CUDAHM: GPU-Accelerated Hierarchical Bayesian Inference with Application to Modeling Cosmic Populations

János M. Szalai-Gindl*
Department of Physics of Complex Systems,
Eötvös Loránd University
and
Department of Applied Mathematics & Statistics,
The Johns Hopkins University
and
Tamás Budavári
Department of Applied Mathematics & Statistics,
The Johns Hopkins University
and
Brandon C. Kelly
Department of Physics,
Broida Hall, University of California
and
Thomas J. Loredo
Cornell Center for Astrophysics & Planetary Science,
Cornell University

April 2, 2017

Abstract

We describe CUDAHM, a C++ framework for hierarchical Bayesian inference with simple (typically single-plate) parametric graphical models that uses graphics processing units (GPUs) to accelerate computations, enabling deployment on very large

This work was supported by the Hungarian Scientific Research Fund via grant OTKA NN 114560. Budavári, Kelly, and Loredo gratefully acknowledge the NSF-funded Statistical and Applied Mathematical Sciences Institute (SAMSI) for support for visits to SAMSI, where this project originated. Loredo's effort was additionally supported by NSF grant AST-1312903.

datasets. CUDAHM exploits conditional independence between instances of a plate, which enables massively parallel exploration of the replication parameter space using the single instruction, multiple data (SIMD) architecture of GPUs. It provides support for constructing Metropolis-within-Gibbs samplers that iterate between GPU-acclerated robust adaptive Metropolis (RAM) sampling of plate-level parameters conditional on upper-level parameter values, and Metropolis-Hastings sampling of upper-level parameters on the host central processing unit (CPU) conditional on the GPU results. The GPU computations are implemented using the Compute Unified Device Architecture (CUDA). CUDAHM is motivated by measurement error problems in astronomy, where density estimation and linear and nonlinear regression problems must be addressed for populations of thousands to millions of objects whose features are measured with possibly complex uncertainties. We demonstrate accurate GPU-accelerated parametric conditional density deconvolution for simulated populations of 10^6 objects in about two hours using a single NVIDIA Tesla K40c GPU.

Keywords: Hierarchical Bayesian models, Metropolis-within-Gibbs sampling, parallel computing, astrostatistics, graphical processing units (GPUs)

1 Introduction

Bayesian inference with graphical models has rapidly grown in popularity and sophistication since the emergence of Markov chain Monte Carlo (MCMC) algorithms for Bayesian computation nearly three decades ago. The work we report here focuses on models with classic, simple graphical structures—directed acyclic graphs (DAGs) that typically have a single plate, i.e., a single level of replication of random variables at the lower level of a hierarchical model. Our work aims to extend the range of application of Bayesian graphical modeling in the direction of increased dataset size, rather than in the direction of increased graphical complexity.

We are motivated by measurement error problems in astronomy: density estimation with measurement error (density deconvolution, or demixing, often of a *conditional* density), and linear and nonlinear regression with measurement errors in both predictors and response. Hierarchical Bayesian modeling is well-suited to such problems, but is relatively new in astronomy (see Kelly 2012 and Loredo 2013 for recent surveys). Although some recent astrostatistical research develops models with rich graphical structure, most astronomers are unfamiliar with hierarchical modeling, and models with simple graphical structure can provide new capability in many areas of astronomy, including basic two-level hierarchical models.[1] But dataset size can be an obstacle to use of such models. Large-scale, automated surveys are providing astronomers with increasingly large datasets for demographic studies of cosmic populations (e.g., categories of stars, galaxies, and planets). Current and emerging surveys are providing measurements for populations with sizes ranging from tens of thousands to 10^8 or even larger. For datasets of these scales, exploration or integration over the latent variables specifying imprecisely measured characteristics of objects in a population can be prohibitive, even for simple models with univariate member characteristics. Yet as population size grows, it becomes increasingly important to account for uncertainty in such latent variables. For example, it is well known that regression and density estimators that ignore measurement error are typically inconsistent, with the ratio

[1] We follow the convention of naming hierarchical DAGs by the number of levels with uncertain nodes, e.g., the number of open nodes in Fig. 2.1, which depicts a model with three levels of random variables, but with the data variables (bottom level) known, i.e., to be conditioned on.

of bias to reported precision growing with sample size (Carroll et al. 2006). Single-plate hierarchical models can account for measurement error in many astronomically interesting scenarios, provided the implementation enables efficient computation for relevant dataset sizes.

In the following section (Sec. 2), we describe the design of the CUDAHM framework, which is motivated by a common computational structure underlying example hierarchical models arising in measurement error problems in astronomy. In Sec. 3, we describe a common astronomical data analysis problem: inferring the luminosity distribution of a class of objects from distance and flux observations of a sample subject to selection effects and measurement error. Such problems may be modeled using latent, thinned marked point processes observed with measurement error; we show that the resulting likelihood function can be cast in a form mirroring the computational structure of single-plate graphical models, enabling implementation with CUDAHM. We present tests of such an implementation, using simulated data, in Sec. **??**. Sec. 5 provides a summary and plans for future work. In the supplementary material, Sec. **??** presents an approximation technique for the computation of an integral appearing in the luminosity function example.

2 CUDAHM motivation and design

2.1 Motivating problem structures

Suppose we observe N members of a large population, with the observed members indexed by $i = 1$ to N. Each object (member) has a property or properties ψ_i (a vector in multiple-property cases); we are interested in estimating the collection of properties, $\{\psi_i\}$, or their distribution, but cannot measure every component of ψ_i with high precision. Instead, for each object, we have observed data, D_i, that provide information about ψ_i. In the following, we sometimes use bold symbols to refer to quantities collectively, e.g., $\boldsymbol{\psi} \equiv \{\psi_i\}$, and $\boldsymbol{D} \equiv \{D_i\}$.

We consider problems where the nature of the observations motivates models that specify a joint sampling distribution for the data, conditional on the member properties, that

factors into a product of conditionally independent *member sampling distributions*,[2]

$$p(\boldsymbol{D} \,|\, \boldsymbol{\psi}) = \prod_{i=1}^{N} p(D_i \,|\, \psi_i). \tag{2.1}$$

We model the member properties, $\boldsymbol{\psi}$, as IID draws from a *population probability density function* (PPDF), $f(\psi_i; \theta)$, with uncertain parameters, θ. Goals of inference may include estimation of the PPDF (i.e., estimation of θ), or estimation of the member properties, $\boldsymbol{\psi}$.

Fig. 2.1 shows the DAG for this type of model, both explicitly and using plate notation. Following standard conventions, open nodes indicate uncertain random variables that are targets of inference, and shaded nodes indicate observed quantities, i.e., random variables that are uncertain a priori, but that become known after observation, and thus may be conditioned on. If we denote the prior PDF for the population distribution parameters by $\pi(\theta)$, this DAG indicates that the joint PDF for all random quantities in this model may be written,

$$p(\theta, \{\psi_i\}, \{D_i\}) = \pi(\theta) \prod_{i=1}^{N} f(\psi_i; \theta)\, p(D_i \,|\, \psi_i) \tag{2.2}$$

$$\propto \pi(\theta) \prod_{i=1}^{N} f(\psi_i; \theta)\, \ell_i(\psi_i), \tag{2.3}$$

where we have defined the *member likelihood functions*,

$$\ell_i(\psi_i) \propto p(D_i|\psi_i).$$

Note that, as likelihood functions (vs. sampling distributions), these functions need only be specified up to proportionality. In particular, any dependence on D_i that does not influence the dependence on ψ_i can be ignored.

The DAG describes a generative model for all of the random variables, including the data. However, when the task is inference of parameters conditional on observed data (versus prediction of unobserved data), specifying member likelihood functions (rather than sampling distributions) can be a significant simplification. In many astronomical applications, estimation of ψ_i from D_i is often a nontrivial inference problem in itself. For

[2]For the sake of simplicity, we denote random variables and their values with the same symbol. Also, $p(\bullet)$ will be used to denote a probability mass function or a probability density function, depending on the type of the argument.

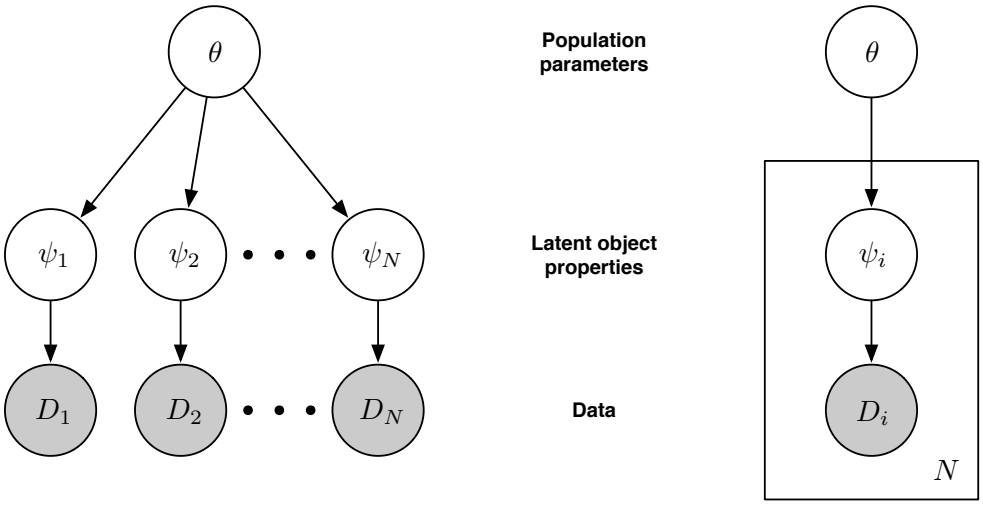


Figure 2.1: Directed acyclic graph (DAG) for a 2-level hierarchical Bayesian model. *Left*: DAG explicitly showing replicated conditionally independent subgraphs. *Right*: DAG depicting replicated elements with a plate.

example, when ψ_i denotes the apparent brightness of a star and D_i denotes image data, inference may involve fitting the complicated point spread function of an imaging instrument to Poisson distributed photon counts in dozens or hundreds of pixels (often marginalizing over an uncertain background or instrument calibration component). The resulting likelihood function for ψ_i (or marginal likelihood function, when there are nuisance parameters) will often be relatively easy to summarize as a function of ψ_i; e.g., it may be well approximated by a Gaussian or multivariate Gaussian function (perhaps after a transformation). On the other hand, the sampling distribution for the data may be quite complicated and high-dimensional. In many circumstances, it may not even be well-defined. Weather or spacecraft conditions may affect the precision, accuracy, and even the quantity of data for a member observation; the repeated sampling distribution may be hard or even impossible to define objectively, while the likelihood function may be well-defined. Most astronomical surveys produce member estimates with heteroscedastic uncertainties, in the sense of producing member likelihood functions with widths that vary from object to object. It may be difficult or impossible to accurately describe the repeated sampling properties of the heteroscedastic uncertainties. But for inference based on *given* observations, only the actually

available member likelihood functions matter. Implementing inference in a manner that requires specifying only the member likelihood functions, rather than the sampling distributions, is a better fit to the nature of astronomical survey catalog data summaries than an implementation requiring unique specification of the lowest level sampling distributions.

A widely-used approach for posterior sampling in the context of two-level hierarchical models is the *Metropolis-within-Gibbs* (MWG) algorithm, where the θ population parameters and the ψ_i member properties are sampled in separate, alternating steps. First, the member properties are sampled by holding the population parameters fixed, then the population parameters are sampled by holding the member properties fixed. These steps may each be implemented with Metropolis or Metropolis-Hastings algorithms; their sequential combination amounts to Gibbs sampling on the joint space. Explicitly, the steps are:

$$\psi_i \sim p(\psi_i \,|\, \theta, D_i), \quad \forall i \in 1:N; \tag{2.4}$$

$$\theta \sim p(\theta \,|\, \boldsymbol{\psi}, M). \tag{2.5}$$

The departure point for CUDAHM is recognition that, since the ψ_i properties are conditionally independent in (2.4), they may be sampled in parallel, making this part of the MWG algorithm suitable for a massively parallel implementation using GPUs. We describe such an implementation further below.

Since ψ_i may be a vector, the ψ_i node in the DAG may admit a factorization leading to further structure within the plate in Fig. 2.1. Fig. 2.2 shows DAGs for several other single-plate modeling scenarios for which inference may be implemented using MWG with massively parallel sampling of member properties.

The DAG in the left panel depicts a frequently arising structure in astronomy, where the object properties ψ_i consist of intrinsic *characteristics* χ_i that, if known, can predict *observables*, $\mathcal{O}_i$, i.e., quantities that can predict observed data.[3] An important example is inference of *number-size distributions* (also known as number counts or $\log N$–$\log S$ distributions). Here the object characteristics are distance, r_i, and luminosity, L_i (amount of

[3]In astronomical parlance, measurements of the observables are used to "characterize" the object, whence our choice of the term "characteristics" here. We reserve the term for properties intrinsic to the source, i.e., not depending on observer-referred quantities such as distance. E.g., luminosity is a characteristic; flux (brightness at the telescope) is an observable.

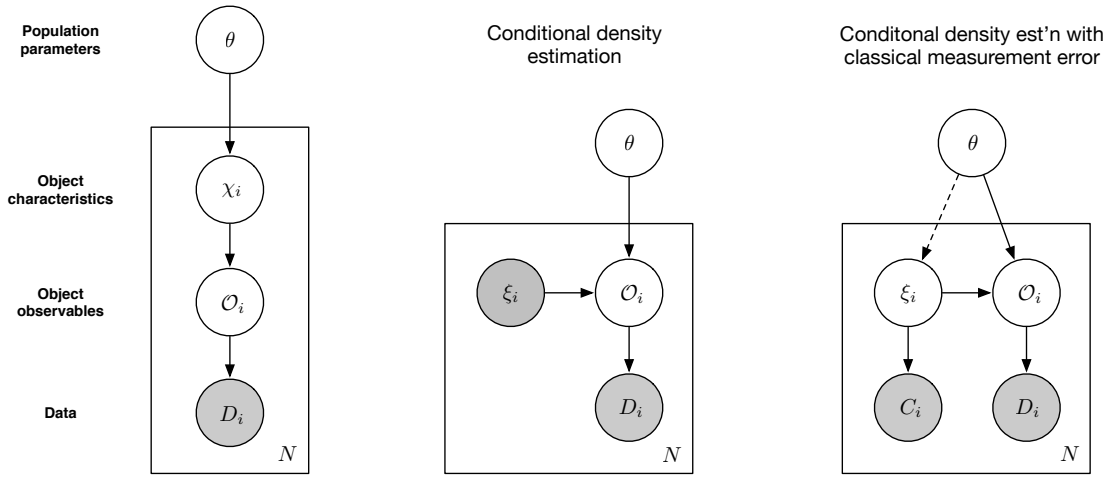


Figure 2.2: Example single-plate DAGs that may be implemented in CUDAHM. *Left*: DAG for a 3-level hierarchical model corresponding to demographic inference for objects with latent characteristics χ_i, related to latent observables $\mathcal{O}_i$. *Center*: DAG for conditional density estimation, expressed via a latent observable $\mathcal{O}_i$, and a precisely measured predictor (covariate), ξ_i. *Right*: DAG for conditional density estimation with classical measurement error, with a latent predictor, ξ_i, measured indirectly via data C_i. The predictor may have an a priori known prior distribution, or it may be parameterized (with parameters included in θ, in which case the dashed edge would be present).

energy emitted per unit time). The observable is flux (rate of energy flow per unit area normal to the line of sight, per unit time, at the telescope), F_i, related to the characteristics via the inverse-square law, $F_i = L_i/(4\pi r_i^2)$ (or its cosmological generalization).

The DAG in the middle panel depicts conditional density estimation, where the properties ψ_i are comprised of precisely measurable predictors (covariates), ξ_i, that, together with the population parameters θ, specify the PDF for observables, $\mathcal{O}_i$; the data provide likelihood functions for the $\mathcal{O}_i$. An important example is inference of a *luminosity function*, which describes the population distribution for the luminosities of a class of sources (say, a stellar or galaxy type). If the PDF for luminosity is $f(L;\theta)$, and the distances to objects may be precisely measured (say, via spectroscopic redshift data), then by a simple change of variables the PDF for the flux observable for a source at distance d is $4\pi d^2 f(4\pi d^2 F;\theta)$ (in Euclidean space). This would be the distribution for the $\mathcal{O}_i$ node in the middle DAG. We treat a more complicated version of this problem below, where the object sample is subject to flux-dependent selection effects.

As a final example, the DAG in the right panel depicts conditional density estimation with measurement error (i.e., uncertainty in the predictors), with a classical measurement error structure (data distributions conditional on latent predictor values). A wide variety of astronomical data analysis problems have this structure. Kelly et al. (2012) describes a noteworthy example studying how the spectrum of infrared emission from heated interstellar dust depends on properties of the dust grains; this is one of the specific problems motivating CUDAHM. Earlier studies, based on maximum likelihood estimates of dust properties (ignoring measurement error), found a surprising negative correlation between dust temperature and a spectral index parameter indicating how the dust properties tilt the infrared spectrum away from a black body spectrum. Accounting for measurement error *reversed the sign* of the inferred correlation, reconciling it with some theoretical models. Kelly et al. (2012) analyzed measurements from $\sim 10^4$ dust regions; CUDAHM dramatically accelerates the calculations and makes such studies feasible with 10 to 100 times larger samples.

2.2 CUDAHM architecture

To sample member propertis in the MWG algorithm, as specified in Eq. 2.4, we use the robust adaptive Metropolis (RAM) algorithm devised by Vihola (2012). It works by adaptively refining a Metropolis algorithm proposal distribution during the sampling process until a target mean acceptance rate α_* is reached. CUDAHM currently uses a multivariate normal distribution as the proposal q, and sets the target mean acceptance probability to a default value of $\alpha_* = 0.4$. Adaptation involves using new samples to adjust the proposal covariance matrix in a manner that decays with time along the Markov chain so as to guarantee correct asymptotic sampling. Specifically, adjustments enter with a decaying weight, $\eta_n = n^{-2/3}$, where n is the iteration number along the Markov chain.

Following Vihola (2012), let S_1 be the identity matrix and X_1 some point in the space to be sampled for which the target density $\pi(X_1) > 0$. Each RAM iteration cycles through the following steps:

1. Compute $Y_n = X_{n-1} + S_{n-1}U_n$, where $U_n \sim q$ is an independent random vector.

2. With probability $\alpha_n \equiv \min\{1, \pi(Y_n)/\pi(X_{n-1})\}$ the step is accepted, and $X_n = Y_n$; otherwise the step is rejected and $X_n = X_{n-1}$.

3. Compute the lower-diagonal matrix S_n with positive diagonal elements satisfying the equation
$$S_n S_n^T = S_{n-1}\left(I + \eta_n(\alpha_n - \alpha_*)\frac{U_n U_n^T}{\| U_n \|^2}\right) S_{n-1}^T \tag{2.6}$$
where I is an identity matrix. The solution for S_n is unique as it is the Cholesky factor of the right hand side.

CUDAHM implements these steps on the GPU, via "kernel" code (code executed on a GPU) written in the CUDA language and taking advantage of optimized CUDA library functions (e.g., for random number generation and linear algebra). For the population sampling step in the MWG algorithm, as specified in Eq. 2.5, parameter updates are computed on the host CPU, using code written in standard C++, but with the log-posterior calculation (using the updated population parameters) executed on the GPU.

•••*[New text from Janos below; this needs some massaging; we also need to provide a link to the main CUDAHM repo here:]*•••

CUDAHM enables one to easily and rapidly construct an MCMC sampler for a simple hierarchical model, requiring the user to supply only a minimimal amount of CUDA code. There are four main classes in CUDAHM that users may need to use directly:

- `DataAugmentation`: This class controls the calculations involving the member properties.
- `PopulationPar`: This class controls the calculations involving the population parameters.
- `GibbsSampler`: This class runs the MWG sampler.
- `GibbsSamplerWithCompactMemoryUsage`: This class also runs a MWG sampler, but with more efficient use of memory. It opens an output file stream (which has a buffer) and writes samples out on the fly.

The simplest use case involves only instantiating the `GibbsSampler` class, since this will internally construct `DataAugmentation` and `PopulationPar` objects, using reasonable default algorithms. However, if one wants to subclass the `DataAugmentation` or `PopulationPar` classes, to customize algorithms to the problem, or to generalize the framework, then pointers to the instances of these classes must be provided to the `GibbsSampler` constructor. In general this is only needed if one wants to override the default methods for setting the initial values, or if one wants to override the default prior on the population parameters (which is an uninformative uniform distribution).

We note the CUDAHM default methods assume *all* member properties are uncertain; kernel functions that are used for updating member properties and the population parameters on the GPU assume all member properties may change in each MWG cycle. This is important to consider for problems where the object properties contain precisely measurable predictors (e.g., covariates for conditional density estimation in the manner of the middle DAG in Fig. 2.2), which should be held fixed over the course of posterior simulation. When this is the case, the user should override the default methods. The `lum_func` implementation, which is related to the luminosity function example case, provides concrete guidance on this issue. We are exploring internal architectural changes to simplify the user API for handling such cases.

There are two functions that the user must provide: a function that computes the logarithm of the probability density of the measurements given the member properties for each object in the sample, and a function that computes the logarithm of the probability density of the member properties given the parent population parameters. These functions execute on the GPU and must be written in CUDA. The file `cudahm_blueprint.cu` under the `cudahm` directory contains a blueprint with documentation that the user may use when constructing their own MCMC sampler.

3 Luminosity function estimation: Framework

The CUDAHM distribution contains some example code implementing basic hierarchical models, such as the normal-normal model, and a realistically complicated astrophysical example handling regression with classical measurement error, with nonlinear models (the interstellar dust problem described above).

As a somewhat more complicated application of CUDAHM, we here consider luminosity function estimation, a parametric conditional density estimation problem arising across many areas of astronomy. We highlight this example, both because of its ubiquity across astronomy, and to illustrate the generality of CUDAHM. We describe luminosity function estimation for a model including not only measurement error, but also *selection effects*. The selection effects make the joint distribution structure more complicated than the conditional density estimation DAG shown in Fig. 2.2; in particular, there are two plates (corresponding to detected and undetected objects). However, the likelihood function can be manipulated to have a conditional dependence structure similar to that of a single-plate DAG, allowing straightforward implementation of the model using CUDAHM.

3.1 Astronomical background

The fundamental observables for localized astronomical sources include position (both direction on the celestial sphere, and distance, r, in some chosen coordinate system), and apparent brightness, quantified in terms of flux, F (energy or photon number per unit time per unit area normal to the line of sight). Astronomers use these observables to study

demographic properties of many classes of sources, including stars and galaxies of various types, minor planets (such as asteroids), and explosive transients (gamma-ray bursts, supernovae). For concreteness, here we focus on observations of nearby galaxies, for which distance may be measured by using spectroscopy to find the *redshift*, z, of spectral lines (i.e., their fractional shift in wavelength from laboratory values). Due to the cosmological expansion, for relatively nearby galaxies outside the local group, distance is proportional to redshift to a good approximation, with

$$r = \frac{cz}{H_0}, \tag{3.1}$$

where c denotes the speed of light, and H_0 is Hubble's constant, measuring the current expansion rate of the universe, with $H_0 \approx 70$ km s^{-1} Mpc^{-1} (with Mpc denoting megaparsecs). H_0 is measured with a precision of several percent, and spectroscopic redshifts for nearby galaxies can be measured to sub-percent precision. For simplicity, here we consider distances to be precisely measured, via spectroscopic redshifts. Often, astronomers use redshift directly as a proxy for distance.

A fundamental intrinsic characteristic of a source is its *luminosity* (emitted power), L, a measure of its intrinsic (vs. apparent) brightness (with units of energy per unit time). For nearby galaxies (i.e., for distances where space is very nearly Euclidean), the inverse-square law relates L to the observables F and r:

$$F = \frac{L}{4\pi r^2}. \tag{3.2}$$

The *luminosity function*, $\phi(L, r)$, describes the distribution of luminosities for a population at a specified distance (or redshift). It is typically defined as the intensity function for a point process, i.e., as specifying the expected number of galaxies per unit volume at distance r, per unit luminosity interval. If we denote the spatial number density of galaxies at distance r by $n(r)$, then $n(r) = \int dL\, \phi(L, r)$. The *luminosity PDF* for galaxies at distance r is then

$$f(L, r) = \frac{\phi(L, r)}{n(r)}. \tag{3.3}$$

Note that $\phi(L, r)$ and $f(L, r)$ specify *conditional* distributions, i.e., distributions for L at a given r.[4] Using (3.2), the *flux PDF* for galaxies at distance r, denoted $\rho(F, r)$, can be

[4]Authors vary on the definition of the luminosity function, many defining it as done here, and others

found by a change of variables from L to F, giving

$$\rho(F, r) = 4\pi r^2 f(4\pi r^2 F, r). \tag{3.4}$$

The galaxy luminosity function carries valuable information about the formation and evolution of galaxies, therefore it is an important target of inquiry in astronomy (see Binggeli et al. 1988 and Johnston 2011 for reviews). Johnston 2011 provides a review of methods developed by astronomers for estimation of galaxy luminosity functions. CUDAHM can address parametric luminosity function inference; this section provides an example as a demonstration of CUDAHM's capability and flexibility. In the next subsection, we describe a useful parametric family of luminosity functions. Then we present a hierarchical Bayesian framework for modeling astronomical survey data using thinned latent marked point process models, developing it specifically for luminosity function inference. The next section presents a simulation study demonstrating CUDAHM luminosity function inference for simulated populations with sizes up to 10^6.

3.2 Parametric luminosity function models

For most cosmic populations, including galaxies, the luminosity function falls very steeply with increasing luminosity. The canonical starting point for parametric modeling of luminosity distributions is the *Schecter function*,

$$\phi(L; \theta) = \frac{A}{L_*} \left(\frac{L}{L_*} \right)^{\beta} e^{-L/L_*}, \tag{3.5}$$

where the parameters $\theta = (L_*, \beta, A)$ comprise a luminosity scale, L_*, a power law index, β, and an amplitude, A.[5] The form of the Schecter function would seem to imply a luminosity distribution that is a gamma distribution (with shape parameter $\alpha = \beta - 1$). However, the observed samples of many populations follow Eq. 3.5 with β in the interval $(-2, -1)$, in which case the integral of $\phi(L; \theta)$ over L is infinite, and the luminosity distribution is formally improper (with α outside of the allowed range for the gamma distribution).

using "luminosity function" to denote what we here call the luminosity PDF.

[5]There are varying conventions for parameterizing the amplitude of the Schecter function. In this parameterization, A has units of space density. In similar parameterizations, A is often denoted ϕ_*, although it neither has the units of ϕ, nor is it equal to $\phi(L_*)$, as the symbol might misleadingly suggest.

Low-luminosity sources are unobservable (due to noise and background, discussed below), so in practice the *observable* luminosity function is truncated at low luminosities, and the impropriety is often ignored. But the actual luminosity function must rise less quickly with decreasing L (corresponding to β becoming larger than -1) or be cut off at low luminosities (corresponding to there being a minimum galaxy size).

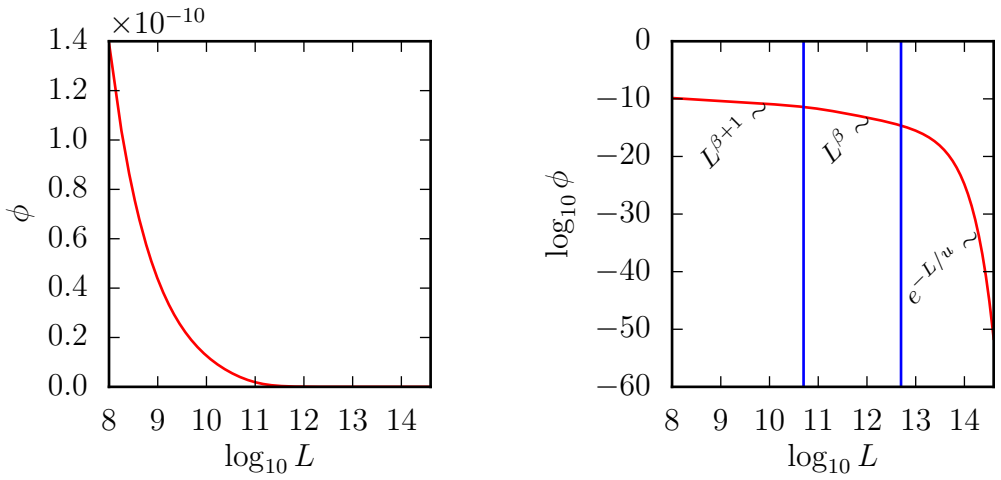
For some populations, an increase in the power law index (i.e., flattening of the logarithmic slope) is in fact observed at low luminosities. This has been observed for quasars (galaxies with a large, actively accreting central black hole; see McGreer et al. 2013). Similarly, the stellar initial mass function (related to the stellar luminosity function, and fit with similar models) has a low-mass (low-luminosity) index that flattens by ≈ 1 (Kroupa 2007). Motivated by such observations, and to keep the luminosity distribution proper, we here adopt a "break-by-one" (BB1) generalization of the Schecter function, with $\phi \propto L^{\beta+1}$ at low luminosities, and thus integrable for $\beta > -2$. Specifically, the BB1 model has a luminosity distribution with three parameters: a mid-luminosity power law index, β, and two parameters defining the mid-luminosity range, (l, u), with $l < u$ and u playing the role of L_* in the Schecter function, and the power law index smoothly breaking to $\beta + 1$ as L decreases below l. The BB1 luminosity PDF has following functional form:

$$f(L;\theta) = \frac{C(\beta,u,l)}{u}\left(1-e^{-L/l}\right)\left(\frac{L}{u}\right)^{\beta} e^{-L/u}, \tag{3.6}$$

where the normalization constant $C(\theta)$ is

$$C(\beta,u,l) = \begin{cases} \dfrac{1}{\Gamma(\beta+1)\cdot\left(1-\frac{1}{\left(1+\frac{u}{l}\right)^{\beta+1}}\right)} & \text{if } \beta > -2 \text{ and } \beta \neq -1; \\ \dfrac{1}{\log\left(1+\frac{u}{l}\right)} & \text{if } \beta = -1. \end{cases} \tag{3.7}$$

Note that as $l \to 0$, the BB1 distribution becomes a gamma distribution (if $\beta > -1$). We designed the BB1 distribution to have smooth power law break behavior at low L, yet also have an analytical normalization constant; it is proper for $\beta > -2$. It can also be sampled from using a straightforward modification of a widely-used algorithm for sampling from the gamma distribution (Ahrens & Dieter 1974). These properties make it useful for simulation experiments.



(a) (b)

Figure 3.1: Example BB1 truncated broken power law PDF, on a log-linear scale (a) and on log-log scale (b), with parameters $(u, l) = (5 \times 10^{10}, 5 \times 10^{12})$ in dimensionless units, and $\beta = -1.5$. In Panel b the blue verticals show the lower limit l and upper limit u of the region where the power law slope is $\approx \beta$.

We define a BB1 luminosity function by multiplying the BB1 luminosity distribution by the galaxy spatial number density, which is simply a constant, n, for a homogeneous population. Fig. 3.1 shows an example BB1 luminosity function, with $\beta = 1.5$, and $(l, u) = (5 \times 10^{10}, 5 \times 10^{12})$ in dimensionless units; it is plotted both with log-linear axes, and with log-log axes, where the varying power law behavior is evident. The local power law index corresponds to the slope, $G(L)$, in log-log space, defined by

$$G(L) \equiv \frac{\mathrm{d}\log f}{\mathrm{d}\log L} = \frac{L}{f}\frac{\mathrm{d}f}{\mathrm{d}L} = g(L) + \beta - \frac{L}{u}, \tag{3.8}$$

with

$$g(L) = \frac{L}{l} \cdot \frac{1}{e^{L/l} - 1}. \tag{3.9}$$

Evidently, $g(L) \to 0$ for $L \gg l$ and $g(L) \to 1$ for $L \ll l$. Thus the logarithmic slope, $G(L)$, corresponds to an exponential cutoff at large L, and at small L, a slope of $\beta + 1$. When $u \gg l$, so there is a range where $L \gg l$ but $L \ll u$, the logarithmic slope is $\approx \beta$ in that range.

Finally, the BB1 cumulative distribution function is

$$F(L;\theta) = C(\theta)\left[\Gamma(\beta+1) - \gamma\left(\beta+1, L/u\right) - \frac{\Gamma(\beta+1) - \gamma\left(\beta+1, L\cdot\left(\frac{1}{u}+\frac{1}{l}\right)\right)}{\left(1+\frac{u}{l}\right)^{\beta+1}}\right], \quad (3.10)$$

where $\Gamma(\cdot)$ and $\gamma(\cdot,\cdot)$ denote the gamma function and the upper incomplete gamma function, respectively.

3.3 Modeling survey selection effects and measurement error

Astronomers estimate luminosity functions and other astronomical distributions using data compiled in *survey catalogs*: tables of estimates (including uncertainties) of object properties, accompanied by a description of selection criteria for the survey that produced the catalog. Catalogs are derived data products; they summarize information in more complex and voluminous raw datasets, such as large collections of images or time series. The nature of astronomical catalog data makes their analysis differ in some respects from analyses of survey data familiar in the statistical survey sampling literature.

Flux measurements are affected by measurement error that is often dominated by Poisson fluctuations in the photon counting rate, including fluctuations from astrophysical and instrumental backgrounds. The measurement error thus approximately scales with the square root of the flux, and is fractionally greater at low flux than at high flux. At low fluxes, photons from the backgrounds can produce false source detections. To prevent this, surveys adopt detection criteria to strongly mitigate against false detections. A simple, representative criterion is to accept sources only if the estimated flux is ν times the flux uncertainty, with $\nu \approx 5$ so that the probability for false detection is low even for large catalogs (i.e., there is strong control of the family-wise error rate).

Detection criteria introduce *selection effects* into catalogs. Most obviously, faint sources (low luminosity sources, or distant high luminosity sources) are excluded; the observable luminosity function is a thinned version of the actual luminosity function. In addition, *measurement error* more subtly but significantly distorts the shape of the observable distribution, a phenomenon well known in the density deconvolution literature, and also recognized in the astronomical literature, where it is sometimes called *Eddington bias*, in reference to early discussions of the distortion by Eddington and Jeffreys (REFS). They noted that an

object with a measured flux of $\hat{F}$ is more likely to be an object with a true flux $F < \hat{F}$ than one with $F > \hat{F}$, even when measurement errors have a symmetric distribution, because dim sources are more numerous than bright sources in most astronomical settings. Selection effects can exacerbate the distortion in the vicinity of a flux threshold, with measurement error and the falling flux distribution conspiring to scatter more below-threshold sources into the observed sample than above-threshold sources out of it, a component of what astronomers call *Malmquist bias* (Binney & Merrifield 1998) (although the term is used inconsistently). Hierarchical modeling can automatically account for such thinning and distortion, in a manner that adapts to the shape of the luminosity function; this is a major motivation for its increasing popularity in astrostatistics.

We have developed a framework for modeling astronomical survey data using *thinned latent marked point process models*. Measurement error is handled in a hierarchical Bayesian manner, by introducing latent member property parameters, with catalog estimates understood as describing member likelihood functions for the properties. We model the population distribution of the latent member properties using marked point processes. We model selection effects through thinning of the latent point process. This framework was originally developed for studying the luminosity function of gamma-ray bursts, powerful explosive transients thought to mark the birth of stellar-mass black holes (Loredo & Wasserman 1995, 1998). It was subsequently adapted to study the luminosity distribution (and through it, the size distribution) of trans-Neptunian objects, asteroid-like minor planets in the outer solar system (Loredo 2004, Petit et al. 2008). We outline the framework here as it applies to luminosity function estimation and similar problems, in contexts where a marked Poisson point process is an appropriate model for the latent member properties. Kelly et al. (2008) independently developed a similar approach for settings where a binomial point process may be appropriate, and applied it to estimating the number density of quasars as a function of redshift.

A somewhat subtle aspect of the problem is the tie between measurement error and selection effects. In astronomical surveys, the raw data are searched to find candidate objects. For candidates that pass detection criteria, the data are used to estimate source properties. The same underlying data are used both for selection (detection) and measure-

ment; these tasks are not independent, as is assumed in many statistical survey methods outside of astronomy. Ignoring the dependence can corrupt inferences.

Object detection is typically implemented via a scanning procedure. For example, for image data, a fixed aperture may be scanned over the image; a detection algorithm determines if an object is present, e.g., by comparing the estimated flux in the aperture to a threshold value (set by background and noise estimates), or by fitting an image model to the data in the aperture and comparing the fitted amplitude to a threshold. For time series data, a window may be scanned over the time series, with an object detected if the estimated flux is above a threshold. If an object is detected, its properties are estimated, e.g., by a likelihood or weighted least squares calculation, with estimation results summarized in the catalog.

Fig. 3.2 illustrates the framework and its relationship to catalog construction. We split the object property parameter space into scan and mark components. The scan component corresponds to the dimensions over which the detection scan operates; the mark component corresponds to the remaining dimensions. For our galaxy luminosity function example, the scan component is the two-dimensional position of the galaxy image on the detector (corresponding to its direction on the sky), and the mark component is the galaxy luminosity and distance, or equivalently, flux and distance (in a more complex case, the mark component might include color and morphological parameters). In the figure, the dots (red) indicate the true properties of seven galaxies; the blue contours depict likelihood functions for the properties, based on noisy image data. The gray region at the bottom is bounded above by the position-dependent detection threshold; an object is detected only if its best-fit (maximum likelihood) flux is above the threshold. Here two of the seven galaxies are not detected.

We model the properties using a (latent) marked Poisson point process, i.e., a Poisson point process for the scanned parameters, and a probability density function for the mark parameters. For concreteness, we focus on the luminosity function example, taking the scan parameter to be object position, x (a 2-vector), and the mark parameters to be flux and distance, (F, r). We suppose that the spatial density of galaxies is approximately constant over the region probed by the survey. There is thus a constant Poisson intensity

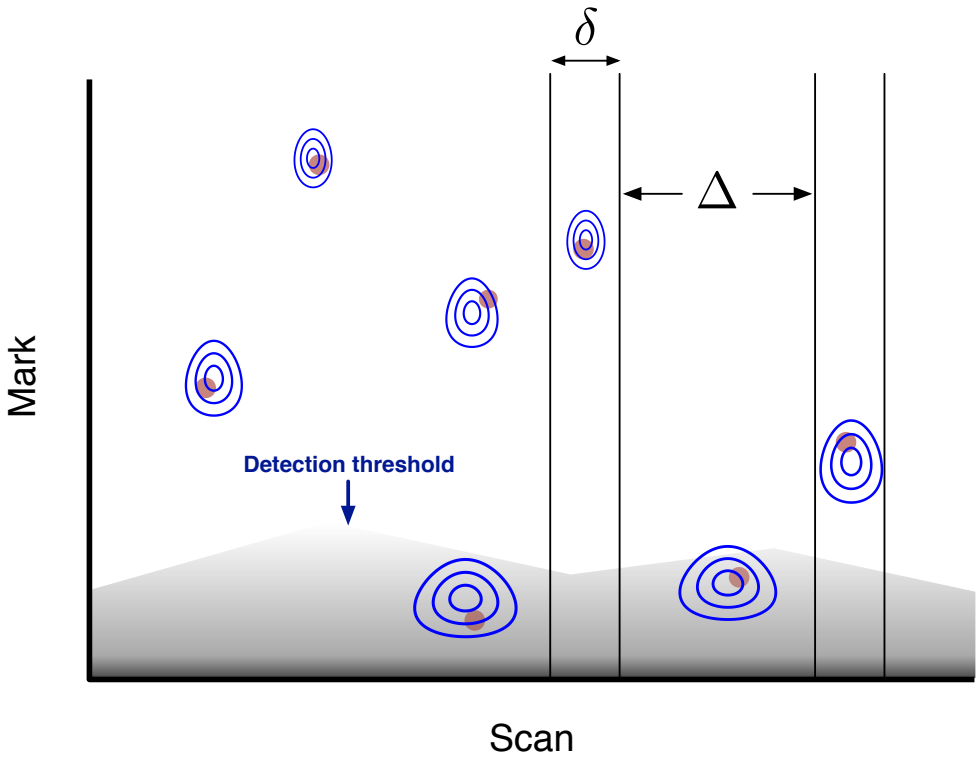


Figure 3.2: Depiction of thinned latent marked point process model for catalog data produced by an astronomical survey. Object properties are split into a scanned subset and a mark subset. Dots (red) show latent (true) values for an object's properties. Contours (blue) depict member likelihood functions from analysis of the raw survey data; catalogs provide summaries of these for detected objects. Gray region at bottom depicts the nondetection region; candidates with estimated mark values below a varying threshold are rejected. δ and Δ denote sizes of detection and nondetection intervals.

parameter, λ, describing the distribution of galaxies in x. We assume a BB1 luminosity PDF, independent of distance (of course, the flux PDF will depend on distance, thanks to the inverse square law). The flux and distance mark PDF is thus a product of PDFs, $h(r)\rho(F,r)$. For $\rho(F,r)$ we adopt the BB1 luminosity PDF form of (3.6), with parameters $\boldsymbol{\zeta} = (\beta, u, l)$, related to $\rho(F,r)$ via (3.4); we write the flux PDF as $\rho(F,r;\boldsymbol{\zeta})$ when we want to display the parameter dependence. For the PDF for galaxy distance, $h(r)$, the assumption of homogeneity implies

$$h(r) = \begin{cases} \dfrac{3r^2}{r_u^3} & \text{if } 0 \le r \le r_u, \\ 0 & \text{otherwise,} \end{cases} \tag{3.11}$$

where r_u is an upper limit on distance chosen to be beyond the surveyed volume.[6] The population model thus has parameters $\theta = (\lambda, \boldsymbol{\zeta})$.

We consider a case where we have precise distance measurements for the galaxies (e.g., from high-resolution spectroscopic data providing precise redshifts). We assume independent errors in the position and flux measurements, so the catalog contains descriptions of separate member likelihood functions for flux and position, denoted $\ell_i(F)$ and $m_i(x)$ for galaxy i, with $i = 1$ to N. Formally, denoting the image data for detected galaxy i by D_i, we are writing

$$p(D_i|x,F,r) = \ell_i(F)\, m_i(x)\, \delta(r - r_i), \tag{3.12}$$

where the Dirac delta function factor represents the precise measurement of distance. We must also describe the survey's selection effects. These are determined by the detection threshold as a function of the scan location. At each scanned location, x, the threshold determines the set, $\mathcal{D}_x$, of possible data (i.e., images) that would be deemed detections. For example, if the detection criterion is that the MLE flux estimate, $\hat{F}(D)$ for data D, must exceed a threshold $F_{\rm th}(x)$, then $\mathcal{D}_x = \{D : \hat{F}(D) > F_{\rm th}(x)\}$. Reporting $\mathcal{D}_x$, or equivalently $F_{\rm th}(x)$, then describes the selection effects. But we will see below that a simpler summary of the detection criteria will suffice.

[6] That is, chosen so that the most luminous galaxies of interest (i.e, with L of order u in the BB1 model) have fluxes comfortably below the lowest flux threshold. In deep surveys (reaching to very dim fluxes), cosmological considerations, including the finite age of the universe and the non-Euclidean geometry of spacetime, ameliorate the growth of $h(r)$ with r.

We now compute the likelihood function for the parameters, based on catalog data describing member likelihood functions and the selection effects. The construction we use is illustrated in Fig. 3.2. We partition the scan space into N detection intervals, δ_i, containing a single detected object, and M nondetection intervals, Δ_j, in which no candidate object passed the detection criterion.[7] The likelihood function is the product of the (conditionally independent) probabilities for these intervals.

We first consider the probability for no detection in one of the Δ_j intervals. We break it up into subintervals of size δx, small enough that the detection threshold is approximately constant over the interval. The probability for seeing no detections in δx is the sum of the probabilities for the following events (conditioned on the population parameters, $(\lambda, \boldsymbol{\zeta})$):

- No objects have x in the interval.
- One object has x in the interval, but it produced data that were not in $\mathcal{D}_x$.
- Two objects have x in the interval, but both produced data that were not in $\mathcal{D}_x$.
- And so on....

Each event is a conjunction of two simpler events, the Poisson probability for the specified number of objects lying in the interval, and the probability for not detecting any events in the interval. We will express the latter probability in terms of the *detection efficiency* at x for objects with flux F,

$$\eta(x, F) = p(D \in \mathcal{D}_x | F) \tag{3.13}$$

$$= p(\hat{F}(D) > F_{\rm th}(x) | F), \tag{3.14}$$

where the condition F denotes that an object is present with flux F. The probability for detecting an object with a given location and distance but unspecified flux and distance, given the population parameters, is then

$$p_x(\boldsymbol{\zeta}) = \int dr \int dF\, \rho(F, r)\, h(r)\, \eta(x, F). \tag{3.15}$$

[7] We are presuming that galaxy images are well-separated, i.e., we do not treat here the *crowded field* case, where the images of distinct objects may strongly overlap.

The probability for *not* detecting an object with a given location is then $1 - p_x(\boldsymbol{\zeta})$.

Now let ν denote the (unknown) number of objects with x in δx. Then the probability for no detections in δx at x is

$$\begin{aligned} q(x) &= \sum_{\nu=0}^{\infty} \frac{(\lambda \delta x)^{\nu}}{\nu!} e^{-\lambda \delta x} \left[1 - p_x(\boldsymbol{\zeta})\right]^{\nu} \\ &= e^{-\lambda \delta x} \sum_{\nu=0}^{\infty} \frac{(\lambda \delta x)^{\nu} \left[1 - p_x(\boldsymbol{\zeta})\right]^{\nu}}{\nu!} \\ &= \exp\left[-\lambda \delta x p_x(\boldsymbol{\zeta})\right]. \end{aligned} \tag{3.16}$$

This is the probability for no detections in a subinterval of a Δ_j interval. The probability for no detections across the entire interval is the product of its subinterval probabilities. The exponents add, so that the nondetection probability becomes

$$q(\Delta_j) = \exp\left[-\lambda \int_{\Delta_j} dx \int dr \int dF\, \eta(x, F)\, h(r)\, \rho(F, r)\right]. \tag{3.17}$$

This is just the Poisson probability for seeing no events, when the expected number of events is λ times the fraction of the population expected to be detected in the interval, given the threshold behavior (encoded in the detection efficiency).

Now consider the probability for the data associated with a detection interval, δ_i; for simplicity, we assume all of these intervals are of the same size, δ, in x. The probability for getting data D_i from detection of an object in δ_i is the sum of the probabilities for the following events:

- One object has x in the interval, and was detected producing data D_i.
- Two objects have x in the interval, one of which was detected producing D_i, with the other undetected.
- And so on....

To simplify the calculation, let us stipulate that the detected object has values of (x, F, r) in small intervals (dx, dF, dr); at the end, we will account for their uncertainty via marginalization.

The first case is simple; the probability for one object in the interval, having the specified properties, and being detected producing D_i, is

$$p_1(\lambda, \boldsymbol{\zeta}) = (\lambda\delta)e^{-\lambda\delta}\left[\frac{dx}{\delta}\, h(r)dr\, \rho(F, r; \boldsymbol{\zeta})dF\right] p(D_i \in \mathcal{D}_x, D_i | x, F, r). \tag{3.18}$$

The final probability is for a conjunction; it may be written

$$p(D_i \in \mathcal{D}_x, D_i | x, F, r) = p(D_i | x, F, r)\, p(D_i \in \mathcal{D}_x | D_i), \tag{3.19}$$

where we have dropped (x, F, r) from the last factor because the values of the properties are irrelevant for determining detection, once the data are in hand. Now note that detection is deterministic given the data, i.e., either the data correspond to a candidate meeting the detection criteria or not. But for a detected object, by definition the data met the criteria, so the last factor is equal to unity. The first factor we recognize as the member likelihood function, defined in (3.12). This completes the computation of $p_1(\lambda, \boldsymbol{\zeta})$.

For cases with $\nu > 1$ objects present, we will have a factor like $p_1(\lambda, \boldsymbol{\zeta})$ for the detected object, and nondection probabilities like the $[1 - p_x(\boldsymbol{\zeta})]$ factor appearing in the Δ_j probability derived above. But in addition, we have to account for not knowing which of the ν objects is detected. The resulting probability for the case of ν objects present can be written as follows:

$$\begin{aligned} p_\nu = & \frac{(\lambda\delta)^\nu}{\nu!} e^{-\lambda\delta} \\ & \times \left(\frac{dx}{\delta}\, h(r)dr\, \rho(F, r; \boldsymbol{\zeta})dF\right) \ell_i(F)\, m_i(x)\, \delta(r - r_i) \\ & \times [1 - p_x(\boldsymbol{\zeta})]^{\nu-1} \\ & \times \nu. \end{aligned} \tag{3.20}$$

Line by line, the factors are:

- the Poisson probability for ν objects being in the interval,
- the probability for one of them having the given properties and producing the detection data, D_i,
- the probability for the remaining objects not being detected,

- a factor of ν from summing over the possibilities for which of the ν objects is detected.

To facilitate summing the p_ν probabilities over ν, we gather the ν-dependent terms in (3.20) as follows:

$$\begin{aligned} p_\nu = (\lambda\delta)\, e^{-\lambda\delta} \left(\frac{dx}{\delta}\, h(r) dr\, \rho(F, r; \boldsymbol{\zeta}) dF \right) \ell_i(F)\, m_i(x)\, \delta(r - r_i) \\ \times \frac{1}{(\nu-1)!} (\lambda\delta)^{\nu-1} \left[1 - p_x(\boldsymbol{\zeta})\right]^{\nu-1} . \end{aligned} \tag{3.21}$$

Upon summing over $\nu \geq 1$, and marginalizing over the uncertain values of (x, F, r), we find that the probability for the detection data in interval δ_i is

$$p(D_i|\lambda, \boldsymbol{\zeta}) = q(\delta_i)\, h(r_i)\, (\lambda\delta) \left[\int_{\delta_i} \frac{dx}{\delta}\, m_i(x) \right] \left[\int dF\, \rho(F, r_i; \boldsymbol{\zeta})\, \ell_i(F) \right], \tag{3.22}$$

where $q(\delta_i)$ is an exponential of an integral of the same form as in the nondetection probability in (3.17).

The likelihood function is the product of detection probabilities (3.22) and nondetection probabilities (3.17) for all of the δ_i and Δ_j intervals. All of these probabilities share an exponential factor resembling (3.17). In the product, there will be a sum of the integrals in the exponents; this corresponds to a single integral over the entire x domain of the survey, of the form:

$$\lambda \int_\Omega dx \int dr \int dF\, \eta(x, F)\, h(r)\, \rho(F, r; \boldsymbol{\zeta}), \tag{3.23}$$

where Ω denotes the full range of positions surveyed (which would be measured in terms of solid angle on the sky). Note that the only x-dependent factor in the integrand is the detection efficiency. This lets us write the integral in simpler manner. Introduce the *average detection efficiency*,

$$\overline{\eta}(F) \equiv \frac{1}{\Omega} \int_\Omega dx\, \eta(x, F). \tag{3.24}$$

Using this, (3.23) can be written as a two-dimensional integral,

$$(\lambda\Omega) \int dr \int dF\, \overline{\eta}(F)\, h(r)\, \rho(F, r; \boldsymbol{\zeta}). \tag{3.25}$$

The factor $(\lambda\Omega)$ is the expected number of objects in the surveyed region, which depends only on the λ parameter. The remaining factor is the fraction of these that are detectable; it depends only on the remaining population parameters, $\boldsymbol{\zeta}$.

Equation (3.25) shows that the average efficiency is a kind of sufficient statistic for the survey's threshold behavior. Although catalog builders must determine the detection efficiency over the entire range of the survey, they need only report the lower-dimensional average efficiency for analysts.

We can now write down the full likelihood function for the luminosity function parameters. Dropping some factors that do not depend on the parameters, the likelihood function is

$$\begin{aligned}\mathcal{L}(\lambda,\boldsymbol{\zeta}) = \lambda^N \exp\left[-(\lambda\Omega)\int dr \int dF\, \overline{\eta}(F)\, h(r)\, \rho(F,r;\boldsymbol{\zeta})\right] \\ \times \prod_{i=1}^{N} h(r_i) \int dF\, \rho(F,r_i;\boldsymbol{\zeta})\, \ell_i(F).\end{aligned} \tag{3.26}$$

This likelihood function is reminiscent of that for an inhomogenous Poisson point process, whose likelihood is proportional to a product of intensity function factors, evaluated at the observed points, and an exponential whose negative argument is the integral of the intensity function over the observed domain. One difference is the integral over the latent observable, F, in the product factor; this accounts for measurement error. A more subtle difference is that integrand in the exponential is not the same function playing the role of the intensity function in the product factor. There is an average efficiency factor in the exponential, but not in the product factor. This is because of a feature of astronomical surveys noted earlier: the data used for characterization (estimating latent parameters) is also used for detection. As a result, were one to insert an efficiency factor into the product terms, the data would be doubly used. This appeared explicitly in our derivation; the text after (3.19). Some heuristic derivations of similar likelihood functions in the astronomical literature have missed this point, instead inserting an $\overline{\eta}(F)$ factor in the detected object integrals in the likelihood function. This corrupts inferences; see Loredo (2004) for further discussion.

Notably, the Poisson process intensity parameter, λ, appears in the likelihood function only as the power, λ^N, and multiplying the integral in the exponential. As a result, if we adopt a conjugate prior for λ (a gamma distribution), we can easily compute the marginal likelihood function for the $\boldsymbol{\zeta}$ parameters. For simplicity, we adopt the limiting case of a uniform prior for λ. Marginalizing over λ and dropping some $\boldsymbol{\zeta}$-independent terms, we find

that the marginal likelihood function for $\boldsymbol{\zeta}$ takes the form

$$\mathcal{L}_m(\boldsymbol{\zeta}) = \prod_{i=1}^{N} \int dF\, \mu(F, r_i; \boldsymbol{\zeta})\, \ell_i(F), \tag{3.27}$$

where we have introduced an *effective density* for the latent observables, F and r,

$$\mu(F, r; \boldsymbol{\zeta}) \equiv \frac{h(r)\rho(F, r; \boldsymbol{\zeta})}{\int dr \int dF\, \overline{\eta}(F)\, h(r)\, \rho(F, r; \boldsymbol{\zeta})}. \tag{3.28}$$

Equation (3.27) resembles the likelihood function for a binomial point process, where the observations have measurement errors described by the member likelihood functions. But the analogy is not exact, because the effective density is not a PDF (it does not integrate to unity).

3.4 Implementation with CUDAHM

The thinned latent Point process framework is a more complicated hierarchical model than those depicted in Fig. 2.1. Fig. **??** shows a schematic DAG for the framework. Separate plates depict the conditional independence structure for parts of the joint distribution describing detected and undetected objects (a more detailed DAG would partition the nondetection among the Δ_j intervals; this would involve nested plates). The number of replications for the detection and nondetection plates, N and $\overline{N}$, are random variables, since the number of objects in the surveyed region is not known a priori, and is informative about the population parameters.

Despite these differences with respect to the simpler model structures discussed previously, the structure of the likelihood functions in (3.26) and (3.27) is essentially the same as that for conditional density estimation with measurement error (the middle DAG in Fig. **??**). This is because the nondetection part of the DAG in Fig. **??** corresponds to a product of exponentials, whose arguments sum in a single integral: the integral on the first line of the likelihood function in (3.26), and in the denominator of the effective densigyt in (3.28). The likelihood or marginal likelihood thus has a single product term, composed of independent factors for each detected object—just the type of structure CUDAHM was designed to sample from.

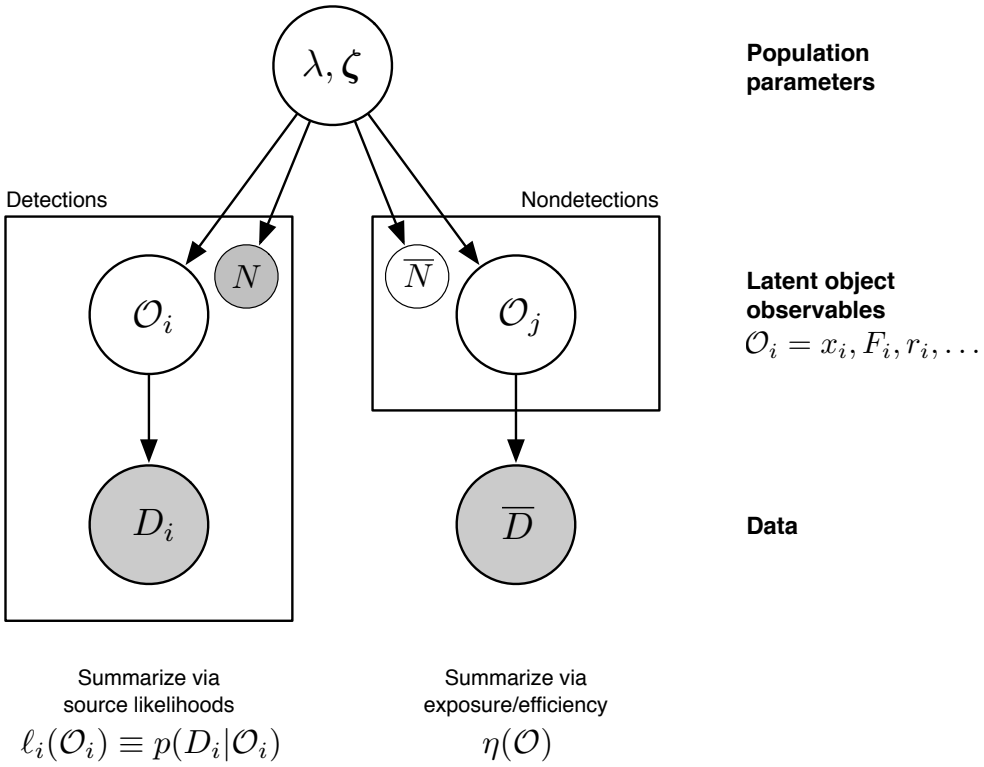


Figure 3.3: Schematic DAG for a thinned latent marked point process model for luminosity function estimation from survey catalog data. The small N and $\overline{N}$ nodes specify the numbers of replications of the detection and nondetection plates, respectively.

4 Luminosity function estimation: Simulation study

To explore the performance of CUDAHM for luminosity function estimation, we implemented the framework just described and applied it to simulated data. To focus on performance, this example ignores important complexities arising in modeling real galaxy catalog data. For example, we ignore cosmological corrections to the inverse square law (which depend on uncertain cosmological parameters). We also ignore diversity in the spectra of galaxies. This is important for real data because instruments gather light in limited spectral ranges, determined by properties of the atmosphere, telescope optics, and detector sensitivity. Among the optical elements are filters that deliberately restrict the spectral passband, so that repeated measurements with different filters can provide broadband information about galaxy spectra. As a result, galaxies with the same bolometric (full-spectrum) luminosity and distance, but different spectral shapes, will have different apparent brightnesses (fluxes). A full analysis would incorporate data from multiple bands. For the study described here, we assume the simulated galaxies have the same spectra, or that the catalog estimates have been adjusted for spectral diversity.

4.1 Simulation setup: population parameters and priors

We simulate observations of a population described by the BB1 model, with parameters chosen so that galaxies with $L > l$ have a distribution similar to that found in the analysis by Blanton et al. (Blanton et al. 2003, B03) of $\approx 150,000$ galaxies with spectroscopic redshifts observed in the Sloan Digital Sky Survey (SDSS). We set $u = 1\times10^{10}L_{\odot}$ (where $L_{\odot}$ denotes the solar luminosity; this is approximately equal to the B03 value of L_* in a Schecter function fit), $\beta = -1.5$ (a bit steeper than the B03 value), and $l = 1\times10^{8}L_{\odot}$, corresponding to the lowest luminosities studied by B03. We choose survey parameters corresponding to a deeper survey than SDSS, thus probing luminosities dimmer than $L = l$. These choices are motivated in part by current and emerging surveys, such as the photometric (broadband) surveys by the Panoramic Survey Telescope and Rapid Response System (Pan-STARRS, current) and the Large Synoptic Survey Telescope (LSST; starting in 2023), and the spectroscopic (redshift) survey by the Dark Energy Spectroscopic Instrument (DESI; starting in 2019).

For distances, we sample values from a spatially homogeneous population extending out to a maximum distance $r_{\max} = 1$ Gpc. Only very luminous sources can be detected from large distances. For the detection criteria we adopt (see below), this maximum distance is such that galaxies are visible beyond $r_{\max}$ only if $L \gtrsim 20u$, an event with negligible probability (which we formally exclude by truncation).

•••*[Janos also truncated below l; is this necessary?]*•••

The upper luminosity scale, u, and maximum distance, $r_{\max}$, together define a fiducial flux value,

$$F_{\rm fid} \equiv \frac{u}{4\pi r_{r_{\max}}^2} \approx 3.2 \times 10^{-13} \left(\frac{u}{10^{10} L_\odot}\right) \left(\frac{r_{\max}}{1 \text{ Gpc}}\right)^{-2} \text{ erg cm}^{-2} \text{ s}^{-1}. \tag{4.1}$$

This is a convenient unit in which to express fluxes. Although $F_{\rm fid}$ is minuscule, modern survey telescopes would detect $\sim 10^4$ to 10^5 photons from a source with this flux.

•••*[Note the parameter value changes! The plots need revision; luminosity axis labels need to be shifted by* $\times 100$*, and the axis label should read* $\log_{10}(L/L_\odot)$*. Check the* $r_{\max}$ *value and make sure it corresponds to the flux limit and luminosity truncation after shifting to the new params.]*•••

Now we consider the choice of prior for the population parameters. For β, we adopt a prior that is flat with respect to the angle in log-log space. This choice has the virtue of not putting a lot of prior probability on the steep slope range, which a flat prior on β would do. Denoting the angle by φ, if we adopt a prior PDF of $h(\varphi)$ on φ, the prior on $\beta = \tan\varphi$ is

$$p(\beta) = \frac{h(\varphi)}{1+\beta^2}. \tag{4.2}$$

For a flat φ prior between two cut-offs φ_L and φ_U,

$$p(\beta) = \frac{1}{\varphi_U - \varphi_L} \cdot \frac{1}{1+\beta^2}. \tag{4.3}$$

This is a truncated Cauchy distribution. The BB2 distribution requires $\beta > -2$, corresponding to $\varphi_L = -1.107$. If we require Eq. 3.6 to be decreasing, the upper limit becomes $\beta < 0$, corresponding to $\varphi_U = 0$. Thus, the prior on β is

$$p(\beta) = \frac{0.903}{1+\beta^2} \qquad \text{for} -2 < \beta < 0. \tag{4.4}$$

For the upper scale we use a log-flat prior, a conventional choice for a scale parameter that must be positive, even though this will be improper on both sides, we can ignore the impropriety and the normalizing constant since the likelihood function will make the posterior proper. A prior flat in $\log u$ corresponds to $p(u) \propto \frac{1}{u}$. The lower scale l must be below the upper scale u, which we can ensure by using a prior factored as $p(l, u) = p(u)p(l \,|\, u)$ and taking $p(l \,|\, u)$ to vanish for $l \geq u$. A log-flat prior also seems appealing for l but the data (i.e. luminosity measurements) do not probe the distribution down to zero due to the flux limit of the telescope, we could not have a proper prior without introduction a lower cut-off. Instead, we simply use a flat prior on l. Hence, the overall prior will be

$$p(\beta, l, u) \propto \frac{l}{u \cdot (1 + \beta^2)} \qquad \text{for} - 2 < \beta < 0, l < u. \tag{4.5}$$

4.2 Simulation setup: detection and measurement

We simulate measurement errors and selection effects using a simplified model commonly adopted in astronomical simulation studies (Fan 1999, Ivezic et al. 2008). Modern astronomical optical detectors, such as cameras using charge-coupled devices (CCDs), count photons. The Poisson distribution accurately describes photon arrival and detection, but there are additional contributions to measurement uncertainty, including from backgrounds and electronic noise. To screen out false detections, only reasonably strong candidate sources are accepted as genuine, so that catalog data are typically in the large-counts regime. Simulation models work in this regime, approximating the Poisson distribution by a normal distribution, and treating the additional contributions to measurement uncertainty also using normal distributions. The overall measurement uncertainty thus is approximately normal, with a variance found by adding the variances of the component processes.

For simulation studies of hierarchical Bayes approaches, it is important to distinguish approximations of the sampling distribution used to generate simulated data, and approximations of the member likelihood functions needed for inference with a particular simulated dataset. As a concrete illustration of the distinction, consider a source with true flux F being measured by an idea photon counting detector, with measurement uncertainty due only to Poisson counting uncertainty associated with the source flux. For an instrument with projected area A observing for a time T, the expected number of photons is $\mu = ATF$,

and the standard deviation in the number of photons is $\mu^{1/2}$. In the high-counts regime, we could simulate an observation by drawing a number of photon counts, n, from a normal distribution $N(\mu, \mu)$, i.e., with

$$p(n|F) \approx \frac{1}{(ATF)^{1/2}\sqrt{2\pi}} \exp\left[-\frac{(n - ATF)^2}{2ATF}\right]. \tag{4.6}$$

Suppose the simulated value of n is $n_{\rm obs}$. For analysis of that observation, we would need to approximate the member likelihood function based on that datum. The exact likelihood function, $\ell(F)$, based on the Poisson sampling distribution, is proportional to a gamma distribution with shape parameter $\alpha = n_{\rm obs} + 1$ and scale parameter $1/AT$, with its mode at $\hat{F} = n_{\rm obs}/(AT)$, which could serve as a convenient point estimator. Expressed in terms of $\mu = ATF$, this gamma distribution has mean $\langle ATF \rangle = n_{\rm obs} + 1$ and variance $n_{\rm obs} + 1$. For large $n_{\rm obs}$, the member likelihood function is thus well approximated by a Gaussian function with mean and variance equal to $n_{\rm obs} + 1 \approx n_{\rm obs}$:

$$\begin{aligned} \ell(F) &\propto \frac{1}{n_{\rm obs}^{1/2}\sqrt{2\pi}} \exp\left[-\frac{(ATF - n_{\rm obs})^2}{2n_{\rm obs}}\right] \\ &\propto \exp\left[-\frac{(F - n_{\rm obs}/(AT))^2}{2n_{\rm obs}/(AT)^2}\right]. \end{aligned} \tag{4.7}$$

Thus for generating simulated data, we would use a normal distribution, (4.6), with variance depending on the true flux. But for analyzing a simulated data set, we would use likelihood functions proportional to normal distributions, (4.6), with variances depending on the simulated data, $n_{\rm obs}$, for each object.

For our simulations of catalog data, we use normal sampling distributions for estimated fluxes, $\hat{F}_i$, with standard deviations that depend on the true fluxes, F_i, according to

$$\begin{aligned} \sigma(F) &= \sqrt{\sigma_0^2 + (\alpha F)^2} \\ &= \sigma_0 \left[1 + \left(\frac{\alpha F}{\sigma_0}\right)^2\right]^{1/2}, \end{aligned} \tag{4.8}$$

where σ_0 characterizes the noise contributions from backgrounds and detector electronics, and α characterizes how Poisson fluctuations in the number of detected photons influence $\hat{F}_i$, relative to the flux-independent noise sources. For an object with simulated best-fit flux $\hat{F}_i$, we use a member likelihood function that is a Gaussian function with mode at $\hat{F}_i$

and standard deviation parameter

$$\hat{\sigma}_i(\hat{F}) = \sqrt{\sigma_0^2 + (\alpha \hat{F}_i)^2}, \tag{4.9}$$

using the same values for σ_0 and α as are used for the sampling distributions. Based on published simulations of existing and anticipated surveys (Fan 1999, Ivezic et al. 2008), we set $\alpha = 10^{-2}$ for our simulations. We set $\sigma_0 = 6.4 \times 10^{-10}$ and $\alpha = 10^{-2}$ and $\sigma_0 = 6.4 \times 10^{-10}$ erg cm^{-2} s^{-1} for our simulations. The latter value was chosen so that sources with $L = 20u$ at $r = r_{\max}$ are just detectable by the error-based detection criterion described below.

•••*[Check these error/threshold params; they should correspond to Janos's values shifted to physical units.]*•••

For detection, we require a candidate object to have estimated flux a factor $\nu = 5$ times the measurement error. This corresponds to a threshold flux satisfying $F_{\rm th} = \nu \hat{\sigma}_i(F_{\rm th})$, which gives

$$F_{\rm th} = \frac{\nu \sigma_0}{\sqrt{1 - \alpha^2}}. \tag{4.10}$$

For the parameters of our simulation, this corresponds to $F_{\rm th} = X$. The detection efficiency is the probability that the measured value, $F_{\rm th}$, for a source with true flux F will be above the threshold,

$$\begin{aligned} \eta(F) &= p(\hat{F} > F_{\rm th}|F) \\ &= \Phi\left(\frac{F - F_{\rm th}}{\sigma(F)}\right), \end{aligned} \tag{4.11}$$

where $\Phi(\cdot)$ denotes the standard normal cumulative distribution function.

4.3 Simulation setup

•••*[IGNORE THIS SUBSECTION.]*•••

In case of real measurements the distance of the objects is known from redshift observations. In our simplified case we assume these distance measurements to be without error. To compile the simulated data set, we assume a spherically symmetric, homogeneous distribution of galaxies and generate the random distances r_i accordingly. We use

the inverse-square law

$$F_i = \frac{L_i}{4\pi r_i^2}. \tag{4.12}$$

The observational noise E of the flux F is modeled as Gaussian with zero mean and a standard deviation of

$$\sigma(F) = \sqrt{\sigma_0^2 + (0.01F)^2}. \tag{4.13}$$

The flux limit of the telescope is the constant T and C will denote the event when an object is detected, i.e. the noisy flux is above the threshold: $D > T$. When generating the random sample, the luminosity is limited between $10 \leq \log_{10} L \leq 14$ which implies the distance limit

$$r_{\max} = \sqrt{\frac{L_{\max}}{4\pi T}}. \tag{4.14}$$

As it was discussed above in Sec. 1, the spatial distribution of galaxies is considered homogeneous, hence the density function becomes

$$\delta(r) = \begin{cases} \dfrac{3r^2}{r_{\max}^3} & \text{if } 0 \leq r \leq r_{\max} \\ 0 & \text{otherwise} \end{cases} \tag{4.15}$$

and the cumulative distribution function:

$$\Delta(r) = \begin{cases} 0 & \text{if } r < 0 \\ \dfrac{r^3}{r_{\max}^3} & \text{if } 0 \leq r \leq r_{\max} \\ 1 & \text{if } r \geq r_{\max} \end{cases} \tag{4.16}$$

To take the selection effect of the flux limit into account, we have to calculate the probability of measuring a galaxy with a given luminosity and distance above the flux limit. By assuming Gaussian noise, as we already did in Sec. **??**, the sought probability

becomes

$$p(C|L,r) = p(C|F) = \Pr(D > T|F) \tag{4.17}$$

$$= \Pr(F + E > T) \tag{4.18}$$

$$= \Pr(\underbrace{F+E}_{\sim \mathcal{N}(F,\sigma(F))} > T) \tag{4.19}$$

$$= \frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{F-T}{\sqrt{2}\sigma(F)}\right)\right) \tag{4.20}$$

$$=: \zeta(F; T, \sigma_0) \tag{4.21}$$

4.4 Case study

•••*[Tables and plots need to be adjusted here to shift to new units.]*•••

We apply the hierarchical Bayesian model to estimate the model parameters of a simulated random sample of 100,000 galaxies. We compare the results of the Bayesian analysis to maximum likelihood and evaluate the performance of the GPU-based implementation. The true values of the population parameters are listed in the first row of Tab. 4.1. The value of σ_0 in Eq. 4.13 is chosen to be $\sigma_0 = 1$. The flux limit is $T = 5.0$ and the distance limit is $r_{\max} \gtrsim 1.12 \times 10^6$.

We executed 1.5M burn-in and the same number of live MCMC steps to sample the probability distribution of the population parameters. The length of the burn-in sequence was chosen by visual inspection of the autocorrelation plots. To reduce autocorrelations in the Markov chain, θ samples were thinned and only every 150^{th} sample was kept, hence the final number of samples was 10,000. Fig. 4.1 shows the traces, histograms and autocorrelation plots of the population parameters whereas Tab. 4.1 lists the result in a numerical format.

4.5 Performance tests

We used NVIDIA Tesla K40c cards for the performance tests. First, we executed tests without imposing a flux limit, which is a much simpler case as it does not contain the time-consuming numerical integration of Eq. **??**. Next, we executed the with the flux limit turned on. Fig. 4.2 shows the elapsed time as a function of iteration number (non-thinned)

	β	l	u
θ_{true}	-1.5	5.0×10^{10}	5.0×10^{12}
θ_{MCMC}	-1.5037	5.0302×10^{10}	5.0000×10^{12}
$\sigma_{\theta,\text{MCMC}}$	0.0059	3.3548×10^{9}	3.1806×10^{10}
θ_{MLE}	-1.5564	7.3222×10^{10}	5.7207×10^{12}
$\theta_{\text{MLE, no noise}}$	-1.5009	4.9341×10^{10}	4.9819×10^{12}
$\lvert\theta_{\text{true}} - \theta_{\text{MLE}}\rvert$	$> 9.5525 \cdot \sigma_{\beta,\text{MCMC}}$	$> 6.9221 \cdot \sigma_{l,\text{MCMC}}$	$> 22.6591 \cdot \sigma_{u,\text{MCMC}}$

Table 4.1: Summary of parameter estimation results. The first row of the table indicates the true values which the simulated data was generated with. The second row shows the mean of the distributions coming of the Bayesian model, whereas the third row contains the standard deviation of them. For reference, we indicate the outcome of the ML estimatior run on the simulated data with and without noise in rows 4 and 5, respectively. To compare the Bayesian model to ML, the last row of the table shows the difference between the ML estimator and the true values in terms of the standard deviation of the posterior from the Bayesian model.

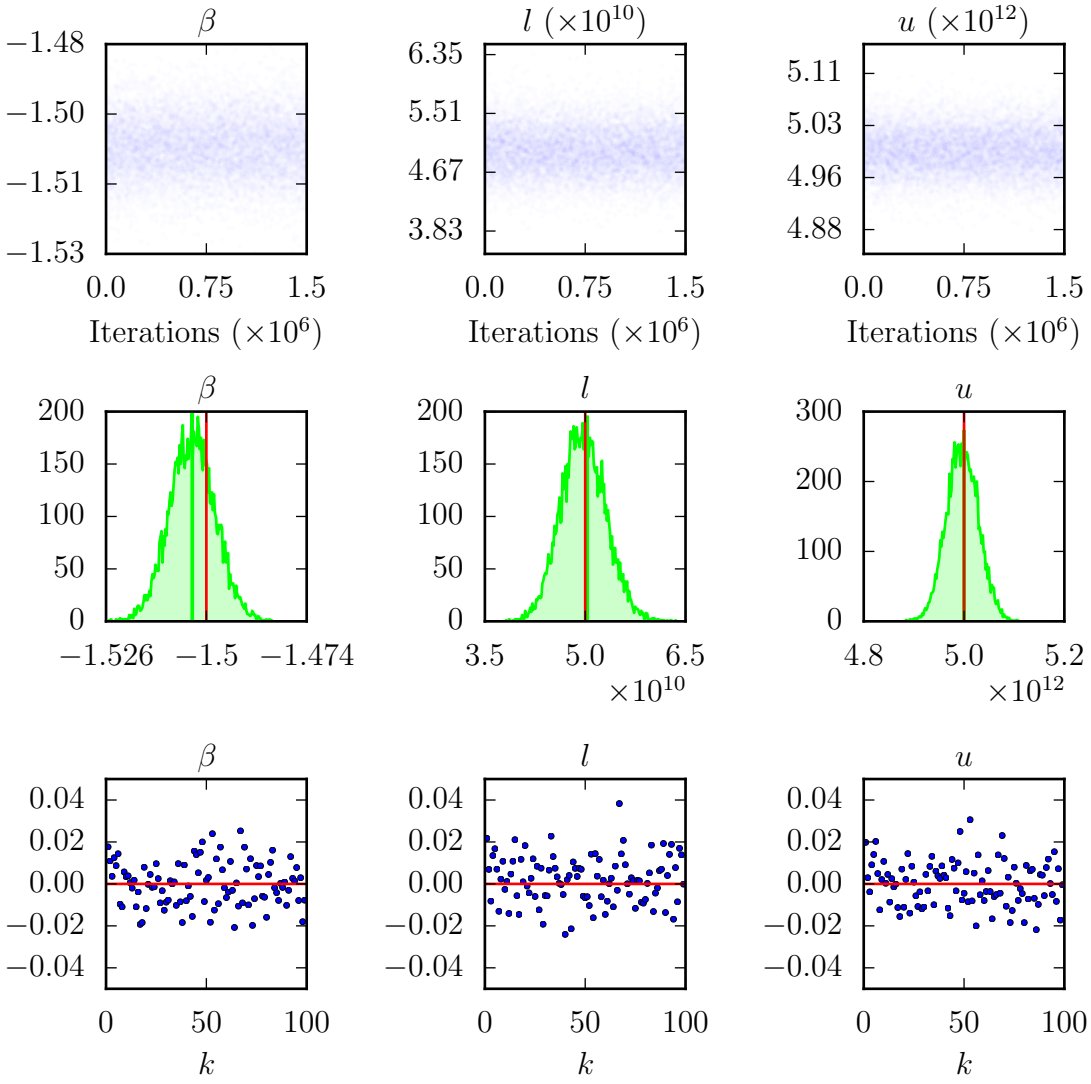


Figure 4.1: Trace (upper row), histogram (middle row) and autocorrelation (lower row) plots of the three population parameters. The red vertical line overplotted the histograms show the true value of the parameter. With the exception of β, Bayesian modelling can recover model parameters with excellent accuracy.

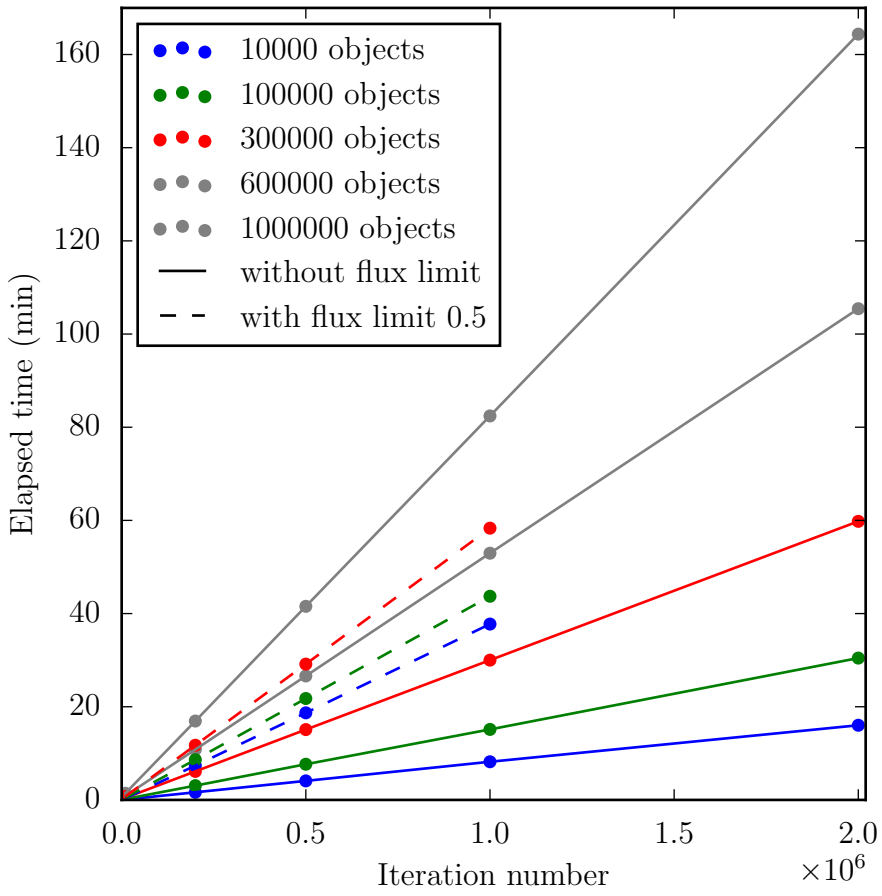


Figure 4.2: Runtime of the GPU code as a function of iterations for different numbers of objects, with (solid lines) and without (dashed lines) considering a flux limit of $T = 0.5$. Taking the flux limit into account results in an approximately fourfold increase of runtime.

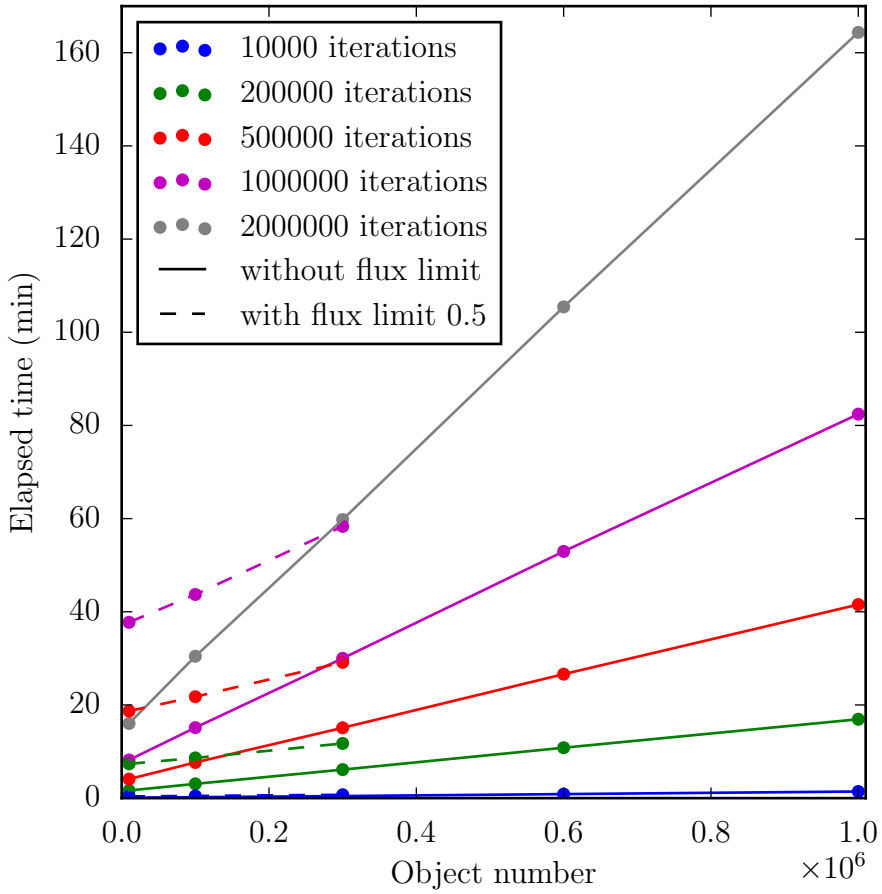


Figure 4.3: Runtime of the GPU code as a function of the number of objects, after a given number of iterations, with (solid lines) and without (dashed lines) considering a flux limit of $T = 0.5$. Taking the flux limit into account results in an approximately fourfold increase of runtime. On the other hand, the scaling of runtime with the number of objects is linear. This is due to the fact that fluxes are statistical independence, hence their distribution can be sampled in parallel on the GPU.

for different number of objects. The functions are trivially linear as iteration steps always take a fixed amount of time. More informative is Fig. 4.3 where we plot the elapsed time as a function of the number of objects for various numbers of iteration steps. The linear scaling of computation time with the number of objects is due to the fact that the characteristics are independent. As it is visible from the figures, turning on the flux limit clearly decreases the performance. Real galaxy catalogs contain objects on the order of 10^8, two magnitudes more than our simulated data set. Extrapolating from out performance numbers, estimating the parameters of the luminosity function with 2×10^5 Markov steps would take about 2000 minutes, a bit less then one and a half days, which makes applying our method to real data feasible.

5 Summary

We have described a usefully general C++ framework for massively parallel implementation of simple hierarchical Bayesian models. We have applied this framework to estimating luminosity functions of objects measured with uncertainty, and subject to (quantified) selection effects. Our implementation shows linear scaling with both, the number of Markov chain iterations and the number of objects, which makes it applicable to real data sets of size up to 10^8 using modest-sized clusters of GPUs.

References

Ahrens, J. H. & Dieter, U. (1974), 'Computer methods for sampling from gamma, beta, poisson and bionomial distributions', *Computing* **12**(3), 223–246.
URL: *https://link.springer.com/article/10.1007/BF02293108*

Binggeli, B., Sandage, A. & Tammann, G. A. (1988), 'The luminosity function of galaxies', *Annual Review of Astronomy and Astrophysics* **26**, 509–560.
URL: *http://adsabs.harvard.edu/abs/1988ARA%26A..26..509B*

Binney, J. & Merrifield, M. (1998), *Galactic Astronomy.*

Blanton, M. R., Hogg, D. W., Bahcall, N. A., Brinkmann, J., Britton, M., Connolly, A. J., Csabai, I., Fukugita, M., Loveday, J., Meiksin, A. et al. (2003), 'The galaxy luminosity function and luminosity density at redshift z= 0.1', *The Astrophysical Journal* **592**, 819–838.

Carroll, R. J., Ruppert, D., Stefanski, L. A. & Crainiceanu, C. M. (2006), *Measurement error in nonlinear models*, Vol. 105 of *Monographs on Statistics and Applied Probability*, second edn, Chapman & Hall/CRC, Boca Raton, FL. A modern perspective.
URL: *http://dx.doi.org/10.1201/9781420010138*

Fan, X. (1999), 'Simulation of Stellar Objects in SDSS Color Space', *The Astronomical Journal* **117**, 2528–2551.
URL: *http://adsabs.harvard.edu/abs/1999AJ....117.2528F*

Ivezic, Z., Tyson, J. A., Abel, B., Acosta, E., Allsman, R., AlSayyad, Y., Anderson, S. F., Andrew, J., Angel, R., Angeli, G., Ansari, R., Antilogus, P., Arndt, K. T., Astier, P., Aubourg, E., Axelrod, T., Bard, D. J., Barr, J. D., Barrau, A., Bartlett, J. G., Bauman, B. J., Beaumont, S., Becker, A. C., Becla, J., Beldica, C., Bellavia, S., Blanc, G., Blandford, R. D., Bloom, J. S., Bogart, J., Borne, K., Bosch, J. F., Boutigny, D., Brandt, W. N., Brown, M. E., Bullock, J. S., Burchat, P., Burke, D. L., Cagnoli, G., Calabrese, D., Chandrasekharan, S., Chesley, S., Cheu, E. C., Chiang, J., Claver, C. F., Connolly, A. J., Cook, K. H., Cooray, A., Covey, K. R., Cribbs, C., Cui, W., Cutri, R., Daubard, G., Daues, G., Delgado, F., Digel, S., Doherty, P., Dubois, R., Dubois-Felsmann, G. P., Durech, J., Eracleous, M., Ferguson, H., Frank, J., Freemon, M., Gangler, E., Gawiser, E., Geary, J. C., Gee, P., Geha, M., Gibson, R. R., Gilmore, D. K., Glanzman, T., Goodenow, I., Gressler, W. J., Gris, P., Guyonnet, A., Hascall, P. A., Haupt, J., Hernandez, F., Hogan, C., Huang, D., Huffer, M. E., Innes, W. R., Jacoby, S. H., Jain, B., Jee, J., Jernigan, J. G., Jevremovic, D., Johns, K., Jones, R. L., Juramy-Gilles, C., Juric, M., Kahn, S. M., Kalirai, J. S., Kallivayalil, N., Kalmbach, B., Kantor, J. P., Kasliwal, M. M., Kessler, R., Kirkby, D., Knox, L., Kotov, I., Krabbendam, V. L., Krughoff, S., Kubanek, P., Kuczewski, J., Kulkarni, S., Lambert, R., Guillou, L. L., Levine, D., Liang, M., Lim, K.-T., Lintott, C., Lupton, R. H., Mahabal, A., Marshall,

P., Marshall, S., May, M., McKercher, R., Migliore, M., Miller, M., Mills, D. J., Monet, D. G., Moniez, M., Neill, D. R., Nief, J.-Y., Nomerotski, A., Nordby, M., O'Connor, P., Oliver, J., Olivier, S. S., Olsen, K., Ortiz, S., Owen, R. E., Pain, R., Peterson, J. R., Petry, C. E., Pierfederici, F., Pietrowicz, S., Pike, R., Pinto, P. A., Plante, R., Plate, S., Price, P. A., Prouza, M., Radeka, V., Rajagopal, J., Rasmussen, A., Regnault, N., Ridgway, S. T., Ritz, S., Rosing, W., Roucelle, C., Rumore, M. R., Russo, S., Saha, A., Sassolas, B., Schalk, T. L., Schindler, R. H., Schneider, D. P., Schumacher, G., Sebag, J., Sembroski, G. H., Seppala, L. G., Shipsey, I., Silvestri, N., Smith, J. A., Smith, R. C., Strauss, M. A., Stubbs, C. W., Sweeney, D., Szalay, A., Takacs, P., Thaler, J. J., Van Berg, R., Berk, D. V., Vetter, K., Virieux, F., Xin, B., Walkowicz, L., Walter, C. W., Wang, D. L., Warner, M., Willman, B., Wittman, D., Wolff, S. C., Wood-Vasey, W. M., Yoachim, P., Zhan, H. & Collaboration, f. t. L. (2008), 'LSST: from Science Drivers to Reference Design and Anticipated Data Products', *arXiv:0805.2366 [astro-ph]* . arXiv: 0805.2366.

URL: *http://arxiv.org/abs/0805.2366*

Johnston, R. (2011), 'Shedding light on the galaxy luminosity function', *Astronomy and Astrophysics Review* **19**, 41.

Kelly, B. C. (2012), Measurement Error Models in Astronomy, *in* E. D. Feigelson & G. J. Babu, eds, 'Statistical Challenges in Modern Astronomy V', Lecture Notes in Statistics, Springer New York, pp. 147–162.

Kelly, B. C., Fan, X. & Vestergaard, M. (2008), 'A flexible method of estimating luminosity functions', *The Astrophysical Journal* **682**(2), 874–895.

Kelly, B. C., Shetty, R., Stutz, A. M., Kauffmann, J., Goodman, A. A. & Launhardt, R. (2012), 'Dust Spectral Energy Distributions in the Era of Herschel and Planck: A Hierarchical Bayesian-fitting Technique', *The Astrophysical Journal* **752**, 55.

Kroupa, P. (2007), The stellar initial mass function, Vol. 241, pp. 109–119.

URL: *http://adsabs.harvard.edu/abs/2007IAUS..241..109K*

Loredo, T. J. (2004), Accounting for Source Uncertainties in Analyses of Astronomical

Survey Data, *in* R. Fischer, R. Preuss & U. V. Toussaint, eds, 'American Institute of Physics Conference Series', Vol. 735 of *American Institute of Physics Conference Series*, pp. 195–206.

Loredo, T. J. (2013), Bayesian Astrostatistics: A Backward Look to the Future, *in* J. M. Hilbe, ed., 'Astrostatistical Challenges for the New Astronomy', number 1 *in* 'Springer Series in Astrostatistics', Springer New York, pp. 15–40.

Loredo, T. J. & Wasserman, I. M. (1995), 'Inferring the spatial and energy distribution of gamma-ray burst sources. 1: Methodology', *The Astrophysical Journal Supplement Series* **96**, 261–301.
URL: *http://adsabs.harvard.edu/abs/1995ApJS...96..261L*

Loredo, T. J. & Wasserman, I. M. (1998), 'Inferring the Spatial and Energy Distribution of Gamma-Ray Burst Sources. II. Isotropic Models', *The Astrophysical Journal* **502**, 75–107.
URL: *http://adsabs.harvard.edu/abs/1998ApJ...502...75L*

McGreer, I. D., Jiang, L., Fan, X., Richards, G. T., Strauss, M. A., Ross, N. P., White, M., Shen, Y., Schneider, D. P., Myers, A. D., Brandt, W. N., DeGraf, C., Glikman, E., Ge, J. & Streblyanska, A. (2013), 'The z = 5 Quasar Luminosity Function from SDSS Stripe 82', *The Astrophysical Journal* **768**, 105.
URL: *http://adsabs.harvard.edu/abs/2013ApJ...768..105M*

Petit, J.-M., Kavelaars, J. J., Gladman, B. & Loredo, T. (2008), Size Distribution of Multikilometer Transneptunian Objects, *in* 'The Solar System Beyond Neptune', pp. 71–87.
URL: *http://adsabs.harvard.edu/abs/2008ssbn.book...71P*

Vihola, M. (2012), 'Robust adaptive metropolis algorithm with coerced acceptance rate', *Statistics and Computing* **22**(5), 997–1008.