

Kernel methods application to compare different rate distribution in Generalized Markov Fluids

José Bavio y Beatriz Marrón

Departamento de Matemática
Universidad Nacional del Sur

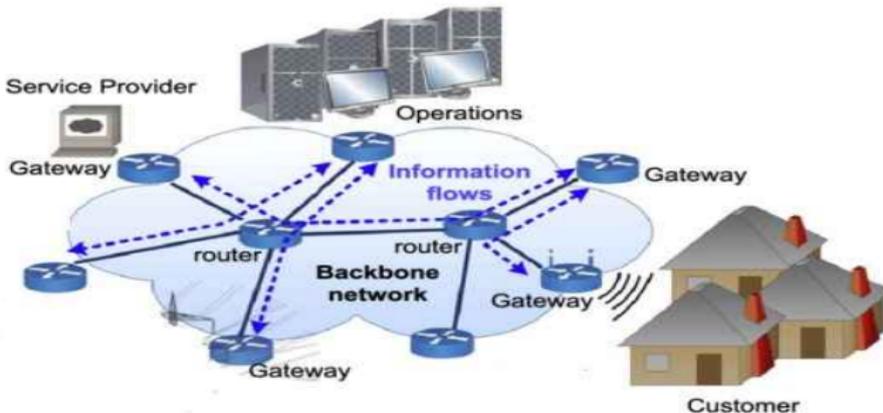


5 de Junio de 2019

Organización de esta presentación

- El desafío tecnológico que inspira el trabajo.
- Definiciones y algunas herramientas para una forma de resolución.
- Simulaciones.

Problema en una red de telecomunicaciones



Objetivos Generales a nivel de un nodo de la red

- 1 Determinar una magnitud para asignar recursos.
- 2 Modelar las fuentes de tráfico.

- Medir trabajo o requerimiento de la fuente. Aditiva

Definición (F. Kelly, 1996)

El Ancho de Banda Efectivo de una fuente estacionaria es

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

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

- X_t se debe modelar.

Modelado de conexiones - Flujo Markoviano Generalizado

Motivación

- Conexión activa varía entre estados de actividad (stream, navegación, descarga, etc)
- Cada estado despacha información en un cierto rango de velocidades (disjuntos) de manera aleatoria
- Modelo
 - Z_t es una cadena de Markov irreducible homogénea con espacio de estados finito $\mathcal{K} = \{1, 2, \dots, k\}$, distribución invariante π , y generador infinitesimal \mathbb{Q}^Z .
 - $Y_t | Z_t = i$ se distribuye según la ley f_i , con soportes disjuntos, esperanza μ_i y varianza σ_i^2 .
 - Finalmente el Flujo Markoviano Generalizado es

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

Ancho de Banda efectivo de Flujo Markoviano Generalizado

Objetivo

Estimaciones basadas en trazas de tráfico.

Teorema

Sea $\{X_t\}_{t \geq 0}$ un Flujo Markoviano Generalizado modulado por la CM irreducible homogénea Z_t con distribución invariante π y generador infinitesimal \mathbb{Q}^Z . Las variables $Y_t | Z_t = i$ con distribución f_i , de soportes conocidos y disjuntos, primer momento finito μ_i , varianza

σ_i^2 , la matriz $\mathbb{H} = \begin{pmatrix} \mu_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mu_k \end{pmatrix} \in M_{k \times k}$, y $\mathbf{1}$ un vector columna

con todas las entradas iguales a 1, entonces:

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

Estimación del ancho de banda efectivo

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

- Estimador para cada elemento de \mathbb{Q} no diagonal

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

$\gamma(i, j, h) = n^\circ$ de transiciones de la cadena de i a j en $[0, h]$,

$\tau(i, h) =$ tiempo que la cadena permanece en el estado i en $[0, h]$. (Lebedev-Lukashuk, 1986)

- Estimador para cada elemento de \mathbb{H}

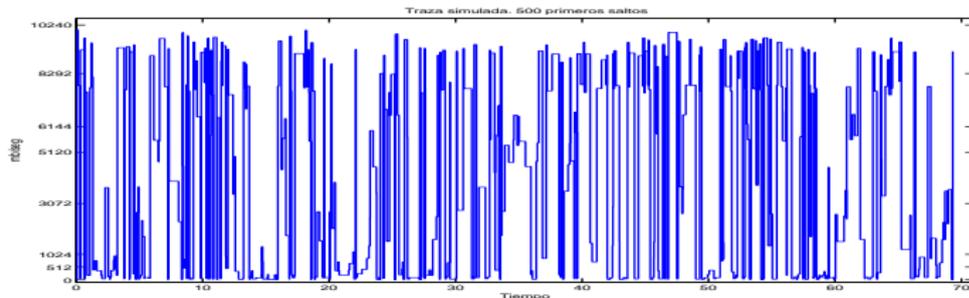
$$\mu_i^{(n)} = \frac{\sum_{r=1}^{N_i(n)} Y_i^r}{N_i(n)},$$

$N_i(n)$ la cantidad de veces que la cadena Z entró en el estado i en el intervalo $[0, n]$ (promedio de las velocidades de despacho en el estado i para $[0, n]$).

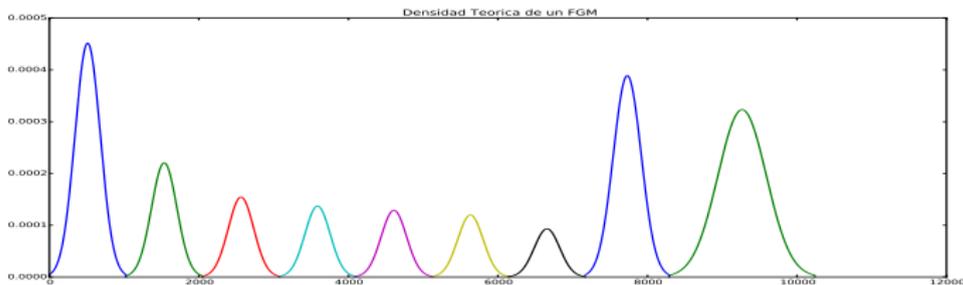
- Estimador para la π (distribución invariante)
Area debajo de la densidad en los extremos de cada intervalo de despacho.

Traza y estimación de un FMG

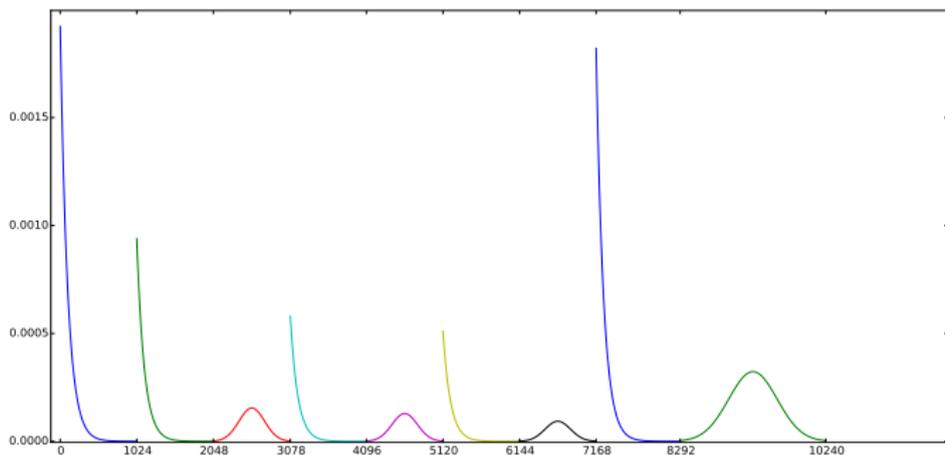
- A partir de una traza de tráfico de un FMG, debemos determinar $\hat{\pi}$, \hat{Q} y \hat{h}_i , sin conocimiento previo.



- Densidad de un FMG: Ejemplo 1



Ejemplo 2



- Cantidad de estados de la cadena \approx cantidad de picos y valles de la densidad estimada
- Soportes para cada estado de transmisión \approx con los mínimos relativos consecutivos
- La distribución invariante $\pi \approx$ area debajo de la densidad estimada entre dos valles.
- Valores medios de transmisión por estado (h_i) \approx Máximos relativos en las simétricas, en las asimétricas...

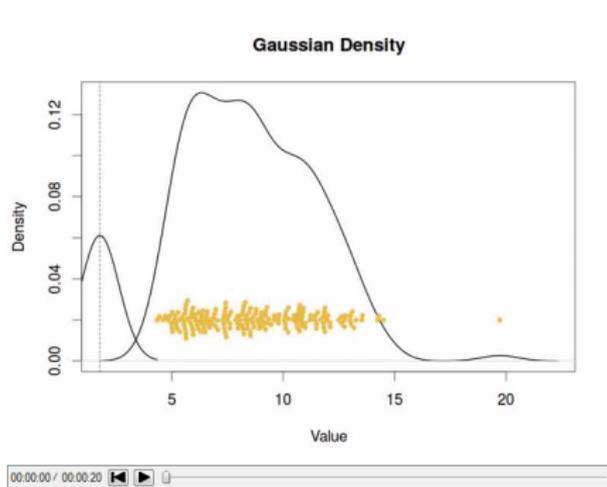
Estimación por Núcleos

- Abordaje por núcleos para estimar la densidad del FGM, que nos permite estimar $\hat{\pi}$, \hat{Q} y \hat{h}_i
- Un núcleo es una función $k(x)$ que:
 - Es real no negativa, integrable y acotada
 - Simétrica alrededor de 0
 - Integral es 1.
- Para una muestra aleatoria x_i de tamaño N con densidad $f(x)$, la estimación por núcleo de f es

$$\hat{f}(x) = \frac{1}{Nh} \sum_{i=1}^N k\left(\frac{x - x_i}{h}\right).$$

Animación de cómo funciona un núcleo

Video <https://www.youtube.com/watch?v=EFVUK-2rcso>



- h es el ancho de ventana importante y hay diversas maneras de estimarlo
- Ventana muy pequeña muchos extremos espureos, ventana muy grande puede colapsar estados consecutivos con rango “chico”.

Input: traza de tráfico de la fuente a estimar

- Realizar la estimación por núcleos de la densidad \hat{f} .
- Con los valores máximos se estima \hat{h}_i , y con la cantidad de estos se determina la cantidad de estados de la cadena.
- Con los mínimos consecutivos se determinan los rangos de transmisión de cada estado.
- Con el área entre estos mínimos estimamos $\hat{\pi}_i$
- Con esta información se reconstruye la cadena modulante (cada instante se asigna el estado correspondiente) y se calculan los estimadores de Lebedev-Lukashuk para Q .

Simulaciones y resultados numéricos I

- Trazas simuladas (datos reales trabajo a futuro)
- La cadena modulante tiene $k = 9$ estados asociados a un intervalo de velocidad de transferencia de datos

State	Transfer speed (Mbps)	Distribution
1	(0, 1024]	exp
2	(1024, 2048]	exp
3	(2048, 3072]	normal
4	(3072, 4096]	exp
5	(4096, 5120]	normal
6	(5120, 6144]	exp
7	(6144, 7168]	normal
8	(7168, 8292]	exp
9	(8292, 10240]	normal

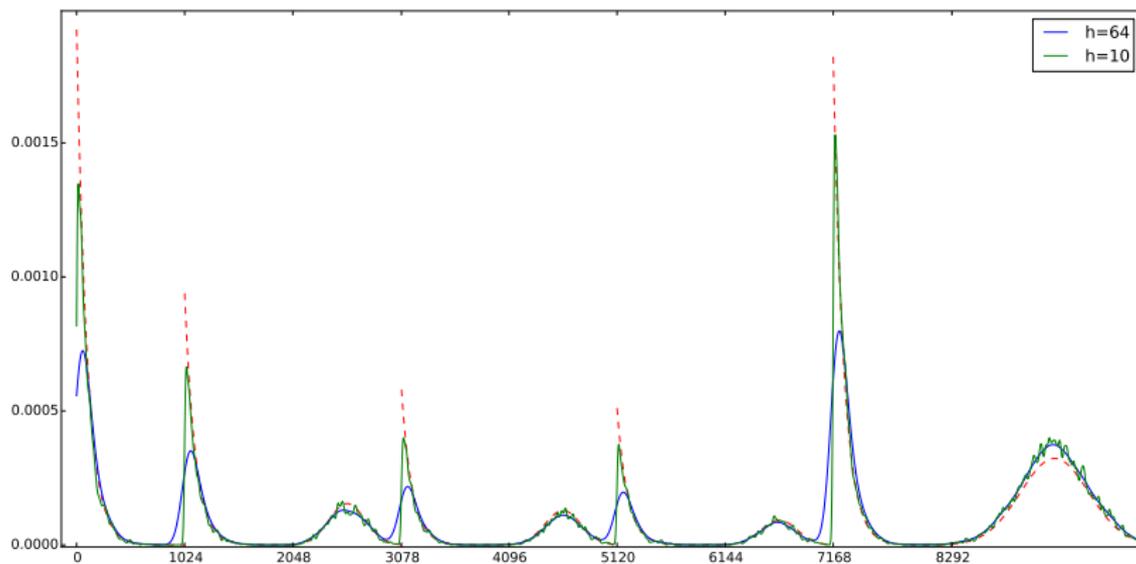
Simulaciones y resultados numéricos II

- El estado más probable es el de la mayor tasa de transferencia disponible en el canal de transmisión.
- Es más habitual saltar a los estados adyacentes o bien al de tasa máxima o mínima de transferencia.

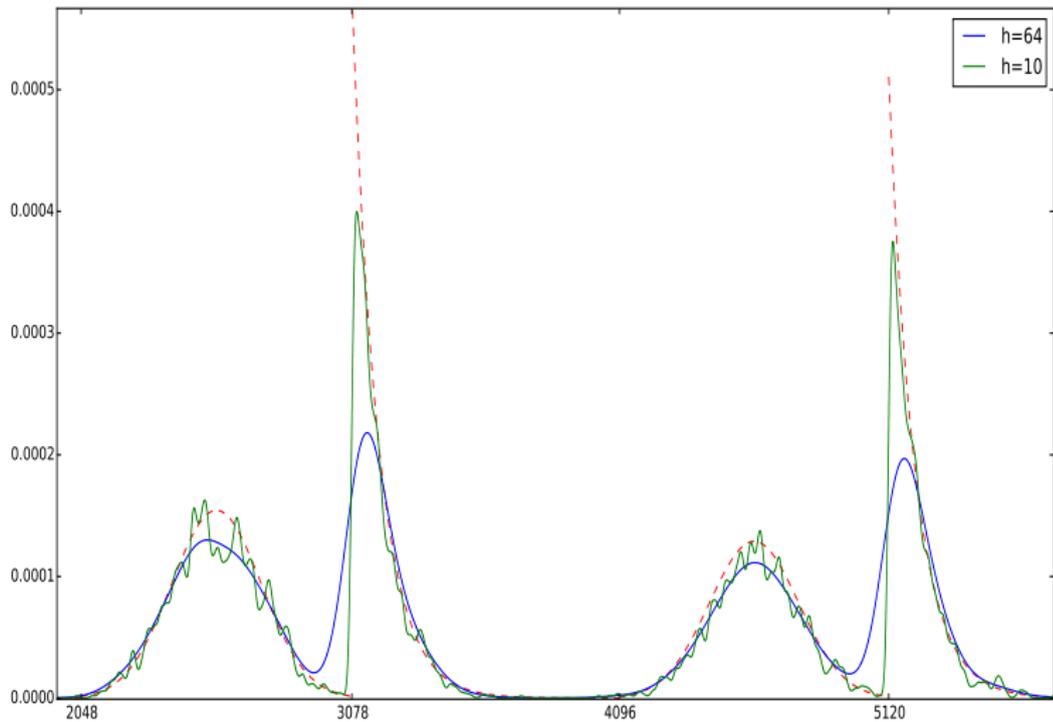
$$Q = \begin{pmatrix} -7 & 2 & 0,13 & 0,13 & 0,13 & 0,13 & 0,13 & 0,13 & 4,22 \\ 2 & -7 & 2 & 0,13 & 0,13 & 0,13 & 0,13 & 0,13 & 2,35 \\ 1 & 2 & -7 & 2 & 0,13 & 0,13 & 0,13 & 0,13 & 1,48 \\ 1 & 0,125 & 2 & -7 & 2 & 0,13 & 0,13 & 0,13 & 1,485 \\ 1 & 0,125 & 0,13 & 2 & -7 & 2 & 0,13 & 0,13 & 1,485 \\ 1 & 0,125 & 0,13 & 0,13 & 2 & -7 & 2 & 0,13 & 1,485 \\ 1 & 0,125 & 0,13 & 0,13 & 0,13 & 2 & -7 & 2 & 1,485 \\ 2 & 0,20 & 0,20 & 0,20 & 0,20 & 0,20 & 0,20 & -8 & 4,8 \\ 2 & 0,30 & 0,30 & 0,30 & 0,30 & 0,30 & 0,30 & 5 & -8,8 \end{pmatrix}$$

Gráficos I Densidad estimada vs Densidad Teórica

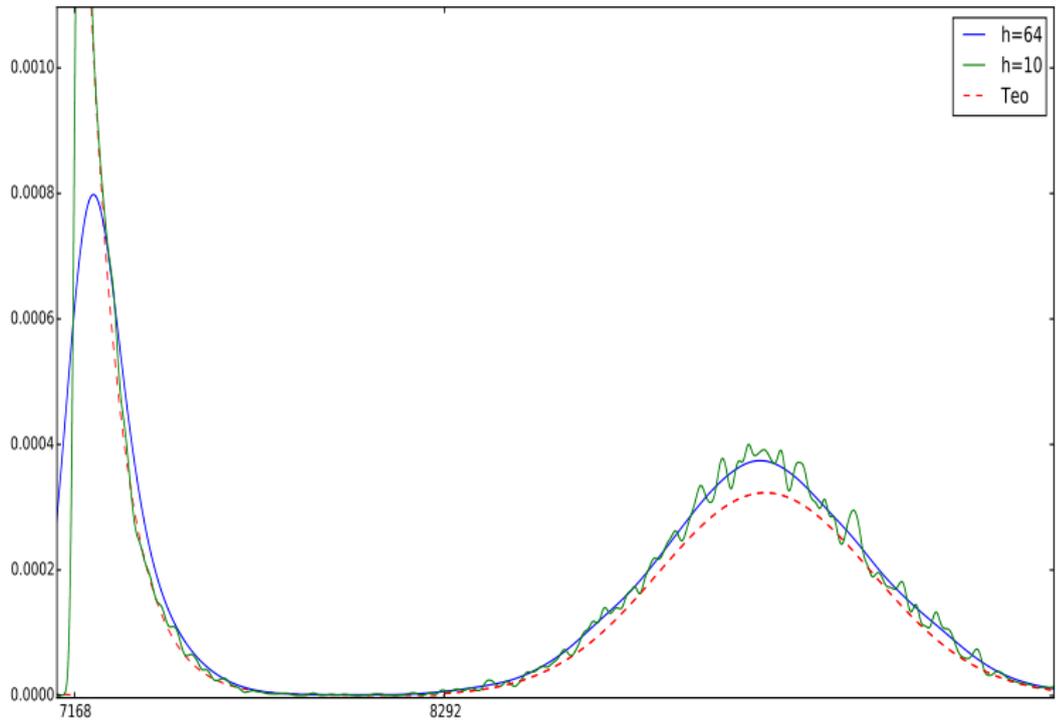
- 50000 saltos, $T_{max} = 7114$ segundos, Python 2.7



Gráficos II



Gráficos III



- Q teórico - Q estimado

-0,026	0,020	0,023	0,001	0,011	-0,014	-0,022	-0,042	0,049
0,137	-0,149	-0,028	-0,051	0,009	0,005	-0,027	0,005	0,100
0,091	-0,033	-0,195	0,007	0,013	-0,010	0,025	0,043	0,058
0,033	-0,038	0,223	-0,054	0,033	-0,092	-0,027	0,025	-0,044
0,012	-0,029	0,035	-0,005	0,309	-0,129	-0,031	0,006	-0,161
-0,038	0,025	-0,042	-0,020	0,132	-0,144	0,104	-0,027	0,054
0,033	0,010	0,025	0,006	-0,014	0,076	0,315	-0,250	-0,200
0,274	0,013	0,046	0,046	0,040	0,018	0,027	-0,936	0,488
0,433	0,073	0,059	0,058	0,041	0,069	0,039	1,066	-1,836

- π teórica vs $\hat{\pi}$

	1	2	3	4	5	6	7	8	9
π	0,192	0,093	0,066	0,057	0,054	0,051	0,039	0,182	0,262
$\hat{\pi}$	0.141	0.085	0.058	0.053	0.050	0.048	0.038	0.193	0.302

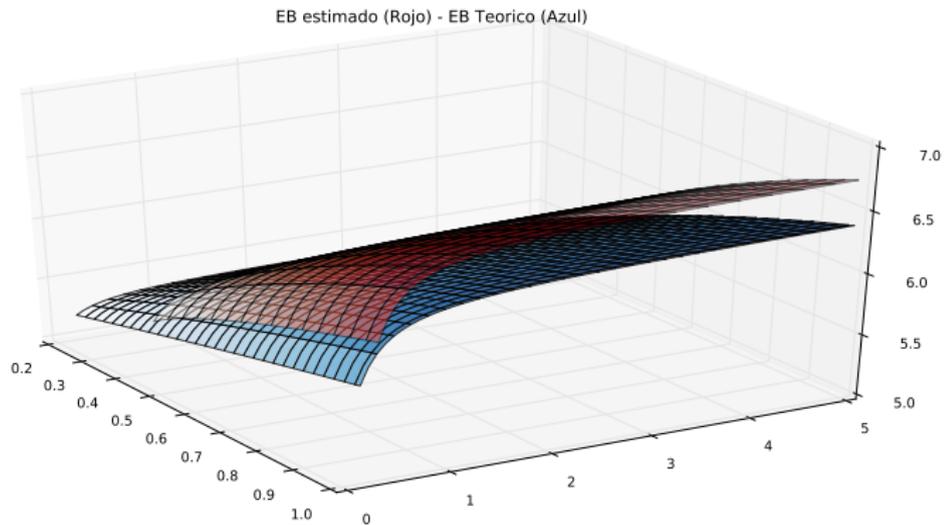
- Rangos teóricos vs Rangos estimados

Estado	Rango(Mbps)	Rango Estimado
1	[0, 1024]	[0,795)
2	(1024, 2048]	(795,1931)
3	(2048, 3072]	1931,2934)
4	(3072, 4096]	(2934,3888)
5	(4096, 5120]	(3888,4972)
6	(5120, 6144]	(4972,5909)
7	(6144, 7168]	(5909,6985)
8	(7168, 8292]	(6985,8009)
9	(8292, 10240]	(8009,10240]

- H teórico vs H estimado

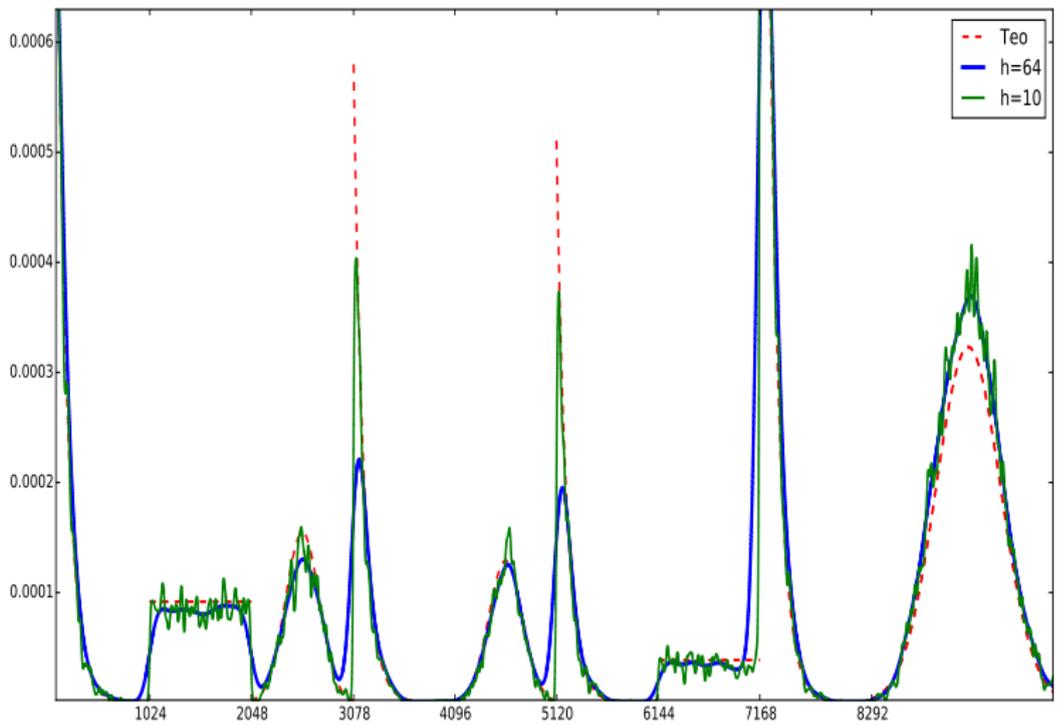
H teo	100	1124	2563	3172	4608	5220	6656	7268	9266
H est1	55	1082	2537	3133	4625	5179	6656	7225	9270
Dif	45	42	26	39	-17	41	0	43	-4

Simulaciones y resultados numéricos



Conclusiones y trabajo futuro

- Conclusiones
 - Modelo ajusta correctamente. Exponenciales
 - No necesita conocimiento previo específico. Aplicabilidad.
 - Determinación de h (ventana del núcleo)
- Trabajos Futuros
 - Distribuciones multimodales o uniforme en cada estado (semi supervisado)
 - Soportes no disjuntos
 - Datos Reales



MUCHAS GRACIAS!

- Bavio, J.M and Marrón, B.S. (2018). Properties of the Estimators for the Effective Bandwidth in a Generalized Markov Fluid Model.OJS, Vol.8 No.1, pp. 69-84.
- Marrón, B.S. (2012). Estadística de Procesos Estocásticos aplicados a Redes de Datos y Telecomunicación. Ph.D. Thesis. Departamento de Matemática, Universidad Nacional del Sur, Argentina.
<http://repositoriodigital.uns.edu.ar/bitstream/123456789/2302/1/Marron-Beatriz-Tesis.pdf>.
- Perera, G. et al. (2002). *Effective bandwidth estimation and testing for Markov sources*. Performance Evaluation 45, p. 1-19.