

Entrega: curso de datos extremos

Laura Montaldo, CI: 3.512.962-7

2024-03-06



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY



Índice

Resumen	3
Motivación y objetivo del estudio	4
Marco teorico	10
El enfoque de conteo de eventos y los modelos de base Poissoniana. . .	10
POT (Peaks Over Treshold) y variantes	15
Estrategia Empírica	16

Resumen

Your abstract goes here.

Motivación y objetivo del estudio

Siguiendo a Perera, Segura, y Crisci (2021), 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. Entonces, en una gran variedad de dominios disciplinares suele ser de gran interés el trabajo con datos extremos, los que admiten diversos enfoques. La teoría más clásica de estadística de datos extremos se basa en los trabajos de Fréchet, Gumbel, Weibull, Fisher, Tippet, Gnedenko, entre otros. En este estudio, el foco va a estar puesto en esquemas que extienden a las distribuciones extremas clásicas.

Los índices de *S&P* son una familia de índices de renta variable¹ diseñados para medir el rendimiento del mercado de acciones en Estados Unidos que cotizan en bolsas estadounidenses. Ésta familia de índices está compuesta por una amplia variedad de índices basados en tamaño, sector y estilo. Los índices están ponderados por el criterio *float-adjusted market capitalization* (FMC). Además, se disponen de índices ponderados de manera equitativa y con límite de capitalización de mercado, como es el caso del *S&P 500*. En este sentido, el *S&P500* entraría en el conjunto de índices ponderados por capitalización bursátil ajustada a la flotación (ver [S&P Dow Jones Indices](#)). El mismo mide el rendimiento del segmento de gran capitalización del mercado estadounidense. Es considerado como un indicador representativo del mercado de renta variable de los Estados Unidos, y está 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

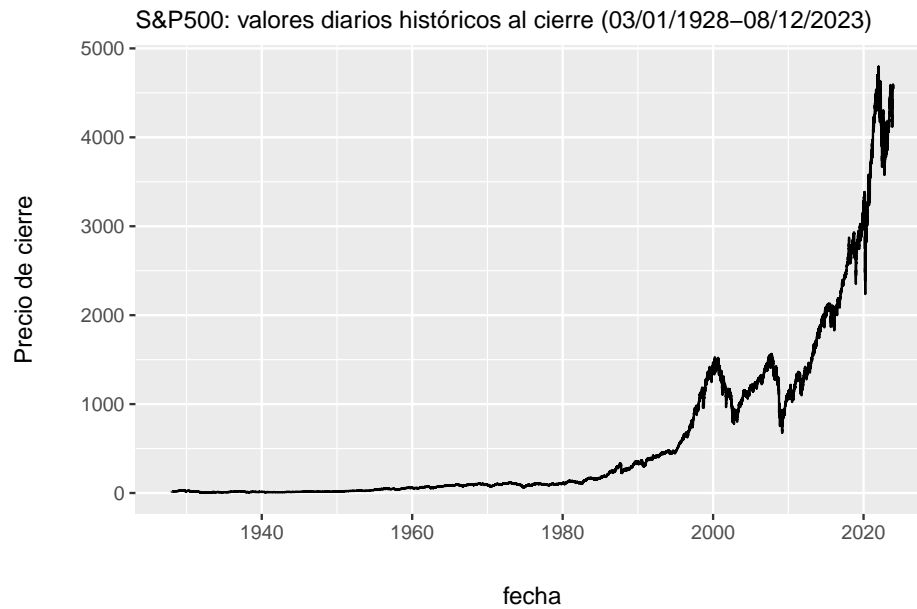
$$Indicador_t = \frac{Precio_{t-1}}{Precio_t}, \quad \text{para } t = 1, \dots, T \quad (1)$$

Interpretación del Indicador:

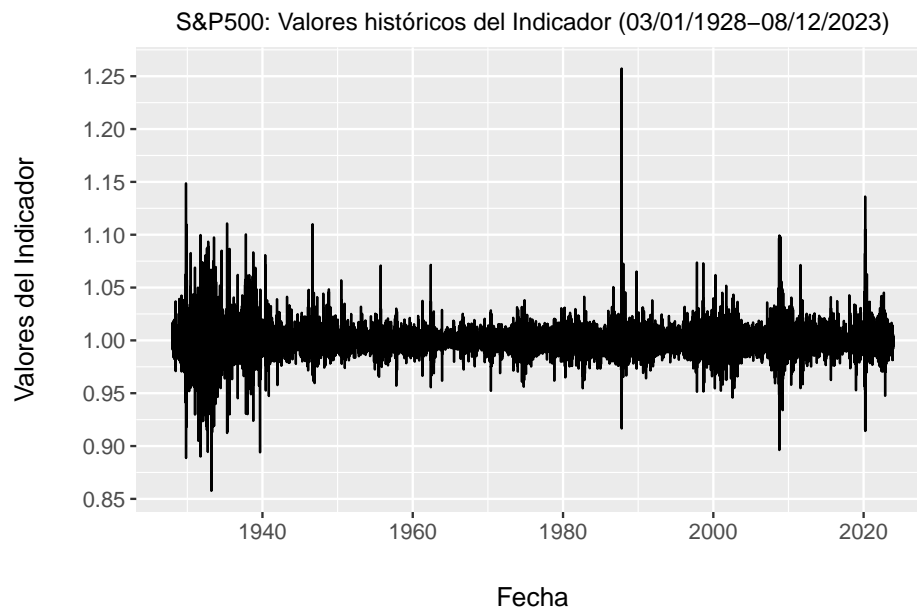
- Si el $Indicador_t \leq 1$, el precio de cierre de hoy es mayor o igual que el de ayer, lo cual podría ser considerado una señal positiva.
- Si el $Indicador_t > 1$, el precio de cierre de hoy es menor que el de ayer, lo cual podría considerarse una señal de alerta.

¹En inglés se llaman equity indices

En las siguiente figuras se muestra la evolución histórica desde la fecha 03/01/1928 hasta 08/12/2023 del precio al cierre del día del indicar S&P 500.



```
## [1] 24100      9
```



Concepto de series de tiempo estacionarias Siguiendo a Enders (2014), se dice que un proceso estocástico z_t con media y varianza finitas es estacionario en covarianza para todo t y $t - s$ cuando

$$E(z_t) = E(z_{t-s}) = \mu \quad (2)$$

$$var(z_t) = var(z_{t-s}) = \sigma_z^2 \quad (3)$$

$$cov(z_t, z_{t-s}) = cov(z_{t-j}, z_{t-j-s}) = \gamma_s \quad (4)$$

donde $\mu, \sigma_z^2, \gamma_s$ con constantes.

En~(3), cuando $s = 0$ se va a tener que γ_0 es la varianza de z_t . Que una serie temporal sea estacionaria en covarianza implica que tanto la media como todas sus autocovarianzas² no están afectadas por cambios en los orígenes temporales³.

Dado que se está en un marco de ecuaciones en diferencias, se tienen que cumplir condiciones de estabilidad en el sistema y esto implica que las raíces características asociadas al sistema de interés caigan dentro del círculo unitario. Cuando una serie temporal no sea estacionaria, es posible que presente una evidente tendencia, medias y varianzas que no son constantes en el tiempo.

Pruebas ADF Se testea para cada serie de datos

$$H_0) \gamma = 0$$

$$H_1) \gamma \neq 0$$

En este sentido, se consideran tres ecuaciones de regresión distintas para poner a prueba H_0 ,

$$\Delta z_t = \gamma z_{t-1} + \sum_{i=2}^p \beta_i \Delta z_{t-i+1} + \varepsilon_t \quad (5)$$

$$\Delta z_t = a_0 + \gamma z_{t-1} + \sum_{i=2}^p \beta_i \Delta z_{t-i+1} + \varepsilon_t \quad (6)$$

$$\Delta z_t = a_0 + \gamma z_{t-1} + a_2 t + \sum_{i=2}^p \beta_i \Delta z_{t-i+1} + \varepsilon_t \quad (7)$$

Existen tres estadísticos τ , τ_μ , τ_τ para probar la hipótesis nula $H_0) \gamma = 0$ en cada caso y además, se tienen otros tres F -estadísticos ϕ_1 , ϕ_2 , ϕ_3 para hacer pruebas conjuntas sobre los coeficientes.

²También se habla de proceso debilmente estacionario o estacionario de segundo orden.

³En modelos multivariados, el término autocovarianza refiere a la covarianza entre z_t y sus propios rezagos mientras que covarianza cruzada refiere a la covarianza entre series temporales

Los estadísticos ϕ_1, ϕ_2, ϕ_3 se construyen como pruebas F :

$$\phi_i = \frac{[SSR_{restringido} - SSR_{no\ restringido}] / r}{SSR_{no\ restringido} / (t - k)}$$

donde SSR es la suma de los cuadrados de los residuos en los modelos restringidos y no restringidos, r es la cantidad de restricciones, T es la cantidad de observaciones, k es la cantidad de parámetros estimados en el modelo irrestricto, $i = 1, 2, 3$. A su vez, $T - k$ van a ser los grados de libertad del modelos sin restricciones. Los valores de los coeficientes estimados se van a comparar con los valores críticos de tablas reportados por Dickey y Fuller(1981). La hipótesis nula indica que el proceso de generación de los datos es la del modelo restringido contra la hipótesis alternativa de que los datos son generados por el modelo sin restringir. Cuando los valores de ϕ_1, ϕ_2, ϕ_3 sean mayores a los valores críticos reportados por Dickey y Fuller(1981) se rechaza la hipótesis nula, cuando sean menores a los valores críticos entonces no se rechaza la hipótesis nula.

En el siguiente cuadro 1 se consideran los tres modelos y cada una de las hipótesis a testear con sus respectivos estadísticos.

Cuadro 1: Modelos del test ADF

	Modelo	H_0)	Estadístico de prueba
c	$\Delta z_t = a_0 + \gamma z_{t-1} + a_2 t + \sum_{i=2}^p \beta_i \Delta z_{t-i+1} + \varepsilon_t$	$\gamma = 0$ $\gamma = a_2 = 0$ $\gamma = a_2 = a_0 = 0$	τ_τ ϕ_3 ϕ_2
b	$\Delta z_t = a_0 + \gamma z_{t-1} + \sum_{i=2}^p \beta_i \Delta z_{t-i+1} + \varepsilon_t$	$\gamma = 0$ $\gamma = a_0 = 0$	τ_μ ϕ_1
a	$\Delta z_t = \gamma z_{t-1} + \sum_{i=2}^p \beta_i \Delta z_{t-i+1} + \varepsilon_t$	$\gamma = 0$	τ

Cuando no se conozca el proceso de generación de los datos, se sugiere realizar las pruebas de Dickey-Fuller Aumentado partiendo del modelo menos restrictivo para cada serie temporal a uno más particular.

Si bien las pruebas de ADF son útiles para detectar la presencia de raíces unitarias, los mismos tienen sus limitaciones. Partiendo del modelo general al particular, cada prueba está condicionada a que las pruebas anteriores sean correctas. Cuando se empieza por el primer paso, es decir, con el modelo (c) con constante y con tendencia, se hace más difícil rechazar H_0), por lo tanto, cuando se rechaza la hipótesis nula en un modelo (c) se tiende a rechazar también la hipótesis nula cuando no se incluyan los términos deterministas. A su vez, establece que el problema principal de las pruebas de Dickey-Fuller es que tanto el intercepto como la pendiente de la tendencia son, con frecuencia, estimados de manera *pobre* bajo la presencia de raíces unitarias. En general, se tiende a no rechazar la hipótesis nula de raíz unitaria incluso cuando el verdadero valor de γ no es cero. Además, la prueba presenta limitaciones también frente a cambios de régimen.

Pruebas KPSS En Shin et al. (1992), se parte de una representación cada serie temporal como la suma de un componente de tendencia determinística, un paseo aleatorio y un error estacionario. En este contexto, se pone a prueba

H_0) la serie es estacionaria alrededor de una tendencia

H_1) la serie es no estacionaria

que se corresponde con la hipótesis de que la varianza del paseo aleatorio (*random walk*) es igual a cero.

Se emplea un estadístico de Multiplicadores de Lagrange (ML) para testear la hipótesis nula de estacionariedad. De esta manera, siendo z_t con $t = 1, 2, \dots, T$ las series a las que se les quiere aplicar el test, se asume que se puede descomponer a la serie en la suma de un componente de tendencia determinística, un paseo aleatorio y un error estacionario se tiene que,

$$z_t = \xi t + r_t + \varepsilon_t \quad (8)$$

Donde r_t es un paseo aleatorio:

$$r_t = r_{t-1} + u_t, \quad (9)$$

donde u_t es $iid(0, \sigma_u^2)$. El valor inicial r_0 es fijo y sirve de intercepto. La hipótesis de estacionariedad es $\sigma_u^2 = 0$ y como se asume que ε_t es estacionario, bajo la hipótesis nula z_t es estacionaria alrededor de una tendencia.

En el caso particular de que en el modelo (8) se tenga $\xi = 0$, bajo la hipótesis nula z_t va a ser estacionaria alrededor de una constante (r_0).

Sean e_t con $t = 1, 2, \dots, T$, los residuos de la regresión z con un intercepto y tendencia. A su vez, sea $\hat{\sigma}_\varepsilon^2$ la estimación del error de la varianza de la regresión (suma de los residuos al cuadrado dividida T). Con lo anterior, se define el proceso de suma parcial de los residuos como

$$S_t = \sum_{i=1}^t r_i, \quad t = 1, \dots, T \quad (10)$$

Entonces el estadístico ML es

$$ML = \sum_{t=1}^T S_t^2 / \hat{\sigma}_\varepsilon^2 \quad (11)$$

En el caso de que se quiera poner a prueba la hipótesis nula de estacionariedad alrededor de una constante se define e_t como los residuos de la regresión z sobre un intercepto ($e_t = z_t - \bar{z}$).

Cabe resaltar que es una prueba de cola superior y se reportan los valores críticos. Además, para este caso se asume que los errores $\varepsilon_t \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2)$. Sin embargo, se puede extender la prueba con supuestos más débiles sobre la distribución de los errores dado que el supuesto anterior puede ser poco realista.

```
ts_relacion=df[,c('relacion')]
```

```
result_adf <- suppressWarnings(adf.test(ts_relacion))  
cat('p-valor adf:', result_adf$p.value ,'\n')
```

```
## p-valor adf: 0.01
```

```
result_kpss <- suppressWarnings(kpss.test(ts_relacion))  
cat('p-valor kpss:', result_kpss$p.value ,'\n')
```

```
## p-valor kpss: 0.1
```

Marco teorico

El enfoque de conteo de eventos y los modelos de base Poissoniana.

Fijaremos un cierto umbral, llamaremos *evento* cuando la variable observada supera ese umbral y dado un cierto intervalo del tiempo J , contaremos

$$N(J) = \text{número de eventos en el intervalo } J.$$

Fijando un $u = 1.025$, el evento es una señal negativa del precio de SP\$500 que estaría dado por la cantidad de días en el mes J que $Indicador_t$ supera el umbral u . Un evento ser que $Indicador_t$ supere un umbral $u > 1.025$. Si el número de eventos está dado por los días de un mes, entonces $N(mes)$ va a ser la cantidad de días que ocurrió el evento en un mes, $N(mes) = 3$ es la cantidad de señales negativas en el mes.

como la cantidad de días (d) que el indicar aumentó un cierto porcentaje de $d - 1$ a d .

2.5% en un Si d es la cantidad de días y J es el intervalo de días en un mes, $N(mes)$ $N(mes) =$ cantidad de períodos de 3 días durante un mes, en que se registró una aumento mayor o igual a 2.5% del índice de un día para el otro.

N es lo que se llama un proceso de conteo o proceso puntua⁴, un tipo de modelos utilizados en Logística, Telecomunicaciones, estudios de Contaminación Atmosférica o Costera, Clima, etc.

El proceso de conteo más simple es el llamado *Proceso de Poisson*, que puede caracterizarse de la siguiente manera.

Definición 0.1 (Proceso de Poisson). *Si N es un proceso de conteo y $\lambda > 0$, diremos que N es un Proceso de Poisson de parámetro λ (y abreviaremos N es $PP(\lambda)$) si se cumple:*

- a) *Para todo intervalo J de los reales positivos, $N(J)$ es una variable aleatoria que tiene distribución de Poisson de parámetro λ longitud(J).*
- b) *Si J, L, M, \dots es una cantidad arbitraria de intervalos de reales positivos disjuntos, entonces $N(J), N(L), N(M), \dots$ son variables aleatorias independientes.*

El siguiente teorema brinda una visualización muy interesante de los Procesos de Poisson, que nos servirá mucho para introducir otros modelos y que es ideal para poder simular computacionalmente Procesos de Poisson.

⁴Counting process, Point process en inglés

Teorema 0.1 (Otra visión de los Procesos de Poisson). Si T_1, \dots, T_n, \dots se supone *iid*, con distribución Exponencial de parámetro $\lambda > 0$ y definimos que ocurre el primer evento en el instante T_1 , el segundo en el instante $T_1 + T_2$, el tercero en el instante $T_1 + T_2 + T_3$ y así sucesivamente, el proceso N de conteo de tales eventos, es un proceso de Poisson.

Dicho de otro modo el Proceso de Poisson representa eventos aislados (“que ocurren de a uno y claramente separados”), con tiempos inter-eventos *iid* y exponenciales. Obviamente, esto muchas veces es *too good to be true*, pero variaciones de este modelo tan simple pueden brindar modelos realistas.

Observación 1: En la práctica, si se toman datos en los instantes $1, \dots, n$ suele reescalar el tiempo dividiendo por n y los instantes quedan en $[0, 1]$. Allí se define un *PP* de manera casi idéntica, obviamente modificando en la definición, tanto en *a*) como en *b*) que los intervalos deben estar contenidos en $[0, 1]$.

Observación 2: Conviene recordar que si X es una *VA* Poisson de parámetro $\lambda > 0$ y T es una *VA* exponencial de parámetro λ , entonces $E(X) = \lambda$ y $E(T) = 1/\lambda$. Si T_1, \dots, T_n, \dots siendo *iid* son los tiempos inter-eventos de un *PP*(λ) se deduce entonces de la ley de los grandes números que

$$\frac{\sum_{i=1}^{i=n} T_i}{n} \rightarrow 1/\lambda \quad \text{cuando } n \rightarrow \infty \quad (12)$$

Es decir que el tiempo promedio entre eventos “a la larga” es $1/\lambda$. Similarmente si J_1, \dots, J_n, \dots son intervalos disjuntos de longitud 1, por la definición 0.1 y la ley de los grandes números se tiene que

$$\frac{\sum_{i=1}^{i=n} N(J_i)}{n} \rightarrow \lambda \quad \text{cuando } n \rightarrow \infty \quad (13)$$

Más aún, puede probarse que

$$\frac{N((0, t))}{t} \rightarrow \lambda \quad \text{cuando } t \rightarrow \infty \quad (14)$$

Esto permite observar una consecuencia del Teorema 0.1, que es una propiedad intuitivamente muy atractiva.

La tasa promedial de incidencia de los eventos en un *PP*(λ) es inversamente proporcional al tiempo promedial inter-eventos.

Ejemplo 1: Propiedades como esta hicieron, en las primeras dos décadas del siglo XX, a un creador genial como Agner Erlang modelar mediante Procesos de Poisson las llamadas que arribaban a una central telefónica, así como (con parámetros muy distintos) el proceso de ocupación de las líneas entre dos centrales. Eso condujo no sólo al desarrollo de las primeras centrales de telefonía conmutada por circuitos por CTC, la filial danesa de Bell, sino además a que Erlang desarrollara su “fórmulas de bloqueo”, fino cálculo por el cual, según los parámetros del proceso de arribo y del proceso de ocupación de líneas, se calcula la probabilidad de “saturación” (no hay ninguna línea disponible) dado el número de líneas entre centrales, o, dada una probabilidad de saturación “tolerable” (ϵ).

DISEÑAR (determinar el mínimo número de líneas necesarias para que la probabilidad de bloqueo no exceda ϵ). Si el tiempo entre arribos de llamadas a la central es Exponencial de parámetro λ , y la duración media de una llamada es Exponencial de parámetro μ , entonces el parámetro crucial de la fórmula de Erlang es

$$\rho = \lambda / \mu$$

$$= \text{“duración media de la llamada”} / \text{“tiempo medio entre llamadas”} \quad (15)$$

y a mayor valor de ρ , mayor probabilidad de saturación para una conectividad dada. Esta fórmula (15) aún sigue en uso en algunos problemas y dió pie al desarrollo de fórmulas de bloqueo más sofisticadas para situaciones más complejas. Con mucha justicia, la unidad en la que se mide la intensidad de tráfico en redes se llama *erlang* y este ejemplo nos parece una clara muestra de cuán útil ha sido el muy sencillo Proceso de Poisson. Sin embargo, en otros problemas, por ejemplo en modernas redes de datos en las que los “eventos” de “demanda de servicio” pueden ocurrir simultáneamente en muy grandes cantidades (“clustering”), aparece un modelo más sofisticado, que puede ser definido a partir del Proceso de Poisson: el Proceso de Poisson Compuesto.

Definición 0.2 (Proceso de Poisson Compuesto). *Si N es un Proceso de Poisson de parámetro $\lambda > 0$, G es una distribución de probabilidad en los naturales $1, 2, 3, \dots$, consideramos un proceso S_1, \dots, S_n iid con distribución G y construimos un nuevo proceso de conteo M de la forma siguiente:*

- *Cuando N tiene su primer evento, M tiene S_1 eventos simultáneos;*
- *Cuando N tiene su segundo evento, M tiene S_2 eventos simultáneos.....
(y así sucesivamente)*

decimos que M es un Proceso de Poisson Compuesto de parámetro $\lambda > 0$ y distribución de eventos G (y abreviaremos M es $PPC(\lambda; G)$)

Ejercicio 1 : Demostrar que para un $PPC(\lambda; G)$ el tiempo medio inter-eventos sigue siendo $1/\lambda$, pero que la tasa de incidencia media de eventos ahora es $\lambda E(G)$.

Observación 3. Para aclarar, si G es una distribución degenerada otorga al 1 probabilidad 1, el correspondiente $PPC(\lambda; G)$ en realidad es un $PPC(\lambda)$. Ergo, el PP es un caso particular de PPC .

Observación 4. Para evitar confusiones frecuentes, distinguiremos explícitamente estos procesos de los llamados Procesos de Poisson no-homogéneos. Para ello recordemos, sin entrar en tecnicismos, que una medida en los reales positivos es una función que a los conjuntos asocia números positivos con las mismas propiedades formales, excepto que no tiene por qué dar a todo el conjunto de los reales positivos (a todo el universo) el valor 1. Dicho de otro modo una probabilidad es una medida particular, que a todo el universo asigna el valor 1. Puede pensarse como ejemplo típico de una medida, la que asigna a un conjunto la integral sobre ese conjunto de una función no negativa (no necesariamente de integral total 1, puede ser incluso infinita). La longitud es el ejemplo más simple de medida (llamada también medida de Lebesgue) y la longitud de todos los reales positivos es infinito. Puede demostrarse que la longitud multiplicada por una constante no negativa son las únicas medidas invariantes por traslaciones, punto importante para la distinción que queremos hacer.

Definición 0.3 (Proceso de Poisson No Homogéneo). *Si N es un proceso de conteo y m es una medida que NO puede expresarse como una constante por la longitud, diremos que N es un Proceso de Poisson No Homogéneo de medida m (y abreviaremos N es $PPNH(m)$) si se cumple:*

- a) *Para todo intervalo J de los reales positivos, $N(J)$ es una variable aleatoria que tiene distribución de Poisson de parámetro $m(J)$.*
- b) *Si J, L, M, \dots es una cantidad arbitraria de intervalos positivos DISJUNTOS, entonces $N(J), N(L), N(M), \dots$ son variables aleatorias independientes.*

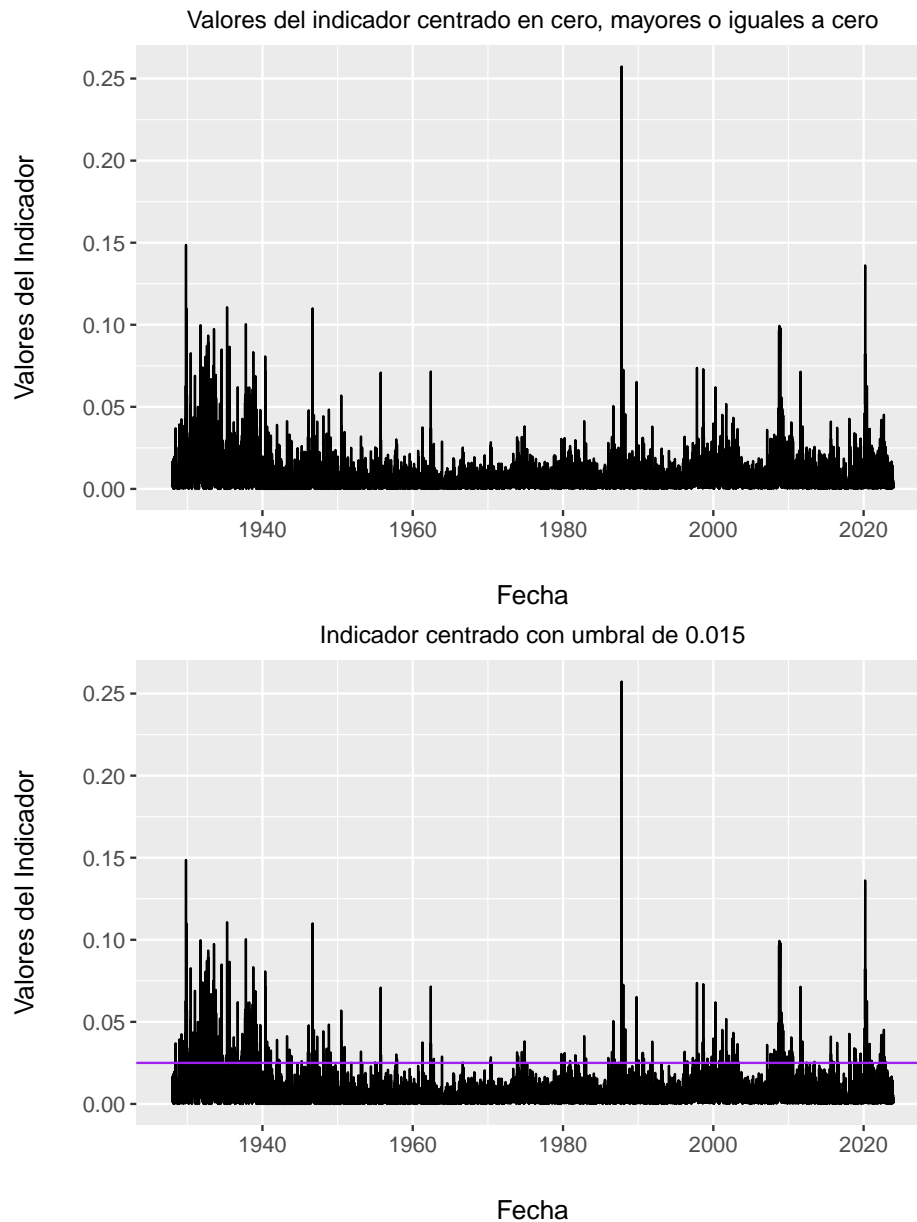
Queda claro que el proceso de Poisson podría verse de la manera a) y b) anterior cuando $m = \text{constante por longitud}$, por eso, para no confundir, se excluye a título expreso que m pueda ser constante por longitud. Para dejar en claro la diferencia entre los PPNH y los PPC (o el simple PP), recordemos que en los PPC, los tiempos inter-eventos son exponenciales de parámetro $\lambda > 0$ e *iid*. El siguiente resultado muestra la diferencia de conceptos. Por su extrema simplicidad, lo detallaremos.

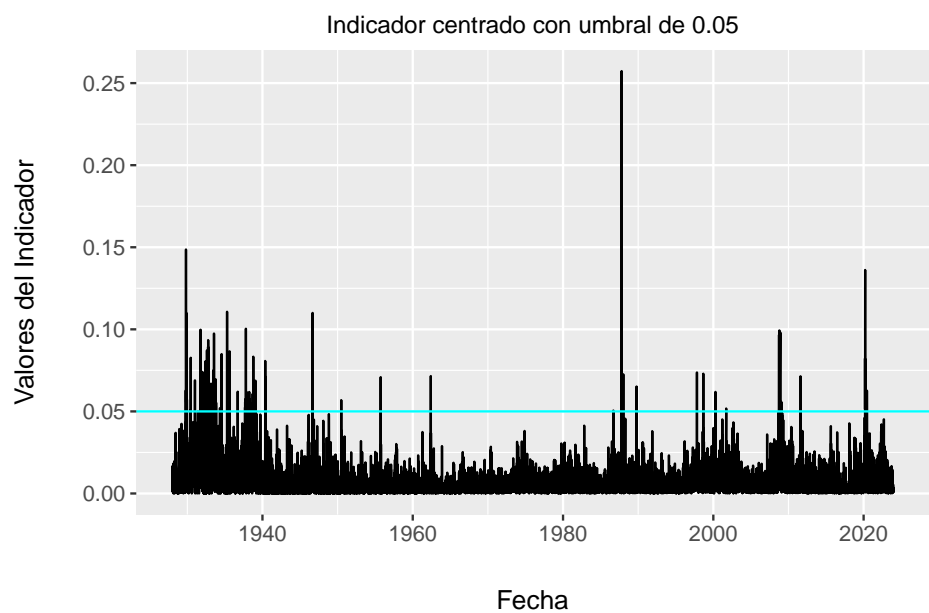
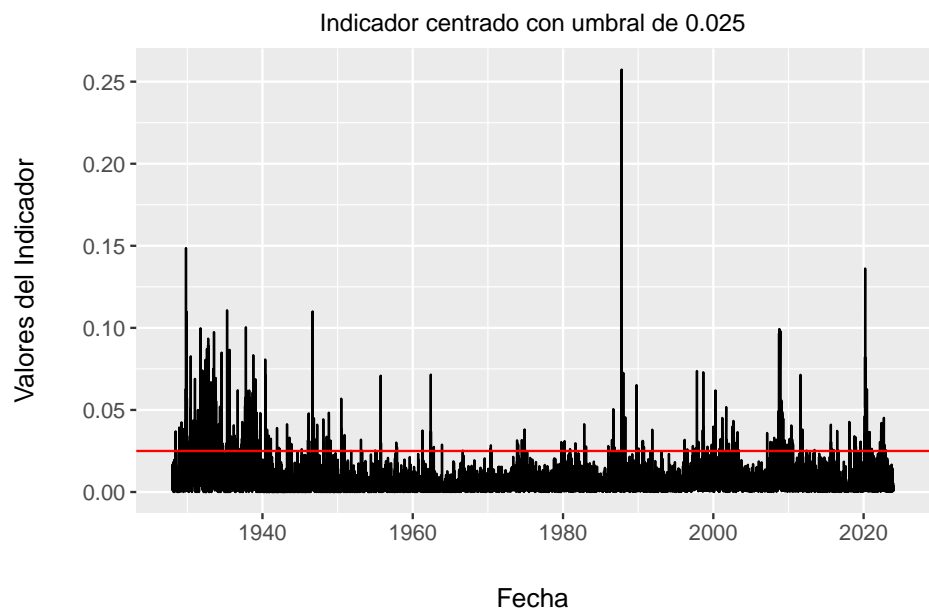
Teorema 0.2 (PPNH no es PPC). *Si N es un PPNH y T_1 es el tiempo del primer evento, la distribución de T_1 no es exponencial. Por lo tanto, un PPNH no es PPC.*

POT (Peaks Over Threshold) y variantes

Estrategia Empírica

A la columna relativa a la relacion de precios se la resta por 1 para tener centrados los valores de la relacion de precios en cero. Y posteriormente analizar si las series, fijando distintos umbrales son estacionarias.





```
filtered_df_0_025 <- df %>%
  filter(rel_cero >= 0.025)
```

```
head(filtered_df_0_025)
```

```
##      Date  Open  High  Low Close Volume Dividends Stock.Splits relacion
```

```
## 1 1928-06-11 18.68 18.68 18.68 18.68      0      0      0 1.036938
## 2 1928-07-11 18.95 18.95 18.95 18.95      0      0      0 1.025330
## 3 1928-12-06 22.91 22.91 22.91 22.91      0      0      0 1.039284
## 4 1929-02-07 24.71 24.71 24.71 24.71      0      0      0 1.031566
## 5 1929-03-25 24.51 24.51 24.51 24.51      0      0      0 1.042432
## 6 1929-04-01 24.88 24.88 24.88 24.88      0      0      0 1.026125
##      rel_cero
## 1 0.03693793
## 2 0.02532979
## 3 0.03928414
## 4 0.03156620
## 5 0.04243162
## 6 0.02612546
```

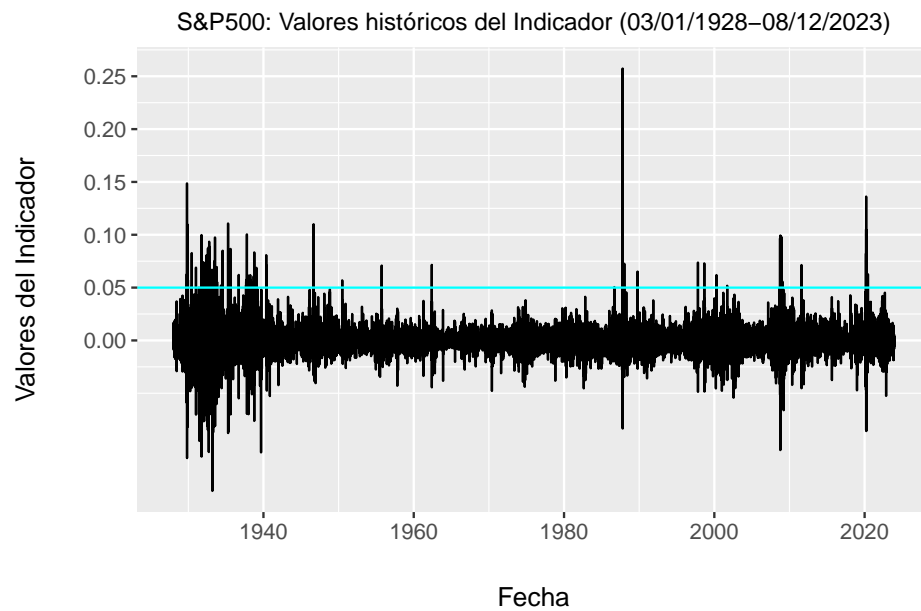
```
data=filtered_df_0_025[,c('Date', 'rel_cero')]
n=dim(data)[1]
```

```
fecha_maxima <- max(data$Date)
# Reescalar el tiempo dividiendo cada fecha por la fecha máxima
data$tiempo_reescalado <- as.numeric(data$Date - min(data$Date)) / as.numeric(fecha_maxima -
head(data)
```

```
##      Date      rel_cero tiempo_reescalado
## 1 1928-06-11 0.03693793      0.0000000000
## 2 1928-07-11 0.02532979      0.0008690614
## 3 1928-12-06 0.03928414      0.0051564311
## 4 1929-02-07 0.03156620      0.0069814600
## 5 1929-03-25 0.04243162      0.0083140209
## 6 1929-04-01 0.02612546      0.0085168019
```

```
ggplot(df, aes(x = Date, y = rel_cero)) +
  geom_line() +
  geom_hline(yintercept = 0.05, linetype = "solid", color = "cyan") + # Add horizontal line
  ggtitle("S&P500: Valores históricos del Indicador (03/01/1928-08/12/2023)") +
  xlab("Fecha") +
  ylab("Valores del Indicador") +
  scale_x_date(limits = date_range) +
  scale_y_continuous(breaks = seq(0, ceiling(max(df$relacion)), by = 0.05)) +
  theme(
    axis.title.x = element_text(margin = margin(t = 20, b = 40)),
    axis.title.y = element_text(margin = margin(r = 20, l = 40)),
    plot.title = element_text(size = 11, hjust = 0.5), # Center the plot title
    axis.text = element_text(size = 10), # Adjust the size of axis text
    axis.title = element_text(size = 12), # Adjust the size of axis titles
    legend.title = element_text(size = 10), # Adjust the size of legend title
```

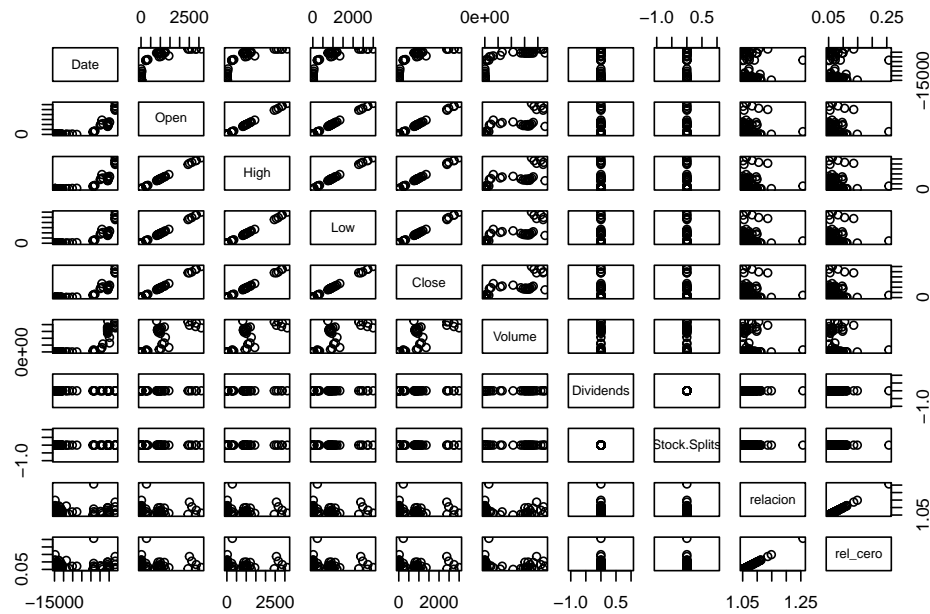
```
legend.text = element_text(size = 8) # Adjust the size of legend text
)
```



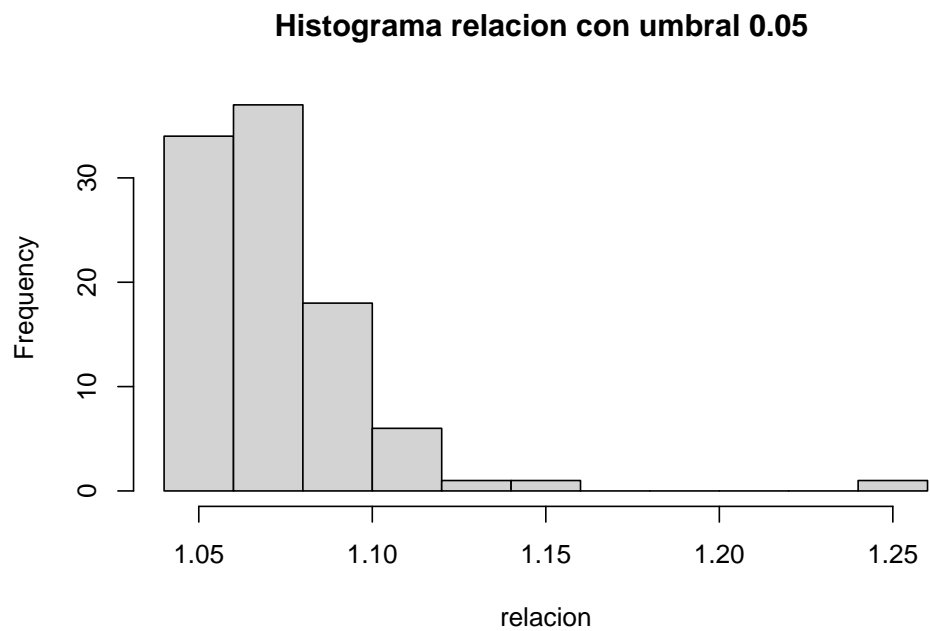
```
filtered_df_0_05 <- df %>%
  filter(rel_cero >= 0.05)
dim(filtered_df_0_05)
```

```
## [1] 98 10
```

```
plot(filtered_df_0_05)
```

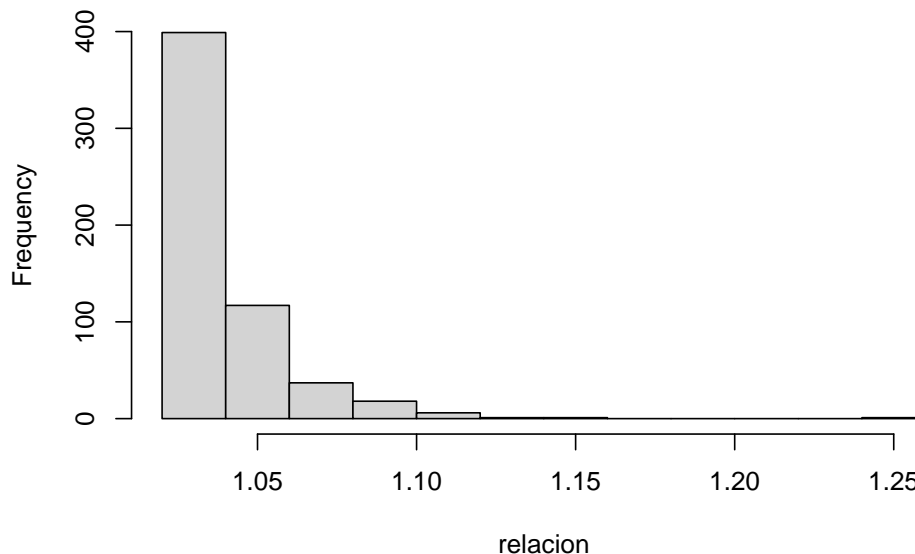


```
hist(filtered_df_0_05$relacion, main = "Histograma relacion con umbral 0.05 ", xlab = "relacion")
```



```
hist(filtered_df_0_025$relacion, main = "Histograma relacion con umbral 0.025 ", xlab = "relacion")
```

Histograma relacion con umbral 0.025



- Enders, W. 2014. *Applied Econometric Time Series*. Wiley Series en Probability y Statistics. Wiley.
- Perera, Gonzalo, Angel Segura, y Carolina Crisci. 2021. *Curso de estadística de datos extremos, cap. 1 a cap. 5*.
- Shin, Yongcheol, Denis Kwiatkowski, Peter Schmidt, y Peter C. B. Phillips. 1992. «Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root: How Sure Are We That Economic Time Series Are Nonstationary?» *Journal of Econometrics* 54 (1-3): 159-78.
- Stephenson, A. G. 2002. «evd: Extreme Value Distributions». *R News* 2 (2): 0. <https://CRAN.R-project.org/doc/Rnews/>.