Large-eruption rates of stratovolcanoes

Jonathan Rougier* Stephen R. Sparks†
Katharine V. Cashman†

Compiled March 31, 2015, from `missing1.tex`

Abstract

The large-eruption rates of stratovolcanoes are modelled exchangeably, in order to derive an informative prior distribution as an input into a subsequent volcano-by-volcano hazard assessment. Derivation and implementation of the likelihood function is not straightforward, necessitating a detailed code verification. Analysis of the LaMEVE database suggests that most of the world's stratovolcanoes are currently capable of producing large eruptions, but that the current large-eruption rate of active volcanoes with no recent record of eruptions is low, about 1.2 /kyr. The model is checked via a visual 'Turing test', and the paper concludes with some reflections on code verification and computation methods.

1 Introduction

Large explosive volcanic eruptions pose a major threat to human populations worldwide, as well as to economic performance; see, e.g., Sparks *et al.* (2013). For our purposes, we define 'large' to be a magnitude of at least 4, following the definition of Mason *et al.* (2004), which is

$$\text{magnitude} = \log_{10}(\text{mass in kg}) - 7.$$

*Department of Mathematics, University of Bristol. Corresponding author, email `j.c.rougier@bristol.ac.uk`
†Department of Earth Sciences, University of Bristol.

Magnitude 4 ('M4') corresponds to 100 million tonnes of ejected matter. See Simkin (1993) for background information about volcanoes, and Cashman and Sparks (2013) for a recent review of volcano physics. This paper focuses on stratovolcanoes, which are conical volcanoes capable of both explosive and effusive eruptions—Mt Fuji outside Tokyo is an iconic stratovolcano. There have been 56 large stratovolcano eruptions since 1900CE, globally.

We are currently engaged in a screening exercise to assess individual stratovolcanoes in terms of their return period curves for large eruptions, which is the first step in assessing the threat they pose. As part of this work we would like an informative prior distribution for the eruption rate of a specific volcano. We do not treat all volcanoes as having the same rate. Instead, we treat these rates as *a priori* exchangeable, modelled as an independent and identically distributed (IID) sample from some population distribution which is itself uncertain. Describing this population distribution is the objective of our analysis in this paper. More details are given in section 4.1.

The LaMEVE database (see section 4.3) contains 263 stratovolcanoes. Volcanologists (e.g. the co-authors of this paper) believe that many of these volcanoes are not currently capable of producing a large explosive eruption without a change of state, i.e. they are 'inactive'. Such a qualitative distinction between inactive and active volcanoes implies that the proportion of inactive volcanoes (i.e. volcanoes with rate zero) need not be almost the same as the proportion of active volcanoes with a very low rate such as 0.001 /yr. This is in contrast to rates further up the scale, where the proportion of volcanoes with a given rate is believed to be a smooth function of the rate. If λ_i is the current large-eruption rate for volcano i, then this discontinuity can be represented by an 'lump' of probability in the distribution for λ_i, at $\lambda_i = 0$; see (1). This lump introduces a complication to what might otherwise be a standard statistical problem, since it takes the model for eruptions outside the standard conjugate framework.

It is natural to model the number of large eruptions as a homogeneous Poisson process with rate λ_i, over a modest time-interval such as a few centuries (modest in geological time). As described above, the λ_i's are modelled as IID from some population, expressed as a family of distributions with a

finite number of uncertain parameters. Without the lump of probability at zero, the natural choice would be the Gamma distribution with shape parameter α and rate parameter β. Integrating out the λ_i's implies an IID Negative Binomial model for the number of large eruptions given (α, β). Then a posterior distribution for (α, β) can easily be derived as a function of a prior distribution and the observations. Possibly this posterior distribution could be effectively summarised in terms of the maximum likelihood estimate.

However, the lump of probability at zero is incompatible with a Gamma distribution: we need a more sophisticated distribution for the λ_i's. But this introduces the potential difficulty that the λ_i's cannot be integrated out in a closed-form expression. In this case they would be nuisance parameters: there are 263 of them, compared to just a handful of 'interesting' (to us) population parameters. Computationally, this would be very challenging.

There is a second complication too. The record of eruptions is incomplete, even going back for a modest time-interval such as 400 years (see, e.g., Simkin and Seibert, 1994; Brown *et al.*, 2014). The model must allow for under-recording in the database. Ideally the under-recording rate would vary by volcano; but, given the general scarcity of large eruptions, we must be satisfied with the approximation of a single value common to all volcanoes. Happily, there is a convenient conjugate form for the under-recording, if the eruptions themselves are a homogeneous Poisson process. If the recording process is IID Bernoulli with probability π, then the process for recorded eruptions is Poisson with rate $\pi\lambda_i$. As will be shown below, this model can easily be extended to different recording rates in different periods. So in fact under-recording is not a major complication, accepting this simple model of the same recording rate for all volcanoes.

Section 2 describes our model for the number of large eruptions of the world's stratovolcanoes, and shows that it has the very attractive feature of having a closed-form integrated likelihood for the population parameters of interest. As this derivation is moderately complicated, and the resulting function likewise, section 3 considers code verification: something that is not usually included in published papers and which, we suspect, is not done as often as it ought to be. Section 4 describes our analysis: section 4.1 clarifies

our research objective; section 4.2 describes our prior beliefs; section 4.3 presents the dataset and our posterior inference; section 4.4 gives model diagnostics; and section 4.5 describes the computation. Section 5 summarises our findings, and reflects on two unusual but complementary features of our analysis.

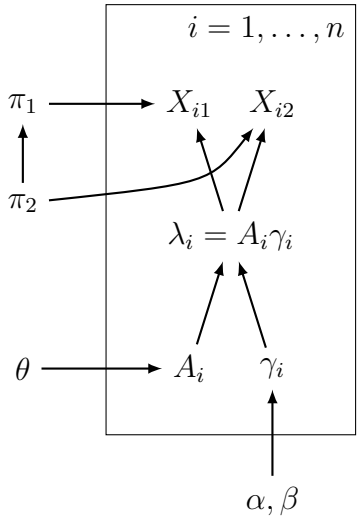
2 The model and the integrated likelihood

Consider a specified stratovolcano, volcano i, with large-eruption rate λ_i. We can model an lump of probability at zero in the prior distribution of λ_i by introducing a latent variable A_i, where $A_i = 0$ if the volcano is 'inactive', and $A_i = 1$ if the volcano is 'active'; an inactive volcano is currently incapable of producing an M4+ eruption without a change of state. We treat A_i as IID Bernoulli with probability θ. If $A_i = 1$, then the volcano has rate γ_i, where γ_i is IID Gamma with shape α and rate β. Then $\lambda_i = A_i\,\gamma_i$, and

$$\Pr(\lambda_i \leq v) = (1-\theta) + F(v;\alpha,\beta)\cdot\theta \tag{1}$$

for $v \geq 0$ and zero otherwise, where $F(\cdot\,;\alpha,\beta)$ is the distribution function of a Gamma distribution. This gives a three-parameter distribution for λ_i, with θ controlling the size of the lump of probability at zero.

For later reference here is the Directed Acyclic Graph (DAG) of the complete model. For simplicity of presentation (and without loss of generality—see below), assume that there are just two time-periods, 1600–1799CE with length $\Delta_1 = 200$, and 1800–2010CE with length $\Delta_2 = 211$. The recording probability is π_1 in the first period and π_2 in the second, with $\pi := (\pi_1, \pi_2)$. The data for each volcano comprise $\boldsymbol{X}_i = (X_{i1}, X_{i2})$, where X_{ij} is the number of large eruptions for volcano i in period j. The total number of volcanoes is $n = 263$.



(2)

This DAG shows π, θ, and (α, β) as mutually independent: we discuss the prior distribution of these four population parameters in section 4.2. This model is only tractable if our choices of marginal and conditional distributions permit the A_i's and the γ_i's to be integrated out in a closed-form expression. Interestingly, perhaps surprisingly, this can be done, even allowing for the recording rate to vary through time.

The key distinction is between volcanoes with no recorded eruptions, $\boldsymbol{x}_i = \mathbf{0}$, which may or may not be active, and volcanoes with at least one recorded eruption, which must be active in our model. Let there be n_z volcanoes with $\boldsymbol{x}_i = \mathbf{0}$, and assume for convenience that they are indexed from 1 to n_z. Let $\boldsymbol{\gamma} := (\gamma_1, \ldots, \gamma_n)$. Summing over the latent A_i's gives

$$\begin{aligned} L(\theta, \pi, \boldsymbol{\gamma}) &= \sum_{\boldsymbol{a} \in \{0,1\}^n} \prod_{i=1}^{n} \mathrm{p}(\boldsymbol{x}_i; a_i, \pi, \gamma_i)\, \mathrm{p}(a_i; \theta) \\ &= \sum_{\boldsymbol{a} \in \{0,1\}^{n_z}} \prod_{i=1}^{n_z} \mathrm{p}\left(\mathbf{0}; a_i, \pi, \gamma_i\right) \mathrm{p}(a_i; \theta) \prod_{i=n_z+1}^{n} \left\{ \mathrm{p}(\boldsymbol{x}_i; 1, \pi, \gamma_i)\, \mathrm{p}(1; \theta) \right\} \\ &= \theta^{n-n_z} \prod_{i=n_z+1}^{n} \mathrm{p}(\boldsymbol{x}_i; 1, \pi, \gamma_i) \sum_{\boldsymbol{a} \in \{0,1\}^{n_z}} \prod_{i=1}^{n_z} \mathrm{p}\left(\mathbf{0}; a_i, \pi, \gamma_i\right) \mathrm{p}(a_i; \theta). \end{aligned} \tag{3}$$

The second equality follows because all the terms in the sum with $a_i = 0$ for $i > n_z$ are zero.

Now integrate out γ_i from $\mathrm{p}(\boldsymbol{x}_i; a_i, \pi, \gamma_i)$ with respect to the Gamma prior distribution, for each i. To avoid clutter, drop the i subscript on $\boldsymbol{x}_i = (x_{i1}, x_{i2})$ and γ_i, and write $s := x_1 + x_2$. For the non-zero $\boldsymbol{x}_i$'s, the Poisson process model implies that

$$\begin{aligned}\mathrm{p}(\boldsymbol{x}; 1, \pi, \gamma) &= \mathrm{Poi}(x_1; \Delta_1\pi_1\gamma) \cdot \mathrm{Poi}(x_2; \Delta_2\pi_2\gamma) \\ &= e^{-(\Delta_1\pi_1 + \Delta_2\pi_2)\gamma} \frac{\gamma^s}{x_1!\, x_2!} (\Delta_1\pi_1)^{x_1} (\Delta_2\pi_2)^{x_2}, \end{aligned} \tag{4}$$

where 'Poi' denotes the Poisson Probability Mass Function (PMF). The Gamma prior distribution for γ_i implies that

$$\mathrm{p}(\boldsymbol{x}; 1, \pi, \gamma)\, \mathrm{p}(\gamma; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{(\Delta_1\pi_1)^{x_1} (\Delta_2\pi_2)^{x_2}}{x_1!\, x_2!} \gamma^{s+\alpha-1}\, e^{-(\Delta_1\pi_1 + \Delta_2\pi_2 + \beta)\gamma}. \tag{5}$$

Integrating out γ then gives

$$\mathrm{p}(\boldsymbol{x}; 1, \pi, \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{(\Delta_1\pi_1)^{x_1} (\Delta_2\pi_2)^{x_2}}{x_1!\, x_2!} \left(\frac{(\Delta_1\pi_1 + \Delta_2\pi_2 + \beta)^{s+\alpha}}{\Gamma(s+\alpha)} \right)^{-1}. \tag{6}$$

For the zero $\boldsymbol{x}_i$'s, take $s = x_1 = x_2 = 0$ and then, for $a = 1$,

$$\mathrm{p}\left(\mathbf{0}; 1, \pi, \alpha, \beta\right) = \left(\frac{\beta}{\Delta_1\pi_1 + \Delta_2\pi_2 + \beta} \right)^\alpha. \tag{7}$$

Obviously $\mathrm{p}(\mathbf{0}; 0, \dots) = 1$, which can also be verified from (7) using the equivalence $\{a = 0\} \sim \{a = 1, \Delta_1 = \Delta_2 = 0\}$.

Finally, integrate $\boldsymbol{\gamma}$ out of (3) using the above results to give

$$\begin{aligned} L(\theta, \pi, \alpha, \beta) = \theta^{n-n_z} &\prod_{i=n_z+1}^{n} \mathrm{p}(\boldsymbol{x}_i; 1, \pi, \alpha, \beta) \\ &\times \sum_{\boldsymbol{a} \in \{0,1\}^{n_z}} \prod_{i=1}^{n_z} \mathrm{p}(\mathbf{0}; a_i, \pi, \alpha, \beta)\, \mathrm{p}(a_i; \theta). \end{aligned} \tag{8}$$

The sum simplifies because the summand is invariant to permutations of $\boldsymbol{a}$, implying that there are only $n_z + 1$ distinct values. Letting n_a denote the number of 1's in $\boldsymbol{a}$, and using that $\mathrm{p}(\mathbf{0}; 0, \dots) = 1$,

$$\begin{aligned}
&\sum_{\boldsymbol{a}\in\{0,1\}^{n_z}} \prod_{i=1}^{n_z} \mathrm{p}(\mathbf{0}; a_i, \pi, \alpha, \beta)\, \mathrm{p}(a_i; \theta) \\
&\quad = \sum_{n_a=0}^{n_z} \binom{n_z}{n_a} (1-\theta)^{n_z - n_a} \left(\frac{\beta}{\Delta_1\pi_1 + \Delta_2\pi_2 + \beta} \right)^{\alpha n_a} \theta^{n_a} \\
&\quad = \sum_{n_a=0}^{n_z} \mathrm{Bin}(n_a; n_z, \theta) \left(\frac{\beta}{\Delta_1\pi_1 + \Delta_2\pi_2 + \beta} \right)^{\alpha n_a}, \qquad (9)
\end{aligned}$$

where 'Bin' denotes the Binomial PMF. Together, (8), (6), and (9) define a closed-form expression for the integrated likelihood of $(\theta, \pi, \alpha, \beta)$, which is nearly costless to evaluate.

The general case where the total interval Δ is divided up into k periods $\Delta_1, \dots, \Delta_k$, each with its own recording rate π_j, can be inferred directly.

In this paper, we just use two periods: 1600–1799CE and 1800–2010CE, with recording rates $\pi = (\pi_1, 0.9)$. Our beliefs about the recording rate over the last four centuries are fairly limited. We believe the recording rate in the C20th to have been nearly 1, and close to 1 in the C19th. We believe it to be substantially lower than this in the C17th and C18th. These beliefs are partly shaped by the frequency of recorded large eruptions in the database. One simple measure of under-recording is to compare the mid-point of the whole period (1805CE) with the date that divides set of of eruptions into two halves. These should be roughly the same under the hypothesis that eruption rates and recording rates are stable, but in fact the two-halves date is much later: 1856CE. We attribute this to a lower rate of recording in the first half of the period. In fact, a very crude calculation based on the above dates suggests that the recording rate for 1600-1799CE is about half of that for 1800-2010CE.

The results of Coles and Sparks (2006), Deligne *et al.* (2010), and Furlan (2010) are also relevant. All three of these papers impose parametric forms on the global recording rate as a function of time and magnitude. Furlan's

analysis allows for a step-change in the recording rate, consistent with our beliefs that it has been nearly one for at least the last century. But she estimates the timing of this step-change to be prior to 1600CE, which is at odds with the analysis above. Therefore we have settled on the simple model of having an uncertain recording rate $\pi_1 \leq 0.9$ for 1600–1799CE, and a recording rate of $\pi_2 = 0.9$ thereafter. We also tried two other plausible models for the recording rate through time, and the results were effectively unchanged.

3 Code verification

When implementing an inference based on the likelihood function given in section 2, three questions arise:

1. Have I done the maths correctly?
2. Have I coded the log-likelihood function correctly?
3. Are other approximations that I have made acceptable?

Together, these questions comprise 'code verification'. They are appropriate for any computation, but are particularly germane for modern statistical inferences with complex models and posterior approximations. As long as it is possible to sample from the model, there is a simple approach to code verification which ought to be applied routinely in statistical computation.

The approach is described in Cook *et al.* (2006), but we outline it here, for clarity. Let $f(\cdot\,;\theta)$ be a model for some observables Y, and suppose it is possible to simulate from this model for any $\theta \in \Omega$. Specify a prior distribution over Ω, say p_Θ. If the objective is a posterior distribution for Θ then this ought to be the actual prior; for verifying the likelihood function anything that is cheap to simulate from and evaluate is fine. Choose a score function S. Then follow this algorithm:

- For $i = 1, \ldots, m$:
    1. Sample $\theta \sim \mathrm{p}_\Theta(\cdot)$.

2. Sample $y \sim f(\cdot\,; \theta)$.
3. Evaluate $\mathrm{p}^*_\Theta := \mathrm{p}_{\Theta|Y}(\cdot \mid y)$.
4. Set $s_i \leftarrow S(\theta, \mathrm{p}^*_\Theta)$.

On completion, analyse the scores $s_1, \ldots, s_m$ according to their distribution under the null hypothesis that the code is correct.

Consider one iteration of the loop. The values (θ, y) are simulated from the joint distribution of (Θ, Y), for which

$$\mathrm{p}(\theta, y) = \mathrm{p}_\Theta(\theta)\, f(y; \theta) = \mathrm{p}(y)\, \mathrm{p}(\theta \mid y).$$

Since y is also a simulation from the marginal distribution $\mathrm{p}(y)$, the value θ ought to be compatible with the posterior distribution of Θ for this y. Simulating (θ, y) in steps 1 and 2 involves simulating from the model, while evaluating the posterior distribution in step 3 involves evaluating the likelihood function. So if the simulation and the likelihood function derive from the same model, then the simulated θ will be compatible with the evaluated posterior distribution. This compatibility is assessed by the score function, in step 4.

Cook *et al.* (2006) suggested using the relative rank of θ in a Monte Carlo simulation from the posterior distribution as the score function S, and then visualising the histogram of scores, which should be uniform under the null hypothesis, according to the Probability Integral Transform (Casella and Berger, 2002, sec. 2.2). (Where θ is a vector, do this for each margin.) Rather than simulate from the posterior distribution, we have represented it on a finite grid (see section 4.5), and hence we use the score function $S(\theta, \mathrm{p}^*_\Theta) = \Pr^*(\Theta \leq \theta)$ applied to each margin. Cook *et al.* suggest measuring uniformity of the scores using the p-value for the test statistic

$$\sum_{i=1}^{m} \left(\Phi^{-1}(s_i)\right)^2,$$

where Φ^{-1} is the standard Normal quantile function, which should be χ^2_m under the null hypothesis. This statistic is designed to emphasise departures

from uniformity at the two extremes, which Cook *et al.* claim is sometimes a hallmark of code errors. Below, we will simply use the p-value of the Kolmogorov-Smirnoff test, which we find to be a helpful statistic when inspecting a QQ plot. We do a separate assessment of power (see below and the Appendix).

As a multi-parameter assessment, we check the coverage of credible sets of different levels, using $S(\theta, \mathrm{p}^*_\Theta) = 1$ if θ is in our level-ν posterior credible set for p^*_Θ, and zero otherwise. We used 'snug' credible sets defined as level sets of the log-likelihood function. Under the null hypothesis, $s_1, \ldots, s_m$ should behave like a random sample from a Bernoulli distribution with parameter ν.

Other approximations, e.g. representing the posterior distribution on a finite grid, are also tested at the same time. The presence of approximations introduces the usual conundrum of p-values, which is that the p-value *ought* to be small, if the sample size m is large enough. An advantage of using the coverage of credible sets is that we can monitor the coverage for increasing sample size, assessed using, e.g., a 95% confidence interval, and satisfy ourselves that the coverage is about right up until the point where the confidence interval is acceptably narrow.

Figures 1 and 2 show these two code verification diagnostics, using the prior distribution described in section 4.2, and using $m = 200$ iterations. Initially we had $m = 100$ but we decided, on inspecting the results (notably the width of the 95% confidence intervals in Figure 2), that more power would be helpful. Both Figures appear to be satisfactory, and we are happy to report that our code verified first time. In order to assess the power, we then made an 'error' in the likelihood function, replacing every α with $\alpha \times 1.1$. This is a tiny error on the scale of the prior distribution for α, which ranges from 0.25 to 100 (see section 4.2). Even so, the verification diagnostics indicated a problem (shown in the Appendix).

We should finish this section on a note of caution. Satisfying the code verification test is a necessary but not sufficient condition for code validity, and the power calculation is suggestive but not conclusive. What we are claiming in this section is that if there is a lurking error in the code, then it is a small one which, we hope, will not substantially affect the results.

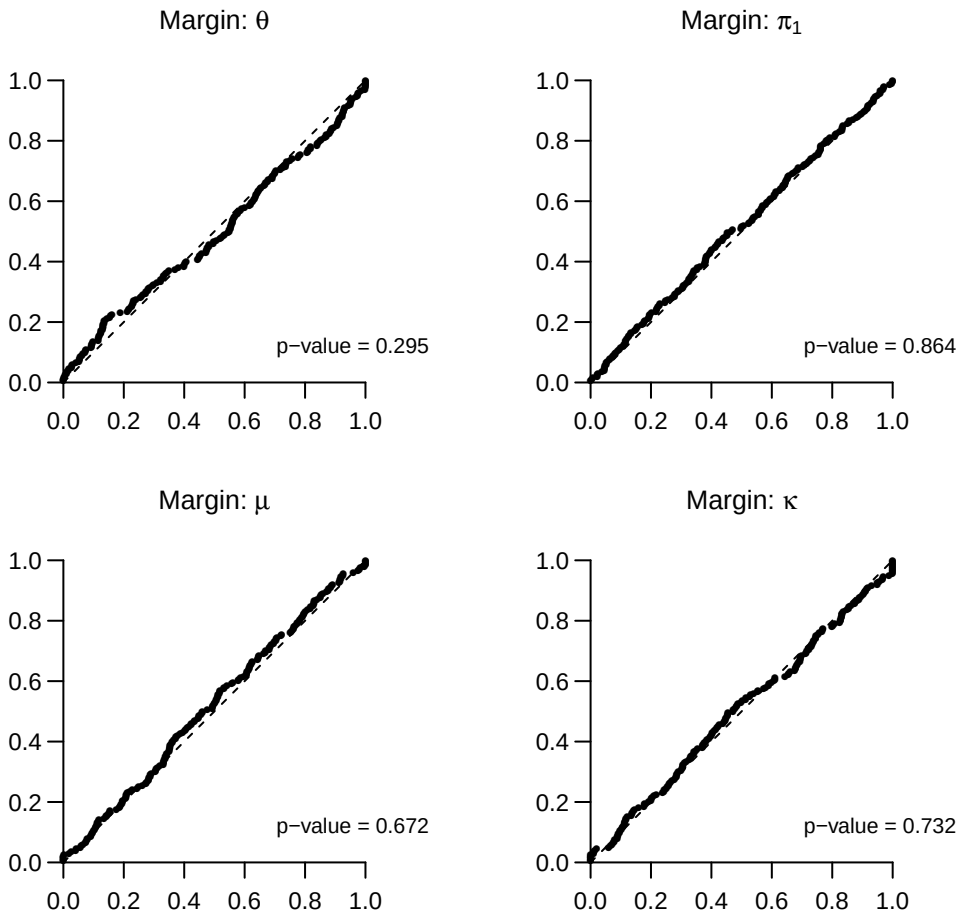


Figure 1: Code verification. Each panel shows the QQ plot for the scores of one of the four parameters. These should be approximately straight lines if there are no errors. The p-values are for the Kolmogorov-Smirnoff test.

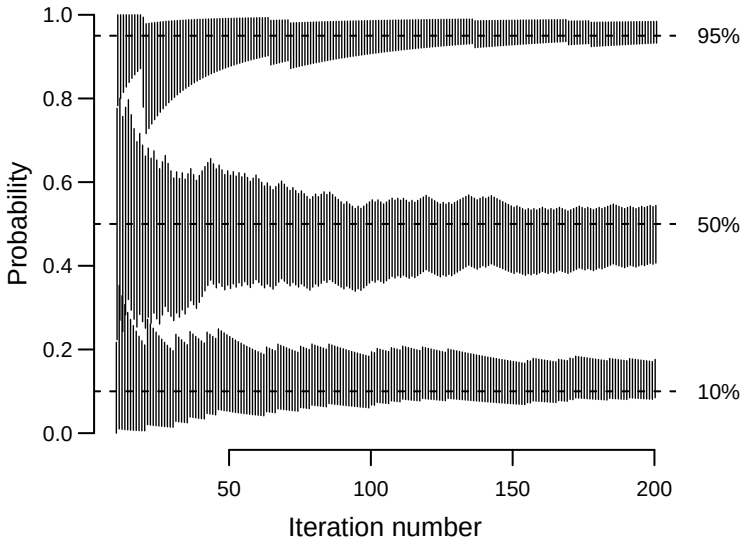


Figure 2: Code verification. 95% confidence intervals for the coverage probability of three different credible sets, with the target values shown as horizontal dashed lines. The vertical bars should converge on the target values if there are no errors. See Brown *et al.* (2001) for computing confidence intervals for Binomial probabilities.

4 Results

4.1 Our target inference

Ultimately, we want an informative prior distribution for the large-eruption rate of a specified active stratovolcano. Imagine an additional active volcano, not in the database, say volcano number $n + 1$. Write the dataset for the n volcanoes in the database as

$$\mathcal{D} := \left\{ \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \right\}.$$

The posterior distribution $\mathrm{p}(\gamma_{n+1} | \mathcal{D})$ provides the informative prior distribution we seek. Moreover, because there are so many volcanoes in the database, the effect of removing one active volcano from $\mathcal{D}$ will have little impact on the posterior distribution of γ_{n+1}. Therefore this posterior distribution can also be used as an informative prior for one of the active volcanoes in the database, or for a small fraction of them. Thus we wish to compute

$$\begin{aligned} F^*_\gamma(v) &:= \Pr(\gamma_{n+1} \leq v \mid \mathcal{D}) \\ &= \iint F(v; \alpha, \beta)\, \mathrm{p}^*(\alpha, \beta)\, \mathrm{d}\alpha\, \mathrm{d}\beta \end{aligned} \tag{10}$$

for values $v \geq 0$, where $F(\cdot\,; \alpha, \beta)$ is the Gamma distribution function, and an asterisk indicates conditioning on $\mathcal{D}$.

Ideally, the marginal posterior distribution $\mathrm{p}^*(\alpha, \beta)$ would be highly concentrated, in which case the approximation

$$F^*_\gamma(v) \approx F(v; \hat{\alpha}, \hat{\beta}) \tag{11}$$

would be extremely convenient, where the hats denote maximum likelihood estimates based on $\mathcal{D}$. But it is clear *a priori* that this will not be the case, because of a lack of identifiability in the likelihood function.

Set aside the issue of under-recording. There are at least two ways of getting lots of volcanoes with no large eruptions over the given period, as is the case in the database. First, most volcanoes might be active, but the rates

of the active volcanoes might be very low, so that the probability of an active volcano having no eruptions is high. This would correspond to a large θ and a small α/β (focusing on the expectation of γ_i, for simplicity). Second, most volcanoes might be inactive, but the rates of the active volcanoes might be much larger. This would correspond to a small θ and a much larger α/β. Of course there is also a spectrum of possibilities between these two extremes. The observations on their own cannot distinguish easily between points on the spectrum because even active volcanoes have only a handful of large eruptions in 400 years. Therefore there will be a ridge in the log-likelihood connecting the points in this spectrum, and the maximum likelihood estimate is unlikely to be a useful estimate, because it does not represent the centre of concentration.

Under-recording adds to the lack of identifiability, because a smaller π_1 can partially compensate for a larger α/β.

This identification problem does not pose any particular challenges for our inference, which is simply a posterior expectation. But it means that we cannot provide a simple closed-form approximation for the function F^*_γ, such as (11). We know, *a priori*, that our F^*_γ will be a non-degenerate mixture of Gamma distributions. So instead we must find a tractable approximation to F^*_γ, one that is, ideally, easy both to sample from and to evaluate. This is given below in Table 2.

4.2 Our choice of prior distribution

There are four population parameters: two probabilities π_1 (the recording rate for 1600–1799CE) and θ (the probability of being active), and the two parameters of the Gamma distribution for γ_i, the shape α and the rate β; see (2). In addition to the identification problem discussed in section 4.1, note that this inference is data-poor: the entire dataset is given below in Table 1. Undoubtedly we will need to provide an informative prior distribution, which represents, at least minimally, our beliefs about large eruptions of stratovolcanoes. This incisive comment from I.J. Good is apposite:

> When a statistician selects a [prior] distribution, he ought to make

> some effort to use the one that corresponds to his own judgement, whether the distribution is physical or intuitive. Since the selection is to some extent arbitrary, the statistician will have an opportunity to cheat. The more honest he tries to be, the more arbitrary and complicated his choice will look, and the more he will open the door to accusations of cheating. There is also the danger of unconscious cheating (wishful thinking). For this reason there is much to be said for minimizing arbitrariness, for compromising between philosophy and politics, between the ideal and the expedient. One method of making such a compromise is to restrict the class of [prior] distributions to a class with a small number of parameters. (Good, 1965, p. 16)

We will implement this suggestion by using only uniform distributions for the four population parameters, although, as discussed immediately below, the mutual independence which facilitates this choice requires a careful reparameterisation.

We will treat π_1, θ, and (α, β) as mutually independent in the prior distribution. We will take the prior distribution of π_1 to be $\mathrm{U}(0, 0.9)$, as we have very limited beliefs about the average recording rate in the period 1600–1799CE. Furthermore, such quantitative beliefs as we have are 'contaminated' by our exposure to previous empirical work using a similar dataset (see the end of section 2). We also take the prior distribution of θ to be $\mathrm{U}(0, 1)$, again reflecting very limited beliefs.

The prior for (α, β) is more complicated. We have beliefs about γ_i, concerning its expectation, μ say, and its coefficient of variation, κ say (defined as the standard deviation divided by the expectation). This parameterisation is chosen because our beliefs about μ and κ are largely independent. But this independence does not translate to independence for α and β. Therefore we transform to the new parameters (μ, κ) with

$$\alpha = \frac{1}{\kappa^2} \quad \text{and} \quad \beta = \frac{1}{\mu\kappa^2} = \frac{\alpha}{\mu} \tag{12}$$

according to the Gamma distribution properties

$$\mathrm{E}(\gamma_i; \alpha, \beta) = \frac{\alpha}{\beta} \quad \text{and} \quad \mathrm{Var}(\gamma_i; \alpha, \beta) = \frac{\alpha}{\beta^2}.$$

We treat μ and κ as independent in the prior distribution, and represent our beliefs as $\mu \sim \mathrm{U}(0.001, 0.01)$ /yr and $\kappa \sim \mathrm{U}(0.1, 2)$. The limits for μ represent our belief that the return period for large eruptions is likely to be between 100 yr and 1000 yr, which is broadly consistent with the assessment of long-term global magma production in White *et al.* (2006). The limits for κ represent that this is a weakly-held belief. A back-of-the-envelope calculation, taking the lower limits as approximately zero in both cases, gives

$$\begin{aligned} \mathrm{E}(\gamma_i) &\approx \frac{0.01}{2} = \frac{1}{2} \cdot 10^{-2} \\ \mathrm{Var}(\gamma_i) &\approx \frac{0.01^2}{3}\,\frac{2^2}{3} + \frac{0.01^2}{12} = \frac{19}{36} \cdot 10^{-4} \approx \frac{1}{2} \cdot 10^{-4}. \end{aligned}$$

Hence the coefficient of variation of γ_i is approximately $\sqrt{2}$, showing that this is a diffuse prior for γ_i; more diffuse, for example, than the Exponential distribution (which has coefficient of variation 1). The actual prior distribution for γ_i implied by these choices is shown as a dashed line in Figure 4.

4.3 Data and posterior distribution

The Large Magnitude Explosive Volcanic Eruptions (LaMEVE) database is an open-access database which represents the most complete record of M4+ eruptions; see Crosweller *et al.* (2012) and Brown *et al.* (2014). The download for this paper was made on 20 Oct 2014. We use all of the volcanoes in this database that are classified as stratovolcanoes, 263 in total. The complete dataset used in our analysis is shown in Table 1.

Our model was described in section 2 and our prior distribution in section 4.2. The code was verified in section 3, using our prior distribution. All calculations were performed in the statistical computing environment R (R Core Team, 2013), and the data and code are freely available at XXXX. Grid-based methods were used to assess the posterior distribution of the

Table 1: Observed frequency of M4+ eruptions for stratovolcanoes, extracted from the LaMEVE database. Value n_{ij} indicates the number of volcanoes with i recorded eruptions in 1600–1799CE and j recorded eruptions in 1800–2010CE.

		Values for j						
		0	1	2	3	4	5	Sum
i	0	182	30	10	2	0	0	224
	1	20	5	3	1	0	1	30
	2	2	1	2	2	1	0	8
	3	0	0	0	0	0	0	0
	4	0	0	1	0	0	0	1
	5	0	0	0	0	0	0	0
	Sum	204	36	16	5	1	1	263

population parameters, and functions of them: see section 4.5 for further details.

Figure 3 shows the posterior margins for the four population parameters. Clearly the data are highly informative with respect to our prior distribution. The recording rate for the period 1600–1799CE is less than 0.9; the posterior range is about $(0.3, 0.8)$ and the posterior expectation is $\mathrm{E}^*(\pi_1) = 0.53$. This result is consistent with the finding in Coles and Sparks (2006), and contradicts that of Furlan (2010), who finds that the recording rate is effectively 1 from about 1600CE. Interestingly, this result is also consistent with our very crude calculation that $\pi_1 \approx 0.9/2$ (see the end of section 2).

The posterior distribution for θ has been pulled towards 1; the probability that volcano $n+1$ is active is $\mathrm{E}^*(\theta) = 0.79$. This distribution function is approximately quadratic, and so its density function is approximately triangular on the interval $(0.4, 1.0)$. The posterior distribution for γ_{n+1}, F^*_γ from (10), is given in Figure 4. This has been pulled towards zero and concentrated, with an expectation of $\mathrm{E}^*(\gamma_{n+1}) = \mathrm{E}^*(\mu) = 0.0024$ /yr and an effective upper bound of about 0.02 /yr. Table 2 gives some posterior quantiles of γ_{n+1}, from

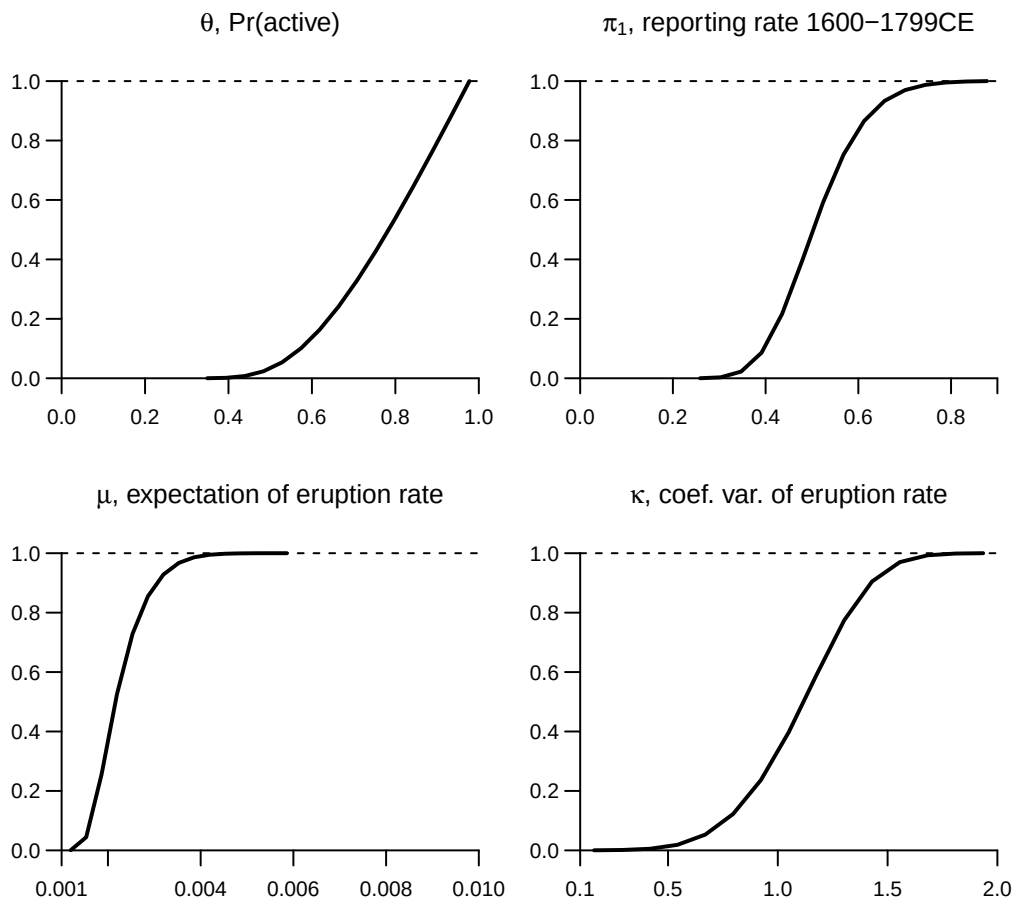


Figure 3: Distribution functions for the posterior margins of the four population parameters. The prior margins are uniform on the horizontal range (i.e. the prior distribution functions are diagonal straight lines). The restricted horizontal range of the curves shows the effect of range refinement (see section 4.5).

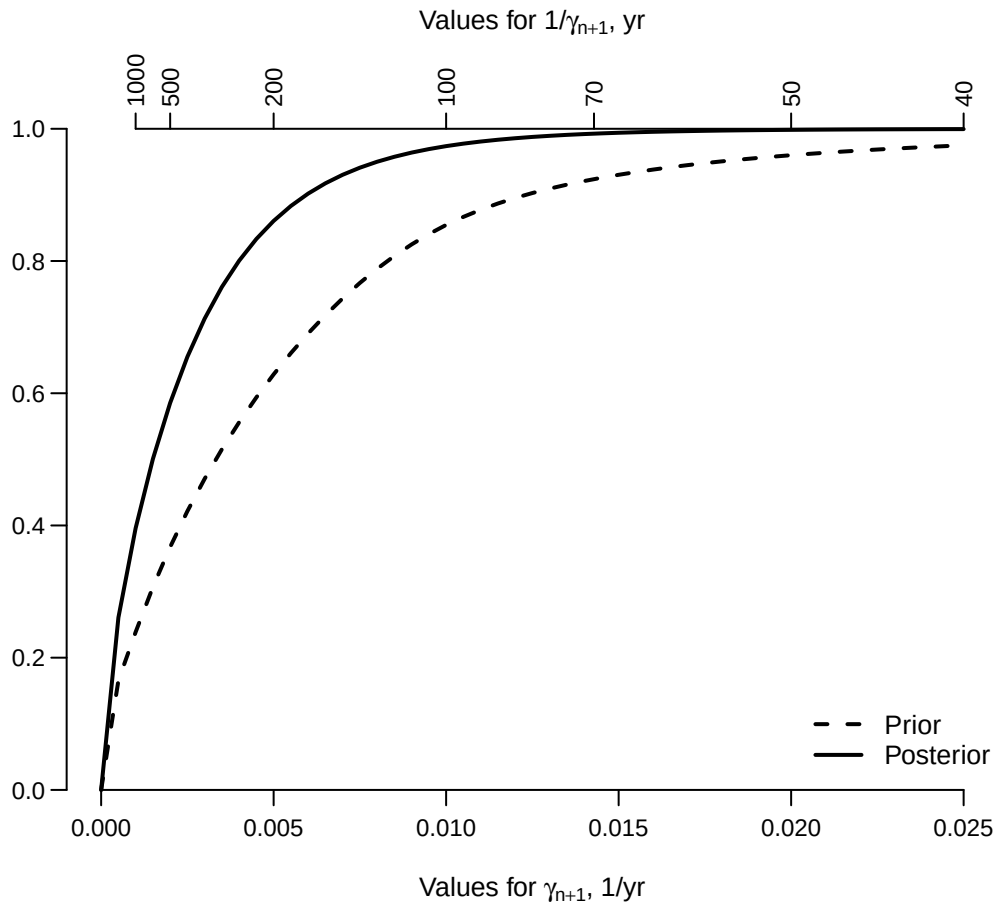


Figure 4: Prior and posterior distribution functions for γ_{n+1}, see (10). Quantiles are given in Table 2.

Table 2: Posterior quantiles of F^*_γ, see (10) and the solid line in Figure 4. The units of γ_{n+1} are /kyr and the units of $1/\gamma_{n+1}$ are kyr.

Prob.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	0.99	0.999
γ_{n+1}	0.00	0.19	0.38	0.65	1.02	1.50	2.10	2.89	3.99	5.95	7.99	13.22	21.91
$1/\gamma_{n+1}$	∞	5.21	2.61	1.55	0.98	0.67	0.48	0.35	0.25	0.17	0.13	0.08	0.05

which approximate density and quantile functions can easily be constructed, using linear or spline interpolation.

4.4 Model checking

Model checking is mandatory whenever statisticians make tractable parametric choices for marginal or conditional distributions. Volcanologists are certainly going to be appropriately sceptical about our choice of Uniform, Gamma, Bernoulli, and Poisson distributions, and the exchangeable framework which knits them into a joint distribution for eruptions over a large set of volcanoes; see (2).

For our model checking, we use a version of the posterior predictive approach, as proposed by Rubin (1984). Imagine a replicated version of the dataset, $\mathcal{D}^{\text{rep}}$, which represents an entirely new set of n volcanoes. The actual dataset $\mathcal{D}$ is visualised among the members of a random sample taken from $\mathcal{D}^{\text{rep}} \,|\, \mathcal{D}$. This is a statistical 'Turing test', in which the model competes with the actual dataset to 'fool' us, the experts, and in so doing establishes whether it is a reasonable representation of our beliefs. This Turing test interpetation of model checking was discussed by McWilliams (2007), in the context of climate modelling.

The algorithm for generating one replication is:

1. Sample $(\pi, \theta, \mu, \kappa)$ from the posterior distribution.
2. Compute α and β from (12).
3. For $i = 1, \ldots, n$:
   (a) Sample $A_i \sim \text{Ber}(\theta)$ and $\gamma_i \sim \text{Ga}(\alpha, \beta)$.
   (b) Set $\lambda_i = A_i \gamma_i$.
   (c) Sample $X_{i1}^{\text{rep}} \sim \text{Poi}(\Delta_1 \pi_1 \lambda_i)$ and $X_{i2}^{\text{rep}} \sim \text{Poi}(\Delta_2 \pi_2 \cdot 0.9 \cdot \lambda_i)$.
4. Set $\mathcal{D}^{\text{rep}} = \bigcup_{i=1}^{n} \{(X_{i1}^{\text{rep}}, X_{i2}^{\text{rep}})\}$.

Technically, the samples in step 3c are from Poisson processes, not from the Poisson distribution.

After k replications, the dataset is $\mathcal{D}, \mathcal{D}_1^{\text{rep}}, \ldots, \mathcal{D}_k^{\text{rep}}$. This combined dataset is too complicated to visualise in its entirety, and so we select a summary visualisation for each element. We prefer visual model checking, rather than a more numerical approach based on a test statistic. A visual approach is more accessible and more convincing, being similar to the model-development approach used widely in earth and environmental sciences. As our summary, we choose the recorded large-eruption sequence of the volcano with the largest number of recorded eruptions over the period 1600–2010CE (crudely, the 'most-active' volcano), which is easy to visualise on a line, allowing us to put a visual summary of all $k + 1$ datasets on the same page, for an easy comparison.

Figure 5 shows the result, using $k = 9$ replications. There is no apparent discrepancy between the eruption sequence of the actual most-active volcano (Kelut, in Indonesia) and the most active volcano from a replicated dataset. This outcome does not indicate that volcanoes follow homogeneous Poisson processes, as per our model, but, more prosaically, that we do not have sufficiently rich dataset for large eruptions to distinguish a homogeneous Poisson process from a more complicated process.

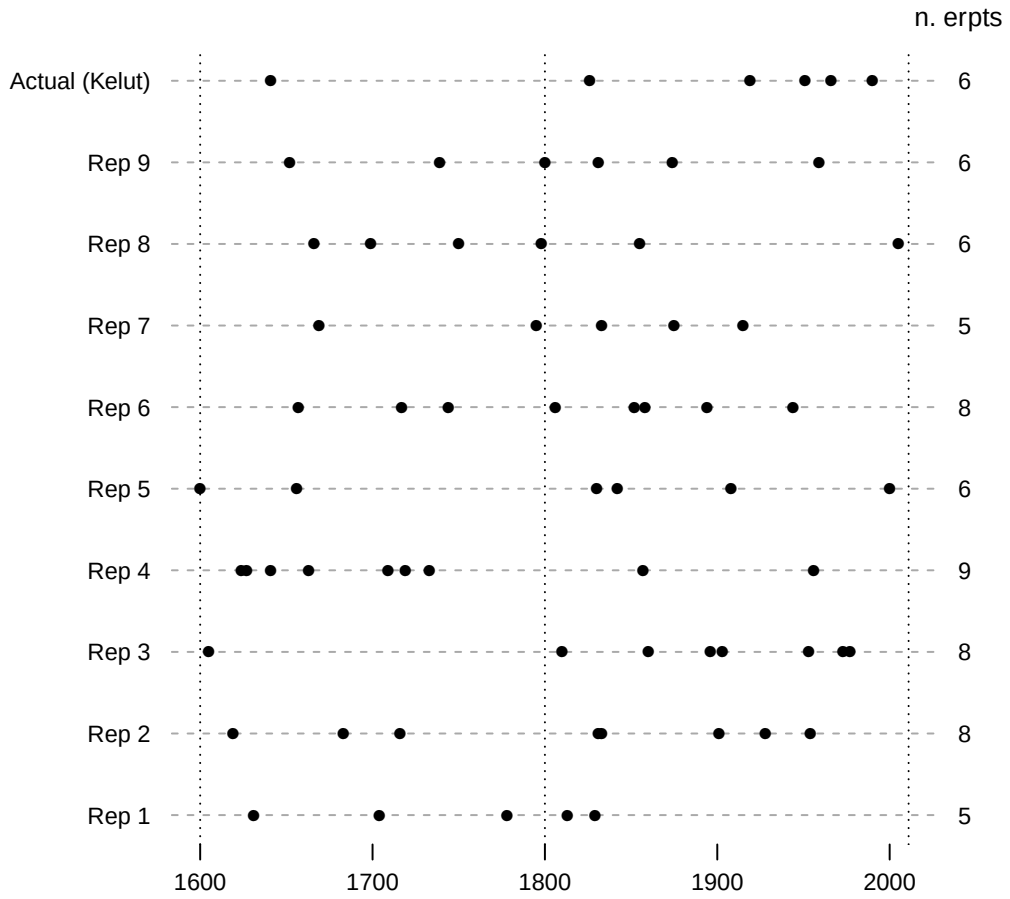


Figure 5: 'Turing test' for model checking. The recorded eruption sequence of the actual most-active volcano (Kelut), and the most-active volcano from a replicated dataset, for nine replications. Under the null hypothesis that the model adequately reflects our beliefs, there should be no apparent discrepancy between the first sequence and the other nine. The vertical dotted lines indicate the periods 1600–1799CE and 1800–2010CE, where the first period has a recording rate of π_1 and the second period has a recording rate of 0.9; from the dataset, $\mathrm{E}^*(\pi_1) = 0.53$.

Figure 5 demonstrates a feature of homogeneous Poisson processes which often catches people out: they can appear 'clustered', especially conditional on there being a large number of events.

4.5 Computation

It is easy to get carried away with Monte Carlo methods, but for problems with only a small number of population parameters, finite grids are more efficient (there is further discussion at the end of section 5). In the simplest case a regular finite grid represents the midpoint integration rule. There are better rules than this (see, e.g., Smith *et al.*, 1987, for a review), but they come with a higher price, in terms of complexity of implementation. Where the parameters are mutually independent in the prior distribution, a non-uniform marginal prior distribution can be incorporated as non-regular spacing, using a change of variables; but in this paper the parameters are mutually independent (after a suitable transformation) and marginally uniform. A generalisation can be used when the parameters are not mutually independent; this is sometimes termed Rosenblatt's transformation (Mardia *et al.*, 1979, sec.2.4).

The main downside with grids is that the likelihood can be highly concentrated with respect to the bounding box of the prior distribution, implying that most of the gridpoints contribute nothing to computing posterior expectations. This can be addressed by an initial grid refinement based on level sets of the log-likelihood function. For this purpose, quasi-random sequences are ideal. If $\{\boldsymbol{t}_1, \boldsymbol{t}_2, \ldots\}$ is a quasi-random sequence and $\hat{\ell}$ is the largest log-likelihood in the sequence, then all points in the sequence satisfying

$$\ell(\boldsymbol{t}_i) \geq \hat{\ell} - \chi_p^{-2}(0.999)/2$$

are kept, where χ_p^{-2} is the quantile function of the χ_p^2 distribution, and p is the number of parameters. The revised bounding box is the smallest box that contains all of the kept points. A similar approach is used in Rue *et al.* (2009) and Scheel *et al.* (2011).

In the application in this paper (four parameters), 10^4 evaluations were

used in the quasi-random sequence (a Sobol sequence), and these reduced the volume of the bounding box by between 50% and 95%, assessed over many simulations of the parameters and dataset. Fifteen points per parameter were used in the grid. As shown in section 3, code verification appeared to be satisfactory, suggesting that the grid refinement step and 15-point quadrature rule were not introducing an appreciable inaccuracy in the computation.

Computing on grids and space-filling designs has got much faster with modern multiple-core CPUs. Because every gridpoint can be evaluated in parallel, the `mclapply` and `mcmapply` functions in the R package `parallel` provide a painless and effective upgrade on a single computer, subject to some restrictions (notably on Windows machines). The package `parallel` is now in the R base. If `X` is a matrix of parameter values, e.g. the result of `as.matrix(expand.grid(...))`, then the log-likelihood function can be evaluated over `X` using

```
grid <- lapply(1L:nrow(X), function(i) X[i, ])
gridllik <- mclapply(X = grid, FUN = llik, ...)
gridllik <- unlist(gridllik)
```

where '`...`' indicates additional arguments to `llik`. The first line is slightly clunky but it preserves the column names of `X` in the names of each element of `grid`, which might be used in `llik(par, ...)`; there might be a small speed-up from using `mclapply` here too. Experienced R programmers in search of more speed are likely to be rather more creative, in an application-dependent way, but there is much to be said for the transparency of the above code.

5 Summary and discussion

These results have met our research objective, of deriving an informative prior distribution for the rate of M4+ eruptions of an active stratovolcano. The model is presented in section 2; in it, the current large-eruption rates of active stratovolcanoes are treated as exchangeable. The model allows for the recording rate in the database to vary with time. In our treatment we have adopted a simple implementation, with an uncertain recording rate over

the period 1600–1799CE, and a recording rate of 0.9 since then. Our results are presented in section 4.3, with the target distribution shown as the solid line in Figure 4 and summarised, for numerical purposes, in Table 2. Model checking is via a posterior predictive 'Turing test', described in section 4.4 and shown in Figure 5.

Some calculations can put our results in perspective. There are 263 stratovolcanoes in the LaMEVE database. Of these, 81 have a recorded large eruption since 1600CE, and are definitely active according to our criterion (see Table 1). We have found that $\mathrm{E}^*(\theta) = 0.79$. Crudely, then, about $0.79 \times 263 \approx 208$ volcanoes are active, and, subtracting the 81 which we know are active, this leaves about 127 which are active but have no recorded large eruptions since 1600CE.

Now consider one such stratovolcano. Direct calculation using the posterior probabilities shows that the number of large eruptions of such a volcano over one year is well approximated by a Poisson distribution with rate 0.00123 /yr. So, over the next decade, the probability that at least one large eruption will be to a stratovolcano with no recent history of large eruptions is approximately

$$1 - \exp(-127 \times 0.00123 \times 10) \approx 0.79,$$

i.e. odds of about four to one. This seems like a fair bet to us. Indeed, in the decade 2001–2010CE there were seven large eruptions, of which four were at volcanoes with no recorded large eruptions since 1600CE: Ruang (2002, mag. 4.0), Reventador (2002, 4.6), Kasatochi (2008, 4.0), and Eyjafjallajökull (2010, 4.0). Note that these are 'large' eruptions (M4+); Eyjafjallajökull, for example, also erupted in 1612 and 1821, but at what were judged to be less than M4.

This forecast is based entirely on the information in one database. Local knowledge could be used to refine this forecast, since an inspection of each volcano would reveal much more information about its current state. Sakurajima, for example, has had several large eruptions in the last one hundred years, and we treat it as active. But it recently changed state from

'closed-conduit' to 'open-conduit', making large eruptions much less likely. For our purposes, though, our crude characterisation of volcanoes suffices, because we will be incorporating local information at the next stage of our volcano-by-volcano risk assessment.

Finally, there are two complementary features of our analysis which are unusual for a typical modern Bayesian statistical application: our detailed code verification (section 3 and the Appendix) and our use of a grid-based method rather than Monte Carlo (MC) (section 4.5). We regard code verification as a crucial part of any meaningful statistical analysis, which should be insisted on by collaborators and stakeholders. The Cook *et al.* (2006) verification approach that we have adopted involves many iterations of the inferential calculation, and is not feasible if this calculation takes more than a few seconds. Grid-based methods can meet this requirement, taking full advantage of the features of modern computers, such as CPUs with multiple cores, and large memories.

For a finite budget of CPU cycles, both MC and grid-based methods are approximations. Whether an approximation is accurate can be settled through code verification. In this paper we are able to confirm that our combination of grid refinement and a 15-point quadrature rule is accurate. This involved 200 iterations of the inferential calculation, plus another 200 to assess the power of our test. In reality, of course, we have done twenty or thirty times as many iterations, while the code was being developed and refined, and the calculations replicated on different computers. This would have been impractical for an MC method taking, say, a few minutes per iteration.

Acknowledgements

We would like to thank Liz Swanson for her suggestions, which were used in section 5, and the Bayesian Cake Club reading group in the Maths Department, for discussions about code verification. This work was supported by the Natural Environment Research Council (NERC) funded Consortium on Risk in the Environment: Diagnostics, Integration, Benchmarking, Learn-

ing and Elicitation (CREDIBLE); grant number NE/J017450/1. Sparks acknowledges funding from the European Research Council in the VOLDIES project which led to the development of the LaMEVE database. Cashman acknowledges support from the AXA Research Fund.

References

L.D. Brown, T.T. Cai, and A. DasGupta, 2001. Interval estimation for a Binomial proportion. *Statistical Science*, **16**(2), 101–117. With discussion, pp 117–133.

S.K. Brown *et al.*, 2014. Characterisation of the Quaternary eruption record: analysis of the Large Magnitude Explosive Volcanic Eruptions (LaMEVE) database. *Journal of Applied Volcanology*, **3**(5). `http://www.appliedvolc.com/content/3/1/5`.

G. Casella and R.L. Berger, 2002. *Statistical Inference*. Pacific Grove, CA: Duxbury, 2nd edition.

C. Cashman and R.S.J. Sparks, 2013. How volcanoes work: A 25 year perspective. *Geological Society of America Bulletin*, **125**(5–6), 664–690.

S. Coles and R.S.J. Sparks, 2006. Extreme value methods for modelling historical series of large volcanic magnitudes. In H.M. Mader, S.G. Coles, C.B. Connor, and L.J. Connor, editors, *Statistics in Volcanology*, pages 47–56. Special Publication of IAVCEI, Geological Society of London.

S.R. Cook, A. Gelman, and D.B. Rubin, 2006. Validation of software for Bayesian models using posterior quantiles. *Journal of Computational and Graphical Statistics*, **15**(3), 675–692.

H.S. Crosweller *et al.*, 2012. Global database on large magnitude explosive volcanic eruptions (LaMEVE). *Journal of Applied Volcanology*, **1**(4). `http://www.appliedvolc.com/content/1/1/4`.

N.I. Deligne, S.G. Coles, and R.S.J. Sparks, 2010. Recurrence rates of large explosive volcanic eruptions. *Journal of Geophysical Research*, **115**, B06203.

C. Furlan, 2010. Extreme value methods for modelling historical series of large volcanic magnitudes. *Statistical Modelling*, **10**(2), 113–132.

I.J. Good, 1965. *The Estimation of Probabilities: An Essay on Modern Bayesian Methods*. The M.I.T. Press, Cambridge MA, USA. Paperback edition, 1968.

K.V. Mardia, J.T. Kent, and J.M. Bibby, 1979. *Multivariate Analysis*. Harcourt Brace & Co., London.

B.G. Mason, D.M. Pyle, and C. Oppenheimer, 2004. The size and frequency of the largest explosive eruptions on earth. *Bulletin of Volcanology*, **66**(8), 735–748.

J.C. McWilliams, 2007. Irreducible imprecision in atmospheric and oceanic simulations. *Proceedings of the National Academy of Sciences*, **104**(21), 8709–8713.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013.

J.C. Rougier, R.S.J. Sparks, and L.J. Hill, editors, 2013. *Risk and Uncertainty Assessment for Natural Hazards*. Cambridge University Press, Cambridge, UK.

D.B. Rubin, 1984. Bayesianly justifiable and relevant frequency calculations for the applied statistician. *The Annals of Statistics*, **12**(4), 1151–1172.

H. Rue, S. Martino, and N. Chopin, 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society, Series B*, **71**(2), 319–392. With discussion.

I. Scheel, P.J. Green, and J.C. Rougier, 2011. A graphical diagnostic for identifying influential model choices in Bayesian hierarchical models. *Scandinavian Journal of Statistics*, **38**(3), 529–550.

T. Simkin, 1993. Terrestrial volcanism in space and time. *Annual Review of Earth and Planetary Sciences*, **21**, 427–452.

T. Simkin and L. Seibert, 1994. *Volcanoes of the World*. Geoscience Press, Tucson AZ, USA.

A.F.M. Smith, A.M. Skene, J.E.H. Shaw, and J.C. Naylor, 1987. Progress with numerical and graphical methods for practical Bayesian statistics. *Journal of the Royal Statistical Society, Series D (The Statistician)*, **36**, 75–82.

R.S.J. Sparks, W.P. Aspinall, H.S. Crosweller, and T.K. Hincks, 2013. Risk and uncertainty assessment of volcanic hazards. In Rougier *et al.* (2013), chapter 11.

S.M. White, J.A. Crisp, and F.J. Spera, 2006. Long-term volumetric eruption rates and magma budgets. *Geochemistry, Geophysics, Geosystems*, **7**(3), Q03010. doi:10.1029/2005GC001002.

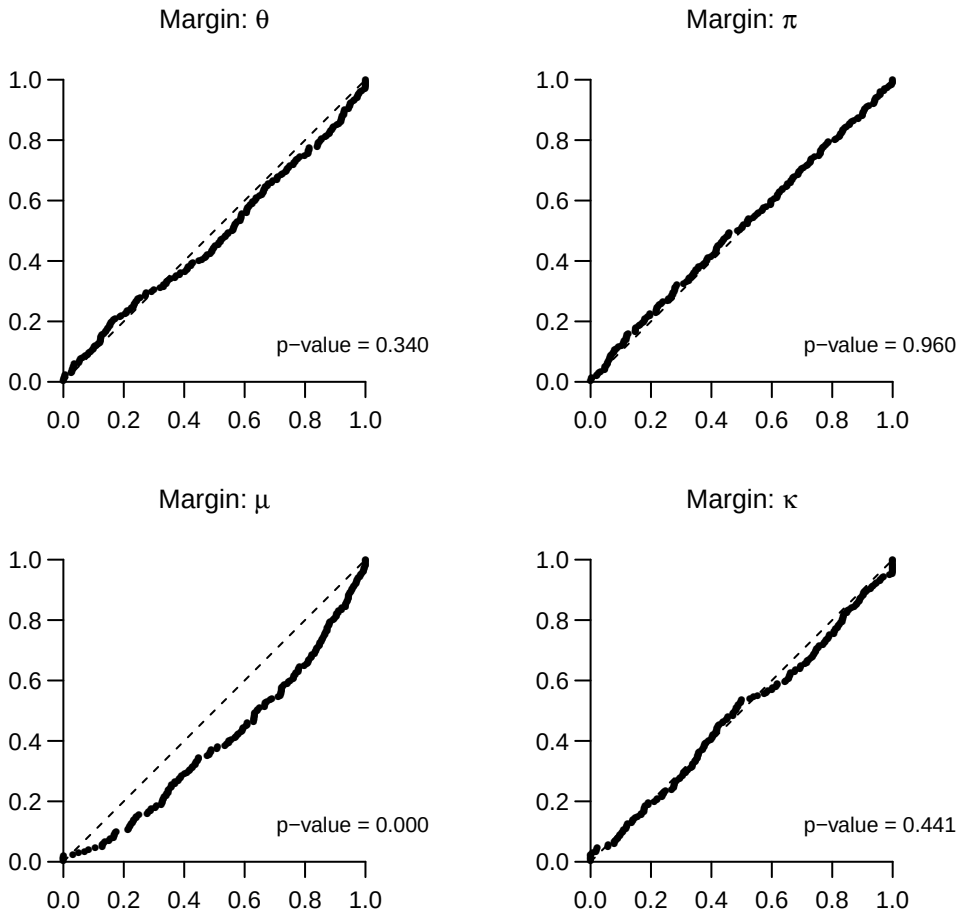


Figure 6: Same as Figure 1, except with a small code error.

Appendix

Here is a simple assessment of the statistical power of the code verification. A small error is introduced into the code, namely that the value of α in the likelihood function is 10% too large (see the end of section 3). The two Figures duplicate Figures 1 and 2, but with this error. Figure 6 is alarming, while Figure 7 is marginal. Clearly, this code verification test has useful power, against this type of error.

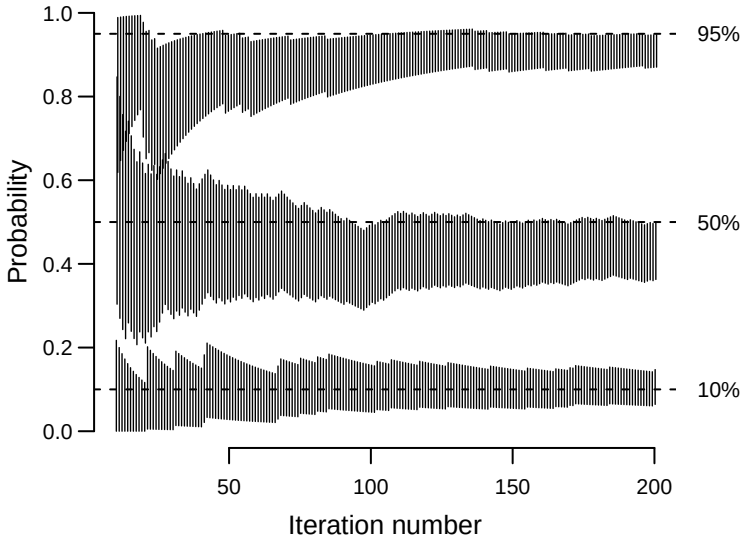


Figure 7: Same as Figure 2, except with a small code error.