

# Efficient brute-force marginalisation of discrete variables in Stan: an epidemic model of sub-critical transmission as a case study

Luiz Max Carvalho<sup>1</sup> Gytis Dudas<sup>2</sup> Andrew Rambaut<sup>3</sup> Tanja Stadler<sup>4</sup> Trevor Bedford<sup>5</sup>

<sup>1</sup>School of Applied Mathematics, Getúlio Vargas Foundation

<sup>2</sup>Independent consultant

<sup>3</sup>Institute of Evolutionary Biology, University of Edinburgh

<sup>4</sup>Department of Biosystems Science and Engineering, ETH Zürich

<sup>5</sup>Vaccine and Infectious Disease Division, Fred Hutch Cancer Centre

## Motivation

As is widely known, the HMC algorithm underlying Stan cannot handle discrete latent variables (parameters). Thus, when for one to be able to fit a model containing latent variables in Stan, these need to be marginalised over, either analytically or by brute-force. Let  $X$  be the discrete latent variables of interest,  $\mathbf{Y}$  be the data and other variables in the model and  $\boldsymbol{\theta}$  be the model parameters. We can write

$$\Pr(\boldsymbol{\theta} \mid \mathbf{Y}) \propto \sum_{i \in \Omega} \Pr(\boldsymbol{\theta} \mid \mathbf{Y}, X = i) = \sum_{i \in \Omega} \Pr(\mathbf{Y}, X = i \mid \boldsymbol{\theta}) \pi(\boldsymbol{\theta}),$$

which usually takes the more familiar form

$$\Pr(\boldsymbol{\theta} \mid \mathbf{Y}) \propto \sum_{i=0}^{\infty} \Pr(\mathbf{Y}, X = i \mid \boldsymbol{\theta}) \pi(\boldsymbol{\theta}). \quad (1)$$

Here we will discuss a few ways to do the marginalisation in (1) and discuss what is, to the best of our knowledge, **a new technique for truncating infinite sums** while controlling approximation error.

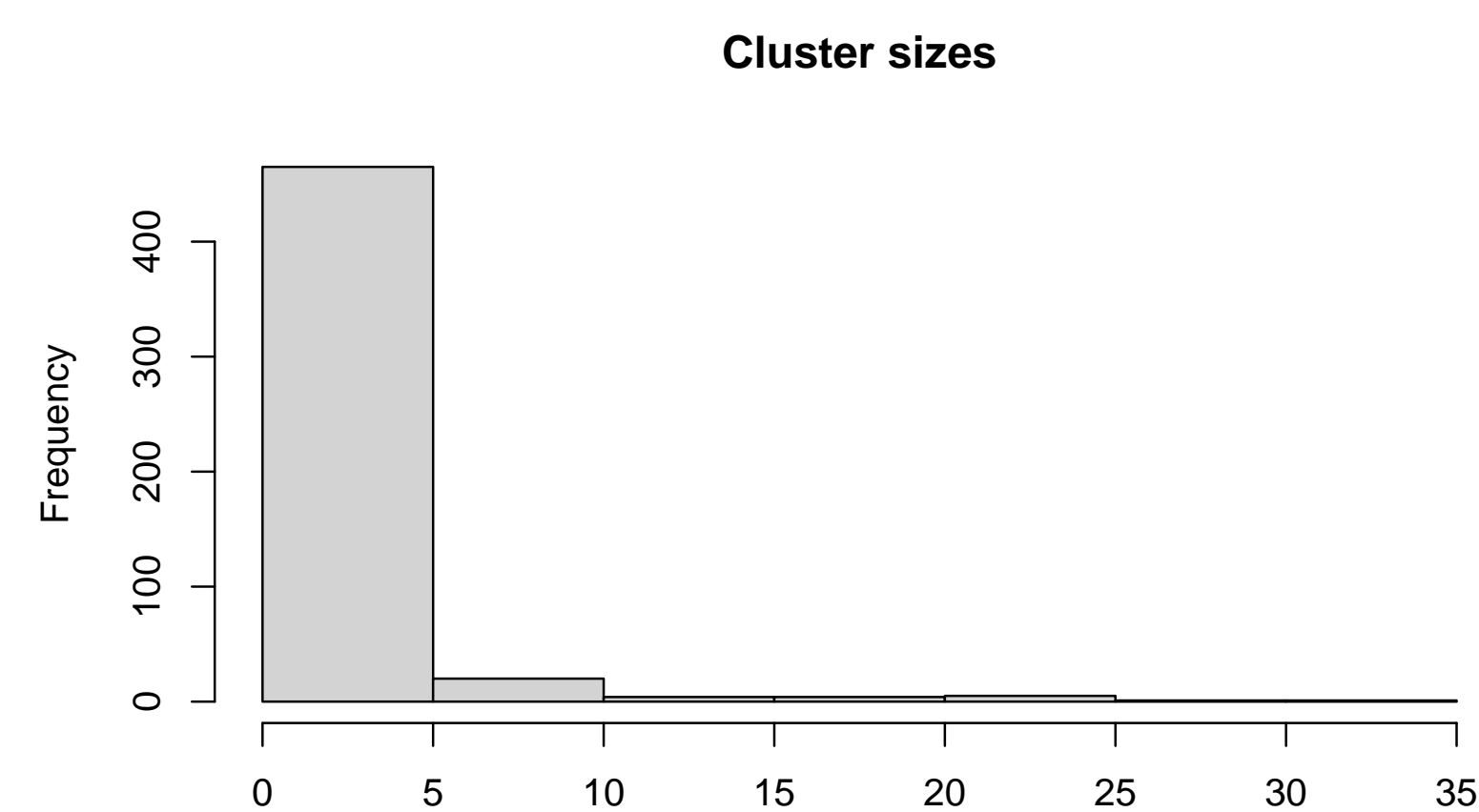
## A motivating example: sub-critical transmission chains

Here are interested in the model [Blumberg & Lloyd-Smith, \(2013a\)](#), in which subcritical transmission leads to “stuttering” chains of transmission.

- **Birth-death.** The model assumes a birth-death process in which an initial single infection results in a distribution of offspring infections.
- **Secondary infection** distribution for each infection follows a negative binomial distribution with mean  $R_0$  and over-dispersion parameter  $\omega$ . When  $\omega \rightarrow \infty$ , the secondary infection distribution is Poisson distributed and when  $\omega = 1$  the secondary infection distribution is an exponential distribution.
- **Total number of cases** within a transmission chain,  $c$ , that are caused by a single introduction (including the index case) has limiting distribution:

$$\Pr(C = c \mid R_0, \omega) = \frac{\Gamma(\omega c + c - 1)}{\Gamma(\omega c) \Gamma(c + 1)} \frac{(R_0/\omega)^{c-1}}{(1 + R_0/\omega)^{\omega c + c - 1}}. \quad (2)$$

- **An outbreak** may be comprised of multiple concurrent transmission chains resulting from multiple introduction events. Each chain  $i$  comprises a count of cases  $c_i$  and the vector of case counts within distinct transmission chains is  $\mathbf{C} = (c_1, \dots, c_K)$ .
- **Total number** of cases in the outbreak by  $N = \sum_{i=1}^K c_i$ .
- **MLE.** When  $\mathbf{C}$  are observed directly, the maximum likelihood estimator of  $R_0$  is  $\widehat{R}_0^{\text{MLE}} = 1 - \frac{K}{N}$ .



## Modelling observation error

The main issue with estimating  $\{R_0, \omega\}$  from data is that the clusters (and cluster sizes) are observed imperfectly. Here we explore the so-called size-independent model of observation error, which motivates the employment of truncation techniques.

### Binomial (size-independent) observation error

Each case in chain/cluster  $c_i$  is observed independently with probability  $\psi$  such that we now observe a set  $\mathbf{S} = (s_1, \dots, s_J)$  of cases. Thus we have

$$P(S = s_i \mid C = c_i, \psi) = \binom{c_i}{s_i} \psi^{s_i} (1 - \psi)^{c_i - s_i}. \quad (3)$$

When  $s_i = 0$ , the cluster is not observed. The conditional probability that a cluster is not observed is  $\Pr(S = 0 \mid C = c_i, \psi) = (1 - \psi)^{c_i}$ , while the marginal probability is

$$\Pr(S = 0 \mid \psi, R_0, \omega) = \sum_{j=1}^{\infty} \Pr(S = 0 \mid C = j, \psi) \Pr(C = j \mid R_0, \omega). \quad (4)$$

After some manipulation we get

$$\sum_{k=1}^{\infty} \kappa(k, \omega) b^k = \frac{a}{1 + a}, \quad (5)$$

with  $a := R_0/\omega$ ,  $b = a/(1 + a)^{(1+\omega)}$  and  $\kappa(c, \omega) := \frac{\Gamma(\omega c + c - 1)}{\Gamma(\omega c) \Gamma(c + 1)}$ . Whilst the mapping  $a \mapsto b$  is not invertible analytically, it can be easily inverted numerically *via* Newton solver. Hence for the case  $s_i = 0$ , we can write

$$P(S = 0 \mid R_0, \omega, \psi) = \sum_{k=1}^{\infty} \frac{1 + a}{a} \kappa(k, \omega) b^k (1 - \psi)^k, \\ = \frac{\omega + R_0}{R_0} \frac{u}{(1 + u)}, \quad (6)$$

where  $b \cdot (1 - \psi) = u/(1 + u)^{(1+\omega)}$  and  $u$  can be obtained by numerical inversion. The marginal probability when  $S > 0$  is not available in closed-form, and we discuss a strategy to marginalisation in **Adaptive truncation**.

## Adaptive truncation

Let  $p(j) := \Pr(\boldsymbol{\theta} \mid \mathbf{Y}, X = j)$  let  $Q = \sum_{j=s}^{\infty} p(j)$  be the sum of interest, with  $\epsilon$  the desired approximation error. Define  $L := \lim_{n \rightarrow \infty} p(n)$ . The algorithm can be summarised as follows:

1. Find  $\tilde{N}$  such that  $p(j+1)/p(j) < 1$  for all  $j > \tilde{N}$ ;
2. Compute  $\dot{Q} = \sum_{j=s}^{\tilde{N}} p(j)$ ;
3. Let  $z_j = \frac{p(j+1)p(j)}{p(j) - p(j+1)}$ ;
4. Find  $\tilde{N} \geq \tilde{N}$  such that  $p(j) \left( \frac{L}{1-L} \right) + z_j =: \Delta_j < 2\epsilon$  for all  $j > \tilde{N}$ ;
5. Compute  $\ddot{Q} = \sum_{j'=\tilde{N}}^{\tilde{N}} p(j')$ ;
6. Compute the estimate  $\hat{Q} = \dot{Q} + \ddot{Q} + \frac{1}{2} \Delta_{\tilde{N}}$ .

For the model discussed here,  $L = [(1 + \omega)/(R_0 + \omega)]^{1+\omega} R_0(1 - \psi)$ , but usually we have  $L = 0$ .

## Priors

- **Independent priors.**  $\pi(R_0, \omega, \psi) = \pi_R(R_0) \pi_H(\omega) \pi_S(\psi)$ ;
- **Reproduction number:**  $\pi_R(R_0)$ , Beta(1, 1) or a boundary-avoiding Beta(2, 2) – reference prior is available when  $\omega$  is fixed;
- **Dispersion:**  $\pi_H(\omega)$ , Gamma(1, 1);
- **Observation probability:**  $\pi_S(\psi)$ , Beta(1, 1);

## Including information about the number of unobserved clusters

- **How many unobserved clusters?** We do not observe the number of size zero clusters,  $n_0 = K - J$ , directly. Since we know the number of observed clusters  $J$ , we can also choose to model the true number of clusters  $K$  instead of  $n_0$ , and this is the route we take here.
- **Marginalising over  $K$** 
  - Let  $p_0 := \Pr(S = 0 \mid R_0, \omega, \psi)$ ;
  - Then for  $k > J$  we have  $\Pr(K = k \mid \boldsymbol{\theta}) = \Pr(K = k \mid p_0) = p_0^{k-J}$ ;
  - Fix a total (case) count sum:  $N = \sum_{i=1}^J c_i$ ;
  - Then  $K \in [J, U]$ , where  $U = J + N - T$  and  $T = \sum_{i=1}^J s_i$ .

Using a doubly-truncated Poisson prior on  $K$ , the marginalisation is now

$$m''(J, U, p_0, \lambda_K) := \sum_{k=J}^U \Pr(K = k \mid p_0) \pi_2(K = k \mid \lambda_K) \\ = \frac{\exp(\lambda_K(p_0 - 1))}{p_0^J}.$$

## Preliminary results

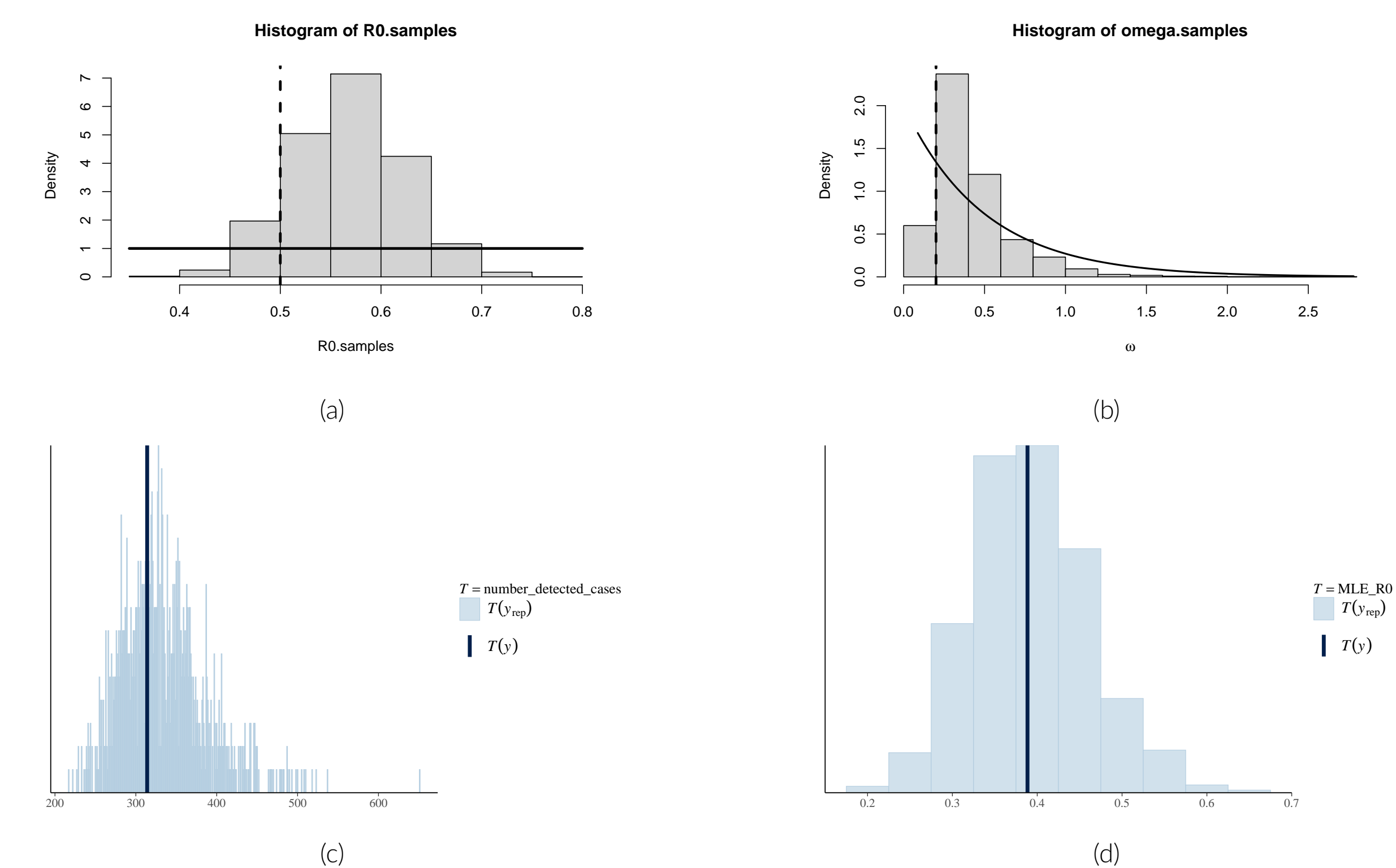


Figure 1. Posterior distributions and posterior predictives for quantities of interest in the model.