

RESEARCH ARTICLE

Application of referenced thermodynamic integration to Bayesian model selection

Iwona Hawryluk¹ , Swapnil Mishra² , Seth Flaxman³, Samir Bhatt^{1,4*}, Thomas A. Mellan¹ 

1 MRC Centre for Global Infectious Disease Analysis, School of Public Health, Imperial College London, London, United Kingdom, **2** Saw Swee Hock School of Public Health and Institute of Data Science, National University of Singapore and National University Health System, Singapore, Singapore, **3** Department of Computer Science, University of Oxford, Oxford, United Kingdom, **4** Department of Public Health, Section of Epidemiology, University of Copenhagen, Copenhagen, Denmark

 These authors contributed equally to this work.

* samir.bhatt@sund.ku.dk


 OPEN ACCESS

Citation: Hawryluk I, Mishra S, Flaxman S, Bhatt S, Mellan TA (2023) Application of referenced thermodynamic integration to Bayesian model selection. PLoS ONE 18(8): e0289889. <https://doi.org/10.1371/journal.pone.0289889>

Editor: Mohammadreza Hadizadeh, Central State University, UNITED STATES

Received: June 6, 2023

Accepted: July 27, 2023

Published: August 14, 2023

Copyright: © 2023 Hawryluk et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All data generated and used in this project can be found in the Github repository (<https://github.com/mrc-ide/referenced-TI>).

Funding: SB and IH acknowledge support from the MRC Centre for Global Infectious Disease Analysis (MR/R015600/1), jointly funded by the UK Medical Research Council (MRC) and the UK Foreign, Commonwealth & Development Office (FCDO), under the MRC/FCDO Concordat agreement, and also part of the EDCTP2 programme supported by the European Union. SB is funded by the National

Abstract

Evaluating normalising constants is important across a range of topics in statistical learning, notably Bayesian model selection. However, in many realistic problems this involves the integration of analytically intractable, high-dimensional distributions, and therefore requires the use of stochastic methods such as thermodynamic integration (TI). In this paper we apply a simple but under-appreciated variation of the TI method, here referred to as *referenced TI*, which computes a single model's normalising constant in an efficient way by using a judiciously chosen reference density. The advantages of the approach and theoretical considerations are set out, along with pedagogical 1 and 2D examples. The approach is shown to be useful in practice when applied to a real problem—to perform model selection for a semi-mechanistic hierarchical Bayesian model of COVID-19 transmission in South Korea involving the integration of a 200D density.

Introduction

The marginalised likelihood, or normalising constant of a model, is a feature central to the principles and pathology of Bayesian statistics. For example—given two models representing two competing hypotheses, the ratio of the normalising constants (known as the Bayes factor), describes the relative probability of the data having been generated by one hypothesis compared to the other. Consequently, at a practical level the estimation of normalising constants is an important topic for model selection in the Bayesian setting [1].

In practice, estimating a normalising constant relies on “integrating out” or marginalising the parameters of the model to get the probability the associated hypothesis produced the data. But in general this is difficult, because we cannot easily integrate arbitrary high-dimensional distributions—certainly analytical or quadrature-based methods are of little help directly. As a result, practitioners turn to a range of approaches, typically based on statistical sampling. Specific examples include bridge sampling [2, 3], stochastic density of

Institute for Health Research (NIHR) Health Protection Research Unit in Modelling and Health Economics, a partnership between the UK Health Security Agency, Imperial College London and LSHTM (grant code NIHR200908). Disclaimer: "The views expressed are those of the author(s) and not necessarily those of the NIHR, UK Health Security Agency or the Department of Health and Social Care." S.B. acknowledges support from the Novo Nordisk Foundation via The Novo Nordisk Young Investigator Award (NNF200C0059309). SB acknowledges support from the Danish National Research Foundation via a chair grant. S.B. acknowledges support from The Eric and Wendy Schmidt Fund For Strategic Innovation via the Schmidt Polymath Award (G-22-63345). IH was funded by a MRC PhD studentship (MR/S502388/1). SM acknowledges support from National Research Foundation via The NRF Fellowship Class of 2023 Award (NRF-NRFF15-2023-0010). SF acknowledges support from the EPSRC (EP/V002910/2). 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.

states based methods [4, 5] and thermodynamic integration [6–9]. This work focuses on the latter —thermodynamic integration—in particular on the development of the practical details for efficient application for Bayesian model selection.

By way of introduction for researchers unfamiliar with thermodynamic integration (TI), it allows us to estimate the ratio of two normalising constants in a general and asymptotically exact way. Instead of marginalising the associated densities explicitly in terms of the high-dimensional integrals, using TI we only have to evaluate a 1-dimensional integral, where the integrand can easily be sampled with Markov Chain Monte Carlo (MCMC). To see how this works, consider two models labelled 1 and 2 with normalising constants z_1 and z_2 . Each z is given by

$$z_i = \int q_i(\boldsymbol{\theta})d\boldsymbol{\theta}, \quad i \in \{1, 2\}, \tag{1}$$

where q_i is a density for model M_i with parameters $\boldsymbol{\theta}$, that gives the model’s Bayesian posterior density as

$$p_i(\boldsymbol{\theta}) = \frac{q_i(\boldsymbol{\theta})}{z_i}, \quad i \in \{1, 2\}.$$

To apply thermodynamic integration we introduce the concept of a path between $q_1(\boldsymbol{\theta})$ and $q_2(\boldsymbol{\theta})$, linking the two densities via a series of intermediate ones. This family of densities, parameterised by the path coordinate λ , is denoted by $q(\lambda; \boldsymbol{\theta})$. An example path in λ is shown in Fig 1A.

The density $q(\lambda; \boldsymbol{\theta})$, linking q_1 to q_2 and defining the intermediate densities, can be chosen to have an optimal or in some way convenient path. A common choice based on convenience is the geometric one

$$q(\lambda; \boldsymbol{\theta}) = q_2^\lambda(\boldsymbol{\theta})q_1^{1-\lambda}(\boldsymbol{\theta}), \quad \lambda \in [0, 1].$$

The important point to note is that for $\lambda = 0$, $q(\lambda; \boldsymbol{\theta})$ returns the first density $q(0; \boldsymbol{\theta}) = q_1(\boldsymbol{\theta})$, for $\lambda = 1$ it gives $q(1; \boldsymbol{\theta}) = q_2(\boldsymbol{\theta})$, and for in-between λ values a log-linear mixture of the end-point densities. Just as we have defined a family of densities, there is an associated normalising constant for any point along the path, that for any value of λ is given by

$$z(\lambda) = \int_{\Omega(\lambda)} q(\lambda; \boldsymbol{\theta})d\boldsymbol{\theta}.$$

A further small but important point to avoid complications is to have densities with common support, for example $\Omega(\lambda = 1) = \Omega(\lambda = 0)$. Hereafter support is denoted by Ω .

Having set up the definitions of $q(\lambda; \boldsymbol{\theta})$ and $z(\lambda)$, the TI expression can be derived, to compute the log-ratio of $z_1 = z(\lambda = 0)$ and $z_2 = z(\lambda = 1)$, while avoiding explicit integrals over the

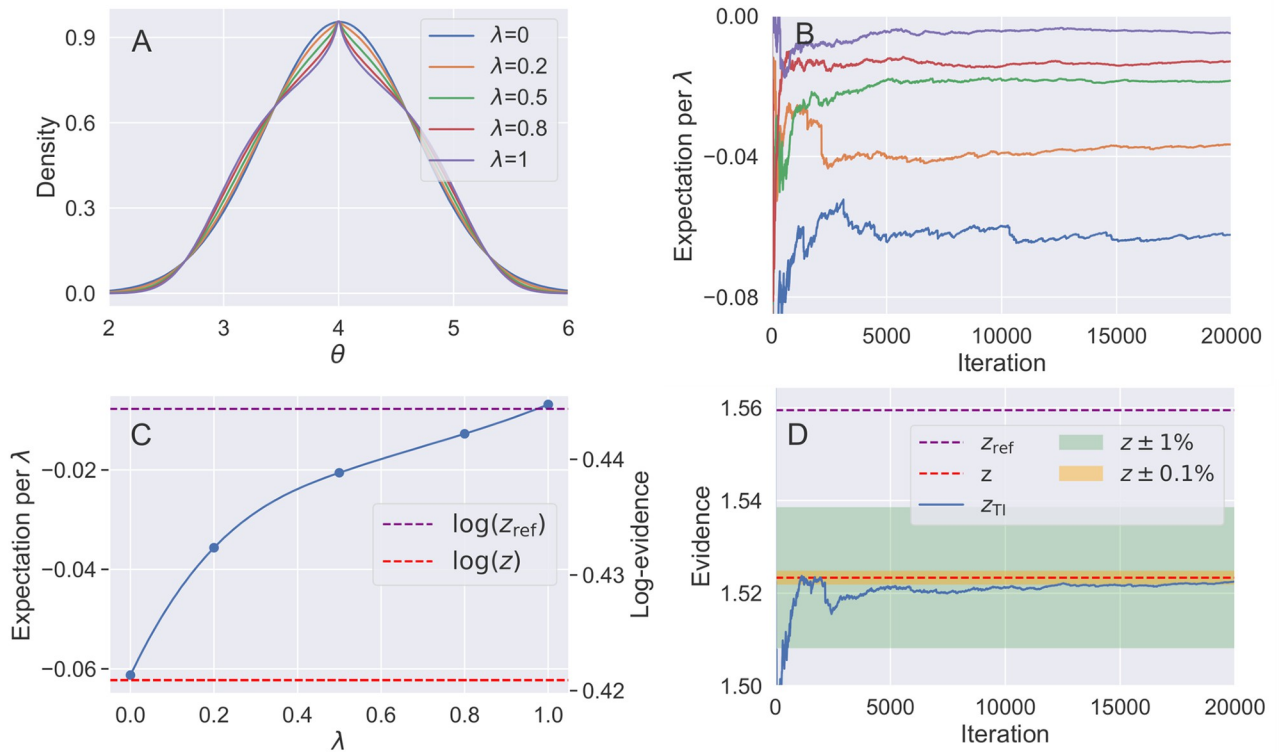


Fig 1. Illustration of steps from the 1D pedagogical example. A) $q^\lambda q_{\text{ref}}^{(1-\lambda)}$ for the 1d example density in parameter θ (Eq 9) at selected λ values along the path. B) Expectation $\mathbb{E}_{q(\lambda;\theta)} \left[\log \frac{q(\theta)}{q_{\text{ref}}(\theta)} \right]$ vs MCMC iteration, shown at each value of λ sampled. C) λ -dependence of $\mathbb{E}_{q(\lambda;\theta)} \left[\log \frac{q(\theta)}{q_{\text{ref}}(\theta)} \right]$, the TI contribution to the log-evidence. D) Convergence of the evidence z , with 1% convergence after 500 iterations and 0.1% after 17,000 iterations per λ .

<https://doi.org/10.1371/journal.pone.0289889.g001>

models' parameters θ . This is laid out as follows:

$$\begin{aligned}
 \log \frac{z_2}{z_1} &= \int_0^1 \partial_\lambda \log z(\lambda) d\lambda \\
 &= \int_0^1 \frac{1}{z(\lambda)} \partial_\lambda z(\lambda) d\lambda \\
 &= \int_0^1 \frac{1}{z(\lambda)} \partial_\lambda \int_\Omega q(\lambda; \theta) d\theta d\lambda \\
 &= \int_0^1 \frac{1}{z(\lambda)} \int_\Omega (\partial_\lambda \log q(\lambda; \theta)) q(\lambda; \theta) d\theta d\lambda \quad (2) \\
 &= \int_0^1 \mathbb{E}_{p(\lambda;\theta)} [\partial_\lambda \log q(\lambda; \theta)] d\lambda \\
 &= \int_0^1 \mathbb{E}_{p(\lambda;\theta)} \left[\log \frac{q_2(\theta)}{q_1(\theta)} \right] d\lambda \\
 &= \int_0^1 \mathbb{E}_{q(\lambda;\theta)} \left[\log \frac{q_2(\theta)}{q_1(\theta)} \right] d\lambda,
 \end{aligned}$$

Here we started with the fundamental theorem of calculus (first step), rules of differentiating logs (second step), definition of $z(\lambda)$ (third step), assumed exchangeability of ∂_λ and $\int d\theta$ and

log differentiation rules again (fourth step), identifying the expectation $\mathbb{E}_{q(\lambda;\theta)}$ from sampling distribution $q(\lambda;\theta)$ (fifth step), differentiation of the geometric path for $q(\lambda)$ (sixth step), and finally equivalence of sampling from q and p . The final line in the expression summarises the usefulness of TI. Instead of having to work with the complicated high-dimensional integrals of Eq 1 to find the log-Bayes factor $\log \frac{z}{z_1}$, which measures the relative probability of getting the data from one hypothesis compared to another, we only need to consider a 1-dimensional integral of an expectation, and that expectation can be readily produced by MCMC.

In our paper we examine the details of a *referenced TI* approach, which is a variation on the TI theme that we find useful to enable fast and accurate normalising constant calculations. Our main contributions are as follows:

- We show how to generate a reference normalising constant from an exactly-integratable reference density, through sampling or gradients, and with parameter constraints. And we present how to use this reference in the TI method to efficiently estimate a normalising constant of an arbitrary high-dimensional density.
- We discuss performance benchmarks for a well-known problem in the statistical literature [10], which shows the method performs favourably in terms of accuracy and the number of iterations to convergence.
- Finally the technique is applied to a hierarchical Bayesian time-series model describing the COVID-19 epidemic in South Korea.

In relation to other work, we recognise using a reference for thermodynamic integration is a topic that has been raised, especially in early theoretically-oriented literature [11–13]. Our additional contribution is to bridge the gap from theory and simple examples to application, which includes choosing the reference using MCMC samples or gradients, examination of reference support, comparisons of convergence, and illustration of the approach for a non-trivial real-world problem.

Methods

Referenced TI

Introducing a reference density and associated normalising constant as

$$\begin{aligned} z &= z_{\text{ref}} \frac{z}{z_{\text{ref}}} \\ &= z_{\text{ref}} \exp \int_0^1 \mathbb{E}_{q(\lambda;\theta)} \left[\log \frac{q(\theta)}{q_{\text{ref}}(\theta)} \right] d\lambda, \end{aligned} \quad (3)$$

yields an efficient approach to compute Bayes factors, or more generally to marginalise an arbitrary density for any application. To clarify notation, here z is the normalising constant of interest with density q , z_{ref} is a reference normalising constant with associated density q_{ref} . In the second line the ratio z/z_{ref} is straightforwardly given by the thermodynamic integral identity in Eq 2.

While the Eq 2 can be directly applied to conduct a pairwise model comparison between two hypotheses, by introducing a reference we can naturally marginalise the density of a single model [11, 12]. This is useful when comparing multiple models as $n > \binom{n}{2}$ for $n > 3$. Another motivation to reference the TI is the MCMC computational efficiency of converging the TI expectation. In Eq 3, with judicious choice of q_{ref} the reference normalising constant z_{ref} can be evaluated analytically and account for most of z . In this case $\log \frac{q(\theta)}{q_{\text{ref}}(\theta)}$ tends to have a small expectation and variance and converges quickly.

This idea of using an exactly solvable reference, to aid in the solution of an otherwise intractable problem, has been a recurrent theme in the computational and mathematical sciences in general [14–17], and variations on this approach have been used to compute normalising constants in various guises in the statistical literature [8, 9, 11, 18–24]. For example, in the generalised stepping stone method a reference is introduced to speed up convergence of the importance sampling at each temperature rung [23, 24]. In the work of [22] a theoretical discussion has been presented that shows the error budget of thermodynamic integration depends on the J-divergence of the densities being marginalised. Noting this, [19] provide an illustration for a 2-dimensional example in their work on recursive pathways to marginal likelihood estimation. And in the power posteriors method, a reference is used but the reference is a prior density and thus $z_{\text{ref}} = 1$ [20]. This approach is elegant as the reference need not be chosen—it is simply the prior—however the downside is that for poorly chosen or uninformative priors, the thermodynamic integral will be slow to converge and susceptible to instability. In particular for complex hierarchical models with weakly informative priors this is found to be an issue.

For referenced TI as presented here, the reference density in Eq 3 can be chosen at convenience, but the main desirable features are that it should be easily formed without special consideration or adjustments and that z_{ref} should be analytically integratable and account for as much of z as possible. Such a choice of z_{ref} ensures the part with expensive sampling is small and converges quickly. An obvious choice in this regard is the Laplace-type reference, where the log-density is approximated with a second-order one, for example a multivariate Gaussian. For densities with a single concentration, Laplace-type approximations are ubiquitous, and an excellent natural choice for many problems. In the following section we consider approaches that can be used to formulate a reference normalising constant z_{ref} from a second-order log-density (though more generally other tractable references are possible). In each referenced TI scenario, we note that even if the reference approximation is poor, the estimate of the normalising constant based on Eq 3 remains asymptotically exact—only the speed of convergence is affected (subject to the assumptions of matching support for end-point densities).

Taylor expansion at the mode Laplace reference

The most straightforward way to generate a reference density is to Taylor expand the log-density to second order about a mode. Noting no linear term is present, we see the reference density is

$$q_{\text{ref}}(\boldsymbol{\theta}) = \exp\left(\log q(\boldsymbol{\theta}_0) + \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T \mathbf{H}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\right), \quad (4)$$

where \mathbf{H} is the Hessian matrix and $\boldsymbol{\theta}_0$ is the vector of mode parameters. The associated normalising constant is

$$\begin{aligned} z_{\text{ref}} &= \int_{-\infty}^{\infty} q_{\text{ref}}(\boldsymbol{\theta}) d\boldsymbol{\theta} \\ &= \int_{-\infty}^{\infty} \exp\left(\log q(\boldsymbol{\theta}_0) + \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T \mathbf{H}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\right) d\boldsymbol{\theta} \\ &= q(\boldsymbol{\theta}_0) \int_{-\infty}^{\infty} \exp\left(\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T \mathbf{H}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\right) d\boldsymbol{\theta} \\ &= q(\boldsymbol{\theta}_0) \sqrt{\det(2\pi\mathbf{H}^{-1})}. \end{aligned} \quad (5)$$

This approach to yield a reference density, using either analytic or finite difference gradients at mode, tends to produce a density close to the true one in the neighbourhood of $\boldsymbol{\theta}_0$. But this is

far from guaranteed, particularly if the density is asymmetric, or has non-negligible high-order moments, or is discontinuous for example exhibiting cusps. In many instances a more reliable choice of reference can be found by using MCMC samples from the whole posterior density.

Sampled covariance Laplace reference

A second straightforward approach to form a reference density, that's often more robust, is by drawing samples from the true density $q(\theta)$ to estimate the mean parameters $\hat{\theta}$ and covariance matrix $\hat{\Sigma}$, such that

$$q_{\text{ref}}(\theta) = q(\hat{\theta}) \exp\left(-\frac{1}{2}(\theta - \hat{\theta})^T \hat{\Sigma}^{-1} (\theta - \hat{\theta})\right) \tag{6}$$

Then the reference normalising constant is

$$z_{\text{ref}} = q(\hat{\theta}) \sqrt{\det(2\pi\hat{\Sigma})}. \tag{7}$$

This method of generating a reference is simple and reliable. It requires sampling from the posterior $q(\theta)$ so is more expensive than gradient-based methods, but the cost associated with drawing enough samples to generate a sufficiently good reference tends to be quite low. In the primary application discussed later, regarding structured high-dimensional Bayesian hierarchical models, we use this approach to generate a reference density and normalising constant.

Though the sampled covariance reference is typically a good approach, it is not in general optimal within the Laplace-type family of approaches —typically another Gaussian reference exists with different parameters that can generate a normalising constant closer to the true one, thus potentially leading to overall faster convergence of the thermodynamic integral to the exact value. Such an optimal reference can be identified variationally, as we show in [S1 Appendix](#)

Reference support

If a model involves a bounded parameter space, for example $\theta_1 \in [0, \infty)$, $\theta_2 \in (-1, \infty)$ etc. as commonly arise in structured Bayesian models, in referenced TI the exact analytic integration for the reference density should be commensurately limited. This is necessary not only so the reference is closer to the true density to speed up convergence, but also so MCMC samples from both densities can be drawn on the same parameter space, as is required for the thermodynamic integrand in [Eq 3](#) to be well-defined. However, the calculation of arbitrary probability density function orthants (an orthant is a specific region in a multi-dimensional space or more precisely a bounded n-dimensional space), even for well-known analytic functions such as the multivariate Gaussian, is in general a difficult problem. High-dimensional orthant computations usually require advanced techniques, the use of approximations, or sampling methods [\[25–30\]](#). Fortunately, we can simplify our reference density to create a reference with tractable analytic integration for limits by using a diagonal approximation to the sampled covariance or Hessian matrix. For example the orthant of a diagonal multivariate Gaussian can be given in terms of the error function [\[31\]](#),

$$z_{\text{ref}} = q(\hat{\theta}) \sqrt{\det(2\pi\Sigma^{\text{diag}})} \prod_{i \in K} \left(1 + \text{erf} \left(\frac{\hat{\theta}_i - a_i}{\sqrt{2\Sigma_i^{\text{diag}}}} \right) \right), \tag{8}$$

where K denotes the set of indices of the parameters with lower limits a_i . Σ^{diag} is a diagonal

covariance matrix, that is one containing only the variance of each of the parameters, without the covariance terms and Σ_i^{diag} denotes the i^{th} element of the diagonal. Restricting our density to a diagonal one is a poorer approximation than using the full covariance matrix. In practice however this has not been a substantial drawback to the convergence of the thermodynamic integral—and again we state that the quality of the reference affects only convergence rather than eventual accuracy of the normalising constant. This behaviour is observed in the practical examples later considered, though we do recognise the distinction between accuracy and convergence and matters of asymptotic consistency using an MCMC estimator with finite iterations are not clear cut.

Technical implementation

Referenced TI was implemented in Python and Stan programming languages. Using Stan enables fast MCMC simulations with Hamiltonian Monte Carlo and No-U-Turn algorithm [32, 33], and portability between other statistical languages. We also provide an example of carrying out referenced-TI in NumPyro [34]. The code for all examples shown in this paper is available at <https://github.com/mrc-ide/referenced-TI>. In the examples shown in Section Applications, we used 4 chains with 20,000 iterations per chain for the pedagogical examples, and 4 chains with 2,000 iterations for the other applications. In all cases, half of the iterations were used for the burn-in (alternatively warm-up [33]). Mixing of the chains and the sampling convergence was checked in each case, by ensuring that the \hat{R} value was ≤ 1.05 and investigating the trace plots. The \hat{R} is a standard MCMC diagnostic that assesses the convergence of multiple parallel chains running in an MCMC simulation. It is computed by comparing the variance between chains to the variance within each chain. If chains have not fully converged, their variances will be larger compared to when they have reached convergence [35].

In all examples, the integral given in Eq 2 was discretised to allow computer simulations. Each expectation $\mathbb{E}_{q(\lambda, \theta)} \left[\log \frac{q_1(\theta)}{q_0(\theta)} \right]$ was evaluated at $\lambda = 0.0, 0.1, 0.2, \dots, 0.9, 1.0$, unless stated otherwise. To obtain the value of the integral in Eq 2, we interpolated a curve linking the expectations using a cubic spline, which was then integrated numerically. The pseudo-code of the algorithm with sampled covariance Laplace reference is shown in Algorithm 1.

Algorithm 1 Referenced thermodynamic integration algorithm

Input q —un-normalised density, q_{ref} —un-normalised reference density, Λ —set of coupling parameters λ , N —number of MCMC iterations

Output z —normalising constant of the density q

- 1: Define un-normalised density q and the reference density q_{ref}
- 2: Calculate z_{ref} analytically by using the determinant of the covariance matrix, as per Eqs 5 or 6 from the main text.
- 3: **for** $\lambda \in \Lambda$ **do**
- 4: Sample N values θ_n from $q^\lambda q_{\text{ref}}^{1-\lambda}$
- 5: **for** $n = 1, 2, \dots, N$ **do**
- 6: Calculate $\log \frac{q(\theta_n)}{q_{\text{ref}}(\theta_n)}$
- 7: **end for**
- 8: Compute the mean, $\mathbb{E}_\lambda = \frac{1}{N} \sum_{n=1}^N \log \frac{q(\theta_n)}{q_{\text{ref}}(\theta_n)}$
- 9: **end for**
- 10: Interpolate between the consecutive \mathbb{E}_λ values to obtain a curve $\partial_\lambda \log(z(\lambda))$
- 11: Integrate $\partial_\lambda \log(z(\lambda))$ over $\lambda \in [0, 1]$ to get $\log \frac{z}{z_{\text{ref}}}$
- 12: Calculate $z = z_{\text{ref}} \cdot \exp\left(\log \frac{z}{z_{\text{ref}}}\right)$

Applications

In this section we present applications of the referenced TI approach. In 1D Pedagogical Example and 2D Pedagogical Example with Constrained Parameters we give 1- and 2-dimensional pedagogical introductions to the approach. In Benchmarks—*Radiata Pine* we select a linear regression model for a well-known problem in the statistical literature, and finally in Model Selection for the COVID-19 Epidemic in South Korea we consider a challenging model selection task for a structured Bayesian model of the COVID-19 epidemic in South Korea.

1D pedagogical example

To illustrate the technique we consider the 1-dimensional density

$$q(\theta) = \exp\left(-\frac{1}{2}\sqrt{|\theta - 4|} - \frac{1}{2}(\theta - 4)^4\right), \theta \in \mathbb{R}, \tag{9}$$

with normalising constant $z = \int_{-\infty}^{\infty} q(\theta)d\theta$. This density has a cusp and it does not have an analytical integral that easily generalises to multiple dimensions.

In this instance the Laplace approximation based on the second-order Taylor expansion at the mode will fail due to the cusp, so we use the more robust covariance sampling method. Sampling from the 1D density $q(\theta)$ we find a variance of $\hat{\sigma}^2 = 0.424$, giving a Gaussian reference density $q_{\text{ref}}(\theta)$ with normalising constant of $z_{\text{ref}} = 1.559$. The full normalising constant, $z = z_{\text{ref}} \frac{z}{z_{\text{ref}}}$, is evaluated by Eq 3, by setting up a thermodynamic integration along the sampling path $q^\lambda q_{\text{ref}}^{(1-\lambda)}$. The expectation, $\mathbb{E}_{q(\lambda,\theta)} \left[\log \frac{q(\theta)}{q_{\text{ref}}(\theta)} \right]$, is evaluated at 5 points along the coupling parameter path $\lambda = 0.0, 0.2, 0.5, 0.8, 1.0$, shown in Fig 1. In this simple example, the integral can be easily evaluated to high accuracy using quadrature [36, 37], giving a value of 1.523. Referenced TI reproduces this value, with convergence of z shown in Fig 1, converging to 1% of z with 500 iterations and 0.1% within 17, 000 iterations.

This example illustrates notable characteristic features of referenced TI. Here the reference $q_{\text{ref}}(\theta)$ is a good approximation to $q(\theta)$, with z_{ref} accounting for most of z ($z_{\text{ref}} = 1.02z$). Consequently $\frac{z}{z_{\text{ref}}}$ is close to 1, and the expectations, $\mathbb{E}_{q(\lambda,\theta)} \left[\log \frac{q(\theta)}{q_{\text{ref}}(\theta)} \right]$, evaluated by MCMC for the remaining part of the integral are small. For the same reasons the variance at each λ is small, leading to favourable convergence within a small number of iterations. And finally $\mathbb{E}_{q(\lambda,\theta)} \left[\log \frac{q(\theta)}{q_{\text{ref}}(\theta)} \right]$ weakly depends on λ , so there is no need to use a very fine grid of λ values or consider optimal paths—satisfactory convergence is easily achieved using a simple geometric path with 4 λ -intervals.

2D pedagogical example with constrained parameters

As a second example, consider a 2-dimensional unnormalised density function with a constrained parameter space:

$$q(\theta_1, \theta_2) = \exp(-\Theta), \tag{10}$$

with

$$\Theta = \frac{1}{4} \sum_{ij \in \{1,2\}} \left(\theta_i + \frac{1}{2}\right)^{2j} + \frac{1}{8} \theta_1 \theta_2^2, \tag{11}$$

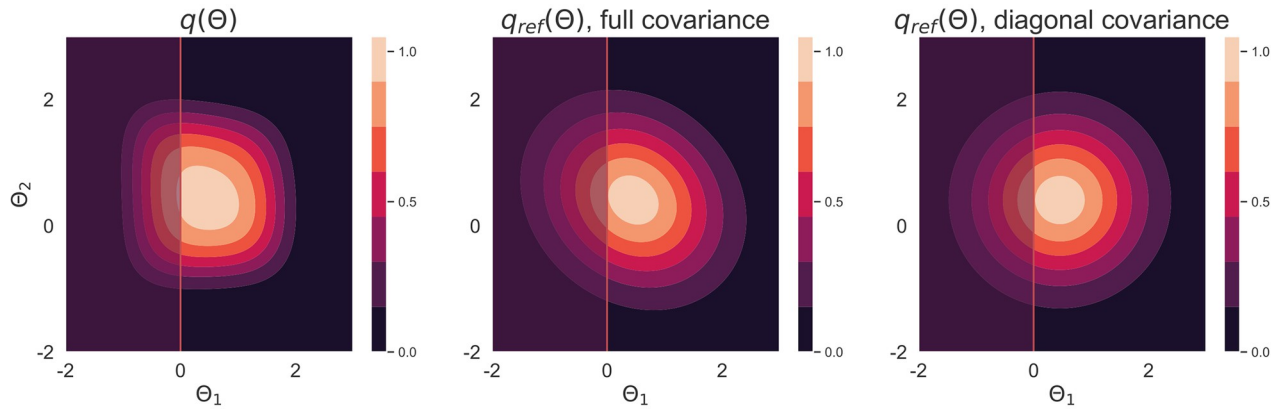


Fig 2. Contour plots of un-normalised densities. Contour plots of the un-normalised density q and its two reference densities q_{ref} , one using a full covariance matrix and another using a diagonal covariance matrix that can be easily marginalised. The red line shows the lower boundary $\theta_1 = 0$ and the shaded $\theta_1 < 0$ region to the left of the line is outside of the support of the density q .

<https://doi.org/10.1371/journal.pone.0289889.g002>

and

$$\theta_1 \in [0, +\infty) \text{ and } \theta_2 \in (-\infty, +\infty).$$

A reference density $q_{ref}(\theta)$ can be constructed from the Hessian at the mode of $q(\theta)$. To marginalise it on a parameter space with restricted support, we use a reference density $q_{ref}^{diag}(\theta)$ based on a diagonal Hessian, that has an exact and easy to calculate orthant. All densities are shown in Fig 2.

To obtain the log-evidence of the model, we calculated the exact value numerically [36, 37], and using a sampled diagonal covariance matrix, as per Eq 8, to account for the lower bound of the parameter θ_1 . Without this restriction the final normalising constant is overestimated— if the support of the parameters in the MCMC is not the same as for the analytic z_{ref} calculation, z_{ref} as shown in Eq 3 does not cancel with the TI reference. Numerical comparison of the referenced TI to quadrature is presented in the S2 Appendix.

Benchmarks—Radiata Pine

To benchmark the application of the referenced TI in the model selection task, two non-nested linear regression models are compared for the *radiata pine* data set [10]. This example has been widely used for testing normalising constant calculating methods, since in this instance the exact value of the model evidence can be computed. The data consists of 42 3-dimensional data-points, expressed as y_i —maximum compression strength, x_i —density and z_i —density adjusted for resin content. In this example, we follow the approach of [21], using the priors from therein, and test which of the two models M_1 and M_2 provides better predictions for the compression strength:

$$M_1 : y_i = \alpha + \beta(x_i - \bar{x}) + \epsilon_i, \epsilon_i \sim N(0, \tau^{-1}), i = 1, \dots, n,$$

$$M_2 : y_i = \gamma + \delta(z_i - \bar{z}) + \eta_i, \eta_i \sim N(0, \rho^{-1}), i = 1, \dots, n.$$

Five methods of estimating the model evidence were used in this example: Laplace approximation using a sampled covariance matrix, model switch TI along a path directly connecting the models [8, 38], referenced TI, power posteriors with equidistant 11 λ -placements (labelled

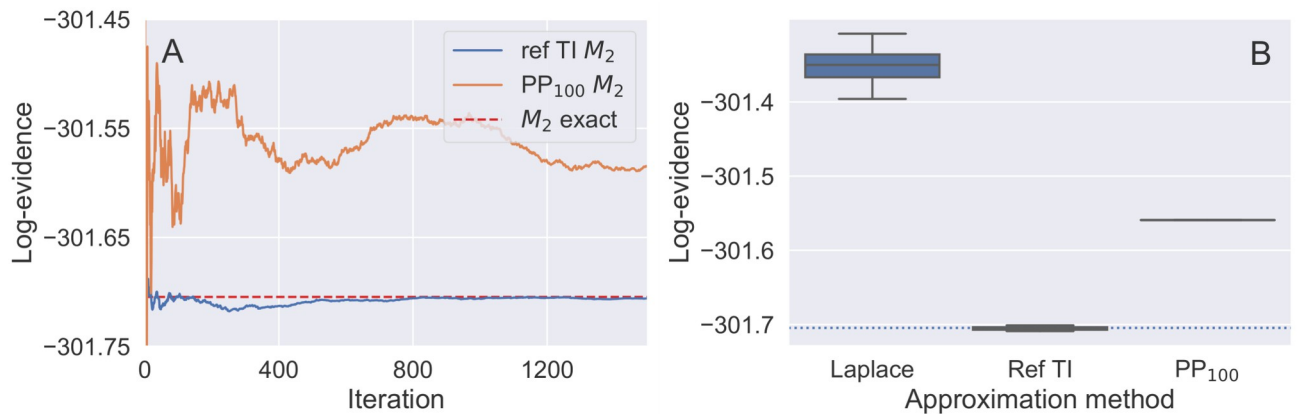


Fig 3. The estimated log-evidence of M_2 from the Radiata Pine benchmark problem. The log-evidence of M_2 from the Radiata Pine benchmark problem is shown estimated using three approaches. (A) shows the rolling mean of log-evidence of M_2 over 1500 iterations per λ obtained by referenced TI (blue line) and PP_{100} (orange line) methods. The exact value is shown with red dashed line. (B) shows the mean log-evidence of the model M_2 evaluated over 15 runs of the three algorithms. The exact value of the log-evidence is shown with the dotted line.

<https://doi.org/10.1371/journal.pone.0289889.g003>

here as PP_{11}) and power posteriors with 100 λ -s (PP_{100}), following the example from [21]. For the model switch TI, referenced TI and PP_{11} we used $\lambda \in \{0.0, 0.1, \dots, 1.0\}$.

The expectation from MCMC sampling per each λ for model switch TI, referenced TI, PP_{11} and PP_{100} and fitted cubic splines between the expectations are shown in the S2 Appendix. Both reference and model switch TI methods eliminate the problem of divergence of expectation for $\lambda = 0$, which is observed with the power posteriors, where samples for $\lambda = 0$ come from the prior density function. And both reference and model switch have smaller residuals for splines in λ fitted to $\mathbb{E}_{q(\lambda;\theta)} \left[\log \frac{q(\theta)}{q_{\text{ref}}(\theta)} \right]$ than power posteriors.

For each approach, the splines fitted to $\mathbb{E}_{q(\lambda;\theta)} \left[\log \frac{q(\theta)}{q_{\text{ref}}(\theta)} \right]$ were integrated to obtain the log-evidence for models M_1 and M_2 , and the log-ratio of the two models' evidences for the model switch TI. The rolling means of the integral over 1500 iterations for referenced TI and PP_{100} for M_2 are shown in Fig 3A. We can see from the plot, that referenced TI presents favourable convergence to the exact value, whereas PP_{100} oscillates around it. Fig 3B shows the distribution of log-evidence for the same model generated by 15 runs of the three algorithms: Laplace approximation with sampled covariance matrix, referenced TI and PP_{100} . Similar figures for model M_1 are given in the S2 Appendix. Although all three methods resulted in a log-evidence satisfactorily close to the exact solution, referenced TI was the most accurate and importantly, converged fastest (308 MCMC draws compared to 55,000 draws needed for the power posterior method to achieve standard error of 0.5%, excluding burn-in, see S2 Appendix).

Model selection for the COVID-19 epidemic in South Korea

The final example of using referenced TI for calculating model evidence is fitting a renewal model to COVID-19 case data from South Korea. The data were obtained from opendata.ecdc.europa.eu/covid19/casedistribution/csv. The model is based on a statistical representation of a stochastic branching process whose expectation mechanistically follows a renewal-type equation. Its derivation and details are explained in [39] and a short explanation of the model is provided in the S2 Appendix. Briefly, the model is fitted to the time-series case data and estimates a number of parameters, including serial interval and the effective reproduction

number, R_t . The number of cases for each day are modelled by a negative binomial likelihood, with location parameter estimated by a renewal equation. Three modifications of the original model are tested here:

- variation of the infection generation interval for values $GI = 5, 6, 7, 8, 9, 10, 20$ —where GI denotes the mean of Rayleigh-distributed generation interval,
- changing the order of the autoregressive model for the reproduction number, for $AR(k)$ with $k = 2, 3, 4$ lags,
- varying the length of the sliding window for estimating the reproduction number for values in $W = 1, 2, 3, 4, 7$ days.

Within each group of models, GI , AR and W , we want to select the best model through the highest evidence method. The dimension of each model was dependent on the modifications applied, but in all the cases the normalising constant was a 40- to 200-dimensional integral. The log-evidence of each model was calculated using the Laplace approximation with a sampled covariance matrix, and then correction to the estimate was obtained using referenced TI method. Values of the log-evidence for each model calculated by both Laplace and referenced TI methods are given in Table 1. Interestingly, the favoured model in each group, that is the model with the highest log-evidence, was different when the evidence was evaluated using the Laplace approximation than when it was evaluated with referenced TI. For example, using the Laplace method, sliding window of length 7 was incorrectly identified as the best model, whereas with referenced TI window of length 2 was chosen to be the best among the tested sliding windows models, which agrees with the previous studies of the window-length selection in H1N1 influenza and SARS outbreaks [40]. This exposes how essential it is to accurately determine the evidence, even good approximations can result in misleading results. Bayes factors for all model pairs are shown in the S2 Appendix.

Table 1. Log-evidence estimated by Laplace and referenced TI approximations. In each section, model with the highest log-evidence estimated by Laplace or referenced TI method is indicated in bold. The credible intervals for log-evidence comes from calculating the quantiles of the integral from Eq 2, where the integral values were obtained from the spline interpolated using running means of the expectations per λ over all iterations.

Model	Log-evidence Laplace	Log-evidence ref TI [95% CrI]
GI = 5	-1274	-716 [-715.6, -715.2]
GI = 6	-1274	-703 [-703.3, -702.7]
GI = 7	-1255	-732 [-732.4, -731.8]
GI = 8	-1245	-685 [-685.5, -684.7]
GI = 9	-1310	-803 [-802.8, -802.3]
GI = 10	-1313	-805 [-805.1, -805.3]
GI = 20	-1170	-796 [-796.3, -795.5]
AR(2)	-1207	-711 [-711.2, -710.6]
AR(3)	-1293	-704 [-704.7, -703.7]
AR(4)	-2166	-821 [-820.6, -819.2.]
W = 1	-1260	-802 [-802.1, -801.6]
W = 2	-1069	-791 [-791.2, -790.7]
W = 3	-1003	-807 [-807.5, -807.2]
W = 4	-940	-811 [-811.1, -810.7]
W = 7	-875	-814 [-813.7, -813.5]

<https://doi.org/10.1371/journal.pone.0289889.t001>

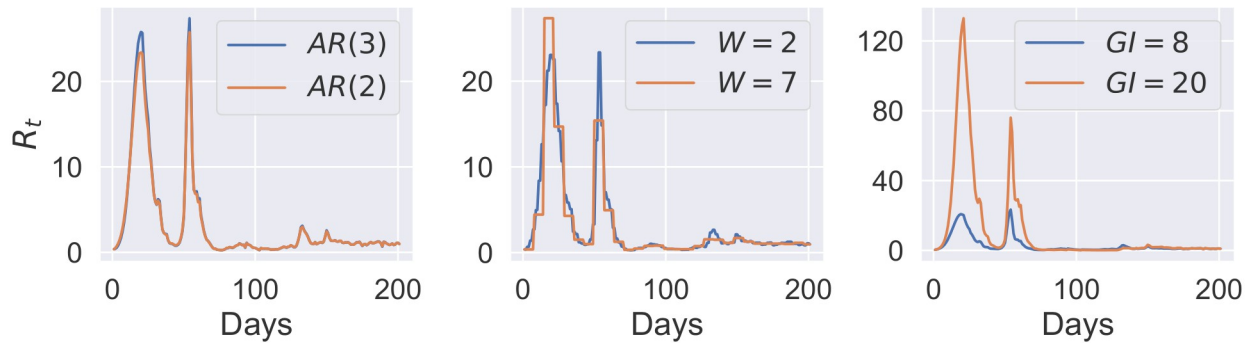


Fig 4. Time-dependent reproduction number. Time-dependent reproduction number generated by models with the highest evidence calculated using the Laplace approximation (orange lines) and referenced TI (blue lines). Note, the fitting data in this example contains superspreading events (which leads to very high values of R_t on certain days) so is not representative of SARS-CoV-2 transmission generally.

<https://doi.org/10.1371/journal.pone.0289889.g004>

Interpretation of the COVID-19 model selection

The importance of performing model selection in a rigorous way is clear from Fig 4, where the generated R_t time-series are plotted for the models favoured by Laplace and referenced TI methods (additional posterior densities are shown in the S2 Appendix). The differences in the R_t time-series show the pitfalls of selecting an incorrect model. The differences between the two favoured models were most extreme for the $GI = 8$ and $GI = 20$ models. While a $GI = 8$ is plausible, even likely for COVID-19, $GI = 20$ is implausible given observed data [41]. This is further supported by observing that for $GI = 20$, favoured by the Laplace method, R_t reached the value of over 125 in the first peak—around 100 more than for the $GI = 8$. The second peak was also largely overestimated, where R_t reached a value of 75.

We find it interesting to note that all models present a similar fit to the confirmed COVID-19 cases data (see S2 Appendix). This makes it impossible to select the best model through visual inspection and comparison of the model fits, or by using model selection methods that do not take the full posterior distributions into account. Although the models might fit the data well, other quantities generated, which are often of interest to the modeller, might be completely incorrect. Moreover, it emphasises the need to test multiple models before any conclusion or inference is undertaken, especially with the complex, hierarchical models.

Although often the Laplace approximation of the normalising constant is sufficient to pick the best model, it was not the case in this epidemiological model selection problem. We can see in Table 1 that the evidence was the highest for the “boundary” models when Laplace approximation was applied. For example, for the sliding window length models, when the Gaussian approximation was applied, the log-evidence was monotonically increasing with the value of W within the range of values that seem reasonable ($W = 1$ to 7). In contrast, with referenced TI, the log-evidence geometry is concave within the range of *a priori* reasonable parameters.

Discussion

The examples shown in Section Applications illustrate the applicability of the referenced TI approach for calculating model evidence. In the *radiata pine* example, referenced TI performed better than the other tested methods in terms of accuracy and speed. When using referenced TI, at $\lambda = 0$ values are sampled from the reference density rather than the prior as in the power posterior method, which should be closer to the original density (in the sense of

Kullback–Leibler or Jensen–Shannon divergence). This leads not only to a more accurate estimate of the normalising constant, but also much faster convergence of the MCMC samples. A detailed theoretical characterisation of rates of convergence is beyond the scope of this article, nonetheless the empirical tests presented consistently show faster convergence than with comparative approaches. This is useful especially for evaluating model evidence in complex hierarchical models where each MCMC iteration is computationally demanding.

Although referenced thermodynamic integration and other methods using path-sampling have theoretical asymptotically exact Monte Carlo estimator limits, in practice a number of considerations affect accuracy. For example, biases will be introduced to the referenced TI estimate if one endpoint density substantially differs from another. An example of this and explanation is included in the [S3 Appendix](#).

Furthermore, the discretisation of the coupling parameter path in λ can introduce a discretisation bias. For the power posteriors method, [9] propose an iterative way of selecting the λ -placements to reduce the discretisation error. [42] test multiple λ -placements for 2- and 20D regression models, and report relative bias for each tested scenario. In the referenced TI algorithm discretisation bias is however negligible—the use of the reference density results in TI expectations that are both small and have low variance, and therefore curvature with respect to λ . In our framework we use geometric paths with equidistant coupling parameters λ between the un-normalised posterior densities, but there are other possible choices of the path constructions, for example a harmonic [7] or hypergeometric path [38]. This optimisation might be worth exploring, however, as illustrated in [Fig 3B](#), the expectations evaluated vs λ are typically near-linear with referenced TI suggesting limited gains, although the extent of this will differ from problem to problem.

In the application to the renewal model for the COVID-19 epidemic in South Korea, we showed that for a complex structured model, hypothesis selection by Laplace approximation of the normalising constant can give misleading results. Using referenced TI, we calculated model evidence for 16 models, which enabled a quick comparison between chosen pairs of competing models. Importantly, the evidence given by the referenced TI was not monotonic with the increase of one of the parameters, which was the case for the Laplace approximation. The referenced TI presented here will similarly be useful in other situations particularly where the high-dimensional posterior distribution is uni-modal but non-Gaussian.

Conclusions

Normalising constants are fundamental in Bayesian statistics. In this paper we give an account of referenced thermodynamic integration (TI), in terms of theoretical consideration regarding the choice of reference, and show how it can be applied to realistic practical problems. We show how referenced TI allows efficient calculation of a single model's evidence by sampling from geometric paths between the un-normalised density of the model and a judiciously chosen reference density—here, a sampled multivariate normal that can be generated and integrated with ease. Referenced TI method has practical utility for substantially challenging problems of model selection in epidemiology and we suggest it has applicability in other fields of applied machine learning that rely on high-dimensional Bayesian models.

Supporting information

S1 Appendix. Variational Laplace reference.
(PDF)

S2 Appendix. Applications.

(PDF)

S3 Appendix. Bias and variance.

(PDF)

Acknowledgments

The authors would like to thank Microsoft Research for the donated cloud compute time, and the Stan Forums community for maintaining the space for collaborative and helpful discussions.

Author Contributions

Conceptualization: Iwona Hawryluk, Swapnil Mishra, Samir Bhatt.

Formal analysis: Iwona Hawryluk.

Funding acquisition: Samir Bhatt.

Investigation: Iwona Hawryluk, Samir Bhatt.

Methodology: Iwona Hawryluk, Swapnil Mishra, Seth Flaxman, Samir Bhatt.

Project administration: Samir Bhatt.

Software: Iwona Hawryluk, Swapnil Mishra, Samir Bhatt.

Supervision: Swapnil Mishra, Samir Bhatt.

Visualization: Iwona Hawryluk.

Writing – original draft: Iwona Hawryluk.

Writing – review & editing: Iwona Hawryluk, Swapnil Mishra, Seth Flaxman, Samir Bhatt.

References

1. Kass RE, Raftery AE. Bayes Factors. *Journal of the American Statistical Association*. 1995; 90(430):773–795. <https://doi.org/10.1080/01621459.1995.10476572>
2. Bennett CH. Efficient estimation of free energy differences from Monte Carlo data. *Journal of Computational Physics*. 1976; 22(2):245–268. [https://doi.org/10.1016/0021-9991\(76\)90078-4](https://doi.org/10.1016/0021-9991(76)90078-4)
3. Meng XL, Wong WH. Simulating ratios of normalising constants via a simple identity: a theoretical exploration. *Statistica Sinica*. 1996; 6(4):831–860.
4. Skilling J. Nested Sampling for General Bayesian Computation. *Bayesian Analysis*. 2006; 1(4):833–860. <https://doi.org/10.1214/06-BA127>
5. Habeck M. Evaluation of marginal likelihoods via the density of states. In: *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics*. vol. 22 of *Proceedings of Machine Learning Research*. La Palma, Canary Islands: PMLR; 2012. p. 486–494. Available from: <https://proceedings.mlr.press/v22/habeck12.html>.
6. Kirkwood JG. Statistical Mechanics of Fluid Mixtures. *The Journal of Chemical Physics*. 1935; 3(5):300–313. <https://doi.org/10.1063/1.1749657>
7. Gelman A, Meng XL. Simulating normalizing constants: from importance sampling to bridge sampling to path sampling. *Statistical Science*. 1998; 13(2):163–185. <https://doi.org/10.1214/ss/1028905934>
8. Lartillot N, Philippe H. Computing Bayes Factors Using Thermodynamic Integration. *Systematic Biology*. 2006; 55(2):195–207. <https://doi.org/10.1080/10635150500433722> PMID: 16522570
9. Friel N, Hurn M, Wyse J. Improving power posterior estimation of statistical evidence. *Statistics and Computing*. 2017; p. 1165–1180. <https://doi.org/10.1007/s11222-016-9678-6>
10. Williams EJ. *Regression Analysis*. Wiley; 1959.
11. Neal RM. *Probabilistic Inference Using Markov Chain Monte Carlo Methods*. 1993;.

12. Diccio TJ, Kass RE, Raftery A, Wasserman L. Computing Bayes Factors by Combining Simulation and Asymptotic Approximations. *Journal of the American Statistical Association*. 1997; 92(439):903–915. <https://doi.org/10.1080/01621459.1997.10474045>
13. Grosse RB, Maddison CJ, Salakhutdinov RR. Annealing between distributions by averaging moments. In: *Advances in Neural Information Processing Systems*. vol. 26; 2013. Available from: <https://proceedings.neurips.cc/paper/2013/file/fb60d411a5c5b72b2e7d3527cfc84fd0-Paper.pdf>.
14. Dirac PAM, Bohr NHD. The quantum theory of the emission and absorption of radiation. *Proceedings of the Royal Society of London Series A, Containing Papers of a Mathematical and Physical Character*. 1927; 114(767):243–265. <https://doi.org/10.1098/rspa.1927.0039>
15. Gell-Mann M, Brueckner KA. Correlation energy of an electron gas at high density. *Physical Review*. 1957; 106(2):364–368. <https://doi.org/10.1103/PhysRev.106.369>
16. Vočadlo L, Alfè D. Ab initio melting curve of the fcc phase of aluminum. *Phys Rev B*. 2002; 65. <https://doi.org/10.1103/PhysRevB.65.214105>
17. Duff AI, Davey T, Korbmacher D, Glensk A, Grabowski B, Neugebauer J, et al. Improved method of calculating ab initio high-temperature thermodynamic properties with application to ZrC. *Physical Review B*. 2015; 91(21):214311. <https://doi.org/10.1103/PhysRevB.91.214311>
18. Xie W, Lewis PO, Fan Y, Kuo L, Chen MH. Improving Marginal Likelihood Estimation for Bayesian Phylogenetic Model Selection. *Systematic Biology*. 2010; 60(2):150–160. <https://doi.org/10.1093/sysbio/syq085> PMID: 21187451
19. Cameron E, Pettitt A. Recursive Pathways to Marginal Likelihood Estimation with Prior-Sensitivity Analysis. *Statistical Science*. 2014; 29(3):397–419. <https://doi.org/10.1214/13-STS465>
20. Friel N, Pettitt AN. Marginal likelihood estimation via power posteriors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2008; 70(3):589–607. <https://doi.org/10.1111/j.1467-9868.2007.00650.x>
21. Friel N, Wyse J. Estimating the evidence: a review. *Statistica Neerlandica*. 2012; 66(3):288–308. <https://doi.org/10.1111/j.1467-9574.2011.00515.x>
22. Lefebvre G, Steele R, Vandal AC. A path sampling identity for computing the Kullback–Leibler and J divergences. *Computational Statistics & Data Analysis*. 2010; 54(7):1719–1731. <https://doi.org/10.1016/j.csda.2010.01.018>
23. Fan Y, Wu R, Chen MH, Kuo L, Lewis PO. Choosing among partition models in Bayesian phylogenetics. *Molecular Biology and Evolution*. 2011; 28(1):523–532. <https://doi.org/10.1093/molbev/msq224> PMID: 20801907
24. Baele G, Lemey P, Suchard MA. Genealogical Working Distributions for Bayesian Model Testing with Phylogenetic Uncertainty. *Systematic Biology*. 2016; 65(2):250–264. <https://doi.org/10.1093/sysbio/syv083> PMID: 26526428
25. Ridgway J. Computation of Gaussian orthant probabilities in high dimension. *Statistics and computing*. 2016; 26(4):899–916. <https://doi.org/10.1007/s11222-015-9578-1>
26. Azzimonti D, Ginsbourger D. Estimating Orthant Probabilities of High-Dimensional Gaussian Vectors with An Application to Set Estimation. *Journal of Computational and Graphical Statistics*. 2016; 27:255–267. <https://doi.org/10.1080/10618600.2017.1360781>
27. Owen DB. Orthant Probabilities. *Wiley StatsRef: Statistics Reference Online*. 2014. <https://doi.org/10.1002/9781118445112.stat01097>
28. Miwa T, Hayter AJ, Kuriki S. The evaluation of general non-centred orthant probabilities. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2003; 65(1):223–234. <https://doi.org/10.1111/1467-9868.00382>
29. Curnow RN, Dunnett CW. The Numerical Evaluation of Certain Multivariate Normal Integrals. *The Annals of Mathematical Statistics*. 1962; 33(2):571–579. <https://doi.org/10.1214/aoms/1177704581>
30. Ruben H. An asymptotic expansion for the multivariate normal distribution and Mills' ratio. *Journal of Research of the National Bureau of Standards Section B Mathematics and Mathematical Physics*. 1964; p. 3. <https://doi.org/10.1017/S1446788700026872>
31. Brown M. A generalized error function in n dimensions. US Naval Missile Center, Theoretical Analysis Division. 1963;.
32. Hoffman MD, Gelman A. The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *J Mach Learn Res*. 2014; 15(1):1593–1623.
33. Carpenter B, Gelman A, Hoffman M, Lee D, Goodrich B, Betancourt M, et al. Stan: A Probabilistic Programming Language. *Journal of Statistical Software, Articles*. 2017; 76(1):1–32. <https://doi.org/10.18637/jss.v076.i01>

34. Phan D, Pradhan N, Jankowiak M. Composable Effects for Flexible and Accelerated Probabilistic Programming in NumPyro. arXiv preprint. 2019. <https://doi.org/10.48550/arXiv.1912.11554>
35. Gelman A, Rubin DB. Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*. 1992; 7(4):457–472. <https://doi.org/10.1214/ss/1177011136>
36. Piessens R, deDoncker Kapenga E, Ueberhuber C, Kahaner D. QUADPACK: A Subroutine Package for Automatic Integration. Springer; 1983.
37. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*. 2020; 17:261–272. <https://doi.org/10.1038/s41592-019-0686-2> PMID: 32015543
38. Vitoratou S, Ntzoufras I. Thermodynamic Bayesian model comparison. *Statistics and Computing*. 2017; p. 1165–1180. <https://doi.org/10.1007/s11222-016-9678-6>
39. Berah T, Mellan TA, Miscouridou X, Mishra S, Parag KV, Pakkanen MS, et al. Unifying the effective reproduction number, incidence, and prevalence under a stochastic age-dependent branching process. arXiv preprint. 2021. <https://doi.org/10.48550/arXiv.2107.05579>
40. Parag KV, Donnelly CA. Using information theory to optimise epidemic models for real-time prediction and estimation. *PLOS Computational Biology*. 2020; 16(7):1–20. <https://doi.org/10.1371/journal.pcbi.1007990> PMID: 32609732
41. Bi Q, Wu Y, Mei S, Ye C, Zou X, Zhang Z, et al. Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. *The Lancet Infectious Disease*. 2020; 20:911–919. [https://doi.org/10.1016/S1473-3099\(20\)30287-5](https://doi.org/10.1016/S1473-3099(20)30287-5) PMID: 32353347
42. Calderhead B, Girolami M. Estimating Bayes factors via thermodynamic integration and population MCMC. *Computational Statistics & Data Analysis*. 2009; 53(12):4028–4045. <https://doi.org/10.1016/j.csda.2009.07.025>