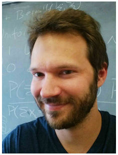
Approximate inference for generative models of astronomical images

Jon McAuliffe
with the DESI collaboration

Statistics Department
UC Berkeley

June 7th, 2016

The DESI collaboration



Jeff Regier

Ryan Giordano

Jon McAuliffe

Andy Miller

Ryan Adams

Debbie Bard

Dustin Lang

Aaron Meisner

David Schlegel

Prabhat

Project goals

1. 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.
2. Identify promising galaxies for spectrograph targeting.
    - Better understand dark energy and the geometry of the universe.
3. Replace Photo, a carefully hand-tuned heuristic for building astronomical catalogs from photometric data
    - It isn't obvious that we can outperform Photo.
4. Develop an extensible model and inference procedure, for use by the astronomical community.
    - Future applications might include finding supernovae and detecting near-Earth asteroids.

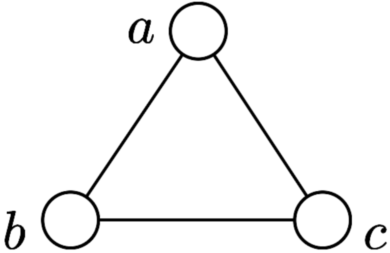
An astronomical image

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.

Outline

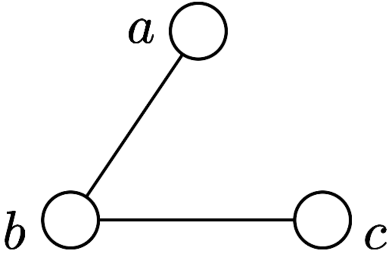
1. review graphical models
2. our graphical model for astronomical images (Celeste)
3. review approximate inference
4. approximate inference for Celeste
5. scaling Celeste to catalog the universe

Our approach: graphical models



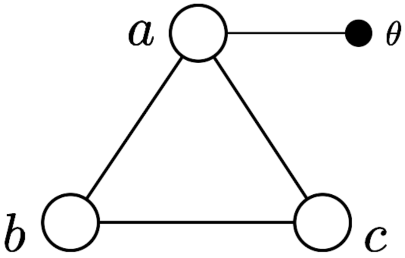
$$p(a, b, c) = p(c|a, b)p(b|a)p(a)$$

Our approach: graphical models



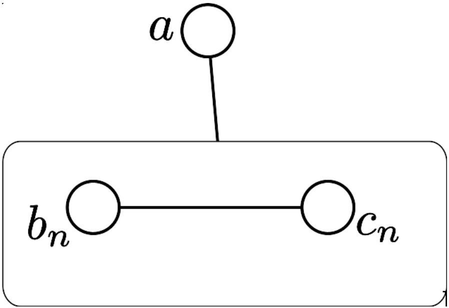
$$p(a, b, c) = p(c|b)p(b|a)p(a)$$

Our approach: graphical models



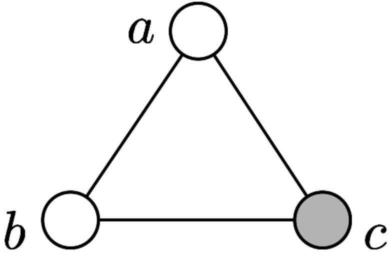
$$p(a, b, c) = p(c|a, b)p(b|a)p_\theta(a)$$

Our approach: graphical models



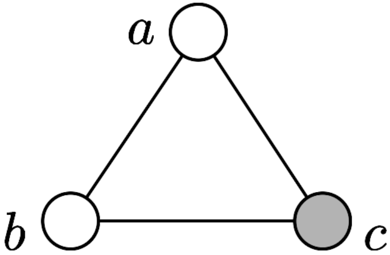
$$p(a, \mathbf{b}, \mathbf{c}) = \prod_{n=1}^{N} \left[p(c_n | a, b_n) p(b_n | a) \right] p(a)$$

Our approach: graphical models



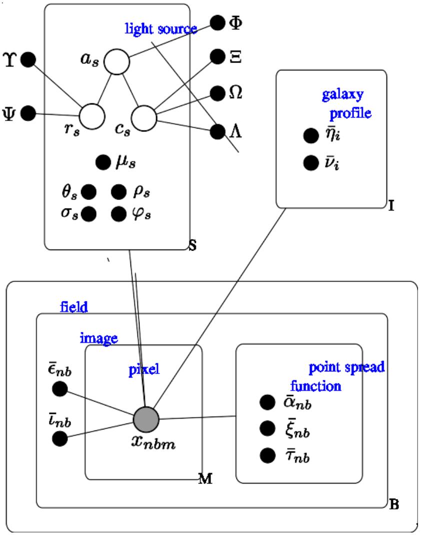
$$p(a, b, c) = p(c|a, b)p(b|a)p(a)$$

Our approach: posterior inference



$$\underbrace{p(a, b|c)}_{\text{posterior}} \propto \underbrace{p(c|a, b)}_{\text{likelihood}} \underbrace{p(b|a)p(a)}_{\text{prior}}$$

The Celeste graphical model



Brightness priors

For light source s, the energy emitted in the r band

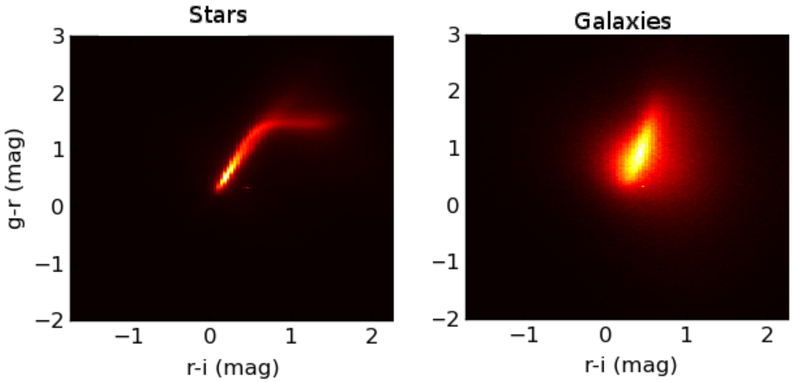
$$r_s|(a_s = i) \sim \mathrm{Gamma}\left(\Upsilon^{(i)}, \Psi^{(i)}\right).$$

The log ratios of brightnesses in adjacent bands (the *colors*)

$$c_s|(a_s = i) \sim \mathrm{MvNormalMix}\left(\Omega^{(i)}, \Lambda^{(i)}, \Xi^{(i)}\right).$$

Then the brightness ℓ_{sb} in each band b is a deterministic function of r_s and c_s.

Scientific color priors



Galaxies: light-density model

The light density for galaxy s is modeled as mixture of two extremal galaxy prototypes:

$$h_s(w) = \theta_s h_{s1}(w) + (1 - \theta_s) h_{s0}(w).$$

Each prototype ($i = 0$ or $i = 1$) is a mixture of bivariate normal distributions:

$$h_{si}(w) = \sum_{j=1}^{J} \bar{\eta}_{ij} \phi(w; \mu_s, \bar{\nu}_{ij} Q_s).$$

Shared covariance matrix Q_s accounts for the scale σ_s, rotation φ_s, and axis ratio ρ_s.

An elliptical galaxy, $\theta_s = 0$

A spiral galaxy, $\theta_s = 1$

Idealized sky view

The brightness for sky position w is

$$G_b(w) = \sum_{s=1}^{S} \ell_{sb} g_s(w)$$

where

$$g_s(w) = \begin{cases} \mathbf{1}\{\mu_s = w\}, & \text{if } a_s = 0 \text{ (``star'')} \\ h_s(w), & \text{if } a_s = 1 \text{ (``galaxy'').} \end{cases}$$

Astronomical images

Images differ from the idealized sky view due to

1. pixelation and point spread

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

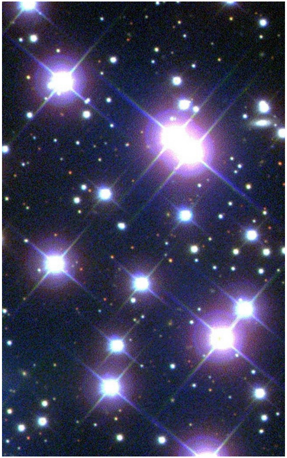
$$G_{nbm} = G_b * f_{nbm}$$

2. background radiation and calibration

$$F_{nbm} = \iota_{nb} \left[\epsilon_{nb} + G_{nbm} \right]$$

3. finite exposure duration

$$x_{nbm} | \left(a_s, r_s, c_s \right)_{s=1}^{S} \sim \mathrm{Poisson} \left(F_{nbm} \right)$$



Intractable posterior

Let $\Theta = (a_s, r_s, c_s)_{s=1}^{S}$. The posterior on Θ is intractable because of coupling between the sources:

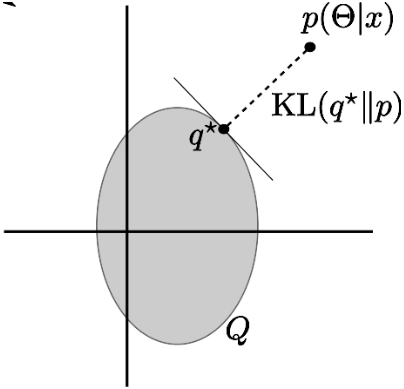
$$p(\Theta|x) = \frac{p(x|\Theta)p(\Theta)}{p(x)}$$

and

$$\begin{aligned} p(x) &= \int p(x|\Theta)p(\Theta)\, d\Theta \\ &= \int \prod_{n=1}^{N} \prod_{b=1}^{B} \prod_{m=1}^{M} p(x_{nbm}|\Theta)p(\Theta)\, d\Theta. \end{aligned}$$

Variational inference

Variational inference approximates the exact posterior p with a simpler distribution $q^\star \in Q$.



Variational inference

The optimization problem can be written without $p(x)$ or $p(\Theta|x)$:

$$\begin{aligned}
q^{\star} &= \arg\min_{q\in Q} \left[\mathrm{KL}(\, q(\theta) \parallel p(\Theta|x)\,)\right] &(1)\\
&= \arg\min_{q\in Q} \left[\mathbb{E}_q \log q(\theta) - \mathbb{E}_q \log p(\Theta|x)\right] &(2)\\
&= \arg\min_{q\in Q} \left[\mathbb{E}_q \log q(\theta) - \mathbb{E}_q \log p(\Theta, x) - \log p(x)\right] &(3)\\
&= \arg\min_{q\in Q} \left[\mathbb{E}_q \log q(\theta) - \mathbb{E}_q \log p(\Theta, x)\right]. &(4)
\end{aligned}$$

Variational inference vs. MCMC

VI advantages:

- potentially orders of magnitude faster than MCMC
- no unknown "mixing time"
- no post-processing of samples—can compute statistics of the approximating distribution almost instantly

VI 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

$$
\begin{aligned}
&\mathcal{L}(\chi, \mu, \kappa, \gamma, \zeta, \beta, \lambda, \theta, \rho, \sigma, \varphi) \\
&= C + \sum_{n=1}^{N} \sum_{b=1}^{B} \sum_{m=1}^{M} \Bigg\{ \sum_{a \in \{0,1\}^S} \prod_{s=1}^{S} \chi_s^{a_s} (1-\chi_s)^{1-a_s} \Bigg\{ \int_{r_1} \int_{c_1} \int_{k_1} \cdots \int_{r_S} \int_{c_S} \int_{k_S} \\
&\qquad x_{nbm} \log \Bigg[\epsilon_{nb} + \sum_{s=1}^{S} r_s \prod_{j=b}^{b_r} \exp\{c_{sb}\} \prod_{j=b_r}^{b-1} \exp\{c_{sb}\} \\
&\qquad\qquad \times \int \sum_{k=1}^{3} \bar{\alpha}_{nk} \phi\left(m - w; \bar{\xi}_{nbk}, \bar{\Sigma}_{nbk}\right) g_{si}(w)\, dw \Bigg] \\
&\qquad - \iota_{nb} \sum_{s=1}^{S} r_s \prod_{j=b}^{b_r} \exp\{c_{sb}\} \prod_{j=b_r}^{b-1} \exp\{c_{sb}\} \int \sum_{k=1}^{3} \bar{\alpha}_{nk} \phi\left(m - w; \bar{\xi}_{nbk}, \bar{\Sigma}_{nbk}\right) g_{si}(w)\, dw \\
&\quad dr_1 \; dc_1 \; dk_1 \; \ldots \; dr_S \; dc_S \; dk_S \Bigg\} \Bigg\} \\
&\quad - \sum_{s=1}^{S} \Bigg\{ D_{\mathrm{KL}}(q(a_s), p(a_s)) + \sum_{i=1}^{2} \chi_s^{a_s} (1-\chi_s)^{1-a_s} \\
&\qquad \times \left[D_{\mathrm{KL}}(q(r_s|a_s = i), p(r_s|a_s = i)) + D_{\mathrm{KL}}\left(q(k_s, c_s|a_s = i), p_s(k_s, c_s|a_s = i)\right)\right] \Bigg\}.
\end{aligned}
$$

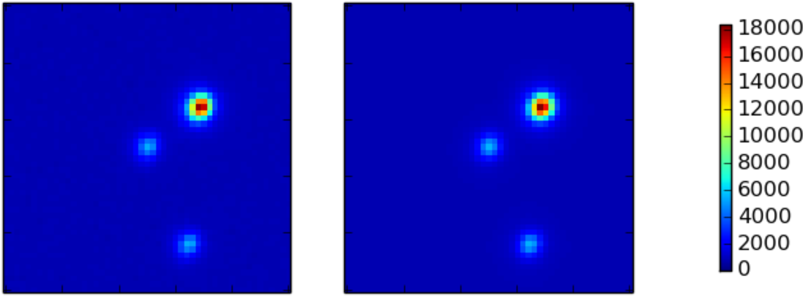
However,

- An approximating distribution that factorizes across light sources (a "structured mean-field" assumption) makes most expectations tractable:

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

- The delta method for moments approximates the remaining expectations.
- Existing catalogs provide good initial settings for the variational parameters.
- Light sources are unlikely to contribute photons to distant pixels.
- The model contains an auxiliary variable indicating the mixture component that generated each source's colors.

Posterior predictive check on a "stamp"



Left is a 51 pixel $\times$ 51 pixel sub-region of an astronomical image, captured through the $\mathtt{r}$ band filter. Each pixel's value corresponds to the number of photons that hit it. The right panel shows $\mathbb{E}_{q^\star}[F_{nbm}]$, the mean of x_{nbm} with respect to our posterior approximation.

Results (June 2015)

	photo	celeste	improve
position	0.22	**0.20**	.02 (.00)
missed gals	28 / 654	**15 / 654**	.02 (.01)
missed stars	**8 / 654**	31 / 654	-.04 (.01)
color u-g	1.10	**0.49**	.61 (.04)
color g-r	0.16	**0.09**	.07 (.01)
color r-i	0.09	**0.06**	.03 (.00)
color i-z	0.25	**0.10**	.15 (.01)
brightness	**0.76**	1.60	-.83 (.12)
profile	0.19	0.23	-.04 (.02)
axis ratio	0.17	**0.13**	.04 (.01)
scale	**0.37**	1.28	-.91 (.17)
angle	19.40	18.10	1.40 (.80)

Jeffrey Regier, Andrew Miller, Jon McAuliffe, Ryan Adams, Matthew Hoffman, Dustin Lang, David Schlegel, and Prabhat. *Celeste: Variational inference for a generative model of astronomical images.* ICML 2015.

Results on synthetic data

	real		synthetic
	photo	celeste	celeste
position	0.22	0.20	**0.08**
missed gals	28 / 654	15 / 654	**14 / 654**
missed stars	8 / 654	31 / 654	**6 / 654**
color u-g	1.10	0.49	**0.20**
color g-r	0.16	0.09	**0.05**
color r-i	0.09	0.06	**0.04**
color i-z	0.25	0.10	**0.08**
brightness	0.76	1.60	**0.29**
profile	0.19	0.23	**0.16**
axis ratio	0.17	0.13	**0.11**
scale	0.37	1.28	**0.23**
angle	19.40	18.10	**14.90**

Model misfit leaves room for improvement. Enhancing the galaxy model is a promising research direction.

Cataloging the universe

SDSS dataset

- 20 TB of images
- 500 million stars and galaxies
- 100 GB catalog–scales linearly with input size
- 5 million CPU-hour budget

Scaling

- exact Hessian for Newton's method with a trust region
  - Ryan Giordano (UCB)
- object-level parallelism (block coordinate descent)
  - Kyle Barbary (Berkeley Center for Cosmological Physics)
- pixel-level parallelism for "Knights Landing" processors
  - Kiran Pramnany (Intel)

github.com/jeff-regier/Celeste.jl