Celeste: Variational inference for a generative model of astronomical images

Jeffrey Regier

Statistics Department
UC Berkeley

July 9, 2015

Joint work with Jon McAuliffe (UCB Statistics), Andrew Miller, Ryan Adams, Matt Hoffman, Dustin Lang, David Schlegel, and Prabhat
An astronomical image

An image from the Sloan Digital Sky Survey, showing a galaxy from the constellation Serpens, 100 million light years from Earth, along with several other galaxies and many stars from our own galaxy.
Project goals

- Catalog all galaxies and stars that are visible through the next generation of telescopes.
  - The Large Synoptic Survey Telescope, for example, will house a 3200-megapixel camera producing 8 terabytes of images nightly.
- Identify promising galaxies for spectrograph targeting.
  - Better understand dark energy and the geometry of the universe.
- Develop an extensible model and inference procedure, for use by the astronomical community.
  - Future applications might include finding supernovae and detecting killer asteroids.
The Celeste graphical model
Brightness and colors

- latent random variable $r_s$ models \textit{brightness} of source $s$ in the reference band
  - the arriving quantity of energy in band “r”

  \[
  r_s | (a_s = i) \sim \text{Gamma} \left( \gamma^{(i)}, \Phi^{(i)} \right).
  \]

- latent random vector $c_s$ models the $B - 1$ \textit{colors} of source $s$
  - color = log ratio of brightness in adjacent bands

  \[
  k_s | (a_s = i) \sim \text{Categorical} \left( \Xi_1^{(i)}, \ldots, \Xi_D^{(i)} \right),
  \]

  \[
  c_s | (k_s = d, a_s = i) \sim \text{MvNormal} \left( \Omega^{(i,d)}, \Lambda^{(i,d)} \right).
  \]

- Then, the brightness $\ell_{sb}$ of each light source $s$ in each band $b$ is a deterministic function of $r_s$ and $c_s$.  

Colors: scientific priors

Stars

Galaxies
Galaxies: light-density model

\[ R_s = \begin{bmatrix} \cos \varphi_s & -\sin \varphi_s \\ \sin \varphi_s & \cos \varphi_s \end{bmatrix}, \]

\[ W_s = R_s^T \begin{bmatrix} \sigma_s^2 & 0 \\ 0 & \sigma_s^2 \rho_s^2 \end{bmatrix} R_s, \]

\[ h_{si} (w') = \sum_{j=1}^{J} \tilde{\eta}_{ij} \phi (w'; \mu_s, \tilde{\nu}_{ij} W_s), \quad i = 0 \text{ or } 1 \]

\[ h_s (w') = \theta_s h_{s1} (w') + (1 - \theta_s) h_{s0} (w'). \]
Idealized sky view

Let $\delta_{\mu s}$ denote the Dirac delta function—the profile of a star. Then brightness for sky position $w'$ is

$$G(w') = \sum_{s=1}^{S} \ell_{sb} g_{sa_s}(w')$$

where

$$g_{si}(w') = \begin{cases} 
\delta_{\mu_s}(w'), & \text{if } i = 0 \text{ ("star")}, \\
h_s(w'), & \text{if } i = 1 \text{ ("galaxy")}.
\end{cases}$$
Point spread function

Credit: Sloan Digital Sky Survey
Point spread function

- PSF $\leftrightarrow$ mixture of $K$ Gaussians:

$$f_{nbm}(w') = \sum_{k=1}^{K} \bar{\alpha}_{nbk} \phi \left( \bm{w}_m; w' + \bar{\xi}_{nbk}, \bar{\tau}_{nbk} \right).$$

- $\phi$ is the bivariate normal density.
- $K$ and the $(\bar{\alpha}_{nb}, \bar{\xi}_{nb}, \bar{\tau}_{nb})$ are the parameters of the PSF, determined a priori.
Likelihood: idealized sky view + PSF

Convolve intermediate sky view w/PSF ⇒ photon arrival rate for pixel \( m \):

\[
G_{nbm} = \int G(w') f(w') \, dw'
\]

\[
= \sum_{s=1}^{S} \ell_{sb} \int g_{s_{as}}(w') f(w') \, dw' .
\]

These normal-normal convolutions are analytic. For stars,

\[
f_{s_{0}}(m) := \int g_{s_{0}}(w') f(w') \, dw'
\]

\[
= \sum_{k=1}^{K} \bar{\alpha}_{nbk} \phi (m; \mu_{s} + \bar{\xi}_{nbk}, \bar{\tau}_{nbk}) .
\]

Let \( \theta_{s_{1}} = \theta_{s} \) and \( \theta_{s_{2}} = 1 - \theta_{s} \). For galaxies,

\[
f_{s_{1}}(m) := \int g_{s_{1}}(w') f(w') \, dw'
\]

\[
= \sum_{k=1}^{K} \bar{\alpha}_{nk} \sum_{i=1}^{2} \theta_{si} \sum_{j=1}^{J} \bar{\eta}_{1j} \phi (m; \mu_{s} + \bar{\xi}_{nbk}, \bar{\tau}_{nbk} + \bar{\nu}_{ij} \mathcal{W}_{s}) .
\]
Likelihood

Let $a = (a_s)_{s=1}^{S}$, $r = (r_s)_{s=1}^{S}$, and $c = (c_s)_{s=1}^{S}$. Then the expected number of photons landing in pixel $m$ is

$$F_{nbm}(a, r, c) = \nu_{nb} [\epsilon_{nb} + G_{nbm}].$$

For $n = 1, \ldots, N$, $b = 1, \ldots, B$, and $m = 1, \ldots, M$, we model

$$x_{nbm}|a, r, c \sim \text{Poisson}(F_{nbm}(a, r, c)).$$
Intractable posterior

Let $\Theta = (a, r, c)$. The posterior on $\Theta$ is intractable because of coupling between the sources:

$$p(\Theta|x) = \frac{p(x|\Theta)p(\Theta)}{p(x)}$$

and

$$p(x) = \int p(x|\Theta)p(\Theta) \, d\Theta$$

$$= \int \prod_{n=1}^{N} \prod_{b=1}^{B} \prod_{m=1}^{M} p(x_{nbm}|\Theta)p(\Theta) \, d\Theta.$$
Variational inference

Let $Q$ be a family of distributions on latent variables $\Theta$. For $q \in Q$, 

$$\log p(x) \geq \mathbb{E}_q [\log p(x, \Theta)] - \mathbb{E}_q [\log q(\Theta)]$$

$$= \log p(x) - D_{KL} (q(\Theta), p(\Theta|x))$$

$$=: \mathcal{L}(q).$$

Variational inference approximates the exact posterior with a simpler distribution 

$$q^* = \arg \max_{q \in Q} \mathcal{L}(q).$$
Variational inference: advantages and limitations

Advantages:
- potentially orders of magnitude faster the MCMC
- no post-processing of samples—can compute statistics of the approximating distribution nearly instantaneously

Limitations:
- bias—and few error bounds are known for statistics based on an approximating distribution rather than the true posterior
- may require modeling changes, to avoid intractable expectations
- may necessitate solving difficult optimization problems, even if all expectations are tractable
Variational optimization... isn’t easy

\[ \mathcal{L} (\chi, \mu, \kappa, \gamma, \zeta, \beta, \lambda, \theta, \rho, \sigma, \varphi) \]

\[ = C + \sum_{n=1}^{N} \sum_{b=1}^{B} \sum_{m=1}^{M} \left\{ \sum_{a \in \{0,1\}} \prod_{s=1}^{S} \chi^a_s (1 - \chi_s)^{1-a_s} \right\} \left\{ \int_{r_1} \int_{c_1} \int_{k_1} \cdots \int_{r_S} \int_{c_S} \int_{k_S} \right. \]

\[ \chi_{n bm} \log \left[ \epsilon_{nb} + \sum_{s=1}^{S} r_s \prod_{j=b}^{b-1} \exp \{c_{sb}\} \prod_{j=b}^{b-1} \exp \{c_{sb}\} \right] \]

\[ \times \left[ \sum_{k=1}^{3} \bar{\alpha}_{nk} \phi \left( m - w; \bar{\xi}_{nbk}, \bar{\Sigma}_{nbk} \right) g_{si} (w) \right] \]

\[ - \nu_{nb} \sum_{s=1}^{S} r_s \prod_{j=b}^{b-1} \exp \{c_{sb}\} \prod_{j=b}^{b-1} \exp \{c_{sb}\} \int_{k=1}^{3} \bar{\alpha}_{nk} \phi \left( m - w; \bar{\xi}_{nbk}, \bar{\Sigma}_{nbk} \right) g_{si} (w) \]

\[ dr_1 \ dc_1 \ dk_1 \ \ldots \ dr_S \ dc_S \ dk_S \right\} \]

\[ + \sum_{s=1}^{S} \left\{ D_{KL} (q(a_s), p(a_s)) + \sum_{i=1}^{2} \chi^a_s (1 - \chi_s)^{1-a_s} \right\} \]

\[ \times \left[ D_{KL} (q(r_s | a_s = i), p(r_s | a_s = i)) + D_{KL} (q(k_s, c_s | a_s = i), p_s (k_s, c_s | a_s = i)) \right] \]
However,

- A structured mean-field assumption that factorizes across light sources makes most expectations tractable:

\[ q(\Theta) = \prod_{s=1}^{S} q(\Theta_s). \]

- The delta method for variational inference approximates the remaining expectations.

- Existing catalogs provide good initial settings for the variational parameters.

- Light sources are unlikely to contribute photons to distant pixels.
Left is a 51 pixel × 51 pixel sub-region of an astronomical image, captured through the $r$ band filter. Each pixel’s value corresponds to the number of photons that hit it. The right panel shows $\mathbb{E}_{q^*}[F_{nbm}]$, the mean of $x_{nbm}$ with respect to our posterior approximation.
# Results

<table>
<thead>
<tr>
<th>dataset model</th>
<th>real</th>
<th>synthetic</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>primary</td>
<td>celeste</td>
<td>improve</td>
</tr>
<tr>
<td>position</td>
<td>0.22</td>
<td>0.20</td>
<td>.02 (.00)</td>
</tr>
<tr>
<td>missed gals</td>
<td>28 / 654</td>
<td>15 / 654</td>
<td>.02 (.01)</td>
</tr>
<tr>
<td>missed stars</td>
<td>8 / 654</td>
<td>31 / 654</td>
<td>-.04 (.01)</td>
</tr>
<tr>
<td>brightness</td>
<td>0.76</td>
<td>1.60</td>
<td>-.83 (.12)</td>
</tr>
<tr>
<td>color u-g</td>
<td>1.10</td>
<td>0.49</td>
<td>.61 (.04)</td>
</tr>
<tr>
<td>color g-r</td>
<td>0.16</td>
<td>0.09</td>
<td>.07 (.01)</td>
</tr>
<tr>
<td>color r-i</td>
<td>0.09</td>
<td>0.06</td>
<td>.03 (.00)</td>
</tr>
<tr>
<td>color i-z</td>
<td>0.25</td>
<td>0.10</td>
<td>.15 (.01)</td>
</tr>
<tr>
<td>profile</td>
<td>0.19</td>
<td>0.23</td>
<td>-.04 (.02)</td>
</tr>
<tr>
<td>eccentricity</td>
<td>0.17</td>
<td>0.13</td>
<td>.04 (.01)</td>
</tr>
<tr>
<td>scale</td>
<td>0.37</td>
<td>1.28</td>
<td>-.91 (.17)</td>
</tr>
<tr>
<td>angle</td>
<td>19.40</td>
<td>18.10</td>
<td>1.40 (.80)</td>
</tr>
</tbody>
</table>