



Universidad Nacional del Sur

TESIS DE DOCTOR EN MATEMÁTICA

*Estadística de Procesos Estocásticos Aplicados a Redes de
Datos y Telecomunicación*

Beatriz Susana Marrón

BAHÍA BLANCA

ARGENTINA

Mayo, 2012

A Daniel, Ignacio, Javier y a mis padres.

Índice general

Prefacio	5
Resumen	6
Abstract	7
Introducción	8
1. Ancho de Banda Efectivo	13
1.1. Introducción	13
1.2. Ancho de Banda Efectivo	15
1.3. Modelos Markovianos	27
1.3.1. Cadenas de Markov	27
1.3.2. Flujos Markovianos	32
1.4. Cálculo del Ancho de Banda Efectivo para Flujos Markovianos	32
2. Una Generalización del Modelo de Flujo Markoviano	35
2.1. Introducción	35
2.2. Flujos Markovianos con tasas de transferencias aleatorias	36
2.3. Cálculo del ancho de banda efectivo para el flujo Markoviano generalizado	37
2.3.1. Caso discreto	37
2.3.2. Caso continuo	40
2.4. Estimación del ancho de banda efectivo	43
2.4.1. Estimadores no paramétricos	43
2.4.2. Estimadores paramétricos	44
2.5. Estimación asintóticamente gaussiana para $\alpha(s, t)$	45
2.6. Consistencia de los estimadores	62
2.7. Intervalo de Confianza para $\alpha(s, t)$	65
3. Punto operacional	74
3.1. Introducción	74
3.2. Teoría de grandes desvíos	76
3.3. Punto operacional	81
3.4. Estimación del punto operacional para el Flujo Markoviano Generalizado .	83

4. Simulaciones y resultados numéricos	89
4.1. Introducción	89
4.2. Parámetros elegidos para la simulación	90
4.3. Estimación del ancho de banda efectivo de trazas	91
4.4. Estimación del punto operacional del enlace	93
4.5. Estimación de los parámetros de calidad de servicio de un enlace	94
4.6. Diseño de un enlace basado en la estimación del ancho de banda efectivo .	96
Conclusiones y trabajos futuros	101
Notación y abreviaturas	102
Glosario	103
Bibliografía	105

Prefacio

Esta tesis es presentada como parte de los requisitos para optar al grado académico de Doctor en Matemática de la Universidad Nacional del Sur y no ha sido presentada previamente para la obtención de otro título en esta Universidad u otras. La misma contiene resultados obtenidos en investigaciones llevadas a cabo en el Departamento de Matemática de la Universidad Nacional de Sur durante el período comprendido entre los meses de setiembre de 2007 y Abril de 2012, bajo la dirección de Gonzalo Perera, Profesor Titular del La Universidad de la República, Uruguay.

Agradecimientos

Deseo expresar mi agradecimiento al Dr. Gonzalo Perera cuya ayuda, amistad y paciente atención hicieron posible la culminación de este trabajo.

Quiero agradecer también el apoyo brindado por las autoridades y personal del Departamento de Matemática de la UNS y del INMABB, y muy especialmente al Lic. José Bavio por su colaboración en la implementación de algunos de los algoritmos que se presentan en este trabajo.

8 de Mayo de 2012

Departamento de Matemática.

Universidad Nacional del Sur.

Resumen

En este trabajo se aborda el problema de compartir recursos en redes de comunicación de banda ancha que puede garantizar una cierta calidad de servicio (QoS), y se desarrollan algunos de los resultados sobre fuentes de datos y modelización del tráfico, especialmente en los aspectos del ajuste del modelo y la estimación de parámetros.

El multiplexado de las fuentes de tasa variable plantea un problema matemático y estadístico: la estimación de las necesidades de recursos de una fuente o un conjunto de fuentes. El método de estimación deberá ser lo suficientemente simple para ponerse en práctica en el control de aceptación de conexiones (CAC).

Se busca realizar un aporte al empleo del concepto de ancho de banda efectivo con el fin de estimar la asignación de recursos o la “ocupación” del canal de cada fuente. En particular para la estimación y cálculo del punto operacional de un enlace dentro de una red, entendiendo como punto operacional al par de valores de los parámetros de tiempo y espacio o multiplexado en que el ancho de banda efectivo da la probabilidad asintótica de desborde del buffer.

Nuestro estudio trata diversos aspectos, por un lado se enfoca el problema del modelado estocástico de las fuentes de tráfico de las redes y sus enlaces y por otro lado se aborda el problema de la estimación de algunos parámetros de la calidad del servicio.

Respecto del primer aspecto se desarrollan ejemplos de una amplia gama de modelos típicos utilizados hasta ahora para modelar las fuentes de tasa variable y además se introduce un nuevo modelo de especial interés.

Para el segundo aspecto se demuestra que, dado un buen estimador del ancho de banda efectivo que obtenemos a partir de las trayectorias de tráfico, se puede estimar con precisión el punto operacional y bajo la imposición de algunas condiciones de regularidad, se prueba que dicho estimador es consistente y se construye un intervalo de confianza.

Además se extendenden estas propiedades a otros parámetros del enlace, como son la probabilidad de pérdida, la estimación de la capacidad del enlace y el tamaño de buffer mínimo necesarios para que el enlace opere con una probabilidad de pérdida para determinada calidad de servicio.

Por último, mediante simulaciones de trazas, se verifican las precisiones de los resultados teóricos obtenidos.

Abstract

This work addresses the problem of sharing network resources in broadband communication that can guarantee a certain quality of service (QoS), and develops some of the results on data sources and modeling of traffic, especially on aspects of model fit and parameter estimation.

The multiplexing of variable rate sources poses a mathematical and statistical problem: estimating the resource requirements of a font or a set of fonts. The estimation method should be simple enough to be implemented in connection acceptance control (CAC).

We try to contribute to the use of the concept of effective bandwidth in order to estimate the allocation of resources or the ρ occupation of the channel from each source. In particular for the estimation and calculation of the operating point of a link within a network, defining operational point as the pair of values of the parameters of time and space or multiplexed on the effective bandwidth gives the asymptotic probability of buffer overflow.

Our study covers various aspects on the one hand focuses on the problem of stochastic modeling of traffic sources and network links and on the other hand addresses the problem of estimating some parameters of quality of service.

For the first aspect, examples of a wide range of typical models used so far to model variable rate sources are developed and a new model of special interest is introduced.

For the second aspect is shown that, given a good estimator of the effective bandwidth that we get from traffic paths can be estimated accurately the operational point, and under imposing some regularity conditions we prove that this estimator is consistent and confidence interval can be developed.

These properties are also extended to other parameters of the link such as the loss probability, the minimum link capacity and the minimum buffer size needed for the link operate with a given quality of service.

Finally, by simulated traces, the details of the theoretical results are checked

Introducción

El creciente interés en transportar servicios en tiempo real y la necesidad de disponer de redes que integren varios servicios de telecomunicaciones motivó el concepto de Red Digital de Servicios Integrados que ahora se denomina de banda estrecha, que consiste en el uso de una única infraestructura para el transporte de datos en forma conjunta, como por ejemplo audio e imágenes.

A mediados de la década de los 80 la Unión Internacional de Telecomunicaciones comenzó la estandarización del modelo de Red Digital de Servicios Integrados de Banda Ancha, la cual permitiría la transmisión de información a altas velocidades.

Un punto de vital importancia a lo largo del de este proceso fue la elección del mecanismo de transferencia, éste se podría definir como el conjunto de herramientas de multiplexado y conmutación que utiliza la red. También lo fue la formulación de ciertos principios de funcionamiento de la red y las formas de medir su desempeño. Es posible que la más importante y de la que nos ocuparemos en este trabajo sea la calidad de servicio garantizada (QoS) como también así la forma de garantizarla.

Se eligió entonces el modelo de redes de modo de transferencia asincrónico ATM (Asynchronous Transfer Mode), que permite multiplexar eficientemente diversas fuentes ya que mediante el concepto de multiplexado estadístico de las fuentes, se obtiene una gran eficiencia en el uso de los recursos de la red.

Si se dispone de diversas fuentes que emiten a una tasa variable, las redes tradicionales reservan un ancho de banda igual a la tasa pico o máxima, malgastando de esta forma el ancho sobrante cuando la tasa resulta inferior. Esta es la ineficiencia que se trata de evitar con el multiplexado de las fuentes sobre enlaces con capacidad menor que las sumas de las tasas picos.

El costo del multiplexado es que, al ser un método estadístico, la posibilidad que los picos de todas las fuentes coincidan no es nula, en tal caso se perderá información por overflow lo que perjudicará la calidad del servicio. Como los servicios de tiempo real son sensibles al retardo y más aún a la pérdida de información, para minimizar los efectos de las pérdidas y mantener la calidad del servicio, tanto para las nuevas conexiones como para las preexistentes, se necesita disponer de controles de admisión de conexiones (CAC) y mecanismos de control de congestión que nos permitan decidir si se puede admitir una nueva fuente de transmisión.

En el marco de la riqueza y diversidad de problemas que estos temas presentan, uno de los objetivos de este trabajo es el análisis de la performance de una red y cuando ese análisis se realiza desde un enfoque paramétrico se debe conocer cómo se comporta el tráfico de la fuente. Para hacerlo es necesario disponer de buenos modelos matemáticos que permitan estudiar y comprender el comportamiento estadístico del tráfico (datos, audio y video) que será transmitido por las redes.

En el caso de las fuentes de datos, los generadores más habituales son los servidores de web, las transferencias de ficheros (FTP), las conexiones remotas y los chats, es decir todos los servicios presentes en las redes IP (Internet). El comportamiento en cada caso

es muy diferente, la web se caracteriza por la gran cantidad de transferencia de paquetes, generalmente de reducido tamaño, en cambio la FTP se caracteriza por presentar largas transferencias de información.

Para las fuentes de audio lo habitual es que la señal transmitida sea de voz, donde se trata de aprovechar las características del habla humana que se presenta por ráfagas (*talk spurts*) con silencios intercalados entre palabras y entre frases, por eso lo más habitual es que los codificadores de voz incorporen detectores de silencio durante los cuales no se transmite información.

Por último las fuentes de video de tasa variable (VBR) se caracterizan por estar fuertemente correlacionadas. Este tipo de tráfico es muy importante en las redes de banda ancha por ser el que más recursos de transmisión necesita y uno de los más sensibles al retardo y a la pérdida de información.

Otra característica de las redes ATM es que son orientadas a conexión, es decir cada vez que se transfiere información, se debe establecer un camino virtual desde el origen al destino y dentro de la llamada se producen ráfagas (*bursts*) en las que se transmiten celdas de información, seguidas de intervalos de silencio. Dentro de cada ráfaga las celdas no necesariamente se transmiten de modo uniforme, por lo tanto las redes ATM presentan un comportamiento diferente en cada una de las escalas de tiempo consideradas: llamadas, ráfagas y celdas.

Las escalas de celdas y ráfagas son las que nos permiten dimensionar los buffers y los mecanismos de admisión de nuevas conexiones ya que uno de los posibles problemas provienen de la probabilidad de que el tráfico agregado exceda la capacidad de salida del sistema provocando pérdida de información y estos *overflows* pueden evitarse mediante mecanismos de control de admisión eficientes.

En el desarrollo de la tesis se ofrece una visión, sin ánimo de ser exhaustiva, de algunos modelos y se propone y estudia uno nuevo para el tema en el cual nos hemos enfocados, el tráfico en redes de banda ancha.

Tradicionalmente se ha asumido que en las redes de conmutación de paquetes de datos se daban las condiciones necesarias para suponer que la generación de celdas sigue un proceso de Poisson o de Bernoulli. Sin embargo, como ya hemos mencionado al describir los servicios de Internet, cada aplicación presenta una tasa y distribución de celdas diferentes que pueden ir desde transmisiones esporádicas y cortas hasta largas transferencia de información.

Para modelar las fuentes de voz los más apropiados son los modelos *on-off* que se basan en una cadena de Markov con dos estados. Estos modelos aproximan el fenómeno *talk spurts* ya que describen una fuente que emite información por ráfagas, generando paquetes en el estado *on* y silencios en el estado *off*. Para modelar mezclas de tráfico de voz y datos se suele emplear los Procesos de Poisson modulados por cadenas de Markov (MMPP) que son procesos estocásticos que utiliza una cadena de Markov que define, según el estado en que se encuentra, la tasa del tráfico, es decir la cadena modula el proceso de generación de la información.

Los modelos de flujos Markovianos son especialmente recomendados para modelar fuentes de video tipo VBR, donde el tráfico se presenta como un flujo continuo, sobre

todo cuando las unidades de tráfico o paquetes son muy pequeños comparados con el tráfico total. En estos modelos el estado de la cadena de Markov modulante determina la tasa del flujo. A cada estado le corresponde una tasa fija. Una fuente de video codificada se podría modelar de manera simplificada como una cadena con dos estados. Cuando la imagen está “quieta” no hay diferencia entre un cuadro y el anterior y no se transfiere información, estado *off* de la cadena con velocidad de transmisión nula. Cuando la imagen varía mucho se debe transmitir todo el cuadro cada vez, se estaría en el estado *on* de la cadena con una velocidad de transmisión alta. En casos reales cuando la imagen está “quieta” en realidad hay poca diferencia y se transmite a baja velocidad, con cierta variabilidad de cuadro a cuadro y probablemente exista más de un estado de transmisión a velocidad alta dependiendo de la magnitud del cambio.

En un tráfico real de datos, voz o video, si se quisiera modelar mediante un flujo Markoviano, la gran cantidad de valores diferentes que se observan para las tasas, se traducirá en el aumento de la cantidad de estados que debería asumir la cadena modulante y por consiguiente aumentará la dimensión del problema haciéndolo muy difícil de manejar.

Esto nos ha llevado a pensar en un modelo alternativo que introducimos en esta tesis y que hemos denominado Flujo Markoviano Generalizado, donde los cambios bruscos de la velocidad de transferencia reportan un cambio de estado en la cadena modulante, pero dentro de un estado se permite que la tasa asuma ya no un valor fijo sino un valor aleatorio sorteado de acuerdo a alguna distribución de probabilidades.

Para interpretar este modelo podemos pensar que el estado de la cadena indica un tipo de actividad en la transferencia como por ejemplo correo, chat, conversación, videoconferencia, entre otros y para cada estado la tasa se sortea, dentro de un rango de valores razonable para ese tipo de actividad y según una ley de probabilidad.

Una vez asignado un modelo a la fuente estudiada, sería interesante disponer de algún criterio que permita valorar la bondad del mismo. La validación consiste en medir qué tan próximo es el modelo a la fuente real de tráfico.

Cuando se realiza ingeniería de tráfico en línea se necesita contar con herramientas que permitan analizar la performance midiendo o estimando los parámetros de la calidad de servicio requerida sobre la red, en particular será necesario estimar la probabilidad de pérdida. Por otra parte este parámetro es usado como criterio de diseño y para aceptar conexiones, por lo tanto resulta técnicamente relevante disponer de buenas estimaciones de dicha probabilidad. Un modelo que se ajuste bien a los datos de tráfico servirá a tal fin.

El desarrollo de las herramientas estadísticas para el estudio del comportamiento de un enlace de una red surge con mayor fuerza a partir de la noción de *ancho de banda efectivo* introducida por Kelly en 1996 [43], que permite encontrar expresiones para estimar la probabilidad de pérdida en un enlace.

La primer pregunta que se presenta es por qué el nombre de ancho de banda efectivo. La idea es que el valor de esta función indicará la cantidad de ancho de banda mínimo del enlace que es necesario reservar para la fuente a los efectos de cumplir con los requerimientos de calidad de servicio. Se busca una función del tráfico de una fuente que, dependiendo del contexto, indique la cantidad de recursos que se debe reservar a la fuente

y el contexto estará dado por la capacidad del enlace, del buffer, de las máximas pérdidas que se desean y de otras fuentes que también alimenten el enlace.

Esta magnitud tiene propiedades interesantes, entre ellas que se encuentra entre la media y el máximo del tráfico de la fuente ya que si se reserva el valor pico de la fuente, no se tendrían pérdidas pero se estaría desperdiciando capacidad del sistema, mientras que si se reserva un valor muy cercano a la media se tendrá un buen aprovechamiento de los recursos, pero las pérdidas probablemente estén por encima de las deseadas.

El concepto de ancho de banda efectivo puede ser aplicado a una fuente o a una agregación de fuentes, eventualmente muy grande como lo es en los enlaces que están en el corazón de una red. En este enfoque, la asintótica de muchas fuentes, el ancho de banda efectivo está estrechamente relacionado con la probabilidad de pérdida por desborde del buffer.

Será necesario entonces disponer de una buena estimación del ancho de banda efectivo para, a partir de ella, poder estimar la probabilidad de pérdida.

Diversos autores como [25, 64, 55] trataron sobre este punto encontrando buenos estimadores para los modelos clásicos como el MMPP o el Flujo Markovianos, nuestra principal motivación es encontrar un buen estimador para el nuevo modelo, el Flujo Markoviano Generalizado.

En este trabajo se encontrará una fórmula para el cálculo del ancho de banda efectivo del tipo de Kesidis, Walrand y Chang [42], desarrollada para el modelo del Flujo Markoviano Generalizado y que es de particular interés porque permite expresar dicha magnitud en función de los parámetros del modelo: el generador infinitesimal de la cadena modulante y las tasas medias de transferencia. Esta fórmula nos permite proponer un estimador del ancho de banda efectivo a partir de trayectorias de tráfico, sobre el cual se prueba su consistencia y su distribución asintótica.

Una vez que se tiene un buen estimador para el ancho de banda efectivo surge la pregunta de si será cierto que usando este estimador en lugar del valor teórico para calcular la probabilidad de pérdida se obtiene un estimador de la misma que herede ambas propiedades, consistencia y normalidad asintótica.

En este trabajo se determina bajo qué condiciones se cumple esto tanto para la probabilidad de pérdida como para la capacidad del enlace y para el tamaño de buffer mínimo necesarios para que el enlace opere con cierta probabilidad de pérdida para determinada calidad de servicio. Estos estimadores son aplicados al análisis y diseño de redes garantizando una cierta calidad de servicio, punto central en el diseño de nuevas redes de telecomunicación. Mediante simulaciones se aplican estos resultados al análisis de performance y diseño y se verifican las precisiones de los resultados teóricos obtenidos.

Esta tesis está organizada de la siguiente manera, en el Capítulo 1 se presentan algunos conceptos fundamentales que permiten desarrollar herramientas estadísticas para el tratamiento de problemas de interés en el modelado y análisis de performance de redes de datos como son el concepto de ancho de banda efectivo, sus propiedades y ejemplos de su cálculo para los modelos clásicos. Se dan las herramientas básicas de Cadena de Markov necesarias para comprender tanto el modelo de Flujo Markoviano como para el modelo de flujo Markoviano Generalizado, por último se calcula el ancho de banda efectivo para

el modelo de Flujo Markoviano.

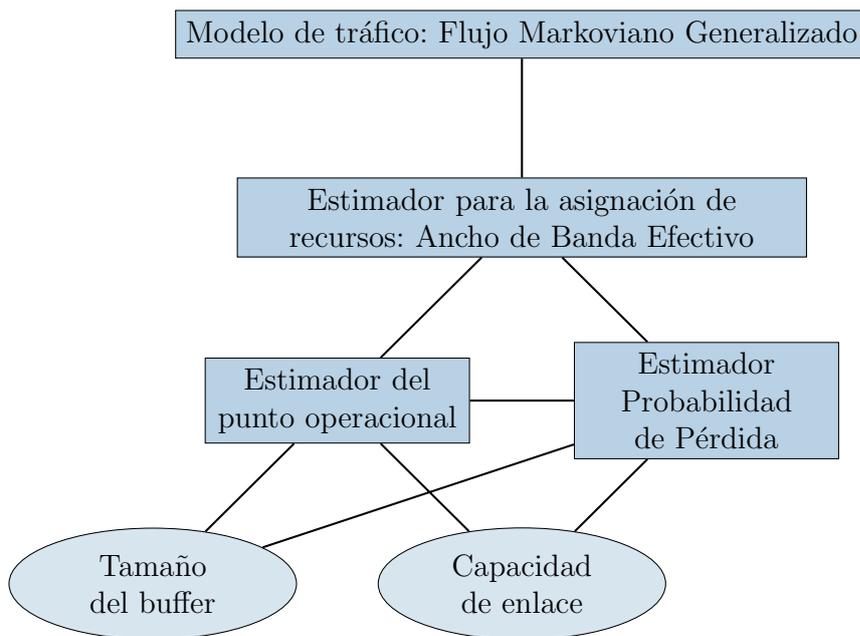
En el Capítulo 2 se estudia el nuevo modelo propuesto en esta tesis para describir el funcionamiento de las fuentes al cual se lo denominó Flujo Markoviano Generalizado. Para este modelo se calcula su ancho de banda efectivo, se encuentra una fórmula del tipo Kesidis, Walrand y Chang para la expresión del mismo, se presenta un estimador para el ancho de banda efectivo a partir de trazas de tráfico y se demuestra su consistencia y su distribución asintótica. Por último se presenta un intervalo de confianza para el mismo.

En el Capítulo 3 se introducen las herramientas de la teoría de grandes desviaciones necesarias en las demostraciones y se presenta el concepto de punto operacional y un estimador para el mismo. Se demuestran su consistencia y distribución asintótica. Se analizan estos conceptos para el modelo de Flujo Markoviano Generalizado.

Por último en el Capítulo 4, se muestran los resultados de las simulaciones con los cuales se validan los resultados teóricos obtenidos en los Capítulos 2 y 3.

Si bien este trabajo está motivado por la ingeniería y el análisis del tráfico sobre redes de flujos de tasas variables, sus resultados son válidos para problemas de uso compartido de recursos limitados, en que se debe mantener ciertos niveles de pérdida, donde la pérdida puede entenderse de diversas formas según el problema.

El siguiente diagrama de bloques explica como se desarrollará el trabajo de esta tesis.



Capítulo 1

Ancho de Banda Efectivo

1.1. Introducción

Esta tesis aborda algunos problemas de interés en el modelado y análisis de performance de redes de datos. Por consiguiente, en este capítulo introducimos algunos conceptos fundamentales que permiten desarrollar herramientas estadísticas para el tratamiento de este tipo de problemas. En particular presentamos la definición y propiedades de una magnitud llamada *ancho de banda efectivo* utilizada en el análisis local de redes y colas de espera. Mostraremos además algunos modelos de tráfico de datos usados en la literatura para los cuales se puede calcular su ancho de banda efectivo, con el objetivo de brindar la introducción necesaria y motivadora de los resultados originales que se desarrollarán en los capítulos siguientes.

Si consideramos un sistema multiplexor donde varios flujos de datos comparten una única salida, es de interés conocer cuánto recurso requiere cada flujo del sistema. Pensamos en el sistema como un servidor que puede procesar el trabajo que recibe de varias fuentes distintas a una cierta tasa constante c y además puede almacenar en un *buffer* hasta una cierta cantidad b de unidades de trabajo. El trabajo que llega cuando el *buffer* está completo se descarta y por lo tanto nunca es procesado.

Un ejemplo de este tipos de redes es la Internet, cuya política tradicional de trabajo es el “best effort”, es decir hacer lo que puede dividiendo sus recursos entre todas las solicitudes. Si en un instante muchas comunicaciones deben usar un mismo enlace de salida, los paquetes se van almacenando en cola (buffers) esperando su turno, esto produce retardo y si el buffer se llena los paquetes se pierden.

Las aplicaciones tradicionales de Internet son los mails, la transferencia de archivos y el uso de la Web donde los retardos no afectan demasiado y las pérdidas se solucionan con retransmisiones pero en los nuevos servicios que puede ofrecer la Internet como son las comunicaciones de voz y video, tanto el retardo como la pérdida de paquetes son críticos.

La causa más significativa de retardo y pérdidas de paquetes en los buffers, es el exceso de tráfico en el canal de transmisión, por lo que es necesario proponer modelos estocásticos y técnicas estadísticas que permitan controlar las probabilidades de congestión o de pérdida de paquetes y establecer condiciones para poder realizar cada tarea demandada.

Conocer entonces, cuánto de los recursos disponibles necesita cada conexión tiene aplicación directa a lo que se llama control de admisión de conexiones (CAC), ya que una nueva conexión puede ser aceptada si hay suficientes recursos disponibles y el precio de la conexión debería ser, de alguna forma, proporcional a los recursos necesarios. Al tomar la decisión de aceptar una nueva conexión, se debe estimar cuanto ocupará esta en el canal de comunicación, para saber si al añadir el nuevo tráfico se puede mantener la calidad de servicio (QoS), tanto de la nueva conexión como de las ya existentes.

El ancho de banda de una red de datos se da en términos del número de bits que se pueden transmitir por segundo. El ancho de banda efectivo de una red no sólo dependerá del ancho de banda físico, sino también de la eficiencia de la codificación de los datos en paquetes, la eficiencia con que opera la red y si el tiempo de la red se desperdicia debido a las colisiones de mensajes. Esto impone restricciones importantes sobre su funcionamiento.

Hay dos límites claros para estos requerimientos, por un lado se podría reservar la tasa máxima de transmisión para cada conexión (lo que da lugar a un multiplexor determinístico), si bien no hay posibilidad de “*buffer overflow*” es decir no queda trabajo sin procesar, la utilización del enlace es muy pobre. Por otro lado, se podría reservar la tasa media de transmisión para cada conexión, que permite utilizar en media el cien por ciento del enlace, pero es inevitable que exista “*buffer overflow*”. Entonces, dada una cierta calidad de servicio requerida (probabilidad de pérdida de trabajo, retardo en el tiempo de procesamiento, etc.) la cantidad de recursos que deberá reservarse para cada conexión, está en algún lugar entre la tasa media y la tasa máxima de dicha conexión. Esta cantidad se conoce generalmente como *ancho de banda efectivo* y es entonces una medida del uso de los recursos.

El concepto de ancho de banda efectivo introducido por Kelly en [43], es una medida útil de la ocupación del canal de transmisión, cuyo sustento teórico proviene de la teoría de grandes desvíos y cuyas propiedades la hacen una medida apropiada. Aplicaciones de esta magnitud a las redes de telecomunicación se presentan en [27, 36] entre otros.

Desarrollamos como ejemplos varios modelos de tráfico y nos enfocamos más particularmente sobre el modelo conocido como Flujo Markoviano que es aceptado en la literatura para datos provientes de voz y de video MPEG. Para este tipo de modelos Keisidis, Walrang y Chang presentaron en [42] una expresión para el ancho de banda efectivo, cuando las tasas de transferencias son constantes.

Este capítulo está organizado de la siguiente manera, en la Sección 1.2 se presenta el concepto de ancho de banda efectivo, sus propiedades y ejemplos de su cálculo para distintos procesos estacionarios, en la Sección 1.3 se presentan los conceptos fundamentales de cadenas de Markov y el modelo de Flujo Markoviano, por último en la Sección 1.4 se presentan los resultados existentes sobre el ancho de banda efectivo para el modelo de Flujo Markoviano.

1.2. Ancho de Banda Efectivo

Los usuarios que transmiten datos en la red, también denominados *fuentes*, son modelados mediante procesos estocásticos caracterizados por algunas propiedades físicas y estadísticas como por ejemplo la capacidad máxima y la tasa de arribo del trabajo entre otras. Además cada nodo puede describirse mediante ciertos parámetros como son su capacidad, el tamaño del buffer y la probabilidad de pérdida u “overflow” que describe la calidad del nodo.

Dado un nodo con tamaño de buffer y capacidad de enlace fija, si se establece una probabilidad de pérdida, que en general será muy pequeña para cumplir con requisitos de calidad, se puede entonces estimar el número de fuentes que pueden ser atendidas por dicho nodo. El multiplexado estadístico se utiliza para agregar tráfico en un nodo y tiene la ventaja de permitir el uso del recurso ocioso que de otra manera se hubiera desperdiciado y la desventaja de no garantizar la calidad de servicio. Tanto los parámetros de la fuente como los del nodo son conocidos y el ancho de banda efectivo estima la cantidad de recurso que cada fuente consume.

La cantidad de trabajo que llega a un switch desde una fuente en un intervalo de tiempo de una longitud t , es claramente una variable aleatoria, en consecuencia un proceso estocástico es la herramienta adecuada para modelarlo.

Sea (Ω, \mathcal{A}, P) un espacio de probabilidad, consideraremos en él un proceso estocástico, es decir una función $X : \Omega \times \mathbb{R}^+ \rightarrow \mathbb{R}$ tal que $X(\omega, t) = X_t(\omega)$ es una variable aleatoria para todo t positivo. Para cada $\omega \in \Omega$ a la función $X(\omega, \cdot) : \mathbb{R}^+ \rightarrow \mathbb{R}$ se le llamará trayectoria del proceso.

Es necesario definir además qué características deben tener los procesos estocásticos que describen las fuentes que concurren a cada nodo de la red.

Un proceso estocástico se dice *estacionario* si dado k positivo, para cualquier conjunto de números reales t_1, t_2, \dots, t_k y h , la variable aleatoria $(X_{t_1}, X_{t_2}, \dots, X_{t_k})$ tiene la misma distribución conjunta que $(X_{t_1+h}, X_{t_2+h}, \dots, X_{t_k+h})$ y se dice de *incrementos estacionario* si el proceso $\{X_t - X_s\}_{t \geq 0}$ es un proceso estacionario para todo valor de s positivo. Además se dice de *incrementos independientes* si dado un número natural k , para cualquier t_1, t_2, \dots, t_k tales que $0 = t_0 \leq t_1 \leq \dots \leq t_k$, se cumple que las variables aleatorias $\{X_{t_i} - X_{t_{i-1}}\}$ son independientes para todo $i \in \{1, \dots, k\}$.

En esta presentación X_t representa la cantidad de trabajo que llega a un *switch* desde una fuente en el intervalo $[0, t]$ y suponemos que es un proceso estocástico estacionario, es decir suponemos que la distribución de la cantidad de trabajo que llega en un intervalo sólo depende de la longitud del mismo. Diremos entonces que las fuentes que cumplen esta propiedad son fuentes estacionarias.

Utilizaremos la definición de ancho de banda efectivo (EB su sigla en inglés) desarrollada por Kelly, [43].

Definición 1.1 *Se define la función ancho de banda efectivo de una fuente estacionaria como*

$$\alpha(s, t) = \frac{1}{st} \log E e^{sX_t} \quad 0 < s, t < \infty, \quad (1.1)$$

donde X_t es el trabajo que produce la fuente en el intervalo $[0, t]$ y s y t son parámetros de escala.

La idea es que el valor de esta función para un cierto punto de operación (s, t) indicará la cantidad de ancho de banda mínimo del enlace que es necesario reservar para la fuente a los efectos de cumplir con los requerimientos de calidad de servicio. Se busca una función del tráfico de una fuente que, dependiendo del contexto, indique la cantidad de recursos que se debe reservar a la fuente. El contexto estará dado por el punto de operación que, como se verá, depende de la capacidad del enlace, del buffer, de las máximas pérdidas que se desean, y de otras fuentes que también alimenten el enlace. Una posibilidad es reservar el valor de pico de la fuente, es decir el máximo, en este caso no se tendrán pérdidas pero se estará desperdiciando capacidad. Si se reserva un valor muy cercano a la media se tendrá un buen aprovechamiento de los recursos, pero las pérdidas probablemente estén por encima de las deseadas.

El ancho de banda efectivo cuenta con propiedades que lo hacen interesante en la aplicación al estudio de redes de datos, la primera muy intuitiva y de fácil comprobación es que el ancho de banda efectivo requerido por una fuente que envía datos a velocidad constante C , es efectivamente C , ya que si la fuente envía datos a una velocidad constante C el trabajo acumulado será $X_t = Ct$ y sustituyendo en (1.1) se obtiene que $\alpha(s, t) = C$.

Probemos ahora que el ancho de banda es una cantidad entre la tasa máxima y la tasa media de transmisión.

Proposición 1.2 Para cada $t > 0$ se tiene que

$$\frac{EX_t}{t} \leq \alpha(s, t) \leq \frac{\bar{X}_t}{t} \quad \forall s > 0,$$

donde $\bar{X}_t := \inf\{x : P(X_t > x) = 0\}$ es el supremo esencial de la variable X_t .

Demostración: La función $\phi(x) = e^{sx}$ es una función convexa para cada $s > 0$, entonces aplicando la desigualdad de Jensen [35] se obtiene que

$$Ee^{sX_t} = E\phi(X_t) \geq \phi(EX_t) = e^{sEX_t},$$

y por lo tanto

$$\alpha(s, t) = \frac{1}{st} \log Ee^{sX_t} \geq \frac{1}{st} \log e^{sEX_t} = \frac{1}{st} sEX_t = \frac{EX_t}{t}.$$

Por otra parte

$$P(X_t > \bar{X}_t) = P\left(\bigcup_{n=1}^{\infty} \left\{X_t > \bar{X}_t + \frac{1}{n}\right\}\right) = \lim_{n \rightarrow \infty} P\left(X_t > \bar{X}_t + \frac{1}{n}\right) = 0,$$

pues $P(X_t > \bar{X}_t + \frac{1}{n}) = 0$ para todo valor n , esto significa que $X_t \leq \bar{X}_t$ c.s..

Aplicando el operador esperanza tenemos que

$$Ee^{sX_t} \leq Ee^{s\bar{X}_t} = e^{s\bar{X}_t},$$

por lo tanto

$$\alpha(s, t) \leq \frac{1}{st} \log e^{s\bar{X}_t} = \frac{\bar{X}_t}{t},$$

lo que prueba la otra desigualdad y concluye la demostración. \square

Esto significa que el parámetro s gradúa la ubicación del ancho de banda efectivo entre la “media” y el “máximo” como lo muestra la Figura 1.1.

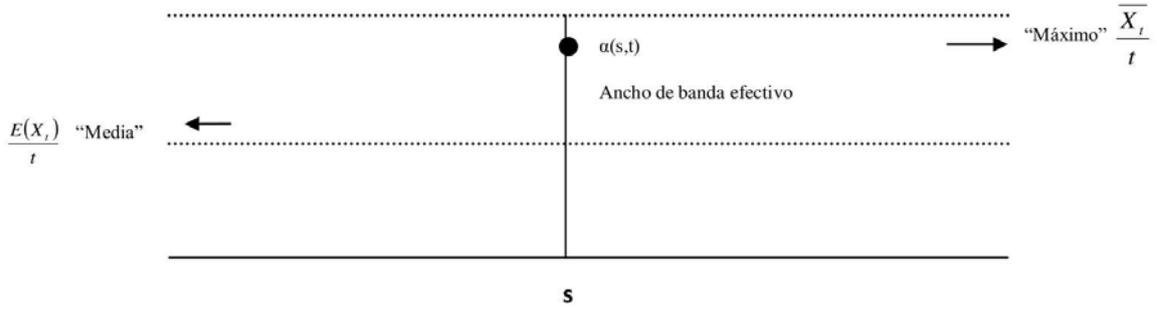


Figura 1.1: Interpretación del parámetro s .

Como ya dijimos, dimensionar en base a la media lleva a enormes pérdidas y dimensionar en base al máximo, a gran ineficiencia por ser un multiplexado pobre, por lo tanto el parámetro s está midiendo el grado de multiplexado que se considera, cuanto mayor grado de multiplexado, mayor seguridad y menor eficiencia.

Si los datos enviados por una fuente se pueden expresar como la suma de los datos enviados por fuentes independientes se puede demostrar que el ancho de banda tiene la propiedad aditiva.

Proposición 1.3 Si $X_t = \sum_{j=1}^n X_t^{(j)}$ donde las variables $\{X_t^{(j)}\}_{j=1, \dots, n}$ son independientes, se cumple que

$$\alpha(s, t) = \sum_{j=1}^n \alpha_j(s, t),$$

donde $\alpha(s, t)$ y $\alpha_j(s, t)$ son los anchos de banda efectivos de X_t y $X_t^{(j)}$ respectivamente.

Demostración: Aplicando propiedades del operador esperanza se tiene que

$$Ee^{sX_t} = Ee^{s(\sum_{j=1}^n X_t^{(j)})} = E \prod_{j=1}^n e^{sX_t^{(j)}} = \prod_{j=1}^n Ee^{sX_t^{(j)}}, \quad (1.2)$$

reemplazando (1.2) en la definición del ancho de banda efectivo se tiene

$$\alpha(s, t) = \frac{1}{st} \log Ee^{sX_t} = \frac{1}{st} \log \left(\prod_{j=1}^n Ee^{sX_t^{(j)}} \right) = \frac{1}{st} \sum_{j=1}^n \log Ee^{sX_t^{(j)}} = \sum_{j=1}^n \alpha_j(s, t).$$

□

Si en cada instante t la fuente despacha una cantidad proporcional a t , entonces el ancho de banda sólo dependerá del producto st y la demostración de la siguiente propiedad es inmediata.

Proposición 1.4 Si existe una variable aleatoria X tal que $X_t = Xt$, cualquiera sea $t > 0$ y $P(X \geq 0) = 1$, entonces

$$\alpha(s, t) = \alpha(st, 1).$$

Con respecto al comportamiento de $\alpha(s, t)$ como función de s , veremos que es creciente y que cuando $s = 0$ está determinada fundamentalmente por la esperanza y la varianza de X_t , mientras que cerca de $s = \infty$ está relacionado con la distribución de X_t cerca de su supremo esencial.

Proposición 1.5 Para cada t fijo, $\alpha(s, t)$ es creciente con s .

Demostración: Sea $0 < s_1 < s_2$, debemos probar que $\alpha(s_1, t) \leq \alpha(s_2, t)$ o lo que es equivalente $\alpha(s_2, t) - \alpha(s_1, t) \geq 0$. Como

$$\alpha(s_2, t) - \alpha(s_1, t) = \frac{1}{s_2 t} \log Ee^{s_2 X_t} - \frac{1}{s_1 t} \log Ee^{s_1 X_t} = \frac{1}{t} \log \left[\frac{(Ee^{s_2 X_t})^{\frac{1}{s_2}}}{(Ee^{s_1 X_t})^{\frac{1}{s_1}}} \right], \quad (1.3)$$

basta probar que el logaritmo en (1.3) es positivo, es decir que

$$\frac{(Ee^{s_2 X_t})^{\frac{1}{s_2}}}{(Ee^{s_1 X_t})^{\frac{1}{s_1}}} \geq 1,$$

y como $s_2 > 0$, alcanza con probar que $\left(\frac{(Ee^{s_2 X_t})^{\frac{1}{s_2}}}{(Ee^{s_1 X_t})^{\frac{1}{s_1}}} \right)^{s_2} = \frac{Ee^{s_2 X_t}}{(Ee^{s_1 X_t})^{\frac{s_2}{s_1}}} \geq 1$.

Por ser $s_2 > s_1$ la función $\phi(x) = x^{\frac{s_2}{s_1}}$ es convexa y aplicando la desigualdad de Jensen [35] se obtiene que

$$(Ee^{s_1 X_t})^{\frac{s_1}{s_2}} \leq E \left[(e^{s_1 X_t})^{\frac{s_2}{s_1}} \right] = Ee^{s_2 X_t},$$

lo que prueba la proposición. □

Proposición 1.6 Si $\alpha(s, t)$ está acotado para algún $s > 0$, entonces para cada t :

$$\alpha(s, t) = \frac{EX_t}{t} + \frac{s}{2t} \text{Var}(X_t) + o(s), \quad \text{cuando } s \rightarrow 0. \quad (1.4)$$

Si $\alpha(s, t)$ está acotado para $s \rightarrow \infty$ y X_t toma el valor \bar{X}_t con probabilidad positiva, entonces para cada t :

$$\alpha(s, t) = \frac{\bar{X}_t}{t} + \frac{1}{st} \log P(X_t = \bar{X}_t) + o\left(\frac{1}{s}\right), \quad \text{cuando } s \rightarrow \infty. \quad (1.5)$$

Demostración: Supongamos que $\alpha(s, t) < \infty$, para algún $s_0 > 0$, por la **Proposición 1.5**, $\alpha(s', t) < \infty$, para todo $s' < s_0$. Utilizando desarrollo en series de potencia, se verifica que

$$Ee^{sX_t} = 1 + sEX_t + s^2 \frac{EX_t^2}{2} + \cdots + s^n \frac{EX_t^n}{n!} + \cdots,$$

para todos los s tales que $Ee^{sX_t} < \infty$. En particular

$$Ee^{sX_t} = 1 + sEX_t + s^2 \frac{EX_t^2}{2} + o(s^2).$$

Recordemos que

$$\log(u + 1) = u - \frac{u^2}{2} + o(u^2), \quad \text{si } |u| < 1, \quad (1.6)$$

y como Ee^{sX_t} es continua y en $s = 0$ toma el valor 1, si tomamos s suficientemente pequeño tal que $|Ee^{sX_t} - 1| < 1$ entonces

$$\log Ee^{sX_t} = \log \left[1 + sEX_t + s^2 \frac{EX_t^2}{2} + o(s^2) \right].$$

Utilizando (1.6) con $u = sEX_t + s^2 \frac{EX_t^2}{2} + o(s^2)$, resulta que

$$\begin{aligned} \log Ee^{sX_t} &= sEX_t + s^2 \frac{EX_t^2}{2} + o(s^2) - \frac{1}{2} (s^2 (EX_t)^2 + o(s^2)) + o(s^2) \\ &= sEX_t + \frac{s^2}{2} (EX_t^2 - (EX_t)^2) + 2o(s^2) - \frac{o(s^2)}{2} \\ &= sEX_t + \frac{s^2}{2} \text{Var}(X_t) + o(s^2). \end{aligned} \quad (1.7)$$

Luego

$$\alpha(s, t) = \frac{1}{st} \left(sEX_t + \frac{s^2}{2} \text{Var}(X_t) + o(s^2) \right) = \frac{EX_t}{t} + \frac{s}{2t} \text{Var}(X_t) + o(s^2),$$

lo que prueba (1.4).

Para probar (1.5) debemos probar que $s \left(\alpha(s, t) - \frac{\bar{X}_t}{t} - \frac{1}{st} P(X_t = \bar{X}_t) \right) \xrightarrow{s \rightarrow \infty} 0$, o lo que es equivalente que

$$s \left(\alpha(s, t) - \frac{\bar{X}_t}{t} \right) \xrightarrow{s \rightarrow \infty} \frac{1}{t} P(X_t = \bar{X}_t).$$

En efecto,

$$\begin{aligned} s \left(\alpha(s, t) - \frac{\bar{X}_t}{t} \right) &= s \left(\frac{1}{st} \log E e^{sX_t} - \frac{\bar{X}_t}{t} \right) \\ &= \frac{1}{t} \log E e^{sX_t} - \frac{s\bar{X}_t}{t} \\ &= \frac{1}{t} \log E e^{sX_t} - \frac{1}{t} \log e^{s\bar{X}_t} \\ &= \frac{1}{t} \log \left[E e^{sX_t} e^{-s\bar{X}_t} \right] \\ &= \frac{1}{t} \log E \left(e^{s(X_t - \bar{X}_t)} \right) \\ &= \frac{1}{t} \log E \left(e^{s(X_t - \bar{X}_t)} \mathbb{I}\{X_t \leq \bar{X}_t\} \right). \end{aligned}$$

Para cada $\omega \in \Omega$ tenemos tres casos posibles

$$X_t(\omega) > \bar{X}_t \implies e^{s(X_t - \bar{X}_t)} \mathbb{I}\{X_t \leq \bar{X}_t\}(\omega) = 0 = \mathbb{I}\{X_t = \bar{X}_t\}(\omega) \quad \text{cualquiera sea } s,$$

$$X_t(\omega) < \bar{X}_t \implies e^{s(X_t - \bar{X}_t)} \mathbb{I}\{X_t \leq \bar{X}_t\}(\omega) \xrightarrow{s \rightarrow \infty} 0 = \mathbb{I}\{X_t = \bar{X}_t\}(\omega),$$

$X_t(\omega) = \bar{X}_t \implies e^{s(X_t - \bar{X}_t)} \mathbb{I}\{X_t \leq \bar{X}_t\}(\omega) = 1 = \mathbb{I}\{X_t = \bar{X}_t\}(\omega)$ cualquiera sea s , y por lo tanto $e^{s(X_t - \bar{X}_t)} \mathbb{I}\{X_t \leq \bar{X}_t\} \xrightarrow{s \rightarrow \infty} \mathbb{I}\{X_t = \bar{X}_t\}$, aplicando entonces el Teorema de Convergencia Dominada, [35], tenemos que

$$E \left[e^{s(X_t - \bar{X}_t)} \mathbb{I}\{X_t \leq \bar{X}_t\} \right] \xrightarrow{s \rightarrow \infty} E \mathbb{I}\{X_t = \bar{X}_t\} = P(X_t = \bar{X}_t).$$

Luego,

$$\frac{1}{t} \log E \left[e^{s(X_t - \bar{X}_t)} \mathbb{I}\{X_t \leq \bar{X}_t\} \right] \xrightarrow{s \rightarrow \infty} \frac{1}{t} \log P(X_t = \bar{X}_t),$$

lo que prueba (1.5). \square

Corolario 1.7 *De la proposición anterior se deduce que:*

$$\lim_{s \rightarrow 0} \alpha(s, t) = \frac{EX_t}{t} := \alpha(0, t),$$

$$\lim_{s \rightarrow \infty} \alpha(s, t) = \frac{\bar{X}_t}{t} := \alpha(\infty, t).$$

Hasta ahora hemos considerado que X_t es un proceso que tiene incrementos estacionarios, veamos ahora que características o propiedades posee esta magnitud si el proceso además tiene incrementos independientes.

Proposición 1.8 *Si X_t tiene incrementos independientes, entonces $\alpha(s, t)$ no depende de t .*

Demostración: Consideremos primero un número natural arbitrario m , dados s y t tenemos que

$$Ee^{sX_{mt}} = Ee^{s(\sum_{k=1}^m X_{kt} - X_{(k-1)t})} = E \prod_{k=1}^m e^{s(X_{kt} - X_{(k-1)t})} = \prod_{k=1}^m Ee^{s(X_{kt} - X_{(k-1)t})}.$$

Dado que X_t es un proceso de incrementos estacionarios, X_t y $X_{kt} - X_{(k-1)t}$ tienen la misma distribución para cualquier valor de k desde 1 hasta m y entonces

$$Ee^{sX_{mt}} = \prod_{k=1}^m Ee^{sX_t} = (Ee^{sX_t})^m,$$

y por lo tanto

$$\alpha(s, mt) = \frac{1}{smt} \log Ee^{sX_{mt}} = \frac{1}{smt} \log (Ee^{sX_t})^m = \frac{1}{st} \log Ee^{sX_t} = \alpha(s, t), \quad (1.8)$$

en particular, para $t = 1$ se cumple que $\alpha(s, m) = \alpha(s, 1)$ para cualquier número natural.

Consideremos ahora t racional, es decir $t = \frac{p}{q}$ con p y $q \in \mathbb{N}$. Utilizando (1.8) con $m = q$ resulta que $\alpha(s, \frac{p}{q}) = \alpha(s, q\frac{p}{q}) = \alpha(s, p) = \alpha(s, 1)$, es decir $\alpha(s, t) = \alpha(s, 1)$ cualquiera sea t racional.

Sea ahora t un real no racional, los procesos con incrementos estacionarios e independientes cumplen con la propiedad que $P(\{\omega : X_t(\omega) \text{ es discontinua en } t\}) = 0$, por lo tanto

$$Ee^{sX_t} = E(e^{sX_t} \mathbb{1}_{\{X_t \text{ continua en } t\}}) = E\left(\lim_n e^{sX_{t_n}} \mathbb{1}_{\{X_t \text{ continua en } t\}}\right),$$

donde t_n es una sucesión de racionales que tiende hacia t en forma creciente. Como $s > 0$, la convergencia es monótona y aplicando el Lema de Fatou, [35], se tiene que

$$Ee^{sX_t} = \lim_n E(e^{sX_{t_n}} \mathbb{1}_{\{X_t \text{ continua en } t\}}) = \lim_n E(e^{sX_{t_n}}),$$

es decir

$$\alpha(s, t) = \frac{1}{st} \log Ee^{sX_t} = \lim_n \frac{1}{st_n} \log Ee^{sX_{t_n}} = \lim_n \alpha(s, t_n) = \alpha(s, 1),$$

ya que los t_n son racionales, se tiene entonces que $\alpha(s, t) = \alpha(s, 1)$, cualquiera sea t real positivo. \square

El ancho de banda efectivo puede calcularse explícitamente para distintas distribuciones de los procesos X_t , veamos algunos ejemplos.

Ejemplo 1.9 *Flujo de velocidad aleatoria.*

Vimos, inmediatamente a continuación de la definición de ancho de banda efectivo, que si el proceso es un flujo de velocidad constante de la forma $X_t = Ct$, entonces $\alpha(s, t) = C$. Una generalización de este ejemplo es considerar $X_t = Xt$, donde X es una variable aleatoria tal que $P(X \geq 0) = 1$, por ejemplo supongamos que X tiene una distribución exponencial del parámetro λ , entonces calculando

$$Ee^{stX} = \int_{k=0}^{+\infty} e^{stx} \lambda e^{-\lambda x} dx = \lambda \int_0^{+\infty} e^{(st-\lambda)x} dx = \frac{\lambda}{\lambda - st},$$

en la región $s, t \geq 0$ y $st \leq \lambda$, resulta

$$\alpha(s, t) = \frac{1}{st} \log \left(\frac{\lambda}{\lambda - st} \right), \quad (1.9)$$

que efectivamente depende sólo del producto st como lo afirma la **Proposición 1.4**.

Para el tráfico telefónico ha sido usado exitosamente el modelo de Poisson desde el trabajo presentado por Erlang en la primera parte del siglo veinte donde se presenta un modelo que describe el tráfico de conversación telefónica con propiedades teóricas que lo hacen fácil de analizar, por lo tanto cualquier trabajo sobre telecomunicación debe abarcar al modelo de Poisson como un caso especial. Veamos ahora como calculamos explícitamente el ancho de banda efectivo para dicho modelo.

Ejemplo 1.10 *Proceso de Poisson.*

Sea X_t un proceso de Poisson, es decir $X_t = \text{máx}\{k : T_1 + \dots + T_k \leq t\}$ donde $\{T_n\}_{n \in \mathbb{N}}$ son variables aleatorias independientes e idénticamente distribuidas con distribución exponencial con parámetro λ , el modelo se puede interpretar como la llegada de una unidad de trabajo cuando ocurre cada una de las exponenciales T_n y equivale a contar el número de eventos ocurridos hasta el tiempo t .

Tenemos entonces que la distribución de X_t es $P(X_t = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$, para $k \in \mathbb{N} \cup \{0\}$ y calculando la esperanza indicada en (1.1) tenemos que

$$Ee^{sX_t} = \sum_{k=0}^{\infty} e^{sk} \frac{(\lambda t)^k}{k!} e^{-\lambda t} = e^{-\lambda t} \sum_{k=0}^{\infty} \frac{(e^s \lambda t)^k}{k!} = e^{-\lambda t} e^{e^s \lambda t} = e^{\lambda t(e^s - 1)},$$

y por lo tanto

$$\alpha(s, t) = \frac{1}{st} \log e^{\lambda t(e^s - 1)} = \frac{1}{st} (\lambda t(e^s - 1)) = \lambda \frac{(e^s - 1)}{s}. \quad (1.10)$$

Generalizando el modelo a la llegada de b unidades de trabajo cuando ocurre cada una de las exponenciales, tendríamos el proceso $Y_t = bX_t$, con X_t definido como antes y en este caso

$$Ee^{sY_t} = Ee^{sbX_t} = \sum_{k=0}^{\infty} e^{sbk} \frac{(\lambda t)^k}{k!} e^{-\lambda t} = e^{\lambda t(e^{sb} - 1)},$$

y por lo tanto

$$\alpha(s, t) = \lambda \frac{(e^{sb} - 1)}{s}. \quad (1.11)$$

Si la cantidad de unidades de trabajo en cada arribo es aleatoria, se obtiene un proceso de Poisson Compuesto, en este caso se tiene una sucesión $\{T_n\}_{n \in \mathbb{N}}$ de variables aleatorias independientes e idénticamente distribuidas con distribución esponencial con parámetro λ y una sucesión $\{W_n\}_{n \in \mathbb{N}}$ de variables aleatorias también independientes e idénticamente distribuidas con distribución F , independientes de las variables $\{T_n\}$. En este caso el proceso de entrada es

$$Y_t = \sum_{k=1}^{X_t} W_k, \quad (1.12)$$

donde X_t es el número de llegadas antes de t , tenemos entonces que

$$e^{sY_t} = e^{s \sum_{k=1}^{X_t} W_k} = \sum_{n=0}^{\infty} e^{s \sum_{k=1}^n W_k} \mathbb{I}_{\{X_t=n\}} = \lim_{N \rightarrow \infty} \sum_{n=0}^N e^{s \sum_{k=1}^n W_k} \mathbb{I}_{\{X_t=n\}},$$

donde la última convergencia es monótona. Teniendo en cuenta la independencia de W_n y X_t y la idéntica distribución se tiene que

$$\begin{aligned} Ee^{sY_t} &= \sum_{n=0}^{\infty} E \left(e^{s \sum_{k=1}^n W_k} \mathbb{I}_{\{X_t=n\}} \right) \\ &= \sum_{n=0}^{\infty} E \left(e^{s \sum_{k=1}^n W_k} \right) P(X_t = n) \\ &= \sum_{n=0}^{\infty} (Ee^{sW_1})^n e^{-\lambda t} \frac{(\lambda t)^n}{n!} \\ &= e^{-\lambda t} \sum_{n=0}^{\infty} (Ee^{sW_1})^n \frac{(\lambda t)^n}{n!} \\ &= e^{-\lambda t} e^{\lambda t (Ee^{sW_1})} \\ &= e^{\lambda t (Ee^{sW_1} - 1)}, \end{aligned}$$

y por lo tanto

$$\alpha(s, t) = \frac{1}{st} \log e^{\lambda t (Ee^{sW_1} - 1)} = \frac{1}{st} \lambda t (Ee^{sW_1} - 1) = \frac{\lambda}{s} (Ee^{sW_1} - 1). \quad (1.13)$$

Observemos que los casos anteriores, son casos particulares tomando $W_n = 1$ o $W_n = b$ para todo n y en el caso particular en que la distribución de W_n sea exponencial con parámetro μ , tendremos que

$$\alpha(s, t) = \frac{\lambda}{\mu - s}.$$

Ejemplo 1.11 *Fuentes Periódicas*

Una fuente periódica es una fuente que envía b unidades de trabajo cada d unidades de tiempo. El proceso se define como

$$X_t = b \max\{n \geq 1 : Ud + (n-1)d \leq t\}, \quad (1.14)$$

donde la variable U tiene distribución uniforme en el intervalo $[0, 1]$. El tiempo T_n de la n -ésima llegada está dado por $T_n = Ud + (n-1)d$ que tiene distribución uniforme en el intervalo $[(n-1)d, nd]$, en particular T_1 tiene distribución uniforme en $[0, d]$.

Calculemos la distribución de X_t ,

$$P(X_t = kb) = P(\max\{n \geq 1 : Ud + (n-1)d \leq t\} = k),$$

pero $\max\{n \geq 1 : Ud + (n-1)d \leq t\} = k$ sólo si $Ud + (n-1)d \leq t$ y $Ud + kd > t$, o lo que es equivalente sólo si $Ud + kd \in (t, t+d]$.

Entonces, recordando que la variable $Ud + kd$ tiene distribución uniforme en el intervalo $[kd, (k+1)d]$ se tiene que $P(X_t = kb) = \frac{\text{long}((t, t+d] \cap [kd, (k+1)d])}{d}$ es decir

$$P(X_t = kb) = \begin{cases} 0 & \text{si } k < \frac{t}{d} - 1 \\ -\frac{t}{d} + (k+1) & \text{si } k \in (\frac{t}{d} - 1, \frac{t}{d}) \\ \frac{t}{d} - (k-1) & \text{si } k \in (\frac{t}{d}, \frac{t}{d} + 1) \\ 0 & \text{si } k > \frac{t}{d} + 1 \end{cases},$$

es decir X_t toma casi seguramente los valores $[\frac{t}{d}]b$ ó $([\frac{t}{d}] + 1)b$ con las siguientes probabilidades,

$$P\left(X_t = \left[\frac{t}{d}\right]b\right) = -\frac{t}{d} + \left(\left[\frac{t}{d}\right] + 1\right) = 1 - \left(\frac{t}{d} - \left[\frac{t}{d}\right]\right),$$

$$P\left(X_t = \left(\left[\frac{t}{d}\right] + 1\right)b\right) = \frac{t}{d} - \left(\left[\frac{t}{d}\right] + 1\right) - 1 = \frac{t}{d} - \left[\frac{t}{d}\right].$$

Entonces

$$\begin{aligned} Ee^{sX_t} &= e^{s[\frac{t}{d}]b} P\left(X_t = \left[\frac{t}{d}\right]b\right) + e^{s([\frac{t}{d}] + 1)b} P\left(X_t = \left(\left[\frac{t}{d}\right] + 1\right)b\right) \\ &= e^{s[\frac{t}{d}]b} \left(1 - \left(\frac{t}{d} - \left[\frac{t}{d}\right]\right)\right) + e^{s([\frac{t}{d}] + 1)b} \left(\frac{t}{d} - \left[\frac{t}{d}\right]\right) \\ &= e^{s[\frac{t}{d}]b} \left[\left(\frac{t}{d} - \left[\frac{t}{d}\right]\right) (e^{sb} - 1) + 1\right], \end{aligned}$$

por lo tanto

$$\alpha(s, t) = \left[\frac{t}{d}\right] \frac{b}{t} + \frac{1}{st} \log \left(1 + (e^{sb} - 1) \left(\frac{t}{d} - \left[\frac{t}{d}\right]\right)\right), \quad (1.15)$$

y además se cumple que

$$\lim_{t \rightarrow 0} \alpha(s, t) = \frac{e^{sb} - 1}{sd}.$$

Ejemplo 1.12 *Proceso de conteo generalizado*

Sea $\{\Delta T_n\}_{n \geq 1}$ una sucesión de variables aleatorias tales que $P(\Delta T_n > 0) = 1$ para todos los valores de n . Definimos las variables $T_n = \sum_{k=1}^n \Delta T_k$ y $T_0 = 0$; se tiene entonces que las variables T_n para $n \geq 1$, que representan los tiempos de llegada de trabajo cumplen que $0 = T_0 < T_1 < T_2 < \dots$, con probabilidad 1. Definimos el proceso de conteo como

$$N_t = \text{máx}\{k : T_k \leq t\},$$

es decir la cantidad de llegadas hasta el tiempo t donde los tiempos de llegada están dados por las variables $\{\Delta T_n\}$. Observemos que este proceso tiene, con probabilidad 1, trayectorias crecientes, constantes a tramos y saltos de altura 1.

Consideremos también la sucesión $\{W_n\}_{n \geq 1}$ de variables aleatorias independientes e idénticamente distribuidas que además son independientes de las $\{\Delta T_n\}$, por lo tanto las variables $\{W_n\}$ son independientes del proceso N_t que es una función de las variables $\{\Delta T_n\}$ y la cantidad de trabajo acumulado hasta t es entonces

$$X_t = \sum_{k=0}^{N_t} W_k,$$

con la convención $W_0 = 0$.

La primera observación que debemos hacer es que si elegimos $\{\Delta T_n\}$ con distribución exponencial con parámetro λ , entonces T_n tendrá la distribución de Erlang, $\Gamma(n, \lambda)$ y N_t será el proceso de Poisson compuesto (1.12).

Para calcular el ancho de banda efectivo tenemos que

$$e^{sX_t} = e^{s \sum_{k=0}^{N_t} W_k} = \sum_{n=0}^{\infty} e^{s \sum_{k=0}^n W_k} \mathbb{I}_{\{N_t=n\}} = \lim_N \sum_{n=0}^N e^{s \sum_{k=0}^n W_k} \mathbb{I}_{\{N_t=n\}},$$

donde la última convergencia es monótona. Por lo tanto

$$\begin{aligned} Ee^{sX_t} &= \sum_{n=0}^{\infty} Ee^{s \sum_{k=0}^n W_k} \mathbb{I}_{\{N_t=n\}} \\ &= \sum_{n=0}^{\infty} Ee^{s \sum_{k=0}^n W_k} E(\mathbb{I}_{\{N_t=n\}}) \\ &= \sum_{n=0}^{\infty} Ee^{s \sum_{k=0}^n W_k} P(N_t = n), \end{aligned}$$

por la independencia de las W_k y N_t , y como además las W_k son independientes e idénticamente distribuidas, y $Ee^{sW_0} = 1$ por ser $e^{sW_0} = 1$ casi seguramente, tenemos que

$$E\left(e^{s \sum_{k=0}^n W_k}\right) = (Ee^{sW})^n,$$

donde W representa una variable W_k arbitraria.

Por lo tanto el ancho de banda efectivo resulta ser

$$\alpha(s, t) = \frac{1}{st} \log \sum_{n=0}^{\infty} P(N_t = n) (Ee^{sW})^n, \quad (1.16)$$

que se puede escribir como

$$\alpha(s, t) = \frac{1}{st} \log (g_t(Ee^{sW})), \quad (1.17)$$

donde g_t es la *función generatriz de probabilidades* de N_t definida por

$$g_t(x) = E[x^{N_t}] = \sum_{n=0}^{\infty} x^n P(N_t = n).$$

Observación 1.13

- a) Si N_t es un proceso de Poisson de parámetro λ , entonces la función generatriz de probabilidades es $g_t(x) = e^{-\lambda t} \sum_{n=0}^{\infty} \frac{(x\lambda t)^n}{n!} = e^{\lambda t(x-1)}$ que al sustituirla en (1.17) nos da el resultado obtenido en el **Ejemplo 1.10**.
- b) Si N_t es el de una fuente periódica, tomando $d = 1$ para simplificar, entonces la función generatriz de probabilidades es $g_t(x) = x^{[t]}(1 - (t - [t])) + x^{[t]+1}(t - [t])$, que al sustituirla en (1.17) nos da el resultado obtenido en el **Ejemplo 1.11**.

Observación 1.14

Lo destacable de la fórmula (1.17) es que separa claramente cómo afecta el proceso de conteo y cómo afectan las variables de trabajo al ancho de banda efectivo.

Veamos ahora cómo calculamos explícitamente el ancho de banda efectivo para fuentes gaussianas ya que el movimiento Browniano ha sido usado como un caso límite del tráfico de paquetes pesados.

Ejemplo 1.15 Fuente gaussiana

Sea $X_t = \lambda t + \sigma W_t$, donde W_t es un proceso de Wiener. W_t es un proceso de incrementos estacionarios e independientes tal que $W_t \sim N(0, t)$, entonces $X_t \sim N(\lambda t, \sigma^2 t)$ y por lo tanto

$$Ee^{sX_t} = Ee^{s(\lambda t + \sigma W_t)} = e^{s\lambda t} Ee^{s\sigma W_t}.$$

Llamando $Z_t = \frac{W_t}{\sqrt{t}}$ tenemos que $Z_t \sim N(0, 1)$ con lo que resulta

$$Ee^{sX_t} = e^{s\lambda t} Ee^{s\sigma\sqrt{t}Z_t} = e^{s\lambda t + \frac{1}{2}s^2\sigma^2 t},$$

y entonces

$$\alpha(s, t) = \frac{1}{st} \log E e^{sX_t} = \frac{1}{st} \left(s\lambda t + \frac{1}{2}s^2\sigma^2 t \right) = \lambda + \frac{1}{2}s\sigma^2, \quad (1.18)$$

que no depende de t , ya que el proceso tiene incrementos independientes, además es lineal en s , verificando así la ecuación (1.4) de la **Proposición 1.6**.

Observación 1.16

En realidad X_t debería ser no negativo, propiedad que no se cumple en el caso gaussiano, sin embargo este modelo es admisible si σ es pequeño en relación con λ de modo que la probabilidad de X_t tome valores negativos sea pequeña.

1.3. Modelos Markovianos

Los procesos de Markov proporcionan medios muy flexibles, potentes y eficientes para la descripción y el análisis de las propiedades de un sistema dinámico ya que, entre otras cosas, nos permiten obtener fácilmente medidas de performance y la fiabilidad del sistema. Por otra parte, los procesos de Markov constituyen la teoría fundamental que subyace en el concepto de sistemas de colas o trabajos en espera, dado que cada sistema de cola puede ser representado mediante un proceso de Markov.

Además de destacar la relación entre el proceso de Markov y sistemas de colas, vale la pena señalar también que las propiedades fundamentales de la teoría de colas son frecuentemente demostradas en términos del proceso de Markov subyacente. En particular las cadenas de Markov son una herramienta para analizar el comportamiento de determinados tipos de procesos estocásticos, procesos que evolucionan en forma no determinística a lo largo del tiempo, en torno a un conjunto de estados.

Vamos a considerar cadenas de Markov a tiempo continuo y aunque varios de los conceptos que se mencionarán para este tipo de procesos son análogos al caso de cadenas a tiempo discreto, existen diferencias en el comportamiento y desarrollo matemático entre ambos modelos. No se pretende presentar y desarrollar la teoría completa de cadenas de Markov a tiempo continuo, sólo se presentarán algunos de los conceptos que serán utilizados a lo largo de este trabajo.

1.3.1. Cadenas de Markov

Una cadena de Markov a tiempo continuo es un proceso estocástico $\{X_t\}_{t \geq 0}$ definido sobre el espacio de probabilidades (Ω, A, P) que toma valores en un espacio numerable E , también llamado espacio de estados, y que cumple la propiedad de Markov

$$P(X_{t+s} = j | X_t = i, X_{t_1} = j_1, \dots, X_{t_k} = j_k) = P(X_{t+s} = j | X_t = i) := P_{ij}(t, s), \quad (1.19)$$

para cualquier k , $0 \leq t_1 < t_2 < \dots < t$ y para todos los posibles valores de $j_1, j_2, \dots, j_k, i, j$ pertenecientes al espacio de estados E de la cadena.

La ecuación (1.19) se puede interpretar como que toda la historia de la cadena de Markov puede ser resumida con lo que ocurre en el presente estado del proceso o, equivalentemente, que dado el pasado, el futuro es condicionalmente independiente del pasado.

Obsérvese que no estamos suponiendo que se conoce la historia del proceso en todo el pasado a tiempo continuo, sino únicamente en una colección arbitraria pero finita de tiempos pasados t_1, t_2, \dots, t_k, t . Toda cadena de Markov a tiempo continuo cada vez que entra en un estado $i \in E$ verifica que la distribución del tiempo que permanece antes de transitar a otro estado, es exponencial y además cuando deja el estado i verifica que $\sum_{j \neq i \in E} P_{ij} = 1$, donde P_{ij} es la probabilidad de que, dado que salió del estado i , pase al estado j .

Supondremos que estas probabilidades de transición son estacionarias en el tiempo. Es decir, no dependen de t , cumpliendo entonces que

$$P(X_{t+s} = j | X_t = i) = P(X_s = j | X_0 = i),$$

si pasa esto la cadena se llama homogénea y notaremos $P_{ij}(t)$ a la probabilidad de transición del estado i al estado j en un tiempo t , en particular para $t = 0$ se define $P_{ij}(0)$ como la función delta de Kronecker, es decir

$$P_{ij}(0) = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases}.$$

Denominamos matriz de transición de la cadena a $P(t) = (P_{ij}(t))_{i,j \in E}$, y a la familia $\{P(t)\}_{t \geq 0}$ se la denomina semigrupo de transición. Por convención, $P(0) = I$ la matriz identidad de dimensión coincidente con la dimensión del espacio de estados E .

Las probabilidades de transición satisfacen también una versión continua de la *Ecuación de Chapman-Kolmogorov*, [52], que en este caso se conoce también como la propiedad de semigrupo, enunciada en la siguiente proposición y que se obtiene partir de la propiedad de Markov

Proposición 1.17 *Para cualquier par de estados i y j y para todo $t \geq 0, s \geq 0$*

$$P_{ij}(t+s) = \sum_{k \in E} P_{ik}(t)P_{kj}(s). \quad (1.20)$$

En notación matricial, $P(t+s) = P(t)P(s)$.

La ecuación de Chapman-Kolmogorov es de interés porque permite expresar a la probabilidad $P_{ij}(t)$, para cualquier tiempo t positivo, en términos de probabilidades infinitesimales, es decir probabilidades de transición en intervalos de tiempo de longitud muy pequeña. Por ejemplo, para cualquier natural n , tomando $\Delta t = \frac{t}{n}$ se tiene que

$$P_{ij}(t) = \sum_{k_1, \dots, k_{n-1}} P_{ik_1}(\Delta t)P_{k_1 k_2}(\Delta t) \cdots P_{k_{n-1} j}(\Delta t),$$

esto quiere decir que es suficiente conocer el comportamiento de $P_{ij}(t)$ en tiempos t pequeños para conocer su comportamiento para cualquier tiempo t positivo.

Si notamos con $\pi(t)$ la distribución de la cadena en el tiempo t , la i -ésima componente de dicho vector de probabilidades está dado por $\pi_i(t) = \sum_{j \in E} \pi_j(0) P_{ji}(t)$, en término de matrices tendremos que $\pi(t) = \pi(0)P(t)$, es decir para todo tiempo t positivo, la distribución de la cadena sólo depende de la distribución inicial de la cadena y del semigrupo de transición.

También se puede probar que para cualquier $0 \leq t_1 < t_2 < \dots < t_k$ y para todo subconjunto j_0, j_1, \dots, j_k del conjunto de estados E , se cumple

$$P(X_{t_1} = j_1, \dots, X_{t_k} = j_k) = \sum_{j_0 \in E} P(X_0 = j_0) \prod_{i=1}^k P_{j_{i-1}j_i}(t_{j_i} - t_{j_{i-1}}),$$

que expresa que las distribuciones conjuntas del proceso también dependen sólo de la distribución inicial y del semigrupo de transición.

A diferencia del caso discreto, el caso continuo no puede resolverse fácilmente usando el cálculo de las probabilidades de los diferentes estados, más bien debe transformarse en un sistema de ecuaciones diferenciales que nos conduzca a los resultados necesarios. Para este fin definimos $q_{ij}(t)$ la tasa de transición instantánea de la cadena de Markov a tiempo continuo del estado i al estado j , para $i \neq j$, relacionándola con la probabilidad condicional de transición.

Consideremos el período $[t, t + \Delta t]$, como $P(t)$ es continua en el origen, ya que cuando t se aproxima a 0 se aproxima a $P(0) = I$, se demuestra la existencia para cada estado y bajo condiciones bastante generales, de las funciones no negativas, continuas y finitas definidas por

$$q_{ij}(t) = \lim_{\Delta t \rightarrow 0} \frac{P_{ij}(t, t + \Delta t) - \delta_{ij}}{\Delta t} \quad i \neq j, \quad (1.21)$$

$$q_{ii}(t) = \lim_{\Delta t \rightarrow 0} \frac{P_{ii}(t, t + \Delta t) - 1}{\Delta t}. \quad (1.22)$$

Como $\sum_{j \in E} P_{ij}(t, t + \Delta t) = 1$, es claro que para cualquier instante t se cumple que $\sum_{j \in E} q_{ij}(t) = 0$.

Se llama generador infinitesimal de la cadena a la matriz $\mathbb{Q} = (q_{ij})_{i,j \in E}$ donde q_{ij} y q_{ii} están definidos como en (1.21) y (1.22) respectivamente y cada uno tiene una interpretación concreta. El valor q_{ij} se interpreta como la cantidad de transiciones del estado i al j por unidad de tiempo, esto es la velocidad con la que la cadena deja el estado i para pasar al estado j en el instante t . La cantidad $q_i = -q_{ii}$ puede interpretarse como la cantidad de transiciones que salen del estado i por unidad de tiempo o la velocidad total a la que es abandonado el estado i para pasar a cualquier otro estado en el instante t . Se tiene entonces que

$$q_{ij} = \left. \frac{dP_{ij}(t)}{dt} \right|_{t=0},$$

en términos de matrices se tiene que \mathbb{Q} es la derivada de la función matricial $P(t)$ evaluada en $t = 0$.

A partir de las tasas de transición se puede reconstruir la cadena. Si un estado i es absorbente entonces $q_i = 0$, si es estable $0 < q_i < \infty$ y si es instantáneo $q_i = \infty$. La cadena deja los estados instantáneos de forma inmediata, así que siempre asumiremos $q_i < \infty$. Si el estado i es absorbente, la cadena permanecerá en i por siempre. Por el contrario, si el estado es estable, la cadena permanecerá en i un tiempo aleatorio distribuido exponencialmente con tasa q_i y luego saltará a otro estado j con probabilidad de transición $P_{ij} = \frac{q_{ij}}{q_i}$.

Usando recursivamente este procedimiento construimos la cadena y como la cadena está determinada por las probabilidades de transición, entonces deberíamos poder calcular estas últimas a partir de las tasas de transición. La formalización de esta construcción se expresa en el siguiente resultado:

Teorema 1.3.1 (*Ecuaciones diferenciales de Kolmogorov*) *Supongamos que $q_i < \infty$ para cada estado $i \in E$, entonces las probabilidades de transición $P_{ij}(t)$ son diferenciables para todo $t \geq 0$ y para cualquier par de estados $i, j \in E$ se tiene*

$$P'_{ij}(t) = \lim_{\Delta t \rightarrow 0} \sum_{k \neq i} q_{ik} P_{kj}(t) - q_i P_{ij}(t), \quad \text{Ecuación Backward}$$

$$P'_{ij}(t) = \lim_{\Delta t \rightarrow 0} \sum_{k \neq j} q_{kj} P_{ik}(t) - q_j P_{ij}(t), \quad \text{Ecuación Forward}$$

y las condiciones iniciales para ambos conjuntos de ecuaciones son $P(0) = I$.

Formalmente, la solución de los conjuntos de ecuaciones diferenciales de Kolmogorov puede ser expresada como

$$P(t) = \exp(\mathbb{Q}t). \quad (1.23)$$

Cuando \mathbb{Q} es una matriz de dimensión finita, la serie anterior es convergente y es la única solución para los dos sistemas de ecuaciones, si es de dimensión infinita no podemos afirmar nada.

Al igual que en el caso discreto, la distribución estacionaria de la cadena de Markov a tiempo continuo es de interés y posee propiedades equivalentes en ambos casos. Para cualquier estado $i \in E$ la probabilidad π_i es independiente del tiempo, esto es $\pi = \pi P(t)$ para todo t , y también de la distribución inicial $\pi(0)$. En el caso que la distribución inicial es una distribución invariante se cumple que $\pi(t) = \pi(0)P(t) = \pi(0)$

Si existe la distribución estacionaria para la cadena, al ser independiente del tiempo se tiene que $\frac{d\pi(t)}{dt} = 0$ y se puede demostrar que si π es una distribución estacionaria y cumple que $\pi = \pi P(t)$ es equivalente a decir que $\pi \mathbb{Q} = 0$.

Algunos conceptos y propiedades de los procesos de Markov a tiempo continuo son similares a los desarrollados para el caso discreto ya que toda cadena de Markov a tiempo continuo X_t tiene asociada una cadena a tiempo discreto que denotamos X_n y que está definida por $X_n = X_{\tau_n}$ donde τ_n , con $n \geq 0$, es la sucesión de los tiempos de transición de la cadena, $\tau_0 = 0$ y $\tau_n = \infty$ si hay menos de n transiciones en $(0, \infty)$.

Algunas características de la cadena a tiempo discreto se trasladan a su versión continua y a partir de una cadena a tiempo discreto puede construirse su versión continua en el tiempo tomando tiempos exponenciales independientes entre salto y salto.

En una cadena de Markov de tiempos discretos un estado i se dice transitorio si, partiendo de i , con probabilidad uno el conjunto de tiempos $\{n \geq 0 : X_n = i\}$ es acotado. En cambio, se dice que es recurrente si, partiendo de i , con probabilidad uno el conjunto de tiempos $\{n \geq 0 : X_n = i\}$ es no acotado. Si el tiempo medio de recurrencia es finito, el estado se dirá recurrente positivo, caso contrario se dirá recurrente nulo. Además si todos los estados se comunican entre si, la cadena se dice irreducible.

Observación 1.18

No existen estados recurrentes nulos en cadenas de Markov finitas.

Generalizamos estos conceptos a cadenas de tiempo continuo, con la ayuda de la cadena discreta asociada, de la siguiente manera.

Definición 1.19 Una cadena de Markov a tiempo continuo X_t se dice irreducible si la cadena X_n a tiempo discreto definida como antes es irreducible.

Definición 1.20 Un estado i del espacio E se dice recurrente para la cadena de Markov a tiempo continuo X_t , si es recurrente para la cadena X_n a tiempo discreto definida como antes. Además la cadena X_t se dice recurrente si todos sus estados son recurrentes.

El siguiente teorema nos da condiciones que nos aseguran cuando esta distribución estacionaria existe y es única en una cadena de Markov a tiempo continuo.

Teorema 1.21 Si X_t es una cadena de Markov homogénea, irreducible y recurrente positiva entonces existe una única distribución invariante y se cumple además que

$$\lim_{t \rightarrow +\infty} P(X_t = x_j) = \pi_j,$$

para cualquier distribución inicial.

También es de interés conocer el comportamiento probabilista de un proceso luego de una larga evolución y una herramienta para ello la da el siguiente teorema,

Teorema 1.22 (Teorema Ergódico) Si X_t es una cadena de Markov homogénea, irreducible y recurrente entonces para toda función $f : E \rightarrow R^+$ se cumple que

$$\lim_{t \rightarrow +\infty} \frac{1}{t} \int_0^t f(X_t) dt = E_\pi(f(X_t)) = \sum_{j \in E} f(j) \pi_j.$$

1.3.2. Flujos Markovianos

Sea Y_t una Cadena de Markov a tiempo continuo, homogénea, irreducible y con espacio de estados finito $E = \{h_1, h_2, \dots, h_m\}$, con $h_i \in \mathbb{R}$ para todo i y tal que se cumple que $0 \leq h_1 < h_2 < \dots < h_m$, definimos Flujo Markoviano modulado por Y_t al proceso

$$X_t = \int_0^t Y_s ds. \quad (1.24)$$

Dado que Y_t es una Cadena de Markov homogénea, irreducible y recurrente positiva por tener una cantidad finita de estados, sabemos por el **Teorema 1.21** que existe una única distribución invariante π , este resultado nos asegura que tomando como distribución inicial la invariante, Y_s resulta un proceso estacionario y por lo tanto X_t es un proceso de incrementos estacionarios y tiene sentido calcular su ancho de banda efectivo. Este modelo es importante ya que es aceptado en la literatura para datos provenientes de fuentes de voz o de video MPEG, [3].

El proceso X_t representa el trabajo acumulado en el intervalo de tiempo $[0, t]$ recibido desde una fuente que envía datos a velocidad Y_t y los estados h_i de la cadena X_t representan la velocidad con que se acumula trabajo. Dichas velocidades son constantes y cambian aleatoriamente, por lo cual el proceso X_t tiene trayectorias lineales, continuas a tramos y con pendientes dadas por los estados h_i . La cadena Y_t se llama cadena modulante.

1.4. Cálculo del Ancho de Banda Efectivo para Flujos Markovianos

En esta sección mostramos la expresión del ancho de banda efectivo para el modelo de Flujo Markoviano presentado en la sección anterior. Consideremos que los nodos de una red de transferencia de datos pueden asumir distintos estados y esta situación se modela mediante una cadena de Markov homogénea a tiempo continuo Y_t , con espacio de estados finito y distribución invariante π y para cada estado i se conoce la velocidad h_i , constante, con la que la fuente despacha datos.

En el siguiente teorema, que daremos sin demostración, Keisidis, Walrang y Chang presentaron en [42] una expresión para el ancho de banda efectivo de un Flujo Markoviano X_t definido en (1.24).

Teorema 1.23 *Sea $\{X_t\}_{t \geq 0}$ un flujo markoviano modulado por la cadena de Markov irreducible homogénea Y_t definida como antes, con generador infinitesimal \mathbb{Q}^Y y espacio de estados $0 \leq h_1 < h_2 < \dots < h_m$. Sea \mathbb{H} una matriz diagonal con $(H_{ij})_{1 \leq i, j \leq m} = \begin{cases} h_i & i = j \\ 0 & i \neq j \end{cases}$, π la distribución invariante de la cadena y $\mathbf{1}$, un vector columna con todas las entradas iguales a 1.*

Entonces

$$\alpha(s, t) = \frac{1}{st} \log \{ \pi \exp [(\mathbb{Q}^Y + s\mathbb{H}) t] \mathbf{1} \}.$$

La expresión depende del generador infinitesimal de la cadena modulante Y_s , de los estados de la misma y de su distribución invariante.

Es importante resaltar que Keisidis, Walrang y Chang en la demostración del teorema anterior presentada en [42], no utilizan en ningún momento que la distribución inicial fuera la invariante, esta hipótesis sólo es necesaria para que el proceso X_t posea incrementos estacionarios y por lo tanto tenga sentido calcular el ancho de banda efectivo.

En las fuentes de audio, lo más habitual es que la señal transmitida sea de voz y para modelarla, se aprovechan las características del habla humana, que se presenta a ráfagas (*talk spurt*) con silencios intercalados entre palabras y entre frases, por eso es común que los codificadores de voz incorporen detectores de silencios, durante los cuales no transmiten información.

Dentro de los modelos de Flujo Markoviano, los que más se aproximan al fenómeno de *talk spurt* son los que se basan en una cadena de Markov de dos estados (*on* y *off*).

Estos modelos describen una fuente que emite información a ráfagas, de manera que en el estado *on* se generan paquetes de voz y en el estado *off* hay silencio.

Veamos ahora como calculamos explícitamente el ancho de banda efectivo para dicho modelo.

Ejemplo 1.24 Procesos On-Off

Supongamos que el proceso de llegada del trabajo de la fuente es un flujo Markoviano modulado por una cadena de Markov con dos estados 0 (*off*) y h (*on*), mientras la cadena está en el estado h se produce trabajo a tasa constante h , mientras está en el estado 0 no se produce trabajo, y la matriz que da la información de las tasas de transferencias es

$$\mathbb{H} = \begin{pmatrix} 0 & 0 \\ 0 & h \end{pmatrix}.$$

Supongamos que el generador infinitesimal de la cadena es

$$\mathbb{Q} = \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix},$$

donde λ se interpreta como la cantidad de transiciones del estado 0 al estado h por unidad de tiempo y μ se interpreta como la cantidad de transiciones al estado 0 por unidad de tiempo.

La distribución invariante de la cadena es un vector de probabilidad π tal que $\pi\mathbb{Q} = \mathbf{0}$, entonces

$$\pi = \left(\frac{\mu}{\lambda + \mu}, \frac{\lambda}{\lambda + \mu} \right).$$

Tomando como distribución inicial de la cadena la distribución invariante, el ancho de banda efectivo resulta ser

$$\alpha(s, t) = \frac{1}{st} \log \left\{ \left(\frac{\mu}{\lambda + \mu}, \frac{\lambda}{\lambda + \mu} \right) \exp \left[\begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu + hs \end{pmatrix} t \right] \mathbf{1} \right\}.$$

Para estimar la fórmula del ancho de banda efectivo para este modelo dada por el teorema anterior, es suficiente estimar el generador infinitesimal \mathbb{Q} de la cadena modulante y su distribución invariante π , ya que los valores de las velocidades de transmisión h_i en cada estado, son determinísticas y conocidas.

Perera *et al.* en [55], encontraron un estimador para el ancho de banda efectivo usando este modelo de tráfico y también una expresión para su intervalo de confianza.

Capítulo 2

Una Generalización del Modelo de Flujo Markoviano

2.1. Introducción

En este capítulo se propone y estudia en detalle un nuevo modelo para describir el comportamiento de las fuentes. En el modelo de flujo modulado por una cadena de Markov, el estado de la cadena determina la tasa del flujo. En consecuencia, la tasa continua es cuantizada a un número finito de niveles equivalente al número de estados de la cadena de Markov y los cambios de la velocidad de transferencia dan cuenta de un cambio de estado.

En un tráfico real de datos, voz o video, la gran cantidad de valores diferentes que se observan para las tasas se traslada en este modelo en la cantidad de estados que asume la cadena modulante y por consiguiente aumenta la dimensión de la matriz generador infinitesimal, haciendo el problema del modelado casi inmanejable.

Esta inconveniencia del modelo nos ha llevado a proponer otro alternativo que hemos denominado Flujo Markoviano Generalizado con tasas de transferencia aleatorias, donde los cambios bruscos de la velocidad de transferencia reportan un cambio de estado en la cadena, pero dentro de un estado se permite que la tasa asuma aleatoriamente cualquier valor de acuerdo a alguna distribución de probabilidades.

Este nuevo planteo puede interpretarse de la siguiente manera: el estado de la cadena indicaría un tipo de actividad en la transferencia como por ejemplo correo, chat, conversación, video-conferencia, entre otros y para cada estado la tasa se sortea, según una ley de probabilidad, dentro de un rango de valores razonable para ese tipo de actividad.

Una vez definido el modelo, nuestro interés se centró en primer lugar en calcular el ancho de banda efectivo para el mismo. Obtuvimos una fórmula explícita del tipo Kesidis, Walrand y Chang para el ancho de banda efectivo. Luego, proponemos un estimador del ancho de banda efectivo a partir de trasa de tráfico, demostramos su consistencia y convergencia débil con el propósito de obtener un intervalo de confianza.

Este capítulo está organizado de la siguiente manera, en la Sección 2.2 se introduce el modelo de Flujo Markoviano Generalizado, en la Sección 2.3 se desarrollan fórmulas

de tipo Kesidis, Walrand y Chang para el modelo de Flujo Markoviano Generalizado propuesto, diferenciando el caso discreto del caso continuo. Para el nuevo modelo, en la Sección 2.4 se presentan los conceptos fundamentales sobre los estimadores, paramétricos y no paramétricos, del ancho de banda efectivo para luego introducir en la Sección 2.5 un estimador gaussiano del ancho de banda efectivo a partir de traza de tráfico. En la Sección 2.6 se demuestra su consistencia y por último en la Sección 2.7 se presenta un intervalo de confianza para la estimación del ancho de banda efectivo.

2.2. Flujos Markovianos con tasas de transferencias aleatorias

El modelo que presentamos en esta sección es una generalización del flujo markoviano. En este caso la velocidad de transferencia desde una fuente es una variable aleatoria cuyo rango y distribución de probabilidades están determinados por el estado de la cadena modulante.

Más precisamente, supondremos que una fuente en una red de datos asume el estado Z_s en el instante s , donde Z_s es una cadena de Markov a tiempo continuo, homogénea e irreducible, con espacio de estados $\mathcal{K} = \{1, \dots, k\}$ y distribución inicial igual a la distribución invariante π . Sean además f_1, f_2, \dots, f_k , k leyes de probabilidad con soportes disjuntos y conocidos. Cuando la cadena Z_s alcanza en el instante s el estado i , la velocidad Y_s con la cual la fuente transfiere datos, ya no toma un valor fijo para ese estado sino que se sortea independientemente de la cadena Z_s , según la ley f_i , es decir la variable aleatoria $Y_s | Z_s = i$ se distribuye según la ley de probabilidad conocida f_i para $i = 1, \dots, k$.

El proceso Y_s toma el valor sorteado todo el tiempo que la cadena Z_s permanece en ese estado, si la cadena modulante cambia de estado, se sortea de nuevo un valor.

En otras palabras si notamos con $\tau_i^{(j)}$ el instante en que Z_s alcanza por j -ésima vez el estado i , y con $\sigma_i^{(j)}$ el instante en que Z_s abandona por j -ésima vez el estado i , con $\sigma_i^{(j)} > \tau_i^{(j)}$, resulta que $Z_s = i$ durante $[\tau_i^{(j)}, \sigma_i^{(j)})$ correspondiente a la j -ésima visita de Z_s al estado i y entonces Y_s toma el valor sorteado según la ley f_i para todo $s \in [\tau_i^{(j)}, \sigma_i^{(j)})$.

Si interpretamos el estado de la cadena modulante como la actividad que realiza un usuario, la velocidad de transferencia asumen rangos distintos de acuerdo al tipo de actividad. Por ejemplo, si el usuario envía un correo electrónico utilizará una tasa de transferencia sensiblemente menor que la que requiere un usuario que interviene en una videoconferencia, por lo tanto parece razonable suponer que el rango de velocidad Y_s , además de depender del estado s en que se encuentra la cadena modulante, sea distinto para cada estado.

Observación 2.1

El proceso Y_s es observable y, dado que los soportes de las distribuciones f_i son disjuntos y conocidos, también resulta observable el proceso Z_s .

El flujo markoviano modulado por la cadena Z_s y que representa el trabajo acumulado en el intervalo $[0, t]$ recibido desde la fuente que despacha información a velocidad Y_s resulta nuevamente

$$X_t = \int_0^t Y_s ds.$$

De aquí en más los resultados que obtendremos estarán referidos a este modelo al que denominaremos *Flujo Markoviano Generalizado*.

2.3. Cálculo del ancho de banda efectivo para el flujo Markoviano generalizado

Nuestro primer objetivo es encontrar fórmulas del tipo de la obtenida por Kesidis, Walrand y Chang [42] para calcular el ancho de banda efectivo asociado al modelo de Flujo Markoviano Generalizado introducido en la sección anterior.

Vamos a considerar separadamente dos casos, uno discreto y el otro continuo. En el caso discreto, la distribución de probabilidades f_i de Y_s cuando la cadena Z_s entra en el estado i está concentrada en un conjunto finito de valores $\{h_i^{(1)}, \dots, h_i^{(L_i)}\}$, mientras que en el caso continuo cada f_i es una función de densidad con soporte $[c_i, c_{i+1}) \subset \mathbb{R}^+$ para $c_i < c_{i+1}$. Tanto los conjuntos finitos como los intervalos de soporte son disjuntos y conocidos para cada estado i .

2.3.1. Caso discreto

Dada Z_s una cadena de Markov irreducible, homogénea, con espacio de estados $\mathcal{K} = \{1, 2, \dots, k\}$ y distribución invariante π , comenzamos considerando el caso en que Y_s se sortea según la ley f_i concentrada en los valores $\mathcal{L}_i = \{h_i^{(1)}, \dots, h_i^{(L_i)}\}$, tomando con probabilidad positiva, uno de estos valores todo el tiempo que la cadena Z_s esté en el estado i . Más precisamente, $f_i(u) = P(Y_s = u / Z_s = i)$, con primer momento μ_i y varianza σ_i^2 .

Llamamos $\mathcal{L} = \{h_i^{(j)} : 1 \leq i \leq k, 1 \leq j \leq L_i\}$, al conjunto de los posibles valores que toma Y_s y definimos la matriz diagonal \mathbb{H} como una matriz de dimensión k , cuyos elementos no nulos son los primeros momentos de cada distribución.

El siguiente teorema proporciona una fórmula para el cálculo del ancho de banda efectivo de un proceso con estas características.

Teorema 2.2 *Sea $\{X_t\}_{t \geq 0}$ un Flujo Markoviano Generalizado modulado por la cadena de Markov irreducible homogénea Z_t con distribución invariante π y generador infinitesimal \mathbb{Q}^Z . Sean las variables aleatorias Y_t y la matriz diagonal \mathbb{H} , de dimensión k , definidas como antes, entonces*

$$\alpha(s, t) = \frac{1}{st} \log \{ \pi \exp [(\mathbb{Q}^Z + s\mathbb{H}) t] \mathbf{1} \},$$

donde $\mathbf{1}$ es un vector columna con todas las entradas iguales a 1.

Demostración: Por la definición de ancho de banda efectivo de una fuente estacionaria presentada en (1.1) basta probar que

$$E(e^{sX_t}) = \pi \exp[(\mathbb{Q}^Z + s\mathbb{H})t] \mathbf{1}.$$

Como $X_t = \int_0^t Y_s ds$, escribiendo la integral como límite de sumas de Riemann tenemos que

$$E(e^{sX_t}) = E\left(e^{s \lim_{n \rightarrow \infty} \sum_{r=1}^n Y_{rt} \frac{t}{n}}\right) = E\left(\lim_{n \rightarrow \infty} e^{s \sum_{r=1}^n Y_{rt} \frac{t}{n}}\right). \quad (2.1)$$

Dado que $|Y_s| \leq \max_{\substack{1 \leq i \leq k \\ 1 \leq j \leq L_i}} h_i^{(j)} \leq \infty$, tenemos que $|X_t| \leq \int_0^t |Y_s| ds \leq t \max_{\substack{1 \leq i \leq k \\ 1 \leq j \leq L_i}} h_i^{(j)} \leq \infty$, y por lo tanto $e^{sX_t} < \infty$.

Aplicando el teorema de convergencia dominada, [35], resulta

$$E(e^{sX_t}) = \lim_{n \rightarrow \infty} E\left(e^{s \frac{t}{n} \sum_{r=1}^n Y_{rt} \frac{t}{n}}\right).$$

La variable $e^{s \frac{t}{n} \sum_{r=1}^n Y_{rt} \frac{t}{n}}$ es discreta por lo tanto

$$E\left(e^{s \frac{t}{n} \sum_{r=1}^n Y_{rt} \frac{t}{n}}\right) = \sum_{(u_1, \dots, u_n) \in \mathcal{L}^n} e^{s \frac{t}{n} \sum_{r=1}^n u_r} P(Y_{\frac{t}{n}} = u_1, \dots, Y_{\frac{tn}{n}} = u_n).$$

Como

$$\begin{aligned} P(Y_{\frac{t}{n}} = u_1, \dots, Y_{\frac{tn}{n}} = u_n) &= \sum_{(i_1, \dots, i_n) \in \mathcal{K}^n} P(Y_{\frac{t}{n}} = u_1, \dots, Y_{\frac{tn}{n}} = u_n / Z_{\frac{t}{n}} = i_1, \dots, Z_{\frac{tn}{n}} = i_n) \\ &\quad \cdot P(Z_{\frac{t}{n}} = i_1, \dots, Z_{\frac{tn}{n}} = i_n) \\ &= \sum_{(i_1, \dots, i_n) \in \mathcal{K}^n} P(Y_{\frac{t}{n}} = u_1 / Z_{\frac{t}{n}} = i_1) \cdots P(Y_{\frac{tn}{n}} = u_n / Z_{\frac{tn}{n}} = i_n) \\ &\quad \cdot P(Z_{\frac{t}{n}} = i_1, \dots, Z_{\frac{tn}{n}} = i_n) \\ &= \sum_{(i_1, \dots, i_n) \in \mathcal{K}^n} \prod_{j=1}^n f_{i_j}(u_j) P(Z_{\frac{t}{n}} = i_1, \dots, Z_{\frac{tn}{n}} = i_n), \end{aligned}$$

y Z_s es una cadena de Markov homogénea con distribución invariante π , tenemos que

$$\begin{aligned}
E \left(e^{s \frac{t}{n} \sum_{r=1}^n Y_{rt}} \right) &= \sum_{(u_1, \dots, u_n) \in \mathcal{L}^n} \prod_{r=1}^n e^{s \frac{t}{n} u_r} \sum_{(i_0, i_1, \dots, i_n) \in \mathcal{K}^{n+1}} \pi(i_0) \prod_{j=0}^{n-1} P_{\frac{t}{n}}^Z(i_j, i_{j+1}) f_{i_{j+1}}(u_{j+1}) \\
&= \sum_{(i_0, i_1, \dots, i_n) \in \mathcal{K}^{n+1}} \pi(i_0) \prod_{j=0}^{n-1} \sum_{(u_1, \dots, u_n) \in \mathcal{L}^n} e^{\frac{st}{n} u_{j+1}} f_{i_{j+1}}(u_{j+1}) P_{\frac{t}{n}}^Z(i_j, i_{j+1}).
\end{aligned}$$

Cada sumatoria en la última expresión representa la función generadora de momentos de la variable con distribución $f_{i_{j+1}}$ que la notaremos

$$\phi_{i_{j+1}} \left(\frac{st}{n} \right) = \sum_{(u_1, \dots, u_n) \in \mathcal{L}^n} e^{\frac{st}{n} u_{j+1}} f_{i_{j+1}}(u_{j+1}),$$

con lo que tenemos que

$$E \left(e^{s \frac{t}{n} \sum_{r=1}^n Y_{rt}} \right) = \sum_{(i_0, i_1, \dots, i_n) \in \mathcal{K}^{n+1}} \pi(i_0) \prod_{j=0}^{n-1} a_{\frac{t}{n}}(i_j, i_{j+1}),$$

donde $a_{\frac{t}{n}}(i, j) = \phi_i \left(\frac{st}{n} \right) P_{\frac{t}{n}}^Z(i, j)$ son los elementos de la matriz $A_{\frac{t}{n}}$ de dimensión k .

El término de la derecha se puede expresar como producto de matrices de la siguiente manera

$$\begin{aligned}
\sum_{(i_0, i_1, \dots, i_n) \in \mathcal{K}^{n+1}} \pi(i_0) \prod_{j=0}^{n-1} a_{\frac{t}{n}}(i_j, i_{j+1}) &= \sum_{(i_0, i_n) \in \mathcal{K}^2} \pi(i_0) (A_{\frac{t}{n}}^n)_{i_0, i_n} \\
&= \sum_{i_0 \in \mathcal{K}} \pi(i_0) \sum_{i_n \in \mathcal{K}} (A_{\frac{t}{n}}^n)_{i_0, i_n} \\
&= \pi A_{\frac{t}{n}}^n \mathbf{1},
\end{aligned}$$

donde π es la distribución invariante de la cadena Z y $\mathbf{1}$ es un vector columna con las k entradas iguales a 1.

A su vez, la matriz $A_{\frac{t}{n}}$ puede expresarse como el producto de la matriz $\left[P_{\frac{t}{n}}^Z \right]$ cuyos elementos son $\left(P_{\frac{t}{n}}^Z \right)_{i,j} = P_{\frac{t}{n}}^Z(i, j)$, y la matriz

$$\left[C_{\frac{t}{n}} \right]_{k \times k} = \begin{bmatrix} \phi_1 \left(\frac{st}{n} \right) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \phi_k \left(\frac{st}{n} \right) \end{bmatrix}.$$

Cada una de estas matrices admite los siguientes desarrollos de Taylor

$$P_{\frac{t}{n}}^Z = P_0^Z + (P_t^Z)'_{t=0} \cdot \frac{t}{n} + o\left(\frac{t}{n}\right) = I + \mathbb{Q}^Z \cdot \frac{t}{n} + o\left(\frac{t}{n}\right),$$

$$C_{\frac{t}{n}} = C_0 + (C_t)'_{t=0} \cdot \frac{t}{n} + o\left(\frac{t}{n}\right) = I + s\mathbb{H} \cdot \frac{t}{n} + o\left(\frac{t}{n}\right),$$

donde I es la matriz identidad de dimensión k . Luego,

$$\left[A_{\frac{t}{n}}\right]^n = \left[I + (\mathbb{Q}^Z + s\mathbb{H}) \cdot \frac{t}{n} + o\left(\frac{t}{n}\right)\right]^n,$$

y puesto que

$$\left[I + (\mathbb{Q}^Z + s\mathbb{H}) \frac{t}{n} + o\left(\frac{t}{n}\right)\right]^n \xrightarrow{n \rightarrow \infty} \exp [(\mathbb{Q}^Z + s\mathbb{H}) t],$$

se tiene que

$$E \left(e^{s \frac{t}{n} \sum_{r=1}^n Y_{rt}^r} \right) = \pi A_{\frac{t}{n}}^n \mathbf{1} \xrightarrow{n \rightarrow \infty} \pi \exp [(\mathbb{Q}^Z + s\mathbb{H}) t] \mathbf{1},$$

lo que concluye el teorema. \square

2.3.2. Caso continuo

Consideremos ahora el caso en que la tasa de transferencia Y_s es una variable aleatoria continua con función de densidad de probabilidades f_i , todo el tiempo que la cadena Z_s esté en el estado i . Supongamos además que tiene soporte $[c_i, c_{i+1}) \subset \mathbb{R}^+$, $c_i < c_{i+1}$, disjuntos y conocidos, primer momento μ_i , varianza σ_i^2 y transformada de Laplace $\phi_i(t)$. Recordemos que la cadena Z tiene como espacio de estados el conjunto $\mathcal{K} = \{1, 2, \dots, k\}$ y definimos la matriz diagonal \mathbb{H} de dimensión k como una matriz cuyos elementos no nulos son $\mu_1, \mu_2, \dots, \mu_k$, los primeros momentos de cada distribución f_i . Un resultado similar al anterior, para el cálculo del ancho de banda efectivo, se obtiene a partir del siguiente teorema, que si bien tiene algunas similitudes con el **Teorema 2.2**, las herramientas que se utilizan en su demotración son distintas.

Teorema 2.3 *Sea $\{X_t\}_{t \geq 0}$ un Flujo Markoviano Generalizado modulado por la cadena de Markov irreducible homogénea Z_t con distribución invariante π y generador infinitesimal \mathbb{Q}^Z . Sean las variables aleatorias Y_t y la matriz diagonal \mathbb{H} , de dimensión k , definidas como antes, entonces*

$$\alpha(s, t) = \frac{1}{st} \log \{ \pi \exp [(\mathbb{Q}^Z + s\mathbb{H}) t] \mathbf{1} \}, \quad (2.2)$$

donde $\mathbf{1}$ es un vector columna con todas las entradas iguales a 1.

Demostración:

Por la definición de ancho de banda efectivo de una fuente estacionaria presentada en (1.1) basta probar que

$$E(e^{sX_t}) = \pi \exp[(\mathbb{Q}^Z + s\mathbb{H})t] \mathbf{1}.$$

Dado que $X_t = \int_0^t Y_s ds$ y escribiendo la integral como límite de sumas de Riemann, tenemos que

$$E(e^{sX_t}) = E\left(e^{s \lim_{n \rightarrow \infty} \sum_{r=1}^n Y_{rt} \frac{t}{n}}\right) = E\left(\lim_{n \rightarrow \infty} e^{s \sum_{r=1}^n Y_{rt} \frac{t}{n}}\right).$$

Puesto que $|Y_s| \leq \max_{1 \leq i \leq k} c_{i+1} \leq \infty$, tenemos que $|X_t| \leq \int_0^t |Y_s| ds \leq t \max_{1 \leq i \leq k} c_{i+1} \leq \infty$, y por lo tanto $e^{sX_t} < \infty$.

Aplicando el teorema de convergencia dominada, [35], resulta

$$E(e^{sX_t}) = \lim_{n \rightarrow \infty} E\left(e^{s \frac{t}{n} \sum_{r=1}^n Y_{rt}}\right).$$

Como la variable $e^{s \frac{t}{n} \sum_{r=1}^n Y_{rt}}$ es continua y además Z_s es una Cadena de Markov Homogénea con distribución invariante π , tenemos que

$$\begin{aligned} & E\left(e^{s \frac{t}{n} \sum_{r=1}^n Y_{rt}}\right) \\ &= \sum_{(i_1, \dots, i_n) \in \mathcal{K}^n} \int_{(\mathbb{R}^+)^n} e^{s \frac{t}{n} \sum_{r=1}^n u_r} P(Z_{\frac{t}{n}} = i_1, \dots, Z_{tn} = i_n) f_{i_1}(u_1) du_1 \dots f_{i_n}(u_n) du_n \\ &= \sum_{(i_0, \dots, i_n) \in \mathcal{K}^{n+1}} \int_{(\mathbb{R}^+)^n} \prod_{r=1}^n e^{\frac{st}{n} u_r} \pi(i_0) \prod_{j=0}^{n-1} \left(P_{\frac{t}{n}}^Z(i_j, i_{j+1}) f_{i_{j+1}}(u_{j+1}) du_{j+1} \right) \\ &= \sum_{(i_0, \dots, i_n) \in \mathcal{K}^{n+1}} \pi(i_0) \left[\prod_{j=0}^{n-1} \left(P_{\frac{t}{n}}^Z(i_j, i_{j+1}) \int_{(\mathbb{R}^+)} e^{\frac{st}{n} u_{j+1}} f_{i_{j+1}}(u_{j+1}) du_{j+1} \right) \right]. \end{aligned}$$

Cada integral en la última expresión representa la función generadora de momento de la distribución que la notaremos $\phi_{i_j}(\frac{st}{n}) = \int_{(\mathbb{R}^+)} e^{\frac{st}{n}u_j} f_{i_j}(u_j) du_j$ y ya que f_{i_0} es una función de densidad, tenemos que

$$E \left(e^{\sum_{r=1}^n Y_{rt}} \right) = \sum_{(i_0, \dots, i_n) \in \mathcal{K}^{n+1}} \pi(i_0) \int_{(\mathbb{R}^+)} \prod_{j=0}^{n-1} a_{\frac{t}{n}}(i_j, i_{j+1}),$$

donde $a_{\frac{t}{n}}(i_j, i_{j+1}) = \phi_{i_{j+1}}(\frac{st}{n}) P_{\frac{t}{n}}^Z(i_j, i_{j+1})$ son los elementos de la matriz cuadrada $A_{\frac{t}{n}}$. El término de la derecha se puede expresar como producto de matrices de la siguiente manera

$$\begin{aligned} \sum_{(i_0, \dots, i_n) \in \mathcal{K}^{n+1}} \pi(i_0) \int_{(\mathbb{R}^+)} \prod_{j=0}^{n-1} a_{\frac{t}{n}}(i_j, i_{j+1}) &= \sum_{(i_0, i_n) \in \mathcal{K}^2} \pi(i_0) (A_{\frac{t}{n}}^n)_{i_0, i_n} \\ &= \sum_{i_0 \in \mathcal{K}} \pi(i_0) \sum_{i_n \in \mathcal{K}} (A_{\frac{t}{n}}^n)_{i_0, i_n} \\ &= \pi A_{\frac{t}{n}}^n \mathbf{1}, \end{aligned}$$

donde π es la distribución invariante de la cadena Z y $\mathbf{1}$ es un vector columna con las k entradas iguales a 1.

A su vez, la matriz $A_{\frac{t}{n}}$ puede expresarse como el producto de la matriz $\left[P_{\frac{t}{n}}^Z \right]$ cuyos elementos son $\left(P_{\frac{t}{n}}^Z \right)_{i,j} = P_{\frac{t}{n}}^Z(i, j)$, y la matriz

$$\left[C_{\frac{t}{n}} \right]_{k \times k} = \begin{bmatrix} \phi_1(\frac{st}{n}) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \phi_k(\frac{st}{n}) \end{bmatrix}.$$

Cada una de estas matrices admiten los siguientes desarrollos de Taylor

$$P_{\frac{t}{n}}^Z = P_0^Z + (P_t^Z)'_{t=0} \cdot \frac{t}{n} + o\left(\frac{t}{n}\right) = I + \mathbb{Q}^Z \cdot \frac{t}{n} + o\left(\frac{t}{n}\right),$$

$$C_{\frac{t}{n}} = C_0 + (C_t)'_{t=0} \cdot \frac{t}{n} + o\left(\frac{t}{n}\right) = I + s\mathbb{H} \cdot \frac{t}{n} + o\left(\frac{t}{n}\right),$$

donde I es la matriz identidad de dimensión k , y la matriz \mathbb{H} contiene en su diagonal las derivadas $\phi_i'(0) = \mu_i$. Luego,

$$\left[A_{\frac{t}{n}} \right]^n = \left[I + (\mathbb{Q}^Z + s\mathbb{H}) \cdot \frac{t}{n} + o\left(\frac{t}{n}\right) \right]^n,$$

y puesto que

$$\left[I + (\mathbb{Q}^Z + s\mathbb{H}) \cdot \frac{t}{n} + o\left(\frac{t}{n}\right) \right]^n \xrightarrow{n \rightarrow \infty} \exp [(\mathbb{Q}^Z + s\mathbb{H}) t],$$

se tiene que

$$E \left(e^{s \frac{t}{n} \sum_{r=1}^n Y_{rt}} \right) = \pi A_{\frac{t}{n}}^n \mathbf{1} \xrightarrow{n \rightarrow \infty} \pi \exp [(\mathbb{Q}^Z + s\mathbb{H}) t] \mathbf{1},$$

lo que concluye el teorema. \square

En las últimas dos secciones se han obtenido fórmulas similares a la obtenida por Kesidis, Walrand y Chang para calcular el ancho de banda efectivo del modelo de Flujo Markoviano Generalizado con tasas de transferencias aleatorias, discretas o continuas.

De forma análoga al caso del flujo markoviano clásico, las fórmulas dependen del generador infinitesimal de la cadena modulante, de su distribución invariante y de una matriz que contiene la información de las tasas de transferencia.

En el caso del Flujo Markoviano clásico la información que contendrá la matriz serán los valores de las tasas de transferencia, ya que esos valores son fijos y conocidos, mientras que para el caso del nuevo modelo, el Flujo Markoviano Generalizado, donde las tasas de transferencias son aleatorias, la información que nos dará la matriz será la tasa media de transferencia con la que se despacha en cada estado.

Los resultados obtenidos en esta sección fueron publicados en [49].

2.4. Estimación del ancho de banda efectivo

En cuanto a la estimación del ancho de banda efectivo hay dos enfoques, el paramétrico y el no paramétrico. En el enfoque paramétrico se asume un modelo para el tráfico de la fuente y a partir de las trazas se estima un conjunto de parámetros de dicho modelo que intervienen en la expresión del ancho de banda efectivo, mientras que en el enfoque no paramétrico no se asume un modelo específico del tráfico y se procura construir un estimador del ancho de banda efectivo calculando el valor esperado que aparece en la función generatriz de momentos a través de promedios temporales en la traza.

Este último enfoque si bien es más general, tiene por desventaja que al no asumir un modelo de tráfico, no se tendrá una expresión analítica del mismo que pueda ser usada para cálculos posteriores, además de ser difícil encontrar un teorema central del límite que permita construir un intervalo de confianza del estimador lo cual es posible en el caso paramétrico. Trataremos primero el caso no paramétrico y luego el paramétrico.

2.4.1. Estimadores no paramétricos

Estimar el ancho de banda efectivo cuando no se asume un modelos de tráfico requiere de un estimador puntual que se pueda calcular sobre cualquier traza a analizar. El más

comúnmente utilizado es el estimador, originariamente propuesto por Amir Dembo en [28] y aplicado en los trabajos de Courcoubetis y Rabinovich [25, 64].

Para su implementación primero se divide la traza en bloques de largo t para la cual la suma sobre los bloques

$$\bar{X}_k = \sum_{i=(k-1)t}^{kt} x(i) \quad 0 \leq k \leq [T/t], \quad (2.3)$$

donde $x(i)$ es la cantidad de trabajo que arriba por intervalo de tiempo, son aproximadamente independientes e idénticamente distribuidas. El ancho de banda efectivo puede estimarse entonces usando el promedio temporal

$$\alpha_n(s, t) = \frac{1}{st} \log \left[\frac{1}{[T/t]} \sum_{j=1}^{[T/t]} e^{s\bar{X}_j} \right]. \quad (2.4)$$

Es claro en este caso que se tendrá una buena estimación cuando los valores de t verifiquen que $t \ll T$. Es decir, cuando el número de muestras del proceso de incrementos dentro de las traza sea suficientemente grande.

2.4.2. Estimadores paramétricos

El estimador paramétrico del ancho de banda efectivo más simple asume que las fuentes siguen una distribución de Poisson y estima el parámetro λ de dicha distribución a partir de la media temporal de la traza. Si bien este modelo tiene la virtud de la simplicidad, es de poca utilidad práctica ya que en general el tráfico no corresponde a un proceso poissoniano.

Los casos paramétricos de mayor interés son los llamados modelos de tráfico markovianos, tanto para tazas de transferencias fijas como aleatorias, presentados en las secciones precedentes.

Recordemos que en ambos casos el proceso es gobernado por una cadena de Markov de tiempo continuo, en estos modelos de tráfico se supone que en cada estado de la cadena la fuente transmite a una velocidad (bit/s) fija y conocida, para el modelo tradicional, o aleatoria, para el modelo que presentamos en la **Sección 2.2**. Cuando la cadena cambia de un estado a otro, cambia la velocidad de transmisión de la fuente.

Para el caso de Flujo Markoviano se presenta en [55] un estimador consistente, para el Flujo Markoviano Generalizado, el ancho de banda efectivo fue presentado en la **Sección 2.3** y en la próxima sección se presenta un estimador a partir de las trayectorias o trazas de tráfico.

Al tratar de aplicar estos estimadores en la práctica sobre trazas de tráfico reales, surgen una serie de preguntas como ¿cuáles son los estados? ¿cómo determinar a que estado pertenece cada punto de la traza? Y sobre todo, para el caso de tazas aleatorias, ¿cómo manejar el hecho que no se sabe en que estado se encuentra la cadena modulante, si sólo se ve el valor de la taza de despacho?

Nosotros proponemos un modelo donde en cada estado de la cadena modulante, al sortear la taza de despacho se hace de conjuntos disjuntos y conocidos. Es decir, las distribuciones que se puede elegir en un estado determinado de la cadena modulante, no se puede elegir cuando la cadena modulante cambia de estado.

Veamos el estimador que presentamos y cuales son sus propiedades.

2.5. Estimación asintóticamente gaussiana para $\alpha(s, t)$

En esta sección presentamos un estimador consistente del ancho de banda efectivo, dado por la ecuación (2.2), a partir de trayectorias de tráfico. Este estimador deberá contener por una parte un estimador del generador infinitesimal \mathbb{Q} , de la cadena modulante y por otro lado un estimador de las tasas medias de despacho que son los elementos de la diagonal de la matriz \mathbb{H} .

A partir del siguiente teorema, cuya demostración se puede ver en [45], se obtiene una estimación gaussiana asintótica de los elementos de \mathbb{Q} en función de las trayectorias o trazas de tráfico.

Teorema 2.4 (Lebedev-Lukashuk) *Sea $(X_t)_{t \in \mathbb{R}^+}$ una cadena de Markov homegénea irreducible con espacio de estados finito $S = \{1, \dots, k\}$ y generador infinitesimal $\mathbb{Q} = (q_{ij})_{i,j=1,\dots,k}$ desconocido y sea $D = \{(i, j) \in S \times S / q_{ij} > 0\}$.*

El estimador de máxima verosimilitud de q_{ij} , dado por

$$q_{ij}^{(n)}(x) = \frac{\gamma(i, j, nx)}{\tau(i, nx)}, \quad (2.5)$$

donde $\gamma(i, j, h) :=$ número de transiciones de la cadena de i a j en el intervalo $[0, h]$ y $\tau(i, h) :=$ tiempo que permaneció la cadena en el estado i durante el intervalo $[0, h]$, cumple que

1. $\lim_{n \rightarrow +\infty} |q_{ij}^{(n)}(u) - q_{ij}| = 0$ c.s.
2. $\left(\sqrt{nu} (q_{ij}^{(n)}(u) - q_{ij}) \right)_{(i,j) \in D} \xrightarrow[n \rightarrow \infty]{w} \left(\sqrt{\frac{q_{ij}}{\pi(i)}} W_{ij}(u) \right)_{(i,j) \in D}$ como proceso estocástico en $u \in [0, 1]$, donde $W = \{W_{ij}\}_{(i,j) \in D}$ denota un proceso de Wiener standard.

Corolario 2.5 *Si $u = 1$ se tiene que la sucesión de variables aleatorias $(q_{ij}^{(n)})_{(i,j) \in D}$ cumple que*

$$\sqrt{n} \left(q_{ij}^{(n)} - q_{ij} \right)_{(i,j) \in D} \xrightarrow[n \rightarrow \infty]{w} N(\mathbf{0}, \Sigma),$$

$$\text{donde } \Sigma = \begin{bmatrix} \frac{q_{11}}{\pi(1)} & \dots & \dots \\ \dots & \ddots & \dots \\ \dots & \dots & \frac{q_{k(k-1)}}{\pi(k)} \end{bmatrix} \in \mathcal{M}_{k(k-1) \times k(k-1)}.$$

Veamos ahora como encontrar, también a partir de las trayectorias de tráfico, un estadístico que nos permita estimar la tasa media de despacho μ_i en cada estado de la cadena modulante.

Notaremos con $Y_i^{(r)}$ la r -ésima velocidad observada que corresponde al rango de Y_t cuando la cadena modulante está en el estado i . Recordemos que como los rangos son disjuntos y conocidos, el estado i queda inequívocamente individualizado a partir de la observación $y_i^{(r)}$.

El valor que puede asumir la variable aleatoria $Y_i^{(r)}$ será alguno de los elementos del conjunto $\mathcal{L}_i = \{h_i^{(1)}, \dots, h_i^{(L_i)}\}$ en los cuales se concentra la función de probabilidad en el caso discreto, o pertenecerá al conjunto $[c_i, c_{i+1}]$ que es el soporte de la función de densidad f_i en el caso continuo, además $E(Y_i^{(r)}) = \mu_i$ y $V(Y_i^{(r)}) = \sigma_i$ puesto que la cadena está en el estado i .

Sea $N_i(n)$ la cantidad de veces que la cadena modulante Z_t entró en el estado i en el intervalo $[0, n]$.

Proponemos entonces como estimador de μ_i a

$$\mu_i^{(n)} = \frac{\sum_{r=1}^{N_i(n)} Y_i^{(r)}}{N_i(n)}, \quad (2.6)$$

que representa el promedio de los valores de las velocidades de despacho en el intervalo $[0, n]$.

Observemos que el número de términos sumados es una cantidad aleatoria $N_i(n)$ que claramente aumenta cuando la longitud del intervalo observado n , aumenta. Como nuestra cadena modulante es finita, irreducible y aperiódica, todos los estados son recurrentes positivos y notamos con $1/\lambda_i$ al tiempo medio de recurrencia de la cadena al estado i . Entonces en el intervalo $[0, n]$, de longitud n , el proceso hará en promedio $n/(1/\lambda_i)$, o equivalentemente $n\lambda_i$ visitas al estado i , por lo tanto cuando n crece el número de visitas al estado i satisface la siguiente relación

$$\frac{N_i(n)}{n} \xrightarrow{c.s.} \lambda_i. \quad (2.7)$$

Analicemos que propiedades tiene el estadístico propuesto. En primer lugar veamos que es un estimador insesgado de la tasa media de despacho, para ello calculemos su valor esperado

$$\begin{aligned}
E(\mu_i^{(n)}) &= E\left(E(\mu_i^{(n)}/N_i(n))\right) \\
&= E\left(E\left(\frac{\sum_{j=1}^{N_i(n)} Y_i^j}{N_i(n)} / N_i(n)\right)\right) \\
&= E\left(\frac{\sum_{j \in \mathbb{N}} E\left(Y_i^{(j)}/N_i(n)\right)}{N_i(n)}\right) \\
&= E(\mu_i/N_i(n)) \\
&= \mu_i.
\end{aligned} \tag{2.8}$$

Calculemos ahora la varianza de este estimador, para ello calculemos el segundo momento

$$\begin{aligned}
E\left((\mu_i^{(n)})^2\right) &= E\left(E(\mu_i^{(n)})^2/N_i(n)\right) \\
&= E\left(E\left(\frac{\sum_{j=1}^{N_i(n)} \sum_{k=1}^{N_i(n)} Y_i^{(j)} Y_i^{(k)}}{N_i^2(n)} / N_i(n)\right)\right) \\
&= E\left(\frac{\sum_{j=1}^{N_i(n)} \sum_{k=1}^{N_i(n)} E\left(Y_i^{(j)} Y_i^{(k)} / N_i(n)\right)}{N_i^2(n)}\right).
\end{aligned} \tag{2.9}$$

Como

$$E\left(Y_i^{(k)} Y_i^{(j)} / N_i(n)\right) = (\sigma_i^2 + \mu_i^2) \delta_{kj} + \mu_i^2 (1 - \delta_{kj}), \tag{2.10}$$

con

$$\delta_{kj} = \begin{cases} 1 & \text{si } k = j \\ 0 & \text{si } k \neq j \end{cases},$$

reemplazando en (2.9) tenemos que

$$\begin{aligned}
E\left(\frac{\sum_{j=1}^{N_i(n)} \sum_{k=1}^{N_i(n)} E\left(Y_i^{(j)} Y_i^{(k)} / N_i(n)\right)}{N_i^2(n)}\right) &= E\left(\frac{1}{N_i^2(n)} (\mu_i^2 N_i^2(n) + \sigma_i^2 N_i(n))\right) \\
&= E\left(\mu_i^2 + \sigma_i^2 \frac{1}{N_i(n)}\right) \\
&= \mu_i^2 + \sigma_i^2 E\left(\frac{1}{N_i(n)}\right),
\end{aligned}$$

y por lo tanto

$$V(\mu_i^{(n)}) = \sigma_i^2 E\left(\frac{1}{N_i(n)}\right), \tag{2.11}$$

que debido a (2.7) podemos decir que $V(\mu_i^{(n)}) \approx \frac{\sigma_i^2}{\lambda_i n}$, cuando $n \rightarrow \infty$.

Observación 2.6

Las propiedades demostradas aseguran que $\mu_i^{(n)}$ es un estimador consistente de la tasa media de despacho μ_i .

Una cadena irreducible y recurrente positiva tiene una distribución única que, a menos que el conjunto de estados consista en un único estado necesariamente absorbente, está dada por

$$\pi_i = \frac{\lambda_i}{q_i}. \quad (2.12)$$

Esta es una fórmula intuitivamente razonable ya que, como vimos, para un intervalo de tiempo grande $[0, n]$, la cadena realiza aproximadamente $n\lambda_i$ visitas al estado i y el tiempo promedio que pasa por visita en dicho estado es $1/q_i$. Por lo tanto el tiempo total empleado en el estado i durante el intervalo de tiempo debe ser $n\lambda_i/q_i$ y la proporción de tiempo dedicado en el estado i debe ser aproximadamente λ_i/q_i .

Nos restaría ahora ver cual es la distribución asintótica del estadístico, para ello presentamos el siguiente teorema

Teorema 2.7 *Sea X_t un Flujo Markoviano Generalizado, $Y_i^{(r)}$ los valores aleatorios que asume Y cuando la cadena modulante Z está en el estado i y $\mu_i^{(n)}$ definido como en (2.6), entonces*

$$\sqrt{n} \left(\mu_i^{(n)} - \mu_i \right) \xrightarrow[n \rightarrow \infty]{w} N \left(0, \frac{\sigma_i^2}{\lambda_i} \right). \quad (2.13)$$

Para demostrarlo utilizaremos el siguiente teorema, cuya demostración se puede encontrar en [13]. Observemos que si X_n e Y_n son sucesiones de variables aleatorias que tienen dominio común, tiene sentido hablar de la distancia $\rho(X_n, Y_n)$. Si el espacio donde cada una toma valores es separable, entonces $\rho(X_n, Y_n)$ es una variable aleatoria y también tiene sentido hablar de su convergencia en probabilidad.

Teorema 2.8 *Si $X_n \xrightarrow{\mathcal{D}} X$ y $\rho(X_n, Y_n) \xrightarrow{\mathcal{P}} 0$, entonces $Y_n \xrightarrow{\mathcal{D}} X$.*

Aplicaremos este teorema para encontrar la distribución del estadístico que proponemos, viendo que se aproxima a otro cuya distribución es conocida.

Ahora estamos en condiciones de demostrar el Teorema 2.7,

Demostración del Teorema 2.7:

Sabemos por (2.8) que $\mu_i^{(n)}$ es un estimador insesgado.

Dadas las variables $Y_i^{(r)}$, aplicando un Teorema Central del Límite clásico, [35], podemos conocer el comportamiento asintótico de $\frac{\sum_{r=1}^{[\lambda_i n]} Y_i^{(r)}}{\lambda_i n}$ y se cumple que

$$\sqrt{n} \left(\frac{\sum_{r=1}^{[\lambda_i n]} Y_i^{(r)}}{\lambda_i n} - \mu_i \right) \xrightarrow[n \rightarrow \infty]{w} N \left(0, \frac{\sigma_i^2}{\lambda_i} \right).$$

Si probamos entonces que $\left| \frac{\sum_{r=1}^{N_i(n)} Y_i^{(r)}}{N_i(n)} - \frac{\sum_{r=1}^{[\lambda_i n]} Y_i^{(r)}}{\lambda_i n} \right| \xrightarrow{\mathcal{P}} 0$, aplicando el teorema anterior tendríamos la distribución asintótica de $\mu_i^{(n)}$. Para ello basta con probar la convergencia en media cuadrática, que implica la convergencia en probabilidad buscada, es decir basta probar que $E \left(\frac{\sum_{r=1}^{N_i(n)} Y_i^{(r)}}{N_i(n)} - \frac{\sum_{r=1}^{[\lambda_i n]} Y_i^{(r)}}{\lambda_i n} \right)^2 \xrightarrow{n \rightarrow \infty} 0$.

En efecto,

$$\begin{aligned} E \left(\frac{\sum_{r=1}^{N_i(n)} Y_i^{(r)}}{N_i(n)} - \frac{\sum_{r=1}^{[\lambda_i n]} Y_i^{(r)}}{\lambda_i n} \right)^2 &= E \left(\frac{\sum_{r=1}^{N_i(n)} Y_i^{(r)}}{N_i(n)} - \frac{\sum_{r=1}^{N_i(n)} Y_i^{(r)}}{\lambda_i n} + \frac{\sum_{r=1}^{N_i(n)} Y_i^{(r)}}{\lambda_i n} - \frac{\sum_{r=1}^{[\lambda_i n]} Y_i^{(r)}}{\lambda_i n} \right)^2 \\ &\leq 2 \left(E \left(\frac{\sum_{r=1}^{N_i(n)} Y_i^{(r)}}{N_i(n)} - \frac{\sum_{r=1}^{N_i(n)} Y_i^{(r)}}{\lambda_i n} \right)^2 + E \left(\frac{\sum_{r=1}^{N_i(n)} Y_i^{(r)}}{\lambda_i n} - \frac{\sum_{r=1}^{[\lambda_i n]} Y_i^{(r)}}{\lambda_i n} \right)^2 \right) \\ &= 2 \left(E \left(\sum_{r=1}^{N_i(n)} Y_i^{(r)} \left(\frac{1}{N_i(n)} - \frac{1}{\lambda_i n} \right) \right)^2 + E \left(\frac{\sum_{r=1}^{N_i(n)} Y_i^{(r)} - \sum_{r=1}^{[\lambda_i n]} Y_i^{(r)}}{\lambda_i n} \right)^2 \right). \end{aligned}$$

Aplicando esperanza condicional para calcular la esperanza en el primer término de la expresión anterior se tiene que

$$E \left(\sum_{r=1}^{N_i(n)} Y_i^{(r)} \left(\frac{1}{N_i(n)} - \frac{1}{\lambda_i n} \right) \right)^2 = E \left(\left(\frac{1}{N_i(n)} - \frac{1}{\lambda_i n} \right)^2 \sum_{k=1}^{N_i(n)} \sum_{j=1}^{N_i(n)} E \left(Y_i^{(k)} Y_i^{(j)} / N_i(n) \right) \right). \quad (2.14)$$

Reemplazando (2.10) en (2.14) tenemos que

$$\begin{aligned} E \left(\left(\frac{1}{N_i(n)} - \frac{1}{\lambda_i n} \right)^2 \sum_{k=1}^{N_i(n)} \sum_{j=1}^{N_i(n)} E \left(Y_i^{(k)} Y_i^{(j)} / N_i(n) \right) \right) &= \frac{(\lambda_i n - N_i(n))^2}{(\lambda_i n N_i(n))^2} (\sigma_i^2 N_i(n) + \mu_i^2 N_i^2(n)) \\ &= \left(1 - \frac{N_i(n)}{\lambda_i n} \right)^2 \left(\frac{1}{N_i(n)} \sigma_i^2 + \mu_i^2 \right), \end{aligned}$$

que tiende a cero cuando $n \rightarrow \infty$.

Veamos ahora el segundo término

$$E \left(\frac{\sum_{r=1}^{N_i(n)} Y_i^{(r)} - \sum_{r=1}^{[\lambda_i n]} Y_i^{(r)}}{\lambda_i n} \right)^2 = E \left(\frac{\sum_{\min(N_i(n), [\lambda_i n])}^{\max(N_i(n), [\lambda_i n])} Y_i^{(r)}}{\lambda_i n} \right)^2, \quad (2.15)$$

calculando nuevamente su esperanza mediante la esperanza condicional, obtenemos

$$\begin{aligned}
E \left(\frac{\sum_{\min(N_i(n), [\lambda_i n])}^{\max(N_i(n), [\lambda_i n])} Y_i^{(r)}}{\lambda_i n} \right)^2 &= E \left(E \left(\frac{\sum_{\min(N_i(n), [\lambda_i n])}^{\max(N_i(n), [\lambda_i n])} Y_i^{(r)}}{\lambda_i n} \right)^2 / N_i(n) \right) \\
&= E \left(\frac{\sum_{\min(N_i(n), [\lambda_i n])}^{\max(N_i(n), [\lambda_i n])} \sum_{\min(N_i(n), [\lambda_i n])}^{\max(N_i(n), [\lambda_i n])} E \left(Y_i^{(k)} Y_i^{(j)} / N_i(n) \right)}{\lambda_i^2 n^2} \right),
\end{aligned}$$

y aplicando el resultado mostrado en (2.10) tenemos que

$$\frac{\sigma_i^2 (\max(N_i(n), [\lambda_i n]) - \min(N_i(n), [\lambda_i n])) + \mu_i^2 (\max(N_i(n), [\lambda_i n]) - \min(N_i(n), [\lambda_i n]))^2}{\lambda_i^2 n^2}. \quad (2.16)$$

Además, como

$$\max(N_i(n), [\lambda_i n]) - \min(N_i(n), [\lambda_i n]) = |N_i(n) - [\lambda_i n]|, \quad (2.17)$$

la ecuación (2.16) se convierte en

$$\frac{|N_i(n) - [\lambda_i n]|}{\lambda_i^2 n^2} \sigma_i^2 + \frac{(N_i(n) - [\lambda_i n])^2}{\lambda_i^2 n^2} \mu_i^2 = \frac{|N_i(n) - [\lambda_i n]|}{\lambda_i^2 n^2} \sigma_i^2 + \left(1 - \frac{N_i(n)}{\lambda_i n}\right)^2 \mu_i^2,$$

con lo que se puede observar que ambos términos tienden a 0 cuando $n \rightarrow \infty$, y queda así demostrado el teorema. \square

Aplicando un razonamiento análogo podremos encontrar la distribución asintótica de $\hat{\sigma}_i^2 = \frac{\sum_{r=1}^{N_i(n)} (Y_i^{(r)})^2}{N_i(n)} - \left(\frac{\sum_{r=1}^{N_i(n)} Y_i^{(r)}}{N_i(n)} \right)^2$ y demostrar que será entonces un estimador consistente de la varianza σ_i^2 .

Proposición 2.5.1 *El estimador $\hat{\sigma}_i^2 = \frac{\sum_{r=1}^{N_i(n)} (Y_i^{(r)})^2}{N_i(n)} - \left(\frac{\sum_{r=1}^{N_i(n)} Y_i^{(r)}}{N_i(n)} \right)^2$ y la varianza muestral $S_i^2 = \frac{\sum_{r=1}^{[\lambda_i n]} (Y_i^{(r)})^2}{\lambda_i n} - \left(\frac{\sum_{r=1}^{[\lambda_i n]} Y_i^{(r)}}{\lambda_i n} \right)^2$ tienen la misma distribución asintótica.*

Demostración:

Basta con probar la convergencia en media cuadrática, que implica la convergencia buscada, es decir basta probar que

$$E \left(\frac{\sum_{r=1}^{N_i(n)} (Y_i^{(r)})^2}{N_i(n)} - \left(\frac{\sum_{r=1}^{N_i(n)} Y_i^{(r)}}{N_i(n)} \right)^2 - \left(\frac{\sum_{r=1}^{[\lambda_i n]} (Y_i^{(r)})^2}{\lambda_i n} - \left(\frac{\sum_{r=1}^{[\lambda_i n]} Y_i^{(r)}}{\lambda_i n} \right)^2 \right) \right)^2 \xrightarrow{n \rightarrow \infty} 0.$$

En efecto

$$E \left(\frac{\sum_{r=1}^{N_i(n)} (Y_i^{(r)})^2}{N_i(n)} - \left(\frac{\sum_{r=1}^{N_i(n)} Y_i^{(r)}}{N_i(n)} \right)^2 - \left(\frac{\sum_{r=1}^{[\lambda_i n]} (Y_i^{(r)})^2}{\lambda_i n} - \left(\frac{\sum_{r=1}^{[\lambda_i n]} Y_i^{(r)}}{\lambda_i n} \right)^2 \right) \right)^2 \quad (2.18)$$

se puede acotar por

$$2 \left[E \left(\frac{\sum_{r=1}^{N_i(n)} (Y_i^{(r)})^2}{N_i(n)} - \frac{\sum_{r=1}^{[\lambda_i n]} (Y_i^{(r)})^2}{\lambda_i n} \right)^2 + E \left(\left(\frac{\sum_{r=1}^{N_i(n)} Y_i^{(r)}}{N_i(n)} \right)^2 - \left(\frac{\sum_{r=1}^{[\lambda_i n]} Y_i^{(r)}}{\lambda_i n} \right)^2 \right) \right]. \quad (2.19)$$

Trabajamos con cada uno de los términos para acotarlos, en el caso del primer término de (2.19) tenemos que

$$\begin{aligned} E \left(\frac{\sum_{r=1}^{N_i(n)} (Y_i^{(r)})^2}{N_i(n)} - \frac{\sum_{r=1}^{[\lambda_i n]} (Y_i^{(r)})^2}{\lambda_i n} \right)^2 &= E \left(\frac{\sum_{r=1}^{N_i(n)} (Y_i^{(r)})^2}{N_i(n)} - \frac{\sum_{r=1}^{N_i(n)} (Y_i^{(r)})^2}{\lambda_i n} \right. \\ &\quad \left. + \frac{\sum_{r=1}^{N_i(n)} (Y_i^{(r)})^2}{\lambda_i n} - \frac{\sum_{r=1}^{[\lambda_i n]} (Y_i^{(r)})^2}{\lambda_i n} \right)^2 \\ &\leq 2 \left[E \left(\frac{\sum_{r=1}^{N_i(n)} (Y_i^{(r)})^2}{N_i(n)} - \frac{\sum_{r=1}^{N_i(n)} (Y_i^{(r)})^2}{\lambda_i n} \right)^2 \right. \\ &\quad \left. + E \left(\frac{\sum_{r=1}^{N_i(n)} (Y_i^{(r)})^2}{\lambda_i n} - \frac{\sum_{r=1}^{[\lambda_i n]} (Y_i^{(r)})^2}{\lambda_i n} \right)^2 \right] \\ &= 2 \left[E \left(\left(\frac{1}{N_i(n)} - \frac{1}{\lambda_i n} \right)^2 \left(\sum_{r=1}^{N_i(n)} (Y_i^{(r)})^2 \right) \right) \right. \\ &\quad \left. + E \left(\frac{1}{\lambda_i^2 n^2} \left(\sum_{r=\min(N_i(n), [\lambda_i n])}^{\max(N_i(n), [\lambda_i n])} (Y_i^{(r)})^2 \right) \right) \right] \quad (2.20) \end{aligned}$$

Calculando la esperanza mediante esperanza condicional, y teniendo en cuenta las relaciones (2.10) y (2.17), el primer término en (2.20) se puede expresar como

$$\begin{aligned}
E \left(\left(\frac{1}{N_i(n)} - \frac{1}{\lambda_i n} \right)^2 \left(\sum_{r=1}^{N_i(n)} (Y_i^{(r)})^2 \right)^2 \right) &= E \left(\frac{1}{N_i(n)} - \frac{1}{\lambda_i n} \right)^2 \sum_{j=1}^{N_i(n)} \sum_{k=1}^{N_i(n)} E \left((Y_i^{(j)})^2 (Y_i^{(k)})^2 / N_i(n) \right) \\
&= \left(\frac{1}{N_i(n)} - \frac{1}{\lambda_i n} \right)^2 \left(E((Y_i^{(j)})^4) N_i(n) \right. \\
&\quad \left. + (E(Y_i^{(j)})^2)^2 N_i(n) (N_i(n) - 1) \right) \\
&= \left(1 - \frac{N_i(n)}{\lambda_i n} \right)^2 \left(\frac{1}{N_i(n)} E((Y_i^{(j)})^4) \right. \\
&\quad \left. + \left(1 - \frac{1}{N_i(n)} \right) (E(Y_i^{(j)})^2)^2 \right),
\end{aligned}$$

y el segundo término de (2.20) se puede expresar como

$$\begin{aligned}
E \left(\frac{1}{\lambda_i^2 n^2} \sum_{r=\min(N_i(n), [\lambda_i n])}^{\max(N_i(n), [\lambda_i n])} (Y_i^{(r)})^2 \right)^2 &= E \left(\frac{1}{\lambda_i^4 n^4} \sum_{j=\min(N_i(n), [\lambda_i n])}^{\max(N_i(n), [\lambda_i n])} \sum_{k=\min(N_i(n), [\lambda_i n])}^{\max(N_i(n), [\lambda_i n])} E((Y_i^{(j)})^2 (Y_i^{(k)})^2 / N_i(n)) \right) \\
&= \frac{1}{\lambda_i^4 n^4} \left(E((Y_i^{(j)})^4) |N_i(n) - [\lambda_i n]| \right. \\
&\quad \left. + (E(Y_i^{(j)})^2)^2 |N_i(n) - [\lambda_i n]| (|N_i(n) - [\lambda_i n]| - 1) \right) \\
&= \left(1 - \frac{N_i(n)}{\lambda_i n} \right) \left(E((Y_i^{(j)})^4) \frac{1}{\lambda_i^3 n^3} \right. \\
&\quad \left. + \left(1 - \frac{N_i(n)}{\lambda_i n} - \frac{1}{\lambda_i n} \right) \frac{1}{\lambda_i^2 n^2} (E(Y_i^{(j)})^2)^2 \right),
\end{aligned}$$

ambos tiende a 0 cuando $n \rightarrow \infty$ ya que la variable tiene momentos finitos.

Por otro lado, el segundo término de (2.19) se puede escribir y acotar de la siguiente manera

$$\begin{aligned}
E \left(\left(\frac{\sum_{r=1}^{N_i(n)} Y_i^{(r)}}{N_i(n)} \right)^2 - \left(\frac{\sum_{r=1}^{[\lambda_i n]} Y_i^{(r)}}{\lambda_i n} \right)^2 \right)^2 &= E \left(\frac{\sum_{j=1}^{N_i(n)} \sum_{k=1}^{N_i(n)} Y_i^{(j)} Y_i^{(k)}}{N_i^2(n)} - \frac{\sum_{j=1}^{[\lambda_i n]} \sum_{k=1}^{[\lambda_i n]} Y_i^{(j)} Y_i^{(k)}}{\lambda_i^2 n^2} \right)^2 \\
&\leq 2 \left[E \left(\frac{\sum_{j=1}^{N_i(n)} \sum_{k=1}^{N_i(n)} Y_i^{(j)} Y_i^{(k)}}{N_i^2(n)} - \frac{\sum_{j=1}^{N_i(n)} \sum_{k=1}^{N_i(n)} Y_i^{(j)} Y_i^{(k)}}{\lambda_i^2 n^2} \right)^2 \right. \\
&\quad \left. + E \left(\frac{\sum_{j=1}^{N_i(n)} \sum_{k=1}^{N_i(n)} Y_i^{(j)} Y_i^{(k)}}{\lambda_i^2 n^2} - \frac{\sum_{j=1}^{[\lambda_i n]} \sum_{k=1}^{[\lambda_i n]} Y_i^{(j)} Y_i^{(k)}}{\lambda_i^2 n^2} \right)^2 \right] \\
&= 2 \left[E \left(\left(\frac{1}{N_i^2(n)} - \frac{1}{\lambda_i^2 n^2} \right) \sum_{j=1}^{N_i(n)} \sum_{k=1}^{N_i(n)} Y_i^{(j)} Y_i^{(k)} \right)^2 \right. \\
&\quad \left. + E \left(\frac{1}{\lambda_i^4 n^4} \sum_{j=\min(N_i(n), [\lambda_i n])}^{\max(N_i(n), [\lambda_i n])} \sum_{k=\min(N_i(n), [\lambda_i n])}^{\max(N_i(n), [\lambda_i n])} Y_i^{(j)} Y_i^{(k)} \right)^2 \right]. \tag{2.21}
\end{aligned}$$

Calculando la esperanza mediante esperanza condicional el primer término en (2.21) se puede expresar como

$$E \left(\left(1 - \frac{N_i^2(n)}{\lambda_i^2 n^2} \right)^2 \left(\frac{1}{N_i^2(n)} \right)^2 \sum_{j=1}^{N_i(n)} \sum_{k=1}^{N_i(n)} \sum_{l=1}^{N_i(n)} \sum_{m=1}^{N_i(n)} E \left(Y_i^{(j)} Y_i^{(k)} Y_i^{(l)} Y_i^{(m)} / N_i(n) \right) \right), \tag{2.22}$$

donde los términos cumplen con la relación

$$E \left(Y_i^{(j)} Y_i^{(k)} Y_i^{(l)} Y_i^{(m)} / N_i(n) \right) = \begin{cases} E \left((Y_i^{(j)})^4 \right) & \text{si } j = k = l = m \\ E \left((Y_i^{(j)})^3 \right) E \left(Y_i^{(k)} \right) & \text{si } j = k = l \neq m \\ E \left((Y_i^{(j)})^2 \right) E \left((Y_i^{(k)})^2 \right) & \text{si } j = l \neq k = m \\ E \left(Y_i^{(j)} \right)^4 & \text{si } j \neq k \neq l \neq m \end{cases}, \tag{2.23}$$

y ninguno aparece más de $N_i^2(n)$ veces, por lo tanto (2.22) tiende a 0 cuando $n \rightarrow \infty$.

El segundo término en (2.21) se puede expresar como

$$E \left(\frac{1}{\lambda_i^4 n^4} \sum_{j=\min(N_i(n), [\lambda_i n])}^{\max(N_i(n), [\lambda_i n])} \sum_{k=\min(N_i(n), [\lambda_i n])}^{\max(N_i(n), [\lambda_i n])} \sum_{l=\min(N_i(n), [\lambda_i n])}^{\max(N_i(n), [\lambda_i n])} \sum_{m=\min(N_i(n), [\lambda_i n])}^{\max(N_i(n), [\lambda_i n])} E \left(Y_i^{(j)} Y_i^{(k)} Y_i^{(l)} Y_i^{(m)} / N_i(n) \right) \right). \tag{2.24}$$

Realizando un análisis similar al del término anterior, como ningún término de (2.24) aparece más de $(N_i(n) - \lambda_i n)^4$ veces, la expresión tiende a 0 cuando $n \rightarrow \infty$, lo que concluye la demostración. \square

Como lo que ocurre en cada estado de la cadena Z es independiente, podemos hallar la distribución conjunta del vector $\Upsilon_n = (\mu_1^{(n)}, \mu_2^{(n)}, \dots, \mu_k^{(n)})$ que será Normal multivariada con vector de media $(\mu_1, \mu_2, \dots, \mu_k)$ y matriz de covarianza

$$\Sigma_{\Upsilon} = \begin{bmatrix} \frac{\sigma_1^2}{\lambda_1} & 0 & \dots & 0 \\ 0 & \frac{\sigma_2^2}{\lambda_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \frac{\sigma_k^2}{\lambda_k} \end{bmatrix}. \quad (2.25)$$

A partir de los estimadores de máxima verosimilitud de q_{ij} y de los estimadores para μ_i introducidos en (2.6), cuyas propiedades se muestran en los **Teoremas 2.4 y 2.7**, vamos a construir un estimador consistente de $\alpha(s, t)$.

Definimos el vector $\Lambda = (q_{ij})_{1 \leq i \neq j \leq k} \in \mathbb{R}^{k(k-1)}$, vector de los elementos no diagonales de la matriz \mathbb{Q} y el vector $\Upsilon = (\mu_i)_{1 \leq i \leq k} \in \mathbb{R}^k$, vector de los elementos no diagonales de la matriz \mathbb{H} .

También definimos algunas funciones que nos ayudarán a reconstruir las matrices a partir de los vectores antes presentados, ellas son $\mathcal{Q} : \mathbb{R}^{k(k-1)} \rightarrow \mathcal{M}_{k \times k}$ tal que $\mathcal{Q}(\Lambda) = \mathbb{Q}$, donde $\mathbb{Q} = (Q_{ij})_{1 \leq i, j \leq k}$ es tal que

$$Q_{ij} = \begin{cases} q_{ij} & \text{si } i \neq j \\ \sum_{j=1, j \neq i}^{j=k} q_{ij} = -q_{ii} & \text{si } i = j \end{cases}, \quad (2.26)$$

que reconstruye \mathbb{Q} a partir de los elementos no diagonales, ya que la cadena es conservativa, y $\mathcal{H} : \mathbb{R}^k \rightarrow \mathcal{M}_{k \times k}$ tal que $\mathcal{H}(\Upsilon) = \mathbb{H}$, donde $\mathbb{H} = (H_{ij})_{1 \leq i, j \leq k}$ es tal que

$$H_{ij} = \begin{cases} \mu_i & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases}, \quad (2.27)$$

que reconstruye la matriz \mathbb{H} .

Vamos a presentar otra función que también nos permite reconstruir una matriz, que llamaremos $\hat{\mathbb{Q}}$ y que nos proporciona la misma información que la matriz \mathbb{Q} , pero con algunas propiedades que la hacen mas ventajosa. Sea $\hat{\mathcal{Q}} : \mathbb{R}^{k(k-1)} \rightarrow \mathcal{M}_{k \times k}$ tal que $\hat{\mathcal{Q}}(\Lambda) = \hat{\mathbb{Q}}$, donde $\hat{\mathbb{Q}} = (\hat{Q}_{ij})_{1 \leq i, j \leq k}$ y sus elementos son de la forma

$$\hat{Q}_{ij} = \begin{cases} q_{ij} & \text{si } j < k \\ 1 & \text{si } j = k \end{cases}. \quad (2.28)$$

Tanto \mathbb{Q}, Λ como $\hat{\mathbb{Q}}$ contienen exactamente la misma información y una de las ventajas que tiene $\hat{\mathbb{Q}}$ sobre \mathbb{Q} es que es inversible. Además, el hecho que la información que da \mathbb{Q}

es la misma que da Λ nos permite pensar cualquier parámetro que dependa de \mathbb{Q} como función de Λ .

En efecto, si π es la distribución invariante se tiene que $\pi\mathbb{Q} = 0$ y $\langle \pi, 1 \rangle = 1$, esta información se puede resumir en la ecuación $\pi\hat{\mathbb{Q}} = e_k$, es decir que π depende de \mathbb{Q} y por lo tanto consideramos $\pi = \pi(\Lambda)$.

Por último presentamos ahora otras funciones que nos permitirán escribir al ancho de banda efectivo encontrado en (2.2), mediante su composición. En efecto, definiendo

$$\mathcal{B} : \mathbb{R}^{k(k-1)} \times \mathbb{R}^k \longrightarrow \mathcal{M}_{k \times k} \text{ tal que } \mathcal{B}(\Lambda, \Upsilon) = \exp[(\mathcal{Q}(\Lambda) + \mathcal{H}(\Upsilon)s)t], \quad (2.29)$$

$$g : \mathbb{R}^{k(k-1)} \times \mathbb{R}^k \longrightarrow \mathbb{R} \text{ tal que } g(\Lambda, \Upsilon) = \pi(\Lambda)\mathcal{B}(\Lambda, \Upsilon)\mathbf{1}, \quad (2.30)$$

$$\Psi : \mathbb{R}^{k(k-1)} \times \mathbb{R}^k \longrightarrow \mathbb{R} \text{ tal que } \Psi(\Lambda, \Upsilon) = \frac{1}{st} \log(g(\Lambda, \Upsilon)), \quad (2.31)$$

resulta que

$$\alpha(s, t) = \Psi(\Lambda, \Upsilon). \quad (2.32)$$

Estamos ahora en condiciones de presentar el siguiente teorema que nos da una estimación asintóticamente gaussiana de $\alpha(s, t)$, uno de los objetivos principales de este capítulo.

Teorema 2.9 *Sea X_t un Flujo Markoviano Generalizado, se considera para s y t fijos, $\alpha(s, t)$ definido como en (2.2), Ψ como en (2.31), $q_{ij}^{(n)}$ como en (2.5) y $\mu_i^{(n)}$ como en (2.6). Entonces definiendo*

$$\Lambda_n = \left(q_{ij}^{(n)} \right)_{1 \leq i \neq j \leq k}, \quad (2.33)$$

$$\Upsilon_n = \left(\mu_i^{(n)} \right)_{1 \leq i \leq k}, \quad (2.34)$$

y

$$\alpha^{(n)}(s, t) = \Psi(\Lambda_n, \Upsilon_n), \quad (2.35)$$

resulta que

$$\sqrt{n} (\alpha^{(n)}(s, t) - \alpha(s, t)) \xrightarrow[n \rightarrow \infty]{w} N(0, \sigma^2), \quad (2.36)$$

con $\sigma^2 = \nabla\Psi(\Lambda, \Upsilon)\Sigma'\nabla\Psi(\Lambda, \Upsilon)^t$, donde la matriz de covarianza Σ' es la matriz por bloques $\Sigma' = \begin{bmatrix} \Sigma & 0 \\ 0 & \Sigma_\Upsilon \end{bmatrix}$, Σ es la matriz de covarianza que se deduce en el **Corolario 2.5** y Σ_Υ es la matriz de covarianza definida en (2.25).

Más precisamente se tiene que

$$\begin{aligned} \sigma^2 &= \frac{1}{(st\pi(\Lambda)\mathcal{B}(\Lambda, \Upsilon)\mathbf{1})^2} \left[\sum_{(i,j) \in D} \frac{q_{ij}}{\pi(i)} \left(\frac{\partial\pi(\Lambda)}{\partial q_{ij}} \mathcal{B}(\Lambda, \Upsilon)\mathbf{1} + \pi(\Lambda) \sum_{l=0}^{\infty} \sum_{r=0}^{l-1} A(\Lambda, \Upsilon)\mathbf{1} \right)^2 \right. \\ &\quad \left. + \sum_{i=1}^k \frac{\sigma_i^2}{\lambda_i} \left(\pi(\Lambda) \sum_{l=0}^{\infty} \sum_{r=0}^{l-1} A^*(\Lambda, \Upsilon)\mathbf{1} \right)^2 \right], \end{aligned}$$

donde

$$A(\Lambda, \Upsilon) = \frac{t^l}{l!} (\mathcal{Q}(\Lambda) + \mathcal{H}(\Upsilon)s)^r V^{ij} (\mathcal{Q}(\Lambda) + \mathcal{H}(\Upsilon)s)^{l-r-1}, \quad (2.37)$$

siendo $V^{ij} \in \mathcal{M}_{k \times k}(\mathbb{R})$ tal que

$$(V^{ij})_{lm} = \begin{cases} 1 & \text{si } i = l \text{ y } j = m \neq i \\ -1 & \text{si } l = i = m \\ 0 & \text{en otro caso} \end{cases},$$

y

$$A^*(\Lambda, \Upsilon) = \frac{t^l s^l}{l!} \left((\mathcal{Q}(\Lambda) + \mathcal{H}(\Upsilon)s)^r U^i (\mathcal{Q}(\Lambda) + \mathcal{H}(\Upsilon)s)^{l-r-1} \right), \quad (2.38)$$

con $U^i \in \mathcal{M}_{k \times k}(\mathbb{R})$ y tal que

$$(U^i)_{lm} = \begin{cases} 1 & \text{si } l = m = i \\ 0 & \text{en otro caso} \end{cases}.$$

Para la demostración de este teorema necesitaremos algunos resultados previos.

Teorema 2.10 (*Representación de Skorohod*) Sean $(X_n)_{n \in \mathbb{N}}$ una sucesión de variables aleatorias reales tales que X_n tiene función de distribución F_n . Sea $(a_n)_{n \in \mathbb{N}}$ una sucesión creciente de números reales positivos, tales que $a_n \xrightarrow{n \rightarrow +\infty} +\infty$. Sean $\mu \in \mathbb{R}$ y X una variable aleatoria real con función de distribución F , continua tal que

$$a_n (X_n - \mu) \xrightarrow[n \rightarrow +\infty]{w} X,$$

entonces existen X_n^* y X^* variables aleatorias tales que X_n^* tiene función de distribución F_n , X^* tiene función de distribución F y

$$a_n (X_n^* - \mu) \xrightarrow[n \rightarrow +\infty]{c.s.} X^*.$$

Demostración:

Sea U una variable aleatoria con distribución Uniforme en el intervalo $[0, 1]$, definimos entonces $X_n^* = F_n^{-1}(U)$, donde F_n^{-1} es la inversa generalizada de la función F_n , es decir $F_n^{-1}(t) = \inf\{s : F_n(s) \geq t\}$. Así definida X_n^* es una variable aleatoria con función de distribución F_n . Análogamente se define $X^* = F^{-1}(U)$, con F^{-1} inversa generalizada de a función F , y así X^* es una variable aleatoria con función de distribución F .

Por la convergencia débil que se tiene por hipótesis, si llamamos $Y_n = a_n(X_n - \mu)$ y $F_{Y_n}(s)$ su función de distribución, se cumple que

$$F_{Y_n}(s) \xrightarrow[n \rightarrow +\infty]{} F(s),$$

para todo s ya que F es continua.

Además la convergencia es uniforme, es decir

$$\sup_{s \in \mathbb{R}} |F_{Y_n} - F(s)| \xrightarrow{n \rightarrow +\infty} 0.$$

Llamando $\sup_{s \in \mathbb{R}} |F_{Y_n} - F(s)| = C_n$, tenemos que para todo s se cumple

$$F(s) - C_n \leq F_{Y_n}(s) \leq F(s) + C_n.$$

Por otro lado tenemos que

$$F_{Y_n}(s) = P(a_n(X_n - \mu) \leq s) = P\left(X_n \leq \mu + \frac{s}{a_n}\right) = F_n\left(\mu + \frac{s}{a_n}\right).$$

Consideremos ahora $t \in [0, 1]$, se cumple que

$$\{s : F(s) - C_n \geq t\} \subseteq \left\{s : F_n\left(\mu + \frac{s}{a_n}\right) \geq t\right\} \subseteq \{s : F(s) + C_n \geq t\},$$

y por lo tanto,

$$\{s : F(s) \geq t + C_n\} \subseteq \left\{s : F_n\left(\mu + \frac{s}{a_n}\right) \geq t\right\} \subseteq \{s : F(s) \geq t - C_n\}.$$

Entonces

$$\left\{\mu + \frac{s}{a_n} : F(s) \geq t + C_n\right\} \subseteq \{s : F_n(s) \geq t\} \subseteq \left\{\mu + \frac{s}{a_n} : F(s) \geq t - C_n\right\},$$

y tomando ínfimo sobre s , resulta que

$$\inf_s \left\{\mu + \frac{s}{a_n} : F(s) \geq t + C_n\right\} \geq \inf_s \{s : F_n(s) \geq t\} \geq \inf_s \left\{\mu + \frac{s}{a_n} : F(s) \geq t - C_n\right\},$$

es decir

$$\mu + \frac{1}{a_n} F^{-1}(t + C_n) \geq F_n^{-1}(t) \geq \mu + \frac{1}{a_n} F^{-1}(t - C_n),$$

y también,

$$F^{-1}(t + C_n) \geq a_n (F_n^{-1}(t) - \mu) \geq F^{-1}(t - C_n).$$

Sabemos que F es continua y creciente, que $t + C_n \downarrow t$ y que $t - C_n \uparrow t$, entonces $F^{-1}(t + C_n) \downarrow F^{-1}(t)$ y $F^{-1}(t - C_n) \uparrow F^{-1}(t)$, por lo tanto

$$a_n (F_n^{-1}(t) - \mu) \xrightarrow{n \rightarrow +\infty} F^{-1}(t),$$

salvo que F sea constante en un intervalo de la forma $(F^{-1}(t), F^{-1}(t) + \delta_t)$ con $\delta_t > 0$.

Consideremos el conjunto $\mathcal{C} = \{t : F \text{ es constante en } (F^{-1}(t), F^{-1}(t) + \delta_t)\}$, sabemos que si $t, t' \in \mathcal{C}$ y $t \neq t'$, entonces

$$(F^{-1}(t), F^{-1}(t) + \delta_t) \cap (F^{-1}(t'), F^{-1}(t') + \delta_{t'}) = \emptyset,$$

ya que caso contrario, debería existir $T \in (F^{-1}(t), F^{-1}(t) + \delta_t) \cap (F^{-1}(t'), F^{-1}(t') + \delta_{t'})$ y por lo tanto

$$t = F(F^{-1}(t)) = F(F^{-1}(T)) = F(F^{-1}(t')) = t',$$

lo que nos conduce al absurdo.

El conjunto \mathcal{C} es unión numerable de conjuntos disjuntos formados por un único elemento, por lo tanto es numerable, con lo cual $P(U \in \mathcal{C}) = 0$. Entonces, probamos que para todo $t \in \mathcal{C}^c$ se cumple

$$a_n(F_n^{-1}(t) - \mu) \xrightarrow[n \rightarrow +\infty]{} F^{-1}(t),$$

es decir $P\left(\left\{t : a_n(F_n^{-1}(t) - \mu) \xrightarrow[n \rightarrow +\infty]{} F^{-1}(t)\right\}\right) = 1$, lo que significa que

$$a_n(X_n^* - \mu) \xrightarrow[n \rightarrow +\infty]{c.s.} X^*,$$

y esto concluye la demostración del teorema. \square

Lema 2.11 Sea $(Z_n)_{n \in \mathbb{N}}$ una sucesión de variables aleatorias en \mathbb{R}^d , $d \geq 1$, $(a_n)_{n \in \mathbb{N}}$ una sucesión de números positivos tal que $a_n \rightarrow \infty$ y $Z \in \mathbb{R}^d$ tal que

$$a_n(Z_n - Z) \xrightarrow[n \rightarrow \infty]{w} N(\mathbf{0}, \Sigma),$$

con Σ matriz de covarianza. Sea $G : \mathbb{R}^d \rightarrow \mathbb{R}$ diferenciable en un entorno de Z , entonces se tiene que

$$a_n(G(Z_n) - G(Z)) \xrightarrow[n \rightarrow \infty]{w} N(\mathbf{0}, \nabla G(Z)\Sigma\nabla G(Z)^t).$$

Demostración:

Por el teorema anterior sabemos que existe $(Z_n^*)_{n \in \mathbb{N}}$ una sucesión de variables aleatorias tal que Z_n^* tiene la misma distribución que Z_n y cumple que

$$a_n(Z_n^* - Z) \xrightarrow[n \rightarrow \infty]{c.s.} Z^*, \quad (2.39)$$

donde Z^* tiene distribución $N(\mathbf{0}, \Sigma)$.

Sea $G : \mathbb{R}^d \rightarrow \mathbb{R}$ diferenciable en un entorno de Z , aplicando desarrollo de Taylor se tiene que

$$a_n(G(Z_n) - G(Z)) = a_n\nabla G(Z)(Z_n^* - Z) + a_n(Z_n^* - Z)^t H_{\xi_n^*}(Z_n^* - Z),$$

con ξ_n^* entre Z_n^* y Z . Por (2.39), resulta que

$$a_n\nabla G(Z)(Z_n^* - Z) = \nabla G(Z)a_n(Z_n^* - Z) \xrightarrow[n \rightarrow \infty]{c.s.} \nabla G(Z)Z^*,$$

además

$$a_n(Z_n^* - Z)^t H_{\xi_n^*}(Z_n^* - Z) = \frac{1}{a_n}[a_n(Z_n^* - Z)]^t H_{\xi_n^*} a_n(Z_n^* - Z),$$

donde

$$[a_n(Z_n^* - Z)]^t H_{\xi_n^*} a_n(Z_n^* - Z) \xrightarrow[n \rightarrow \infty]{c.s.} (Z^*)^t H_Z Z^*.$$

Como $\frac{1}{a_n} \xrightarrow[n \rightarrow \infty]{} 0$, se tiene que $a_n (Z_n^* - Z)^t H_{\xi_n^*} (Z_n^* - Z) \xrightarrow[n \rightarrow \infty]{c.s.} 0$, y por lo tanto

$$a_n (G(Z_n^*) - G(Z)) \xrightarrow[n \rightarrow \infty]{c.s.} \nabla G(Z) Z^*.$$

Dado que $Z^* \sim N(\mathbf{0}, \Sigma)$, se cumple que $\nabla G(Z) Z^* \sim N(\mathbf{0}, \nabla G(Z) \Sigma \nabla G(Z)^t)$ y resulta que

$$a_n (G(Z_n^*) - G(Z)) \xrightarrow[n \rightarrow \infty]{w} N(\mathbf{0}, \nabla G(Z) \Sigma \nabla G(Z)^t).$$

Para concluir con la demostración del teorema sólo basta recordar que Z_n^* tiene la misma distribución que Z_n , con lo que se obtiene que

$$a_n (G(Z_n) - G(Z)) \xrightarrow[n \rightarrow \infty]{w} N(\mathbf{0}, \nabla G(Z) \Sigma \nabla G(Z)^t). \quad (2.40)$$

□

Lema 2.12 Considerando Ψ como en (2.31), g como en (2.30) y \mathcal{B} como en (2.29), se tiene que

1. $\frac{\partial \Psi(\Lambda, \Upsilon)}{\partial q_{ij}} = \frac{1}{st g(\Lambda, \Upsilon)} \frac{\partial g(\Lambda, \Upsilon)}{\partial q_{ij}}.$
2. $\frac{\partial \Psi(\Lambda, \Upsilon)}{\partial \mu_i} = \frac{1}{st g(\Lambda, \Upsilon)} \frac{\partial g(\Lambda, \Upsilon)}{\partial \mu_i}.$
3. $\frac{\partial g(\Lambda, \Upsilon)}{\partial q_{ij}} = \frac{\partial \pi(\Lambda)}{\partial q_{ij}} \mathcal{B}(\Lambda, \Upsilon) \mathbf{1} + \pi(\Lambda) \left(\sum_{l=0}^{\infty} \sum_{r=0}^{l-1} \frac{t^l}{l!} (\mathcal{Q}(\Lambda) + \mathcal{H}(\Upsilon)s)^r V^{ij} (\mathcal{Q}(\Lambda) + \mathcal{H}(\Upsilon)s)^{l-r-1} \right) \mathbf{1},$
con $(V^{ij})_{lm} = \begin{cases} 1 & \text{si } i = l \text{ y } j = m \neq i \\ -1 & \text{si } l = i = m \\ 0 & \text{en otro caso} \end{cases}.$
4. $\frac{\partial g(\Lambda, \Upsilon)}{\partial \mu_i} = \pi(\Lambda) \left(\sum_{l=0}^{\infty} \sum_{r=0}^{l-1} \frac{t^l s^l}{l!} (\mathcal{Q}(\Lambda) + \mathcal{H}(\Upsilon)s)^r U^i (\mathcal{Q}(\Lambda) + \mathcal{H}(\Upsilon)s)^{l-r-1} \right) \mathbf{1},$ con
 $(U^i)_{lm} = \begin{cases} 1 & \text{si } l = m = i \\ 0 & \text{en otro caso} \end{cases}.$

Demostración:

Definimos $\Psi(\Lambda, \Upsilon) = \frac{1}{st} \log(g(\Lambda, \Upsilon))$, entonces fácilmente se puede ver que se cumple

$$\frac{\partial \Psi(\Lambda, \Upsilon)}{\partial q_{ij}} = \frac{1}{st g(\Lambda, \Upsilon)} \frac{\partial g(\Lambda, \Upsilon)}{\partial q_{ij}},$$

y

$$\frac{\partial \Psi(\Lambda, \Upsilon)}{\partial \mu_i} = \frac{1}{st g(\Lambda, \Upsilon)} \frac{\partial g(\Lambda, \Upsilon)}{\partial \mu_i},$$

con lo que queda demostrado 1 y 2.

Por definición

$$g(\Lambda, \Upsilon) = \pi(\Lambda) \mathcal{B}(\Lambda, \Upsilon) \mathbf{1}, \quad (2.41)$$

derivando (2.41) respecto del q_{ij} tenemos

$$\frac{\partial g(\Lambda, \Upsilon)}{\partial q_{ij}} = \frac{\partial \pi(\Lambda)}{\partial q_{ij}} \mathcal{B}(\Lambda, \Upsilon) \mathbf{1} + \pi(\Lambda) \frac{\partial (\mathcal{B}(\Lambda, \Upsilon) \mathbf{1})}{\partial q_{ij}}. \quad (2.42)$$

Si notamos $\mathcal{B}(\Lambda, \Upsilon) = (b_{ij})_{1 \leq i, j \leq k}$, se tiene entonces que $\mathcal{B}(\Lambda, \Upsilon) \mathbf{1} = (\sum_{j=1}^k b_{ij})_{1 \leq i \leq k}$, y por lo tanto

$$\frac{\partial (\mathcal{B}(\Lambda, \Upsilon) \mathbf{1})}{\partial q_{ij}} = \left(\frac{\partial (b_{11} + b_{12} + \cdots + b_{1k})}{\partial q_{ij}}, \dots, \frac{\partial (b_{k1} + b_{k2} + \cdots + b_{kk})}{\partial q_{ij}} \right) = \frac{\partial \mathcal{B}(\Lambda, \Upsilon)}{\partial q_{ij}} \mathbf{1},$$

entonces (2.42) se convierte en

$$\frac{\partial g(\Lambda, \Upsilon)}{\partial q_{ij}} = \frac{\partial \pi(\Lambda)}{\partial q_{ij}} (\mathcal{B}(\Lambda, \Upsilon) \mathbf{1}) + \pi(\Lambda) \frac{\partial \mathcal{B}(\Lambda, \Upsilon)}{\partial q_{ij}} \mathbf{1}. \quad (2.43)$$

Por otro lado, usando que $D(\exp(A))(B) = \sum_{l=0}^{\infty} \sum_{r=0}^{l-1} \frac{A^r B A^{l-1-r}}{l!}$, al aplicar el operador diferencial resulta que

$$\begin{aligned} \frac{\partial \mathcal{B}(\Lambda, \Upsilon)}{\partial q_{ij}} &= D(\mathcal{B}(\Lambda, \Upsilon))(e_{ij}) \\ &= D(\exp[(\mathcal{Q}(\Lambda) + \mathcal{H}(\Upsilon)s)t]) D((\mathcal{Q}(\Lambda) + \mathcal{H}(\Upsilon)s)t)(e_{ij}) \\ &= D(\exp[(\mathcal{Q}(\Lambda) + \mathcal{H}(\Upsilon)s)t]) \left(\frac{\partial \mathcal{Q}(\Lambda)}{\partial q_{ij}} \right) t \\ &= \sum_{l=0}^{\infty} \sum_{r=0}^{l-1} \frac{1}{l!} [(\mathcal{Q}(\Lambda) + \mathcal{H}(\Upsilon)s)t]^r \left(\frac{\partial \mathcal{Q}(\Lambda)}{\partial q_{ij}} \right) t [(\mathcal{Q}(\Lambda) + \mathcal{H}(\Upsilon)s)t]^{l-1-r} \\ &= \sum_{l=0}^{\infty} \sum_{r=0}^{l-1} \frac{t^l}{l!} [(\mathcal{Q}(\Lambda) + \mathcal{H}(\Upsilon)s)t]^r V^{ij} [(\mathcal{Q}(\Lambda) + \mathcal{H}(\Upsilon)s)t]^{l-1-r}, \end{aligned} \quad (2.44)$$

con $V^{ij} = \frac{\partial \mathcal{Q}(\Lambda)}{\partial q_{ij}}$ y $(V^{ij})_{lm} = \begin{cases} 1 & \text{si } i = l \text{ y } j = m \neq i \\ -1 & \text{si } l = i = m \\ 0 & \text{en otro caso} \end{cases}$,

y 3 se obtiene sustituyendo (2.44) en (2.43).

Derivando ahora (2.41) respecto del μ_i tenemos que

$$\frac{\partial g(\Lambda, \Upsilon)}{\partial \mu_i} = \pi(\Lambda) \frac{\partial (\mathcal{B}(\Lambda, \Upsilon))}{\partial \mu_i} \mathbf{1}. \quad (2.45)$$

Aplicando nuevamente el operador diferencial resulta que

$$\begin{aligned} \frac{\partial \mathcal{B}(\Lambda, \Upsilon)}{\partial \mu_i} &= D(\mathcal{B}(\Lambda, \Upsilon))(e_{ij}) \\ &= D(\exp[(\mathcal{Q}(\Lambda) + \mathcal{H}(\Upsilon)s)t]) D((\mathcal{Q}(\Lambda) + \mathcal{H}(\Upsilon)s)t)(e_{ij}) \\ &= D(\exp[(\mathcal{Q}(\Lambda) + \mathcal{H}(\Upsilon)s)t]) \left(\frac{\partial \mathcal{H}(\Upsilon)}{\partial \mu_i} \right) st \\ &= \sum_{l=0}^{\infty} \sum_{r=0}^{l-1} \frac{1}{l!} [(\mathcal{Q}(\Lambda) + \mathcal{H}(\Upsilon)s)t]^r \left(\frac{\partial \mathcal{H}(\Upsilon)}{\partial \mu_i} \right) st [(\mathcal{Q}(\Lambda) + \mathcal{H}(\Upsilon)s)t]^{l-1-r} \\ &= \sum_{l=0}^{\infty} \sum_{r=0}^{l-1} \frac{t^l s^l}{l!} [(\mathcal{Q}(\Lambda) + \mathcal{H}(\Upsilon)s)t]^r U^i [(\mathcal{Q}(\Lambda) + \mathcal{H}(\Upsilon)s)t]^{l-1-r}, \end{aligned} \quad (2.46)$$

con $U^i = \frac{\partial \mathcal{H}(\Lambda)}{\partial \mu_i}$, cuyos elementos son de la forma $(U^i)_{lm} = \begin{cases} 1 & \text{si } l = i = m \\ 0 & \text{en otro caso} \end{cases}$ y sustituyendo (2.46) en (2.45) se obtiene $\not\propto$. \square

Observación 2.13

Del Lema anterior se tiene que $\nabla \Psi(\Lambda, \Upsilon) = (\nabla_{\Lambda}, \nabla_{\Upsilon}) \Psi(\Lambda, \Upsilon) = \left(\frac{\partial \Psi(\Lambda, \Upsilon)}{\partial q_{ij}}, \frac{\partial \Psi(\Lambda, \Upsilon)}{\partial \mu_i} \right) \in \mathbb{R}^{k^2}$.

Ahora estamos en condiciones de demostrar el teorema que se propone en esta tesis.

Demostración del Teorema 2.9 :

Por el **Corolario 2.5** $\sqrt{n} \left(q_{ij}^{(n)} - q_{ij} \right) \xrightarrow[n \rightarrow \infty]{w} N(\mathbf{0}, \Sigma)$, donde $\Sigma = \frac{q_{ij}}{\pi(i)} \mathbb{I}_{k \times (k-1)}$ y por la definición de Λ_n , se tiene que

$$\sqrt{n} (\Lambda_n - \Lambda) \xrightarrow[n \rightarrow \infty]{w} N(\mathbf{0}, \Sigma). \quad (2.47)$$

Por otro lado, por el **Teorema 2.7** sabemos que

$$\sqrt{n} (\Upsilon_n - \Upsilon) \xrightarrow[n \rightarrow \infty]{w} N(\mathbf{0}, \Sigma_{\Upsilon}). \quad (2.48)$$

Aplicando entonces el **Lema 2.11**, con $G = \Psi(\Lambda, \Upsilon)$, se obtiene que

$$\sqrt{n} (\alpha^{(n)}(s, t) - \alpha(s, t)) \xrightarrow[n \rightarrow \infty]{w} N(0, \sigma^2), \quad (2.49)$$

con

$$\sigma^2 = \nabla \Psi(\Lambda, \Upsilon) \Sigma' \nabla \Psi(\Lambda, \Upsilon)^t,$$

$$y \Sigma' = \begin{bmatrix} \Sigma & 0 \\ 0 & \Sigma_{\Upsilon} \end{bmatrix}.$$

Veamos ahora que expresión tendrá σ^2 . Por la **Observación (2.13)** tenemos que

$$\begin{aligned} \nabla \Psi(\Lambda, \Upsilon) \Sigma' \nabla \Psi(\Lambda, \Upsilon)^t &= \frac{\partial \Psi(\Lambda, \Upsilon)^2}{\partial q_{11}} \frac{q_{11}}{\pi(1)} + \dots + \frac{\partial \Psi(\Lambda, \Upsilon)^2}{\partial q_{1(k-1)}} \frac{q_{1(k-1)}}{\pi(1)} + \dots \\ &+ \frac{\partial \Psi(\Lambda, \Upsilon)^2}{\partial q_{k1}} \frac{q_{k1}}{\pi(k)} + \dots + \frac{\partial \Psi(\Lambda, \Upsilon)^2}{\partial q_{k(k-1)}} \frac{q_{k(k-1)}}{\pi(k)} + \dots \\ &+ \frac{\partial \Psi(\Lambda, \Upsilon)^2}{\partial \mu_1} \frac{\sigma_1^2}{\lambda_1} + \dots + \frac{\partial \Psi(\Lambda, \Upsilon)^2}{\partial \mu_k} \frac{\sigma_k^2}{\lambda_k} \\ &= \sum_{i=1}^k \sum_{j=1}^{k-1} \frac{\partial \Psi(\Lambda, \Upsilon)^2}{\partial q_{ij}} \frac{q_{ij}}{\pi(i)} + \sum_{i=1}^k \frac{\partial \Psi(\Lambda, \Upsilon)^2}{\partial \mu_i} \frac{\sigma_i^2}{\lambda_i}, \end{aligned}$$

y por lo demostrado en el **Lema (2.12)** resulta que

$$\begin{aligned} \sigma^2 &= \frac{1}{(st\pi(\Lambda)\mathcal{B}(\Lambda, \Upsilon)\mathbf{1})^2} \left[\sum_{(i,j) \in D} \frac{q_{ij}}{\pi(i)} \left(\frac{\partial \pi(\Lambda)}{\partial q_{ij}} \mathcal{B}(\Lambda, \Upsilon)\mathbf{1} + \pi(\Lambda) \sum_{l=0}^{\infty} \sum_{r=0}^{l-1} A(\Lambda, \Upsilon)\mathbf{1} \right)^2 \right. \\ &\left. + \sum_{i=1}^k \frac{\sigma_i^2}{\lambda_i} \left(\pi(\Lambda) \sum_{l=0}^{\infty} \sum_{r=0}^{l-1} A^*(\Lambda, \Upsilon)\mathbf{1} \right)^2 \right], \end{aligned}$$

lo que concluye con la demostración. \square

2.6. Consistencia de los estimadores

En esta sección presentamos estimadores consistentes para cada uno de los parámetros que intervienen en la fórmula de la varianza σ^2 obtenida en el **Teorema 2.9**. Comencemos con una propiedad importante del generador infinitesimal de la cadena modulante en el Flujo Markoviano Generalizado que daremos sin demostración

Como ya dijimos tanto la matriz \mathbb{Q} y $\hat{\mathbb{Q}}$ que construimos con sendas funciones \mathcal{Q} y $\hat{\mathcal{Q}}$ aplicadas al vector de los elementos del generador infinitesimal Λ , nos brindan la misma información. En el siguiente Lema, que daremos sin demostración, presentamos algunas de las propiedades que posee $\hat{\mathbb{Q}}$ y que nos ayudarán a encontrar los estimadores consistentes que buscamos.

Lema 2.14 *La matriz $\hat{\mathbb{Q}} = \hat{\mathcal{Q}}(\Lambda)$ admite inversa $\hat{\mathbb{Q}}^{-1}(\Lambda)$ que es diferenciable y cumple*

$$D(\hat{\mathbb{Q}}^{-1})(\Lambda)(x) = -\hat{\mathbb{Q}}^{-1}(\Lambda)D(\mathcal{Q})(\Lambda)\hat{\mathbb{Q}}^{-1}(\Lambda)(x).$$

Veamos ahora el resultado principal que nos permite encontrar estimadores consistentes para cada parámetro

Proposición 2.15 Sea $\Lambda_n = (q_{ij}^{(n)})_{1 \leq i \neq j \leq k}$, definido en (2.33) y $\Upsilon_n = (\mu_1^{(n)})_{1 \leq i \leq k}$ definido (2.34), entonces

1. $p_n = e_k \hat{\mathcal{Q}}^{-1}(\Lambda_n)$ es un estimador consistente de $\pi(\Lambda)$.
2. $dp_n^{ij} = -e_k \hat{\mathcal{Q}}^{-1}(\Lambda_n) \frac{\partial \hat{\mathcal{Q}}(\Lambda)}{\partial q_{ij}} \hat{\mathcal{Q}}^{-1}(\Lambda_n)$ es un estimador consistente de $\frac{\partial \pi(\Lambda)}{\partial q_{ij}}$.
3. $B_n = \sum_{l=0}^{m_n} \frac{t^l (\mathcal{Q}(\Lambda_n) + \mathcal{H}(\Upsilon_n)s)^l}{l!}$ es un estimador consistente de $B(\Lambda, \Upsilon) = \exp[(\mathcal{Q}(\Lambda) + \mathcal{H}(\Upsilon)s)t]$.
4. $S_n = stp_n B_n \mathbf{1}$ es un estimador consistente de $S = st\pi(\Lambda)B(\Lambda, \Upsilon)\mathbf{1}$.

Demostración:

1. Recordemos que $\hat{\mathcal{Q}} = \hat{\mathcal{Q}}(\Lambda)$. Como ya vimos $\pi \hat{\mathcal{Q}} = e_k$, entonces por el Lema anterior $\pi = e_k \hat{\mathcal{Q}}^{-1}$. Como $\Lambda_n \xrightarrow[n \rightarrow \infty]{c.s.} \Lambda$ y $\hat{\mathcal{Q}}(\Lambda)$ es una función continua, entonces

$$\hat{\mathcal{Q}}(\Lambda_n) \xrightarrow[n \rightarrow \infty]{c.s.} \hat{\mathcal{Q}}(\Lambda).$$

Por otro lado $\hat{\mathcal{Q}}^{-1}$ también es una función continua, entonces para n suficientemente grande, $\hat{\mathcal{Q}}(\Lambda_n)$ admite inversa y se cumple que

$$\hat{\mathcal{Q}}^{-1}(\Lambda_n) \xrightarrow[n \rightarrow \infty]{c.s.} \hat{\mathcal{Q}}^{-1}(\Lambda),$$

por lo tanto $p_n = e_k \hat{\mathcal{Q}}^{-1}(\Lambda_n)$ es un estimador consistente de $\pi(\Lambda)$.

2. Sabemos que $\pi(\Lambda) = e_k \hat{\mathcal{Q}}^{-1}(\Lambda)$, entonces

$$\frac{\partial \pi(\Lambda)}{\partial q_{ij}} = e_k \frac{\partial \hat{\mathcal{Q}}^{-1}}{\partial q_{ij}},$$

aplicando el resultado del **Lema 2.14** resulta

$$\frac{\partial \hat{\mathcal{Q}}^{-1}}{\partial q_{ij}} = -\hat{\mathcal{Q}}^{-1}(\Lambda) \frac{\partial \hat{\mathcal{Q}}^{-1}}{\partial q_{ij}} \hat{\mathcal{Q}}^{-1}(\Lambda).$$

Además sabemos que $\hat{\mathcal{Q}}^{-1}(\Lambda_n)$ es un estimador consistente de $\hat{\mathcal{Q}}^{-1}(\Lambda)$, entonces $dp_n^{ij} = -e_k \hat{\mathcal{Q}}^{-1}(\Lambda_n) \frac{\partial \hat{\mathcal{Q}}(\Lambda_n)}{\partial q_{ij}} \hat{\mathcal{Q}}^{-1}(\Lambda_n)$ es un estimador consistente de $\frac{\partial \pi(\Lambda)}{\partial q_{ij}}$.

3. Probemos ahora que $\mathcal{B}_n = \sum_{l=0}^{m_n} \frac{t^l}{l!} (\mathcal{Q}(\Lambda_n) + \mathcal{H}(\Upsilon_n)s)^l$ es un estimador consistente de $\mathcal{B}(\Lambda, \Upsilon) = \exp[(\mathcal{Q}(\Lambda) + \mathcal{H}(\Upsilon)s)t]$. Es decir probemos que $\mathcal{B}_n \xrightarrow[n \rightarrow \infty]{c.s.} \mathcal{B}(\Lambda, \Upsilon)$, o lo que es equivalente que $|\mathcal{B}_n - \mathcal{B}(\Lambda, \Upsilon)| \xrightarrow[n \rightarrow \infty]{} 0$.

En primer lugar recordemos que la matriz $\mathcal{B}(\Lambda, \Upsilon)$ se puede escribir de dos manera equivalentes como sigue

$$\mathcal{B}(\Lambda, \Upsilon) = \exp[(\mathcal{Q}(\Lambda) + \mathcal{H}(\Upsilon)s)t] = \sum_{l=0}^{\infty} \frac{t^l}{l!} \left((\mathcal{Q}(\Lambda) + \mathcal{H}(\Upsilon)s)^l \right).$$

Veamos entonces como podemos expresar la diferencia entre \mathcal{B}_n y $\mathcal{B}(\Lambda, \Upsilon)$,

$$\begin{aligned} \mathcal{B}_n - \mathcal{B}(\Lambda, \Upsilon) &= \sum_{l=0}^{m_n} \frac{t^l}{l!} (\mathcal{Q}(\Lambda_n) + \mathcal{H}(\Upsilon_n)s)^l - \sum_{l=0}^{\infty} \frac{t^l}{l!} (\mathcal{Q}(\Lambda) + \mathcal{H}(\Upsilon)s)^l \\ &= \sum_{l=0}^{m_n} \frac{t^l}{l!} \left[(\mathcal{Q}(\Lambda_n) + \mathcal{H}(\Upsilon_n)s)^l - (\mathcal{Q}(\Lambda) + \mathcal{H}(\Upsilon)s)^l \right] \\ &\quad + \sum_{l=m_n+1}^{\infty} \frac{t^l}{l!} (\mathcal{Q}(\Lambda) + \mathcal{H}(\Upsilon)s)^l. \end{aligned}$$

El segundo término es la cola de una serie convergente, por lo tanto tiende a 0 cuando $n \rightarrow \infty$.

Resta ver lo que ocurre con el primer término, para lo cual aplicaremos el Teorema del Valor Medio definiendo la función $f : \mathcal{M}_{k \times k} \rightarrow \mathcal{M}_{k \times k}$ tal que $f(M) = M^l$ entonces

$$f(\mathcal{B}_n) - f(\mathcal{B}) = Df(\tilde{\mathcal{B}}_n) \cdot (\mathcal{B}_n - \mathcal{B}), \quad (2.50)$$

donde $\tilde{\mathcal{B}}_n$ está entre \mathcal{B}_n y \mathcal{B} , o en forma equivalente

$$\begin{aligned} f(\mathcal{Q}(\Lambda_n) + \mathcal{H}(\Upsilon_n)s) - f(\mathcal{Q}(\Lambda) + \mathcal{H}(\Upsilon)s) &= Df(\tilde{\mathcal{B}}_n) \cdot (\mathcal{Q}(\Lambda_n) + \mathcal{H}(\Upsilon_n)s - (\mathcal{Q}(\Lambda) + \mathcal{H}(\Upsilon)s)) \\ &= Df(\tilde{\mathcal{B}}_n) \cdot ((\mathcal{Q}(\Lambda_n) - \mathcal{Q}(\Lambda)) + (\mathcal{H}(\Upsilon_n) - \mathcal{H}(\Upsilon))s). \end{aligned}$$

Recordando la definición del operador diferencial de la función $f(A)$, se tiene que $D(f(A))(B) = \sum_{i=0}^{l-1} A^i B A^{l-i-1}$, entonces aplicando esta definición, la ecuación (2.50) se convierte en

$$f(\mathcal{B}_n) - f(\mathcal{B}) = \sum_{i=0}^{l-1} \tilde{\mathcal{B}}_n^i ((\mathcal{Q}(\Lambda_n) - \mathcal{Q}(\Lambda)) + (\mathcal{H}(\Upsilon_n) - \mathcal{H}(\Upsilon))s) \tilde{\mathcal{B}}_n^{l-i-1},$$

y tenemos así que

$$\begin{aligned}
& \left\| \sum_{l=0}^{m_n} \frac{t^l}{l!} \left[(\mathcal{Q}(\Lambda_n) + \mathcal{H}(\Upsilon_n)s)^l - (\mathcal{Q}(\Lambda) + \mathcal{H}(\Upsilon)s)^l \right] \right\| \\
&= \left\| \sum_{l=0}^{m_n} \frac{t^l}{l!} \sum_{i=0}^{l-1} \tilde{\mathcal{B}}_n^i ((\mathcal{Q}(\Lambda_n) - \mathcal{Q}(\Lambda)) + (\mathcal{H}(\Upsilon_n) - \mathcal{H}(\Upsilon))s) \tilde{\mathcal{B}}_n^{l-i-1} \right\| \\
&\leq \sum_{l=0}^{m_n} \frac{t^l}{l!} \sum_{i=0}^{l-1} \|\tilde{\mathcal{B}}_n\|^i \|(\mathcal{Q}(\Lambda_n) - \mathcal{Q}(\Lambda)) + (\mathcal{H}(\Upsilon_n) - \mathcal{H}(\Upsilon))s\| \|\tilde{\mathcal{B}}_n\|^{l-i-1} \\
&\leq \sum_{l=0}^{m_n} \frac{t^l}{l!} \sum_{i=0}^{l-1} \|\tilde{\mathcal{B}}_n\|^{l-1} (\|(\mathcal{Q}(\Lambda_n) - \mathcal{Q}(\Lambda))\| + \|(\mathcal{H}(\Upsilon_n) - \mathcal{H}(\Upsilon))s\|) \\
&\leq \sum_{l=0}^{m_n} \frac{t^l}{l!} l \|\tilde{\mathcal{B}}_n\|^{l-1} (\|(\mathcal{Q}(\Lambda_n) - \mathcal{Q}(\Lambda))\| + \|(\mathcal{H}(\Upsilon_n) - \mathcal{H}(\Upsilon))s\|) \\
&= (\|(\mathcal{Q}(\Lambda_n) - \mathcal{Q}(\Lambda))\| + \|(\mathcal{H}(\Upsilon_n) - \mathcal{H}(\Upsilon))s\|) t \sum_{l=0}^{m_n} \frac{t^{l-1}}{(l-1)!} \|\tilde{\mathcal{B}}_n\|^{l-1}.
\end{aligned}$$

Con un argumento análogo al que ya utilizamos, se tiene que $\sum_{l=0}^{m_n} \frac{t^{l-1}}{(l-1)!} \|\tilde{\mathcal{B}}_n\|^{l-1}$ está acotada por ser la suma parcial de una serie convergente y que cada uno de los términos del primer factor tienden a cero porque tanto Λ_n como Υ_n son estimadores consistentes de Λ y Υ respectivamente, lo que deja demostrado que \mathcal{B}_n es un estimador consistente de \mathcal{B} .

4. Este punto se deduce directamente del punto 1, donde probamos que p_n es un estimador consistente de $\pi(\Lambda)$ y del punto 3, donde probamos que \mathcal{B}_n es un estimador consistente de $\mathcal{B}(\Lambda, \Upsilon)$. \square

2.7. Intervalo de Confianza para $\alpha(s, t)$

El propósito de esta sección es construir un intervalo de confianza para el ancho de banda efectivo $\alpha(s, t)$, para esto debemos previamente presentar un estimador consistente de la varianza obtenida en el **Teorema 2.9**, a partir de los estimadores consistentes que hemos encontrado en la sección anterior para cada uno de los parámetros que intervienen en la expresión de la varianza.

Teorema 2.16 Sean $q_{ij}^{(n)}$ los estimadores de máxima verosimilitud presentados en el **Teorema 2.4**, S_n, p_n, dp_n y \mathcal{B}_n los estimadores vistos en la **Proposición 2.15** y m_n una

sucesión de números reales positivos tales que $m_n \xrightarrow[n \rightarrow \infty]{} \infty$, entonces

$$\begin{aligned} \sigma_n^2 = & \frac{1}{S_n^2} \left[\sum_{i=1}^k \sum_{j=1}^{k-1} \frac{q_{ij}^{(n)}}{p_n(i)} \left(dp_n^{ij} \mathcal{B}_n \mathbf{1} + p_m^{ij} \sum_{l=0}^{m_n} \sum_{r=0}^{l-1} A(\Lambda_n, \Upsilon_n) \mathbf{1} \right)^2 \right. \\ & \left. + \sum_{l=1}^k \frac{\widehat{\sigma}_i^2}{\widehat{\lambda}_i} \left(p_n^{ij} \sum_{l=0}^{m_n} \sum_{r=0}^{l-1} A^*(\Lambda_n, \Upsilon_n) \mathbf{1} \right)^2 \right], \end{aligned} \quad (2.51)$$

es un estimador consistente de σ^2 , con $A(\cdot, \cdot)$ definida como en (2.37) y $A^*(\cdot, \cdot)$ definida como en (2.38).

Demostración:

Por los resultados obtenidos sobre la consistencia de algunos estimadores en las **Proposiciones 2.15** y **2.5.1**, sólo habría que probar que

$$\sum_{l=0}^{m_n} \sum_{r=0}^{l-1} A(\Delta_n, \Upsilon_n) \xrightarrow[n \rightarrow \infty]{c.s.} \sum_{l=0}^{\infty} \sum_{r=0}^{l-1} A(\Delta, \Upsilon), \quad (2.52)$$

y

$$\sum_{l=0}^{m_n} \sum_{r=0}^{l-1} A^*(\Delta_n, \Upsilon_n) \xrightarrow[n \rightarrow \infty]{c.s.} \sum_{l=0}^{\infty} \sum_{r=0}^{l-1} A^*(\Delta, \Upsilon). \quad (2.53)$$

Teniendo en cuenta la definición de $A(\cdot, \cdot)$, para probar que vale (2.52) definimos $T(l, r) : \mathbb{R}^{k^2} \rightarrow \mathcal{M}_{k \times k}$ tal que

$$T(l, r)(\Lambda, \Upsilon) = (\mathcal{Q}(\Lambda) + \mathcal{H}(\Upsilon)s)^r V^{ij} (\mathcal{Q}(\Lambda) + \mathcal{H}(\Upsilon)s)^{l-r-1},$$

y por lo tanto se tiene que

$$T(l, r)(\Lambda_n, \Upsilon_n) = (\mathcal{Q}(\Lambda_n) + \mathcal{H}(\Upsilon_n)s)^r V^{ij} (\mathcal{Q}(\Lambda_n) + \mathcal{H}(\Upsilon_n)s)^{l-r-1},$$

entonces bastaría probar que

$$\sum_{l=0}^{m_n} \sum_{r=0}^{l-1} \frac{t^l}{l!} T(l, r)(\Lambda_n, \Upsilon_n) - \sum_{l=0}^{\infty} \sum_{r=0}^{l-1} \frac{t^l}{l!} T(l, r)(\Lambda, \Upsilon) \xrightarrow[n \rightarrow \infty]{c.s.} 0.$$

Para esto veamos que la resta anterior se puede expresar de la forma

$$\sum_{l=0}^{m_n} \sum_{r=0}^{l-1} \frac{t^l}{l!} (T(l, r)(\Lambda_n, \Upsilon_n) - T(l, r)(\Lambda, \Upsilon)) - \sum_{l=m_n+1}^{\infty} \sum_{r=0}^{l-1} \frac{t^l}{l!} T(l, r)(\Lambda, \Upsilon), \quad (2.54)$$

y para probar que dicha expresión tiende casi seguramente a cero debemos ver que ocurre con cada término. En primer lugar analicemos el segundo término de la ecuación (2.54)

$$\begin{aligned}
\left\| \sum_{l=m_n+1}^{\infty} \sum_{r=0}^{l-1} \frac{t^l}{l!} T(l,r)(\Lambda, \Upsilon) \right\| &\leq \sum_{l=m_n+1}^{\infty} \sum_{r=0}^{l-1} \frac{t^l}{l!} \|T(l,r)(\Lambda, \Upsilon)\| \\
&\leq \sum_{l=m_n+1}^{\infty} \sum_{r=0}^{l-1} \frac{t^l}{l!} \|\mathcal{Q}(\Lambda) + \mathcal{H}(\Upsilon)s\|^r \|V^{ij}\| \|\mathcal{Q}(\Lambda) + \mathcal{H}(\Upsilon)s\|^{l-r-1} \\
&= \|V^{ij}\| \sum_{l=m_n+1}^{\infty} \sum_{r=0}^{l-1} \frac{t^l}{l!} \|\mathcal{Q}(\Lambda) + \mathcal{H}(\Upsilon)s\|^{l-1},
\end{aligned}$$

que tiende a 0 ya que el primer factor está acotado y el otro factor es la cola de la serie convergente $e^{t\|\mathcal{Q}(\Lambda)+\mathcal{H}(\Upsilon)s\|}$.

Veamos ahora el primer término de la ecuación (2.54)

$$\left\| \sum_{l=0}^{m_n} \sum_{r=0}^{l-1} \frac{t^l}{l!} (T(l,r)(\Lambda_n, \Upsilon_n) - T(l,r)(\Lambda, \Upsilon)) \right\| \leq \sum_{l=0}^{m_n} \sum_{r=0}^{l-1} \frac{t^l}{l!} \|(T(l,r)(\Lambda_n, \Upsilon_n) - T(l,r)(\Lambda, \Upsilon))\|, \quad (2.55)$$

aplicando Teorema del Valor Medio con $(\tilde{\Lambda}_n, \tilde{\Upsilon}_n)$ entre (Λ_n, Υ_n) y (Λ, Υ) , y teniendo en cuenta que $(\Lambda_n, \Upsilon_n) - (\Lambda, \Upsilon) = (\Lambda_n - \Lambda, \Upsilon_n - \Upsilon)$ tenemos que

$$T(l,r)(\Lambda_n, \Upsilon_n) - T(l,r)(\Lambda, \Upsilon) = DT(l,r)(\tilde{\Lambda}_n, \tilde{\Upsilon}_n)(\Lambda_n - \Lambda, \Upsilon_n - \Upsilon),$$

y por lo tanto la ecuación (2.55) es equivalente a

$$\sum_{l=0}^{m_n} \sum_{r=0}^{l-1} \frac{t^l}{l!} \left\| DT(l,r)(\tilde{\Lambda}_n, \tilde{\Upsilon}_n)(\Lambda_n - \Lambda, \Upsilon_n - \Upsilon) \right\|. \quad (2.56)$$

Definiendo las funciones

$$F(l,r)(\Lambda, \Upsilon) = (\mathcal{Q}(\Lambda) + \mathcal{H}(\Upsilon)s)^r, \quad (2.57)$$

$$G(l,r)(\Lambda, \Upsilon) = (\mathcal{Q}(\Lambda) + \mathcal{H}(\Upsilon)s)^{l-r-1}, \quad (2.58)$$

la función $T(l,r)(\Lambda, \Upsilon)$ se puede expresar de la siguiente manera

$$T(l,r)(\Lambda, \Upsilon) = F(l,r)(\Lambda, \Upsilon)V^{ij}G(l,r)(\Lambda, \Upsilon),$$

entonces, al aplicar el operador diferencial tenemos que

$$\begin{aligned}
DT(l,r)(\tilde{\Lambda}_n, \tilde{\Upsilon}_n)(\Lambda_n - \Lambda, \Upsilon_n - \Upsilon) &= DF(l,r)(\tilde{\Lambda}_n, \tilde{\Upsilon}_n)(\Lambda_n - \Lambda, \Upsilon_n - \Upsilon)V^{ij}G(l,r)(\tilde{\Lambda}_n, \tilde{\Upsilon}_n) \\
&\quad + F(l,r)(\tilde{\Lambda}_n, \tilde{\Upsilon}_n)V^{ij}DG(l,r)(\tilde{\Lambda}_n, \tilde{\Upsilon}_n)(\Lambda_n - \Lambda, \Upsilon_n - \Upsilon),
\end{aligned}$$

y cada término en la ecuación (2.56) se puede acotar de la siguiente manera

$$\begin{aligned} \left\| DT(l, r)(\tilde{\Lambda}_n, \tilde{\Upsilon}_n)(\Lambda_n - \Lambda, \Upsilon_n - \Upsilon) \right\| &\leq \left\| DF(l, r)(\tilde{\Lambda}_n, \tilde{\Upsilon}_n)(\Lambda_n - \Lambda, \Upsilon_n - \Upsilon) V^{ij} G(l, r)(\tilde{\Lambda}_n, \tilde{\Upsilon}_n) \right\| \\ &\quad + \left\| F(l, r)(\tilde{\Lambda}_n, \tilde{\Upsilon}_n) V^{ij} DG(l, r)(\tilde{\Lambda}_n, \tilde{\Upsilon}_n)(\Lambda_n - \Lambda, \Upsilon_n - \Upsilon) \right\|. \end{aligned} \quad (2.59)$$

Teniendo en cuenta que se puede pensar que $F(l, r)(\Lambda, \Upsilon) = G_r(\mathcal{A}(\Lambda, \Upsilon))$, donde $G_r(B) = B^r$ y $\mathcal{A}(\Lambda, \Upsilon) = \mathcal{Q}(\Lambda) + \mathcal{H}(\Upsilon)s$, al aplicar el operador diferencial se tendrá que

$$\begin{aligned} DF(l, r)(\tilde{\Lambda}_n, \tilde{\Upsilon}_n)(\Lambda_n - \Lambda, \Upsilon_n - \Upsilon) &= \\ &= \sum_{j=0}^{r-1} \left(\mathcal{Q}(\tilde{\Lambda}) + \mathcal{H}(\tilde{\Upsilon})s \right)^j \left((\mathcal{Q}(\Lambda_n) + \mathcal{H}(\Upsilon_n)s) - (\mathcal{Q}(\Lambda) + \mathcal{H}(\Upsilon)s) \right) \left(\mathcal{Q}(\tilde{\Lambda}) + \mathcal{H}(\tilde{\Upsilon})s \right)^{r-j-1} \\ &= \sum_{j=0}^{r-1} \left(\mathcal{Q}(\tilde{\Lambda}) + \mathcal{H}(\tilde{\Upsilon})s \right)^j \left(\mathcal{Q}(\Lambda_n - \Lambda) + \mathcal{H}(\Upsilon_n - \Upsilon)s \right) \left(\mathcal{Q}(\tilde{\Lambda}) + \mathcal{H}(\tilde{\Upsilon})s \right)^{r-j-1}, \end{aligned}$$

y en consecuencia

$$\begin{aligned} \left\| DF(l, r)(\tilde{\Lambda}_n, \tilde{\Upsilon}_n)(\Lambda_n - \Lambda, \Upsilon_n - \Upsilon) \right\| &= \left\| \sum_{j=0}^{r-1} \left(\mathcal{Q}(\tilde{\Lambda}) + \mathcal{H}(\tilde{\Upsilon})s \right)^j \left((\mathcal{Q}(\Lambda_n) + \mathcal{H}(\Upsilon_n)s) - \right. \right. \\ &\quad \left. \left. (\mathcal{Q}(\Lambda) + \mathcal{H}(\Upsilon)s) \right) \left(\mathcal{Q}(\tilde{\Lambda}) + \mathcal{H}(\tilde{\Upsilon})s \right)^{r-j-1} \right\| \\ &\leq \sum_{j=0}^{r-1} \left\| \mathcal{Q}(\Lambda_n - \Lambda) + \mathcal{H}(\Upsilon_n - \Upsilon)s \right\| \left\| \mathcal{Q}(\tilde{\Lambda}) + \mathcal{H}(\tilde{\Upsilon})s \right\|^{r-1} \\ &\leq r \left(\left\| \mathcal{Q}(\Lambda_n - \Lambda) \right\| + \left\| \mathcal{H}(\Upsilon_n - \Upsilon)s \right\| \right) \left\| \mathcal{Q}(\tilde{\Lambda}) + \mathcal{H}(\tilde{\Upsilon})s \right\|^{r-1}. \end{aligned} \quad (2.60)$$

Aplicando un razonamiento análogo para la función $G(l, r)$ tenemos que

$$\begin{aligned} DG(l, r)(\tilde{\Lambda}_n, \tilde{\Upsilon}_n)(\Lambda_n - \Lambda, \Upsilon_n - \Upsilon) &= \\ &= \sum_{j=0}^{l-r-2} \left(\mathcal{Q}(\tilde{\Lambda}) + \mathcal{H}(\tilde{\Upsilon})s \right)^j \left((\mathcal{Q}(\Lambda_n) + \mathcal{H}(\Upsilon_n)s) - (\mathcal{Q}(\Lambda) + \mathcal{H}(\Upsilon)s) \right) \left(\mathcal{Q}(\tilde{\Lambda}) + \mathcal{H}(\tilde{\Upsilon})s \right)^{l-r-j-2} \\ &= \sum_{j=0}^{l-r-2} \left(\mathcal{Q}(\tilde{\Lambda}) + \mathcal{H}(\tilde{\Upsilon})s \right)^j \left(\mathcal{Q}(\Lambda_n - \Lambda) + \mathcal{H}(\Upsilon_n - \Upsilon)s \right) \left(\mathcal{Q}(\tilde{\Lambda}) + \mathcal{H}(\tilde{\Upsilon})s \right)^{l-r-j-2}, \end{aligned}$$

y

$$\begin{aligned}
\left\| DG(l, r)(\tilde{\Lambda}_n, \tilde{\Upsilon}_n)(\Lambda_n - \Lambda, \Upsilon_n - \Upsilon) \right\| &\leq (l - r - 1) \|(\mathcal{Q}(\Lambda_n) + \mathcal{H}(\Upsilon_n)_s) \\
&\quad - (\mathcal{Q}(\Lambda) + \mathcal{H}(\Upsilon)_s)\| \left\| \mathcal{Q}(\tilde{\Lambda}) + \mathcal{H}(\tilde{\Upsilon})_s \right\|^{l-r-2} \\
&= (l - r - 1) \|(\mathcal{Q}(\Lambda_n - \Lambda) \\
&\quad + \mathcal{H}(\Upsilon_n - \Upsilon)_s)\| \left\| \mathcal{Q}(\tilde{\Lambda}) + \mathcal{H}(\tilde{\Upsilon})_s \right\|^{l-r-2} \quad (2.61)
\end{aligned}$$

Sustituyendo (2.61) y (2.60) en (2.59) resulta que

$$\begin{aligned}
&\left\| DT(l, r)(\tilde{\Lambda}_n, \tilde{\Upsilon}_n)(\Lambda_n - \Lambda, \Upsilon_n - \Upsilon) \right\| \leq \\
&\leq r \|\mathcal{Q}(\Lambda_n - \Lambda) + \mathcal{H}(\Upsilon_n - \Upsilon)_s\| \left\| \mathcal{Q}(\tilde{\Lambda}) + \mathcal{H}(\tilde{\Upsilon})_s \right\|^{r-1} \|V^{ij}\| \|G(l, r)(\tilde{\Lambda}_n, \tilde{\Upsilon}_n)\| \\
&\quad + \|F(l, r)(\tilde{\Lambda}_n, \tilde{\Upsilon}_n)\| \|V^{ij}\| (l - r - 1) \|\mathcal{Q}(\tilde{\Lambda}) + \mathcal{H}(\tilde{\Upsilon})_s\|^{l-r-2} \|\mathcal{Q}(\Lambda_n - \Lambda) + \mathcal{H}(\Upsilon_n - \Upsilon)_s\| \\
&= r \|\mathcal{Q}(\Lambda_n - \Lambda) + \mathcal{H}(\Upsilon_n - \Upsilon)_s\| \left\| \mathcal{Q}(\tilde{\Lambda}) + \mathcal{H}(\tilde{\Upsilon})_s \right\|^{r-1} \|V^{ij}\| \left\| \mathcal{Q}(\tilde{\Lambda}) + \mathcal{H}(\tilde{\Upsilon})_s \right\|^{l-r-1} \\
&\quad + \left\| \mathcal{Q}(\tilde{\Lambda}) + \mathcal{H}(\tilde{\Upsilon})_s \right\|^r \|V^{ij}\| (l - r - 1) \|\mathcal{Q}(\tilde{\Lambda}) + \mathcal{H}(\tilde{\Upsilon})_s\|^{l-r-2} \|\mathcal{Q}(\Lambda_n - \Lambda) + \mathcal{H}(\Upsilon_n - \Upsilon)_s\| \\
&= r \|\mathcal{Q}(\Lambda_n - \Lambda) + \mathcal{H}(\Upsilon_n - \Upsilon)_s\| \|\mathcal{Q}(\tilde{\Lambda}) + \mathcal{H}(\tilde{\Upsilon})_s\|^{l-2} \|V^{ij}\| + (l - r - 1) \|\mathcal{Q}(\Lambda_n - \Lambda) \\
&\quad \mathcal{H}(\Upsilon_n - \Upsilon)_s\| \|\mathcal{Q}(\tilde{\Lambda}) + \mathcal{H}(\tilde{\Upsilon})_s\|^{l-2} \|V^{ij}\| \\
&= (l - 1) \|\mathcal{Q}(\Lambda_n - \Lambda) + \mathcal{H}(\Upsilon_n - \Upsilon)_s\| \|\mathcal{Q}(\tilde{\Lambda}) + \mathcal{H}(\tilde{\Upsilon})_s\|^{l-2} \|V^{ij}\|,
\end{aligned}$$

por lo tanto el primer término de la ecuación (2.56) se terminará acotando de la siguiente manera

$$\begin{aligned}
&\left\| \sum_{l=0}^{m_n} \sum_{r=0}^{l-1} \frac{t^l}{l!} (T(\Lambda_n, \Upsilon_n) - T(\Lambda, \Upsilon)) \right\| \leq \\
&\leq \sum_{l=0}^{m_n} \sum_{r=0}^{l-1} \frac{t^l}{l!} (l - 1) \|\mathcal{Q}(\Lambda_n - \Lambda) + \mathcal{H}(\Upsilon_n - \Upsilon)_s\| \|\mathcal{Q}(\tilde{\Lambda}) + \mathcal{H}(\tilde{\Upsilon})_s\|^{l-2} \|V^{ij}\| \\
&= \|V^{ij}\| \|\mathcal{Q}(\Lambda_n - \Lambda) + \mathcal{H}(\Upsilon_n - \Upsilon)_s\| t^2 \sum_{l=0}^{m_n} \frac{t^{l-2}}{l!} (l - 1) l \|\mathcal{Q}(\tilde{\Lambda}) + \mathcal{H}(\tilde{\Upsilon})_s\|^{l-2} \\
&= \|V^{ij}\| \|\mathcal{Q}(\Lambda_n - \Lambda) + \mathcal{H}(\Upsilon_n - \Upsilon)_s\| t^2 \sum_{l=2}^{m_n} \frac{t^{l-2}}{(l-2)!} \|\mathcal{Q}(\tilde{\Lambda}) + \mathcal{H}(\tilde{\Upsilon})_s\|^{l-2} \\
&= \|V^{ij}\| \|\mathcal{Q}(\Lambda_n - \Lambda) + \mathcal{H}(\Upsilon_n - \Upsilon)_s\| t^2 \sum_{v=0}^{m_n+2} \frac{t^v}{v!} \|\mathcal{Q}(\tilde{\Lambda}) + \mathcal{H}(\tilde{\Upsilon})_s\|^v.
\end{aligned}$$

Dado que tanto \mathcal{Q} como \mathcal{H} son funciones continuas y que Λ_n y Υ_n son estimadores consistentes de Λ y Υ respectivamente, se cumple que

$$\mathcal{Q}(\Lambda_n) \xrightarrow[n \rightarrow \infty]{cs} \mathcal{Q}(\Lambda),$$

$$\mathcal{H}(\Upsilon_n) \xrightarrow[n \rightarrow \infty]{c.s.} \mathcal{H}(\Upsilon),$$

y entonces

$$\|\mathcal{Q}(\Lambda_n) + \mathcal{H}(\Upsilon_n)s\| \xrightarrow[n \rightarrow \infty]{c.s.} \|\mathcal{Q}(\Lambda) + \mathcal{H}(\Upsilon)s\|, \quad (2.62)$$

sólo bastaría probar que $\sum_{v=0}^{m_n+2} \frac{t^v}{v!} \|\mathcal{Q}(\tilde{\Lambda}) + \mathcal{H}(\tilde{\Upsilon})s\|^v$ está acotada.

En efecto, por (2.62) toda sucesión convergente está acotada, por lo tanto existe $M > 0$ tal que

$$\|\mathcal{Q}(\tilde{\Lambda}) + \mathcal{H}(\tilde{\Upsilon})s\| \leq M,$$

y resulta que

$$\sum_{v=0}^{m_n+2} \frac{t^v}{v!} \|\mathcal{Q}(\tilde{\Lambda}) + \mathcal{H}(\tilde{\Upsilon})s\|^{l-2} \leq \sum_{v=0}^{m_n+2} \frac{t^v}{v!} M^v = \sum_{v=0}^{m_n+2} \frac{(tM)^v}{v!} \leq e^{tM} = C,$$

lo que concluye la demostración de la validez de (2.52).

Teniendo en cuenta la definición de $A^*(\cdot, \cdot)$, para probar que vale (2.53) definimos $T^*(l, r) : \mathbb{R}^{k^2} \rightarrow \mathcal{M}_{k \times k}$ tal que

$$T^*(l, r)(\Lambda, \Upsilon) = (\mathcal{Q}(\Lambda) + \mathcal{H}(\Upsilon)s)^r U^i (\mathcal{Q}(\Lambda) + \mathcal{H}(\Upsilon)s)^{l-r-1},$$

y por lo tanto se tiene que

$$T^*(l, r)(\Lambda_n, \Upsilon_n) = (\mathcal{Q}(\Lambda_n) + \mathcal{H}(\Upsilon_n)s)^r U^i (\mathcal{Q}(\Lambda_n) + \mathcal{H}(\Upsilon_n)s)^{l-r-1},$$

entonces bastaría probar que

$$\sum_{l=0}^{m_n} \sum_{r=0}^{l-1} \frac{t^l s^l}{l!} T^*(l, r)(\Lambda_n, \Upsilon_n) - \sum_{l=0}^{\infty} \sum_{r=0}^{l-1} \frac{t^l s^l}{l!} T^*(l, r)(\Lambda, \Upsilon) \xrightarrow[n \rightarrow \infty]{c.s.} 0.$$

Para esto veamos que la resta anterior se puede expresar de la forma

$$\sum_{l=0}^{m_n} \sum_{r=0}^{l-1} \frac{t^l s^l}{l!} (T^*(l, r)(\Lambda_n, \Upsilon_n) - T^*(l, r)(\Lambda, \Upsilon)) - \sum_{l=m_n+1}^{\infty} \sum_{r=0}^{l-1} \frac{t^l s^l}{l!} T^*(l, r)(\Lambda, \Upsilon), \quad (2.63)$$

y para probar que dicha expresión tiende casi seguramente a cero debemos ver que ocurre con cada término. En primer lugar analicemos el segundo término de la ecuación (2.63)

$$\begin{aligned} \left\| \sum_{l=m_n+1}^{\infty} \sum_{r=0}^{l-1} \frac{t^l s^l}{l!} T^*(l, r)(\Lambda, \Upsilon) \right\| &\leq \sum_{l=m_n+1}^{\infty} \sum_{r=0}^{l-1} \frac{t^l s^l}{l!} \|T^*(l, r)(\Lambda, \Upsilon)\| \\ &\leq \sum_{l=m_n+1}^{\infty} \sum_{r=0}^{l-1} \frac{t^l s^l}{l!} \|\mathcal{Q}(\Lambda) + \mathcal{H}(\Upsilon)s\|^r \|U^i\| \|\mathcal{Q}(\Lambda) + \mathcal{H}(\Upsilon)s\|^{l-r-1} \\ &= \|U^i\| \sum_{l=m_n+1}^{\infty} \sum_{r=0}^{l-1} \frac{t^l s^l}{l!} \|\mathcal{Q}(\Lambda) + \mathcal{H}(\Upsilon)s\|^{l-1}, \end{aligned}$$

que tiende a 0 ya que el primer factor está acotado y el otro factor es la cola de la serie convergente $e^{ts\|\mathcal{Q}(\Lambda)+\mathcal{H}(\Upsilon)s\|}$.

Veamos ahora el primer término de la ecuación (2.63)

$$\left\| \sum_{l=0}^{m_n} \sum_{r=0}^{l-1} \frac{t^l s^l}{l!} (T^*(l, r)(\Lambda_n, \Upsilon_n) - T^*(l, r)(\Lambda, \Upsilon)) \right\| \leq \sum_{l=0}^{m_n} \sum_{r=0}^{l-1} \frac{t^l s^l}{l!} \|(T^*(l, r)(\Lambda_n, \Upsilon_n) - T^*(l, r)(\Lambda, \Upsilon))\|, \quad (2.64)$$

aplicando Teorema del Valor Medio con $(\tilde{\Lambda}_n, \tilde{\Upsilon}_n)$ entre (Λ_n, Υ_n) tenemos que

$$T^*(l, r)(\Lambda_n, \Upsilon_n) - T^*(l, r)(\Lambda, \Upsilon) = DT^*(l, r)(\tilde{\Lambda}_n, \tilde{\Upsilon}_n)(\Lambda_n - \Lambda, \Upsilon_n - \Upsilon),$$

y por lo tanto la ecuación (2.64) es equivalente a

$$\sum_{l=0}^{m_n} \sum_{r=0}^{l-1} \frac{t^l s^l}{l!} \left\| DT^*(l, r)(\tilde{\Lambda}_n, \tilde{\Upsilon}_n)(\Lambda_n - \Lambda, \Upsilon_n - \Upsilon) \right\|. \quad (2.65)$$

La función $T^*(l, r)(\Lambda, \Upsilon)$ se puede expresar de la siguiente manera

$$T^*(l, r)(\Lambda, \Upsilon) = F(l, r)(\Lambda, \Upsilon)U^i G(l, r)(\Lambda, \Upsilon),$$

donde F y G están definidas como en (2.57) y (2.58) respectivamente. Aplicando el operador diferencial tenemos que

$$DT^*(l, r)(\tilde{\Lambda}_n, \tilde{\Upsilon}_n)(\Lambda_n - \Lambda, \Upsilon_n - \Upsilon) = DF(l, r)(\tilde{\Lambda}_n, \tilde{\Upsilon}_n)(\Lambda_n - \Lambda, \Upsilon_n - \Upsilon)U^i G(l, r)(\tilde{\Lambda}_n, \tilde{\Upsilon}_n) + F(l, r)(\tilde{\Lambda}_n, \tilde{\Upsilon}_n)U^i DG(l, r)(\tilde{\Lambda}_n, \tilde{\Upsilon}_n)(\Lambda_n - \Lambda, \Upsilon_n - \Upsilon),$$

y teniendo en cuenta las cotas obtenidas en (2.60) y (2.61) para $DF(l, r)$ y $DG(l, r)$ respectivamente, cada término en la ecuación (2.65) se puede acotar de la siguiente manera

$$\begin{aligned} & \left\| \sum_{l=0}^{m_n} \sum_{r=0}^{l-1} \frac{t^l s^l}{l!} (T^*(\Lambda_n, \Upsilon_n) - T^*(\Lambda, \Upsilon)) \right\| \leq \\ & \leq \sum_{l=0}^{m_n} \sum_{r=0}^{l-1} \frac{t^l s^l}{l!} (l-1) \|\mathcal{Q}(\Lambda_n - \Lambda) + \mathcal{H}(\Upsilon_n - \Upsilon)s\| \|\mathcal{Q}(\tilde{\Lambda}) + \mathcal{H}(\tilde{\Upsilon})s\|^{l-2} \|U^i\| \\ & = \|U^i\| \|\mathcal{Q}(\Lambda_n - \Lambda) + \mathcal{H}(\Upsilon_n - \Upsilon)s\| (ts)^2 \sum_{l=0}^{m_n} \frac{(ts)^{l-2}}{l!} (l-1)l \|\mathcal{Q}(\tilde{\Lambda}) + \mathcal{H}(\tilde{\Upsilon})s\|^{l-2} \\ & = \|U^i\| \|\mathcal{Q}(\Lambda_n - \Lambda) + \mathcal{H}(\Upsilon_n - \Upsilon)s\| (ts)^2 \sum_{l=2}^{m_n} \frac{(ts)^{l-2}}{(l-2)!} \|\mathcal{Q}(\tilde{\Lambda}) + \mathcal{H}(\tilde{\Upsilon})s\|^{l-2} \\ & = \|U^i\| \|\mathcal{Q}(\Lambda_n - \Lambda) + \mathcal{H}(\Upsilon_n - \Upsilon)s\| (ts)^2 \sum_{v=0}^{m_n+2} \frac{(ts)^v}{v!} \|\mathcal{Q}(\tilde{\Lambda}) + \mathcal{H}(\tilde{\Upsilon})s\|^v. \end{aligned}$$

Al igual que antes, sólo bastaría probar que $\sum_{v=0}^{m_n+2} \frac{(ts)^v}{v!} \|\mathcal{Q}(\tilde{\Lambda}) + \mathcal{H}(\tilde{\Upsilon})s\|^v$ está acotada. En efecto, por (2.62) toda sucesión convergente está acotada, por lo tanto existe $M > 0$ tal que

$$\left\| \mathcal{Q}(\tilde{\Lambda}) + \mathcal{H}(\tilde{\Upsilon})s \right\| \leq M,$$

y resulta que

$$\sum_{v=0}^{m_n+2} \frac{(ts)^v}{v!} \|\mathcal{Q}(\tilde{\Lambda}) + \mathcal{H}(\tilde{\Upsilon})s\|^{l-2} \leq \sum_{v=0}^{m_n+2} \frac{(ts)^v}{v!} M^v = \sum_{v=0}^{m_n+2} \frac{(tsM)^v}{v!} \leq e^{tsM} = C^*,$$

lo que concluye la demostración. \square

Corolario 2.17 *Tomando $\alpha(s, t)$, $\alpha^{(n)}(s, t)$ y σ_n^2 definidas como en (2.2), (2.35) y (2.51) respectivamente, se tiene que*

$$\frac{\sqrt{n} (\alpha^{(n)}(s, t) - \alpha(s, t))}{\sigma_n^2} \xrightarrow[n \rightarrow \infty]{w} N(0, 1). \quad (2.66)$$

Demostración:

En el **Teorema 2.9**, se probó que

$$\sqrt{n} (\alpha^{(n)}(s, t) - \alpha(s, t)) \xrightarrow[n \rightarrow \infty]{} N(0, \sigma^2),$$

o equivalentemente

$$\frac{\sqrt{n} (\alpha^{(n)}(s, t) - \alpha(s, t))}{\sigma^2} \xrightarrow[n \rightarrow \infty]{} N(0, 1).$$

Además en el **Teorema 2.16** se probó que σ_n^2 es un estimador consistente de σ^2 , de donde se obtiene lo que queríamos probar, es decir que

$$\frac{\sqrt{n} (\alpha^{(n)}(s, t) - \alpha(s, t))}{\sigma_n^2} \xrightarrow[n \rightarrow \infty]{} N(0, 1).$$

\square

Ahora estamos en condiciones de encontrar cual será el intervalo de confianza para el ancho de banda efectivo.

Proposición 2.18 *Tomando $\alpha(s, t)$, $\alpha^{(n)}(s, t)$ y σ_n^2 definidas como en (2.2), (2.35) y (2.51) respectivamente y considerando el intervalo*

$$I_\alpha(n) = \left[\alpha^{(n)}(s, t) - \frac{z_\epsilon \sigma_n}{\sqrt{n}}, \alpha^{(n)}(s, t) + \frac{z_\epsilon \sigma_n}{\sqrt{n}} \right], \quad (2.67)$$

donde z_ϵ es tal que $P(Z > z_\epsilon) = \frac{\epsilon}{2}$ con $Z \sim N(0, 1)$, resulta que

$$\lim_{n \rightarrow \infty} P(\alpha(s, t) \in I_\alpha(n)) = 1 - \epsilon.$$

Demostración:

Por la definición de $I_\alpha(n)$ se tiene que

$$P(\alpha(s, t) \in I_\alpha(n)) = P\left(\alpha^{(n)}(s, t) - \frac{z_\epsilon \sigma_n}{\sqrt{n}} \leq \alpha(s, t) \leq \alpha^{(n)}(s, t) + \frac{z_\epsilon \sigma_n}{\sqrt{n}}\right) \quad (2.68)$$

$$= P\left(\frac{\sqrt{n}}{\sigma_n} |\alpha^{(n)}(s, t) - \alpha(s, t)| \leq z_\epsilon\right), \quad (2.69)$$

y por el **Corolario 2.17**, se tiene que

$$\lim_{n \rightarrow \infty} P\left(\frac{\sqrt{n}}{\sigma_n} |\alpha^{(n)}(s, t) - \alpha(s, t)| \leq z_\epsilon\right) = P(|Z| \leq z_\epsilon) = 1 - \epsilon,$$

entonces

$$\lim_{n \rightarrow \infty} P(\alpha(s, t) \in I_\alpha(n)) = 1 - \epsilon.$$

□

Capítulo 3

Punto operacional

3.1. Introducción

Las redes de telecomunicación fueron modeladas históricamente utilizando herramientas de la teoría de colas clásica. Desde hace algunos años diversos investigadores comenzaron a utilizar otras herramientas estadísticas para abordar el problema del modelado y el análisis de performance de redes de telecomunicaciones, basándose en que muchos de los fenómenos que interesa estudiar en un enlace de una red son lo que podríamos llamar “eventos raros”. Cuando decimos raros hacemos referencia a eventos tales como la pérdida de paquetes de información transferida en un enlace de red cuya probabilidad real puede tomar valores del orden de 10^{-6} o menos.

Estos fenómenos están asociados a las colas de las distribuciones de probabilidad de interés, como por ejemplo a la distribución de probabilidad de la ocupación del buffer del enlace. Dichos eventos son grandes desvíos respecto de los valores medios de las mencionadas distribuciones de probabilidad. Por este motivo, es la teoría de grandes desvíos (desviaciones), [29], la que proporciona las herramientas para el análisis de redes de telecomunicación. Esta teoría busca encontrar la velocidad con que tiende a cero la probabilidad de un evento, por ejemplo la velocidad con que tiende a cero la probabilidad de pérdida en un enlace cuando el tamaño del buffer tiende a infinito.

La teoría de grandes desvíos tiene la ventaja que los resultados obtenidos son en general analíticamente manejables para colas de tipo muy general, sin embargo tienen limitaciones en cuanto a la precisión de los resultados que se obtienen cuando se estima, por ejemplo, la probabilidad de pérdida a partir de esos resultados.

Podemos decir que la teoría de grandes desvíos dice cual es la asíntota, a escala logarítmica, con la que tiende a cero la probabilidad de pérdida y sin embargo una estimación de la pérdida a partir de esta asíntota a escala logarítmica comete errores que pueden, en algunos casos, llegar a ser importantes.

En la aplicación de grandes desvíos a las telecomunicaciones se han estudiado dos tipos de regímenes, el régimen asintótico de “buffer grande” y el régimen asintótico “de muchas fuentes”. El régimen asintótico de buffer grande estudia el comportamiento de un enlace cuando el tamaño del buffer tiende a infinito. La asíntota de muchas fuentes

estudia el comportamiento cuando el número de fuentes tiende a infinito, y se escala proporcionalmente al número de fuentes, la capacidad y el tamaño del buffer del enlace.

En un backbone de internet, que es el interés principal de nuestro trabajo, es mucho más adecuado el modelo de la asintótica de muchas fuentes ya que en el corazón de internet es más razonable suponer que arriban muchas fuentes a un enlace de la red que asumir que el buffer del enlace tiene un tamaño muy grande, sin embargo en un enrutamiento de acceso puede ser más adecuado el enfoque de buffer grande.

En el estudio de redes de datos, dependiendo del tipo de aplicación, la probabilidad de pérdida de paquetes de información puede ser el índice de performance más restrictivo del cual depende la viabilidad de brindar determinado tipo de servicio, por ello es necesario estimarla y estudiar su variación frente a ciertos cambios tales como cambios en la estructura de la red, cambios en el tamaño del buffer, cambios en la capacidad de enlace, entre otros.

El creciente interés en estudiar la probabilidad de pérdida se justifica por el aumento de la incidencia de aplicaciones de tiempo real en las redes, lo que ha aumentado el interés por el desarrollo de técnicas de estimación y predicción del funcionamiento de un enlace desde el punto de vista de la calidad de servicio QoS.

En este capítulo mostraremos como se puede emplear el concepto de ancho de banda efectivo en la estimación y cálculo del punto operacional de un enlace dentro de una red digital en el caso particular de que el trabajo total que recibe el sistema es la suma de procesos independiente e idénticamente distribuidos.

Como “punto operacional” nos referimos al par de valores de los parámetros de tiempo t y espacio o multiplexado s , en el que el ancho de banda efectivo da la probabilidad de desborde del buffer. Veremos también que si se tiene un buen estimador del ancho de banda efectivo, podremos obtener un buen estimador del punto operacional que heredará la consistencia de dicho estimador. Así mismo derivaremos fórmulas para calcular intervalos de confianza o bien estimarlos en forma empírica.

La mayoría de los estimadores del ancho de banda efectivo, como el estimador promedio presentado por Courcoubetis en [23] ó el del modelo de flujo markoviano presentado en [55] cumplen las condiciones para que el punto operacional y su intervalo de confianza puedan ser bien estimados.

Nosotros demostraremos que nuestro estimador, introducido para el modelo propuesto en el capítulo anterior, también cumple estas condiciones.

Este capítulo está organizado de la siguiente manera, en la Sección 3.2 presentamos los resultados fundamentales de la teoría de grandes desvíos que utilizaremos en la Sección 3.3 donde presentamos el concepto de punto operacional, sus propiedades y ejemplos de cálculo, en la Sección 3.4 se introducen los resultados fundamentales que permitirán encontrar un estimador consistente del punto operacional para el modelo de Flujo Markoviano Generalizado introducido en el capítulo anterior.

3.2. Teoría de grandes desvíos

Consideremos un sistema multiplexor donde el trabajo arriba desde N fuentes independientes y todas con idéntica distribución de probabilidades y cada una con tamaño de buffer b y capacidad de enlace c , entonces el tamaño total del buffer es igual a Nb y la capacidad total de enlace es igual a Nc como se muestra en la figura (3.1)



Figura 3.1: Régimen de muchas fuentes

Más formalmente, sea $X_t^{(j)}$ una sucesión de procesos estocásticos independientes e idénticamente distribuidos y con valor inicial igual a cero, que representa el trabajo acumulado que se recibe desde cada fuente. El trabajo total que recibe el sistema para $t > 0$ es entonces,

$$X_t = \sum_{j=1}^N X_t^{(j)},$$

y la cantidad de trabajo del sistema con N fuentes en estado estacionario se puede expresar como

$$W_t^N = X_t - Ct = \sum_{j=1}^N (X_t^{(j)} - ct), \quad (3.1)$$

donde c es la capacidad de cada fuente del servidor y $C = Nc$ es la capacidad total.

Para analizar cual sería el tamaño de la cola, tengamos en cuenta que si consideramos una cola del tipo “first in first out” (FIFO), el comportamiento de la cola es gobernado por la ecuación de Lindley, [29],

$$Q_{n+1} = (Q_n + X_n)^+,$$

donde Q_n para $n \in \mathbf{Z}$, denota el tamaño de la cola al instante n , X_n denota la diferencia entre la cantidad de trabajo que arriba en el instante n y la cantidad de trabajo que puede ser procesado en dicho instante y $(\cdot)^+$ es el máximo entre (\cdot) y 0.

La pregunta que surge es si cuando n crece este proceso tendrá límite, es decir en qué casos el trabajo de la cola “en régimen” tendrá una distribución estacionaria y cuando esta distribución será independiente del estado inicial de la cola.

Estas preguntas las responde el teorema de Lyones que presentamos a continuación y cuya demostración puede encontrarse en [7]. La formulación que presentamos de dicho teorema no es la original pero es de importancia para nuestro trabajo.

Teorema 3.1 (Lyones) *Si se cumple que el proceso X_t es estacionario, ergódico y tal que $E(X_t) < 0$, entonces para cualquier condición inicial de la cola Q , se verifica que*

$$\lim_{n \rightarrow \infty} P(Q_n \leq x) = P(Q_0 \leq x), \quad (3.2)$$

donde $Q_0 = \sup_{n > 0} \sum_{i=-n}^{-1} X_i$ es casi seguramente finita.

A partir de este resultado se puede estudiar la probabilidad de overflow de un enlace mirando la distribución en régimen de la cola.

En nuestro caso donde la cantidad de trabajo del sistema está definida en (3.1), el tamaño de la cola según el **Teorema 3.1** es

$$Q_N = \sup_{t \geq 0} W_t^N,$$

y buscaremos una estimación para $P(Q_N > B) = P(Q_N > Nb)$ donde b es el tamaño del buffer disponible para cada fuente y $B = Nb$ es el tamaño total del buffer. Esta cantidad da una idea del valor de la probabilidad de que el sistema supere su capacidad de almacenamiento y por lo tanto haya pérdida de datos u overflow.

La idea básica de la teoría de grandes desvíos es poder establecer resultados asintóticos de la forma $P(X_n \in \mathcal{B}) \approx e^{nI(\mathcal{B})}$ para una sucesión de variables aleatorias $\{X_n\}_{n \in \mathbb{N}}$ cuando $n \rightarrow \infty$, para un boreliano $\mathcal{B} \in \mathbb{R}$ y donde $I(\mathcal{B})$ es un coeficiente que depende del conjunto \mathcal{B} y de la distribución de las variables X_n .

Comencemos por definir más formalmente estos conceptos.

Definición 3.2 *Una función de velocidad I es una función $I : \mathbb{R} \rightarrow [0, \infty]$ semicontinua inferior, esto es que para cualquier $\alpha \in \mathbb{R}$ el conjunto de nivel $\Psi_I(\alpha) = \{x \in \mathbb{R} : I(x) \leq \alpha\}$ es cerrado.*

Diremos que I es una buena función de velocidad si además sus conjuntos de nivel $\Psi_I(\alpha)$ son compactos para todo $\alpha \in \mathbb{R}$.

Además, llamamos dominio efectivo de I al conjunto de puntos de velocidad finita, es decir al conjunto $\mathcal{D}_I = \{x \in \mathbb{R} : I(x) < \infty\}$

Definición 3.3 *Diremos que una sucesión de variables aleatorias $\{X_n\}_{n \in \mathbb{N}}$ satisface un Principio de Grandes Desviaciones si existe una función de velocidad $I : \mathbb{R} \rightarrow \mathbb{R}^+$ para la cual se cumple*

$$-\inf_{x \in \mathcal{B}^\circ} I(x) \leq \liminf_{n \rightarrow \infty} \frac{1}{n} \log P(X_n \in \mathcal{B}) \leq \limsup_{n \rightarrow \infty} \frac{1}{n} \log P(X_n \in \mathcal{B}) \leq -\inf_{x \in \bar{\mathcal{B}}} I(x), \quad (3.3)$$

donde \mathcal{B} es un boreliano en \mathbb{R} .

En general, la ecuación (3.3) no implica un límite preciso, sin embargo si los extremos de la desigualdad coinciden, es decir existe un boreliano \mathcal{B} que es conjunto de continuidad de I , se deduce de dicha ecuación que podemos extender la definición de la función I a los borelianos de la siguiente manera

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log P(X_n \in \mathcal{B}) = I(\mathcal{B}),$$

y se tiene que, para n suficientemente grande, es válida la aproximación

$$P(X_n \in \mathcal{B}) \approx e^{nI(\mathcal{B})}. \quad (3.4)$$

Intuitivamente podemos interpretar la función de velocidad como una función que indica el orden o tasa con la que decae la probabilidad del conjunto. Un principio de grandes desviaciones establece que dichas probabilidades decaen exponencialmente y que dicha tasa de decaimiento en cada conjunto está determinada por aquel punto del conjunto que tenga una tasa de decaimiento más lenta, es decir la “menor” de todas la $I(x)$, con $x \in \mathcal{B}$.

Una forma equivalente de establecer un principio de grandes desviaciones es presentado en la siguiente proposición.

Proposición 3.4 *La sucesión de variables aleatorias $\{X_n\}_{n \in \mathbb{N}}$ satisface un principio de grandes desviaciones si y sólo si existe una función de velocidad $I : \mathbb{R} \rightarrow \mathbb{R}^+$ para la cual se cumple*

$$\lim_{n \rightarrow \infty} \sup \frac{1}{n} \log P(X_n \in F) \leq - \inf_{x \in F} I(x) \quad \forall F \text{ cerrado} \quad (3.5)$$

y

$$\lim_{n \rightarrow \infty} \inf \frac{1}{n} \log P(X_n \in F) \geq - \inf_{x \in G} I(x) \quad \forall G \text{ abierto}. \quad (3.6)$$

Demostración:

Demostremos primero que si las variables cumplen un principio de grandes desvíos se cumplen (3.5) y (3.6). En efecto, si F es cerrado, tomando $\mathcal{B} = F$ en (3.3) se obtiene (3.5) ya que $F = \bar{F}$. Análogamente se obtiene (3.6) tomando $\mathcal{B} = G$ en (3.3) pues al ser G abierto, $G = G^\circ$.

Para la otra implicación, si \mathcal{B} es un boreliano de \mathbb{R} , se cumple que $\mathcal{B}^\circ \subset \mathcal{B} \subset \bar{\mathcal{B}}$ y por lo tanto

$$P(X_n \in \mathcal{B}^\circ) \leq P(X_n \in \mathcal{B}) \leq P(X_n \in \bar{\mathcal{B}}),$$

y como la función logaritmo es creciente y $n > 0$ se tiene que

$$\frac{1}{n} \log P(X_n \in \mathcal{B}^\circ) \leq \frac{1}{n} \log P(X_n \in \mathcal{B}) \leq \frac{1}{n} \log P(X_n \in \bar{\mathcal{B}}).$$

Tomando límites se tiene

$$\begin{aligned} \liminf_{n \rightarrow \infty} \frac{1}{n} \log P(X_n \in \mathcal{B}^\circ) &\leq \liminf_{n \rightarrow \infty} \frac{1}{n} \log P(X_n \in \mathcal{B}) \\ &\leq \limsup_{n \rightarrow \infty} \frac{1}{n} \log P(X_n \in \mathcal{B}) \\ &\leq \limsup_{n \rightarrow \infty} \frac{1}{n} \log P(X_n \in \bar{\mathcal{B}}), \end{aligned}$$

y finalmente se obtiene (3.3) utilizando las ecuaciones (3.5) y (3.6) en los conjuntos \mathcal{B}° y $\bar{\mathcal{B}}$. \square

Uno de los primeros resultados de la teoría es el establecer un principio de grandes desviaciones para el promedio de variables aleatorias independientes e idénticamente distribuidas lo que nos permitiría determinar la velocidad de la convergencia de la sucesión de variables $\{\bar{X}_n\}$ a μ , asegurada por la *ley débil de los grandes números*. El resultado se conoce como Teorema de Crámer, lo presentamos a continuación y luego lo aplicaremos a nuestro proceso. Su demostración puede encontrarse en [29].

Teorema 3.5 (*Cramér*) Sea $(X_n)_{n \in \mathbb{N}}$ una sucesión de variables aleatorias independientes e idénticamente distribuidas y $\bar{X}_n = \frac{1}{n} \sum_{j=1}^n X_j$ el promedio de dichas variables. Sea μ_n la medida inducida en \mathbb{R} por la variable aleatoria \bar{X}_n , es decir $\mu_n(\mathcal{B}) = P(\bar{X}_n \in \mathcal{B})$. Se cumple entonces que la familia $\{\mu_n\}$ satisface el siguiente principio de grandes desviaciones

$$-\inf_{x \in \Gamma^\circ} \Delta^*(x) \leq \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(\Gamma) \leq \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(\Gamma) \leq -\inf_{x \in \Gamma} \Delta^*(x), \quad (3.7)$$

donde Γ es un boreliano en \mathbb{R} , $\Delta^*(x) = \sup_{s \in \mathbb{R}} \{sx - \Delta(s)\}$ y $\Delta(s) = \log E e^{sX_t}$ es el logaritmo de la función generadora de momentos común a todas las variables X_j y se denomina función de velocidad.

Observación 3.6

El Teorema de Crámer se puede generalizar a variables no independientes ni idénticamente distribuidas mediante el Teorema de Gärtner-Ellis, [29].

Observación 3.7

Si consideramos el boreliano $\mathcal{B} = [a, +\infty]$ con $a \geq E(X)$, como la función no negativa Δ^* es convexa y alcanza su mínimo en $E(X)$ se tiene que

$$\inf_{x \in \mathcal{B}^\circ} \Delta^*(x) = \inf_{x \in \bar{\mathcal{B}}} \Delta^*(x) = \Delta^*(a),$$

y por lo tanto

$$\lim_{n \rightarrow \infty} \frac{\log P(\bar{X}_n \geq a)}{n} = -\Delta^*(a),$$

o bien, para n suficientemente grande, si $a \geq E(X)$, se puede aproximar la probabilidad como

$$P(\bar{X}_n \geq a) \approx e^{n\Delta^*(a)}.$$

Aplicando el resultado del teorema anterior a las variables W_t^N definidas en (3.1), y que por el **Teorema 3.1** determinan el tamaño de la cola, se obtiene la siguiente proposición

Proposición 3.8 Sean las variables aleatorias W_t^N definidas como en (3.1) con t fijo. La sucesión $\frac{1}{N}W_t^N$ verifica un principio de grandes desviaciones con función de velocidad $\Delta_t^*(x) = \sup_{s \geq 0} \{s(x + ct) - \log Ee^{sX_t}\}$. En particular, si $b + ct > E(X_t)$ se cumple que

$$-\Delta_t^*(b^+) \leq \lim_{N \rightarrow \infty} \inf \frac{1}{N} \log P(W_t^N > Nb) \leq \lim_{N \rightarrow \infty} \sup \frac{1}{N} \log P(W_t^N > Nb) \leq -\Delta_t^*(b). \quad (3.8)$$

Demostración:

La demostración consiste en aplicar el **Teorema 3.5** e identificar los términos que aparecen en la ecuación (3.7).

Sabemos que $W_t^N = \sum_{j=1}^N Y_t^{(j)}$, con $Y_t^{(j)} = (X_t^{(j)} - ct)$ es una sucesión de variables aleatorias independientes e idénticamente distribuidas, ya que las variables $X_t^{(j)}$ también lo son. Entonces aplicando el **Teorema 3.5** a las variables $Y_t^{(j)}$ con t fijo y observando que $\bar{Y}_t^{(j)} = \frac{1}{N}W_t^N$ se tiene que

$$-\inf_{x \in \Gamma^\circ} \Delta^*(x) \leq \lim_{N \rightarrow \infty} \inf \frac{1}{N} \log P\left(\frac{1}{N}W_t^N \in \Gamma\right) \leq \lim_{N \rightarrow \infty} \sup \frac{1}{N} \log P\left(\frac{1}{N}W_t^N \in \Gamma\right) \leq -\inf_{x \in \bar{\Gamma}} \Delta^*(x). \quad (3.9)$$

Se puede probar además que si $\Delta(s) = \log Ee^{sY_t} < \infty$ para algún $s > 0$, entonces $E(Y_t) < \infty$ y para cualquier $x \geq E(Y_t)$, la expresión $\Delta^*(x)$ se puede escribir como $\Delta^*(x) = \sup_{s \geq 0} \{sx - \Delta(s)\}$ que es no decreciente en $x > E(Y_t)$. Entonces tomando el boreliano $\Gamma = (b, +\infty)$, con $b > E(Y_t) = E(X_t - ct)$, se tiene que

$$\inf_{x \in \bar{\Gamma}} \Delta_t^*(x) = \inf_{x \geq b} \Delta_t^*(x) = \Delta_t^*(b), \quad (3.10)$$

$$\inf_{x \in \Gamma^\circ} \Delta_t^*(x) = \inf_{x > b} \Delta_t^*(x) = \Delta_t^*(b^+), \quad (3.11)$$

donde

$$\begin{aligned} \Delta_t^*(x) &= \sup_{s \geq 0} \{sx - \Delta(s)\} \\ &= \sup_{s \geq 0} \{sx - \log E(e^{sY_t})\} \\ &= \sup_{s \geq 0} \{sx - \log E(e^{s(X_t - ct)})\} \\ &= \sup_{s \geq 0} \{sx + sct - \log E(e^{sX_t})\} \\ &= \sup_{s \geq 0} \{s(x + ct) - \log E(e^{sX_t})\}. \end{aligned}$$

Sustituyendo (3.10) y (3.11) en (3.9) y observando que $P(\frac{1}{N}W_t^N \in \Gamma) = P(W_t^N > Nb)$ se obtiene el resultado buscado. \square

Corolario 3.9 En las hipótesis de la proposición anterior si $\Delta_t^*(x)$ es continua en b , se tiene que

$$\lim_{N \rightarrow \infty} \frac{1}{N} \log P(W_t^N > Nb) = -\Delta_t^*(b) = -\sup_{s > 0} \{s(b + ct) - \log E(e^{sX_t})\}.$$

Veamos entonces la relación entre la probabilidad de pérdida y el ancho de banda efectivo.

Teorema 3.10 Sean las variables aleatorias W_t^N definidas según la ecuación (3.1) y $Q_N = \sup_{t \geq 0} W_t^N$ el tamaño de la cola en estado estacionario del sistema. Supongamos que existe un valor positivo s_0 que cumple $\alpha(s_0, t) \leq c' < c$ para todo $t \geq t_0$ y que $\limsup_{\epsilon \rightarrow 0} \log E(sX_{t,\epsilon}) = 0$, con $X_{t,\epsilon} = \sup_{0 \leq t \leq \epsilon} X_t$. Entonces se cumple que

$$-I(b^+) \leq \lim_{N \rightarrow \infty} \inf \frac{1}{N} \log P(Q_N > Nb) \leq \lim_{N \rightarrow \infty} \sup \frac{1}{N} \log P(Q_N > Nb) \leq -I(b), \quad (3.12)$$

donde

$$I(x) = \inf_{t \geq 0} \sup_{s \geq 0} \{s(x + ct) - \log E(e^{sX_t})\}. \quad (3.13)$$

Observación 3.11

La expresión (3.13) se puede escribir en forma equivalente como

$$I(x) = \inf_{t \geq 0} \sup_{s \geq 0} \{s(x + ct) - st\alpha(s, t)\}. \quad (3.14)$$

3.3. Punto operacional

Como ya hemos visto, en la asintótica de muchas fuentes, el ancho de banda efectivo está relacionado con la probabilidad de pérdida estacionaria por desborde del buffer en un enlace a través de la llamada fórmula del ínfimo-supremo (*inf sup*) que da el **Teorema 3.10**, ya que si Q_N es la cantidad de trabajo estacionario en la cola, la probabilidad de overflow o de desborde está dada por

$$\lim_{N \rightarrow \infty} \frac{1}{N} \log P(Q_N > B) = -\gamma, \quad (3.15)$$

con

$$-\gamma = \inf_{t \geq 0} \sup_{s \geq 0} \{(b + ct)s - st\alpha(s, t)\}, \quad (3.16)$$

donde c es la capacidad de cada enlace, b es el tamaño o capacidad de cada buffer y N es la cantidad de fuentes que intervienen en el enlace y tal que su ancho de banda efectivo es igual a $\alpha(s, t)$, tal como lo demuestra Courcourbetis en [23].

Llamaremos t^* y s^* a los valores numéricos de los parámetros t y s en que se alcanza el *inf sup* de la ecuación (3.16). Ese par de valores, t^* y s^* , constituye el llamado *punto operacional* de un determinado enlace.

Resulta técnicamente relevante disponer de buenas estimaciones de t^* y s^* dado que la probabilidad de pérdida es uno de los criterios más importantes tanto para el diseño de las redes como para el control de admisión de nuevas conexiones (CAC).

Cuando existe un modelo para la fuente de tráfico es posible un enfoque paramétrico, por eso veamos primero algunos ejemplos del cálculo del punto operacional para dos de los modelos introducidos en el primer capítulo.

Ejemplo 3.12 *Punto operacional en el proceso de Poisson.*

Como vimos en el **Ejemplo 1.10**, para este modelo de tráfico el ancho de banda efectivo está dado por

$$\alpha(s, t) = \frac{1}{st} \log e^{\lambda t(e^s - 1)} = \frac{1}{st} (\lambda t(e^s - 1)) = \lambda \frac{(e^s - 1)}{s},$$

entonces para determinar el punto operacional cuando tenemos una única fuente debemos optimizar la ecuación (3.16) que en este caso en particular tendrá la expresión

$$\inf_{t \geq 0} \sup_{s \geq 0} \{(b + ct)s - t\lambda(e^s - 1)\}.$$

Para t fijo, consideremos la función $f(s) = (b + ct)s - t\lambda(e^s - 1)$, entonces

$$\frac{df(s)}{ds} = b + ct - t\lambda e^s, \quad (3.17)$$

igualando a 0 obtenemos que $s^* = \log\left(\frac{b+ct}{t\lambda}\right)$. Notemos que para que la cola sea estable debe cumplirse que $c > \lambda$, esto hace al argumento del logaritmo de la ecuación (3.17) mayor que 1, y a $s^* \geq 0$.

Calculamos la segunda derivada de la función $\frac{d^2 f(s)}{ds^2} = -t\lambda e^s$ y al evaluarla en s^* vemos que es negativa por lo tanto s^* maximiza la función f .

Consideremos ahora para realizar la optimización respecto de t la función

$$\begin{aligned} g(t) &= s^*(b + ct) - t\lambda(e^{s^*} - 1) \\ &= -b - (c - \lambda)t + (b + ct) \log\left(\frac{b}{\lambda t} + \frac{c}{\lambda}\right), \end{aligned}$$

esta ecuación puede ser minimizada analíticamente, pero se obtiene un valor de t negativo, por lo se minimiza numéricamente para encontrar t^* .

Ejemplo 3.13 *Punto operacional para una fuente Gaussiana.*

Como vimos en el **Ejemplo 1.15**, para este modelo de tráfico el ancho de banda efectivo está dado por

$$\alpha(s, t) = \frac{1}{st} \log Ee^{sXt} = \frac{1}{st} \left(s\lambda t + \frac{1}{2}s^2\sigma^2 t \right) = \lambda + \frac{1}{2}s\sigma^2,$$

entonces para determinar el punto operacional cuando tenemos una única fuente debemos optimizar la ecuación (3.16) que en este caso en particular tendrá la expresión

$$\inf_{t \geq 0} \sup_{s \geq 0} \left\{ (b + ct)s - st \left(\lambda + \frac{1}{2}s\sigma^2 \right) \right\}. \quad (3.18)$$

Para t fijo, consideremos la función $f(s) = (b + ct)s - st(\lambda + \frac{1}{2}s\sigma^2)$, entonces

$$\frac{df(s)}{ds} = b + t(c - \lambda - s\sigma^2), \quad (3.19)$$

igualando a 0 obtenemos que $s^* = \frac{b+t(c-\lambda)}{t\sigma^2}$ que es positivo y es efectivamente un mínimo ya que $\frac{d^2f(s)}{ds^2} = -t\sigma^2$ es negativa para cualquier valor de s .

Sustituyendo s^* en (3.18) obtenemos la ecuación

$$\inf_{t \geq 0} \frac{b + t(c - \lambda)}{t\sigma^2}(b + ct) - \frac{b + t(c - \lambda)}{t\sigma^2}t \left(\lambda + \left(\frac{b + t(c - \lambda)}{t\sigma^2} \right) \frac{\sigma^2}{2} \right),$$

que al simplificarla se convierte en

$$\inf_{t \geq 0} \frac{(b + t(c - \lambda))^2}{2t\sigma^2}.$$

Para realizar la optimización respecto de t consideremos la función $g(t) = \frac{(b+t(c-\lambda))^2}{2t\sigma^2}$, derivando obtenemos

$$\frac{dg(t)}{dt} = \frac{-b^2 + (c - \lambda)^2 t^2}{2t^2\sigma^2},$$

y al igualar la expresión a 0, la solución positiva es

$$t^* = \frac{b}{c - \lambda},$$

y

$$s^* = \frac{2(c - \lambda)}{\sigma^2}.$$

3.4. Estimación del punto operacional para el Flujo Markoviano Generalizado

En el caso del flujo markoviano, el proceso se modela mediante una cadena de Markov de tiempo continuo, donde el estado de la cadena determina la velocidad a la cual produce trabajo la fuente y Kesidis, Walrand y Chang muestran en [42] que en este caso el ancho de banda efectivo se puede calcular explícitamente en términos del generador infinitesimal de la cadena, pudiéndose hallar un estimador consistente para el ancho de banda efectivo así como un intervalo de confianza.

Para el modelo de flujo markoviano generalizado introducido en la **Sección 2.2** del presente trabajo, vimos que también podemos calcular explícitamente el ancho de banda efectivo en términos del generador infinitesimal de la cadena y de la velocidad media a la cual produce trabajo la fuente [49] y además hallamos un estimador consistente y un intervalo de confianza para el ancho de banda efectivo.

Dado un estimador de $\alpha(s, t)$, deseamos estimar γ y el punto operacional (s^*, t^*) . Una manera de hacerlo es sustituyendo $\alpha(s, t)$ por su estimador $\alpha^{(n)}(s, t)$ en la ecuación (3.16). Resolviendo el problema de optimización correspondiente se obtienen como resultado los valores de γ_n y (s_n^*, t_n^*) , entonces la pregunta es bajo qué condiciones estos valores son buenos estimadores de los verdaderos valores γ y (s^*, t^*) .

Entonces planteamos dos problemas diferentes para la estimación. El primero consiste en, dado un buen estimador $\alpha^{(n)}(s, t)$ del ancho de banda efectivo $\alpha(s, t)$, encontrar condiciones necesarias bajo las cuales los estimadores γ_n y (s_n^*, t_n^*) obtenidos a partir de la resolución de la ecuación *inf sup* en (3.16) son buenos estimadores del punto operacional (s, t) y de la tasa de decaimiento de la probabilidad de overflow γ . Esta afirmación no es obvia porque γ_n , s_n^* y t_n^* deben estimarse a partir de una función no lineal e implícita. El segundo problema consiste en encontrar estos buenos estimadores del ancho de banda efectivo y determinar cuándo se verifican las condiciones para que el punto operacional se pueda estimar usando la ecuación (3.16).

Para el modelo de flujo markoviano, el primer problema fue tratado en [5] donde se presenta una respuesta que incluye propiedades de consistencia y un teorema central del límite para los estimadores, basados en condiciones de regularidad del ancho de banda efectivo. Nuestro objetivo es encontrar resultados similares para el modelo de flujo markoviano generalizado presentado en la **Sección 2.2**, para esto presentamos el siguiente teorema que atiende el primer problema.

Consideremos la función $g(s, t) = s(b + ct) - s\alpha(s, t)$, si llamamos $\Lambda(s, t) = E(e^{sXt})$ podemos escribirla como

$$g(s, t) = s(b + ct) - \log \Lambda(s, t).$$

Bajo hipótesis de regularidad de la función $\Lambda(s, t)$ demostraremos que es posible expresar el punto operacional (s^*, t^*) como función $\Lambda(s, t)$ aplicando el teorema de la función implícita, [69]. De forma análoga se obtiene un estimador (s_n^*, t_n^*) como función del estimador $\Lambda_n(s, t)$ de $\Lambda(s, t)$ y agregando hipótesis sobre el estimador $\Lambda_n(s, t)$ será posible demostrar propiedades de consistencia y teorema central del límite para (s_n^*, t_n^*) .

Teorema 3.14 *Sea $\alpha^{(n)}(s, t)$ un estimador de $\alpha(s, t)$, ambas funciones de clase \mathcal{C}^1 , es decir continuas y con derivada primera continua tal que*

$$\alpha^{(n)}(s, t) \xrightarrow[n \rightarrow \infty]{c.s.} \alpha(s, t), \quad (3.20)$$

$$\frac{\partial}{\partial s} \alpha^{(n)}(s, t) \xrightarrow[n \rightarrow \infty]{c.s.} \frac{\partial}{\partial s} \alpha(s, t), \quad (3.21)$$

$$\frac{\partial}{\partial t} \alpha^{(n)}(s, t) \xrightarrow[n \rightarrow \infty]{c.s.} \frac{\partial}{\partial t} \alpha(s, t), \quad (3.22)$$

uniformemente en intervalos acotados. Dadas la probabilidad de pérdida γ , la capacidad del enlace c y el tamaño del buffer b , si s_n^ y t_n^* son las soluciones de*

$$\frac{\partial}{\partial s} \alpha^{(n)}(s_n^*, t_n^*) - \frac{\gamma}{s_n^{*2} t_n^*} = 0, \quad (3.23)$$

$$\frac{\partial}{\partial t} \alpha^{(n)}(s_n^*, t_n^*) - \frac{\gamma}{s_n^* t_n^{*2}} - \frac{b}{t_n^{*2}} = 0, \quad (3.24)$$

entonces (s_n^*, t_n^*) es un estimador consistente de (s^*, t^*) . Además si vale un teorema central del límite funcional para $\alpha^{(n)}(s, t)$, esto es si

$$\sqrt{n} (\alpha^{(n)}(s, t) - \alpha(s, t)) \xrightarrow[n \rightarrow \infty]{w} \mathcal{G}(s, t), \quad (3.25)$$

donde $\mathcal{G}(s, t)$ es un proceso gaussiano continuo, entonces

$$\sqrt{n} ((s_n^*, t_n^*) - (s^*, t^*)) \xrightarrow[n \rightarrow \infty]{w} \mathcal{N}(\mathbf{0}, \Sigma), \quad (3.26)$$

con $\mathcal{N}(\mathbf{0}, \Sigma)$ denotando a la distribución normal bivariada, centrada y con matriz de covarianza Σ .

Recordemos algunos conceptos que necesitaremos en el desarrollo de la demostración. En primer lugar, si E y F son espacios normados o seminormados, una función $f : E \rightarrow F$ es diferenciable en $e \in E$ si existe una transformación lineal continua $df(e) : E \rightarrow F$ tal que para todo e' en un entorno de e se verifica que

$$f(e') = f(e) + df(e)(e' - e) + o(\|e' - e\|).$$

En segundo lugar, si E_1 , E_2 y F son espacios normados y $f : E_1 \times E_2 \rightarrow F$ es diferenciable, también lo son para cada $e_1 \in E_1$ y $e_2 \in E_2$, las funciones $f_{e_1} : E_2 \rightarrow F$ definida como $f_{e_1}(e_2) = f(e_1, e_2)$ y $f^{e_2} : E_1 \rightarrow F$ definida como $f^{e_2}(e_1) = f(e_1, e_2)$.

Por último recordemos la formulación general para la derivada de una función implícita. Si f es diferenciable, y para cada $e_2 \in E_2$ existe un único elemento $v(e_2)$ de E_1 tal que $f(v(e_2), e_2) = \vec{0}$ y $df^{e_2}(v(e_2))$ admite inversa, entonces $df^{e_2}(v(e_2))dv(e_2) + df_{v(e_2)}(e_2) = 0$, de donde

$$dv(e_2) = - (df^{e_2}(v(e_2)))^{-1} df_{v(e_2)}(e_2). \quad (3.27)$$

Aplicaremos (3.27) al conjunto $E_1 = \mathbb{R}^{+2}$ con la norma Euclídea y el conjunto $E_2 = \left\{ f : \mathbb{R}^{+2} \rightarrow \mathbb{R} \text{ de clase } C^1 \right\}$ con la seminorma

$$\|f\| = \sum_{n=1}^{\infty} \frac{1}{2^n} \left(\frac{\|f\|_n}{1 + \|f\|_n} \right),$$

donde $\|f\|_n$ es la norma de tipo Sobolev

$$\sup \left\{ |f(s, t)| + \left| \frac{\partial}{\partial s} f(s, t) \right| + \left| \frac{\partial}{\partial t} f(s, t) \right| : (s, t) \in [\delta, n]^2, \delta > 0 \right\}.$$

Con esta norma se cumple que toda sucesión $\{f_m\}$ de funciones de E_2 es tal que $\|f_m\| \rightarrow 0$ cuando $m \rightarrow \infty$ si y sólo si $f_m \rightarrow 0$, $\frac{\partial}{\partial s} f_m \rightarrow 0$, $\frac{\partial}{\partial t} f_m \rightarrow 0$ uniformemente sobre intervalos acotados cuando $m \rightarrow \infty$.

Un resultado del análisis funcional que también utilizaremos lo enunciamos en el siguiente lema, cuya demostración se puede encontrar en [12].

Lema 3.15 Sea X_n y X una sucesión de variables aleatorias y una variable aleatoria respectivamente definidas sobre S y sea $u \in \mathbb{R}^d$. Si la función f es diferenciable en ambas coordenadas con derivadas continuas y acotadas entonces se cumple que

$$f(X_n, u) \xrightarrow[n \rightarrow \infty]{w} f(X, u).$$

Ahora si estamos en condiciones de abordar la demostración,

Demostración del Teorema 3.14: De las ecuaciones (3.21) y (3.22), se tiene que (s^*, t^*) es la solución de la ecuación $K((s, t), \alpha) = \vec{0}$ donde

$$K((s, t), \alpha) = \left(\begin{array}{c} \frac{\partial}{\partial s} \alpha(s, t) - \frac{\gamma}{s^2 t} \\ \frac{\partial}{\partial t} \alpha(s, t) - \frac{\gamma}{s t^2} - \frac{b}{t^2} \end{array} \right), \quad (3.28)$$

y (s_n^*, t_n^*) es la solución de $K((s_n, t_n), \alpha^{(n)}) = \mathbf{0}$.

Como la función K es diferenciable, para cada $\alpha \in E_2$ existe un único elemento (s, t) de E_1 tal que $v(\alpha) = (s, t)$ y podemos pensar que (s^*, t^*) como $(s^*, t^*)(\alpha) = v(\alpha)$, entonces utilizando la notación antes presentada y (3.27) tenemos que $K(v(\alpha), \alpha) = \mathbf{0}$ y

$$dv(\alpha) = - (dK^\alpha(v(\alpha)))^{-1} dK_{v(\alpha)}(\alpha).$$

Como $(s_n^*, t_n^*) = v(\alpha^{(n)})$ tenemos que

$$(s_n^*, t_n^*) - (s^*, t^*) = v(\alpha^{(n)}) - v(\alpha) = dv(\alpha)(\alpha^{(n)} - \alpha) + o(\|\alpha^{(n)} - \alpha\|). \quad (3.29)$$

Por las hipótesis del teorema $\|\alpha^{(n)} - \alpha\| \xrightarrow[n \rightarrow \infty]{c.s} 0$ y $dv(\alpha)$ es continua, por lo tanto está acotada sobre un conjunto compacto y como trabajamos con $s > 0$ y $t > 0$ tenemos que

$$\|(s_n^*, t_n^*) - (s^*, t^*)\| \xrightarrow[n \rightarrow \infty]{} 0, \quad (3.30)$$

con lo que queda demostrado que (s_n^*, t_n^*) es un estimador consistente de (s^*, t^*) .

Utilizando el Lema 3.15, tenemos que $(\alpha^{(n)} - \alpha)$ verifica un teorema central del límite funcional y se cumple que

$$\sqrt{n} (\alpha^{(n)}(s, t) - \alpha(s, t)) \xrightarrow[n \rightarrow \infty]{} \mathcal{G}(s, t), \quad (3.31)$$

donde $\mathcal{G}(s, t)$ es un proceso Gaussiano C^1 , resulta entonces que

$$\sqrt{n} ((s_n^*, t_n^*) - (s^*, t^*)) = dv(\alpha) (\sqrt{n}(\alpha^{(n)} - \alpha)) + o(\sqrt{n}\|\alpha^{(n)} - \alpha\|). \quad (3.32)$$

Como $\sqrt{n}\|\alpha^{(n)} - \alpha\|$ está acotado en probabilidad y entonces $o(\sqrt{n}\|\alpha^{(n)} - \alpha\|) \xrightarrow[n \rightarrow \infty]{\mathcal{P}} 0$.

Por otra parte como $dv(\alpha)$ es continua se verifica que

$$\sqrt{n} dv(\alpha)(\alpha^{(n)} - \alpha) \xrightarrow[n \rightarrow \infty]{w} dv(\alpha),$$

con lo que queda demostrado que

$$\sqrt{n}((s_n^*, t_n^*) - (s^*, t^*)) \xrightarrow[n \rightarrow \infty]{w} dv(\alpha). \quad (3.33)$$

Como $dv(\alpha)$ es una transformación lineal de \mathbb{R}^2 , $dv(\alpha)\mathcal{G}$ es una variable aleatoria $\mathcal{N}(\mathbf{0}, \Sigma)$, donde Σ se puede calcular en términos de la covarianza de \mathcal{G} , $\sigma_{\mathcal{G}}^2$ y de la transformación $dv(\alpha)$ de la siguiente manera

$$\Sigma = \begin{pmatrix} \frac{\partial s}{\partial \alpha} \\ \frac{\partial t}{\partial \alpha} \end{pmatrix} \sigma_{\mathcal{G}}^2 \begin{pmatrix} \frac{\partial s}{\partial \alpha} & \frac{\partial t}{\partial \alpha} \end{pmatrix} \quad (3.34)$$

$$= \sigma_{\mathcal{G}}^2 \begin{pmatrix} \frac{\partial s}{\partial \alpha} & \frac{\partial s}{\partial \alpha} \frac{\partial t}{\partial \alpha} & \frac{\partial s}{\partial \alpha} \frac{\partial t}{\partial \alpha} \\ \frac{\partial t}{\partial \alpha} & \frac{\partial t}{\partial \alpha} \frac{\partial s}{\partial \alpha} & \frac{\partial t}{\partial \alpha} \frac{\partial t}{\partial \alpha} \end{pmatrix}, \quad (3.35)$$

lo que concluye la demostración dando una fórmula explícita para la matriz de covarianza. \square

Observación 3.16

- a) El método usado en la demostración anterior se conoce como método δ . Es utilizado para obtener extensiones de teoremas centrales del límite mediante funcionales diferenciales. Se debe elegir la topología apropiada para garantizar diferenciabilidad ya que no cualquier funcional de un proceso asintóticamente gaussiano es asintóticamente gaussiano.
- b) El cálculo de la matriz de covarianza Σ no es trivial, sin embargo el resultado de la normalidad asintótica siempre permite la estimación de Σ en términos de las covarianzas empíricas, aunque pueda no disponerse de una expresión explícita para la varianza.
- c) Como el **Teorema 3.14** asegura convergencia uniforme sobre intervalos acotados, también asegura que γ_n dada por

$$\gamma_n = s_n^*(b + ct_n^*) - \alpha^{(n)}(s_n^*, t_n^*), \quad (3.36)$$

hereda las propiedades del estimador (s_n^*, t_n^*) , que permite construir intervalos de confianza para la probabilidad de pérdida y utilizando las mismas ideas que en la demostración del teorema se tiene que

$$\sqrt{n}(\gamma_n - \gamma) \xrightarrow[n \rightarrow \infty]{w} \mathcal{N}(0, \sigma^2).$$

En el régimen de muchas fuentes, las expresiones para el tamaño total del buffer B y la capacidad total del enlace C en función de γ obtenidas son similares a las obtenidas por Courcoubetis [25], también derivadas de la teoría de grandes desviaciones y del mismo estilo de la ecuación *inf-sup* (3.16).

Con un procedimiento análogo al del **Teorema 3.14** se pueden probar resultados de consistencia y un teorema central del límite para los estimadores del punto operacional (s^*, t^*) en estos casos.

Si se quiere estimar el punto operacional para el tamaño mínimo de buffer b requerido dada una determinada probabilidad de pérdida a partir de un estimador consistente del ancho de banda efectivo, la función a optimizar será

$$b_n = \sup_{s \geq 0} \inf_{t \geq 0} \left\{ \frac{st\alpha^{(n)}(s, t) + \gamma}{s} - ct \right\}, \quad (3.37)$$

y si se quiere estimar el punto operacional para la capacidad del enlace c necesaria a partir del estimador consistente del ancho de banda efectivo, la función a optimizar será

$$c_n = \sup_{s \geq 0} \inf_{t \geq 0} \left\{ \frac{st\alpha^{(n)}(s, t) + \gamma}{st} - \frac{b}{t} \right\}. \quad (3.38)$$

Como el **Teorema 3.14** asegura convergencia uniforme sobre intervalos acotados, también asegura que el estimador b_n dado por

$$b_n = \frac{s_n^* t_n^* \alpha^{(n)}(s_n^*, t_n^*) + \gamma}{s_n^*} - ct_n^*,$$

y el estimador de c_n dado por

$$c_n = \frac{s_n^* t_n^* \alpha^{(n)}(s_n^*, t_n^*) + \gamma}{s_n^* t_n^*} - \frac{b}{t_n^*},$$

heredan las propiedades del estimador (s_n^*, t_n^*) , que permite construir intervalos de confianza para ambos parámetros de diseño, utilizando las mismas ideas que en la demostración de dicho teorema.

Capítulo 4

Simulaciones y resultados numéricos

4.1. Introducción

Hemos concentrado nuestro trabajo en redes como las que se pueden encontrar en el corazón de internet que es donde, por su complejidad y alto tráfico, se justifica aplicar herramientas sofisticadas de análisis y diseño.

Como ya hemos mencionado en la sección previa, en estas redes la hipótesis de muchas fuentes es más razonable que la hipótesis de buffer grande. La razón es que en un backbone de este tipo arriban una gran cantidad de flujos de múltiples usuarios cada uno con capacidades importantes de enlace aunque la capacidad del buffer para cada fuente es, por lo general, pequeña ya que las redes se dimensionan para atender la simultaneidad de arribo de paquetes y no de ráfagas.

En la **Sección 3.3** vimos que en el régimen asintótico de muchas fuentes, la probabilidad de pérdida está dada por la optimización de la fórmula *inf sup* (3.16). El primer problema que se presenta cuando trabajamos con casos reales es que no siempre se cuenta con una fórmula teórica para el ancho de banda efectivo, excepto que se simplifique el problema adoptando un modelo específico para la fuente de tráfico.

En el caso general se cuenta con estimadores basados en trazas de tráfico y se debe resolver la ecuación (3.16), no para $\alpha(s, t)$ sino para su estimador $\alpha^{(n)}(s, t)$. Del análisis de la **Sección 3.4**, sabemos que el estimador del punto operacional obtenido a partir de un buen estimador del ancho de banda efectivo, es consistente y verifica un Teorema Central del Límite.

En esta sección realizaremos el análisis con trazas de tráfico generadas mediante simulaciones a partir de un modelo teórico conocido para poder hacer comparaciones.

Este capítulo está organizado de la siguiente manera: en la Sección 4.2 se presentan los parámetros del modelo simulado, en la Sección 4.3 se presenta la estimación del ancho de banda efectivo para el modelo simulado, en la Sección 4.4 se presenta el método de optimización con el cual se obtiene el punto operacional del enlace, en la Sección 4.5 se presenta la estimación de la probabilidad de pérdida y finalmente en la Sección 4.6 se presentan los estimadores de los parámetros de diseño, el tamaño mínimo de buffer y la capacidad mínima de enlace para operar con una cierta calidad de servicio.

4.2. Parámetros elegidos para la simulación

Para validar los resultados obtenidos en los Capítulos 2 y 3 y aplicarlos al análisis y diseño de redes, realizamos diversas simulaciones de tráfico según el modelo de Flujo Markoviano Generalizado presentado en la **Sección 2.2** para una fuente.

En el modelo que elegimos, la cadena de Markov modulante tiene $k = 13$ estados y cada estado está asociado a un intervalo de velocidad de transferencia de datos como se muestra en la siguiente tabla

Estado	Velocidad de transferencia
1	(0, 64]
2	(64, 128]
3	(128, 256]
4	(256, 512]
5	(512, 1024]
6	(1024, 2048]
7	(2048, 3072]
8	(3072, 4096]
9	(4096, 5120]
10	(5120, 6144]
11	(6144, 7168]
12	(7168, 8292]
13	(8292, 10240]

Por su diseño, uno espera que el estado usual sea el de la mayor tasa de transferencia disponible en el canal de transmisión, por eso el estado más probable es el 13. También es más habitual saltar de un estado a los estados adyacentes o bien al de tasa máxima de transferencia o al de transferencia mínima o nula. Con estas consideraciones se diseña el generador infinitesimal de la cadena modulante que está dado por la matriz

$$Q = \begin{pmatrix} -7 & 2 & 0,13 & 0,13 & 0,13 & 0,13 & 0,13 & 0,13 & 0,13 & 0,13 & 0,13 & 0,13 & 0,13 & 3,75 \\ 2 & -7 & 2 & 0,13 & 0,13 & 0,13 & 0,13 & 0,13 & 0,13 & 0,13 & 0,13 & 0,13 & 0,13 & 1,88 \\ 1 & 2 & -7 & 2 & 0,13 & 0,13 & 0,13 & 0,13 & 0,13 & 0,13 & 0,13 & 0,13 & 0,13 & 1 \\ 1 & 0,125 & 2 & -7 & 2 & 0,13 & 0,13 & 0,13 & 0,13 & 0,13 & 0,13 & 0,13 & 0,13 & 1 \\ 1 & 0,125 & 0,13 & 2 & -7 & 2 & 0,13 & 0,13 & 0,13 & 0,13 & 0,13 & 0,13 & 0,13 & 1 \\ 1 & 0,125 & 0,13 & 0,13 & 2 & -7 & 2 & 0,13 & 0,13 & 0,13 & 0,13 & 0,13 & 0,13 & 1 \\ 1 & 0,125 & 0,13 & 0,13 & 0,13 & 2 & -7 & 2 & 0,13 & 0,13 & 0,13 & 0,13 & 0,13 & 1 \\ 1 & 0,125 & 0,13 & 0,13 & 0,13 & 0,13 & 2 & -7 & 2 & 0,13 & 0,13 & 0,13 & 0,13 & 1 \\ 1 & 0,125 & 0,13 & 0,13 & 0,13 & 0,13 & 0,13 & 2 & -7 & 2 & 0,13 & 0,13 & 0,13 & 1 \\ 1 & 0,125 & 0,13 & 0,13 & 0,13 & 0,13 & 0,13 & 0,13 & 2 & -7 & 2 & 0,13 & 0,13 & 1 \\ 2 & 0,20 & 0,20 & 0,20 & 0,20 & 0,20 & 0,20 & 0,20 & 0,20 & 0,20 & 0,20 & 0,20 & -8 & 4 \\ 2 & 0,30 & 0,30 & 0,30 & 0,30 & 0,30 & 0,30 & 0,30 & 0,30 & 0,30 & 0,30 & 0,30 & 5,00 & -10 \end{pmatrix}$$

Dentro de cada uno de estos intervalos se sortea cuánto efectivamente se despacha por medio de una distribución normal truncada al intervalo con media igual al punto medio del intervalo y desvío igual a un sexto de la longitud del intervalo. La matriz diagonal \mathbb{H} contendrá en los elementos de su diagonal los valores medios de estas distribuciones.

En las simulaciones realizadas se generaron trazas de largo 1.000 y un ejemplo de traza simulada se puede ver en la Figura 4.1

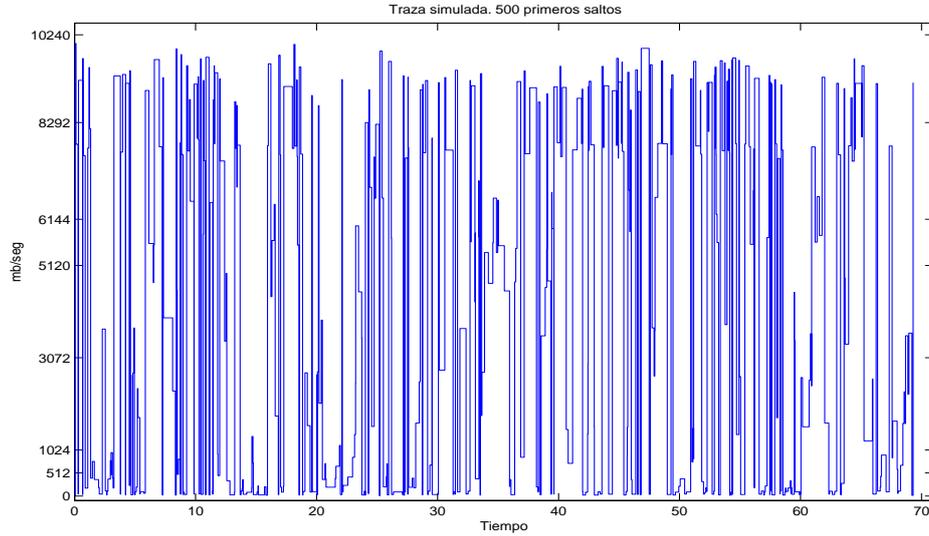


Figura 4.1: Traza generada con el modelo de Flujo Markoviano Generalizado.

4.3. Estimación del ancho de banda efectivo de trazas

Nuestro primer objetivo en calcular el ancho de banda efectivo del modelo teórico presentado, para lo cual usaremos el resultado mostrado en **Teorema 2.2**, el ancho de banda efectivo se calcula entonces según la fórmula

$$\alpha(s, t) = \frac{1}{st} \log \{ \pi \exp [(Q^Z + s\mathbb{H}) t] \mathbf{1} \}.$$

En la Figura 4.2 se muestra el ancho de banda efectivo calculado para el modelo de Flujo Markoviano Generalizado según la ecuación anterior.

Para cada traza simulada estimamos el ancho de banda efectivo utilizando el estimador presentado en el **Teorema 2.9** según la ecuación (2.35) y en la Figura 4.3 se muestra la comparación del ancho de banda efectivo estimado para una traza con el valor teórico.

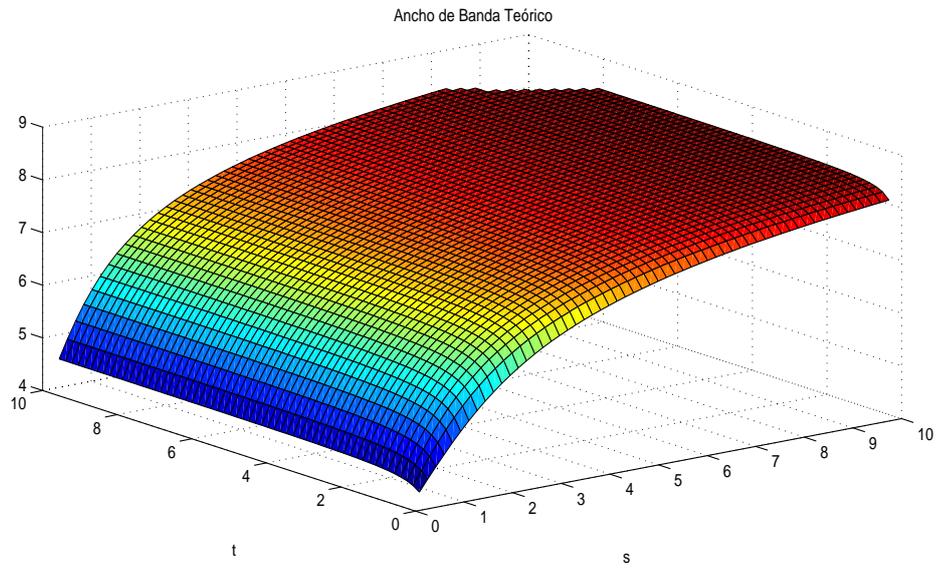


Figura 4.2: Ancho de banda efectivo para el modelo de Flujo Markoviano Generalizado.

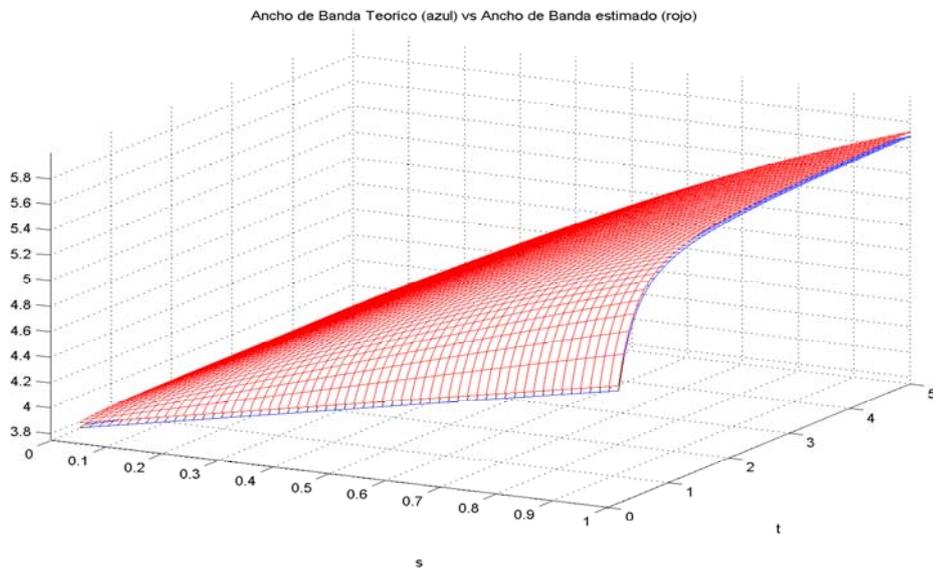


Figura 4.3: Ancho de banda teórico vs. ancho de banda estimado.

4.4. Estimación del punto operacional del enlace

Una vez que obtenemos un buen estimador del ancho de banda efectivo del tráfico que arriba a un enlace, tratamos ahora de estimar los parámetros de calidad de servicio a partir de dicho estimador.

Como ya mencionáramos, el estimador usado verifica las hipótesis requeridas en el **Teorema 3.14** y por lo tanto el estimador del punto operacional será consistente y verificará Teorema Central del Límite.

Calculamos el punto operacional (s^*, t^*) del modelo teórico del Flujo Markoviano Generalizado y su estimador (s_n^*, t_n^*) para cada traza simulada. Para ello se debe resolver el problema de optimización de la fórmula *inf-sup* de la ecuación (3.16), con $\alpha(s, t)$ como en el **Teorema 2.3** y el estimador $\alpha^{(n)}(s, t)$ como en el **Teorema 2.9**.

Para resolver la ecuación (3.16), se procede numéricamente en dos pasos. En el primer paso, para t fijo se encuentra el punto $s^*(t)$ que maximiza $g(s, t) = (c + bt)s - st\alpha(s, t)$ como función de s .

La maximización $g(s, t)$ puede resolverse numéricamente de una manera eficaz al tener en cuenta que la función $st\alpha(s, t)$ es convexa en s , mientras que $(ct + b)s$ es lineal en s . Debido a esto, $g_t(s) = g(s, t)$ es una función unimodal de s y el máximo es único. Por lo tanto, para encontrar dicho máximo se puede empezar desde un primer intervalo de incertidumbre $[s_a, s_b]$ que contiene al máximo. Para determinarlo es suficiente que, para algún $x \in [s_a, s_b]$ se cumpla que $g_t(x) > g_t(s_a)$ y $g_t(x) > g_t(s_b)$. Luego se reduce el intervalo de incertidumbre usando la técnica de “búsqueda de oro” (*golden section search*) como se muestra en [25], de la siguiente manera:

1. Teniendo en cuenta el intervalo $[s_a, s_b]$, dos puntos de prueba s_l, s_r se seleccionan de manera que $s_r - s_a = s_b - s_l = \mathfrak{g}(s_b - s_a)$, donde \mathfrak{g} es la proporción áurea, que es igual a $(\sqrt{5} - 1)/2 \approx 0,618$.
2. Evaluamos $g_t(s_l)$ y $g_t(s_r)$, e identificamos los tres casos:
 - a) si $g_t(s_l) > g_t(s_r)$ el intervalo de búsqueda se convierte en $[s_a, s_r]$,
 - b) si $g_t(s_l) < g_t(s_r)$ el intervalo de búsqueda se convierte en $[s_l, s_b]$,
 - c) si $g_t(s_l) = g_t(s_r)$ el intervalo de búsqueda se convierte en $[s_l, s_r]$.
3. Se repiten los pasos 1 y 2 hasta que el intervalo de incertidumbre tenga una longitud menor que un valor pequeño ϵ .

La “búsqueda de oro” minimiza el número máximo de pasos necesarios para reducir el intervalo de incertidumbre en cierta longitud preestablecida.

Luego de calcular $s^*(t)$ para cada t , es necesario minimizar la función $g(s^*(t), t)$ y encontrar t^* . Para este segundo problema de optimización no hay propiedades generales que permitan hacer el algoritmo de búsqueda de forma eficiente ya que en efecto, $g(s^*(t), t)$ puede tener más de un mínimo local por esta razón, la minimización se resuelve mediante una estrategia de búsqueda lineal de t .

El problema de optimización se resolvió mediante programas desarrollados en el entorno MATLAB (R2011a). Como la optimización involucra la exponencial de una matriz, se reescalaron los valores de las trazas para sortear el problema que se ocasiona con los valores grandes, que MATLAB interpreta como infinito.

Estimamos el punto operacional (s^*, t^*) y su intervalo de confianza, para lo cual se simularon $K = 300$ trazas de longitud $T = 10000$ muestras y para cada traza simulada indexada por $i = 1, \dots, K$, se construyó el correspondiente estimador $(s_n^*(i), t_n^*(i))$.

Por el **Teorema 3.14**, el vector $\sqrt{n}((s_n^*, t_n^*) - (s^*, t^*))$ es asintóticamente normal bivariado con media $\vec{0}$ y matriz de covarianza Σ que estimamos usando las covarianzas empíricas de la siguiente manera,

$$\Sigma_K = \frac{n}{K} \begin{pmatrix} \sum_{i=1}^K (s_n^*(i) - \bar{s}_n^*)^2 & \sum_{i=1}^K (s_n^*(i) - \bar{s}_n^*)(t_n^*(i) - \bar{t}_n^*) \\ \sum_{i=1}^K (s_n^*(i) - \bar{s}_n^*)(t_n^*(i) - \bar{t}_n^*) & \sum_{i=1}^K (t_n^*(i) - \bar{t}_n^*)^2 \end{pmatrix}, \quad (4.1)$$

donde $\bar{s}_n^* = \frac{1}{K} \sum_{i=1}^K s_n^*(i)$ y $\bar{t}_n^* = \frac{1}{K} \sum_{i=1}^K t_n^*(i)$.

Luego, podremos decir que aproximadamente

$$(s_n^*, t_n^*) \approx N \left((s^*, t^*), \frac{1}{K} \Sigma_K \right),$$

de donde la región de confianza de nivel α se puede obtener como

$$R_\alpha = (s_n^*, t_n^*) + \frac{A_K^t \mathcal{B}(\vec{0}, \sqrt{\chi_\alpha^2(2)})}{\sqrt{n}},$$

con A_K la matriz que verifica que $A_K^t A_K = \Sigma_K$ y $\mathcal{B}(x, r)$ es la bola de centro x y radio r .

Para verificar los resultados obtenidos se calculó el punto operacional (s^*, t^*) de otras 300 trazas, independientes de las usadas para estimar Σ_K , y se construyó la región de confianza del 95% donde se espera que aproximadamente el 95% de los puntos (s_n^*, t_n^*) calculados en cada una de las trazas simuladas se encuentren dentro de la región $R_{0,05} = (s^*, t^*) + \frac{1}{\sqrt{n}} A_K^t \mathcal{B}(\vec{0}, \sqrt{\chi_{0,05}^2(2)})$.

Los resultados numéricos se muestran en la Figura 4.4, y se encontró que el 95,16% de los puntos estimados se ubicaron dentro de la región predicha.

4.5. Estimación de los parámetros de calidad de servicio de un enlace

El objetivo fundamental de la estimación del punto operacional es la estimación de la probabilidad de pérdida del enlace según la ecuación (3.16) y de otros parámetros de la calidad del servicio como por ejemplo el retardo de los paquetes de información.

Respecto a este último es importante notar que, en el régimen asintótico de muchas fuentes, el retardo real que sufren los paquetes que atraviesan un enlace coincide asintóticamente con el retardo virtual de los mismos como se puede ver en [60], entendiendo por

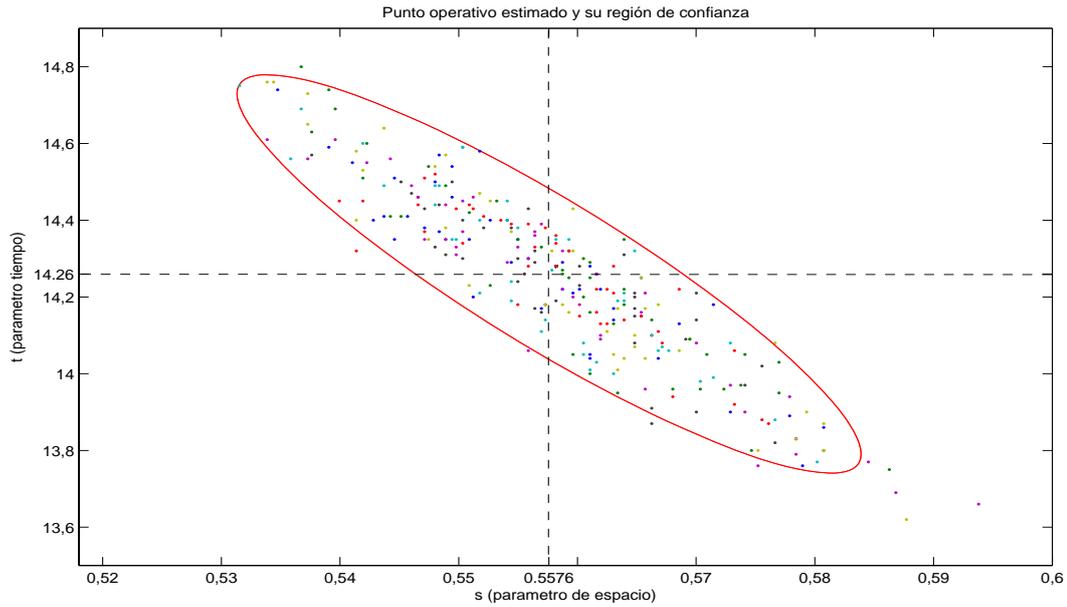


Figura 4.4: Punto operacional estimado y región de confianza del 95 %.

retardo virtual el que se obtiene a través del tamaño de la cola. En otras palabras, si el enlace envía C paquetes por unidad de tiempo y la probabilidad de que el tamaño de la cola sea superior a B es q , entonces la probabilidad de que el retardo sea superior a B/C será q , por lo tanto en dicho régimen, si se estima la probabilidad de que el tamaño de la cola supere a B , se obtiene directamente un estimador de la distribución real del retardo.

Nos concentraremos en la estimación de la probabilidad de pérdida, ya que el análisis del retardo se deduce de la misma ecuación. Recordemos que el estimador del ancho de banda efectivo que se utiliza cumple con las hipótesis del **Teorema 3.14**, entonces el estimador de la probabilidad de pérdida será

$$\gamma_n = \inf_{t \geq 0} \sup_{s \geq 0} \{ (c + bt)s - st\alpha^{(n)}(s, t) \}, \quad (4.2)$$

que es consistente y verifica un Teorema Central del Límite.

Una vez obtenido este estimador, la probabilidad de pérdida se podrá aproximar por

$$q_n = P_n(Q_N > B) \approx e^{-N\gamma_n}, \quad (4.3)$$

donde Q_N es el tamaño de la cola y N es la cantidad de fuentes del sistema.

En la Figura 4.5 se muestra la estimación de γ_n para 600 trazas simuladas, su valor teórico y su intervalo de confianza. En este caso los resultados numéricos muestran que el 95,5 % de los valores se encuentran dentro del intervalo de confianza del 95 %

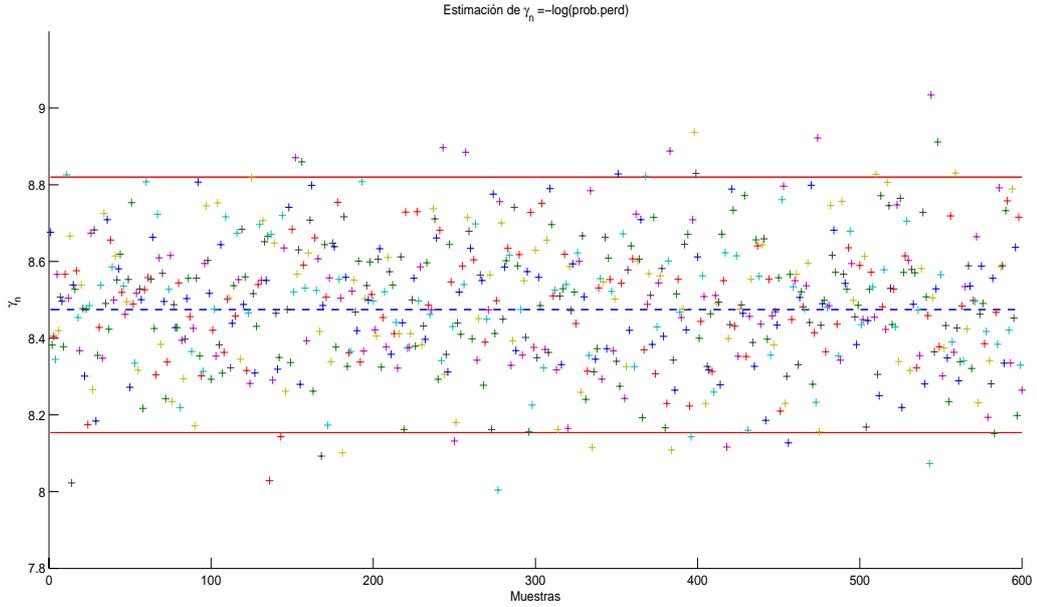


Figura 4.5: γ_n estimado, γ teórico e intervalo de confianza del 95 %.

4.6. Diseño de un enlace basado en la estimación del ancho de banda efectivo

Los resultados anteriores pueden extenderse al caso del diseño de un enlace de una red que requiere cierta calidad de servicio. Se podría conocer el tamaño del buffer mínimo $B = Nb$ de un enlace dada su capacidad $C = Nc$, el tráfico que lo atraviesa y la probabilidad de pérdida máxima que se desea tener o en su defecto el retardo máximo deseado.

Análogamente, si se tiene la misma información que antes pero, definido el tamaño del buffer que se desea, es posible calcular la capacidad mínima necesaria del enlace para asegurar la probabilidad de pérdida requerida.

La respuesta a estos problemas de diseño se obtienen resolviendo ecuaciones similares a las de la optimización *inf-sup* vista anteriormente.

El tamaño de buffer mínimo para asegurar una probabilidad de pérdida γ está dado por la ecuación

$$B_n = \sup_{s \geq 0} \inf_{t \geq 0} \left\{ \frac{Nst\alpha^{(n)}(s, t) + N\gamma}{s} - Ct \right\}. \quad (4.4)$$

Por otro lado la capacidad mínima necesaria para asegurar una probabilidad de pérdida γ es

$$C_n = \sup_{s \geq 0} \inf_{t \geq 0} \left\{ \frac{Nst\alpha^{(n)}(s, t) + N\gamma}{st} - \frac{B}{t} \right\}. \quad (4.5)$$

Las expresiones (4.4) y (4.5) se optimizaron de manera análoga a lo realizado en la **Sección 4.4** para la probabilidad de pérdida, trabajando con una única fuente, es decir $N = 1$. Para cada caso se realizaron 600 simulaciones.

En la Figura 4.6 se muestran los resultados numéricos de la estimación del punto operacional para el tamaño mínimo del buffer. Éstos se interpretan como los valores de los parámetros s y t donde se alcanza el valor del buffer que aseguran una determinada probabilidad de pérdida cuando la capacidad del enlace es c y dado un valor de ancho de banda efectivo. Además en la Figura 4.7 se muestran las estimaciones del tamaño de buffer mínimo, su valor teórico y la región de confianza del 95 %, comprobando que el 97,66 % de los valores se encuentran dentro de la región. Los valores negativos de tamaño de buffer para alguna traza indican simplemente que no se necesitó buffer para cumplir con los requerimientos de calidad de servicio establecidos.

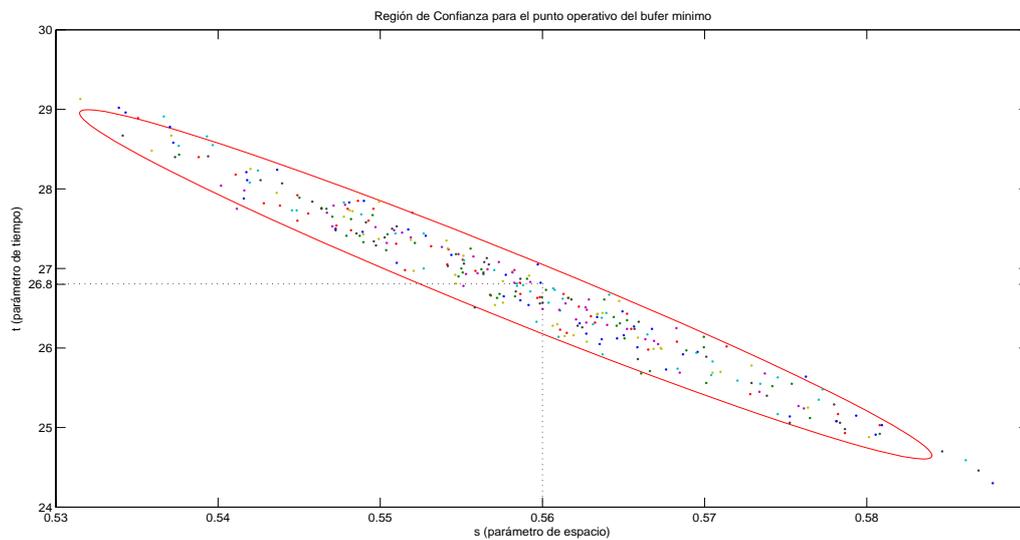


Figura 4.6: Punto operacional estimado y región de confianza del 95 % para tamaño mínimo de buffer.

De manera idéntica procedemos con los puntos operacionales para la capacidad del enlace interpretándolos como los valores de los parámetros s y t donde se alcanza la capacidad mínima requerida para asegurar una determinada probabilidad de pérdida con tamaño de buffer b y dado un valor de ancho de banda efectivo. En la Figura 4.8 se muestran los resultados numéricos de la estimación del punto operacional mientras que las estimaciones de la capacidad mínima del enlace, el valor teórico y el intervalo de confianza del 95 % se muestran en la Figura 4.9 y el 96,6 % de los valores se encuentran dentro del intervalo.

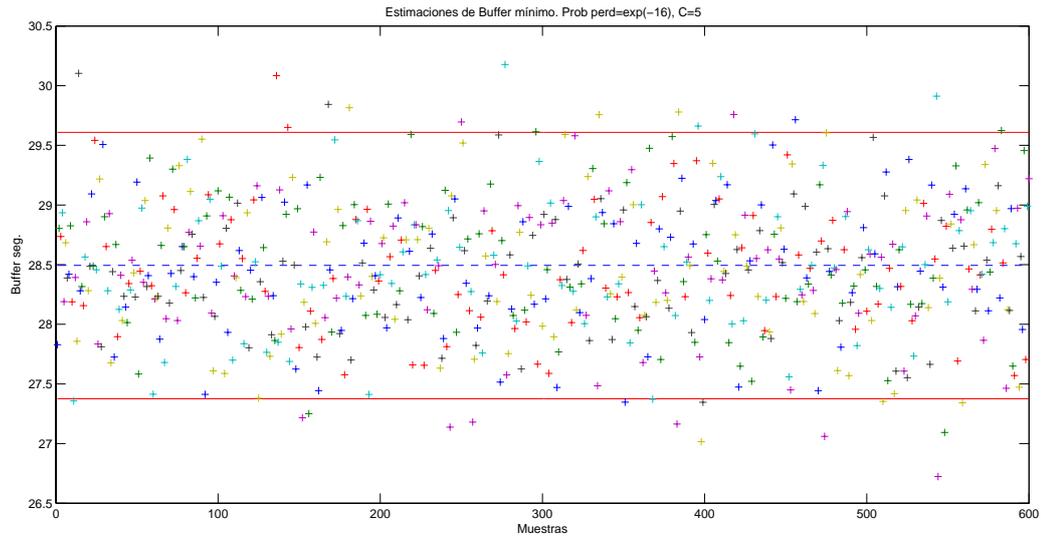


Figura 4.7: B_n estimado, B teórico e intervalo de confianza del 95 %.

El tamaño del buffer es casi siempre un dato duro, es decir un dato impuesto, por eso habitualmente se hace diseño sobre la capacidad del enlace.

Es de interés analizar como varía la probabilidad de pérdida cuando cambia la capacidad c del enlace. Para ello recurrimos a la Figura 4.10 donde graficamos $-\gamma$ vs. c . Se puede observar que la pendiente de dicha curva decrece más rápidamente a medida que c aumenta y cuando la capacidad del enlace se aproxima al valor pico de la fuente la probabilidad de pérdida tiende a 0 ($-\gamma \rightarrow -\infty$). En nuestro caso, por la forma en que lo diseñamos, el modelo se encuentra trabajado casi al límite de la capacidad del enlace.

En la Figura 4.11 se puede observar una aplicación de la Figura 4.10 para apreciar mejor la diferencia entre la curva con el modelo teórico y con el modelo estimado ya que ambas se encuentran muy próximas.

Si en cambio la probabilidad de pérdida tolerada no varía, es decir se mantiene γ fijo, la variación de c es menor, esto es motivado por la poca dispersión de c como se puede observar en la Figura 4.9.

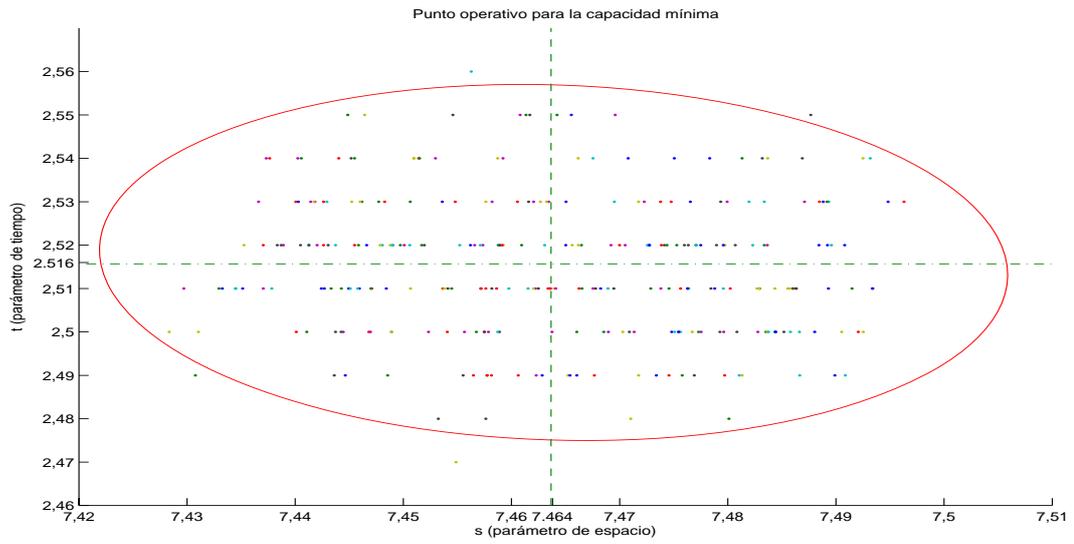


Figura 4.8: Punto operacional estimado y región de confianza del 95 % para la capacidad del enlace.

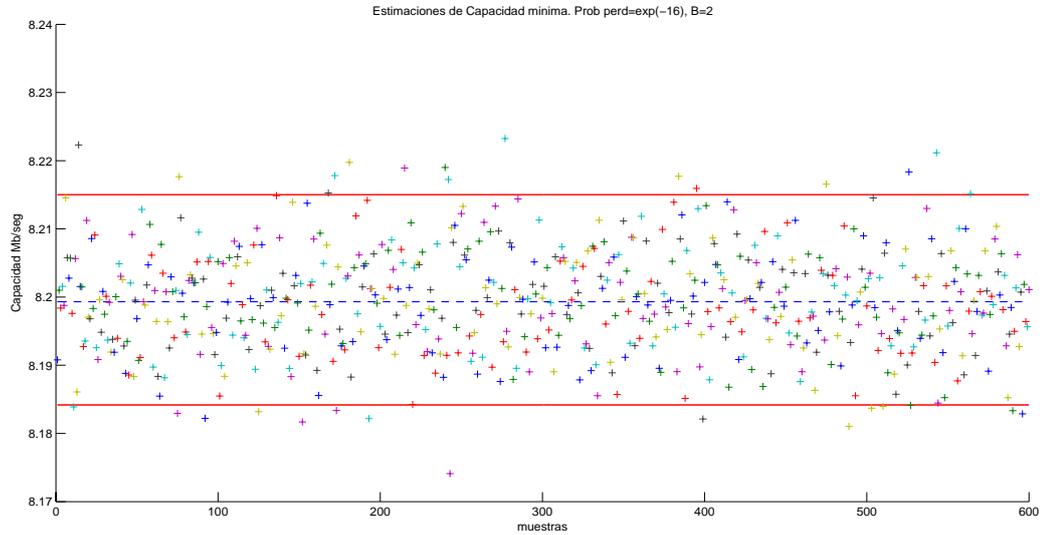


Figura 4.9: C_n estimado, C teórico e intervalo de confianza del 95 %.

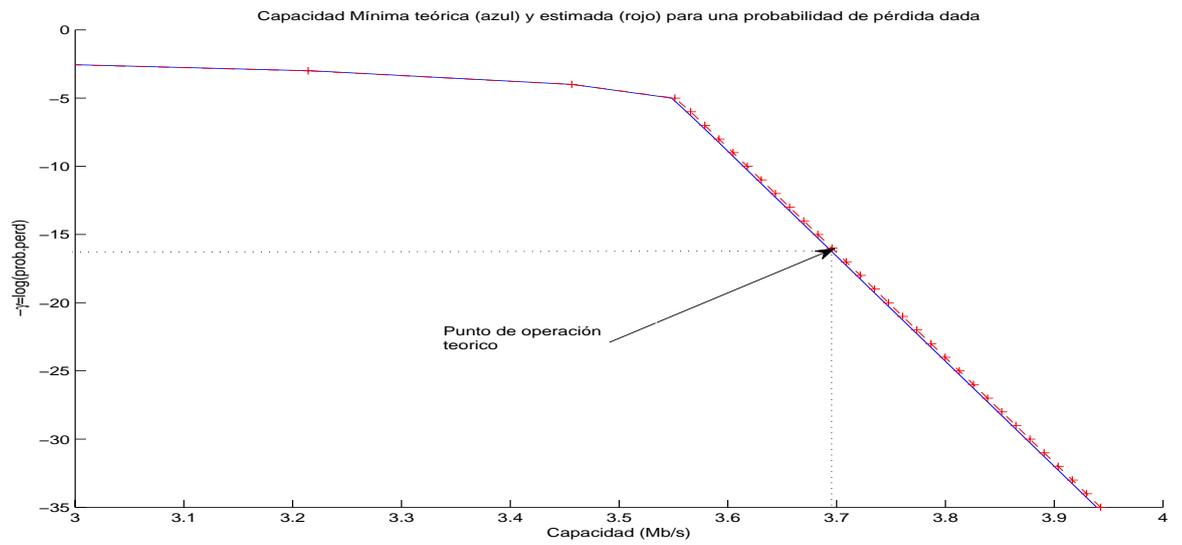


Figura 4.10: $-\gamma$ vs. capacidad del enlace.

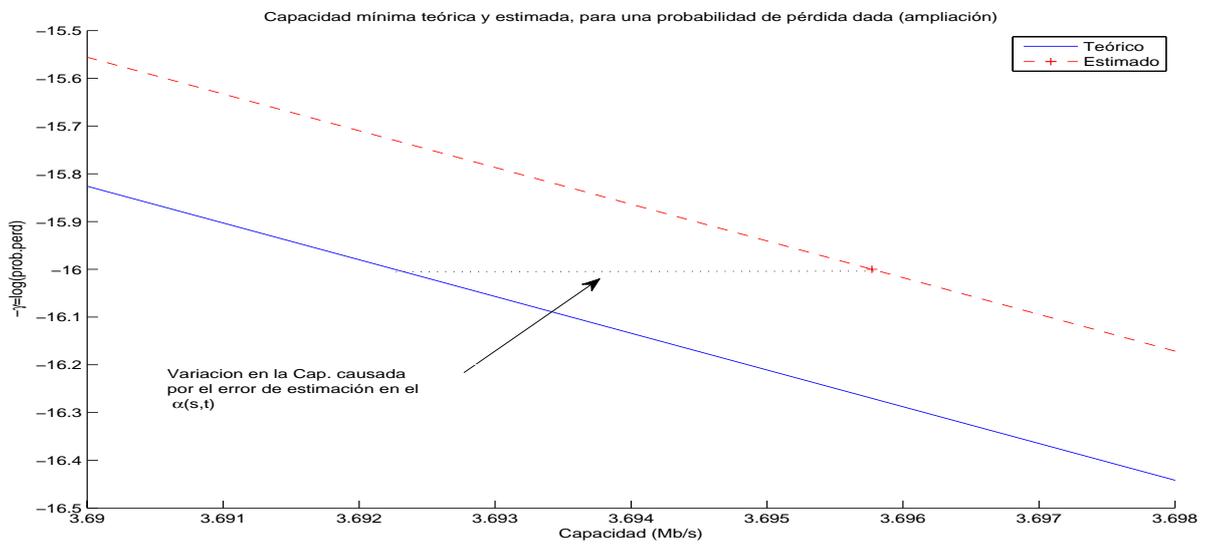


Figura 4.11: Ampliación de la imagen en la Figura 4.10.

Conclusiones y trabajos futuros

En esta tesis resumimos el estado del arte en estas dos áreas y en cada una de ellas hemos procurado realizar un aporte:

- Modelado de las trazas de tráfico: hemos propuesto un nuevo modelo teórico en el cual es posible aplicar refinadas herramientas y resultados de estadística matemática. El modelo presenta además la ventaja de ser muy realista para los actuales requerimientos de las redes de telecomunicación. Realizamos un estudio riguroso del modelo y hemos encontrado una fórmula para el ancho de banda efectivo donde se puede visualizar el papel que cumple los parámetros del modelo.
- Estimación de parámetros: hemos propuesto una metodología para estimar el ancho de banda efectivo para una traza de tráfico de una fuente en el modelo introducido. Esta posee las propiedades esperadas para asegurar que cumple con un Teorema Central de Límite y poder así construir un intervalo de confianza. Hemos encontrado también bajo qué condiciones estimadores del punto operacional de un enlace y de la probabilidad de pérdida basados en el ancho de banda que hemos estimado son consistentes y cuándo es válido un Teorema Central del Límite para dichos estimadores. Hemos procurado condiciones bajo las cuales se pueda asegurar consistencia de estimadores de los parámetros de diseño de un enlace, el tamaño mínimo de buffer y la capacidad mínima de enlace.

El tema abordado es de interés actual y muy diverso. En el desarrollo de este trabajo han surgido algunos aspectos que se pueden analizar con mayor detalle. Por ejemplo, en el área de la estimación de la calidad de servicio en una red aún existen diversos aspectos que quedan abiertos. Algunas de las preguntas que surgen son: ¿Es posible relajar el supuesto de que los soportes de las medidas son disjuntos? De ser así el modelo de Flujo Markoviano Generalizado podría tratarse como un modelo de Markov Escondido. ¿Es esto conveniente? Es decir, ¿ajusta mejor un tráfico real o sólo agrega complicaciones a los cálculos? ¿Cuál sería entonces el estimador a partir de trayectorias de tráfico para estimar el ancho de banda efectivo? Se podría proponer una comparación de ventajas y desventajas de ambos modelos y sus estimadores.

Los puntos anteriores son sólo algunos de los temas que podríamos encarar como trabajo futuro.

Notación y abreviaturas

En esta sección presentamos la notación y abreviaturas que utilizaremos a lo largo de la tesis.

\mathbb{I}_A	Indicatriz del conjunto A .
$[x]$	Mayor entero menor o igual que x .
I_k	Matriz identidad de orden k .
\vec{v}	Vector v .
$E(X)$	Esperanza matemática de la variable X .
$Var(X)$	Varianza de la variable X .
CAC	Control de aceptación de conexiones.
QoS	Calidad de servicio.
$\mathcal{M}_{m \times n}$	Clase de matrices de orden $m \times n$.
$\mathcal{N}(\mu, \sigma^2)$	Distribución normal con media μ y varianza σ^2 .
$\mathcal{N}(\vec{\mu}, \Sigma)$	Distribución normal bivariada con vector de media $\vec{\mu}$ y matriz de covarianza Σ .
\mathcal{G}	Proceso gaussiano continuo.
W	Proceso de Wiener standard.
C^1	Clase de funciones continuas, con derivada primera continua.
$X \xrightarrow{c.s.} Y$	Convergencia casi segura.
$X \xrightarrow{w} Y$	Convergencia débil.
$X \xrightarrow{\mathcal{D}} Y$	Convergencia en distribución.
$X \xrightarrow{\mathcal{P}} Y$	Convergencia en probabilidad.

Glosario

Ancho de Banda	Cantidad de información o de datos que se puede enviar a través de una conexión de red en un período dado. Se indica generalmente en bits por segundo (bit/s), kilobits por segundo (kbit/s), o megabits por segundo (Mbit/s).
Backbone	La parte de una red que se utiliza como la ruta primaria para transportar tráfico entre segmentos de la misma.
Buffer	Dispositivo digital que está reservada para el almacenamiento temporal de información. Mientras los datos están en el buffer, aguardan para ser procesados.
Buffer overflow (Desborde del buffer)	Sucede cuando el tamaño de los datos entrantes exceden el tamaño del buffer, dando como resultando pérdida de información.
Calidad de servicio (QoS)	Conjunto de reglas para priorizar el tráfico de datos en una red basándose en garantías específicas de disponibilidad y rendimiento de red.
Control de admisión de una conexión (CAC)	Colección de acciones tomadas por la red durante la fase de instalación para establecer si un camino/canal virtual puede ser aceptado por la red.
Fuente	Entidad que envía unidades de datos del servicio por una conexión.
Ingeniería de tráfico	Proceso de mapear la demanda de tráfico sobre la topología de la red para controlar el flujo de tráfico.
MPEG	Estándar de compresión de Video creado por el grupo MOVING PICTURE EXPERT GROUP.
Multiplexar	Combinar dos o más canales de información en un solo medio de transmisión.
Multiplexado estadístico	Método para combinar varios canales en un único medio de transmisión que da prioridad en función del volumen de datos que cada canal intentan transmitir en cualquier momento dado.

Nodo	Punto de interconexión a una red.
Overflow	Exceso de datos que pueden ser perdidos o transferidos.
Protocolo	Conjunto de reglas para usar una interconexión o una red de manera que la información transmitida en la interconexión pueda ser interpretada correctamente por todas las partes que intervienen en la comunicación.
Red	Conjunto de nodos interconectados para establecer una comunicación.
Switch o conmutador	Dispositivo de interconexión de redes informáticas.
Tráfico	Conjunto de datos que circula a través de la red, en un momento determinado.
Trazas o Trayectorias de tráfico	Medición de datos que circula a través de la red, a lo largo del tiempo.

Bibliografía

- [1] Adler, R.J. (1981). *The Geometry of Random Fields*. New York. John Wiley & Sons.
- [2] Albert, A. (1962). *Estimating the Infinitesimal Generator of a Continuous Time, Finite State Markov Process*. Ann. Math. Statist. Volume 33, Number 2, pp. 727-753.
- [3] Alzate, M. (1990). *Multiplexaje de voz y datos*. Tesis de Maestria MIE-90-II-1, Departamento de Ingeniería Eléctrica, Universidad de los Andes.
- [4] Antoniadis, P., Courcoubetis, C., Mason, R. (2004). *Comparing economic incentives in peer-to-peer networks*. Computer Networks, 46, pp. 133-146.
- [5] L. Aspirot, P. Belzarena, P. Bermolen, A. Ferragut, G. Perera and M. Simon. (2005) *Quality of service parameters and link operating point estimation based on effective bandwidths*. Performance Evaluations, 59 , pp. 103-120.
- [6] L. Aspirot, P. Belzarena, G. Perera and B. Bazzano. (2005) *End-To-End quality of service prediction based on functional regression*. HET-NET, 59 , pp. 103-120.
- [7] Baccelli, F. and Brémaud, P. (2003). *Elements of queueing theory. Palm martingal calculus and stochastic recurrences*, 2nd Ed. Sringer-Verlag.
- [8] Bartlett, M.S. (1967). *Inference and Stochastic Processes. Monographs on Applied Statistic*. London. Chapman and Hall.
- [9] Bartlett, M.S. (1975). *The Statistical Analysis of Spatial Pattern*. London. Chapman and Hall.
- [10] Bernshtein, S. N. (1944). Extension of the Central Limit Theorem of Probability Theory to Sums of Dependent Variables. *Uspehi Mat. Nauk*, **10**, 65-114. (Russian).
- [11] Billingsley, P. (1961). *Statistical Inference for Markov Process*. Chicago. University of Chicago Press.
- [12] Billingsley, P. (1968). *Convergence of Probability Measures*. New York. Wiley & Sons.

- [13] Billingsley, P. (1979). *Probability and Measures*. New York. Wiley & Sons.
- [14] Bitsakia, M., Stamoulis, D., Courcoubetis, C. (2006). *An efficient auction-based mechanism for hierarchically structured bandwidth markets*. Computer Communications, 29, pp. 911-921.
- [15] Bolch, G., Greiner, S., de Meer, H. and Trivedi, K. S. (2006). *Queueing Networks and Markov Chains. Modeling and Performance Evaluations with Computer Science Applications*, 2nd Edition. Wiley & Sons.
- [16] Bremaud, P. (1999). *Markov Chain*. New York. Springer.
- [17] Brèmand, P. (1999). *Markov Chain, Gibbs Field, Montecarlo Simulation and Queues*. Text in Applied Mathematics N°31. New York. Springer.
- [18] Buschiazzo, D., Ferragut, A. y Vázquez, A. (2005). *Fast overflow probability estimation tool for MPLS networks*. LANC'05 Proceedings of the 3rd international IFIP/ACM Latin American conference on Networking.
- [19] Cao, J., Cleveland, W.S., Sun, D.X. (2004). *Bandwidth Estimation for Best-Effort Internet Traffic*. Statistical Science. Vol.19, N°3, pp. 518-543.
- [20] Chang, C.S. (1994). *Stability Queue Length and Delay of the Deterministic and Stochastic Queueing Networks*. IEEE. Trans. AC. 39, pp. 913-931.
- [21] Chang, C.S. (2000). *Performance Guarantees in Communications Network*. Springer.
- [22] Chung, K.L. (1967). *Markov Chain with stationary transition probabilities*. Berlin. Springer.
- [23] Courcoubetis, C., Weber, R. (1996). *Buffer overflow asymptotics for a switch handling many traffic sources*. Journal of Applied Probability, 33. http://www.ics.forth.gr/netgroup/publications/CW95_largeN.ps.gz.
- [24] Courcoubetis, C., Weber, R. (2003). *Pricing Communications Networks*. England. John Wiley & Sons.
- [25] Courcoubetis, C., Siris, V.A. (2002). *Procedures and tools for analysis of network traffic measurements*. Performance Evaluation, 48, pp.5-23.
- [26] Cressie, N.A.C. (1993). *Statistics for Spatial Data*. New York. John Wiley & Sons.
- [27] G. de Veciana, Walrand, J. (1995). *Effective bandwidths. Call admission, traffic policing and filtering for ATM networks*. Queueing Systems, 20, pp.37-59.
- [28] Dembo, A.; Zanjic, T. (1995). *Large Deviations: From Empirical Mean and Measure to Partial Sum Process*. Stochastic process Applications, 57, pp. 191-224.

- [29] Dembo, A.; Zeitouni, O. (1993). *Large Deviations Techniques and its Applications*. Jones and Barlett. New York.
- [30] Dobrushin, P.L. (1968). *The Description of a Random Field by Means of Conditional Probabilities and Conditions of its Regularity*. Theory of Probability and its Applications. Vol XIII, N°2, pp. 197-224.
- [31] Doukhan, P. (1995). *Mixing: Properties and Examples*. Lectures Notes in Statistics 85. Springer-Verlag.
- [32] Doukhan, P., Louhichi, S. (1999). A New Weak Dependence Condition and Applications to Moment Inequalities. *Stochastic Processes and their Applications*, 84, pp. 313-342.
- [33] Elwalid, A. I. and Mitra, D. (1993). Effective bandwidth of general Markovian traffic sources and admission control of high speed networks. *IEEE/ACM Transactions on Networking*. Vol. 1, 3, pp. 329-343.
- [34] Eun, D. Y. and Shroff, N.B. (2004). Network Decomposition in the Many-Sources Regime. *Advances in Applied Probability*, Vol. 36, No. 3, pp. 893-918.
- [35] Feller, W. (1978). *Introducción a la Teoría de Probabilidades y sus Aplicaciones*. Vol.II. México. Limusa.
- [36] Gibbens, R.J. (1996). *Traffics characterisation and effective bandwidths for broadband network traces*. *Stochastics Networks: Theory and applications*. Oxford University Press. Oxford. <http://www.statslab.cam.ac.uk/Reports/1996/1996-9.ps>.
- [37] Gilks, W.R., Richardson, S. and Spiegelhalter, D.L. (1995). Introducing Markov Chain Monte Carlo. In *Markov Chain Monte Carlo in Practice* (eds. Gilks, W.R., Richardson, S. and Spiegelhalter, D.L.), pp. 1-19. London. Chapman & Hall.
- [38] Guyón, X. (1995) *Random Fields on a Network. Modeling, Statistics, and Applications*. Spinger-Verlag.
- [39] Hoel, P.G., Port, S.C., Stone, C.J. (1972). *Introduction to Stochastic Processes*. Houghton Mifflin.
- [40] Kaj, I. (1999). *Stochastic Modeling in Broadband Communications Systems*. Lectures Notes Series of the Laboratory for Research in Statistics and Probability. Carleton University. N° 327.
- [41] Kalashnikov, V. (1994). *Mathematical Methods in Queueing Theory*. Klüver.
- [42] Kesidis, G., Walrand, J. and Chang, C.S. (1993). *Effective bandwidth for multiclass Markov fluids and other ATM sources*. IEEE ACM Trans. Networking 1,4, pp. 424-428.

- [43] Kelly, F. (1996). *Notes on effective bandwidths. Stochastic Networks: Theory and applications.* Oxford University Press. Oxford. pp. 141-168. <http://www.statslab.cam.ac.uk/~frank/eb.ps>
- [44] Kendal, M. G. and Stuart, A. (1961). *The Advanced Theory of Statistics*, Vol.2. London. Griffin.
- [45] Lebedev, E. A., Lukashuk, L. I. (1986). *Maximum likelihood estimation of the infinitesimal matrix of a Markov chain with continuous time* (in Russian), Dokl. Akad. Nauk Ukr. SSR Ser. A 1, pp.12-14.
- [46] Lewis, J.T. and Russell, R. (1996). An introduction to Large Deviations for Teletraffic Engineers. <http://www.stp.dias.ie/DAPG/LDtut96.ps>.
- [47] Likhanov, N., Mazumdar, R., Öztürk, O. (2004). Many Sources Asymptotics for Networks with Small Buffers. *Queueing Systems Theory and Applications (QUESTA)*, Vol. 46(1-2), pp.129-147.
- [48] Marrón, B.(2004). *Modelos Markovianos en Segmentación de Imágenes.* Tesis Magister. Universidad Nacional del Sur. Bahía Blanca. Argentina.
- [49] Marrón, B.(2009). *Cálculo del Ancho de Banda Efectivo para un Flujo Markoviano con tasas de transferencia continuas.* Proceedings of the X Congreso Antonio Monteiro, Univ. Nacional del Sur, Bahía Blanca, Argentina. pp. 127-134. ISSN: 0327-9170. <http://www.matematica.uns.edu.ar/XCongresoMonteiro/Actas/MARRON.pdf>
- [50] Marrón, B.(2012). Co-occurrence Matrix and Fractal Dimension for Image Segmentation. *Revista de Matemática: Teoría y Aplicaciones.* Vol. 19(1), pp. 49-63. ISSN: 1409-2433. <http://revista.emate.ucr.ac.cr/index.php/revista/article/view/369/287>.
- [51] Newell, G.F. (1982). *Applications of Queueing Theory.* Chapman and Hall.
- [52] Norris, D. (1997). *Markov Chain.* Cambridge.
- [53] O'Connell, N. (1999). Large Deviations with Applications to Telecommunications. <http://www.maths.lth.se/mathstat/research//asn/ldcourse99notes.ps>.
- [54] Ozturk, O., Mazumdar, R. y Likhanov, N. (2004). Many sources asymptotics in networks with small buffers. *Queueing Systems*, 46, pp.129-147.
- [55] Pechiar, J., Perera, G. and Simón, M. (2002). *Effective bandwidth estimation and testing for Markov sources.* Performance Evaluation, 45, pp. 157-175.
- [56] Perera, G. (1994). *Estadística Espacial y Teoremas Centrales del Límites.* Tesis Doctoral. Centro de Matemática. Universidad de la República. Uruguay.
- [57] Perera, G. (1994) Spatial Statistics, Central Limit Theorems for Mixing Random Fields and the Geometry of \mathbf{Z}^d . *C.R.Acad. Sci. Paris*, t.319, Série I, pp. 1083-1088.

- [58] Perera, G. (1997). Geometry of \mathbf{Z}^d and the Central Limit Theorem for Weakly Dependent Random Fields. *Journal of Theoretical Probability*. Vol. 10, N°3, pp. 581-603.
- [59] Perera, G. (2000). *Random Fields on \mathbf{Z}^d , Limit Theorems and Irregular Sets*. Lecture Notes in Statistics N°159: 57-78. Centre de Recherches Mathématiques. Springer.
- [60] Shakkottai, S. and Srikant, R. (2001). Many sources Delay Asymptotics with Applications to Priority Queues. *Queueing Systems Theory and Applications (QUESTA)*, Vol. 39, pp.183-200.
- [61] Shwartz, A. and Weiss, A. (1995). *Large Deviations for Performance Analysis: Queues, Communications and Computing*. Chapman Hall.
- [62] Siris, V.A. (1999): Large Deviations Techniques for Traffic Engineering. <http://www.ics.forth.gr/netgroup/msa>.
- [63] Smith, A.F.M, and Roberts, G.O. (1993). Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods. *J.R.S.S. Serie B*, 55, pp. 3-25.
- [64] Rabinovich, P. (2000). *Statistical Estimation of Effective Bandwidth*. M.Sc. Thesis. University of Cambridge.
- [65] Robertazzi, T.G. (2000). *Computer Networks and Systems- Queueing Theory and Performance Evaluation*. Third Editions. Springer.
- [66] Rosen, E., Viswanathan, A. Callon, R. (2001). *Multiprotocol Label Switching Architecture*. RFC3031, IETF.
- [67] Ripley, B.D. (1981). *Spatial Statistics*. London. John Wiley & Sons.
- [68] Ripley, B.D. (1988). *Spatial Inference for Spatial Processes*. Cambridge University Press.
- [69] Rudin, W. (1980). *Principios de Análisis Matemático*. 3er Edition. McGraw Hill
- [70] Wischik, D. (1999). Large Deviations and Internet Congestion. <http://www.statslab.cam.ac.uk/~djw1005/Stats/Research/phd.html>.
- [71] Wischik, D. (1999). The Output of a Switch or Effective Bandwidths for Networks. *Queueing Systems*, 32, pp. 383-396.
- [72] Wischik, D. (2000). Sample path large deviations for queues with many inputs. *Annals of Applied Probability*, 11, pp. 389-404.