Moment-generating Function for Bayesian Computation

with an application in X-ray source intensity estimation

Siyang Li [1] David van Dyk [1] and Maximilian Autenrieth [1,2]

[1]Imperial College London
jl1023@imperial.ac.uk

[2]Cambridge University

19 December, 2025

Problem: the model marginalisation integral

Notation:

- $\boldsymbol{\xi}$: population (hyper-)parameters;
- $\boldsymbol{\theta}$: individual (unit-level) parameters;
- $\mathbf{y}$: 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.

Problem: the model marginalisation integral

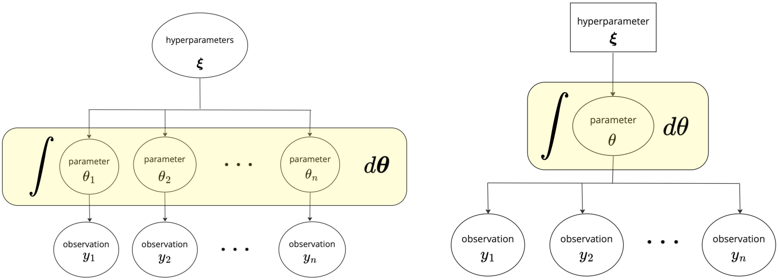


Figure: Hierarchical model marginalisation vs. marginal likelihood computation.

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

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

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

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

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

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

Derivatives of prior moment-generating function

with Poisson likelihoods, multivariate, m observations

Corollary

For $\boldsymbol{\lambda} := \mathbf{r}\boldsymbol{\theta}$ *and Poisson likelihood:*

$$p(\mathbf{y}|\boldsymbol{\lambda}) = \prod_{j=1}^{m} \frac{(\zeta_j \lambda_j)^{y_j}}{y_j!} e^{-\zeta_j \lambda_j},$$

under certain trivial conditions, the mgf-marginalisation is

$$p(\mathbf{y}|\boldsymbol{\xi}) = \left[\prod_{j=1}^{m} \frac{1}{y_j!} \zeta_j^{y_j}\right] \frac{\partial^{\sum_{s=1}^{m} y_s}}{\partial t_1^{y_1} \partial t_2^{y_2} \cdots \partial t_m^{y_m}} \prod_{i=1}^{n} M_{\theta_i|\boldsymbol{\xi}}((\mathbf{t}^\intercal \mathbf{r})_i) \Big|_{\mathbf{t}=-\boldsymbol{\zeta}}. \qquad (2)$$

Limitations

Nested Sampling:

- can deal with complicated models;
- (almost) always practical;
- sophisticated.

Moment-generating Function marginalisation:

- only works for certain likelihoods;
- sometimes not as practical as theory promises;
- underexplored.

Example: estimating X-ray Source Intensities

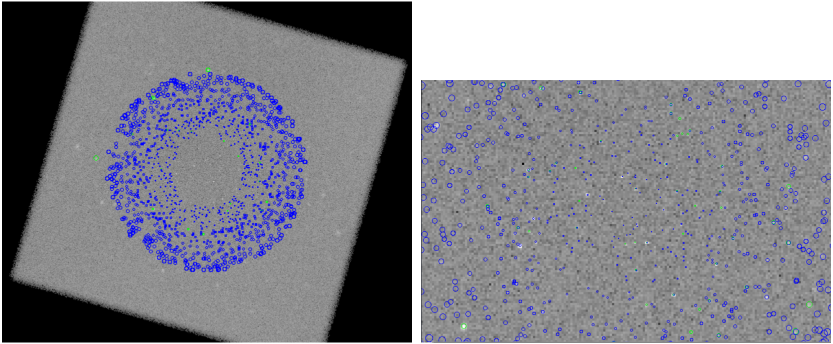


Figure: X-ray photon counts in NGC2516 Southern Beehive, Chandra telescope

Goal: Estimate distribution of source count rates

Issues:

- background contamination
- overlapping source regions

Types of sources

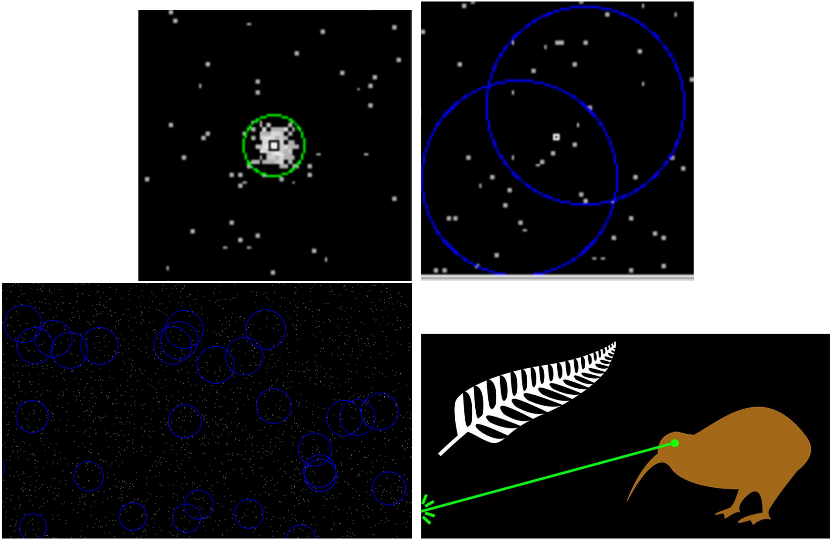


Figure: Isolated, binary-overlapping, multiple overlapping and kiwi-laser sources

Estimating X-ray Source Intensities

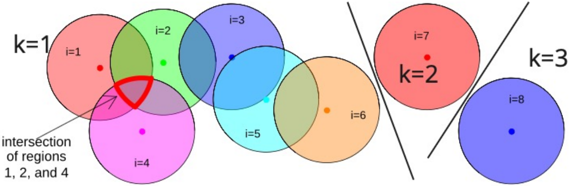
model construction

Observation	Segment counts Y_s and background count X
Likelihood	$Y_s \vert (\underline{\lambda}_i, \underline{\xi}_s) \overset{\text{indep}}{\sim} \text{Poisson}\left(\sum_i (r_{s,i} e_s \underline{\lambda}_i + \underline{a}_s \underline{\xi}_s)\mathcal{T}\right)$
Population(source)	$\underline{\lambda}_i \vert \alpha, \beta \overset{\text{i.i.d.}}{\sim} \text{Gamma}(\alpha, \beta)$
Population(background)	$\underline{\xi}_s \overset{\text{i.i.d.}}{\sim} \text{Gamma}\left(\alpha_\xi, \beta_\xi\right)$
Population parameters	$\frac{\alpha}{\beta} \sim \mathbb{1}[\frac{\alpha}{\beta} > 0]$, $\frac{\alpha}{\beta^2} \sim \mathbb{1}[\frac{\alpha}{\beta^2} > 0]$
Population parameters	$\alpha_\xi = 10^{-6} + x$, $\beta_\xi = A\mathcal{T}(10^{-12} + 1)$

- More parameters than observations! Want $\iint L(\alpha, \beta, \underline{\boldsymbol{\xi}}, \underline{\boldsymbol{\lambda}}; \mathbf{Y}) d\underline{\boldsymbol{\xi}} d\underline{\boldsymbol{\lambda}}$.
- $\mathbf{Y}$ linearly dependent $\implies L(\alpha, \beta, \underline{\boldsymbol{\xi}}, \underline{\boldsymbol{\lambda}}; \mathbf{Y})$ is not a product in $\mathbf{Y}$.

$\implies \int\int \prod \cdots d\underline{\boldsymbol{\xi}} d\underline{\boldsymbol{\lambda}} \neq \prod \int\int \cdots d\underline{\boldsymbol{\xi}} d\underline{\boldsymbol{\lambda}}$.

High-order Derivative to Compute



But how?

$$
\begin{aligned}
& p(\mathbf{y}_{\mathcal{G}_k} | \alpha, \beta) \\
&= \left[\prod_{s \in \mathcal{G}_k} \frac{1}{y_s!} t_s^{y_s} \right] \\
&\quad \times \frac{\partial^{\sum_{s \in \mathcal{G}_k} y_s}}{\partial t_{s_1}^{y_{s_1}} \partial t_{s_2}^{y_{s_2}} \cdots \partial t_{s \in \mathcal{G}_k}^{y_s}} \left\{ \prod_{i \in \mathcal{G}_k} \left(\frac{\beta}{\beta - (\mathbf{t}^\top \tilde{\mathbf{r}})_i} \right)^{\alpha} \prod_{s \in \mathcal{G}_k} \left(\frac{\beta_\xi}{\beta_\xi - (\mathbf{t}^\top \tilde{\mathbf{r}})_{I+s}} \right)^{\alpha_\xi} \right\} \Bigg|_{\mathbf{t} = -\zeta}
\end{aligned}
$$

gives

$$
L(\alpha, \beta; \mathbf{Y}) = \prod_k p(\mathbf{y}_{\mathcal{G}_k} | \alpha, \beta)
$$

Moment generating function for transformed parameters

linear dependency

$\lambda_s := \sum_{i \in s} r_{s,i} \underline{\lambda}_i$.

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

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

Obtain a closed form of the marginal distribution like *no other method*!

Posterior Contours

Figure: Contours of marginal posterior (isolated + binary-overlapping sources)

Model evidence calculation

with Poisson likelihoods, univariate, n observations

For $\lambda = \zeta\theta$, with certain trivial assumptions,

$$p(\mathbf{y}) = \left[\prod_{i=1}^{n} \frac{1}{y_i!}\right] \left(\frac{\partial}{\partial t}\right)^{\sum_{i=1}^{n} y_i} M_\theta(t\zeta)\bigg|_{t=-n} \tag{3}$$

or

$$p(\mathbf{y}) = \left[\prod_{i=1}^{n} \frac{1}{y_i!}\right] \zeta^{\sum_{i=1}^{n} y_i} \left(\frac{\partial}{\partial t}\right)^{\sum_{s=1}^{n} y_s} M_\theta(t)\bigg|_{t=-n\zeta}. \tag{4}$$

Derivatives of prior moment-generating function

with gamma likelihoods, univariate, 1 observation, known shape α

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

Facts:

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

Caputo fractional derivative

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

and

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

Derivatives of prior moment-generating function

with gamma likelihoods, multivariate, n observation, known shape $\boldsymbol{\alpha}$

$$p(\mathbf{y}|\boldsymbol{\xi}) = \left[\prod_{i=1}^{n} \frac{y_i^{\alpha_i - 1}}{\Gamma(\alpha_i)}\right] \frac{\partial^{\sum_{s=1}^{n} \alpha_s}}{\partial t_1^{\alpha_1} \partial t_2^{\alpha_2} \cdots \partial t_n^{\alpha_n}} M_{\boldsymbol{\theta}|\boldsymbol{\xi}}(\mathbf{t}) \bigg|_{\mathbf{t}=-\mathbf{y}}. \tag{5}$$

If $\mathbf{r}$ is diagonal in $\boldsymbol{\beta} := \mathbf{r}\boldsymbol{\theta} > \mathbf{0}$, then

$$p(\mathbf{y}|\boldsymbol{\xi}) = \left[\prod_{i=1}^{n} \frac{y_i^{\alpha_i - 1}}{\Gamma(\alpha_i)}\right] \prod_{i=1}^{n} \frac{\partial^{\alpha_i}}{\partial t_i^{\alpha_i}} M_{\theta_i|\boldsymbol{\xi}}((\mathbf{t}^\top \mathbf{r})_i) \bigg|_{\mathbf{t}=-\mathbf{y}}, \tag{6}$$

Model evidence calculation

with gamma likelihoods, univariate, n observations, known shape α

$\beta := r\theta > 0$,

$$p(\mathbf{y}) = \left[\prod_{i=1}^{n} y_i^{\alpha-1}\right] \frac{r^{n\alpha}}{\Gamma(\alpha)^n} \left(\frac{\partial}{\partial t}\right)^{n\alpha} M_\theta(t)\bigg|_{t=-r\sum_{i=1}^{n} y_i}$$

or

$$p(\mathbf{y}) = \left[\prod_{i=1}^{n} y_i^{\alpha-1}\right] \frac{1}{\Gamma(\alpha)^n} \left(\frac{\partial}{\partial t}\right)^{n\alpha} M_\theta(rt)\bigg|_{t=-\sum_{i=1}^{n} y_i}$$

for (Liouville-)Caputo fractional derivative $\left(\frac{\partial}{\partial t}\right)^{n\alpha}$, lower limit $-\infty$.

Extension: other Likelihoods

$Y \sim$	Transformation	Result	MGF derivative
Laplace($\mu = 0, \tau$)	$\|Y\|$	Gamma($1, \tau$)	1
Normal($\mu = 0, \tau$)	Y^2	Gamma$\left(\frac{1}{2}, \frac{\tau}{2}\right)$	0.5
Rayleigh(b)	Y^2	Gamma$\left(1, \frac{1}{2b^2}\right)$	1
Maxwell-Boltzmann(a)	Y^2	Gamma$\left(\frac{3}{2}, \frac{1}{2a^2}\right)$	1.5
Inv-Gamma(α, β)	$\frac{1}{Y}$	Gamma(α, β)	α
Lévy($\mu = 0, c$)	$\frac{1}{Y}$	Gamma$\left(\frac{1}{2}, \frac{c}{2}\right)$	0.5
Weibull($\rho = 1, \lambda$)	Y	Gamma($1, \lambda$)	1
BurrXII($c = 1, k$)	$\log(Y + 1)$	Gamma($1, k$)	1
Pareto(α, y_m)	$\log\left(\frac{Y}{y_m}\right)$	Gamma($1, \alpha$)	1
Dagum($a = 1, b = 1, q$)	$\log\left(\frac{1}{Y} + 1\right)$	Gamma($1, q$)	1
Gompertz($b = 1, \eta$)	$1 - e^Y$	Gamma($1, \eta$)	1

Extension: posterior moments/mgf

with Poisson likelihoods, univariate, 1 observation

For canonical parameter $\eta = \log \theta$,

$$\mathbb{E}_\theta(\theta^t | y) = M_{\eta|y}(t) = \frac{\left(\frac{d}{dr}\right)^{y+t} M_\theta(r)\Big|_{r=-1}}{\left(\frac{d}{dr}\right)^{y} M_\theta(r)\Big|_{r=-1}}, \tag{7}$$

which is also known as 'transfer function'.

Extension: posterior mgf

with gamma likelihoods, univariate, 1 observation, known shape α

$$M_{\theta|y}(t) = \frac{\left(\frac{d}{dr}\right)^{\alpha}_{(-\infty)+} M_\theta(r)\Big|_{r=-y+t}}{\left(\frac{d}{dr}\right)^{\alpha}_{(-\infty)+} M_\theta(r)\Big|_{r=-y}}. \tag{8}$$

Appendix

Derivative-expectation exchanging conditions

Proposition

The conditions for exchanging derivative and expectation are given by

1. $\kappa \in \mathbb{N}_0$;
2. $\upsilon(\theta, t)$ *is defined for* $0 \leq \theta < \infty$ *and* $c \leq t \leq d$*, where* $[c, d]$ *is any arbitrary subset of the radius of convergence for the prior mgf* $M_\theta(t)$;
3. *For the prior measure* $\mu(d\theta)$, $\upsilon(\theta, t)$ *is integrable with respect to* μ *for each* $t \in [c, d]$;
4. *Let* $s \in (c, d)$. *Then for every* $\epsilon > 0$ *there exists a* $\delta > 0$ *such that*

$$\left| \left(\frac{d}{dt}\right)^{\kappa} \upsilon(\theta, t) - \left(\frac{d}{dt}\right)^{\kappa} \upsilon(\theta, s) \right| < \epsilon$$

for all $\theta \in [0, \infty)$ *and for all* $t \in (s - \delta, s + \delta)$.

Poisson derivative-expectation exchanging

1. κ is the observed Poisson count, so $\kappa \in \mathbb{N}_0$;
2. $\left(\frac{d}{dt}\right)^{\kappa} \upsilon(\theta, t) = \theta^{\kappa} e^{t\theta}$ with $t = -1$, so $\upsilon(\theta, t) = e^{\theta t}$;
3. Prior distributions with mgf's exist and is finite for all t in $[c, d]$ for $c < -1 < d \implies \forall t \in [c, d]$, $\upsilon(\theta, t) = e^{t\theta}$ is $\mu(d\theta)$-integrable;
4. $\theta^{\kappa} e^{t\theta}$ is continuous with respect to both $\theta \in [0, \infty)$ and $t \in [c, d]$.

Derivatives of prior moment-generating function

with Poisson likelihoods, multivariate, m observations

Corollary

Suppose $\boldsymbol{\lambda} := \mathbf{r}\boldsymbol{\theta}$ *with known* $\mathbf{r} \in \mathbb{R}^{m\times n}$ *for* $m \geq n$ *and* $y_j \overset{indep}{\sim} Poisson(\lambda_j)$, *such that*

$$p(\mathbf{y}|\boldsymbol{\lambda}) = \prod_{j=1}^{m} \frac{(\zeta_j \lambda_j)^{y_j}}{y_j!} e^{-\zeta_j \lambda_j},$$

where $\boldsymbol{\zeta} \in \mathbb{R}^m$ *are known constants. Suppose the population 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*

$$p(\mathbf{y}|\boldsymbol{\xi}) = \left[\prod_{j=1}^{m} \frac{1}{y_j!} \zeta_j^{y_j}\right] \frac{\partial^{\sum_{s=1}^{m} y_s}}{\partial t_1^{y_1} \partial t_2^{y_2} \cdots \partial t_m^{y_m}} \prod_{i=1}^{n} M_{\theta_i|\boldsymbol{\xi}}((\mathbf{t}^\intercal \mathbf{r})_i)\Big|_{\mathbf{t}=-\boldsymbol{\zeta}}, \quad (9)$$

if $M_{\boldsymbol{\theta}|\boldsymbol{\xi}}(\mathbf{t})$ *is continuous and differentiable up to the suitable order at* $-\boldsymbol{\zeta}$*.*

Bell Polynomials

Computing High-order Derivatives, **Isolated Sources**

$B_0 := 1$ and

$$B_{n+1}(x_1, \ldots, x_{n+1}) = \sum_{i=0}^{n} \binom{n}{i} B_{n-i}(x_1, \ldots, x_{n-i})\, x_{i+1}, \tag{10}$$

For $\underline{K}(t) := \sum_{i=1}^{n} K_{\theta_i|\boldsymbol{\xi}}(t)$, Using Faà di Bruno's formula,

$$\begin{aligned}
&\left(\frac{d}{dt}\right)^{y} \prod_{i=1}^{n} M_{\theta_i|\boldsymbol{\xi}}(t) \\
&= \left(\frac{d}{dt}\right)^{y} \exp\left[\sum_{i=1}^{n} K_{\theta_i|\boldsymbol{\xi}}(t)\right] \\
&= \exp\left[\underline{K}(t)\right] B_y\left[\underline{K}'(t), \underline{K}''(t), \ldots, \underline{K}^{(y)}(t)\right] \\
&= \left[\prod_{i=1}^{n} M_{\theta_i|\boldsymbol{\xi}}(t)\right] B_y\left[\underline{K}'(t), \underline{K}''(t), \ldots, \underline{K}^{(y)}(t)\right]
\end{aligned} \tag{11}$$

Generalised Leibniz Rule

Computing High-order Derivatives, **Overlapping Sources**

For $\mathbf{k}_i! := \prod_{j=1}^m k_{i,j}!$, $|\mathbf{k}| := \sum_{j=1}^m k_j$, $D^{\mathbf{k}} := \frac{\partial^{|\mathbf{k}|}}{\partial t_1^{k_1} \cdots \partial t_m^{k_m}}$ and multinomial coefficient

$$\binom{\mathbf{y}}{\mathbf{k}_1, \ldots, \mathbf{k}_n} = \frac{\mathbf{y}!}{\prod_{i=1}^n \mathbf{k}_i!} = \frac{y_1! \cdots y_m!}{\prod_{i=1}^n \prod_{j=1}^m k_{i,j}!},$$

GLR is

$$D^{\mathbf{y}} \left(\prod_{i=1}^n M_{\theta_i|\boldsymbol{\xi}} \right) (\zeta) = \sum_{\mathbf{k}_1 + \cdots + \mathbf{k}_n = \mathbf{y}} \binom{\mathbf{y}}{\mathbf{k}_1, \ldots, \mathbf{k}_n} \prod_{i=1}^n D^{\mathbf{k}_i} M_{\theta_i|\boldsymbol{\xi}}(\zeta), \tag{12}$$

Generalised Leibniz Rule

Computing High-order Derivatives, **Binary Overlapping Sources**

For

$$
\begin{aligned}
L_A &:= \beta - r_1 e_u t_u^0 - r_2 e_w t_w^0, \\
L_B &:= \beta - r_3 e_v t_v^0 - r_4 e_w t_w^0, \\
M_u &:= \beta_2 - a_u t_u^0, \qquad M_v := \beta_2 - a_v t_v^0, \qquad M_w := \beta_2 - a_w t_w^0,
\end{aligned}
$$

$$
\begin{aligned}
&\frac{\partial^{y_u+y_v+y_w}}{\partial t_u^{y_u} \partial t_v^{y_v} \partial t_w^{y_w}} f(t_u, t_v, t_w)\Big|_{(t^0)} \\
=&y_u! y_v! y_w! \sum_{m=0}^{y_u} \sum_{p=0}^{y_v} \sum_{n=0}^{y_w} \sum_{q=0}^{y_w - n} \left(\frac{\beta}{L_A}\right)^{\alpha} \frac{\Gamma(\alpha + m + n)}{\Gamma(\alpha) m! n!} \left(\frac{r_1 e_u}{L_A}\right)^{m} \left(\frac{r_2 e_w}{L_A}\right)^{n} \\
&\times \left(\frac{\beta}{L_B}\right)^{\alpha} \frac{\Gamma(\alpha + p + q)}{\Gamma(\alpha) p! q!} \left(\frac{r_3 e_v}{L_B}\right)^{p} \left(\frac{r_4 e_w}{L_B}\right)^{q} \\
&\times \left(\frac{\beta_2}{M_u}\right)^{\alpha_2} \frac{\Gamma(\alpha_2 + (y_u - m))}{\Gamma(\alpha_2)(y_u - m)!} \left(\frac{a_u}{M_u}\right)^{(y_u - m)} \left(\frac{\beta_2}{M_v}\right)^{\alpha_2} \frac{\Gamma(\alpha_2 + (y_v - p))}{\Gamma(\alpha_2)(y_v - p)!} \left(\frac{a_v}{M_v}\right)^{(y_v} \\
&\times \left(\frac{\beta_2}{M_w}\right)^{\alpha_2} \frac{\Gamma(\alpha_2 + (y_w - n - q))}{\Gamma(\alpha_2)(y_w - n - q)!} \left(\frac{a_w}{M_w}\right)^{(y_w - n - q)} .
\end{aligned} \tag{13}
$$

Example 2

Pump failure, gamma-prior hierarchical model

From **?**. 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 2

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 [**?**]: $\hat{\alpha} = 1.27$ 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 3

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

Example 3

Pump failure, Pareto-prior non-hierarchical model

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

$$
\begin{aligned}
p(\mathbf{y}|\alpha, k) &= \prod_{i=1}^{10} p(y_i(t_i)|\alpha, k) = \alpha k^{\sum_{i=1}^{10} y_i} \prod_{i=1}^{10} \frac{t_i^{y_i}}{y_i!} E_{\alpha-y_i+1}(kt_i) \\
&= \left[\prod_{i=1}^{10} \frac{t_i^{y_i}}{y_i!}\right] \alpha k^{\sum_{i=1}^{10} y_i} E_{\alpha+1-\sum_{i=1}^{10} y_i}\left(k \sum_{i=1}^{10} t_i\right) \\
&= [2.799194 \times 10^{48}] \alpha k^{75} E_{\alpha-74}(350.032k),
\end{aligned}
$$