Appendix A:
Overview of CUDAHM API

CUDAHM enables C++ programmers to rapidly construct a MWG sampler for a simple hierarchical model, requiring the user to supply only a minimimal amount of CUDA code. A template file in the distribution, `cudahm_blueprint.cu`, contains a heavily commented example application that users may copy and customize. Here we briefly highlight key elements of the application programming interface (API).

For basic applications, the user must define two functions, instantiate a single class that manages the computations, and create a configuration file read by the compiled application that specifies parameters defining an MCMC run. The functions compute probabilities needed by the two steps in the MWG algorithm. One computes the logarithm of the probability density for the measurements given the member properties for each object in the sample (the product of the $\ell(\cdot)$ member likelihood functions). The second computes the logarithm of the probability density for the member properties given the parent population parameters (the product of $f(\cdot)$ population distribution factors). Both functions execute on the GPU and must be written as CUDA kernels (i.e., using the CUDA extensions to the `C` language).

To manage the computations, the user must instantiate one of two classes. A `GibbsSampler` instance runs the MWG algorithm, storing all samples in memory; the user can write out the samples, or a subset, as needed upon completion. Alternatively, a `GibbsSamplerWithCompactMemoryUsage` instance runs the algorithm, but with more efficient use of memory; it opens an output file stream (which has a buffer) and writes samples out on the fly.

Internally, these sampler classes instantiate two classes that users may wish to customize (by subclassing) for advanced use cases (in which case pointers to the subclasses must be provided to the sampler's constructor). The `DataAugmentation` class controls calculations involving the member properties. The `PopulationPar` class controls calculations involving the population parameters. Users may wish to subclass these classes in order to customize one or both steps of the MWG algorithm. For example, evaluation of the prior for population parameters occurs in a `PopulationPar` instance, which implements a uniform prior as a default choice. To override this default prior, the `PopulationPar` class should be subclassed.

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.

CUDAHM is open-source software maintained on GitHub at `https://github.com/`

`tloredo/CUDAHM`; the branch named `paper1` contains the version described in this paper and can be obtained via the following Git command:

```
git clone -b paper1 https://github.com/tloredo/CUDAHM [directory]
```

where `[directory]` denotes an optional name for the destination directory (overriding the default `CUDAHM`). The CUDAHM distribution contains 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 briefly described in the main text at the end of § 2.1).

Appendix B: Thinned latent marked point process likelihood function

In this Appendix we derive the likelihood function for the parameters of a luminosity function model, based on catalog data composed of member likelihood functions for detected objects, and appropriate summaries of the detection criteria to account for selection effects.

Object detection is typically implemented via a scanning procedure. For example, for image data, a fixed aperture may be scanned over the image. As the scan proceeds, a detection algorithm determines if an object is present at each candidate location, 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 more carefully estimated, e.g., by a likelihood-based or weighted least-squares calculation, with estimation results summarized in the catalog.

Fig. B.1 illustrates the process 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. The region is depicted with a gradient to depict that, as a function of *true* object properties, the resulting selection is probabilistic. Here two of the seven objects 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

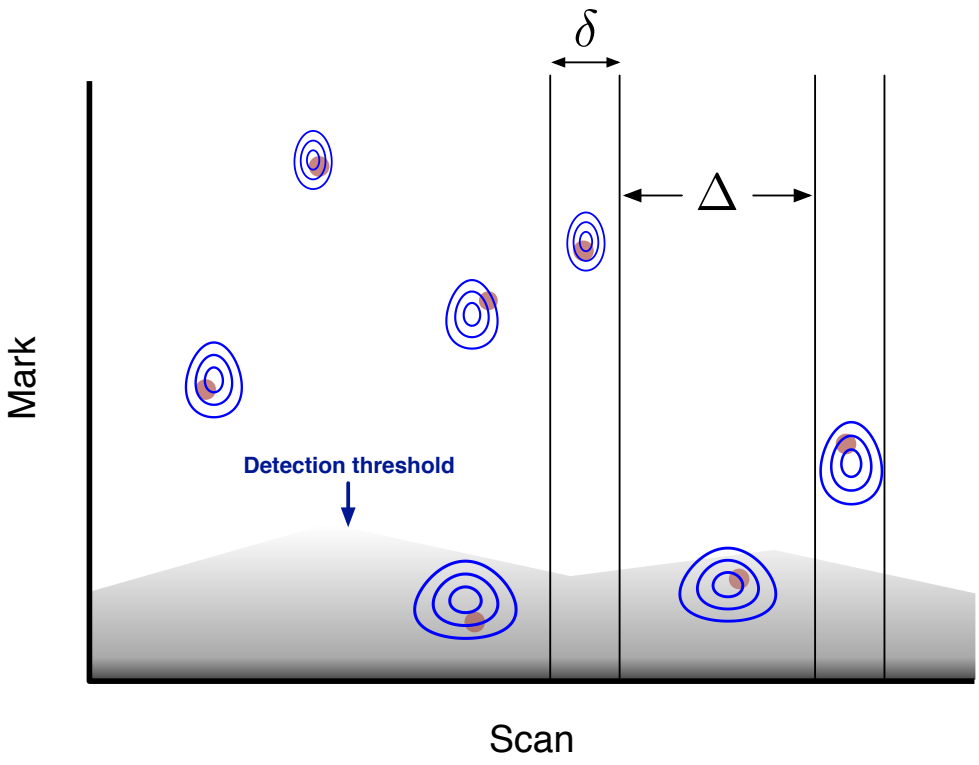

Figure B.1: Depiction of thinned latent marked point process model for catalog data produced by an astronomical survey. Object properties are split into a scanned subset and a mark subset. Dots (red) show latent (true) values for an object's properties. Contours (blue) depict member likelihood functions 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.

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 parameter, λ, describing the distribution of galaxies in x. We assume a luminosity PDF that is 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 a distance PDF, $h(r)$, and a conditional flux PDF, $\rho(F, r)$. We use $\boldsymbol{\zeta}$ to denote the flux PDF parameters, writing it as $\rho(F, r; \boldsymbol{\zeta})$ when we want to display the parameter dependence.

For the PDF for galaxy distance, $h(r)$, we assume homogeneity, which implies

$$h(r) = \begin{cases} \dfrac{3r^2}{r_u^3} & \text{if } 0 \leq r \leq r_u, \\ 0 & \text{otherwise,} \end{cases} \tag{B.1}$$

where r_u is an upper limit on distance chosen to be beyond the surveyed volume. (That is, r_u is chosen so that the most luminous galaxies of interest 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.) 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.[1] Formally, denoting the image data for detected galaxy i by D_i, we are writing

$$p(D_i|x, F, r) = \ell_i(F)\, m_i(x)\, \delta(r - r_i), \tag{B.2}$$

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., arrangements of counts in the pixels in a scanned aperture) that would be pass detection criteria. 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 more compact summary of the detection criteria will suffice.

We now compute the likelihood function for the parameters, based on catalog data describing member likelihood functions and the selection effects. For simplicity, we here consider "pure catalog" settings with stringent detection criteria (e.g., high thresholds), so that it is unlikely there are any false detections in the catalog (it is straightforward to generalize to settings with nonnegligible false detection rate). Fig. B.1 includes depictions

[1] Independence of flux and position estimates is almost universally assumed for astronomical catalogs, but for dim sources there can be significant dependence. We depict this in Fig. B.1.

of elements of our construction. 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.[2] 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,

$$\eta(x, F) \equiv p(D \in \mathcal{D}_x | F) \tag{B.3}$$

$$= p(\hat{F}(D) > F_{\rm th}(x) | F), \tag{B.4}$$

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

$$p_x(\boldsymbol{\zeta}) = \int \mathrm{d}r \int \mathrm{d}F \, \rho(F, r) \, h(r) \, \eta(x, F). \tag{B.5}$$

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

$$\begin{aligned} q(x) &= \sum_{\nu=0}^{\infty} \frac{(\lambda \delta x)^\nu}{\nu!} e^{-\lambda \delta x} \left[1 - p_x(\boldsymbol{\zeta})\right]^\nu \\ &= e^{-\lambda \delta x} \sum_{\nu=0}^{\infty} \frac{(\lambda \delta x)^\nu \left[1 - p_x(\boldsymbol{\zeta})\right]^\nu}{\nu!} \\ &= \exp\left[-\lambda \delta x p_x(\boldsymbol{\zeta})\right]. \end{aligned} \tag{B.6}$$

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.

[2] 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 exponents add, so that the nondetection probability becomes

$$q(\Delta_j) = \exp\left[-\lambda \int_{\Delta_j} \mathrm{d}x \int \mathrm{d}r \int \mathrm{d}F\, \eta(x,F)\, h(r)\, \rho(F,r)\right]. \tag{B.7}$$

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

$$p_1(\lambda, \boldsymbol{\zeta}) = (\lambda\delta)e^{-\lambda\delta}\left[\frac{dx}{\delta}\, h(r)dr\, \rho(F,r;\boldsymbol{\zeta})dF\right] p(D_i \in \mathcal{D}_x, D_i|x,F,r). \tag{B.8}$$

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

$$p(D_i \in \mathcal{D}_x, D_i|x,F,r) = p(D_i|x,F,r)\, p(D_i \in \mathcal{D}_x|D_i), \tag{B.9}$$

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 passing the detection criteria or not. But for a detected object, by definition the data passed the criteria, so the last factor is equal to unity. The first factor we recognize as the member likelihood function, defined in (B.2). 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:

$$\begin{aligned} p_\nu = & \frac{(\lambda\delta)^\nu}{\nu!} e^{-\lambda\delta} \\ & \times \left(\frac{dx}{\delta}\, h(r) dr\, \rho(F, r; \boldsymbol{\zeta}) dF \right) \ell_i(F)\, m_i(x)\, \delta(r - r_i) \\ & \times [1 - p_x(\boldsymbol{\zeta})]^{\nu-1} \\ & \times \nu. \end{aligned} \tag{B.10}$$

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 (B.10) as follows:

$$\begin{aligned} p_\nu = & (\lambda\delta)\, e^{-\lambda\delta} \left(\frac{dx}{\delta}\, h(r) dr\, \rho(F, r; \boldsymbol{\zeta}) dF \right) \ell_i(F)\, m_i(x)\, \delta(r - r_i) \\ & \times \frac{1}{(\nu - 1)!} (\lambda\delta)^{\nu-1} [1 - p_x(\boldsymbol{\zeta})]^{\nu-1}\,. \end{aligned} \tag{B.11}$$

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

$$p(D_i|\lambda, \boldsymbol{\zeta}) = q(\delta_i)\, h(r_i)\, (\lambda\delta) \left[\int_{\delta_i} \frac{\mathrm{d}x}{\delta}\, m_i(x) \right] \left[\int \mathrm{d}F\, \rho(F, r_i; \boldsymbol{\zeta})\, \ell_i(F) \right], \tag{B.12}$$

where $q(\delta_i)$ is an exponential of an integral, the same function appearing in the nondetection probability of (B.7).

The likelihood function is the product of detection probabilities (B.12) and nondetection probabilities (B.7) for all of the δ_i and Δ_j intervals. All of these probabilities share an exponential factor resembling (B.7). 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:

$$\lambda \int_\Omega \mathrm{d}x \int \mathrm{d}r \int \mathrm{d}F\, \eta(x, F)\, h(r)\, \rho(F, r; \boldsymbol{\zeta}), \tag{B.13}$$

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*,

$$\overline{\eta}(F) \equiv \frac{1}{\Omega} \int_\Omega \mathrm{d}x\, \eta(x, F). \tag{B.14}$$

Using this, (B.13) can be written as a two-dimensional integral,

$$(\lambda\Omega) \int \mathrm{d}r \int \mathrm{d}F\, \overline{\eta}(F)\, h(r)\, \rho(F, r; \boldsymbol{\zeta}). \tag{B.15}$$

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 (B.15) 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

$$\begin{aligned} \mathcal{L}(\lambda, \boldsymbol{\zeta}) = \lambda^N \exp\left[-(\lambda\Omega) \int \mathrm{d}r \int \mathrm{d}F\, \overline{\eta}(F)\, h(r)\, \rho(F, r; \boldsymbol{\zeta})\right] \\ \times \prod_{i=1}^{N} h(r_i) \int \mathrm{d}F\, \rho(F, r_i; \boldsymbol{\zeta})\, \ell_i(F). \end{aligned} \tag{B.16}$$

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 the 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; see the text after (B.9). 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 in two places: in the factor in front, 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

$$\mathcal{L}_m(\boldsymbol{\zeta}) = \prod_{i=1}^{N} \int \mathrm{d}F\, \mu(F, r_i; \boldsymbol{\zeta})\, \ell_i(F), \tag{B.17}$$

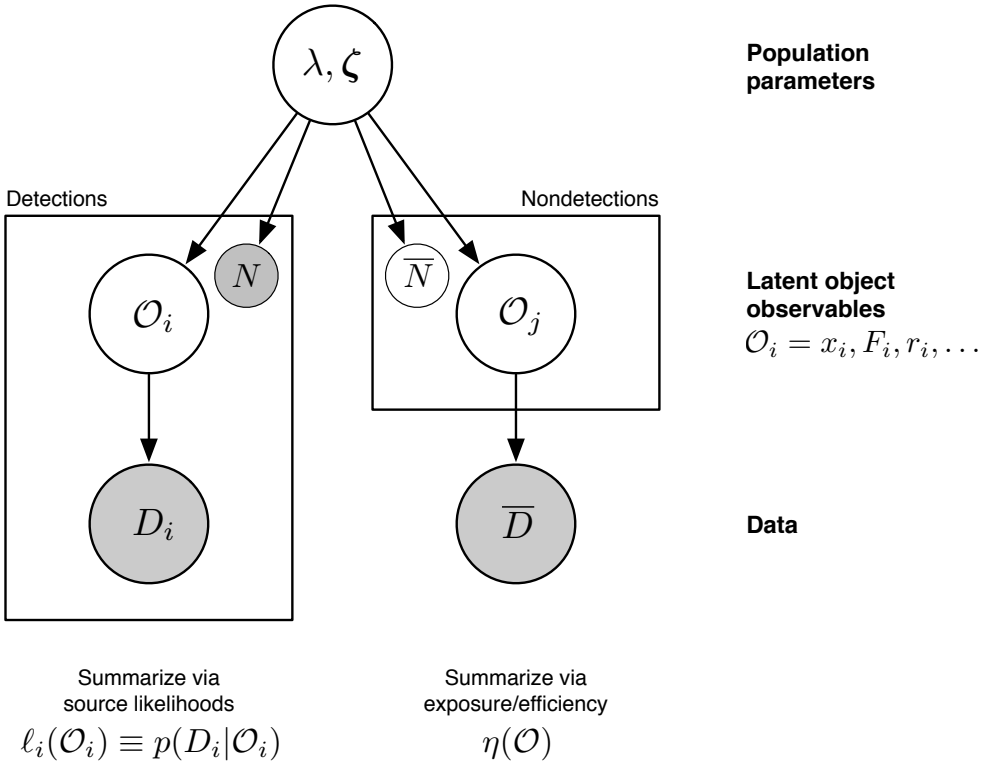

Figure B.2: Schematic DAG for a thinned latent marked point process model for luminosity function estimation from survey catalog data. The small N and $\overline{N}$ nodes specify the numbers of replications of the detection and nondetection plates, respectively.

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

$$\mu(F, r; \boldsymbol{\zeta}) \equiv \frac{h(r)\rho(F, r; \boldsymbol{\zeta})}{\int \mathrm{d}r \int \mathrm{d}F\, \overline{\eta}(F)\, h(r)\, \rho(F, r; \boldsymbol{\zeta})}. \tag{B.18}$$

Equation (B.17) 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 for the latent member parameters (F, r) (it does not integrate to unity); rather, it is a probability distribution for the data (up to proportionality).

The thinned latent Point process framework is a more complicated hierarchical model than those depicted in Fig. 2.2 (in the main text). Fig. B.2 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 (B.16) and (B.17) is essentially the same

as that for conditional density estimation with measurement error (the middle DAG in Fig. B.2). This is because the nondetection part of the DAG in Fig. B.2 corresponds to a product of exponentials, whose arguments sum in a single integral: the integral on the first line of the likelihood function in (B.16), and in the denominator of the effective density in (B.18). 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 is designed to sample from.

Appendix C: Exponential-cutoff break-by-one power law luminosity function

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*, a power law that is smoothly truncated at large luminosities by an exponential decay factor:

$$\phi(L;\theta) = \frac{A}{L_*}\left(\frac{L}{L_*}\right)^{\beta} e^{-L/L_*}, \tag{C.19}$$

where the parameters $\theta = (L_*, \beta, A)$ comprise a luminosity scale, L_*, a nominal mid-luminosity power law index, β, and an amplitude, A. (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.) The Schechter function follows from a basic physical model of galaxy formation via self-similar gravitational condensation (Press and Schechter 1974). Many astrophysical processes are approximately self-similar over a wide range of scales, leading to power law distributions. The index provides information about the scale-free behavior; upper and/or lower breaks provide information about key physical scales. Astronomers often seek to distill these kinds of salient features from the data.

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. C.19 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). Low-luminosity sources are unobservable (due to noise and background, discussed in the main text), 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 observable luminosities. For example, this is the case 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:

$$f(L;\theta) = \frac{C(\beta,u,l)}{u}\left(1 - e^{-L/l}\right)\left(\frac{L}{u}\right)^{\beta} e^{-L/u}, \tag{C.20}$$

where the normalization constant $C(\beta, u, l)$ is

$$C(\beta,u,l) = \begin{cases} \dfrac{1}{\Gamma(\beta+1)\cdot\left(1 - \frac{1}{\left(1+\frac{u}{l}\right)^{\beta+1}}\right)} & \text{if } \beta > -2 \text{ and } \beta \neq -1; \\ \dfrac{1}{\log\left(1+\frac{u}{l}\right)} & \text{if } \beta = -1. \end{cases} \tag{C.21}$$

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 $\beta > -2$. We generate samples from the BB1 distribution using a straightforward modification of a widely-used algorithm for sampling from the gamma distribution (Ahrens and Dieter 1974). These properties make it useful for simulation experiments.

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. 4.1 (in the main text) shows an example BB1 luminosity function, with $\beta = 1.5$, and $(l, u) = (1 \times 10^8, 1 \times 10^{10})$ in solar luminosity ($L_\odot$) 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

$$G(L) \equiv \frac{\mathrm{d}\log f}{\mathrm{d}\log L} = \frac{L}{f}\frac{\mathrm{d}f}{\mathrm{d}L} = g(L) + \beta - \frac{L}{u}, \tag{C.22}$$

with

$$g(L) = \frac{L}{l}\cdot\frac{1}{e^{L/l} - 1}. \tag{C.23}$$

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

$$F(L;\theta) = C(\theta)\left[\Gamma(\beta+1) - \gamma\left(\beta+1, L/u\right) - \frac{\Gamma(\beta+1) - \gamma\left(\beta+1, L\cdot\left(\frac{1}{u}+\frac{1}{l}\right)\right)}{\left(1+\frac{u}{l}\right)^{\beta+1}}\right], \quad \text{(C.24)}$$

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

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*

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.

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*

Press, W. H. & Schechter, P. (1974), 'Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation', *The Astrophysical Journal* **187**, 425–438.
URL: *http://adsabs.harvard.edu/abs/1974ApJ...187..425P*