UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
ESCUELA DE POSTGRADO

TESIS PARA OPTAR AL GRADO DE
DOCTOR EN CIENCIAS DE LA INGENIERÍA
MENCIÓN MODELACIÓN MATEMÁTICA
EN COTUTELA CON LA UNIVERSIDAD DE PARIS 5

CUTOFF PARA N-MUESTRAS DE PROCESOS ESTOCÁSTICOS EXPONENCIALMENTE CONVERGENTES Y PARTICIONES ALEATORIAS DEL INTERVALO $[0, 1]$

JAVIERA PAULINA BARRERA MARTÍNEZ

2005

UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
ESCUELA DE POSTGRADO

Cutoff para n-muestras de procesos estocásticos exponencialmente convergentes y particiones aleatorias del intervalo $[0, 1]$

JAVIERA PAULINA BARRERA MARTÍNEZ

MIEMBROS DE LA COMISIÓN EVALUADORA

Sr. Servet Martínez	Profesor Guía	Universidad de Chile
Sr. Bernard Ycart	Profesor Co-Guía	Université de Paris 5
Sr. Thierry Huillet	Profesor Integrante	Université de Cergy-Pontoise
Sr. Jean-René Chazottes	Profesor Integrante	Ecole Polytechnique
Sr. Joaquín Fontbona	Profesor Integrante	Universidad de Chile
Sr. Alejandro Maass	Profesor Integrante	Universidad de Chile
Sr. Jaime San Martín	Profesor Integrante	Universidad de Chile
Sr. Roberto Fernández	Profesor Integrante	Université de Rouen

TESIS PARA OPTAR AL GRADO DE
DOCTOR EN CIENCIAS DE LA INGENIERÍA
MENCIÓN MODELACIÓN MATEMÁTICA
EN COTUTELA CON LA UNIVERSIDAD DE PARIS 5

SANTIAGO DE CHILE
DICIEMBRE 2005

Cuando acabe este verso que canto
yo no sé, yo no sé, madre mía
si me espera la paz o el espanto;
si el ahora o si el todavía.
Pues las causas me andan cercando
cotidianas, invisibles.
Y el azar se me viene enredando
poderoso, invencible.

Causas y azares
Silvio Rodriguez

Agradecimientos

Esta tesis resume el trabajo de cuatro años de doctorado en co-tutela entre la Universidad de Chile y la Universidad de París 5. De estos cuatro años, dos transcurrieron en Chile y dos en Francia. Este trabajo trata dos temas tanto en lo formal como en lo personal; para saber qué es lo que trata formalmente basta leer los próximos 5 capítulos. En lo personal, trata sobre dejar el país que uno quiere para aprender a querer otro.

Primero quiero agradecer a mis dos profesores guías Servet Martínez y Bernard Ycart. Servet fue quien me motivó en el estudio de las probabilidades en su curso de cadenas de Markov y más tarde, calibrando mis cualidades como matemática, me desafió con problemas muy interesantes. Bernard, por su parte, fue un profesor guía cuidadoso, exigiendo tanto pulir los resultados como presentarlos con elegancia. A ambos les agradezco mucho el haber guiado esta tesis.

Les agradezco también a Thierry Huillet, Christian Paroissin y Béatrice Lachaud que colaboraron conmigo en este trabajo de investigación y con quienes fue un placer discutir.

Les agradezco a los profesores Roberto Fernández y Danièle Gardy por darse el tiempo de leer y emitir su opinión sobre esta tesis.

Le agradezco al Departamento de Ingeniería Matemática de la Universidad de Chile donde transcurrieron mis primeros años del doctorado y tuve excelentes cursos y grandes profesores. También les doy las gracias a los funcionarios quienes siempre han sido muy cariñosos conmigo. En la facultad que alberga a este departamento hice mis mejores amigos. Algunos estudiando física un verano: Eduardo, Vicho, Andrés, Jordi, Eduardo, Pedro y Rocío; otros durante la carrera: Isabel, Matías, Ángela, Álvaro, Nicolás, Mariel, Joaquín, Anneli y la Mariela; y aquellas amistades que hice en la cancha, jugando a la pelota: Daren, Ma Alejandra, Ai-ling, Daniela, Andrea, Lorna y Natalia. A todos ellos les agradezco de corazón.

En septiembre del 2003 me fui a Francia con sentimientos encontrados. Se hacía difícil dejar Chile, pero al mismo tiempo Francia despertaba mi curiosidad. En la Universidad de París 5 me esperaba el profesor Bernard Ycart, quien junto a los funcionarios me ayudaron a solucionar todo tipo de problemas. También conocí a los tesistas de la U. de París 5: Béatrice, Olivier, Raphaël y David. Luego encontré a las tesistas del 4 piso Gwendoline, Claire y Amandine. Todos ellos fueron excelentes amigos y compañeros con los que resolvimos puzzles, enigmas y cultivamos plantas a la hora del café. Gracias a Pedro conocí a los chicos del Laboratorio PSE: Facundo, Dimitris, Hector, Michael y Gregory con quienes visitamos diferentes restaurantes de París. También estudiaban en esa época por allá Eduardo y Anneli con quienes compartí esta experiencia. Gracias a todos ellos por hacer de París un hogar.

Los últimos 8 años de mi vida no se pueden separar de Pedro. Juntos partimos a vivir en París y descubrir Europa. Le agradezco de corazón el apoyo y aliento que me brindó para llevar a cabo esta tesis. También le agradezco a la familia de Pedro y a Bárbara por el apoyo incondicional que nos dieron y visitas que nos hicieron.

Desde el fondo de mi alma les agradezco a mi padres, a mis hermanos y al resto mi familia por estar siempre orgullosos de mi trabajo, no sería tan fácil gozar de esta disciplina sin su compresión.

Agradezco al Estado de Chile y de Francia por financiar mi doctorado y al Núcleo Milenio P01-005 y proyecto Mecesup por el apoyo brindado.

Índice general

RESUMEN DEL INFORME FINAL
PARA OPTAR AL TÍTULO DE
DOCTOR EN CIENCIAS DE LA INGENIERÍA
POR: JAVIERA BARRERA MARTÍNEZ
PROF. GUÍA: SR. SERVET MARTÍNEZ
FECHA DE EXAMEN: 6 DICIEMBRE 2005

Cutoff para n-muestras de procesos estocásticos exponencialmente convergentes y particiones aleatorias del intervalo $[0, 1]$

En este trabajo abordamos dos temas de probabilidades uno es el fenómeno de Cutoff en n-tuplas y el otro es sobre particiones aleatorias del intervalo $[0, 1]$. Los resultados de Cutoff corresponden al trabajo realizado en conjunto con B. Lachaud y B. Ycart. Los resultados en el tema de particiones aleatorias en colaboración con T. Huillet y Ch. Paroissin.

El Cutoff es una propiedad definida para familias de procesos de Markov que refleja la convergencia abrupta a la distribución estacionaria. En esta tesis abordamos el tema de las distancias entre distribuciones de probabilidad, motivados por el fenómeno de Cutoff. Luego estudiamos este fenómeno para n-tuplas de procesos independientes, tanto idénticamente distribuidos como no. El resultado principal de la primera parte da condiciones bajo las cuales una n-tupla de procesos que convergen exponencialmente tiene un Cutoff en el sentido de las distancias de Variación total, Hellinger, Chi-cuadrado y Kullback.

La segunda parte de los resultados se centran en las particiones aleatorias. Estudiamos las características de la partición de Dirichlet (D) que después de una permutación sesgada por tamaño es una aproximación de una partición GEM (Griffiths-Engen-McCloskey). También estudiamos la partición del modelo de fragmentación de la vara que aproxima una partición GEM de parámetro 1. Ambos análisis se centran en la función generadora de momentos conjunta de las particiones.

En este trabajo consideramos dos estrategias para organizar ítemes: *desplazar al frente* (MtF) que organiza una lista y *desplazar a la raíz* (MtR) que organiza un árbol binario de búsqueda. Los resultados en torno a las dos estrategias se centran en el análisis del costo de búsqueda en régimen estacionario en el contexto de particiones aleatorias generadas por renormalización. El resultado principal para la regla MtF es la distribución asintótica en el número de ítemes del costo de búsqueda y para la estrategia MtR se encontró el comportamiento asintótico del costo de búsqueda para dos particiones particulares.

Introducción

Como lo expresa el título, en esta tesis abordamos dos temas de probabilidades uno es el fenómeno de Cutoff en n-tuplas y el otro es sobre particiones aleatorias del intervalo $[0, 1]$. Los Capítulos 2 y 3 corresponden al trabajo realizado sobre Cutoff en conjunto con B. Lachaud y B. Ycart [6]. Los Capítulos 4 y 5 abordan los trabajos realizado en el tema de particiones aleatorias en colaboración con T. Huillet [3], con Ch. Paroissin [7] y [8] y con ambos autores [4] y [5].

El Cutoff es una propiedad definida para familias de procesos de Markov que refleja la convergencia abrupta a la distribución estacionaria. En el Capítulo 2 abordamos el tema de las distancias entre distribuciones de probabilidad, motivados por el fenómeno de Cutoff. Básicamente definimos las distancias de Variación total, Hellinger, Chi-cuadrado y Kullback. Resumimos aquellas propiedades relevantes para el desarrollo de nuestros resultados, las que se pueden encontrar en [51] o [29] y calculamos algunos ejemplos. El concepto de Cutoff busca describir la convergencia abrupta que se observa en ciertos procesos estocásticos y está ligado una distancia. Tradicionalmente ha sido la distancia en Variación total, pero no ha sido la única, como discutiremos en el Capítulo 3.

El objetivo del Capítulo 3 es estudiar el fenómeno de Cutoff para n-tuplas de procesos independientes, tanto idénticamente distribuidos como no. Para esto extendemos la definición de Cutoff para otras distancias (Definición 3.1.1) y damos la definición de convergencia exponencial a taza ρ (Definición 3.2.1) que utilizaremos donde buscamos que el logaritmo de la distancia en el instante t se comporte como $-\rho t$. Es en la Sección 3.3 donde exponemos nuestro resultado principal de Cutoff, el Teorema 3.3.1, que da condiciones bajo las cuales una n-tupla de procesos exponencialmente convergentes tiene un Cutoff en el sentido de las distancias de Variación total, Hellinger, Chi-cuadrado y Kullback. En conjunto estos dos capítulos constituyen el artículo "Cutoff for exponentially converging processes" que fue enviado a una revista para su eventual publicación (ver [6]).

El segundo tema de esta tesis es abordado en los Capítulos 4 y 5. En el primero de estos en la Sección 4.2 introducimos dos modelos generales para generar particiones aleatorias. El primero que llamaremos partición por renormalización es generado por n variables independientes que son normalizadas por la suma total de ellas. El segundo, que llamaremos de localización aleatoria, es generado por n variables independientes con soporte en $(0, 1)$ donde los segmentos se originan como el producto de las variables. En la Sección 4.3 introducimos formalmente la partición aleatoria de Dirichlet, Poisson-Dirichlet y GEM. Recordamos varias de sus propiedades y recopilamos algunos resultados nuevos que fueron publicados en el artículo "Size-biased permutation of Dirichlet Partitions and search-cost distribution" [4]. Finalmente en la Sección 4.4 definimos y estudiamos las propiedades del modelo de partición de *fragmentación de la vara* que aproxima la partición de GEM para cierto parámetro. Los resultados descritos en esta sección fueron objeto de la publicación "On the random spliting of the interval" [3].

En el Capítulo 5 presentamos dos estrategias para ordenar ítemes: las estrategias de Move-to-front para ordenar en una lista y Move-to-root para ordenar en un árbol. En la Sección 5.1 se describen ambas estrategias y sus propiedades, entre ellas el costo de búsqueda (Definiciones 5.1.1 y 5.1.5) que mide desempeño de ambas estrategias en el instante t. Ambas reglas han sido ampliamente estudiadas en el contexto de cadenas de Markov donde las probabilidades de transición dependen de la probabilidad (p_i) de solicitar cada uno de los ítemes ver para el Move-to-front los trabajos [60] y [31] y más recientemente [28] y [26] para el Move-to-root ver [20] y [19]. El análisis del tema es diferente al realizado en los trabajos anteriores, pues las probabilidades de requerir un ítem están dadas por una partición aleatoria por renormalización. En la Sección 5.2 estudiamos la regla Move-to-front, nuestro resultado principal para esta estrategia es el Teorema 5.2.7 donde encontramos la distribución del costo de búsqueda estacionario normalizado cuando el número de ítemes tiende a infinito. Para el caso de la regla de Move-to-root, estudiada en la Sección 5.3, también analizamos el comportamiento asintótico en el número de ítemes, pero nuestro resultado se restringe al primer y segundo momento del costo de búsqueda estacionario. Los resultado sobre la estrategia Move-to-front trabajos se pueden encontrar en el artículo "On the distribution of the search cost for the move-to-front rule with random weights" ([7]) y su continuación el artículo "Limiting search cost distribution for move-to-front rule with random request probabilities" ([5]). Así como los resultados de Move-to-root fueron expuestos en el poster "On the stationary search cost for the move-to-root rule with random weight" ([8]).

DISTANCIAS ENTRE DISTRIBUCIONES DE PROBABILIDAD

Una pregunta fundamental en el ámbito de las probabilidades y la estadística ha sido la convergencia de familias de distribuciones de probabilidad. En las aplicaciones surge la necesidad de cuantificar esta convergencia en términos de una distancia. En la literatura existen muchas métricas y otras funciones que nos permiten tener una noción de "distancia" entre distribuciones. Por simplicidad nos referiremos a ellas como distancias, aún cuando en el sentido matemático no lo sean. Estas funciones, o métricas, han probado ser de gran utilidad en diferentes contextos ya sea por su interpretación, sus propiedades teóricas o las relaciones existentes entre ellas. Por ejemplo, a fines de los años 40 y durante los 50 las distancias entre distribuciones de probabilidad se utilizan en el diseño de test estadísticos (ver e.g. [32]). Desde los años 70 grandes esfuerzos se han realizado en el ámbito de la teoría de los valores extremos para establecer el rango de validez de las aproximaciones con respecto a diferentes distancias (ver las notas bibliográficas de los Capítulos 4 y 5 de [51]). En los años 90 se desarrollan los métodos de Monte-Carlo (ver [27]), que consisten en aproximar una probabilidad ν a partir de una cadena de Markov reversible con ley estacionaria ν. En la aplicación de esta técnica es natural que surja la necesidad de imponer una distancia para determinar cuándo detener el algoritmo.

Nuestra motivación para abordar el tema de las distancias es el fenómeno de Cutoff. El concepto de Cutoff busca describir la convergencia abrupta que se observa en ciertos procesos estocásticos y está ligado una distancia. Tradicionalmente esta ha sido la distancia en Variación total, pero no ha sido la única, como discutiremos en el capítulo correspondiente.

Abordar el tema de las distancias entre leyes de probabilidad nos tomarías más que un capítulo de esta tesis. Escogimos aquellas que nos permiten estimar la distancia entre dos medidas producto: la distancia de Variación total (VT), Hellinger (H), Chi-cuadrado (χ^2) y Kullback (K). En la primera de cuatro secciones tratamos la distancia en VT para describir sus diferentes caracterizaciones y propiedades. La segunda está dedicada a las otras tres distancias, a sus propiedades y a estudiar como se relacionan entre ellas y la distancia en VT. En la tercera sección tratamos las distancias entre medias productos y, finalmente, en la última sección proporcionamos el cálculo explícito de las distancias entre dos medidas de probabilidad generadas como un producto tensorial de una misma distribución para el caso de dos productos de Bernoulli, de Poisson y de Gaussianas.

Si bien algunas de las distancias están definidas en un contexto más amplio que el de medidas de probabilidad, nos restringiremos a este último espacio para estudiarlas, por lo que toda afirmación deberá considerarse en este contexto.

2.1. Distancia en Variación total

Consideremos el espacio medible $(E, \mathcal{F})$. Sean μ y ν dos medidas de probabilidad definidas en este espacio. Sea λ una distribución que domina a ambas medidas y denotemos por g (resp. h) la densidad de μ (resp. ν) con respecto a λ.

La distancia en VT entre las dos medidas de probabilidad μ y ν se define como:

$$d_{VT}(\mu,\nu) = \sup_{A\in\mathcal{F}} |\mu(A) - \nu(A)| \ . \tag{2.1}$$

De la definición se desprende que $d_{VT}(\mu,\nu)$ toma valores entre 0 y 1. De la ecuación (2.1) vemos que la distancia en VT entre dos medidas μ y ν es el peor error que cometeríamos al aproximar la medida $\mu(A)$ de un conjunto cualquiera por $\nu(A)$. Esta interpretación puede ser una de las razones que la ha llevado a ser la más utilizada en la literatura. Otra razón puede ser sus otras múltiples caracterizaciones que han permitido desarrollar diferentes técnicas para acotarla. La siguiente propiedad enumera tres de las más conocidas.

Proposición 2.1.1.

1. *Sea $\mu(\varphi) = \int \varphi \, d\mu$ entonces la distancia en VT entre μ y ν vale*

$$d_{VT}(\mu,\nu) = \frac{1}{2}\sup\{|\mu(\varphi) - \nu(\varphi)| \ : \ |\varphi| \leq 1 \ y \ \varphi \ medible \ \} \ .$$

2. *La distancia en VT entre μ y ν es la norma $L^1(\lambda)$ entre las densidades con respecto a λ de μ y ν,*

$$d_{VT}(\mu,\nu) = \frac{1}{2}\int_E |g-h| d\lambda \ .$$

3. *Sean X e Y dos variables aleatorias distribuidas según μ y ν respectivamente, la distancia en VT entre μ y ν vale el ínfimo de la probabilidad que X e Y sean distintas sobre todas las posibles leyes conjuntas entre X e Y. Esta propiedad es conocida como la caracterización de "acoplamiento" (en inglés coupling).*

$$d_{VT}(\mu,\nu) = \inf\{P(X \neq Y) \ : \ X, Y \ t.q. \ \mathcal{L}(X) = \mu \ y \ \mathcal{L}(Y) = \nu\} \ .$$

La caracterización (2) permite trabajar la distancia en VT usando las técnicas de las normas L^p. Por otros lado, la construcción de acoplamientos entre medidas se ha convertido en una importante técnica para estimar la distancia en VT (ver [40]).

Aun cuando la distancia en VT posee estas y otras propiedades, otras distancias han hecho su aparición en la literatura. Algunas porque son más simples de calcular y permiten acotar la distancia en VT. Mientras otras simplemente para reemplazarla, pues la distancia en VT no es capaz de captar ciertos fenómenos como por ejemplo la convergencia de una sucesión de variables aleatorias discretas a una variable continua. La discusión realizada por Gibbs y Su en [29] ilustra bien los motivos generales que llevan a considerar otras distancias y las relaciones entre ellas. En la siguiente sección introducimos las distancias de Hellinger, Chi-cuadrado y Kullback que permiten obtener las cotas del orden correcto para la distancias entre medidas productos de la Sección 2.3.

2.2. Las distancias de Helliger, Chi-cuadrado y Kullback

De las diferentes distancias entre distribuciones que hay en la literatura escogimos estas tres porque tienen la propiedad de permitir trabajar la distancia entre medidas productos en términos de la distancia entre sus marginales. La distancia de χ^2 debe su nombre al test estadístico pues, como veremos, el cuadrado de la distancia normalizado por por la talla de la muestra tiene la misma forma que el estadístico

χ^2. La distancia de K es también conocida como la distancia de Entropía relativa y fue definida por Kullback y Leibler en 1951 como la generalización de la noción de entropía de Shannon. La distancia de H habría sido definida y popularizada por Kakutani en 1948. Hellinger, quien dio el nombre a la distancia, utilizó una cantidad relacionada a esta distancia en la teoría de operadores. Para más detalles sobre el uso de estas distancias ver el artículo de [29] y las referencias citadas en él.

Al igual que en la sección precedente, consideremos el espacio medible $(E, \mathcal{F})$. Sean μ y ν dos medidas de probabilidad definidas en este espacio. Sea λ una distribución que domina a ambas medidas y denotemos por g (resp. h) la densidad de μ (resp. ν) con respecto a λ. Sea S_μ (resp. S_ν) el soporte de μ (resp. ν).

Definición 2.2.1.

1. *La distancia de Hellinger entre μ y ν es*

$$d_H(\mu,\nu) = \frac{1}{\sqrt{2}}\Big(\int_E (\sqrt{f} - \sqrt{g})^2\, d\lambda\Big)^{1/2} = \Big(1 - \int_E \sqrt{fg}\, d\lambda\Big)^{1/2}.$$

2. *La distancia de $L^2(\lambda)$ entre μ y ν es*

$$d_{L^2(\lambda)}(\mu,\nu) = \left[\int (f-g)^2\, d\lambda\right]^{1/2}.$$

3. *La distancia de Kullback entre μ y ν es*

$$d_K(\mu,\nu) = \left[\int_{S_\mu} f \log(f/g)\, d\lambda\right]^{1/2}.$$

Las distancias de H y K son independientes de la medida dominante λ. La distancia de H es una métrica y toma valores entre 0 y 1. La distancia de K no es simétrica en sus argumentos (μ, ν) y toman valores entre 0 e ∞ (ver [48] pág. 61). Si μ es absolutamente continua con respecto a ν luego $d^2_{L^2(\nu)}(\mu, \nu)$ es la distancia de Chi-cuadrado usual entre μ y ν. Como trataremos principalmente este caso, y para asegurar la homogeneidad de los resultados, notaremos por $d_{\chi^2}(\mu, \nu)$ la distancia de Chi-cuadrado (χ^2) entre μ y ν a la distancia $L^2(\nu)$:

$$d_{\chi^2}(\mu, \nu) = d_{L^2(\nu)}(\mu, \nu) \ .$$

La distancia de χ^2 tampoco es simétrica y toma valores entre 0 e ∞.

La siguiente proposición resume las cotas clásicas que relacionan las cuatro distancias (ver figura 1 de [58]).

Proposición 2.2.2.

1. $d_H(\mu, \nu)^2 \leq d_{TV}(\mu, \nu)$
2. $d_{TV}(\mu, \nu) \leq d_H(\mu, \nu)\sqrt{2 - d_H(\mu, \nu)^2} \leq \sqrt{2} d_H(\mu, \nu)$
3. $d_{TV}(\mu, \nu) \leq 2 d_{\chi^2}(\mu, \nu)$
4. $d_{TV}(\mu, \nu) \leq 2/\sqrt{2} d_K(\mu, \nu)$
5. $d_H(\mu, \nu) \leq \sqrt{2} d_K(\mu, \nu)$
6. $d_K(\mu, \nu) \leq \sqrt{\log(1 + d_{\chi^2}(\mu, \nu)^2)} \leq d_{\chi^2}(\mu, \nu)$

De las dos primeras desigualdades de esta proposición se desprende que la distancia en VT y de H son equivalentes en el sentido de que inducen la misma topología.

2.3. Distancia entre medidas productos

Como mencionamos en la introducción del capítulo estamos interesados en la distancia entre medidas producto. Sea $(E^{(n)}, \mathcal{F}^{(n)})$ el espacio que resulta del producto de los espacios $(E_i, \mathcal{F}_i)$ con $i = 1, \ldots, n$, es decir, $E^{(n)} = E_1 \times \ldots \times E_n$ y $\mathcal{F}^{(n)} = \mathcal{F}_1 \otimes \ldots \otimes \mathcal{F}_n$. Consideremos las medidas de probabilidad $\mu^{(n)}$ y $\nu^{(n)}$ definidas en $(E^{(n)}, \mathcal{F}^{(n)})$ como el producto de medidas $\mu^{(n)} = \mu_1 \otimes \ldots \otimes \mu_n$ donde cada μ_i esta definida en $(E_i, \mathcal{F}_i)$. La siguiente proposición nos permite relacionar las distancias entre cada par de medidas μ_i y ν_i con la distancia entre las medidas producto $\mu^{(n)}$ y $\nu^{(n)}$.

Proposición 2.3.1.

1. *Variación total:*

$$1 - \exp\left(-\frac{1}{2}\sum_{i=1}^{n} d_{TV}^2(\mu_i, \nu_i)\right) \leq d_{TV}(\mu^{(n)}, \nu^{(n)}) \leq \sum_{i=1}^{n} d_{TV}(\mu_i, \nu_i)\ . \tag{2.2}$$

2. *Hellinger:*

$$d_H^2(\mu^{(n)}, \nu^{(n)}) = 1 - \prod_{i=1}^{n}\Big(1 - d_H^2(\mu_i, \nu_i)\Big)\ .$$

$$1 - \exp\left(-\sum_{i=1}^{n} d_H^2(\mu_i, \nu_i)\right) \leq d_H^2(\mu^{(n)}, \nu^{(n)}) \leq \sum_{i=1}^{n} d_H^2(\mu_i, \nu_i)\ . \tag{2.3}$$

3. *Chi-cuadrado:*

$$d_{\chi^2}^2(\mu^{(n)}, \nu^{(n)}) = \prod_{i=1}^{n}\Big(1 + d_{\chi^2}^2(\mu_i, \nu_i)\Big) - 1\ .$$

$$\sum_{i=1}^{n} d^2_{\chi^2}(\mu_i,\nu_i) \leq d^2_{\chi^2}(\mu^{(n)},\nu^{(n)}) \leq \exp\left(\sum_{i=1}^{n} d^2_{\chi^2}(\mu_i,\nu_i)\right) - 1\ . \tag{2.4}$$

4. *Kullback:*

$$d^2_K(\mu^{(n)},\nu^{(n)}) = \sum_{i=1}^{n} d^2_K(\mu_i,\nu_i)\ . \tag{2.5}$$

Esta proposición muestra que todas las distancias, salvo la distancia en variación total, se comportan como $\left(\sum d(\mu_i,\nu_i)^2\right)^{1/2}$ cuando la distancia en cada coordenada $d(\mu_i,\nu_i)$ es pequeña. Basta recordar que si $x_1,\ldots,x_n$ se comportan aproximadamente como $K_i\ n$ con K_i constante, entonces se puede hacer la siguiente aproximación:

$$\prod_{i=1}^{n}(1+x_i) \sim 1 + \sum_{i=1}^{n} x_i\ .$$

Las relaciones expuestas en la Proposición 2.3.1 para la cota superior de VT y las expresiones para la distancias de H, χ^2 y K son conocidas y se pueden encontrar en el Lema 3.3.10 p. 100 en [51]. La demostración de la cota inferior de la distancia en VT es una simple combinación de las Proposiciones 2.3.1 y 2.2.2.

Demostración:

$$\begin{aligned}
d_{VT}(\mu^{(n)},\nu^{(n)}) &\geq d^2_H(\mu^{(n)},\nu^{(n)}) \\
&= 1 - \prod_{i=1}^{n}\left(1 - d^2_H(\mu_i,\nu_i)\right) \\
&\geq 1 - \exp\left(-\sum_{i=1}^{n} d^2_H(\mu_i,\nu_i)\right) \\
&\geq 1 - \exp\left(-\frac{1}{2}\sum_{i=1}^{n} d^2_{VT}(\mu_i,\nu_i)\right)\ .
\end{aligned}$$

□

2.4. Ejemplos

El objetivo de esta sección es dar algunos ejemplos de estimaciones precisas de la distancia entre medidas producto. En general si consideramos dos medidas de la misma familia paramétrica es razonable pensar que la distancia entre ellas será una función suave de la diferencia entre parámetros. En este caso consideraremos las distribuciones de Bernoulli, Poisson y Normal. En estos resultados, ε y δ deben ser interpretados como funciones de n. Observemos que en las Proposiciones 2.4.1 y 2.4.3 tanto las cotas en el caso de VT como las expresiones para las otras distancias convergen a constantes positivas para $\varepsilon = n^{-1/2}$. De manera similar en la Proposición 2.4.6, para $\text{máx}\{\varepsilon, \delta\} = n^{-1/2}$. Los cálculos para las distancias de H, χ^2 y K son simples y la mayor dificultad que presentan es realizar algunos desarrollos de orden. En el caso de la distancia en VT para las distribuciones de Poisson y Normal los cálculos son más complejos y para demostrarlas utilizamos el mismo argumento que da Pollard en [48] (pág. 63) para la distancia entre dos Gaussianas de misma varianza.

2.4.1. Bernoulli

En esta subsección μ y ν son dos distribuciones de Bernoulli.

Proposición 2.4.1. *[6] Sea p y ε dos reales entre* 0 *y* 1*. Sea μ (respectivamente ν) una distribución de Bernoulli de parámetro $p(1-\varepsilon)$ (respectivamente p). Sea μ^n*

(y respectivamente ν^n) el producto tensorial de n copias de μ (respectivamente ν). Luego,

1. *Variación total*

$$1-\exp\left(-\frac{n}{2}p^2\varepsilon^2\right) \leq d_{VT}(\mu^n,\nu^n) \leq \left[1-\exp\left(-n\left(\frac{p}{4(1-p)}\varepsilon^2+o(\varepsilon^2)\right)\right)\right]^{1/2}.$$

2. *Hellinger*

$$d_H(\mu^n,\nu^n) = \left[1-\exp\left(-n\left(\frac{p}{8(1-p)}\varepsilon^2+o(\varepsilon^2)\right)\right)\right]^{1/2}.$$

3. *Chi-cuadrado*

$$d_{\chi^2}(\mu^n,\nu^n) = \left[\left(1+\frac{p}{1-p}\varepsilon^2\right)^n-1\right]^{1/2}.$$

4. *Kullback*

$$d_K(\mu^n,\nu^n) = \sqrt{n}\left[\frac{p}{2(1-p)}\varepsilon^2+o(\varepsilon^2)\right]^{1/2}.$$

Demostración: Para la distancia de Hellinger de su definición tenemos

$$d_H(\mu,\nu)^2 \;=\; 1-\left(p\sqrt{1-\varepsilon}+(1-p)\sqrt{1+\frac{p}{1-p}\varepsilon}\right).$$

Utilizando una aproximación de la función $\sqrt{1+x}$ encontramos que

$$\begin{aligned}
d_H(\mu,\nu)^2 &= 1 - p\left[1 - \frac{\varepsilon}{2} - \frac{\varepsilon^2}{8} + o(\varepsilon^2)\right] \\
&\quad -(1-p)\left[1 + \frac{p}{2(1-p)}\varepsilon - \frac{p^2}{8(1-p)^2}\varepsilon^2 + o(\varepsilon^2)\right] \\
&= \frac{p}{8(1-p)}\varepsilon^2 + o(\varepsilon^2)\ .
\end{aligned}$$

Luego la distancia de H entre las dos distribuciones de Bernoulli vale

$$d_H(\mu,\nu) = \frac{1}{2}\sqrt{\frac{p}{2(1-p)}}|\varepsilon| + o(\varepsilon)\ .$$

Utilizando la proposición 2.2.2 podemos calcular el producto.

$$\begin{aligned}
d_H(\mu^n,\nu^n) &= \left[1 - \left(1 - \frac{p}{8(1-p)}\varepsilon^2 + o(\varepsilon^2)\right)^n\right]^{1/2} \\
&= \left[1 - \exp\left(-n\left(\frac{p}{8(1-p)}\varepsilon^2 + o(\varepsilon^2)\right)\right)\right]^{1/2}\ .
\end{aligned}$$

Realizamos el mismo calculo para la distancia de χ^2:

$$\begin{aligned}
d_{\chi^2}(\mu,\nu) &= \left((1-\varepsilon)^2 p + (1-p) + 2\varepsilon p + \frac{p^2\varepsilon^2}{1-p} - 1\right)^{1/2} \\
&= \left(\varepsilon^2 p + \frac{p^2\varepsilon^2}{1-p}\right)^{1/2} \\
&= \sqrt{\frac{p}{1-p}}\ |\varepsilon|\ .
\end{aligned}$$

Utilizando la proposición 2.3.1 podemos calcular la distancia entre las medidas producto μ^n y ν^n

$$\begin{aligned} d_{\chi^2}(\mu^n,\nu^n) &= \left[\left(1+\frac{p}{1-p}\varepsilon^2\right)^n - 1\right]^{1/2} \\ &= \left[\exp\left(n\left(\frac{p}{1-p}\varepsilon^2+o(\varepsilon^2)\right)\right)-1\right]^{1/2} . \end{aligned}$$

Para la distancia de K tenemos

$$d_K(\mu,\nu)^2 = (1-\varepsilon)p\log(1-\varepsilon)+(1-p+\varepsilon p)\log\left(1+\frac{p}{1-p}\varepsilon\right) .$$

Para obtener una expresión similar a la obtenida para las otras distancias utilizamos la aproximación de Taylor de $\log(1+x)$ en torno a 0,

$$\begin{aligned} d_K(\mu,\nu)^2 &= (1-\varepsilon)p\left(-\varepsilon-\frac{\varepsilon^2}{2}+o(\varepsilon^2)\right) \\ &\quad +(1-p+\varepsilon p)\left(\frac{p}{1-p}\varepsilon-\frac{1}{2}\left(\frac{p}{1-p}\right)^2\varepsilon^2+o(\varepsilon^2)\right) \\ &= \frac{p}{2(1-p)}\varepsilon^2+o(\varepsilon^2) . \end{aligned}$$

Entonces la distancia de K entre las dos leyes de Bernoulli esta dada por:

$$d_K(\mu,\nu)=\sqrt{\frac{p}{2(1-p)}}|\varepsilon|+o(\varepsilon) .$$

Utilizando la proposición 2.3.1 obtenemos que para el producto de n copias de Bernoulli la distancia de K es:

$$d_K(\mu,\nu)=\sqrt{n}\left[\sqrt{\frac{p}{2(1-p)}}|\varepsilon|+o(\varepsilon)\right] .$$

Un sencillo cálculo nos permite obtener la distancia en VT:

$$\begin{aligned} d_{VT}(\mu,\nu) &= \frac{1}{2}|p - p' + (1-p') - (1-p)| \\ &= p|\varepsilon| \ . \end{aligned}$$

Utilizando la proposición 2.3.1 podemos acotar inferiormente la distancia en VT entre μ^n y ν^n y como cota superior utilizaremos la relación entre VT y H de la Proposición 2.2.2.

$$1 - \exp\left(-\frac{n}{2}p^2\varepsilon^2\right) \le d_{VT}(\mu^n,\nu^n) \le \left[1 - \exp\left(-n\left(\frac{p}{4(1-p)}\varepsilon^2 + o(\varepsilon^2)\right)\right)\right]^{1/2} \ .$$

$\square$

Observación 2.4.2. *Es importante notar que para los cuatro casos la distancia entre μ y ν se comporta de manera lineal con respecto a ε cuando este valor es pequeño:*

$$\begin{aligned} d_{VT}(\mu,\nu) &= p|\varepsilon| \ , \\ d_H(\mu,\nu) &= \frac{1}{2}\sqrt{\frac{p}{2(1-p)}}|\varepsilon| + o(\varepsilon) \ , \\ d_{\chi^2}(\mu,\nu) &= \sqrt{\frac{p}{1-p}} \ |\varepsilon| \ , \\ d_K(\mu,\nu) &= \sqrt{\frac{p}{2(1-p)}}|\varepsilon| + o(\varepsilon) \ . \end{aligned}$$

2.4.2. Poisson

En esta subsección μ y ν son dos distribuciones de Poisson.

Proposición 2.4.3. *[6] Sea α un real positivo y ε un real en el intervalo* $(-\infty, 1)$.

Sea μ (respectivamente ν) una distribución de Poisson de parámetro $\alpha(1-\varepsilon)$ (respectivamente α). Sea μ^n (y respectivamente ν^n) el producto tensorial de n copias de μ (respectivamente ν). Luego,

1. *Variación total*

$$1-\exp\left(-\frac{n}{2}\left(R_\alpha^2\varepsilon^2+o(\varepsilon^2)\right)\right)\leq d_{TV}(\mu^n,\nu^n)\ ,$$

   y

$$d_{TV}(\mu^n,\nu^n)\leq\left[1-\exp\left(-n\left(\frac{\alpha}{4}\varepsilon^2+o(\varepsilon^2)\right)\right)\right]^{1/2}\ ,$$

   con $R_\alpha=\frac{\alpha^{\lfloor\alpha\rfloor+1}e^{-\alpha}}{\lfloor\alpha\rfloor!}$.

2. *Hellinger*

$$d_H(\mu^n,\nu^n)=\left[1-\exp\left(-n\alpha\left(1-\frac{\varepsilon}{2}-\sqrt{1-\varepsilon}\right)\right)\right]^{1/2}\ .$$

3. *Chi-cuadrado*

$$d_{\chi^2}(\mu^n,\nu^n)=\left[\exp\left(n\alpha\varepsilon^2\right)-1\right]^{1/2}\ .$$

4. *Kullback*

$$d_K(\mu^n,\nu^n)=\sqrt{n}\left[\alpha(1-\varepsilon)\log(1-\varepsilon)+\alpha\varepsilon\right]^{1/2}\ .$$

Demostración: Los cálculos realizados para las distancias de H, χ^2 y K son similares a los desarrollados en el caso de Bernoulli. Sea $\beta = \alpha(1-\varepsilon)$, para la distancia de H:

$$\begin{aligned} d_H(\mu,\nu)^2 &= 1 - \sum_{i=0}^{\infty} \frac{\alpha^{i/2}\exp\left(-\frac{\alpha}{2}\right)}{(i!)^{1/2}} \frac{\beta^{i/2}\exp\left(-\frac{\beta}{2}\right)}{(i!)^{1/2}} \\ &= 1 - \exp\left(-\frac{\alpha+\beta}{2}\right)\exp\sqrt{\alpha\beta}\ . \end{aligned}$$

Reemplazando β obtenemos la siguiente expresión:

$$d_H(\mu,\nu) = \left[1 - \exp\left(-\alpha + \frac{\alpha}{2}\varepsilon + \alpha\sqrt{1-\varepsilon}\right)\right]^{1/2}\ .$$

Para simplificar esta expresión utilizamos la aproximación de Taylor de $\sqrt{1+x}$ y encontramos:

$$\begin{aligned} d_H(\mu,\nu) &= \left[1 - \exp\left(-\alpha + \frac{\alpha}{2}\varepsilon + \alpha\left(1 - \frac{\varepsilon}{2} - \frac{\varepsilon^2}{8} + o(\varepsilon^2)\right)\right)\right]^{1/2} \\ &= \left[1 - \exp\left(-\frac{\alpha}{8}\varepsilon^2 + o(\varepsilon^2)\right)\right]^{1/2}\ . \end{aligned}$$

Por último aproximamos $\exp(x)$ con lo que obtenemos:

$$d_H(\mu,\nu) = \frac{\sqrt{\alpha}}{2\sqrt{2}}|\varepsilon| + o(\varepsilon)\ .$$

La Proposición 2.3.1 nos permite obtener directamente la distancia entre μ^n y ν^n

$$d_H(\mu^n,\nu^n) = \left[1 - \exp\left(n\left(-\frac{\alpha}{8}\varepsilon^2 + o(\varepsilon^2)\right)\right)\right]^{1/2}\ .$$

Realizamos el mismo cálculo para la distancia de χ^2, nuevamente sea $\beta = (1-\varepsilon)\alpha$.

Luego:

$$\begin{aligned} d_{\chi^2}(\mu,\nu) &= \left[\sum_{i=0}^{\infty} \frac{\beta^{2i}\exp(-2\beta)}{(i!)^2}\frac{i!}{\alpha^i\exp(-\alpha)} - 1\right]^{1/2} \\ &= \left[\exp\left(\frac{(\alpha-\beta)^2}{\alpha}\right) - 1\right]^{1/2} . \end{aligned}$$

Reemplazando β

$$d_{\chi^2}(\mu,\nu) = \left[\exp(\alpha\varepsilon^2) - 1\right]^{1/2} ,$$

la distancia entre μ^n y ν^n será

$$d_{\chi^2}(\mu^n,\nu^n) = \left[\exp(n\alpha\varepsilon^2) - 1\right]^{1/2} .$$

Por último para la distancia de K para $\beta = (1-\varepsilon)\alpha$ un cálculo análogo produce:

$$\begin{aligned} d_K(\mu,\nu) &= \left[\sum_{i=0}^{\infty} \frac{\beta^{i}\exp(-\beta)}{i!}\log\left(\frac{\beta^i\exp(-\beta)}{\alpha^i\exp(-\alpha)}\right)\right]^{1/2} \\ &= \left[\beta\log\left(\frac{\beta}{\alpha}\right) + \alpha - \beta\right]^{1/2} . \end{aligned}$$

Reemplazando β y aproximando en torno a 0 la función $\log(1+x)$ y $\sqrt{1+x}$ tenemos

$$\begin{aligned} d_K(\mu,\nu) &= ((1-\varepsilon)\alpha\log(1-\varepsilon) + \alpha\varepsilon)^{1/2} \\ &= \left((1-\varepsilon)\alpha(-\varepsilon - \frac{\varepsilon^2}{2} + o(\varepsilon^2)) + \alpha\varepsilon\right)^{1/2} \\ &= \left(-\frac{\alpha}{2}\varepsilon^2 + \alpha\varepsilon^2 + o(\varepsilon^2)\right)^{1/2} \\ &= \sqrt{\frac{\alpha}{2}}|\varepsilon| + o(\varepsilon) . \end{aligned}$$

Obtenemos la distancia entre μ^n y ν^n:

$$\begin{aligned} d_K(\mu^n,\nu^n) &= \left(n\left[(1-\varepsilon)\alpha\log(1-\varepsilon)+\alpha\varepsilon\right]\right)^{1/2} \\ &= \sqrt{n}\left(\sqrt{\frac{\alpha}{2}}|\varepsilon|+o(\varepsilon)\right) \ . \end{aligned}$$

Usando la segunda caracterización de la distancia en VT de la Proposición 2.1.1 tenemos

$$\begin{aligned} d_{VT}(\mu,\nu) &= \frac{1}{2}\sum_{i=0}^{\infty}\frac{\alpha^i\exp(-\alpha)}{i!}|(1-\varepsilon)^i\exp(\alpha\varepsilon)-1| \\ &= \frac{1}{2}\mathbb{E}_\nu\left(|\exp(\alpha\varepsilon+X\log(1-\varepsilon))-1|\right) \ . \end{aligned}$$

Sea $f(\varepsilon,x) = \exp(\alpha\varepsilon+x\log(1-\varepsilon))$, luego $d_{VT}(\mu,\nu) = \mathbb{E}_\nu(|f(\varepsilon,X)-1|)$. Para estimar esta cantidad realizaremos un desarrollo de Taylor de $f(\varepsilon,x)$ como función de ε.

$$\begin{aligned} \frac{\partial f}{\partial\varepsilon}(\varepsilon,x) &= f(\varepsilon,x)\left(\alpha-\frac{x}{1-\varepsilon}\right) \ , \\ \frac{\partial^2 f}{\partial\varepsilon^2}(\varepsilon,x) &= f(\varepsilon,x)\left(\left(\alpha-\frac{x}{1-\varepsilon}\right)^2+\frac{x}{(1-\varepsilon)^2}\right) \ . \end{aligned}$$

luego obtenemos la siguiente aproximación de Taylor para f:

$$f(\varepsilon,x)=1+(\alpha-x)\varepsilon+\int_0^1 e^{\alpha\theta\varepsilon+x\log(1-\theta\varepsilon)}\left((\alpha-\frac{x}{1-\theta\varepsilon})^2+\frac{x}{(1-\theta\varepsilon)^2}\right)\frac{\varepsilon^2}{2}(1-\theta)d\theta \ .$$

Llamemos $\epsilon(x,\varepsilon)$ al termino integral. Podemos acotar la $d_{VT}(\mu,\nu)$ por

$$\mathbb{E}_\nu|\alpha-X|\,|\varepsilon|-\mathbb{E}_\nu|\epsilon(X,\varepsilon)|\le 2d_{VT}(\mu,\nu)\le\mathbb{E}_\nu|\alpha-X|\,|\varepsilon|+\mathbb{E}_\nu|\epsilon(X,\varepsilon)| \ .$$

Vamos a probar que la esperanza $\mathbb{E}_\nu \left|\epsilon(X,\varepsilon)\right|$ se comporta como $o(\varepsilon)$ y que por lo tanto $d_{VT}(\mu,\nu) \sim \frac{1}{2}\mathbb{E}_\nu\left(|\alpha - X|\right)|\varepsilon|$. El teorema de Fubini nos permite intercambiar la esperanza con respecto X y la integral con respecto a θ.

$$\begin{aligned}
\mathbb{E}_\nu \left|\epsilon(X,\varepsilon)\right| &= \mathbb{E}_\nu\left[\int_0^1 e^{\alpha\theta\varepsilon + X\log(1-\theta\varepsilon)}\left((\alpha - \frac{X}{1-\theta\varepsilon})^2 + \frac{X}{(1-\theta\varepsilon)^2}\right)\frac{\varepsilon^2}{2}(1-\theta)d\theta\right] \\
&= \int_0^1 e^{\alpha\theta\varepsilon}\mathbb{E}_\nu\left[e^{X\log(1-\theta\varepsilon)}\left((\alpha - \frac{X}{1-\theta\varepsilon})^2 + \frac{X}{(1-\theta\varepsilon)^2}\right)\right]\frac{\varepsilon^2}{2}(1-\theta)d\theta \ .
\end{aligned}$$

El siguiente lema nos permitirá calcular la esperanza al interior de la integral

Lema 2.4.4. *Sea K un real, α y m dos reales positivos. Sea Y una variable de Poisson de parámetro α. Entonces*

$$\mathbb{E}(e^{YK}Y^m) = \mathbb{E}(\tilde{Y}^m)\exp\left(-\alpha\left(1-e^K\right)\right) \ ,$$

donde $\widetilde{Y}$ es una variable de Poisson de parámetro αe^K.

Este lema nos permite obtener:

$$\mathbb{E}_\nu\left[e^{X\log(1-\theta\varepsilon)}\left(\left(\alpha - \frac{X}{1-\theta\varepsilon}\right)^2 + \frac{X}{(1-\theta\varepsilon)^2}\right)\right] = 2\frac{\alpha}{1-\theta\varepsilon}e^{-\alpha\theta\varepsilon} \ .$$

Reemplazando este valor en la expresión integral de $\mathbb{E}_\nu(\epsilon(\varepsilon, X))$ obtenemos

$$\begin{aligned}
\mathbb{E}_\nu \left|\epsilon(X,\varepsilon)\right| &= \int_0^1 \left(\frac{\alpha}{1-\theta\varepsilon}\right)\varepsilon^2(1-\theta)d\theta \\
&= \alpha\left(\frac{-\log(1-\varepsilon)\varepsilon - \varepsilon - \log(1-\varepsilon)}{\varepsilon^2}\right)\varepsilon^2 \\
&= \frac{3}{2}\alpha\varepsilon^2 + o(\varepsilon^2) \ .
\end{aligned}$$

Deducimos entonces que $d_{VT}(\mu,\nu)$ se comporta como $\frac{1}{2}\mathbb{E}|\alpha - X||\varepsilon|$ para ε pequeño.

Aún debemos calcular $\mathbb{E}_\nu|X-\alpha|$. Como $\mathbb{E}_\nu(X)=\alpha$, tendremos que $\mathbb{E}_\nu|X-\alpha| = 2\mathbb{E}_\nu|X-\alpha|_+$ luego

$$\begin{aligned}
\mathbb{E}_\nu|X-\alpha| &= 2\sum_{i=0}^{\lfloor\alpha\rfloor}(\alpha-i)\frac{\alpha^i}{i!}e^{-\alpha} \\
&= 2\left(\alpha\sum_{i=0}^{\lfloor\alpha\rfloor}\frac{\alpha^i}{i!}e^{-\alpha} - \alpha\sum_{i=1}^{\lfloor\alpha\rfloor}\frac{\alpha^{i-1}}{(i-1)!}e^{-\alpha}\right) \\
&= 2\alpha P_\nu(X=\lfloor\alpha\rfloor)\ .
\end{aligned}$$

Con esto concluimos que la distancia en variación total entre μ y ν vale

$$d_{VT}(\mu,\nu) = \frac{\alpha^{1+\lfloor\alpha\rfloor}}{\lfloor\alpha\rfloor!}e^{-\alpha}|\varepsilon| + o(\varepsilon)\ .$$

Utilizando las cotas de la Proposición 2.3.1 podemos acotar $d_{VT}(\mu^n,\nu^n)$ por

$$1-\exp\left(-\frac{n}{2}\left(R_\alpha^2\varepsilon^2+o(\varepsilon^2)\right)\right) \le d_{VT}(\mu^n,\nu^n) \le \left[1-\exp\left(-n\left(\frac{\alpha}{4}\varepsilon^2+o(\varepsilon^2)\right)\right)\right]^{1/2}\ ,$$

con $R_\alpha = \frac{\alpha^{\lfloor\alpha\rfloor+1}e^{-\alpha}}{\lfloor\alpha\rfloor!}$. $\square$

Observación 2.4.5. *Al igual que en la subsección anterior para los cuatro casos la distancia entre μ y ν se comporta de manera lineal con respecto a ε cuando este valor es pequeño:*

$$\begin{aligned}
d_{VT}(\mu,\nu) &= \frac{\alpha^{1+\lfloor\alpha\rfloor}}{\lfloor\alpha\rfloor!}e^{-\alpha}|\varepsilon| + o(\varepsilon)\ , \\
d_H(\mu,\nu) &= \frac{\sqrt{\alpha}}{2\sqrt{2}}|\varepsilon| + o(\varepsilon)\ , \\
d_{\chi^2}(\mu,\nu) &= \sqrt{\alpha}\ \varepsilon + o(\varepsilon)\ , \\
d_K(\mu,\nu) &= \sqrt{\frac{\alpha}{2}}|\varepsilon| + o(\varepsilon)\ .
\end{aligned}$$

2.4.3. Normal

En esta subsección μ y ν son dos distribuciones Normales.

Proposición 2.4.6. *[6] Sea m, ε dos reales, v un real positivo y δ un real mayor que -1. Sea μ (respectivamente ν) una distribución Normal de esperanza $m+v\varepsilon$ (respectivamente m) y varianza $v^2(1+\delta)$ (respectivamente v^2). Sea μ^n (y respectivamente ν^n) el producto tensorial de n copias de μ (respectivamente ν). Luego,*

1. *Variación total:*

$$1 - \exp\left(-\frac{n}{16\pi}\left(C_1(z)|\delta| + C_2(z)|\varepsilon| + o(||\varepsilon, \delta||)\right)^2\right) \leq d_{TV}(\mu^n, \nu^n) \ ,$$

$$d_{TV}(\mu^n, \nu^n) \leq \left(1 - \left(\frac{\sqrt{1+\delta}}{1+\delta/2}\right)^n \exp\left(-\frac{n\varepsilon^2}{2(2+\delta)}\right)\right)^{1/2} \ .$$

Donde $z = |\varepsilon|/|\delta|$, y C_1, C_2 son tales que $C_1(\infty) = 1$, $C_2(\infty) = 2$, $C_1(0) = 2e^{-1/2}$, y $C_2(0) = e^{-1/2}$; más precisamente, $C_1(z)$ y $C_2(z)$ toman valores en los intervalos $[e^{-1/2}, 1 + e^{-1/2}]$ y $[e^{-1/2}, 2]$ respectivamente.

2. *Hellinger:*

$$d_H(\mu^n, \nu^n) = \left(1 - \left(\frac{\sqrt{1+\delta}}{1+\delta/2}\right)^{n/2} \exp\left(-\frac{n\varepsilon^2}{4(2+\delta)}\right)\right)^{1/2} \ .$$

3. *Chi-cuadrado:*

$$d_{\chi^2}(\mu^n, \nu^n) = \left(\left(1-\delta^2\right)^{-n/2} \exp\left(\frac{n\varepsilon^2}{1-\delta}\right) - 1\right)^{1/2} \ .$$

4. Kullback:

$$d_K(\mu^n, \nu^n) = \sqrt{n}\left(\log(1+\delta) + \left(\frac{\varepsilon^2 - 1}{2(1+\delta)}\right)\right)^{1/2}.$$

Demostración: En primer lugar calcularemos la distancia entre dos distribuciones Normales para el caso particular de $\mu_0 \sim N(m_0, v_0^2)$ y $\nu_0 \sim N(0,1)$ y luego generalizaremos el resultado. Repetiremos el mismo argumento para cada distancia. En el caso de la distancia de H de su definición tenemos

$$d_H(\mu,\nu) = \left[1 - \int \frac{1}{\sqrt{2\pi}\sqrt{v_0}} \exp\left(-\frac{1}{4}\left(\frac{x - m_0}{v_0}\right)^2 - \frac{1}{4}x^2\right) dx\right]^{1/2}.$$

Utilizando la siguiente igualdad

$$\frac{1}{4}\left(\left(\frac{x - m_0}{v_0}\right)^2 + x^2\right) = \frac{1}{2}\left[x - \frac{m}{v_0^2 + 1}\right]^2 \frac{1 + v_0^2}{2v^2} + \frac{1}{4}\frac{m_0^2}{v_0^2 + 1},$$

podemos concluir que

$$d_H(\mu,\nu) \quad = \quad \left[1 - \sqrt{2}\left(\frac{v_0}{v_0^2 + 1}\right)^{1/2} \exp\left(-\frac{1}{4}\frac{m_0^2}{v_0^2 + 1}\right)\right]^{1/2}.$$

Ahora consideremos dos variables normales cualesquiera $\mu \sim N(m_1, v_1^2)$ y $\nu \sim N(m_2, v_2^2)$. Sean $m_0 = \frac{m_1 - m_2}{v_2}$ y $v_0 = \frac{v_1}{v_2}$, tendremos que la distancia de H entre dos normales se puede obtener como $d_H(\mu,\nu) = d_H(N(m_0, v_0^2), N(0,1))$. En nuestro caso $m_1 = m + v\varepsilon$, $v_1 = v\sqrt{1+\delta}$ $m_2 = m$ y $v_2 = v$ luego $m_0 = \varepsilon$ y $v_0 = \delta$, por lo tanto

$$d_H(\mu,\nu) = \left[1 - \left(\frac{\sqrt{1+\delta}}{1 + \delta/2}\right)^{1/2} \exp\left(-\frac{1}{4}\frac{\varepsilon^2}{(2+\delta)}\right)\right]^{1/2}.$$

Reemplazando esta expresión en la Propiedad 2.3.1 obtenemos

$$d_H(\mu^n,\nu^n) \;=\; \left[1-\left(\frac{\sqrt{1+\delta}}{1+\delta/2}\right)^{n/2}\exp\left(-\frac{n}{4}\frac{\varepsilon^2}{(2+\delta)}\right)\right]^{1/2} .$$

Utilizando las aproximaciones de Taylor de $(1+x)^{1/k}$ para simplificar la expresión de $d_H(\mu,\nu)$ obtenemos el siguiente desarrollo de orden

$$\begin{aligned}\left(\frac{\sqrt{1+\delta}}{1+\delta/2}\right)^{1/2} &= \left(1+\frac{\delta}{4}-\frac{3}{32}\delta^2+o(\delta^2)\right)\left(1-\frac{\delta}{4}+\frac{3}{32}\delta^2+o(\delta^2)\right)\\ &= 1-\frac{1}{16}\delta^2+o(\delta^2)\ .\end{aligned}$$

Aproximando la segunda expresión

$$\begin{aligned}\frac{1}{4}\frac{\varepsilon^2}{(2+\delta)} &= \frac{\varepsilon^2}{8}+o(||(\varepsilon,\delta)||^2)\\ \exp\left(-\frac{1}{4}\frac{\varepsilon^2}{2+\delta}\right) &= 1-\frac{\varepsilon^2}{8}+o(||(\varepsilon,\delta)||^2)\ .\end{aligned}$$

Reemplazando ambas aproximaciones en la última expresión de la distancia de H obtenemos,

$$d_H(\mu,\nu)=\frac{1}{2\sqrt{2}}\left[\frac{\delta^2}{2}+\varepsilon^2+o(||(\varepsilon,\delta)||^2)\right]^{1/2} .$$

Para el caso de la distancia de χ^2, al igual que antes, primero consideremos la distancia entre $\mu_0 \sim N(m_0,v_0^2)$ y $\nu_0 \sim N(0,1)$

$$\begin{aligned}d_{\chi^2}(\mu_0,\nu_0) &= \left[\int \frac{1}{2\pi v_0^2}\exp\left(-\frac{(x-m_0)^2}{v^2}\right)\sqrt{2\pi}\exp\left(\frac{1}{2}x^2\right)dx-1\right]^{1/2}\\ &= \left[\int \frac{1}{\sqrt{2\pi}v_0^2}\exp\left(-\frac{(x-m_0)^2}{v_0^2}+\frac{x^2}{2}\right)dx-1\right]^{1/2}\end{aligned}$$

considerando la siguiente igualdad,

$$2\frac{(x-m_0)^2}{v_0^2}-x^2=\left[x-\frac{m_0}{\sqrt{2-v_0^2}}\right]^2\frac{2-v_0^2}{v^2}-\frac{2m_0^2}{2-v_0^2},$$

la expresión se reduce a

$$d_{\chi^2}(\mu_0,\nu_0) \;=\; \left[\frac{1}{v_0\sqrt{2-v_0^2}}\exp\left(\frac{m_0^2}{2-v_0^2}\right)-1\right]^{1/2}.$$

Sean $\mu\sim N(m_1,v_1^2)$ y $\nu\sim N(m_2,v_2^2)$ con $m_2=m$, $v_2=v$, $m_1=m+v\varepsilon$, $v_1=v\sqrt{1+\delta}$. Al igual que para H podemos escribir la distancia de χ^2 entre μ y ν como $d_{\chi^2}(\mu,\nu)=d_{\chi^2}(N(m,v^2),N(0,1))$. Luego tendremos

$$d_{\chi^2}(\mu,\nu) \;=\; \left[\frac{1}{\sqrt{1-\delta^2}}\exp\left(\frac{\varepsilon^2}{1-\delta}\right)-1\right]^{1/2}.$$

Pasando al producto con la Proposición 2.2.2

$$d_{\chi^2}(\mu^n,\nu^n) \;=\; \left[\left(\frac{1}{1-\delta^2}\right)^{n/2}\exp\left(n\frac{\varepsilon^2}{1-\delta}\right)-1\right]^{1/2}.$$

Utilizando una aproximación $(1+x)^{-1}$ y de $\exp(x)$

$$d_{\chi^2}(\mu,\nu)=\left[\frac{\delta^2}{2}+\varepsilon^2+o(||(\varepsilon,\delta)||^2)\right]^{1/2}.$$

Finalmente consideremos una vez más μ_0 una distribución Normal $N(m_0,v_0^2)$ y ν_0

una $N(0,1)$. De la definición de la distancia de K obtenemos:

$$
\begin{aligned}
d_K(\mu_0,\nu_0) &= \left[\int \log\left(\frac{v\exp\left(-\frac{1}{2}x^2\right)}{\exp\left(-\frac{1}{2}\left[\frac{x-m}{v}\right]^2\right)}\right)\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}dx\right]^{1/2} \\
&= \left[\log(v)-\frac{1}{2}+\frac{1+m^2}{2v^2}\right]^{1/2} .
\end{aligned}
$$

Nuevamente consideramos $m_1 = m + v\varepsilon$, $v_1 = v\sqrt{1+\delta}$, $m_2 = m$ y $v_2 = v$. Al igual que para la distancia de H podemos escribir la distancia entre μ y ν como $d_K(\mu,\nu) = d_K(N(m,v^2), N(0,1))$. Luego,

$$
d_K(\mu,\nu) = \left[\frac{1}{2}\log(1+\delta)+\frac{\varepsilon^2}{2(1+\delta)}-\frac{\delta}{2(1+\delta)}\right]^{1/2} .
$$

Combinando la aproximación de $x(1+x)^{-1}$ y $\log(1+x)$:

$$
\log(1+\delta)-\frac{\delta}{1+\delta}=\frac{\delta^2}{2}+o(\delta^2) ,
$$

y por otro lado de la aproximación de $(1+x)^{-1}$:

$$
\frac{\varepsilon^2}{1+\delta}=\varepsilon^2+o(||(\varepsilon,\delta)||^2) .
$$

De estos dos desarrollos de orden deducimos:

$$
d_K(\mu,\nu) = \left[\frac{\delta^2}{4}+\frac{\varepsilon^2}{2}+o(||(\varepsilon,\delta)||^2)\right]^{1/2} .
$$

De la Proposición 2.3.1 se desprende que la distancia de K entre dos μ^n y ν^n vale:

$$
d_K(\mu^n,\nu^n) = \left[n\left(\frac{1}{2}\log(1+\delta)+\frac{1+\varepsilon^2}{2(1+\delta)}-\frac{1}{2}\right)\right]^{1/2} .
$$

Utilizando la aproximación

$$d_K(\mu^n,\nu^n) = \left[n \left(\frac{\delta^2}{4} + \frac{\varepsilon^2}{2} + o(||(\varepsilon,\delta)||^2) \right) \right]^{1/2} .$$

Usando la segunda caracterización para la distancia en VT de la Proposición 2.1.1 obtenemos la siguiente expresión:

$$\begin{aligned} d_{VT}(\mu,\nu) &= \frac{1}{2}\int \left| \frac{1}{\sqrt{2\pi}\sqrt{1+\delta}} \exp\left(-\frac{(x-\varepsilon)^2}{2(1+\delta)} \right) - \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2} \right) \right| dx \\ &= \frac{1}{2}\mathbb{E}\left[\left| \frac{1}{\sqrt{1+\delta}} \exp\left(-\left(\frac{1}{2(1+\delta)}(X-\varepsilon)^2 - \frac{X^2}{2} \right) \right) - 1 \right| \right] . \end{aligned}$$

Donde X esta distribuida como una $N(0,1)$. Definimos la función f como:

$$f(\varepsilon,\delta,x) = \frac{1}{\sqrt{1+\delta}} \exp\left[-\frac{1}{2}\left(\frac{1}{1+\delta}(x-\varepsilon)^2 - x^2 \right) \right] .$$

Vamos a utilizar un desarrollo de Taylor en ε y δ para obtener una cota. Para esto calculamos sus derivadas de primer y segundo orden:

$$\begin{aligned} \frac{\partial}{\partial\varepsilon} f(\varepsilon,\delta,x) &= f(\varepsilon,\delta,x)\, \frac{x-\varepsilon}{1+\delta} , \\ \frac{\partial}{\partial\delta} f(\varepsilon,\delta,x) &= f(\varepsilon,\delta,x)\frac{1}{2}\left(\left(\frac{x-\varepsilon}{1+\delta} \right)^2 - \frac{1}{1+\delta} \right) , \\ \frac{\partial^2}{\partial\varepsilon^2} f(\varepsilon,\delta,x) &= f(\varepsilon,\delta,x)\left[\left(\frac{x-\varepsilon}{1+\delta} \right)^2 - \frac{1}{1+\delta} \right] , \\ \frac{\partial^2}{\partial\delta^2} f(\varepsilon,\delta,x) &= f(\varepsilon,\delta,x)\left[\frac{1}{4}\left(\frac{(x-\varepsilon)^2}{(1+\delta)^2} - \frac{1}{1+\delta} \right)^2 - \frac{(x-\varepsilon)^2}{(1+\delta)^3} + \frac{1}{2(1+\delta)^2} \right] , \\ \frac{\partial^2}{\partial\varepsilon\partial\delta} f(\varepsilon,\delta,x) &= f(\varepsilon,\delta,x)\left[\frac{1}{2}\left(\frac{(x-\varepsilon)^2}{(1+\delta)^2} - \frac{1}{1+\delta} \right)\left(\frac{x-\varepsilon}{1+\delta} \right) - \frac{x-\varepsilon}{(1+\delta)^2} \right] . \end{aligned}$$

Luego obtenemos la siguiente aproximación de Taylor en torno a $(0,0)$:

$$f(\varepsilon,\delta,x) = 1 + x\varepsilon + \frac{x^2\delta}{2} - \frac{\delta}{2} + \epsilon(\varepsilon,\delta,x) \ ,$$

donde $\epsilon(\varepsilon,\delta,x)$ es el error y se define como

$$\epsilon(\varepsilon,\delta,x) = \int_0^1 \left(\frac{\partial^2}{\partial\varepsilon^2} f(\theta\varepsilon,\theta\delta)\varepsilon^2 + \frac{\partial^2}{\partial\delta^2} f(\theta\varepsilon,\theta\delta)\delta^2 + 2\frac{\partial^2}{\partial\varepsilon\partial\delta} f(\theta\varepsilon,\theta\delta)\varepsilon\delta \right)(1-\theta)d\theta \ .$$

Como $d_{VT}(\mu,\nu) = \frac{1}{2}\mathbb{E}|1 - f(\varepsilon,\delta,X)|$ podemos aproximar la distancia en VT entre μ y ν por $\frac{1}{2}\mathbb{E}\left|X\varepsilon + \frac{X^2\delta}{2} - \frac{\delta}{2}\right|$, el error que cometemos esta acotado por:

$$\left| \frac{1}{2}\mathbb{E}\left|X\varepsilon + \frac{X^2\delta}{2} - \frac{\delta}{2}\right| - d_{VT}(\mu,\nu) \right| \leq \frac{1}{2}\mathbb{E}|\epsilon(\varepsilon,\delta,X)| \ , \tag{2.6}$$

donde X se distribuye como una normal $N(0,1)$. Al igual que en el caso de la distribución de Poisson, demostraremos que $\mathbb{E}|\epsilon(\varepsilon,\delta,X)|$ es despreciable frente a $\frac{1}{2}\mathbb{E}\left|X\varepsilon + \frac{X^2\delta}{2}\right|$; luego, la aproximación será valida. Demostraremos que $\mathbb{E}|\epsilon(\varepsilon,\delta,X)|$ se comporta como $o(||(\varepsilon,\delta)||)$, la última expresión será una buena aproximación de la distancia que estamos estimando. Sean

$$\begin{aligned}
\eta_1(\varepsilon,\delta) &= \mathbb{E}\left[|\frac{\partial^2}{\partial\varepsilon^2} f(\varepsilon,\delta,X)| \right] \ , \\
\eta_2(\varepsilon,\delta) &= \mathbb{E}\left[|\frac{\partial^2}{\partial\delta^2} f(\varepsilon,\delta,X)| \right] , \\
\eta_3(\varepsilon,\delta) &= \mathbb{E}\left[|\frac{\partial^2}{\partial\varepsilon\partial\delta} f(\varepsilon,\delta,X)| \right] \ .
\end{aligned}$$

Luego de intercambiar la esperanza con respecto a X con la integral con respecto a

θ tenemos

$$\mathbb{E}\left[|\epsilon(\varepsilon,\delta,X)|\right] \leq \int_0^1 \left(\eta_1(\theta\varepsilon,\theta\delta)\varepsilon^2 + +\eta_2(\theta\varepsilon,\theta\delta)\delta^2 + 2\eta_3(\theta\varepsilon,\theta\delta)\varepsilon\delta\right)(1-\theta)d\theta \ .$$

Un simple cambio de variables nos permite afirmar que para cualquier función g

$$\mathbb{E}_X\left(f(\varepsilon,\delta,X)g\left(\frac{X-\varepsilon}{\sqrt{1+\delta}}\right)\right) = \mathbb{E}_Y(g(T)) \ .$$

Con X e Y dos variables Normales $N(0,1)$. Utilizando esta propiedad y recordando que X se distribuye como una normal $N(0,1)$, calculamos η_1, η_2 y η_3:

$$\begin{aligned}
\eta_1(\varepsilon,\delta) &= \mathbb{E}\left[f(\varepsilon,\delta,X)\left|\left(\frac{X-\varepsilon}{1+\delta}\right)^2 - \frac{1}{1+\delta}\right|\right] \\
&= \frac{1}{1+\delta}\mathbb{E}\left[\left|X^2-1\right|\right] \\
&= \frac{1}{1+\delta}\left[\frac{4}{\sqrt{2\pi}}e^{-1/2}\right] \\
&= K_1\frac{1}{1+\delta} \ .
\end{aligned}$$

Para η_2 consideremos $x_1 = \sqrt{3-\sqrt{6}}$ y $x_2 = \sqrt{3+\sqrt{6}}$ de modo que $\{-x_2, -x_1, x_1, x_2\}$

son las raíces del polinomio $x^4 + 6x^2 + 3$. Sea $A = [-x_2, -x_1] \bigcup [x_1, x_2]$ obtenemos

$$\begin{aligned}
\eta_2(\varepsilon, \delta) &= I\!\!E\left[f(\varepsilon, \delta) \left| \frac{1}{4}\left(\frac{(x-\varepsilon)^2}{(1+\delta)^2} - \frac{1}{1+\delta} \right)^2 - \frac{(x-\varepsilon)^2}{(1+\delta)^3} + \frac{1}{2(1+\delta)^2} \right| \right] \\
&= \frac{1}{4(1+\delta)^2} I\!\!E\left[\left| \left(X^2 - 1\right)^2 - 4X^2 + 2 \right| \right] \\
&= \frac{1}{2(1+\delta)^2} \left(\int_A (x^4 - 6x^2 + 3) \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \right) \\
&= \frac{1}{(1+\delta)^2} \sqrt{\frac{3}{\pi}} e^{-3/2} \left(\sqrt{3+\sqrt{6}} \exp\left(\frac{\sqrt{6}}{2} \right) - \sqrt{3-\sqrt{6}} \exp\left(\frac{-\sqrt{6}}{2} \right) \right) \\
&= K_2 \frac{1}{(1+\delta)^2} > \frac{4}{3^{\frac{1}{4}} \sqrt{\pi}} e^{-\frac{3}{2}} \ .
\end{aligned}$$

Finalmente acotamos el último término η_3. Sea $x_0 = \sqrt{3}$ luego $\{-x_0, 0, x_0\}$ son las raíces del polinomio $x^3 + 3x$. Definimos el conjunto $B = [|\infty, -x_0] \bigcup [0, x_0]$

$$\begin{aligned}
\eta_3(\varepsilon, \delta) &= I\!\!E\left[f(\varepsilon, \delta, X) \left| \frac{1}{2}\left(\frac{(x-\varepsilon)^2}{(1+\delta)^2} - \frac{1}{1+\delta} \right) \left(\frac{X-\varepsilon}{1+\delta} \right) - \frac{X-\varepsilon}{(1+\delta)^2} \right| \right] \\
&= \frac{1}{2(1+\delta)^{3/2}} I\!\!E\left[\left| X^3 - 3X \right| \right] \\
&= \frac{1}{(1+\delta)^{3/2}} \left(-\int_B (x^3 - 3x) \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \right) \\
&= \frac{1}{(1+\delta)^{3/2}} \frac{1}{\sqrt{2\pi}} \left(1 + 4e^{-3/2} \right) \\
&= K_3 \frac{1}{(1+\delta)^{3/2}} \ .
\end{aligned}$$

Entonces el error queda acotado por

$$I\!\!E\left(|\epsilon(\varepsilon, \delta, X)| \right) \leq \int_0^1 \left(K_1 \frac{1}{1+\theta\delta} \varepsilon^2 + K_2 \frac{1}{(1+\theta\delta)^2} \delta^2 + K_3 \frac{1}{(1+\theta\delta)^{3/2}} \varepsilon\delta \right) (1-\theta) d\theta$$

Estimamos el comportamiento asintótico de cada uno de los tres términos

$$\begin{aligned}
K_1\varepsilon^2\int_0^1 \eta_1(\theta\varepsilon,\theta\delta)(1-\theta)d\theta &= K_1\varepsilon^2\frac{\delta\log(1+\delta)+\log(1+\delta)-\delta}{\delta^2} \\
&= K_1\varepsilon^2\frac{\delta\,(\delta+o(\delta))+\delta-\delta^2/2+o(\delta^2)-\delta}{\delta^2} \\
&= K_1\varepsilon^2\frac{1}{2}(1+o_\delta(1)) \ .
\end{aligned}$$

Un calculo similar para los otros dos términos nos da

$$\begin{aligned}
K_2\delta^2\int_0^1 \eta_2(\theta\varepsilon,\theta\delta)(1-\theta)d\theta &= K_2\delta^2\frac{1}{2}(1+o_\delta(1)) \ , \\
2K_3\varepsilon\delta\int_0^1 \eta_3(\theta\varepsilon,\theta\delta)(1-\theta)d\theta &= K_3\varepsilon\delta(1+o_\delta(1)) \ .
\end{aligned}$$

La combinación de los tres últimos términos nos permite afirmar que $\mathbb{E}|\epsilon(\varepsilon,\delta,X)|$ es $o(||(\varepsilon,\delta)||^2)$:

$$\mathbb{E}|\epsilon(\varepsilon,\delta,X)| \leq \left(K_1\varepsilon^2+K_2\delta^2+2K_3\varepsilon\delta\right)\frac{1}{2}(1+o_\delta(1)) = o(||(\varepsilon,\delta)||^2) \ .$$

Luego, de la desigualdad (2.6) podemos concluir que $d_{VT}(\mu,\nu)$ se puede aproximar para δ y ε pequeños por

$$d_{VT}(\mu,\nu) = \frac{1}{2}\mathbb{E}\left|X\varepsilon+\frac{X^2\delta}{2}-\frac{\delta}{2}\right| + o(||(\varepsilon,\delta)||)$$

Para finalizar debemos estimar $\mathbb{E}\left|X\varepsilon+\frac{X^2\delta}{2}-\frac{\delta}{2}\right|$. Recordemos que X se distribuye como una $N(0,1)$. Sean x_1 y x_2 las raíces de $x^2 + \frac{2\varepsilon}{\delta}x - 1$ con $x_1 < x_2$.

$$\begin{aligned}
I\!E\left(\left|X\varepsilon + \frac{X^2\delta}{2} - \frac{\delta}{2}\right|\right) &= |\delta|\, I\!E\left(|X^2 + \tfrac{2\varepsilon}{\delta}X - 1|_-\right) \\
&= \frac{|\delta|}{\sqrt{2\pi}}\left((x_2 + \tfrac{2\varepsilon}{\delta})e^{-x_2^2/2} - (x_1 + \tfrac{2\varepsilon}{\delta})e^{-x_1^2/2}\right) .
\end{aligned}$$

Sea $\tilde{z} = \varepsilon/\delta$ entonces las raíces del polinomio valen $x_1 = -\tilde{z} - \sqrt{\tilde{z}^2+1}$ y $x_2 = -\tilde{z} + \sqrt{\tilde{z}^2+1}$. Reemplazando en la expresión se tiene

$$\begin{aligned}
I\!E\left(\left|X\varepsilon + \frac{X^2\delta}{2} - \frac{\delta}{2}\right|\right) &= \frac{|\delta|}{\sqrt{2\pi}}\Big[\left(\tilde{z} + \sqrt{\tilde{z}^2+1}\right)\exp\left(-\frac{1}{2}(\sqrt{\tilde{z}^2+1} - \tilde{z})^2\right) \\
&\quad + \left(\sqrt{\tilde{z}^2+1} - \tilde{z}\right)\exp\left(-\frac{1}{2}(\sqrt{\tilde{z}^2+1} + \tilde{z})^2\right)\Big] \\
&= \frac{|\delta|}{\sqrt{2\pi}}\Big[\frac{1}{\sqrt{\tilde{z}^2+1} + |\tilde{z}|}\exp\left(-\frac{1}{2}(\sqrt{\tilde{z}^2+1} + |\tilde{z}|)^2\right) \\
&\quad + \left(\sqrt{\tilde{z}^2+1} + |\tilde{z}|\right)\exp\left(-\frac{1}{2}\frac{1}{(\sqrt{\tilde{z}^2+1} + |\tilde{z}|)^2}\right)\Big] .
\end{aligned}$$

Sea $z = |\varepsilon/\delta|$ y $y(z) = \sqrt{z^2+1} + z$. Luego $d_{VT}(\mu,\nu)$ se puede aproximar para δ y ε pequeños por

$$d_{VT}(\mu,\nu) = \frac{|\delta|}{2\sqrt{2\pi}}\left[\frac{1}{y(z)}\exp\left(-\frac{1}{2}(y(z))^2\right) + y(z)\exp\left(-\frac{1}{2}(y(z))^{-2}\right)\right] + o(||(\varepsilon,\delta)||) ,$$

Nos gustaría obtener una aproximación más sencilla para poder obtener una cota inferior de $d_{VT}(\mu^n,\nu^n)$ utilizaremos la desigualdad de la Proposición 2.2.2 y la relación

$y(z) - 1 = z(y(-z^{-1}) + 1)$:

$$\begin{aligned} d_{VT}(\mu,\nu) &= \frac{|\delta|}{\sqrt{8\pi}}\left[\frac{1}{y(z)}\exp\left(-\frac{1}{2}(y(z))^2\right) + \exp\left(-\frac{1}{2}(y(z))^{-2}\right)\right] \\ &\quad + \frac{|\varepsilon|}{\sqrt{8\pi}}(y(-z^{-1})+1)\exp\left(-\frac{1}{2}(y(z))^{-2}\right) + o(||(\varepsilon,\delta)||)\ . \end{aligned}$$

Definimos $C_1(z)$ (respectivamente $C_2(z)$) como el coeficiente de $|\delta|$ (respectivamente $|\varepsilon|$).

$$\begin{aligned} C_1(z) &= \frac{1}{y(z)}\exp\left(-\frac{1}{2}(y(z))^2\right) + \exp\left(-\frac{1}{2}(y(z))^{-2}\right) \\ C_2(z) &= (y(-z^{-1})+1)\exp\left(-\frac{1}{2}(y(z))^{-2}\right)\ . \end{aligned}$$

En consecuencia la distancia en VT entre μ y ν vale

$$d_{VT}(\mu,\nu) = \frac{|\delta|}{\sqrt{8\pi}}C_1(z) + \frac{|\varepsilon|}{\sqrt{8\pi}}C_2(z) + o(||(\varepsilon,\delta)||)\ ,$$

donde el comportamiento de C_1 y C_2 esta dado por

$$\begin{aligned} C_1(0) = 2e^{-1/2} \quad &; C_1(\infty) = 1 \quad ; e^{-1/2} \le C_1(z) \le 1 + e^{-1/2}\ , \\ C_2(0) = e^{-1/2} \quad &; C_2(\infty) = 2 \quad ; e^{-1/2} \le C_2(z) \le 2\ . \end{aligned}$$

Finalmente podemos acotar $d_{VT}(\mu^n,\nu^n)$ por

$$1 - 2\exp\left(\frac{n}{16\pi}\left(C_1(z)|\delta| + C_2(z)|\varepsilon| + o(||(\varepsilon,\delta)||)\right)^2\right) \le d_{VT}(\mu^n,\nu^n)$$

$$d_{VT}(\mu^n,\nu^n) \le \left[1 - \left(\frac{\sqrt{1+\delta}}{1+\delta/2}\right)^n \exp\left(-\frac{n}{2}\frac{\varepsilon^2}{(2+\delta)}\right)\right]^{1/2}\ .$$

$\square$

Observación 2.4.7. *En el caso de la distribución Normal vemos el mismo comportamiento lineal que en los casos de Bernoulli y de Poisson para las distancias de H, χ^2 y K y para la distancia en VT vemos que esta acotada entre dos funciones lineales. La distancia entre μ y ν se comporta de manera lineal cuando ambos parámetros ε y δ son pequeños:*

$$
\begin{aligned}
d_{VT}(\mu,\nu) &= \frac{|\delta|}{\sqrt{8\pi}}C_1(z) + \frac{|\varepsilon|}{\sqrt{8\pi}}C_2(z) + o(||(\varepsilon,\delta)||)\ , \\
d_H(\mu,\nu) &= \frac{1}{2\sqrt{2}}\left[\frac{\delta^2}{2} + \varepsilon^2 + o(||(\varepsilon,\delta)||^2)\right]^{1/2}\ , \\
d_{\chi^2}(\mu,\nu) &= \left[\frac{\delta^2}{2} + \varepsilon^2 + o(||(\varepsilon,\delta)||^2)\right]^{1/2}\ , \\
d_K(\mu,\nu) &= \left[\frac{\delta^2}{4} + \frac{\varepsilon^2}{2} + o(||(\varepsilon,\delta)||^2)\right]^{1/2}\ .
\end{aligned}
$$

Donde $z = |\varepsilon/\delta|$ y C_1 y C_2 son dos funciones positivas acotadas.

Cutoff para n-tuplas de procesos exponencialmente convergentes

El objetivo de este capítulo es estudiar el fenómeno de Cutoff para n-tuplas de procesos independientes, tanto idénticamente distribuidos como no. Nuestro resultado principal, el Teorema 3.3.1, da condiciones bajo las cuales una n-tupla tiene un Cutoff en el sentido de las distancias de VT, H, χ^2 y K. Para cada $i = 1, ..., n$ asumimos que la i-ésima coordenada converge a una velocidad exponencial de tasa ρ_i a su medida de equilibrio o, de un modo más preciso, el logaritmo de la distancia al tiempo t es equivalente a $-\rho_i t$ (Definición 3.2.1). Se prueba que bajo condiciones adecuadas sobre las tasas de convergencias (ρ_i) la n-tupla tiene un Cutoff al tiempo:

$$t_n = \text{máx}\left\{ \frac{\log i}{2\rho_{(i,n)}} \, ; \; i = 1, \ldots, n \right\} ,$$

donde $\rho_{(1,n)}, \ldots, \rho_{(n,n)}$ son los valores de $\rho_1, \ldots, \rho_n$ ordenados de manera creciente. Resultados más precisos se pueden probar para coordenadas i.i.d. (Teoremas 3.3.4 y 3.3.5): si ρ es la tasa de convergencia exponencial común, no solo hay Cutoff para la n-tupla al tiempo $\log n/(2\rho)$, si no que además para u fijo estimaciones precisas para la distancia al equilibrio al tiempo $\log n/(2\rho) + u$ se pueden hacer. Más precisamente, se puede mostrar que el Cutoff ocurre en un intervalo de tiempo de largo $O(1)$ en torno a $\log n/(2\rho)$. Las demostraciones se basan en desigualdades que acotan la distancia entre dos medidas producto en función de la distancias entre las marginales (Proposición 2.3.1). La mayoría son consecuencia directa de desigualdades clásicas que se encuentran en libros de estadística (e.g. [51], [38]) y que hemos recopilado en el Capítulo 2.

En este capítulo se abordan los resultados del artículo "Cutoff for n-tuples of exponentially converging process" ([6]). En la primera sección daremos un ejemplo sencillo donde se observa el fenómeno, luego daremos la definición formal que utilizaremos y discutiremos los resultados particulares previos que se encuentran en la literatura para Cutoff de n-tuplas. En la Sección 3.2 discutimos la noción de convergencia exponencial que usaremos y el comportamiento de la sumatoria de distancias exponencialmente convergentes (Lema 3.2.4). En la Sección 3.3 se encuentra nuestro resultado principal de Cutoff: el Teorema 3.3.1 para n-tuplas de procesos exponencialmente convergentes independientes pero no idénticamente distribuidos así como otros resultados para el caso i.i.d.. En la Sección 3.4 se estudian n-tuplas de procesos i.i.d. para los casos en que las coordenadas corresponde a Procesos de Markov Binarios, procesos de Nacimiento y muerte M/M/∞ y procesos de difusión de Ornstein-Uhlenbeck.

3.1. Definición y algunos ejemplos clásicos

Primero explicaremos el Cutoff con el clásico ejemplo de revolver cartas de un maso. Luego propondremos una definición formal del fenómeno y citaremos algunos trabajos de Cutoff de la literatura y veremos como se relacionan con nuestro trabajo.

3.1.1. ¿Qué es el Cutoff?

El fenómeno de Cutoff es la formalización de la convergencia abrupta al equilibrio que se observa en algunas cadenas de Markov. El Cutoff fue identificado por primera vez por Aldous y Diaconis en [1] para algunos paseos aleatorios en grupos de permutaciones. Después el fenómeno ha sido observado en muchas otras cadenas de Markov.

Intuitivamente el fenómeno se puede entender con un ejemplo cotidiano donde se observa el Cutoff: barajando un mazo de 52 cartas. Una de las maneras de barajar un mazo es realizando el siguiente proceso, se divide el mazo en dos grupos y se mezclan en uno dejando caer aleatoriamente una carta del primer o segundo mazo con una probabilidad proporcional al numero de cartas que quedan sin mezclar en cada grupo. Si el proceso se realiza 7 veces todas las posibles permutaciones de las 52 cartas serán aproximadamente igualmente probables. Si el proceso se realiza 5 veces el mazo no estará suficientemente barajado y se podrá reconocer ciertos patrones del orden original. Este proceso es una cadena de Markov cuyos estados son las permutaciones de $\{1,\dots,52\}$ y cuya ley estacionaria es una distribución uniforme. El hecho que barajar 5 veces el mazo no es suficiente y que hacerlo 7 veces si lo es, es la característica del fenómeno de convergencia abrupta: Cutoff. En general tendremos que para un mazo de n cartas se necesita barajar al menos $3\log_2 n/2$ veces para obtener un mazo bien mezclado. Este comportamiento se ha observado no solo en esta cadena, sino que en varios otros procesos de Markov ver, por ejemplo, los trabajos [16, 54, 42, 63, 55, 49]. Una completa revisión de Cutoff en paseos aleatorios en grupos es realizada por Saloff-Coste en [56]. En particular se puede encontrar la descripción formal de diferentes modelos para barajar cartas.

3.1.2. Definición y un poco de historia

La definición clásica de Cutoff esta ligada a la distancia en VT (ecuación 2.1). Diremos que un fenómeno de Cutoff ocurre cuando antes de cierto "tiempo de Cutoff" la cadena esta lejos del equilibrio, es decir que la distancia en VT entre la ley del proceso al tiempo t y la medida de equilibrio es cercana a 1 y después de este instante la distancia decae exponencialmente a 0. Aparte de la distancia en VT existen otras distancias para caracterizar la convergencia, en el Capítulo 2 revisamos aquellas que son relevantes para este trabajo. Utilizaremos la siguiente definición de Cutoff la cual es independiente de la distancia utilizada entre distribuciones de probabilidad.

Definición 3.1.1. *Para cada $n \geq 0$, sea $E^{(n)}$ un espacio medible, sea $X^{(n)} = \{X^{(n)}(t)\,;\ t \geq 0\}$ un proceso estocástico en $E^{(n)}$, convergente en distribución a una distribución ley de probabilidad $\nu^{(n)}$. Sea d una distancia entre distribuciones de probabilidad. Para $t \geq 0$, sea $d^{(n)}(t)$ la distancia entre la distribución de $X^{(n)}(t)$ y $\nu^{(n)}$:*

$$d^{(n)}(t) = d(\mathcal{L}X^{(n)}(t), \nu^{(n)}) \ .$$

Sea (t_n) una sucesión de reales positivos. Se dirá que la sucesión de procesos $(X^{(n)})$ tiene un Cutoff al tiempo (t_n) en el sentido de la distancia d si para $c > 0$:

$$\begin{aligned} c < 1 &\implies \lim_{n\to\infty} d^{(n)}(ct_n) = M \ , \\ c > 1 &\implies \lim_{n\to\infty} d^{(n)}(ct_n) = 0 \ , \end{aligned}$$

donde M es el máximo valor que puede tomar la distancia escogida.

La definición clásica se diferencia de esta en que utiliza la distancia en VT y el por ende debe ser igual a 1. Nuestra definición no es el primer intento por extender la definición de Cutoff a otras distancias, Saloff-Coste en [56] introduce la noción de L^p-cutoff. En el mismo artículo, el autor introduce la noción de Precutoff como un medio para capturar el orden de magnitud de un posible Cutoff.

Casos particulares de los Teoremas 3.3.1 y 3.3.4 se pueden encontrar en la literatura. Uno de los primeros ejemplos de Cutoff fue estudiado por Diaconis y Shahshahani en [18] y por Diaconis *et al.* en [17] para los paseos aleatorios en el hipercubo. Como lo hizo notar Ycart en [63] p. 91, ese paseo aleatorio puede interpretarse como una versión a tiempo discreto de una n-tupla de cadenas de Markov Binarias i.i.d. a tiempo continuo. El caso más general de n-tuplas de cadenas de Markov reversibles i.i.d. en espacios de estados finitos fue estudiado por Ycart en [63] y sus aplicaciones a tiempos de parada para métodos MCMC fueron descritas por el mismo autor en [64]. Bon y Păltănea consideran en [12] el caso de procesos de Markov Binarios a tiempo continuo e independientes, pero no necesariamente idénticamente distribuidos, en el contexto de teoría de la fiabilidad. Las demostraciones originales de Cutoff para el paseo aleatorio en [18] y [17], así como para n-tuplas de procesos de Markov reversibles en [63], se basan en el análisis espectral de la matriz de transición. En este trabajo decidimos abordar el problema relacionando el fenómeno de Cutoff al modo en que las distancias dan cuenta de la concentración de las medidas producto. Una comparación en el uso de las diferentes distancias para medir Cutoffs de paseos aleatorios puede encontrarse en [58].

3.2. Procesos exponencialmente convergentes

Primero precisaremos la noción de convergencia exponencial que usaremos y veremos como se relaciona con otras definiciones.

Si d es una distancia entre distribuciones de probabilidad, llamaremos $d(t)$ a la distancia entre un proceso X al tiempo t y su limite ν.

$$d(t) = d(\mathcal{L}X(t), \nu) \ .$$

Para asegurarnos que la distancia de χ^2 está bien definida asumiremos de ahora en adelante que $\mathfrak{L}X(t)$ es absolutamente continua con respecto a ν para todo $t > 0$.

Definición 3.2.1. *Sea $X = \{X(t)\,;\ t \geq 0\}$ un proceso estocástico, ν una distribución de probabilidad y ρ un real positivo. Diremos que el proceso X converge a ν a tasa exponencial ρ según la distancia d si*

$$\lim_{t \to +\infty} \frac{\log d(t)}{t} = -\rho \ . \tag{3.1}$$

Esta definición es coherente con la usualmente utilizada para la tasa de convergencia exponencial, por ejemplo para los procesos de nacimiento y muerte (ver e.g. [61]). A veces se entiende por comportamiento exponencial a convergencias del tipo $d(t) \sim Re^{-\rho t}$. Podemos decir que nuestra definición es más general en el sentido en que toma en cuenta aquellos procesos para los cuales la distancia decae como $d(t) \sim R(t)e^{-\rho t}$, donde $R(t)$ tiene un crecimiento subexponencial.

Las diferentes desigualdades de la Proposición 2.2.2 llevan a implicaciones obvias entre la convergencia exponencial en una u otra distancia. En general, estas implicaciones no son equivalencias. Los casos de las distancias en VT y de χ^2 han sido ampliamente estudiados en el contexto de procesos de Markov. Sin embargo, las definiciones clásicas difieren de la Definición 3.2.1. Las definiciones de Ergodicidad exponencial y de L^2-ergodicidad exponencial que siguen fueron tomadas de [14] (ver Cap. 4 pág. 144 para Ergodicidad exponencial y Cap. 9 p. 311 para L^2-Ergodicidad exponencial).

Definición 3.2.2. *Con la notación anterior, X es* exponencialmente ergódico *si existe dos constantes positivas R y ρ tales que para $t > 0$,*

$$d_{TV}(t) \leq Re^{-\rho t} \ .$$

Definición 3.2.3. *Sea X un proceso de Markov y $\{P(t)\,, t \geq 0\}$ su semigrupo. El proceso X tiene una* L^2-convergencia exponencial *si existe un real positivo ρ, tal que para todo f en $L^2(\nu)$*

$$\|P(t)f - \nu(f)\|_{L^2(\nu)} \leq R_f e^{-\rho t} \ ,$$

para una constante positiva R_f.

Precisemos como se relaciona la definición de L^2-convergencia exponencial a nuestra definición de convergencia exponencial en el sentido de la distancia de χ^2. Asumamos que ν es *reversible* con respecto a la acción del semigrupo. Luego la acción de $P(t)$ en una medida con signo μ en $L^2(\nu)$ es equivalente a la acción de $P(t)$ en f en $L^2(\nu)$. Más aún, sea μ la distribución de $X(0)$ y $f = \frac{d\mu}{d\nu}$ luego para todo $t > 0$,

$$d_{\chi^2}(t) = \|P(t)f - \nu(f)\|_{L^2(\nu)} \ .$$

La comparación entre convergencia exponencial y $L^2(\nu)$-convergencia exponencial ha sido investigada por varios autores entre ellos Rosenthal, Roberts y Tweedie en [52] y [53]. En el Teorema 2.1 de [52] y Teorema 2 de [53] los autores demuestran que las dos nociones son equivalentes para una cadena de Markov ergódica a tiempo discreto, si la medida invariante ν es reversible y si el espacio de estados ha sido proveído de una σ-álgebra generada de manera numerable. En [13] Chen extiende los resultados para cadenas de Markov en tiempo continuo. Si el espacio de estados es numerable, el autor reemplaza la reversibilidad por una condición más débil. Como es sugerido por todos estos autores, el rol de la reversibilidad en esos resultados sería esencialmente técnico.

Al menos en el caso markoviano, cuando el proceso converge a tasa exponencial para una de las cuatro distancias consideradas aquí, pensamos que es razonable esperar que convergerá a la misma tasa exponencial para las otras tres distancias. Este es evidentemente el caso para una cadena de Markov irreducible (a tiempo continuo) en un espacio de estados finito, donde la tasa de convergencia común ρ será el gap, i.e. el más pequeño en valor absoluto de los valores propios no nulos del generador infinitesimal (ver e.g. Sección 2.1 en [41]). También es el caso para el proceso de nacimiento y muerte M/M/∞ y para el proceso de Ornstein-Uhlenbeck (ver Sección 3.4).

Sea (X_i) una sucesión de procesos independientes exponencialmente convergentes. Notemos por $d_i(t)$ la distancia al equilibrio de la i-ésima coordenada al tiempo t. Asumiremos que todos los procesos son independientes entre si. Consideremos la n-tupla $X^{(n)} = (X_1, \dots, X_n)$, y notemos por $d^{(n)}(t)$ su distancia al equilibrio al tiempo t. Como ya se comentó en la Proposición 2.3.1, cualquiera sea la distancia que se está considerando, $d^{(n)}(t)$ se comporta esencialmente como $\sum_{i=1}^{n} d_i^2(t)$. El siguiente lema técnico relaciona esta sumatoria a la tasa de convergencia exponencial, y será la clave para obtener el Teorema 3.3.1.

Lema 3.2.4. *[6] Para $i = 1, 2, \ldots$, sea d_i una función positiva definida en $\mathbb{R}^+$, y ρ_i un real positivo. Para $n \geq 1$, notemos por $\rho_{(1,n)}, \ldots, \rho_{(n,n)}$ los valores de $\rho_1, \ldots, \rho_n$ ordenados de manera creciente, y llamemos τ_n al siguiente real:*

$$\tau_n = \max \left\{ \frac{\log i}{\rho_{(i,n)}} \, , \; i = 1, \ldots, n \right\} \, . \tag{3.2}$$

Asumamos que se satisfacen la siguientes hipótesis.

1. *Existe una función decreciente positiva g, convergente a 0 cuando t tiende a infinito, y un real positivo t_0 tal que para todo $t \geq t_0$ y para todo $i \geq 1$,*

$$\left| \frac{\log d_i(t)}{t} + \rho_i \right| \leq g(t) \, . \tag{3.3}$$

2. $$\lim_{n \to \infty} \rho_{(1,n)} \, \tau_n = +\infty \, . \tag{3.4}$$

3. *Para cualquier real positivo c,*

$$\lim_{n \to \infty} \frac{g(c\tau_n)}{\rho_{(1,n)}} = 0 \, . \tag{3.5}$$

Luego para cualquier entero positivo k, cualquier real positivo c y cualquier secuencia

(τ_n') *tal que* $\lim \tau_n'/\tau_n = 1$,

$$c < 1 \implies \lim_{n\to\infty} \sum_{i=1}^{n} d_i \left(c\frac{\tau_n'}{k} \right)^k = +\infty \ ,$$
$$c > 1 \implies \lim_{n\to\infty} \sum_{i=1}^{n} d_i \left(c\frac{\tau_n'}{k} \right)^k = 0 \ .$$

De las tres hipótesis, la primera es obviamente la más importante. Nos dice que no basta con que los $d_i(t)$ converjan a cero a tasa exponencial ρ_i, sino que además lo deben hacer con cierta uniformidad en i. Las otras hipótesis no son difíciles de satisfacer.

Observación 3.2.5. *La segunda y tercera hipótesis involucran $\rho_{(1,n)}$, el cual se define como el mínimo entre $\rho_1, \ldots, \rho_n$. Si la sucesión (ρ_i) no diverge a $+\infty$, $\rho_{(1,n)}$ estará acotado y por ende τ_n tenderá a infinito luego la primera hipótesis será satisfecha. Si (ρ_i) esta acotado inferiormente por un valor mayor estricto que 0 y no diverge a infinito, luego ambas hipótesis (3.4) y (3.5) son satisfechas. Sin embargo, puede ocurrir que una subsucesión de (ρ_i) tienda a 0 en ese caso es posible que (τ_n) tienda a $+\infty$, y $g(c\tau_n)$ tiende a 0 suficientemente rápido de modo que todavía se pueden satisfacer las hipótesis para tener Cutoff.*

Demostración: Demostraremos primero el resultado para τ_n. Llamemos

$$g_i(t) = \frac{\log d_i(t)}{t} + \rho_i \ .$$

Luego:
$$\sum_{i=1}^{n} d_i \left(c\frac{\tau_n}{k}\right)^k = \sum_{i=1}^{n} \exp\Big(-\rho_i c\tau_n + c\tau_n g_i\left(c\frac{\tau_n}{k}\right)\Big) \ .$$
Utilizando (3.3), los valores g_i están uniformemente acotados:

$$\forall t \geq t_0, \quad \forall i \ , \quad |g_i(t)| \leq g(t) \ .$$

Entonces para n suficientemente grande:

$$S_n \exp\Big(-c\tau_n g\left(c\frac{\tau_n}{k}\right)\Big) \leq \sum_{i=1}^{n} d_i \left(c\frac{\tau_n}{k}\right)^k \leq S_n \exp\Big(c\tau_n g\left(c\frac{\tau_n}{k}\right)\Big) \ ,$$

con
$$S_n = \sum_{i=1}^{n} \exp(-\rho_i c\tau_n) = \sum_{i=1}^{n} \exp(-\rho_{(i,n)} c\tau_n) \ .$$

Primero demostraremos la cota superior, para $c > 1$. Observemos que para todo $i = 1, \ldots, n$, $\exp(-\rho_{(i,n)} c\tau_n) \leq i^{-c}$, pues $\tau_n \geq \frac{\log i}{\rho_{(i,n)}}$. Para todo $l = 1, \ldots, n-1$, se puede escribir:

$$\begin{aligned} S_n &\leq l e^{-\rho_{(1,n)} c\tau_n} + \sum_{i=l+1}^{n} i^{-c} \\ &\leq l e^{-\rho_{(1,n)} c\tau_n} + \int_l^n x^{-c}\, dx \\ &= l e^{-\rho_{(1,n)} c\tau_n} + \frac{1}{c-1}\left(l^{-(c-1)} - n^{-(c-1)}\right) \ . \end{aligned}$$

Esta desigualdad también es válida para $l = n$. Definamos $l_n = \lfloor e^{\rho_{(1,n)}\tau_n} \rfloor$, donde $\lfloor \cdot \rfloor$ denota la parte integral; tenemos que l_n es menor o igual a n, como consecuencia de la definición de τ_n. Tenemos:

$$
\begin{aligned}
S_n &\leq l_n e^{-\rho_{(1,n)} c\tau_n} + \frac{l_n^{-(c-1)}}{c-1} \\
&\leq e^{-\rho_{(1,n)}(c-1)\tau_n} + \frac{(e^{\rho_{(1,n)}\tau_n}-1)^{-(c-1)}}{c-1} \\
&= e^{-\rho_{(1,n)}(c-1)\tau_n}\left(1+\frac{(1-e^{-\rho_{(1,n)}\tau_n})^{-(c-1)}}{c-1}\right) .
\end{aligned}
$$

Entonces:

$$
\begin{aligned}
&\sum_{i=1}^{n} d_i \left(c\frac{\tau_n}{k}\right)^k \\
&\leq e^{-\rho_{(1,n)}(c-1)\tau_n}\left(1+\frac{(1-e^{-\rho_{(1,n)}\tau_n})^{-(c-1)}}{c-1}\right)\exp\left(c\tau_n g\left(c\frac{\tau_n}{k}\right)\right) \\
&= \left(1+\frac{(1-e^{-\rho_{(1,n)}\tau_n})^{-(c-1)}}{c-1}\right)\exp\left(-\rho_{(1,n)}\tau_n\left((c-1)-c\frac{g(\frac{c\tau_n}{k})}{\rho_{(1,n)}}\right)\right) ,
\end{aligned}
$$

el cual tiende a 0 cuando n tiende a infinito, usando las hipótesis (3.4) y (3.5).

Demostremos ahora la cota inferior para $0 < c < 1$. Para cada n, escogemos i_n^* tal que $\tau_n = \log i_n^* / \rho_{(i_n^*,n)}$, i.e. $i_n^* = \exp(\tau_n \rho_{(i_n^*,n)}) \geq \exp(\tau_n \rho_{(1,n)})$. Se obtiene:

$$
\begin{aligned}
S_n &\geq \sum_{i=1}^{i_n^*} \exp(-c\rho_{(i,n)}\tau_n) \\
&\geq \exp((1-c)\rho_{(i_n^*,n)}\tau_n) \\
&\geq \exp((1-c)\rho_{(1,n)}\tau_n) .
\end{aligned}
$$

Luego:

$$\sum_{i=1}^{n} d_i \left(c\frac{\tau_n}{k}\right)^k \geq \exp\left(\rho_{(1,n)}\tau_n \left((1-c) - c\frac{g(\frac{c\tau_n}{k})}{\rho_{(1,n)}}\right)\right) \ ,$$

el cual tiende a $+\infty$ cuando n tiende a infinito, usando (3.4) y (3.5).

Consideremos ahora otra sucesión (τ'_n), equivalente a (τ_n). La nueva suma puede ser acotada como antes por:

$$S'_n \exp\left(-c\tau'_n g\left(c\frac{\tau'_n}{k}\right)\right) \leq \sum_{i=1}^{n} d_i \left(c\frac{\tau'_n}{k}\right)^k \leq S'_n \exp\left(c\tau'_n g\left(c\frac{\tau'_n}{k}\right)\right) \ ,$$

con

$$S'_n = \sum_{i=1}^{n} \exp(-\rho_i c\tau'_n) \ .$$

Veamos el caso de la cota superior. Fijemos $c' < 1$ tal que $cc' > 1$. Para n suficientemente grande, $c' \leq \tau'_n/\tau_n \leq 1/c'$. Entonces tenemos que:

$$S'_n \leq \sum_{i=1}^{n} \exp(-\rho_i cc'\tau_n) \ ,$$

y

$$\exp\left(c\tau'_n g\left(c\frac{\tau'_n}{k}\right)\right) \leq \exp\left((c/c')\tau_n g\left(cc'\frac{\tau_n}{k}\right)\right) \ ,$$

pues g es decreciente. La cota superior de S_n se puede aplicar a S'_n, reemplazando c por cc'. Se obtiene:

$$\begin{aligned}
&\sum_{i=1}^{n} d_i \left(c\frac{\tau'_n}{k}\right)^k \\
&\leq e^{-\rho_{(1,n)}(cc'-1)\tau_n}\left(1+\frac{(1-e^{-\rho_{(1,n)}\tau_n})^{-(cc'-1)}}{cc'-1}\right)\exp\left((c/c')\tau_n g\left(cc'\frac{\tau_n}{k}\right)\right) \\
&= \left(1+\frac{(1-e^{-\rho_{(1,n)}\tau_n})^{-(cc'-1)}}{cc'-1}\right)\exp\left(-\rho_{(1,n)}\tau_n\left((cc'-1)-(c/c')\frac{g(\frac{cc'\tau_n}{k})}{\rho_{(1,n)}}\right)\right) \ ,
\end{aligned}$$

el cual tiende a 0 cuando n tiende a infinito. Para la cota inferior, la demostración es análoga y será omitida. □

3.3. Tiempos de Cutoff

Sea $(X_i)_{i\in\mathbb{N}}$ una secuencia de procesos independientes, cada uno convergiendo a tasa exponencial a su distribución estacionaria de acuerdo a una distancia d (Definición 3.2.1). Notemos por $d_i(t)$ la distancia al equilibrio de X_i al tiempo t y por ρ_i la tasa exponencial de convergencia. Para $n \geq 1$, consideremos la n-tupla $X^{(n)} = (X_1, \ldots, X_n)$. En vista de la Proposición 2.3.1 y del Lema 3.2.4, es natural para la secuencia $(X^{(n)})$ tener una tiempo de Cutoff en $\tau_n/2$, cuando τ_n está definida por (3.2). Primero lo probaremos, y luego lo ilustraremos en el caso particular de procesos binarios. Además presentaremos otros tiempos de Cutoff. Finalmente, para el caso i.i.d. se darán resultados más precisos.

Teorema 3.3.1. *[6]*

1. *Sea d la distancia de H, de χ^2 o de K. Suponemos que $d_i(t)$ y ρ_i satisfacen las hipótesis* (3.3), (3.4) *y* (3.5) *del Lema 3.2.4. Entonces la sucesión de procesos $(X^{(n)})$ tiene un Cutoff de acuerdo a la distancia d al tiempo*

$$t_n = \text{máx}\left\{\frac{\log i}{2\rho_{(i,n)}}\,,\ i = 1, \ldots, n\right\}\,, \tag{3.6}$$

donde $\rho_{(1,n)}, \ldots, \rho_{(n,n)}$ son los valores $\rho_1, \ldots, \rho_n$ ordenados de manera creciente.

2. Si hay un Cutoff al tiempo t_n para la distancia de H entonces hay Cutoff al mismo instante para la distancia en VT.

3. Suponiendo que cada proceso X_i tiene la misma tasa de convergencia exponencial para la distancia en VT y la distancia de χ^2 o bien la de K. Entonces hay Cutoff al tiempo t_n de acuerdo a la distancia en VT si (3.3), (3.4) *y* (3.5) *se satisfacen para ambas distancias.*

Demostración: Usando el Lema 3.2.4, la suma $\sum_{i=1}^{n} d_i^2(ct_n)$ tiende a $+\infty$ para $0 < c < 1$, a 0 para $c > 1$. El resultado se obtiene para la distancia H, χ^2 y K usando (2.3), (2.4) y (2.5) de la Proposición 2.3.1. Si d es la distancia VT, y las hipótesis del Lema 3.2.4 se satisfacen, entonces de (2.2), sólo podemos deducir que la distancia al equilibrio al tiempo ct_n tiende a 1 para $0 < c < 1$, a 0 para $c > 2$. Pero de las dos primeras desigualdades de la Proposición 2.2.2, si un Cutoff ocurre para la distancia de Hellinger, entonces también ocurre para la distancia en VT. Asumiendo que el Lema 3.2.4 se puede aplicar a las distancias de VT y χ^2, con las mismas tasas de convergencia $\{\rho_i\}$, entonces para $0 < c < 1$ la distancia en VT al equilibrio tiende a 1 al tiempo ct_n, usando la cota inferior de (2.2). Tiende a 0 para $c > 1$, usando la tercera desigualdad de la Proposición 2.2.2. El mismo argumento se puede aplicar si uno utiliza la distancia de K en vez de la distancia χ^2, usando la cuarta desigualdad de la Proposición 2.2.2. □

Observemos que las conclusiones del teorema 3.3.1 se satisfacen si t_n es reemplazada por t'_n cuando $\lim t'_n/t_n = 1$: esta conclusión se obtiene directamente del Lema 3.2.4.

Para ilustrar el Teorema 3.3.1, consideremos la sucesión de procesos de Markov de salto binarios independientes. Para $i \geq 1$, sean α_i y ρ_i dos reales positivos tales que $0 < \alpha_i < \rho_i$. El proceso X_i toma sus valores en $\{0, 1\}$. Salta de 0 a 1 con tasa α_i y de 1 a 0 con tasa $\rho_i - \alpha_i$. Sabemos que la distribución de $X_i(t)$ partiendo de 0 al tiempo 0 es una Bernoulli (ver e.g. Sección 7.5 en [10]) de parámetro:

$$p_i(t) = \frac{\alpha_i}{\rho_i}\left(1 - e^{-\rho_i t}\right) .$$

La distancia entre la ley de $X_i(t)$ y la medida estacionaria esta dada por (ver Observación 2.4.2):

$$\begin{array}{lll} \text{Variación total:} & d_i(t) & = \frac{\alpha_i}{\rho_i}\, e^{-\rho_i t} \\ \text{Hellinger:} & d_i(t) & = \sqrt{\frac{\alpha_i}{8(\rho_i-\alpha_i)}}\, e^{-\rho_i t}\,(1+o(1)) \\ \text{Chi-cuadrada:} & d_i(t) & = \sqrt{\frac{\alpha_i}{\rho_i-\alpha_i}}\, e^{-\rho_i t} \\ \text{Kullback:} & d_i(t) & = \sqrt{\frac{\alpha_i}{2\rho_i-\alpha_i}}\, e^{-\rho_i t}\,(1+o(1)) \ . \end{array}$$

Por ejemplo tomemos $\alpha_i = \rho_i/2$. Luego la hipótesis (3.3) de convergencia uniforme es trivialmente satisfecha, puesto que $\log\left(d_i(t)/t\right) + \rho_i$ puede ser acotada por $g(t) = K/t$, escogiendo una constante K apropiada. Luego $g(c\tau_n)/\rho_{(1,n)} = K/(c\tau_n\rho_{(1,n)})$ y las hipótesis (3.4) y (3.5) son equivalentes. El hecho que las hipótesis sean o no satisfechas solo depende en la sucesión (ρ_i). Como ya hicimos notar, si $0 < \liminf \rho_i < +\infty$, entonces τ_n tiende a infinito y $\rho_{(1,n)}$ estará acotado lejos 0, entonces podemos aplicar el Teorema 3.3.1. Si ambos ρ_i y $\log i/\rho_i$ son crecientes y tienden a infinito (e.g. $\rho_i = \log(\log(i+2)))$ luego $t_n = \log n/(2\rho_n)$ es un tiempo de Cutoff. La sucesión (ρ_i) también puede tender a 0. Por ejemplo, tomemos $\rho_i = 1/\log(i+1)$: nuevamente podemos aplicar el Teorema 3.3.1; en este caso el tiempo de Cutoff t_n es equivalente a $(\log(n))^2/2$.

Si se tiene (3.3) y si las tasas de convergencia ρ_i convergen a $\rho > 0$, luego se puede aplicar el Teorema 3.3.1. Como veremos más adelante, el tiempo de Cutoff t_n es equivalente a $\log n/(2\rho)$, como si todas las tasas fueran iguales a ρ. Es natural buscar una condición más general bajo la cual $\log n/(2\rho)$ es un tiempo de Cutoff. En el caso de procesos binarios, Bon y Păltănea [12] proponen condiciones suficientes para que el Cutoff ocurra al tiempo $\log n/(2 \liminf \rho_i)$. Su resultado puede verse como caso particular del Teorema 3.3.1 y de la Proposición 3.3.2 que sigue.

Proposición 3.3.2. *[6] Para cualquier ρ positivo, notemos por $N(\rho, n)$ el número de tasas menores o iguales que ρ entre $\rho_1, \ldots, \rho_n$:*

$$N(\rho, n) = \sum_{i=1}^{n} \mathbb{I}_{[0,\rho]}(\rho_i) \ ,$$

donde $\mathbb{I}_A$ es la función indicatriz sobre el conjunto A. Para $n \geq 1$, definimos ρ_n^* como:

$$\rho_n^* = \min \left\{ \frac{\rho_i \log n}{\log N(\rho_i, n)} \ ; \ i = 1, \ldots, n \right\} \ ,$$

con $1/\log(1) = +\infty$. El tiempo de Cutoff t_n definido en (3.6) es asintóticamente equivalente a $t'_n = \log n/(2\rho)$ si y sólo si la sucesión (ρ_n^*) converge a $\rho > 0$.

Demostración: Observemos que t_n puede expresarse en función de $N(\rho_i, n)$ del siguiente modo:

$$t_n = \max \left\{ \frac{\log i}{2\rho_{(i,n)}} \ ; \ i = 1, \ldots, n \right\} = \max \left\{ \frac{\log N(\rho_i, n)}{2\rho_i} \ ; \ i = 1, \ldots, n \right\} \ .$$

El cuociente t'_n/t_n tiende a 1 cuando n tiende a infinito si y sólo si

$$\begin{aligned}\rho &= \lim_{n\to\infty} \frac{\log n}{2t_n} \\ &= \lim_{n\to\infty} \frac{\log n}{\max\left\{\frac{\log N(\rho_i,n)}{\rho_i}\,;\ i=1,\ldots,n\right\}} \\ &= \lim_{n\to\infty} \min\left\{\frac{\rho_i \log n}{\log N(\rho_i,n)}\,;\ i=1,\ldots,n\right\} \ .\end{aligned}$$

$\square$

La Proposición 3.3.2 se puede entender del siguiente modo. Para $\rho > 0$, $N(\rho,n)$ es el número de coordenadas de la n-tupla que convergen más lentamente que $e^{-\rho t}$. Si este número es de un tamaño importante (en el sentido que $\log n/\log N(\rho,n)$ esté acotado, entonces la subtupla de coordenadas correspondientes convergerán solo después del instante $\log N(\rho,n)/(2\rho)$. Este será el tiempo de Cutoff para la n-tupla completa si es el mayor de los tiempos de convergencia de todas las subtuplas de tamaño importante. Pensamos que es interesante ilustrar la idea de Cutoff para subtuplas para un caso más general. En la proposición que sigue a continuación tratamos el caso donde la sucesión (ρ_i) tiene un número finito de puntos de acumulación.

Proposición 3.3.3. *[6] Sea A un entero fijo. Para $a = 1,\ldots,A$, sea $k \mapsto \varphi_a(k)$ una función a valores enteros creciente. Notemos por $m_a(n)$ el número de valores de $\varphi_a(k)$ entre 1 y n:*

$$m_a(n) = \sum_k \mathbb{I}_{[1,n]}(\varphi_a(k)) \ .$$

Asumamos que $m_1(n) + \cdots + m_A(n) = n$ y que los $\varphi_a(k)$ son diferentes dos a dos. Aún más asumamos que para $a = 1,\ldots,A$, la subsucesión $(\rho_{\varphi_a(k)})$ converge a $\varrho_a > 0$.

notemos por s_n *el siguiente real:*

$$s_n = \text{máx}\left\{ \frac{\log m_a(n)}{2\varrho_a} \; ; \; a = 1, \ldots, A \right\} \; ,$$

con $\log 0 = -\infty$. *Luego* (s_n) *y* (t_n) *(definida en* (3.6)*) son asintóticamente equivalentes.*

Demostración: La hipótesis implica que cualquier valor ρ_i pertenece a solo una de las subsucesiones $(\rho_{\varphi_1(n)}), \ldots, (\rho_{\varphi_A(n)})$. Sin perdida de generalidad, asumiremos que $\varrho_1, \ldots, \varrho_A$ son todos distintos y que están ordenados de manera creciente. Para $a = 1, \ldots, A$, sea $m_a^*(n) = m_1(n) + \cdots + m_a(n)$. Sea

$$s_n^* = \text{máx}\left\{ \frac{\log m_a^*(n)}{2\varrho_a} \; ; \; a = 1, \ldots, A \right\} \; .$$

Primero probaremos que (t_n) y (s_n^*) son equivalentes. Usaremos la misma expresión de t_n que usamos en la demostración de la Proposición 3.3.2.

$$\begin{aligned} t_n &= \text{máx}\left\{ \frac{\log N(\rho_i, n)}{2\rho_i} \; ; \; i = 1, \ldots, n \right\} \\ &= \underset{a=1,\ldots,A}{\text{máx}} \; \text{máx}\left\{ \frac{\log N(\rho_{\varphi_a(k)}, n)}{2\rho_{\varphi_a(k)}} \; ; \; k = 1, \ldots, m_a(n) \right\} \; . \end{aligned}$$

Fijemos $\epsilon > 0$, suficientemente pequeño de modo que todos los intervalos $(\varrho_a - \epsilon \, , \, \varrho_a + \epsilon)$ son disjuntos. Para i suficientemente grande $\rho_i \in (\varrho_a - \epsilon \, , \, \varrho_a + \epsilon)$ si $i = \varphi_a(k)$. Entonces existe un entero K tal que para n suficientemente grande,

$$\text{máx}\left\{ \frac{\log N(\rho_{\varphi_a(k)}, n)}{2\rho_{\varphi_a(k)}} \; ; \; k = 1, \ldots, m_a(n) \right\} \leq \frac{\log(m_a^*(n) + K)}{2(\varrho_a - \epsilon)} \; . \tag{3.7}$$

Tomemos ahora n tal que $\rho_{\varphi_a(m_a(n)/2)}, \ldots, \rho_{\varphi_a(m_a(n))}$ son más pequeños que $\varrho_a + \epsilon$, y consideremos el mayor de entre estos $m_a(n)/2$ valores. Entonces podemos establecer que:

$$\max \left\{ \frac{\log N(\rho_{\varphi_a(k)}, n)}{2\rho_{\varphi_a(k)}} \; ; \; k = 1, \ldots, m_a(n) \right\} \geq \frac{\log(m^*_{a-1}(n) + m_a(n)/2 - K')}{2(\varrho_a + \epsilon)} \; , \tag{3.8}$$

para algún entero fijo K'. De las ecuaciones (3.7) y (3.8) se concluye que $t_n \sim s^*_n$.

Solo falta demostrar que $s^*_n \sim s_n$. Obviamente, $s_n \leq s^*_n$. En la definición de s^*_n, el máximo es alcanzado para $a = 1$, o para algún $a > 1$ tal que:

$$\frac{\log m^*_a(n)}{2\varrho_a} \geq \frac{\log m^*_{a-1}(n)}{2\varrho_{a-1}} \iff \frac{\log m^*_a(n)}{\log m^*_{a-1}(n)} \geq \frac{\varrho_a}{\varrho_{a-1}} > 1 \; .$$

Si n es suficientemente grande, esto implica:

$$\begin{aligned} & \log m^*_a(n) > \log m^*_{a-1}(n) + \log 2 \\ \iff & m^*_a(n) > 2m^*_{a-1}(n) \\ \iff & m_a(n) > m^*_{a-1}(n) \\ \iff & 2m_a(n) > m^*_a(n) \; . \end{aligned}$$

Por lo tanto:

$$s^*_n \leq \max \left\{ \frac{\log(2m_a(n))}{2\varrho_a} \; ; \; a = 1, \ldots, A \right\} \; ,$$

de donde se concluye el resultado. □

La Proposición 3.3.3 se entender como sigue. Para $A = 1$, la sucesión de tasas converge a ϱ_1, y el Cutoff ocurrirá como si todas las tasas fueran iguales a ϱ_1. Para

$A > 1$, la n-tupla que consideramos se compone de A subtuplas independientes, con cardinalidades $m_1(n), \dots, m_A(n)$ respectivamente. La a-ésima subtupla tiene un Cutoff al tiempo $\log m_a(n)/(2\varrho_a)$. El tiempo de Cutoff s_n para la n-tupla completa es el último de estos tiempos. Esto puede tener algunas consecuencias inesperadas. Por ejemplo tomemos $A = 2$ y $\varphi_1(k) = k^2$. Uno tiene $m_1(n) = \lfloor\sqrt{n}\rfloor$ y $m_2(n) = n - m_1(n) \sim n$. Tomemos $\varrho_1 = 1$ y $\varrho_2 = 3$. El Cutoff para la n-tupla ocurre en el instante $s_n = \log n/4$, y no $\log n/2$ o $\log n/6$ como uno podría haber pensado originalmente.

El caso particular donde todas las coordenadas convergen a la misma tasa exponencial ρ, las condiciones para el Cutoff (3.4) y (3.5) son automáticamente satisfechas y la condición (3.3) sólo requiere que $(\log d_i(t)/t)$ converja uniformemente en i a $-\rho_i$. Esto se satisface automáticamente si $d_i(t)$ es el mismo para todo i (en particular si las coordenadas son procesos i.i.d.). Aun más asumiremos que $d_i(t)$ converge exponencialmente en un sentido más estricto que el de la Definición 3.2.1: existen dos reales positivos R y ρ tales que para todo i,

$$\lim_{t\to+\infty} d_i(t)e^{\rho t} = R\,. \tag{3.9}$$

Bajo esta hipótesis, la Proposición 2.3.1 produce estimaciones más precisas de la distancia $d^{(n)}(t)$ para t alrededor del instante de Cutoff. El siguiente Teorema muestra estos resultados para las distancias de Hellinger, Chi-cuadrada y Kullback (las demostraciones son sencillas y serán omitidas).

Teorema 3.3.4. *[6]*

1. Asumamos que d es la distancia Hellinger y que (3.9) *se satisface.*

$$\lim_{n\to\infty} d^{(n)}\left(\frac{\log n}{2\rho}+u\right)=\left(1-\exp\left(-R^2e^{-2\rho u}\right)\right)^{1/2} .$$

2. Asumamos que d es la distancia Chi-cuadrado y que (3.9) *se satisface.*

$$\lim_{n\to\infty} d^{(n)}\left(\frac{\log n}{2\rho}+u\right)=\left(\exp\left(R^2e^{-2\rho u}\right)-1\right)^{1/2} .$$

3. Asumamos que d es la distancia Kullback y que (3.9) *se satisface.*

$$\lim_{n\to\infty} d^{(n)}\left(\frac{\log n}{2\rho}+u\right)=Re^{-\rho u} .$$

Como ya observamos la distancia en VT es particular. Aun si asumimos que se tiene (3.9) para ambas distancias, VT y otra distancia, la Proposición 2.3.1 no podemos concluir que $d^{(n)}(\log n/(2\rho)+u)$ converge. Sólo podemos obtener cotas que se obtiene fácilmente de combinar Proposición 2.3.1 con la segunda relación de la Proposición 2.2.2.

Teorema 3.3.5. *[6] Notemos por $d_{TV,i}(t)$ y $d_{H,i}(t)$ la distancias al equilibrio de la i-ésima componente, medidas en las distancias en Variación total y de Hellinger respectivamente. Asumamos que existen reales positivos R_{TV}, R_H y ρ tales que para todo i,*

$$\lim_{t\to+\infty} d_{TV,i}(t)e^{\rho t}=R_{TV} \quad \textit{and} \quad \lim_{t\to+\infty} d_{H,i}(t)e^{\rho t}=R_H .$$

Notemos por $d_{TV}^{(n)}(t)$ la distancia en VT al equilibrio de la n-tupla $X^{(n)}(t)$. Entonces *se satisfacen las siguientes desigualdades:*

$$1 - \exp\left(-\frac{1}{2}R_{TV}^2 e^{-2\rho u}\right) \leq \liminf_{n\to\infty} d_{TV}^{(n)}(\log n/(2\rho) + u)$$

y

$$\limsup_{n\to\infty} d_{TV}^{(n)}(\log n/(2\rho) + u) \leq \left(1 - \exp(-2R_H^2 e^{-2\rho u})\right)^{1/2} .$$

El Teorema 3.3.5 sugiere que la distancia en VT al equilibrio de la n-tupla se comporta como una exponencial doble cuando u tiende a $-\infty$,

$$\limsup_{n\to\infty} 1 - d_{TV}^{(n)}(t_n + u) \leq \exp\left(-\frac{1}{2}R_{TV}^2 e^{-2\rho u}\right) + o(u) ,$$

y una exponencial simple cuando u tiende a $+\infty$,

$$\limsup_{n\to\infty} d_{TV}^{(n)}(t_n + u) \leq \sqrt{2}\, R_H e^{-\rho u} + o(u) .$$

Esto comportamiento es coherente con las cotas obtenidas para el paseo aleatorio en el hipercubo n-dimensional por Diaconis y Shahshahani en [18] y Diaconis *et al.* en [17], y para cadenas de Markov reversibles definidas sobre espacios de estados finitos por Ycart en [63]. En la siguiente sección los Teoremas 3.3.4 y 3.3.5 serán ilustrados por otros casos particulares.

3.4. Ejemplos de procesos i.i.d.

En esta sección presentamos tres ejemplos. En cada uno de los tres casos, consideramos una n-tupla de procesos $(X_1, \ldots, X_n)$. Las coordenadas X_i son copias independientes de X, donde X es un proceso Binario, el proceso de nacimiento y muerte M/M/∞ o el proceso de difusión de Ornstein-Uhlenbeck. En cada caso los Teoremas 3.3.4 y 3.3.5 se pueden aplicar.

3.4.1. Proceso Binario

El proceso X comienza en 0 al tiempo 0, saltará de 0 a 1 con tasa α, y de 1 a 0 con tasa $\rho - \alpha$, como en el ejemplo de la sección anterior. Luego $X(t)$ es una variable de Bernoulli de parámetro

$$p(t) = \frac{\alpha}{\rho}\left(1 - e^{-\rho t}\right) \ ,$$

y su distribución asintótica ν es también una variable Bernoulli, de parámetro α/ρ. La distancia al equilibrio puede ser calculada usando la Proposición 2.4.1 o aplicando los Teoremas 3.3.4 y 3.3.5, con (ver Observación 2.4.2):

$$R_{TV} = \frac{\alpha}{\rho}\ ,\ R_H = \sqrt{\frac{\alpha}{8(\rho-\alpha)}}\ ,\ R_{\chi^2} = \sqrt{\frac{\alpha}{\rho-\alpha}}\ ,\ R_K = \sqrt{\frac{\alpha}{2(\rho-\alpha)}}\ .$$

Obtenemos:

1. Variación total:

$$1 - \exp\left(-\frac{\alpha^2}{2\rho^2}e^{-2\rho u}\right) \le \liminf_{n\to\infty} d_{TV}^{(n)}\left(\frac{\log n}{2\rho} + u\right) \ ,$$

$$\limsup_{n\to\infty} d_{TV}^{(n)}\left(\frac{\log n}{2\rho} + u\right) \le \left(1 - \exp\left(-\frac{\alpha}{4(\rho-\alpha)}e^{-2\rho u}\right)\right)^{1/2} \ .$$

2. Hellinger:

$$\lim_{n\to\infty} d_H^{(n)}\left(\frac{\log n}{2\rho}+u\right) = \left(1-\exp\left(-\frac{\alpha}{8(\rho-\alpha)}e^{-2\rho u}\right)\right)^{1/2} .$$

3. Chi-cuadrado:

$$\lim_{n\to\infty} d_{\chi^2}^{(n)}\left(\frac{\log n}{2\rho}+u\right) = \left(\exp\left(\frac{\alpha}{\rho-\alpha}e^{-2\rho u}\right)-1\right)^{1/2} .$$

4. Kullback:

$$\lim_{n\to\infty} d_K^{(n)}\left(\frac{\log n}{2\rho}+u\right) = \sqrt{\frac{\alpha}{2(\rho-\alpha)}}e^{-\rho u} .$$

Estos resultados se relacionan con cotas similares obtenidas por Ycart en [63] en el contexto de cadenas de Markov en espacios discretos, y por Diaconis y Shahshahani en [18] y Diaconis *et al.* en [17] para el paseo aleatorio en el hipercubo n-dimensional, y Bon y Păltănea en [12] para procesos Binarios independientes no idénticamente distribuidos.

3.4.2. La cola M/M/∞

El proceso X es un proceso de nacimiento y muerte con tasa de nacimiento contante α (de k a $k+1$) y tasa de muerte lineal $k\rho$, de k a $k-1$ (ver e.g. Sección 7a Cap. XVII de [23]). Si $X(0)=0$, la distribución de $X(t)$ es una variable de Poisson de parámetro

$$\alpha(t) = \frac{\alpha}{\rho}\left(1-e^{-\rho t}\right) ,$$

y su distribución asintótica ν también es una variable de Poisson, de parámetro α/ρ. La distancia al equilibrio puede ser calculada usando la Proposición 2.4.3 o aplicando los Teoremas 3.3.4 y 3.3.5, con (ver Observación2.4.5):

$$R_{TV} = R\left(\frac{\alpha}{\rho}\right) \,,\; R_H = \sqrt{\frac{\alpha}{8\rho}} \,,\; R_{\chi^2} = \sqrt{\frac{\alpha}{\rho}} \,,\; R_K = \sqrt{\frac{\alpha}{2\rho}} \,,$$

donde

$$R(a) = \frac{e^{-a}}{\lfloor a \rfloor !} \, a^{\lfloor a \rfloor + 1} \,.$$

Obtenemos:

1. Variación total:

$$1 - \exp\left(-\frac{R(\alpha/\rho)^2}{2} e^{-2\rho u}\right) \leq \liminf_{n\to\infty} d_{TV}^{(n)}\left(\frac{\log n}{2\rho} + u\right) \quad ,$$

$$\limsup_{n\to\infty} d_{TV}^{(n)}\left(\frac{\log n}{2\rho} + u\right) \leq \left(1 - \exp\left(-\frac{\alpha}{4\rho} e^{-2\rho u}\right)\right)^{1/2} \quad .$$

2. Hellinger:

$$\lim_{n\to\infty} d_H^{(n)}\left(\frac{\log n}{2\rho} + u\right) = \left(1 - \exp\left(-\frac{\alpha}{8\rho} e^{-2\rho u}\right)\right)^{1/2} \quad .$$

3. Chi-cuadrado:

$$\lim_{n\to\infty} d_{\chi^2}^{(n)}\left(\frac{\log n}{2\rho} + u\right) = \left(\exp\left(\frac{\alpha}{\rho} e^{-2\rho u}\right) - 1\right)^{1/2} \quad .$$

4. Kullback:

$$\lim_{n\to\infty} d_K^{(n)}\left(\frac{\log n}{2\rho} + u\right) = \sqrt{\frac{\alpha}{2\rho}} e^{-\rho u} \,.$$

El fenómeno de Cutoff para la familia de procesos M/M/∞ indexada por el estado inicial n fue estudiado por Martínez e Ycart en [42] en el contexto de procesos de nacimiento y muerte en árboles. En la Proposición 6.1 los autores encontraron cotas análogas a aquellas de la Proposición anterior para la distancia en VT (ver también p. 293 en [65]).

3.4.3. El proceso Ornstein-Uhlenbeck

El proceso X es la solución de la siguiente ecuación diferencial estocástica (ver e.g. ejemplo 4(b) Cap. X de [24]):

$$\begin{cases} dX(t) &= \alpha\sqrt{2\rho}\; dB_t - \rho X(t)\; dt\;, \\ X(0) &= x_0\;, \end{cases}$$

Donde $\alpha, \rho > 0$ y $\{B_t\,, t \geq 0\}$ es el movimiento Browniano estándar. La distribución de $X(t)$ es una variable Normal de parámetros

$$m(t) = x_0 e^{-\rho t} \quad \text{and} \quad v(t) = \alpha^2 \left(1 - e^{-2\rho t}\right)\;,$$

y su distribución asintótica ν es también una variable Normal, de parámetros 0 y α^2. La distancia al equilibrio puede calcularse usando la Proposición 2.4.6 o aplicando los Teoremas 3.3.4 y 3.3.5, con

$$R_{TV} = \frac{|x_0|}{\alpha\sqrt{2\pi}}\;,\; R_H = \frac{|x_0|}{\alpha\sqrt{8}}\;,\; R_{\chi^2} = \frac{|x_0|}{\alpha}\;,\; R_K = \frac{|x_0|}{\alpha\sqrt{2}}\;.$$

Obtenemos:

1. Variación total:

$$1 - \exp\left(-\frac{x_0^2}{4\pi\alpha^2} e^{-2\rho u}\right) \leq \liminf_{n\to\infty} d_{TV}^{(n)}\left(\frac{\log n}{2\rho} + u\right) \ ,$$

$$\limsup_{n\to\infty} d_{TV}^{(n)}\left(\frac{\log n}{2\rho} + u\right) \leq \left(1 - \exp\left(-\frac{x_0^2}{4\alpha^2} e^{-2\rho u}\right)\right)^{1/2} \ .$$

2. Hellinger:

$$\lim_{n\to\infty} d_H^{(n)}\left(\frac{\log n}{2\rho} + u\right) = \left(1 - \exp\left(-\frac{x_0^2}{8\alpha^2} e^{-2\rho u}\right)\right)^{1/2} \ .$$

3. Chi-cuadrado:

$$\lim_{n\to\infty} d_{\chi^2}^{(n)}\left(\frac{\log n}{2\rho} + u\right) = \left(\exp\left(\frac{x_0^2}{\alpha^2} e^{-2\rho u}\right) - 1\right)^{1/2} \ .$$

4. Kullback:

$$\lim_{n\to\infty} d_K^{(n)}\left(\frac{\log n}{2\rho} + u\right) = \frac{|x_0|}{\sqrt{2}\,\alpha} e^{-\rho u} \ .$$

El cutoff para el proceso de Ornstein-Uhlenbeck ha sido estudiado por Lachaud en [39], en él el autor lo relaciona con la distribución del tiempo (de parada) que le toma a la media empírica de la n-tupla alcanzar el valor 0.

PARTICIONES ALEATORIAS DEL INTERVALO $[0,1]$

Una partición del intervalo $[0,1]$ se construye a partir de cualquier distribución de probabilidad discreta. Sea $\mathbf{p} = (p_1, p_2, \ldots)$ una distribución de probabilidad discreta. Definiremos la partición del intervalo $[0,1]$ asociada a $\mathbf{p}$ como los subintervalos I_i con $i = 1, 2, \ldots$ tales que:

$$I_i = [s_{i-1}, s_i) \ ,$$

donde $s_0 = 0$ y $s_i = \sum_{i=1}^{n} p_i$. En esta partición el subintervalo i-ésimo tiene largo p_i.

Las particiones dan lugar a algunos problemas que enunciamos a continuación. Definimos un muestreo sesgado por tamaño como aquel donde la probabilidad de muestrear el subitervalo i-ésimo es proporcional a su tamaño. Tiremos independientemente k muestreos sesgados por tamaño y llamemos k-muestra a los intervalos muestreados. Dos preguntas naturales son: ¿La k-muestra contiene dos o más veces un mismo subintervalo? ¿han sido todos los subintervalos visitados?. Estas dos preguntas corresponden a dos problemas clásicos de probabilidades la paradoja del cumpleaños y el problema del coleccionista de cupones (ver Feller [23] pág. 47-48 para la definición de los problemas). Otros problemas similares quizás menos conocidos que se pueden plantear en una partición son: las reglas de organización mover al frente (move-to-front MtF), desplazar hacia la raíz (move-to-root MtR), el menos utilizado (Least recently used LRU), regla de transposición (transposition rule) y la permutación sesgada por tamaño (Size-biased permutation S-BP) (ver [22]). Como veremos en este capítulo y el siguiente, varios autores han abordado estos problemas, entre ellos se puede destacar el trabajo de Flajolet *et al.* [28] por utilizar el marco teórico común de los lenguajes regulares para abordar los problemas: de la paradoja del cumpleaños, el coleccionista de cupones, MtF y LRU.

Hay muchas aplicaciones que involucran la descomposición de un ítem de masa unitaria en componentes de tamaño aleatorio (partición aleatoria), por ejemplo la asignación de memoria de un computador, la frecuencia de un gen en una población, la fragmentación de una roca. En estos modelos puede tener tanto o más sentido hacerse las mismas preguntas que hemos descrito para las particiones no aleatorias. Como veremos a lo largo capítulo hay varias maneras de generar una partición aleatoria, presentaremos dos (Sección 4.2): la partición por renormalización y el modelo de localización aleatoria.

Entre las particiones aleatoria una de las más utilizadas es la partición aleatoria de Poisson-Dirichlet (PD) introducida por Kingman (ver[35]). La partición de PD resulta de ordenar de manera decreciente la partición de Griffiths-Engen-McCloskey (GEM) y esta última se obtiene de una permutación aleatoria de la PD. Además la partición GEM se obtiene como el límite de una permutación aleatoria de la partición de Dirichlet (D). La partición GEM y su versión ordenada, la partición PD, constituyen un modelo que tiene aplicaciones en distintos ámbitos como: en ecología, en genética, en estadística bayesiana y teoría de números (ver sec. 6 en [59] y las referencias citadas ahí). La relación entre esta tres particiones, PD, GEM y D, está descrita por Kingman en [35] (pág.90-99) y los detalles de ellas que son relevantes para esta tesis están descritos en la Sección 4.3. La relevancia de estas tres particiones nos motiva a estudiar algunas nuevas propiedades de la partición de D y otra partición que, al igual que la permutación aleatoria de la partición de D, también aproximan la partición GEM.

En este capítulo en la Sección 4.1 recordamos algunas cantidades importantes definidas para una partición (no aleatoria) del intervalo unitario. Luego en la Sección 4.2 introducimos dos modelos generales para generar particiones aleatorias. El primero que llamaremos partición por renormalización es generado por n variables independientes, donde el i-ésimo segmento se define como la i-ésima variable renormalizada por la suma total. El segundo, que llamaremos de localización aleatoria que es generado por n variables independientes con soporte en $(0, 1)$ donde el i-ésimo segmentos se genera como el producto de i variables independientes. En la Sección 4.3 introducimos formalmente la partición aleatoria de Dirichlet, recordamos varias de sus propiedades, su relación con las particiónes PD y GEM. También los resultados realizados en colaboración con T. Huillet y Ch. Paroissin publicados en el artículo [4] en la revista "Probability in the engineering and informational science". En la Sección 4.4 definimos y estudiamos las propiedades del modelo de partición de *fragmentación de la vara* que aproxima la partición de GEM para cierto parámetro. Este trabajo [3] realizado en colaboración T. Huillet fue publicado en la revista "Statistics & probability letters".

4.1. Particiones del intervalo $[0, 1]$

En esta sección definiremos algunas cantidades para una partición no aleatoria como: la media de Rényi, el muestreo sesgado por tamaño, el costo de búsqueda en una lista, la permutación sesgada por tamaño. Identificaremos la partición con la distribución discreta que la genera $\mathbf{p_n}$ en el caso finito de n segmentos y simplemente $\mathbf{p}$ en el caso de una partición numerable.

Definición 4.1.1. *Para $\beta \in I\!R$ la media de Rényi es*

$$\langle \mathbf{p_n} \rangle_\beta := \left(\sum_{m=1}^{n} p_m^{\beta+1} \right)^{1/\beta} .$$

Observemos que esta cantidad tiende al tamaño del intervalo más grande de la partición cuando β tiende a ∞ y al tamaño del intervalo más pequeño cuando β tiende a $-\infty$.

Definición 4.1.2. *Definimos el muestreo sesgado por tamaño a partir de una variable aleatoria V distribuida como uniforme en $[0,1]$, como:*

$$I(V) = \sum_{i=1}^{n} i \mathbb{1}_{[s_{i-1}, s_i]}(V) ,$$

Donde s_i son las sumas parciales de $\mathbf{p_n}$. La variable $I(V)$ es una variable discreta que toma el valor i con probabilidad p_i.

Supongamos que $\mathbf{p_n}$ corresponde a la probabilidad de requerir unos objetos numerados de 1 a n. Imaginemos que disponemos estos objetos en una lista en una permutación ς desconocida. La búsqueda lineal de un objeto consiste en revisar la lista desde un extremo hasta encontrarlo, podemos decir, que el costo de encontrar el objeto en la k-ésima posición es $k-1$. Ahora si el objeto que se requiere se pide aleatoriamente de acuerdo a un muestreo sesgado por tamaño, llamaremos S al costo (aleatorio) de buscarlo en la lista.

Proposición 4.1.3. *Sea ς una permutación de $\{1,\dots,n\}$. Consideremos una lista donde los objetos están ordenado según ς. La distribución del costo S de buscar un objeto requerido según un muestreo sesgado por tamaño en esta lista está dada por*

$$P(S=k+1)=p_{\varsigma_k}\ .$$

Donde donde p_k es la probabilidad de requerir el objeto k.

Definición 4.1.4. *La permutación sesgada por tamaño (size-biased permutation S-BP) de una partición $\mathbf{p_n}$, es la permutación aleatoria $p_{\sigma_1},\dots p_{\sigma_n}$ de la partición, donde $\sigma=(\sigma_1,\dots,\sigma_n)$ se define como: σ_1 es un muestreo sesgado por tamaño de $\mathbf{p_n}$, se remueve el intervalo σ_1 escogido y se renormalizan los intervalos restantes para obtener una partición $\mathbf{p_{n-1}}$, se repite la operación para definir iterativamente $\sigma_2,\dots,\sigma_n$ hasta que no queden intervalos.*

La S-BP es una permutación aleatoria, y la probabilidad de obtener una permutación ς es:

$$P(\sigma=\varsigma)=p_{\varsigma_1}\frac{p_{\varsigma_2}}{1-p_{\varsigma_1}}\cdots\frac{p_{\varsigma_{n-1}}}{1-p_{\varsigma_1}-p_{\varsigma_2}-\cdots-p_{\varsigma_{n-2}}}\ . \tag{4.1}$$

El tamaño del primer segmento de la S-BP satisface la siguiente propiedad.

Proposición 4.1.5. *Sea $\mathcal{L}=p_{\sigma_1}$ y sea g una función real luego*

$$I\!E\left(\frac{g(\mathcal{L})}{\mathcal{L}}\right)=\sum_{i=1}^{n}g(p_i)\ .$$

En particular para $g(y) = y\mathbb{1}_{y\leq x}$ obtenemos la distribución de $\mathcal{L}$:

$$F_{\mathcal{L}}(x) = \sum_{m=1}^{n} p_m \mathbb{1}_{\{p\leq x\}} \ . \tag{4.2}$$

4.2. Particiones Aleatorias

En esta sección introducimos dos modelos generales para construir una partición aleatoria: la partición por renormalización y la partición del modelo de localización aleatoria. Además definimos la propiedad de intercambiabilidad de una partición aleatoria.

4.2.1. El modelo de partición por renormalización

Una manera de generar una partición aleatoria es a partir de sucesión $W = (W_i)_{i\in\mathbb{N}}$ de variable aleatorias independientes no negativas. Definimos la partición $\mathbf{p_n}$ de n elementos como

$$\forall i \in \{1, ..., n\} \ , \qquad p_i = \frac{W_i}{\Omega_n} \quad \text{donde} \quad \Omega_n = \sum_{i=1}^{n} W_i \ . \tag{4.3}$$

Notemos que una misma partición puede ser engendrada por secuencias ω y ω' de variables aleatorias positivas i.i.d. diferentes en distribución.

Observación 4.2.1. *Si W y W' son dos sucesiones de variables aleatorias independientes estrictamente positivas i.i.d. tales que para todo i se tiene $W_i' \stackrel{(d)}{=} kW_i$. Sea n un entero cualquiera, consideremos los n primero términos para generar las* *particiones aleatorias* $\mathbf{p_n}$ *y* $\mathbf{p'_n}$. *Sea* $\Omega_n = W_1 + \cdots + W_n$ *y* $\Omega'_n = W'_1 + \cdots + W'_n$. *Luego las dos* $\mathbf{p_n}$ *y* $\mathbf{p'_n}$ *son iguales en distribución,*

$$p'_i \overset{(d)}{=} \frac{kW_i}{\sum_{j=1}^{n} kW_j} = \frac{kW_i}{k\Omega_n} = p_i \ .$$

Sea $f(p_1, \ldots, p_n)$ la densidad de la partición $\mathbf{p_n}$ con respecto a la medida de Lebesgue. Una partición se dice intercambiable si la densidad es simétrica en su argumentos, es decir, si σ es una permutación de $1, ..., n$

$$f(p_1, \ldots, p_n) = f(p_{\sigma_1}, \ldots, p_{\sigma_n}) \ . \tag{4.4}$$

Un simple cálculo nos permiten demostrar que la esperanza de $\mathbf{p_n}$ vale $(\frac{1}{n}, \ldots, \frac{1}{n})$, dependiendo únicamente del numero de intervalos que tiene la partición n.

Proposición 4.2.2. *Sea* $\mathbf{p_n}$ *una partición construida a partir de una secuencia* $W = (W_i)_{i \in \mathbb{N}}$ *de variable aleatorias no negativas i.i.d utilizando la fórmula (4.3). La partición* $\mathbf{p_n}$ *es intercambiable.*

4.2.2. El modelo de localización aleatoria

Supongamos que una especie llega y ocupa una fracción aleatoria del espacio disponible. Luego llega una segunda especie e independientemente de la primera utiliza una fracción aleatoria del espacio que quedó disponible. Se repite el proceso para n especies, donde cada una de ellas ocupa una fracción aleatoria del espacio disponible. Este modelo de partición aleatoria es conocido como el modelo de localización aleatoria (Random alocation model RAM).

Definamos el modelo formalmente. Consideremos una sucesión $\mathbf{U} = (U_j)_{j\in I\!N}$ de variables aleatorias independientes con soporte en el intervalo $[0,1]$. Sea $\overline{U}_j = 1 - U_j$. Se define la partición $\mathbf{p_{n+1}} = (p_1, \ldots, p_{n+1})$ como

$$p_i = U_i \prod_{j=1}^{i-1} \overline{U}_j \quad \forall i = 1, \ldots, n \quad \text{ y } \quad p_{n+1} = \prod_{j=1}^{n} \overline{U}_j \; . \tag{4.5}$$

Donde p_{n+1} es el espacio desocupado y p_i es el espacio ocupado por la especie i para $i \leq n$. También se puede definir la partición $\mathbf{p} = (p_i)_{i\in I\!N}$ con una cantidad numerable de especies como:

$$p_i = U_i \prod_{j=1}^{i-1} \overline{U}_j \quad \forall i \in I\!N \, . \tag{4.6}$$

4.3. La familia de particiones de Dirichlet

Las particiones GEM y PD son una de las particiones más estudiada y utilizada como modelo. La partición GEM debe su nombre a McCloskey y Egen quienes la utilizaron para modelar el tamaño de poblaciones en ecología y a Griffiths quien utilizo el modelo en genética (ver [59] y las referencias citadas). Como veremos en este capítulo la partición PD debe a su nombre a que se puede obtener como límite de la partición finita de Dirichlet (D) que ha sido ordenada decrecientemente (ver cap. 9 de [35]).

En esta sección primero introducimos formalmente la partición de D, veremos que puede ser generado con modelos de partición renormalizada y la partición RAM y estableceremos su relación con las particiones numerables PD y GEM. Luego presentamos algunos resultados nuevos para la S-BP, como la función de momentos para cada segmento y la función de momentos conjunta. Finalmente utilizando estas propiedades estudiamos la distribución del costo de búsqueda en una lista (ver Sección 4.1 Propiedad 4.1.3) para la partición de D y su permutación sesgada por tamaño.

4.3.1. Introducción del modelo

La partición D cuenta con muchas propiedades, enumeraremos aquellas que sean relevantes para el desarrollo de los resultados de esta tesis.

Consideremos n constantes estrictamente positivas $\alpha_1, \ldots, \alpha_n$ y notemos por $\overrightarrow{\alpha}$ al vector $(\alpha_1, ..., \alpha_n)$. La densidad de una partición de D de parámetros $\overrightarrow{\alpha}$ esta dada por:

$$f_{\mathbf{p_n}}(p_1, ..., p_n) = \frac{\Gamma(\alpha_1 + \alpha_2 + \ldots + \alpha_n)}{\Gamma(\alpha_1)\Gamma(\alpha_2) \ldots \Gamma(\alpha_n)} p_1^{\alpha_1 - 1} p_2^{\alpha_2 - 1} \ldots p_n^{\alpha_n - 1} \tag{4.7}$$

donde $p_1 + \ldots + p_n = 1$ y Γ es la función Gamma. Una de las características que hace a la partición D popular es que se puede construir a partir de una secuencia de variables aleatorias independientes $(W_i)_{i \in I\!N}$ utilizando el modelo de renormalización (ecuación (4.3)) donde W_i se distribuye como una variable Gamma de parámetro α_i (ver [35] pág. 91-92). Cuando las variables $(W_i)_{i \in I\!N}$ son i.i.d. decimos que la partición se distribuye como una partición de D simétrica y la llamaremos $D_n(\alpha)$.

En este sección nos interesaremos en la partición de D simétrica la cual tiene la importante propiedad de que su S-BP (ver Definición 4.1.4) puede describirse con el modelo localización aleatoria. Sean $U_1, \ldots, U_{n-1}$ variables aleatorias independientes con U_i distribuida como una variable Beta de parámetros $(\alpha+1, (n-i)\alpha)$. Si llamamos $\mathbf{l_n}$ a la partición que resulta de aplicar una permutación sesgada por tamaño, tendremos que

$$
\begin{aligned}
l_m &= U_m \prod_{k=1}^{m-1} \overline{U}_k \text{ con } m = 1, \ldots, n-1 \ , \qquad (4.8)\\
l_n &= \prod_{k=1}^{m} \overline{U}_k \ . \qquad (4.9)
\end{aligned}
$$

Si tomamos el limite cuando el numero de intervalos n tiende a infinito y $n\alpha$ tiende a γ, obtenemos la partición GEM de parámetro γ. Este procedimiento es conocido como el límite de Kingman. Notemos que la partición GEM(γ) también responde al modelo de localización aleatoria dado por la ecuación (4.6) construida a partir de una secuencia de variables aleatorias $(U_i')_{i\in I\!N}$ i.i.d. distribuidas como una variable Beta de parámetros $(1, \gamma)$ (para más detalles ver nuevamente [35] pág. 98-99). Ordenando la partición GEM(γ) en intervalos decrecientes se obtiene la partición de PD de parámetro γ (PD(γ)).

Finalmente, recordemos que la ley Beta de parámetros a y b ($B_{a,b}$) estrictamente positivos con soporte en $[0, 1]$ tiene densidad:

$$
f_{a,b}(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1} \ ,
$$

y su función de generadora de momentos es

$$\mathbb{E}\left(B_{a,b}^q\right) = \frac{\Gamma(a+q)\Gamma(a+b)}{\Gamma(a)\Gamma(a+b+q)} \; . \tag{4.10}$$

4.3.2. Muestreo y permutación sesgada por tamaño

Consideremos $\mathbf{l_n}$ la partición aleatoria resultante de aplicar una S-BP a $\mathbf{D_n}(\alpha)$. A lo largo de la sección encontraremos algunas de propiedades de $\mathbf{l_n}$ entre ellas función de momentos conjunta.

Proposición 4.3.1. *[4] La variable $\mathcal{L} = l_1$ es un muestreo sesgado por tamaño, $\mathcal{L}$ se distribuye como una variable Beta de parámetros $(1+\alpha, (n-1)\alpha)$. Luego $\mathcal{L}$ satisface las siguiente desigualdad:*

$$\mathcal{L} \succcurlyeq_{st} p_m \; .$$

Demostración: De la ecuación 4.2 tenemos que $F_L(x) = n\int_0^x yF_{p_n}(y)\;dt$. Luego para demostrar la desigualdad estocástica basta probar que $n\int_x^1 yF_{p_n}(y)\;dt \geq \overline{F}_{p_n}(x)$ o en otras palabras $\mathbb{E}(p_m \mid p_m > x) \geq 1/n = \mathbb{E}(p_m)$. Esta última desigualdad (y por ende la proposición) es cierta para cualquier partición simétrica. □

No deja de ser paradójico que si escogemos un intervalo al azar de $\mathbf{p_n}$ sea estocásticamente más grande que cada uno de los intervalos de la partición. Esta aparente contradicción puede entenderse si notamos que en un muestre sesgado por tamaño los segmentos más grandes son preferidos a los más pequeños (paradoja discutida por Feller en [24] pág. 22-23 para $\alpha = 1$ y también fue investigada por Hawkes en [30] pág. 294-295).

Corolario 4.3.2. *[4] Sea B_n una variable aleatoria distribuida como una Bernoulli de parámetro $1/n$ y sea $B_{\alpha,1}$ otra variable aleatoria distribuida como una ley Beta de parámetros $(\alpha, 1)$ independiente de B_n. Definimos la variable R_n con soporte en $[0, 1]$ como*

$$R_n \stackrel{(d)}{=} B_n + (1 - B_n) \cdot B_{\alpha,1} \ .$$

Luego, $\mathcal{L}$ y p_n satisfacen la siguiente igualdad

$$R_n \ \mathcal{L} \stackrel{(d)}{=} p_n \ ,$$

con R_n y $\mathcal{L}$ independientes.

Demostración: De la definición de R_n deducimos que la función de momentos vale:

$$\mathbb{E}[R_n^q] = \frac{1}{n} + \left(1 - \frac{1}{n}\right) \frac{\alpha}{\alpha + q}$$

Tomando $g(x) = x^q$ en la Proposición 4.1.5 obtenemos:

$$\mathbb{E}[\mathcal{L}^q] = n\mathbb{E}\left[p_m^{q+1}\right] = \frac{n\Gamma(n\alpha)\Gamma(\alpha + q + 1)}{\Gamma(\alpha)\Gamma(n\alpha q + 1)} \ .$$

Recordando que la función generadora de momentos de p_m es:

$$\mathbb{E}[p_m^q] = (\Gamma(n\alpha)\Gamma(\alpha + q)) \,/\, (\Gamma(\alpha)\Gamma(n\alpha q)) \ .$$

Podemos encontrar la siguiente factorización:

$$\mathbb{E}[p_m^q] = \frac{n\alpha + q}{n(\alpha + q)} \mathbb{E}[\mathcal{L}^q] = \mathbb{E}[R_n^q]\, \mathbb{E}[\mathcal{L}^q] \ .$$

□

Encontramos que la partición $\mathbf{l_n}$ es decreciente en el sentido de estocástico o más precisamente el siguiente resultado:

Teorema 4.3.3. *[4]*

1. *La ley del m-ésimo fragmento de* $\mathbf{l_n}$ *esta caracterizada por:*

$$\begin{aligned} \mathbb{E}[L_m^q] =& \frac{\Gamma(1+\alpha+q)\Gamma((n-m+1)\alpha+1)}{\Gamma(\alpha+1)\Gamma((n-m+1)\alpha+1+q)} \\ & \times \prod_{k=1}^{m-1} \frac{\Gamma((n-k)\alpha+q)\Gamma((n-k+1)\alpha+1)}{\Gamma((n-k)\alpha)\Gamma((n-k+1)\alpha+1+q)} \ , \end{aligned} \tag{4.11}$$

y para $m = n$

$$\mathbb{E}[l_n^q] = \prod_{k=1}^{n-1} \frac{\Gamma((n-k)\alpha+q)\Gamma((n-k+1)\alpha+1)}{\Gamma((n-k)\alpha)\Gamma((n-k+1)\alpha+1+q)} \ .$$

2. *Sea* $B_{(n-m)\alpha,1}$ *una variable aleatoria distribuida como una Beta de parámetros* $((n-m+1)\alpha, 1)$*. Luego*

$$l_m \overset{(d)}{=} B_{(n-m)\alpha,1} l_{m-1} \text{ para } m = 1, \ldots, n \ ,$$

donde para cada $m = \{1, \ldots, n\}$ *la variable* $B_{(n-m)\alpha,1}$ *es independiente de* L_{m-1}*.*

3. $l_1 \succcurlyeq_{st} \ldots \succcurlyeq_{st} l_n$

Demostración: La parte (1) es consecuencia directa de la construcción del modelo. La caracterización de $\mathbf{l_n}$ con el modelo de localización aleatoria en la ecuación (4.8) nos permite deducir que para $m = \{1, \ldots, n-1\}$

$$\mathbb{E}[l_m^q] = \mathbb{E}[U_m^q] \prod_{k=1}^{m-1} \mathbb{E}\left[\overline{U}_k^q\right] ,$$

y para $m = n$

$$\mathbb{E}[l_n^q] = \prod_{k=1}^{n} \mathbb{E}\left[U_k^q\right] .$$

Sabemos que cada U_m se distribuye como una variable Beta de parámetros $(1 + \alpha, (n-m)\alpha)$ y que son independientes entre si. Además $\overline{U}_m$ también se distribuye una variable Beta pero de parámetros $((n-m)\alpha, 1+\alpha)$. Recordando la función de momentos de una ley Beta (ver ecuación 4.10) se obtiene el resultado. La parte (3) del teorema es consecuencia directa de (2) por lo que solo necesitamos probar (2). La función de momentos de $B_{(n-m+1)\alpha,1}$ está dada por:

$$\mathbb{E}\left[B^q_{(n-m+1)\alpha,1}\right] = \frac{\Gamma((1+\alpha+q)\Gamma((n-m+1)\alpha+1)}{\Gamma(\alpha+1)\Gamma((n-m+1)\alpha+1+q)} .$$

Luego reagrupando términos en la relación de la parte (1) del teorema encontramos, $\mathbb{E}[l_m^q] = \mathbb{E}\left[B^q_{(n-m+1)\alpha,1}\right] \mathbb{E}[l_{m-1}^q]$. $\square$

El resultado (2) del Teorema 4.3.3 se puede encontrar también en [15] con una demostración similar.

De este último resultado podemos deducir el siguiente corolario para la esperanza de l_m:

Corolario 4.3.4. *[4] Sea $\theta := 1/\alpha$, luego*

$$\mathbb{E}[l_m] = \frac{(\theta+1)\Gamma(n)}{\Gamma(\theta+n+1)} \frac{\Gamma(\theta+n-m+1)}{\Gamma(n-k+1)} \ , \ \textit{para } k = \{1, \ldots, n\} \ ,$$

con $\sum_{i=1}^{n} \mathbb{E}[l_m] = 1$.

Demostración: Reemplazando en la expresión (4.11) $q = 1$ y $\theta = 1/\alpha$ se obtiene el resultado:

$$\begin{aligned}
\mathbb{E}[l_m] &= \prod_{k=1}^{m} \frac{(n-i)\alpha}{(n-i+1)\alpha+1} \frac{1+\theta}{(n-k+1)\theta+1} \\
&= (\theta+1)\frac{(n-1)!}{(n-k)!} \frac{1}{\prod_{k=0}^{m-1}(\theta+n-k)} \\
&= (\theta+1)\frac{\Gamma(n)}{\Gamma(n-k+1)} \frac{\Gamma(\theta+n-k+1)}{\Gamma(\theta+n+1)} \ .
\end{aligned}$$

□

Calcularemos ahora la función generadora de momentos conjunta de la partición $\mathbf{l_n}$. De la caracterización de la ecuación 4.8 tenemos

$$\begin{aligned}
\mathbb{E}\left[\prod_{m=1}^{n} l_m^{q_m}\right] &= \mathbb{E}\left[\prod_{m=1}^{n} U_m^{q_m} \prod_{k=1}^{m-1} \overline{U}_k^{q_m}\right] \\
&= \prod_{k=1}^{n} \mathbb{E}\left[U_k^{q_k} \overline{U}_k^{q_{k+1}+\ldots+q_n}\right] \ .
\end{aligned}$$

Luego usando la función de momentos de U_k tenemos el siguiente teorema:

Teorema 4.3.5. *[4] La función de momentos conjunta de l_n está dada por*

$$\begin{aligned} & I\!\!E\left[\prod_{m=1}^{n} l_m^{q_m}\right] \\ & = \prod_{k=1}^{n}\left\{\frac{\Gamma(1+(n-k+1)\alpha)}{\Gamma(1+\alpha)\Gamma(m-k)\alpha}\frac{\Gamma(1+\alpha+q_k)\Gamma((n-k)\alpha)+q_{k+1}+\ldots+q_n}{\Gamma(1+(n-k+1)\alpha+q_k\ldots+q_n)}\right\} \end{aligned} \tag{4.12}$$

Demostración: Sea V una variable distribuida como una ley Beta de parámetros (a,b). Sea $\overline{V} := 1-V$. Luego

$$\begin{aligned} I\!\!E\left[V^{q_1}\overline{V}^{q_2}\right] &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\int_0^1 v^{a+q_1-1}(1-v)^{b+q_2-1}\,dv \\ &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\frac{\Gamma(a+q_1)\Gamma(b+q_2)}{\Gamma(a+b+q_1+q_2)}\,. \end{aligned}$$

Utilizando esta expresión para calcular $I\!\!E\left[U_k^{q_k}\overline{U}_k^{q_{k+1}+\ldots+q_n}\right]$ se tiene la conclusión. □

Observemos que si $q_k = q/n$ para $k = 1,\ldots,n$ en el Teorema 4.3.5 se obtiene el promedio geométrico de la partición $\mathbf{l_n}$.

4.3.3. Una comparación del costo de búsqueda en la partición de Dirichlet y su permutación sesgada por tamaño

En este modelo las probabilidad de requerir el i-ésimo ítem es p_i, ahora consideremos dos posibilidades para ordenar los ítemes en una lista: dejar los ítem en el orden original o reordenarlos de acuerdo a una permutación sesgada por tamaño. El costo de búsqueda del ítem i esta dado por su posición en la lista, luego si se requiere un ítem de acuerdo a un muestreo sesgado por tamaño ¿cual es el costo de buscar en cada una de las listas?.

Llamemos C_n al costo de búsqueda en la lista en el orden original $\mathbf{p_n}$ y S_n el costo de búsqueda en la lista permutada $\mathbf{l_n}$ (ver Proposición 4.1.3). Primero para $\mathbf{p_n}$ tenemos el costo de buscar el ítem i es $i-1$ luego la función generadora de momentos de C_n es:

$$\mathbb{E}\left[e^{-sC_n}\right] = \frac{1}{n}\sum_{m=1} ne^{-s(m-1)} = \frac{1}{n}\frac{1-e^{-sn}}{1-e^{-s}} \ .$$

En consecuencia el costo esperado vale $\mathbb{E}[C_n] = (n-1)/2$, $\mathbb{E}[C_n^2] = (n-1)(2n-1)/6$ y $\mathrm{VAR}[C_n] = (n-1)(n+1)/12$. Podemos encontrar el siguiente comportamiento asintótico para C_n.

Proposición 4.3.6. *[4] Sea U una variable distribuida como una ley Uniforme en $[0,1]$ luego*

$$\frac{C_n}{n} \xrightarrow[n\to\infty]{(d)} U \ .$$

Demostración: La demostración es directa tomando límite en la función generadora de momentos

$$\lim_{n\to\infty} \mathbb{E}\left[e^{-sC_n}\right] = \frac{1-e^{-s}}{s} \ ,$$

y reconociendo la transformada de Laplace de una distribución Uniforme en $[0,1]$. Fill encuentra el mismo comportamiento para el costo de búsqueda de una lista en otro contexto [25]. □

Ahora estudiemos el costo de búsqueda S_n en la lista permutada.

Lema 4.3.7. *[4]*

1. La distribución de S_n es:

$$P(S_n = k) = \frac{(\theta+1)\Gamma(n)}{\Gamma(\theta+n+1)} \frac{\Gamma(\theta+n-k)}{n-k} \text{ para } k = 0, \ldots, n \; .$$

es unimodal con moda en $k = 0$.

2. El primer y segundo momento son

$$\mathbb{E}[S_n] = \frac{n-1}{2-\theta} \quad y \quad \mathbb{E}\left[S_n^2\right] = \frac{(n-1)(2n+\theta-1)}{(\theta+2)(\theta+3)} \; .$$

Demostración: La primera parte es consecuencia directa de la expresión de $\mathbb{E}[l_m]$ en el Corolario 4.3.4 pues $P(S_n = k) = l_{k+1}$. Notemos que $P(S_n = 0) = (\theta+1)/(1+n\theta)$, y para cada $k = 0, \ldots, n-1$ tenemos

$$\frac{P(S_n = k+1)}{P(S_n = k)} = \frac{n-k-1}{\theta+n-k-1} < 1,$$

que nos indica que la moda de la distribución es $k = 0$. Para (2) tenemos que

$$\begin{aligned} \mathbb{E}[S_n^q] &= \sum_{k=1}^{n-1} k^q \mathbb{E}[l_k] = \frac{(\theta+1)\Gamma(n)}{\Gamma(\theta+n+1)} \sum_{k=1}^{n-1} k^q \frac{\Gamma(\beta+n-k)}{\Gamma(n-k)} \\ &= \frac{(\theta+1)\Gamma(n)}{\Gamma(\theta+n+1)} \sum_{i=1}^{n-1} (n-i)^q \frac{\Gamma(\beta+i)}{\Gamma(i)} \; . \end{aligned}$$

Sea $A := \Gamma(\theta+n+1)/\left((\theta+1)\Gamma(n)\right)$, como $\sum_{k=0}^{n-1} P(S_n = k) = 1$ podemos expresar A en términos de la siguiente sumatoria:

$$A = \sum_{i=1}^{n} \frac{\Gamma(\beta+i)}{\Gamma(i)} \; .$$

Sea $B := \Gamma(\theta+n+1)/\left((\theta+2)\Gamma(n-1)\right)$, utilizando el mismo razonamiento que el usado para A (basta tomar $(\theta+1, n-1)$ en vez de (θ, n)) expresamos B como:

$$
\begin{aligned}
B &= \sum_{i=1}^{n-1} \frac{\Gamma(\beta+1+i)}{\Gamma(i)} = \sum_{i=1}^{n-1} i \frac{\Gamma(\beta+1+i)}{\Gamma(i+1)} \\
&= \sum_{i=2}^{n} (i-1) \frac{\Gamma(\beta+i)}{\Gamma(i)} = \sum_{i=1}^{n} i \frac{\Gamma(\beta+i)}{\Gamma(i)} - A\,.
\end{aligned}
$$

Tomando $q=1$ obtenemos el valor esperado del costo de búsqueda S_n

$$
\begin{aligned}
\mathbb{E}[S_n] &= \frac{1}{A}\left(n \sum_{i=1}^{n} i \frac{\Gamma(\beta+i)}{\Gamma(i)} - \sum_{i=1}^{n} i \frac{\Gamma(\beta+i)}{\Gamma(i)}\right) \\
&= \frac{1}{A}\left(n\left(A - \frac{\Gamma(\theta+n)}{\Gamma(n)}\right) - \left(A + B - n\frac{\Gamma(\theta+n)}{\Gamma(n)}\right)\right) \\
&= n - 1 - \frac{(n-1)(\theta+1)}{\theta+1} \\
&= \frac{n-1}{2+\theta}\,.
\end{aligned}
$$

Usando argumentos similares podemos encontrar el segundo momento ($q=2$). Sin ahondar en muchos detalles sea

$$
C := \frac{\Gamma(\theta+n+1)}{(\theta+3)\Gamma(n-2)}\,.
$$

Similarmente a lo hecho para B, podemos expresar C en función de A y B:

$$
C = \frac{(n-1)(n-2)(\theta+1)}{\theta+3} A = \sum_{i=1}^{n} i^2 \frac{\Gamma(\theta+i)}{i} - 3B - A\,.
$$

Luego para $q = 2$ obtenemos

$$\begin{aligned}
\mathbb{E}\left[S_n^2\right] &= \frac{1}{A}(n^2A - 2n(A+B) + C + 3B + A) \\
&= \frac{(n-1)(2n+\theta-1)}{(\theta+2)(\theta+3)} \ .
\end{aligned}$$

□

El resultado para la esperanza de S_n ya había sido obtenido por Kingman en [34], con otras técnicas. El resultado para la varianza también fue obtenido en el artículo [7] que también forma es parte de esta tesis, (ver Capítulo 5 en Sección 5.2.3). Nos parece que la distribución de S_n (parte (1) del Lema 4.3.7) es nueva.

Podemos obtener el comportamiento asintótico de S_n al igual que lo hicimos para C_n.

Teorema 4.3.8. *[4] Sea $B_{1,1+1/\alpha}$ una variable siguiendo una ley Beta de parámetros $(1, 1+1/\alpha)$ luego*

$$\frac{S_n}{n} \xrightarrow[n\to\infty]{(d)} B_{1,1+1/\alpha} \ .$$

Demostración: De la expresión de la función de momentos de S_n,

$$\mathbb{E}\left[\left(\frac{S_n}{n}\right)^q\right] = \frac{(\theta+1)\Gamma(n)}{\Gamma(\theta+n+1)} \sum_{i=1}^{n-1} \left(1 - \frac{i}{n}\right) \frac{\Gamma(\theta+i)}{\Gamma(i)} \ .$$

Para n grande, esta cantidad se puede aproximar por

$$\mathbb{E}\left[\left(\frac{S_n}{n}\right)^q\right] \sim \frac{(\theta+1)\Gamma(n)}{\Gamma(\theta+n+1)} \int_0^1 (1-x)^q \frac{\Gamma(\theta+nz)}{nz} \, dz \ .$$

De la formula de Stirling con $a > 0$, podemos aproximar $\Gamma(a+z)/\Gamma(z)$ por z^a cuando z tiende a infinito. Esto nos permite aproximar la función de momentos por

$$\begin{aligned} \mathbb{E}\left[\left(\frac{S_n}{n}\right)^q\right] &\sim \frac{\theta+1}{n^\theta}\int_0^1 (1-z)^q (nz)^\theta \, dz \\ &\sim (\theta+1)\int_0^1 (1-z)^q (z)^\theta \, dz \ , \end{aligned}$$

la cual es la función de momentos de una variable $B_{1,1+\theta}$ de media $1/(2+\theta)$ menor que $1/2$. Este resultado fue generalizado para cualquier S-BP de una partición obtenida por renormalización simétrica en [4] resultado que forma parte de esta tesis en el Capítulo 5 en la Sección 5.2.2. $\square$

Tenemos que en el caso en que los ítemes se quedan en el orden original el costo de búsqueda normalizado por el número de ítemes C_n/n tiende a una Uniforme, mientras que en el caso de la lista permutada (permutación sesgada por tamaño) el costo de búsqueda normalizado S_n/n tiene una distribución Beta de parámetros $(1, 1+1/\alpha)$. Claramente tenemos $U \succcurlyeq_{st} B_{1,1+1/\alpha}$ expresando el hecho que C_n es asintóticamente mayor que S_n, luego $\mathbf{l_n}$ organiza mejor los ítemes que $\mathbf{p_n}$.

Recordemos que la partición GEM(γ) se obtiene de realizar el límite de Kingman sobre $\mathbf{l_n}$, luego podemos estudiar el comportamiento asintótico del costo de búsqueda S_n para este límite.

Proposición 4.3.9. *[4] Sea γ una constante positiva y $\alpha(n) = \gamma/n$, el costo S_n en la partición $\mathbf{l_n}$ con parámetro $\alpha(n)$ converge a la ley Geométrica de parámetro γ (G).*

$$S_n \xrightarrow[n\to\infty]{(d)} G \ .$$

Demostración: Sea $\mathbf{l}_\infty^*$ la partición GEM(γ) a la que converge $\mathbf{l_n}$ cuando tomamos el límite de Kingman. Tenemos que para $\mathbf{l}_\infty^* = (l_i^*)_{i \in I\!N}$ se tiene

$$I\!E[l_i^*] = \left(\frac{\gamma}{1+\gamma}\right)^{k-1} \frac{1}{1+\gamma} \ ,$$

Luego la función generadora de momentos del costo de búsqueda de esta partición S^* vale

$$I\!E\left[e^{sS^*}\right] = \sum_{i \geq 1} e^{si-1} I\!E[l_i^*] = \frac{1}{1+\gamma(1-e^{-s})} \ .$$

La cual es la transformada de Laplace de una distribución geométrica. Notemos que $P(S^* = 0) = E(l_1^*)$. $\square$

4.4. El modelo de fragmentación de la vara

Consideremos una partición GEM(1) y la partición de PD(1). En conjunto satisfacen las siguientes importantes propiedades de invariancia:

- El modelo de partición GEM(1) es invariante para la operación de S-BP (ver [57], [45], [62]).
- Si se ordena por tamaño decrecientes la partición GEM(1) se obtiene la partición PD(1) y se se realiza una S-BP en una partición PD(1) se obtiene una partición GEM(1) (ver [47]).
- La partición de PD(1) es invariante ante las operaciones de inserción y supresión

de un intervalo (ver [47]) y para las operaciones de fragmentación y fusión de dos intervalos (ver [46]).

Para comprender mejor la tercera propiedad describiremos las operaciones. La operación de inserción consiste en generar un segmento de largo X distribuido según una ley Beta en insertarlo en la partición previamente reescalada por un factor $1 - X$ y reordenar, la operación de supresión en una partición consiste en suprimir un intervalo escogido de acuerdo a un muestreo sesgado por tamaño y renormalizar. La operación de fragmentación y fusión toma dos muestreos sesgados por tamaño, los divide en un punto aleatorio distribuido uniformemente si ambos muestreos caen en el mismo intervalo y los fusiona en uno solo de largo la suma de los dos si caen en dos intervalos distintos y luego los ordena de manera decreciente.

Como ya mencionamos la partición GEM(1) se puede aproximar por la S-BP de una partición de D simétrica $D_n(1/n)$. En esta sección nos interesamos en otra aproximación de esta partición y sus propiedades; consideremos la partición del intervalo unitario $\mathbf{p_{n+1}}$ de $n+1$ fragmentos generada con el modelo de fracturación de la vara a partir de una secuencia de variables i.i.d. $(U_i)_{i \in \mathbb{N}}$ distribuidas como uniformes en $[0, 1]$. Cuando n tiende a infinito este modelo converge débilmente a la partición GEM(1). La función densidad de $\mathbf{p_{n+1}}$ está dada por:

$$f_{\mathbf{p_{n+1}}}(p_1, ..., p_n) = \frac{1}{\prod_{m=1}^{n} \left(1 - \sum_{k=1}^{m} p_k\right)} \mathbb{1}_{\{\sum_{m=1}^{n} p_m < 1\}} \, .$$

Estudiaremos algunos aspectos estadísticos de este modelo de partición. Los resultados que presentaremos a continuación fueron publicados en el artículo [3].

Recordemos que la densidad de una variable aleatoria Gamma de parámetro α con soporte en $I\!R^+$ es

$$f_\alpha(x) = \frac{x^{\alpha-1}e^x}{\Gamma(\alpha)} \ ,$$

donde Γ es la función Gamma.

4.4.1. El promedio geométrico de los intervalos ocupados

Primero calculamos la función de momentos de la cual derivaremos algunas propiedades.

Lema 4.4.1. *[3] La función conjunta de momentos de la partición* $\mathbf{p_{n+1}}$ *está dada por:*

$$I\!E\left(\prod_{m=1}^{n} p_m^{q_m}\right) = \frac{1}{1+q_1+\ldots+q_n}\prod_{m=1}^{n}\frac{\Gamma(1+q_m)}{1+q_m+\ldots+q_n} \ .$$

Demostración: De la definición de $\mathbf{p_{n+1}}$ como una partición RAM podemos calcular la función generadora de momentos conjunta:

$$\begin{aligned}
I\!E\left(\prod_{m=1}^{n} p_m^{q_m}\right) &= I\!E\left(\prod_{m=1}^{n}\prod_{k=1}^{n}\bar{U}_k^{q_m}U_m^{q_m}\right) = \prod_{m=1}^{n} I\!E\left(U_m^{q_m}U_m^{q_{m+1}+\ldots+q_n}\right) \\
&= \prod_{m=1}^{n}\frac{\Gamma(1+q_m)\Gamma(1+q_{m+1}+\ldots+q_n)}{\Gamma(2+q_m+\ldots+\ldots+q_n)} \\
&= \frac{1}{\Gamma(1+q_1+\ldots+q_n)}\prod_{m=1}^{n}\frac{\Gamma(1+q_m)}{1+q_m+\ldots+\ldots+q_n} \ .
\end{aligned}$$

□

Este resultado nos permite obtener el comportamiento del promedio geométrico de los intervalos. Definimos la función F como $F(q) := (1+q^{-1})\log(1+q) - 1$ con $F(0) = 0$. Tendremos que $F'(q)$ tiende a $1/2$ cuando q tiende a 0. Luego

Proposición 4.4.2. *[3] Se tiene el siguiente límite c.s.*

$$\lim_{n\to\infty} \prod_{m=1}^{n} p_m^{1/n} = e^{-F'(0)} = 0,6065\ldots\,;$$

y

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n} \log P\left(\prod_{m=1}^{n} p_m^{1/n} > e^{-x} \right) &= f(x) \ \textit{con } x > \tfrac{1}{2}\,, \\
\lim_{n\to\infty} \frac{1}{n} \log P\left(\prod_{m=1}^{n} p_m^{1/n} \leq e^{-x} \right) &= f(x) \ \textit{con } x \in (0, \tfrac{1}{2}]\,,
\end{aligned}$$

donde $f(x) = \inf_{q>-1}(qx - F(q))$ es una cantidad menor o igual a 0. La función f corresponde a la trasformada de Legendre cóncava de F con $f(1/2) = 0$.

Demostración: Tomamos la ecuación del Lema 4.4.1 con $q_m = q/n$ para $m = 1, \ldots, n$ siendo $q > -1$, así obtenemos

$$\mathbb{E}\left[\left(\prod_{m=1}^{n} p_m^{1/n} \right)^{q} \right] = \frac{\Gamma(1+q/n)^n}{\Gamma(1+q)} \prod_{m=1}^{n} \frac{1}{1+mq/n}\,. \tag{4.13}$$

Si consideramos $q > 0$ tenemos

$$\mathbb{E}\left[\left(\prod_{m=1}^{n} p_m^{1/n} \right)^{q} \right] = \frac{\Gamma(1+q/n)^n}{\Gamma(1+q)(q/m)^n} \frac{\Gamma(1+n/q)}{\Gamma(1+n+n/q)}\,.$$

Usando la fórmula de Stirling obtenemos el siguiente equivalente asintótico cuando

n tiende a infinito, donde $\mathfrak{E}$ es la constante de Euler:

$$\mathbb{E}\left[\left(\prod_{m=1}^{n} p_m^{1/n}\right)^q\right] \sim \frac{e^{\mathfrak{E}q}}{\Gamma(1+q)(1+q)^{1/2}} \left(e(1+q)^{-(1+1/q)}\right)^n \ , \text{ con } q > 0 \ .$$

Por otro lado si $q > -1$ se tiene el siguiente límite:

$$\lim_{n\to\infty} -\frac{1}{n} \log\left(\prod_{m=1}^{n} \frac{1}{1+mq/n}\right) = \int_0^1 \log(1+qx)dx = F(q) \ .$$

Aplicando esta aproximación a la ecuación (4.13) tendremos la siguiente convergencia puntual para el promedio geométrico de la partición,

$$\lim_{n\to\infty} -\frac{1}{n} \log \mathbb{E}\left[\left(\prod_{m=1}^{n} p_m^{1/n}\right)^q\right] = F(q) \ .$$

□

El siguiente lema relaciona el promedio geométrico de la partición $\mathbf{p_{n+1}}$ con el promedio geométrico de una partición de Dirichlet simétrica.

Lema 4.4.3. *[3] Sea* $\mathbf{d_n}$ *una partición de Dirichlet simétrica de* n *intervalos de parámetro* $\alpha = 1$*. Sea* $B_{1,n-1}$ *una variable aleatoria de ley Beta de parámetros* $(1, n-1)$*, independiente de* $\prod_{m=1}^{n} p_m^{1/n}$*. Sean* $\{B_{n/m,1} \ : \ m = 1, \ldots, n\}$ *variables i.i.d. con* $B_{n/m,1}$ *de ley Beta de parámetros* $(n/m, 1)$*. Luego tenemos:*

$$B_{1,n-1} \prod_{m=1}^{n} p_m^{1/n} \overset{(d)}{=} \prod_{m=1}^{n} d_m^{1/n} \prod_{m=1}^{n} B_{n/m,1}, \tag{4.14}$$

con $\{B_{n/m,1} \ : \ m = 1, \ldots, n\}$ *independiente de* $\prod_{m=1}^{n} d_m^{1/n}$*.*

Demostración: La función conjunta de momentos de $\mathbf{d_n}$ es

$$\mathbb{E}\left[\prod_{m=1}^{n} d_m^{q_m}\right] = \frac{\Gamma(n)}{\Gamma(n+q_1+\ldots+q_n)}\prod_{m=1}^{n}\Gamma(1+q_m)\ .$$

Imponiendo $q_m = q/n$ para $m = 1, \ldots, n$ tenemos

$$\mathbb{E}\left[\left(\prod_{m=1}^{n} d_m^{1/n}\right)^q\right] = \frac{\Gamma(n)}{\Gamma(n+q)}\Gamma(1+q/n)^n\ .$$

La función de momentos de $B_{n/m,1}$ es $(1+mq/n)^{-1}$ luego la función de momentos de $\prod_{m=1}^{n} B_{n/m,1}$ es

$$\mathbb{E}\left[\prod_{m=1}^{n} B_{n/m,1}^q\right] = \prod_{m=1}^{n}\frac{1}{1+mq/n}\ .$$

Luego recordando que la función de momento de $B_{1,n-1}$ es $\Gamma(1+q)\Gamma(n)/\Gamma(n+q)$, basta con reconocer los términos de la ecuación (4.13) para concluir. $\square$

4.4.2. La función partición

Definimos la función de partición de $\mathbf{p_{n+1}}$ como $Z_n(\beta) := \sum_{m=1}^{n+1} p_m^\beta$ para $\beta > -1$. El rango de esta variable es $[(n+1)^{1-\beta}, 1]$ para $\beta \geq 1$ y $[1, (n+1)^{1-\beta}]$ si $\beta < 1$. Luego el comportamiento de $Z_n(\beta)$ cuando n tiende a infinito está dado por la siguiente proposición,

Proposición 4.4.4. *[3] Sea $\beta > 0$. Luego $Z_n(\beta)$ converge en distribución cuando n tiende a infinito,*

$$Z_n(\beta) \xrightarrow[n\to\infty]{d} Z_\infty(\beta)\ ,$$

donde $Z_\infty(\beta) \in [0,1]$ para $\beta > 1$ y $Z_\infty(\beta) \in (1,\infty)$ si $\beta \in (0,1)$. La variable $Z_\infty(\beta)$ está caracterizada por la siguiente ecuación donde U_1 sigue la ley Uniforme en $[0,1]$ y $Z_\infty(\beta)$ y $Z'_\infty(\beta)$ son i.i.d:

$$Z_\infty(\beta) \overset{(d)}{=} U_1^\beta + \overline{U}_1^\beta Z_\infty(\beta)' \, . \tag{4.15}$$

Demostración: Tenemos que $Z_1(\beta) = U_1^\beta + \overline{U}_1^\beta$ luego para $n \geq 2$ se tiene

$$Z_n(\beta) = U_1^\beta + \overline{U}_1^\beta (U_2^\beta + \overline{U}_2^\beta U_3^\beta + \ldots + \overline{U}_2^\beta \ldots \overline{U}_n^\beta) = U_1^\beta + \overline{U}_1^\beta Z'_{n-1}(\beta) \, ,$$

con $Z'_{n-1}(\beta) \overset{(d)}{=} Z_{n-1}(\beta)$ e independiente de U_1. Luego para $\beta > 0$ tenemos que $Z_n(\beta)$ converge en distribución a $Z_\infty(\beta)$ definido en la ecuación (4.15). Notemos por $m_q(\beta) = \mathbb{E}[Z_\infty(\beta)^q]$ el q-ésimo momento de $Z_\infty(\beta)$. La función m_q queda definida recursivamente para $q \in \mathbb{N}$ por:

$$m_q(\beta) = \sum_{k=1}^{q} \binom{q}{k} \mathbb{E}\left[U_1^{\beta(q-k)} \overline{U}_1^{\beta k}\right] m_k(\beta) \, ,$$

con $\mathbb{E}\left[U_1^{\beta(q-k)} \overline{U}_1^{\beta k}\right] = \Gamma((q-k)\beta+1)\Gamma(k\beta+1)/\Gamma(q\beta+2)$. En particular $m_1(\beta) = 1/\beta$ y $m_2(\beta) = (2\beta+1)^{-1}[1 + \Gamma(\beta)^2/\Gamma(2\beta)]$ son el primer y segundo momento. $\square$

A continuación enumeramos algunas de las propiedades consecuencias directas de la Proposición 4.4.4:

1. se tiene $m(1) = 1$ y $m_2(1) = 1$,

2. Si $\beta \in (-1,0)$ luego $(\mathbb{E}Z_n(\beta))^{1/n}$ converge a $1/(1+\beta)$ cuando n tiende a

infinito y se deduce que $Z_n(\beta)$ no converge.

3. Cuando $\beta > 0$, la función generadora de momentos de $Z_\infty(\beta)$ puede obtenerse de (4.15) elevando a la potencia λ, desarrollando el lado derecho en una serie de potencia de λ

$$\mathbb{E}\left[Z_\infty^\lambda\right] = \sum_{a_1,a_2\geq 0} (\lambda)_{a_1+a_2} \mathbb{E}\left[(U_1^\beta - 1)^{a_1}\overline{U}_1^{\beta a_2}\right] \frac{m_{a_2}(\beta)}{a_1!a_2!} ,$$

con $(\lambda)_a = \lambda(\lambda-1)\ldots(\lambda-a+1)$, $(\lambda)_0 := 1$ y

$$\begin{aligned}\mathbb{E}\left[(U_1^\beta - 1)^{a_1}\overline{U}_1^{\beta a_2}\right] &= \sum_{r=0}^{a_1}(-1)^{a_1-r}\binom{a_1}{r}\mathbb{E}\left[U_1^{\beta r}\overline{U}_1^{\beta a_2}\right] \\ &= \sum_{r=0}^{a_1}(-1)^{a_1-r}\binom{a_1}{r}\frac{\Gamma(r\beta+1)\Gamma(a_2\beta+1)}{\Gamma((r+a_2)\beta+2)} .\end{aligned}$$

Tomando $\lambda = q/\beta$ con $q \geq 0$ nos da $\mathbb{E}\left[Z_\infty^{q/\beta}\right]$ que es límite de la función de momentos de la β-norma de $\mathbf{p_{n+1}}$: $Z_n(\beta)^{1/\beta}$ cuyo rango está en el intervalo $[(n+1)^{(1-\beta)/\beta}, 1]$ considerando $\beta > 0$.

4. El β-promedio de Rényi de los fragmentos de $\mathbf{p_{n+1}}$ se puede calcular como:

$$\langle \mathbf{p_{n+1}} \rangle_\beta := Z_n(\beta+1)^{1/\beta} \text{ con } \beta \in \mathbb{R} .$$

El rango de esta variable es $[1/(n+1), 1]$ para cualquier $\beta \in (-1, \infty)$ y $[0, 1/(n+1)]$ para cualquier $\beta \in (-\infty, -1)$. Notemos que si $p_{1:n+1}$ es el intervalo más grande y $p_{n+1:n+1}$ es el intervalo más pequeño de $\mathbf{p_{n+1}}$, se tiene que $\langle \mathbf{p_{n+1}} \rangle_\beta$ converge a $p_{1:n+1}$ cuando β tiende a ∞ y converge a $p_{n+1:n+1}$ cuando β tiende a $-\infty$. Para el promedio de Rényi de $\mathbf{p_{n+1}}$ obtenemos el siguiente resultado

Corolario 4.4.5. *[3] Si $\beta > -1$ y $\beta \neq 0$ la variable $\langle \mathbf{p_{n+1}} \rangle_\beta$ converge en distribución a $\langle \mathbf{p_\infty} \rangle_\beta \in [0,1]$. Donde la función de momentos de $\langle \mathbf{p_{n+1}} \rangle_\beta$ está dada por:*

$$\mathbb{E}\left[\langle \mathbf{p_{n+1}} \rangle_\beta^q\right] = \sum_{a_1,a_2 \geq 0} \binom{q}{\beta}_{a_1+a_2} \mathbb{E}\left[(U_1^{\beta+1}-1)^{a_1}\overline{U}_1^{(\beta+1)a_2}\right] \frac{m_{a_2}(\beta+1)}{a_1!a_2!} \,.$$

4.4.3. La ley unidimensional de un segmento

Consideremos la ley de un intervalo p_m con $m \in \{1,\ldots,n\}$. Es fácil demostrar que p_m se distribuye como una variable log-Gamma, cuya función densidad y distribución están dadas por (ver [24] pág. 47):

$$\begin{aligned} f_{p_m}(x) &= \frac{1}{(m-1)!}(-\log x)^{m-1} \,, \qquad (4.16)\\ F_{p_m}(x) &= x\sum_{k=0}^{m-1}\frac{1}{k!}(-\log x)^k \text{ con } x \in [0,1] \,. \end{aligned}$$

La distribución de la variable p_m se puede caracterizar también por su función de momentos $\mathbb{E}(p_m^q) = (1+q)^{-m}$ con $q > -1$ y tiene media 2^{-m}.

Notemos que el espacio restante p_{n+1} y el espacio ocupado por la última especie p_n son iguales en distribución luego $\mathbb{E}(p_{n+1}^q) = (1+q)^{-n}$ con $q > -1$ y tiene media 2^{-n}. Utilizando la ley de los grandes números y el teorema central del límite (versiones multiplicativas) tenemos la siguiente proposición.

Proposición 4.4.6. *[3] El espacio restante p_{n+1} (que es igual en distribución al* *espacio ocupado por la última especie p_n) satisface los siguientes limites*

$$p_{n+1} \xrightarrow[n\to\infty]{c.s.} e^{-1} \ , \quad \left[ep_{n+1}^{1/n}\right] \xrightarrow[n\to\infty]{(d)} L \ ,$$

donde L es una distribución Lognormal.

Observemos que si consideramos el espacio ocupado por las m primeras especies J_m, tendremos que $1 - J_m$ satisface el mismo resultado que acabamos de presentar pues equivale a considerar la partición $\mathbf{p_{m+1}}$.

4.4.4. Orden decreciente en tamaño: la distribución del más pequeño y del más largo de los intervalos

Sea $\mathbf{p_{n+1}}^{\downarrow} = (p_{1:n+1}, \ldots, p_{n+1:n+1})$ la partición obtenida al ordenar por tamaño decreciente la partición $\mathbf{p_{n+1}}$ incluyendo el espacio restante p_{n+1}. En particular, para el intervalo más pequeño y para el intervalo más largo encontramos la siguiente proposición:

Proposición 4.4.7. *[3] Sea $\overline{F}_{p_{n+1:n+1}}(x) = P(p_{n+1:n+1} > x)$. Luego podemos definir por recurrencia la distribución del más pequeño de los segmentos $p_{n+1:n+1}$ por*

$$\overline{F}_{p_{2:2}}(x) = (1 - 2x) \ , \ \textit{con } x \in (0, 1/2)$$

y para $n \geq 2$

$$\overline{F}_{p_{n+1:n+1}}(x) = \int_x^{1-sx} \overline{F}_{p_{n:n}}\left(\frac{x}{1-u}\right) du \ , \ \textit{con } x \in (0, 1/(n+1)) \ .$$

Sea $F_{p_{1:n+1}}(x) = P(p_{1:n+1} \leq x)$. Luego también podemos definir por recurrencia la distribución del intervalo más grande segmento $p_{1:n+1}$ por

$$F_{p_{1:2}}(x) = (2x-1)\ ,\ \ \text{con } x \in (1/2,1)\ ,$$

y para $n \geq 2$

$$F_{p_{1:n+1}}(x) = \int_{(1-sx)_+}^{x} F_{p_{1:n}}\left(\frac{x}{1-u}\right)\,du\ ,\ \ \text{con } x \in (1/(n+1),1)\quad .$$

Demostración: La demostración de ambas recurrencias es similar, basta condicionar en el tamaño del primer p_1 de la partición $\mathbf{p_{n+1}}$. Será suficiente demostrar la recurrencia el segmento más largo.

Cuando $n=1$ tenemos dos segmentos que suman 1, luego es evidente que el más grande de los intervalos tiene un tamaño más grande que 1/2.

$$F_{p_{1:2}}(x) = P(U_1 \leq x, 1-U_1 \leq x) = 2x-1\ ,\ \ \text{con } x \in (1/2,1)\quad .$$

Para $n \geq 2$, condicionamos en U_1:

$$\begin{aligned}
F_{p_{1:n+1}}(x) &= P\left(p_1 \leq x, \ldots, p_{n+1} \leq x\right) \\
&= P\left(U_1 \leq x, \overline{U}_1 U_2 \leq x, \ldots, \prod_{m=1}^{n} \overline{U}_m \leq x\right) \\
&= \int_{(1-nx)\vee 0}^{x} P\left(U_2 \leq \frac{x}{1-u}, \ldots, \prod_{m=2}^{n} \overline{U}_m \leq \frac{x}{1-u}\right) du \\
&= \int_{(1-nx)_+}^{x} F_{p_{1:n}}\left(\frac{x}{1-u}\right) du\ ,
\end{aligned}$$

donde el límite inferior de la integral es $(1-nx)_+$ porque el soporte de la variable $p_{1:n}$ es $(1/n, 1)$. Para $n=2$ la variable $p_{1:3}$ tiene soporte en $(1/3, 1)$ y distribución:

$$F_{p_{1:3}}(x) = -2x \log\left(\frac{1-x\wedge(1-x)}{1-(1-2x)}\right) + x - x\wedge(1-x)\ .$$

La distribución y la densidad de $p_{1:3}$ son continuas, en cambio la derivada de la densidad presenta una discontinuidad en $x=1/2$. □

Proposición 4.4.8. *[3] Sea ζ_1 una variable aleatoria tal que $1/\zeta_1$ se distribuye como una ley de Dickman. Luego el más pequeños de los segmentos $p_{1:n+1}$ converge en distribución a ζ_1 cuando n tiende a infinito.*

Demostración: Asumamos que $p_{1:n+1}$ converge en distribución a una variable que llamaremos $p_{1:\infty}$. Supongamos que $p_{1:\infty}$ es una variable no degenerada y sea F_∞ su función distribución. Luego de la definición recursiva de $F_{p_{1:n+1}}$ tenemos que F_∞ debe ser solución funcional de

$$F_\infty(x) = \int_0^x F_\infty\left(\frac{x}{1-u}\right)\ du \text{ con } x\in(0,1)\ dt\ .$$

Esta ecuación equivale a decir que la variable $p_{1:\infty}$ satisface la relación $p_{1:\infty} \overset{(d)}{=} U_1 \vee \mathbf{U}_1 p'_{1:\infty}$ donde $p'_{1:\infty}$ y $p_{1:\infty}$ están idénticamente distribuidas y ambas son independiente de U_1. Haciendo un cambio de variable $x/(1-u)$ obtenemos:

$$F_\infty(x) = x\int_{(1-x)/x}^{1/x} F_\infty\left(\frac{1}{t}\right)\ dt\ .$$

Finalmente sea G la función distribución de $1/p_{1:\infty}$, luego $\overline{G}(t) = F_\infty(1/t)$ con $t>1$.

Reemplazando G en la relación anterior,

$$x(1-G(x)) = \int_{(1-x)/x}^{1/x} G(t)\ dt\ .$$

Derivando obtenemos la relación

$$x\overline{G}'(x) + \overline{G}(x-1) = 0,$$

que muestra que G es una función de Dickman (ver [33] y las referencia citadas ahí). Luego la distribución de $p_{1:\infty}$ es:

$$F_\infty(x) = 1 + \sum m = 1^{[1/s]} \frac{(\log x)^m}{m!}\ ,\ \text{con } x \in (0,1)\ .$$

$\square$

Para el intervalo más pequeños encontramos el siguiente comportamiento asintótico,

Proposición 4.4.9. *[3] Cuando n tiende a infinito*

$$\liminf_{n\to\infty} e^n p_{n+1:n+1} = 0 \quad y \quad \limsup_{n\to\infty} e^n p_{n+1:n+1} = \infty\ .$$

Demostración: Sea $\mathbf{d} = (d_i)_{m\in\mathbb{N}}$ una partición PD de parámetro 1. Kingman estableció que $e^m d_m$ converge a 1 c.s. cuando m tiende a infinito, luego podríamos especular que existe una variable no degenerada $\mathcal{C}$ estrictamente positiva tal que,

$$e^n p_{n+1:n+1} \xrightarrow{(d)} \mathcal{C}\ .$$

Si fuera así sea F_∞ la distribución de la variable $\mathcal{C}$. De la recursión para el segmento más pequeño de la Proposición 4.4.7, tenemos que $\overline{F}_\infty = 1 - F_\infty$ debería satisfacer la siguiente relación:

$$\overline{F}_\infty(x) = \int_0^1 \overline{F}_\infty\left(\frac{x}{e(1-u)}\right)\, du \ .$$

Esto significa que $\mathcal{C}$ satisface la siguiente igualdad $\mathcal{C} \overset{(d)}{=} e\overline{U}\mathcal{C}'$ donde $\overline{U}$ y $\mathcal{C}'$ son independientes, $\overline{U}$ sigue una ley Uniforme y $\mathcal{C}'$ se distribuye como $\mathcal{C}$. Después de un cambio de variable en la integral obtenemos:

$$\overline{F}_\infty(x) = \frac{x}{e}\int_0^{e/x} \overline{F}_\infty\left(\frac{1}{t}\right)\, dt \ .$$

Derivando esta expresión obtenemos la relación

$$x\overline{F}'_\infty(x) = \overline{F}_\infty(x) - \overline{F}_\infty(x/e) \ .$$

Sea $\overline{G}(z) = \overline{F}_\infty(e^z)$ con $z \in I\!R$, luego $\overline{G}$ debe satisfacer la ecuación de diferencia-diferencial:

$$\overline{G}'(z) = \overline{G}(z) - \overline{G}(z-1) \quad \text{con } \overline{G}(-\infty) = 1 \text{ y } \overline{G}(\infty) = 0 \ .$$

Bajo estos supuestos las únicas soluciones podrían ser las funciones constantes $\overline{G}(z) = 0$ o $\overline{G}(z) = 1$. Luego las soluciones para la ecuación funcional original son $\overline{F}(x) = 0$ para $x > 0$ ó $\overline{F}(x) = 0$ para $x \geq 0$. $\square$

Ahora si consideramos el más pequeños y más largo de los intervalos ocupados sin tomar en cuenta p_{n+1}, es decir, $p^+_{n:n+1} := p_1 \vee \ldots \vee p_n$ y $p^+_{1:n+1} := p_1 \wedge \ldots \wedge p_n$.

Obtenemos una relación de recurrencia similar pero más simple:

$$F_{p^+_{1:2}}(x) = s \quad \text{con } x \in (0,1) \quad ,$$

y para $n \geq 2$

$$F_{p^+_{1:n+1}}(x) = \int_0^x F_{p^+_{1:n}} \left(\frac{x}{1-u} \right) \, du \ , \ \text{con } x \in (1/(n+1), 1) \quad .$$

Luego para $n = 2$ obtenemos

$$F_{p^+_{1:2}}(x) = s \log \left(1 - s \wedge (1-s)\right) + s - s \wedge (1-s) \ , \quad \text{con } x \in (0,1) \ .$$

La función distribución y la densidad de $p^+_{1:2}$ son continuas pero la derivada de la densidad tiene una discontinuidad en $s = 1/2$. Usando el mismo razonamiento que utilizamos para $p_{1:n+1}$ tenemos que $p^+_{1:n+1}$ converge en distribución a L cuando n tiende a infinito donde $1/L$ se distribuye como una ley de Dickman.

Análogamente para el mínimo tenemos:

$$F_{p^+_{1:2}}(x) = 1 - s \quad \text{con } x \in (0,1) \quad ,$$

y para $n \geq 2$

$$\overline{F}_{p^+_{n:n+1}}(x) = \int_x^1 \overline{F}_{p^+_{n-1:n}} \left(\frac{x}{1-u} \right) \, du \ , \ \text{con } x \in (1/(n+1), 1) \quad .$$

La cantidad $\overline{F}_{p^+_{n:n+1}}(x)$ se interpreta en el modelo como la probabilidad que todas las especies ocupen al menos una fracción x del espacio (la cual podría ser una condición de supervivencia en algún medio para algún valor de x).

De mismo modo podemos definir la distribución conjunta del segmento más pequeño y más grande de $\mathbf{p_n}$. Sea

$$G_n(x_1, x_2) := P\left(p_{1:n+1} > x_1, p_{n+1:n+1} \leq x_2\right) \ .$$

Luego

$$G_1(x_1, x_2) = x_2 \wedge (1 - x_1) - x_1 \vee (1 - x_2) \ , \text{ con } x_1 < \tfrac{1}{2} \leq x_2$$

y para $n \geq 2$

$$G_n(x_1, x_2) = \int_{x_1(1-x_2)}^{x_2(1-x_1)} G_{n-1}\left(\frac{x_1}{1-u}, \frac{x_2}{1-u}\right) du \ \text{ con } x_1 < \frac{1}{n+1} \geq s_2 \ .$$

Análogamente, sea

$$G_n^+(x_1, x_2) := P\left(p^+_{1:n+1} > x_1, p^+_{n,n+1} \leq x_2\right) \ .$$

Luego para G_1^+ tenemos

$$G_1^+(x_1, x_2) = x_2 - x_1 \ , \text{ con } x_1 < x_2$$

y para $n \geq 2$

$$G_n^+(x_1, x_2) = \int_{x_1}^{x_2} G_{n-1}\left(\frac{x_1}{1-u}, \frac{x_2}{1-u}\right) du \ \text{ con } x_1 < x_2 \ .$$

Esta cantidad nos permite estudiar la regularidad de los espacios ocupados por las especies. Si $x_1 = x$ y $x_2 = 2x$ tenemos que $G_n^+(x, 2x)$ es la probabilidad que cada una de las especies ocupe una fracción mayor que x y menor que $2x$ luego los gaps $|p_m - x|$ serán menores que x para esta configuración.

4.4.5. Permutación sesgada por tamaño

Sea $\mathbf{l_{n+1}} = (l_1, \ldots, l_{n+1})$ la partición resultante de realizar una S-BP de $\mathbf{p_{n+1}}$. De la Proposición 4.1.5 deducimos la siguiente relación para la distribución del primer intervalo l_1, utilizando la función $g(l) = l\,\mathbb{1}_{l>x}$:

$$
\begin{aligned}
\overline{F}_{l_1}(x) &= \sum_{m=1}^{n+1} \mathbb{E}\left[p_m \mathbb{1}_{p_m>x}\right] \\
&= \sum_{m=1}^{n+1} \int_x^1 t f_{p_m}(t)\, dt\ .
\end{aligned}
$$

Reemplazando el valor de la densidad f_{p_m} dado en la ecuación (4.16) en la última expresión:

$$
\overline{F}_{l_1}(x) = \sum_{m=0}^{n-1} \int_x^1 \frac{t(-\log t)^m}{m!}\, dt + \int_s^1 \frac{t(-\log t)^{n-1}}{(n-1)!}\, dt\ . \tag{4.17}
$$

Podemos proceder análogamente para la función de momentos de l_1 y obtenemos

$$
\begin{aligned}
\mathbb{E}\,[l_1^q] &= \sum_{m=0}^{n+1} \mathbb{E}\left[p_m^{q+1}\right] \\
&= \sum_{m=0}^{n} \frac{1}{(1+q)^m} + \frac{1}{(1+q)^n} \\
&= \frac{1+q(2+q)^{-n}}{1+q}\ .
\end{aligned}
$$

Notemos que $\mathbb{E}[l_1] = \mathbb{E}[\langle \mathbf{p_{n+1}} \rangle_1] \geq \mathbb{E}(p_1)$, luego podríamos conjeturar que l_1 es estocásticamente mayor que p_1 que, estocásticamente, es más grande de los segmentos de la partición.

Proposición 4.4.10. *[3] Para el primer segmento de* $\mathbf{l_{n+1}}$ *se tiene*

$$l_1 \succcurlyeq_{st} p_1 \ .$$

Demostración: Debemos demostrar que para todo x en $[0,1]$ se tiene $\overline{F}_{l_1}(x) \geq \overline{F}_{p_1}(x)$. Tenemos que en $\overline{F}_{l_1}(0) = \overline{F}_{p_1}(0) = 1$ y $\overline{F}_{l_1}(1) = \overline{F}_{p_1}(1) = 0$. Además derivando dos veces la expresión de $\overline{F}_{l_1}$ en la ecuación (4.17) obtenemos

$$\overline{F}''_{l_1}(x) = -\frac{2}{(n-1)!}(\log s)^{(n-1)}\left[1 + \frac{n-1}{2\log s}\right] \ .$$

El único punto de inflexión de $\overline{F}_{l_1}$ es $x = e^{(n-1)}$. Luego como $\overline{F}_{l_1}(0) = 0$ y $\overline{F}_{l_1}(1) = -1$ se tiene que $\overline{F}_{l_1}(x) \geq 1 - x = \overline{F}_{p_1}(x)$ la desigualdad deseada. $\square$

Finalmente notemos que $\mathbb{E}[l_1^q]$ tiende a $\mathbb{E}[p_1^q] = 1/(1+q)$ cuando n tiende a infinito es natural si consideramos la propiedad de invarianza de la partición límite GEM(1).

Consideremos ahora la función de momentos conjunta de $\mathbf{l_{n+1}}$, que recordemos es la S-BP de $\mathbf{p_{n+1}}$. Consideremos la ecuación (4.1) que es la distribución de la S-BP de una partición no aleatoria. Sea Ω el conjunto de las permutación de $\{1, \ldots, n+1\}$. Luego

$$\mathbb{E}\left[\prod_{m=1}^{n+1} l_1^q\right] = \sum_{\sigma\in\Omega} \mathbb{E}\left[\prod_{k=1}^{n} \frac{p_{\sigma_k}^{q_k+1}}{1 - \sum_{j=1}^{k} p_{\sigma_j}} p_{\sigma_{n+1}}^{q_{n+1}+1}\right] \ .$$

Aunque calculable en teoría, algo de combinatoria es necesaria para evaluar el lado derecho de la última ecuación. Recordando que la S-BP se define recursivamente como un muestreo sesgado tamaño, la extracción del segmento escogido y la renormalización de los segmentos restantes, la última expresión se puede encontrar. Podemos calcular la función de momentos del i-ésimo segmento de la partición $\mathbf{l_{n+1}}$ condicional a los primeros i segmentos. Sean $p_{\sigma_1}, \ldots, p_{\sigma_i}$ los tamaños de los primeros $i-1$ segmentos de $\mathbf{l_{n+1}}$ y sea $\Xi = \{1, \ldots, n+1\} - \{\sigma_1, \ldots, \sigma_i\}$:

$$\begin{aligned} I\!\!E\left[l_{i+1}^q \mid l_1 = p_{\sigma_1}, \ldots, l_{i-1} = p_{\sigma_i}\right] &= I\!\!E\left[I\!\!E\left[l_{i+1}^q \mid \mathbf{p_{n+1}}, l_1 = p_{\sigma_1}, \ldots, l_{i-1} = p_{\sigma_i}\right]\right] \\ &= I\!\!E\left[\frac{1}{1 - \sum_{j=1}^{k} p_{\sigma_j}} \sum_{\sigma_{i+1} \in \Xi} p_{\sigma_{i+1}}\right] . \end{aligned}$$

Es una expresión recursiva de los momentos un poco más sencilla que la conjunta.

Reglas de autorganización con popularidades aleatorias

En este capítulo estudiaremos las estrategias para ordenar objetos, la regla de Move-to-Front (MtF) y de Move-to-Root (MtR). Imaginemos que tenemos una lista de n ítemes y que a cada instante t un objeto es requerido el cual, una vez usado, es desplazado al comienzo de la lista, esta es la estrategia MtF. Si en vez de considerar una lista se considera un árbol (de búsqueda binaria) para almacenar los ítemes, de modo que una vez que se esta disponible el ítem será desplazado a la raíz. Luego tendremos a groso modo la estrategia MtR. Pensando solo en la lista, supongamos que el objeto i es solicitado una fracción p_i del total de pedidos con $\sum_i p_i = 1$ formando $\mathbf{p_n} = (p_1, \ldots, p_n)$ el vector de popularidades. Si para encontrar el objeto vamos revisando la lista desde adelante hacia atrás, pasaremos menos tiempo buscando en la lista si ordenamos en una permutación ς tal que (p_{ς_i}) sea decreciente. Pero, si por algún motivo, no conocemos los valores de (p_i), podemos intuir que la regla MtF será una buena estrategia. Una situación similar ocurre para un árbol; si se conocen los valores de $\mathbf{p_n}$ Knuth propone (ver [36] pág. 433-477) un árbol de búsqueda binaria de costo mínimo. En el caso que no se conocen $\mathbf{p_n}$ se puede utilizar la regla MtR.

Supongamos que los objetos son solicitados independientemente de las solicitudes pasadas y que la probabilidad p_i de requerir el objeto i es constante en el tiempo. En este caso se puede modelar el problema como una cadena de Markov que itera sobre las permutaciones de los n objetos. Esta cadena de Markov también es conocida como como la librería de Tsetlin o proceso de Montón (“Heap process”). En 1965 McCabe estudia en [43] el proceso de MtF desde el punto de vista probabilista, motivado por sus aplicaciones para ordenar. Por su parte Tsetlin ya en el 1963 represento este proceso como un autómata finito reconociendo que se podía describir como una cadena de Markov (ver [60]). Por su parte la regla MtR fue propuesta por Allen y Munro en 1978 en [2], la cual explicaremos en detalle en la primera sección. Una revisión completa de ambas estrategias y sus generalizaciones se puede encontrar en la tesis de Bodell [11]. Las propiedades del modelo MtF fueron estudiadas por diferentes autores obteniendo el costo de búsqueda en régimen estacionario por McCabe en [43], la distribución estacionaria de la cadena por Hendricks en [31] (y Tsetlin [60]) y la función generatriz del costo de búsqueda por Flajolet *et al.* en [28] entre otras propiedades. En el caso de la regla MtR el análisis resulta más complejo, sin embargo Dobrow y Fill en [20] y [21] logran hacer un análisis sobre el proceso de Markov con estados en los posibles árboles binarios, su costo de búsqueda, distribución estacionaria y velocidad de convergencia.

Para ambas estrategias es difícil hacer un análisis más profundo sin hacer algún supuesto sobre $\mathbf{p_n}$, Fill estudia en [25] la regla MtF suponiendo ciertas distribuciones para las popularidades $\mathbf{p_n}$. Análogamente Dobrow y Fill hacen lo mismo en [21] para la regla de MtR. Esto motiva a enfrentar las estrategias a un escenario donde se tenga alguna noción del comportamiento de las popularidades sin conocer su valor exacto y preguntarnos que podemos decir del comportamiento de las reglas MtF y MtR. Es eso precisamente lo que hacemos al enfrentar a las estrategias de al caso general de las particiones aleatorias por renormalización definidas en el Capítulo 4. Nuestros resultados se reducen a estudiar el costo de búsqueda estacionario y su comportamiento asintótico cuando el numero de objetos n tiende a infinito.

Nuestro trabajo es similar al realizado por Papanicolaou *et al.* en [44] que estudiaron el problema del coleccionista de cupones para la misma partición aleatoria.

En este capítulo se recopilan los resultados del trabajo realizado en colaboración con Ch. Paroissin [7]) publicado en el "Journal of Applied Probability" y su continuación ([5]) realizada con la colaboración de Ch. Paroissin y T. Huillet a sido aceptadad por "Operating Research Letters". Así como los resultados obtenidos con Ch. Paroissin del poster ([8]) presentados en el "Third Colloquium on Mathematics and Computer Science Algorithms, Trees, Combinatorics and Probabilities". En la Sección 5.1 presentamos la motivación del problema, los modelos de las estrategias y las propiedades relevantes para nuestros resultados. En la Sección 5.2 estudiamos el comportamiento asintótico cuando el numero de elementos tiende a infinito y finalmente proveemos algunos ejemplos. Finalmente en la Sección 5.3 se derivan expresiones para el primer y segundo momentos del costo de búsqueda para la estrategia MtR (BS_n) y se dan algunos ejemplos con el comportamiento asintótico cuando es posible.

5.1. Reglas de auto-organización de datos

En esta sección veremos primero que las ventajas de la regla se pueden intuir en un modelo simple y luego introduciremos formalmente los modelos de las dos estrategias en el caso de popularidades deterministas.

5.1.1. Motivación: ¿Cómo y dónde almacenar libros?

Imaginemos que tenemos una repisa con nuestros libros (podría ser cualquier otro tipo de objeto). Cada vez que un libro es requerido recorremos la estantería de izquierda a derecha comparando uno a uno los títulos de los libros con el título de aquel que buscamos hasta dar con el libro requerido. Podemos decir que el tiempo que nos toma encontrar el libro es proporcional al número de comparaciones que tuvimos que hacer, es decir, la posición del menos uno.

Si consideramos que hay libros que consultaremos más frecuentemente que otros, nos gustaría ordenar los libros de modo de pasar poco tiempo en la búsqueda de aquel que necesitamos. Por supuesto que si supiéramos la frecuencia con que consultamos cada libro bastaría que los ordenemos de manera decreciente con respecto a su popularidad y minimizaríamos nuestro tiempo de búsqueda pues mientras más frecuentemente necesitemos un libro menos comparaciones haremos para encontrarlo. ¿Que pasa si no conocemos la frecuencia con que necesitaremos el libro? o ¿si cambia en el tiempo? o si estimarlas tiene un costo, dado este costo ¿vale la pena estimarlas?. Una posibilidad es que una vez que desocupamos un libro lo dejemos en su lugar original, pero esto no tomará en cuenta que hay libros más populares. Una solución intermedia es que cada vez que terminamos de utilizar un libro lo ponemos en la primera posición (a la izquierda de la estantería). De este modo si un libro se pide frecuentemente tenderá a quedar en el lado izquierdo y si un libro es rara vez solicitado tenderá a quedar a la derecha, esta es la regla de move-to-front (mover al frente). Sin embargo no es la única estrategia que sin conocer explícitamente los valores de las popularidades de los libros las tome en cuenta, también esta la regla de permutar (transposing o move-ahead-1) donde una vez que se desocupa un libro lo ponemos una posición más adelante de la que tenía cuando lo encontramos. Por supuesto que de la combinación de estas dos reglas pueden surgir muchas otras estrategias que pueden ser más o menos valiosas dependiendo del contexto del problema.

Si consideramos el mismo problema pero ahora nos permitimos almacenar nuestros libros en una estructura más sofisticada que una lista (representada por la repisa) podríamos disminuir aun más nuestro tiempo de búsqueda. Es así como surge la regla de Move-to- Root que ordena los libros en un árbol de búsqueda binaria.

5.1.2. El modelo de Move-to-Front

Consideremos una lista de n archivos (o libros) y denotemos el conjunto de archivos por $\{1, \ldots, n\}$ donde la popularidad de los elementos es constante en el tiempo. A cada instante de tiempo t se requiere un archivo, el i-ésimo archivo se requerirá con probabilidad p_i, como en un muestreo sesgado por tamaño en $\mathbf{p_n}$ (ver Definición 4.1.4). La lista de archivos es actualizada de acuerdo a la regla de MtF, es decir, cada vez que un archivo es requerido, una vez utilizado será desplazado a la primera posición de la lista. Luego el proceso se puede entender como una cadena de Markov donde el espacio de estados es el conjunto de las permutaciones del conjunto $\{1, \ldots, n\}$ y la probabilidad de transición entre dos permutaciones ς y ς' está dado por:

$$P(\sigma(t+1) = \varsigma' \mid \sigma(t) = \varsigma) = \begin{cases} p_i & \text{si } \varsigma' \text{ se deduce } \varsigma \text{ moviendo } i \text{ al frente ,} \\ 0 & \text{en otro caso .} \end{cases}$$

Este modelo también es conocido como la librería de Tsetlin o "Heap process" y se puede considerar el problema a tiempo continuo si entre un pedido y otro pasa un tiempo exponencial. Este proceso tiene una única distribución estacionaria: *la permutación sesgada por tamaño* de $\mathbf{p_n}$ definida en el Capítulo 4 en la Definición 4.1.4. La distribución estacionaria fue obtenida por Hendricks [31] pero ya había sido obtenida por Tsetlin en [60] en el contexto de autómatas (ver también [22] y [26]).

El desempeño que alcanza esta estructura esta dado por el tiempo que nos demoramos en encontrar el archivo requerido, que es el tiempo que demoramos en hacer una búsqueda lineal en la lista.

Definición 5.1.1. *El costo de búsqueda lineal $S_n(t)$ en la estrategia MtF se define el costo de búsqueda lineal en la permutación que genero la estrategia en el instante en que se solicita el ítem (ver Proposición 4.1.3).*

Luego si el objeto requerido al instante t se encuentra en la posición i el costo de buscarlo será $i-1$.

Este concepto de costo fue introducido por McCabe [43]. Notemos por S_n el costo de búsqueda cuando el proceso está en régimen estacionario. En la siguiente proposición se entrega el valor de la transformada de Laplace de S_n en el contexto de popularidades determinista y corresponde al Teorema 2 de [26] (también se puede encontrar en [11]):

Proposición 5.1.2. *La transformada de Laplace del costo S_n esta dado por,*

$$\mathbb{E}[\exp(-sS_n)\,|\,\omega] = \int_0^\infty e^{-t}\sum_{i=1}^n p_i^2 \prod_{\substack{k=1\\k\neq i}}^n \big(1+e^{-s}(e^{tp_k}-1)\big)\,dt\ , \tag{5.1}$$

para todo $s \geqslant 0$.

De está expresión se pueden obtener la esperanza y varianza de S_n:

$$\begin{aligned}\mathbb{E}[S_n] &= \sum_{i\neq j}\frac{p_ip_j}{p_i+p_j}\ ,\\ \mathrm{VAR}[S_n] &= \mathbb{E}[S_n]\,(1-\mathbb{E}[S_n]) + \\ &\qquad 4\sum_{i<j<k}\frac{p_ip_jp_k}{p_i+p_j+p_k}\left(\frac{1}{p_i+p_j}+\frac{1}{p_i+p_k}+\frac{1}{p_k+p_j}\right)\ .\end{aligned}$$

5.1.3. Relación con otras estructuras aleatorias

Como lo hicieron notar Flajolet *et al.* en [28](o también Fill y Holst en [26]), el costo de búsqueda $S_n(t)$ de la estrategia MtF está relacionado con la probabilidad de falla de la memoria Caché cuando se usa la estrategia "el menos recientemente usado" (en ingles Least-Recently-Used LRU). La memoria Caché es una memoria de rápido acceso de un computador que almacena algunos archivos que son frecuentemente requeridos. Si la memoria Caché puede almacenar k archivos de n, la estrategia LRU consiste en almacenar los k primeros de la la lista de n archivos ordenados de acuerdo a la regla MtF. De aquí se desprende que la probabilidad de que la memoria Caché de tamaño k fallé es $P(S_n(t) > k)$. Otra problema clásico de particiones que se relaciona con la estrategia MtF es el problema del coleccionista de cupones. Es fácil comprender por un argumento de acoplamiento que una vez que cada archivo sea requerido al menos una vez la estructura markoviana de la estrategia habrá "olvidado" el orden inicial por lo tanto estará en régimen estacionario. Determinar el instante en que todos los archivos han sido requeridos al menos una vez es el problema del coleccionista de cupones. Por último, como ya mencionamos la S-BP de la partición $\mathbf{p_n}$ definida en la Definición 4.1.4 es la distribución estacionaria de la estrategia MtF (ver [22]).

5.1.4. El Modelo de Move-to-Root

Consideremos un árbol binario que tiene la propiedad que de cada nodo solo salen dos ramas. Ordeno mis libros en este árbol de la siguiente forma: tomo el primer libro del estante y lo coloco en la raíz del árbol, el segundo libro lo cuelgo de la rama derecha si lexicográficamente el título del libro es mayor que el título del libro de la raíz y de la rama izquierda si es menor. Aplicamos esta regla hasta encontrar una rama vacía. De este modo en cada nodo del árbol todos los libros que cuelgan de la rama izquierda tienen títulos "mayores" que el título del libro del nodo considerado y los que cuelgan de la rama izquierda tienen títulos menores. El tiempo que nos demoraremos en encontrar un libro en esta estructura será, nuevamente, proporcional al número de comparaciones que debemos hacer que en el caso del árbol es la distancia a la raíz menos 1. Cuando necesitamos un libro lo tomamos y cuando lo guardamos lo ponemos en la raíz pero para no perder el orden lexicográfico del árbol y que siga siendo un árbol binario debemos realizar una serie de cambios en la estructura. Esta manera de ordenar los datos es conocida como MtR (desplazar hacia la raíz).

Formalmente un árbol con raíz es un grafo dirigido con un único nodo llamado raíz donde existe un único camino desde la raíz a cualquier nodo j. Cada nodo i, distinto de j, de este camino es considerado un ancestro de j, siendo el ancestro más cercano el padre del nodo. El subárbol de raíz i consiste de i y todos sus descendientes. Un árbol binario es un árbol ordenado en el cual cada nodo tiene a lo más dos hijos (uno a la izquierda y/o el en la derecha). Un nodo sin hijo es una hoja o nodo terminal.

Un árbol de búsqueda binaria es un árbol binario en el cual todos los ítemes etiquetados con valores menores que el etiquetado de la raíz quedan almacenados en el subárbol de la izquierda y aquellos ítemes etiquetados con valores mayores que la raíz se almacenan en el subárbol derecho. Luego, para una secuencia de ítemes, tenemos el siguiente algoritmo para construir un árbol binario de búsqueda (ver [36]):

Observación 5.1.3.

1. *Si no hay raíz, insertar el ítem como raíz;*
2. *Si el etiquetado es menor que él de la raíz, insertar el ítem en subárbol izquierdo;*
3. *Si el etiquetado es mayor que él de la raíz, insertar el ítem en subárbol derecho.*

Notemos que la construcción de árbol de búsqueda binaria depende del etiquetado pero también del orden en que se insertan los ítemes.

La siguiente manera de iterar el orden en un árbol binario de búsqueda es conocida como la regla de Move-To-Root y fue sugerida por Allen y Munro en [2]. Consideremos n ítemes que están ordenados en un árbol binario de búsqueda, a cada instante de tiempo (discreto) un ítem es requerido independiente de los requeridos anteriormente y es desplazado a la raíz del árbol binario de búsqueda respetando la estructura binaria y de orden del árbol. Esta estructura será actualizada de acuerdo al siguiente algoritmo que repetiremos hasta que el ítem requerido llegue a la raíz:

Observación 5.1.4. *Sea a el ítem requerido:*

1. *Si a es la raíz, no hacer nada;*
2. *si a es un hijo izquierdo sea r su padre, luego modificar el subárbol cuya raíz es r como se describe a continuación:*
   - *intercambiar a con r de modo que a será ahora la raíz del subárbol;*
   - *el antiguo subárbol izquierdo de a continuara como subárbol izquierdo de a;*
   - *el antiguo subárbol derecho de a pasará a ser el subárbol izquierdo de r;*
   - *el antiguo subárbol derecho r continuará subárbol derecho de r.*
3. *Si a es un hijo derecho sea r su padre, se modifica el subárbol cuya raíz es r con una transformación análoga.*

El objetivo de esta heurística es mantener el árbol de búsqueda binaria cercano a su forma optimal. La cadena de Markov asociada fue estudiada por Dobrow y Fill en [20, 21]. Dobrow extendió algunos de los resultados en [19] al caso en que los ítemes la probabilidad de requerir un ítem depende de t de forma markoviana.

Definición 5.1.5. *El costo de búsqueda lineal $BS_n(t)$ en la estrategia MtR se define como si al tiempo t el último ítem que se requirió fue i luego*

$$BS_n(t) = d(i, R) - 1 ,$$

donde $d(i, R)$ es la distancia del ítem i a la raíz.

Notaremos por BS_n por el costo de búsqueda en régimen estacionario.

Proposición 5.1.6. *El primer momento de BS_n es:*

$$\mathbb{E}[BS_n] = 2 \sum_{1 \leqslant i < j \leqslant n} \frac{p_i p_j}{p_i + \cdots + p_j} , \tag{5.2}$$

y el segundo

$$\mathbb{E}[S_n^2] = \mathbb{E}[S_n] + 4V , \tag{5.3}$$

donde V vale:

$$V = \sum_{1 \leqslant i < j < k \leqslant n} \frac{p_i p_j p_k}{p_i + \cdots + p_k} \left(\frac{1}{p_i + \cdots + p_j} + \frac{1}{p_j + \cdots + p_k} \right) .$$

Ver Teorema 3.1 en [2] (ver también [11]).

5.2. Move-to-Front con popularidades aleatorias

En esta sección primero derivaremos a partir de los resultados para particiones finitas, el comportamiento del costo de búsqueda estacionario y su comportamiento asintótico.

A lo largo de la sección consideremos el modelo de partición por normalización de la Sección 4, luego sea $\omega = \{\omega_i \ : \ i \in I\!N\}$ una sucesión de variables independientes, luego las popularidades aleatorias estarán dadas por:

$$p_i = \frac{\omega_i}{W_n} \quad \text{con } W_n = \sum_{i=1}^{n} \omega_i \ .$$

5.2.1. Expresión exacta para la transformada de Laplace

En esta subsección, estudiaremos la transformada de Laplace ϕ_{S_n} del costo de búsqueda estacionario para la estrategia Move-To-Front con popularidades aleatorias. El siguiente teorema entrega una representación integral exacta:

Teorema 5.2.1. *[7] Para la sucesión ω de variables aleatorias i.i.d.,*

$$\forall s \geqslant 0, \quad \phi_{S_n}(s) = n \int_0^\infty \int_t^\infty \phi''(r) \psi_{t,s}(r)^{n-1} \, dr \, dt \ , \tag{5.4}$$

donde:

$$\forall t \geqslant 0, \forall r \geqslant t, \quad \psi_{t,s}(r) = \phi(r) + e^{-s}(\phi(r-t) - \phi(r)) \ . \tag{5.5}$$

Demostración: De la ecuación 5.1 uno puede obtener la transformada de Laplace de S_n condicional a ω: $\forall s \geqslant 0$,

$$I\!E[\exp(-sS_n) \,|\, \omega] = \int_0^\infty e^{-t} \sum_{i=1}^{n} p_i^2 \prod_{\substack{k=1 \\ k \neq i}}^{n} \left(1 + e^{-s}(e^{tp_k} - 1)\right) dt \ .$$

Reemplazando $p_1, \ldots, p_n$ por su expresión en función de $\omega_1, \ldots, \omega_n$, obtenemos:

$$\mathbb{E}[\exp(-sS_n) \,|\, \omega] = \sum_{i=1}^{n} \int_0^\infty e^{-tW_n} \frac{w_i^2}{W_n} \prod_{\substack{k=1 \\ k \neq i}}^{n} \big(1 + e^{-s}(e^{tw_k} - 1)\big) \; dt \; .$$

Luego podemos deducir que la transformada de Laplace de S_n vale:

$$\mathbb{E}[\exp(-sS_n)] = \sum_{i=1}^{n} \int_0^\infty \mathbb{E}\left[e^{-tW_n} \frac{w_i^2}{W_n} \prod_{\substack{k=1 \\ k \neq i}}^{n} \big(1 + e^{-s}(e^{tw_k} - 1)\big) \right] \; dt \; . \tag{5.6}$$

En el caso de variables aleatorias idénticamente distribuidas, la ecuación (5.6) se reduce a:

$$\mathbb{E}[\exp(-sS_n)] = n \int_0^\infty \mathbb{E}\left[e^{-tW_n} \frac{w_n^2}{W_n} \prod_{k=1}^{n-1} \big(1 + e^{-s}(e^{tw_k} - 1)\big) \right] \; dt \; .$$

Sea Q la esperanza en la integral:

$$Q = \mathbb{E}\left[e^{-tW_n} \frac{w_n^2}{W_n} \prod_{k=1}^{n-1} \big(1 + e^{-s}(e^{tw_k} - 1)\big) \right] \; .$$

Podemos reescribir Q como:

$$\begin{aligned} Q &= \mathbb{E}\left[\int_t^\infty e^{-rW_n} \, dr \; w_n^2 \prod_{k=1}^{n-1} \big(1 + e^{-s}(e^{tw_k} - 1)\big) \right] \\ &= \int_t^\infty \mathbb{E}\left[e^{-rW_n} w_n^2 \prod_{k=1}^{n-1} \big(1 + e^{-s}(e^{tw_k} - 1)\big) \right] dr \\ &= \int_t^\infty \phi''(r) \left(\phi(r) + e^{-s}(\phi(r-t) - \phi(r))\right)^{n-1} \; dr \; , \end{aligned}$$

puesto que:

$$\exp(-tW_n) = \prod_{k=1}^{n} \exp(-tw_k) \; .$$

Reemplazando Q en la expresión correspondiente deducimos que la transformada de Laplace del costo de búsqueda vale:

$$\mathbb{E}[\exp(-sS_n)] = \int_0^{\infty} \int_t^{\infty} \phi''(r)\psi_{t,s}(r)^{n-1}\, dr\, dt \; .$$

$\square$

Cálculos similares a los recién expuestos se pueden hacer en el caso de pesos independientes, en este caso uno obtiene el siguiente resultado:

Teorema 5.2.2. *[7] Para la sucesión ω de variables aleatorias independientes,*

$$\forall s \geqslant 0, \quad \phi_{S_n}(s) = \sum_{i=1}^{n} \int_0^{\infty} \left(\int_t^{\infty} \phi_i''(r) \prod_{\substack{k=1 \\ k\neq i}}^{n} \psi_{t,s,k}(r)\, dr \right) dt \; ,$$

donde para todo $i \in \{1, \ldots, n\}$:

$$\forall t \geqslant 0, \forall r \geqslant t, \quad \psi_{t,s,i}(r) = \phi_i(r) + e^{-s}(\phi_i(r-t) - \phi_i(r)) \; .$$

Del Teorema 5.2.1, podemos derivar una representación integral para los momentos de orden 1 y 2 del costo de búsqueda estacionario:

Corolario 5.2.3. *[7] Para la sucesión ω de variables aleatorias i.i.d.,*

$$\mathbb{E}[S_n] = n(n-1) \int_0^{\infty} \phi(r)^{n-2} \left(\int_r^{\infty} (\phi'(t))^2\, dt \right) dr \tag{5.7}$$

y

$$\mathbb{E}[S_n^2] = n(n-1)(2n-3)\int_0^\infty \phi(r)^{n-2}\left(\int_r^\infty (\phi'(t))^2\,dt\right)dr$$
$$-2n(n-1)(n-2)\int_0^\infty \phi(r)^{n-3}\left(\int_r^\infty \phi(t)(\phi'(t))^2\,dt\right)dr\;. \tag{5.8}$$

Demostración:

1. De la ecuación (5.4), obtenemos que, para cualquier $s \geqslant 0$,

$$\phi'_{S_n}(s) = -n(n-1)\int_0^\infty\int_t^\infty \phi''(r)\psi_{t,s}(r)^{n-2}e^{-s}(\phi(r-t)-\phi(r))\,dr\,dt\;. \tag{5.9}$$

   Evaluado esta expresión en $s=0$:

$$\begin{aligned}
\phi'_{S_n}(0) &= -n(n-1)\int_0^\infty\int_t^\infty \phi''(r)\phi(r-t)^{n-2}(\phi(r-t)-\phi(r))\,dr\,dt \\
&= -n(n-1)\int_0^\infty\int_0^\infty \phi''(r+t)\phi(r)^{n-2}(\phi(r)-\phi(r+t))\,dr\,dt \\
&= -n(n-1)\int_0^\infty \Big[\phi'(r)\phi(r)^{n-1} \\
&\qquad -\phi(r)^{n-2}\int_0^\infty \phi''(r+t)\phi(r+t)\,dt\Big]\,dr \\
&= -n(n-1)\int_0^\infty \phi(r)^{n-2}\left(\int_0^\infty (\phi'(r+t))^2\,dt\right)dr \\
&= -n(n-1)\int_0^\infty \phi(r)^{n-2}\left(\int_r^\infty (\phi'(t))^2\,dt\right)dr\;.
\end{aligned}$$

   Luego obtenemos la ecuación (5.7), pues $\mathbb{E}[S_n] = -\phi'_{S_n}(0)$.

2. Derivando ϕ'_{S_n} una vez más de la ecuación (5.9) obtenemos:

$$\begin{aligned}\phi''_{S_n}(s) = n(n-1)(n-2)&\int_0^\infty\int_t^\infty \phi''(r)\psi_{t,s}(r)^{n-3}e^{-s}(\phi(r-t)-\phi(r))^2\,dr\,dt\\ +n(n-1)&\int_0^\infty\int_t^\infty \phi''(r)\psi_{t,s}(r)^{n-2}e^{-s}(\phi(r-t)-\phi(r))\,dr\,dt\ .\end{aligned}$$

Evaluando en $s=0$:

$$\begin{aligned}\phi''_{S_n}(0) = n(n-1)(n-2)&\int_0^\infty\int_t^\infty \phi''(r)\phi(r-t)^{n-3}(\phi(r-t)-\phi(r))^2\,dr\,dt\\ +n(n-1)&\int_0^\infty\int_t^\infty \phi''(r)\phi(r-t)^{n-2}(\phi(r-t)-\phi(r))\,dr\,dt\ .\end{aligned}$$

Notemos por A y B estas dos integrales. Ya hemos calculado B en primera parte de esta demostración. Calculemos A:

$$\begin{aligned}A &= \int_0^\infty\int_t^\infty \phi''(r)\phi(r-t)^{n-3}(\phi(r-t)-\phi(r))^2\,dr\,dt\\ &= \int_0^\infty\int_0^\infty \phi''(r+t)\phi(r)^{n-3}(\phi(r)-\phi(r+t))^2\,dr\,dt\\ &= \int_0^\infty\int_0^\infty \phi''(r+t)\phi(r)^{n-3}(\phi(r)^2-2\phi(r)\phi(r+t)+\phi(r+t)^2)\,dr\,dt\\ &= 2\int_0^\infty \phi(r)^{n-2}\left(\int_r^\infty(\phi'(t))^2\,dt\right)dr\\ &\qquad -2\int_0^\infty \phi(r)^{n-3}\left(\int_r^\infty \phi(t)(\phi'(t))^2\,dt\right)dr\ .\end{aligned}$$

Combinando las expresiones de A y B deducimos la expresión de la ecuación (5.8), puesto que $\mathbb{E}[S_n^2]=\phi''_{S_n}(0)$.

$\square$

Cálculos análogos se pueden hacer para los momentos de ordenes mayores de S_n. Las expresiones para los dos primeros momentos de S_n del Corolario 5.2.3 se pueden calcular directamente usando los resultados de McCabe en [43]. En el caso en que los pesos no son idénticamente distribuidos pero siguen siendo independientes, uno puede realizar cálculos similares para encontrar resultados análogos a los del Corolario 5.2.3.

Teorema 5.2.4. *[7] Para una sucesión $\omega = (\omega_i)_{i \in I\!N}$ de variables aleatorias independientes,*

$$I\!E[S_n] = \sum_{\substack{i,j=1 \\ i \neq j}}^{n} \int_0^\infty \left(\int_t^\infty \phi_i'(r) \phi_j'(r) \, dr \right) \prod_{\substack{k=1 \\ k \neq i,j}}^{n} \phi_k(t) \, dt \ . \tag{5.10}$$

Como ya dijimos uno puede obtener esta expresión del costo de búsqueda esperado ya sea usando el Teorema 5.2.2 o directamente de los resultados de McCabe.

5.2.2. Fórmula asintótica para la transformada de Laplace

Nos interesa conocer el comportamiento asintótico del costo de búsqueda cuando el número de archivos es creciente. Primero, daremos un equivalente asintótico puntual de la transformada de Laplace de S_n. Este resultado será usado para obtener los límites de los dos primeros momentos de S_n.

Teorema 5.2.5. *[7] Para una sucesión $\omega = (\omega_i)_{i \in I\!N}$ variables aleatorias i.i.d.,*

$$\forall s \geqslant 0, \quad \phi_{S_n}(s) \sim - \int_0^\infty \frac{\phi''(r) \psi_{r,s}(r)^n}{\psi_{r,s}'(r)} \, dr \ , \tag{5.11}$$

cuando n tiende a infinito.

Demostración: Sea $s \geqslant 0$ fijo:

$$\phi_{S_n}(s) = n \int_0^\infty I_n(t)\, dt \ ,$$

donde:

$$I_n(t) = \int_t^\infty \phi''(r)\psi_{t,s}(r)^{n-1}\, dr \ .$$

Sea $t \geqslant 0$ fijo. Hemos calculado un equivalente asintótico de $I_n(t)$ cuando n tiende a infinito. Primero reformulemos $I_n(t)$ como:

$$I_n(t) = \int_t^\infty h_1(r) e^{(n-1)h_2(r)}\, dr \ ,$$

donde $h_1(r) = \phi''(r)$ y $h_2(r) = \log \psi_{t,s}(r)$, para $r \geqslant t$. Las integrales como $I_n(t)$ son llamadas integrales de Laplace generalizadas ([9], section 6.4, p. 261-276). Como ϕ' es creciente y negativa, $\psi_{t,s}$ es decreciente para cualquier valor de t y s. Luego h_2 es también decreciente y alcanza su máximo en $r = t$. Luego su principal contribución en $I_n(t)$ será en una vecindad de $r = t$ siempre que $h_1(t) \neq 0$. Esto último es cierto pues:

$$\forall t \geqslant 0, \quad h_1(t) = \phi''(t) = I\!E\left[w_1^2 e^{-tw_1}\right] > 0 \ .$$

Sea $\varepsilon \in (0,1)$ fijo. Tenemos la siguiente descomposición:

$$I_n(t) = \int_t^{t+\varepsilon} h_1(r) e^{(n-1)h_2(r)}\, dr \times \left(1 + \frac{\displaystyle\int_{t+\varepsilon}^\infty h_1(r) e^{(n-1)h_2(r)}\, dr}{\displaystyle\int_t^{t+\varepsilon} h_1(r) e^{(n-1)h_2(r)}\, dr} \right) \ . \tag{5.12}$$

Cuando h_1 es una función no negativa,

$$\frac{\displaystyle\int_{t+\varepsilon}^{\infty} h_1(r)e^{(n-1)h_2(r)}\,dr}{\displaystyle\int_{t}^{t+\varepsilon} h_1(r)e^{(n-1)h_2(r)}\,dr} \leqslant \frac{\displaystyle\int_{t+\varepsilon}^{\infty} h_1(r)e^{(n-1)h_2(r)}\,dr}{\displaystyle\int_{t}^{t+\varepsilon/2} h_1(r)e^{(n-1)h_2(r)}\,dr} \,.$$

Luego h_2 es decreciente, luego tenemos:

$$\frac{\displaystyle\int_{t+\varepsilon}^{\infty} h_1(r)e^{(n-1)h_2(r)}\,dr}{\displaystyle\int_{t}^{t+\varepsilon/2} h_1(r)e^{(n-1)h_2(r)}\,dr} \leqslant e^{(n-1)(h_2(t+\varepsilon)-h_2(t+\varepsilon/2))} \times \frac{\displaystyle\int_{t+\varepsilon}^{\infty} h_1(r)\,dr}{\displaystyle\int_{t}^{t+\varepsilon/2} h_1(r)\,dr} \,.$$

El lado derecho de está desigualdad tiende a 0 cuando n tiende a infinito pues que h_2 es decreciente implica que $h_2(t+\varepsilon) - h_2(t+\varepsilon/2) < 0$. Luego obtenemos:

$$I_n(t) \sim \int_{t}^{t+\varepsilon} h_1(r)e^{(n-1)h_2(r)}\,dr \ ,$$

cuando n tiende a infinito. Usando una expansión de Taylor de h_2 en torno a $r = t$, obtenemos el siguiente equivalente asintótico:

$$I_n(t) \sim \int_{t}^{t+\varepsilon} h_1(t)e^{(n-1)(h_2(t)+(r-t)h_2'(t))}\,dr \ ,$$

cuando n tiende a infinito. Esto es posible debido a que $h_1(t) \neq 0$. Recordemos que $h_1(t) = \phi''(t)$ y $h_2'(t) = \psi_{t,s}'(t)/\psi_{t,s}(t)$ el cual es negativo. Luego obtenemos:

$$I_n(t) \sim \phi''(t)\psi_{t,s}(t)^{n-1} \int_{t}^{t+\varepsilon} e^{(n-1)(r-t)\psi_{t,s}'(t)/\psi_{t,s}(t)}\,dr \ ,$$

cuando n tiende a infinito. Usando el mismo tipo de descomposición que en la ecuación (5.12), obtenemos:

$$\begin{aligned}\int_t^{t+\varepsilon} e^{(n-1)(r-t)h_2'(t)}\,dr &= \int_t^{\infty} e^{(n-1)(r-t)h_2'(t)}\,dr \times \left(1 - \frac{\int_{t+\varepsilon}^{\infty} e^{(n-1)(r-t)h_2'(t)}\,dr}{\int_t^{\infty} e^{(n-1)(r-t)h_2'(t)}\,dr}\right) \\ &= \int_t^{\infty} e^{(n-1)(r-t)h_2'(t)}\,dr \times \left(1 - e^{(n-1)h_2'(t)\varepsilon}\right) \ .\end{aligned}$$

Finalmente se obtiene la siguiente aproximación:

$$I_n(t) \sim -\frac{\phi''(t)\psi_{t,s}(t)^n}{(n-1)\psi'_{t,s}(t)} \ ,$$

cuando n tiende a infinito. Luego deducimos que:

$$\phi_{S_n}(s) \sim -\int_0^{\infty} \frac{\phi''(r)\psi_{r,s}(r)^n}{\psi'_{r,s}(r)}\,dr \ ,$$

cuando n tiende a infinito. □

De este expresión podemos calcular expresiones asintóticas para los momentos de orden 1 y 2 del costo de búsqueda estacionario:

Corolario 5.2.6. *[7] Para la sucesión ω de variables aleatorias i.i.d.,*

$$\frac{1}{n-1}\mathbb{E}[S_n] \xrightarrow[n\to\infty]{} \frac{1}{\mu}\int_0^{\infty} (\phi'(r))^2\,dr \ , \qquad (5.13)$$

y

$$\frac{1}{(n-1)^2}\mathbb{E}[S_n^2] \xrightarrow[n\to\infty]{} \frac{2}{\mu}\int_0^{\infty} (1-\phi(r))(\phi'(r))^2\,dr \ , \qquad (5.14)$$

Como S_n toma valores en $\{0,\ldots,n-1\}$, tiene más sentido poner $n-1$ en vez de n en el corolario.

Demostración:

1. De la ecuación (5.11) podemos obtener, para cualquier $s \geqslant 0$,

$$\phi'_{S_n}(s) \sim \int_0^\infty \phi''(r)e^{-s}\left[\frac{n(1-\phi(r))\psi_{r,s}(r)^{n-1}}{\psi'_{r,s}(r)} + \frac{(\mu+\phi'(r))\psi_{r,s}(r)^n}{(\psi'_{r,s}(r))^2}\right] dr\ . \tag{5.15}$$

   Evaluando la última expresión en $s=0$

$$\phi'_{S_n}(0) \sim \int_0^\infty \phi''(r)\left[\frac{-n(1-\phi(r))}{\mu} + \frac{(\mu+\phi'(r))}{\mu^2}\right] dr\ .$$

   Cuando n tiende a infinito, el segundo termino en la integral es despreciable, ya que:

$$\int_0^\infty (\phi'(r))^2\, dr \neq 0\ .$$

   Luego,

$$\phi'_{S_n}(0) \sim -\frac{n}{\mu}\int_0^\infty \phi''(r)(1-\phi(r))\, dr\ .$$

   Integrando por partes la última expresión:

$$\int_0^\infty \phi''(r)(1-\phi(r))\, dr = \int_0^\infty (\phi'(r))^2\, dr\ .$$

   Como $\mathbb{E}[S_n] = -\phi'_{S_n}(0)$ obtenemos (5.13).

2. Derivando la expresión de la ecuación (5.15):

$$\phi''_{S_n}(s) \sim -\phi'_{S_n}(s) + \int_0^\infty \phi''(r)e^{-s}\left[-\frac{n(n-1)(1-\phi(r))^2\psi_{r,s}(r)^{n-2}}{\psi'_{r,s}(r)}\right.$$
$$\left.-\frac{n(1-\phi(r))(\mu+\phi'(r))\psi_{r,s}(r)^{n-1}}{(\psi'_{r,s}(r))^2}(1+e^{-s}) + \frac{2e^{-s}(\mu+\phi'(r))^2\psi_{r,s}(r)^n}{(\psi'_{r,s}(r))^4}\right] dr\ .$$

Evaluando en $s = 0$:

$$\phi''_{S_n}(0) \sim -\phi'_{S_n}(0) + \int_0^\infty \phi''(r)\left[\frac{n(n-1)(1-\phi(r))^2}{\mu}\right.$$
$$\left.-\frac{2n(1-\phi(r))(\mu+\phi'(r))}{\mu^2} + \frac{2(\mu+\phi'(r))^2}{\mu^4}\right] dr\ .$$

Solo el primer termino de la segunda integral domina pues

$$\int_0^\infty (1-\phi(r))(\phi'(r))^2\, dr \neq 0\ .$$

Luego,

$$\phi''_{S_n}(0) \sim \frac{n(n-1)}{\mu}\int_0^\infty \phi''(r)(1-\phi(r))^2\, dr\ .$$

Integrando por partes la expresión anterior concluimos la aproximación de $\mathbb{E}(S_n^2)$ en (5.14).

$\square$

El resultado presentado en el Corolario 5.2.6 donde se dan los valores de los límites cuando n tiende a infinito de los dos primeros momentos de S_n/n, sugiere que quizás el límite cuando n tiende a infinito de la variable S_n/n puede existir. Este resultado se exploró en el artículo [4] y es la continuación de esta sección.

De el Teorema 5.2.1, podemos obtener la siguiente expresión exacta para la función densidad de la distribución límite de S_n/n:

Teorema 5.2.7. *[5] Para una sucesión $\omega = (\omega_i)_{i \in I\!N}$ de pesos aleatorios i.i.d. con media finita μ,*

$$\frac{S_n}{n} \xrightarrow[n \to \infty]{d} S ,$$

donde S es una variable aleatoria continua con la siguiente función densidad f_S:

$$f_S(x) = -\frac{1}{\mu} \frac{\phi''\left(\phi^{-1}(1-x)\right)}{\phi'\left(\phi^{-1}(1-x)\right)} \mathbb{1}_{[0,1-p_0]}(x) , \tag{5.16}$$

Donde $p_0 = I\!P(\omega_i = 0)$ y ϕ^{-1} es la función inversa de ϕ.

Observación 5.2.8. *El hecho que el intervalo $[0, 1 - p_0]$ es el soporte de S puede entenderse de manera intuitiva: la cantidad p_0 es la probabilidad que un ítem nunca sea requerido. En régimen estacionario, uno esperaría que un ítem que nunca fue requerido se ubique al final de la lista; np_0 es el número esperados de ítemes jamás requeridos. Luego no es sorprendente que el soporte de S no sea todo el intervalo $[0, 1]$. Notemos que en el caso en que la distribución de los pesos es continua $p_0 = 0$.*

Demostración: Debemos demostrar que S_n/n converge en distribución, cuando n tiende a infinito, a cierta variable aleatoria que vamos a notar por S. Primero, observemos que:

$$\forall s \geqslant 0 , \quad \phi_{S_n/n}(s) = \phi_{S_n}\left(\tfrac{s}{n}\right) .$$

Luego estamos interesados en el límite de $\phi_{S_n}(s/n)$.

Para cualquier par de reales a y b tal que $0 \leqslant a \leqslant b \leqslant \infty$, sea:

$$I_n(a,b) = \int_a^b \phi''(r)\left[\phi(r) + e^{-s/n}(\phi(r-t) - \phi(r))\right]^{n-1} dr\,.$$

Si $b = \infty$, luego omitiremos este parámetro, i.e. $I_n(a) = I_n(a,\infty)$. Usando esta notación, el Teorema 5.2.1 queda:

$$\phi_{S_n}\left(\tfrac{s}{n}\right) = n\int_0^\infty I_n(t)\,dt\,. \tag{5.17}$$

Ahora podemos descomponer $I_n(t)$ en dos partes: $I_n(t) = I_n(t, t+\varepsilon) + I_n(t+\varepsilon)$. Vamos a demostrar que $nI_n(t+\varepsilon)$ tiende a 0 cuando n tiende a infinito:

$$\begin{aligned} nI_n(t+\varepsilon) &= n\int_{t+\varepsilon}^\infty \phi''(r)\left[e^{-s/n}\phi(r-t) + (1-e^{-s/n})\phi(r)\right]^{n-1} dr\,,\\ &\leqslant n\int_{t+\varepsilon}^\infty \phi''(r)\phi(r-t)^{n-1}\,dr\,,\\ &\leqslant -n\phi(\varepsilon)^{n-1}\phi'(t+\varepsilon)\,,\end{aligned}$$

debido a que ϕ es decreciente. Luego $\lim_{n\to\infty} nI_n(t+\varepsilon) = 0$, para todo $\varepsilon > 0$.

Ahora estimaremos $I_n(t, t+\varepsilon)$. Sea $h_n(r,t) = \phi(r) + e^{-s/n}(\phi(r-t) - \phi(r))$. Para un valor fijo de t, la función $h_n(\cdot, t)$ se comporta como ϕ. En particular $\frac{\partial h_n}{\partial r}$ es una función creciente para $r \in [t, t+\varepsilon]$. Luego obtenemos la siguiente cota:

$$\frac{\partial h_n}{\partial r}(t,t) \leqslant \frac{\partial h_n}{\partial r}(r,t) \leqslant \frac{\partial h_n}{\partial r}(t+\varepsilon,t)\,,$$

y

$$\phi''(t+\varepsilon) \leqslant \phi''(r) \leqslant \phi''(t)\,.$$

Luego, podemos acotar $I_n(t, t+\varepsilon)$ por:

$$\begin{aligned} I_n(t, t+\varepsilon) &= \int_t^{t+\varepsilon} \phi''(r)\,(h_n(r,t))^{n-1}\,\frac{\partial h_n}{\partial r}(r,t)\frac{\partial h_n}{\partial r}(r,t)^{-1}\,dr \\ &\leqslant \phi''(t)\frac{\partial h_n}{\partial r}(t,t)^{-1}\int_t^{t+\varepsilon} (h_n(r,t))^{n-1}\,\frac{\partial h_n}{\partial r}(r,t)\,dr \\ &\leqslant \phi''(t)\frac{\partial h_n}{\partial r}(t,t)^{-1}\frac{1}{n}\left[(h_n(t+\varepsilon,t))^n - (h_n(t,t))^n\right] \ . \end{aligned}$$

Estas cotas son validas para cualquier $\varepsilon > 0$; tomando el límite cuando ε tiende a 0 deducimos que límite de $nI_n(t)$ existe y esta dado por:

$$\lim_{n\to\infty} nI_n(t) = \frac{\phi''(t)}{\mu}\exp\left(-(1-\phi(t))s\right) \ .$$

Procediendo de un modo similar, podemos encontrar una cota inferior:

$$I_n(t, t+\varepsilon) \geqslant \phi''(t+\varepsilon)\frac{\partial h_n}{\partial r}(t+\varepsilon,t)^{-1}\frac{1}{n}\left[(h_n(t+\varepsilon,t))^n - (h_n(t,t))^n\right] \ .$$

Luego, para cualquier $\varepsilon > 0$, uno puede probar que se tiene el siguiente límite:

$$\begin{aligned} \lim_{n\to\infty} (h_n(t+\varepsilon,t))^n &= 0 \ , \\ \lim_{n\to\infty} (h_n(t,t))^n &= \exp\left[-s(1-\phi(t))\right] \ , \\ \lim_{n\to\infty} \frac{\partial h_n}{\partial r}(t,t) &= \phi'(0) \ , \\ \lim_{n\to\infty} \frac{\partial h_n}{\partial r}(t+\varepsilon,t) &= \phi'(\varepsilon) \ . \end{aligned}$$

Reemplazando estos límites en las ecuaciones previas y recordando que $nI_n(t+\varepsilon)$ es 0 para todo $\varepsilon > 0$, obtenemos que los valores límites de $nI_n(t)$ cuando n tiende a infinito puede ser acotados:

$$\begin{aligned}
\liminf_{n\to\infty} nI_n(t) &= \liminf_{n\to\infty} nI_n(t,t+\varepsilon) \geqslant -\frac{\phi''(t+\varepsilon)}{\phi'(0)} \exp\left(-(1-\phi(t))s\right) \ , \\
\limsup_{n\to\infty} nI_n(t) &= \limsup_{n\to\infty} nI_n(t,t+\varepsilon) \leqslant -\frac{\phi''(t)}{\phi'(\varepsilon)} \exp\left(-(1-\phi(t))s\right) \ .
\end{aligned}$$

Esto es cierto para todo $\varepsilon > 0$; luego tomando ε tendiendo a 0, tenemos:

$$\lim_{n\to\infty} nI_n(t) = \frac{\phi''(t)}{\mu} \exp\left(-(1-\phi(t))s\right) \ .$$

Reemplazando este límite en la ecuación (5.17) obtenemos

$$\lim_{n\to\infty} \phi_{S_n/n}(s) = \frac{1}{\mu} \int_0^\infty \phi''(t) e^{-(1-\phi(t))s}\, dt \ , \tag{5.18}$$

el cual denotaremos por $\phi_S(s)$. Aunque este límite a priori no es necesariamente la transformada de Laplace de una variable aleatoria, de acuerdo al Teorema de Continuidad (pág. 431 Ch. XIII en [24]), será suficiente verificar que $\lim_{s\to 0} \phi_S(s) = 1$, lo que puede ser demostrado usando el teorema de convergencia dominada.

Realizando el cambio de variable apropiado $y = 1 - \phi(r)$ en la ecuación (5.18) nos da:

$$\phi_S(s) = -\frac{1}{\mu} \int_0^{1-p_0} \frac{\phi''\left(\phi^{-1}(1-y)\right)}{\phi'\left(\phi^{-1}(1-y)\right)} e^{-ys}\, dr,$$

donde para los límites de la integral usamos la propiedad siguiente $\phi(\infty) = p_0$ (ver [24] observación en el Teorema 1(a) pág. 439 Ch. XIII). Entonces, tenemos que:

$$f_S(y) = -\frac{1}{\mu} \frac{\phi''\left(\phi^{-1}(1-y)\right)}{\phi'\left(\phi^{-1}(1-y)\right)} \mathbb{1}_{[0,1-p_0]}(y)$$

es la función densidad de la variable S. $\square$

Como corolario del teorema, podemos calcular el q-ésimo momento y la función de distribución de S:

Corolario 5.2.9. *[5] Para $q \in \mathbb{R}$,*

$$E[S^q] = \frac{1}{\mu}\int_0^\infty (1-\phi(t))^q \phi''(t)\, dt \ ,$$

y, para todo $x \in [0,1]$,

$$\mathbb{P}(S \leqslant x) = \left(1 + \frac{1}{\mu}\phi'(\phi^{-1}(1-x))\right) \mathbb{1}_{[0,1-p_0]}(x) + \mathbb{1}_{(1-p_0,1]}(x) \ .$$

La función de acumulación de S despierta un interés particular la regla de MtF está relacionada con la estrategia de LRU (ver por ejemplo [28]). Como ya explicamos en la Subsección 5.1.3 una de las estrategias utilizadas para decidir que elementos se almacenan en la memoria Caché es la estrategia LRU. En el contexto de esta estrategia de organización de la memoria Caché una pregunta importante surge: ¿cual es la probabilidad que el archivo requerido no esté en la memoria caché? La probabilidad de este evento es llamada "page defaul" y en lo que sigue la notaremos por π_k. A causa de la relación existente entre la regla de MtF y la estrategia LRU, como se mencionó, tenemos que $\pi_k = \mathbb{P}(S_n \geqslant k)$. Luego, si asumimos que el tamaño de la memoria Caché es proporcional al número de archivos, digamos $k = \alpha n$ con $\alpha \in [0,1]$ fijo, para un gran número de archivos, se tiene la siguiente aproximación:

$$\pi_{\alpha n} \simeq 1 + \frac{1}{\mu}\phi'(\phi^{-1}(1-\alpha))$$

si $\alpha < p_0$ y $\pi_{\alpha n} \simeq 1$.

5.2.3. Ejemplos y algunas propiedades

En esta sección, estudiamos algunos ejemplos para los cuales podemos encontrar expresiones explícitas para los cálculos. Consideraremos ambos casos, distribuciones continuas y discretas para los pesos aleatorios.

Ejemplo 5.2.10. *Supongamos que los pesos tienen una distribución de Dirac con masa en* 1 *(en otras palabras, los pesos son deterministas y son igualmente populares). Entonces* $\phi(r) = e^{-r}$, $\mu = 1$ *y* $p_0 = 0$*; deducimos que:*

$$f_{S_1}(x) = \mathbb{1}_{[0,1]}(x) \ .$$

Luego, S_1 *tiene la distribución uniforme con soporte en* $[0,1]$*: este resultado fue demostrado por Fill (Teorema 4.2, p. 198 de [25]). El k-ésimo momento (con* $k \in \mathbb{R}_+$*) y la función de acumulación de* S_1 *es:*

$$\mathbb{E}[S_1^k] = \frac{1}{k+1} \qquad y \qquad \forall x \in [0,1]\,, \quad F_{S_1}(x) = \mathbb{P}(S_1 \leqslant x) = x \ .$$

Ejemplo 5.2.11. *Consideremos un ejemplo que difiere poco del anterior. Supongamos que los pesos tiene una distribución de Bernoulli con parámetro* $p \in (0,1]$*. Luego,* $\phi(r) = pe^{-r} + (1-p)$*, la esperanza vale* $\mu = p$ *y* $p_0 = 1-p$*. Para una gran cantidad de archivos, la proporción* p *de archivo que será requerida con la misma frecuencia, los restantes nunca serán requeridos. Luego, uno puede pensar que uno tendrá una distribución límite para* S *no muy distinta que aquella del ejemplo* *anterior. En efecto, un calculo simple nos muestra que:*

$$f_{S_2}(x) = \frac{1}{p}\mathbb{1}_{[0,p]}(x) \; .$$

Luego, S_2 tiene una distribución uniforme en el intervalo $[0,p]$. Obviamente, tenemos que $S_2 \preccurlyeq_{st} S_1$ (donde $\preccurlyeq_{st}$ denota el orden estocástico usual).

Ejemplo 5.2.12. *Supongamos que los pesos tienen una distribución Gamma de parámetro $\alpha > 0$. En este ejemplo, el vector aleatorio $(p_1,\ldots,p_n)$ tiene una distribución de Dirichlet simétrica $D_n(\alpha)$ (ver Sección 4). En este caso, $\mu = \alpha$, $p_0 = 0$ y $\phi(r) = (1+r)^{-\alpha}$. Nuestros cálculos dan:*

$$f_{S_3}(x) = \left(1 + \frac{1}{\alpha}\right)(1-x)^{1/\alpha}\mathbb{1}_{[0,1]}(x) \; ,$$

la cual es la función densidad de una variable Beta de parámetros $(1, 1+1/\alpha)$. Notemos que este resultado ya fue demostrado en la Sección 4.3 Teorema 4.3.8 [4] usando las propiedades de la distribución de Dirichlet (en este caso calculamos la distribución del costo de búsqueda estacionario para cualquier n finito). El k-ésimo momento (con $k \in \mathbb{R}_+$) de S_3 es:

$$\mathbb{E}[S_3^k] = \frac{\Gamma(k+1)\Gamma(2+\frac{1}{\alpha})}{\Gamma(2+k+\frac{1}{\alpha})} \; .$$

En particular $\mathbb{E}[S_3] = \alpha/(2\alpha+1)$ y $Var[S_3] = ((\alpha+1)\alpha^2)/((3\alpha+1)(2\alpha+1)^2)$. Uno puede calcular también la función de acumulación de la variable S_3, para cualquier $x \in [0,1]$:

$$F_{S_3}(x) = 1 - (1-x)^{1+1/\alpha} \; .$$

Fácilmente podemos deducir que, para cualquier $x \in [0,1]$, $\bar{F}_{S_3}(x) \leqslant \bar{F}_{S_1}(x)$. Luego *encontramos la siguiente desigualdad $S_3 \preceq_{st} S_1$.*

Observemos que si el parámetro α tiende a infinito, luego el vector aleatorio de las popularidades aleatorias $(p_1, \ldots, p_n)$ converge en distribución a $(\frac{1}{n}, \ldots, \frac{1}{n})$: mientras más grande el valor de α, más se concentran $(p_1, \ldots, p_n)$ en torno a su esperanza. Luego es razonable que S_3 converja en distribución a S_1 cuando α tiende a infinito. De hecho, uno puede probar fácilmente que este resultado es cierto en el caso general:

Proposición 5.2.13. *[5] Sea $\omega(\alpha) = (w_i(\alpha))_{i\in\mathbb{N}^*}$ una secuencia de pesos aleatorios i.i.d. tal que $w_1(\alpha)$ converge en distribución a w cuando α tiende a infinito. Luego, $S(\alpha)$ converge en distribución a S.*

Ejemplo 5.2.14. *Supongamos que los pesos tienen una distribución Geométrica en $\mathbb{N}$ de parámetro $p \in (0,1)$. En este caso $\mu = (1-p)/p$, $p_0 = p$ y $\phi(r) = p/(1-(1-p)e^{-r})$. Un simple cálculo nos da:*

$$f_{S_4}(x) = \frac{2(1-x)-p}{1-p} \mathbb{1}_{[0,1-p]}(x) \ .$$

El k-ésimo momento (con $k \in \mathbb{R}_+$) de S_3 es:

$$\mathbb{E}[S_4^k] = \frac{(2+pk)(1-p)^k}{(k+1)(k+2)} \ .$$

En particular, tenemos $\mathbb{E}[S_4] = \frac{(2+p)(1-p)}{6}$ y $Var[S_4] = \frac{(1-p)^2(2+2p-p^2)}{36}$. Uno puede también calcular la función de acumulación de S_4, para cualquier $x \in [0,1]$, tenemos:

$$F_{S_4}(x) = \frac{x(2-p-x)}{1-p} \mathbb{1}_{[0,1-p]}(x) + \mathbb{1}_{(1-p,1]}(x) \ .$$

De esta última expresión podemos verificar la desigualdad $S_4 \preccurlyeq_{st} S_1$.

Ejemplo 5.2.15. *Supongamos que los pesos siguen una distribución de Poisson de* *parámetro λ. En este caso, $\mu = \lambda$, $p_0 = e^{-\lambda}$ y $\phi(r) = \exp(\lambda e^{-r} - 1)$. Un sencillo cálculo muestra que:*

$$f_{S_5}(x) = \frac{\ln(1-x) + \lambda + 1}{\lambda} 1\!\!1_{[0,1-e^{-\lambda}]}(x) \ .$$

Usando las formula 1.6.5.3 de [50] (pág. 244), uno puede calcular el k-ésimo momento (con $k \in I\!N$) de S_5:

$$I\!E[S_5^k] = \frac{1}{\lambda(k+1)} \left[\lambda + (1-e^{-\lambda})^{k+1} - \sum_{i=1}^{k+1} \frac{(1-e^{-\lambda})^i}{i} \right] \ .$$

En particular, tenemos $I\!E[S_5] = \frac{1}{2} - \frac{1-e^{-2\lambda}}{4\lambda}$. Uno también puede calcular la función de acumulación de S_5, para cualquier $x \in [0,1]$, obtenemos:

$$F_{S_5}(x) = (x - \frac{1}{\lambda}(1-x)\ln(1-x)) 1\!\!1_{[0,1-e^{-\lambda}]}(x) + 1\!\!1_{(1-e^{-\lambda},1]}(x) \ .$$

Luego de esta última expresión uno puede deducir que $S_5 \preccurlyeq_{st} S_1$.

Ejemplo 5.2.16. *Como se comentó en la introducción de esta sección, Fill [25] consideró el problema de calcular el límite del costo de búsqueda lineal S_n para diferentes popularidades deterministas. Entre los casos que analizó, Fill estudió los dos siguientes:*

1. *$\omega_i \sim i^{-\alpha}$, con $\alpha \in (0,1)$. Él encontró:*

$$\frac{S_n}{n} \xrightarrow[n\to\infty]{d} A(\alpha) \ ,$$

   donde $A(\alpha)$ es una variable aleatoria cuya distribución está descrita en el Apéndice A de [25].

2. $\omega_i \sim i^s$, con $s > 0$. Él encontró:

$$\frac{S_n}{n} \xrightarrow[n\to\infty]{d} B(s) \; ,$$

donde $B(s)$ es una variable aleatoria cuya distribución está también descrita en el Apéndice A de [25].

Consideremos ahora el problema de calcular la distribución del costo de búsqueda límite usando el modelo de popularidades aleatorias. Trataremos dos casos, una ley de Pareto y una ley Beta:

1. *Asumamos que la distribución que genera la partición es una Pareto de parámetro* $1/\alpha$, *con densidad:*

$$f_\omega(x) = \frac{1}{\alpha} x^{-(1/\alpha+1)} \mathbb{1}_{(1,\infty)}(x) \; .$$

En este caso, nuestros cálculo usando la ecuación (5.18) dan que la transformada de Laplace la densidad del f_{S_6} *es precisamente la de* $A(\alpha)$ *como fue caracterizada por Fill en el Lema 4.5, ecuación (4.4).*

2. *Asumiendo que la distribución que genera la partición es una Beta de parámetros* $(1/s, 1)$. *En esto caso, nuestro resultado de la ecuación (5.16) nos da la densidad* f_{S_7} *la cual es precisamente el de la densidad* $B(s)$ *dado en el Corolario A.2, de la ecuación (A.7).*

Al menos de para estos dos ejemplos es posible encontrar una modelo de popularidades determinista y uno de popularidades aleatorias donde el límite del costo de búsqueda *normalizado tiene la misma distribución.*

Del análisis hecho de estos cinco ejemplos, podemos observar que S_2, S_3, S_4 y S_5 son estocásticamente menores que S_1. Podemos intuir que el caso en que los pesos se distribuyen como una Dirac corresponde al peor caso, dado que todos los ítemes tienen la misma popularidad. Podemos probar que esto es cierto en el caso general:

Proposición 5.2.17. *[5] Sea S la distribución límite del costo de búsqueda asociado a la sucesión ω de variables aleatorias positivas i.i.d. con media finita μ. Luego, $S \preceq_{st} S_1$.*

Demostración: Sea w una variable aleatoria positiva con densidad f y esperanza μ. Sea $\widehat{w}$ otra variable positiva con densidad g definida como $g(x) = xf(x)/\mu$. Sea $\phi(t) = \mathbb{E}[e^{-tw}]$ (respectivamente $\hat{\phi}$) la transformada de Laplace de w (respectivamente $\hat{w}$). Con un cálculo sencillo se puede verificar que $w \preccurlyeq_{st} \widehat{w}$, luego por ser e^{-tx} una función decreciente de x se tiene que $\hat{\phi}(t) \leq \phi(t)$. Como la transformada de Laplace de $\hat{w}$ vale

$$\widehat{\phi}(t) = -\frac{\phi'(t)}{\mu} ,$$

se tiene que $-\frac{\phi'(t)}{\mu} \leqslant \phi(t)$ para todo $t \geqslant 0$.

Aplicando está desigualdad en este contexto, tenemos que para cualquier $x \in [0, 1 - p_0]$,

$$\mathbb{P}(S \leqslant x) = 1 + \frac{1}{\mu}\phi'(\phi^{-1}(1-x)) \geqslant x .$$

Luego, para cualquier $x \in [0,1]$, tenemos $\mathbb{P}(S \geqslant x) \leqslant \mathbb{P}(S_1 \geqslant x)$, donde S_1 es una variable aleatoria de distribución uniforme en el intervalo unitario. Entonces, $S \preccurlyeq_{st} S_1$. $\square$

Como corolario de este resultado uno puede deducir que $\mathbb{E}[S] \leqslant \mathbb{E}[S_1] = \frac{1}{2}$.

5.3. Move-to-root con popularidades aleatorias

5.3.1. Los dos primeros momentos de orden del costo de búsqueda estacionario

En esta subsección, deducimos de los resultados de Allen y Munro [2] los dos primeros momentos del costo estacionario de búsqueda cuando los ítemes tienen pesos aleatorios i.i.d. para la estrategia. Notemos por BS_n el costo de búsqueda estacionario cuando utilizamos la estructura de árbol binario para almacenar los datos.

Teorema 5.3.1. *[8] Para una sucesión $\omega = (\omega_i)_{i\in\mathbb{N}}$ de pesos aleatorios i.i.d. se tiene:*

$$\mathbb{E}[BS_n] = 2\sum_{i=1}^{n-1}\int_0^\infty\int_t^\infty (n-i)\phi'(u)^2\phi(u)^{i-1}\phi(t)^{n-i-1}\,dudt\,. \tag{5.19}$$

$$\begin{aligned}\mathbb{E}[BS_n^2] = \mathbb{E}[BS_n] - 8\sum_{i=1}^{n-2}\sum_{j=1}^{n-i-1}(n-i-j)\times \\ \int_0^\infty\int_t^\infty\int_u^\infty \phi'(v)^2\phi'(u)\phi(v)^{i-1}\phi(u)^{j-1}\phi(t)^{n-i-j-1}\,dvdudt\,.\end{aligned} \tag{5.20}$$

Demostración:

De la expresión de la esperanza de BS_n para el caso de popularidades deterministas en la ecuación 5.2 podemos calcular la esperanza del costo de búsqueda

BS_n en función de ω

$$\begin{aligned}
\mathbb{E}[BS_n] &= \mathbb{E}[\mathbb{E}[BS_n \,|\, \omega]] \\
&= 2 \sum_{1 \leqslant i < j \leqslant n} \mathbb{E}\left[\frac{p_i p_j}{p_i + \cdots + p_j}\right] \\
&= 2 \sum_{a=1}^{n-1} (n-a) \mathbb{E}\left[\frac{p_1 p_{a+1}}{p_1 + \cdots + p_{a+1}}\right] \\
&= 2 \sum_{a=1}^{n-1} (n-a) \mathbb{E}\left[\frac{w_1 w_{a+1}}{(w_1 + \cdots + w_{a+1}) W_n}\right] ,
\end{aligned}$$

pues los pesos están idénticamente distribuidos. Notemos por E_a la esperanza

$$E_a = \mathbb{E}\left[\frac{w_1 w_{a+1}}{(w_1 + \cdots + w_{a+1}) W_n}\right]$$

donde a toma los valores $\{1, \ldots, n-1\}$. Sea $W_n' = w_2 + \cdots + w_a$ y $W_n'' = W_n - W_n' - w_1 - w_{a+1}$. Luego E_a se puede expresar como una función de W_n', W_n'', w_1 y w_{a+1} las cuales son cuatro variables aleatorias independientes:

$$\begin{aligned}
E_a &= \mathbb{E}\left[\frac{w_1 w_{a+1}}{(w_1 + W_n' + w_{a+1})(w_1 + W_n' + w_{a+1} + W_n'')}\right] \\
&= \int_0^\infty \mathbb{E}\left[\frac{w_1 w_{a+1}}{w_1 + W_n' + w_{a+1}} e^{-(w_1 + W_n' + w_{a+1})t}\right] \phi(t)^{n-a-1} \, dt \\
&= \int_0^\infty \int_t^\infty \mathbb{E}\left[w_1 w_{a+1} e^{-(w_1 + w_{a+1})u}\right] \phi(u)^{a-1} \phi(t)^{n-a-1} \, dt du \\
&= \int_0^\infty \int_t^\infty \phi'(u)^2 \phi(u)^{a-1} \phi(t)^{n-a-1} \, du dt \, .
\end{aligned}$$

Reemplazando la expresión de E_a que acabamos de encontrar en la sumatoria, obtenemos:

$$\mathbb{E}[S_n] = 2\sum_{a=1}^{n-1}\int_0^\infty\int_t^\infty (n-a)\phi'(u)^2\phi(u)^{a-1}\phi(t)^{n-a-1}\,dudt\ .$$

De en [2] o también en [11], tenemos que el segundo momento condicional a la partición vale:

$$\mathbb{E}[BS_n^2] = \mathbb{E}[\mathbb{E}[BS_n^2\,|\,\omega]] = \mathbb{E}[BS_n] + 4V\ , \tag{5.21}$$

donde V está dada por:

$$V = \sum_{1\leqslant i<j<k\leqslant n}\mathbb{E}\left[\frac{p_ip_jp_k}{p_i+\cdots+p_k}\left(\frac{1}{p_i+\cdots+p_j}+\frac{1}{p_j+\cdots+p_k}\right)\right]\ .$$

La esperanza de BS_n ya la hemos calculado, solo debemos estimar V. Podemos simplificar la expresión de V utilizando la propiedad de intercambiabilidad de la partición:

$$V = \sum_{a=1}^{n-2}\sum_{b=1}^{n-a-1}(n-a-b)\left[V_1(a,b)+V_2(a,b)\right]\ ,$$

donde:

$$V_1(a,b) = \mathbb{E}\left[\frac{p_1p_{a+1}p_{a+b+1}}{(p_1+\cdots+p_{a+b+1})(p_1+\cdots+p_{a+1})}\right]\ ,$$

y:

$$V_2(a,b) = \mathbb{E}\left[\frac{p_1p_{a+1}p_{a+b+1}}{(p_1+\cdots+p_{a+b+1})(p_{a+1}+\cdots+p_{a+b+1})}\right]\ .$$

Notemos que $V_2(a,b) = V_1(b,a)$. Luego obtenemos la siguiente expresión para V:

$$V = 2\sum_{a=1}^{n-2}\sum_{b=1}^{n-a-1}(n-a-b)V_1(a,b)\ .$$

Realizando un cálculo similar al que hicimos podemos obtener la siguiente expresión integral para V_1:

$$V_1(a,b) = -\int_0^\infty \int_t^\infty \int_u^\infty \phi'(v)^2 \phi'(u)\phi(v)^{a-1}\phi(u)^{b-1}\phi(t)^{n-a-b-1}\, dvdudt\ .$$

Reemplazando esta V_1 en la expresión de V y, a su vez, V en la ecuación (5.21) obtenemos el segundo momento del costo de búsqueda BS_n. □

5.3.2. Ejemplos

En esta subsección ilustraremos el comportamiento del costo de búsqueda cuando los pesos son generados de acuerdo a diferentes distribuciones de probabilidad. Estudiaremos los casos de pesos deterministas y exponenciales. Para estas dos distribuciones estimaremos los dos primeros momentos de BS_n.

En los ejemplos que presentaremos, usaremos los números armónicos (ver [37], p. 73-76). Por lo que, antes de abordar los ejemplos, recordaremos su definición y algunas propiedades:

Definición 5.3.2. *Para cualquier $n \in \mathbb{N}^*$, el n-ésimo número armónico H_n es:*

$$H_n = \sum_{k=1}^{n} \frac{1}{k}\ .$$

Cuando n tiende a infinito es sabido que H_n se comporta como $\log n$, más precisamente se puede aproximar H_n por:

$$H_n = \log n + \mathfrak{C} + \frac{1}{2n} - \frac{1}{12n^2} + \frac{1}{120n^4} - \epsilon \ ,$$

donde:

$$0 < \epsilon < \frac{1}{252n^6} \ ,$$

y $\mathfrak{C}$ denota la constante de Euler.

Recordemos las siguientes dos relaciones que involucran los números armónicos:

$$\sum_{k=1}^{n} H_k = (n+1)H_n - n \ .$$

$$\sum_{k=1}^{n} \frac{1}{k} H_k = \frac{1}{2}(H_n^2 + H_n^{(2)}) \ ,$$

donde:

$$H_n^{(2)} = \sum_{k=1}^{n} \frac{1}{k^2} \ .$$

La expresión $H_n^{(2)}$ satisface el siguiente límite:

$$\lim_{n\to\infty} H_{n-1}^{(2)} = \frac{\pi^2}{6} \ .$$

Finalmente, se puede probar fácilmente el siguiente resultado:

Lema 5.3.3. *Para todo $n \in \mathbb{N}^*$ y $c \in \mathbb{Z}$,*

$$\sum_{k=1}^{n} \frac{k}{k+c} = n - c(H_{n+c} - H_c) \ .$$

Si $c = 1$, la expresión se reduce a:

$$\sum_{k=1}^{n} \frac{k}{k+1} = (n+1) - H_{n+1} \ .$$

Con las propiedades aquí expuestas podemos abordar los ejemplos.

Ejemplo 5.3.4. *Consideremos una sucesión de pesos deterministas: para cualquier $i \in \{1, \ldots, n\}$, $w_i = 1$. La transformada de Laplace ϕ vale:*

$$\phi(s) = e^{-s} \ .$$

Luego uno obtiene:

$$\mathbb{E}[BS_n] = \frac{2(n+1)H_n}{n} - 4 \ . \tag{5.22}$$

Cuando el número n de ítemes tiende al infinito, obtenemos la fórmula asintótica siguiente:

$$\mathbb{E}[BS_n] \sim 2 \log n \ , \tag{5.23}$$

pues $H_n \sim \log n$. Luego encontramos las expresiones (5.22) y (5.23) que ya habían sido encontradas por por Allen y Munro en ([2], pág. 529).

Calculemos el segundo momento de BS_n. Obtenemos:

$$\mathbb{E}[(BS_n)^2] = 8\left(1+\frac{1}{n}\right) H_n H_{n-1} - 4\left(1+\frac{2}{n}\right)\left(H_{n-1}^2 + H_n^{(2)}\right) - \left(14 - \frac{10}{n}\right) H_n$$
$$+20 - \frac{16}{n} \ .$$

Usando las propiedades de los números armónicos obtenemos el siguiente equivalente *asintótico para el momento de orden 2:*

$$\mathbb{E}[(BS_n)^2] \sim 4\log^2 n \ .$$

Ejemplo 5.3.5. *Consideremos la sucesión de variables aleatorias i.i.d. distribuidas como una Gamma de parámetros α y λ. Sin perdida de generalidad, podemos asumir $\lambda = 1$. La transformada de Laplace ϕ del peso aleatorio vale:*

$$\phi(s) = (1+s)^{-\alpha} \ .$$

Luego obtenemos:

$$\begin{aligned} \mathbb{E}[BS_n] &= \frac{2\alpha}{n}\sum_{i=1}^{n-1}\frac{n-i}{(i+1)\alpha+1} \qquad (5.24) \\ &= 2\left(\alpha + \frac{\alpha+1}{n}\right)\sum_{i=1}^{n-1}\frac{1}{(i+1)\alpha+1} - \frac{2(n-1)}{n} \ . \end{aligned}$$

Utilizando el siguiente equivalente asintótico:

$$\sum_{a=1}^{n-1}\frac{1}{(a+1)\alpha+1} \sim \frac{1}{\alpha}\sum_{a=2}^{n}\frac{1}{a} \ .$$

Obtenemos la siguiente aproximación:

$$\mathbb{E}[BS_n] \sim 2\log n \ .$$

Para $\alpha = 1$, los pesos están distribuidos como una variable exponencial *de parámetro* 1 *y de la ecuación (5.24) obtenemos:*

$$\mathbb{E}[BS_n] = 2\left(1 + \frac{2}{n}\right) H_{n+1} - 5 - \frac{4}{n} \ .$$

Cuando el número de ítemes n *tiende a infinito, obtenemos la siguiente aproximación asintótica:*

$$\mathbb{E}[BS_n] \sim 2 \log n \ ,$$

Calculamos ahora el momento de orden 2 *y obtenemos:*

$$\mathbb{E}[(BS_n)^2] = \mathbb{E}[BS_n] + \frac{8\alpha^2}{n} \sum_{a=1}^{n-2} \sum_{b=1}^{n-a-1} \frac{n-a-b}{(\alpha(a+b+1)+1)(\alpha(a+1)+1)} \ .$$

En el caso $\alpha = 1$ *(distribución exponencial), esta expresión se reduce a:*

$$\mathbb{E}[(BS_n)^2] = 8\left(1 + \frac{2}{n}\right) H_{n+1} H_n - 4\left(1 + \frac{3}{n}\right) H_n^2 - 14\left(\frac{1}{3} + \frac{1}{n}\right) H_{n+1} - \frac{8(n+1)}{n} H_n - 4\left(1 + \frac{3}{n}\right) H_n^{(2)} - 29 - \frac{52}{3n} \ .$$

Luego obtenemos el siguiente equivalente asintótico para el segundo momento:

$$\mathbb{E}[(BS_n)^2] \sim 4 \log^2 n \ .$$

Tanto para el ejemplo 5.3.4 como para el ejemplo 5.3.5 en el caso $\alpha = 1$ que presentamos tenemos que:

$$\frac{BS_n}{\mathbb{E}[BS_n]} \xrightarrow[n \to \infty]{Pr} 1 \ .$$

pues por la desigualdad de Chebychev (ver p. 233 [23]) es suficiente demostrar:

$$\frac{\mathrm{Var}[BS_n]}{\mathbb{E}[BS_n]^2} \xrightarrow[n\to\infty]{} 0 \ .$$

No deja de ser sorprendente que los dos ejemplos tengan el mismo comportamiento asintótico para la estrategia MtR y un comportamiento absolutamente diferente en el caso de MtF.

Ejemplo 5.3.6. *Consideremos una secuencia de variables aleatoria i.i.d. distribuidas como una Poisson de parámetro λ. La transformada de Laplace ϕ del peso vale:*

$$\phi(s) = e^{\lambda(e^{-t}-1)} \ .$$

Luego uno puede calcular el siguiente equivalente asintótico:

$$\mathbb{E}[BS_n] \sim 2\log n \ .$$

El comportamiento asintótico del costo esperado es igual en el caso determinista y exponencial.

Conclusiones

Los resultados presentes en esta tesis sobre Cutoff para n-tuplas de procesos se diferencian de los resultados anteriores en tres aspectos: en que de manera general toma espacios no necesariamente finitos, considera diferentes distancias y permite que los procesos converjan a distintas tasas exponenciales. En el Teorema 3.3.1 damos condiciones muy generales para que un n-tupla de procesos que convergen exponencialmente a tasas ρ_i tengan un Cutoff en el instante

$$t_n = \max\left\{ \frac{\log i}{2\rho_{(i,n)}},\ i = 1, \ldots, n \right\} ,$$

donde $\rho_{(1,n)}, \ldots, \rho_{(n,n)}$ son los valores $\rho_1, \ldots, \rho_n$ ordenados de manera creciente. En la Proposición 3.3.2 se agregan algunas condiciones para encontrar un instante de Cutoff de la forma misma forma que en los resultados previos de n-tuplas de Ycart [63] $t'_n = \log n/2\rho$. La demostración del teorema principal se basa en el Lema 3.2.4 y podemos decir que es una combinación entre como las distancias dan cuenta del fenómeno de concentración de la medida y la convergencia a tasa exponencial de los procesos al régimen estacionario. En consecuencia, se podría extender este resultado a n-tuplas de procesos no independientes pero donde siga habiendo concentración de la medida.

Parte de los resultados de esta tesis se centran en las particiones aleatorias que son un modelo que como ya mencionamos tiene diversas aplicaciones en ecología, genética, estadística bayesiana y teoría de números. Entre las particiones más utilizadas están las particiones GEM y PD; relacionadas a través de las operaciones de S-BP y de reordenamiento de manera decreciente. Estudiamos las características de la partición de D que después de una S-BP es una aproximación de GEM y de la partición del modelo de fragmentación de la vara que aproxima una partición GEM de parámetro 1. Ambos análisis se basan principalmente en la función generadora de momentos conjunta de las particiones Teorema 4.3.5 y Lema 4.4.1.

Los resultados de este trabajo en torno a las estrategias de organización MtF y MtR se centran en el análisis del costo de búsqueda en régimen estacionario S_n para MtF y BS_n para MtR con n el numero de ítemes a ordenar. Este trabajo se diferencia de los anteriores en que estudia ambas reglas de auto-organización en el contexto de particiones aleatorias y es similar al desarrollado por [44] quien estudia en una partición aleatoria el problema del coleccionista de cupones. Para el estudio se consideraron las particiones generadas por renormalización y se estudiaron ambos costos de búsqueda en régimen estacionario. El resultado principal para la regla MtF es la distribución asintótica del costo de búsqueda normalizado por el numero de ítemes S_n/n en el Teorema 5.2.7. En la estrategia MtR se encontró para dos particiones particulares que el costo de búsqueda estacionario se comporta como $2\log n$, lo que propone la interesante pregunta de saber cuan general es este comportamiento. Se pueden plantear aquí diversas ideas para extender este trabajo: podría hacerse el mismo análisis para el régimen no estacionario al menos para la estrategia de MtF, también podría analizarse otras estrategias de organización y/o otras particiones aleatorias.

Referencias

[1] D. Aldous and P. Diaconis, *Shuffling cards and stopping times*, Amer. Math. Monthly **93** (1986), no. 5, 333–348. MR 841111 (88a:60021)

[2] B. Allen and I. Munro, *Self-organizing binary search trees*, J. Assoc. Comput. Mach. **25** (1978), no. 4, 526–535. MR 508699 (80a:68034)

[3] J. Barrera and T. Huillet, *On random splitting of the interval*, Statist. Probab. Lett. **66** (2004), no. 3, 237–250. MR 2044909 (2005c:60063)

[4] J. Barrera, T. Huillet, and C. Paroissin, *Size-biased permutation of Dirichlet partitions and search-cost distribution*, Probab. Engrg. Inform. Sci. **19** (2005), no. 1, 83–97. MR 2104552 (2005h:60029)

[5] J. Barrera, T. Huillet, and Ch. Paroissin, *Limiting search cost distribution for move-to-front rule with random request probabilities*, Accepted, 2005.

[6] J. Barrera, B. Lachaud, and B. Ycart, *Cutoff for exponentially converging processes*, Submited, 2005.

[7] J. Barrera and C. Paroissin, *On the distribution of the search cost for the move-to-front rule with random weights*, J. Appl. Probab. **41** (2004), no. 1, 250–262. MR 2036286 (2004k:68034)

[8] ———, *On the stationary search cost for the move-to-root rule with random weights*, Mathematics and computer science. III, Trends Math., Birkhäuser, Basel, 2004, pp. 147–148. MR 2090503

[9] C.M. Bender and S.A. Orszag, *Advanced mathematical methods for scientists and engineers: Asymptotic methods and perturbation theory*, Springer, New-York, 1999.

[10] U.N. Bhat, *Elements of applied stochastic processes*, John Wiley & Sons Inc., New York, 1972, Wiley Series in Probability and Mathematical Statistics. MR 0322976 (48 #1334)

[11] J. Bodell, *Cost of searching - probablistic analysis of the self organizing Move-to-Front and Move-to-Root sorting rules*, Ph.D. thesis, Mathematics Department, Royal Institute of Technology, Sweden., 1997.

[12] J-L. Bon and E. Păltănea, *Convergence of the number of failed components in a Markov system with nonidentical components*, J. Appl. Probab. **38** (2001), no. 4, 882–897. MR 1876546 (2003a:60116)

[13] M-F. Chen, *Equivalence of exponential ergodicity and L^2-exponential convergence for Markov chains*, Stochastic Process. Appl. **87** (2000), no. 2, 281–297. MR 1757116 (2002a:60120)

[14] ———, *From Markov chains to non-equilibrium particle systems*, second ed., World Scientific Publishing Co. Inc., River Edge, NJ, 2004. MR 2091955

[15] P. Collet, T. Huillet, and S. Martínez, *Size-bias picking for finite random partitions of the interval*, 2003 Preprint CMM-B-04/11-116.

[16] P. Diaconis, *The cutoff phenomenon in finite Markov chains*, Proc. Natl. Acad. Sci. USA **93** (1996), 1659–1664.

[17] P. Diaconis, R.L. Graham, and J.A. Morrison, *Asymptotic analysis of a random walk on a hypercube with many dimensions*, Random Structures Algorithms **1** (1990), no. 1, 51–72. MR 1068491 (91g:60078)

[18] P. Diaconis and M. Shahshahani, *Time to reach stationarity in the Bernoulli-Laplace diffusion model*, SIAM J. Math. Anal. **18** (1987), no. 1, 208–218. MR 871832 (88e:60014)

[19] R.P. Dobrow, *The move-to-root rule for self-organizing trees with Markov dependent requests*, Stochastic Anal. Appl. **14** (1996), no. 1, 73–87. MR 1373410 (96j:68041)

[20] R.P. Dobrow and J.A. Fill, *On the Markov chain for the move-to-root rule for binary search trees*, Ann. Appl. Probab. **5** (1995), no. 1, 1–19. MR 1325037 (96d:60100)

[21] ———, *Rates of convergence for the move-to-root Markov chain for binary search trees*, Ann. Appl. Probab. **5** (1995), no. 1, 20–36. MR 1325038 (96d:60101)

[22] P. Donnelly, *The heaps process, libraries, and size-biased permutations*, J. Appl. Probab. **28** (1991), no. 2, 321–335. MR 1104569 (92f:60116)

[23] W. Feller, *An introduction to probability theory and its applications. Vol. I*, Third edition, John Wiley & Sons Inc., New York, 1968. MR 0228020 (37 #3604)

[24] ———, *An introduction to probability theory and its applications. Vol. II.*, Second edition, John Wiley & Sons Inc., New York, 1971. MR 0270403 (42 #5292)

[25] J.A. Fill, *Limits and rates of convergence for the distribution of search cost under the move-to-front rule*, Theoret. Comput. Sci. **164** (1996), no. 1-2, 185–206. MR 1411204 (97f:68023)

[26] J.A. Fill and L. Holst, *On the distribution of search cost for the move-to-front rule*, Random Structures Algorithms **8** (1996), no. 3, 179–186. MR 1603279 (99b:60118)

[27] G.S. Fishman, *Monte Carlo*, Springer Series in Operations Research, Springer-Verlag, New York, 1996, Concepts, algorithms, and applications. MR 1392474 (97g:65019)

[28] P. Flajolet, D. Gardy, and L. Thimonier, *Birthday paradox, coupon collectors, caching algorithms and self-organizing search*, Discrete Appl. Math. **39** (1992), no. 3, 207–229. MR 1189469 (93i:68107)

[29] A.L. Gibbs and F.E. Su, *On choosing and bounding probability metrics*, Int. Statist. Review **70** (2002), no. 3, 419–435.

[30] J. Hawkes, *On the asymptotic behaviour of sample spacings*, Math. Proc. Cambridge Philos. Soc. **90** (1981), no. 2, 293–303. MR 620739 (82h:60026)

[31] W.J. Hendricks, *The stationary distribution of an interesting Markov chain*, J. Appl. Probability **9** (1972), 231–233. MR 0292178 (45 #1265)

[32] W. Hoeffding, *The collected works of Wassily Hoeffding*, Springer Series in Statistics: Perspectives in Statistics, Springer-Verlag, New York, 1994, Edited and with a preface by N. I. Fisher and P. K. Sen. MR MR1307621 (96c:62003)

[33] L. Holst, *The Poisson-Dirichlet distribution and its relatives revisited*, Preprint of the Royal Institute of Technology, Sweden., 2001.

[34] J.F.C. Kingman, *Random partitions in population genetics*, Proc. Roy. Soc. London Ser. A **361** (1978), no. 1704, 1–20. MR 0526801 (58 #26167)

[35] ———, *Poisson processes*, Oxford Studies in Probability, vol. 3, The Clarendon Press Oxford University Press, New York, 1993, Oxford Science Publications. MR 1207584 (94a:60052)

[36] D.E. Knuth, *The art of computer programming*, vol. 3: Sorting and Searching, Addison-Wesley Publishing Co., Reading, Massachusetts, 1973.

[37] ———, *The art of computer programming*, second ed., Addison-Wesley Publishing Co., Reading, Mass.-London-Amsterdam, 1975, Volume 1: Fundamental algorithms, Addison-Wesley Series in Computer Science and Information Processing. MR 0378456 (51 #14624)

[38] A.P. Korostelëv and A.B. Tsybakov, *Minimax theory of image reconstruction*, Lecture Notes in Statistics, vol. 82, Springer-Verlag, New York, 1993. MR 1226450 (95a:62028)

[39] B. Lachaud, *Cutoff and hitting times for a sample of Ornstein-Uhlenbeck processes and its average*, J. Appl. Probab. (2005), no. 4, to appear.

[40] T. Lindvall, *Lectures on the coupling method*, Dover Publications Inc., Mineola, NY, 2002, Corrected reprint of the 1992 original. MR 1924231

[41] F. Martinelli, *Relaxation times of Markov chains in statistical mechanics and combinatorial structures*, Probability on discrete structures, Encyclopaedia Math. Sci., vol. 110, Springer, Berlin, 2004, pp. 175–262. MR 2023653 (2005b:60260)

[42] S. Martínez and B. Ycart, *Decay rates and cutoff for convergence and hitting times of Markov chains with countably infinite state space*, Adv. in Appl. Probab. **33** (2001), no. 1, 188–205. MR 1825322 (2002e:60117)

[43] J. McCabe, *On serial files with relocatable records*, Operations Res. **13** (1965), 609–618. MR 0182458 (31 #6681)

[44] V. G. Papanicolaou, G. E. Kokolakis, and S. Boneh, *Asymptotics for the random coupon collector problem*, J. Comput. Appl. Math. **93** (1998), no. 2, 95–105. MR MR1638006

[45] J. Pitman, *Random discrete distributions invariant under size-biased permutation*, Adv. in Appl. Probab. **28** (1996), no. 2, 525–539. MR 1387889 (97d:62020)

[46] ———, *Poisson-Dirichlet and GEM invariant distributions for split-and-merge transformation of an interval partition*, Combin. Probab. Comput. **11** (2002), no. 5, 501–514. MR 1930355 (2004d:60253)

[47] J. Pitman and M. Yor, *The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator*, Ann. Probab. **25** (1997), no. 2, 855–900. MR 1434129 (98f:60147)

[48] D. Pollard, *A user's guide to measure theoretic probability*, Cambridge Series in Statistical and Probabilistic Mathematics, vol. 8, Cambridge University Press, Cambridge, 2002. MR 1873379 (2002k:60003)

[49] U. Porod, *The cut-off phenomenon for random reflections*, Ann. Probab. **24** (1996), no. 1, 74–96. MR 1387627 (97e:60012)

[50] A. P. Prudnikov, Yu. A. Brychkov, and O. I. Marichev, *Integrals and series. Vol. 1*, Gordon & Breach Science Publishers, New York, 1986, Elementary functions,

Translated from the Russian and with a preface by N. M. Queen. MR MR874986 (88f:00013)

[51] R.-D. Reiss, *Approximate distributions of order statistics*, Springer Series in Statistics, Springer-Verlag, New York, 1989, With applications to nonparametric statistics. MR 988164 (90e:62001)

[52] G.O. Roberts and J.S. Rosenthal, *Geometric ergodicity and hybrid Markov chains*, Electron. Comm. Probab. **2** (1997), no. 2, 13–25 (electronic). MR 1448322 (99b:60122)

[53] G.O. Roberts and R.L. Tweedie, *Geometric L^2 and L^1 convergence are equivalent for reversible Markov chains*, J. Appl. Probab. **38A** (2001), 37–41, Probability, statistics and seismology. MR 1915532 (2003g:60122)

[54] J.S. Rosenthal, *On generalizing the cut-off phenomenon for random walks on groups*, Adv. in Appl. Math. **16** (1995), no. 3, 306–320. MR 1342831 (96j:60007)

[55] L. Saloff-Coste, *Lectures on finite Markov chains*, Lectures on probability theory and statistics (Saint-Flour,1996) (Berlin), Lecture Notes in Math., vol. 1665, Springer, 1997, pp. 301–413. MR 1490046 (99b:60119)

[56] ———, *Random walks on finite groups*, Probability on discrete structures, Encyclopaedia Math. Sci., vol. 110, Springer, Berlin, 2004, pp. 263–346. MR 2023654 (2004k:60133)

[57] M. Sibuya and H. Yamato, *Ordered and unordered random partitions of an integer and the GEM distribution*, Statist. Probab. Lett. **25** (1995), no. 2, 177–183. MR 1365035 (97k:62021)

[58] F.E. Su, *Methods for quantifying rates of convergence for random walks on groups*, Ph.D. thesis, Harvard University, 1995.

[59] S. Tavaré and W.J. Ewens, *Multivariate ewens distribution*, Discrete multivariate distributions, Wiley Series in Probability and Statistics: Applied Probability and Statistics, John Wiley & Sons Inc., New York, 1997, pp. 232–246. MR 1429617 (99h:62014)

[60] M.L. Tsetlin, *Finite automata and the simulation of the simplest forms of behavior*, Uspehi Mat. Nauk **18** (1963), no. 4 (112), 3–28. MR 0159734 (28 #2951)

[61] E.A. Van Doorn, *Representations for the rate of convergence of birth-death processes*, Teor. Ĭmovīr. Mat. Stat. (2001), no. 65, 33–38. MR 1936126 (2003i:60145)

[62] H. Yamato, M. Sibuya, and T.Ñomachi, *Ordered sample from two-parameter GEM distribution*, Statist. Probab. Lett. **55** (2001), no. 1, 19–27. MR 1860188 (2002k:62030)

[63] B. Ycart, *Cutoff for samples of Markov chains*, ESAIM Probab. Statist. **3** (1999), 89–106 (electronic). MR 1716128 (2000j:60085)

[64] ______, *Stopping tests for Markov chain Monte-Carlo methods*, Methodol. Comput. Appl. Probab. **2** (2000), no. 1, 23–36. MR 1783151 (2001h:60125)

[65] ______, *Cutoff for Markov chains: some examples and applications*, Complex systems (Santiago, 1998) (Dordrecht), Nonlinear Phenom. Complex Systems, vol. 6, Kluwer Acad. Publ., 2001, pp. 261–300. MR 1886358 (2004a:60126)