University of Alberta

HOMEOMORPHIC METHODS FOR SEPARATING MEASURES IN HISTORICAL FILTERING

by

David Ballantyne ©

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of
Master of Science

in

Mathematics

Department of Mathematical and Statistical Sciences
Edmonton, Alberta
Fall 2005

0-494-09119-3

University of Alberta

Library Release Form

Name of Author: David Ballantyne
Title of Thesis: Homeomorphic Methods for Separating Measures in Historical Filtering
Degree: Master of Science
Year This Degree Granted: 2005

Department of Mathematical and Statistical Sciences

Date: October 3rd, 2005

To Marcus and Reem,
without whom this thesis could not have been completed,
Kim and Maciek,
without whom this thesis could not have been continued,
and Judy and Bill,
without whom this thesis could not have been begun.

ABSTRACT

This work poses a stochastic filtering problem in which there are discontinuities in the signal evolution and in which the form of the observations is such that solutions require *historical* or *path-space* filtering. In combination, these elements exceed the theory from Ethier & Kurtz (1986) and Bhatt & Karandikar (1993). Some explanation is given for terms and techniques in the theories of weak convergence and of collections of functionals on topological spaces, and then the application of homeomorphic methods yields a separation result. With this result, measures on the space of signal paths of the type under consideration can be mutually distinguished. This is a step towards establishing uniqueness results required to prove convergence of approximations to equations in a filtering theory and, ultimately, a complete and applicable theory of approximations to the optimal filter for this class of problems.

ACKNOWLEDGEMENT

The author gratefully acknowledges assistance and funding from the Natural Sciences and Engineering Research Council of Canada, the Alberta Informatics Circle of Research Excellence, and the Prediction in Interacting Systems Centre of Excellence within the Mathematics of Information Technology and Complex Systems network, as well as the Department of Mathematical and Statistical Sciences at the University of Alberta.

Most especially, the author thanks Michael Kouritzin for his supreme efforts, of many kinds, to encourage completion of the thesis.

Table of Contents

List of Figures

Chapter 1

Introduction

1.1 Motivating problem

Consider the problem of tracking a submarine. The submarine follows some *path*, over time, through the ocean. Its position x can be taken as a function of time t through a function $x(t)$ or x_t defined for all $t \in [0, \infty)$. However, the submarine can take any one of a number of paths, and it is not known beforehand which path the submarine will take.

A method to handle this situation is to model the submarine by a random process $X_t(\omega) = X(t, \omega)$. That is, for each $\omega \in \Omega$, $X_t(\omega)$ is a path through the ocean. The probabilities of the various paths are then determined by the probability space $(\Omega, \mathcal{F}, P)$ on which the random process is defined. It is assumed that the signal process is *Markov*, meaning that the probabilities of future motion are determined solely by the current position of the signal, without needing to refer to what path the signal took to enter its current state.

For example, suppose that the submarine position in three-space at any time is given by

$$X_t = \begin{bmatrix} x_t \\ y_t \\ z_t \end{bmatrix}, \tag{1.1}$$

or, if the additional elements of lateral orientation θ and velocity v are included,

$$X_t = \begin{bmatrix} x_t \\ y_t \\ z_t \\ \theta_t \\ v_t \end{bmatrix}. \tag{1.2}$$

Here there must be constraints on the variables in order to properly model the physical situation, say, $-100 \leq z \leq 0$, $0 \leq \theta < 2\pi$, and $0 \leq v \leq 10$. That is, the submarine stays in the ocean between a depth of 100 units and the water surface, its orientation is given in radians, and it moves between a speed of 0 and 10 units of distance per unit time.

This says nothing so far about how the submarine moves. Since the path that it takes is to be random, there must be a stochastic definition of its progress in time. The submarine motion can be modeled as an Itô stochastic differential equation (SDE) as follows:

$$dX_t = \begin{bmatrix} dx_t = v_t \cos(\theta_t)dt + dW_t^x \\ dy_t = v_t \sin(\theta_t)dt + dW_t^y \\ dz_t = (-50 - z_t)dt + \sqrt{(0 - z_t)(z_t - (-100))}dW_t^z \\ d\theta_t = \frac{1}{10}dW_t^\theta \\ dv_t = (5 - v_t)dt + \sqrt{(10 - v_t)(v_t - 0)}dW_t^v \end{bmatrix}. \tag{1.3}$$

Here, each of W^x, W^y, W^z, W^θ, and W^v are independent Brownian motions that drive the overall random process X. Intuitively, in a small time period dt, each submarine coordinate will move a small deterministic amount given by the function beside the factor dt in its equation and an additional small random amount given by the function beside the factor dW_t. It is understood that the values for θ are taken mod 2π, and the equations for change in depth dz_t and change in velocity dv_t are formed to allow each value to wander only within the allowable range while drifting back to the central value within the range. If the underlying Brownian motions are all almost-surely continuous versions then, by the theory of Itô integration, the path of the submarine defined by this SDE is almost-surely continuous.

By this equation, the submarine will travel in the (x, y) plane given by its orientation and velocity with some small diffusion, will maintain its orientation subject to some diffusion, and will wander randomly within the possible ranges of depths and velocities. There must also be some distribution X_0 on S for the initial submarine position to complete the definition of the stochastic process $X = \{X_t : t \in [0, \infty)\}$.

The space S of possible states of the submarine is then a subset of $\mathbb{R}^5$, that is, $X_t \in S \subset \mathbb{R}^5$ for all $t \in [0, \infty)$. By taking d to be the Euclidean metric on S, (S, d) is made into a complete, separable metric space. Then the space of all (almost-surely) possible paths of the submarine lies in the more complicated set $C_S[0, \infty)$, the space of continuous paths in the space (S, d). Handling probability measures defined on $C_S[0, \infty)$ may require significantly more mathematical machinery than is needed for such measures on (S, d) itself. The details about function spaces such as $C_S[0, \infty)$ are given in Section 2.2.

A practical filtering problem develops if it is assumed that there is imperfect information regarding the state $\{X_t\}_{t \geq 0}$ of the submarine to be tracked. In the language of filtering, the signal, X, is to be tracked given only the information from the observations $Y = \{Y_s\}_{s \geq 0}$ for times s up to some time t. The observation process Y is taken to be functionally but stochastically determined by the state of the process X, so that a perfect reconstruction of the state of X is not possible and instead only a distribution for the state of X_t over the space S can be determined from $\{Y_s\}_{s \leq t}$.

As an example, the observation could be the approximate distance from the submarine to each of some number M of surface vessels, as given by noisy sonar readings. (In this case, the single "observation" is used to describe the reception of more than one reading. That is, the observation could be a vector of read values.) In practical problems, usually the observations are taken to be discrete in time, so that $Y_t = Y_{t_k}$ for $t \in [t_k, t_{k+1})$, or more often by simply defining the observations as a sequence $Y = \{Y_k\}_{k=1}^{\infty}$ only at discrete times t_k for $k \in \{1, 2, \ldots\}$.

So, suppose that there are M surface vessels that follow the deterministic paths $\{(x_t^j, y_t^j)\}_{t \geq 0}$, $1 \leq j \leq M$ through time along the ocean surface. Take

d_3 to be the Euclidean metric on $\mathbb{R}^3$, and let the M independent random sequences V_k^j, $1 \leq j \leq M$ be the sequentially independent noise in the sonar values each given by a normal distribution $\mathcal{N}(0, \sigma)$. Then the observations are given by

$$Y_k = \begin{bmatrix} Y_k^1 \\ Y_k^2 \\ \vdots \\ Y_k^M \end{bmatrix} = \begin{bmatrix} d_3((x_{t_k}, y_{t_k}, z_{t_k}), (x_{t_k}^1, y_{t_k}^1, 0)) + V_k^1 \\ d_3((x_{t_k}, y_{t_k}, z_{t_k}), (x_{t_k}^2, y_{t_k}^2, 0)) + V_k^2 \\ \vdots \\ d_3((x_{t_k}, y_{t_k}, z_{t_k}), (x_{t_k}^M, y_{t_k}^M, 0)) + V_k^M \end{bmatrix}. \tag{1.4}$$

The problem is then to find an estimate for the state of the signal given the information from the observations. Precisely, the requirement is to determine

$$P(X_{t_k} \in A \mid \sigma\{Y_j : j \leq k\}) \tag{1.5}$$

for Borel subsets A of S and for each time t_k in a practical manner that can be implemented on a computer. Note that the SDE defining $\{X_t\}_{t \geq 0}$ is nonlinear, so the linear filtering theory of Kalman and Bucy [12] will not suffice.

For ease, define the σ-algebra $\mathcal{F}_k^Y \doteq \sigma\{Y_j : j \leq k\}$. (This is the *information contained in the observations up to time* t_k.) Then, define $C_b(S)$ to be the *bounded, continuous functionals* $\varphi : S \to \mathbb{R}$. (Full descriptions of collections of functionals like $C_b(S)$ are provided in Section 1.2, following.) Take the (operator, domain) pair $(\mathcal{L}, \mathcal{D}(\mathcal{L}))$ to be the *weak generator* for the Markov signal process $\{X_t\}_{t \geq 0}$. The theory of such generators will not be dealt with in this thesis, so for novice readers its use in the following development can be taken as the outer framework for the elaboration of the thesis, a framework which need not itself be precisely understood.

In order to compute $P(X_{t_k} \in A \mid \mathcal{F}_k^Y)$, methods must be developed which approximate the solution to the true filtering equations, where the approximation is in the sense of *weak convergence*, which is described in Section 2.1. However, weak convergence is defined in terms of bounded continuous functionals. So, rather than operating on probabilities that the signal is within a set A, this is expressed instead as

$$E[1_A(X_{t_k}) \mid \mathcal{F}_k^Y] = P(X_{t_k} \in A \mid \mathcal{F}_k^Y) \tag{1.6}$$

and then this is approximated by

$$E[\varphi(X_{t_k}) \mid \mathcal{F}_k^Y] \tag{1.7}$$

for functionals $\varphi \in C_b(S)$, since the collection of bounded continuous functionals can approximate (necessarily bounded) indicator functions 1_A on the Borel sets.

Now, filtering equations that involve the evolution of $\varphi(X_t)$ for $t \in [0, \infty)$ as the process $\{X_t\}_{t\geq 0}$ evolves over time can only be defined if φ is also in the domain of the weak generator of the signal $\{X_t\}_{t\geq 0}$. Satisfying this second condition requires a subcollection $\mathcal{M} \subset C_b(S)$ which is rich enough to approximate the functionals in $C_b(S)$ and which is also a subset of the domain of the weak generator for X.

It is necessary that the collection $\mathcal{M}$ be at least rich enough to distinguish probability measures from each other, in order to obtain results on the uniqueness of solutions to various applicable filtering equations. By Ethier and Kurtz (1986) [9] Theorem 3.4.5(a), since (S, d) is a complete, separable metric space, as long as $\mathcal{M} \subset C_b(S)$, $\mathcal{M}$ *separates points*, and $\mathcal{M}$ is an *algebra*, then the collection $\mathcal{M}$ can *separate* probability measures, that is, distinguish one from another. (The precise definition of *separates points* is given in Definition 2.33, *algebra* in Definition 3.2, and *separating collection* in Definition 2.7.)

In this case, it can simply be required of the signal that the domain of its weak generator, $\mathcal{D}(\mathcal{L})$, be contained in the space of bounded continuous functionals on S, that $\mathcal{D}(\mathcal{L})$ separate points, and that $\mathcal{D}(\mathcal{L})$ form an algebra. These conditions turn out not to be onerous. Then, for such signals, the collection $\{\varphi : \varphi \in \mathcal{D}(\mathcal{L})\}$ separates probability measures, and the process of defining and approximating filtering equations can proceed. (See, for example, Zakai (1969) [19] for the continuation of the process of defining the filtering equations.) Approximations that provably converge to the filtering equations and can be implemented in computer software are candidate solutions to the practical problem.

Figure 1.1 and Figure 1.2 are pictures of an example simulation of a submarine tracking problem similar to the one described above. Each figure is composed of three frames, the leftmost of which displays the actual true posi-

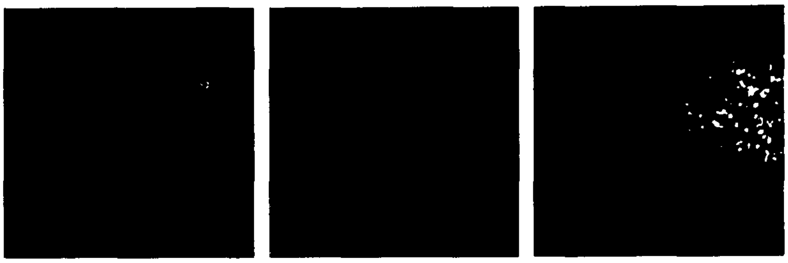

Figure 1.1: Example filtering simulation near $t = 0$.

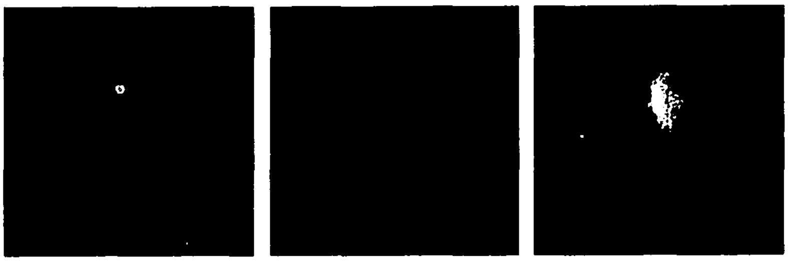

Figure 1.2: Example filtering simulation at a later time.

tions in the (x, y) plane of the submarine (white circle) and the four tracking surface vessels (coloured squares), the central of which records the observations that have been received by each tracking vessel, and the rightmost of which displays a visualisation of the filter estimate of the (x, y) position of the submarine, given the information from the observations up to the current time. The estimate has more probability mass in sections which are more heavily covered in white, while the red circle is the true location of the submarine and is only included for reference purposes. The filter estimate is constructed from a finite approximation method which provably converges to the optimal filter in the sense of weak convergence. The first figure depicts the simulation at the early stages, when few observations have been received, and the second figure depicts a later stage in the simulation when the observations have provided enough information that the track has become more precise. It is evident that this is a workable solution to the tracking problem.

Consider, now, a slightly different and more difficult to analyse problem. In this case, the submarine is controlled purposely to avoid detection. To do this, the submarine captain takes sharp maneuvers from one set of control settings to another at random times. For example, the captain may suddenly set a new course, speed, and depth through the submarine control mechanism, and the submarine, while not perfectly responsive, will begin to drift towards these new settings. Since it is unknown beforehand which maneuvers the captain will take, they will also be modeled as part of the stochastic process that is to be filtered.

Let the new stochastic process $\{X_t\}_{t\geq 0}$ reside in a new state space given by

$$X_t = \begin{bmatrix} x_t \\ y_t \\ z_t \\ \theta_t \\ v_t \\ \widehat{z}_t \\ \widehat{\theta}_t \\ \widehat{v}_t \end{bmatrix}, \tag{1.8}$$

where $\widehat{z}_t$, $\widehat{\theta}_t$, and $\widehat{v}_t$ are control settings that are constrained to lie within the open interior of the range of each corresponding state coordinate. Then, define the stochastic evolution by the SDE

$$dX_t = \begin{bmatrix} dx_t = v_t \cos(\theta_t)dt + dW_t^x \\ dy_t = v_t \sin(\theta_t)dt + dW_t^y \\ dz_t = (\widehat{z}_t - z_t)dt + \sqrt{(0 - z_t)(z_t - (-100))}dW_t^z \\ d\theta_t = \frac{1}{2}([(\widehat{\theta}_t - \theta_t + \pi) \bmod 2\pi] - \pi)dt + \frac{1}{10}dW_t^\theta \\ dv_t = (\widehat{v}_t - v_t)dt + \sqrt{(10 - v_t)(v_t - 0)}dW_t^v \end{bmatrix} \tag{1.9}$$

for the first five state coordintates, and let the other three coordinates switch as defined by a continuous state-space Markov chain with rate λ in which transitions, when they occur, are equally likely to bring the chain into any of the other possible states. That is, at transitions, the chain will enter a new state selected uniformly from among the three-dimensional domain $(-100, 0) \times [0, 2\pi) \times (0, 10)$ representing the new submarine depth ($\widehat{z}$), heading ($\widehat{\theta}$), and speed ($\widehat{v}$) control settings.

This new signal resides in a new state space $S \subset \mathbb{R}^8$ and, more importantly, it now has discontinuous paths even if the underlying Brownian motions are continuous. The reason, of course, is because of the jumps at the times at which the control settings are switched. Paths of this new signal lie in the function space $D_S[0,\infty)$ rather than in $C_S[0,\infty)$, a space that is also described in Section 2.2. An example problem in this class, where the signal model switches based on a Markov chain, is described in Ballantyne, Chan, and Kouritzin (2000) [1].

As an additional difficulty, the observations will also be modified so that they depend on the past path of the submarine. Perhaps the sonar soundings differ depending on what exact path the submarine took through the layers of ocean water, or the instrumentation can pick up information from the wake of the submarine. As a practical matter, sound waves will always take some time to travel from the submarine to the tracking vessels, and this physical delay alone means that reference to some amount of data on the historical position of the signal is unavoidable. This more realistic situation could be modeled, in a possible example, by observations of the following form. First, define $\overline{X}_t \doteq (x_t, y_t, z_t)$ to be the three-space position of the submarine at time t and define $\overline{x}_t^j \doteq (x_t^j, y_t^j, 0)$ to be the surface position of the j^{th} tracking vessel at time t. Then the observations Y_k are defined by

$$Y_k^j = d_3(\overline{X}_{t_k-3\delta}, \overline{x}_{t_k}^j) + d_3(\overline{X}_{t_k-2\delta}, \overline{x}_{t_k}^j) + d_3(\overline{X}_{t_k-\delta}, \overline{x}_{t_k}^j) + V_k^j \tag{1.10}$$

for $1 \leq j \leq M$, $k \geq 1$, and some δ small enough that $3\delta < t_1$. An example filtering problem with observations that depend on the historical path of the signal is described in Kouritzin et al. (2005) [14].

The goal is the same in this new situation, that is, to evaluate

$$P(X_{t_k} \in A \mid \mathcal{F}_k^Y) \tag{1.11}$$

for Borel sets A and for each time t_k. As in the previous problem, this is expressed as expectations of indicator functions. However, because of the new form of the observations, it is necessary to take account of the past path of the signal in the filtering equations. To do so, a new path-space variant of the weak generator $(\mathcal{L}_{[0,t]}, \mathcal{D}(\mathcal{L}_{[0,t]}))$ is defined for a process $t \to X_{[0,t]} \in D_S[0,\infty)$

that describes the entire path of the signal up to that time at each given point in time t. Necessarily, $X_{[0,t]}$ is Markov.

Functionals defined on path-space are of the form $\Phi : D_S[0,\infty) \to \mathbb{R}$, and could be used to approximate the expectations of indicator functions as before. However, the domain of the path-space generator cannot consist solely of bounded, continous functionals. They are, in general, *bounded and measurable* instead, that is, elements of $B(D_S[0,\infty))$, as will be explained in Section 2.2 in the discussion after Proposition 2.32. Because of this, a new collection of approximating functionals is required that are only bounded and measurable, not bounded and continuous, and can approximate indicator functions on the set $D_S[0,\infty)$.

A candidate approximating collection is defined as follows. Let $\{I_m\}_{m\geq 1}$, where $I_m = \{t_1,\ldots,t_m\} \subset [0,\infty)$, be a *dense timepoint mesh* (as given in Definition 3.15) and let π_t be the projection function from $D_S[0,\infty)$ to S at time t given by $\pi_t(X) = X_t$. Then, define the functionals $\Phi_{I_m} : D_S[0,\infty) \to \mathbb{R}$ by

$$\Phi_{I_m}(X) = \varphi_1(\pi_{t_1}(X)) \cdots \varphi_m(\pi_{t_m}(X)) \tag{1.12}$$

for some functionals $\varphi_i \in \mathcal{D}(\mathcal{L})$, where $\mathcal{D}(\mathcal{L})$ is the domain of the non-path-space operator for the signal X. Next, define $\mathcal{D}(\mathcal{L}_{[0,t]})$ to be the linear span of functionals of the type Φ_{I_m} and take $\mathcal{L}_{[0,t]}$ to be an appropriate set of operators on $B(D_S[0,\infty))$ defined on $\mathcal{D}(\mathcal{L}_{[0,t]})$. Again, a subcollection $\mathcal{M} \subset B(D_S[0,\infty))$ will be required that is rich enough to separate probability measures.

Since it is no longer the case that $\mathcal{M} \subset C_b(D_S[0,\infty))$, Theorem 3.4.5 from [9] can no longer be used. Some new theory needs to be developed, potentially with new requirements on the weak generator $(\mathcal{L}, \mathcal{D}(\mathcal{L}))$ of the signal that will ensure properties of $(\mathcal{L}_{[0,t]}, \mathcal{D}(\mathcal{L}_{[0,t]}))$ such that probability measures on paths can be separated.

1.2 Notation and preliminaries

The set $\mathbb{N}$ will always be taken to include zero, that is, $\mathbb{N} = \{0, 1, 2, \ldots\}$. When zero is to be excluded, the notation is $\mathbb{N}^+ = \{1, 2, \ldots\}$.

The pair $(S, \mathcal{T}_S)$ will be used to denote a topological space, while (E, d_E) will denote a metric space. For metric spaces, $\mathcal{T}_E$ is the topology generated by (E, d_E). The notation $\mathcal{B}(S)$ refers to the Borel σ-algebra generated by $\mathcal{T}_S$ on any topological space S.

Definition 1.1. Suppose that (E, d_E) is a metric space. Then the m^{th} *product metric space* E^m is the collection $\{(x_1, \dots, x_m) : x_i \in E \text{ for } i = 1, \dots, m\}$. The metric on this collection is

$$d((x_1, \dots, x_m), (y_1, \dots, y_m)) = \sum_{i=1}^{m} d_E(x_i, y_i). \tag{1.13}$$

It is possible to put a topology on a space which is constructed as the arbitrary product of another topological space, and this underlying space need not even be a metric space. This canonical topology, called the *product topology*, will be assumed in product spaces (especially $\mathbb{R}^\infty$) unless otherwise noted.

Definition 1.2. If $(S, \mathcal{T}_S)$ is a topological space and J is some index set, the *product topology* on S^J, denoted $\mathcal{T}_{S^J}$, is defined to be the topology generated by the basis

$$\mathbb{B}_{S^J} = \left\{ \left\{ x \in S^J : x_{j_i} \in U_{j_i},\ 1 \le i \le m \right\} : m \in \mathbb{N}^+,\ j_1, \dots, j_m \in J \right\}. \tag{1.14}$$

This definition is compatible with the definition of product metric spaces in that the product metric defined in equation (1.13) generates the product topology on $(E, d_E)^m$. In addition, even for a countably infinite product, if the underlying space E is a metric space then so is the product space E^J.

Proposition 1.3. *Suppose S^J is a product space on the topological space $(S, \mathcal{T}_S)$. Then the net $x_\alpha \to x$ in S^J if and only if $x_{\alpha,j} \to x_j$ in S for all $j \in J$.*

Proof: Suppose $x_\alpha \to x$ in S^J. Then for any $j \in J$ and any neighbourhood U_j of x_j in S, the set $V = \left\{ x \in S^J : x_j \in U_j \right\}$ is a neighbourhood of x in the product topology, so there exists an $\mathcal{A}$ such that $\alpha \succeq \mathcal{A} \Rightarrow x_\alpha \in V \Rightarrow x_{\alpha,j} \in U_j$, and thus $x_{\alpha,j} \to x_j$.

Now suppose $x_{\alpha,j} \to x_j$ for all $j \in J$ and let O_x be a basic neighbourhood of x in S^J. Then $O_x = \left\{ x \in S^J : x_{j_i} \in U_{j_i},\ 1 \le i \le m \right\}$ for some m indices

$j_1, \dots j_m \in J$. Let $\mathcal{A}_i$ be such that $\alpha \succeq \mathcal{A}_i \Rightarrow x_{\alpha,j_i} \in U_{j_i}$, $1 \le i \le m$. Then there exists an $\mathcal{A}$ such that $\mathcal{A} \succeq \mathcal{A}_i$, $1 \le i \le m$, so that $\alpha \succeq \mathcal{A} \Rightarrow x_\alpha \in O_x$, and $x_\alpha \to x$ in S^J. ■

For this reason, the product topology is often called the topology of pointwise convergence.

Basic to the analysis of probabilities on topological spaces is the consideration of functionals mapping these spaces into the real line. Recall the following definitions for functions and functionals on a metric space E.

Definition 1.4. A function $f : E \to S$ is *continuous* with respect to the topology $\mathcal{T}_S$ on S if, for each set $U \in \mathcal{T}_S$, $f^{-1}(U) \in \mathcal{T}_E$. For a *functional* $f : E \to \mathbb{R}$, continuity refers to the topology generated by the Euclidean metric on $\mathbb{R}$; that is, a functional f is continuous if for all U which are open in $\mathbb{R}$ under the standard topology $\mathcal{T}_{\mathbb{R}}$, $f^{-1}(U) \in \mathcal{T}_E$.

Definition 1.5. A function $f : E \to S$ is *measurable* with respect to the σ-algebra $\mathcal{S}$ on S if, for each set $A \in \mathcal{S}$, $f^{-1}(A) \in \mathcal{B}(E)$. For a functional $f : E \to \mathbb{R}$, measurability refers to the Borel σ-algebra on $\mathbb{R}$; that is, a functional f is measurable if for all $A \in \mathcal{B}(\mathbb{R})$, $f^{-1}(A) \in \mathcal{B}(E)$.

Note that all continuous functions are (Borel-)measurable.

Definition 1.6. A function $f : E \to S$ is *bounded* with respect to the metric d_S on S if there exists an $M > 0$ for which $d_S(f(x), f(y)) < M \ \forall x, y \in E$. For a functional $f : E \to \mathbb{R}$, boundedness refers to the Euclidean metric on $\mathbb{R}$.

Important classes of functionals are given specific identifications.

Definition 1.7. The collections $M(E)$, $B(E)$, $C(E)$, and $C_b(E)$ denote respectively the measurable, bounded and measurable, continuous, and bounded and continuous functionals $f : E \to \mathbb{R}$.

Recall also that the *supremum norm* $\|\cdot\|_\infty$, defined by

$$\|f\|_\infty \doteq \sup_{x \in E} |f(x)|, \tag{1.15}$$

is a well-defined (finite) norm on $B(E)$ and $C_b(E)$. Under this norm, $B(E)$ and $C_b(E)$ are complete, that is, $(B(E), \|\cdot\|_\infty)$ and $(C_b(E), \|\cdot\|_\infty)$ are Banach spaces.

Once product spaces have been formed, it is often important to examine one individual component of an element in the product. Sometimes, a finite number of components need to be referenced. In these cases, a notion of *projection* provides access to the element or elements of the basic space.

Definition 1.8. If S^J is a product space on the topological space $(S, \mathcal{T}_S)$, then the *projection* $\pi_j : S^J \to S$ for $j \in J$ is given by $\pi_j(x) = x_j$ for all $x \in S^J$. For any finite set of indices $j_1, \ldots, j_m$ in J, the *finite-dimensional projection* $\pi_{j_1,\ldots,j_m} : S^J \to S^m$ is defined to be the element $(x_{j_1}, \ldots, x_{j_m})$ for all $x \in S^J$.

Projections from a product space with the product topology have extremely nice topological properties.

Proposition 1.9. *The finite-dimensional projection* $\pi_{j_1,\ldots,j_m} : S^J \to S^m$ *is both continuous and measurable in the product topology.*

Proof: If U is an open subset of S^m, say $U = \{x \in S^m : x_i \in U_i,\ 1 \leq i \leq m\}$ for some $m \in \mathbb{N}^+$ and $U_1, \ldots, U_m \in \mathcal{T}_S$, then

$$\pi_{j_1,\ldots,j_m}^{-1}(U) = \left\{x \in S^J : x_{j_i} \in U_i,\ 1 \leq i \leq m\right\} \tag{1.16}$$

is a basic set in the product topology, so that $\pi_{j_1,\ldots,j_m}$ is continuous. Since $\pi_{j_1,\ldots,j_m}$ is continuous, it is also measurable. ■

By taking $m = 1$ in the above proposition, single-dimensional projections π_j are also proved to be continuous and measurable.

The next proposition introduces a mathematical object that will have a central role in this thesis.

Proposition 1.10. *If* $g_j : S \to \mathbb{R}$ *for* $j \in J$, *and* $\Gamma : S \to \mathbb{R}^J$ *is defined by* $\Gamma(x) = \prod_{j \in J} g_j(x)$, *then*

i) Γ *is continuous if and only if* g_j *is continuous for each* $j \in J$.

ii) Γ *is measurable if and only if* g_j *is measurable for each* $j \in J$.

Proof: First, both forward implications follow from the continuity or measurability of the composition $\pi_j \circ \Gamma(x) = g_j(x)$.

For the reverse implication in the continuous case, assume that all of the g_j are continuous, that is, that for all $j \in J$, $V \in \mathcal{T}_{\mathbb{R}} \Rightarrow g_j^{-1}(V) \in \mathcal{T}_S$. Then take any $U \in \mathbb{B}_{\mathbb{R}^J}$, the standard basis for the product topology in $\mathbb{R}^J$. The set U can be expressed as $\{x \in \mathbb{R}^J : x_{j_i} \in U_i,\ 1 \leq i \leq m\}$ for some $m \in \mathbb{N}^+$, $j_1, \ldots, j_m \in J$, and $U_1, \ldots, U_m \in \mathcal{T}_{\mathbb{R}}$, so $\Gamma^{-1}(U) = \bigcap_{i=1}^{m} g_{j_i}^{-1}(U_i) \in \mathcal{T}_S$.

If each of the g_j is measurable, then to show that Γ is measurable, it is enough to show that inverse images of the measurable rectangles are in $\mathcal{B}(S)$, since the measurable rectangles generate $\mathcal{B}(\mathbb{R}^J)$. Let

$$A = \{x \in \mathbb{R}^J : x_{j_i} \in A_i,\ 1 \leq i \leq m\} \tag{1.17}$$

for some $m \in \mathbb{N}^+$, $j_1, \ldots, j_m \in J$, and $A_1, \ldots, A_m \in \mathcal{B}(\mathbb{R})$ be such a measurable rectangle. Then, similarly to the above argument, $B \in \mathcal{B}(\mathbb{R}) \Rightarrow g_j^{-1}(B) \in \mathcal{B}(S)$ for each $j \in J$, and so $\Gamma^{-1}(A) = \bigcap_{i=1}^{m} g_{j_i}^{-1}(A_i) \in \mathcal{B}(S)$. ■

For now, note that if each of the component functions g_j of Γ are continuous, and if the inverse of the function Γ is defined and is continuous, then Γ is a *homeomorphism* from S to $\Gamma(S)$, that is, a continuous mapping with a continuous inverse.

Chapter 2

Convergence and Homeomorphic Techniques

2.1 Weak convergence

One of the core notions of modern filtering theory is the concept of an *approximate filter*, that is, a class of filters indexed by n such that as n increases, the approximate filter associated with n converges, in some sense, to the optimal filter. Approximate filters are useful in situations in which the optimal filter cannot be readily implemented in a computer algorithm. The notion of convergence used, however, needs not to be so strong that no computable task can converge. The *weak convergence* of probability measures is an appropriate type of convergence for this requirement, and weak convergence therefore plays a central role in filtering theory.

Definition 2.1. If P is a probability measure on the topological space $(S, \mathcal{T}_S)$ (with Borel sets $\mathcal{B}(S)$) and $\{P_n\}$ is a sequence of probability measures on the same space, then P_n *converges weakly* to P, written $\operatorname*{wk-lim}_{n\to\infty} P_n = P$ or $P_n \xrightarrow[n\to\infty]{\text{wk}} P$, if $P_n f \to Pf$ for all $f \in C_b(S)$. Here

$$Pf \doteq \int_S f\,dP \doteq \int_S f(x)P(dx) \tag{2.1}$$

for measures P and functionals f. It is also said that P is the *weak limit* of $\{P_n\}$. In the literature, weak convergence is often denoted by $P_n \Rightarrow_{n\to\infty} P$.

Recall that the standard notation is to supress the dependence of the Borel sets on the topology of the space, so that $\mathcal{B}(S)$ will be the Borel sets generated by the topology $\mathcal{T}_S$ of the topological space $(S, \mathcal{T}_S)$, unless otherwise indicated.

To deal readily with the probability measures, it will be helpful to select those which are from a nice class.

Definition 2.2. A probability measure P on a topological space $(S, \mathcal{T}_S)$ is *regular* if for every $A \in \mathcal{B}(S)$ and every $\varepsilon > 0$, there exist a closed set F and an open set G such that $F \subset A \subset G$ and $P(G - F) < \varepsilon$,

or equivalently,

a probability measure P on a topological space $(S, \mathcal{T}_S)$ is *regular* if, for every $A \in \mathcal{B}(S)$,

$$P(A) = \inf\{P(G) : G \supset A,\ G \text{ open}\} \tag{2.2}$$

and

$$P(A) = \sup\{P(F) : F \subset A,\ F \text{ closed}\}. \tag{2.3}$$

It is good that regular probability measures are the norm, especially in common topological spaces, as in the following result.

Proposition 2.3. *If E is a metric space, then each probability measure on E is regular.*

The proof, stated in different language, is in Billingsley (1999) [5] as Theorem 1.1.

Recall that probability measures on a topological space $(S, \mathcal{T}_S)$ are defined on the Borel sets of S, so that for measures P and Q, $P = Q$ means that, for all $B \in \mathcal{B}(S)$, $P(B) = Q(B)$. Since all of the probability measures required for the application will be regular, the following proposition provides a useful characterization of the equality of such measures through the use of functionals on the space. This will be the basis for all later work to differentiate probability measures.

Proposition 2.4. *If P and Q are regular probability measures on a topological space $(S, \mathcal{T}_S)$ and $Qf = Pf$ for all $f \in C_b(S)$, then $P = Q$.*

The proof is in Theorem 1.2 of Billingsley (1999) [5].

Definition 2.5. A set A in $\mathcal{B}(S)$ is a *P-continuity set* if $P(\partial A) = 0$.

Note that ∂A is closed and so it is Borel-measurable. Also, $\partial S = \varnothing$, so for all probability measures P, S is a P-continuity set. For metric spaces, the following theorem relates the definitional characterisation of weak convergence using functionals with characterisations using measurable sets.

Theorem 2.6 (Alexandroff's "Portmanteau" Theorem). *If P_n and P are probability measures on the metric space (E, d_E), then the following are equivalent:*

i) $\operatorname*{wk-lim}_{n\to\infty} P_n = P$.

ii) $P_n f \to Pf$ *for all uniformly continuous* $f \in C_b(E)$.

iii) $\limsup_{n\to\infty} P_n F \leq PF$ *for all closed sets* F.

iv) $\liminf_{n\to\infty} P_n G \geq PG$ *for all open sets* G.

v) $P_n A \to PA$ *for all P-continuity sets* A.

This useful result is proved in Billingsley (1999) [5] as Theorem 2.1.

The vital notions of what it means to be *convergence-determining* and *separating* are made precise in the next definition.

Definition 2.7. Let $\mathcal{A}$ be a subclass of $\mathcal{B}(S)$ that includes S and let P, Q, and $\{P_n\}$ all be probability measures on S. The subclass $\mathcal{A}$ is a *separating class* if, for all P and Q,

$$PA = QA \ \ \forall A \in \mathcal{A} \ \Rightarrow \ P = Q, \tag{2.4}$$

and $\mathcal{A}$ is a *convergence-determining class* if, for all $\{P_n\}_{n=0}^{\infty}$ and P,

$$P_n A \to PA \ \ \forall P\text{-continuity sets } A \in \mathcal{A} \ \Rightarrow \ P_n \xrightarrow[n\to\infty]{\text{wk}} P. \tag{2.5}$$

Also, if $\mathcal{M}$ is a collection of functionals on S, then $\mathcal{M}$ is *separating* if, for all P and Q,

$$Pf = Qf \ \ \forall f \in \mathcal{M} \ \Rightarrow \ P = Q, \tag{2.6}$$

and $\mathcal{M}$ is *convergence-determining* if, for all $\{P_n\}_{n=0}^{\infty}$ and P,

$$P_n f \to Pf \ \ \forall f \in \mathcal{M} \ \Rightarrow \ P_n \xrightarrow[n\to\infty]{\text{wk}} P. \tag{2.7}$$

Example 2.8. By Proposition 2.3, the closed sets of a metric space are a separating class and, by an argument on pages 17 and 18 of Billingsley (1999) [5], the open sets of a separable metric space are a convergence-determining class.

An easy corollary of Proposition 2.4 is that the bounded continuous functionals on a metric space are a separating collection, and by the definition of weak convergence, this collection is also convergence-determining.

The above examples are of very large, general collections of sets and functionals that possess the qualities almost or entirely by definition. However, they do suggest a relationship between the two illustrated concepts which is stated in the following proposition.

Proposition 2.9. *If every probability measure on S is regular, then any class of sets which is convergence-determining is also separating, and any collection of functionals which is convergence-determining is also separating.*

Proof: Suppose $\mathcal{A}$ is convergence-determining and that $PA = QA$ for all $A \in \mathcal{A}$. Let $\{P_n\}_{n=0}^{\infty}$ be the sequence for which $P_n = Q$ $\forall n \in \mathbb{N}$. Then trivially $P_n A \to PA$ for all P-continuity sets in $\mathcal{A}$, and then, since $\mathcal{A}$ is convergence-determining, $P_n \xrightarrow[n\to\infty]{\text{wk}} P$. By definition, for all $f \in C_b(S)$, $P_n f \to Pf$, which means that for all $f \in C_b(S)$, $Qf = Pf$. By Proposition 2.4, $P = Q$.

The final portion of the above proof, along with the definition of weak convergence, proves the case for collections of functionals. ∎

More illuminating examples of separating and convergence-determining classes can be found among the *finite-dimensional sets.*

Definition 2.10. The *finite-dimensional sets for* $\mathbb{R}^\infty$ are defined by

$$\pi^{-1}_{k_1,\ldots,k_m} H \tag{2.8}$$

for finite-dimensional projections $\pi_{k_1,\ldots,k_m}$, $k_1, \ldots, k_m \in \mathbb{N}$, and Borel sets H in $\mathcal{B}(\mathbb{R}^m)$.

Proposition 2.11. *The finite-dimensional sets on $\mathbb{R}^\infty$ are a convergence-determining class, and the projections on $\mathbb{R}^\infty$ are a convergence-determining collection.*

This is Example 2.4 in Billingsley (1999) [5] for sets, and the extension to projections is clear.

Definition 2.12. The space $C[0,1]$ is the space of real-valued continuous functions on the set $[0,1]$. The *finite-dimensional sets for* $C[0,1]$ are defined by

$$\pi^{-1}_{t_1,\dots,t_k} H \tag{2.9}$$

for finite-dimensional projections $\pi_{t_1,\dots,t_k}$ and sets H in $\mathcal{B}(\mathbb{R}^k)$.

Proposition 2.13. *The finite-dimensional sets on $C[0,1]$ are a separating class, but are* not *a convergence-determining class, and the same holds for the collection of projections.*

For the proof, see Example 1.3 and Example 2.5 in Billingsley (1999) [5]. This is an example of a separating class (collection) that is not a convergence-determining class (collection), proving that the two concepts are not equivalent. The subtlety of weak convergence in function spaces is largely due to this discrepancy. Much of the work of Chapter 3 is to resolve this difficulty.

The final theorem in this section allows useful transformations of weak convergence from one space to another if the two are linked by a continuous function, or at least, one that is nearly continuous.

Definition 2.14. For a measurable function h, let D_h be the set of its discontinuities.

Theorem 2.15 (Mapping Theorem). *If P_n and P are probability measures on S, $P_n \xrightarrow[n\to\infty]{\text{wk}} P$, and $P(D_h) = 0$, then $P_n h^{-1} \xrightarrow[n\to\infty]{\text{wk}} Ph^{-1}$.*

This is Theorem 2.7 in Billingsley (1999) [5].

As a consequence of the Mapping Theorem, convergence of probability measures on a product space (with the product topology, as is conventional in this thesis) implies the convergence of all finite-dimensional distributions on that product space. For the space $\mathbb{R}^\infty$, the converse also holds.

Proposition 2.16. *If $\{P_n\}$ and P are probability measures on $\mathbb{R}^\infty$, then $P_n \xrightarrow[n\to\infty]{\text{wk}} P$ if and only if $P_n\pi^{-1}_{l_1,\dots,l_k} \xrightarrow[n\to\infty]{\text{wk}} P\pi^{-1}_{l_1,\dots,l_k}$ for all finite-dimensional projections $\pi_{l_1,\dots,l_k}$ from $\mathbb{R}^\infty$ to $\mathbb{R}^k$, or equivalently, if and only if the convergence holds for all finite-dimensional projections $\pi_{0,\dots,k}$ of the first $k+1$ coordinates of elements of $\mathbb{R}^\infty$.*

The proof is given in Example 2.6 in Billingsley (1999) [5], with the equivalence a consequence of Theorem 2.8(ii) from the same source. This Theorem 2.8(ii) from [5] implies that, when the space is separable, the weak convergence for single-dimensional distributions of a sequence is equivalent to weak convergence for arbitrary finite-dimensional distributions of the sequence, and also equivalent to weak convergence for all initial finite-dimensional distributions of the type $P_n\pi^{-1}_{0,\dots,k}$.

2.2 Measures on function spaces

Recall from Section 2.1 that the space $C[0,1]$ was defined to be the set of continuous real-valued functions on $[0,1]$. This definition can be extended to include functions on $[0,1]$ that take values in a metric space E.

Definition 2.17. For any metric space (E, d_E), the space $C_E[0,1]$ is defined to be the set of all continous functions on $[0,1]$ that take values in E. This space is metrized by the *uniform metric* or *supremum norm*

$$d(x,y) \doteq \sup_{t\in[0,1]} d_E(x(t), y(t)). \tag{2.10}$$

If $\{X_t,\ t \in [0,1]\}$ is a continuous random process on the probability space $(\Omega, \mathcal{F}, P)$ that takes values in E, then $X(\omega)$ is an element of $C_E[0,1]$ and $X_t \doteq \pi_t X \in E$, where π_t is the projection of Definition 1.8. If the element of $C_E[0,1]$ describes the motion of some object of interest over the time $t = 0$ to $t = 1$, then it will often be called the *path* of that object.

Note that the topology defined on $C_E[0,1] \subset E^{[0,1]}$ is not the subspace topology of the product topology on $E^{[0,1]}$. A new topology has specifically been defined by the uniform metric. Since this is a new topology, topological properties must be assessed anew.

Proposition 2.18. *If the metric space (E, d_E) is separable and complete, then the metric space $C_E[0,1]$ with the uniform metric is separable and complete.*

This is proved in Billingsley (1999) [5] in Example 1.3 for the specific case of $E = \mathbb{R}$, and the generalization is clear.

Proposition 2.19. *The projections $\pi_t : C_E[0,1] \to E$ are continuous for each $t \in [0,1]$.*

Proof: Let $B(a, \varepsilon_1)$, for some $a \in E$ and $\varepsilon_1 > 0$, be an open ball in the base for the topology $\mathcal{T}_E$ on E, and let $y \in \pi_t^{-1}(B(a,\varepsilon_1))$, so that $d_E(y(t), a) < \varepsilon_1$. Now let $\varepsilon_2 = \varepsilon_1 - d_E(y(t), a)$, let $O_y = \{x \in C_E[0,1] : d(x,y) < \varepsilon_2\}$, and note that

$$\begin{aligned} x \in O_y &\Rightarrow d_E(x(t), y(t)) < \varepsilon_2 \\ &\Rightarrow d_E(x(t), a) \leq d_E(x(t), y(t)) + d_E(y(t), a) < \varepsilon_1 \\ &\Rightarrow x \in \pi_t^{-1}(B(a, \varepsilon_1)). \end{aligned} \tag{2.11}$$

Since O_y is an open set such that $y \in O_y \subset \pi_t^{-1}(B(a,\varepsilon_1))$, and y was arbitrary, the set $\pi_t^{-1}(B(a,\varepsilon_1))$ is open in $C_E[0,1]$ and π_t is continuous. ■

Proposition 2.20. *The collection of projections $\{\pi_t : t \in [0,1]\}$ is separating on $C_{\mathbb{R}}[0,1]$.*

This is given in Billingsley (1999) [5] in Example 1.3 for finite-dimensional sets, and the extension to finite-dimensional projections is clear. Note that it is vitally important to understand that this collection of projections is not a collection of functionals unless $E = \mathbb{R}$. To obtain functionals on $C_E[0,1]$ from projections, it is necessary to compose the projections with functionals $\phi : E \to \mathbb{R}$.

To handle discontinuous paths, some notion of reasonable discontinuity is required. If the paths could take any value in, say $\mathbb{R}^{[0,1]}$, then a number of pathogenic cases could occur. The following definition restricts the discontinuities in a common way.

Definition 2.21. A function $f : \mathbb{R} \to E$ that is, at any point $x \in \mathbb{R}$, continuous from the right at x and has a limit which exists from the left at x is called *càdlàg* (continu à droit, limites à gauche). If the domain of the function f is the set $[u, v]$, and the function satisfies the continuity and limit properties at every $x \in [u, v]$, then the function is said to be *càdlàg on* $[u, v]$.

Definition 2.22. For any metric space (E, d_E), the space $D_E[0, 1]$ is defined to be the set of all càdlàg functions on $[0, 1]$ that take values in E.

Note that no metric was defined on $D_E[0, 1]$ in the above definition. It is difficult to arrange for a metric on the càdlàg functions that will have useful convergence properties; the naive definition using the uniform metric will suffer from the following circumstance.

Example 2.23. Let $x(t) = I_{\{t=1\}}$ and $x_n(t) = I_{\{t \in [1-\frac{1}{n},1]\}}$ be elements of $D_{\mathbb{R}}[0, 1]$. Note that the lack of right continuity at the point 1 is irrelevant, since only the space $[0, 1]$ is at issue. For all n, there exists a point $t \in (1-\frac{1}{n}, 1)$ such that

$$|x(t) - x_n(t)| = 1, \tag{2.12}$$

so that x_n cannot converge to x in the uniform metric.

This is unfortunate, since it seems that the functions x_n are getting close to x in some sense, in particular if the parameter t is taken to be time and the evaluation of the timing of events is understood to be imperfect, say, because of process noise. A new metric for $D_E[0, 1]$ that fixes this problem is the *Skorohod metric.*

Definition 2.24. Let Λ be the collection of strictly increasing, continuous functions from $[0, 1]$ onto $[0, 1]$. Then the *Skorohod metric* on the space $D_E[0, 1]$ is defined by

$$d_{[0,1]}(x, y) = \inf_{\lambda \in \Lambda} \left\{ \sup_{s<t} \left| \log \frac{\lambda t - \lambda s}{t - s} \right| \vee \sup_t d_E(x(t) - y(\lambda t)) \right\}. \tag{2.13}$$

The argument that the function so defined is a metric is outlined in Billingsley (1999) [5] on pages 124 to 126. The Skorohod metric defined here will always be used as the metric in the space $D_E[0, 1]$. Note, again, that the product topology has been explicitly abandoned in favour of the topology generated by the Skorohod metric.

While the Skorohod metric looks horrible, it reduces to a nice metric on spaces in which discontinuities are not an issue.

Proposition 2.25. *The Skorohod metric applied solely to elements of* $C_E[0,1]$ *is equivalent to the uniform metric.*

This is in Billingsley (1999) [5] in the discussion on page 124.

As in the case for $C_E[0,1]$, if the underlying space E has tractable properties then they transfer over to the space $D_E[0,1]$.

Proposition 2.26. *If the space* (E, d_E) *is a complete separable metric space then the Skorohod metric on* $D_E[0,1]$ *is separable and complete.*

This is Theorem 12.2 in Billingsley (1999) [5].

Proposition 2.27. *If* (E, d_E) *is a separable metric space then the finite-dimensional sets* $\pi^{-1}_{t_1,\ldots,t_k} H$ *for* $H \in \mathcal{B}(\mathbb{R}^k)$ *are separating on* $D_E[0,1]$ *and the projections on* $D_E[0,1]$ *are measurable and, if* $E = \mathbb{R}$*, separating.*

The proof is given in Theorem 12.5 in Billingsley (1999) [5].

The previously defined spaces $C_E[0,1]$ and $D_E[0,1]$ only allow functions which, if the parameter t is taken to be time, exist for some finite duration and are then done. If there is no preconceived end point for a random process, it needs to reside in a larger domain.

Definition 2.28. For any metric space (E, d_E), the space $D_E[0,\infty)$ is defined to be the set of all càdlàg functions on $[0,\infty)$ that take values in E.

This is a simple extension of the definition of $D_E[0,1]$ to an infinite time horizon. However, a new Skorohod metric is required on $D_E[0,\infty)$. First, let $D_E[0,m]$ be, as one would expect, the collection of càdlàg functions on $[0,m]$ for integers m, and define an analogous Skorohod metric as follows.

Definition 2.29. For each $m \in \mathbb{N}^+$, let Λ_m be the collection of strictly increasing, continuous functions from $[0,m]$ onto $[0,m]$, and define the Skorohod metric on the space $D_E[0,m]$ by

$$d_{[0,m]}(x,y) = \inf_{\lambda \in \Lambda_m} \left\{ \sup_{s<t} \left| \log \frac{\lambda t - \lambda s}{t-s} \right| \vee \sup_t d_E(x(t) - y(\lambda t)) \right\}. \qquad (2.14)$$

Definition 2.30. For each $m \in \mathbb{N}^+$, define

$$g_m(t) = \begin{cases} 1 & \text{if } t \leq m-1, \\ m-t & \text{if } m-1 \leq t \leq m, \\ 0 & \text{if } t \geq m, \end{cases} \tag{2.15}$$

let x^m be the element of $D_E[0,\infty)$ defined by $x^m(t) = g_m(t)x(t)$ for all $t \geq 0$, and define the Skorohod metric on $D_E[0,\infty)$ by

$$d_{[0,\infty)}(x,y) = \sum_{m=1}^{\infty} 2^{-m}(d_{[0,m]}(x^m, y^m) \wedge 1). \tag{2.16}$$

That this is a metric is proved in Billingsley (1999) [5] on pages 166 to 168. Nice properties from the underlying metric space E also transfer to the new space $D_E[0,\infty)$.

Proposition 2.31. *If the space (E, d_E) is a complete separable metric space then the Skorohod metric on $D_E[0,\infty)$ is separable and complete.*

This is Theorem 16.3 in Billingsley (1999) [5].

Proposition 2.32. *If (E, d_E) is a separable metric space then the finite-dimensional sets $\pi^{-1}_{t_1,\ldots,t_k}H$ for $H \in \mathcal{B}(\mathbb{R}^k)$ are separating on $D_E[0,\infty)$ and the projections on $D_E[0,\infty)$ are measurable and, if $E = \mathbb{R}$, separating.*

The proof is given in Theorem 16.6 in Billingsley (1999) [5].

Note that on $D_E[0,1]$ and on $D_E[0,\infty)$, the projections are only measurable, rather than being continuous. This is because projections at the exact time of a jump in the process are not necessarily continuous. It is for this reason that the weak generator of a càdlàg historical process must be defined on the set of bounded functionals, rather than bounded continuous functionals.

2.3 Homeomorphic methods

One of the core techniques of this thesis is to operate with a subset of one of the basic collections of functionals, but a subset that is rich enough to maintain some property or condition that is true for the entire collection. The following definitions are of such useful subsets of functionals on a topological space $(S, \mathcal{T}_S)$.

Definition 2.33. Let the collection $\mathcal{M} \subset M(S)$.

i) $\mathcal{M}$ *separates points* (SP) if for any $x, y \in S$ with $x \neq y$ there is a $g \in \mathcal{M}$ for which $g(x) \neq g(y)$.

ii) $\mathcal{M}$ *strongly separates points* (SSP) if for every $x \in S$ and neighbourhood O_x of x, there is a finite collection $\{g_1, \ldots, g_k\} \subset \mathcal{M}$ such that

$$\inf_{y \notin O_x} \max_{1 \leq l \leq k} |g_l(x) - g_l(y)| > 0. \tag{2.17}$$

iii) $\mathcal{M}$ is *pointwise convergence determining* (PWCD) if for any net $\{x_i\}_{i \in I} \subset S$ and point $x \in S$, whenever $g(x_i) \to g(x)$ for all $g \in \mathcal{M}$, necessarily $x_i \to x$ in S.

The concept of SSP is introduced in Ethier and Kurtz (1986) [9], and the concept of PWCD is introduced in Bhatt and Karandikar (1993a) [3], although Bhatt and Karandikar label their definition of PWCD with the name SSP. This quirk will be explained subsequently.

Note that "separation" and "convergence determining" in these definitions have no necessary connection with the separation of probability measures and determination of weak convergence in Definition 2.7. (A connection is possible, but it will not be detailed in this thesis.)

Example 2.34. Take d to be the distance function of the metric space (S, d), let $g_x : S \to \mathbb{R}$ be defined by $g_x(y) = d(x, y)$, and define the collection $\mathcal{M}$ by $\mathcal{M} \doteq \{g_x\}_{x \in S}$. Then the collection $\mathcal{M}$ SP, SSP, and is PWCD.

Example 2.35. Let $S = C([0,1])$ with the usual convention that $d(f, g) \doteq \|f - g\|_\infty$. Define $\pi_x(f) \doteq f(x)$ for all $x \in [0,1]$ and $f \in S$ and let $\mathcal{M} = \{\pi_x\}_{x \in [0,1]}$. Then, if $f, g \in S$ and $f \neq g$ there exists an $x \in [0,1]$ for which $f(x) \neq g(x)$, that is, $\pi_x(f) \neq \pi_x(g)$, so that $\mathcal{M}$ SP.

Now let $f = 0$, the zero function on $[0,1]$, and define

$$f_n(x) = \begin{cases} nx & \text{if } \ 0 \leq x \leq \frac{1}{n} \\ 2 - nx & \text{if } \ \frac{1}{n} \leq x \leq \frac{2}{n} \\ 0 & \text{if } \ \frac{2}{n} \leq x \leq 1 \end{cases} . \tag{2.18}$$

Note that these functions are all in S and that $\|f_n - f\|_\infty = 1\ \forall n$. Suppose that $\mathcal{M}$ were to SSP. Then for $f = 0 \in S$ given above, and a neighbourhood taken small enough to exclude all of the f_n, there would have to be a finite collection $\{\pi_{x_1}, \ldots, \pi_{x_k}\} \subset \mathcal{M}$ such that $\min_{n\in\mathbb{N}} \max_{1\leq l\leq k} |\pi_{x_l}(f) - \pi_{x_l}(f_n)| > 0$. But if $x_* = \min_{1\leq l\leq k} x_l$, then there exists an N such that $\frac{2}{N} < x_*$, and so $\pi_{x_l}(f_N) = 0$, $1 \leq l \leq k$, a contradiction. So $\mathcal{M}$ does not SSP. Also, $f_n(x) \to 0 = f(x)$ $\forall x \in [0,1]$, that is, $\pi_x(f_n) \to \pi_x(f)$ $\forall x \in [0,1]$, but $f_n \nrightarrow f$ in S, so $\mathcal{M}$ is not PWCD.

Recall that a topological space $(S, \mathcal{T}_S)$ is T_1 if, for all x and y in S such that $x \neq y$, there exists a neighbourhood $O_x \in \mathcal{T}_S$ of x such that $y \notin O_x$. Note that a metric space E is always T_1. The following proposition states that, not only is it possible to SSP or be PWCD without SP, but in fact SP is a consequence of SSP and PWCD.

Proposition 2.36. *Suppose S is a T_1 topological space and let $\mathcal{M} \subset M(S)$.*

i) If $\mathcal{M}$ SSP, then $\mathcal{M}$ SP.

ii) If $\mathcal{M}$ is PWCD, then $\mathcal{M}$ SP.

Proof: If $\mathcal{M}$ SSP, then for any $x \in S$ and any neighbourhood O_x of x, there exist an $\varepsilon > 0$ and a finite collection $\{g_1, \ldots, g_k\} \subset \mathcal{M}$ such that

$$\max_{1\leq l\leq k} |g_l(x) - g_l(y)| < \varepsilon \ \Rightarrow\ y \in O_x. \tag{2.19}$$

So if $y \neq x$ then, since S is T_1, there is an O_x such that $y \notin O_x$, and then one of the above functions g_l will satisfy $g_l(x) \neq g_l(y)$.

Now, say that $\mathcal{M}$ is PWCD and let $x, y \in S$ with $x \neq y$. Let $x_i = y$, $\forall i$. Then since S is T_1, $x_i \nrightarrow x$, so that by the definition of PWCD, $\exists g \in \mathcal{M}$ such that $g(x_i) \nrightarrow g(x)$, that is, $g(y) \neq g(x)$. ■

The following fundamental theorem unites the apparently unconnected concepts of SSP and PWCD.

Theorem 2.37. *If $\mathcal{M} \subset M(S)$, then $\mathcal{M}$ SSP if and only if $\mathcal{M}$ is PWCD.*

Proof: Suppose $\mathcal{M}$ SSP; it is the case that any net $\{x_\alpha\}$ which does not converge to a point x cannot satisfy the condition that $g(x_\alpha) \to g(x)$, $\forall g \in \mathcal{M}$, so that $\mathcal{M}$ must be PWCD. To see this, let $x \in S$ and let the net $\{x_\alpha\}_{\alpha \in J} \subset S$ be such that $x_\alpha \nrightarrow x$. Then there exists a neighbourhood O_x of x, an $\varepsilon > 0$, and a finite set $\{g_1, \ldots, g_k\} \subset \mathcal{M}$ such that for any $\beta \in J$, there is an $\alpha \succeq \beta$ satisfying $x_\alpha \notin O_x$, and then

$$\max_{1 \leq l \leq k} |g_l(x_\alpha) - g_l(x)| \geq \varepsilon > 0. \tag{2.20}$$

So, not all of the functional values $\{g(x)\}_{g \in \mathcal{M}}$ are converging, and thus SSP implies PWCD.

The reverse implication will be proven with the contrapositive. Suppose that $\mathcal{M}$ does not SSP, that is, there exist a point $x \in S$ and a neighbourhood O_x of x for which all finite collections $\{g_1, \ldots, g_k\}$ from $\mathcal{M}$ satisfy

$$\{g_1, \ldots, g_k\} \subset \mathcal{M} \Rightarrow \inf_{y \notin O_x} \max_{1 \leq l \leq k} |g_l(x) - g_l(y)| = 0. \tag{2.21}$$

If $\mathcal{M}$ is finite, then taking $\{g_1, \ldots, g_k\} = \mathcal{M}$ and choosing a sequence of points $\{y_n\}_{n=0}^{\infty}$ in $S \setminus O_x$ such that $\max_{g \in \mathcal{M}} |g(x) - g(y_n)| < \frac{1}{n}$ generates a sequence for which $g(y_n) \to g(x)$, $\forall g \in \mathcal{M}$, but $y_n \nrightarrow x$, so that $\mathcal{M}$ is not PWCD. So, assume that $\mathcal{M}$ is infinite.

Define a directed set $(J, \preceq)$ by taking J to be the collection of finite subsets of $\mathcal{M}$ and ordering by inclusion, that is, if α_1 and α_2 are both finite subsets of $\mathcal{M}$ then $\alpha_1 \preceq \alpha_2$ if $\alpha_1 \subset \alpha_2$. Next, define a net $\{y_\alpha\}_{\alpha \in J}$ on S as follows: if $\alpha \in J$ is a subset with k elements, say $\alpha = \{g_1, \ldots, g_k\}$, then choose y_α to be an element of $S \setminus O_x$ such that $\max_{1 \leq l \leq k} |g_l(x) - g_l(y_\alpha)| < \frac{1}{k}$. Naturally, such an element exists since $\inf_{y \notin O_x} \max_{1 \leq l \leq k} |g_l(x) - g_l(y)| = 0$.
Now, for any $g \in \mathcal{M}$ and any $\varepsilon > 0$, take $\beta \in J$ such that $g \in \beta$ and β contains a finite number $K > \frac{1}{\varepsilon}$ of functions in $\mathcal{M}$. Then $\alpha \succeq \beta \Rightarrow |g(x) - g(y_\alpha)| \leq \frac{1}{K} < \varepsilon$, and thus $g(y_\alpha) \to g(x)$. However, $y_\alpha \nrightarrow x$, since $y_\alpha \notin O_x$, $\forall \alpha \in J$. So $g(y_\alpha) \to g(x)$, $\forall g \in \mathcal{M}$, but $y_\alpha \nrightarrow x$, and thus $\mathcal{M}$ is not PWCD.

Whether finite or infinite, $\mathcal{M}$ is not PWCD, so the contrapositive is shown; that is, PWCD implies SSP. ■

This explains the terminology of Bhatt and Karadikar in their 1993 papers. In those papers, Bhatt and Karandikar also introduce and make use of a homeomorphism between difficult spaces S and subsets of $\mathbb{R}^\infty$, working out difficulties in $\mathbb{R}^\infty$ and then passing the managed objects back into S. The following is a version of a proof that their methods are viable.

Theorem 2.38. *Suppose S is T_1 and $\mathcal{M} \subset C(S)$. Then $\Gamma(x) \doteq \{g(x)\}_{g\in\mathcal{M}}$ is a homeomorphism between S and $\Gamma(S) \subset \mathbb{R}^{\mathcal{M}}$ if and only if $\mathcal{M}$ is PWCD.*

Proof: First, suppose that Γ as defined above is a homeomorphism between S and $\Gamma(S) \subset \mathbb{R}^{\mathcal{M}}$, let $\{x_i\}_{i\in I}$ be any net and x be any point in S, and suppose that $g(x_i) \to g(x)$, $\forall g \in \mathcal{M}$. Let O_x be any neighbourhood of x. Then $\Gamma(O_x)$ is an open neighbourhood of $\Gamma(x)$ in $\Gamma(S)$ and, by Proposition 1.10, $\{g(x_i)\}_{g\in\mathcal{M}} \to \{g(x)\}_{g\in\mathcal{M}} = \Gamma(x)$ in $\mathbb{R}^{\mathcal{M}}$, so there exists a $j \in I$ such that $i \succeq j \Rightarrow \{g(x_i)\}_{g\in\mathcal{M}} \in \Gamma(O_x) \Rightarrow x_i \in O_x$. So $\{x_i\}_{i\in I}$ converges to x in S, and thus $\mathcal{M}$ is PWCD.

Next, suppose that $\mathcal{M}$ is PWCD and define Γ as above. This function Γ is continuous since each component function $g \in \mathcal{M}$ is, and Γ^{-1} is well-defined because $\mathcal{M}$ SP by Proposition 2.36. It only remains to show that Γ^{-1} is continuous. Assume that $\Gamma(x_i) \to \Gamma(x)$ for some net $\{x_i\}_{i\in I}$ and point x in S. Then necessarily $g(x_n) \to g(x)$, $\forall g \in \mathcal{M}$, and thus $x_n \to x$ since $\mathcal{M}$ is PWCD. So Γ^{-1} is continuous and Γ is a homeomorphism between S and $\Gamma(S) \subset \mathbb{R}^{\mathcal{M}}$. ■

Note that the above proof that $\mathcal{M}$ is PWCD requires only that Γ^{-1} be continuous, and also note that if $\mathcal{M}$ is PWCD then Γ^{-1} is continuous regardless of whether $\mathcal{M} \subset C(S)$ or $\mathcal{M} \subset B(S)$. So, if the condition that $\mathcal{M} \subset C(S)$ is reduced to $\mathcal{M} \subset B(S)$, then it remains true that Γ has a continuous inverse if and only if $\mathcal{M}$ is PWCD.

Naturally, $\mathcal{M} \subset C(S)$ strongly separating points is also equivalent to the above function Γ being a homeomorphism because of Theorem 2.37. Because of this, a proof of this statement is strictly unnecessary, but to complete the circle of ideas an explicit proof will be provided. First, some lemmas will be required which are also relevant to later results.

Definition 2.39. For any finite subcollection $\{g_1, \ldots, g_k\} = \mathcal{M}_k \subset \mathcal{M} \subset M(S)$ define the *$(\mathcal{M}_k,\varepsilon)$-ball about the point x*, $B_{\mathcal{M}_k}(x, \varepsilon)$, by

$$B_{\mathcal{M}_k}(x,\varepsilon) \doteq \left\{ y \in S : \max_{1\le l\le k} |g_l(x) - g_l(y)| < \varepsilon \right\}. \tag{2.22}$$

The following lemma provides a convenient equivalent definition of SSP.

Lemma 2.40. *The collection $\mathcal{M} \subset M(S)$ SSP if and only if for each x and neighbourhood O_x of x, there exists a finite subcollection $\{g_1, \ldots, g_k\} = \mathcal{M}_k \subset \mathcal{M}$ such that $B_{\mathcal{M}_k}(x, \varepsilon) \subset O_x$.*

Proof: If $\mathcal{M}$ SSP, then by definition, for each $x \in S$ and neighbourhood O_x of x there is a finite subcollection $\{g_1, \ldots, g_k\} = \mathcal{M}_k \subset \mathcal{M}$ such that

$$\inf_{y\notin O_x} \max_{1\le l\le k} |g_l(x) - g_l(y)| > 0, \tag{2.23}$$

or equivalently, by taking an ε value equal to this infimum, there exists a collection $\{g_1, \ldots, g_k\} = \mathcal{M}_k \subset \mathcal{M}$ and an $\varepsilon > 0$ such that

$$\max_{1\le l\le k} |g_l(x) - g_l(y)| < \varepsilon \Rightarrow y \in O_x. \tag{2.24}$$

This is the statement that $B_{\mathcal{M}_k}(x, \varepsilon) \subset O_x$.

For the reverse implication, let x be in S, let O_x be a neighbourhood of x, and let $\{g_1, \ldots, g_k\} = \mathcal{M}_k \subset \mathcal{M}$ be the finite collection for which, by hypothesis, $B_{\mathcal{M}_k}(x, \varepsilon) \subset O_x$. Then for any $y \notin O_x$, also $y \notin B_{\mathcal{M}_k}(x, \varepsilon)$, so that $\max_{1\le l\le k} |g_l(x) - g_l(y)| \ge \varepsilon$. Since this is true for any $y \notin O_x$,

$$\inf_{y\notin O_x} \max_{1\le l\le k} |g_l(x) - g_l(y)| \ge \varepsilon > 0, \tag{2.25}$$

so that $\mathcal{M}$ SSP. ■

Definition 2.41. Define the sets $\mathbb{B}_{\mathcal{M}}$ by

$$\mathbb{B}_{\mathcal{M}} \doteq \{B_{\mathcal{M}_k}(x,\varepsilon) : \mathcal{M}_k \subset \mathcal{M}, \mathcal{M}_k \text{ finite}, x \in S, \varepsilon > 0\}. \tag{2.26}$$

Lemma 2.42. *For any $\mathcal{M} \subset M(S)$, the sets $\mathbb{B}_{\mathcal{M}}$ form a basis for a topology $\mathcal{T}_{\mathcal{M}}$ on S.*

Proof: To show that $\mathbb{B}_{\mathcal{M}}$ is a basis, it is necessary to show that for any B_1 and B_2 in $\mathbb{B}_{\mathcal{M}}$, and any $x \in B_1 \cap B_2$, there exists a $B_3 \in \mathbb{B}_{\mathcal{M}}$ such that $x \in B_3 \subset B_1 \cap B_2$. So, let $B_{\mathcal{M}_{k_1}}(x_1, \varepsilon_1)$, with $\mathcal{M}_{k_1} = \{g_1, \ldots, g_{k_1}\}$, and $B_{\mathcal{M}_{k_2}}(x_2, \varepsilon_2)$, with $\mathcal{M}_{k_2} = \{h_1, \ldots, h_{k_2}\}$, be in $\mathbb{B}_{\mathcal{M}}$. Suppose that $x \in B_{\mathcal{M}_{k_1}}(x_1, \varepsilon_1) \cap B_{\mathcal{M}_{k_2}}(x_2, \varepsilon_2)$, that is, that

$$\max_{1 \le l \le k_1} |g_l(x_1) - g_l(x)| < \varepsilon_1 \text{ and } \max_{1 \le l \le k_2} |h_l(x_2) - h_l(x)| < \varepsilon_2. \tag{2.27}$$

Define

$$\varepsilon = \left(\varepsilon_1 - \max_{1 \le l \le k_1} |g_l(x_1) - g_l(x)|\right) \wedge \left(\varepsilon_2 - \max_{1 \le l \le k_2} |h_l(x_2) - h_l(x)|\right), \tag{2.28}$$

$$f_l = \begin{cases} g_l & : \ 1 \le l \le k_1 \\ h_{(l-k_1)} & : \ k_1 + 1 \le l \le k_1 + k_2 \end{cases}, \tag{2.29}$$

and $\mathcal{M}_k = \{f_1, \ldots, f_{k_1+k_2}\}$. Then

$$x \in B_{\mathcal{M}_k}(x, \varepsilon) \subset B_{\mathcal{M}_{k_1}}(x_1, \varepsilon_1) \cap B_{\mathcal{M}_{k_2}}(x_2, \varepsilon_2), \tag{2.30}$$

so $\mathbb{B}_{\mathcal{M}}$ is a basis. ∎

For obvious reasons, the sets $\mathbb{B}_{\mathcal{M}}$ will be called the *basis generated by* $\mathcal{M}$, and the topology $\mathcal{T}_{\mathcal{M}}$ will be called the *topology generated by* $\mathcal{M}$.

Lemma 2.43. *If $\mathcal{M} \subset M(S)$ SSP then the topology $\mathcal{T}_{\mathcal{M}}$ generated by the basis $\mathbb{B}_{\mathcal{M}}$ is finer than the topology $\mathcal{T}_S$, that is, $\mathcal{T}_S \subset \mathcal{T}_{\mathcal{M}}$.*

Proof: It suffices to show that for each $x \in S$ and neighbourhood O_x of x, there exists a $B \in \mathcal{T}_{\mathcal{M}}$ such that $x \in B \subset O_x$. So, taking an $x \in S$ and neighbourhood O_x, there exists by Lemma 2.40 a finite collection $\mathcal{M}_k \subset \mathcal{M}$ such that $B_{\mathcal{M}_k}(x, \varepsilon) \subset O_x$, and this set is in the basis that generates $\mathcal{T}_{\mathcal{M}}$. ∎

Theorem 2.44. *Let $(S, \mathcal{T}_S)$ be a T_1 topological space and suppose $\mathcal{M} \subset M(S)$ and that $\Gamma : S \to \mathbb{R}^{\mathcal{M}}$ is defined by $\Gamma(x) \doteq \{g(x)\}_{g \in \mathcal{M}}$. Then $\Gamma(x)$ has a continuous inverse $\Gamma^{-1} : \Gamma(S) \to S$ if and only if $\mathcal{M}$ SSP. Further, Γ is a homeomorphism if and only if $\mathcal{M} \subset C(S)$ and $\mathcal{M}$ SSP.*

Proof: First, suppose $\mathcal{M}$ SSP. Then by Proposition 2.36, $\mathcal{M}$ SP, so Γ^{-1} exists. By Lemma 2.43, $\mathcal{T}_S \subset \mathcal{T}_{\mathcal{M}}$, so Γ^{-1} is continuous.

Next, suppose Γ has a continuous inverse. Take any $x \in S$ and neighbourhood O_x of x. Since Γ^{-1} is continuous, $\Gamma(O_x)$ is open in $\mathbb{R}^{\mathcal{M}}$. Now, the sets $\{\eta \in \mathbb{R}^{\mathcal{M}} : \eta_{g_l} \in B(a_l, \varepsilon_l), 1 \le l \le k\}$ for $g_1, \ldots, g_k \in \mathcal{M}$, $a_1, \ldots, a_k \in \mathbb{R}$, and $\varepsilon_1, \ldots, \varepsilon_k > 0$ form a basis in the product topology on $\mathbb{R}^{\mathcal{M}}$, so there exists a set $G = \{\eta \in \mathbb{R}^{\mathcal{M}} : \eta_{g_l} \in B(a_l, \varepsilon_l), 1 \le l \le k\} \subset \Gamma(O_x)$ such that $\Gamma(x) \in G$. Thus

$$\inf_{\xi \notin G} \max_{1 \le l \le k} |(\Gamma(x))_{g_l} - \xi_{g_l}| > 0, \tag{2.31}$$

or equivalently,

$$\inf_{\xi \notin G} \max_{1 \le l \le k} \left| g_l(x) - g_l(\Gamma^{-1}(\xi)) \right| > 0. \tag{2.32}$$

Then, since $\Gamma^{-1}(G) \subset O_x$ means that $\xi \notin G \Rightarrow \Gamma^{-1}(\xi) \notin O_x$,

$$\inf_{y \notin O_x} \max_{1 \le l \le k} |g_l(x) - g_l(y)| > 0, \tag{2.33}$$

so that $\mathcal{M}$ SSP.

If it is already known that Γ^{-1} is continuous, then, by Proposition 1.10(i), Γ is a homeomorphism if and only if each component function g is continuous, that is, if and only if $\mathcal{M} \subset C(S)$. ■

Given a collection $\mathcal{M} \subset M(S)$ that might not SSP, it is still possible to define a topology $\mathcal{T}_{\mathcal{M}}$ through the basis $\mathbb{B}_{\mathcal{M}}$ of $(\mathcal{M}_k, \varepsilon)$-balls as defined in equation (2.26), although it may be that $(S, \mathcal{T}_{\mathcal{M}})$ will be strictly coarser than $(S, \mathcal{T}_S)$.

Definition 2.45. A function ρ on a space E is a *pseudometric* if ρ is a metric *except* that ρ may have points $x, y \in E$ for which $\rho(x, y) = 0$ but $x \neq y$.

Example 2.46. Let $S = C([0,1])$ and define $\pi_x(f) \doteq f(x)$ for all $x \in [0,1]$ and $f \in S$, as in Example 2.35. If $x_1, \ldots, x_k \in [0,1]$ then

$$\rho(f, g) \doteq \sum_{l=1}^{k} |\pi_{x_l}(f) - \pi_{x_l}(g)| \tag{2.34}$$

is a pseudometric on S: it is clearly positive and symmetric, and the triangle inequality holds since it holds on each of the k dimensions. However, any two functions that are identical at the points $x_1, \ldots, x_k$ but differ elsewhere will have ρ-distance zero.

Definition 2.47. For $\mathcal{M} = \{g_k\}_{k=0}^{\infty} \subset M(S)$ define the *countable $\mathcal{M}$-pseudometric* $\rho_{\mathcal{M}}$ by

$$\rho_{\mathcal{M}}(x,y) = \sum_{k=0}^{\infty} 2^{-k}(|g_k(x) - g_k(y)| \wedge 1). \tag{2.35}$$

The following proposition shows that if $\mathcal{M}$ is countable, then the countable $\mathcal{M}$-pseudometric generates a topology identical to the topology generated by $\mathcal{M}$.

Proposition 2.48. *Suppose $\mathcal{M} = \{g_k\}_{k=0}^{\infty} \subset M(S)$. Then $\mathcal{T}_{\mathcal{M}}$ is equal to $\mathcal{T}_{\rho_{\mathcal{M}}}$, the topology generated by the open balls in the countable $\mathcal{M}$-pseudometric $\rho_{\mathcal{M}}$.*

Proof: Define

$$B_{\rho_{\mathcal{M}}}(x,\varepsilon) \doteq \{y \in S : \rho_{\mathcal{M}}(x,y) < \varepsilon\}. \tag{2.36}$$

Then $\mathbb{B}_{\rho_{\mathcal{M}}} = \{B_{\rho_{\mathcal{M}}}(x,\varepsilon) : x \in S, \varepsilon > 0\}$ is a basis for the topology of open balls generated by $\rho_{\mathcal{M}}$. By the definition of $\mathbb{B}_{\mathcal{M}}$ and of $\mathbb{B}_{\rho_{\mathcal{M}}}$ and by the triangle inequality, for every $x \in B \in \mathbb{B}_{\mathcal{M}}$ there exists a $B_{\mathcal{M}_k}(x,\varepsilon) \in \mathbb{B}_{\mathcal{M}}$ such that $B_{\mathcal{M}_k}(x,\varepsilon) \subset B$, and for every $x \in B \in \mathbb{B}_{\rho_{\mathcal{M}}}$ there exists a $B_{\rho_{\mathcal{M}}}(x,\varepsilon) \in \mathbb{B}_{\rho_{\mathcal{M}}}$ such that $B_{\rho_{\mathcal{M}}}(x,\varepsilon) \subset B$. Thus, to show that $\mathcal{T}_{\rho_{\mathcal{M}}}$ and $\mathcal{T}_{\mathcal{M}}$ are equal, it suffices to show that for each $x \in S$, finite $\mathcal{M}_k \subset \mathcal{M}$, and $\varepsilon > 0$, there exists a $\delta > 0$ such that $B_{\rho_{\mathcal{M}}}(x,\delta) \subset B_{\mathcal{M}_k}(x,\varepsilon)$, and that for each $x \in S$ and $\varepsilon > 0$, there exists a finite $\mathcal{M}_k \subset \mathcal{M}$ and $\delta > 0$ such that $B_{\mathcal{M}_k}(x,\delta) \subset B_{\rho_{\mathcal{M}}}(x,\varepsilon)$.

So, first, let $x \in S$, $\mathcal{M}_k = \{g_{k_1}, \ldots, g_{k_m}\} \subset \mathcal{M}$, and $\varepsilon > 0$. Let $N =$

$\max_{1 \le l \le m} k_l + 1$ and let $\delta = 2^{-N}(\varepsilon \wedge 1)$. Then

$$\begin{aligned}
y \in B_{\rho_{\mathcal{M}}}(x,\delta) &\Rightarrow \rho_{\mathcal{M}}(x,y) < \delta \\
&\Rightarrow \sum_{k=0}^{\infty} 2^{-k}\left(|g_k(x) g_k(y)| \wedge 1\right) < 2^{-N}(\varepsilon \wedge 1) \\
&\Rightarrow \sum_{l=1}^{m} 2^{-N}\left(|g_{k_l}(x) - g_{k_l}(y)| \wedge 1\right) < 2^{-N}(\varepsilon \wedge 1) \\
&\Rightarrow \max_{1 \le l \le m} |g_{k_l}(x) - g_{k_l}(y)| < \varepsilon \wedge 1 \le \varepsilon \\
&\Rightarrow y \in B_{\mathcal{M}_k}(x,\varepsilon). \qquad (2.37)
\end{aligned}$$

Next, let $x \in S$ and $\varepsilon > 0$. Let $N \in \mathbb{N}$ be such that $2^{-N} < \frac{\varepsilon}{2}$, let $\mathcal{M}_N = \{g_0, g_1, \ldots, g_N\}$, and let $\delta = \frac{\varepsilon}{4}$. Note that $\sum_{k=0}^{N} 2^{-k}\delta < 2\delta < \frac{\varepsilon}{2}$ and that $\sum_{k=N+1}^{\infty} 2^{-k} = 2^{-N} < \frac{\varepsilon}{2}$. Then $y \in B_{\mathcal{M}_k}(x,\delta) \Rightarrow \max_{0 \le k \le N} |g_k(x) - g_k(y)| < \delta$, so that

$$\begin{aligned}
\rho_{\mathcal{M}}(x,y) &= \sum_{k=0}^{\infty} 2^{-k}\left(|g_k(x) - g_k(y)| \wedge 1\right) \\
&= \sum_{k=0}^{N} 2^{-k}\left(|g_k(x) - g_k(y)| \wedge 1\right) + \sum_{k=N+1}^{\infty} \left(|g_k(x) - g_k(y)| \wedge 1\right) \\
&\le \sum_{k=0}^{N} 2^{-k}\delta + \sum_{k=N+1}^{\infty} 2^{-k} \\
&< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon, \qquad (2.38)
\end{aligned}$$

and so $y \in B_{\rho_{\mathcal{M}}}(x,\varepsilon)$. ■

If, in addition, $\mathcal{M}$ SP, then the countable $\mathcal{M}$-pseudometric contains a lot of additional structure.

Proposition 2.49. *If $\mathcal{M} = \{g_k\}_{k=0}^{\infty} \subset M(S)$ and $\mathcal{M}$ SP, then $\rho_{\mathcal{M}}$ is a metric and $\mathcal{M}$ SSP on $(S, \mathcal{T}_{\mathcal{M}})$.*

Proof: Suppose $x \ne y \in S$. Then, since $\mathcal{M}$ SP, there exists a $j \in \mathbb{N}$ such that $g_j(x) \ne g_j(y)$, so that

$$\rho_{\mathcal{M}}(x,y) = \sum_{k=0}^{\infty} 2^{-k}(|g_k(x) - g_k(y)| \wedge 1) \ge 2^{-j}\left(|g_j(x) - g_j(y)| \wedge 1\right) > 0. \qquad (2.39)$$

Thus, $\rho_{\mathcal{M}}$ is a metric. Now, for any $x \in S$, the basis element $B_x \in \mathbb{B}_{\mathcal{M}}$ is defined by

$$B_x = B_{\mathcal{M}_k}(x, \varepsilon) = \left\{ y \in S : \max_{1 \leq l \leq k} |g_l(x) - g_l(y)| < \varepsilon \right\} \tag{2.40}$$

for some $\varepsilon > 0$ and $\{g_1, \dots, g_k\} = \mathcal{M}_k \subset \mathcal{M}$, so there trivially exists this finite subcollection $\mathcal{M}_k \subset \mathcal{M}$ that satisfies the requirement $B_{\mathcal{M}_k}(x, \varepsilon) \subset B_x$ of Lemma 2.40. ■

Example 2.50. Let $S = \mathbb{R}^\infty$ and let $g_k = \pi_k$, $k \in \mathbb{N}$. The collection $\pi = \{\pi_k\}_{k=0}^{\infty} \subset M(S)$ generates a topology $\mathcal{T}_\pi$ which is identical to $\mathcal{T}_{\mathbb{R}^\infty}$, the product topology on $\mathbb{R}^\infty$ with the standard topology. To see this, note that for any finite $\mathcal{M}_k \subset \pi$, the $(\mathcal{M}_k, \varepsilon)$-ball is open in the product topology, and that for any open set $B = \left\{x \in \mathbb{R}^\infty : x_{k_j} \in U_{k_j},\ 1 \leq j \leq m\right\}$ in the product topology and $x \in B$, an ε can be chosen so that $B(x_{k_j}, \varepsilon) \subset U_{k_j}$, $1 \leq j \leq m$, so that

$$\begin{aligned} B_{\mathcal{M}_k}(x, \varepsilon) &= \left\{ y \in \mathbb{R}^\infty : \max_{1 \leq j \leq m} \left|\pi_{k_j}(x) - \pi_{k_j}(y)\right| < \varepsilon \right\} \\ &= \left\{ y \in \mathbb{R}^\infty : \max_{1 \leq j \leq m} \left|x_{k_j} - y_{k_j}\right| < \varepsilon \right\} \subset B. \end{aligned} \tag{2.41}$$

Further, let ρ_π be as given in equation (2.35). Then, by Proposition 2.48, $\mathcal{T}_\pi = \mathcal{T}_{\rho_\pi}$, so that $\mathcal{T}_{\rho_\pi} = \mathcal{T}_{\mathbb{R}^\infty}$. Since the projections separate points on $\mathbb{R}^\infty$, by Proposition 2.49, ρ_π is a metric. Because $(\mathbb{R}^\infty, \mathcal{T}_{\mathbb{R}^\infty})$ is complete and separable, $(\mathbb{R}^\infty, \rho_\pi)$ is a complete, separable metric space.

One final useful lemma is provided here. Recall from basic topology that a topological space is *second-countable* if it has a countable base.

Lemma 2.51. *If S is a T_1 second-countable topological space, $\mathcal{M} \subset C(S)$, and $\mathcal{M}$ SSP, then there is a countable collection $\{g_k\}_{k=0}^{\infty} \subset \mathcal{M}$ that will also SSP.*

Proof: By the homeomorphism of Theorem 2.44, the collection $\mathbb{B}_{\mathcal{M}}$ defined in equation (2.26) forms a basis for $\mathcal{T}_S$. However, since S is second-countable, any basis contains a countable basis (see, for example, [15, exercise 4-1.5]), so only a countable number of the elements of $\mathcal{M}$ are required, say $\{g_k\}_{k=0}^{\infty}$. ■

Note that the countable collection $\{g_k\}_{k=0}^{\infty}$ in Lemma 2.51 can be taken closed under (countable) multiplication or addition if the containing collection $\mathcal{M}$ is closed under these operations.

Chapter 3

Separation of Historical Processes

3.1 Preliminary theorems

Three basic theorems are listed here that are of use in the proofs in this chapter. The first, Tychonoff's Theorem, allows the construction of compact spaces as a product of compact spaces.

Theorem 3.1 (Tychonoff). *Any product of compact spaces is compact.*

Proof: See Theorem 12.4 of [10]. ■

The Stone-Weierstrass theorem allows the approximation of continuous functions on a compact Hausdorff space $\mathcal{K}$ by functions from some subalgebra of $C(\mathcal{K})$. In order to understand this theorem, it is necessary first to define the term *algebra*.

Definition 3.2. An *algebra* over $\mathbb{R}$ is a vector space $\mathcal{A}$ over $\mathbb{R}$ which has a multiplication defined between its elements such that it is multiplicatively closed and that, for any $f, g, h \in \mathcal{A}$, $(f \cdot g) \cdot h = f \cdot (g \cdot h)$, $(f+g) \cdot h = f \cdot h + g \cdot h$, and $f \cdot (g+h) = f \cdot g + f \cdot h$, and for which scalars $c \in \mathbb{R}$ satisfy $c(f \cdot g) = (cf) \cdot g = f \cdot (cg)$. A *subalgebra* of an algebra $\mathcal{A}$ is a subset of $\mathcal{A}$ which is itself an algebra under the same addition and multiplication as $\mathcal{A}$.

The algebras this thesis is concerned with are those of functionals over a topological space, that is, algebras within $M(S)$, $C_b(S)$, and such. The following example is canonical.

Example 3.3. The collection of functions $C(S)$ forms an algebra over $\mathbb{R}$ if, for $f, g \in C(S)$, $(f+g)(x) \doteq f(x) + g(x)$ and $(f \cdot g)(x) \doteq f(x)g(x)$ for all $x \in S$. This is the *algebra of pointwise addition and multiplication.*

Theorem 3.4 (Stone-Weierstrass). *Suppose that $\mathcal{K}$ is a compact Hausdorff space and $\mathcal{A}$ is a subalgebra of $C(\mathcal{K})$ which includes a non-zero constant function. Then $\mathcal{A}$ is dense in $C(\mathcal{K})$ if and only if $\mathcal{A}$ separates points on $\mathcal{K}$.*

Proof: The proof is in many textbooks, for example, see Theorem A in section 36 of Simmons (1983) [18]. ■

The next result provides an amazing and powerful method to map measures on complete, separable metric spaces to each other.

Theorem 3.5 (Kuratowski). *Let (E_1, d_1) and (E_2, d_2) be complete, separable metric spaces. If $A_1 \in \mathcal{B}(E_1)$ and $\phi : A_1 \to E_2$ is measurable and one-to-one, then $A_2 \doteq \phi(A_1) \in \mathcal{B}(E_2)$ and ϕ^{-1} is a measurable function from A_2 onto A_1.*

Proof: See Theorem 3.9 and Corollary 3.3 of Chapter I of Parthasarathy (1967) [16]. ■

3.2 Main results

The two main theorems of this chapter provide sufficient conditions for collections of functionals to be convergence-determining and to be separating. They generalize results in Ethier and Kurtz (1986) [9], specifically Theorem 3.4.5. A number of lemmas will be required to prove the first theorem.

Lemma 3.6. *If $G \subset \mathbb{R}^\infty$ and the values of G in each component are bounded, then $\overline{G}$ is compact.*

Proof: For each k, let $\overline{R_k(G)}$ be the closure of the range of values for the k^{th} coordinate of G, that is,

$$\overline{R_k(G)} \doteq \overline{\{x_k : x \in G\}} \subset \mathbb{R}. \tag{3.1}$$

Then $\overline{R_k(G)}$ is compact for each k, so that their product $\overline{R_0(G)} \times \overline{R_1(G)} \times \cdots$ is compact by Tychonoff's Theorem, Theorem 3.1. Since $\overline{G}$ is closed and $\overline{G} \subset \overline{R_0(G)} \times \overline{R_1(G)} \times \cdots$, $\overline{G}$ is compact. ∎

Definition 3.7. If $G \subset \mathbb{R}^\infty$, define a *G-enlarged Borel set* $A \subset \mathbb{R}^\infty$ by $A = B \cup B'$, where $B \in \mathcal{B}(G)$ (using the subspace topology) and $B' \cap G = \varnothing$. Also, define the *G-enlarged Borel σ-algebra* $\mathcal{G}$ to contain all G-enlarged Borel sets.

Note that, since $B \subset G$ and $B' \subset (\mathbb{R}^\infty \setminus G)$, the two components B and B' of the decomposition of a given enlarged Borel set will be unique.

Lemma 3.8. *The G-enlarged Borel σ-algebra $\mathcal{G}$ is a σ-algebra.*

Proof: A verification of the axioms of a σ-algebra follows.

i) If $B = \varnothing \in \mathcal{B}(G)$ and $B' = \varnothing$ (which satisfies $\varnothing \cap G = \varnothing$), then $B \cup B' = \varnothing \in \mathcal{G}$.

ii) Let $A \in \mathcal{G}$, say $A = B \cup B'$ with $B \in \mathcal{B}(G)$ and $B' \cap G = \varnothing$. Then

$$\begin{aligned} A^c &= (B \cup B')^c = B^c \cap (B')^c \\ &= ((G \setminus B) \cup G^c) \cap (G \cup B''), \text{ where } B'' \cap G = \varnothing \\ &= ((G \setminus B) \cap (G \cup B'')) \cup (G^c \cap (G \cup B'')) \\ &= ((G \setminus B) \cap G) \cup ((G \setminus B) \cap B'') \cup (G^c \cap G) \cup (G^c \cap B'') \\ &= (G \setminus B) \cup B'' \in \mathcal{G}, \text{ since } G \setminus B \in \mathcal{B}(G). \end{aligned} \tag{3.2}$$

iii) Let $\{A_n\} \subset \mathcal{G}$, say $A_n = B_n \cup B_n'$ with $B_n \in \mathcal{B}(G)$ and $B_n' \cap G = \varnothing$. Then

$$\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} (B_n \cup B_n') = \bigcup_{n=1}^{\infty} B_n \cup \bigcup_{n=1}^{\infty} B_n' \in \mathcal{G}, \tag{3.3}$$

since $\bigcup_{n=1}^{\infty} B_n \in \mathcal{B}(G)$ and $(\bigcup_{n=1}^{\infty} B_n') \cap G = \varnothing$. ∎

Lemma 3.9. *If $G \in \mathcal{B}(\mathbb{R}^\infty)$, then the G-enlarged Borel σ-algebra includes the Borel sets on $\mathbb{R}^\infty$.*

Proof: Let A be any set in $\mathcal{B}(\mathbb{R}^\infty)$. Then $A = (A \cap G) \cup (A \setminus G)$, where $A \cap G \in \mathcal{B}(G)$ and $(A \setminus G) \cap G = \varnothing$. Thus, $\mathcal{B}(\mathbb{R}^\infty) \subset \mathcal{G}$. ∎

Recall that the restriction $\widehat{f}$ of a function $f : S \to T$ to the subset $S_0 \subset S$ is defined by the values $\widehat{f}(x) \doteq f(x)$ for all $x \in S_0$. If $(S, \mathcal{T}_S)$ is a topological space with Borel sets $\mathcal{B}(S)$, then by the argument on pages 243-244 of Billingsley (1999) [5], whenever $S_0 \in \mathcal{B}(S)$ and $(S_0, \mathcal{T}_{S_0})$ is given the relative topology it follows that $\mathcal{B}(S_0) \subset \mathcal{B}(S)$, so that restriction can be defined for probability measures on $\mathcal{B}(S)$.

Lemma 3.10. *Let $(S, \mathcal{T}_S)$ be a topological space and let $\{P_n\}_{n=0}^\infty$ and P be probability measures defined on $(S, \mathcal{B}(S))$. Let $S_0 \in \mathcal{T}_S$, assume that $P_n(S_0) = P(S_0) = 1$ for all $n \in \mathbb{N}$, and define the topological space $(S_0, \mathcal{T}_{S_0})$ to have the relative topology. If $\left\{\widehat{P}_n\right\}_{n=0}^\infty$ and $\widehat{P}$ are the restrictions of $\{P_n\}_{n=0}^\infty$ and P to the space $(S_0, \mathcal{B}(S_0))$, then $P_n \xrightarrow[n\to\infty]{\text{wk}} P$ in $(S, \mathcal{T}_S)$ implies that $\widehat{P}_n \xrightarrow[n\to\infty]{\text{wk}} \widehat{P}$ in $(S_0, \mathcal{T}_{S_0})$.*

Proof: Take $G_0 = G \cap S_0$ to be any open set in S_0, where G is open in S. Then $\widehat{P}_n G_0 = P_n G$ and $\widehat{P} G_0 = PG$, so that

$$\liminf_{n\to\infty} P_n G \geq PG \;\Rightarrow\; \liminf_{n\to\infty} \widehat{P}_n G_0 \geq \widehat{P} G_0, \tag{3.4}$$

and the result follows by the Portmanteau Theorem, Theorem 2.6. ∎

Theorem 3.11. *Suppose $(S, \mathcal{T}_S)$ is a T_1 separable topological space and $\mathcal{M} \subset C_b(S)$ satisfies $\mathcal{M}$ SSP and $\mathcal{M}$ is closed under multiplication. Then, $\mathcal{M}$ is convergence-determining; that is, for all P and P_n which are probability measures on S,*

$$P_n g \to Pg,\ \forall g \in \mathcal{M} \;\Rightarrow\; P_n \xrightarrow[n\to\infty]{\text{wk}} P. \tag{3.5}$$

Proof: By Lemma 2.51, the collection $\mathcal{M}$ can be assumed without loss of generality to be countable, say $\mathcal{M} = \{g_k\}_{k=0}^\infty$. Now, using Theorem 2.44

define the homeomorphism $\Gamma : S \to \Gamma(S) \subset \mathbb{R}^\infty$ by $\Gamma(x) \doteq (g_0(x), g_1(x), \ldots)$ and endow $G \doteq \Gamma(S)$ with the subspace topology.

Given any probability measures P, P_n on $(S, \mathcal{B}(S))$, use them to define new probability measures Q and Q_n on $(G, \mathcal{B}(G))$ by $Q \doteq P\Gamma^{-1}$ and $Q_n \doteq P_n\Gamma^{-1}$. Next, let $\mathcal{G}$ be the extended Borel σ-algebra formed from G and define probability measures $\widehat{Q}$ and $\widehat{Q}_n$ on $\mathcal{G}$ by $\widehat{Q}(A) \doteq Q(B)$, $\widehat{Q}_n(A) \doteq Q_n(B)$, where A is the extended Borel set with first component $B \in \mathcal{B}(G)$.

Since the functions $\{g_k\}$ are bounded, by Lemma 3.6, the closure $K = \overline{G}$ is compact. Let $K \cap \mathcal{G} \doteq \{K \cap A : A \in \mathcal{G}\}$ and note that $\mathcal{B}(K) \subset K \cap \mathcal{G} \subset \mathcal{G}$. Then $\widehat{Q}$ and $\widehat{Q}_n$ define, by restriction, probabilities on $\mathcal{B}(K)$ by $\widehat{Q}(K \cap A) \doteq \widehat{Q}(A)$ and $\widehat{Q}_n(K \cap A) \doteq \widehat{Q}_n(A)$ which are equal to Q and Q_n on $\mathcal{B}(G)$. Using the facts that $G \subset K$ and $G \in \mathcal{G}$, note that $\widehat{Q}(G) = \widehat{Q}(K) = \widehat{Q}(G \cup (K \setminus G)) = Q(G) = 1$, and $\widehat{Q}(K \setminus G) = 0$, and similarly for Q_n.

Now, since $Q = P\Gamma^{-1}$, $Q_n = P_n\Gamma^{-1}$, Γ is continuous, and $\int_S g_k dP_n \to \int_S g_k dP$ by hypothesis, by change of variable

$$\begin{aligned}
\int_K \pi_k d\widehat{Q}_n = \int_G \pi_k d\widehat{Q}_n = \int_G \pi_k dQ_n = \int_G g_k \circ \Gamma^{-1} d(P_n\Gamma^{-1}) \\
= \int_S g_k dP_n \to \int_S g_k dP \\
= \int_G g_k \circ \Gamma^{-1} d(P\Gamma^{-1}) = \int_G \pi_k dQ = \int_G \pi_k d\widehat{Q} = \int_K \pi_k d\widehat{Q} \qquad (3.6)
\end{aligned}$$

for each projection function $\pi_k : \mathbb{R}^\infty \to \mathbb{R}$. Since $\mathcal{M}$ is closed under multiplication, by the same argument as in equation (3.6) but using the π_{k_*} such that $g_{k_1} \cdot g_{k_2} = g_{k_*}$, $\int_K (\pi_{k_1} \cdot \pi_{k_2}) d\widehat{Q}_n \to \int_K (\pi_{k_1} \cdot \pi_{k_2}) d\widehat{Q}$ for each pair of projection functions π_{k_1} and π_{k_2}, and similarly for higher orders of multiplication. As well, $\int_K 1 d\widehat{Q}_n = 1 \to 1 = \int_K 1 d\widehat{Q}$. Thus, by linearity of the integral,

$$\int_K f \circ \pi_{k_1,\ldots,k_j} d\widehat{Q}_n \to \int_K f \circ \pi_{k_1,\ldots,k_j} d\widehat{Q} \qquad (3.7)$$

for all polynomials f in j variables (from among the countably many projection functions) on $\mathbb{R}^j$.

Let $F = \pi_{0,\ldots,j}(K) \subset \mathbb{R}^{j+1}$; then, F is compact because the projections are continuous and by Tychonoff's Theorem, Theorem 3.1. The polynomials are a subalgebra of $C(F)$ that include the constant functions and SP, so by the Stone-Weierstrass Theorem, Theorem 3.4, they are dense in $C(F)$. Thus, for any $g \in C_b(\mathbb{R}^{j+1})$, there exists a polynomial f defined on F such that $\sup_{x\in F} |f(x) - g(x)| < \varepsilon$. If the function g is restricted to F then, because g has $\widehat{Q}_n$-measure zero outside of F,

$$\sup_{n\in\mathbb{N}} \left| \int_F g d\widehat{Q}_n \pi_{0,\ldots,j}^{-1} - \int_F f d\widehat{Q}_n \pi_{0,\ldots,j}^{-1} \right| < \varepsilon \tag{3.8}$$

and similarly

$$\left| \int_F g d\widehat{Q} \pi_{0,\ldots,j}^{-1} - \int_F f d\widehat{Q} \pi_{0,\ldots,j}^{-1} \right| < \varepsilon. \tag{3.9}$$

Therefore, by the triangle inequality and equation (3.7),

$$(\widehat{Q}_n \pi_{0,\ldots,j}^{-1}) g \to (\widehat{Q} \pi_{0,\ldots,j}^{-1}) g \tag{3.10}$$

for all $g \in C_b(\mathbb{R}^{j+1})$, so that

$$\widehat{Q}_n \pi_{0,\ldots,j}^{-1} \xrightarrow[n\to\infty]{\text{wk}} \widehat{Q} \pi_{0,\ldots,j}^{-1} \tag{3.11}$$

on $\mathbb{R}^{j+1}$ for each $j \in \mathbb{N}$. That is, the finite-dimensional distributions of $\widehat{Q}_n$ converge to $\widehat{Q}$, so that by Proposition 2.16, $\widehat{Q}_n \xrightarrow[n\to\infty]{\text{wk}} \widehat{Q}$ on $\mathbb{R}^\infty$. Then, since $\widehat{Q}_n(K) = \widehat{Q}(K) = 1$, by Lemma 3.10 (with $S = \mathbb{R}^\infty$ and $S_0 = K$), $\widehat{Q}_n \xrightarrow[n\to\infty]{\text{wk}} \widehat{Q}$ when restricted to K.

Now take any uniformly continuous functional h on G. It has a uniquely defined continuous extension $\widehat{h}$ on K, and then $Q_n h = \widehat{Q}_n \widehat{h} \to \widehat{Q}\widehat{h} = Qh$ on G. Since $Q = P\Gamma^{-1}$, $Q_n = P_n \Gamma^{-1}$, and Γ^{-1} is continuous, by the Mapping Theorem, Theorem 2.15, $P_n \xrightarrow[n\to\infty]{\text{wk}} P$. ■

The above theorem and proof idea is taken from a preprint of a paper by Douglas Blount and Michael Kouritzin.

Lemma 3.12. *If $\mathcal{M} = \{g_k\}_{k=0}^\infty \subset M(S)$ and $\mathcal{M}$ SP, then each $g \in \mathcal{M}$ is continuous on the metric space $(S, \rho_\mathcal{M})$ and on $(S, \mathcal{T}_\mathcal{M})$; specifically, $\mathcal{M} \subset C(S, \mathcal{T}_\mathcal{M}) = C(S, \mathcal{T}_{\rho_\mathcal{M}})$.*

Proof: First, $\rho_{\mathcal{M}}$ is a metric and $\mathcal{T}_{\mathcal{M}} = \mathcal{T}_{\rho_{\mathcal{M}}}$ by Proposition 2.48 and Proposition 2.49. Let the ε-ball $B(a, \varepsilon)$ be in the base for the standard topology on $\mathbb{R}$, and take any $g_k \in \mathcal{M}$. Note that $g_k^{-1}(B(a,\varepsilon)) = \{y \in S : |a - g_k(y)| < \varepsilon\}$. Now let $x \in g_k^{-1}(B(a,\varepsilon))$ and let δ be such that $0 < \delta < \varepsilon - |a - g_k(x)|$. Then, if $|g_k(x) - g_k(y)| < \delta$,

$$|a - g_k(y)| \le |1 - g_k(x)| + |g_k(x) - g_k(y)| < |a - g_k(x)| + \varepsilon - |a - g_k(x)| = \varepsilon, \tag{3.12}$$

so that $y \in g_k^{-1}(B(a,\varepsilon))$, and thus

$$B_{\{g_k\}}(x,\delta) = \{y \in S : |g_k(x) - g_k(y)| < \delta\} \subset g_k^{-1}(B(a,\varepsilon)). \tag{3.13}$$

Since $B_{\{g_k\}}(x,\delta) \in \mathcal{T}_{\mathcal{M}}$ and x was any element of $g_k^{-1}(B(a,\varepsilon))$, $g_k^{-1}(B(a,\varepsilon)) \in \mathcal{T}_{\mathcal{M}} = \mathcal{T}_{\rho_{\mathcal{M}}}$. ■

This last proof is very similar to the proof of Proposition 2.19.

Proposition 3.13. *Suppose (E, d_E) is a complete, separable metric space, $\mathcal{M} = \{g_k\}_{k=0}^{\infty} \subset M(E)$, $\mathcal{M}$ SP, and $\rho_{\mathcal{M}}$ is the countable $\mathcal{M}$-pseudometric of Definition 2.47. Then $\mathcal{B}(E, d_E) = \mathcal{B}(E, \rho_{\mathcal{M}})$.*

Proof: Define $\Gamma : (E, d_E) \to (\mathbb{R}^\infty, \rho_\pi)$ by $\Gamma(x) \doteq (g_0(x), g_1(x), \ldots)$, where $(\mathbb{R}^\infty, \rho_\pi)$ is the complete, separable metric space defined in Example 2.50. By Proposition 1.10, Γ is measurable, and since $\mathcal{M}$ SP it is one-to-one, so by Kuratowski's Theorem, Theorem 3.5, for any $A \in \mathcal{B}(E, d_E)$, $\Gamma(A) \in \mathcal{B}(\mathbb{R}^\infty, \rho_\pi)$, and also $\Gamma^{-1} : (\Gamma(E), \rho_\pi) \to (E, d_E)$ is Borel measurable. Additionally, since $\mathcal{M}$ is countable and SP, by Lemma 3.12 Γ is continuous from $(E, \rho_{\mathcal{M}})$ to $(\mathbb{R}^\infty, \rho_\pi)$, and by Proposition 2.48 and Proposition 2.49 $\mathcal{M}$ SSP on $(E, \mathcal{T}_{\mathcal{M}}) = (E, \mathcal{T}_{\rho_{\mathcal{M}}})$. By Theorem 2.44, Γ is a homeomorphism between $(E, \rho_{\mathcal{M}})$ and $(\Gamma(E), \rho_\pi)$. Thus

$$\mathcal{B}(E, d_E) = \left\{\Gamma^{-1}(B) : B \in \mathcal{B}(\Gamma(E), \rho_\pi)\right\} = \mathcal{B}(E, \rho_{\mathcal{M}}). \tag{3.14}$$

■

Theorem 3.14. *Suppose (E, d_E) is a complete, separable metric space and $\mathcal{M} = \{g_k\}_{k=0}^{\infty} \subset B(E)$ satisfies $\mathcal{M}$ SP and $\mathcal{M}$ is closed under multiplication. Then, $\mathcal{M}$ is a separating collection; that is, for all P and Q which are probability measures on S,*

$$Pg = Qg, \ \forall g \in \mathcal{M} \ \Rightarrow \ P = Q. \tag{3.15}$$

Proof: First, define the countable $\mathcal{M}$-pseudometric $\rho_{\mathcal{M}}$ as in Definition 2.47. By Proposition 2.48 and Proposition 2.49, $(E, \rho_{\mathcal{M}})$ is a metric space, $\mathcal{T}_{\mathcal{M}} = \mathcal{T}_{\rho_{\mathcal{M}}}$, and $\mathcal{M}$ SSP on $(E, \mathcal{T}_{\mathcal{M}})$. Also, each $g \in \mathcal{M}$ is continuous on $(E, \mathcal{T}_{\mathcal{M}})$ by Lemma 3.12. Since $\mathcal{M}$ is bounded, this means that $\mathcal{M} \subset C_b(E, \mathcal{T}_{\mathcal{M}})$. Now, by Theorem 3.11, $\mathcal{M}$ is convergence-determining on $(E, \mathcal{T}_{\mathcal{M}})$, and so by Proposition 2.9, $\mathcal{M}$ is separating on $(E, \mathcal{T}_{\mathcal{M}})$.

Suppose by hypothesis that $Pg = Qg$, $\forall g \in \mathcal{M}$. Then, since $\mathcal{M}$ is separating on $(E, \mathcal{T}_{\mathcal{M}})$, $P = Q$ on $\mathcal{B}(E, \mathcal{T}_{\mathcal{M}})$ and thus on $\mathcal{B}(E, \mathcal{T}_{\rho_{\mathcal{M}}})$. Then by Proposition 3.13, $P = Q$ on $\mathcal{B}(E, d_E)$. Thus, $\mathcal{M}$ is separating on (E, d_E).
■

Definition 3.15. If $I_m = \{t_1, \ldots, t_m\} \subset [0, \infty)$ for all $m \in \mathbb{N}^+$, $I_m \subset I_{m+1}$ for all m, and $\bigcup_{m=1}^{\infty} I_m$ is dense in $[0, \infty)$, then the collection $\{I_m\}_{m=0}^{\infty}$ is a *dense increasing mesh of timepoints.* For brevity, this will be shortened to *dense timepoint mesh.* Such an object is often called a *partition increasing to the identity*, but note that in this definition the timepoints in the mesh need not be ordered, that is, there is no requirement that $t_i < t_j$ for $i < j$. (See Remark 3.18 below.)

With these two main results, it is now possible to proceed with the construction of the compound functionals that will provide the separating class for probability measures on spaces of càdlàg paths.

Definition 3.16. Let $\mathcal{C}$ be a collection of functionals on E, and define the functional $\varphi : E^m \to \mathbb{R}$ by

$$\varphi(z) \doteq \varphi_1(z_1)\varphi_2(z_2) \cdots \varphi_m(z_m) \tag{3.16}$$

for some fixed $\varphi_1, \ldots, \varphi_m \in \mathcal{C}$ and each variable $z = (z_1, \ldots, z_m)$. Then $f_{\varphi, I_m} : D_E[0, \infty) \to \mathbb{R}$, the *functional φ evaluated at the timepoints I_m*, is defined as

$$f_{\varphi, I_m}(x) \doteq \varphi\left(\pi_{t_1}(x), \ldots, \pi_{t_m}(x)\right) = \varphi_1(\pi_{t_1}(x)) \cdots \varphi_m(\pi_{t_m}(x)), \qquad (3.17)$$

where the t_i are those from the timepoint mesh I_m. The notation $\varphi \in \mathcal{C}^m$ is used to denote that the functional φ is formed from functionals $\varphi_1, \ldots, \varphi_m \in \mathcal{C}$.

The overall purpose is to have a collection of functionals defined on the finite dimensional distributions of elements $x \in D_E[0, \infty)$, where the particular timepoint coordinates from which projections are formed are taken from the increasing mesh $\{I_m\}_{m=0}^{\infty}$.

Proposition 3.17. *Suppose $\mathcal{C} \subset C_b(E)$ and let $\mathcal{M}$ be the collection of functionals $\mathcal{M} = \{f_{\varphi, I_m} : m \in \mathbb{N}^+, \varphi \in \mathcal{C}^m\}$ for some dense timepoint mesh $\{I_m\}_{m=1}^{\infty}$. Then the following holds:*

i) The collection $\mathcal{M} \subset B(E)$.

ii) If $\mathcal{C}$ is countable, then $\mathcal{M}$ is countable.

iii) If $\mathcal{C}$ is closed under multiplication, then $\mathcal{M}$ is closed under multiplication.

iv) If $\mathcal{C}$ contains the constant 1 functional and SP on E, then $\mathcal{M}$ SP on $D_E[0, \infty)$.

Proof:

i) Each π_{t_i} is measurable by Proposition 2.32, and each φ_i is measurable, so each $\varphi_i \circ \pi_{t_i}$ is measurable and finite products of such functionals are measurable. Further, for any φ, each of $\varphi_1, \ldots, \varphi_m$ is bounded, so their finite product is bounded.

ii) First, countably index $\mathcal{M}$ by the number m of points in the timepoint mesh. For each such m, there are associated a number of elements in $\mathcal{M}$ given by the cardinality of $\mathcal{C}$, which is countable, to the power of this m, the number of timepoints to select an element of $\mathcal{C}$ from, which is finite. So $\mathcal{M}$ is exhausted by associating with each m a countable number of elements of $\mathcal{M}$, and so all of $\mathcal{M}$ is countable.

iii) If f_{φ,I_m} and f_{γ,I_n} are in $\mathcal{M}$ then, assuming without loss of generality that $n \geq m$,

$$\begin{aligned} f_{\varphi,I_m} \cdot f_{\gamma,I_n}(x) &= f_{\varphi,I_m}(x) f_{\gamma,I_n}(x) = \prod_{i=1}^{m} \varphi_i(\pi_{t_i}(x)) \prod_{i=1}^{n} \gamma_i(\pi_{t_i}(x)) \\ &= \prod_{i=1}^{m} \varphi_i(\pi_{t_i}(x))\gamma_i(\pi_{t_i}(x)) \prod_{i=m+1}^{n} \gamma_i(\pi_{t_i}(x)) \\ &= \prod_{i=1}^{m} (\varphi_i \cdot \gamma_i)(\pi_{t_i}(x)) \prod_{i=m+1}^{n} \gamma_i(\pi_{t_i}(x)), \end{aligned} \tag{3.18}$$

so that $f_{\varphi,I_m} \cdot f_{\gamma,I_n} \in \mathcal{M}$.

iv) Let $x \neq y \in D_E[0,\infty)$, say $x(t) \neq y(t)$ at some t. Since $\bigcup_{m=1}^{\infty} I_m$ is dense in $[0,\infty)$, there exists a subsequence $\{t_{n_k}\}_{k=1}^{\infty}$ of the timepoint mesh that converges to t from the right. Since both x and y are continuous from the right and $x(t) \neq y(t)$, there exists a $t_* > t$ such that

$$t \leq s \leq t_* \Rightarrow |x(s) - y(s)| > \frac{|x(t) - y(t)|}{2}. \tag{3.19}$$

Then, there exists a $K \in \mathbb{N}^+$ such that $t < t_{n_K} < t_*$, and then $x(t_{n_K}) \neq y(t_{n_K})$. Let $n = n_K$. Since $\mathcal{C}$ SP, there exists a $\varphi_* \in \mathcal{C}$ such that $\varphi_*(\pi_{t_n}(x)) \neq \varphi_*(\pi_{t_n}(y))$.

Given this particular n and φ_*, let $f \in \mathcal{M}$ be f_{φ,I_n} where

$$\varphi = 1 \cdot 1 \cdots 1 \cdot \varphi_*. \tag{3.20}$$

Note that there are exactly n functionals φ_i in this product, and it is exactly at coordinate n in I_n that the above property of φ_* is true. Then

$$f_{\varphi,I_n}(x) = \varphi_*(\pi_{t_n}(x)) \neq \varphi_*(\pi_{t_n}(y)) = f_{\varphi,I_n}(y), \tag{3.21}$$

so this is a functional in $\mathcal{M}$ that separates x from y.

■

Remark 3.18. From the note in part (iv) above, there is no need for a set of functionals $\cdot 1 \cdots 1$ at the end of the definition of φ (as may be required in other developments in which the timepoints in the mesh are ordered).

Chapter 4

Conclusion

4.1 Application to the motivating problem

In the second submarine tracking problem, the discontinuities in the signal model and the requirement for historical filtering because of the observation model complicated the traditional filtering problem. Instead of needing a collection of functionals from $C_b(S)$ which could separate probability measures, the collection from the new problem was only known to be bounded, that is, to be a subset of $B(D_S[0,\infty))$. However, by using the weak generator for the signal $(\mathcal{L}, \mathcal{D}(\mathcal{L}))$ and the results from this thesis, it is still possible to use this collection of bounded functionals to separate probability measures.

In particular, require of the signal model that $\mathcal{D}(\mathcal{L}) \subset C_b(S)$, and that $\mathcal{D}(\mathcal{L})$ is countable, closed under multiplication, contains the constant 1 functional, and SP on S. Then take some dense timepoint mesh $\{I_m\}_{m=1}^{\infty}$, and find by Proposition 3.17 that the collection $\mathcal{M} \doteq \{f_{\varphi,I_m} : m \in \mathbb{N}^+, \varphi \in \mathcal{C}^m\}$ satisfies $\mathcal{M} \subset B(S)$ and $\mathcal{M}$ is countable, closed under multiplication, and SP on $D_S[0,\infty)$. So, since (S, d_S) is a complete, separable metric space, by Theorem 3.14, $\mathcal{M}$ is a separating collection on $D_S[0,\infty)$.

Once a separating collection of functionals is available, filtering equations can be constructed and this collection can be used in determinations of uniqueness.

4.2 Summary

An initial filtering problem was introduced which was tractable using known methods, but was then extended to a second filtering problem in which the published methods were not sufficient to begin the process of forming a solution. In particular, no method was available to separate probability measures, that is, distinguish one from another, using only a collection of functionals available from the problem definition.

After outlining the problem and providing some background notation and definitions, the essential notion of weak convergence was introduced and developed for function spaces including $D_E[0,\infty)$, the space of càdlàg paths on the set $[0,\infty)$. The properties of collections of functionals were then explored, particularly focussed on the notions of separating points, strongly separating points, and being pointwise convergence determining, and the connections between these concepts. Some propositions that were important to later developments were here introduced, including useful knowledge about the basis $\mathbb{B}_{\mathcal{M}}$ and the pseudometric $\rho_{\mathcal{M}}$. Most importantly, the homeomorphism Γ inspired by Bhatt and Karandikar was explained and related to the other concepts.

In the core of the thesis, three fundamental theorems from Tychonoff, Stone and Weierstrass, and Kuratowski were provided and used to prove the two main theorems. The first theorem, using the homeomorphism Γ, described conditions under which a collection of functionals on a topological space would be convergence determining. These conditions were too strict to be used directly in the application; however, the second main theorem used this weak convergence under stringent conditions to prove that a collection would be separating under less severe requirements.

These latter, separating requirements were satisfied in the second filtering problem, showing that probability measures on paths in the space in question could be determined from each other, and allowing the possibility of continuing the process of defining filtering equations and attempting to approximate these equations.

4.3 Future possibilities

Having a collection that will separate probability measures is only the first step towards constructing an applicable filter for a given problem. The next main task is to find further useable conditions under which a collection will be convergence-determining, so that a start can be made to proving the existence of solutions to various filtering equations. The results from Chapter 3, in particular the two main theorems, provide a basis from which such work could proceed.

Bibliography

[1] David J. Ballantyne, Hubert Y. Chan, and Michael A. Kouritzin. Novel branching particle method for tracking. In Oliver E. Drummond, editor, *Signal and Data Processing of Small Targets 2000*, volume 4048, pages 277–287. SPIE, 2000.

[2] Abhay G. Bhatt and Rajeeva L. Karandikar. Invariant measures and evolution equations for Markov processes characterized via martingale problems. *Ann. Probab.*, 21(4):2246–2268, 1993.

[3] Abhay G. Bhatt and Rajeeva L. Karandikar. Weak convergence to a Markov process: the martingale approach. *Probab. Theory Related Fields*, 96(3):335–351, 1993.

[4] Patrick Billingsley. *Probability and measure.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, third edition, 1995. A Wiley-Interscience Publication.

[5] Patrick Billingsley. *Convergence of probability measures.* Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second edition, 1999. A Wiley-Interscience Publication.

[6] Douglas Blount and Michael A. Kouritzin. On tightness for probability measures on $D_E[0,\infty)$ and $D_E[0,T]$. 2005. Under final preparation.

[7] Douglas Blount and Michael A. Kouritzin. On weak convergence and separation of probability measures. 2005. Under final preparation.

[8] Richard M. Dudley. *Real analysis and probability.* The Wadsworth & Brooks/Cole Mathematics Series. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA, 1989.

[9] Stewart N. Ethier and Thomas G. Kurtz. *Markov processes.* Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986. Characterization and convergence.

[10] Theodore W. Gamelin and Robert Everist Greene. *Introduction to topology.* Saunders Series. Saunders College Publishing, Philadelphia, Pa., 1983.

[11] Ronald F. Gariepy and William P. Ziemer. *Modern real analysis.* PWS Publishing Company, Boston, Massachusetts, 1995.

[12] R. E. Kalman and R. S. Bucy. New results in linear filtering and prediction theory. *Trans. ASME Ser. D. J. Basic Engrg.*, 83:95–108, 1961.

[13] Ioannis Karatzas and Steven E. Shreve. *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics.* Springer-Verlag, New York, second edition, 1991.

[14] Michael A. Kouritzin, David J. Ballantyne, Hyukjoon Kim, and Yaozhong Hu. On sonobuoy placement for submarine tracking. In Ivan Kadar, editor, *Signal Processing, Sensor Fusion, and Target Recognition XIV*, volume 5809, pages 244–252. SPIE, 2005.

[15] James R. Munkres. *Topology: a first course.* Prentice-Hall Inc., Englewood Cliffs, N.J., 1975.

[16] Kalyanapuram R. Parthasarathy. *Probability measures on metric spaces.* Probability and Mathematical Statistics, No. 3. Academic Press Inc., New York, 1967.

[17] Sidney I. Resnick. *A probability path.* Birkhäuser Boston Inc., Boston, MA, 1999.

[18] George F. Simmons. *Introduction to topology and modern analysis.* Robert E. Krieger Publishing Co. Inc., Melbourne, Fla., 1983. Reprint of the 1963 original.

[19] Moshe Zakai. On the optimal filtering of diffusion processes. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 11:230–243, 1969.