Imputation of Time Series with Missing Values under Heavy-Tailed AR Model via Stochastic EM

Prof. Daniel P. Palomar
The Hong Kong University of Science and Technology (HKUST)

Invited Talk - International Workshop on Mathematical Issues in Information Sciences (MIIS)
The Chinese University of Hong Kong, Shenzhen
21 December 2018
Outline

1 Motivation

2 Imputation: State of the art
   - Different formulations
   - Basics on imputation

3 Approach for statistical time series imputation
   - Step 1: Estimation of parameters
   - Step 2: Imputation of missing values

4 Numerical simulations

5 Summary
Outline

1 Motivation

2 Imputation: State of the art
   - Different formulations
   - Basics on imputation

3 Approach for statistical time series imputation
   - Step 1: Estimation of parameters
   - Step 2: Imputation of missing values

4 Numerical simulations

5 Summary
In theory, data is typically assumed complete and algorithms are designed for complete data.

In practice, however, **often data has missing values**, due to a variety of reasons.

Then the algorithms designed for complete data **can be disastrous!**

Missing values typically happen during the data observation or recording process:

1. values may not be measured,
2. values may be measured but get lost, or
3. values may be measured but are considered unusable.

---

Some real-world cases where missing values occur:

- some stocks may suffer a lack of liquidity resulting in no transaction and hence no price recorded
- observation devices like sensors may break down during the measurement
- weather or other conditions may disturb sample taking schemes
- in industrial experiments some results may be missing because of mechanical breakdowns unrelated to the experimental process
- in an opinion survey some individuals may be unable to express a preference for one candidate over another
- respondents in a survey may not answer every question
- countries may not collect statistics every year
- archives may simply be incomplete
- subjects may drop out of panels.
What is imputation?

- How can we cope with data with missing values?
- One option is to design processing algorithms that can accept missing values, but has to be done in a case by case basis and is expensive.
- Another option is **imputation**: filling in those missing values based on some properties of the data. After that, processing algorithms for complete data can be safely used.

- However, magic cannot be done to impute missing values. One has to rely on some structural properties like some temporal structure.

- There are many imputation techniques, many heuristic (can do more harm than good) and some with a sound statistical foundation.

- Many works assume a Gaussian distribution, which doesn’t hold in many applications.

👉 We will focus on statistically sound methods for time series imputation under heavy-tailed distributions.
Outline

1 Motivation

2 Imputation: State of the art
   - Different formulations
   - Basics on imputation

3 Approach for statistical time series imputation
   - Step 1: Estimation of parameters
   - Step 2: Imputation of missing values

4 Numerical simulations

5 Summary
1 Motivation

2 Imputation: State of the art
   - Different formulations
   - Basics on imputation

3 Approach for statistical time series imputation
   - Step 1: Estimation of parameters
   - Step 2: Imputation of missing values

4 Numerical simulations

5 Summary
In big data system analytics, it is often the case that the high-dimensional data matrix lies in a low-dimensional subspace.

A popular example is the Netflix problem where the data matrix contains movie ratings by users and is extremely sparse:

$$X = \begin{bmatrix}
2 & 3 & ? & ? & 5 & ? \\
1 & ? & ? & 4 & ? & 3 \\
? & ? & 3 & 2 & ? & 5 \\
4 & ? & 3 & ? & 2 & 4 \\
\end{bmatrix}$$

In 2009, the Netflix prize of US$1M was awarded to a method based, among other techniques, on the **low-rank property**:

$$X = AB^T$$

where both $A$ and $B$ are extremely thin matrices.

This low-rank property can be used to impute missing values.
Inpainting in image processing

- In image processing, a popular problem is that of images with missing blocks of pixels:

  ![Inpainting Example](image)

- In this case, one can use the natural structure of images, e.g., small gradient or a dictionary of small structures commonly appearing.

- **Total variation** is a common technique that imputes the missing pixels by ensuring a small $\ell_1$-norm of the gradient.

- Learning an **overcomplete dictionary** allows for imputing blocks of pixels based on the dictionary.
A somehow related problem in signal processing and wireless communications is frugal sensing and compressive covariance sensing where one wants to obtain the complete knowledge of a covariance matrix.

In frugal sensing,\(^2\) one wants to obtain the matrix \(X\) from knowledge of the value of some cuts \(c_i^T X c_i = v_i\) or even just one bit of information of the cuts \(c_i^T X c_i \leq t\).

More generally, in compressive covariance sensing,\(^3\) one wants to reconstruct \(X\) from the smaller matrix \(C^T X C\), where \(C\) is some tall compression or selection matrix that exploits structural information or sparsity in some domain.

---


Time series with structure

- In some applications, one of the dimensions of the data matrix is time.
- The time dimension sometimes has some specific structure on the distribution of the missing values, like the monotone missing pattern.\(^4\)
- The time dimension can also follow some structural model that can be effectively used to fill in the missing values.
- One simple example of **time structure** is the **random walk**, which is pervasive in financial applications (e.g., log-returns of stocks):\(^5\)

\[
y_t = \phi_0 + y_{t-1} + \epsilon_t.
\]

- Another example of **time structure** is the **AR(p) model** (e.g., traded log-volume of stocks):

\[
y_t = \phi_0 + \sum_{i=1}^{p} \phi_i y_{t-i} + \epsilon_t.
\]


Outline

1 Motivation

2 Imputation: State of the art
   - Different formulations
   - Basics on imputation

3 Approach for statistical time series imputation
   - Step 1: Estimation of parameters
   - Step 2: Imputation of missing values

4 Numerical simulations

5 Summary
Some simple and naive approaches for imputation are:

- **hot deck imputation**: recorded units in the sample are used to substitute values
- **mean imputation**: means from sets of recorded values are substituted

Sounds like a good idea but it distorts the empirical distribution of the sampled values (biases in variances and covariances)

$$
y_1, y_2, \ldots, y_{k_1}, NA_1, \ldots, NA_{k_2} \rightarrow \hat{\sigma}^2 = \frac{1}{k_1} \sum_{i=1}^{k_1} (y_i - \hat{\mu})^2$$

$$
y_1, y_2, \ldots, y_{k_1}, \hat{\mu}, \ldots, \hat{\mu} \rightarrow \hat{\sigma}^2 = \frac{1}{k_1 + k_2} \left( \sum_{i=1}^{k_1} (y_i - \hat{\mu})^2 + \sum_{i=1}^{k_2} 0 \right)$$
Naive vs sound imputation methods

- Ad-hoc methods of imputation can lead to serious biases in variances and covariances.

- Examples are:
  - mean imputation
  - constant interpolation
  - linear interpolation
  - polynomial interpolation
  - spline interpolation

- A sound imputation method should preserve the statistics of the observed values.

  The statistical way is to first estimate the distribution of the missing values conditional on the observed values \( f(\mathbf{y}_{\text{miss}} | \mathbf{y}_{\text{obs}}) \) and then impute based on that posterior distribution.
Naive vs sound time series imputation methods

Illustration of different naive imputation methods and a sound statistical method that preserves the statistics:

- Imputation with LOCF (last observation carried forward)
- Imputation with linear interpolation
- Imputation with splines
- Imputation with proposed method (statistics preserved)
Suppose we have somehow estimated the conditional distribution \( f(y_{\text{miss}} | y_{\text{obs}}) \).

At this point it is trivial to randomly generate the missing values from that distribution:

\[
y_{\text{miss}} \sim f(y_{\text{miss}} | y_{\text{obs}}).
\]

This only gives you one realization of the missing values.

In some applications, one would like to have multiple realizations of the missing values to properly test the performance of some subsequent methods or algorithms.

Multiple imputation (MI) consists of generating multiple realizations of the missing values:

\[
y_{\text{miss}}^{(k)} \sim f(y_{\text{miss}} | y_{\text{obs}}) \quad \forall k = 1, \ldots, K.
\]
Motivation

Imputation: State of the art
- Different formulations
- Basics on imputation

Approach for statistical time series imputation
- Step 1: Estimation of parameters
- Step 2: Imputation of missing values

Numerical simulations

Summary
Outline

1 Motivation

2 Imputation: State of the art
   - Different formulations
   - Basics on imputation

3 Approach for statistical time series imputation
   - Step 1: Estimation of parameters
   - Step 2: Imputation of missing values

4 Numerical simulations

5 Summary
Estimation of parameters for iid Gaussian

- Suppose a univariate random variable $y$ follows a Gaussian distribution:
  
  $$y \sim \mathcal{N}(\mu, \sigma^2).$$

- We have $T$ incomplete samples $\{y_t\}$ and the missing mechanism is ignorable (aka MAR), the ML estimation problem is formulated as
  
  $$\text{maximize} \quad \log \left( \prod_{t \in C_{\text{obs}}} f_G(y_t; \mu, \sigma^2) \right),$$

  where $C_{\text{obs}}$ is the set of the indexes of the observed samples, and $f_G(\cdot)$ is the pdf of the Gaussian distribution.

- **Closed-form solution:**

  $$\hat{\mu} = \frac{1}{n_{\text{obs}}} \sum_{t \in C_{\text{obs}}} y_t$$

  and

  $$\hat{\sigma}^2 = \frac{1}{n_{\text{obs}}} \sum_{t \in C_{\text{obs}}} (y_t - \hat{\mu})^2.$$
In many applications, the **Gaussian distribution is not appropriate** and a more realistic heavy-tailed distribution is necessary. An example is in the financial returns of stocks:
Estimation of parameters for iid Student’s $t$

- The Student’s $t$-distribution is a widely used heavy-tailed distribution.
- Suppose $y$ follows a Student’s $t$-distribution: $y \sim t(\mu, \sigma^2, \nu)$ with pdf
  \[
  f_t(y; \mu, \sigma^2, \nu) = \frac{\Gamma \left( \frac{\nu+1}{2} \right)}{\sqrt{\nu \pi \sigma \Gamma \left( \frac{\nu}{2} \right)}} \left( 1 + \frac{(y - \mu)^2}{\nu \sigma^2} \right). 
  \]
- Given the incomplete data set, the ML estimation problem for $\theta = (\mu, \sigma^2, \nu)$ can be formulated as
  \[
  \maximize_{\mu, \sigma^2, \nu} \log \left( \prod_{t \in C_{\text{obs}}} f_t(y_t; \mu, \sigma^2, \nu) \right) .
  \]
- **No closed-form solution. 😞**
- Interestingly, the Student’s $t$-distribution can be represented as a Gaussian mixture:
  \[
  y_t | \tau_t \sim \mathcal{N} \left( \mu, \frac{\sigma^2}{\tau_t} \right), \quad \tau_t \sim \text{Gamma} \left( \frac{\nu}{2}, \frac{\nu}{2} \right),
  \]
  where $\tau_t$ is the mixture weight.
We can use the expectation-maximization (EM) algorithm to solve this ML estimation problem by regarding $\tau_{\text{obs}} = \{\tau_t\}_{t \in \mathcal{C}_{\text{obs}}}$ as latent variables:

- **Expectation (E) - step**: compute the expected complete data log-likelihood given the current estimates

$$Q\left(\theta | \theta^{(k)}\right) = \mathbb{E}_f(\tau_{\text{obs}} | y_{\text{obs}}, \theta^{(k)}) \left[ \log \left( f\left(y_{\text{obs}}, \tau_{\text{obs}} | \theta \right) \right) \right]$$

$$= - \sum_{t \in \mathcal{C}_{\text{obs}}} \frac{w_t^{(k)}}{2\sigma^2} \left( y_t - \mu \right)^2 - \frac{n_{\text{obs}}}{2} \log \left( \sigma^2 \right) + n_{\text{obs}} \frac{\nu}{2} \log \left( \frac{\nu}{2} \right)$$

$$- n_{\text{obs}} \log \left( \Gamma \left( \frac{\nu}{2} \right) \right) + \frac{\nu}{2} \sum_{t \in \mathcal{C}_{\text{obs}}} \left( \delta_t^{(k)} - w_t^{(k)} \right) + \text{const.},$$

where $w_t^{(k)} = \mathbb{E}[\tau_t] = \frac{\nu^{(k)} + 1}{\nu^{(k)} + \left(y_t - \mu^{(k)}\right)^2 / \left(\sigma^{(k)}\right)^2}$,

$$\delta_t^{(k)} = \mathbb{E}[\log(\tau_t)] = \psi \left( \frac{\nu^{(k)} + 1}{2} \right) - \log \left( \frac{\nu^{(k)} + \left(y_t - \mu^{(k)}\right)^2 / \left(\sigma^{(k)}\right)^2}{2} \right).$$
Estimation of parameters for iid Student’s $t$

- **Maximization (M) - step**: update the estimates as

$$\theta^{(k+1)} = \arg\max_{\theta} Q(\theta | \theta^{(k)})$$

and has **closed-form solution**: 😊

$$\mu^{(k+1)} = \frac{\sum_{t \in C_{obs}} w_t^{(k)} y_t}{\sum_{t \in C_{obs}} w_t^{(k)}},$$

$$\left(\sigma^{(k+1)}\right)^2 = \frac{\sum_{t \in C_{obs}} w_t^{(k)} (y_t - \mu^{(k+1)})^2}{n_{obs}},$$

$$\nu^{(k+1)} = \arg\max_{\nu > 0} n_{obs} \left( \frac{\nu}{2} \log \left( \frac{\nu}{2} \right) - \log \left( \Gamma \left( \frac{\nu}{2} \right) \right) \right) + \frac{\nu}{2} \sum_{t \in C_{obs}} \left( \delta_t^{(k)} - w_t^{(k)} \right).$$
Algorithm

Stochastic EM algorithm for iid Student’s \( t \):

Initialize \( \mu^{(0)} \), \( (\sigma^{(0)})^2 \), and \( \nu^{(0)} \). Set \( k = 0 \).

repeat

\[
W_t^{(k)} = \frac{\nu^{(k)} + 1}{\nu^{(k)} + (y_t - \mu^{(k)})^2 / (\sigma^{(k)})^2},
\]

\[
\mu^{(k+1)} = \frac{\sum_{t \in C_{\text{obs}}} W_t^{(k)} y_t}{\sum_{t \in C_{\text{obs}}} W_t^{(k)}},
\]

\[
(\sigma^{(k+1)})^2 = \frac{\sum_{t \in C_{\text{obs}}} W_t^{(k)} (y_t - \mu^{(k+1)})^2}{n_{\text{obs}}},
\]

\[
\nu^{(k+1)} = \arg\max_{\nu > 0} Q\left(\mu^{(k+1)}, (\sigma^{(k+1)})^2, \nu | \mu^k, (\sigma^k)^2, \nu^k\right)
\]

\( k \leftarrow k + 1 \)

until convergence
Consider an AR(1) time series with innovations following a Student’s \(t\)-distribution:

\[ y_t = \varphi_0 + \varphi_1 y_{t-1} + \varepsilon_t, \]

where \( \varepsilon_t \sim t(0, \sigma^2, \nu) \).

The ML estimation problem for \( \theta = (\varphi_0, \varphi_1, \sigma^2, \nu) \) is formulated as

\[
\text{maximize } \log \left( \int \Pi_{t=2}^T f_t(y_t; \varphi_0 + \varphi_1 y_{t-1}, \sigma^2, \nu) \, dy_{\text{miss}} \right)
\]

The objective function involves an integral, and we have no closed-form expression.

As before, we can represent \( \varepsilon_t \) as

\[ \varepsilon_t | \tau_t \sim \mathcal{N} \left( 0, \frac{\sigma^2}{\tau_t} \right), \quad \tau_t \sim \text{Gamma} \left( \frac{\nu}{2}, \frac{\nu}{2} \right), \]

and use the EM type algorithm to solve this optimization problem by regarding \( y_{\text{miss}} \) and \( \tau = \{\tau_t\}_{t=1,\ldots,T} \) as the latent variables.
Estimation of parameters for AR(1) Student’s $t$

The complete data log-likelihood $f(y_{obs}, y_{miss}, \tau \mid \theta)$ belongs to the exponential family:

$$f(y_{obs}, y_{miss}, \tau \mid \theta) = h(y_{obs}, y_{miss}, \tau) \exp(-\psi(\theta) + \langle s(y_{obs}, y_{miss}, \tau), \phi(\theta) \rangle)$$

where

$$h(y_{obs}, y_{miss}, \tau) = \prod_{t=2}^{T} \tau_t^{-\frac{1}{2}},$$

$$\psi(\theta) = -(T-1) \left\{ \frac{\nu}{2} \log \left( \frac{\nu}{2} \right) - \log \left( \Gamma \left( \frac{\nu}{2} \right) \right) - \frac{1}{2} \log (\sigma^2) - \frac{1}{2} \log (2\pi) \right\},$$

$$s(y_{obs}, y_{miss}, \tau) = \sum_{t=2}^{T} \left[ \log (\tau_t) - \tau_t y_t^2, \tau_t y_{t-1}^2, \tau_t y_t, \tau_t y_{t-1}, \tau_t y_{t-2} \right],$$

$$\phi(\theta) = \left[ \frac{\nu}{2}, -\frac{1}{2\sigma^2}, -\frac{\varphi_0^2}{2\sigma^2}, -\frac{\varphi_1^2}{2\sigma^2}, \varphi_0, \frac{\varphi_1}{\sigma^2}, -\frac{\varphi_0 \varphi_1}{\sigma^2} \right].$$
Thus the expected complete data log-likelihood can be expressed as

\[ Q(\theta | \theta^{(k)}) = \mathbb{E}_f(y_{\text{miss}}, \tau | y_{\text{obs}}, \theta^{(k)}) \left[ \log \left( f(y_{\text{obs}}, y_{\text{miss}}, \tau | \theta) \right) \right] \]

\[ = - \psi(\theta) + \left\langle \bar{s}(\theta^{(k)}), \phi(\theta) \right\rangle + \text{const.}, \]

where \( \bar{s}(\theta^{(k)}) = \mathbb{E}_f(y_{\text{miss}}, \tau | y_{\text{obs}}, \theta^{(k)}) [s(y_{\text{obs}}, y_{\text{miss}}, \tau)]. \)

The computation of \( Q(\theta | \theta^{(k)}) \) is reduced to that of \( \bar{s}(\theta^{(k)}). \)

However, \( f(y_{\text{miss}}, \tau | y_{\text{obs}}, \theta^{(k)}) \) is very complicated, and we cannot even obtain \( \bar{s}(\theta^{(k)}) \) in closed form. 😱

We can take a stochastic approximation of the expectation but, still, drawing samples from \( f(y_{\text{miss}}, \tau | y_{\text{obs}}, \theta^{(k)}) \) is very complicated! 😱
We will use a Markov chain Monte Carlo (MCMC) process.

In particular, we consider the Gibbs sampling method to generate the Markov chain: we divide the latent variables \((y_{\text{miss}}, \tau)\) into two blocks \(\tau\) and \(y_{\text{miss}}\) and then generate a Markov chain of samples from the conditional distributions

\[
\begin{align*}
&f(\tau|y_{\text{miss}}, y_{\text{obs}}, \theta^{(k)}) \\
&f(y_{\text{miss}}|\tau, y_{\text{obs}}, \theta^{(k)})
\end{align*}
\]

alternatively:

\begin{itemize}
\item \textbf{Drawing from} \(f(\tau|y_{\text{miss}}, y_{\text{obs}}, \theta^{(k)})\) \textit{is trivial since the elements of} \(\tau\) \textit{are iid following a univariate gamma distribution.}
\item \textbf{Drawing from} \(f(y_{\text{miss}}|\tau, y_{\text{obs}}, \theta^{(k)})\) \textit{is trivial since it’s just a Gaussian distribution.}
\end{itemize}
**Algorithm**

### Stochastic EM algorithm for AR(1) Student’s t:

Initialize latent variables and set $k = 0$.

repeat

- **Simulation step:** generate the samples $\left( \tau^{(k,l)}, y^{(k,l)}_m \right)$ ($l = 1, 2 \ldots, L$) from $f \left( y_{\text{miss}}, \tau \mid y_{\text{obs}}; \theta^{(k)} \right)$ via Gibbs sampling ($L$ parallel chains).

- **Approximation step:**

$$
\hat{s}^{(k)} = \hat{s}^{(k-1)} + \gamma^{(k)} \left( \frac{1}{L} \sum_{l=1}^{L} s \left( y_{\text{obs}}, y^{(k,l)}_{\text{miss}}, \tau^{(k,l)} \right) - \hat{s}^{(k-1)} \right).
$$

- **Maximization step:**

$$
\theta^{(k+1)} = \arg\max_{\theta} \hat{Q} (\theta, \hat{s}^{(k)}).
$$

$k \leftarrow k + 1$

until convergence
Maximization step

The maximization step $\theta^{(k+1)} = \arg\max_\theta \hat{Q}(\theta, \hat{s}^{(k)})$ can be obtained in closed form:

$$
\varphi^{(k+1)}_0 = \frac{\hat{s}_5^{(k)} - \varphi_1^{(k+1)} \hat{s}_7^{(k)}}{\hat{s}_3^{(k)}},
$$

$$
\varphi^{(k+1)}_1 = \frac{\hat{s}_3^{(k)} \hat{s}_6^{(k)} - \hat{s}_5^{(k)} \hat{s}_7^{(k)}}{\hat{s}_3^{(k)} \hat{s}_4^{(k)} - (\hat{s}_7^{(k)})^2},
$$

$$
\left(\sigma^{(k+1)}\right)^2 = \frac{1}{T-1} \left(\hat{s}_2^{(k)} + \left(\varphi_0^{(k+1)}\right)^2 \hat{s}_3^{(k)} + \left(\varphi_1^{(k+1)}\right)^2 \hat{s}_4^{(k)} - 2\varphi_0^{(k+1)} \hat{s}_5^{(k)} - 2\varphi_1^{(k+1)} \hat{s}_6^{(k)} + 2\varphi_0^{(k+1)} \varphi_1^{(k+1)} \hat{s}_7^{(k)}\right),
$$

$$
\nu^{(k+1)} = \arg\max_{\nu > 0} (T-1) \left\{ \frac{\nu}{2} \log \left(\frac{\nu}{2}\right) - \log \left(\Gamma \left(\frac{\nu}{2}\right)\right) \right\} + \frac{\nu \hat{s}_1^{(k)}}{2}.
$$
Convergence

The previous algorithm is very simple but does it converge? 😐

Theorem:

The sequence \( \{ \theta^{(k)} \} \) generated by the algorithm has the following
asymptotic property: with probability 1, \( \lim_{k \to +\infty} d \left( \theta^{(k)}, \mathcal{L} \right) = 0 \), where
\( d \left( \theta^{(k)}, \mathcal{L} \right) \) denotes the distance from \( \theta^{(k)} \) to the set of stationary points
of observed data log-likelihood \( \mathcal{L} = \left\{ \theta \in \Theta, \frac{\partial l(\theta; y_{\text{obs}})}{\partial \theta} = 0 \right\} \).\(^a\)

\(^a\) J. Liu, S. Kumar, and D. P. Palomar, “Parameter estimation of heavy-tailed
1 Motivation

2 Imputation: State of the art
   • Different formulations
   • Basics on imputation

3 Approach for statistical time series imputation
   • Step 1: Estimation of parameters
   • Step 2: Imputation of missing values

4 Numerical simulations

5 Summary
Step 2: Imputation of missing values

- Given the conditional distribution \( f(y_{\text{miss}}|y_{\text{obs}}) \), it is trivial to randomly generate the missing values (multiple realizations can be drawn for multiple imputation):
  \[
  y_{\text{miss}} \sim f(y_{\text{miss}}|y_{\text{obs}}).
  \]

- However, in our case, we don’t have the conditional distribution \( f(y_{\text{miss}}|y_{\text{obs}}) \) in closed form:
  \[
  f(y_{\text{miss}}|y_{\text{obs}}) = \int f(y_{\text{miss}}|y_{\text{obs}}, \theta)f(\theta|y_{\text{obs}})d\theta
  \]

- An improper way of imputing (which is acceptable in many cases with small percentage of missing values) is with \( f(y_{\text{miss}}|y_{\text{obs}}, \theta^{ML}) \), but even that expression is not available.

- We can instead draw from \( f(y_{\text{miss}}, \tau|y_{\text{obs}}, \theta^{ML}) \) and discard \( \tau \), but that expression is not available either. 😞

👉 We can generate the samples from the joint based on Markov chains.
Step 2: Imputation of missing values

- In particular, we consider the Gibbs sampling method to generate the Markov chain: we divide the latent variables \((y_{\text{miss}}, \tau)\) into two blocks \(\tau\) and \(y_{\text{miss}}\) and then generate a Markov chain of samples from the conditional distributions \(f(\tau|y_{\text{miss}}, y_{\text{obs}}, \theta)\) and \(f(y_{\text{miss}}|\tau, y_{\text{obs}}, \theta)\) alternatively.

- ** Drawing from \(f(\tau|y_{\text{miss}}, y_{\text{obs}}, \theta)\) is trivial since the elements of \(\tau\) are iid, so it is just a univariate gamma distribution for each element.**

- ** Drawing from \(f(y_{\text{miss}}|\tau, y_{\text{obs}}, \theta)\) is just a Gaussian distribution.**

- If multiple imputation is needed, then the Markov chain has to be generated multiple times. But this is not the correct way to do multiple imputation! 😱

  - **The correct way is via a Bayesian characterization of \(\theta\) instead a point estimation like ML.**
Outline

1 Motivation

2 Imputation: State of the art
   - Different formulations
   - Basics on imputation

3 Approach for statistical time series imputation
   - Step 1: Estimation of parameters
   - Step 2: Imputation of missing values

4 Numerical simulations

5 Summary
Numerical simulations

Estimation of parameters for AR(1) Student’s $t$ with real data (S&P 500):

- $\phi_0$: Initial values 0.006, 0.012
- $\phi_1$: Initial values 0.9980, 0.9988
- $\sigma^2$: Initial values $3.5 \times 10^{-05}$, $5.0 \times 10^{-05}$
- $\nu$: Initial values 3.0, 4.0, 5.0

Graphs showing the evolution of $\phi_0$, $\phi_1$, $\sigma^2$, and $\nu$ over iterations.
Numerical simulations: imputed or real?
Numerical simulations: imputed or real?

S&P 500

2012–01–03 / 2015–07–07

S&P 500

2012–01–03 / 2015–07–07

D. Palomar (HKUST)

Imputation of Time Series

39 / 42
Outline

1. Motivation

2. Imputation: State of the art
   - Different formulations
   - Basics on imputation

3. Approach for statistical time series imputation
   - Step 1: Estimation of parameters
   - Step 2: Imputation of missing values

4. Numerical simulations

5. Summary
We have introduced the issue of missing values in observations.

Imputation is the mechanism by which one fills in those missing values.

Many methods are ad-hoc with no good statistical results.

Other methods are based on some properly defined formulation based on some structural properties of the data matrix.

Time series contain special temporal structure that can be employed for imputation.

Sound statistical method:

1. estimate the statistics of the underlying distribution function and construct the conditional distribution
2. impute based on the conditional distribution either a single time or multiple times (multiple imputation)
Thanks

For more information visit:

https://www.danielppalomar.com