Bayesian Registration using Hamiltonian Monte Carlo.

Henrique Cheng [1] & John Fricks [2]

[1] *School of Mathematical and Statistical Sciences, Arizona State University, USA hcheng43@asu.edu*
[2] *School of Mathematical and Statistical Sciences, Arizona State University, USA & SISTM team, Bordeaux Public Health, Université de Bordeaux, France jfricks@asu.edu*

Résumé. Dans cette présentation, une nouvelle méthode d'enregistrement de courbe dans le contexte de l'estimation de la fonction moyenne est présenté. Un schéma numérique Hamiltonien de Monte Carlo est utilisé pour échantillonner la distribution a posteriori des fonctions de déformation étant donné les données fonctionnelles. Une comparaison numérique avec un algorithme de Metropolis-Hastings plus traditionnel est présentée. La conclusion préliminaire de l'étude numŕique est que l'approche informatique Hamiltonienne de Monte Carlo accélère considérablement la convergence vers le postérieur de haute dimension requis pour l'estimation de la fonction de courbe.

Mots-clés. Données fonctionnelles, Enregistrement de courbe, Statistique bayésienne.

Abstract. In this presentation, a novel method for curve registration in the context of mean function estimation is presented. A Hamiltonian Monte Carlo numerical scheme is used to sample for the posterior distribution of the warping functions given the functional data. A numerical comparison with a more traditional Metropolis-Hastings algorithm is presented. The preliminary conclusion of the numerical study is that the Hamiltonian Monte Carlo computational approach substantially speeds convergence to the high dimensional posterior that is required for curve function estimation.

Keywords. Functional Data Analysis, Curve Registration, Bayesian Statistics.

1 Introduction

Registration is an important topic of functional data analysis. Intuitively, registration attempts to remove phase variation in a sample of observed functions. For an extended discussion on registration see Ramsay and Silverman (2019) or Marron et al (2015). One goal of registration can be to infer the mean function of such a group of functions in the presence of phase variation. We will largely be following Earls and Hookers approach to simultaneously register and estimate a mean function (2017).

We write a more specific statistical model as follows.

$$Y_j\left(t_{i,j}\right) = \mu\left(h_j\left(t_{i,j}\right)\right) + X_j\left(h_j\left(t_{i,j}\right)\right) + \epsilon_{i,j} \tag{1}$$

for time index $i = 1, ..., I_j$, subject index $j = 1, ..., J$, and with $0 \leq t_{i,j} \leq T$. The goal for the purposes of this study, then, will be to estimate $\mu(\cdot)$.

The function $h_j(\cdot)$ is a subject-specific increasing time warping function with $h_j(0) = 0$ and which will be represented here as

$$h_j(t) = \int_0^t e^{w_j(s)} ds \tag{2}$$

with $w_j(\cdot)$ modeled as a zero-mean Gaussian, Matérn process.

2 Simulation Setup and Conclusion

A Bayesian registration method was implemented using a Hamiltonian Monte Carlo method for sampling from the posterior distribution of $h_j|Y_j$ following Betancourt (2017). For testing purposes, a mean model consisting of a finite combination of Fourier components with Gaussian amplitude error and a time-warping function described above. The full hierarchical Bayes model is given here:

$$\begin{aligned}
Y_j(h_j(t))|f(t|\boldsymbol{\beta}, \tau) &\overset{iid}{\sim} GP(\mu(t), \Sigma_f(t, u|\sigma_f^2)); \quad t, u \in \mathcal{T}, \quad j = 1, ..., J \\
h_j(t)|w_j(t) \qquad w_j(t) &\overset{iid}{\sim} GP(0, \Sigma_w(t, u|\boldsymbol{\gamma})) \\
\boldsymbol{\gamma} &= (\gamma_1, \gamma_2)' \\
\mu(t|\boldsymbol{\beta}, \tau) &= \beta_0 + \sum_{q=1}^{Q} \beta_{1q} \cos(\frac{2\pi q}{\tau} t) + \beta_{2q} \sin(\frac{2\pi q}{\tau} t) \\
\boldsymbol{\beta} &\sim \mathrm{N}_{2Q+1}(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}}) \\
\tau &\sim N(\mu_\tau, \sigma_\tau^2) \\
\sigma_f^2 &\sim IG(a/2, 2/b) \\
\boldsymbol{\gamma} &\sim N_2(\boldsymbol{\mu}_{\boldsymbol{\gamma}}, \Sigma_{\boldsymbol{\gamma}})
\end{aligned}$$

For comparison, a Metropolis-Hastings within Gibbs Monte Carlo method for sampling the posterior distribution was also developed. Here we present some preliminary simulation results. In Figure 1, we see three realizations simulated from the model in Equation 1 with

$$\mu(t) = -1 \cos(\frac{2\pi}{12} t) + 2 \cos(\frac{2\pi}{6} t) - 2 \cos(\frac{2\pi}{4} t)$$

and $\sigma_\epsilon^2 = 0.55^2$, as well as with the warping function defined by Equation 2 with $w(\cdot)$ being a Matérn process with $(\sigma_w^2, \rho, \nu)' = (0.2, 1, 3)$.

These three simulations were used as data and the parameters of the Bayesian hierarchical model presented above were estimated by using the Hamiltonian Monte Carlo method to draw from the posterior distributions. A more traditional Metropolis within Gibbs was also used for comparison. Posterior means can be found in Table 1.

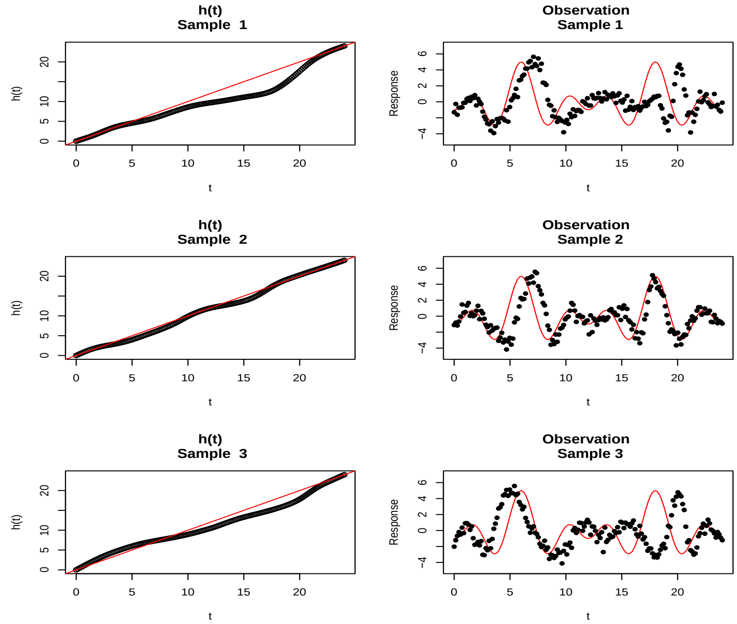

Figure 1: Three simulated warping functions whose base values are independently sampled from a MVN distribution with Matérn covariance function parameters $(\sigma_w^2, \rho, \nu)' = (0.2, 1, 3)$ on the LHS. On the RHS, each respective warping function is applied to 170 points along the reference function $-1\cos(\frac{2\pi}{12}t) + 2\cos(\frac{2\pi}{6}t) - 2\cos(\frac{2\pi}{4}t)$ (red line) and iid $N(0, .55^2)$ amplitude noise is added to generate observations (black points).

S1	τ	$\boldsymbol{\beta}$	σ_f^2	$\exp(\boldsymbol{\gamma})$
True	12	$(0, -1, 0, 2, 0, -2, 0)$	0.55^2	$(0.2, 1)$
MH	12.69	$(-0.09, -0.57, 0.07, 0.94, 0.23, -0.51, -0.17)$	1.87^2	$(0.43, 0.72)$
HMC	12.08	$(-0.05, -0.93, -0.03, 1.98, -.01, -1.93, .01)$	0.53^2	$(0.28, 0.91)$

Table 1: A comparison of true parameter values and point estimates under the strict Metropolis within Gibbs and HMC adapted sampler.

In addition to the point estimates, the posterior warping functions can be sampled. In Figure 2, the simulated as well as the posterior warping function for both the HMC and Metropolis within Gibbs samplers for each of the three replicates. Note that the posterior means of the HMC are a substantially closer fit to the true warping function compared to the Metropolis within Gibbs sampler for comparable computational cost.

To see the effect of registration on the fit of the reference function, Figure 3 shows simulated data along with the registered reference functions. The last column shows the comparison using the HMC algorithm with the Metropolis within Gibbs shown in column 2. The plots show substantial improvement of the fit using the HMC algorithm with comparable computing cost, bolstered by the root mean squared error of roughly 0.5 for HMC with root means squared error for the Metropolis within Gibbs close to 2.

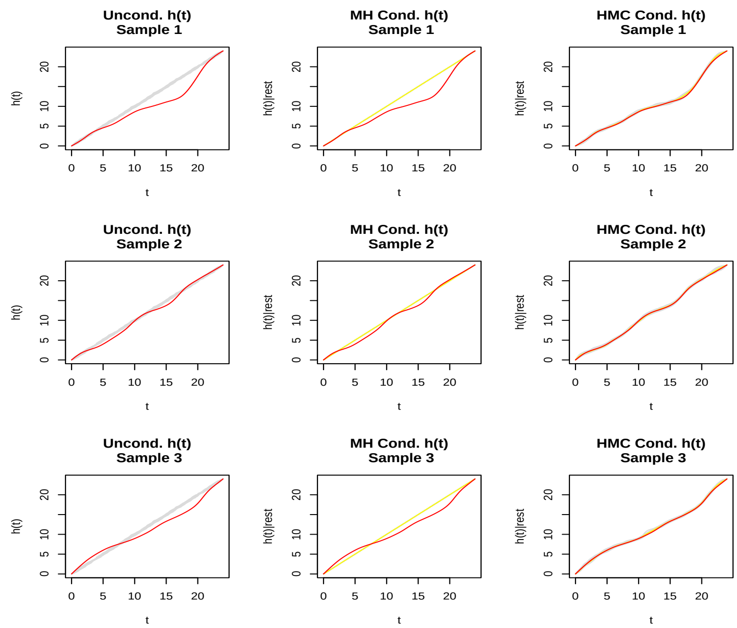

Figure 2: Linearly interpolated draws from the unconditional (first column) and conditional (second and third columns) warping functions for each sample in the study. The true warping function is in red for each plot, the yellow lines in the second column are the estimated warping functions under the strict Metropolis within Gibbs sampler for each sample, and the yellow lines in the third column are likewise but for the HMC adapted sampler. In the second column, each 8000^{th} draw was taken from the posterior distribution for each warping function and in the third column, each 2400^{th} draw was taken.

The primary conclusion from the numerical results is that similar results can be obtained from the two methods; however, the more traditional Metropolis-Hastings within Gibbs sampler required substantially more computational resources to accomplish a similar statistical accuracy to HMC.

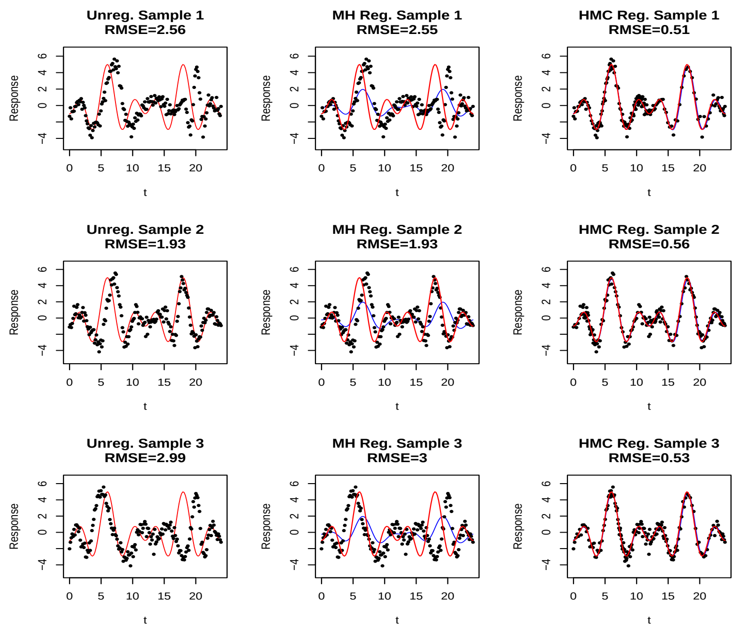

Figure 3: Points corresponding to the unregistered (left-most column) and registered values under the strict Metropolis within Gibbs and HMC adapted samplers (center and right-most columns) for the study. The red line corresponds to the true reference function, whereas the blue lines in the second and third columns correspond to the estimated reference function under the Metropolis and HMC samplers, respectively. The RMSE values were computed by taking the square root of the average squared deviances of response values from the true reference function.

Bibliography

Marron, James Stephen, James O. Ramsay, Laura M. Sangalli, and Anuj Srivastava. "Functional data analysis of amplitude and phase variation." Statistical Science (2015): 468-484.

Ramsay, J. and B. Silverman. (2005). Functional data analysis (Second ed.). New York: Springer.

Earls, C. and G. Hooker. (2017). Variational Bayes for functional data registration, smoothing, and prediction. Bayesian Analysis 12(2): 557-582.

Betancourt, M. (2017). A conceptual introduction to Hamiltonian Monte Carlo. arXiv:1701.02434.