Robust and accurate inference via a mixture of Gaussian and terrors

Hyungsuk Tak

ICTS/SAMSI Time Series Analysis for Synoptic Surveys and Gravitational Wave Astronomy

22 March 2017

Joint work with Justin Ellis (Caltech, JPL)

Motivation

- Gaussian error, $\epsilon_j \sim \mathrm{N}_1(0,\ V_j)$

  - Pros : (Thin tails) Efficient inference when $\nexists$ outliers.
  - Cons: Can bias inference when $\exists$ outliers.

- Student's t_ν error, $\epsilon_j \sim V_j^{0.5} t_\nu$

  - Pros: (Heavy tails) Robust to outliers.
  - Cons: Can be less efficient due to unnecessarily heavy tails for most of the normally observed data.

- Why not use both? A mixture of Gaussian and Student's t_ν errors,

$$\begin{aligned} \epsilon_j &\sim \mathrm{N}_1(0,\ V_j) && \text{with probability } 1-\theta_j, \\ &\sim V_j^{0.5} t_\nu && \text{with probability } \theta_j, \end{aligned}$$

enables a robust & accurate estimation.

The mixture error

A p-dimensional mixture error with *'known'* (or very accurately estimated) variance component $\mathbf{V}_j$ is

$$
\begin{aligned}
\epsilon_j \mid z_j, \alpha_j &\sim \mathrm{N}_p(\mathbf{0},\ {\alpha_j}^{z_j}\mathbf{V}_j),\\
z_j \mid \theta_j &\sim \mathrm{Bernoulli}(\theta_j),\\
\theta_j &\sim \mathrm{Uniform}(0,1),\\
\alpha_j &\sim \mathrm{Inv.Gamma}(\nu/2,\ \nu/2).
\end{aligned}
$$

- z_j, a latent outlier indicator.
- θ_j, probability of datum j being an outlier.
- α_j, an auxiliary variable to express t_ν as a scale mixture of Gaussian.

  E.g., if $x \mid \alpha \sim \mathrm{N}(0, \alpha)$ & $\alpha \sim \mathrm{Inv.Gamma}(\nu/2, \nu/2)$, then $x \sim t_\nu$.

  (West, 1987; Peel and McLachlan, 2000; Gelman et al., 2013)

Relationship with other errors

$$
\begin{aligned}
\boldsymbol{\epsilon}_j \mid z_j, \alpha_j &\sim \mathrm{N}_p(\mathbf{0},\ \alpha_j^{z_j}\mathbf{V}_j),\\
z_j \mid \theta_j &\sim \mathrm{Bernoulli}(\theta_j),\\
\theta_j &\sim \mathrm{Uniform}(0,1),\\
\alpha_j &\sim \text{inverse-Gamma}(\nu/2,\ \nu/2).
\end{aligned}
$$

This proposed mixture error is marginally equivalent to

- a Gaussian error if $z_j = 0$ for all j.
- a t_ν error if $z_j = 1$ for all j.
- a mixture of Gaussian & t_ν error.
- a mixture of two Gaussians if α_j is fixed at a constant (e.g., MLE) (Aitkin and Wilson, 1980; Hogg et al., 2010; Vallisneri and van Haasteren, 2017).

Converting Gaussian error to mixture error

Conversion: We simply multiply $\alpha_j^{z_j}$ to V_j with prior distributions on the additional parameters:

$$
\begin{aligned}
\text{From} \qquad \epsilon_j &\sim \mathrm{N}_p(0,\ V_j) \\
\text{to} \quad \epsilon_j \mid z_j, \alpha_j &\sim \mathrm{N}_p(0,\ \alpha_j^{z_j} V_j), \\
z_j \mid \theta_j &\sim \text{Bernoulli}(\theta_j), \\
\theta_j &\sim \text{Uniform}(0, 1), \\
\alpha_j &\sim \text{Inv.Gamma}(\nu/2,\ \nu/2).
\end{aligned}
$$

Implementation: We can use any sampler derived from a Gaussian error model ($V_j \rightarrow \alpha_j^{z_j} V_j$), additionally updating z_j, θ_j, and α_j.

Example 1: Unknown location

Three data sets:

- Original Data: $y_j \overset{i.i.d.}{\sim} \mathrm{N}(0,\ 1)$ for $j = 1, 2, \ldots, 20$.
- Data with an outlier: The same data except $y_{20} = -10$ or $y_{20} = 10$.

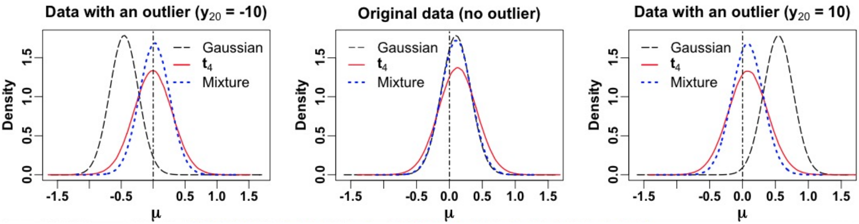
Suppose the mean (μ) of the generative Gaussian distribution is unknown.

Three error models with an improper flat prior on μ:

- Gaussian error: $y_j \mid \mu = \mu + \epsilon_j, \ \ \epsilon_j \sim \mathrm{N}(0,\ 1)$
- t_4 error (Chp 17, Gelman et al., 2013):
$$y_j \mid \mu = \mu + \epsilon_j, \ \ \epsilon_j \sim t_4$$
- Mixture error:
$$\begin{aligned} y_j \mid \mu, z_j, \alpha_j = \mu + \epsilon_j, &\ \ \ \epsilon_j \sim N(0, \alpha_j^{z_j}), \\ z_j &\sim \mathrm{Bernoulli}(0.1), \\ \alpha_j &\sim \mathrm{Inv.Gamma}(2, 2). \end{aligned}$$

Example 1: Unknown location (cont.)

Marginal posterior distribution of μ based on a million posterior samples.



- Without outlier, the dotted blue curve (mixture) passes in-between the dashed (Gaussian) and solid red (t_4) curves, i.e., a mixture effect.
- With outlier, the mixture error robustly maintains the mixture effect, enabling a robust and more accurate inference.
- CPU time (seconds): 11 for t_4 and 27 for mixture.

Example 3: Pulsar timing data

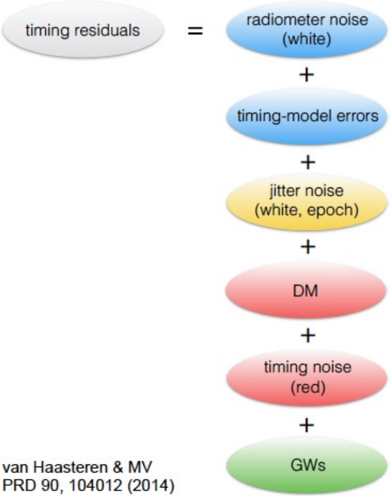
Pulse timing array for detecting gravitational waves

Credit: John Rowe, Swinbourne

"Interacting black holes in merging galaxies generates low frequency gravitational waves. As these waves propagate through space, they cause coordinated changes in the arrival times of radio signals from pulsars, the universe's most stable natural clocks. These telltale variations can be detected by powerful radio telescopes," NRAO Outreach

Example 3: Pulsar timing data (cont.)

$$\mathrm{TOA}_j - \tau_j^{\mathrm{TM}} = \epsilon_j^{\mathrm{WN}} + \delta t_j{}^{\mathrm{TM}} + \tau_j^{\mathrm{EC}} + \tau_j^{\mathrm{DM}} + \tau_j^{\mathrm{RN}} + \tau_j^{\mathrm{GW}}$$



Example 3: Pulsar timing data (cont.)

Statistical modeling: With $\phi = (A, \gamma)$

- $\boldsymbol{\varepsilon}^{\mathrm{WN}} \sim \mathrm{Normal}_n(\mathbf{0}, N)$
- $\delta {t_j}^{\mathrm{TM}} + \tau_j^{\mathrm{EC}} + \tau_j^{\mathrm{DM}} + \tau_j^{\mathrm{RN}} + \tau_j^{\mathrm{GW}} \mid \phi \sim \mathrm{Normal}_n(\mathbf{0}, T'B(\phi)T)$

Probability distribution:

Let $\boldsymbol{\delta}\mathbf{t}$ be the observed residuals.

$$\boldsymbol{\delta}\mathbf{t} \mid \phi \sim \mathrm{Normal}_n(\mathbf{0},\ V + T'B(\phi)T)$$

- Likelihood function of ϕ:

$$L(\phi) \propto \exp\left(-0.5 \times \boldsymbol{\delta}\mathbf{t}^{\top}(V + T'B(\phi)T)^{-1}\boldsymbol{\delta}\mathbf{t}\right) \times |V + T'B(\phi)T|^{-0.5}$$

- Prior dist. (f) of ϕ: $\log_{10}(A) \sim \mathrm{Unif}(-18, -10)$ and $\gamma \sim \mathrm{Unif}(0, 7)$.

- Posterior dist. (π) of ϕ:

$$\pi(\phi \mid \boldsymbol{\delta}\mathbf{t}) \propto L(\phi) \times f(\phi)$$

Example 3: Pulsar timing data (cont.)

Original model: $\boldsymbol{\delta}\mathbf{t} \mid \phi \sim \text{Normal}_n(\mathbf{0},\ V + T'B(\phi)T)$

- (Ellis, 2016) Equivalent hierarchical model:

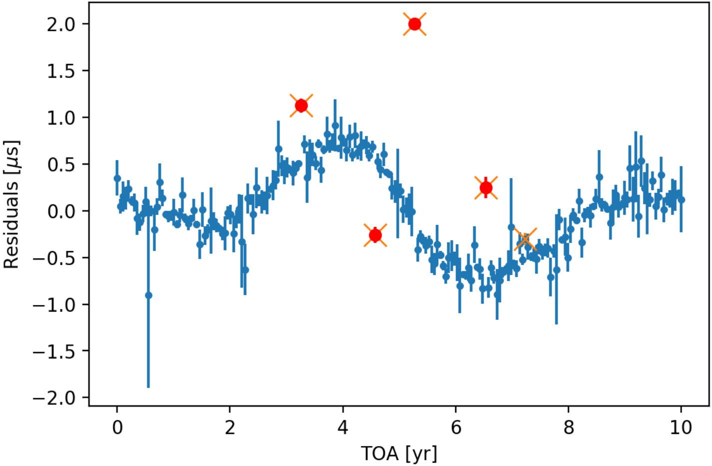
$$\boldsymbol{\delta}\mathbf{t} \mid \mathbf{b} \sim \text{Normal}_n(T'\mathbf{b},\ V)$$
$$\mathbf{b} \mid \phi \sim \text{Normal}_k(\mathbf{0},\ B(\phi))$$

- Converting the hierarchical model to outlier model:

$$\boldsymbol{\delta}\mathbf{t} \mid \mathbf{b} \sim \text{Normal}_n(T'\mathbf{b},\ \alpha^z V)$$
$$\mathbf{b} \mid \phi \sim \text{Normal}_k(\mathbf{0},\ B(\phi))$$
$$z_j \mid \theta_j \sim \text{Bernoulli}(\theta_j), \quad \theta_j \sim \text{Uniform}(0,1), \quad \alpha_j \sim \text{Inv.Gamma}(2,2).$$

Example 3: Pulsar timing data (cont.)

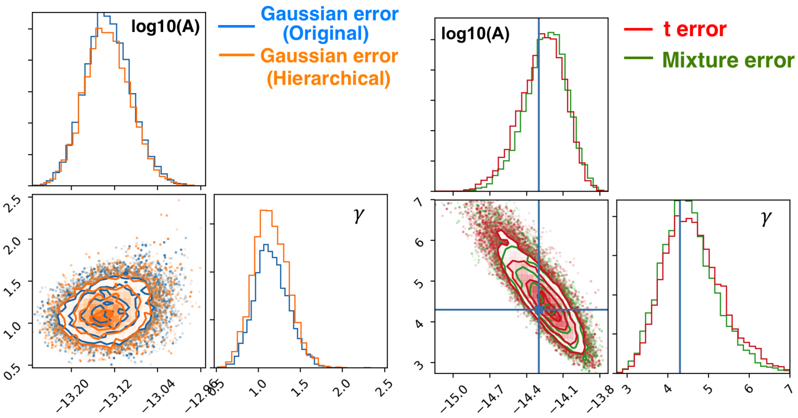
A simulation study (done by Justin Ellis)



Yellow crosses for synthetic outliers and red dots for the data that our model considers as outliers.

Example 3: Pulsar timing data (cont.)

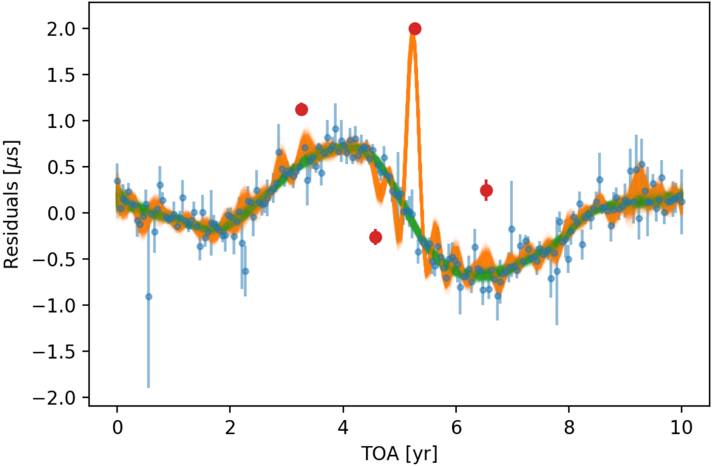
Marginal posterior distributions of $\log_{10}(A)$ and γ with their scatter plot.
- Left three panels: The original and hierarchical Gaussian error models.



- Right three panels: Red from t_4 errors, green from mixture errors, and Blue lines for true values.

Example 3: Pulsar timing data (cont.)

Orange curve for the fit with Gaussian errors, and green curve for the fit with mixture errors.



Conclusion

A mixture error model can result in a robust and more accurate parameter estimation in the presence of outliers than a t_4 error model.

It is simple and always possible to convert a Gaussian error to a mixture error by multiplying $\alpha_j^{z_j}$ to the (known) variance component V_j.

Additional computational cost is not expensive.

Detecting outliers is another venue.

References

1. Aitkin, M. and Wilson, G. T. (1980) “Mixture Models, Outliers, and the EM Algorithm” *Technometrics*, **22**, 3, 325–331.
2. Geha et al. (2003) “Variability-Selected Quasars in MACHO Project Magellanic Cloud Fields” *The Astrophysical Journal*, **125**, 1, 1.
3. Gelman et al. (2013). Bayesian Data Analysis. *CRC press*.
4. Kelly et al. (2009) “Are the Variations in Quasar Optical Flux Driven by Thermal Fluctuations?” *The Astrophysical Journal*, **698**, 1, 895.
5. Peel et al. (2000) “Robust Mixture Modelling using the t Distribution” *Statistics and Computing*, **10**, 4, 339–348.
6. Hogg et al. (2010) “Data Analysis Recipes: Fitting a Model to Data” *arXiv:1008.4686*.
7. Vallisneri, M. and van Haasteren, R. (2017) “Taming Outliers in Pulsar-Timing Datasets with Hierarchical Likelihoods and Hamiltonian Sampling” *Monthly Notices of the Royal Astronomical Society*, **466**, 4954.
8. West, M. (1987) “On Scale Mixtures of Normal Distributions” *Biometrika*, **74**, 3, 646–648.
9. Ellis, J. (2016) “Pulsar Timing Data Analysis”, *SAMSI ASTRO WG3*