



## 1. DATOS BÁSICOS DEL TFG:

**Título:** TÉCNICAS DE PROBABILIDAD EN LA GENERACIÓN DE DATOS EN FENÓMENOS COMPLEJOS

**Descripción general** (resumen y metodología):

### RESUMEN:

El TFG comienza con una breve introducción y revisión sobre los contenidos relativos a vectores aleatorios analizados en la asignatura de Probabilidad de segundo del Grado de Matemáticas, contemplando su extensión a nuevas familias de distribuciones multidimensionales en el plano discreto y continuo. Se implementarán técnicas de simulación estocástica de dichos vectores (basadas en distribuciones paramétricas y no paramétricas) para la obtención de datos sintéticos. Posteriormente, se abordará la generación de datos sintéticos mediante modelos de regresión no lineal, basados en modelos físicos y biológicos. Esta técnica de generación de datos es bastante potente, siendo utilizada en múltiples disciplinas para simular y analizar fenómenos complejos. (Nótese que este aspecto requiere la ampliación de los contenidos impartidos en el Tema 4 de la asignatura de Probabilidad del Grado de Matemáticas). Posteriormente, se extenderán dichas técnicas de generación al contexto más general de los sistemas dinámicos a partir de la Teoría básica de Procesos Estocásticos, tras una breve revisión bibliográfica. En particular, se generarán modelos clásicos de procesos estocásticos univariantes y multivariantes, que nuevamente permitirán la obtención de datos sintéticos en sistemas aleatorios gobernados por ecuaciones de evolución de la Física y la Biología, incluyendo la modelización de sistemas de partículas.

### METODOLOGÍA:

La metodología se basa en los conocimientos teóricos sobre vectores aleatorios y distribuciones condicionadas adquiridos en la asignatura de Probabilidad, incluyendo la regresión mínimo cuadrática, extendidos al ámbito de los Procesos Estocásticos y Ecuaciones en Derivadas Parciales Estocásticas. Se desarrollará asimismo la implementación de algoritmos para la generación de datos sintéticos, basados en la simulación de distribuciones de probabilidad paramétricas y no paramétricas. Alternativamente, en la generación de datos sintéticos, se aplicarán modelos de regresión no lineal, en cuya implementación se seguirán los siguientes pasos:

- (1) Definición del modelo. Establecer las ecuaciones no lineales que describen el fenómeno.
- (2) Utilizar datos reales o literatura para estimación de los parámetros del modelo.
- (3) Generación de datos sintéticos mediante simulaciones numéricas basadas en el modelo definido con perturbaciones aleatorias modelizadas mediante vectores aleatorios, o bien, procesos estocásticos, en el caso de datos sintéticos infinito-dimensionales.

**Tipología:** Estudio de casos, teóricos o prácticos, relacionados con la temática del Grado.

### Objetivos planteados:

#### OBJETIVOS:

Los principales objetivos de este trabajo son:

- O1. Revisión de los contenidos de las asignaturas de Probabilidad y Procesos Estocásticos del Grado de Matemáticas.
- O2. Revisión bibliográfica sobre métodos usuales de generación de modelos aleatorios vectoriales e infinito-dimensionales basados en distribuciones.
- O3. Revisión bibliográfica sobre algoritmos de implementación.
- O4. Implementación práctica de los algoritmos de generación de datos sintéticos en las dos vertientes analizadas: Distribuciones de probabilidad y modelos de regresión no lineal.

O5. Análisis sobre la eficiencia de los métodos programados en la reproducción de comportamientos y patrones de sistemas físicos y biológicos conocidos. Aplicación al problema de predicción.

**Bibliografía básica:**

BIBLIOGRAFÍA BÁSICA:

Asmussen, S and Glynn, P. W. (2007). Stochastic Simulation: Algorithms and Analysis. Springer Science+Business Media, LLC.

Balakrishnan, N., Melas, V. B. y Ermakov, S. (2000). Advances in Stochastic Simulation Methods. Springer Science+Business Media, LLC.

Binder, K., Kinder, K. y Heermann, D.W. (2002). Monte Carlo Simulation in Statistical Physics: An Introduction. Springer.

Koralov, L. y Sinai, Y.G. (2010). Theory of Probability and Random Processes. Springer.

Kay, S.M. (2019). Intuitive Probability and Random Processes using Matlab. Springer.

Landriscina, F. (2013). Simulation and Learning. Springer.

Nikolenko, S.I. (2021). Synthetic Data for Deep Learning. Springer.

Murray, J.D. (2002). Mathematical Biology. An Introduction. Springer.

Ripley, B.D. (2006). Stochastic Simulation. John Wiley.

Seber, G.A.F. and Wild, C.J. (2003). Nonlinear Regression. Wiley-Interscience.

**Recomendaciones y orientaciones para el estudiante:**

Se requiere poseer conocimientos de Probabilidad, Procesos Estocásticos y Lenguajes de Programación.

**Plazas:** 1

**2. DATOS DEL TUTOR/A:**

**Nombre y apellidos:** MARÍA DOLORES RUIZ MEDINA

**Ámbito de conocimiento/Departamento:** ESTADÍSTICA E INVESTIGACIÓN OPERATIVA

**Correo electrónico:** mruiz@ugr.es

**3. COTUTOR/A DE LA UGR (en su caso):**

**Nombre y apellidos:**

**Ámbito de conocimiento/Departamento:**

**Correo electrónico:**

**4. COTUTOR/A EXTERNO/A (en su caso):**

**Nombre y apellidos:**

**Correo electrónico:**

**Nombre de la empresa o institución:**

**Dirección postal:**

**Puesto del tutor en la empresa o institución:**

**Centro de convenio Externo:**

**5. DATOS DEL ESTUDIANTE:**

**Nombre y apellidos:**

**Correo electrónico:**