

Tarea 4 - Computación Científica y Ciencia de los Datos

Fluctuaciones en el precio de las acciones de Microsoft

Prof: Pablo Román

26 de Junio 2025

1 Objetivos

Desarrollar un modelo de computación probabilística para describir grandes fluctuaciones en el mercado accionario. Se revisa el precio diario de las acciones de Microsoft entre los años 2015 y 2021. Se espera lograr extraer un entendimiento estadístico de un problema sobre el cual se dispone de escasa información sobre su dinámica real. Específicamente se espera:

- Aplicar herramientas de programación probabilística para modelar cambios estructurales (change-points) en series temporales financieras. Esto son en base a modelos estadísticos Bayesianos interpretables, usando PyMC para realizar inferencias sobre eventos disruptivos en mercados.
- Diagnosticar modelos jerárquicos.
- Generar pronósticos probabilísticos a partir del modelo inferido y visualizarlos con intervalos de confianza.
- Interpretación de los resultados obtenidos por un modelo de programación probabilística.

2 Contexto: Base de datos de precios accionarios

Durante el primer trimestre del año 2020, los mercados financieros globales fueron severamente afectados por la irrupción del COVID-19. Las acciones de Microsoft, al igual que otras empresas del índice NASDAQ, sufrieron una fuerte caída seguida de una recuperación relativamente rápida, lo que constituye un evento de cambio estructural en su serie temporal de precios. Este tipo de eventos puede modelarse como una ruptura en el comportamiento estadístico de la serie. Modelos clásicos como el del desastre de minas de carbón (https://www.pymc.io/projects/docs/en/stable/learn/core_notebooks/pymc_overview.html#case-study-2-coal-mining-disasters) y modelos de regresión lineal (https://www.pymc.io/projects/docs/en/stable/learn/core_notebooks/GLM_linear.html#linear-regression), ofrecen una base para extender estos enfoques a datos financieros.

Se utilizará como punto de partida el análisis exploratorio disponible en la base de datos Microsoft Stock EDA (Clement Wouters (Kaggle, 2022): <https://www.kaggle.com/code/clemwo/microsoft-stock-eda>). Allí se presenta un análisis cualitativo de los datos. Dicha base de datos se puede obtener de <https://www.kaggle.com/datasets/vijayvenkitesh/microsoft-stock-time-series-analysis>. Esta tiene las siguientes columnas:

- Date: Fecha de la muestra, en formato mes/día/año hh:mm:ss
- Open: Precio de la acción al inicio de las transacciones del día.
- High: Precio máximo de la acción en todo el día.
- Low: Precio mínimo de la acción en todo el día.
- Close: Precio de la acción al final de las transacciones del día.
- Volume: Número total de transacciones realizadas en el día.

En la figura (1) se observa la evolución del precio de las acciones según la base de datos Microsoft Stock. Al inicio del año 2020 se observa una brusca caída de este precio con una posterior recuperación de la tendencia. Esto tendría como explicación el inicio de la pandemia de COVID. También pueden observarse otros períodos, en los cuales se observan fenómenos análogos. Por ejemplo, a finales de 2018 e inicio de 2019.

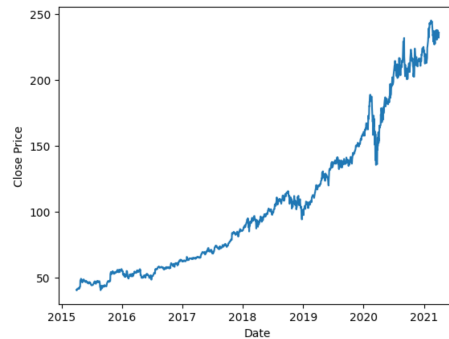


Figure 1: Precio del cierre de día de las acciones de Microsoft

3 Metodología: Programación probabilística

La programación probabilística (https://www.pymc.io/blog/chris_F_pydata2022.html) permite construir modelos estadísticos de alta expresividad. Esto significa que se puede entender de manera mucho más sencilla un fenómeno complejo debido a la simpleza del modelo que lo representa. El proceso de modelamiento (https://juanitorduz.github.io/html/pyconco22_orduz.html#/title-slide) consiste en integrar diversas fuentes de incertidumbre, partiendo por los datos. En este caso, al observar el gráfico de la evolución de los precios, puede modelarse un segmento de estos por una recta. Una recta cuya pendiente e intercepto son variables aleatorias, explícita en los parámetros mediante distribuciones a priori y realizando inferencias mediante algoritmos como NUTS.

4 Trabajo a realizar

Debe presentar un jupyter notebook que presente el modelamiento de quiebres en el comportamiento de precios de las acciones y que **describa claramente su interpretación de los resultados** según lo visto en clases.

1. (1 pto) Documente en forma clara y prolija en jupyter notebook la resolución de su tarea. Describa en detalle su enfoque utilizado e interprete claramente sus resultados y supuestos. Limite su documentación, aproximadamente a 8 hojas carta sin gráficos. Ejemplifique mediante gráficos y animaciones (matplotlib.animate). Un gráfico bien desarrollado vale más que mil palabras (o nulo si no es autoexplicativo).
2. (2.0 pt) Modele de la mejor manera posible el precio de una acción en un período en el cual el comportamiento observado es mas o menos creciente. Utilizando dicho modelo prediga los valores que se observarían en los intervalos donde se observan grandes quiebres de la tendencia. Evalúe su modelo según la metodología que aparece en el sitio de Pymc. Interprete sus resultados.
3. (3 pts) Modele de la mejor manera posible los quiebres de tendencia que aparecen en distintos tiempos. Dicho modelo debe funcionar al menos para los datos de quiebre del 2020 y del 2018-2019. Interprete sus resultados.
4. (1 pts Bonus) Modelar de la mejor manera posible la serie de tiempo completa. Evalúe e interprete sus resultados.

5 Entrega

Debe subir su Jupyter Notebook de colab en classroom con plazo al día Viernes 11 de Julio antes de las 23:55 hrs. Se descuenta 1 punto por día de atraso.