Inférence Bayésienne et problèmes inverses

Quentin Legros

Laboratoire Jean Kuntzmann
Université Grenoble Alpes

March 13, 2023
Example 2 - the presumption of innocence

The case of Sally Clark

- In 1999 (UK), a mother is accused of the murder of her two infant sons.
- The prosecutor defense speech, based on the expert pediatrician statistics, was overwhelming.
- It convinced all of the audience, and the mother was imprisoned.
Example 2 - the presumption of innocence

Prosecutor defense
- The prosecution case relied on statistical evidence, based on Sudden infant death syndrome (SIDS) probability.
- Prosecution expert argued $p(\text{SIDS}) = 1/8543$
- $p(\text{both SIDS}) = p(\text{SIDS}) \times p(\text{SIDS}) \approx 1/73 \text{ millions.}$
- Concluded $p(\text{innocence}) \approx 1/73 \text{ millions}$

Royal Statistical Society reaction
- In October 2001, the RSS expressed its concern at the "misuse of statistics in the courts".
- In January 2002: "The calculation leading to 1/73 millions is false"
Example 2 - the presumption of innocence

What went wrong

- Two SIDS in the same family are not independent.
- Genetic or environmental factors.
- However, the reasoning is also wrong.

Hypothesis:

Innocence: \[ P(\text{ Skull } | \text{ Angel }) = \frac{1}{73 \text{ millions}} \]

Guilty: \[ P(\text{ Skull } | \text{ Devil }) = 100\% \]
Example 2 - the presumption of innocence

Bayesian Inference

- $P(\text{🥳})$: Innocence prior
- $P(\text{😭})$: Guilty prior
- $P(\text{💀})$: Guilty likelihood
- $P(\text{ //= })$: Innocence likelihood
- $P(\text{ //= }\text{🥳})$: Probability of innocence
- $P(\text{ //= }\text{😭})$: Probability of guilt
Example 2 - the presumption of innocence

Bayesian Inference

P( Skull ) | Happy  |
Innocence prior

P( Skull ) | Devil   |
Guilty prior

P( Happy | Skull )  |
Innocence likelihood

P( Devil | Skull )  |
Guilty likelihood

P( Skull ) | Happy   |
Probability of innocence

P( Skull ) | Devil   |
Probability of guilt
Example 2 - the presumption of innocence

Bayesian analysis

- Two SIDS in the same family is rare.

- But infanticidal mothers are also very rare (1/500 millions).

- All hypothesis should be compared.

- Presumption of innocence $\Rightarrow$ a priori, peoples are statistically innocents.

- Guilt : evidence should be higher than the innocence prior belief.

- Sally Clark was convicted after evidence of the statistics misrepresentation.
Plan

1. Introduction
2. In practice
3. Lidar analysis
4. Pseudo-Bayesian approach
5. Conclusion
Frequentist vs Bayesian statistics

**Probability of events**
Knowing a given theory

**Frequentist statistics**

**Probability of theories**
Knowing given events

**Bayesian statistics**
Let's consider 5 different dices.

**First questions**

- Given dice and a result (i.e. D12 and 7), is this result extraordinary?
- Frequentist approach: depends of what is extraordinary (p value).
- Here $p \approx 8\% \Rightarrow$ not extraordinary.
- Different conclusion if we obtain a 10 times successive 7
Frequentist vs Bayesian statistics

Second questions

- We obtain a 7, what dice has been thrown?
- We can discard D4 and D6.
- Hard question using frequentist inference.
- However the answer is simple when adopting a Bayesian approach.
Frequentist vs Bayesian statistics

Bayes’ theorem

\[ P(A)P(B|A) = P(B)P(A|B) \]

- with A and B two events.
- Allows to express the union of two events.
- Laplace rewrites it several years later:

\[ P(B|A) = \frac{P(A|B)P(B)}{P(A)} \]

- Before we could only compute the probability of an event if knowing the cause.
- Laplace wrote "Mémoire sur la probabilité des causes par les événements", 1771
Frequentist vs Bayesian statistics

**Second questions**

- Allows for the evaluation of each hypothesis
- We are not rejecting hypothesis anymore, we compare them.
- Require an a prior belief on hypothesis before testing them.
Second questions
- Posterior belief: D8 is the most probable
- If another dice is thrown, we can update the prior belief.
Plan

1. Introduction
2. In practice
3. Lidar analysis
4. Pseudo-Bayesian approach
5. Conclusion
A summary of Bayes' theorem is provided, along with a detailed explanation of its application.

**Bayes’ theorem**

\[
P(\text{variable}|\text{data}) = \frac{P(\text{data}|\text{variable})P(\text{variable})}{P(\text{data})}
\]

- We want to estimate a set of variables from a posterior distribution.
- Need a likelihood function and prior models.
- We can work on:
  - The observation model \(\Rightarrow\) likelihood.
  - The priors model \(\Rightarrow\) regularization.
  - The estimation strategy \(\Rightarrow\) computational tractability, performances.
Likelihood

- Likelihood is a function of parameters of a statistical model.

- Computed from the observations.

⇒ Observation model.
  - What we (almost) always do.
  - Traducing a real-life problem into a mathematical equation.
  - Maths are useful to solve problem.

- Note: maximum likelihood estimation can be enough.
Prior model

Meaning

- **Prior**: probability knowledge before observation.

- Can inform on the level of knowledge (Uniform, Spike and Slab).

- Regularization (Markov chain, physical constraints)

- Can be adapted to estimation strategies (assumed density filtering)
Bayes’ theorem

\[ P(var_1, var_2, var_3|\text{data}) \propto P(\text{data}|var_1, var_2, var_3) P(var_1) P(var_2) P(var_3) \]

Meaning

- Posterior proportional to the product of likelihood and priors.
- Different type of estimator (MMSE, MAP)
- Convexity is preferred.
- Alternative: Markov chain Monte Carlo algorithms.
- Link with optimization techniques.
Summary

- Define an observation model.
- Compute the likelihood using the model and the observations.
- Assign a prior model to each model parameter.
- Compute the posterior distribution.
- Infer estimates from the posterior distribution.
Single-photon Lidar

Motivation

- Registering of first incident photon.
- Excellent temporal resolution (few ps).
- Limitation to ambient illumination.
Single-photon Lidar

Motivation
- The imaging conditions determine the reconstruction problem difficulty.
Multispectral analysis

Motivation

- Complex imaging system required.
- Longer acquisition time.
- Data size dependent to the number of wavelengths.
Motivation

- All colors acquired with a single detector \(^1\).
- Acquisition time / data size \(\Rightarrow\) independent to spectrum size

\(^1\)X. Ren, Y. Altmann, R. Tobin, A. McCarthy, S. McLaughlin and G. S. Buller, "Wavelength-time coding for multispectral 3D imaging using single-photon Lidar", Optics Express 2018
How and what to do

**Objective**

- 3D imaging from single waveform multispectral Lidar data.
- Estimation of depth, reflectivity and background profiles.
- Fast estimation process.
- Uncertainty quantification.

**Outline**

- Observation model.
- Prior models.
- Estimation strategy.
- Results.
- Conclusion and future work.
Observation model

Standard observation model

\[
y_t | (r, b, t_0) \sim \mathcal{P} (rg(t - t_0) + b),
\]  

- $\mathcal{P}$: Poisson distribution.
- $y_t$: photon counts in time bin $t$.
- $r$: surface reflectivity in pixel.
- $b$: background level in pixel.
- $t_0$: surface depth in pixel.
- $g$: instrumental response function.
Standard observation model

\[ y_{n,t}|(r_n, b_n, t_n) \sim \mathcal{P} \left( b_n + \sum_{\ell=1}^{L} r_{n,\ell} g_\ell(t - t_n) \right), \]

- \( \mathcal{P} \): Poisson distribution.
- \( y_{n,t} \): photon counts of the \( n \)th photon in the \( t \)th temporal bin
- \( r_n \): spectral reflectivity of the \( n \)th pixel
- \( b_n \): background level of the \( n \)th pixel
- \( L \): number of wavelengths
- \( t_n \): depth profile in the \( n \)th pixel
- \( g_\ell \): impulse response function (IRF) associated to the \( \ell \)th wavelength
- \( T \): histograms length
Bayesian model

Joint likelihood

\[ p(\mathbf{Y}|\mathbf{R}, \mathbf{b}, \mathbf{t}) = \prod_{n} p(y_n|r_n, t_n, b_n). \] (3)

Prior models

- Depth: Total variation
  \[ p(t|\epsilon) = \exp[-\epsilon \text{TV}(t)]. \] (4)

- Background level: Gamma
  \[ f(b|\alpha, \beta) = \prod_{n} \mathcal{G}(b_n|\alpha, \beta). \] (5)

\(^2\)Y. Altmann, A. Maccarone, A. McCarthy, G. Newstadt, G. S. Buller, S. McLaughlin and A. Hero, "Robust spectral unmixing of sparse multispectral Lidar waveforms using gamma Markov random fields", IEEE TCI 2018
Bayesian model

Prior models

- Reflectivity: Cluster-Gamma

\[
f(R|k_c,\ell, \theta_c,\ell) = \prod_{c=1}^{C} \prod_{n \in l_c} \prod_{\ell=1}^{L} G(r_n,\ell; k_c,\ell, \theta_c,\ell), \tag{6}\]

- Hyperparameters, hierarchical model: gamma and inverse gamma

\[
f(k_c,\ell, \theta_c,\ell) = \prod_{c,\ell} G(k_c,\ell; 2, 0.5) 1_{(1,\infty)}(k_c,\ell) IG(\theta_n,\ell; 1.01, 0.5). \tag{7}\]

---

Prior models

- Joint posterior distribution (with $\Phi = (\{k_{c,\ell}, \theta_{c,\ell}\}_{c,\ell}, \alpha, \beta)$):

$$f(R, b, t, \Phi | Y) \propto p(Y | R, b, t)p(t)f(R, b | \Phi)f(\Phi),$$

- Challenging joint estimation: multimodal likelihood function
- Sequential estimation $\Rightarrow$ stochastic EM algorithm
  
  $$(\hat{R}, \hat{b}, \hat{\Phi}) = \arg\max_{R, b, \Phi} \sum_t f(R, b, t, \Phi | Y)$$

- Marginal maximum a posteriori depth estimation

$$\hat{t}_n = \arg\max_{t_n} \sum_{t \setminus n} p(t | Y, \hat{R}, \hat{b}, \hat{\Phi}), \text{ with } t \setminus n \text{ is } t \text{ expect } t_n.$$
Estimation strategy

Computational bottlenecks
- Computation of the likelihood function.
- Depth sampling in the stochastic step.

Likelihood function computed on a coarser grid
- Discrete range grid $\Rightarrow$ used to compute marginal probability.
- Depth grid subsampled by a factor $T_s$.
- Allows satisfactory depth sampling if $T_s$ remains lower than the delays between each IRF.

Reducing the number of depth samples
- Only one sample.
- Multimodal nature of the likelihood will help for robust sampling.
Set \( G_\ell = \sum_{t=1}^{T} g_\ell(t - t_n) \).

- SBR: Signal-to-background ratio.
- MSC: Mean signal photon counts.
- Comparison with C-Dirichlet \(^3\).
- Different values for \( T_s \) and \( N_t \).
- Denoted as C-Gamma \( \{ T_s, N_t \} \).
- \( \text{MCS} = \frac{1}{N} \sum_{n,\ell} r_{n,\ell} G_\ell \)
- \( \text{SBR} = \frac{1}{N} \sum_{n,\ell} \frac{r_{n,\ell} G_\ell}{Tb_n} \)

---

\(^3\) Q. Legros, S. Meignen, S. McLaughlin, and Y. Altmann, "Expectation-Maximization based approach to 3D reconstruction from single-waveform multispectral Lidar data", IEEE TCI 2020
Results - Reflectivity

MSC : 10 photons

MSC : 44 photons

- C-Dirichlet
- C-Gamma
- C-Gamma\{1,100\}
- C-Gamma\{4,1\}
- C-Gamma\{10,1\}
Results - Depth

MSC : 10 photons

MSC : 44 photons

Fraction of pixel

Absolute error

Fraction of pixel

Absolute error

C-Dirichlet
C-Gamma
C-Gamma{1,100}
C-Gamma{4,1}
C-Gamma{10,1}
### Results - Computational complexity

Computational cost of competing approaches for synthetic SW-MSL data analysis.

<table>
<thead>
<tr>
<th>MSC/SBR</th>
<th>10 / (\infty)</th>
<th>10 / 0.046</th>
<th>44 / (\infty)</th>
<th>44 / 0.426</th>
</tr>
</thead>
<tbody>
<tr>
<td>C-Gamma{1,1}</td>
<td>104s</td>
<td>132s</td>
<td>350s</td>
<td>493s</td>
</tr>
<tr>
<td>C-Gamma{4,1}</td>
<td>43s</td>
<td>49s</td>
<td>55s</td>
<td>69s</td>
</tr>
<tr>
<td>C-Gamma{10,1}</td>
<td>34s</td>
<td>36s</td>
<td>38s</td>
<td>42s</td>
</tr>
<tr>
<td>C-Dirichlet</td>
<td>530s</td>
<td>592s</td>
<td>629s</td>
<td>662s</td>
</tr>
</tbody>
</table>

- Similar estimation performance.
- Significant speed-up.
Results - Example

- Depth/RGB reconstruction with 1.1 (bottom) and 5.7 (top) signal photons per pixel.
- Non-negligible background (SBR = 1.4).
- Comparison with existing approach $^4$.

---

$^4$ X. Ren, Y. Altmann, R. Tobin, A. McCarthy, S. McLaughlin and G. S. Buller, "Wavelength-time coding for multispectral 3D imaging using single-photon Lidar", Optics Express 2018
Recent Bayesian model for 3D reconstruction from single-waveform multipectral Lidar data.

Bayesian approaches can be relatively fast.

Perform well with few information.
Plan

1. Introduction
2. In practice
3. Lidar analysis
4. Pseudo-Bayesian approach
5. Conclusion
Focus on MultiComponent Signal (MCS).

\[ x(t) = \sum_{k=1}^{K} x_k(t), \text{ with } x_k(t) = a_k(t) e^{i\phi_k(t)}, \]

Investigated approaches
- Mixture of \( K \) superimposed components.
- \( a_k(t) \) and \( \phi_k(t) \) the time-varying amplitude and phase of component \( k \).
Motivation

- Variety of application.
- Audio, medical, astronomical, echolocation,…
- Purpose: extracting the components.
Introduction

Limitations

- Challenge in the presence of noise.
- Acquisition condition, recording device, presence of outliers,...
Introduction: Time-frequency tools

Frequency analysis

- Classical approach: frequency analysis.
- The Fourier transform of a signal $x$ writes
  \[
  \hat{x}(\omega) = \int_{\mathbb{R}} x(u) e^{-j\omega u} \, du,
  \]
  with $j^2 = -1$ and $\hat{x}$ the Fourier transform of $x$.
- However: Signal frequency can vary over time.
- $\Rightarrow$ Short-Time Fourier Transform (STFT).
  \[
  \mathcal{F}_{x}^{h}(t, \omega) = \int_{\mathbb{R}} x(u)h(t - u)^* e^{-j\omega u} \, du,
  \]
  with $z^*$ the complex conjugate of $z$.
- $h$ the analysis window providing temporal information.
Introduction: Time-frequency tools

Spectrograms of multi-components signals

- Provide time and frequency content.
- Signals components appear as ridges in the time frequency plane.
- Interest: ridges position.
Our model

\[ s_{n,m}|\bar{m}_n \sim g(m - \bar{m}_n). \]

- Model 1D signals \( s_n \).
- Known data distribution.

- \( m \): frequency in \([0, M - 1]\).
- \( \bar{m}_n \): ridge position in the \( n \)-th time bin.
- \( g(m) = \frac{2\sqrt{\pi} L}{M} e^{-\left(\frac{2\pi m L}{M}\right)^2} \).
- \( L \): time spread of the analysis window.
Observation model

Our model

\[ s_{n,m} | \bar{m}_n \sim g(m - \bar{m}_n). \]

Limitations

- Simple model.
- Computationally attractive.
- Presence of noise neglected.
- Assumes for the presence of a single component.
Observation model

Limitations

- Lack of generality of the postulated model.
- Discrepancies with noisy observations.
- Or in the presence of multiple components.
- Inefficiency of Maximum Likelihood Estimation (MLE).

Proposed approach

- Estimation performance does not only depend on the model quality.
- Modification of the similarity measure\(^5\).

---

Note that

- Performing MLE $\Leftrightarrow$ minimizing Kullback-Leibler Divergence (KLD) between model and observations.
- KLD not suitable when the postulated model is inaccurate.
- Implies model mismatch.
  - In the presence of external spurious noise.
  - When observing multicomponent signals.

Alternative variational objective

- KLD replaced by the Rényi and $\beta$ divergences.
- Allow respectively for mode seeking character and robustness.
Estimation strategy

Alternative inference

- Variational inference based on alternative divergences.
- Need for the divergences cross entropy\(^2\).

Cross entropy

- For the \(\beta\)-divergence (\(\beta\)-d), \(\beta > 0\)
  \[
  CE_{\beta}(\bar{m}_n) = -\frac{1 + \beta}{\beta} \sum_m p(s_n, m|\bar{m}_n)^\beta + \int p(m|\bar{m}_n)^{1+\beta} dm. \tag{10}
  \]

- For the Rényi divergence (R-d), \(\alpha > 0, \alpha \neq 1\)
  \[
  CE_{\alpha}(\bar{m}_n) = \frac{1}{\alpha - 1} \log \left( \sum_m s_{n,m}^\alpha p(s_n, m|\bar{m}_n)^{1+\alpha} \right). \tag{11}
  \]

Estimation strategy

Pseudo-Bayesian estimation

- Approximate posterior distribution obtained by maximizing the evidence lower-bound (ELBO)$^6$
  \[ p(\tilde{m}_n | s_n) \propto e^{-MCE(\tilde{m}_n)} p(\tilde{m}_n). \]  
  \( (12) \)

- Spatial prior model \( p(\tilde{m}_n) \) discussed hereafter.
- Plug cross entropy for alternative objectives.
- Ridge position estimated by minimum mean squared error (MMSE).

---

Estimation strategy

Online estimation

- Ridge extraction.
- Iterating on time axis.
- Sequential propagation of the information.
- Spatial prior: Gaussian random walk.
- **Complexity:** Variational inference.
- **Accuracy:** Backward correction.
Estimation strategy

Sequential demodulation
- Extraction of a ridge.
- Update of the data.
- Removing energy associated with the estimated component.
- Stopping criterion: method of choice\(^7\).

Results

Numerical experiments

- Reconstruction performance of a MCS.
- Reconstruction quality factor: $RQF = 10 \log_{10} \left( \frac{||x||^2}{||x-\hat{x}||^2} \right)$.
- Assessment: Component-wise RQF.

Three components

- Sinusoidal frequency modulated (FM).
- Linear chirp.
- Sinusoid.

Figure: Spectrogram of the analyzed multicomponent signal.
Table: RQF of each components (averaged over 100 realizations) for the different competing approaches for a SNR = 10 dB.

<table>
<thead>
<tr>
<th></th>
<th>Sinusoid</th>
<th>Linear chirp</th>
<th>Sin. FM chirp</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brevdo</td>
<td>16.10</td>
<td>15.46</td>
<td>2.86</td>
<td>11.47</td>
</tr>
<tr>
<td>Brevdo-Synchrosqueezing</td>
<td>16.43</td>
<td>15.34</td>
<td>5.24</td>
<td>12.34</td>
</tr>
<tr>
<td>Proposed $\beta$-d, $\beta = 0.5$</td>
<td><strong>16.71</strong></td>
<td>15.22</td>
<td>9.13</td>
<td>13.69</td>
</tr>
<tr>
<td>Proposed $\beta$-d, $\beta = 0.8$</td>
<td>16.45</td>
<td>14.92</td>
<td>5.49</td>
<td>12.29</td>
</tr>
<tr>
<td>Proposed-KLD</td>
<td>2.46</td>
<td>2.65</td>
<td>1.18</td>
<td>2.10</td>
</tr>
<tr>
<td><strong>Proposed R-d, $\alpha = 0.5$</strong></td>
<td>16.59</td>
<td>15.24</td>
<td><strong>9.57</strong></td>
<td><strong>13.80</strong></td>
</tr>
<tr>
<td>Proposed R-d, $\alpha = 0.8$</td>
<td>15.44</td>
<td>15.22</td>
<td>7.84</td>
<td>12.83</td>
</tr>
</tbody>
</table>

Numerical experiments

- Our method obtains the best averaged RQF using R-d ($\alpha = 0.5$).
- Efficient recovery of the sinusoidally FM chirp.
- Alternative divergences circumvent the lack of generality of our model.
Numerical experiments

- Single linear chirp component.
- SNR = -15dB.
- Rényi divergence, $\alpha = 0.2$. 

![Graphs showing ground truth, observations, and mask with data points and curves.](image-url)
Results - examples 2

Numerical experiments

- Single linear chirp component.
- SNR = -15dB.
- Rényi divergence, $\alpha = 0.2$. 

![Ground truth](image1.png)
![Observations](image2.png)
![Mask](image3.png)
Conclusions

- A novel pseudo-Bayesian estimation procedure to demodulate MCS.

- Inaccurate model.

- Performances improved through estimation strategy.

- Adaptive prior model.
Conclusions

- Bayesian approaches can be adapted to a wide range of problems.
- Do not require a large amount of data.
- Allow for interpretation and confidence estimation of predictions.
- Not necessarily computationally expensive.
- Different ways to address a problem: models, priors, estimation strategy...
Thanks for your attention!

legros.quentin2@hotmail.fr