From Delaporte to Poisson or Negative Binomial: MH-within-Gibbs Inference and Limits

Leonidas Ntrekos
Email: l.ntrekos@gmail.com
GitHub: LNtrekos

August 29, 2025

1 Setup

This short note presents a reproducible *MH-within-Gibbs* sampler for the Delaporte *Distribution* in R, not so efficient but with nice presentation of the whole methodology. The goal is to estimate (posterior) (r, p, λ) from observed (simulated) counts $X_1, \ldots, X_n$ and to illustrate (in R) the process where the posterior approaches a Negative Binomial distribution when the Poisson component vanishes or vice versa depending on sample X. The code lives in `src/delaporte_sampler.R` figures are auto-generated to `results/figures/`.

2 Model

Let $X_i = Y_{1i} + Y_{2i}$ with $Y_{1i} \perp\!\!\!\perp Y_{2i}$ and

$$\begin{aligned} Y_{1i} &\sim \mathcal{NB}(r, p), & \Pr(Y_{1i} = y) &= \frac{\Gamma(y+r)}{\Gamma(r)\,\Gamma(y+1)}\, p^{r}\,(1-p)^{y}, \\ Y_{2i} &\sim \mathcal{P}(\lambda), & \Pr(Y_{2i} = z) &= e^{-\lambda}\,\frac{\lambda^{z}}{z!}, \end{aligned}$$

where $r > 0$, $p \in (0, 1)$, $\lambda > 0$. This yields the Delaporte distribution for X_i.

Priors. We assume a priori independence,

$$\pi(r, p, \lambda) = \pi(r)\,\pi(p)\,\pi(\lambda),$$

and place weakly-informative priors

$$r \sim \text{Gamma}(a_r, b_r), \qquad p \sim \text{Beta}(a_p, b_p), \qquad \lambda \sim \text{Gamma}(a_\lambda, b_\lambda),$$

where the Gamma is in *rate* parameterization,

$$f(r \mid a, b) = \frac{b^{a}}{\Gamma(a)}\, r^{a-1} e^{-br}, \qquad r > 0,$$

and the Beta density is

$$f(p \mid a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\, p^{a-1}(1-p)^{b-1}, \qquad p \in (0, 1).$$

For the examples in this report we take

$$a_r = b_r = 0.01, \quad a_p = b_p = 1, \quad a_\lambda = b_\lambda = 0.01 \text{ (and } n = 100)$$

but any other weakly-informative choices can be used.

3 Inference (MH-within-Gibbs)

Latent counts $Y_{1i} \in \{0, \dots, X_i\}$ are introduced. One sweep consists of:

1. **Update $\lambda \mid \cdot$.** Given Y_1, the full conditional is
$$\lambda \mid \cdot \sim \mathrm{Gamma}\Big(\underbrace{\sum_i (X_i - Y_{1i})}_{\text{Poisson totals}} + a_\lambda,\ n + b_\lambda \Big).$$

2. **Update $p \mid \cdot$.** Given Y_1 and r,
$$p \mid \cdot \sim \mathrm{Beta}\big(nr + a_p,\ \sum_i Y_{1i} + b_p \big).$$

3. **Update $r \mid \cdot$ by MH.** With a log-normal random-walk proposal $r' \sim \log \mathcal{N}(\log r, v^2)$, accept with probability $\min\{1, \alpha\}$ where
$$\begin{aligned}\log \alpha = &\left[\sum_{i=1}^{n} \big(\log \Gamma(Y_{1i} + r') - \log \Gamma(r') \big) + n\, r' \log p + a_r \log r' - b_r\, r' \right] \\ &- \left[\sum_{i=1}^{n} \big(\log \Gamma(Y_{1i} + r) \big) - n \log \Gamma(r) + n\, r \log p + a_r \log r - b_r\, r \right].\end{aligned}$$

4. **Update $Y_{1i} \mid \cdot$.** For each i, enumerate $y = 0, \dots, X_i$ and sample with weights proportional to
$$\Pr(Y_{1i} = y \mid \cdot) \propto \frac{\Gamma(y + r)}{\Gamma(r)\, \Gamma(y + 1)}\, p^{r} (1 - p)^{y} \cdot \frac{\lambda^{X_i - y}}{\Gamma(X_i - y + 1)}.$$

   Stable enumeration for Y_{1i}. For each observation i and candidate value $y \in \{0, \dots, x_i\}$ define the log–weight
$$\begin{aligned}\ell_{iy} = &\log \Gamma(y + r) - \log \Gamma(r) - \log \Gamma(y + 1)\ +\ r \log p\ +\ y \log(1 - p) \\ &+\ (x_i - y) \log \lambda\ -\ \log \Gamma(x_i - y + 1)\ -\ \lambda.\end{aligned}$$

   The terms $r \log p$ and $-\lambda$ do not depend on y and can be dropped when normalizing. To compute probabilities stably, set $m_i = \max_{0 \le j \le x_i} \ell_{ij}$ and use
$$\Pr(Y_{1i} = y \mid r, p, \lambda, x_i) = \frac{\exp\{\ell_{iy} - m_i\}}{\sum_{j=0}^{x_i} \exp\{\ell_{ij} - m_i\}} = \mathrm{softmax}\left(\ell_{i0}, \dots, \ell_{ix_i} \right)_y.$$

   Numerical stability (clamping). To avoid undefined logs when p is arbitrarily close to 0 or 1 and when λ is close to 0, we clamp with a tiny ε:
$$\tilde{p} = \mathrm{clamp}_{01}(p; \varepsilon) := \min\{\max(p, \varepsilon),\ 1 - \varepsilon\}, \qquad \tilde{\lambda} = \max(\lambda, \varepsilon),$$
   with $\varepsilon = 10^{-12}$ in our implementation.

4 Chain behaviour across simulated datasets.

Because $X_i \sim \text{Delaporte}(r, p, \lambda)$ satisfies

$$\mathbb{E}[X] = \lambda + \underbrace{\frac{r(1-p)}{p}}_{\mu_{\text{NB}}},$$

many (r, p, λ) combinations produce almost the same mean when n is moderate, but with quite different distribution each time. As a result, the MH-within-Gibbs sampler often exhibits visible "jumps" between two interpretations of the data: (i) a *Poisson-like* regime with r small, $p \approx 1$, and λ explaining most of the mean; and (ii) an *NB-like* regime with larger r, moderate p, and λ near 0. The traces in plot below illustrate these switches; the density panels show the long right tail of r and the concentration of p near the boundary, both typical under weak priors.

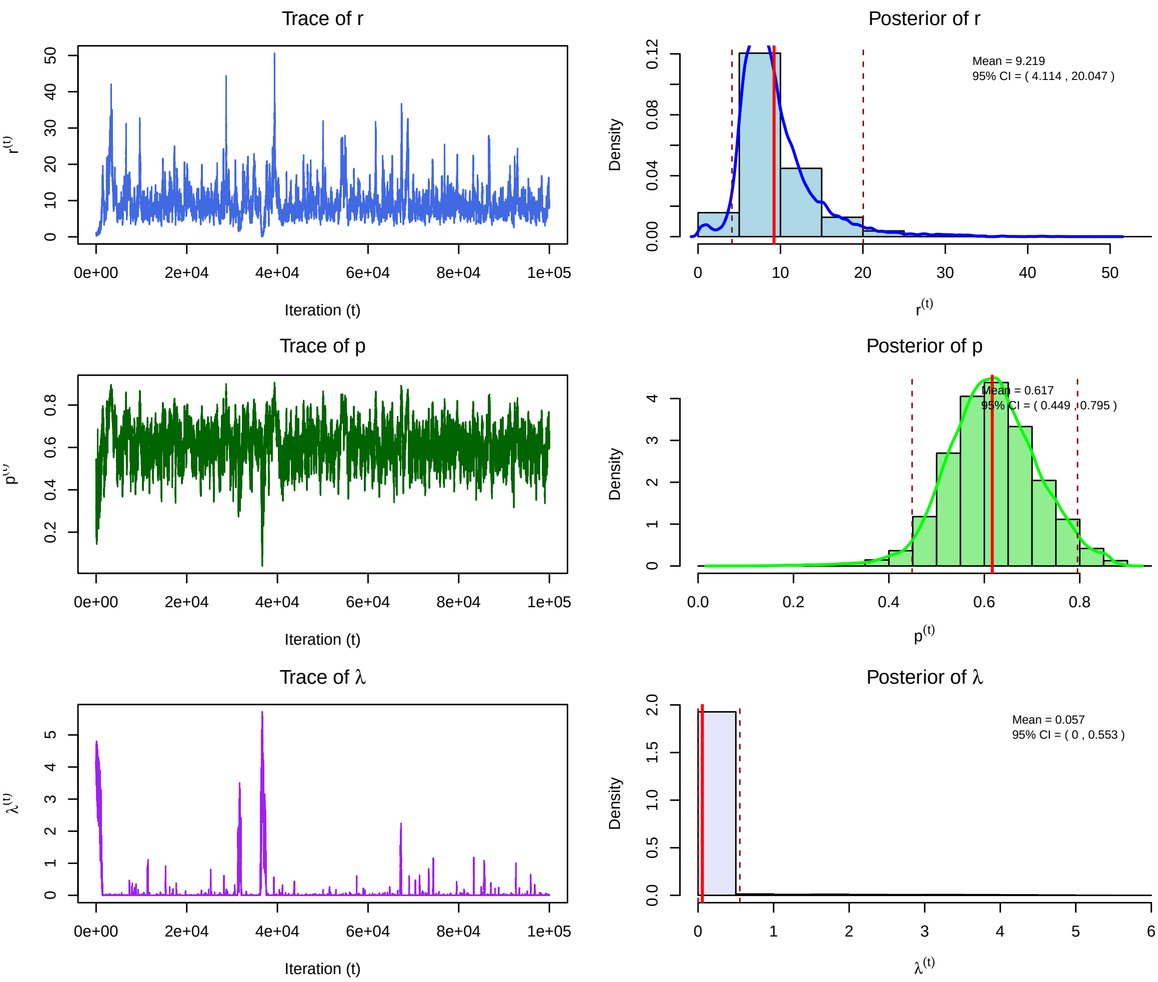

Figure 1: Posterior diagnostics from the repository example: trace plots (left) and marginal posterior densities (right) for r, p, and λ; solid lines mark posterior means and dashed lines the 95% credible intervals. is effectively zero (posterior mean 0.057), implying that the Poisson component vanishes in practice; the Delaporte reduces to a near–Negative Binomial fit.

Practical notes. The observed switching is *posterior multimodality/flatness*, not necessarily non-convergence. We mitigate numerical issues by clamping p and λ (Sec. 3): $\tilde{p} = \min\{\max(p, \varepsilon), 1-\varepsilon\}$ and $\tilde{\lambda} = \max(\lambda, \varepsilon)$ with $\varepsilon = 10^{-12}$, and by tuning the r proposal to a 20–40% acceptance rate.

Below is the pseudoalgorithm summarising one MH-within-Gibbs sweep using the updates above.

Algorithm 1 MH-within-Gibbs for the Delaporte model

Require: Data $X_{1:n}$; hyperparameters (a_r, b_r), (a_p, b_p), (a_λ, b_λ); proposal scale v; iterations T

1: Initialize $r^{(0)} > 0$, $p^{(0)} \in (0,1)$, $\lambda^{(0)} > 0$, and $Y_{1i}^{(0)} \in \{0, \dots, X_i\}$ for all i

2: **for** $t = 1, \dots, T$ **do**

3: **Update** $\lambda \mid \cdot$: $\lambda^{(t)} \sim \text{Gamma}\left(\sum_{i=1}^{n}(X_i - Y_{1i}^{(t-1)}) + a_\lambda,\ n + b_\lambda\right)$

4: **Update** $p \mid \cdot$: $p^{(t)} \sim \text{Beta}\left(n\, r^{(t-1)} + a_p,\ \sum_{i=1}^{n} Y_{1i}^{(t-1)} + b_p\right)$

5: **Update** $r \mid \cdot$ **by MH**: Propose $r' \sim \log\mathcal{N}(\log r^{(t-1)}, v^2)$ and accept with probability $\min\{1, e^{\log\alpha}\}$ where

$$\log\alpha = \Big[\sum_i (\log\Gamma(Y_{1i}^{(t-1)} + r') - \log\Gamma(r')) + n\, r' \log p^{(t)} + a_r \log r' - b_r r'\Big] - \Big[\sum_i (\log\Gamma(Y_{1i}^{(t-1)} + r^{(t-1)}) - \log\Gamma(r^{(t-1)})) + n\, r^{(t-1)} \log p^{(t)} + a_r \log r^{(t-1)} - b_r r^{(t-1)}\Big].$$

If accepted set $r^{(t)} \leftarrow r'$, else $r^{(t)} \leftarrow r^{(t-1)}$.

6: **Update each** $Y_{1i} \mid \cdot$ **by finite enumeration**: For $y = 0, \dots, X_i$ compute

$$\ell_{iy} = \log\Gamma(y + r^{(t)}) - \log\Gamma(r^{(t)}) - \log\Gamma(y+1) + r^{(t)} \log p^{(t)} + y\log(1 - p^{(t)}) + (X_i - y)\log\lambda^{(t)} - \log\Gamma(X_i - y + 1).$$

Set $m_i = \max_y \ell_{iy}$ and $w_{iy} = \exp(\ell_{iy} - m_i)$; sample $Y_{1i}^{(t)}$ from Categorical$(w_{i0}, \dots, w_{iX_i})$.

7: **end for**