The Astronomer’s Theory of Everything

David W. Hogg

Center for Cosmology and Particle Physics, New York University

and

Max-Planck-Institut für Astronomie, Heidelberg

2013 March 5
XDQSO target selection (Bovy et al., 1011.6392)
lots of bad data is equal to a bit of good data
Software is going to be much more productive, per dollar, than anything else in astrophysics!
the new reality

- all our goals are getting substantially more ambitious every ten years
- funding is flat
- the era of under-budgeting and over-designing (and getting away with it) is over
- many observational programs will lead to upper limits or bare detections
  - example: *JWST* will take low $s/n$ spectra of a *couple* of Earth-like exoplanets (in the most optimistic scenario) over its entire mission lifetime.
- existing (and aging) data will become far more valuable
  - example: The sky has been imaged thousands of times already in many bands; these data have never seriously been assembled and analyzed coherently.
we need new tools

- How do we move all of the information from all of the data (ever taken) to the quantities of interest?
- How do we get high $s/n$ information about what we want when every individual datum is very noisy?
- It’s all about tools.
conclusions

- software, software, software, and applied math
- lots of bad data is equal to a bit of good data
- modeling beats supervised classification
- point estimates are bad, models are good
- hierarchical modeling will require MCMC
- we can find populations, no member of which is individually detectable
modeling beats supervised classification
XDQSO target selection (Bovy et al., 1011.6392)
XDQSO target selection (Bovy et al., 1011.6392)
XDQSO target selection (Bovy et al., 1011.6392)
XDQSO target selection (Bovy et al., 1011.6392)
what’s wrong with supervised classification?

- support vector machines, boosting, deep learning
- these are all awesome
- they require that test data have the same statistical and error properties as training data
  - never true!
- they require that all features be measured for all data points
  - never true!
  - (If you know enough about your data to fix this problem, then just write down a likelihood!)
point estimates are bad, models are good
faint proper motions (Lang et al. 0808.4004)
faint proper motions (Lang et al. 0808.4004)
faint proper motions (Lang et al. 0808.4004)
faint proper motions (Lang et al. 0808.4004)
faint proper motions (Lang et al. 0808.4004)

- If we had only a catalog, we would have failed.
- If we had only a coadd, we would have failed.
what’s wrong with LSST and PanSTARRS?

- reducing data with point estimates
- building catalogs from “co-adds” with point estimates
- catalog matching
- All of these throw away information. Does it matter?
  - Lang and I are betting it does: theTractor.org
The *Tractor* (Lang *et al.*)
The *Tractor* (Lang *et al.*
heirarchical modeling will require MCMC
the Exoplanet Theory of Everything

- different techniques find different planets
  - radial velocity, transit, direct detection, astrometry, microlensing
- completeness or selection functions are smooth functions of exoplanet and host-star properties
- most observations show no clearly detectable planet
- Earth-like planets are in a “bitter spot” for all observational techniques
- how do we take the (literally) billions of data points and obtain the best possible picture of the full exoplanet population?
  - distributions in orbital and planetary (composition, size) parameters
  - distributions for multiplicity and “architecture”
  - all as a function of host-star properties
  - plus all the individual systems measured as well as possible
observations
planets
$r$
$f_\theta(t_n)$
$f_{\text{obs}; n}$
$t_0, a, e, \varpi, \delta i, M_p$
$R_*, f_*, \{I_k\}$
$i$
$\delta^2$
$t_n$
$\sigma_n^2$
$t_{\exp n}$
$M_*$
Generated using Bart v0.0.2-2f9b755

- Short cadence
- Long cadence

Time [Hours Since Transit]
Bart (Foreman-Mackey et al., forthcoming)

- built on very successful emcee package (Foreman-Mackey et al., 1202.3665)
- designed for exoplanet measurement and discovery of false positives
- very easy to use

```python
import bart

# Initialize a planet.
planet = bart.Planet(r=0.01, a=21.3, t0=3.85)
planet.parameters += [bart.parameters.Parameter(r"$r$", "r"),
                      bart.parameters.LogParameter(r"$a$", "a")]

# Initialize the star.
ldp = bart.kepler.fiducial_ldp(teff=6438, logg=4.28, feh=0.0)
star = bart.Star(mass=planet.get_mstar(12.4138), ldp=ldp)

# Set up the system.
system = bart.PlanetarySystem(star)
system.parameters.append(bart.parameters.CosParameter(r"$i$", "iobs"))
system.add_planet(planet)

# Add data and fit.
system.add_dataset(bart.KeplerDataset("path/to/kepler/data/lc.fits"))
system.fit(2000)
```
exoplanets around white dwarfs (Schiminovich, Lang, Hogg)
exoplanets around white dwarfs (Schiminovich, Lang, Hogg)
exoplanets around white dwarfs (Schiminovich, Lang, Hogg)
exoplanets around white dwarfs (Schiminovich, Lang, Hogg)
the undetectable can be measured
ersatz 007

radial velocity (kms⁻¹)

time (d)
ersatz stacked and binned
family of priors $p(\omega_n|\alpha)$, parameterized by some $\alpha$

\[
p(\{D_n\}_{n=1}^{N} | \alpha) = \prod_{n=1}^{N} \int d\omega_n p(D_n|\omega_n) p(\omega_n|\alpha) \quad (1)
\]

if you believe there can be a likelihood, then you believe there can be a **marginalized likelihood**

the fact that each internal $p(D_n|\omega_n)$ contains no clear peak (no clear object detection at all) doesn’t change anything!
hierarchical inference: What does it require?

- accurate likelihood functions
  - accurate noise models, or parameterized noise models
- fast inference
  - self-tuning MCMC (like *emcee*; Foreman-Mackey *et al.*, 1202.3665)
  - robustness to multimodal likelihood functions
- concept of self-calibration
  - calibration and noise parameters are not different from astrophysical parameters
- racks and racks of metal
  - (it can’t be done in “map–reduce” framework)
hierarchical inference: Why does it work?

- The marginalized likelihood is large when there is high prior probability in locations where there is high likelihood.
- When likelihoods are broad, the best prior is the most concentrated prior that is “consistent with” all individual-object likelihood functions.
- The operation is a **heteroskedastic deconvolution**.
  - (in modern parlance, a “deconvolution” is always the result of fitting a generative or forward model)
conclusions

- software, software, software, and applied math
- lots of bad data is equal to a bit of good data
- modeling beats supervised classification
- point estimates are bad, models are good
- hierarchical modeling will require MCMC
- we can find populations, no member of which is individually detectable