A New Tone of EM algorithm in the Universe: Analysis of MMT/Megacam Data

Zhan Li (With f.b.bianco and p. protopapas)
Dec. 16
Agenda

- Backgrounds
- Models
- Some Numerical Results
- Discussion
Background

- In observing the objects in the space, there is a gap between the observable objects by direct observations and the observable objects by X-ray.
- Our Analysis of MMT/Megacam data is trying to fill the gap
Background

MMT/Megacam survey

Target objects 200m-1km (currently >=600m)
Background

MMT is a 6.5 meter telescope on the summit of Mt. Hopkins, Arizona

Fred Lawrence Whipple Observatory
Background

Megacam continuous readout

Sky is added at every exposure (x2304)
Background

- The data from MMT/Megacam is two dimensional time series data: we have two dimensional observations of the stars and we also have a time horizon.
- We will indirectly observe the targeted objects via the stars.
- Want to find out the "events" when the targeted objects pass the stars.
Background

• How to identify the “events”? By the fluctuation of the flux of the stars

• Need to de-convolute the effects from the stars and the background
Background
Agenda

- Backgrounds
- Models
- Some Numerical Results
- Discussion
Models

- We will utilize the EM algorithm in the de-convolution
- Will present models with different assumptions/approaches
- Currently focus only on the de-convolution problem
Models

- We have binned data of photons, from both the background and the stars (bin size- a pixel or so )

- Notations and Setups of the question:
  - Stars: \( i = 1, \cdots, n \)
  - Bins: \( j = 1, \cdots, m \)
  - Observed Data: \( Y_{\text{obs}} = \{N_1, \cdots, N_m\} \) observed counts from each bin
  - Missing Data: \( Y_{\text{mis}} = \{Z_{ij}\} \), \( i = 1, \cdots, n+1, j = 1, \cdots, m \) the photons from star \( i \) to bin \( j \). the subscript \( n+1 \) means background
Models

- Model 1: Fixed bin counts with Poisson Backgrounds
Models

- We do not place distribution assumptions on the number of photons in each bin and we assume the background in each bin are i.i.d. poisson.
- Within each bin, the number of photons from each star (background excluded) follow a multinomial distribution ($\tilde{N}_j$ total photons from stars):

$$(Z_{1j}, \cdots, Z_{nj}) \sim \text{multinomial}(\tilde{N}_j, p_{1j}, \cdots, p_{nj})$$

- Normal parameterization for PSF (point spread function):

$$p_{ij} = \frac{q_i \phi(x_j; \mu_i, \sigma_i)}{\sum_{i=1}^{n} q_i \phi(x_j; \mu_i, \sigma_i)}$$
Models

- Take the background into account

\[
Z_{n+1,j} | N_j \sim \text{Pois}(\lambda) | \text{Pois}(\lambda) \leq N_j
\]

\[
Z_1, \cdots, Z_n, Z_{n+1,j} | N_j, Z_{n+1,j} \sim \text{multinomial}(N_j - Z_{n+1,j}, p_1,j, \cdots, p_{n,j})
\]
Models

• Pros of the model: We do have the closed form solution for the updating equation:

\[ \lambda' = \sum_j M_j \]

\[ q' = \sum_j \tilde{q}_j (N_j - M_j) / \sum_j (N_j - M_j) \]

\[ \mu_1' = \sum_j \tilde{q}_j (N_j - M_j) x_j / \sum_j \tilde{q}_j (N_j - M_j) \]

\[ \mu_2' = \sum_j (1 - \tilde{q}_j) (N_j - M_j) x_j / \sum_j (1 - \tilde{q}_j) (N_j - M_j) \]

\[ \sigma_1' = \left( \frac{\sum_j \tilde{q}_j (N_j - M_j) (x_j - \mu_1')^2}{\sum_j \tilde{q}_j (N_j - M_j)} \right)^{1/2} \]

\[ \sigma_2' = \left( \frac{\sum_j (1 - \tilde{q}_j) (N_j - M_j) (x_j - \mu_1')^2}{\sum_j (1 - \tilde{q}_j) (N_j - M_j)} \right)^{1/2} \]

\[ M_j = \sum_{k=0}^{N_j} \frac{k \times e^{-\lambda} \lambda^k / k!}{\sum_{l=0}^{N_j} e^{-\lambda} \lambda^l / l!} \]
Models

- Cons of the model:
  - Tend to underestimate the background and over estimate the dispersion of the normal distribution
  - Not explicitly estimating the intensity of flux: we only estimate the proportion in each normal
Models

• Model II: Poisson bin counts with Poisson Backgrounds
Models

- We will assume that the total number of photons from a star is following a Poisson distribution (Esch and et al. 2004)

\[ Y_j | \lambda_i, \lambda_B \sim Poisson\left[\left(\sum_i P_{ij}\lambda_i\right) + \lambda_B\right] \]

- And we will incorporate the location and dispersion of the stars through parameterization of PSF (point spread function)
Models

- Then, within the same bin (pixel), we have

\[(Z_{1,j}, \cdots, Z_{nj}, Z_{n+1,j}) \sim \text{multinomial}(N_j, \frac{\lambda_1 P_{1j}}{\sum_i \lambda_i P_{ij} + \lambda_B}, \cdots, \frac{\lambda_n P_{nj}}{\sum_i \lambda_i P_{ij} + \lambda_B}, \frac{\lambda_B}{\sum_i \lambda_i P_{ij} + \lambda_B})\]

- The parameterization of the PSF:

\[p_{ij} \propto \phi(x_j; \mu_i, \sigma_i)\]
Models

- **Pros:**
  - Explicitly model the intensity or flux of the star through the poisson parameter
  - Better acknowledged in the research community

- **Cons:**
  - We do not have the closed form solution for EM iteration, which is especially undesirable for the large scale problem we have
Models

- Model 3: Hierarchical Bayes
Models

- We will assume that the total number of photons from a star is following a Poisson distribution (Esch and et al. 2004).
- We will use Hierarchical Bayes instead of EM algorithm to sample the posterior distribution of intensity and the PSF.
Models

- Pros: Very effective in accounting for the uncertainty of parameters
- Cons: not conjugate prior, computational concerns...
Numerical Results

- **Model I:**

<table>
<thead>
<tr>
<th>$p$</th>
<th>$\mu_1$</th>
<th>$\mu_2$</th>
<th>$\sigma_1$</th>
<th>$\sigma_2$</th>
<th>$\lambda_B$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5(.49)</td>
<td>80(79.4)</td>
<td>30(30.0)</td>
<td>10(9.8)</td>
<td>10(11.5)</td>
<td>100(83.5)</td>
</tr>
<tr>
<td>0.5(.27)</td>
<td>80(83.2)</td>
<td>50(52.2)</td>
<td>10(8.0)</td>
<td>10(19.7)</td>
<td>100(64.4)</td>
</tr>
<tr>
<td>.3(.32)</td>
<td>80(79.1)</td>
<td>30(29.9)</td>
<td>10(10.4)</td>
<td>10(11.4)</td>
<td>100(75.5)</td>
</tr>
</tbody>
</table>
Discussion and Future work

• Need a computational effective way to de-convolute the stars with certain accuracy
• Next step: look at the time series data
Thank you and Happy Holiday!