
Inferencia bayesiana del movimiento propio de las
estrellas sobre GPU



Trabajo Fin de Máster

Jose Etxeberria Méndez

Trabajo de investigación para el

Máster Universitario en Inteligencia Artificial Avanzada

Universidad Nacional de Educación a Distancia (UNED)

Escuela Técnica Superior de Ingeniería Informática

Dirigido por

Dr. Luis Sarro Baro

Presentado en

Junio 2020

Resumen

La búsqueda de los movimientos propios de las estrellas es uno de los focos de la astronomía, al permitir identificar grupos de movimiento propio común que puedan compartir su origen y por tanto edad y composición química. Tradicionalmente la estimación del movimiento propio se ha realizado mediante métodos del paradigma frecuentista, principalmente regresión por mínimos cuadrados. Este tipo de métodos presentan problemas ante bajos volúmenes de datos y datos con altas irregularidades. En este trabajo se defiende la preferencia del paradigma bayesiano para la estimación de los movimientos propios y se desarrolla una herramienta que permite realizar la estimación utilizando modelos bayesianos jerárquicos de manera robusta ante la presencia de espurios y datos sesgados. La implementación se realiza sobre GPU, optimizando al máximo la paralelización del procesamiento.

Palabras clave Modelo bayesiano jerárquico, movimiento propio, programación GPU

Abstract

Measuring the proper motion of stars is one of the studies where astronomy focuses because it leads to identify star groups with common proper motion, that may share their origin and so their age and chemical composition. Traditionally, proper motions estimation has been performed with methods based on the frequentist paradigm, mainly with least squares regression. These methods may arise problems with low data volumes and irregular data. In this project, preference for bayesian paradigm is advocated for proper motions estimate and it is developed a tool that provides the estimate robust to outliers and offsets in data. Development is performed on GPU, maximizing parallelization processing capabilities.

Key words Hierarchical bayesian model, stars proper motion, GPU-computing

Laburpena

Astronomiaren oinarrietako bat, izarren mugimendu propioen bilaketa da, izarren mugimendu propio taldeen identifikazioa baimenduz. Jatorria

bera izanda, adina eta konposizio kimiko ere partekatzen dute. Tradizional-ki mugimendu propioen estimazioa paradigma frekuentistaren bitartez egin izan da, karratu txikieneko erregresioaren bitartez batez ere. Metodo mota hauek arazoak ematen dituzte irregulartasun handia duten eta kantitate txikian dauden datuen aurrean. Lan honen bitartez bayestar paradigmaren lehenespena defendatzen da mugimendu propioen estimazioa aurrera eramateko, datu faltsu eta partzialen aurrean era indartsuan hierarkizatutako bayestar ereduak erabiltzea bermatzen duen lanabes bat garatuz. Implementazioa GPU-aren bitartez egin da, prozesaketaren paralelizazioa ahalik eta gehien optimizatuz.

Hitz gakoak Hierarkizatutako bayestar ereduak, izarren mugimendu propioa, GPU prozesaketa

Índice general

1. Introducción	1
1.1. Motivación	1
1.2. Propuesta y objetivos	1
1.3. Proyecto COSMIC-DANCE	2
1.4. Movimientos propios en astronomía	3
1.5. Modelos bayesianos	8
1.6. Estructura del documento	10
2. Revisión de la bibliografía	12
2.1. Estimación de movimientos propios	12
2.2. Inferencia bayesiana	15
3. Modelo bayesiano jerárquico para inferencia de movimientos propios	19
3.1. Introducción a la inferencia bayesiana	19
3.1.1. Fundamentos de la inferencia bayesiana	19
3.1.2. Fundamentos de los modelos bayesianos jerárquicos	22
3.2. Aplicación a la inferencia de movimientos propios a partir de medidas astrométricas	23
3.2.1. El modelo bayesiano jerárquico para movimientos propios	24
3.2.2. Muestreador utilizado para la estimación	28
4. Implementación	34
4.1. Entorno de trabajo	34
4.1.1. Hardware	34
4.1.2. Software	35
4.2. Implementación del algoritmo	37

4.2.1. Entrada de datos	38
4.2.2. Preprocesado	38
4.2.3. Caracterización	39
4.2.4. Postprocesado	42
5. Evaluación	44
5.1. Metodología de evaluación	44
5.2. Colecciones de evaluación	45
5.3. Métricas de evaluación	45
5.4. Gráficas	46
6. Resultados	49
6.1. Datos sintéticos	49
6.1.1. Análisis de un experimento individual	50
6.1.2. Análisis estadístico de varios experimentos	60
6.2. Datos reales	76
6.2.1. Análisis de un experimento individual	76
6.2.2. Análisis estadístico de varios experimentos	80
6.3. Escalabilidad	86
7. Conclusiones y líneas de futuro	89
7.1. Conclusiones	89
7.2. Líneas de futuro	91
8. Bibliografía	92
A. Manual de uso	97
B. Propuesta de estrellas con movimiento propio	99

Índice de Figuras

1.1.	Posiciones de una estrella en ascensión recta (eje Y) a lo largo del tiempo (eje X). Colección: muestra de DANCe.	6
1.2.	Posiciones de una estrella en ascensión recta (eje Y) a lo largo del tiempo (eje X). Colección: muestra de DANCe.	6
1.3.	Posiciones de una estrella en ascensión recta (eje Y) a lo largo del tiempo (eje X). Colección: muestra de DANCe.	7
1.4.	Posiciones de una estrella en declinación (eje Y) a lo largo del tiempo (eje X). Colección: muestra de DANCe.	7
3.1.	Representación gráfica del modelo bayesiano jerárquico propuesto	28
3.2.	Resumen de los parámetros intervinientes en el cálculo de verosimilitud.	29
6.1.	Posiciones de la estrella en ascensión recta (eje Y) a lo largo del tiempo (eje X) incluyendo los modelos lineales inferidos. Colección: Sintética sin movimientos propios. Configuración: $it = 8000, pop = 1500$	51
6.2.	Verosimilitud (eje Y) de los modelos lineales en función del valor que tienen (eje X) en cada parámetro, m_{denorm_α} (arriba) y b_{denorm_α} (abajo). La escala cromática indica el orden de muestreo (i.e., la verosimilitud). Colección: Sintética sin movimientos propios. Configuración: $it = 8000, pop = 1500$	52
6.3.	Dispersión de valores de los parámetros m (eje Y) frente a b (eje X). La escala cromática indica el orden de muestreo (i.e., la verosimilitud). Colección: Sintética sin movimientos propios. Configuración: $it = 8000, pop = 1500$	53

6.4. Histograma de los parámetros m_{denorm_α} (arriba) y b_{denorm_α} (abajo). Colección: Sintética sin movimientos propios. Configuración: $it = 8000, pop = 1500$	54
6.5. Distribución a posteriori de los parámetros, m_{denorm_α} (arriba) y b_{denorm_α} (abajo). Colección: Sintética sin movimientos propios. Configuración: $it = 8000, pop = 1500$	55
6.6. Verosimilitud (eje Y) de los modelos lineales en función del valor que tienen (eje X) en cada parámetro normalizado, m (arriba) y b (abajo). La escala cromática indica el orden de muestreo (i.e., la verosimilitud). Colección: Sintética sin movimientos propios. Configuración: $it = 8000, pop = 1500$	56
6.7. Posiciones de una estrella en declinación (eje Y) a lo largo del tiempo (eje X) incluyendo los modelos lineales inferidos. Colección: Sintética con movimientos propios. Configuración: $it = 8000, pop = 1500$	57
6.8. Verosimilitud (eje Y) de los modelos lineales en función del valor que tienen (eje X) en cada parámetro, m_{denorm_α} (arriba) y b_{denorm_α} (abajo). La escala cromática indica el orden de muestreo (i.e., la verosimilitud). Colección: Sintética con movimientos propios. Configuración: $it = 8000, pop = 1500$	58
6.9. Dispersión de valores de los parámetros m (eje Y) frente a b (eje X). Colección: Sintética con movimientos propios. Configuración: $it = 8000, pop = 1500$	60
6.10. Distribución a posteriori de los parámetros, m_{denorm_α} (arriba) y b_{denorm_α} (abajo). Colección: Sintética con movimientos propios. Configuración: $it = 8000, pop = 1500$	61
6.11. Verosimilitud (eje Y) de los modelos lineales en función del valor que tienen (eje X) en cada parámetro normalizado, m (arriba) y b (abajo). La escala cromática indica el orden de muestreo (i.e., la verosimilitud). Colección: Sintética con movimientos propios. Configuración: $it = 8000, pop = 1500$	62
6.12. Posiciones normalizadas de una estrella en declinación (eje Y) a lo largo del tiempo (eje X) incluyendo los modelos lineales inferidos. Colección: Sintética con movimientos propios. Configuración: $it = 8000, pop = 1500$	63

6.13. Nivel del factor Bayes (coordenada radial) a lo largo de 10 experimentos (coordenada angular). Colección: Sintética sin movimientos propios. Configuración: $it = 8000, pop = 1500$. . .	64
6.14. Nivel del factor Bayes (coordenada radial) a lo largo de 10 experimentos (coordenada angular). Colección: Sintética con movimientos propios. Configuración: $it = 8000, pop = 1500$. . .	65
6.15. Especificidad en la predicción (eje Y) a lo largo de 10 experimentos (eje X). Colección: Sintética sin movimientos propios. Configuración: $it = 8000, pop = 1500$	66
6.16. Sensibilidad en la predicción (eje Y) a lo largo de 10 experimentos (eje X). Colección: Sintética con movimientos propios. Configuración: $it = 8000, pop = 1500$	67
6.17. Sensibilidad y especificidad medias en la predicción (eje Y), promediadas entre 10 experimentos, variando el tamaño de la población (eje X). Colecciones: Sintética con movimientos propios (azul) y sin movimientos propios (verde). Configuración: $it = 2000$	68
6.18. Sensibilidad y especificidad medias en la predicción (eje Y), promediadas entre 10 experimentos, variando el tamaño de la población (eje X). Colecciones: Sintética con movimientos propios (azul) y sin movimientos propios (verde). Configuración: $it = 4000$	69
6.19. Sensibilidad y especificidad en la predicción (eje Y) variando el tamaño de la población (eje X). Colecciones: Sintética con movimientos propios (azul) y sin movimientos propios (verde). Configuración: $it = 6000$	69
6.20. Sensibilidad y especificidad medias en la predicción (eje Y), promediadas entre 10 experimentos, variando el tamaño de la población (eje X). Colecciones: Sintética con movimientos propios (azul) y sin movimientos propios (verde). Configuración: $it = 8000$	69
6.21. Sensibilidad y especificidad medias en la predicción (eje Y), promediadas entre 10 experimentos, variando el tamaño de la población (eje X). Colecciones: Sintética con movimientos propios (azul) y sin movimientos propios (verde). Configuración: $it = 10000$	70

6.22. Especificidad en la predicción (eje Y) a lo largo de 10 experimentos (eje X). Colección: Sintética sin movimientos propios. Configuración: $it = 8000, pop = 2000$	71
6.23. Sensibilidad en la predicción (eje Y) a lo largo de 10 experimentos (eje X). Colección: Sintética con movimientos propios. Configuración: $it = 8000, pop = 2000$	72
6.24. Especificidad en la predicción (eje Y) a lo largo de 10 experimentos (eje X). Colección: Sintética sin movimientos propios ni sesgos ni espurios. Configuración: $it = 8000, pop = 2000$. .	73
6.25. Especificidad en la predicción (eje Y) a lo largo de 10 experimentos (eje X). Colección: Sintética sin movimientos propios ni espurios pero sí sesgos. Configuración: $it = 8000, pop = 2000$	73
6.26. Sensibilidad en la predicción (eje Y) a lo largo de 10 experimentos (eje X). Colección: Sintética con movimientos propios sin espurios pero sí sesgos. Configuración: $it = 8000, pop = 2000$	74
6.27. Especificidad en la predicción (eje Y) a lo largo de 10 experimentos (eje X). Colección: Sintética sin movimientos propios ni sesgos pero sí espurios. Configuración: $it = 8000, pop = 2000$	74
6.28. Sensibilidad en la predicción (eje Y) a lo largo de 10 experimentos (eje X). Colección: Sintética con movimientos propios sin sesgos pero sí espurios. Configuración: $it = 8000, pop = 2000$	75
6.29. Posiciones de una estrella en ascensión recta (eje Y) a lo largo del tiempo (eje X) incluyendo los modelos inferidos. Colección: muestra de DANCe. Configuración: $it = 6000, pop = 1000$	77
6.30. Verosimilitud (eje Y) de los modelos lineales en función del valor que tienen (eje X) en cada parámetro, m_{denorm_α} (arriba) y b_{denorm_α} (abajo). La escala cromática indica el orden de muestreo (i.e., la verosimilitud). Colección: muestra de DANCe. Configuración: $it = 6000, pop = 1000$	78
6.31. Dispersión de valores de los parámetros m (eje Y) frente a b (eje X). Colección: muestra de DANCe. Configuración: $it = 6000, pop = 1000$	79
6.32. Distribución a posteriori de los parámetros, m_{denorm_α} (arriba) y b_{denorm_α} (abajo). Colección: muestra de DANCe. Configuración: $it = 6000, pop = 1000$	79

6.33. Sensibilidad en la predicción (eje Y) variando el tamaño de la población (eje X). Colección: muestra de DANCe. Configuración: $it = 2000$	81
6.34. Sensibilidad en la predicción (eje Y) variando el tamaño de la población (eje X). Colección: muestra de DANCe. Configuración: $it = 4000$	81
6.35. Sensibilidad en la predicción (eje Y) variando el tamaño de la población (eje X). Colección: muestra de DANCe. Configuración: $it = 6000$	82
6.36. Sensibilidad en la predicción (eje Y) variando el tamaño de la población (eje X). Colección: muestra de DANCe. Configuración: $it = 8000$	82
6.37. Matrices de confusión de las predicciones de movimiento propio (eje X) respecto a la referencia (eje Y) de estrellas en las zonas blanca (arriba) y gris(abajo). Colección: muestra de DANCe. Configuración: $it = 8000pop = 1500$	84
6.38. Nivel del factor Bayes (coordenada radial) a lo largo de 10 experimentos (coordenada angular). Colección: muestra de DANCe. Configuración: $it = 8000, pop = 1500$	85
6.39. Tiempo consumido (eje Y) por la GPU (arriba) y por todo el programa (abajo) a lo largo de experimentos con distinto tamaño de población (eje X). Colección: Sintéticos con movimiento. Configuración: $it = 10000$	87
6.40. Tiempo consumido (eje Y) por la GPU (arriba) y por todo el programa (abajo) a lo largo de experimentos con distinto número de iteraciones (eje X). Colección: Sintéticos con movimiento. Configuración: $pop = 250$	88
6.41. Tiempo consumido (eje Y) por la GPU (arriba) y por todo el programa (abajo) a lo largo de experimentos con distinto número de estrellas (eje X). Colección: Sintéticos con movimiento. Configuración: $it = 50pop = 250$	88

Índice de Tablas

5.1. Interpretación del factor de Bayes basada en las propuestas de Jeffreys (1998) y Kass and Raftery (1995) y adaptada a la nomenclatura del problema en estudio.	45
B.1. Listado de estrellas de una muestra del catálogo DANCe, sobre las cuales se estima la presencia de movimientos propios. Se omiten los ratios en caso de que dicho eje no haya sido estimado con movimiento propio, o no con la suficiente evidencia sobre la hipótesis contraria.	100

Capítulo 1

Introducción

1.1. Motivación

Desde la existencia de civilizaciones primitivas la observación de los fenómenos celestes ha sido un pilar fundamental en la ciencia. Desde la predicción de la llegada del invierno que motivó el inicio del sedentarismo, hasta la medición de la edad de la tierra o la curvatura espacio-temporal inducida por las ondas gravitacionales, la observación del espacio nos brinda la hoja de ruta que ha guiado, guía y guiará nuestro mundo.

La observación de galaxias, sistemas estelares y planetas externos ayuda a comprender las leyes que rigen el comportamiento del universo en el que nos encontramos inmersos, promoviendo un desarrollo tecnológico sin precedentes que posteriormente se materializa en la evolución de la civilización.

1.2. Propuesta y objetivos

El conocimiento del movimiento propio de las estrellas es fundamental para validar las leyes que definen la naturaleza del universo. Nos permite identificar grupos de movimiento propio común que, presumiblemente, tuvieron origen por el colapso de una nube de gas y polvo y por lo tanto comparten la misma edad y composición química. Por lo que su observación es uno de los principales propósitos de la astronomía.

El presente trabajo tiene como objetivo el desarrollo de un herramienta de software para la estimación robusta frente a sesgos y espurios del movimiento propio de millones de estrellas, ayudándose para ello de la capacidad de paralelización de las tarjetas gráficas.

Está integrado en el marco del proyecto Cosmic-DANCe, el cual busca ampliar nuestro conocimiento del cosmos valiéndose de las más precisas técnicas de medición.

A pesar de haber realizado un gran avance en los últimos años, nuestro desconocimiento del cosmos es equiparable a su envergadura. Debido a las grandes escalas temporales y astronómicas distancias a las que se producen los cambios cosmológicos, los datos de los que se dispone son relativamente escasos para concluir comportamientos globales de los astros.

Se propone la utilización de un modelo bayesiano jerárquico para inferir la velocidad de los movimientos propios de las estrellas. Se parte de datos de su posición en sucesivos instantes de tiempo. Esta técnica permite obtener de forma eficiente la distribución de probabilidad a posteriori de modelos complejos de elevado número de parámetros.

1.3. Proyecto COSMIC-DANCE

El proyecto Cosmic-Dance, promovido por la Universidad de Burdeos, Francia, se vale de las más precisas técnicas de medición, comparables al telescopio espacial Gaia, pero quintuplicando su horizonte de observación, para profundizar en los clusters de estrellas más jóvenes oscurecidas por polvo espacial.

En un periodo de hasta 60 años, se extraen datos de posiciones y fotometría de longitud de onda múltiple y mediante modernas técnicas de minería de datos se catalogan para censar miembros de hasta 20 clústers jóvenes (< 100 Myr, million years) y cercanos (< 500 pc, parsec). Los datos obtenidos por Gaia corresponden a estrellas de alta luminosidad, trabaja en el rango visible del espectro lumínico, mientras que los recabados por DANCe son infrarrojos y permiten observar zonas donde se concentra la formación de nuevas estrellas. En estas zonas Gaia se ciega por la elevada emisión nebular en el rango visible. El trabajo complementario de Cosmic-DANCe y GAIA permite cubrir los astros desde el límite de fragmentación hasta las estrellas masivas OB.

El proyecto trata de localizar asociaciones de estrellas por la similitud que presentan sus movimientos propios. Al estar las estrellas muy lejos, las mínimas velocidades observables son del orden de 20 a 50 mas/año, milisegundos de arco por año.

Con el siguiente proceso de análisis de datos se obtienen precisiones de 10 a 25 mas/año y combinando imágenes de numerosos años se alcanzan precisiones de 1 mas/año:

Se recopilan y catalogan imágenes de distintos instrumentos. Se hacen dos correcciones, la primera de distorsiones de grado bajo por comparación con el catálogo externo 2MASS, la segunda de distorsiones de mayor grado, empleando la superposición de imágenes de distintos instrumentos. Una vez se dispone de las imágenes sin distorsión, se aplica una solución basada en fotometría para cada instrumento fotométrico y finalmente, se infieren los movimientos propios a partir de un ajuste lineal de mínimos cuadrados a las posiciones en distintas épocas (Bouy et al., 2013). Se dan más detalles sobre el ajuste en la Sección 2.1.

La caracterización de los clústeres de estrellas es un proceso anidado ya que las características de un clúster dependen de los miembros que lo forman, y la determinación de los miembros que lo forman depende de las características del clúster. La búsqueda de resultados requiere un análisis de cantidades ingentes de datos en espacios de alta dimensionalidad por lo que se hace patente la necesidad de usar técnicas avanzadas de minería de datos, como son los métodos bayesianos. Estos métodos se valen de unas distribuciones preliminares para acotar los valores probables de los parámetros a estimar, esta información preliminar adquiere mayor transcendencia conforme evoluciona el conocimiento del espacio a lo largo de los años.

1.4. Movimientos propios en astronomía

Sistema de coordenadas A la hora de realizar mediciones es necesaria la existencia de un sistema de referencia. En las observaciones astronómicas, el sistema de referencia habitual es el sistema de coordenadas ecuatoriales heliocéntrico. Éste es un sistema de coordenadas esféricas, donde el punto de observación es el sol y el sistema de coordenadas es una corteza esférica que lo rodea, denominada esfera celeste. Los ejes de coordenadas son la ascensión recta (α) y declinación (δ), equivalentes a la longitud y latitud terrestres, que indican una posición unívoca en la esfera.

Referencia Dado el sistema de coordenadas ecuatorial, se debe determinar una referencia respecto a los ejes de coordenadas. Ésta referencia la marcan los cuásares (Colom, 2005), que debido a estar extremadamente lejos,

su posición observable no cambia, ni pasados cientos de años, manteniendo estable el sistema de referencia a lo largo del tiempo. Algunas constelaciones son ejemplo de estos cuásares, el cambio en su posición es imperceptible a pesar de comparar observaciones antiquísimas. El movimiento propio de las estrellas no tan lejanas es cuantizado gracias a la permanencia del sistema de coordenadas.

Movimiento propio Las estrellas realizan diversas trayectorias como consecuencia de la inercia heredada de su formación y la influencia gravitatoria de las galaxias en las que se encuentran, orbitando en torno a sus centros. El movimiento en el espacio, respecto al sol, se descompone en velocidad radial y velocidad transversal. El movimiento producido por la velocidad radial es aquel que aleja o acerca a la estrella respecto al sol y se obtiene por la diferencia de frecuencia de la luz en base al efecto Doppler. El movimiento propio es el desplazamiento producido a consecuencia de la velocidad transversal y se mide al variar la posición del astro en el sistema de coordenadas ecuatorial a lo largo del tiempo. Como consecuencia del movimiento de traslación de la tierra, las observaciones presentan un movimiento cíclico denominado paralaje. En este trabajo se asume que los datos de entrada tienen aplicada la corrección del paralaje.

Por mucho que las estrellas se muevan a velocidades muy elevadas y orbiten en torno al centro de las galaxias, los tiempos de traslación son tan astronómicos que en siglos de observación no se percibe una curvatura de su trayectoria y por tanto no se perciben modificaciones en las velocidades. El modelo que define su movimiento se puede aproximar sin pérdida de generalidad a una línea recta en sistema de coordenadas ecuatorial.

Tomando imágenes de las estrellas a lo largo de los años se aprecian sutiles modificaciones en su ascensión recta y declinación, pudiendo utilizar estas posiciones para inferir la velocidad del movimiento propio de la estrella en cada coordenada. La velocidad percibida en cada coordenada es constante y por tanto, en la representación de cada coordenada frente al tiempo también se espera una recta. De modo que la estimación se realiza individualmente para α (ascensión recta) y δ (declinación).

$$\begin{aligned}\alpha &= m_\alpha \cdot t + b_\alpha \\ \delta &= m_\delta \cdot t + b_\delta\end{aligned}\tag{1.1}$$

Siendo

α	Posición en el eje de ascensión recta.
δ	Posición en el eje de declinación.
m	Pendiente del modelo lineal de cada eje.
b	Intersección del modelo lineal con cada eje en el origen.

La pendiente de estas rectas, representa la velocidad de la estrella en cada eje, la búsqueda que pretende el presente estudio.

Para ejemplificar la naturaleza de los datos con los que se desea trabajar, en las figuras 1.1, 1.2, 1.3 y 1.4 se muestran observaciones de varias estrellas del catálogo PPM extraídas del proyecto DANCe.

En estas figuras se puede observar un carácter disperso de las medidas y bastante diferencia entre distintas estrellas, no desde el punto de vista de su aparente movimiento, que evidentemente cada estrella tendrá el suyo, si no respecto a la calidad de los datos. Mientras hay estrellas con numerosas observaciones y algunas de ellas relativamente distantes en el tiempo (Fig. 1.1), la mayoría de ellas reducen las observaciones a pocos años.

Se ha indicado con una escala cromática el instrumento con el que se ha tomado cada medida, se puede observar como varios instrumentos han extraído datos de todas las estrellas (por ejemplo: 3, 11, 21) mientras que otros instrumentos solo aportan observaciones de ciertas estrellas. Cada uno de estos instrumentos acarrea un sesgo instrumental, comúnmente conocidos como *offsets*. En la medida de lo posible estos sesgos se reducen durante la observación y agregación de los datos, pero siempre hay un componente no controlado que marca diferencias entre las observaciones de cada instrumento (ver Fig. 1.3).

A su vez se aprecia una dispersión y error característicos de cada instrumento, y en varias ocasiones, se aprecian observaciones que se alejan considerablemente del resto, entendidas como datos espurios, comúnmente conocidos como *outliers*, pueden verse claramente en las figuras 1.2 y 1.4.

De forma generalizada se aprecian huecos de tiempo irregulares en las observaciones y en ocasiones una cantidad muy reducida de mediciones (Fig. 1.4).

Como es inherente de toda observación, las precisiones de los equipos de medición son limitadas, al igual que su disponibilidad, por lo que es frecuente encontrarse datos espurios, sesgos, huecos de datos irregulares y momentos

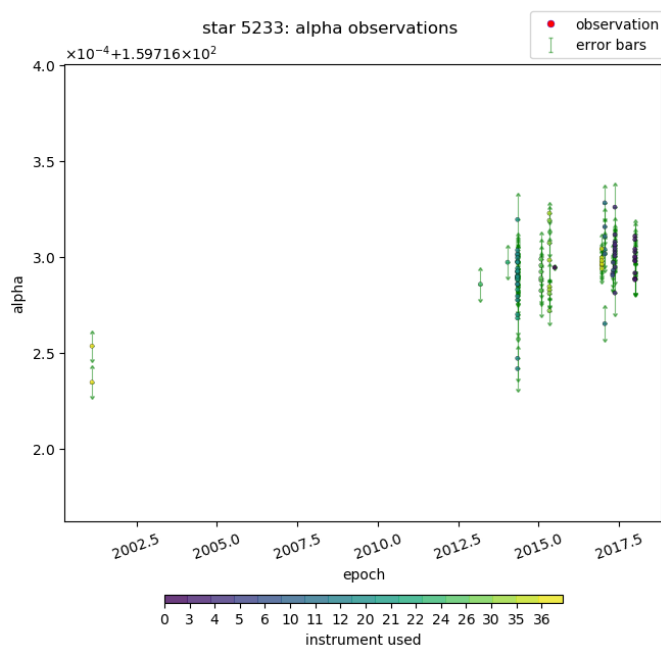


Figura 1.1: Posiciones de una estrella en ascensión recta (eje Y) a lo largo del tiempo (eje X). Colección: muestra de DANCe.

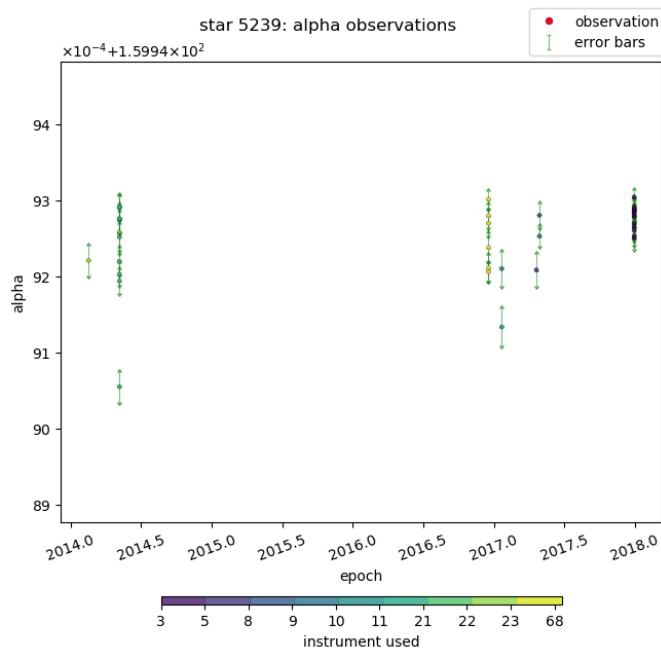


Figura 1.2: Posiciones de una estrella en ascensión recta (eje Y) a lo largo del tiempo (eje X). Colección: muestra de DANCe.

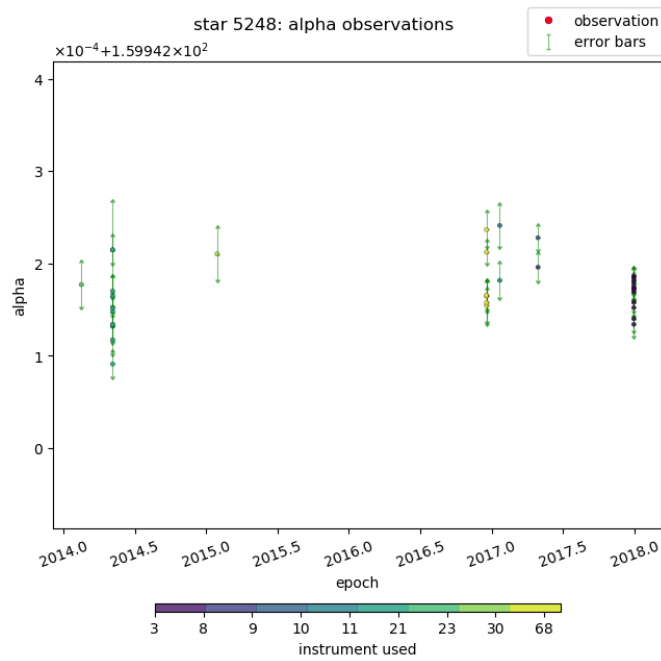


Figura 1.3: Posiciones de una estrella en ascensión recta (eje Y) a lo largo del tiempo (eje X). Colección: muestra de DANCe.

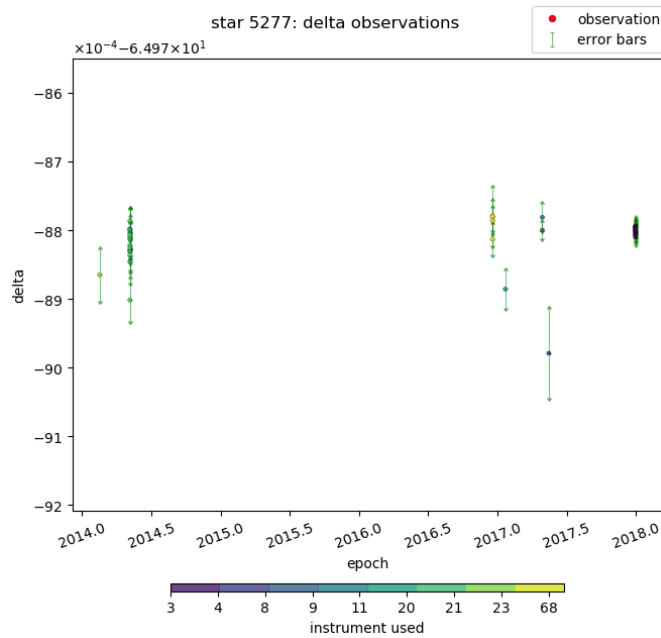


Figura 1.4: Posiciones de una estrella en declinación (eje Y) a lo largo del tiempo (eje X). Colección: muestra de DANCe.

de mayor concentración de datos. Estos factores son un inconveniente para el tratamiento de los datos y muchas técnicas de modelado son sensibles a ellos. Como se presentará a continuación, en este estudio se hace uso de modelos bayesianos por su robustez ante la presencia de este tipo de irregularidades en los datos.

1.5. Modelos bayesianos

En cualquier problema real donde se realice un estudio en detalle, intervienen numerosas variables para conformar un resultado y conviene analizar la probabilidad de que el resultado esté condicionado por cada variable. La teoría bayesiana ofrece un marco de trabajo para cuantificar dichas probabilidades. Son cada vez más las disciplinas de estudio en las que se aplican las técnicas bayesianas para inferir el conocimiento a partir de los datos.

Es necesario aclarar el concepto de probabilidad, para evitar la confusión con el usado por el paradigma frecuentista. El Teorema de Bernoulli puntualiza la distinción, y a su vez relación, entre la probabilidad de que ocurra un suceso en un único experimento, llámese p y la frecuencia relativa de numerosas repeticiones del experimento, la cual converge en p (Loredo, 1990). Estas interpretaciones de la probabilidad son adoptadas por los paradigmas bayesiano y frecuentista respectivamente. En el primero se entiende como probabilidad la certeza de que ocurra un evento, sin necesidad de reproducirlo, en el segundo se refiere a la probabilidad como la citada frecuencia de que ocurra el evento.

El paradigma bayesiano ofrece un planteamiento de la solución que resulta natural en el entorno científico, se proponen hipótesis que satisfagan las observaciones. La falta de datos es una característica más del problema a modelar. En el cálculo bayesiano intervienen todas las observaciones, formando parte de la hipótesis incluso la presencia de aquellas observaciones espurias. Los datos no mienten, son la realidad observacional, lo que sí cambian son nuestras hipótesis sobre lo que esperamos de ellos. Si tienen desviaciones, espurios (*outliers*), sesgos (*offsets*) o huecos, son consecuencias inherentes de la observación, forman parte de la naturaleza del experimento, y por tanto deben incorporarse a las hipótesis.

Los métodos bayesianos obtienen sus soluciones iterando sobre el espacio de parámetros de las hipótesis, a diferencia de los métodos del paradigma

frecuentista, los cuales se basan en la iteración sobre un espacio de datos hipotéticos (Loredo, 1992).

Diseñando una hipótesis adecuada, se permite considerar la presencia de los sesgos no controlados y espurios. Estas hipótesis, pueden expresarse en espacios de parámetros de muy alta dimensionalidad, implicando un alto coste computacional. El desarrollo de técnicas de muestreo avanzadas como las MCMC supone un importante empuje para el uso de técnicas bayesianas al realizar un muestreo eficiente del espacio de parámetros.

Habitualmente se genera confusión al referirse a un modelo por no aclararse si se está haciendo referencia a la función de probabilidad, a la expresión definida por los parámetros que se desean inferir, o a los valores que pueden tomar los parámetros que la definan. En este estudio se refiere como modelo bayesiano a la forma de cálculo de la probabilidad y como modelo a la expresión a inferir, definida por un conjunto de parámetros. Se habla de caracterización de un modelo para referirse a unos valores concretos que puedan tomar sus parámetros.

Los modelos bayesianos buscan obtener la probabilidad de que un modelo tome unos valores que lo hagan encajar con los datos extraídos en la observación. Ésta es la conocida como probabilidad a posteriori de que dados los datos, D , los n parámetros $\theta = (\theta_1, \theta_2, \theta_3, \dots, \theta_n)$ que caracterizan un modelo M obtengan unos valores específicos. Para obtenerla, valiéndose de la relación establecida por el teorema de Bayes (Ec. 3.5), se multiplica la verosimilitud de los datos por el conocimiento a priori, definido por los *priors*, y se normaliza entre la distribución de probabilidad de los datos D . Estos conceptos se profundizan en la sección 3.1.1.

Estos modelos bayesianos presentan varios beneficios. Permiten realizar la comparación de modelos de forma sencilla. Proporcionan mejores soluciones en espacios muestrales pequeños (Ruiz-Benito et al., 2018). Permiten aplicar conocimiento a priori sobre el problema y en los determinados modelos bayesianos jerárquicos, evaluar distintos niveles de variación, como se verá en la Sección 3.1.2. Son capaces de inferir modelos complejos en espacios de parámetros de alta dimensionalidad relacionados, son robustos ante la falta de datos (Bolker, 2008), soportan la presencia de espurios (Kruschke, 2013) y permiten realizar de forma eficiente la agregación de datos de distintas fuentes (Clark, 2005). También proporcionan un resultado completo al dar la distribución de probabilidad a posteriori sobre todo el espacio de

parámetros, no se limitan a concluir una única caracterización óptima del modelo debido a la falta de precisión y desperdicio de información que eso supone.

En el paradigma frecuentista la validación de un modelo se realiza individualmente, no se requieren alternativas. En cambio en el paradigma bayesiano, la validación de un modelo se realiza por comparación con otro alternativo, mediante el factor de Bayes. La solución que aporta sigue la regla de la Navaja de Ockham; modelos simples son preferidos a menos que los complejos sean considerablemente mejores (Loredo, 1990). La comparación de modelos nos permite confrontar un modelo en el que las estrellas tengan movimiento propio frente a otro en el que no lo tengan.

Estos beneficios han llevado a la utilización de técnicas bayesianas en el presente estudio, donde se desea hacer la comparación de modelos inferidos sobre observaciones que por su naturaleza y cuestiones técnicas presentan sesgos y espurios, distribuidos de forma no equitativa a lo largo del tiempo. La proliferación de su uso en astrofísica ratifica su conveniencia para el análisis de datos observacionales, caracterizados por estas irregularidades.

1.6. Estructura del documento

Se incluye una breve descripción de los capítulos del trabajo.

Capítulo 1. Introducción. Este capítulo introduce los principales motivos y objetivos que han llevado a la realización de este trabajo. Se incluyen los conceptos básicos que engloba el trabajo, la importancia de los movimientos propios en astronomía, los detalles del proyecto Cosmic-DANCe y las bondades de los modelos bayesianos en el proceso de inferencia.

Capítulo 2. Estado del arte. Este capítulo describe los pasos realizados hasta el momento en la medición de movimientos propios. Se muestran las técnicas actuales más utilizadas para la caracterización de los movimientos propios, así como sus debilidades. Se analiza la evolución de los modelos bayesianos jerárquicos y su papel en la astrofísica.

Capítulo 3. Modelo bayesiano para medición de movimientos propios. En este capítulo se describe en profundidad el fundamento teóri-

co bayesiano y el modelo bayesiano propuesto para la medición de los movimientos propios.

Capítulo 4. Entorno de desarrollo. En este capítulo se describen las herramientas hardware y software utilizadas para el desarrollo del trabajo y se dan detalles sobre la implementación.

Capítulo 5. Evaluación. Este capítulo detalla la metodología utilizada para evaluar el rendimiento del programa desarrollado, las métricas y colecciones utilizadas en el análisis de resultados, así como las gráficas utilizadas para la representación de los mismos.

Capítulo 6. Discusión. Este capítulo presenta, analiza y discute en profundidad los resultados obtenidos con base en la metodología de evaluación presentada en el capítulo anterior.

Capítulo 7. Conclusiones y trabajo futuro. Este capítulo recopila las diferentes conclusiones extraídas del trabajo realizado y propone algunas líneas de trabajo futuro.

Capítulo 2

Revisión de la bibliografía

Se estudian por un lado las técnicas actuales de medición de movimientos propios y por otro la introducción del paradigma bayesiano en el campo de la astrofísica.

2.1. Estimación de movimientos propios

En la correcta estimación de movimientos propios intervienen numerosos factores desde la componente instrumental en la medición (englobada en los citados *offsets*) hasta la técnica de evaluación de las posiciones para la extracción de la estimación, pasando ineludiblemente por la agregación de distintas fuentes y calibraciones. Todos estos factores se diseñan minuciosamente para minimizar la incertidumbre sobre las estimaciones finales y cada uno de ellos abarca todo un campo de investigación. Este estudio se centra en el procesado de las posiciones para la inferencia de los movimientos propios, sin entrar en detalles sobre el trabajo previo de tratamiento de datos.

Tradicionalmente en astronomía, las técnicas utilizadas para la estimación de movimientos propios se limitan a la búsqueda de mínimos cuadrados y su caso particular de estimación de mínimo χ^2 , al asumir que los errores de los datos siguen una distribución Gaussiana, análogo a maximizar la verosimilitud. No es hasta este mismo año cuando se han observado el uso de técnicas bayesianas para su estimación.

Se han consultado los métodos de estimación de los movimientos propios en diversos catálogos astronómicos encontrados en la literatura.

En los catálogos Hipparcos y Tycho-1 ([Perryman et al., 1997](#)) y UKIDSS

(Lodieu et al., 2012) se utilizan mínimos cuadrados ponderados para la estimación de los parámetros de los modelos, en el primer caso ponderados por una matriz de covarianzas y en el segundo en función del brillo.

En el catálogo Tycho-2 (Høg et al., 2000b) los movimientos propios se estiman con mínimos cuadrados ponderados y se hace una purga de espurios. Para las medidas de la mayoría de las fuentes, los pesos de cada medida son la inversa de su varianza, en otros casos su desviación estándar (Høg et al., 2000a).

Los catálogos PPM (Positions and Proper Motions) (Röser and Bastian, 1991) y PPMX (Positions and Proper Motions Extended), utilizan mínimos cuadrados ponderados para la estimación de los movimientos propios, con el peso de cada medida siendo la inversa de la varianza (Röser et al., 2008). Se realiza una eliminación automática de espurios en base a los residuos normalizados.

En el citado estudio de Bouy et al. (2013) del proyecto DANCe, se hace uso del software SCAMP para obtener la solución astrométrica, el cual hace el cálculo de movimientos propios por minimización de χ^2 . En el cálculo se combinan observaciones de distintos instrumentos y se incluyen sus distorsiones instrumentales (Bertin, 2006).

Aparte de los catálogos habituales, investigaciones independientes sobre los movimientos propios de las estrellas también recurren al método de mínimos cuadrados para la estimación de sus parámetros, como puede observarse desde investigaciones maduras como la de Lyne et al. (1982) hasta las más recientes como Brown et al. (2015).

Se pueden encontrar soluciones análogas, al buscar la optimización del logaritmo de la verosimilitud. Por ejemplo, el desarrollo de Lang et al. (2009). La implementación considera varias estrellas a la vez en la optimización. No obstante no considera la presencia de información a priori y requiere que los datos espurios sean previamente identificados para su tratamiento.

La presencia de técnicas bayesianas en la astronomía suponen una alternativa que está ganando gran protagonismo, como se profundizará en el siguiente apartado. En los últimos años se han implementado modelos de inferencia bayesiana en el contexto del proyecto DANCe para la clasificación de las estrellas del catálogo Pleiades (Olivares et al., 2018).

Durante la escritura del presente estudio se ha publicado una extensión del catálogo GAIA, GPS1+ (Tian et al., 2020), donde el cálculo de los

movimientos propios se realiza combinando la astrometría de los catálogos Gaia, PS1, SDSS y 2MASS. Sobre este conjunto, extraen los movimientos propios con métodos de mínimos cuadrados, en línea con la técnica de cálculo en GPS1. Además, en aquellos casos que disponen de fuentes de Gaia con movimientos propios, se utilizan éstos como información a priori para la inferencia de los movimientos propios empleando modelos bayesianos. De esta manera se recalculan los movimientos propios del 40% de GPS1.

Se utiliza un modelo lineal a comparar con los datos fuente sometidos a corrección de sesgo, la información a priori que se incluye en el modelo bayesiano se reduce a establecer la probabilidad a priori de la pendiente como la resultante de una gaussiana de media μ_{gaia} y varianza ϵ_{gaia} , siendo estos parámetros la media de movimientos propios estimados para cada fuente en Gaia y su error respectivamente. El modelo generativo utilizado no considera la presencia de espurios, en una fase previa se realiza una purga de los datos.

El artículo de [Tian et al. \(2020\)](#) utiliza la misma idea que se propone en el presente estudio pero difiere en la implementación, dado que se espera dar una solución en un espacio multivariable que considere espurios y sesgos no controlados, haciendo que el modelo generativo sea distinto. Los detalles de la implementación se incluyen en la Sección 3.

Las técnicas de mínimos cuadrados se fundamentan en minimizar el valor acumulado del cuadrado de los errores entre la estimación y la referencia. En un ajuste de unos datos $D = \{x_i, y_i\}_{i=1}^N$, a un modelo lineal $f_\theta(x_i) = m \cdot x_i + b$, donde $\theta \equiv \{m, b\}$, siendo m la inclinación y b la intersección con el origen, se busca la combinación de valores de esos parámetros θ que acumule un menor error evaluado sobre los datos D :

$$\theta = \arg \min_{\theta} \sum_{i=1}^N w_i (f_\theta(x_i) - y_i)^2 \quad (2.1)$$

w_i representa el peso de cada medida, el uso de estos pesos da lugar a la técnica de mínimos cuadrados ponderados. La técnica que excluye este término es una formulación particular de la Ec. 2.1 donde $w_i = 1 \quad \forall i \in \{1, \dots, N\}$. Este término habitualmente es la inversa de la desviación estándar de la medida, considerando de este modo las incertidumbres particulares de cada medida. Para su resolución se utilizan técnicas analíticas, el error $(f_\theta(x_i) - y_i)$ se deriva entre cada parámetro y se iguala a 0, obteniendo un sistema de ecuaciones a resolver, habitualmente de manera matricial, donde

el resultado son unos valores únicos de los parámetros. La minimización de χ^2 es equivalente siendo el peso de cada medida su varianza, la cual debe seguir una distribución gaussiana. La maximización de la verosimilitud es proporcional a la minimización de la estimación de χ^2 (Hogg et al., 2010)

La contrapartida fundamental de estas técnicas de estimación, es que dan una estimación correcta solo si el modelo propuesto se aproxima lo suficiente a la realidad. De no ser así, estamos ante un caso de mala especificación y la estimación puede ser incorrecta dando lugar a falsas interpretaciones. En ese caso las técnicas de comparación entre caracterizaciones del modelo como el ratio de verosimilitudes no son válidas (Feigelson and Babu, 2012b) y por ende la comparación entre modelos.

Este problema puede verse acentuado en situaciones donde la proporción entre el volumen de datos y el número de parámetros sea baja. Debido a la presencia de los errores en la medición, cuanto más baja sea esta proporción, mayor será la desviación que habrá entre la estimación y la realidad (Glusker et al., 1995). Poniendo por ejemplo la estimación del presente estudio, la presencia de errores unido al bajo volumen de datos puede hacer que la estimación sugiera movimientos propios siendo falso. Por otro lado, son sensibles a la presencia de datos espurios, requiriendo comúnmente una purga o un descarte de valores extremos como se ha visto a lo largo de los catálogos consultados.

El uso del marco bayesiano permite la comparación de modelos, de hecho es un requerimiento intrínseco para la valoración de un modelo y permite la inferencia de modelos complejos ante presencia de pocos datos y con irregularidades (sesgos y espurios). A continuación se analiza su evolución e introducción en la astrofísica.

2.2. Inferencia bayesiana

El paradigma bayesiano se basa en la Teoría Bayesiana de Probabilidad, introducida por Bayes (1763) y desarrollada por el astrofísico Laplace en Laplace (1820). A pesar de ello su uso en la evaluación científica, particularmente en la astrofísica, ha tardado en manifestarse.

El uso de estas técnicas comenzó a experimentar un creciente protagonismo en la astrofísica hace ya 40 años (Loredo, 1990) pero ha sido en los últimos 20 años cuando se ha incrementado con rapidez. Sin embargo, este

paradigma sigue siendo altamente infrautilizado, y en muchas de ocasiones en las que se utiliza, los modelos son muy simples (Loredo, 2013).

Durante mucho tiempo, este paradigma, ha generado rechazo en científicos de diversas disciplinas, debido posiblemente a varias razones entre las que se distinguen razones conceptuales y técnicas.

Por un lado, desde el punto de vista conceptual, la correlación entre la observación científica, con su inherente repetitividad de los experimentos, y la frecuencia de un acontecimiento, ha propiciado la adopción de la concepción frecuentista de la probabilidad para el análisis de sus resultados.

La confusión en el distinto tratamiento del concepto de probabilidad ha llevado a muchos científicos a considerar erróneamente el paradigma bayesiano como una forma más de calcular lo mismo, añadiendo simplemente la modulación de los priors. Mientras que la diferencia entre los dos paradigmas es radical, se calculan conceptos distintos de manera distinta.

La probabilidad frecuentista es aquella frecuencia relativa del cumplimiento a la que tiende el experimento tras infinitas repeticiones. A la hora de proporcionar la probabilidad de una hipótesis, se realizan numerosas repeticiones del experimento hasta obtener una confianza de que la frecuencia con la que ocurre un evento, o no, se estabiliza. Requiere de la existencia de un número finito de observaciones para establecer la probabilidad de cada observación por comparación de su frecuencia con las demás. Es un valor que varía conforme se evoluciona en la experimentación. Este concepto de probabilidad parece contrario a la intuición a la hora de hablar de una hipótesis que representa el mundo real. Ya que una hipótesis será cierta o falsa, o mejor dicho, tendrá un cierto grado de plausibilidad, pero siempre el mismo, no dependerá del número de repeticiones que se hayan hecho del experimento. El grado de plausibilidad de una hipótesis per se, es precisamente el concepto de probabilidad que adopta el paradigma bayesiano.

En el paradigma frecuentista, generalmente el resultado vienen dado por la máxima verosimilitud, mientras que en el paradigma bayesiano viene dado por la distribución de probabilidad a posteriori, la cual depende de la verosimilitud, es un término más de su cálculo. En problemas simples, con errores gaussianos, y con priors no informativos, el resultado de ambas vertientes es similar (VanderPlas, 2014). Sin embargo, la diferente asunción de la probabilidad les lleva a calcular factores distintos. En el paradigma frecuentista se cuantifican las probabilidades de las observaciones ante unos parámetros

fijos de un modelo. Mientras que en el bayesiano se cuantifican las probabilidades de parámetros desconocidos de un modelo ante unos datos observados fijos.

En el paradigma bayesiano, la probabilidad indica el grado de certeza sobre una hipótesis, por mucho que la probabilidad de los datos dada una hipótesis sea un valor real bajo, individualmente no permite concluir que sea falsa, requiere la comparación con otras hipótesis para las cuáles la probabilidad sea *suficientemente mayor*. En muchas hipótesis el espacio de muestreo es tan grande que todos los eventos tienen baja probabilidad (Loredo, 1994).

La distinción de si una hipótesis es *suficientemente mayor* la marca la información a priori, siendo por tanto un requisito indispensable en la estimación (Loredo, 1992).

Los priors no solo modulan la verosimilitud, permiten reportar la probabilidad de una distribución de modelos a partir de las verosimilitudes de cada uno. “Priors are important, not so much as modulators of likelihoods, but as converters from intensities (likelihoods) to measures (probabilities)” (Loredo, 2013).

De esta manera, el resultado de los métodos bayesianos, no da una respuesta única al problema, una hipótesis verdadera única, sino que resulta una distribución de probabilidad de las hipótesis, lo que muchos ven como una contrapartida, pero los bayesianos defienden por asemejarse en mayor medida a la realidad experimental.

Por otro lado, también se encuentran dificultades desde el punto de vista técnico. Tal y como se ha introducido, el cálculo bayesiano implica la búsqueda en el espacio de hipótesis, los distintos valores que pueden tomar los parámetros que identifican al modelo. Este espacio de búsqueda puede ser muy amplio, de manera que la obtención de la distribución de probabilidad a posteriori, que implica el cálculo de la integral sobre los parámetros, suponga un gran esfuerzo computacional.

El desarrollo de técnicas de muestreo de Markov Chain Monte Carlo permite abordar el cálculo integral de una distribución con un reducido coste computacional, lo cual ha sido aprovechado por las técnicas bayesianas para el cálculo de la probabilidad a posteriori. Esta integración ha aliviado las complicaciones técnicas abriendo las puertas a una auténtica revolución en la ciencia de tratamiento de datos (Sharma, 2017).

En la búsqueda de una justificación convincente de la conveniencia del

paradigma bayesiano, en 1939, Harold Jeffrey presenta una serie de soluciones prácticas a problemas estadísticos (Jeffreys, 1998), sentando las bases de la inferencia científica fundamentada en ideas estadísticas bayesianas. Esta conveniencia se ve reforzada durante las siguientes décadas, entre otros, por los trabajos de R.T.Cox y E.T.Jaynes (Loredo, 1990), haciendo reavivar el interés de la astrofísica en estas técnicas.

Estos estudios han permitido reducir las barreras conceptuales y acelerar la aceptación práctica del paradigma bayesiano, hasta el punto de ser uno de los principales temas presentes en la conferencia *Statistical Challenges in Modern Astronomy V* (Feigelson and Babu, 2012a). Con dos sesiones dedicadas a la inferencia bayesiana tanto en SCMA V como en SCMA VI. Dentro de la astrofísica, sus ramas de aplicación son muy diversas, consultando por ejemplo las presentaciones de SCMA V, se encuentran trabajos fundamentados en el paradigma bayesiano que van desde la modelización de la formación de galaxias (Lu et al., 2012) hasta la medición de astrosismología (Benomar, 2012), mapeado de la reverberación en núcleos galácticos activos (Brewer, 2012) o identificación de fuentes cruzadas (Budavári, 2012). Muchos de los artículos que se encuentran en (Feigelson and Babu, 2012a) utilizan métodos jerárquicos bayesianos y técnicas de MCMC.

En base a las ventajas destacadas de las técnicas bayesianas en la Sección 1.5 y a los precedentes citados se adopta la búsqueda de un modelo bayesiano jerárquico, robusto frente a sesgos y espurios, para la estimación de los movimientos propios por comparación entre modelos con movimiento y sin movimiento. Para abordar las limitaciones computacionales y permitir una elevada escalabilidad de los cálculos la implementación se realiza sobre GPU. Los detalles del modelo teórico se explican en la Sección 3 y los detalles de la implementación técnica en la Sección 4.

Capítulo 3

Modelo bayesiano jerárquico para inferencia de movimientos propios

En este capítulo se describe en profundidad el fundamento bayesiano y se detallan las particularidades del modelo bayesiano jerárquico implementado.

3.1. Introducción a la inferencia bayesiana

En esta sección se introducen los fundamentos de la inferencia bayesiana y los modelos bayesianos jerárquicos.

3.1.1. Fundamentos de la inferencia bayesiana

La teoría bayesiana de la probabilidad se fundamenta en dos axiomas, la regla de la suma (Ec. 3.1) y la del producto (Ec. 3.2).

$$P(A|C) + P(\bar{A}|C) = 1 \quad (3.1)$$

$$P(AB|C) = P(A|BC) \cdot P(B|C) \quad (3.2)$$

A partir del segundo axioma y dado que el orden de una operación lógica no altera su resultado, $P(AB) = P(BA)$, se obtiene la relación conocida como el Teorema de Bayes (Ec. 3.3). Se adecua la nomenclatura al caso que nos ocupa para ir introduciendo el problema. Se consideran dos sucesos

aleatorios que son la tenencia de unos datos, D , y la propuesta de una hipótesis, H . Dicho teorema establece que dados los dos sucesos aleatorios:

$$P(H|D) \cdot P(D) = P(D|H) \cdot P(H) \quad (3.3)$$

Siendo

$P(H|D)$ Probabilidad de H dado D .

$P(D)$ Probabilidad de D .

$P(D|H)$ Probabilidad de D dado H .

$P(H)$ Probabilidad de H .

La hipótesis H engloba el modelo M , determinado por m parámetros referidos como $\theta = \{\theta_0, \theta_1, \dots, \theta_m\}$, y la información a priori del problema I , que incluye las incertidumbres, los instrumentos utilizados para las observaciones, conocimiento sobre datos espurios y cualquier otro aspecto que influya en el problema.

El objetivo es conocer la probabilidad a posteriori de que un modelo propuesto M caracterizado por θ sea cierto dadas las observaciones D y la información a priori del problema. A partir del teorema de Bayes su expresión atiende a la Ecuación 3.5.

La comparación entre distintos modelos se realiza como fase final del algoritmo del presente estudio, en las explicaciones próximas se asume que el modelo es fijo y se hace alusión a los parámetros θ asumiendo que corresponden al modelo M . Se hace distinción entre los parámetros del modelo θ y el resto de información del problema I .

$$\begin{aligned} P(\theta, I, D) &= P(\theta|D, I)P(D, I) \\ &= P(D|\theta, I)P(\theta, I) \end{aligned} \quad (3.4)$$

$$P(\theta|D, I) = \frac{P(D|\theta, I) \cdot P(\theta, I)}{P(D, I)} = \frac{P(D|\theta, I) \cdot P(\theta|I) \cdot P(I)}{P(D|I) \cdot P(I)} \quad (3.5)$$

Siendo

$P(\theta D, I)$	Distribución de probabilidad a posteriori de los parámetros del modelo θ dados los datos D y el conocimiento del problema I . Conocida como <i>full a posteriori</i> .
$P(D I)$	Distribución de probabilidad de los datos D , dado el conocimiento del problema I . Conocida como evidencia.
$P(D \theta, I)$	Distribución de probabilidad de los datos D dados los parámetros del modelo θ y el conocimiento del problema I . Conocida como verosimilitud.
$P(\theta I)$	Distribución de probabilidad a priori de los parámetros del modelo θ , dado el conocimiento del problema I .
$P(I)$	Distribución de probabilidad a priori del conocimiento del problema I .

La verosimilitud, \mathcal{L} , es la probabilidad de medir los datos observados dado un conjunto θ de parámetros de M . Los priors representan el conocimiento a priori del problema, la probabilidad que un parámetro tome un determinado valor, independientemente de los datos del problema.

La evidencia es el resultado de la marginalización del numerador de la Ec. 3.5 respecto a todo el espacio de parámetros de los modelos. En la estimación de parámetros, juega un papel meramente normalizador, siendo prescindible en la mayoría de ocasiones. Debido a esto, el numerador de la Ec. 3.5 se conoce como la distribución de probabilidad full a posteriori sin normalizar. Haciendo la distinción con la distribución de probabilidad full a posteriori marginalizada, aquella que ha sido normalizada entre la evidencia. Se incluye *full* para indicar que relaciona a todos los parámetros del problema.

En muchas presentaciones se omite el término referente al conocimiento del problema I . En otras se hace referencia a él, pero no se muestra su probabilidad a priori. Resumiendo la ecuación Ec. 3.5 en:

$$P(\theta|D, I) = \frac{P(D|\theta, I) \cdot P(\theta|I)}{P(D|I)} \quad (3.6)$$

Por mucho que se omita, siempre está presente, cada experimento tiene unas características inherentes que se engloban en este conocimiento.

A no ser que el prior tenga gran variación en la región de interés, la

verosimilitud será muy similar a la distribución de probabilidad a posteriori. En caso de no tener conocimiento sobre la probabilidad a priori, se asigna un valor no informativo en el rango explorado, por ejemplo una distribución uniforme. En ese caso, la solución bayesiana es proporcional a la verosimilitud (VanderPlas, 2014). En esta asociación se basa el paradigma frecuentista para reducir su análisis a la búsqueda de la máxima verosimilitud, ignorándose los beneficios del paradigma bayesiano presentados en la introducción.

No obstante, se debe alertar de que el uso de priors no informativos no está recomendado, en caso de no satisfacer que la integral en el espacio de los priors sea un número finito, pueden llevar al muestreador MCMC a muestreos sesgados (Tak et al., 2018).

Por ello, en caso de no poder determinar los priors por falta de conocimiento preliminar, conviene aplicar técnicas que modelen los priors con utilidad. Con esta premisa se hace uso de los modelos bayesianos jerárquicos.

3.1.2. Fundamentos de los modelos bayesianos jerárquicos

Un modelo se define jerárquico cuando los priors que modulan la distribución de probabilidad no son fijos, si no que dependen a su vez de otra distribución con unos determinados parámetros que pueden seguir sus propios priors. Estos parámetros son conocidos como hiperparámetros y sus priors como hiperpriors.

Partiendo de la Ec. 3.6, y suponiendo un modelo definido por un parámetro θ_1 , se obtiene la expresión detallada Ec. 3.7

$$P(\theta_1|D, I) = \frac{P(D|\theta_1, I) \cdot P(\theta_1|I)}{P(D|I)} \quad (3.7)$$

$P(\theta_1|I)$ es el prior de θ_1 . En los modelos bayesianos sencillos, este prior generalmente es una distribución fija, ya sea no informativa (por ejemplo una distribución uniforme en el rango esperado de valores $[a, b]$, $P(\theta_1|I) \sim U(a, b)$), o basado en otra evidencia (por ejemplo en otro estudio que ha evidenciado que sigue una distribución normal con unos valores específicos μ_{θ_1} y σ_{θ_1} , $P(\theta_1|I) \sim \mathcal{N}(\mu_{\theta_1}, \sigma_{\theta_1})$).

El uso de priors no informativos puede dar lugar a distribuciones a posteriori inadecuadas (Hobert and Casella, 1996; Carlin and Louis, 2010). Siem-

pre se sabe algo sobre el problema como para limitar los priors en un rango que descarte valores extremos. Un problema bien formulado requiere añadir suficiente información para evitar asignaciones ambiguas en el cálculo (Loredo, 1990).

Estos priors pueden depender de una distribución de probabilidad caracterizada por unos parámetros variables μ_v y σ_v . De esta manera el prior de θ_1 no se define por una distribución de probabilidad con parámetros fijos, si no que ésta está supeditada a los valores de μ_v y σ_v . Por ejemplo, el prior puede seguir una distribución normal con parámetros variables, $P(\theta_1|I) \sim \mathcal{N}(\mu_v, \sigma_v)$ donde μ_v siga una distribución binomial $\mu_v \sim \mathcal{B}(10, 0,5)$ y σ_v una normal $\sigma_v \sim \mathcal{N}(0, 1)$. En este ejemplo, μ_v y σ_v son hiperparámetros, parámetros variables que modulan el prior de θ_1 y las distribuciones de probabilidad $\mathcal{B}(10, 0,5)$ y $\mathcal{N}(0, 1)$ son sus respectivos hiperpriors, las distribuciones que modulan los hiperparámetros.

El caso anteriormente presentado de GPS1+ (Tian et al., 2020) es un ejemplo de modelo bayesiano jerárquico donde el prior de la inclinación del movimiento propio μ proviene de una distribución gaussiana $\mu \sim \mathcal{N}(\mu_{gaia}, \epsilon_{gaia})$, cuyos hiperparámetros μ_{gaia} y ϵ_{gaia} toman valores en función de la fuente de observación, estimados en experimentos previos de Gaia.

Esta técnica permite hacer una evaluación indirecta del espacio de parámetros, estableciendo distintos niveles de variabilidad, de forma que la búsqueda de los priors adecuados se integra como parte del propio proceso de inferencia de la probabilidad a posteriori. De este modo los priors son poco restrictivos sin llegar al extremo de que sean no informativos.

3.2. Aplicación a la inferencia de movimientos propios a partir de medidas astrométricas

Introducido ya el concepto de inferencia bayesiana y los modelos bayesianos jerárquicos, se profundiza en el modelo bayesiano desarrollado en el presente estudio. Para terminar esta sección se incide sobre el muestreador utilizado para la estimación de las probabilidades.

3.2.1. El modelo bayesiano jerárquico para movimientos propios

Anteriormente ya se ha indicado que el movimiento propio de cada eje se busca individualmente. Así pues, se muestra el desarrollo para la estimación del movimiento propio de una estrella en el eje de ascensión recta, α , siendo igualmente válido para la declinación δ .

Supongamos que los datos D están compuestos por N tríos de medidas $d_i = \{\alpha_i, \sigma_{\alpha_i}, t_i\}$ que representan la posición en el eje de ascensión recta, el error e instante de tiempo en la medición respectivamente. Se propone un modelo lineal con los parámetros, pendiente, m , e intersección con el origen, b .

$$\alpha_i = m \cdot t_i + b \quad (3.8)$$

Reformulando la Ec. 3.5 con $\theta \equiv \{m, b\}$ y $D \equiv \{\alpha_i\}_{i=1}^N$:

$$P(\theta | \{\alpha_i\}_{i=1}^N, I) = \frac{P(\{\alpha_i\}_{i=1}^N | \theta, I) \cdot P(\theta | I)}{P(\{\alpha_i\}_{i=1}^N | I)} \quad (3.9)$$

Siendo

I	Conocimiento a priori, es decir, $\{t_i\}_{i=1}^N$ y $\{\sigma_{\alpha_i}\}_{i=1}^N$.
$P(\theta \{\alpha_i\}_{i=1}^N, I)$	Distribución de probabilidad a posteriori.
$P(\{\alpha_i\}_{i=1}^N m, b, I)$	Verosimilitud.
$P(\theta I)$	Prior.
$P(\{\alpha_i\}_{i=1}^N I)$	Evidencia.

Dado que cada dato se toma independientemente de los demás, la verosimilitud se obtiene por multiplicación de las probabilidades individuales de cada medida, (Ec. 3.10).

$$P(\{\alpha_i\}_{i=1}^N | \theta, I) = \mathcal{L} = \prod_{i=1}^N p(\alpha_i | \theta, I) \quad (3.10)$$

La forma que tome el cálculo de la probabilidad determinará la complejidad del modelo.

La verosimilitud evalúa cómo de plausible es que el modelo propuesto haya generado los datos observados, en este sentido, el uso de estos modelos

bayesianos recibe el nombre de modelos generativos.

Se desea hacer un modelo generativo que sea robusto por un lado ante la presencia de datos espurios y por otro ante la presencia de un sesgo instrumental indeterminado. Para abordar el primer requisito se hace uso de los modelos mixtos, los cuales consisten en la conjunción de distintas distribuciones de probabilidad. Se sigue el modelo generativo propuesto por [Hogg et al. \(2010\)](#). Las distribuciones intervinientes se denominan *foreground*, p_{fg} y *background*, p_{bg} .

La distribución de probabilidad foreground cuantifica la adecuación del modelo lineal propuesto como generador de los datos observados, se incluye en este punto el término de corrección de sesgo de cada medida α_i , a continuación se explica la forma de obtenerlo.

$$p_{fg}(\alpha_i|\theta_{fg}, I) = \frac{1}{\sqrt{2\pi\sigma_{\alpha i}^2}} e^{-\frac{[\alpha_i - mt_i - b - o_i]^2}{2\sigma_{\alpha i}^2}} \quad (3.11)$$

Siendo $\theta_{fg} \equiv \{m, b, o_i\}$ y $I \equiv \{t_i, \sigma_{\alpha i}\}$

La distribución background representa a los datos espurios. Modelada como una gaussiana con media Y_o y desviación estándar un incremento de V_{oi} sobre la desviación estándar de la medida.

$$p_{bg}(\alpha_i|\theta_{bg}, I) = \frac{1}{\sqrt{2\pi[V_{oi} + \sigma_{\alpha i}^2]}} e^{-\frac{[\alpha_i - Y_o]^2}{2[V_{oi} + \sigma_{\alpha i}^2]}} \quad (3.12)$$

Siendo $\theta_{bg} \equiv \{Y_o, V_{oi}\}$ y $I \equiv \sigma_{\alpha i}$

La distribución de probabilidad combinada es una ponderación de cada distribución en función de la probabilidad de espurios, P_o .

$$p(\{\alpha_i\}_{i=1}^N|\theta, I) = \prod_{i=1}^N (1 - P_o) \cdot p_{fg}(\alpha_i|\theta_{fg}, I) + P_o \cdot p_{bg}(\alpha_i|\theta_{bg}, I) \quad (3.13)$$

Siendo, por tanto,

$$\theta \equiv \{m, b, \{o_i\}_{i=1}^N, P_o, Y_o, \{V_{oi}\}_{i=1}^N\}$$

$$I \equiv \{\{t_i\}_{i=1}^N, \{\sigma_{\alpha i}\}_{i=1}^N\}$$

Se debe matizar que la corrección de sesgo aplicada sobre cada medida i no es independiente para cada observación, dependen del instrumento l utilizado para realizar la medición. Esta información se incluye como dato de

entrada de modo que $d_i = \{\alpha_i, \sigma_{\alpha_i}, t_i, l_i\}$ con $l_i \in \{1, \dots, L\} \quad \forall i \in \{1, \dots, N\}$. El número de instrumentos distintos utilizados, L , es considerablemente menor al número de observaciones realizadas con dichos instrumentos, $L \ll N$.

Con este planteamiento se tendría un espacio de parámetros de $N + L + 4$ dimensiones. Lo cual puede llegar a ser muy elevado, y más preocupante, escala con el aumento de los datos. Se hacen una serie de asunciones para reducir la dimensionalidad.

Por un lado, respecto a la distribución de background, no se realiza una estimación de los parámetros. Los valores de P_o y Y_o se establecen por configuración y el valor de V_{oi} se obtiene en función del error de cada medida, $V_{oi} = f(\sigma_i) = 10\sigma_i$. Por lo que se reduce la dimensionalidad del problema a $L + 2$ parámetros, lo cual, a pesar de seguir siendo una alta dimensionalidad, ya no escala con el aumento de datos. Además, con esta asignación se está supeditando la presencia de espurios a los propios datos, y se está particularizando la probabilidad de espurio a la amplitud del error de cada medida individualmente, tratando de minimizar la arbitrariedad en la selección de parámetros.

Por otro lado, en a la distribución de *foreground* se incluye la estimación de los parámetros m y b . Pero además, para abordar el requisito de robustez ante el sesgo, se hace uso del conocimiento del instrumento de medida utilizado. En un nivel de variación superior se obtiene la magnitud del sesgo de cada medida o_i , siguiendo una distribución normal con media 0 y desviación estándar λ_{l_i} , es decir $o_i \sim \mathcal{N}(0, \lambda_{l_i})$.

Para cada instrumento, l , se obtiene un valor lambda, λ_l , aplicando un factor k , sobre el rango del conjunto de observaciones realizadas con dicho instrumento, $range_l$:

$$\lambda_l = k \cdot range_l. \quad (3.14)$$

He aquí la presencia de la jerarquía, λ_{l_i} es un hiperparámetro de la distribución normal, y ésta es el prior del sesgo. Para establecer $range_l$ se obtiene la diferencia entre el valor máximo y mínimo medidos con ese instrumento l , es decir $range_l \sim f(\{\alpha_i\}_{i=1}^N, l)$. El factor k es un valor en el rango $[0, 1]$ que se establece por configuración. Dado que cada observación i ha sido obtenida con uno de los instrumentos l , el prior del sesgo de cada observación sigue una distribución característica para su instrumento l .

De esta manera, los L parámetros del sesgo, se supeditan a los propios

datos. A pesar de desconocer su probabilidad a priori, no se asignan unos priors no informativos, ni unos arbitrarios, en su lugar se configura una distribución de valores esperados. Es decir, la jerarquía permite estimar en un nivel los L parámetros del sesgo y en otro nivel estimar directamente los dos que caracterizan al modelo lineal buscado (m, b) , siendo los restantes calculados (V_{oi}) indirectamente a partir de los datos observacionales o bien prefijados por configuración (P_o, Y_o)

El prior de la Ec. 3.9, con todos los parámetros del modelo a inferir, siendo estos independientes entre sí, se expresa:

$$P(m, b, \{o_i\}_{i=1}^N) = P(m) \cdot P(b) \cdot P(\{o_i\}_{i=1}^N) \quad (3.15)$$

Siendo

$P(m)$	Una distribución $\mathcal{N}(0, \sigma_m)$.
$P(b)$	Una distribución $\mathcal{U}(-5, 5)$.
$P(\{o_i\}_{i=1}^N)$	Una distribución $\mathcal{N}(0, \lambda_{l_i})$ para cada i .

Donde σ_m se establece por configuración y λ_{l_i} se determina por la Ec. 3.14 en función del instrumento utilizado para cada medición l_i .

Una representación del modelo bayesiano jerárquico según los estándares de la literatura puede verse en la figura 3.1. Por un lado están los parámetros “de la población” m y b , a la hora de calcular la verosimilitud de un modelo, estos parámetros se mantienen constantes, son los mismos aplicados sobre todas las observaciones. Por otro lado, el parámetro o_i depende del instrumento con el que se haya obtenido cada observación, por tanto varía a lo largo de la evaluación del conjunto de N observaciones, se denomina parámetro “individual”.

Se incluye también a modo de resumen el cuadro 3.2, que relaciona todos los parámetros intervinientes en el cálculo de la verosimilitud, ya sean parámetros a estimar, los calculados de forma indirecta y los fijados por configuración.

La clave de este modelo generativo está en su capacidad para incluir no solo los datos correctos con sus incertidumbres si no también los datos espurios y la corrección de sesgo, de forma que al reproducir el modelo se generan tanto los datos del modelo lineal con su sesgo como posibles medidas espurias que contengan las observaciones.

Llegado a este punto ya se está en disposición de calcular la probabilidad

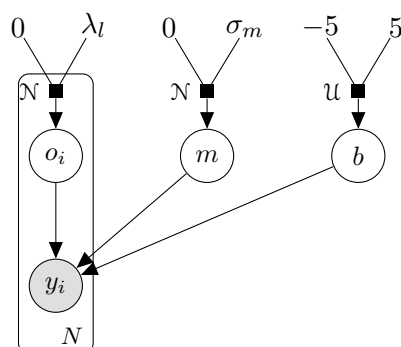


Figura 3.1: Representación gráfica del modelo bayesiano jerárquico propuesto

a posteriori. Sin embargo, debido a una alta dimensionalidad, el cálculo de la evidencia puede presentar un coste computacional intratable. Para abordar este problema entran en escena los métodos de Markov Chain Monte Carlo (MCMC).

3.2.2. Muestreador utilizado para la estimación

Para obtener la evidencia y finalmente la probabilidad a posteriori de una manera eficiente se hace una aproximación mediante técnicas de muestreo, particularmente con las técnicas MCMC.

El método de Monte Carlo se introduce en [Metropolis and Ulam \(1949\)](#) y consiste en realizar el cálculo de la integral de una función considerando solo una muestra aleatoria de puntos del espacio de parámetros de la función que satisfacen una condición de aceptación. Tiene especial interés en el cálculo de integrales multidimensionales que pueden volverse intratables en términos computacionales.

Las cadenas de Markov son métodos estadísticos en los que la probabilidad de que ocurra un evento depende únicamente del inmediatamente anterior, requisito conocido como la propiedad de Markov. Expresado matemáticamente, dada una secuencia de variables aleatorias $X \equiv \{X_0, X_1, \dots, X_n\}$:

$$P(X_n|X_{n-1}) = P(X_n|X_0, X_1, \dots, X_{n-1}) \quad (3.16)$$

X_i y X_j en lugar de variables individuales pueden representar un conjunto de variables $X_i \equiv X_{1i}, X_{2i}, X_{3i}$, asemejándose al caso de este estudio.

La integración de las dos técnicas da lugar a MCMC, una técnica de

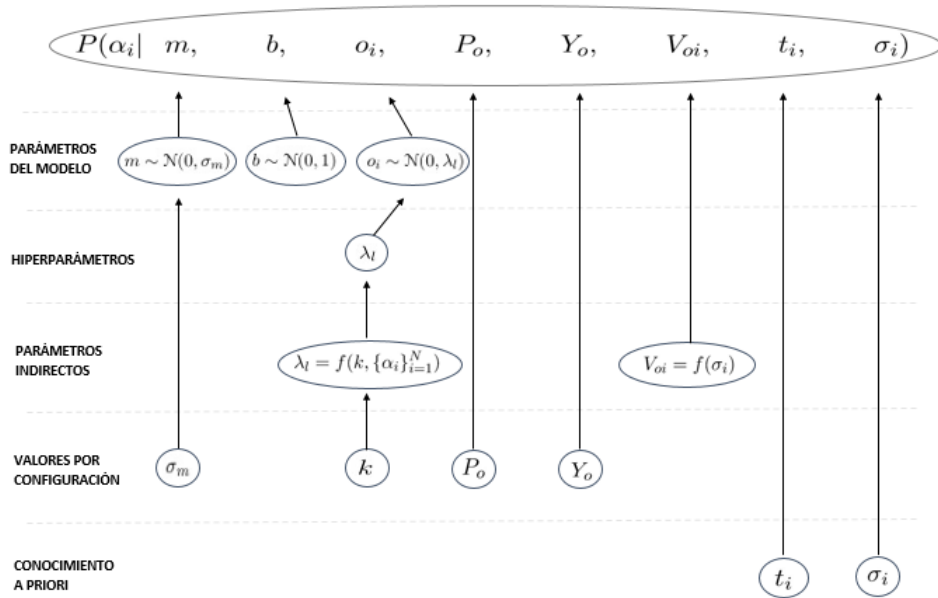


Figura 3.2: Resumen de los parámetros intervinientes en el cálculo de verosimilitud.

muestreo del espacio de parámetros que no busca obtener el mejor punto del espacio (una combinación de parámetros θ) si no obtener una muestra de su distribución, particularmente útil en la búsqueda de la distribución a posteriori de los parámetros que caracterizan al modelo.

En el muestreo de una distribución de probabilidad $P(X)$, Metropolis considera una distribución de probabilidad de transiciones entre un punto y otro $Q(X_j|X_i) \equiv q_{ji}$ que sea simétrica ($q_{ji} = q_{ij}$). Partiendo de un punto inicial X_i , referido como pivote, acepta un nuevo punto X_j , referido como candidato, sujeto a una probabilidad igual al ratio entre las densidades de probabilidad del candidato y del pivote. El ratio permite utilizar funciones proporcionales a la densidad de probabilidad, referidas como $f(X)$, en lugar de la propia $P(X)$ que se desea muestrear.

En su estudio, [Hastings \(1970\)](#) propone una generalización de la idea de Metropolis para casos donde la distribución de transiciones q_{ji} sea asimétrica. Para ello, en el cálculo del ratio de funciones, añade el ratio de probabilidades de transición q_{ji} de la cadena de Markov. Esta generalización da lugar al algoritmo Metropolis-Hastings MCMC (MH-MCMC).

El algoritmo MH-MCMC se compone de los siguientes pasos:

1. Se parte de un punto aleatorio del espacio de parámetros X_i (el pivote)
2. Se selecciona un nuevo punto del espacio de parámetros X_j (el candidato), de acuerdo a la distribución $Q(X_j|X_i)$,
3. Se muestrea un valor U de una distribución aleatoria acotada en $[0, 1]$
4. El pivote se sustituye por el candidato en caso de que $U < \frac{f(X_j) \cdot q_{ij}}{f(X_i) \cdot q_{ji}}$
5. Vuelta al punto 3 hasta que se cumpla una condición de parada

MH-MCMC y la versión inicial simétrica de Metropolis, son los primeros y más extendidos métodos MCMC. A partir de los años 80 empiezan a surgir nuevos algoritmos como el recocido simulado (Kirkpatrick et al., 1983), muestreo de Gibbs (Geman and Geman, 1984), esperanza-maximización (EM) (Dempster et al., 1977) y el de aumento de datos (DA) (Tanner and Wong, 1987). En su estudio, Tanner and Wong (2010) destacan que MH-MCMC y el muestreo de Gibbs son particularmente adecuados para su uso en el paradigma bayesiano (Sharma, 2017).

Distintas formas de acometer cada punto han dado lugares a diversas variantes del algoritmo. Ya sea el distinto modo de muestreo de puntos del espacio de parámetros, tanto los de partida como los candidatos, la forma del ratio de aceptación o la condición de parada. Ésta última es la más variada, bien sea por un número prefijado de iteraciones, una reducción en la diversidad de los nuevos puntos, o un bajo índice de sustitución, entre otros.

En este estudio se ha implementado la propuesta de Metropolis, donde $P(X)$ es la distribución de probabilidad a posteriori que se desea muestrear, $f(X_j)$ es la función de densidad del prior, proporcional a la función de densidad de $P(X)$, que se utiliza para el cálculo del ratio de aceptación y $Q(X_j|X_i)$ es la función de generación de candidatos a partir del incremento (o decremento) de un *paso*. El paso es una cantidad que se autoregula en función del ratio entre aceptaciones y rechazos de candidatos.

3.2.2.1. Nested sampling

En 2006, Skilling sugiere el algoritmo nested sampling (Skilling, 2006), un método basado en MCMC que proporciona una solución optimizada para

la estimación de la evidencia de un problema. Se trata de un factor clave en los problemas de comparación de modelos bayesianos, donde la superioridad de un modelo se determina por el ratio entre la evidencia de dos modelos. Fundamental para el presente estudio en el que se compara un modelo con movimientos propios frente a otro sin movimientos propios.

La evidencia del problema se calcula de la siguiente forma:

$$P(D|I) = \frac{P(D|\theta, I) \cdot P(\theta|I)}{P(\theta|D, I)} \quad (3.17)$$

Marginalizando sobre todo el espacio de parámetros θ se obtiene la Ec. 3.18.

$$P(D|I) = \int P(D|\theta, I) \cdot P(\theta|I) d\theta \quad (3.18)$$

La exploración de todo el espacio de parámetros en problemas de alta dimensionalidad como el del presente estudio puede suponer de un excesivo coste computacional. Skillings propone el muestreo sobre el espacio de probabilidades a priori, con una restricción basada en el valor de verosimilitud.

Se realiza un intercambio de variable para separar el cálculo de la densidad de masa del prior y el de la evidencia. De modo que la evidencia es la suma de todas las verosimilitudes por una amplitud que es determinada por la densidad del prior X .

$$P(D|I) = \int_0^1 P(D|X, I) dX \quad (3.19)$$

$$dX = P(\theta|I) d\theta$$

El cálculo de la densidad del prior (Ec. 3.20) se hace de tal forma que el espacio de parámetros está restringido por el valor de verosimilitud \mathcal{L}_0 .

$$X(\mathcal{L}_0) = \int_{\mathcal{L}(\theta) > \mathcal{L}_0} P(\theta|I) d\theta \quad (3.20)$$

En la aproximación de la integral de la Ec. 3.19 se realiza un proceso iterativo. Se parte de un conjunto Θ de N caracterizaciones de los parámetros del modelo $\theta_1, \dots, \theta_N$ inicializados aleatoriamente y sus correspondientes verosimilitudes $\mathcal{L}(\theta_1), \dots, \mathcal{L}(\theta_N)$, la evidencia siendo nula y el primer valor del muestreo al máximo $X_0 = 1$.

En cada iteración i :

1. La caracterización que peor verosimilitud tenga del conjunto Θ se escoge como nominada i a ser sustituida, hará también la función de pivote inicial en la exploración de la cadena de Markov.
2. Se le asigna un valor $X_i = e^{-i/N}$ que decrece exponencialmente de 1 a 0 conforme aumentan las iteraciones.
3. Se calcula el peso propio de la iteración, a partir de la relación con el anterior valor de X , $w_i = X_{i-1} - X_i$, siendo también decreciente de forma exponencial desde $1/N$ a 0.
4. Se incrementa el valor de evidencia en $\mathcal{L}_i \cdot w_i$.
5. Se realiza la exploración de una nueva caracterización candidata por MCMC que sustituirá a la nominada.
 - a) Establecer un pivote para la exploración, ya sea la caracterización nominada u otra nueva (El algoritmo nested sampling no entra a valorar cómo realizar la exploración MCMC).
 - b) Inicializar una nueva caracterización del modelo, $\theta_{candidate}$, la candidata. Se obtiene en función del pivote $Q(\theta_{candidate}|\theta_{pivot})$, aplicando un *paso* al valor de sus parámetros
 - c) Aceptar la candidata como nuevo pivote en función de dos condiciones. La primera que su verosimilitud sea superior a la de la nominada \mathcal{L}_i . La segunda es una probabilidad de aceptación en proporción al ratio entre los priors de ambas, como se ha explicado al inicio de este apartado.
 - d) El *paso* se inicializa por configuración y se autoregula en base al ratio entre candidatas aceptadas y rechazadas. Si predominan los candidatos aceptados, el tamaño del paso se reduce de forma proporcional al número de aceptados, en caso contrario aumenta de forma proporcional al número de rechazados. A lo largo de la exploración este ratio se equilibra.
6. Tras un determinado número de exploraciones, el lugar de la nominada es ocupado por el último pivote y se inicia de nuevo el proceso desde el primer punto hasta completar todas las iteraciones.

Finalmente se hace un último incremento a la evidencia de $(L(\theta_1) + \dots + L(\theta_n)) \cdot X_j/N$ para cubrir el rango de X_j a 0 e incluir los últimos puntos presentes en el conjunto. Con gran número de iteraciones este factor es despreciable.

La distribución que compone el conjunto de muestras θ_i que han sido nominadas representa una muestra de la distribución a posteriori del problema. Esta distribución es la que se desea almacenar como solución, y se obtiene como elemento colateral en el cálculo de la evidencia.

Debido a los beneficios computacionales de los algoritmos MCMC y en particular de nested sampling, esta metodología se aplica en la búsqueda de la distribución de movimientos propios así como de su evidencia, para realizar posteriormente la comparación entre un modelo con movimientos propios y otro sin movimientos propios.

En el próximo capítulo se aportan detalles técnicos de la implementación del algoritmo.

Capítulo 4

Implementación

Parte del objetivo del presente trabajo es poder hacer la evaluación de los movimientos propios sobre millones de estrellas, siendo así, se trata de un problema altamente paralelizable. En la búsqueda de una alta eficiencia computacional, además de utilizar métodos de estimación optimizados del paradigma bayesiano, como es la aplicación de nested sampling sobre un modelo bayesiano jerárquico, uno de los esfuerzos centrales en el desarrollo del presente trabajo ha sido realizar la implementación del código para su ejecución sobre GPU, de forma totalmente paralelizada.

4.1. Entorno de trabajo

A continuación se detallan las características del hardware y software utilizado.

4.1.1. Hardware

Un buen rendimiento del hardware es imprescindible para obtener resultados en tiempos aceptables. Mayoritariamente se utilizan dos tipos de procesadores para computación, las CPU (Central Processing Unit) y las GPU (Graphical Processing Unit). Las primeras contienen un número reducido de núcleos que desarrollan buenas tareas individualmente en propósitos generales, mientras que las segundas contienen gran número de núcleos (más de 4000 en los últimos productos) y destacan en propósitos específicos que impliquen operaciones matriciales centrándose fundamentalmente en el procesado de imágenes.

Por lo general en un mismo equipo conviven procesadores CPU y GPU, ambos son necesarios, encargándose el primero de las tareas secuenciales y dejando al segundo las operaciones matriciales.

En el paradigma de la computación, emerge un modo de programación de nominado GPGPU (General Purpose - GPU) que se caracteriza por utilizar las GPU con propósitos generales de cálculo matemático más allá del procesado gráfico. El uso de algoritmos altamente paralelizables para el tratamiento de datos permite aprovechar el potencial de este tipo de procesadores, por lo que este modo de programación se denomina programación paralelizada.

El equipo sobre el que se ha desarrollado el estudio consta de:

- CPU Intel Core i7-8700K 3.7GHz x86_64
- GPU GeForce GTX 1060 con 6 GB de memoria.
- Memoria RAM: 8 GB, 2.1 GHz

Esta GPU dispone de 1280 núcleos CUDA. La terminología núcleos CUDA hace referencia a los procesadores multi-núcleo que componen las GPU, matizando que tiene una arquitectura con la que se interactúa a través de la plataforma CUDA.

Adoptando la filosofía de programación paralelizada, la implementación se ha planteado de la forma más paralelizable posible, cada núcleo se encarga de la búsqueda de la distribución a posteriori de los movimientos propios de una estrella. De forma que se pueden procesar en paralelo tantas estrellas como unidades de procesamiento tenga la GPU.

4.1.2. Software

Se detalla a continuación la plataforma y librerías utilizadas para la implementación.

El lenguaje utilizado para la implementación es Python debido a la extensa comunidad que lo respalda, a su conveniencia para el desarrollo científico y a su facilidad de interpretación, aspectos que favorecen tanto el desarrollo de la implementación como la extensión del proyecto por futuros colaboradores.

4.1.2.1. Programación sobre GPU

CUDA es una plataforma de computación paralelizada que está especialmente diseñada por Nvidia para el uso con propósito general (GPGPU). Comprende diversos paquetes de desarrollo y diagnóstico. CUDA se desarrolló inicialmente para ser utilizado mediante C/C++, pero en los últimos años se han publicado interfaces para su uso en otros lenguajes como Python (PyCUDA). PyCUDA es un entorno de programación para GPUs de NVIDIA que permite hacer la implementación del código con Python, pero solo del código que se ejecuta sobre CPU y la interfaz con la GPU. Las funciones a ejecutar en la GPU, denominadas *kernels*, se deben programar en C. Otra librería importante proporcionada es CuPY, ofrece un interfaz para la realización de operaciones sobre la memoria de la GPU con un API análogo a Numpy. Numpy es una librería de referencia en Python para la realización de operaciones matemáticas sobre vectores y matrices.

De forma paralela se ha desarrollado la suite de librerías de código abierto RAPIDS, promovido por Anaconda, Inc. y soportado por NVIDIA, Incluye librerías para el uso íntegro del API de Python, de entre las que se han utilizado Numba y CuDF. CuDF ofrece un API análogo al de Pandas, para el manejo de tablas etiquetadas sobre GPU. Pandas ofrece el tratamiento de datos en forma de tablas, añadiendo al potencial de Numpy el manejo de los vectores a través de etiquetas.

CuPY y CuDF, son útiles para el manejo de la memoria directamente en la GPU y minimizar las transferencias de datos entre CPU y GPU, sin embargo, estas librerías permiten realizar operaciones únicas distribuidas sobre todos los nodos, no están pensadas para ejecutar un flujo de operaciones entero sobre un único núcleo. Numba es un compilador compatible con Numpy, que utiliza el compilador LLVM siendo compatible tanto con CUDA-GPUs como con AMD-GPUs. Este entorno permite el uso de Python exclusivamente, alcanzando eficiencias similares a la programación directa sobre C o FORTRAN. Mediante Numba, los *kernels* de CUDA se programan también en Python, permitiendo el uso de un mismo lenguaje en la implementación, un factor importante para la portabilidad y colaboración futura. Esta librería permite implementar un flujo de operaciones enlazando *kernels*, con un mayor control de la memoria, lo que permite dirigir a cada núcleo los datos de una estrella y que ejecute el flujo de operaciones entero en cada uno.

Se ha adoptado el entorno de trabajo de Anaconda, que permite la creación de entornos virtuales de Python donde instalar las dependencias necesarias. Los paquetes CuDF y Numba están especialmente optimizados para funcionar en estos entornos de Anaconda. Se ha hecho uso de la librería Numba para compilar el algoritmo nested sampling y ejecutarlo sobre la GPU. Se han utilizado los paquetes CuPY y CuDF para preparación de los datos en la memoria de la GPU, ambos paquetes compatibles con Numba. Toda la programación de la GPU se realiza a través del interfaz que proporciona Numba.

Aparte de su evidente beneficio a la hora de facilitar la implementación, comprensión y colaboración del código, un aspecto fundamental para el desarrollo científico, Numba también presenta sus limitaciones. Se trata de un proyecto joven y colaborativo de código abierto, que está creciendo rápidamente pero por el momento las funcionalidades que se soportan en GPU son reducidas, por ejemplo, no soporta las operaciones vectoriales dentro de un mismo núcleo o la definición de clases, siendo necesario hacer un desarrollo más extenso. No obstante, este proyecto sustentado por Anaconda y Nvidia tiene muchas garantías de convertirse en el entorno de referencia para la programación de GPUs con Python. La intención de este estudio es liberar la herramienta desarrollada, por lo que, pensando en la evolución colaborativa del código en futuros desarrollos, Numba se ha considerado una herramienta adecuada.

4.2. Implementación del algoritmo

El código ha sido desarrollado desde cero valiéndose de las librerías genéricas citadas. Se ha tomado como referencia la implementación del algoritmo nested sampling de [Skilling \(2006\)](#) en Python, *mininest*, realizada por Issac Trots en 2007 y accesible desde la web de John Skilling:

<http://www.inference.org.uk/bayesys/>

Las aportaciones de la presente implementación respecto a la de referencia son por un lado la integración con un modelo bayesiano jerárquico y por otro la implementación de todo el flujo de forma paralelizable sobre GPU.

Para el desarrollo se ha seguido el estándar de Python PEP8. Se asegura la limpieza del código con *pylint* y se incluye una suite de test para validar cambios con *pytest*. Se ha utilizado GitHub como herramienta de control de

versiones.

4.2.1. Entrada de datos

El fichero de entrada debe contener una información mínima, para cada estrella debe haber al menos dos entradas con datos del experimento y la medición:

- Identificador de la estrella, `SOURCE_NUMBER`, ej: 5233
- Identificador del instrumento utilizado, `ASTR_INSTRUM`
- Valores de ascensión recta y declinación, `ALPHA_J2000` [grados] y `DELTA_J2000` [grados]
- Sus errores, `ERRA_WORLD` y `ERRB_WORLD`
- El instante de tiempo de la medición, `EPOCH` [años]

En la astrofísica es habitual utilizar el formato de archivos FITS. Consisten en una tabla etiquetada de datos similar a las mallas de datos (*dataframe*) de CuDF pero con información contextual adicional. El programa implementado soporta únicamente la carga de estos ficheros. El paquete *astropy* permite cargar estos ficheros y proporciona una interfaz para el volcado del contenido en un *dataframe* de Pandas.

4.2.2. Preprocesado

Normalización Los datos de entrada se someten a una normalización, cada estrella se normaliza independientemente.

- Se obtienen la media y desviación estándar de las medidas de ascensión recta (α), declinación (δ) y época
- A cada medida se le resta la media y se divide por la desviación estándar del campo que corresponda (α , δ o la época)
- Los errores de α y δ se dividen entre la desviación estándar del campo que corresponda (α o δ)

Rango de los instrumentos Se calcula el rango de medición de cada instrumento de medida l para cada coordenada α y δ , $range_{\alpha l}$ y $range_{\delta l}$. El rango es la diferencia entre las medidas máxima y mínima de entre todas las realizadas con un mismo instrumento. Cada medida i se asocia con el rango que le corresponde en cada coordenada en función del instrumento l con el que haya sido extraída.

Construcción de matrices El entorno de Numba no soporta el uso de *dataframes* por lo que es necesario transformar los datos en matrices de un mismo formato de dato. Se construyen las matrices de punto flotante con un orden determinado consistente a lo largo del programa para el acceso a los distintos campos.

Para aprovechar la paralelización al máximo, ya se ha comentado que cada núcleo CUDA va a procesar una estrella, de modo que cada núcleo trabajará con una matriz 2D con las medidas de entrada. La primera función que se ejecuta en la GPU es responsable del direccionamiento de cada matriz a su núcleo, por lo que la matriz que se pasa a la GPU es 3D.

Por lo tanto, el *dataframe* 2D de Pandas de entrada se agrupa por estrellas en una matriz 3D de CuPY de punto flotante. En la primera dimensión se suceden las distintas estrellas, la segunda enumera las observaciones y la tercera sus atributos. El campo “ASTR_INSTRUM” se descarta a cambio de incluir los rangos recién calculados del instrumento utilizado en “INSTRUM_ALPHA_RANGE” = $range_{\alpha l}$ y “INSTRUM_DELTA_RANGE” = $range_{\delta l}$, asociando a cada estrella su instrumento l . CuDF se utiliza de interludio para esta transformación.

Las matrices deben ser regulares pero el número de medidas de todas las estrellas no son las mismas, por lo que para cada estrella se han rellenado las mediciones con una entrada anómala, hasta alcanzar el máximo número de medidas que haya para una misma estrella, para así completar la cuadratura de la matriz 3D. En el momento de hacer el direccionamiento a los núcleos se selecciona solo el número de medidas reales tomadas de cada estrella para trabajar únicamente con los datos válidos.

4.2.3. Caracterización

Una vez preparados todos los datos, se realiza la caracterización del movimiento propio. Como se ha indicado anteriormente, se analizan inde-

pendientemente las coordenadas α y δ . Además para cada coordenada, para poder validar si un modelo de movimiento propio es más adecuado que uno sin movimiento propio, se hace una comparación del resultado de cada modelo. Por lo que la caracterización de una estrella consta de 4 ejecuciones del algoritmo, variando el eje y variando el modelo.

En cada ejecución se seleccionan los datos de interés (α o δ), una función instancia desde la CPU al algoritmo de *nested sampling* y a la desnormalización de los resultados, ambos procesados en la GPU. Para terminar se guardan los resultados en memoria CPU hasta recabar todas las ejecuciones. Posteriormente se hace la comparación de distribuciones con y sin movimiento y se vuelcan los resultados en disco.

Inicialización de matrices El procesado en GPU requiere que se construyan de antemano los contenedores de resultados y otros que puedan utilizarse en el proceso. Se inicializan 3 matrices 3D y una 2D con todo ceros, las cuatro distribuyen las estrellas en la primera dimensión. Una contendrá la muestra (las combinaciones de valores de los parámetros del modelo muestreados), la matriz 2D contendrá los estadísticos de la muestra, una tercera contendrá la población Θ de caracterizaciones y la cuarta contendrá dicha población en su estado inicial Θ_0 .

Para las caracterizaciones de los modelos, la segunda y tercera dimensión suceden los elementos de la muestra, tantos como iteraciones del algoritmo, junto con sus atributos, que son 6: los 2 valores aleatorios utilizados para su inicialización (en caso de ser caracterizaciones de la población inicial), el ángulo (m), la intersección (b), el logaritmo de la verosimilitud, el logaritmo del peso, la pendiente desnormalizada y la intersección desnormalizada.

Para las características tan solo se almacenan 4 atributos en una segunda dimensión: el logaritmo de la evidencia, su desviación, el valor de entropía H y su desviación. La entropía es una medida de la cantidad de información asociada a un valor particular de una variable aleatoria y se almacena con vistas a análisis posteriores que se deseen hacer de los resultados.

La matriz de la población Θ hace referencia al conjunto de caracterizaciones de un modelo que se actualiza de forma iterativa, sustituyendo miembros nominados por candidatos mejor adaptados. Este conjunto se inicializará como primer paso del algoritmo, dando lugar a Θ_0 , que se guarda en la matriz de población inicial con objetivos diagnósticos. Estas dos ma-

trices distribuyen a lo largo de la segunda y tercera dimensión los elementos de la muestra y sus atributos, los mismos de la primera matriz descrita.

Al empezar el procesamiento de la GPU, lo primero que se hace es fraccionar las matrices por la primera dimensión, para distribuir cada estrella en un hilo de procesamiento distinto enviando a cada uno tan solo los datos y contenedores que competen a su estrella: la matriz de datos de entrada, la de resultados de las muestras y sus estadísticas, la de la población y la de la población inicial. A partir de este punto, el procesado que se describe es ejecutado por un mismo núcleo.

La implementación del algoritmo requiere por un lado la estructura iterativa de nested sampling que acumula la evidencia de la distribución, fundamentada en la exploración de parámetros por MCMC y por otro lado la definición del modelo bayesiano jerárquico.

Nested sampling Se indican brevemente detalles de implementación sobre el algoritmo explicado en la sección 3.2.2.1. En primer lugar se inicializa la población Θ con N valores de los parámetros del modelo $\theta_n \equiv \{m_n, b_n\}$ muestreados de las distribuciones a priori de cada parámetro. m_n representa el ángulo ϕ (en radianes) y su distribución a priori es una distribución normal centrada en 0 con desviación estándar establecida por configuración. El ángulo se trunca en $[-1,5, 1,5]$, debido a que es un margen suficiente para el rango de pendiente esperada, $\tan(\phi)$, y se evitan así caracterizaciones con pendientes tan pronunciadas que el cambio de escala no permita analizar otras más adecuadas. b_n es la intersección con la ordenada en el origen y su distribución a priori es uniforme escalada al rango $[-5, 5]$.

Tras obtener la verosimilitud de la muestra inicial, se comienza el proceso iterativo y de paso se guarda para diagnóstico. El número de iteraciones I se fija por configuración.

Se selecciona el miembro de la población Θ con peor verosimilitud, el *nominado*, se calcula su peso y se actualiza la evidencia con su aportación. Se almacena como elemento de la muestra, éste es el punto clave para obtener el resultado buscado, ya que el conjunto de nominados conforman una muestra de la distribución a posteriori. Este nominado se debe sustituir por un nuevo candidato así que a continuación se selecciona otro elemento de la población, para ser el *pivote* y se realiza la exploración de MCMC para actualizarlo por un nuevo *candidato*, el último *pivote* sustituye al *nomina-*

do. El proceso continúa iterando hasta alcanzar el número de iteraciones fijado por configuración. Los peores elementos de la población van siendo seleccionados y sustituidos por otros mejores, de manera que al final toda la población acaba contenida en una región pequeña del espacio a posteriori.

Mediante la exploración de MCMC se proponen nuevos candidatos con valores vecinos a los priors de los parámetros del pivote. A cada parámetro del pivote m, b se le incrementa una magnitud aleatoria en el rango $[-paso, paso]$. Se realiza un truncamiento del valor de ambos parámetros m y b de igual forma que en la inicialización. El número de intentos que conlleva la exploración se determina por configuración.

Modelo bayesiano jerárquico Para el cálculo de verosimilitud se implementa la Ec. 3.13 sustituyendo en ella la Ec. 3.11 y Ec. 3.12. La esencia jerárquica del algoritmo se manifiesta en cada cálculo de verosimilitud. El prior del sesgo se muestrea de una distribución gaussiana centrada en 0 y con desviación de acuerdo a la Ec. 3.14. Los priors de los parámetros m, b se van muestreando a lo largo de las iteraciones, conforme se van generando nuevos candidatos y sustituyen al peor miembro de la población. Al sumar las verosimilitudes de cada nominado y aplicarles su peso dependiente del prior, se está marginalizando sobre los parámetros estimados. Es decir, se está obteniendo la evidencia y de paso, una muestra de la distribución full a posteriori marginalizada.

Detalles sobre el manual de uso, las opciones de configuración de la implementación y el acceso al código se pueden consultar en el Apéndice A.

4.2.4. Postprocesado

Al terminar las iteraciones, las caracterizaciones muestreadas se someten a un proceso de desnormalización, procesado también por la GPU.

- Para cada estrella, se obtienen la media y desviación estándar de los datos en cada eje (α, δ) y en la época, identificados respectivamente como: $\mu_\alpha, \sigma_\alpha^2, \mu_\delta, \sigma_\delta^2, \mu_{epoch}, \sigma_{epoch}^2$.
- Para cada caracterización del modelo, se obtienen la pendiente e intersección desnormalizadas de cada eje $m_{denorm_\alpha}, b_{denorm_\alpha}, m_{denorm_\delta}, b_{denorm_\delta}$.

$$\begin{aligned}
m_{denorm\alpha} &= \tan(m_\alpha) \cdot \frac{\sigma_\alpha^2}{\sigma_{epoch}^2} \\
m_{denorm\delta} &= \tan(m_\delta) \cdot \frac{\sigma_\delta^2}{\sigma_{epoch}^2} \\
b_{denorm\alpha} &= b_\alpha \cdot \sigma_\alpha^2 + \mu_\alpha - m_{denorm\alpha} \cdot \mu_{epoch} \\
b_{denorm\delta} &= b_\delta \cdot \sigma_\delta^2 + \mu_\delta - m_{denorm\delta} \cdot \mu_{epoch}
\end{aligned} \tag{4.1}$$

Dado que en la entrada de datos las unidades correspondientes son grados y años, la pendiente desnormalizada toma las unidades de $[\frac{degrees}{year}]$, escalándolo por $3,6 \cdot 10^6$ se obtienen milisegundos de arco por año [*mas*], la unidad oficial utilizada.

Tras la desnormalización de la muestra, se procede al cálculo de los estadísticos resumen de la distribución que conforman. Se obtienen la media, desviación estándar y percentiles 25, 50 y 75, en todos los casos ponderado con los pesos de los modelos. Esta operación se realiza en CPU debido a que, en el momento de desarrollo del estudio, no se ha hallado una solución para la interpolación en GPU.

Una vez se hayan realizado las ejecuciones del algoritmo para los 4 escenarios, con movimiento y sin movimiento, tanto en α como en δ , se analiza cada eje. Se comparan las evidencias de las muestras de modelos con movimiento y sin movimiento en función del factor Bayes (K).

Se hace referencia como la ganadora a la distribución que mayor evidencia muestre, y se anota el valor del factor Bayes. Los resultados se guardan en un fichero JSON y se da opción al usuario a permitir el guardado exclusivo de las características de la muestra del modelo ganador o guardar todos los resultados. Esto incluye para cada estrella, cada una de las 4 ejecuciones, la muestra de la distribución a posteriori del modelo, los estadísticos de la distribución, las caracterizaciones de la población inicial y las estadísticas de evidencia e información.

A continuación se entra en detalle en el proceso de evaluación.

Capítulo 5

Evaluación

5.1. Metodología de evaluación

Ya se ha introducido el uso del factor Bayes para la comparación de las hipótesis. Este factor es el ratio entre la evidencia de la distribución con movimiento y sin movimiento. Determina cuánto mayor (o menor) es la evidencia de una distribución frente a la otra (Jeffreys, 1998).

Jeffreys propone la consideración que se debe dar a cada una de las hipótesis en función de su ratio. En este estudio la implementación se ha realizado en base logarítmica por conveniencia matemática, por lo que resulta más apropiada la propuesta por Kass and Raftery (1995), popularmente citada y adaptada en la Tabla 5.1 para mantener la distinción de 5 niveles propuesta originalmente por Jeffreys. Considérese que se está contrastando la evidencia de la distribución con movimiento propio frente a la de la otra. En el problema en estudio, las hipótesis nula y alternativa representan la validez de los modelos con movimiento y sin movimiento. Es decir son mutuamente excluyentes, o se da una o se da la otra, por lo que la evidencia en contra de una de las hipótesis es evidencia a favor de la otra. Definiendo el ratio como la relación de evidencia de la hipótesis de movimiento propio frente a la de carencia de movimiento, se personaliza la descripción de la evidencia en la Tabla 5.1.

Para valorar la eficacia del algoritmo, se compara la hipótesis ganadora (en base al resultado del factor Bayes), con los movimientos propios calculados por otras fuentes. De cara a la evaluación, en el texto se hace referencia a un modelo lineal, y en las gráficas simplemente a un modelo, para hablar de unos valores que puedan tomar sus parámetros, también referido como

Nivel	Ratio	Log ratio	Interpretación
1	$1 < K \leq 3,2$	$0 < \log K \leq 1/2$	Evidencia a favor del movimiento, pero no merece más mención
2	$3,2 < K \leq 10$	$3,2 < \log K \leq 10$	Evidencia sustancial del movimiento propio
3	$10 < K \leq 32$	$1 < \log K \leq 1,5$	Evidencia fuerte del movimiento propio
4	$32 < K \leq 100$	$1,5 < \log K \leq 2$	Evidencia muy fuerte del movimiento propio
5	$K > 100$	$\log K > 2$	Evidencia decisiva del movimiento propio

Tabla 5.1: Interpretación del factor de Bayes basada en las propuestas de [Jeffreys \(1998\)](#) y [Kass and Raftery \(1995\)](#) y adaptada a la nomenclatura del problema en estudio.

caracterización de un modelo.

5.2. Colecciones de evaluación

Se evalúan datos de 2 tipos de colecciones.

En primer lugar y siendo las colecciones más extensas, se han generado 8 colecciones de datos sintéticas para hacer una primera valoración. Se distinguen entre colecciones que tienen y no tienen movimiento propio, tienen y no tienen sesgo y tienen y no tienen espurios. El modelo ganador de la inferencia se compara con la referencia conocida a priori dada la fabricación de los datos de entrada. Cada una de estas colecciones consta de 1000 estrellas con hasta 200 medidas por estrella.

En segundo lugar, se ha evaluado un conjunto de datos de 210 estrellas del proyecto DANCe. Estas estrellas tienen menos medidas siendo la estrella 5254 la que más tiene con 103.

5.3. Métricas de evaluación

Se ha implementado un módulo para facilitar el análisis de resultados. Se realiza el análisis de dos métricas. En primer lugar la constancia de los resultados, un aspecto fundamental para poder confiar en ellos. En segundo

lugar su certeza en comparación con la información disponible de cada una de las colecciones evaluadas. Todas las pruebas realizadas se han repetido 10 veces para que las conclusiones tengan mayor significación estadística.

En el análisis de la constancia se ha inspeccionado la variabilidad del factor de Bayes a lo largo de varias ejecuciones del programa con la misma configuración. Se aporta una gráfica radial (i.e. Fig. 6.14) donde se distribuyen factores Bayes de cada experimento para visualizar la constancia en este factor.

El análisis de la confusión se diferencia entre una colección y otra. En la colección de datos sintéticos, la referencia de que los datos tengan o no movimiento propio se conoce de antemano. En el caso de la colección de DANCe, se tienen los movimientos propios calculados mediante mínimos cuadrados y el error de la estimación. En función de la relación de estos valores se puede establecer un margen que distinga aquellas estrellas de las cuales hay certeza suficiente de que tienen movimiento propio. Son aquellas para las que el valor de movimiento propio proporcionado en los catálogos es *superior a 5 veces el error* de la estimación. Se toma esta franja como referencia para hacer una validación inicial. Para todas las estrellas que cumplan con esta regla, el programa debe estimar una evidencia certera de tener movimiento propio. Teniendo en cuenta esta referencia, para cada experimento se compone una matriz de confusión (i.e. Fig. 6.37), a partir de la cual se analizan los resultados. En caso de que el movimiento del catálogo para una estrella no sea superior a 5 veces su error de medición, no se puede asegurar la presencia o ausencia de movimiento para esa estrella, por lo que solo se trabaja con la parte positiva de las citadas matrices de confusión.

5.4. Gráficas

Se ha implementado otro módulo para la visualización de los resultados. La generación de las gráficas de resultados se puede incluir como opción en la ejecución del programa principal o se puede ejecutar posteriormente partiendo del fichero de resultados de cada experimento. Estas gráficas se generan para cada estrella, en cada una se comparan los resultados de la distribución con movimiento y sin movimiento, e incluyen:

- Dispersión de coordenadas de la estrella junto con los modelos lineales muestreados de la distribución a posteriori. (Ej.: Fig. 6.1).

- Distribución de la inclinación m y la intersección b de la muestra. (Ej.: Fig. 6.2).
- Dispersión de la inclinación m frente a la intersección b de la muestra. (Ej.: Fig. 6.3).
- Histograma de la inclinación m y la intersección b de la muestra. (Ej.: Fig. 6.5).
- Histograma de la población inicial.
- Distribución a posteriori de la inclinación m y la intersección b de la muestra. (Ej.: Fig. 6.5).

Por defecto, las gráficas se generan con los datos ya desnormalizados, excepto las de dispersión de parámetros e histograma de la población inicial. Ejecutado como un módulo aparte, permite la generación de todas las gráficas para los valores normalizados.

En todas las gráficas, salvo en la de dispersión de parámetros, se identifica la distribución que mayor evidencia presenta con un fondo verdoso en la gráfica. Se apilan las gráficas correspondientes a la distribución con movimiento propio a la izquierda y las otras a la derecha. En las figuras con 3 o 4 gráficas, se hace una distinción por parámetros, la parte superior se reserva para la pendiente y la parte inferior para la intersección. En las figuras con 1 o 3 gráficas se omite la representación del modelo sin movimiento propio por ser trivial.

En las gráficas de dispersión y en la de distribución de los parámetros, los modelos se identifican por una escala de colores, consistente entre ellas, que representa el orden en el que han sido muestreadas, es decir, representa el valor de verosimilitud. La escala cromática de la gráfica de verosimilitud de los parámetros es redundante pero aporta una visualización de la correspondencia entre las tres.

En estas mismas gráficas se incluye con color negro, indicando aquellas caracterizaciones de la muestra que pertenecen a la población inicial, independientemente de su valor de verosimilitud. Estas escalas ayudan en la interpretación de los resultados. La gráfica de dispersión de coordenadas incluye además una segunda escala de colores, aplicada sobre los datos, que distingue el instrumento utilizado en la observación.

La presentación y discusión de los resultados se incluye en el siguiente apartado.

Capítulo 6

Resultados

En este capítulo se presentan los resultados a la par que se discuten. Se diferencia el análisis entre las colecciones de datos sintéticos y la de datos reales. Con todas las colecciones se han realizado numerosas pruebas en el análisis, con diversos valores del número de iteraciones del algoritmo y el tamaño de la población de exploración. El resto de parámetros configurables del algoritmo se mantienen fijos:

- Desviación estándar del prior de la pendiente, σ_m : $\pi/8$
- Valor medio de la distribución de espurios, Y_o : 0
- Factor de ponderación de la distribución de espurios, P_o : 0.3
- Proporción del rango del instrumento que determina la desviación estándar del prior del sesgo, k : 5 %
- Pasos de exploración: 20
- Tamaño inicial del paso de exploración: 0.3

Entre todas las pruebas se analiza un total de 1 TB de resultados.

6.1. Datos sintéticos

Se ha ejecutado el programa sobre cada una de las 8 colecciones de datos sintéticos. De cara a la presentación de resultados aclarar que las 8 colecciones son independientes, por mucho que coincidan los identificadores de las estrellas usados.

En esta sección se analizan los resultados de la ejecución del algoritmo sobre un conjunto de 50 estrellas de las distintas colecciones. Se realizan dos tipos de análisis, en primer lugar se hace un análisis individual de dos estrellas (Sección 6.1.1) y en segundo lugar se hacen varios análisis grupales de la precisión del resultado a lo largo de los 10 experimentos (Sección 6.1.2).

En el análisis individual se extraen dos estrellas de las colecciones con sesgos y espurios, por ser el caso en el que la inferencia es más compleja. Una estrella se extrae de la colección sin movimientos propios (Sección 6.1.1.1) y la otra de la colección con movimientos propios (Sección 6.1.1.2). De cada estrella se explican y analizan las gráficas de la muestra de modelos lineales resultantes del algoritmo.

En los análisis grupales se comparan los resultados de los 10 experimentos valiéndose del módulo desarrollado para tal efecto. En primer lugar, se analizan los indicadores de constancia y confusión en la predicción, comparando las colecciones con y sin movimiento propio ante una configuración fija (Sección 6.1.2.1). Un segundo análisis evalúa el efecto del cambio de los valores del tamaño de la población (en adelante *pop*) y del número de iteraciones (en adelante *it*) sobre una colección fija con movimiento propio, sesgos y espurios (Sección 6.1.2.2). Un último análisis compara los resultados sobre distintas colecciones para ver el efecto de la presencia de espurios y sesgos en la estimación (Sección 6.1.2.3).

La configuración particular de los experimentos realizados para el análisis individual y los dos primeros análisis grupales es de $it = 8000$ y $pop = 1500$.

6.1.1. Análisis de un experimento individual

Se ejecuta el módulo encargado de generar las gráficas de resultados del algoritmo, se explican y analizan las gráficas que corresponden a cada una de las dos estrellas consideradas.

6.1.1.1. Colección sin movimiento propio

En la figura Fig. 6.1 se muestran las observaciones de una estrella con distintos instrumentos y los modelos lineales inferidos. Se observa la distinción entre los modelos lineales que pertenecen a la población inicial, en color negro, y los que han sido resultado de la exploración, escala cromática izquierda.

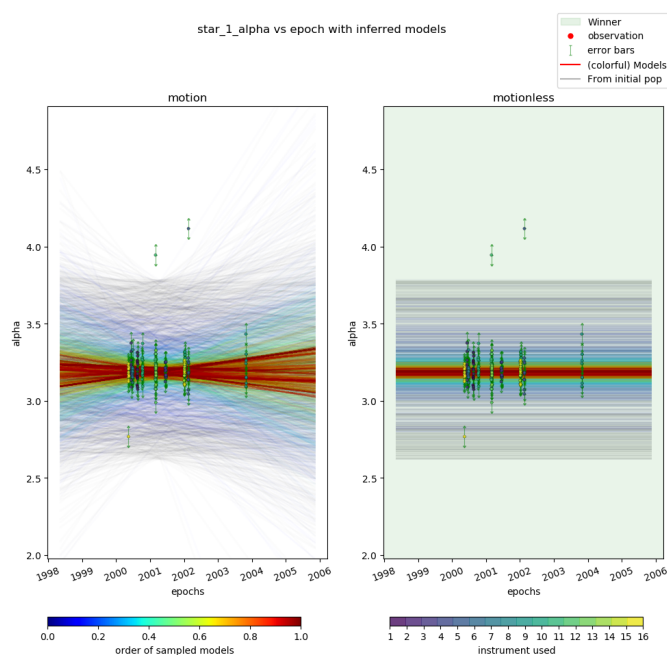


Figura 6.1: Posiciones de la estrella en ascensión recta (eje Y) a lo largo del tiempo (eje X) incluyendo los modelos lineales inferidos. Colección: Sintética sin movimientos propios. Configuración: $it = 8000$, $pop = 1500$

Se observa como conforme se procede en las iteraciones, muestreando nuevos candidatos, se van restringiendo a un área más pequeña hasta converger en una región localizada. En esta gráfica, este comportamiento se observa claramente en la distribución del modelo sin movimiento propio, sin embargo, en la distribución del modelo con movimiento propio no hay una clara convergencia. En este ejemplo, sin movimiento propio, se ha predicho correctamente la carencia de dicho movimiento, tal y como puede observarse por el color verdoso del fondo de las gráficas sin movimiento propio, a pesar de la presencia de sesgos y espurios.

Para analizar con más detalle la convergencia de los modelos en una región del espacio se pueden consultar las figuras Fig. 6.2 y Fig. 6.3. En la primera se observa el valor de verosimilitud para cada modelo lineal, distinguiendo cada parámetro. En la segunda se enfrentan los parámetros. En ambas gráficas, los modelos lineales que han pertenecido a la población inicial se marcan en negro, independientemente de su verosimilitud.

En la figura Fig. 6.2, se aprecia la evolución que ha seguido cada parámetro en la iteración. Para la distribución sin movimiento propio se observa

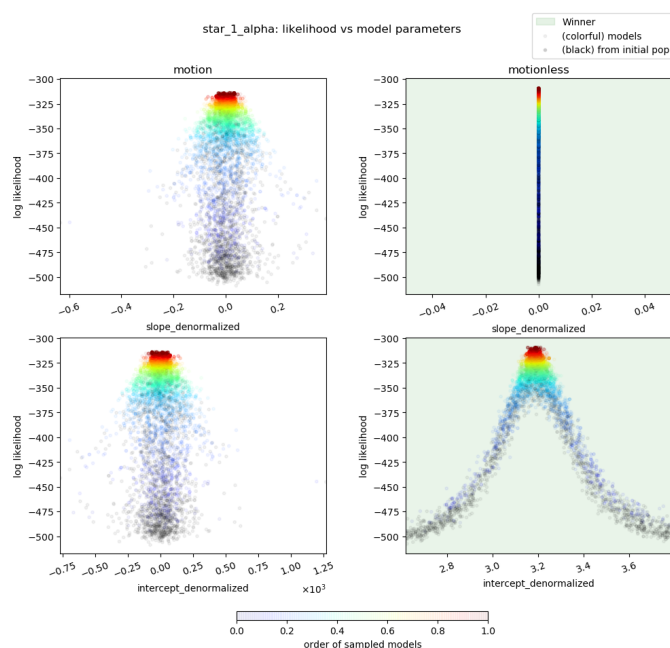


Figura 6.2: Verosimilitud (eje Y) de los modelos lineales en función del valor que tienen (eje X) en cada parámetro, m_{denorm_α} (arriba) y b_{denorm_α} (abajo). La escala cromática indica el orden de muestreo (i.e., la verosimilitud). Colección: Sintética sin movimientos propios. Configuración: $it = 8000$, $pop = 1500$

como en la intersección, inicializada uniformemente en el rango $[-5, 5]$, los valores de la población inicial se reparten equitativamente por dicho rango. En el caso del modelo con movimiento los parámetros inicializados m presentan una mayor densidad centrada en 0, acorde con la inicialización gaussiana establecida. Se inicia el muestreo guardando primero los candidatos más alejados al centro. Conforme van siendo sustituidos en la exploración, comienza a aumentar la selección de modelos resultantes de dicha exploración. Llega un punto en el que prácticamente todos los elementos de la población inicial han sido ya seleccionados, pudiendo salvarse alguno que haya sido inicializado con parámetros particularmente adecuados. La población se ha renovado y los actuales miembros han sido generados durante la exploración.

La exploración converge hacia el centro de la gráfica, dado que se trabaja con datos normalizados y centrados. Respecto a la pendiente, en la distribución sin movimiento propio, siempre va a valer 0, por eso se ha descartado en el resto de gráficas. Se ha mantenido en la Fig. 6.2 por motivos comparati-

vos con su vecina, principalmente para las estrellas con movimiento propio, donde se observará una notable diferencia en el valor de verosimilitud (por ejemplo: Fig. 6.8).

En el caso de la distribución con movimientos propios, en la Fig. 6.2 se observa como la pendiente no ha terminado de converger en una región concentrada. Se aprecia una misma densidad de medidas, y con similar verosimilitud, a ambos lados del 0. En la gráfica de las observaciones (Fig. 6.1) también puede apreciarse la dispersión en los últimos modelos lineales muestreados. No se ha encontrado una región de los parámetros con movimiento que tenga suficiente verosimilitud.

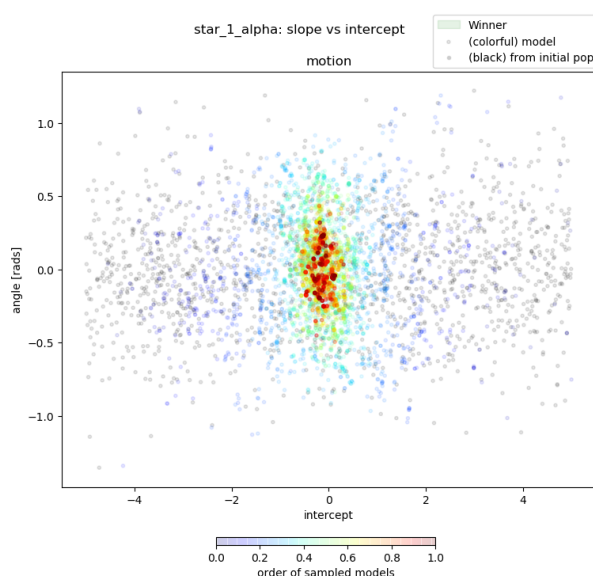


Figura 6.3: Dispersión de valores de los parámetros m (eje Y) frente a b (eje X). La escala cromática indica el orden de muestreo (i.e., la verosimilitud). Colección: Sintética sin movimientos propios. Configuración: $it = 8000$, $pop = 1500$

La figura 6.3 muestra la dispersión de los parámetros m y b entre sí. Se trata de los parámetros originales sin desnormalizar. Se ve claramente como no hay una región definida que concentre los modelos, los valores de ambos parámetros se reparten en torno al punto $(0, 0)$. Se aprecia muy bien también, como la población inicial (puntos negros) está distribuida uniformemente a lo largo de la intersección y los valores del ángulo tienden a una mayor densidad en el centro, al seguir una distribución gaussiana.

En la figura 6.4 se muestra la frecuencia de cada parámetro en los mode-

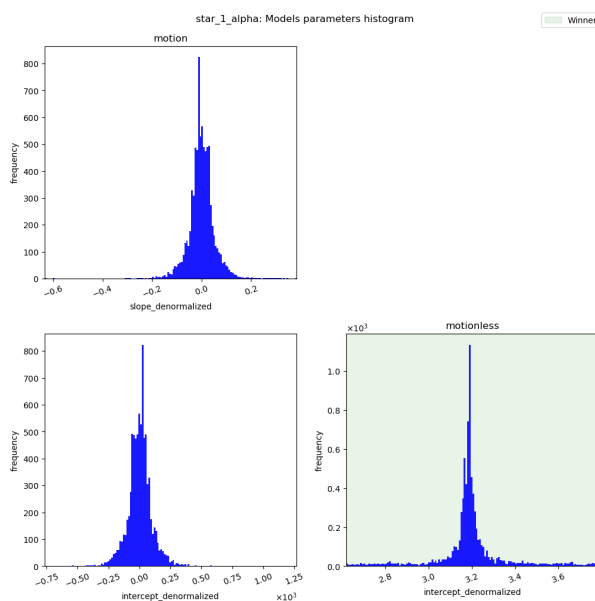


Figura 6.4: Histograma de los parámetros m_{denorm_α} (arriba) y b_{denorm_α} (abajo). Colección: Sintética sin movimientos propios. Configuración: $it = 8000$, $pop = 1500$

los. En concordancia con las gráficas anteriores, hay una mayor frecuencia de los parámetros en torno al 0. No obstante, esta gráfica no nos aporta mucha información, al no estar marginalizados los parámetros. Se debe consultar la distribución a posteriori para valorar el conocimiento a priori del problema. Ésta se muestra en la figura 6.5. Comparando ambas figuras se observa que tras la marginalización hay un cambio considerable en los resultados. La probabilidad a posteriori de los modelos lineales en regiones con priors poco probables se ve reducida y se destacan aquellos con mayor densidad de probabilidad a priori. De forma coherente, en este ejemplo la distribución se sitúa en torno a 0, no hay una región a posteriori clara por la que decantarse.

La simetría entre las gráficas de la pendiente y de la intersección es debida a la naturaleza de las medidas. La escala real de las épocas es al menos un orden de magnitud mayor que la de la intersección, haciendo que una variación en la inclinación normalizada suponga un efecto mucho mayor en la escala desnormalizada que el que pueda suponer una variación de la intersección normalizada. Por esta razón el valor de intersección desnormalizada recibe su influencia principalmente de la pendiente desnormalizada. El efecto espejo viene de la resta implicada en la desnormalización.

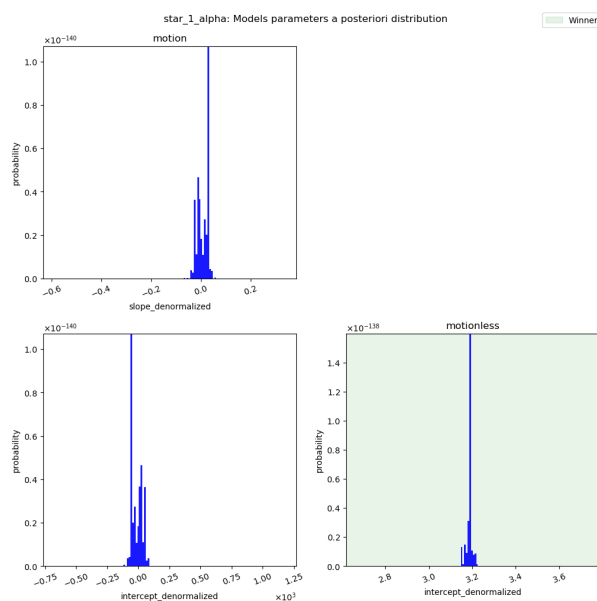


Figura 6.5: Distribución a posteriori de los parámetros, m_{denorm_α} (arriba) y b_{denorm_α} (abajo). Colección: Sintética sin movimientos propios. Configuración: $it = 8000, pop = 1500$

Se muestra a modo de ejemplo la gráfica de verosimilitud de los parámetros sin desnormalizar en la figura Fig. 6.6. En el caso de m , evoluciona con una tímida tendencia hacia 0, sin focalizarse en una región definida. En el caso de la intersección, se puede observar un comportamiento similar en la evolución del parámetro para ambas distribuciones, sin embargo, sutilmente se aprecia una mayor densidad de puntos azules en las faldas de la distribución con movimientos propios. Esto es debido a que la exploración tiene más facilidad de valorar positivamente modelos lineales con el parámetro de intersección más separado del centro, ya que la inclinación juega su papel haciendo que sea una caracterización en cierta medida plausible. Esta observación evidencia también la necesidad de una mayor exploración, para converger en una región definida, ante un espacio de búsqueda de mayor dimensionalidad; en este caso el parámetro extra m respecto a la distribución sin movimiento propio.

A continuación se presentan estas mismas gráficas de un ejemplo con movimiento para dar paso en adelante al análisis estadístico grupal.

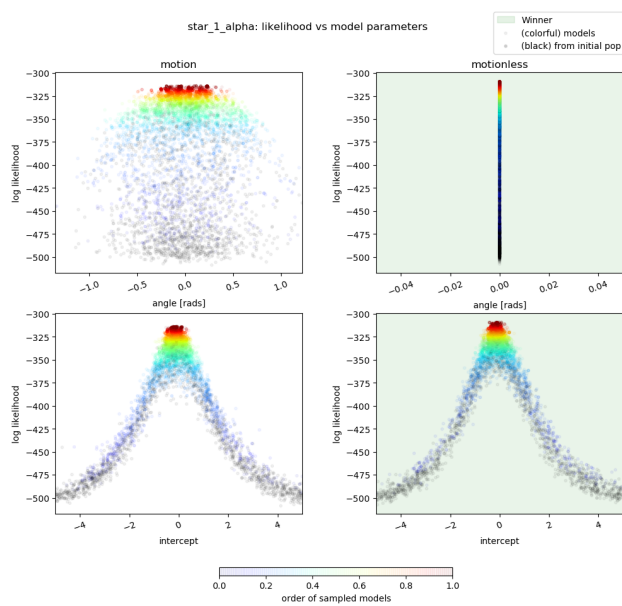


Figura 6.6: Verosimilitud (eje Y) de los modelos lineales en función del valor que tienen (eje X) en cada parámetro normalizado, m (arriba) y b (abajo). La escala cromática indica el orden de muestreo (i.e., la verosimilitud). Colección: Sintética sin movimientos propios. Configuración: $it = 8000$, $pop = 1500$

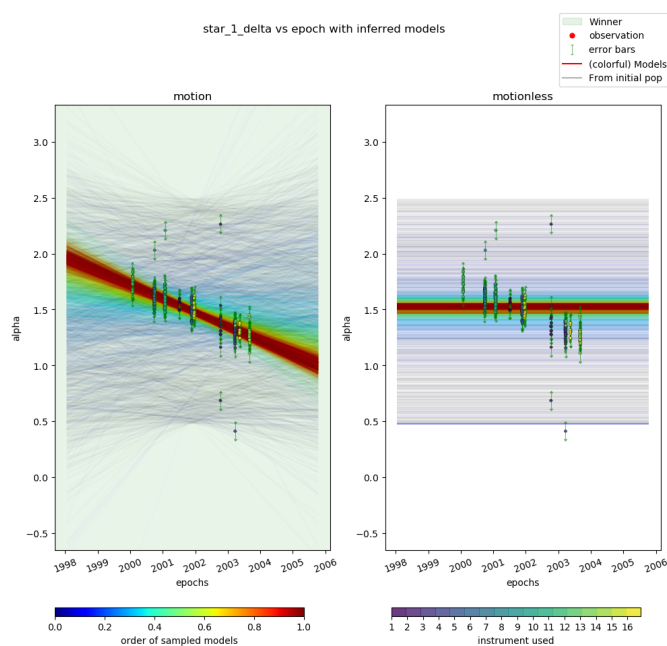


Figura 6.7: Posiciones de una estrella en declinación (eje Y) a lo largo del tiempo (eje X) incluyendo los modelos lineales inferidos. Colección: Sintética con movimientos propios. Configuración: $it = 8000, pop = 1500$

6.1.1.2. Colección con movimiento propio

En la Figura 6.7, se ven los modelos inferidos sobre los datos de una estrella arbitraria de la colección con movimientos propios, sesgo y espurios. Los datos espurios se ven claramente y los sesgos de varios instrumentos también se aprecian gracias a la escala cromática derecha. Se ve por el fondo verdoso que la distribución de mayor evidencia es la de movimientos propios. Se observa cómo la banda cromática va reduciendo su amplitud conforme aumenta la verosimilitud. Partiendo de unos modelos azules muy dispersos, se avanza en la generación de candidatos más adecuados y la banda cian ya empieza a posicionarse. Las bandas verdes y más remarcada la amarilla ya marcan una tendencia y a partir de aquí los candidatos generados ya se aglutinan en una región bastante definida. La distribución sin movimiento se posiciona a mitad de altura de los datos, pero las observaciones tienen una clara tendencia y a simple vista se ve que los modelos lineales no secundan los datos.

Para analizar en más detalle la convergencia en la región de los parámetros se analizan las figuras Fig. 6.8 y Fig. 6.9. En la primera se ve un com-

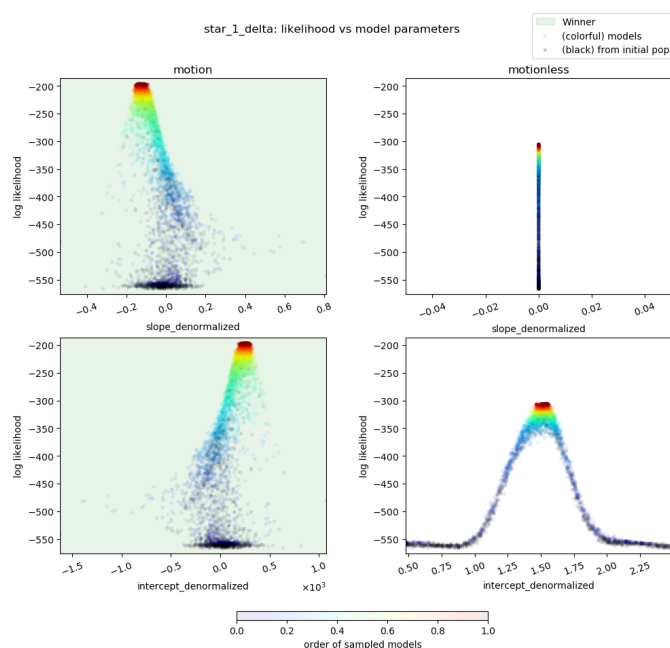


Figura 6.8: Verosimilitud (eje Y) de los modelos lineales en función del valor que tienen (eje X) en cada parámetro, m_{denorm_α} (arriba) y b_{denorm_α} (abajo). La escala cromática indica el orden de muestreo (i.e., la verosimilitud). Colección: Sintética con movimientos propios. Configuración: $it = 8000$, $pop = 1500$

portamiento muy distinto al de su homóloga de la Fig. 6.2.

En la distribución sin movimiento, el parámetro de intersección evoluciona hasta un máximo a partir del cual no puede avanzar. La diferencia de verosimilitud entre ambas distribuciones es notable.

En la distribución con movimiento, se observa como la mayoría de las muestras de la población inicial tienen una baja verosimilitud. A pesar de tener un valor de pendiente que en futuras iteraciones tenga mayor verosimilitud, esto se debe a que dichos modelos lineales tienen una intersección particularmente mala. Esto se aprecia en la versión normalizada de la figura (Fig. 6.11).

La banda de mayor densidad de los parámetros que se forma es debida a que éstos se van generando en las posiciones cercanas a las ya presentes. Todos aquellos modelos lineales con ángulo mayor a 0 han sido progresivamente sustituidos por modelos lineales generados en la vecindad de un miembro de la población. Al ser al inicio abundante la cantidad de mode-

los lineales con ángulos mayores a 0, tienen una probabilidad alta de ser escogidos como pivotes. Conforme se van explorando nuevos candidatos, se van aglutinando en torno a los miembros más abundantes. A la altura de $\text{loglikelihood} = -350$ los peores miembros de la población se reducen a una banda con $m \approx 0$. Mirando a la figura normalizada (Fig. 6.11), los valores de intersección de estos modelos lineales están en su mayoría aglutinados en los costados de la gráfica, debido a la misma razón que la banda de m de la Fig. 6.8. Al proseguir en las iteraciones, a medida que la intersección se va centrando, las pendientes van alejándose del 0. Los modelos con pendiente en la región final empiezan a aflorar a partir de $\text{loglikelihood} = -300$, poblando paulatinamente la región y tendiendo a su punto medio. Con esta explicación se desea recalcar que la exploración compete a los dos parámetros conjuntamente. Por mucho que uno tenga un valor óptimo, el segundo es imprescindible para que la verosimilitud sea buena.

Para observar la relación entre ambos parámetros se incluye la Fig. 6.9. En esta figura se observa que partiendo de una población inicial debidamente distribuida (uniforme en b , gaussiana en m), las caracterizaciones del modelo con movimiento terminan convergiendo en una región definida y claramente separada de la pendiente nula.

La distribución a posteriori (Fig. 6.10) muestra la región del espacio de parámetros donde se agrupan caracterizaciones ponderadas por su aportación en el cálculo de la evidencia. La distribución con movimiento propio apunta a una región en torno a $m \approx -0,1$ y $b \approx 250$, mientras que la distribución sin movimiento se centra en $b \approx 150$, precisamente un valor centrado en las propias coordenadas, al no tener inclinación sus modelos lineales. Se observa que la diferencia de probabilidad entre ambas distribuciones es de 40 órdenes de magnitud. Con esta comparación se enfatiza la naturaleza bayesiana del estudio, un valor reducido de probabilidad a posteriori no significa per se que la hipótesis sea falsa, la confianza la aporta la comparación con otra hipótesis. 40 órdenes de magnitud se antojan suficientes para apoyar la hipótesis con movimiento.

Analizando la evolución de los parámetros normalizados (Fig. 6.11). El hecho de que el valor de intersección normalizado tienda a 0 es causa de la normalización. Se centran los datos con el origen en la media de las épocas y la media de las coordenadas (ver los modelos normalizados en la Fig. 6.12). Se observan unas colas muy aplastadas en los extremos del parámetro de

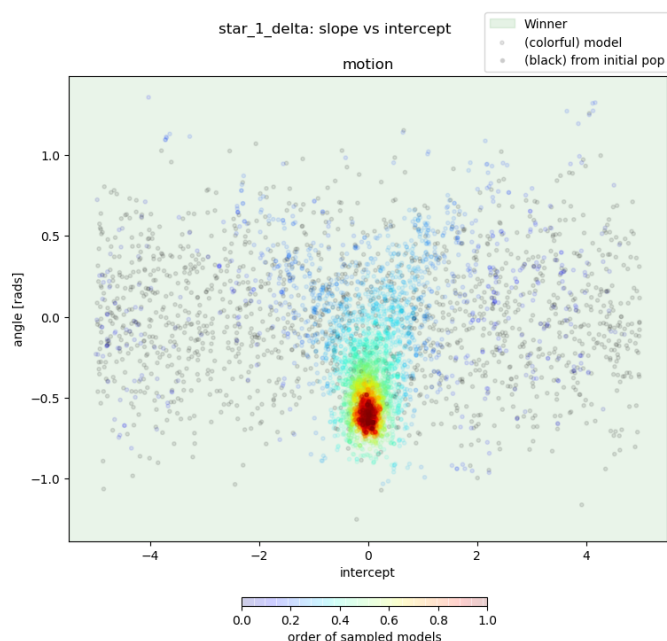


Figura 6.9: Dispersión de valores de los parámetros m (eje Y) frente a b (eje X). Colección: Sintética con movimientos propios. Configuración: $it = 8000$, $pop = 1500$

intersección. Esto es debido a que para esos valores no existe prácticamente ninguna pendiente que se acerque a los datos. La exploración que se da en esa región es residual. Conforme avanza la iteración, los valores de las colas se aglutinan en los lomos y a partir de $\loglikelihood = -400$ se observa cómo los nuevos candidatos van poblando paulatinamente la que será la región de convergencia. Respecto a la distribución del parámetro m se observa que es un calco escalado de la pendiente desnormalizada de la Fig. 6.8.

6.1.2. Análisis estadístico de varios experimentos

A continuación se presenta el análisis estadístico de los resultados sobre distintas colecciones sintéticas y con distintas configuraciones. Se evalúan la constancia en los resultados y la certeza de los mismos por comparación con la referencia. En este caso la referencia es el propio nombre del fichero, donde se indican las características con las que han sido generados los datos sintéticos (con/sin movimiento propio, con/sin espurios, con/sin sesgo). En cada colección y configuración se realizan 10 experimentos sobre 50 estrellas.

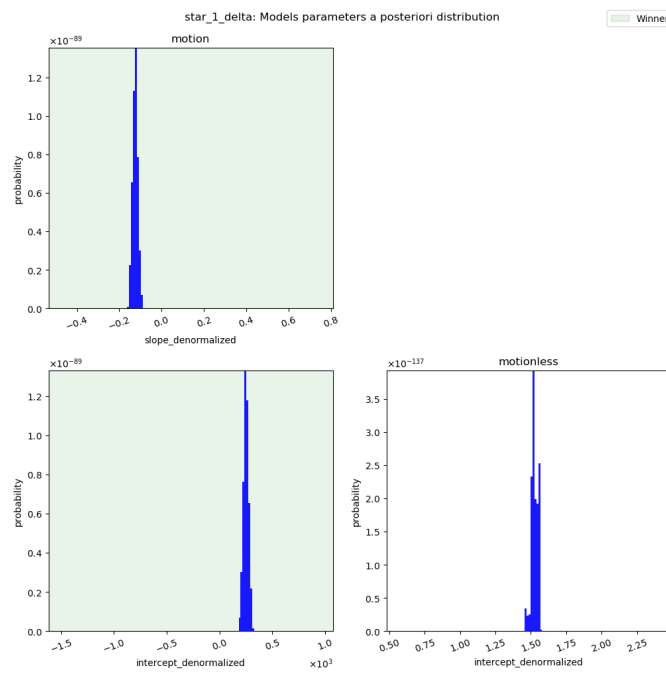


Figura 6.10: Distribución a posteriori de los parámetros, m_{denorm_α} (arriba) y b_{denorm_α} (abajo). Colección: Sintética con movimientos propios. Configuración: $it = 8000$, $pop = 1500$

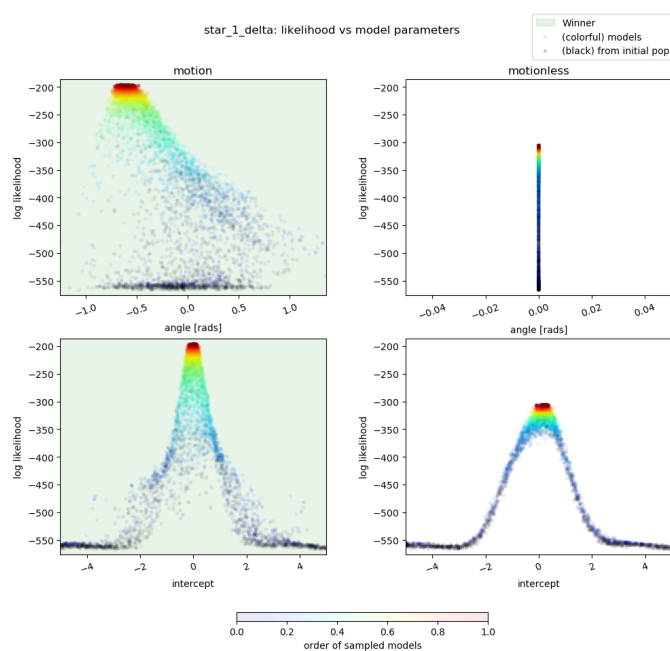


Figura 6.11: Verosimilitud (eje Y) de los modelos lineales en función del valor que tienen (eje X) en cada parámetro normalizado, m (arriba) y b (abajo). La escala cromática indica el orden de muestreo (i.e., la verosimilitud). Colección: Sintética con movimientos propios. Configuración: $it = 8000$, $pop = 1500$

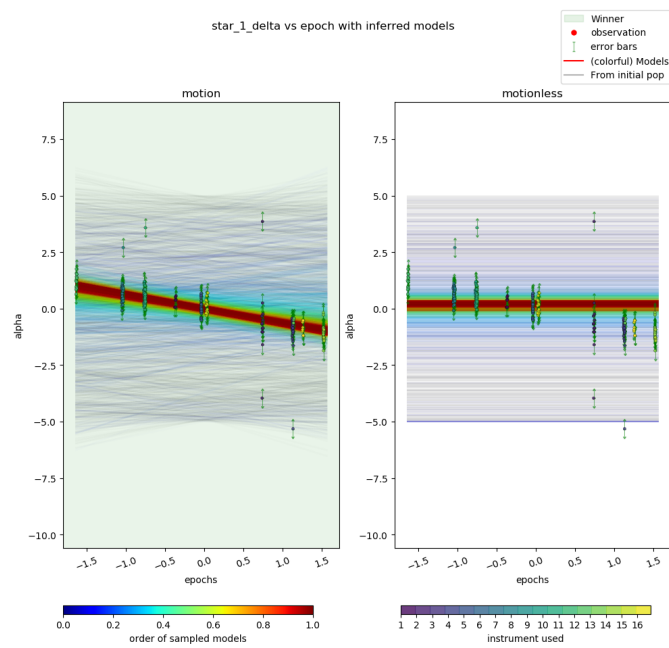


Figura 6.12: Posiciones normalizadas de una estrella en declinación (eje Y) a lo largo del tiempo (eje X) incluyendo los modelos lineales inferidos. Colección: Sintética con movimientos propios. Configuración: $it = 8000$, $pop = 1500$

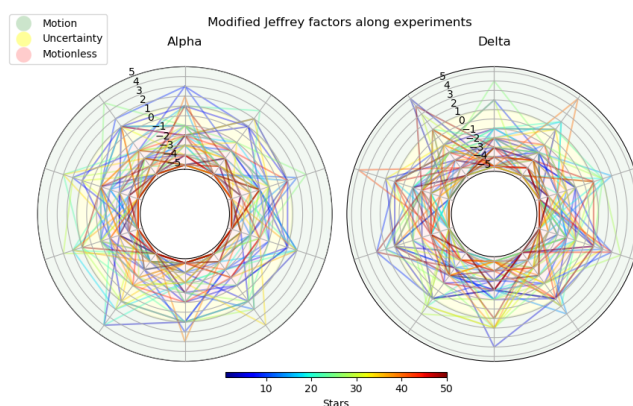


Figura 6.13: Nivel del factor Bayes (coordenada radial) a lo largo de 10 experimentos (coordenada angular). Colección: Sintética sin movimientos propios. Configuración: $it = 8000$, $pop = 1500$

6.1.2.1. Comparación entre colecciones con y sin movimiento propio

Se muestran y discuten los resultados de los 10 experimentos sobre 50 estrellas de las colecciones con movimiento y sin movimiento, ambas con sesgos y espurios. Se mantiene la configuración del análisis anterior, $it = 8000$ y $pop = 1500$.

La constancia de los resultados en las colecciones sin movimiento y con movimiento se visualiza en las figuras Fig. 6.13 y Fig. 6.14 respectivamente. En estas gráficas se distribuyen en la coordenada angular las 10 ejecuciones realizadas y en la coordenada radial el nivel del factor Bayes. Se representa una adaptación al presentado en la Tabla 5.1, en caso de que el ratio sea inferior a 1 (a 0 en logaritmo), la evidencia se acumula a favor de la hipótesis contraria. Se ha indicado con números negativos aquellos factores que secundan la ausencia de movimiento propio.

En la colección sin movimientos propios la variación entre resultados es alta mientras que la misma configuración sobre la colección con movimientos propios da unos resultados muy consistentes entre experimentos.

En el caso de la colección sin movimientos propios (Fig. 6.13), en muchas de las estrellas los resultados a lo largo de distintas ejecuciones presentan poca variabilidad, pero en otras tantas, el cambio del factor Jeffreys entre ejecuciones es sustancial. Esto refleja un resultado negativo ya que ocasionalmente se asevera la presencia de movimiento en estrellas de la colección

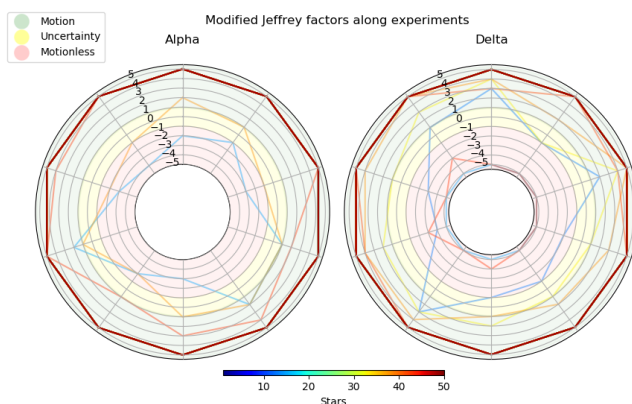


Figura 6.14: Nivel del factor Bayes (coordenada radial) a lo largo de 10 experimentos (coordenada angular). Colección: Sintética con movimientos propios. Configuración: $it = 8000$, $pop = 1500$

sin movimientos propios. No obstante, en la gran mayoría de los experimentos, la predicción se mantiene en la región de factores Bayes negativos, es decir, con evidencia en la ausencia de movimiento propio.

El escenario cambia drásticamente al analizar la colección con movimientos propios (Fig. 6.14). Prácticamente la totalidad de las estrellas tienen valores constantes a lo largo de los experimentos, como se puede ver por la superposición de casi todas las estrellas. Además, al situarse en el nivel 5, se indica que la evidencia de presencia de movimiento propio es muy fuerte, consistente con los datos introducidos. Hay 3 estrellas para las cuales los resultados son variantes, llegando incluso a predecir la ausencia de movimiento en alguna ocasión.

Para analizar la precisión en la clasificación se genera una gráfica que agrupa las precisiones en las estimaciones de entre las 10 ejecuciones del conjunto bajo análisis. Puede verse en la Fig. 6.15 la predicción sobre la colección sin movimientos propios. Solo se tienen TN (verdaderos sin movimiento) y FP (falsos con movimiento), por lo que solo aplica hablar de especificidad, la fracción de estimaciones de ausencia de movimiento (TN) entre los reales sin movimiento (TN + FP). El valor medio de esta especificidad se sitúa en 0,86 para α y 0,84 para δ .

La misma gráfica se muestra para la colección con movimientos propios en la Fig. 6.16. En este caso solo aplica la sensibilidad al no haber referencias sin movimiento, tan solo TP (verdaderos con movimiento) y FN (falsos

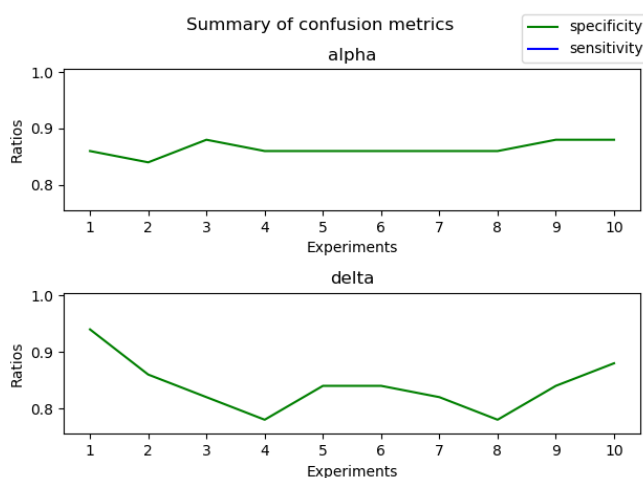


Figura 6.15: Especificidad en la predicción (eje Y) a lo largo de 10 experimentos (eje X). Colección: Sintética sin movimientos propios. Configuración: $it = 8000$, $pop = 1500$

sin movimiento). La sensibilidad es la fracción de predicciones con movimiento (TP) entre los reales con movimiento (TP + FN). En esta gráfica la sensibilidad toma un valor medio de 0.98 para α y 0.95 para δ .

6.1.2.2. Efectos del cambio de configuración

A continuación se analizan los resultados de los 10 experimentos de nuevo sobre las colecciones con movimientos propios y sin movimientos propios, en ambos casos con sesgos y espurios.

Se discute por un lado el efecto de la variación del tamaño de la población y por otro su relación con el número de iteraciones.

Dado un número concreto de iteraciones, si la población es muy grande, la mejora de verosimilitud que proporciona la exploración tarda más en manifestarse. Deben irse seleccionando todos los miembros malos e ir siendo sustituidos paulatinamente. Cuantos más miembros en la población, más se tarda en comenzar la selección de miembros generados en la exploración y menos renovaciones experimentará cada modelo lineal. Por lo tanto una población excesivamente grande puede dificultar la convergencia del algoritmo.

Sin embargo, por otro lado, si el tamaño de la población es demasiado pequeño, es más fácil que los modelos lineales se queden en posiciones estancas. Al haber una población pequeña, a medida que evoluciona el algoritmo

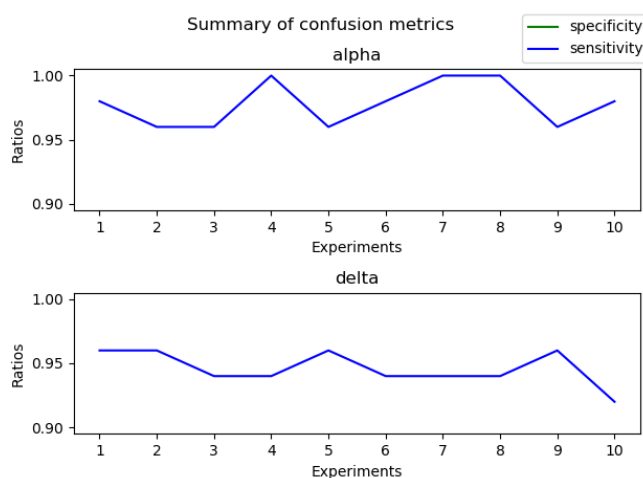


Figura 6.16: Sensibilidad en la predicción (eje Y) a lo largo de 10 experimentos (eje X). Colección: Sintética con movimientos propios. Configuración: $it = 8000$, $pop = 1500$

y las caracterizaciones se aglutinan en una misma región, su diversidad disminuye, dificultando la exploración. Cuando la exploración no consigue dar con un candidato más adaptado, se guarda en la muestra el pivote, si en la población destaca la presencia de unas caracterizaciones concretas, tienen más posibilidad de ser escogidas como pivotes. Los pivotes pueden situarse en posiciones en torno a las cuales no se encuentren mejores individuos (mínimos locales). Tener pocos miembros en la población, es decir, potencialmente poca diversidad de pivotes, hace que sea más fácil que éstos se estanquen y no queden miembros libres de mínimos locales. Puede llegar un punto en que la población está conformada por copias de muy pocos modelos o incluso uno solo, a partir de donde una exploración más profunda va a ser inútil.

Por el razonamiento presentado, se debe escoger un tamaño de población suficientemente grande como para garantizar una diversidad en las caracterizaciones, pero no tan grande que se hagan pocas renovaciones de la población. Se debe garantizar una diversidad en la población y una renovación de la misma. Un aumento del ratio it/pop (en adelante ratio K) lleva a una exploración más profunda y por tanto a unas caracterizaciones de los modelos mejor adaptadas a los datos.

Sin embargo, el número de iteraciones y el tamaño de la población no

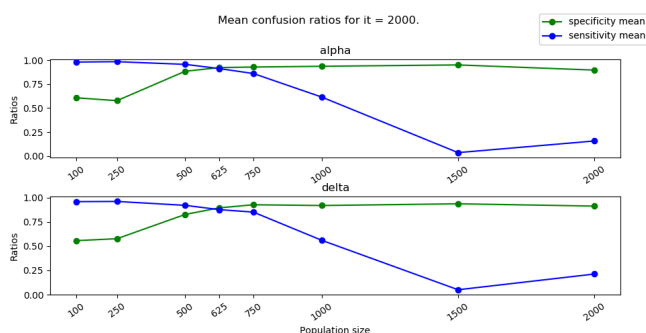


Figura 6.17: Sensibilidad y especificidad medias en la predicción (eje Y), promediadas entre 10 experimentos, variando el tamaño de la población (eje X). Colecciones: Sintética con movimientos propios (azul) y sin movimientos propios (verde). Configuración: $it = 2000$

se pueden escoger arbitrariamente altos ya que tienen una importante implicación en el rendimiento. El aumento del número de iteraciones, implica un considerable coste computacional, no tanto por la duración del algoritmo, que evaluado sobre GPU presenta tiempos ridículamente pequeños, si no por el hecho de estar guardando todos los miembros de la muestra para dar la respuesta. Cada ejecución escala fácilmente al orden de GBs en los resultados y con ello el coste de RAM asociado a manejar la memoria. La gestión de grandes bloques de memoria es uno de los principales problemas de la implementación, imponiendo un cuello de botella en el cómputo.

Se ha analizado por un lado, dejando fijo el número de iteraciones, el cambio de pop con varios valores entre 250 y 3000. Conforme aumenta pop disminuye el ratio K . Se ha repetido este análisis para los números de iteraciones 2000, 4000, 6000, 8000, 10000, se puede consultar el resultado en las figuras 6.17, 6.18, 6.19, 6.20 y 6.21 respectivamente para cada número de iteraciones.

En estas gráficas, cada punto representa la media de la métrica de precisión (especificidad o sensibilidad) a lo largo de 10 experimentos. Se han combinado en cada gráfica los resultados medios sobre las colecciones con movimiento propio (sensibilidad media) y sin movimiento propio (especificidad media).

En todos los análisis se comparte el siguiente comportamiento. Se ha observado como poblaciones muy pequeñas, es decir ratios K altos, aportan valores de sensibilidad igual a 1 en algunos casos, pero al mismo tiempo tie-

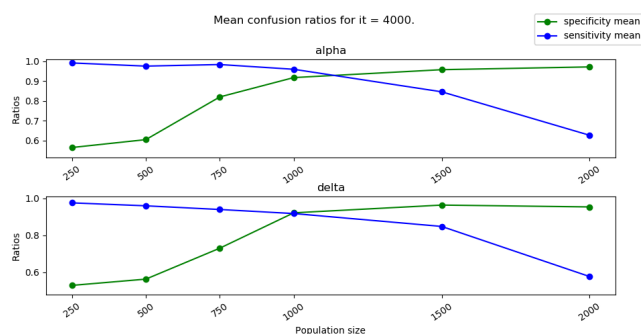


Figura 6.18: Sensibilidad y especificidad medias en la predicción (eje Y), promediadas entre 10 experimentos, variando el tamaño de la población (eje X). Colecciones: Sintética con movimientos propios (azul) y sin movimientos propios (verde). Configuración: $it = 4000$

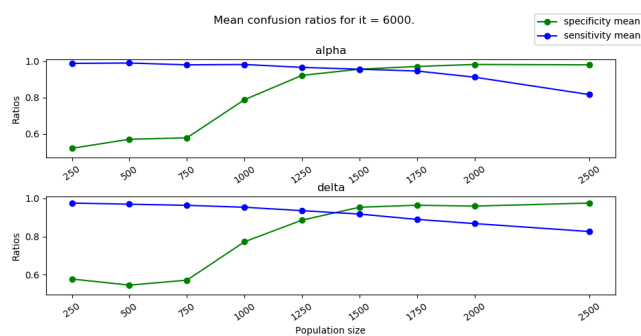


Figura 6.19: Sensibilidad y especificidad en la predicción (eje Y) variando el tamaño de la población (eje X). Colecciones: Sintética con movimientos propios (azul) y sin movimientos propios (verde). Configuración: $it = 6000$

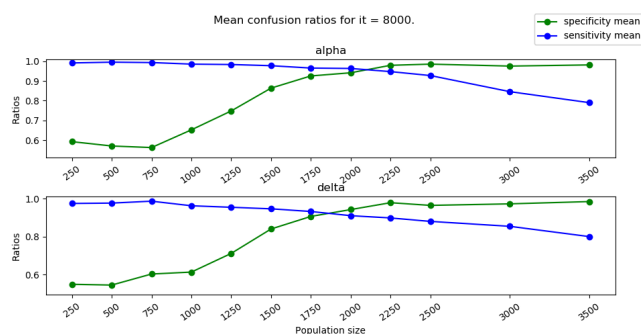


Figura 6.20: Sensibilidad y especificidad medias en la predicción (eje Y), promediadas entre 10 experimentos, variando el tamaño de la población (eje X). Colecciones: Sintética con movimientos propios (azul) y sin movimientos propios (verde). Configuración: $it = 8000$

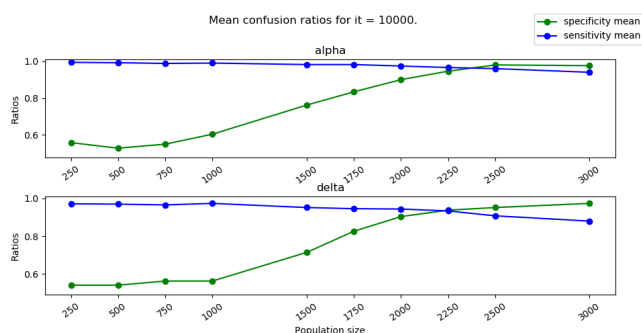


Figura 6.21: Sensibilidad y especificidad medias en la predicción (eje Y), promediadas entre 10 experimentos, variando el tamaño de la población (eje X). Colecciones: Sintética con movimientos propios (azul) y sin movimientos propios (verde). Configuración: $it = 10000$

nen valores de especificidad media de 0,6, es decir, sobre estiman la presencia de movimiento con bastante frecuencia. Por otro lado poblaciones muy grandes, ratios K bajos, aportan altos valores de especificidad media pero bajos de sensibilidad media. Fuera del rango de ratio $K \in [2, 6]$ los resultados no son válidos, bien la especificidad media es muy baja (K altos) o lo es la sensibilidad media (K bajos). Se puede observar a lo largo de estas figuras que un aumento del número de iteraciones, produce una menor reducción de la sensibilidad media ante ratios K más bajos, permitiendo convivir con la región de aumento de especificidad media y alcanzando por tanto unas relaciones altas de especificidad media y sensibilidad media en un abanico más amplio de tamaños de población, siendo así más sencillo escoger unos valores adecuados.

El valor de ratio con cada número de iteraciones que ha dado la mejor relación entre sensibilidad media y especificidad media es aproximadamente 4. Las mejores combinaciones con ratios 3,2, 4, 4, 4 y 4,4 respectivamente son $\{it, pop\} \in \{\{2000, 625\}, \{4000, 1000\}, \{6000, 1500\}, \{8000, 2000\}, \{10000, 2250\}\}$. Se percibe un aumento del ratio con mayores iteraciones.

Se comparan entre sí los análisis de dichos experimentos. El resultado de sensibilidad es consistente entre todos ellos, con 0.95 para α y 0.92 para δ . En el caso de la especificidad media los mejores valores son para $it = 8000$ y $it = 6000$ con 0.95 en α y δ , le sigue $it = 4000$ con 0.94 y 0.92 en sendas coordenadas, $it = 10000$ con 0.91 en ambas y en último lugar $it = 2000$ con 0.9 y 0.81 de especificidad media en α y δ respectivamente.

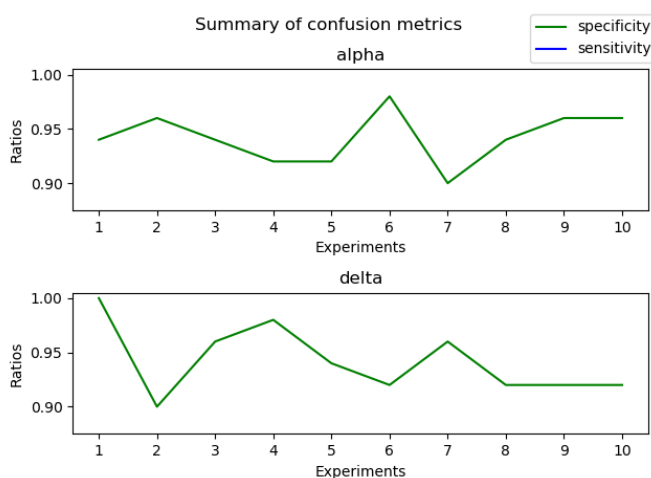


Figura 6.22: Especificidad en la predicción (eje Y) a lo largo de 10 experimentos (eje X). Colección: Sintética sin movimientos propios. Configuración: $it = 8000$, $pop = 2000$

No hay una diferencia significativa entre los mejores resultados con cada it . La sensibilidad media es consistente entre los distintos casos y la especificidad media sí se ve mermada con los números más extremos de iteraciones ($it = 10000$ y $it = 2250$).

Recapitulando, los mejores resultados se obtienen con ratios de iteraciones respecto al tamaño de la población en torno a $4 \pm 0,5$. Valores de este ratio superiores o inferiores producen resultados que sobre estiman la presencia de movimiento propio o su ausencia respectivamente. De entre las distintas iteraciones, los mejores resultados se obtienen con $it = 8000$ y $pop = 2000$ teniendo especificidad media de 0.95 en ambos ejes (Fig. 6.22) y una sensibilidad media de 0.95 y 0.92 en α y δ (Fig. 6.23) respectivamente.

A continuación se hace un análisis del efecto de la presencia de espurios y sesgos en los datos.

6.1.2.3. Efectos de la presencia de espurios y sesgos

Se utilizan los análisis de experimentos con mejor ratio $K = 4$, para comparar el efecto de la inclusión de sesgos y espurios en los datos. Estos análisis se recaban y comparan individualmente para cada valor de it , tratando de distinguir un comportamiento común entre todos ante la inclusión de estas irregularidades. El siguiente comportamiento se comparte entre los

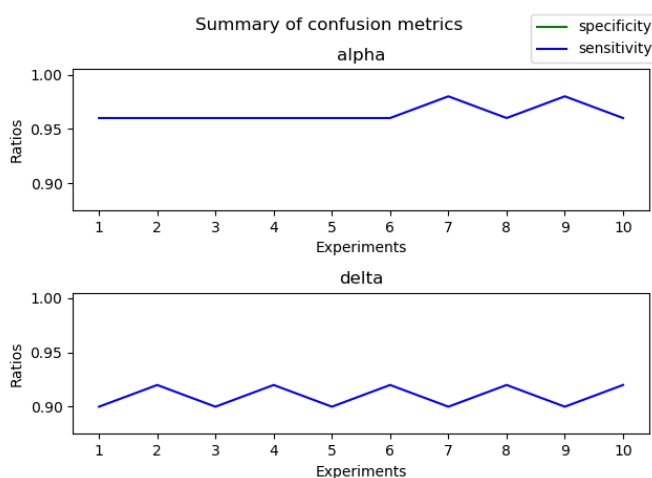


Figura 6.23: Sensibilidad en la predicción (eje Y) a lo largo de 10 experimentos (eje X). Colección: Sintética con movimientos propios. Configuración: $it = 8000$, $pop = 2000$

distintos valores de it evaluados:

En el caso de colecciones sin espurios ni sesgo (ni movimiento propio), el ratio de especificidad de α se mantiene alto (ver Fig. 6.24). Las gráficas correspondientes a otras iteraciones son similares. La especificidad de α no baja de 0,975 y la de δ se mantiene en 0,95. Por su parte la sensibilidad en ambas coordenadas se mantiene perfecta (= 1, se omite la gráfica).

Al añadir sesgos a la colección, se reducen los FP de ambas coordenadas, ver figura 6.25. En α se realiza una estimación casi perfecta (~ 1) y en δ también es elevada, manteniéndose en torno a 0,98. Sin embargo, por otro lado, aumentan los FN, disminuyendo por tanto la sensibilidad un 1% (Fig. 6.26).

En caso de añadir únicamente datos espurios, los FP disminuyen en δ hasta alcanzar especificidades plenas, dando lugar a las mejores estimaciones para esta coordenada (~ 1), pero empeora en α (Fig. 6.27), con caídas de hasta 4 centésimas en la especificidad media respecto a la estimación sobre la colección sin irregularidades. La mejora de especificidad en δ se ve compensada por un considerable aumento de FN con su consiguiente caída en la sensibilidad de $\sim 0,05$ y $\sim 0,07$ en α y δ respectivamente, ver Fig. 6.28.

Finalmente, en caso de añadir tanto espurios como sesgos, la estimación es peor que en la colección sin irregularidades, tanto en sensibilidad como

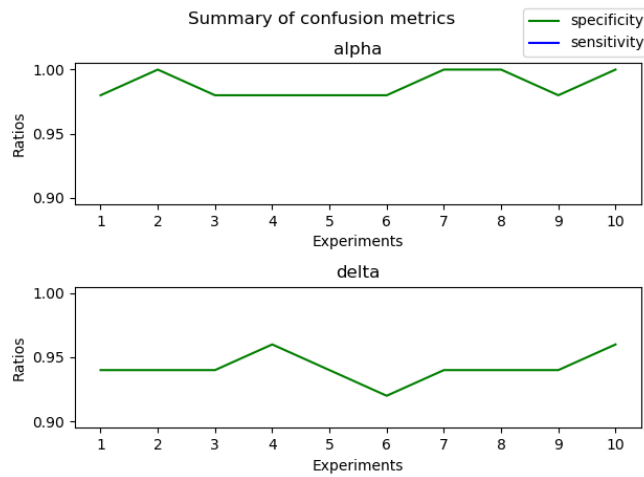


Figura 6.24: Especificidad en la predicción (eje Y) a lo largo de 10 experimentos (eje X). Colección: Sintética sin movimientos propios ni sesgos ni espurios. Configuración: $it = 8000, pop = 2000$

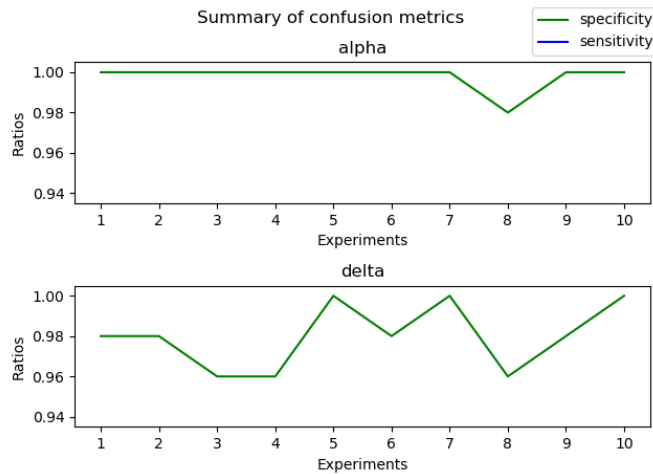


Figura 6.25: Especificidad en la predicción (eje Y) a lo largo de 10 experimentos (eje X). Colección: Sintética sin movimientos propios ni espurios pero sí sesgos. Configuración: $it = 8000, pop = 2000$

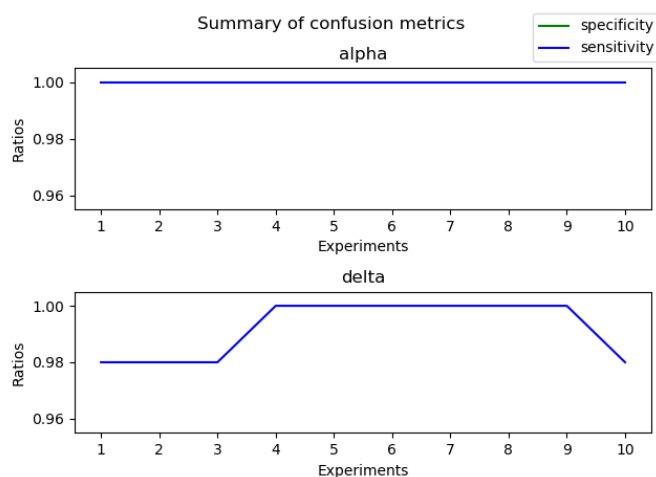


Figura 6.26: Sensibilidad en la predicción (eje Y) a lo largo de 10 experimentos (eje X). Colección: Sintética con movimientos propios sin espurios pero sí sesgos. Configuración: $it = 8000$, $pop = 2000$

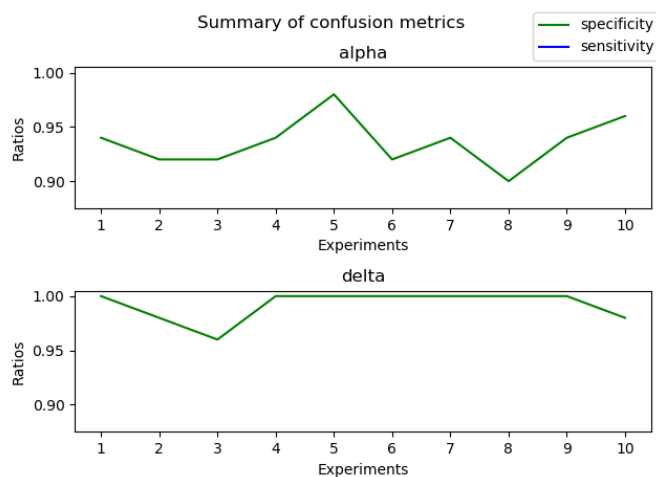


Figura 6.27: Especificidad en la predicción (eje Y) a lo largo de 10 experimentos (eje X). Colección: Sintética sin movimientos propios ni sesgos pero sí espurios. Configuración: $it = 8000$, $pop = 2000$

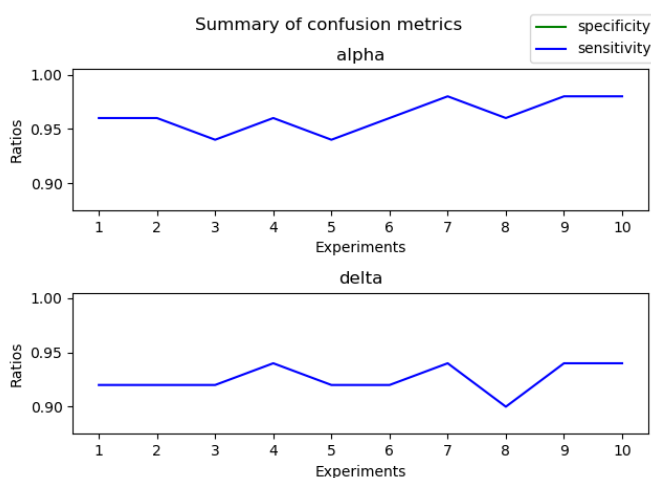


Figura 6.28: Sensibilidad en la predicción (eje Y) a lo largo de 10 experimentos (eje X). Colección: Sintética con movimientos propios sin sesgos pero sí espurios. Configuración: $it = 8000$, $pop = 2000$

en especificidad, en ambos ejes. La especificidad baja principalmente en α ($\sim 4\%$) y solo ligeramente ($\sim 1\%$) en δ (ver Figura 6.22). Por contra, la sensibilidad baja considerablemente en δ ($\sim 9\%$), y un 5% en α (Fig. 6.23).

En resumen, el análisis sobre los datos sintéticos muestra que se debe encontrar un compromiso entre el tamaño de la población y el número de iteraciones para obtener un resultado con bajos errores de estimación. Valores pequeños de esta relación dan lugar a sobre estimación de movimiento propio y valores grandes dan lugar a la sobre estimación de ausencia de movimiento. El compromiso idóneo se sitúa en 4 veces más iteraciones que miembros en la población, con tendencia a aumentar esta relación con volúmenes grandes de iteraciones.

La presencia de cualquier tipo de irregularidad (sesgos o espurios) compromete la sensibilidad del sistema aumentando el número de FN. La presencia de sesgos tiene un impacto menor en la reducción de sensibilidad, compensado con la reducción de FP. La presencia de espurios tiene un impacto importante en el aumento de FN, con mayor efecto en el eje de declinación. Por último, la presencia tanto de espurios como de sesgos provoca un aumento tanto de FP como de FN, en mayor medida que ante cada una de estas irregularidades individualmente, alcanzando bajadas de hasta un 9% en la precisión.

En los escenarios con todas las irregularidades contempladas, el mejor resultado se obtiene con $it = 8000$ y $pop = 2000$, alcanzando precisiones de 0,95 en ambos ejes como se ha comentado anteriormente.

A continuación se hace un análisis de la colección de datos reales del proyecto DANCe.

6.2. Datos reales

Por un lado se muestran las observaciones de una estrella real y los resultados de la inferencia sobre las mismas y por otro un análisis estadístico de los resultados sobre múltiples experimentos con distintas configuraciones.

6.2.1. Análisis de un experimento individual

Se ha seleccionado arbitrariamente una estrella que tenga movimiento en uno de los ejes. La ejecución en la que se ha evaluado ha constado de 6000 iteraciones y una población de 1000 miembros.

En la Fig. 6.29 se muestran las observaciones de la estrella 5235 con los modelos inferidos. Tiene mayor evidencia la presencia de movimiento propio.

La evolución de los parámetros se puede consultar en la Fig. 6.30, destacan dos aspectos de esta gráfica:

Por un lado, la mitad de las iteraciones dan modelos comprendidos en la región final de la exploración, en torno a $\loglikelihood = -70$, mientras que los modelos nominados durante la primera mitad del algoritmo abarcan todo el resto del rango de valores de verosimilitud. Esto puede ser una señal de que sobran iteraciones, hay que encontrar un equilibrio entre el volumen de iteraciones necesario para adquirir la suficiente evidencia y el coste en RAM que supone. Cabe recordar que el número de observaciones reales es por lo general la mitad que el de los datos sintéticos evaluados y en ocasiones bastante menor. Cuanto menor sea el volumen de datos menos iteraciones son necesarias para que el algoritmo converja en una región óptima entre todas ellas.

Por otro lado, se recalca la banda inferior de valores. Se ha visto en gráficas anteriores (Fig. 6.8 y Fig. 6.11), pero no se ha hecho mención de por qué existe esa banda inferior. La aportación de la distribución de espurios en el cálculo de la verosimilitud (p_{bg}) es la responsable de que se forme ese límite inferior, esta probabilidad es independiente del valor de los

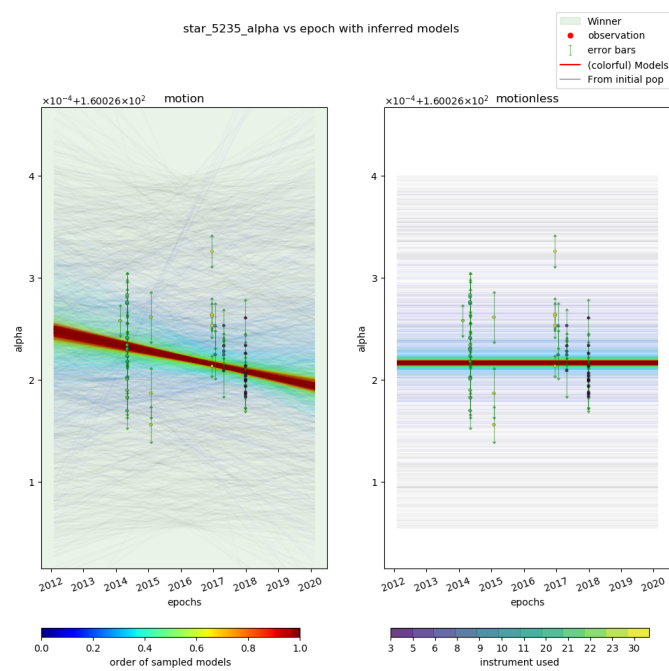


Figura 6.29: Posiciones de una estrella en ascensión recta (eje Y) a lo largo del tiempo (eje X) incluyendo los modelos inferidos. Colección: muestra de DANCe. Configuración: $it = 6000$, $pop = 1000$

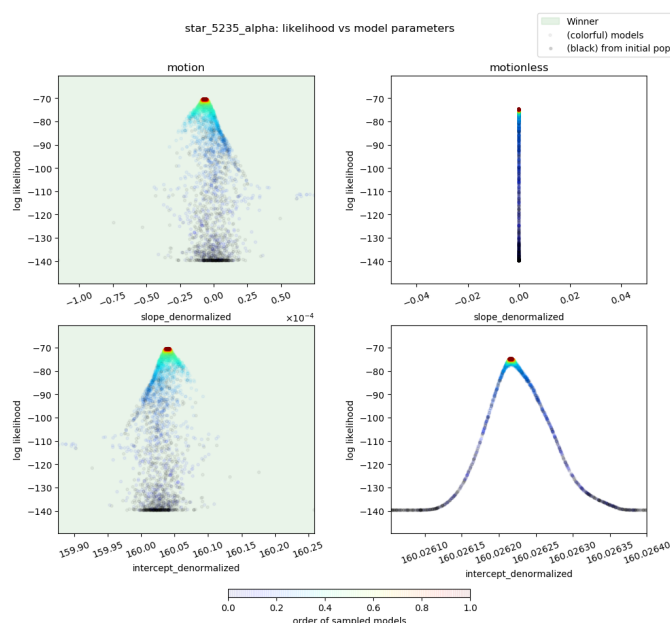


Figura 6.30: Verosimilitud (eje Y) de los modelos lineales en función del valor que tienen (eje X) en cada parámetro, m_{denorm_α} (arriba) y b_{denorm_α} (abajo). La escala cromática indica el orden de muestreo (i.e., la verosimilitud). Colección: muestra de DANCe. Configuración: $it = 6000$, $pop = 1000$

parámetros evaluados. Todas las caracterizaciones tienen un mínimo de verosimilitud por la posibilidad de que los datos sean espurios. La presencia de este término determina el mínimo de verosimilitud, ya que en caso contrario, ciertos modelos que se adapten muy mal, tendrían verosimilitudes tremendamente bajas. Esto ocurre cuando se analiza el algoritmo sin considerar la población de espurios, ciertas caracterizaciones tienen verosimilitudes extremadamente bajas (grandes y negativas), varios órdenes de magnitud por encima de las verosimilitudes de la inmensa mayoría.

En la gráfica de dispersión de los parámetros (Fig. 6.31) se ve aún más claro cómo la mayoría de las iteraciones se agrupan en una región del espacio muy pequeña.

En la distribución a posteriori (Fig. 6.32), se percibe como, a pesar de manejar valores muy pequeños ($\approx 10^{-4}$), la probabilidad final de la pendiente se distingue de 0, prediciendo un sutil movimiento propio. El máximo a posteriori se sitúa en torno a 36 mas/year .

Se procede con el análisis estadístico de los resultados. De igual manera que antes, se evalúan distintas configuraciones, al menos 10 veces cada una,

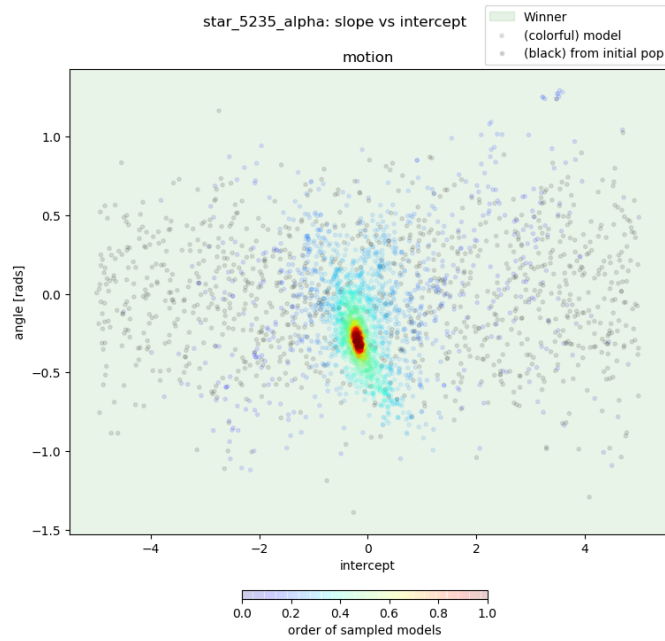


Figura 6.31: Dispersión de valores de los parámetros m (eje Y) frente a b (eje X). Colección: muestra de DANCe. Configuración: $it = 6000, pop = 1000$

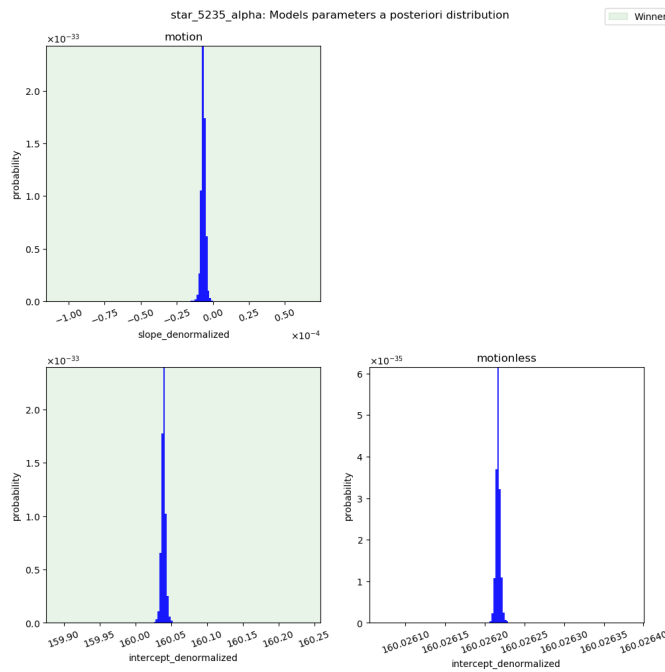


Figura 6.32: Distribución a posteriori de los parámetros, m_{denorm_α} (arriba) y b_{denorm_α} (abajo). Colección: muestra de DANCe. Configuración: $it = 6000, pop = 1000$

de donde sacar métricas de constancia y confusión.

6.2.2. Análisis estadístico de varios experimentos

Se recuerda que la referencia de si una estrella en cuestión tiene movimiento propio, se determina por la relación entre los valores de movimiento propio y error extraídos del propio catálogo consultado. A diferencia del análisis con datos sintéticos, en los que se separaban las referencias con movimiento y sin movimiento, en los catálogos reales conviven estrellas con movimiento claro, otras con sospecha de movimiento bajo una alta incertidumbre y otras tan lejanas que su movimiento es imperceptible a lo largo de siglos.

De cara a hacer la evaluación de la estimación se consideran dos clases entre las estrellas de entrada. Aquellas por encima de la regla de ($PM > 5 \cdot error$) y aquellas por debajo. Podrían etiquetarse estas clases como clase de movimiento propio seguro y clase de movimiento propio desconocido. Una estrella predicha con movimiento propio y perteneciente a la clase de movimiento propio seguro será considerada un TP, en caso de predecirla carente de movimiento se considera un FN. Este valor debe ser el menor posible, ya que es seguro que las estrellas por encima de la regla fijada tendrán movimiento propio perceptible. Respecto a las estrellas de la clase de movimiento propio desconocido, no se sabe si tienen movimiento propio o no. Simplemente por mantener la dinámica de la matriz de confusión, esta clase se etiqueta como “negativa”, de modo que una predicción de movimiento para una estrella de esta región se etiqueta como FP y una predicción de ausencia de movimiento como TN. No obstante, por lo recién comentado, la referencia en esta clase es desconocida, por lo que el análisis de la confusión se limita a la clase “positiva” (con movimiento propio). Muchas de las estrellas de la clase de movimiento propio desconocido sí van a tener un movimiento perceptible, así que se espera un alto número de FP. Lo cual no refleja un error en la predicción por no haber ninguna certeza de la ausencia de movimiento en las estrellas de esta clase. Es decir, la especificidad que se pueda extraer de la matriz de confusión no es válida.

Se ha analizado el efecto de fijar el valor de las iteraciones y se ha variado el tamaño de la población. Se ha repetido el análisis para los distintos valores de las iteraciones, obteniendo unas gráficas de la evolución de los errores de predicción conforme cambia el tamaño de la población. Se pueden consultar

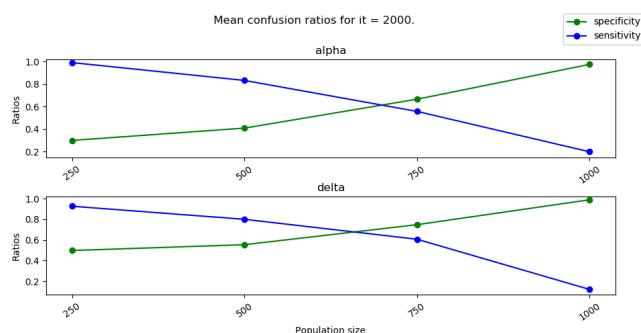


Figura 6.33: Sensibilidad en la predicción (eje Y) variando el tamaño de la población (eje X). Colección: muestra de DANCe. Configuración: $it = 2000$

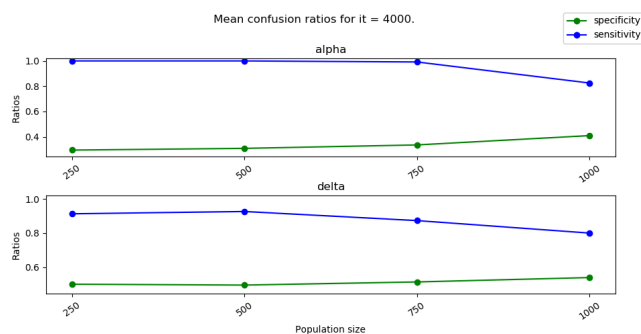


Figura 6.34: Sensibilidad en la predicción (eje Y) variando el tamaño de la población (eje X). Colección: muestra de DANCe. Configuración: $it = 4000$

en las Fig. 6.33, 6.34, 6.35 y 6.36 para 2000, 4000, 6000 y 8000 iteraciones.

En el caso de la declinación hay al menos un FN en todos los experimentos. Con poblaciones de 500 y 750 la predicción de movimiento en el eje de ascensión recta tiene 0 FN en todos los experimentos y con $pop = 250$ también salvo en un experimento con un FN. En la misma línea con el comportamiento observado en el análisis sobre datos sintéticos, las poblaciones más pequeñas dan altos valores de sensibilidad. Aunque en estas gráficas no se pueda apreciar por el desconocimiento de cuándo hay un negativo real, ya se ha mostrado previamente como la alta sensibilidad de ratios K altos viene acompañada de una sobre estimación de la presencia de movimiento propio. Conforme disminuye el ratio K se reduce la sensibilidad y aumenta la especificidad; se estima un menor número de estrellas con movimiento propio a cambio de ser más certero en dicha estimación. El mejor compromiso de sensibilidad sin ceder a la sobre estimación se consigue con $it = 6000$ y

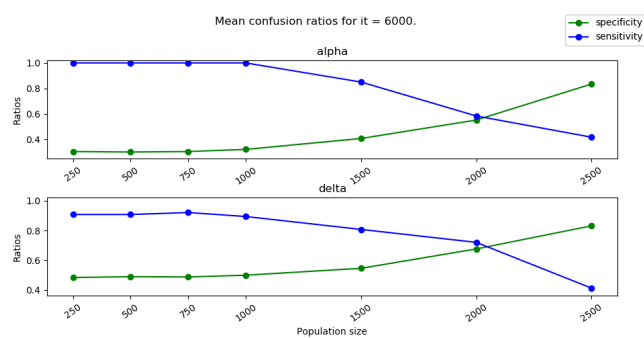


Figura 6.35: Sensibilidad en la predicción (eje Y) variando el tamaño de la población (eje X). Colección: muestra de DANCe. Configuración: $it = 6000$

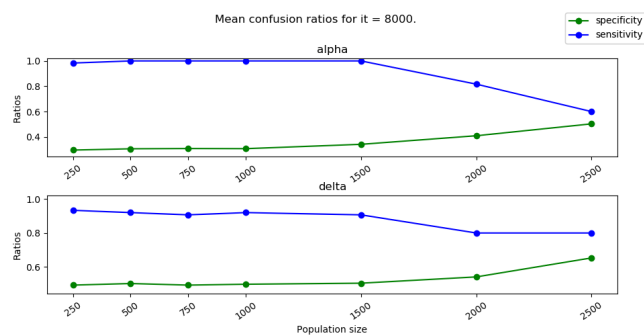


Figura 6.36: Sensibilidad en la predicción (eje Y) variando el tamaño de la población (eje X). Colección: muestra de DANCe. Configuración: $it = 8000$

$pop = 1500$ con una sensibilidad de 0,8.

Zona gris Aparte de las estrellas de la clase con movimiento propio seguro, referida en este análisis como “zona blanca”, se hace una búsqueda de las estrellas que se encuentran en la clase de movimiento propio desconocido. Así como se puede asegurar que con un valor de movimiento propio por encima de $5 \cdot error$, sí hay movimiento propio, no existe un umbral inferior que permita aseverar que una estrella no tiene movimiento propio perceptible por nuestros instrumentos. Se hace una nueva clasificación entre aquellas cuyo valor se comprende en el rango $[2 \cdot error, 5 \cdot error]$ y aquellas fuera de él. Este rango es denominado “zona gris”, en referencia a la incertidumbre sobre si una estrella tiene movimiento propio o no. Se establece el límite inferior en $2 \cdot error$ por acotar la selección entre las que más probablemente tengan movimiento propio.

A efectos de comparación, se hace la consideración de que las estrellas de la zona gris son la clase “positiva” y tienen movimiento propio, de modo que si la estimación lo secunda sería un TP, si la estimación no lo secunda se identificará como un FN. Dado que se desconoce la realidad sobre si tienen o no movimiento propio las estrellas de esta zona, los términos TP y FN hacen referencia a que una estrella de la zona gris ha sido estimada con movimiento y sin movimiento respectivamente.

De igual modo que en el análisis anterior, la clase “negativa” incluye la estrellas con movimiento propio $< 2 \cdot error$, entre las cuáles, si bien con menor frecuencia que en la zona gris, también habrá estrellas con movimiento propio, por lo que se espera la estimación de FPs. Pero en este caso además, la clase negativa incluye todo lo que no pertenece a la “positiva”, por lo que comprende también las estrellas de la “zona blanca”, que serán FPs. Por esta razón, no tiene sentido analizar el valor de FP e igual que en el anterior, este análisis se centra en la sensibilidad.

Se muestran los resultados de la estimación para los dos tipos de clasificaciones contempladas. Para ello se selecciona una de las combinaciones de parámetros que previamente han aportado buenos resultados en la estimación del movimiento propio: $it = 6000$ y $pop = 1500$.

En la figura 6.37 se agrupan las matrices de confusión de la clasificación anterior de zona blanca / no zona blanca (las dos superiores) y la clasificación actual de zona gris / no zona gris (las dos inferiores). Centrándose

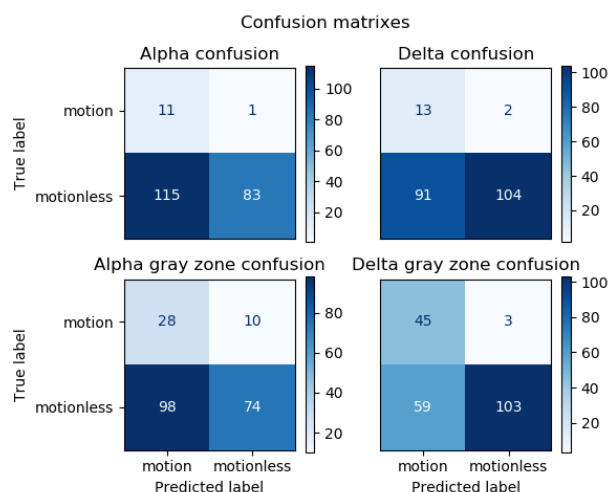


Figura 6.37: Matrices de confusión de las predicciones de movimiento propio (eje X) respecto a la referencia (eje Y) de estrellas en las zonas blanca (arriba) y gris (abajo). Colección: muestra de DANCe. Configuración: $it = 8000$ $pop = 1500$

en las dos inferiores, en la ascensión recta hay 28 estrellas clasificadas como poseedoras de movimiento, con 10 clasificadas sin movimiento. Por su lado en la declinación estos números son 45 y 3 respectivamente.

En la Fig. 6.38 se observa como con $it = 8000$ y $pop = 1500$ el resultado no solo es consistente en la predicción de movimiento o no movimiento a lo largo de los experimentos, si no que además se mantiene muy estable el ratio reportado.

En el estudio sobre datos sintéticos se ha visto, que en los casos en los que no hay movimiento propio, puede haber varios FP, marcados por la sensibilidad y esta puede reducirse a 0.8 (ver Fig. 6.19), por lo que, si bien muchas de las estrellas de la zona gris clasificadas con movimiento serán ciertas, puede haber varias que no lo sean. No obstante, cabe destacar que en el citado análisis, los resultados son muy variables entre sí (el caso con $it = 6000$ y $pop = 1500$ resulta una gráfica muy similar a la Fig. 6.13), lo cual no tiene nada que ver con el análisis actual, donde la constancia de los resultados es casi plena (ver Fig. 6.38). Esta consideración aporta mayor confianza en los resultados.

Por último destacar, que si bien el valor de FP de Fig. 6.37 no es representativo por convivir estrellas de la zona blanca y estrellas con $PM < 2 \cdot error$,

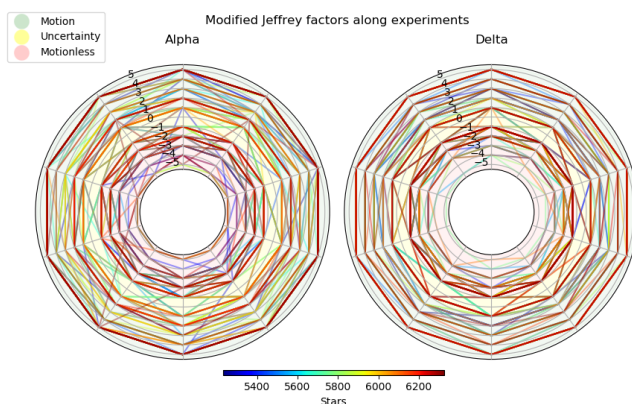


Figura 6.38: Nivel del factor Bayes (coordenada radial) a lo largo de 10 experimentos (coordenada angular). Colección: muestra de DANCe. Configuración: $it = 8000$, $pop = 1500$

el número de estrellas de la zona blanca es conocido gracias a las matrices superiores, en el caso de α son 12 y en δ son 15, de modo que descartando estas estrellas de los FP, se tienen 86 (α) y 44 (δ) estrellas respectivamente, en las que se ha predicho movimiento pero su valor de referencia de movimiento propio es inferior a dos veces su error. Es decir, en muchas de las estrellas más dudosas se predice movimiento propio, y con elevada consistencia entre experimentos. En algunos casos la estimación será fundada y en otros sobre estimada.

En el Anexo B, se muestra a modo de ejemplo una selección de estrellas con TP en la zona gris, en la mayoría de estas estrellas es probable que haya movimiento propio perceptible. Dado el ingente número de estrellas disponibles en los catálogos, conocer cuáles presumiblemente tendrán movimiento propio es de enorme utilidad. Tener esta selección permite focalizar observaciones futuras en estas estrellas con mayor profundidad.

Como resumen se destaca que se ha alcanzado una estimación de datos sintéticos con el 95 % de acierto. Se ha localizado las combinaciones de parámetros que reportan los mejores resultados. Con las mejores combinaciones se han obtenido predicciones de estrellas de datos de DANCe cuya referencia actual no permite aseverar la presencia de movimiento propio o no y se han ratificado las predicciones sobre las estrellas con movimiento propio seguro.

6.3. Escalabilidad

El tiempo de procesado utilizado por la GPU es muy pequeño en comparación con el tiempo que le cuesta ejecutarse al programa entero. La ejecución del algoritmo nested sampling con todas sus iteraciones se realiza en décimas de segundo. Las mayores limitaciones se encuentran en el procesado en CPU y el uso de memoria RAM. El guardado de todos los modelos, uno por iteración y estrella, suponen un uso intensivo de la memoria durante el programa y también en su volcado a disco. El manejo de esta cantidad de memoria ralentiza los tiempos del algoritmo.

Se han identificado los principales cuellos de botella en el cómputo en la carga de los datos, el guardado de la muestra y el cálculo de sus estadísticas. Uno de los puntos de mayor coste computacional es la interpolación para extraer los percentiles ponderados de los modelos, que ha sido necesario implementar en CPU por carencia de las funciones necesarias en GPU.

Se muestra el tiempo consumido por el procesado del algoritmo en GPU y el tiempo total del programa (CPU + GPU) a lo largo de distintas pruebas. Se observa por un lado la variación del tamaño de la población, por otro la variación del número de iteraciones y finalmente la variación del número de estrellas en la colección.

La variación del tamaño de la población se espera que tenga un impacto no muy elevado en GPU, dado que solo influye en el consumo de RAM durante la ejecución del algoritmo y no sobre las fases anterior y posterior, donde se localizan los cuellos de botella mencionados.

En la gráfica superior de la figura 6.39 se observa como la diferencia de tiempo de la GPU a lo largo de distintas iteraciones es baja, aumentando ligeramente con poblaciones más grandes, siendo el rango de resultados de 4,5 *ms*. En el caso del tiempo total del programa (gráfica inferior) es 4 órdenes de magnitud superior y muestra una tendencia incremental conforme aumenta el tamaño de la población, no obstante la diferencia entre $pop = 250$ y $pop = 3000$ son 1,25 minutos, por tanto no tiene una influencia significativa.

Se muestra en la figura 6.40 el aumento de tiempo consumido fijando la población en 250 y variando el número de iteraciones. Se observa como el aumento del tiempo de programa es linealmente proporcional al número de iteraciones. Por esto se debe encontrar un valor razonable que permita una correcta estimación sin comprometer las capacidades de memoria de la

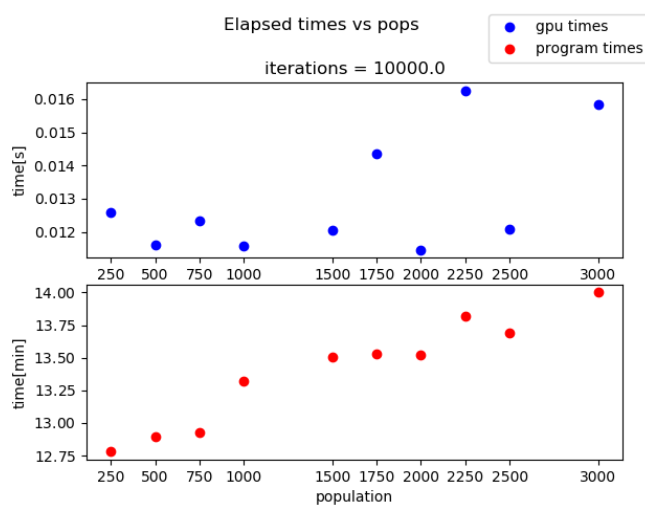


Figura 6.39: Tiempo consumido (eje Y) por la GPU (arriba) y por todo el programa (abajo) a lo largo de experimentos con distinto tamaño de población (eje X). Colección: Sintéticos con movimiento. Configuración: $it = 10000$

máquina donde se ejecute. Por otro lado el uso de distintas iteraciones no tiene un impacto significativo en el uso de GPU debido a los bajos tiempos que se manejan.

Respecto a la variación del tamaño de la colección. Por el momento no se permiten grandes volúmenes por colapso de la memoria de la GPU, un contratiempo que se va a solventar con la integración del algoritmo en el clúster de computación del proyecto DANCe. Se muestran los tiempos consumidos al evaluar colecciones de distinto tamaño (Fig. 6.41). Las variaciones son pequeñas pero se puede observar que se mantiene el comportamiento creciente lineal del tiempo del programa y una aparente baja influencia en el tiempo de GPU.

En el momento de integrar la herramienta desarrollada en el cluster del proyecto DANCe, donde se tienen mayores recursos, se podrá hacer uso del potencial de la GPU sin estar tan limitados por el uso de memoria.

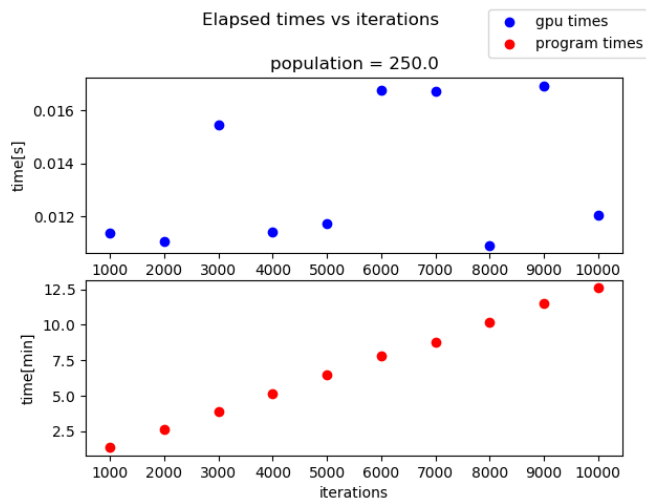


Figura 6.40: Tiempo consumido (eje Y) por la GPU (arriba) y por todo el programa (abajo) a lo largo de experimentos con distinto número de iteraciones (eje X). Colección: Sintéticos con movimiento. Configuración: $pop = 250$.

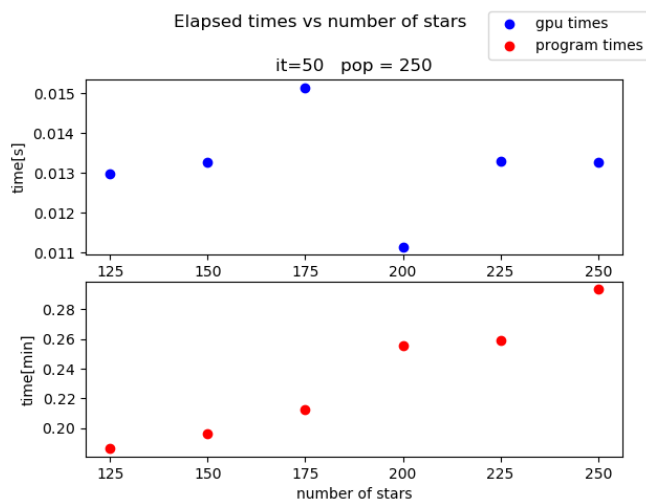


Figura 6.41: Tiempo consumido (eje Y) por la GPU (arriba) y por todo el programa (abajo) a lo largo de experimentos con distinto número de estrellas (eje X). Colección: Sintéticos con movimiento. Configuración: $it = 50$ $pop = 250$.

Capítulo 7

Conclusiones y líneas de futuro

7.1. Conclusiones

- La estimación de los movimientos propios tradicionalmente se ha realizado mediante métodos del paradigma frecuentista. En los últimos años el paradigma bayesiano presenta un amplio crecimiento en la astrofísica y este mismo año se ha realizado una ampliación de los movimientos propios estimados de GAIA utilizando estos métodos.
- Se ha desarrollado íntegramente una herramienta para la inferencia de los movimientos propios mediante el uso de modelos bayesianos jerárquicos.
- Se adquiere robustez ante datos espurios al usar un modelo generativo mixto y ante datos sesgados al usar la jerarquía bayesiana.
- Para decidir qué modelo es más plausible que haya generado los datos, bien uno con movimiento propio u otro sin él, se realiza una comparación entre sus respectivas evidencias. Para obtenerla se implementa el algoritmo nested sampling, que de forma colateral permite obtener una muestra de la distribución a posteriori buscada.
- La variación del número de iteraciones y el tamaño de la población juega un papel fundamental en el éxito en la estimación. Ante un número mayor de iteraciones, los resultados tienen menores errores y son más consistentes entre experimentos.

- Un aumento del número de iteraciones debe venir acompañado de un aumento del tamaño de la población, asegurando que el ratio entre el primero y el segundo se encuentra en torno a $4 \pm 0,5$. El ratio óptimo aumenta ligeramente con el tamaño de la población.
- Con las configuraciones adecuadas se alcanzan sensibilidades y especificidades de 0.95 en ambos ejes α y δ , evaluado sobre la colección de datos sintéticos, y sensibilidades de 0.8 sobre colecciones de datos reales.
- La región de muestreo se restringe por la verosimilitud del último elemento muestreado. Si el tamaño de la población se acerca al número de iteraciones la restricción del espacio de búsqueda en una región pequeña no llegará a hacerse efectiva. Por contra, si la población es demasiado pequeña en comparación con las iteraciones, tiene mayor facilidad para que sus miembros se estanquen en mínimos locales.
- Si la exploración no es exitosa, el elemento de la población elegido como pivote se duplica, reduciendo la diversidad de la población. La falta de diversidad favorece que el algoritmo converja en mínimos locales y tiene mayor impacto en poblaciones pequeñas.
- Se han procesado observaciones de estrellas reales extraídas del proyecto Cosmic DANCe, validando el movimiento propio de referencia y proponiendo nuevas estrellas con movimiento entre aquellas en las que la referencia no es clara.
- La implementación se ha realizado sobre GPU aprovechando la máxima paralelización posible, al ejecutar cada núcleo individualmente el algoritmo entero, lo que permite procesar tantas estrellas a la vez como núcleos tenga la GPU.
- El tiempo de cómputo invertido en la GPU es bajo, es posible sacarle mucho potencial. El elevado consumo de RAM supone un cuello de botella en el procesamiento. Queda pendiente habilitar el procesamiento de volúmenes grandes de estrellas.
- La herramienta implementada va dirigida a la comunidad científica y en especial a su integración en el proyecto Cosmic DANCe. Por lo que

se liberará para aprovechar el potencial de la inferencia bayesiana y las GPUs en el aumento del conocimiento científico.

- Se han implementado dos módulos para la visualización y análisis de los resultados por correspondencia con una referencia dada que facilitan el estudio de las capacidades de la herramienta.

7.2. Líneas de futuro

- Implementación de la herramienta en el cluster de computación de Cosmic DANCe para la paralelización simultánea a través de múltiples GPUs.
- Reducir el uso de RAM, por ejemplo, mediante escritura directa a disco o la separación de la estimación de los modelos con movimiento propio y sin movimiento propio en distintas ejecuciones.
- Utilizar la herramienta para procesar los datos de millones de estrellas disponibles en el proyecto Cosmic DANCe.
- Liberar el código y mantener un plan de soporte de la herramienta.
- El uso de esta herramienta permite explorar el rendimiento de las técnicas bayesianas sobre un marco eficiente como es el de las GPUs, facilitando realizar investigaciones profundas que den lugar a comparaciones sólidas entre configuraciones del algoritmo o extensiones del mismo.

Capítulo 8

Bibliografía

- Bayes, T. (1763). LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, FRS communicated by Mr. Price, in a letter to John Canton, AMFR S. *Philosophical transactions of the Royal Society of London*, (53):370–418.
- Benomar, O. (2012). Asteroseismology: Bayesian analysis of solar-like oscillators. In *Statistical Challenges in Modern Astronomy V*, pages 171–176. Springer.
- Bertin, E. (2006). Automatic astrometric and photometric calibration with SCAMP. In *Astronomical Data Analysis Software and Systems XV*, volume 351, page 112.
- Bolker, B. M. (2008). *Ecological models and data in R*. Princeton University Press.
- Bouy, H., Bertin, E., Moraux, E., Cuillandre, J.-C., Bouvier, J., Barrado, D., Solano, E., and Bayo, A. (2013). Dynamical analysis of nearby clusters—Automated astrometry from the ground: precision proper motions over a wide field. *Astronomy & Astrophysics*, 554:A101.
- Brewer, B. J. (2012). Bayesian Analysis of Reverberation Mapping Data. In *Statistical Challenges in Modern Astronomy V*, pages 189–195. Springer.
- Brown, W. R., Anderson, J., Gnedin, O. Y., Bond, H. E., Geller, M. J., and Kenyon, S. J. (2015). Proper Motions and Trajectories for 16 Extreme Runaway and Hypervelocity Stars. *The Astrophysical Journal*, 804(1):49.

- Budavári, T. (2012). On Statistical Cross-Identification in Astronomy. In *Statistical Challenges in Modern Astronomy V*, pages 291–302. Springer.
- Carlin, B. P. and Louis, T. A. (2010). *Bayes and empirical Bayes methods for data analysis*. Chapman and Hall/CRC.
- Clark, J. S. (2005). Why environmental scientists are becoming Bayesians. *Ecology letters*, 8(1):2–14.
- Colom, A. A. (2005). *Hablando de física a la salida del cine*. Universitat Illes Balears.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22.
- Feigelson, E. and Babu, J. (2012a). *Statistical Challenges in Modern Astronomy V*. Lecture Notes in Statistics. Springer New York.
- Feigelson, E. D. and Babu, G. J. (2012b). *Modern statistical methods for astronomy: with R applications*. Cambridge University Press.
- Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence*, (6):721–741.
- Glusker, J. P., Lewis, M., Rossi, M., and Muller, G. (1995). *Crystal structure analysis for chemists and biologists*, volume 34. Weinheim/Bergstr.: Verlag Chemie; New York: Academic Press, c1962-.
- Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications.
- Hobert, J. P. and Casella, G. (1996). The effect of improper priors on Gibbs sampling in hierarchical linear mixed models. *Journal of the American Statistical Association*, 91(436):1461–1473.
- Høg, E., Fabricius, C., Makarov, V. V., Bastian, U., Schwkendiek, P., Wicenec, A., Urban, S., Corbin, T., and Wycoff, G. (2000a). Construction and verification of the Tycho-2 Catalogue. *Astronomy and Astrophysics*, 357:367–386.

- Høg, E., Fabricius, C., Makarov, V. V., Urban, S., Corbin, T., Wycoff, G., Bastian, U., Schwekendiek, P., and Wicenec, A. (2000b). The Tycho-2 Catalogue of the 2.5 million brightest stars. Technical report.
- Hogg, D. W., Bovy, J., and Lang, D. (2010). Data analysis recipes: Fitting a model to data. *arXiv:1008.4686 [astro-ph.IM]*.
- Jeffreys, H. (1998). *The theory of probability*. OUP Oxford.
- Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the american statistical association*, 90(430):773–795.
- Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P. (1983). Optimization by simulated annealing. *science*, 220(4598):671–680.
- Kruschke, J. K. (2013). Bayesian estimation supersedes the t test. *Journal of Experimental Psychology: General*, 142(2):573.
- Lang, D., Hogg, D. W., Jester, S., and Rix, H.-W. (2009). Measuring the undetectable: Proper motions and parallaxes of very faint sources. *The Astronomical Journal*, 137(5):4400.
- Laplace, P. S. (1820). *Théorie analytique des probabilités*, volume 7. V. Courcier.
- Lodieu, N., Deacon, N. R., and Hambly, N. C. (2012). Astrometric and photometric initial mass functions from the UKIDSS Galactic Clusters Survey–I. The Pleiades. *Monthly Notices of the Royal Astronomical Society*, 422(2):1495–1511.
- Loredo, T. J. (1990). From Laplace to supernova sn 1987a: Bayesian inference in astrophysics. In *Maximum entropy and bayesian methods*, pages 81–142. Springer.
- Loredo, T. J. (1992). Promise of Bayesian inference for astrophysics. In *Statistical challenges in modern astronomy*, pages 275–297. Springer.
- Loredo, T. J. (1994). The return of the prodigal: Bayesian inference for astrophysics. *Bayesian Statistics*, 5:121.
- Loredo, T. J. (2013). Bayesian astrostatistics: a backward look to the future. In *Astrostatistical challenges for the new astronomy*, pages 15–40. Springer.

- Lu, Y., Mo, H. J., Weinberg, M. D., and Katz, N. (2012). Bayesian Predictions from the Semi-analytic Models of Galaxy Formation. In *Statistical Challenges in Modern Astronomy V*, pages 523–525. Springer.
- Lyne, A. G., Anderson, B., and Salter, M. J. (1982). The proper motions of 26 pulsars. *Monthly Notices of the Royal Astronomical Society*, 201:503–520.
- Metropolis, N. and Ulam, S. (1949). The monte carlo method. *Journal of the American statistical association*, 44(247):335–341.
- Olivares, J., Sarro, L. M., Moraux, E., Berihuete, A., Bouy, H., Hernández-Jiménez, S., Bertin, E., Galli, P. A. B., Huelamo, N., Bouvier, J., and Barrado, D. (2018). "The seven sisters DANCe IV. Bayesian hierarchical model". *Astronomy & Astrophysics*, 617:A15.
- Perryman, M. A. C. et al. (1997). The HIPPARCOS and TYCHO catalogues. *star*, 2:2pil.
- Röser, S. and Bastian, U. (1991). *PPM Star Catalogue. Positions and proper motions of 181731 stars north of -2.5 degrees declination for equinox and epoch J2000. 0. Vol. I: Zones + 80° to + 30°. Vol. II: Zones + 20° to -0°.*
- Röser, S., Schilbach, E., Schwan, H., Kharchenko, N. V., Piskunov, A. E., and Scholz, R.-D. (2008). PPM-Extended (PPMX) – a catalogue of positions and proper motions. *Astronomy & Astrophysics*, 488(1):401–408.
- Ruiz-Benito, P., Andivia, E., A. J., Astigarraga, J., Barrientos, R., Cruz-Alonso, V., Florencio, M., Gómez, D., Martínez-Baroja, L., Quiles, P., Rohrer, Z., Santos, A., Velado, E., Villén-Pérez, S., and Morales-Castilla, I. (2018). Ventajas de la estadística bayesiana frente a la frecuentista: ¿por qué nos resistimos a usarla? *Revista Ecosistemas*, 27(2):136–139.
- Sharma, S. (2017). Markov Chain Monte Carlo methods for Bayesian data analysis in astronomy. *Annual Review of Astronomy and Astrophysics*, 55:213–259.
- Skilling, J. (2006). Nested sampling for general Bayesian computation. *Bayesian analysis*, 1(4):833–859.

- Tak, H., Ghosh, S. K., and Ellis, J. A. (2018). How proper are Bayesian models in the astronomical literature? *Monthly Notices of the Royal Astronomical Society*, 481(1):277–285.
- Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distributions by Data Augmentation. *Journal of the American statistical Association*, 82(398):528–540.
- Tanner, M. A. and Wong, W. H. (2010). From EM to Data Augmentation: the emergence of MCMC Bayesian computation in the 1980s. *Statistical science*, 25(4):506–516.
- Tian, H. J., Xu, Y., Liu, C., Rix, H. W., Sesar, B., and Goldman, B. (2020). The extended Gaia-PS1-SDSS (GPS1+) proper motion catalog. *arXiv:2004.12641 [astro-ph.GA]*.
- VanderPlas, J. (2014). Frequentism and Bayesianism: A Python-driven primer. *arXiv:1411.5018 [astro-ph.IM]*.

Apéndice A

Manual de uso

El programa puede descargarse del repositorio utilizado para su desarrollo: <https://github.com/luis-sarro/ppmatic>. ¹

Se recomienda ejecutar el programa en un entorno virtual de Anaconda. Se proporciona en el código el listado de dependencias y la función del Makefile “make env-create” para la creación del entorno de Anaconda, tan solo es necesario tener instalado conda para su ejecución.

El programa se inicia al ejecutar el fichero *astronest_app.py*. Se requiere proporcionar como argumentos de entrada:

1. La ruta al fichero de configuración.
2. La ruta al directorio del espacio de trabajo
3. Nombre del fichero FITS que contenga los datos de las estrellas y esté localizado en la carpeta *datasets* del espacio de trabajo.

El espacio de trabajo sigue una estructura determinada que debe conservarse, debe constar de las carpetas *datasets* y *results*, estando los ficheros FITS en la primera.

Este fichero debe contener una tabla con al menos los siguientes parámetros:

1. **SOURCE_NUMBER**: identificador de la estrella.

¹Actualmente el repositorio es privado a la espera de su integración en el cluster del proyecto DANCe. Se ha creado un usuario temporal con acceso al mismo para uso del tribunal de evaluación del proyecto. usuario: unedforgithub@gmail.com contraseña: astronestUNED1

2. ALPHA_J2000: coordenada de ascensión recta en grados.
3. DELTA_J2000: coordenada de declinación en grados.
4. ERRA_WORLD: error en la observación de ascensión recta.
5. ERRB_WORLD: error en la observación de declinación.
6. EPOCH: instante de captura de la observación en años.
7. ASTR_INSTRUM: identificador del instrumento utilizado para la observación.

Si se denomina `REPO_PATH` a la carpeta contenedora del programa, *astronest*, y los espacios de trabajo, *workspaces*, y se considera que existe un *workspace* llamado *main_workspace* y en su carpeta *datasets* hay un fichero *dataset_name.fits*, el programa se ejecuta con la siguiente instrucción:

```
PYTHONPATH=$REPO_PATH/astronest \
python $REPO_PATH/astronest/astronest_app.py \
--config $REPO_PATH/astronest/config.yml \
--workspace $REPO_PATH/workspaces/main_workspace \
--dataset dataset_name.fits
```

Se permite también la ejecución desde una sesión de Python importando la función principal del módulo *astronest_app.py* y aportándole los argumentos en una lista.

La configuración del programa se debe indicar a partir de un fichero YAML. Se permite el cambio de ciertos parámetros de la configuración también mediante argumentos de entrada. Los de uso más habitual son el número de iteraciones (*-it*), el tamaño de la población (*-pop*), el número de intentos de exploración (*-e*), si generar el fichero de resultados completo (*-l*) o si generar las gráficas de resultados para cada estrella al terminar (*-all*).

Para más detalles sobre los parámetros de configuración, los argumentos de entrada, las características del espacio de trabajo y el contenido de los ficheros de resultados se dirige al lector a la documentación del programa disponible en el repositorio.

Apéndice B

Propuesta de estrellas con movimiento propio

En este apéndice se lista un conjunto de estrellas de las que actualmente no se tiene una certeza plena de si tienen movimiento propio o no; las contenidas en la denominada “zona gris”. De entre todas las predichas con movimiento propio por el algoritmo, se citan solo aquellas que tienen ratios de Bayes que aporten una evidencia sustancial ($\log \text{ratio} > 3,2$ de acuerdo a la Tabla [5.1](#))

Identificador	Log ratio en α	Log ratio en δ
5233	13.7	-
5260	4.4	-
5293	-	3.8
5303	18	-
5233	13.7	-
5316	12.5	5.5
5334	10.3	6.1
5405	3.6	-
5442	7.3	-
5480	-	7.2
5728	-	4.6
5917	4.1	-
5946	7.1	-
5958	6.8	-
5975	-	4.8
5986	4.8	3.6
6056	-	3.9
6061	23.6	-
6120	-	5.7
6166	10.4	-
6189	6	-
6304	3.4	-
6319	22.6	-

Tabla B.1: Listado de estrellas de una muestra del catálogo DANCe, sobre las cuales se estima la presencia de movimientos propios. Se omiten los ratios en caso de que dicho eje no haya sido estimado con movimiento propio, o no con la suficiente evidencia sobre la hipótesis contraria.