

Índice general

1. Introducción al análisis de datos funcionales	1
1.1. Desarrollo del curso	1
1.2. Aprender más sobre FDA	3
1.3. Bibliografía	3

Capítulo 1

Introducción al análisis de datos funcionales

En muchos campos científicos, tales como en las ciencias económicas, medioambientales, salud, etc. aparecen magnitudes (temperaturas, grado de concentración polínica, cotizaciones bursátiles, nivel de estrés, etc.) caracterizadas por la evolución de una variable a lo largo del tiempo (proceso estocástico), de modo que los datos de que se dispone son funciones en lugar de los vectores del análisis multivariante clásico. La imposibilidad de medir la mayoría de estas variables continuamente en el tiempo, unida a la complejidad teórica de muchos de los métodos estadísticos disponibles para su análisis llevan a que se manejen resúmenes periódicos que constituyen las series temporales contenidas normalmente en los anuarios estadísticos. Aunque existen muchas técnicas para la modelización y predicción de datos temporales discretos, la mayoría de ellas, como por ejemplo la teoría clásica de Box-Jenkins (1970), imponen que se verifiquen hipótesis bastante restrictivas como estacionariedad, observaciones igualmente espaciadas o pertenencia a una clase de procesos específica.

Estos problemas han planteado la necesidad de desarrollar metodologías potentes que permitan, en la práctica, la modelización y predicción a partir de datos funcionales dando lugar a una especialidad estadística muy reciente conocida con el nombre de Análisis de Datos Funcionales (FDA). Esta terminología ha sido acuñada gracias a los libros de Ramsay y Silverman (1997, 2002, 2005), que son el referente bibliográfico para cualquiera interesado en FDA. A esta línea de investigación pertenecen los profesores que imparten este curso, que han formulado en los últimos años modelos teóricos basados en la técnica de reducción de dimensión Análisis en Componentes Principales Funcional (FPCA), que han sido aplicados con éxito en distintos campos (turismo, bolsa, medioambiente) en cumplimiento de los objetivos de distintos proyectos de investigación financiados.

1.1. Desarrollo del curso

El primer problema que hay que abordar cuando se trabaja con datos funcionales es la reconstrucción de la verdadera forma funcional de las trayectorias muestrales a partir de observaciones en un conjunto finito de instantes de

tiempo que pueden ser diferentes para los distintos individuos muestrales. Para solucionar este problema se recurre a técnicas de aproximación numérica de funciones que serán estudiadas en el Tema 2 del curso.

El Tema 3 está dedicado a la generalización de la técnica de Análisis en Componentes principales al caso funcional (FPCA). La estimación de las componentes principales a partir de funciones muestrales independientes del proceso, conlleva problemas de aproximación matemática de las soluciones de una ecuación integral de núcleo la función de covarianza muestral del proceso, que se complican, aún más, cuando en la práctica se dispone sólo de observaciones en tiempo discreto de las curvas muestrales del proceso. Una primera solución se obtiene realizando un ACP clásico de los datos discretos observados. Hay muchos ejemplos que ponen de manifiesto que los resultados así obtenidos no son robustos en el caso de observaciones desigualmente espaciadas. En este tema se propondrán métodos numéricos de aproximación de las componentes principales cuya eficiencia y precisión serán puestas de manifiesto mediante simulaciones y aplicaciones sobre datos reales de turismo y bolsa.

Una vez resuelto el problema de aproximación del FPCA a partir de observaciones discretas de una muestra de funciones muestrales, en el Tema 3 nos planteamos aplicar el FPCA para resolver el problema de predicción de un proceso estocástico en tiempo continuo en el futuro a partir de su pasado reciente. Para ello se estudiarán los modelos PCP de predicción lineal en componentes principales como una extensión de la Regresión en Componentes Principales al caso en que se quiere predecir una variable (el proceso en el futuro) en función de las componentes principales asociadas a un número infinito de predictores (variables del proceso en el pasado). Formularemos además los modelos PCP ponderados como una adaptación de los modelos PCP para predecir la evolución de una serie temporal en un periodo continuo de tiempo a partir de observaciones discretas en el pasado, proponiendo además un criterio óptimo de selección de las componentes principales basado tanto en su varianza explicada como en la correlación entre componentes del pasado y del futuro. Se estudiarán aplicaciones de estos modelos para predecir la evolución de las cotizaciones en bolsa de Madrid del grupo banca.

En el Tema 4 se introducirán los modelos de regresión lineal funcional para explicar una variable de respuesta a partir de un predictor funcional relacionado. Se estudiará tanto el caso de respuesta escalar como el de respuesta funcional, y se establecerá su relación con los modelos PCP.

Finalmente, el Tema 6 aborda el problema de predecir una variable binaria a partir de la evolución temporal de una variable continua (predictor funcional). Para resolverlo se estudiarán modelos logit funcionales de respuesta binaria de una gran utilidad práctica puesto que son muchos los campos de aplicación en los que aparecen datos de este tipo. En el área de Ciencias de la Salud y en Epidemiología es prioritario determinar como uno o varios factores de riesgo inciden en la probabilidad de desarrollar una determinada enfermedad. El problema de estimación de estos modelos se resolverá de nuevo usando como predictores las componentes principales del predictor funcional. Para ilustrar estos modelos utilizaremos datos temporales de evolución del grado de estrés en una muestra de pacientes de la enfermedad autoinmune LUPUS para modelizar la probabilidad de tener un brote en esta enfermedad.

1.2. Aprender más sobre FDA

Para completar esta introducción sobre FDA el alumno debe consultar el sitio web sobre FDA mantenido por el profesor J. Ramsay que es uno de los investigadores de más prestigio en este tema a nivel mundial

<http://ego.psych.mcgill.ca/misc/fda/index.html>.

En esta página web el alumno interesado en aprender más sobre este tema puede encontrar

- Ejemplos sobre como funcionan las técnicas de FDA en la solución de problemas con datos reales.
- Bibliografía tanto para expertos como usuarios de FDA sobre investigaciones relevantes en este tema.
- Software em Mathlab, R o S-Plus para llevar a cabo sus propios análisis con datos funcionales.
- Conjuntos de diferentes tipos de datos reales de tipo funcional.
- Información sobre congresos y seminarios sobre FDA en el mundo.

Información complementaria sobre ejemplos, objetivos y primeros pasos en FDA se puede encontrar en el Capítulo 1 del libro de Ramsay y Silverman (1997, 2005).

Una amplia gama de ejemplos de distintos tipos de análisis de datos funcionales con datos reales puede verse en el libro de Ramsay y Silverman (2002) sobre FDA aplicado.

1.3. Referencias bibliográficas

- **Ramsay J.O. y Silverman B.W. (1997, 2005):** Functional Data Analysis. (First and Second editions). Springer-Verlag.
- **Ramsay J.O. y Silverman B.W. (2002):** Applied Functional Data Analysis. Springer-Verlag.