



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

**POSGRADO EN CIENCIAS
MATEMÁTICAS**

FACULTAD DE CIENCIAS

**El proceso Fleming-Viot de mutación neutral
desde una perspectiva Bayesiana**

TESIS

QUE PARA OBTENER EL GRADO ACADÉMICO DE

MAESTRA EN CIENCIAS

PRESENTA

ISADORA ANTONIANO VILLALOBOS

DIRECTOR DE TESIS: DR. RAMSÉS HUMBERTO MENA CHÁVEZ

MÉXICO, D.F.

MAYO, 2008



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

A Rubén, Ivonne y Enrique.

Índice general

Resumen.	v
Agradecimientos.	VII
Introducción.	IX
1. Marco teórico.	1
1.1. Convergencia.	2
1.1.1. Convergencia débil y topología de Prokhorov.	2
1.1.2. Topología de Skorokhod para trayectorias «càdlàg».	5
1.2. Procesos de Markov.	9
1.2.1. Semigrupos y generadores.	9
1.2.2. Procesos de Markov.	12
1.2.3. El problema de martingala.	15
1.2.4. Ejemplo: un proceso de muerte puro.	19
1.3. Intercambiabilidad y modelo bayesiano no paramétrico	20
1.3.1. Inferencia bayesiana y método de Monte Carlo.	24
1.3.2. Método de muestreo de Gibbs.	26
1.4. Procesos de Markov vía procesos latentes	27
1.4.1. Cadenas de Markov vía variables latentes.	27
1.4.2. Generalización del método: procesos markovianos en tiempo continuo.	29
2. Modelos de Urnas y el proceso Dirichlet.	33
2.1. Urnas de Pólya generalizadas	34
2.2. El proceso Dirichlet.	38
3. Proceso Fleming-Viot.	47
3.1. Modelos de difusión y genética poblacional.	48
3.2. Cadenas de Markov con valores en un espacio de medidas.	50

3.2.1. Modelo diploide.	50
3.2.2. Modelo de Wright-Fisher.	52
3.2.3. Modelo de Moran.	53
3.3. Caracterización del proceso Fleming-Viot.	53
3.4. Modelo de mutación neutral.	56
4. Fleming-Viot vía urnas.	59
4.1. Proceso de partículas intercambiables.	60
4.2. Ecuaciones de Chapman-Kolmogorov.	63
4.3. Modelo de mutación neutral.	66
Conclusiones.	73
Bibliografía.	76

Resumen.

El objeto principal de esta tesis es la presentación de un método para la construcción de funciones de transición homogéneas en el tiempo, correspondientes a procesos de Markov con valores en espacios de medidas de probabilidad. Dicho método se basa en la introducción de procesos latentes y en las ideas que fundamentan el método de simulación conocido como muestreo de Gibbs y se ilustra en este trabajo por medio de un ejemplo.

Para ello, se define el proceso de difusión conocido como Fleming-Viot, que en el contexto de genética poblacional se utiliza para modelar la distribución de tipos genéticos en una población y su evolución en el tiempo. Se enuncia algunas propiedades conocidas para un caso particular, conocido como modelo de mutación neutral para alelos infinitos o simplemente Fleming-Viot de mutación neutral. Tanto en su forma más general como en el caso más sencillo, el proceso se define formalmente a través de su generador infinitesimal.

Posteriormente, se presenta una construcción alternativa para la función de transición correspondiente al proceso Fleming-Viot de mutación neutral, a través de un proceso latente cuyas realizaciones son vectores aleatorios con entradas distribuidas de acuerdo con un modelo bayesiano no paramétrico. Se utiliza resultados desarrollados para el proceso Dirichlet en el ámbito de la estadística bayesiana no paramétrica, en particular su caracterización en términos de una urna de Pólya generalizada, para derivar de forma sencilla algunas propiedades del proceso Fleming-Viot así construido.

A modo de conclusión, se propone la posibilidad de generalizar el método para la definición de otros procesos de Markov con valores en espacios de medidas de probabilidad.

Agradecimientos.

Agradezco a los maestros que con sus enseñanzas y sus pláticas, además de enseñarme, me aligeraron el camino.

A Ramsés, sin cuyo apoyo y guía, este trabajo no existiría y quien me ha ayudado a dirigir mis pasos hacia el futuro.

A los amigos, los de ahora y los de antes, por la alegría sin la cual lo demás pierde importancia.

A mi familia, por todo, pues cualquier otra palabra sería insuficiente.

A Conacyt, que con una beca, me permitió dedicar toda mi energía a aprender.

Introducción.

Un modelo básico muy utilizado por la estadística bayesiana, considera los datos como realizaciones de un conjunto de variables aleatorias, con una distribución común que no es completamente conocida y condicionalmente independientes dada la información desconocida. En el caso paramétrico, se supone que todo aquello que se desconoce acerca de la distribución, puede ser representado por un parámetro θ de dimensión finita, el cual toma valores en un espacio Θ completamente conocido. La incertidumbre acerca del fenómeno de estudio al momento de plantear el modelo, se incorpora en una distribución inicial $\Pi(\theta)$. Este modelo bayesiano paramétrico se puede escribir como

$$\begin{aligned} X_k | \mu_\theta &\stackrel{iid}{\sim} \mu_\theta; \quad k = 1, \dots, n \\ \theta &\sim \Pi, \end{aligned}$$

donde X_1, \dots, X_n representan las observaciones y μ_θ es una distribución perteneciente a una familia indexada por el parámetro desconocido θ . La distribución Π definida sobre Θ , induce a su vez una distribución sobre la familia de distribuciones $\{\mu_\theta\}_{\theta \in \Theta}$. En el modelo no paramétrico, se supone que la información desconocida es demasiada para ser representada por un parámetro de dimensión finita. Por lo tanto, el modelo anterior se transforma en

$$\begin{aligned} X_k | \mu &\stackrel{iid}{\sim} \mu; \quad k = 1, \dots, n \\ \mu &\sim \Pi, \end{aligned}$$

donde $\mu \in \mathcal{P}(E)$ es cualquier distribución o medida de probabilidad sobre el espacio E en el cual toman valores las variables X_1, \dots, X_n . Esta vez, Π es una distribución definida sobre todo $\mathcal{P}(E)$, no únicamente sobre un subespacio indexado por algún parámetro de dimensión finita. La existencia de esta distribución queda garantizada por el teorema de representación de de Finetti, bajo el supuesto de intercambiabilidad de las observaciones, que se convierte así en un supuesto fundamental de muchos modelos bayesianos no paramétricos.

En el contexto del modelo no paramétrico, μ juega el papel de una «variable aleatoria» que toma valores sobre el espacio $\mathcal{P}(E)$ de medidas de probabilidad sobre E , por lo que suele llamársele medida aleatoria. El espacio $\mathcal{P}(E)$ es muy grande y en general, difícil de trabajar. Sin embargo, el desarrollo de la estadística bayesiana no paramétrica ha dado origen a resultados teóricos y herramientas prácticas para el manejo de medidas aleatorias, que pueden aprovecharse en otras áreas. El presente trabajo es un ejemplo de ello.

La genética poblacional debe en gran parte su desarrollo a la utilización de modelos de probabilidad, en los cuales el estado de la población estudiada en un momento dado, se representa por medio de una distribución de probabilidad. Así pues, la evolución de una población se modela por medio de procesos estocásticos que toman valores en un espacio $\mathcal{P}(E)$ de medidas de probabilidad. Una importante familia de procesos de este tipo fue introducida en 1979 por Fleming y Viot [FV79]. El ahora conocido proceso de Fleming-Viot es un proceso de Markov a tiempo continuo, definido a través de su generador infinitesimal y abarca una amplia gama de modelos. Uno de los más simples es el modelo de mutación neutral para alelos infinitos, estudiado entre otros, por Ethier y Griffiths [EG93] y Ethier y Kurtz [EK86, EK93]. Aún tratándose de un caso relativamente sencillo, el estudio de las propiedades de este proceso ha dependido, hasta ahora, de un intenso desarrollo teórico y un amplio conocimiento de la teoría de procesos de Markov, martingalas y generadores infinitesimales.

El presente trabajo, está basado en los resultados de Walker, Hatjispuros y Nicolieris [WHN07], que extienden las ideas de Pitt, Chatfield y Walker [PCW02] y Mena y Walker [MW05, MW07] para la construcción de procesos de Markov por medio de procesos latentes, explotando las propiedades del método de simulación conocido como muestreo de Gibbs.

Se presenta una construcción alternativa del modelo de mutación neutral para alelos infinitos, a través de un proceso latente cuyas realizaciones son vectores aleatorios con entradas distribuidas de acuerdo con un modelo bayesiano no paramétrico. Utiliza resultados desarrollados para el proceso Dirichlet en el ámbito de la estadística bayesiana no paramétrica, en particular su caracterización en términos de una urna de Pólya generalizada, para derivar de forma sencilla algunas propiedades del proceso Fleming-Viot así construido.

En el primer capítulo se introduce brevemente los principales conceptos

implicados en el desarrollo del trabajo. Se define dos conceptos de convergencia débil para medidas de probabilidad; el primero inducido por la métrica de Prokhorov y el segundo por la métrica de Skorokhod para el espacio $\mathcal{D}_E[0, \infty)$ de trayectorias «càdlàg» sobre E . Se presenta definiciones y resultados importantes involucrados en la teoría de procesos de Markov y particularmente, su caracterización a través de generadores, semigrupos y el problema de martingala de Stroock y Varadhan. Se explica un modelo bayesiano no paramétrico, el concepto de intercambiabilidad y el teorema de de Finetti que los relaciona. Se menciona algunas dificultades que surgen en torno a la inferencia bayesiana y se introduce el algoritmo de muestreo de Gibbs. Finalmente, se ilustra una idea general para la construcción de procesos de Markov a través de procesos latentes, en la cual se basa el resultado principal de este trabajo.

En el segundo capítulo se presenta los modelos conocidos como urnas de Pólya generalizadas, su relación con el concepto de intercambiabilidad para sucesiones de variables aleatorias y algunos resultados generales. Se pone particular énfasis en el proceso Dirichlet que, además de ser muy importante para la estadística bayesiana no paramétrica, es esencial para el desarrollo posterior del trabajo.

En el tercer capítulo se define el proceso Fleming-Viot en el contexto de genética poblacional. Se define algunos conceptos relacionados con la teoría de la evolución, cuya incorporación a modelos probabilísticos ha hecho de la genética matemática un área muy importante para el avance de la investigación en genética evolutiva. Se presenta algunos modelos a tiempo discreto para poblaciones finitas, que pueden ser reinterpretados como aproximaciones al Fleming-Viot. Se define dicho proceso en su sentido más amplio, a través de su generador infinitesimal. Se introduce como caso particular el modelo de mutación neutral, para el cual la distribución estacionaria y la función de transición son conocidas.

El último capítulo contiene el resultado principal del trabajo. Se plantea un modelo a tiempo discreto para la construcción de una cadena de Markov con valores en el espacio $\mathcal{P}(E)$, cuya distribución estacionaria es un proceso Dirichlet. Dicho modelo se extiende para introducir una dependencia con respecto al tiempo, construyéndose así un proceso de Markov a tiempo continuo definido a través de una función de transición y cuya existencia se prueba por la verificación de las condiciones de Chapman-Kolmogorov. Gracias a las propiedades del proceso Dirichlet y en especial a su representación como una urna de Pólya generalizada, estas condiciones pueden ser reexpresadas en términos de un proceso con espacio de estados discreto. Finalmente, se exhibe un proceso de este tipo, el cual es solución para dichas

condiciones. El proceso de Markov construido en este caso, coincide con el modelo Fleming-Viot de mutación neutral.

Capítulo 1

Marco teórico.

El presente capítulo contiene algunas definiciones y resultados que sirven de marco teórico para el desarrollo de este trabajo.

En la primera sección, se define dos conceptos de convergencia, el primero en el espacio $\mathcal{P}(E)$ de medidas de probabilidad sobre un espacio métrico (E, r) y el segundo sobre el espacio $\mathcal{D}_E[0, \infty)$ de trayectorias continuas por la derecha con límites por la izquierda. Para ello, se define las métricas de las cuales se deriva la topología de convergencia débil en cada uno de dichos espacios. Se define a partir de esto, la convergencia en distribución para variables aleatorias con valores en E y para procesos aleatorios con trayectorias en $\mathcal{D}_E[0, \infty)$.

La segunda sección presenta algunos conceptos y resultados importantes de la teoría de procesos de Markov. Introduce definiciones básicas relativas a semigrupos de operadores, generadores y el problema de martingala de Stroock y Varadhan para la caracterización de procesos de Markov.

En la tercera sección se da una breve explicación de las ideas en que se basa la estadística bayesiana y específicamente la bayesiana no paramétrica. Se define la propiedad de intercambiabilidad para sucesiones de variables aleatorias y se enuncia el teorema de de Finetti, como posible justificación teórica para los modelos bayesianos no paramétricos. Se introduce el concepto de medida aleatoria y su papel dentro de un modelo bayesiano no paramétrico básico. Se menciona algunas dificultades relacionadas con la inferencia bayesiana y la consecuente necesidad de algoritmos de simulación para distribuciones multivariadas. En este contexto se presenta el algoritmo de muestreo de Gibbs y su relación con las cadenas de Markov.

Finalmente, en la cuarta sección se presenta un método para definir cadenas de Markov con distribuciones estacionarias dadas, basado en la idea

que da lugar al muestreo de Gibbs. Se menciona algunas maneras de generalizar este método y en particular, se ilustra con un ejemplo la aplicación del método para generar procesos de Markov a tiempo continuo a través de variables latentes.

No se pretende desarrollar en un capítulo todos estos temas de forma exhaustiva. Únicamente se presenta conceptos y resultados que son de utilidad para la comprensión del resto del trabajo. Para una lectura más profunda de los temas de convergencia y procesos de Markov, se puede consultar [EK86, Lam77]. Los temas relacionados con la inferencia bayesiana se pueden profundizar en [Sch95, BS04]. La sección sobre construcción de cadenas y procesos de Markov a tiempo continuo está basada en [PCW02, MW05, MW07].

1.1. Convergencia.

1.1.1. Convergencia débil y topología de Prokhorov.

Sea (E, r) un espacio métrico. Denótese por $\mathcal{B}(E)$ al álgebra de subconjuntos de Borel de E , por $\mathcal{P}(E)$ al espacio de medidas de probabilidad sobre E y por $\mathcal{C}_b(E)$ al espacio de funciones reales continuas y acotadas en (E, r) , dotado con la norma del supremo, es decir $\|f\| = \sup_{x \in E} |f(x)|$.

Definición 1.1.1 (Convergencia débil). Una sucesión $\{\mu_n\}_{n \geq 0} \subset \mathcal{P}(E)$ converge débilmente a $\mu \in \mathcal{P}(E)$ si, para toda $f \in \mathcal{C}_b(E)$

$$\lim_{n \rightarrow \infty} \int f d\mu_n = \int f d\mu \quad (1.1)$$

y se escribirá $\mu_n \Rightarrow \mu$.

Sean $(\Omega, \mathcal{F}, \mathbb{P})$ un espacio de probabilidad y $X : \Omega \rightarrow E$ una variable aleatoria con valores en E . Denótese por $\mu_X \in \mathcal{P}(E)$ la distribución de X , es decir, $\mu_X(B) = \mathbb{P}[X \in B]$ para cualquier $B \in \mathcal{B}(E)$.

Definición 1.1.2 (Convergencia en distribución). Una sucesión $\{X_n\}_{n \geq 0}$ de variables aleatorias con valores en E converge en distribución a la variable aleatoria X con valores en E si $\mu_{X_n} \Rightarrow \mu_X$ o equivalentemente, si

$$\lim_{n \rightarrow \infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)] \quad (1.2)$$

para toda $f \in \mathcal{C}_b(E)$. En este caso, se denotará $X_n \xrightarrow{\mathcal{D}} X$.

Definición 1.1.3 (Métrica de Prokhorov). Sea (E, r) un espacio métrico. Si se define $\rho : \mathcal{P}(E) \times \mathcal{P}(E) \rightarrow \mathbb{R}^+$ como

$$\rho(\mu, \nu) = \inf\{\epsilon > 0 : \mu(F) \leq \nu(F^\epsilon) + \epsilon, \forall F \subset E \text{ cerrado}\}, \quad (1.3)$$

donde

$$F^\epsilon = \{x \in E : \inf_{y \in F} r(x, y) < \epsilon\}, \quad (1.4)$$

entonces ρ es una métrica sobre $\mathcal{P}(E)$, conocida como la métrica de Prokhorov (también escrito Prohorov por algunos autores [EK86]).

La métrica de Prokhorov, está definida en general para cualesquiera dos medidas finitas sobre $(E, \mathcal{B}(E))$, pero en el caso de medidas de probabilidad se tiene que $\rho(\mu, \nu) \leq 1$ para cualesquiera $\mu, \nu \in \mathcal{P}(E)$. Se trata de una extensión de la medida de Lévy para espacios de funciones de distribución y al igual que ella, induce la topología de convergencia débil. Es decir, $\mathcal{P}(E)$ con la topología de convergencia débil, es metrizable por ρ , como muestra el siguiente resultado.

Teorema 1.1.4. Sean (E, r) un espacio métrico, $\{\mu_n\}_{n \geq 0} \subset \mathcal{P}(E)$ una sucesión de medidas de probabilidad sobre E y $\mu \in \mathcal{P}(E)$. Entonces, las condiciones, (b) a (f) son equivalentes e implicadas por (a). Si además E es separable, las seis condiciones son equivalentes:

- (a) $\lim_{n \rightarrow \infty} \rho(\mu_n, \mu) = 0$.
- (b) $\mu_n \Rightarrow \mu$.
- (c) $\lim_{n \rightarrow \infty} \int f d\mu_n = \int f d\mu$, para cualquier $f \in \mathcal{C}_b(E)$ uniformemente continua.
- (d) $\limsup_{n \rightarrow \infty} \mu_n(F) \leq \mu(F)$ para todo $F \subset E$ cerrado.
- (e) $\liminf_{n \rightarrow \infty} \mu_n(G) \geq \mu(G)$ para todo $G \subset E$ abierto.
- (f) $\lim_{n \rightarrow \infty} \mu_n(A) = \mu(A)$ para todo $A \subset E$ para el cual μ es continua, es decir $A \in \mathcal{B}(E)$ y $\mu(\partial A) = 0$, donde ∂A es la frontera de A .

Gracias a esto, es posible llevar el concepto de convergencia débil al ámbito de convergencia en espacios métricos, al tiempo que se conserva algunas propiedades importantes, de acuerdo con el siguiente teorema.

Teorema 1.1.5. Si E es separable, entonces $\mathcal{P}(E)$ también lo es. Si además (E, r) es completo, $(\mathcal{P}(E), \rho)$ es completo.

Cuando se desea verificar la convergencia de una sucesión $\{x_n\}_{n \geq 0}$ en un espacio métrico, una práctica común consiste en verificar que la sucesión está contenida en algún conjunto compacto y probar después que cada subsucesión convergente de $\{x_n\}_{n \geq 0}$ debe converger al mismo elemento x . De ello se deriva que $\lim_{n \rightarrow \infty} x_n = x$. Para aplicar este argumento al espacio $(\mathcal{P}(E), \rho)$, es conveniente contar con caracterizaciones de los subconjuntos compactos de $\mathcal{P}(E)$. Una forma de hacerlo es a través del teorema de Prokhorov, que relaciona los conceptos de tensión y compacidad.

Definición 1.1.6 (Tensión). Sea (E, r) espacio métrico. Una medida $\mu \in \mathcal{P}(E)$ se dice tensa si para cada $\epsilon > 0$ existe un conjunto compacto $K \subset E$ tal que $\mu(K) \geq 1 - \epsilon$.

Una familia $\mathcal{M} \subset \mathcal{P}(E)$ se dice tensa si para cada $\epsilon > 0$ existe un conjunto compacto $K \subset E$ tal que

$$\inf_{\mu \in \mathcal{M}} \mu(K) \geq 1 - \epsilon.$$

Lema 1.1.7. Si (E, r) es un espacio métrico completo y separable, entonces toda $\mu \in \mathcal{P}(E)$ es tensa.

Definición 1.1.8 (Compacidad relativa). Una familia de medidas de probabilidad $\mathcal{M} \subset \mathcal{P}(E)$ se dice relativamente compacta si su cerradura es compacta. Equivalentemente, \mathcal{M} es relativamente compacta si para cada sucesión $\{\mu_n\}_{n \geq 0} \subset \mathcal{M}$, existe una subsucesión $\{\mu_{n_i}\}_{i \geq 0}$ y una medida $\mu \in \mathcal{P}(E)$ (no necesariamente en \mathcal{M}), tal que $\mu_{n_i} \Rightarrow \mu$.

Teorema 1.1.9 (Prokhorov). Sean (E, r) completo y separable y $\mathcal{M} \subset \mathcal{P}(E)$. Entonces, las siguientes condiciones son equivalentes:

- (a) \mathcal{M} es tensa.
- (b) Para cada $\epsilon > 0$, existe un conjunto compacto $K \subset E$ tal que

$$\inf_{\mu \in \mathcal{M}} \mu(K^\epsilon) \geq 1 - \epsilon, \tag{1.5}$$

donde K^ϵ se define como en la expresión (1.4).

- (c) \mathcal{M} es relativamente compacta.

Corolario 1.1.10. Para (E, r) arbitrario, si $\mathcal{M} \subset \mathcal{P}(E)$ es tensa, entonces es relativamente compacta.

Si $(E_n, r_n)_{n \geq 1}$ es una sucesión de espacios métricos, se define el espacio métrico producto (E, r) como

$$E = \prod_{n=1}^{\infty} E_n, \quad (1.6)$$

$$r(x, y) = \sum_{n=1}^{\infty} \frac{1}{2^n} \min\{1, r_n(x_n, y_n)\}, \quad x, y \in E. \quad (1.7)$$

Las siguientes proposiciones permiten una generalización de los resultados de tensión y compacidad, para espacios producto.

Proposición 1.1.11. *Si E_n es separable para cada $n \geq 1$, entonces E es separable y $\mathcal{B}(E) = \prod_{n=1}^{\infty} \mathcal{B}(E_n)$.*

Proposición 1.1.12. *Sea $\{\mu_\alpha\}_{\alpha \in \mathcal{I}} \subset \mathcal{P}(E)$, donde \mathcal{I} es un conjunto de índices. Para cada α y $n \geq 1$, sea μ_α^n la n -ésima distribución marginal de μ_α , es decir, $\mu_\alpha^n = \mu_\alpha \pi_n^{-1}$, donde $\pi_n(x) = x_n$ es la proyección sobre E_n . Entonces, $\{\mu_\alpha\}_{\alpha \in \mathcal{I}}$ es tensa si y sólo si $\{\mu_\alpha^n\}_{\alpha \in \mathcal{I}}$ es tensa para cada n .*

1.1.2. Topología de Skorokhod para trayectorias «càdlàg».

Para estudiar límites de procesos estocásticos, no es suficiente considerar la convergencia en tiempos fijos; es necesario definir qué se entiende por convergencia a nivel de procesos. Para ello, se aprovecha que la mayoría de los procesos estocásticos pueden modificarse para obtener procesos con trayectorias continuas por la derecha y con límites por la izquierda (sin modificar sus distribuciones finito-dimensionales).

Sea (E, r) un espacio métrico. Se denotará por $\mathcal{D}_E[0, \infty)$ al espacio de funciones «càdlàg» (abreviación del francés «continue à droite avec des limites à gauche») en E . Es decir, si $x \in \mathcal{D}_E[0, \infty)$, entonces $x : [0, \infty) \rightarrow E$ y para cada $t \geq 0$ se tiene:

- (a) Continuidad por la derecha

$$\lim_{s \rightarrow t^+} x(s) = x(t).$$

- (b) Límites por la izquierda

$$\lim_{s \rightarrow t^-} x(s) \equiv x(t^-).$$

Por convención $\lim_{s \rightarrow 0^-} x(s) \equiv x(0^-) = x(0)$.

Lema 1.1.13. Si $x \in \mathcal{D}_E[0, \infty)$, entonces x tiene a lo más una cantidad numerable de puntos de discontinuidad.

Definición 1.1.14 (Convergencia de Skorokhod). Se dice que una sucesión $\{x_n\}_{n \geq 0} \subset \mathcal{D}_E[0, \infty)$ converge a $x \in \mathcal{D}_E[0, \infty)$ en el sentido de Skorokhod o en sentido funcional, si y solo si cada sucesión $\{t_n\}_{n \geq 0} \subset [0, \infty)$ tal que $\lim_{n \rightarrow \infty} t_n = t$, satisface las siguientes condiciones:

(a) La sucesión $\{x_n(t_n)\}$ tiene a lo más dos puntos límite:

$$\lim_{n \rightarrow \infty} \min\{r(x_n(t_n), x(t)), r(x_n(t_n), x(t^-))\} = 0.$$

(b) Continuidad por la derecha de x : Si $\lim_{n \rightarrow \infty} r(x_n(t_n), x(t)) = 0$, entonces para cualquier sucesión $\{s_n\}$ de tiempos tal que $s_n \geq t_n$ y $\lim_{n \rightarrow \infty} s_n = t$, se tiene que

$$\lim_{n \rightarrow \infty} r(x_n(s_n), x(t)) = 0.$$

(c) Límites por la izquierda de x : Si $\lim_{n \rightarrow \infty} r(x_n(t_n), x(t^-)) = 0$, entonces para cualquier sucesión $\{s_n\}$ de tiempos tal que $0 \leq s_n \leq t_n$ y $\lim_{n \rightarrow \infty} s_n = t$, se tiene

$$\lim_{n \rightarrow \infty} r(x_n(s_n), x(t^-)) = 0.$$

La topología que define esta convergencia es conocida como la topología de Skorokhod (también escrito Skorohod [EK86]) y es inducida por una métrica d que se define a continuación.

Sea Λ' el conjunto de funciones $\lambda : [0, \infty) \rightarrow [0, \infty)$ estrictamente crecientes. En particular, se considera las funciones $\lambda \in \Lambda'$ tales que λ es continua, $\lambda(0) = 0$ y $\lim_{t \rightarrow \infty} \lambda(t) = \infty$. Sean $q := r \wedge 1$ y $\Lambda \subset \Lambda'$ el conjunto de funciones Lipschitz continuas tales que

$$\gamma(\lambda) := \sup_{s > t \geq 0} \left| \log \frac{\lambda(s) - \lambda(t)}{s - t} \right| < \infty. \quad (1.8)$$

Para $x, y \in \mathcal{D}_E[0, \infty)$, se define

$$d(x, y) = \inf_{\lambda \in \Lambda} \left[\gamma(\lambda) \vee \int_0^\infty e^{-u} d(x, y, \lambda, u) du \right], \quad (1.9)$$

donde

$$d(x, y, \lambda, u) = \sup_{t \geq 0} q(x(t \wedge u), y(\lambda(t) \wedge u)). \quad (1.10)$$

Proposición 1.1.15. Sean $\{x_n\}_{n \geq 0} \subset \mathcal{D}_E[0, \infty)$ y $x \in \mathcal{D}_E[0, \infty)$. Entonces, las siguientes cuatro afirmaciones son equivalentes:

(a) $\lim_{n \rightarrow \infty} d(x_n, x) = 0$.

(b) x_n converge a x en el sentido de Skorokhod.

(c) Existe $\{\lambda_n\}_{n \geq 0} \subset \Lambda$ tal que

$$\lim_{n \rightarrow \infty} \gamma(\lambda_n) = 0 \quad (1.11)$$

y para toda $T > 0$

$$\lim_{n \rightarrow \infty} \sup_{0 \leq t \leq T} r(x_n(t), x(\lambda_n(t))) = 0. \quad (1.12)$$

(d) Para cada $T > 0$, existe $\{\lambda_n\}_{n \geq 0} \subset \Lambda'$ (posiblemente dependiente de T) tal que

$$\lim_{n \rightarrow \infty} \sup_{0 \leq t \leq T} |\lambda_n(t) - t| = 0, \quad (1.13)$$

$$\lim_{n \rightarrow \infty} \sup_{0 \leq t \leq T} r(x_n(t), x(\lambda_n(t))) = 0. \quad (1.14)$$

De este modo, la métrica de Skorokhod permite convertir el problema de convergencia para sucesiones de trayectorias «càdlàg» en un problema de convergencia en espacios métricos, conservando algunas propiedades.

Teorema 1.1.16. Si E es separable, entonces $\mathcal{D}_E[0, \infty)$ también lo es. Además, si (E, r) es completo, entonces $(\mathcal{D}_E[0, \infty), d)$ es completo.

Una variable aleatoria $X = \{X_t\}_{t \geq 0}$ con valores en $\mathcal{D}_E[0, \infty)$ es un proceso estocástico con trayectorias en $\mathcal{D}_E[0, \infty)$ (aunque si E no es separable el inverso no es necesariamente cierto). Para estudiar la convergencia en distribución de dichos procesos o equivalentemente, la convergencia débil en $\mathcal{P}(\mathcal{D}_E[0, \infty))$, es importante tener más información acerca de la σ -álgebra de Borel para $\mathcal{D}_E[0, \infty)$ asociada a la topología de Skorokhod. El siguiente resultado establece que, cuando E es separable, $\mathcal{B}(\mathcal{D}_E[0, \infty))$ es simplemente la σ -álgebra generada por las proyecciones π_t para cada tiempo $t \geq 0$ (que pueden ser interpretadas como las «coordenadas» del proceso, pues $\pi_t(X) = X_t$).

Proposición 1.1.17. Para cada $t \geq 0$, sea $\pi_t : \mathcal{D}_E[0, \infty) \rightarrow E$, la proyección definida por $\pi_t(x) = x(t)$. Entonces, para cualquier $D \subset [0, \infty)$ denso

$$\mathcal{B}(\mathcal{D}_E[0, \infty)) \supset \sigma(\pi_t : 0 \leq t < \infty) = \sigma(\pi_t : t \in D), \quad (1.15)$$

además, si E es separable, $\mathcal{B}(\mathcal{D}_E[0, \infty)) = \sigma(\pi_t : t \in D)$.

Sea $\{X_\alpha\}_{\alpha \in \mathcal{I}}$, con \mathcal{I} un conjunto de índices, una familia de procesos estocásticos con trayectorias en $\mathcal{D}_E[0, \infty)$. Si E no es separable, supóngase que cada X_α es una variable aleatoria con valores en $\mathcal{D}_E[0, \infty)$. Denótese por $\{\mu_\alpha\}_{\alpha \in \mathcal{I}} \subset \mathcal{P}(\mathcal{D}_E[0, \infty))$ la familia de distribuciones de probabilidad asociadas. Es decir, $\mu_\alpha(B) = \mathbb{P}[X_\alpha \in B]$ para toda $B \in \mathcal{B}(\mathcal{D}_E[0, \infty))$. Se dice que $\{X_\alpha\}$ es relativamente compacta si $\{\mu_\alpha\}$ lo es, es decir, si la cerradura de $\{\mu_\alpha\}$ en $\mathcal{P}(\mathcal{D}_E[0, \infty))$ es compacta. Los siguientes teoremas presentan condiciones para la compacidad relativa de $\{\mu_\alpha\}$, en términos de un *módulo de continuidad* w' .

Para cada $x \in \mathcal{D}_E[0, \infty)$, $\delta > 0$ y $T > 0$ se define

$$w'(x, \delta, T) = \inf_{\{t_i\}} \max_i \sup_{s, t \in [t_{i-1}, t_i]} r(x(s), x(t)), \quad (1.16)$$

donde $\{t_i\}$ varía sobre las particiones de la forma $0 = t_0 < t_1 < \dots < t_m = T$, con $m \geq 1$ y $\min_{1 \leq i \leq m} (t_i - t_{i-1}) > \delta$. Entonces $w'(x, \delta, T)$ es no decreciente en δ y T , continuo por la derecha en δ y

$$\lim_{\delta \rightarrow 0^+} w'(x, \delta, T) = 0. \quad (1.17)$$

Teorema 1.1.18. Sea (E, r) un espacio métrico completo. Entonces, un conjunto $A \subset \mathcal{D}_E[0, \infty)$ es relativamente compacto si y solo si satisface las siguientes condiciones:

- (a) Para cada $t \geq 0$ racional, existe un conjunto compacto $\Gamma_t \subset E$ tal que $x(t) \in \Gamma_t$ para toda $x \in A$.
- (b) Para cada $T > 0$,

$$\lim_{\delta \rightarrow 0^+} \sup_{x \in A} w'(x, \delta, T) = 0. \quad (1.18)$$

Teorema 1.1.19. Sean (E, r) completo y separable y $\{X_\alpha\}_{\alpha \in \mathcal{I}}$ una familia de procesos con trayectorias en $\mathcal{D}_E[0, \infty)$. Entonces $\{X_\alpha\}$ es relativamente compacta si y solo si satisface las siguientes condiciones:

(a) Para cada $\epsilon > 0$ y $t \geq 0$ racional, existe un compacto $\Gamma_{\epsilon,t} \subset E$ tal que

$$\inf_{\alpha} \mathbb{P}[X_{\alpha}(t) \in \Gamma_{\epsilon,t}] \geq 1 - \epsilon. \quad (1.19)$$

(b) Para cada $\epsilon > 0$ y $T > 0$, existe $\delta > 0$ tal que

$$\sup_{\alpha} \mathbb{P}[w'(X_{\alpha}, \delta, T) \geq \epsilon] \leq \epsilon. \quad (1.20)$$

Es difícil trabajar con las condiciones establecidas por estos teoremas. Por lo tanto, en ocasiones resulta útil llevar las condiciones de compacidad relativa al ámbito de los procesos con valores reales, a través de los siguientes teoremas.

Teorema 1.1.20. Sea $\{X_n\}_{n \geq 1} \subset \mathcal{D}_{\mathbb{R}}[0, \infty)$, tal que:

(a) Para cada $t \geq 0$, $\{X_n(t)\}_{n \geq 1}$ es tensa.

(b) Dada una sucesión de tiempos de paro τ_n , acotada por T , para cada $\epsilon > 0$ existen $\delta > 0$ y n_0 tales que

$$\sup_{n \geq n_0} \sup_{t \in [0, \delta]} \mathbb{P}[|X_n(\tau_n + t) - X_n(\tau_n)| > \epsilon] \leq \epsilon. \quad (1.21)$$

Entonces $\{X_n\}_{n \geq 1}$ es tensa.

Teorema 1.1.21. Sean (E, r) completo y separable y $\{X_n\}_{n \geq 1}$ una sucesión de procesos con trayectorias en $\mathcal{D}_E[0, \infty)$. Supóngase que se cumple la siguiente condición de contención compacta: para cada $\epsilon > 0$ y cada $T > 0$ existe un compacto $\Gamma_{\epsilon,T} \subset E$ tal que

$$\inf_n \mathbb{P}[X_n(t) \in \Gamma_{\epsilon,T} : 0 \leq t \leq T] \geq 1 - \epsilon. \quad (1.22)$$

Sea $C \subset \mathcal{C}_b(E)$ denso (con la topología de convergencia uniforme sobre compactos). Entonces $\{X_n\}_{n \geq 1}$ es relativamente compacta si y solo si para cada $f \in C$, $\{f(X_n)\}_{n \geq 1}$ es relativamente compacta como familia de procesos en $\mathcal{D}_{\mathbb{R}}[0, \infty)$.

1.2. Procesos de Markov.

1.2.1. Semigrupos y generadores.

A lo largo de esta sección, \mathcal{L} representa un espacio de Banach, es decir, un espacio vectorial normado y completo con respecto a la topología inducida por la norma.

Definición 1.2.1 (Semigrupo). Una familia uniparamétrica $\{T(t) : t \geq 0\}$ de operadores lineales acotados en \mathcal{L} (es decir, $T(t) : \mathcal{L} \rightarrow \mathcal{L}$ para $t \geq 0$) se llama semigrupo si $T(0) = I$ (el operador identidad) y para cualesquiera $s, t \geq 0$, $T(s+t) = T(s)T(t)$.

Definición 1.2.2 (Semigrupo de contracción). Sea $\{T(t) : t \geq 0\}$ un semigrupo en \mathcal{L} . Se dice que es de contracción si $\|T(t)\| \leq 1$ para toda $t \geq 0$, donde $\|T(t)\| := \sup\{\|T(t)f\| : f \in \mathcal{L}, \|f\| \leq 1\}$.

Definición 1.2.3 (Semigrupo fuertemente continuo). Un semigrupo $\{T(t) : t \geq 0\}$ en \mathcal{L} se dice fuertemente continuo si $\lim_{t \rightarrow 0} T(t)f = f$ para cada $f \in \mathcal{L}$.

Proposición 1.2.4. Si $\{T(t) : t \geq 0\}$ es un semigrupo de contracción fuertemente continuo en \mathcal{L} , entonces para cada $f \in \mathcal{L}$, $t \mapsto T(t)f$ es una función continua de $[0, \infty)$ en \mathcal{L} .

Un operador lineal A en \mathcal{L} es un mapeo lineal con rango $\mathcal{R}(A) \subset \mathcal{L}$ y cuyo dominio, $\mathcal{D}(A)$ es un subespacio de \mathcal{L} . La gráfica de A está dada por

$$\mathcal{G}(A) := \{(f, Af) : f \in \mathcal{D}(A)\} \subset \mathcal{L} \times \mathcal{L}. \quad (1.23)$$

El espacio $\mathcal{L} \times \mathcal{L}$ es a su vez de Banach, con suma y producto escalar definidos componente a componente y con norma $\|(f, g)\| = \|f\| + \|g\|$.

Definición 1.2.5 (Operador lineal cerrado). Un operador lineal A en \mathcal{L} se dice cerrado si su gráfica $\mathcal{G}(A)$ es un subespacio cerrado de $\mathcal{L} \times \mathcal{L}$.

Definición 1.2.6 (Generador infinitesimal). El generador infinitesimal de un semigrupo $\{T(t)\}$, también llamado simplemente el generador, es un operador lineal A definido por

$$Af = \lim_{t \rightarrow 0} \frac{1}{t} \{T(t)f - f\}, \quad (1.24)$$

cuyo dominio $\mathcal{D}(A)$ es el subespacio de todas las $f \in \mathcal{L}$ tales que este límite existe.

Definición 1.2.7 (Operador disipativo). Se dice que un operador lineal A sobre \mathcal{L} es disipativo si $\|\lambda f - Af\| \geq \lambda \|f\|$ para cada $f \in \mathcal{D}(A)$ y $\lambda > 0$.

Definición 1.2.8 (Semigrupo medible). Un semigrupo $\{T(t) : t \geq 0\}$ en \mathcal{L} se dice medible si para cada $f \in \mathcal{L}$, $T(\cdot)f$ es medible como función en $([0, \infty), \mathcal{B}([0, \infty)))$.

Definición 1.2.9 (Generador completo). Se define el generador completo de un semigrupo de contracción medible $\{T(t)\}$ en \mathcal{L} como el conjunto

$$\hat{A} = \left\{ (f, g) \in \mathcal{L} \times \mathcal{L} : T(t)f - f = \int_0^t T(s)g ds, \quad t \geq 0 \right\}. \quad (1.25)$$

En general, \hat{A} puede ser multivaluado. Por ejemplo, si $\mathcal{L} = \mathcal{B}(\mathbb{R})$, el espacio de funciones acotadas en \mathbb{R} con la norma del supremo y se define $T(t)f(x) = f(x+t)$, entonces $(0, g) \in \hat{A}$ para toda $g \in \mathcal{B}(\mathbb{R})$ tal que $g(x) = 0$ casi seguramente (c.s.) con respecto a la medida de Lebesgue.

Sea (E, r) un espacio métrico. La bp-cerradura de $M \subset \mathcal{B}(E)$, una colección de funciones acotadas sobre E , es el menor subconjunto $\bar{M} \subset \mathcal{B}(E)$ para el cual $M \subset \bar{M}$ y para cada $\{f_n\} \subset \bar{M}$ tal que $\sup_n \|f_n\| < \infty$ y $\lim_{n \rightarrow \infty} f_n(x) = f(x)$ con $f \in \mathcal{B}(E)$, se tiene que $f \in \bar{M}$. El prefijo «bp» corresponde a la abreviatura del inglés para acotada y puntual («bounded-pointwise-closure»).

Definición 1.2.10 (Operador conservativo). Se dice que un operador $A \subset \mathcal{B}(E) \times \mathcal{B}(E)$ es conservativo si $(1, 0)$ pertenece a la bp-cerradura de A .

Considérese un espacio métrico (E, r) separable y localmente compacto (cada punto tiene una vecindad contenida en un compacto). Sea $\mathcal{C}(E)$ el espacio de funciones continuas que se desvanecen en el infinito, es decir, $f \in \mathcal{C}(E)$ si para toda $\epsilon > 0$ existe $K \subset E$ compacto tal que $\|f(x)\| < \epsilon$ para cualquier $x \notin K$. Entonces $\mathcal{C}(E)$, con la norma del supremo, es un espacio de Banach.

Definición 1.2.11 (Semigrupo positivo). Un semigrupo $\{T(t) : t \geq 0\}$ en $\mathcal{C}(E)$ es positivo si $T(t)$ es un operador positivo para toda $t \geq 0$. Es decir, si $T(t)$ lleva funciones no negativas a funciones no negativas.

Definición 1.2.12 (Semigrupo de Feller). Un semigrupo de contracción $\{T(t)\}$ en $\mathcal{C}(E)$, fuertemente continuo, positivo y cuyo generador es conservativo, se llama semigrupo de Feller.

El siguiente resultado permite caracterizar la clase de operadores lineales que son generadores de semigrupos de Feller.

Teorema 1.2.13. *Un operador lineal A en \mathcal{L} es el generador de un semigrupo de contracción fuertemente continuo sobre \mathcal{L} si y solo si:*

- (a) $\mathcal{D}(A)$ es denso en \mathcal{L} .

(b) A es disipativo.

(c) $\mathcal{R}(\lambda - A) = \mathcal{L}$ para alguna $\lambda > 0$.

Proposición 1.2.14. Sean $\{T(t)\}$ y $\{S(t)\}$ dos semigrupos de contracción fuertemente continuos sobre \mathcal{L} , con generadores A y B respectivamente. Si $A = B$, entonces $T(t) = S(t)$ para toda $t \geq 0$.

Por lo tanto, un semigrupo de Feller queda determinado de manera única por su generador.

1.2.2. Procesos de Markov.

Definición 1.2.15 (Proceso de Markov). Sea $\{X(t), t \geq 0\}$ un proceso estocástico definido en un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$ y con valores en un espacio de estados E . Sea $\mathcal{F}_t^X = \sigma(X(s) : s \leq t)$, entonces X es un Proceso de Markov si

$$\mathbb{P}[X(t+s) \in \Gamma \mid \mathcal{F}_t^X] = \mathbb{P}[X(t+s) \in \Gamma \mid X(t)], \quad (1.26)$$

para cualesquiera $s, t \geq 0$ y $\Gamma \in \mathcal{B}(E)$.

Nótese que la condición (1.26) implica

$$\mathbb{E}[f(X(t+s)) \mid \mathcal{F}_t^X] = \mathbb{E}[f(X(t+s)) \mid X(t)], \quad (1.27)$$

para cualesquiera $s, t \geq 0$ y $f \in \mathcal{B}(E)$.

Definición 1.2.16 (Función de transición). Una función $P(t, x, \Gamma)$ definida en $[0, \infty) \times E \times \mathcal{B}(E)$ es una función de transición homogénea en el tiempo si:

- (a) $P(t, x, \cdot) \in \mathcal{P}(E)$, para $(t, x) \in [0, \infty) \times E$.
- (b) $P(0, x, \cdot) = \delta_x$, para $x \in E$, donde $\delta_x \in \mathcal{P}(E)$ denota la medida de probabilidad que acumula toda su masa en x .
- (c) $P(\cdot, \cdot, \Gamma) \in \mathcal{B}([0, \infty) \times E)$, para $\Gamma \in \mathcal{B}(E)$.
- (d) Satisface las ecuaciones de Chapman-Kolmogorov:

$$P(t+s, x, \Gamma) = \int_E P(s, y, \Gamma) P(t, x, dy), \quad (1.28)$$

para cualesquiera $s, t \geq 0$, $x \in E$ y $\Gamma \in \mathcal{B}(E)$.

En el caso en que $E = \mathbb{R}^k$, se puede expresar las ecuaciones de Chapman-Kolmogorov en términos de integrales con respecto a λ , la medida de Lebesgue, como

$$P(t+s, x, \Gamma) = \int_{\mathbb{R}} \dots \int_{\mathbb{R}} P(s, y, \Gamma) P(t, x, y) d\lambda(y_1) \dots d\lambda(y_k), \quad (1.29)$$

donde $y = (y_1, \dots, y_k)$. Cuando E es un espacio más general, la solución de estas ecuaciones se vuelve más compleja, pues involucra una integral con respecto a una medida en $\mathcal{P}(E)$.

Una función de transición $P(t, x, \Gamma)$ es la función de transición para un proceso de Markov X homogéneo en el tiempo, si

$$\mathbb{P}[X(t+s) \in \Gamma \mid \mathcal{F}_t^X] = P(s, X(t), \Gamma), \quad (1.30)$$

para cualesquiera $s, t \geq 0$ y $\Gamma \in \mathcal{B}(E)$. O equivalentemente, si

$$\mathbb{E}[f(X(t+s)) \mid \mathcal{F}_t^X] = \int f(y) P(s, X(t), dy), \quad (1.31)$$

para $s, t \geq 0$ y $f \in \mathcal{B}(E)$.

La medida de probabilidad $\nu \in \mathcal{P}(E)$ dada, para cada $\Gamma \in \mathcal{B}(E)$, por $\nu(\Gamma) = \mathbb{P}[X(0) \in \Gamma]$ se conoce como *distribución inicial* de X .

Una función de transición $P(t, x, \Gamma)$ y una distribución inicial $\nu \in \mathcal{P}(E)$ determinan de manera única las distribuciones finito-dimensionales de un proceso de Markov X , como:

$$\begin{aligned} \mathbb{P}[X(0) \in \Gamma_0, X(t_1) \in \Gamma_1, \dots, X(t_n) \in \Gamma_n] = \\ \int_{\Gamma_0} \dots \int_{\Gamma_n} P(t_n - t_{n-1}, y_{n-1}, \Gamma_n) P(t_{n-1} - t_{n-2}, y_{n-2}, dy_{n-1}) \\ \dots P(t_1, y_0, dy_1) \nu(dy_0) \end{aligned} \quad (1.32)$$

para cualesquiera $t_0, \dots, t_n \geq 0$ y $\Gamma_0, \dots, \Gamma_n \in \mathcal{B}(E)$.

El siguiente teorema establece la existencia de dicho proceso.

Teorema 1.2.17. Sean $P(t, x, \Gamma)$ una función de transición y $\nu \in \mathcal{P}(E)$. Si para cada $t \geq 0$ la medida de probabilidad $\int P(t, x, \cdot) \nu(dx)$ es tensa (lo cual es siempre cierto, si (E, r) es completo y separable, de acuerdo con el lema 1.1.7), entonces existe un proceso de Markov X , con valores en E , cuyas distribuciones finito-dimensionales están determinadas de manera única por la ecuación (1.32).

Si P_x denota la medida definida en el teorema anterior, con $\nu = \delta_x$ y X es el correspondiente proceso canónico, es decir $X(t, \omega) = \omega(t)$, se sigue de la ecuación (1.32) que para cualquier $B = \{X(0) \in \Gamma_0, \dots, X(t_n) \in \Gamma_n\}$, donde $0 < t_1 < \dots < t_n$ y $\Gamma_0, \dots, \Gamma_n \in \mathcal{B}(E)$, $P_x(B)$ es una función Borel-medible de x . El siguiente teorema establece que esto es cierto para toda $B \in \mathcal{B}(E)^{[0, \infty)}$.

Proposición 1.2.18. *Sea P_x la medida en $\mathcal{B}(E)^{[0, \infty)}$ dada en el teorema 1.2.17, con $\nu = \delta_x$. Entonces $P_x(B)$ es una función Borel-medible para cada $B \in \mathcal{B}(E)^{[0, \infty)}$.*

Por lo tanto, si (E, r) es completo y separable, una función de transición $P(t, x, \Gamma)$ define, para cada $x \in E$, un proceso de Markov X_x que empieza en x ($X_x(0) = x$) c.s.

Definición 1.2.19 (Cadena de Markov y función de transición a un paso). Sea $\{X_k\}_{k \geq 1}$ un proceso estocástico a tiempo discreto definido en un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$ y con valores en E . Sea $\mathcal{F}_k^X = \sigma(X_j : j \leq k)$. Entonces X es una cadena de Markov si

$$\mathbb{P}[X_{k+j} \in \Gamma \mid \mathcal{F}_k^X] = \mathbb{P}[X_{k+j} \in \Gamma \mid X_k], \quad (1.33)$$

para cualesquiera $j, k \geq 0$ y $\Gamma \in \mathcal{B}(E)$.

Una función $P(x, \Gamma)$ definida en $E \times \mathcal{B}(E)$ es una función de transición a un paso si

$$P(x, \cdot) \in \mathcal{P}(E), \quad x \in E, \quad (1.34)$$

$$P(\cdot, \Gamma) \in \mathcal{B}(E), \quad \Gamma \in \mathcal{B}(E). \quad (1.35)$$

Una función de transición a un paso $P(x, \Gamma)$ es la función de transición para una cadena de Markov X , homogénea en el tiempo si

$$\mathbb{P}[X_{k+1} \in \Gamma \mid \mathcal{F}_k^X] = P(X_k, \Gamma), \quad (1.36)$$

para cada $k \geq 0$ y $\Gamma \in \mathcal{B}(E)$.

Nótese que

$$\mathbb{P}[X_{k+j} \in \Gamma \mid \mathcal{F}_k^X] = \int \dots \int P(y_{j-1}, \Gamma) P(y_{j-2}, dy_{j-1}) \dots P(X_k, dy_1) \quad (1.37)$$

Resultados equivalentes al teorema 1.2.17 y la proposición 1.2.18 se satisfacen para cadenas de Markov.

Definición 1.2.20 (Proceso adaptado y progresivo). Sea $\{\mathcal{F}_t\}$ una filtración. Un proceso X es $\{\mathcal{F}_t\}$ -adaptado si $X(t)$ es \mathcal{F}_t -medible para cada $t \geq 0$. Si además, para cada $t \geq 0$ la restricción de X a $[0, t] \times (\Omega, \mathcal{F}, \mathbb{P})$ es $(\mathcal{B}([0, t]) \times \mathcal{F}_t)$ -medible, entonces X es $\{\mathcal{F}_t\}$ -progresivo.

En general, no es fácil encontrar expresiones correspondientes a funciones de transición. Como consecuencia, definir procesos de Markov a través de éstas no es siempre práctico. Se suele recurrir a las propiedades de semigrupos, dado que existe una relación entre las funciones de transición homogéneas en el tiempo y los semigrupos de operadores.

Si $P(t, x, \Gamma)$ es una función de transición homogénea en el tiempo, entonces

$$T(t)f(x) = \int f(y)P(t, x, dy) \quad (1.38)$$

define un semigrupo de contracción medible en $\mathcal{B}(E)$, gracias a la propiedad de Chapman-Kolmogorov (1.28).

Definición 1.2.21 (Proceso de Markov correspondiente a un semigrupo). Sea $\{T(t)\}$ un semigrupo sobre un subespacio cerrado $\mathcal{L} \subset \mathcal{B}(E)$. Se dice que un proceso de Markov X con valores en E corresponde a $\{T(t)\}$ si para cualesquiera $s, t \geq 0$ y $f \in \mathcal{L}$,

$$\mathbb{E}[f(X(t+s)) | \mathcal{F}_t^X] = T(s)f(X(t)). \quad (1.39)$$

Si $\{T(t)\}$ está definido a través de una función de transición como en la expresión (1.38), entonces esta condición se convierte en la igualdad (1.27), que caracteriza a los procesos de Markov.

Proposición 1.2.22. *Sea X un proceso de Markov con valores en un espacio separable E , con distribución inicial $\nu \in \mathcal{P}(E)$ y correspondiente a un semigrupo $\{T(t)\}$ sobre un subespacio cerrado $\mathcal{L} \subset \mathcal{B}(E)$. Supóngase que \mathcal{L} separa puntos, es decir, para cualesquiera $\mu_1, \mu_2 \in \mathcal{P}(E)$ tales que $\mu_1 \neq \mu_2$, existe $f \in \mathcal{L}$ tal que $\int f d\mu_1 \neq \int f d\mu_2$. Entonces $\{T(t)\}$ y ν determinan de manera única las distribuciones finito-dimensionales de X .*

1.2.3. El problema de martingala.

La última proposición de la sección anterior establece que, bajo ciertas condiciones, las distribuciones finito-dimensionales de un proceso de Markov X , quedan determinadas de manera única por una distribución inicial y el generador correspondiente $\{T(t)\}$. Éste a su vez, queda determinado por su generador completo \hat{A} o por un subconjunto «suficientemente grande»

$A \subset \hat{A}$. Para saber qué significa que un conjunto sea suficientemente grande, se puede recurrir al problema de martingala de Stroock y Varadhan, basado en el siguiente resultado.

Proposición 1.2.23. Sean $\{\mathcal{F}_t\}$ una filtración y X un proceso de Markov $\{\mathcal{F}_t\}$ -progresivo con valores en E y función de transición $P(x, t, \Gamma)$. Sea $\{T(t)\}$ el semigrupo correspondiente a X , con generador completo \hat{A} . Si $(f, g) \in \hat{A}$ entonces

$$M(t) := f(X(t)) - \int_0^t g(X(s)) ds \quad (1.40)$$

es una $\{\mathcal{F}_t^X\}$ -martingala.

La idea de Stroock y Varadhan consiste en usar esta propiedad de martingala como medio para caracterizar el proceso de Markov asociado con un generador A dado.

En lo que resta de esta sección, los procesos tomarán valores en un espacio métrico (E, r) . En ocasiones, A denotará un operador multivaluado, por lo que se puede pensar en él como un subconjunto, no necesariamente lineal, de $\mathcal{B}(E) \times \mathcal{B}(E)$.

Una *solución al problema de martingala para A* es un proceso estocástico X con valores en E y definido en un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$, tal que para cada $(f, g) \in A$, la igualdad (1.40) define una martingala con respecto a la filtración

$$*\mathcal{F}_t^X := \mathcal{F}_t^X \vee \sigma \left(\int_0^s h(X(u)) du : s \leq t, h \in \mathcal{B}(E) \right). \quad (1.41)$$

En general, cada evento en $*\mathcal{F}_t^X$ difiere de un evento en \mathcal{F}_t^X por un evento de probabilidad cero. Si X es progresivo, en particular si es continuo por la derecha, entonces $*\mathcal{F}_t^X = \mathcal{F}_t^X$.

Si $\{\mathcal{G}_t\}$ es una filtración, con $\mathcal{G}_t \supset *\mathcal{F}_t^X$ para cada $t \geq 0$ y el proceso $\{M(t)\}_{t \geq 0}$ definido por la expresión (1.40) es una $\{\mathcal{G}_t\}$ -martingala para cada $(f, g) \in A$, se dice que X es una *solución al problema de martingala para A , con respecto a $\{\mathcal{G}_t\}$* . Cuando se especifica una distribución inicial $\nu \in \mathcal{P}(E)$, se dice que una solución X al problema de martingala para A es *solución al problema de martingala para (A, ν)* si ν es la distribución de $X(0)$.

En general, se puede pensar que X tiene trayectorias en $\mathcal{D}_E[0, \infty)$. Se dice que una medida de probabilidad $\mu \in \mathcal{D}_E[0, \infty)$ es *solución al problema de martingala para A* (o para (A, ν)) si el *proceso de coordenadas* o *proceso canónico*, definido en $(\mathcal{D}_E[0, \infty), \mathcal{B}(\mathcal{D}_E[0, \infty)), \mu)$ por

$$X(t, \omega) := \omega(t), \quad \omega \in \mathcal{D}_E[0, \infty), \quad t \geq 0, \quad (1.42)$$

es solución al problema de martingala para A (o para (A, ν)).

Un proceso medible X es solución al problema de martingala para A si y sólo si

$$\begin{aligned} 0 &= \mathbb{E} \left[\left(f(X(t_{n+1})) - f(X(t_n)) - \int_{t_n}^{t_{n+1}} g(X(s)) ds \right) \prod_{k=1}^n h_k(X(t_k)) \right] \\ &= \mathbb{E} \left[f(X(t_{n+1})) \prod_{k=1}^n h_k(X(t_k)) \right] - \mathbb{E} \left[f(X(t_n)) \prod_{k=1}^n h_k(X(t_k)) \right] \\ &\quad - \int_{t_n}^{t_{n+1}} \mathbb{E} \left[g(X(s)) \prod_{k=1}^n h_k(X(t_k)) \right] ds, \end{aligned}$$

para cualesquiera $0 \leq t_1 < t_2 < \dots < t_{n+1}$, $(f, g) \in A$ y $h_1, \dots, h_n \in \mathcal{B}(E)$ (o equivalentemente en $\mathcal{C}(E)$). Por lo tanto, decir que un proceso es solución a un problema de martingala, es una afirmación acerca de sus distribuciones finito-dimensionales. En particular, cualquier modificación medible de una solución a un problema de martingala para A es también solución. Además, si $A_1 \subset A_2$, entonces cualquier solución al problema de martingala para A_2 es también solución para A_1 , aunque el inverso no es necesariamente cierto.

Se dice que las soluciones al problema de martingala para (A, ν) satisfacen la *condición de unicidad* si cualesquiera dos soluciones tienen las mismas distribuciones finito-dimensionales, lo cual es únicamente una formalización de la idea de igualdad de procesos.

Definición 1.2.24 (Problema de martingala bien planteado). Se dice que el problema de martingala para (A, ν) está bien planteado si existe una solución y se satisface la condición de unicidad. Si esto es cierto para toda $\nu \in \mathcal{P}(E)$, se dice que el problema de martingala para A está bien planteado.

Se dice que el problema de martingala para (A, ν) está bien planteado en $\mathcal{D}_E[0, \infty)$ (o alternativamente en $\mathcal{C}_E[0, \infty)$, el espacio de trayectorias continuas) si existe una solución única con distribución $\mu \in \mathcal{P}(\mathcal{D}_E[0, \infty))$ ($\mu \in \mathcal{P}(\mathcal{C}_E[0, \infty))$).

Usualmente, si el problema de martingala para (A, δ_x) está bien planteado para cada $x \in E$, entonces el problema de martingala para (A, ν) está bien planteado para toda $\nu \in \mathcal{P}(E)$. Nótese que un problema de martingala puede estar bien planteado en $\mathcal{D}_E[0, \infty)$, sin estar bien planteado en general, ya que la unicidad puede darse únicamente bajo la restricción de que la solución tenga trayectorias «càdlàg».

El siguiente teorema establece en esencia, que un proceso de Markov se puede caracterizar como la solución única al problema de martingala para su generador.

Teorema 1.2.25. *Sea $A \subset \mathcal{B}(E) \times \mathcal{B}(E)$ un operador lineal disipativo, con (E, r) métrico y separable. Supóngase que existe un operador lineal $A' \subset A$ tal que $\mathcal{L} := \mathcal{D}(A') = \mathcal{R}(\lambda - A')$ para alguna $\lambda > 0$ y \mathcal{L} separa puntos. Sea $\nu \in \mathcal{P}(E)$ y supóngase que X es una solución al problema de martingala para (A, ν) . Entonces X es un proceso de Markov correspondiente al semigrupo en \mathcal{L} generado por la cerradura de A' y el problema de martingala para (A, ν) satisface la condición de unicidad.*

Bajo las condiciones del teorema anterior, cada solución al problema de martingala para A es un proceso de Markov. El siguiente resultado establece que la unicidad de soluciones para el problema de martingala implica la propiedad de Markov.

Teorema 1.2.26. *Sean E separable y $A \subset \mathcal{B}(E) \times \mathcal{B}(E)$. Supóngase que para cada $\nu \in \mathcal{P}(E)$ cualesquiera dos soluciones X, Y del problema de martingala para (A, ν) tienen las mismas distribuciones unidimensionales, es decir, para cada $t > 0$,*

$$\mathbb{P}[X(t) \in \Gamma] = \mathbb{P}[Y(t) \in \Gamma], \quad \Gamma \in \mathcal{B}(E) \quad (1.43)$$

Entonces, cualquier solución del problema de martingala para A con respecto a una filtración $\{\mathcal{G}_t\}$ es un proceso de Markov con respecto a $\{\mathcal{G}_t\}$ y cualesquiera dos soluciones al problema de martingala para (A, ν) tienen las mismas distribuciones finito-dimensionales. Es decir, la condición (1.43) implica la unicidad.

Corolario 1.2.27. *Sean E separable y $A \subset \mathcal{B}(E) \times \mathcal{B}(E)$. Supóngase que para cada $\nu \in \mathcal{P}(E)$, cualesquiera dos soluciones X, Y al problema de martingala para (A, ν) con trayectorias en $\mathcal{D}_E[0, \infty)$ (o en $\mathcal{C}_E[0, \infty)$) satisfacen la condición (1.43) para cada $t \geq 0$. Entonces, para cada $\nu \in \mathcal{P}(E)$, cualesquiera dos soluciones al problema de martingala para (A, ν) con trayectorias en $\mathcal{D}_E[0, \infty)$ ($\mathcal{C}_E[0, \infty)$) tienen la misma distribución en $\mathcal{D}_E[0, \infty)$ ($\mathcal{C}_E[0, \infty)$).*

Las condiciones para la existencia de soluciones a un problema de martingala requieren un desarrollo más extenso, que rebasa los propósitos del presente trabajo. Por tanto, sólo se presentará un último resultado en esta sección, que muestra una forma simple de obtener una solución a un problema de martingala como límite débil de soluciones de problemas de martingala que lo aproximen.

Lema 1.2.28. Sean $A \subset \mathcal{C}_E[0, \infty) \times \mathcal{C}_E[0, \infty)$ y $A_n \subset \mathcal{B}(E) \times \mathcal{B}(E)$, para $n \geq 1$. Supóngase que para cada $(f, g) \in A$ existen $(f_n, g_n) \in A_n$ tales que

$$\lim_{n \rightarrow \infty} \|f_n - f\| = 0, \quad \lim_{n \rightarrow \infty} \|g_n - g\| = 0. \quad (1.44)$$

Si para cada n , X_n es una solución al problema de martingala para A_n con trayectorias en $\mathcal{D}_E[0, \infty)$ y $X_n \Rightarrow X$, entonces X es una solución al problema de martingala para A .

1.2.4. Ejemplo: un proceso de muerte puro.

En esta sección se presenta un proceso de Markov que resultará de utilidad más adelante.

Considérese un sistema cuyo estado en cada tiempo $t \geq 0$ representa el número de individuos en una población, en ese momento. Supóngase que cuando hay n individuos en la población,

- (a) Nuevos individuos nacen con una tasa exponencial β_n .
- (b) Los individuos mueren con una tasa exponencial λ_n .

Es decir, si en un momento dado hay n individuos en la población, el tiempo que debe transcurrir hasta que ocurra (exactamente) un nacimiento es una variable aleatoria exponencial con media $1/\beta_n$ e independiente del tiempo que debe transcurrir hasta que ocurra (exactamente) una muerte, el cual es a su vez una variable aleatoria exponencial con media $1/\lambda_n$. Un sistema de este tipo se llama *proceso de nacimiento y muerte*. Los parámetros $\{\beta_n\}_{n \geq 0}$ y $\{\lambda_n\}_{n \geq 0}$ se llaman tasas de nacimiento y de muerte, respectivamente. Por lo tanto, un proceso de nacimiento y muerte es un proceso de Markov a tiempo continuo con valores en $E = \{0, 1, 2, \dots\}$, para el cual las transiciones desde el estado n pueden ir, de forma directa, únicamente a los estados $n - 1$ si ocurre una muerte antes que un nacimiento y $n + 1$ en caso contrario.

Cuando $\beta_n = 0$ para toda $n \geq 0$, se tiene un proceso de muerte puro. Tavaré [Tav84], en el contexto de modelos de genética poblacional, estudia un proceso de muerte puro $\{D_t^N\}$ con población inicial $D_0^N = N$ y tasas de muerte $\lambda_n = \frac{1}{2}n(n - 1 + \theta)$ para $n = 0, 1, \dots, N$, $\theta > 0$. El generador de este proceso es un operador lineal $Q = (q_{i,j})$ definido por

$$\begin{aligned} q_{n,n} &= -\lambda_n, & n = 0, 1, \dots, N, \\ q_{n,n-1} &= \lambda_n, & n = 1, \dots, N, \\ q_{n,k} &= 0 & k \neq n, n - 1. \end{aligned} \quad (1.45)$$

Después, haciendo N tender a infinito, muestra la existencia de un proceso de muerte puro $\{D_t\}$ con población inicial (límite) $D_0 = \infty$ y generador

$$\begin{aligned} q_{n,n} &= -\lambda_n, & n \geq 0, \\ q_{n,n-1} &= \lambda_n, & n \geq 1, \\ q_{n,k} &= 0 & k \neq n, n-1. \end{aligned} \tag{1.46}$$

La función de transición $d_{rn}^\theta(t) := P(t, n, r)$ para este proceso, está dada por

$$d_{nr}^\theta(t) = \begin{cases} 1 - \sum_{m=1}^n (-1)^{m-1} \binom{n}{m} \frac{(\theta)_{m-1\uparrow}}{(\theta+n)_{m\uparrow}} \gamma_{m,t,\theta} & \text{si } r = 0, \\ \sum_{m=n}^{\infty} (-1)^{m-n} \binom{m}{r} \binom{n}{m} \frac{(\theta+r)_{m-1\uparrow}}{(\theta+n)_{m\uparrow}} \gamma_{m,t,\theta} & \text{si } 1 \leq r \leq n, \end{cases}$$

donde

$$\gamma_{m,t,\theta} = (2m-1+\theta)e^{-\lambda_m t}. \tag{1.47}$$

Aquí y en lo que resta del trabajo, $(a)_{k\uparrow} := a(a+1)\cdots(a+k-1)$ para $k = 1, 2, \dots$, con $(a)_{0\uparrow} := 1$ denota el factorial ascendente generalizado, definido para cualquier $a \in \mathbb{R}$. También se utilizará más adelante la notación $(a)_{k\downarrow} := a(a-1)\cdots(a-k+1)$, con $(a)_{0\downarrow} := 1$ para el factorial generalizado.

Además, haciendo n tender a infinito, Tavaré [Tav84] calcula las probabilidades $d_n^\theta(t) := \mathbb{P}[D_t = n]$ dadas por

$$d_n^\theta(t) = \begin{cases} 1 - \sum_{m=1}^{\infty} (-1)^{m-1} (\theta)_{m-1\uparrow} m!^{-1} \gamma_{m,t,\theta} & \text{si } n = 0, \\ \sum_{m=n}^{\infty} (-1)^{m-n} \binom{m}{n} (\theta+n)_{m-1\uparrow} m!^{-1} \gamma_{m,t,\theta} & \text{si } n > 0. \end{cases}$$

En el capítulo 3 se verá la relación entre estas probabilidades y el proceso de difusión conocido como Fleming-Viot de mutación neutral.

1.3. Modelo bayesiano no paramétrico, intercambiabilidad y medidas aleatorias.

Supóngase que se desea obtener información acerca de una población o fenómeno, a partir de un conjunto de observaciones $\{X_k\}_{k=1}^n$. Para ello, se puede utilizar la estadística, siendo necesario asumir un modelo y algunos supuestos que permitan utilizar los datos disponibles para inferir la información deseada. En el ámbito de la estadística clásica paramétrica, una práctica generalizada consiste en representar las observaciones como variables aleatorias independientes y con una distribución común $\mu \in \mathcal{P}(E)$.

Conocer de forma completa dicha distribución implicaría, en términos estadísticos, conocer todo lo que se desea acerca del fenómeno de estudio, por lo que los datos no aportarían ninguna información adicional relevante. Se asume por lo tanto que μ no es completamente conocida, pero pertenece a una familia de distribuciones específica, indexada por un parámetro de dimensión finita θ . Dicho parámetro tiene un valor fijo, pero desconocido, dentro de un espacio de parámetros Θ y representa, por lo tanto, todo lo que no se conoce acerca del modelo. El problema se convierte entonces en un problema de estimación del parámetro o de otras propiedades de interés asociadas con él.

La estadística bayesiana introduce una variante, relacionada con el concepto de aleatoriedad. En este enfoque, cualquier cantidad desconocida es susceptible de afirmaciones acerca de la incertidumbre que sobre ella se tiene, a través de una medida de probabilidad. Con esto en mente, se puede definir una medida de probabilidad sobre el espacio de parámetros Θ que refleje la información subjetiva acerca del fenómeno, representada por θ , previa al inicio del análisis estadístico. Por lo tanto, se puede asignar al parámetro una *distribución inicial* de probabilidad Π , obteniéndose un modelo bayesiano paramétrico de la siguiente forma

$$\begin{aligned} X_k | \mu_\theta &\stackrel{iid}{\sim} \mu_\theta; \quad k = 1, \dots, n \\ \theta &\sim \Pi \end{aligned} \tag{1.48}$$

El teorema de Bayes provee un mecanismo para incorporar la información contenida en los datos. De este modo se obtiene la *distribución posterior* del parámetro, a partir de la cual se estimará las cantidades de interés.

$$\Pi(B | \mathbb{X}^n \in \Gamma) := \mathbb{P}[\theta \in B | \mathbb{X}^n \in \Gamma] = \frac{\prod_{k=1}^n \mu_\theta(\Gamma_k) \Pi(B)}{\int_{\Theta} \prod_{k=1}^n \mu_\theta(\Gamma_k) d\Pi(\theta)}, \tag{1.49}$$

donde $\mathbb{X}^n := (X_1, \dots, X_n)$ es un vector aleatorio con distribución conjunta, condicional al parámetro, $\mathbb{P}[\mathbb{X}^n \in \Gamma | \theta \in B] = \prod_{k=1}^n \mu_\theta(\Gamma_k) \Pi(B)$, definida para $B \in \mathcal{B}(\Theta)$ y $\Gamma := \Gamma_1 \times \dots \times \Gamma_n \in \mathcal{B}(E)$. En la sección 1.3.1 se explica brevemente la forma en que la distribución posterior para θ se utiliza para hacer inferencia, es decir, para obtener información acerca del fenómeno de estudio a través del modelo planteado, incorporando los datos disponibles. El modelo (1.48) no es el único modelo bayesiano paramétrico posible, pero basta para introducir la idea que da origen al modelo bayesiano no paramétrico que sirve de base para el desarrollo del capítulo 4.

En ambos enfoques paramétricos (clásico y bayesiano), el supuesto distribucional resulta muy restrictivo, pues impone ciertas propiedades a las

variables estudiadas, por ejemplo unimodalidad o algún tipo de relación entre los momentos de su distribución. Como consecuencia, especialmente en casos en los que no se cuenta con suficiente información acerca del fenómeno de estudio, es conveniente relajar los supuestos. Una manera natural de hacerlo es aumentar la dimensión del parámetro utilizado para caracterizar a la distribución, aumentando así el «tamaño» del espacio que representa la incertidumbre. En el extremo, se puede pensar en parámetros de dimensión infinita, surgiendo la llamada estadística no paramétrica.

En el caso bayesiano, esto implica una dificultad que detuvo por mucho tiempo el desarrollo de la teoría no paramétrica. Aumentar la dimensión del espacio de parámetros hasta hacerla infinita es equivalente a aumentar el espacio de distribuciones en el que se trabaja. Por lo tanto, en lugar de definir una distribución inicial sobre un espacio Θ de dimensión finita, se debe asignar una distribución inicial sobre el espacio de todas las distribuciones posibles para las variables aleatorias de interés. La existencia de tal distribución sobre el espacio de distribuciones se puede garantizar por medio del teorema de representación de de Finetti, a cambio de asumir cierta relación entre los datos, conocida como intercambiabilidad.

Definición 1.3.1 (Intercambiabilidad). Un conjunto finito X_1, \dots, X_n de variables aleatorias con valores en E (es decir, $X_j : \Omega \rightarrow E$ medible para cada $j \in \{1, \dots, n\}$) se dice intercambiable si cada permutación de (X_1, \dots, X_n) tiene la misma distribución conjunta que cualquier otra permutación. Una sucesión $\{X_k\}_{k \geq 1}$ de variables aleatorias con valores en E es intercambiable si cada subconjunto finito lo es.

En esta definición y en el resto de este capítulo, $(\Omega, \mathcal{F}, \mathbb{P})$ es un espacio de probabilidad y (E, r) un espacio métrico completo y separable, con σ -álgebra de Borel, $\mathcal{B}(E)$. Al igual que en las secciones anteriores, $\mathcal{P}(E)$ denota al conjunto de medidas de probabilidad sobre E que dotado con la métrica de convergencia débil es a su vez un espacio métrico completo y separable, con σ -álgebra de Borel $\mathcal{B}(\mathcal{P}(E))$. Es posible entonces definir una medida de probabilidad Π sobre $(\mathcal{P}(E), \mathcal{B}(\mathcal{P}(E)))$. Sea E^∞ el espacio producto (infinito) de E , con σ -álgebra de Borel $\mathcal{B}(E^\infty) = (\mathcal{B}(E))^\infty$. Si $\mu \in \mathcal{P}(E)$, entonces $\mu^\infty \in \mathcal{P}(E^\infty)$ denota la correspondiente medida producto sobre E^∞ . Para cualquier $x \in E$, $\delta_x \in \mathcal{P}(E)$ denota la medida de probabilidad con toda su masa acumulada en x .

Teorema 1.3.2 (Representación de de Finetti). *Una sucesión de variables aleatorias $\{X_k\}_{k \geq 1}$ con valores en E es intercambiable si y solo si existe una medida de probabilidad Π sobre $(\mathcal{P}(E), \mathcal{B}(\mathcal{P}(E)))$ tal que para*

cualquier boreliano $B = \{B_k\}_{k \geq 1} \in \mathcal{B}(E^\infty)$ (i.e. $B_k \in \mathcal{B}(E)$ para cada $k \geq 1$),

$$\mathbb{P}[X^\infty \in B] := \mathbb{P}[X_k \in B_k; k \geq 1] = \int_{\mathcal{P}(E)} \mu^\infty(B) \Pi(d\mu) \quad (1.50)$$

Además Π , llamada en ocasiones la medida de de Finetti de la sucesión, es única e igual al límite de las distribuciones empíricas:

$$\Pi = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \delta_{X_k} \quad (1.51)$$

Una interpretación útil del teorema de representación de de Finetti consiste en considerar que una sucesión infinita de variables aleatorias $\{X_n\}_{n \geq 1}$ definida sobre un espacio métrico (E, r) completo y separable, es intercambiable si y solo si las variables aleatorias son condicionalmente independientes e idénticamente distribuidas (iid) como μ , dada la medida $\mu \in \mathcal{P}(E)$. Esta interpretación, bajo el supuesto de intercambiabilidad de las observaciones, da lugar al siguiente *modelo bayesiano no paramétrico* para sucesiones intercambiables

$$\begin{aligned} X_k | \mu &\stackrel{iid}{\sim} \mu; \quad k = 1, \dots, n \\ \mu &\sim \Pi. \end{aligned} \quad (1.52)$$

Nótese que, si se restringe el espacio $\mathcal{P}(E)$ a una familia paramétrica $\{\mu_\theta\} \subset \mathcal{P}(E)$, con $\theta \in \Theta$, entonces Π se puede reexpresar como una medida sobre el espacio de parámetros. Así se recupera el modelo bayesiano paramétrico (1.48).

El supuesto de intercambiabilidad es una afirmación acerca de la relación entre las variables u observaciones modeladas. Consiste en suponer que el orden en que la información se recibe o incorpora al modelo es irrelevante y, como se explicó arriba, equivale en la práctica a hacer un supuesto de independencia condicional. Esto puede no ser realista en todos los casos, pero resulta aceptable para muchas aplicaciones y es una forma de justificar modelos bayesianos como (1.48) y (1.52), que aportan una gran flexibilidad al análisis estadístico.

En la práctica, la utilización de modelos del tipo (1.52) para realizar inferencia estadística, depende de la posibilidad de definir una distribución inicial Π sobre $\mathcal{P}(E)$ que resulte manejable y que pueda ser actualizada a partir de la información contenida en los datos, a través del teorema de Bayes. Es decir, que permita calcular la distribución posterior

$$\Pi(B | \mathbb{X}^n \in \Gamma) := \mathbb{P}[\mu \in B | \mathbb{X}^n \in \Gamma] = \frac{\prod_{k=1}^n \mu(\Gamma_k) \Pi(B)}{\int_{\mathcal{P}(E)} \prod_{k=1}^n \mu(\Gamma_k) d\Pi(\mu)}, \quad (1.53)$$

definida para toda $B \in \mathcal{B}(\mathcal{P}(E))$.

Ferguson [Fer73], al introducir el proceso Dirichlet como herramienta para la inferencia bayesiana no paramétrica, mencionó dos características deseables para una distribución inicial no paramétrica:

- (a) El soporte debe ser grande, con respecto al espacio de medidas $\mathcal{P}(E)$ sobre el cual se define.
- (b) La distribución posterior dada la muestra debe ser manejable analíticamente.

El fuerte avance en recursos y métodos computacionales en años posteriores al artículo de Ferguson ha restado importancia al segundo punto, facilitando la utilización de distribuciones iniciales más generales y que permiten satisfacer de forma más amplia el primer punto. Se ha desarrollado resultados teóricos en torno a procesos estocásticos con valores en $\mathcal{P}(E)$, que pueden ser utilizados como distribuciones iniciales no paramétricas, así como métodos de simulación que permiten el cálculo de cantidades de interés, sin la necesidad de encontrar una expresión analítica explícita para la distribución posterior (1.53). Una importante familia de distribuciones iniciales no paramétricas, de la cual el proceso Dirichlet es un caso particular, se deriva del llamado modelo de urnas, que se presentará en el capítulo 2.

1.3.1. Inferencia bayesiana y método de Monte Carlo.

Considérese el modelo bayesiano paramétrico (1.48), con $E = \mathbb{R}$ y espacio de parámetros $\Theta = \mathbb{R}^d$. El planteamiento requiere la especificación de una familia paramétrica $\{\mu_\theta : \theta \in \Theta\}$ que establezca la forma en que la incertidumbre acerca del fenómeno de estudio se relaciona con el parámetro. Supóngase que $\mu_\theta(B) := \mathbb{P}[X \in B|\theta]$, con $B \in \mathcal{B}(E)$, tiene densidad asociada $p(X|\theta)$, la cual vista como función de θ se suele llamar *verosimilitud*. Se requiere igualmente la definición de una distribución inicial o «a priori» para θ , $\Pi \in \mathcal{P}(\Theta)$, totalmente conocida, es decir $\Pi(B) = \mathbb{P}[\theta \in B]$ se conoce para cualquier $B \in \mathcal{B}(\Theta)$ y con densidad asociada $p(\theta)$. (También se podría suponer que Π es dependiente a su vez de parámetros desconocidos para los cuales se debe definir una distribución inicial, obteniéndose un *modelo jerárquico*).

La información de los datos $\mathbb{X}^n := (X_1, \dots, X_n)$ se incorpora al modelo a través del teorema de Bayes para obtener la distribución posterior $\Pi(\cdot|\mathbb{X}^n) \in \mathcal{P}(\Theta)$ definida por $\Pi(B|\mathbb{X}^n) := \mathbb{P}[\theta \in B|\mathbb{X}^n]$, para cualquier

$B \in \mathcal{B}(\Theta)$. Su densidad asociada se calcula como

$$p(\theta|\mathbb{X}^n) = \frac{p(\theta)p(\mathbb{X}^n|\theta)}{\int_{\Theta} p(\theta)p(\mathbb{X}^n|\theta)d\theta}, \quad (1.54)$$

donde $p(\mathbb{X}^n|\theta) = \prod_{j=1}^n p(X_j|\theta)$. En este punto, se plantea un primer problema, pues la integral en la expresión de arriba puede resultar muy difícil de calcular, especialmente en el caso en que θ es un parámetro multidimensional. Por lo tanto, en muchos casos, la distribución posterior o final en un modelo bayesiano paramétrico se conoce salvo una constante de proporcionalidad.

En principio, el conocimiento de la distribución posterior para θ permite hacer inferencia sobre cualquier función $\phi = \phi(\theta)$ medible, por ejemplo, un estimador puntual se puede calcular como

$$\mathbb{E}[\phi(\theta)|\mathbb{X}^n] = \int_{\Theta} \phi(\theta)p(\theta|\mathbb{X}^n)d\theta. \quad (1.55)$$

O se puede hacer inferencia sobre observaciones futuras, mediante la llamada *distribución predictiva*

$$p(X|\mathbb{X}^n) = \int_{\Theta} p(X|\theta)p(\theta|\mathbb{X}^n)d\theta. \quad (1.56)$$

El cálculo de estas dos integrales puede ser muy complejo, aún en el caso en que se conoce la constante de proporcionalidad para $p(\theta|\mathbb{X}^n)$.

Cuando las distribuciones u otras cantidades no pueden ser calculadas explícitamente, es necesario recurrir a aproximaciones analíticas o numéricas. Una de las alternativas más utilizadas es el llamado *método de Monte Carlo*, que permite aproximar, con base en las leyes de grandes números, el valor esperado $\mathbb{E}[\phi(X)] < \infty$ para cualquier función medible ϕ sobre E y $X \sim \mu \in \mathcal{P}(E)$, por

$$\mathbb{E}[\phi(X)] = \int_E \phi(x)\mu(dx) \approx \frac{1}{k} \sum_{j=1}^k \phi(x_k), \quad (1.57)$$

donde (x_1, \dots, x_k) son realizaciones independientes de una variable aleatoria con distribución μ . Por lo tanto, para obtener una aproximación a las integrales que aparecen en las expresiones (1.54), (1.55) y (1.56) basta poder simular observaciones independientes de $\theta \sim \Pi(\theta|\mathbb{X}^n)$. Sin embargo, esto aún puede ser complejo, sobre todo en el caso en que la dimensión de θ es mayor a uno. Entonces, es necesario contar con un algoritmo que permita la simulación de variables aleatorias multivariadas. Un algoritmo de simulación muy utilizado en estos casos es el llamado muestreo de Gibbs.

1.3.2. Método de muestreo de Gibbs.

Sea $g : A \rightarrow A$ una función continua para la cual se desea encontrar un punto fijo, es decir, un punto $x \in A$ tal que $g(x) = x$. A partir de un punto inicial $x_0 \in A$ se puede definir la sucesión $\{x_k\}_{k \geq 1}$ haciendo $x_k = g(x_{k-1})$. Si la sucesión converge, el límite x es el punto fijo deseado. Esta es la base para el muestreo de Gibbs, también llamado muestreo de sustitución sucesiva (*Successive Substitution Sampling [Sch95]*) o muestreo condicional alternante (*Alternating Conditional Sampling [GCSR04]*). La idea explota algunas propiedades de las cadenas de Markov para construir un proceso estacionario con distribución marginal conocida.

Actualmente, el uso más común del muestreo de Gibbs es quizá la simulación de variables aleatorias multivariadas y es particularmente útil en el contexto de inferencia bayesiana.

Sea $\mathbb{Y}^n := (Y_1, \dots, Y_n)$ un vector de variables aleatorias y para cada i , sea $f_{Y_i|\mathbb{Y}^{-i}}$ (donde $\mathbb{Y}^{-i} := \{Y_j : j \neq i\}$) la densidad correspondiente a la distribución condicional de Y_i dadas las demás, con respecto a una medida μ_i . Se denota por $f_{\mathbb{Y}^n}$ la densidad correspondiente a la distribución conjunta de \mathbb{Y}^n con respecto a la medida producto $\mu = \mu_1 \times \dots \times \mu_n$, de la cual se desea simular un valor.

Supóngase que se observa $v := (v_1, \dots, v_n)$, realización de una densidad $f_{\mathbb{Y}^n}$ con respecto a la medida μ . Es posible definir una nueva densidad $f_{\mathbb{X}^n}$ para $\mathbb{X}^n := (X_1, \dots, X_n)$ de la siguiente forma. La densidad de X_1 con respecto a μ_1 es $f_{Y_1|\mathbb{Y}^{-1}}(\cdot | v_2, \dots, v_n)$; la densidad condicional, con respecto a μ_2 , de X_2 dado $X_1 = x_1$ es $f_{Y_2|\mathbb{Y}^{-2}}(\cdot | x_1, v_3, \dots, v_n)$ y así sucesivamente hasta tener que $f_{Y_n|\mathbb{Y}^{-n}}(\cdot | x_1, \dots, x_{n-1})$ es la densidad condicional de X_n dadas X_1, \dots, X_{n-1} . Entonces

$$f_{\mathbb{X}^n}(x) = \int \left[\prod_{i=1}^n f_{Y_i|\mathbb{Y}^{-i}}(x_i | z_i) \right] f_{\mathbb{Y}^n}(v) d\mu(v)$$

donde $z_1 = (v_2, \dots, v_n)$, $z_i = (x_1, \dots, x_{i-1}, v_{i+1}, \dots, v_n)$ y $z_n = (x_1, \dots, x_{n-1})$.

Si se define el operador

$$T(f)(x) := \int \left[\prod_{i=1}^n f_{Y_i|Y_{-i}}(x_i | z_i) \right] f(v) d\mu(v), \quad (1.58)$$

entonces $T(f_{\mathbb{Y}^n}) = f_{\mathbb{Y}^n}$, de modo que la densidad conjunta para \mathbb{Y}^n es un punto fijo de T . Por lo tanto, si se parte de una densidad inicial f_0 , se puede utilizar el método de sustitución sucesiva, haciendo $f_k = T(f_{k-1})$. Si la sucesión $\{f_k\}_{k \geq 1}$ es convergente, su límite es $f_{\mathbb{Y}^n}$.

En la práctica, no es necesario resolver en cada paso una integral para obtener cada una de las densidades conjuntas, basta obtener muestras de las distribuciones condicionales involucradas en la definición del operador. A este procedimiento se le conoce comúnmente como muestreo de Gibbs:

(a) Inicialización.

Se fija $\mathbb{X}_0^n = (x_1(0), \dots, x_n(0))$, que se puede considerar como una observación de una distribución inicial f_0 .

(b) Muestreo condicional alternante.

Dado $\mathbb{X}_{k-1}^n = (x_1(k-1), \dots, x_n(k-1))$, se simula $\mathbb{X}_k^n = (x_1(k), \dots, x_n(k))$ a partir de las distribuciones condicionales, es decir, se genera $x_i(k)$ de la densidad $f_{Y_i|Y^{-i}}(x_i(k) | z_i(k))$ para cada $i = 1, \dots, n$.

La sucesión $\{\mathbb{X}_k^n\}_{k \geq 1}$ es una cadena de Markov, ya que la densidad conjunta de \mathbb{X}_k^n , para cada $k \geq 1$, depende únicamente de \mathbb{X}_{k-1}^n . Por este motivo, se dice que el muestreo de Gibbs es un caso particular de un conjunto más general de métodos conocidos como métodos de Monte Carlo vía Cadenas de Markov (MCMC). Además, la convergencia de $\{f_k\}_{k \geq 1}$ a f_{Y^n} es suficiente para garantizar la convergencia de $\{\mathbb{X}_k^n\}_{k \geq 1}$ a una realización de la variable aleatoria \mathbb{X}^n con la distribución conjunta deseada, incluso si el orden de actualización de las variables se modifica en cada paso (por ejemplo, si el orden de actualización se elige aleatoriamente).

Las cadenas de Markov construidas por el método de muestreo de Gibbs son reversibles en el tiempo.

1.4. Construcción de procesos de Markov estacionarios a través de procesos latentes.

1.4.1. Cadenas de Markov vía variables latentes.

En el año 2000, Pitt *et al* [PCW02] introdujeron, en el contexto de modelos autorregresivos del tipo AR(1), un método para la construcción de cadenas de Markov estrictamente estacionarias, con distribuciones marginales conocidas. Esta construcción se logra de forma conceptualmente sencilla explotando propiedades del muestreo de Gibbs.

Supóngase que se desea construir una cadena de Markov $\{X_k\}_{k \geq 1}$, cuya distribución marginal pertenece a una familia paramétrica, con densidad f_X . Es posible definir el proceso a través de su función de transición, de modo que la distribución marginal deseada resulte invariante en el tiempo. Para

ello, se incorpora la dependencia entre X_k y X_{k-1} a través de un proceso latente Y (que juega el papel de un parámetro para la función de transición de $\{X_k\}$), con densidad condicional $f_{Y|X}(y|x)$. La función de transición a un paso para el proceso se define como

$$P(x_{k-1}, x_k) = \int f_{X|Y}(x_k|y) f_{Y|X}(y|x_{k-1}) \mu_1(dy), \quad (1.59)$$

donde

$$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x) f_X(x)}{\int f_{Y|X}(y|x) f_X(x) \mu_2(dx)}. \quad (1.60)$$

Aquí, μ_1 y μ_2 denotan medidas de referencia en los espacios en los que toman valores X y Y respectivamente. En la práctica pueden coincidir con las medidas de Lebesgue o de conteo.

De este modo se satisface

$$\int P(x_{k-1}, x_k) f_X(x_{k-1}) \mu_2(dx_{k-1}) = f_X(x_k), \quad (1.61)$$

por lo que f_X resulta invariante para la transición definida, es decir, es la distribución estacionaria del proceso $\{X_k\}$.

Es importante señalar que, al introducir variables latentes en la construcción del proceso $\{X_k\}_{k \geq 1}$, implícitamente se está construyendo también el proceso $\{Y_k\}_{k \geq 1}$.

Mena y Walker [MW05] generalizaron este método al adoptar un enfoque no paramétrico en la definición de la función de transición. La idea se basa en reinterpretar a esta última como un valor esperado con respecto a la distribución condicional $f_{Y|X}$, que toma el papel de distribución posterior (en el sentido bayesiano) para la inicial f_Y con verosimilitud $f_{X|Y}$, es decir

$$P(x_{k-1}, x_k) = \mathbb{E}[f_{X|Y}(x|y) | x_{k-1}]. \quad (1.62)$$

Si en lugar de pensar en Y como una variable latente se considera la introducción de una función de densidad aleatoria f con densidad inicial $\Pi(df)$, se puede definir la función de transición a un paso como

$$P(x_{k-1}, x_k) = \mathbb{E}[f(x) | x_{k-1}] = \int f(x) \Pi(df | x_{k-1}), \quad (1.63)$$

donde $\Pi(df | x_{k-1})$ es la densidad posterior para f dada la observación x_{k-1} .

El atractivo principal de esta propuesta reside en la posibilidad de utilizar métodos desarrollados en el área de la estadística bayesiana no paramétrica, junto con las propiedades del muestreo de Gibbs, para la construcción de

cadena de Markov reversibles y estacionarias con distribuciones marginales generales. Es importante destacar que la teoría que fundamenta el muestreo de Gibbs y por tanto, las propiedades de los procesos así construidos, funciona para procesos que toman valores en espacios muy generales. En particular, se puede pensar en el proceso latente $\{f_k\}_{k \geq 1}$ con valores en el espacio de densidades sobre el espacio en el que toma valores X . Esta idea será retomada en el capítulo 4 como base para la construcción del proceso Fleming-Viot a través de un proceso latente.

1.4.2. Generalización del método: procesos markovianos en tiempo continuo.

Otra posible generalización del método presentado por Pitt *et al* [PCW02] se refiere a la construcción de procesos de Markov a tiempo continuo. Mena y Walker ([MW07]) proponen una forma de hacerlo, a través de un modelo que incluya algún parámetro dependiente del tiempo. La manera más sencilla de explicar el método es a través de un ejemplo.

Se parte del ejemplo básico de Pitt *et al* [PCW02] para la construcción de una cadena de Markov $\{X_k\}_{k \geq 1}$ con distribución estacionaria $\text{Ga}(\cdot | a, b)$. Las distribuciones condicionales que relacionan al proceso de interés con el proceso latente, son elegidas teniendo en cuenta la marginal deseada y para facilitar los cálculos, se elige una distribución conjugada (en el sentido bayesiano).

Considérese las distribuciones condicionales siguientes

$$X|Y \sim \text{Ga}(X|a + Y, b + \phi), \quad (1.64)$$

$$Y|X \sim \text{Po}(Y|\phi X), \quad (1.65)$$

donde $\phi > 0$. Entonces, la densidad conjunta para (X, Y) está dada por

$$f_{X,Y}(x, y) = \text{Po}(y|\phi x) \text{Ga}(x|a, b) \quad (1.66)$$

y la distribución marginal para X ,

$$f_X(x) = \text{Ga}(x|a, b) \quad (1.67)$$

coincide con la distribución estacionaria deseada para el proceso $\{X_k\}_{k \geq 0}$.

Se usa el método del muestreo de Gibbs, iniciando con $X_0 \sim \text{Ga}(X|a, b)$ para construir una sucesión,

$$X_0 \longrightarrow Y_1 \longrightarrow X_1 \longrightarrow Y_2 \cdots$$

De este modo, como se mencionó al principio de esta sección, se construye simultáneamente dos cadenas de Markov, reversibles, gracias a las propiedades del muestreo de Gibbs. Además, por la igualdad (1.61), $\{X_k\}_{k \geq 1}$ es una cadena de Markov estacionaria con densidad f_X .

El siguiente paso es llevar la cadena a tiempo continuo. Para ello, es conveniente observar que el parámetro ϕ controla la correlación en el proceso $\{X_k\}_{k \geq 1}$. Si ϕ es pequeño (cercano a cero), entonces es más probable que Y_1 tome valores pequeños, de modo que la distribución de X_1 será «cercana» a la $\text{Ga}(\cdot | a, b)$. Si ϕ es grande, es más probable que Y_1 tome valores grandes y en consecuencia que X_1 resulte cercana a X_0 . Por lo tanto tiene sentido que, para llevar el proceso a tiempo continuo, se sustituya ϕ por ϕ_t , una función determinista estrictamente creciente dependiente del tiempo. Puesto que el proceso que se desea construir toma valores en \mathbb{R} , un espacio completo y separable, basta encontrar una función $\phi_t = \phi(t)$ que satisfaga las ecuaciones de Chapman-Kolmogorov, para garantizar la existencia del proceso de Markov a tiempo continuo $\{X_t\}_{t \geq 0}$.

En este caso, Mena y Walker [MW07] demuestran que todas las soluciones son de la forma

$$\phi_t = \frac{b}{e^{ct} - 1} \quad (1.68)$$

para alguna $c > 0$. Se obtiene un proceso de Markov $\{X_t\}_{t \geq 0}$ homogéneo en el tiempo, con densidad de transición

$$P(t, x_0, x_t) = \sum_{y=0}^{\infty} \text{Ga}(x_t | y + a, \phi_t + b) \text{Po}(y | x_0 \phi_t). \quad (1.69)$$

Demuestran además, que $\{X_t\}_{t \geq 0}$ equivale (en distribución) a un proceso de difusión conocido como proceso Gamma-Poisson (o modelo Cox-Ingersoll-Ross reparametrizado) que se puede definir como la solución a una ecuación diferencial estocástica dada por

$$dX_t = c(a/b - X_t)dt + \sqrt{\frac{2c}{b}} X_t dW_t,$$

donde $\{W_t\}$ denota un movimiento browniano. La utilización de este método para construir procesos de difusión es de particular interés para el presente trabajo.

Debido a las propiedades del muestreo de Gibbs, todos los procesos obtenidos de esta forma son reversibles, sin embargo, es posible utilizarlo para construir procesos de Markov no estacionarios [MW07].

Se puede generalizar el método aún más para construir modelos con funciones de transición dependientes de cualquier conjunto, por ejemplo de covariables, en vez de fijar la dependencia únicamente en el tiempo.

Otra alternativa consiste en permitir que la dependencia del parámetro ϕ , que regula la correlación del proceso con respecto al tiempo, sea estocástica. Es decir, en vez de definir una función determinista ϕ_t , puede asignarse una medida de probabilidad $\mathbb{P}[\phi_t = \phi]$ sobre el espacio de posibles valores del parámetro.

Finalmente, es posible combinar la construcción a tiempo continuo con la construcción de Mena y Walker [MW05], que utiliza el enfoque no paramétrico. En este caso, en vez de un proceso de variables latentes $\{Y_k\}_{k \geq 1}$ se tiene un proceso de densidades o medidas.

Estas últimas dos ideas serán esenciales para la construcción de un proceso de difusión, conocido como Fleming-Viot de mutación neutral, en el capítulo 4.

Capítulo 2

Modelos de Urnas y el proceso Dirichlet.

Los modelos de urnas se basan en la siguiente idea. Imagínese una urna que contiene, en un principio, cierto número fijo de bolas de diferentes colores o tipos, donde todos los posibles colores son conocidos y forman un espacio E , llamado el *espacio de estados o tipos*. Se extrae una bola y dependiendo de su color, se modifica el contenido de la urna para la siguiente extracción, repitiendo este procedimiento indefinidamente.

Aunque conceptualmente simple, esta idea da lugar a una muy amplia familia de modelos probabilísticos. En los casos más sencillos, E es finito pero se puede extender la misma idea para un espacio de estados infinito, incluso no numerable. El número de bolas de cada color contenidas en la urna es, en el caso básico, un entero no negativo, pero el esquema se puede generalizar para números reales no negativos. La probabilidad de extraer de la urna una bola de un tipo dado es, en el caso completamente aleatorio (cada bola tiene la misma probabilidad de ser elegida), igual a la proporción de bolas de dicho tipo dentro de la urna al momento de la extracción. Además, en este caso básico, la modificación de la urna después de la extracción consiste en devolver la bola extraída junto con una nueva del mismo tipo. Sin embargo, en los modelos generales es posible introducir mecanismos de extracción y reemplazo más complejos, como dependencia del pasado, extracción de más de una bola a la vez, aparición de nuevos colores, entre otros. También es común la definición de modelos en que varias urnas interactúan entre sí.

Su gran flexibilidad, ha hecho de los modelos de urnas una herramienta muy importante, en particular para la estadística bayesiana no paramétrica, pues bajo ciertas condiciones permiten la representación y manipulación de

distribuciones para sucesiones de variables aleatorias intercambiables.

A lo largo de este capítulo se presentará una familia de modelos de urnas que dan lugar a sucesiones intercambiables. Se pondrá especial énfasis en el modelo de Blackwell y MacQueen [BM73], que constituye una importante caracterización para el proceso Dirichlet propuesto por Ferguson [Fer73] y utilizado ampliamente desde entonces. Esta caracterización resulta esencial para la construcción del proceso Fleming-Viot en el capítulo 4.

El recuento que se presenta aquí de los modelos de urnas de Pólya y urnas de Pólya generalizadas no es exhaustivo y su propósito consiste únicamente en servir de base para la construcción desarrollada en el último capítulo de este trabajo. Los resultados aquí presentados provienen en su mayoría de [BM73, IZ03, Pem07].

2.1. Urnas de Pólya generalizadas

Las urnas de Pólya deben su nombre al matemático húngaro Pólya György quien las introdujo por primera vez para modelar el comportamiento de enfermedades contagiosas. En el modelo original sólo hay bolas de dos colores, rojo y negro, y la urna contiene inicialmente una bola de cada color. En un primer momento, se extrae aleatoriamente de la urna una bola y se devuelve junto con otra del mismo color, repitiendo el procedimiento indefinidamente. Este esquema se puede representar por un proceso $\{X_n\}_{n \geq 1}$ con espacio de estados $E = \{0, 1\}$, donde $X_n = 1$ si en la n -ésima extracción se obtiene una bola roja y $X_n = 0$ si se obtiene una negra. Se denota por N_{1n} y N_{0n} la cantidad de bolas rojas y negras respectivamente al tiempo n , con N_{10} y N_{00} , las cantidades iniciales, no necesariamente iguales a 1. Entonces, la proporción de bolas rojas dentro de la urna antes de la extracción $n + 1$ es $p_{1n} = N_{1n}/(N_{1n} + N_{0n})$ y $p_{0n} = 1 - p_{1n}$ es la proporción de bolas negras. En este modelo básico, la probabilidad de extracción es «uniforme», en el sentido de que las bolas se extraen con igual probabilidad, de modo que la probabilidad de extraer una bola roja en la extracción $n + 1$ es igual a p_{1n} . Una propiedad sobresaliente de este modelo es que las proporciones p_{1n} convergen casi seguramente a un límite aleatorio $p_1 \sim \text{Be}(p_1|a, b)$, donde $a = N_{10}$ y $b = N_{00}$.

El modelo de la urna de Pólya se puede generalizar para un espacio de estados finito $E = \{0, 1, \dots, k\}$, con $k \geq 2$. Además, es posible modificar el esquema de reemplazo después de cada extracción. Denótese por N_{jn} , con $j \in E$, el número de bolas del tipo j contenidas en la urna al tiempo n . Supóngase que cada vez que se extrae una bola del color i , se devuelve

a la urna junto con A_{ij} bolas de color j , para $j \in E$, donde la colección $\{A_{ij} : i, j \in E\}$ está formada por números reales fijos, tales que $A_{ij} \geq 0$ para $i \neq j$ y $A_{ii} \geq -1$, de modo que el número de bolas introducidas cada vez no es necesariamente entero. Supóngase además que la probabilidad de extraer una bola de color i al tiempo $n+1$ es $p_{in} := N_{jn} / \sum_{j \in E} N_{jn}$.

El vector de probabilidades $p_n = (p_{1n}, \dots, p_{kn})$ pertenece al espacio $\Delta_E := \{(p_{in})_{i \in E} \in [0, 1]^E : \sum_{i \in E} p_{in} = 1\}$, para cada $n \geq 1$. Observando esto, resulta natural generalizar al caso en que E es infinito, sustituyendo Δ_E por $\mathcal{P}(E)$, el espacio de medidas de probabilidad sobre E , y p_n por $\mu_n \in \mathcal{P}(E)$. Partiendo de una medida inicial μ_0 , la probabilidad de extraer una bola del tipo $x \in E$ al tiempo $n+1$ queda determinada por μ_n . En el caso más sencillo, tras cada extracción, la bola es devuelta a la urna, junto con otra del mismo tipo. Entonces, la actualización de la medida está definida en términos del tipo de bola extraída como

$$\mu_{n+1} = \frac{\theta_n \mu_n + \delta_{X_n}}{\theta_n + 1}, \quad (2.1)$$

con $\theta_n = \mu_n(E)$, donde δ_x denota la medida de probabilidad con toda su masa acumulada en x . Si $\{X_n\}_{n \geq 1}$ es el proceso definido por los tipos de bolas extraídas en cada tiempo, la expresión (2.1) para μ_n se puede reexpresar en términos de la medida inicial como

$$\mu_{n+1} = \frac{\mu_0 + \sum_{i=1}^n \delta_{X_i}}{1+n}, \quad (2.2)$$

pues $\theta_0 = \mu_0(E) = 1$. A un modelo de este tipo se le llama comúnmente *urna de Pólya generalizada*.

Para el presente trabajo, resultan de particular interés los modelos de urnas de Pólya generalizadas que dan origen a sucesiones $\{X_n\}_{n \geq 1}$ intercambiables.

Sea $\{X_n\}_{n \geq 1}$ una sucesión de variables aleatorias con valores en un espacio E . Si la sucesión es generada a partir de una urna de Pólya generalizada, es claro que los valores observados se pueden repetir (en virtud de la componente discreta $\sum \delta_{X_i}$ en la expresión (2.2)). Denótese por X_1^*, X_2^*, \dots a los valores únicos (no repetidos) de X_1, X_2, \dots , en orden de aparición.

Se denota por π_n a la partición de $[n] := \{1, 2, \dots, n\}$ inducida por las primeras n extracciones, $\{X_j\}_{j=1}^n$. Es decir, $\pi_n = \{C_{jn} : j = 1, \dots, k_n^*\}$, donde k_n^* es el número de valores únicos en X_1, \dots, X_n , $X_i = X_j^*$ para cada $i \in C_{jn}$, y

$$\begin{aligned} C_{in} \cap C_{jn} &= \emptyset, \quad i \neq j \\ \cup_{j=1}^{k_n^*} C_{jn} &= [n]. \end{aligned}$$

La cardinalidad de C_{jn} , denotada por k_{jn}^* es el número de veces que X_j^* se repite en X_1, \dots, X_n .

Sea $\mu \in \mathcal{P}(E)$ una medida de probabilidad no nula sobre E . Considérese las sucesiones $\{X_n\}_{n \geq 1}$ definidas por

$$\mathbb{P}[X_1 \in \cdot] = \mu(\cdot) \quad (2.3)$$

$$\mathbb{P}[X_{n+1} \in \cdot | X_1, \dots, X_n] = \frac{w_{0n}}{\sum_{j=0}^{k_n^*} w_{jn}} \mu(\cdot) + \sum_{j=1}^{k_n^*} \frac{w_{jn}}{\sum_{j=0}^{k_n^*} w_{jn}} \delta_{X_j^*}(\cdot), \quad (2.4)$$

donde $w_{jn} = w_{jn}(n, k_{1n}^*, \dots, k_{k_n^*n}^*)$ para $j = 0, \dots, k_n^*$ son funciones simétricas que dependen únicamente de $\{n, k_{1n}^*, \dots, k_{k_n^*n}^*\}$.

Ishwaran y Zarepour [IZ03] establecen una condición suficiente (aunque no necesaria) para que una sucesión definida por las distribuciones (2.3) y (2.4) sea intercambiable.

Teorema 2.1.1. *Sea $\mu \in \mathcal{P}(E)$ no nula y no atómica, es decir, $\mu(E) > 0$ y $\mu(\{x\}) = 0$ para cualquier $x \in E$. Supóngase que para cada $n \geq 1$, $w_{jn} = \psi(k_{jn}^*)$ y $w_{0n} = \psi_0(k_n^*)$ donde ψ y ψ_0 son funciones reales no negativas fijas, y para cada partición π_i de $[i]$*

$$\sum_{j=0}^{k_n^*} w_{jn} = \xi(i) > 0 \quad (2.5)$$

con ξ una función real positiva fija. Entonces la sucesión $\{X_n\}_{n \geq 1}$ definida a partir de las distribuciones (2.3) y (2.4) es intercambiable.

Corolario 2.1.2. *Bajo las condiciones del Teorema 2.1.1,*

- (a) *Sea μ_{n+1} la distribución condicional para X_{n+1} definida por (2.4). Entonces $\mu_{n+1} \rightarrow \mu^*$ c.s. cuando $n \rightarrow \infty$, donde μ^* es la medida de probabilidad aleatoria definida por*

$$\mu^* = \sum_j p_j \delta_{X_j^*} + (1 - \sum_j p_j) \mu, \quad (2.6)$$

con $p_j = \lim_{n \rightarrow \infty} k_{jn}^*/n$.

- (b) $\{X_j^*\}$ son i.i.d. con distribución μ e independientes de $\{p_j\}$.
- (c) Dada μ^* , $\{X_j\}_{j \geq 1}$ son condicionalmente independientes con distribución μ^* .

- (d) Si $w_{0n}/\xi(n) \rightarrow 0$ c.s. cuando $n \rightarrow \infty$, entonces μ^* es discreta con probabilidad uno. Es decir,

$$\mu^* = \sum_j p_j \delta_{X_j^*}. \quad (2.7)$$

Algunos casos particulares de urnas de Pólya generalizadas que satisfacen estas condiciones son los siguientes:

- (a) Una muestra de observaciones independientes e idénticamente distribuidas.

$$X_n \stackrel{iid}{\sim} \mu, \quad n \geq 1. \quad (2.8)$$

Es la forma más evidente de intercambiabilidad y se obtiene haciendo $w_{0n} = 1$ y $w_{jn} = 0$.

- (b) N valores elegidos al azar. Sea $N > 1$ un entero positivo y defínase $w_{0n} = (N - k_n^*) \mathbb{I}_{\{k_n^* < N\}}$, $w_{jn} = 1$. Cuando $k_n^* \geq N$, w_{0n} se anula, lo cual restringe a la sucesión a tener a lo más N valores distintos. Además, la condición (2.5) se satisface porque

$$\sum_{j=0}^{k_n^*} w_{jn} = N > 0 \quad (2.9)$$

- (c) La sucesión de Blackwell y MacQueen con parámetro μ . Corresponde a la elección $w_{0n} = \mu(E)$ y $w_{jn} = k_{jn}^*$. En este caso

$$\sum_{j=0}^{k_n^*} w_{jn} = \mu(E) + n > 0. \quad (2.10)$$

Este modelo es particularmente importante por caracterizar al proceso Dirichlet, como se verá en la próxima sección.

- (d) El proceso Poisson-Dirichlet con parámetros (ν, α, θ) . Corresponde a la elección $w_{0n} = \theta + \alpha k_n^*$ y $w_{jn} = k_{jn}^* - \alpha$ para $0 \leq \alpha < 1$ y $\theta > -\alpha$. En este caso

$$\sum_{j=0}^{k_n^*} w_{jn} = \theta + n > 0. \quad (2.11)$$

La expresión (2.4) resulta ser la distribución predictiva (en el sentido bayesiano de la sección 1.3) para una medida aleatoria discreta $\tilde{\mu}$ definida por

$$\tilde{\mu}(\cdot) = V_0 \delta_{Z_0}(\cdot) + \sum_{k \geq 1} \left[\prod_{i=0}^{k-1} (1 - V_i) V_k \right] \delta_{Z_k}(\cdot), \quad (2.12)$$

donde $\{V_k\}_{k \geq 0}$ son iid con distribución $\text{Be}(\cdot | 1 - \alpha, \theta + k\alpha)$ e independientes de $\{Z_k\}_{k \geq 0}$ que son a su vez iid con distribución ν . El proceso Dirichlet con parámetro μ es un caso particular de este modelo y se obtiene tomando $\alpha = 0$ y $\mu = \theta\nu$.

- (e) Modelo de distribuciones Dirichlet finito dimensionales de Fisher con parámetros (θ, ν) . Sea $N > 1$ y defínase $w_{0n} = \theta(1 - k_n^*/N) \mathbb{I}_{\{k_n^* < N\}}$, $w_{jn} = k_{jn}^* + \theta/N$, para $\theta > 0$. Entonces

$$\sum_{j=0}^{k_n^*} w_{jn} = \theta + n > 0. \quad (2.13)$$

Además, la elección de w_{0n} evita que el proceso tenga más de N valores distintos. En este caso, la expresión (2.4) corresponde a la distribución predictiva para una distribución aleatoria $\tilde{\mu}$ definida por

$$\tilde{\mu}(\cdot) = \sum_{k=1}^N \frac{G_k}{\sum_{k=1}^N G_k} \delta_{Z_k}(\cdot), \quad (2.14)$$

donde $\{g_k\}$ son variables aleatorias iid $\text{Ga}(\cdot | \theta/N, 1)$ e independientes de $\{Z_k\}_{k \geq 0}$ que son iid con distribución ν .

2.2. El proceso Dirichlet.

En 1973, Ferguson [Fer73] introdujo el proceso Dirichlet, definiéndolo a partir de sus distribuciones finito dimensionales y demostrando la existencia del proceso a través de las condiciones de consistencia de Kolmogorov. Para ello se basó en las propiedades de la distribución Dirichlet, conocida en el ámbito de la estadística bayesiana por ser la distribución inicial conjugada para los parámetros de la distribución multinomial.

Definición 2.2.1 (Distribución Dirichlet). Sean Z_1, \dots, Z_n variables aleatorias independientes tales que $Z_j \sim \text{Ga}(z|a_j, 1)$, donde $a_j \geq 0$ para

toda j y $a_j > 0$ para algunas j 's. La distribución Dirichlet con parámetro (a_1, \dots, a_n) es la distribución de (Y_1, \dots, Y_n) , donde para $j = 1, \dots, n$,

$$Y_j = \frac{Z_j}{\sum_{i=1}^n Z_i}. \quad (2.15)$$

En este caso, se denota $(Y_1, \dots, Y_n) \sim \mathcal{D}(\cdot | a_1, \dots, a_n)$.

Definición 2.2.2 (Proceso Dirichlet). Sea ν una medida no nula, no negativa y finitamente aditiva sobre un espacio medible (E, \mathcal{F}) , tal que $\nu(E) < \infty$. Se dice que una medida aleatoria μ sobre (E, \mathcal{F}) se distribuye de acuerdo con un proceso Dirichlet con parámetro ν si para cada $n = 1, 2, \dots$ y para cada partición medible (B_1, \dots, B_n) de E , el vector $(\mu(B_1), \dots, \mu(B_n))$ tiene distribución Dirichlet, $\mathcal{D}(\cdot | \nu(B_1), \dots, \nu(B_n))$. En este caso, se usa la notación $\mu \sim \mathcal{D}(\mu | \nu)$.

A partir de su introducción, el proceso Dirichlet ha sido ampliamente estudiado y utilizado. Parte de su importancia se deriva de las diferentes formas de caracterizarlo que, a través de generalizaciones, han dado origen a nuevos procesos o distribuciones para medidas aleatorias. Una de estas caracterizaciones, presentada por Ferguson [Fer73] y retomada por Pitman en el contexto de muestreo de especies [Pit96], se basa en la llamada distribución Poisson-Dirichlet.

Definición 2.2.3 (Distribución Poisson-Dirichlet). Sea $\{Z_t\}_{t \geq 0}$ un proceso Gamma con medida de intensidad $(e^{-x}/x) dx$. Es decir, un proceso estocástico con incrementos independientes tal que para cada t , $Z_t \sim \text{Ga}(\cdot | t, 1)$. Sean $\Gamma_1 > \Gamma_2 > \dots$ los tamaños ordenados de los saltos del proceso para $t \in [0, \theta]$ y defínase

$$p_n = \frac{\Gamma_n}{Z_\theta}. \quad (2.16)$$

Entonces se dice que $\{p_n\}_{n \geq 1}$ tiene la distribución Poisson-Dirichlet con parámetro θ que se denota $\text{PD}(\cdot | \theta)$.

Teorema 2.2.4. Sea ν una medida como en la definición 2.2.2. Defínase $\theta := \nu(E)$ y $\bar{\nu} := \nu/\theta$ y sea $\{p_n\}_{n \geq 1} \sim \text{PD}(\cdot | \theta)$. Si

$$\mu := \sum_{i=1}^{\infty} p_i \delta_{Z_i}, \quad (2.17)$$

con $Z_i \stackrel{iid}{\sim} \bar{\nu}$ e independientes de $\{p_n\}_{n \geq 1}$, entonces $\mu \sim \mathcal{D}(\cdot | \nu)$.

Una caracterización alternativa para la distribución Poisson-Dirichlet da lugar a una definición constructiva para el proceso Dirichlet.

Teorema 2.2.5. *Sea $\{W_n\}_{n \geq 1}$ una sucesión de variables aleatorias iid con distribución $\text{Be}(\cdot | 1, \theta)$ y defínase*

$$\tilde{p}_n = \left[\prod_{j=1}^{n-1} (1 - W_j) \right] W_n \quad (2.18)$$

Sea $\{p_n\}_{n \geq 1}$ una sucesión obtenida al ordenar $\{\tilde{p}_n\}$ en orden descendente. Entonces $\{p_n\}_{n \geq 1} \sim \text{PD}(\cdot | \theta)$

En este caso, $p_1 > p_2 > \dots > 0$ y $\sum_i p_i = 1$ c.s. Por lo tanto, si se sustituye p_n por \tilde{p}_n en la igualdad (2.17), se obtiene la llamada representación de asignación proporcional o «stick-breaking» para el proceso Dirichlet.

Como se mencionó en la sección anterior, Blackwell y MacQueen [BM73] introdujeron una caracterización del proceso Dirichlet a través de una urna de Pólya generalizada, la cual permite obtener de forma iterativa una muestra de un proceso Dirichlet a través de la distribución predictiva (2.4). El resultado se presenta a continuación.

Teorema 2.2.6. *Sea $\{X_n\}_{n \geq 1}$ una sucesión con distribución definida por una urna de Pólya generalizada con parámetro ν , tal que $\nu(E) = \theta < \infty$. Es decir*

$$\mathbb{P}[X_1 \in \cdot] = \nu(\cdot) \quad (2.19)$$

$$\mathbb{P}[X_{n+1} \in \cdot | X_1, \dots, X_n] = \frac{\nu(\cdot) + \sum_{j=1}^n \delta_{X_j}(\cdot)}{\theta + n}. \quad (2.20)$$

Entonces,

(a) *La sucesión $\{\mu_n\}_{n \geq 1} \subset \mathcal{P}(E)$ definida por*

$$\mu_n = \frac{\nu + \sum_{j=1}^n \delta_{X_j}}{\theta + n} \quad (2.21)$$

converge c.s. a una medida discreta $\mu \in \mathcal{P}(E)$.

(b) $\mu \sim \mathcal{D}(\cdot | \nu)$.

(c) *Dada μ las variables X_1, X_2, \dots son condicionalmente independientes con distribución μ .*

A continuación se presenta una serie de resultados relacionados con el proceso Dirichlet que serán útiles para el desarrollo del capítulo 4.

Lema 2.2.7. Si $\mu \sim \mathcal{D}(\cdot | \nu)$, entonces μ es una medida discreta casi seguramente.

Lema 2.2.8. Sea X_1, \dots, X_n una muestra de tamaño n de un proceso Dirichlet μ con parámetro ν . Es decir,

$$X_i \stackrel{iid}{\sim} \mu, \quad i = 1, \dots, n \quad (2.22)$$

$$\mu \sim \mathcal{D}(\cdot | \nu) \quad (2.23)$$

Entonces, la distribución condicional de μ dada la muestra es a su vez un proceso Dirichlet,

$$\mu | X_1, \dots, X_n \sim \mathcal{D}(\cdot | \nu + \sum_{i=1}^n \delta_{X_i}) \quad (2.24)$$

Obsérvese que de este resultado se sigue que

$$\mathcal{D}(\cdot | \nu) = \int_{E^n} \mathcal{D}(\cdot | \nu + \sum_{i=1}^n \delta_{X_i}) \mu(X_1) \dots \mu(X_n) \quad (2.25)$$

Lema 2.2.9. Sea \mathbb{X}^n un vector aleatorio de dimensión n con valores en un espacio E^n y distribución

$$\mathbb{P}(\mathbb{X}^n | \mu) = \prod_{j=1}^n \mu(X_j), \quad (2.26)$$

donde, para $\nu \in \mathcal{P}(E)$ y $\theta > 0$ fijas

$$\mu \sim \mathcal{D}(\cdot | \theta \nu). \quad (2.27)$$

Entonces, se puede integrar μ para obtener

$$\mathbb{P}[\mathbb{X}^n] := \int_{\mathcal{P}(E)} \mathbb{P}(\mathbb{X}^n | \mu) \mathcal{D}(d\mu | \theta \nu) = \prod_{j=1}^n \left(\frac{\theta \nu + \sum_{i=1}^{j-1} \delta_{X_i}}{\theta + j - 1} \right). \quad (2.28)$$

Este lema nos dice que la distribución marginal de un vector aleatorio formado por observaciones condicionalmente independientes e idénticamente distribuidas de acuerdo a una medida aleatoria con distribución Dirichlet, es igual a la distribución de un vector aleatorio obtenido a partir del modelo de urna de Pólya generalizada:

$$X_1 \sim \nu \quad (2.29)$$

y para $j = 2, \dots, n$,

$$X_j \sim \frac{\theta\nu + \sum_{i=1}^{j-1} \delta_{X_i}}{\theta + j - 1}. \quad (2.30)$$

El siguiente teorema, presentado por Walker, *et al.* [WHN07], es esencial en la construcción del proceso Fleming-Viot a través de un proceso latente, que se verá en el capítulo 4. Se presenta en esta sección por tratarse de una propiedad de un modelo cuya distribución inicial no paramétrica es un proceso Dirichlet, la cual se satisface en general y no únicamente en el contexto del capítulo 4.

Teorema 2.2.10. Sean $\mathbb{X}^n := (X_1, \dots, X_n)$ y $\mathbb{Y}^m := (Y_1, \dots, Y_m)$ dos vectores aleatorios con valores en E^n y E^m respectivamente, y distribuidos de acuerdo al modelo

$$X_j \stackrel{iid}{\sim} \mu \quad j \in \{1, \dots, n\} \quad (2.31)$$

$$Y_j | \tilde{\mu} \stackrel{iid}{\sim} \tilde{\mu} \quad j \in \{1, \dots, m\} \quad (2.32)$$

$$\tilde{\mu} | \mathbb{X}^n \sim \mathcal{D}(\cdot | \theta\nu + \sum_{j=1}^n \delta_{X_j}), \quad (2.33)$$

para cualesquiera $\theta > 0$ y $\nu, \mu \in \mathcal{P}(E)$ no atómicas. Sean $\{Y_1^*, \dots, Y_m^*\}$ el conjunto de valores únicos de \mathbb{Y}^m y R el número de elementos de dicho conjunto que coinciden con algún elemento de $\{X_1, \dots, X_n\}$. Entonces, la densidad de R está definida, para $r \in \{0, \dots, \min\{n, m\}\}$, por

$$\mathbb{P}[R = r | \mathbb{X}^n] = \mathbb{P}[R = r] = \frac{\binom{n}{r}_\downarrow (\theta + r)_{m-r\uparrow}}{(\theta + n)_{m\uparrow}} \binom{m}{r} \quad (2.34)$$

o equivalentemente,

$$\mathbb{P}[R = r] = r! \binom{n}{r} \binom{m}{r} \frac{(\theta)_{n\uparrow} (\theta)_{m\uparrow}}{(\theta)_{n+m\uparrow} (\theta)_{r\uparrow}} \quad (2.35)$$

Para la demostración se requiere el siguiente resultado de combinatoria, tomado del mismo artículo [WHN07].

Lema 2.2.11. Sea $\phi \geq 0$. Entonces, para $m \geq r > 0$ y definiendo $0! \equiv 1$, se tiene que

$$\sum_{k=0}^{m-r} k! \binom{k+r-1}{k} \binom{m-r}{k} (\phi)_{m-r-k\uparrow} = (\phi + r)_{m-r\uparrow}. \quad (2.36)$$

Demostración. (Teorema 2.2.10.) Supóngase, sin pérdida de generalidad, que las r observaciones de \mathbb{X}^n que reaparecen en $\{Y_1^*, \dots, Y_{m^*}^*\}$ son las primeras. Denótese como s_i a la multiplicidad o número de veces que se repite la observación X_i en la muestra \mathbb{Y}^m (sin contar la primera). Entonces $k = s_1 + s_2 + \dots + s_r$ es el número total de observaciones repetidas de \mathbb{Y}^m que coinciden con alguna observación de \mathbb{X}^n , con $0 \leq k \leq m-r$ y $0 \leq s_i \leq m-r$.

El modelo de urnas que rige la distribución condicional de $[\mathbb{Y}^m | \mathbb{X}^n]$, de acuerdo con el lema 2.2.9, permite recurrir a las distribuciones condicionales para calcular la probabilidad de obtener una configuración específica para \mathbb{Y}^m . Considérese los siguientes eventos

$$A_1 := [Y_1 = X_1]$$

$$A_j := A_{j-1} \cap [Y_j = X_j], \quad 2 \leq j \leq r.$$

$$A_{r+j} := A_{r+j-1} \cap [Y_{r+j} = X_1], \quad 1 \leq j \leq S_1.$$

$$A_{r+s_1+j} := A_{r+s_1+j-1} \cap [Y_{r+s_1+j} = X_2], \quad 1 \leq j \leq S_2.$$

⋮

$$A_{r+k-s_r+j} := A_{r+k-s_r+j-1} \cap [Y_{r+k-s_r+j} = X_r], \quad 1 \leq j \leq S_r.$$

$$A_{r+k+j} := A_{r+k+j-1} \cap [Y_{r+k+j} \in E \setminus \{X_1, \dots, X_n\}], \quad 1 \leq j \leq m-r-k.$$

Entonces,

$$\begin{aligned} \mathbb{P}[Y_1 = X_1 | \mathbb{X}^n] &= \frac{1}{\theta + n} \\ \mathbb{P}[Y_j = X_j | \mathbb{X}^n, A_{j-1}] &= \frac{1}{\theta + n + j - 1} \\ \mathbb{P}[Y_{r+j} = X_1 | \mathbb{X}^n, A_{r+j-1}] &= \frac{1 + j}{\theta + n + r + j - 1} \\ \mathbb{P}[Y_{r+s_1+j} = X_2 | \mathbb{X}^n, A_{r+s_1+j-1}] &= \frac{1 + j}{\theta + n + r + s_1 + j - 1} \\ &\vdots \\ \mathbb{P}[Y_{r+k-s_r+j} = X_r | \mathbb{X}^n, A_{r+k-s_r+j-1}] &= \frac{1 + j}{\theta + n + r + k - s_r + j - 1} \\ \mathbb{P}[Y_{r+k+j} \in E \setminus \{X_1, \dots, X_n\} | \mathbb{X}^n, A_{r+k+j-1}] &= \frac{\theta + j - 1}{\theta + n + r + k + j - 1} \end{aligned}$$

y la probabilidad de ocurrencia simultánea de estos eventos es

$$\mathbb{P}[A_m | \mathbb{X}^n] = \prod_{i=1}^m \mathbb{P}[A_i | A_{i-1}] = \frac{(s_1 + 1)! \cdots (s_r + 1)! (\theta)_{m-r-k \uparrow}}{(\theta + n)_{m \uparrow}}, \quad (2.37)$$

donde $A_0 = E$. Ahora, como $Y_i | \tilde{\mu} \stackrel{iid}{\sim} \tilde{\mu}$, por el teorema de representación de de Finetti (teorema 1.3.2), las observaciones y por tanto los eventos, son intercambiables. Así pues, la probabilidad de que la muestra \mathbb{Y}^m contenga al valor X_i , $s_i + 1$ veces (la primera más las repeticiones), para cada $i = 1 \dots r$, con k y $\{s_1, \dots, s_r\}$ fijos, sin importar el orden de aparición en \mathbb{Y}^n , es igual a $\mathbb{P}[A_m | \mathbb{X}^n]$ multiplicado por el número de formas en que se puede elegir k observaciones con $\{s_1, \dots, s_r\}$ repeticiones, del total de m observaciones en \mathbb{Y}^m . Dicha probabilidad es

$$\left[\frac{m!}{\prod_{i=1}^r (s_i + 1)! (m - r - k)!} \right] \left[\frac{\prod_{i=1}^r (s_i + 1)! (\theta)_{m-r-k \uparrow}}{(\theta + n)_{m \uparrow}} \right] = \frac{m! (\theta)_{m-r-k \uparrow}}{(m - r - k)! (\theta + n)_{m \uparrow}} \quad (2.38)$$

Si se permite que k y s_1, \dots, s_r varíen, se obtiene la probabilidad de que exactamente las primeras r componentes de \mathbb{X}^n aparezcan en $\{Y_1^*, \dots, Y_m^*\}$ cualquier número de veces, dada por

$$\sum_{k=0}^{m-r} \sum_{\{s_1 + \dots + s_r = k\}} \frac{m! (\theta)_{m-r-k \uparrow}}{(m - r - k)! (\theta + n)_{m \uparrow}} = \sum_{k=0}^{m-r} \frac{m! (\theta)_{m-r-k \uparrow}}{(m - r - k)! (\theta + n)_{m \uparrow}} \binom{k + r - 1}{k}, \quad (2.39)$$

donde la igualdad se debe a que los sumandos no dependen de s_1, \dots, s_r , por lo tanto la suma interior simplemente es igual al número de formas en que k se puede descomponer como suma de r enteros positivos. Reacomodando términos, esto es igual a

$$\frac{m!}{(m - r)!} \sum_{k=0}^{m-r} k! \binom{k + r - 1}{k} \binom{m - r}{k} \frac{(\theta)_{m-r-k \uparrow}}{(\theta + n)_{m \uparrow}} \quad (2.40)$$

y aplicando el lema 2.2.11, esto es igual a

$$\frac{m! (\theta + r)_{m-r \uparrow}}{(m - r)! (\theta + n)_{m \uparrow}} \quad (2.41)$$

Finalmente, para obtener la probabilidad de que r valores únicos $\{Y_1^*, \dots, Y_m^*\}$ sean iguales a cualesquiera elementos de \mathbb{X}^n , es necesario multiplicar por el número de formas en que se puede elegir r de las n X_j 's, sin repetición. Así se llega al resultado deseado

$$\mathbb{P}[r(t+s) = r | n(t) = n, n(t+s) = m] = \binom{n}{r} \frac{m!(\theta+r)_{m-r\uparrow}}{(m-r)!(\theta+n)_{m\uparrow}} = \frac{(n)_{r\downarrow}(\theta+r)_{m-r\uparrow}}{(\theta+n)_{m\uparrow}} \binom{m}{r} \quad (2.42)$$

□

Lema 2.2.12. *Bajo las condiciones del teorema 2.2.10, la distribución de \mathbb{Y}^m dado R ,*

$$\mathbb{P}[\mathbb{Y}^m | R] := \int_{E^n} \mathbb{P}[\mathbb{Y}^m | \mathbb{X}^n, R] d\mu(X_1), \dots, d\mu(X_n) \quad (2.43)$$

corresponde al modelo

$$Y_j \stackrel{iid}{\sim} \mu, \quad j \in \{1, \dots, R\} \quad (2.44)$$

$$Y_j | \tilde{\mu}_Y \stackrel{iid}{\sim} \tilde{\mu}_Y, \quad j \in \{R+1, \dots, m\} \quad (2.45)$$

$$\tilde{\mu}_Y | \mathbb{Y}^R \sim \mathcal{D}(\cdot | \theta\nu + \sum_{j=1}^R \delta_{Y_j}) \quad (2.46)$$

De acuerdo con el lema 2.2.7, el proceso Dirichlet, usado como distribución inicial no paramétrica tiene fuertes limitaciones, pues asigna probabilidad 1 al espacio $\mathcal{P}_d(E) \subset \mathcal{P}(E)$ de medidas de probabilidad discretas. Antoniak [Ant74] extendió los resultados de Ferguson [Fer73], dando origen a un modelo de mezclas de procesos Dirichlet. Definió así una nueva distribución inicial no paramétrica, que puede ser interpretada como un proceso Dirichlet $\mathcal{D}(\cdot | \nu)$, donde ν es a su vez una medida aleatoria.

Definición 2.2.13 (Mezcla de procesos Dirichlet). Sean (E, \mathcal{F}) un espacio medible y (U, \mathcal{A}, H) un espacio de probabilidad. Se dice que una medida aleatoria μ sobre (E, \mathcal{F}) se distribuye de acuerdo con una mezcla de procesos Dirichlet con distribución de mezclas H sobre el espacio de índices (U, \mathcal{A}) y medida de transición α , si para cada partición medible B_1, \dots, B_k de E y $k \geq 1$ se tiene que

$$\begin{aligned} \mathbb{P}[\mu(B_1) \leq y_1, \dots, \mu(B_k) \leq y_k] \\ = \int_U \mathcal{D}((y_1, \dots, y_k) | \alpha(u, B_1), \dots, \alpha(u, B_k)) dH(u), \end{aligned}$$

donde $\alpha : U \times \mathcal{F} \rightarrow [0, \infty)$ es tal que

- (a) Para cada $u \in U$, $\alpha(u, \cdot)$ es una medida finita no nula y no negativa sobre (E, \mathcal{F}) .
- (b) Para cada $B \in \mathcal{F}$, $\alpha(\cdot, B)$ es medible en (U, \mathcal{A}) .

En este caso, se denota

$$\mu \sim \mathcal{MD}(\cdot | \alpha, H) := \int_U \mathcal{D}(\cdot | \alpha(u, \cdot)) dH(u).$$

Antoniak demostró propiedades analíticas análogas a las del proceso Dirichlet, por ejemplo la propiedad de cerradura dada por el lema 2.2.8.

La idea de las mezclas de procesos Dirichlet permite también definir un modelo no paramétrico del tipo (1.52) que resulta más realista en el caso en que las variables X_k son continuas. Esto se logra a través de la introducción de una distribución paramétrica conocida como Kernel de mezclas. Aunque dicho modelo, propuesto por Lo en 1984 [Lo84] no es la única inicial no paramétrica que asigna probabilidad positiva a las medidas de probabilidad continuas, es probablemente la más utilizada, pues su relación con el proceso Dirichlet la hace muy manejable. Además, los algoritmos de simulación desarrollados para el proceso Dirichlet pueden extenderse fácilmente al modelo de mezclas, el cual puede ser también representado como una urna de Pólya generalizada.

Otro ejemplo de extensión del proceso Dirichlet lo constituyen los llamados procesos Dirichlet dependientes. En pocas palabras, se trata de modelos jerárquicos en los que el parámetro del proceso Dirichlet es a su vez una medida aleatoria que se distribuye de acuerdo con un proceso Dirichlet.

En resumen, aunque el proceso Dirichlet por sí mismo no resulta útil para el modelado de muchos problemas que se pueden plantear en el ámbito de la estadística bayesiana no paramétrica, ha sido ampliamente estudiado y utilizado. Ha dado origen a diversas extensiones y generalizaciones a partir de las cuales han surgido nuevas familias de distribuciones iniciales no paramétricas. Tiene, por lo tanto una importancia tanto histórica, pues dio origen a la inferencia bayesiana no paramétrica en la práctica, como actual, pues es la base de una variedad de familias de modelos muy utilizados.

Capítulo 3

Proceso Fleming-Viot.

En los últimos años se ha desarrollado intensamente la teoría en torno a una amplia familia de procesos estocásticos que toman valores en espacios de medidas. Se trata de procesos que surgen como una extensión de la idea de aproximaciones de difusión de Feller («Feller diffusion approximation») para plantear modelos poblacionales que incluyan tanto el tamaño de la población como la situación de los individuos que la forman. Dos ejemplos básicos de este tipo de modelos son el proceso de Dawson-Watanabe, en que el tamaño de la población evoluciona de acuerdo con un modelo de difusión de Feller, y el proceso de Fleming-Viot en que la composición de la población cambia, pero su tamaño permanece constante.

El proceso Fleming-Viot fue introducido en 1979 por Fleming y Viot, como un modelo de genética poblacional [FV79]. Constituye una generalización del proceso de difusión de Wright-Fisher para un espacio de estados finito. El proceso, que toma valores en un espacio de medidas de probabilidad $\mathcal{P}(E)$ sobre un espacio de estados (o tipos) E infinito, modela la distribución de tipos dentro de una población de tamaño fijo. El proceso se puede construir como el límite en distribución de algunas sucesiones de cadenas de Markov ampliamente utilizadas en el contexto de genética poblacional. En este capítulo, tras una breve introducción, se presenta en la segunda sección tres modelos que pueden dar origen a dicha construcción. Posteriormente, se da la forma general del generador del proceso y su caracterización como solución a un problema de martingala, planteado originalmente por Fleming y Viot. Finalmente, en la última sección se introduce un caso particular, el modelo de mutación neutral para alelos infinitos («Infinitely-many-neutral-alleles model»), al que se llamará Modelo Fleming-Viot de mutación neutral, en el resto del trabajo. Se menciona algunas propiedades de este proceso,

como su función de transición markoviana y su distribución estacionaria, las cuales se recuperan en la construcción planteada en el capítulo 4.

Una introducción a los superprocesos, aunque enfocada al proceso de Dawson-Watanabe, puede encontrarse en el libro de A. Etheridge [Eth00]. Para un recuento de propiedades y casos particulares del proceso Fleming-Viot se puede consultar [EK93], de donde se tomaron los resultados presentados en este capítulo.

3.1. Modelos de difusión y genética poblacional.

La teoría moderna de la evolución, descansa fuertemente sobre los fundamentos de la genética poblacional que permitió integrar la teoría de selección natural con la genética mendeliana. Ronald A. Fisher [Fis99], J.B.S. Haldane [Hal24] y Sewall Wright [Wri32, Wri84] son reconocidos como los fundadores del área, pues a través de la formalización de conceptos en modelos matemáticos, lograron explicar de forma satisfactoria, cómo interactúan los mecanismos de selección natural y herencia genética en el desarrollo de las poblaciones.

En términos generales, la genética poblacional utiliza modelos probabilísticos, con mecanismos de transición, para estudiar la distribución de frecuencias de alelos en una población con desarrollo evolutivo. La frecuencia de un alelo es la proporción o porcentaje de veces, dentro de una población, que un alelo dado se encuentra en una posición específica de un cromosoma. La importancia del estudio de dichas frecuencias en genética poblacional, consiste en la posibilidad de utilizarlos como una medida de la diversidad genética en la población estudiada.

Algunas de las fuerzas evolutivas principales conocidas y susceptibles de ser incorporadas en un modelo de este tipo son:

- **Selección natural.** El término fue introducido por Charles Darwin en 1859, en su conocido libro «El origen de las especies» [Dar59]. Se refiere al mecanismo por el cual las diferencias entre individuos se reflejan de alguna forma en su apariencia externa o funcional (fenotipo), afectando su capacidad reproductiva. De este modo, los individuos con características menos favorables para adaptarse a su medio ambiente, pueden tener menor oportunidad de engendrar una progenie (selección por fecundidad) o pueden tener menor oportunidad para sobrevivir y por tanto para tener descendencia (selección por viabilidad). En ambos casos, una menor capacidad para enfrentarse a las condiciones internas y externas se refleja en una menor capacidad para transmitir dichas

características a través de la herencia. Sin embargo, debe tenerse en cuenta que, aunque este mecanismo afecta la dinámica de los tipos (fenotipos y genotipos) en la población, no necesariamente implica que las características menos favorables tenderán a desaparecer.

- **Deriva genética.** Su nombre preciso debería ser deriva de alelos, pues se refiere al cambio en frecuencias de alelos de una generación a otra, debido a factores que determinan al azar, qué tipos de alelos (variantes genéticas) se preservan y cuáles desaparecen. A diferencia de la selección natural, la deriva no se relaciona con la capacidad de adaptación u otras características del individuo. Se refiere a cambios aleatorios que ocurren de un individuo a otro y que, especialmente en poblaciones pequeñas, afectan la distribución de frecuencias de alelos.
- **Mutación.** Se refiere a un mecanismo aleatorio, no necesariamente relacionado con la deriva, que produce cambios en la estructura genética de un individuo y cuyo efecto acumulado se puede reflejar en la distribución de frecuencias de alelos de la población.
- **Migración.** Se refiere al movimiento de individuos en el espacio, ocasionando la salida y llegada de individuos nuevos a una población, modificando la distribución de tipos genéticos.
- **Recombinación.** Se refiere al efecto de la reproducción «sexual», es decir, aquella en la que las características de dos individuos se mezclan en un nuevo individuo o descendiente.

La teoría matemática o probabilista desarrollada en torno a la genética poblacional («mathematical population genetics») permite modelar una población cuyos individuos pueden ser clasificados por «tipos», cada uno de los cuales se representa por un elemento x en algún conjunto E . A través de estos modelos, se hace posible el estudio de la distribución de los tipos en la población total a través del tiempo, por medio de procesos estocásticos. Existen diversos procesos de esta naturaleza, dependiendo del espacio de tipos E (si es finito o infinito) y de la forma de incorporar los mecanismos evolutivos que modifican la estructura de tipos en la población al paso del tiempo. Uno de ellos es el proceso Fleming-Viot, cuya importancia se debe en gran medida a que incorpora, en su forma general, una amplia gama de modelos poblacionales.

3.2. Cadenas de Markov con valores en un espacio de medidas.

El proceso Fleming-Viot surge como el límite, en distribución, de algunas sucesiones de cadenas de Markov que aparecen en el contexto de genética de población. Esta sección introduce tres de ellos. En adelante, se supondrá que (E, r) es un espacio métrico separable y completo. En la mayoría de las aplicaciones E es compacto o localmente compacto, de modo que además se supondrá compacidad cuando resulte conveniente.

3.2.1. Modelo diploide.

Se define una relación de equivalencia en $E^2 = E \times E$ por $(x, y) \sim (z, w)$ si y sólo si $(x, y) = (z, w)$ o $(x, y) = (w, z)$. El espacio $E^{(2)}$ de clases de equivalencia puede pensarse como el conjunto de todos los pares no ordenados $\{x, y\}$ de elementos de E . En una población diploide, los cromosomas aparecen en pares, de modo que el genotipo que describe a un individuo se representa como un elemento de $E^{(2)}$.

Es posible definir una métrica $r^{(2)}$ en $E^{(2)}$ como

$$r^{(2)}(\{x, y\}, \{z, w\}) := (r(x, z) + r(y, w)) \wedge (r(x, w) + r(y, z)). \quad (3.1)$$

y el mapeo $\rho : E^2 \rightarrow E^{(2)}$ dado por $\rho(x, y) = \{x, y\}$ resulta continuo bajo esta métrica. Además, si f es una función simétrica y Borel-medible en E^2 , entonces la función g en $E^{(2)}$ definida por $g(\{x, y\}) := f(x, y) = f(y, x)$ es también Borel-medible.

Dada $\mu \in \mathcal{P}(E^{(2)})$, una medida de probabilidad sobre $E^{(2)}$, su *simetrización*, $\hat{\mu} \in \mathcal{P}(E^2)$ es la medida de probabilidad sobre E^2 definida por

$$\hat{\mu}(\Gamma) = \int_{E^{(2)}} \frac{1}{2} (\delta_{(x,y)}(\Gamma) + \delta_{(y,x)}(\Gamma)) \mu(d\{x, y\}). \quad (3.2)$$

Por ejemplo, si $\mu = \delta_{\{x,y\}}$, es la medida que acumula toda su masa en $\{x, y\} \in E^{(2)}$, entonces $\hat{\mu} = \frac{1}{2}(\delta_{(x,y)} + \delta_{(y,x)})$ es la medida cuya masa se distribuye a partes iguales entre los puntos $(x, y), (y, x) \in E^2$. Obsérvese que $\hat{\mu}\rho^{-1} = \mu$, de modo que μ puede ser recuperada a partir de su simetrización. Se dice que $\mu \in \mathcal{P}(E^{(2)})$ está en la *forma de Hardy-Weinberg* si $\hat{\mu}$ es una medida producto, es decir, si $\hat{\mu} = \hat{\mu}\pi^{-1} \times \hat{\mu}\pi^{-1} = (\hat{\mu}\pi^{-1})^2$, donde π es la proyección de E^2 sobre su primera coordenada.

Este modelo y los dos siguientes dependen de ciertos parámetros que se introducen a continuación.

Para cada entero positivo M , sea w_M una función Borel-medible en E^2 , positiva, simétrica y acotada. Sea R_M una función de transición a un paso en $E^2 \times \mathcal{B}(E^2)$ tal que

$$R_M((x, y), dx' \times dy') = R_M((y, x), dy' \times dx'). \quad (3.3)$$

Sea Q_M una función de transición a un paso en $E \times \mathcal{B}(E)$. Las funciones w_M , R_M y Q_M incorporan al modelo los mecanismos de selección, recombinación y mutación, respectivamente.

Sea $N \geq 1$ el tamaño total de la población y denótese por $\eta_N : (E^{(2)})^N \rightarrow \mathcal{P}(E^{(2)})$ la distribución empírica,

$$\eta_N(\{x_1, y_1\}, \dots, \{x_N, y_N\}) = \frac{1}{N} \sum_{i=1}^N \delta_{\{x_i, y_i\}}. \quad (3.4)$$

Dada $\mu \in \mathcal{P}(E^{(2)})$, se define $\hat{\mu}_{2N}^*$, $\hat{\mu}_{2N}^{**}$ y $\hat{\mu}_{2N}^{***}$ en $\mathcal{P}(E^{(2)})$ como

$$\hat{\mu}_{2N}^*(dx \times dy) = \frac{w_{2N}(x, y)(\hat{\mu}\pi^{-1})^2(dx \times dy)}{\langle w_{2N}, (\hat{\mu}\pi^{-1})^2 \rangle}, \quad (3.5)$$

$$\hat{\mu}_{2N}^{**}(dx' \times dy') = \int_{E^2} R_{2N}((x, y), dx' \times dy') \hat{\mu}_{2N}^*(dx \times dy), \quad (3.6)$$

$$\hat{\mu}_{2N}^{***}(dx' \times dy') = \int_{E^2} Q_{2N}(x, dx') Q_{2N}(y, dy') \hat{\mu}_{2N}^{**}(dx \times dy), \quad (3.7)$$

donde $\langle f, \mu \rangle = \int_E f d\mu$ es el producto escalar, definido para cualquier función f acotada en E y $\mu \in \mathcal{P}(E)$. La condición (3.3) garantiza que $\hat{\mu}_{2N}^{**}$ es la simetrización de una medida en $\mathcal{P}(E^{(2)})$.

El modelo diploide se define como una cadena de Markov con espacio de estados:

$$\mathcal{P}_N(E^{(2)}) := \eta_N((E^{(2)})^N) \subset \mathcal{P}(E^{(2)}) \quad (3.8)$$

y función de transición a un paso $P_N(\mu, d\nu)$ en $\mathcal{P}_N(E^{(2)}) \times \mathcal{B}(\mathcal{P}_N(E^{(2)}))$ dada por

$$P_N(\mu, \cdot) = \int_{(E^{(2)})^N} \delta_{\eta_N(\{x_1, y_1\}, \dots, \{x_N, y_N\})}(\cdot) (\hat{\mu}_{2N}^{***})^N(d\{x_1, y_1\}, \dots, d\{x_N, y_N\}), \quad (3.9)$$

donde $\mu^N = \mu \times \mu \times \dots \times \mu \in \mathcal{P}((E^{(2)})^N)$ es la correspondiente medida producto.

La interpretación del modelo es la siguiente. Si $\mu \in \mathcal{P}_N(E^{(2)})$ es la distribución empírica de los N genotipos en la generación padre, entonces la distribución empírica de los N genotipos en la generación de descendientes se determina a partir de μ siguiendo cuatro pasos:

- (a) Reproducción y selección, regida por la ecuación (3.5).
- (b) Recombinación, regida por la ecuación (3.6).
- (c) Mutación, regida por la ecuación (3.7).
- (d) Deriva aleatoria, regulada por la transición a un paso (3.9).

3.2.2. Modelo de Wright-Fisher.

En una población diploide de tamaño N , hay $M = 2N$ gametos. Este modelo considera únicamente la distribución empírica de los gametos y es matemáticamente más simple que el anterior, pero biológicamente menos razonable. No es necesario que M sea par, por lo que se trabaja con la distribución empírica, $\eta_M : E^M \rightarrow \mathcal{P}(E)$ para cualquier $M \in \mathbb{Z}_+$,

$$\eta_M(x_1, \dots, x_M) = \frac{1}{M} \sum_{i=1}^M \delta_{x_i}. \quad (3.10)$$

Dada $\mu \in \mathcal{P}(E)$, se define $\hat{\mu}_M^* \in \mathcal{P}(E^2)$ y $\hat{\mu}_M^{**}, \hat{\mu}_M^{***} \in \mathcal{P}(E)$ como

$$\hat{\mu}_M^*(dx \times dy) = \frac{w_M(x, y)\mu^2(dx \times dy)}{\langle w_M, \mu^2 \rangle}, \quad (3.11)$$

$$\hat{\mu}_M^{**}(dx') = \int_{E^2} R_M((x, y), dx' \times E) \hat{\mu}_M^*(dx \times dy), \quad (3.12)$$

$$\hat{\mu}_M^{***}(dx') = \int_E Q_M(x, dx') \hat{\mu}_M^{**}(dx), \quad (3.13)$$

para w_M, R_M y Q_M definidas como en la sección 3.2.1.

El modelo de Wright-Fisher se define como una cadena de Markov con espacio de estados:

$$\mathcal{P}_M(E) := \eta_M((E)^M) \subset \mathcal{P}(E) \quad (3.14)$$

y función de transición a un paso $P_M(\mu, d\nu)$ en $\mathcal{P}_M(E) \times \mathcal{B}(\mathcal{P}_M(E))$ dada por

$$P_M(\mu, \cdot) = \int_{(E)^M} \delta_{\eta_M(x_1, \dots, x_M)}(\cdot) (\hat{\mu}_M^{***})^M(dx_1, \dots, dx_M). \quad (3.15)$$

La interpretación del modelo es equivalente a la del modelo anterior. A pesar de que el modelo de Wright-Fisher no se apega a la realidad biológica, al combinar características diploides ((3.11) y (3.12)) y haploides ((3.13)), es un modelo estocástico discreto muy utilizado en genética poblacional.

Es importante destacar que la función de transición $P_M(\mu, d\nu)$ se puede extender de $\mathcal{P}_M(E) \times \mathcal{B}(\mathcal{P}_M(E))$ a $\mathcal{P}(E) \times \mathcal{B}(\mathcal{P}_M(E))$. Por otro lado, si se define $\gamma : \mathcal{P}(E^{(2)}) \rightarrow \mathcal{P}(E)$ como $\gamma(\mu) = \hat{\mu}\pi^{-1}$, la imagen del modelo diploide bajo γ no coincide necesariamente con el modelo Wright-Fisher con $M = 2N$. La coincidencia se da cuando la función de selección es multiplicativa (es decir, $w_{2N}(x, y) = v_{2N}(x)v_{2N}(y)$) y no hay recombinación (es decir, $R_M((x, y), \cdot) = \delta_{(x,y)}(\cdot)$).

3.2.3. Modelo de Moran.

Existen diversas variantes del modelo de Moran. Se presenta a continuación una de las más sencillas.

A diferencia de los modelos anteriores, en este caso las generaciones se traslapan. Las transiciones consisten en la muerte de un individuo y el nacimiento de otro. Sea $M \geq 1$ el número de gametos en la población. El modelo se define como una cadena de Markov con espacio de estados $\mathcal{P}_M(E)$ y función de transición a un paso $P_M(\mu, d\nu)$ en $\mathcal{P}_M(E) \times \mathcal{B}(\mathcal{P}_M(E))$ dada por

$$P_M(\mu, \cdot) = \int_E \int_E \delta_{\mu - M^{-1}\delta_x + M^{-1}\delta_{x'}}(\cdot) \hat{\mu}_M^{***}(dx') \mu(dx), \quad (3.16)$$

donde $\hat{\mu}_M^{***} \in \mathcal{P}(E)$ está definida en términos de $\mu \in \mathcal{P}(E)$ como en el modelo anterior.

La interpretación es clara, dados los mecanismos explicados en los dos modelos anteriores, aplicados a un único individuo de la población, que en cada transición es reemplazado por un único descendiente.

3.3. Caracterización del proceso Fleming-Viot.

Los modelos presentados en la sección anterior se pueden extender, de manera natural, al caso en que E es infinito pero numerable. Sin embargo, los fenómenos poblacionales estudiados pueden ser muy complejos, de forma que aún dichas extensiones no bastan. Existen diversas razones por las cuales se requiere un modelo para poblaciones con espacio de tipos E no numerable. Algunas de ellas son:

- (a) La característica o características que definen el tipo de un individuo pueden ser variables cuantitativas en un espacio continuo, requiriéndose un espacio de tipos no numerable. En este caso, se puede utilizar el llamado *modelo de mutación por pasos con espacio de estados continuo*.
- (b) Supóngase que se desea modelar una población en la que cada individuo del tipo x , independientemente de los demás, muta es decir, es sustituido por un nuevo individuo cada cierto tiempo, de acuerdo con una distribución exponencial con parámetro $\frac{1}{2}\theta$. El tipo del individuo que resulta de la mutación está distribuido de acuerdo a una función de transición a un paso $P(x, d\xi)$. El supuesto básico del modelo es que cada mutante pertenece a un tipo nuevo, por lo que $P(x, \cdot)$ debe ser no atómica para cualquier $x \in E$ y consecuentemente E debe ser no numerable. A un modelo de este tipo se le llama *modelo de mutación neutral para alelos infinitos*.
- (c) Supóngase que se desea estudiar la descendencia de cada individuo en una población. Por ejemplo, se puede utilizar el intervalo $[0, 1]$ para representar el conjunto de posiciones en un cromosoma y se dice que un individuo es del tipo $\mathbf{x} = (x_0, x_1, \dots) \in E = [0, 1]^{\mathbb{Z}^+}$ si x_0, x_1, \dots es la sucesión de posiciones en las que han ocurrido mutaciones en la descendencia del individuo. Incluso si $[0, 1]$ fuera sustituido por un conjunto finito con al menos dos elementos, el conjunto E seguiría siendo no numerable. Para este tipo de análisis se ha utilizado el llamado *modelo de mutación para sitios infinitos sin recombinación*.

La extensión de los modelos aquí presentados al caso en que E es no numerable es un problema complejo. La propuesta de Fleming y Viot [FV79] consiste en definir un proceso de Markov con espacio de estados $\mathcal{P}(E)$, el espacio de medidas de probabilidad sobre $(E, \mathcal{B}(E))$, dotado con la topología de convergencia débil. Dicha definición se hace a través de su generador y su existencia se prueba mostrando la existencia y unicidad de la solución al problema de martingala correspondiente.

En su forma general, el proceso Fleming-Viot se describe en términos de un generador \mathcal{L} sobre $\mathcal{B}(\mathcal{P}(E))$ dado por

$$\begin{aligned}
\mathcal{L}\varphi(\mu) &= \frac{1}{2} \int_E \int_E \mu(dx)(\delta_x(dy) - \mu(dy)) \frac{\partial^2 \varphi(\mu)}{\partial \mu(x) \partial \mu(y)} \\
&+ \int_E \mu(dx) A \left(\frac{\partial \varphi(\mu)}{\partial \mu(\cdot)} \right) (x) \\
&+ \int_E \int_E \mu(dx) \mu(dy) B \left(\frac{\partial \varphi(\mu)}{\partial \mu(\cdot)} \right) (x, y) \\
&+ \int_E \int_E \mu(dx) \mu(dy) (\sigma(x, y) - \langle \sigma, \mu^2 \rangle) \frac{\partial \varphi(\mu)}{\partial \mu(x)}, \quad (3.17)
\end{aligned}$$

donde (E, r) es un espacio métrico compacto, llamado el *espacio de tipos*, $\mathcal{P}(E)$ es el espacio de medidas de probabilidad sobre E dotado con la topología de convergencia débil, A es el generador de un semigrupo de Feller sobre $\mathcal{B}(E)$, llamado el *operador de mutación*, B es un operador sobre $\mathcal{B}(E^2)$ (posiblemente no acotado) llamado *operador de recombinación* con *intensidad de recombinación* $\alpha \geq 0$, definido por

$$Bg(x, y) = \alpha \int_E (g(x') - g(x)) R((x, y), dx'), \quad (3.18)$$

R es una función de transición a un paso sobre $E^2 \times \mathcal{B}(E)$, $\sigma \in \mathcal{B}_s(E^2)$ es una función acotada y simétrica llamada *función de intensidad de selección* y el operador $\frac{\partial \varphi(\mu)}{\partial \mu(\cdot)}$ se define como

$$\frac{\partial \varphi(\mu)}{\partial \mu(x)} = \lim_{\epsilon \rightarrow 0^+} \frac{\varphi(\mu + \epsilon \delta_x) - \varphi(\mu)}{\epsilon}. \quad (3.19)$$

El dominio $\mathcal{D}(\mathcal{L})$ se define como el conjunto de funciones $\varphi \in \mathcal{B}(\mathcal{P}(E))$ de la forma

$$\varphi(\mu) = F(\langle f_1, \mu \rangle, \dots, \langle f_k, \mu \rangle) := F(\langle \mathbf{f}, \mu \rangle), \quad (3.20)$$

donde $k \geq 1$, $f_i \in \mathcal{D}(A)$ para cada $i = 1, \dots, k$ y $F \in \mathcal{C}^2(\mathbb{R}^k)$. Para este tipo de funciones, la expresión para \mathcal{L} se reduce a

$$\begin{aligned}
\mathcal{L}\varphi(\mu) &= \frac{1}{2} \sum_{i,j=1}^k (\langle f_i f_j, \mu \rangle - \langle f_i, \mu \rangle \langle f_j, \mu \rangle) F_{z_i z_j}(\langle \mathbf{f}, \mu \rangle) \\
&+ \sum_{i=1}^k (\langle A f_i, \mu \rangle + \langle B f_i, \mu^2 \rangle) F_{z_i}(\langle \mathbf{f}, \mu \rangle) \\
&+ \sum_{i=1}^k (\langle (f_i \circ \pi), \mu^2 \rangle - \langle f_i, \mu \rangle \langle \sigma, \mu^2 \rangle) F_{z_i}(\langle \mathbf{f}, \mu \rangle). \quad (3.21)
\end{aligned}$$

Teorema 3.3.1. *Sea E localmente compacto y supóngase que la cerradura de A genera un semigrupo de Feller sobre $\mathcal{C}(E)$. Entonces el problema de martingala para \mathcal{L} , con distribución inicial dada, tiene solución única. Si además $B \equiv 0$, entonces el problema de martingala para \mathcal{L} en $\mathcal{C}_{\mathcal{P}(E)}[0, \infty)$ está bien planteado.*

Sea $\{\nu_{\tau}^{(M)}, \tau \in \mathbb{Z}_+\}_{M \geq 1}$ una sucesión de modelos de Wright-Fisher como los descritos en la sección 3.2.2. Bajo ciertas condiciones sobre E y las sucesiones $\{w_M\}$, $\{R_M\}$, $\{Q_M\}$, y asumiendo convergencia débil de las distribuciones iniciales, se puede demostrar ([EK93]) que

$$\{\nu_{[Mt]}^{(M)}, t \geq 0\} \longrightarrow \{\mu_t, t \geq 0\} \quad \text{cuando } M \rightarrow \infty, \quad (3.22)$$

donde $\{\mu_t, t \geq 0\}$ es un proceso Fleming-Viot.

3.4. Modelo de mutación neutral.

Un importante caso particular del proceso Fleming-Viot se obtiene cuando (E, r) es un espacio métrico compacto, el generador \mathcal{L} es de la forma

$$\begin{aligned} (\mathcal{L}\varphi)(\mu) &= \frac{1}{2} \int_E \int_E \mu(dx)(\delta_x(dy) - \mu(dy)) \frac{\delta^2 \varphi(\mu)}{\delta \mu(x) \delta \mu(y)} \\ &+ \int_E \mu(dx) A \left(\frac{\delta \varphi(\mu)}{\delta \mu(\cdot)} \right) (x) \end{aligned} \quad (3.23)$$

y el operador de mutación A está dado por

$$Af(x) = \frac{1}{2} \theta \int_E (f(\xi) - f(x)) P(x, d\xi), \quad (3.24)$$

donde $\theta > 0$ y $P(x, d\xi)$ es una función de transición a un paso sobre $E \times \mathcal{B}(E)$ tal que

$$P(x, \{\xi\}) = 0, \quad x, \xi \in E. \quad (3.25)$$

El proceso Fleming-Viot resultante se conoce como *modelo de mutación neutral para alelos infinitos*. La interpretación corresponde a una población en la cual ocurren mutaciones con una tasa $\frac{1}{2}\theta$. La función $P(x, \cdot)$ es la distribución del tipo de alelo que resulta de la mutación de un individuo de tipo x y la condición (3.25) modela el supuesto básico según el cual cada mutante pertenece a un nuevo tipo, no presente en la población al momento de la mutación.

Aunque algunos de los resultados presentados a continuación son válidos para esta clase de modelos o incluso para versiones más generales del proceso Fleming-Viot, por simplicidad, en adelante se hablará siempre de un caso particular del modelo de mutación neutral para alelos infinitos, dado por

$$Af(x) = \frac{1}{2}\theta \int_E (f(\xi) - f(x))\nu_0(d\xi), \quad (3.26)$$

donde $\theta > 0$ y $\nu_0 \in \mathcal{P}(E)$. En el resto de este trabajo, se usará el término *modelo de mutación neutral* para referirse a este proceso. En este caso, se supone que la distribución del tipo de alelo que resulta de la mutación de un individuo de tipo x no depende de x , de modo que $P(x, \cdot) \equiv \nu_0(\cdot)$ para toda $x \in E$.

Si se supone que E es finito, es posible identificar el espacio de estados en que toma valores el proceso con el espacio

$$\Delta_E := \left\{ (p_i)_{i \in E} \in [0, 1]^E : \sum_{i \in E} p_i = 1 \right\}, \quad (3.27)$$

donde p_i se puede interpretar como la proporción de la población que pertenece al tipo $i \in E$. Si además el soporte de ν_0 es todo E , el generador del proceso Fleming-Viot de mutación neutral es

$$L = \frac{1}{2} \sum_{i, j \in E} p_i (\delta_{ij} - p_i) \frac{\partial^2}{\partial p_i \partial p_j} + \sum_{j \in E} \left(\sum_{i \in E} q_{ij} p_i \right) \frac{\partial}{\partial p_j}, \quad (3.28)$$

donde $Q := (q_{ij})_{i, j \in E}$ es la matriz infinitesimal de transiciones para un proceso de Markov en E y para cada $i, j \in E, i \neq j$,

$$q_{ij} = \frac{1}{2}\theta_j \equiv \frac{1}{2}\theta \nu_0(\{j\}) > 0 \quad (3.29)$$

se interpreta como la intensidad de mutación del tipo i al tipo j . En este caso, el proceso tiene una única distribución estacionaria $\Pi \in \mathcal{P}(\Delta_E)$, que coincide con la distribución Dirichlet con parámetro $(\theta_i)_{i \in E}$, cuya densidad es $(|E| - 1)$ -dimensional y proporcional a $\prod_{i \in E} p_i^{\theta_i - 1}$.

Ethier y Kurtz [EK86] generalizaron este resultado para E no numerable, determinando la distribución estacionaria para el modelo de mutación neutral.

Teorema 3.4.1. *El proceso Fleming-Viot con espacio de tipos E y operador de mutación A definido por la ecuación (3.26), con $\theta > 0$ y $\nu_0 \in \mathcal{P}(E)$, tiene*

una distribución estacionaria única, $\Pi(\cdot | \theta, \nu_0) \in \mathcal{P}(\mathcal{P}(E))$ dada por

$$\Pi(d\mu | \theta, \nu_0) = \mathbb{P} \left[\sum_{i=1}^{\infty} \rho_i \delta_{\xi_i} \in d\mu \right], \quad (3.30)$$

donde $\{\rho_i\}_{i \geq 1}$ tiene la distribución Poisson-Dirichlet con parámetro θ y $\{\xi_i\}_{i \geq 1}$ son iid con distribución ν_0 e independientes de $\{\rho_i\}_{i \geq 1}$.

Esto es equivalente a afirmar que el modelo de mutación neutral tiene una distribución estacionaria única, $\Pi(\cdot | \theta, \nu_0) = \mathcal{D}(\cdot | \theta \nu_0)$ (recuérdese el teorema 2.2.4). Además, el modelo de mutación neutral es un proceso de Markov reversible en el tiempo.

Teorema 3.4.2. *El proceso Fleming-Viot del teorema 3.4.1 es reversible con respecto a $\Pi(\cdot | \theta, \nu_0)$. Es decir, si $\{\mu_t\}_{t \in \mathbb{R}}$ es un proceso Fleming-Viot estacionario de mutación neutral, entonces*

$$\{\mu_t\}_{t \in \mathbb{R}} \stackrel{\mathcal{D}}{=} \{\mu_{-t}\}_{t \in \mathbb{R}} \quad (3.31)$$

Finalmente, existe una forma explícita para la función de transición de este proceso, calculada por Ethier y Griffiths [EG93].

Teorema 3.4.3. *El proceso Fleming-Viot con espacio de estados E y operador de mutación A definido por la ecuación (3.26) tiene función de transiciones $P(t, \mu, d\nu)$ para $t > 0$ y $\mu \in \mathcal{P}(E)$ dada por*

$$P(t, \mu, d\nu) = \sum_{i=1}^n d_n^\theta(t) \int_{E^n} \Pi \left(d\nu | \theta \nu_0 + \sum_{i=1}^n \delta_{X_i} \right) \mu(dX_1) \dots \mu(dX_n), \quad (3.32)$$

donde $d_n^\theta(t) = \mathbb{P}[D_t = n]$, $n \geq 0$ son las probabilidades correspondientes al proceso de muerte puro presentado en la sección 1.2.4.

En particular, este teorema muestra que, para $t > 0$ y $\mu \in \mathcal{P}(E)$, $P(t, \mu, \cdot)$ es una mezcla de procesos Dirichlet.

Capítulo 4

Fleming-Viot vía urnas.

En el capítulo anterior se presentó la definición del proceso Fleming-Viot a partir de su generador. Aún el modelo de mutación neutral, su forma más sencilla, es un proceso complejo que toma valores en un espacio muy grande. El estudio de sus propiedades resulta complicado y los resultados de Ethier y Kurtz [EK93] para la distribución estacionaria y de Ethier y Griffiths [EG93] para la función de transición, se basan en un extenso conocimiento de la teoría de procesos de Markov y el desarrollo de resultados intermedios. Las propiedades adicionales necesarias para realizar inferencia estadística a partir de este modelo, resultan igualmente difíciles de analizar.

Walker, Hatjispyros y Nicolieris [WHN07] proponen una construcción alternativa para el proceso Fleming-Viot de mutación neutral, basada en el método presentado en la sección 1.4.2 para la definición de procesos de Markov a tiempo continuo a través de procesos latentes. Dicha construcción será desarrollada en el presente capítulo.

Se presenta primero la base para la construcción de un proceso a tiempo discreto $\{\mu_k\}_{k \geq 0}$ con valores en $\mathcal{P}(E)$, el espacio de medidas de probabilidad sobre un espacio métrico, completo y separable, E . Para ello se utiliza un proceso latente $\{\mathbb{X}_k^n\}_{k \geq 0}$ con valores en E^n , donde \mathbb{X}_k^n puede interpretarse como una muestra de tamaño n de la distribución μ_k . Esta muestra se utiliza para definir la transición de μ_{k-1} a μ_k . Posteriormente, para construir el proceso a tiempo continuo $\{\mu_t\}_{t \geq 0}$, se aleatoriza el tamaño de muestra η_t utilizado para la transición de μ_s a μ_{s+t} , introduciendo la dependencia del tiempo a través de una densidad $p_t(n) := \mathbb{P}[\eta_t = n]$. Finalmente, se establece las condiciones para que el proceso $\{\mu_t\}_{t \geq 0}$ construido satisfaga las ecuaciones de Chapman-Kolmogorov. Gracias a la representación del proceso Dirichlet en términos de un modelo de urnas, se muestra que para ello

basta fijar una condición sobre la densidad $p_t(n)$. Se demuestra después que, en el caso particular en que $p_t(n) = d_n^\theta(t)$ (las probabilidades asociadas al proceso de muerte definido en la sección 1.2.4), se satisface dicha condición y se obtiene el proceso Fleming-Viot de mutación neutral.

4.1. Proceso de partículas intercambiables.

Considérese el siguiente modelo no paramétrico, tal como se describe en la sección 1.3:

$$X_k | \mu \stackrel{iid}{\sim} \mu, \quad k \geq 1 \quad (4.1)$$

$$\mu \sim \mathcal{D}(\cdot | \theta \nu_0), \quad (4.2)$$

donde $\{X_k\}_{k \geq 1}$ es un proceso estocástico a tiempo discreto con valores en un espacio E métrico, completo y separable; $\nu_0 \in \mathcal{P}(E)$ es no atómica, $\theta > 0$ y $\mathcal{D}(\cdot | \theta \nu_0)$ denota al proceso Dirichlet, definido en la sección 2.2. Sea $\mathbb{X}^n := (X_1, \dots, X_n)$, con $n \geq 0$ fija. El lema 2.2.8 establece que la distribución condicional de μ dado \mathbb{X}^n es de la forma

$$\mathbb{P}[\mu \in \cdot | \mathbb{X}^n] = \mathcal{D}(\cdot | \theta \nu_0 + \sum_{i=1}^n \delta_{X_i}). \quad (4.3)$$

Las X_i 's son condicionalmente independientes e idénticamente distribuidas dada $\mu \in \mathcal{P}(E)$. Por lo tanto, existe una distribución conjunta para (\mathbb{X}^n, μ) , definida sobre $E \times \mathcal{P}(E)$ como,

$$\mathbb{P}[d\mathbb{X}^n, d\mu] = \mathbb{P}[dX_1, \dots, dX_n, d\mu] = \mathcal{D}(d\mu | \theta \nu_0) \prod_{i=1}^n \mu(dX_i). \quad (4.4)$$

Con base en lo anterior, se puede aplicar el método presentado en la sección 1.4 para la construcción de un proceso de Markov a tiempo discreto, $\{\mu_k\}_{k \geq 0}$ con valores en el espacio de medidas $\mathcal{P}(E)$, a través de un proceso latente $\{\mathbb{X}_k^n\}_{k \geq 0}$ con valores en E^n . Para ello, se parte de las distribuciones condicionales

$$\mu | \mathbb{X}^n \sim \mathcal{D}(d\mu | \theta \nu_0 + \sum_{i=1}^n \delta_{X_i}) =: \Pi(d\mu | \mathbb{X}^n, n), \quad (4.5)$$

$$\mathbb{X}^n | \mu \sim \prod_{i=1}^n \mu(X_i) =: \mathbb{P}(\mathbb{X}^n | n, \mu), \quad (4.6)$$

donde la sustitución de $d\mathbb{X}^n$ por \mathbb{X}^n se justifica porque μ , al provenir de un proceso Dirichlet, es discreta casi seguramente (lema 2.2.7).

Se define así, para cada medida inicial $\mu_0 \in \mathcal{P}(E)$, una cadena de Markov reversible $\{\mu_k\}_{k \geq 0}$ con función de transición a un paso,

$$\begin{aligned} P(\mu, B) &= \int_{E^n} \Pi(B|\mathbb{X}^n, n) \mathbb{P}(\mathbb{X}^n|n, \mu) \\ &= \int_{E^n} \mathcal{D}(B|\theta\nu_0 + \sum_{i=1}^n \delta_{X_i}) \mu(X_1) \dots \mu(X_n), \end{aligned} \quad (4.7)$$

definida para $\mu \in \mathcal{P}(E)$ y $B \in \mathcal{B}(\mathcal{P}(E))$. Aquí, $\mathcal{B}(\mathcal{P}(E))$ se refiere a los borelianos de $\mathcal{P}(E)$, con la topología de convergencia débil (sección 1.1.1). Además, la cadena es estacionaria desde su inicio, con distribución estacionaria,

$$\Pi(d\mu) = \int_{E^n} \mathbb{P}(\mathbb{X}^n, d\mu) = \mathcal{D}(d\mu|\theta\nu_0) \prod_{i=1}^n \left[\int_E \mu(dX_i) \right] = \mathcal{D}(d\mu|\theta\nu_0), \quad (4.8)$$

es decir, la distribución estacionaria coincide con la marginal para μ , correspondiente a la distribución conjunta (4.4).

Siguiendo el razonamiento de Mena y Walker [MW07], presentado en la sección 1.4.2, se observa que la relación entre μ_k y μ_{k-1} depende del valor de n . Por ejemplo, para valores de n pequeños, es probable que μ_1 sea cercana a ν_0 . Conforme n crece, el parámetro del proceso Dirichlet acumula masa en \mathbb{X}_0^n o más específicamente, en la distribución empírica correspondiente a una muestra de tamaño n de μ_0 , de modo que μ_1 resulta con mayor probabilidad, cercana a μ_0 . Por lo tanto, para llevar el proceso a tiempo continuo, resulta natural imponer una dependencia con respecto al tiempo sobre n .

Para aclarar esta idea, considérese que se desea ampliar el modelo definido por las distribuciones condicionales (4.5) y (4.6) para construir un proceso a tiempo continuo $\{\mu_t\}_{t \geq 0}$. La ampliación se dará a través del parámetro n , el número de observaciones independientes e idénticamente distribuidas de la medida μ que determinan el parámetro del proceso Dirichlet en la función de transición a un paso (4.7). Para el proceso a tiempo continuo, la función de transición $P(t, \mu, B)$ (que se dará en forma explícita más adelante) dependerá también de un tamaño de muestra, esta vez variable y dependiente de t , que se denotará por η_t . Para dar mayor flexibilidad al modelo, en lugar de imponer una relación funcional determinista del tipo $\eta_t = \phi(t)$, se introduce una relación a nivel distribucional.

Denótese por $p_t(n) := \mathbb{P}[\eta_t = n]$ a la probabilidad de utilizar n observaciones para definir una transición en un periodo de tiempo $t \geq 0$ (la transición del tiempo s al tiempo $t + s$ para cualquier $s \geq 0$), con $\sum_{n=1}^{\infty} p_t(n) = 1$ para todo $t \geq 0$. Entonces, se puede definir un proceso a tiempo continuo a partir de las condicionales,

$$\mu|\mathbb{X}^n, n \sim \mathcal{D}(d\mu|\theta\nu_0 + \sum_{i=1}^n \delta_{X_i}) =: \Pi(d\mu|\mathbb{X}^n, n) \quad (4.9)$$

$$\mathbb{X}^n|\mu, n \sim \prod_{i=1}^n \mu(X_i) =: \mathbb{P}(\mathbb{X}^n|n, \mu) \quad (4.10)$$

$$\eta_t|\mu, \mathbb{X}^n \sim p_t(\cdot), \quad (4.11)$$

donde la notación pone en evidencia que la dependencia con respecto al tiempo se establece únicamente a través de η_t , por lo que el proceso construido será homogéneo en el tiempo.

Más específicamente, se puede aplicar la idea sugerida por el método de muestreo de Gibbs, como en la sección 1.4, para construir un proceso en tiempo continuo $\{\mu_t\}_{t \geq 0}$ de la siguiente forma:

0. Inicialización.

Se parte de una medida inicial no atómica, $\mu_0 \in \mathcal{P}(E)$.

1. Transición.

La transición de μ_s a μ_{t+s} para cada $t, s \geq 0$ se lleva a cabo en dos pasos:

(a) Se obtiene $\eta_t = \eta_{(t+s)-s}$ con distribución

$$\mathbb{P}[\eta_t = n] = p_t(n). \quad (4.12)$$

(b) Se genera $\mathbb{X}^{\eta_t} := \mathbb{X}^{\eta_t}(t + s)$ con distribución condicional

$$\mathbb{P}(\mathbb{X}^{\eta_t}|\eta_t, \mu_s) = \prod_{i=1}^{\eta_t} \mu_s(X_i). \quad (4.13)$$

Finalmente, μ_{t+s} se genera a partir de la distribución condicional

$$\Pi(d\mu_{t+s}|\mathbb{X}^{\eta_t}, \eta_t) = \mathcal{D}(d\mu_{t+s}|\theta\nu_0 + \sum_{i=1}^{\eta_t} \delta_{X_i}). \quad (4.14)$$

De este modo se construye simultáneamente dos procesos: el proceso de medidas $\{\mu_t\}_{t \geq 0}$ con valores en $\mathcal{P}(E)$ y un proceso latente de observaciones $\{\mathbb{X}^{M_t}(t)\}_{t \geq 0}$, con valores en $\{E^n : n = 0, 1, 2, \dots\}$, donde M_t denota el número (aleatorio) de elementos en la muestra al tiempo t . (Nótese que $\{M_t\}$ es, a su vez, un proceso de Markov con espacio de estados $\{0, 1, 2, \dots\}$).

La función de transición para el proceso $\{\mu_t\}_{t \geq 0}$ es

$$\begin{aligned} P(t, \mu, B) &= \sum_{n=0}^{\infty} p_t(n) \int_{E^n} \Pi(B|\mathbb{X}^n, n) \mathbb{P}(d\mathbb{X}^n | n, \mu) \\ &= \sum_{n=0}^{\infty} p_t(n) \int_{E^n} \mathcal{D}(B|\theta\nu_0 + \sum_{i=1}^n \delta_{X_i}) \mu(dX_1) \dots \mu(dX_n), \quad (4.15) \end{aligned}$$

para $\mu \in \mathcal{P}(E)$ y $B \in \mathcal{B}(\mathcal{P}(E))$.

Si E es un espacio métrico completo y separable, entonces $\mathcal{P}(E)$, con la topología de convergencia débil también lo es (sección 1.1.1). De modo que para garantizar la existencia de un proceso de Markov homogéneo en el tiempo $\{\mu_t\}_{t \geq 0}$, con función de transición (4.15), basta verificar las ecuaciones de Chapman-Kolmogorov. Para hacer esto de manera directa, sería necesario definir una medida sobre $\mathcal{P}(E)$ (equivalente a la medida de Lebesgue en el caso en que el proceso toma valores en \mathbb{R}^d , como se indicó en la sección 1.2.2) e integrar con respecto a ella. Este es un problema complejo que se puede simplificar gracias a la estructura de construcción a través del proceso latente de observaciones $\{\mathbb{X}^{M_t}(t)\}_{t \geq 0}$. La clave, como se verá en la siguiente sección, es que la dependencia con respecto al tiempo en la función de transición $P(t+s, \mu_0, d\mu_{t+s})$ se establece únicamente a través del número de observaciones iid de μ_0 involucradas en la distribución condicional para μ_{t+s} , en la igualdad (4.14).

4.2. Ecuaciones de Chapman-Kolmogorov.

Sea δ_{μ_0} la medida de probabilidad sobre $(\mathcal{P}(E), \mathcal{B}(\mathcal{P}(E)))$ que acumula toda su masa sobre una medida $\mu_0 \in \mathcal{P}(E)$. Recordemos de la sección 1.2.2 que, si la expresión (4.15) define una función de transición homogénea en el tiempo, entonces δ_{μ_0} y $P(t, \mu, d\nu)$ definen un proceso de Markov $\{\mu_t\}_{t \geq 0}$ con valores en $\mathcal{P}(E)$. Debido a la definición de $P(t, \mu, B)$, en términos de la distribución conjunta de $(\mu_t, \mathbb{X}^{\eta_t}, \eta_t)$, para garantizar que es una función de transición homogénea en el tiempo, basta verificar las ecuaciones de

Chapman-Kolmogorov,

$$P(t+s, \mu_0, B) = \int_{\mathcal{P}(E)} P(t, \mu_s, B) P(s, \mu_0, d\mu_s) \quad (4.16)$$

para cualesquiera $B \in \mathcal{B}(\mathcal{P}(E))$ y $t, s \geq 0$. Por lo tanto, con el fin de establecer las condiciones para la existencia del proceso $\{\mu_t\}_{t \geq 0}$, se analizará estas ecuaciones a través del proceso latente.

Partiendo directamente de la definición (4.15), el lado izquierdo de la igualdad (4.16) se puede escribir como

$$\begin{aligned} \sum_{r=0}^{\infty} \mathbb{P}[\eta_{t+s} = r] \int_{E^r} \Pi(B|\mathbb{X}^r, r) \mathbb{P}(d\mathbb{X}^r|r, \mu_0) = \\ \sum_{r=0}^{\infty} p_{t+s}(r) \int_{E^r} \mathcal{D}(B|\theta\nu_0 + \sum_{j=1}^r \delta_{X_j}) \mu_0(dX_1) \dots \mu_0(dX_r). \end{aligned} \quad (4.17)$$

Haciendo la misma sustitución para las funciones de transición, el lado derecho se puede escribir como

$$\begin{aligned} \int_{\mathcal{P}(E)} \left[\sum_{m=0}^{\infty} p_t(m) \int_{E^m} \Pi(B|\mathbb{Y}^m, m) \mathbb{P}(d\mathbb{Y}^m|m, \mu_s) \right] \\ \times \left[\sum_{n=0}^{\infty} p_s(n) \int_{E^n} \Pi(d\mu_s|\mathbb{X}^n, n) \mathbb{P}(d\mathbb{X}^n|n, \mu_0) \right], \end{aligned} \quad (4.18)$$

donde $\mathbb{Y}^m := (Y_1, \dots, Y_m)$ tiene un papel análogo al de \mathbb{X}^n . Reordenando términos, esto es igual a

$$\begin{aligned} \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} p_t(m) p_s(n) \int_{E^m} \int_{E^n} \Pi(B|\mathbb{Y}^m, m) \mathbb{P}(d\mathbb{X}^n|n, \mu_0) \\ \times \left[\int_{\mathcal{P}(E)} \mathbb{P}(d\mathbb{Y}^m|m, \mu_s) \Pi(d\mu_s|\mathbb{X}^n, n) \right]. \end{aligned} \quad (4.19)$$

Por el lema 2.2.9, la última integral es igual a

$$\mathbb{P}(d\mathbb{Y}^m|m, \mathbb{X}^n, n) = \prod_{j=1}^m \left(\frac{\theta\nu_0 + \sum_{i=1}^n \delta_{X_i} + \sum_{k=1}^{j-1} \delta_{Y_k}}{\theta + n + j - 1} \right). \quad (4.20)$$

Por lo tanto el lado derecho de (4.16) se puede reescribir como

$$\sum_{m=0}^{\infty} \sum_{n=0}^{\infty} p_t(m) p_s(n) \int_{E^m} \int_{E^n} \Pi(B|\mathbb{Y}^m, m) \mathbb{P}(d\mathbb{Y}^m|m, \mathbb{X}^n, n) \mathbb{P}(d\mathbb{X}^n|n, \mu_0). \quad (4.21)$$

Para proseguir a partir de este punto, es conveniente analizar la interpretación de las últimas dos expresiones e introducir algo de notación. Supóngase que, partiendo de μ_0 y para $\eta_s = M_s = n$, se genera una muestra $\mathbb{X}^n := \mathbb{X}^n(s)$, con $X_i \stackrel{iid}{\sim} \mu_0$. Después, para $\eta_t = M_{t+s} = m$, se genera el vector de observaciones $\mathbb{Y}^m = \mathbb{X}^m(t+s)$, no como observaciones iid con distribución μ_s , sino directamente de la urna de Pólya generalizada:

$$Y_1 \sim \frac{\theta\nu + \sum_{i=1}^n \delta_{X_i}}{\theta + n} \quad (4.22)$$

y para $j = 2, \dots, m$,

$$Y_j \sim \frac{\theta\nu + \sum_{i=1}^n \delta_{X_i} + \sum_{i=1}^{j-1} \delta_{Y_i}}{\theta + n + j - 1}. \quad (4.23)$$

Siguiendo la notación de la sección 2.1, denótese por (Y_1^*, \dots, Y_k^*) a los valores únicos de \mathbb{Y}^m . Entonces, para cada $j = 1, \dots, k$, se tiene que $Y_j^* \in \{X_1, \dots, X_n\}$ o $Y_j^* \sim \nu_0$. Por lo tanto, habrá un número r de observaciones Y_j^* que coinciden con alguna X_i , para alguna $0 \leq r \leq \min\{n, m\}$.

Denótese en general por R_{t+s} al número de observaciones únicas de $\mathbb{X}^{M_{t+s}}(t+s)$ que coinciden con alguna observación de $\mathbb{X}^{M_s}(s)$. Entonces, la distribución (4.20) se puede escribir, condicionando sobre R_{t+s} , como

$$\mathbb{P}(d\mathbb{Y}^m|m, \mathbb{X}^n, n) = \sum_{r=0}^{\infty} \mathbb{P}(d\mathbb{Y}^m|m, \mathbb{X}^n, n, r) \mathbb{P}[R_{t+s} = r|n, m]. \quad (4.24)$$

Sustituyendo en la expresión (4.21) e invirtiendo el orden de las sumas, se obtiene

$$\begin{aligned} & \sum_{r=0}^{\infty} \sum_{m=r}^{\infty} \sum_{n=r}^{\infty} p_t(m) p_s(n) \mathbb{P}[R_{t+s} = r|n, m] \\ & \quad \times \int_{E^m} \Pi(B|\mathbb{Y}^m, m) \int_{E^n} \mathbb{P}(d\mathbb{Y}^m|m, \mathbb{X}^n, n, r) \mathbb{P}(d\mathbb{X}^n|n, \mu_0). \end{aligned} \quad (4.25)$$

Aplicando el resultado del lema 2.2.12, para resolver la integral interna, la integral externa se convierte en

$$\int_{E^m} \mathcal{D}(B|\theta\nu_0 + \sum_{j=1}^m \delta_{Y_j}) \mu_0(dY_1) \dots \mu_0(dY_r) \nu(dY_{r+1}) \dots \nu(dY_m), \quad (4.26)$$

donde $\nu \sim \mathcal{D}(\cdot | \theta\nu_0 + \sum_{j=1}^r \delta_{Y_j})$. Por el lema 2.2.8,

$$\begin{aligned} \int_{E^{m-r}} \mathcal{D}(B | \theta\nu_0 + \sum_{j=1}^r \delta_{Y_j} + \sum_{j=1}^{m-r} \delta_{Y_j}) \nu(dY_{r+1}) \dots, \nu(dY_m) \\ = \mathcal{D}(B | \theta\nu_0 + \sum_{j=1}^r \delta_{Y_j}). \end{aligned} \quad (4.27)$$

Así, la expresión (4.25) es igual a

$$\sum_{r=0}^{\infty} \sum_{m=r}^{\infty} \sum_{n=r}^{\infty} p_t(m) p_s(n) \mathbb{P}[R_{t+s} = r | n, m] \int_{E^r} \Pi(B | \mathbb{Y}^r, r) \mathbb{P}(d\mathbb{Y}^r | r, \mu_0). \quad (4.28)$$

Comparando con la ecuación (4.17), se observa que una condición suficiente para que las ecuaciones de Chapman-Kolmogorov para el proceso $\{\mu_t\}_{t \geq 0}$ se satisfagan es:

$$p_{t+s}(r) = \sum_{n=r}^{\infty} \sum_{m=r}^{\infty} \mathbb{P}[R_{t+s} = r | \eta_s = n, \eta_t = m] p_t(m) p_s(n). \quad (4.29)$$

Finalmente, sustituyendo el resultado del teorema 2.2.10, se concluye que las ecuaciones de Chapman-Kolmogorov para el proceso $\{\mu_t\}_{t \geq 0}$ se satisfacen si para toda $t, s > 0$ y $r \in \{0, 1, 2, \dots\}$,

$$p_{t+s}(r) = \sum_{n=r}^{\infty} \sum_{m=r}^{\infty} r! \binom{n}{r} \binom{m}{r} \frac{(\theta)_{n \uparrow} (\theta)_{m \uparrow}}{(\theta)_{n+m \uparrow} (\theta)_{r \uparrow}} p_s(m) p_t(n). \quad (4.30)$$

4.3. Modelo de mutación neutral.

Recuérdese el proceso de muerte puro $\{D_t\}_{t \geq 0}$ presentado en la sección 1.2.4. En el capítulo 3 se mencionó la relación entre este proceso y la función de transición de un caso particular del proceso Fleming-Viot, llamado modelo de mutación neutral. Dicha relación se establece a través de las probabilidades $d_n^\theta(t) := \mathbb{P}[D_t = n]$, dadas por

$$d_n^\theta(t) = \begin{cases} 1 - \sum_{m=1}^{\infty} (-1)^{m-1} (\theta)_{m-1 \uparrow} m!^{-1} \gamma_{m,t,\theta} & \text{si } n = 0, \\ \sum_{m=n}^{\infty} (-1)^{m-n} \binom{m}{n} (\theta+n)_{m-1 \uparrow} m!^{-1} \gamma_{m,t,\theta} & \text{si } n > 0, \end{cases}$$

donde

$$\gamma_{m,t,\theta} = (2m-1+\theta)e^{-\lambda_m t}$$

y $\lambda_m = \frac{1}{2}m(m + \theta - 1)$, $m \geq 0$ son las tasas de muerte del proceso $\{D_t\}_{t \geq 0}$.

Se probará que $\{p_t(n) = d_n^\theta(t) : n \in \{0, 1, 2, \dots\}, t \geq 0\}$ es una solución para la ecuación (4.30). De ello se sigue que, haciendo $p_t(n) = d_n^\theta(t)$ en la expresión (4.11), se construye un proceso de Markov a tiempo continuo $\{\mu_t\}_{t \geq 0}$ con función de transición

$$P(t, \mu, B) = \sum_{n=0}^{\infty} d_n^\theta(t) \int_{E^n} \mathcal{D}(B | \theta \nu_0 + \sum_{i=1}^n \delta_{X_i}) \mu(dX_1) \dots \mu(dX_n). \quad (4.31)$$

Esta es exactamente igual a la función de transición (3.32), de modo que en este caso, el proceso $\{\mu_t\}_{t \geq 0}$ construido coincide con el proceso Fleming-Viot de mutación neutral.

Antes de proceder a la demostración, es conveniente presentar algunos resultados.

Lema 4.3.1. *Para toda $\phi > 0$, $k \geq 1$ y $r \in \{1, \dots, k\}$ se tiene que*

$$\sum_{l=0}^k (-1)^{k-l} \binom{k}{l} (\phi + l)_{k-r \uparrow} = 0 \quad (4.32)$$

Demostración. Se probará el resultado por inducción. Por claridad, se usará la notación

$$f(k, r) := \sum_{l=0}^k (-1)^{k-l} \binom{k}{l} (\phi + l)_{k-r \uparrow},$$

para $k \geq 1$ y $r \in \{1, \dots, k\}$. Entonces, $f(2, 1) = \phi - 2(\phi + 1) + (\phi + 2) = 0$. Además, como $(a)_{0 \uparrow} = 1$ para cualquier $a \in \mathbb{R}$, usando el desarrollo del binomio de Newton, se tiene que para cualquier $k \geq 1$

$$f(k, k) = \sum_{l=0}^k (-1)^{k-1} \binom{k}{l} = (-1 + 1)^k = 0.$$

Ahora, supóngase que $f(k, r) = 0$ para cualesquiera $r \in \{1, \dots, k\}$ y

$1 \leq k < K$ y que $f(K, r) = 0$ para $r \in \{R, \dots, K\}$ con $R > 1$. Entonces

$$\begin{aligned}
f(K, R-1) &= \sum_{l=0}^K (-1)^{K-l} \binom{K}{l} (\phi + l)_{K-R+1\uparrow} = \\
&= \sum_{l=0}^K (-1)^{K-l} \binom{K}{l} (\phi + l)(\phi + l + 1)_{K-R\uparrow} = \\
&= \phi \sum_{l=0}^K (-1)^{K-l} \binom{K}{l} (\phi + l + 1)_{K-R\uparrow} \\
&+ K \sum_{l=1}^K (-1)^{K-l} \binom{K-1}{K-l} (\phi + l + 1)_{K-R\uparrow} = \\
\phi f(K, R) + K \sum_{l=1}^K (-1)^{K-l} \binom{K-1}{l-1} (\phi + l + 1)_{K-1-R+1\uparrow} &= \\
\phi f(K, R) + K f(K-1, R-1) &= 0.
\end{aligned}$$

□

Lema 4.3.2. Para cualesquiera $s > 0$, $\theta > 0$ y $n \in \{0, 1, 2, \dots\}$,

$$\sum_{m=n}^{\infty} \frac{(m)_{n\downarrow}}{(\theta + m)_{n\uparrow}} d_m^\theta(s) = e^{-\lambda_n s} \quad (4.33)$$

Demostración. Obsérvese primero que

$$\frac{(m)_{0\downarrow}}{(\theta + m)_{0\uparrow}} d_m^\theta(s) = e^{-\lambda_n s} = 1.$$

Como $\lambda_0 = 0$,

$$\sum_{m=0}^{\infty} d_0^\theta(s) = 1 = e^{-\lambda_0 s}.$$

Ahora, para $n \geq 1$, sustituyendo el valor de $d_m^\theta(s)$ en el lado izquierdo de la igualdad (4.33) se obtiene

$$\sum_{m=n}^{\infty} \frac{(m)_{n\downarrow}}{(\theta + m)_{n\uparrow}} \sum_{k=m}^{\infty} (-1)^{k-m} \binom{k}{m} (\theta + m)_{k-1\uparrow} k!^{-1} \gamma_{k,s,\theta},$$

que intercambiando el orden de las sumas, es igual a

$$\begin{aligned} \sum_{k=n}^{\infty} \sum_{m=n}^k \frac{(m)_{n\downarrow}}{(\theta+m)_{n\uparrow}} (-1)^{k-m} \binom{k}{m} (\theta+m)_{k-1\uparrow} k!^{-1} \gamma_{k,s,\theta} = \\ \sum_{k=n+1}^{\infty} \sum_{m=n}^k \frac{(m)_{n\downarrow}}{(\theta+m)_{n\uparrow}} (-1)^{k-m} \binom{k}{m} (\theta+m)_{k-1\uparrow} k!^{-1} \gamma_{k,s,\theta} \\ + \frac{(n)_{n\downarrow} \gamma_{n,s,\theta}}{n!(\theta+2n-1)}. \end{aligned}$$

Recuérdese que $\gamma_{n,s,\theta} = (2n-1+\theta)e^{\lambda_n t}$, de modo que

$$\frac{(n)_{n\downarrow} \gamma_{n,s,\theta}}{n!(\theta+2n-1)} = e^{-\lambda_n s}.$$

Observando que para toda $k > n$

$$\begin{aligned} (m)_{n\downarrow} k!^{-1} \binom{k}{m} &= \frac{1}{(k-n)!} \binom{k-n}{m-n}, \\ \frac{(\theta+m)_{k-1\uparrow}}{(\theta+m)_{n\uparrow}} &= (\theta+m+n)_{k-n-1\uparrow}, \end{aligned}$$

el resto de la suma se puede escribir como

$$\begin{aligned} \sum_{k=n+1}^{\infty} \frac{\gamma_{k,s,\theta}}{(k-n)!} \sum_{m=n}^k (-1)^{k-m} \binom{k-n}{m-n} (\theta+m+n)_{k-n-1\uparrow} = \\ \sum_{k=n+1}^{\infty} \frac{\gamma_{k,s,\theta}}{(k-n)!} \sum_{l=0}^{k-n} (-1)^{k-l-n} \binom{k-n}{l} (\theta+l+2n)_{k-n-1\uparrow} = \\ \sum_{k=n+1}^{\infty} \frac{\gamma_{k,s,\theta}}{(k-n)!} \sum_{l=0}^k (-1)^{k-l} \binom{k}{l} (\theta+2n+l)_{k-1\uparrow} = 0, \end{aligned}$$

donde la última igualdad se obtiene aplicando el lema 4.3.1. \square

Lema 4.3.3. Para cualesquiera $s > 0$, $\theta > 0$ y $n \in \{1, 2, \dots\}$,

$$\sum_{m=n-1}^{\infty} \frac{(m)_{n-1\downarrow}}{(\theta+m)_{n\uparrow}} d_m^\theta(s) = \frac{1}{2} \frac{1}{\lambda_n - \lambda_{n-1}} (e^{-\lambda_{n-1}s} - e^{\lambda_n s}). \quad (4.34)$$

Demostración. Obsérvese que, para $n = 1$, el lado derecho de esta igualdad es

$$\frac{1}{2} \frac{1}{\lambda_1 - \lambda_0} (e^{-\lambda_0 s} - e^{\lambda_1 s}) = \frac{1}{\theta} (1 - e^{\lambda_1 s}),$$

y el lado izquierdo,

$$\begin{aligned} \sum_{m=0}^{\infty} \frac{1}{\theta + m} d_m^\theta(s) &= \frac{1}{\theta} d_0^\theta(s) + \sum_{m=1}^{\infty} \frac{1}{\theta + m} d_m^\theta(s) = \\ &= \frac{1}{\theta} \left(1 - \sum_{m=1}^{\infty} d_m^\theta(s) \right) + \sum_{m=1}^{\infty} \frac{1}{\theta + m} d_m^\theta(s) = \frac{1}{\theta} \left(1 - \sum_{m=1}^{\infty} \frac{m}{\theta + m} d_m^\theta(s) \right). \end{aligned}$$

Además, por el lema 4.3.2,

$$\sum_{m=1}^{\infty} \frac{m}{\theta + m} d_m^\theta(s) = \sum_{m=1}^{\infty} \frac{(m)_{1\downarrow}}{(\theta + m)_{1\uparrow}} d_m^\theta(s) = e^{-\lambda_1 s}$$

Ahora, para $n \geq 2$, sustituyendo el valor de $d_m^\theta(s)$ en el lado izquierdo de la igualdad (4.34) que se desea demostrar, se obtiene

$$\begin{aligned} \sum_{m=n-1}^{\infty} \frac{(m)_{n-1\downarrow}}{(\theta + m)_{n\uparrow}} \sum_{k=m}^{\infty} (-1)^{k-m} \binom{k}{m} (\theta + m)_{k-1\uparrow} k!^{-1} \gamma_{k,s,\theta} &= \\ \sum_{k=n-1}^{\infty} k!^{-1} \gamma_{k,s,\theta} \sum_{m=n-1}^k (-1)^{k-m} \binom{k}{m} (m)_{n-1\downarrow} \frac{(\theta + m)_{k-1\uparrow}}{(\theta + m)_{n\uparrow}} &= \\ \frac{\gamma_{n-1,s,\theta}}{(n-1)!} (n-1)_{n-1\downarrow} \frac{(\theta + n-1)_{n-2\uparrow}}{(\theta + n-1)_{n\uparrow}} &+ \\ + \frac{\gamma_{n,s,\theta}}{n!} \left[(n)_{n-1\downarrow} \frac{(\theta + n)_{n-1\uparrow}}{(\theta + n)_{n\uparrow}} - n (n-1)_{n-1\downarrow} \frac{(\theta + n-1)_{n-1\uparrow}}{(\theta + n-1)_{n\uparrow}} \right] &+ \\ + \sum_{k=n+1}^{\infty} k!^{-1} \gamma_{k,s,\theta} \sum_{m=n-1}^k (-1)^{k-m} \binom{k}{m} (m)_{n-1\downarrow} \frac{(\theta + m)_{k-1\uparrow}}{(\theta + m)_{n\uparrow}}. & \end{aligned}$$

Observando que $2\lambda_n - \lambda_{n-1} = [n(n-1+\theta) - (n-1)(n+\theta)] = 2n + \theta - 2$, el primer sumando es igual a

$$\frac{(2n + \theta - 3)e^{\lambda_{n-1}s} (n-1)!}{(n-1)! (\theta + 2n - 3) (\theta + 2n - 2)} = \frac{e^{\lambda_{n-1}s}}{2(\lambda_n - \lambda_{n-1})},$$

y el segundo sumando igual a

$$\frac{(2n + \theta - 1)e^{\lambda_n s}}{n!} \left[\frac{n!}{(\theta + 2n - 1)} - \frac{n!}{(\theta + 2n - 2)} \right] = -\frac{e^{\lambda_n s}}{2(\lambda_n - \lambda_{n-1})}.$$

Finalmente, aplicando el lema 4.3.1, el resto de la suma se anula, pues

$$\begin{aligned}
 & \sum_{m=n-1}^k (-1)^{k-m} \binom{k}{m} (m)_{n-1\downarrow} \frac{(\theta+m)_{k-1\uparrow}}{(\theta+m)_{n\uparrow}} = \\
 & \sum_{l=0}^{k-n+1} (-1)^{k-l-n+1} \binom{k}{l+n-1} (l+n-1)_{n-1\downarrow} \frac{(\theta+l-n+1)_{k-1\uparrow}}{(\theta+l-n+1)_{n\uparrow}} = \\
 & \sum_{l=0}^k (-1)^{k-l} \binom{k+n-1}{l+n-1} (l+n-1)_{n-1\downarrow} (\theta+l+1)_{k-2\uparrow} = \\
 & \frac{(k+n-1)!}{k!} \sum_{l=0}^k (-1)^{k-l} \binom{k}{l} (\theta+1+l)_{k-2\uparrow} = 0.
 \end{aligned}$$

□

Recuérdese ahora la ecuación (4.29):

$$p_{t+s}(r) = \sum_{n=r}^{\infty} \sum_{m=r}^{\infty} \mathbb{P}[R_{t+s} = r | \eta_s = n, \eta_t = m] p_t(m) p_s(n).$$

En ella, R_{t+s} representa el número de observaciones independientes e idénticamente distribuidas de μ_0 (en el proceso de partículas) y η_s representa el mismo tamaño de muestra pero al tiempo s . Por lo tanto, se puede reinterpretar la ecuación en términos del proceso $\{R_t\}_{t \geq 0}$ como

$$\mathbb{P}[R_{t+s} = r] = \sum_{n=r}^{\infty} \mathbb{P}[R_{t+s} = r | R_s = n] \mathbb{P}[R_s = n]. \quad (4.35)$$

Equivalentemente, se puede partir de la ecuación (4.30) y observar que la distribución condicional está dada por

$$\mathbb{P}[R_{t+s} = r | R_s = n] = \sum_{m=r}^{\infty} r! \binom{n}{r} \binom{m}{r} \frac{(\theta)_{n\uparrow} (\theta)_{m\uparrow}}{(\theta)_{n+m\uparrow} (\theta)_{r\uparrow}} \mathbb{P}[R_t = m]. \quad (4.36)$$

Haciendo $\mathbb{P}[R_t = m] = d_m^\theta(t)$ y aplicando el lema 4.3.2,

$$\begin{aligned}
 \mathbb{P}[R_{t+s} = n | R_s = n] &= \sum_{m=n}^{\infty} n! \binom{m}{n} \frac{(\theta)_{m\uparrow}}{(\theta)_{n+m\uparrow}} d_m^\theta(t) = \\
 & \sum_{m=n}^{\infty} \frac{(m)_{n\downarrow}}{(\theta+m)_{n\uparrow}} d_m^\theta(t) = e^{-\lambda_n t} = 1 - \lambda_n t + o(t). \quad (4.37)
 \end{aligned}$$

Del mismo modo, aplicando el lema 4.3.3,

$$\begin{aligned}
\mathbb{P}[R_{t+s} = n-1 | R_s = n] &= \\
&= \sum_{m=n-1}^{\infty} n(n-1)! \binom{m}{n-1} \frac{(\theta)_{m\uparrow}(\theta+n-1)}{(\theta)_{n+m\uparrow}} d_m^\theta(t) = \\
&= n(n+\theta-1) \sum_{m=n-1}^{\infty} \frac{(m)_{n-1\downarrow}}{(\theta+m)_{n\uparrow}} d_m^\theta(t) = \\
&= \frac{\lambda_n}{\lambda_n - \lambda_{n-1}} (e^{-\lambda_{n-1}t} - e^{-\lambda_n t}) = \lambda_n t + o(t). \quad (4.38)
\end{aligned}$$

Finalmente,

$$\begin{aligned}
\mathbb{P}[R_{t+s} \leq n-2 | R_s = n] &= \\
&= 1 - \mathbb{P}[R_{t+s} = n | R_s = n] - \mathbb{P}[R_{t+s} = n-1 | R_s = n] = o(t). \quad (4.39)
\end{aligned}$$

Por lo tanto, el generador infinitesimal para el proceso $\{R_t\}_{t \geq 0}$ es

$$\lim_{t \rightarrow 0} \frac{\mathbb{P}[R_{t+s} = r | R_s = n] - \delta_n(r)}{t} = \begin{cases} -\lambda_n & \text{si } r = n \\ \lambda_n & \text{si } r = n-1 \\ 0 & \text{si } r \leq n-2 \end{cases}$$

En la sección 1.2.4 se estableció que este es el generador del proceso de muerte $\{D_t\}_{t \geq 0}$, de modo que en el caso en que $\mathbb{P}[R_t = m] = d_m^\theta(t)$, el proceso $\{R_t\}_{t \geq 0}$ está bien definido y coincide con el proceso de muerte, el cual satisface la condición de Chapman-Kolmogorov,

$$d_{t+s}^\theta(r) = \sum_{n=r}^{\infty} \mathbb{P}[D_{t+s} = r | D_s = n] d_s^\theta(n). \quad (4.40)$$

De modo que el proceso $\{R_t\}_{t \geq 0}$ con distribución condicional,

$$\mathbb{P}[R_{t+s} = r | R_s = n] = \sum_{m=r}^{\infty} r! \binom{n}{r} \binom{m}{r} \frac{(\theta)_{n\uparrow}(\theta)_{m\uparrow}}{(\theta)_{n+m\uparrow}(\theta)_{r\uparrow}} d_s^\theta(m), \quad (4.41)$$

la debe satisfacer también. Reemplazando esta última igualdad en la anterior, se tiene que, para $p_t(n) = d_t^\theta(n)$, las ecuaciones de Chapman-Kolmogorov para el proceso $\{\mu_t\}_{t \geq 0}$, expresadas en la ecuación (4.30), se cumplen. Quedando demostrado que $\{\mu_t\}_{t \geq 0}$ es el proceso Fleming-Viot de mutación neutral.

Conclusiones.

En la sección 1.2.2 se dijo que, ante la dificultad para definir procesos de Markov a través de funciones de transición, suele recurrirse a su caracterización por medio de generadores infinitesimales. El proceso Fleming-Viot presentado en el capítulo 3, constituye un ejemplo de esta caracterización, que en su forma más general se define a partir del generador \mathcal{L} dado por la expresión (3.17). Sin embargo, para que dicho generador caracterice al proceso, es necesario probar que el problema de martingala de Strook y Varadhan para el operador \mathcal{L} está bien planteado. Con este fin, Fleming y Viot [FV79] especificaron el dominio $\mathcal{D}(\mathcal{L})$ como un conjunto de funciones de la forma (3.20), para el cual, el operador toma la forma dada por la expresión (3.21) y demostraron que \mathcal{L} se puede aproximar por una sucesión de generadores para procesos de Wright-Fisher reescalados (expresión 3.22). Como el problema de martingala correspondiente a cada proceso de Wright-Fisher está bien planteado, su límite también lo está (lema 1.2.28). Este procedimiento es complicado, incluso en el caso más simple del modelo de mutación neutral.

En la sección 1.4 se introdujo una idea de Pitt *et al.* [PCW02], generalizada por Mena y Walker [MW05, MW07], para definir procesos de Markov a partir de los conceptos en que se basa el método de muestreo de Gibbs. Es un procedimiento que permite definir la función de transición del proceso de interés a partir de transiciones intermedias para un proceso latente. En el capítulo 4 se utilizó esta idea para construir el proceso Fleming-Viot de mutación neutral, a partir de su función de transición.

Aunque los cálculos pueden resultar complejos, la idea es conceptualmente sencilla. Se busca construir un proceso de Markov $\{\mu_t\}_{t \geq 0}$ con valores en el espacio $\mathcal{P}(E)$ de medidas de probabilidad sobre un espacio métrico (E, r) completo y separable, y con distribución estacionaria $\Pi(\cdot) = \mathcal{D}(\cdot | \theta \nu_0)$. En lugar de intentar definir la función de transición $P(t, \mu_0, B)$ de forma directa, lo cual equivaldría a definir las distribuciones condicionales $\mathbb{P}[\mu_t \in B | \mu_0]$ para toda $t \geq 0$, se introduce un «paso intermedio». Es decir,

se introduce un proceso latente a partir de las distribuciones (4.9) a (4.11). Como se trata de las distribuciones condicionales para una distribución conjunta bien definida, la función $P(t, \mu_0, B) := \mathbb{P}[\mu_t \in B | \mu_0]$ dada por la expresión (4.15) está también bien definida y basta que satisfaga las ecuaciones de Chapman-Kolmogorov para garantizar que sea una función de transición homogénea en el tiempo. Así, para cada $\mu_0 \in \mathcal{P}(E)$ existe un proceso de Markov con función de transición $P(t, \mu_0, B)$ y distribución inicial δ_{μ_0} , que además será reversible y estacionario con distribución invariante Π . Gracias a las propiedades del proceso Dirichlet y en particular, a su caracterización como urna de Pólya generalizada, las ecuaciones de Chapman-Kolmogorov se pueden transformar en una condición sobre las $p_t(n)$ (ecuación (4.30)), la cual se satisface haciendo $p_t(n) = d_n^\theta(t)$ (las probabilidades para el proceso de muerte puro introducido en la sección 1.2.4). El proceso $\{\mu_t\}_{t \geq 0}$ que se obtiene coincide con el Fleming-Viot de mutación neutral.

La definición del proceso Fleming-Viot a través de su generador infinitesimal tiene la ventaja de definir, a partir de una única expresión, una amplia familia de modelos. Sin embargo arroja poca luz sobre las propiedades del proceso definido, aún en el caso particular del modelo de mutación neutral. La construcción aquí propuesta, además de mostrar de forma directa la existencia del proceso, su función de transición, su distribución estacionaria y la propiedad de reversibilidad, aporta una caracterización del proceso más manejable desde los puntos de vista estadístico y probabilista. Por ejemplo, sugiere una forma de simular realizaciones aproximadas del proceso, aprovechando los algoritmos de simulación existentes para el proceso Dirichlet (ver por ejemplo [IJ01]).

En la sección 4.3 se demostró que $p_t(n) = d_n^\theta(t)$ es una solución para la condición (4.30), pero no parece evidente que sea la única. Una posible extensión del presente trabajo consistiría en buscar una caracterización para todas las distribuciones $p_t(n)$ que satisfacen dicha condición. De este modo, se obtendría una familia de procesos de Markov reversibles, con distribución estacionaria $\Pi(\cdot) = \mathcal{D}(\cdot | \theta \nu_0)$ y función de transición dada por la expresión (4.15). En caso de que la solución fuese única, se demostraría que el proceso Fleming-Viot de mutación neutral es el único que puede ser construido a partir de este modelo.

Ruggiero y Walker [RW07] presentan una construcción basada en la teoría bayesiana no paramétrica, para el proceso Fleming-Viot con selección por fertilidad (otro caso particular del modelo general). En su propuesta, no introducen una dependencia del tiempo a través de un parámetro; definen el proceso de Markov a tiempo continuo a través de un límite de cadenas de Markov, de forma similar a la aproximación de Fleming y Viot [FV79].

Sin embargo, la construcción parte de un proceso de partículas semejante al definido en la sección 4.1, basado en un modelo de mezclas de procesos Dirichlet, en lugar del proceso Dirichlet. De aquí surge una idea interesante. Es posible generalizar la construcción del capítulo 4 sustituyendo al proceso Dirichlet en la definición de la distribución (4.9), por otro modelo que, en principio, podría ser cualquier urna de Pólya generalizada que diese origen a una sucesión intercambiable. Por ejemplo, se podría probar con el proceso Estable o más generalmente, con el proceso Poisson-Dirichlet (ver [IJ01]). Aunque la propiedad de cerradura o conjugación dada por el lema 2.2.8 no se cumple para estos procesos, es posible encontrar expresiones análogas para la distribución posterior, dada la muestra. Por lo tanto, aparentemente el procedimiento desarrollado en la sección 4.2 para el planteamiento de las condiciones de Chapman-Kolmogorov podría replicarse. De ser esto posible, bastaría encontrar un resultado análogo al teorema 2.2.10 para llegar a una condición para un proceso con espacio de estados discreto, como en la ecuación (4.30).

Finalmente, se podría introducir la dependencia temporal a través de algún otro parámetro de las distribuciones del modelo de partículas, en lugar de involucrarla únicamente en el tamaño de muestra. Se tendría que analizar entonces la posibilidad de replantear de manera sencilla las ecuaciones de Chapman-Kolmogorov para la función de transición definida, así como la existencia y caracterización de las soluciones.

En conclusión, la idea planteada en este trabajo e ilustrada por medio de la construcción del proceso Fleming-Viot de mutación neutral, puede ser una herramienta poderosa para definir procesos con valores en un espacio $\mathcal{P}(E)$, a través de funciones de transición. Constituye una liga entre la estadística bayesiana no paramétrica y los procesos de difusión, cuyo alcance aún está por explorarse.

Bibliografía.

- [Ant74] C. E. Antoniak. Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *The Annals of Statistics*, 2(6):1152–1174, November 1974.
- [BM73] D. Blackwell and J. B. MacQueen. Ferguson distributions via Pólya urn schemes. *The Annals of Statistics*, 2(1):353–355, March 1973.
- [BS04] J. M. Bernardo and A. F. M. Smith. *Bayesian Theory*. Wiley Series in Probability and Statistics. John Wiley & Sons, 2004.
- [Cha05] C. A. Charalambides. *Combinatorial methods in Discrete Distributions*. Wiley Series in Probability and Statistics. John Wiley and Sons, Hoboken, NJ, 2005.
- [Dar59] C. Darwin. *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. John Murray, November 1859.
- [Dud04] R. M. Dudley. *Real Analysis and Probability*. Cambridge University Press, 2004.
- [EG93] S.N. Ethier and R. C. Griffiths. The transition function of a Fleming-Viot process. *The Annals of Probability*, 21(3):1571–1590, 1993.
- [EK93] S.N. Ethier and T. G. Kurtz. Fleming-Viot processes in population genetics. *SIAM J. Control and Optimization*, 31(2):345–386, March 1993.
- [EK86] S.N. Ethier and T. G. Kurtz. *Markov Processes. Characterization and Convergence*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, New York, 1986.

- [Eth00] A. M. Etheridge. *An Introduction to Superprocesses*, volume 20 of *University Lecture Series*. American Mathematical Society, 2000.
- [Fer73] T. S. Ferguson. A Bayesian analysis of some nonparametric problems. *Annals of Statistics*, 1(2):209–230, March 1973.
- [Fis99] R. A. Fisher. *A Genetical Theory of Natural Selection*. Oxford Univ. Press, 3th edition, 1999. Edición original, 1930.
- [FV79] W. H. Fleming and M. Viot. Some measure-valued Markov processes in population genetics theory. *Indiana Univ. Math. J.*, 28:817–843, 1979.
- [GCSR04] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian Data Analysis*. Texts in Statistical Science. Chapman & Hall/CRC, 2nd edition, 2004.
- [Hal24] J. B. S. Haldane. A mathematical theory of natural and artificial selection. *Transactions of the Cambridge Philosophical Society*, 23:19–41, 1924. Primer artículo de una serie de diez.
- [IJ01] H. Ishwaran and L. F. James. Gibbs sampling methods for stick breaking priors. *Journal of the American Statistical Association*, 96(453):161–173, March 2001.
- [IZ03] H. Ishwaran and M. Zarepour. Random probability measures via Pólya sequences: revisiting the Blackwell-MacQueen urn scheme, September 2003. <http://arxiv.org/abs/math/0309041>.
- [Lam77] J. Lamperti. *Stochastic Processes. A Survey of the Mathematical Theory*. Springer-Verlag, New York, 1977.
- [Lo84] A. Y. Lo. On a class of Bayesian nonparametric estimates: I. Density estimates. *The Annals of Statistics*, 12(1):351–357, 1984.
- [MW05] R.H. Mena and S.G. Walker. Stationary autoregressive models via a Bayesian nonparametric approach. *Journal of Time Series Analysis*, 26(6):789–805, November 2005.
- [MW07] R. H. Mena and S. G. Walker. Construction of Markov processes in continuous time. Trabajo sometido, 2007.

-
- [PCW02] M.K. Pitt, C. Chatfield, and S. G. Walker. Constructing first order stationary autoregressive models via latent processes. *Scandinavian Journal of Statistics*, 29(4):657–663, 2002.
- [Pem07] R. Pemantle. A survey of random processes with reinforcement. *Probability Surveys*, 4:1–79, Feb 2007.
- [Pit96] J. Pitman. Some developments of the Blackwell-MacQueen urn scheme. In Ferguson, L. S. Shapley, and MacQueen, editors, *Statistics, probability and game theory, Papers in honor of David Blackwell*, volume 30 of *Institute of Mathematical Statistics, Lecture Notes - Monograph Series*, pages 245–267. Institute of Mathematical Statistics, Hayward, CA, August 1996.
- [RW07] M. Ruggiero and S. G. Walker. Bayesian nonparametric construction of the Fleming-Viot process with fertility selection. *Statistica Sinica*. Preimpresión, 2007.
- [Sch95] M. J. Schervish. *Theory of Statistics*. Springer Series in Statistics. Springer, first edition, 1995.
- [Tav84] S. Tavaré. Line-of-descent and genealogical processes, and their applications in population genetics models. *Theoretical Population Biology*, 26(2):119–164, October 1984.
- [WHN07] S. G. Walker, S. J. Hatjispayros, and T. Nicolieris. A Fleming-Viot process and Bayesian nonparametrics. *The Annals of Applied Probability*, 17(1):67–80, February 2007.
- [Wri32] S. Wright. The roles of mutation, inbreeding, crossbreeding and selection in evolution. In *Proc. 6th Int. Cong. Genet.*, volume 1, pages 356–366, 1932.
- [Wri84] S. Wright. *Evolution and the Genetics of Populations: Genetics and Biometric Foundations*. Univ. of Chicago Press, 1984.