Moment-generating Function methods in Estimations of the Luminosity Function

in the presence of "Dark" sources

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

[1]Imperial College London

February 05, 2025

Overview

1. Luminosity Functions with Dark Sources
2. The MGF method in Statistical Marginalisation
   - Transfer functions in Bayesian models and exact Bayesian inference
   - Evidence computation
   - Random Stopping-Bayesian equivalence
   - Marginal Likelihoods in GLMM

Luminosity Functions with Dark Sources

Luminosity Functions with Dark Sources

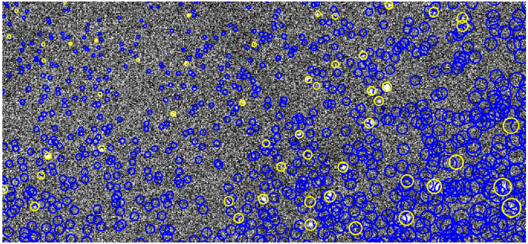


Figure: X-ray sources in a part of Chandra Deep Field South. Yellow=sources detected in the X-ray catalogue, blue=optical sources.

Objective: to **efficiently** estimate the distribution of X-ray flux among sources by statistically-marginalising out the intensity parameters.

Challenges:

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

mgf-marginalisation is currently the only analytical method.

Structure: a Bayesian hierarchical model

Source intensity $\underline{\lambda}_i$(count/s/cm^2): (rescaled) expected source count.
Very large number of iid X-ray sources $\underline{\boldsymbol{\lambda}}$:

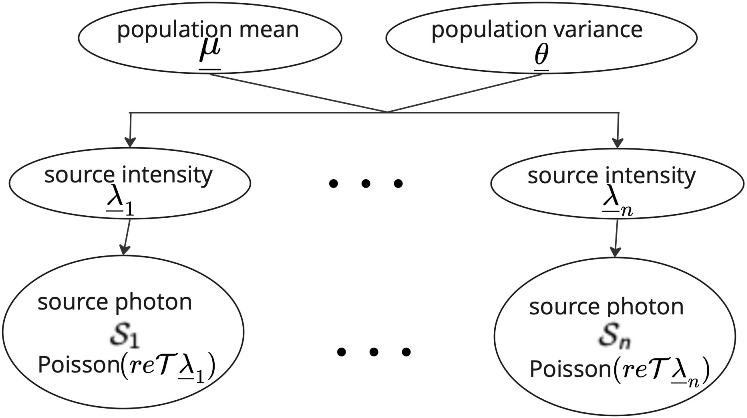


Figure: Hierarchical structure of the population of source intensity parameters.

Astronomical Concepts and Instrumental Variables

Likelihoods:

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

- Point spread function (PSF): radius for source region i which $\sim 90\%$ of the photons from source i will be observed.
- Luminosity function: distributions of source intensities in a population.

- $\underline{a}_i$: area of source region i.
- $\mathcal{T}$: 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: telescope effective area at the source location.
- A: area of which the background count is collected.
- The location of each source.

Problem: non-homogeneous background contamination

and the ill-behaved background subtraction

Likelihoods:

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

1. The universe is 3-D, but telescopic images are 2-D.
2. Observed photons are inhomogeneously background-contaminated.
3. $\mathcal{B}_i$: background count, $\mathcal{S}_i$: source count in source region i.
4. We **only** observe $Y_i = \mathcal{S}_i + \mathcal{B}_i$.
5. **$\mathcal{S}_i$ and $\mathcal{B}_i$ are not directly observable! X and Y are observations.**

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

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

Previous likelihoods:

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

1. Consider rates $\underline{\boldsymbol{\xi}} = (\underline{\xi}_1, \ldots, \underline{\xi}_K)$ for different background regions $k = 1, \cdots, K$.
2. Observe pure backgrounds $\mathbf{X} = (X_1, \ldots, X_K)$:

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

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

$\mathcal{S}_i$ 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

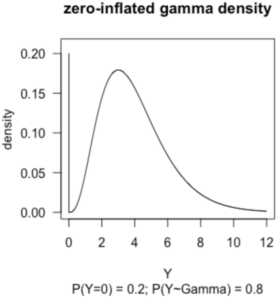


Figure: Zero-inflated gamma density

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

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

Problem: overlapping source regions

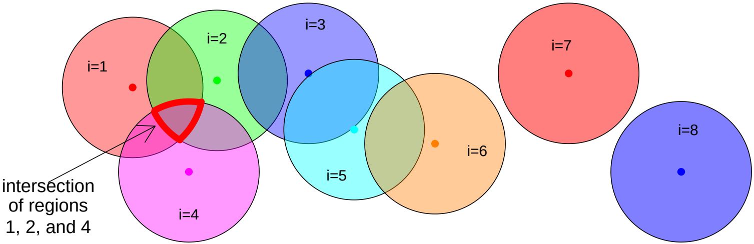


Figure: Overlapping sources.[1] The highlighted area is $s = \{1, 2, 4\}$.

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.

[1] image source: Wang et al. (2024)

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, $e_s(\text{cm}^2)$;
- 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:

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

Solution: adjustments in the likelihoods

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

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

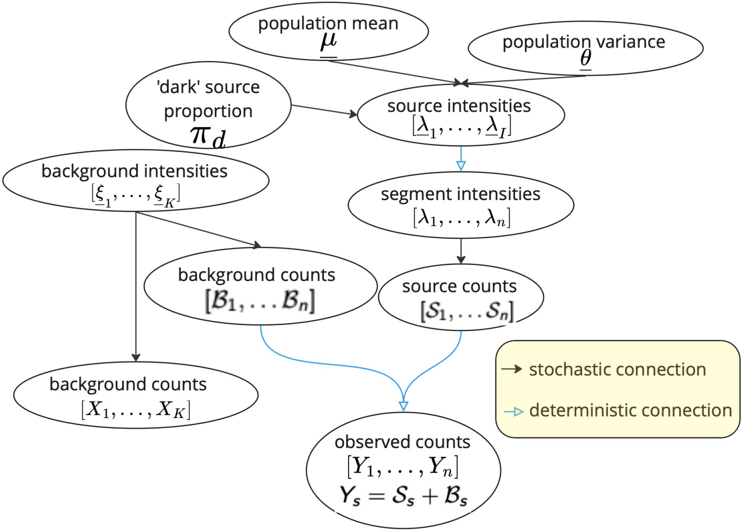
$$(Y_s | \underline{\xi}_k, \underline{\boldsymbol{\lambda}}) \overset{\text{indep}}{\sim} \text{Poisson}\left(\left(a_s \underline{\xi}_k + \sum_{i \in s} r_{s,i} e_s \underline{\lambda}_i \right) \mathcal{T} \right)$$

$$\overset{d}{=} \text{Poisson}\left(\left(a_s \underline{\xi}_k + e_s \lambda_s \right) \mathcal{T} \right)$$

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 $\underline{\boldsymbol{\xi}}$.
3. *Some X-ray sources can be X-ray 'dark'.* Zero-inflated population distributions for source intensities $\underline{\boldsymbol{\lambda}}$.
4. *Some source regions overlap.* Source intensity likelihood modified accordingly.

DAG of the statistical model



This model can be summarised in a nutshell.

Marginalising the population of source intensity parameters

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

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

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

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

The MGF method in Statistical Marginalisation

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:

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

The 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 theorem [Poisson likelihoods]

Theorem (mgf-marginalisation (Poisson likelihood))

Suppose $Y_i|\theta_i \overset{indep}{\sim} Poisson(\theta_i)$ *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).

mgf-marginalisation corollary [Poisson likelihoods]

Corollary

Suppose $\boldsymbol{\lambda} := \mathbf{r}\boldsymbol{\theta}$, *where* $\mathbf{r} \in \mathbb{R}^{m \times n}$ *and* $m \geq n$, $m \in \mathbb{R}$. *Suppose* $Y_j | \lambda_j \overset{indep}{\sim} \text{Poisson}(\lambda_j)$, *and 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.

Astro Example

Photon counts in overlapping sources

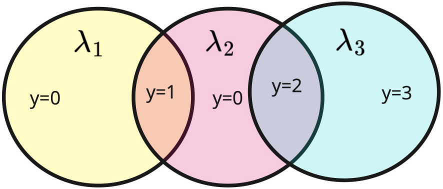


Figure: Examplar X-ray photon counts for three overlapping sources in each segment.

$$A = [A_{(1)}, A_{(2)}, A_{(3)}] = \begin{bmatrix} 0.1 & 0 & 0 \\ 0.9 & 0.1 & 0 \\ 0 & 0.1 & 0 \\ 0 & 0.8 & 0.1 \\ 0 & 0 & 0.9 \end{bmatrix}$$

Astro Example

Photon counts in overlapping sources

- An equivalent Poisson identity-link random-effect GLM:

$$\begin{aligned}\mu_i &= A_{(1)i}\lambda_1 + A_{(2)i}\lambda_2 + A_{(3)i}\lambda_3, \\ Y_i &\sim \text{Poisson}(\mu_i),\end{aligned}$$

- $\mathbf{t}^{\intercal} A = (0.1t_1 + 0.9t_2, 0.1t_2 + 0.1t_3 + 0.8t_4, 0.1t_4 + 0.9t_5)$.
- $p(\mathbf{Y} = (0, 1, 0, 2, 3) | \alpha = 4.5, \beta = 2) = 0.005745693$.
- **No other analytical methods available.**
- Hypothesis-testing simulation: 5694 out of the 10^6 iterations agree with $\mathbf{Y} = (0, 1, 0, 2, 3)$.
- Under H_0, the number of iterations having the simulated counts agreeing with $\mathbf{Y} = (0, 1, 0, 2, 3)$ follows binomial($n = 10^6, p = 0.005745693$), with a central 95% credible interval of $(5598, 5894)$.

Derivatives of prior moment-generating function

with gamma likelihoods, univariate, 1 observation

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

Facts: [a]

$$p_{\text{gamma}}(y|\theta) = \frac{\theta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\theta y},$$

$$\left(\frac{d}{dt}\right)^\alpha_{(-\infty)+} e^{\theta y} = \theta^\alpha e^{\theta y}$$

and

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

[a] $\left(\frac{d}{dt}\right)^\alpha_{(-\infty)+}$ is the Riemann-Liouville fractional derivative with a lower limit of $-\infty$.

mgf-marginalisation theorem [gamma likelihoods]

Theorem

Suppose $\boldsymbol{\theta} > \mathbf{0}$ *a.s., with* $Y_i \overset{indep}{\sim} \text{Gamma}(\alpha_i, \theta_i)$ *for some known* $\boldsymbol{\alpha} \in \mathbb{R}_+^n$*. Suppose the prior mgf exists and satisfies* $M_{\boldsymbol{\theta}|\boldsymbol{\xi}}(-\mathbf{y}) < \infty$*. Then*

$$p(\mathbf{y}|\boldsymbol{\xi}) = \left[\prod_{i=1}^{n} \frac{y_i^{\alpha_i - 1}}{\Gamma(\alpha_i)}\right] \frac{\partial^{\sum_{s=1}^{n} \alpha_s}}{\partial t_1^{\alpha_1} \partial t_2^{\alpha_2} \cdots \partial t_n^{\alpha_n}} M_{\boldsymbol{\theta}|\boldsymbol{\xi}}(\mathbf{t})\Big|_{\mathbf{t}=-\mathbf{y}}, \tag{2}$$

where $\frac{\partial^{\alpha_s}}{\partial t_s^{\alpha_s}} := D_{z+}^{\alpha_s}$ *is the RL fractional derivative of order* α_s *with the lower limit* $z = -\infty$*.*

mgf-marginalisation theorem [likelihood-specific]

Theorem (likelihood-specific mgf-marginalisation)

Suppose $Y|\theta,\mu \sim \mathcal{D}_l(\mu,\theta)$ *such that, for some* $s \in \mathbb{R}_+$, $\rho \in \mathbb{R}$ *and known* μ*, the likelihood function*

$$L(\theta; y|\mu) = f_0(y,\mu) \left(\frac{\partial}{\partial t}\right)_{k+}^{s} e^{h(y,\mu)(\rho+\theta)} \mathbb{1}[\theta \geq 0]\bigg|_{t=h(y,\mu)}, \qquad (3)$$

and some other conditions *hold. Then for fixed* μ*,*

$$p(y|\mu,\xi) = f_0(y,\mu) \left(\frac{\partial}{\partial t}\right)_{k+}^{s} e^{t\rho} M_{\theta|\xi}(t)\bigg|_{t=h(y,\mu)}. \qquad (4)$$

An overview of mgf-marginalisation theorems

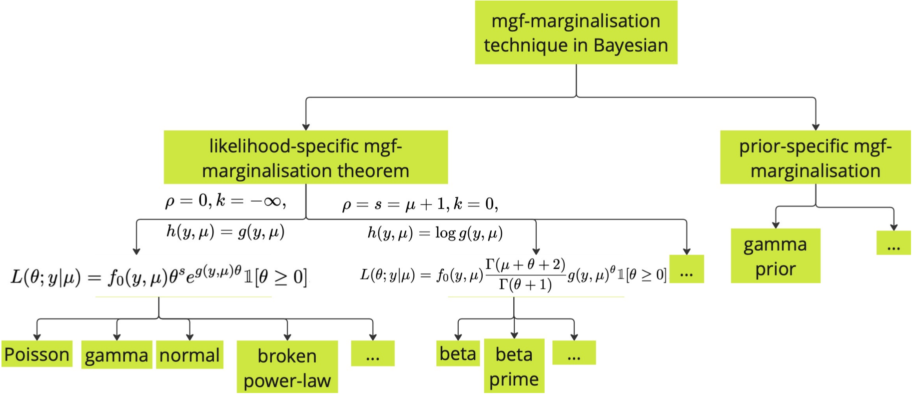


Figure: The use of mgf-marginalisation theorems in Bayesian inference

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

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

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

The source-intensity model in a nutshell

- Problem: the intensities for $\mathcal{S}_{s,i}$ have been marginalised, but we only observe $Y_s = \sum_{i \in s} \mathcal{S}_{s,i} + \mathcal{B}_s$.
- Solution: convolution in the sampler, or marginalise the intensities for $\mathcal{B}_s$ altogether:

$$\tilde{\mathbf{Y}} \sim \text{Poisson}(\mathcal{T}\tilde{\boldsymbol{\lambda}}), \tag{5}$$

where

$$\tilde{\boldsymbol{\lambda}} = \tilde{\mathbf{r}}\tilde{\boldsymbol{\theta}},\ \tilde{\mathbf{Y}} = \left[\begin{array}{c} \mathbf{Y} \\ \hline \mathbf{X} \end{array}\right],\ \tilde{\mathbf{r}} = \left[\begin{array}{c|c} \mathbf{er} & \underline{\mathbf{a}} \\ \hline \mathbf{0} & \mathbf{A} \end{array}\right],\ \tilde{\boldsymbol{\theta}} = \left[\begin{array}{c} \underline{\boldsymbol{\lambda}} \\ \hline \underline{\boldsymbol{\xi}} \end{array}\right],\ \underline{\boldsymbol{\lambda}} = \begin{bmatrix} \underline{\lambda}_1 \\ \vdots \\ \underline{\lambda}_I \end{bmatrix},$$

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix},\ \mathbf{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_K \end{bmatrix},\ \mathbf{e} = \begin{bmatrix} e_1 & 0 & \cdots & 0 \\ 0 & e_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & e_n \end{bmatrix},\ \underline{\mathbf{a}} = \begin{bmatrix} \underline{a}_1 & 0 & \cdots & 0 \\ 0 & \underline{a}_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \underline{a}_n \end{bmatrix},$$

The source-intensity model in a nutshell

$$\mathbf{r} = \begin{bmatrix} r_{1,1} & r_{1,2} & \cdots & r_{1,I} \\ r_{2,1} & r_{2,2} & \cdots & r_{2,I} \\ \vdots & \vdots & & \vdots \\ r_{n,1} & r_{n,2} & \cdots & r_{n,I} \end{bmatrix}, \mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 & \cdots & \mathbf{0} \\ \vdots & & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{A}_K \end{bmatrix}, \underline{\boldsymbol{\xi}} = \begin{bmatrix} \underline{\boldsymbol{\xi}}_1 \\ \vdots \\ \underline{\boldsymbol{\xi}}_K \end{bmatrix},$$

$$\mathbf{A}_k = \begin{bmatrix} A_k \\ 0 \\ \vdots \\ 0 \end{bmatrix}^\intercal, \underline{\boldsymbol{\xi}}_k = \begin{bmatrix} \underline{\xi}_k \\ \vdots \\ \underline{\xi}_k \end{bmatrix}$$

for all $k \in \{1, \ldots, K\}$, where the lengths of $\mathbf{A}_k$ and $\underline{\boldsymbol{\xi}}_k$ are determined by how many segments s are in the background region k.

- e.g. if there are 3 source segments in background region $k = 5$, then $\mathbf{A}_5 = [A_5, 0, 0]$ and $\underline{\boldsymbol{\xi}}_5 = [\underline{\xi}_5, \underline{\xi}_5, \underline{\xi}_5]^\intercal$.

An equivalent identity-link Poisson random-effect model for this model.

The remaining work for the source-intensity model

The remaining work to be done, for $m := n + K$ being the total length of the response vector $\tilde{\mathbf{y}}$ and $\boldsymbol{\zeta} = [\mathcal{T} \cdots \mathcal{T}]^\intercal$, is to find

$$
\begin{aligned}
& p(\tilde{\mathbf{y}} | \alpha, \beta, \pi_d) \\
= & \left[\prod_{j=1}^{m} \frac{1}{\tilde{y}_j!} t_j^{\tilde{y}_j} \right] \frac{\partial^{\sum_{j=1}^{m} \tilde{y}_j}}{\partial t_1^{\tilde{y}_1} \partial t_2^{\tilde{y}_2} \cdots \partial t_m^{\tilde{y}_m}} \\
& \left\{ \prod_{i=1}^{I} \left[\pi_d + (1 - \pi_d) \left(\frac{\beta}{\beta - (\mathbf{t}^\intercal \tilde{\mathbf{r}})_i} \right)^{\alpha} \right] \prod_{k=1}^{K} \left(\frac{\beta_\xi}{\beta_\xi - (\mathbf{t}^\intercal \tilde{\mathbf{r}})_{I+k}} \right)^{\alpha_\xi} \right\} \Bigg|_{\mathbf{t} = -\boldsymbol{\zeta}},
\end{aligned}
$$

where it is known that

$$
\beta_\xi = \frac{\frac{10^6}{A\mathcal{T}}}{\frac{10^{18}}{(A\mathcal{T})^2}} = 10^{-12} A\mathcal{T} \text{ and } \alpha_\xi = \frac{\left(\frac{10^6}{A\mathcal{T}} \right)^2}{\frac{10^{18}}{(A\mathcal{T})^2}} = 10^{-6}.
$$

Extension 1: with Tweedie's formula

- By specifying the marginal distribution instead of the prior, one can also find the posterior distribution.
- Tweedie's formula formulates this for exponential-family likelihood.

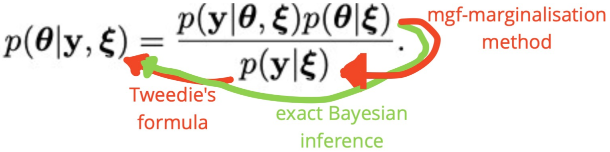


Figure: A method to achieve exact Bayesian inference in simple cases

Extension 1: for the posterior cumulant generating function

For the natural parameter η, suppose the likelihood is

$$L(\eta; y) = f_0(y)\exp[\eta y - \kappa(\eta)],$$

then Equation (2.4) in Efron (2011) gives the posterior cumulant generating function

$$K_{\eta|y}(t) = \kappa(y+t) - \kappa(y) = \log\left[\frac{p(y+t)}{f_0(y+t)}\right] - \log\left[\frac{p(y)}{f_0(y)}\right]$$

for the marginal density $p(y)$.

Extension 1: the Poisson-likelihood posterior moments

as a function of prior mgf

For Poisson likelihoods, $\eta = \log(\theta)$:

$$
\begin{aligned}
\mathbb{E}_\theta(\theta^t|y) =& M_{\eta|y}(t) \\
=& \frac{p(y+t) f_0(y)}{p(y) f_0(y+t)} \\
=& \frac{\left[\frac{1}{\Gamma(y+t+1)} \left(\frac{d}{dr}\right)^{y+t} M_\theta(r) \Big|_{r=-1} \right] \frac{1}{y!}}{\left[\frac{1}{y!} \left(\frac{d}{dr}\right)^{y} M_\theta(r) \Big|_{r=-1} \right] \frac{1}{\Gamma(y+t+1)}} \\
=& \frac{\left(\frac{d}{dr}\right)^{y+t} M_\theta(r) \Big|_{r=-1}}{\left(\frac{d}{dr}\right)^{y} M_\theta(r) \Big|_{r=-1}},
\end{aligned}
$$

Extension 2: Exact calculations for evidences

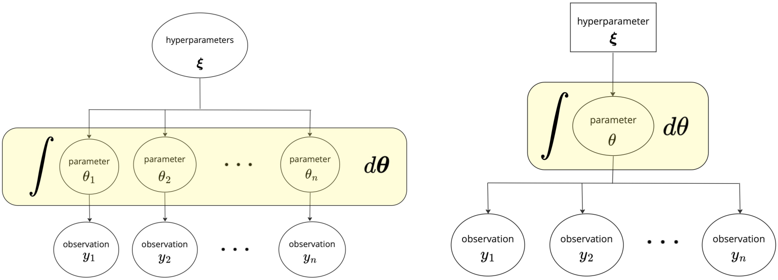


Figure: Hierarchical model marginalisation vs. evidence computation.

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

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

Extension 2: Evidence computation [Poisson likelihoods]

Theorem (mgf marginal likelihood calculation (Poisson likelihood))

Let $Y_i|\theta \overset{iid}{\sim}$ Poisson(θ). 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 3: random stopping-time models

A Poisson-process example:

- $T \sim \mathcal{D}_p$, the independent random stopping-time. 'Prior'.
- $N(T) \sim \text{Poisson}(\lambda T)$, the value of Poisson-process at random time T. 'Likelihood'.
- marginal distribution of N: 'evidence'.
- $T|N = n$: infer the random stopping time using the random stopping-time using the observation from Poisson process. 'Posterior'.
- Maximum likelihood: "find the optimal fixed stopping time";
- Bayesian: "infer the random stopping time".

e.g. Cox (1960) gives the analytical formula for the number of renewals in a gamma-length random interval, i.e. evidences for models with gamma priors.

Extension 4: marginal likelihood calculation in GLMMs

mgf methods requires linear transforms of parameters.
mgf methods can find marginal likelihoods in:

1. log-link Poisson HGLM: $\boldsymbol{\lambda} = \boldsymbol{\theta} e^{\mathbf{Xa}+\mathbf{b}}$;
2. identity-link Poisson GLMM and HGLM: $\boldsymbol{\lambda} = \mathbf{Xa} + \mathbf{b} + \mathbf{Z}\boldsymbol{\theta}$;
3. inverse-link gamma GLMM: $\boldsymbol{\beta} = \alpha\mathbf{Xa} + \alpha\mathbf{b} + \alpha\mathbf{Z}\boldsymbol{\theta}$;
4. inverse-identity-link gamma HGLM: $\boldsymbol{\beta} = \alpha\mathbf{Xa} + \alpha\mathbf{b} + \alpha\boldsymbol{\theta}$;
5. log-negative-log-link gamma HGLM: $\boldsymbol{\beta} = \alpha\boldsymbol{\theta} e^{-\mathbf{Xa}-\mathbf{b}}$;

mgf methods can not find marginal likelihoods in:

1. log-link Poisson GLMM: $\boldsymbol{\lambda} = e^{\mathbf{Xa}+\mathbf{Z}\boldsymbol{\theta}}$;
2. identity- and log-link gamma GLMM;
3. identity-, inverse- and log-link gamma HGLM.

Example 2

cake baking

- 3 recipes for making cakes;
- 15 batches of cake mix are made for each recipe, 45 batches in total;
- Each batch divided into 6 cakes baked at 6 different temperatures;
- 6 baking temperatures are 10°C apart from 175°C to 225°C;
- response: breaking angle of the cake;
- 270 observations in total.

random effect: replications of the cakes;
fixed effects: the temperature and recipe, with their interactions.

Example 2

cake baking

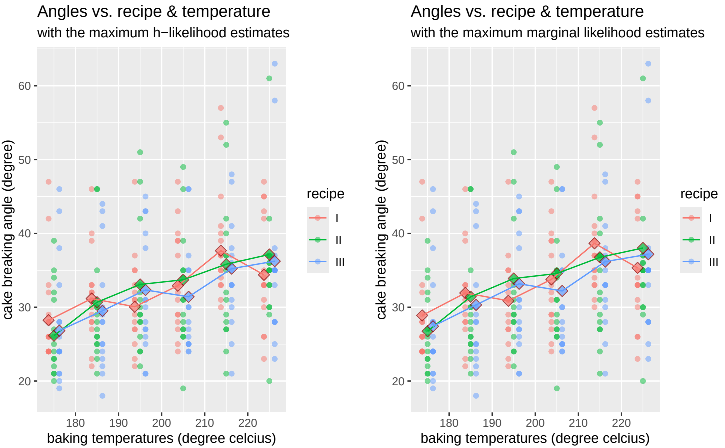


Figure: Comparison of the model fits using maximum h-likelihoods in Lee and Nelder (1996) with maximum marginal likelihoods via mgf.

Cox, D. (1960). On the number of renewals in a random interval. *Biometrika*, 47(3/4):449–452.

Efron, B. (2011). Tweedie's formula and selection bias. *Journal of the American Statistical Association*, 106(496):1602–1614.

Gaver, D. P. and O'Muircheartaigh, I. G. (1987). Robust empirical bayes analyses of event rates. *Technometrics*, 29(1):1–15.

Jordanova, P., Dušek, J., and Stehlík, M. (2013). Microergodicity effects on ebullition of methane modelled by mixed poisson process with pareto mixing variable. *Chemometrics and Intelligent Laboratory Systems*, 128:124–134.

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*.

Lee, Y. and Nelder, J. A. (1996). Hierarchical generalized linear models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(4):619–656.

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

Milgram, M. (1985). The generalized integro-exponential function. *Mathematics of computation*, 44(170):443–458.

Wang, L., Kashyap, V. L., van Dyk, D. A., and Zeras, A. (2024). Bayesian methods for modeling source intensities. *[Manuscript in preparation]*.

Appendix

Conditions for general mgf-marginalisation

Other conditions for the general-likelihood mgf-marginalisation to hold include:

1. the functions $f_0(y, \mu)$ and $h(y, \mu)$ are real and finite;
2. the prior distribution is $\theta|\xi \sim \mathcal{D}_p(\xi)$, such that the prior mgf $M_{\theta|\xi}(t)$ exists and is finite for all $t \in [c, d]$ where c, d are some real constants;
3. $M_{\theta|\xi}(t)$ is $\lceil s \rceil$-th order differentiable at $t = h(y, \mu)$;
4. $L(\theta; y|\mu) \geq 0$ is a continuous function of θ on $[0, \infty)$ for all values of $h(y, \mu) = t \in [c, d]$;
5. $p(y|\mu, \xi) < \infty$ exists for all values of $h(y, \mu) = t \in [c, d]$;
6. $\left(\frac{\partial}{\partial t}\right)_{k+}^{s}$ is the Riemann-Liouville fractional derivative with a lower limit of $k \in \mathbb{R} \cup \{-\infty, \infty\}$.

Example 3

Pump failure, gamma-prior hierarchical model

From Gaver and O'Muircheartaigh (1987). Number of pump failures y_i and the operating times t_i of pump i:

i	1	2	3	4	5	6	7	8	9	10
t_i	94.32	15.72	62.88	125.76	5.24	31.44	1.048	1.048	2.096	10.48
y_i	5	1	5	14	3	19	1	1	4	22

Table: Pump failure data

Model:

$$(\lambda_i|\alpha,\beta) \overset{\text{iid}}{\sim} \text{Gamma}(\alpha,\beta)$$

$$Y_i|\lambda_i \overset{\text{indep.}}{\sim} \text{Poisson}(\lambda_i t_i).$$

Example 3

Pump failure, gamma-prior hierarchical model

- Equivalent GLMM:

$$\log(\mathbb{E}(Y_i|\lambda_i)) = \log(\mu_i) = \tilde{\eta}_i = \log(t_i) + \log(\lambda_i),$$

where $Y_i|\lambda_i \overset{\text{indep.}}{\sim} \text{Poisson}(\mu_i)$, $\log(t_i)$: offsets, random effects: $\log(\lambda_i)$, no fixed effects.

-

$$\begin{aligned} p(\mathbf{y}|\alpha,\beta) &= \prod_{i=1}^{10} \frac{t_i^{y_i}}{y_i!} \left(\frac{\partial}{\partial s_i}\right)^{y_i} M_{\lambda_i|\alpha,\beta}(s_i)\bigg|_{s_i=-t_i} \\ &= \prod_{i=1}^{10} \frac{t_i^{y_i}}{y_i!} \frac{\Gamma(\alpha+y_i)}{\Gamma(\alpha)} \frac{\beta^\alpha}{(\beta+t_i)^{\alpha+y_i}}. \end{aligned}$$

- Empirical Bayesian [Gaver and O'Muircheartaigh 1987]: $\hat{\alpha} = 1.27$ and $\hat{\beta} = 0.82$, so $p(\mathbf{y}|\alpha = 1.27, \beta = 0.82) = 2.766569 \times 10^{-16}$.
- Verification: $(\lambda_i|\alpha,\beta) \sim \text{NegBin}(\alpha, \frac{\beta}{\beta+t_i})$, so $p(\mathbf{y}|\alpha = 1.27, \beta = 0.82) = 2.766569 \times 10^{-16}$.

Example 4

Pump failure, Pareto-prior non-hierarchical model

$$(\lambda|\alpha,\beta) \sim \text{Pareto}(\alpha, k),$$
$$y_i|\lambda \overset{\text{iid}}{\sim} \text{Poisson}(\lambda t_i).$$

- The exponential integral function [Milgram 1985]:
$$E_r(z) = z^{r-1}\Gamma(1-r, z),$$
where $\Gamma(1-r,z) = \int_z^\infty t^{-r}e^{-t}dt$.
-
$$\begin{aligned} p(\mathbf{y}|\alpha, k) &= \left[\prod_{i=1}^{10} \frac{t_i^{y_i}}{y_i!}\right] \left(\frac{d}{ds}\right)^{\sum_{i=1}^{10} y_i} M_{\lambda|\alpha,k}(s)\bigg|_{s=-\sum_{i=1}^{10} t_i} \\ &= \left[\prod_{i=1}^{10} \frac{t_i^{y_i}}{y_i!}\right] k^{75}\alpha E_{\alpha+1-75}\left(k\sum_{i=1}^{10} t_i\right) \\ &= [2.799194 \times 10^{48}]\alpha k^{75} E_{\alpha-74}(350.032k). \end{aligned}$$

Example 4

Pump failure, Pareto-prior non-hierarchical model

- Verification: Marginal density of mixed Poisson-Pareto [Jordanova et al. 2013]: $p(y(t)|\alpha, k) = \frac{\alpha(kt)^y}{y!} E_{\alpha-y+1}(kt)$.
-

$$p(\mathbf{y}|\alpha, k) = \prod_{i=1}^{10} p(y_i(t_i)|\alpha, k) = \alpha k^{\sum_{i=1}^{10} y_i} \prod_{i=1}^{10} \frac{t_i^{y_i}}{y_i!} E_{\alpha-y_i+1}(kt_i)$$

$$= \left[\prod_{i=1}^{10} \frac{t_i^{y_i}}{y_i!}\right] \alpha k^{\sum_{i=1}^{10} y_i} E_{\alpha+1-\sum_{i=1}^{10} y_i} \left(k \sum_{i=1}^{10} t_i\right)$$

$$= [2.799194 \times 10^{48}] \alpha k^{75} E_{\alpha-74}(350.032k),$$

Fractional derivatives

The derivative $\left(\frac{d}{dt}\right)^{\alpha}_{(-\infty)+}$ is of fractional order α.

- In general, fractional derivative operators are neither commutative nor additive:

$$\frac{\partial^{\alpha_1}}{\partial t^{\alpha_1}}\frac{\partial^{\alpha_2}}{\partial t^{\alpha_2}}f(t) \neq \frac{\partial^{\alpha_2}}{\partial t^{\alpha_2}}\frac{\partial^{\alpha_1}}{\partial t^{\alpha_1}}f(t) \text{ and } \frac{\partial^{\alpha_1}}{\partial t^{\alpha_1}}\frac{\partial^{\alpha_2}}{\partial t^{\alpha_2}}f(t) \neq \frac{\partial^{\alpha_1+\alpha_2}}{\partial t^{\alpha_1+\alpha_2}}f(t).$$

- We need to select the class of fractional derivatives that preserve the exchangability:

$$\frac{\partial^{\alpha_1}}{\partial t_1^{\alpha_1}}\frac{\partial^{\alpha_2}}{\partial t_2^{\alpha_2}}f(t) = \frac{\partial^{\alpha_2}}{\partial t_2^{\alpha_2}}\frac{\partial^{\alpha_1}}{\partial t_1^{\alpha_1}}f(t).$$

- For $\boldsymbol{\beta} = \mathbf{r}\boldsymbol{\theta}$ for $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$ independent parameters, if $\mathbf{r}$ is diagonal, then Riemann-Liouville (RL) derivatives preserve exchangability.

RL fractional derivatives

- RL fractional derivatives:

$$(D^{\alpha}_{u+}f)(x) = \frac{\partial^{(\langle\alpha\rangle+1)}}{\partial x^{(\langle\alpha\rangle+1)}}\frac{1}{\Gamma(\gamma)}\int_u^x (x-y)^{\gamma-1}f(y)dy,$$

  $\langle x\rangle$ is the largest integer strictly smaller than x,
  $\langle\alpha\rangle + 1 - \alpha =: \gamma \in [0,1)$ is the fractional part of differentiation.
- Initial condition $D^{\alpha_i}_{z+}e^{t_i\beta_i}\big|_{t_i=-y_i} = \beta_i^{\alpha_i}\exp[-\beta_i y_i]$ gives $z=-\infty$.
- $D^{\alpha_i}_{(-\infty)+}$ is the desired operator.

mgf-marginalisation corollary [gamma likelihoods]

Corollary

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*

$$p(\mathbf{y}|\boldsymbol{\beta}) = \prod_{i=1}^{n} \frac{\beta_i^{\alpha_i}}{\Gamma(\alpha_i)} y_i^{\alpha_i - 1} e^{-\beta_i y_i},$$

and the prior mgf exists and satisfies $M_{\theta_i|\boldsymbol{\xi}}((-\mathbf{t}^\intercal \mathbf{r})_i) < \infty$ *for each* $i \in \{1, 2, \ldots, n\}$*, if* $M_{\theta_i|\boldsymbol{\xi}}((\mathbf{t}^\intercal \mathbf{r})_i)$ *is continuous and differentiable up to the* $\langle \alpha_i \rangle + 1$*-th order at* $\mathbf{t} = -\mathbf{y}$*, then*

$$p(\mathbf{y}|\boldsymbol{\xi}) = \left[\prod_{i=1}^{n} \frac{y_i^{\alpha_i - 1}}{\Gamma(\alpha_i)}\right] \prod_{i=1}^{n} \frac{\partial^{\alpha_i}}{\partial t_i^{\alpha_i}} M_{\theta_i|\boldsymbol{\xi}}((\mathbf{t}^\intercal \mathbf{r})_i)\Big|_{\mathbf{t}=-\mathbf{y}},$$

where $\frac{\partial^{\alpha_s}}{\partial t_s^{\alpha_s}} := D_{z+}^{\alpha_s}$ *is the RL fractional derivative of order* α_s *with the lower limit* $z = -\infty$*.*

Calculating RL fractional derivatives

Remark

Under the same assumptions in Corollary 6, if $M_{\theta_i|\boldsymbol{\xi}} \in L^1[-\infty, -r_{ii}y_i + \epsilon_i]$ *and* $M_{\theta_i|\boldsymbol{\xi}} * K^{n-\alpha} \in W^{n,1}([-\infty, -r_{ii}y_i + \epsilon_i])$ *for some* $\epsilon_i > 0$*,*

$$p(\mathbf{y}|\boldsymbol{\xi}) = \left[\prod_{i=1}^{n} \frac{y_i^{\alpha_i - 1}}{\Gamma(\alpha_i)}\right] \prod_{i=1}^{n} \frac{1}{\Gamma(\gamma_i)} \frac{\partial^{\langle\alpha_i\rangle+1}}{\partial t_i^{\langle\alpha_i\rangle+1}} \{\mathcal{M} Q_{\theta_i|\boldsymbol{\xi}}\}(\gamma_i)\Big|_{t_i=-y_i}, \tag{6}$$

where $Q_{\theta_i|\boldsymbol{\xi}}(l) := M_{\theta_i|\boldsymbol{\xi}}(r_i(t_i - l))$ *is the moment-generating function for* $l_i := t_i - x$, $\gamma_i = \langle\alpha_i\rangle + 1 - \alpha_i$ *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 defined in Equation (2.1) in Luchko and Kiryakova (2013).*

mgf-marginalisation corollary [ineger-shape gamma GLMM]

Corollary

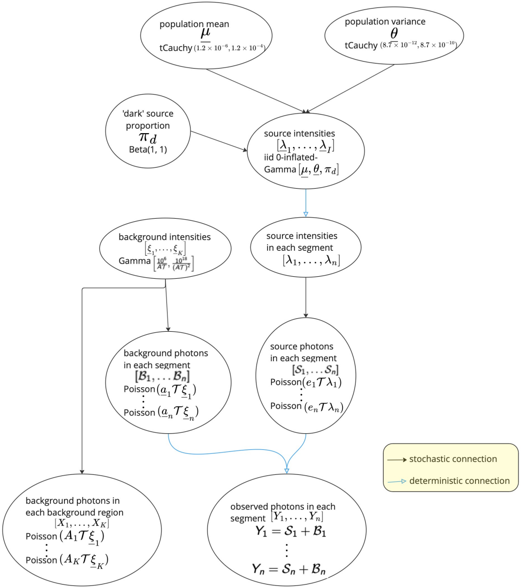
Suppose $\boldsymbol{\alpha}_j \in \mathbb{N}_0^m$. *Suppose* $\boldsymbol{\beta} := \mathbf{r}\boldsymbol{\theta} > \mathbf{0}$ *is a linear transformation of the independent parameters* $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_n) > \mathbf{0}$, *with* $\mathbf{r} \in \mathbb{R}^{m\times n}$ *for* $m \geq n$, *and suppose* $Y_j|\beta_j \overset{indep.}{\sim} Gamma(\alpha_j, \zeta_j\beta_j)$,

$$p(\mathbf{y}|\boldsymbol{\beta}) = \prod_{j=1}^{m} \frac{(\zeta_j\beta_j)^{\alpha_j}}{\Gamma(\alpha_j)} y_j^{\alpha_j - 1} e^{-\zeta_j\beta_j y_j},$$

where $\boldsymbol{\zeta} \in \mathbb{R}^m$ *are known constants, and the prior mgf exists and satisfies* $M_{\theta_i|\boldsymbol{\xi}}((-\mathbf{t}^\intercal\mathbf{r})_i) < \infty$ *for each* $i \in \{1, 2, \ldots, n\}$, *and if* $M_{\theta_i|\boldsymbol{\xi}}((\mathbf{t}^\intercal\mathbf{r})_i)$ *is continuous and differentiable up to the appropriate order at* $-((\mathbf{y}\odot\boldsymbol{\zeta})^\intercal\mathbf{r})_i$, *then*

$$p(\mathbf{y}|\boldsymbol{\xi}) = \left[\prod_{j=1}^{m} \frac{y_j^{\alpha_j-1}\zeta_j^{\alpha_j}}{\Gamma(\alpha_j)}\right] \frac{\partial^{\sum_{j=1}^m \alpha_j}}{\partial t_1^{\alpha_1}\partial t_2^{\alpha_2}\cdots\partial t_m^{\alpha_m}} \prod_{i=1}^{n} M_{\theta_i|\boldsymbol{\xi}}((\mathbf{t}^\intercal\mathbf{r})_i)\Bigg|_{\mathbf{t}=-\mathbf{y}\odot\boldsymbol{\zeta}}.$$

DAG of the statistical model



A simple simulation study

without overlapping sources and with homogeneous background

Figure: NGC 2516 Southern Beehive

Simulation steps for photon counts[2]:

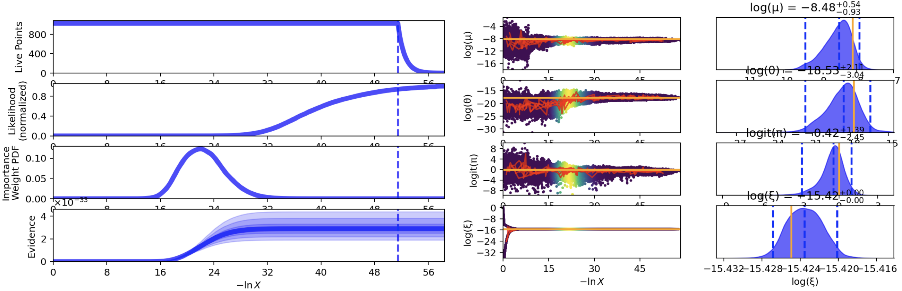
1. Simulate the background count $X \sim \text{Poisson}(A\xi\mathcal{T} = 2.5 \times 10^5)$.
2. Simulate $[re\mathcal{T}\lambda_1, \ldots, re\mathcal{T}\lambda_n]$ with $\lambda_i \overset{\text{indep}}{\sim}$ zero-inflated Gamma$[re\mathcal{T}\mu, (re\mathcal{T})^2\theta]$ with $p(\lambda_i = 0) = \pi_d$.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

[2] as in Wang et al. (2024)

Nested sampling (full posterior) diagnostics

Dynesty (Koposov et al. (2023)) used. A NS on $(\mu, \theta, \pi_d, \xi, \boldsymbol{\lambda})$.
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

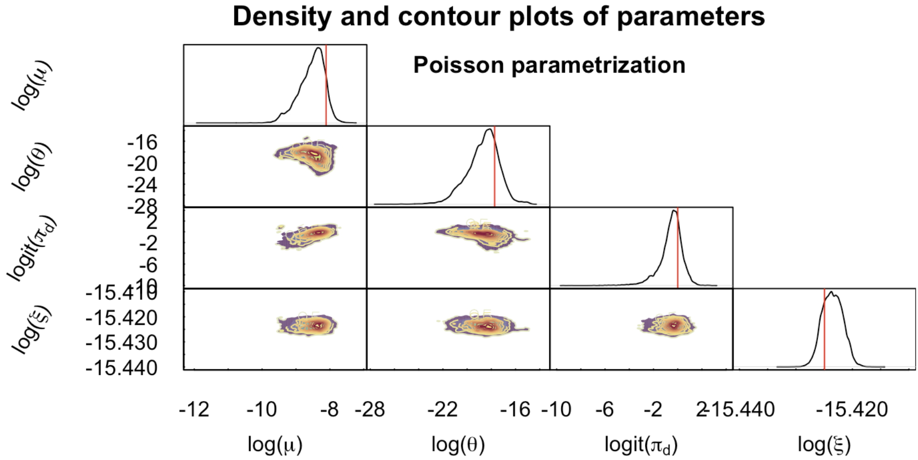
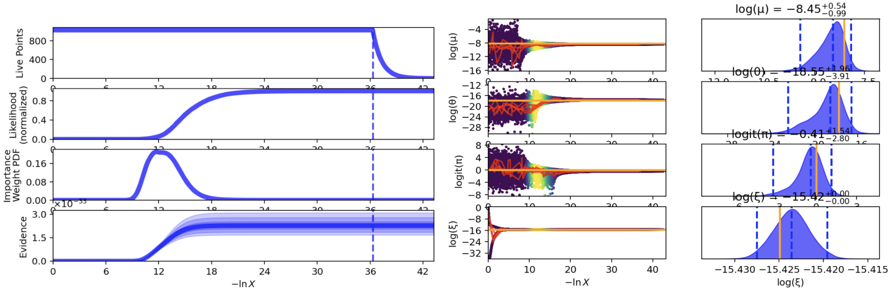


Figure: NS posterior samples with no overlapping sources.

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

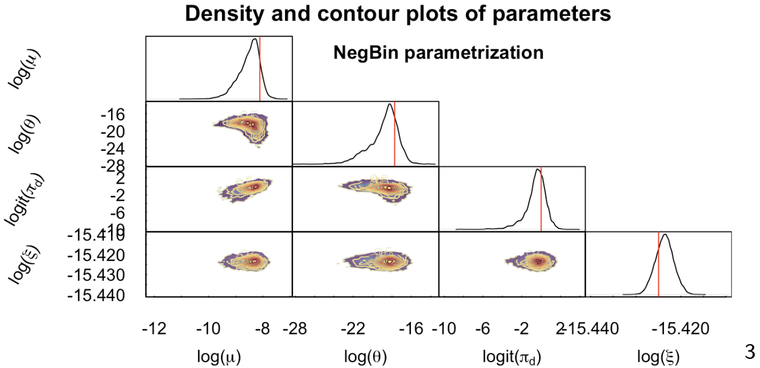


Figure: NS (negative-binomial parametrised) posterior samples under model without overlapping sources.

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

[3] 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

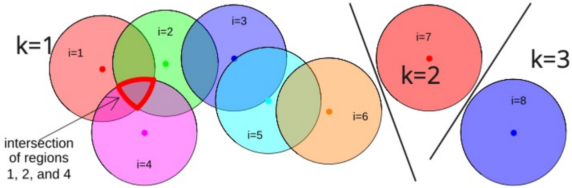


Figure: The overlap structure of sources used for simulation study [4]

[4] source of base picture and data: Wang et al. (2024)

A more complicated simulation study

with overlapping sources and nonhomogeneous background

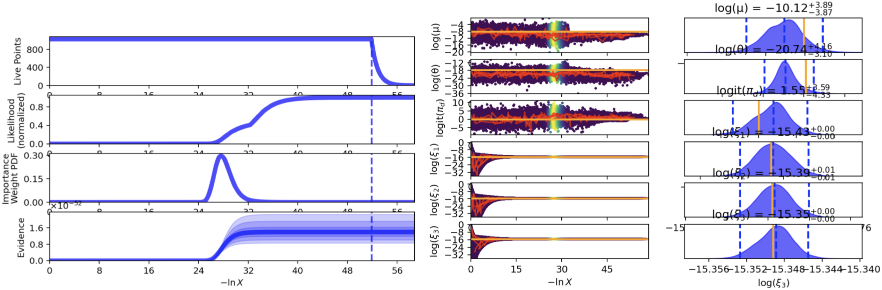
Suppose true values: $\mu = 3 \times 10^{-4}, \theta = 2 \times 10^{-8}, \pi_d = 0.5$.

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 $[\lambda_1, \ldots, \lambda_n]$ 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 $\mathcal{B}_s \overset{\text{indep}}{\sim} \text{Poisson}(a_s\hat{\xi}_s\mathcal{T})$, $\mathcal{S}_s \sim \text{Poisson}(e\mathcal{T}\rho_s)$, $Y_s = \mathcal{B}_s + \mathcal{S}_s$.

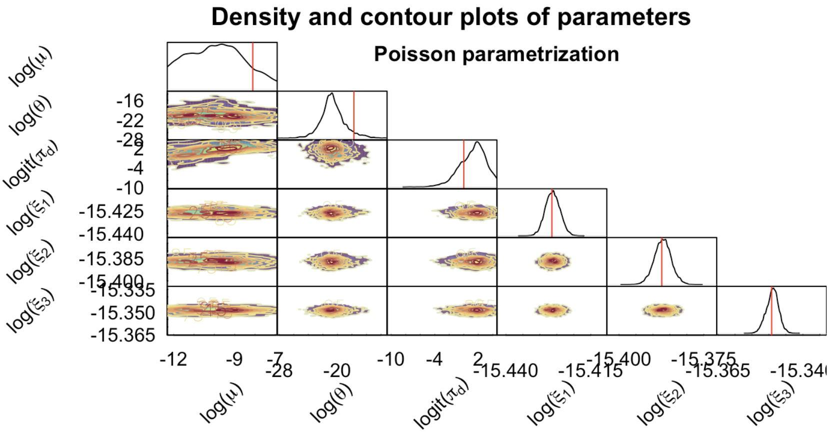
Nested sampling diagnostics

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.