

UNIVERSIDAD NACIONAL DE EDUCACIÓN A DISTANCIA

ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA INFORMÁTICA

Master en Ingeniería de Sistemas y Control

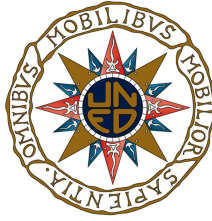
**DETECCIÓN DE CAMBIOS EN TIEMPO REAL EN
FLUJOS DE DATOS**

TOMÁS CRUZ BARRIOS

Dirigido por: Dr. JESÚS ANTONIO VEGA SÁNCHEZ

Supervisado por: Dr. SEBASTIÁN DORMIDO CANTO

Curso: 2013/2014 (Convocatoria Junio)



MASTER EN INGENIERÍA DE SISTEMAS Y CONTROL

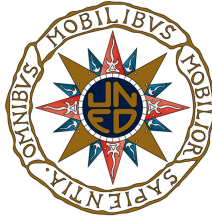
**DETECCIÓN DE CAMBIOS EN TIEMPO REAL EN FLUJOS
DE DATOS**

Proyecto de Fin de Master de modalidad oferta específica propuesto por un
profesor

Realizado por: TOMÁS CRUZ BARRIOS

Dirigido por: Dr. JESÚS ANTONIO VEGA SÁNCHEZ

Supervisado por: Dr. SEBASTIÁN DORMIDO CANTO



DETECCIÓN DE CAMBIOS EN TIEMPO REAL EN FLUJOS DE DATOS

Master en Ingeniería de Sistemas y Control

Realizado por: TOMÁS CRUZ BARRIOS

Dirigido por: Dr. JESÚS ANTONIO VEGA SÁNCHEZ

Supervisado por: Dr. SEBASTIÁN DORMIDO CANTO

Tribunal calificador:

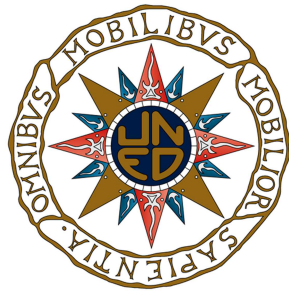
Presidente: D./D^a.
(firma)

Secretario: D./D^a.
(firma)

Vocal: D./D^a.
(firma)

Fecha de lectura y defensa:

Calificación:



Autorización

Autorizamos a la Universidad Complutense y a la UNED a difundir y utilizar con fines académicos, no comerciales y mencionando expresamente a sus autores, tanto la memoria de este Trabajo Fin de Máster, como el código, la documentación y/o el prototipo desarrollado.

Firmado: Tomás Cruz Barrios

Firma del alumno

A mi padre...

Agradecimientos

Quiero agradecer al Dr. Sebastián Dormido Canto el haberse cruzado en mi camino y convencerme para que me enrolara en la aventura de cursar este Master. Gracias a ello he conseguido progresar personal y profesionalmente. Nunca se lo agradeceré lo suficiente.

Ha sido para mí un privilegio haber sido discípulo de una persona del prestigio y el conocimiento del Dr. Jesús A. Vega Sánchez. Gracias a su guía he adquirido un método de trabajo que me ha permitido conseguir nuevos logros que nunca antes imaginé alcanzar. Adquiriendo además nuevas habilidades que me permiten afrontar nuevos retos profesionales con la seguridad y confianza necesarios para alcanzar el éxito.

Gracias a mi familia, Silvia, Tomás y Valeria, por los momentos de ausencia que habéis tenido que sufrir mientras realizaba este Master. Silvia, sin tu apoyo y aliento nunca lo hubiera conseguido.

Gracias a mis padres por educarme en la cultura del esfuerzo.

Resumen

Los plasmas de fusión son sistemas físicos regidos por complejas interacciones altamente no lineales y que necesitan diagnosticarse de la mejor manera posible. La base de datos de cualquier dispositivo actual se compone de millones de señales. La mayoría de estas señales son de evolución temporal, con un número de muestras entre 10.000 y varios millones. Por ejemplo, en el tokamak JET (con una duración de descarga de 40 s), se han realizado más de 70.000 disparos y el número de señales que se adquieren en cada descarga puede ser superior a 10.000. En la actualidad, el tamaño de la base de datos de JET es de 40 Tbytes y su tendencia de crecimiento refleja que la cantidad de datos adquiridos se duplica cada dos años. Por tanto, es necesaria la inclusión de sistemas automáticos que permitan un tratamiento eficiente y global de las bases de datos de fusión.

Por lo expuesto y dada la enorme cantidad de información a tratar, las técnicas de aprendizaje automático (*Machine Learning*) resultan de gran ayuda en el análisis de señales. En particular, las técnicas de reconocimiento de cambios en flujos de datos son útiles tanto en análisis en tiempo real como off-line. Este proyecto consiste en el desarrollo de una aplicación que muestre la efectividad de varios algoritmos, basados en martingalas, para la detección de cambios. El sistema automático de reconocimiento es general y no depende un evento particular o representación específica (formas de onda, imágenes o señales multidimensionales). La metodología desarrollada facilita buenos resultados para la detección de cambios, tanto en la evolución temporal de perfiles, como la distribución espacial en 2-D de la intensidad de las emisiones de volumen.

Lista de palabras clave:

Detección de cambios, tiempo-real, machine learning, plasmas de fusión nuclear, martingalas.

Índice general

Índice de tablas	iv
Índice de figuras	ix
1. Fusión nuclear: la solución energética del futuro.	1
1.1. Necesidades energéticas mundiales y el problema del cambio climático	1
1.2. Soluciones energéticas viables: Renovables y Energía nuclear. .	2
1.3. Reacciones nucleares	4
1.4. Plasmas de fusión	7
1.5. Confinamiento Magnético	8
1.6. TJ-II y JET	9
1.6.1. Diagnósis	10
2. Objetivo y alcance del proyecto.	13
2.1. Minería de datos aplicada a la ciencia e ingeniería	13
2.1.1. Minería de datos aplicada al estudio de la fusión nuclear	14

2.2. Objetivo del trabajo de fin de máster	14
2.2.1. Introducción	14
2.2.2. Descripción del objetivo	15
3. Fundamentos teóricos y relación con otros trabajos.	19
3.1. Clasificación y definición de los diferentes tipos de cambio . . .	19
3.2. Requerimientos de un algoritmo para el análisis de flujos de datos	20
3.3. Escenarios estándares de aprendizaje automático	21
3.4. Predicciones inductivas y transductivas	22
3.5. Predictores conformales	23
3.5.1. Introducción	23
3.5.2. Medida de no conformidad	24
3.5.3. Medidas de <i>extrañeza</i>	25
3.5.4. Cálculo on-line de p-valores	27
3.6. Test de Intercambiabilidad On-Line	28
3.6.1. Conceptos de aleatoriedad e intercambiabilidad	28
3.6.2. Martingalas	29
3.6.3. Supermartingalas aleatorias e intercambiables	29
3.6.4. Martingalas intercambiables aleatorias	30
3.6.5. Martingalas basadas en p-valores	30
3.7. Test de martingalas para la detección de cambios.	40

3.7.1.	Algoritmo para el Test de Martingalas	41
3.7.2.	Rendimiento del test de martingala	43
3.8.	Test de múltiples martingalas utilizando multi-vistas	43
4.	Prueba de capacidades del test de martingalas	45
4.1.	Descripción de los experimentos	45
4.2.	Análisis de sensibilidad	46
4.3.	Experimentos de detección de cambios en perfiles	46
4.3.1.	Experimento 1	49
4.3.2.	Experimento 2	54
4.3.3.	Experimento 3	59
4.3.4.	Experimento 4	65
4.3.5.	Experimento 5	70
4.3.6.	Experimento 6	75
4.3.7.	Experimento 7	83
4.3.8.	Experimento 8	91
4.3.9.	Experimento 9	99
4.3.10.	Experimento 10	104
4.3.11.	Experimento 11	109
4.3.12.	Experimento 12	114
4.4.	Conclusiones	119

5. Capacidad de detección de cambios en perfiles de fusión nuclear	121
5.1. Perfiles de Bolometría del TJ-II (disparo 21319)	121
5.1.1. Randomized Power	122
5.1.2. Simple Mixture	123
5.1.3. Plug-in Martingale	123
5.1.4. Preprocesamiento de datos	123
5.2. Conclusiones	124
6. Conclusiones y futuros trabajos	125
6.1. Conclusiones	125
6.2. Futuros Trabajos	126
Bibliografía	126

Índice de cuadros

3.1. Ejemplos de Kernels para la estimación de la densidad.	37
4.1. Media de retardos, FA y cambios no detectados para un conjunto de 100 ejecuciones de la Randomized Power Martingale y varios valores de ϵ	51
4.2. Media de retardos, FA y cambios no detectados para un conjunto de 100 ejecuciones de la Simple Mixture Martingale.	52
4.3. Media de retardos, FA y cambios no detectados para un conjunto de 100 ejecuciones de la Plug-in Martingale.	53
4.4. Media de retardos, FAR's y cambios no detectados para un conjunto de 100 ejecuciones de la Randomized Power Martingale y varios valores de ϵ	56
4.5. Media de retardos, FAR's y cambios no detectados para un conjunto de 100 ejecuciones de la Simple Mixture Martingale.	57
4.6. Media de retardos, FAR's y cambios no detectados para un conjunto de 100 ejecuciones de la Plug-in Martingale.	58
4.7. Media de retardos, FA y cambios no detectados para un conjunto de 100 ejecuciones de la Randomized Power Martingale y varios valores de ϵ	61
4.8. Media de retardos, FA y cambios no detectados para un conjunto de 100 ejecuciones de la Simple Mixture Martingale.	62

4.9. Media de retardos, FA y cambios no detectados para un conjunto de 100 ejecuciones de la Plug-in Martingale.	63
4.10. Media de retardos, FA y cambios no detectados para un conjunto de 100 ejecuciones de la Randomized Power Martingale y varios valores de ϵ	67
4.11. Media de retardos, FA y cambios no detectados para un conjunto de 100 ejecuciones de la Simple Mixture Martingale.	68
4.12. Media de retardos, FA y cambios no detectados para un conjunto de 100 ejecuciones de la Plug-in Martingale.	69
4.13. Media de retardos, FA y cambios no detectados para un conjunto de 100 ejecuciones de la Randomized Power Martingale y varios valores de ϵ	72
4.14. Media de retardos, FA y cambios no detectados para un conjunto de 100 ejecuciones de la Simple Mixture Martingale.	73
4.15. Media de retardos, FA y cambios no detectados para un conjunto de 100 ejecuciones de la Plug-in Martingale.	74
4.16. Media de retardos para el primer cambio, FA y cambios no detectados para un conjunto de 100 ejecuciones de la Randomized Power Martingale y varios valores de ϵ	77
4.17. Media de retardos para el segundo cambio, FA y cambios no detectados para un conjunto de 100 ejecuciones de la Randomized Power Martingale y varios valores de ϵ	78
4.18. Media de retardos, FA y cambios no detectados, para el primer cambio, para un conjunto de 100 ejecuciones de la Simple Mixture Martingale.	79
4.19. Media de retardos, FA y cambios no detectados, para el segundo cambio, para un conjunto de 100 ejecuciones de la Simple Mixture Martingale.	80

4.20. Media de retardos, FA y cambios no detectados, para el primer cambio, para un conjunto de 100 ejecuciones de la Plug-in Martingale.	81
4.21. Media de retardos, FA y cambios no detectados, para el segundo cambio, para un conjunto de 100 ejecuciones de la Plug-in Martingale.	81
4.22. Media de retardos para el primer cambio, FA y cambios no detectados, para el primer cambio, para un conjunto de 100 ejecuciones de la Randomized Power Martingale y varios valores de ϵ	85
4.23. Media de retardos para el primer cambio, FA y cambios no detectados, para el segundo cambio, para un conjunto de 100 ejecuciones de la Randomized Power Martingale y varios valores de ϵ	86
4.24. Media de retardos, FA y cambios no detectados, para el primer cambio, para un conjunto de 100 ejecuciones de la Simple Mixture Martingale.	87
4.25. Media de retardos, FA y cambios no detectados, para el segundo cambio, para un conjunto de 100 ejecuciones de la Simple Mixture Martingale.	88
4.26. Media de retardos, FA y cambios no detectados, para el primer cambio, para un conjunto de 100 ejecuciones de la Plug-in Martingale.	89
4.27. Media de retardos, FA y cambios no detectados, para el segundo cambio, para un conjunto de 100 ejecuciones de la Plug-in Martingale.	89
4.28. Media de retardos para perfiles sin ruido, FA y cambios no detectados para un conjunto de 100 ejecuciones de la Randomized Power Martingale y varios valores de ϵ	94

4.29. Media de retardos para perfiles con ruido, FA y cambios no detectados para un conjunto de 100 ejecuciones de la Randomized Power Martingale y varios valores de ϵ para perfiles con ruido.	95
4.30. Media de retardos para perfiles sin ruido, FA y cambios no detectados para un conjunto de 100 ejecuciones de la Simple Mixture Martingale.	96
4.31. Media de retardos con ruido, FA y cambios no detectados para un conjunto de 100 ejecuciones de la Simple Mixture Martingale.	97
4.32. Media de retardos para perfiles sin ruido, FA y cambios no detectados para un conjunto de 100 ejecuciones de la Plug-in Martingale.	97
4.33. Media de retardos, FA y cambios no detectados para un conjunto de 100 ejecuciones de la Plug-in Martingale.	98
4.34. Media de retardos para perfiles suavizados, FAR's y cambios no detectados para un conjunto de 100 ejecuciones de la Randomized Power Martingale y varios valores de ϵ	101
4.35. Media de retardos para perfiles suavizados, FAR's y cambios no detectados para un conjunto de 100 ejecuciones de la Simple Mixture Martingale.	102
4.36. Media de retardos para perfiles suavizados, FAR's y cambios no detectados para un conjunto de 100 ejecuciones de la Plug-in Martingale.	103
4.37. Media de retardos para perfiles sin ruido, FA y cambios no detectados para un conjunto de 100 ejecuciones de la Randomized Power Martingale y varios valores de ϵ	106
4.38. Media de retardos para perfiles sin ruido, FA y cambios no detectados para un conjunto de 100 ejecuciones de la Simple Mixture Martingale.	107

4.39. Media de retardos para perfiles sin ruido, FAR's y cambios no detectados para un conjunto de 100 ejecuciones de la Plug-in Martingale.	108
4.40. Media de retardos, FA y cambios no detectados para un conjunto de 100 ejecuciones de la Randomized Power Martingale y varios valores de ϵ	111
4.41. Media de retardos, FA y cambios no detectados para un conjunto de 100 ejecuciones de la Simple Mixture Martingale. . .	112
4.42. Media de retardos para perfiles sin ruido, FA y cambios no detectados para un conjunto de 100 ejecuciones de la Plug-in Martingale.	113
4.43. Media de retardos, FA y cambios no detectados para un conjunto de 100 ejecuciones de la Randomized Power Martingale y varios valores de ϵ	116
4.44. Media de retardos, FA y cambios no detectados para un conjunto de 100 ejecuciones de la Simple Mixture Martingale. . .	117
4.45. Media de retardos para perfiles sin ruido, FA y cambios no detectados para un conjunto de 100 ejecuciones de la Plug-in Martingale.	118
5.1. Media de retardos, desviación estándar y cambios no detectados para un conjunto de 100 ejecuciones de la Randomized Power Martingale y varios valores de ϵ	122
5.2. Media de retardos, desviación estándar y cambios no detectados para un conjunto de 20 ejecuciones de la Simple Mixture Martingale.	123
5.3. Media de retardos, desviación estándar y cambios no detectados para un conjunto de 20 ejecuciones de la Plug-in Martingale.	123

Índice de figuras

1.1. Gráfico de las energías de enlace nuclear según la masa atómica. Se puede observar como el rendimiento para la fusión es mucho mayor que para la fisión.	5
1.2. Representación esquemática de la reacción de fusión D-T. Un núcleo de deuterio y otro de tritio se unen para dar lugar a un núcleo de helio y un neutrón, liberándose una energía de 17.58 MeV.	7
1.3. Representación esquemática de un tokamak. El plasma se representa en color rosado.	9
1.4. Stellarator TJ-II. El plasma se representa en color fucsia. Se pueden apreciar las ventanas de observación para diagnosis del plasma.	10
1.5. Localización y representación esquemática de los principales diagnósticos del stellarator TJ-II. También se muestran las dos líneas de calentamiento de ECRH (GR1 y GR2), así como las líneas de inyección de partículas neutras para calentamiento (NBI).	11
1.6. Localización y representación esquemática de los principales diagnósticos del Tokamak JET.	11
2.1. Simulación de cambios en flujos de datos. Los cambios mostrados son: (a) cambio a una amplitud mayor; (b) mayor ancho; (c) más ruido	15

2.2. Simulación de líneas de contorno de emisión del plasma con diferentes niveles de fluctuación.	16
2.3. Evolución temporal de perfiles de bolometría. Perfiles de bolometría del stellarator TJ-II (disparo 21319).	17
2.4. (a) Idealmente, el vector de características debería ser seleccionado de forma que los dos estados sean separables en el espacio de características. (b) Los datos son leídos de forma secuencial, uno a uno.	18
3.1. Predicción inductiva y transductiva.	22
3.2. Valores de función de apuesta que se utilizan para calcular Randomized Power y Simple Mixture Martingale.	32
3.3. Cuatro estimaciones del kernel de densidad utilizando $n = 100$ y variables generadas aleatoriamente siguiendo distribución normal. observese que cuanto menor es el valor de h , la estimación es más tosca.	36
3.4. Gráficos de varios kernels que pueden utilizarse para estimar la función de densidad. a) Triángulo, b) Epanechnikov, c) Biweight, d) Triweight, e) Normal o Gaussiano, f) k_2 Silverman	39
4.1. Ejemplos de efectos de la variable α_1 de la función 4.1.	46
4.2. Ejemplos de efectos de la variable α_2 de la función 4.1.	47
4.3. Ejemplos de efectos de la variable A_0 en la función 4.1.	48
4.4. Ejemplos de efectos de ruido simulado para la función 4.1.	48
4.5. Ejemplos de suavizado para los casos de la figura 4.4 con $\alpha_1 = 4, \alpha_2 = 4$	48
4.6. Perfiles Experimento 1.	49

4.7. Ejemplos de detección del experimento 1. Se indica en cada gráfico el momento en el que la arquitectura de martingalas identifica el cambio.	49
4.8. Ejemplos de detección del experimento 1. Se indica en cada gráfico el momento en el que la arquitectura de martingalas identifica el cambio.	50
4.9. Gráfico Box Plot para 100 ejecuciones de Ramdomized Power Martingales por cada una de las 10 ϵ	51
4.10. Gráfico Box Plot para 100 ejecuciones.	52
4.11. Gráfico Box Plot para 100 ejecuciones de Plug-In Martingale y 5 factores (FCT) diferentes.	53
4.12. Perfiles Experimento 2.	54
4.13. Ejemplos de detección del experimento 2. Se indica en cada gráfico el momento en el que la arquitectura de martingalas identifica el cambio.	54
4.14. Ejemplos de detección del experimento 2. Se indica en cada gráfico el momento en el que la arquitectura de martingalas identifica el cambio.	55
4.15. Gráfico Box Plot para 100 ejecuciones de Ramdomized Power Martingale por cada una de las 10 ϵ	56
4.16. Gráfico Box Plot para 100 ejecuciones.	57
4.17. Gráfico Box Plot para 100 ejecuciones de Plug-In Martingale y 5 factores (FCT) diferentes de multiplicación de h (Normal Reference Rule).	58
4.18. Perfiles Experimento 3.	59
4.19. Ejemplos de detección del experimento 3. Se indica en cada gráfico el momento en el que la arquitectura de martingalas identifica el cambio.	59

4.20. Ejemplos de detección del experimento 3. Se indica en cada gráfico el momento en el que la arquitectura de martingalas identifica el cambio.	60
4.21. Gráfico Box Plot para 100 ejecuciones de Ramdomized Power Martingale por cada una de las 10 ϵ	61
4.22. Gráfico Box Plot para 100 ejecuciones.	62
4.23. Gráfico Box Plot para 100 ejecuciones de Plug-In Martingale y 5 factores (FCT) diferentes de multiplicación de h (Normal Reference Rule).	63
4.24. Perfiles Experimento 4.	65
4.25. Ejemplos de detección del experimento 4. Se indica en cada gráfico el momento en el que la arquitectura de martingalas identifica el cambio.	65
4.26. Ejemplos de detección del experimento 4. Se indica en cada gráfico el momento en el que la arquitectura de martingalas identifica el cambio.	66
4.27. Gráfico Box Plot para 100 ejecuciones de Ramdomized Power Martingale por cada una de las 10 ϵ	67
4.28. Gráfico Box Plot para 100 ejecuciones.	68
4.29. Gráfico Box Plot para 100 ejecuciones de Plug-In Martingale y 5 factores (FCT) diferentes de multiplicación de h (Normal Reference Rule).	69
4.30. Perfiles Experimento 5.	70
4.31. Ejemplos de detección del experimento 5. Se indica en cada gráfico el momento en el que la arquitectura de martingalas identifica el cambio.	70
4.32. Ejemplos de detección del experimento 5. Se indica en cada gráfico el momento en el que la arquitectura de martingalas identifica el cambio.	71

4.33. Gráfico Box Plot para 100 ejecuciones de Ramdomized Power Martingale por cada una de las 10 ϵ	72
4.34. Gráfico Box Plot para 100 ejecuciones.	73
4.35. Gráfico Box Plot para 100 ejecuciones de Plug-In Martingale y 5 factores (FCT) diferentes de multiplicación de h (Normal Reference Rule).	74
4.36. Perfiles Experimento 6.	75
4.37. Ejemplos de detección del experimento 6. Se indica en cada gráfico el momento en el que la arquitectura de martingalas identifica el cambio.	75
4.38. Ejemplos de detección del experimento 6. Se indica en cada gráfico el momento en el que la arquitectura de martingalas identifica el cambio.	76
4.39. Gráfico Box Plot, para el primer cambio, para 100 ejecuciones de Ramdomized Power Martingale por cada una de las 10 ϵ	77
4.40. Gráfico Box Plot, para el segundo cambio, para 100 ejecuciones de Ramdomized Power Martingale por cada una de las 10 ϵ	78
4.41. Gráfico Box Plot del primer cambio para 100 ejecuciones.	79
4.42. Gráfico Box Plot del segundo cambio para 100 ejecuciones.	79
4.43. Gráfico Box Plot del primer cambio para 100 ejecuciones de Plug-In Martingale y 5 factores (FCT) diferentes de multiplicación de h (Normal Reference Rule).	80
4.44. Gráfico Box Plot del segundo cambio para 100 ejecuciones de Plug-In Martingale y 5 factores (FCT) diferentes de multiplicación de h (Normal Reference Rule).	81
4.45. Perfiles Experimento 7.	83

4.46. Ejemplos de detección del experimento 7. Se indica en cada gráfico el momento en el que la arquitectura de martingalas identifica el cambio.	83
4.47. Ejemplos de detección del experimento 7. Se indica en cada gráfico el momento en el que la arquitectura de martingalas identifica el cambio.	84
4.48. Gráfico Box Plot del primer cambio para 100 ejecuciones de Randomized Power Martingale por cada una de las 10 ϵ	85
4.49. Gráfico Box Plot del segundo cambio para 100 ejecuciones de Randomized Power Martingale por cada una de las 10 ϵ	86
4.50. Gráfico Box Plot del primer cambio para 100 ejecuciones. . . .	87
4.51. Gráfico Box Plot del segundo cambio para 100 ejecuciones. . . .	87
4.52. Gráfico Box Plot del primer cambio, para el primer cambio, para 100 ejecuciones de Plug-In Martingale y 5 factores (FCT) diferentes de multiplicación de h (Normal Reference Rule). . . .	88
4.53. Gráfico Box Plot del segundo cambio para 100 ejecuciones de Plug-In Martingale y 5 factores (FCT) diferentes de multiplicación de h (Normal Reference Rule).	89
4.54. Perfiles Experimento 7.	91
4.55. Ejemplos de detección del experimento 7. Se indica en cada gráfico el momento en el que la arquitectura de martingalas identifica el cambio.	91
4.56. Ejemplos de detección del experimento 7. Se indica en cada gráfico el momento en el que la arquitectura de martingalas identifica el cambio.	92
4.57. Perfiles Experimento 7.	92
4.58. Ejemplos de detección del experimento 7. Se indica en cada gráfico el momento en el que la arquitectura de martingalas identifica el cambio.	93

4.59. Ejemplos de detección del experimento 7. Se indica en cada gráfico el momento en el que la arquitectura de martingalas identifica el cambio.	93
4.60. Gráfico Box Plot para 100 ejecuciones de Ramdomized Power Martingale por cada una de las 10 ϵ y perfiles sin ruido.	94
4.61. Gráfico Box Plot para 100 ejecuciones de Ramdomized Power Martingale por cada una de las 10 ϵ para perfiles con ruido.	95
4.62. Gráfico Box Plot para 100 ejecuciones.	96
4.63. Gráfico Box Plot para 100 ejecuciones.	96
4.64. Gráfico Box Plot para 100 ejecuciones de Plug-In Martingale y 5 factores (FCT) diferentes de multiplicación de h (Normal Reference Rule).	97
4.65. Gráfico Box Plot para 100 ejecuciones de Plug-In Martingale y 5 factores (FCT) diferentes de multiplicación de h (Normal Reference Rule).	98
4.66. Perfiles Experimento 9.	99
4.67. Ejemplos de detección del experimento 9. Se indica en cada gráfico el momento en el que la arquitectura de martingalas identifica el cambio.	99
4.68. Ejemplos de detección del experimento 9. Se indica en cada gráfico el momento en el que la arquitectura de martingalas identifica el cambio.	100
4.69. Gráfico Box Plot para 100 ejecuciones de Ramdomized Power Martingale por cada una de las 10 ϵ y perfiles suavizados.	101
4.70. Gráfico Box Plot para 100 ejecuciones.	102
4.71. Gráfico Box Plot para 100 ejecuciones de Plug-In Martingale y 5 factores (FCT) diferentes de multiplicación de h (Normal Reference Rule).	103

4.72. Perfiles Experimento 10.	104
4.73. Ejemplos de detección del experimento 10. Se indica en cada gráfico el momento en el que la arquitectura de martingalas identifica el cambio.	104
4.74. Ejemplos de detección del experimento 10. Se indica en cada gráfico el momento en el que la arquitectura de martingalas identifica el cambio.	105
4.75. Gráfico Box Plot para 100 ejecuciones de Ramdomized Power Martingale por cada una de las 10 ϵ y perfiles sin ruido.	106
4.76. Gráfico Box Plot para 100 ejecuciones.	107
4.77. Gráfico Box Plot para 100 ejecuciones de Plug-In Martingale y 5 factores (FCT) diferentes de multiplicación de h (Normal Reference Rule).	108
4.78. Perfiles Experimento 12.	109
4.79. Ejemplos de detección del experimento 11. Se indica en cada gráfico el momento en el que la arquitectura de martingalas identifica el cambio.	109
4.80. Ejemplos de detección del experimento 11. Se indica en cada gráfico el momento en el que la arquitectura de martingalas identifica el cambio.	110
4.81. Gráfico Box Plot para 100 ejecuciones de Ramdomized Power Martingales por cada una de las 10 ϵ	111
4.82. Gráfico Box Plot para 100 ejecuciones.	112
4.83. Gráfico Box Plot para 100 ejecuciones de Plug-In Martingale y 5 factores (FCT) diferentes de multiplicación de h (Normal Reference Rule).	113
4.84. Perfiles Experimento 12.	114

4.85. Ejemplos de detección del experimento 12. Se indica en cada gráfico el momento en el que la arquitectura de martingalas identifica el cambio.	114
4.86. Ejemplos de detección del experimento 12. Se indica en cada gráfico el momento en el que la arquitectura de martingalas identifica el cambio.	115
4.87. Gráfico Box Plot para 100 ejecuciones de Ramdomized Power Martingales por cada una de las 10 ϵ	116
4.88. Gráfico Box Plot para 100 ejecuciones.	117
4.89. Gráfico Box Plot para 100 ejecuciones de Plug-In Martingale y 5 factores (FCT) diferentes de multiplicación de h (Normal Reference Rule).	118
5.1. Perfil de bolometría del TJ-II (disparo 21319).	122
5.2. Perfiles obtenidos a partir del cálculo de pendientes de grupos de 3 perfiles.	124

Capítulo 1

Fusión nuclear: la solución energética del futuro.

1.1. Necesidades energéticas mundiales y el problema del cambio climático

La Agencia Internacional de la Energía (IEA), en su informe estadístico anual de 2013 [1], estima que, de seguir la tendencia actual, las necesidades de abastecimiento de energía para 2035 podrían llegar hasta 11.750 Mtpe¹. Bajo el escenario de referencia de la IEA² *World Energy Outlook*, los combustibles fósiles seguirán dominando como fuente primaria de energía, representando el 77 % de dicho incremento. El 90 % de esta demanda vendrá de países no pertenecientes a la OECD³, liderados por China, India y Países de Oriente Medio.

Estas tendencias, son inviables en relación con el medio ambiente, la seguridad energética y el desarrollo económico. Muchos países y regiones serán más dependientes de las importaciones de petróleo, aunque los combustibles obtenidos de gas de esquistos pueden cubrir gran parte de las necesidades energéticas, tal y como se tiene previsto para el caso de EE.UU. El sector del transporte concentrará gran parte de este crecimiento de la demanda, donde

¹Millones de toneladas de petróleo equivalente

²International Energy Agency

³Organisation for Economic Co-operation and Development

la elasticidad del precio es bajo, y los riesgos incrementarán la volatilidad del mercado. Se producirá un rápido incremento de las emisiones de CO_2 , como consecuencia del aumento del consumo de combustibles fósiles, alcanzando 37.037 Mt en 2035, lo que incrementará la temperatura global $3,6^\circ C$. Asimismo se producirá una subida del precio de los combustibles fósiles. Por ejemplo, se asume que los precios del petróleo llegarán a 128 USD ⁴ por barril (en precios del 2012) para el 2035.

Por otro lado, los últimos acontecimientos de Fukushima, tras el terremoto de Japón de 2011, han puesto de nuevo en entredicho la seguridad y viabilidad económica de las instalaciones nucleares de fisión nuclear y su posible contribución a la seguridad del suministro, estabilidad de precios de la energía y disminución de gases de efecto invernadero. Por lo que el escenario anteriormente descrito puede ser aún más desfavorable.

Tanto productores como consumidores de energía se enfrentan a la urgente necesidad de ir a sistemas globales de energía más sostenibles.

1.2. Soluciones energéticas viables: Renovables y Energía nuclear.

En cuanto a las soluciones viables, se destacan dos: las energías renovables y la energía nuclear. Las energías renovables son en general bien aceptadas por la sociedad al no plantear aparentemente riesgos para los trabajadores y ser consideradas amigables con el medioambiente. A pesar de ello, la base para la construcción de paneles solares, por ejemplo, para la producción de energía fotovoltaica es el silicio (basado en el procesado del sílice), cuya extracción es altamente costosa. Los molinos para la producción de energía eólica son poco rentables, su instalación poco agradable a la vista y empieza a no estar claro que no tengan un impacto medioambiental. Además, no garantizan la seguridad del suministro ya que su producción está sujeta a fenómenos de la naturaleza que no podemos controlar (la energía eólica depende de la velocidad del viento, la solar, térmica y fotovoltaica, de las horas de sol disponibles anualmente, la minihidráulica del agua embalsada, etc.).

Otra solución es la energía nuclear que se presenta como una solución

⁴United States Dollar

viable ya que puede producir grandes cantidades de energía. Esta puede ser generada mediante procesos de fisión nuclear, en las que se dividen núcleos de alto número másico, o de fusión nuclear, donde se juntan núcleos atómicos ligeros. La energía producida mediante fisión, descubierta en 1939 por los investigadores alemanes Hahn y Strassmann, se genera como una reacción en cadena. Este tipo de reacciones, hoy en día, pueden ser controladas con seguridad. Siendo su principales ventajas

1. Garantiza el abastecimiento eléctrico.
2. Frena las emisiones contaminantes de CO_2 .
3. Reduce la dependencia energética exterior.
4. Produce electricidad de forma constante con precios estables y predecibles.

Es por ello por lo que es considerada una parte muy importante en la generación de energía eléctrica en el mundo, actualmente existen 437 reactores nucleares en operación comercial, que generan aproximadamente el 35 % de la energía eléctrica consumida. Las centrales nucleares están instaladas principalmente en los países desarrollados, y dentro de la Unión Europea se encuentran en operación 148 reactores nucleares que proporcionan, aproximadamente, el 34 % de la electricidad consumida por los Estados Miembros. Sin embargo, la extracción, enriquecimiento y utilización de materiales de alto peso atómico, como puede ser el uranio, para los procesos de fisión nuclear generan residuos radioactivos que tardan miles de años en perder su toxicidad para el medioambiente. La gestión y almacenamiento de esta basura radiactiva plantea además importantes retos técnicos y económicos. Todo ello genera un fuerte rechazo a este tipo de energías por parte de la población que, tal y como se ha comentado con anterioridad, ponen en entredicho el desarrollo futuro de este tipo de energía.

Otra forma de conseguir energía nuclear sería mediante la fusión termónuclear, aunque para esta solución todavía existen barreras que deben soslayarse, tales como la complejidad tecnológica de los dispositivos de fusión. El reto de la fusión consiste en reproducir en la Tierra, las reacciones que se producen en las estrellas, mediante la construcción de reactores de fusión capaces de satisfacer una gran parte de la demanda energética del planeta en el medio plazo. El combustible requerido, está disponible en abundancia:

el deuterio está presente en el agua de los océanos y los mares ($30g/m^3$), y el tritio puede ser producido a partir del litio, en el mismo reactor, en lo que comúnmente se denomina la *cría del manto*. El sistema es intrínsecamente seguro debido al hecho de que un reactor sólo contiene, en un momento dado, una cantidad de combustible suficiente para unos pocos segundos de operación, además, es respetuoso con el medio ambiente, ya que no se producen escapes contaminantes. Al seleccionar elementos con un bajo grado de actividad no será necesario almacenar durante cientos de años residuos altamente radiactivos.

El objetivo de la investigación internacional en el campo de la fusión es diseñar un prototipo de planta de energía de fusión que esté de acuerdo con las demandas de la sociedad actual: a saber, que sea seguro, fiable, sostenible y amigable con el medio ambiente y económicamente viable. En los últimos diez años, se han realizado importantes progresos científicos y tecnológicos que nos intentan acercar a este objetivo.

1.3. Reacciones nucleares

Los núcleos atómicos consisten en protones y neutrones que se empaquetan en una forma extremadamente compacta en una pequeña región del espacio. Ambos tienen una masa casi idéntica, que es 1.836 veces más grande que la masa del electrón. El tamaño (radio) de un núcleo viene dado aproximadamente por la expresión:

$$R = r\sqrt[3]{a}$$

donde $r=1,5 \times 10^{-13}$ cm y a es el número de nucleones en el núcleo.

Z es el número de protones en el núcleo, la variación de $A-Z$ para cualquier Z dado, da lugar al fenómeno de los isótopos. Dos isótopos del hidrógeno: el deuterio y el tritio, son muy relevantes para la fusión, como se verá más adelante. A menudo, los isótopos de un elemento natural son radiactivos y se transforman de manera espontánea en otros elementos. Por ejemplo, el tritio tiene una emisión beta débil (con un promedio de energía de 5,7 keV y un máximo de energía de 18,6 keV) y con una vida media de 12,35 años. Sin embargo, este fenómeno no es de gran importancia para la fusión ya que

los tiempos característicos en los que el proceso de fusión se produce son lo suficientemente cortos como para que este decaimiento ocurra muy a menudo.

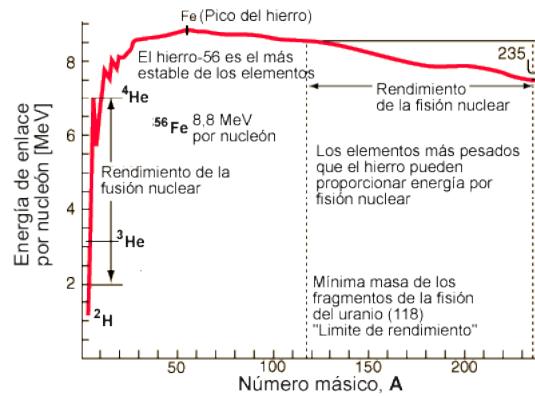


Figura 1.1: Gráfico de las energías de enlace nuclear según la masa atómica. Se puede observar como el rendimiento para la fusión es mucho mayor que para la fisión.

Dentro de los núcleos las fuerzas de repulsión entre los protones cargados positivamente, son mucho menores que las fuerzas nucleares atractivas que mantienen unido el núcleo. Es necesario invertir una cantidad de energía con el fin de extraer un nucleón del núcleo, y cuando un nucleón es capturado por un núcleo, esta misma energía se libera. En la figura 1.1 podemos ver que la energía requerida depende del número másico A , pero tiene un valor máximo de alrededor de 8 MeV para los núcleos próximos al hierro. En la fisión, elementos con fuertes energías de enlace menores de 8 MeV se rompen en núcleos más ligeros, por lo tanto la liberación de energía desde los núcleos resultantes están más cerca del hierro. Por otro lado, en la fusión, dos núcleos ligeros se combinan produciendo uno más pesado, que también está más cerca del hierro. Este último proceso es el que hace que las estrellas tengan un núcleo de hierro hacia el final de sus vidas. Tanto en la fisión como en la fusión, se libera una considerable cantidad de energía, equivalente a la disminución de la masa total de los núcleos del producto con respecto a la masa inicial. La diferencia de masa se libera como energía, de acuerdo con la famosa ecuación de Einstein:

$$e = mc^2$$

Para formar un núcleo de más de un elemento, es necesario que estén lo suficientemente juntos para que las fuerzas atractivas nucleares funcionen. Con este fin, tendrán que superar la fuerza repulsiva de Coulomb, que actúa como una barrera, antes de que las fuerzas nucleares atractivas pueden entrar en juego. La fuerza de repulsión de Coulomb es proporcional a la carga del núcleo, y por tanto, los plasmas formados por los núcleos de la familia de hidrógeno son los más adecuados para conseguir la fusión. Existen diversas reacciones de fusión que pueden ser útiles desde el punto de vista energético. De todas las reacciones que implican isótopos de hidrógeno, la reacción deuterio-tritio tiene la mayor sección eficaz para temperaturas relativamente bajas y, por tanto, es más fácil de conseguir de una manera controlada. La reacción produce un neutrón de alta energía y helio uno, tal y como se muestra en la figura 1.2

Actualmente, se investiga principalmente en dos vías con el objetivo puesto en conseguir reactores nucleares de fusión:

1. **Confinamiento inercial.** El confinamiento inercial produce reacciones nucleares de fusión tras el calentamiento y compresión de bolas de hidrógeno (alcanzándose densidades de $\simeq 10^{32}m^{-3}$ y temperaturas de decenas de keV) durante periodos de tiempo muy cortos (nanosegundos). Para comprimir las bolas a altas densidades y calentarlas se emplean láseres de alta potencia, por lo que para el progreso del confinamiento inercial se requiere un fuerte desarrollo de los mismos.
2. **Confinamiento magnético.** En los dispositivos de confinamiento magnético se emplean campos magnéticos para mantener aislado el plasma de la cámara de vacío. El plasma, normalmente formado a partir de átomos de hidrógeno y sus isótopos (con densidades de $10^{20}m^{-3}$), se calienta a temperatura muy elevada (decenas de keV) y con un tiempo de confinamiento de la energía alto (del orden de segundos). El confinamiento magnético se consigue porque el plasma, aunque macroscópicamente neutro, está compuesto por partículas cargadas (iones y electrones), que se desplazan helicoidalmente a lo largo de las líneas de campo.

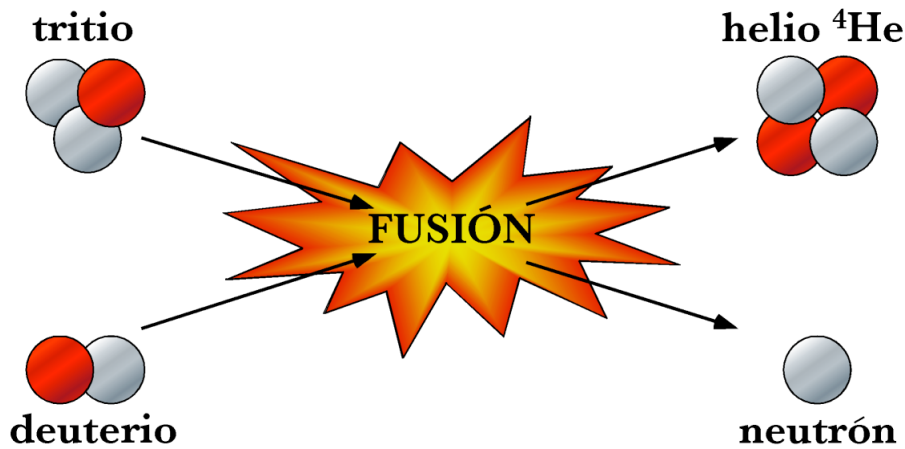


Figura 1.2: Representación esquemática de la reacción de fusión D-T. Un núcleo de deuterio y otro de tritio se unen para dar lugar a un núcleo de helio y un neutrón, liberándose una energía de 17.58 MeV.

1.4. Plasmas de fusión

Con el fin de superar la repulsión electrostática entre los núcleos que se quiere fusionar en condiciones de laboratorio, es necesario que invertir mucha energía, esto se puede lograr mediante su calentamiento a temperaturas muy altas. En tales condiciones, la materia se encuentra en un estado gaseoso altamente ionizado y exhibe un comportamiento colectivo que llamamos plasma. El 99% del Universo está formado por esos plasmas, o materia ionizada, pero en la naturaleza existe una gran variedad de plasmas, que pueden no poseer las condiciones de altas temperaturas necesarias para conseguir la fusión termonuclear.

Un gas ordinario, como el helio, es un conjunto de átomos eléctricamente neutros, en los que cada átomo contiene un número de electrones con carga negativa que orbitan un pequeño núcleo central compuesto por un número igual de protones cargados positivamente y varios neutrones sin carga eléctrica. Los electrones que orbitan están encerrados junto con el núcleo por las fuerzas atómicas, formando un solo átomo eléctricamente neutro. Debido a sus átomos neutros, un gas ordinario muestra poca respuesta a las fuerzas eléctricas y magnéticas, y es un pobre conductor de electricidad.

El gas se convierte en plasma cuando se calienta lo suficiente como para causar que los electrones queden libres del átomo. Las características de un

plasma son muy diferente de las de un gas ordinario. A diferencia de los gases ordinarios, los plasmas son buenos conductores que no solo son afectados significativamente por las fuerzas electromagnéticas, sino que también pueden llegar a generarlas. Esta autogeneración de fuerzas, como resultado de los movimientos colectivos de partículas cargadas y las corrientes eléctricas en el plasma, producen un comportamiento muy complejo que aún no se ha llegado a comprender en su totalidad. Comprender y describir los complejos sistemas dinámicos de los plasmas es un reto científico fundamental.

1.5. Confinamiento Magnético

La idea básica es la de confinar las partículas en un espacio pequeño por medio de campos magnéticos. Esto es factible debido al hecho de que las partículas del plasma que se quieren confinar están cargadas eléctricamente, de modo que un campo magnético puede ejercer una fuerza sobre ellas. Los dispositivos toroidales son los que han logrado los mejores rendimientos en confinamiento. Dentro de estos dispositivos destacan dos clases de dispositivos: el TOKAMAK y el STELLARATOR.

El tokamak (ver figura 1.3), propuesto por Sajarov y Tamm en 1951, es el dispositivo con el que se han conseguido los mejores resultados en el campo de la investigación en fusión nuclear controlada. Es simétrico con respecto al eje central, con bobinas externas alrededor del toro que generan un campo magnético toroidal muy alto (varios teslas).

El stellarator (ver figura 1.4) es un dispositivo toroidal de confinamiento magnético que fue propuesto por el astrofísico Lyman Spitzer en 1951. Los dispositivos de la clase stellarator no tienen simetría toroidal, y los campos magnéticos necesarios para confinar el plasma son generados exclusivamente por bobinas externas, sin requerir una corriente eléctrica toroidal en el plasma. Esta característica hace que su confinamiento sea inherentemente estacionario, siempre que se mantengan las corrientes en las bobinas externas. Salvando estas diferencias, la física de los plasmas producidos en stellarators y tokamaks es similar. Debido a que el diseño y la construcción de los stellarators es más complicado, su desarrollo ha ido siempre una generación por detrás de los tokamaks, no alcanzando el rendimiento de estos últimos.

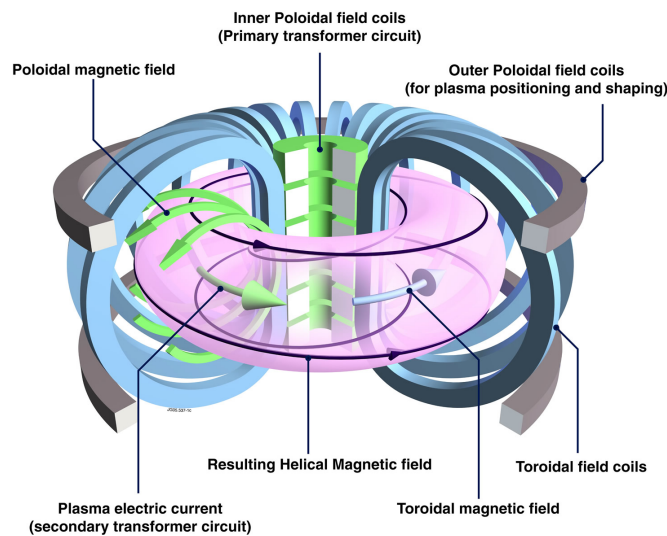


Figura 1.3: Representación esquemática de un tokamak. El plasma se representa en color rosado.

1.6. TJ-II y JET

En Madrid se encuentra en funcionamiento el “Stellarator” más grande de Europa, el TJ-II. Es de tipo heliac flexible de tamaño medio. Sus objetivos principales son el estudio del confinamiento y del transporte de energía y de partículas.

El Joint European Torus (JET), situado en el centro para fusión nuclear de Culham, es el mayor y más potente tokamak del mundo además de ser el principal objetivo del programa Europeo de investigación en fusión nuclear. Este dispositivo está diseñado para estudiar la fusión en las condiciones que serían necesarias en una planta de generación, siendo el único dispositivo que opera en la actualidad utilizando una mezcla de deuterio-tritio como combustible el cual será utilizado en los futuros reactores comerciales de fusión.

Desde que JET comenzó a operar en 1983, se han conseguido importantes avances en la ciencia e ingeniería de fusión, incrementando la confianza y la viabilidad del tokamak como modelo de construcción de futuras plantas de producción de energía.

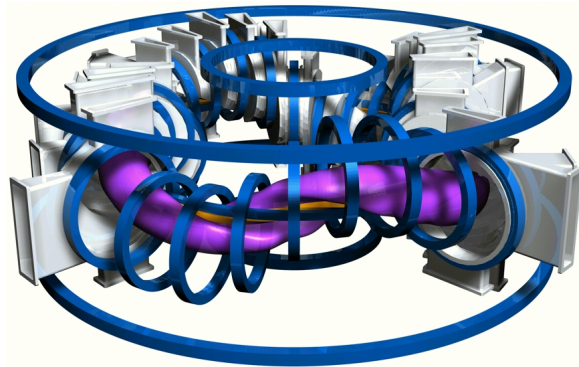


Figura 1.4: Stellarator TJ-II. El plasma se representa en color fucsia. Se pueden apreciar las ventanas de observación para diagnosis del plasma.

1.6.1. Diagnosis

El principal objetivo de la diagnosis del plasma es deducir información de su estado a partir de observaciones experimentales de los procesos físicos y efectos asociados que en él se producen. Para conseguir un rendimiento científico óptimo, se dispone de un conjunto de diversos diagnósticos con el que se pueden determinar las características fundamentales del plasma producido. La figura 1.5 muestra los principales elementos de diagnosis del TJ-II[2]. Del mismo modo podemos ver los diagnósticos de JET en la figura 1.6.

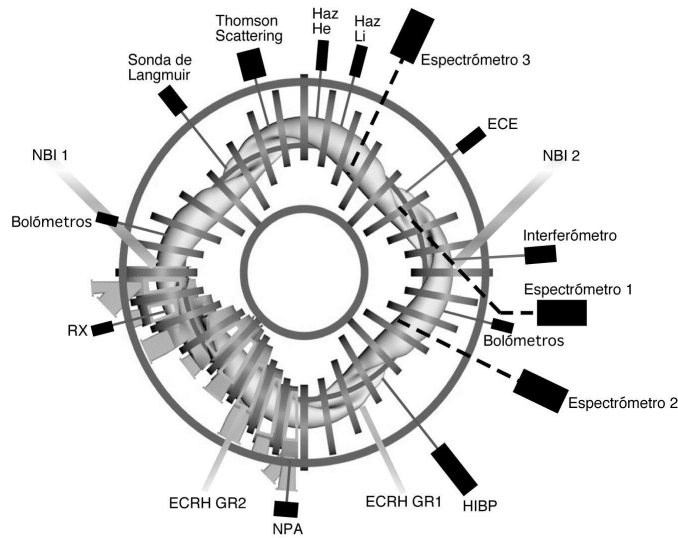


Figura 1.5: Localización y representación esquemática de los principales diagnósticos del stellarator TJ-II. También se muestran las dos líneas de calentamiento de ECRH (GR1 y GR2), así como las líneas de inyección de partículas neutras para calentamiento (NBI).

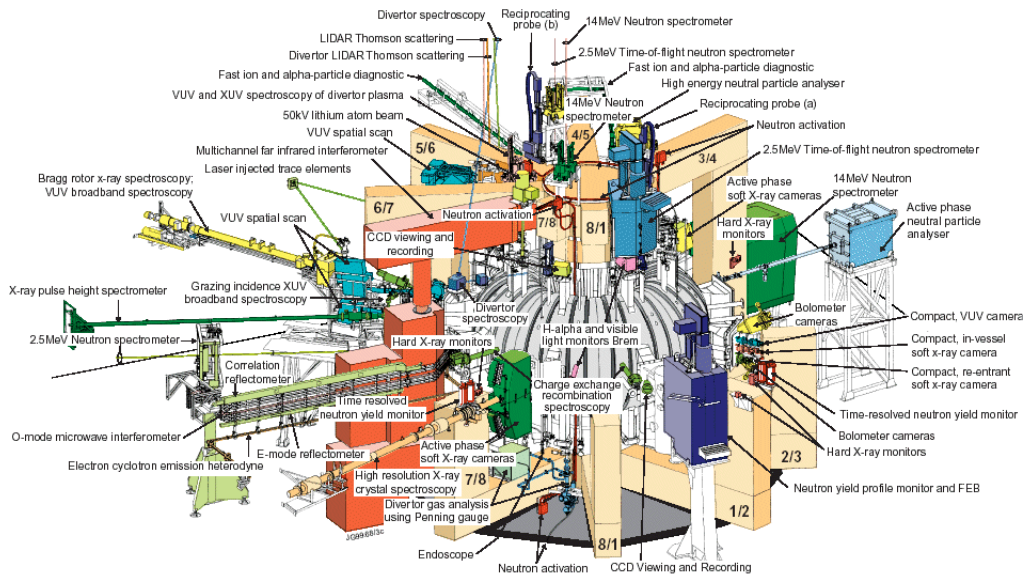


Figura 1.6: Localización y representación esquemática de los principales diagnósticos del Tokamak JET.

Capítulo 2

Objetivo y alcance del proyecto.

2.1. Minería de datos aplicada a la ciencia e ingeniería

Probablemente uno de los rasgos más característicos de la ciencia e ingeniería actual es la explosión de información que está disponible para su almacenamiento y posterior análisis. Experimentos en física, como el Gran Colisionador de Hadrones, que busca profundizar en el conocimiento de la materia y sus interacciones, genera 15 PetaBytes de datos cada año. En neurología, *The human brain project* cuyo objetivo es obtener un conocimiento profundo del cerebro humano, serán necesarios 50 PetaBytes de almacenamiento para simular el cerebro de un ratón, y se llegará a 200 PetaBytes cuando se pretenda simular el cerebro humano. En astronomía, el proyecto Gaia está enfocado a realizar el censo astronómico más exhaustivo realizado jamás de la vía láctea, donde se espera recabar más de 200 TeraBytes de información. En biología, se estima que para 2014 haya disponible 1 ExaByte de datos genómicos.

Esta explosión de información se ha hecho posible gracias al aumento de la complejidad de los sistemas de instrumentación y a la mayor capacidad de almacenamiento y procesamiento de los actuales sistemas de computación, que están ayudando a generar un mayor conocimiento científico y técnico.

No obstante, analizar toda esta información conlleva una ardua tarea que no puede ser aproximada con las metodologías tradicionales de análisis de datos científicos. Nuevos métodos automáticos de análisis de datos son esenciales para transformar grandes bases de datos en conocimiento científico.

2.1.1. Minería de datos aplicada al estudio de la fusión nuclear

Un ejemplo de lo anteriormente descrito son los plasmas generados en dispositivos de confinamiento magnético. Como ejemplo tenemos JET que puede generar descargas de hasta 60 segundos, alcanzándose hasta 0,5 Terabytes de datos por descarga. Un amplio conjunto de señales como son formas de onda, imágenes y líneas de nivel, son adquiridas en tiempo real a través de los diagnósticos descritos en el capítulo anterior. Esta necesidad de almacenamiento y análisis se ve incrementada ante la duración prevista de las descargas (30 minutos) en futuras máquinas tales como ITER o Wendelstein 7X. Por lo que tales descargas producirán un volumen de datos significativamente mayor al actual. [3]

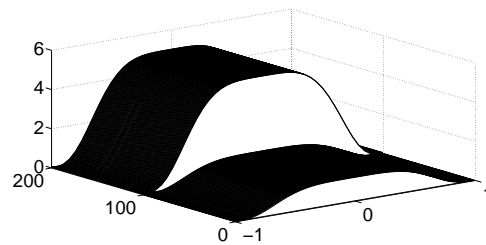
2.2. Objetivo del trabajo de fin de máster

2.2.1. Introducción

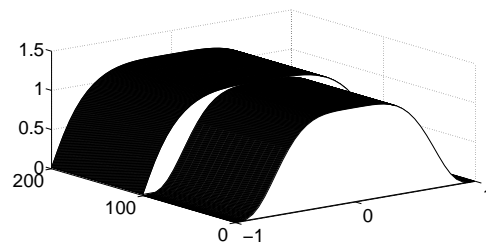
En la operación de plasmas de pulso largo, como por ejemplo ITER, el reconocimiento automático puede ser una tarea de gran importancia. El término *evento* se utiliza para identificar cualquier clase de cambio en el flujo de datos, donde la palabra *cambio* se refiere a una diferencia cualitativa en la secuencia de evolución temporal. El reconocimiento automático de cambios es útil en el análisis de datos en tiempo real (por ejemplo activando y desactivando la adquisición de señales o incrementando/decrementando los ratios de muestreo) y off-line (centrando la atención en los intervalos de tiempo relevantes).

2.2.2. Descripción del objetivo

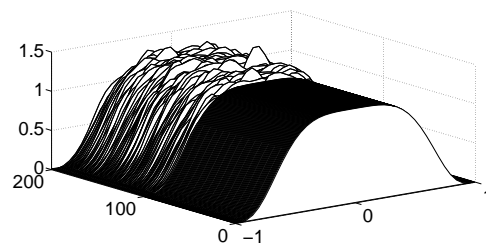
Este trabajo describe la aplicación a entornos de fusión nuclear de una técnica de propósito general para la detección de cambios en flujos de datos variables en el tiempo, sin la necesidad de generar desarrollos particulares para cada problema específico. Dicha técnica está basada en una metodología de aprendizaje automático [4]. Con este método, en cada instante de tiempo los puntos de datos tienen una representación multidimensional. Por ejemplo, cada punto puede ser un vector cuyos componentes son la amplitud de diferentes formas de onda o el píxel de intensidad de una imagen.



(a)



(b)



(c)

Figura 2.1: Simulación de cambios en flujos de datos. Los cambios mostrados son: (a) cambio a una amplitud mayor; (b) mayor ancho; (c) más ruido

Los cambios en flujos de datos son transiciones desde un estado inicial a uno final, ambos caracterizados por su evolución particular.

En la figura 2.1 podemos ver simulaciones de cambios en el perfil radial a lo largo del tiempo. Estos perfiles pueden corresponder, por ejemplo, a cambios en diferentes magnitudes físicas del plasma, como la temperatura, densidad o radiación. Por otro lado, los cambios simulados pueden ser debidos a la inyección adicional de calor, el incremento de impurezas o la presencia de una mayor actividad MHD en el plasma.

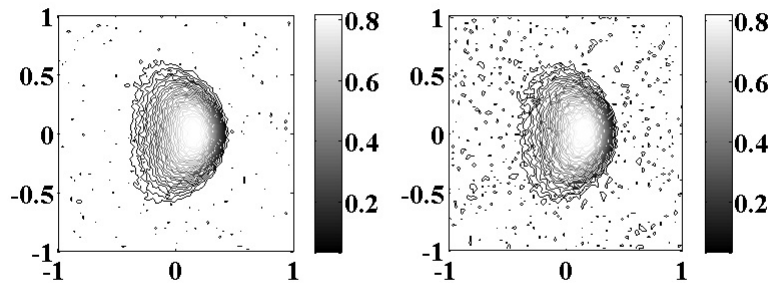


Figura 2.2: Simulación de líneas de contorno de emisión del plasma con diferentes niveles de fluctuación.

La figura 3.1 muestra el cambio del flujo de datos en un problema de dimensión mayor. El objetivo es la detección de variaciones en la distribución espacial de dos dimensiones de la intensidad de emisión del volumen del plasma. El termino emisión lo utilizamos en un sentido amplio y de hecho puede representar diferentes emisiones del plasma tales como emisiones de rayos X blandos, de bolometría o de neutrones. Las emisiones del plasma pueden evolucionar con un nivel de fluctuación particular, por ejemplo el aumento de la actividad de MHD ¹ puede provocar el efecto de elevar el nivel de fluctuación.

La figura 2.3 muestra la evolución de los perfiles experimentales del stellarator TJ-II. En el estado inicial, la amplitud se incrementa de forma monóto-

¹Magnetohidrodinámica. Este tipo de actividad permite estimar el nivel de turbulencia del plasma de fusión.

na y, a un cierto instante de tiempo, las cuerdas centrales muestran un decremento de la amplitud, manteniendo un valor constante.

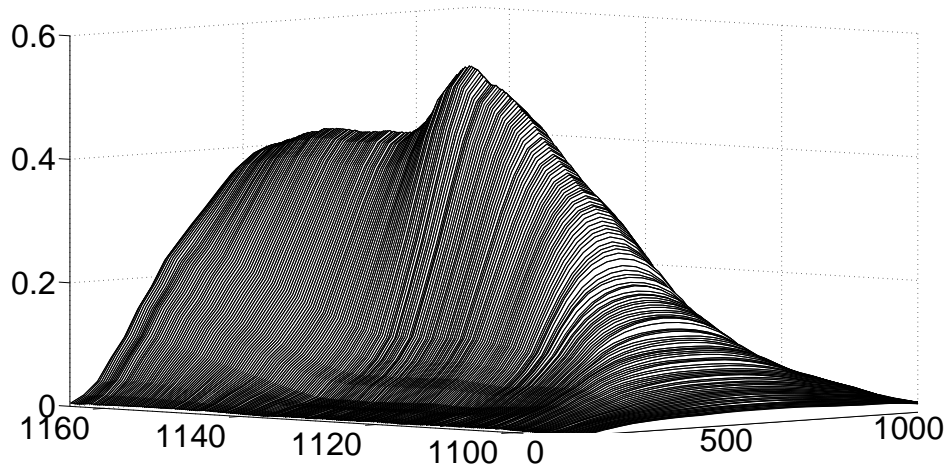


Figura 2.3: Evolución temporal de perfiles de bolometría. Perfiles de bolometría del stellarator TJ-II (disparo 21319).

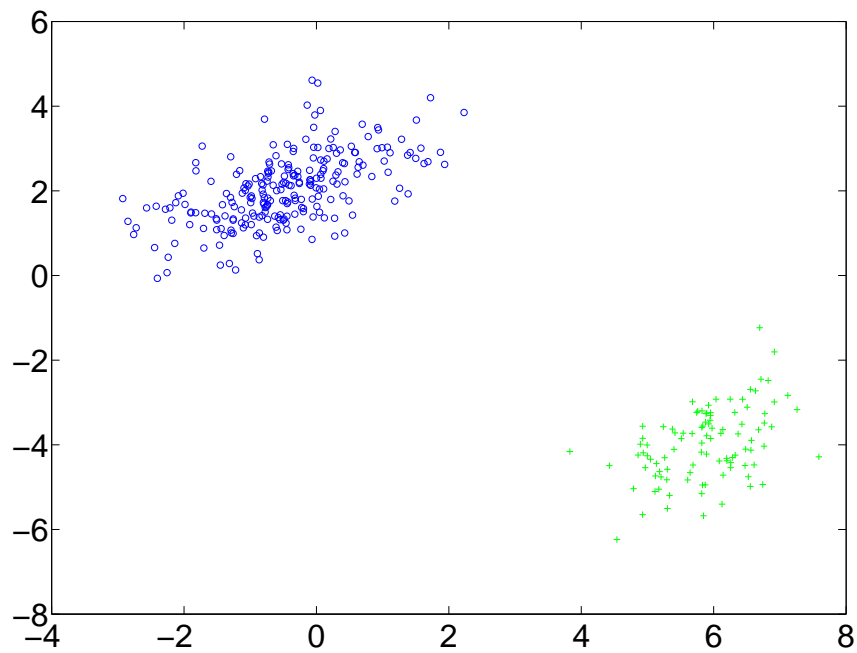
Solución Conceptual

La identificación de un algoritmo para detectar cambios en flujo de datos debe cumplir los siguientes requerimientos:

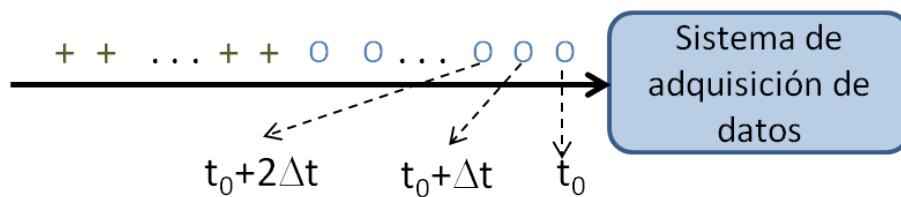
1. El tiempo de retardo entre un punto de cambio y su detección debe ser mínimo.
2. El número de cambios verdaderos sin detectar debe ser mínimo.
3. El número de falsas detecciones o falsas alarmas (FA) de cambios debe ser mínimo.
4. Debe tratar eficientemente flujos de datos.
5. La secuencia de datos debe ser leída una sola vez.

La detección del cambio es más efectiva si la descripción de los puntos forman matemáticamente clusters separados en el espacio de características.

La figura 2.4 representa datos que se han leído uno a uno, primero todos los de un determinado cluster y luego los del otro. En el momento del cambio (lectura del primer elemento del segundo cluster), el proceso de detección tiene que reconocer que los datos leídos de ahí en adelante son *lejanos* a los datos leídos con anterioridad, momento en el que debe dispararse una alarma de advertencia de dicho cambio.



(a)



(b)

Figura 2.4: (a) Idealmente, el vector de características debería ser seleccionado de forma que los dos estados sean separables en el espacio de características. (b) Los datos son leídos de forma secuencial, uno a uno.

Capítulo 3

Fundamentos teóricos y relación con otros trabajos.

3.1. Clasificación y definición de los diferentes tipos de cambio

Tal y como se define en [5] y [6] podemos distinguir los siguientes tipos de cambios:

1. término del cambio (*concept change*).
 - a) término de deriva (*concept drift*).
 - b) término de desplazamiento (*concept shift*).
2. cambio de distribución o muestreo.

Como *término* nos referimos a la variable objetivo que el modelo intenta predecir. El término de deriva describe un cambio gradual del término y el de desplazamiento ocurre cuando un cambio entre dos términos es más abrupto.

El cambio de distribución, también conocida como cambio de muestreo, desplazamiento, o desplazamiento virtual, se refiere al cambio en la distribución de los datos. Incluso si el término se mantiene igual, el cambio a

menudo suele llevarnos a revisar el actual modelo debido a que el error de dicho modelo puede no ser aceptable con la nueva distribución de datos.

Diversos autores como Stanley [7], han sugerido que desde un punto de vista práctico, no es esencial distinguir entre término del cambio y distribución del cambio.

La detección de cambios, como hemos visto en el apartado 2.2.2, no es una tarea fácil, ya que los modelos que se propongan deben desarrollar un compromiso entre la detección de verdaderos cambios y la detección de falsas alarmas. Cuantos más cambios verdaderos queramos detectar mayor será el ratio de falsas alarmas (FAR) y viceversa.

Siguiendo a [4] en este trabajo definiremos el cambio como el proceso de transición desde un estado o forma a otro. En un flujo de datos, el cambio se produce cuando el proceso de generación de datos se desplaza desde un modelo de generación de dichos datos a otro. Este desplazamiento suele producirse por un cambio en los parámetros del modelo. Por ejemplo, si los datos observados son generados siguiendo una distribución Gaussiana $N(\mu, \sigma^2)$ con media μ y varianza σ^2 desde $N(0, 1)$ a $N(2, 1)$.

3.2. Requerimientos de un algoritmo para el análisis de flujos de datos

Tal y como describe Bifet en [8] cualquier algoritmo de proceso de flujo de datos debe cumplir las siguientes características:

1. **Procesar un ejemplo en cada instante de tiempo e inspeccionarlo una sola vez.** La característica principal de un flujo de datos es que los datos entran de uno en uno. Una vez que el dato ha sido tratado se desecha sin la posibilidad de recuperarlo posteriormente. Esta condición puede ser relajada en los casos en los que el flujo de datos es reenviado para que un segundo algoritmo refine el modelo que ha aprendido. No obstante, para que un algoritmo sea lo suficientemente flexible para que pueda aplicarse de forma universal a cualquier flujo de datos deberá ser de un solo paso.

2. **Utilizar una cantidad de memoria limitada.** Probablemente la principal motivación para emplear un algoritmo de flujo de datos es que permite el proceso de datos en aquellos casos en los que estos son de un tamaño mayor que la memoria disponible.
3. **Necesidad de tiempo de proceso limitado.** Para que un algoritmo sea escalable para cualquier número de ejemplos, su tiempo de ejecución debe ser lineal. Para que un algoritmo pueda trabajar en *tiempo real* debe tratar los ejemplos tan rápido, o más, que la velocidad de los datos que le llegan. Lo contrario significaría la pérdida de datos. Esta característica puede relajarse en el caso de que realicemos procesamientos *off-line* o simplemente si el usuario de dicho algoritmo solo quiera obtener resultados en un tiempo razonable.
4. **Estar preparado para realizar predicciones en cualquier momento.** Un algoritmo ideal debe ser capaz de producir el mejor modelo a partir de los datos observados después de ver un número determinado de ejemplos.

Para el caso particular de un algoritmo secuencial y/o en tiempo real efectivo para detectar cambios requiere adicionalmente que:

- El tiempo medio entre que se produzca un cambio y su detección sea mínimo.
- El número de cambios reales no detectados y falsas alarmas (FA) sea mínimo.

3.3. Escenarios estándares de aprendizaje automático

La detección de cambios se puede considerar en los tres escenarios estándares del aprendizaje automático (*machine learning*):

1. **Cambio en los modelos de clasificación.** En este caso se asume que el modelo de generación del flujo de datos lleva inherente en su patrón

de distribución solo unas pocas clases. Como mínimo tendrá dos clases que determinarán el cambio.

2. **Cambio en el modelo de *cluster*.** En este escenario se suele asumir que el número de los *clusters* es desconocido o los datos sin etiquetar no están agrupados en un solo *clusters*.
3. **Cambios en modelos de regresión.** En este escenario, se asume que los datos están descritos por una función: $f : \mathbb{R}^n \rightarrow \mathbb{R}$.

3.4. Predicciones inductivas y transductivas

Vapnik [9] formuló una distinción entre la inducción y la transducción aplicado al problema de predicción. En la *predicción inductiva*, se obtienen reglas o modelos más o menos generales a partir de un conjunto de entrenamiento, que llamaremos ejemplos y que son representativos de la distribución de los datos a predecir. En la *predicción transductiva*, se toma un acceso directo, por lo que se pasa de los ejemplos anteriores a directamente la predicción de un nuevo ejemplo.

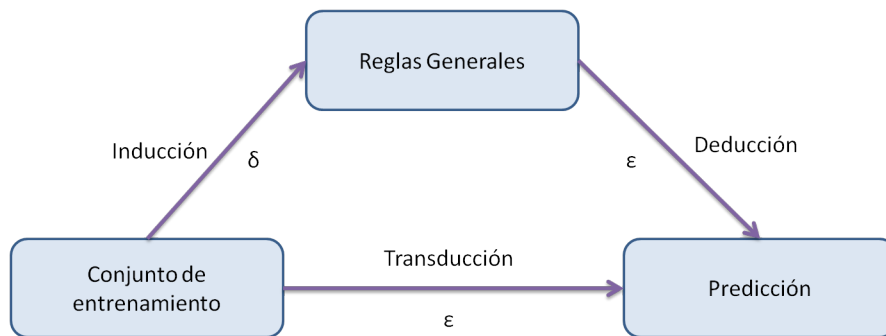


Figura 3.1: Predicción inductiva y transductiva.

En el caso de predicciones sencillas, la distinción entre la inducción y la transducción resulta ser menos nítida.

Un método para realizar la transducción On-Line, es un método para predecir y_n de $x_1, y_1, \dots, x_{n-1}, y_{n-1}, x_n$. Este método proporciona una predicción para cualquier ejemplo que pueda ser representado como x_n , y así lo

define, al menos implícitamente, por regla general, podría ser extraído de $x_1, y_1, \dots, x_{n-1}, y_{n-1}$

3.5. Predictores conformales

En este apartado describiremos el concepto de predictor conformal que utilizaremos más adelante en nuestra descripción del modelo de detección de cambios. Lo descrito en este apartado se basa en la descripción que se hace de los mismos en [10].

3.5.1. Introducción

Los predictores conformales son un método de predicción que proporciona una evaluación de su precisión y fiabilidad y que pueden ser utilizados con cualquier método de predicción para clasificación, clustering y regresión incluyendo, entre otros, máquinas de vector soporte (SVM), árboles de decisión, redes neuronales y modelos de regresión.

Para estimar la confianza en las predicciones con métodos de aprendizaje automático, uno de los métodos más típicos es el viejo e ingenuo método del estimador *Holdout*. Con el propósito de calcular este estimador, los ejemplos disponibles son divididos en dos grupos: un conjunto de entrenamiento y un conjunto de test. El conjunto de entrenamiento es utilizado para encontrar una regla de predicción (proceso de aprendizaje) y luego esta regla de predicción es aplicada al conjunto de test. El ratio de observaciones correctas en el conjunto de test es utilizado como una medida de confianza de las reglas de predicción cuando es aplicada a los nuevos ejemplos.

Los predictores conformales tienen varias ventajas sobre los estimadores *Holdout*:

1. Los predictores conformales no hacen una separación rígida entre aprendizaje y estimación. Todos los objetos son tratados simultáneamente y, por tanto, aprendizaje y predicción se combinan.

2. Como consecuencia de esta mezcla, los predictores conformales aprenden y predicen al mismo tiempo, mejorando continuamente su rendimiento tal y como va realizando cada nueva predicción y buscando cómo ajustarse. De esta forma, no es necesario realizar reentrenamientos.
3. Las predicciones son más fiables y precisas porque una nueva medida de confianza es obtenida en cada clasificación, sin confiarla a valores obtenidos con un conjunto fijo de ejemplos (conjunto de entrenamiento).
4. La confianza en la etiqueta predicha en cada nueva clasificación se toma de los objetos previos y todas las etiquetas posibles.
5. Un predictor conformal puede detectar ambigüedades en sus tareas de clasificación, por ejemplo, cuando una única clase no puede ser asignada a un nuevo ejemplo. En algunas circunstancias, detectar la ambigüedad puede ser más importante que producir una mala estimación.

3.5.2. Medida de no conformidad

Para determinar el nivel preciso de confianza en las predicciones, los predictores conformales utilizan la experiencia obtenida de los ejemplos tratados con anterioridad. Para cada nuevo ejemplo a clasificar, es necesario medir cuán diferentes es el ejemplo actual de los anteriores. Para ello es necesario definir una medida de no conformidad y se debe determinar una puntuación de no conformidad para estimar cómo de diferente es un nuevo ejemplo respecto a una *bolsa* con los antiguos ejemplos [11].

Definimos una *bolsa* de tamaño $n \in \mathbb{N}$ como una colección de n elementos alguno de los cuales pueden estar repetidos. Es como un conjunto en la medida en la que sus elementos no están ordenados y como una lista en la medida en la que un elemento puede ocurrir más de una vez. Se escribirá $\{a_1, \dots, a_N\}$ para la lista a_1, \dots, a_N eliminando la información sobre el orden.

Dada una medida de no conformidad A y una bolsa $\{z_1, \dots, z_n\}$ es posible calcular el valor de no conformidad

$$\alpha_1 = A(\{z_i, \dots, z_{i-1}, z_{i+1}, \dots, z_n\}, z_i) \quad (3.1)$$

para cada ejemplo z_i en la bolsa.

Por regla general, cada ejemplo es asignado a una medida de no conformidad α_i tal y como se muestra en la ecuación 3.1 mediante la aplicación de algún método predictivo. En general existen varios tipos de medida de no conformidad, también llamadas *extrañezas*, relacionadas con métodos predictivos de clasificación, clustering y regresión [4]. Detallaremos algunos ejemplos de estos métodos en el siguiente apartado.

3.5.3. Medidas de *extrañeza*

k-NN: k vecinos más cercanos

Dada una secuencia de distancias entre los miembros de un conjunto de entrenamiento dado, que cuantifica en qué medida es probable la etiqueta de un ejemplo. En este caso se define la medida de no conformidad del ejemplo i con la etiqueta y en relación con el resto de los ejemplos de entrenamiento como:

$$s_i = \frac{\sum_{j=1}^k d_{ij}^y}{\sum_{j=1}^k d_{ij}^{-y}} \quad (3.2)$$

donde d_{ij}^y es la j -enésima distancia más corta en la secuencia de distancias ordenada ascendentemente del ejemplo i a los otros ejemplos con la misma etiqueta de clase y , y d_{ij}^{-y} es la j -enésima distancia más corta de la secuencia ordenada de distancias del ejemplo i desde los ejemplo con etiqueta de clase diferente de y . La medida de no conformidad en este caso será el resultante del ratio de la suma de las k distancias más cercanas a la misma clase, a la suma de las k distancias más cercanas a las otras clases. La no conformidad de un ejemplo con etiqueta y será mayor cuando la distancia desde los ejemplos de la misma clase sea más amplia y cuando la distancia a las otras clases es menor.

SVM: Máquinas de vector soporte

Dado el conjunto de entrenamiento $T = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_n, y_n)\}$ donde $y_i \in \{-1, 1\}$, el SVM busca el hiperplano que produce el margen máximo.

Cuando un nuevo ejemplo \mathbf{x}_{n+1} es incluido con una etiqueta y_{n+1} en el conjunto de entrenamiento, que emplea los multiplicadores de Lagrange $\lambda_1, \lambda_2, \dots, \lambda_n, \lambda_{n+1}$ asociados con los ejemplos en el conjunto de entrenamiento y $(\mathbf{x}_{n+1}, y_{n+1})$ como la medida de no conformidad cuando utilizamos el SVM.

La relación entre medida de no conformidad y multiplicadores de Lagrange puede explicarse de la siguiente forma: Los ejemplos fuera del margen y en el lado correcto del hiperplano no tiene ningún multiplicador de Lagrange. Para los ejemplos en el margen, los valores de los multiplicadores de Lagrange están entre 0 y C . Todos los ejemplos dentro del margen tienen multiplicadores de Lagrange con valor C . Los ejemplos con mayor medida de no conformidad son aquellos dentro del margen y en el lado erróneo del hiperplano, mientras que los que están correctamente etiquetados fuera del margen son los que tienen una medida de no conformidad menor.

Modelos de cluster

Dado el conjunto de datos sin etiquetar $T = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$, se define la medida de no conformidad de un ejemplo \mathbf{x}_i en relación con un modelo de *cluster* como:

$$s(T, \mathbf{x}_i) = \|\mathbf{x}_i - c\|, \quad (3.3)$$

donde c es el centroide del *cluster* y $\|\cdot\|$ es una distancia métrica.

Modelos de regresión

Dado un conjunto de entrenamiento $T = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_n, y_n)\}$ donde $y_i \in \mathbb{R}$, se define la medida de no conformidad de un ejemplo (\mathbf{x}_i, y_i) con respecto a un modelo de regresión como:

$$s(T, (\mathbf{x}_i, y_i)) = \frac{|y_i - f(\mathbf{x}_i)|}{e^{g(\mathbf{x}_i)}} \quad (3.4)$$

donde f es la función de regresión y g es la estimación de la función de error para f en \mathbf{x}_i . En nuestro caso, la función de estimación g está basada en un modelo de regresión de los valores $\ln(|y_i - f(\mathbf{x}_i)|)$ para \mathbf{x}_i , donde $i = 1, \dots, n$.

3.5.4. Cálculo on-line de p-valores

Utilizando las medidas de no conformidad α_i obtenidas con algunos de los métodos descritos en el apartado 3.5.2 para todos los ejemplos observados, el p-valor p_n que corresponde al ejemplo z_n se obtiene calculando:

$$p_n = \frac{\#\{i : \alpha_i > \alpha_n\} + \theta_n \#\{n : \alpha_n = \alpha_n\}}{n} \quad (3.5)$$

donde θ_i es seleccionada de forma aleatoria dentro del intervalo $[0, 1]$ para cada valor de i . El símbolo $\#$ corresponde a la cardinalidad de un conjunto.

El siguiente algoritmo describe un proceso de cálculo on-line de los p-valores [12].

Un resultado estándar de la teoría de los predictores conformales es el teorema que se define a continuación [13]:

Teorema 3.1. *Sea $N \in \mathbb{N}$ y los ejemplos $z_n \in \mathbf{Z}, n = 1, \dots, N$, son generados con una distribución de probabilidad P sobre \mathbf{Z}^N de acuerdo con una compresión N -modelo on-line. Cualquier transductor conformal suavizado en dicho modelo es exactamente válido (producirá p-valores independientes $p_n, n = 1, \dots, N$, distribuidos uniformemente en $[0, 1]$).*

Algoritmo 1 Generación de p-valores on-line

Entrada: (z_1, z_2, \dots) datos a comprobar.**Salida:** (p_1, p_2, \dots) secuencia de p-valores.

- 1: **para** $i = 1, 2, \dots$ **hacer**
 - 2: observar un nuevo ejemplo z_i
 - 3: **para** $j = 1$ hasta i **hacer**
 - 4: $\alpha_j = A(\{z_1, \dots, z_i\}, z_j)$
 - 5: **fin para**
 - 6: $p_i = \frac{\#\{j:\alpha_j > \alpha_i\} + \theta_n \#\{j:\alpha_j = \alpha_i\}}{i}$
 - 7: **fin para**
-

La propiedad de que los ejemplos generados por una distribución intercambiabile genera p-valores independientes y uniformemente distribuidos nos permite comprobar la intercambiabilidad mediante el cálculo de martingalas como funciones de los p-valores.

3.6. Test de Intercambiabilidad On-Line

Siguiendo [14], vamos a definir los diferentes elementos necesarios para construir un test de intercambiabilidad On-Line.

3.6.1. Conceptos de aleatoriedad e intercambiabilidad

Sean los elementos z_1, z_2, \dots , que llamaremos *ejemplos*, de una medida del espacio \mathbf{Z} . La hipótesis de *aleatoriedad* es aquella en la que cada $z_n, n = 1, 2, \dots$, es generado de acuerdo con una distribución de probabilidad P en \mathbf{Z} , y los ejemplos son independientes unos de otros (de modo que la secuencia z_1, z_2, \dots es generada por la distribución de potencia P^∞). Una hipótesis muy parecida es la de *intercambiabilidad* que es aquella en la que los ejemplos z_1, z_2, \dots son generados de acuerdo a una probabilidad de distribución *intercambiabile* Q en \mathbf{Z}^∞ , por ejemplo, si Q está formado por la permutación de ejemplos $z_{\pi(1)}, \dots, z_{\pi(n)}$, que están distribuidos como los ejemplos originales z_1, z_2, \dots , para cualquier n y π de $\{1, \dots, n\}$. Es inmediato que *a priori* la hipótesis de intercambiabilidad es tan o más débil que la hipótesis de aleatoriedad, puesto que todas las distribuciones de potencia

son intercambiables.

3.6.2. Martingalas

Para testar la hipótesis de aleatoriedad/intercambiabilidad *on-line*: después de observar cada uno de los ejemplos z_n el método de testeo debe generar un número M_n que refleje la fortaleza de la evidencia hallada frente a la hipótesis. La forma más natural de hacerlo es mediante el uso de una Martingala no negativa que comience con el valor 1. Supongamos primero que queremos testar simplemente la hipótesis de que z_1, z_2, \dots son generados desde una probabilidad de distribución Q en \mathbf{Z}^∞ . Diremos que una secuencia de variables aleatorias M_0, M_1, \dots, ∞ es una *Q-supermartingala* si, para todo $n = 0, 1, \dots, M_n$ es una función medible de z_1, \dots, z_n (en particular, M_0 es una constante) y

$$M_n \geq \mathbb{E}^Q(M_{n+1} \mid M_1, \dots, M_n). \quad (3.6)$$

Si $M_0 = 1$ y $\inf_n M_n \geq 0$, M_n puede ser estimado como el capital de un jugador que comienza en 1, y que nunca llega a la bancarrota, al comienzo de cada mano n apuesta a z_n . Si tal supermartingala M nunca toma un valor grande, nuestra creencia en Q es indeterminada; esta intuición es formalizada por la desigualdad de Doob, la cual implica que

$$Q\{\exists n : M_n \geq C\} \leq 1/C, \forall C \geq 1 \quad (3.7)$$

donde C es una constante positiva constante.

3.6.3. Supermartingalas aleatorias e intercambiables

Cuando probamos una *hipótesis compuesta* \mathcal{P} (por ejemplo una familia de distribuciones de probabilidad), utilizaremos una *P-supermartingalas*. Por ejemplo, secuencias de variables aleatorias M_0, M_1, \dots , las cuales son *Q-supermartingalas* para todas $Q \in \mathcal{P}$. Si P es el conjunto de todas las distribuciones de potencia P^∞ , P oscila sobre la distribución de probabilidad en \mathbf{Z} ,

llamaremos a \mathcal{P} -*supermartingalas supermartingalas aleatorias*. Estaremos más interesados en una amplia familia \mathcal{P} formada por todas las distribuciones de probabilidad intercambiable Q en \mathbf{Z}^∞ ; en este caso usaremos el término *supermartingala intercambiable* para \mathcal{P} – *supermartingalas*.

El teorema de Finetti y el hecho de que los espacios de Borel son cerrados bajo productos contables, implica que cada distribución intercambiable Q en \mathbf{Z}^∞ es una mezcla de distribuciones de potencia P^∞ que proporciona \mathbf{Z} como Borel. Esto muestra que la noción de supermartingalas aleatorias e intercambiables no negativas coincide en el caso Borel. Pero incluso sin la presunción de que \mathbf{Z} sea Borel, todas las supermartingalas aleatorias son supermartingalas intercambiables.

3.6.4. Martingalas intercambiables aleatorias

Las martingalas intercambiables aleatorias son una secuencia de funciones medibles $M_n(z_1, \theta_1, \dots, z_n, \theta_n)$ (A cada ejemplo z_n se le añade un número aleatorio $\theta_n \in [0, 1]$) de modo que, para cualquier distribución de probabilidad intercambiable Q en \mathbf{Z}^∞ ,

$$M_n = \mathbb{E}^{QxU^\infty}(M_{n+1} | M_1, \dots, M_n), \quad (3.8)$$

donde U es una distribución uniforme en $[0, 1]$.

3.6.5. Martingalas basadas en p-valores

Dada una secuencia de p-valores veremos la forma de calcular una función de dichos p-valores utilizando para ello una martingala.

Sea $f_i : [0, 1]_i \rightarrow [0, \infty)$, para cada $i \in \{1, 2, \dots\}$. Sea (p_1, p_2, \dots) la secuencia de p-valores generada por el algoritmo 1. Consideraremos las mar-

tingalas S_n de la forma:

$$S_n = \prod_{i=1}^n f_i(p_i), n = 1, 2, \dots, \quad (3.9)$$

donde designaremos a $f_i(p) = f_i(p_1, \dots, p_{i-1}, p$ y llamaremos a $f_i(p)$ *función de apuesta* [12].

Para asegurarnos de que 3.9 es en realidad una martingala, necesitamos la siguiente restricción en la *función de apuesta* f_i .

$$\int_0^1 f_i(p) dp = 1, \quad n = 1, 2, \dots \quad (3.10)$$

Por tanto podemos ver que:

$$\mathbf{E}(S_{n+1} | S_0, \dots, S_n) = \int_0^1 \prod_{i=1}^n f_i(p_i) f_{n+1}(p) dp =$$

$$\prod_{i=1}^n (f_i(p_i)) \int_0^1 f_{n+1}(p) dp = \prod_{i=1}^n f_i(p_i) = S_n$$

Utilizando la ecuación 3.9 podemos actualizar la martingala de forma on-line. Para ello calculamos el p-valor p_i para un nuevo ejemplo mediante el algoritmo 1, llegando a ser el valor actualizado de la martingala $S_i = S_{i-1} \cdot f_i(p_i)$. Con el fin de definir las martingalas completamente es necesario definir la *función de apuesta* f_i . A continuación definiremos varios tipos de martingalas a partir de la definición de la *función de apuesta* f_i .

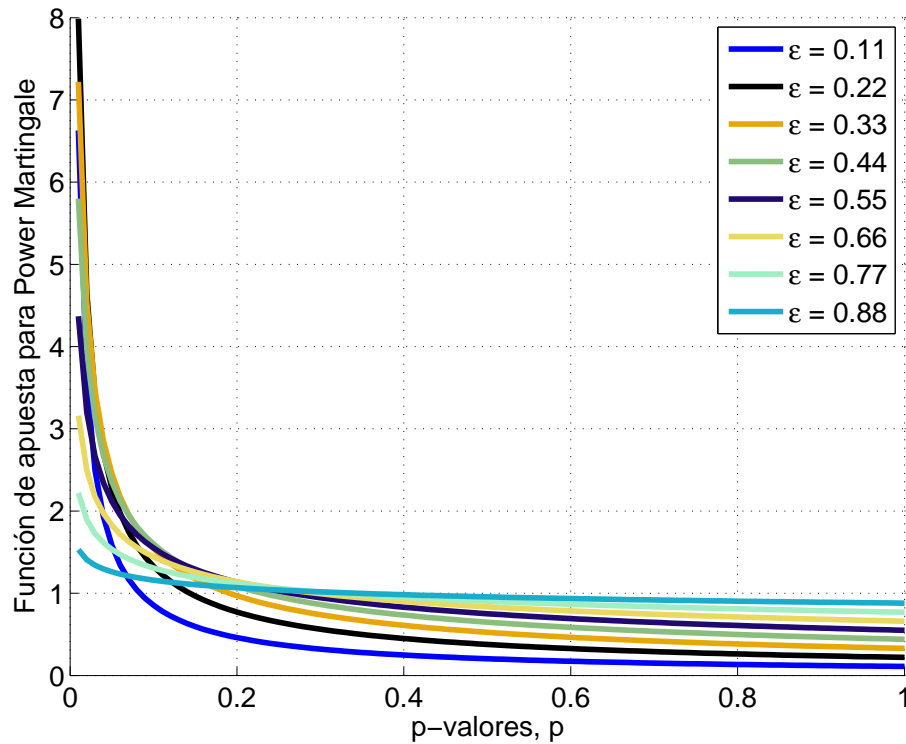


Figura 3.2: Valores de función de apuesta que se utilizan para calcular Randomized Power y Simple Mixture Martingale.

Como podemos ver en la figura 3.2 las martingalas construidas de esta forma serán crecientes si se analiza una secuencia suficiente de ejemplos cuyos p-valores tengan valores pequeños.

Randomized Power Martingale

Tal y como se propone en [13] y basado en el teorema 3.1, los p-valores pueden ser utilizados para construir supermartingalas intercambiables y martingalas intercambiables aleatorias. Por tanto tenemos que la *función de apuesta* tendría la siguiente forma [12]:

$$\forall i : f_i(p) = \epsilon p^{\epsilon-1} \quad (3.11)$$

donde $\epsilon \in [0, 1]$. Varias martingalas pueden ser construidas a partir de

esta *función de apuesta*. Entre ellas tenemos que:

$$M_n^\epsilon = \prod_{i=1}^n \epsilon p_i^{\epsilon-1} = \epsilon^n \left(\prod_{i=1}^n p_i \right)^{\epsilon-1} \quad (3.12)$$

Donde el valor inicial de esta martingala será 1 y que llamaremos *randomized power martingale*.

Simple Mixture Martingale

A partir de la randomized power martingale y con el fin de eliminar la dependencia de ϵ , podemos utilizar la siguiente martingala intercambiable aleatoria:

$$M_n = \int_0^1 M_n^\epsilon d\epsilon \quad (3.13)$$

A la ecuación 3.13 le llamaremos *Simple mixture martingale*.

Plug-In Martingale

ESTIMACIÓN KERNEL DE LA FUNCIÓN DE DENSIDAD DE PROBABILIDAD

Varias son las referencias disponibles para definir estimadores de la densidad por kernel. Seguiremos la descripción que se da en [15], y que también es conocida como **Parzen window**:

- Estimador de Kernel Univariable

En este caso el estimador kernel es dado por:

$$f_{Ker}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \quad (3.14)$$

donde $K(t)$ es el **kernel**, n es el número de ejemplos y h el ancho de ventana. El kernel debe satisfacer la condición $\int K(t)dt = 1$ para asegurar que nuestra estimación en la ecuación 3.14 es un buen estimador de la densidad. Si definimos $K_h(t) = K(t/h)/h$, entonces podemos también escribir el estimador kernel de la siguiente forma:

$$f_{Ker}(x) = \frac{1}{n} \sum_i^n K_h(x - X_i) \quad (3.15)$$

Habitualmente, el kernel es una función de densidad de probabilidad que suele utilizar la densidad normal estándar. No obstante veremos que se pueden utilizar otros tipos de kernels que pueden ser más adecuados para determinados fines.

A continuación se describe un procedimiento para generar el kernel:

1. Elegir un kernel, un parámetro de suavizado h y el dominio (conjunto de valores x) sobre los que evaluar $f(x)$.
2. Para cada X_i , evaluar el siguiente kernel para todo el dominio:

$$K_i = K\left(\frac{x - X_i}{h}\right) \quad i = 1, \dots, n. \quad (3.16)$$

El resultado de esta fórmula será un conjunto de n curvas, una para cada dato de X_i .

3. Peso de cada curva para $1/h$.
4. Para cada x , tomar la media de las curvas ponderadas.

El parámetro h determina el volumen de suavizado que tenemos en el estimador $f_{Ker}(x)$. un valor pequeño de h genera una curva rugosa, mientras

que un valor grande de h genera una curva suavizada. Cuando el ancho de la ventana es pequeño, los resultados del estimador son poco precisos principalmente por la incorporación de ruido. Para valores grandes de h obtenemos estimadores suavizados que pueden obviar variaciones importantes e interesantes de la estimación. Dado que diferentes valores de h generan diferentes tipos de clasificadores, es necesario examinar diferentes tipos de estimaciones del kernel de densidad para diferentes tipos de anchos de ventana, y de esta forma decidir cuál es el mejor.

A continuación veremos la forma de determinar los valores óptimos del parámetro h . Bajo ciertas condiciones, se sabe que el AMISE ¹ para un estimador kernel de densidad univariable es:

$$AMISE_{Ker}(h) = \frac{R(K)}{nh} + \frac{1}{4}\sigma_k^4 h^4 R(f'') \quad (3.17)$$

Donde el kernel K es una función de densidad de probabilidad continua con $\mu_k = 0$ y $0 < \sigma_k^2 < \infty$. El ancho de ventana que minimiza la ecuación 3.17 viene dada por

$$h_{Ker}^* = \left(\frac{R(K)}{n\sigma_k^4 R(f'')} \right)^{1/5} \quad (3.18)$$

Para el caso de un kernel que es igual a la densidad normal tenemos que $R(f'') = 3/(8\sqrt{\pi}\sigma^5)$

¹Asymtotic Mean Integrated Absolute Error

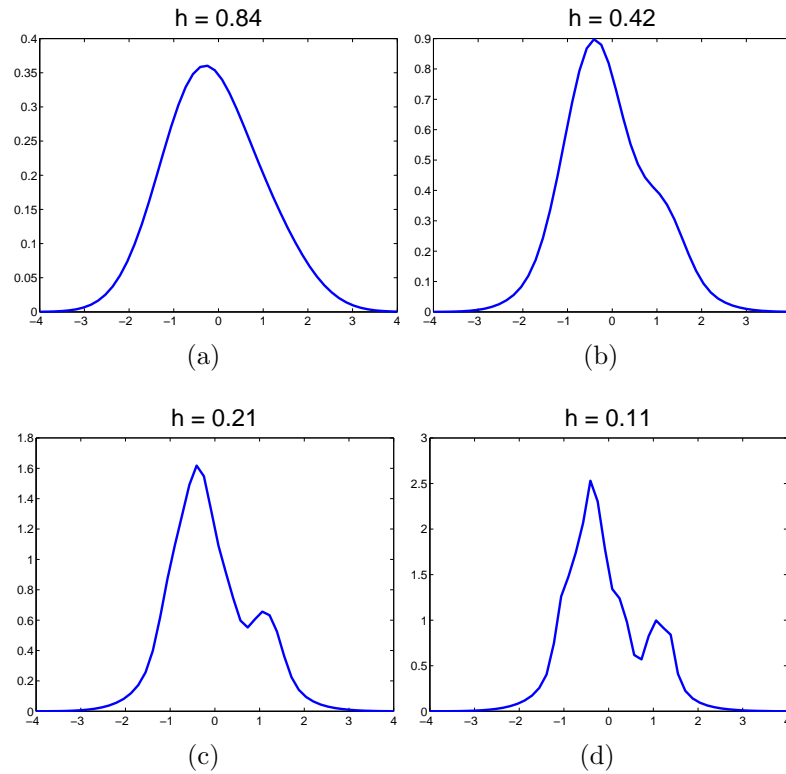


Figura 3.3: Cuatro estimaciones del kernel de densidad utilizando $n = 100$ y variables generadas aleatoriamente siguiendo distribución normal. observese que cuanto menor es el valor de h , la estimación es más tosca.

- Normal Reference Rule - Kernels

También conocida como regla del pulgar de Silverman.

$$h_{Ker}^* = \left(\frac{4}{3}\right)^{1/5} \sigma n^{-1/5} \approx 1.06 \sigma n^{-1/5} \quad (3.19)$$

Podemos utilizar alguna estimación idónea para σ , de modo que la desviación estándar tenga la forma $\sigma = IQR/1.348$. Con esto tendríamos un ancho de ventana de

$$h_{Ker}^* = 0.786 \times IQR \times n^{-1/5}$$

La selección del tipo de kernel para el estimador es otro problema que se debe abordar. En el cuadro 3.1 se pueden ver ejemplos de tipos de kernels.

Nombre del Kernel	Ecuación
Triángulo	$K(t) = (1 - t) \quad -1 \leq t \leq 1$
Epanechnikov	$K(t) = \frac{3}{4}(1 - t^2) \quad -1 \leq t \leq 1$
Cuártico o Biweight de Tukey	$K(t) = \frac{15}{16}(1 - t^2)^3 \quad -1 \leq t \leq 1$
Triweight	$K(t) = \frac{35}{32}(1 - t^2)^3 \quad -1 \leq t \leq 1$
Normal o Gaussiano	$K(t) = \frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}} \quad -\infty < t < \infty$
Coseno	$K(t) = \frac{\pi}{4} \cos\left(\frac{\pi}{2}t\right) \quad -1 \leq t \leq 1$
K_2 de Silverman	$K(t) = \frac{3}{\pi}(1 - t^2)^2 \quad -1 \leq t \leq 1$

Cuadro 3.1: Ejemplos de Kernels para la estimación de la densidad.

Es conocido que la elección de un kernel para el cálculo de la función de densidad es menos importante que la selección adecuada del parámetro h . Esto es así porque el efecto que puede producir la selección de un kernel queda suavizado por la media del proceso de cálculo. Nos centraremos por tanto en la eficiencia de los diferentes kernels que tiene que ver sobretudo con el coste computacional y la cantidad de diferenciabilidad requerida en la

estimación. El más eficiente es el kernel de Epanechnikov [16].

$$f(n) = \begin{cases} \frac{3}{4}(1 - t^2) & \text{si } -1 \leq t \leq 1 \\ 0 & \text{si cualquier otro valor} \end{cases} \quad (3.20)$$

En el cuadro 3.1 podemos ver diferentes tipos de kernels. Estos kernels tienen una eficiencia muy parecida al kernel de Epanechnikov, siendo el kernel normal o Gaussiano el menos eficiente de todos.

NUEVA METODOLOGÍA PLUG-IN

Utilizaremos como estimación de densidad de probabilidad la función de apuesta $f_i(p)$. En cada iteración la estimación de la función de densidad de probabilidad es calculada usando los p-valores acumulados:

$$p_i(p) = \hat{p}(p_1, \dots, p_{i-1}, p) \quad (3.21)$$

donde $\hat{p}(p_1, \dots, p_{i-1}, p)$ es la estimación de la probabilidad de la función de densidad usando los p-valores p_1, \dots, p_{i-1} generados por el algoritmo 1.

Sustituyendo estas funciones de apuesta en 3.9 obtenemos una nueva martingala llamadas *plug-in* martingale.

Un ejemplo de implementación de este tipo de martingalas es el que a continuación se detalla. Los p-valores se encuentran siempre en el intervalo $[0, 1]$, el método estándar de Parzen window da resultados pobres para valores que están cerca del límite. Con el fin de obtener mejores resultados para los puntos cercanos al límite la secuencia de p-valores es reflejada a la izquierda del cero y a la derecha del uno. De modo que la estimación kernel de la densidad se calcula con la muestra extendida $U_{i=1}^n \{-p_i, p_i, 2 - p_i\}$. La función de estimación de la densidad es inicializada a cero fuera del intervalo de la unidad y normalizada mediante la integración a uno. Para los resultados presentados en este trabajo se utilizarán como parámetros el kernel Gaussiano y la regla del pulgar de Silverman (Normal reference rule) para el cálculo del

valor de h . Tal y como se comenta en [12] no se han encontrado diferencias utilizando otro tipo de kernels.

Los valores S_n de los plug-in martingale pueden ser actualizados recursivamente. Esto supone calcular los valores de no conformidad $(\alpha_1, \dots, \alpha_n)$ desde (z_1, \dots, z_n) para lo cual emplea el tiempo $g(n)$ y la evaluación de 3.21 toma un tiempo $h(n)$. Con lo cual es fácil colegir que calcular el valor α_n en el intervalo $\alpha_1, \dots, \alpha_n$ tomará el tiempo $\Theta(n)$.

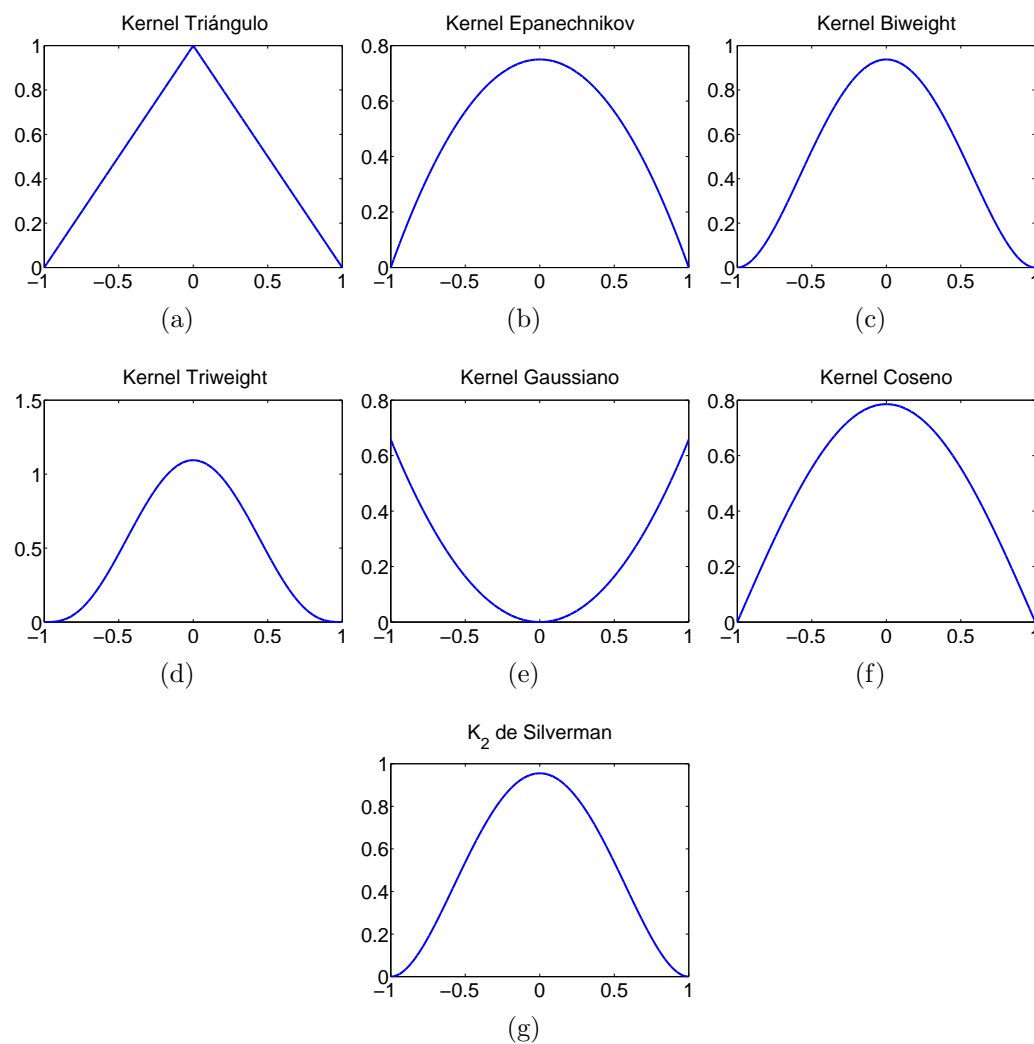


Figura 3.4: Gráficos de varios kernels que pueden utilizarse para estimar la función de densidad. a) Triángulo, b) Epanechnikov, c) Biweight, d) Triweight, e) Normal o Gaussiano, f) k_2 Silverman

3.7. Test de martingalas para la detección de cambios.

En esta sección se describirá una arquitectura basada en martingalas para la detección de cambios. En esta arquitectura cuando se observa un nuevo dato se realiza el test de hipótesis para decidir si ha ocurrido un cambio en el modelo que genera el flujo de datos. La decisión está basada en si se produce o no una violación de la presunción de intercambiabilidad utilizando para ello cualquiera de las martingalas descritas en el apartado 3.6.5. Como hemos comentado con anterioridad, vamos a considerar el siguiente test de hipótesis:

H_0 : "no hay cambios en el flujo de datos", frente a la alternativa H_1 : "ha ocurrido un cambio en el flujo de datos". Este test se estaría realizando mientras que:

$$0 < M_n^{(\epsilon)} < \lambda \quad (3.22)$$

donde λ es un número positivo. El test rechaza la hipótesis de nulidad H_0 cuando $M_n^{(\epsilon)} \geq \lambda$.

Supongamos que $\{M_k : 0 \leq k < \infty\}$ es una martingala no negativa. Si $E(M_n) - E(M_0) = 1$, entonces y a partir de la máxima desigualdad de Doob,

$$\lambda P \left(\max_{0 \leq k \leq n} M_k \geq \lambda \right) \leq E(M_n), \quad (3.23)$$

y para cualquier valor de $\lambda > 0$ y $n \in \mathbb{N}$, se tiene que:

$$P \left(\max_{k \leq n} M_k \geq \lambda \right) \leq \frac{1}{\lambda} \quad (3.24)$$

Esta desigualdad nos indica que es improbable que el valor de M_k sea grande. Se rechazará la hipótesis nula cuando el valor de la martingala sea mayor que λ . La desigualdad de la ecuación 3.24 es una cota superior para el ratio de falsas alarmas (FAR ²), que es una medida de cambios detectados

²False Alarm Rate

cuando realmente no los hay. El valor de λ es determinado por el FAR que se está dispuesto a aceptar.

3.7.1. Algoritmo para el Test de Martingalas

La metodología de detección de cambios que se propone en el Algoritmo 2, es una consecuencia directa de la máxima desigualdad de Doob para martingalas. Se puede ver que $M_i^{(\epsilon)} = \epsilon p_i^{\epsilon-1} M_{i-1}^{(\epsilon)}$. Por tanto, no es necesario recalcular $M(i)$ en el paso 9 para $p_j, j = 1, \dots, i - 1$.

Se observa que la desigualdad máxima de Doob de la ecuación 3.23 es en general suficiente para que la ecuación 3.24 sea verdad para cualquier k entera positiva. Esto significa que la ecuación 3.23 es verdadera mientras que el modelo que genere los datos se mantenga estable.

Sea α el tamaño del test decidiendo a favor de la hipótesis alternativa H_1 y cuando la hipótesis nula H_0 es verdadera y $1 - \beta$ es la potencia del test decidiendo a favor de una hipótesis alternativa cuando es verdadera, el test de martingalas de acuerdo a la ecuación 3.24 que es una aproximación a SPRT ³, con

$$\lambda \leq \frac{1 - \beta}{\alpha} \quad (3.25)$$

y el tiempo medio de retardo $E(m)$. Es decir, el número de ejemplos observados esperados antes de que un cambio es detectado, es aproximado por SPRT de la siguiente forma:

$$E(m) \approx \frac{(1 - \beta) \log \lambda}{E(\mathcal{L})} \quad (3.26)$$

donde

$$\mathcal{L} = \log \epsilon p_i^{\epsilon-1} \quad (3.27)$$

³sequential probability ratio test

Algoritmo 2 Test de Martingalas

Entrada: $M(0) \leftarrow 1; i \leftarrow 1; T \leftarrow \{\}$; Inicializar valor λ

```
1: repetir
2:   Un nuevo ejemplo  $x_i$  es observado
3:   si  $T \leftarrow \{\}$  entonces
4:      $x \leftarrow 0$ 
5:   si no
6:     Calcular la medida de extrañeza (ver sección 3.5.2) de  $x_i$  respecto a
       los elementos procesados de  $T$ .
7:   fin si
8:   Calcular los p-valores  $p_i$  utilizando la función 3.5
9:   Calcular  $M(i)$  mediante la función 3.9
10:  si  $M(i) > \lambda$  entonces
11:    DETECTADO CAMBIO
12:     $M(i) = 1$ ;
13:    Reinicializar  $T$  a conjunto vacío.
14:  si no
15:    Añadir  $x_i$  a  $T$ .
16:  fin si
17:   $i \leftarrow i + 1$ ;
18: fin repetir
```

3.7.2. Rendimiento del test de martingala

Una parte importante para un algoritmo que quiera analizar un flujo de datos es el rendimiento del mismo. A continuación analizaremos el rendimiento del algoritmo 2. Este rendimiento está relacionado con el paso 6 de dicho algoritmo, paso en el que se calcula la *extrañeza* o medida de no conformidad. El cálculo de los p-valores en el paso 8 puede realizarse con un montículo ordenado cuya complejidad sería $O(n \log n)$ y cuyo coste es despreciable frente al coste del cálculo de la *extrañeza* del paso 6.

Un flujo de datos muy grande en el que no se producen cambios hace que la medida de *extrañeza* del p-valor p_i tenga que realizarse contra una bolsa $\{z_1, \dots, z_n\}$ con muchos elementos, lo cual es muy costoso. El coste además está relacionado con la medida de no conformidad que se seleccione. Una posible solución a este problema podría ser el uso de dos o más test de martingalas las cuales arrancarían y reorganizarían en distintos momentos del tiempo antes de que se produzca un desbordamiento de memoria o tiempo de ejecución (time-out). Un criterio de detección de cambio podría ser o bien el del primer cambio que se produzca o una decisión de conjunto.

3.8. Test de múltiples martingalas utilizando multi-vistas

A continuación se sigue lo que se describe en [17]. En el problema de aprendizaje multi-vista, un ejemplo z es representado por un número de subconjunto de características. Cada subconjunto de características describe una vista de los ejemplos. El escenario de multi-vistas está íntimamente relacionado con el coentrenamiento (co-training).

Para el test de múltiples martingalas, consideraremos el escenario multi-vista de modo que cada martingala construida intenta identificar cambios con respecto a un subconjunto particular del conjunto de características. Además el hecho de que las características son extraídas de los datos originales, el conjunto de características debería ser independiente uno de otro con el objeto de minimizar la redundancia.

Apuntaremos que de acuerdo con la teoría de martingalas, la representa-

Algoritmo 3 Test de múltiples martingalas con M-vistas, M=2

Entrada: $M_1(0) \leftarrow 1; M_2(0) \leftarrow 1; i \leftarrow 1; T \leftarrow \{\}$; Inicializar valor λ

- 1: **repetir**
 - 2: Un nuevo ejemplo sin etiquetar x_i es observado
 - 3: Construir las dos vistas/características desde x_i , ej.: $\bar{x}_i = \{f_{i1}, f_{i2}\}$
 - 4: Calcular la medida de extrañeza \bar{s}_1 y \bar{s}_2 utilizando la función 3.5 a partir de $\{f_{11}, \dots, f_{i1}\}$ y $\{f_{12}, \dots, f_{i2}\}$.
 - 5: Calcular p-valores p_1 y p_2 a partir de \bar{s}_1 y \bar{s}_2 , respectivamente, utilizando 3.5
 - 6: Calcular $M_1(i)$ y $M_2(i)$ a partir de p_1 y p_2 mediante la función 3.9
 - 7: **si** $M_1(i) > \lambda$ **O** $M_2(i) > \lambda$ **entonces**
 - 8: **DETECTADO CAMBIO**
 - 9: $M_1(i) = M_2(i) = 1$;
 - 10: Reinicializar T a conjunto vacío.
 - 11: **si no**
 - 12: Añadir x_i a T .
 - 13: **fin si**
 - 14: $i \leftarrow i + 1$;
 - 15: **fin repetir**
-

ción de los datos no afecta al límite de probabilidad.

Corolario 1. *Cuando el test de múltiples martingalas con M-vistas es utilizado para la detección de cambios, el número esperado de datos, m , observados antes de que se detecte un cambio es,*

$$E_{\mathcal{M}}^*(m) \leq E(m) \quad (3.28)$$

Por lo que podemos concluir que el número de detección pasadas por alto utilizando un test de múltiples martingalas está acotado superiormente por el número de detecciones pasadas por alto que se obtiene utilizando el método de la martingala original. Además, el número de falsas alarmas utilizando el test de múltiples martingalas está acotado inferiormente por el número de falsas alarmas que se obtiene utilizando el método de martingalas original. El algoritmo 3 describe el algoritmo de test de múltiples martingalas utilizando dos vistas. El algoritmo puede ser extendido de forma trivial a $M > 2$.

Capítulo 4

Prueba de capacidades del test de martingalas

En este apartado vamos a realizar una serie de baterías de experimentos que nos permitan discernir la bondad de los desarrollos elaborados para el análisis con la arquitectura de martingalas que se ha fijado en la sección 3.7. Concretamente probaremos las capacidades y limitaciones de detección de cambios en flujos de datos.

4.1. Descripción de los experimentos

Para la realización de los diferentes experimentos vamos a generar perfiles de evolución parecidos a los de la figura 2.1. Estos perfiles teóricos nos van a servir para evaluar la sensibilidad de la arquitectura de martingalas, y validar la posible aplicabilidad a la detección de cambios en los perfiles experimentales de evolución temporal del stellarator TJ-II. Para ello jugaremos con la altura, anchura y ruido de los perfiles que generemos, de forma que se generarán grupos de perfiles con diferentes características. La unión de estos grupos simulará un flujo de datos con cambios (cada vez que finaliza un grupo de señales y comienza otro con características diferentes) que nuestra arquitectura de martingalas deberá detectar.

4.2. Análisis de sensibilidad

En los experimentos que se van a mostrar en las siguientes secciones se tomará $\lambda = 20$. Según la máxima desigualdad de Doob, ecuación 3.24, con este valor tendríamos un error del 5%, lo cual sería suficientemente preciso para el tipo de problemas que estamos analizando.

4.3. Experimentos de detección de cambios en perfiles

Para la simulación de los perfiles que utilizaremos en los experimentos se ha desarrollado un programa en Matlab que simula los perfiles siguiendo la siguiente ecuación:

$$f(x) = A_0(1 - |r^{\alpha_1}|)^{\alpha_2} \quad (4.1)$$

Dónde,

A_0 = altura de los perfiles. α_1 = concentración de los valores centrales de los perfiles o nivel de *achatamiento* de los perfiles. α_2 = longitud de las *colas* de los perfiles. En la figura 4.2 se puede observar varios de los efectos que producen estas variables.

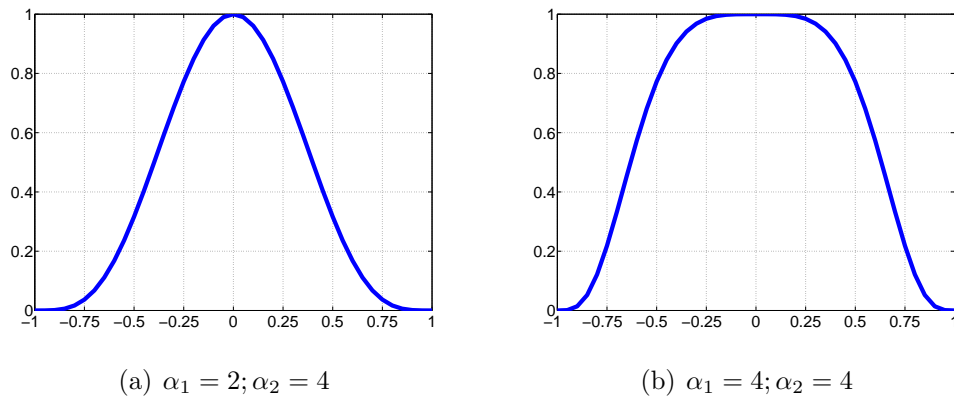


Figura 4.1: Ejemplos de efectos de la variable α_1 de la función 4.1.

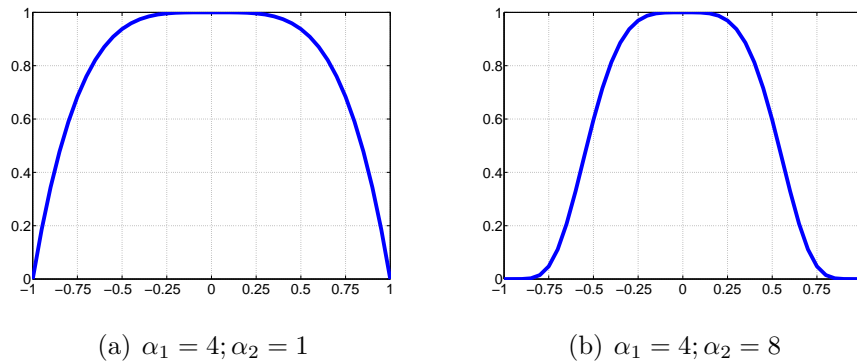
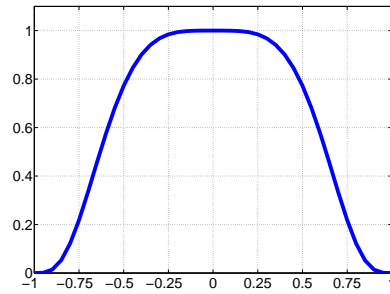
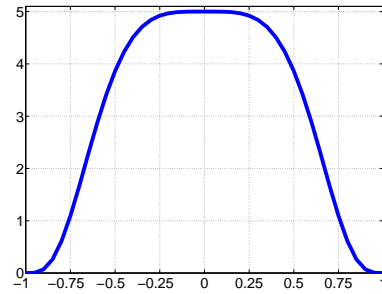
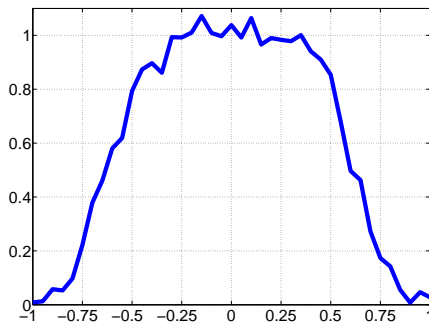


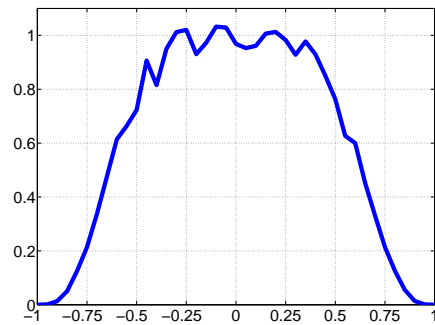
Figura 4.2: Ejemplos de efectos de la variable α_2 de la función 4.1.

Cada perfil consta de 41 puntos. Además de esto implementaremos una serie de variables y funciones que permitirán modificar la generación de perfiles simulados. A continuación se detallaremos las variables y funciones con las que podemos jugar:

- A_0 : Altura de los perfiles.
- Factor de ruido. Nos permite simular ruido en los perfiles simulados. Este ruido puede ser a su vez dependiente o independiente de la amplitud de los perfiles simulados.
 - El ruido independiente se genera añadiendo a cada perfil el valor resultante de la función $f(x) = fct \cdot randn$, donde fct es el factor de ruido y $randn$ es una función que genera número aleatorios que siguen un distribución Gaussiana.
 - El ruido dependiente se genera añadiendo a cada perfil el valor resultante de la función $f(x) = 1 + fct \cdot randn$.
- Factor de suavizado: suaviza perfiles simulados, a los cuales previamente se les ha añadido un ruido dependiente o independiente de la amplitud de los perfiles. Para ello utilizamos splines cúbicos.

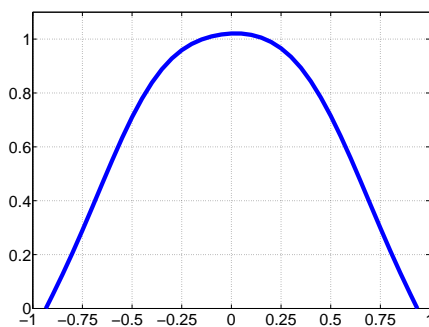
(a) $A_0 = 1$ (b) $A_0 = 5$ Figura 4.3: Ejemplos de efectos de la variable A_0 en la función 4.1.

(a) Con ruido dependiente

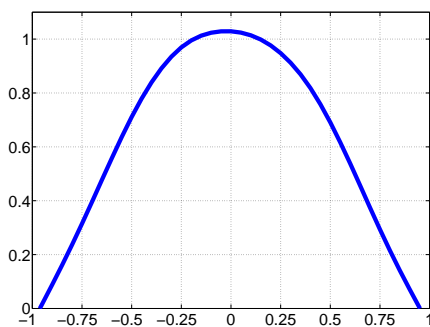


(b) Con ruido independiente

Figura 4.4: Ejemplos de efectos de ruido simulado para la función 4.1.



(a) Con ruido dependiente



(b) Con ruido independiente

Figura 4.5: Ejemplos de suavizado para los casos de la figura 4.4 con $\alpha_1 = 4$, $\alpha_2 = 4$.

4.3.1. Experimento 1

En este caso se generan 2 grupos de perfiles cada uno con una amplitud en el centro diferente. A dichos perfiles además se les añade un factor de ruido gaussiano dependiente, sin factor de suavizado posterior.

- Grupo 1: $\alpha_1 = \alpha_2 = 4$; fct = 0,01; $A_0 = 1, 2$
- Grupo 2: $\alpha_1 = \alpha_2 = 4$; fct = 0,01; $A_0 = 4$

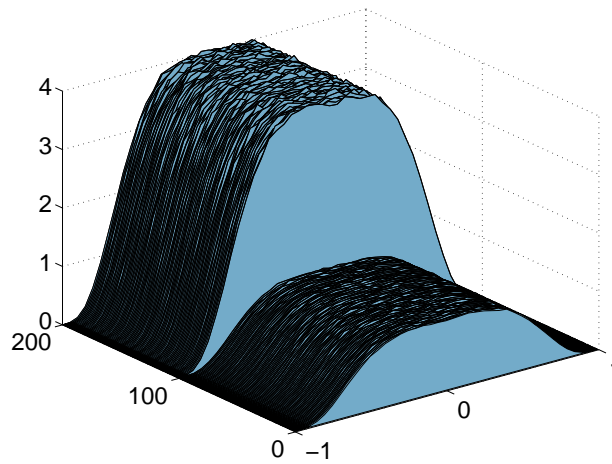


Figura 4.6: Perfiles Experimento 1.

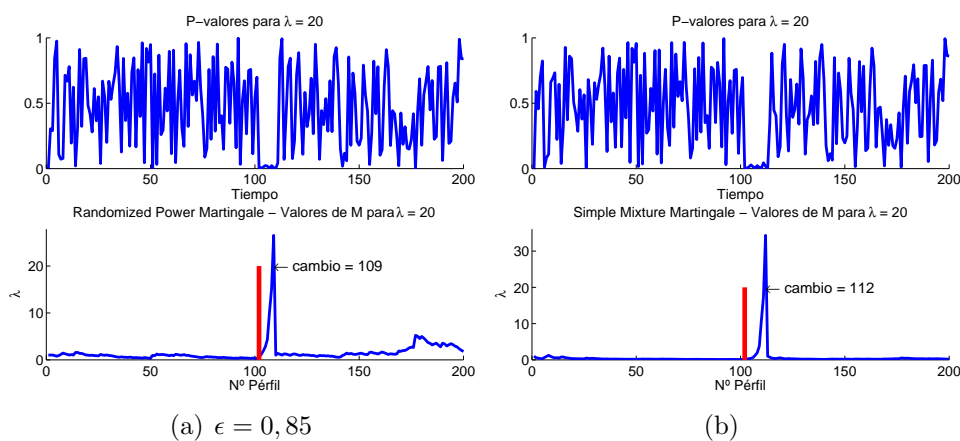


Figura 4.7: Ejemplos de detección del experimento 1. Se indica en cada gráfico el momento en el que la arquitectura de martingalas identifica el cambio.

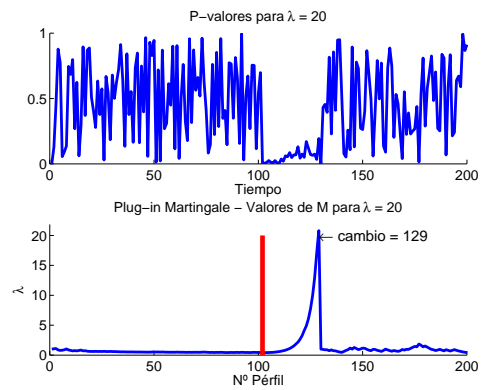


Figura 4.8: Ejemplos de detección del experimento 1. Se indica en cada gráfico el momento en el que la arquitectura de martingalas identifica el cambio.

En este caso podemos observar en las figuras 4.7 y 4.8 el momento en el que se detecta el cambio con un retardo. La línea roja marca el momento del cambio real. Debido al carácter estocástico de las martingalas, detallamos valores obtenidos después de ejecutar el algoritmo 100 veces para cada una de estas martingalas.

Randomized Power Martingale

Para esta martingala se han obtenido resultados para 10 ϵ distintas. Podemos ver el resultado obtenido en el cuadro 4.1. En la figura 4.9 podemos ver el resultado de detección de cambios, para 100 ejecuciones, mediante un gráfico Box Plot de resultados. Se han eliminado las falsas alarmas (FA) para evitar desvirtuar los resultados.

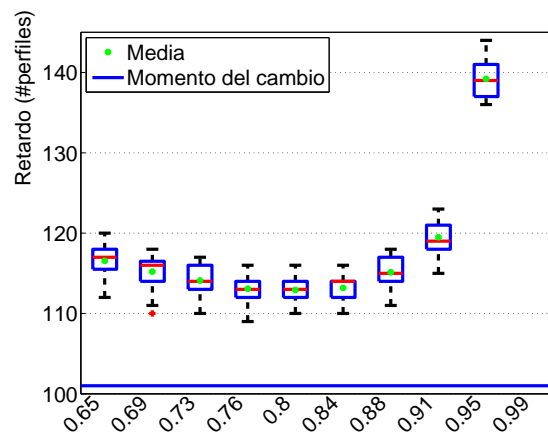


Figura 4.9: Gráfico Box Plot para 100 ejecuciones de Randomized Power Martingales por cada una de las 10 ϵ .

ϵ	Retardo (Nº perfiles)	FA	Sin detectar
0,65	15,54	2	-
0,69	14,20	3	-
0,73	13,11	5	-
0,76	12,08	8	-
0,8	11,92	3	-
0,84	12,19	-	-
0,88	14,13	-	-
0,91	18,49	-	-
0,95	38,21	-	-
0,99	-	-	100

Cuadro 4.1: Media de retardos, FA y cambios no detectados para un conjunto de 100 ejecuciones de la Randomized Power Martingale y varios valores de ϵ .

Simple Mixture

A continuación mostramos resultados para 100 ejecuciones de Simple Mixture.

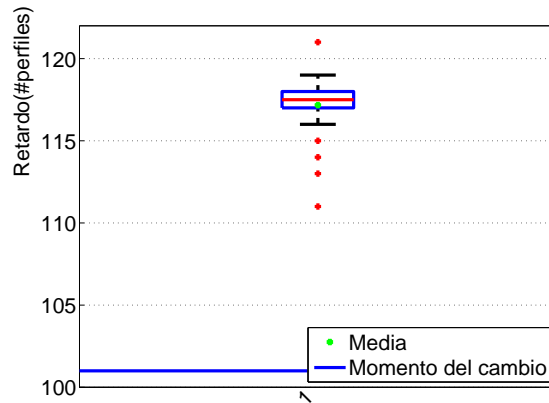


Figura 4.10: Gráfico Box Plot para 100 ejecuciones.

Retardo (Nº perfiles)	FA	Sin detectar
16,17	2	-

Cuadro 4.2: Media de retardos, FA y cambios no detectados para un conjunto de 100 ejecuciones de la Simple Mixture Martingale.

Plug-in Martingale

En este caso veremos los resultados para esta martingala para 100 ejecuciones y 5 valores de FCT distintas. FCT es un factor que multiplica el valor de h (Normal Reference Rule).

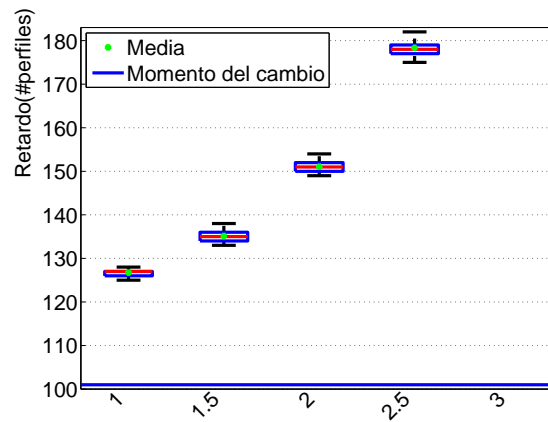


Figura 4.11: Gráfico Box Plot para 100 ejecuciones de Plug-In Martingale y 5 factores (FCT) diferentes.

FCT	Retardo (N° perfiles)	FA	Sin detectar
1	25,74	-	-
1,5	34,16	-	-
2	50,06	-	-
2,5	77,35	-	-
3	-	-	100

Cuadro 4.3: Media de retardos, FA y cambios no detectados para un conjunto de 100 ejecuciones de la Plug-in Martingale.

Conclusiones

Los menores retardos se obtienen para Randomized Power Martingale con un valor de $\epsilon \in [0,75, 0,85]$, seguido de Simple Mixture Martingale. Plug-in es la que arroja un tiempo de retardo peor. No obstante, es la que produce menos FA.

4.3.2. Experimento 2

En este caso se replica el experimento anterior pero con un cambio más suave:

- Grupo 1: $\alpha_1 = \alpha_2 = 4$; $fct = 0,01$; $A_0 = 1, 2$
- Grupo 2: $\alpha_1 = \alpha_2 = 4$; $fct = 0,01$; $A_0 = 2, 2$

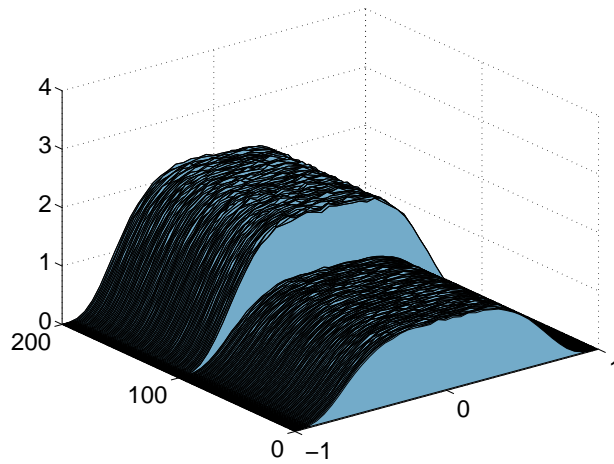


Figura 4.12: Perfiles Experimento 2.

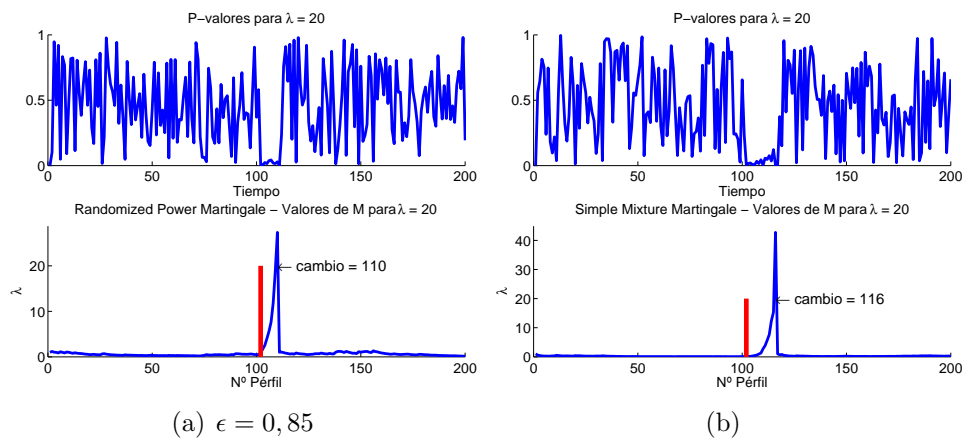
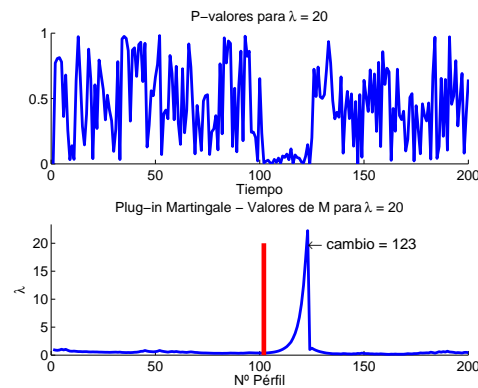


Figura 4.13: Ejemplos de detección del experimento 2. Se indica en cada gráfico el momento en el que la arquitectura de martingalas identifica el cambio.



(a) Factor de multiplicación de h (FCT)
 $= 1$

Figura 4.14: Ejemplos de detección del experimento 2. Se indica en cada gráfico el momento en el que la arquitectura de martingalas identifica el cambio.

En este caso podemos observar en las figuras 4.13 y 4.14 el momento en el que se detecta el cambio con un retardo. La línea roja marca el momento del cambio real. Debido al carácter estocástico de las martingalas, detallamos valores obtenidos después de ejecutar el algoritmo 100 veces para cada una de estas martingalas.

Randomized Power Martingale

Para esta martingala se han obtenido resultados para 10 ϵ distintas. Podemos ver el resultado obtenido en el cuadro 4.4. En la figura 4.15 podemos ver el resultado de detección de cambios, para 100 ejecuciones, mediante un gráfico Box Plot de resultados. Se han eliminado las falsas alarmas (FA) para evitar desvirtuar los resultados.

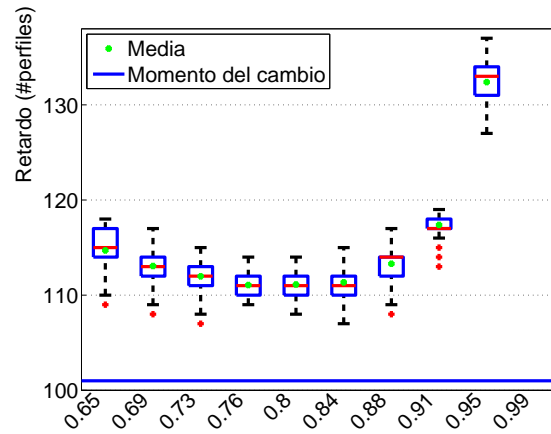


Figura 4.15: Gráfico Box Plot para 100 ejecuciones de Randomized Power Martingale por cada una de las 10 ϵ .

ϵ	Retardo (N° perfiles)	FAR	Sin detectar
0,65	13,70	-	-
0,69	12,06	1	-
0,73	10,98	-	-
0,76	10,06	-	-
0,8	10,12	-	-
0,84	10,35	-	-
0,88	12,30	-	-
0,91	16,38	-	-
0,95	31,40	-	-
0,99	-	-	100

Cuadro 4.4: Media de retardos, FAR's y cambios no detectados para un conjunto de 100 ejecuciones de la Randomized Power Martingale y varios valores de ϵ .

Simple Mixture

A continuación mostramos resultados para 100 ejecuciones de Simple Mixture.

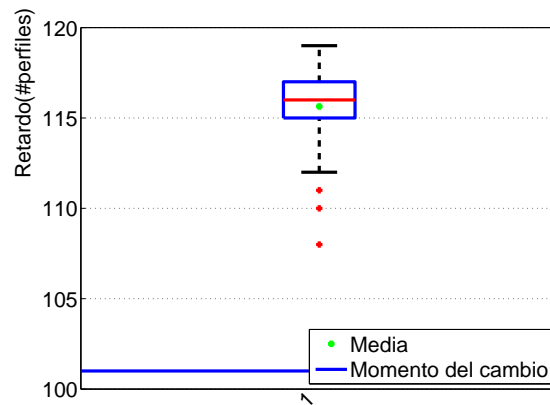


Figura 4.16: Gráfico Box Plot para 100 ejecuciones.

Retardo (Nº perfiles)	FAR	Sin detectar
14,64	-	-

Cuadro 4.5: Media de retardos, FAR's y cambios no detectados para un conjunto de 100 ejecuciones de la Simple Mixture Martingale.

Plug-in Martingale

En este caso veremos los resultados para esta martingala para 100 ejecuciones y 5 valores de FCT distintas. FCT es un factor que multiplica el valor de h (Normal Reference Rule).

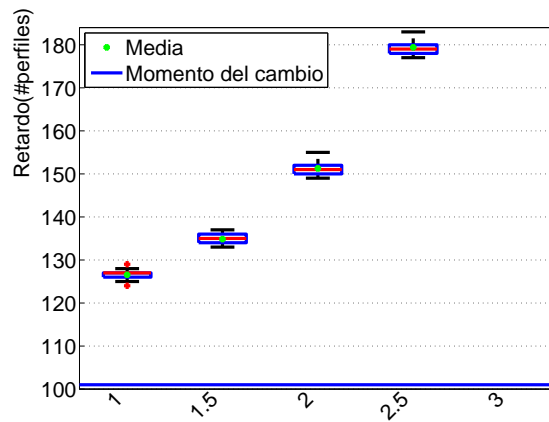


Figura 4.17: Gráfico Box Plot para 100 ejecuciones de Plug-In Martingale y 5 factores (FCT) diferentes de multiplicación de h (Normal Reference Rule).

FCT	Retardo (N° perfiles)	FAR	Sin detectar
1	25,55	-	-
1,5	33,86	-	-
2	50,27	-	-
2,5	78,40	-	-
3	-	-	100

Cuadro 4.6: Media de retardos, FAR's y cambios no detectados para un conjunto de 100 ejecuciones de la Plug-in Martingale.

Conclusiones

No parece que haya una relación directa entre el retardo y el tamaño de la diferencia del cambio.

4.3.3. Experimento 3

En este caso se replica el experimento anterior pero los perfiles son generados *sin ruido*:

- Grupo 1: $\alpha_1 = \alpha_2 = 4$; $A_0 = 1, 2$
- Grupo 2: $\alpha_1 = \alpha_2 = 4$; $A_0 = 2, 2$

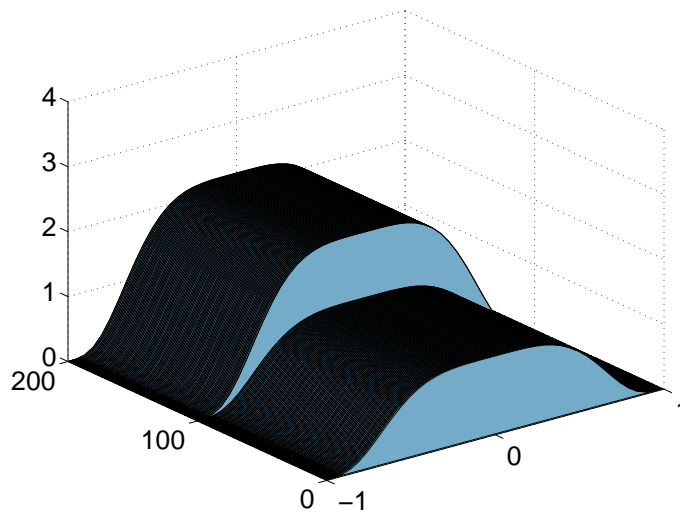


Figura 4.18: Perfiles Experimento 3.

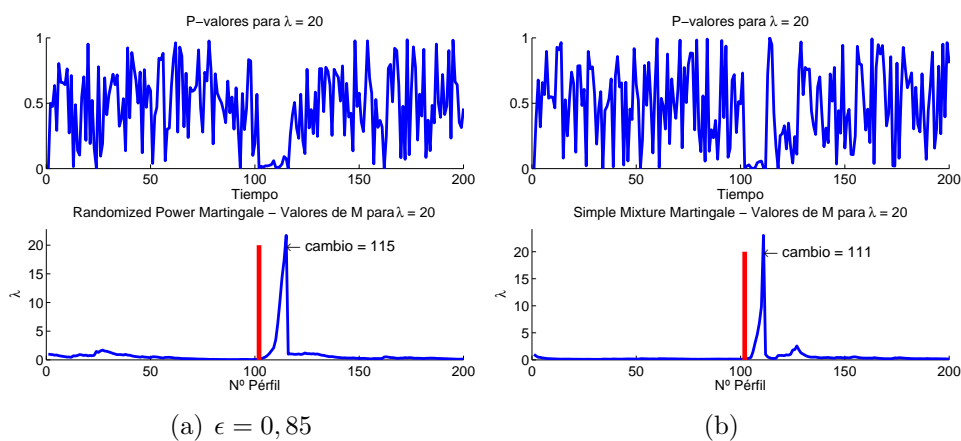
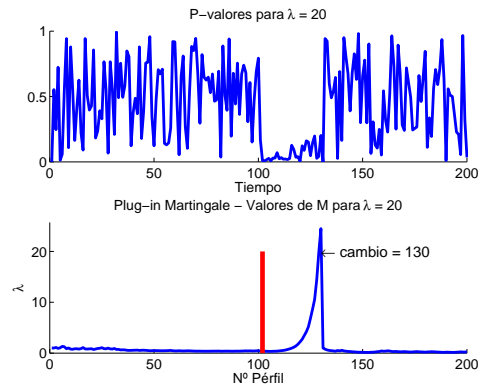


Figura 4.19: Ejemplos de detección del experimento 3. Se indica en cada gráfico el momento en el que la arquitectura de martingalas identifica el cambio.



(a) Factor de multiplicación de h (FCT)
 $= 1$

Figura 4.20: Ejemplos de detección del experimento 3. Se indica en cada gráfico el momento en el que la arquitectura de martingalas identifica el cambio.

En este caso podemos observar en las figuras 4.19 y 4.20 el momento en el que se detecta el cambio con un retardo. La línea roja marca el momento del cambio real. Debido al carácter estocástico de las martingalas, detallamos valores obtenidos después de ejecutar el algoritmo 100 veces para cada una de estas martingalas.

Randomized Power Martingale

Para esta martingala se han obtenido resultados para 10 ϵ distintas. Podemos ver el resultado obtenido en el cuadro 4.7. En la figura 4.21 podemos ver el resultado de detección de cambios, para 100 ejecuciones, mediante un gráfico Box Plot de resultados. Se han eliminado las falsas alarmas (FA) para evitar desvirtuar los resultados.

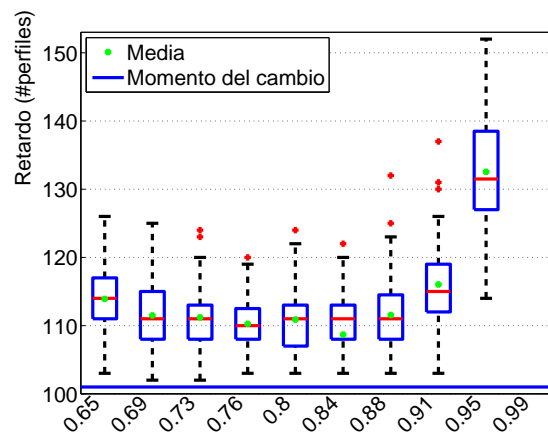


Figura 4.21: Gráfico Box Plot para 100 ejecuciones de Randomized Power Martingale por cada una de las 10 ϵ .

ϵ	Retardo (Nº perfiles)	FAR	Sin detectar
0,65	12,90	8	-
0,69	10,49	11	-
0,73	10,21	5	-
0,76	9,25	4	-
0,8	9,88	2	-
0,84	7,68	5	-
0,88	10,56	1	-
0,91	15,03	-	-
0,95	31,55	-	-
0,99	-	-	100

Cuadro 4.7: Media de retardos, FA y cambios no detectados para un conjunto de 100 ejecuciones de la Randomized Power Martingale y varios valores de ϵ .

Simple Mixture

A continuación mostramos resultados para 100 ejecuciones de Simple Mixture.

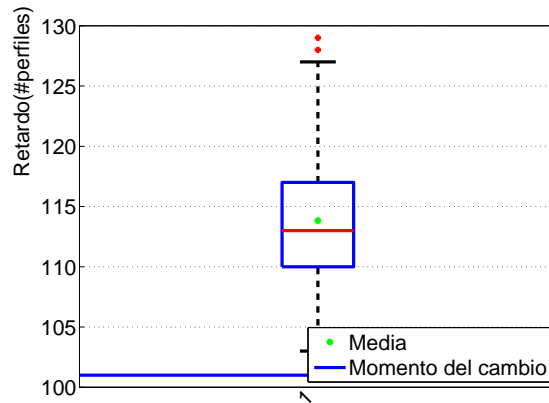


Figura 4.22: Gráfico Box Plot para 100 ejecuciones.

Retardo (Nº perfiles)	FA	Sin detectar
12,83	-	-

Cuadro 4.8: Media de retardos, FA y cambios no detectados para un conjunto de 100 ejecuciones de la Simple Mixture Martingale.

Plug-in Martingale

En este caso veremos los resultados para esta martingala para 100 ejecuciones y 5 valores de FCT distintas. FCT es un factor que multiplica el valor de h (Normal Reference Rule).

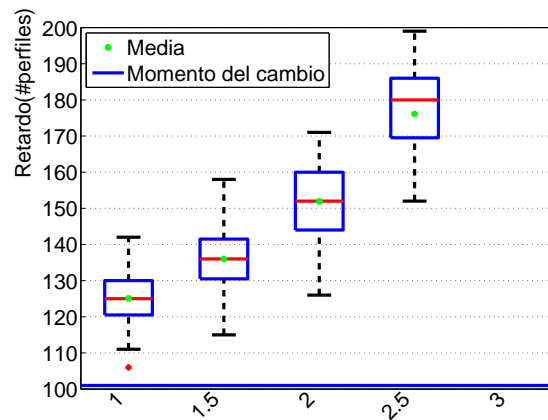


Figura 4.23: Gráfico Box Plot para 100 ejecuciones de Plug-In Martingale y 5 factores (FCT) diferentes de multiplicación de h (Normal Reference Rule).

FCT	Retardo (N° perfiles)	FA	Sin detectar
1	24,08	2	-
1,5	34,99	-	-
2	50,93	-	-
2,5	75,09	-	-
3	-	-	100

Cuadro 4.9: Media de retardos, FA y cambios no detectados para un conjunto de 100 ejecuciones de la Plug-in Martingale.

Conclusión

Comparando estos resultados con los de la sección 4.3.2 vemos que no hay diferencias apreciables para el nivel de ruido que hemos introducido en los perfiles.

Existe cierta dificultad a la hora de comparar resultados porque la dispersión es diferente (mayor para el caso de perfiles sin ruido). Para el caso sin ruido llama la atención que el número de FA sea mayor que para el caso de con ruido. Como en casos anteriores las FA se dan mucho menos en el caso de la Plug-In.

Desde el punto de vista del retardo, Randomized es la que mejor comportamiento tiene junto a la Simple Mixture. El retardo de Plug-In es significativamente mayor que para el resto de martingalas.

4.3.4. Experimento 4

En este caso vamos a probar el grado de sensibilidad para cambios sutiles en la altura de perfiles con ruido:

- Grupo 1: $\alpha_1 = \alpha_2 = 4$; fct = 0,01; $A_0 = 1, 2$
- Grupo 2: $\alpha_1 = \alpha_2 = 4$; fct = 0,01; $A_0 = 1, 21$

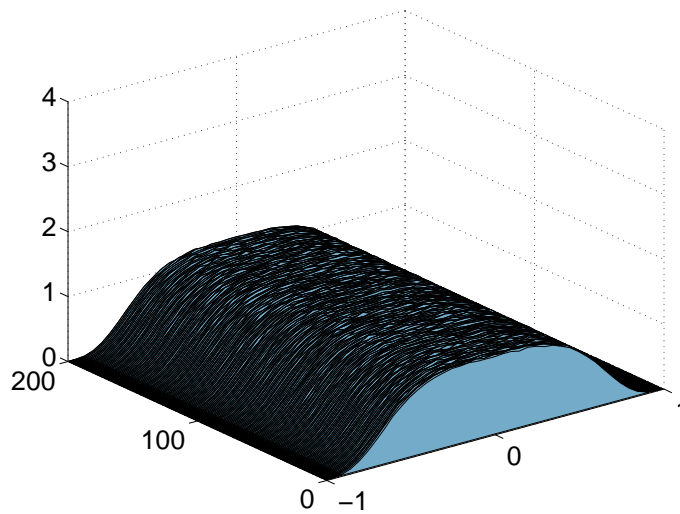


Figura 4.24: Perfiles Experimento 4.

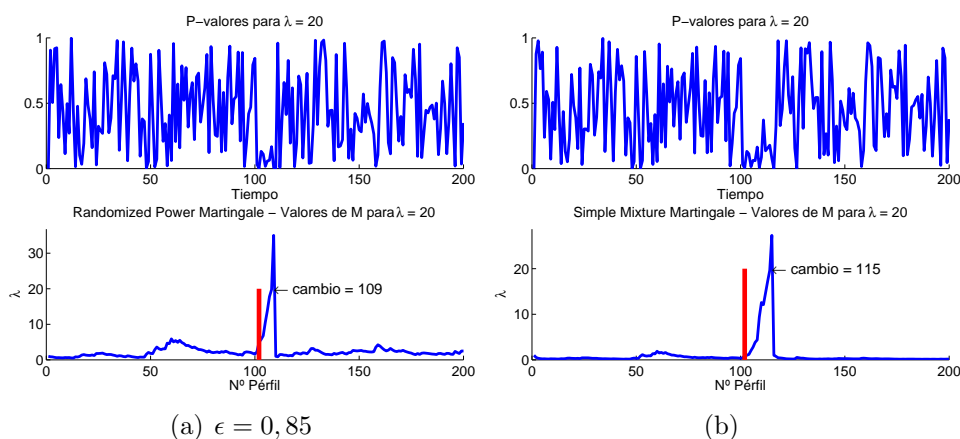
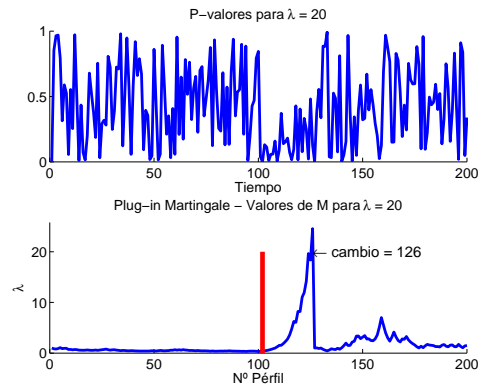


Figura 4.25: Ejemplos de detección del experimento 4. Se indica en cada gráfico el momento en el que la arquitectura de martingalas identifica el cambio.



(a) Factor de multiplicación de h (FCT)
= 1

Figura 4.26: Ejemplos de detección del experimento 4. Se indica en cada gráfico el momento en el que la arquitectura de martingalas identifica el cambio.

En este caso podemos observar en las figuras 4.25 y 4.26 el momento en el que se detecta el cambio con un retardo. La línea roja marca el momento del cambio real. Debido al carácter estocástico de las martingalas, detallamos valores obtenidos después de ejecutar el algoritmo 100 veces para cada una de estas martingalas.

Randomized Power Martingale

Para esta martingala se han obtenido resultados para 10 ϵ distintas. Podemos ver el resultado obtenido en el cuadro 4.10. En la figura 4.27 podemos ver el resultado de detección de cambios, para 100 ejecuciones, mediante un gráfico Box Plot de resultados. Se han eliminado las falsas alarmas (FA) para evitar desvirtuar los resultados.

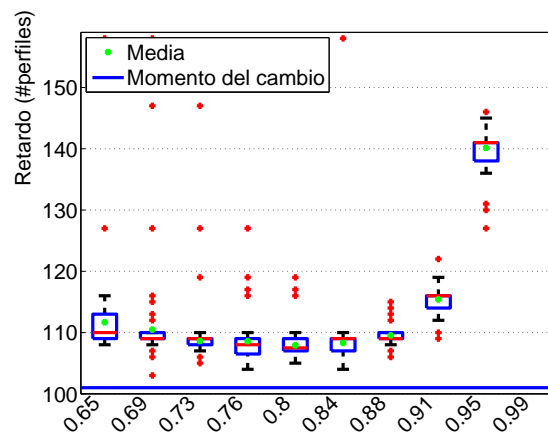


Figura 4.27: Gráfico Box Plot para 100 ejecuciones de Randomized Power Martingale por cada una de las 10 ϵ .

ϵ	Retardo (Nº perfiles)	FA	Sin detectar
0,65	10,67	4	-
0,69	9,47	10	-
0,73	7,61	14	-
0,76	7,60	15	-
0,8	6,92	8	-
0,84	7,29	-	-
0,88	8,54	1	-
0,91	14,40	-	-
0,95	39,14	-	-
0,99	-	-	100

Cuadro 4.10: Media de retardos, FA y cambios no detectados para un conjunto de 100 ejecuciones de la Randomized Power Martingale y varios valores de ϵ .

Simple Mixture

A continuación mostramos resultados para 100 ejecuciones de Simple Mixture.

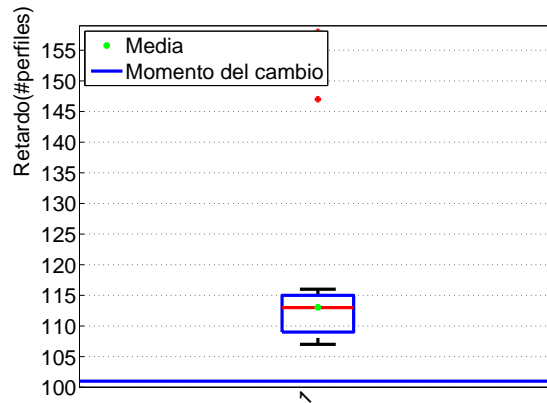


Figura 4.28: Gráfico Box Plot para 100 ejecuciones.

Retardo (Nº perfiles)	FA	Sin detectar
12,05	5	-

Cuadro 4.11: Media de retardos, FA y cambios no detectados para un conjunto de 100 ejecuciones de la Simple Mixture Martingale.

Plug-in Martingale

En este caso veremos los resultados para esta martingala para 100 ejecuciones y 5 valores de FCT distintas. FCT es un factor que multiplica el valor de h (Normal Reference Rule).

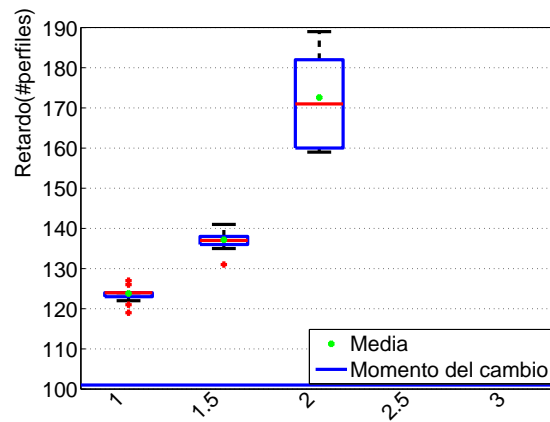


Figura 4.29: Gráfico Box Plot para 100 ejecuciones de Plug-In Martingale y 5 factores (FCT) diferentes de multiplicación de h (Normal Reference Rule).

FCT	Retardo (N° perfiles)	FA	Sin detectar
1	22,79	-	-
1,5	36,16	-	-
2	71,59	-	-
2,5	-	-	100
3	-	-	100

Cuadro 4.12: Media de retardos, FA y cambios no detectados para un conjunto de 100 ejecuciones de la Plug-in Martingale.

Conclusión

Para cambios muy pequeños se sigue detectando el cambio, lo que demuestra la sensibilidad del método independientemente de la martingala que utilicemos.

Se sigue cumpliendo que Randomized Power y Simple Mixture son las martingalas con menor retardo en la detección del cambio y Plug-In la que da menos FA.

Parece por los resultados que una diferencia de cambio menor implica un número mayor de FA y cambios no detectados.

4.3.5. Experimento 5

En este caso se replica el experimento anterior pero los perfiles son generados con una diferencia de altura aún menor *sin ruido*:

- Grupo 1: $\alpha_1 = \alpha_2 = 4$; $\text{fct} = 0,01$; $A_0 = 1, 2$
- Grupo 2: $\alpha_1 = \alpha_2 = 4$; $\text{fct} = 0,01$; $A_0 = 1, 201$

Para este caso, no se detectan cambios.

Ahora vamos a probar para el mismo caso pero sin ruido.

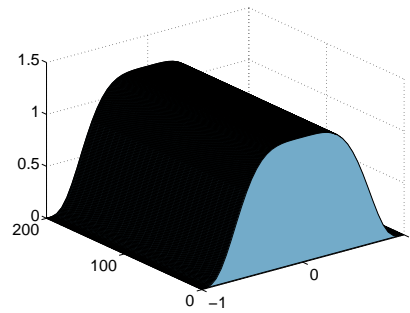


Figura 4.30: Perfiles Experimento 5.

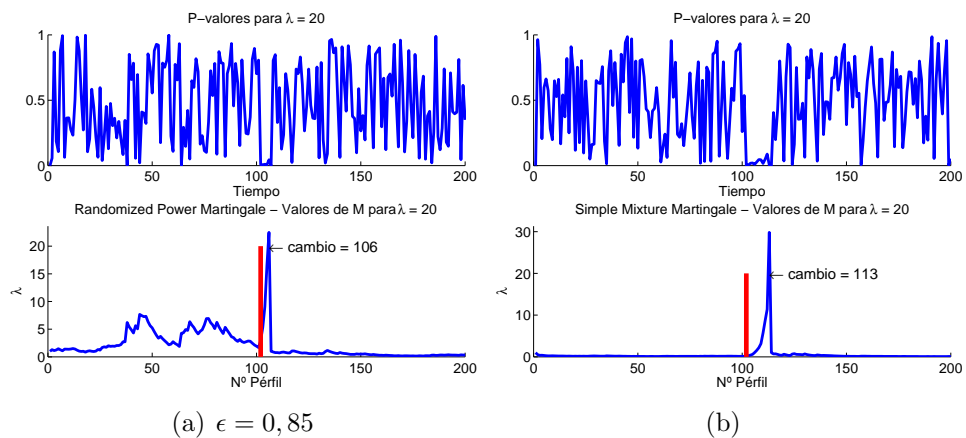
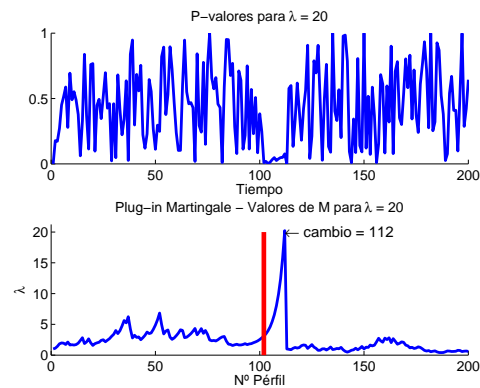


Figura 4.31: Ejemplos de detección del experimento 5. Se indica en cada gráfico el momento en el que la arquitectura de martingalas identifica el cambio.



(a) Factor de multiplicación de h (FCT)
= 1

Figura 4.32: Ejemplos de detección del experimento 5. Se indica en cada gráfico el momento en el que la arquitectura de martingalas identifica el cambio.

En este caso podemos observar en las figuras 4.31 y 4.32 el momento en el que se detecta el cambio con un retardo. La línea roja marca el momento del cambio real. Debido al carácter estocástico de las martingalas, detallamos valores obtenidos después de ejecutar el algoritmo 100 veces para cada una de estas martingalas.

Randomized Power Martingale

Para esta martingala se han obtenido resultados para 10 ϵ distintas. Podemos ver el resultado obtenido en el cuadro 4.13. En la figura 4.33 podemos ver el resultado de detección de cambios, para 100 ejecuciones, mediante un gráfico Box Plot de resultados. Se han eliminado las falsas alarmas (FA) para evitar desvirtuar los resultados.

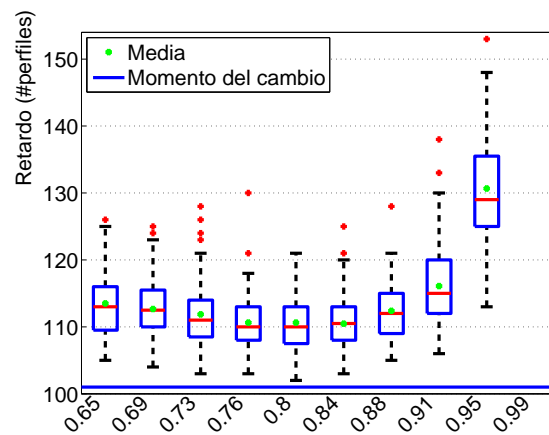


Figura 4.33: Gráfico Box Plot para 100 ejecuciones de Randomized Power Martingale por cada una de las 10 ϵ .

ϵ	Retardo (Nº perfiles)	FA	Sin detectar
0,65	12,05	7	1
0,69	11,67	1	-
0,73	10,87	4	-
0,76	9,72	5	-
0,8	9,63	5	-
0,84	10	1	1
0,88	11,39	1	-
0,91	15,09	-	-
0,95	29,67	0	-
0,99	-	-	100

Cuadro 4.13: Media de retardos, FA y cambios no detectados para un conjunto de 100 ejecuciones de la Randomized Power Martingale y varios valores de ϵ .

Simple Mixture

A continuación mostramos resultados para 100 ejecuciones de Simple Mixture.

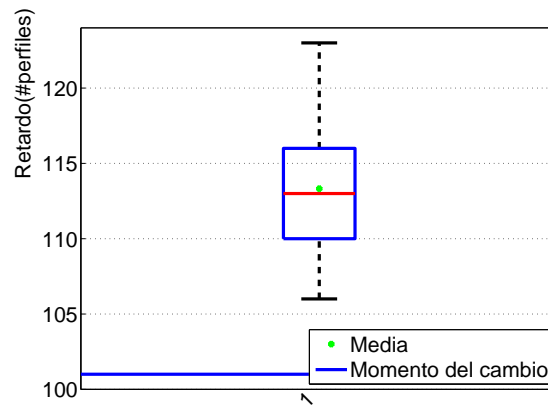


Figura 4.34: Gráfico Box Plot para 100 ejecuciones.

Retardo (N° perfiles)	FA	Sin detectar
12,32	4	-

Cuadro 4.14: Media de retardos, FA y cambios no detectados para un conjunto de 100 ejecuciones de la Simple Mixture Martingale.

Plug-in Martingale

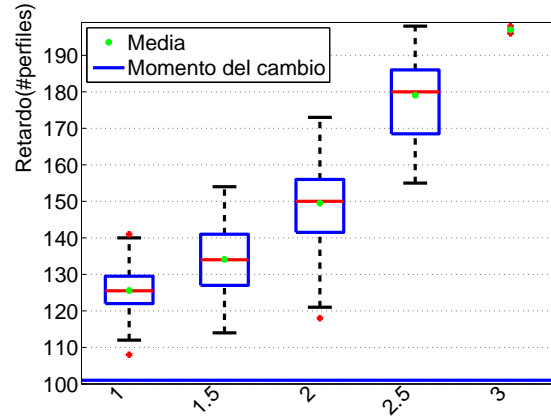


Figura 4.35: Gráfico Box Plot para 100 ejecuciones de Plug-In Martingale y 5 factores (FCT) diferentes de multiplicación de h (Normal Reference Rule).

FCT	Retardo (Nº perfiles)	FA	Sin detectar
1	24,58	-	-
1,5	33,14	-	-
2	48,59	-	-
2,5	78,09	-	6
3	96	-	97

Cuadro 4.15: Media de retardos, FA y cambios no detectados para un conjunto de 100 ejecuciones de la Plug-in Martingale.

Conclusión

Para los casos de perfiles con ausencia de ruido el grado de sensibilidad del método es muy alto.

El nivel de ruido marca el umbral de sensibilidad del método. A mayor ruido menor grado de sensibilidad y viceversa.

En el caso del retardo no parece que haya una relación entre este y el ruido que tengan los perfiles.

4.3.6. Experimento 6

En este caso generamos varios grupos de perfiles con diferente altura. Además los perfiles son generados sin ruido :

- Grupo 1: $\alpha_1 = \alpha_2 = 4$; $A_0 = 1$
- Grupo 2: $\alpha_1 = \alpha_2 = 4$; $A_0 = 3$
- Grupo 3: $\alpha_1 = \alpha_2 = 4$; $A_0 = 7$

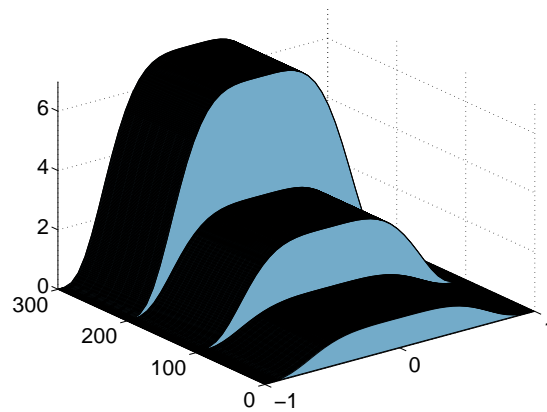


Figura 4.36: Perfiles Experimento 6.

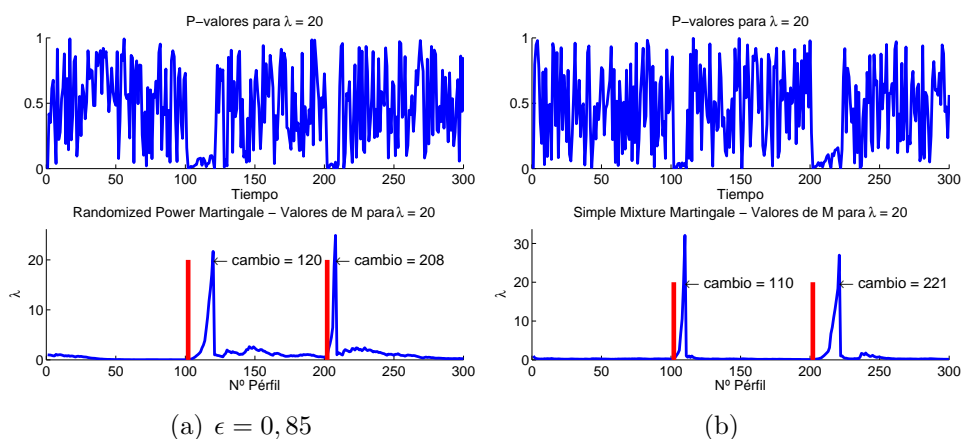
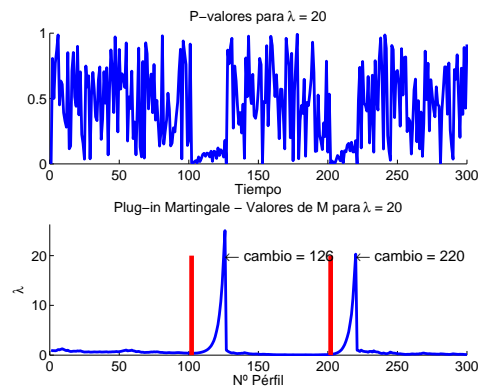


Figura 4.37: Ejemplos de detección del experimento 6. Se indica en cada gráfico el momento en el que la arquitectura de martingalas identifica el cambio.



(a) Factor de multiplicación de h (FCT)
 $= 1$

Figura 4.38: Ejemplos de detección del experimento 6. Se indica en cada gráfico el momento en el que la arquitectura de martingalas identifica el cambio.

En este caso podemos observar en las figuras 4.37 y 4.38 el momento en el que se detecta el cambio con un retardo. La línea roja marca el momento del cambio real. Debido al carácter estocástico de las martingalas, detallamos valores obtenidos después de ejecutar el algoritmo 100 veces para cada una de estas martingalas.

Randomized Power Martingale

Para esta martingala se han obtenido resultados para 10 ϵ distintas. Podemos ver el resultado obtenido en el cuadro 4.16 y 4.17. En la figura 4.39 y 4.40 podemos ver el resultado de detección de cambios, para 100 ejecuciones, mediante un gráfico Box Plot de resultados. Se han eliminado las falsas alarmas (FA) para evitar desvirtuar los resultados. Por claridad, para cada uno de los cambios aparece una tabla y gráfico por separado.

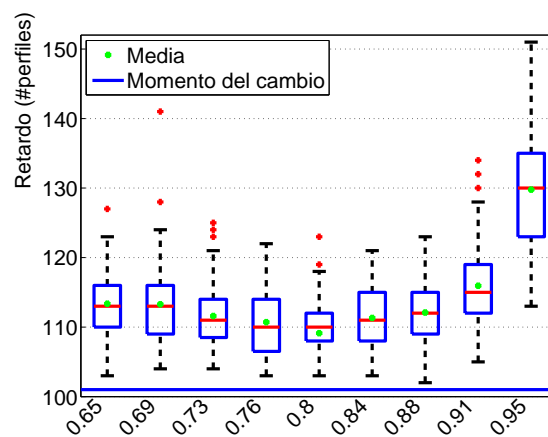


Figura 4.39: Gráfico Box Plot, para el primer cambio, para 100 ejecuciones de Randomized Power Martingale por cada una de las 10 ϵ .

ϵ	Retardo (Nº perfiles)	FA	Sin detectar
0,65	12,36	8	-
0,69	12,28	12	-
0,73	10,60	1	-
0,76	9,71	5	-
0,8	9,23	6	-
0,84	10,31	1	-
0,88	11,09	-	-
0,91	14,94	-	-
0,95	28,78	-	-
0,99	-	-	100

Cuadro 4.16: Media de retardos para el primer cambio, FA y cambios no detectados para un conjunto de 100 ejecuciones de la Randomized Power Martingale y varios valores de ϵ .

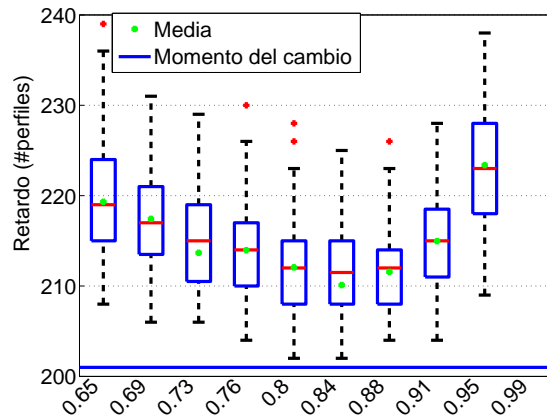


Figura 4.40: Gráfico Box Plot, para el segundo cambio, para 100 ejecuciones de Randomized Power Martingale por cada una de las 10 ϵ .

ϵ	Retardo (Nº perfiles)	FA	Sin detectar
0,65	18,92	5	-
0,69	16,43	8	-
0,73	12,67	7	-
0,76	12,97	5	-
0,8	11,07	-	-
0,84	10,95	3	-
0,88	10,56	-	-
0,91	13,97	-	-
0,95	22,38	-	-
0,99	-	-	100

Cuadro 4.17: Media de retardos para el segundo cambio, FA y cambios no detectados para un conjunto de 100 ejecuciones de la Randomized Power Martingale y varios valores de ϵ .

Simple Mixture

A continuación mostramos resultados para 100 ejecuciones de Simple Mixture.

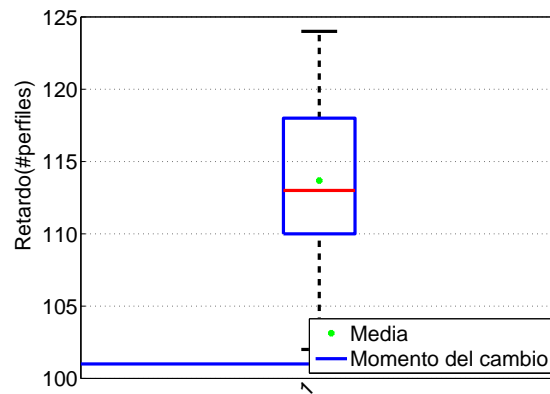


Figura 4.41: Gráfico Box Plot del primer cambio para 100 ejecuciones.

Retardo (N° perfiles)	FA	Sin detectar
12,68	4	-

Cuadro 4.18: Media de retardos, FA y cambios no detectados, para el primer cambio, para un conjunto de 100 ejecuciones de la Simple Mixture Martingale.

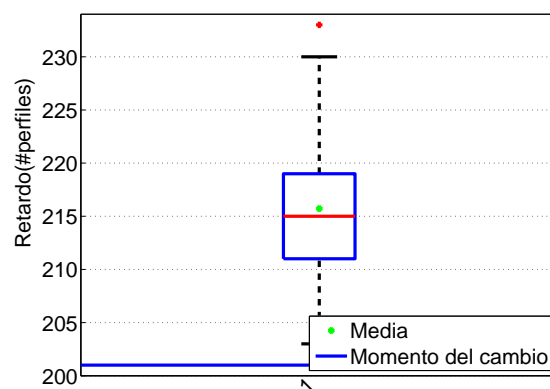


Figura 4.42: Gráfico Box Plot del segundo cambio para 100 ejecuciones.

Retardo (N° perfiles)	FA	Sin detectar
14,72	2	-

Cuadro 4.19: Media de retardos, FA y cambios no detectados, para el segundo cambio, para un conjunto de 100 ejecuciones de la Simple Mixture Martingale.

Plug-in Martingale

En este caso veremos los resultados para esta martingala para 100 ejecuciones y 5 valores de FCT distintas. FCT es un factor que multiplica el valor de h (Normal Reference Rule).

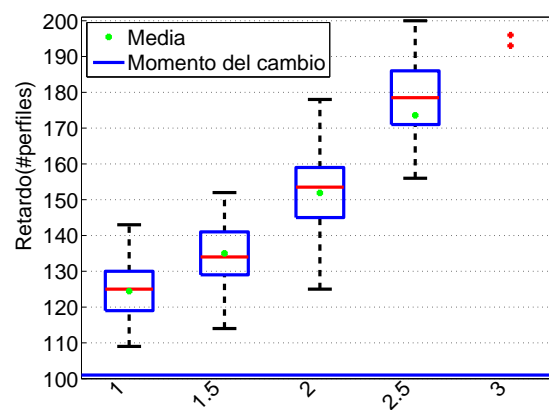


Figura 4.43: Gráfico Box Plot del primer cambio para 100 ejecuciones de Plug-In Martingale y 5 factores (FCT) diferentes de multiplicación de h (Normal Reference Rule).

FCT	Retardo (N° perfiles)	FA	Sin detectar
1	23,45	2	-
1,5	34,01	-	-
2	50,89	-	3
2,5	72,57	-	6
3	-	2	98

Cuadro 4.20: Media de retardos, FA y cambios no detectados, para el primer cambio, para un conjunto de 100 ejecuciones de la Plug-in Martingale.

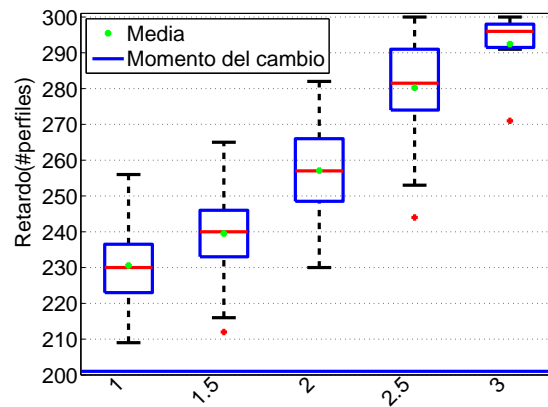


Figura 4.44: Gráfico Box Plot del segundo cambio para 100 ejecuciones de Plug-In Martingale y 5 factores (FCT) diferentes de multiplicación de h (Normal Reference Rule).

FCT	Retardo (N° perfiles)	FA	Sin detectar
1	29,75	1-	-
1,5	38,51	-	-
2	56,09	-	-
2,5	28,78	-	18
3	177,73	-	93

Cuadro 4.21: Media de retardos, FA y cambios no detectados, para el segundo cambio, para un conjunto de 100 ejecuciones de la Plug-in Martingale.

Conclusión

El método es igual de efectivo cuando hay varios cambios en el flujo de datos.

4.3.7. Experimento 7

En este caso generamos varios grupos de perfiles con diferente altura. Además los perfiles son generados sin ruido :

- Grupo 1: $\alpha_1 = \alpha_2 = 4$; $A_0 = 7$
- Grupo 2: $\alpha_1 = \alpha_2 = 4$; $A_0 = 3$
- Grupo 3: $\alpha_1 = \alpha_2 = 4$; $A_0 = 7$

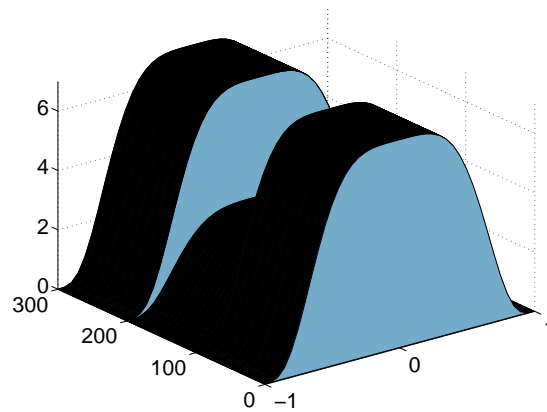


Figura 4.45: Perfiles Experimento 7.

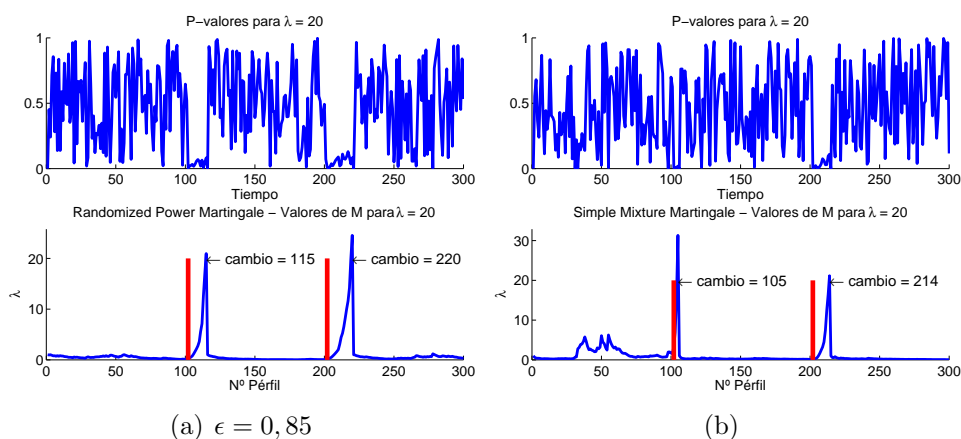
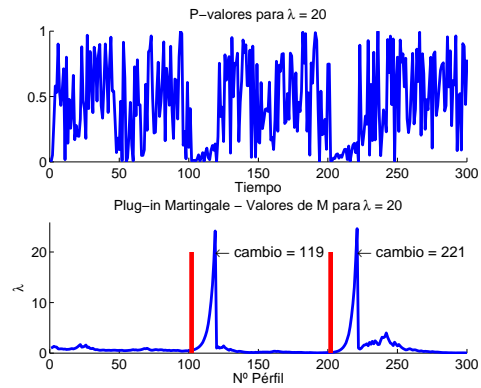


Figura 4.46: Ejemplos de detección del experimento 7. Se indica en cada gráfico el momento en el que la arquitectura de martingalas identifica el cambio.



(a) Factor de multiplicación de h (FCT)
 $= 1$

Figura 4.47: Ejemplos de detección del experimento 7. Se indica en cada gráfico el momento en el que la arquitectura de martingalas identifica el cambio.

En este caso podemos observar en la figura 4.46 y 4.47 el momento en el que se produce el cambio, que se detecta con un retardo. La línea roja marca el momento del cambio real. Se detallan valores obtenidos después de ejecutar el algoritmo 100 veces para cada una de estas martingalas.

Randomized Power Martingale

Para esta martingala se han obtenido resultados para 10 ϵ distintas. Podemos ver el resultado obtenido en el cuadro 4.22 y 4.23. En la figura 4.48 y 4.49 podemos ver el número de perfiles que se deben leer desde que el cambio se produce hasta que se detecta (retardo). Para cada uno de los cambios aparece un gráfico y una tabla de resultados por separado.

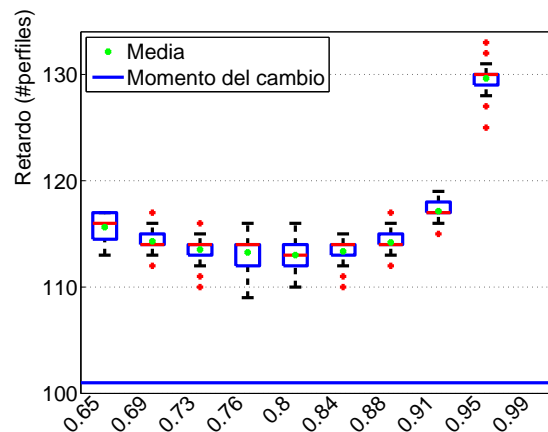


Figura 4.48: Gráfico Box Plot del primer cambio para 100 ejecuciones de Randomized Power Martingale por cada una de las 10 ϵ .

ϵ	Retardo (Nº perfiles)	FA	Sin detectar
0,65	12,04	2	-
0,69	11,45	5	-
0,73	10,43	5	-
0,76	9,14	4	-
0,8	9,75	1	-
0,84	9,85	2	-
0,88	11,32	-	-
0,91	14,77	-	-
0,95	30,38	-	-
0,99	-	-	100

Cuadro 4.22: Media de retardos para el primer cambio, FA y cambios no detectados, para el primer cambio, para un conjunto de 100 ejecuciones de la Randomized Power Martingale y varios valores de ϵ .

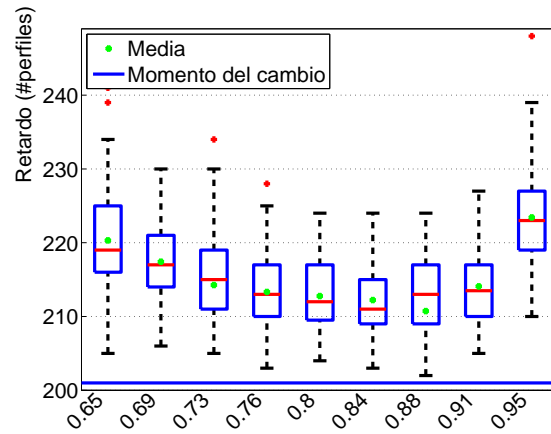


Figura 4.49: Gráfico Box Plot del segundo cambio para 100 ejecuciones de Randomized Power Martingale por cada una de las 10 ϵ .

ϵ	Retardo (Nº perfiles)	FA	Sin detectar
0,65	19,32	9	-
0,69	16,42	11	-
0,73	13,27	8	-
0,76	12,32	9	-
0,8	11,77	3	-
0,84	11,23	2	-
0,88	11,87	3	1
0,91	13,09	-	-
0,95	22,43	-	1
0,99	-	-	100

Cuadro 4.23: Media de retardos para el primer cambio, FA y cambios no detectados, para el segundo cambio, para un conjunto de 100 ejecuciones de la Randomized Power Martingale y varios valores de ϵ .

Simple Mixture

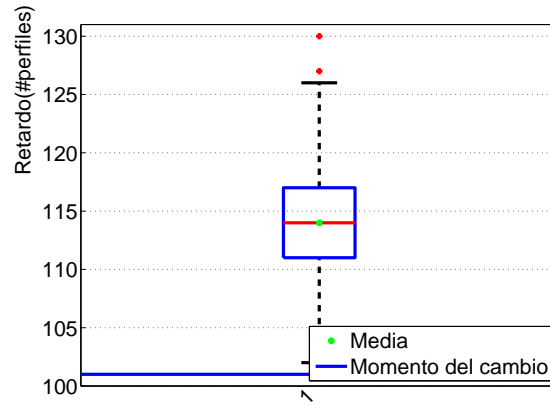


Figura 4.50: Gráfico Box Plot del primer cambio para 100 ejecuciones.

Retardo (Nº perfiles)	FA	Sin detectar
12,99	4	-

Cuadro 4.24: Media de retardos, FA y cambios no detectados, para el primer cambio, para un conjunto de 100 ejecuciones de la Simple Mixture Martingale.

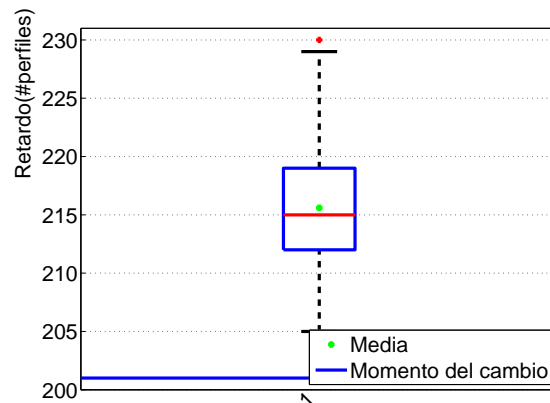


Figura 4.51: Gráfico Box Plot del segundo cambio para 100 ejecuciones.

Retardo (N° perfiles)	FA	Sin detectar
14,6	2	-

Cuadro 4.25: Media de retardos, FA y cambios no detectados, para el segundo cambio, para un conjunto de 100 ejecuciones de la Simple Mixture Martingale.

Plug-in Martingale

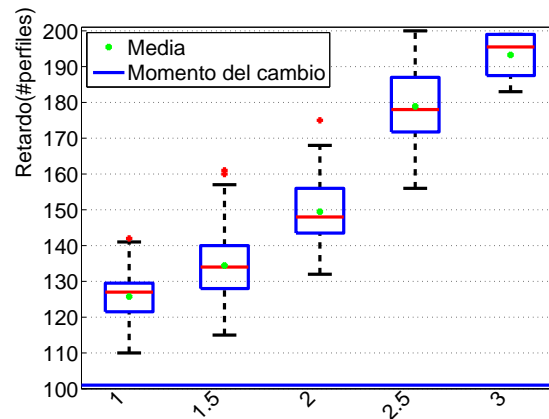


Figura 4.52: Gráfico Box Plot del primer cambio, para el primer cambio, para 100 ejecuciones de Plug-In Martingale y 5 factores (FCT) diferentes de multiplicación de h (Normal Reference Rule).

Los valores de $FCT = 3$, corresponden a dos valores que se consideran FA por la distancia excesiva respecto al cambio inicial (detección en lectura n° 198).

FCT	Retardo (N° perfiles)	FA	Sin detectar
1	24,76	3	-
1,5	33,43	-	-
2	48,45	-	-
2,5	77,89	-	-
3	92,25	-	-

Cuadro 4.26: Media de retardos, FA y cambios no detectados, para el primer cambio, para un conjunto de 100 ejecuciones de la Plug-in Martingale.

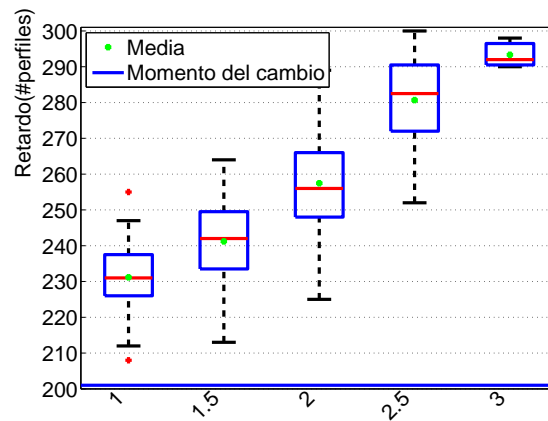


Figura 4.53: Gráfico Box Plot del segundo cambio para 100 ejecuciones de Plug-In Martingale y 5 factores (FCT) diferentes de multiplicación de h (Normal Reference Rule).

FCT	Retardo (N° perfiles)	FA	Sin detectar
1	30,16	1	-
1,5	40,20	1	-
2	56,45	-	-
2,5	79,69	-	18
3	92,33	-	-

Cuadro 4.27: Media de retardos, FA y cambios no detectados, para el segundo cambio, para un conjunto de 100 ejecuciones de la Plug-in Martingale.

Conclusión

Comparando los resultados con los obtenidos en la sección 4.3.6 podemos concluir que el método es eficiente independientemente de si el cambio de los perfiles es del tipo bajo-alto o alto-bajo o el número de ellos, no apreciándose ninguna característica diferencial en los resultados de ambos tipos de cambios.

4.3.8. Experimento 8

En este experimento vamos a probar resultados para un caso en el que los perfiles no tiene ruido y otro en el que si tendrá un ruido más significativo que lo visto hasta ahora. Para el caso del perfil sin ruido tenemos.

- Grupo 1: $\alpha_1 = \alpha_2 = 4$; $A_0 = 10$
- Grupo 2: $\alpha_1 = \alpha_2 = 4$; $A_0 = 5$

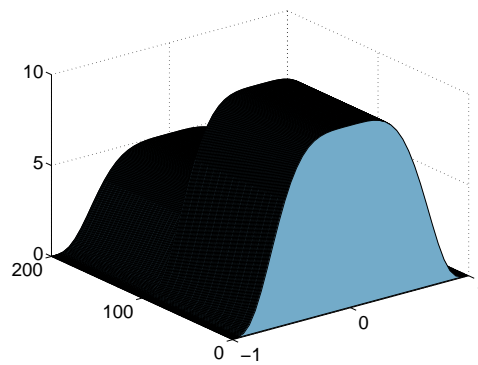


Figura 4.54: Perfiles Experimento 7.

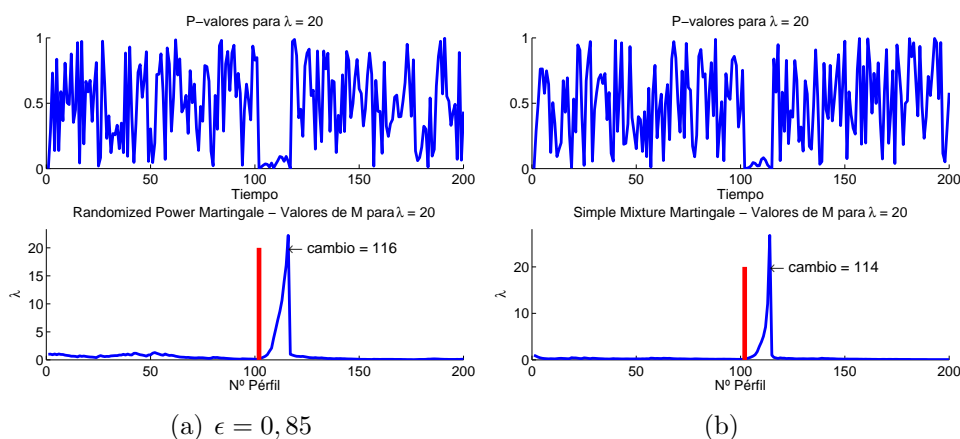
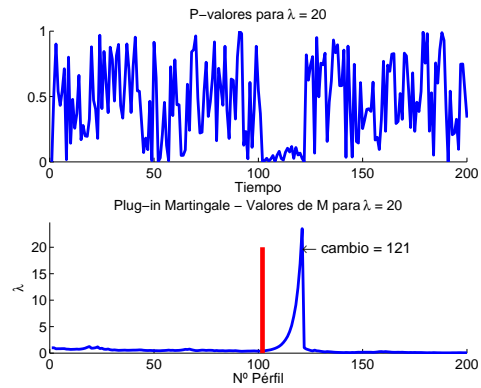


Figura 4.55: Ejemplos de detección del experimento 7. Se indica en cada gráfico el momento en el que la arquitectura de martingalas identifica el cambio.



(a) Factor de multiplicación de h (FCT) = 1

Figura 4.56: Ejemplos de detección del experimento 7. Se indica en cada gráfico el momento en el que la arquitectura de martingalas identifica el cambio.

A continuación se muestran los perfiles con los que se va a comparar los resultados de los perfiles sin ruido. Como se puede apreciar los perfiles son iguales pero con ruido.

- Grupo 1: $\alpha_1 = \alpha_2 = 4$; $fct = 0,1$; $A_0 = 10$
- Grupo 2: $\alpha_1 = \alpha_2 = 4$; $fct = 0,1$; $A_0 = 5$

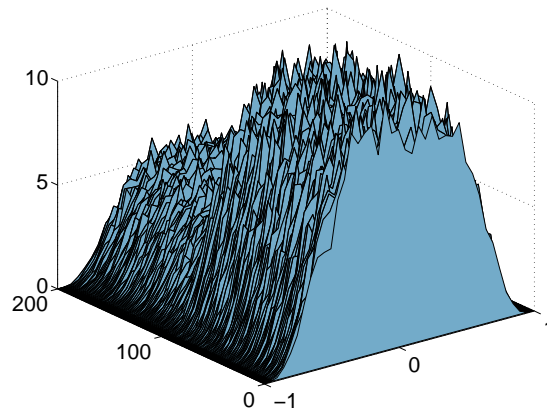


Figura 4.57: Perfiles Experimento 7.

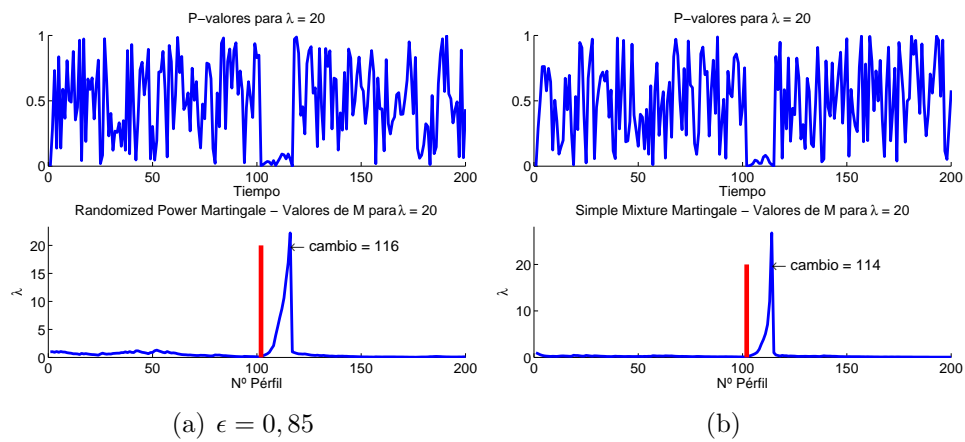


Figura 4.58: Ejemplos de detección del experimento 7. Se indica en cada gráfico el momento en el que la arquitectura de martingalas identifica el cambio.

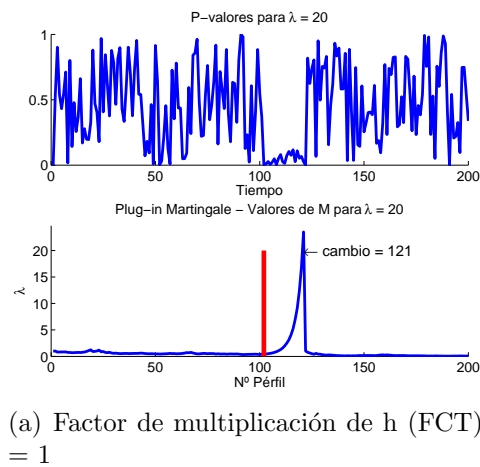


Figura 4.59: Ejemplos de detección del experimento 7. Se indica en cada gráfico el momento en el que la arquitectura de martingalas identifica el cambio.

Para los dos tipos de perfiles, con ruido y sin él, se han generado 100 ejecuciones de cada una de las martingalas.

Randomized Power Martingale

A continuación se muestra el resultado sendas rondas de 100 ejecuciones para perfiles sin ruido y con ruido

RESULTADOS SIN RUIDO

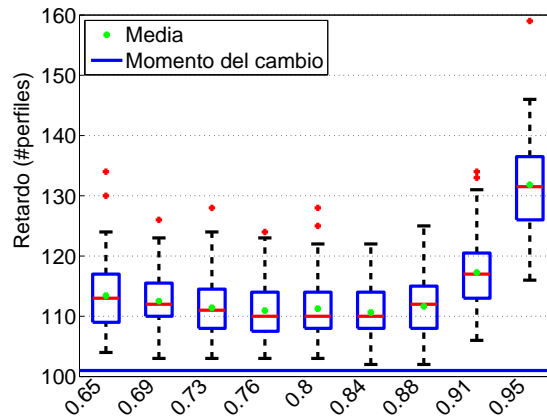


Figura 4.60: Gráfico Box Plot para 100 ejecuciones de Ramdomized Power Martingale por cada una de las 10 ϵ y perfiles sin ruido.

ϵ	Retardo (Nº perfiles)	FA	Sin detectar
0,65	12,43	8	-
0,69	11,51	3	-
0,73	10,41	7	-
0,76	9,94	7	-
0,8	10,25	4	-
0,84	9,63	2	-
0,88	10,65	1	-
0,91	16,26	-	-
0,95	30,81	-	-
0,99	-	-	100

Cuadro 4.28: Media de retardos para perfiles sin ruido, FA y cambios no detectados para un conjunto de 100 ejecuciones de la Randomized Power Martingale y varios valores de ϵ .

RESULTADOS CON RUIDO

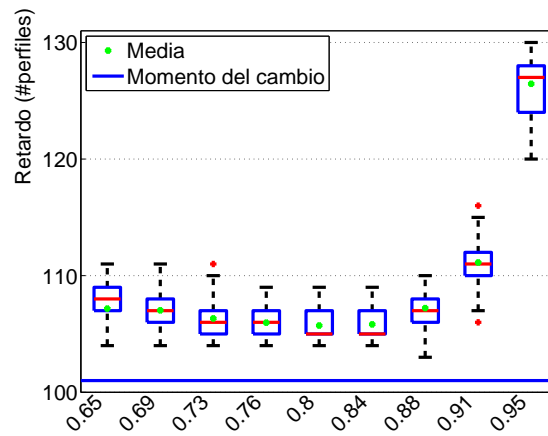


Figura 4.61: Gráfico Box Plot para 100 ejecuciones de Randomized Power Martingale por cada una de las 10 ϵ para perfiles con ruido.

ϵ	Retardo (Nº perfiles)	FA	Sin detectar
0,65	6,17	7	1
0,69	6,03	9	-
0,73	5,33	19	-
0,76	4,97	11	-
0,8	4,72	12	-
0,84	4,82	1	-
0,88	6,22	-	-
0,91	10,13	-	-
0,95	25,45	-	-
0,99	-	-	100

Cuadro 4.29: Media de retardos para perfiles con ruido, FA y cambios no detectados para un conjunto de 100 ejecuciones de la Randomized Power Martingale y varios valores de ϵ para perfiles con ruido.

Simple Mixture

RESULTADOS SIN RUIDO

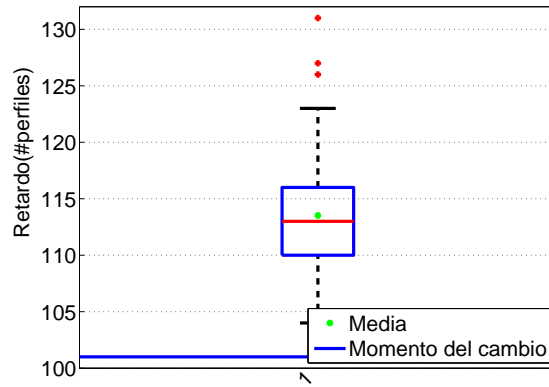


Figura 4.62: Gráfico Box Plot para 100 ejecuciones.

Retardo (N° perfiles)	FA	Sin detectar
12,52	2	-

Cuadro 4.30: Media de retardos para perfiles sin ruido, FA y cambios no detectados para un conjunto de 100 ejecuciones de la Simple Mixture Martingale.

RESULTADOS CON RUIDO

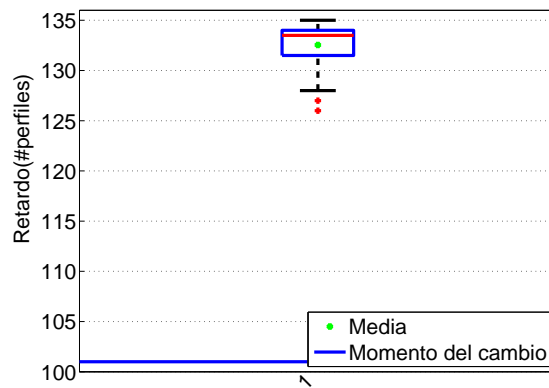


Figura 4.63: Gráfico Box Plot para 100 ejecuciones.

Retardo (N° perfiles)	FA	Sin detectar
12,36	-	-

Cuadro 4.31: Media de retardos con ruido, FA y cambios no detectados para un conjunto de 100 ejecuciones de la Simple Mixture Martingale.

Plug-in Martingale

RESULTADOS SIN RUIDO

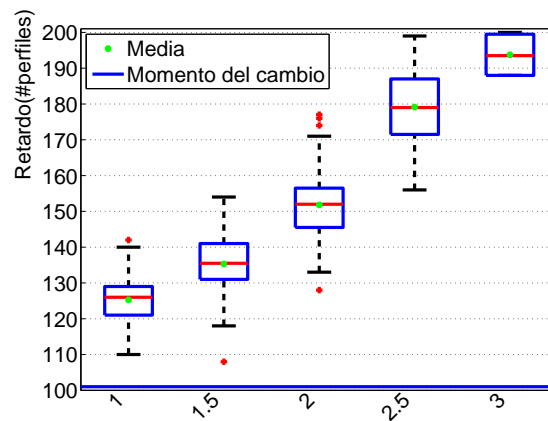


Figura 4.64: Gráfico Box Plot para 100 ejecuciones de Plug-In Martingale y 5 factores (FCT) diferentes de multiplicación de h (Normal Reference Rule).

FCT	Retardo (N° perfiles)	FA	Sin detectar
1	24,29	-	-
1,5	34,30	-	-
2	50,78	-	-
2,5	78,15	-	-
3	92,75	-	-

Cuadro 4.32: Media de retardos para perfiles sin ruido, FA y cambios no detectados para un conjunto de 100 ejecuciones de la Plug-in Martingale.

RESULTADOS CON RUIDO

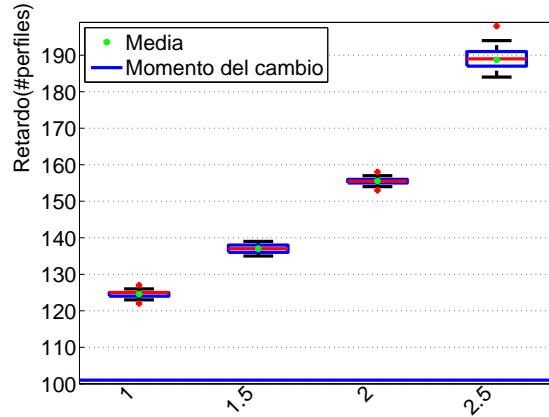


Figura 4.65: Gráfico Box Plot para 100 ejecuciones de Plug-In Martingale y 5 factores (FCT) diferentes de multiplicación de h (Normal Reference Rule).

FCT	Retardo (Nº perfiles)	FA	Sin detectar
1	23,56	1	-
1,5	36,01	-	-
2	54,55	-	-
2,5	87,70	-	-
3	-	-	100

Cuadro 4.33: Media de retardos, FA y cambios no detectados para un conjunto de 100 ejecuciones de la Plug-in Martingale.

Conclusión

Como hemos visto con anterioridad no parece que el ruido afecte de forma significativa al retardo, ni al número de FA ni a las no detecciones de cambios.

Plug-In sigue siendo la que mayor retardo tiene y la que obtiene menor número de FA.

4.3.9. Experimento 9

Probaremos a continuación el efecto que se produce mediante el suavizado de las señales que tengan ruido.

- Grupo 1: $\alpha_1 = \alpha_2 = 4$; $fct = 0.1$; $A_0 = 10$
- Grupo 2: $\alpha_1 = \alpha_2 = 4$; $fct = 0.1$; $A_0 = 5$

Como factor de suavizado utilizamos, tal y como explicamos al principio del capítulo, splines cúbicos.

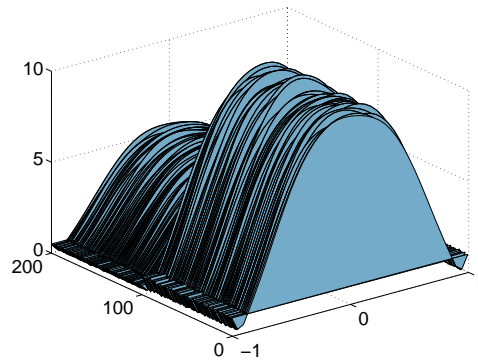


Figura 4.66: Perfiles Experimento 9.

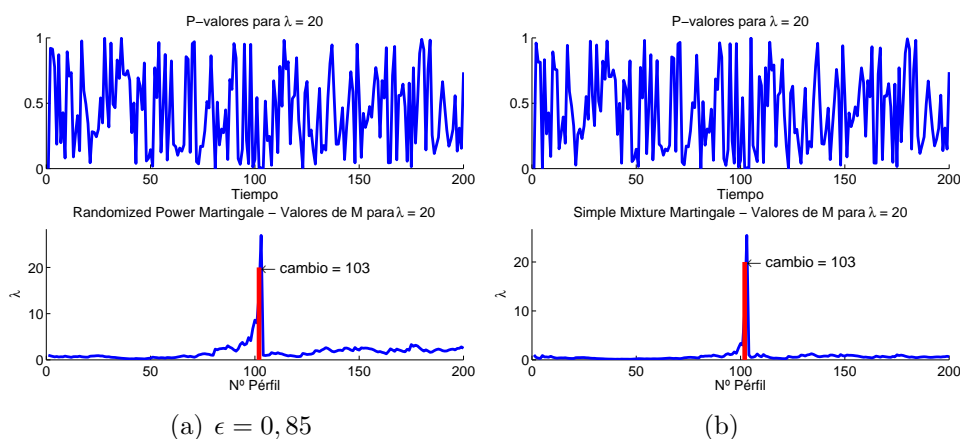
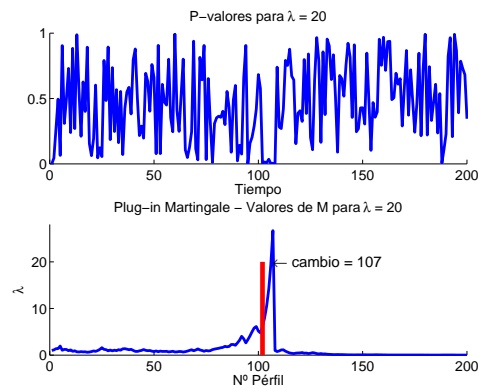


Figura 4.67: Ejemplos de detección del experimento 9. Se indica en cada gráfico el momento en el que la arquitectura de martingalas identifica el cambio.



(a) Factor de multiplicación de h (FCT)
 $= 1$

Figura 4.68: Ejemplos de detección del experimento 9. Se indica en cada gráfico el momento en el que la arquitectura de martingalas identifica el cambio.

En este caso podemos observar en las figuras 4.67 y 4.68 el momento en el que se detecta el cambio con un retardo. La línea roja marca el momento del cambio real. Debido al carácter estocástico de las martingalas, detallamos valores obtenidos después de ejecutar el algoritmo 100 veces para cada una de estas martingalas.

Randomized Power Martingale

Para esta martingala se han obtenido resultados para 10 ϵ distintas. Podemos ver el resultado obtenido en el cuadro 4.34. En la figura 4.69 podemos ver el resultado de detección de cambios, para 100 ejecuciones, mediante un gráfico Box Plot de resultados. Se han eliminado las falsas alarmas (FA) para evitar desvirtuar los resultados.

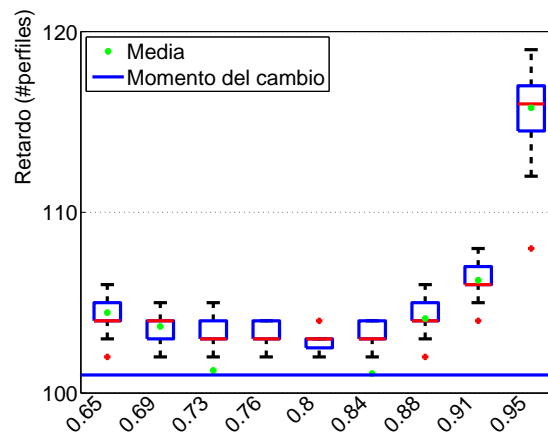


Figura 4.69: Gráfico Box Plot para 100 ejecuciones de Randomized Power Martingale por cada una de las 10 ϵ y perfiles suavizados.

ϵ	Retardo (Nº perfiles)	FAR	Sin detectar
0,65	3,45	5	-
0,69	2,69	8	-
0,73	2,31	2	2
0,76	2,04	4	8
0,8	1,96	8	2
0,84	2,14	2	-
0,88	3,12	-	-
0,91	5,24	-	-
0,95	14,78	-	-
0,99	-	-	100

Cuadro 4.34: Media de retardos para perfiles suavizados, FAR's y cambios no detectados para un conjunto de 100 ejecuciones de la Randomized Power Martingale y varios valores de ϵ .

Simple Mixture

A continuación mostramos resultados para 100 ejecuciones de Simple Mixture.

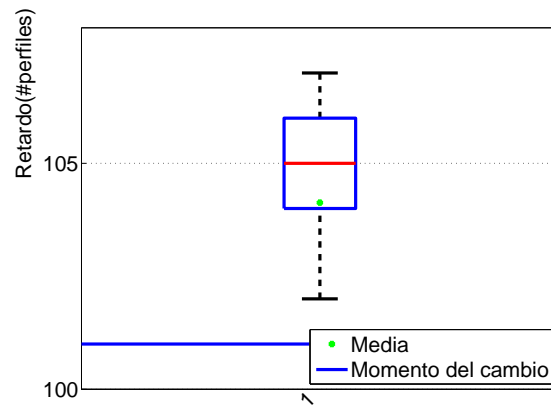


Figura 4.70: Gráfico Box Plot para 100 ejecuciones.

Retardo (N° perfiles)	FAR	Sin detectar
3,13	5	-

Cuadro 4.35: Media de retardos para perfiles suavizados, FAR's y cambios no detectados para un conjunto de 100 ejecuciones de la Simple Mixture Martingale.

Plug-in Martingale

En este caso veremos los resultados para esta martingala para 100 ejecuciones y 5 valores de FCT distintas. FCT es un factor que multiplica el valor de h (Normal Reference Rule).

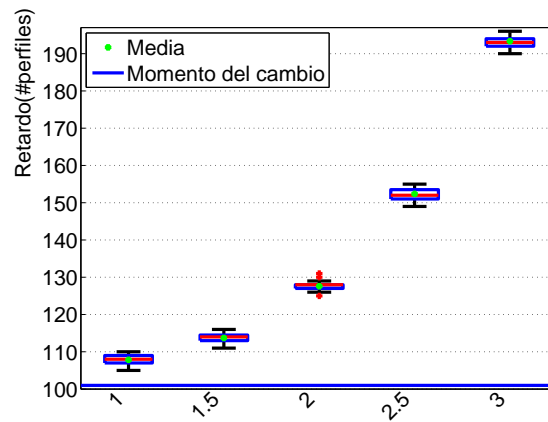


Figura 4.71: Gráfico Box Plot para 100 ejecuciones de Plug-In Martingale y 5 factores (FCT) diferentes de multiplicación de h (Normal Reference Rule).

FCT	Retardo (Nº perfiles)	FAR	Sin detectar
1	6,84	-	-
1,5	12,70	-	-
2	26,59	-	-
2,5	51,36	-	-
3	92,37	-	-

Cuadro 4.36: Media de retardos para perfiles suavizados, FAR's y cambios no detectados para un conjunto de 100 ejecuciones de la Plug-in Martingale.

Conclusión

El suavizado de señales produce una mejora en los resultados.

4.3.10. Experimento 10

Probaremos a continuación el efecto de señales con ruido independiente:

- Grupo 1: $\alpha_1 = \alpha_2 = 4$; $\text{fct} = 0.1$; $A_0 = 10$
- Grupo 2: $\alpha_1 = \alpha_2 = 4$; $\text{fct} = 0.1$; $A_0 = 5$

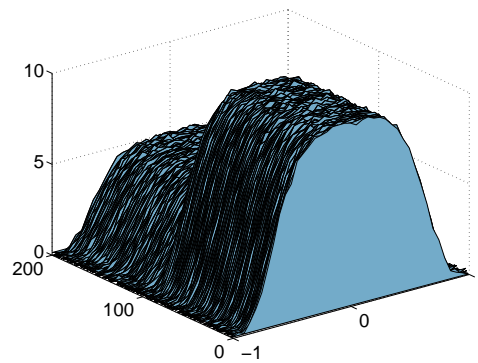


Figura 4.72: Perfiles Experimento 10.

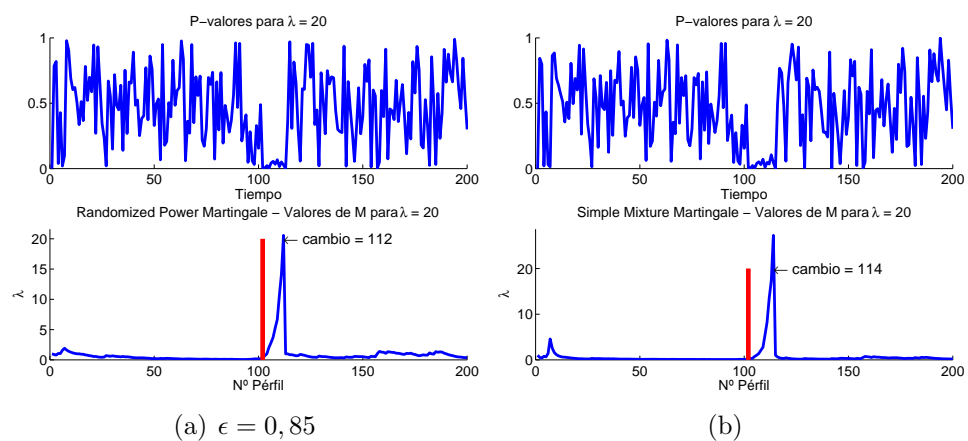
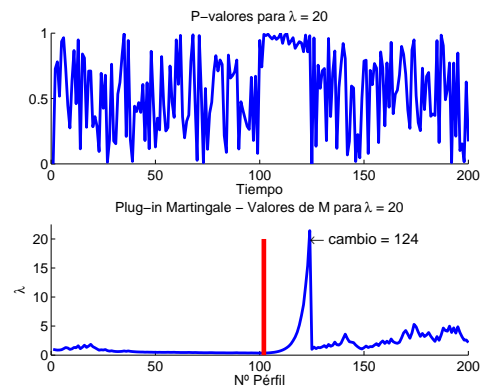


Figura 4.73: Ejemplos de detección del experimento 10. Se indica en cada gráfico el momento en el que la arquitectura de martingalas identifica el cambio.



(a) Factor de multiplicación de h (FCT)
 $= 1$

Figura 4.74: Ejemplos de detección del experimento 10. Se indica en cada gráfico el momento en el que la arquitectura de martingalas identifica el cambio.

En este caso podemos observar en las figuras 4.73 y 4.74 el momento en el que se detecta el cambio con un retardo. La línea roja marca el momento del cambio real. Debido al carácter estocástico de las martingalas, detallamos valores obtenidos después de ejecutar el algoritmo 100 veces para cada una de estas martingalas.

Randomized Power Martingale

Para esta martingala se han obtenido resultados para 10 ϵ distintas. Podemos ver el resultado obtenido en el cuadro 4.37. En la figura 4.75 podemos ver el resultado de detección de cambios, para 100 ejecuciones, mediante un gráfico Box Plot de resultados. Se han eliminado las falsas alarmas (FA) para evitar desvirtuar los resultados.

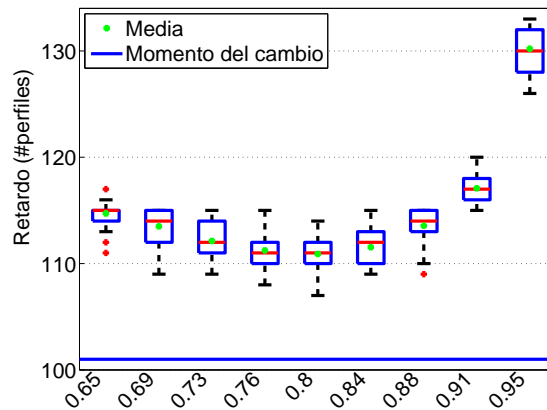


Figura 4.75: Gráfico Box Plot para 100 ejecuciones de Randomized Power Martingale por cada una de las 10 ϵ y perfiles sin ruido.

ϵ	Retardo (Nº perfiles)	FA	Sin detectar
0,65	13,72	5	-
0,69	12,50	1	-
0,73	11,12	2	-
0,76	10,24	1	-
0,8	9,90	8	2
0,84	10,54	2	-
0,88	12,55	-	-
0,91	16,08	-	-
0,95	29,21	-	-
0,99	-	-	100

Cuadro 4.37: Media de retardos para perfiles sin ruido, FA y cambios no detectados para un conjunto de 100 ejecuciones de la Randomized Power Martingale y varios valores de ϵ .

Simple Mixture

A continuación mostramos resultados para 100 ejecuciones de Simple Mixture.

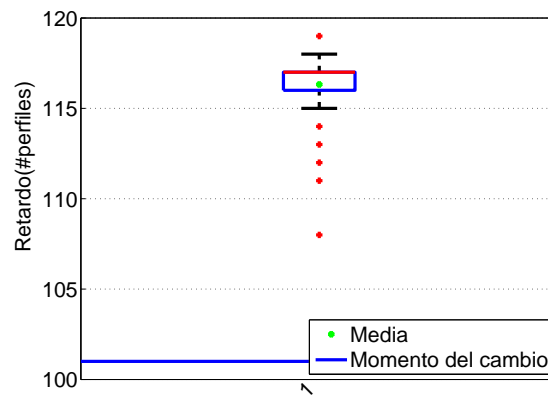


Figura 4.76: Gráfico Box Plot para 100 ejecuciones.

Retardo (Nº perfiles)	FA	Sin detectar
15,32	1	-

Cuadro 4.38: Media de retardos para perfiles sin ruido, FA y cambios no detectados para un conjunto de 100 ejecuciones de la Simple Mixture Martingale.

Plug-in Martingale

En este caso veremos los resultados para esta martingala para 100 ejecuciones y 5 valores de FCT distintas. FCT es un factor que multiplica el valor de h (Normal Reference Rule).

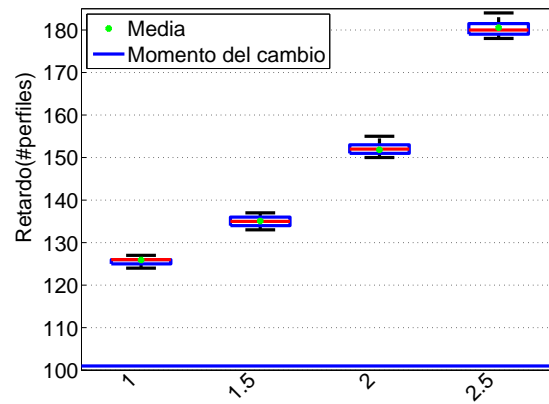


Figura 4.77: Gráfico Box Plot para 100 ejecuciones de Plug-In Martingale y 5 factores (FCT) diferentes de multiplicación de h (Normal Reference Rule).

FCT	Retardo (N° perfiles)	FAR	Sin detectar
1	24,91	-	-
1,5	34,07	-	-
2	50,83	-	-
2,5	79,54	-	-
3	-	-	-

Cuadro 4.39: Media de retardos para perfiles sin ruido, FAR's y cambios no detectados para un conjunto de 100 ejecuciones de la Plug-in Martingale.

Conclusión

No hay diferencias apreciables si el método se ejecuta sobre perfiles con ruido independiente o independiente.

4.3.11. Experimento 11

Probaremos a continuación el efecto de señales con diferentes anchuras:

- Grupo 1: $\alpha_1 = \alpha_2 = 5$; $fct = 0.01$; $A_0 = 10$
- Grupo 2: $\alpha_1 = \alpha_2 = 4$; $fct = 0.01$; $A_0 = 10$

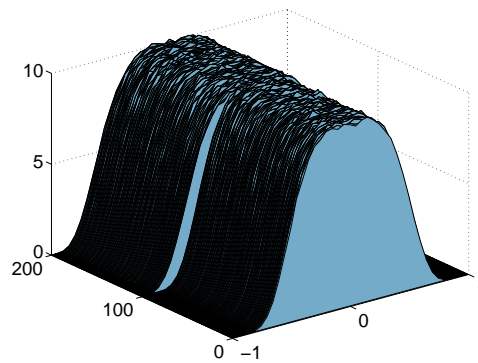


Figura 4.78: Perfiles Experimento 12.

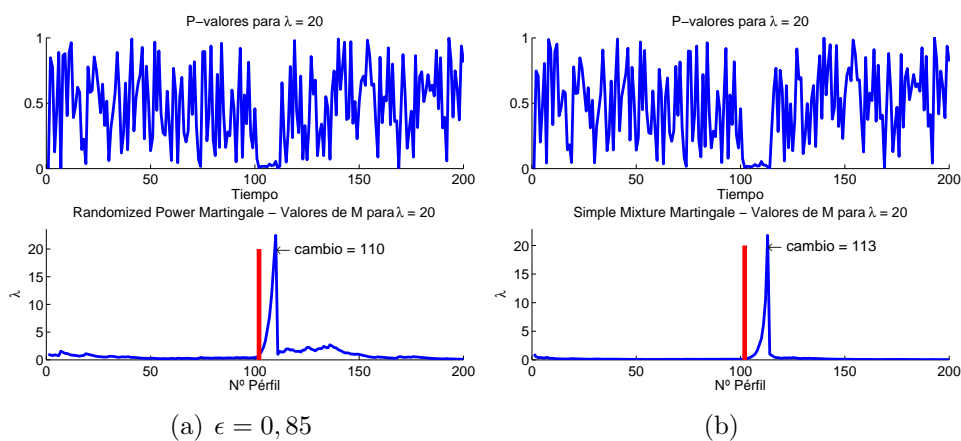
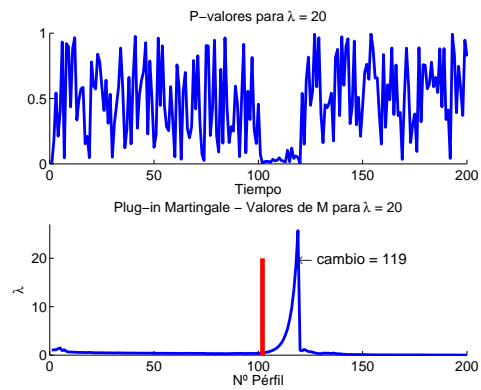


Figura 4.79: Ejemplos de detección del experimento 11. Se indica en cada gráfico el momento en el que la arquitectura de martingalas identifica el cambio.



(a) Factor de multiplicación de h (FCT)
 $= 1$

Figura 4.80: Ejemplos de detección del experimento 11. Se indica en cada gráfico el momento en el que la arquitectura de martingalas identifica el cambio.

En este caso podemos observar en las figuras 4.79 y 4.80 el momento en el que se detecta el cambio con un retardo. La línea roja marca el momento del cambio real. Debido al carácter estocástico de las martingalas, detallamos valores obtenidos después de ejecutar el algoritmo 100 veces para cada una de estas martingalas.

Randomized Power Martingale

Para esta martingala se han obtenido resultados para 10 ϵ distintas. Podemos ver el resultado obtenido en el cuadro 4.40. En la figura 4.81 podemos ver el resultado de detección de cambios, para 100 ejecuciones, mediante un gráfico Box Plot de resultados. Se han eliminado las falsas alarmas (FA) para evitar desvirtuar los resultados.

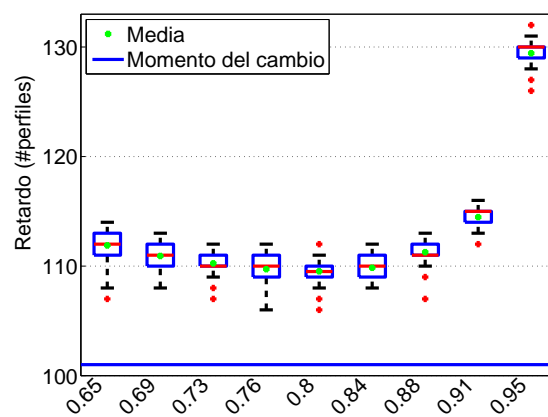


Figura 4.81: Gráfico Box Plot para 100 ejecuciones de Randomized Power Martingales por cada una de las 10 ϵ .

ϵ	Retardo (Nº perfiles)	FA	Sin detectar
0,65	10,89	2	-
0,69	9,92	-	-
0,73	9,25	-	-
0,76	8,72	-	-
0,8	8,53	-	-
0,84	8,84	-	-
0,88	10,27	-	-
0,91	13,47	-	-
0,95	28,43	-	-
0,99	-	-	100

Cuadro 4.40: Media de retardos, FA y cambios no detectados para un conjunto de 100 ejecuciones de la Randomized Power Martingale y varios valores de ϵ .

Simple Mixture

A continuación mostramos resultados para 100 ejecuciones de Simple Mixture.

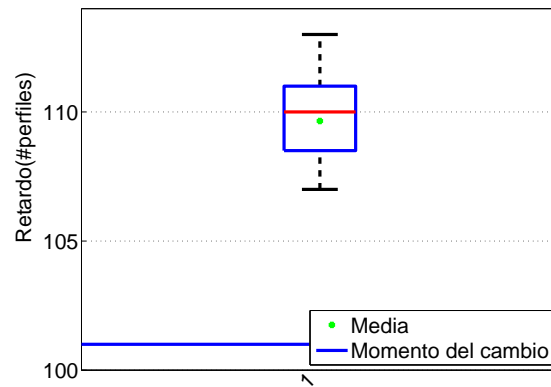


Figura 4.82: Gráfico Box Plot para 100 ejecuciones.

Retardo (Nº perfiles)	FA	Sin detectar
8,65	1	-

Cuadro 4.41: Media de retardos, FA y cambios no detectados para un conjunto de 100 ejecuciones de la Simple Mixture Martingale.

Plug-in Martingale

En este caso veremos los resultados para esta martingala para 100 ejecuciones y 5 valores de FCT distintas. FCT es un factor que multiplica el valor de h (Normal Reference Rule).

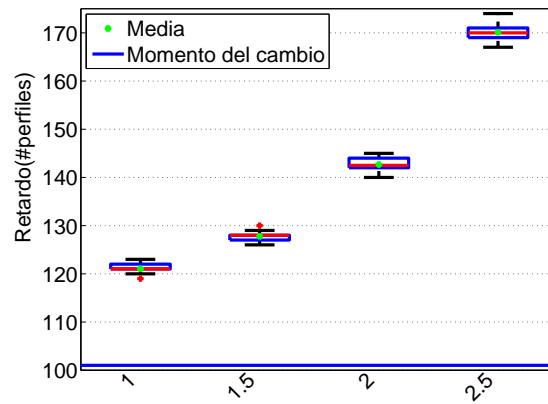


Figura 4.83: Gráfico Box Plot para 100 ejecuciones de Plug-In Martingale y 5 factores (FCT) diferentes de multiplicación de h (Normal Reference Rule).

FCT	Retardo (N° perfiles)	FA	Sin detectar
1	20,05	-	-
1,5	26,78	-	-
2	41,66	-	-
2,5	69,12	-	-
3	-	-	-

Cuadro 4.42: Media de retardos para perfiles sin ruido, FA y cambios no detectados para un conjunto de 100 ejecuciones de la Plug-in Martingale.

Conclusión

El modelo detecta cambios en anchura sin dificultades.

4.3.12. Experimento 12

Probaremos a continuación el efecto de señales con diferente nivel de ruido:

- Grupo 1: $\alpha_1 = \alpha_2 = 2$; $fct = 0.01$; $A_0 = 10$
- Grupo 2: $\alpha_1 = \alpha_2 = 4$; $fct = 0.01$; $A_0 = 10$

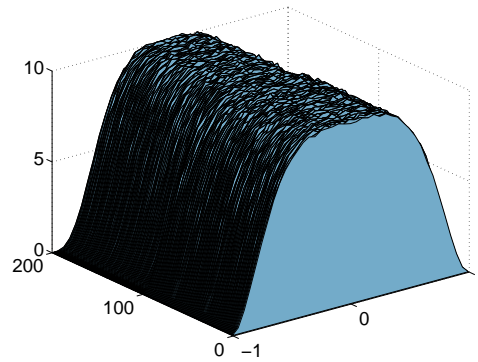


Figura 4.84: Perfiles Experimento 12.

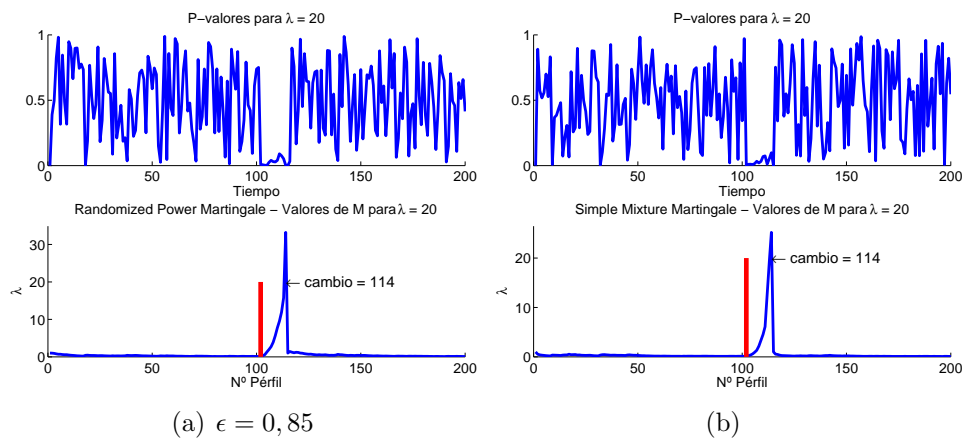
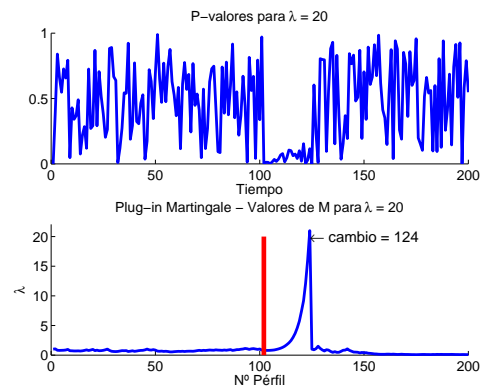


Figura 4.85: Ejemplos de detección del experimento 12. Se indica en cada gráfico el momento en el que la arquitectura de martingalas identifica el cambio.



(a) Factor de multiplicación de h (FCT)
 $= 1$

Figura 4.86: Ejemplos de detección del experimento 12. Se indica en cada gráfico el momento en el que la arquitectura de martingalas identifica el cambio.

En este caso podemos observar en las figuras 4.85 y 4.86 el momento en el que se detecta el cambio con un retardo. La línea roja marca el momento del cambio real. Debido al carácter estocástico de las martingalas, detallamos valores obtenidos después de ejecutar el algoritmo 100 veces para cada una de estas martingalas.

Randomized Power Martingale

Para esta martingala se han obtenido resultados para 10 ϵ distintas. Podemos ver el resultado obtenido en el cuadro 4.43. En la figura 4.87 podemos ver el resultado de detección de cambios, para 100 ejecuciones, mediante un gráfico Box Plot de resultados. Se han eliminado las falsas alarmas (FA) para evitar desvirtuar los resultados.

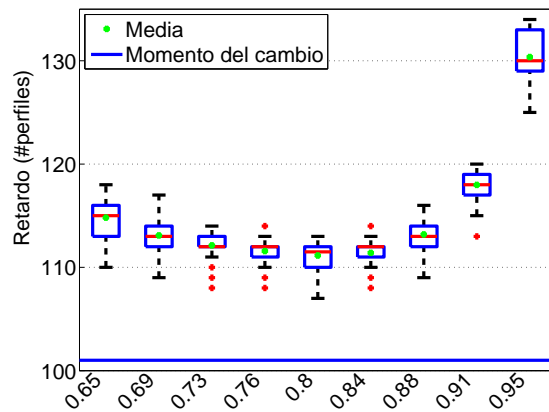


Figura 4.87: Gráfico Box Plot para 100 ejecuciones de Randomized Power Martingales por cada una de las 10 ϵ .

ϵ	Retardo (Nº perfiles)	FA	Sin detectar
0,65	13,81	-	-
0,69	12,09	-	-
0,73	11,13	-	-
0,76	10,58	-	-
0,8	10,15	-	-
0,84	10,40	-	-
0,88	12,19	-	-
0,91	16,98	-	-
0,95	29,36	-	-
0,99	-	-	100

Cuadro 4.43: Media de retardos, FA y cambios no detectados para un conjunto de 100 ejecuciones de la Randomized Power Martingale y varios valores de ϵ .

Simple Mixture

A continuación mostramos resultados para 100 ejecuciones de Simple Mixture.

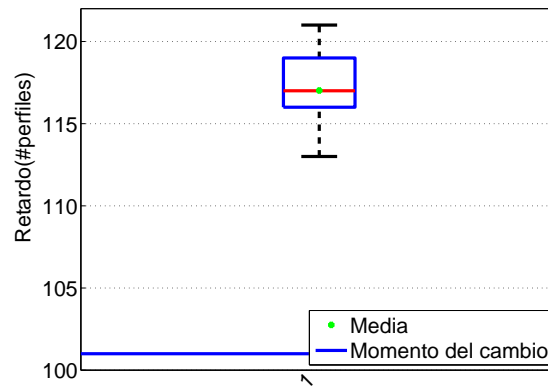


Figura 4.88: Gráfico Box Plot para 100 ejecuciones.

Retardo (N° perfiles)	FA	Sin detectar
16,02	1	-

Cuadro 4.44: Media de retardos, FA y cambios no detectados para un conjunto de 100 ejecuciones de la Simple Mixture Martingale.

Plug-in Martingale

En este caso veremos los resultados para esta martingala para 100 ejecuciones y 5 valores de FCT distintas. FCT es un factor que multiplica el valor de h (Normal Reference Rule).

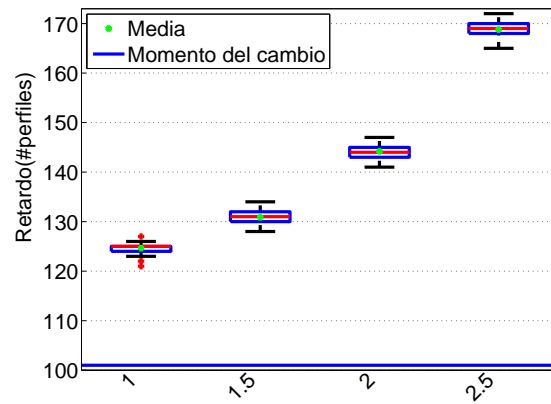


Figura 4.89: Gráfico Box Plot para 100 ejecuciones de Plug-In Martingale y 5 factores (FCT) diferentes de multiplicación de h (Normal Reference Rule).

FCT	Retardo (N° perfiles)	FA	Sin detectar
1	23,65	-	-
1,5	29,88	-	-
2	43,20	-	-
2,5	67,79	-	-
3	-	-	-

Cuadro 4.45: Media de retardos para perfiles sin ruido, FA y cambios no detectados para un conjunto de 100 ejecuciones de la Plug-in Martingale.

Conclusión

Con esta prueba y la de la sección 4.3.12, podemos decir que el modelo identifica cambios independientemente de que el cambio sea ancho-estrecho o estrecho-ancho.

4.4. Conclusiones

Una vez ejecutados todos los experimentos podemos llegar a las siguientes conclusiones:

- Desde el punto de vista del retardo, Randomized y Simple Mixture, por este orden, son las martingalas que tienen menor retardo. Por el contrario Plug-In da un retardo significativamente superior a las dos anteriores.
- Respecto a Falsas Alarmas: Plug-In es la que con diferencia se comporta mejor. En este sentido no hay diferencias significativas entre Randomized y Simple Mixture.
- Para el caso de Randomized los valores de ϵ que mejor se comportan son aquellos que pertenecen al intervalo $[0,8 \ 0,95]$.
- Plug-In se comporta mejor para valores de FCT dentro del intervalo $[1 \ 2]$.
- En la detección de cambios en amplitud el método detecta tanto los cambios bajo-alto como alto-bajo e independiente del número de cambios que haya. Es igualmente efectivo para cambios en anchura, en este caso tampoco importa si el cambio es ancho-estrecho o estrecho-ancho.
- El ruido marca el límite de sensibilidad del modelo. A mayor ruido menor sensibilidad para la detección del cambio y viceversa.
- El hecho de que el ruido de los perfiles sea independiente o dependiente no afecta al resultado final.
- En general, el suavizado de señales con ruido mejora los resultados finales obtenidos.

Capítulo 5

Capacidad de detección de cambios en perfiles de fusión nuclear

En este apartado se realizarán una serie de experimentos con grupos de perfiles procedentes del TJ-II. El objetivo será identificar posibles cambios dentro de un grupo de perfiles. Con este fin generaremos un espacio de características de dos dimensiones compuesto por la desviación estándar y la media de los puntos de cada perfil.

5.1. Perfiles de Bolometría del TJ-II (disparo 21319)

Para este caso vamos a intentar detectar cambios en los perfiles de bolometría del TJ-II, concretamente del disparo 21319. Como puede verse en la figura 2.3 los perfiles crecen suavemente de derecha a izquierda, sin embargo hay un punto en el que aparece un cambio más evidente justo en la cúspide en el que se produce una caída brusca que intentaremos detectar mediante nuestro test de martingalas.

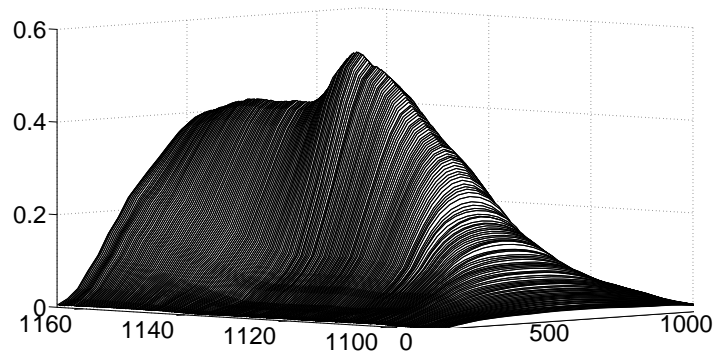


Figura 5.1: Perfil de bolometría del TJ-II (disparo 21319).

5.1.1. Randomized Power

Para varios valores de ϵ y $\lambda = 20$ obtenemos los siguientes resultados:

ϵ	μ (Nº perfiles)	σ^2	Sin detectar
0,65	13,36	2,20	9
0,69	15,64	4,68	6
0,73	13,33	5,52	5
0,76	14,82	4,37	3
0,8	18,50	5,49	2
0,84	17,58	8,51	8
0,88	20,11	5,01	11
0,91	-	-	20
0,95	-	-	20
0,99	-	-	20

Cuadro 5.1: Media de retardos, desviación estándar y cambios no detectados para un conjunto de 100 ejecuciones de la Randomized Power Martingale y varios valores de ϵ .

5.1.2. Simple Mixture

A continuación mostramos resultados para 20 ejecuciones de Simple Mixture.

ϵ	μ (Nº perfiles)	σ^2	Sin detectar
21,14	4,88	13	-

Cuadro 5.2: Media de retardos, desviación estándar y cambios no detectados para un conjunto de 20 ejecuciones de la Simple Mixture Martingale.

5.1.3. Plug-in Martingale

En este caso veremos los resultados para esta martingala para 20 ejecuciones y 5 valores de FCT distintas. FCT es un factor que multiplica el valor de h (Normal Reference Rule).

ϵ	μ (Nº perfiles)	σ^2	Sin detectar
1	21,08	10,95	8
1,5	-	-	-
2	-	-	-
2,5	-	-	-
3	-	-	-

Cuadro 5.3: Media de retardos, desviación estándar y cambios no detectados para un conjunto de 20 ejecuciones de la Plug-in Martingale.

5.1.4. Preprocesamiento de datos

Para poder conseguir los datos obtenidos es necesario realizar un preprocesado de los perfiles. En este caso se ha realizado un tratamiento a los perfiles consistente en el cálculo de la pendiente de un grupo formado por 3

perfiles. El resultado para cada punto de los 3 perfiles ha generado un nuevo perfil y que podemos ver en la figura 5.1.4.

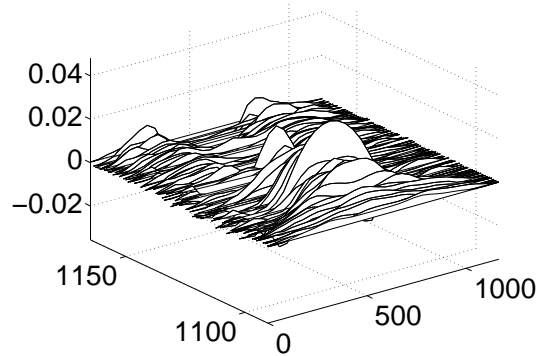


Figura 5.2: Perfiles obtenidos a partir del cálculo de pendientes de grupos de 3 perfiles.

Se ha realizado la misma tarea para grupos de 5 a 10 perfiles no detectándose diferencias apreciables en los resultados.

5.2. Conclusiones

Con los resultados mostrados en la sección 5.1.4 se demuestra que el test de martingalas es una herramienta efectiva para su aplicación en la detección de cambios en flujos de datos variables en el tiempo procedentes de fenómenos físicos de fusión nuclear. Como característica destacable de este test tenemos la característica de propósito general que nos evita realizar desarrollos específicos para cada caso particular.

Se demuestra además la importancia de realizar un preprocesado efectivo que genere unos clusters lo suficientemente separables para hacer que el método sea lo suficientemente efectivo. En este caso se ha optado por el cálculo de la pendiente para grupos de 3 perfiles, obteniéndose resultados aceptables.

Capítulo 6

Conclusiones y futuros trabajos

Los resultados de este trabajo fueron presentados en el 9º congreso de la Agencia Internacional de Energía Atómica, sobre control, adquisición de datos y participación remota para la investigación en fusión realizado en Mayo de 2013, bajo el título *Real-time change detection in data streams with FPGAs*. Con tal motivo se completó este trabajo con la implementación en una FPGA del test de martingalas descrita en este trabajo. Como martingala se implementó la Randomized Power Martingale obteniéndose los mismos resultados del capítulo 4. Esta implementación fue realizada por un equipo de la Universidad Politécnica de Madrid, quedando dicha implementación fuera del alcance de este trabajo.

6.1. Conclusiones

Como conclusiones principales de la arquitectura de martingalas propuesta tenemos las siguientes:

- Desde el punto de vista del retardo, Randomized y Simple Mixture, por este orden, son las martingalas que tienen menor retardo. Por el contrario Plug-In da un retardo significativamente superior a las dos anteriores.
- Respecto a Falsas Alarmas: Plug-In es la martingala que con diferencia

se comporta mejor. En este sentido no hay diferencias significativas entre Randomized y Simple Mixture.

- Para el caso de Randomized los valores de ϵ que mejor se comportan son aquellos que pertenecen al intervalo $[0, 80, 95]$.
- Plug-In se comporta mejor para valores de FCT dentro del intervalo $[1, 2]$.
- Se demuestra que el modelo propuesto es efectivo para la detección de cambios en flujos de datos procedentes de medidas de fenómenos físicos de fusión nuclear.
- El modelo propuesto es de propósito general no necesitándose desarrollos específicos para cada aplicabilidad particular. No obstante es preciso realizar un preprocesado efectivo de los datos a analizar para construir un vector de características adecuado al fenómeno que se desee analizar.
- Es aplicable tanto a detección de cambios en tiempo real como en procesado *off-line*.
- Como punto a mejorar de este modelo tenemos el excesivo tiempo de proceso que puede dificultar su aplicación a casos de tiempo real.

6.2. Futuros Trabajos

A continuación se detallan futuros trabajos que podrían realizarse alrededor de este trabajo:

- Mejora del rendimiento para aplicaciones en tiempo real. una posibilidad es generar una solución parecida a la planteada en la sección 3.7.2.
- Estudio de su posible aplicabilidad como modelo matemático para la demostración de la estacionalidad o no de determinadas señales de interés científico-técnico.
- Desarrollo de otras medidas de *extrañeza* más robustas que doten al modelo de mayor sensibilidad y que contribuya a evitar falsas alarmas.

Bibliografía

- [1] International Energy Agency. Key world energy statistics. Technical report, International Energy Agency, 9, rue de la Fédération; 75739 Paris Cedex 15, 2013.
- [2] David Rapisarda Socorro. *Estudio de la dinámica de iones por métodos espectroscópicos en el TJ-II*. PhD thesis, Universidad Complutense - CIEMAT, 2007.
- [3] Giuseppe A. Rattá Gutiérrez. *Técnicas de Minería de datos aplicadas a fusión nuclear: Predicción en tiempo real y clasificación*. PhD thesis, UNED - CIEMAT, 2010.
- [4] Shen-Shyang Ho and Harry Wechsler. A martingale framework for detecting changes in data streams by testing exchangeability. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(12):2113–2127, 2010.
- [5] Alexey Tsymbal, Mykola Pechenizkiy, Pádraig Cunningham, and Seppo Puuronen. Dynamic integration of classifiers for handling concept drift. *Inf. Fusion*, 9(1):56–68, January 2008.
- [6] Albert Bifet, Geoss Holmes, Richard Kirkby, and Bernhard Pfahringer. *DATA STREAM MINING: A Practical Approach*. The University of Waikato, 2011.
- [7] Kenneth O. Stanley. Learning concept drift with a committee of decision trees. Technical Report AI03-302, Department of Computer Sciences, The University of Texas at Austin, 2003.
- [8] Albert Bifet, Geoff Holmes, Bernhard Pfahringer, Philipp Kranen, Hardy Kremer, Timm Jansen, and Thomas Seidl. Moa: Massive online analysis, a framework for stream classification and clustering. In *Journal of Machine Learning Research (JMLR) Workshop and Conference*

- Proceedings, Volume 11: Workshop on Applications of Pattern Analysis*, pages 44–50. Journal of Machine Learning Research, 2010.
- [9] Vladimir N. Vapnik. *Statistical Learning Theory*. Wiley-Interscience, 1998.
- [10] J. Vega, A. Murari, A. Pereira, S. Gonzalez, and I. Pastor. Accurate and reliable image classification by using conformal predictors in the t_j -ii thomson scattering. *REVIEW OF SCIENTIFIC INSTRUMENTS*, 81(10):10E118, Oct 2010.
- [11] Glenn Shafer and Vladimir Vovk. A tutorial on conformal prediction. *J. Mach. Learn. Res.*, 9:371–421, June 2008.
- [12] Valentina Fedorova, Alexander Gammerman, Ilia Nouretdinov, and Vladimir Vovk. Plug-in martingales for testing exchangeability on-line. *CoRR*, abs/1204.3251, 2012.
- [13] Vladimir Vovk, Alex Gammerman, and Glenn Shafer. *Algorithmic Learning in a Random World*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005.
- [14] Vladimir Vovk, Ilia Nouretdinov, and Alexander Gammerman. Testing exchangeability on-line. In Tom Fawcett and Nina Mishra, editors, *ICML*, pages 768–775. AAAI Press, 2003.
- [15] W. Martinez. *Computational statistics handbook with MATLAB*. Chapman & Hall/CRC, 2001.
- [16] V. Epanechnikov. Nonparametric estimates of a multivariate probability density. *Theory of probability and its applications*, 14:153–158, 1969”.
- [17] Shen-Shyang Ho and Harry Wechsler. Detecting changes in unlabeled data streams using martingale. In *Proceedings of the 20th International Joint Conference on Artificial Intelligence, IJCAI’07*, pages 1912–1917, San Francisco, CA, USA, 2007. Morgan Kaufmann Publishers Inc.
- [18] Johan Himberg. A som based cluster visualization and its application for false coloring. *Helsinki University of Technology. Laboratory of Computer and Information Science.*, 1999.