

Nested Sampling for Bayesian Model Comparison in the Context of *Salmonella* Disease Dynamics

Richard Dybowski*, Trevelyan J. McKinley, Pietro Mastroeni, Olivier Restif

Department of Veterinary Medicine, University of Cambridge, Cambridge, United Kingdom

Abstract

Understanding the mechanisms underlying the observed dynamics of complex biological systems requires the statistical assessment and comparison of multiple alternative models. Although this has traditionally been done using maximum likelihood-based methods such as Akaike's Information Criterion (AIC), Bayesian methods have gained in popularity because they provide more informative output in the form of posterior probability distributions. However, comparison between multiple models in a Bayesian framework is made difficult by the computational cost of numerical integration over large parameter spaces. A new, efficient method for the computation of posterior probabilities has recently been proposed and applied to complex problems from the physical sciences. Here we demonstrate how nested sampling can be used for inference and model comparison in biological sciences. We present a reanalysis of data from experimental infection of mice with *Salmonella enterica* showing the distribution of bacteria in liver cells. In addition to confirming the main finding of the original analysis, which relied on AIC, our approach provides: (a) integration across the parameter space, (b) estimation of the posterior parameter distributions (with visualisations of parameter correlations), and (c) estimation of the posterior predictive distributions for goodness-of-fit assessments of the models. The goodness-of-fit results suggest that alternative mechanistic models and a relaxation of the quasi-stationary assumption should be considered.

Citation: Dybowski R, McKinley TJ, Mastroeni P, Restif O (2013) Nested Sampling for Bayesian Model Comparison in the Context of *Salmonella* Disease Dynamics. PLoS ONE 8(12): e82317. doi:10.1371/journal.pone.0082317

Editor: Andrew J. Yates, Albert Einstein College of Medicine, United States of America

Received: August 15, 2013; **Accepted:** October 22, 2013; **Published:** December 20, 2013

Copyright: © 2013 Dybowski et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: RD was funded by the Biotechnology and Biological Sciences Research Council (BBSRC) (grant number BB/1002189/1). TJM was funded by the Biotechnology and Biological Sciences Research Council (BBSRC) (grant number BB/1012192/1). OR was funded by the Royal Society. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: rd460@cam.ac.uk

Introduction

Model comparison

Model-based inference is widely used in life sciences in order to assess the plausibility of hypothesised biological mechanisms based on data from observations or experiments. One of the most common approaches to compare competing models representing alternative hypotheses relies on Akaike's Information Criterion (AIC) [1]. For a given data set \mathcal{D} , the plausibility of the candidate models \mathcal{M}_i is assessed by calculating their respective AIC values, AIC_i :

$$AIC_i = -2 \ln p(\mathcal{D} | \hat{\theta}_{MLE,i}, \mathcal{M}_i) + 2n_{df,i}. \quad (1)$$

In (1), $\hat{\theta}_{MLE,i}$ is the maximum likelihood estimate of the set parameters associated with model \mathcal{M}_i , and n_{df} is the corresponding number of degrees of freedom. If $AIC_1 < AIC_2$ then \mathcal{M}_1 is more plausible than \mathcal{M}_2 , with respect to \mathcal{D} , in the sense that the Kullback-Liebler divergence of \mathcal{M}_1 from the true model is smaller [2].

An important drawback to the classic approach to model choice is that it is based on a single point estimate $\hat{\theta}_{MLE,i}$ of θ , the uncertainty in θ being ignored. In contrast, the Bayesian approach considers a probability distribution for θ , with $p(\theta_{(i)} | \mathcal{D}, \mathcal{M}_i)$ expressing the uncertainty in $\theta_{(i)}$ given \mathcal{D} (for a model $\mathcal{M}_{(i)}$).

Suppose that we wish to select a model from a set of candidate models $\{\mathcal{M}_1, \dots, \mathcal{M}_m\}$ given our observation of data \mathcal{D} . We can express this goal probabilistically by stating that the aim is to determine the most probable model: $\arg \max_{\mathcal{M}_i} p(\mathcal{M}_i | \mathcal{D})$.

From Bayes' theorem, we have

$$p(\mathcal{M}_i | \mathcal{D}) = \frac{p(\mathcal{M}_i)p(\mathcal{D} | \mathcal{M}_i)}{\sum_{j=1}^m p(\mathcal{M}_j)p(\mathcal{D} | \mathcal{M}_j)}; \quad (2)$$

therefore, if $p(\mathcal{M}_i)$ is known, or considered to be equal for all \mathcal{M}_i then the focus is on the model evidence $p(\mathcal{D} | \mathcal{M}_i)$.

If $\theta_{(i)}$ is the set of parameters associated with model \mathcal{M}_i , the Bayesian approach to $p(\mathcal{D} | \mathcal{M}_{(i)})$ is to integrate over all possible values of $\theta_{(i)}$:

$$p(\mathcal{D} | \mathcal{M}_{(i)}) = \int_{\theta_{(i)}} p(\mathcal{D} | \theta_{(i)}, \mathcal{M}_i) p(\theta_{(i)} | \mathcal{M}_i) d\theta_{(i)}. \quad (3)$$

In addition to allowing for parameter uncertainty, (3) intrinsically penalizes against models that are better able to fit to observed data because of their complexity [3], thereby removing the need for an explicit complexity penalization term.

The integral of (3) can be estimated analytically or numerically. In analytical approaches, the integral is approximated by the adoption of simplifying assumptions; for example, as used for

derivation of the Bayes Information Criterion [4]. Numerical approaches are based on some form of Monte Carlo sampling such as Gibbs Sampling [5].

One approach to estimating the integral

$$\int_{\mathbf{q}} \underbrace{\mathcal{L}(\theta)}_{p(\mathcal{D}|\theta, \mathcal{M})/t} / \underbrace{\pi(\theta)}_{p(\theta|\mathcal{M})/t} d\theta$$

numerically is to sample θ randomly from its prior,

$$\int_{\mathbf{q}} \mathcal{L}(\theta)\pi(\theta) d\theta \approx \frac{1}{n} \sum_{k=1}^n \mathcal{L}(\theta_k), \text{ where } \theta_k \sim \pi; \quad (4)$$

however, the prior $\pi(\theta)$ is often concentrated in places where the likelihood $\mathcal{L}(\theta)$ is relatively low. This problem becomes more severe in high-dimensional parameter (θ) spaces, or in problems where the likelihood function $\mathcal{L}(\theta)$ is concentrated in a very small region.

To overcome the problem, Skilling [6,7] proposed a means of estimating $\int_{\mathbf{q}} \mathcal{L}(\theta)\pi(\theta)d\theta$ that, by design, samples θ sparsely from the θ space where the likelihood $\mathcal{L}(\theta)$ is low, and densely where $\mathcal{L}(\theta)$ is high, by means of ‘nested sampling’, which is the focus of this paper. A recent addition to the Bayesian arsenal, nested sampling has been used in cosmology to compare alternative models of the universe against observed data [8]. Outside of physics, it has, so far, received little attention [9,10].

Within-host dynamics of a bacterial infection

Quantitative research on infectious disease dynamics has undergone rapid development over the last two decades, motivated by concerns about emerging infections that can spread globally and about the evolution of pathogens resistant to existing control measures such as antimicrobials and vaccines. Bayesian computation has become the method of choice to fit stochastic dynamic models to epidemiological [11] or experimental datasets [12]. This is in large part due to the appeal of being able to produce measures of uncertainty and correlation for the model parameters based on their posterior probability distributions. Similarly, models for within-host dynamics of infection have more recently started to benefit from Bayesian inference approaches [13].

Salmonella enterica causes systemic diseases (typhoid and paratyphoid fever) [14], food-borne gastroenteritis and non-typhoidal septicaemia (NTS) [15] in humans and in many other animal species world-wide, which also cause a very serious problem for the food industry. The global burden of typhoid fever is estimated at ca. 22 million cases with a mortality estimated at ca. 200,000 deaths per year [14,16]. Paratyphoid has an estimated 5.4 million illnesses worldwide [16]. The high incidence of these diseases, that affect both travellers to and residents in endemic areas, and threaten infants, children and immunodeficient patients, dictates the urgent need for more efficacious preventive and therapeutic measures.

In the mouse model of systemic infection, *Salmonella* reside and proliferate mainly within phagocytic cells of the spleen liver, bone marrow and lymph nodes [17–19]. Observation of *Salmonella* by fluorescence microscopy in the tissues of mice has revealed that a key feature of systemic infections with wild type bacteria is the presence, on average, of low bacterial numbers within individual phagocytes irrespective of net bacterial growth rate and time since infection [20–23].

In an effort to understand the dynamics that underpin the intracellular numerical distributions of *Salmonella* within the host cells, and to capture the essential traits of the cell-to-cell spread of the bacteria, we have used mathematical model frameworks for the intensity of intracellular infection that links the quasi-stationary distribution of bacteria to bacterial and cellular demography. An example of this the work done by Brown et al. [24], who compared the observed distribution $\{C_n\}$, where C_n is the number of cells with n bacteria, across 16 candidate infection models. The models under consideration were as follows: (a) one homogeneous model, in which, for every cell, burst occurred only when the number of bacteria n in a cell reached a single burst threshold \mathcal{N}_{max} ; (b) five heterogeneous models having a probability distribution of burst thresholds; and (c) eight stochastic models for which there is a probability that a given cell will undergo burst. Two datasets were analysed, one for a virulent strain of bacteria and the other for an attenuated strain. Brown et al. [24] computed the maximum likelihood estimates of the parameters of each model, and selected the ‘best’ model based on the corresponding AIC values.

In order to overcome the issues raised by AIC discussed above, we decided to re-analyse the datasets and re-assess the models within a Bayesian framework.

Methods

What follows is an elaboration of the description of nested sampling given by Skilling [6,7].

Nested sampling

The expected value of a function g of a random variable X is given by

$$\mathbb{E}(g(\mathbf{X})) = \int_{\mathbf{x}} g(\mathbf{x})f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}$$

where $f_{\mathbf{X}}$ is the pdf of \mathbf{X} . On comparing this expression with the target integral $\int_{\theta} \mathcal{L}(\theta)\pi(\theta)d\theta$, it is clear that

$$\int_{\mathbf{q}} \mathcal{L}(\theta)\pi(\theta) d\theta = \mathbb{E}(\mathcal{L}(\theta)); \quad (5)$$

that is to say, the expected value of the likelihood under the prior. The cumulative distribution function $F_X(x)$ with respect to a random variable X is defined by

$$F_X(x) = p(X < x) = \int_{-\infty}^x f_X(y) dy$$

and is related to the expectation $\mathbb{E}(X)$ by

$$\mathbb{E}(X) = \int_0^{\infty} (1 - F_X(x)) dx,$$

[25]; consequently, from (5), we obtain the important relationship

$$\int_{\mathbf{q}} \mathcal{L}(\theta)\pi(\theta)d\theta = \int_0^{\infty} (1 - F_{\mathcal{L}}(\lambda))d\lambda = \int_0^1 (1 - F_{\mathcal{L}}(\lambda))d\lambda, \quad (6)$$

where λ is likelihood, and \mathcal{L} in the right-hand integral is equal to $\mathcal{L}(\theta)$. The reason why (6) is important is that the multivariate integral on the left-hand side has been equated to a univariate integral.

Since θ has a distribution defined by prior π , and $\mathcal{L} = \mathcal{L}(\theta)$, it follows that \mathcal{L} has a probability distribution and thus a cumulative distribution function,

$$F_{\mathcal{L}}(\lambda) = p(\mathcal{L} < \lambda) = \int_0^\lambda f_{\mathcal{L}}(y) dy, \tag{7}$$

which is present in the integrand of the right-hand integral of (6).

We can replace $\int_0^1 (1 - F_{\mathcal{L}}(\lambda)) d\lambda$ in (6) with a more accessible integral by the following steps. First, since the pdf of \mathcal{L} is connected to the pdf of θ via $\mathcal{L}(\theta)$, we can write

$$\int_0^\lambda f_{\mathcal{L}}(y) dy = \int_{\theta: \mathcal{L}(\theta) < \lambda} \pi(\theta) d\theta; \tag{8}$$

thus, from (6), (7) and (8), we can write

$$\begin{aligned} \int_{\mathbf{q}} \mathcal{L}(\theta) \pi(\theta) d\theta &= \int_{\lambda=0}^{\lambda=1} \left(1 - \int_{\theta: \mathcal{L}(\theta) < \lambda} \pi(\theta) d\theta \right) d\lambda \\ &= \int_{\lambda=0}^{\lambda=1} \int_{\theta: \mathcal{L}(\theta) > \lambda} \pi(\theta) d\theta d\lambda. \end{aligned} \tag{9}$$

It will be convenient to rewrite the inner integral of (9) as $\phi(\lambda)$ to give

$$\int_{\theta} \mathcal{L}(\theta) \pi(\theta) d\theta = \int_0^1 \phi(\lambda) d\lambda, \tag{10}$$

where $\phi(\lambda)$ is the probability of selecting θ from the prior π such that $\mathcal{L}(\theta) > \lambda$:

$$\phi(\lambda) = \int_{\theta: \mathcal{L}(\theta) > \lambda} \pi(\theta) d\theta. \tag{11}$$

Introducing $\xi = \phi(\lambda)$, hence $\lambda = \phi^{-1}(\xi)$, we can rewrite the previous integral as

$$\int_{\mathbf{q}} \mathcal{L}(\theta) \pi(\theta) d\theta = \int_0^1 \phi^{-1}(\xi) d\xi, \tag{12}$$

where $\phi^{-1}(\xi)$ is that likelihood λ such that $p(\mathcal{L}(\theta) > \lambda) = \xi$ (cf. Equation (11)); for example, if $\phi^{-1}(0.9) = 0.0042$ then 90% of θ drawn from the prior $\pi(\theta)$ will have likelihoods greater than 0.0042.

The algorithm

The main steps of the nested sampling technique are as follows. First, n points θ (i.e., parameter vectors) are sampled from the prior π , and their corresponding likelihoods $\mathcal{L}(\theta)$ determined. The point $\theta_{min,1}$ having the smallest likelihood is determined and its likelihood $\lambda_{min,1}$ is recorded. Furthermore, the probability ξ_1 that $\mathcal{L}(\theta) > \lambda_{min,1}$ is also recorded. Point $\theta_{min,1}$ is replaced by a new θ drawn from the prior π but restricted to those θ for which $\mathcal{L}(\theta) > \lambda_{min,1}$. In other words, a *restricted prior* is used: $\pi(\mathcal{L}(\theta) > \lambda_{min,1})$. If Ω is the set of all possible θ then the set $\Omega_1 = \{\theta | \mathcal{L}(\theta) > \lambda_{min,1}\}$ is a subset of Ω .

The above sequence of determining θ_{min} and the corresponding ξ is performed on the new set of points, giving rise to $\lambda_{min,2}$ and ξ_2 . Point $\theta_{min,2}$ is replaced by a θ drawn from the new restricted prior $\pi(\mathcal{L}(\theta) > \lambda_{min,2})$. In other words, θ is sampled from $\Omega_2 = \{\theta | \mathcal{L}(\theta) > \lambda_{min,2}\}$, for which $\Omega_2 \subset \Omega_1$.

This cycle is repeated until some stopping criterion has been reached. If this termination occurs at the J -th iteration then the resulting values of $\lambda_{min,i}$ and ξ_i will be

$$\lambda_{min,J} > \lambda_{min,J-1} > \dots > \lambda_{min,1},$$

$$\xi_J < \xi_{J-1} < \dots < \xi_1,$$

and the resulting sequence of Ω subsets is

$$\Omega_J \subset \Omega_{J-1} \subset \dots \subset \Omega_1 \subset \Omega;$$

hence the term *nested sampling*.

Model evidence $\mathcal{Z} = \int_{\mathbf{q}} \mathcal{L}(\theta) \pi(\theta) d\theta$ can be estimated from the recorded $\lambda_{min,i}$ and ξ_i values by means of the approximation

$$\mathcal{Z} = \int_0^1 \phi^{-1}(\xi) d\xi \approx \sum_{i=1}^J \Delta \widehat{\mathcal{Z}}_i, \tag{13}$$

where J is the number of iterations used, and $\Delta \widehat{\mathcal{Z}}$ is a vertical rectangular segment under the curve of Figure 1.

Algorithm 1 (Table 1) describes the above process in pseudocode.

Practical adjustments to the algorithm

We will now consider how some of the aspects of Algorithm 1 can be implemented.

Segment $\Delta \widehat{\mathcal{Z}}_i$ used in (13) could be evaluated by the trapezoidal approach

$$\Delta \widehat{\mathcal{Z}}_i = \lambda_{min,i} (\xi_{i-1} - \xi_{i+1}) / 2$$

but Sivia and Skilling [26] have found

$$\Delta \widehat{\mathcal{Z}}_i = \lambda_{min,i} (\xi_{i-1} - \xi_i)$$

to be adequate (line 9 in Algorithm 1).

Line 7 in Algorithm 1 used the assignment $\xi_i \leftarrow P(\mathcal{L}(\theta) > \lambda_{min,i})$, but an alternative approach is to replace this assignment with $\xi_i \leftarrow \mathbb{E}[\xi_i]$. An approximation of $\mathbb{E}[\xi_i]$ is derived as follows. Let τ_i denote the ratio ξ_i / ξ_{i-1} , with $\xi_0 = 1$. At the k th iteration we have

$$0 < \xi_J < \xi_{J-1} < \dots < \xi_1 < 1,$$

and so

$$\xi_k = \tau_k \xi_{k-1} = \prod_{i=1}^k \tau_i,$$

therefore,

$$\mathbb{E}[\log \xi_k] = \sum_{i=1}^k \mathbb{E}[\log \tau_i]. \tag{14}$$

Now,

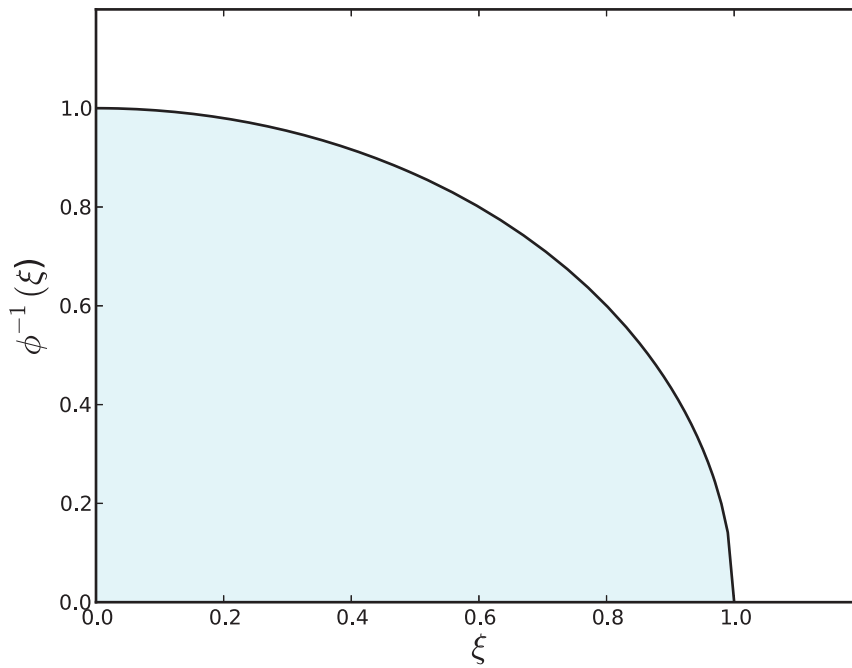


Figure 1. The shaded area below the curve for $\phi^{-1}(\xi)$ is equal to $\int_{\theta} \mathcal{L}(\theta)\pi(\theta) d\theta$. See Equation (12).
doi:10.1371/journal.pone.0082317.g001

$$\begin{aligned} \mathbb{E}[\log \tau_i] &= \int_0^1 \log \tau p(\tau) d\tau &&= [\tau^n \log \tau]_0^1 - \left[\frac{\tau^n}{n}\right]_0^1 \\ &= \int_0^1 \log \tau n\tau^{n-1} d\tau &&= \frac{-1}{n}, \end{aligned}$$

[27]

therefore, from (14),

Table 1. Algorithm 1: The nested sampling algorithm.

Input: (a) likelihood function $\mathcal{L}(\theta)$; (b) prior $\pi(\theta)$; (c) number n of active parameter vectors in use during nested sampling.

Out put: an estimate $\hat{\mathcal{Z}}$ of \mathcal{Z} .

1: Let S be a set of n parameter vectors $\theta_1, \dots, \theta_n \stackrel{iid}{\sim} \pi$

2: $\hat{\mathcal{Z}} \leftarrow 0$

3: $i \leftarrow 1$

4: **while** terminating condition not satisfied **do**

5: $\theta_{min,i} \leftarrow \arg \min_{\theta \in S} \mathcal{L}(\theta)$

6: $\lambda_{min,i} \leftarrow \mathcal{L}(\theta_{min,i})$

7: $\xi_i \leftarrow P(\mathcal{L}(\theta) > \lambda_{min,i})$

8: **if** $i > 1$ **then**

9: $\Delta \mathcal{Z}_i \leftarrow \lambda_{min,i}(\xi_{i-1} - \xi_i)$ ▷ Estimated segment of \mathcal{Z}

10: $\hat{\mathcal{Z}} \leftarrow \hat{\mathcal{Z}} + \Delta \mathcal{Z}_i$

11: $\theta_{new} \sim \pi | (\mathcal{L}(\theta) > \lambda_{min,i})$ ▷ Restricted prior

12: $S \leftarrow S - \{\theta_{min,i}\}$

13: $S \leftarrow S \cup \{\theta_{new}\}$

return $\hat{\mathcal{Z}}$

doi:10.1371/journal.pone.0082317.t001

Table 2. Algorithm 2: An implementation of Algorithm 1 in which practical adjustments are included.

Input (a) likelihood function $\mathcal{L}(\theta)$; (b) prior $\pi(\theta)$; (c) number n of active parameter vectors in use during nested sampling; (d) procedure for determining a region $\mathcal{R}(\mathbf{S})$ of parameter space that encloses a set of parameter vectors \mathbf{S} ; (e) fraction f of \mathcal{Z} to be estimated.

Output: an estimate $\hat{\mathcal{Z}}$ of \mathcal{Z} .

```

1: Let  $\mathbf{S}$  be a set of  $n$  parameter vectors  $\theta_1, \dots, \theta_n \stackrel{iid}{\sim} \pi$ 
2:  $\hat{\mathcal{Z}} \leftarrow 0$ 
3:  $i \leftarrow 1$ 
4: Repeat
5:  $\theta_{min,i} \leftarrow \arg \min_{\theta \in \mathbf{S}} \mathcal{L}(\theta)$ 
6:  $\lambda_{min,i} \leftarrow \mathcal{L}(\theta_{min,i})$ 
7:  $\xi_i \leftarrow P(\mathcal{L}(\theta) > \lambda_{min,i})$ 
8: if  $i > 1$  then
9:  $\Delta \mathcal{Z}_i \leftarrow \lambda_{min,i}(\xi_{i-1} - \xi_i)$  ▷ Estimated segment of  $\mathcal{Z}$ 
10:  $\hat{\mathcal{Z}} \leftarrow \hat{\mathcal{Z}} + \Delta \mathcal{Z}_i$ 
11:  $\theta_{new} \sim \pi(\theta \in \mathcal{R}(\mathbf{S}) \wedge \mathcal{L}(\theta) > \lambda_{min,i})$  ▷ Restricted prior
12:  $\mathbf{S} \leftarrow \mathbf{S} - \{\theta_{min,i}\}$ 
13:  $\mathbf{S} \leftarrow \mathbf{S} \cup \{\theta_{new}\}$ 
14: until:  $\lambda_{min,i} \xi_i < f \hat{\mathcal{Z}}$  ▷ The stopping condition
return  $\hat{\mathcal{Z}}$ 

```

doi:10.1371/journal.pone.0082317.t002

$$\mathbb{E}[\log \xi_k] = \sum_{i=1}^k \left(\frac{-1}{n} \right) = \frac{-k}{n}. \qquad \mathbb{E}[\xi_k] \approx \exp\left(\frac{-k}{n}\right).$$

Since the logarithm function is strictly increasing and concave, we have, from Jensen’s inequality, that

$$\log \mathbb{E}[\xi_k] \leq \mathbb{E}[\log \xi_k]$$

and thus

$$\mathbb{E}[\xi_k] \leq \exp\left(\frac{-k}{n}\right);$$

however, Sivia and Skilling [26, p. 186] drop the inequality and use the approximation

As regards the termination of Algorithm 1, there is no rigorous criterion as to when the algorithm should be stopped, but Skilling [7] and Feroz and Hobson [28] have found

$$\lambda_{min,i} \xi_i < f \hat{\mathcal{Z}}$$

to be an effective stopping condition, where f is the fraction of \mathcal{Z} that will not significantly contribute to the estimate of \mathcal{Z} (according to a user-defined value).

Chopin and Robert [29] have shown that the asymptotic variance of the nested sampling approximation typically grows linearly with parameter dimensions.

Table 3. Probability distributions $d(\mathcal{N}|\theta)$ for the burst thresholds \mathcal{N} .

Model	Distribution	Parameters, θ
1	$d(\mathcal{N} \theta) = \delta_{\mathcal{N}, \mathcal{N}_{max}}$	$\mathcal{N}_{max} \in \{2, \dots, 40\}$
2	$d(\mathcal{N} \theta) = r \delta_{\mathcal{N}, \mathcal{N}_1} + (1-r) \delta_{\mathcal{N}, \mathcal{N}_2}$	$\mathcal{N}_1 \in \{2, \dots, 5\}, \mathcal{N}_2 \in \{2, \dots, 30\}, r \in [0, 1]$
3	$d(\mathcal{N} \theta) = (\lambda^{\mathcal{N}} / \mathcal{N}!) \exp(-\lambda)$	$\lambda \in (0, 40]$
4	$d(\mathcal{N} \theta) = \binom{t}{\mathcal{N}} p^{\mathcal{N}} (1-p)^{t-\mathcal{N}}$	$t \in \{30, \dots, 45\}, p \in [0, 1]$
5	$d(\mathcal{N} \theta) = \binom{\mathcal{N}+r-1}{\mathcal{N}} p^{\mathcal{N}} (1-p)^r$	$r \in \{1, \dots, 30\}, p \in (0, 1)$
6	$d(\mathcal{N} \theta) = p(1-p)^{\mathcal{N}-1}$	$p \in (0, 1]$

(1) Unimodal Kronecker, (2) bimodal Kronecker, (3) Poisson, (4) binomial, (5) negative binomial, and (6) geometric.

doi:10.1371/journal.pone.0082317.t003

Table 4. Parameters used for the eight stochastic models based on (15) and (16).

Model	Parameters, θ				
	μ_0	μ_1	μ_2	α_0	α_e
7	μ_0	0	0	1	0
8	0	μ_1	0	1	0
9	0	0	μ_0	1	0
10	μ_0	μ_1	μ_2	1	0
11	μ_0	0	0	1	α_e
12	0	μ_1	0	1	α_e
13	0	0	μ_2	1	α_e
14	μ_0	μ_1	μ_2	1	α_e

For each model, some of the parameters were set equal to constant values, which effectively removed the parameters from the model. The range of values considered were $\mu_0, \mu_1, \mu_2 \in [0, 2]$ and $\alpha_e \in [0, 1]$.
doi:10.1371/journal.pone.0082317.t004

Finally, there is the structure of the restricted priors. Each new point θ_{new} for a set \mathbf{S} of active points is sampled from prior π conditioned on the restriction that $\mathcal{L}(\theta_{new}) > \lambda_{min}$. Rather than searching across the entire θ -space for such a point, it is more computationally efficient to restrict the search to a region $\mathcal{R}(\mathbf{S})$ that contains \mathbf{S} . We have used rectangular cuboids for $\mathcal{R}(\mathbf{S})$.

Incorporating the above points into Algorithm 1 leads to Algorithm 2 (Table 2). Before applying the algorithm to our experimental datasets, we tested it on a simple two-parameter likelihood function $\mathcal{L}(\alpha, \beta) = \alpha^{30} (1 - \alpha)^{30} \beta^{30} (1 - \beta)^{30}$. The analyses and results are presented in Methods S1.

The Salmonella models

Evidence $p(\mathcal{D}|\mathcal{M})$ was estimated by nested sampling with respect to two groups of models associated with within-host *S. enterica* infection, where each model \mathcal{M} provides an expression for the probability $q(n|\theta, \mathcal{M})$ that a host cell contains n bacteria.

In the first group of models, infected cells are assumed to burst when the number of bacteria they contain reach a fixed threshold

Table 5. Algorithm 3: Estimation of $q(n|\theta, \mathcal{M})$ using an iterative estimation of the infection rate constant γ .

Input: parameters θ for model \mathcal{M} .
Output: an estimate of probabilities $q(1 \theta, \mathcal{M}), q(2 \theta, \mathcal{M}), \dots, q(n_{max} \theta, \mathcal{M})$.
1: $\gamma \leftarrow 1$ Initial value for γ
2: $\gamma_{old} \leftarrow 10\gamma$
3: while: $ (\gamma - \gamma_{old})/\gamma_{old} > 0.01$ do
4: $\gamma_{old} \leftarrow \gamma$
5: $F \leftarrow [f_1, f_2, f_3, \dots, f_2 \dots f_{max}]$
where $f_n = (\alpha_{n-1}(n-1))/(\gamma + \alpha_n n + \mu(n; \mu_0, \mu_1, \mu_2))$ ▷Equation (21)
6: $q_1 \leftarrow (\sum_{n=1}^{n_{max}} F[n])^{-1}$ ▷Estimate of $q(1 \theta, \mathcal{M})$
7: $P \leftarrow q_1 F$ ▷Estimates of $q(n \theta, \mathcal{M})$ where $n = 1, \dots, n_{max}$
8: $P \leftarrow P / \sum_{n=1}^{n_{max}} P[n]$ ▷Normalisation of the estimated probabilities
9: $\gamma \leftarrow (\sum_{j=1}^{n_{max}} j^2 q(j \theta, \mathcal{M})) / (\sum_{j=1}^{n_{max}} j q(j \theta, \mathcal{M}))$
return: $P = [q(1 \theta, \mathcal{M}), q(2 \theta, \mathcal{M}), \dots, q(n_{max} \theta, \mathcal{M})]$

doi:10.1371/journal.pone.0082317.t005

Table 6. The number C_n of cells containing n bacteria when virulent (SL5560) and attenuated (SL3261) strains of bacteria were used.

n	C_n	
	Virulent	Attenuated
1	655	1189
2	250	396
3	87	104
4	86	70
5	54	40
6	42	25
7	13	8
8	30	10
9	8	9
10	19	3
11	5	7
12	12	4
13	5	3
14	1	4
15	6	0
16	3	2
17	2	1
18	0	2
19	1	1
20	4	0
21	0	0
22	0	0
23	0	0
24	0	1
25	1	0
26	0	0
27	0	0
28	0	0
29	1	0

doi:10.1371/journal.pone.0082317.t006

\mathcal{N} . The probability distributions considered for \mathcal{N} are shown in Table 3.

For the second group of models, the assumption is that, instead of pre-programmed burst thresholds \mathcal{N} , there is burst rate μ that is a function of the number of bacteria n in a cell. For these models, the general relationship is

$$\mu(n; \mu_0, \mu_1, \mu_2) = \mu_0 + \mu_1 n + \mu_2 n^2 \tag{15}$$

where $\mu_0, \mu_1, \mu_2 \in [0, \infty)$. Furthermore, the rate of bacterial replication α_n is assumed to be related to n by

$$\alpha_n = \alpha_0 \exp(-\alpha_e n) \tag{16}$$

where $\alpha_0, \alpha_e \in [0, \infty)$. As explained in Brown et al. [24], in the dynamic model, time can be re-scaled by the baseline replication rate α_0 , therefore this parameter cannot be estimated using the quasi-stationary distribution. For convenience, we set $\alpha_0 = 1$, so

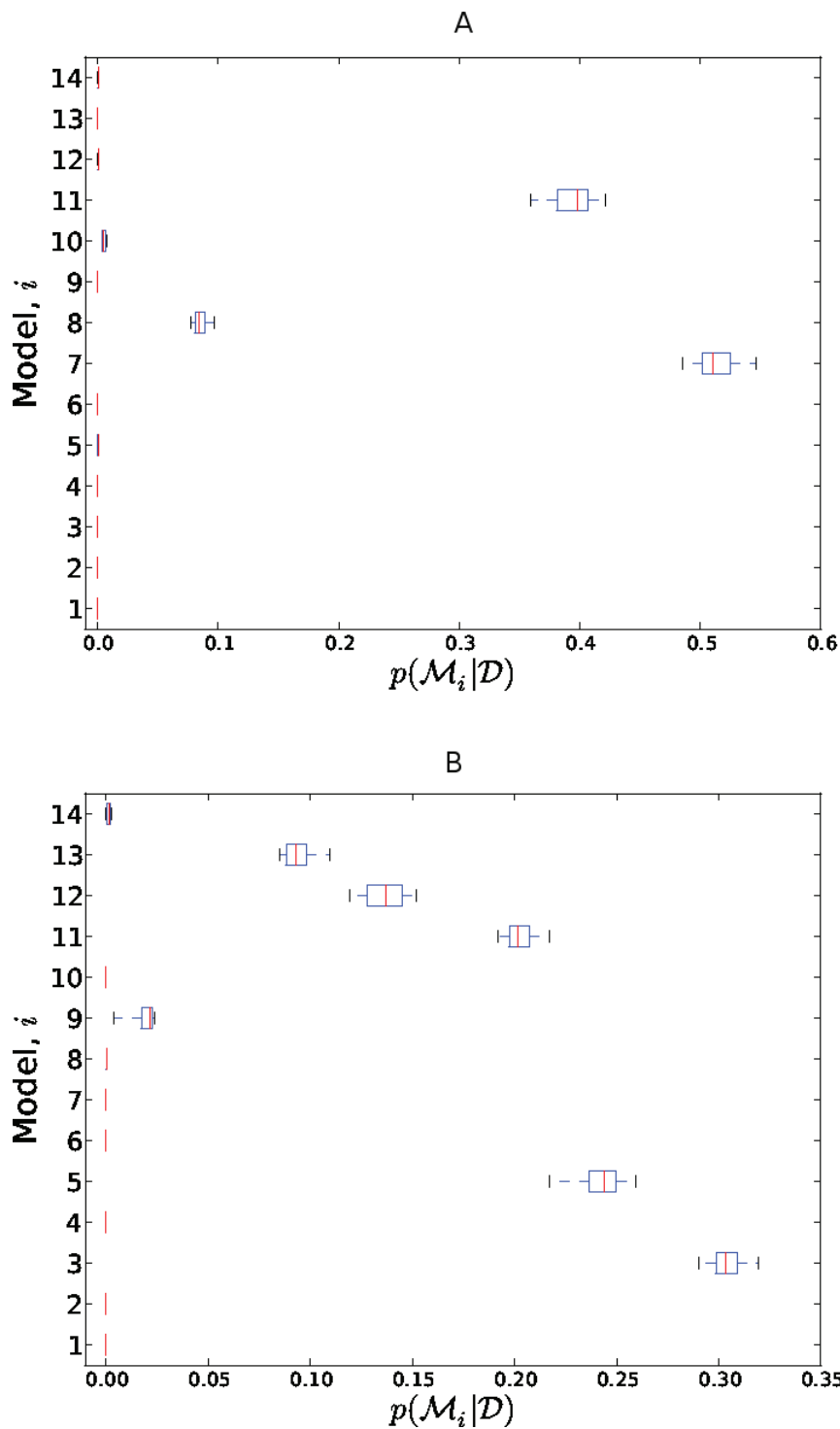


Figure 2. Estimates of the posterior model probabilities $p(\mathcal{M}|\mathcal{D})$ when using data from (A) the attenuated strain and (B) the virulent strain.

doi:10.1371/journal.pone.0082317.g002

that the values of other parameters are relative to the baseline replication rate. The parameters of the eight stochastic models considered are shown in Table 4.

Under the assumption that the number of host cells infected by n bacteria reaches a quasi-stationary distribution, the probability $q(n|\theta, \mathcal{M})$ that a cell contains n bacteria can be derived for the 14

models [30]. For Model 1, we have the relationship

$$q(n|\theta, \mathcal{M}) = \frac{\mathcal{N}_{max}}{(\mathcal{N}_{max} - 1)n(n + 1)}. \quad (17)$$

For Models 2 to 6, the relationship is

Table 7. Median $-\log_{10}(\hat{\mathcal{Z}})$ estimated for Models 1 to 6.

Model	Distribution	Attenuated	Virulent
1	$d(\mathcal{N} \theta) = \delta_{\mathcal{N}, \mathcal{N}_{max}}$	77.59	38.56
2	$d(\mathcal{N} \theta) = r\delta_{\mathcal{N}, \mathcal{N}_1} + (1-r)\delta_{\mathcal{N}, \mathcal{N}_2}$	69.49	92.79
3	$d(\mathcal{N} \theta) = (\lambda \mathcal{N} / \mathcal{N}!) \exp(-\lambda)$	53.75	34.09
4	$d(\mathcal{N} \theta) = \binom{t}{\mathcal{N}} p^{\mathcal{N}} (1-p)^{t-\mathcal{N}}$	245.87	281.46
5	$d(\mathcal{N} \theta) = \binom{\mathcal{N}+r-1}{\mathcal{N}} p^{\mathcal{N}} (1-p)^r$	30.26	34.18
6	$d(\mathcal{N} \theta) = p(1-p)^{\mathcal{N}-1}$	84.26	79.97

The highest model evidence $\hat{\mathcal{Z}}$ (bold) and second highest model evidence (italic) models are highlighted.

doi:10.1371/journal.pone.0082317.t007

$$q(n|\theta, \mathcal{M}) \propto \frac{1}{n(n+1)} \sum_{\mathcal{N}=n+1}^{\infty} d(\mathcal{N}|\theta). \quad (18)$$

For Models 7 to 16, we have the recursive relationship

$$q(n|\theta, \mathcal{M}) = \frac{\alpha_{n-1}(n-1)}{\gamma + \alpha_n n + \mu(n; \mu_0, \mu_1, \mu_2)} q(n-1|\theta, \mathcal{M}), \quad (19)$$

where the infection rate constant γ is given by

$$\gamma = \frac{\sum_{j=1}^{\infty} j \alpha_j q(j|\theta, \mathcal{M})}{\sum_{j=1}^{\infty} j q(j|\theta, \mathcal{M})}. \quad (20)$$

The value for $q(1|\theta, \mathcal{M})$ can be handled as follows. Let

$$f_n = \frac{\alpha_{n-1}(n-1)}{\gamma + \alpha_n n + \mu(n; \mu_0, \mu_1, \mu_2)}, \quad (21)$$

so that (19) can be written as $q(n|\theta, \mathcal{M}) = f_n q(n-1|\theta, \mathcal{M})$, then

$$\begin{aligned} \sum_{n=1}^{\infty} q(n|\theta, \mathcal{M}) &= q(1|\theta, \mathcal{M}) + q(2|\theta, \mathcal{M}) + q(3|\theta, \mathcal{M}) + \dots \\ &= q(1|\theta, \mathcal{M}) + f_2 q(1|\theta, \mathcal{M}) + f_2 f_3 q(1|\theta, \mathcal{M}) + \dots \\ &= q(1|\theta, \mathcal{M}) (1 + f_2 + f_2 f_3 + \dots) \end{aligned}$$

but $\sum_{n=1}^{\infty} q(n|\theta, \mathcal{M}) = 1$; therefore,

$$q(1|\theta, \mathcal{M}) = (1 + f_2 + f_2 f_3 + \dots)^{-1}.$$

When bacterial replication is not dependent on n , $\alpha_e = 0$, in which case $\gamma = 1$, but when replication is density dependent, (19) and (20) need to be solved self-consistently. This can be done by assuming an initial value for γ , computing $q(n|\theta, \mathcal{M})$ from (19), updating γ using (20), and repeating this iteratively until γ no longer changes significantly. This process is shown in Algorithm 3 (Table 5).

Table 8. Median $-\log_{10}(\hat{\mathcal{Z}})$ estimated for stochastic Models 7 to 14.

Model	Parameters, θ					Attenuated	Virulent
	μ_0	μ_1	μ_2	α_0	α_e		
7	μ_0	0	0	1	0	27.21	38.56
8	0	μ_1	0	1	0	28.00	36.93
9	0	0	μ_2	1	0	38.80	35.24
10	μ_0	μ_1	μ_2	1	0	29.21	39.21
11	μ_0	0	0	1	α_e	27.32	34.27
12	0	μ_1	0	1	α_e	30.13	34.43
13	0	0	μ_2	1	α_e	41.25	34.60
14	μ_0	μ_1	μ_2	1	α_e	30.04	36.34

The highest model evidence $\hat{\mathcal{Z}}$ (bold) and second highest model evidence (italic) models are highlighted.

doi:10.1371/journal.pone.0082317.t008

Likelihood function

With expressions for $q(n|\theta, \mathcal{M})$ established for all the models, we can now determine the likelihood $\mathcal{L}(\theta)$ required for Algorithm 2. Following Brown et al. [30], we can express the likelihood function by a multinomial distribution:

$$\mathcal{L}(\theta) = p(\mathcal{D}|\theta, \mathcal{M}) \quad (22)$$

$$= p(\{C_n\} | \{q(n|\theta, \mathcal{M})\}) \quad (23)$$

$$= Y! \prod_{n=1}^{n_{max}} \frac{q(n|\theta, \mathcal{M})^{C_n}}{C_n!} \quad (24)$$

where $\{C_n\}$ is the observed distribution of C_n (the number of cells with n bacteria), and $Y = \sum_n C_n$, if observations are assumed to be

Table 9. $-\log_{10}(\hat{\mathcal{Z}})$ estimates for all models.

Model	Attenuated			Virulent		
	min	median	max	min	median	max
1	77.56	77.59	77.63	38.55	38.56	38.58
2	69.38	69.49	69.66	92.66	92.79	92.88
3	53.71	53.75	53.79	34.07	34.09	34.10
4	245.83	245.87	245.91	281.36	281.46	281.50
5	29.93	30.26	30.52	34.16	34.18	34.24
6	84.23	84.26	84.30	79.93	79.97	80.01
7	27.19	27.21	27.24	38.52	38.56	38.58
8	27.94	28.00	28.02	36.88	36.93	36.97
9	38.78	38.80	38.85	35.20	35.24	35.98
10	29.06	29.21	29.39	38.66	39.21	43.12
11	27.28	27.32	27.38	34.24	34.27	34.28
12	29.99	30.13	30.36	34.39	34.43	34.50
13	40.93	41.25	41.84	34.53	34.60	34.63
14	29.86	30.04	30.48	36.19	36.34	39.65

doi:10.1371/journal.pone.0082317.t009

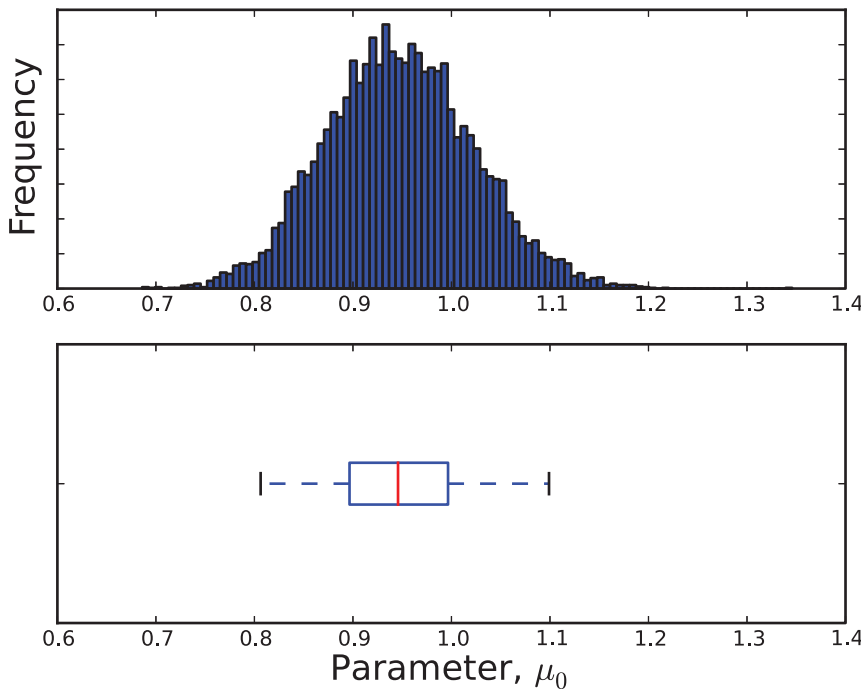


Figure 3. An estimate of the marginal probability distribution $p(\mu_0|\mathcal{D}, \text{Model7})$. \mathcal{D} is data from the attenuated strain. doi:10.1371/journal.pone.0082317.g003

independent. Garca-Pérez [31] provides an algorithm for the accurate computation of multinomial probabilities.

As regards the prior $\pi(\theta)$ for a model \mathcal{M} , it will be assumed to be uniform across the parameter space of interest for that model; consequently, the prior will be set equal to the reciprocal of the size of the parameter space. More precisely,

$$\pi(\theta) = \pi(\theta_1, \dots, \theta_D) = \left(\prod_{i=1}^D \max(\theta_i) - \min(\theta_i) \right)^{-1}.$$

A continuation approach

The theory underlying nested sampling assumes that all the parameters for a model have continuous values, however, this will not necessarily be the case in practice. For example, the binomial model (Model 3) has a discrete parameter n and a continuous parameter p .

It is possible to formulate a theory of nested sampling for discrete parameters by replacing integrals with summations, but modifications to Algorithm 2 would be required to take account of the fact that, if θ is discrete, several points could occupy the same location in parameter space.

An alternative response to the presence of discrete parameters is to use a type of *continuation approach* [32]; in other words, if $f(x)$ is a function defined only for integer values of x , replace it with another function $g(x)$ that takes real values, but for which $g(x) = f(x)$ when $x \in \mathbb{N}$ (or \mathbb{N}_0).

For Model 2, the Kronecker delta $\delta_{\mathcal{N}, \mathcal{N}_i}$ can be replaced with a narrow Gaussian function $\exp(-(\mathcal{N} - \mathcal{N}_i)^2)$ with $\mathcal{N}_i \in [1, \infty)$. In the case of Model 1, continuation can be applied directly to (17) by allowing $\mathcal{N}_{max} \in [1, \infty)$.

For those models using a factorial of a parameter (i.e., Models 4 and 5), we can replace $x!$ with $\Gamma(x+1)$ since Γ is a function of a real value.

The data

The data \mathcal{D} consisted of the number C_n of mice cells observed (via fluorescence microscopy) to contain n *S. enterica* bacteria: $\mathcal{D} = \{C_n\}_{n=1}^{29}$. One dataset was used for a virulent bacterial strain (SL5560); another for an attenuated strain (SL3261). The infected cells were taken randomly from various locations in the liver. The observed C_n values are shown in Table 6.

The data was pooled. If $C_n^{[t]}$ denotes the number of cells having n bacteria on day t then, for the virulent strain, Brown et al. [24] used $C_n = C_n^{[3]} + C_n^{[4]}$, and for the attenuated strain they used $C_n = C_n^{[4]} + C_n^{[6]} + C_n^{[10]} + C_n^{[12]}$.

Posterior model probabilities

If we assume that the set of candidate models is exhaustive, we can apply (2) to estimate the posterior probability $p(\mathcal{M}|\mathcal{D})$ for each model. Furthermore, if $p(\mathcal{M}_i)$ is assumed to be equal for all models, we can use

$$\hat{p}(\mathcal{M}_i|\mathcal{D}) = \frac{\hat{\mathcal{Z}}_i}{\sum_{j=1}^{14} \hat{\mathcal{Z}}_j}. \tag{25}$$

There are 14 models, each arbitrarily having 10 estimates of $\hat{\mathcal{Z}}$, but it is impractical to systematically apply each of the 10^{14} possible combinations of $\hat{\mathcal{Z}}$ to [25]; therefore, the $\hat{\mathcal{Z}}$ values were

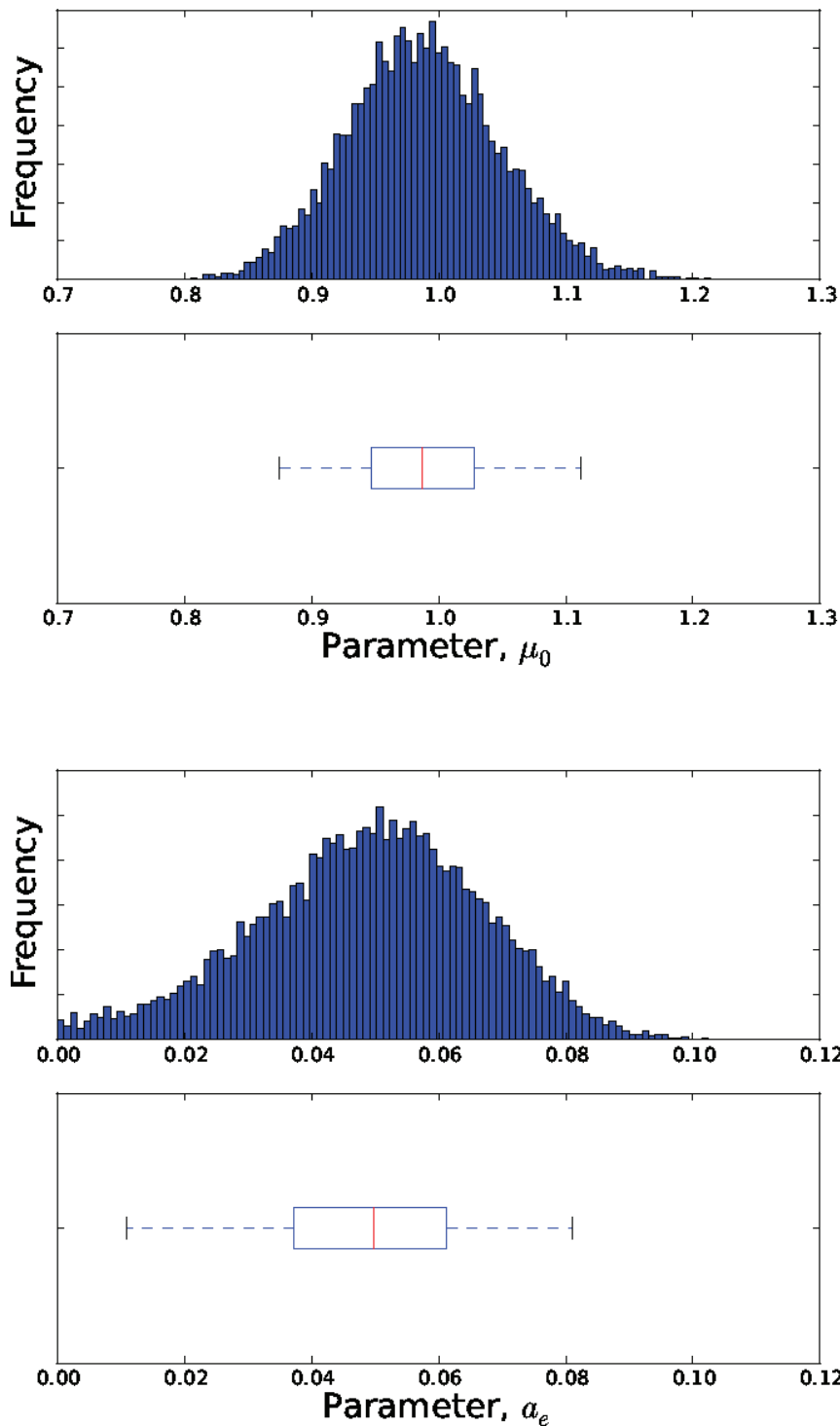


Figure 4. Estimates of the marginal probability distributions $p(\mu_0|\mathcal{D}, \text{Model11})$ and $p(a_e|\mathcal{D}, \text{Model11})$. \mathcal{D} is data from the attenuated strain. doi:10.1371/journal.pone.0082317.g004

chosen randomly in order to obtain distributions for $p(\mathcal{M}|\mathcal{D})$. The resulting distributions are shown in Figure 2.

An alternative approach to Bayesian model comparison is to use the Bayes factor $p(\mathcal{D}|\mathcal{M}_i)/p(\mathcal{D}|\mathcal{M}_j)$. This provides a relative comparison of models \mathcal{M}_i and \mathcal{M}_j but not the absolute values of their posterior probabilities $p(\mathcal{M}|\mathcal{D})$.

Results

The estimated model-evidence values $\hat{\mathcal{Z}}$ obtained by nested sampling for each model is shown in Tables 7 and 8. The ranges are shown in Table 9.

With respect to the data from the attenuated strain, the most probable model was Model 7 (μ_0 only) followed by Model 11 (μ_0 and a_e). With respect to the data from the virulent strain, the most

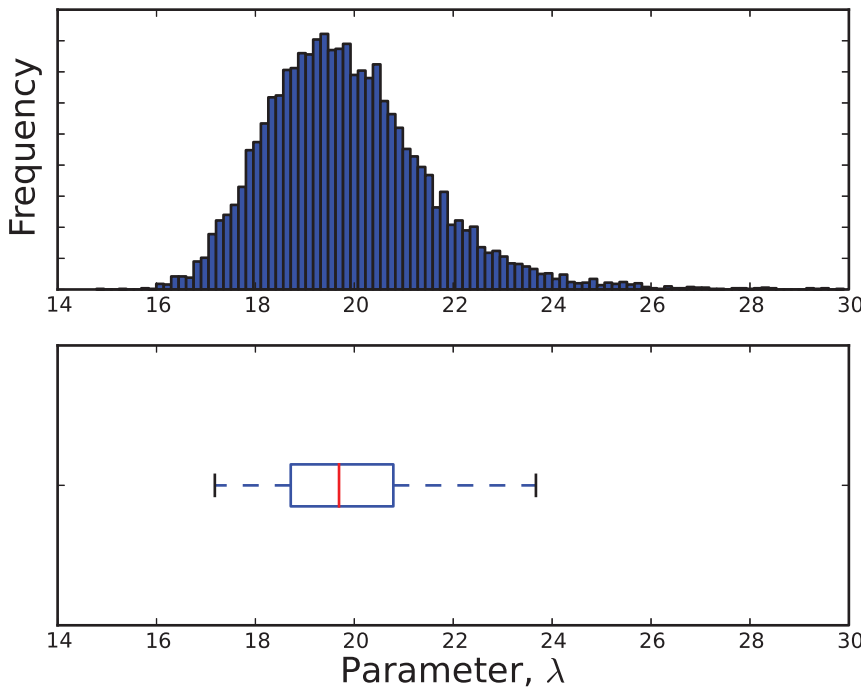


Figure 5. An estimate of the marginal probability distribution $p(\lambda|\mathcal{D}, \text{Model3})$. \mathcal{D} is data from the virulent strain. doi:10.1371/journal.pone.0082317.g005

probable model was Model 3 (Poisson) followed by Model 5 (negative binomial).

Parameter distributions

After having estimated the most probable model, \mathcal{M}^* , it is of interest to estimate the posterior joint probability of the parameters θ with respect to \mathcal{D} and \mathcal{M}^* : $p(\theta|\mathcal{D}, \mathcal{M}^*)$.

From Bayes’ theorem, we can write

$$p(\theta|\mathcal{D}, \mathcal{M}^*) = \frac{\mathcal{L}(\theta|\mathcal{M}^*)\pi(\theta)}{\int_{\theta} \mathcal{L}(\theta|\mathcal{M}^*)\pi(\theta)}, \tag{26}$$

and the denominator of Eqn (26) can be estimated by nested sampling:

$$\hat{p}(\theta|\mathcal{D}, \mathcal{M}^*) = \frac{\mathcal{L}(\theta|\mathcal{M}^*)\pi(\theta)}{\hat{\mathcal{Z}}}. \tag{27}$$

Parameter estimation via reject sampling

Distribution $p(\theta|\mathcal{D}, \mathcal{M})$ can be estimated using reject sampling with approximation (27). As part of this process, the maximum of $\hat{p}(\theta|\mathcal{D}, \mathcal{M})$ can be determined by performing Nelder-Mead simplex optimisation with respect to this distribution over parameter space.

The estimated parameter distributions obtained by reject sampling for Models 3, 5, 7 and 11, are shown in Figures 3, 4, 5, and 6, respectively. In each case, the sample size n was 10000. The samples obtained by reject sampling were also used to construct density scatter plots (Figures 7 and 8), which provide a visualisation of the correlations between the parameters.

Parameter estimation directly from nested sampling

The parameter sequence $\{\theta_{min,1}, \theta_{min,2}, \dots, \theta_{min,J}\}$ is produced during nested sampling. Can this set of parameters be regarded as a random sample from $p(\theta|\mathcal{D}, \mathcal{M})$? Sivia and Skilling [26] proposed using $\{\theta_{min,k}\}_{k=1}^J$ for this purpose so long as it is weighted by $w_k = \Delta \hat{\mathcal{Z}}_k / \hat{\mathcal{Z}}$, where $\Delta \hat{\mathcal{Z}}_k = \lambda_{min,k}(\zeta_{k-1} - \zeta_k)$, on the basis that $w_k \approx p(\theta_{min,k}|\mathcal{D}, \mathcal{M})$. A theoretical justification for this is given by Chopin and Robert [29].

The appropriateness of regarding $\{\theta_{min,k}\}_{k=1}^J$ as a random sample from $p(\theta|\mathcal{D}, \mathcal{M})$, was ascertained empirically using the Kolomogorov-Smirnov test, as follows.

The Kolmogorov-Smirnov statistic D_n is given by

$$D_n = \sup_{x \in \mathbb{R}} |(x) - F_0(x)|,$$

where $F_0(x)$ is the cdf of the null-hypothesis pdf, and $F(x)$ is the empirical cdf obtained from a sample $\{X_i\}$:

$$F(x) = \frac{1}{n} \sum_{i=1}^n 1\{X_i \leq x\}. \tag{28}$$

This definition can be generalized to a weighted Kolmogorov-Smirnov statistic by replacing (28) with a weighted cdf:

$$F(x) = \sum_{i=1}^n w_i 1\{X_i \leq x\}.$$

This allows us to take account of the weights $\{w_k\}_{k=1}^J$ on $\{\theta_{min,k}\}_{k=1}^J$.

Applying this method to the toy model \mathcal{M}_{toy} presented in Methods S1, a sample $\{\theta_{min,k}\}_{k=1}^J$, with $J = 10586$, was obtained by performing nested sampling for the evaluation of evidence

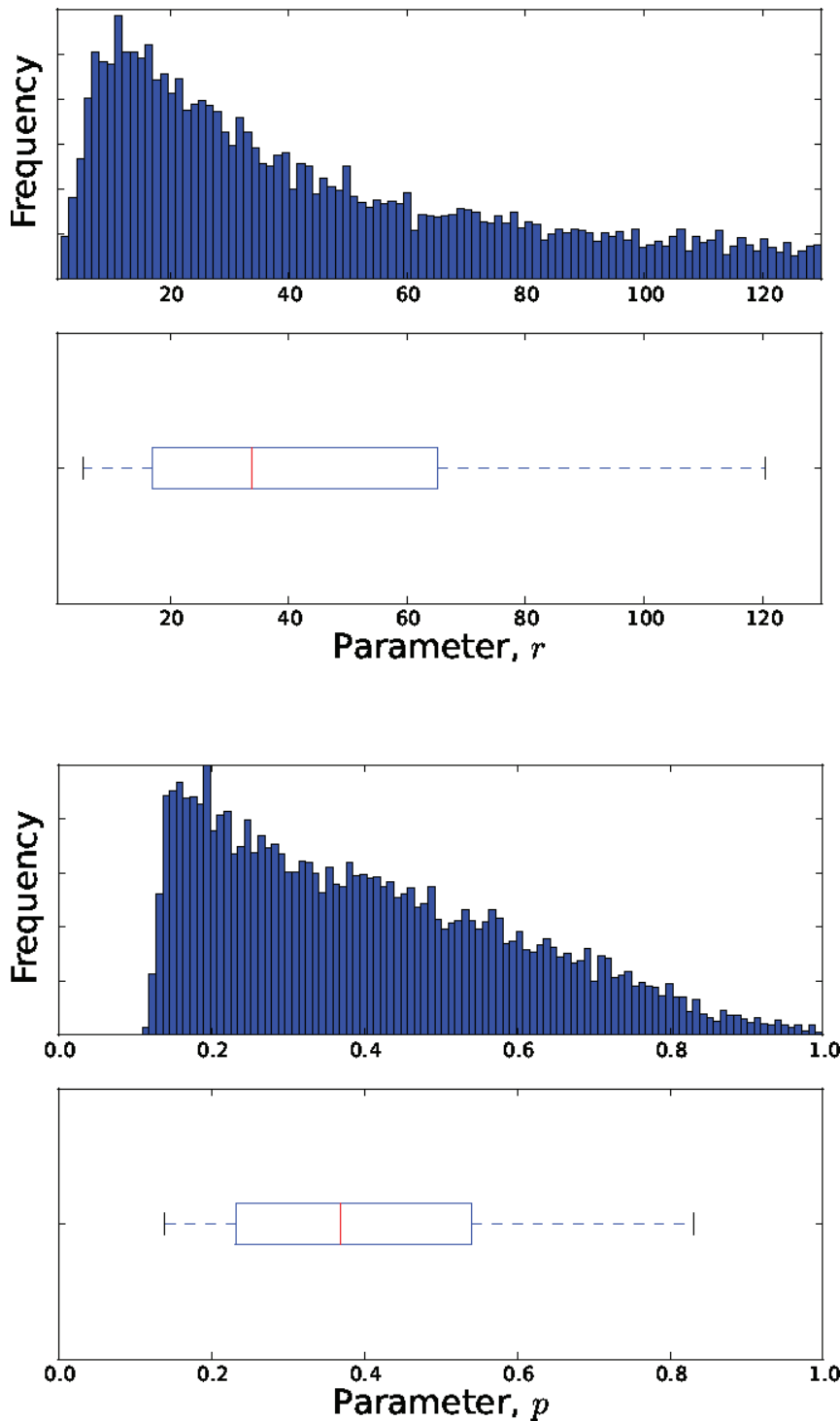


Figure 6. Estimates of the marginal probability distributions $p(r|\mathcal{D}, \text{Model5})$ and $p(p|\mathcal{D}, \text{Model5})$. \mathcal{D} is data from the virulent strain. doi:10.1371/journal.pone.0082317.g006

$p(\mathcal{D}|\mathcal{M}_{toy})$, where $\theta = [\alpha, \beta]T$. The corresponding sample $\{\alpha_{min,k}\}_{k=1}^J$ was compared with the marginal beta distribution,

$$p(\alpha|\cdot, \mathcal{M}_{toy}) = \alpha^{30}(1-\alpha)^{30},$$

using the weighted Kolmogorov-Smirnov statistic D_J . This statistic was equal to 0.01298. In order to obtain a frequentist p -value for

the statistic, an empirical probability distribution for D_J was obtained by randomly selecting a set $\{\tilde{\alpha}_k\}_{k=1}^J$ of α values from $p(\alpha|\cdot, \mathcal{M}_{toy})$ and determining D_J for the set, this being done 10000 times. On comparing 0.01298 with this empirical distribution, the p -value for $\{\alpha_{min,k}\}_{k=1}^J$ was found to be 0.0276. In contrast, when a sample of size J was obtained by reject sampling from

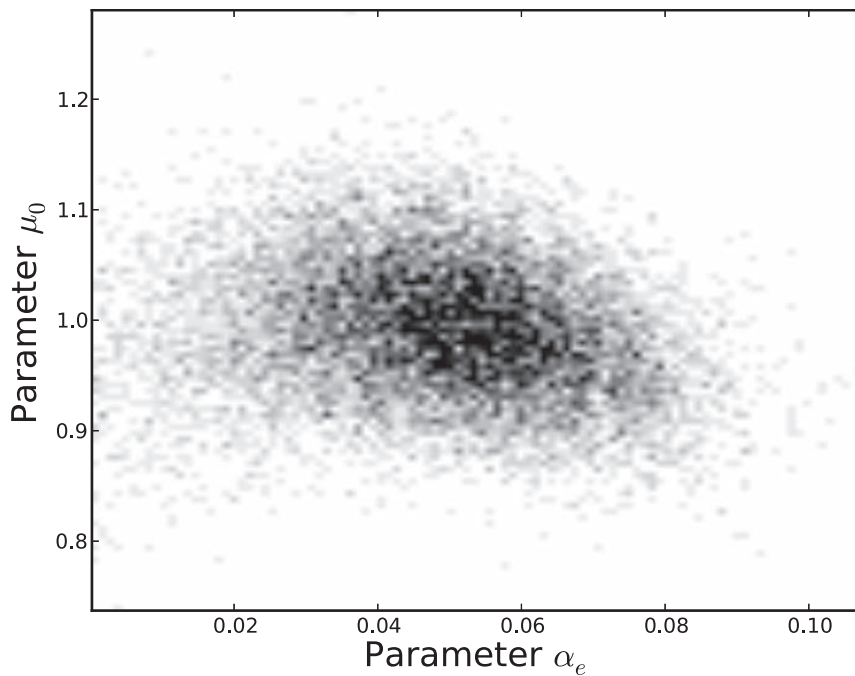


Figure 7. Density scatter plot of the estimated joint probability distribution $p(\mu_0, \alpha_e | \mathcal{D}, \text{Model11})$. \mathcal{D} is data from the attenuated strain. doi:10.1371/journal.pone.0082317.g007

$p(\alpha | \cdot, \mathcal{M}_{toy})$, the value of unweighted D_J was 0.00630, which has a p -value of 0.5772.

As a result of this experiment, it was decided not to use $\{\theta_{min,k}\}_{k=1}^J$ for estimating parameter distributions.

Model checking

It does not follow that the most probable model from a set of candidate models is necessarily an acceptable model: the most probable model may be the least worst of a set of poor models. What is required is an assessment of the fit of the most probable models to the observed data.

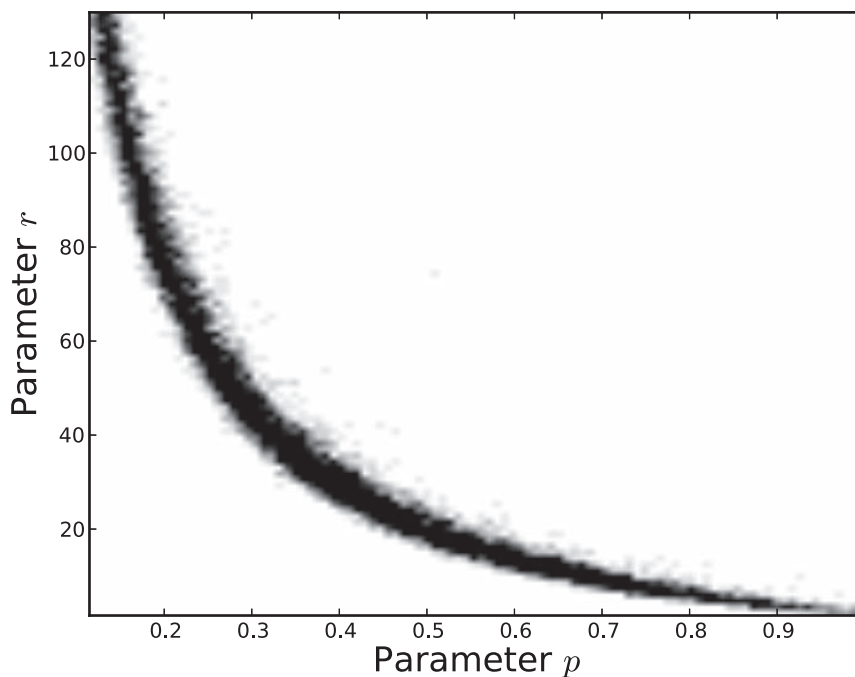


Figure 8. Density scatter plot of the estimated joint probability distribution $p(r, p | \mathcal{D}, \text{Model5})$. \mathcal{D} is data from the virulent strain. doi:10.1371/journal.pone.0082317.g008

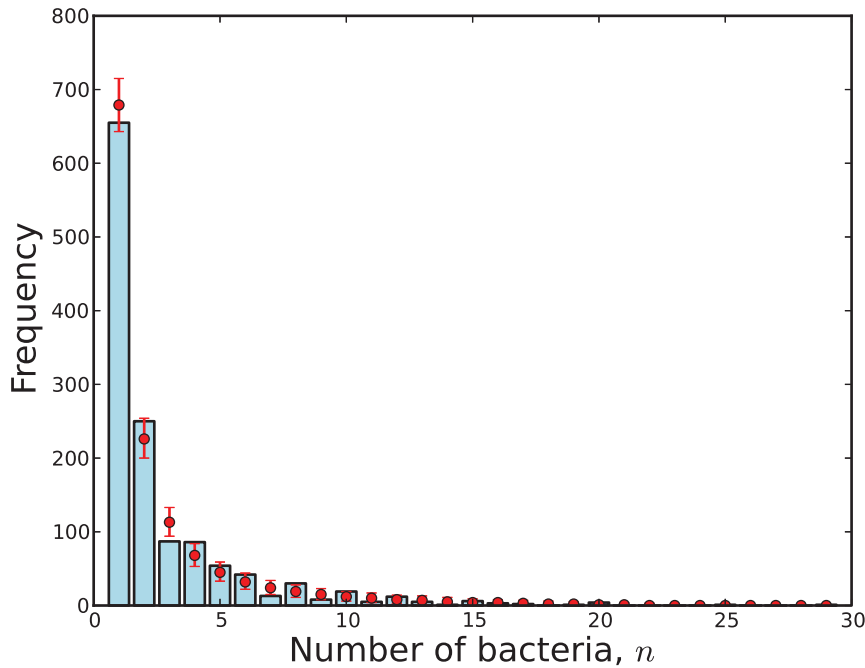


Figure 9. The observed number of cells with n bacteria (blue) compared with 95% credibility intervals (red) predicted by Model 3 with respect to the virulent strain.
doi:10.1371/journal.pone.0082317.g009

A common approach to assessing the fit of a model to data is to use a p -value with respect to some statistic $T(\mathbf{y})$, where \mathbf{y} is observed data. More formally, the classical p -value is given by

$$p(T(\mathbf{y}', \theta) \geq T(\mathbf{y}, \theta) | \theta, \mathcal{M}), \quad (29)$$

where \mathbf{y}' is a possible future value, and the probability is taken over the distribution of \mathbf{y}' given θ , a single parameter estimate.

A drawback of (29) is that it does not take account of the uncertainty in θ expressed by the posterior distribution $p(\theta | \mathbf{y}, \mathcal{M})$. In contrast, the Bayesian *posterior predictive p-value* [33,34]

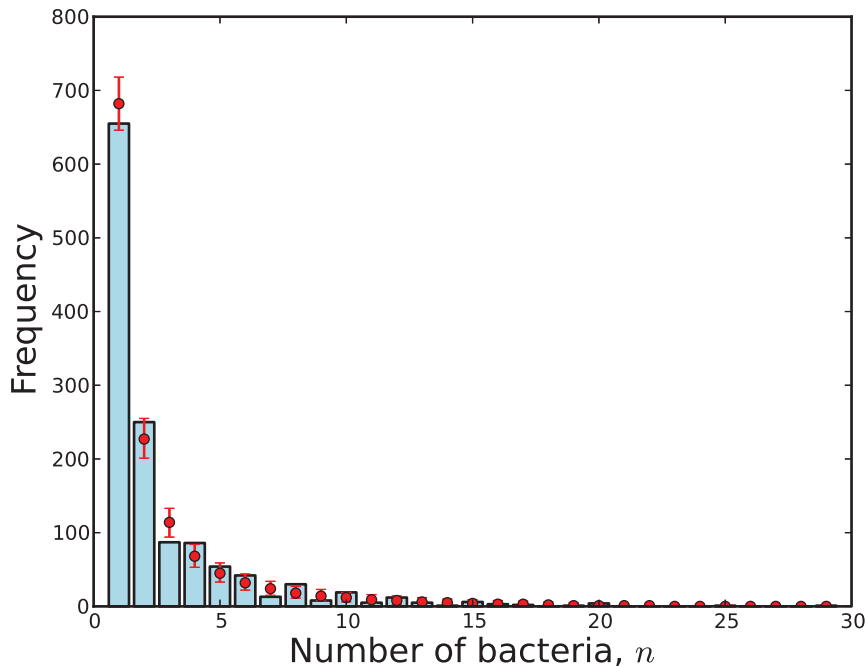


Figure 10. The observed number of cells with n bacteria (blue) compared with 95% credibility intervals (red) predicted by Model 5 with respect to the virulent strain.
doi:10.1371/journal.pone.0082317.g010

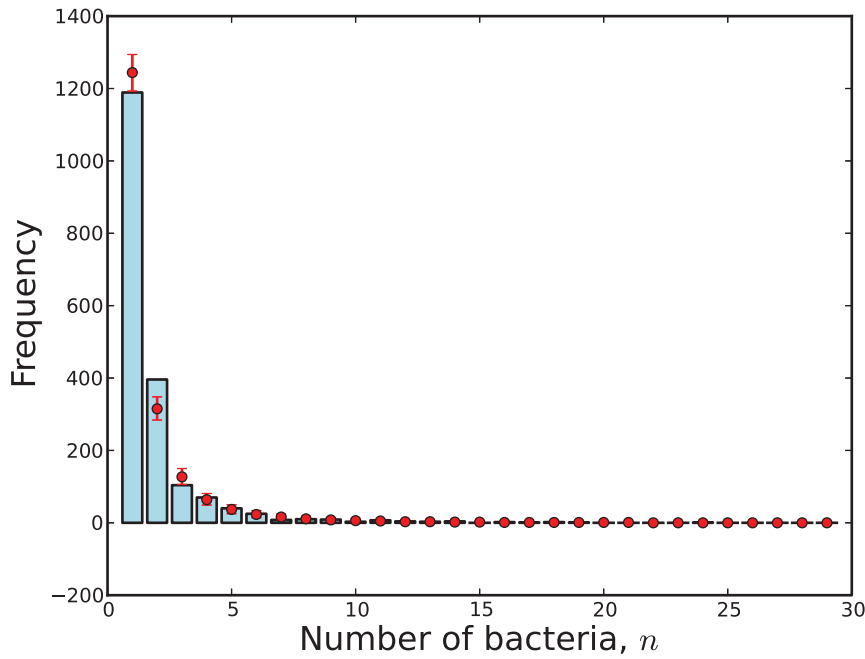


Figure 11. The observed number of cells with n bacteria (blue) compared with 95% credibility intervals (red) predicted by Model 7 with respect to the attenuated strain.
doi:10.1371/journal.pone.0082317.g011

$$p(T(\mathbf{y}') \geq T(\mathbf{y}) | \mathbf{y}, \mathcal{M}), \quad (30) \quad = \int_{\mathbf{q}} p(\mathbf{y}' | \theta, \mathcal{M}) p(\theta | \mathbf{y}, \mathcal{M}) d\theta.$$

overcomes the problem by using the *posterior predictive distribution*:

$$p(\mathbf{y}' | \mathbf{y}, \mathcal{M}) = \int_{\mathbf{q}} p(\mathbf{y}' | \theta | \mathbf{y}, \mathcal{M}) d\theta$$

The posterior distribution can be simulated by drawing m values $\tilde{\theta}$ from $p(\theta | \mathbf{y}, \mathcal{M})$, and then, for each $\tilde{\theta}$, sampling a \mathbf{y}' from

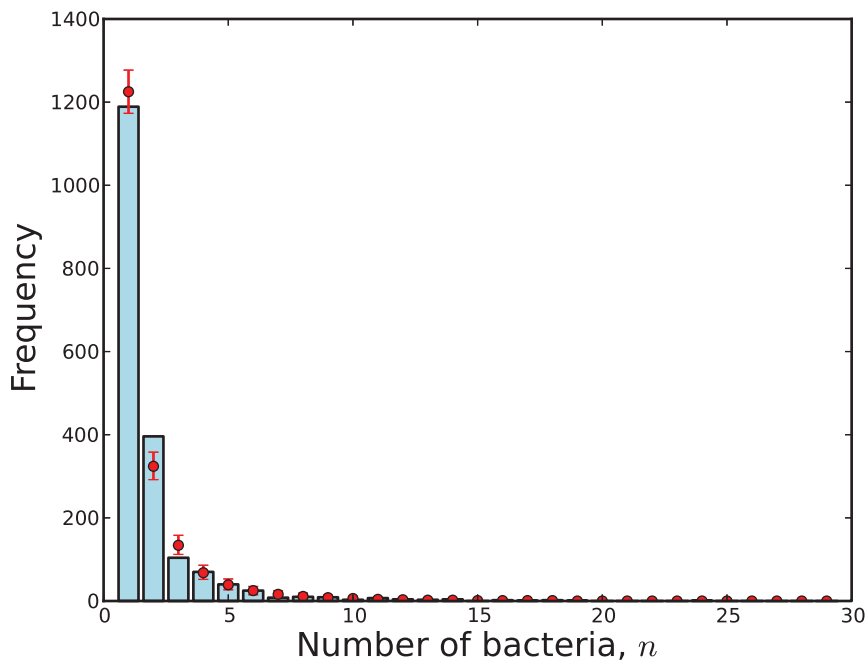


Figure 12. The observed number of cells with n bacteria (blue) compared with 95% credibility intervals (red) predicted by Model 11 with respect to the attenuated strain.
doi:10.1371/journal.pone.0082317.g012

$p(\mathbf{y}'|\tilde{\theta},\mathcal{M})$. The resulting m values of \mathbf{y}' represent draws from $p(\mathbf{y}'|\mathbf{y},\mathcal{M})$.

In the context of the *Salmonella* study, $\tilde{\theta}$ was provided by the parameter estimates obtained for $p(\theta|\mathcal{D},\mathcal{M})$, m was set to 10000, and $p(\mathbf{y}'|\tilde{\theta},\mathcal{M})$ was modelled as a multinomial distribution

$$p(\{y'_n\}_{n=1}^{29}|\{q(n|\tilde{\theta},\mathcal{M})\}_{n=1}^{29})=Y!\prod_{n=1}^{29}\frac{q(n|\tilde{\theta},\mathcal{M})^{y'_n}}{y'_n!}, \quad (31)$$

where Y is the total number of counts (cf. (22)).

In order to obtain m values of $T(\mathbf{y}')$ drawn from $p(\mathbf{y}'|\mathbf{y},\mathcal{M})$, each \mathbf{y}' drawn from $p(\mathbf{y}'|\tilde{\theta},\mathcal{M})$ is mapped to $T(\mathbf{y}',\tilde{\theta})$.

We used the G -statistic for the test statistic T [35]. The G -statistic is proportional to the Kullback-Leibler measure of distribution divergence, and is given by

$$G(\theta,\mathcal{M})=2\sum_{n=1}^{29}O_n\ln\left(\frac{O_n}{E_n(\theta,\mathcal{M})}\right), \quad (32)$$

where $O_n=C_n$, and $E_n(\theta,\mathcal{M})$ is the expected value for y'_n : $E_n(\theta,\mathcal{M})=Yq(n|\theta,\mathcal{M})$.

Applying the above approach for estimating the distribution of G under a given model \mathcal{M} , the posterior predictive p -values for $\{C_n\}_{n=1}^{29}$ were found to be 0.005 for Model 7 and 0.006 for Model 11 (with respect to the attenuated strain), $<10^{-4}$ for Model 3 and $<10^{-4}$ for Model 5 (with respect to the virulent strain). This suggests a poor fit of the models to the data.

A visual representation of the fit of data to a model \mathcal{M} can be provided by comparing the observed count C_n (the number of cells containing n bacteria) to the distribution of m possible count values y'_n obtained via (31). This visualisation is shown in Figures 9, 10, 11 and 12.

Discussion

The AIC is a common maximum-likelihood approach to model comparison, but nested sampling enables a Bayesian approximation of model evidence $p(\mathcal{D}|\mathcal{M})$ to be computed, along with the advantages of adopting the Bayesian approach. These include integration across parameters; estimation of the posterior parameter distributions (with visualisation of parameter correlations); and estimation of the posterior predictive distributions for goodness-of-fit assessments of the models.

Under the assumptions used, the most probable models with respect to the virulent and attenuated strains of *S. enterica* were burst-threshold Model 3 (Poisson) and burst-rate Model 7 (μ_0 only), respectively. The next two most probable models were

References

- Anderson D (2008) Model based inference in the life sciences: a primer on evidence. New York, NY: Springer Science+Business Media, LLC.
- Akaike H (1974) A new look at statistical model identification. IEEE Transactions on Automatic Control AU-19: 195–223.
- Bishop C (2006) Pattern Recognition and Machine Learning. New York: Springer.
- Schwarz G (1978) Estimating the dimension of a model. Annals of Statistics 6: 461–464.
- Geman S, Geman D (1984) Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence 6: 721–741.
- Skilling J (2004) Nested sampling. AIP Conference Proceedings 735: 395–405.
- Skilling J (2006) Nested sampling for general Bayesian computation. Bayesian Analysis 1(4): 833–859.

burst-threshold Model 5 (negative binomial) and burst-rate Model 11 (μ_0 plus a_c), respectively. However, the Bayesian posterior predictive p -values indicate that alternative models and/or a relaxation of the quasi-stationary assumption adopted by Brown et al. [24] should be considered. It may be the case that one of the candidate models is correct but the use of pooled data was detrimental.

Other assumptions of the underlying mechanistic model may also be wrong; in particular, the absence of bacterial death and the assumption that each released bacterium infects a new macrophage.

For both the attenuated and virulent strains, the data \mathcal{D} was recorded over a number of days following infection and then pooled, with $\mathcal{D}=\{C_n\}_{n=1}^{29}$. If time-dependent data is to be retained and nested sampling is to be applied then a method is required to estimate the likelihood function $p(\{\mathcal{D}^{[t]}\}|\theta,\mathcal{M})$, where $\mathcal{D}^{[t]}=\{C_n^{[t]}\}$ and $C_n^{[t]}$ is the number of cells containing n bacteria on the t -th day. Branching processes have been used to model a variety of biological systems [36], and we will investigate the potential of estimating $p(\{\mathcal{D}^{[t]}\}|\theta,\mathcal{M})$ through the use of Bellman-Harris processes to model within-host infection dynamics.

We have demonstrated that a visualisation of the marginal and joint posterior parameter distributions $p(\theta|\mathcal{D},\mathcal{M})$ is readily obtainable once model evidence \mathcal{Z} has been estimated by nested sampling. The estimated joint posterior distributions provided a visualisation of the correlations between the parameters. Through the use of a weighted Kolmogorov-Smirnov test, we also found that the parameter sequence $\{\theta_{min,k}\}_{k=1}^J$ resulting from nested sampling could not be regarded as a random sample from the posterior parameter distribution $p(\theta|\mathcal{D},\mathcal{M})$.

One drawback of Algorithm 2 is that the restricted priors will converge to a single mode when a likelihood is multi-modal, and this will cause the evidence \mathcal{Z} to be underestimated. This issue can be resolved by implementing a multi-modal version of nested sampling, such as that proposed by Feroz et al. [37] for comparing cosmological models.

Supporting Information

Methods S1 Toy example of nested sampling.
(PDF)

Acknowledgments

We wish to thank Dr Andrew Grant and Dr Chris Coward for their helpful contributions during discussions.

Author Contributions

Conceived and designed the experiments: RD TJM. Performed the experiments: RD. Analyzed the data: RD TJM OR. Wrote the paper: RD TJM PM OR. Statistical analysis: RD TJM.

- Mukherjee P, Parkinson D, Liddle A (2006) A nested sampling algorithm for cosmological model selection. Astrophysical Journal Letters 638: L51–L54.
- Murray I, Ghahramani Z, Mackay D, Skilling J (2006) Nested sampling for Potts models. In: Weiss Y, Scholkopf B, Platt J, editors. Advances in Neural Information Processing Systems (NIPS) 19. Cambridge, MA: MIT Press. pp. 947–954.
- Jasa T, Xiang N (2012) Nested sampling applied in Bayesian room-acoustics decay analysis. Journal of the Acoustical Society of America 132: 3251–3262.
- O'Neill P (2002) A tutorial introduction to Bayesian inference for stochastic epidemic models using Markov chain Monte Carlo methods. Mathematical Biosciences 180: 103–114.
- Charleston B, Bankowski B, Gubbins S, Chase-Topping M, Schley D, et al. (2011) Relationship between clinical signs and transmission of an infectious disease and the implications for control. Science 332: 726–729.

13. Miller M, Raberg L, Read A, Savill N (2010) Quantitative analysis of immune response and erythropoiesis during rodent malarial infection. *PLoS Computational Biology* 6: e1000946.
14. Crump J, Luby S, Mintz E (2004) The global burden of typhoid fever. *Bulletin of the World Health Organization* 82: 346–353.
15. Mulholland E, Adegbola R (2005) Bacterial infections - a major cause of death among children in Africa. *New England Journal of Medicine* 352: 75–77.
16. Crump J, Mintz E (2010) Global trends in typhoid and paratyphoid fever. *Clinical Infectious Diseases* 50: 241–246.
17. Mastroeni P, Grant A, Restif O, Maskell D (2009) A dynamic view of the spread and intracellular distribution of *Salmonella enterica*. *Nature Reviews Microbiology* 7: 73–80.
18. Mastroeni P, Grant A (2011) Spread of *Salmonella enterica* in the body during systemic infection: unravelling host and pathogen determinants. *Expert Reviews in Molecular Medicine* 13: e12.
19. Richter-Dahlfors A, Buchan A, Finlay B (1997) Murine salmonellosis studied by confocal microscopy: *Salmonella typhimurium* resides intracellularly inside macrophages and exerts a cytotoxic effect on phagocytes in vivo. *Journal of Experimental Medicine* 186: 569–580.
20. Grant A, Foster G, McKinley T, Brown S, Clare S, et al. (2009) Bacterial growth rate and host factors as determinants of intracellular bacterial distributions in systemic *Salmonella enterica* infections. *Infection and Immunity* 77: 5608–5611.
21. Grant A, Morgan F, McKinley T, Foster G, Maskell D, et al. (2012) Attenuated *Salmonella Typhimurium* lacking the pathogenicity island-2 type 3 secretion system grow to high bacterial numbers inside phagocytes in mice. *PLOS Pathogens* 8: e1003070.
22. Grant A, Sheppard M, Deardon R, Brown S, Foster G, et al. (2008) Caspase-3-dependent phagocyte death during systemic *Salmonella enterica* serovar Typhimurium infection of mice. *Immunology* 125: 28–37.
23. Sheppard M, Webb C, Heath F, Mallovs V, Emilianus R, et al. (2003) Dynamics of bacterial growth and distribution within the liver during *Salmonella* infection. *Cellular Microbiology* 5: 593–600.
24. Brown S, Cornell S, Sheppard M, Grant A, Maskell D, et al. (2006) Intracellular demography and the dynamics of *Salmonella enterica* infections. *PLoS Biology* 4: e349.
25. Dudewicz E, Mishra S (1988) *Modern Mathematical Statistics*. New York: John Wiley.
26. Sivia D, Skilling J (2006) *Data Analysis: A Bayesian Tutorial*, 2nd edition. Oxford: Oxford University Press.
27. Larson H (1982) *Introduction to Probability Theory and Statistical Inference*, 3rd edition. New York: John Wiley.
28. Feroz F, Hobson M (2008) Multimodal nested sampling: an efficient and robust alternative to MCMC methods for astronomical data analysis. *Monthly Notices of the Royal Astronomical Society* 2: 449–463.
29. Chopin N, Robert C (2010) Properties of nested sampling. *Biometrika* 97: 741–755.
30. Brown S, Cornell S, Sheppard M, Grant A, Maskell D, et al. (2006) Protocol S1: Details of model constructions and statistical analyses for “Intracellular demography and the dynamics of *Salmonella enterica* infections”. *PLoS Biology* 4: e349.
31. Garcia-Perez M (1999) MPROB: computation of multinomial probabilities. *Behaviour Research Methods, Instruments and Computers* 31: 701–705.
32. Ng KM (2002) *A Continuation Approach for Solving Nonlinear Optimization Problems with Discrete Variables*. Ph.D. thesis, Department of Management Science and Engineering, Stanford University, Stanford, CA.
33. Meng XL (1994) Posterior predictive p-values. *The Annals of Statistics* 22: 1142–1160.
34. Gelman A, Carlin J, Stern H, Rubin D (1995) *Bayesian Data Analysis*. London: Chapman & Hall.
35. Sokal R, Rohlf F (1995) *Biometry*, 3rd edition. New York: Freeman.
36. Kimmel M, Axelrod D (2002) *Branching Processes in Biology*. New York: Springer-Verlag.
37. Feroz F, Hobson M, Bridges M (2009) MultiNest: an efficient and robust Bayesian inference tool for cosmology and particle physics. *Monthly Notices of the Royal Astronomical Society* 398: 1601–1614.