UMass Bayesian Inference Engine

Martin D. Weinberg
UMass Astronomy
weinberg@astro.umass.edu

May 5, 2008

Current team

- Neal Katz (Astro)
- Michael Lavine (Math)
- Houjun Mo (Astro)
- Eliot Moss (CS)
- Byn Choi (CS)
- Joerg Colberg (Astro)
- Ilsang Yoon (Astro)
- Lu Yu (Astro)

Motivation

- Multi-terabyte data catalogs (2MASS, SDSS, GOODS, etc.)
- In-principle solution to the inference problem: MCMC
  - Incorporate data from multiple catalogs
  - Can merge data sources with different attributes
- Current packages—*Bayespack, BUGS, S-Plus, R*
  - Not production oriented
  - Although good for proof of concept

Motivation

- Multi-terabyte data catalogs (2MASS, SDSS, GOODS, etc.)
- In-principle solution to the inference problem: MCMC
  - Incorporate data from multiple catalogs
  - Can merge data sources with different attributes
- Current packages—*Bayespack, BUGS, S-Plus, R*
  - Not production oriented
  - Although good for proof of concept

Motivation

- Multi-terabyte data catalogs (2MASS, SDSS, GOODS, etc.)
- In-principle solution to the inference problem: MCMC
  - Incorporate data from multiple catalogs
  - Can merge data sources with different attributes
- Current packages—*Bayespack, BUGS, S-Plus, R*
  - Not production oriented
  - Although good for proof of concept

Motivation

- Multi-terabyte data catalogs (2MASS, SDSS, GOODS, etc.)
- In-principle solution to the inference problem: MCMC
  - Incorporate data from multiple catalogs
  - Can merge data sources with different attributes
- Current packages—*Bayespack, BUGS, S-Plus, R*
  - Not production oriented
  - Although good for proof of concept

$\Rightarrow$Bayesian Inference Engine

General Goals for the BIE

1. Perform inference & hypothesis testing on large-volume survey data
2. Apply advanced computational techniques to optimize Bayesian methodology
   - Exploit the intrinsic parallelism in MCMC
   - Embed structures optimized for mapped data
   - Multiple data sets incorporated by general data stream architecture (consumer-producer chains)
   - Library, with front-end parser (like Octave)
3. Platform for future development and statistical research

General Goals for the BIE

1. Perform inference & hypothesis testing on large-volume survey data
2. Apply advanced computational techniques to optimize Bayesian methodology
   - Exploit the intrinsic parallelism in MCMC
   - Embed structures optimized for mapped data
   - Multiple data sets incorporated by general data stream architecture (consumer-producer chains)
   - Library, with front-end parser (like Octave)
3. Platform for future development and statistical research

General Goals for the BIE

1. Perform inference & hypothesis testing on large-volume survey data
2. Apply advanced computational techniques to optimize Bayesian methodology
   - Exploit the intrinsic parallelism in MCMC
   - Embed structures optimized for mapped data
   - Multiple data sets incorporated by general data stream architecture (consumer-producer chains)
   - Library, with front-end parser (like Octave)
3. Platform for future development and statistical research

General Goals for the BIE

Motivation
Goals
ProbStat
HEP
Advantages
Software features
Killer applications
Semi-analytic models
GALPHAT
Galphat results
Status
M-H #1
M-H #2
Tempered-states #1
Tempered-states #2
Tempered-states #3
HEP details

1. Perform inference & hypothesis testing on large-volume survey data
2. Apply advanced computational techniques to optimize Bayesian methodology
   - Exploit the intrinsic parallelism in MCMC
   - Embed structures optimized for mapped data
   - Multiple data sets incorporated by general data stream architecture (consumer-producer chains)
   - Library, with front-end parser (like Octave)
3. Platform for future development and statistical research

General Goals for the BIE

1. Perform inference & hypothesis testing on large-volume survey data
2. Apply advanced computational techniques to optimize Bayesian methodology
   - Exploit the intrinsic parallelism in MCMC
   - Embed structures optimized for mapped data
   - Multiple data sets incorporated by general data stream architecture (consumer-producer chains)
   - Library, with front-end parser (like Octave)
3. Platform for future development and statistical research

   Statistical, computational and astronomical research on the same platform!

Specific Goal: Inference!

- Parametric modelling
- Semi-parametric modelling
  - Parametric in dimensions with strong astronomical prior knowledge (an exponential disk, Sérsic bulge)
  - Non-parametric in dimensions with little prior knowledge (asymmetry analysis)
- Full non-parametric modeling (basis sets, Polya-tree priors)
- General hypothesis testing
  - Without nested models . . .
  - More than model selection
  - Complex hypotheses

Specific Goal: Inference!

- Parametric modelling
- Semi-parametric modelling
  - Parametric in dimensions with strong astronomical prior knowledge (an exponential disk, Sérsic bulge)
  - Non-parametric in dimensions with little prior knowledge (asymmetry analysis)
- Full non-parametric modeling (basis sets, Polya-tree priors)
- General hypothesis testing
  - Without nested models . . .
  - More than model selection
  - Complex hypotheses

Specific Goal: Inference!

Motivation
Goals
ProbStat
HEP
Advantages
Software features
Killer applications
Semi-analytic models
GALPHAT
Galphat results
Status
M-H #1
M-H #2
Tempered-states #1
Tempered-states #2
Tempered-states #3
HEP details

- Parametric modelling
- Semi-parametric modelling
  - Parametric in dimensions with strong astronomical prior knowledge (an exponential disk, Sérsic bulge)
  - Non-parametric in dimensions with little prior knowledge (asymmetry analysis)
- Full non-parametric modeling (basis sets, Polya-tree priors)
- General hypothesis testing
  - Without nested models . . .
  - More than model selection
  - Complex hypotheses

Specific Goal: Inference!

- Parametric modelling
- Semi-parametric modelling
  - Parametric in dimensions with strong astronomical prior knowledge (an exponential disk, Sérsic bulge)
  - Non-parametric in dimensions with little prior knowledge (asymmetry analysis)
- Full non-parametric modeling (basis sets, Polya-tree priors)
- General hypothesis testing
  - Without nested models . . .
  - More than model selection
  - Complex hypotheses

Probability and Statistics

- Standard Metropolis-Hastings Details
  - Bayes: $\pi \rightarrow P(H|D) \propto P(H)P(D|H)$
  - Detailed balance: $\pi(x)T(x,x') = \pi(x')T(x',x)$
- Tempered-states annealing (R. Neal) Details
- Multiple chains:
  - Dispersed chains, same temperature
  - Parallel tempered chains
  - Differential evolution
- Particle filter
- Convergence analysis

Probability and Statistics

- Standard Metropolis-Hastings Details
  - Bayes: $\pi \rightarrow P(H|D) \propto P(H)P(D|H)$
  - Detailed balance: $\pi(x)T(x, x') = \pi(x')T(x', x)$
- Tempered-states annealing (R. Neal) Details
- Multiple chains:
  - Dispersed chains, same temperature
  - Parallel tempered chains
  - Differential evolution
- Particle filter
- Convergence analysis

- Standard Metropolis-Hastings Details
  - Bayes: $\pi \rightarrow P(H|D) \propto P(H)P(D|H)$
  - Detailed balance: $\pi(x)T(x, x') = \pi(x')T(x', x)$
- Tempered-states annealing (R. Neal) Details
- Multiple chains:
  - Dispersed chains, same temperature
  - Parallel tempered chains
  - Differential evolution
- Particle filter
- Convergence analysis

Probability and Statistics

- Standard Metropolis-Hastings Details
  - Bayes: $\pi \rightarrow P(H|D) \propto P(H)P(D|H)$
  - Detailed balance: $\pi(x)T(x,x') = \pi(x')T(x',x)$
- Tempered-states annealing (R. Neal) Details
- Multiple chains:
  - Dispersed chains, same temperature
  - Parallel tempered chains
  - Differential evolution
- Particle filter
- Convergence analysis

Probability and Statistics

- Standard Metropolis-Hastings Details
  - Bayes: $\pi \rightarrow P(H|D) \propto P(H)P(D|H)$
  - Detailed balance: $\pi(x)T(x, x') = \pi(x')T(x', x)$
- Tempered-states annealing (R. Neal) Details
- Multiple chains:
  - Dispersed chains, same temperature
  - Parallel tempered chains
  - Differential evolution
- Particle filter
- Convergence analysis

Probability and Statistics

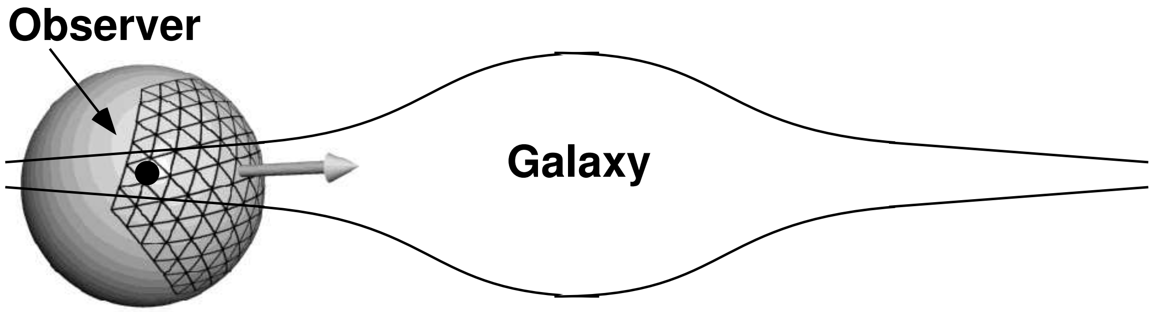
- Standard Metropolis-Hastings Details
  - Bayes: $\pi \to P(H|D) \propto P(H)P(D|H)$
  - Detailed balance: $\pi(x)T(x,x') = \pi(x')T(x',x)$
- Tempered-states annealing (R. Neal) Details
- Multiple chains:
  - Dispersed chains, same temperature
  - Parallel tempered chains
  - Differential evolution
- Particle filter
- Convergence analysis

Probability and Statistics

- Standard Metropolis-Hastings Details
  - Bayes: $\pi \rightarrow P(H|D) \propto P(H)P(D|H)$
  - Detailed balance: $\pi(x)T(x,x') = \pi(x')T(x',x)$
- Tempered-states annealing (R. Neal) Details
- Multiple chains:
  - Dispersed chains, same temperature
  - Parallel tempered chains
  - Differential evolution
- Particle filter
- Convergence analysis

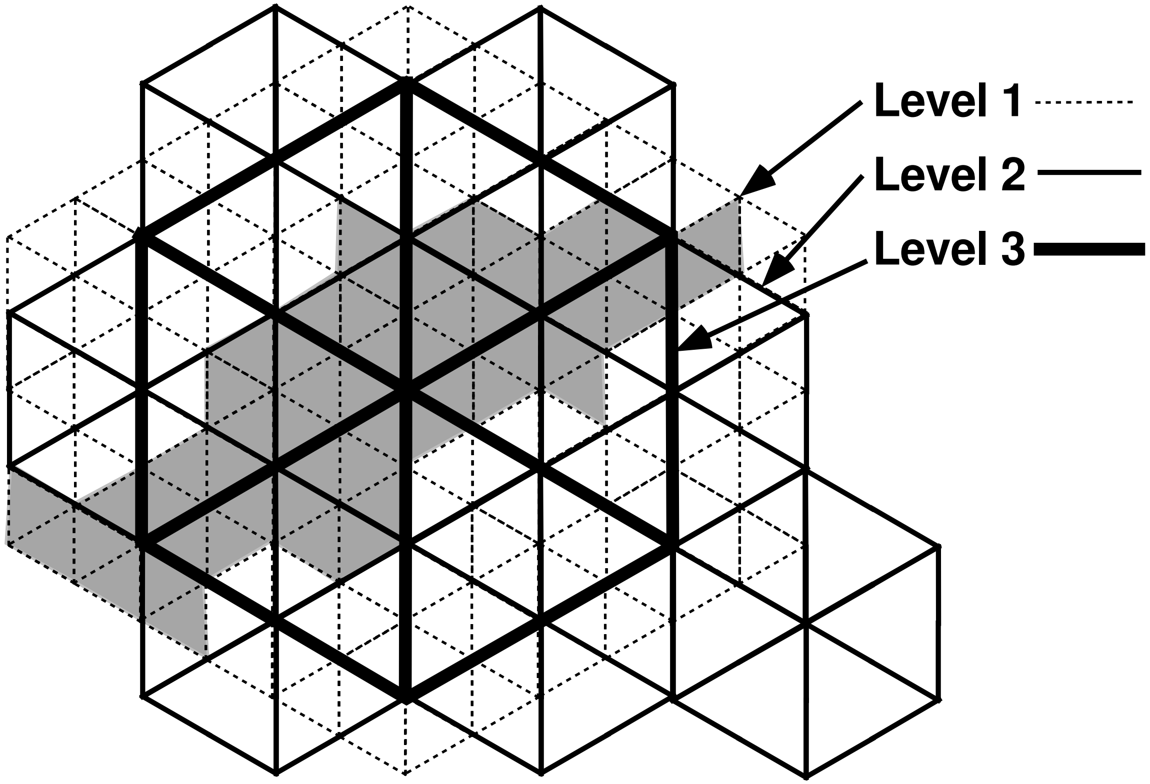
HEP

- Multilevel resolution: Hierarchical Empirical Priors



HEP

- Multilevel resolution: Hierarchical Empirical Priors



Details

Summary: advantages

- Multilevel hierarchical prior:
  - Eliminates high-dim volume at low resolution
  - Degrades the influence of prior on convergence
  - Improves MCMC convergence
  - Rigorously conserves probability
- Multiple temperatures, “Loosened constraints” (with M. Lavine)
  - Accelerates convergence and mixing
  - Prevents getting “stuck” in local minima

Summary: advantages

Motivation
Goals
ProbStat
HEP
Advantages
Software features
Killer applications
Semi-analytic models
GALPHAT
Galphat results
Status
M-H #1
M-H #2
Tempered-states #1
Tempered-states #2
Tempered-states #3
HEP details

- Multilevel hierarchical prior:
  - Eliminates high-dim volume at low resolution
  - Degrades the influence of prior on convergence
  - Improves MCMC convergence
  - Rigorously conserves probability
- Multiple temperatures, “Loosened constraints” (with M. Lavine)
  - Accelerates convergence and mixing
  - Prevents getting “stuck” in local minima

Software features

- Automated model integration and convergence testing

Software features

- Automated model integration and convergence testing
- MPI implementation for clusters and supercomputers

Software features

- Automated model integration and convergence testing
- MPI implementation for clusters and supercomputers
- Object-oriented design, easily extensible

Software features

- Automated model integration and convergence testing
- MPI implementation for clusters and supercomputers
- Object-oriented design, easily extensible
- Persistence system: save, checkpoint, recall objects!

Software features

- Automated model integration and convergence testing
- MPI implementation for clusters and supercomputers
- Object-oriented design, easily extensible
- Persistence system: save, checkpoint, recall objects!
- Non-Bayesian “data mining”

Software features

- Automated model integration and convergence testing
- MPI implementation for clusters and supercomputers
- Object-oriented design, easily extensible
- Persistence system: save, checkpoint, recall objects!
- Non-Bayesian "data mining"

Screen shots: GUI Visualizer Strip Contour

Killer applications

1. Star counts
2. Semi-analytic models (SAMS)
3. Galaxy image analysis

Semi-analytic models

- BIE-SAM: incorporates features from all major SAM groups
- Current practice: adjust parameters by hand to fit observed summary data (e.g. luminosity or mass function)
- Problem: no confidence estimates/can not achieve goal of rejecting phenomenological models
- Example: 8 parameter model given galaxy mass function

Semi-analytic models

- BIE-SAM: incorporates features from all major SAM groups
- Current practice: adjust parameters by hand to fit observed summary data (e.g. luminosity or mass function)
- Problem: no confidence estimates/can not achieve goal of rejecting phenomenological models
- Example: 8 parameter model given galaxy mass function

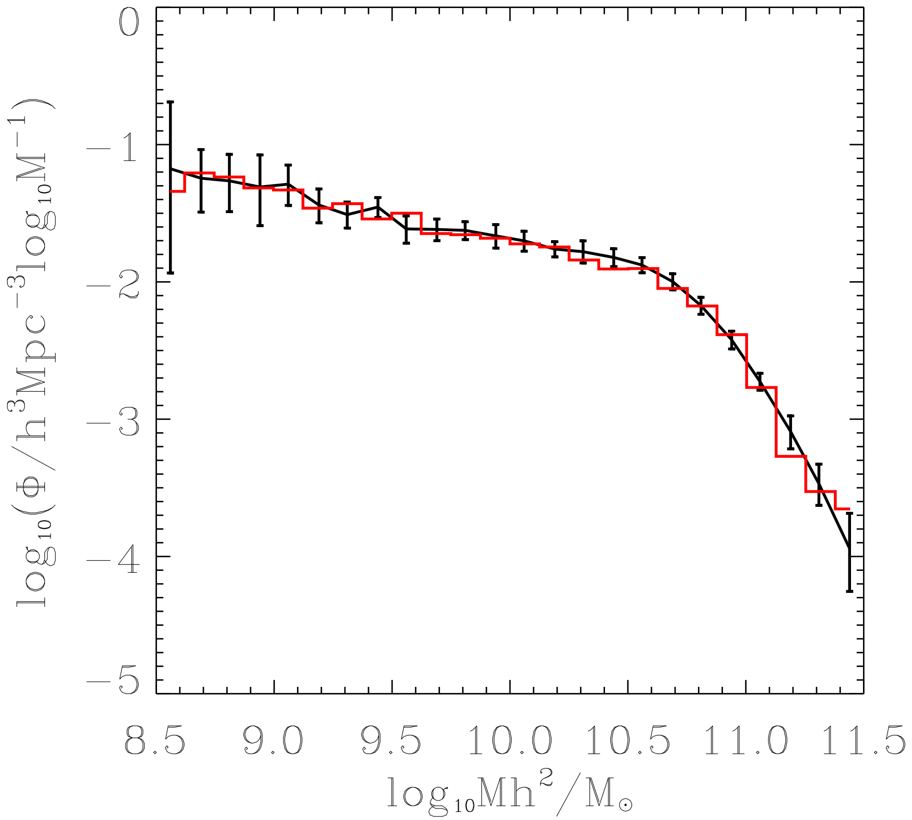
Semi-analytic models

- BIE-SAM: incorporates features from all major SAM groups
- Current practice: adjust parameters by hand to fit observed summary data (e.g. luminosity or mass function)
- Problem: no confidence estimates/can not achieve goal of rejecting phenomenological models
- Example: 8 parameter model given galaxy mass function

Semi-analytic models

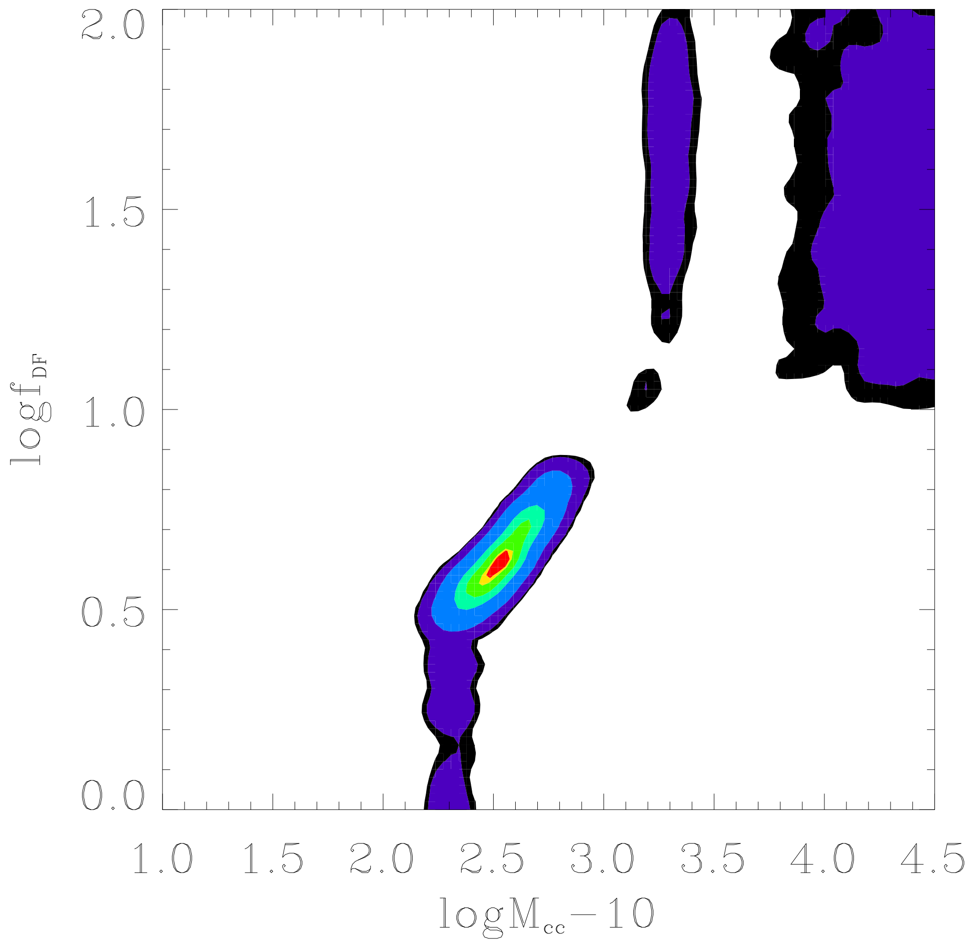
- BIE-SAM: incorporates features from all major SAM groups
- Current practice: adjust parameters by hand to fit observed summary data (e.g. luminosity or mass function)
- Problem: no confidence estimates/can not achieve goal of rejecting phenomenological models
- Example: 8 parameter model given galaxy mass function

Semi-analytic models



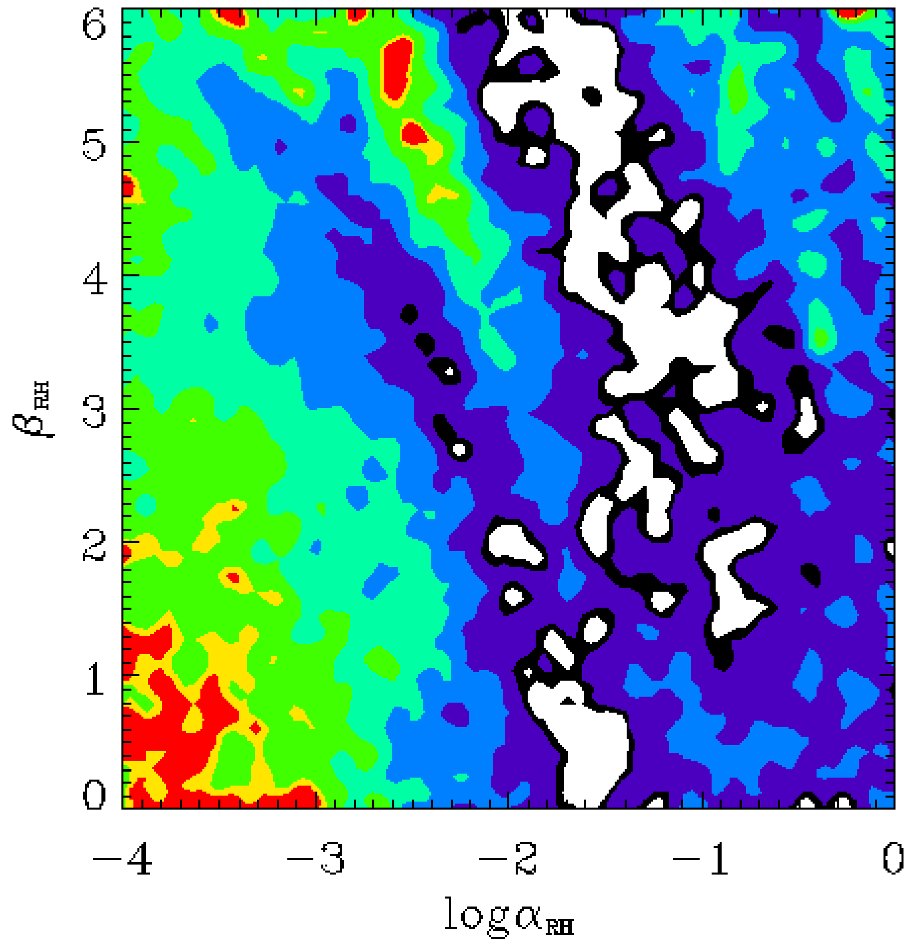
Predicted stellar mass function

Semi-analytic models



Cooling cut-off (M_H) vs. merging timescale

Semi-analytic models



Amplitude vs. power index of supernova reheating

Galaxy photometric attributes

- Multicomponent modeling (typically 12 parameters, more)
- Background estimation
- Adaptive integration, optimized with two-dimensional interpolation on cumulative distributions
- Rotation by FFT shear algorithm
- Scientific goals

GALaxy PHotometric ATtributes

- Multicomponent modeling (typically 12 parameters, more)
- Background estimation
- Adaptive integration, optimized with two-dimensional interpolation on cumulative distributions
- Rotation by FFT shear algorithm
- Scientific goals

GALPHAT: main features

- Multicomponent modeling (typically 12 parameters, more)
- Background estimation
- Adaptive integration, optimized with two-dimensional interpolation on cumulative distributions
- Rotation by FFT shear algorithm
- Scientific goals

GALPHAT: main features

- Multicomponent modeling (typically 12 parameters, more)
- Background estimation
- Adaptive integration, optimized with two-dimensional interpolation on cumulative distributions
- Rotation by FFT shear algorithm
- Scientific goals

GALPHAT: main features

- Multicomponent modeling (typically 12 parameters, more)
- Background estimation
- Adaptive integration, optimized with two-dimensional interpolation on cumulative distributions
- Rotation by FFT shear algorithm
- Scientific goals

GALPHAT: main features

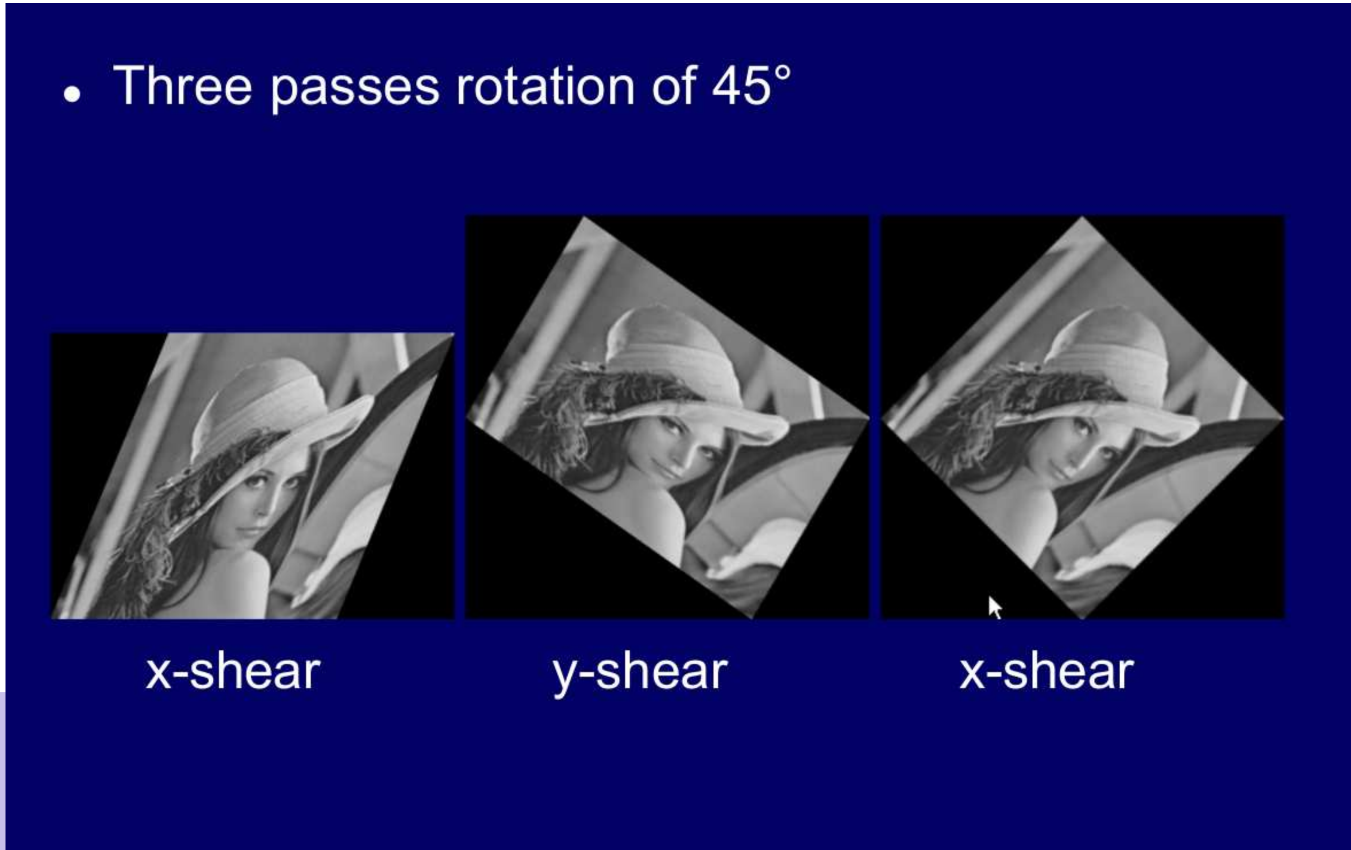
- Multicomponent modeling (typically 12 parameters, more)
- Background estimation
- Adaptive integration, optimized with two-dimensional interpolation on cumulative distributions
- Rotation by FFT shear algorithm
- Scientific goals

GALPHAT: main features

- Multicomponent modeling (typically 12 parameters, more)
- Background estimation
- Adaptive integration, optimized with two-dimensional interpolation on cumulative distributions
- Rotation by FFT shear algorithm
- Scientific goals

GALPHAT: main features

- Multicomponent modeling (typically 12 parameters, more)
- Background estimation
- Adaptive integration, optimized with two-dimensional interpolation on cumulative distributions
- Rotation by FFT shear algorithm



GALPHAT: main features

- Multicomponent modeling (typically 12 parameters, more)
- Background estimation
- Adaptive integration, optimized with two-dimensional interpolation on cumulative distributions
- Rotation by FFT shear algorithm
- Scientific goals
  1. Evaluation of galaxy evolution theories (model selection)
  2. Look for correlations between inferred parameters and everything else to gain insight (knowledge discovery)

GALPHAT: main features

- Multicomponent modeling (typically 12 parameters, more)
- Background estimation
- Adaptive integration, optimized with two-dimensional interpolation on cumulative distributions
- Rotation by FFT shear algorithm
- Scientific goals
  - Accurate recovery of bulge/disk ratios from 2MASS & SDSS
  - Extend to higher redshift (GOODS, GEMS): evolution of bulge/disk ratio
  - Hypothesis testing with full posterior probabilities

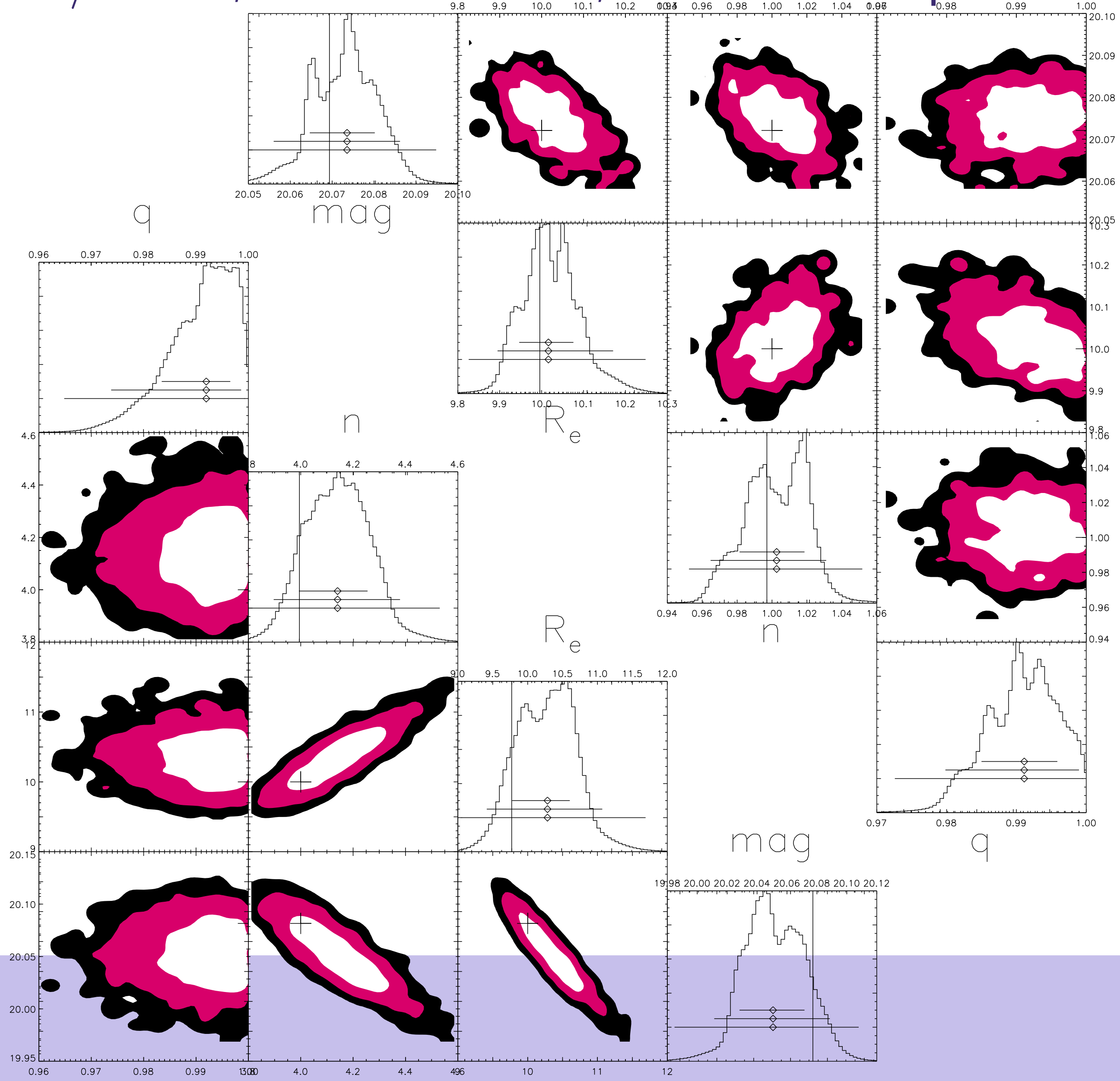
GALPHAT: results

- S/N=15, Sersic index: n, 1- and 2-component models
- 2MASS data

GALPHAT: results

- Motivation
- Goals
- ProbStat
- HEP
- Advantages
- Software features
- *Killer* applications
- Semi-analytic models
- GALPHAT
- **Galphat results**
- Status
- M-H #1
- M-H #2
- Tempered-states #1
- Tempered-states #2
- Tempered-states #3
- HEP details

- S/N=15, Sersic index: n, 1- and 2-component models

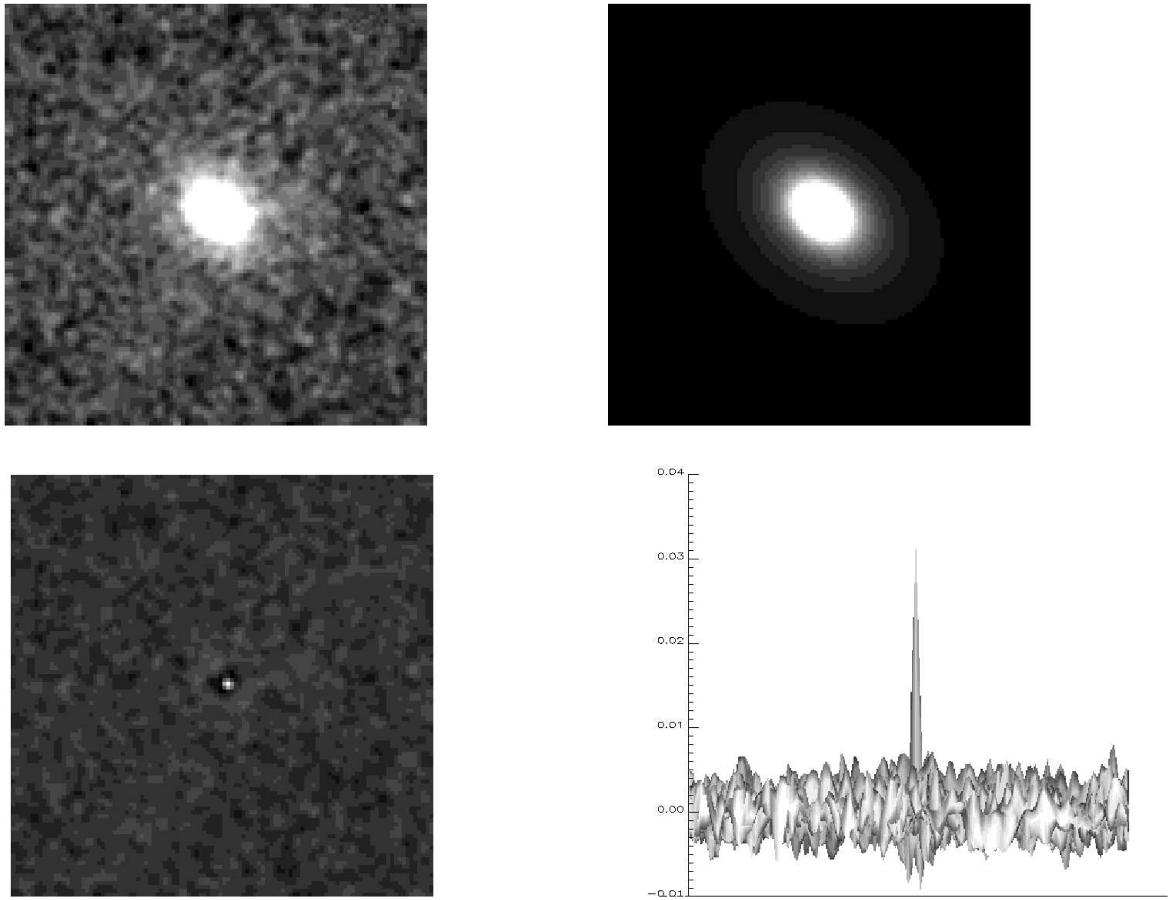


GALPHAT: results

- S/N=15, Sersic index: n, 1- and 2-component models
- 2MASS data

GALPHAT: results

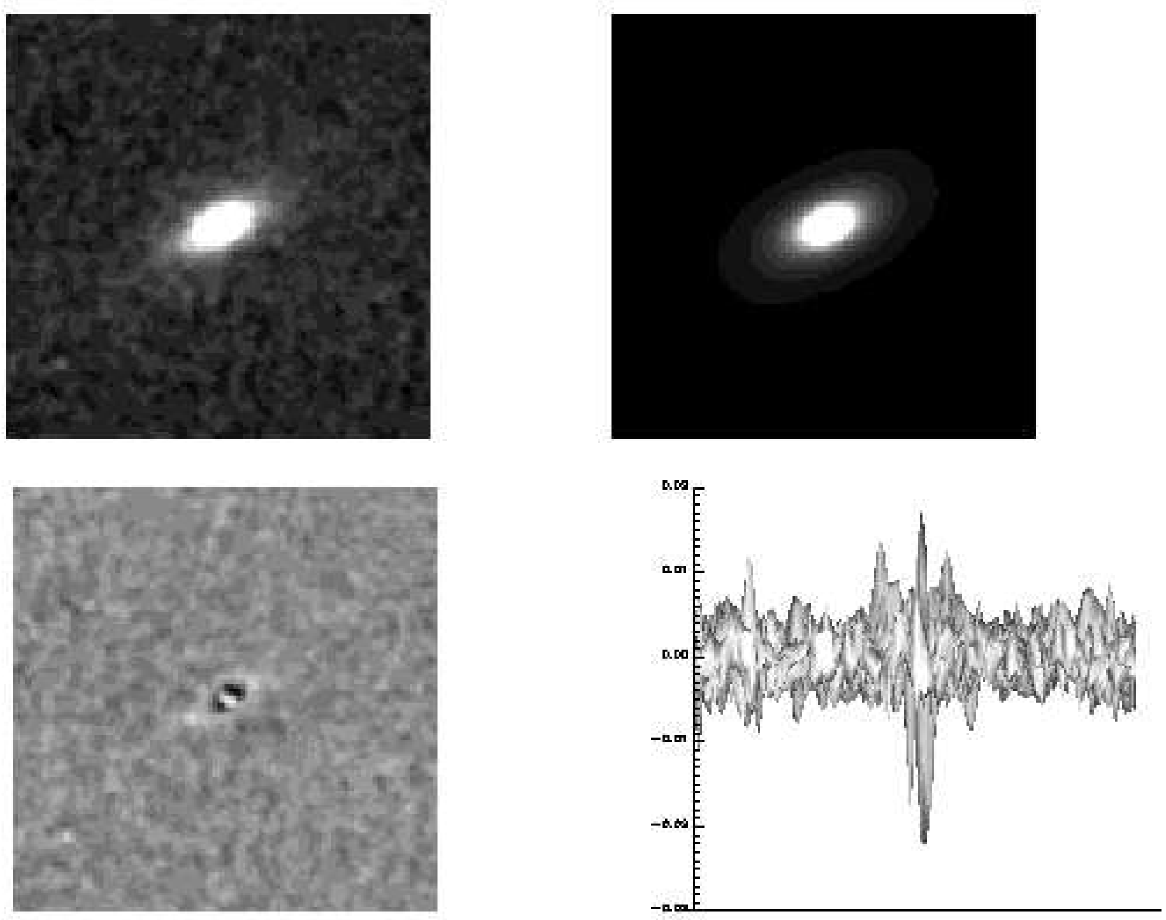
- S/N=15, Sersic index: n, 1- and 2-component models
- 2MASS data



NGC 137 [S0]

GALPHAT: results

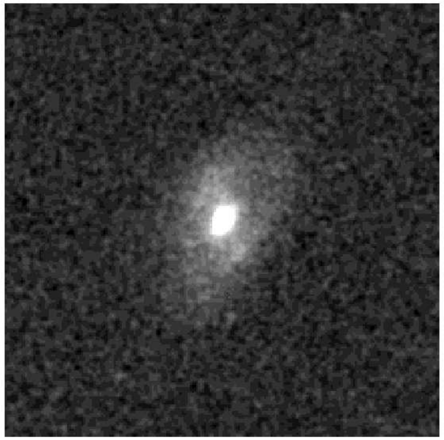
- S/N=15, Sersic index: n, 1- and 2-component models
- 2MASS data

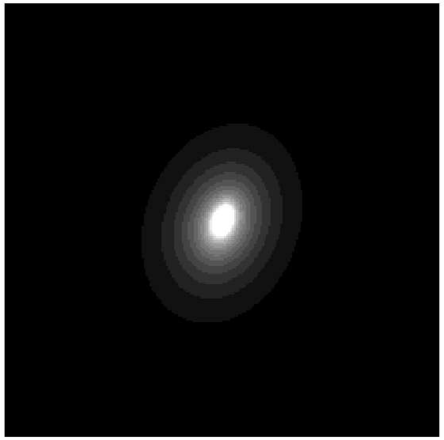


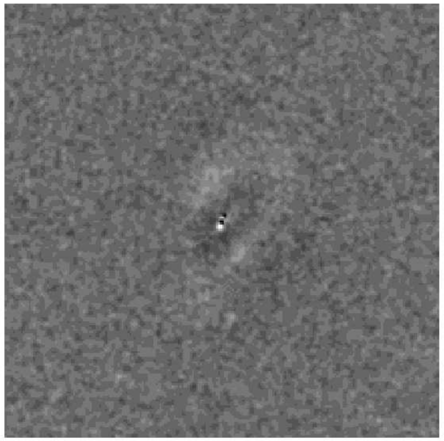
NGC 311 [S0]

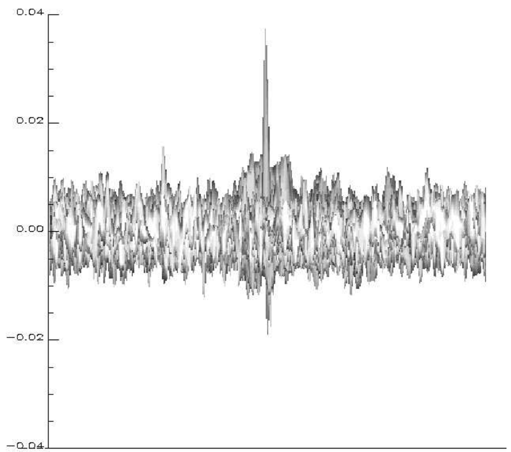
GALPHAT: results

- S/N=15, Sersic index: n, 1- and 2-component models
- 2MASS data





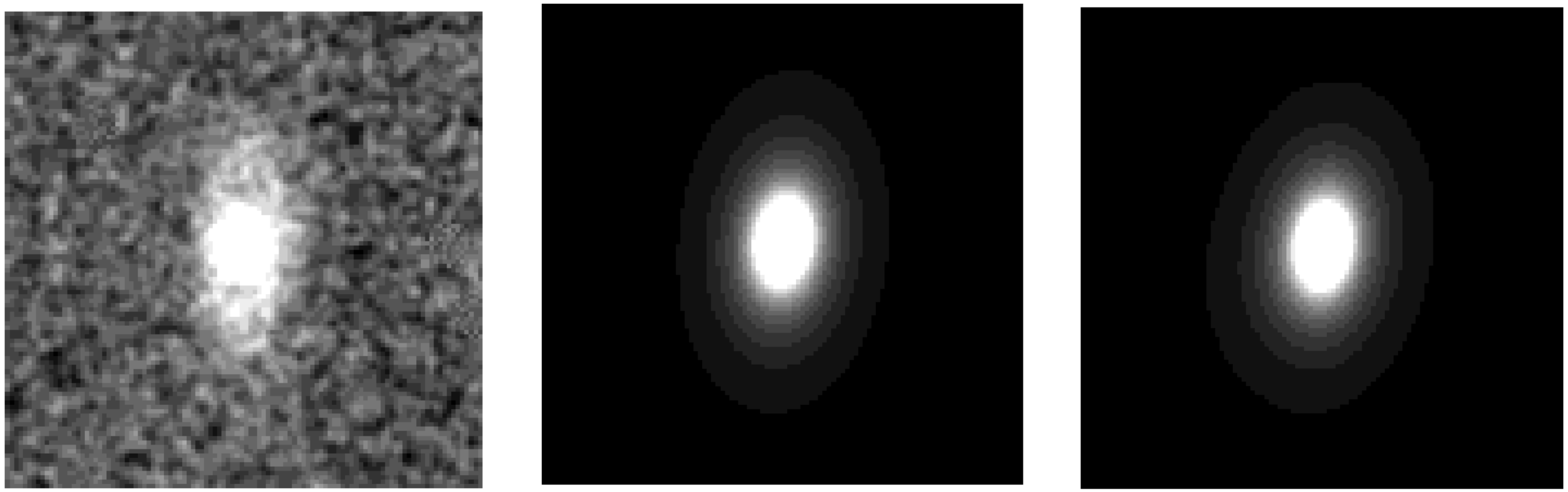




NGC 470 [SA(rs)b]

GALPHAT: results

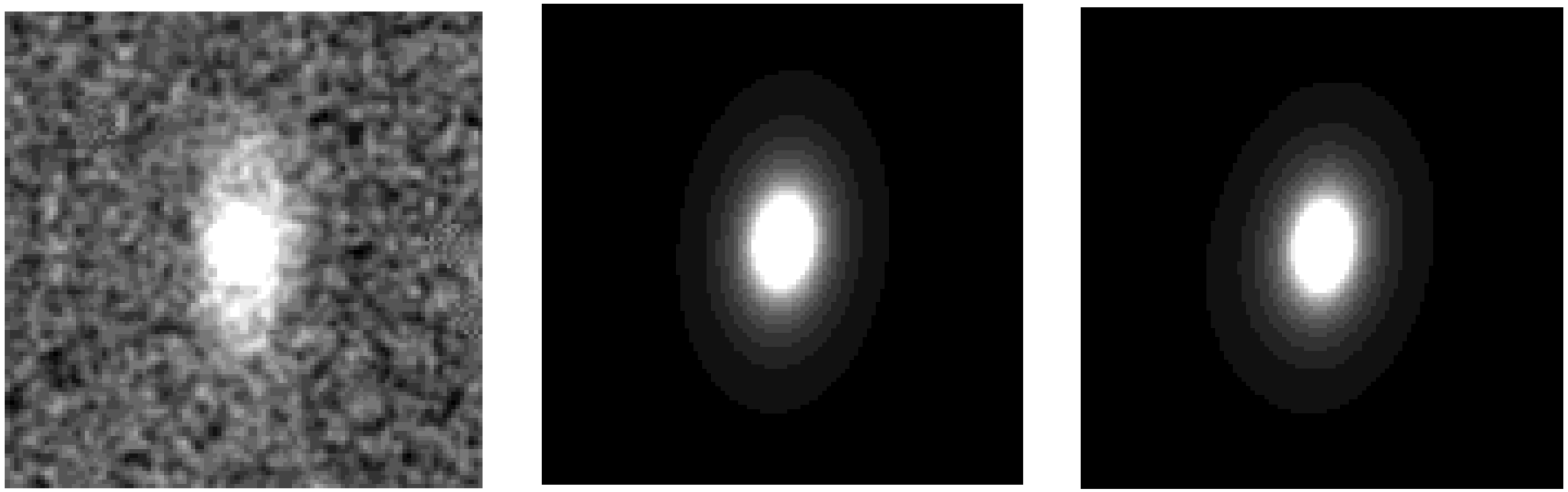
- S/N=15, Sersic index: n, 1- and 2-component models
- 2MASS data



NGC 374 [S0/a]
left: data, middle: 1-comp, right: 2-comp

GALPHAT: results

- S/N=15, Sersic index: n, 1- and 2-component models
- 2MASS data



NGC 374 [S0/a]
left: data, middle: 1-comp, right: 2-comp

$\log(B_{21}) = 0.056 \rightarrow$ minimal evidence to reject 1-comp

Current status

- Listed on Astro Stat web site
- Project web site: `www.astro.umass.edu/~weinberg/bie`
- 2008 release (Summer?) including persistence
- 2008/2009 release of GALPHAT stand-alone
- 2009 release of BIE-SAM stand-alone
- Interim releases (e.g. now) contact me!

Current status

- Listed on Astro Stat web site
- Project web site: `www.astro.umass.edu/~weinberg/bie`
- 2008 release (Summer?) including persistence
- 2008/2009 release of GALPHAT stand-alone
- 2009 release of BIE-SAM stand-alone
- Interim releases (e.g. now) contact me!

The End

Metropolis-Hastings: details [1]

- Begin with state $x = X_i$
- Select new state with transition probability: $q(x, x')$
- Accept $X_{i+1} = x'$; otherwise reject x' and set $X_{i+1} = x$.
- This is a Markov chain with transition probabilities given by:

$$P(x, x') = q(x, x')\alpha(x, x')$$

if $x \neq x'$, or

$$P(x, x') = 1 - \sum_{x \neq x'} q(x, x')\alpha(x, x')$$

otherwise.

Metropolis-Hastings: details [2]

- If we now set:

$$\alpha(x, x') = \min\left[1, \frac{\pi(x')q(x, x')}{\pi(x)q(x', x)}\right]$$

  if $\pi(x)q(x', x) > 0$ and $\alpha(x, x') = 1$ otherwise.

- It is now easy to check that $\pi(x)p(x, x') = \pi(x')p(x', x)$.

Metropolis-Hastings: details [2]

- If we now set:

$$\alpha(x, x') = \min\left[1, \frac{\pi(x')q(x, x')}{\pi(x)q(x', x)}\right]$$

if $\pi(x)q(x', x) > 0$ and $\alpha(x, x') = 1$ otherwise.

- It is now easy to check that $\pi(x)p(x, x') = \pi(x')p(x', x)$.

Theorem:
One can then show that $\pi(x)$ is the equilibrium distribution of the Markov chain if $q(x, x')$ is aperiodic and irreducible (Hastings 1970).

Tempered-states annealing: details [1]

Want: sample from a distribution, $p_0(x)$, which may have many isolated modes

Strategy:

- Define a series of n other distributions, $p_1(x), \ldots, p_n(x)$
- p_i being easier to sample from than p_{i-1}
- For each i, define a pair of base transitions, $\hat{T}_i$ and $\check{T}_i$
- Each p_i as an invariant distribution
- Satisfy the following mutual reversibility condition for all x and x':
  $p_i(x)\hat{T}_i(x, x') = \check{T}_i(x', x)p_i(x')$
- Detailed balance if $\hat{T}_i = \check{T}_i$
- Find candidate state by transitions in the sequence $\hat{T}_1 \cdots \hat{T}_n \check{T}_n \cdots \check{T}_1$

Tempered-states annealing: details [2]

- Accepted or rejected based on ratios of probabilities involving intermediate states. Choose levels to be successively broader, *higher temperature*.
- Begin in state $\hat{x}_0$, generate candidate state, $\check{x}_0$, as follows:

  Generate $\hat{x}_1$ from $\hat{x}_0$ using $\hat{T}_1$.
  Generate $\hat{x}_1$ from $\hat{x}_1$ using $\hat{T}_2$.
  $\vdots$
  Generate $\bar{x}_n$ from $\hat{x}_{n-1}$ using $\hat{T}_n$.
  Generate $\check{x}_{n-1}$ from $\bar{x}_0$ using $\check{T}_n$.
  $\vdots$
  Generate $\check{x}_1$ from $\check{x}_2$ using $\check{T}_2$.
  Generate $\check{x}_0$ from $\check{x}_1$ using $\check{T}_1$.

Tempered-states annealing: details [3]

Motivation
Goals
ProbStat
HEP
Advantages
Software features
Killer applications
Semi-analytic models
GALPHAT
Galphat results
Status
M-H #1
M-H #2
Tempered-states #1
Tempered-states #2
Tempered-states #3
HEP details

- The candidate state is then accepted with probability

$$\alpha = \min \left[1, \frac{p_1(\hat{x}_0)}{p_0(\hat{x}_0)} \cdots \frac{p_n(\hat{x}_{n-1})}{p_{n-1}(\hat{x}_{n-1})} \cdot \frac{p_{n-1}(\check{x}_{n-1})}{p_n(\check{x}_{n-1})} \cdots \frac{p_0(\check{x}_0)}{p_1(\check{x}_0)} \right]$$

- Each p_i occurs an equal number of times in the numerator and denominator of the above product of ratios $\Rightarrow$ Don't need normalization for p_i
- Need fine spacing of levels to have high acceptance probability

return

HEP details

- Standard Bayes: $P(H|D, I) = P(H|I)P(D|H, I)/P(D|I)$

HEP details

- Standard Bayes: $P(H|D,I) = P(H|I)P(D|H,I)/P(D|I)$
- Binning at level $n-1$ contains all of the observations but with less information than at level n:
  $D_{n-1} = \bigcup D_n$

HEP details

- Standard Bayes: $P(H|D,I) = P(H|I)P(D|H,I)/P(D|I)$
- Binning at level $n-1$ contains all of the observations but with less information than at level n: $D_{n-1} = \bigcup D_n$
- Hierarchical update:

$$\begin{aligned} P(H|D_0, D_1, \ldots, D_n, I) \quad &\propto \quad P(H|D_0, \ldots, D_{n-1}, I) \times \\ & \quad [P(D_n|H,I)/P(D_{n-1}|H,I)] \end{aligned}$$

HEP details

- Standard Bayes: $P(H|D,I) = P(H|I)P(D|H,I)/P(D|I)$
- Binning at level $n-1$ contains all of the observations but with less information than at level n:
  $D_{n-1} = \bigcup D_n$
- Hierarchical update:

$$\begin{aligned} P(H|D_0, D_1, \ldots, D_n, I) \quad &\propto \quad P(H|D_0, \ldots, D_{n-1}, I) \times \\ &[P(D_n|H,I)/P(D_{n-1}|H,I)] \end{aligned}$$

  - Prior is posterior at previous level

HEP details

- Standard Bayes: $P(H|D, I) = P(H|I)P(D|H, I)/P(D|I)$
- Binning at level $n-1$ contains all of the observations but with less information than at level n:
  $D_{n-1} = \bigcup D_n$
- Hierarchical update:

$$\begin{aligned} P(H|D_0, D_1, \ldots, D_n, I) &\propto P(H|D_0, \ldots, D_{n-1}, I) \times \\ &\quad [P(D_n|H, I)/P(D_{n-1}|H, I)] \end{aligned}$$

  - Prior is posterior at previous level
  - Likelihood becomes the Likelihood *ratio*

return

HEP details

- Standard Bayes: $P(H|D,I) = P(H|I)P(D|H,I)/P(D|I)$
- Binning at level $n-1$ contains all of the observations but with less information than at level n:
  $D_{n-1} = \bigcup D_n$
- Hierarchical update:

$$P(H|D_0, D_1, \ldots, D_n, I) \propto P(H|D_0, \ldots, D_{n-1}, I) \times [P(D_n|H, I)/P(D_{n-1}|H, I)]$$

  - Prior is posterior at previous level
  - Likelihood becomes the Likelihood *ratio*

return

HEP details

- Standard Bayes: $P(H|D,I) = P(H|I)P(D|H,I)/P(D|I)$
- Binning at level $n-1$ contains all of the observations but with less information than at level n:
  $D_{n-1} = \bigcup D_n$
- Hierarchical update:

$$P(H|D_0, D_1, \ldots, D_n, I) \propto P(H|D_0, \ldots, D_{n-1}, I) \times [P(D_n|H,I)/P(D_{n-1}|H,I)]$$

  - Prior is posterior at previous level
  - Likelihood becomes the Likelihood *ratio*

return