Informative Priors and Bayesian Computation

Shirin Golchi
University of
British Columbia, Statistics
Kelowna, BC V1V 1V7
Email: golchi.shirin@gmail.com

***Abstract*—The use of prior distribution is often a controversial topic in Bayesian inference. Informative priors are often avoided at all costs. However, when prior information is available informative priors are an appropriate way of introducing this information into the model. Furthermore, informative priors, when used properly and creatively, can provide solutions to computational issues and improve modeling efficiency. Through three examples in different applications we demonstrate the importance and usefulness of informative priors in incorporating external information into the model and overcoming computational difficulties.**

keywords: Aggregated relational data; Constraints; Differential Equation Models; Particle discovery; Sequential Monte Carlo; Uncertainty.

I. Introduction

In most scientific problems, in addition to the observed data, the researcher has some prior or external information about the unknowns, e.g., the range of plausible values. In classical statistics such information is incorporated into the model in form of (hard) constraints. However, hard constraints do not allow expressing uncertainty about the available information. Moreover, presence of constraints often results in inference and computational challenges and in many cases relaxation of the constraints or slight departure from available information facilitates statistical analysis.

In the Bayesian framework constraints are formulated into prior distributions. In [1] it is argued that the common Bayesian approach for incorporating hard constraints is not equivalent to imposing the constraint in frequentist approaches. They consider the constraint as a set that the parameter is believed to belong to and argue that defining an even flat prior on the constraint set is a much stronger assumption than the frequentist alternative that assumes no more than mere membership of the set since a flat prior assumes that all values in the constraint set are equally likely. In [2] similar arguments are made about imposing hard constraints using "non-informative" priors that can result in misleading the posterior. They consider inference about a continuous monotonic function based on discretized observations and explain mathematically why imposing the monotonicity constraints strictly can result in biased estimates.

Bayesian inference is often criticized for its reliance on prior distributions [3]. We emphasize the role of prior distributions in presence of modeling constraints and external information. In this paper, we consider scenarios where hard constraints need to be replaced by soft constraints due to either prior uncertainty or methodological/computational challenges created by the constraints. We showcase the use of informative priors in a Bayesian framework to formally introduce uncertainty about the prior knowledge or temporarily departing from the constraints for computational purposes through three different examples.

Our first example (Section II) is a Bayesian hierarchical model to analyze aggregated relational data proposed by [4]. External information about some of the model parameters is required to identify the model. We show that informative priors can be used to realistically incorporate external information about some of the model parameters to overcome nonidentifiability, thereby addressing efficient computation, accessibility of the models to the general user and interpretability of the results.

In our second example (Section III) we consider Bayesian analysis of the data used to discover the Higgs boson. We review the Bayesian hierarchical model proposed by [5] and illustrate the role of the informative prior used in their model in providing the statistical power to detect the signal by comparing the results to those obtained by using an "objective" prior.

In the third example (Section IV) we consider inference for a system of ordinary differential equations (ODE) where adherence to the ODE is a hard constraint that creates computational difficulties. By introducing a prior distribution that allows departure from the ODE model one can overcome computational challenges while leaving the results unaffected by relaxation of the con-

straints. Section V follows with concluding remarks.

II. Analysis of aggregated relational data

Aggregated relational data (ARD) in social surveys are collected by asking the respondents about the number of their acquaintances in various groups that are chosen either due to their significance to the researcher or as auxiliary groups with known demographic information (male and female names). While most ARD has been collected to infer social structure they are perceived as a summary of full network data (nodes and edges) and therefore can arise in other areas of science where networks appear. Full network data is transformed into ARD by aggregating the links of a set of sample nodes over specific communities and ignoring the community membership of the sample nodes. While such aggregation results in loss of information it appears necessary in various contexts for preservation of privacy.

Suppose that y_{ik} is the number of acquaintances of individual i in group k. The goal is to recover as much information as possible about the underlying network from these counts. [4] propose the following model

$$y_{ik} \sim Poisson(\lambda_{ik}) \tag{1}$$

with $\lambda_{ik} = \gamma_{ik} \exp(\alpha_i + \beta_k)$ where α_i is the degree or network size of respondent i, β_k is the gregariousness of group k, and γ_{ik} is the individual propensity parameter that represents the tendency of respondent i to make ties with group k. Due to computational limitations, [4] were not able to fit this highly parametrized model. They integrate over the propensity parameters with respect to a gamma prior distribution with mean 1 and variance, $var(\gamma_{ik}) = \frac{\omega_k - 1}{\exp(\alpha_i + \beta_k)}$, where ω_k represents the overdispersion associated with group k. The prior variance of the propensity parameters is assumed to be smaller if the corresponding network sizes and gregariousness parameters are larger. [4] integrate the γ_{ik} parameters out of the likelihood with respect to this prior distribution that results in the following negative binomial likelihood;

$$\begin{aligned} y_{ik} \sim \mathcal{NB}(&mean = \exp(\alpha_i + \beta_k), \\ &var = \omega_k \exp(\alpha_i + \beta_k)), \end{aligned} \tag{2}$$

This simplification reduces the number of parameters by $K(I-1)$. We refer to the above negative binomial likelihood as the overdispersed model. However, even after this simplification the model is non-identifiable since the likelihood depends on α_i and β_k through their sum. Therefore the value of the likelihood remains unchanged for infinite number of values of α and β. [4] address non-identifiability by renormalizing the model parameters according to known demographics of certain groups at each iteration of their Gibbs sampling algorithm.

The demographic information that is available about a subset of groups is the key component of this model since without this information the renormalization step in [4] is entirely arbitrary that could result in unreliable parameter estimates. We argue that in a Bayesian framework any additional information should be used formally, i.e., in form of a prior distribution that would allow incorporation of uncertainty.

In the following we explain the choice of informative priors that allow us to incorporate this information into the model to overcome non-identifiability. The resulting model is fit using Hamiltonian Monte Carlo and is implemented in Stan (http://mc-stan.org/). The same strategy can be used to gain computational efficiency and improvements in the inference for related models in [6], [7], [8].

Consider the original Poisson model in (1). The main advantage of this model is interpretability of the model parameters: the individual propensity parameters can be used as a relative distance metric between the respondents and the subpopulations. In addition to the data the subpopulation sizes for the male and female names in the US population are available. Following [4] we use normal prior distributions for the log group size parameters β_k. However, we incorporate the additional information as the mean of the normal priors for the corresponding β_k. Small variances are chosen for these normal priors that allow slight deviations from the mean allowing for introducing uncertainty. For the subset of β_k that such information is not available we use diffuse normal priors centered at the mean of the known means:

$$\beta_k \sim \mathcal{N}(\log(\frac{N_k}{N}), \sigma_\beta^2), k = 1, \ldots, K_1, \tag{3}$$

$$\beta_k \sim \mathcal{N}(\frac{1}{K_1} \sum_{k=1}^{K_1} \log(\frac{N_k}{N}), \tau_\beta^2), k = K_1 + 1, \ldots, K. \tag{4}$$

The prior distribution for α_i is,

$$\alpha_i \sim \mathcal{N}(\mu_\alpha, \sigma_\alpha^2), \tag{5}$$

where μ_α and σ_α^2 are assigned non-informative priors.

For the propensity parameters γ_{ik} we define a prior that assumes that individuals are equally likely to form ties with all groups ($E(\gamma_{ik} = 1)$) with enough flexibility that would let the data decide otherwise. The following gamma distribution represents this assumption

$$\gamma_{ik} \sim \Gamma(a_\gamma, b_\gamma) \tag{6}$$

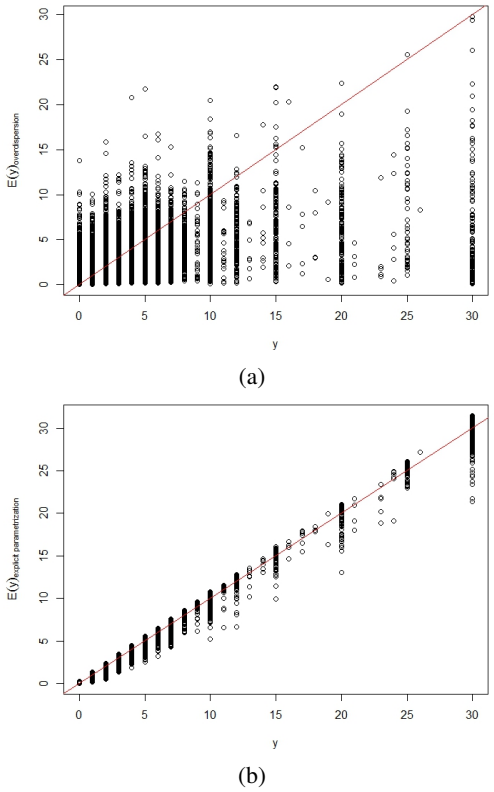


Fig. 1: Posterior predictive checks for (a) the overdispersed model and (b) the proposed model.

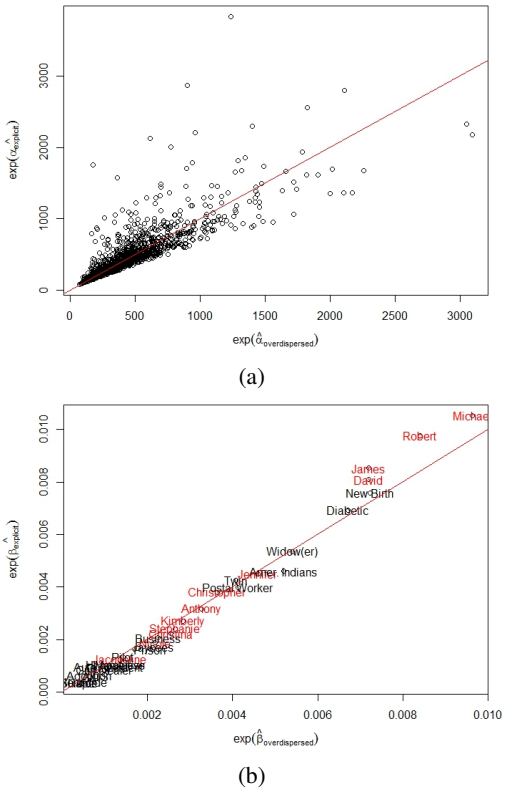


Fig. 2: Comparison of the estimates for (a) degrees and (b) group sizes obtained by the overdispersed model and the explicit parametrization of individual propensities - the groups in red are those with informative priors.

if $a_\gamma = b_\gamma$ are chosen such that the prior variance is relatively large. We choose, $a_\gamma = b_\gamma = 0.25$ that is equivalent to a variance of 4.

Having resolved non-identifiability, we are able to implement the above model in Stan, taking advantage of the efficiency of the Hamiltonian Monte Carlo sampler. The Stan implementation of the model is given in Appendix A. The Stan implementation removes the computational obstacles in making the model accessible to the users.

We fit the two models to the data collected by [9] that was also used in [4]. Figures 1b and 1a present the plots of data generated from the posterior predictive distribution against the observed data for the overdispersed model and the proposed model, respectively. The proposed model clearly provides a better fit to the data.

The estimated values of the shared parameters between the two models, i.e., α_i and β_k are plotted in Figure 2. For the sake of interpretability we plot $\exp(\alpha_i)$, the network sizes or degrees and $\exp(\beta_k)$, the proportion of subpopulations for the two models.

In Figure 2a the degree estimates are for the most part in agreement under the two models. For the group proportions (Figure 2b), the estimates are aligned except for a few larger subpopulations (popular male names). This discrepancy is mostly due to the recall bias, i.e., the fact that respondents often under/over-report the number of their acquaintances in larger/smaller groups. In the original overdispersed model in [4] this bias is ignored. A correction function is proposed for the overdispersed model in [6]. The full model on the other hand corrects for the recall bias implicitly by allowing individuals to vary in their propensity to make ties with different groups.

III. A Bayesian model for particle detection

In this section we consider a Bayesian analysis of the data generated by particle detctors for the purpose of discovering the Higgs particle. The Bayesian model proposed by [5] relies on information provided by the

theory and Monte Carlo studies that are independent of the data being analyzed. This information is incorporated through informative priors. In the following we review the model and emphasize the role of the informative prior in a comparison with a more "objective" prior.

The Standard Model (SM) of particle physics describes the dynamics of subatomic particles. The Higgs particle is an essential component of the SM ([10], [11], [12], [13]). The existence of the Higgs boson needs to be confirmed by experiments run at the Large Hadron Collider (LHC) at the European organization for nuclear research, known as CERN, a high energy collider that is specifically designed and constructed to detect the Higgs particle.

A simplified description of the experiment given by [5] is as follows. Two beams of protons circulating at very high speeds in the LHC collide inside two detectors (ATLAS and CMS). Collisions or "events" result in generation of new particles, possibly including the Higgs boson. The Higgs particle cannot be detected directly since, if generated, decays extremely quickly into other known SM particles. Therefore, the existence of the Higgs particle is inferred by the combinations of detectable particles predicted by the SM. Once a Higgs particle is created there are several different "decay modes" through which it may decay. The decay process can be reconstructed based on the detected collision byproducts. Events with reconstructed processes that match one of the possible Higgs decay modes are recorded as "Higgs candidates" and the invariant mass of the unobserved particle is computed from the reconstruction. A histogram of the estimator of the mass is then created. However, there are other processes, not involving the Higgs boson, that can result in the generation of Higgs event byproducts which also pass the cuts; these are called "background events". Therefore, the histogram created is either a histogram of background-only events if the Higgs particle does not exist or a histogram of background-plus-signal events, otherwise.

Figure 3 shows a typical background-plus-signal data that is simulated by computer models that simulate the behavior of particle detectors. The slight overflow of events about $m = 125$ shows the level of evidence that generally appears in such data. A glance at these data suggests that without any prior knowledge about the distribution of the background events little statistical power is available in signal detection.

In the current practice the background function is described as a smooth forth order polynomial whose parameters are estimated using background-only data

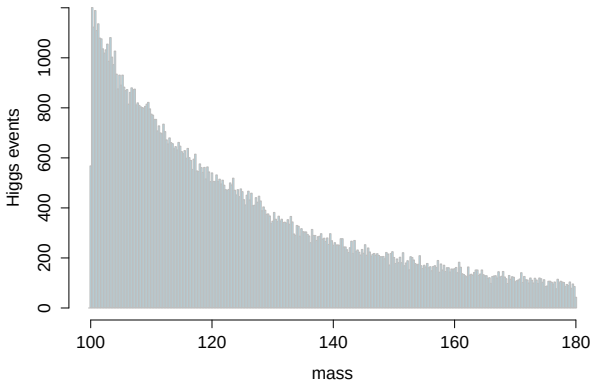


Fig. 3: Simulated data representing the invariant masses of events.

that are simulated using Monte Carlo. The estimated background model is used in the analysis. Using plug-in estimates in this manner would not allow incorporation of uncertainty about the background model.

[5] proposed a Bayesian hierarchical model for the analysis of the output of the particle detector. They incorporate the additional information about the background function through a Gaussian process prior centered at the polynomial model used in the current practice. The Gaussian process allows deviations from this mean function.

In the following we review the model given by [5] and compare it with a variation of the model with an "objective" prior over the background function. In doing so we make the point that an appropriate informative prior is a formal and statistically justifiable tool for incorporating external knowledge to the model that can be crucial in detecting the signal.

The data presented in Figure 3 are modeled as realizations of a Poisson process whose intensity function is given by the sum of a background process $\Lambda(m)$ and a signal function $s_{m_H}(m)$. The shape of the signal function is given by the theory and its location is determined by the parameter of interest, i.e., the unknown mass of the Higgs particle $m_H \in \mathcal{M}$, where $\mathcal{M} = \{\emptyset\} \cup (m_0, m_n)$ $((m_0, m_n) \subset \mathcal{R}^+ - \{0\})$. $m_H \in (m_0, m_n)$ means that the Higgs boson has a mass in the search window, (m_0, m_n), while $m_H = \emptyset$ refers to the case that the particle does not exist, at least not with a mass in (m_0, m_n). The uncertainty about the background

$\Lambda(m)$ is modeled by a log-Gaussian process,

$$\log \Lambda_{\eta,\sigma^2}(m) \sim \mathcal{GP}(\xi(m), \rho_{\eta,\sigma_2}(m, m')), \quad m \in (m_0, m_n). \quad (7)$$

with mean function, $\xi(m)$, and covariance function given by,

$$\rho_{\eta,\sigma_2}(m, m') = \sigma^2 \exp(-\eta(m - m')^2), \quad (8)$$

where σ^2 is the variance parameter and η is the correlation parameter that controls the smoothness of the background function.

The signal function is chosen as a Gaussian probability density function with the location parameter m_H,

$$s_{m_H}(m) = c_{m_H}\, \phi(\frac{m - m_H}{\epsilon}) \qquad m_H \in (m_0, m_n),$$
$$s_\emptyset(m) = 0, \quad (9)$$

where c_{m_H} is a scaling constant, and ϕ is the normal probability density function with standard deviation ϵ.

The likelihood is given by,

$$\pi(\mathbf{y}|\Lambda, m_H) = \prod_{i=1}^{n} \frac{\exp(-\Gamma_i)\Gamma_i^{y_i}}{y_i!}, \quad (10)$$

where

$$\Gamma_i = \int_{m_{i-1}}^{m_i} [\Lambda(m) + s_{m_H}(m)]dm. \quad (11)$$

The grid $\mathbf{m} = (m_0, m_1, \ldots, m_n)$ is the vector of bin boundaries over the search window.

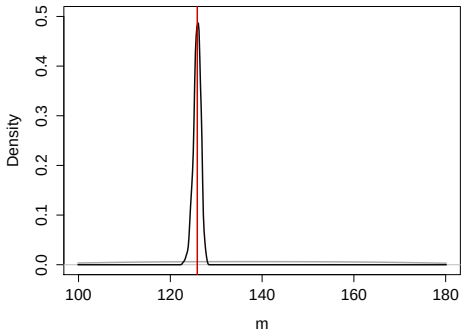
The posterior distribution of the model parameters $\theta = (\eta, \sigma^2, \Lambda, m_H)$ given the data $\mathbf{y}$ is as follows,

$$\pi(\theta \mid \mathbf{y}) = \frac{\pi(\theta)\pi(\mathbf{y}|\theta)}{\int \pi(\theta)\pi(\mathbf{y}|\theta)d\theta}, \quad (12)$$

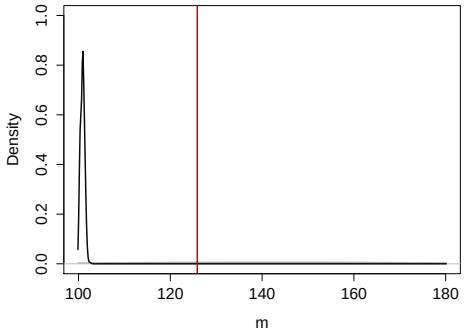
with independent prior, $\pi(\theta) = \pi(\eta)\pi(\sigma^2)\pi(m_H)\pi(\Lambda)$. The prior distribution, $\pi(m_H)$ is a mixture of a point mass at $m_H = \emptyset$ and a continuous distribution on (m_0, m_n). The hyper-priors on η and σ^2 are inverse Gamma distributions with shape and scale parameters equal to one.

[5] suggest that the prior mean of the background $\xi(m)$ should be determined from Monte Carlo studies that are used in the current practice [11] to obtain the functional form of the background model and the uncertainties associated with it. In absence of such information [5] replace $\xi(m)$ with a polynomial fit to the simulated data provided by the CMS group and described in [14] and [15].

In addition to this choice of prior we fit the same model with a constant mean log-Gaussian prior on the background. We use the sequential Monte Carlo algorithm outlined in Algorithm 1 in [5] to sample from the posterior distribution (12) with the two different choices of prior mean for the background. Figures 4a and 4b show kernel density estimates of marginal posterior distribution of the mass over the search window with the informative and "non-informative" priors over the background. The estimated posterior probability of $m_H = \emptyset$ is zero under the model with informative background prior suggesting discovery while it is .05 under the objective background prior. Clearly, without using the information available about the distribution of background events the signal is not detectable as seen in Figure 4b where the background fluctuations for small masses are mistaken for signal.



(a)



(b)

Fig. 4: Kernel density estimates of the posterior distribution of the mass of the Higgs particle with (a) informative prior and (b) objective prior on the background function. The vertical red lines are drawn at the reported mass of the Higgs particle

IV. Inference for ordinary differential equations

Consider the problem of estimating the parameters of a system of ordinary differential equations (ODE) from noisy observations of the system. In most real problems the ODE cannot be solved analytically and evaluation of the likelihood relies on numerically solving the ODE that in turn relies on initial state values. Inference for these type of problems is challenging due to sensitivity of the likelihood to small changes in the parameters. In Bayesian inference the challenges are translated into highly peaked and often multimodal posterior surfaces that are very difficult to explore using MCMC [16], [17].

[18] address Bayesian inference for ODEs by relaxation of the assumption of adherence ot the ODE. The data is modeled as the summation of the ODE solution and a discrepancy term that facilitates computation by fitting the deviation of the data from the ODE solution for "incorrect" parameter values. [18] propose a sequential Monte Carlo based algorithm that reduces the role of the discrepancy term at each step until it completely diminishes from the model.

The discrepancy term used by [18] is a kernel smoother. In this section we propose replacing the kernel smoother with a zero mean Gaussian process (GP) that essentially performs similarly in relaxation of adherence to the ODE but is also interpretable as a probability distribution around the ODE solution. In other words, Bayesian computation is facilitated by a prior distribution about the ODE that is made more informative at each step converging to a point mass at the ODE solution.

The example used in [18] is a Susceptible-Infected-Removed (SIR) epidemiological model for the deaths occurred during the black plague epidemic of 1666, in the village of Eyam, UK, that was quarantined. The grave digger kept records of the deaths. The data refers to the second outbreak of the plague from June 19, 1666 to November 1, 1666, with a population of fixed size $N = 261$. At time ν the population is split into groups of Susceptible $S(\nu)$, Infected, $I(\nu)$, and Removed, $R(\nu)$ [19], [20]. Since there is no recovery from the plague, $R(\nu)$ is the number of deaths up to time ν. The epidemiological model for the rates of change of states $S(\nu), I(\nu)$ and $R(\nu)$ is given by:

$$\begin{aligned} \frac{dS(\nu)}{dt} &= -\beta S(\nu) I(\nu), \\ \frac{dI(\nu)}{dt} &= \beta S(\nu) I(\nu) - \alpha I(\nu), \\ \frac{dR(\nu)}{dt} &= \alpha I(\nu) \end{aligned} \tag{13}$$

where β describes the plague transmissivity and α describes the rate of death once an individual is infected.

At time 0 the population only consists of susceptible and infectious individuals therefore we have, $R(0) = 0$ and $S(0) = N - I(0)$. Consequently $I_0 = I(0)$ is included in the vector of parameters to estimate: $\theta = (\alpha, \beta, I_0)$.

Based on a finite population of size N, the likelihood for the $n = 136$ observed cumulative deaths $y(\nu)$ at times $\{\nu_1, \ldots, \nu_n\}$ are modeled as a binomial with expected value equal to the solution to $R_\theta(\nu)$ from (13):

$$P(\mathbf{y} \mid R_\theta(\nu)) = \prod_{i=1}^{n} \binom{N}{y_i} \left(\frac{R_\theta(\nu_i)}{N}\right)^{y_i} \times \left(1 - \frac{R_\theta(\nu_i)}{N}\right)^{(N-y_i)}.$$

Similarly to the approach by [18] we replace $R_\theta(\nu)$ in the likelihood by $R_\theta(\nu)$ plus a discrepancy term. However, we use a zero mean GP instead of the kernel smoother used by [18]. Every evaluation of the likelihood requires solving (13) numerically to obtain $R(\theta, \nu)$ and fitting a zero mean GP to the residuals, $y(\nu) - R(\theta, \nu)$. More explicitly, for each set of parameter values θ a sample path is given by

$$\hat{R}_{\theta,\xi}(\nu) = R_\theta(\nu) + z_\xi(\nu), \tag{14}$$

where $z_\xi(\nu)$ is a zero mean GP with a covariance function parametrized by $\xi = (\eta, \sigma^2, \tau^2)$,

$$cov(z_\xi(\nu), z_\xi(\nu')) = \begin{cases} \sigma^2 \exp(-\eta(\nu - \nu')^2) & \text{if } \nu \neq \nu' \\ \sigma^2 + \tau^2 & \text{if } \nu = \nu' \end{cases}. \tag{15}$$

The GP is then interpreted as a prior distribution about the ODE solution that can be tuned to allow relaxation of adherence to the ODE model. The sequentially constrained Monte Carlo (SCMC) algorithm of [18] can be used to sample from the target posterior distribution $\pi_T(\boldsymbol{\theta} \mid \mathbf{y})$. The intermediate densities are defined through covariance parameter values of the GP,

$$\pi_t\left(\boldsymbol{\theta}, \hat{R} \mid \boldsymbol{\theta}\right) \propto P\left(\mathbf{y} | \hat{R}_{\boldsymbol{\theta}}\right) \pi\left(\hat{R}_{\boldsymbol{\theta},\boldsymbol{\xi_t}} \mid R_{\boldsymbol{\theta}}\right) \pi(\boldsymbol{\theta}), \tag{16}$$

where $\pi\left(\hat{R}_{\boldsymbol{\theta},\boldsymbol{\xi_t}} \mid R_{\boldsymbol{\theta}}\right)$ represents the GP prior in the posterior. As t increases the parameters ξ_t are adjusted to make the prior more focused about the ODE solution by making the GP sample paths smoother (decreasing η_t) and the variance smaller (decreasing σ_2). Eventually, the GP prior converges to a point mass at the ODE solution, i.e., $\pi\left(\hat{R}_{\boldsymbol{\theta},\boldsymbol{\xi_t}} \mid R_{\boldsymbol{\theta}}\right) = 1$.

Following [18] the prior on α and β is gamma $(1, 1)$ and $I_0 \sim binomial(N, 5/N)$. Figure 5 shows the final sample of the joint posterior of α and β in form of separate clouds that correspond to initial values I_0 that closely resembles the results of [18].

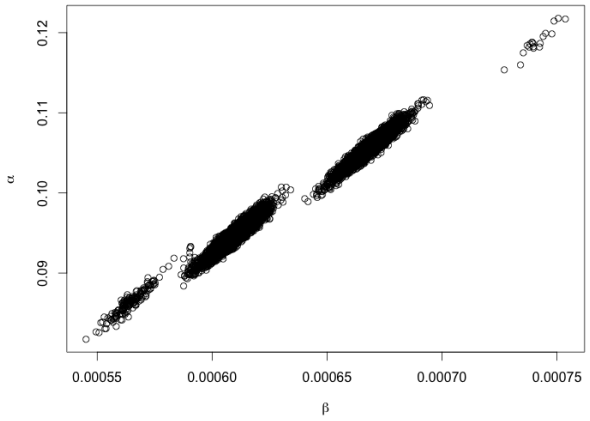


Fig. 5: Samples from the joint posterior distribution of the model parameters. The three large clouds of particles correspond to $I_0 = 6$, $I_0 = 5$ and $I_0 = 4$, respectively, from left to right.

V. Conclusion

In this article we have demonstrated the use of informative priors in Bayesian inference and computation through three examples in different contexts. In the first example the choice of informative priors play a crucial role from both inference and computational point of views – external information that is required to identify the model are incorporated in a formal and interpretable manner resulting in a hierarchical model that can be implemented into Stan and made accessible to users.

In our second example, the informative prior is used to incorporate external information provided by theory and experiments together with appropriate level of uncertainty. Discovery of new physics would not have been possible without proper usage of all the information. We demonstrate the importance of an informative prior by comparing the results to the case that an objective prior is used.

In the third example, we use an informative prior from a purely computational perspective. In fact, the prior is completely eliminated from the model in the final step of the sampling algorithm. The role of the prior is to provide means of relaxation of model assumptions that create challenges in Bayesian computation. By reformulating the model adherence assumptions into a probability distribution posterior sampling is facilitated. However, inference results are not affected since the additional uncertainty is sequentially eliminated from the model.

Appendix A
Stan implementation for the model proposed in Section II

```
data {
 int<lower=0> I;
 int<lower=0> K;
 vector[K] mu_beta;
 vector<lower=0>[K] sigma_beta;
 int  y[I,K];
}
parameters {
 vector[I] alpha;
 vector[K] beta;
 matrix<lower=0>[I,K] gam;
 real mu_alpha;
 real<lower=0> sigma_alpha;
}
model {
 alpha ~ normal(mu_alpha, sigma_alpha);
 beta ~ normal(mu_beta, sigma_beta);
 for (i in 1:I)
  for (k in 1:K)
   gam[i,k] ~ gamma(.25, .25);
 mu_alpha ~ normal(0 , 25);
 sigma_alpha ~ normal(0 , 5);
 for (k in 1:K) {
  for (i in 1:I) {
   real xi_i_k;
   xi_i_k <-
    gam[i,k] * exp(alpha[i] + beta[k]);
   y[i,k] ~ poisson(xi_i_k);
  }
 }
}
```


Acknowledgment

This research is partially supported by NSF grant SES-1023176 and the Natural Sciences and Engineering Research Council of Canada (NSERC). The author would

like to thank Dr Andrew Gelman and Dr Tian Zheng for their support and invaluable comments.

REFERENCES

[1] P.B. Stark, (2015). *Constraints versus Priors*, SIAM/ASA J. Uncertainty Quantification, 3(1), pp. 586–598.

[2] A. Gelman, (1996). *Bayesian model-building by pure thought: some principles and examples*, Stat. Sinica, 6(1), pp. 215–232.

[3] A. Gelman, (2008). *Objections to Bayesian Statistics*, Bayesian Analysis, 3 (3), pp. 445–450.

[4] T. Zheng, M.J. Salganik, A. Gelman, (2006). *How many people do you know in prison?: Using Overdispersion in Count Data to Estimate Social Structure in Networks*, Journal of the American Statistical Association, 101, pp. 409–423.

[5] S. Golchi, R. Lockhart (2016). *A Bayesian Search for the Higgs Particle*, arXiv:1501.02226 [stat.AP].

[6] T.H. Mccormick, T. Zheng (2007). *Adjusting for Recall Bias in How many Xs do you know? Surveys*, in Proceedings of the Joint Statistical Meetings.

[7] T.H. McCormick, M.J. Salganik, T. Zheng, (2010) *How many people do you know?: Efficiently estimating personal network size*, Journal of the American Statistical Association, 105 (489), 59–70.

[8] T.H. Mccormick, T. Zheng (2012). *Latent demographic profile estimation in hard-to-reach groups*, The Annals of Applied Statistics, 6 (4), 1795–1813.

[9] C. McCarty, P.D. Killworth, H.R. Bernard, E.C. Johnsen, G.A. Shelley (2001). *Comparing Two Methods for Estimating Network Size*, Human Organization, 60, pp. 28–39.

[10] ATLAS Collaboration (2012). *Observation of a new particle in the search for the standard model Higgs boson with the ATLAS detector at the LHC*, Physics Letters B, 716(1), pp. 1–29.

[11] ATLAS Collaboration (2014). *Measurement of Higgs boson production in the diphoton decay channel in pp collisions at center-of-mass energies of 7 and 8 TeV with the ATLAS detector*, Phys. Rev. D, 90:112015.

[12] CDF Collaboration (2011). *Search for high mass resonances decaying to muon pairs in root s = 1.96 TeV pp collisions.*, Phys. Rev. Lett., 106:121801.

[13] CMS Collaboration (2012). *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Physics Letters B, 716(1), pp. 30–61.

[14] CMS Collaboration (2014). *Observation of the diphoton decay of the Higgs boson and measurement of its properties*, Eur. Phys. J. C, 74:3076.

[15] CMS Collaboration (2013). *Updated measurements of the Higgs boson at 125 GeV in the two photon decay channel*, CERN, Geneva, CMS-PAS-HIG-13-001.

[16] D.A. Campbell, R.J. Steele,(2009). *Smooth functional tempering for nonlinear differential equation models*, Statistics and Computing, 22(2), pp. 429–443.

[17] B. Calderhead, M. Girolami (2009). *Estimating Bayes Factors via Thermodynamic Integration and Population MCMC*, Computational Statistics and Data Analysis, 53, pp. 4028–4045.

[18] S. Golchi, D. Campbell (2016). *Sequentially Constrained Monte Carlo*, Computational Statistics and Data Analysis, 97, pp. 98–113.

[19] G.F. Raggett, (1982). *Modelling the Eyam plague*, Institue of Mathematics and its Applications, 18, pp. 221–226.

[20] D.A. Campbell, S. Lele,(2014). *An ANOVA test for parameter estimability using data cloning with application to statistical inference for dynamic systems*, Computational Statistics and Data Analysis, 70, pp. 257–267.