Entrega: curso de datos extremales

Laura Montaldo, CI: 3.512.962-7

2024-01-11

MEDIA

Modelización Estadística de
Datos e Inteligencia Artificial

Índice

Resumen

Your abstract goes here.

Motivación y objetivo del estudio

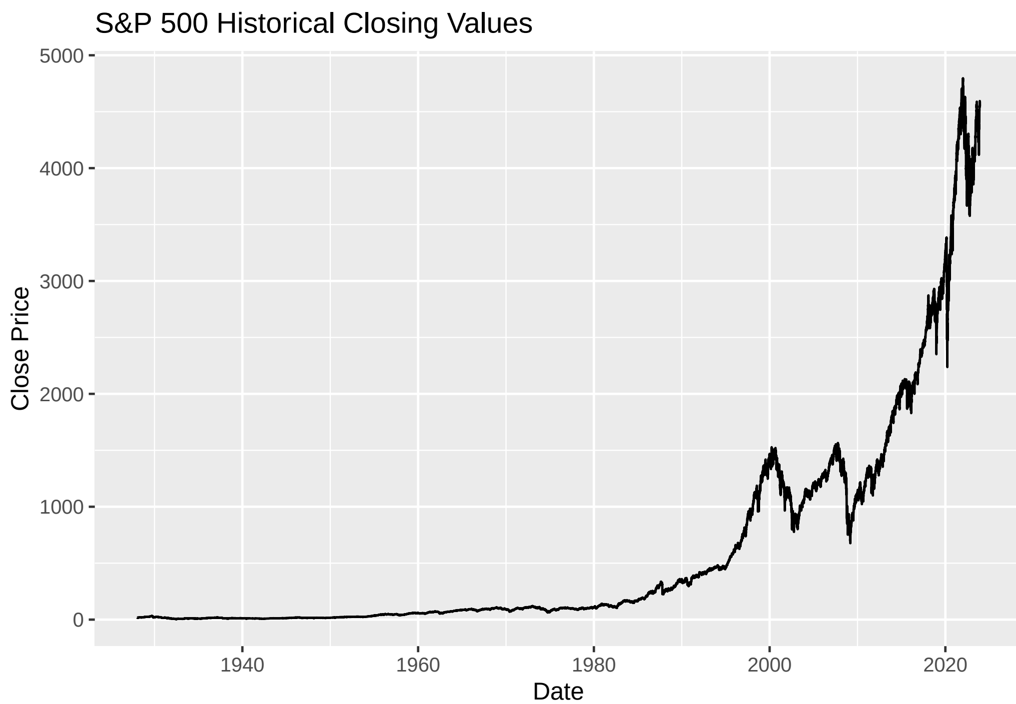
Los índices de $S\&P$ constituyen una familia de índices de renta variable[1] diseñados para evaluar el rendimiento del mercado de acciones de empresas estadounidenses que cotizan en bolsas de Estados Unidos. Esta familia de índices abarca una amplia variedad de categorías basadas en tamaño, sector y estilo. La ponderación de los índices se realiza conforme a la capitalización bursátil ajustada al *free float* (FMC). Además, se dispone de índices ponderados de manera equitativa y con límite de capitalización de mercado, como es el caso del $S\&P:500$. En este contexto, el $S\&P500$ forma parte de los índices ponderados por capitalización bursátil ajustada a la flotación (consultar $S\&P$ Dow Jones Indices). Este índice mide el rendimiento del segmento de gran capitalización del mercado estadounidense y es considerado un indicador representativo del mercado de renta variable en Estados Unidos, compuesto por 500 empresas constituyentes.

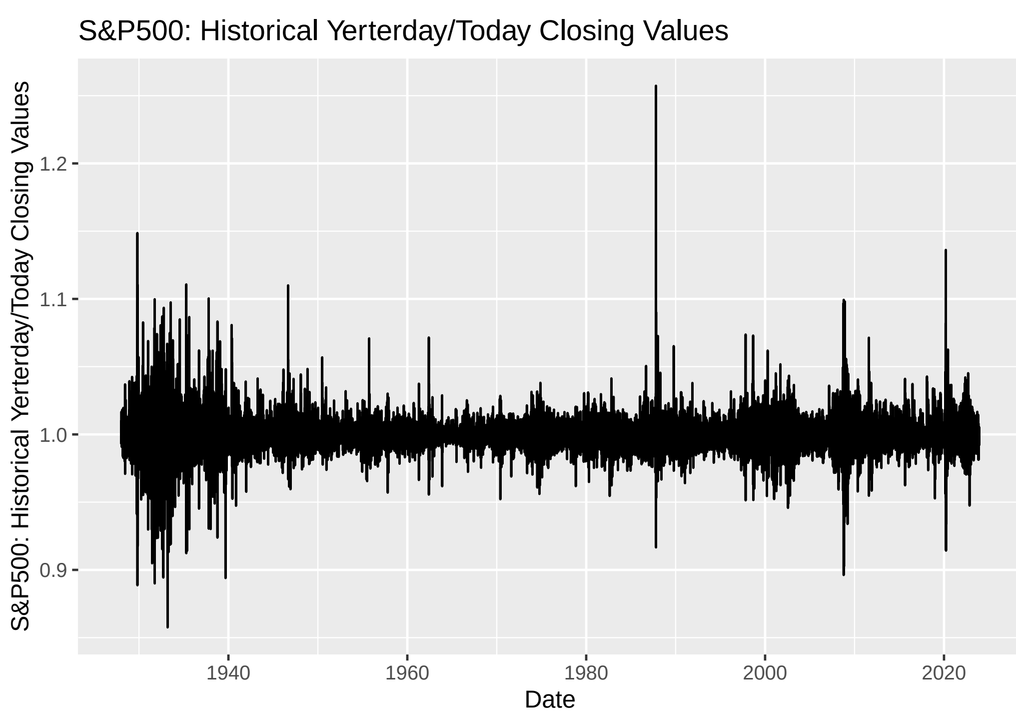
Se busca crear un indicador de una posible crisis bursátil. Como variable de referencia de toma la relación de precios al cierre de ayer sobre la de hoy $Indicador = Precio_{t-1}/Precio_t$.

Interpretación del Indicador:

- Si el indicador es menor a 1, el precio de cierre de hoy es mayor que el de ayer, lo cual podría ser considerado un signo positivo.
- Si el indicador es mayor a 1, el precio de cierre de hoy es menor que el de ayer, lo cual podría considerarse una señal de alerta.

[1] En inglés se les denomina equity indices


S&P 500 Historical Closing Values
Close Price
0
1000
2000
3000
4000
5000
1940
1960
1980
2000
2020
Date


S&P500: Historical Yerterday/Today Closing Values
S&P500: Historical Yerterday/Today Closing Values
0.9
1.0
1.1
1.2
1940
1960
1980
2000
2020
Date

Marco Teórico

Teoría asintótica clásica y las distribuciones extremales y sus dominios de atracción

Siguiendo las notas del curso, se dice que tenemos datos extremos cuando cada dato corresponde al máximo o mínimo de varios registros. Son un caso particular de evento raro o gran desviación respecto a la media.

Asumiremos que nuestros datos son iid (independientes e idénticamente distribuidos, son dos suposiciones juntas). Esta doble suposición suele no ser realista en aplicaciones concretas (ninguna de sus dos componentes, incluso) pero para comenzar a entender la teoría clásica, la utilizaremos por un tiempo.

Si tenemos datos $X_1, ..., X_n$ *iid* con distribución F, entonces $X_n^* = max(X_1, ..., X_n)$ tiene distribución F_n^* dada por $F_n^*(t) = F(t)_n$. Si conocemos la distribución F conoceríamos la distribución F_n^* , pero en algunos casos la lectura que queda registrada es la del dato máximo y no la de cada observación que dio lugar al mismo, por lo que a veces ni siquiera es viable estimar F. Pero aún en los casos en que F es conocida o estimable, si n es grande, la fórmula de F_n^* puede resultar prácticamente inmanejable. En una línea de trabajo similar a la que aporta el Teorema Central del Límite en la estadística de valores medios, un teorema nos va a permitir aproximar F_n^* por distribuciones más sencillas. Este es el Teorema de Fischer-Tippet-Gnedenko (FTG, para abreviar) que presentaremos en breve.

Como $X_1, ..., X_n$ iid, definimos $Y_i = -X_i$ para todo valor de i, entonces $Y_1, ..., Y_n$ iid y además $min(X_1, ..., X_n) = -max(Y_1, ..., Y_n)$ la teoría asintótica de los mínimos de datos iid se reduce a la de los máximos, razón por la que nos concentramos aquí en estudiar el comportamiento asintótico de los máximos exclusivamente.

Las distribuciones extremales

Las distribuciones extremales son tres: la distribución de Gumbel; la distribución de Weibull; la distribución de Fréchet.

Distribución de Gumbel

Se dice que una variable tiene distribución de Gumbel si su distribución es:

$$\Lambda(x) = exp\{-e^{-x}\} \quad \text{para todo } x \text{ real}$$

Distribución de Weibull

Se dice que una variable tiene distribución de Weibull de orden $\alpha > 0$ si su distribución es:

$$\Psi_\alpha(x) = \begin{cases} exp-(-x)^\alpha & si\ x < 0 \\ 1 & \text{en otro caso} \end{cases}$$

Distribución de Fréchet

Se dice que una variable tiene distribución de Fréchet de orden $\alpha > 0$ si su distribución es:

$$\Phi_\alpha(x) = \begin{cases} exp\{-x^{-\alpha}\} & si\ x > 0 \\ 0 & \text{en otro caso} \end{cases}$$

Teorema 1: Relaciones entre las versiones standard de las distribuciones extremales

X tiene distribución $\Phi_\alpha(x)$ si y sólo si $(-1/X)$ tiene distribución $\Psi_\alpha(x)$ si y sólo si $log(X^\alpha)$ tiene distribución Λ.

Teorema 2: Algunos datos de las distribuciones extremales

Parte 1. Si X tiene distribución $\Lambda^{(\mu,\beta)}$ entonces tiene:

a) Valor esperado $E(X) = \mu + \beta\gamma$, donde γ es la constante de Euler-Mascheroni, cuyo valor aproximado es 0.5772156649.

b) Moda: μ

c) Mediana: $\mu - \beta \log(\log 2) \approx \mu - 0.36651\beta$

d) Desviación estándar: $\beta\pi\sqrt{6} \approx 1.2825\beta$

e) Si $X^+ = \max(X, 0)$, entonces $E(X + k)$ es finito para todo valor de k natural.

f) Para simular computacionalmente X, se puede tomar U uniforme en $(0, 1)$ y hacer $X = \mu - \beta \log(-\log U)$.