
THE PRINCIPLE OF LEAST ACTION
IN STOCHASTIC PROCESSES

MARIANO CARUSO



UNIVERSIDAD
DE GRANADA

Editor: Universidad de Granada. Tesis Doctorales

Autor: Mariano Guillermo Caruso

ISBN: 978-84-9163-557-4

URI: <http://hdl.handle.net/10481/48334>

IN MEMORIAM

De los verdaderos oponentes, cuya concienzuda lectura y posterior crítica de este trabajo dará lugar a mejores ideas.

DEGRADECIMIENTOS

Sería injusto comenzar agradeciendo, puesto que lo que hoy tengo no me ha sido dado gratuitamente. Ninguna palabra de mi parte podrá devolver todo que se me ha dado para que pueda llegar a esta instancia. Por tanto, antes que enunciar palabras de gran carga emotiva quisiera pedir perdón. Mis más sentidas disculpas por todo lo que he hecho pasar a mi esposa, a mis hijas y a mi familia, para que este trabajo pueda llegar a término.

Algo respetable y merecedor de agradecimiento en este tiempo han sido los problemas. Tengo claro que éstos han sido, son y serán los grandes motores de los cambios; de lo contrario, a falta de problemas todo sería similar a un estado muy simple de describir: el equilibrio, i.e. la muerte.

Agradezco de manera especial a mi directora de tesis, la Dra. Cecilia Jarne, quién le ha confiado a este trabajo el carácter de *habeas data*. También al jurado de este trabajo, más conocido como el *tribunal pentagonal*, o trinidad de cinco personas, mucho más que un *oxímoron cardinal*, quién me ha otorgado el *habeas corpus*.

Estoy realmente agradecido, eso sí, por la historia que se fue tejiendo en este tiempo y quisiera aludir a aquellos que forman parte de esta entretela y que son hijos de *la gratuidad*. Principalmente a mi familia, mi comunidad, quienes me padecen de forma sincera, esto es poniéndome en mi lugar; y finalmente a todos los demás, que podrán conocer solo mi cara amable y sonriente.

PRÓLOGO

El temario que desarrollaremos en esta tesis se resume bajo el lema: *Todo lo que usted nunca quiso saber sobre procesos estocásticos y, sin embargo, le contaron.*

Comenzamos el trabajo con una introducción para orientar la temática del presente trabajo, justificándola histórica y lógicamente, en el capítulo 1.

En el capítulo 2 daremos una descripción sobre cómo pueden los principios variacionales englobar una gran cantidad de problemas, e incluso abordar otros nuevos. Utilizaremos como modelo la propia historia. Mostraremos concretamente la secuencia de ideas que llevaron a Fermat a enunciar lo que podríamos llamar hoy el *principio optimal primigenio* sobre la propagación de la luz, y que posteriormente inspirase a Bernoulli la forma de resolver otro problema, ahora de la mecánica. Esto servirá para establecer el *dispositivo* del que haremos uso en el resto de la tesis, a fin de construir lo que nos hemos propuesto, insinuado ya desde el título de este trabajo.

En el capítulo 3 daremos una breve introducción de la teoría de procesos estocásticos *a la* Kolmogorov. También abordaremos los procesos de Markov, pues éstos describen una grandiosa variedad de sistemas físicos, químicos, biológicos y económicos.

En el capítulo 4 daremos una breve introducción al cálculo variacional y a una formulación lagrangiana estándar. Probaremos que es posible utilizarla para reformular la dinámica que proviene de la teoría de procesos estocásticos de Kolmogorov. Nos referimos a aquellos procesos que satisfacen las ecuaciones de Chapman-Kolmogorov. Daremos una forma concreta para el lagrangiano, comenzando con el caso en que las variables aleatorias toman valores en un conjunto numerable. Incluiremos una

descripción dual de las ecuaciones de Kolmogorov, denominadas *forward* y *backward*, unificadas en un mismo objeto lagrangiano. Esta formulación nos lega una estructura simpléctica, i.e. una posible formulación geométrica de los procesos estocásticos. Junto con ella será posible construir incluso un formalismo hamiltoniano de los procesos estocásticos de esta clase, con un breve paso por la geometría simpléctica. Finalmente daremos los pasos para la construcción de densidades lagrangianas para el caso en que las variables aleatorias sean continuas, lo que constituirá una teoría de campos de los *procesos estocásticos kolmogorovianos*.

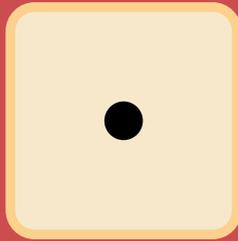
En el capítulo 5 introduciremos los conceptos de covarianza de las ecuaciones de evolución, heredadas de la formulación lagrangiana. Estudiaremos la posibilidad de conectar dos procesos cualesquiera de esta categoría vía una transformación local o gauge. En otras palabras, probaremos que los procesos *kolmogorovianos* son equivalentes uno a uno. Daremos paso a las ideas de nuestro artículo *Equivalent Markov processes under gauge group*, *Phys. Rev. E.* **92**, 052132.

En el capítulo 6 construiremos un modelo, con aplicación a la biología, para poner a prueba la hipótesis acerca del uso práctico de las transformaciones gauge. Dicho modelo trata sobre el cálculo del número de ancestros de una especie que se reproduce de manera biparental. Haremos referencia a nuestro artículo *Markov chain approach to the distribution of ancestors in species of biparental reproduction*, *Phys. Rev. E.* **90**, 022125.

Ambos artículos poseen un *impact factor* igual a 2.252.

Como corolario de estos 6 capítulos finalmente se presentan las conclusiones que dan cierre a estas facetas o *caras*. También expondremos algunas de sus implicaciones, esto es, nuevas cuestiones que quedarán en suspensión para un futuro siempre abierto y que darán lugar a la construcción de nuevos cuerpos de conocimiento.

INTRODUCCIÓN



1

INTRODUCCIÓN

Una de las principales tareas de la física teórica y de la ciencia en general, trata sobre la unificación de sus partes. No en tanto la reunión de teorías (o ideas dispersas) en otra que las contenga, sino en un sentido más amplio y trivial. Debemos, en primer lugar, conocer para poder condensar ideas. Esto explica que, en general, la ciencia atraviese primero un período de análisis para luego poder desarrollar la síntesis. Toda *comprensión* se basa en una posterior *compresión* de ideas. Pero ¿qué es comprender? Al menos para las ciencias, con la inclusión de la matemática, diremos en primer lugar lo que no es: comprender no es describir. Eso lo hacemos cuando no tenemos más remedio. En términos comparativos agregaremos también que *comprensión* (de comprender) se corresponde a una *compresión* (de comprimir). Se trata de ganar inteligibilidad sobre el objeto de estudio, separando la esencia de sus matices.

Supongamos dos listas que surgen de conjugar los elementos de $\{-1, 1\}$. En primer lugar la secuencia alternada *ad infinitum*: $\{1, -1, 1, -1, \dots\}$, la cual es susceptible de comprensión, o comprensión, a través de la regla $\{(-1)^n : n \in \mathbb{N}\}$. En segundo lugar la secuencia al azar, e.g. $\{1, -1, 1, -1, 1, 1, -1, \dots\}$, que no puede comprimirse por tanto no podemos más que describir la lista por extensión, no tenemos regla capaz de decir *y así sucesivamente*.¹ La primera secuencia tiene un patrón de naturaleza determinista, la

¹Estos dos tipos de secuencias presentados refieren también a las dos formas conocidas y generales de expresar un conjunto, a saber: comprensión y extensión, respectivamente.

segunda quizá también tenga algún patrón o regla, pero podemos afirmar que de distinta naturaleza. En cualquier caso, lo que buscamos son patrones, incluso en procesos que ocurran al azar. Diremos además que cuanto mayor sea el grado de *compresibilidad* de una afirmación, esta requiere un menor número de *bits* para describirla, tal y como ocurre en el ejemplo anterior de las dos secuencias de números. A fines comunicativos, el grado de compresibilidad de una afirmación está ligado al grado de complejidad.

Comprender es separar por *decantación* todo lo superfluo de todo lo que no se comparte. En este sentido, los principios variacionales han dado a las teorías físicas actuales el poder de conjugar las reglas que surgen de una mínima expresión que explica lo máximo compartido. Llegamos así a la concepción de los principios variacionales como compresores de las leyes, al menos y hasta el presente trabajo, para las teorías de la física.

Un concepto importante en este contexto es el de *funcional*. Por un lado, tenemos a las funciones, aquellos objetos definidos para cada valor de una variable que pertenece a cierta región, de manera que para cada valor de la variable le corresponde un valor de la función; sabemos también que existen funciones de varias variables. Por el otro, las funcionales son, por extensión nominal, aquellos objetos que quedan determinados a partir de una o más funciones de cierta clase [1].

Respecto de la física, grandes mentes han *analizado* las reglas del juego que otras mentes semejantes han *sintetizado*, posteriormente, en forma de *principios optimales*. Estos, a su vez, han surgido de un área propia de la matemática: el cálculo de variaciones.

Los primeros indicios del uso de tales principios aparecen con P. Fermat en 1662, respecto al problema de la propagación de la luz: “*la trayectoria que sigue la luz al propagarse desde un punto A hacia otro B, ambos fijos, es tal que el tiempo empleado en recorrerla es mínimo*”. En aquellos años no fue enunciado utilizando el aparato matemático del cálculo de variaciones. Sin embargo este princi-

pio comprende el germen de un área de la física y la matemática, pues plantea la posibilidad de deducción de un conjunto de reglas (leyes) invocando la optimización de alguna cantidad. En física esta cantidad recibió posteriormente el nombre de funcional *acción*; que para el caso de Fermat corresponde precisamente al tiempo de propagación de la luz, dado un camino posible entre los dos puntos fijos. Esto es, la acción de Fermat puede ser medida experimentalmente.

Posteriormente el cálculo de variaciones fue desarrollándose, siempre a partir de problemas concretos, normalmente inspirados por la física. En 1696, J. Bernoulli² abordó analíticamente algunas de estas ideas a partir del problema de la curva *braquistócrona* (gr. *brákhistos* “el más corto”, *khronos* “tiempo”). Este último surge de la pregunta sobre la curva que una dos puntos fijos A y B (no alineados verticalmente) que posea la propiedad de que un punto material, sometido solo a un campo gravitatorio constante, se deslice en el menor tiempo posible. De nuevo se intenta optimizar una funcional que, en este caso, es también el tiempo que toma el punto material en deslizarse de A a B. La solución a este problema es la curva denominada cicloide [1].

Más tarde L. Euler (1733) con su *Elementa Calculi Variationum* otorgó un marco concreto, dando nombre y métodos propios a esta disciplina matemática, también J. L. Lagrange contribuyó extensamente a la teoría [2]. El *modus operandi* consiste en obtener las denominadas ecuaciones de Euler-Lagrange, cuya solución optimiza a la funcional dada.

En la física contamos con un principio de optimización subyacente que afirma que dado un sistema físico, de todas sus posibles formas de evolución temporal, compatibles con ciertas condiciones que detallaremos posteriormente, éste adopta o *escoge* aquella(s) que hace(n) estacionaria a la funcional acción, *S*. Permitimos una pluralidad de funciones como solución al problema

²Nos referimos a Johann (1667 – 1748), de la familia Bernoulli de grandes artistas y científicos.

de optimizar la acción S , puesto que puede no ser única.

Respecto de la dinámica de sistemas (en tanto evolución temporal), estos principios permiten *comprimir* las leyes que gobiernan su evolución temporal. Diremos que de todas las formas de evolución temporal de un sistema, solo ha(n) de ser posible(s) aquella(s) que haga(n) que la funcional acción sea estacionaria. A partir de este principio podemos deducir las leyes que rigen la evolución temporal del sistema bajo estudio. Las otras formas de evolución no corresponden a la realidad del mundo. Este principio provee, por tanto, una forma de selección de leyes fundamentales, o *selección fundamental* [3], para abreviar.

Diremos además que dada una acción S , lo que hace un principio variacional $\delta S = 0$, con ella es dotar de equilibrio a las leyes fundamentales, que se deducen de éste. Nos referimos aquí al equilibrio de la ley *per se* en tanto a su conservación, no a las eventuales propiedades de una solución de esa ley. El principio de *acción estacionaria* se suele enunciar como de *acción mínima*, i.e. la acción adquiere un valor mínimo para aquellas candidatas a leyes fundamentales. En ese caso diremos que los principios variacionales dotan de estabilidad a las leyes que de él se deduzcan; nos referimos de nuevo a la ley en sí misma, no a las situaciones reguladas por la ley [3]. En resumen, $\delta S = 0$ puede ser visto como la invocación de la conservación de leyes fundamentales. Esto explica el mecanismo de *innovación* utilizado hasta el presente en la física fundamental, basado preferentemente en la modificación de S . Tomemos como ejemplo una categoría: la teoría cuántica de campos, toda innovación allí mediante se realiza mediante la incorporación de los denominados términos de *nueva física*, que afectan a la acción primigenia sobre la que queremos realizar algún tipo de innovación.

Dada la potencia de los principios variacionales, gran parte de las teorías actuales de la física los han convertido, a su tiempo, en parte de sus fundamentos³. La primera de ellas, debida en gran

³De hecho toda teoría que admita algún principio variacional debería, quizá,

parte a Lagrange, es el caso de la mecánica analítica, una formulación alternativa a la de Newton, pero ciertamente más poderosa. En la mecánica analítica se parte de una funcional acción, la cual está definida como la integral de otra cantidad denominada *lagrangiana*. Esta formulación fue ampliamente extendida a las teorías clásicas de campos, como el electromagnetismo de Maxwell, la gravitación de Einstein, la mecánica de los medios continuos y la mecánica de fluidos. Incluso a la mecánica cuántica, en sus versiones relativista y no relativista, y finalmente a las teorías cuánticas de campos.

Existe en la naturaleza una idea trivial, no por elemental, sino por *fundamental*, sobre la cual están montadas todas las áreas de la física, y que nos permite comprenderla, pues *comprender en ciencia es encontrar lo común entre lo diverso* [3]. Estamos hablando del principio variacional $\delta S = 0$ que se halla en el tronco del árbol de las ciencias físicas. En [3] se ha discutido la extensión de $\delta S = 0$ al caso en que S represente otra cantidad: la entropía, a fin de incluir a la termodinámica como otra de las ramas de este gran árbol.⁴ De esta forma podemos incluir otras ramas tales como la termodinámica del equilibrio, termodinámica lineal del no equilibrio, termodinámica no lineal del no equilibrio, teoría cinética de los gases, mecánica estadística clásica del equilibrio, mecánica estadística clásica del no equilibrio, mecánica estadística cuántica, mecánica estadística relativista, etc.

Confiar el rol de *codificadores de leyes fundamentales* a los principios variacionales de esta clase, consigna ciertos privilegios implicados. Centraremos nuestra atención en una formulación *lagrangiana*. Automáticamente obtenemos lo que se denomina *covarianza* de las leyes subsidiarias, es decir que éstas son independientes de las coordenadas que se elijan para describir al sistema. Eventualmente tendremos cantidades conservadas que provenirán de cierto tipo de invariancia de la teoría [4] y existe una

incluirlo en su axiomática.

⁴En tal caso $\delta S = 0$ representa un principio diferencial.

forma de describir la interacción de subsistemas de una categoría dada.

Hemos intentado justificar los *porqué* utilizar principios variacionales en la física mostrando algunos *como*, es decir, denotando las bondades en cuanto a unificación y deducción automática de las leyes o ecuaciones de evolución de la teoría de la que sean invocados. Aunque podemos hablar de uno de los *porqué* directamente, acerca de la estacionaridad (o extremalidad) de la acción en mecánica clásica. Esto fue gracias a otra *compresión*, en este caso por conexión entre la mecánica clásica y la mecánica cuántica, inspirada por Dirac [6] y desarrollada completamente por R. Feynman desde su tesis doctoral y posteriormente en [5]. Se trata del formalismo *path integral* que hizo su entrada en mecánica cuántica y se utiliza hoy en las teorías cuánticas de campos. El principio de acción estacionaria (o extremal) utilizado sin más en la teoría clásica, encuentra su justificación desde la teoría cuántica. Según lo expuesto podemos hablar con toda justicia de *la naturaleza variacional de la ley física*[7].

No todo lo que brilla es oro, pues debemos decir que los principios variacionales en si mismos se suceden de las reglas. Ya hemos dicho que éstos pueden ser considerados, en cierta medida, como parte de la síntesis del proceso de formalización de una teoría. De hecho los principios variacionales están supeditados en gran parte a las reglas que consiguen comprimir. Normalmente los fundamentos provienen posteriormente al conjunto de reglas que se han verificado por la observación. Precisamente por el hecho de que, en general, primero se *observa*, luego se *legaliza* y más tarde se *fundamenta*. Tenemos entonces una sujeción o dependencia de los principios variacionales al conjunto de reglas. Esto no es más que la manifestación del diálogo entre experimento⁵ y teoría, que debe mantenerse abierto, siempre y cuando querramos

⁵Cuando hablamos de experimento no nos referimos a la naturaleza, esta última está ahí afuera y comprenderla, en algún sentido, es objeto de la ciencia a través de la experimentación.

formar parte de un sistema científico vivo.⁶

Para poner un ejemplo citemos nuevamente el trabajo de Fermat. El principio de *tiempo mínimo* que él enunció explica las reglas, de reflexión y refracción entre otros, pero no constituye en sí mismo una idea más fundamental, en el sentido en que deja un sin número de preguntas abiertas, por ejemplo: ¿cómo *sabe* la luz qué camino tomar frente a un obstáculo? El principio de *tiempo mínimo* no parece ser tan intuitivo. Nosotros establecemos una condición, fijando los puntos A y B, y es la luz la que *se las arregla* para minimizar el tiempo de vuelo, incluso colocando objetos en medio. Ello se debe a que estamos tratando con lo que se denomina *óptica geométrica*, que es una aproximación dentro de una teoría más fundamental. Estamos pues, haciendo preguntas en la periferia del dominio de tal teoría, cuyas respuestas saltan y caen fuera de ésta. Para explicar a un nivel más profundo estas cuestiones hemos de adentrarnos en los extraordinarios paisajes de la dinámica cuántica, donde reinan las probabilidades [8, 9].

No pretendemos axiomatizar las teorías *solamente* a partir de principios variacionales, pues es lógicamente imposible ya que se requieren definir más elementos. Por ejemplo, las teorías actuales de la física, incluyen una serie de axiomas sobre los objetos que contienen, además de una semántica propia y una interpretación [4]. Por tanto, no es posible hacer una reducción íntegra de ninguna teoría a principios variacionales, no obstante podremos considerarlos como parte del conjunto de tales axiomas.

Respecto de los fundamentos de la propia de teoría de probabilidades, debemos indicar que uno de los 23 problemas del milenio, planteados por Hilbert⁷ en el año 1900, fue el de axiomatizar aquellas ramas de la física en los que la matemática juega un papel importante [10]. En aquella época había dos grandes

⁶Es más, hoy en día es necesario además un diálogo *inter experimental e inter teórico*.

⁷Realmente en la conferencia del 8 de Agosto de 1900 los problemas publicados fueron 10, los 13 restantes fueron agregados posteriormente.

candidatas: la mecánica clásica y la teoría de probabilidades.⁸

La primera de ellas fue iniciado con la formulación variacional por extraordinarios matemáticos: Lagrange, Euler, Hamilton, Jacobi, Poisson entre otros. La formulación axiomática más basta se debió a G. Hamel, alrededor de 1901. Pero lo que aquellos conformaron fueron las bases de una axiomatización *geométrica* de la mecánica, en la que contribuyeron además otros grandes como Poincaré y Cartan. Actualmente todas las teorías físicas actuales descritas por principios variacionales pueden abordarse desde la geometría.

Mientras que la axiomatización de la teoría de probabilidades fue alcanzada con éxito por el matemático ruso Andrei Kolmogorov [11] (alrededor del año 1933), lo que permitió incluirla como un área formal de las matemáticas, cuyo dominio de aplicación esta hoy presente en todas las ciencias.

Históricamente el estudio hacia lo fundamental fue tomando una dirección hacia escalas espaciales de la naturaleza del mínimo tamaño posible. El hecho de que tal escala mínima represente una posibilidad que conquistar, abrió un largo capítulo respecto de la indagación hacia escalas cada vez menores. Parece que lo fundamental está asociado con aquello trivial o minimal. El estudio de los fenómenos en tales escalas, cada vez más pequeñas, dieron lugar al descubrimiento de objetos con propiedades no bien definidas, es decir, no locales o extendidas en todo el espacio. Este es el caso de la posición de objetos cuánticos, por citar un ejemplo. En todos ellos encontramos que la forma de responder a determinadas preguntas es a través de una medida de probabilidad. La propia teoría nos dice cómo evoluciona dicha medida en el tiempo. Sin embargo, la concepción de probabilidad no es exclusiva de las teorías cuánticas, sino que subyace a muchas teorías físicas. En el caso de las teorías clásicas, las probabilidades tienen su origen en la mecánica estadística.

⁸Esto ilustra que, por aquella época, la teoría de probabilidades no era considerada como una rama de la matemática, sino de la física.

Anteriormente hemos dicho que a fin de cuentas queremos entender más acerca del mundo que nos rodea, sea este conformado por entes tanto inanimados como vivientes. Siempre desde el punto de vista de la optimización podemos considerar también otros desplazamientos, ahora de seres vivientes. Sea este el caso de un pez que se desplaza en el agua de manera óptima, esto es, sin invertir prácticamente energía alguna en generar espuma. Por el contrario, los peces que generan espuma desperdician no solo parte de esa energía en calentar el agua, sino que, aún peor, definen muy bien un camino que puede conectarlos con un posible predador. En este contexto, las razones que permiten tal nivel de optimización quizá sean distintas a las que gobiernan el mundo inanimado, sin embargo subyace un principio de optimización en la naturaleza toda. Un observante del mundo natural podrá comprobar que, a partir de otras reglas quizá, la naturaleza escoge caminos que optimizan alguna cantidad. Esto no solo ha de verificarse en teorías del *mundo inanimado*, sino también en una enorme y diversa cantidad de fenómenos del *mundo viviente*.

Llegamos así al tema medular de esta tesis: una serie de trabajos elaborados por Kolmogorov y paralelamente por el matemático británico Sydney Chapman, dieron lugar a unas ecuaciones fundamentales en el campo de los procesos estocásticos. Cuya solución describe la evolución temporal de las medidas de probabilidad asociada al proceso estocástico. Pretendemos dar un origen variacional, o como hemos señalado de selección fundamental, de las leyes que rigen los procesos estocásticos.

Introduciremos por primera vez una descripción lagrangiana para procesos estocásticos, tomando como objetos dinámicos a las medidas de probabilidad. Esto implica no solo una construcción formal alternativa a la propuesta por Kolmogorov, sino que además será posible abordar esta teoría con nuevas preguntas y herramientas: tomar parte del abanico de técnicas y métodos de otras teorías de la física para traducirlas y exportarlas hacia la teoría de procesos estocásticos. Este procedimiento es hijo la des-

cripción de ambas teorías a través de un lenguaje común, permitirán abordar problemas desde otra perspectiva diferente y sus frutos habrán de recogerse en lo sucesivo. Emplearemos los conceptos de simetrías, en particular simetrías locales (gauge), herramientas poderosas sobre las que están diseñados, los modelos actuales que describen las interacciones fundamentales de la materia. En este caso, se trata de otros grupos locales que permiten conectar todos los procesos estocásticos de una categoría dada.

Estos conceptos pueden ser ahora aplicados con toda rigurosidad para ampliar las fronteras actuales de esta disciplina. Tal es el caso del estudio de la interacción de procesos estocásticos, que puede implementarse haciendo uso del lagrangiano.

Los procesos estocásticos son inherentes a cualquier sistema complejo. Se encuentran en gran variedad de situaciones en áreas que trascienden a la física, tales como la economía y la biología matemáticas. Por este motivo pretendemos dar una introducción a esta formulación lagrangiana de los procesos estocásticos, sin perder de vista ejemplos específicos de aplicación. Trataremos una situación particular con aplicación a la biología, en la que condensarán algunas de estas ideas.

El camino hacia una comprensión más fundamental de la naturaleza nos ha conformado dos visiones muy concretas: **1** una imagen probabilística de la misma. Parece que la forma de abordarla a ese nivel sea dando lugar a la contingencia. También hemos señalado otro aspecto sustancial de la misma a saber **2** los principios variacionales atraviesan todas las áreas formales de la física y dejan cierto vestigio hacia algunos aspectos de las ciencias de la vida. Ambas visiones fundamentales admiten un lenguaje matemático dado y esta tesis pretende vincularlas, arguyendo sobre los *fundamentos variacionales* de la propia *teoría de probabilidades*. Para ser más precisos una descripción de los **procesos estocásticos** vía **principios variacionales**.

El siguiente principio subyace a esta investigación, a saber: dadas dos disciplinas, mientras esperamos que una nueva idea

transforme a alguna de estas áreas en cuestión, podemos mezclar conceptos estableciendo puentes entre ambas. Como corolario de ello tenemos la posibilidad de transitar en ambos sentidos trasladando interpretaciones y conceptos de una a otra. Dado un concepto de una disciplina podemos transformarlo para ver cómo luce, i.e. reinterpretarlo, desde la otra.

De la misma forma que en los sistemas complejos tenemos propiedades emergentes, que surgen de la mutua influencia (esto es, cuando dos o más partes de un sistema interactúan), es de esperar que en el sistema en el cual los científicos están inmersos (que es también complejo), por tanto, surjan nuevas ideas a partir de la interacción de dos o más áreas de la ciencia. Eventualmente estas ideas podrán decantar en otra área de la ciencia.

Finalmente otro punto de vista, quizá más nostálgico, completa esta introducción. En su trabajo "Space-Time Approach to Non-Relativistic Quantum Mechanics" [5], Richard Feynman escribió "*existe cierto placer en reconocer viejas cosas desde un nuevo punto de vista*", en el contexto de una nueva formulación de la mecánica cuántica. Lo cierto es que sus ideas trajeron, no solo la posibilidad de realizar cálculos con mayor eficiencia, sino una nueva forma de poder visualizar los procesos cuánticos. Estas palabras permiten delinear otro aspecto de la filosofía de la presente investigación puesto que se trata de una alternativa a la descripción estándar de los procesos estocásticos a la Kolmogorov, pero con toda una riqueza que los métodos y enfoque variacionales poseen. La introducción de una *geometrización simpléctica* de la teoría de éstos procesos junto con el concepto de *simetría* dará un marco diferente desde donde contemplarla y estudiarla.

PRIMEROS PRINCIPIOS OPTIMALES



2

PRIMEROS PRINCIPIOS OPTIMALES

Normalmente el camino hacia la construcción de reglas (leyes naturales) surge de la observación experimental o teórica. Posteriormente quizá se tenga un conjunto de reglas que explican toda una fenomenología. Este conjunto se puede *comprimir* cada para ganar una mayor *comprensión* del problema. La forma de comprensión, hemos dicho, representa algún tipo de unificación que reúne el conjunto de esas reglas, en el sentido de que éstas podrán ser deducidas a partir de algo más fundamental.

Utilizando un ejemplo histórico mostraremos la transición de obtener una regla a partir de la observación. Finalmente obtendremos un principio variacional que englobe la regla, es decir que permite deducirla, y por último que ésta regla explica aquella primera observación. En resumen, mostraremos un ejemplo para la secuencia:

Observación \mapsto **Regla** \mapsto **Principio Fundamental**

Comenzaremos este capítulo a partir de los ya mentados trabajos de Fermat y de Bernoulli, primeros balbuceos del cálculo variacional, para satisfacer lo que nos hemos propuesto anteriormente.

El principio de Fermat, tal como lo hemos enunciado de manera simplificada¹, encierra gran parte de la óptica *geométrica*.

¹Este principio nos permite *entender* a la luz y podemos verificarlo experimentalmente, incluso en los casos no triviales, i.e. si en la trayectoria que ésta siga tenemos cambios del medio, i.e. interfaces. Pero no explica la *aparente magia* subyacente, en la que fijamos los puntos A y B en cada medio y la luz es *quién sabe* qué camino tomar para minimizar el tiempo de vuelo.

Por otro lado, el primero de los problemas optimales para la evolución temporal de puntos materiales aparece con Bernoulli.

En la terminología y formulación moderna del cálculo de variaciones, los enunciados de Fermat y Bernoulli corresponden sintéticamente a $\delta S = 0$, donde la acción S puede ser el tiempo de desplazamiento de la luz o del punto material, respectivamente.

Existe una interesante conexión histórica entre los problemas de la luz y de la curva braquistócrona, a partir de la propiedades sobre la curva cicloide. Fue Bernoulli quién utilizó las ideas de Fermat, buscando una curva tal que se verifique una suerte de *ley de las refracciones* de Snell, a fin de emular el comportamiento *óptimo* que la luz exhibe naturalmente.

Veamos en primer lugar el problema de Fermat. Si bien hoy sabemos que este problema pertenece al dominio de la óptica geométrica, que es una aproximación de una teoría más fundamental, no vamos a describirlo con los resultados del *diario del lunes*², sino de forma cronológica o histórica.

Se ha observado que cuando la luz se desplaza de un punto a otro lo hace siguiendo líneas rectas, en ausencia de obstáculos. No parecen haber interferencias entre los rayos de un haz de luz. Entonces podríamos preguntarnos por lo que ocurre con ella cuando golpea otros objetos, es decir, cuando tenemos un cambio de medio. El primer objeto a considerar por los antiguos fue el espejo. Se observa que la luz llega al espejo y no continúa en línea recta sino que se refleja en una nueva línea recta que cambia a merced de la inclinación del espejo respecto del haz de luz. Este problema es el de la *reflexión* de la luz y fue resuelto antes que Fermat. Se puede probar que los ángulos (medidos desde la normal a la superficie del espejo) de los rayos incidente (θ_i) y reflejado (θ_r) satisfacen la siguiente relación

$$\theta_i = \theta_r \tag{2.1}$$

²Frase alusiva a la argumentación *banal* sobre el pasado a partir de resultados en el presente, preferentemente utilizada en el fútbol. Donde diga *diario* debe decirse *periódico*.

En la figura 2.1 vemos un esquema sobre la llamada ley de reflexión.

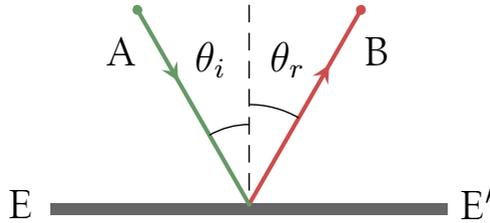


Figura 2.1: Los rayos *incidente* y *reflejado* forman ángulos θ_i y θ_r , respecto de la normal a la superficie E–E', tales que $\theta_i = \theta_r$.

El problema se torna más complicado cuando la luz se propaga de un medio a otro, por ejemplo desde el agua hacia el aire, aquí tampoco observamos una única línea recta, sino dos; cada una de las cuales define un ángulo respecto de la línea normal a la interfaz, ver figura 2.2. Si el ángulo de incidencia en el agua es pequeño, la desviación del rayo en el aire no es considerablemente grande. Pero si el ángulo de incidencia aumenta, entonces la desviación se hace muy grande. La relación entre estos ángulos fue un interrogante para los antiguos. Claudio Tolomeo elaboró una lista de pares correspondientes de ángulos en agua y en aire. Aunque dicha lista fuera elaborada alrededor del año 140, no fue sino hasta 1621 cuando el matemático holandés Willebord Snell encontró la *regla* que los conecta. En general, tendremos velocidades de propagación de la luz distintas (v_1, v_2) , para cada uno de los medios (μ_1, μ_2) , los ángulos serán también diferentes en general (θ_1, θ_2) y la regla que los involucra es

$$\frac{\text{sen } \theta_1}{v_1} = \frac{\text{sen } \theta_2}{v_2}. \quad (2.2)$$

La figura 2.2 esquematiza el problema de la refracción.

Como corolario de la regla de Snell tenemos que a mayores velocidades de propagación y mayores desviaciones respecto de

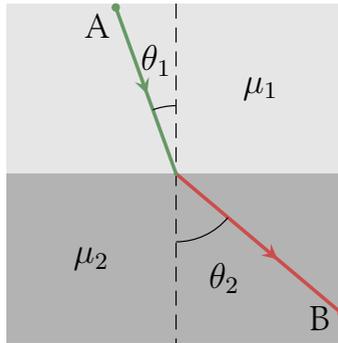


Figura 2.2: El rayo **incidente** tiene una inclinación menor que el rayo **refractado** puesto que la velocidades de propagación (v_1, v_2) en los respectivos medios (μ_1, μ_2) satisfacen $v_1 < v_2$.

la normal y también que el cociente $\text{sen}(\theta)/v$ es un número fijo e igual entre dos medios cualesquiera.

Fermat ensayó la idea del *tiempo mínimo*. Este principio explica la regla de las reflexiones y refracciones con éxito. Tal y como hemos adelantado existe una forma de deducir la ley reflexión (2.1) y la ley de refracción (2.2) a partir del principio establecido posteriormente por Fermat.

Existen muchas formas de probar esto, daremos una muy sencilla que unifica ambas reglas de *reflexión* y *refracción*. Un *hipotético* rayo de luz que se propague de A hacia B es *reflejado* en la interfaz, definida por la intromisión del espejo o *refractado* en la interfaz, definida por la intromisión de otro medio. Dado que los puntos A y B están fijos en ambos casos, el camino que seguirá ese *hipotético* rayo de luz depende de un único punto sobre la interfaz, digamos X, al que podemos asignar las coordenadas $(x, 0)$. Llamemos $\varrho(x)$ y $\varrho'(x)$ a las distancias entre A y X y entre B y X, respectivamente. De esta forma, el tiempo de vuelo de A hacia B, viene dado por una función solo de las coordenadas del punto X, denotada por $t(x)$. Para el problema de la reflexión $t(x) = [\varrho(x) + \varrho'(x)]/v$, mientras que para el problema de la re-

fracción $t(x) = \varrho(x)/v_1 + \varrho'(x)/v_2$. Ambas funciones $t(x)$ presentan un *mínimo* tal que se verifican las leyes de reflexión (2.1) y refracción (2.2), respectivamente.

Unos 30 años más tarde, Bernoulli ensayó una solución para el problema de la braquistócrona utilizando las ideas de Fermat. Pensó que si la luz misma minimiza el tiempo de vuelo, algo que imite este comportamiento sería una posible solución del problema. Según hemos dicho, si la luz se mueve de A a B, a través de un solo medio lo hace en línea recta, minimizando el tiempo y la distancia entre ambos puntos. Pero la curva que buscaba Bernoulli no era algo tan simple como el segmento que une A con B, de hecho ya mucho tiempo antes Galileo había encontrado que un arco de circunferencia garantizaba un menor tiempo que el segmento recto. Respecto a la propagación de la luz, hemos visto que al colocar un obstáculo en su camino, la trayectoria recta se fractura. Si el objeto interpuesto es tal que el rayo refractado se aparta aún más que el incidente $\theta_2 > \theta_1$ entonces tendremos que la velocidad de la luz en los medios μ_2 y μ_1 satisface $v_2 > v_1$, tal y como reza un corolario de la ley de Snell.

Si en alguna parte del camino desde A a B interpusiéramos un número finito de obstáculos para la luz, de manera que en cada interfaz ésta se mueva cada vez más rápido, quizá se consiga emular el comportamiento de un punto material cuya velocidad va *in crescendo*, pues está sometido solo a un campo gravitatorio constante. En el caso límite tendremos que la curva que buscamos es tal que una suerte de ley de Snell se satisface en *cada* punto de la misma, es decir $\text{sen}(\theta)/v$ es una constante para todos los puntos de la curva. En particular sean (x, y) las coordenadas para cada punto de la curva buscada y llamamos θ (por extensión del problema de la luz) al ángulo comprendido entre la vertical (normal a la superficie) y la recta tangente en ese punto (dirección del hipotético rayo). Además sabemos que a partir de la conservación de la energía mecánica tenemos que $v = \sqrt{2gy}$, siendo v el módulo del vector velocidad de la punto material (solo depende de

y) y g la aceleración provista por el campo gravitatorio uniforme. Bernoulli demostró que la curva que satisface que $\text{sen}(\theta)/\sqrt{y}$ sea un invariante para todo punto es la denominada *cicloide*.

La secuencia cronológica $\mathcal{O} \mapsto \mathcal{R} \mapsto \mathcal{PF}$, para este caso fue la siguiente **Tolomeo 140** \mapsto **Snell 1621** \mapsto **Fermat 1662**. Fue Bernoulli quién, en 1696, se inspiró en el comportamiento de luz para resolver el problema de la curva braquistócrona.

Este es solo un ejemplo de tantos en los que puede hacerse uso de las ideas de un área de la física o de la matemática en otra área, a través de un proceso de *deconstrucción*, o despiece, de sus conceptos y símbolos. Hemos ejemplificado este dispositivo con el problema de la propagación de luz y la curva bracistócrona, cuyas partes corresponden la una a la óptica geométrica y la otra a un problema de la mecánica clásica, respectivamente.

Para ver más claramente los elementos que tienen en común, podemos enunciar ambos problemas en un lenguaje moderno provisto por el cálculo de variaciones. El principio de Fermat³ puede reducirse a la optimización de la siguiente funcional

$$S = \int_{\gamma} n \, ds = \int_a^b n(\mathbf{x}(\zeta)) \|\dot{\mathbf{x}}(\zeta)\| \, d\zeta \quad (2.3)$$

donde \mathbf{x} representa las coordenadas de un punto de la curva γ , que está parametrizada por $\zeta \in [a, b]$, los puntos A y B tienen coordenadas $\mathbf{x}(a)$ y $\mathbf{x}(b)$, respectivamente y $\dot{\mathbf{x}}(\zeta)$ es la derivada total de $\mathbf{x}(\zeta)$ respecto de ζ .

La expresión (2.3) corresponde a una integral de línea, con s la longitud de arco, del índice de refracción del medio $n(\mathbf{x})$ en el que la luz se propaga. En el enunciado de Fermat según (2.3), la acción S es la longitud óptica, pero el cociente S/c es el tiempo de vuelo, puesto que $n(\mathbf{x}) = c/v(\mathbf{x})$, siendo $v(\mathbf{x})$ un campo de velocidades tal que para cada punto del espacio devuelve el módulo de

³Constituye en realidad versión extendida de este principio, válida para medios no homogéneos e isotrópicos en general, donde la luz se puede propagar en todas las direcciones.

la velocidad de la luz en un entorno de \mathbf{x} . Vale decir que $v(\mathbf{x})$ devuelve solamente el módulo puesto que estamos bajo la hipótesis de isotropía espacial.⁴

Los casos elementales de las figuras (2.1) y (2.2) podemos incluirlos trivialmente considerando a γ como una curva continua y no derivable en el punto de la interfaz, es decir el espejo, para la reflexión y además otra discontinuidad del campo $n(\mathbf{x})$ en la interfaz de las dos regiones μ_1 y μ_2 para la refracción. Haciendo uso del cálculo de variaciones sobre la acción (2.3) podemos deducir trivialmente las leyes de reflexión (2.1) y refracción (2.2).

El problema de Bernoulli también corresponde a la optimización una funcional similar a (2.3). La función escalar $n(\mathbf{x})$ sería una suerte de índice de refracción muy particular, de hecho $n = 1/v$, donde v es el módulo de la velocidad que solo depende de la coordenada vertical y . Reescribimos a $n(\mathbf{x})$ como $n(y) = 1/\sqrt{2gy}$, tal como hemos dicho en la página 25. El problema se trata de encontrar una curva plana que recorrerá el punto material, por tanto el problema se reduce a buscar cierta función $y(x)$. En ese caso la curva queda parametrizada por la misma coordenada x entonces $ds = \sqrt{1 + [y'(x)]^2} dx$. De esta forma la acción (2.3) toma la forma

$$S = \int_{x_a}^{x_b} \sqrt{\frac{1 + [y'(x)]^2}{2g y(x)}} dx. \quad (2.4)$$

En ambos casos la acción S es proporcional al tiempo que emplea la luz o el punto material para ir de A a B. Puesto que ambos provienen de una integral de línea de la forma

$$S \sim \int_{\gamma} \frac{ds}{v(\mathbf{x})}.$$

En resumen, lo que Bernoulli hizo fue abordar el problema del punto material a partir de una analogía con la luz, donde vale

⁴No confundir $v(\mathbf{x})$ con $\|\dot{\mathbf{x}}(\zeta)\|$, pues ζ no es necesariamente el tiempo, aunque sí son iguales para los \mathbf{x} sobre la curva γ , en el caso en que $\zeta = t$, i.e. bajo una parametrización temporal de γ .

un principio fundamental como el de Fermat. Continuando con lo expuesto en la página 12, en este contexto diremos que el principio de Fermat es *más* fundamental, no por ser históricamente anterior, sino por tener un dominio de aplicabilidad en toda la óptica geométrica, mientras que la búsqueda de Bernoulli corresponde solo a *una* trayectoria dentro de un contexto muy especial de la mecánica clásica con campo gravitatorio uniforme, etc. Este es otro de los puntos que debemos resaltar: el de la jerarquía de los sistemas descritos por principios variacionales, según el dominio de validez de las reglas que de éstos se deducen. Veamos un ejemplo extremo, si buscamos tales principios para teorías obsoletas por erróneas, como el caso de la física aristotélica, no podremos decir que ello la convierta o restituya en fundamental. En ciencia aquello fundamental está asociado a la noción de verdad científica, verdad que podemos derribar o falsear posteriormente. Sin embargo que la ciencia funcione o avance haciendo uso de la falseabilidad, no *mitiga* el trabajo por fundamentar nuestras reglas; por el contrario lo *motiva*. No dejamos de excavar para construir cada vez desde mejores bases, sabiendo que luego podrán ser derribadas. Tenemos aquí otra imagen, ahora funcional, acerca de los principios fundamentales que es la de ser soporte o *base* sobre la cual construir, o deducir, las reglas para las cuales fue invocado.

Hemos dicho que lo fundamental suele estar orientado allí donde las probabilidades juegan un rol principal. Por tanto, tenemos aquí otro enfoque del estudio de lo fundamental a través del estudio de lo estocástico. Debemos agregar que existen algunas áreas de la física y de la matemática que no han sido abordadas con un formalismo *lagrangiano* a fin de describir la dinámica planteada por el programa de Kolmogorov [11]; este es el caso de la mecánica estadística y el de los procesos estocásticos en general. Hacia esta clase de problemas va dirigido este trabajo de tesis. Se pretende incluir a la teoría de procesos estocásticos bajo el tipo de teorías que pueden describirse a partir de principios va-

riacionales, que son comunes al resto de las ya mentadas teorías actuales de la física.

En el capítulo siguiente daremos una breve introducción a los procesos estocásticos *kolmogorovianos*, i.e. aquellos descritos a través de las ecuaciones de Kolmogorov. Mostraremos que estas ecuaciones hacen referencia a una dualidad presente en la descripción de todo proceso estocástico: **prospección** y **retrospección**. En líneas generales, estamos hablando de sistemas cuyos estados (elementos de cierto conjunto E) son tales que dado uno de ellos, digamos a un dado instante **actual**, existe una gran variedad de formas de arribar desde (hacia) estados **anteriores** (**posteriores**).

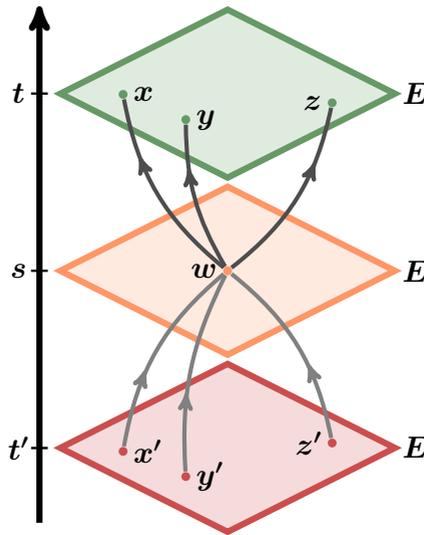


Figura 2.3: La figura muestra que dado un estado w del sistema al instante s podemos tener la inquietud de preguntarnos por su estado al tiempo $t > s$, o bien su estado al tiempo $t' < s$, i.e. construir su futuro o bien reconstruir su pasado, respectivamente.

Este es simplemente un adelanto de los temas que trataremos en lo sucesivo.

PROCESOS
ESTOCÁSTICAMENTE
DETERMINADOS



3

PROCESOS ESTOCÁSTICAMENTE DETERMINADOS

DOS CARAS DE LA MISMA MONEDA

Una de las pretensiones centrales en la física, e incluso de cualquier rama del conocimiento, trata sobre el estudio de los cambios de los objetos o mejor dicho los cambios de la propiedades de los objetos. En particular, sobre los cambios en el tiempo de las propiedades de los objetos.

Desde tiempos inmemoriales una de las cuestiones que ha preocupado al ser humano es el devenir, el futuro, desconocido para todos. Por esta razón quizá una de las pretensiones más fundamentales en este sentido sea la de predicción temporal, para intentar adelantar aquello que vendrá y mitigar el temor y la angustia. En otras palabras conocer el estado futuro de los elementos que forman parte de cierto sistema. Aunque el pasado quizá también necesite ser recuperado. Esto es importante, no solamente por nuestra precaria memoria ni por un sentimiento nostálgico, sino porque un intervalo de tiempo suficientemente largo hacia atrás nos deja también con un sin fin de preguntas difíciles de contestar. Lo que podríamos denominar el temor sobre lo que fue, que junto con el temor a lo que será, tienen una raíz común: el temor a lo desconocido. En este sentido, no sabemos muy bien cómo es que el universo¹ ha venido a *ser*, como continúa *siendo*, ni como dejará, eventualmente, de *serlo*. De esta

¹Cuando decimos *universo* nos referimos no necesariamente a éste como un todo, sino también a los subsistemas que dentro de éste podemos concebir como nuestro universo.

forma surgen los conceptos de **prospección** y **retrospección** que atañen al tiempo, i.e. a los fenómenos que ocurren *en* el tiempo. Ambos están ligados al de una secuencia de eventos, es decir que introducen igualmente a la noción de proceso. Estas son pues las dos caras del mismo proceso.

La formalización de la mecánica clásica² ha resuelto este problema pudiendo ejercer la predicción y retrodicción con igual eficacia, dado el estado actual de nuestro sistema. En estos sistemas existe cierta noción de determinismo que permite revelar tanto el futuro como el pasado del sistema [38].

Por otro lado, en formulación estándar de la mecánica cuántica tenemos cierto tipo de indeterminación, pues aunque la función de onda queda perfectamente determinada por la ecuación de Schrödinger, ésta en sí misma no puede determinar las propiedades todas del estado del sistema, por ejemplo no puede determinar aquellas propiedades denominadas *incompatibles* que se corresponden con operadores no conmutativos [6]. Sin embargo, sí es posible ofrecer cierto tipo de determinación en mecánica cuántica, la que proviene de la ecuación de Schrödinger sobre la función de onda, que es similar a la que pueden hacer las ecuaciones de Euler-Lagrange para una trayectoria, e.g. coordenadas y momentos, en mecánica clásica. En ambos casos se trata de ecuaciones diferenciales que pueden determinar perfectamente los estados del sistema: **clásico** e.g. *coordenadas y momentos* o bien **cuántico** e.g. *función de onda*; dado su conocimiento para un tiempo actual, todos ellos quedan perfectamente determinados para todo tiempo anterior y posterior. En términos más amplios, la descripción la evolución temporal de sistemas a través de ecuaciones diferenciales está obligada a definir a priori el conjunto de los **estados** y las **propiedades** de los objetos del sistema, así como designar también la manera en la que pueden variar en el tiempo los primeros y obtenerse a partir de ellos los segundos.

²Nos referimos aquí al mundo mesoscópico o intermedio y además simple. Un mundo que contiene bolas de billar, osciladores armónicos, etc.

De estas cavilaciones podemos decir alguna cuestión acerca del determinismo o indeterminismo. A primer vista, la visión del universo que cada científico pueda tener no es del todo independiente del dominio de investigación en el que trabaja. Distinguiamos ciertos niveles en la observación y experimentación científica: el *perceptible* por nuestros sentidos y simple, que es el que describe, por ejemplo, la mecánica analítica. En este nivel ya hemos dicho que la tendencia es hacia el determinismo. El *imperceptible* por pequeño, por grande o por complejo. En el fondo de todo científico hay una orientación hacia cierto tipo de determinismo, aunque sea por oficio, se cree en la reproducibilidad de las leyes. Pero eso atañe al método científico, donde se toman como hipótesis de trabajo la idea de que la naturaleza es inteligible, la idea de que se puede separar el sujeto del objeto, el fenómeno de aquel que lo observa. Por último un proceso dialéctico, un diálogo, entre experiencia y teoría que permite avanzar la ciencia; mientras que las teorías científicas sí que pueden contener ciertos rasgos de indeterminación.

Desde un punto de vista epistemológico tenemos a Laplace quién asociara la idea de azar a algo subjetivo, es decir a una incapacidad de cálculo o de nuestra ignorancia. Fue Poincaré quién más tarde comenzó a hablar de azar objetivo, refiriéndose a algo que es propio e irreductible en las teorías de la física, en particular, la mecánica estadística.

En general, las preguntas sobre la predicción y la retrodicción no corresponden al mismo tipo de respuestas. Para los procesos estocásticos, conocido el estado actual del sistema, no podemos garantizar cuál será el estado del mismo en un tiempo futuro. La misma pregunta vale intercambiando los términos futuro por pasado, aunque con una respuesta no trivial, es decir, existe una gran variedad de formas de arribar desde estados pasados al estado actual.

En términos abstractos, si $\xi(t)$ representa el estado del sistema al tiempo t . Supongamos que conocemos $\xi(s)$ y tenemos una regla

que nos permite conocer $\xi(t)$, para $t > s$. Luego no es cierto que usando siempre la misma regla, podamos afirmar que $\xi(t')$ con $t' < s$, sea el estado del cual proviene el sistema para luego llegar a $\xi(s)$ y finalmente a $\xi(t)$. Es decir que, en general, la regla para la **prospección** no es la misma que para la **retrospección**. Esta es la base de una descripción *dual* de los procesos estocásticos.

A lo largo de este trabajo veremos que **predicción** y **retrodicción** conforman dos posibles descripciones del mismo proceso estocástico y corresponderán a las ecuaciones **forward** y **backward** de Kolmogorov [13].

TEORÍA DE LA PROBABILIDAD

Las primeras nociones de probabilidad surgen esencialmente como una aplicación de un problema fundamental. Confieren al ser humano la respuesta sobre la predicción, adelantarle a lo que vendrá. Quizá incluso sea propicia para mitigar su temor, aunque sea el de perder en el *escolazo*³. Es ésta una solución tanto más racional que recurrir a nigromantes o adivinos, pero aplicada a problemas más banales respecto de las grandes cuestiones que atraviesan al ser humano por cierto.

El estudio de los problemas de juegos de azar surge con la correspondencia entre Pascal y, nuevamente, Fermat alrededor de 1654. Aunque se podían calcular probabilidades no podemos hablar aún de una teoría de probabilidades en esa época. Los métodos que dieron ambos permitieron responder más preguntas que las generadas por los jugadores.

La definición primigenia de probabilidad, entendida como cociente de número de casos favorables sobre número de casos totales presenta algunas deficiencias, por ejemplo implica cierta circularidad. Pues es válida para el caso en que los sucesos o eventos sobre los cuales se trata de asignar una noción de probabilidad

³Vocablo argentino lunfardo, *escolazo* léase *juego de azar*.

sean *equiprobables*. En aquellos tiempos se recurría a la matemática ante todo para sugerir o prever estrategias que permitan *mejorar* la suerte en el juego, bajo el concepto intuitivo de que jugando un número suficientemente grande de veces se puede inclinar la balanza a nuestro favor. Que los eventos en cuestión sean igualmente posibles, se debe mayormente al hecho de que las reglas de los juegos de azar sean justas.

La teoría de probabilidad estuvo fuertemente asociada con el cálculo combinatorio. Un resultado importante provino de J. Bernoulli⁴ en su obra póstuma *Ars Conjectandi* en 1713, que incluía un teorema importante que se conoce hoy como la *ley fuerte de los grandes números*. Este teorema dió lugar a otro muy importante el *teorema central del límite*.

Aun en el año 1900 la noción misma de probabilidad era confusa. Henri Poincaré afirmó en su libro *Calcul des probabilité* (1900) que no se podía dar una definición satisfactoria de probabilidad, en cada caso había que realizar una serie de convenciones.

Más tarde fue utilizada por Maxwell y Boltzman en mecánica estadística. Otras teorías como la de Mendel [14] sobre la herencia genética y la de Darwin [15] sobre la evolución de las especies, emplearon el concepto de probabilidad y de azar como parte fundamental. A principios del siglo XX se incorpora a la lista la mecánica cuántica, de forma que la probabilidad cubrió gran parte de las ciencias de la época.

El lugar que hoy tiene la teoría de probabilidades es dentro de la teoría de la medida. A continuación daremos alguna indicación desde un punto de vista moderno. El punto de partida es la asunción de un conjunto ω y una clase \mathfrak{S} de subconjuntos de ω ; suele utilizarse la nomenclatura $\mathfrak{S}(\omega)$, para hacer explícita tal relación. De manera que todo evento, para el cual tenga sentido hablar de una probabilidad, puede ser interpretado como cierto subconjunto de ω que pertenece a la clase \mathfrak{S} . Un evento arbitrario puede ser considerado como la unión de ciertos subconjuntos

⁴Nos referimos al Jacob (1654-1705), el hermano mayor de Johann.

de ω , de esta forma el conjunto ω recibe el nombre de espacio de eventos elementales. Un evento que se da con certeza se corresponde con el conjunto ω . La incompatibilidad de una clase de eventos se asocia a que la intersección de los correspondientes conjuntos sea el conjunto vacío. La reunión, la concurrencia y la discriminación de dos o más eventos se asocian con la unión, intersección y diferencia de los respectivos conjuntos. Por lo tanto, resulta natural que el conjunto de eventos, aquella clase \mathfrak{S} de subconjuntos de ω , sea una σ -álgebra de conjuntos sobre ω :

1. $\omega \in \mathfrak{S}$,
2. $\forall A \in \mathfrak{S} \implies A^c \in \mathfrak{S}$,
3. $\forall A_k \in \mathfrak{S}, k = 1, 2, \dots \implies \bigcup_n A_n \in \mathfrak{S}$,

vemos que \mathfrak{S} es cerrada bajo el complemento y bajo la unión numerable. Solo falta definir una función que asigne lo que denominamos probabilidad para los eventos, es decir para los elementos de la σ -álgebra \mathfrak{S} , es decir, tomamos una medida P sobre \mathfrak{S} al intervalo $[0, 1]$, tal que $P(\omega) = 1$. Queda entonces conformado el *espacio de probabilidad* como la terna $(\omega, \mathfrak{S}, P)$.

Veamos dos conceptos más, relevantes para este trabajo, uno es el de *variable aleatoria* X como función medible sobre (ω, \mathfrak{S}) en algún conjunto numerable o no numerable E . Esta definición abstracta permite definir eventos asociados a conjuntos abstractos, asociados al segundo de los conceptos importantes: el del *probabilidad condicional*. Para dos conjuntos A_1, A_2 en \mathfrak{S} , la probabilidad condicional de A_2 relativa a (o bajo la hipótesis de, o dado) A_1 está dada por

$$P(A_2|A_1) = \frac{P(A_2 \cap A_1)}{P(A_1)} \quad (3.2)$$

con tal que $P(A_1) \neq 0$, si $P(A_1) = 0$ entonces (3.2) está indefinida. Suele decirse que A_2 es el evento *condicionado* por el *condicionante* A_1 . Eventualmente usaremos la notación $P(A_2 \cap A_1) = P(A_2, A_1)$ para la probabilidad de la intersección.

A modo de ejemplo, si se lanzan dos monedas al aire, los eventos elementales son subconjuntos ordenados de la forma $\{\omega_1\omega_2\}$, cada moneda (ω_i) toma valores cara (c) o cruz (x), de manera que el espacio de eventos elementales es el conjunto de pares ordenados $\omega = \{cc, cx, xc, xx\}$. Supongamos una variable aleatoria $X : \omega \rightarrow \mathbb{R}$, que de cuenta del número de cruces que se obtienen en cada realización, $X(cc) = 0$, $X(cx) = 1 = X(xc)$, $X(xx) = 2$. La repetición de lanzamientos conforma una secuencia de realizaciones $\mathcal{R}(X) = \{X_1, X_2, \dots\}$, donde cada elemento de este conjunto es una variable aleatoria del mismo tipo que toma valores del conjunto $\{0, 1, 2\}$, según sea el resultado de las dos monedas en cada lanzamiento. En este ejemplo las variables aleatorias del conjunto $\mathcal{R}(X)$ son independientes, i.e. $P(X_i, X_j) = P(X_i)P(X_j)$ para cualesquiera dos variables X_i, X_j de $\mathcal{R}(X)$, pero podemos pensar en otro tipo de realizaciones en las que esto no ocurra, en tal caso hablaremos de cierto tipo de correlación entre estas variables aleatorias.

La teoría de procesos estocásticos surge a partir de la teoría de probabilidades, trata sobre conjuntos de variables aleatorias para determinados instantes temporales que están correlacionadas, que *concatena* dichos conjuntos. Formalmente un proceso estocástico definido sobre un espacio de probabilidad $(\omega, \mathfrak{G}, P)$ es toda familia de funciones medibles $\{x_t\}_{t \in T}$ sobre ω a valores en E , donde T es un cierto conjunto de índices, discreto o continuo. Dado el espacio de eventos elementales $\omega \in \omega$, la aplicación $t \mapsto x_t(\omega)$ define una trayectoria del proceso estocástico. Normalmente t representa al tiempo, aunque pueden considerarse diferentes situaciones. Por ejemplo, si t representa la distancia espacial respecto a un punto denominado de origen, podríamos considerar a x_t como el número de puntos aleatoriamente distribuidos cuya distancia al origen es menor que t . En este trabajo nos referiremos a t como el *tiempo*.

Esta breve introducción permite vislumbrar que la teoría de probabilidades quedó formalizada con rotundo éxito a partir de

la teoría de la medida. Posteriormente el estudio de los procesos estocásticos permitió trabajar con la teoría axiomatizada de probabilidades y avanzar matemáticamente hablando, i.e. sin hacer caso prácticamente a situaciones del mundo natural. Si bien es posible utilizar la teoría de probabilidades para modelar algún fenómeno aleatorio y a partir de esto recorrer el camino opuesto: obtener de las características del fenómeno ciertos conceptos teóricos, debemos decir que para muchos matemáticos los elementos mismos del fenómeno o sus propiedades no forman parte del corpus teórico⁵; lo que se observa del fenómeno podría motivar introducir algún concepto o buscar algún resultado dentro del modelo matemático que nos ayude a entender el fenómeno o darnos una idea de sus propiedades, pero las observaciones mismas no forman parte del modelo axiomático. Desde este punto de vista, es discutible hablar de probabilidades en un contexto tan abstracto, a lo que los matemáticos responden a ellos con la teoría de la medida. Esta dicotomía se fue acentuando al punto tal de hablar de probabilidades teórico—prácticas de la ciencias y probabilidades formales producto de la axiomática [16, 17, 18].

Cabe destacar que la formalización de la teoría de la probabilidad y los procesos estocásticos, a partir de la teoría de la medida, es de vital importancia para garantizar una correcta operativa de los conceptos que se utilizan en la resolución de procesos concretos, asegurando la consistencia lógica entre otras cosas.

Debemos decir también que a fines de resolución de problemas particulares, normalmente no se utiliza la notación formal que aparece en la axiomática. Utilizaremos una notación ligeramente diferente a fin de ganar mayor intuición y para trabajar posteriormente las ideas respecto del cálculo variacional. Denotaremos por x_t o $x(t)$ a una variable aleatoria al tiempo t . Aunque tendremos en mente que ésta sea en realidad una función sobre

⁵Esta concepción puramente formal sin hacer alusión a fenómenos, aunque inspirándose en ellos, no forma parte de la estructura de una teoría física en la que no se puede ni debe hacerse tal escisión.

el espacio medible (ω, \mathfrak{S}) a cierto espacio E , por ejemplo \mathbb{R} , de manera que para cada $\omega \in \omega$ tendríamos formalmente $x_t(\omega)$ o $x(t, \omega)$.

Comenzaremos tratando el problema de variables aleatorias en tiempo continuo. Una descripción de los procesos en tiempo discreto, puede hacerse a partir de lo que se denomina una *esqueletización* del proceso estocástico (*skeleton process*) [20].

Para describir matemáticamente un proceso estocástico podemos optar por una de sus dos versiones. Una de ellas utiliza el cálculo estocástico, la cual se basa en construir ecuaciones de evolución sobre la propia variable aleatoria, con operaciones bien definidas. Mientras que en esta tesis pretendemos describir cómo cambia en el tiempo una variable aleatoria, a través de su medida de probabilidad o propensión y de esta forma tendremos ecuaciones deterministas para tal medida.

PROCESOS ESTOCÁSTICAMENTE DETERMINADOS

Supongamos que estamos interesados en estudiar como cambian en el tiempo un conjunto de estados de cierto sistema.

Si el sistema al tiempo t el sistema se encuentra en el estado y y está únicamente determinado por su estado x al tiempo s , diremos que el sistema será *puramente determinista*. Matemáticamente traducimos esta afirmación diciendo que existe una única función ζ tal que

$$y = \zeta(t, x, s). \quad (3.3)$$

Esta descripción atañe, a la mecánica analítica, por mencionar un ejemplo, aunque un estado del sistema x corresponde a coordenadas y momentos.

Tenemos también otra categoría de procesos en la cual el estado x al tiempo s solo determina una cierta probabilidad de encontrar al sistema en algún estado y al tiempo t . En ese caso diremos que el sistema será *estocásticamente determinista*.

Matemáticamente expresamos esto con una medida de probabilidad P

$$P(y, t, x, s). \quad (3.4)$$

Hemos dicho que $P(y, t, x, s)$ asigna una medida de probabilidad al caso en que *si* el sistema se encontrara en el estado x al tiempo s , *entonces* el sistema se encuentre en algún estado y al tiempo t . Por esta razón se dice que $P(y, t, x, s)$ mide la propensión de la transición $(x, s) \mapsto (y, t)$, donde $s \leq t$.

La notación (3.4) introducida en [13] es algo antigua. Continuaremos con esta exposición *a la* Kolmogorov [13] con una notación moderna, aunque distinta a la notación utilizada para formalizar la teoría de probabilidades.

En notación actual (3.4) corresponde a una probabilidad condicional $P(y, t|x, s)$, sobre dos miembros de la familia de variables aleatorias $\{x_t : \omega \rightarrow E\}_{t \in T}$, con T intervalo real, dados dos instantes $s, t \in T$, con $s \leq t$ y dos estados x, y de E . En particular si E es numerable $P(y, t|x, s)$ corresponde a $P(x_t = y|x_s = x)$. Mientras que para el caso en que E sea no numerable $P(y, t|x, s)$ corresponde $P(x_t \leq y|x_s \leq x)$, que es una medida condicional acumulada, aunque también podremos referirnos a la densidad condicional.

La teoría de procesos estocásticos de variable continua fue ampliamente desarrollada. Sobre este tipo de procesos volveremos con algunos comentarios. Cabe adelantar que en ese caso la evolución de la medida de probabilidad estará descrita por ecuaciones diferenciales no ordinarias.

En particular, en esta tesis desarrollaremos los procesos en los que el número de estados es finito o numerable, esto dará lugar a ecuaciones diferenciales ordinarias que determinarán la evolución temporal, es decir que cada variable aleatoria toma *valores discretos*.

TRANSICIÓN A UNA NOTACIÓN ESTADO-TEMPORAL

Antes de continuar, queremos dar cuenta de una transición entre nomenclaturas. Insistiremos sobre este punto, ya que tratamos con otra de las trivialidades que encierran algo fundamental. Sobre una nomenclatura se *piensa y habla* de aquellas cosas de las que suele ser complejo enhebrar palabras, sin temor de incurrir en el error. No obstante, una nomenclatura que no refleje todos los aspectos del objeto de estudio de interés o incluso que refleje aún más aspectos que no formen parte del estudio de este trabajo no nos serán de utilidad. En particular, queremos hacer implícita la dependencia con el conjunto de eventos elementales ω y centrar la atención sobre una descripción *estado-temporal*, es decir en términos de elementos de los conjuntos de *estados* E y *tiempos* T .

Queremos pasar de la notación usual [16, 17, 18], para eso haremos una breve revisión de las ideas y su nomenclatura sobre teoría de probabilidades ahora aplicada a procesos estocásticos. Consideremos una familia de variables aleatorias, para la cual cada integrante de la misma está temporalmente etiquetado; todas ellas tienen un dominio en común, ω , llamado *espacio de eventos elementales*, también poseen también el mismo conjunto imagen, denominado *espacio de estados*, $E = \{x_i\}_{i \in I \subset \mathbb{N}}$. Denotamos por $\{x_t : \omega \rightarrow E\}_{t \in T}$ a esta familia, donde T es intervalo real que etiqueta a los integrantes de la familia.

Vale aclarar que en la sucesión E no hay elementos coincidentes, allí sólo están los estados posibles para cada evento. Esta idea puede ser formalizada de la siguiente manera. Para cada instante fijo t los eventos que ocurren y conllevan a un mismo estado, e.g. x_i , a través de x_t , están completamente definidos por dicho estado x_i , precisamente por la preimagen del estado dada por x_t . Definimos el conjunto de los eventos que por x_t van a parar a un mismo estado x_i como

$$\omega_{\langle t, i \rangle} := x_t^{-1}(\{x_i\}). \quad (3.5)$$

Luego para cada $i \in I$ tenemos que $\emptyset \neq \omega_{\langle t, i \rangle} \subseteq \omega$. Notemos que

para dos estados distintos $x_i \neq x_j$, estos conjuntos son disjuntos. Es decir,

$$\forall i, j \in I : i \neq j \implies \omega_{\langle t, i \rangle} \cap \omega_{\langle t, j \rangle} = \emptyset. \quad (3.6)$$

Además la unión disjunta para cada t

$$\bigcup_{i \in I} \omega_{\langle t, i \rangle} = \omega \quad (3.7)$$

De todo esto se sigue que la familia $\{\omega_{\langle t, i \rangle}\}_{i \in I}$ es una partición relativa a ω e incluida en $\mathcal{P}(\omega)$, el conjunto de partes ω .

De (3.5) tenemos que $x_t(\omega_{\langle t, i \rangle}) = \{x_i\}$, o bien en otros términos $\forall \omega \in \omega_{\langle t, i \rangle} : x_t(\omega) = x_i$, en palabras: dado un instante de tiempo t el estado x_i queda completamente definido a partir de aquellos eventos $\omega \in \omega_{\langle t, j \rangle}$.

Podemos tratar con eventos más generales que no correspondan necesariamente a conjuntos unitarios del tipo $\{x_i\}$ vía x_t^{-1} , como se hiciera en (3.5), sino a un conjunto $E' = \{x_i\}_{i \in I'}$ con $I' \subseteq I$ y $E' \subseteq E$. En tal caso construimos un conjunto de eventos, $\{\omega_{\langle t, i \rangle}\}_{i \in I'}$ cuya unión denotaremos por

$$\omega_{\langle t, E' \rangle} = \bigcup_{i \in I'} \omega_{\langle t, i \rangle} \quad (3.8)$$

el conjunto $\omega_{\langle t, E' \rangle}$ está asociado a la preimágen del conjunto E' vía x_t que denotamos por

$$\omega_{\langle t, E' \rangle} := x_t^{-1}(E'). \quad (3.9)$$

Luego la medida de probabilidad P sobre estos conjuntos de eventos (3.8) y (3.9), que son elementos de la σ -álgebra, \mathfrak{S} , escribiremos

$$P(\omega_{\langle t, E' \rangle}) := \sum_{i \in I'} P(\omega_{\langle t, i \rangle}) \quad (3.10)$$

Abordaremos ahora la notación de la medida de probabilidad sobre un espacio de probabilidad $(\omega, \mathfrak{S}, P)$. Comenzaremos con

el caso de subconjuntos unitarios $\{x_i\}$ de E para arribar finalmente a la que utilizaremos en este trabajo. Para este caso tenemos las siguiente formas de expresar a $P(x_t(\omega) = x_i)$

$$P(x_t(\omega) \in \{x_i\}) := P(x_t^{-1}(\{x_i\})) := P(\omega_{\langle t, i \rangle}). \quad (3.11)$$

Repetimos el procedimiento para conjuntos que no sean necesariamente unitarios, i.e. $E' = \{x_i\}_{i \in I'}$ con $I' \subseteq I$ y $E' \subseteq E$

$$P(x_t(\omega) \in E') := P(x_t^{-1}(E')) := P(\omega_{\langle t, E' \rangle}). \quad (3.12)$$

En consecuencia

$$P(x_t(\omega) \in E) = P(\omega_{\langle t, E \rangle}) = P(\omega) = 1. \quad (3.13)$$

Finalmente de todo lo anterior se inducen las siguientes relaciones biunívocas para cada $t \in T$ y $E' \subseteq E$

$$\omega_{\langle t, E' \rangle} := x_t^{-1}(E') \sim (E', t). \quad (3.14)$$

luego el cálculo de la medida de probabilidad P sobre $\omega_{\langle t, E' \rangle}$, hace alusión solo al par ordenado (E', t) . De forma tal que $P(E', t)$ significa $P(\omega_{\langle t, E' \rangle})$.

Cuando E' consiste en un conjunto unitario $\{x_i\}$ tendremos

$$\omega_{\langle t, i \rangle} := x_t^{-1}(\{x_i\}) \sim (x_i, t). \quad (3.15)$$

De manera tal que $P(x_i, t)$ significa $P(\omega_{\langle t, i \rangle})$.

En resumen, a partir de $\omega_{\langle t, E' \rangle} \sim (E', t)$ en (3.14) y $\omega_{\langle t, i \rangle} \sim (x_i, t)$ en (3.15), conformamos el siguiente **nomenclátor**

$$P(x_i, t) \sim P(\omega_{\langle t, i \rangle}), \quad P(x_j, t | x_i, s) \sim P(\omega_{\langle t, j \rangle} | \omega_{\langle s, i \rangle}), \quad (3.16)$$

$$P(E', t) \sim \sum_{j \in I'} P(\omega_{\langle t, j \rangle}), \quad P(E', t | x_i, s) \sim \sum_{j \in I'} P(\omega_{\langle t, j \rangle} | \omega_{\langle s, i \rangle}),$$

donde $s \leq t$.

Hemos logrado nuestro cometido, dar una notación que no dependa explícitamente de ω , pero que está conectada, vía x_t^{-1} , a la notación usual de conjuntos sobre nuestra σ -álgebra, \mathfrak{S} .

En [13] Kolmogorov trata directamente con probabilidades sobre los estados. Aunque subyace allí la idea de probabilidad como medidas de conjuntos de una σ -álgebra, que a propósito, el mismo Kolmogorov ideó en su axiomática de la teoría de probabilidades [11]. El recorrido que sugiere [13] puede mapearse vía la nomenclatura propuesta hacia los elementos de σ -álgebra.

En primer lugar asume que la medida P de (3.4) satisface [13]

$$P(E, t|x_i, s) = 1. \quad (3.17)$$

La expresión (3.17) es una propuesta de exhaustividad, en el sentido en que los estados de E agotan la totalidad de posibilidades en el lapso de tiempo de $|t-s|$. También tendremos una expresión complementaria a (3.17) que de cuenta sobre imposibilidad del conjunto *vacío*. Diremos que dado el par (x_i, s) , para todo tiempo posterior $t > s$ el sistema debe estar en algún estado de E , i.e. es imposible que los estados del sistema *vivan fuera* de E .

Para el caso de eventos más generales que se correspondan con un subconjunto de estados $E' = \{x_j\}_{j \in I'}$ con $I' \subseteq I$ y $E' \subseteq E$

$$P(E', t|x_i, s) = \sum_{j \in I'} P(x_j, t|x_i, s). \quad (3.18)$$

con $s, t \in T$, $s \leq t$, $x_i \in E$, que se sigue de nuestro **nomenclátor** (3.16).

En 1931, el matemático ruso A. N. Kolmogorov sentó las bases de la dinámica que subyace a los procesos estocásticos. Estaba interesado en obtener una ecuación para la medida de probabilidad condicional. Para el caso que nos compete en esta tesis, i.e. un espacio de estados numerable, y dados 3 instantes ordenados $t \geq r \geq s$, tenemos que el miembro izquierdo de (3.18) *debe* satisfacer [13]

$$P(E', t|x_i, s) = \sum_{n \in I} P(E', t|x_n, r)P(x_n, r|x_i, s). \quad (3.19)$$

donde la suma cubre todo el conjunto de estados $E = \{x_n\}_{n \in I}$. La expresión (3.19) es la denominada ecuación fundamental de Kolmogorov. Anteriormente hemos insistido en que también el matemático Sydney Chapman ha trabajado paralelamente en estas ecuaciones, por lo que suele decirse que (3.19) es la ecuación de Chapman-Kolmogorov para espacios numerables.

ECUACIONES DE CHAPMAN KOLMOGOROV

Queremos dar cuenta de los argumentos o hipótesis de plausibilidad de (3.19). Haremos un recorrido hacia la ecuación de Chapman - Kolmogorov (3.19) aplicando los axiomas de la teoría de probabilidad sobre sistemas *estocásticamente deterministas*.

La expresión (3.19) involucra 3 tipos de eventos

$$\alpha = \omega_{\langle t, E' \rangle}, \quad \beta_n = \omega_{\langle r, n \rangle}, \quad \gamma = \omega_{\langle s, i \rangle}; \quad (3.20)$$

que se corresponden, vía nuestro **nomenclátor**, al trío

$$\alpha \sim (E', t), \quad \beta_n \sim (x_n, r), \quad \gamma \sim (x_i, s), \quad (3.21)$$

asociados a cada uno de los tiempos t, s y r con tal que $t \geq r \geq s$. Los eventos α y γ corresponden a subconjuntos de E , vía (3.5), (3.8) y (3.9), mientras que $\{\beta_n\}_{n \in I}$ corresponde a una partición de E , vía (3.5), (3.6) y (3.7).

Consideremos el siguiente conjunto $\alpha^* = \alpha \cap \gamma$, de manera que podremos escribir $\alpha^* = \bigcup_n \alpha^* \cap \beta_n$, luego por aditividad de P sobre α^* tendremos excluyentes

$$P(\alpha^*) = \sum_{n \in I} P(\alpha^*, \beta_n), \quad (3.22)$$

reemplazando $\alpha^* = \alpha \cap \gamma$ en (3.22)

$$P(\alpha, \gamma) = \sum_{n \in I} P(\alpha, \beta_n, \gamma), \quad (3.23)$$

multiplicamos por $P(\gamma)$ a ambos miembros y multiplicandoy dividiendo por $P(\beta_n, \gamma)$ en cada sumando del miembro derecho

$$P(\alpha|\gamma) = \sum_{n \in I} P(\alpha|\beta_n, \gamma)P(\beta_n|\gamma), \quad (3.24)$$

finalmente sustituyendo α , β_n y γ de (3.21) en (3.24)

$$P(E', t|x_i, s) = \sum_{n \in I} P(E', t|x_n, r, x_i, s)P(x_n, r|x_i, s), \quad (3.25)$$

que se asemeja a la ecuación fundamental de Chapman-Kolmogorov (3.19).

Una condición suficiente para que (3.25) se reduzca a la ecuación fundamental de Chapman-Kolmogorov (3.19) es

$$P(E', t|x_n, r, x_i, s) = P(E', t|x_n, r). \quad (3.26)$$

Más adelante mostraremos que la condición (3.26) es satisfecha por un proceso de Markov. Por esta razón se dice que para todo proceso de Markov la ecuación fundamental de Chapman-Kolmogorov resulta automáticamente satisfecha, o bien es considerada una identidad.

Introduciremos ahora una notación general, que posteriormente nos permitirá definir con claridad un proceso de Markov. Consideremos entonces explícitamente al proceso estocástico sobre una espacio de probabilidad $(\omega, \mathfrak{S}, P)$ y variable aleatoria $x(t)$, para cada t de cierto intervalo real T . Consideremos los conjuntos ∇_t y Δ_t de todas las posibles historias anteriores y posteriores⁶ a cierto t . Estos conjuntos son elementos de la σ -álgebra, \mathfrak{S} , que están generados por las preimágenes de las variables aleatorias de los conjuntos $\{x_u : u < t\}$ y $\{x_u : u > t\}$, respectivamente. [19, 20]

⁶A modo de regla nemotécnica, la notación ∇_t y Δ_t está asociada a un *pasado* apunta hacia *abajo* y un *futuro* apunta hacia *arriba*, tal como ocurre si colocásemos un eje vertical de coordenadas temporales.

Usando esta notación, la ecuación de Chapman-Kolmogorov (3.19) puede escribirse como

$$P(\Delta_t | \nabla_s) = \sum_{n \in I} P(\Delta_t | x_n, r) P(x_n, r | \nabla_s), \quad (3.27)$$

con $t > r > s$. La validez de esta última expresión se basa en que podemos repetir el procedimiento anterior, que dio lugar a (3.24), considerando a los eventos α y γ , como el conjunto de todas las historias futuras y pasadas del sistema. En este sentido diremos que estos conjuntos de historias pasadas y futuras corresponden a *eventos abstractos*, los cuales no se corresponden necesariamente *un* estado del sistema, ya que en general éstos pueden corresponderse a subconjuntos de E asociados a distintos tiempos. En otras palabras, tales *eventos abstractos* no pueden ser siempre considerados como *un* estado del sistema, pues éste último no puede siquiera *estar* al mismo tiempo en dos elementos de E .

Por otro lado, los estados del sistema tienen la propiedad de la *mutua exclusividad* o *incompatibles*, i.e. dados dos estados $x_i \neq x_j$ de $E = \{x_i\}_{i \in I \subset \mathbb{N}}$, la probabilidad de encontrar al sistema en ambos estados *simultáneamente* es nula

$$\forall x_i, x_j \in E : x_i \neq x_j \implies P(x_i, x_j) = 0 \quad (3.28)$$

la cual se sigue directamente de (3.6).⁷ En (3.28) hemos omitido la referencia temporal en favor de acentuar el efecto sobre lo que ocurre con los estados simultáneos.

Como tratamos con un espacio E numerable podemos definir

$$\mathcal{P}_{ji}(t, s) := P(x_j, t | x_i, s), \quad (3.29)$$

la cual asigna una medida de propensión de la transición entre estados $(x_i, s) \mapsto (x_j, t)$, para $s \leq t$. Estamos utilizando aquí el

⁷Existen otras teorías que también hacen uso de la probabilidad como cantidad fundamental y que no necesariamente satisfacen esta propiedad de exclusividad mutua. Estamos hablando de la mecánica cuántica. Allí podemos hablar de *superposición de estados* [6].

término *transición* entre estados con cierta licencia, pues el caso $i = j$ más bien corresponde a una *auto-transición*.

Hagamos un comentario al respecto. Si queremos incluir el caso $i = j$ dentro de la categoría de transiciones propiamente dichas, i.e. $i \neq j$, debemos notar que en éstas últimas tenemos un *cambio* de estado. Tenemos, por tanto, que considerar un ámbito en el que se hable de cambios de estado $\forall i, j$, incluyendo el caso $i = j$. Podemos optar por una extensión de la noción de estado, incluyendo la etiqueta temporal. En rigor de verdad los estados son los elementos de E . De manera que consideraremos al par (x_j, t) como un *estado enriquecido*. El conjunto de tales cantidades se forma tomando los $x_j \in E$ indexados con un instante de tiempo $t \in T$. En tal caso podremos decir que (x_i, s) y (x_i, t) corresponden a *estados distintos*; en cuyo caso sería correcto hablar de una transición entre estados enriquecidos, aún en el caso $i = j$.

En este contexto, podemos reescribir el pedido de exhaustividad sobre E (3.17) como

$$\sum_{j \in I} \mathcal{P}_{ji}(t, s) = 1 \quad (3.30)$$

la cual es válida para todo $t > s$ y para todo estado x_i de E . La expresión (3.30) incluye el hecho de que los elementos de E son excluyentes uno a uno y que todos ellos forman un conjunto *exhaustivo*, en el sentido en que la suma sobre todos ellos agota la totalidad de posibilidades.

Por otro lado, tenemos también una medida para el caso

$$\mathcal{P}_{ji}(u, u) = \delta_{ji}, \quad \forall u \in T \quad (3.31)$$

donde δ_{ji} a la función de Kronecker, $\delta_{ii} := 1$ y $\delta_{ji} := 0$ $j \neq i$. La expresión (3.31) está relacionada con la incompatibilidad mutua de (3.28)⁸; y se puede interpretar como una *certeza de permanencia*

⁸Pues por (3.29)

$$\mathcal{P}_{ji}(t, t) = P(x_j, t | x_i, t) = \frac{P(x_j, t, x_i, t)}{P(x_i, t)},$$

estado-temporal o bien bajo alguna licencia *imposibilidad de transiciones instantáneas*, en el sentido en que para un *lapso* de tiempo de duración nula, solo tenemos la certeza de permanencia del sistema en el estado x_i o bien resulta *imposible* (leáse probabilidad nula) cambiar del estado x_i a cualquier otro estado distinto x_j de manera instantánea. Veremos también que (3.31) puede ser útil como condición inicial o final de las ecuaciones diferenciales de Kolmogorov, pero eso es tema de la próxima sección.

Si nos remitimos al estudio de la propensión de los estados de E , la ecuación de Chapman-Kolmogorov general (3.27) toma la forma *elemental*⁹

$$\mathcal{P}_{ji}(t, s) = \sum_{k \in I} \mathcal{P}_{jk}(t, r) \mathcal{P}_{ki}(r, s), \quad (3.32)$$

con $t \geq r \geq s$ y nuevamente la suma cubre todos los estados de E . Esta versión de la ecuación de Chapman-Kolmogorov es la que trataremos en adelante.

En ambos casos podemos trabajar con una nomenclatura matricial formal más cómoda, teniendo en cuenta que todo producto de matrices infinitas requiere el estudio de la convergencia de las series pertinentes. Podemos trabajar indistintamente con una nomenclatura matricial que comprime la notación anterior. La primera explicitaba a los estados involucrados y la segunda puede ganar en simplicidad, precisamente por hacer lo contrario.

Finalmente las ecuaciones de Chapman-Kolmogorov (3.32) para las probabilidades de transición entre estados del sistema, de una cantidad numerable de estados, pueden escribirse utilizando una nomenclatura matricial [21, 22] formal como

$$\mathcal{P}(t, s) = \mathcal{P}(t, r) \mathcal{P}(r, s), \quad (3.33)$$

con $t \geq r \geq s$.

luego por (3.28) tenemos (3.31)

⁹La expresión (3.32) regula la dinámica sobre los *estados* de E .

Tenemos una diferencia sutil de la formulación general de la ecuación de Chapman-Kolmogorov (3.27) o (3.32) con respecto a su forma *elemental* (3.32). Esta última se trata de relaciones de un *mismo objeto* matricial, evaluado en distintos instantes de tiempo, mientras que en el caso general (3.27) el objeto es una probabilidad condicional mixta, en el sentido en que condicionado y condicionante son estados (x_n, r) y eventos abstractos ∇_t y Δ_t . Estos provienen respectivamente de un partición *elemental* en un instante r y de las historias pasadas y futuras a cierto instante t .

Desde este punto de vista trabajar con (3.27) o (3.32) nos provee, como ya hemos dicho, la posibilidad de tratar con objetos del *mismo tipo*, de hecho hasta el punto tal que forman un *semigrupo*, en el sentido matemático del término¹⁰. La operación del semigrupo será el producto usual matricial. Incluso para el caso en que E no sea finito, el producto matricial resulta ser asociativo para matrices no negativas, que es el caso de $\mathcal{P}(t, s)$: la estructura de semigrupo se mantiene para todo E numerable [22].

ECUACIONES DE KOLMOGOROV

Procederemos ahora a mostrar una derivación del par de ecuaciones diferenciales *fundamentales* de Kolmogorov, también conocido como *forward – backward*. Podemos encontrar una deducción completa en [13]. Allí se asume la continuidad y diferenciabilidad de todas las probabilidades de transición $\mathcal{P}_{ji}(t, s)$ con respecto a t y s , con $t \neq s$.

No pretendemos repetir el procedimiento de [13], por eso daremos una breve deducción del par de ecuaciones de evolución para estos procesos, relajando levemente la condición de diferen-

¹⁰Un semigrupo es una estructura algebraica formada por un conjunto no vacío A y una operación binaria o interna \otimes definida sobre $A \times A$, tal que para todos $x, y, z \in A$ tenemos: **(1)** $x \otimes y \in A$ (*cierre*); **(2)** $(x \otimes y) \otimes z = x \otimes (y \otimes z)$ (*asociatividad*). Si además tenemos un único elemento *neutro* (e), tal que: $e \otimes x = x = x \otimes e$, en ese caso la estructura se denomina *monoide*.

ciabilidad incluyendo el caso $t = s$.¹¹

Recordemos que la medida de probabilidad condicional \mathcal{P} satisface 3 condiciones, a saber:

1. *exhaustividad de los estados* (3.30),
2. *certeza de permanencia estado-temporal* (3.31),
3. ecuaciones de *Chapman-Kolmogorov* (3.32).

Comenzamos definiendo el siguiente objeto matricial

$$\Omega(u) := [\partial_t \mathcal{P}(t, s)]|_{(t,s)=(u,u)}. \quad (3.34)$$

También podemos escribir $\Omega(u) = [D_{\hat{\mathbf{i}}} \mathcal{P}(t, s)]|_{(t,s)=(u,u)}$, donde $D_{\hat{\mathbf{i}}}$ corresponde a la derivada respecto de la primera variable de \mathcal{P} . Esta notación distingue perfectamente entre el *valor* que toma una variable y su *ubicación*. También nos permite arribar de manera clara al siguiente corolario: consideremos un elemento de matriz $\mathcal{P}_{ji}(t, s)$, que corresponde a una función de dos variables temporales (t, s) , e.g. de \mathbb{R}^2 en \mathbb{R} . A partir de la condición (3.31) que reescribimos $\mathcal{P}_{ji}(u, u) = \delta_{ji}$, significa que $\mathcal{P}_{ji}(t, s)$ es constante sobre la recta $t = s$ del plano (t, s) . De forma que su variación en la dirección de *aquella* recta, y sobre los puntos de la misma, es nula, i.e. $D_{\hat{\mathbf{v}}} \mathcal{P}_{ji}(t, s)|_{(u,u)} = 0$, con $\hat{\mathbf{v}} = \frac{1}{\sqrt{2}}(\hat{\mathbf{i}} + \hat{\mathbf{j}})$ y también $\hat{\mathbf{i}} = (1, 0)$, $\hat{\mathbf{j}} = (0, 1)$ son los dos vectores unitarios canónicos. Puesto que $\mathcal{P}_{ji}(t, s)$ es diferenciable podemos escribir esta *derivada direccional* como el producto escalar de un vector $\hat{\mathbf{v}}$ por el gradiente de \mathcal{P}_{ji} , i.e. $D_{\hat{\mathbf{v}}} \mathcal{P}_{ji} = \frac{1}{\sqrt{2}}(D_{\hat{\mathbf{i}}} \mathcal{P}_{ji} + D_{\hat{\mathbf{j}}} \mathcal{P}_{ji})$, sobre los puntos de la recta identidad del plano (t, s) [23]. De aquí se sigue que

$$[D_{\hat{\mathbf{i}}} \mathcal{P}(u, t)]_{u=t} + [D_{\hat{\mathbf{j}}} \mathcal{P}(t, u)]_{u=t} = 0. \quad (3.35)$$

¹¹En [13] no se exige que $\mathcal{P}_{ji}(t, s)$ sea diferenciable y continua para todo t, s , pues en el caso en que la variable aleatoria tome valores en un conjunto no numerable, puede darse el caso de que exista discontinuidad en $t = s$. Nosotros estamos trabajando con el caso numerable, por lo que relajaremos esa condición.

El primer sumando corresponde a la definición de $\Omega(t)$ según (3.34). De manera tal que $\Omega(t)$ presenta las siguientes **dos** variantes de escritura

$$[\partial_u \mathcal{P}(u, t)]_{u=t} = \Omega(t) = - [\partial_u \mathcal{P}(t, u)]_{u=t} \quad (3.36)$$

que nos será de mucha utilidad.

La matriz $\Omega(t)$ (3.34) corresponde al generador de traslaciones temporales infinitesimales o también en este contexto generador de transiciones infinitesimales (temporales). Para entender esto estudiemos la evolución temporal infinitesimal $t \mapsto t + \epsilon$, con ϵ pequeño, para \mathcal{P}_{ji} . Consideremos un desarrollo de $\mathcal{P}_{ji}(t + \epsilon, s)$ alrededor del punto (t, s) y para un ϵ pequeño :

$$\mathcal{P}_{ji}(t + \epsilon, s) = \mathcal{P}_{ji}(t, s) + \epsilon \partial_t \mathcal{P}_{ji}(t, s) + \mathcal{O}(\epsilon). \quad (3.37)$$

con $\mathcal{O}(\epsilon)$ una función¹² tal que $\lim_{\epsilon \rightarrow 0} \mathcal{O}(\epsilon)/\epsilon = 0$ [23]. Luego tomando $s = t$, tenemos

$$\mathcal{P}_{ji}(t + \epsilon, t) = \delta_{ji} + \epsilon \Omega_{ji}(t) + \mathcal{O}(\epsilon), \quad (3.38)$$

la función $\Omega_{ji}(t)$ domina las transiciones de una evolución temporal infinitesimal $t \mapsto t + \epsilon$.

Como corolario de la condición *exhaustiva* para las probabilidades condicionales $\{\mathcal{P}_{ji}(t, s)\}_{i,j \in I}$ (3.30), tenemos que los generadores $\{\Omega_{ji}(t)\}_{i,j \in I}$ heredan la siguiente propiedad

$$\sum_j \Omega_{ji}(t) = 0, \quad (3.39)$$

válida para cada t y cada estado x_i de E .

Finalmente para derivar el par de ecuaciones diferenciales de Kolmogorov, podemos optar por hacerlo en elementos $\mathcal{P}_{ji}(t, s)$ o bien en forma matricial más compacta $\mathcal{P}(t, s)$. A fin de simplificar

¹²Dicha función $\mathcal{O}(\epsilon)$ se denomina *resto* de Taylor y eventualmente será distinta para cada par de estados (x_i, x_j) y tiempos (s, t) , a pesar de que la notación simplificada no lo explicita.

la notación de subíndices utilizaremos la segunda alternativa. A partir de la ecuación de Chapman - Kolmogorov (3.33) y $\Omega(t)$ de (3.36), calculemos la derivada parcial respecto de t y evaluando finalmente en $r = t$, podemos deducir la ecuación *forward*

$$\partial_t \mathcal{P}(t, s) = \Omega(t) \mathcal{P}(t, s), \quad (3.40)$$

$$\partial_t \mathcal{P}_{ji}(t, s) = \sum_{k \in I} \Omega_{jk}(t) \mathcal{P}_{ki}(t, s)$$

Un procedimiento similar a partir de (3.33) y $\Omega(s)$ de (3.36), calculamos la derivada parcial respecto de s y evaluando finalmente en $r = s$ nos conduce a la ecuación *backward*

$$\partial_s \mathcal{P}(t, s) = -\mathcal{P}(t, s) \Omega(s), \quad (3.41)$$

$$\partial_s \mathcal{P}_{ji}(t, s) = -\sum_{k \in I} \mathcal{P}_{jk}(t, s) \Omega_{ki}(s)$$

A pesar del aspecto formal con derivadas parciales, el par *forward-backward* corresponde a ecuaciones diferenciales ordinarias, ya que para (3.40), o (3.41), el parámetro s , o t , está fijo y aparece solo en la condición: $\mathcal{P}_{ji}(s, s) = \delta_{ji}$, *inicial*, o $\mathcal{P}_{ji}(t, t) = \delta_{ji}$, *final*, respectivamente [24].

Esto nos lleva a la siguiente cuestión, hemos dicho que los adjetivos **forward** y **backward** resultan estar asociados a las nociones de **prospección** y **retrospección**, respectivamente. Entender el motivo de tal asociación nos ayuda a comprender mejor la razón de ser del par de ecuaciones (3.40)-(3.41). La primera de ellas es una ecuación en t , a s fijo, que es resuelta integrándola en orden cronológico, diremos que (3.40) responde a una *prospección*, con condición inicial en s , i.e en el pasado. En sentido inverso, algo similar ocurre con la segunda de ellas, a t fijo, que es resuelta integrándola en orden acronológico, entonces diremos que (3.41) responde a una *retrospección*, con condición final en t , i.e en el futuro.

PREGUNTAS PROSPECTIVAS Y RETROSPECTIVAS

Destacamos algunas preguntas importantes. Las probabilidades de transición responden a la siguiente pregunta *condicional prospectiva* o *condicional forward*

c.f. Si el sistema se encuentra al tiempo s en el estado x_i , ¿cuál es la probabilidad de que éste se **hubiere** encontrado al tiempo t en el estado x_j ?

También podría realizarse la pregunta *condicional retrospectiva* o *condicional backward*

c.b. Si el sistema se encuentra al tiempo t en el estado x_j , ¿cuál es la probabilidad de que éste se **hubiera** encontrado al tiempo s en el estado x_i ?

Debemos decir que para el caso en que el estado (x_i, s) sea la condición inicial, o bien que (x_j, t) sea la condición final, las probabilidades de transición podrán describir la evolución de la medida de probabilidad de un estado (x_j, t) , o bien de (x_i, s) . Esto es, de las dos entradas que tiene una probabilidad de transición $P(x_j, t|x_i, s)$, podremos estudiar la evolución sobre una de ellas, fijando antes la otra. Pero las respuestas de **c.f.** y **c.b.** son siempre entre estados *puntuales*, no nos dan respuesta a la evolución temporal del sistema como un todo.

Veamos el siguiente caso que ilustra este punto. Supongamos que la condición es inicial y al tiempo s , donde tengo una distribución inicial sobre E . Entonces no podría utilizar solo una $\mathcal{P}_{ji}(t, s)$, sino que necesitaría todas las probabilidades condicionales, para cada x_i , que haría las veces de condicionante inicial, i.e. todo el conjunto $\{\mathcal{P}_{ji}(t, s) : \forall x_i \in E\}$. En general, con una sola probabilidad condicional no podría estudiar los procesos que tengan por condición inicial o final una distribución sobre E [24]. Existen, por tanto, numerosos casos que quedan fuera: todos aquellos en los que la condición que fijamos para resolver (3.40) o

(3.41) no es sobre un *único* estado de E en particular. Si tenemos por condición inicial (o bien final), a una distribución sobre E , denotada de manera genérica por $\{\alpha_k\}_{k \in I}$ (o bien $\{w_k\}_{k \in I}$), deberíamos considerar una suma ponderada sobre los estados iniciales (finales) tal como:

$$\sum_{i \in I} \mathcal{P}_{ji}(t, s) \alpha_i(s) \quad (3.42)$$

$$\sum_{j \in I} w_j(t) \mathcal{P}_{ji}(t, s) \quad (3.43)$$

para estudiar la evolución temporal de la probabilidad de (x_j, t) a s fijo, o bien de (x_i, s) a t fijo, respectivamente. Se trata de dos combinaciones convexas¹³ de las medidas condicionales $\mathcal{P}_{ji}(t, s)$ sobre una distribución inicial $\{\alpha_k\}_{k \in I}$ o final $\{w_k\}_{k \in I}$, respectivamente [24]. Dado que $s \leq t$, la dependencia temporal de α_i con s en (3.42) o de w_j con t en (3.43), responde sencillamente a una convención que refleja el carácter inicial o final de las respectivas distribuciones. En cualquier caso cabe destacar que estos instantes de tiempo están fijos: s está fijo en el primero y t está fijo en el segundo.

Tomemos el objeto (3.42) para s fijo, podemos comprobar que satisface la ecuación forward de Kolmogorov (3.40). Esto se hereda del hecho que $\mathcal{P}(t, s)$ satisface dicha ecuación. Lo mismo ocurre con el objeto (3.43) para t fijo, satisface la ecuación backward de Kolmogorov (3.41).

Más adelante probaremos que (3.42) y (3.43) corresponden a las medidas de probabilidad absolutas, que pueden obtenerse como marginalizaciones de la medida de probabilidad conjunta¹⁴

$$P(x(t), x(s)). \quad (3.44)$$

¹³*Combinación convexa* puesto que ambas $\{\alpha_k\}_{k \in I}$ y $\{w_k\}_{k \in I}$ satisfacen que $\alpha_k \geq 0$, $w_k \geq 0$, $\forall k \in I$ y $\sum_{k \in I} \alpha_k = 1$, $\sum_{k \in I} w_k = 1$, como toda distribución *honest*a que representa una probabilidad.

¹⁴Hemos simplificado la notación, en pos de maximizar la claridad sobre el objeto del que queremos hablar. Hemos tomado la notación $x(t)$ para la variable aleatoria x_t . Por supuesto entendemos que una notación más apropiada para

Cuando introdujimos las ecuaciones forward y backward, en la página 55, hemos dicho que se trataba de ecuaciones ordinarias sobre las probabilidades de transición, puesto que se consideraba una de sus entradas fija y la otra variable.

Aunando todo lo expuesto en esta sección, a saber: la construcción de las preguntas **c.f.** y **c.b.** y la construcción de distribuciones absolutas a partir de las condicionadas en (3.42) y (3.43), surge la posibilidad de traducir las preguntas sobre una proyección y retrospección a términos absolutos, en otras palabras a *probabilidades absolutas*. Presentamos pues el otro par de preguntas que han surgido de esta discusión, ahora en términos *absolutos*

a.f. ¿Cuál es la probabilidad de que el sistema se **hubiere** encontrado al tiempo t en el estado x_j ?

Y su pregunta *dual* o *conjugada*, también en términos *absolutos*

a.b. ¿Cuál es la probabilidad de que el sistema se **hubiera** encontrado al tiempo s en el estado x_i ?

Las preguntas **a.f.** y **a.b.** omiten explícitamente aquello que fijamos como condición inicial o final. En otras palabras, buscamos una respuesta que no esté supeditada a dos tiempos, sino solo a uno, puesto que nos interesa conocer una medida de probabilidad absoluta.

Podría parecer que las preguntas **a.f.** y **b.f.** admiten la misma probabilidad absoluta, puesto que parecen ser idénticas, debido a la omisión de aquello que se ha fijado como condición inicial o final. Pero esto es solo un artificio, esta apariencia se desvanece al observar la riqueza del lenguaje que nos deja dos tiempos verbales distintos: **hubiere** para referirnos al futuro y **hubiera** para referirnos al pasado, que conectan con el carácter predictivo y retrodictivo respectivamente. También podemos distinguir res-

(3.44) debería ser $P(x(t) = x_j, x(s) = x_i)$. La marginalización de ésta medida sobre $x(t)$ corresponde a sumar sobre los x_j , *idem* $x(s)$ (\dots) sobre x_i .

puestas desde el punto de vista matemático, pues tenemos que (3.42) y (3.43) no son a priori iguales. Por tanto, según esta discusión podemos decir que las probabilidades absolutas asociadas a las ecuaciones forward (3.40) y backward (3.41) de Kolmogorov no serán necesariamente las mismas.

Introduciremos por vez primera el concepto de *proceso coordinizable*, que será aquel que admita una descripción en términos de ciertas *coordenadas generalizadas*. Probaremos entonces que éstas coordenadas estarán asociadas unívocamente a las ecuaciones de Kolmogorov forward y backward.

PROCESOS ESTOCÁSTICOS COORDINALIZABLES

En esta sección veremos una forma de introducir coordenadas generalizadas para procesos estocásticos de esta clase: estocásticamente determinados y sobre espacios de estados numerables.

La mecánica analítica ya nos ha dado lecciones de que no solo son coordenadas aquellas que hacen mención a la *ubicación* espacial o espacio-temporal del objeto de estudio. Dicha teoría está formulada bajo la idea de *coordenadas generalizadas*. Las coordenadas de un sistema deben ser capaces de describir su estado en todo, o *casi* todo, momento. De hecho, en física podemos tener campos escalares, vectoriales, tensoriales, operadores, etc. como objetos que cumplan con este rol.

¿Qué ocurre con teorías en las que predomina la contingencia? ¿Se puede acaso hablar en esos casos de estado del sistema? Para el caso que nos ocupa podemos hablar de estados del sistema, hemos dicho que éstos son los elementos del conjunto numerable E , pero lo que no podemos asignar es *una* trayectoria en E , solo podemos hablar en términos de medidas de probabilidad de cada estado de E . No tiene sentido considerar a los propios elementos de E como coordenadas.

Tal y como hemos adelantado al final de la sección anterior, estamos interesados en describir la evolución temporal de los es-

tados del sistema, a través de la medida de probabilidad absoluta, denotada genéricamente por $P(x(t))$ y a la que podemos concebir como objeto fundamental que nos habla del estado del sistema. Satisface las demandas que hemos impuesto en tanto que es lo más cercano a una trayectoria del sistema. En otras palabras, no podemos considerar *una* trayectoria, entonces solo podemos dar un paso atrás y describir la propensión de una trayectoria $x(t)$. Esta medida de probabilidad será una función de E en el intervalo $[0, 1]$ de \mathbb{R} .

El conocimiento de $P(x(t))$ puede permitirnos calcular la *función generadora de momentos* y con ella obtener sus momentos, por ejemplo, el valor esperado, la desviación estándar y en general, los momentos de todo orden [24]. Es decir, las cantidades que podrían observarse directa o indirectamente en el experimento.

Comenzaremos a desarrollar estas ideas a partir de una mejor comprensión de la ecuación fundamental de Chapman - Kolmogorov (3.32).

La figura 3.1 nos provee una representación esquemática de la misma.

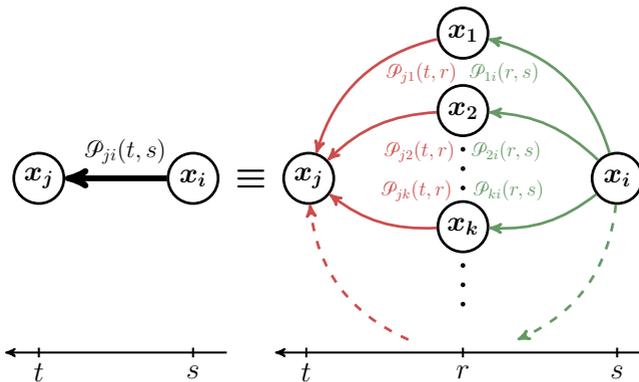


Figura 3.1: Representación esquemática de la ecuación (3.32). El elemento de matriz asociado a la transición $(x_i, s) \mapsto (x_j, t)$ se puede entender como la suma de contribuciones de elementos de matriz asociados a transiciones temporalmente intermedias $(x_i, s) \mapsto (x_k, r)$ y $(x_k, r) \mapsto (x_j, t)$.

Para calcular una transición particular entre dos estados de E y dos instantes de tiempo (s, t) , siempre existirá un tiempo *intermedio* a estos, que llamaremos r . La transición de (x_i, s) a (x_j, t) puede realizarse por cualquiera de los estados de E en el instante intermedio r , digamos (x_k, r) . Recordando que los estados de E son mutuamente excluyentes, por lo que cada camino de la figura 3.1, por ejemplo, $(x_j, t) \leftarrow (x_k, r) \leftarrow (x_i, s)$ está medido por $\mathcal{P}_{jk}(t, r)\mathcal{P}_{ki}(r, s)$. El sistema tomará *una* de ellas, pero la probabilidad total considera las probabilidades de todas las opciones o historias intermedias excluyentes. Esto es precisamente la ecuación de Chapman - Kolmogorov (3.32), que reescribimos en colores

$$\mathcal{P}_{ji}(t, s) = \sum_{k \in I} \mathcal{P}_{jk}(t, r)\mathcal{P}_{ki}(r, s). \quad (3.45)$$

En la nomenclatura que estamos usando para representar las probabilidades de transición y los diagramas asociados a ella de la figura 3.1, el futuro está a *izquierda* y el pasado a *derecha*

$$\mathcal{P}_{ji}(t, s) := P(x_j, t | x_i, s), \quad (3.46)$$

con $t \geq s$. En general estamos habituados a considerar lo contrario. De hecho, en la bibliografía abunda la nomenclatura opuesta, es decir

$$P_{ij}(s, t) := P(x_j, t | x_i, s), \quad (3.47)$$

con $t \geq s$. Tres son los motivos por los cuales sacrificamos la notación (3.47). La primera proviene de la propia teoría de probabilidades, el condicionante está a derecha y el condicionado a izquierda por lo que adoptamos (3.46). La segunda de ellas descansa en la rápida comparación de símbolos, es decir, esta nomenclatura es utilizada en otras teorías físicas, como la mecánica cuántica, que nos será de utilidad posteriormente. La tercera proviene simplemente de ir contra la corriente y porqué no justificarla desde un lugar y una época lejanos: la cultura aborigen *aymara*; la cual considera que el futuro se encontraría detrás, puesto que

éste aún no habría sido desvelado, mientras que el pasado está ahí delante, a la vista.

En la secciones subsiguientes continuaremos con el camino de la introducción de coordenadas en procesos estocásticos. Haremos una parada técnica y utilizaremos como funciona esto es una teoría conocida, aún hoy controversial: la mecánica cuántica. Igualmente mentaremos aquellos elementos de la teoría consensuados por la comunidad científica; sin hacer mención, por ejemplo, al problema de la medida. En resumen, solo nos remitiremos a la evolución temporal determinista de esa teoría.

ESTRUCTURA PARCIAL DE UNA MECÁNICA CUÁNTICA

Desarrollaremos entonces la idea de coordenadas para estos procesos estocásticos. Para tal menester haremos uso de una teoría consagrada: la mecánica cuántica, en particular, utilizaremos parcialmente como referencia su estructura matemática, a modo de muleta para este recorrido. A pesar de las diferencias interpretativas de los símbolos, utilizaremos la semejanza de algunas estructuras compartidas por estas teorías para introducir un concepto fundamental en procesos estocásticos: el de coordenadas. Como toda muleta nos será de utilidad por un tiempo, en cuanto la estructura que pretendemos construir posea cierta autonomía todo andamio desaparecerá.

El título expresa el hecho de que hablaremos de *una* mecánica cuántica, en el sentido en que trataremos espacios de Hilbert numerables.

Comencemos con uno de los elementos fundamentales en la mecánica cuántica y también en las teorías cuánticas de campos, la denominada *amplitud de probabilidad*, que satisfacen ecuaciones similares a la de Chapman-Kolmogorov y, por tanto, diagramas semejantes a la figura 3.1 [5]. Existe una similitud entre probabilidades de transición clásicas y amplitudes de probabilidad cuánticas, en el sentido en que para calcular una de estas últimas también se puede hacer uso de la suma sobre historias.

No obstante, también tendremos algunas diferencias, las amplitudes de probabilidad toman valores sobre \mathbb{C} y no son en sí mismas probabilidades. El cuadrado del módulo de estas amplitudes es interpretado como probabilidades de transición.

Estas semejanzas y diferencias estructuran una de las maneras teóricas de introducir nuevos fenómenos en mecánica cuántica como la superposición y la interferencia [5].

Recordemos que estamos tratando con un conjunto de estados E numerable. Consideremos, por tanto, un espacio de Hilbert numerable, que tiene un conjunto completo de vectores base ¹⁵ denominados *kets* [6] y denotados por $|k\rangle$, que hacen las veces de los estados x_k . El conjunto de estos vectores base del espacio de Hilbert los resumimos como $\{|k\rangle : k \in \mathcal{I} \subseteq \mathbb{N}\}$.

Debemos aclarar que un sistema cuántico puede *estar* en una superposición de estos vectores base $|k\rangle$. Mientras que un sistema clásico hemos dicho que, si bien podemos calcular la probabilidad de un evento abstracto que involucre varios estados de E , éste siempre *estará* en uno y solo uno de los estados de E .¹⁶

Un vector $|\varphi\rangle$ del espacio de Hilbert al tiempo t se denota por $|\varphi, t\rangle$ [6]. Consideremos la amplitud de probabilidad asociada a la transición $|i, s\rangle \mapsto |j, t\rangle$ entre dos vectores base a distintos tiempos:

$$\mathcal{A}_{ji}(t, s) := \langle j, t | i, s \rangle \quad (3.48)$$

Vamos a probar que las amplitudes (3.48) pueden satisfacer una ecuación similar a la de Chapman - Kolmogorov (3.32).

Volvamos a las similitudes, la noción de estados clásicos mutuamente excluyentes se traduce a vectores mutuamente ortonormales, que denotamos por el producto interno

$$\langle j | i \rangle = \delta_{ji}. \quad (3.49)$$

¹⁵También llamados *estados*, en este caso, *cuánticos*.

¹⁶Aunque no seamos capaces de decir en cuál estado específicamente, de ahí la necesidad de asignar medidas de probabilidad en los sistemas clásicos.

En esta teoría la condición de ortonormalidad se conserva para todo t , con lo cual

$$\langle j, t | i, t \rangle = \delta_{ji}, \quad (3.50)$$

que corresponde simbólicamente a $\mathcal{P}_{ji}(t, t) = \delta_{ji}$ en (3.31), aunque su interpretación no sea la misma.

De la completitud de la base podemos expresar el operador identidad del espacio de Hilbert en cuestión como

$$\hat{I} = \sum_{k \in \mathcal{I}} |k\rangle \langle k|. \quad (3.51)$$

Puesto que (3.49) vale para todo t , dando lugar a (3.50), lo mismo ocurre con (3.51), que también vale para todo t y dando lugar a

$$\hat{I} = \sum_{k \in \mathcal{I}} |k, t\rangle \langle k, t|. \quad (3.52)$$

Finalmente, insertando el operador identidad (3.52) para $t = r$, en la amplitud de probabilidad (3.48), i.e. $\langle j, t | \hat{I} | i, s \rangle$, tendremos

$$\mathcal{A}_{ji}(t, s) = \sum_{k \in \mathcal{I}} \mathcal{A}_{jk}(t, r) \mathcal{A}_{ki}(r, s). \quad (3.53)$$

Esto prueba la estructura compartida con la ecuación de Chapman - Kolmogorov (3.33), aunque sin restricción en el orden temporal [25].

No es casual que una teoría como esta tenga una estructura similar, en cuanto a su evolución temporal determinista, que la que hemos descrito sobre procesos estocásticos. En este contexto se puede introducir un operador lineal y unitario, $\hat{U}(t, s)$, que hace las veces de $\mathcal{P}(t, s)$. Podemos representar a $\hat{U}(t, s)$ en la base completa que hemos introducido, de manera que cada *elemento de matriz*

$$U_{ji}(t, s) = \langle j | \hat{U}(t, s) | i \rangle. \quad (3.54)$$

Toda transición de vectores base $|i, s\rangle \mapsto |j, t\rangle$ está mediada por el operador $\hat{U}(t, s)$. La afirmación anterior vale en caso de que

al sistema no se le perturbe inoportunamente, por ejemplo con alguna medición.

De esta forma una amplitud de probabilidad (3.48) corresponde a los elementos de matriz (3.54), i.e. $\mathcal{A}_{ji}(t, s) = U_{ji}(t, s)$.

Puesto que un vector $|\varphi\rangle$ se puede escribir como una combinación lineal o superposición de vectores base, éste hereda el mediador $\hat{U}(t, s)$ como conexión entre la transición de vectores a distintos tiempos. Es decir, que $|\varphi, s\rangle \mapsto |\varphi, t\rangle$ también corresponde a la transición

$$|\varphi, t\rangle = \hat{U}(t, s)|\varphi, s\rangle. \quad (3.55)$$

De esta forma tendremos que $U(t, s)$ satisface

$$\hat{U}(t, s) = \hat{U}(t, r) \hat{U}(r, s), \quad (3.56)$$

que es semejante a (3.33), aunque aquí no tenemos necesariamente la restricción temporal $t \geq r \geq s$ [25]. De hecho, podemos repetir el procedimiento incluso con el orden temporal inverso $t \leq r \leq s$. Este es un aspecto fundamentalmente diferente de $\hat{U}(t, s)$ respecto de las probabilidades condicionales *a la* Kolmogorov $\mathcal{P}(t, s)$.

Por otro lado todo vector $|\varphi, t\rangle$ puede ser escrito a partir de sus coordenadas en la base completa

$$\varphi_j(t) := \langle j|\varphi, t\rangle, \quad (3.57)$$

para cada j . Estas funciones $t \mapsto \varphi_j(t)$ toman valores sobre \mathbb{C} .

Pero también tenemos otro tipo de coordenadas, igualmente importantes. En esta versión de la mecánica cuántica sobre espacios de Hilbert numerables contamos con coordenadas φ_k y sus conjugadas φ_k^* , ambas independientes para una descripción completa del problema.

Veamos esto a partir del cálculo del *valor medio* de un operador A sobre un vector $|\varphi\rangle$, para ello utilizaremos la siguiente expresión [4]

$$\langle A \rangle_\varphi = \sum_{k, l \in \mathcal{I}} \varphi_k^* A_{kl} \varphi_l. \quad (3.58)$$

Para calcular $\langle A \rangle_\varphi$ requerimos los elementos de matriz A_{kl} y además *ambos* conjuntos de coordenadas del vector $|\varphi\rangle$, i.e. $\{\varphi_k^*, \varphi_k\}$, todos ellos expresados en la misma base. Incluso en el caso en que el operador A resulte diagonal en esta base, de igual forma necesitaremos de los dos conjuntos de coordenadas para calcular $\langle A \rangle_\varphi$, pues en ese caso tendré que la suma doble (3.58) se reduce a una suma simple sobre k de elementos $\varphi_k^* A_{kk} \varphi_k$.

A partir de (3.55) podemos escribir una relación de coordenadas a distintos tiempos

$$\varphi_j(t) = \sum_{i \in \mathcal{I}} U_{ji}(t, s) \varphi_i(s). \quad (3.59)$$

Esta estructura deja entrever algo interesante: las amplitudes probabilidad U_{ji} y un vector del espacio de Hilbert como condición inicial, determinan la dinámica del sistema para todo tiempo¹⁷.

Podemos obtener también una relación similar a (3.59) pero temporalmente inversa, para las coordenadas conjugadas $\{\varphi_i^*\}$

$$\varphi_i^*(s) = \sum_{j \in \mathcal{I}} \varphi_j^*(t) U_{ji}(t, s). \quad (3.60)$$

De igual manera similar las amplitudes probabilidad U_{ji} y un vector del espacio de Hilbert como condición final, determinan la dinámica del sistema para todo tiempo.

Las coordenadas $\{\varphi_j\}$ evolucionan temporalmente como vectores (columnas), mientras que las coordenadas $\{\varphi_i^*\}$ evolucionan temporalmente como co-vectores (filas) [26].

Las expresiones (3.59) y (3.60) tienen una estructura similar a (3.42) y (3.43), que reescribimos

$$\sum_{i \in \mathcal{I}} \mathcal{P}_{ji}(t, s) \alpha_i(s) \quad (3.61)$$

$$\sum_{j \in \mathcal{I}} w_j(t) \mathcal{P}_{ji}(t, s) \quad (3.62)$$

¹⁷Como siempre dejando al sistema cuántico librado a sí mismo, sin hacer alusión a ninguna interacción foránea al sistema, por ejemplo una medida.

pero con una diferencia sutil. En mecánica cuántica la condición inicial o final es impuesta sobre φ o φ^* , en general distintos, aunque relacionados por conjugación compleja.

Tomando el conjugado de complejo de (3.60), usando que $\hat{U}(t, s)$ es unitario y satisface (3.56), luego

$$\varphi_i(s) = \sum_{j \in \mathcal{I}} U_{ij}(s, t) \varphi_j(t), \quad (3.63)$$

que corresponde a una ecuación idéntica a (3.59) pero intercambiando el orden temporal sin problema alguno. Es decir, intercambia el tipo de preguntas prospectivas y retrospectivas. Diremos que la mecánica cuántica contesta a preguntas prospectivas y retrospectivas con igual eficacia e idéntica respuesta. Utilizamos el mismo operador unitario para hacer evolucionar temporalmente a φ en un sentido dado y en el sentido inverso.

Esta es una diferencia fundamental entre la mecánica cuántica y los procesos estocásticos. En es caso una retrospección no corresponde lisa y llanamente a invertir el orden temporal en las probabilidades de transición, puesto que $\mathcal{P}(t, s)$ está definida para $t \geq s$. También podemos encontrar otra diferencia entre proyección y retrospección puesto que no siempre existe una operación que nos permita conectar (3.61) con (3.62), esencialmente porque no siempre existe conexión entre las distribuciones inicial $\{\alpha_i(s)\}_{i \in I}$ y final $\{w_j(t)\}_{j \in I}$.

La mecánica cuántica tiene aún más matices, la base de vectores puede estar degenerada, por nombrar uno de ellos [6]. Tal y como hemos adelantado un sistema cuántico puede estar en una superposición de estados, incluso estar entrelazados. Estos conceptos no tienen correlato en el mundo clásico, en el sentido en que hemos desarrollado nuestro cúmulo de similitudes.

Hemos comenzado a encontrar diferencias entre ambas teorías y, por otro lado, también cierta autonomía : es el momento de retirar el andamiaje cuántico, que nos ha ayudado en esta construcción.

Reanudando el curso específico hacia procesos estocásticos *kolmogorovianos*, ahora con una clara orientación desarrollada en las dos secciones anteriores, planteamos el siguiente problema: queremos asociar probabilidades absolutas a las ecuaciones de Kolmogorov forward (3.40) y backward (3.41), que hagan las veces de coordenadas.

Veremos que la forma natural en que esto puede llevarse a cabo, desde el contexto de la teoría de probabilidades, arroja resultados similares a la que hemos visto para la mecánica cuántica.

Consideremos la medida conjunta $P(x(t), x(s))$ con $s \leq t$ que posteriormente marginalizaremos sobre cada variable. Por razones que luego quedarán más claras, llamaremos contexto *prospectivo* (o forward) al caso en que marginalicemos sobre $x(s)$ y contexto *retrospectivo* (o backward) cuando hagamos lo propio sobre $x(t)$.

Para esta tarea debemos tener en cuenta que la idea de *probabilidades condicionales*, puede dar lugar a algunas confusiones, en particular porque implica los términos *condicionante* y *condicionado* que a secas pueden remitir a cierto orden temporal estricto; por lo que limitaremos su uso para no incurrir en un desatino. Puede que sea mejor reemplazarlo por *probabilidades de transición*. Sobre esto volveremos con un importante comentario al final de este capítulo.

Cuando trabajemos en un contexto prospectivo no tendremos este tipo de problemas, pero cuando hagamos lo propio en el contexto retrospectivo, pueden surgir inconsistencias propias del lenguaje. Precisamente para evitar estos problemas hemos decidido iniciar el camino a partir de la medida conjunta $P(x(t), x(s))$.

Dada la probabilidad conjunta $P(x(t), x(s))$, procedamos a marginalizar sobre $x(s)$ para obtener una distribución unidimensional sobre $x(t)$. Denotemos esto de manera simplificada como

sigue

$$P(x(t)) = \sum_{x(s)} P(x(t), x(s)), \quad (3.64)$$

incorporando (3.2) tendremos

$$P(x(t)) = \sum_{x(s)} P(x(t)|x(s))P(x(s)) \quad (3.65)$$

Hemos dicho ya que el procedimiento de marginalización corresponde a *sumar* (o integrar en el caso en que E sea denso, por ejemplo algún subconjunto de \mathbb{R}) sobre la variable *marginal*, en este caso $x(s)$. Realmente esto es una suma sobre los valores que $x(s)$ pueda tomar, es decir, sobre todos los estados de $E = \{x_i\}_{i \in I}$.

Repitiendo el procedimiento anterior, ahora con una notación más precisa se tiene

$$P(x(t) = x_j) = \sum_{i \in I} P(x(t) = x_j, x(s) = x_i) \quad (3.66)$$

Denotando por $q_j(t)$ a la probabilidad absoluta $P(x(t) = x_j)$ en el contexto prospectivo tendremos

$$q_j(t) = \sum_{i \in I} \mathcal{P}_{ji}(t, s) q_i(s). \quad (3.67)$$

En la figura 3.2 hemos esquematizado el espacio de estados E . No estamos insinuando que E conforme un plano, simplemente representamos un *lugar* de pertenencia de los elementos $\{x_i\}_{i \in I}$. No hemos querido utilizar una línea para representar a E , puesto que tampoco queremos insinuar relación de orden alguna entre sus elementos.

Las probabilidades de transición $\{\mathcal{P}_{ji}(t, s)\}$, con $t \geq s$ conectan probabilidades absolutas en s (pasado) con probabilidades absolutas en t (futuro), lo cual corresponde a una *prospección*.

Al realizar un procedimiento similar para el contexto backward, debemos tener precaución en el último paso. En términos más precisos no podemos asignar $P(x_i, s|x_j, t)$ a $\mathcal{P}_{ij}(s, t)$, pues

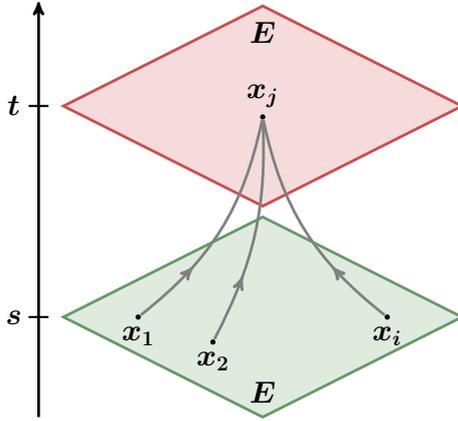


Figura 3.2: Esquema prospectivo: a partir de las distribuciones absolutas de la variable aleatoria $x(s)$, denotada por $\{q_i\}_{i \in I}$ al instante s , puede construirse q_j en t , según (3.67).

ambos objetos están definidos para $t \geq s$, en particular el segundo (3.46). De manera que t estuviese en el estado condicionado y no en el condicionante. Este mismo problema fue observado por Kolmogorov en un contexto denominado por él como *probabilidades de transición inversas*. Aunque fue tratado de manera lateral y ligeramente diferente a la forma que presentamos en este trabajo. En tal ocasión, Kolmogorov abordó el problema backward definiendo probabilidades de transición inversas, denotadas como $\Pi_{ji}(t, s)$ y en general distintas a las $\mathcal{P}_{ji}(t, s)$ [27]. En el presente trabajo haremos lo propio de manera equivalente con una ventaja adicional: probaremos que todo proceso backward proviene de cierta inversión temporal de un proceso forward. Esto apunta a resaltar la dualidad *prospección-retrospección* que subyace en los procesos estocásticamente deterministas.

Recordemos que las preguntas forward y backward no son respondidas, a priori, por la misma distribución de probabilidad, de igual forma que $\sum_{i \in I} \mathcal{P}_{ji}(t, s) \alpha_i(s)$ y $\sum_{j \in I} w_j(t) \mathcal{P}_{ji}(t, s)$, según (3.42) y (3.43), no son necesariamente iguales. Por tal motivo, utilizaremos una medida de probabilidad absoluta distinta

para la descripción backward. Repetimos el procedimiento forward, pero marginalizando la distribución conjunta ahora sobre $x(t)$, por lo que tendremos $P^\star(x(s))$, otra medida de probabilidad absoluta¹⁸. En general, P^\star es una medida distinta de P .

$$P^\star(x(s)) = \sum_{x(t)} P^\star(x(s)|x(t))P^\star(x(t)) \quad (3.68)$$

Definimos entonces el objeto válido para $s \leq t$

$$\mathcal{P}_{ij}^\star(s, t) := P^\star(x(s) = x_i | x(t) = x_j). \quad (3.69)$$

Denotando por $p_i(s)$ a la probabilidad absoluta $P^\star(x(s) = x_i)$ en el contexto backward tendremos

$$p_i(s) = \sum_{j \in I} \mathcal{P}_{ij}^\star(s, t) p_j(t), \quad (3.70)$$

Las probabilidades de transición $\{\mathcal{P}_{ij}^\star(s, t)\}$, con $t \geq s$ conectan probabilidades absolutas en t (futuro) con probabilidades absolutas en s (pasado), lo cual corresponde a una *retrospección*. Es la pregunta por la medida de probabilidad la que va hacia atrás, no el proceso estocástico. En la figura 3.3 podemos asociar una imagen que representa la retrodicción que hemos introducido.

El miembro derecho de la expresión (3.70) no corresponde aún a (3.43), de la página 57, que reescribimos

$$\sum_{j \in I} w_j(t) \mathcal{P}_{ji}(t, s), \quad (3.71)$$

incluso en el supuesto caso en que nuestra distribución final $w_j(t)$ sea igual a $p_j(t)$ para todo j . En tal caso, una condición suficiente para ello sería

$$\mathcal{P}_{ij}^\star(s, t) = \mathcal{P}_{ji}(t, s), \quad (3.72)$$

¹⁸Colocamos el superíndice \star en P^\star , pues queremos denotar que \star conjuga la medida P , dejando P^\star ; de manera semejante al rol de la conjugación compleja $*$, que relaciona φ con φ^* . Por supuesto \star no corresponde a la conjugación compleja $*$ sino a otra clase de conjugación cuyo sentido quedará más claro en las secciones subsiguientes.

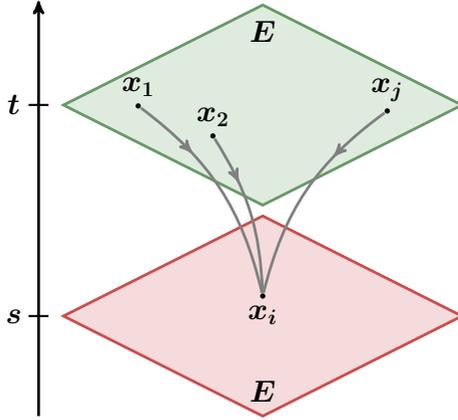


Figura 3.3: Esquema retrospectivo: a partir de las distribuciones absolutas de la variable aleatoria $x(t)$, denotada por $\{p_j\}_{j \in I}$ al instante t , puede reconstruirse p_i en s , según (3.70).

la cual no tiene una interpretación directa y sencilla. En caso en que (3.72) fuese verdadera ya podríamos afirmar que $\mathcal{P}_{ij}^*(s, t)$ y $p_i(s)$ (o P^*) responderán perfectamente a las preguntas backward **c.b.** y **a.b.** de la página 56. En la sección siguiente confeccionaremos el contexto para probar (3.72).

Antes de terminar esta sección, podemos extraer más información de la estructura que surge de (3.70), pues permite probar la siguiente propiedad *kolmogorviana* para $\mathcal{P}_{ij}^*(s, t)$

$$\mathcal{P}_{ij}^*(s, t) = \sum_{k \in I} \mathcal{P}_{ik}^*(s, r) \mathcal{P}_{kj}^*(r, t), \quad (3.73)$$

$$\mathcal{P}^*(s, t) = \mathcal{P}^*(s, r) \mathcal{P}^*(r, t) \quad (3.74)$$

válidas para $s \leq r \leq t$. La demostración (3.74) es simple: escribimos (3.70) en forma compacta como

$$p(s) = \mathcal{P}^*(s, t)p(t) \quad (3.75)$$

que vale para todo par de tiempos $s \leq t$. Tomemos la misma expresión pero para $s \leq r$ y otra para $r \leq t$

$$p(s) = \mathcal{P}^*(s, r)p(r), \quad p(r) = \mathcal{P}^*(r, t)p(t) \quad (3.76)$$

a partir de estas dos ecuaciones (3.76), reemplazamos la segunda en la primera, luego igualamos con (3.75) y tenemos (3.74).

También podríamos definir un generador infinitesimal $\Omega^*(s)$ asociado a $\mathcal{P}^*(s, t)$

$$\Omega^*(s) := [\partial_s \mathcal{P}^*(s, t)]_{t=s}. \quad (3.77)$$

A partir de (3.74) y (3.77) podemos incluso deducir una ecuación para $\mathcal{P}^*(s, t)$

$$\partial_s \mathcal{P}^*(s, t) = \Omega^*(s) \mathcal{P}^*(s, t). \quad (3.78)$$

Todo esto no es casualidad, ya hemos dicho que $\mathcal{P}^*(s, t)$ corresponde a una evolución temporal *inversa*, i.e. hacia el *pasado*, de las probabilidades absolutas P^* . En la sección siguiente demostraremos que (3.78) satisface la ecuación backward de Kolmogorov (3.41), vía una *inversión temporal* que conecta $\mathcal{P}^*(s, t)$ con $\mathcal{P}(t, s)$.

Inicialmente hemos designado una nomenclatura para el par (P, P^*) a las distribuciones (*forward, backward*), por ser éstas mutuamente conjugadas, sin embargo no explicitamos aún el contexto de esta afirmación.

Posteriormente hemos elegido la nomenclatura de coordenadas específica¹⁹ (q_j, p_j) , de la cual veremos que tendrá especial sentido en el capítulo siguiente. Allí desarrollaremos de manera precisa el contexto que sostiene la afirmación acerca de la conjugación entre estas probabilidades absolutas. Probaremos que cada par de coordenadas (q_j, p_j) prospectivas y retrospectivas, resultan ser variables conjugadas vía cierto lagrangiano L [30, 31].

Tampoco hemos dado cierre a la pregunta ¿quién o qué es $\mathcal{P}_{ij}^*(s, t)$? Solo hemos impuesto que cada una de las $\mathcal{P}_{ij}^*(s, t)$ y $p_i(s)$ deben contestar a las preguntas retrospectivas **c.b.** y **a.b.** de la sección 3.6. Esta cuestión será el motivo de la próxima sección, donde buscaremos una relación entre $\mathcal{P}_{ij}^*(s, t)$ y $\mathcal{P}_{ji}(t, s)$.

¹⁹En este momento, podemos justificar la elección de los símbolos p y q pues poseen un orden alfabético *antes-después*, que requerimos para denotar una *retrodicción-predicción*.

INVERSIÓN TEMPORAL, UN INTERCAMBIO PRO \rightleftharpoons RETRO

El título de esta sección ha logrado sintetizar el propósito de la misma. Desarrollaremos la idea intuitiva acerca de que la inversión temporal sobre una prospección corresponde a una retrospección y viceversa. La inversión temporal corresponde a un intercambio en los prefijos *pro* y *retro*.

En la sección anterior hemos visto que (3.67), que reescribimos

$$q_j(t) = \sum_{i \in I} \mathcal{P}_{ji}(t, s) q_i(s), \quad (3.79)$$

responde a un proceso forward, en el sentido en que satisface $q_j(t)$ satisface la ecuación forward de Kolmogorov (3.40) para s fijo.

Aún nos resta probar que las probabilidades absolutas p_i de (3.70) responden al proceso backward y que una *inversión temporal* hará las veces de puente entre estas dos descripciones *duales* del mismo proceso estocástico.

Siempre que tratamos al tiempo o al concepto de inversión temporal nuestro cerebro, mejor dicho el cerebro del *autor* de este trabajo, posee grandes dificultades para procesarlo. Por esta razón necesitamos construir esta idea de inversión temporal, intentando no perder la cordura (cosa que ya ha hecho el *autor* de este trabajo), haciendo uso de unas bases lo suficientemente estables.

Utilizaremos entonces una *cronología*²⁰, i.e. una teoría del tiempo local, expuesta en [4, 29]. Allí se aborda el concepto de la *anisotropía del tiempo*. Esta teoría analiza el concepto de *función tiempo* a partir de tres conceptos abstractos: suceso²¹, marco de

²⁰*Cronología: khrónos* (tiempo)–*logos* (tratado o estudio).

²¹En [4, 29] se utiliza el término *evento* en lugar de *suceso*. En este trabajo hemos utilizado este último término con toda premeditación para evitar confusiones con los elementos de una σ -álgebra sobre el conjunto de eventos elementales ω . Además el término suceso se podrá aplicar tanto eventos ($\omega \in \Omega$) como a estados ($x_i \in E$) del proceso estocástico.

referencia y escala cronométrica. Los últimos dos los supondremos fijos, por lo que podemos simplificar la notación de [4, 29]. Consideraremos una función tiempo T que asigna un número real a dos sucesos (s, s') del conjunto \mathcal{S} , $T : \mathcal{S} \times \mathcal{S} \rightarrow \mathbb{R}$. Denotamos por $t = T(s, s')$ a la duración o lapso de tiempo entre dos sucesos (s, s') .

Para fijar ideas veamos uno de los axiomas de esta teoría del tiempo [29], adaptados a esta notación

A. *Las duraciones contiguas son aditivas*, en símbolos:

$$\forall s, s', s'' \subset E \Rightarrow T(s, s') + T(s', s'') = T(s, s'').$$

De aquí surge el siguiente teorema

T. *La duración es un intervalo orientado*, en símbolos:

$$\forall s, s' \subset E \Rightarrow -T(s, s') = T(s', s). \quad (3.80)$$

La demostración de **T** es elemental. A partir de $T(s, s) = 0$ ²² y por el axioma **A** junto con $s'' = s$. En este sentido, la función duración T es *antisimétrica* respecto a los sucesos.

Una lectura literal de la tesis del teorema **T** puede ser: la inversión del signo de $t = T(s, s')$ corresponde a una inversión del orden de aparición de los sucesos (s, s') . En otras palabras, si inicialmente s precedía a s' , ahora será al revés: una inversión temporal corresponde a una inversión de procesos [29]. Denotemos s precede a s' como $s \rightsquigarrow s'$.

Dado un proceso estocástico, una inversión temporal sobre éste corresponde también a otro proceso estocástico que, al igual que el primero, evoluciona *hacia adelante* en el tiempo; aunque el orden de una secuencia de sucesos $s \rightsquigarrow s'$ se invierte por $s' \rightsquigarrow s$.

Pasemos ahora a procesos estocásticos, queremos construir un proceso *time reversal* asociado a un proceso dado. Intuitivamente

²²En [29] tenemos que $T(s, s) = 0$ es un axioma de la teoría. En general dos sucesos (s, s') se dicen *simultáneos* si y solo si $T(s, s') = 0$ [4].

tenemos que las probabilidades de transición ordenan los sucesos, de manera propensiva por cierto. En este contexto, podemos considerar que los sucesos sean precisamente elementos de nuestro conjunto de estados $E = \{x_1, x_2, \dots\}$. Cuando decimos que el suceso s procede al suceso s' requerimos también asignar a cada uno de ellos una coordenada temporal digamos s y t , a cada uno de ellos, con tal que $s \leq t$.

Tenemos entonces tres pares de objetos: sucesos (s, s') , estados (x_i, x_j) y tiempos (s, t) . La afirmación $s \rightsquigarrow s'$ corresponde entonces a la transición $(x_i, s) \mapsto (x_j, t)$, que está mediada por $\mathcal{P}_{ji}(t, s)$, con $s \leq t$. Denotemos por \mathbf{p} al *tipo* de preguntas **prospectivas** sobre el proceso estocástico. La inversión temporal de esta pregunta prospectiva \mathbf{p} atañe a la afirmación $s' \rightsquigarrow s$, la cual corresponde a la transición $(x_j, s) \mapsto (x_i, t)$, que está medida por $\mathcal{P}_{ij}(t, s)$. Denotemos por \mathbf{p}^* al *tipo* de preguntas **retrospectivas** sobre el proceso estocástico. Tal y como hemos adelantado: en ambos casos conservamos el orden temporal en la transición, i.e. $s \leq t$, pero invertimos el orden de precedencia de los sucesos.

En la página 56 hemos presentado dos clases complementarias de preguntas absolutas, a saber: **a.f.** y **a.b.** (o bien las condicionales **c.f.** y **c.b.**); las cuales desde ahora quedarán identificadas por las clases \mathbf{p} y \mathbf{p}^* , respectivamente. No obstante, ambas clases refieren al mismo proceso estocástico.

Tenemos entonces que la inversión temporal del proceso descrito por $\mathcal{P}_{ji}(t, s)$ corresponde a $\mathcal{P}_{ij}(t, s)$. Definimos una operación de inversión temporal \mathfrak{I} que afecta a estos elementos de matriz como sigue:

$$\mathfrak{I}(\mathcal{P}_{ji}(t, s)) := \mathcal{P}_{ij}(t, s). \quad (3.81)$$

En términos matriciales \mathfrak{I} corresponde a una trasposición t de filas por columnas:

$$\mathfrak{I}(\mathcal{P}(t, s)) := \mathcal{P}^t(t, s). \quad (3.82)$$

Por otro lado, hemos dicho que un proceso backward evolucionaria, también, hacia adelante en el tiempo. Precisamente pre-

servamos esto en cada una de las probabilidades de transición $\mathcal{P}(t, s)$ y $\mathcal{P}^*(s, t)$, ambas bien definidas según el mismo orden temporal $s \leq t$.

En la sección anterior hemos planteado argumentos de plausibilidad acerca de que la marginalización de $P(x(t), x(s))$ sobre $x(t)$ (3.70) pudiera estar asociada al proceso backward, en el sentido señalado por (3.71). Podremos asociar las coordenadas $\{p_i\}$ a la ecuación backward, si exigimos que $\mathcal{P}_{ij}^*(s, t)$ corresponda a la inversión temporal de $\mathcal{P}_{ij}(t, s)$, esto es

$$\mathcal{P}_{ij}^* = \mathfrak{I}(\mathcal{P}_{ij}) \quad (3.83)$$

tenemos automáticamente por (3.82) que la relación a distintos tiempos entre coordenadas $\{p_i\}$ (3.70) toma la forma

$$p_i(s) = \sum_{j \in I} p_j(t) \mathcal{P}_{ji}(t, s), \quad (3.84)$$

que es precisamente la estructura (3.71) que queríamos conseguir.

La evolución temporal está gobernada por las probabilidades de transición $\mathcal{P}(t, s)$, de manera que las coordenadas $\{q_j\}$ evolucionan temporalmente como vectores (columnas), mientras que las coordenadas $\{p_i\}$ evolucionan temporalmente como co-vectores (filas) [26].

Además hemos conseguido dar un significado a $\mathcal{P}_{ij}^*(s, t)$ en términos de $\mathcal{P}_{ji}(t, s)$, ambas bien definidas para $s \leq t$. Ya hemos demostrado que las medidas condicionales $\{\mathcal{P}_{ij}^*(s, t)\}_{i,j \in I}$ satisfacen una ecuación similar a la de Chapman - Kolmogorov (3.74), que reescribimos:

$$\mathcal{P}_{ij}^*(s, t) = \sum_{k \in I} \mathcal{P}_{ik}^*(s, r) \mathcal{P}_{kj}^*(r, t), \quad (3.85)$$

con $s \leq r \leq t$. Por otro lado, en esta sección hemos conectado a las medidas de transición \mathcal{P} y \mathcal{P}^* , vía una inversión temporal

$$\mathfrak{I} : \mathcal{P} \longmapsto \mathcal{P}^*. \quad (3.86)$$

Podemos utilizar esta conexión para comprobar la buena definición de la inversión temporal que hemos introducido. En primer lugar y en virtud de la conexión (3.86) podemos demostrar (de manera alternativa) la ecuación (3.85), a partir de que \mathcal{P} satisfaga la ecuación de Chapman - Kolmogorov (3.32). Para finalizar estas comprobaciones, calculemos ahora el generador infinitesimal Ω^* asociado a \mathcal{P}^* , que hemos definido en (3.77); introduciendo ahora la conexión (3.83) tenemos que

$$\Omega^*(s) = -\Omega^t(s) \quad (3.87)$$

con lo cual podemos ver que $\mathcal{P}^*(s, t)$ junto con $\Omega^*(s)$ satisfacen la ecuación backward de Kolmogorov (3.41).

En resumen, derivando respecto de t (o s) $\mathcal{P}(t, s)$ obtenemos la ecuación forward (o backward) de Kolmogorov. Por otro lado derivando respecto de s (o t) $\mathcal{P}^*(s, t)$ obtenemos la ecuación backward (o forward) de Kolmogorov.

En general estamos habituados a entender la inversión temporal, buscando en los objetos alguna flecha que, asociada a la dinámica, sea susceptible de ser invertida. Queremos ser enfáticos en esto: no invertimos el *curso* del tiempo, en tal caso, invertimos el orden de cierta secuencia de sucesos que tiene lugar *en* el tiempo. De hecho, en la página 75 hemos dicho que dada una secuencia de sucesos $s \rightarrow s'$ un inversión temporal de esta secuencia corresponde a considerar otra secuencia, i.e. $s' \rightarrow s$, que no es otra cosa que una inversión de la flecha de la primera de ellas, esto es $s \leftarrow s'$.

Vamos a presentar otra forma de abordar el tema de la inversión temporal de una forma más gráfica, utilizando una forma de esquematizar estos procesos estocásticos.

Comenzamos por un estado x_i , que estará representado por $\textcircled{x_i}$. Dados estados, podemos conectarlos mediante líneas. Cada línea tiene asignada una probabilidad de transición \mathcal{P}_{ji} , el primer índice hace referencia al estado de llegada. Debemos distinguir a las líneas, según el sentido de la conexión, ya que en general

$\mathcal{P}_{ji} \neq \mathcal{P}_{ij}$: las líneas entre estados distintos tienen una *dirección*, en adelante flechas y que representamos como $\textcircled{x_i} \longrightarrow \textcircled{x_j}$.

Tenemos también probabilidades de *auto-transición* o *permanencia*, dadas por \mathcal{P}_{ii} . Los estados representados por $\textcircled{x_i}$ pueden por tanto estar conectados a sí mismos mediante un bucle, i.e. una línea que se cierra sobre el mismo estado y que representamos como $\textcircled{x_i} \curvearrowright$.

Para cada estado x_i existen tres tipos de conexiones: las que *salen* de x_i a algún estado distinto: **transiciones salientes de x_i** , las que *entran* a x_i desde algún estado distinto: **transiciones entrantes a x_i** y las que se pliegan sobre *sí mismas*: auto-transiciones en x_i . La figura 3.4 resume estas ideas.

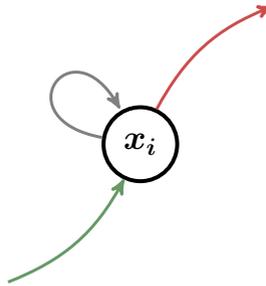


Figura 3.4: Conexiones para un estado x_i .

Dado un proceso estocástico podemos entenderlo como un conjunto de puntos, interconectados uno a uno mediante elementos de matriz de transición, como este objeto tiene dos entradas solo basta con describir la *interacción* de dos estados, distintos, digamos x_i y x_j . Resumimos estas nociones en la figura 3.5.

Puesto que cada una de las líneas del diagrama está ponderada por una probabilidad de transición y ello implica una direccionalidad, tendremos entonces un *digrafo ponderado*. El término *digrafo* proviene de la fusión **directed-graph**, que al castellano nos deja *grafo dirigido*. Los procesos estocásticos descritos hasta ahora pueden representarse como un *digrafo ponderado* [32].

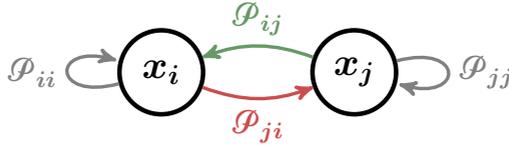


Figura 3.5: Diagrama esquemático de la interacción de dos estados distintos x_i y x_j asociado al proceso estocástico, en el que se responden preguntas de carácter prospectivo (\mathbf{p}).

Intuitivamente una inversión temporal de un proceso se construye tomando su digrafo ponderado asociado e invirtiendo el sentido de todas sus aristas y ponderándolas con la medida \mathcal{P}^* . Se siguen las mismas reglas: dados dos nodos, podemos conectarlos mediante aristas que representamos como $(x_i) \rightarrow (x_j)$. Cada arista tiene asignada una probabilidad de transición \mathcal{P}_{ji}^* , el primero de sus índices, j , hace referencia al estado de llegada mientras que el segundo, i , hace referencia al estado de salida. En la figura 3.6, podemos ver un esquema de interacción de dos nodos del grafo asociado al proceso temporalmente inverso.

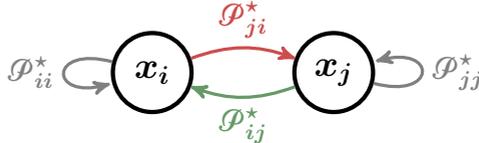


Figura 3.6: Diagrama esquemático de la interacción de dos estados distintos x_i y x_j asociados al proceso estocástico en el que se responden preguntas de carácter retrospectivo (\mathbf{p}^*).

Presentamos dos formas de abordar el problema de la inversión temporal a partir de su digrafo ponderado.

1. Forma constructiva

Dado proceso estocástico con medida \mathcal{P} , construimos su proceso estocástico temporalmente inverso, invirtiendo el sentido de todas las aristas de su digrafo y reponderándolas con la medida inversa $\mathfrak{I}(\mathcal{P})$.

2. Forma comparativa

Un proceso estocástico con medida condicional \mathcal{P}^* y preguntas tipo \mathbf{p}^* corresponde a una inversión temporal del proceso estocástico con medida condicional \mathcal{P} y preguntas tipo \mathbf{p} si y solo si las aristas del grafo asociadas a \mathcal{P}^* están ponderadas con $\mathfrak{I}(\mathcal{P})$, i.e. $\mathcal{P}_{ij}^* = \mathfrak{I}(\mathcal{P}_{ij})$.

Concluimos esta sección con un comentario sobre la invariancia bajo inversión temporal \mathfrak{I} . La cual podemos definir como

$$\mathfrak{I}(\mathcal{P}) = (\mathcal{P}), \quad (3.88)$$

esto ocurre si y solo si para cada par de estados (x_i, x_j) de E

$$\mathcal{P}_{ji} = \mathcal{P}_{ij}. \quad (3.89)$$

El generador infinitesimal Ω satisface en este caso

$$\Omega_{ji} = \Omega_{ij}. \quad (3.90)$$

DISTRIBUCIONES PROSPECTIVAS Y RETROSPECTIVAS

Iniciamos este capítulo hablando de predicción y retrodicción como formas duales en la descripción de un proceso. Sobre este punto y para procesos estocásticamente determinados tenemos una medida de probabilidad condicional que satisface las ecuaciones de Chapman-Kolmogorov. De allí podemos derivar el par de ecuaciones de Kolmogorov forward-backward para la probabilidad condicional (3.40)-(3.41), que reescribimos

$$\partial_t \mathcal{P}_{ji}(t, s) = \sum_{k \in I} \Omega_{jk}(t) \mathcal{P}_{ki}(t, s) \quad (3.91)$$

$$\partial_s \mathcal{P}_{ji}(t, s) = - \sum_{k \in I} \mathcal{P}_{jk}(t, s) \Omega_{ki}(s) \quad (3.92)$$

Las ecuación forward y backward hace referencia a un predicción y retrodicción, de allí que se las denominara ecuaciones prospectivas y retrospectivas [24].

Hemos logrado introducir un conjunto de distribuciones unidimensionales P y P^* a partir de marginalización la distribución conjunta.

Hemos utilizado luego una notación que explicita el estado particular que puede tomar la variable aleatoria x , sobre el que estamos interesados en calcular su medida de probabilidad:

$$q_j(t) = P(x(t) = x_j), \quad p_i(s) = P^*(x(s) = x_i). \quad (3.93)$$

Finalmente en la sección anterior, mostramos que $\mathcal{P}^*(s, t)$ satisface la ecuación backward de Kolmogorov (3.92) a través de una inversión temporal \mathfrak{J} : $\mathcal{P}^*(s, t) = \mathfrak{J}(\mathcal{P}(t, s))$.

Las distribuciones unidimensionales $q_j(t)$ y $p_i(s)$ evolucionan en el tiempo según

$$q_j(t) = \sum_{i \in I} \mathcal{P}_{ji}(t, s) q_i(s), \quad (3.94)$$

$$p_i(s) = \sum_{j \in I} p_j(t) \mathcal{P}_{ji}(t, s). \quad (3.95)$$

De forma que (3.94) y (3.95) satisfacen las ecuaciones forward y backward de Kolmogorov (3.91) y (3.92)

$$\dot{q}_j(t) = \sum_{i \in I} \Omega_{ji}(t) q_i(t), \quad (3.96)$$

$$\dot{p}_i(s) = - \sum_{j \in I} p_j(s) \Omega_{ji}(s). \quad (3.97)$$

denotaremos a la derivada total de cierta función $f(\gamma)$ sobre \mathbb{R} como $\dot{f}(\gamma)$ y otras como $d_\gamma f(\gamma)$.

Estas son las coordenadas prospectivas y retrospectivas que requerimos para describir un procesos estocástico determinista sobre un espacio de estados E numerable.

Podemos comprimir la notación en forma matricial

$$\dot{q}(t) = \Omega(t) q(t), \quad (3.98)$$

$$\dot{p}(s) = -\Omega^t(s) p(s). \quad (3.99)$$

ambas distribuciones como vectores columna.

Un comentario más para finalizar esta sección. Hemos dicho que una inversión temporal corresponde a una inversión de toda secuencia de estados, eso nos ha llevado a que la operación de inversión temporal definida sobre las probabilidad de transición traspone sus elementos, i.e. $\mathcal{T} : \mathcal{P}_{ji} \mapsto \mathcal{P}_{ij}$. Podemos notar que las ecuaciones (3.98) y (3.99) difieren entre si de la trasposición de Ω y un signo. En la sección anterior concluimos que una inversión temporal corresponde a una trasposición de Ω y puesto que el problema de inversión temporal responde a la pregunta retrospectiva **a.b.** de la página 56 (la cual va hacia atrás en el tiempo) esto conlleva a invertir el orden temporal en la derivada para la distribución backward. Este puede verse también a partir de (3.95), allí tenemos dos tiempos ordenados $s \leq t$, cuando tomamos la derivada \dot{p} para construir (3.97) o (3.99) aparece el signo $-$ y solo uno de los tiempos.

PROCESOS DE MARKOV

Llegamos aquí a la clase de procesos estocásticos que abarca la más amplia variedad de sistemas físicos, biológicos, económicos, etcétera.²³ Quizá la raíz de tal cuestión quede clarificada inspeccionando la definición de proceso de Markov.

²³En el sentido en que los describe de manera idónea.

En 1906 el matemático ruso A. Markov introdujo una nueva línea de investigación para una sucesión de variables aleatorias no independientes, sino *encadenadas*, de acuerdo con el término que él utilizó. Supongamos un proceso estocástico sobre el espacio de probabilidad (E, \mathfrak{S}, P) con una familia de variables aleatorias $x(t)$, para $t \in T$ un intervalo real. Se trata de la siguiente cuestión: asignar una medida de probabilidad a un suceso, dada su historia anterior y su estado actual. Los procesos de Markov tienen la propiedad fundamental de que esa medida no depende explícitamente de la historia previa, sino que dependen explícitamente del estado actual.

Utilizando la notación de la sección anterior, podemos condensar la propiedad de Markov formalmente como

$$P(\Delta_t | x(t), \nabla_t) = P(\Delta_t | x(t)), \quad (3.100)$$

donde ∇_t (Δ_t) contienen a todas las historias pasadas (futuras) del sistema respecto del presente, que está fijado en t [19, 20].

Todo proceso estocástico que satisface (3.100) se denomina de Markov. Algunos autores han denominado a (3.100) como *principio de Markov*, en el sentido en que un proceso estocástico puede realizarse una petición de *este* principio por hipótesis de trabajo.

Una profunda lectura de esta propiedad es la siguiente: la medida de un suceso, referido al futuro del sistema, dado su estado actual y su pasado, es igual a la medida de ese suceso referido al futuro del sistema dado su estado actual.

Muchos autores interpretan esta propiedad diciendo que los procesos de Markov no tienen memoria, que no dependen de su historia pasada, pero esto es semántica y lógicamente incorrecto, o bien, siendo algo más diplomáticos, es parcialmente correcto.

Para fijar ideas, podremos ayudarnos en la comprensión de la propiedad de Markov (3.100) utilizando el siguiente recurso, pensemos en un reloj particular de arena. La cual, como sabemos, fluye de un lado al otro del reloj, siempre atravesando la región central. Lo interesante es que en un reloj de este tipo la noción

temporal y espacial se entremezclan. Desarrollamos la metáfora como sigue: toda la arena que se encuentra en la cavidad superior (pasado), tiene que pasar por el centro (presente) para acumularse en la cavidad inferior (futuro). Para acercar la metáfora aún más a cierto proceso estocástico podemos asignar alguna irregularidad en la superficie del reloj o bien en el movimiento de los granos de arena, para dar lugar a la contingencia. No daremos más detalles acerca de la naturaleza del proceso, solo pensar en aquello que garantice aleatoriedad en el flujo de arena.

Realizando cortes transversales, i.e. *horizontales*, al reloj de manera que se obtengan una serie de planos etiquetados por un parámetro τ . Cado uno de los cuales interseca al reloj, de manera que cada τ refiere a cierta región *espacial* dentro del reloj, en la cual tenemos una sección plana, que llamaremos \mathbf{R}_τ con una cantidad de arena. En cada una de estas secciones planas podemos definir una densidad de arena que podemos cuantificar con la variable aleatoria $x(\tau)$. Supongamos que el centro del reloj se corresponde a cierto $\tau = t$. Cabe entonces la pregunta de asignar una medida de probabilidad $P(\Delta_t|x(t), \nabla_t)$, para entender la correlación entre las distintas regiones. En la figura 3.7 mostramos precisamente la correspondencia entre 3 instantes ordenados $s < t < u$ y 3 regiones del reloj, por éstos etiquetadas, denotadas por \mathbf{R}_s (*pasado*), \mathbf{R}_t (*presente*) y \mathbf{R}_u (*futuro*). En la me-

táfora propuesta, la arena se conglomerada en la región t y dada su estrechez, la densidad futura depende fuertemente o exclusivamente de la densidad en $x(t)$, por lo que se satisface (3.100) para cada t , es decir para cada *presente*. De esta forma los sucesos se concatenan linealmente, en el sentido de que lo que ocurra en el *futuro* del reloj, i.e. en las regiones \mathbf{R}_τ con $\tau > t$, depende del estado actual $x(t)$. Luego los sucesos del *pasado* del reloj, i.e. las regiones \mathbf{R}_τ con $\tau < t$, pueden *afectar*²⁴ a los del *futuro* del reloj

²⁴Debemos decir que los eventos o estados, o en general sucesos, no pueden considerarse como causas de las que surja *afectación* alguna, pero en el caso de

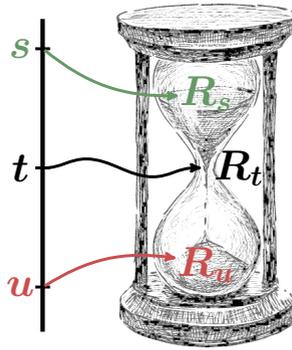


Figura 3.7: Proceso de Markov vía la metáfora del reloj de arena.

solo a través del estado actual. Los sucesos pasados $t' < t$ han de *afectar* al futuro solo cuando les toque ser presente, es decir cuando traslade la pregunta por la medida $P(\Delta_{t'}|x(t'), \nabla_{t'})$ a un nuevo presente en t' .

Al igual que en la metáfora del reloj de arena, podemos decir que en los procesos de Markov todas las historias pasadas y todas las historias futuras están *desconectadas* en el sentido de que son independientes o que no están correlacionadas directamente entre si. Aunque sí lo están de manera transitiva, pues ambas se encuentran *mediadas* por el presente, que las *interconecta*.²⁵

Queremos comprender mejor el tipo de procesos dentro de esta categoría *markoviana* definida por (3.100). Una forma interesante de comprimir lo que *es* puede ser comprendiendo aún mejor lo que *no es*. Por ejemplo, veamos que (3.100) no implica que el conjunto de historias pasadas sea independiente del estado presente y del conjunto de historias futuras. En símbolos: dado un t cualquiera, decir que $\{\Delta_t, x(t)\}$ sea independiente de $\{\nabla_t\}$. Esta

la arena sí tenemos algún tipo de determinación. Sobre esto volveremos en la sección siguiente. Por esta razón lo hemos resaltado en *cursiva*.

²⁵Esto nos remite a un ejemplo en la física, en la que esta metáfora del reloj de arena encuentra más similitudes. En la teoría de la relatividad tenemos una estructura denominada *cono de luz* cuyo vértice define el *presente* de una partícula masiva que se desplaza sobre el espacio-tiempo.

última afirmación es realmente fuerte, en el sentido en que nos deja un conjunto muy pequeño de procesos que la satisface. Intuitivamente parecer ser que si todo pasado está *desconectado* del presente y a su vez también de todo futuro del sistema, tendríamos una independencia de lo que ocurra en cualquier instante de tiempo con respecto a lo que ocurra en cualquier otro instante de tiempo. Veamos formalmente esta idea a partir del miembro izquierdo de la propiedad de Markov (3.100), que por definición de probabilidad de condicional (3.2), se puede escribir como

$$P(\Delta_t|x(t), \nabla_t) := \frac{P(\Delta_t, x(t), \nabla_t)}{P(x(t), \nabla_t)},$$

incorporando la hipótesis $\{\Delta_t, x(t)\}$ sea independiente de $\{\nabla_t\}$, entonces

$$P(\Delta_t|x(t), \nabla_t) = \frac{P(\Delta_t, x(t))}{P(x(t), \nabla_t)} P(\nabla_t)$$

$$P(\Delta_t|x(t), \nabla_t) = P(\Delta_t|x(t)) \frac{P(x(t))P(\nabla_t)}{P(x(t), \nabla_t)}.$$

Dado que el proceso es de Markov, es decir, se satisface (3.100), luego tenemos

$$P(x(t), \nabla_t) = P(x(t))P(\nabla_t), \quad (3.101)$$

donde pasado y presente terminan siendo independientes, con lo cual tenemos un proceso estocástico trivial. De (3.101) tenemos

$$P(x(t)|\nabla_t) = P(x(t)). \quad (3.102)$$

Toda la historia pasada del sistema está *desconectada* del estado presente.

Concluimos esta cuestión diciendo que exigir a un proceso de Markov que la historia pasada sea independiente del estado presente y la historia futura, nos deja procesos no correlacionados.

Retomando la metáfora, esto se entiende como cortar al reloj horizontalmente y separar las regiones R_t de las R_s , para todo $t \neq s$, de manera que la distribución de arena en una de ellas no esté correlacionada con la cantidad de arena en cualquiera de las otras. En otras palabras *destrozar* el reloj y con él nuestra metáfora.

Digamos además que un proceso de Markov no es aquel cuya propiedad fundamental es la *falta de memoria* (o *memoryless*) sino más bien aquel que requiere una *memoria a corto plazo*. Se trata de conservar el conocimiento actual del sistema para calcular una medida del estado futuro del sistema. Dado el presente y el pasado, los procesos de Markov solo requieren la información sobre el presente.

Cuando estemos interesados en estudiar la evolución temporal sobre un conjunto de estados numerable $E = \{x_i\}_{i \in I}$, dada una historia pasada, la propiedad general de Markov (3.100) toma la siguiente forma *puntual*

$$P(x(t)|x(s), \nabla_s) = P(x(t)|x(s)), \quad (3.103)$$

con $s \leq t$. O bien explicitando los elementos de E :

$$P(x(t) = x_j | x(s) = x_i, \nabla_s) = P(x(t) = x_j | x(s) = x_i), \quad (3.104)$$

con lo cual recuperamos la idea de que en todo proceso de Markov, las probabilidades de transición, presentes en el miembro derecho de (3.103) o (3.104), i.e. $P(x_j, t | x_i, s)$, juegan un rol esencial para describir, junto con la condición inicial, la dinámica de tal proceso.

TRANSICIONES NO NECESARIAMENTE CAUSALES

Hemos señalado que uno de los objetos relevantes en estos procesos estocásticos son las probabilidades de transición, entendidas como probabilidades condicionales. Debemos subrayar que la probabilidad condicional $P(y|x)$ no adjudica a x como causa de

y , siquiera en un sentido débil. Veamos por el absurdo, si decimos que x sea *probablemente* la causa de y , o bien que dado x tengamos *probablemente* un efecto y medido por $P(y|x)$, pues siempre puedo invertirlos vía el teorema de Bayes [24]. Dicho teorema permite escribir $P(x|y)$ en función de $P(y|x)$, el primero asigna a y como causa x , mientras que el segundo asigna a x como causa de y , lo cual no tiene sentido.

Es más, si quisiéramos tomar cierta distancia temporal entre los estados, es decir, considerar variables aleatorias (x_s, y_t) con $s < t$, no podremos tampoco considerar que $P(y_t|x_s)$ (o $P(x_s|y_t)$) consolide causas en ninguno de ellos. Pues siempre cabe el par de preguntas *prospectivas-retrospectivas* que intercambian causas por efectos, invalidando cualquier consideración inicial.

Hemos utilizado un lenguaje capaz de evitar el uso de los términos condicionante y condicionado, salvo en caso de extrema necesidad. Pues, en general, no tiene sentido hablar de condicionante causal ubicado en un tiempo t y condicionado causal ubicado en un tiempo s , con $s < t$. Evitaremos caer en problemas si hablamos en términos de transiciones.

Cada estado es la consecuencia de un conjunto de determinantes (causales y no causales); en otras palabras, los valores de los parámetros que especifican un estado provienen como consecuencias tanto de procesos internos como de vinculaciones externas. No puede haber, por tanto, acciones de un estado sobre otro del sistema. Considerar a los estados como causas es una falacia, puesto que éstos no poseen la virtud productiva presente en la causalidad. En otras palabras no producen nada solo son parte de la historia del sistema [38, 39].

Como conclusión de este capítulo diremos que todo proceso *estocásticamente determinista* puede ser descrito de manera equivalente en sus versiones *duales* forward-backward.

La puesta en escena se da a través de las probabilidades absolutas, que *coordinan* toda la obra y que responden a preguntas absolutas **a.f.** y **a.b.** en términos relativos o propensivos.

FORMULACIÓN VARIACIONAL



4

FORMULACIÓN VARIACIONAL

LUZ, CÁMARA Y ACCIÓN

El título hace referencia a la *terna* que trataremos en este capítulo, a saber: lagrangiano, coordenadas y acción.

Toda intención de ir hacia los fundamentos es hija de un trabajo *detrás de las escenas*. Consideremos esta divertida e ilustrativa metáfora: una obra de teatro. Conforman la obra 4 áreas bien ensambladas, que listaremos como: **A**, **C**, **G**, **L** y **P**, i.e. ordenados alfabéticamente y según las jerarquías de *esencial a fundamental* o de *suficiente a necesario*.¹ Tenemos, por supuesto, a los **artistas** (**A**), e.g. actores, músicos, etc., quienes ceñidos a ciertas reglas internas, tienen la potestad de improvisar y la posibilidad de errar, respecto del libreto. Con el fin de garantizar que la ejecución esté dentro de cierto canon, al cabo de un número de funciones, tenemos a los **coordinadores** (**C**), e.g. productores, directores, etc. Los cuales se ciñen a un **guión** (**G**) o algún tipo de documentación que regule la obra. El guión surge a partir de un **libro** (**L**) original mediante un **proceso de adaptación** (**P**). Por último, el libro original proviene de la inspiración del compositor.²

¹Suelen simbolizarse por **A**, **C**, **G** y **T** a los nucleótidos de una molécula de ADN, pero esto solo es una feliz *casi* coincidencia.

²Para los menos exigentes daremos la correspondencia explícita: los **artistas** a las *variables aleatorias*; los **coordinadores** a las *medidas de probabilidad absolutas* (coordenadas) de esas variables aleatorias; el **guión** a las *ecuaciones de evolución*; el **libro** al *lagrangiano*, o bien a la *acción*, que también surge de la inspiración del compositor. El **proceso de adaptación** (que va del libro al guión) al *principio variacional*. Esta metáfora se aplica con similar eficacia a otras situa-

De esta forma todo comienza con intentar confeccionar el libro y a partir de ahí adaptarlo a un guión, luego vendrán los demás roles de la obra.³

Esta idea de inferir el principio fundamental a partir de la puesta en escena, rememora a la metáfora de Richard Feynman acerca del ajedrez, en su libro *The Character of the Physical Law*. Allí se establece una analogía entre intentar descifrar cuáles son las reglas de un juego y cuáles las leyes de la naturaleza, ambas desconocidas, a partir de la observación del propio juego o de la naturaleza [33]. De manera similar y volviendo a nuestra metáfora, tenemos una situación muy particular, a saber: la obra ya ha comenzado, puesto que tenemos un **guión**, o *ecuaciones de evolución* que refieren a algún proceso estocástico. Solo que no se sabe a partir de qué libro está adaptado. Este es el tema del presente capítulo: confeccionar un **libro** o *lagrangiano*, sabiendo que éste *debe* corresponderse vía un **proceso de adaptación** al **guión** de la obra en curso. A modo de adelanto daremos solo el título de este libro: *Procesos estocásticamente determinados en tiempo continuo*. El cual será desarrollado de manera completa en las siguientes secciones.

Antes de ello, hablaremos del método, i.e. del **proceso de adaptación** que, como hemos dicho, se corresponde con un *principio variacional*. El cual, a su vez, está dentro del área con la que titulamos la siguiente sección.

ciones descritas a partir de principios variacionales.

³Podríamos continuar preguntándonos, pues toda vez que se inicia una metáfora tenemos la tentación de explorar hasta dónde pueden *estirarse* sus límites, dejaremos correr libremente la imaginación del lector para contestar a preguntas tales como quién es el **caramelero**, el **acomodador**, el **colado**, etc.

CÁLCULO VARIACIONAL

En líneas generales, estamos buscando definir un lagrangiano L y a partir de ella otra funcional acción S dada por

$$S = \int_T L dt \quad (4.1)$$

con $T = [a, b] \subseteq \mathbb{R}$ sobre el cual puede tener lugar el proceso estocástico. En general trataremos con lagrangianas que dependerán de la coordenada ξ , su derivada temporal $\dot{\xi}$ y eventualmente el tiempo t . También se puede extender al caso en que ξ puede ser un vector de componentes (ξ_1, ξ_2, \dots) . La funcional acción S dependerá del camino $\xi(t)$, es decir, para cada función $\xi(t)$ tenemos un valor de la acción S sobre ella igual a $S[\xi(t)]$ [1].

Consideremos brevemente algunos elementos del cálculo variacional [1]. Comenzamos considerando la siguiente funcional $F[\xi(t)]$ general. Definimos variación $\delta\xi$ del argumento $\xi(t)$ a la diferencia entre dos funciones $\delta\xi = \xi(t) - \xi'(t)$; aquí una de las dos se considera que varía dentro de cierta clase de funciones.

A partir de $\delta\xi$ se puede definir el concepto de proximidad de funciones de orden k si y solo si las derivadas temporales l -ésimas $\delta\xi^{(l)}$ son pequeñas para $l = 0, \dots, k$. De aquí puede precisarse la definición de continuidad de una funcional[1]. Para el caso de funcionales de la forma de (4.1) serán continuas sólo en casos excepcionales, dado que el integrando contiene a $\dot{\xi}$. En muchos casos es más natural exigir que la proximidad de funciones se defina a partir de que $\delta\xi$ y $\delta\dot{\xi}$ sean pequeñas [1].

La *variación* de una funcional $F[\xi]$, $\delta F[\xi]$, se puede definir a partir de $F[\xi + \gamma \delta\xi] - F[\xi]$

$$\delta F[\xi] = \partial_\gamma F[\xi + \gamma \delta\xi] \Big|_{\gamma=0} \quad (4.2)$$

donde γ es un parámetro y para ξ y $\delta\xi$ fijos.

De la misma forma que podemos estudiar si una función posee extremos en algún punto, a partir de la anulación de su primera derivada en ese punto; en el cálculo variacional tendremos

la noción de función *extremal* de la funcional F , cuyo estudio comprende la búsqueda de aquellas funciones $\xi(t)$ que anulan la variación de F . Esto es, la búsqueda de las curvas $\xi(t)$ tales que

$$\delta F[\xi(t)] = 0. \quad (4.3)$$

Consideraremos funcionales de la forma (4.1)

$$F[\xi(t)] = \int_a^b f(\xi, \dot{\xi}, t) dt. \quad (4.4)$$

Existen unas ecuaciones diferenciales para el integrando f , cuya resolución comprende el conjunto de funciones *extremales* de F . Tales funciones $\xi(t)$ toman ciertos valores fijos sobre la frontera de $T = [a, b]$, i.e. $\xi(a) = \alpha$ y $\xi(b) = \beta$.

Daremos ahora una breve deducción de tales ecuaciones. Sea una familia de curvas parametrizadas por γ como

$$\xi_\gamma = \xi + \gamma \delta\xi, \quad (4.5)$$

tales que $\xi_\gamma(a) = \alpha$ y $\xi_\gamma(b) = \beta$, para cada γ , i.e. todas ellas toman los mismos valores en los extremos del intervalo $[a, b]$. Luego $\delta\xi(a) = 0 = \delta\xi(b)$, en otras palabras $\delta\xi(t)$ se anula en la frontera del intervalo T .

Consideremos una cierta función del parámetro γ , a través de F y para ξ y $\delta\xi$ fijos, como sigue

$$\varphi(\gamma) = F[\xi_\gamma]. \quad (4.6)$$

La variación de la funcional F dada por (4.2), se puede reescribir como una derivada usual de la función $\varphi(\gamma)$, mediante la conexión (4.6), explícitamente como

$$\dot{\varphi}(0) = \partial_\gamma F[\xi_\gamma]|_{\gamma=0}. \quad (4.7)$$

Por esta razón calculamos entonces la derivada primera de $\varphi(\gamma)$

para construir con ella la variación de F vía

$$\begin{aligned}
 \dot{\varphi}(\gamma) &= \int_a^b \partial_\gamma f(\xi_\gamma, \dot{\xi}_\gamma, t) dt, \\
 &= \int_a^b \left[\partial_{\xi_\gamma} f(\xi_\gamma, \dot{\xi}_\gamma, t) \partial_\gamma \xi_\gamma + \partial_{\dot{\xi}_\gamma} f(\xi_\gamma, \dot{\xi}_\gamma, t) \partial_\gamma \dot{\xi}_\gamma \right] dt, \quad (4.8) \\
 &= \int_a^b \left[\partial_{\xi_\gamma} f(\xi_\gamma, \dot{\xi}_\gamma, t) \delta \xi + \partial_{\dot{\xi}_\gamma} f(\xi_\gamma, \dot{\xi}_\gamma, t) \delta \dot{\xi} \right] dt,
 \end{aligned}$$

hemos utilizado que $\partial_\gamma \xi_\gamma = \delta \xi$ y $\partial_\gamma \dot{\xi}_\gamma = \delta \dot{\xi}$, por definición de la familia (4.5). La segunda integral puede expresarse como

$$\begin{aligned}
 \int_a^b \partial_{\dot{\xi}_\gamma} f \delta \dot{\xi} dt &= \int_a^b d_t \left(\partial_{\dot{\xi}_\gamma} f \delta \xi \right) dt - \int_a^b d_t \left(\partial_{\dot{\xi}_\gamma} f \right) \delta \xi dt, \\
 &= - \int_a^b d_t \left(\partial_{\dot{\xi}_\gamma} f \right) \delta \xi dt.
 \end{aligned} \quad (4.9)$$

la primer integral del miembro derecho se anula idénticamente pues $\delta \xi(a) = 0 = \delta \xi(b)$. Hemos simplificado la notación, f está evaluada en $(\xi_\gamma, \dot{\xi}_\gamma, t)$.

Finalmente tenemos que

$$\dot{\varphi}(\gamma) = \int_a^b \left\{ \partial_{\xi_\gamma} f(\xi_\gamma, \dot{\xi}_\gamma, t) - d_t \left[\partial_{\dot{\xi}_\gamma} f(\xi_\gamma, \dot{\xi}_\gamma, t) \right] \right\} \delta \xi dt. \quad (4.10)$$

Según hemos dicho, la variación de F , δF , se expresa como

$$\dot{\varphi}(0) = \int_a^b \left[\partial_\xi f - d_t(\partial_{\dot{\xi}} f) \right] \delta \xi dt \quad (4.11)$$

cuyas extremales se obtienen de anular la integral anterior para todo $\delta \xi$. Nuevamente hemos simplificado la notación, f está evaluada ahora en $(\xi, \dot{\xi}, t)$. Por el *lema fundamental del cálculo variacional* [1] aplicado a $\dot{\varphi}(0) = 0$, según (4.11), tenemos las ecuaciones de Euler-Lagrange para el integrando f

$$0 = \partial_\xi f - d_t(\partial_{\dot{\xi}} f). \quad (4.12)$$

De manera similar, se puede hacer una derivación de las ecuaciones de *Euler-Lagrange* para el caso en que el tratemos con un vector $\xi = (\xi_1, \xi_2, \dots)$ y sus derivadas temporales primeras, concluiríamos con

$$0 = \partial_{\xi_j} f - d_t(\partial_{\dot{\xi}_j} f), \quad (4.13)$$

las cuales se obtienen tomando variaciones sólo de ξ_j y dejándolo a las otras $\{\xi_i\}_{i \neq j}$ invariables, en ese caso la funcional $F[\xi]$ se reduce $F_j[\xi_j]$, i.e. solo depende de ξ_j [1].

Volviendo a nuestra metáfora teatral, el proceso de adaptación, es decir el procedimiento que va del libro al guión, se corresponde con el principio variacional $\delta S = 0$ y el guión que resulte para los coordinadores $\{q_j, p_i : j, i \in I\}$ deben ser las ecuaciones de Kolmogorov para las distribuciones unidimensionales. Estas serán las pautas para la construcción de L . Anteriormente hemos adelantado su título; su forma y su contenido serán tratados en la siguiente sección.

FUNCIONAL LAGRANGIANA

En el capítulo anterior nos hemos centrado precisamente en el objeto intermedio de esta terna: la noción de coordenadas. De la discusión del capítulo anterior hemos visto que todo proceso estocástico es susceptible de ser descrito de manera *dual*, de manera de realizar prospección (predicción) o retrospectión (retrodicción).

Hemos demostrado cómo todos estos conceptos están relacionados a partir de una inversión temporal. Hemos visto además que podemos introducir probabilidades absolutas asociadas a cada descripción, que nos permite considerarlas como coordenadas en el sentido que hemos señalado en la sección de la página 81.

Escribimos una vez más el par forward - backward de Kolmogorov, de la siguiente forma:

$$\dot{q}_j(t) = \sum_{i \in I} \Omega_{ji}(t) q_i(t), \quad (4.14)$$

$$\dot{p}_j(t) = - \sum_{i \in I} p_i(t) \Omega_{ij}(t). \quad (4.15)$$

O bien comprimidas mediante una notación matricial

$$\dot{q} = \Omega q, \quad (4.16)$$

$$\dot{p} = -\Omega^t p. \quad (4.17)$$

Se trata de ecuaciones diferenciales lineales y de primer orden.

Consideremos un producto *interno* entre vectores columna α y β de componentes $\{\alpha_i : i \in I\}$ y $\{\beta_i : i \in I\}$ como

$$\alpha \cdot \beta := \sum_{k \in I} \alpha_k \beta_k \quad (4.18)$$

también nos será de utilidad escribir este producto interno de forma matricial como

$$\alpha \cdot \beta = \alpha^t \beta. \quad (4.19)$$

Cada una de las ecuaciones (4.14) y (4.15) corresponde a un *sistema* de ecuaciones diferenciales acopladas. Pero no *totalmente* acopladas, pues podemos observar que las mismas no resultan acopladas entre sí, en el sentido en que en las prospectivas (4.14) solo aparece q , mientras que en las retrospectivas (4.15) solo aparece p .

Podemos estar *tentados* de construir un lagrangiano L_q que responda a las ecuaciones forward (4.14) y otro L_p que responda a las backward (4.15). Sucumbamos pues a la *tentación de disgregar* el problema en dos partes y centremos la atención en la construcción del supuesto L_q . Notemos que debe depender de (q, \dot{q}) y ser cuadrática en q . Que sea cuadrática en q nos garantiza al menos

una contribución lineal cuando tomemos las derivadas parciales $\partial_{q_j} L_q$ de las ecuaciones de Euler-Lagrange (4.13).

$$L_q = \sum_{k \in I} q_k \dot{q}_k - \sum_{k, l \in I} q_k A_{kl} q_l. \quad (4.20)$$

Notemos que de la forma cuadrática $\sum_{k, l} q_k A_{kl} q_l$ solo sobreviven las contribuciones simétricas de A , independientemente de la estructura de la matriz A . Explicitamos esta afirmación considerando las ecuaciones de Euler-Lagrange (4.13) sobre L_q y para cada q_j

$$\dot{q}_j = \frac{1}{2} \sum_{i \in I} (A_{ji} + A_{ij}) q_i. \quad (4.21)$$

Si quisiéramos asociar esta ecuación a la ecuación forward de Kolmogorov para q_j (4.14), esto equivaldría a considerar una matriz Ω simétrica. En nuestro primer atisbo de funcional lagrangiana L_q , no importa qué tipo de matriz sea A , solo sobrevive su *parte simétrica*⁴.

Por esta razón vemos que L_q no es apropiada, salvo que estemos interesados en describir procesos invariantes bajo una inversión temporal.

El mismo *defecto* ocurre si consideramos lo propio con L_p y otra matriz A'

$$L_p = \sum_{k \in I} p_k \dot{p}_k + \sum_{k, l \in I} p_k A'_{kl} p_l. \quad (4.22)$$

Esto parece indicar que si el proceso estocástico es invariante bajo inversión temporal, el problema se disgrega perfectamente en dos partes.⁵

⁴Toda matriz A puede escribirse como la suma de su parte simétrica, A_σ , y su parte asimétrica A_α : $A = A_\alpha + A_\sigma$, $A_\sigma = \frac{1}{2}(A + A^t)$ y $A_\alpha = \frac{1}{2}(A - A^t)$

⁵Dos partes que estarán más estrechamente relacionadas, si Ω es simétrica las ecuaciones forward y backward son esencialmente las mismas; pues ya hemos dicho que las backward están integradas en el orden temporal inverso a las forward. De ahí el signo – que acompaña a la ecuación backward y la distingue de la forward.

Veamos ahora dónde está el problema de haber sucumbido a aquella *tentación* sobre una construcción separada de L_q y L_p , para identificarlo y eventualmente arrancarlo de raíz. Escribiendo a A como la suma de su parte simétrica A_σ y antisimétrica A_α de manera que la forma cuadrática dentro de ambos lagrangianos es tal que $q.A_\alpha q = 0 = p.A'_\alpha p$. En otras palabras, las funcionales L_q o L_p no pueden distinguir, o no son sensibles a, la parte antisimétrica de A o A' , respectivamente.

Entonces si consideramos separadamente a L como la suma de estas dos funcionales L_q y L_p , el problema estructural de ambas funcionales es heredado por L .

No obstante, la estructura con la que hemos iniciado nuestro libro L , tiene *casi* todo lo que necesitamos. Solo hará falta encontrar una forma para L que contenga la *dualidad* (q, p) , de manera que no sea posible separarlas como suma de dos funcionales L_q y L_p .

Hemos aprendido de lo expuesto anteriormente para construir L , que comenzar por el término cuadrático en q o p , i.e. el término *no cinético* nos ha llevado a buen término. Este procedimiento es el inverso al que se puede hacer en mecánica clásica, donde el término cinético puede deducirse de primeros principios [30]. Esto podría deberse a que allí la noción de sistema libre, i.e. *partícula(s) libre(s)*, es anterior al de sistema interactuante. Sobre este punto podemos agregar que en procesos estocásticos la idea de término *no cinético* o de *interacción* está asociada a todo aquello que promueva la propensión de transiciones, las cuales se encuentran condensadas en Ω o \mathcal{P} . La noción de sistema libre en procesos estocásticos corresponde a la *imposibilidad* (léase probabilidad nula) de transición a otros estados, i.e. $\mathcal{P}_{ji}(t, s) = \delta_{ji}$ que corresponde a la sola existencia de *auto*-transiciones (que implica $\Omega_{ji}(t) = 0$) para todo par temporal (t, s) y todo par de estados (x_j, x_i) de E , pues es la que nos conduce a ecuaciones

$$\dot{q} = 0, \quad \dot{p} = 0. \tag{4.23}$$

Si hubiéramos decidido comenzar por este camino habríamos tenido una gran variedad de formas de construir el término cinético de L que diera lugar a (4.23).

Continuaremos con lo que hemos aprendido, esto es, idear el término *no cinético* de formas bilineales y de allí buscar el término cinético que nos permita arribar a nuestra meta: escribir de una vez por todas el libro L (o S), que *adaptado* por las ecuaciones de Euler-Lagrange para L (o el principio variacional $\delta S = 0$) nos deje nuestro *guión* de ecuaciones de Kolmogorov forward-backward.

Ensayamos nuestra segunda versión del lagrangiano. Insistimos nuevamente en que éste contenga una forma bilineal, del tipo

$$\sum_{k,l} p_k \Omega_{kl} q_l, \quad (4.24)$$

o de manera compacta $p.\Omega q$. La cual no tiene el *defecto* anteriormente señalado ⁶ y de hecho satisface precisamente

$$\partial_p(p.\Omega q) = \Omega q, \quad (4.25)$$

$$\partial_q(p.\Omega q) = \Omega^t p.$$

Ambas tienen el aspecto deseado, pues conforman una parte de las ecuaciones forward y backward, (4.16) y (4.17) respectivamente. Hemos usado una notación compacta para ∂_q de manera que aplicado sobre un campo escalar f , $\partial_q f$ se corresponde a un vector columna de componentes $\{\partial_{q_i} f : i \in I\}$ [26]. Idem con $\partial_p f$ cuyas componentes serán $\{\partial_{p_i} f : i \in I\}$.⁷

⁶Nos referimos a que la forma bilineal $p.\Omega q$ sí es sensible a la parte antisimétrica de la matriz que la conforma: en este caso Ω .

⁷Sin esta nomenclatura compacta la expresión (4.25) se habría escrito

$$\partial_{p_j} \left(\sum_{k,l \in I} p_k \Omega_{kl} q_l \right) = \sum_{l \in I} \Omega_{jl} q_l$$

$$\partial_{q_i} \left(\sum_{k,l \in I} p_k \Omega_{kl} q_l \right) = \sum_{k \in I} p_k \Omega_{ki}$$

A partir de (4.25) podemos observar que cuando derivamos respecto de p , i.e. las coordenadas retrospectivas, tenemos una parte de las ecuaciones prospectivas y viceversa.

Centremos por un momento la atención en reproducir la dinámica forward. Busquemos entonces el término cinético, que derivado respecto de p nos deje a \dot{q} , postulamos

$$p \cdot \dot{q}, \quad (4.26)$$

una vez más lo trivial conlleva a lo fundamental [3].

Uniendo estos ingredientes (4.25) y (4.26), tenemos justificada la invocación, por *inspiración del compositor*, de nuestro lagrangiano

$$L_\mu = \sum_{k \in I} p_k \dot{q}_k - \sum_{k, l \in I} p_k \Omega_{kl} q_l. \quad (4.27)$$

el subíndice μ remite a la inspiración por alguna *musa*.

Pongamos a prueba las bondades de L_μ (4.27) utilizando la notación compacta:

$$L_\mu = p \cdot \dot{q} - p \cdot \Omega q. \quad (4.28)$$

Puede comprobarse *trivialmente* (término utilizado con toda premeditación) que de las ecuaciones de Euler-Lagrange, que surgen al permitir variaciones de la acción solo a través de p , i.e. dejando invariable a q , se deducen las ecuaciones de Kolmogorov prospectivas o forward.

Aunque se trata de un procedimiento trivial veamos como funciona. Dado que L_μ no contiene a \dot{p} las ecuaciones de Euler-Lagrange (4.13) para L_μ , variando p y fijando q , nos dejan

$$\partial_p L_\mu = 0 \implies \dot{q} - \Omega q = 0, \quad (4.29)$$

que corresponden a la ecuación forward de Kolmogorov (4.16).

Calculemos ahora el momento conjugado a cada q_j , por definición es $\partial_{\dot{q}_j} L_\mu$ [30, 31] y según (4.27) es igual a

$$\partial_{\dot{q}_j} L_\mu = p_j. \quad (4.30)$$

Este recorrido demuestra lo que habíamos adelantado antes en la discusión de la página 73, a saber: las coordenadas retrospectivas p_j resultan ser variables conjugadas de las coordenadas prospectivas, q_j , vía L_μ según (4.27).

Además el lagrangiano L_μ tiene la bondad de incluir las ecuaciones backward (4.17), las cuales serán las ecuaciones de Euler-Lagrange para L_μ , tomando ahora variaciones de la acción respecto de q , dejándolo invariable a p . Para hacer aún más simple tal deducción, podemos escribir de manera equivalente⁸ a L_μ de (4.28) como

$$L_\mu \equiv \dot{q} \cdot p - q \cdot \Omega^t p. \quad (4.31)$$

Las ecuaciones de Euler-Lagrange para L_μ , variando q y fijando p , nos dejan

$$d_t(\partial_{\dot{q}} L_\mu) - \partial_q L_\mu = 0 \implies \dot{p} + \Omega^t p = 0.$$

Hemos completado nuestro propósito en cuanto a definir un lagrangiano tal que sus ecuaciones de Euler - Lagrange correspondan a las ecuaciones forward - backward.

Continuando con el recorrido canónico, podemos definir un hamiltoniano a partir de una *transformación de Legendre* sobre L_μ [30, 31, 35]. Considerando a las coordenadas $\{q_j\}_{j \in E}$ y sus momentos $\{p_j\}_{j \in E}$ y el lagrangiano L_μ propuesto según (4.27) tenemos

$$\begin{aligned} H_\mu &:= \sum_{i \in I} p_i \dot{q}_i - L_\mu \\ H_\mu &= \sum_{i, j \in I} p_j \Omega_{ji} q_i \\ H_\mu &= p \cdot \Omega(t) q, \end{aligned} \quad (4.32)$$

donde hemos comprimido la expresión usando la notación matricial. El hamiltoniano H_μ depende de (q, p) y eventualmente del tiempo de manera explícita a través de $\Omega(t)$.

⁸Recordemos que $q \cdot \Omega^t p \equiv p \cdot \Omega q$.

Podemos reescribir las ecuaciones forward - backward de Kolmogorov (4.16) y (4.17) en forma hamiltoniana. Para ello usaremos la forma de H_μ (4.32) y teniendo en cuenta (4.25)

$$\dot{q} = \partial_p H_\mu \quad (4.33)$$

$$\dot{p} = -\partial_q H_\mu. \quad (4.34)$$

A pesar de que el lagrangiano L_μ permita deducir, vía las ecuaciones de Euler - Lagrange, tanto las forward como las backward, nótese que L_μ según (4.27) o (4.28) no trata en pie de igualdad a q y a p , en el sentido en que L_μ contiene a \dot{q} pero no a \dot{p} . Necesitamos recurrir una vez más a nuestra musa, en otras palabras requerimos pues un *extra de musa*.

El defecto de nuestra primera inspiración, L_μ , radica en que su término cinético $p \cdot \dot{q}$ solo responde correctamente a la pregunta *prospectiva* (asociada a la ecuación *forward*), a saber: ese objeto es tal que derivado respecto de p nos deje a \dot{q} . En resumen, cuando construimos L_μ nos hemos enfocado solo en el aspecto prospectivo, omitiendo el retrospectivo. Aunque debemos observar que la forma bilineal (4.25) responde perfectamente bien a ambos requerimientos. Por lo tanto, reconstruiremos ahora el término cinético propio de la pregunta retrospectiva. Invertimos los términos de la pregunta, es decir, formulamos la pregunta retrospectiva (asociada a la ecuación *backward*): buscamos ahora el término cinético que derivado respecto de q nos deje a $-\dot{p}$. De nuevo ensayamos una trivialidad

$$-q \cdot \dot{p}. \quad (4.35)$$

Eso nos conduce a un lagrangiano complementario al anterior, i.e. a L_ν , que escribimos directamente de forma compacta como

$$L_\nu = -q \cdot \dot{p} - q \cdot \Omega^t p, \quad (4.36)$$

el subíndice ν remite a la *nueva* (ν -eva) intervención de nuestra musa (μ -sa). Por otro lado, L_ν posee las mismas bondades que L_μ

en el sentido en que pueden deducirse las ecuaciones forward y backward. Para L_ν resulta ahora elemental la deducción de las ecuaciones backward a partir de variar la acción respecto de q .

Calculemos ahora el momento conjugado a cada p_j , por definición es $\partial_{\dot{p}_j} L_\nu$ [30, 31] y según (4.36) es igual a

$$\partial_{\dot{p}_j} L_\nu = -q_j. \quad (4.37)$$

A partir de una transformada de Legendre sobre L_ν definimos el hamiltoniano H_ν

$$H_\nu := - \sum_{i \in I} q_i \dot{p}_i - L_\nu \quad (4.38)$$

$$H_\nu = p \cdot \Omega(t) q,$$

que es idénticamente el mismo que el que se obtenía a partir de L_μ (4.32). En este caso las coordenadas son $\{p_j\}_{j \in E}$ y sus momentos conjugados $\{-q_j\}_{j \in E}$ y el lagrangiano L_ν propuesto según (4.36)

Ambas formulaciones lagrangianas L_μ y L_ν están relacionadas a través de la operación $q \longleftrightarrow p$, que intercambia ecuaciones prospectivas con retrospectivas, corresponde a una inversión temporal, tal y como hemos probado en 3.8. Por lo tanto eso induce la siguiente transformación sobre el operador diferencial $d_t \longleftrightarrow -d_t$ y $\Omega \longleftrightarrow \Omega^t$, que corresponden a una inversión temporal asociada al intercambio de q por p .

Ahora bien, hemos entendido que los lagrangianos L_μ y L_ν satisfacen perfectamente los requerimientos en cuanto a que de ellos se pueden deducir, vía las ecuaciones de Euler - Lagrange, la dinámica forward y backward de Kolmogorov. Cada una posee (*coordenadas, momentos*) escritas de manera ordenada: para L_μ son (q, p) mientras que para L_ν son $(p, -q)$.

Aún así ambos lagrangianos L_μ y L_ν según (4.28) y (4.36) no tratan en pie de igualdad a q y a p , en el sentido en que L_μ contiene a \dot{q} pero no a \dot{p} y L_ν contiene a \dot{p} pero no a \dot{q} .

Surge entonces la siguiente pregunta: ¿cómo unificar ambas descripciones dadas por L_μ y L_ν ? Una solución pragmática sería considerar una combinación lineal que trate en pie de igualdad ambas L_μ y L_ν

$$L = \frac{1}{2}(L_\mu + L_\nu). \quad (4.39)$$

la *linealidad* de la combinación (4.39) se comprende naturalmente a fin de *conservar el contexto*, i.e. las ecuaciones de Euler - Lagrange, que son lineales en f (4.13). Es decir se trata de una suerte de promedio, semisuma o combinación convexa uniforme entre L_μ y L_ν y que resumimos con el lema: la *conjunción* de L_μ y su *complemento* L_ν conforman la *totalidad* L .

Otra forma de obtener el mismo resultado proviene del hecho de que dos lagrangianos que difieran entre sí de una derivada total de una función, proveen la mismas ecuaciones de Euler - Lagrange [30, 31].

Explícitamente construimos L a partir de la suma L_ν y la derivada total $\frac{1}{2}d_t(p.q)$

$$L_\nu + \frac{1}{2}d_t(p.q), \quad (4.40)$$

o también podríamos haberlo hecho a partir de la suma de L_μ y la derivada total $-\frac{1}{2}d_t(p.q)$

$$L_\mu - \frac{1}{2}d_t(p.q), \quad (4.41)$$

ambos casos (4.40) y (4.41) nos dejan un lagrangiano que es idénticamente igual a L según (4.39).

Aunque el procedimiento de sumar o restar una derivada total a cierto lagrangiano, no modifica las ecuaciones de evolución, sí que cambia la forma del mismo. No obstante, más adelante veremos que el producto $p.q$ es un invariante de la teoría. Nos referimos, por supuesto, al producto $p(t).q(t)$ que se conserva idéntico para todo instante de tiempo. Con lo cual, el procedimiento de agregar o sustraer $d_t(p.q)$ a L_ν o L_μ , respectivamente, es inocuo o *neutro*, no solo por dejar invariantes las ecuaciones de Euler - Lagrange, sino porque $d_t(p.q)$ es en sí mismo *el elemento neutro bajo la suma* para todo t .

Escribamos explícitamente nuestro lagrangiano final L

$$L = \frac{1}{2}(p.\dot{q} - q.\dot{p}) - p.\Omega q \quad (4.42)$$

Ahora calculemos los momentos conjugados asociados a las coordenadas (q, p) , que resultaran ser por definición [30, 31]

$$(\partial_{\dot{q}}L, \partial_{\dot{p}}L) = \frac{1}{2}(p, -q) \quad (4.43)$$

tal y como ocurriera antes de manera separada para L_μ y L_ν . Puesto que, desde el principio, podríamos haber considerado un lagrangiano $L_\mu = \frac{1}{2}(p.\dot{q} - p.\Omega q)$ y $L_\nu = -\frac{1}{2}(q.\dot{p} + q.\Omega^t p)$, en ese caso $L = L_\mu + L_\nu$, que sería idénticamente igual a (4.39), solo cambia la forma original de L_μ y L_ν . En cualquier caso, dado L y un número real a , el producto de ambos nos deja otro lagrangiano, aL , igualmente válido, en tanto que ambos comparten las mismas ecuaciones de Euler - Lagrange.

La transformación de Legendre de L define el mismo hamiltoniano H que (4.32) y (4.38)

$$H = p.\Omega q. \quad (4.44)$$

Puesto que H es el mismo, las ecuaciones de Hamilton que hemos mostrado en (4.33) y (4.34) permanecen idénticas

$$\dot{q} = \partial_p H \quad (4.45)$$

$$\dot{p} = -\partial_q H. \quad (4.46)$$

Un comentario más antes de terminar la sección, acerca de la conservación de la normalización en el tiempo de las coordenadas (q_j, p_j) . No hará falta agregar condicional adicional en L , a modo de *multiplicador de Lagrange* [30, 31], que fije de manera externa la normalización de las coordenadas (q_j, p_j) , en el sentido en que la suma de todas ellas sea igual 1. Esta condición resulta implícita en la propiedad exhaustiva sobre \mathcal{P} , en (3.30) del capítulo

anterior, que reescribimos

$$\sum_j \mathcal{P}_{ji}(t, s) = 1, \quad (4.47)$$

Veamos como recobrar la conservación de la normalización a partir de (4.47) y junto con $q_j(t) = \sum_i \mathcal{P}_{ji}(t, s) q_i(s)$ tenemos

$$\begin{aligned} \sum_j q_j(t) &= \sum_{j \in I} \sum_{i \in I} \mathcal{P}_{ji}(t, s) q_i(s) \\ &= \sum_{i \in I} \sum_{j \in I} \mathcal{P}_{ji}(t, s) q_i(s) \\ \sum_j q_j(t) &= \sum_{i \in I} q_i(s). \end{aligned}$$

Concluimos que la normalización de q se conserva en el tiempo.

También puede deducirse esta situación a partir de Ω , que aparece explícitamente en el lagrangiano L . La matriz Ω hereda la propiedad exhaustiva de \mathcal{P} , que toma la forma $\sum_j \Omega_{ji} = 0$. A partir de las ecuaciones forward para cada q_j (4.14), sumemos entonces sobre j a todas ellas

$$\begin{aligned} \sum_{j \in I} \dot{q}_j(t) &= \sum_{j \in I} \sum_{i \in I} \Omega_{ji}(t) q_i(t), \\ &= \sum_{i \in I} \sum_{j \in I} \Omega_{ji}(t) q_i(t), \\ \sum_{j \in I} \dot{q}_j(t) &= 0, \end{aligned}$$

que es otra forma de afirmar que la normalización se conserva en el tiempo.

De manera semejante ocurre para p , cuando tengamos una propiedad exhaustiva, similar a (4.47), para \mathcal{P}^* . Solo basta pedir normalización por ejemplo para la condición inicial o final para garantizar su conservación en todo momento. Esto nos permite deducir que dichas sumas se conservan en el tiempo. Por lo que si

la distribución inicial o final está normalizada a 1, la distribución prospectiva y retrospectiva también lo estará para cualquier otro instante de tiempo.

Para culminar esta sección diremos que siempre podemos considerar un lagrangiano L y una constante real, de cuya multiplicación tendríamos otro lagrangiano igualmente válido, en tanto que ambos comparten las mismas ecuaciones de Euler - Lagrange. Esta arbitrariedad nos deja cierto tipo de indeterminación. Ahora bien, esto tiene un límite en cuanto consideramos la aditividad de dos lagrangianos [30], en este caso por ejemplo hemos construido L a partir de la suma de L_μ y L_ν . De esta forma, solo nos quedará por fijar una constante que puede cambiar las unidades o el signo global de L . Respecto de las unidades, es una *suerte* quizá que S no posea unidades. Esto nos deja tranquilos, puesto que estamos en un área de la matemática que se aplica, i.e. forma parte, de distintas ramas de las ciencias como la física, la química, la biología, la economía (aún a riesgo de error), etc. En otras palabras, tratamos con medidas de probabilidad que no poseen unidades de ninguna de estas ciencias. De esta forma solo nos queda analizar la cuestión del signo global de L .

ESPACIO DE PROBABILIDADES ABSOLUTAS

En esta sección daremos algunas precisiones acerca del *lugar matemático* donde residen las probabilidades absolutas q y p .

Para el caso en que E sea de dimensión finita n , las probabilidades absolutas (q_1, \dots, q_n) *viven* en una región del hiperplano de dimensión $n - 1$ formada por $\sum_{j=1}^n q_j = 1$, con tal que $q_j \geq 0$. Lo mismo ocurre con las probabilidades absolutas (p_1, \dots, p_n) . Se trata pues de un espacio plano, que ha sido tratado extensamente, denominado *simplex*, en este caso *probability simplex*, definido

como [40]

$$\Sigma = \left\{ (\vartheta_1, \dots, \vartheta_n) : \sum_{j=1}^n \vartheta_j = 1, \vartheta_j \geq 0 \forall j \right\}. \quad (4.48)$$

La condición $\sum_{j=1}^n \vartheta_j = 1$ expresa que el vector columna ϑ de coordenadas $(\vartheta_1, \dots, \vartheta_n)$ es norma-1. En ese sentido diremos que ϑ como medida de probabilidad está *correctamente normalizada* o que es una medida *decente* de probabilidad. Podemos compactar más la notación para el caso en que ϑ sea norma-1. Consideramos para ello el vector columna u cuyos elementos sean todos iguales a 1; podemos reescribir entonces que ϑ es norma-1 como

$$u \cdot \vartheta = 1. \quad (4.49)$$

Denotemos también el caso de vectores no negativos de manera más compacta: un vector x de componentes (x_1, \dots, x_n) es no negativo sii $x_j \geq 0$ para todo $j = 1, \dots, n$ o de manera comprimida $x \succeq 0$ [40].

De esta forma reescribimos el *simplex* Σ como

$$\Sigma = \left\{ (\vartheta_1, \dots, \vartheta_n) : u \cdot \vartheta = 1, \vartheta \succeq 0 \right\}, \quad (4.50)$$

de forma que las medidas q y p *cohabitan* en el *simplex* Σ (4.50).

En la sección siguiente haremos una transición de este mundo de la *simplexidad* al mundo de la *simplecticidad*, esto es, veremos que es posible dar una formulación de estas ideas en un espacio simpléctico denotado por (V, ω) , donde $\omega : V \times V \rightarrow \mathbb{R}$ será la métrica de ese espacio: una forma bilineal, no degenerada y antisimétrica [35, 36] y $V = \Sigma \times \Sigma$.

Quién será la matriz canónica asociada a la forma bilineal ω y cómo quedarán reformuladas nuestras ideas bajo esta concepción, son preguntas que conciernen a la siguiente sección.

FORMULACIÓN SIMPLÉCTICA

También podríamos haber construido L directamente a partir de otro camino. Sabemos que el par de ecuaciones de Kolmogorov forward - backward para q y p , resultan ser complementarias y describen un proceso estocásticamente determinista. Siguiendo esta noción podríamos haber considerado un único un vector ξ que tenga todas las coordenadas $\{q_j, p_i\}_{j,i \in I}$ dispuestas en forma de columna. Podemos fusionar ambas ecuaciones forward y backward, escritas en la forma (4.45) y (4.46) y que reescribimos

$$\dot{q} = \partial_p H, \quad \dot{p} = -\partial_q H,$$

en un único sistema de ecuaciones para $\xi := (q, p)^t$ de la forma

$$\dot{\xi} = v(\xi, t) \tag{4.51}$$

donde $v(\xi, t)$ es un campo vectorial lineal en ξ . Veremos que, en general, el sistema de ecuaciones (4.51) es *no autónomo* [34]. El campo vectorial v puede escribirse de manera *caprichosa* como

$$v(\xi, t) = J\mathcal{U}(t)\xi, \tag{4.52}$$

donde las matrices J y $\mathcal{U}(t)$ vienen dadas por

$$J := \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}, \quad \mathcal{U}(t) := \begin{pmatrix} 0 & \Omega^t(t) \\ \Omega(t) & 0 \end{pmatrix} \tag{4.53}$$

con I matriz identidad. La matriz J satisface que $J^t = -J$ y además es ortogonal $J^t = J^{-1}$.

Buscamos un lagrangiano de la forma $\xi \cdot \alpha \dot{\xi} + \xi \cdot \beta \xi$. Necesariamente β no puede ser antisimétrica, de lo contrario $\xi \cdot \beta \xi = 0$. Mejor dicho la parte antisimétrica de β no sobrevive en L . Por otro lado, observando la forma completa de L según (4.42) requerimos que la matriz α debe ser antisimétrica.

Realizando comparaciones de $\xi.\alpha\dot{\xi} + \xi.\beta\xi$ con L según (4.42) de la sección anterior, tenemos que α y β definen el siguiente lagrangiano, ahora dependiente de $\{\xi, \dot{\xi}\}$, también dependerá eventual y explícitamente de t .⁹

$$L = -\frac{1}{2} \left[\xi.J\dot{\xi} + \xi.U(t)\xi \right] \quad (4.54)$$

que es idénticamente igual al lagrangiano L dado por (4.42). También podemos expresar equivalentemente a L como

$$L = \frac{1}{2} \left[\dot{\xi}.J\xi - \xi.U(t)\xi \right] \quad (4.55)$$

pues $J^t = -J$, luego $-\xi.J\dot{\xi} = -\dot{\xi}.J^t\xi = \dot{\xi}.J\xi$.

En este caso, sí es lícito incluir en L el término no cinético a partir una forma cuadrática $\xi.U(t)\xi$, pues $U(t)$ no es antisimétrica; de hecho, ésta es simétrica por definición (4.53). Mientras que el término cinético de L es de la forma $\xi.J\dot{\xi}$.

A partir de esta última expresión (4.55) para el lagrangiano L se puede obtener el momento conjugado asociado a ξ , que denotamos por p_ξ y es definimos $p_\xi := \partial_{\dot{\xi}}L$ [30, 31, 35, 36], toma la forma

$$p_\xi = \frac{1}{2}J\xi, \quad (4.56)$$

que está de acuerdo con (4.43).

El hamiltoniano, definido vía una transformación de Legendre del lagrangiano L , es $H = p_\xi.\dot{\xi} - L$, toma la forma

$$H = p_\xi.JU\xi, \quad (4.57)$$

es idénticamente igual al hamiltoniano $H = p.\Omega q$ según (4.44) de la sección anterior.¹⁰

⁹La dependencia de L con t proviene de la eventual dependencia del generador infinitesimal de las traslaciones Ω con el parámetro t .

¹⁰Recordemos que H es, a su vez, igual a H_μ (4.32) o bien a H_ν (4.38).

Las ecuaciones de Hamilton, obtenidas para (q, p) en (4.33) y (4.34), se pueden condensar según la forma usual de una formulación hamiltoniana de una dinámica [31, 36], las cuales escribimos de manera compacta y en componentes como

$$\dot{\xi} = J\partial_{\xi}H \quad (4.58)$$

$$\dot{\xi}_j = \sum_{i \in I} J_{ji} \partial_{\xi_i} H. \quad (4.59)$$

La matriz antisimétrica J (4.53) define lo que se denomina una forma bilineal *canónica*, antisimétrica y no degenerada, de un espacio lineal simpléctico. Estábamos esperando esta estructura, ya que estos espacios conforman el escenario de una formulación lagrangiana de una dinámica [26, 31, 35, 36].

En este sentido podemos hablar de una interpretación geométrica haciendo uso del concepto de *espacio físico* o *espacio de fases* [30], que corresponde al espacio formado por todas las coordenadas generalizadas y sus momentos conjugados. El espacio de fases corresponde al espacio simpléctico de vectores de la forma $\xi = (q, p)^t$.

El término *simpléctico* fue introducido por primera vez por el matemático Hermann Weyl; cuyo significado, del griego, trae la noción de aquello *entrelazado* [37]. Weyl utilizó ese término para distinguirlo y evitar confusiones con el término *complejo*, que tiene un significado similar, aunque de origen latino.¹¹

De hecho los espacios vectoriales simplécticos son isomorfos a los espacios vectoriales complejos. Esto puede verse a partir de una operación denominada *decomplexificación* que permite escribir espacios simplécticos a partir de espacios complejos y otra denominada *complexificación* para el camino inverso [34]. Esta camino de doble sentido de circulación permite establecer incluso

¹¹El término *simpléctico* es de hecho una manera de escribir el término “complejo”, este último proviene del latín *complexus*, mientras que “simpléctico” viene del correspondiente término griego *symplektikos*.

una equivalencia entre sistemas cuánticos sobre espacios complejos, como los que hemos tratado en la página 62, y sistemas clásicos en espacios simplécticos [41].

En el contexto de este trabajo, el término simpléctico merece cierto elogio a su etimología, que hace referencia a aquello *entrelazado*. Requerimos de un espacio simpléctico para formular correctamente estas ideas. En la sección anterior hemos visto que el intento de formalizar la descripción de la dinámica de los procesos estocásticamente determinados, a partir de dos lagrangianos L_q y L_p que forma *separada* hagan su papel, es en general infructuoso. Requerimos, por tanto, de un único lagrangiano que conjugue ambas coordenadas y velocidades y eso nos deja *necesariamente* en un espacio vectorial simpléctico. En resumen, para una formulación general y consistente, parecería que no podemos separar en general aquello que el formalismo propuesto ha *entrelazado*. Tal y como se dijo desde el título de la sección de la página 33: *prospección* y *retrospección* corresponden a dos aspectos, generalmente complementarios, del mismo proceso. Ahora agregamos, que dicho proceso no puede en general describirse como dos subprocesos regulados por sendos lagrangianos L_q y L_p .

De la misma forma que una *moneda* no puede *desmembrarse* en dos *partes* de manera que cada una de ellas contenga solo una *cara*. Esto es debido a que el concepto de *cara* refiere al de *moneda*, tal como *prospección* y *retrospección* atañen al de *proceso estocástico*.

Esto precisamente lo que venimos afirmando sobre la descripción dual para procesos estocásticos. En este caso, el espacio vectorial simpléctico V corresponde al producto cartesiano de dos simplexes ordenados $\Sigma \times \Sigma$, en el sentido en que uno refiere a q y el otro a p . En las ecuaciones de Hamilton podemos observar esta idea de entrelazamiento, o bien, enlazamiento; para no hacer alusión a ciertos fenómenos característicos de la mecánica cuántica. Puesto que la derivada total de una de las variables, q o p , es igual a la derivada parcial del hamiltoniano respecto de la otra

variable canónica, p o q , que reescribimos

$$\dot{q} = \partial_p H, \quad \dot{p} = -\partial_q H.$$

Esta noción de enlazamiento también subyace en el espacio en el que viven las medidas (q, p) . Desde un punto de vista constructivo, ya hemos dicho y redicho que el espacio simpléctico V que estamos tratando se compone, entre otras cosas, por el producto cartesiano $\Sigma \times \Sigma$, donde Σ es un *simplex* introducido en (4.50) de la página 111. Cuando se consideran conjuntamente a las coordenadas $\{\xi_j\}$ y la matriz J , que define una forma bilineal antisimétrica, queda conformado un espacio lineal simpléctico, tal y como ocurre en la formulación de la mecánica clásica [35]. Allí *conviven* las probabilidades absolutas prospectivas y retrospectivas (q, p) .

La noción de coordenadas remite a cierta geometría. Se podrían extender estas ideas, para tratarse sobre el escenario de *variedades simplécticas* [35, 36]. Incluso podría tratarse el problema sobre una variedad de dimensión n para las coordenadas, a partir de otra de dimensión $n+1$ que incluya al tiempo. De manera que una *foliación* de tal variedad nos deje subvariedades de dimensión n , etiquetadas con t [36], en las cuales *viven* las medidas de probabilidad absoluta.

Sobre este punto cabe agregar un comentario acerca de una extensión a variedades diferenciales: una estructura más rica que nuestro elemental *simplex*. Para ello recordemos que, esencialmente, la condición que define el lugar en el que *viven* las medidas de probabilidad *honestas* [24] es

$$\sum_{j=1}^n \vartheta_j = 1. \tag{4.60}$$

dicha condición (4.60) junto con que cada una de las ϑ_j sean no negativas, definen a nuestro *simplex* Σ . Ahora bien, cabe la posibilidad de debilitar la primer condición (corromper, sería más

apropiado en este caso) y llevarla a

$$\sum_{j=1}^n \vartheta_j < 1. \quad (4.61)$$

Toda medida de probabilidad *corrupta* o *deshonesta* es, naturalmente, expulsada a vivir fuera del simplex Σ , son aquellas medidas que no satisfacen (4.60) sino (4.61). Denotamos a esa región hostil por ϑ definida como

$$\vartheta = \left\{ (\vartheta_1, \dots, \vartheta_n) : \sum_{j=1}^n \vartheta_j < 1, \vartheta_j \geq 0 \forall j \right\}, \quad (4.62)$$

la cual tiene por espacio límite al *simplex* Σ .

Una medida de probabilidad que satisfaga (4.61) se consideraba antiguamente mal definida, impropia o indecente. Ahora bien, es posible eliminar toda perniciosa adjetivación sobre tales medidas, pues realmente no tiene nada de malo vivir allí. En primer lugar la región ϑ encierra una riqueza mayor en tanto variedad diferencial. Actualmente, la expresión (4.61) tiene una interpretación concreta [24], la clave está en que estas medidas de probabilidad evolucionan temporalmente. El miembro izquierdo de (4.61) representa la probabilidad de que solo pueden tener lugar un número finito de transiciones a un tiempo finito t [24].

SOBRE CORCHETES E INVARIANCIAS TEMPORALES

Al igual que ocurre en la formulación de hamiltoniana de la mecánica clásica y cuántica, H corresponde al generador infinitesimal de las traslaciones temporales [30, 31]. Para procesos estocásticos, este papel es representado por Ω . Veamos como están conectados estos dos elementos. Para procesos estocásticos temporalmente homogéneos o estacionarios, i.e. aquellos en los que existe invariancia bajo traslaciones temporales [26]

$$\mathcal{P}(t, s) = \mathcal{P}(t', s'), \quad t - s = t' - s' \quad (4.63)$$

con $t \geq s$ y $t' \geq s'$. Luego de (4.63), tomando $s' = 0$, tenemos que $\mathcal{P}(t, s) = \mathcal{P}(t - s, 0)$, i.e. depende de la diferencia de tiempos $t - s$; entonces Ω será independiente del tiempo. En ese caso, el hamiltoniano $H = p.\Omega q$ de (4.32), o (4.38), se conserva en el tiempo. Podemos probar esto calculando explícitamente la derivada total

$$\dot{H} = \partial_t H + \sum_i \partial_{q_i} H \dot{q}_i + \partial_{p_i} H \dot{p}_i, \quad (4.64)$$

insertando allí las ecuaciones (4.33) y (4.34) tenemos $\dot{H} \equiv \partial_t H$ y para este caso en particular en que Ω no depende del tiempo ocurre que $\partial_t H = 0$, luego $\dot{H} = 0$.

Notemos que aún en el caso en que H (o bien Ω) sea invariante en el tiempo esto no implica aún invariancia bajo la inversión temporal, tal y como la hemos presentado. Esto último se consigue cuando Ω es simétrica. Existen diferencias matemáticas y semánticas (e.g. las que provienen de alguna interpretación física), entre la operación de traslación temporal y la de inversión temporal. Desde un punto de vista formal si el sistema es invariante frente a la primera transformación no implica necesariamente que lo sea también frente a la segunda y viceversa.

A partir de (4.64) podemos ver que para cualquier función f que dependa de (q, p) , y eventualmente de t , podemos expresar su variación temporal total como

$$\dot{f} = \partial_t f + \sum_{i \in I} \partial_{q_i} f \dot{q}_i + \partial_{p_i} f \dot{p}_i,$$

reemplazando aquí las ecuaciones forward - backward (4.33) y (4.34)

$$\dot{f} = \partial_t f + \sum_{i \in I} \partial_{q_i} f \partial_{p_i} H - \partial_{p_i} f \partial_{q_i} H, \quad (4.65)$$

que permite definir de manera natural una operación denominada *corchete de Poisson*, sobre dos funciones f y g de (q, p) y even-

tualmente de t [26, 31, 35, 36]

$$[f, g] := \sum_{i \in I} \partial_{q_i} f \partial_{p_i} g - \partial_{p_i} f \partial_{q_i} g. \quad (4.66)$$

De forma compacta podemos redefinir el corchete de Poisson como

$$[f, g] := \partial_q f \cdot \partial_p g - \partial_p f \cdot \partial_q g. \quad (4.67)$$

De esta forma la variación temporal de f se expresa de manera compacta

$$\dot{f} = \partial_t f + [f, H]. \quad (4.68)$$

A partir de ésta última expresión, podemos probar algo que habíamos comentado en la sección anterior: el producto $p(t) \cdot q(t)$ es un invariante de la teoría. Definimos $f(p, q) = p \cdot q$. En primer lugar, notamos que f no depende explícitamente de t , i.e. $\partial_t f = 0$, con lo cual de (4.68) tendremos $d_t f = [f, H]$. Verifiquemos entonces la nulidad del corchete de Poisson $[f, H]$ para todo instante t , usando su expresión compacta (4.67) junto con $p \cdot \Omega q = q \cdot \Omega^t p$.

$$\begin{aligned} [p \cdot q, H] &= \partial_q(p \cdot q) \cdot \partial_p H - \partial_p(p \cdot q) \cdot \partial_q H, \\ &= p \cdot \partial_p H - q \cdot \partial_q H, \\ &= p \cdot \Omega q - q \cdot \Omega^t p \\ [p \cdot q, H] &= 0, \end{aligned}$$

que es lo que queríamos probar: $p \cdot q$ es un invariante de la teoría, es decir, se conserva en el tiempo.

Podemos llegar a la misma conclusión desde la propia teoría de procesos estocásticos. En otras palabras, el producto $p(t) \cdot q(t)$ es igual a $p(s) \cdot q(s)$. Sin pérdida de generalidad, tomamos $t \geq s$, de forma que $q(s)$ evoluciona a $q(t)$ y que $p(t)$ involuciona a $p(s)$

a través de $\mathcal{P}(t, s)$ como sigue

$$q_j(t) = \sum_{i \in I} \mathcal{P}_{ji}(t, s) q_i(s), \quad (4.69)$$

$$p_i(s) = \sum_{j \in I} p_j(t) \mathcal{P}_{ji}(t, s). \quad (4.70)$$

esto lo sabemos de (3.94) y (3.95) del capítulo anterior. De manera que $p(t).q(t)$ puede escribirse como

$$\begin{aligned} \sum_{j \in I} p_j(t) q_j(t) &= \sum_{j \in I} p_j(t) \left[\sum_{i \in I} \mathcal{P}_{ji}(t, s) q_i(s) \right], \\ &= \sum_{i \in I} \left[\sum_{j \in I} p_j(t) \mathcal{P}_{ji}(t, s) \right] q_i(s), \quad (4.71) \\ \sum_{j \in I} p_j(t) q_j(t) &= \sum_{i \in I} p_i(s) q_i(s). \end{aligned}$$

de forma que obtenemos que $p(t).q(t) = p(s).q(s)$. Para el caso en que E sea numerable e infinito, podemos igualmente hacer uso de esta prueba puesto que el producto matricial resulta ser asociativo para matrices no negativas, que es el caso de \mathcal{P} , p y q [22].

Hablemos ahora de otra cantidad conservada, la norma de la distribución absoluta. Comencemos, por ejemplo, con la *forward* q . Ya hemos señalado que no hará falta incluir un multiplicador de Lagrange que fuerce esta condición, pues está implícita en la propiedad de exhaustividad sobre \mathcal{P} o bien la que hereda Ω . Consideremos una breve demostración de que utilizando las ideas de esta sección. Usaremos la notación compacta para la normalización de q

$$u.q = 1 \quad (4.72)$$

donde u es un vector columna cuyas componentes sean todas iguales a 1. Consideramos la función $g(q) = u.q$ como no depende

explícitamente del tiempo, i.e. $\partial_t g = 0$, calculemos el corchete de Poisson

$$\begin{aligned} [u.q, H] &= \partial_q(u.q) \cdot \partial_p H \\ &= u \cdot \Omega q \\ &= u^t \Omega q \end{aligned}$$

a partir de la propiedad de exhaustividad heredada sobre Ω como $u^t \Omega = 0$, con lo cual $[u.q, H] = 0$, la normalización de q se conserva en el tiempo.

También podemos escribir el corchete de Poisson (4.67) en función de ξ y J como la forma bilineal

$$[f, g] := \partial_\xi f \cdot J \partial_\xi g \tag{4.73}$$

Hemos presentado brevemente esta importante operación binaria, que no solo permite verificar si una cantidad es temporalmente invariante sino también permite construir nuevos invariantes a partir de otros dos. Sean f y g dos funciones de (q, p) y eventualmente t , tales que $d_f = 0 = d_t g$, entonces ocurre que $d_t[f, g] = 0$ [30, 31].

Los corchetes de Poisson constituyen una operación fundamental en la formulación geométrica de la mecánica en física [30, 31] y en la geometría diferencial en matemáticas [36], pues son parte esencial de la evolución temporal de tales sistemas físicos y en general de sistemas dinámicos. Permiten distinguir cierta clase de transformaciones denominadas *canónicas*. Sobre este tema volveremos en el capítulo siguiente.

PRINCIPIO DE ACCIÓN ESTACIONARIA

Llegamos ahora al principio variacional para procesos estocásticos que estábamos buscando. Hemos visto que todo proceso estocásticamente determinista, según Kolmogorov [13], está caracterizado por una función L que contiene a las medidas absolutas de probabilidad *prospectivas* (q) y *retrospectivas* (p), que

coordinan todo proceso en el sentido que hemos señalado en el capítulo 3.¹² A partir de L construimos una funcional acción S como la integral temporal de L en sobre un intervalo real T

$$S = \int_T L dt. \quad (4.74)$$

La evolución temporal de la propensión de tal proceso estocástico está gobernada por las ecuaciones de Kolmogorov forward y backward para (q, p) , las cuales se pueden obtener equivalentemente a partir de

$$\delta S = 0. \quad (4.75)$$

específicamente, tomando variaciones de S respecto de p y dejando invariable a q obtenemos las ecuaciones de evolución forward para q . Mientras que tomando variaciones de S respecto de q y dejando invariable a p obtenemos las ecuaciones de evolución backward para p . O bien a partir de $\delta_p S = 0$ obtendremos las ecuaciones forward o propectivas para q y a partir de $\delta_q S = 0$ obtendremos las ecuaciones backward o retrospectivas para p .

Hemos dicho que el principio variacional enunciado de manera resumida como $\delta S = 0$ tiene implícito el carácter estacionario de S . Veamos explícitamente tal conexión, para ello podemos expresar la acción de manera compacta en las coordenadas del espacio de fases $\xi = (q, p)^t$. Usando la notación del capítulo 4.2, evaluamos la acción S en una cierta función del parámetro γ , para ξ y $\delta\xi$ fijos, como sigue

$$\varphi(\gamma) = S[\xi_\gamma]. \quad (4.76)$$

La variación de la funcional S , i.e. $\delta S[\xi]$, se puede escribir como una derivada usual de la función $\varphi(\gamma)$, mediante la conexión (4.76), explícitamente como

$$\dot{\varphi}(0) = \partial_\gamma S[\xi_\gamma]|_{\gamma=0}. \quad (4.77)$$

¹²Eventualmente L contiene también a \dot{q} y \dot{p} .

Desarrollamos $\varphi(\gamma)$ en potencias de γ y alrededor de $\gamma = 0$ como:

$$\varphi(\gamma) = \varphi(0) + \dot{\varphi}(0)\gamma + R(\gamma), \quad (4.78)$$

donde $R(\gamma)$ representa el resto de los términos del desarrollo, i.e. los de orden γ^n con $n \geq 2$. Luego por (4.76) tenemos

$$S[\xi_\gamma] = S[\xi] + \delta S[\xi]\gamma + R(\gamma). \quad (4.79)$$

En el caso en que cierta curva $\xi(t)$ satisfaga $\delta S[\xi] = 0$, ξ será una *extremal* de S . Entonces la acción evaluada en curvas cercanas a la curva extremal, $S[\xi_\gamma]$, se mantiene prácticamente idéntica a $S[\xi]$, es decir, *estacionaria*.

Podemos enunciar el principio de acción estacionaria como

Dado un proceso estocástico kolmogoroviano, la evolución temporal de su propensión es tal que la integral S se mantiene estacionaria.

En otras palabras, de entre todas aquellas medidas de probabilidad, que toman idénticos valores en los extremos del intervalo T , un principio variacional de estas características permite distinguir las *correctas*, i.e. aquellas que responden a la dinámica *kolmogoroviana*.

A este respecto, agregamos un comentario acerca del título de esta tesis sobre la acción *mínima*. Este adjetivo es más bien histórico, puesto que, en la formulación lagrangiana de la mecánica clásica podemos encontrar gran variedad de funcionales acción que han de ser mínimas para aquellas trayectorias que realmente formarán parte de la historia del sistema. Pero lo cierto es que para deducir las ecuaciones de Euler-Lagrange que rigen tal dinámica, solo basta exigir la anulación de la primera variación de, en este caso, S : $\delta S = 0$, es decir exigir que la acción sea *estacionaria*. Esto representa una condición necesaria de mínimo, pero no lo garantiza [30, 31, 35].

En nuestro caso, las soluciones del problema variacional anulan idénticamente al lagrangiano y por consiguiente a la acción.

Esto puede verse directamente si reemplazamos las ecuaciones de evolución temporal forward $\dot{q} = \Omega q$ y backward $\dot{p} = -\Omega^t p$ en $L = \frac{1}{2}(p.\dot{q} - q.\dot{p}) - p.\Omega q$, obtendremos $L = 0$. O equivalentemente tomar las soluciones del problema variacional compacto $\dot{\xi} = J\mathcal{U}(t)\xi$ y reemplazarlas en $L = -\frac{1}{2}(\xi.J\dot{\xi} + \xi.\mathcal{U}\xi)$, junto con $J^2 = -I$, implica $L = 0$, i.e. S evaluada sobre las soluciones del problema variacional es idénticamente nula.

Un estudio completo que involucre condiciones suficientes de la existencia de mínimos, máximos o incluso puntos de silla de S , implica investigar el signo de la variación de segundo orden de S [1].

INTEGRAL ACCIÓN COMO UN PRODUCTO INTERNO

Comenzamos esta sección con la notación matricial de las ecuaciones forward y backward para las medidas absolutas de probabilidades.

$$\dot{q} = \Omega q, \tag{4.80}$$

$$\dot{p} = -\Omega^t p. \tag{4.81}$$

Se trata de ecuaciones diferenciales lineales y de primer orden.

Definimos dos operadores *diferenciales matriciales forward* \mathcal{D}_f y *backward* \mathcal{D}_b

$$\mathcal{D}_f := d_t - \Omega(t), \tag{4.82}$$

$$\mathcal{D}_b := -d_t - \Omega^t(t), \tag{4.83}$$

donde d_t corresponde a tomar la primera derivada respecto de t y tiene implícita una matriz identidad.

De esta forma, las ecuaciones forward (4.80) y backward (4.81) pueden escribirse como la aplicación de \mathcal{D}_f y \mathcal{D}_b sobre vectores

columna de manera compacta

$$\mathcal{D}_{\mathbf{f}}(q(t)) = 0 \quad (4.84)$$

$$\mathcal{D}_{\mathbf{b}}(p(t)) = 0, \quad (4.85)$$

Recordamos la notación para el producto interno entre vectores columna u, v como

$$u \cdot v := \sum_{k \in I} u_k v_k, \quad (4.86)$$

donde $\{(u_k, v_k) : k \in I\}$ son sus respectivas componentes.

Para el caso en que tratemos con funciones sobre el intervalo temporal T a valores vectoriales, podemos considerar el siguiente producto interno

$$(u, v) := \int_T u(t) \cdot v(t) dt \quad (4.87)$$

en ese caso las componentes de cada vector u y v serán funciones de T sobre un cuerpo, e.g. \mathbb{R} o \mathbb{R}^+ . La notación $[u, v]$ no hace mención al tiempo t , pues a pesar de que ambos vectores u, v sean funciones del tiempo, el resultado de la integral es un número real fijo.

El producto escalar (u, v) satisface las siguientes propiedades, con $\alpha, \beta \in \mathbb{R}$:

- (1) *lineal* $(u, \alpha v + \beta v') = \alpha(u, v) + \beta(u, v')$,
- (2) *simétrico* $(u, v) = (v, u)$,
- (3) *no negativo* $(u, u) \geq 0$, en particular $(u, u) = 0 \Leftrightarrow u = 0$.

De (1) y (2) tenemos *linealidad* sobre el primer argumento:

$$(\alpha u + \beta u', v) = \alpha(u, v) + \beta(u', v).$$

Veamos ahora el producto

$$\begin{aligned} (u, \mathcal{D}_f(v)) &= \int_T u \cdot d_t v - u \cdot \Omega v \, dt \\ &= \int_T -d_t u \cdot v - (\Omega^t u) \cdot v \, dt + \phi \cdot \varphi|_{\partial T} \end{aligned} \quad (4.88)$$

$$(u, \mathcal{D}_f(v)) = (\mathcal{D}_b(u), v) + u \cdot v|_{\partial T}$$

donde ∂T corresponde a la frontera de T y $u \cdot v|_{\partial T}$ representa un término de borde, e.g. si $T = [a, b]$ entonces $\partial T = \{a, b\}$ y $u \cdot v|_{\partial T} = u(b) \cdot v(b) - u(a) \cdot v(a)$.

De esta forma podemos escribir al lagrangiano $L_\mu = p \cdot \dot{q} - p \cdot \Omega q$ como

$$L_\mu = p \cdot \mathcal{D}_f(q) \quad (4.89)$$

De manera análoga tenemos podemos escribir a $L_\nu = -q \cdot \dot{p} - q \cdot \Omega^t p$

$$L_\nu = q \cdot \mathcal{D}_b(p) \quad (4.90)$$

Este producto producto interno (ϕ, φ) (4.87) junto con los operadores diferenciales \mathcal{D}_f (4.82) y \mathcal{D}_b (4.83) permiten escribir el lagrangiano $L = \frac{1}{2}(L_\mu + L_\nu)$ y con ella, la acción S . Recordamos que S depende de la curva $\xi(t)$ o bien de $q(t)$ y $p(t)$

$$S = \frac{1}{2}(p, \mathcal{D}_f(q)) + \frac{1}{2}(\mathcal{D}_b(p), q). \quad (4.91)$$

Concluimos entonces que, a menos de una constante real, el operador adjunto de \mathcal{D}_f , denotado por \mathcal{D}_f^\dagger , corresponde a \mathcal{D}_b . Es más, para el caso de procesos estocásticos, el término de borde es idénticamente nulo: $p \cdot q|_{\partial T} = 0$. En este caso los operadores diferenciales forward y backward resultan ser *mutuamente adjuntos* de manera exacta: $\mathcal{D}_f^\dagger = \mathcal{D}_b$ y $\mathcal{D}_b^\dagger = \mathcal{D}_f$.

Por otro lado, sea A un operador, en caso que exista operador adjunto, denotado por A^\dagger , se cumplirá entonces que $\llbracket u, A(v) \rrbracket = \llbracket A^\dagger(u), v \rrbracket$ [25]. De forma que todo esto puede ser escrito usando un único operador diferencial, digamos \mathcal{D}_f y su adjunto.

$$S = \frac{1}{2}(p, \mathcal{D}_f(q)) + \frac{1}{2}(\mathcal{D}_f^\dagger(p), q). \quad (4.92)$$

aunque ya hemos señalado que $(p, \mathcal{D}_f(q)) - (\mathcal{D}_f^\dagger(p), q)$ es igual a un *término de borde*, que para nuestro caso es nulo. Por lo que la acción también puede escribirse como

$$S = (p, \mathcal{D}_f(q)). \quad (4.93)$$

Este es otro aspecto de la *dualidad forward - backward* de las medidas absolutas q y p .

Esta forma de definir escribir a L_μ y L_ν a partir de (4.89) y (4.90), i.e. haciendo uso de los operadores diferenciales de las ecuaciones forward y backward aplicados sobre q y p y contraídos escalarmente con p y q , respectivamente, nos será de utilidad para la siguiente sección.

COMENTARIOS ACERCA DE LA NOTACIÓN

Insistimos nuevamente con este punto, pues creemos que una notación puede ser una verdadera ayuda o realmente un problema que induzca a errores lógicos o conceptuales. Este trabajo conjuga ideas de distintas disciplinas, cada una de las cuales posee una notación nominal. Al momento de introducir un concepto de una de ellas, hemos conservado su notación y eventualmente hemos descuidado al resto o parte de este. No queremos que esto lleve a problemas mayores, de ahí el motivo de esta sección.

Hemos introducido la noción de coordenadas a partir de las medidas absolutas de probabilidad. Puesto que dado un proceso estocástico, contamos con una descripción dual prospectiva-retrospectiva, esto nos ha llevado a considerar a (q, p) como el conjunto de coordenadas que describen, respectivamente, estos dos aspectos de un proceso estocásticos. Ahora bien la notación propuesta, para tales cantidades, fue construida siguiendo el siguiente esquema: comenzando por el dispositivo prospectivo, de-

signamos por q a las probabilidades absolutas, de manera de asociarnos a la notación usualmente utilizada en mecánica clásica [30, 31]. Pero se tenía pensado que las retrospectivas serían designadas por p , no solamente porque p y q poseen un orden alfabético *antes-después*, sino porque realmente ocurre que p resulta ser la variable conjugada de q vía nuestro lagrangiano.

Sin embargo, en física las q 's suelen representar a las coordenadas generalizadas, dentro del espacio de configuración. De manera que la formulación lagrangiana se desarrolla en el conjunto en el que viven (q, \dot{q}) . Por otro lado, se definen las coordenadas conjugadas p 's, vía el lagrangiano L , como $p = \partial_{\dot{q}}L$. Finalmente la formulación hamiltoniana se desarrolla sobre el conjunto en el que viven (q, p) , denominado espacio de fases. En resumen tenemos que el lagrangiano L depende de (q, \dot{q}) y que el hamiltoniano H depende de (q, p) . A su vez L y H pueden depender explícitamente de t .

Ahora bien puede parecer que se mezclan dos conceptos el de espacio de configuración y el de espacio de fases, en otras palabras puede causar cierta *molestia* que nuestro lagrangiano contenga a su vez a q y a p .

Podemos resolver este aparente problema realizando un cambio de notación que nos permitirá tratar con los objetos de la teoría sin incurrir en errores conceptuales.

Para las medidas forward mantenemos la notación inicialmente propuesta $\{q_j\}_{j \in I}$ que cambian temporalmente a través de

$$q_j(t) = \sum_{j \in I} \mathcal{P}_{ji}(t, s) q_i(s). \quad (4.94)$$

Proponemos un cambio de notación para las medidas backward con q^* , tal como hemos usado para las probabilidades condicionales $\mathcal{P}^*(s, t)$, de manera que

$$q_i^*(s) = \sum_{j \in I} \mathcal{P}_{ij}^*(s, t) q_j^*(t). \quad (4.95)$$

De nuevo, al igual que para la medida condicional, el uso del \star no indica conjugación compleja \star .

Hemos dicho ya que ambas descripciones prospectiva y retrospectiva son independientes en el sentido en que responden a preguntas distintas de un procesos estocástico. Ahora bien, matemáticamente esta idea puede desarrollarse como sigue, debemos garantizar que (q, q^\star) sean *buenas* coordenadas del problema, esto requiere que ambas describan el problema que se plantea y que sean independientes. Dado que nuestro problema es la descripción de la dinámica *forward* y *backward*, entonces este primer punto queda satisfecho por q y q^\star , definidas sus componentes por (4.94) y (4.95), respectivamente. Para el segundo punto veamos dos argumentos, en primer lugar (q, q^\star) son linealmente independientes: dados $\alpha, \beta, \in \mathbb{R}$ tales que $\alpha q + \beta q^\star = 0$ implica $\alpha = 0 = \beta$. Probemos esto como sigue, tomamos el producto escalar de la combinación lineal anterior con q , i.e. $\alpha q \cdot q + \beta q \cdot q^\star = 0$, derivando respecto de t tenemos $\alpha q \cdot \dot{q} = 0$, pues $d_t(q \cdot q^\star) = 0$. Finalmente tendremos $\alpha q \cdot \Omega q = 0$ y puesto que la forma cuadrática no es idénticamente nula, nos queda $\alpha = 0$. Repitiendo el procedimiento tomando el producto escalar con q^\star , tendremos $\beta = 0$.

Intentemos ahora buscar alguna relación entre (q, q^\star) . Tal relación debe ser lineal, digamos de la forma $q^\star = \Lambda q$, pues se tiene que preservar la forma de las ecuaciones prospectivas y retrospectivas. En otras palabras, lo que buscamos es una transformación dada por una matriz real y no singular Λ , tal que Ω y $-\Omega^t$ resulten ser similares vía Λ , i.e. buscamos que $-\Omega^t = \Lambda \Omega \Lambda^{-1}$. Esto implica que $tr(\Omega) = 0$, que no es cierta para todas las matrices con las que estamos tratando, pues $\Omega_{ii} \leq 0$ [21]. En tal caso, si ocurre que $tr(\Omega) = 0$ implica $\Omega_{ii} = 0 \forall i \in I$. Por otro lado, de la propiedad de exhaustividad sobre los elementos de $\{\mathcal{P}_{ji} : j, i \in I\}$, tenemos que los elementos de $\{\Omega_{ji} : j, i \in I\}$ satisfacen, por herencia, una propiedad similar (3.39), que reescribimos

$$\sum_{j \in I} \Omega_{ji} = 0. \quad (4.96)$$

Puesto que los elementos de Ω fuera de la diagonal principal son positivos [21], $\Omega_{ji} \geq 0, \forall i, j \in I$ tales que $i \neq j$, concluimos que para este tipo de matrices Ω pedir que su traza sea nula implica que tales matrices son idénticamente nulas. En tal caso las medidas q y q^* son fijas, i.e. estacionarias, independientes del tiempo, por tanto, es trivial hallar una relación lineal entre ambas. En otras palabras, tal conexión entre q existe solo en los casos estacionarios de manera elemental y para cualquier Ω no nula tal relación no tiene lugar. Esto prueba que en general (q, q^*) corresponden a descripciones matemáticamente independientes.

Definimos nuestro espacio de configuración por el conjunto en el que viven las coordenadas prospectivas y retrodictivas (q, q^*) , que denotaremos por q . En tal caso sería conveniente revisar el procedimiento de obtención de L .

El estudio y posterior confección de L_μ y L_ν se mantiene igual, pero reescribamos ambos dos usando la nueva notación de coordenadas (q, q^*) del *espacio de configuración*.

$$L_\mu = q^* \cdot \mathcal{D}_f(q) \tag{4.97}$$

$$L_\nu = q \cdot \mathcal{D}_b(q^*).$$

Ambos lagrangianos L_μ y L_ν dependen de (q, q^*) , además L_μ depende de \dot{q} y L_ν depende de \dot{q}^* . Tenemos una relación entre L_μ y L_ν , como sigue

$$L_\nu = L_\mu - d_t(q^* \cdot q). \tag{4.98}$$

La búsqueda de igualdad nos ha llevado a construir, en el capítulo anterior, un lagrangiano que dependa de todas las coordenadas $q=(q, q^*)$ y sus derivadas $\dot{q}=(\dot{q}, \dot{q}^*)$. Proponemos un lagrangiano a partir de una combinación lineal general

$$L = aL_\mu + bL_\nu, \tag{4.99}$$

junto con la exigencia de que las ecuaciones de Euler -Lagrange para q y q^* sean precisamente las *forward* y *backward* de Kolmo-

gorov, i.e. $\mathcal{D}_f(q) = 0$ y $\mathcal{D}_b(q^*) = 0$, implica $a = b$.¹³ En esta oportunidad tomaremos¹⁴ $a = 1 = b$ y escribamos explícitamente a L en función de las coordenadas q y sus velocidades \dot{q}

$$L = \dot{q} \cdot Jq - q \cdot \mathcal{U}q, \quad (4.100)$$

donde J y \mathcal{U} vienen dadas por

$$J := \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}, \quad \mathcal{U} := \begin{pmatrix} 0 & \Omega^t \\ \Omega & 0 \end{pmatrix}.$$

El lagrangiano L (4.100) queda definido como función de (q, \dot{q}) y eventualmente de t a través de Ω . Las ecuaciones de Euler-Lagrange toman la forma

$$d_t(\partial_{\dot{q}}L) - \partial_q L = 0. \quad (4.101)$$

Calculemos ahora la variable conjugada a $q = (q, q^*)$, denotada por $p = (p, p^*)$ y definida a través de L como

$$p = \partial_{\dot{q}}L, \quad (4.102)$$

de donde se obtiene $p = (q^*, -q)$, que expresamos sintéticamente como $p = Jq$.

Para la formulación hamiltoniana tendremos

$$H := p \cdot \dot{q} - L. \quad (4.103)$$

donde debemos reemplazar en L toda eventual dependencia de \dot{q} como función de p . De manera que H quede definida como función de (q, p) y eventualmente de t a través de Ω , como

$$H = p \cdot J\mathcal{U}q. \quad (4.104)$$

las ecuaciones de Hamilton pueden escribirse como

$$\dot{q} = \partial_p H, \quad \dot{p} = -\partial_q H. \quad (4.105)$$

¹³En lugar de (4.99) podría haberse postulado algo más general de la forma $L = \alpha(q^* \cdot \dot{q}) - \beta(\dot{q}^* \cdot q) - \gamma(q^* \cdot \Omega q)$, en tal caso habríamos concluido que $\alpha + \beta = \gamma$, lo que nos lleva a (4.99).

¹⁴En el capítulo anterior hemos utilizado $a = \frac{1}{2} = b$, lo cual implicó que las variables conjugadas de (q, q^*) , vía nuestro antiguo L , fueran $\frac{1}{2}(q^*, -q)$.

EXTENSIÓN A ESPACIOS NO NUMERABLES

Queremos presentar una extensión de la noción de procesos estocásticos al caso no numerable. Estudiaremos el caso en que la variable aleatoria tome valores sobre un espacio E no numerable, e.g. algún intervalo de \mathbb{R} o bien todo el conjunto \mathbb{R} . En tal caso hablaremos de variables aleatorias *continuas*, en relación a a su conjunto imagen ahora no numerable.

Podemos trabajar directamente con dos densidades de probabilidad como cantidades fundamentales, asociadas a las ecuaciones forward y backward. En este caso iremos directamente a la forma de las ecuaciones. Aprovecharemos las ideas de las secciones precedentes para dar forma al lagrangiano y posteriormente a la acción pertinente.

Dado un proceso estocástico cuya variable aleatoria esté definida sobre un espacio E ahora no numerable, podemos conjurar un análisis similar al anterior, que permitiera hablar de medidas de probabilidad absoluta prospectivas y retrospectivas como coordenadas que lo describen. Simplemente tendremos algunos cambios de notación, puesto que en lugar de contemplar sumas que recorran los estados de E , tendremos ahora integrales.

Debemos notar que las ecuaciones prospectivas y retrospectivas en variable continua que desarrolló Kolmogorov se aplica para densidades de probabilidad condicional [13]. También veremos que estas ecuaciones resultarán ser mutuamente adjuntas

Con estas ideas en mente, consideremos nuevamente los trabajos de Kolmogorov. Precisamente en [13] dedujo lo que denominamos como ecuaciones prospectivas y retrospectivas o *ecuación fundamental forward* y *ecuación fundamental backward*, para las densidades condicionales. Se trata de dos ecuaciones diferenciales, ahora no ordinarias, e.g. a derivadas parciales.

La primera de ellas tiene la forma

$$\partial_t u(y, t) = -\partial_y [f_1(y, t)u(y, t)] + \partial_y^2 [f_2(y, t)u(y, t)], \quad (4.106)$$

la cual se denomina ecuación de *Fokker-Planck*, en honor a Adriaan Fokker [43] y Max Planck [44], quiénes trabajaron en ella, antes de que Kolmogorov la redescubriera de manera independiente.

La segunda de ellas tiene la forma

$$\partial_s v(x, s) = -f_1(x, s)\partial_x v(x, s) - f_2(x, s)\partial_x^2 v(x, s). \quad (4.107)$$

La densidad condicional $P(y, t|x, s)$ satisface (4.106) para (x, s) fijo y satisface (4.107) para (y, t) fijo.

En la sección anterior hemos degustado las bondades de introducir un producto interno para el caso numerable (4.87), entre otras fueron la de permitirnos escribir a la acción S de una forma compacta y conveniente. Consideremos entonces el producto interno, ahora para campos escalares u y v definidos sobre la región $E \times T$ como

$$\llbracket u, v \rrbracket = \int_{E \times T} u(x, t)v(x, t) dx dt, \quad (4.108)$$

que satisface las propiedades de *linealidad* (1), *simetría* (2) y *no negatividad* (3), de la sección anterior.

Con este aparato matemático podremos describir de manera simple una clase más amplia de ecuaciones diferenciales que las de Kolmogorov (4.106) y (4.107).

Consideremos un operador diferencial lineal, \mathfrak{D} sobre un campo escalar $u(x, t)$, como

$$\mathfrak{D}(u) := \partial_t u + \sum_{k=0}^n f_k(x, t)\partial_x^k u. \quad (4.109)$$

El operador diferencial adjunto a \mathfrak{D} , i.e. \mathfrak{D}^\dagger , respecto del producto (4.108), definido sobre un campo escalar $v(x, t)$ es [42]

$$\mathfrak{D}^\dagger(v) := -\partial_t v + \sum_{k=0}^n (-1)^k \partial_x^k [f_k(x, t)v]. \quad (4.110)$$

Al igual que en la sección anterior habíamos definido un operador diferencial *forward* \mathcal{D}_f y otro *backward* \mathcal{D}_b asociados a las

ecuaciones homónimas. Aquí podemos hacer lo propio: definir dos operadores \mathfrak{D}_f y \mathfrak{D}_b como

$$\mathfrak{D}_b(v) := \partial_t u + f_1(x, t) \partial_x u + f_2(x, t) \partial_x^2 u. \quad (4.111)$$

$$\mathfrak{D}_f(u) := -\partial_t u - \partial_x [f_1(x, t) u] + \partial_x^2 [f_2(x, t) u],$$

Los operadores \mathfrak{D}_b y \mathfrak{D}_f pueden obtenerse como casos particulares de \mathfrak{D} y \mathfrak{D}^\dagger , respectivamente. Ambos operadores \mathfrak{D}_b y \mathfrak{D}_f se obtiene tomando \mathfrak{D} de en (4.109) y \mathfrak{D}^\dagger de (4.110) bajo las condiciones $f_0 \equiv 0$ y $f_k \equiv 0$ para $k \geq 3$.

Nótese que hemos comenzado con el operador \mathfrak{D} (4.109) que resultó estar asociado a la ecuación backward (4.107), se ha procedido de manera inversa respecto de la que hemos mostrado en la sección anterior para el caso numerable. Podríamos haber optado por comenzar con la descripción forward y de ahí, mediante su operador adjunto inferir el operador backward tal y como hicimos en la sección anterior. Aunque en este contexto resulta más natural comenzar con la descripción backward, en cuanto la forma del operador diferencial asociado. De hecho la ecuación backward requiere menos condiciones para la existencia y unicidad de la solución [13].

Recordemos que para el caso numerable teníamos un vector de probabilidad absoluta para cada tiempo t , que hacían las veces de coordenadas, cuyas componentes $\{q_i(t), p_i(t)\}_{i \in I}$ referían al valor x_i de la variable aleatoria que denotamos por $x(t)$ o bien x_t . El proceso estocástico *discreto* estaba definido a partir de la familia de funciones medibles $\{x_t\}_{t \in T}$ sobre cierto espacio de estados numerable. En el caso continuo tendremos algo similar, pero el conjunto de funciones medibles, denotadas por $\{\xi_t\}_{t \in T}$, tomarán valores en cierto espacio E ahora no numerable.

En tal caso tendremos *densidades de probabilidad* que denominaremos, por extensión conceptual, forward y backward $\varrho(x, t)$ y $\varrho^*(x, t)$, respectivamente. Notemos que x representa un valor del conjunto de estados E que puede tomar $\{\xi_t\}$, i.e. la variable

aleatoria ahora continua. Matemáticamente $\varrho(x, t)$ y $\varrho^*(x, t)$ corresponden a campos escalares que toma valores en el intervalo real $[0, 1]$ y están definidos sobre una región de \mathbb{R}^2 , $E \times T$, que denominamos *estado-tiempo*.

De manera similar a como hemos desarrollado en el capítulo anterior, podemos definir densidades prospectivas y retrospectivas, a partir de un procedimiento de marginalización sobre la densidad conjunta de probabilidad $P(y, t, x, s)$. Notemos que ambos estados $\{x, y\}$ corresponden a valores de $\{\xi_s, \xi_t\}$, respectivamente.

Comenzamos con el aparato prospectivo, marginalizando sobre la variable aleatoria ξ_s

$$P(y, t) = \int_E P(y, t, x, s) dx, \quad (4.112)$$

reescribimos el integrando como $P(y, t, x, s) = P(y, t|x, s)P(x, s)$.

Repetiremos entonces el procedimiento realizado en el capítulo anterior, i.e. consideramos al miembro izquierdo de (4.112) como la densidad absoluta, en adelante denotada por $\varrho(y, t)$

$$\varrho(y, t) = \int_E P(y, t|x, s)\varrho(x, s) dx, \quad (4.113)$$

recordemos que la densidad condicional $P(y, t|x, s)$ está definida para $s \leq t$, con lo cual (4.113) corresponde a una *prospección* temporal.

De forma similar podemos marginalizar sobre la variable aleatoria ξ_t , lo que implica entonces una nueva densidad condicional, a la que distinguiremos con el superíndice \star .¹⁵

$$\varrho^*(x, s) = \int_E P^*(x, s|y, t)\varrho^*(y, t) dy, \quad (4.114)$$

recordemos que $P^*(x, s|y, t)$ está definida para $s \leq t$, con lo cual (4.114) corresponde a una *retrospección* temporal.

¹⁵Insistimos en que esta notación con superíndice \star no corresponde a la conjugación compleja, la cual hemos distinguido por $*$ a lo largo de todo el trabajo.

Las densidades condicionales P y P^* corresponden a preguntas de la forma **c.f.** y **c.b.**

Las densidades (4.113) y (4.114) cumplen el mismo rol que para el caso discreto, en el sentido en que corresponden a preguntas de la forma **a.f.** y **a.b.**, también podremos calcular su función generadora de momentos. De esta obtener las cantidades que podrían observarse directa o indirectamente en el experimento como por ejemplo, el valor esperado, la desviación estándar y en general, los momentos de todo orden [24].

Suponiendo que las densidades P y P^* satisfacen sendas ecuaciones prospectivas y retrospectivas generalizadas, que informalmente expresamos como $\mathfrak{D}^\dagger(P) = 0$ y $\mathfrak{D}(P^*) = 0$, lo mismo ocurrirá entonces con ϱ y ϱ^* , respectivamente. La idea general sobre la que basamos tal afirmación hace uso de la estructura de las expresiones (4.113) y (4.114).

Comencemos con ϱ según (4.113). En primer lugar veamos que (4.113) está bien definida, esto se obtiene a partir de que el integrando es una densidad de *probabilidad*, luego es *integrable*. En segundo lugar, probemos que ϱ según (4.113) es solución de la ecuación prospectiva generalizada, i.e. $\mathfrak{D}^\dagger(\varrho) = 0$. Para probarlo, denotemos al operador diferencial lineal \mathfrak{D}^\dagger escrito en coordenadas $\langle y, t \rangle$ como: $\mathfrak{D}^\dagger_{\langle y, t \rangle}$. Por hipótesis $\mathfrak{D}^\dagger_{\langle y, t \rangle}(P(y, t|x, s)) = 0$. Con lo cual $\mathfrak{D}^\dagger_{\langle y, t \rangle}$ aplicado al integrando de (4.113) es trivialmente continuo en todo $E \times E$, de hecho es idénticamente nulo. Con lo cual es posible intercambiar las operaciones de integración y diferenciación de forma que $\mathfrak{D}^\dagger_{\langle y, t \rangle}(\varrho(y, t)) = 0$. Una conclusión similar ocurre con $\varrho^*(x, s)$: $\mathfrak{D}_{\langle x, s \rangle}(\varrho^*(x, s)) = 0$.

Considerando ambos campos ϱ, ϱ^* escritos en las mismas coordenadas (x, t) , podemos continuar con el esquema presentado en las secciones precedentes y construir un lagrangiano. En este caso, trataremos con una *densidad lagrangiana* \mathcal{L}_μ , similar a L_μ (4.89), definida como

$$\mathcal{L}_\mu = \varrho^* \mathfrak{D}_f(\varrho), \quad (4.115)$$

notemos que \mathcal{L}_μ no depende explícitamente de $\{\partial_t \varrho^*, \partial_x^k \varrho^*\}_{k \geq 1}$, de manera similar a lo que ocurría con L_μ . Al igual que hemos hecho en las secciones precedentes, podemos complementar a \mathcal{L}_μ con otra *densidad lagrangiana* \mathcal{L}_ν , similar a a L_ν de (4.90), definida como

$$\mathcal{L}_\nu = \varrho \mathfrak{D}_b(\varrho^*), \quad (4.116)$$

notemos que \mathcal{L}_ν no depende explícitamente de $\{\partial_t \varrho, \partial_x^k \varrho\}_{k \geq 1}$, de manera similar a lo que ocurría con L_ν .

La densidad lagrangiana del sistema será entonces

$$\mathcal{L} = \frac{1}{2}(\mathcal{L}_\mu + \mathcal{L}_\nu) \quad (4.117)$$

Por último, la acción del sistema será entonces

$$S = \int_{E \times T} \mathcal{L} \, dx \, dt, \quad (4.118)$$

y puede escribirse como

$$S = \frac{1}{2} \llbracket \varrho^*, \mathfrak{D}_f(\varrho) \rrbracket + \frac{1}{2} \llbracket \mathfrak{D}_b(\varrho^*), \varrho \rrbracket \quad (4.119)$$

usando el hecho de que los operadores diferenciales forward \mathfrak{D}_f y backward \mathfrak{D}_b resultan ser *mutuamente adjuntos*, tenemos

$$S = \llbracket \varrho^*, \mathfrak{D}_f(\varrho) \rrbracket, \quad (4.120)$$

la cual tiene la misma forma que (4.93). También podemos escribir la ecuación (4.120) como

$$S = \llbracket \mathfrak{D}_b(\varrho^*), \varrho \rrbracket. \quad (4.121)$$

Veamos como funciona, a partir de un caso emblemático por 3 razones: forma parte del pasado (condición necesaria), ha constituido un *annus mirabilis* y fue protagonizado por A. Einstein, quién en 1905, utilizó una de las ecuaciones de Kolmogorov, la prospectiva, para dar cuenta de la función densidad correspondiente al movimiento browniano [78]

$$\partial_t \varrho(x, t) = D \partial_x^2 \varrho(x, t). \quad (4.122)$$

Escribamos también la ecuación retrospectiva asociada, para completar el par de ecuaciones de Kolmogorov

$$-\partial_t \varrho^*(x, t) = D \partial_x^2 \varrho^*(x, t). \quad (4.123)$$

Este par de ecuaciones pueden obtenerse a partir de la siguiente densidad lagrangiana

$$\mathcal{L} = \frac{1}{2}(\varrho^* \partial_t \varrho - \varrho \partial_t \varrho^* - D \varrho^* \partial_x^2 \varrho - D \varrho \partial_x^2 \varrho^*). \quad (4.124)$$

puesto que las ecuaciones (4.122) y (4.123) están incluidas en la familia de operadores diferenciales \mathfrak{D} y \mathfrak{D}^\dagger de (4.109) y (4.110), para el caso en que todos los f_k son idénticamente nulos, excepto el caso $k = 2$.

El formalismo presentado, permite incluir con facilidad algunos casos cuyos operadores diferenciales sean algo más generales que los que hemos discutido en (4.111) y ejemplificado con el movimiento browniano en (4.122) y (4.123). Nos referimos a los operadores (4.109) y (4.110). Este caso corresponde a una expansión de Kramers-Moyal [45, 46] de manera que la ecuación forward de Kolmogorov (o ecuación de Fokker-Planck) y su ecuación *dual* backward puedan generalizarse. Tal situación también está incluida en este formalismo lagrangiano, dada la generalidad del operador \mathfrak{D} o también de su adjunto \mathfrak{D}^\dagger . Simplemente se trata de reemplazar el operador \mathfrak{D}_b por \mathfrak{D} y \mathfrak{D}_f por \mathfrak{D}^\dagger y construir una densidad lagrangiana más general \mathcal{L} , que adquiere la forma

$$\mathcal{L} = \frac{1}{2} \varrho^* \mathfrak{D}^\dagger(\varrho) + \frac{1}{2} \varrho \mathfrak{D}(\varrho^*), \quad (4.125)$$

y su correspondiente acción puede escribirse de dos formas alternativas

$$\llbracket \mathfrak{D}(\varrho^*), \varrho \rrbracket = S = \llbracket \varrho^*, \mathfrak{D}^\dagger(\varrho) \rrbracket. \quad (4.126)$$

Las ecuaciones de evolución para ϱ se obtienen haciendo variar la acción respecto de ϱ^* , i.e. $\delta_{\varrho^*} S = 0$ dejando ϱ fija, y las ecuaciones de evolución para ϱ^* se obtienen haciendo variar la acción respecto de ϱ , i.e. $\delta_{\varrho} S = 0$ dejando ϱ^* fija. En tal caso, debe

utilizarse una extensión de la versión que hemos presentado del cálculo de variaciones para el caso de funcionales de varias variables independientes, en este caso (x, t) , junto con las derivadas de los campos ϱ y ϱ^* , respecto de éstas [1].

A modo de resumen de este capítulo, hemos querido construir nuestro *libro* argumentando a favor y en contra de cada una de sus *páginas*. A este respecto, es importante señalar también y en la medida de lo posible, cuales fueron las ideas que nos han llevado a la forma final de nuestro lagrangiano. Puesto que a partir del contenido simbólico-matemático de una teoría pueden y deben esgrimirse argumentos sobre la interpretación de tales símbolos y sus derivaciones lógicas, es importante que se delimite lo mejor posible cuál fue el camino para la construcción de nuestro *libro*. Cuanto más transparente sea el procedimiento de esta construcción, e.g más evidentes se hagan sus puntos fuertes y débiles, más vulnerable se torna, para ser fácilmente derrumbado o bien utilizado como base de construcciones e interpretaciones ulteriores.

TRANSFORMACIONES
INTER-ESTOCÁSTICAS



5

TRANSFORMACIONES INTER-ESTOCÁSTICAS

En este capítulo trataremos con uno de los artículos publicados en *Physical Review*, en el cual hemos probado que todos los procesos estocásticos kolmogorovianos sobre espacios de estados numerables están conectados a través de una transformación local, i.e. dependiente del tiempo. El título hace referencia a las transformaciones que pueden efectuarse sobre una medida de probabilidad absoluta de manera de arribar a otra medida con tal que ambas correspondan a procesos estocásticos *kolmogorovianos*, en otras palabras, una transformación tal que conserve la *clase* de procesos estocásticos *kolmogorovianos*. De allí la motivación por titular a esta clase de transformaciones con el prefijo *inter* seguido de *estocásticas*, porque conectan a todos los integrantes de la clase de procesos *kolmogorovianos*. Quizá también se podría haber titulado *trasformaciones intra-kolmogorovianas*, puesto que son éstas las que no nos dejan *salir* de esta *clase* de procesos.

Estas ideas serán expuestas transcribiendo el artículo, que se encuentra en *Phys. Rev. E.* **92**, 052132, luego de eso haremos un comentario de cómo puede tratarse esta idea desde el punto de vista de una transformación de simetría sobre el lagrangiano.

Creemos que esta fuerte propiedad tiene un atractivo práctico al momento de abordar procesos complicados, i.e. que provengan de ecuaciones de evolución de mayor dificultad en su resolución.

EQUIVALENT MARKOV PROCESSES UNDER GAUGE GROUP

ABSTRACT

We have studied Markov processes on denumerable state space and continuous time. We found that all these processes are connected via gauge transformations. We have used this result before as a method for resolution of equations, included the case where the sample space is time dependent in a previous work *Phys. Rev. E* **90**, 022125 (2014). We found a general solution through a dilation of the state space, although the prior probability distribution of the states defined in this new space takes smaller values with respect to the one in the initial problem. The gauge (local) group of dilations modifies the distribution on the dilated space to restore the original process. In this work we show how Markov process in general could be linked via gauge (local) transformations and we present some illustrative examples for this results.

INTRODUCTION

Continuous-time Markov process are used to describe a variety of stochastic complex processes. They have been widely used in mathematical physics to describe the properties of important models in equilibrium and non-equilibrium, such as the Ising model [47].

Other important example of application is the use of continuous – time Markov chains in queueing theory [68]. Regarding biology, Markov chains are used to explain the properties of reaction networks, chemical system involving multiple reactions and chemical species [49] and kinetics of linear arrays of enzymes [50].

In this work, we show a way to connect a given pair of Markov processes via gauge transformations. The link between diffe-

rent processes is a mathematical observation that enrich the description of the stochastic process. In addition, in some cases this observation could become an useful tool to study a particularly complex Markov problem using a simpler auxiliary Markov process and proposing an adequate transformation to link both of them.

This approach was heuristically explored on our recent work [51] as an alternative method for the resolution of equations of Markov process on denumerable state spaces and continuous time. Nevertheless, in order to obtain a phenomenological or approximate solution to certain Markov process involves some knowledge of the system parameters and depends. This is precisely the case of [51].

In the following we present a mathematical description of a general stochastic system. The structure of the paper is: in section 5.2, a motivation of the problem in section 5.3, the formal aspects of the equivalence of the process in section 5.4 and finally the conclusions and final comments in section 5.5.

DENUMERABLE STATE SPACE AND CONTINUOUS TIME

We start by reviewing the basics aspects of this class of stochastic processes. Let's consider a stochastic system described by a Markov process with a random variable $x(t)$, which takes values from the state space at the *instant* t

$$\mathcal{S} = \{x_n : n \in \ell\}, \quad (5.1)$$

where t represent a *time variable*, some parameter used to describe the evolution of the process, that takes values from a real interval \mathcal{T} , ℓ is the countable set of *labels* for the states, such that $\ell \subseteq \mathbb{Z}_0^+$.

We defined the conditional probability to find the system in the state x_l , at the instant t , given that at instant s was in the state x_k , denoted by

$$\mathcal{P}_{lk}(t, s) = \mathbb{P}(x(t) = x_l | x(s) = x_k). \quad (5.2)$$

We understand this conditional probability as a transition element between the states $x_k \mapsto x_l$ and with a temporal evolution $s \mapsto t$. These conditional probabilities describes the time evolution of the stochastic system, in the sense that they allow us to connect any two ordered pairs (x_k, s) , (x_l, t) .

The time evolution of a Markov process, is determined by the knowledge of a prior probability distribution for each t , denoted by

$$p_n(t) = \mathbb{P}(x(t) = x_n), \quad (5.3)$$

for all $(t, n) \in \mathcal{T} \times \mathcal{L}$.

An equivalent way to describe this process is through an initial value $p_n(0)$ and a conditional probabilities $\mathcal{P}_{nm}(t, s)$, which represents the transition matrix elements of the states $x_m \mapsto x_n$. For each t the events are mutually exclusive, then

$$p_n(t) = \sum_{m \in \mathcal{L}} \mathcal{P}_{nm}(t, s) p_m(s). \quad (5.4)$$

Consequently at the time $t + \epsilon$ the probability to find the system in x_n , is given by the transition from x_m at time t , in this way

$$p_n(t + \epsilon) = \sum_{m \in \mathcal{L}} \mathcal{P}_{nm}(t + \epsilon, t) p_m(t). \quad (5.5)$$

After some elementary operations, we get:

$$d_t p_n(t) = \sum_{m \in \mathcal{L}} \mathbf{Q}_{nm}(t) p_m(t), \quad (5.6)$$

where d_t is a notation of total time derivative, and $\mathbf{Q}_{nm}(t)$ is given by

$$\mathbf{Q}_{nm}(t) = \partial_t \mathcal{P}_{nm}(t, s)|_{s=t}. \quad (5.7)$$

where $\mathbf{Q}_{nm}(t)$ is called the *infinitesimal generator*.

The equation (6.5) is named the Kolmogorov equation, the foundational work [13]. Another authors are referred later to (6.5) as the forward Kolmogorov equations [24].

We define $\boldsymbol{\varphi}(t)$ as an $|\mathcal{S}|$ -tuple of the probability distribution as $\boldsymbol{\varphi}(t) = (p_0(t), p_1(t), \dots)^\top$. And also we used a notation for the cardinal number of a set \mathcal{S} is given by $|\mathcal{S}|$ and $^\top$ represents the transposition operation.

The evolution equation for the process can be expressed in a *matrix form* as [21]

$$d_t \boldsymbol{\varphi}(t) = \mathbf{Q}(t) \boldsymbol{\varphi}(t). \quad (5.8)$$

In this way we have a mathematical description of a Markov process in terms of a set of prior probabilities $\{p_0(t), p_1(t), \dots\}$ and an infinitesimal generator $\mathbf{Q}(t)$.

MOTIVATION OF THE PROBLEM

In this section we present a motivational example to show how a given pair of Markov processes could be linked via gauge transformations. Let's consider two particular stochastic processes, with their respective infinitesimal generators

$$\mathbf{Q}_+ = \begin{pmatrix} -\nu & 0 \\ \nu & 0 \end{pmatrix}, \quad \mathbf{Q}_- = \begin{pmatrix} 0 & \mu \\ 0 & -\mu \end{pmatrix}. \quad (5.9)$$

This matrices \mathbf{Q}_+ and \mathbf{Q}_- corresponds to a *pure birth process* and a *pure death process*, whereas ν and μ are the birth and death rates, respectively. Note that $\mathbf{Q}_+ + \mathbf{Q}_-$, from (5.9), is equal to another infinitesimal generator which corresponds to a finite 2–state birth-death process. Also, to be more explicit, we can represent each of these processes through the diagrams in Fig. 5.1 and 5.2.

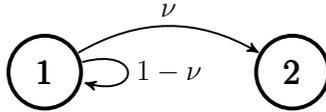


Figure 5.1: Diagram associated to the finite pure birth process with the infinitesimal generator \mathbf{Q}_+ .

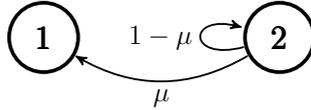


Figure 5.2: Diagram associated to the finite pure death process with the infinitesimal generator \mathbf{Q}_- .

The differential equation from (6.6) applied for each of these processes are summarized by:

$$d_t \varphi_+(t) = \mathbf{Q}_+(t) \varphi_+(t), \quad (5.10)$$

$$d_t \varphi_-(t) = \mathbf{Q}_-(t) \varphi_-(t).$$

We demonstrate that exist a 2×2 matrix λ that connect the solutions φ_+ and φ_- in a following way:

$$\varphi_+ = \lambda \varphi_-. \quad (5.11)$$

First of all only for the particular case $\nu = \mu$ we will have a constant matrix λ

$$\lambda = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad (5.12)$$

which corresponds to an interchange of the states $\textcircled{1} \longleftrightarrow \textcircled{2}$. In other words, (5.12) corresponds to a reflection that interchange the Fig. 5.1 by the Fig. 5.2 and viceversa. Explicitey for the case $\nu = \mu$, using (5.12), we have

$$\mathbf{Q}_+ = \lambda \mathbf{Q}_- \lambda^{-1}. \quad (5.13)$$

The last equation together with (5.10) involve (5.11). The matrix λ from (5.12) is a time independent change of coordinates between equations (5.10) for the case $\nu = \mu$.

We have noticed that λ is not a constant matrix, for the case $\nu \neq \mu$. For the present case we obtain the solutions with a non trivial initial conditions are

$$\varphi_+ = \begin{pmatrix} e^{-\nu t} \\ 1 - e^{-\nu t} \end{pmatrix}, \quad \varphi_- = \begin{pmatrix} 1 - e^{-\mu t} \\ e^{-\mu t} \end{pmatrix}. \quad (5.14)$$

Using the explicit solutions (5.14), the proof that the equation (5.11) is true is straightforward, since there is a matrix

$$\lambda = \begin{pmatrix} 0 & e^{(\mu-\nu)t} \\ 1 & 1 - e^{(\mu-\nu)t} \end{pmatrix}, \quad (5.15)$$

such that the equation (5.11) is true.

This example shows, in pedagogical way, that it is possible to write the solution of a stochastic process starting from another process. The bridge between φ_+ and φ_- is built through a local transformation λ . Also the correspond link between \mathbf{Q}_+ and \mathbf{Q}_- through a local transformation λ is given by

$$\mathbf{Q}_+ = \lambda \mathbf{Q}_- \lambda^{-1} + d_t \lambda \lambda^{-1}, \quad (5.16)$$

for all ν and μ . The expression (5.16) is almost equal to (5.13) but with an added term $d_t \lambda \lambda^{-1}$.

We will see that the group of this kind of transformations is structured as a gauge group. We can formalize and generalize this idea in the following sections and we prove that is possible to connect any pair of infinitesimal generator $(\mathbf{Q}, \mathbf{Q}')$ and any pair of prior distributions of probability (φ, φ') , associated to these infinitesimal generators, in a similar way that (5.16) and (5.11),

respectively

$$\mathbf{Q}' = \boldsymbol{\lambda} \mathbf{Q} \boldsymbol{\lambda}^{-1} + d_t \boldsymbol{\lambda} \boldsymbol{\lambda}^{-1}, \quad (5.17)$$

$$\varphi' = \boldsymbol{\lambda} \varphi.$$

FORMAL ASPECTS OF EQUIVALENT MARKOV PROCESSES

We considered a map $\boldsymbol{\Gamma}_\lambda$, given a non singular matrix $\boldsymbol{\lambda}$, which transforms a matrix \mathbf{Q} as

$$\boldsymbol{\Gamma}_\lambda(\mathbf{Q}) = \boldsymbol{\lambda} \mathbf{Q} \boldsymbol{\lambda}^{-1} + d_t \boldsymbol{\lambda} \boldsymbol{\lambda}^{-1}, \quad (5.18)$$

where $\boldsymbol{\lambda}, \mathbf{Q} \in \mathbb{R}^{|\ell| \times |\ell|}$ are t -dependent differentiable matrices. Thereby, $\boldsymbol{\lambda}$ is a local transformation and we will prove that $\boldsymbol{\Gamma}_\lambda$ form a group of local (gauge) transformations and $\boldsymbol{\Gamma}_\lambda \in \mathbb{R}^{|\ell| \times |\ell|}$.

In addition, in this section we study the possibility that for all pair of matrices \mathbf{Q} and \mathbf{Q}' , t -dependent and differentiable, there is a non singular matrix $\boldsymbol{\lambda}$, t -dependent and differentiable, that connect \mathbf{Q} and \mathbf{Q}' as

$$\mathbf{Q}' = \boldsymbol{\Gamma}_\lambda(\mathbf{Q}). \quad (5.19)$$

If we composed two transformation $\boldsymbol{\Gamma}_\lambda \circ \boldsymbol{\Gamma}_{\lambda'}$ with $\boldsymbol{\lambda}$ and $\boldsymbol{\lambda}'$ are non singular, we see that

$$\boldsymbol{\Gamma}_\lambda \circ \boldsymbol{\Gamma}_{\lambda'}(\mathbf{Q}) = \boldsymbol{\Gamma}_{\boldsymbol{\lambda} \boldsymbol{\lambda}'}(\mathbf{Q}). \quad (5.20)$$

From (5.20) we see that if

$$[\boldsymbol{\lambda}, \boldsymbol{\lambda}'] = \mathbf{0} \implies \boldsymbol{\Gamma}_\lambda \circ \boldsymbol{\Gamma}_{\lambda'}(\mathbf{Q}) = \boldsymbol{\Gamma}_{\lambda'} \circ \boldsymbol{\Gamma}_\lambda(\mathbf{Q}). \quad (5.21)$$

Using this properties of composition (5.20) and (5.21), we give an expression for the inverse map $\boldsymbol{\Gamma}_\lambda^{-1}$. First of all, we have trivially

$$\boldsymbol{\Gamma}_1(\mathbf{Q}) = \mathbf{Q}, \quad (5.22)$$

where $\mathbf{1}$ is the identity matrix. If we consider the composed transform $\lambda'' = \lambda\lambda'$ such that $\lambda\lambda' = \mathbf{1} = \lambda'\lambda$ then from (5.21) we have

$$\Gamma_\lambda \circ \Gamma_{\lambda'} = \mathbf{1} = \Gamma_{\lambda'} \circ \Gamma_\lambda. \quad (5.23)$$

Finally from (5.23) the inverse of Γ_λ is *unique* and given by

$$\Gamma_\lambda^{-1} = \Gamma_{\lambda^{-1}}, \quad (5.24)$$

for more details of the properties of composition (5.20) and inverse transformation (5.24) see Appendix A1.

We will demonstrate that for any pair of t -dependent differentiable matrices \mathbf{Q} and \mathbf{Q}' both of $|\ell| \times |\ell|$ there exist a non-singular t -dependent differentiable matrices λ of $|\ell| \times |\ell|$ that connect them. For that we can define the following *equivalence relation*:

$$\mathbf{Q}' \sim \mathbf{Q} \iff \exists \lambda : \mathbf{Q}' = \Gamma_\lambda(\mathbf{Q}), \quad (5.25)$$

more details that \sim is a well defined *equivalence relation* see Appendix A1. From the equivalence relation (5.25) then λ satisfies the differential equation:

$$d_t \lambda = \mathbf{Q}' \lambda - \lambda \mathbf{Q}. \quad (5.26)$$

First of all, the solution of (5.26) exist for the trivial cases $\mathbf{Q} = \mathbf{0}$ and $\mathbf{Q}' = \mathbf{0}$, i.e. we denoted by λ_1 and λ_2 the respective solutions for each case

$$d_t \lambda_1 = \mathbf{Q}' \lambda_1, \quad (5.27)$$

$$d_t \lambda_2 = -\lambda_2 \mathbf{Q}. \quad (5.28)$$

We can obtain (λ_1, λ_2) as an iterative non singular solutions, for more details of this solution see Appendix A2. The existence of solutions for (5.27) and (5.28) implies that λ_1 and λ_2 connects $\mathbf{Q}' \sim \mathbf{0}$ and $\mathbf{0} \sim \mathbf{Q}$, respectively. This implication is true from the

definition of the equivalence relation. From the existence of solutions for (5.27) and (5.28) then we have

$$\exists \lambda_1 : \mathbf{Q}' = \Gamma_{\lambda_1}(\mathbf{0}) \iff \mathbf{Q}' \sim \mathbf{0}, \quad (5.29)$$

$$\exists \lambda_2 : \mathbf{0} = \Gamma_{\lambda_2}(\mathbf{Q}) \iff \mathbf{0} \sim \mathbf{Q}, \quad (5.30)$$

and from transitivity of the equivalence relation (5.25) we have $\mathbf{Q}' \sim \mathbf{Q}$, this means that there is a given λ that $\mathbf{Q}' = \Gamma_{\lambda}(\mathbf{Q})$.

We express the solution λ as a function of the solutions of (5.27) and (5.28), (λ_1, λ_2) , respectively. We say that a solution λ built in this way is a *transitive solution*, or *composite solution*, the name will be clear in the construction procedure of the solution λ . From (5.29) and (5.30) we see that the *transitivity solution* is constructed from the composition of transformations $\mathbf{Q}' = \Gamma_{\lambda_1}(\mathbf{0})$ and $\mathbf{0} = \Gamma_{\lambda_2}(\mathbf{Q})$ as follows

$$\mathbf{Q}' = \Gamma_{\lambda_1}(\Gamma_{\lambda_2}(\mathbf{Q})) \quad (5.31)$$

from the composition rule (5.20) applied to (5.31)

$$\mathbf{Q}' = \Gamma_{\lambda_1 \lambda_2}(\mathbf{Q}) \quad (5.32)$$

where the *transitive solution* is given by

$$\lambda = \lambda_1 \lambda_2. \quad (5.33)$$

We have demonstrated that for any pair of matrices $(\mathbf{Q}, \mathbf{Q}')$, t -dependent and differentiable, there is a non singular matrix λ , t -dependent and differentiable, that connect \mathbf{Q} and \mathbf{Q}' through the map Γ_{λ} , given by the expression (5.18)

$$\mathbf{Q}' = \Gamma_{\lambda}(\mathbf{Q}). \quad (5.34)$$

Suppose now that \mathbf{Q} and \mathbf{Q}' are the infinitesimal generators, t -dependent and differentiable, of the following differential equa-

tions

$$d_t \varphi = \mathbf{Q} \varphi, \quad (5.35)$$

$$d_t \varphi' = \mathbf{Q}' \varphi',$$

finally from (5.34) and (5.35) we have

$$\varphi' = \lambda \varphi. \quad (5.36)$$

We found that for any pair infinitesimal generators, t -dependent and differentiable, $(\mathbf{Q}, \mathbf{Q}')$ associated to (5.35), exist another t -dependent and differentiable matrix λ that connect the distribution of probability φ and φ' .

Until now we have considered the equivalence of Markov processes of the same dimension, i.e. the state spaces of every couple of processes have the same cardinality. We will go one step further, we will prove the equivalence of all continuous-time Markov processes on a denumerable state space.

Without loss of generality we define $(\mathbf{Q}, \mathcal{S})$ and $(\mathbf{Q}', \mathcal{S}')$ as the respective infinitesimal generators and state spaces, such that $n = |\mathcal{S}| < |\mathcal{S}'| = n'$. We can construct another process associated with \mathbf{Q} , such that the infinitesimal generator, \mathbf{Q} , is given by

$$\mathbf{Q}_{ij} = \begin{cases} \mathbf{Q}_{ij}, & \forall i, j \in [1, n] \subseteq \mathbb{N} \\ 0, & \forall i, j \in [n+1, n'] \subseteq \mathbb{N} \end{cases} \quad (5.37)$$

or in block form

$$\mathbf{Q} = \begin{pmatrix} \mathbf{Q} & 0 & \dots \\ 0 & 0 & \dots \\ \vdots & \vdots & \ddots \end{pmatrix}. \quad (5.38)$$

The matrix \mathbf{Q} corresponds to a new process on a state space \mathbf{S} which have the same cardinality of \mathcal{S}' . We have completed the process on \mathcal{S} with a number of *redundant* states, such that the

resulting state space \mathbf{S} satisfy $|\mathbf{S}| = n' = |\mathcal{S}'|$. For illustrative purposes the Fig. 5.3 shows the composed state space \mathbf{S} from \mathcal{S} and a set of isolated and absorbing states $\{\textcircled{k} : k \in [n+1, n'] \subseteq \mathbb{N}\}$:

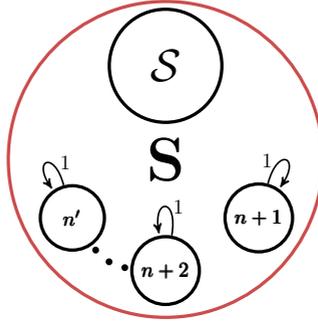


Figura 5.3: (Color online) This diagram shows the composed process on the state space $\mathbf{S} \approx \mathcal{S}'$, in the sense that $|\mathbf{S}| = |\mathcal{S}'|$, with the infinitesimal generator \mathbf{Q} .

All the states of $\{\textcircled{k} : k \in [n+1, n'] \subseteq \mathbb{N}\}$ are isolated or mutually disconnected and also from each state of \mathcal{S} ; they are all absorbing states. If the process start in some \textcircled{k} of this redundant set, it stays there forever.

A final comment is related to the case where the state space is time dependent Markov process. This is in the sense that its cardinality is a function of time n_t , meaning that for a given generation time t (this is $t \in \mathbb{N}$) the state space $\mathcal{S}_t : |\mathcal{S}_t| = n_t$. Explicitly we have

$$\mathcal{S}_t = \{x_1, x_2, \dots, x_{n_t}\}, \quad (5.39)$$

for a given t . If \mathbf{Q}_t is the infinitesimal generator for each t , we can construct another process associated with \mathbf{Q}_t , such that the infinitesimal generator, \mathbf{Q} , is constructed in a similar way that (5.37)

$$\mathbf{Q}_{ij} = \begin{cases} (\mathbf{Q}_t)_{ij}, & \forall i, j \in [1, n_t] \subseteq \mathbb{N} \\ 0, & \forall i, j \in [n_t + 1, N] \subseteq \mathbb{N} \end{cases} \quad (5.40)$$

for sufficiently big number $N \in \mathbb{Z}_0^+$. In other terms we can write

$$\mathbf{Q}_{ij} = (\mathbf{Q}_t)_{ij} u(n_t - i)u(n_t - j), \quad (5.41)$$

where $u(x)$ is a Heaviside step function

$$u(x) := \begin{cases} 0 & : x < 0 \\ 1 & : x \geq 0 \end{cases}. \quad (5.42)$$

In a matrix form we express the dilution of \mathcal{S}_t inside \mathbf{S} , such that $S = \{x_1, \dots, x_N\}$, for finite value of N or $N \rightarrow \infty$:

$$\mathbf{Q} = \begin{pmatrix} \mathbf{Q}_t & 0 & \dots \\ 0 & 0 & \dots \\ \vdots & \vdots & \ddots \end{pmatrix} \quad (5.43)$$

The matrix \mathbf{Q} corresponds to a new process on a state space \mathbf{S} , which it is equivalent to any other.

CONCLUSION AND FINAL OBSERVATIONS

The aim of present work is to demonstrate that there is a way to modify the solution for a simple or known process, which is represented by the infinitesimal generator \mathbf{Q} , in order to get another process partially known, or at least with a very profound difficulty to be resolved, and represented by \mathbf{Q}' .

We have shown how for a given pair of Markov processes $(\mathbf{Q}, \mathcal{S})$ and $(\mathbf{Q}', \mathcal{S}')$ they could be linked via gauge (local) transformations λ , that allow us obtain \mathbf{Q}' from \mathbf{Q} via Γ_λ .

Even when the state space of each process has different cardinality, it is still possible to establish a link via a local transformation.

This connection also could be explored when the state space is time dependent, in the sense that the number of states change with time, that was used in [51], intuitively.

In addition, allows us to address a new problem from another known one, through a non-local modulation of the well known solution, following the expression (5.36). We have not only shown that this is feasible to do through formal and constructive proof of existence of that λ , but also we indicated what is the right way to do it, should be across a linear and local (time-dependent) operation.

Future research through a Lagrangian description (working process) may find novel applications of the present proposal. In this approach the role of this kind of transformation λ will be a symmetry of the Lagrangian. A gauge theory of stochastic processes can be improved formally through a variational principle.

ACKNOWLEDGMENTS

We thank our respective institutions, UNQ-IFIBA, UGR and CONICET and a special mention to Fernando Cornet. Recall also our anonymous readers and reviewers for their contribution to this work, including Federico G. Vega. Gabriel Lio, Micaela Moretton and María Clara Caruso have shared with us the labor as local coaches.

APPENDIX

A1. SOME RELEVANT PROPERTIES OF MAP Γ_λ

• COMPOSITE MAPS

We composed two transformation $\Gamma_\lambda \circ \Gamma_{\lambda'}$ with λ and λ' are non singular and then prove that

$$\Gamma_\lambda \circ \Gamma_{\lambda'}(\mathbf{q}) = \Gamma_{\lambda\lambda'}(\mathbf{q}). \quad (5.1)$$

We calculate directly

$$\begin{aligned}
\Gamma_\lambda \circ \Gamma_{\lambda'}(\mathbf{Q}) &= \Gamma_\lambda(\lambda' \mathbf{Q} \lambda'^{-1} + d_t \lambda' \lambda'^{-1}) \\
&= \lambda[\lambda' \mathbf{Q} \lambda'^{-1} + d_t \lambda' \lambda'^{-1}] \lambda^{-1} + d_t \lambda \lambda^{-1} \\
&= \lambda \lambda' \mathbf{Q} (\lambda \lambda')^{-1} + \lambda d_t \lambda' (\lambda \lambda')^{-1} + d_t \lambda \lambda' (\lambda \lambda')^{-1} \\
&= \lambda \lambda' \mathbf{Q} (\lambda \lambda')^{-1} + d_t (\lambda \lambda') (\lambda \lambda')^{-1} \\
&= \Gamma_{\lambda \lambda'}(\mathbf{Q}),
\end{aligned}$$

where in the second line the term $d_t \lambda' \lambda'^{-1} = d_t \lambda' (\lambda \lambda^{-1}) \lambda'^{-1}$. This complete the demonstration that expression (5.20) is satisfied.

• **INVERSE MAP**

We calculate explicitly $\Gamma_{\lambda^{-1}}(\mathbf{Q})$ and then prove that

$$\Gamma_{\lambda^{-1}}(\mathbf{Q}) = \Gamma_\lambda^{-1}(\mathbf{Q}), \quad (5.2)$$

for all \mathbf{Q} . Let's calculate the left hand of (5.2)

$$\begin{aligned}
\Gamma_{\lambda^{-1}}(\mathbf{Q}) &= \lambda^{-1} \mathbf{Q} \lambda + d_t (\lambda^{-1}) \lambda \\
&= \lambda^{-1} \mathbf{Q} \lambda + d_t (\lambda^{-1}) \lambda \\
&= \lambda^{-1} \mathbf{Q} \lambda - \lambda^{-1} d_t \lambda.
\end{aligned}$$

Finally we check directly that $\Gamma_{\lambda^{-1}}$ is equal to Γ_λ^{-1}

$$\begin{aligned}
\Gamma_\lambda \circ \Gamma_{\lambda^{-1}}(\mathbf{Q}) &= \Gamma_\lambda(\lambda^{-1} \mathbf{Q} \lambda - \lambda^{-1} d_t \lambda) \\
&= \lambda(\lambda^{-1} \mathbf{Q} \lambda - \lambda^{-1} d_t \lambda) \lambda^{-1} + d_t \lambda \lambda^{-1} \\
&= \lambda \lambda^{-1} \mathbf{Q} \lambda \lambda^{-1} - \lambda \lambda^{-1} d_t \lambda \lambda^{-1} + d_t \lambda \lambda^{-1} \\
&= \mathbf{Q} - d_t \lambda \lambda^{-1} + d_t \lambda \lambda^{-1} \\
&= \mathbf{Q},
\end{aligned}$$

where we used $d_t (\lambda^{-1} \lambda) = \mathbf{0} \implies d_t (\lambda^{-1}) \lambda = -\lambda^{-1} d_t \lambda$. This complete the demonstration that (5.24) is true, i.e. $\Gamma_{\lambda^{-1}} = \Gamma_\lambda^{-1}$.

• Γ_λ AS AN EQUIVALENCE RELATION

We say that the map Γ_λ define an equivalence relation between the space of between of the vector space of matrices of the same dimension. For a given two matrices $(\mathbf{Q}, \mathbf{Q}')$ we can define a relation between them

$$\mathbf{Q}' \sim \mathbf{Q} \iff \exists \lambda : \mathbf{Q}' = \Gamma_\lambda(\mathbf{Q}), \quad (5.3)$$

where $\Gamma_\lambda(\mathbf{Q}) := \lambda \mathbf{Q} \lambda^{-1} + d_t \lambda \lambda^{-1}$ and λ is a non singular matrix. This relation \sim is an equivalence in the sense that $\forall \mathbf{Q}, \mathbf{Q}', \mathbf{Q}''$ the following properties are true:

- (R) $\mathbf{Q} \sim \mathbf{Q}$ (*reflexivity*)
- (S) $\mathbf{Q} \sim \mathbf{Q}' \Rightarrow \mathbf{Q}' \sim \mathbf{Q}$ (*symmetry*)
- (T) $\mathbf{Q}'' \sim \mathbf{Q}' \wedge \mathbf{Q}' \sim \mathbf{Q} \Rightarrow \mathbf{Q}'' \sim \mathbf{Q}$ (*transitivity*)

The first assertion (R) is true from the identity matrix $\lambda = \mathbf{1}$ and by definition $\Gamma_\lambda(\mathbf{Q}) = \mathbf{Q}$. The assertion (S) is also true from the existence of the inverse matrix λ^{-1} and construct through (5.2) the *inverse* connection $\mathbf{Q}' \sim \mathbf{Q}$. The last assertion (T) is true from the composed transformation of non singular matrices $\lambda = \lambda'' \lambda'$ and (5.1), such that $\mathbf{Q}'' = \Gamma_{\lambda''}(\mathbf{Q}')$ and $\mathbf{Q}' = \Gamma_{\lambda'}(\mathbf{Q})$, then $\mathbf{Q}'' = \Gamma_{\lambda''}(\Gamma_{\lambda'}(\mathbf{Q})) = \Gamma_{\lambda'' \lambda'}(\mathbf{Q}) = \Gamma_\lambda(\mathbf{Q})$. Finally, we get to $\mathbf{Q}'' \sim \mathbf{Q}$.

A2. ALTERNATIVE EXPRESSION FOR λ

In the present work we said that in order to give an expression for the solution of (5.26), we see that the *transitivity solution* is constructed from the composed transformation $\lambda = \lambda_1 \lambda_2$ and (5.1)

$$\Gamma_{\lambda_1 \lambda_2}(\mathbf{Q}) = \Gamma_{\lambda_1} \circ \Gamma_{\lambda_2}(\mathbf{Q}) = \mathbf{Q}', \quad (5.4)$$

where λ_1 and λ_2 are solution of (5.27) and (5.28), respectively

$$\mathbf{Q}' \sim \mathbf{0} \iff d_t \lambda_1 = \mathbf{Q}' \lambda_1, \quad (5.5)$$

$$\mathbf{0} \sim \mathbf{Q} \iff d_t \lambda_2 = -\lambda_2 \mathbf{Q}.$$

We express the solutions of (5.5) as a formal iterative solution

$$\begin{aligned} \lambda_1(t) = & \left[1 + \int_0^t \mathbf{Q}(t_1) dt_1 \right. \\ & \left. + \int_0^t \int_0^{t_1} \mathbf{Q}(t_1) \mathbf{Q}(t_2) dt_1 dt_2 + \dots \right] \lambda_1(0), \end{aligned} \quad (5.6)$$

$$\begin{aligned} \lambda_2(t) = & \lambda_2(0) \left[1 - \int_0^t \mathbf{Q}'(t_1) dt_1 \right. \\ & \left. + \int_0^t \int_0^{t_1} \mathbf{Q}'(t_1) \mathbf{Q}'(t_2) dt_1 dt_2 + \dots \right]. \end{aligned}$$

We obtain a general expression of the iterative solution (5.6) through a Magnus series [52]

$$\lambda_1(t) = \sum_{n \in \mathbb{N}} \frac{1}{n!} \mathbf{\Lambda}_n(t), \quad (5.7)$$

$$\lambda_2(t) = \sum_{n \in \mathbb{N}} \frac{(-1)^n}{n!} \mathbf{\Lambda}'_n(t),$$

where each $\mathbf{\Lambda}_n(t)$ and $\mathbf{\Lambda}'_n(t)$ are given by

$$\mathbf{\Lambda}_n(t) = \int_0^t \mathbf{Q}(t_1) dt_1 \int_0^{t_1} \mathbf{Q}(t_2) dt_2 \cdots \int_0^{t_{n-1}} \mathbf{Q}(t_{n-1}) dt_n, \quad (5.8)$$

$$\mathbf{\Lambda}'_n(t) = \int_0^t \mathbf{Q}'(t_1) dt_1 \int_0^{t_1} \mathbf{Q}'(t_2) dt_2 \cdots \int_0^{t_{n-1}} \mathbf{Q}'(t_{n-1}) dt_n.$$

ALBORES PROBABILÍSTICOS DEL GRUPO GAUGE

La presente sección pretende ser un anexo al artículo anteriormente transcrito. Simplemente es un comentario para argumentar desde un punto de vista de la teoría de probabilidades una variante que garantice la forma de este tipo de transformaciones. Es decir, existe un procedimiento constructivo de este tipo de transformaciones a partir de argumentos propios de la teoría de probabilidades.

Para fijar ideas, supongamos dos variables aleatorias $x(t)$ e $x'(t)$, definidas en espacios de estados numerables E y E' respectivamente. Sean $q(t)$ y $q'(t)$ las coordenadas prospectivas asociadas los procesos estocásticos $\{x(t)\}_{t \in T}$ y $\{x'(t)\}_{t \in T'}$, respectivamente. El artículo que hemos transcrito ha dejado claro que la forma general de vincular q con q' es mediante una matriz λ no singular y local, i.e. dependiente de t : $q' = \lambda q$.¹

La pregunta por conocer una relación *entre* la medida $P(x)$ y la medida $P(x')$ puede responderse de manera suficiente si conocemos su distribución *conjunta*: $P(x, x')$. Esta cantidad podrá ser vista como un *nexo* entre ambas medidas $P(x)$ y $P(x')$, a través de una adecuada marginalización.

Veamos como funciona a partir de la probabilidad conjunta $P(x(t), x'(t))$, procedamos a marginalizar sobre $x(t)$ para obtener una distribución unidimensional sobre $x'(t)$. Denotemos esto de manera simplificada como sigue

$$P(x'(t)) = \sum_{x(t)} P(x'(t), x(t)), \tag{5.9}$$

$$P(x'(t)) = \sum_{x(t)} P(x'(t)|x(t))P(x(t)).$$

¹En esta sección haremos omisión, según convenga, de la dependencia con t de los objetos o bien de los estados de las variables aleatorias a fin de simplificar la notación y agilizar el discurso.

Reescribimos ahora (5.9) haciendo referencia explícita a los estados de $E = \{x_i\}_{i \in I}$ y $E' = \{x'_j\}_{j \in I'}$, junto con las definiciones $q_i(t) := P(x(t) = x_i)$ y $q'_j(t) := P(x'(t) = x'_j)$

$$q'_j(t) = \sum_{i \in I} P(x'(t) = x'_j | x(t) = x_i) q_i(t), \quad (5.10)$$

de la cual podemos observar la relación lineal entre q y q' . Las probabilidades de transición *entre* procesos kolmogorovianos denotadas por $P(x'(t) = x'_j | x(t) = x_i)$ hacen las veces de elementos de matriz $\lambda_{ji}(t)$.

Hemos visto que este tipo de transformaciones de coordenadas, al menos en lo que atañe a las *prospectivas*, resultan ser de la forma $f_t : q \mapsto q'$. En particular, todas ellas son lineales, i.e. la forma $f_t(q) = \lambda(t)q$. Por tanto, la eventual dependencia explícita con t proviene de la matriz $\lambda(t)$. Estas transformaciones pueden ser contempladas dentro de una variedad mayor de transformaciones, que son propias de una formulación lagrangiana o hamiltoniana, conocida como *transformaciones canónicas*. La clase $q' = f_t(q)$ corresponde a las denominadas transformaciones puntuales [30, 31, 35], en tales textos se denotan como $q' = F(q, t)$.

Para terminar esta sección diremos que aún nos resta describir qué tipo de transformaciones requerimos para poder convertir medidas *retrospectivas*, i.e. p en otras p' que nos dejen, de nuevo, en un proceso estocástico *kolmogoroviano*. En otras palabras, si λ conecta dos medidas prospectivas q con q' , mediante la regla: $q' = \lambda q$, surge la pregunta de qué tipo de transformaciones hará lo propio entre p y p' , de manera que las coordenadas (q, p) asociadas a un proceso kolmogoroviano dado, resulte entonces que las coordenadas (q', p') estén ambas asociadas a otro proceso kolmogoroviano vía λ .

TRANSFORMACIONES CANÓNICAS

En esta sección conjugaremos las ideas expuestas anteriormente en el artículo Phys. Rev. E. 92 en este capítulo a la luz del formalismo lagrangiano y hamiltoniano que hemos presentado en el capítulo anterior.

Hemos mostrado que las transformaciones locales, reguladas por la matriz $\lambda(t)$, permiten definir clases de equivalencia entre procesos estocásticos. Veremos ahora que éstas corresponden a *transformaciones canónicas* [30, 31, 35]. Dichas transformaciones reciben ese nombre, puesto que conservan los corchetes de Poisson y la forma *canónica* de las ecuaciones de evolución, aun cuando la propia forma del hamiltoniano no permanezca invariante. Este es el objetivo de la presente sección.

En el artículo que hemos transcrito nos hemos enfocado en estudiar las transformaciones *gauge* de la ecuación forward. Nos resta ahora hacer lo propio con la ecuación backward, es decir, queremos saber cómo se conectan las medidas absolutas backward asociadas a dos procesos *kolmogorovianos*.

Retomamos nuestra nomenclatura usual empleada en los capítulos anteriores, al menos la que hemos utilizado para las distribuciones forward (q) y backward (p). Veamos el siguiente argumento, dados dos procesos estocásticos kolmogorovianos, asociados a dos variables aleatorias $x(t)$ y $x'(t)$ cuyos espacios de estados E y E' tengan la misma *cantidad* de elementos (igual cardinal); las respectivas ecuaciones forward serán de la forma

$$\dot{q} = \Omega q \quad (5.11)$$

$$\dot{q}' = \Omega' q' \quad (5.12)$$

pueden conectarse a través de una matriz $\lambda(t)$, tal que $q' = \lambda q$, incluso para el caso en que las matrices Ω y Ω' dependan de t . Esto ha quedado demostrado en el artículo anterior, pues para cada par de matrices (Ω, Ω') existe una matriz no singular λ tal

que

$$\Omega' = \Gamma_\lambda(\Omega) \quad (5.13)$$

A partir de la relación (5.13) entre las matrices Ω y Ω' que-remos derivar la relación entre p y p' , i.e. buscar la transforma-ción local que permite establecer la equivalencia de las ecuacio-nes backward. A fines de simplificar el cálculo, escribamos las ecuaciones backward, asociadas al mismo par de procesos esto-cásticos, asociados a (Ω, Ω') , como

$$\dot{p}_* = -p_*\Omega \quad (5.14)$$

$$\dot{p}'_* = -p'_*\Omega' \quad (5.15)$$

donde $p_* = p^t$ y $p'_* = (p')^t$ son covectores, i.e. vectores fila, asocia-dos a p y p' , i.e. vectores columna, respectivamente.

Veamos si existe una matriz $\lambda_*(t)$ tal que $p'_* = p_*\lambda_*$. Espe-ramos también que λ_* esté relacionada con λ , pues ésta última conecta el par (Ω, Ω') a través de (5.13), para ello calculemos

$$\begin{aligned} d_t(p_*\lambda_*) &= d_t p_* \lambda_* + p_* d_t \lambda_* \\ &= -p_*\Omega\lambda_* + p_* d_t \lambda_* \\ &= -(p_*\lambda_*) (\lambda_*^{-1}\Omega\lambda_*) + (p_*\lambda_*) \lambda_*^{-1} d_t \lambda_* \quad (5.16) \\ &= -(p_*\lambda_*) \left[\lambda_*^{-1}\Omega\lambda_* - \lambda_*^{-1} d_t \lambda_* \right] \end{aligned}$$

$$d_t(p_*\lambda_*) = -(p_*\lambda_*)\Gamma_{\lambda_*^{-1}}(\Omega)$$

hemos omitido la dependencia de Ω con t a fin de simplificar y hemos utilizado que p_* satisface (5.14). Si buscamos que la ma-triz transformada $\Gamma_{\lambda_*^{-1}}(\Omega)$ sea igual a Ω' , o bien que $p_*\lambda_* = p'_*$, en cualquiera de los dos casos por comparación con (5.15) y hacien-do uso de (5.13) tenemos

$$\Gamma_{\lambda_*^{-1}}(\Omega) = \Omega' = \Gamma_\lambda(\Omega), \quad (5.17)$$

con lo cual se concluye que

$$\lambda_* = \lambda^{-1}. \quad (5.18)$$

Podemos resumir lo siguiente dado un par de variables aleatorias $(x(t), x'(t))$, asociadas procesos estocásticos kolmogorovianos, que toman valores en espacios de estados numerables (E, E') de igual cardinal, podemos mapear sus respectivas medidas absolutas prospectivas (q, q') y retrospectivas (p, p') a través de cierta matriz no singular y local $\lambda(t)$

$$q' = \lambda q \quad (5.19)$$

$$p' = [\lambda^{-1}]^t p. \quad (5.20)$$

Esta última expresión puede escribirse como $p'_* = p_* \lambda^{-1}$.

De aquí podemos agregar otro invariante ahora *universal* de la teoría. Ya hemos dicho que el producto $p.q$ es invariante en todo instante para un dado proceso; ahora podemos extrapolar su carácter invariante a *todos* los procesos estocásticos equivalentes entre sí. En símbolos, tenemos que

$$p'(t).q'(t) = p(t).q(t). \quad (5.21)$$

Esto corresponde a una invariancia de este producto, no solo para cada instante de tiempo, es decir, en *cada* proceso estocástico; sino también *entre* los procesos estocásticos que pertenecen a la misma clase de equivalencia.

El tipo de cambio de coordenadas que induce una matriz no singular y local λ corresponde a las denominadas *canónicas*. Para probar esta afirmación, comprimimos las transformaciones (5.19) y (5.20) con la notación simpléctica $\xi = (q, p)^t$ y $\xi' = (q', p')^t$. De manera que el cambio de coordenadas $\xi \mapsto \xi'$ es lineal y mediada por λ como sigue

$$\xi' = \Lambda \xi \quad (5.22)$$

como la transformación $\Lambda : \xi \mapsto \xi'$ es lineal, según (5.22), la matriz *jacobiana* asociada² $\partial_\xi \xi'$ es igual a la propia Λ

$$\Lambda := \begin{pmatrix} \lambda & 0 \\ 0 & [\lambda^{-1}]^t \end{pmatrix}. \quad (5.23)$$

Si ocurre que la matriz *jacobiana* $\partial_\xi \xi'$ es tal que

$$(\partial_\xi \xi')^t J \partial_\xi \xi' = J, \quad (5.24)$$

se dice que tal transformación $\Lambda : \xi \mapsto \xi'$ se denomina *canónica* [30, 31, 35]. Dicha condición (5.24) se verifica idénticamente

$$\begin{pmatrix} \lambda^t & 0 \\ 0 & \lambda^{-1} \end{pmatrix} \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix} \begin{pmatrix} \lambda & 0 \\ 0 & [\lambda^{-1}]^t \end{pmatrix} \equiv \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix} \quad (5.25)$$

esta condición nos lleva a la invariancia del corchete de Poisson de dos funciones f y g , respecto de un cambio de coordenadas. En otras palabras el corchete de Poisson es invariante bajo transformaciones canónicas. Para probar esto escribimos $[f, g]_\xi$ en forma compacta usando (4.73)

$$[f, g]_\xi = \partial_\xi f \cdot J \partial_\xi g \quad (5.26)$$

hemos agregado el subíndice ξ para denotar que estamos calculando el corchete en esas coordenadas. Utilizando el cambio de variables sugerido por (5.22), expresemos $\partial_\xi f = \partial_\xi \xi' \partial_{\xi'} f$, idem con g , obtenemos

$$\begin{aligned} [f, g]_\xi &= \partial_{\xi'} f \cdot [(\partial_\xi \xi')^t J \partial_\xi \xi'] \partial_{\xi'} g \\ &= \partial_{\xi'} f \cdot J \partial_{\xi'} g \end{aligned} \quad (5.27)$$

$$[f, g]_\xi = [f, g]_{\xi'}$$

de la primera a la segunda línea hemos usado (5.24).

²Nos referimos a la matriz asociada a la transformación (5.22), digamos M , que se puede denotar de manera compacta como $\partial_\xi \xi'$ cuyos elementos son $M_{ij} = \partial_{\xi_j} \xi'_i$.

También puede probarse que el lagrangiano L que hemos introducido en el capítulo anterior, permanece invariante en la clase de equivalencia que define λ . Es decir, dado un proceso kolmogoroviano de coordenadas (q', p') y matriz de transiciones infinitesimales Ω' , su correspondiente lagrangiano L' es idénticamente igual al lagrangiano L , el cual proviene otro proceso kolmogoroviano de coordenadas (q, p) y matriz de transiciones infinitesimales Ω , tal que $\Gamma_\lambda(\Omega) = \Omega'$.

Según lo expuesto en el capítulo anterior podemos trabajar equivalentemente con un lagrangiano³ $L = p \cdot \dot{q} - p \cdot \Omega q$. Utilizando esta forma para L , la demostración es inmediata

$$\begin{aligned} L' &= p' \cdot \dot{q}' - p' \cdot \Omega' q' \\ &= p \cdot [\lambda^{-1}(\dot{\lambda}q + \lambda\dot{q})] - p \cdot [\lambda^{-1}(\lambda\Omega\lambda^{-1} + \dot{\lambda}\lambda^{-1})\lambda q] \\ L &= p \cdot \dot{q} - p \cdot \Omega q = L. \end{aligned}$$

Explicítando aún más, hemos probado que $L'(\xi', \dot{\xi}', t) = L(\xi, \dot{\xi}, t)$, para todo par $(\xi', \dot{\xi}')$ de coordenadas conectadas mediante una transformación canónica $\xi' = \Lambda\xi$, dada por (5.23). Dados 2 lagrangianos asociados a 2 procesos *kolmogorovianos*, pertenecientes a la misma clase de equivalencia⁴, son idénticamente iguales.

Llegados a esta instancia cabe mencionar que el volúmen del *espacio fásico* de coordenadas (q, p) se conserva en el tiempo. No obstante, tenemos una diferencia sutil respecto a lo que ocurre en la mecánica clásica. En nuestro caso cada punto del *espacio fásico* corresponde a la *propensión* de los estados del sistema. La evolución temporal del sistema puede ser vista como una transformación canónica y se corresponde al *movimiento* de este punto sobre lo que se denomina una *trayectoria fásica*. Tendremos entonces un teorema de *Louville*, similar a la mecánica clásica: el volúmen de ese espacio fásico se conserva en el tiempo. [30, 31]

³Pues es idénticamente igual a $L = \frac{1}{2}(p \cdot \dot{q} - q \cdot \dot{p}) - p \cdot \Omega q$, ya que la diferencia entre ambos es $d_t(q \cdot p)$, que es idénticamente nulo.

⁴Nos referimos a dos procesos estocásticos que pueden conectarse mediante una matriz no singular y local $\lambda(t)$.

ANCESTROS,
UN MODELO A ESCULPIR



6

ANCESTROS, UN MODELO A ESCULPIR

El presente capítulo pretende mostrar un modelo de aplicación a la biología. Nos interesa calcular el número de ancestros de una especie que requiere de 2 progenitores para reproducirse. El modelo inicialmente planteado presenta notorias dificultades para ser resuelto, puesto que la cantidad de estados varía con el tiempo. Es pues un complejo escenario para poner a prueba alguna de las ideas presentadas en los capítulos anteriores.

El procedimiento que utilizaremos para dar una distribución de ancestros a una generación dada es análogo a lo que Miguel Angel Buonarroti afirmó en su famoso soneto¹. En este caso, el *mármol* viene a ser una progresión geométrica, i.e. la *totalidad*. En cada generación quitamos aquello que *sobra* y que no formará parte de la obra. En otras palabras, describiremos un proceso estocástico que de cuenta de aquellos que *no sean* ancestros, respecto de esa *totalidad*.

Los detalles fueron publicados en *Phys. Rev. E. 90, 022125* y lo que sigue es una transcripción del mismo.

¹Dejamos al lector solo una pieza del mismo:

*No tiene el gran artista ni un concepto
Que un mármol sólo en sí no contenga
En su extensión, mas solo a tal llega
La mano que obedece al intelecto (1-4)*

MARKOV CHAIN APPROACH TO THE DISTRIBUTION OF ANCESTORS IN SPECIES OF BIPARENTAL REPRODUCTION

ABSTRACT

We studied how to obtain a distribution for the number of ancestors in species of sexual reproduction. Present models concentrate on the estimation of distributions repetitions of ancestors in genealogical trees. It has been shown that is not possible to reconstruct the genealogical history of each species along all its generations by means of a geometric progression. This analysis demonstrates that it is possible to rebuild the tree of progenitors by modeling the problem with a Markov chain. For each generation, the maximum number of possible ancestors is different. This brings huge problems for the resolution. We found a solution through a dilation of the sample space, although the distribution defined there takes smaller values respect to the initial problem. In order to correct the distribution for each generation, we introduced the invariance under gauge (local) group of dilations. These ideas can used to study the interaction of several processes and provide a new approach on the problem of the common ancestor. In the same direction, this model also provide some elements that can be used to improve models of animal reproduction.

INTRODUCTION

Up till now, previous attempts aiming to calculate the number of ancestors in species of sexual reproduction have not been totally successful. Present models concentrate on the estimation of distributions of ancestors repetitions in genealogical trees [56, 57, 58]. It has been shown that is not possible to reconstruct the genealogical history of each species along all its generations by means of a geometric progression [59]. The reason for that is the

geometric progression is determined by a sequence of independent events. We postulate that *blood relationship* is a kind of interaction that connects the events. It is possible to re-build the tree of progenitors by modelling the problem with a Markov chain. If we consider a random variable which represents the number of ancestors present in a given generation, the size of the sample space depends of each generation. This brings serious complications on the solution of the problem, not only of mathematical nature. We propose submerge the original sample space into a larger one. This *dilution* modifies the probability distribution. We show the need to implement a covariant derivative, due to a gauge transformation, which leaves the evolution equation invariant and correct the probability distribution.

The main goal of present work is describe the distribution of ancestors for species with sexual reproduction, but also show the novel method used here to solve other stochastic problems.

There are two important assumptions about the biology of the considered system. The first one is that the species described here has not specific behavior of sexual partner selection (random mating reproduction) [60, 61]. Many species or population groups exhibit this kind of reproduction. This is the most simple case to perform the calculation. The second assumption is about the population size. The distribution of ancestors for a given generation is contained in a population large enough to not force the selection of sexual partners blood related. The partners could be blood related or not, randomly. Current model presents a random mating in non-overlapping generations with negligible mutation and selection. These two assumptions are common to develop population genetic models, in particular, these are present in the Hardy-Weinberg principle [63, 62].

In this work we show a way to calculate a probability distribution to get a certain number of ancestors for each generation. We have obtained its first two cumulants: the expected value of the number and its dispersion.

Specific conditions about small size populations, or specific sexual behavior can be considered later as modifications of the general case described here.

Present work will be useful in order to understand the origin of species extrapolating the individual genealogy for all members at the beginning of the species. It is possible to go one step further, to establish how populations can be affected by certain conditions, such as isolation or migration of individuals, by studying population groups with different genetic pool [64]. These ideas can be used to perform more realistic models in animal populations and also, improve estimations about extinction processes.

A MARKOVIAN APPROACH TO THE ANCESTORS PROBLEM

To calculate the number of ancestors of an individual it is necessary to use a statistical approach. If we simply accept that 2^{t+1} allow us to calculate the number of ancestors in the t -generation, where $t = 0$ is the generation of progenitors of the first order (or parents for short) and so forth, we arrive at an absurdity. As we turn to past generations, the probability that some ancestors have been relatives is significantly larger [56, 57, 58]. This implies a restriction on the number of ancestors with respect to 2^{t+1} . This last quantity corresponds to the maximum possible number of ancestors in each t -generation.

There are several examples showing different ways in which the number of ancestors is reduced with respect to the maximum number in each generation. As stated in [59], the reduction of the number of ancestors, compared to 2^{t+1} , is caused by blood relationship. Figure 6.1 shows, as an example, only the three first generations of genealogical tree with two different ways to constrain the number of ancestors. There is a way to weight the blood relationship using a statistical approach that includes all possible kinds of relationship in each generation. In this approach the only constraint in the number of ancestors is caused by random

blood relationship between individuals of the same generation. We considered a population of ancestors whose maximum size in each t -generation is given by the geometric progression 2^{t+1} .

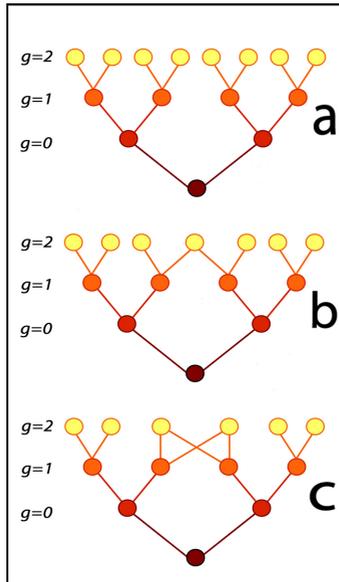


Figura 6.1: Examples of three kinds of genealogical trees (only the first few g -generations) **a**: No restrictions by blood relationship. **b** and **c**: Two kind of restriction in third generation, ancestors sharing one (**b**) or two parents (**c**). The restriction by blood relationship increases according to the degree of endogamy.

We did not consider any restriction for the number of ancestors generated by issues related to culture, in the human case, ethological in the animal case, or isolation of populations, etc. If we want to study the distributions of ancestors of individuals from populations where there are less or equal individuals than 2^{t+1} for the t -generation, there is an additional restriction on the number of ancestors. Blood relationship interconnects the events in the original process that leads to 2^{t+1} , which was generated by independent events and no relation between ancestors of each generation.

Derrida's model [56, 57] is based on numerical simulation under the same assumptions (closed population evolving under sexual reproduction with non-overlapping generations). The population size is fixed for all generations and equally divided into two groups, representing males and females. At every generation, they form random heterosexual pairs and assign them a certain number of descendants according to a Poisson distribution. This is done by choosing for each male or female one pair of parents at random in the previous generation.

In our work the population size is not fixed, but always bigger than 2^{t+1} for each t -generation.

We defined two random variables $y(t)$ and $x(t)$ which represents the number of individuals who are inside and outside to the set of ancestors, with respect to the maximum possible number of ancestors in each t -generation. For this definition we have

$$x(t) + y(t) = 2^{t+1}. \quad (6.1)$$

We considered each generation as a link of the chain which form a first order Markov process [65, 21]. This process is constructed on a given set of individuals ordered by generations. We take the current generation and we count its parents. Then we take all selected individuals and remake the previous question and so forth. There exists a generation in which the question or previous classification makes no more sense, in which case the process ends after a given generation. This kind of process is widely used to describe the evolution of traits that adopt a finite number of states [66].

From equation (6.1) we have

$$y(t) = 2^{t+1} - x(t), \quad (6.2)$$

if $y(t)$ describes a Markov process that implies $x(t)$ describes another Markov process.

We do not distinguish the different kinds of blood relationship between the ancestors of a particular generation such as brot-

hers or cousins, and so forth. We simply consider them as indistinguishable and we just count how many there are. For the purpose of the calculations we consider t as a continuous variable. Finally we associate a discrete-time Markov process to the continuous-time Markov process $\{x(t) : t \geq 0\}$ called a *skeleton process* [67] defined as $\{x(g) : g \geq 0\}$, where g is the generation number.

The time evolution of this process is determined by the knowledge of the probability distribution in each t -generation, denoted by

$$p_n(t) = \mathbb{P}[x(t) = n] \quad (6.3)$$

for all $(n, t) \in \mathbb{S}_t \times \mathbb{R}$, where \mathbb{S}_t is the sample space of $x(t)$ which corresponds to the interval $[0, \mathbf{n}_t]$, $\mathbf{n}_t = \lfloor 2^{t+1} \rfloor - 2$ and $\lfloor z \rfloor$ is the integer part of a real number z .

An equivalent way to describe the process it is through an initial value $p_n(0)$ and the conditional probability given by $\mathcal{P}_{nm}(t, s) = \mathbb{P}[x(t) = n | x(s) = m]$, which represents the transition matrix elements of the states $(m, s) \mapsto (n, t)$.

For each generation the events are mutually exclusive. Consequently at the time $t + \epsilon$ the probability of find n restrictions is given by to the transition from m restrictions at the time t , in this way

$$p_n(t + \epsilon) = \sum_{m \in \mathbb{S}_t} \mathcal{P}_{nm}(t + \epsilon, t) p_m(t). \quad (6.4)$$

After some elementary operations (see appendix section **A1**) we get

$$d_t p_n(t) = \sum_{m \in \mathbb{S}_t} \mathbf{Q}_{nm}(t) p_m(t), \quad (6.5)$$

where d_t denotes the total time derivative $\frac{d}{dt}$, $\mathbf{Q}_{nm}(t)$ is called the *infinitesimal generator* and $\mathbf{Q}_{nm}(t) = \partial_t \mathcal{P}_{nm}(t, s)|_{s=t}$.

We define $\boldsymbol{\varphi}(t)$ as an $|\mathbb{S}_t|$ -tuple of the probability distribution as $\boldsymbol{\varphi}(t) = (p_0(t), p_1(t), \dots, p_{\mathbf{n}_t}(t))^{\top}$, where $|\mathbb{S}_t|$ denotes the cardinal number of \mathbb{S}_t and \top represents the transposition.

The evolution equation for the process can be expressed in a *matrix form* as

$$d_t \boldsymbol{\varphi}(t) = \mathbf{Q}(t) \boldsymbol{\varphi}(t) \quad (6.6)$$

We denote the expectation number of ancestors by $\alpha(t) = \langle y(t) \rangle$ and from the equation (6.2)

$$\alpha(t) = 2^{t+1} - \langle x(t) \rangle \quad (6.7)$$

where $\langle x^k(t) \rangle$ is the expectation value of $x(t)$ raised to the positive integer power k (or k -moment for short) of the distribution $p_n(t)$ and by definition is $\langle x^k(t) \rangle = \sum_n n^k p_n(t)$. The quantity $\langle x(t) \rangle$ represents a constraint caused by blood relationship, which affects the expectation number of ancestors in each generation.

DILUTION OF SAMPLE SPACE VIA GAUGE GROUP

The sample space of $x(t)$ is different for each t -generation, thus there is enormous difficulty to solve the equation (6.6). We considered a *dilution* of \mathbb{S}_t within a larger set $\mathbb{S} \supseteq \mathbb{S}_t$, for all t , consisting of replacing the endpoint \mathbf{n}_t by a huge number N . This dilution can be viewed as a *dilution* represented with the substitution rule $\mathbf{n}_t \longmapsto N$, such that $\mathbb{S} = [0, N]$. On the other hand we know that there exist a certain T -generation that can be considered as the end of the process. The existence of a limit generation, T , allows us to choose $N = \mathbf{n}_T$. Consequently we can solve the problem in this *dilated sample space* and then recover the lost endpoint caused by the dilution through a suitable transformation. The price to pay for it is the need of renormalization of the distribution defined on \mathbb{S} to compensate the dilution effect. The renormalization takes place by a linear transformation which modifies the norm of the distribution for each generation. This *local* transformation (i.e. depends of each t), is structured as a *gauge group*, specifically the *group of local dilations*. Essentially, the distribution defined on \mathbb{S}_t is equivalent to the renormalized distribution which is defined on the dilated sample space \mathbb{S} .

In summary, we can interpret that the process on \mathbb{S}_t is the result of a process on this larger set \mathbb{S} which interacts with another process on the complement of \mathbb{S}_t , denoted by $\mathbb{S} - \mathbb{S}_t$. This interaction is represented by the renormalization of the distribution defined on \mathbb{S} , in an effective theory context. As long as the process on \mathbb{S} becomes much simpler, the description on $\mathbb{S} - \mathbb{S}_t$ will be more complex. This is the basis for the dilation transformation, which is discussed in the appendix sections **A5** and **A6**.

We considered a version in which the sample space \mathbb{S}_t is dilated to the set of natural numbers \mathbb{N} , including the 0 element. Then we have only one boundary condition for the state $n = 0$. This allows us to focus on *time homogeneous processes*, i.e. the infinitesimal generator is independent of t . Another consideration is the *spatial homogeneity*, i.e. the case where the infinitesimal generator does not depend on the state of the random variable $X(t)$.

The Markov process in this larger sample space \mathbb{N} requires to consider two new random variables $\{X, Y\}$ defined on \mathbb{N} and related in a similar way to the old random variables $\{x, y\}$ from (6.2). The associated probability distribution is denoted by $P_n(t) = \mathbb{P}[X(t) = n]$ and defines $\phi(t) = (P_0(t), P_1(t), \dots)^\top$ which satisfies the equation

$$d_t \phi(t) = \mathbf{Q} \phi(t). \quad (6.8)$$

Knowing the initial conditions $\phi(0) = (1, 0, \dots)^\top$ and the infinitesimal generator \mathbf{Q} we can write the formal solution of (6.8) as

$$\phi(t) = \exp(t\mathbf{Q}) \phi(0). \quad (6.9)$$

In order to establish the matrix \mathbf{Q} , we study the time evolution $t \mapsto t + \epsilon$ for small value of ϵ . Therefore, only transitions to the nearest states are allowed, because the infinitesimal time evolution only has a finite variety of transition states. For $n \neq 0$ these transitions are $n \mapsto \{n - 1, n, n + 1\}$ and $n \mapsto \{n, n + 1\}$, for $n = 0$.

Considering this brief discussion, the dynamics described by the equation (6.8) and the imposed conditions represents a time

homogeneous *birth-death process*. A naive way to picture the process in the context of *queueing theory* [68], is through one queue representing all ancestors waiting to be classified if they are blood related or not by one server.

In the appendix section **A4**, we show how to choose a numerical matrix \mathbf{Q} . Finally the evolution equations takes the form

$$d_t P_n(t) = P_{n+1}(t) - 2P_n(t) + P_{n-1}(t), \tag{6.10}$$

$$d_t P_0(t) = P_1(t) - P_0(t),$$

together with the initial condition which is $P_n(0) = \delta_{n0}$, we obtain the explicit solution [68]

$$P_n(t) = e^{-2t} [I_n(2t) + I_{n+1}(2t)] \tag{6.11}$$

where $I_n(x)$ is the modified Bessel function [81]. A brief description to obtain the solution (6.11) is also present in [68]. There is a construction of the generatrix function $g(t, z) = \sum_{n \in \mathbb{N}} P_n(t) z^n$, see (6.22), and from (6.8) derive an equation for $g(t, z)$.

The equation (6.10) is the generic expression for all Markov processes on denumerable sample spaces and continuous time with a particular values of \mathbf{Q} .

THE GAUGED DISTRIBUTION OF ANCESTORS

As we have previously argued, before using this distribution to calculate the moments, it is necessary to perform a renormalization process. The reason is that the solution given by (6.11) is normalized over \mathbb{N} . We perform a gauge transformation [69], denoted by \mathfrak{g}_t , which is applied to the probability distributions as

$$\mathfrak{g}_t : P_n(t) \longrightarrow \lambda(t) P_n(t). \tag{6.12}$$

The transformation (6.12) leaves the evolution equation (6.8) invariant and allows both distributions to describe a Markov pro-

cess. We denote $\mathfrak{p}_n(t) = \lambda(t) P_n(t)$ the gauge transformed distribution of $P_n(t)$. The action of the group \mathfrak{g}_t applied to the distribution $P_n(t)$ leads to a distribution $\mathfrak{p}_n(t)$ defined over \mathbb{S}_t . This idea can be understood in the context of conditional probabilities, with which we can obtain a projection of the distribution on \mathbb{N} into \mathbb{S}_t , keeping the correct normalization.

In other words, the transformation \mathfrak{g}_t leads to a new random variable \mathbf{X} , which is the *gauge transformed* of X .

To preserve the invariance of (6.8) under \mathfrak{g}_t , we introduce a covariant derivative

$$D_t = d_t - \omega(t) \quad (6.13)$$

where $\omega(t) = d_t \lambda(t) [\lambda(t)]^{-1}$. See the appendix for a more extensive explanation.

The expectation value of \mathbf{X} raised to a positive integer power k is $\langle \mathbf{X}^k(t) \rangle = \sum_n n^k \mathfrak{p}_n(t)$. This allows us to write a general relation between $\langle X^k(t) \rangle$ and $\langle \mathbf{X}^k(t) \rangle$

$$\langle \mathbf{X}^k(t) \rangle = \lambda(t) \langle X^k(t) \rangle. \quad (6.14)$$

Rescaling the process described by $X(t)$ and using the solution (6.11) we calculated the first two cumulants

$$\langle \mathbf{X}(t) \rangle = \lambda(t) \langle X(t) \rangle, \quad (6.15)$$

$$\langle [\mathbf{X}(t) - \langle \mathbf{X}(t) \rangle]^2 \rangle = \lambda(t) [2t - \langle X(t) \rangle - \langle X(t) \rangle^2].$$

where

$$\langle X(t) \rangle = e^{-2t} \left[2t I_1(2t) + \left(2t + \frac{1}{2} \right) I_0(2t) \right] - \frac{1}{2}. \quad (6.16)$$

From equation (6.2) the variance of x is equal to the variance of y . The same argument is valid for X and Y .

We define the *standard deviation* of $\mathbf{Y}(t) = 2^{t+1} - \mathbf{X}(t)$, denoted by $\sigma(t)$, as the square root of the second equation of (6.15), which quantifies the statistical fluctuation.

As we consider a constant function $\omega(t)$, then

$$\lambda(t) = 2^{a+b}. \quad (6.17)$$

We have obtained a family of functions for the expectation number of ancestors

$$\alpha(t) = 2^{t+1} - \lambda(t)\langle X(t) \rangle \quad (6.18)$$

parametrized by the real numbers \mathbf{a} and \mathbf{b} of (6.17).

If the expected value satisfies $\alpha(t_1) = \alpha_1$ and $\alpha(t_2) = \alpha_2$, for two generations t_1 and t_2 such that $t_1 \neq 0 \neq t_2$, the parameters \mathbf{a} and \mathbf{b} can be obtained by

$$\mathbf{a} = \frac{1}{t_2 - t_1} \log_2 \left[\frac{2^{t_2+1} - \alpha_2 \langle X(t_1) \rangle}{2^{t_1+1} - \alpha_1 \langle X(t_2) \rangle} \right] \quad (6.19)$$

$$\mathbf{b} = \frac{1}{t_2 - t_1} \left\{ t_2 \log_2 \left[\frac{2^{t_1+1} - \alpha_1}{\langle X(t_1) \rangle} \right] - t_1 \log_2 \left[\frac{2^{t_2+1} - \alpha_2}{\langle X(t_2) \rangle} \right] \right\}$$

where naturally $\alpha_i \leq 2^{t_i+1}$, for $i = 1, 2$, to ensure good definition of \mathbf{a} and \mathbf{b} .

The gauge transformation modulates the amplitude of $\langle X(t) \rangle$. This allows us to define the notion of *horizontal* and *vertical range* of $\alpha(t)$. One important point of the curve $\alpha(t)$ is the *maximum generation range*, this is a nonzero generation T in which α becomes equal to 2. Another interesting point is the maximum of $\alpha(t)$, which determines the *intensity of the process*. Without loss of generality we can choose $t_2 = T$, in which case $\alpha(t_2) = 2$, and $\alpha(t_1) = \sup\{\alpha(t) : t \in [0, T]\}$. For any pair of different points, (t_1, α_1) and (t_2, α_2) , considered relevant, we select one and only one curve of the family, parameterized by \mathbf{a} and \mathbf{b} given by (6.19). The gauge transformation \mathbf{g}_t , through the \mathbf{a} and \mathbf{b} parameters, controls both *horizontal range* and *vertical range* of the process.

This T maybe not be a realistic value, but fix a maximum number of generations of a particular species may have.

For illustrative purposes, in figure 6.2, we have selected 4 curves to the expectation number of ancestors $\alpha(t)$ given by (6.18) and parametrized by different values of $\{a, b\}$. We include a geometric progression 2^{t+1} , which corresponds to the maximum possible number of ancestors in each t -generation.

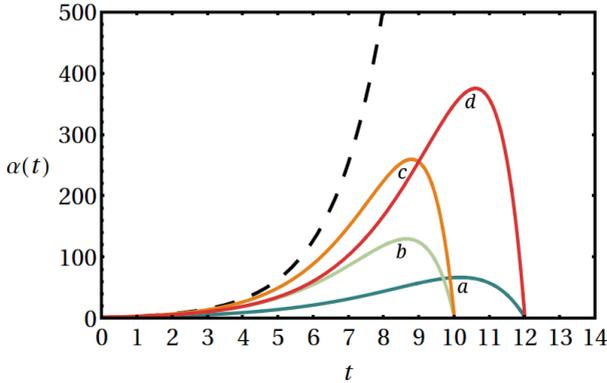


Figure 6.2: The 4 values of $\{a, b\}$ are obtained from (6.19) and parametrized from $(t_1, \alpha_1, t_2, \alpha_2) = (t_1, \xi 2^{t_1+1}, T, 2)$, where $0 \leq \xi \leq 1$. In t_1 the curve reaches the fraction ξ of the total number of possible ancestors for that generation, while t_2 defines the maximum generation range denoted by T . The geometric progression 2^{t+1} is in dash black line. The $\{a, b, c, d\}$ (blue, green, orange, red) lines can be obtained respectively from $(t_1, \alpha_1, t_2, \alpha_2) \in \{(3, 0, 4 \times 2^4, 12, 2), (3, 0, 7 \times 2^4, 10, 2), (3, 0, 9 \times 2^4, 10, 2), (3, 0, 7 \times 2^4, 12, 2)\}$.

Figure 6.3 shows three realizations of the number of ancestors in terms of the expectation value $\alpha(t)$ and a measure of the dispersion given by $\sigma(t)$, for a particular values of a and b .

This model may be employed in order to recognize a possible threshold to identify high endogamic populations as well as its possible causes. Using the genealogical tree, the model can be used to indicate which living species may be near to the extinction.

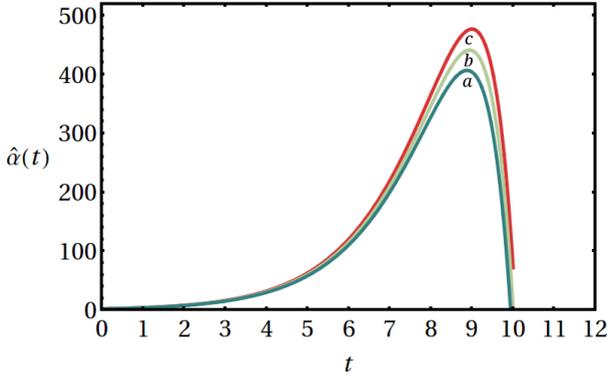


Figura 6.3: A band of curves, defined by the set $\mathcal{B} = \{\hat{\alpha}(t) : \hat{\alpha}(t) \in [\alpha(t) - \sigma(t), \alpha(t) + \sigma(t)]\}$, contains the expectation value $\alpha(t)$ given by the equation (6.18) and its statistical fluctuation at $1-\sigma$. The $\{a,b,c\}$ (blue, green, red) lines corresponds to the realization of $\{\alpha(t) - \sigma(t), \alpha(t), \alpha(t) + \sigma(t)\}$, with the values of $(t_1, \alpha_1, t_2, \alpha_2) = (6, 0,9 \times 2^7, 10, 2)$, respectively.

FINAL COMMENTS AND POSSIBLE MODEL EXTENSIONS

The model explained above allows to calculate the expectation number of ancestors in each generation, considering the possibility of blood relationship between individual of the same generation and a population of ancestors which maximum size is 2^{t+1} . But there are two possible generalizations. The model can be extended to take into account relationships between individuals of other adjacent generations using a similar idea, simply considering higher order Markov chains. By introducing the corresponding terms in the infinitesimal generator \mathbf{Q} , for example an absorbing barrier [24], the same model can be used to calculate the expectation number of ancestors in a specific population with additional restrictions such as isolation, immigration, specific reproductive behavior, or cultural restrictions for the human case.

In future applications we could generalize the model through

a new constraint to fix the maximum number of individuals at certain generation. This proposal implies a generalization of this work in which the maximum number of ancestors will be given by a piecewise function $\gamma(t)$ instead of 2^{t+1} . This leads to slightly modify the process defined at the beginning in (6.2) as $y(t) = \gamma(t) - x(t)$ and the endpoint of the sample space for $x(t)$ becomes $\mathbf{n}_t = \gamma(t) - 2$. This generalization includes a time inhomogeneity in the infinitesimal generator \mathbf{Q} and preserves an appropriate renormalization.

We can include these possible extensions using the process $\{X(t)\}$ and study a most general gauge \mathbf{g}_t transform given by a linear transformation

$$\mathbf{g}_t : \{P_m(t)\}_{m \in \mathbb{N}} \longrightarrow \mathbf{p}_n(t) = \sum_{m \in \mathbb{N}} \lambda_{nm}(t) P_m(t) \quad (6.20)$$

where $\boldsymbol{\lambda} = \{\lambda_{nm}\}$ is a non singular matrix. We can express the last expression in a matrix form

$$\mathbf{g}_t : \boldsymbol{\phi} \longrightarrow \boldsymbol{\phi}' = \boldsymbol{\lambda} \boldsymbol{\phi} \quad (6.21)$$

where $\boldsymbol{\phi}'(t) = (\mathbf{p}_0(t), \mathbf{p}_1(t), \dots)^\top$.

In order to preserve the invariance of (6.8) under this generalization of \mathbf{g}_t , we introduce the corresponding covariant derivative

$$D_t = d_t - \boldsymbol{\omega}(t) \quad (6.22)$$

where $\boldsymbol{\omega}(t) = d_t \boldsymbol{\lambda}(t) [\boldsymbol{\lambda}(t)]^{-1}$.

The evolution equation for $\boldsymbol{\phi}'(t)$ is also invariant under the local dilation group.

$$D_t \boldsymbol{\phi}'(t) = \mathbf{Q}'(t) \boldsymbol{\phi}'(t) \quad (6.23)$$

and the gauged infinitesimal generator is now

$$\mathbf{Q}'(t) = \boldsymbol{\lambda}(t) \mathbf{Q} [\boldsymbol{\lambda}(t)]^{-1} \quad (6.24)$$

which corresponds to a similarity transformation of \mathbf{Q} .

This model itself can be applied to describe other biological or physical systems with similar dynamics. Statistical models of biparental reproduction have already been compared with physical systems before, such as, spin-glass systems [70]. In this regard the *evolutionary graph theory* is an approach to study how topology affects the evolution of a population [71].

Other analogous processes to the biparental reproduction in physics are described with similar statistical or markovian models [65]. In high energy physics the production of a cascade by a cosmic ray is described by the Heitler model [72]. Although this model is different from the one presented here we could compare the number of ancestors with the number of particles in each generation and reinterpret this results in terms of these kinds of phenomena.

It is possible to estimate the maximum generation range, T , searching in the fossil record the first time that a particular species appears and use its reproductive rate. In this way we are classifying each species not in terms of life time on earth (time units), but according to the notions of generational patterns.

The interaction of various of these processes can be combined with universal common ancestor's models [73] to understand the development of a certain species. The ideas in the current model can be used in biology, population ecology and genetics. An important achievement of the model is that based on the previous knowledge of the life time of a certain species, we can calculate the number of ancestors in each generation of this species.

More fundamental uses of this ideas can be found in mathematics, in particular in theory of stochastic processes and physics area connected with the theory of stochastic processes. Future research through a Lagrangian description may find novel applications of the present proposal. In this case we will consider the probabilities $\{p_n(t)\}$ as the set of generalized coordinates.

ACKNOWLEDGMENTS

We thank our respective PhD advisors: Fernando Cornet and María Teresa Dova, Hernán Wahlberg. Also we thank to Carlos García Canal and Huner Fanchiotti for their helpful criticism. We are indebted to Federico Agnolin and Ben Page for reading the manuscript and providing advice. Recall also our anonymous readers and reviewers for their contribution to this work. And finally a special mention to Micaela Moretton, María Clara Caruso and Gabriel Lio for local support.

APPENDIX

A1. ON THE EVOLUTION EQUATION IN \mathbb{S}

We defined the random variable $y(t)$ associated with the number of ancestors as

$$y(t) = 2^{t+1} - x(t) \quad (6.1)$$

where \mathbb{S}_t is the sample space of $x(t)$ and t is a continuous variable. Then, according to a discretization process, the distribution of ancestors is obtained and the variable t will be the number of generations.

We denoted by $\alpha(t)$ the expectation number of $y(t)$ as

$$\alpha(t) = 2^{t+1} - \langle x(t) \rangle \quad (6.2)$$

where $\langle x(t) \rangle$ constrains the number of ancestors under the blood relationship hypothesis. We will focus on the process of this random variable $x(t)$ and our probability distribution denoted by $p_n(t) = \mathbb{P}[x(t) = n]$.

In the problem of counting ancestors the sample space \mathbb{S}_t is different for each t -generation. For this reason we introduced a

fictitious process on a larger sample space \mathbb{S} than the original \mathbb{S}_t . Therefore we have solved the problem in \mathbb{S} and renormalized the distribution to take into account of the interaction with the lost boundary of \mathbb{S}_t . In other words, we have considered a *dilution* of \mathbb{S}_t into \mathbb{S} . In particular we used the set of natural numbers $\mathbb{S} = \mathbb{N}$, including the 0 element.

We considered the study of the evolution over a generic denumerable sample space $\mathbb{S} = [0, N] \subset \mathbb{N}$ with the random variable $X(t)$ and probability distribution $P_n(t) = \mathbb{P}[X(t) = n]$. This evolution is governed by the conditional probability given by

$$\mathcal{P}_{nm}(t, s) = \mathbb{P}[X(t) = n | X(s) = m] \quad (6.3)$$

The matrix $\mathcal{P}(t, s)$ satisfies the *Chapman-Kolmogorov* equation [21, 24]

$$\mathcal{P}(t, s) = \mathcal{P}(t, u)\mathcal{P}(u, s) \quad (6.4)$$

for $0 \leq s \leq u \leq t$. Also the sum of the elements of each column is

$$\sum_{n \in \mathbb{S}} \mathcal{P}_{nm}(t, s) = 1. \quad (6.5)$$

For the general case we develop a power series of the matrix $\mathcal{P}(t + \epsilon, s)$, for a fixed value of s , we have

$$\mathcal{P}_{nm}(t + \epsilon, s) = \mathcal{P}_{nm}(t, s) + \epsilon \partial_t \mathcal{P}_{nm}(t, s) + \dots \quad (6.6)$$

where ∂_t is a simplified notation of partial time derivative $\frac{\partial}{\partial t}$

In order to obtain the equation (6.4) we study the time evolution $t \mapsto t + \epsilon$, for small value of ϵ . We need to know $\mathcal{P}_{nm}(t + \epsilon, t)$ then (6.6) becomes

$$\mathcal{P}_{nm}(t + \epsilon, t) = \delta_{nm} + \epsilon \partial_t \mathcal{P}_{nm}(t, s)|_{s=t} + \dots \quad (6.7)$$

we recognize the second term of (6.8) as the infinitesimal generator $Q_{nm}(t)$

$$Q_{nm}(t) = \lim_{\epsilon \rightarrow 0} \frac{\mathcal{P}_{nm}(t + \epsilon, t) - \delta_{nm}}{\epsilon} \quad (6.8)$$

We assumed that t is continuous. This allows us to evaluate the process at any t between two generations, but it also reduces the number of possible states in an infinitesimal evolution. Therefore only transitions to the nearest states are allowed. For the ancestry problem, the infinitesimal time evolution has a finite number of transition states. These transitions are denoted by $n \mapsto n'$, where $n' \in T_n$, i.e. n' depends on the initial state n . We write explicitly $T_0 = \{0, 1\}$, $T_N = \{N - 1, N\}$ and for $n \neq 0, N$: $T_n = \{n - 1, n, n + 1\}$. Furthermore, if $|n' - n| > 1$ the corresponding transition probability is zero.

From equation (6.5) the matrix $\mathcal{P}(t + \epsilon, t)$ is normalized for all t and small ϵ as

$$\sum_{n' \in T_n} \mathcal{P}_{n'n}(t + \epsilon, t) = 1 \quad (6.9)$$

note that n' runs over T_n , depending on whether n is equal to 0, N or any other value of $\mathbb{S} - \{0, N\}$.

We express $\mathcal{P}_{n'n}(t + \epsilon, t)$ for these three cases from (6.8)

$$\begin{aligned} \mathcal{P}_{n-1n}(t + \epsilon, t) &= \mu_n(t)\epsilon + \mathcal{O}_t(\epsilon), \\ \mathcal{P}_{n+1n}(t + \epsilon, t) &= \nu_n(t)\epsilon + \mathcal{O}_t(\epsilon), \\ \mathcal{P}_{n'n}(t + \epsilon, t) &= 0, \quad |n' - n| > 1, \\ \mathcal{P}_{nn}(t + \epsilon, t) &= 1 - [\nu_n(t) + \mu_n(t)]\epsilon + \mathcal{O}_t(\epsilon), \end{aligned} \quad (6.10)$$

where $\mathcal{O}_t(x)$ represents a type of function that goes to zero with x faster than x , for a given t , that is

$$\lim_{x \rightarrow 0} \frac{\mathcal{O}_t(x)}{x} = 0. \quad (6.11)$$

The fourth equation of (6.10) is obtained through the first three of them. In the general case we proceed as follows from (6.5)

$$\mathcal{P}_{nn}(t, s) = 1 - \sum_{n' \in \mathbb{S} - \{n\}} \mathcal{P}_{n'n}(t, s) \quad (6.12)$$

then from an infinitesimal time evolution $s \equiv t \mapsto t + \epsilon$ and (6.9)

$$\begin{aligned} \mathcal{P}_{nn}(t + \epsilon, t) &= 1 - \sum_{n' \in T_n - \{n\}} \mathcal{P}_{n'n}(t + \epsilon, t) \\ \mathcal{P}_{nn}(t + \epsilon, t) &= 1 - \mathcal{P}_{n+1n}(t + \epsilon, t) - \mathcal{P}_{n-1n}(t + \epsilon, t) \end{aligned}$$

which is the fourth equation of (6.10).

Replacing (6.10) in (6.4) and written for $n \neq 0$, we have

$$\begin{aligned} P_n(t + \epsilon) &= [\nu_{n-1}(t)\epsilon + \mathcal{O}_t(\epsilon)] P_{n-1}(t) \\ &\quad + [\mu_{n+1}(t)\epsilon + \mathcal{O}_t(\epsilon)] P_{n+1}(t) \\ &\quad + [1 - \nu_n(t)\epsilon - \mu_n(t)\epsilon + \mathcal{O}_t(\epsilon)] P_n(t) \end{aligned}$$

then

$$\begin{aligned} \frac{P_n(t + \epsilon) - P_n(t)}{\epsilon} &= \left[\nu_{n-1}(t) + \frac{\mathcal{O}_t(\epsilon)}{\epsilon} \right] P_{n-1}(t) \\ &\quad + \left[\mu_{n+1}(t) + \frac{\mathcal{O}_t(\epsilon)}{\epsilon} \right] P_{n+1}(t) \\ &\quad - \left[\nu_n(t) + \mu_n(t) + \frac{\mathcal{O}_t(\epsilon)}{\epsilon} \right] P_n(t) \end{aligned}$$

taking the limit $\epsilon \rightarrow 0$

$$\begin{aligned} d_t P_n(t) &= \nu_{n-1}(t) P_{n-1}(t) + \mu_{n+1}(t) P_{n+1}(t) \\ &\quad - [\nu_n(t) + \mu_n(t)] P_n(t) \end{aligned} \tag{6.13}$$

The stochastic process described by equation (6.13) corresponds to the general class of stochastic dynamics called *birth and death process*, which includes the *queueing process* [21, 68].

The functions $\mu_n(t)$ and $\nu_n(t)$ are part of the infinitesimal generator $\mathbf{Q}(t)$. From the first equation of (6.10) we have

$$\mu_n(t) = \lim_{\epsilon \rightarrow 0} \frac{\mathcal{P}_{n-1n}(t + \epsilon, t)}{\epsilon} \tag{6.14}$$

which is exactly the element $Q_{n-1n}(t)$ of (6.8).

In summary we list all the elements of $\mathbf{Q}(t)$

$$\mathbf{Q}_{nn}(t) = -[\nu_n(t) + \mu_n(t)], \quad \mathbf{Q}_{n-1n}(t) = \mu_n(t) \quad (6.15)$$

$$\mathbf{Q}_{n+1n}(t) = \nu_n(t), \quad \mathbf{Q}_{n'n}(t) = 0, \quad |n' - n| > 1$$

We write (6.13) in a matrix form as

$$d_t \boldsymbol{\phi}(t) = \mathbf{Q}(t) \boldsymbol{\phi}(t) \quad (6.16)$$

where $\boldsymbol{\phi}(t) = (P_0(t), \dots, P_N(t))^\top$ and $\mathbf{Q}(t) = \{\mathbf{Q}_{n'n}(t)\}$ is given by (6.15).

It should also be pointed out that the coefficients μ_0 and ν_N must be zero, otherwise we require more states than $[0, N]$ in \mathbb{S} . In other words, if μ_0 or ν_N are not equal to zero the left side of (6.9) is not equal to one.

A2. HOMOEGENEOUS HYPOTHESIS

In this section we show how the hypothesis of spatial and temporal homogeneity are used working in the dilated sample space.

We have already said that if there exists a certain T -generation that can be considered as the stop of the process, the upper limit of the dilated space \mathbb{S} , denoted by N , can be chosen as

$$N = \mathbf{n}_T, \quad (6.17)$$

where $\mathbf{n}_T = \sup\{n : n \in \mathbb{S}_t, \forall t \in [0, T]\}$.

Also we considered a dilution of \mathbb{S}_t into \mathbb{N} , i.e. the generic dilated space \mathbb{S} is equal to \mathbb{N} , or $N \rightarrow \infty$. This assumption is true from $\mathbb{S}_t \subseteq \mathbb{N}$, no matter how big is \mathbf{n}_T . In this case the space-time on the process is infinite and we have an infinitesimal generator on \mathbb{N} independent of the state of the random variable X and time-independent, and the process is *space-time homogeneous*.

However the effect of these hypotheses can be compensated with an interaction with the process on $\mathbb{N} - \mathbb{S}_t$, see the section A4 and for more details.

Essentially we will say that the renormalized distribution defined on \mathbb{N} is equivalent to the distribution defined on \mathbb{S}_t . This equivalence is based on the invariance of evolution equation. In this way both distribution corresponds to a Markov process, see section A4.

A3. ON THE SPACE-TIME HOMOGENEOUS SOLUTION IN \mathbb{N}

The evolution equation in the dilated sample space \mathbb{N} under the space-time homogeneity is

$$d_t \phi(t) = \mathbf{Q} \phi(t), \quad (6.18)$$

with $\phi(t) = (P_0(t), P_1(t), \dots)$ and identify the matrix \mathbf{Q} as

$$\mathbf{Q} = \begin{pmatrix} -\nu & \mu & 0 & \dots \\ \nu & -\mu - \nu & \mu & \dots \\ 0 & \nu & -\mu - \nu & \dots \\ 0 & 0 & \nu & \dots \\ 0 & 0 & 0 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \quad (6.19)$$

Under the initial condition $\phi(0) = (1, 0, \dots)^T$, the solution of (6.18) with (6.19) is

$$P_n(t) = e^{-(\nu+\mu)t} \left[\rho^{n/2} I_n(\zeta t) + \rho^{(n-1)/2} I_{n+1}(\zeta t) + (1 - \rho) \rho^n \sum_{j=n+2}^{\infty} \rho^{-j/2} I_j(\zeta t) \right]. \quad (6.20)$$

where $I_n(x)$ is the modified Bessel function [81], $\rho = \nu/\mu$ and $\zeta = 2\sqrt{\nu\mu}$. The first solution of (6.18) appeared in the 1950's, see [74, 75, 76, 77]. A description to obtain the solution (6.20) is also presented in [68].

A4. MOMENTS OF THE DISTRIBUTION

In this section we calculate the first two cumulants of the distribution obtained in (6.20).

First of all we demonstrated the existence of all k -moment of the distribution $P_n(t)$ defined by

$$\langle X^k(t) \rangle = \sum_{n \in \mathbb{N}} n^k P_n(t) \quad (6.21)$$

with $k \in \mathbb{N}$. It is possible to demonstrate that all series defined above converge uniformly $\forall t$. To demonstrate this, we define the *generatrix function*

$$g(t, z) = \sum_{n \in \mathbb{N}} P_n(t) z^n \quad (6.22)$$

for $z \in \mathbb{R}$. If it converges, $g(t, z)$ is well defined. For our case, we know that

$$\sum_{n \in \mathbb{N}} I_n(x) \longrightarrow \frac{1}{2}[e^x + I_0(x)], \quad (6.23)$$

converges uniformly [81]. This allows us to write the distribution's norm and demonstrate that converge uniformly

$$\sum_{n \in \mathbb{N}} P_n(t) \longrightarrow 1. \quad (6.24)$$

Another argument for the general birth-death process, based on the nature of the coefficients $\{(\nu_n, \mu_{n+1}) : n \in \mathbb{N}\}$ leads us to the same conclusion [24].

For each t , we demonstrated that $g(t, 1) \longrightarrow 1$ uniformly, then for the Abel's theorem [23]

$$\sum_{n \in \mathbb{N}} P_n(t) z^n \longrightarrow g(t, z) \quad (6.25)$$

uniformly for each $z \in [0, 1]$.

Therefore the equation (6.22) can be derived term by term keeping the uniform convergence $\forall(t, k)$

$$\sum_{n \in \mathbb{N}} \frac{n!}{(n-k)!} P_n(t) \longrightarrow \partial_{z^k} g(t, z)|_{z=1}. \quad (6.26)$$

Finally $\langle X^k(t) \rangle$ can be obtained as a combination of the series given by (6.26), which converges to $\partial_{z^k} g(t, z)|_{z=1}$, and this completes the demonstration.

We can use two methods for the calculation of the first two moments, given the solution (6.20).

The first one is by definition (6.21) and using the identity [81]

$$nI_n(x) = \frac{x}{2}[I_{n-1}(x) - I_{n+1}(x)]. \quad (6.27)$$

The second one is based on the series (6.21) converges uniformly. We can derive term by term the series (6.21). Using the evolution equation (6.18) with (6.19), we can obtain a differential equation for the expectation value $\langle X(t) \rangle$ and $\langle X^2(t) \rangle$

$$d_t \langle X(t) \rangle = (\nu - \mu) + \mu P_0(t), \quad (6.28)$$

$$d_t \langle X^2(t) \rangle = 2\nu - d_t \langle X(t) \rangle + 2(\nu - \mu) \langle X(t) \rangle. \quad (6.29)$$

For the initial condition $P_n(0) = \delta_{n0}$ then $\langle X^k(0) \rangle = 0$ and integrate the last two equations

$$\langle X(t) \rangle = (\nu - \mu)t + \mu \int_0^t P_0(\tau) d\tau, \quad (6.30)$$

$$\langle X^2(t) \rangle = 2\nu t - \langle X(t) \rangle + 2(\nu - \mu) \int_0^t \langle X(\tau) \rangle d\tau. \quad (6.31)$$

The expression (6.30) shows that to determine $\langle X(t) \rangle$ is sufficient to know the distribution of probability of no blood relationship $P_0(t)$ from (6.20). Also the expression (6.31) shows that to determine $\langle X^2(t) \rangle$ is sufficient to know $\langle X(t) \rangle$.

This two methods arrives at the same result for $\langle X(t) \rangle$.

We are interested in computing the first two cumulants. The first one is the expectation value and the second one is defined as function of the first two moments by $\langle [X(t) - \langle X(t) \rangle]^2 \rangle = \langle X^2(t) \rangle - \langle X(t) \rangle^2$ respectively.

We have said that this is due to dilation of the sample space $\mathbb{S}_t \mapsto \mathbb{N}$. This implies that distributions defined on \mathbb{N} takes smaller values than they should take on \mathbb{S}_t .

Also we show how $\langle X(t) \rangle$ is small compared to 2^{t+1} , for any value of $\nu \geq 0$ and $\mu \geq 0$.

Using (6.30) and $0 \leq P_0(t) \leq 1$ we can express

$$|\langle X(t) \rangle| \leq |\nu - \mu|t + \mu \int_0^t P_0(\tau) d\tau \leq |\nu - \mu|t + \mu t. \quad (6.32)$$

This shows that $\langle X(t) \rangle$ is subordinated to a linear function in t , for all μ, ν , then the expected value of ancestors $\langle Y(t) \rangle = 2^{t+1} - \langle X(t) \rangle$ it will grow indefinitely with t as 2^{t+1} . We can consider by ignorance $\nu = \mu$, in the sense of not knowing the functional form of the trend of the number of ancestors. Although the knowledge of any particular trend can be introduced in the gauge transformation. And finally without loss of generality we can take $\mu = 1 = \nu$, since the problem of non saturation will be solved by dilate the distribution to compensate for the dilution of the sample space, as we shall see in the next section. This dilation can be seen as a renormalization. In the first place we considered the case where $X(t)$ is defined on \mathbb{N} with no renormalization at all. In this way, the renormalization is interpreted as an operation where the correct scale of the interaction is retrieved modifying $\langle X(t) \rangle$, as if we had solved the problem in the original sample space \mathbb{S}_t .

A5. ON THE INTERACTION WITH A FICTICIOUS ENVIROMENT

We analyzed one of the main concepts: the dilution of the sample space involves the study of an interaction between the

initial sample space with the fictitious environment. The introduction of the dilution takes into account the interaction with the lost boundary of \mathbb{S}_t , under the condition of \mathbb{S} is large enough to include \mathbb{S}_t for all t .

We can view the process on \mathbb{S}_t as the result of a process on \mathbb{S} which interact with another process defined on the complement set $\mathbb{R}_t = \mathbb{S} - \mathbb{S}_t$. This point of view can be described in a mathematical precise sense. We define the associated vector spaces $\{S, S_t, R_t\}$ to the sets $\{\mathbb{S}, \mathbb{S}_t, \mathbb{R}_t\}$, where ϕ is a vector in S , which $\dim(S) = N+1$. If the sample space \mathbb{S}_t has $n_t + 1$ number of states, we define φ as the first $n_t + 1$ component of ϕ , i.e. φ is vector of S_t . We expressed these vectors in a canonical basis $\{\mathbf{e}_n\}_{n=0, \dots, N}$ such that

$$\phi(t) = \sum_{n \in \mathbb{S}} P_n(t) \mathbf{e}_n \quad (6.33)$$

$$\phi(t) = \sum_{n \in \mathbb{S}_t} P_n(t) \mathbf{e}_n + \sum_{n \in \mathbb{R}_t} P_n(t) \mathbf{e}_n$$

then we have

$$\phi(t) = \varphi(t) + \psi(t) \quad (6.34)$$

where $\mathbf{e}_0 = (1, 0, \dots, 0)^\top$, $\mathbf{e}_1 = (0, 1, \dots, 0)^\top$ and so forth. For construction S_t is orthogonal to R_t . The dimensions of these vector spaces are determined by $\dim(S) = N + 1$ and $\dim(S_t) = n_t + 1$.

From the equation (6.34) we see, roughly speaking, that the process on \mathbb{S}_t is the result of interaction between the process on \mathbb{S} and \mathbb{R}_t through

$$\varphi(t) = \phi(t) - \psi(t) \quad (6.35)$$

We can write down the evolution equation (6.16) in the form

$$d_t \varphi(t) = \mathbf{Q}_{ss}(t) \varphi(t) + \mathbf{Q}_{sr}(t) \psi(t) \quad (6.36)$$

$$d_t \psi(t) = \mathbf{Q}_{rs}(t) \varphi(t) + \mathbf{Q}_{rr}(t) \psi(t) \quad (6.37)$$

where \mathbf{Q}_{ab} is the $a \times b$ block matrix of \mathbf{Q} , for $a, b \in \{s, r\}$, $s = \mathbf{n}_t + 1$ and $r = N - \mathbf{n}_t$, explicitly

$$\mathbf{Q} = \begin{pmatrix} \mathbf{Q}_{ss} & \mathbf{Q}_{sr} \\ \mathbf{Q}_{rs} & \mathbf{Q}_{rr} \end{pmatrix} \quad (6.38)$$

We see that the equations for $\boldsymbol{\varphi}$ and $\boldsymbol{\psi}$ are coupled, hence the interaction character which was noted above. To be more specific, we write a relation of two *partial solutions* $\boldsymbol{\psi}$ and $\boldsymbol{\varphi}$

$$\boldsymbol{\psi}(t) = \int_{\mathbb{R}} \mathbf{K}_{rs}(t, t') \boldsymbol{\varphi}(t') dt' \quad (6.39)$$

where (6.39) satisfies (6.37) and $\mathbf{K}_{rs}(t, t')$ is the kernel of this transformation defined as

$$\mathbf{K}_{rs}(t, t') = \mathcal{T} \left\{ \exp \left[\int_{t'}^t \mathbf{Q}_{rr}(\tau) d\tau \right] \right\} \mathbf{Q}_{rs}(t') \theta(t - t') \quad (6.40)$$

and $\theta(t)$ are the *unit step* distribution and \mathcal{T} is the time-ordered operator defined as

$$\mathcal{T}[\mathbf{A}(t)\mathbf{A}(u)] = \begin{cases} \mathbf{A}(t)\mathbf{A}(u) & : t > u \\ \mathbf{A}(u)\mathbf{A}(t) & : u > t \end{cases} \quad (6.41)$$

We can obtain another expression for equation (6.36)

$$d_t \boldsymbol{\varphi}(t) = \int_{\mathbb{R}} \mathbf{K}_{ss}(t, t') \boldsymbol{\varphi}(t') dt' \quad (6.42)$$

where \mathbf{K}_{ss} is the kernel of the integro-differential equation (6.42) defined as

$$\mathbf{K}_{ss}(t, t') = \mathbf{Q}_{ss}(t') \delta(t - t') + \mathbf{Q}_{sr}(t) \mathbf{K}_{rs}(t, t') \quad (6.43)$$

and $\delta(t)$ is the *delta* distribution.

The same argument allows us to obtain an inverse relation of (6.39)

$$\boldsymbol{\varphi}(t) = \int_{\mathbb{R}} \mathbf{K}_{sr}(t, t') \boldsymbol{\psi}(t') dt' \quad (6.44)$$

simply interchange in (6.40), (6.42) and (6.43) the quantities $\varphi \longleftrightarrow \psi$, $r \longleftrightarrow s$.

The mathematical construction presented here shows how a process in \mathbb{S} can be described by the interaction of two sub-processes in \mathbb{S}_t and \mathbb{R}_t . Specifically this interaction can be viewed in the relation (6.39) or its inverse (6.44).

In other words, the process on \mathbb{S}_t can be seen as the interaction between the processes on \mathbb{S} and $\mathbb{R}_t = \mathbb{S} - \mathbb{S}_t$, this interaction emerges from the elements of infinitesimal transition probabilities present in $\mathbf{K}_{ss}(t, t')$ through (6.43) and (6.40).

A6. DILATION-DILUTION TRANSFORMATION

The *dilution* operation of \mathbb{S}_t into a larger sample space \mathbb{S} , can also be understood as an *dilation* represented by the substitution rule $\mathbf{n}_t \mapsto N$. In a general sense, every *dilation-dilution* transformation involves a renormalization of the distribution obtained above.

To illustrate this point let us consider two distributions $\{p_n : n \in \mathfrak{s}\}$ and $\{P_n : n \in \mathbb{S}\}$ defined over the sample spaces \mathfrak{s} and \mathbb{S} respectively, such that $\mathfrak{s} \subset \mathbb{S}$ and both are normalized in each samples spaces. By definition we have

$$\sum_{n \in \mathbb{S}} P_n = \sum_{n \in \mathbb{S} - \mathfrak{s}} P_n + \sum_{n \in \mathfrak{s}} P_n = 1. \quad (6.45)$$

Since all terms of (6.45) are positive, there exist a subset of \mathfrak{s} in which the distribution P_n , restricted to \mathfrak{s} , is smaller than $\{p_n : n \in \mathfrak{s}\}$. We can see it in this from

$$\sum_{n \in \mathfrak{s}} P_n < \sum_{n \in \mathfrak{s}} p_n \quad (6.46)$$

We want to describe the process $\{p_n : n \in \mathfrak{s}\}$ through the process of $\{P_n : n \in \mathbb{S}\}$ *projecting* the distribution P_n over \mathfrak{s} . Since there are values of \mathfrak{s} for which P_n is less than p_n , that projection

should amplify P_n to improve the stated description. This amplification corresponds to a dilation transformation.

We had mentioned that the renormalization is performed through a linear time-dependent transformation. As we conserve the linearity of the equation (6.16) and since the original sample space \mathbb{S}_t is time-dependent, the norm must be corrected locally. This *local* transformation is structured as a *gauge group*, specifically a *group of local dilations*.

On the other hand, this transformation can be viewed through a *projection* of the distribution $P_n(t)$ on \mathbb{S}_t , which filter a subset of states and renormalizes the distribution. We used the concept of conditional probability to relate and motivate the renormalization through a gauge transformation. To be more clear, we distinguish the gauge transformed distribution $\mathfrak{p}_n(t)$ and the projected distribution $q_n(t)$.

We used a dilation on a generic sample space $\mathbb{S} = [0, N]$, for a natural number N .

Lets consider the following example. If $q_n(t)$ is defined as

$$q_n(t) = \mathbb{P}[X(t) = n | X(t) \in \mathbb{S}_t] \quad (6.47)$$

this distribution is equal to zero for all state such that $n \in \mathbb{S} - \mathbb{S}_t$. Applying the Bayes identity [24] we have

$$\mathbb{P}[A|B]\mathbb{P}[B] = \mathbb{P}[B|A]\mathbb{P}[A] \quad (6.48)$$

we can express $q_n(t)$ as

$$q_n(t) = \Lambda(t)P_n(t), \quad (6.49)$$

where

$$\Lambda(t) = \frac{\mathbb{P}[X(t) \in \mathbb{S}_t | X(t) = n]}{\mathbb{P}[X(t) \in \mathbb{S}_t]}. \quad (6.50)$$

The numerator of (6.50) is only equal to 0 or 1 depending on whether $n \in \mathbb{S} - \mathbb{S}_t$ or $n \in \mathbb{S}_t$, respectively. The denominator of (6.50) is $\mathbb{P}[X(t) \in \mathbb{S}_t] = \mathbb{P}[X(t) \leq \mathfrak{n}_t]$ and for the mutually exclusive events we have

$$\mathbb{P}[X(t) \in \mathbb{S}_t] = \sum_{n \in \mathbb{S}_t} P_n(t) < 1 \quad (6.51)$$

this implies $q_n(t) \geq P_n(t)$, for $n \in \mathbb{S}_t$. The expression (6.49) involves a probability fraction which can be viewed as a projection operator which transforms distributions defined on \mathbb{S} into distributions defined on \mathbb{S}_t , by filtering selected states and redefining the correct normalization.

The last discussion about the conditional probability as a projection operation motivates the following construction. Given a process in \mathbb{S} described by $\boldsymbol{\phi}(t) = (P_0(t), \dots, P_N(t))^T$ which satisfies a general equation, similar to (6.16)

$$d_t \boldsymbol{\phi}(t) = \mathbf{Q}(t) \boldsymbol{\phi}(t) \quad (6.52)$$

and the dilation transformation

$$\mathbf{g}_t : P_n(t) \longrightarrow \mathfrak{p}_n(t), \quad (6.53)$$

where $\mathfrak{p}_n(t) = \lambda(t)P_n(t)$ are the components of the column vector $\boldsymbol{\phi}'(t)$. We showed that the covariance requirement is satisfied if \mathbf{g}_t is a *gauge transformation*, this implies the addition of a covariant derivative. To prove this statement we transform a general evolution equation (6.52) through a local dilation group \mathbf{g}_t

$$d_t \boldsymbol{\phi}'(t) - d_t \lambda(t) [\lambda(t)]^{-1} \boldsymbol{\phi}'(t) = \mathbf{Q}(t) \boldsymbol{\phi}'(t). \quad (6.54)$$

In order to keep the shape of (6.16) when we transformed by \mathbf{g}_t , we need to introduce an affine connection which can be implemented through a covariant derivative given by

$$D_t = d_t - \omega(t) \quad (6.55)$$

where $\omega(t) = d_t \lambda(t) [\lambda(t)]^{-1}$ is the *gauge function*.

In this case, the evolution equation for $\boldsymbol{\phi}'(t)$ is

$$D_t \boldsymbol{\phi}'(t) = \mathbf{Q}(t) \boldsymbol{\phi}'(t) \quad (6.56)$$

which is invariant under the local dilation group.

The definition (6.55) allows the distribution $P_n(t)$ and $\mathfrak{p}_n(t)$, related by gauged corresponds to a Markov process.

For the case of this work, the gauge function $\omega(t)$ is independent of the t -generation, then $\lambda(t) = 2^{\mathbf{a}t+\mathbf{b}}$, for \mathbf{a}, \mathbf{b} are real constants. Otherwise, the time dependence of $\omega(t)$ introduces a time inhomogeneity in the process.

The example used to motivate the definition (6.53) can be extended to describe a rich variety of cases. If we generalized $q_n(t)$ as

$$q_n(t) = \mathbb{P}[X(t) = n | X(t) \in \mathbb{S}_t, Z], \quad (6.57)$$

where Z is an extra condition and eventually space-time dependent. The distribution (6.57) is equal to zero for $n \in \mathbb{S} - \mathbb{S}_t$. Then for the Bayes identity (6.48) applied to (6.57)

$$q_n(t) = \Lambda_n(t)P_n(t) \quad (6.58)$$

where

$$\Lambda_n(t) = \frac{\mathbb{P}[X(t) \in \mathbb{S}_t, Z | X(t) = n]}{\mathbb{P}[X(t) \in \mathbb{S}_t, Z]}. \quad (6.59)$$

In a more abstract sense, we consider a partition of $\mathbb{S} = \{B_n : n \in \mathbb{S}\}$ and let us consider the event A_n write in the following way:

$$A_n = \bigcup_{m \in \mathbb{S}} A_n \cap B_m. \quad (6.60)$$

Therefore we use the law of total probability

$$\mathbb{P}[A_n] = \sum_{m \in \mathbb{S}} \mathbb{P}[A_n | B_m] \mathbb{P}[B_m] \quad (6.61)$$

for a particular case of the partition $B_m = \{m\}$ and $\mathbb{P}[A_n] = q_n(t)$, then

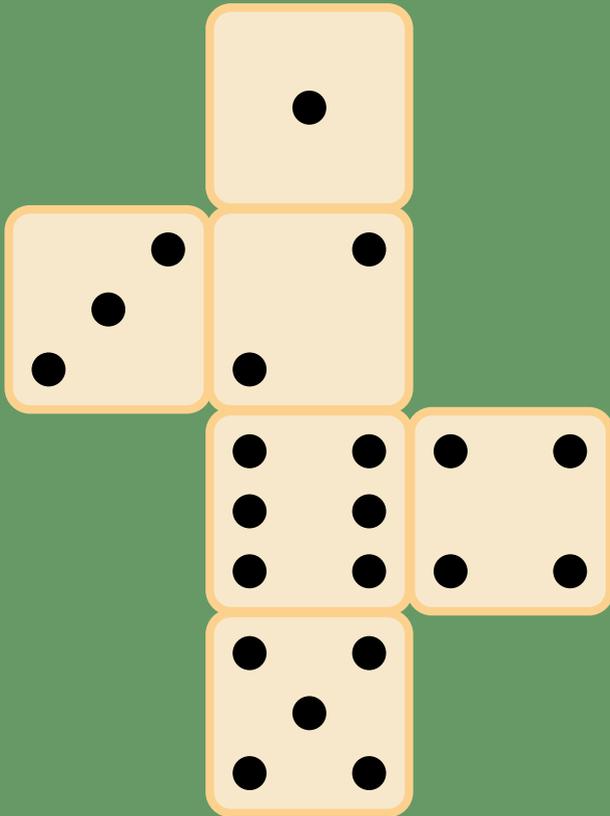
$$q_n(t) = \sum_{m \in \mathbb{S}} \Lambda_{nm}(t)P_m(t) \quad (6.62)$$

where

$$\Lambda_{nm}(t) = \mathbb{P}[A_n | X(t) = m]. \quad (6.63)$$

But of course, the expression (6.62) contains the example presented in this work if we take $\Lambda_{nm}(t) \equiv \lambda(t)\delta_{nm}$. The equation (6.62) motivates the natural generalization of a gauge transformation proposed in (6.20).

CONCLUSIONES



CONCLUSIONES

Hemos decidido no numerar esta sección, siguiendo el orden de los **6 capítulos** anteriormente expuestos. Pues entendemos que las conclusiones tratan no de realizar una lista de lo expuesto en estas **6 caras**, sino de sus implicancias y emergentes. Aunque sí deben exhibir al objeto como un todo que, en este caso, encierra un cuerpo de caras identificadas, i.e. una suerte de *dado*, fiel representante del azar. Entendemos el cierre de de este cuerpo como una representación de la armonía, fruto de la convivencia de las caras que lo componen, i.e. producto de la consistencia de las ideas que componen tal *corpus*. En general estas ideas o *caras*, que lo componen, son deformables y admiten la integración de otras más. Eventualmente podrá ocurrir que las implicaciones de estas ideas permitan incorporar *nuevas caras*, en cuyo caso no podremos cerrar el mismo cuerpo de conocimiento. Alguna de estas implicaciones serán enunciadas aquí como elementos fuera del programa de estudio de esta tesis.

El objetivo general del presente trabajo, acerca de un enfoque alternativo de los procesos estocásticos, fue abordado importando ideas y herramientas utilizadas en otras áreas de la física y la matemática. La introducción del cálculo de variaciones sobre una funcional *acción* se ha llevado a cabo con éxito con el fin de describir la dinámica *kolmogorivana*. Esto fue realizado, primeramente, para el caso en que la variable aleatoria del proceso estocástico bajo estudio tome valores sobre un conjunto de estados numerable. Seguidamente, hemos incluido la posibilidad de que el espacio de estados sea no numerable. Esto permite albergar también a los procesos de difusión o el movimiento browniano, entre otros, que pueden describirse por las ecuaciones de Fokker-Planck. En este contexto hemos desarrollado las ideas para incluir otros ti-

pos de ecuaciones diferenciales más generales. Del concepto de *acción* tenemos otro muy importante y es el de *lagrangiano*, cuyo rol puede ser elevado al de objeto fundamental de la teoría de procesos estocásticos *kolmogorovianos*.

A fin de realizar esta construcción fue preciso hablar de la noción de coordenadas para procesos kolmogorovianos. Hemos observado que, de la misma manera que tenemos dos ecuaciones de Kolmogorov para las medidas asociadas a probabilidades condicionales, persiste una dualidad descriptiva acerca del mismo proceso, cuyas partes difieren en el tipo de preguntas que responden, y que hemos denominado como *dualidad-PR*. Se trata de preguntas complementarias de carácter predictivo o retrodictivo.

Fue posible identificar el concepto de *inversión temporal* de un proceso estocástico. Hemos definido una operación que aplicada sobre la matriz de transiciones temporales, correspondiente a una prospección, nos deja otra matriz de transiciones temporales asociada a una retrospección y viceversa.

Ambas descripciones forward-backward coexisten necesariamente en un mismo lagrangiano, el cual contiene a ambas medidas absolutas de probabilidad como coordenadas de un espacio *simpléctico*. También hemos identificado cada una de las dos ecuaciones de Hamilton, con las ecuaciones *forward* y *backward* de Kolmogorov. Hemos tomado ventaja del formalismo lagrangiano y hamiltoniano para introducir los corchetes de Poisson, con el fin de identificar una serie de invariantes de la teoría de procesos estocásticos. Tal construcción puede realizarse para el caso de variables aleatorias continuas, haciendo uso de densidades de probabilidad absoluta.

Esto ha pretendido ser solo una introducción formal de algunas herramientas, en principio extranjeras, que podemos importar hacia procesos estocásticos. Quedan abiertas algunas cuestiones como, por ejemplo, el estudio de procesos estocásticos en los que intervengan varias variables aleatorias. En tal caso, sería primordial estudiar las medidas de probabilidad o funciones

densidad conjunta. Para empezar se requiere un estudio de la interacción de dos procesos estocásticos utilizando el formalismo lagrangiano que hemos descrito.

Hemos probado que cualesquiera dos procesos estocásticos en tiempo continuo y conjunto de estados numerables, pueden *mapearse* completamente vía una transformación local o gauge. Construimos una clase de equivalencia de todos los procesos de Kolmogorov en tiempo continuo y espacio de estados numerable, vía ese grupo de transformaciones locales. Estas transformaciones constituyen un conjunto de simetrías del lagrangiano, i.e. que dejan invariante la clase de estos procesos estocásticos.

Se ha mostrado el uso concreto de estas ideas con la construcción de un modelo con aplicación a biología. Tratamos un modelo para calcular el número de ancestros para especies de reproducción biparental. Este modelo presentaba el inconveniente de que el espacio de estados de la variable aleatoria cambiaba en el tiempo. En particular el número de estados crecía siguiendo una regla determinista, i.e. una función concreta del tiempo. Esto lo ha convertido en un candidato idóneo para hacer uso de las transformaciones gauge. Hemos resuelto un problema más sencillo que involucraba series de funciones especiales, en particular funciones de Bessel modificadas [81], de manera que modificando adecuadamente esa solución puede obtenerse como solución de nuestro problema ancestral original.

Sobre este punto debemos decir que hemos trabajado en un algoritmo independiente al modelo presentado, capaz de simular árboles de ancestros para contrastar con el modelo propuesto. Este algoritmo intenta generalizar el modelo analítico que hemos presentado. Permite incluir mayor información que característica de la especie particular bajo estudio. Permitiendo describir el grado de endogamia en cada generación y el grado de preferencia en la selección de la pareja. Fue posible realizar una descripción analítica del modelo algorítmico, para evidenciar aquellos puntos propensos de ser extendidos a una mayor variedad de casos. El

mencionado trabajo se encuentra en prensa pero hemos decidido no incluirlo aquí, puesto que daría lugar a un sin fin de preguntas muy diferentes de las que nos hemos propuesto estudiar.

La utilización del cálculo de variaciones en los procesos estocásticos admiten, como hemos dicho, la introducción formal de un lagrangiano y una acción. Junto con el requerimiento de que ésta última sea estacionaria, nos permiten deducir las ecuaciones de Kolmogorov para probabilidades absolutas. Con todo ello se tienen dos conclusiones secundarias. La primera de ellas es la formalización de la dinámica estocástica kolmogoroviana, cuyo escenario de aplicación comprende la física, la biología, la economía, entre otros. La segunda de ellas es la de abrir un campo hacia la geometrización de la teoría de los procesos estocásticos.

Del concepto de probabilidad dependen otras teorías. La primera de ellas tiene sus raíces en el campo de las comunicaciones, se trata de la *teoría clásica de la información*. La segunda de ellas es la *teoría cuántica de la información*. Las teorías cuánticas (e.g. de la mecánica, de campos, de la información, etc.) requieren fundamentalmente el uso del concepto de probabilidad. Su interpretación es la misma que en las teorías clásicas, pero lo que cambia radicalmente es el método empleado para calcularlas [53]. Existen también una serie de extensiones de la teoría de probabilidades clásicas a otras *no kolmogorovianas* [54] [55].

Por otro lado, el cálculo de variaciones ha renovado a la física, no solamente porque gran parte de sus teorías, sino todas, pueden entenderse a través de los mismos principios fundamentales; sino también porque es posible conectar disciplinas conceptualmente diferentes. Tal es el caso de la mecánica clásica y la mecánica cuántica. Esta hipótesis fue desarrollada por R. Feynman en su tesis doctoral de 1942, a través del formalismo hoy conocido como *path integral* que permite relacionar una teoría clásica lagrangiana (junto con su principio variacional) y una teoría cuántica de manera explícita. De hecho, esta vinculación ha permitido ganar una mejor comprensión de ambas teorías y su interrelación.

En este trabajo hemos emitido los primeros balbuceos de una formulación lagrangiana y hamiltoniana de los procesos estocásticos *kolmogorovianos*. Pretendemos dejar un planteamiento claro de un punto inicial a fin de emprender un nuevo camino hacia una posible extensión de la teoría de los procesos estocásticos. Pues cabe la pregunta sobre cuál es el tipo de teorías a las que se puede arribar utilizando el formalismo *path integral* sobre nuestro lagrangiano *clásico*. Sobre ese punto hemos iniciado ya una serie de cálculos que exceden los límites impuestos en esta tesis y dan cuenta de nuevos motivos para futuras investigaciones.

BIBLIOGRAFÍA

- [1] Elsgoltz, L., *Ecuaciones diferenciales y cálculo variacional*, Editorial MIR (1969).
- [2] Basdevant, J. L., *Variational Principles in Physics*, Springer (2007).
- [3] Wagensberg, J., *Las raíces triviales de lo fundamental*, Tusquets Editores (2010).
- [4] Bunge, M., *Foundations of physics*, Springer-Verlag (1967).
- [5] Feynman, R. P., *Space-Time Approach to Non-Relativistic Quantum Mechanics*. Rev. Mod. Phys. **20** (2), 367 – 387 (1953).
- [6] Dirac, P. A., *Principles of Quantum Mechanics*. Oxford (1935).
- [7] García Pérez, P. L. *Sobre la naturaleza variacional de la ley física*. Real Academia de Ciencias Exactas, Físicas y Naturales-Discurso leído en el acto de recepción como académico de número, (2008).
- [8] Feynman, R. P.; Leighton, R. B.; Sands, M., *The Feynman lectures on physics*, Addison-Wesley Publishing Company, Volume 1 (1963).
- [9] Feynman, R. P., *QED: The strange theory of light and matter*, Princeton University Press, (2006).
- [10] D. Hilbert, *Mathematical Problems*. Bull. Amer. Math. Soc., **8** (1901-1902), 437 – 479 (Versión en inglés, autorizada por Hilbert, del original aparecido en Göttinger Nachrichten en 1900 y en Archiv der Mathematik und Physik, **1** (1901), pp. 44 – 63 y 213 – 237).

- [11] Kolmogorov, A., *Trudy Kommunist*, Akad. Razd. Mat. **1**, 8 – 21 (1929). O bien *General measure theory and probability calculus*. Selected Works of A. N. Kolmogorov, Vol. II, Springer (1992).
- [12] Levi, M., *The mathematical mechanic: using physical reasoning to solve problems*, Princeton University Press (2010).
- [13] Kolmogorov, A., *Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung*. Math. Ann. **104** (1), 415 – 458 (1931). O bien *On Analytical Methods in Probability Theory*. Selected Works of A. N. Kolmogorov, Vol. II, Springer (1992).
- [14] Mendel, G., *Experiments in plant hybridization*, Brünn Natural History Society (1865). Traducción al inglés Druery, C. T., William Bateson, J. Royal Horticultural Soc. **26**, 1 – 32 (1901).
- [15] Darwin C., *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*, John Murray, London(1859).
- [16] Doob, J. L., *Stochastic Processes*, John Wiley & Sons, (1953)
- [17] Chung, K. L., *Elementary probability theory with stochastic processes*, Springer-Verlag (1979).
- [18] Gikhman, I. I., Skorokhod, A. V., *Introduction to the theory of random process*, W. B. Saunders Company (1969).
- [19] Dynkin, E. B., *Markov Process*, Springer-Verlag (1965).
- [20] Chung, K. L., *Markov Chains: with stationary transitions probabilities*. Springer-Verlag (1960).
- [21] Kijima, M., *Markov Processes for Stochastic Modeling* (Chapman & Hall 1997).
- [22] Kemeny, J. G., Snell, J. L., Knapp, A. W., *Denumerable Markov Chains*, Springer-Verlag (1976).

- [23] Apostol, T. M., *Mathematical analysis*, Addison-Wesley Pub. Co. (1974).
- [24] Feller, W., *An Introduction to Probability Theory and Its Applications*. Volume I & Volume 2, John Wiley & Son Inc (1968).
- [25] Reed, M., Simon, B., *Methods of Modern Mathematical Physics*. Volume 1-4, Academic Press (1975).
- [26] Marsden, J., Ratiu, T., Abraham, R. *Manifolds, Tensor Analysis, and Applications*. Springer-Verlag (2010).
- [27] Kolmogorov, A., *Zur Theorie der Markoffschen Ketten*. Math. Ann. **112**, 155 – 160(1936). O bien *On the theory of Markov chains*. Selected Works of A. N. Kolmogorov, Vol. II, Springer (1992).
- [28] Bunge, M., *Physical Time: The Objective and Relational Theory*. Phil. Sci. **35**, (4), 355 – 388, (1968).
- [29] Bunge, M., *Controversias en Física*, Editorial Tecnos, (1983).
- [30] Landau, L. D., Lifshitz, E. M., *Curso de física teórica: Mecánica* Editorial Reverté (1994).
- [31] Goldstein, H., Poole C., Safko, J., *Classical mechanics* Addison Wesley (2001).
- [32] Bang-Jensen, J., Gutin, G. Z., *Digraphs: Theory, Algorithms and Applications*, Springer, (2010)
- [33] Feynman, R. P., *The Character of the Physical Law*, MIT Press Paperback Edition, 12° ed. (1985). O bien *El carácter de la ley física*, Tusquets (2015).
- [34] Arnold, V. I., *Ordinary differential equations*, Springer-Verlag, (1992).

- [35] Arnold, V. I., *Mathematical Methods of Classical Mechanics*, Springer-Verlag, (1992).
- [36] Abraham, R., Marsden, J. *Foundations of mechanics*. Addison-Wesley Publishing Company Inc., (1987).
- [37] Weyl, H.; *The Classical Groups. Their Invariants and Representations*, Princeton University Press, (1939).
- [38] Bunge, M., *La causalidad: el principio de causalidad en la ciencia moderna*, Editorial Sudamericana (1997), O bien *Causality. The place of the causal principle in modern science*, Cambridge, Massachusetts, Harvard University Press (1959).
- [39] Bunge, M., *The revival of causality*, Contemporary philosophy: Chroniques nouvelles / A new survey Springer Netherlands 133 – 155 (1982).
- [40] Boyd, Stephen P., *Convex Optimization*, Cambridge University Press, (2011).
- [41] Caruso, M., Fanchiotti, H., García Canal, C. A. *Equivalence between classical and quantum dynamics. Neutral kaons and electric circuits*, Ann. Phys. **326**, 2717 – 2736 (2011).
- [42] Jones, C. K. R. T., Kirchgraber U., Walther, H. O. *Dynamics Reported: Expositions in Dynamical Systems 5*, Springer-Verlag (1996).
- [43] Fokker, A. *Die mittlere Energie rotierender elektrischer Dipole im Strahlungsfeld*, Ann. Phys. **348**, 810–820 (1914).
- [44] Planck, M. *Über einen Satz der statistischen Dynamik und seine Erweiterung in der Quantentheorie*, Sitzungsber. Preuss. Akad. Wiss. **24** (1917)
- [45] Kramers, H. A., Physica **7**, 284 (1940)
- [46] Moyal, J. E., J. Roy. Stat. Soc. B **11**, 150 (1949).

- [47] Ising E., Contribution to the Theory of Ferromagnetism. *Z.Phys.* **31** 253 – 258 (1925).
- [48] Kleinrock, L., *Queueing Theory* (Volume I, Wiley-Interscience Publication 1975).
- [49] D. Anderson, T. Kurtz, *Design and Analysis of Biomolecular Circuits*, Continuous time markov chain models for chemical reaction networks, Springer New York (2011).
- [50] L. Shapiro, D. Zeilberger, A Markov chain occurring in enzyme kinetics. *Jour. Math. Biol.* **15** 351 – 357 (1982)
- [51] Caruso, M., Jarne, C., Markov chain approach to the distribution of ancestors in species of biparental reproduction. *Phys. Rev. E.* **90**, 022125 (2014).
- [52] W. Magnus, On the exponential solution of differential equations for a linear operator. *Comm. Pure and Appl. Math.* **VII 4**, 649 – 673 (1954).
- [53] Feynman, R. P., *The concept of probability in quantum mechanics*, University of California Press (1951).
- [54] Holik, F., Sáenz M., Plastino A., *A discussion on the origin of quantum probabilities*, *Ann. Phys.* **340**, 293 – 310(2014).
- [55] Holik, F., Bosyk, G. M., Bellomo, G. , *Quantum Information as a Non-Kolmogorovian Generalization of Shannon's Theory*, *Entropy* **17** (11), 7349 – 7373 (2015).
- [56] Derrida, B., Manrubia, S. C., Zanette, D.H., Statistical Properties of Genealogical Trees. *Phys. Rev. Lett.* **82**, 9, 1987 – 1999 (1999).
- [57] Derrida, B., Manrubia, S. C., Zanette, D.H., Distribution of repetitions of ancestors in genealogical trees. *Physica A* **281** 1 – 16 (2000).

- [58] Derrida, B., Manrubia, S. C., Zanette, D.H., On the genealogy of a population of biparental individual. *J. theor. Biol.* **203**, 303 – 315 (2000).
- [59] Ohno, S., The Malthusian parameter of ascents: What prevents the exponential increase of one's ancestors?. *Proc. Natl. Acad. Sci.* **93**, 15276 – 15278 (1993).
- [60] Bennet, J. H., On the theory of random mating. Department of Genetics, University of Cambridge, **17**, 1, 311 – 317 (1952).
- [61] Bennet, J. H., Random mating and sex linkage. *J. Theor. Biol.* **4**, 1, 28 – 36 (1963).
- [62] Hardy, G. H., Mendelian Proportions in a Mixed Population. *Sci. N. S.* **28**, 49 – 50 (1908).
- [63] Weinberg, W., *Über den Nachweis der Vererbung beim Menschen*. Jahresh. Ver. Vaterl Naturkd. Württomb. **64**: 369 – 382 (1908), (English translations in Boyer 1963 and Jamenson 1977).
- [64] Begon, M., Townsend, C. R., Harper, J. L., *Ecology - From Individuals to Ecosystems* (Blackwell Publishing, 4th edition 2006).
- [65] Barucha-Reid, A.T., *Elements of the Theory of Markov Processes and Their Applications* (McGraw-Hill, Series in Probability and Statistics 1960).
- [66] Pagel, M., Inferring the historical patterns of biological evolution. *Nature* **401**, 877 – 844 (1999).
- [67] Chung, K. L., *Markov Chains With Stationary Transition Probabilities* (Springer, New York, 2^o edition 1967).
- [68] Kleinrock, L., *Queueing Theory* (Volume I, Wiley-Interscience Publication 1975).

- [69] Feynman, R., Mathematical Formulation of the Quantum Theory of Electrodynamics Interaction. *Phys. Rev.* **80**, 3, 440 – 457 (1950).
- [70] Serva, M. & Peliti, L., A statistical model of an evolving population with sexual reproduction, *J. Phys. A: Math. Gen.* **24**, 705 – 707 (1991).
- [71] Liberman, E.; Hauert, C. and Nowak, M. A., Evolutionary dynamics on graphs. *Nature* **433**, 312 – 316 (2005).
- [72] Heitler, W., *The Quantum Theory of Radiation* (Oxford University Press London, 3rd edition 1954).
- [73] Douglas L. T. Rohde, Steve Olson, Joseph T. Chang, Modelling the recent common ancestry of all living humans. *Nature* **431**, 562 – 566 (2004).
- [74] N.T.J. Bailey, A continuous time treatment of a simple queue using generating functions, *J. Roy. Statist. Soc. Ser. B* **16** (1954) 288 – 291 (1954).
- [75] D.G. Champernowne, An elementary method of solution of the queueing problem with a single server and constant parameter, *J. Roy. Statist. Soc. Ser. B* **18** 125 – 128 (1956).
- [76] A.B. Clark, The time dependent waiting line problem, *Univ. Michigan*, Report M729-IR39, (1953).
- [77] W. Ledermann & G.E.H. Reuter, Spectral theory for the differential equations of simple birth and death processes, *Philos. Trans. Roy. Soc. London Ser. A* **246** 321 – 369 (1954).
- [78] A. Einstein, *Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen*, *Annalen der Physik* (ser. 4), **17**, 549 – 560. O bien *On the Motion of Small Particles Suspended in Liquids at Rest Required by the Molecular-Kinetic Theory of Heat*, Princeton University Press.

- [79] Onsager, L., Machlup, S., *Fluctuations and Irreversible Processes*. Phys. Rev. **91** (6), 1505 – 1512 (1953).
- [80] Bunge, M., *Lagrangian formulation and mechanical interpretation*. Am. J. Phys. **25**, (4), 211 – 218, (1957).
- [81] Abramowitz, M. Stegun, I. A., *Handbook of mathematical functions*, Dover books on mathematics (1972).