e-ISSN 2320 –7876 www.ijfans.org
Vol.11, Iss.9, Dec 2022

CONTRASTIVE LEARNING AND NEW COMPUTATIONAL TOOLS FOR BAYESIAN INFERENCE

NIYAZ AHMED. A
Research Scholar
M.Phil Mathematics
Bharath Institute Of Higher Education And Research
Mail Id : anbu.cs10@gmail.com

Guide Name: **Dr. M. KAVITHA**
Assistant Professor, Department Of Mathematics
Bharath Institute Of Higher Education And Research

Address for Correspondence

NIYAZ AHMED. A
Research Scholar
M.Phil Mathematics
Bharath Institute Of Higher Education And Research
Mail Id : anbu.cs10@gmail.com

Abstract

Bayesian inference is the process of fitting a probability model to a set of data and summarizing the result by a probability distribution on the parameters of the model and on unobserved quantities such as predictions for new observations. Bayesian prediction plays a significant part in different extents of applied statistics. Bayesian approach has many benefits in statistical modelling and data analysis. It offers a system of validating the method of knowledge from data to update beliefs in accord with recent notions of knowledge synthesis. Bayesian approaches usually need less sample data to attain the same quality of implications than approaches based on sampling theory, which become very significant in the case of expensive testing processes. Bayesian inference has been used in various fields such as computer science, reliability analysis, etc.Humans, being a part of the ecosystem, have their own roles to be carried out. The human body is an integrated system, which performs different

functions excretion, respiration, circulation, digestion, endocrine, intellectual and locomotion. The homeostasis is well maintained so that every organ performs its respective functions. Lungs chiefly help in the oxygenation of blood. Kidney excretes the metabolic end products. Heart pumps so that blood transfers oxygen to tissues and takes up carbon dioxide, which is then excreted through the lungs. The nervous system chiefly coordinates all the functions, makes one perceive sensations and also carry out movements. The food one eats must be digested and absorbed to give energy for one's daily needs.Hence when any of these function fail, the entire system gets collapsed as they are closely interrelated.

INTRODUCTION

The human system has its own immune barriers to protect itself against infections, but once the infection sets in, the immune mechanisms come into action. The immune system principally includes lymphocytes and other leucocytes, antigen presenting cells and their chemical mediators. One of the most important infectious diseases with high mortality rate in developing countries is Tuberculosis. Tuberculosis (TB) is the seventh most common disease in the world. India ranks first in the absolute number of incident TB cases diagnosed every year. Tuberculosis is instigated by the bacteria named Mycobacterium tuberculosis. The disease is highly contagious which usually spreads by air droplets and is frequently encountered in immune compromised individuals and in lower socio-economic classes due to overcrowding and malnutrition.

The changing aspects of contagious diseases depends on the probability of coincidence of hosts and pathogens and the spatial distribution among them. The communication of contagious pathogens from affected to vulnerable masses decreases when the space between persons increase. The disease Tuberculosis, falls in the same category. It depends on spatial accumulation or gathering. The spatial correlation depends on the amount of mingling of the population in big cities with huge population of highly movable entities. In this thesis, Bayesian approach is implemented in modelling the effects of TB and forecasting the spatial distribution of TB using various methods.

Previous disease mapping works were based on collating, mapping and analysing prevalence or incidence data with conventional statistical approaches, which are affected by random variation due to population variability and a loss of statistical power when cases are assigned to subgroups (e.g. several geographic subareas). Differences in geographic distribution due to chance may be incorrectly interpreted as true variation of epidemiological interest. The observed extreme values may not reflect the true spatial distribution of the disease, instead may reflect those of the population area. The Bayesian method can overcome these problems as it can model the random and true variation separately and is an attractive alternative to the frequentist approach. Bayesian methods can provide some shrinkage and spatial smoothing of raw standardized incidence ratio estimates, which are strongly influenced by the size of the population at risk, resulting in a noisy and blurred picture of the true unobserved risks.

Consider that the data values z = (z_1,.....z_n) are found independently. The likelihood function is given by

$$L(\phi|z) = p(z_1, z_2, \ldots\ldots, z_n|\phi)$$

Once the data have been observed, in order to obtain the posterior distribution p(ϕ/z) and the probability distribution

The subjective probability is based on the past experiences, and it might be unrealistic. Bayesian approach takes into account any prior knowledge of the experiment that the statistician has, and it is one application of the principle of statistical inference that may be called Bayesian statistics. The prior distribution reflects the subjective belief of the random variable before the sample is drawn, while the posterior distribution is the conditional distribution if the random variable regulates probabilities of events in the following manner:

$P\ (H_0 \mid A) = P(\ A|\ H_0\ )\ P\ (H_0\ )|P\ (A)$ where H_0denotes the null hypothesis, that was associated before the new event A becomes existing. $P(H_0)$ is known as the prior probability of H_0. $P\ (A|\ H_0\ )$ is the conditional probability of considering the event A given that the hypothesis H_0 is true. It is known as the likelihood function when $P(A|H_0)$ is expressed

as a function of H_0 given A. The marginal probability of P(A) of A is determined as the sum of the product of all probabilities of mutually exclusive hypotheses and corresponding conditional probabilities:

$$\sum P(A|H_i)P(H_i)$$

$P(H_0 | A)$ is the posterior probability of H_0 given A.

$$\text{i.e. } P(H_0|A) = \frac{p(A|H_0)}{\sum_i P(A|H_i)P(H_i)}$$

For continuous case, the posterior distribution is

$$P(\phi|z) = \frac{p(\phi)p(z|\phi)}{\int p(\phi)p(z|\phi)d\phi} = \frac{p(\phi)L(\phi|z)}{p(z)} \propto p(\phi)L(\phi|z)$$

BAYESIAN INFERENCE

In reliability analysis, hazard rate plays an indispensable role to characterize life phenomena. In fact, the hazard rate usually is more informative about the underlying mechanism of failure than the other representatives of a lifetime distribution. Technically, failure or hazard rate represents the propensity of a device of age t to fail in the small interval of time t to t + dt. The parametric models, such as gamma, Weibull, and log-normal distribu tions, which are commonly used lifetime distributions display monotone fail ure rates. However, many physical phenomena exhibit failure rates which are non-monotonic. For example, the failure pattern of many mechanical and elec tronic components comprise of three stages: initial stage (or burn-in) where failure is high at the beginning of the product life cycle due to design and manufacturing problems, and decreases towards a constant level, the middle stage with an approximately constant failure rate, which is followed by a final stage (or wear-out phase), from where the failure rate starts to increase. Such failure rates are usually termed as bathtub (BT) or U shaped.

The aforemen tioned models which allow only monotone failure rates are unable to produce bathtub curves and thus cannot adequately interpret data with this character. Bathtub models are possibly more realistic models than monotone failure rate models and have been widely accepted in the field of medicine and are partic ularly useful in reliability related decision making and cost analysis.

Characterization of failure rate function

The role of the parameter γ in determining different shapes of the failure rate function can be studied under two situations:

Case 1: $\gamma \geq 1$

i For any $t > 0$, $h^0(t) > 0$, thus, $h(t)$ is an increasing function.

ii $h(t) \to +\infty$ as $t \to +\infty$.

Case 2: $0 < \gamma < 1$

i Letting $h'(t_0) = 0$, we obtain $t_0 = \alpha\left(\frac{1-\gamma}{\gamma}\right)^{1/\gamma}$.
It is evident that when $0 < \gamma < 1$, t_0 exists and is finite. For $t < t_0$, $h(t)$ is decreasing while it is increasing for $t > t_0$ showing a bathtub shaped property.

ii $h(t) \to \infty$ for $t \to 0$ and $t \to +\infty$.

Thus, we see that the hazard function is exponentially increasing for large t and has a bathtub-shape with achieving a minimum value at t_0 when $0 < \gamma < 1$. These two properties make it a useful alternative to Weibull distribu tion for modeling lifetimes. Figure 2.2 illustrates $h(t)$ with various values of parameters γ and α.

e-ISSN 2320 –7876 www.ijfans.org
Vol.11, Iss.9, Dec 2022

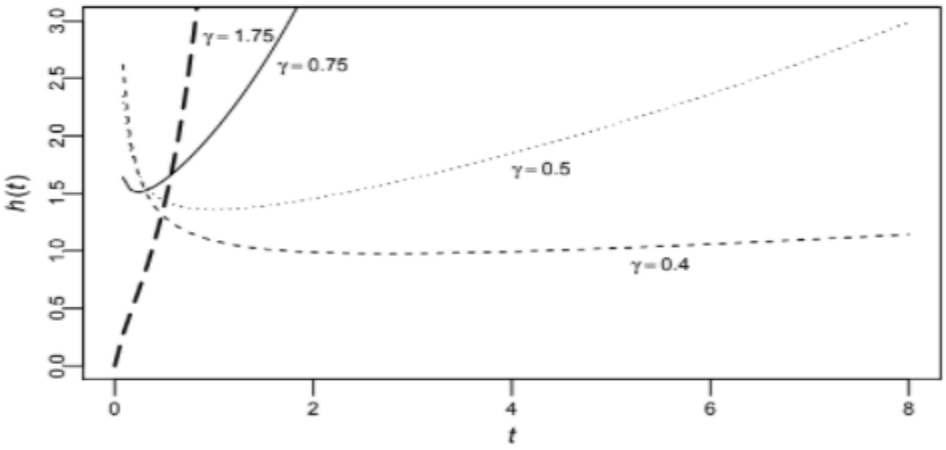

Figure 2.2: *Plot of the failure rate function* $h(t)$ *with* $\alpha = 1$ *and* γ *changing from* 0.4 *to* 1.75. *It is evident that the hazard function for* $\gamma < 1$ *assumes a bathtub shape while for* $\gamma \geq 1$, *it increases exponentially.*

2.3 Model formulation

The Bayesian analysis of concerned reliability model begins with the specifica tion of the likelihood function. For this, let us assume that $t : t_1, t_2, \ldots, t_n$ be the observed lifetimes from exponential power model (2.1). The corresponding likelihood function can be defined as

$$p(\boldsymbol{t} \mid \gamma, \alpha) = \left(\frac{\gamma}{\alpha^{\gamma}}\right)^{n} \prod_{i=1}^{n} (t_i)^{\gamma-1} \exp\left[\sum_{i=1}^{n} \left(\frac{t_i}{\alpha}\right)^{\gamma}\right] \exp\left(n - \sum_{i=1}^{n} e^{(t_i/\alpha)^{\gamma}}\right) \quad (2.2)$$

The next step in Bayesian analysis is to choose a prior distribution that expresses the uncertainty about the parameters of the model, before the data is observed. We considered an independent and weakly informative prior distributions for the parameters. Both the positive parameters are assumed to be half-Cauchy distributed according to their hyperparameters, scale = 25 and are denoted by

$$\begin{aligned} \gamma &\sim \text{half-Cauchy}(25) \\ \alpha &\sim \text{half-Cauchy}(25) \end{aligned} \quad (2.3)$$

3.2 Terms and concepts

We here define some of the important terms and concepts frequently used in the design of experiments;

(a) *Experiment* is a test or series of runs in which purposeful changes are made to the input variables of a process or system so that we may observe and identify the reasons for changes that may be observed in the response.

(b) *Run* is an experimental condition or factor level combination at which responses are measured.

(c) *Experimental units* are the recipients of the experimental treatments.

(d) *Treatments* are the different procedures we want to compare. These could be different voltages to which some electronic devices are subjected or the different drug therapies given to a set of patients.

(e) *Factor* is an explanatory variable that can be manipulated by the exper imenter. Each factor has two or more levels (i.e., different values of the factor). Technically, a combination of factor levels that is assigned to the experimental units is termed as treatment.

(f) *Response* is the outcome that we observe after applying the treatment to an experimental unit.

(g) *Randomization* is a schedule for allocating treatment material and for conducting treatment combinations in a designed experiment, such that the conditions in one run neither depend on the conditions of the previous run nor predict the conditions in the subsequent runs. Other aspects of an experiment can also be randomized: for example, the order in which units are evaluated for their responses.

(h) *Experimental error* is the random variation present in all experimental results. Different experimental units will give different responses to the same treatment, and it is often

true that applying the same treatment over and over again to the same unit will result in different responses in different trials.

Next, we consider the Bayesian regression analysis of a reliability experiment when there is a single factor involved.

BAYESIAN COMPUTATIONAL TOOLS

Some computational challenges

The starting point of a Bayesian analysis being the posterior distribution, let us recall that it is defined by the product

$$\pi(\theta|x) \propto \pi(\theta) f(x|\theta)$$

where θ denotes the parameter and x the data. (The symbol $\propto$ means that the functions on both sides of the symbol are proportional as functions of θ, the missing constant being a function of x, $m(x)$.) The structures of both θ and x can vary in complexity and dimension, although we will not discuss the non parametric case when θ is infinite dimensional, referring the reader to Holmes et al. (2002) for an introduction. The prior distribution is most often available in closed form, being chosen by the experimenter, while the likelihood function $f(x|\theta)$ may be too involved to be computed even for a given pair (x, θ). In special cases where $f(x|\theta)$ allows for a demarginalisation representation

$$f(x|\theta) = \int f(x, z|\theta)\, \mathrm{d}z\,,$$

where $g(x, z|\theta)$ is a (manageable) probability density, we will call z the missing data. However, the existence of such a representation does not necessarily implies it is of any use in computations. (We will encounter both cases in Sections 4 and 5.)

Since the posterior distribution is defined by

$$\pi(\theta|x) = \pi(\theta)f(x|\theta) \Big/ \int_\Theta \pi(\theta)f(x|\theta)\,\mathrm{d}\theta$$

a first difficulty occurs because of the normalising constant: the denominator is very rarely available in closed form. This is an issue only to the extent that the posterior density is defined up to a constant. In cases where the constant does not matter, inference can be easily conducted without the constant. Cases when the constant matters include testing and model choice, since the marginal likelihood

$$m(x) = \int_\Theta \pi(\theta)f(x|\theta)\,\mathrm{d}\theta$$

is central to the Bayesian procedures addressing this inferential problem. Indeed, when comparing two models against the same dataset x, the prefered Bayesian solution (see, e.g., Robert, 2001, Chapter 5, or Jeffreys, 1939) is to use *the Bayes factor*, defined as the ratio of marginal likelihoods

$$\mathfrak{B}_{12}(x) = \frac{m_1(x)}{m_2(x)} = \frac{\int_{\Theta_1} \pi(\theta_1)f(x|\theta_1)\,\mathrm{d}\theta_1}{\int_{\Theta_2} \pi(\theta_2)f(x|\theta_2)\,\mathrm{d}\theta_2}\,,$$

and compared to 1 to decide which model is most supported by the data (and how much). Such a tool—quintessential for running a Bayesian test—means that for almost any inference problem—barring the very special case of conjugate priors— there is a computational issue, not the most promising feature for promoting an inferential method. This aspect has obviously been addressed by the community, see for instance Chen et al. (2000) that is entirely dedicated to the problem of approximating normalising constants or ratios of normalising constants, but I regret the issue is not spelled out much more clearly as one of the major computational challenges of Bayesian statistics (see also Marin and Robert, 2011).

Example 1 As a benchmark, consider the case (Marin et al., 2011a) when a sample (x_1, . . . , x_n) can be issued either from a normal $N(\mu, 1)$ distribution or from a double-exponential $L(\mu, 1/\sqrt{2})$ distribution with density

$$f_0(x|\mu) = \frac{1}{\sqrt{2}} \exp\{-\sqrt{2}|x-\mu|\}\,.$$

e-ISSN 2320 –7876 www.ijfans.org
Vol.11, Iss.9, Dec 2022

(This case was suggested to us by a referee of Robert et al., 2011, however I should note that a similar setting opposing a normal model to (simple) ex ponential data used as a benchmark in Ratmann (2009) for ABC algorithms.) Then, as it happens, the Bayes factor $B_{01}(x_1, \ldots, x_n)$ is available in closed form, since, under a normal $\mu \sim N(0, \sigma^2)$ prior, the marginal likelihood for the normal model is given by

$$\begin{aligned} m_1(x_1,\ldots,x_n) &= \int (2\pi)^{-n/2} \prod_{i=1} \exp\{-(x_i-\mu)^2/2\} \exp\{-\mu^2/2\sigma^2\}\, \mathrm{d}\mu/\sqrt{2\pi}\sigma \\ &= (2\pi)^{-n/2} \exp\{-\sum_{i=1}^{n}(x_i-\bar{x}_n)^2/2\} \\ &\quad \times \int \exp[-\{(n+\sigma^{-2})\mu^2 - 2n\mu\bar{x}_n + n(\bar{x}_n)^2\}/2]\, \mathrm{d}\mu/\sqrt{2\pi}\sigma \\ &= (2\pi)^{-n/2} \exp\{-\sum_{i=1}^{n}(x_i-\bar{x}_n)^2/2\} \\ &\quad \times \exp\{-n\sigma^{-2}(\bar{x}_n)^2/2(n+\sigma^{-2})\}/\sigma\sqrt{n+\sigma^{-2}} \end{aligned}$$

and, for the double-exponential model, by (assuming the sample is sorted)

$$\begin{aligned} m_0(x_1,\ldots,x_n) &= \int 2^{-n/2} \prod_{i=1}^{n} \exp\{-\sqrt{2}|x_i-\mu|\} \exp\{-\mu^2/2\sigma^2\}\, \mathrm{d}\mu/\sqrt{2\pi}\sigma \\ &= \frac{2^{-n/2}}{\sqrt{2\pi}\sigma} \sum_{i=0}^{n} \int_{x_i}^{x_{i+1}} \prod_{j=1}^{i} e^{\sqrt{2}x_j - \sqrt{2}\mu} \prod_{j=i+1}^{n} e^{-\sqrt{2}x_j + \sqrt{2}\mu} e^{-\mu^2/2\sigma^2}\, \mathrm{d}\mu \\ &= \frac{2^{-n/2}}{\sqrt{2\pi}\sigma} \sum_{i=0}^{n} \int_{x_i}^{x_{i+1}} e^{\sqrt{2}\sum_{j=1}^{i} x_j - \sqrt{2}\sum_{j=i+1}^{n} x_j + \sqrt{2}(n-2i)\mu} e^{-\mu^2/2\sigma^2}\, \mathrm{d}\mu \\ &= 2^{-n/2} \sum_{i=0}^{n} e^{\sqrt{2}\sum_{j=1}^{i} x_j - \sqrt{2}\sum_{j=i+1}^{n} x_j + 2(n-2i)^2\sigma^2/2} \\ &\quad \times \int_{x_i}^{x_{i+1}} e^{-\{\mu - \sqrt{2}(n-2i)\sigma^2\}^2/2\sigma^2}\, \mathrm{d}\mu/\sqrt{2\pi}\sigma \\ &= 2^{-n/2} \sum_{i=0}^{n} e^{\sqrt{2}\sum_{j=1}^{i} x_j - \sqrt{2}\sum_{j=i+1}^{n} x_j + (n-2i)^2\sigma^2} \\ &\quad \times \left[\Phi(\{x_{i+1} - \sqrt{2}(n-2i)\sigma^2\}/\sigma) - \Phi(\{x_i - \sqrt{2}(n-2i)\sigma^2\}/\sigma)\right] \end{aligned}$$

with obvious conventions when $i = 0$ ($x_0 = -\infty$) and $i = n$ ($x_{n+1} = +\infty$). To illustrate the consistency of the Bayes factor in this setting, Figure 1 represents the distributions of the Bayes factors associated with 100 normal and 100 double exponential samples of sizes 50 and 200, respectively. While the smaller samples see much overlay in the repartition of the Bayes factors, for 200 observations, in both models, the log-Bayes factor distribution concentrates on the proper side of zero, meaning that it discriminates correctly between the two distributions for a large enough sample size.

Another recurrent difficulty with using posterior distributions for infer ence is the derivation of credible sets— the Bayesian version of confidence sets (see, e.g., Robert, 2001)— since they are usually defined as highest poste rior density regions:

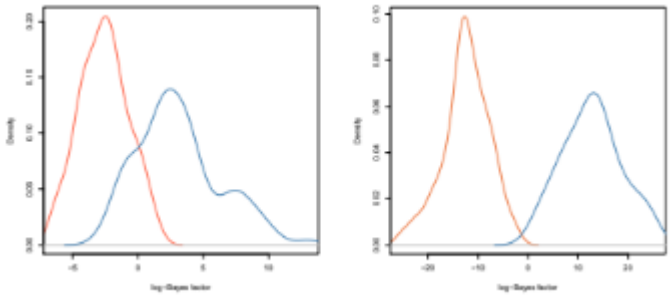

Fig 1. *Repartition of the values of the log Bayes factors associated with 100 normal (orange) and 100 double-exponential sam ples (blue) of size 50 (left) and 200 (right), estimated by the default R density estima tor.*

$$C_\alpha(x) = \{\theta;\, \pi(\theta|x) \geq \kappa_\alpha(x)\}\,,$$

where the bound k_α is determined by the credibility of the set $\mathrm{P}(\theta \in C_\alpha(x)|x) = \alpha$.

While the normalisation constant is irrelevant in this problem, determining the collection of parameter values such that $\pi(\theta)f(x|\theta) \geq \kappa_\alpha(x)$ and calibrating the lower bound $\kappa_\alpha(x)$ on the product $\pi(\theta)f(x|\theta)$ to achieve proper coverage are non-trivial problems that require advanced simulation methods. Once again, the issue is somehow overlooked in the literature.

e-ISSN 2320 –7876 www.ijfans.org
Vol.11, Iss.9, Dec 2022

While one of the major appeals of Bayesian inference is that it is not reduced to an estimation technique—but on the opposite offers a whole range of inferen tial tools to analyse the data against the proposed model—, the computation of Bayesian estimates is nonetheless certainly one of the better addressed compu tational issues. This is especially true for posterior moments like the posterior mean $\mathrm{E}^{\pi}[\theta|x]$ since they are directly represented as ratios of integrals

$$\mathbb{E}^{\pi}[\theta|x] = \frac{\int_{\Theta} \theta\pi(\theta) f(x|\theta)\,\mathrm{d}\theta}{\int_{\Theta} \pi(\theta) f(x|\theta)\,\mathrm{d}\theta}\,.$$

The computational problem may however get involved for several reasons, in cluding for instance

– the space Θ is not Euclidean and the problem imposes shape constraints (as in some time series models);

– the dimension of Θ is large (as in non-parametrics);

– the estimator is the solution to a fixed point problem (as in the credible set definition);

– simulating from $\pi(\theta|x)$ is delicate or even impossible;

the latter case being in general the most challenging and thus the most studied, as the following sections will show.

3. Monte Carlo methods

Monte Carlo methods have been introduced by physicists in Los Alamos, namely Ulam, von Neumann, Metropolis, and their collaborators in the 1940's (see Robert and Casella, 2011). The idea behind Monte Carlo is a straightforward ap plication of the *law of large numbers*, namely that, when x_1, x_2, are i.i.d. from the distribution f, the empirical average converges (almost surely) to $\mathrm{E}_f[h(X)]$

$$\frac{1}{T}\sum_{t=1}^{T} h(x_t)$$

e-ISSN 2320 –7876 www.ijfans.org
Vol.11, Iss.9, Dec 2022

when T goes to $+\infty$. While this per spective sounds too simple to apply to complex problems—either because the simulation from f itself is intractable or because the variance of the empirical average is too large to be manageable—, there exist more advanced exploitations of this result that lead to efficient simulation solutions.

Example 1 (bis) Consider computing the Bayes factor

$B_{01}(x_1, \ldots, x_n) = m_0(x_1, \ldots, x_n)/m_1(x_1, \ldots, x_n)$

by simulating a sample $(\mu_1, \ldots, \mu_T)$ from the prior distribution, $N(0, \sigma^2)$. The approximation to the Bayes factor is then provided by

$$\widehat{\mathfrak{B}_{01}} = \sum_{t=1}^{T} \prod_{i=1}^{n} f_0(x_i|\mu_t) \Big/ \sum_{t=1}^{T} \prod_{i=1}^{n} f_1(x_i|\mu_t)\,,$$

given that in this special case the *same* prior and the *same* Monte Carlo samples can be used.Figure 2 shows the convergence over $T = 10^5$iterations, along with the true value. The method exhibits convergence.

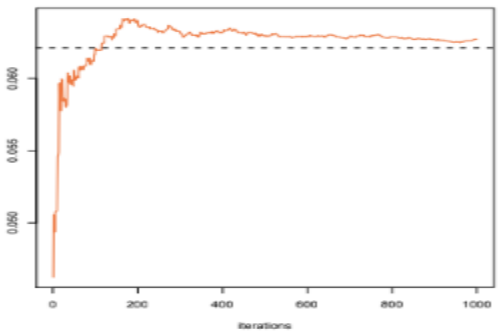

Fig 2. Convergence of a Monte Carlo approximation of B01(x1, . . . , xn) for a normalsample of size n = 19, along with the true value (dash line).

e-ISSN 2320 –7876 www.ijfans.org
Vol.11, Iss.9, Dec 2022

Research Paper

The above example can also be interpreted as an illustration of importance sampling, in the sense that the prior distribution is used as an importance function in both inte grals. We recall that importance sampling is a Monte Carlo method where the quantity of interest $E_f[h(X)]$ is expressed in terms of an expectation under the importance density g,

$$E_f[h(X)] = E_g[h(X)f(X)/g(X)],$$

which allows for the use of Monte Carlo samples distributed from g. Although importance sampling is at the source of the particle method (Doucet et al., 2001), I will not develop this useful sequential method any further, but in stead briefly introduce the notion of bridge
sampling (Meng and Wong, 1996) as it applies to the approximation of Bayes factors

$$\mathfrak{B}_{01}(x) = \int_{\Theta_0} f_0(x|\theta_0)\pi_1(\theta_0)\,\mathrm{d}\theta_0 \Big/ \int_{\Theta_1} f_1(x|\theta_1)\pi_1(\theta_1)\,\mathrm{d}\theta_1$$

(and to other ratios of integrals). This method handles the approximation of ratios of integrals over identical spaces (a severe constraint), by reweighting two samples from both posteriors, through a well-behaved type of harmonic average.

More specifically, when $\Theta_0 = \Theta_1$, possibly after a reparameterisation of both models to endow θ withthe same meaning, we have

$$\begin{aligned}\mathfrak{B}_{01}(x) &= \int_{\Theta_0} f_0(x|\theta)\pi_0(\theta)\alpha(\theta)\pi_1(\theta|x)\mathrm{d}\theta \Big/ \int_{\Theta_1} f_1(x|\theta)\pi_1(\theta)\alpha(\theta)\pi_0(\theta|x)\mathrm{d}\theta \\ &\approx \frac{n_1^{-1}\sum_{j=1}^{n_1} f_0(x|\theta_{1,j})\pi_0(\theta_{1,j})\alpha(\theta_{1,j})}{n_0^{-1}\sum_{j=1}^{n_0} f_1(x|\theta_{0,j})\pi_1(\theta_{0,j})\alpha(\theta_{0,j})}\end{aligned}$$

where $\theta_{0,1}, \ldots, \theta_{0,n0}$ and $\theta_{1,1}, \ldots, \theta_{1,n1}$ are two independent samples coming from the posterior distributions $\pi_0(\theta|x)$ and $\pi_1(\theta|x)$, respectively. (This identity holds for any function α guaranteeing the integrability of the products.) How ever, there exists a quasi-optimal solution, as provided by Gelman and Meng (1998):

$$\alpha^\star(\theta) \propto \frac{1}{n_0\pi_0(\theta|x) + n_1\pi_1(\theta|x)}\,.$$

While this optimum cannot be used—given that it relies on the normalising constants of both $\pi_0(\cdot|x)$ and $\pi_1(\cdot|x)$—, a practical implication of the result resorts to an iterative construction of $\alpha^?$. We gave in Chopin and Robert (2010) an alternative representation of the bridge factor that bypasses this difficulty (if difficulty there is!). While the number of roots is always p, the number of (non-conjugate) complex roots varies between 0 (meaning no complex root) and $b^p/2c$. This representation of the model thus induces a variable dimension structure in that the parameter space is then the product $(-1,$ distributions on $(-1,\ 1)$ and $B(0,\ 1)$, respectively. $1)^r \times B(0,\ 1)^{p-r/2}$, where $B(0,\ 1)$ denotes the complex unit ball and r is the number of real valued roots $\lambda_i B$. The prior distributions on the real and (non-conjugate) complex roots are the uniform.

REFERENCES

1. Andrieu, C., Doucet, A. and Holenstein, R. (2011). Particle Markov chain Monte Carlo (with discussion). *J. Royal Statist. Society Series B*, 72 (2) 269–342.
2. Beaumont, M. (2008). Joint determination of topology, divergence time and immigration in population trees. In *Simulations, Genetics and Human Prehis tory* (S. Matsumura, P. Forster and C. Renfrew, eds.). Cambridge: (McDon ald Institute Monographs), McDonald Institute for Archaeological Research, 134–154.
3. Beaumont, M. (2010). Approximate Bayesian computation in evolution and ecology. *Annual Review of Ecology, Evolution, and Systematics*, 41 379–406. Beaumont, M., Zhang, W. and Balding, D. (2002). Approximate Bayesian
4. computation in population genetics. *Genetics*, 162 2025–2035. Beskos, A., Papaspiliopoulos, O., Roberts, G. and Fearnhead, P. (2006). Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes (with discussion). *J. Royal Statist. Society Series B*, 68 333–382.
5. Blum, M. and Franc¸ois, O. (2010). Non-linear regression models for approx imate Bayesian computation. *Statist. Comput.*, 20 63–73.

6. Breslow, N. and Clayton, D. (1993). Approximate inference in generalized linear mixed models. *J. American Statist. Assoc.*, 88 9–25.
7. Brooks, S., Gelman, A., Jones, G. and Meng, X. (2011). *Handbook of Markov Chain Monte Carlo*. Taylor & Francis.
8. Casella, G. and George, E. (1992). An introduction to Gibbs sampling. *The American Statistician*, 46 167–174.
9. Chen, M., Shao, Q. and Ibrahim, J. (2000). *Monte Carlo Methods in Bayesian Computation*. Springer-Verlag, New York.
10. Chib, S. (1995). Marginal likelihood from the Gibbs output. *J. American Statist. Assoc.*, 90 1313–1321.
11. Chopin, N. and Robert, C. (2010). Properties of nested sampling. *Biometrika*, 97 741–755.
12. Congdon, P. (2006). Bayesian model choice based on Monte Carlo estimates of posterior model probabilities. *Comput. Stat. Data Analysis*, 50 346–357. Cornuet, J.-M., Santos, F., Beaumont, M., Robert, C., Marin, J.-M.,
13. Balding, D., Guillemaud, T. and Estoup, A. (2008). Inferring popula tion history with DIYABC: a user-friendly approach to Approximate Bayesian Computation. *Bioinformatics*, 24 2713–2719.
14. Del Moral, P., Doucet, A. and Jasra, A. (2006). Sequential Monte Carlo samplers. *J. Royal Statist. Society Series B*, 68 411–436.
15. Dickens, C. (1859). *A Tale of Two Cities*. London: Chapman & Hall. Doucet, A., de Freitas, N. and Gordon, N. (2001). *Sequential Monte Carlo Methods in Practice*. Springer-Verlag, New York.
16. Fearnhead, P. and Prangle, D. (2012). Semi-automatic approximate Bayesian computation. *J. Royal Statist. Society Series B*, 74 419–474. (With discussion.).
17. Gelman, A. and Meng, X. (1998). Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. *Statist. Science*, 13 163–185.
18. Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 6 721–741.
19. Gourieroux, C. ´ , Monfort, A. and Renault, E. (1993). Indirect inference. *J. Applied Econometrics*, 8 85–118.

20. Green, P. (1995). Reversible jump MCMC computation and Bayesian model determination. *Biometrika*, 82 711–732.
21. Hastings, W. (1970). Monte Carlo sampling methods using Markov chains and their application. *Biometrika*, 57 97–109.
22. Hjort, N., Holmes, C., Muller, P. ¨ and Walker, S. (2010). *Bayesian nonparametrics*. Cambridge University Press.
23. Hobert, J. and Casella, G. (1996). The effect of improper priors on Gibbs sampling in hierarchical linear models. *J. American Statist. Assoc.*, 91 1461– 1473.
24. Holmes, C., Denison, D., Mallick, B. and Smith, A. (2002). *Bayesian methods for nonlinear classification and regression*. John Wiley, New York. Jaakkola, T. and Jordan, M. (2000). Bayesian parameter estimation via variational methods. *Statistics and Computing*, 10 25–37.
25. Jeffreys, H. (1939). *Theory of Probability*. 1st ed. The Clarendon Press, Oxford.
26. Lauritzen, S. (1996). *Graphical Models*. Oxford University Press, Oxford. imsart-generic ver. 2009/02/27
27. Lee, K., Marin, J.-M., Mengersen, K. and Robert, C. (2009). Bayesian inference on mixtures of distributions. In *Perspectives in Mathematical Sci ences I: Probability and Statistics* (N. N. Sastry, M. Delampady and B. Rajeev, eds.). World Scientific, Singapore, 165–202.
28. Marin, J., Pillai, N., Robert, C. and Rousseau, J. (2011a). Relevant statistics for Bayesian model choice. Tech. Rep. arXiv:1111.4700. Marin, J., Pudlo, P., Robert, C. and Ryder, R. (2011b). Approximate Bayesian computational methods. *Statistics and Computing* 1–14. Marin, J. and Robert, C. (2007). *Bayesian Core*. Springer-Verlag, New York. Marin, J. and Robert, C. (2011). Importance sampling methods for Bayesian discrimination between embedded models. In *Frontiers of Statistical Decision Making and Bayesian Analysis* (M.-H. Chen, D. Dey, P. M¨uller, D. Sun and K. Ye, eds.). Springer-Verlag, New York, 000–000.