0$$

$$\int_{-\infty}^{+\infty}\mu_1^2 e^{-\frac{x_{41}^2}{2}}dx_{41}=\mu_1^2\sqrt{2\pi}$$

带入（7）：

$$
\begin{aligned}
\frac{1}{K}\int_{-\infty}^{+\infty}\frac{\left(x_{41}-\mu_1\right)^2}{2\sigma_1^2}p\left(x_{41}\middle|\boldsymbol{\theta}_0, x_{42}\right)dx_{41} &= \frac{e^{-8}}{4\pi\sigma_1^2}\times\frac{1}{K}\times\left(\sqrt{2\pi}+\mu_1^2\sqrt{2\pi}\right) \\
&= \frac{\sqrt{2\pi}e^{-8}}{4\pi\sigma_1^2}\times\frac{\sqrt{2\pi}}{e^{-8}}\times\left(1+\mu_1^2\right) \\
&= \frac{1}{2\sigma_1^2}\left(1+\mu_1^2\right)
\end{aligned}
\tag{9}
$$

（9）带入（6）：

$$
\begin{aligned}
Q\left(\boldsymbol{\theta};\boldsymbol{\theta}_0\right) &= \sum_{i=1}^{3}\ln p\left(\mathbf{x}_i \middle| \boldsymbol{\theta}\right)-\ln\left(2\pi\sigma_1\sigma_2\right)-\frac{\left(x_{42}-\mu_2\right)^2}{2\sigma_2^2}-\frac{\left(1+\mu_1^2\right)}{2\sigma_1^2} \\
&= \sum_{i=1}^{3}\left[-\ln\left(2\pi\sigma_1\sigma_2\right)-\frac{\left(x_{i1}-\mu_1\right)^2}{2\sigma_1^2}-\frac{\left(x_{i2}-\mu_2\right)^2}{2\sigma_2^2}\right] \\
&\quad -\ln\left(2\pi\sigma_1\sigma_2\right)-\frac{\left(x_{42}-\mu_2\right)^2}{2\sigma_2^2}-\frac{\left(1+\mu_1^2\right)}{2\sigma_1^2}
\end{aligned}
$$

M 步：

$$\frac{\partial Q}{\partial \mu_1}=\sum_{i=1}^{3}\frac{\left(x_{i1}-\mu_1\right)}{\sigma_1^2}-\frac{\mu_1}{\sigma_1^2}=0$$

$$\mu_1=\frac{1}{4}\sum_{i=1}^{3}x_{i1}=\frac{3}{4}$$

$$\frac{\partial Q}{\partial \mu_2}=\sum_{i=1}^{3}\frac{\left(x_{i2}-\mu_2\right)}{\sigma_2^2}+\frac{\left(x_{42}-\mu_2\right)}{\sigma_2^2}=0$$

$$\mu_2=\frac{1}{4}\left(\sum_{i=1}^{3}x_{i1}+x_{42}\right)=\frac{8}{4}$$

$$\frac{\partial Q}{\partial \sigma_1}=\sum_{i=1}^{3}\left[-\frac{1}{\sigma_1}+\frac{\left(x_{i1}-\mu_1\right)^2}{\sigma_1^3}\right]-\frac{1}{\sigma_1}+\frac{\left(1+\mu_1^2\right)}{\sigma_1^3}=0$$

$$\sigma_1^2=\frac{1}{4}\left(\sum_{i=1}^{3}\left(x_{i1}-\mu_1\right)^2+\left(1+\mu_1^2\right)\right)=\frac{3.75}{4}$$

$$\frac{\partial Q}{\partial \sigma_2}=\sum_{i=1}^{3}\left[-\frac{1}{\sigma_2}+\frac{\left(x_{i2}-\mu_2\right)^2}{\sigma_2^3}\right]-\frac{1}{\sigma_2}+\frac{\left(x_{42}-\mu_2\right)^2}{\sigma_2^3}=0$$

$$\sigma_2^2=\frac{1}{4}\left(\sum_{i=1}^{3}\left(x_{i2}-\mu_2\right)^2+\left(x_{42}-\mu_2\right)\right)=\frac{8}{4}$$

- ## 一维高斯分布均值的贝叶斯估计

样本集 $D=\{x_1,\cdots,x_n\}$ 来自于 1 维高斯分布 $N\left(\mu,\sigma^2\right)$，其中方差 σ^2 是已知的，计算均值 μ 的贝叶斯估计。假设均值的先验 $p(\mu)\sim N\left(\mu_0,\sigma_0^2\right)$ 是以 μ_0 为均值，σ_0^2 为方差的高斯分布。

首先计算 μ 的后验概率密度 $p(\mu|D)$，根据贝叶斯公式：

$$p(\mu|D)=\frac{p(D|\mu)p(\mu)}{p(D)}=\frac{p(D|\mu)p(\mu)}{\int p(D|\mu)p(\mu)d\mu}$$

由于 $p(D)=\int p(D|\mu)p(\mu)d\mu$ 是与 μ 及 x 无关的常数，因此令：

$$\alpha=\frac{1}{p(D)}=\frac{1}{\int p(D|\mu)p(\mu)d\mu}$$

样本集 $D=\{x_1,\cdots,x_n\}$ 是独立同分布的，因此：

$$\begin{aligned}
p(\mu|D)&=\alpha p(D|\mu)p(\mu)\\
&=\alpha\prod_{i=1}^{n}p(x_i|\mu)p(\mu)\\
&=\alpha\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x_i-\mu)^2}{2\sigma^2}\right]\times\frac{1}{\sqrt{2\pi}\sigma_0}\exp\left[-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}\right]\\
&=\frac{\alpha}{\left(\sqrt{2\pi}\sigma\right)^n\sqrt{2\pi}\sigma_0}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2-\frac{1}{2\sigma_0^2}(\mu-\mu_0)^2\right]\\
&=\alpha'\exp\left[-\frac{1}{2}\left(\frac{1}{\sigma^2}\sum_{i=1}^{n}x_i^2-\frac{2}{\sigma^2}\mu\sum_{i=1}^{n}x_i+\frac{n}{\sigma^2}\mu^2+\frac{1}{\sigma_0^2}\mu^2-\frac{2}{\sigma_0^2}\mu_0\mu+\frac{1}{\sigma_0^2}\mu_0^2\right)\right]\\
&=\alpha'\exp\left[-\frac{1}{2}\left(\frac{1}{\sigma^2}\sum_{i=1}^{n}x_i^2+\frac{1}{\sigma_0^2}\mu_0^2\right)\right]\exp\left\{-\frac{1}{2}\left[\left(\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}\right)\mu^2-2\left(\frac{1}{\sigma^2}\sum_{i=1}^{n}x_i+\frac{\mu_0}{\sigma_0^2}\right)\mu\right]\right\}\\
&=\alpha''\exp\left\{-\frac{1}{2}\left[\left(\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}\right)\mu^2-2\left(\frac{1}{\sigma^2}\sum_{i=1}^{n}x_i+\frac{\mu_0}{\sigma_0^2}\right)\mu\right]\right\}
\end{aligned}\qquad（1）$$

上述过程中分 2 次对与 μ 无关项进行了归并，其中：

$$\alpha'=\frac{\alpha}{\left(\sqrt{2\pi}\sigma\right)^n\sqrt{2\pi}\sigma_0}，\quad\alpha''=\alpha'\exp\left[-\frac{1}{2}\left(\frac{1}{\sigma^2}\sum_{i=1}^{n}x_i^2+\frac{1}{\sigma_0^2}\mu_0^2\right)\right]$$

由上面的推导结果可以看出，$p(\mu|D)$ 的指数部分是关于 μ 的二次函数，由此可以断定 $p(\mu|D)$ 服从高斯分布：$p(\mu|D)\sim N\left(\mu_n,\sigma_n^2\right)$。

$$\begin{aligned}
p(\mu|D)&=\frac{1}{\sqrt{2\pi}\sigma_n}\exp\left[-\frac{1}{2}\left(\frac{\mu-\mu_n}{\sigma_n}\right)^2\right]\\
&=\frac{1}{\sqrt{2\pi}\sigma_n}\exp\left[-\frac{1}{2}\left(\frac{1}{\sigma_n^2}\mu^2-\frac{2\mu_n}{\sigma_n^2}\mu+\frac{\mu_n^2}{\sigma_n^2}\right)\right]
\end{aligned}\qquad（2）$$

对比（F.1）式和（F.2）式，可以得到：

$$\frac{1}{\sigma_n^2}=\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}$$

$$\frac{\mu_n}{\sigma_n^2}=\frac{1}{\sigma^2}\sum_{i=1}^{n}x_i+\frac{\mu_0}{\sigma_0^2}$$

因此：

$$\sigma_n^2=\frac{\sigma^2\sigma_0^2}{n\sigma_0^2+\sigma^2}$$

$$\mu_n=\left(\frac{1}{\sigma^2}\sum_{i=1}^{n}x_i+\frac{\mu_0}{\sigma_0^2}\right)\frac{\sigma^2\sigma_0^2}{n\sigma_0^2+\sigma^2}=\frac{\sigma_0^2}{n\sigma_0^2+\sigma^2}\sum_{i=1}^{n}x_i+\frac{\sigma^2\mu_0}{n\sigma_0^2+\sigma^2}$$

简化符号，令：$\hat{\mu}_n=\frac{1}{n}\sum_{i=1}^{n}x_i$，则有：

$$p(\mu|D)\sim N\left(\frac{n\sigma_0^2}{n\sigma_0^2+\sigma^2}\hat{\mu}_n+\frac{\sigma^2}{n\sigma_0^2+\sigma^2}\mu_0,\frac{\sigma^2\sigma_0^2}{n\sigma_0^2+\sigma^2}\right)$$

这就是 1 维高斯分布均值μ的贝叶斯估计后验概率密度。有了分布参数μ的后验概率$p(\mu|D)$，下面来计算待识样本x的后验概率：

$$\begin{aligned}
p(x|D)&=\int p(x|\mu)p(\mu|D)d\mu\\
&=\int\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]\frac{1}{\sqrt{2\pi}\sigma_n}\exp\left[-\frac{1}{2}\left(\frac{\mu-\mu_n}{\sigma_n}\right)^2\right]d\mu\\
&=\frac{1}{2\pi\sigma\sigma_n}\int\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2-\frac{1}{2}\left(\frac{\mu-\mu_n}{\sigma_n}\right)^2\right]d\mu\\
&=\frac{1}{2\pi\sigma\sigma_n}\int\exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma^2}-\frac{2x\mu}{\sigma^2}+\frac{\mu^2}{\sigma^2}+\frac{\mu^2}{\sigma_n^2}-\frac{2\mu\mu_n}{\sigma_n^2}+\frac{\mu_n^2}{\sigma_n^2}\right)\right]d\mu\\
&=\frac{1}{2\pi\sigma\sigma_n}\int\exp\left[-\frac{\left(\sigma_n^2+\sigma^2\right)\mu^2-2\left(\sigma_n^2x+\sigma^2\mu_n\right)\mu+\left(\sigma_n^2x^2+\sigma^2\mu_n^2\right)}{2\sigma^2\sigma_n^2}\right]d\mu\\
&=\frac{1}{2\pi\sigma\sigma_n}\exp\left[\frac{\left(\sigma_n^2x+\sigma^2\mu_n\right)^2}{2\sigma^2\sigma_n^2\left(\sigma_n^2+\sigma^2\right)}-\frac{\left(\sigma_n^2x^2+\sigma^2\mu_n^2\right)}{2\sigma^2\sigma_n^2}\right]\int\exp\left[-\frac{\sigma_n^2+\sigma^2}{2\sigma^2\sigma_n^2}\left(\mu-\frac{\sigma_n^2x+\sigma^2\mu_n}{\sigma_n^2+\sigma^2}\right)^2\right]d\mu\\
&=\frac{f(\sigma,\sigma_n)}{2\pi\sigma\sigma_n}\exp\left[\frac{\sigma_n^4x^2+2\sigma_n^2\sigma^2\mu_nx+\sigma^4\mu_n^2-\sigma_n^4x^2-\sigma^2\sigma_n^2x^2-\sigma_n^2\sigma^2\mu_n^2-\sigma^4\mu_n^2}{2\sigma^2\sigma_n^2\left(\sigma_n^2+\sigma^2\right)}\right]\\
&=\frac{f(\sigma,\sigma_n)}{2\pi\sigma\sigma_n}\exp\left[\frac{-\sigma^2\sigma_n^2x^2+2\sigma^2\sigma_n^2\mu_nx-\sigma^2\sigma_n^2\mu_n^2}{2\sigma^2\sigma_n^2\left(\sigma_n^2+\sigma^2\right)}\right]\\
&=\frac{f(\sigma,\sigma_n)}{2\pi\sigma\sigma_n}\exp\left[-\frac{1}{2}\frac{x^2-2x\mu_n+\mu_n^2}{\sigma_n^2+\sigma^2}\right]\\
&=\frac{f(\sigma,\sigma_n)}{2\pi\sigma\sigma_n}\exp\left[-\frac{1}{2}\frac{\left(x-\mu_n\right)^2}{\sigma_n^2+\sigma^2}\right]
\end{aligned}$$

其中的积分项简记为关于σ和σ_n的函数形式：

$$f(\sigma,\sigma_n)=\int\exp\left[-\frac{\sigma_n^2+\sigma^2}{2\sigma^2\sigma_n^2}\left(\mu-\frac{\sigma_n^2x-\sigma^2\mu_n}{\sigma_n^2+\sigma^2}\right)^2\right]d\mu \qquad (3)$$

注意到被积函数是关于μ的二次指数函数，因此是一个高斯函数，而（3）式为高斯积分，其值的大小只与σ和σ_n有关，与μ，x和μ_n无关。根据上面的推导结果可以看出，样本x的后验概率密度$p\left(x\middle|D\right)$服从高斯分布：

$$p\left(x\middle|D\right) \sim N\left(\mu_n, \sigma_n^2 + \sigma^2\right)$$

$f\left(\sigma, \sigma_n\right)$只是一个归一化因子，不需要计算积分即可得到：

$$f\left(\sigma, \sigma_n\right) = \frac{\sqrt{2\pi}\sigma\sigma_n}{\sqrt{\sigma_n^2 + \sigma^2}}$$