EM algorithm — Maximum likelihood estimation of the non-standard t-distribution

Lucas Støjko Andersen

October 13, 2022

1 The non-standard t-distribution

The non-standard t-distribution has density

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

Maximum likelihood estimates have no closed analytic solutions. It has negative log-likelihood

$$\begin{aligned} &- \log \Gamma\left(\frac{\nu+1}{2}\right) + \log \Gamma(\nu/2) + \frac{1}{2}\log \nu + \frac{1}{2}\log \sigma^2 \\ &+ \frac{\nu+1}{2}\log\left(1 + \frac{(x-\mu)^2}{\nu\sigma^2}\right). \end{aligned}$$

For n observations the negative mean log-likelihood is

$$\begin{aligned} &- \log \Gamma\left(\frac{\nu+1}{2}\right) + \log \Gamma(\nu/2) + \frac{1}{2}\log \nu + \frac{1}{2}\log \sigma^2 \\ &+ \frac{\nu+1}{2n}\sum_{i=1}^{n}\log\left(1 + \frac{(x_i-\mu)^2}{\nu\sigma^2}\right). \end{aligned}$$

We have the gradients

$$\begin{aligned} \nabla_\mu \ell &= -\frac{\nu+1}{n\nu\sigma^2}\sum_{i=1}^{n}\frac{x_i-\mu}{1+\frac{(x_i-\mu)^2}{\nu\sigma^2}} \\ \nabla_{\sigma^2}\ell &= \frac{1}{2\sigma^2} - \frac{\nu+1}{2n\nu(\sigma^2)^2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{1+\frac{(x_i-\mu)^2}{\nu\sigma^2}} \end{aligned}$$

$$\nabla_\nu \ell = -\frac{1}{2}\psi\left(\frac{\nu+1}{2}\right) + \frac{1}{2}\psi(\nu/2) + \frac{1}{2\nu} + \frac{1}{2n}\sum_{i=1}^{n} \log\left(1 + \frac{(x_i-\mu)^2}{\nu\sigma^2}\right) - \frac{\nu+1}{2n\nu^2\sigma^2}\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{1+\frac{(x_i-\mu)^2}{\nu\sigma^2}}$$

We have the Hessian for (μ, σ^2)

$$\nabla_\mu^2 \ell = -\frac{\nu+1}{2n\nu\sigma^2}\sum_{i=1}^{n} \frac{1}{1+\frac{(x_i-\mu)^2}{\nu\sigma^2}} + \frac{\nu+1}{n(\nu\sigma^2)^2}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{1+\frac{(x_i-\mu)^2}{\nu\sigma^2}}\right)^2$$

$$\nabla_{\sigma^2}^2 \ell = -\frac{1}{2(\sigma^2)^2} + \frac{\nu+1}{n\nu(\sigma^2)^3}\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{1+\frac{(x_i-\mu)^2}{\nu\sigma^2}} - \frac{\nu+1}{2n\nu^2(\sigma^2)^4}\sum_{i=1}^{n}\left(\frac{(x_i-\mu)^2}{1+\frac{(x_i-\mu)^2}{\nu\sigma^2}}\right)^2$$

$$\nabla_\mu\nabla_{\sigma^2} \ell = \frac{\nu+1}{n\nu(\sigma^2)^2}\sum_{i=1}^{n} \frac{x_i-\mu}{1+\frac{(x_i-\mu)^2}{\nu\sigma^2}} - \frac{\nu+1}{n\nu(\sigma^2)^4}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{1+\frac{(x_i-\mu)^2}{\nu\sigma^2}}\right)^2 (x_i-\mu)$$

2 Expectation Maximization

Consider (X, W) with joint density

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

Then $W \mid X$ has density

$$h(w \mid x) \propto w^{\frac{\nu+1}{2}-1} e^{-\frac{w}{2}\left(1+\frac{(x-\mu)^2}{\nu\sigma^2}\right)} \tag{3}$$

and so $W \mid X \sim \Gamma(\alpha, \beta)$ where

$$\alpha = \frac{\nu+1}{2} \qquad \beta = \frac{1}{2}\left(1 + \frac{(X-\mu)^2}{\nu\sigma^2}\right). \tag{4}$$

Therefore

$$E(W \mid X) = \frac{\alpha}{\beta} = \frac{\nu+1}{1+\frac{(X-\mu)^2}{\nu\sigma^2}} \tag{5}$$

and

$$E(\log W \mid X) = \psi(\alpha) - \log(\beta) = \psi\left(\frac{\nu+1}{2}\right) - \log\left(\frac{1}{2}\left(1+\frac{(X-\mu)^2}{\nu\sigma^2}\right)\right) \tag{6}$$

where ψ is the digamma function

$$\psi(x) = \frac{d}{dx} \log \Gamma(x). \tag{7}$$

It can easily be shown that the marginal distribution of X has a non-standard t-distribution by multiplying the joint density by $\Gamma(\alpha)/\beta^\alpha$.

2.1 Expectation step

Define H as the negative log-likelihood of the joint density:

$$\begin{aligned} H(\theta) =& \frac{1}{2} \log \pi + \frac{1}{2} \log \nu + \frac{1}{2} \log \sigma^2 + \frac{\nu + 1}{2} \log 2 + \log \Gamma(\nu/2) \\ &- \frac{\nu - 1}{2} \log w + \frac{w}{2} - \frac{w}{2} \frac{(x-\mu)^2}{\nu \sigma^2} \end{aligned}$$

where $\theta = (\mu, \sigma^2, \nu)$. Define $Q(\theta \mid \theta') = E_{\theta'}(H(X, W)|X)$ we have

$$\begin{aligned} Q(\theta \mid \theta') =& \frac{1}{2} \log \pi + \frac{1}{2} \log \nu + \frac{1}{2} \log \sigma^2 + \frac{\nu + 1}{2} \log 2 + \log \Gamma(\nu/2) \\ &- \frac{\nu - 1}{2} C_1(\theta', X) \\ &+ \frac{C_2(\theta', X)}{2} + C_2(\theta', X) \frac{(X-\mu)^2}{2\nu\sigma^2} \end{aligned}$$

where

$$C_1(\theta, x) = \psi\left(\frac{\nu+1}{2}\right) - \log\left(\frac{1}{2} + \frac{(x-\mu)^2}{2\nu\sigma^2}\right) \quad C_2(\theta, x) = \frac{\nu + 1}{1 + \frac{(x-\mu)^2}{\nu\sigma^2}}.$$

2.2 Maximization (minimization) step

For n observations (X_i) we seek to minimize

$$Q(\theta \mid \theta') = \frac{1}{n} \sum_{i=1}^{n} Q_i(\theta \mid \theta')$$

in θ. Getting rid of the terms that do not contain θ

$$\begin{aligned} Q(\theta \mid \theta') \simeq& \log \nu + \log \sigma^2 + \nu \log 2 + 2 \log \Gamma(\nu/2) \\ &- \frac{\nu}{n} \sum_{i=1}^{n} C_1(\theta', X_i) + \frac{1}{n\nu\sigma^2} \sum_{i=1}^{n} C_2(\theta', X_i)(X_i - \mu)^2. \end{aligned}$$

The gradient is

$$\partial_\mu Q(\theta \mid \theta') = -\frac{2}{n\nu\sigma^2}\sum_{i=1}^{n} C_2(\theta', X_i)(X_i - \mu)$$
$$\partial_{\sigma^2} Q(\theta \mid \theta') = \frac{1}{\sigma^2} - \frac{1}{n\nu(\sigma^2)^2}\sum_{i=1}^{n} C_2(\theta', X_i)(X_i - \mu)^2$$
$$\partial_\nu Q(\theta \mid \theta') = \frac{1}{\nu} + \log 2 + \psi(\nu/2) - \frac{1}{n}\sum_{i=1}^{n} C_1(\theta', X_i) - \frac{1}{n\nu^2\sigma^2}\sum_{i=1}^{n} C_2(\theta', X_i)(X_i - \mu)^2.$$

Fixed ν

For fixed ν the minimization problem is equivalent to minimizing weighted least squares. Hence the minimizers are

$$\hat{\mu}(\theta') := \frac{\sum_{i=1}^{n} C_2(\theta', X_i)X_i}{\sum_{i=1}^{n} C_2(\theta', X_i)}, \qquad \hat{\sigma}^2_\nu(\theta') := \frac{1}{n\nu}\sum_{i=1}^{n} C_2(\theta', X_i)(X_i - \hat{\mu}(\theta))^2. \tag{8}$$

Therefore the EM-algorithm is described as the updating scheme

$$\theta_{n+1} = (\hat{\mu}(\theta_n), \hat{\sigma}^2_\nu(\theta_n)). \tag{9}$$

Estimating ν

When ν is no longer fixed then there no longer exists nice closed form solutions. However, if any global minimizer exists then $\hat{\mu}(\theta')$ must be part of the solution.