

Tiempos de extinción del modelo logístico estocástico

Santiago Rodríguez Newton¹, Dr. Ehyter Matías Martín González²

Universidad de Guanajuato, DEMAT
santiago.rodriguez@cimat.mx¹
ehyter.martin@ugto.mx²

Resumen

El conocido modelo logístico modela poblaciones tomando en cuenta la población inicial, la tasa de crecimiento y la población máxima; una característica de este modelo son las soluciones constantes correspondientes a la extinción y a la saturación de la población. La versión estocástica de este modelo considera la población como un proceso de Itô donde se añade ruido gaussiano a la tasa de crecimiento. Haciendo uso de simulaciones y del método de exceso de umbrales se analiza la cola de distribución de los tiempos de extinción, la cual no es calculable de manera analítica y es poco lo que se conoce sobre su comportamiento, al menos en el sentido asintótico.

Palabras clave: Ecuaciones Diferenciales Estocásticas, Modelo Logístico, Teoría de Extremos, Simulaciones.

Introducción

Los modelos logísticos estocásticos no son algo reciente, incluso (Levins, 1969) empezó a considerar efectos aleatorios en distintos modelos de población. Desde entonces mucha investigación y aplicaciones se han hecho en el tema, por ejemplo, (Shah, 2013) creó nuevos modelos logísticos para estudiar poblaciones de peces, (Calatayud, 2020) usó un modelo logístico con capacidad de carga variante para estudiar la biomasa de bacterias en la piel y (Roy, 2021) usó un modelo logístico con tasa de crecimiento variante para estudiar la proliferación de células.

En este tipo de modelos, los cuales usualmente son aplicados para modelar tamaños de poblaciones a lo largo del tiempo, es de interés conocer la distribución del tiempo al cual la población llega a cero, conocido en este contexto como tiempo de extinción.

Existen pocos resultados sobre la distribución de probabilidad exacta de tales tiempos, algunos de los cuales pueden encontrarse en la literatura clásica de procesos estocásticos o en artículos especializados [ver, por ejemplo, (Norden, 1982) y las referencias dadas]. En dicho trabajo, el autor obtuvo fórmulas para calcular los momentos de los tiempos de extinción de un modelo logístico. Posteriormente, en (Golec, 2003) se propone una variante del modelo logístico sobre la que, hasta dónde entendemos, no se ha dicho mucho acerca de la distribución del tiempo de extinción asociado.

Dado lo anterior, en este artículo estudiamos la distribución del tiempo de extinción del modelo dado en (Golec, 2003) utilizando simulaciones numéricas y la Teoría de Valores Extremos, la cual ha demostrado a lo largo de diversos estudios, ser una herramienta útil para hacer aproximaciones de diversas cantidades utilizando cantidades consideradas como extremas (cabe mencionar que el tiempo de extinción en sí, ya es considerado un valor extremo en la literatura de dinámica de poblaciones).

El trabajo está organizado de la siguiente manera: en la sección 2 introducimos el modelo logístico de interés y sus características principales, en la sección 3 se aclara lo que es un tiempo de extinción y como simularlo. Por su parte, la sección 4 presenta una breve introducción a la Teoría de Valores Extremos necesaria para entender el método de exceso de umbrales. Posteriormente, en la sección 5 se presenta el análisis realizado utilizando dicha metodología y los resultados obtenidos. Finalmente, en la sección 6 presentamos las conclusiones del estudio realizado y discutimos una avenida para más investigación en el futuro.

Modelo logístico

Versión determinista

Recordemos que el modelo logístico de población está dado por la ecuación diferencial ordinaria

$$P'(t) = r(t)P(t)[M - P(t)], \quad P_0 = P(0)$$

donde $P(t)$ es la población en el tiempo t , $r(t)$ es la tasa de crecimiento y M es la población máxima o "capacidad de carga". La ecuación anterior tiene dos soluciones constantes, $P(t) = 0$ correspondiente a la extinción de la población y $P(t) = M$ correspondiente a la saturación de la población. Haciendo uso de la transformación $y(t) = \frac{P(t)}{M} - 1$ obtenemos la ecuación diferencial

$$y'(t) = -Mr(t)y(t)[1 + y(t)], \quad y_0 = \frac{P_0}{M} - 1.$$

Ahora las soluciones constantes son $y(t) = -1$, correspondiente a la extinción y $y(t) = 0$ correspondiente a la saturación. En otras palabras, hemos "normalizado" las soluciones constantes. Suponiendo que r es continua, la ecuación anterior tiene solución

$$y(t) = \frac{y_0 e^{-MR(t)}}{1 + y_0 [1 - e^{-MR(t)}]}, \quad R(t) = \int_0^t r(s) ds.$$

En el caso particular donde se tiene una tasa de crecimiento constante, la solución se simplifica a

$$y(t) = \frac{y_0 e^{-Mrt}}{1 + y_0 [1 - e^{-Mrt}]}$$

Con la solución en mano, es aparente que si $r > 0$ entonces $y(t) = 0$ y, similarmente, si $r < 0$ entonces $y(t) = -1$. En términos del modelo, lo anterior dice que si la población crece, esta crecerá hasta saturarse; si la población decrece, esta disminuirá hasta extinguirse.

Versión estocástica

Para obtener la versión estocástica del modelo logístico, agregaremos un término aleatorio a la tasa de crecimiento. Siendo precisos, supongamos que $r(t) = \underline{r}(t) + \eta_t$ donde \underline{r} es una función real y η_t es ruido gaussiano con intensidad $\sigma(t)$, i.e., la colección de variables aleatorias $\{\eta_t, t \geq 0\}$ es independiente y $\eta_t \sim N(0, \sigma^2(t))$. Como $\sigma(t)$ es un factor de escala en la distribución normal, obtenemos el diferencial estocástico $\eta_t dt = \sigma(t) dB_t$ donde B_t es un movimiento Browniano. Si sustituimos r en el modelo logístico, expandimos y usamos el diferencial estocástico mencionado, obtenemos el problema de valor inicial de Itô¹

$$dy_t = -M\underline{r}(t)y_t[1 + y_t] dt - M\sigma(t)y_t[1 + y_t] dB_t.$$

De igual manera que la versión determinista, el modelo logístico estocástico tiene soluciones constantes $y_t = -1$ y $y_t = 0$, sin embargo, la convergencia a dichas soluciones requiere condiciones más fuertes descritas en los Teoremas 1 y 2. Por simplicidad denotaremos $y(t, y_0)$ como el proceso de Itô anteriormente descrito con condición inicial y_0 .

¹ Se puede considerar también el problema de valor inicial de Stratonovič, este tiene propiedades distintas y es discutido en (Golec, 2003).

Teorema 1: Si existe $\varepsilon > 0$ tal que para todo $t \geq 0$

$$\sigma^2(t) + \varepsilon \leq \frac{2r(t)}{M},$$

entonces se cumple que para todo $y_0 \in (-1,0)$, $y(t, y_0) \rightarrow 0$ casi seguramente conforme $t \rightarrow \infty$.

En términos del modelo, el Teorema 1 pide una sencilla condición que relaciona la intensidad del ruido, la tasa de crecimiento y la población inicial para garantizar que la población aumente hasta saturarse independientemente de la población inicial.

Teorema 2: Si existe $\varepsilon > 0$ tal que para todo $t \geq 0$

$$-\sigma^2(t) - \varepsilon \geq \frac{2r(t)}{M},$$

entonces se cumple que para todo $y_0 \in (-1,0)$, $y(t, y_0) \rightarrow -1$ casi seguramente conforme $t \rightarrow \infty$.

Análogamente al Teorema 1, el Teorema 2 pide una sencilla relación para garantizar que la población disminuya hasta extinguirse sin importar la población inicial. Si se cumple ninguna de las condiciones de los Teoremas 1 o 2, (Golec, 2003) argumenta que entonces y_t converge casi seguramente a -1 o 0 , es decir, la población puede saturarse o extinguirse.

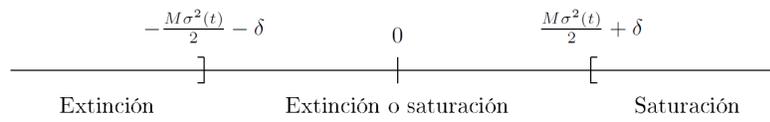


Figura 1. La recta real representa el valor de $r(t)$, los distintos sectores indican el estado hacia el cual converge la población, $\delta > 0$ es arbitrario.

Simulación numérica y tiempos de extinción

Simulación determinista y estocástica

Simular el modelo logístico determinista es sencillo dado que tenemos su solución explícita vista en la sección 2.1. Por su parte, el modelo logístico estocástico no tiene soluciones explícitas conocidas y aunque las tuviera, tendríamos que conocer la distribución de alguna integral de Itô, lo cual puede ser muy complicado. Por lo tanto, simularemos el proceso numéricamente usando el esquema de Milshstein de segundo orden. Suponiendo que r y σ son constantes, este esquema está dado por la relación recursiva

$$y_{k+1} = y_k - My_k[1 + y_k] \left[h \left(r + \frac{M\sigma^2(1 + 2y_k)}{2} \right) - \sigma(1 + y_k)N_k + \frac{M\sigma^2(1 + 2y_k)N_k^2}{2} \right]$$

donde $h > 0$ es el tamaño de salto de la simulación y $N_k \sim N(0, h)$ son variables aleatorias independientes.

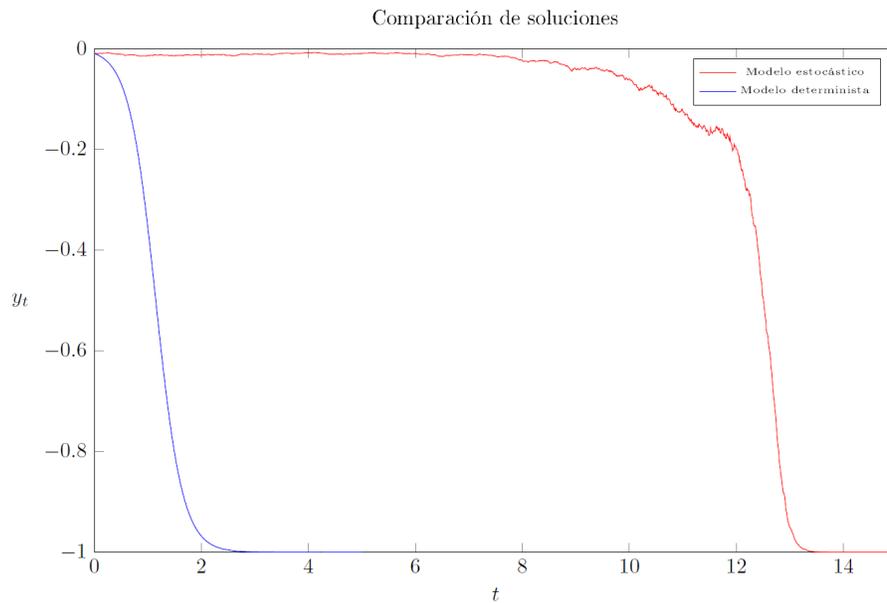


Figura 2. Simulación hecha con parámetros $r = -0.2$, $\sigma = 0.14$, $M = 20$, $y_0 = 0.01$, $h = 0.01$. Se invita al lector verificar que estos parámetros satisfacen la condición pedida en el Teorema 2.

Tiempos de extinción

En un contexto real, aún si sabemos que una especie se extinguirá debido a su negativa tasa de crecimiento, no podemos predecir realísticamente el momento en el cual ocurrirá la extinción, por lo tanto, tiene sentido considerar el modelo logístico estocástico y sus simulaciones para estudiar la distribución de los tiempos de extinción. Como en teoría nunca se alcanza realmente la extinción o saturación, definimos una cota $\varepsilon > 0$ para juzgar si la población se extinguió o no. Formalmente definimos el tiempo de extinción τ_ε como

$$\tau_\varepsilon = \inf\{t \geq 0 \mid y_t \leq \varepsilon - 1\}.$$

En otras palabras, τ_ε es la variable aleatoria cuyo valor nos indica el primer tiempo en el cual el proceso y_t es menor o igual a $\varepsilon - 1$. En lo que resta del documento, se usará $\varepsilon = 0.01$ para definir los tiempos de extinción.

Introducción a la Teoría de Extremos

Es de natural interés estudiar la ocurrencia de eventos extremos y su probabilidad, tales como fuertes terremotos, sequías largas, etc. En pos a este resultado se desarrolló la Teoría de Extremos, la cual estudia las propiedades asintóticas de los máximos de alguna sucesión de variables aleatorias idénticamente distribuidas e independientes. A continuación se verá una breve y corta introducción a esta teoría y en particular se verán los resultados necesarios para entender el método de exceso de umbrales el cual aplicaremos en la sección 5.

Distribuciones de extremos y dominios de atracción

Supongamos que tenemos una sucesión de variables aleatorias X_n idénticamente distribuidas e independientes (iid por sus siglas en inglés) con función de distribución común F . Si definimos $M_n = \max\{X_k, k \leq n\}$ el máximo de las primera n variables aleatorias, se cumple que M_n converge en distribución a $\omega_F = \sup\{x \mid F(x) < 1\}$, el extremo derecho de F . Lo anterior no nos da mucha información acerca del comportamiento asintótico de los máximos, entonces se estudia la convergencia de los máximos normalizados $\frac{M_n - b_n}{a_n}$ por medio del Teorema de Fisher – Tippet.

Teorema 3 (Fisher – Tippet): Sea $\{X_n\}$ una sucesión de variables aleatorias iid. Si existen sucesiones de constantes $a_n > 0$ y $b_n \in \mathbb{R}$ tales que $\frac{M_n - b_n}{a_n}$ converge en distribución a una variable aleatoria L , entonces L tiene una de las siguientes funciones de distribución:

Distribución Fréchet:

$$H_F(x) = \exp\left(-\left[-\frac{x - \mu}{\sigma}\right]^{-\alpha}\right) 1_{(\mu, \infty)}(x), \quad \mu \in \mathbb{R}, \alpha > 0, \sigma > 0.$$

Distribución Gumbel:

$$H_G(x) = \exp\left(-\exp\left[-\frac{x - \mu}{\sigma}\right]\right), \quad \mu \in \mathbb{R}, \sigma > 0.$$

Distribución Weibull:

$$H_W(x) = 1_{[\mu, \infty)}(x) + \exp\left(-\left[-\frac{x - \mu}{\sigma}\right]^\alpha\right) 1_{(-\infty, \mu)}(x), \quad \mu \in \mathbb{R}, \alpha > 0, \sigma > 0.$$

Las distribuciones anteriormente mencionadas se conocen como las distribuciones de extremos. Considerando lo anterior, la siguiente definición tiene sentido ya que clasifica a las distribuciones que cumplen las condiciones del Teorema de Fisher – Tippet.

Definición 1: Sea H una función de distribución de extremos, definimos su dominio de atracción maximal $D(H)$ como el conjunto de funciones de distribución

$$D(H) = \left\{F \mid \lim_{n \rightarrow \infty} F^n(a_n x + b_n) = H(x)\right\},$$

donde $a_n > 0$ y $b_n \in \mathbb{R}$. Nótese que $F^n(a_n x + b_n)$ es la función de distribución de $\frac{M_n - b_n}{a_n}$.

Características de los dominios de atracción

Determinar el dominio de atracción de una distribución (si es que tiene) es útil ya que se conocen muchas propiedades de cada dominio obteniendo así más información de la distribución. Esto es especialmente útil en un contexto práctico donde no se conoce explícitamente la distribución por tratar como es nuestro caso. A continuación se listan tres teoremas con la propiedades más importantes de cada dominio de atracción.

Teorema 4 (Dominio de atracción Fréchet): Una función de distribución F pertenece al dominio de atracción Fréchet si y solo si $\omega_F = \infty$ y $F \in RV_\alpha$ es de cola de variación regular con índice $\alpha > 0$, es decir, para todo $x > 0$,

$$\lim_{t \rightarrow \infty} \frac{\bar{F}(tx)}{\bar{F}(t)} = x^{-\alpha}.$$

Teorema 5 (Dominio de atracción Weibull): Una función de distribución F pertenece al dominio de atracción Weibull si y solo si $\omega_F < \infty$ y la función de distribución $F_*(x) = F(\omega_F - x^{-1})$ es de cola de variación regular.

Teorema 6 (Dominio de atracción Gumbel): Una función de distribución F pertenece al dominio de atracción Gumbel si y solo si existe $z < \omega_F$ tal que

$$\bar{F}(x) = c(x) \exp\left(-\int_z^x \frac{g(t)}{a(t)} dt\right), \quad x \in (z, \omega_F)$$

donde c y g son funciones medibles tales que $\lim_{x \rightarrow \omega_F} c(x) > 0$ y $\lim_{x \rightarrow \omega_F} g(x) = 1$. a es una función positiva, absolutamente continua respecto a la medida de Lebesgue y con densidad a' tal que $\lim_{x \rightarrow \omega_F} a'(x) = 0$.

Función de excesos promedio

Los Teoremas 4, 5 y 6 nos dan maneras para verificar si alguna función de distribución tiene algún dominio de atracción, sin embargo, en nuestro contexto no son útiles dado que se requiere saber cuál es la distribución de manera explícita, por lo tanto, se introduce la función de excesos promedio (FEP).

Definición 2: Dada una variable aleatoria no negativa X con función de distribución F , definimos su función de exceso promedio e_F como

$$e_F(u) = \mathbb{E}[X - u \mid X > u], \quad \bar{F}(u) > 0.$$

Esta función es muy importante porque se comporta de manera distinta asintóticamente en cada dominio de atracción como lo muestran los Teoremas 7, 8 y 9 a continuación.

Teorema 7 (FEP Fréchet): Si F es una función de distribución con dominio de atracción Fréchet, se cumple que

$$\lim_{u \rightarrow \infty} \frac{e_F(u)}{u} > 0.$$

Teorema 8 (FEP Weibull): Si F es una función de distribución con dominio de atracción Weibull, se cumple que

$$\lim_{u \rightarrow \omega_F} \frac{e_F(u)}{u} = 0.$$

Teorema 9 (FEP Gumbel): Si F es una función de distribución con dominio de atracción Gumbel y $\omega_F = \infty$, se cumple que

$$\lim_{u \rightarrow \infty} \frac{e_F(u)}{u} = 0.$$

Si $\omega_F < \infty$, se cumple que

$$\lim_{u \rightarrow \omega_F} \frac{e_F(u)}{\omega_F - u} = 0.$$

Aunque los resultados anteriores son relativamente sencillos de verificar, en la práctica no se sabe cuál es la función de excesos promedio por lo que se emplea la función de excesos promedio empírica (FEPE).

Definición 3: Dada una muestra $\{X_k, k \leq n\}$ de variables aleatorias iid con distribución F , definimos la función de excesos promedio empírica como

$$e_{n,F}(u) = \frac{1}{|\Delta_n(u)|} \sum_{k \in \Delta_n(u)} (X_k - u)$$

donde $\Delta_n(u) = \{k \mid X_k > u\}$. Si sucede que $\Delta_n(u) = \emptyset$, definimos $e_{n,F}(u) = 0$.

Se ha de tener precaución al usar esta herramienta, los Teoremas 7, 8 y 9 no dicen que si la función de excesos promedio presenta el comportamiento asintótico adecuado, la distribución asociada pertenece al dominio de atracción correspondiente. La función de excesos promedio solo sirve para descartar posibles dominios de atracción.

Exceso de umbrales

Aunque conocer el dominio de atracción de una distribución desconocida es útil, no nos da nada cuantitativo para calcular probabilidades asociadas a un conjunto de datos, por suerte, la Teoría de Extremos tiene un resultado que nos ayudará con este problema pero antes veamos dos distribuciones importantes.

Definición 4: Definimos la distribución generalizada de valores extremos (DGVE) como

$$H_{\xi,a,b}(x) = \begin{cases} \exp\left(-\left[1 + \xi\left(\frac{x-b}{a}\right)\right]^{-\xi^{-1}}\right), & \xi \neq 0 \\ \exp\left(\exp\left[-\frac{x-b}{a}\right]\right) & , \quad \xi = 0 \end{cases}$$

donde x es tal que $1 + \frac{\xi(x-b)}{a} > 0$.

Esta distribución es importante dado que encapsula las tres distribuciones de extremos. Si $\xi > 0$, $\xi = 0$ o $\xi < 0$, la DGVE puede ser reparametrizada para obtener las distribuciones Fréchet, Gumbel o Weibull respectivamente.

Definición 5: Definimos la distribución generalizada de Pareto (DGP) como

$$P_{\xi,a,b}(x) = \begin{cases} 1 - \left(1 + \xi\left[\frac{x-b}{a}\right]\right)^{-\xi^{-1}}, & \xi \neq 0 \\ 1 - \exp\left(-\frac{x-b}{a}\right) & , \quad \xi = 0 \end{cases}$$

donde $x \geq b$ si $\xi \geq 0$ y $0 \leq x \leq -\frac{a}{\xi} + b$ si $\xi < 0$.

La distribución anterior es la protagonista del Teorema de Pickands – Balkema – de Hann explicado a continuación.

Teorema 10 (Pickands – Balkema – de Hann): $F \in D(H_{\xi,a,b})$ si y solo si existe una función positiva y medible a tal que

$$\lim_{u \rightarrow \omega_F} \sup \left| \frac{\bar{F}(x+u)}{\bar{F}(u)} - \bar{P}_{\xi,a(u),0}(x) \right| = 0$$

donde el supremo se toma para $x \in [0, \omega_F - u)$.

En términos aplicables, el Teorema de Pickands – Balkema – de Hann nos dice que para umbrales u suficientemente grandes, es posible aproximar $\bar{F}(x+u) \approx \bar{F}(u)P_{\xi,a,0}(x)$.

El método de exceso de umbrales aplicado

Explicada la teoría, procedemos a aplicarla para analizar la distribución de los tiempos de extinción. Siendo precisos, usaremos la FEPE y el extremo derecho para descartar dominios de atracción, fijaremos un umbral u y luego estimaremos parámetros adecuados para usar el Teorema de Pickands – Balkema – de Hann y aproximar la cola de la distribución con la DGP.

Usando el lenguaje de programación R, simulamos 100,000 tiempos de extinción con los parámetros descritos en la figura 2 y un tiempo máximo de simulación 100.

Histograma y gráfica de dispersión

Nuestro primer objetivo será determinar si la distribución de los tiempos de extinción tiene extremo derecho finito o infinito. Para este propósito se muestra un histograma y una gráfica de dispersión de los tiempos de extinción.

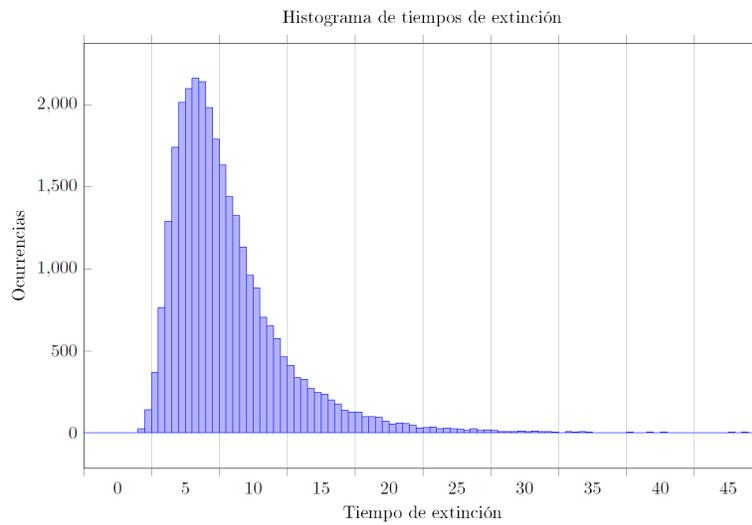


Figura 3. Muestra de 30,000 tiempos de extinción con los mismos parámetros que en la figura 2. Se simuló hasta el tiempo 100 para capturar virtualmente todos los tiempos de extinción, sin embargo, solo se muestran los tiempos menores o iguales a 50.

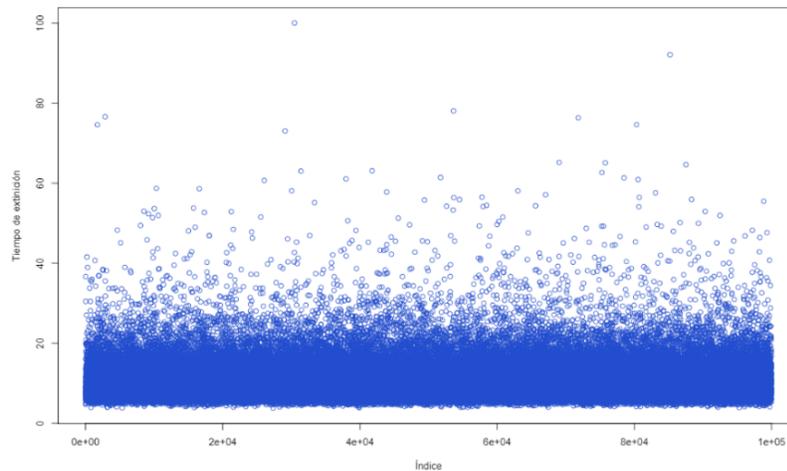


Figura 4. Gráfica de dispersión de los 100,000 tiempos de extinción

Como podemos apreciar, no aparenta haber un valor máximo que los tiempos de extinción puedan tomar, por lo tanto, supondremos que $\omega_F = \infty$ descartando así el dominio de atracción Weibull.

Función de excesos promedio empírica

Para seguir descartando dominios de atracción, graficaremos la función de excesos promedio empírica $e_{n,F}(u)$ y su versión "reescalada" $\frac{e_{n,F}(u)}{u}$ (en nuestro caso, n es 100,000).

De la primera gráfica de la figura 5 podemos apreciar que la FEPE obtenida parece ser una función creciente que diverge a infinito, sin embargo, es la figura 6 la que nos interesa realmente. En la primera gráfica de la figura 6, podemos apreciar que la FEPE reescalada decrece. Puede parecer que decrece hasta converger a 0 lo cual descartaría el dominio de atracción Fréchet, sin embargo, si observamos la segunda gráfica podemos apreciar que se llega a un tipo de mínimo indicado por la línea horizontal. Esto sugiere que la FEPE reescalada converge a alguna constante positiva descartando así el dominio de atracción Gumbel, en otras palabras, no hemos encontrado evidencia en contra del dominio de atracción Fréchet y por lo tanto, es sensato pensar que en efecto la distribución está en el dominio de atracción Fréchet.

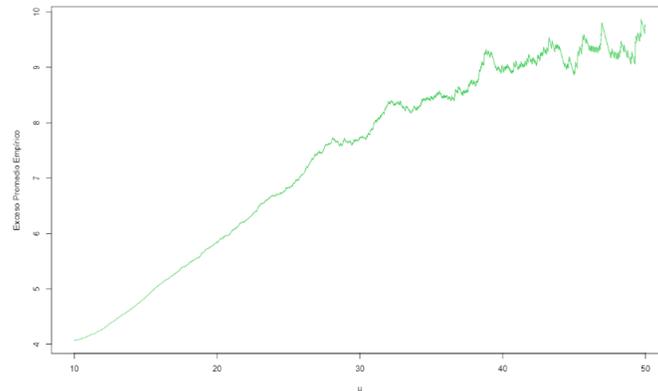


Figura 5. FEPE con dominio [10,50].

Se ha de notar que se puede graficar hasta un dominio [0,100], sin embargo, esto es poco recomendable dada la poca cantidad de datos mayores a 50 (se pueden apreciar datos individuales en la segunda gráfica en forma de 'picos') y el hecho que la FEPE reescalada puede diverger conforme u tiende a 0, además, no hay tiempos de extinción menores a ~ 3 .

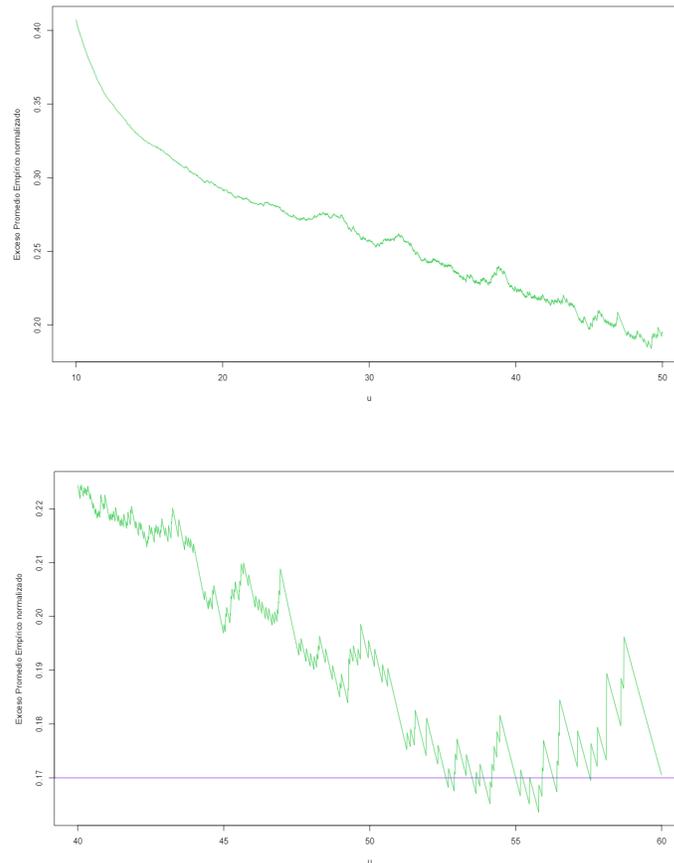


Figura 6. FEPE reescalado con dominios $[10,50]$ y $[40,60]$.

Selección de umbral

Ahora es momento de seleccionar un umbral para luego ajustar una distribución generalizada de Pareto. Esto debe hacerse con algo de cuidado; si por un lado se escoge un umbral demasiado pequeño, la DGP no va a aproximar bien la cola posterior al umbral, si por otro lado se escoge un umbral demasiado grande, no se van a tener suficientes datos para ajustar la DGP de manera correcta causando demasiado sesgo y varianza en la estimación de sus parámetros. Agregado a esto, siempre se desea un umbral lo menor posible para así captar una mayor proporción de los datos.

Una manera de hacer la selección es graficar los parámetros de forma y escala estimados de la DGP variando el umbral, hecho esto, se busca alguna sección donde los parámetros sean relativamente constantes respecto al umbral y se selecciona como umbral el menor de los valores de esta sección, se suele sugerir buscar umbrales entre los cuantiles empíricos $q_{0.95}$ y $q_{0.99}$.

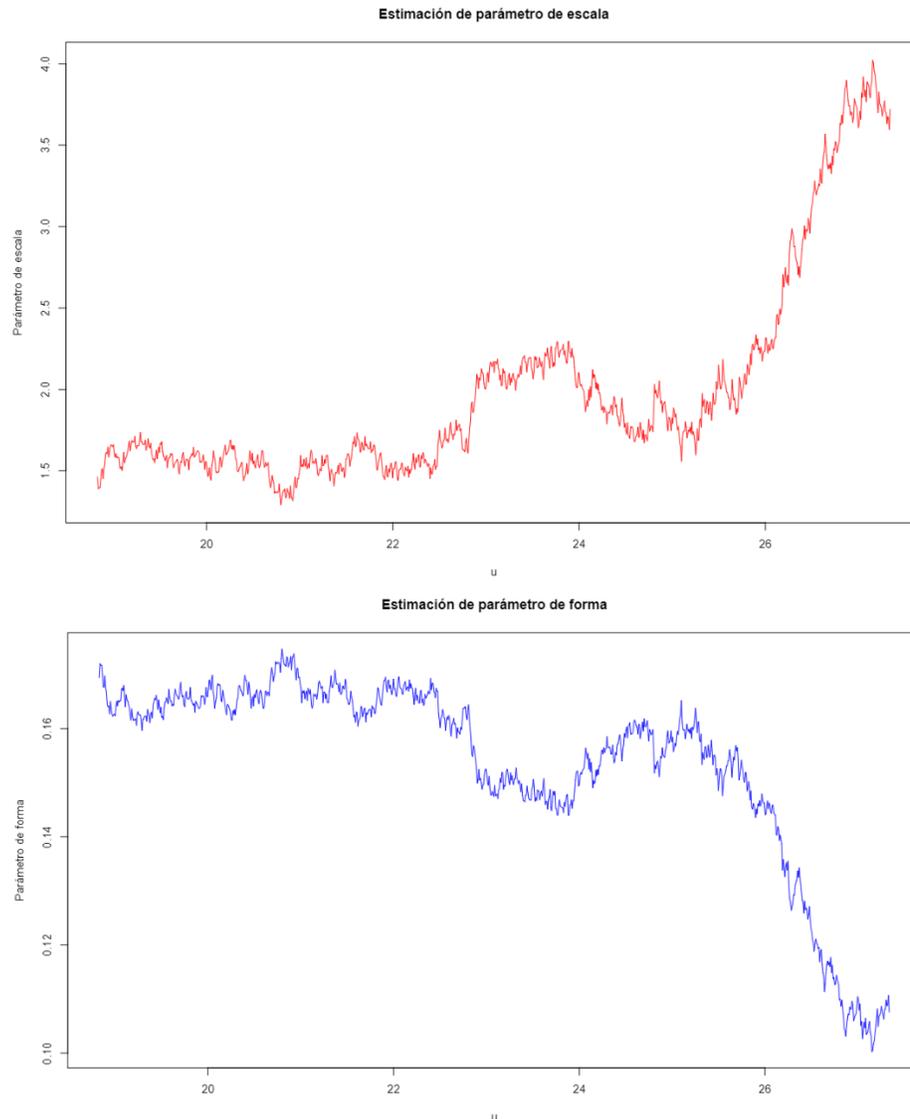


Figura 7. Estimación de parámetros de la DGP para umbrales entre 18.83 y 27.34.

En nuestro caso sucede que $q_{0.95} = 18.83$ y $q_{0.99} = 27.34$ lo cual explica el dominio usado en la figura 7. Como podemos apreciar, los parámetros de forma² y escala son relativamente constantes en los subintervalos 18.83 y 22 por lo que escogemos como umbral $u = q_{0.95} = 18.83$.

Verificación de ajuste

Con nuestro umbral seleccionado, podemos hacer el ajuste de la DGP a los datos posteriores al umbral usando una vez más el paquete ismev en R. Haciendo esto obtenemos parámetros de forma y escala

$$\xi = 0.169 \quad y \quad a = 4.652$$

respectivamente. Nótese que en el Teorema de Pickands – Balkema – de Hann, el parámetro de forma ξ es el mismo para la DGP y para la DGVE, por lo tanto, nuestra estimación del parámetro de forma coincide con

² Las gráficas en la figura 7 fueron hechas con el paquete ismev de R. Para estimar los parámetros, ismev usa el método de máxima verosimilitud (MLE por su sigla en inglés).

nuestra sospecha que la distribución de los tiempos de extinción pertenece al dominio de atracción Fréchet. Para verificar la precisión de nuestro ajuste, veamos las gráficas PP y QQ en la figura 8.

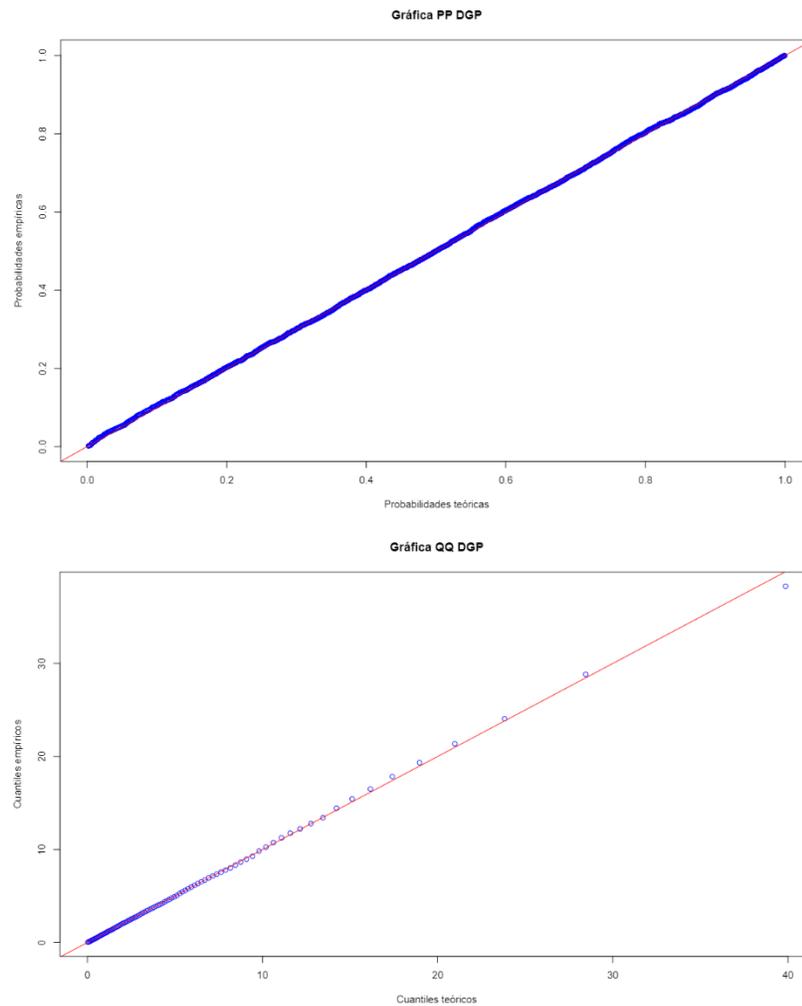


Figura 8. Gráficas PP y QQ de los datos sobre el umbral seleccionado menos dicho umbral comparado con la DGP ajustada. Como muestra para la gráfica PP se usaron todos los datos procesados, como muestra de probabilidades para la gráfica QQ se usaron 0, 1%, ..., 100%.

Podemos notar que las gráficas PP y QQ son casi totalmente rectas, es decir, las probabilidades y cuantiles teóricos y empíricos coinciden casi perfectamente lo que indica un muy buen ajuste del modelo a los datos.

Como última prueba, en la figura 9 se compara la distribución empírica de una muestra de 1,000,000 de tiempos de extinción con la aproximación $\underline{F}(x + u) \approx \underline{F}(u)P_{\xi,a,0}(x)$, es decir, $F(x) \approx 0.95 + 0.05P_{\xi,a,0}(x - u)$.

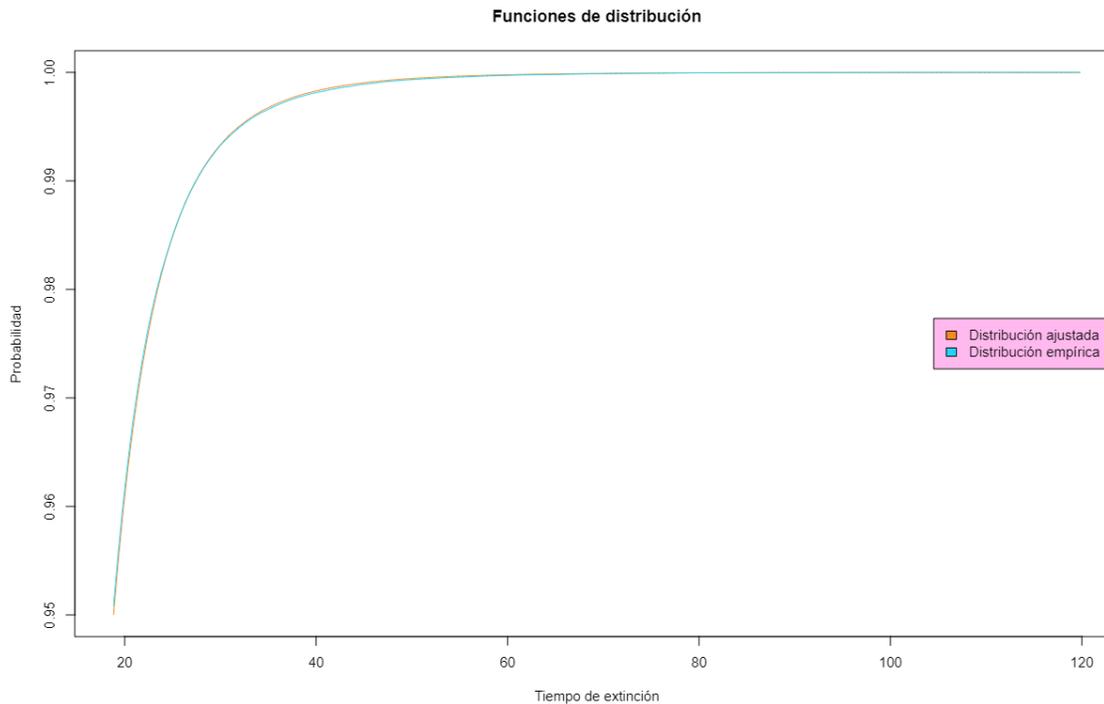


Figura 9. Comparación de la distribución empírica y la aproximación, nótese como ambas gráficas son casi indistinguibles.

Esta prueba muestra que la aproximación mencionada es capaz de extrapolar tiempos de extinción por encima del umbral $u = 18.83$, es decir, es una aproximación muy útil.

Conclusiones

Considerando los resultados descritos en la sección 5.4, se puede concluir que hay suficiente evidencia empírica para decir que:

1. El tiempo de extinción del modelo logístico estocástico con los parámetros descritos en la figura 2 tiene una distribución perteneciente al dominio de atracción Fréchet, es decir, es una distribución con cola de variación regular.
2. Con el umbral correspondiente al cuantil empírico del 95%, se puede aproximar eficazmente la distribución de los tiempos de extinción más grandes (que exceden el umbral) con la distribución generalizada de Pareto con parámetros de forma y escala $\xi = 0.169$ y $a = 4.652$ respectivamente.

Se ha de notar que como los tiempos de extinción son producto de un objeto puramente matemático, la conclusión 1 es una proposición matemática, por lo tanto, se requiere de una demostración formal para aseverar con total seguridad la conclusión 1; dicha demostración podría ser un camino para más investigación en el futuro.

Por otro lado, también resultaría interesante considerar un modelo basado en una ecuación diferencial estocástica en la que la integración sea respecto a un proceso distinto del movimiento browniano.

Referencias

- Calatayud, J., Cortés, J. C., Dorini F. A. & Jornet M. (2020). On a stochastic logistic population model with time-varying carrying capacity. *Computational and Applied Mathematics*, 39(288). <https://doi.org/10.1007/s40314-020-01343-z>
- Golec, J. & Sathanathan, S. (2003). Stability analysis of a stochastic logistic model. *Mathematical and Computer Modelling*, 38(5 – 6). [https://doi.org/10.1016/S0895-7177\(03\)90029-X](https://doi.org/10.1016/S0895-7177(03)90029-X)
- Heffernan, J. E. & Stephenson, A. G. (2018). *ismev: An Introduction to Statistical Modeling of Extreme Values*. cran. <https://cran.r-project.org/package=ismev>
- Leadbetter, M. R., Lindgren, G. & Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer Science & Business Media.
- Levins, R. (1969). The Effect of Random Variations of Different Types on Population Growth. *Proceedings of the National Academy of Sciences of the United States of America*, 62(4). <http://www.jstor.org/stable/59358>
- Norden, R. H. (1982). On the Distribution of the Time to Extinction in the Stochastic Logistic Population Model. *Advances in Applied Probability*, 14(4). <https://doi.org/10.2307/1427019>
- Paolo, B. (2017). *Stochastic Calculus: An Introduction Through Theory and Exercises*. Springer.
- Ribatet, M., Singleton, R. & R Core Team (2022). *SpatialExtremes: Modelling Spatial Extremes*. <https://cran.r-project.org/package=SpatialExtremes>.
- Roy, T., Ghosh, S., Saha, B. & Bhattacharya S. (2022). A noble extended stochastic logistic model for cell proliferation with density-dependent parameters, *Scientific reports*, 12. <https://doi.org/10.1038/s41598-022-12719-y>
- Shah, A. (2013). Stochastic Logistic Model for Fish Growth. *Open Journal of Statistics*, 4(1). <http://dx.doi.org/10.4236/ojs.2014.41002>