Parameter Estimation of Heavy-tailed Random Walk Model from Incomplete Data

Junyan Liu, Sandeep Kumar, and Daniel P. Palomar

Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Hong Kong

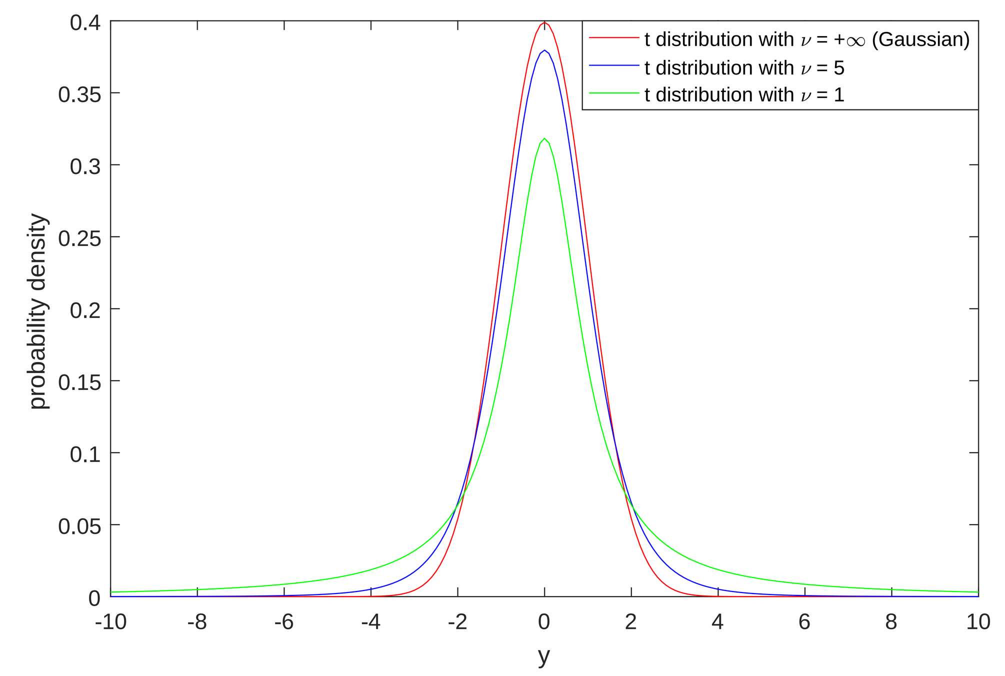
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Background

- In the recent era of data deluge, many applications collect and process large amount of time series data for inference, learning, parameter estimation and decision making.
- Missing values frequently occur in the data recording process, e.g., some stocks may suffer a lack of liquidity resulting in no price recorded, observation devices like sensors breakdown, and weather or other conditions disturb sample taking schemes.
- Traditionally, the parameter estimation for time series from incomplete data has been considered under Gaussian noise. However, many real-world data follow heavy-tailed distributions, e.g., financial time series, brain fMRI, and animals movement.

Heavy-tailed Random Walk Model

- Student's t-distribution: $f_t\left(y;\mu,\sigma^2,\nu\right)=\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\sigma\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{(y-\mu)^2}{\nu\sigma^2}\right)^{-\frac{\nu+1}{2}}$.



- A univariate time series $Y_1, Y_2, \ldots, Y_T$ that follows a Student's t random walk model:

$$Y_t - Y_{t-1} \overset{i.i.d.}{\sim} t\left(\mu,\sigma^2,\nu\right).$$

Problem Formulation

- An observation of this time series with D missing blocks:

$$y_1,\ldots,y_{t_1},\text{NA},\ldots,\text{NA},y_{t_1+n_1+1},\ldots y_{t_d},\text{NA},\ldots,\text{NA},$$
$$y_{t_d+n_d+1},\ldots,y_{t_D},\text{NA},\ldots,\text{NA},y_{t_D+n_D+1},\ldots,y_T.$$

- Maximum likelihood estimation (MLE) problem for μ, σ^2, and ν:

$$\underset{\mu,\sigma^2,\nu>0}{\text{maximize}}\ \ l^{obs}\left(\{y_t\}_{t\in C_{obs}}|\mu,\sigma^2,\nu\right), \quad (1)$$

where

$$l^{obs}\left(\{y_t\}_{t\in C_{obs}}|\mu,\sigma^2,\nu\right)$$
$$=\log\left(\prod_{d=0}^{D}\prod_{t=t_d+n_d+2}^{t_{d+1}} f_t\left(y_t;\mu+y_{t-1},\sigma^2,\nu\right)\right)$$
$$+\log\left(\prod_{d=1}^{D}\int\!\!\cdots\!\!\int\prod_{t=t_d+1}^{t_d+n_d+1} f_t\left(y_t;\mu+y_{t-1},\sigma^2,\nu\right)\mathrm{d}y_{t_d+1}\cdots\mathrm{d}y_{t_d+n_d}\right).$$

- The objective function involves multiple integrals, and has no closed-form expression. It is difficult to optimize this problem directly.

Expectation-Maximization Algorithm

- A general MLE problems with missing data $\mathbf{Z}$, observed data $\mathbf{X}$, parameter $\boldsymbol{\theta}$:

$$\max_{\boldsymbol{\theta}}\ \log p(\mathbf{X}|\boldsymbol{\theta}).$$

- EM algorithm is a very popular iterative algorithm to solve this kind of problem.
  - E step: $Q\left(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}\right)=\mathrm{E}_{\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(k)}}\log p\left(\mathbf{X},\mathbf{Z}|\boldsymbol{\theta}\right)$.
  - M step: $\boldsymbol{\theta}^{(k+1)}=\arg\max_{\boldsymbol{\theta}} Q\left(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}\right)$.

Stochastic Expectation-Maximization Algorithm

- Stochastic EM (when E step is intractable)
  - Simulation step (S step): draw a realization $\mathbf{Z}^{(k)}\sim\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(k)}$.
  - M step: $\boldsymbol{\theta}_1^{(k+1)}=\arg\max_{\boldsymbol{\theta}}\log p\left(\mathbf{X},\mathbf{Z}^{(k)}|\boldsymbol{\theta}\right)$, and $\boldsymbol{\theta}^{(k+1)}=\left(1-\gamma^{(k)}\right)\boldsymbol{\theta}_1^{(k+1)}+\gamma^{(k)}\boldsymbol{\theta}^{(k)}$, where $\sum\gamma^{(k)}=\infty$ and $\sum\left(\gamma^{(k)}\right)^2<\infty$.

Gaussian Mixture Representation of Student's t-Distribution

- The Student's t-distribution can be regarded as a Gaussian mixture.

$$Y_t-Y_{t-1}\overset{i.i.d.}{\sim}t\left(\mu,\sigma^2,\nu\right)\Longleftarrow\begin{cases}Y_t-Y_{t-1}|\mu,\sigma^2,\tau_t\sim\mathcal{N}\left(\mu,\sigma^2/\tau_t\right)\\ \tau_t\sim Gamma\left(\nu/2,\ \nu/2\right)\end{cases}$$

- We regard $\{\tau_t\}$ and $\{y_t\}_{t\in C_{mis}}$ as missing variables, and apply the stochastic EM algorithm to solve problem (1).
- If we only regard $\{y_t\}_{t\in C_{mis}}$ as missing data, there would be no closed-form maximizer in the M step.

Posterior Distribution of Missing Data

- The posterior distribution of missing data (complicated):

$$f\left(y_{t_d+1},y_{t_d+2},\ldots,y_{t_d+n_d},\tau_{t_d+1},\tau_{t_d+2},\ldots,\tau_{t_d+n_d+1}|\left\{y_{t\in C_{obs}}\right\},\mu^k,\left(\sigma^k\right)^2,\nu^k\right)$$
$$\propto\prod_{t=t_d+1}^{t_d+n_d+1}\tau_t^{\frac{\nu^k-1}{2}}\exp\left(-\frac{\nu^k}{2}\tau_t-\frac{\left(\sigma^k\right)^{-2}\left(y_t-y_{t-1}-\mu^k\right)^2}{2}\tau_t\right).$$

- Gibbs sampling: instead of drawing the all components of the missing data jointly, draws realizations of each component sequentially based on its distribution conditional on all the other components.
- Conditional distributions (much simpler):

$$Y_t|\mu^{(k)},\sigma^{(k)},\nu^{(k)},\mathbf{Y}_{-t},\{\tau_t\}\sim\mathcal{N}\left(\frac{\tau_t\left(\mu^{(k)}+y_{t-1}\right)+\tau_{t+1}\left(y_{t+1}-\mu^{(k)}\right)}{\tau_t+\tau_{t+1}},\frac{\left(\sigma^{(k)}\right)^2}{\tau_t+\tau_{t+1}}\right),$$

$$\tau_t|\mu^{(k)},\sigma^{(k)},\nu^{(k)},\{y_t\},\mathcal{T}_{-t}\sim\text{Gamma}\left(\frac{\nu^{(k)}+1}{2},\frac{\left(\sigma^{(k)}\right)^{-2}\left(y_t-\mu^{(k)}-y_{t-1}\right)^2+\nu^{(k)}}{2}\right),$$

where $\mathcal{T}_{-t}$ is the set of all the mixture weights except τ_t,and $\mathbf{Y}_{-t}$ is the set of all the samples except y_t.

M step

- The resulting log-likelihood of the simulated complete data is

$$l\left(\left\{y_t^{(k)}\right\},\left\{\tau_t^{(k)}\right\}|\mu,\sigma^2,\nu\right)$$
$$=\sum_{t=2}^{T}\left\{-\frac{\tau_t^{(k)}}{2\sigma^2}\left(y_t^{(k)}-\mu-y_{t-1}^{(k)}\right)^2-\log\left(\sigma\right)\right.$$
$$\left.-\frac{\nu}{2}\tau_t^{(k)}+\frac{\nu}{2}\log\left(\frac{\nu}{2}\right)+\frac{\nu-1}{2}\log\left(\tau_t^{(k)}\right)-\log\left(\Gamma\left(\frac{\nu}{2}\right)\right)\right\}.$$

- Maximizer:

$$\mu_1^{(k+1)}=\frac{\sum_{t=2}^{T}\tau_t^{(k)}\left(y_t^{(k)}-y_{t-1}^{(k)}\right)}{\sum_{t=2}^{T}\tau_t^{(k)}}, \quad (2)$$

$$\left(\sigma_1^{(k+1)}\right)^2=\frac{\sum_{t=2}^{T}\tau_t^{(k)}\left(y_t^{(k)}-\mu^{(k+1)}-y_{t-1}^{(k)}\right)^2}{T-1}, \quad (3)$$

$$\nu_1^{(k+1)}=\arg\max_{\nu>0}l\left(\left\{y_t^{(k)}\right\},\left\{\tau_t^{(k)}\right\}|\mu^{(k)},\left(\sigma^k\right)^2,\nu\right). \quad (4)$$

Algorithm

1. Initialize $\mu^{(0)}$ and $\sigma^{(0)}$ as an arbitrary number, $\nu^{(0)}$ as an arbitrary positive number, and $k=0$.
2. Draw one realization $\left\{y_t^{(k)}\right\}_{t\in C_{mis}}$ and $\left\{\tau_t^{(k)}\right\}$ via Gibbs sampling method.
3. Compute $\mu_1^{(k+1)}$, $\left(\sigma_1^{(k+1)}\right)^2$, and $\nu_1^{(k+1)}$ according to (2)-(4), and then update

$$\mu^{(k+1)}=\mu^{(k)}+\gamma^{(k)}\left(\mu_1^{(k+1)}-\mu^{(k)}\right),$$
$$\left(\sigma^2\right)^{(k+1)}=\left(\sigma^2\right)^{(k)}+\gamma^{(k)}\left(\left(\sigma_1^2\right)^{(k+1)}-\left(\sigma^2\right)^{(k)}\right),$$
$$\nu^{(k+1)}=\nu^{(k)}+\gamma^{(k)}\left(\nu_1^{(k+1)}-\nu^{(k)}\right),$$

where $\gamma^{(k)}=\frac{1}{k}$.

4. Return to step 2 or stop if the stopping criterion is satisfied.

Numerical Results

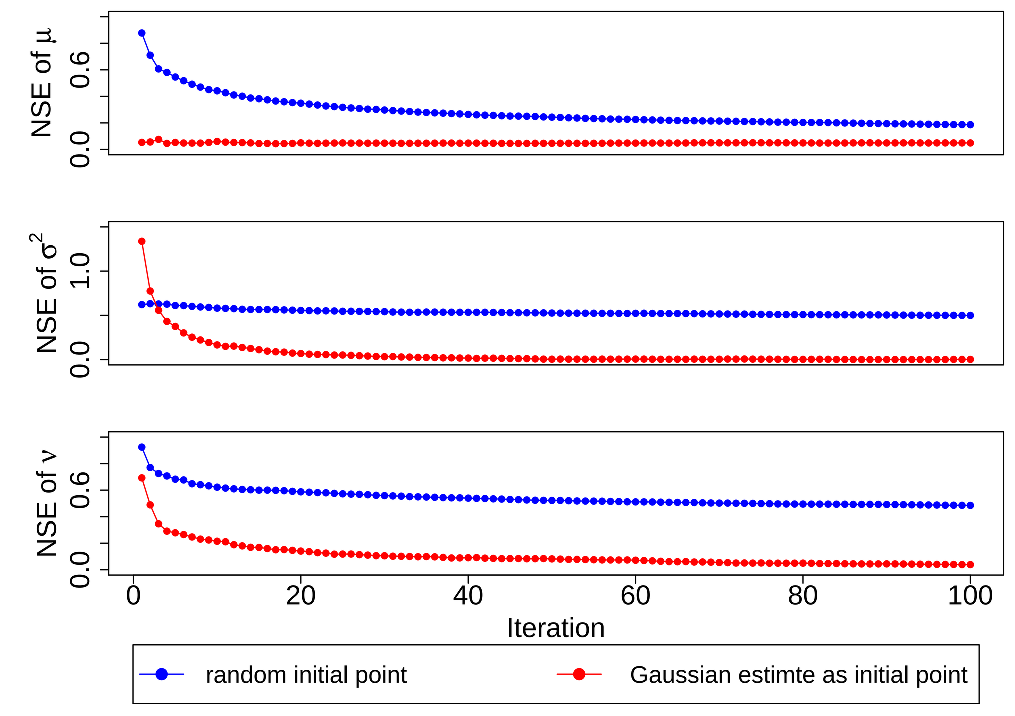


Figure 1: Estimation errors of parameters versus iterations.

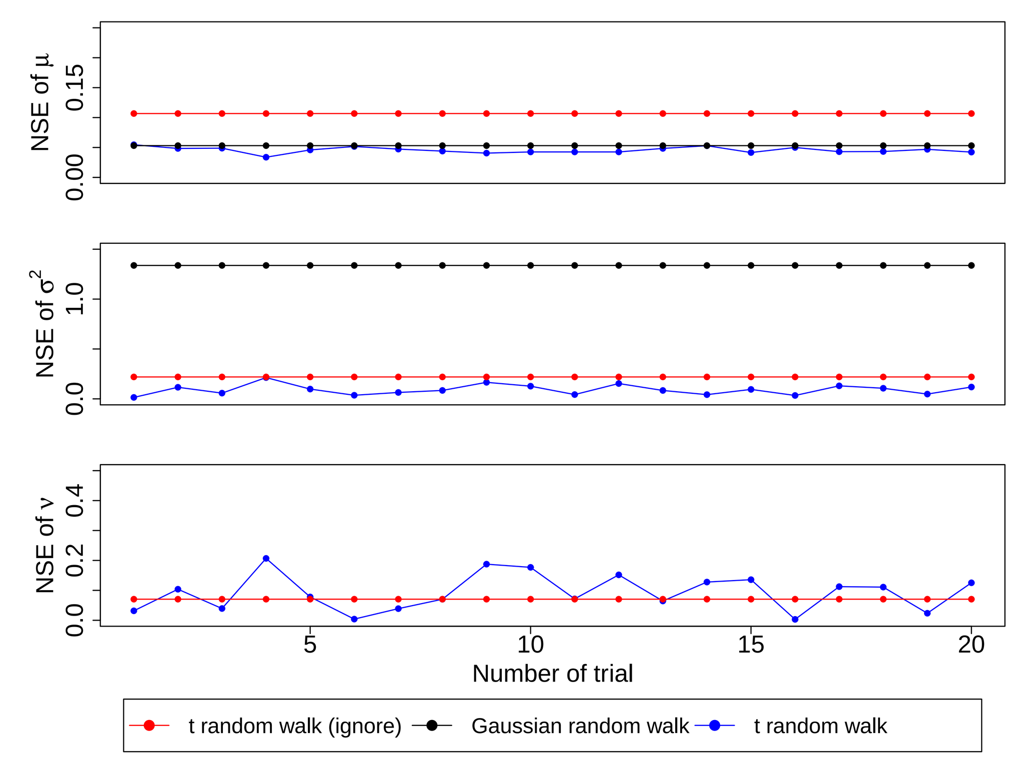


Figure 2: Final estimation errors of three different estimation approaches. t random walk (ignore) means MLE of Student's t random walk by ignoring the missing values (only use the available differences bwtween adjacent samples).

Conclusion

For heavy-tailed data, the traditional methods based on Gaussian distribution are too inefficient, and significant performance gain can be achieved by designing the algorithm under heavy-tailed model.

Email: jliubl@ust.hk