An Invitation to Statistics in Wasserstein Space

Shen Ziliang 2021213218

Shanghai University of Finance and Economics

April 24, 2022

Outline

Content

Prior Instructions

In this section, we illustrate the estimation framework put forth in this chapter by considering an example of a structural mean λ with a bimodal density on the real line. The unwarped point patterns Π originate from Poisson processes with mean measure λ and, consequently, the warped points $\tilde{\Pi}$ are Cox processes (see Sect. 4.1.2).

Explicit Classes of Warp Maps

As a first step, we introduce a class of random warp maps satisfying Assumptions 2, that is, increasing maps that have as mean the identity function. For any integer k define $\zeta_k : [0,1] \to [0,1]$ by

$$\zeta_0(x) = x, \qquad \zeta_k(x) = x - \frac{\sin(\pi k x)}{|k|\pi}, \quad k \in \mathbb{Z} \setminus \{0\}.$$

Clearly $\zeta_k(0) = 0$, $\zeta_k(1) = 1$ and ζ_k is smooth and strictly increasing for all k. Figure 4.4a plots ζ_k for $k = -3, \ldots, 3$. To make ζ_k a random function, we let k be an integer-valued random variable. If the latter is symmetric, then we have

$$\mathbb{E}[\zeta_k(x)] = x, \qquad x \in [0,1].$$

Explicit Classes of Warp Maps

By means of mixtures, we replace this discrete family by a continuous one: let $J > 1$ be an integer and $V = (V_1, \ldots, V_J)$ be a random vector following the flat Dirichlet distribution (uniform on the set of non-negative vectors with $v_1 + \cdots + v_J = 1$). Take independently k_j following the same distribution as k and define

$$T(x) = \sum_{j=1}^{J} V_j \zeta_{k_j}(x).$$

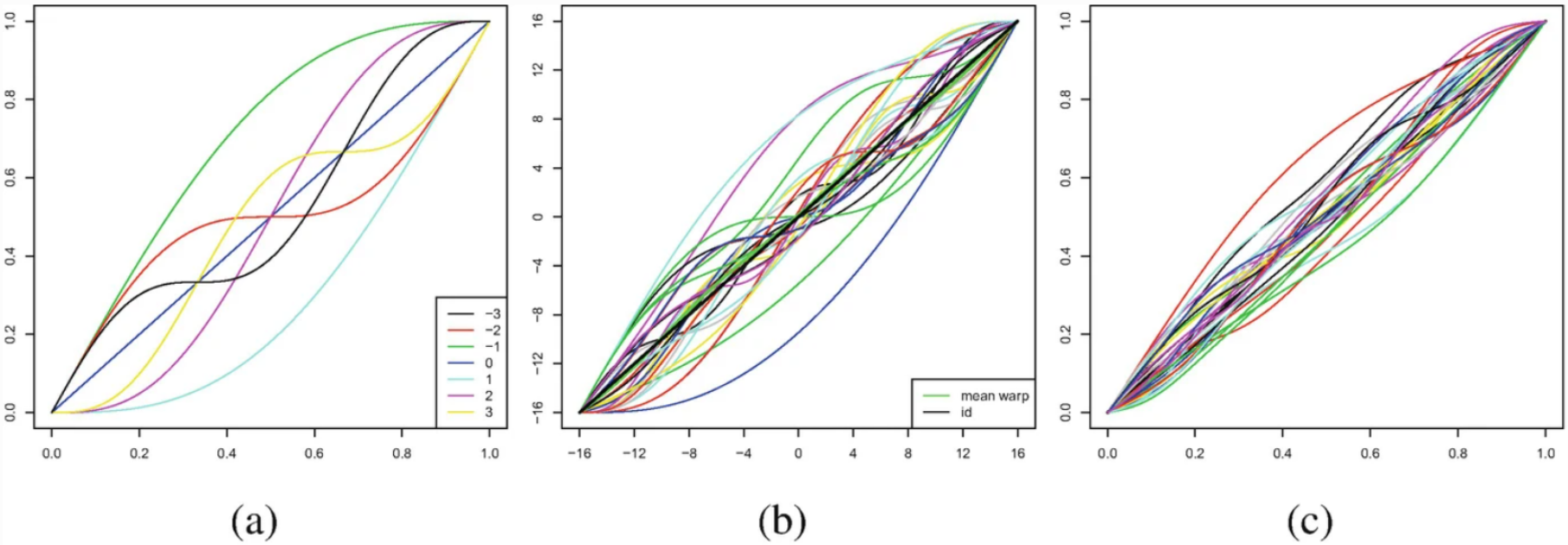
Since V_j is positive, T is increasing and as (V_j) sums up to unity T has mean identity. Realisations of these warp functions are given in Fig. 4.4b and c for $J = 2$ and $J = 10$, respectively. The parameters (k_j) were chosen as symmetrised Poisson random variables: each k_j has the law of XY with X Poisson with mean 3 and $\mathbb{P}(Y = 1) = \mathbb{P}(Y = -1) = 1/2$ for Y and X independent.

Explicit Classes of Warp Map

When $J = 10$ is large, the function T deviates only mildly from the identity, since a law of large numbers begins to take effect. In contrast, $J = 2$ yields functions that are quite different from the identity. Thus, it can be said that the parameter J controls the variance of the random warp function T.

Fig. 4.4

From: Phase Variation and Fréchet Means



(**a**) The functions $\{\zeta_{-3}, ..., \zeta_3\}$; (**b**) realisations of T defined by (4.5) with $J = 2$ and k_j symmetrisations of Poisson random variables with mean 3; (**c**) realisations of T defined by (4.5) with $J = 10$ and k_j as in (**b**)

Bimodal Cox Processes

Let the structural mean measure λ be a mixture of a bimodal Gaussian distribution (restricted to $K = [16, 16]$) and a beta background on the interval $[12, 12]$, so that mass is added at the centre of K but not near the boundary. In symbols this is given as follows. Let ϕ be the standard Gaussian density and let $\beta_{\alpha,\beta}$ denote the density of a the beta distribution with parameters α and β. Then λ is chosen as the measure with density.

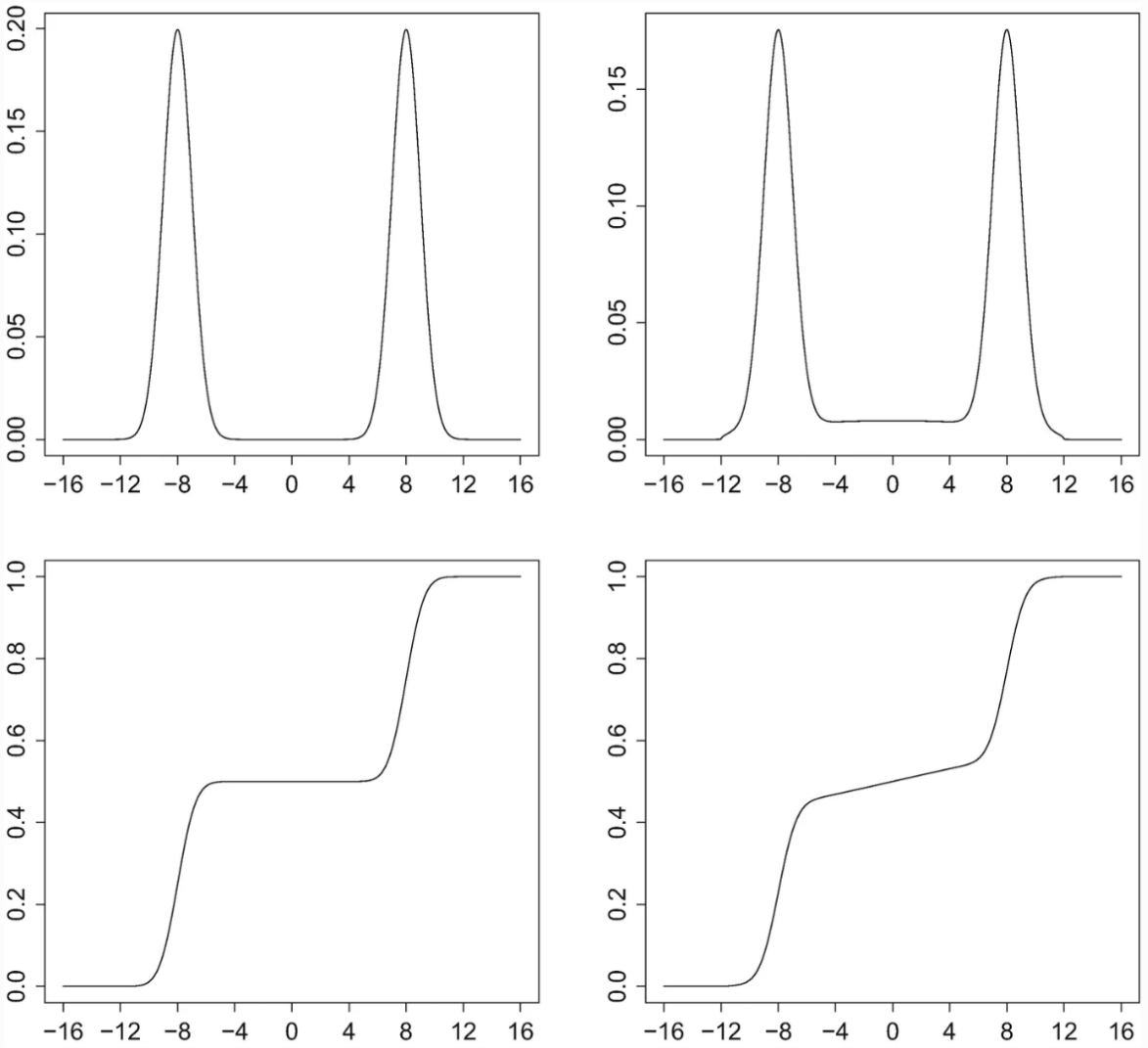
$$f(x) = \frac{1-\epsilon}{2}[\varphi(x-8) + \varphi(x+8)] + \frac{\epsilon}{24}\beta_{1.5,1.5}\left(\frac{x+12}{24}\right), \qquad x \in [-16, 16],$$

where $\epsilon \in [0, 1]$ is the weight of the beta background. (We ignore the loss of a negligible amount of mass due to the restriction of the Gaussians to $[16, 16]$.) Plots of the density and distribution functions are given in Fig. 4.5.

Bimodal Cox Processes

Fig. 4.5

From: Phase Variation and Fréchet Means



Density and distribution functions corresponding to (4.6) with $\epsilon = 0$ and $\epsilon = 0.15$

Bimodal Cox Processes

The main criterion for the quality of our regularised Fréchet–Wasserstein estimator will be its success in discerning the two modes at ± 8; these will be smeared by the phase variation arising from the warp functions. We next simulated 30 independent Poisson processes with mean measure $\lambda, \epsilon = 0.1$, and total intensity (expected number of points) $\tau = 93$. In addition, we generated warp functions as in (4.5) but rescaled to $[-16, 16]$; that is, having the same law as the functions

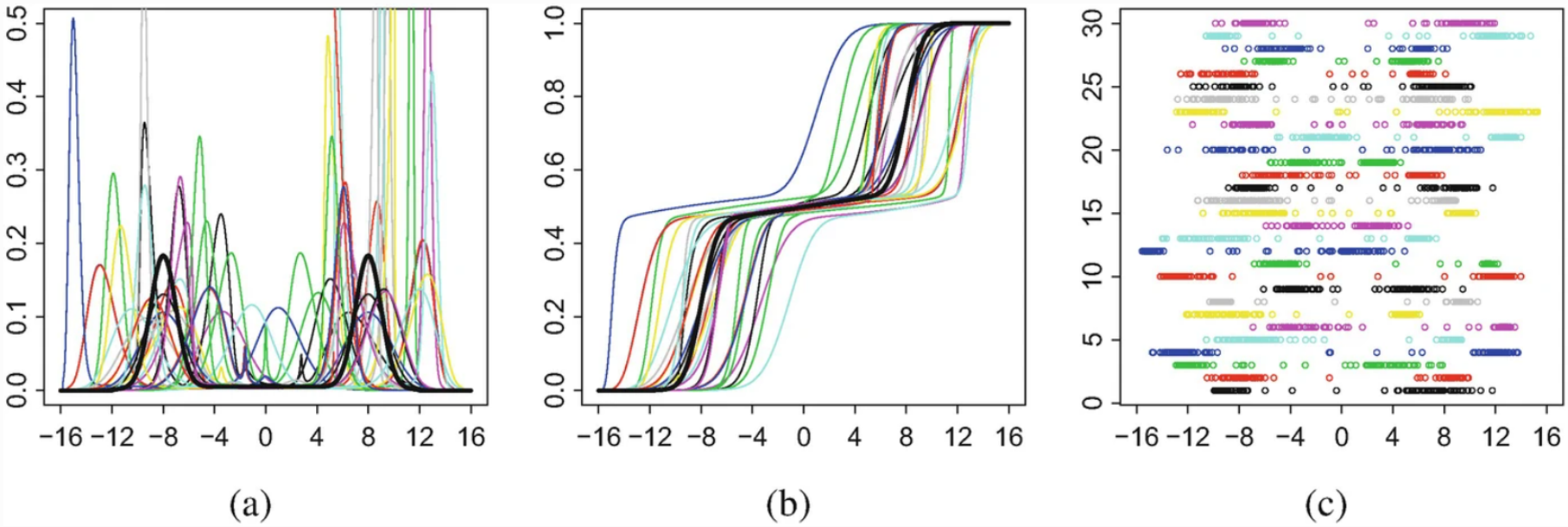
$$x \mapsto 32\, T\left(\frac{x+16}{32}\right) - 16$$

from K to K. These cause rather violent phase variation, as can be seen by the plots of the densities and distribution functions of the conditional measures $\Pi = T\#\lambda$ presented in Fig. 4.6a and b; the warped points themselves are displayed in Fig. 4.6c.

Bimodal Cox Processes

Fig. 4.6

From: Phase Variation and Fréchet Means



(**a**) 30 warped bimodal densities, with density of λ given by (4.6) in solid black; (**b**) their corresponding distribution functions, with that of λ in solid black; (**c**) 30 Cox processes, constructed as warped versions of Poisson processes with mean intensity $93f$ using as warp functions the rescaling to $[-16.16]$ of (4.5)

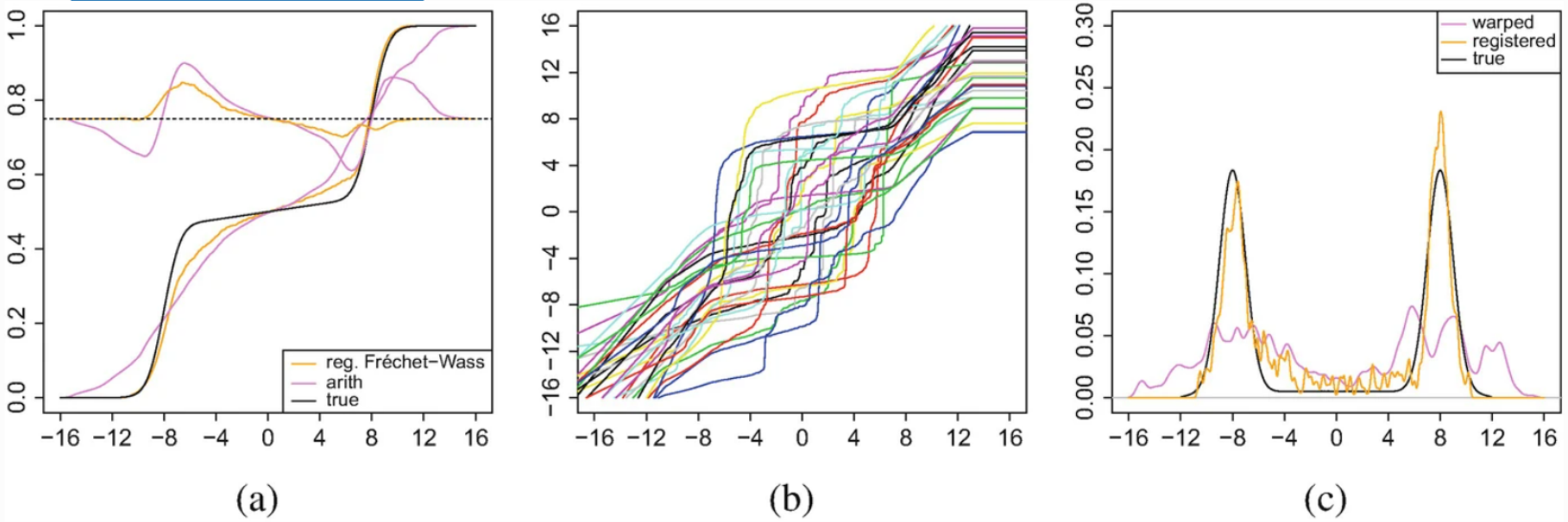
Bimodal Cox Processes

Using these warped point patterns, we construct the regularised Fréchet–Wasserstein estimator employing the procedure described in Sect. 4.3. Each $\tilde{\Pi}_i$ was smoothed with a Gaussian kernel and bandwidth chosen by unbiased cross validation. We deviate slightly from the recipe presented in Sect. 4.3 by not restricting the resulting estimates to the interval $[-16, 16]$, but this has no essential effect on the finite sample performance. The regularised Fréchet–Wasserstein estimator $\hat{\lambda}_n$ serves as the estimator of the structural mean λ and is shown in Fig. 4.7a. It is contrasted with λ at the level of distribution functions, as well as with the empirical arithmetic mean; the latter, the naive estimator, is calculated by ignoring the warping and simply averaging linearly the (smoothed) empirical distribution functions across the observations. We notice that $\hat{\lambda}_n$ is rather successful at locating the two modes of λ, in contrast with the naive estimator that is more diffuse. In fact, its distribution function increases approximately linearly, suggesting a nearly constant density instead of the correct bimodal one.

Bimodal Cox Processes

Fig. 4.7

From: Phase Variation and Fréchet Means



(**a**) Comparison between the regularised Fréchet–Wasserstein estimator, the empirical arithmetic mean, and the true distribution function, including residual curves centred at $y = 3/4$; (**b**) The estimated warp functions; (**c**) Kernel estimates of the density function f of the structural mean, based on the warped and registered point patterns

Bimodal Cox Processes

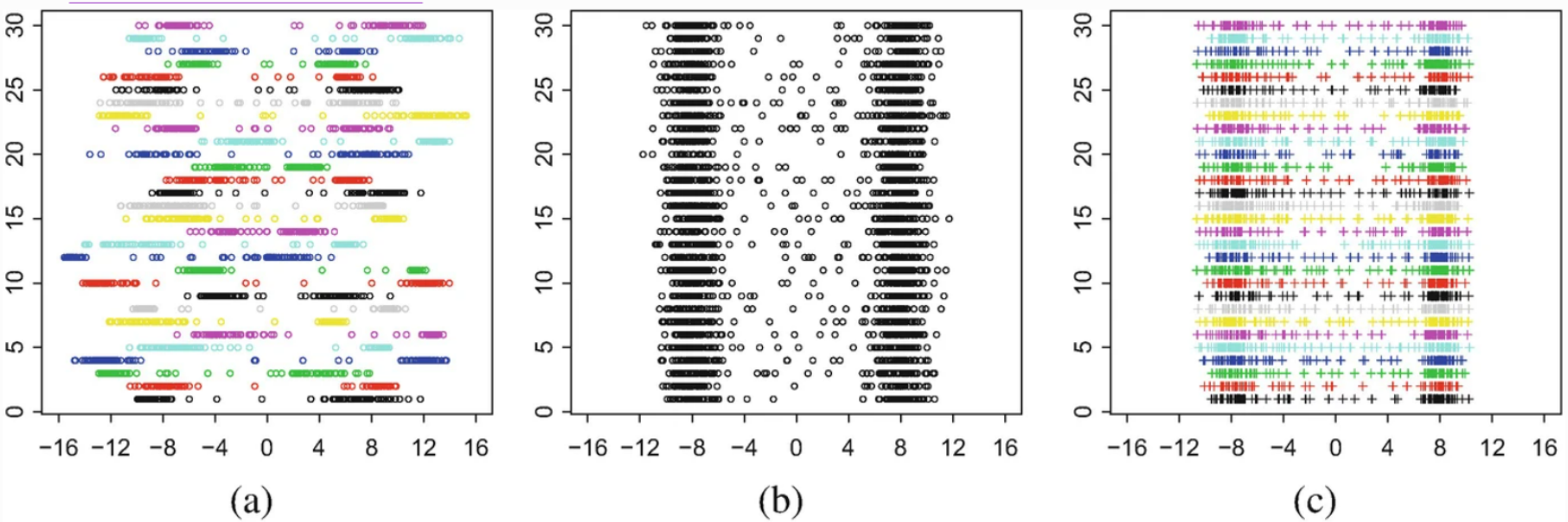
Estimators of the warp maps $\hat{T}_i$, depicted in Fig. 4.7b, and their inverses, are defined as the optimal maps between $\hat{\lambda}_n$ and the estimated conditional mean measures, as explained in Sect. 4.3.4. Then we register the point patterns by applying to them the inverse estimators $\hat{T_i^{-1}}$ (Fig. 4.8). Figure 4.7c gives two kernel estimators of the density of λ constructed from a superposition of all the warped points and all the registered ones. Notice that the estimator that uses the registered points is much more successful than the one using the warped ones in discerning the two density peaks. This is not surprising after a brief look at Fig. 4.8, where the unwarped, warped, and registered points are displayed. Indeed, there is very high concentration of registered points around the true location of the peaks, ± 8. This is not the case for the warped points because of the phase variation that translates the centres of concentration for each individual observation.

Bimodal Cox Processes

It is important to remark that the fluctuations in the density estimator in Fig. 4.7c are not related to the registration procedure, and could be reduced by a better choice of bandwidth. Indeed, our procedure does not attempt to estimate the density, but rather the distribution function.

Fig. 4.8

From: Phase Variation and Fréchet Means



Bimodal Cox processes: (**a**) the observed warped point processes; (**b**) the unobserved original point processes; (**c**) the registered point processes

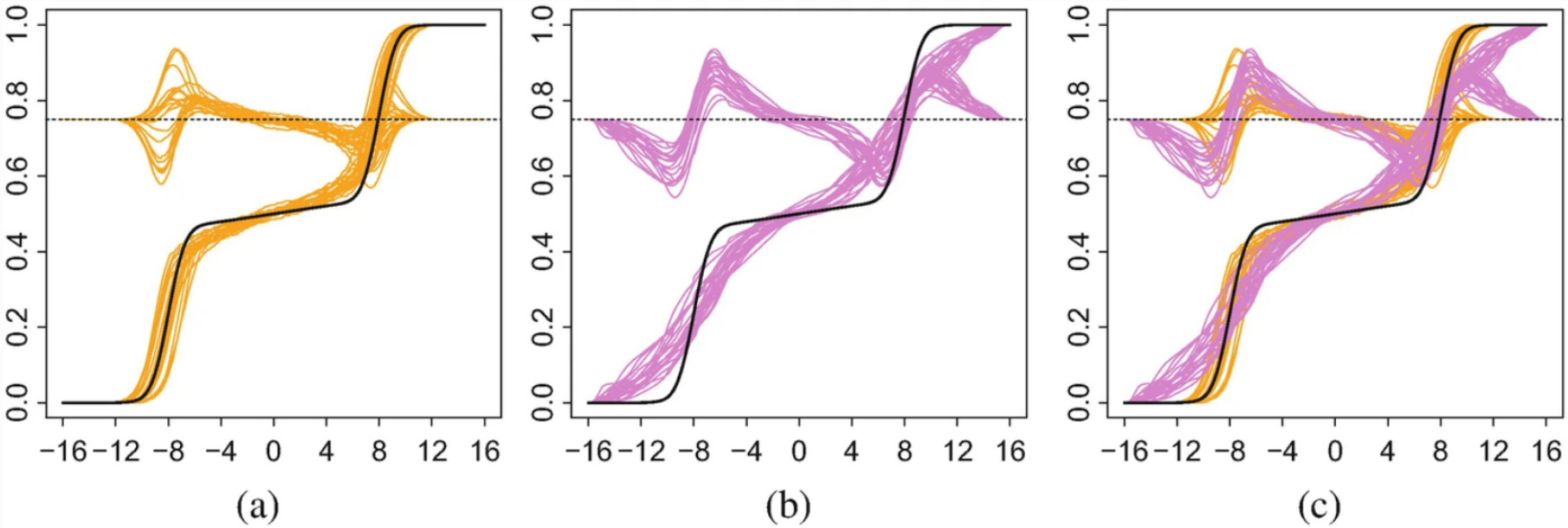
Bimodal Cox Processes

Figure 4.9 presents a superposition of the regularised Fréchet–Wasserstein estimators for 20 independent replications of the experiment, contrasted with a similar superposition for the naive estimator. The latter is clearly seen to be biased around the two peaks, while the regularised Fréchet–Wasserstein seems approximately unbiased, despite presenting fluctuations. It always captures the bimodal nature of the density, as is seen from the two clear elbows in each realisation.

Bimodal Cox Processes

Fig. 4.9

From: Phase Variation and Fréchet Means



(**a**) Sampling variation of the regularised Fréchet–Wasserstein mean $\widehat{\lambda}_n$ and the true mean measure λ for 20 independent replications of the experiment; (**b**) sampling variation of the arithmetic mean, and the true mean measure λ for the same 20 replications; (**c**) superposition of (**a**) and (**b**). For ease of comparison, all three panels include residual curves centred at $y = 3/4$

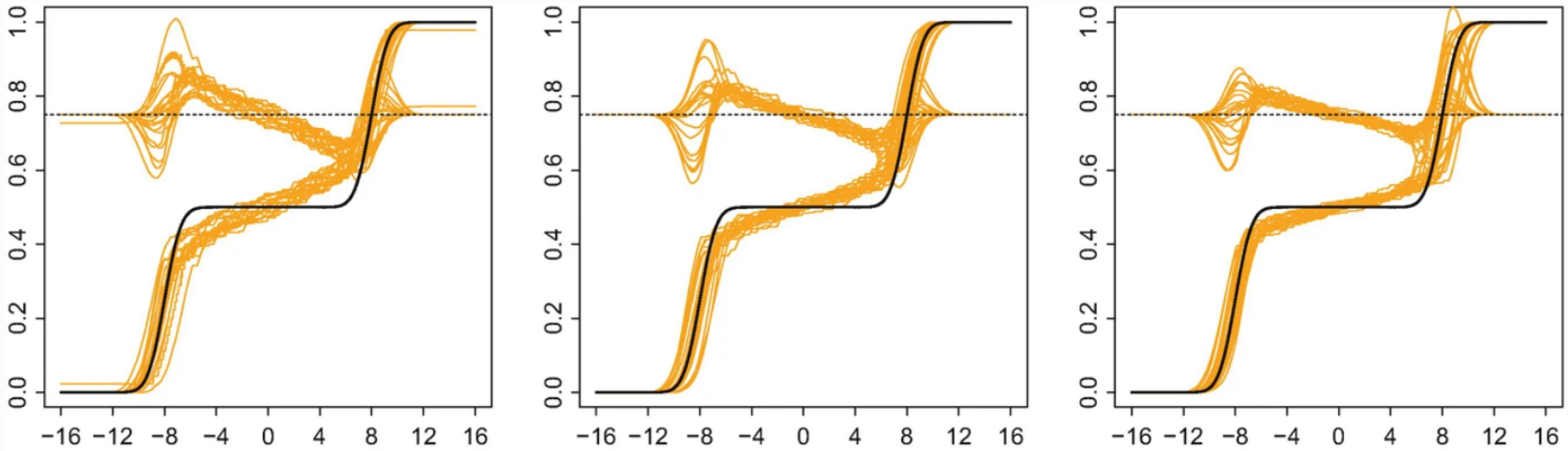
Bimodal Cox Processes

To illustrate the consistency of the regularised Fréchet–Wasserstein estimator $\hat{\lambda}_n$ for λ as shown in Theorem 4.4.1, we let the number of processes n as well as the expected number of observed point per process τ to vary. Figures 4.10 and 4.11 show the sampling variation of $\hat{\lambda}_n$ for different values of n and τ. We observe that as either of these increases, the realisations $\hat{\lambda}_n$ indeed approach λ. The figures suggest that, in this scenario, the amplitude variation plays a stronger role than the phase variation, as the effect of τ is more substantial.

Bimodal Cox Processes

Fig. 4.10

From: Phase Variation and Fréchet Means

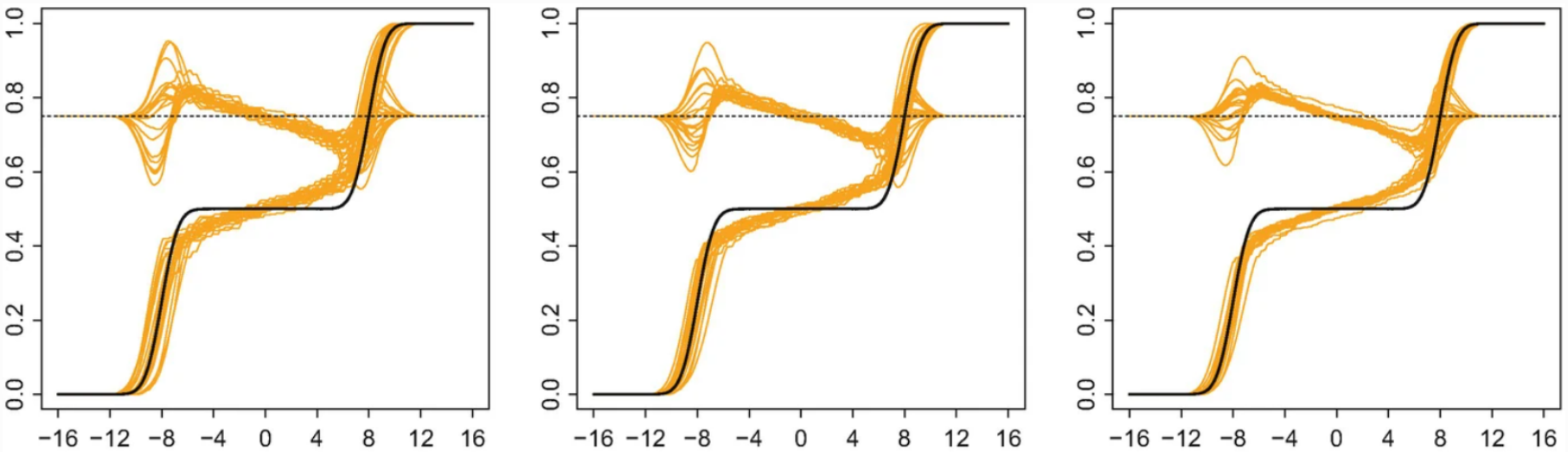


Sampling variation of the regularised Fréchet–Wasserstein mean $\widehat{\lambda}_n$ and the true mean measure λ for 20 independent replications of the experiment, with $\epsilon = 0$ and $n = 30$. Left: $\tau = 43$; middle: $\tau = 93$; right: $\tau = 143$. For ease of comparison, all three panels include residual curves centred at $y = 3/4$

Bimodal Cox Processes

Fig. 4.11

From: Phase Variation and Fréchet Means



Sampling variation of the regularised Fréchet–Wasserstein mean $\widehat{\lambda}_n$ and the true mean measure λ for 20 independent replications of the experiment, with $\epsilon = 0$ and $\tau = 93$. Left: $n = 30$; middle: $n = 50$; right: $n = 70$. For ease of comparison, all three panels include residual curves centred at $y = 3/4$.

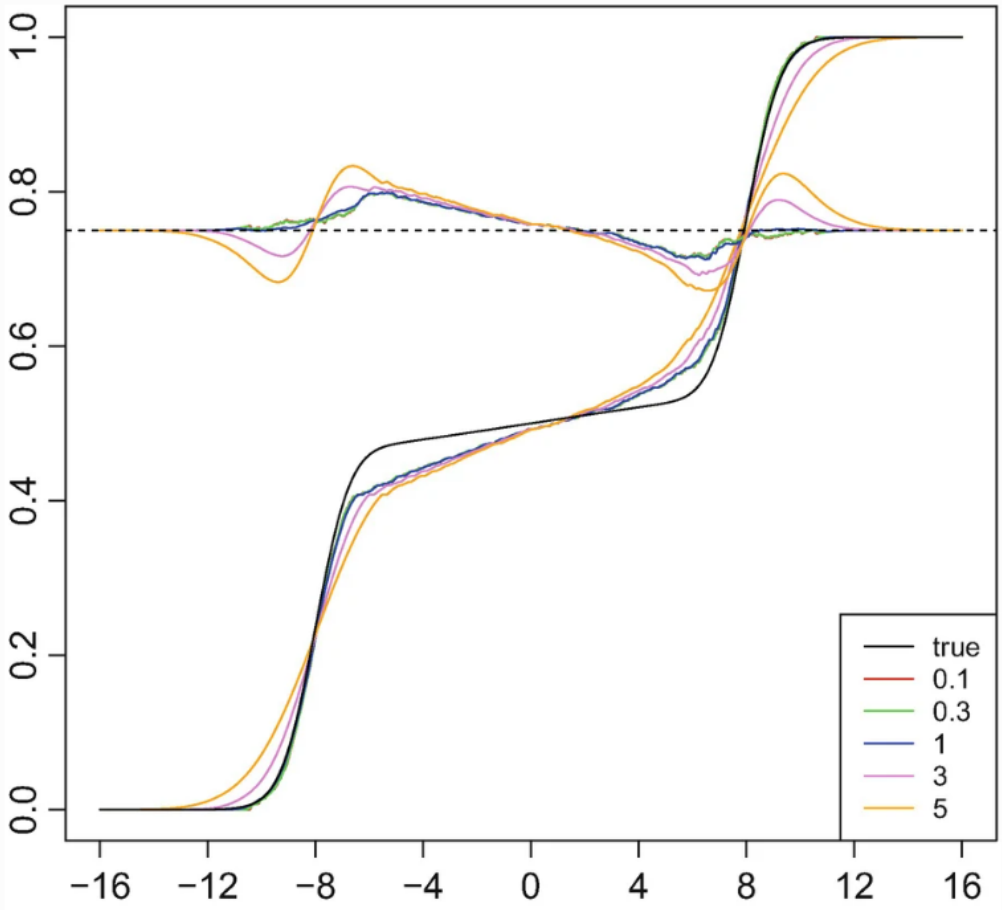
Effect of the Smoothing Parameter

In order to work with measures of strictly positive density, the observed point patterns have been smoothed using a kernel function. This necessarily incurs an additional bias that depends on the bandwidth σ_i. The asymptotics (Theorem 4.4.1) guarantee the consistency of the estimators, in particular the regularised Fréchet–Wasserstein estimator $\hat{\lambda}_n$, provided that $\max_{i=1}^{n} \sigma_i \to 0$. In our simulations, we choose σ_i in a data-driven way by employing unbiased cross validation. To gauge for the effect of the smoothing, we carry out the same estimation procedure but with σ_i multiplied by a parameter s. Figure 4.12 presents the distribution function of $\hat{\lambda}_n$ as a function of s. Interestingly, the curves are nearly identical as long as $s \leq 1$, whereas when $s > 1$, the bias becomes more substantial.

Effect of the Smoothing Parameter

Fig. 4.12

From: Phase Variation and Fréchet Means



Regularised Fréchet–Wasserstein mean as a function of the smoothing parameter multiplier s, including residual curves. Here, $n = 30$ and $\tau = 143$

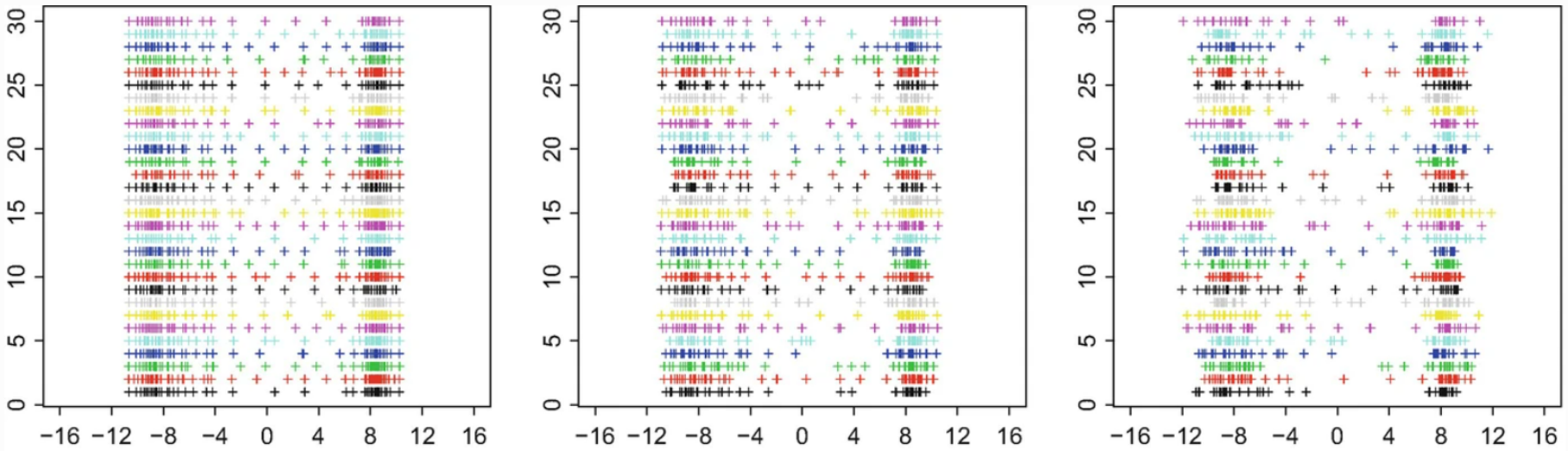
Effect of the Smoothing Parameter

These findings are reaffirmed in Fig. 4.13 that show the registered point processes again as a function of s. We see that only minor differences are present as s varies from 0.1 to 1, for example, in the grey (8), black (17), and green (19) processes. When $s = 3$, the distortion becomes quite more substantial. This phenomenon repeats itself across all combinations of n, τ, and s tested.

Effect of the Smoothing Parameter

Fig. 4.13

From: Phase Variation and Fréchet Means



Registered point processes as a function of the smoothing parameter multiplier s. Left: $s = 0.1$; middle: $s = 1$; right: $s = 3$. Here, $n = 30$ and $\tau = 43$