Registration of functional data for likelihood-free inference

Anthony Ebert, Paul Wu, Kerrie Mengersen, Fabrizio Ruggeri

16 October 2018

Contents

1 Abstract 1

2 Introduction 1

3 Approximate Bayesian Computation 2

4 Explanatory example 3
- 4.1 Method . . . 3
- 4.2 Results and Discussion . . . 4

5 Motivating example: Passenger processing at an international airport 4
- 5.1 Method . . . 4
- 5.2 Results . . . 5

6 Counter example: Hydrological modelling 5
- 6.1 Methods . . . 5

7 References 6
- 7.1 Results and Discussion . . . 6

8 Discussion 6

9 System Information 6

References 6

1 Abstract

Likelihood-free inference on functional data requires specification of sensible distance measures on these functions. We consider the problem of functional data displaying not only amplitude variation but also phase variation. We show how current methods in curve registration can be applied to approximate Bayesian computation with the help of kernel-based methods recently developed for metrics for probability measures.

2 Introduction

Functional data analysis (FDA) address statistical problems involving a functional dataset, where the data comprise a set of functions (Hsing and Eubank 2015). A functional random variable y is a random variable taking values in a function space $\mathsf{f} \in \mathcal{F}$ with associated probability measure p. We consider function spaces of the form $\mathbb{R}^{[0,1]}$, that is $\mathsf{f} : [0,1] \mapsto \mathbb{R}$, transformations from different domain intervals is straight-forward. A functional dataset is a collection of observed functional random variables $\mathsf{f}^1, \mathsf{f}^2, \cdots$. In practice the functional random variables are observed with error on a countable subset of the domain of the function and this is what is referred to as a functional dataset, which contains empirical functional random variables. An empirical functional random variable (EFRV) f is itself a set of pairs $f_1, f_2, \cdots$ of the form $f_i = (t_i^f, y_i^f)$ representing sampling location t_i^f and corresponding function output y_i^f. Therefore f is of the form $\{(t_1^f, y_1^f), \cdots, (t_j^f, y_j^f), \cdots, (t_n^f, y_n^f)\}$, for fixed sampling points $t^f = \{t_1^f, \cdots, t_n^f\}$ and corresponding observed functional outputs $y^f = y_1^f, \cdots, y_n^f$. We append the superscript f to keep track of which sampling points and functional outputs refer to which EFRV.

Special-cases of functional datasets include longitudinal data and time series, however the field of FDA is much broader than this (Wang, Chiou, and Müller 2016). In analysis of functional data, a statistic of interest is an estimate of the mean function $\mu(t) = \mathop{\mathsf{E}}_{\mathsf{f}\sim p}[\mathsf{f}](t)$. In the case of EFRVs with a common set of sampling locations ($t^{f_1} = \cdots = t^{f_i} = \cdots$) this is usually estimated with $\hat{\mu}(t_j) = 1/n_i \sum_{i=1}^{n_i} y_j^{f_i}$. In either case it is possible that the resulting function is not an element of $\mathcal{F}$ due to phase variation. For instance suppose $\mathcal{F}$ is the set of Gaussian functions $\phi(\mu, \sigma)$ with $\mu \sim U(0,5)$, $\sigma = 1$. The mean function, in this case, is not Gaussian. In the next section we learn how problems of this sort are addressed in literature.

There are a variety of approaches in the literature which seek to address the provide statistical tools in the presence of phase variation, these include: dynamic time warping (Padoy et al. 2012,Wang, Gasser, and others (1997)); the Frechet distance (Rote 2007); curve registration and elastic functions (Srivastava et al. 2011). Applications of curve registration include: growth curves (Cheng et al. 2016); surgical workflow

(Padoy et al. 2012).

We standardise the domain of $\mathsf{f}^* \in \mathbb{R}^{[0,T]}$ from $[0,T]$ to $[0,1]$ by setting $\mathsf{f}(t) = \mathsf{f}^*(t \times T)$. The idea of curve registration is to align elements of $\mathcal{F}$ with warping functions $\gamma^{\mathsf{f}_i} : [0,1] \mapsto [0,1]$, such that elements of the set $\mathcal{G} := \{\mathsf{f}^i \circ \gamma^{\mathsf{f}_i} | \mathsf{f}^i \in \mathcal{F}\}$ have features which are aligned. We adopt the elastic functions approach of Srivastava et al. (2011), their approach is to find γ which minimises the Fisher-Rao metric. The distance used for this approach is the Fisher-Rao metric:

$$d_{\mathrm{FR}}(\mathsf{f}, \mathsf{g}) = \int [q^{\mathsf{f}}(t) - q^{\mathsf{g}}(t)]^2 \mathrm{d}t,$$

where $q^f(t) = \mathrm{sign}(f'(t)) \times \sqrt{|f'(t)|}$. We align a curve f to g by defining a warping function $\gamma \in \Gamma$ where $\Gamma = \{\gamma \in [0,1]^{[0,1]} | \gamma \text{ is invertible and } \gamma(0) = 0\}$ to minimise $\rho_{\mathrm{FR}}(\mathsf{f} \circ \gamma, \tilde{\mathsf{g}})$. The advantage of using ρ_{FR} for curve registration is that:

$$d_{\mathrm{FR}}(\mathsf{f}, \mathsf{g}) = d_{\mathrm{FR}}(\mathsf{f} \circ \gamma, \mathsf{g} \circ \gamma)$$

in otherwords the Fisher-Rao metric is invariant to a shared warping. This implies that the discrepancy of amplitudes (amplitude distance) between f and g, defined as

$$d_{\mathrm{amp}}(\mathsf{f}, \mathsf{g}) := \inf_{\gamma \in \Gamma} d_{\mathrm{FR}}(\mathsf{f} \circ \gamma, \mathsf{g})$$

is symmetric (Srivastava et al. 2011). In practice EFRVs f are used in place of true unobserved functions f and the same approach proceeds with straightforward approximations.

For simplicity and without loss of generality we consider problems where the functional dataset consists of a single observed EFRV f. We extend curve registration to the likelihood-free domain with approximate Bayesian computation (ABC). We proceed now with a brief background of ABC, for a detailed exposition on the subject see Sisson, Fan, and Beaumont (2018).

3 Approximate Bayesian Computation

Suppose we can to draw samples from, but not evaluate a probability distribution $p(x|\theta)$ and furthermore we wish to draw samples $\pi(\theta|x)$ such that we can conduct inference on $\theta|x$. We must proceed via simulation-based approaches. This is useful where $p(x|\theta)$ represents a complicated simulator for generating realisations $x_\theta \sim p(\cdot|\theta)$.

The theoretical foundation of ABC rests on the fact that if we draw samples θ independently from the prior $\pi(\theta)$ and use each θ to generate a corresponding x_θ then the set $\{\theta | x_\theta = x\}$ are sampled from $\pi(\theta|x)$. Such an approach would never work in practice as the proportion of x_θ samples equal to x for models with continuous support is zero. For this reason a dissimilarity d defined on the sample space is used to accept or weight realisations x_θ in relation to x according to some threshold $\epsilon > 0$. Furthermore the dissimilarity is often defined in terms of lower dimensional summary statistics of x to avoid the curse of dimensionality.

Dissimilarity, as defined by Jousselme and Maupin (2012), is a weaker notion than distance without the triangle inequality and where $x{=}y \implies d(x,y){=}0$ is true but the converse $d(x,y){=}0 \implies x{=}y$ is not true in general. For instance a common approach is to define d as a distance on lower dimensional summary statistics, this is equivalent to a dissimilarity on the sample space.

A dissimilarity that we here is Maximum mean discrepancy (MMD), as introduced by Gretton, Sejdinovic, et al. (2012), is a metric between probability measures p, q on a common probability space $\mathcal{Q}$:

$$d_{\mathrm{MMD}}(p, q, \mathcal{H}) = \sup_{h \in \mathcal{H}} \left(\int h(x)[p(x) - q(x)] \mathrm{d}x \right),$$

where $\mathcal{H}$ is a function space defined on the same domain as the probability space $\mathcal{Q}$. The function space is taken by Gretton, Borgwardt, et al. (2012) to be set of functions which integrate to 1 on the probability space of interest. Gretton, Borgwardt, et al. (2012) also developed an estimator for this metric based on observed data $\mathbf{x} \sim p$, $\mathbf{y} \sim q$. In our case $\mathbf{x}$ and $\mathbf{y} \sim q$ represent functional datasets which we label as f and g respectively. The MMD estimator is:

$$\begin{aligned} \hat{d}_{\mathrm{MMD}}(f, g) = &\frac{1}{m^2} \sum_{j=1}^{m} \sum_{j'=1}^{m} k(f_j, f_{j'}) + \\ &\frac{1}{n^2} \sum_{j=1}^{n} \sum_{j'=1}^{n} k(g_j, g_{j'}) \\ &- \frac{2}{mn} \sum_{j=1}^{m} \sum_{j'=1}^{n} k(f_j, g_{j'}), \end{aligned}$$

where m is the cardinality of f, n is the cardinality of g and k is a kernel function. A common choice of kernel function is the Gaussian kernel, $k(f_j, g_{j'}) = \exp\left[-0.5\sqrt{(t_j^f - t_{j'}^g)^T S^{-1} (y_j^f - y_{j'}^g)}\right]$, where S is a fixed tuning covariance matrix. It has been shown

that $\hat{\rho}_{\text{MMD}}$ is equivalent to a kernel-smoothed L2 norm between EFRVs. We can therefore use $\hat{\rho}_{\text{MMD}}$ as a dissimilarity on EFRVs rather than probability measures.

In summary ABC samplers return samples of θ depending on the value of $d(f, f_\theta)$. There exist many varieties of ABC samplers in the literature including: accept-reject, replenishment (Drovandi and Pettitt 2011); simulated annealing (Albert, Künsch, and Scheidegger 2015) , ... We find it useful in this paper to introduce an additional abstraction called a loss function (Algorithm 1), which helps to simplify the clarify the language used later.

Algorithm 1 Loss function

1: **function** LOSS($\theta, f|p, d$)
2: $f_\theta \sim p(f|\theta)$
3: $d_\theta = d(f, f_\theta)$
4: **return** d_θ
5: **end function**

This notion makes ABC samplers easier to describe, for instance the rejection sampler can be written as a for loop over i with $(\theta_i \sim \pi(\cdot), d_i = \text{loss}(\theta_i, x))$ keeping only pairs with d_i lower than ϵ. We adopt the ABC sampler as defined by Drovandi and Pettitt termed the replenishment ABC sampler (Appendix).

To develop an ABC sampler for functional data we develop a loss function on θ and functional datasets $\mathsf{f}f \sim \mathcal{F}$ based on the elastic dissimilarity of Srivastava et al. (2011) combined with a discrepancy. In the next section we compare four loss functions, the first two are inelastic loss functions based on MMD and FR discrepancies. The last two are registered loss functions (Algorithm 2), where we sample f_θ from $p(f|\theta)$, register f_θ to f to minimise the FR dissimilarity and then compute the disimilarity between f and the registered version of f_θ.

Algorithm 2 Registered loss function

1: **function** LOSS($\theta, f|p, d$)
2: $f_\theta \sim p(f|\theta)$
3: $\gamma = \arg\inf_{\gamma\in\Gamma} d_{\text{FR}}(f, f_\theta \circ \gamma)$
4: $d_\theta = d(f, f_\theta \circ \gamma)$
5: **return** d_θ
6: **end function**

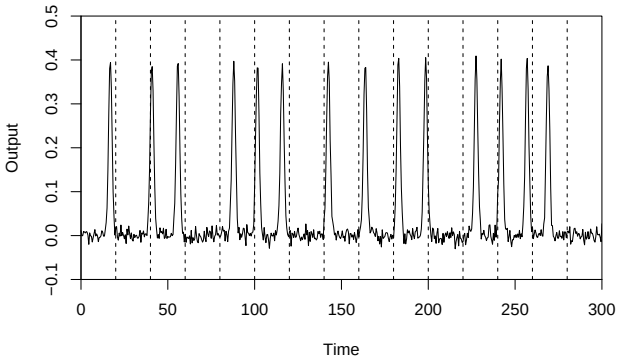


Figure 1: Example of a draw from the Gaussian peak shift model with $\sigma_a = 5, \sigma_\phi = 1$, and $\sigma_\epsilon = 0.01$. The dotted lines are the α values. The distribution of peak shifts is controlled by σ_a, the peak widths are controlled by σ_ϕ and the noise is controlled by σ_ϵ.

4 Explanatory example

4.1 Method

We first construct an artificial problem to demonstrate the method we suggest. The mean function $\mu(t)$ is a sum over 14 Gaussian functions i.e. $\mu(t) = \sum_{k=1}^{14} \phi(t|\mu_k, \sigma_\phi)$. The observed functional output corresponding to fixed sampling locations $t = (0, 0.5, \cdots, 300)$, are independent and normally distributed $y_j \sim N(\mu(t_j), \sigma_\epsilon)$. What makes this problem of interest here is that the mean parameters μ_k are latent random variables $\mu_k \sim N(a_i, \sigma_a)$ making inelastic dissimilarities inappropriate.

The prior distributions for the parameters of interest are $\sigma_\phi \sim U(0, 10)$, and $\sigma_\epsilon \sim U(0, 0.1)$. An example of a model realisation from this distribution is shown in Figure 1.

To test the effectiveness of registered loss functions in this situation we compare all four combinations i.e. MMD and FR with and without registration. We set the covariance matrix for MMD to be the 2×2 diagonal matrix with elements 9 and 1×10^{-4}. We use the R package fdasrvf (Tucker 2017) to align the simulated and observed functional datsets. To compute the elastic distance between functions as defined by Srivastava et al. (2011) we use the implementation contained in the R package fdasrvf (Tucker 2017).

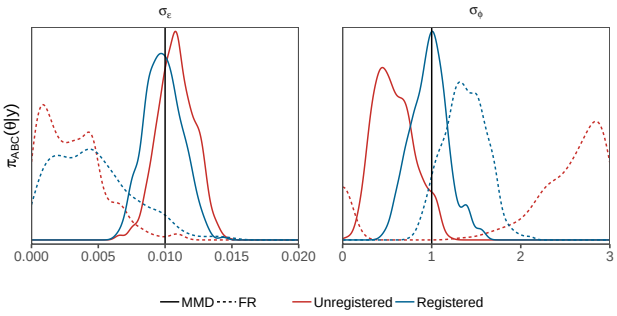


Figure 2: Density plots of posterior samples arising from the replenishment ABC sampler (Drovandi and Pettitt 2011) for the Gaussian peak shift example. Distances shown include MMD (Maximum mean discrepancy) and FR (Fisher-Rao) in on both registered and unregistered data. The true values are shown in the vertical black solid lines.

4.2 Results and Discussion

Density plots of posterior samples arising from the ABC sampler with all four loss functions , registered and unregistered, (Figure 2) show that the registered loss functions outperform their unregistered counterparts. Interestingly the distance we use for registration, FR, when used as the distance for ABC is outperformed by MMD which leads us to believe that the best dissimilarity for registration and the best dissimilarity for ABC do not always coincide. The reason MMD outperforms FR for σ_ϵ in particular is that MMD the first term in Equation can be thought of a penalty for concentrating y^{f_θ} such that the algorithm is lead to a higher value of θ_ϵ to match the noise in y^f. We are interested in the effect of registration so in further examples we proceed only with MMD.

We demonstrate now a more complex example with a dynamic queueing network model of an international airport terminal.

5 Motivating example: Passenger processing at an international airport

Rising demand for air travel and enhanced security screening places pressure on existing airport infrastructure. Airport terminal infrastructure for financial and geographical reasons is difficult to upgrade at such a rate as to match this demand. With this in mind operational planners at airport seek to optimise day-to-day operations. However any optimisation framework requires a model with many parameters, to be a realistic and useful model is will necessarily become complicated and this combined with the data collection scheme leads inevitably to intractable likelihoods. On the other hand simulators are straight-forward to construct although computationally demanding ruling out simulation-based approaches such as ABC. This means that parameter inference is often conducted in an ad-hoc and unprincipled manner.

However a new simulation algorithm for queueing based algorithms called QDC (Ebert et al. 2017) was used by Ebert et al. (2018) to construct an ABC sampler for this problem for the first time, however there were problems with missalignment as noted in the discussion of this paper. We consider a similar model and data as Ebert et al. (2018). The purpose of this model is to predict passenger flows through an airport terminal in response to particular flight schedules and staff rosters. The data comprise records of passenger numbers passing through certain checkpoints for each minute of the day. The flight schedule and staff rosters can be thought of as explanatory variables since they are known and effect the response variable (passenger flows) to some degree which we wish to infer.

After passengers are deplaned from an arriving flight they walk to the immigration system and so a wave of customers crashes down on the shoreline of immigration.

5.1 Method

To model passenger flows we simulate the movements of each passenger j from each flight i. Flight i, begin passenger disembarkation at a_i, passenger j from this flight disembarks from the aircraft at time $d_{ij}^{\text{dis}} = a_i + t_i^{\text{dis}}$, leaves the arrivals concourse at time $d_{ij}^{\text{ac}} = d_{ij}^{\text{dis}} + t_{ij}^{\text{imm}}$, chooses a route r_{ij} through immigration and finishes immigration at time $d_{ij}^{\text{imm}} = d_{ij}^{\text{ac}} + w_{ij}^{\text{imm}} + s_{ij}^{\text{imm}}$. Distributions for disembarkation times t_i^{dis} are fitted prior to analysis, however the other random variables are distributed according to unknown parameters $\alpha, \beta, \lambda_{\text{SG}}, \lambda_{\text{MG}}$ in the following

manner:

Flight disembarkation start	$a_i \sim \mathrm{U}(A_i - 20$
Passenger disembarkation time	$t_{ij} \sim \mathrm{G}(\alpha_i^{\mathrm{dis}}, \beta_i^{\mathrm{di}}$
Walking times	$t_{ij}^{\mathrm{ac}} \sim \mathrm{Gamma}\left\{ ($
Nationality (local or foreign)	$\mathrm{nat}_{ij} \sim \mathrm{Bern}(p_i^{\mathrm{nat}})$
Passenger route (SG or MG)	$r_{ij}\vert\mathrm{nat}_{ij} \sim \mathrm{Bern}(p_{\mathrm{nat}_{ij}}^{\mathrm{imm}}$
Service times	$s_{ij}\vert r_{ij} \sim \mathrm{Exp}\left(\lambda_{r_{ij}}\right)$

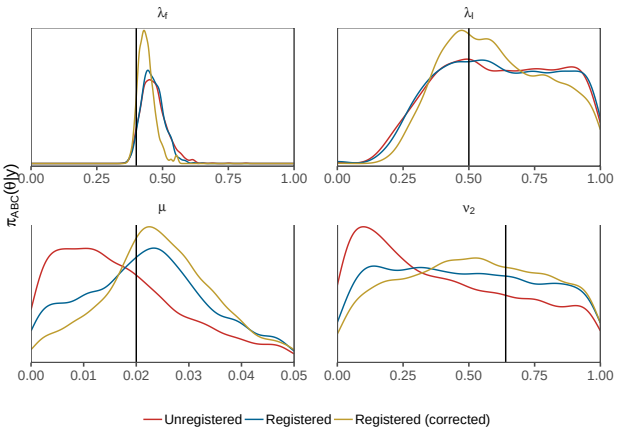
where A_i is the scheduled time for the flight to begin passenger disembarkation. This is a hierarchical model which consists of flight effects, such as a_i and passenger effects, such as s_i.

The last step to recreate the data is to bin the departure vectors $\vec{d}^{\mathrm{ac}}, \vec{d}^{\mathrm{imm}}$, and $\vec{d}_{\mathrm{SG}}^{\mathrm{imm}}$ by minute. We refer to these model realisations as $\vec{x}^{\mathrm{ac}}, \vec{x}^{\mathrm{imm}}$, and $\vec{x}_{\mathrm{SG}}^{\mathrm{imm}}$ which resemble the observed data $\vec{y}^{\mathrm{ac}}, \vec{y}^{\mathrm{imm}}$, and $\vec{y}_{\mathrm{SG}}^{\mathrm{imm}}$. A realisation from this model is shown below.

We compare three distances:

$$\mathbf{D}_0 = \hat{\rho}_{\mathrm{MMD}}(\vec{x}^{\mathrm{ac}}, \vec{y}^{\mathrm{ac}}) + \hat{\rho}_{\mathrm{MMD}}(\vec{x}^{\mathrm{imm}}, \vec{y}^{\mathrm{imm}}) + \hat{\rho}_{\mathrm{MMD}}(\vec{x}_{\mathrm{SG}}^{\mathrm{imm}}, \vec{y}_{\mathrm{SG}}^{\mathrm{imm}})$$
$$\mathbf{D}_1 = \hat{\rho}_{\mathrm{EMM}}(\vec{x}^{\mathrm{ac}}, \vec{y}^{\mathrm{ac}}) + \hat{\rho}_{\mathrm{MMD}}(\vec{x}^{\mathrm{imm}}, \vec{y}^{\mathrm{imm}}) + \hat{\rho}_{\mathrm{MMD}}(\vec{x}_{\mathrm{SG}}^{\mathrm{imm}}, \vec{y}_{\mathrm{SG}}^{\mathrm{imm}})$$
$$\mathbf{D}_2 = \hat{\rho}_{\mathrm{EMM}}(\vec{x}^{\mathrm{ac}}, \vec{y}^{\mathrm{ac}}) + \hat{\rho}_{\mathrm{MMD}}(\vec{z}^{\mathrm{imm}}, \vec{y}^{\mathrm{imm}}) + \hat{\rho}_{\mathrm{MMD}}(\vec{z}_{\mathrm{SG}}^{\mathrm{imm}}, \vec{y}_{\mathrm{SG}}^{\mathrm{imm}}),$$

5.2 Results



6 Counter example: Hydrological modelling

Water flows within catchment areas in response to rainfall and other weather-events are forecasted with run-off models. Runoff models have been used successfully by .... in particular the GR4J (Génie Rural

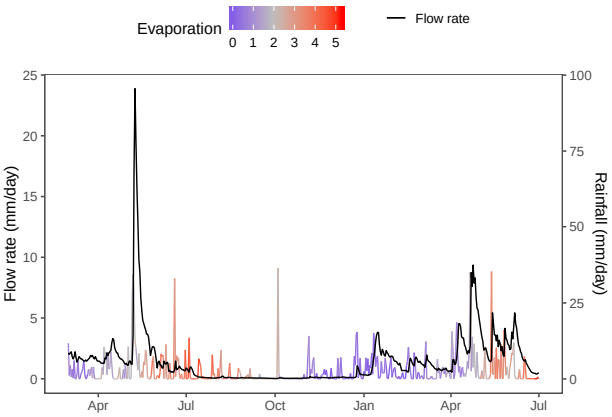


Figure 3: Hydrograph with associated rain and evaporation data from airGR package

à 4 paramètres Journalier) model of Perrin, Michel, and Andréassian (2003) is widely used.

Water flows in hydrology are represented by hydrographs representing volumetric flow-rates at a particular point (see Figure 3).

Parameter estimation proceeds with either expert or automatic calibration. Automatic calibration is typically performed with a goodness-of-fit function such as sum of squared errors rather than a likelihood function. The NSE function is a goodness-of-fit function specifically developed for hydrology. Parameter estimation for hydrological methods is an active area of research Kavetski (2018). It is argued in the literature that once parameters are estimated the parameter uncertainty itself contributes little to overall uncertainty McInerney et al. (2018) compared to the residual error structure.

We apply approximate Bayesian computation for this application for the first time and apply the methodology of curve-registration.

6.1 Methods

The simulation model we use is the GR4J model of Perrin, Michel, and Andréassian (2003). This model contains four parameters labelled: the first three $(\theta_1, \theta_2, \theta_3)$ are capacities in units of length (mm); and the fourth parameter (θ_4) represents a lagged effect in units of time (days).

Predictions from the GR4J given parameters $\theta = (\theta_1, \theta_2, \theta_3, \theta_4)$ and explanatory variables (X) are denoted as $\mathcal{H}(\theta, X)$. Like many run-off models the GR4J

model is deterministic, the practice is to add an error structure on top of this deterministic model, the model in its entirity is then:

$$Y = \mathcal{H}(\theta, X) + \epsilon. \tag{1}$$

7 References

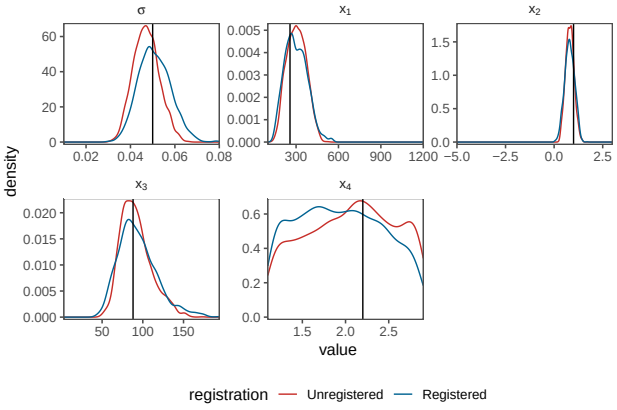
Run-off predictions are positively valued, and it is well known that heteroscedasticity exists in the errors (Kavetski 2018). A common strategy is to perform a Box-Cox transformation $Z(Y, \lambda) = (Y^\lambda - 1)/\lambda$ to transformed predictions with normalised errors η:

$$Y_Z = Z(\mathcal{H}(\theta, X)) + \eta, \tag{2}$$
$$\eta \sim N(0, \sigma^2) \tag{3}$$

We apply the MMD, Registered-MMD and elastic distances as shown previously. Once again we test all three distances.

7.1 Results and Discussion

In this example we find that registered distances perform more poorly than their unregistered counterparts.



8 Discussion

We have compared likelihood-free inference for three problems using three distances. The distances were the elastic distance (as proposed by . . .), maximum mean discrepancy (Gretton et al. . .) and a distance incorporating techniques from both registered maximum mean discrepancy.

The first two problems were amenable to this new distance and the last problem was not. Although all three problems incorporate some aspect of peak position uncertainty, the degrees of freedom for the first two problems was much larger than the last.

9 System Information

```
## R version 3.5.1 (2018-07-02)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Debian GNU/Linux 9 (stretch)
##
## Matrix products: default
## BLAS: /usr/lib/libblas/libblas.so.3.7.0
## LAPACK: /usr/lib/lapack/liblapack.so.3.7.0
##
## locale:
##  [1] LC_CTYPE=en_AU.UTF-8       LC_NUMERIC=C
##  [3] LC_TIME=en_AU.UTF-8        LC_COLLATE=en_AU.UTF-8
##  [5] LC_MONETARY=en_AU.UTF-8    LC_MESSAGES=en_AU.UTF-8
##  [7] LC_PAPER=en_AU.UTF-8       LC_NAME=C
##  [9] LC_ADDRESS=C               LC_TELEPHONE=C
## [11] LC_MEASUREMENT=en_AU.UTF-8 LC_IDENTIFICATION=C
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  m
##
## other attached packages:
## [1] airGR_1.0.14.1          bindrcpp_0.2.2          ggp
## [4] dplyr_0.7.6             CurveRegistration_0.0.2
##
## loaded via a namespace (and not attached):
##  [1] latex2exp_0.4.0  Rcpp_0.12.19     pillar_1.3.0
##  [5] plyr_1.8.4       highr_0.7        bindr_0.1.1
##  [9] digest_0.6.16    evaluate_0.11    tibble_1.4.2
## [13] pkgconfig_2.0.2  rlang_0.2.2      yaml_2.2.0
## [17] stringr_1.3.1    knitr_1.20       rprojroot_1.3-2
## [21] tidyselect_0.2.4 glue_1.3.0       R6_2.2.2
## [25] purrr_0.2.5      tidyr_0.8.1      magrittr_1.5
## [29] scales_1.0.0     codetools_0.2-15 htmltools_0.3.6
## [33] assertthat_0.2.0 colorspace_1.3-2 labeling_0.3
## [37] lazyeval_0.2.1   munsell_0.5.0    crayon_1.3.4
```

References

Albert, Carlo, Hans R. Künsch, and Andreas Scheidegger. 2015. "A Simulated Annealing Approach to Approximate Bayes Computations." *Statistics and*

Computing 25 (6): 1217–32. http://link.springer.com/article/10.1007/s11222-014-9507-8.

Cheng, Wen, Ian L Dryden, Xianzheng Huang, and others. 2016. "Bayesian Registration of Functions and Curves." *Bayesian Analysis* 11 (2). International Society for Bayesian Analysis: 447–75.

Drovandi, Christopher C., and Anthony N. Pettitt. 2011. "Estimation of Parameters for Macroparasite Population Evolution Using Approximate Bayesian Computation." *Biometrics* 67 (1): 225–33. http://onlinelibrary.wiley.com/doi/10.1111/j.1541-0420.2010.01410.x/full.

Ebert, Anthony, Ritabrata Dutta, Paul Wu, Kerrie Mengersen, Fabrizio Ruggeri, and Antonietta Mira. 2018. "Likelihood-Free Parameter Estimation for Dynamic Queueing Networks." *arXiv:1804.02526*.

Ebert, Anthony, Paul Wu, Kerrie Mengersen, and Fabrizio Ruggeri. 2017. "Computationally Efficient Simulation of Queues: The R Package Queuecomputer." *arXiv:1703.02151*. https://arxiv.org/abs/1703.02151.

Gretton, Arthur, Karsten M Borgwardt, Malte J Rasch, Bernhard Schölkopf, and Alexander Smola. 2012. "A Kernel Two-Sample Test." *Journal of Machine Learning Research* 13 (Mar): 723–73.

Gretton, Arthur, Dino Sejdinovic, Heiko Strathmann, Sivaraman Balakrishnan, Massimiliano Pontil, Kenji Fukumizu, and Bharath K. Sriperumbudur. 2012. "Optimal Kernel Choice for Large-Scale Two-Sample Tests." In *Advances in Neural Information Processing Systems*, 1205–13. http://papers.nips.cc/paper/4727-optimal-kernel-choice-for-large-scale-two-sample-tests.

Hsing, Tailen, and Randall Eubank. 2015. *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. John Wiley & Sons.

Jousselme, Anne-Laure, and Patrick Maupin. 2012. "Distances in Evidence Theory: Comprehensive Survey and Generalizations." *International Journal of Approximate Reasoning* 53 (2). Elsevier: 118–45.

Kavetski, Dmitri. 2018. "Parameter Estimation and Predictive Uncertainty Quantification in Hydrological Modelling." *Handbook of Hydrometeorological Ensemble Forecasting*. Springer, 1–42.

McInerney, David, Mark Thyer, Dmitri Kavetski, Bree Bennett, Julien Lerat, Matthew Gibbs, and George Kuczera. 2018. "A Simplified Approach to Produce Probabilistic Hydrological Model Predictions." *Environmental Modelling & Software* 109. Elsevier: 306–14.

Padoy, N., T. Blum, S.-A. Ahmadi, H. Feussner, M.-O. Berger, and N. Navab. 2012. "Statistical Modeling and Recognition of Surgical Workflow." *Medical Image Analysis* 16 (3): 632–41. doi:10.1016/j.media.2010.10.001.

Perrin, Charles, Claude Michel, and Vazken Andréassian. 2003. "Improvement of a Parsimonious Model for Streamflow Simulation." *Journal of Hydrology* 279 (1-4). Elsevier: 275–89.

Rote, Günter. 2007. "Computing the Fréchet Distance Between Piecewise Smooth Curves." *Computational Geometry* 37 (3). Elsevier: 162–74.

Sisson, Scott A, Yanan Fan, and Mark Beaumont. 2018. *Handbook of Approximate Bayesian Computation*. Chapman; Hall/CRC.

Srivastava, Anuj, Wei Wu, Sebastian Kurtek, Eric Klassen, and JS Marron. 2011. "Registration of Functional Data Using Fisher-Rao Metric." *arXiv:1103.3817*.

Tucker, J. Derek. 2017. *Fdasrvf: Elastic Functional Data Analysis*. https://CRAN.R-project.org/package=fdasrvf.

Wang, Jane-Ling, Jeng-Min Chiou, and Hans-Georg Müller. 2016. "Functional Data Analysis." *Annual Review of Statistics and Its Application* 3. Annual Reviews: 257–95.

Wang, Kongming, Theo Gasser, and others. 1997. "Alignment of Curves by Dynamic Time Warping." *The Annals of Statistics* 25 (3). Institute of Mathematical Statistics: 1251–76.