EM algorithm for t-distribution

Anders G. Kristensen

09-11-2022

1 Full model and likelihood

Given $Y = (X, W) \in \mathbb{R} \times (0, \infty)$ with joint density

$$f(x, w) = \frac{1}{\sqrt{\pi\nu\sigma^2} 2^{(\nu+1)} \Gamma(\frac{\nu}{2})} w^{\frac{\nu-1}{2}} e^{-\frac{w}{2}\left(1+\frac{(x-\mu)^2}{\nu\sigma^2}\right)}$$

With location parameter, $\mu \in \mathbb{R}$, scale parameter $\sigma > 0$ and shape parameter $\nu > 0$.

Theorem 1.1. *X has the t-distribution as its marginal density*

Proof. By definition the marginal density of X is given as

$$g(x) = \int f(x, w) dw = \int \frac{1}{\sqrt{\pi\nu\sigma^2} 2^{(\nu+1)/2} \Gamma(\frac{\nu}{2})} w^{\frac{\nu-1}{2}} e^{-\frac{w}{2}\left(1+\frac{(x-\mu)^2}{\nu\sigma^2}\right)} dw$$

$$= \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi\nu\sigma^2} 2^{(\nu+1)/2} \Gamma(\frac{\nu}{2}) \left(\frac{1}{2}\left(1+\frac{(x-\mu)^2}{\nu\sigma^2}\right)\right)^{\left(\frac{\nu+1}{2}\right)}} \int \frac{\frac{1}{2}\left(1+\frac{(x-\mu)^2}{\nu\sigma^2}\right)^{\left(\frac{\nu+1}{2}\right)}}{\Gamma(\frac{\nu+1}{2})} w^{\frac{\nu-1}{2}} e^{-\frac{w}{2}\left(1+\frac{(x-\mu)^2}{\nu\sigma^2}\right)} dw$$

$$= \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi\nu\sigma^2} \Gamma(\frac{\nu}{2}) \left(1+\frac{(x-\mu)^2}{\nu\sigma^2}\right)^{\left(\frac{\nu+1}{2}\right)}} = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu\sigma^2}} \left(1+\frac{(x-\mu)^2}{\nu\sigma^2}\right)^{-\left(\frac{\nu+1}{2}\right)}$$

$\square$

Consider the likelihood of the full data.

$$\sum_{i=1}^{N} \log\left(f(x_i, w_i)\right) = \sum_{i=1}^{N} \log\left(\frac{1}{\sqrt{\pi\nu\sigma^2} 2^{(\nu+1)} \Gamma(\frac{\nu}{2})} w_i^{\frac{\nu-1}{2}} e^{-\frac{w_i}{2}\left(1+\frac{(x_i-\mu)^2}{\nu\sigma^2}\right)}\right)$$

$$= \sum_{i=1}^{N} -\frac{1}{2}\log\left(\pi\nu\sigma^2\right) - (\nu+1)\log(2) - \log\left(\Gamma\left(\frac{\nu}{2}\right)\right) + \log(w_i)\left(\frac{\nu-1}{2}\right) - \frac{w_i}{2}\left(1+\frac{(x_i-\mu)^2}{\nu\sigma^2}\right)$$

$$= -\frac{N}{2}\log\left(\pi\nu\sigma^2\right) - N(\nu+1)\log(2) - N\log\left(\Gamma\left(\frac{\nu}{2}\right)\right) + \sum_{i=1}^{N} \log(w_i)\left(\frac{\nu-1}{2}\right) - \frac{w_i}{2}\left(1+\frac{(x_i-\mu)^2}{\nu\sigma^2}\right)$$

To get the maximum-likelihood estimator we set the gradient of the likelihood equal to zero. First we calculate the entries

$$\frac{\partial}{\partial\sigma^2} \sum_{i=1}^{N} \log\left(f(x_i, w_i)\right) = -\frac{N}{2\sigma^2} + \sum_{i=1}^{N} \frac{w_i(x_i-\mu)^2}{2\nu\sigma^4} \tag{1}$$

And

$$\frac{\partial}{\partial\mu}\sum_{i=1}^{N}\log\big(f(x_i,w_i)\big)=\sum_{i=1}^{N}\frac{w_i(x_i-\mu)}{\nu\sigma^2} \tag{2}$$

Setting (2) to zero gives

$$\sum_{i=1}^{N}w_i(x_i-\mu)=0$$
$$\mu=\bar{wx}$$

And from (2) we see

$$0=-N+\sum_{i=1}^{N}\frac{w_i(x_i-\mu)^2}{\nu\sigma^2}$$
$$\sigma^2=\frac{1}{\nu N}\sum_{i=1}^{N}w_i(x_i-\mu)^2$$

We are going to need the conditional distribution $W|X$. We see that $W|X$ is $\propto f(x,w)$, whence we see that $W|X\sim\Gamma\left(\frac{\nu+1}{2},\frac{1}{2}\left(1+\frac{(x-\mu)^2}{\nu\sigma^2}\right)\right)$. Note also that this implies

$$E\left[W|X\right]=\frac{\frac{\nu+1}{2}}{\frac{1}{2}\left(1+\frac{(x-\mu)^2}{\nu\sigma^2}\right)},\qquad E\left[\log\left(W\right)|X\right]=-\log\left(\frac{1}{2}\left(1+\frac{(x-\mu)^2}{\nu\sigma^2}\right)\right)+\psi\left(\frac{\nu+1}{2}\right)$$

2 The EM-algorithm

To implement the EM-algorithm we calculate the Q-function

$$\begin{aligned}Q(\theta|\theta')&=E_{\theta'}\left[\log\big(f(x,w)\big)\Big|X=x\right]\\&=-\frac{1}{2}\log\left(\pi\nu\sigma^2\right)+E_{\theta'}\left[\log\left(w\right)\big|X=x\right]\left(\frac{\nu-1}{2}\right)-\frac{1}{2}E_{\theta'}\left[w|X=x\right]\left(1+\frac{(x-\mu)^2}{\nu\sigma^2}\right)\end{aligned}$$

Where we have removed the terms that don't depend on σ^2,μ. Differentiating w.r.t. σ^2,μ yields

$$\frac{\partial}{\partial\mu}\big(Q(\theta|\theta')\big)=E_{\theta'}\left[w|X=x\right]\frac{(x-\mu)}{\nu\sigma^2}$$
$$\frac{\partial}{\partial\sigma^2}\big(Q(\theta|\theta')\big)=-\frac{1}{2\sigma^2}+\frac{1}{2}E_{\theta'}\left[w|X=x\right]\frac{(x-\mu)^2}{\nu\sigma^4}$$

Setting them each equal to zero gives

$$\frac{\partial}{\partial\mu}\big(Q(\theta|\theta')\big)=0$$
$$\hat{\mu}=\frac{\sum_{i=1}^{N}x_iE_{\theta'}\left[w_i|X=x_i\right]}{\sum_{i=1}^{N}E_{\theta'}\left[w_i|X=x_i\right]}$$

And

$$\begin{aligned}\frac{\partial}{\partial\sigma^2}\big(Q(\theta|\theta')\big) &= 0\\ N &= \sum_{i=1}^{N} E_{\theta'}\left[w_i|X=x_i\right]\frac{(x_i-\mu)^2}{\nu\sigma^2}\\ \hat{\sigma}^2 &= \frac{1}{N\nu}\sum_{i=1}^{N} E_{\theta'}\left[w_i|X=x_i\right](x_i-\mu)^2\end{aligned}$$

3 The Fisher-information

We find the fisher-information theoretically first

$$\frac{\partial}{\partial\mu^2}\big(Q(\theta|\theta')\big) = -\frac{E_{\theta'}\left[w|X=x\right]}{\nu\sigma^2}$$

$$\frac{\partial}{\partial\mu\sigma^2}\big(Q(\theta|\theta')\big) = -\frac{E_{\theta'}\left[w|X=x\right](x-\mu)}{\nu\sigma^2}$$

$$\frac{\partial}{\partial(\sigma^2)^2}\big(Q(\theta|\theta')\big) = \frac{1}{2\sigma^4} - E_{\theta'}\left[w|X=x\right]\frac{(x-\mu)^2}{\nu\sigma^6}$$