Estimating the Luminosity Function in the presence of "Dark" sources

with a new method for statistical marginalisation

Siyang Li [1]
Supervisors: Prof. David van Dyk [1] and Dr. Maximilian Autenrieth [1]

[1] Imperial College London

April 24, 2024

Overview

Luminosity Functions with Dark Sources

Luminosity Functions with Dark Sources

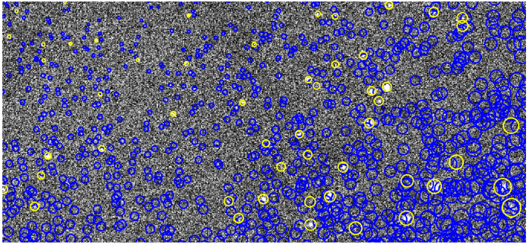
 [1]

Figure: X-ray sources in a part of Chandra Deep Field South. Yellow=sources detected in the X-ray catalogue, blue=optical sources.

Objective: to estimate the distribution of X-ray flux among sources.
Challenges:

1. Large number of X-ray sources with common characteristics.
2. Observed number of photon count Y_i contaminated by background.
3. Some sources are X-ray 'dark'.
4. Some source regions overlap.

[1] Image source: Autenrieth (2023)

Intro to some relevant astrophysical concepts

For each X-ray source i:

- Point spread function (PSF): specifies the radius for source region i which $\sim 90\%$ of the photons from source i will be observed.
- Source intensity λ_i(count/s/cm^2): (rescaled) expected source count from source i.
- Luminosity function: specifies the distributions of source intensities in a population.

Inference for such is formulated via $\mathcal{S}_i$, the number of photons from source i.

Problem: a large number of iid X-ray sources

- very large numbers of X-ray sources in populations
- sources can have independent intensities
- source intensities identicaly distributed

Solution: a Bayesian hierarchical model

Model structure for **source** intensity parameters $\boldsymbol{\lambda}$:

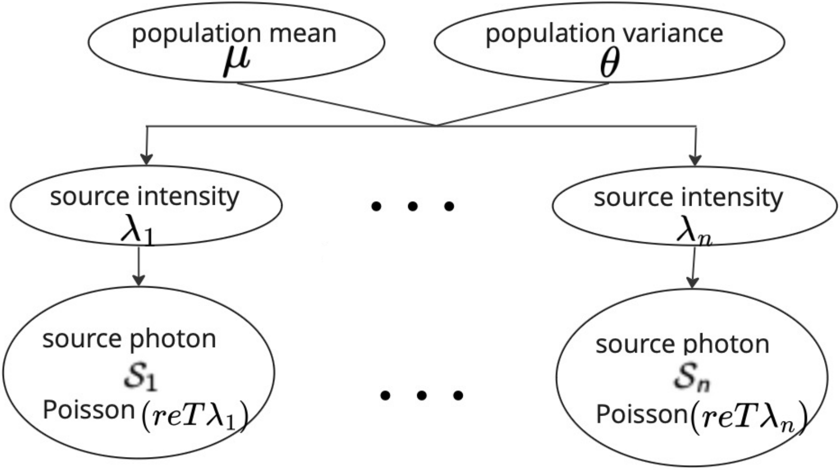


Figure: Hierarchical structure of the population of source intensity parameters.

Instrumental (deterministic) variables to account for

Likelihoods: [2]

$$(Y_i|\xi,\lambda_i) \overset{\text{indep}}{\sim} \text{Poisson}\left((a_i\xi + r_i e_i \lambda_i)\mathcal{T}\right)$$
$$(X|\xi) \sim \text{Poisson}(A\xi\mathcal{T})$$

- a_i(pixels): area of source region i.
- $\mathcal{T}$(s): exposure time for the pure background and source observations.
- r_i: proportion of photons from the source that are expected to fall in the source region.
- e_i(cm^2): telescope effective area at the source location.
- A(pixel): area of which the background count is collected.
- The location of each source.

[2] This work is a continuation from Wang et al. (2023).

Problem: non-homogeneous background contamination

Likelihoods:

$$(Y_i|\xi, \lambda_i) \overset{\text{indep}}{\sim} \text{Poisson}\left((a_i\xi + r_i e_i \lambda_i)\mathcal{T}\right)$$
$$(X|\xi) \sim \text{Poisson}(A\xi\mathcal{T})$$

1. The universe is 3-D, but telescopic images are 2-D.
2. Observed photons are background contaminated.
3. The background contamination is inhomogeneous (projected angle).
4. $\mathcal{B}_i$: the number of photons from background in source region i.
5. This makes $\mathcal{S}_i$ not directly observable.
6. We **only** observe the total photon counts in each source region i, $Y_i = \mathcal{S}_i + \mathcal{B}_i$.
7. **$\mathcal{S}_i$ and $\mathcal{B}_i$ are not directly observable! X and Y are observations.**

Previous solution: background subtraction

Consider $\mathcal{S}_i = Y_i - \mathcal{B}_i$.
When $\mathcal{B}_i$ is large but the source is faint - 'negative' $\mathcal{S}_i$?

New solution: background contamination parameters $\boldsymbol{\xi}$

Previous likelihoods:

$$(Y_i|\xi, \lambda_i) \overset{\text{indep}}{\sim} \text{Poisson}\left((a_i\xi + r_i e_i \lambda_i)\mathcal{T}\right)$$
$$(X|\xi) \sim \text{Poisson}(A\xi\mathcal{T})$$

1. Consider rates $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_K)$(count/s/pixel) for different background regions $k = 1, \cdots, K$ (depending on projected angles).
2. Observe pure backgrounds $\mathbf{X} = (X_1, \ldots, X_K)$:

$$X_k|\xi_k \overset{\text{indep}}{\sim} \text{Poisson}(A_k\xi_k\mathcal{T})$$

   to get information on $\boldsymbol{\xi}$.
3. Then latent variables $\mathcal{B}_i|\xi_k \overset{\text{indep}}{\sim} \text{Poisson}(a_i\xi_k\mathcal{T})$.

$\mathcal{S}_i$ and $\mathcal{B}_i$ are not directly observable! X and Y are observations.

Problem: X-ray ‘dark’ sources

- Weak X-ray sources are lost in the background.
- loads of such sources observed $\implies$ some X-ray photons detected
- a single such source is observed $\implies$ rare to detect X-ray photons
- It is possible that some optical sources don’t emit X-rays.

Solution: zero-inflated distributions

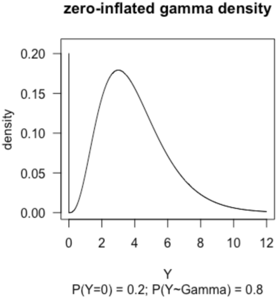


Figure: Zero-inflated gamma density

For the population of distributions for source intensities, with the proportion of dark sources being π_d,

$$\lambda_i|\mu, \theta, \pi_d \begin{cases} = 0 & \text{with probability } \pi_d, \\ \sim \text{Gamma}[\mu, \theta] & \text{with probability } 1 - \pi_d. \end{cases}$$

Problem: overlapping source regions

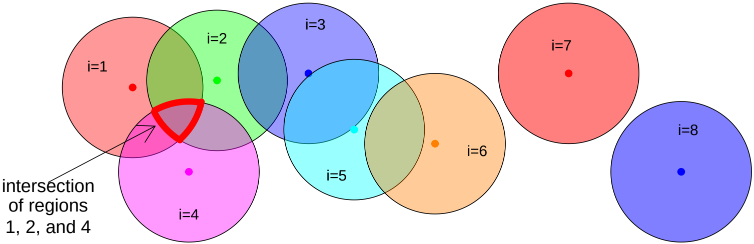


Figure: Overlapping sources.[3] The highlighted area is $s = \{1, 2, 4\}$.

1. The X-ray source regions overlap.
2. The source rates in intersections are not independent of each other.
3. We do not observe Y_i directly, but only Y_s for each segment s.

[3] image source: Wang et al. (2023)

Solution: adjustments in the likelihoods

Re-parametrise the likelihood as the following:

- The area of the segment, a_s(pixels);
- The effective area of the segment, e_s(cm^2);
- The expected proportion of photons from source $i \in s$ that are recorded in segment s, $r_{s,i}$.

Source counts per source per segment:

$$\mathcal{S}_{s,i}|\lambda_i \overset{\text{indep}}{\sim} \text{Poisson}(r_{s,i} e_s \lambda_i \mathcal{T})$$

Solution: adjustments in the likelihoods

Define $\rho_s := \sum_{i \in s} r_{s,i}\lambda_i$.
Observed counts per segment s if segment s is in the background region k:

$$Y_s = \sum_{i \in s} \mathcal{S}_{s,i} + \mathcal{B}_s \implies$$

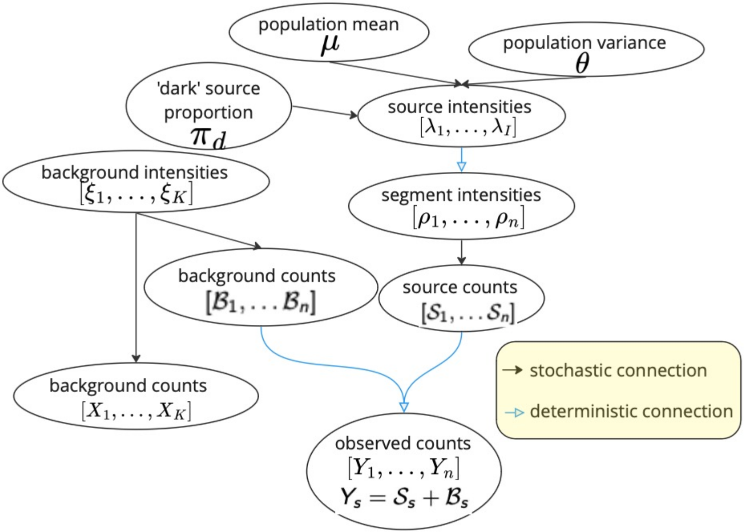
$$(Y_s | \xi_k, \boldsymbol{\lambda}) \overset{\text{indep}}{\sim} \text{Poisson}\left(\left(a_s\xi_k + \sum_{i \in s} r_{s,i} e_s \lambda_i\right)\mathcal{T}\right)$$

$$\overset{d}{=} \text{Poisson}\left(\left(a_s\xi_k + e_s\rho_s\right)\mathcal{T}\right)$$

Key features of the statistical model in use

1. *Large number of X-ray sources with common characteristics.* A Bayesian hierarchical model.
2. *Observed number of photon count Y_i contaminated by background.* Background intensity parameters $\boldsymbol{\xi}$.
3. *Some X-ray sources can be X-ray 'dark'.* Zero-inflated population distributions for source intensities $\boldsymbol{\lambda}$.
4. *Some source regions overlap.* Source intensity likelihood modified accordingly.

DAG of the statistical model



A simple simulation study

without overlapping sources and with homogeneous background

 4

Figure: NGC 2516 Southern Beehive

- Simulate each λ_i and Y_i to mimic Chandra observation of open cluster NGC 2516:
  - $\mathcal{T} = 5 \times 10^4$, $A = 2.5 \times 10^7$, $\xi = 2 \times 10^{-7}$
- Reduced cluster size: $n = 10$.
- dim(parameter)=14.
- Suppose true values: $\mu = 3 \times 10^{-4}, \theta = 2 \times 10^{-8}, \pi_d = 0.5, \xi = 2 \times 10^{-7}$.

[4] Image source: https://www.astrobin.com/full/hleuhx/0/

A simple simulation study

without overlapping sources and with homogeneous background

Figure: NGC 2516 Southern Beehive

Simulation steps for photon counts[5]:

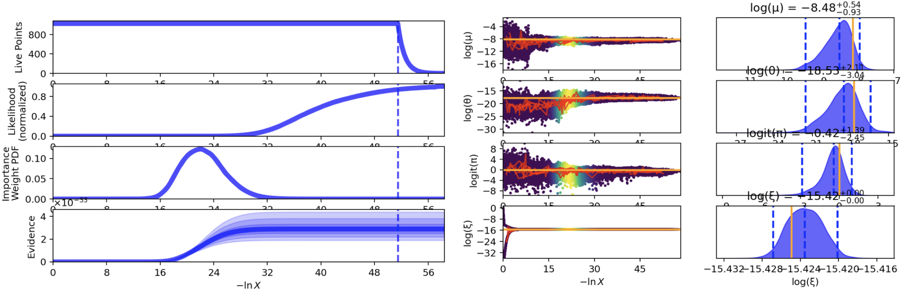
1. Simulate the background count $X \sim \text{Poisson}(A\xi\mathcal{T} = 2.5 \times 10^5)$.
2. Simulate $[re\mathcal{T}\lambda_1, \ldots, re\mathcal{T}\lambda_n]$ with $\lambda_i \overset{\text{indep}}{\sim}$ zero-inflated Gamma$[re\mathcal{T}\mu, (re\mathcal{T})^2\theta]$ with $p(\lambda_i = 0) = \pi_d$.
3. Set $\mathcal{B}_i \overset{\text{indep}}{\sim} \text{Poisson}(a\xi\mathcal{T})$, $\mathcal{S}_i \sim \text{Poisson}(re\mathcal{T}\lambda_i)$ and $Y_i = \mathcal{B}_i + \mathcal{S}_i$.

[5] as in Wang et al. (2023)

Nested sampling (full posterior) diagnostics

Dynesty (Koposov et al. (2023)) used. A NS on $(\mu, \theta, \pi_{\boldsymbol{d}}, \xi, \boldsymbol{\lambda})$.
Stopping criteria: posterior weight per iteration Dlogz $\leq 10^{-10}$.
Results from a typical run:

- 52723 iterations, 703 seconds.
- log marginal likelihood estimate: -74.92 ± 0.1394.



Nested sampling (full posterior) results

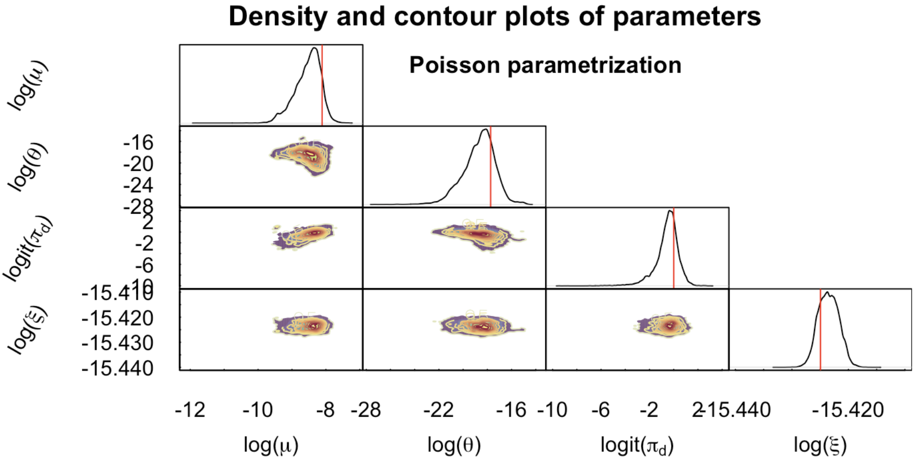


Figure: NS posterior samples with no overlapping sources.

Marginalising the population of source intensity parameters

- dim(parameter space) = n+4
- Marginalise out the population of parameters:

$$p(\mu, \theta, \pi_d, \xi \big| \mathbf{D}) = \int_{\mathbb{R}_+^n} p(\mu, \theta, \pi_d, \xi, \boldsymbol{\lambda} | \mathbf{D}) d\boldsymbol{\lambda}$$

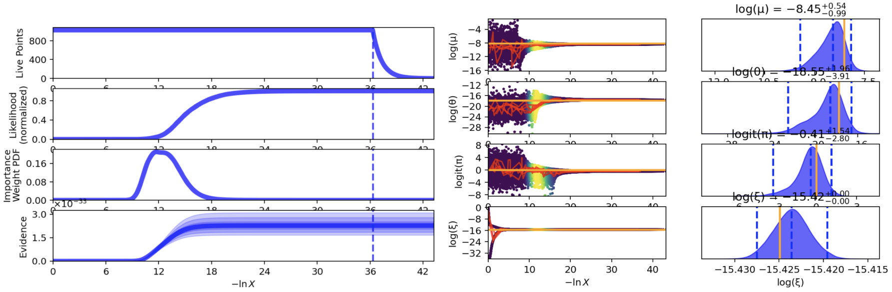
$$\propto p(\mu)p(\theta)p(\xi)p(\pi_d) \int_{\mathbb{R}_+^n} L(\xi, \boldsymbol{\lambda} | \mathbf{D}) p(\boldsymbol{\lambda} | \mu, \theta, \pi_d) d\boldsymbol{\lambda}$$

- pros: dim(parameter space) is fixed at 4, improves sampler efficiency.
- cons: no direct information on $\boldsymbol{\lambda}$ available. A second sampler is needed to infer $\boldsymbol{\lambda}$.

Nested sampling (marginal posterior) diagnostics

Stopping criteria: posterior weight per iteration Dlogz $\leq 10^{-10}$.
By using a new statistical marginalisation method (more on this later), I can construct a NS on $(\mu, \theta, \pi_d, \xi)$ **only**.
Results from a typical run:

- 37174 iterations, 135 seconds.
- log marginal likelihood estimate: -75.16 ± 0.1026.



Nested sampling (marginal posterior) results

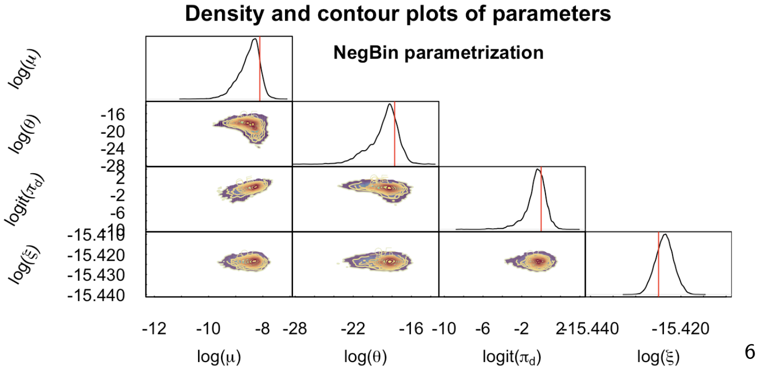


Figure: NS (negative-binomial parametrised) posterior samples under model without overlapping sources.

This density-and-contour plot is under the same scale as the previous density-and-contour plot.

[6] Gamma-Poisson mixing gives a negative binomial distribution.

A more complicated simulation study

with overlapping sources and nonhomogeneous background

Table: Background counts and average background counts per pixel in different regions in the Chandra/HRC-I observation of the open cluster NGC 2516.

Projected Angle	Count	Area (pixels)	Average count per pixel
0-6 (k=1)	219962	22029408	0.0100
6-8 (k=2)	146332	14093856	0.0104
8-16 (k=3)	285300	26448800	0.0108

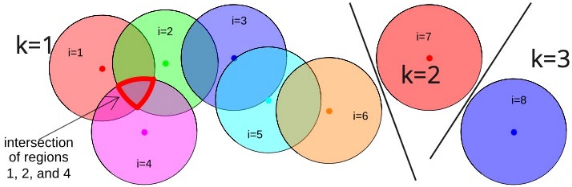


Figure: The overlap structure of sources used for simulation study [7]

[7] source of base picture and data: Wang et al. (2023)

A more complicated simulation study

with overlapping sources and nonhomogeneous background

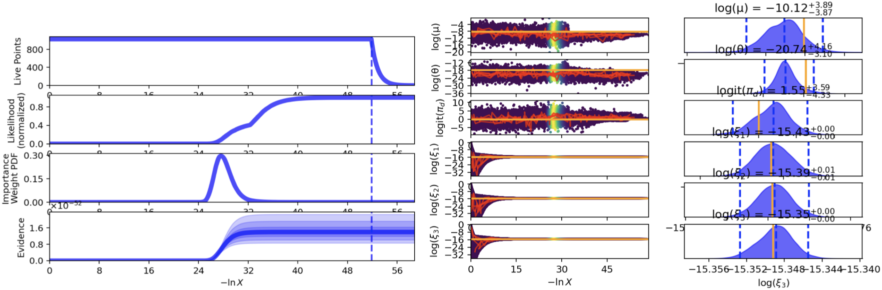
Suppose true values: $\mu = 3 \times 10^{-4}, \theta = 2 \times 10^{-8}, \pi_d = 0.5$.

1. Estimate $\hat{\boldsymbol{\xi}}$ using real data and mle: $\hat{\xi}_{\text{mle}} = \frac{X_k}{A_k \mathcal{T}}$.
2. Transform $\hat{\boldsymbol{\xi}}$ from per bkgd region (ξ_k) to per source segment (ξ_s).
3. Simulate $[\lambda_1, \dots, \lambda_n]$ from zero-inflated Gamma.
4. Set segment areas a_s, segment effective areas e_s, proportion of photons from source $r_{s,i}$.
5. Transform source intensity parameters from per source to per segment, $e\mathcal{T}\rho = e\mathcal{T}\sum_{i \in s} r_{s,i}\lambda_i$.
6. Simulate $\mathcal{B}_s \overset{\text{indep}}{\sim} \text{Poisson}(a_s \hat{\xi}_s \mathcal{T})$, $\mathcal{S}_s \sim \text{Poisson}(e\mathcal{T}\rho_s)$, $Y_s = \mathcal{B}_s + \mathcal{S}_s$.

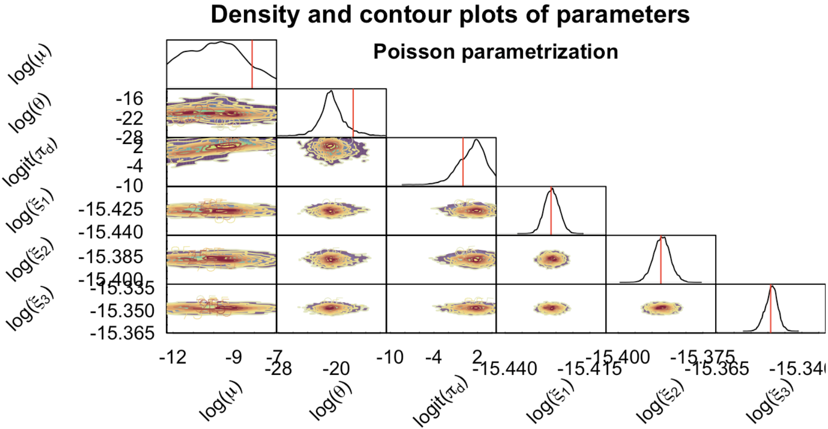
Nested sampling diagnostics

Stopping criteria: posterior weight per iteration Dlogz $\leq 10^{-10}$.
A NS on $(\mu, \theta, \pi_d, \boldsymbol{\xi}, \boldsymbol{\lambda})$, $\boldsymbol{\lambda}$ not marginalised out.
Results from a typical run:

- 53107 iterations, 791 seconds.
- log marginal likelihood estimate: -119.4 ± 0.1605.



Nested sampling results



This density-and-contour plot is under the same scale as the previous density-and-contour plot.

Conclusion and computational issues

- A sophisticated statistical model for astronomers' need.
- Possible to implement NS for model and obtain sensible inferences.
- Parameter-space dimension increases with number of sources / overlapping structure.
- The sampler / inference can run into trouble if there is too much overlap.

A general statistical marginalisation method is useful.

The statistical marginalisation method

The model marginalisation integral

From now on:

- $\boldsymbol{\xi}$ denotes hyperparameters;
- $\boldsymbol{\theta}$ denotes parameters;
- $\mathbf{y}$ denotes observations.

Bayes' formula for the full posterior:

$$p(\boldsymbol{\xi}, \boldsymbol{\theta} | \mathbf{y}) \propto p(\boldsymbol{\xi}) p(\boldsymbol{\theta} | \boldsymbol{\xi}) p(\mathbf{y} | \boldsymbol{\theta}).$$

Law of total probability:

$$p(\boldsymbol{\xi} | \mathbf{y}) = \int_{\Omega_{\boldsymbol{\theta}}} p(\boldsymbol{\xi}, \boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta}.$$

Combining the two:

$$p(\boldsymbol{\xi} | \mathbf{y}) \propto p(\boldsymbol{\xi}) \int_{\Omega_{\boldsymbol{\theta}}} p(\mathbf{y} | \boldsymbol{\theta}) p(\boldsymbol{\theta} | \boldsymbol{\xi}) d\boldsymbol{\theta} = p(\boldsymbol{\xi}) p(\mathbf{y} | \boldsymbol{\xi}), \tag{1}$$

A marginalised Bayes' formula.

Evaluating the model marginalisation integral

Facts:

$$p_{\text{Poisson}}(y|\theta) = \frac{\theta^y}{y!}e^{-\theta}, \text{ and } \frac{d^y}{dt^y}e^{t\theta} = \theta^y e^{t\theta}$$

and

$$M_\theta(t) = \mathbb{E}(e^{t\theta}), \text{ for suitable } t \in \mathbb{R}.$$

Derivatives of prior moment-generating function

with Poisson likelihoods, univariate, 1 observation

$$
\begin{aligned}
&\int_{\Omega_\theta} p(y|\theta)p(\theta|\xi)d\theta \\
=&\mathbb{E}_{\theta|\xi}[p(y|\theta)] \\
=&\frac{1}{y!}\mathbb{E}_{\theta|\xi}[\theta^y e^{-\theta}] \\
=&\frac{1}{y!}\mathbb{E}_{\theta|\xi}[\theta^y e^{t\theta}]\Big|_{t=-1} \\
=&\frac{1}{y!}\mathbb{E}_{\theta|\xi}\left[\frac{d^y}{dt^y}e^{t\theta}\right]\Bigg|_{t=-1} \\
=&\frac{1}{y!}\frac{d^y}{dt^y}\mathbb{E}_{\theta|\xi}\left[e^{t\theta}\right]\Bigg|_{t=-1} \\
=&\frac{1}{y!}\frac{d^y}{dt^y}M_{\theta|\xi}(t)\Bigg|_{t=-1}
\end{aligned}
$$

Facts:

$$p_{\text{Poisson}}(y|\theta) = \frac{\theta^y}{y!}e^{-\theta},$$

$$\frac{d^y}{dt^y}e^{t\theta} = \theta^y e^{t\theta}$$

and

$$M_\theta(t) = \mathbb{E}(e^{t\theta}), \text{ for suitable } t \in \mathbb{R}.$$

mgf-marginalisation with Poisson likelihoods

Theorem (mgf-marginalisation (Poisson likelihood))

Let the length of $\boldsymbol{\theta}$ be $n \in \mathbb{R}$. For $i \in \{1, 2, \ldots, n\}$, suppose each θ_i is the parameter for one y_i.
Suppose the likelihood is Poisson and the prior mgf exists and satisfies $M_{\boldsymbol{\theta}|\boldsymbol{\xi}}(-\mathbf{1}) < \infty$. Then the model marginalisation integral is given by

$$p(\mathbf{y}|\boldsymbol{\xi}) = \frac{1}{y_1! \cdots y_n!} \frac{\partial^{\sum_{s=1}^{n} y_s}}{\partial t_1^{y_1} \cdots \partial t_n^{y_n}} M_{\boldsymbol{\theta}|\boldsymbol{\xi}}(\mathbf{t}) \Big|_{\mathbf{t}=-\mathbf{1}}.$$

This is the result used for marginalising source intensity parameters with no overlapping sources.
Without zero-inflation, here $\mathbf{y}|\boldsymbol{\xi}$ is negative binomial (easy check).

Moment generating function for zero-inflated gamma

$$
\begin{aligned}
& M_{\lambda_i}(t) \\
=& \mathbb{E}[e^{t\lambda_i}] \\
=& \pi_d e^0 + (1-\pi_d)\mathbb{E}_{\text{Gamma}}(e^{t\lambda_i}) \\
=& \pi_d + (1-\pi_d) M_{\lambda_i}^{\text{Gamma}}(t) \\
=& \pi_d + (1-\pi_d)\left(\frac{\beta}{\beta - t}\right)^{\alpha},
\end{aligned}
$$

Moment generating function for transformed parameters

Recall $\rho_s = \sum_{i \in s} r_{s,i} \lambda_i$.

$$M_{\boldsymbol{\lambda}}(\mathbf{t}) = \mathbb{E}(e^{\mathbf{t}^\intercal \boldsymbol{\lambda}}) = \mathbb{E}(e^{\sum_{i=1}^{I} t_i \lambda_i}) = \prod_{i=1}^{I} \mathbb{E}(e^{t_i \lambda_i}) = \prod_{i=1}^{I} M_{\lambda_i}(t_i). \implies$$

$$M_{\boldsymbol{\rho}}(\boldsymbol{\zeta}) = \mathbb{E}(e^{\boldsymbol{\zeta}^\intercal \boldsymbol{\rho}}) = \mathbb{E}(e^{\boldsymbol{\zeta}^\intercal \mathbf{r} \boldsymbol{\lambda}}) = M_{\boldsymbol{\lambda}}((\boldsymbol{\zeta}^\intercal \mathbf{r})^\intercal) = \prod_{i=1}^{I} M_{\lambda_i}((\boldsymbol{\zeta}^\intercal \mathbf{r})_i)$$

mgf-marginalisation with Poisson likelihoods

Corollary

Suppose $\boldsymbol{\lambda} := \mathbf{r}\boldsymbol{\theta}$*, where* $\mathbf{r} \in \mathbb{R}^{m \times n}$ *is a linear transformation of the independent parameters* $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_n)$*, and* $m \geq n$*,* $m \in \mathbb{R}$*, and suppose each* λ_j *is the parameter for one* y_j *for* $j \in \{1, 2, \ldots, m\}$ *with a Poisson likelihood. Suppose the prior mgf exists and satisfies* $M_{\theta_i|\boldsymbol{\xi}}((-\boldsymbol{\zeta}^\intercal \mathbf{r})_i) < \infty$ *for each* $i \in \{1, 2, \ldots, n\}$*. Then*

$$p(\mathbf{y}|\boldsymbol{\xi}) = \frac{1}{y_1! \cdots y_n!} \left[\prod_{s=1}^{m} \zeta_s^{y_s} \right] \frac{\partial^{\sum_{s=1}^{m} y_s}}{\partial t_1^{y_1} \partial t_2^{y_2} \cdots \partial t_m^{y_m}} \prod_{i=1}^{n} M_{\theta_i|\boldsymbol{\xi}}((\mathbf{t}^\intercal \mathbf{r})_i) \Big|_{\mathbf{t}=-\boldsymbol{\zeta}}.$$

This is the result needed for marginalising source intensity parameters with overlapping sources.

Here $\mathbf{y}|\boldsymbol{\xi}$ is no longer as simple as negative binomial.

Exact calculations for marginal likelihoods

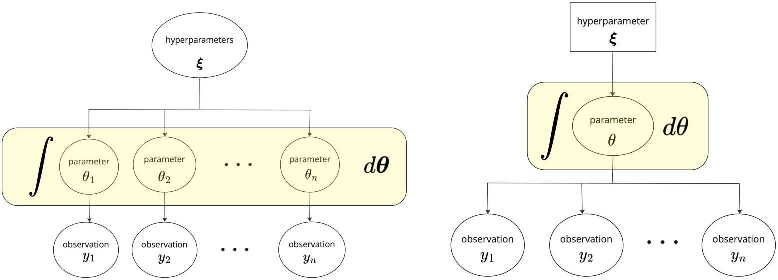


Figure: Hierarchical model marginalisation vs. marginal likelihood computation.

The model marginalisation integral $p(\mathbf{y}|\boldsymbol{\xi})$ is also a marginal likelihood (for sub-models in the hierarchical structure):

$$p(\boldsymbol{\theta}|\mathbf{y}, \boldsymbol{\xi}) = \frac{p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\xi})p(\boldsymbol{\theta}|\boldsymbol{\xi})}{p(\mathbf{y}|\boldsymbol{\xi})}.$$

Marginal likelihood computation with Poisson likelihoods

Theorem (mgf marginal likelihood calculation (Poisson likelihood))

Let the θ be the only parameter in the likelihood indexed by the independent sample $\mathbf{y}$ of length n.
Suppose the likelihood is Poisson. Furthermore, suppose the prior mgf exists and satisfies $M_{\theta|\boldsymbol{\xi}}(-n) < \infty$. Then the model marginalisation integral is given by

$$p(\mathbf{y}|\boldsymbol{\xi}) = \frac{1}{y_1! \cdots y_n!} \left(\frac{\partial}{\partial t} \right)^{\sum_{s=1}^{n} y_s} M_{\theta|\boldsymbol{\xi}}(t) \Big|_{t=-n}.$$

Extension to gamma likelihoods

Theorem (mgf-marginalisation (gamma likelihood))

Suppose $\boldsymbol{\beta} := \mathbf{r}\boldsymbol{\theta} > \mathbf{0}$*, where* $\mathbf{r} \in \mathbb{R}^{n\times n}$ *is a diagonal matrix that scales the independent parameters* $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_n) > \mathbf{0}$*. Suppose the likelihood is gamma. Suppose the prior mgf exists and satisfies* $M_{\theta_i|\boldsymbol{\xi}}((-\mathbf{r}^\intercal \mathbf{y})_i) < \infty$ *for each* $i \in \{1, 2, \ldots, n\}$*. Then if* $M_{\theta_i|\boldsymbol{\xi}} \in L^1[-\infty, u_i y_i]$ *and* $M_{\theta_i|\boldsymbol{\xi}} * K^{n-\alpha} \in W^{n,1}([-\infty, u_i y_i])$*,*

$$p(\mathbf{y}|\boldsymbol{\xi}) = \prod_{i=1}^{n} \frac{1}{\Gamma(\gamma_i)} \frac{\partial^{\langle \alpha_i \rangle + 1}}{\partial t_i^{\langle \alpha_i \rangle + 1}} \{\mathcal{M} L_{\theta_i|\boldsymbol{\xi}}\}(\gamma_i)\Big|_{t_i = -y_i},$$

where $L_{\theta_i|\boldsymbol{\xi}}(l) := M_{\theta_i|\boldsymbol{\xi}}(r_i(t_i - l))$ *is the moment generating function,* $\gamma = \langle \alpha_i \rangle + 1 - \alpha_i$ *is the fraction part in the fractional derivative, and* $\frac{\partial^{\alpha_i}}{\partial t_i^{\alpha_i}} = D_{z+}^{\alpha_i}$ *for* $z = -\infty$ *is the RL fractional derivative operator in use.* $\mathcal{M}$ *is the Mellin transform as defined in Equation (2.1) in Luchko and Kiryakova (2013).*

Bayesian interpretations on moments of distributions?

The moment-calculating equation

$$\mathbb{E}[\theta^k] = \frac{d^k}{dt^k} M_\theta(t)\bigg|_{t=0}$$

is just a special (or limiting) case of the mgf marginal likelihood calculation equation under Poisson likelihood

$$p(\mathbf{y}|\boldsymbol{\xi}) = \frac{1}{y_1! \cdots y_n!} \left(\frac{\partial}{\partial t}\right)^{\sum_{s=1}^{n} y_s} M_{\theta|\boldsymbol{\xi}}(t)\bigg|_{t=-n}.$$

(or the equivalent result under gamma likelihood).

Autenrieth, M. (2023). Principled bayesian modeling and statistical learning with non-representative data in astrophysics.

Koposov, S., Speagle, J., Barbary, K., Ashton, G., Bennett, E., Buchner, J., Scheffler, C., Cook, B., Talbot, C., Guillochon, J., Cubillos, P., Ramos, A. A., Johnson, B., Lang, D., Ilya, Dartiailh, M., Nitz, A., McCluskey, A., and Archibald, A. (2023). *joshspeagle/dynesty: v2.1.3*.

Luchko, Y. and Kiryakova, V. (2013). The mellin integral transform in fractional calculus. *Fractional calculus and applied analysis*, 16:405–430.

Wang, L., Kashyap, V. L., van Dyk, D. A., and Zeras, A. (2023). Bayesian methods for modeling source intensities. *[Manuscript in preparation]*.

DAG of the statistical model