EM algorithm

Maximum Likelihood Estimation of the t-distribution

Lucas Støjko Andersen

University of Copenhagen

October 18, 2022

The t-distribution

The density of the non-standard t-distribution

x The density of the t-distribution is given by

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

with parameters $\mu \in \mathbf{R}$, $\nu > 0$, $\sigma^2 > 0$. Here, Γ is the gamma function.

For independent observations $X_1, X_2, \ldots, X_n$ with density as in (1) there exist no nice analytic solutions to the MLE — even with ν known.

There is another way

A joint approach with latent variables

Consider (X, W) with density

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

Notice that $W \mid X = x \sim \Gamma(\alpha, \beta)$ — the gamma distribution — with parameters

$$\alpha = \frac{\nu+1}{2} \qquad \beta = \frac{1}{2}\left(1 + \frac{(x-\mu)^2}{\nu\sigma^2}\right). \tag{3}$$

This can be used to show that X indeed has the marginal density of the t-distribution with parameters μ, σ^2, ν.

The Full Maximum Likelihood Estimate

MLE with fixed ν

Assume $\nu > 0$ known. I.i.d. $(X_1, W_1), (X_2, W_2), \ldots, (X_n, W_n)$ have log-likelihood

$$\ell(\theta) \simeq -\frac{n}{2} \log \sigma^2 - \frac{1}{2\nu\sigma^2} \sum_{i=1}^{n} W_i (X_i - \mu)^2. \tag{4}$$

The solution is that of the weighted least squares:

$$\hat{\mu} = \frac{\sum_{i=1}^n W_i X_i}{\sum_{i=1}^n W_i}, \qquad \hat{\sigma}^2 = \frac{1}{n\nu} \sum_{i=1}^{n} W_i (X_i - \hat{\mu})^2. \tag{5}$$

The Full Maximum Likelihood Estimate

MLE with fixed ν — implemented in R

We can sample (X, W) by sampling W from a χ^2_ν distribution and then sampling X from a $\mathcal{N}(\mu, \nu\sigma^2/W)$.

```
simulate_X <- function(N, mu, sigma, nu) {
    W <- rchisq(N, nu)
    X <- rnorm(N, mu, sqrt(nu * sigma / W))

    list(x = X, w = W)
}

full_mle <- function(X, W, nu) {
    mu <- sum(X * W) / sum(W)
    sigma <- sum(W * (X - mu)^2) / (nu * (length(X)))
    list(mu = mu, sigma = sigma)
}

set.seed(3939392)
samples <- simulate_X(100000, 5, 1.5, 3)
full_mle(samples$x, samples$w, 3)
```

We obtain the estimates $\hat{\mu}_{\text{full}} = 4.997852$ and $\hat{\sigma}^2_{\text{full}} = 1.500886$.

EM algorithm

MLE of the marginal likelihood for fixed ν

Only X is observed and so the full MLE cannot be computed. Using the EM algorithm we iteratively maximize the quantity

$$Q(\theta \mid \theta') = E_{\theta'}(\log f_\theta(X, W) \mid X) \tag{6}$$

where $\theta = (\mu, \sigma^2)$. Using the log-likelihood from earlier

$$Q(\theta \mid \theta') \simeq -\frac{n}{2}\log\sigma^2 - \frac{1}{2\nu\sigma^2}\sum_{i=1}^{n} E_{\theta'}(W_i \mid X_i)(X_i - \mu)^2. \tag{7}$$

The maximizer is then

$$\hat{\mu}_{\theta'} = \frac{\sum_{i=1}^{n} E_{\theta'}(W_i \mid X_i)X_i}{\sum_{i=1}^{n} E_{\theta'}(W_i \mid X_i)}, \quad \hat{\sigma}^2_{\theta'} = \frac{1}{n\nu}\sum_{i=1}^{n} E_{\theta'}(W_i \mid X_i)(X_i - \hat{\mu})^2.$$

EM algorithm

E-step and M-step

The E-step boils down to computing $E_{\theta'}(W_i \mid X_i)$

$$E_{\theta'}(W_i \mid X_i) = \frac{\alpha}{\beta} = \frac{\nu' + 1}{2} \frac{1}{\frac{1}{2}\left(1 + \frac{(X_i - \mu')^2}{\nu' \sigma'^2}\right)} = \frac{\nu' + 1}{1 + \frac{(X_i - \mu')^2}{\nu' \sigma'^2}} \tag{8}$$

and doing the M-step by computing

$$\hat{\mu}_{\theta'} = \frac{\sum_{i=1}^n E_{\theta'}(W_i \mid X_i) X_i}{\sum_{i=1}^n E_{\theta'}(W_i \mid X_i)}, \quad \hat{\sigma}^2_{\theta'} = \frac{1}{n\nu} \sum_{i=1}^n E_{\theta'}(W_i \mid X_i)(X_i - \hat{\mu})^2.$$

EM algorithm

Implementation of marginal MLE

```
E_step <- function(x, nu) {
    force(x)
    force(nu)
    function(par) {
        # mu = par[1]
        # sigma^2 = par[2]
        (nu + 1) / (1 + ((x - par[1])^2) / (nu * par[2]))
    }
}

M_step <- function(x, nu) {
  force(x)
  force(nu)
  function(EW) {
    mu <- sum(EW * x) / sum(EW)
    sigma <- mean(EW * (x - mu)^2) / nu
    c(mu, sigma)
  }
}
```

EM algorithm

Implementation of marginal MLE

```r
EM <- function(par, x, nu, maxit = 500, min.eps = 1e-7) {
    E <- E_step(x, nu)
    M <- M_step(x, nu)
    for(i in 1:maxit) {
        EW <- E(par)
        new_par <- M(EW)
        if(sum((new_par - par)^2) < min.eps * (sum(par^2) + min.eps)) {
            par <- new_par
            break
        }
        par <- new_par
        if(i == maxit) warning("Maximum number of itertaions reached.")
    }
    names(par) <- c("mu", "sigma")
    list(par = c(par, nu = nu), iterations = i)
}
```

EM algorithm

Comparison of marginal MLE and full MLE

For starting values $\theta' = (1, 2)$ we obtain the estimates

$$\hat{\mu}_{\mathsf{EM}} = 4.995958, \qquad \hat{\sigma}^2_{\mathsf{EM}} = 1.504293, \qquad \nu = 3. \tag{9}$$

in 9 iterations of the EM algorithm. Recall the full MLE

$$\hat{\mu}_{\mathsf{full}} = 4.997852, \qquad \hat{\sigma}^2_{\mathsf{full}} = 1.500886, \qquad \nu = 3. \tag{10}$$

Robustness of the initial parameters

Properties of the t-distribution

The t-distribution has first moment if $\nu > 1$ and second moment if $\nu > 2$. When the moments exist then

$$EX = \mu, \qquad VX = \sigma^2 \frac{\nu}{\nu + 2}. \tag{11}$$

The calculations in the E-step may be unstable

$$\hat{\mu}_{\theta'} = \frac{\frac{1}{n}\sum_{i=1}^{n} E_{\theta'}(W_i \mid X_i) X_i}{\frac{1}{n}\sum_{i=1}^{n} E_{\theta'}(W_i \mid X_i)}, \quad \hat{\sigma}^2_{\theta'} = \frac{1}{n\nu}\sum_{i=1}^{n} E_{\theta'}(W_i \mid X_i)(X_i - \hat{\mu})^2.$$

Robustness of the intial parameters

Gaussian resemblence of the t-distribution for large ν

When $\nu \longrightarrow \infty$ the density of the t-distribution approaches the density of a normal distribution with mean μ and variance σ^2. It does so very quickly!

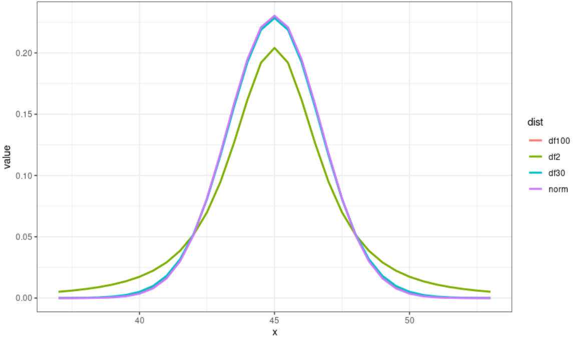

Figure: $\mu = 45$, $\sigma^2 = 3$.