



# **Universidad de Granada**

## **Estudio de algoritmos de optimización aplicados a procesos de difusión**

Alumno:

Luis Miguel Jáimez Rodríguez

Director:

Francisco de Asís Torres Ruiz

## **Máster oficial en Estadística Aplicada**



# Índice general

<b>1. INTRODUCCIÓN</b>	<b>5</b>
<b>2. PROCESOS DE DIFUSIÓN ASOCIADOS A LAS CURVAS GOMPERTZ Y LOGÍSTICO</b>	<b>9</b>
2.1. BREVES NOTAS SOBRE LAS CURVAS GOMPERTZ Y LOGÍSTICA . . . . .	9
2.1.1. Curva Gompertz . . . . .	9
2.1.2. Curva Logística . . . . .	14
2.2. PROCESOS DE DIFUSIÓN ASOCIADOS A LAS CURVAS TIPO GOMPERTZ Y LOGÍSTICO	17
2.2.1. Proceso lognormal con factores exógenos . . . . .	17
2.2.2. Características del modelo . . . . .	22
2.2.3. Adaptación del proceso lognormal a los procesos Gompertz y Logístico . . . . .	25
<b>3. ALGORITMOS ESTOCÁSTICOS DE OPTIMIZACIÓN</b>	<b>31</b>
3.1. CONCEPTOS PREVIOS . . . . .	31
3.1.1. Sobre algoritmos . . . . .	31
3.1.2. Sobre problemas de optimización . . . . .	32
3.2. ALGORITMO SIMULATED ANNEALING (S.A.) . . . . .	32
3.2.1. Presentación del algoritmo . . . . .	33
3.2.2. Problemas discretos frente a continuos . . . . .	36
3.2.3. Resultados de convergencia . . . . .	36
3.3. ALGORITMO VNS . . . . .	41
3.3.1. Extensiones de la VNS . . . . .	45
<b>4. INFERENCIA EN LOS PROCESOS DE DIFUSIÓN TIPO GOMPERTZ Y LOGÍSTICO</b>	<b>49</b>
4.1. CASO LOGARÍTMICO NORMAL CON FACTORES EXÓGENOS . . . . .	50
4.2. INFERENCIA EN EL MODELO DE GOMPERTZ . . . . .	52
4.3. INFERENCIA EN EL MODELO LOGÍSTICO . . . . .	54
<b>5. APLICACIÓN DE LOS ALGORITMOS</b>	<b>57</b>
5.1. APLICACIÓN CON DATOS SIMULADOS . . . . .	57
5.1.1. Simulación de los procesos . . . . .	57
5.2. ESTIMACIÓN EN EL PROCESO DE GOMPERTZ CON DATOS SIMULADOS: . . . . .	59
5.2.1. Estimaciones con Simulated Annealing . . . . .	59

5.2.2. Estimaciones con VNS . . . . .	63
5.3. ESTIMACIÓN EN EL PROCESO LOGÍSTICO CON DATOS SIMULADOS: . . . . .	64
5.3.1. Estimaciones con VNS . . . . .	66
5.4. CONCLUSIONES . . . . .	68
<b>A. DEFINICIONES Y RESULTADOS QUE SE USARÁN</b>	<b>69</b>
A.1. SOBRE PROCESOS DE DIFUSIÓN . . . . .	69
A.1.1. Ecuaciones de Kolmogorov y de Focker-Planck en procesos de difusión . . . . .	70
A.2. SOBRE LA DISTRIBUCIÓN NORMAL MULTIVARIANTE Y LOGARÍTMICO-NORMAL	74
<b>B. CÓDIGOS DE R</b>	<b>77</b>
B.1. SIMULACIÓN DE PROCESOS GOMPERTZ Y LOGÍSTICO . . . . .	77
B.1.1. Simulación del proceso de Gompertz . . . . .	77
B.1.2. Simulación del proceso logístico . . . . .	78
B.2. CÓDIGO DEL SIMULATED ANNEALING EN R ADAPTADO AL PROBLEMA DE ESTI- MACIÓN EN EL MODELO DE GOMPERTZ . . . . .	79
B.3. CÓDIGO DEL SIMULATED ANNEALING EN R ADAPTADO AL PROBLEMA DE ESTI- MACIÓN EN EL MODELO LOGÍSTICO . . . . .	83
B.4. CÓDIGO DEL VNS PARA DISTINTAS VERSIONES: . . . . .	85
B.4.1. VNS reducida . . . . .	86
B.4.2. VNS descendente . . . . .	86

## Capítulo 1

# INTRODUCCIÓN

La modelización y estudio de fenómenos asociados a sistemas dinámicos ha experimentado un gran auge en las últimas décadas. El motivo principal para ello radica en la necesidad de comprender los mecanismos de evolución de los sistemas con el fin de dar una explicación a su comportamiento, permitiendo predecir el mismo sin perder de vista la posible inclusión de influencias ajenas a las variables en estudio que permitan alterar dicho comportamiento y, con ello, tener la posibilidad de controlar externamente la evolución del fenómeno en consideración. Tradicionalmente los modelos empleados para estos fines son determinísticos y nacieron a partir de la modelización mediante ecuaciones diferenciales ordinarias cuyas soluciones son curvas que describen distintos tipos de esquemas de crecimiento. Entre ellos podemos citar el exponencial, asociado al modelo malthusiano, así como el logístico y el gompertziano, relacionados con crecimientos de tipo sigmoidal. Aunque estos modelos son los más extensamente empleados (de hecho fueron los primeros que aparecieron en el estudio de fenómenos de crecimiento), posteriormente han ido apareciendo otros muchos, entre los que podemos citar el modelo de Von Bertalanffy, el de Richards, monomolecular, Blumberg, Weibull, doble sigmoidal. . . , cada uno de los cuales presenta alguna particularidad que le hace más propicio para estudiar determinados esquemas de crecimiento.

Sin embargo, los fenómenos en estudio suelen presentar una elevada complejidad que suele conllevar la especificación de múltiples factores que no siempre son conocidos o cuantificables. Este inconveniente se puede evitar mediante el uso de modelos estocásticos entre los que destacan los procesos de difusión, parte de los cuales están asociados a modelos determinísticos y surgen al introducirles una fluctuación aleatoria en la ecuación diferencial asociada.

En consecuencia, la introducción y estudio de procesos de difusión, con la finalidad del ajuste de modelos a situaciones reales, es una línea de investigación que ha alcanzado una gran importancia. Ello ha llevado a desarrollar diversas líneas de índole teórica relacionadas tanto con el estudio inferencial de los modelos como con el estudio de variables temporales asociadas a ellos, entre otras.

En lo que se refiere a la inferencia, una de las líneas fundamentales en las que más se ha profundizado es la derivada de la consideración de muestreo discreto de trayectorias, tratando la estimación máximo verosímil de los parámetros y de funciones paramétricas de interés (con fines predictivos) como la funciones media, moda y percentiles del proceso. La estimación máximo verosímil de los parámetros conlleva a la consideración de sistemas de ecuaciones cuya solución explícita no está garantizada a priori, por lo que para resolverlos se hace

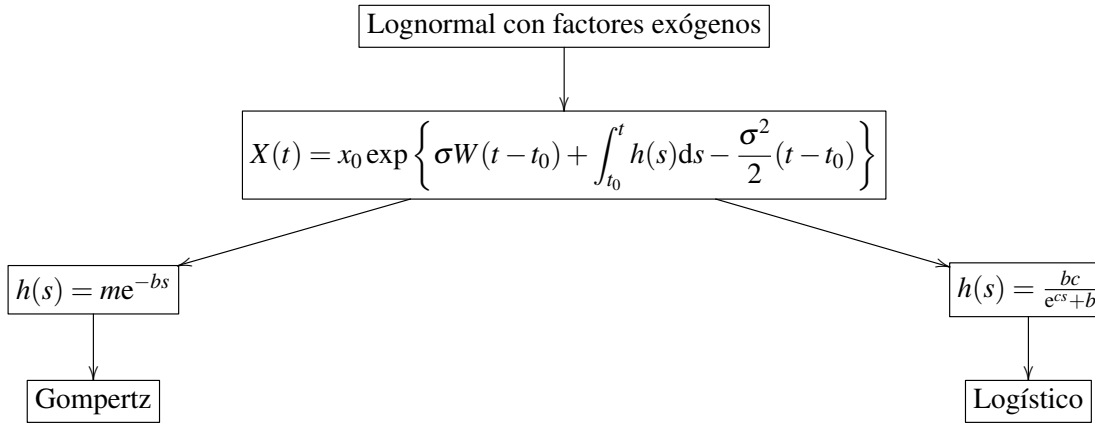
necesario el empleo de procedimientos numéricos. No obstante, puesto que las ecuaciones que se obtienen dependen de los valores muestrales de las trayectorias observadas, no se puede garantizar siempre que los métodos empleados puedan ser utilizados. Por ejemplo, se han descrito casos concretos en los que métodos como el de Newton-Raphson o el de bisección presentan problemas al no verificarse las hipótesis que permiten su empleo. Por lo tanto la estimación máximo verosímil de los parámetros debe abordarse de forma directa maximizando la función de verosimilitud.

Aquí entra en juego una clase de algoritmos ampliamente usados en una gran variedad de problemas de muy distinta naturaleza entre sí: las metaheurísticas, que se definen como estrategias generales para diseñar procedimientos heurísticos para resolver un problema de optimización mediante un proceso de búsqueda en un cierto espacio de soluciones alternativas. En computación dos objetivos fundamentales son encontrar algoritmos con buenos tiempos de ejecución y buenas soluciones, siendo una heurística un algoritmo que abandona uno ó ambos objetivos, y son usados normalmente cuando no existe una solución óptima bajo las restricciones dadas. Se puede decir que un problema de optimización es difícil de resolver cuando no puede hacerse garantizando máxima optimalidad, o conociendo una cota de error mediante un método exacto (determinístico) con un tiempo límite razonable.

Las metaheurísticas se aplican generalmente a los problemas para los que no existe algoritmo específico satisfactorio para resolverlos. Casi todas las metaheurísticas comparten las siguientes características: están inspiradas en la naturaleza (basado en algunos principios de la física, la biología o la etología), hacen uso de componentes estocásticos (que involucran variables aleatorias), no utilizan el gradiente o matriz hessiana de la función objetivo y tienen varios parámetros que deben introducirse apropiadamente en el problema en cuestión.

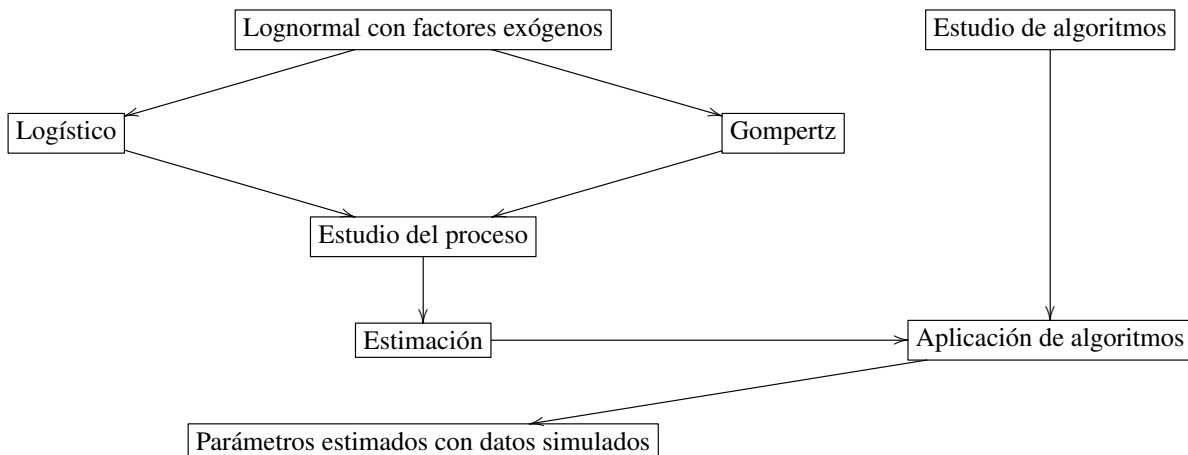
Un metaheurística tendrá éxito en un problema de optimización dado si se puede proporcionar un equilibrio entre la exploración (diversificación) y la explotación (intensificación). Se necesita explotación para identificar las partes de el espacio de búsqueda con soluciones de alta calidad. En la explotación es importante intensificar la búsqueda en algunas áreas prometedoras de la experiencia de búsqueda acumulada. Las principales diferencias entre las metaheurísticas existentes se refieren a la forma particular en que ellos tratan de lograr este equilibrio.

En este trabajo nos centraremos en el Simulated Annealing (S.A.) y en el Variable Neighborhood Search (V.N.S.) descendente híbrido con el S.A. para aplicarlo al problema de estimación máximo verosímil de los parámetros en los procesos de tipo Gompertz y Logístico. Estos procesos pueden ser obtenidos como casos particulares del proceso *lognormal con factores exógenos*. Por ello, en el Capítulo 2 se hace una breve descripción del proceso lognormal con factores exógenos y de su adaptación para obtener los procesos tipo Gompertz y logístico. El esquema es el siguiente:



En el Capítulo 3 se hace un breve resumen de las principales características de los algoritmos de S.A. y V.N.S. y de sus variantes más usuales. En el Capítulo 4 se plantea el problema de la estimación máximo versímil de los parámetros de los procesos tipo gompertz y logístico. Se parte del caso general del proceso lognormal con factores exógenos y se particulariza posteriormente para los procesos considerados. En ambos casos se presentan los sistemas de ecuaciones de verosimilitud, mostrando su complejidad y comentando cómo no es posible asegurar que los métodos clásicos de resolución puedan ser empleados. Finalmente, en el Capítulo 5 se procede a la aplicación de los algoritmos de optimización para la estimación de los parámetros de los procesos. Los datos que se han usado han sido obtenidos mediante la simulación de los procesos, cuestión que se ha realizado a partir de la expresión de ambos como transformados del proceso Wiener. Posteriormente se han adaptado los algoritmos al caso particular de cada proceso. Particular atención ha merecido la acotación de los espacios paramétricos para la aplicación de S.A., cuestión que se ha resuelto a partir de la información muestral proporcionada por las trayectorias. Una vez planteados los algoritmos, estos se han implementado en R, programas que pueden ser consultados en el apéndice B del trabajo.

De esta forma, el esquema general del trabajo es el siguiente:







## Capítulo 2

# PROCESOS DE DIFUSIÓN ASOCIADOS A LAS CURVAS GOMPERTZ Y LOGÍSTICO

### 2.1. BREVES NOTAS SOBRE LAS CURVAS GOMPERTZ Y LOGÍSTICA

El crecimiento es una importante característica en muchos campos de aplicación. El estudio de este fenómeno fue originalmente asociado con la evolución de poblaciones animales. Sin embargo, actualmente es considerado en muchos más contextos como Economía, Biología, Ecología y muchos otros. Por esta razón, se han hecho muchos intentos por construir modelos matemáticos que describan este tipo de comportamientos.

Muchas y diversas representaciones del crecimiento han sido propuestas, con una amplia variedad de curvas asociadas a modelos determinísticos describiendo el comportamiento mostrado por cada una de ellas. La más representativa en sus inicios fue la *exponencial*, que es monótona, no acotada, y cóncava. El modelo determinístico asociado es conocido como *modelo malthusiano*, propuesto para estudiar patrones de comportamiento de poblaciones humanas. El problema es que se suponía no-acotación en el crecimiento cuyo radio depende de la densidad de población, aunque este modelo puede ser apropiado a corto plazo, cuando una población comience a desarrollarse (habrá recursos de sobra para los individuos). El problema surge cuando el tiempo empieza a crecer, la densidad de población es mayor y los recursos empiezan a ser limitados, lo que hace que la mortalidad aumente. De aquí que se planteen otro tipo de modelos de crecimiento.

Es a partir de aquí cuando se proponen modelos de crecimiento acotados surgiendo, particularmente, modelo de *Gompertz* y *Logístico*. Este tipo de curvas se conocen como *sigmoidales* (con forma de S) porque son monótonas y presentan un punto de inflexión donde cambian de cóncavas a convexas. Ambas tienen un patrón similar de crecimiento, cuya principal diferencia está en la localización del punto de inflexión: en la curva tipo Gompertz se sitúa más inicialmente (sobre 35 % al 40 %), mientras que en el modelo logístico se encuentra en medio del crecimiento total.

#### 2.1.1. Curva Gompertz

Su nombre procede del matemático Benjamin Gompertz, el cual la introduce por primera vez en un artículo que trataba sobre la naturaleza de la expresión de la función de la ley de la mortalidad humana. En este trabajo

aparece la función:

$$L_x = dg^{q^x} \quad \text{con} \quad \begin{cases} \log(g) = \frac{mq^{-a}}{1-q^r} \\ q = p^{1/r} \\ m = \log(L_a) - \log(L_{a+r}) \\ d = \frac{L_a}{\epsilon} \\ \log(\epsilon) = \frac{m}{1-q^r} \end{cases}$$

siendo  $L_x$  el tamaño de la población en el instante  $x$ ;  $a$  instante inicial;  $r$  unidad de salto considerada en el tiempo y  $p$  la razón de la progresión geométrica que hay entre el número de personas vivas en el tiempo.

Fue intensamente usado para modelizar y describir patrones de comportamiento en varios campos de aplicación. En particular, es usado en fenómenos de crecimiento con dos características distintivas: *crecimiento acotado* y *tendecia sigmoidal*. Estas características se encuentran, por ejemplo, en el estudio de crecimiento de poblaciones, ya que no crecen indefinidamente. En su lugar, su crecimiento máximo tiende a ser alrededor de la capacidad máxima de su entorno. Cuando se estudia la difusión de innovaciones, el crecimiento es normalmente rápido al principio, pero se ralentiza con el tiempo, cuanto más saturado esté.

Desde que fue introducida por Gompertz, para modelizar la mortalidad humana han surgido varias versiones de esta curva llamadas genéricamente '*tipo Gompertz*'.

Tras la primera forma introducida de este tipo de curva de crecimiento, posteriormente se introducen variantes debido a que se presentan limitaciones sobre su aplicación al trasladarlo a distintos campos (Economía, Biología,...).

Uno de los grandes descubrimientos sobre este modelo fue explicar el crecimiento de tumores en lugar de usar el modelo exponencial que es el que se había considerado siempre, ya que anteriormente fue comúnmente aceptado que el crecimiento de un tumor bajo condiciones ideales seguía una tendencia siempre exponencial hasta el fin del crecimiento (muerte del sujeto). Se observó que aunque el ajuste era bueno en unas primeras fases, posteriormente dejaba de serlo ya que se observaba que el crecimiento era cada vez más y más lento. En este sentido se puede ver en [2] como se empieza a tratar de explicar este crecimiento usando el modelo de Gompertz, en vista de observaciones experimentales.

En la siguiente figura se representan curvas de ambos tipos, junto a una trayectoria simulada del proceso Gompertz, y se ve que el ajuste inicial aceptable en ambos, deja de serlo en tiempos posteriores. Este tipo de crecimiento fue el que se observó en crecimiento de tumores.

En esta misma línea también puede modelizarse el crecimiento de tumores bajo el efecto de un tratamiento, como puede verse en [3]. En principio se asume que el crecimiento es de tipo Gompertz, y el efecto del tratamiento se añade a través de una función  $C(t)$ . El principal objetivo es el ajuste de dicha función (en principio desconocida). Se observa que bajo distintos tratamientos la función puede ser ajustada, lo que daría una medida de la bondad de dichos tratamientos sobre la cura de tumores.

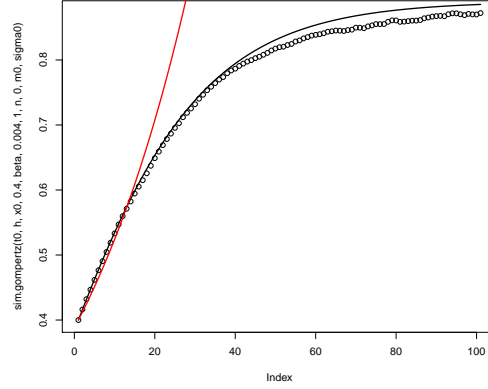


Figura 2.1: Comparación de curva exponencial y de tipo Gompertz

Pasando al estudio de este tipo de curvas, algunas variantes a considerar:

1. **Winsor (1932)**: reescribe la curva original transformándola en una doble exponencial:

$$\left. \begin{array}{l} d = k \\ g = e^{-a} \\ q = e^{-b} \end{array} \right\} \Rightarrow f(t) = ke^{-e^{a-bt}} \quad t \in \mathbb{R}$$

que para valores de  $t \geq t_0 \geq 0$  puede escribirse de la forma

$$f(t) = k \exp \left\{ \log \left( \frac{x_0}{k} \right) e^{-b(t-t_0)} \right\} \quad t \geq t_0.$$

Esta es una curva monótona creciente con un punto de inflexión cuando  $k > x_0 e$ , que viene dado por

$$\left( \frac{\log \left( \log \left( \frac{k}{x_0} \right) \right)}{b} + t_0, \frac{k}{e} \right)$$

2. **Capocelli y Ricciardi (1974)**: usan la curva

$$f(t) = \exp \left\{ \frac{\alpha}{\beta} \left( 1 - e^{-\beta(t-t_0)} \right) - \log(x_0) e^{-\beta(t-t_0)} \right\}$$

con  $\beta > 0$ ,  $0 < x_0 < e^{\alpha/\beta}$ . Curva sigmoideal, con cota  $e^{\alpha/\beta}$  independiente del valor inicial, y con punto de inflexión en

$$\left( \frac{\log \left( \frac{\alpha}{\beta} - \log(x_0) \right)}{\beta} + t_0, e^{\alpha/\beta-1} \right)$$

cuando  $x_0 < e^{\alpha/\beta-1}$ .

3. **Skiadas (1994)**: consideró la expresión

$$f(t) = \exp \left\{ \log(x_0) e^{-\beta(t-t_0)} \right\}$$

caso particular del anterior para  $\alpha = 0$ .

4. **Tan (1986)**: al introducir el proceso de Gompertz de nacimiento y muerte consideró la expresión

$$f(t) = x_0 \exp \left\{ \frac{\alpha}{\beta} \left( 1 - e^{-\beta(t-t_0)} \right) \right\} \quad t \geq t_0 \geq 0$$

con  $\alpha > \beta > 0$  y  $x_0 > 0$ . Es una curva sigmoideal con cota  $k = x_0 e^{\alpha/\beta}$  y punto de inflexión

$$\left( \frac{\log \left( \frac{\alpha}{\beta} \right)}{\beta} + t_0, x_0 e^{-\alpha/\beta - 1} \right)$$

En este trabajo se parte de una curva de crecimiento del último tipo antes mencionado, pero si se llama  $m = \alpha e^{t_0}$ , puede escribirse como

$$x(t) = x_0 \exp \left\{ \frac{m}{\beta} \left( e^{-\beta t_0} - e^{-\beta t} \right) \right\} \quad t \geq t_0 \geq 0$$

que tiene un punto de inflexión en

$$\left( \frac{\log \left( \frac{m}{\beta} \right)}{\beta}, x_0 \exp \left\{ \frac{m}{\beta} (e^{-\beta t_0} - 1) \right\} \right).$$

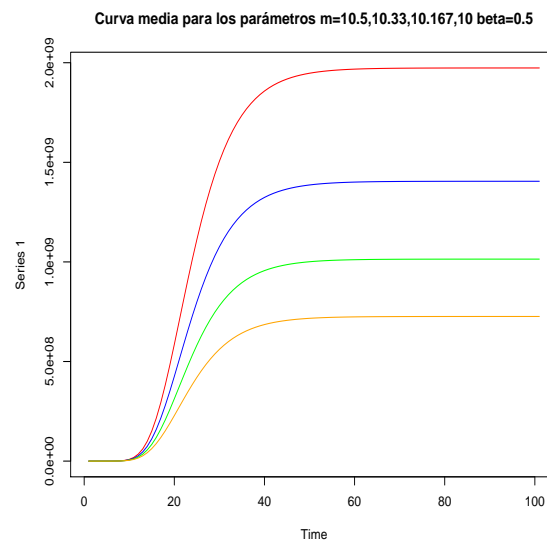
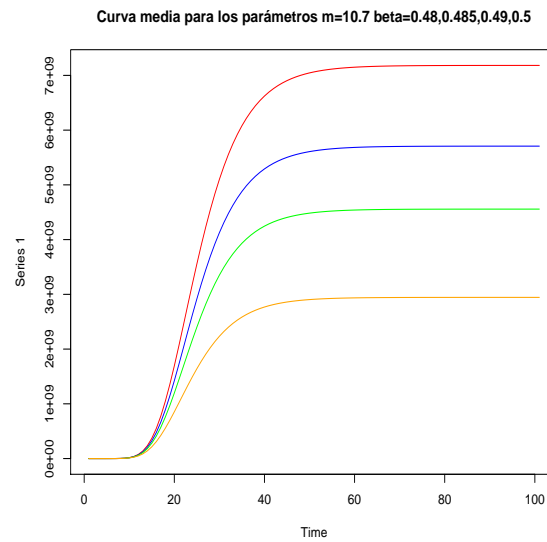
y es solución de la ecuación diferencial

$$\begin{cases} \frac{dx(t)}{dt} = m e^{-\beta t} x(t) \\ x(t_0) = x_0 \end{cases}$$

modelo determinístico asociado a esta clase de problemas de crecimiento.

**Algunas representaciones gráficas de la curva:**

A continuación mostramos algunos ejemplos de la curva. En la cabecera aparecen los valores de los parámetros, donde se ha de entender que cuando sólo aparece un valor en alguno de los parámetros, es que las cuatro curvas toman el mismo, y se corresponden respectivamente con los colores rojo, azul, verde y naranja.



### 2.1.2. Curva Logística

Esta curva fue introducida por Verlhust en el siglo XIX con el objetivo de estudiar crecimiento de poblaciones. En 1920 el interés fue reavivado y muchas vías de investigación fueron enfocadas en este modelo, debido a que es muy apropiado para el estudio del desarrollo y evolución de muchos fenómenos de crecimiento.

Existen muchas situaciones en las que inicialmente se experimenta un crecimiento exponencial, y luego en cierto punto el crecimiento es cada vez más lento, y siempre acotado. Este modelo se adapta bien a problemas de este tipo.

Los orígenes del modelo logístico se encuentran en Ecología, donde sirvió para explicar el crecimiento de poblaciones, también en Demografía, y Biología y Medicina para el análisis del crecimiento de bacterias, tumores, y varios tipos de especies de animales y plantas. En los últimos años, sus aplicaciones han sido extendidas a otros campos. Por ejemplo ha sido usada para estudiar a:

1. *Velocidad en comunicaciones móviles [11]*: se estudia el impacto del efecto de retardo de servicio de telecomunicaciones móviles en UE. Se ha demostrado en varios estudios que el tiempo de retardo entre el conocimiento y la fase de adopción de servicios móviles de los usuarios potenciales determina la velocidad de la difusión móvil de servicios de telecomunicaciones y se puede utilizar de forma efectiva para propósitos de ranking o clúster en los casos en los que la difusión de un nuevo producto se realiza en diferentes países. El enfoque de modelado propuesto se origina a partir del proceso Logístico, bien conocido, que no tiene lugar de forma inmediata, pero sí después de un cierto periodo de tiempo. Una modificación adecuada de el modelo propuesto descrito por un lapso de tiempo a través de ecuaciones diferenciales ordinarias puede ser resuelto analíticamente y sus propiedades para la combinación de varios parámetros se investigan.
2. *Producción global de aceite [10]*: Se usan funciones de tipo logístico para caracterizar la producción máxima y última del crudo mundial y derivados del petróleo.
3. *Evolución de premios Nobel [9]*: aquí se usa la ecuación logística, que puede verse como caso particular de las ecuaciones de Lokta-Volterra.

Pasando al estudio de esta curva, en principio, el modelo determinístico es definido en términos de la siguiente ecuación diferencial

$$\begin{cases} \frac{dx(t)}{dt} = \alpha x(t) - \beta x^2(t) \\ x(t_0) = x_0 \end{cases}$$

donde  $\alpha, \beta$  constantes positivas tal que  $\alpha$  define la *tasa de crecimiento*, y  $-\beta x^2(t)$  factor de inhibición de dicha tasa. De aquí, si  $x(t)$  pequeño entonces  $x'(t)$  elevado, mostrando un crecimiento inicial exponencial, y a medida que  $x(t)$  es mayor,  $x'(t)$  va decreciendo, siendo  $\lim_{t \rightarrow \infty} x'(t) = 0$ .

Se tiene que la solución de la ecuación, que nos da la curva del modelo logístico, es

$$x(t) = \frac{\alpha/\beta}{1 + \left(\frac{\alpha/\beta}{x_0} - 1\right) e^{-\alpha(t-t_0)}} \quad t \geq t_0$$

que verifica que

1. es creciente cuando  $\alpha > \beta x_0$ .
2.  $\lim_{t \rightarrow \infty} x(t) = \frac{\alpha}{\beta}$  independiente del valor inicial.
3. Tiene un punto de inflexión en

$$t_I = t_0 + \frac{1}{\alpha} \log \left( \frac{\alpha/\beta}{x_0} - 1 \right).$$

El hecho de que  $\lim_{t \rightarrow \infty} x(t) = \frac{\alpha}{\beta}$  no dependa del valor inicial  $x_0$  puede hacer que el modelo no sea adecuado en ciertas situaciones, por lo que puede convenir considerar una generalización de la curva, que haga que dicho límite dependa de  $x_0$ :

$$f(t) = \frac{a}{1 + be^{-ct}} \quad a, b, c > 0$$

con  $f(t_0) = x_0 > 0$ , entonces es  $a = x_0(1 + be^{-ct_0})$ , siendo finalmente la expresión de la curva:

$$f(t) = x_0 \frac{1 + be^{-ct_0}}{1 + be^{-ct}} \quad t \geq t_0 ; \quad b, c > 0$$

que es del primer tipo descrito cuando  $b = \left( \frac{\alpha/\beta}{x_0} - 1 \right) e^{\alpha t_0}$  y  $c = \alpha$ .

Esta curva verifica que:

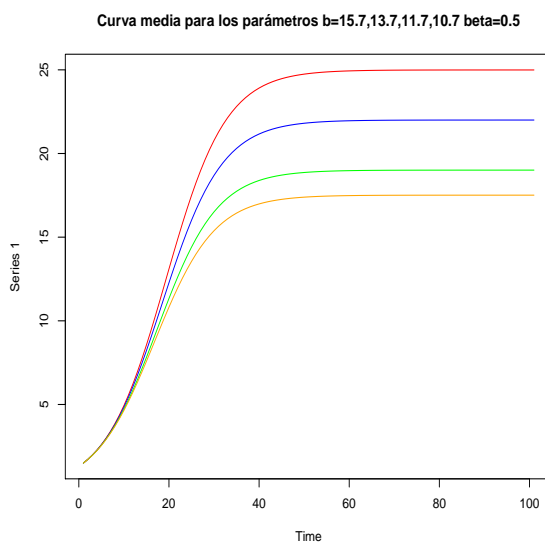
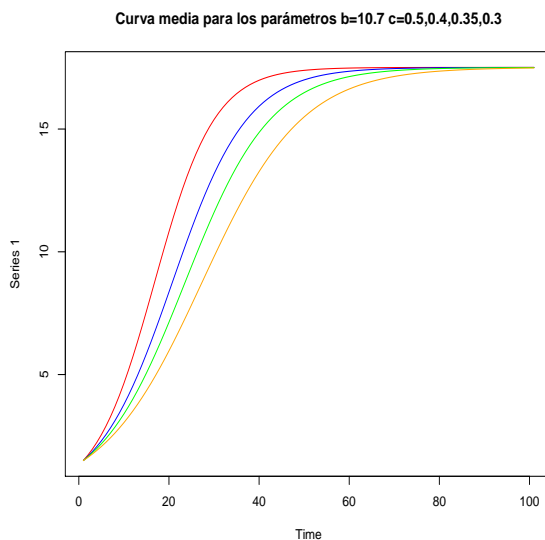
1. Es estrictamente creciente para  $b, c > 0$ .
2.  $\lim_{t \rightarrow \infty} f(t) = x_0(1 + be^{-ct_0})$  dependiente del valor inicial.
3. Se tiene un punto de inflexión en

$$\left( \frac{\log(b)}{c}, \frac{x_0(1 + be^{-ct_0})}{2} \right)$$

y además  $t_I > t_0$  si, y sólo si,  $b > e^{ct_0}$ , y se ve que el valor de la función en el punto de inflexión es la mitad del crecimiento total.

**Algunas representaciones gráficas de la curva:**

A continuación mostramos algunos ejemplos de la curva. En la cabecera aparecen los valores de los parámetros, donde se ha de entender que cuando sólo aparece un valor en alguno de los parámetros, es que las cuatro curvas toman el mismo, y se corresponden respectivamente con los colores rojo, azul, verde y naranja.





## 2.2. PROCESOS DE DIFUSIÓN ASOCIADOS A LAS CURVAS TIPO GOMPERTZ Y LOGÍSTICO

### 2.2.1. Proceso lognormal con factores exógenos

El proceso de difusión lognormal ha sido usado frecuentemente como modelo probabilístico en varios campos donde la variable bajo consideración muestra una tendencia exponencial. La motivación que conduce a incluir factores exógenos en el modelo es introducir una explicación sobre el comportamiento de la variable estudiada por la difusión (variable endógena) en términos de un conjunto de variables externas. Su comportamiento en el tiempo se asume como conocido y deben contribuir a la descripción de la evolución del proceso, así como su control externo con fines de predicción.

Normalmente, la vía de introducir variables externas en el modelo es a través de una función  $h$  dependiente del tiempo que sea continua en el intervalo donde el proceso sea observado.

Este tipo de procesos engloba a los que serán objeto de estudio y aplicación de los algoritmos de optimización, que son los de tipo Gompertz y Logístico, y de ahí su importancia y estudio en este trabajo, además de ser el punto de partida para muchos procesos de difusión de gran interés en la práctica.

**Definición 2.2.1** (Proceso de difusión lognormal con factores exógenos). *Se trata de un proceso con momentos infinitesimales dados por*

$$\begin{aligned} A_1(x, t) &= h(t)x \\ A_2(x) &= \sigma^2 x^2 \end{aligned}$$

con  $\sigma > 0$  y  $h$  función continua y acotada, generalización del modelo Malthusiano donde  $h(t)$  era constante en ese caso.

El estudio de este proceso puede abordarse de distintas formas:

1. *Ecuaciones de Fokker-Planck y Kolmogorov*: estudio desde el punto de vista de ecuaciones en derivadas parciales, con ecuación adelantada (ó de Fokker-Planck) dada por

$$\frac{\partial f(x, t|y, s)}{\partial t} = -h(t) \frac{\partial [xf(x, t|y, s)]}{\partial x} + \frac{\sigma^2}{2} \frac{\partial^2 [x^2 f(x, t|y, s)]}{\partial x^2}$$

y la ecuación atrasada dada por

$$\frac{\partial f(x, t|y, s)}{\partial s} + h(s)y \frac{\partial f(x, t|y, s)}{\partial y} + \frac{\sigma^2}{2} y^2 \frac{\partial^2 f(x, t|y, s)}{\partial y^2} = 0$$

Estas ecuaciones verifican las condiciones iniciales de existencia y unicidad de solución respecto condiciones iniciales  $\lim_{t \downarrow s} f(x, t|y, s) = \delta(x - y)$  y  $\lim_{s \uparrow t} f(x, t|y, s) = \delta(x - y)$ .

2. *Ecuación de Langevin*: En vista del modelo de crecimiento determinístico que se satisface, dado por

$$\begin{cases} \frac{dx(t)}{dt} = h(t)x(t) \\ x(t_0) = x_0 \end{cases}$$

puede considerarse su ecuación de Langevin asociada, que está dada por

$$\frac{dX(t)}{dt} = h(t)X(t) + \sigma X(t)W(t)$$

donde  $W(t)$  es el movimiento browniano estándar en el instante  $t$ .

3. *Ecuaciones diferenciales estocásticas*: Alternativamente, puede considerarse la ecuación diferencial estocástica de Ito, dada por

$$dX(t) = h(t)X(t)dt + \sigma X(t)dW(t)$$

y distribución inicial  $X(t_0) = X_0$  independiente de  $W(t)$  para  $t \geq t_0$ .

Nos centramos en esta última opción: primero se justifica la existencia y unicidad en virtud del Teorema (A.1.5). Usando la misma notación que en ese resultado, es  $a(x, t) = h(t)x$ ,  $b(x, t) = b(x) = \sigma x$ , y se tiene:

Condición de Hölder:  $|a(x, t) - a(y, t)| + |b(x) - b(y)| = (|h(t)| + \sigma)|x - y| \leq \left( \max_{t \in [t_0, T]} |h(t)| + \sigma \right) |x - y|$

Restricción sobre el crecimiento:  $|a(x, t)|^2 + |b(x)|^2 = (h(t)^2 + \sigma^2)x^2 \leq \left( \max_{t \in [t_0, T]} h(t)^2 + \sigma^2 \right) (1 + x^2)$

y si llamamos  $k_1 = \max_{t \in [t_0, T]} |h(t)| + \sigma$  y  $k_2 = \max_{t \in [t_0, T]} h(t)^2 + \sigma^2$ , se toma la constante  $k = \max\{k_1, k_2\}$ , obteniéndose existencia y unicidad de la ecuación en  $[t_0, T]$ , para  $T > t_0$  cualquiera.

Ahora se pasa a resolver la ecuación

Se tiene que, por teoría, aplicando la generalización de la fórmula de Itô, para  $a(x, t) = h(t)x$  y  $b(x, t) = \sigma x$  y suponiendo que existe inversa para cada  $t$  respecto de  $x$  de  $f(x, t)$ , llamada  $g(x, t)$ , se tiene por fórmula de Itô

$$d\xi(t) = df(X(t), t) = \bar{a}(\xi(t), t)dt + \bar{b}(\xi(t), t)dW(t)$$

donde

$$\bar{a}(x, t) = \frac{\partial f(g(x, t), t)}{\partial t} + a(g(x, t), t) \frac{\partial f(g(x, t), t)}{\partial x} + \frac{b^2(g(x, t), t)}{2} \frac{\partial^2 f(g(x, t), t)}{\partial x^2}$$

$$\bar{b}(x, t) = b(g(x, t), t) \frac{\partial f(g(x, t), t)}{\partial t}$$

y si existe  $f$  tal que se verifican las dos expresiones anteriores con  $\bar{a}(x, t) = \bar{a}(t)$ ,  $\bar{b}(x, t) = \bar{b}(t)$  independiente de  $x$ , puede tomarse  $g(x, t) = x$ , y reducirse a la nueva ecuación, que se resuelve fácilmente.

Se comprueba que se satisface la siguiente expresión, que es condición suficiente para obtener la reducibilidad de las ecuaciones

$$\frac{\partial}{\partial x} \left\{ b(x, t) \left[ \frac{\partial b(x, t)}{\partial t} - \frac{\partial}{\partial x} \left[ \frac{a(x, t)}{b(x, t)} \right] + \frac{1}{2} \frac{\partial^2 b(x, t)}{\partial x^2} \right] \right\} = 0 \quad (2.1)$$

ya que

$$\frac{\frac{\partial b(x,t)}{\partial t}}{b^2(x,t)} = 0 \quad - \frac{\partial}{\partial x} \left[ \frac{a(x,t)}{b(x,t)} \right] = - \frac{\partial}{\partial x} \left[ \frac{h(t)}{\sigma} \right] = 0 \quad \frac{1}{2} \frac{\partial^2 b(x,t)}{\partial x^2} = 0$$

y de aquí puede tomarse  $\bar{b}(t)$  tal que  $\bar{b}'(t) = 0$  siendo  $\bar{b}(t) = c$  constante a determinar, luego es

$$\frac{\partial f(x,t)}{\partial x} = \frac{\bar{b}(t)}{b(x,t)} = \frac{c}{\sigma x}$$

siendo

$$\frac{\partial^2 f(x,t)}{\partial x^2} = - \frac{c}{\sigma x^2}$$

y

$$f(x,t) = \frac{c}{\sigma} \log(x) + V(t)$$

y de aquí puede calcularse  $\bar{a}(x,t)$  mediante:

$$\begin{aligned} \bar{a}(x,t) &= \frac{\partial f(g(x,t),t)}{\partial t} + a(g(x,t),t) \frac{\partial f(g(x,t),t)}{\partial x} + \frac{b^2(g(x,t),t)}{2} \frac{\partial^2 f(g(x,t),t)}{\partial x^2} = \\ &= V'(t) + h(t)x \frac{c}{\sigma x} + \frac{\sigma^2 x^2}{2} \left( - \frac{c}{\sigma x^2} \right) = V'(t) + \frac{c}{\sigma} \left( h(t) - \frac{\sigma^2}{2} \right) \end{aligned}$$

donde por el razonamiento anterior puede tomarse  $g(x,t) = x$ , que es lo que se ha hecho, y de aquí se tiene que

$$\bar{a}(t) = V'(t) + \frac{c}{\sigma} \left( h(t) - \frac{\sigma^2}{2} \right)$$

$$\bar{b}(t) = c$$

siendo la ecuación diferencial estocástica de partida, transformada en

$$dZ(t) = \left( V'(t) + \frac{c}{\sigma} \left( h(t) - \frac{\sigma^2}{2} \right) \right) dt + c dW(t)$$

$$Z(t_0) = z_0$$

cuya solución se conoce aplicando Teorema 2.8.1:

$$Z(t) = z_0 + \int_{t_0}^t \left( V'(s) + \frac{c}{\sigma} \left( h(s) - \frac{\sigma^2}{2} \right) \right) ds + cW(t) = z_0 + V(t) - V(t_0) + \frac{c}{\sigma} \left( \int_{t_0}^t h(s) ds - \frac{\sigma^2}{2} (t - t_0) \right) + cW(t)$$

y basta aplicar el cambio de variable inverso para obtener la solución:

$$Z(t) = \frac{c}{\sigma} \log(X(t)) + V(t) \quad X(t) = \exp \left( \frac{\sigma}{c} (Z(t) - V(t)) \right)$$

de donde

$$X(t) = \exp \left\{ \frac{\sigma}{c} \left[ z_0 - V(t_0) + \frac{c}{\sigma} \left( \int_{t_0}^s h(s) ds - \frac{\sigma^2}{2} (t - t_0) \right) c(W(t) - W(t_0)) \right] \right\}$$

y como se impone  $X(t_0) = x_0 \sim \Lambda[\mu_0, \sigma_0^2]$ , se tiene:

$$X(t_0) = x_0 = \exp \left( \frac{\sigma}{c} (z_0 - V(t_0)) \right)$$

siendo  $V(t_0) = z_0 - \frac{c}{\sigma} \log(x_0)$ , luego

$$X(t) = x_0 \exp \left( \int_{t_0}^t h(s) ds - \frac{\sigma^2}{2} (t - t_0) + \sigma(W(t) - W(t_0)) \right)$$

Para obtener su densidad de transición y características nos basamos en el siguiente resultado, que es el fundamental de este capítulo y donde se usan muchos de los resultados y propiedades citadas en el apéndice A:

**Proposición 2.2.1.** Sea  $\{X(t) / t \geq t_0\}$  un proceso de difusión con momentos infinitesimales  $A_1(x, t) = h(t)x$  y  $A_2(x) = \sigma^2 x^2$ , donde  $\sigma > 0$ ,  $x \in \mathbb{R}^+$  y  $h$  una función continua acotada. Si además suponemos que la distribución inicial es constante, o sea  $X(t) = x_{t_0}$  casi seguramente, este proceso coincide con el proceso dado por

$$Y(t) = x_{t_0} \exp \left( \sigma W(t - t_0) + \int_{t_0}^t h(u) du - \frac{\sigma^2}{2} (t - t_0) \right), \quad t \geq t_0,$$

donde  $\{W(t) / t \geq 0\}$  es el proceso de Wiener estándar, o sea, con momentos infinitesimales  $A_1(x, t) = 0$  y  $A_2(x) = 1$ .

*Demostración.* Si llamamos

$$Z(t) = \sigma W(t - t_0) + \int_{t_0}^t h(u) du - \frac{\sigma^2}{2} (t - t_0),$$

se tiene que  $\{Z(t) / t \geq 0\}$  es un proceso de Wiener trasladado, claramente ya que lo es  $\{W(t - t_0) / t \geq 0\}$ , luego es un proceso de Markov por las propiedades de los procesos de Wiener. Y además como  $Y(t) = g(Z(t))$  con  $g(x) = x_{t_0} e^x$  función medible con inversa  $h(x) = \log \left( \frac{x}{x_{t_0}} \right)$ , por propiedades generales de los procesos de Markov se tiene que  $\{Y(t) / t \geq 0\}$  es un proceso de Markov.

Como  $Z(t) = \log(Y(t))$ , se tiene entonces que  $Z(t) \sim \mathcal{N} \left[ \log(x_{t_0}) + \int_{t_0}^t h(u) du - \frac{\sigma^2}{2} (t - t_0), \sigma^2 (t - t_0) \right]$ , y para un vector bidimensional se tiene que

$$\begin{pmatrix} Z(t) \\ Z(s) \end{pmatrix} \sim \mathcal{N}_2 \left[ \begin{pmatrix} \log(x_{t_0}) + \int_{t_0}^t h(u) du - \frac{\sigma^2}{2} (t - t_0) \\ \log(x_{t_0}) + \int_{t_0}^s h(u) du - \frac{\sigma^2}{2} (s - t_0) \end{pmatrix}, \sigma^2 \begin{pmatrix} s - t_0 & s - t_0 \\ s - t_0 & t - t_0 \end{pmatrix} \right],$$

y de aquí se tiene

$$\begin{pmatrix} Y(t) \\ Y(s) \end{pmatrix} \sim \Lambda_2 \left[ \begin{pmatrix} \log(x_{t_0}) + \int_{t_0}^t h(u)du - \frac{\sigma^2}{2}(t-t_0) \\ \log(x_{t_0}) + \int_{t_0}^s h(u)du - \frac{\sigma^2}{2}(s-t_0) \end{pmatrix}, \sigma^2 \begin{pmatrix} s-t_0 & s-t_0 \\ s-t_0 & t-t_0 \end{pmatrix} \right]$$

distribución lognormal bidimensional a partir de la cual puede calcularse la distribución de transición que viene dada por

$$(Y(t)|Y(s)=y) \sim \Lambda_1[\mu_{(2)} + \Sigma_{21}\Sigma_{11}^{-1}(\log(y) - \mu_{(1)}), \Sigma_{22.1}]$$

usando la notación multivariante. Calculando los parámetros de esta distribución se tiene que:

$$\mu_{(2)} + \Sigma_{21}\Sigma_{11}^{-1}(\log(y) - \mu_{(1)}) = \log(y) + \int_s^t h(u)du - \frac{\sigma^2}{2}(t-s)$$

$$\Sigma_{22.1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} = \sigma^2(t-s)$$

entonces se tiene la distribución de transición  $(Y(t)|Y(s)=y) \sim \Lambda_1 \left[ \log(y) + \int_s^t h(u)du - \frac{\sigma^2}{2}(t-s), \sigma^2(t-s) \right]$  con densidad dada por

$$f(x, t|y, s) = \frac{1}{\sqrt{2\pi(t-s)}\sigma_x} \exp \left\{ -\frac{1}{2\sigma^2(t-s)} \left( \log\left(\frac{x}{y}\right) - \int_s^t h(u)du + \frac{\sigma^2}{2}(t-s) \right)^2 \right\}$$

Para ver la existencia y unicidad se aplica el teorema (3.1.4):  $A_1(x, t) = h(t)x$ ,  $A_2(x) = \sigma^2 x^2$ , con  $h$  una función continua,  $\sigma > 0$ ,  $t \geq t_0 > 0$ , y donde o bien  $t \in [t_0, T]$  o bien  $t > t_0 > 0$  con  $h$  acotada.

Para  $A_2(x) = \sigma^2 x^2$  tenemos que

$$\varepsilon < \left( \frac{\partial A_2(x)}{\partial x} \right)^{1/2} = \sqrt{2\sigma x} < k\sqrt{1+x^2}$$

ya que como  $x > 0$  fijo, existe  $\varepsilon > 0$  tal que  $\varepsilon < \sigma x < k\sqrt{1+x^2}$  tomándose como  $\sigma_0 = \varepsilon$ , y  $k = \max\{\varepsilon, |m|\}$ , y

$$\left| \left( \frac{\partial A_2(x)}{\partial x} \right)^{1/2} - \left( \frac{\partial A_2(y)}{\partial x} \right)^{1/2} \right| = \sqrt{2}\sigma|\sqrt{x} - \sqrt{y}| \leq \sqrt{2}\sigma|x - y|$$

Como  $|\sqrt{x} - \sqrt{y}| \leq |x - y|^{1/2} \Leftrightarrow x + y - 2\sqrt{xy} \leq x - y \Leftrightarrow \sqrt{y}(\sqrt{y} - \sqrt{x}) \leq 0 \Rightarrow y \leq x$  basta elegir de forma que  $y \leq x$  verificandose la anterior igualdad buscada siendo  $f(x, t|y, s) = \frac{\partial F(x, t|y, s)}{\partial x}$  la única solución fundamental de la ecuación adelantada.

Además si  $k = \sup_{t \geq 0} |h(t)|$  (que existe tanto si  $h$  acotada como si es continua en  $[t_0, T]$ ) se tiene que:

$$|A_1(x, t)| = |h(t)x| \leq k|x| \leq k\sqrt{1+x^2}$$

y  $\frac{\partial A_1(x,t)}{\partial x} = |h(t)| \leq k \leq k\sqrt{1+x^2}$ , y de aquí es  $f(x,t|y,s) = \frac{\partial F(x,t|y,s)}{\partial x}$  la única solución fundamental de la ecuación adelantada.

Basta solamente ver que se verifica la ecuación de Fokker-Planck con la anterior densidad de transición para concluir. En este caso la ecuación viene dada por (sustituyendo  $A_i$  y simplificando):

$$\frac{\partial f(x,t|y,s)}{\partial t} = \frac{\sigma^2 x^2}{2} \frac{\partial^2 f(x,t|y,s)}{\partial x^2} + (2\sigma^2 - h(t))x \frac{\partial f(x,t|y,s)}{\partial x} + (\sigma^2 - h(t))f(x,t|y,s)$$

Para concluir ha de comprobarse que se verifica la ecuación con la densidad de transición antes obtenida, y por la unicidad de solución deben coincidir ambos procesos. Se comprueba la identidad anterior:

$$\begin{aligned} \frac{\partial f(x,t|y,s)}{\partial t} &= -\frac{1}{2(t-s)}f(x,t|y,s) + \left[ \frac{1}{2\sigma^2(t-s)^2} \left( \log\left(\frac{x}{y}\right) + \int_s^t h(u)du - \frac{\sigma^2}{2}(t-s) \right)^2 - \right. \\ &\quad \left. - \frac{1}{2\sigma^2(t-s)} \left( h(t) - \frac{\sigma^2}{2} \right) \left( \log\left(\frac{x}{y}\right) + \int_s^t h(u)du - \frac{\sigma^2}{2}(t-s) \right) \right] f(x,t|y,s) \\ \frac{\partial f(x,t|y,s)}{\partial x} &= -\frac{1}{x}f(x,t|y,s) - \frac{1}{\sigma^2(t-s)x} \left( \log\left(\frac{x}{y}\right) + \int_s^t h(u)du - \frac{\sigma^2}{2}(t-s) \right) f(x,t|y,s) \\ \frac{\partial^2 f(x,t|y,s)}{\partial x^2} &= \frac{1}{x^2} \left\{ 1 + \frac{1}{\sigma^2(t-s)} \left[ \log\left(\frac{x}{y}\right) + \int_s^t h(u)du - \frac{\sigma^2}{2}(t-s) - 1 \right] \right\} f(x,t|y,s) - \\ &\quad - \frac{1}{x} \left[ 1 + \frac{1}{\sigma^2(t-s)} \left( \log\left(\frac{x}{y}\right) + \int_s^t h(u)du - \frac{\sigma^2}{2}(t-s) \right) \right] \frac{\partial f(x,t|y,s)}{\partial x} \end{aligned}$$

Basta ahora volver a sustituir  $\frac{\partial f(x,t|y,s)}{\partial x}$  en la última igualdad, sustituir estas expresiones en la ecuación y comprobar que se verifica.

□

En el caso en que  $x_{t_0}$  tenga alguna distribución concreta (no necesariamente degenerada), también es válido este resultado.

## 2.2.2. Características del modelo

En vista de que las distribuciones finito-dimensionales son conocidas, y lognormales, se conocen las características del proceso. Pueden distinguirse los casos en que la distribución inicial  $X(t_0)$  sea degenerada, ó sea una  $\Lambda_1(\mu_0, \sigma_0^2)$ . El caso de que sea lognormal incluye al degenerado, ya que basta tomar  $\sigma_0 = 0$ , y se tiene que  $\mu_0 = \log(x_0)$ .

Si  $\sigma_0^2 > 0$  entonces es

$$\mu = \begin{pmatrix} \mu_0 \\ \log(x_0) + \int_{t_0}^t h(s)ds - \frac{\sigma^2}{2}(t-t_0) \end{pmatrix} \quad \Sigma = \sigma_0^2 \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} + \sigma^2 \begin{pmatrix} 0 & 0 \\ 0 & (t-t_0) \end{pmatrix}$$

Análogamente es  $(X(t), X(s)) \sim \Lambda_2[\mu, \Sigma]$  con

$$\mu = \begin{pmatrix} \log(x_0) + \int_{t_0}^t h(u)du - \frac{\sigma^2}{2}(t-t_0) \\ \log(x_0) + \int_{t_0}^s h(u)du - \frac{\sigma^2}{2}(s-t_0) \end{pmatrix}$$

Ahora se calculan algunas características principales aplicando la expresión conocida de éstas por ser una distribución lognormal. Se escribe conjuntamente casos en que la distribución inicial  $X(t_0)$  sea degenerada o lognormal:

**1. Función media:**

$$m(t) = E[X(t_0)] \exp \left\{ \int_{t_0}^t h(u)du \right\}$$

**2. Función media condicionada:**

$$m(t|s) = E[X(t)|X(s) = x_s] = x_s \exp \left\{ \int_s^t h(u)du \right\}$$

**3. Función moda:**

$$\text{Moda}(t) = E[X(t_0)] \exp \left\{ \int_{t_0}^t h(u)du - \frac{3\sigma^2}{2}(t-t_0) \right\}$$

**4. Función moda condicionada:**

$$\text{Moda}(t|s) = x_s \exp \left\{ \int_{t_0}^t h(u)du - \frac{3\sigma^2}{2}\sigma^2(t-s) \right\}$$

**5. Función cuartil de orden  $\alpha$  en  $X(t)$ :**

$$P_\alpha(t) = x_0 \exp \left\{ \int_{t_0}^t h(u)du - \frac{\sigma^2}{2} + z_\alpha \sigma \sqrt{t-t_0} \right\}$$

**6. Función cuartil de  $\alpha$  en  $X(t)$  condicionada a  $X(s)$ :**

$$P_\alpha(t|s) = E[X(s)] \exp \left\{ \int_s^t h(u)du - \frac{\sigma^2}{2} + z_\alpha \sigma \sqrt{t-s} \right\}$$

7. **Momentos de orden  $k > 1$ :**

$$\begin{aligned}
 E[X(t)^k] &= E[E[X(t)^k|X(t_0)]] = E\left[X(t_0)^k \exp\left\{k\left(\int_{t_0}^t h(u)du - \frac{\sigma^2}{2}(t-t_0) + \frac{k}{2}\sigma^2(t-t_0)\right)\right\}\right] = \\
 &= E[X(t_0)^k] \exp\left\{k\left(\int_{t_0}^t h(u)du - \frac{\sigma^2}{2}(t-t_0) + \frac{k}{2}\sigma^2(t-t_0)\right)\right\} \\
 &= E[X(t_0)^k] \exp\left\{k\int_{t_0}^t h(u)du + \frac{k\sigma^2}{2}(t-t_0)(k-1)\right\}
 \end{aligned}$$

8. **Función de covarianzas:** en general es  $Cov(X(t), X(s)) = E[X(t)X(s)] - m(t)m(s)$ , que suponiendo  $s < t$  como es

$$\begin{aligned}
 E[X(t)X(s)] &= E[E[X(t)X(s)|X(s)]] = E[X(s)E[X(t)|X(s)]] = E[X(s)^2] \exp\left\{\int_s^t h(u)du\right\} \\
 &= E[X(t_0)^2] \exp\left\{\int_{t_0}^s h(u)du + \int_{t_0}^t h(u)du + \sigma^2(s-t_0)\right\}
 \end{aligned}$$

y de aquí la covarianza es

$$\begin{aligned}
 Cov(X(t), X(s)) &= E[X(t)X(s)] - m(t)m(s) = \\
 &= E[X(t_0)^2] \exp\left\{\int_{t_0}^s h(u)du + \int_{t_0}^t h(u)du + \sigma^2(s-t_0)\right\} - \\
 &\quad - E[X(t_0)]^2 \exp\left\{\int_{t_0}^t h(u)du + \int_{t_0}^s h(u)du\right\} = \\
 &= \exp\left\{\int_{t_0}^t h(u)du + \int_{t_0}^s h(u)du\right\} \left(E[X(t_0)^2]e^{\sigma^2(s-t_0)} - E[X(t_0)]^2\right)
 \end{aligned}$$

En general, sin suponer  $s < t$  la siguiente expresión es válida

$$Cov(X(s), X(t)) = \exp\left\{\int_{t_0}^t h(u)du + \int_{t_0}^s h(u)du\right\} \left(E[X(t_0)^2]e^{\sigma^2(t \wedge s - t_0)} - E[X(t_0)]^2\right)$$

y en particular, la varianza tiene la forma:

$$Var(X(t)) = \exp\left\{2\int_{t_0}^t h(u)du\right\} \left(E[X(t_0)^2]e^{\sigma^2(t-t_0)} - E[X(t_0)]^2\right)$$

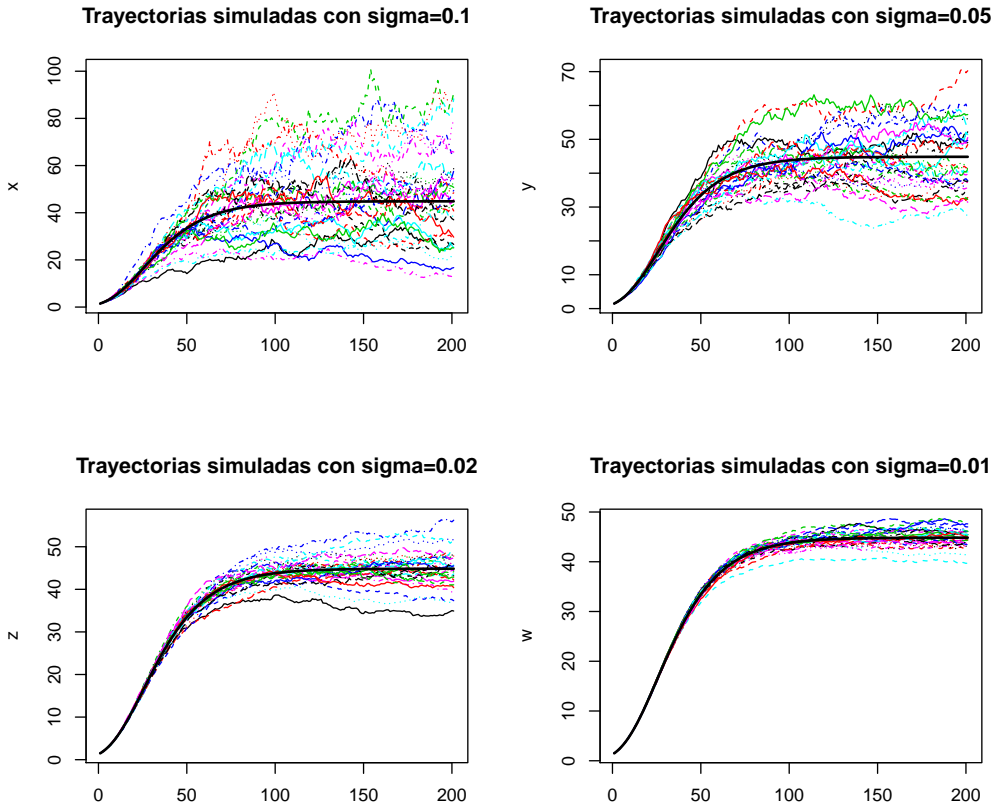


### 2.2.3. Adaptación del proceso lornormal a los procesos Gompertz y Logístico

#### 2.2.3.1. Proceso Gompertz

Se observa que  $\frac{df(t)}{dt} = me^{-\beta t}$ , de donde la media infinitesimal de este proceso es de la forma  $A_1(x, t) = h(t)x$  con  $h(t) = me^{-\beta t}$ , caso particular del proceso lognormal con factores exógenos, el cual ya ha sido estudiado, y basta con sustituir  $h(t)$  obteniéndose la densidad de transición y las características principales.

Se representan, junto a su función media, algunas trayectorias simuladas de este proceso. Los valores de los parámetros son  $m = 1.7$ ,  $\beta = 0.5$  y la varianza cambia en cada gráfica, apreciándose que la dispersión que se produce con el tiempo haga pensar que estos modelos no son recomendables para valores altos de  $\sigma$ :



#### 1. Densidad de transición y características

Este modelo viene dado por un proceso que es caso particular del estudiado anteriormente, y que tiene por momentos infinitesimales:

$$A_1(x, t) = me^{-\beta t}x$$

$$A_2(x) = \sigma^2 x^2$$

siendo en este caso  $h(t) = me^{-\beta t}$ . A partir de lo anterior, teniendo en cuenta cómo son los momentos de una distribución lognormal, puede presentarse la densidad de transición así como las características de

este proceso (media, varianza, moda). Así, este proceso coincide con

$$X(t) = x_{t_0} \exp \left\{ \sigma W(t - t_0) + \frac{m}{\beta} (e^{-\beta t_0} - e^{-\beta t}) - \frac{\sigma^2}{2} (t - t_0) \right\}$$

para  $t \geq t_0$  con  $W(t)$  proceso de Wiener standard.

La densidad de transición es:

$$f(x, t | y, s) = \frac{1}{\sqrt{2\pi(t-s)}\sigma x} \exp \left\{ -\frac{1}{2\sigma^2(t-s)} \left( \log \left( \frac{x}{y} \right) - \frac{m}{\beta} (e^{-\beta s} - e^{-\beta t}) + \frac{\sigma^2}{2} (t-s) \right)^2 \right\}$$

luego se tiene que

$$X(t) | X(s) = y \sim \Lambda_1 \left[ \log(y) + \frac{m}{\beta} (e^{-\beta s} - e^{-\beta t}) - \frac{\sigma^2}{2} (t-s), \sigma^2(t-s) \right]$$

## 2. Características del modelo:

### a) Distribución inicial degenerada:

	Característica
<b>Media</b>	$m(t) = x_0 \exp \left\{ \frac{m}{\beta} (e^{-\beta t_0} - e^{-\beta t}) \right\}$
<b>Moda</b>	$\text{Moda}(t) = x_0 \exp \left\{ \frac{m}{\beta} (e^{-\beta t_0} - e^{-\beta t}) - \frac{3\sigma^2}{2} (t - t_0) \right\}$
<b>Cuantil de orden <math>\alpha</math></b>	$P_\alpha(t) = x_0 \exp \left\{ \frac{m}{\beta} (e^{-\beta t_0} - e^{-\beta t}) - \frac{\sigma^2}{2} (t - t_0) + z_\alpha \sigma \sqrt{t - t_0} \right\}$
<b>Momentos de orden <math>k &gt; 1</math></b>	$E[X(t)^k] = x_0^k \exp \left\{ k \left( \frac{m}{\beta} (e^{-\beta t_0} - e^{-\beta t}) + \frac{k\sigma^2}{2} (t - t_0)(k-1) \right) \right\}$
<b>Varianza</b>	$\text{Var}(X(t)) = x_0^2 \exp \left\{ \frac{2m}{\beta} (e^{-\beta t_0} - e^{-\beta t}) \right\} (e^{\sigma^2(t-t_0)} - 1)$

b) **Distribución inicial**  $\Lambda_1(\mu_0, \sigma_0^2)$ : se tiene en cuenta que  $E[X(t_0)^k] = e^{k\mu_0 + \frac{k^2}{2}\sigma_0^2}$

	<b>Característica</b>
<b>Media</b>	$m(t) = \exp \left\{ \mu_0 + \frac{\sigma_0^2}{2} + \frac{m}{\beta} (e^{-\beta t_0} - e^{-\beta t}) \right\}$
<b>Moda</b>	$\text{Moda}(t) = \exp \left\{ \mu_0 + \frac{\sigma_0^2}{2} + \frac{m}{\beta} (e^{-\beta t_0} - e^{-\beta t}) - \frac{3\sigma^2}{2} (t - t_0) \right\}$
<b>Cuantil de orden <math>\alpha</math></b>	$P_\alpha(t) = \exp \left\{ \mu_0 + \frac{\sigma_0^2}{2} + \frac{m}{\beta} (e^{-\beta t_0} - e^{-\beta t}) - \frac{\sigma^2}{2} (t - t_0) + z_\alpha \sigma \sqrt{t - t_0} \right\}$
<b>Momentos de orden <math>k &gt; 1</math></b>	$E[X(t)^k] = \exp \left\{ k\mu_0 + \frac{k^2}{2}\sigma_0^2 + k\frac{m}{\beta} (e^{-\beta t_0} - e^{-\beta t}) + \frac{k\sigma^2}{2} (t - t_0)(k - 1) \right\}$
<b>Varianza</b>	$\text{Var}(X(t)) = e^{2(\mu_0 + \sigma_0^2)} \exp \left\{ \frac{2m}{\beta} (e^{-\beta t_0} - e^{-\beta t}) \right\} (e^{\sigma_0^2 + \sigma^2(t - t_0)} - 1)$

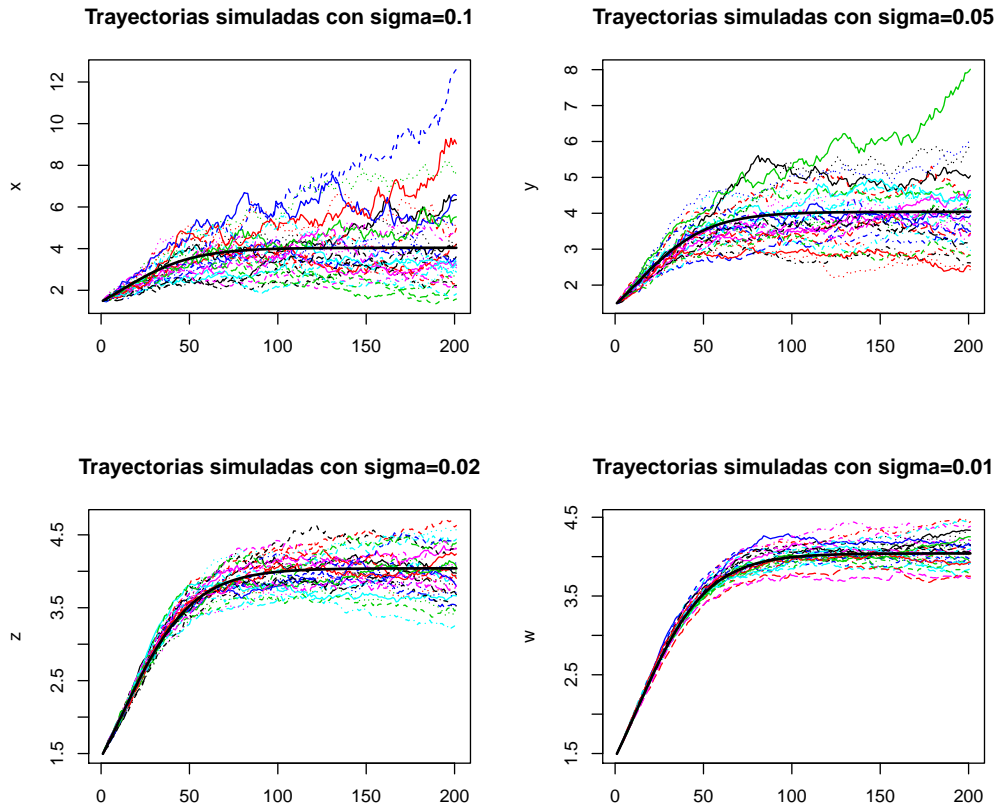
c) **Característica condicionada:**

	<b>Característica condicionada</b>
<b>Media</b>	$m(t s) = x_s \exp \left\{ \frac{m}{\beta} (e^{-\beta s} - e^{-\beta t}) \right\}$
<b>Moda</b>	$\text{Moda}(t s) = x_s \exp \left\{ \frac{m}{\beta} (e^{-\beta s} - e^{-\beta t}) - \frac{3\sigma^2}{2} (t - s) \right\}$
<b>Cuantil de orden <math>\alpha</math></b>	$P_\alpha(t s) = E[X(s)] \exp \left\{ \frac{m}{\beta} (e^{-\beta s} - e^{-\beta t}) - \frac{\sigma^2}{2} (t - t_0) + z_\alpha \sigma \sqrt{t - s} \right\}$

### 2.2.3.2. Proceso Logístico

Se observa que  $\frac{df(t)}{dt} = \frac{bce^{-ct}}{1 + be^{-ct}}f(t) = \frac{bc}{e^{ct} + b}f(t)$  de donde el momento infinitesimal de este proceso es de la forma  $A_1(x, t) = h(t)x$  con  $h(t) = \frac{bc}{e^{ct} + b}$ , caso particular del proceso lognormal con factores exógenos, el cual ya ha sido estudiado, y basta con sustituir  $h(t)$  obteniéndose la densidad de transición y las características principales.

Se representan, junto a su función media, algunas trayectorias simuladas de este proceso. Los valores de los parámetros son  $b = 1.7$ ,  $c = 0.5$  y la varianza cambia en cada gráfica, apreciándose que la dispersión que se produce con el tiempo haga pensar que estos modelos no son recomendables para valores altos de  $\sigma$ :



#### 1. Densidad de transición y características

Este modelo viene dado por un proceso que es caso particular del estudiado anteriormente, y que tiene por momentos infinitesimales:

$$A_1(x, t) = \frac{bc}{e^{ct} + b}x$$

$$A_2(x) = \sigma^2 x^2$$

siendo en este caso  $h(t) = \frac{bc}{e^{ct} + b}$ . A partir de lo anterior, teniendo en cuenta cómo son los momentos de una distribución lognormal, puede presentarse la densidad de transición así como las características de

este proceso (media, varianza, moda). Así, este proceso coincide con

$$X(t) = x_{t_0} \exp \left\{ \sigma W(t - t_0) + \log \left( \frac{1 + be^{-ct_0}}{1 + be^{-ct}} \right) - \frac{\sigma^2}{2} (t - t_0) \right\}$$

para  $t \geq t_0$  con  $W(t)$  proceso de Wiener standard.

La densidad de transición es:

$$f(x, t|y, s) = \frac{1}{\sqrt{2\pi(t-s)}\sigma_x} \exp \left\{ -\frac{1}{2\sigma^2(t-s)} \left( \log \left( \frac{x}{y} \right) - \log \left( \frac{1 + be^{-cs}}{1 + be^{-ct}} \right) + \frac{\sigma^2}{2} (t-s) \right)^2 \right\}$$

luego se tiene que

$$X(t)|X(s) = y \sim \Lambda_1 \left[ \log(y) + \log \left( \frac{1 + be^{-cs}}{1 + be^{-ct}} \right) + \frac{\sigma^2}{2} (t-s) - \frac{\sigma^2}{2} (t-s), \sigma^2(t-s) \right]$$

## 2. Características del modelo:

### a) Distribución inicial degenerada:

	<b>Característica</b>
<b>Media</b>	$m(t) = x_0 \frac{1 + be^{-ct_0}}{1 + be^{-ct}}$
<b>Moda</b>	$\text{Moda}(t) = x_0 \frac{1 + be^{-ct_0}}{1 + be^{-ct}} \exp \left\{ -\frac{3\sigma^2}{2} (t - t_0) \right\}$
<b>Cuantil de orden <math>\alpha</math></b>	$P_\alpha(t) = x_0 \frac{1 + be^{-ct_0}}{1 + be^{-ct}} \exp \left\{ -\frac{\sigma^2}{2} (t - t_0) + z_\alpha \sigma \sqrt{t - t_0} \right\}$
<b>Momentos de orden <math>k &gt; 1</math></b>	$E[X(t)^k] = x_0^k \left( \frac{1 + be^{-ct_0}}{1 + be^{-ct}} \right)^k e^{\frac{k\sigma^2}{2} (t-t_0)(k-1)}$
<b>Varianza</b>	$\text{Var}(X(t)) = x_0^2 \log \left( \frac{(1 + be^{-ct_0})^2}{(1 + be^{-ct})(1 + be^{-cs})} \right) (e^{\sigma^2(t-t_0)} - 1)$

b) **Distribución inicial**  $\Lambda_1(\mu_0, \sigma_0^2)$ : se tiene en cuenta que  $E[X(t_0)^k] = e^{k\mu_0 + \frac{k^2}{2}\sigma_0^2}$

	<b>Característica</b>
<b>Media</b>	$m(t) = e^{\mu_0 + \frac{\sigma_0^2}{2}} \frac{1 + be^{-ct_0}}{1 + be^{-ct}}$
<b>Moda</b>	$\text{Moda}(t) = e^{\mu_0 + \frac{\sigma_0^2}{2}} \frac{1 + be^{-ct_0}}{1 + be^{-ct}} \exp\left\{-\frac{3\sigma^2}{2}(t - t_0)\right\}$
<b>Cuantil de orden <math>\alpha</math></b>	$P_\alpha(t) = e^{\mu_0 + \frac{\sigma_0^2}{2}} \frac{1 + be^{-ct_0}}{1 + be^{-ct}} \exp\left\{-\frac{\sigma^2}{2}(t - t_0) + z_\alpha \sigma \sqrt{t - t_0}\right\}$
<b>Momentos de orden <math>k &gt; 1</math></b>	$E[X(t)^k] = \left(\frac{1 + be^{-ct_0}}{1 + be^{-ct}}\right)^k \exp\left\{k\mu_0 + \frac{k^2}{2}\sigma_0^2 + \frac{k\sigma^2}{2}(t - t_0)(k - 1)\right\}$
<b>Varianza</b>	$\text{Var}(X(t)) = e^{2\mu_0 + \sigma_0^2} \frac{(1 + be^{-ct_0})^2}{(1 + be^{-ct})(1 + be^{-cs})} (e^{\sigma_0^2 + \sigma^2(t - t_0)} - 1)$

c) **Característica condicionada:**

	<b>Característica condicionada</b>
<b>Media</b>	$m(t s) = x_s \frac{1 + be^{-cs}}{1 + be^{-ct}}$
<b>Moda</b>	$\text{Moda}(t s) = x_s \frac{1 + be^{-cs}}{1 + be^{-ct}} \exp\left\{-\frac{3\sigma^2}{2}(t - s)\right\}$
<b>Cuantil de orden <math>\alpha</math></b>	$P_\alpha(t s) = E[X(s)] \frac{1 + be^{-cs}}{1 + be^{-ct}} \exp\left\{-\frac{\sigma^2}{2}(t - s) + z_\alpha \sigma \sqrt{t - s}\right\}$

## Capítulo 3

# ALGORITMOS ESTOCÁSTICOS DE OPTIMIZACIÓN

### 3.1. CONCEPTOS PREVIOS

#### 3.1.1. Sobre algoritmos

A lo largo de las dos secciones siguientes se hará uso de dos conceptos que se definen a continuación:

1. **Heurística:** En computación, dos objetivos fundamentales son encontrar algoritmos con buenos tiempos de ejecución y buenas soluciones, usualmente las óptimas. Una heurística es un algoritmo que abandona uno o ambos objetivos; por ejemplo, normalmente encuentran buenas soluciones, aunque no hay pruebas de que la solución no pueda ser arbitrariamente errónea en algunos casos; o se ejecuta razonablemente rápido, aunque no existe tampoco prueba de que siempre será así. Las heurísticas generalmente son usadas cuando no existe una solución óptima bajo las restricciones dadas (tiempo, espacio, etc.), o cuando no existe del todo.
2. **Metaheurística:** método heurístico para resolver un tipo de problema computacional general, usando los parámetros dados por el usuario sobre unos procedimientos genéricos y abstractos de una manera que se espera eficiente. Normalmente, estos procedimientos son heurísticos. El nombre combina el prefijo griego 'meta' ('más allá', aquí con el sentido de 'nivel superior') y 'heurístico' (de , heuriskein, 'encontrar').

Las metaheurísticas generalmente se aplican a problemas que no tienen un algoritmo o heurística específica que dé una solución satisfactoria; o bien cuando no es posible implementar ese método óptimo. La mayoría de las metaheurísticas tienen como objetivo los problemas de optimización combinatoria, pero por supuesto, se pueden aplicar a cualquier problema que se pueda reformular en términos heurísticos.

El tipo de problemas a los que se aplicarán los algoritmos que se estudian, encajan dentro de los conceptos anteriores, ya que de forma general no puede saberse si dada una función arbitraria, partiendo de una solución inicial para buscar un extremo absoluto, no se caiga en un extremo relativo, obteniendo una solución incorrecta.

### 3.1.2. Sobre problemas de optimización

Sea  $\Omega$  el *espacio de soluciones*,  $f : \Omega \rightarrow \mathbb{R}$  la *función objetivo* definida sobre el espacio de soluciones. El objetivo es encontrar  $\omega^*$  mínimo local, esto es

$$\omega^* \in \Omega \text{ tal que } f(\omega^*) \leq f(\omega) \forall \omega \in \Omega$$

Se supone que la función  $f$  es acotada, para que exista dicho mínimo. Se define  $N(\omega)$  como un entorno de la función para  $\omega \in \Omega$ , habiendo asociados para cada solución entornos  $N(\omega)$ , que pueden ser alcanzados en una sola iteración.

## 3.2. ALGORITMO SIMULATED ANNEALING (S.A.)

Es una metaheurística, y se trata de un algoritmo de búsqueda local, que parte de la idea de poder escapar de óptimos locales. Sus propiedades de convergencia, y de ‘escape’ de extremos locales ha hecho que se popularice este método. Es usado normalmente en problemas de optimización (tanto discretos como continuos). Fue introducida en la década de los 80 como una nueva herramienta para en la solución de grandes problemas de combinatoria, y se trata de una técnica de optimización combinatorial usada para afrontar problemas de gran complejidad matemática, de modo que se obtengan soluciones cercanas a la próxima.

La idea original que dio lugar al S.A. es llamada algoritmo de Metrópolis, basado a su vez en el método de Montecarlo, con el cual se estudian propiedades de equilibrio en el análisis del comportamiento microscópico de los cuerpos.

Recibe este nombre por la analogía del proceso físico de recocido con sólidos, en el cual un sólido cristalino es calentado y luego enfriado muy lentamente hasta que se obtenga una configuración de red lo más regular posible, obteniendo cristales de gran pureza. Si el enfriamiento es suficientemente lento, la configuración final resulta un sólido con una integridad estructural superior. Simulated Annealing establece la conexión entre este tipo de comportamiento termodinámico y la búsqueda de mínimo global en problemas de optimización.

Cada iteración del algoritmo es aplicada a un problema de optimización, y se van comparando las soluciones obtenidas en ellos. Las mejores soluciones son las que siempre se van aceptando, mientras que una parte de las que no son mejores son aceptadas esperando escapar de un óptimo local. La probabilidad de aceptar una solución no buena depende del parámetro temperatura, que normalmente es no decreciente en cada iteración del algoritmo.

Lo novedoso respecto otros algoritmos de búsqueda local descendente es que proporciona medios para escapar de óptimos locales mediante ‘movimientos ascendentes’. Cuando la temperatura decrece a cero estos movimientos ascendentes ocurren con menor frecuencia, y la distribución asociada a la solución con la cadena de Markov no homogénea que modela el comportamiento del algoritmo converge a una forma la cual está concentrada en el conjunto de soluciones globales óptimas.

La clave del algoritmo es la determinación de la temperatura inicial  $T_0$ , velocidad de enfriamiento, longitud  $N_k$  de la cadena. Estos parámetros han de adaptarse al tipo de problema que se esté tratando.

1. **Temperatura:** debe ser adecuada para una exploración completa en las etapas del S.A.. Existen distintas



formas para calcular la temperatura inicial, siendo una de ellas simular el proceso para la 1ª cadena de Markov.

Ha de definirse también una constante de enfriamiento  $0 < B < 1$ , donde normalmente se toma  $0.8 \leq B \leq 0.99$ .

La temperatura final de enfriamiento es aquella en la que ya es muy poco probable que se acepten soluciones que empeoren la actual.

2. **Longitud de la cadena:**  $N_k$  define el número de tentativas para el nivel de temperatura  $T_k$ . Como regla general puede tomarse una longitud creciente a medida que van sucediendo las etapas con el fin de obtener una mayor exploración.

La longitud de la cadena puede ser escogida de forma que sea proporcional al tamaño del problema en algunos casos (en otros, cuando el número de iteraciones es algo grande no es conveniente ya que el tiempo de ejecución puede incrementarse altamente, ya que las últimas etapas alcanzarían un número muy grande de iteraciones).

3. **Criterio de parada:** Puede ser uno de los siguientes:

- a) *Número determinado de niveles de temperatura (usualmente entre 6 y 50).*
- b) *Si la función objetivo no mejora para varios niveles consecutivos.*
- c) *Si no se cumple con un mínimo de aceptaciones en el nivel de temperatura.*

4. **Generación de una nueva configuración:** ha de definirse una estructura de entornos que contengan a las soluciones factibles, formada por aquellas configuraciones que puedan generarse mediante pequeñas modificaciones (sacar elemento activo, agregar uno, ó intercambiar un elemento que está presente en la configuración por otro que no lo está).

La elección de la nueva configuración se hace aleatoriamente, pero se puede incrementar la probabilidad de selección para aquellas configuraciones que tienen un índice de sensibilidad que las identifica como atractivas para conseguir una mejora en la función objetivo.

### 3.2.1. Presentación del algoritmo

Sea  $\Omega$  el *espacio de soluciones*,  $f : \Omega \rightarrow \mathbb{R}$  la *función objetivo* definida sobre el espacio de soluciones. El objetivo es encontrar  $\omega^*$  mínimo local, esto es

$$\omega^* \in \Omega \text{ tal que } f(\omega^*) \leq f(\omega) \forall \omega \in \Omega$$

Se supone que la función  $f$  es acotada, para que exista dicho mínimo. Se define  $N(\omega)$  como un entorno de la función para  $\omega \in \Omega$ , habiendo asociados para cada solución entornos  $N(\omega)$ , que pueden ser alcanzados en una sola iteración.

El algoritmo parte de una solución inicial  $\omega \in \Omega$ , y se genera una solución  $\omega' \in (\omega)$  (generada aleatoriamente o mediante alguna regla). Simulated Annealing está basado en el *criterio de aceptación de Metropolis*,

el cual se modeliza como un sistema termodinámico que mueve la actual solución (estado)  $\omega \in \Omega$  a una solución candidata  $\omega' \in N(\omega)$ , en la cual la energía está siendo minimizada.  $\omega'$  es aceptada como actual solución basándose en la siguiente probabilidad de aceptación

$$P\{\text{Aceptar } \omega' \text{ como nueva solución}\} = \begin{cases} \exp\left[-\frac{f(\omega') - f(\omega)}{t_k}\right] & \text{si } f(\omega') - f(\omega) > 0 \\ 1 & \text{si } f(\omega') - f(\omega) \leq 0 \end{cases}$$

Se define  $t_k$  como el parámetro temperatura en la iteración  $k$ , tal que

$$t_k > 0 \quad \forall k \quad ; \quad \lim_{k \rightarrow \infty} t_k = 0.$$

Esta probabilidad de aceptación es el elemento básico en el mecanismo de búsqueda del simulated annealing. Si la temperatura es reducida de forma suficientemente lenta, entonces el sistema puede alcanzar equilibrio en cada iteración  $k$ . Sean  $f(\omega)$  y  $f(\omega')$  funciones objetivo asociadas con las soluciones  $\omega \in \Omega$ ,  $\omega' \in N(\omega)$  respectivamente (se llaman energías). Este equilibrio sigue la distribución de Boltzmann, que puede ser descrita como la probabilidad del sistema estando en el estado  $\omega \in \Omega$  con en energía  $f(\omega)$  a temperatura  $T$  tal que

$$P\{\text{sistema en el estado } \omega \text{ con temperatura } T\} = \frac{\exp\left(\frac{-f(\omega)}{t_k}\right)}{\sum_{\omega'' \in \Omega} \exp(-f(\omega'')/t_k)}.$$

Si la probabilidad de generar una solución candidata  $\omega'$  de los entornos de soluciones  $\omega \in \Omega$  es  $g_k(\omega, \omega')$ , donde

$$\sum_{\omega' \in \mathcal{N}(\omega)} g_k(\omega, \omega') = 1 \quad \forall \omega \in \Omega, \quad k = 1, 2, \dots$$

entonces una matriz cuadrada estocástica no negativa  $\mathbf{P}_k$  puede ser definida con probabilidades de transición

$$P_k(\omega, \omega') = \begin{cases} g_k(\omega, \omega') \exp(-\Delta_{\omega, \omega'} / t_k) & \omega' \in \mathcal{N}(\omega), \quad \omega' \neq \omega \\ 0 & \omega' \in \mathcal{N}(\omega), \quad \omega' \neq \omega \\ 1 - \sum_{\omega'' \in \mathcal{N}(\omega), \omega'' \neq \omega} \mathbf{P}_k(\omega, \omega'') & \omega' = \omega \end{cases}$$

para toda solución  $\omega \in \Omega$ , y todas las iteraciones  $k = 1, 2, \dots$  y con  $\Delta_{\omega, \omega'} \equiv f(\omega') - f(\omega)$ . Estas probabilidades de transición definen una sucesión de soluciones generadas desde una cadena de Markov no homogénea.

El algoritmo S.A. entra dentro de una clase más amplia de algoritmos llamados *Threshold Algorithms* (de uso difundido debido a que parecen funcionar bien con una amplia gama de problemas reales, y a que algunos como el S.A. tienen características que permiten hacer un análisis de la convergencia), para implementar un T.A. es necesario:

1. función entorno  $\mathcal{N}$  de  $\Omega$  en partes de  $\Omega$ .
2. una sucesión  $t_k$  (los llamados *thresholds*).

En general se parte de  $x \in \Omega$ , genero  $x' \in \mathcal{N}(x)$ , y mediante el valor que toma  $\Delta = f(x') - f(x)$  puedo elegir entre 3 criterios que son los que definen los T.A.:

1. **Local search improvement:**  $t_k = 0 \forall k$ , y si  $\Delta = 0$  entonces acepto  $x'$ .
2. **Threshold accepting:** fijo  $t_k$  sucesión tal que sea decreciente y tienda a cero cuando  $k \rightarrow \infty$ , y entonces si  $\Delta < t_k$  acepto  $x'$ .
3. **Simulated annealing:** considero los  $t_k$  como en el threshold accepting pero el criterio de aceptación es probabilístico:

$$\Delta \leq 0 \implies \text{acepto } x'$$

$$\Delta > 0 \implies \text{acepto } x' \text{ con probabilidad } \exp\left\{\frac{f(x) - f(x')}{t_k}\right\}$$

En este caso, cada entorno de una solución tiene una probabilidad positiva de reemplazar a la solución actual. Los  $t_k$  se eligen de forma tal que a medida que avanzan las iteraciones, aceptar soluciones con grandes incrementos en el costo es menos probable (pero sigue habiendo una probabilidad positiva de aceptarlos).

Algunas ventajas de este tipo de algoritmos es poder escapar de mínimos locales, además de ser aplicable a gran variedad de problemas de distinta naturaleza como alternativa a otros métodos que requieren que la función objetivo verifique ciertas propiedades, a veces muy restrictivas.

Siempre se ha de tener en cuenta en la implementación de S.A. los siguientes puntos:

1. Considerar una representación de posibles soluciones (puede que el conjunto de soluciones factibles sea no acotado, aunque puede conocerse a priori un conjunto acotado donde buscar el óptimo global).
2. Generador de soluciones aleatorias.
3. Poder evaluar la función objetivo para comparar soluciones.
4. Temperatura inicial y regla de enfriamiento.

Se presenta el algoritmo general:

---

**Algoritmo Simulated Annealing**


---

- 1 Seleccionar solución inicial  $\omega \in \Omega$
  - 2 Iniciar  $k = 0$  contador para el cambio de temperatura
  - 3 Elegir regla de enfriamiento (geométrica,...)
  - 4 Elegir temperatura inicial  $T_0 \geq 0$
  - 5 Elegir longitud de la cadena  $M_k$  para cada temperatura
  - 6 **repetir**

---

  - 7  $m = 0$
  - 8 **repetir**

---

  - 9 generar solución  $\omega' \in \mathcal{N}(\omega)$
  - 10 calcular  $\Delta_{\omega, \omega'} = f(\omega') - f(\omega)$
  - 11 si  $\Delta_{\omega, \omega'} \leq 0$  entonces  $\omega \leftarrow \omega'$
  - 12 si  $\Delta_{\omega, \omega'} > 0$  entonces  $\omega \leftarrow \omega'$  con probabilidad  $\exp(-\Delta_{\omega, \omega'}/t_k)$
  - 13  $m \leftarrow m + 1$

---

  - hasta que**  $m = M_k$

---

  - 14  $k \leftarrow k + 1$
- Salida**  $\omega$  cuando se cumpla condición de parada
- 

La formulación dada resulta en  $M_0 + M_1 + \dots + M_k$  iteraciones en total a ser ejecutadas, donde  $k$  corresponde al valor  $t_k$  por el cual algún criterio de parada es introducido (por ejemplo, una temperatura mínima alcanzada). Además, para  $M_k = 1 \forall k$ , la temperatura irá cambiando en cada iteración.

### 3.2.2. Problemas discretos frente a continuos

La mayoría de las aplicaciones del S.A. han sido para problemas de tipo discreto, aunque también tiene aplicación en problemas con dominio continuo (búsqueda de extremos globales en funciones, en aplicaciones estadísticas).

En optimización de funciones continuas ha de partir de una aproximación inicial de la solución y un tamaño de paso para tomar en esta dirección y evaluar un nuevo candidato. Si el nuevo candidato mejora al anterior, entonces me quedaré con éste. Esta migración a través de mínimos locales en la búsqueda de un mínimo global continúa hasta que éste es encontrado, ó algún criterio de parada es alcanzado.

Se sabe que para una clase de algoritmos de tipo S.A. para optimización global continua presenta convergencia el método. El S.A. es construido usando una cadena de Markov en el algoritmo para generar puntos uniformemente distribuidos en una región acotada arbitraria, viéndose la convergencia a un óptimo global.

### 3.2.3. Resultados de convergencia

Los resultados de convergencia asintótica para el S.A. han seguido tradicionalmente dos direcciones: el algoritmo ha sido modelizado como una sucesión de cadenas de Markov homogéneas, o como una única cadena

de Markov no homogénea.

### 3.2.3.1. Aproximación como una sucesión de cadenas de Markov homogéneas:

Se asume que cada temperatura  $t_k$  es constante para un número suficiente de iteraciones  $m$  tal que la matriz  $\mathbf{P}_k$  puede alcanzar distribución estacionaria  $\pi_k$ . La existencia de distribución estacionaria en cada iteración  $k$  se sigue del siguiente resultado:

**Teorema 3.2.1.** *Sea  $\mathbf{P}_k(\omega, \omega')$  la probabilidad de moverse de  $\omega$  a  $\omega'$  en una iteración de la  $k$ -ésima etapa, y sea  $\mathbf{P}_k^{(m)}(\omega, \omega')$  la probabilidad de ir de  $\omega$  a  $\omega'$  en  $m$  iteraciones. Si la cadena de Markov asociada con  $\mathbf{P}_k^{(m)}(\omega, \omega')$  es irreducible y aperiódica, entonces:*

$$\lim_{m \rightarrow \infty} \mathbf{P}_k^{(m)}(\omega, \omega') = \pi_k(\omega')$$

*existe  $\forall \omega, \omega' \in \Omega$  e iteraciones  $k$ . Además  $\pi_k(\omega')$  es la única solución estrictamente positiva del sistema*

$$\begin{cases} \pi_k(\omega') = \sum_{\omega \in \Omega} \pi_k(\omega) \mathbf{P}_k(\omega, \omega') & \forall \omega' \in \Omega \\ \sum_{\omega \in \Omega} \pi_k(\omega) = 1 \end{cases}$$

Los requisitos para la existencia de distribución estacionaria y para la convergencia de la sucesión de vectores  $\pi_k$  incluye las siguientes:

1. Irreducibilidad de la matriz de transición.
2. Aperiodicidad.
3. Distribución estacionaria no nula (prob. positiva de alcanzar cualquier estado desde cualquier otro).

Todas las pruebas de convergencia conocidas están basadas en teoría de cadenas de Markov homogéneas, en las cuáles se usa (explícita ó implícitamente) la condición suficiente de reversibilidad dada por la igualdad

$$\pi_k(\omega) \mathbf{P}_k(\omega, \omega') = \pi_k(\omega') \mathbf{P}_k(\omega', \omega)$$

$\forall \omega, \omega' \in \Omega$ , y toda iteración  $k$ .

Esta es la única condición requerida para garantizar la existencia y unicidad de solución en las ecuaciones de equilibrio del sistema dadas en el Teorema anterior. Una condición necesaria para la reversibilidad es que se cumpla la igualdad

$$\kappa_k(\Delta_{\omega, \omega''}) = \kappa_k(\Delta_{\omega, \omega'}) \kappa_k(\Delta_{\omega', \omega''})$$

para todo  $\omega, \omega', \omega'' \in \Omega$  tal que  $f(\omega) \leq f(\omega') \leq f(\omega'')$  donde  $\kappa_k(\Delta_{\omega, \omega''})$  es la probabilidad de aceptación la transición de  $\omega$  a  $\omega''$  en la  $k$ -ésima etapa.

Las demostraciones de convergencia de las cadenas de Markov conocidas requieren el uso de la anterior propiedad para mantener la solución y entonces abordar condiciones suficientes para la matriz  $\mathbf{P}_k$ .

S.A. y la teoría de convergencia homogénea están basados en el trabajo de *Metropolis et Al*, el cual aborda problemas de equilibrio en Mecánica Estadística. Para ver esta relación, se considera un sistema en equilibrio térmico en su ambiente, en el estado  $S$  con energía  $F(S)$ . La densidad de probabilidad en el espacio fase del punto que representa a  $S$  es proporcional a

$$\exp\left(-\frac{F(s)}{bT}\right)$$

donde  $b$  es la constante de Boltzmann, y  $T$  la temperatura absoluta del ambiente. Por esto, la proporción de tiempo que el sistema emplea en la solución es proporcional a la última expresión y de aquí la densidad de probabilidad del equilibrio  $\forall S \in \Omega$  es

$$\pi_S = \frac{\exp\left(-\frac{F(s)}{bT}\right)}{\int \exp\left(-\frac{F(s)}{bT}\right) dS}$$

con media para cualquier función solución válida  $f(S)$  dada por

$$E[f] = \frac{\int f(S) \exp\left(-\frac{F(s)}{bT}\right) dS}{\int \exp\left(-\frac{F(s)}{bT}\right) dS}.$$

El problema es que para muchas funciones solución, esta expresión no puede ser evaluada analíticamente, teniendo que recurrir a técnicas Monte Carlo para estimar el valor de las dos integrales. Sin embargo, esto a menudo falla en la práctica ya que el factor exponencial hace que una porción importante de las integrales esté concentrada en una región muy pequeña del espacio de soluciones  $\Omega$ . Metropolis resuelve esto discretizando el espacio de soluciones, siendo reemplazadas las integrales por sumatorios sobre  $\omega' \in \Omega$  soluciones discretas, y construyendo una cadena de Markov irreducible y aperiódica con probabilidades de transición  $\mathbf{P}(\omega, \omega')$  tal que

$$\pi(\omega') = \sum_{\omega \in \Omega} \pi(\omega) \mathbf{P}(\omega, \omega') \quad \forall \omega' \in \Omega$$

donde

$$\pi(\omega') = \frac{\exp\left(-\frac{F(\omega')}{bT}\right)}{\sum_{\omega \in \Omega} \exp\left(-\frac{F(\omega)}{bT}\right)} \quad \forall \omega' \in \Omega$$

Se observa que en el cálculo de la distribución de equilibrio, el denominador de la anterior expresión es una constante siempre (constante de normalización). Los ratios  $\frac{\pi(\omega')}{\pi(\omega)}$  es lo único que necesita ser calculado (liberándose del cálculo de dicho denominador) y la matriz de probabilidades de transición  $\mathbf{P}$ , que se sabe que

viene dada (Hammersley y Handscomb) por

$$\mathbf{P}(\omega, \omega') = \begin{cases} g(\omega, \omega') \frac{\pi(\omega')}{\pi(\omega)} & \frac{\pi(\omega')}{\pi(\omega)} < 1, \quad \omega' \neq \omega \\ g(\omega, \omega') & \frac{\pi(\omega')}{\pi(\omega)} \geq 1, \quad \omega' \neq \omega \\ g(\omega, \omega') + \Delta & \omega = \omega' \end{cases}$$

con  $\Delta = \sum_{\omega'' \in \Omega, \pi(\omega'') < \pi(\omega)} g(\omega, \omega'') \left(1 - \frac{\pi(\omega'')}{\pi(\omega)}\right)$  donde  $g(\omega, \omega') \geq 0$ , y  $\sum_{\omega' \in \Omega} g(\omega, \omega') = 1$ ,  $g(\omega, \omega') = g(\omega', \omega)$   $\forall \omega, \omega' \in \Omega$ , función simétrica para generación de probabilidades de la solución.

El uso de ratios de la probabilidad estacionaria para definir la solución de las probabilidades de aceptación, combinado con la solución simétrica de la generación de probabilidades, hace posible el uso de condición de reversibilidad para resolver el problema.

Puede que no se verifique la condición de reversibilidad, lo que dificulta la expresión explícita de la distribución estacionaria  $\pi_k$  para cada etapa  $k$ , aunque sí que existe solución mientras que  $\mathbf{P}_k$  sea irreducible y aperiódica.

Se sabe que la ecuación que se satisface es:

$$\begin{cases} \pi_k(\mathbf{I} - \mathbf{P}_k) = 0 \\ \pi_k \mathbf{e}^T = 1 \end{cases}$$

con  $\mathbf{e}$  vector fila de unos, que se puede reescribir como un sistema de ecuaciones lineales independientes dadas por

$$\pi_k(\mathbf{I} - \mathbf{P}_k)^{[i]} = \mathbf{e}_i$$

donde  $(\mathbf{I} - \mathbf{P}_k)^{[i]}$  es obtenida por sustitución de la  $i$ -ésima columna de  $(\mathbf{I} - \mathbf{P}_k)$  con un vector columna de unos. Resolviendo este sistema se obtienen las distribuciones estacionarias en las sucesivas etapas  $k$ .

### 3.2.3.2. Aproximación como una cadena de Markov no homogénea:

Aquí no se necesita alcanzar la distribución estacionaria para cada etapa  $k$ , pero por otro lado han de ser examinadas una sucesión infinita de iteraciones, con la condición de que la temperatura  $t_k$  decrezca suficientemente lento.

La prueba dada por Mirtra et al. está basada en las condiciones de cadena de Markov no homogénea con ergodicidad fuerte y débil, y se requieren 4 condiciones:

1. La cadena de Markov no homogénea asociada al S.A. debe ser débilmente ergódica (la dependencia con la solución inicial se desvanece en el límite).

2. Un autovector  $\pi_k$  con autovalor 1 debe existir para la ecuación de equilibrio del Teorema anterior para cada iteración  $k$ .
3. La cadena de Markov debe ser fuertemente ergódica (la cadena de Markov debe ser débilmente ergódica y la sucesión de autovectores debe converger a una forma que limita), es decir

$$\sum_{k=0}^{\infty} \|\pi_k - \pi_{k+1}\| < +\infty.$$

4. La sucesión de autovalores debe converger a una forma donde toda masa de probabilidad esté concentrada en el conjunto de soluciones  $\omega^*$  globalmente óptimas. Por lo tanto,

$$\lim_{k \rightarrow +\infty} \pi_k = \pi^{\text{opt}},$$

donde  $\pi^{\text{opt}}$  es la distribución equilibrio donde solo óptimos globales tienen probabilidades estrictamente positivas. (Se sabe que ergodicidad débil y fuerte son equivalentes en el caso de cadenas de Markov homogéneas).

En [4] satisfacen la condición primera (ergodicidad débil) dando primero una cota inferior en la probabilidad de alcanzar alguna solución desde un mínimo local, y luego mostrando que esta cota no se aproxima a cero demasiado rápido. Por ejemplo definen la cota inferior para las probabilidades de transición en la 1ª ecuación del Teorema como

$$P^{(m)}(\omega, \omega') > w^m \exp\left(-\frac{m\Delta_L}{t_{k_{m-1}}}\right) = +\infty,$$

y por tanto lo que se ha de hacer es reducir la temperatura de forma suficientemente lenta. En general, para  $k = 1, 2, \dots$  la sucesión  $\{t_k\}$  debe satisfacer

$$t_k \geq \frac{\beta}{\log(k)}$$

cuando  $\lim_{k \rightarrow \infty} t_k = 0$ ,  $\beta$  constante dependiente del problema, y  $k$  número de iteraciones.

Esta elección de temperatura es suficiente para asegurar la condición de convergencia, habiendo algunos valores de  $\beta$  que hagan la regla logarítmica condición necesaria. Además hay una única elección de  $\beta$  para el que la condición es necesaria y suficiente para la convergencia del S.A. en el conjunto de óptimos globales.



### 3.3. ALGORITMO VNS

Se trata de una metaheurística usada en problemas de optimización global cuya idea básica es cambiar de entorno sistemáticamente para hallar el óptimo local. Habrá versiones, también dependiendo del problema tratado, donde no sea necesaria una etapa de descenso local. Fue diseñado en sus orígenes para aproximar soluciones de problemas de optimización combinatoria, y posteriormente extendido en programación entera, no lineal, y entera no lineal. También ha sido usado como herramienta para teoría de grafos.

El número de aplicaciones ha crecido rápidamente, y pertenecen a varios campos: teoría de localización, análisis cluster, diseño de redes, inteligencia artificial, ingeniería, biología, fiabilidad, geometría, diseño en telecomunicaciones.

La VNS se basa en 3 hechos simples:

1. Un mínimo local con una estructura de entornos no lo es necesariamente con otra.
2. Un mínimo global es mínimo local con todas las posibles estructuras de entornos.
3. Para muchos problemas, los mínimos locales con la misma estructura o distinta de entornos están relativamente cerca.

Estos hechos sugieren el empleo de varias estructuras de entornos en las búsquedas locales para abordar un problema de optimización. El cambio de estructura de entornos se puede realizar de forma determinística, estocástica o determinística y estocástica a la vez.

Se puede definir una función llamada CambioDeEntorno mediante el siguiente algoritmo:

---

#### Cambio de entorno

---

**Función** CambioDeEntorno( $x, x', k$ )

1 **si**  $f(x') < f(x)$  **entonces**

2      $x \leftarrow x'$

3      $k \leftarrow 1$

**si no**

4      $k \leftarrow k + 1$

**Salida**  $x, k$

---

También se tiene la función Agitar para generar soluciones aleatorias:

---

#### Función para obtener nuevas soluciones aleatorias

---

**Función** Agitar( $x, x', k$ )

1    $\omega \leftarrow [1 + \text{Uniforme}(0, 1) \times |\mathcal{N}_k(x)|]$

2    $x' \leftarrow x^\omega$

**salida**  $x'$

---

Pueden usarse estas funciones para definir varias versiones del algoritmo VNS:

1. **VNS descendente:** una búsqueda local determina una mejor solución del entorno de la solución actual. La clásica búsqueda *greedy* descendente consiste en reemplazar iterativamente la solución actual por el resultado de la búsqueda local, mientras se produzca mejora. Si se realiza un cambio de estructura de entornos de forma determinística cada vez que se llega a un mínimo local, se obtiene la **búsqueda de entorno variable descendente** (VND), cuyos pasos son

---

**Algoritmo VNS descendente**

---

**Función**  $\text{VND}(x, k_{\text{máx}})$

1  $k \leftarrow 1$

2 **Repetir**

3  $x' \leftarrow \min_{y \in \mathcal{N}_k(x)} f(y)$  para encontrar óptimo en  $\mathcal{N}_k(x)$

4  $x, k \leftarrow \text{CambioDeEntorno}(x, x', k)$

**hasta que**  $k = k_{\text{máx}}$

**Salida**  $x$

---

La solución final proporcionada por el algoritmo es un mínimo local con respecto a todas las  $k_{\text{máx}}$  estructuras de entornos, y por tanto la probabilidad de alcanzar un mínimo global es mayor que usando una sola estructura.

Además del orden *secuencial* de las estructuras de entornos en la VND anterior, se puede desarrollar una estrategia anidada.

2. **VNS reducida:** se obtiene si se seleccionan soluciones aleatorias de  $\mathcal{N}_k(x)$ , sin aplicarles a continuación un descenso. Los pasos de la RVNS son:

---

**Algoritmo VNS reducido**


---

**Función**  $\text{VND}(x, k_{\text{máx}}, t_{\text{máx}})$ 

 1 **repetir**


---

 2      $k \leftarrow 1$ 


---

 3     **Repetir**


---

 4          $x' \leftarrow \text{Agitar}(x, k)$ 


---

 5          $x, k \leftarrow \text{CambioDeEntorno}(x, x', k)$ 


---

        **hasta que**  $k = k_{\text{máx}}$ 


---

 6      $\text{TiempoCPU}()$ 


---

        **hasta que**  $t > t_{\text{máx}}$ 


---

**Salida**  $x$ 


---

La RVNS es útil para instancias muy grandes de problemas en las que la búsqueda local es muy costosa. Se observa que el mejor valor para el parámetro  $k_{\text{máx}}$  es 2 ó 3. Como condición de parada se usa generalmente el máximo número de iteraciones entre dos mejoras. La RVNS se ha mostrado superior a algunos métodos de Monte-Carlo aplicados a ciertos problemas concretos.

3. **VNS básico:** combina cambios determinísticos y aleatorios de estructura de entornos. Se definen dos funciones que se usarán en este algoritmo. Los pasos que siguen dichas funciones son respectivamente:

---

**Heurística para mayor mejora**


---

**Función**  $\text{MayorMejora}(x)$ 

 1 **Repetir**


---

 2      $x' \leftarrow x$ 


---

 3      $x' \leftarrow \min_{y \in \mathcal{N}_k(x)} f(y)$  para encontrar óptimo en  $\mathcal{N}_k(x)$  (aquí puede usarse S.A.)
 

---

        **hasta que**  $f(x) \geq f(x')$ 


---

**Salida**  $x$ 


---

---

**Heurística para primera mejora**


---

**Función** PrimeraMejora( $x$ )

1    **repetir**


---

2         $x' \leftarrow x; i \leftarrow 0$ 


---

3        **Repetir**


---

4             $i \leftarrow i + 1$ 


---

5             $x \leftarrow \min\{f(x), f(x^i)\}, x^i \in \mathcal{N}(x)$ 


---

          **hasta que**  $(f(x) < f(x') \text{ ó } i = \mathcal{N}(x))$ 

          **hasta que**  $f(x) \geq f(x')$ 
**Salida**  $x$ 


---

Los pasos de la VNS básica son:

---

**VNS básico**


---

**Función** BVNS( $x, k_{\text{máx}}, t_{\text{máx}}$ )

1     $t \leftarrow 0$ 

2    **mientras que**  $t < t_{\text{máx}}$  **hacer**


---

3         $k \leftarrow 1$ 

4        **Repetir**


---

5             $x' \leftarrow \text{Agitar}(x)$ 

6             $x'' \leftarrow \text{MayorMejora}(x')$  (búsqueda local)

7             $x, k \leftarrow \text{CambioDeEntorno}(x, x'', k)$  (cambio de entorno)

---

          **hasta que**  $k = k_{\text{máx}}$ 

8         $\text{TiempoCPU}()$ 


---

**Salida**  $x$ 


---

La condición de parada puede ser, por ejemplo, máximo tiempo de CPU permitido, el máximo número de iteraciones, o el máximo número de iteraciones entre dos mejoras. Frecuentemente los entornos  $\mathcal{N}_k$  sucesivos están anidados. Se observa que la solución  $x'$  se genera al azar para evitar el ciclado, que puede ocurrir si se usa cualquier regla determinística.

4. **VNS general:** Ha sido el que ha dado lugar a las aplicaciones más exitosas aparecidas recientemente. Los pasos de este algoritmo son:

---

**Algoritmo VNS general**

---

**Función** GVNS( $x, l_{\text{máx}}, k_{\text{máx}}, t_{\text{máx}}$ )

1 **repetir**

2  $k \leftarrow 1$

3 **repetir**

4  $x' \leftarrow \text{Agitar}(x, k)$

5  $x'' \leftarrow \text{VND}(x', l_{\text{máx}})$

6  $x, k \leftarrow \text{CambioDeEntorno}(x, x'', k)$

**hasta que**  $k = k_{\text{máx}}$

7  $t \leftarrow \text{TiempoCPU}()$

**hasta que**  $t > t_{\text{máx}}$

**Salida**  $x$

---

### 3.3.1. Extensiones de la VNS

Se han propuesto algunas formas de extender la VNS, constituyendo mejoras prácticas que han permitido resolver con éxito problemas muy grandes.

#### 3.3.1.1. Extensiones básicas:

La BVNS es un método descendente de la 1ª mejora con aleatorización. Sin mucho esfuerzo adicional se transforma un método ascendente-descendente: en los pasos 6 y 7 hacer también  $x \leftarrow x'$  con alguna probabilidad.

También puede transformarse en una búsqueda de la mejor mejora: aplicar un movimiento al mejor  $k^*$  entre los  $k_{\text{máx}}$ . Otras extensiones serían encontrar la solución  $x'$  en el paso 5 como la mejor entre  $b$  soluciones generadas aleatoriamente en el  $k$ -ésimo entorno.

1. **VNS con descomposición:** extiende a la VNS en un esquema de entorno variable en dos niveles basado en la descomposición del problema. Sus pasos son:

---

**Algoritmo VNS con descomposición**


---

**Función** VNDS( $x, k_{\text{máx}}, t_{\text{máx}}, t_d$ )

 1 **repetir**


---

 2      $k \leftarrow 1$ 


---

 3 **repetir**


---

 4      $x' \leftarrow \text{Agitar}(x, k); y \leftarrow x'|x$ 

 5      $y' \leftarrow \text{BVNS}(y, k, t_d); x'' = (x'|y) \cup y'$ 

 6      $x''' \leftarrow \text{PrimeraMejora}(x'')$ 

 7      $x, k \leftarrow \text{CambioDeEntorno}(x, x''', k)$ 


---

**hasta que**  $k = k_{\text{máx}}$ 


---

**hasta que**  $k = k_{\text{máx}}$ 


---

**Salida**  $x$ 


---

$x'|x$  denota un conjunto de atributos presentes en  $x'$  pero no en  $x$ ; y cuando se escribe  $y'$  se denota una solución en el espacio de  $y$ , mientras que  $(x'|y) \cup y' = x''$  denota una solución en el correspondiente espacio completo.

Se observa que la única diferencia entre la VNS básica y la VNDS está en que en lugar de aplicar algún método de búsqueda local, lo que se hace en cada iteración es resolver un subproblema en un subespacio  $X_k \subset \mathcal{N}_k(x)$  con  $x' \in \mathcal{N}_k(x)$ . Cuando la búsqueda local usada en este paso es también VNS aparece un esquema VNS en dos niveles.

La VNDS puede verse como la inclusión del esquema de aproximación sucesiva, que ha sido históricamente muy utilizada en optimización.

2. **VNS sesgada:** una vez que se ha encontrado la mejor solución en una gran región es necesario alejarse bastante para obtener una mejor. La **búsqueda de entorno variable sesgada** afronta la exploración de valles alejados de la solución actual. La VNS sesgada incorpora una compensación por la distancia desde la solución actual para evitar que soluciones generadas en entornos muy alejados que puedan diferenciarse de la solución actual hagan que degenere el VNS.

Se usa una función  $\rho(x, x'')$  que mide la distancia entre la solución actual  $x$  y el óptimo local encontrado  $x''$ . La distancia usada para definir los entornos  $\mathcal{N}_k$  puede también usarse con este propósito. Ha de elegirse correctamente un parámetro  $\alpha$  debe permitir la exploración de valles lejanos a  $x$  cuando  $f(x'')$  es algo mayor que  $f(x)$ , pero no demasiado (en otro caso siempre se abandonaría la solución  $x$ ) y evitar movimientos frecuentes desde  $x$  a una solución cercana para volver a  $x$ . Un buen valor de  $\alpha$  tiene que determinarse experimentalmente en cada caso.

Se han de definir previamente unas funciones que se usarán en este algoritmo:

a) *Función que compare dos soluciones:*

---

**Algoritmo para guardar la mejor solución**

---

**Función** Compara( $x, k_{\text{máx}}$ )

1 **si**  $f(x') < f(x)$  **entonces**

2      $x \leftarrow x'$

**Salida**  $x$

---

b) *Función de cambio de entorno para el VNS sesgado:*

---

**Algoritmo para el cambio de entorno en el VNS sesgado**

---

**Función** CambioDeEntornoS( $x, k_{\text{máx}}$ )

1 **si**  $f(x') - \alpha \rho(x, x') < f(x)$  **entonces**

2      $x \leftarrow x'$

3      $k \leftarrow 1$

**si no**  $k \leftarrow k + 1$

**Salida**  $x, k$

---

Ahora puede darse el algoritmo:

---

**Algoritmo VNS con descomposición**

---

**Función** VNDs( $x, k_{\text{máx}}, t_{\text{máx}}, t_d$ )

1 **repetir**

2      $k \leftarrow 1$

3 **repetir**

4      $x' \leftarrow \text{Agitar}(x, k); y \leftarrow x'|x$

5      $y' \leftarrow \text{BVNS}(y, k, t_d); x'' = (x'|y) \cup y'$

6      $x''' \leftarrow \text{PrimeraMejora}(x'')$

7      $x, k \leftarrow \text{CambioDeEntorno}(x, x''', k)$

**hasta que**  $k = k_{\text{máx}}$

**hasta que**  $k = k_{\text{máx}}$

**Salida**  $x$

---

Podrían considerarse otras extensiones como la **VNS paralela**, donde simultáneamente pueden explorarse entornos del conjunto de soluciones en paralelo para optimizar la búsqueda.

### 3.3.1.2. Híbridos

Otra forma de extender la VNS ha sido incorporarla a otros metaheurísticas, dando lugar a diversas metaheurísticas híbridas:

1. **Busqueda tabú (Tabu Search, TS):** generalmente usa una estructura de entornos con respecto a la que se ejecuta movimientos de ascenso y descenso explotando distintos tipos de memoria. Hay dos maneras de hacer híbridos: usar algún tipo de memoria para orientar la búsqueda dentro del VNS, ó usar la VNS dentro de la TS.
2. **GRASP (Greedy Randomized Adaptive Search Procedure):** consta de dos fases: construcción de soluciones usando un procedimiento *greedy* aleatorizado, y en la segunda fase que es de mejora de la solución por búsqueda local puede usarse el VNS.
3. **Búsqueda multiarranque (MultiStart-Search, MS):** es una metaheurística que trata de evitar el estancamiento de un procedimiento descendente en un óptimo local, simplemente reiniciando desde otra solución. Puede usarse VNS como búsqueda local e hibridizarse con MS.

### 3.3.1.3. VNS para optimización continua

El VNS básico es una de las alternativas entre las diseñadas para la resolución de problemas de optimización, por evitar caer en mínimos locales al ir moviéndose en la estructura de entornos.

Puede verse el algoritmo ya escrito anteriormente donde en el paso de *búsqueda local* se puede aplicar algún método como el S.A., como se va a hacer en este trabajo.



## Capítulo 4

# INFERENCIA EN LOS PROCESOS DE DIFUSIÓN TIPO GOMPERTZ Y LOGÍSTICO

Partiendo de un muestreo discreto, se describe el método de máxima verosimilitud, para lo que es necesario conocer las distribuciones finito-dimensionales del proceso, ya que se requiere la distribución conjunta de las observaciones.

Las distribuciones finito-dimensionales quedan determinadas por las densidades de transición y la distribución inicial, en virtud de la propiedad de Markov que en nuestro caso se cumple.

Si  $\{X(t) / t \geq 0\}$  proceso de difusión con  $t_1, \dots, t_n$  instantes para los que se han observado  $x_1, \dots, x_n$  con  $x_k = X(t_k)$  realizaciones del proceso, la función de verosimilitud es

$$\mathbb{L}_{\mathbf{x}}(\theta_1, \theta) = f_1(x_1) \prod_{i=2}^n f(x_i, t_i | x_{i-1}, t_{i-1})$$

función de  $\Theta' = \Theta_1 \times \Theta$  con  $\Theta_1$  espacio paramétrico correspondiente a la distribución inicial, y  $\Theta$  espacio paramétrico correspondiente a las distribuciones de transición (que vamos a suponer independientes). El objetivo es establecer un criterio para hallar una estimación de los parámetros en función de los valores muestrales observados, del tipo  $\hat{\theta}_1 = \hat{\theta}_1(\mathbf{x})$  y  $\hat{\theta} = \hat{\theta}(\mathbf{x})$  estimaciones de los parámetros. Este criterio será

$$\mathbb{L}_{\mathbf{x}}(\hat{\theta}_1, \hat{\theta}) = \sup_{\theta_1, \theta} \mathbb{L}_{\mathbf{x}}(\theta_1, \theta)$$

es decir, maximizar la función de verosimilitud.

Para el caso de una trayectoria con distribución inicial degenerada, será  $P[X(t_1) = x_1] = 1$ , siendo la función de verosimilitud

$$\mathbb{L}_{\mathbf{x}}(\theta) = \prod_{i=2}^n f(x_i, t_i | x_{i-1}, t_{i-1})$$

Cuando se dispone de más de una trayectoria (éstas pueden corresponderse con la evolución del proceso sobre distintas unidades), el valor inicial no tiene porqué ser siempre el mismo, sino que habrá una distribución

para el instante inicial. En este caso, para la trayectoria  $i$ -ésima con  $i = 1, \dots, d$ , y dados los instantes de tiempo  $t_{ij}$  con  $j = 1, \dots, n_i$  con función de verosimilitud dada por

$$\mathbb{L}_{\mathbf{x}}(\theta_1, \theta) = \prod_{i=1}^d f_1(x_{i1}) \prod_{j=2}^{n_i} f(x_{ij}, t_{ij} | x_{i,j-1}, t_{i,j-1})$$

con  $f_1$  la densidad de la distribución inicial, que se supone conocida.

En caso de que la distribución inicial sea degenerada será la función de verosimilitud:

$$\mathbb{L}_{\mathbf{x}}(\theta) = \prod_{i=1}^d \prod_{j=2}^{n_i} f(x_{ij}, t_{ij} | x_{i,j-1}, t_{i,j-1})$$

y el logaritmo de la función de verosimilitud en el caso general viene dado por

$$\log(\mathbb{L}_{\mathbf{x}}(\theta_1, \theta)) = \sum_{i=1}^d \log(f_1(x_{i1})) + \sum_{i=1}^d \sum_{j=2}^{n_i} \log(f(x_{ij}, t_{ij} | x_{i,j-1}, t_{i,j-1}))$$

con la parte de  $\theta_1$  independiente de  $\theta$ , por lo que su estimación puede hacerse de forma independiente también.

## 4.1. CASO LOGARÍTMICO NORMAL CON FACTORES EXÓGENOS

En este caso se sustituirá por la expresión de la densidad de transición ya calculada para este proceso en concreto. Sean  $\theta_1 = (\mu_0, \sigma_0^2)$  los parámetros de la distribución inicial, y  $\theta$  los parámetros correspondientes a la distribución de transición. La función de verosimilitud será, por tanto:

$$\begin{aligned} \mathbb{L}_{\mathbf{x}}(\theta_1, \theta) &= \prod_{i=1}^d f_1(x_{i1}) \prod_{j=2}^{n_i} f(x_{ij}, t_{ij} | x_{i,j-1}, t_{i,j-1}) = \\ &= \prod_{i=1}^d \frac{1}{\sigma_0 x_{i1} \sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \frac{(\log(x_{i1}) - \mu_0)^2}{\sigma_0^2} \right\} \times \\ &\times \prod_{j=2}^{n_i} \frac{1}{\sigma_0 x_{ij} \sqrt{2\pi(t_{ij} - t_{i,j-1})}} \exp \left\{ -\frac{1}{2} \frac{\left( \log \left( \frac{x_{ij}}{x_{i,j-1}} \right) - \int_{t_{i,j-1}}^{t_{ij}} h(u) du + \frac{\sigma^2}{2} (t_{ij} - t_{i,j-1}) \right)^2}{\sigma^2 (t_{ij} - t_{i,j-1})^2} \right\} \end{aligned}$$

y ahora se toma logaritmo:

$$\begin{aligned} \log(\mathbb{L}_{\mathbf{x}}(\theta_1, \theta)) &= -\sum_{i=1}^d \log(\sigma_0 x_{i1} \sqrt{2\pi}) - \frac{1}{2} \sum_{i=1}^d \frac{(\log(x_{i1}) - \mu_0)^2}{\sigma_0^2} - \\ &- \sum_{i=1}^d \sum_{j=2}^{n_i} \log \left( \sigma_0 x_{ij} \sqrt{2\pi(t_{ij} - t_{i,j-1})} \right) - \frac{1}{2} \sum_{i=1}^d \sum_{j=2}^{n_i} \frac{\left( \log \left( \frac{x_{ij}}{x_{i,j-1}} \right) - \int_{t_{i,j-1}}^{t_{ij}} h(u) du + \frac{\sigma^2}{2} (t_{ij} - t_{i,j-1}) \right)^2}{\sigma^2 (t_{ij} - t_{i,j-1})^2} \end{aligned}$$

Puede escribirse  $\mathbb{L}_{\mathbf{x}}(\theta_1, \theta) = \mathbb{L}_{\mathbf{x}}^1(\theta_1) \mathbb{L}_{\mathbf{x}}^2(\theta)$  donde se ha llamado :

$$\mathbb{L}_{\mathbf{x}}^1(\theta_1) = \prod_{i=1}^d \frac{1}{\sigma_0 x_{i1} \sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \frac{(\log(x_{i1}) - \mu_0)^2}{\sigma_0^2} \right\} = \frac{1}{(2\pi)^{d/2} \sigma_0^d} \left( \prod_{i=1}^d \frac{1}{x_{i1}} \right) \exp \left\{ -\frac{1}{2\sigma_0^2} \sum_{i=1}^d (\log(x_{i1}) - \mu_0)^2 \right\}$$

$$\mathbb{L}_{\mathbf{x}}^2(\theta) = \prod_{i=1}^d \prod_{j=2}^{n_i} \frac{1}{\sigma_0 x_{ij} \sqrt{2\pi(t_{ij} - t_{i,j-1})}} \exp \left\{ -\frac{1}{2} \frac{\left( \log \left( \frac{x_{ij}}{x_{i,j-1}} \right) - \int_{t_{i,j-1}}^{t_{ij}} h(u) du + \frac{\sigma^2}{2} (t_{ij} - t_{i,j-1}) \right)^2}{\sigma^2 (t_{ij} - t_{i,j-1})^2} \right\}$$

En cualquier caso, como debido a la factorización de la función de verosimilitud la estimación de  $\theta_1$  y  $\theta$  son independientes, puede hacerse para  $\theta_1 = (\mu_0, \sigma_0^2)$ :

$$\log \mathbb{L}_{\mathbf{x}}^1(\mu_0, \sigma_0) = -\frac{d}{2} \log(2\pi) - d \log(\sigma_0) - \sum_{i=1}^d \frac{1}{x_{i1}} - \frac{1}{2\sigma_0^2} \sum_{i=1}^d (\log(x_{i1}) - \mu_0)^2$$

y ahora derivo e igualo a cero obteniendo las ecuaciones de verosimilitud:

$$\frac{\partial \log(\mathbb{L}_{\mathbf{x}}^1(\mu_0, \sigma_0))}{\partial \mu_0} = \frac{1}{\sigma_0^2} \sum_{i=1}^d (\log(x_{i1}) - \mu_0) = 0$$

y de aquí es:

$$\hat{\mu}_0 = \frac{1}{d} \sum_{i=1}^d \log(x_{i1})$$

y por otro lado

$$\frac{\partial \log(\mathbb{L}_{\mathbf{x}}^1(\mu_0, \sigma_0))}{\partial \sigma_0} = -\frac{d}{\sigma_0} + \frac{1}{\sigma_0^3} \sum_{i=1}^d (\log(x_{i1}) - \mu_0)^2 = 0$$

y de aquí es:

$$\hat{\sigma}_0^2 = \frac{1}{d} \sum_{i=1}^d (\log(x_{i1}) - \mu_0)^2 = \frac{1}{d} \sum_{i=1}^d \log(x_{i1})^2 - \mu_0^2$$

que expresado en términos de las observaciones es

$$\hat{\sigma}_0^2 = \frac{1}{d} \sum_{i=1}^d \log(x_{i1})^2 - \frac{1}{d^2} \left( \sum_{i=1}^d \log(x_{i1}) \right)^2$$

Ahora se derivaría e igualaríamos  $\mathbb{L}_{\mathbf{x}}^2(\theta)$  a cero para obtener un sistema de ecuaciones, de donde se obtiene la estimación máximo verosímil. Puesto que en la función  $h(t)$  están involucrados los parámetros, este paso se hará en los casos concretos a estudiar.

## 4.2. INFERENCIA EN EL MODELO DE GOMPERTZ

Llamando  $a = \frac{m}{\beta}$  y  $b = e^{-\beta}$ , si se deriva e iguala a cero el logaritmo de  $\mathbb{L}_x^2(a, b, \sigma^2)$ , se obtiene el siguiente sistema de ecuaciones

$$\sum_{i=1}^d \sum_{j=2}^{n_i} \frac{\left[ \log \frac{x_{ij}}{x_{i,j-1}} + a(b^{t_{ij}} - b^{t_{i,j-1}}) + \frac{\sigma^2}{2}(t_{ij} - t_{i,j-1}) \right]}{(t_{ij} - t_{i,j-1})} [b^{t_{ij}} - b^{t_{i,j-1}}] = 0,$$

$$(k-d)\sigma^2 + \sigma^2 \sum_{i=1}^d \sum_{j=2}^{n_i} \left[ \log \frac{x_{ij}}{x_{i,j-1}} + a(b^{t_{ij}} - b^{t_{i,j-1}}) + \frac{\sigma^2}{2}(t_{ij} - t_{i,j-1}) \right] = \sum_{i=1}^d \sum_{j=2}^{n_i} \frac{\left[ \log \frac{x_{ij}}{x_{i,j-1}} + a(b^{t_{ij}} - b^{t_{i,j-1}}) + \frac{\sigma^2}{2}(t_{ij} - t_{i,j-1}) \right]^2}{(t_{ij} - t_{i,j-1})},$$

$$\sum_{i=1}^d \sum_{j=2}^{n_i} \frac{\left[ \log \frac{x_{ij}}{x_{i,j-1}} + a(b^{t_{ij}} - b^{t_{i,j-1}}) + \frac{\sigma^2}{2}(t_{ij} - t_{i,j-1}) \right]}{(t_{ij} - t_{i,j-1})} [b^{t_{ij}} - b^{t_{i,j-1}}] = 0,$$

para el cual no hay solución explícita, teniendo que recurrir a procedimientos numéricos para obtener una aproximación de ésta. En el caso particular de que  $t_{ij} - t_{i,j-1} = h \forall i = 1, \dots, d$  y  $j = 1, \dots, n_i$  (lo más usado normalmente en la práctica), se pueden simplificar estas ecuaciones a:

$$\begin{cases} A_{3,b} + a(b^h - 1)A_{2,b} + \frac{\sigma^2 h}{2}A_{1,b} = 0, & (G1) \end{cases}$$

$$\begin{cases} A_{4,b} + a^2(b^h - 1)^2 A_{2,b} + 2a(b^h - 1)A_{3,b} - \frac{\sigma^4 h^2}{4}(k-d) - (k-d)\sigma^2 h = 0, & (G2) \end{cases}$$

$$\begin{cases} A_{3,b}^* + a(b^h - 1)A_{2,b}^* + \frac{\sigma^2 h}{2}A_{1,b}^* = 0, & (G3) \end{cases}$$

donde

$$A_{1,b} = \sum_{i=1}^d \sum_{j=2}^{n_i} b^{t_{i,j-1}}, \quad A_{2,b} = \sum_{i=1}^d \sum_{j=2}^{n_i} b^{2t_{i,j-1}}, \quad A_{1,b} = \sum_{i=1}^d \sum_{j=2}^{n_i} b^{t_{i,j-1}} \log \frac{x_{ij}}{x_{i,j-1}}$$

$$A_{1,b}^* = \sum_{i=1}^d \sum_{j=2}^{n_i} t_{i,j-1} b^{t_{i,j-1}}, \quad A_{2,b}^* = \sum_{i=1}^d \sum_{j=2}^{n_i} t_{i,j-1} b^{2t_{i,j-1}},$$

$$A_{3,b} = \sum_{i=1}^d \sum_{j=2}^{n_i} t_{i,j-1} b^{t_{i,j-1}} \log \frac{x_{ij}}{x_{i,j-1}}, \quad A_{3,b}^* = \sum_{i=1}^d \sum_{j=2}^{n_i} t_{i,j-1} \log^2 \frac{x_{ij}}{x_{i,j-1}}$$

Denotando

$$C_b = \frac{A_{1,b}A_{3,b}^* - A_{1,b}^*A_{3,b}}{A_{1,b}^*A_{2,b} - A_{1,b}A_{2,b}^*} \quad y \quad D_b = \frac{A_{2,b}^*A_{3,b} - A_{2,b}A_{3,b}^*}{A_{1,b}^*A_{2,b} - A_{1,b}A_{2,b}^*}$$

y tras algunos cálculos, del sistema de ecuaciones se obtiene que

$$a_b = \frac{C_b}{b^h - 1} \quad y \quad \sigma_b^2 = \frac{2D_b}{h}$$

y sustituyendo estas expresiones en (G2) se obtiene la ecuación en  $b$ :

$$A_{4,b} + A_{2,b}C_b^2 + 2A_{3,b}C_b - (k-d)D_b^2 - 2(k-d)D_b = 0$$

que debe ser tratada por métodos numéricos y cuya solución proporcionará la estimación MV de  $b$ . Una vez obtenido  $\hat{b}$ , la estimación MV de  $a$  y  $\sigma^2$  es  $\hat{a} = a_{\hat{b}}$  y  $\hat{\sigma}^2 = \sigma_{\hat{b}}^2$ . Y aplicando el Teorema de Invarianza de Zenha pueden obtenerse estimaciones de funciones dependientes de los parámetros estimados.

En vista de la complejidad de la ecuación, la cual aumenta cuando lo hace el tamaño de la muestra, se requiere necesariamente el uso de procedimientos numéricos. Estudiando la función  $g$  que define este sistema de ecuaciones, dada por

$$g(b) = A_{4,b} + A_{2,b}C_b^2 + 2A_{3,b}C_b - (k-d)D_b^2 - 2(k-d)D_b$$

se observa que procedimientos clásicos no son viables, por la dependencia de la función con la muestra. En particular, hay ejemplos en los que el método de Newton-Raphson falla, ya que las condiciones que garantizan la convergencia a la solución no se cumplen.

También se podría aplicar el método de bisección, ya que es una única función con la que se resuelve todo el sistema, pero en este caso habría de tomarse dos valores iniciales que determinen un entorno de la solución para  $b$ , y se observa que:

1. La función  $g$  no está definida para  $b = 1$ , por lo que el valor inicial a la derecha ha de ser estrictamente menor que uno.
2. Si el valor de  $b$  es muy próximo a 1, el procedimiento se realentizará mucho, y además no hay seguridad de haber escogido el extremo derecho de forma que deje al valor de  $b$  dentro del intervalo de bisección.

Como una alternativa puede verse un procedimiento iterativo en [8] para el caso en que el valor de la cota de cada muestra puede ser visualizado. El problema es que su convergencia no ha sido probada teóricamente, aunque sí ha sido validada empíricamente.

Todas estas consideraciones sugieren el uso de otro tipo de procedimientos, como los que se describen en este trabajo, los cuales han demostrado ser herramientas útiles para resolver una amplia variedad de problemas de optimización, y puede ser directamente aplicadas, como en este caso, al logaritmo de la función objetivo.

### 4.3. INFERENCIA EN EL MODELO LOGÍSTICO

LLamando  $\gamma = e^{-c}$  y  $\eta = \frac{1}{b}$ , si se deriva e iguala a cero el logartimo de  $\mathbb{L}_{\mathbf{x}}^2(\gamma, \eta, \sigma^2)$ , se obtiene el siguiente sistema de ecuaciones

$$\sum_{i=1}^d \sum_{j=2}^{n_i} \frac{\log \frac{x_{ij}}{x_{i,j-1}} - T_{ij}^{\eta, \gamma}}{(t_{ij} - t_{i,j-1}) S_{ij}^{\eta, \gamma}} W_{ij}^{\eta, \gamma} + \frac{\sigma^2}{2} \sum_{i=1}^d \sum_{j=2}^{n_i} \frac{W_{ij}^{\eta, \gamma}}{S_{ij}^{\eta, \gamma}} = 0,$$

$$\sum_{i=1}^d \sum_{j=2}^{n_i} \frac{\log \frac{x_{ij}}{x_{i,j-1}} - T_{ij}^{\eta, \gamma}}{(t_{ij} - t_{i,j-1}) S_{ij}^{\eta, \gamma}} V_{ij}^{\eta, \gamma} + \frac{\sigma^2}{2} \sum_{i=1}^d \sum_{j=2}^{n_i} \frac{V_{ij}^{\eta, \gamma}}{S_{ij}^{\eta, \gamma}} = 0,$$

$$\sigma^4 \sum_{i=1}^d (t_{i,n_i} - t_{i,1}) + 4\sigma^2(n-d) - 4 \sum_{i=1}^d \sum_{j=2}^{n_i} \frac{\log^2 \frac{x_{ij}}{x_{i,j-1}}}{t_{ij} - t_{i,j-1}} - 4 \sum_{i=1}^d \sum_{j=2}^{n_i} \frac{(T_{ij}^{\eta, \gamma})^2}{t_{ij} - t_{i,j-1}} + 8 \sum_{i=1}^d \sum_{j=2}^{n_i} \frac{\log \frac{x_{ij}}{x_{i,j-1}} T_{ij}^{\eta, \gamma}}{t_{ij} - t_{i,j-1}} = 0,$$

donde

$$S_{ij}^{\eta, \gamma} = (\eta + \gamma^{t_{i,j-1}-1}(\eta - \gamma^{t_{ij}})), \quad T_{ij}^{\eta, \gamma} = \log \left( \frac{\eta + \gamma^{t_{i,j-1}}}{\eta + \gamma^{t_{ij}}} \right),$$

$$V_{ij}^{\eta, \gamma} = t_{i,j-1} \gamma_{i,j-1-1} (\eta + \gamma^{t_{ij}} - t_{ij} \gamma^{t_{ij}-1} (\eta + \gamma^{t_{i,j-1}})),$$

$$W_{ij}^{\eta, \gamma} = \gamma^{t_{i,j-1}} - \gamma^{t_{ij}}$$

y el sistema de ecuaciones se puede expresar como

$$\begin{cases} \sigma^2 X_1^{\eta, \gamma} + 2X_2^{\eta, \gamma} - 2X_3^{\eta, \gamma} = 0 & (L1) \\ \sigma^2 X_4^{\eta, \gamma} + 2X_5^{\eta, \gamma} - 2X_6^{\eta, \gamma} = 0 & (L2) \\ \sigma^4 Z_2 + 4\sigma^2(n-d) - 4Y_2^{\eta, \gamma} + 8Y_1^{\eta, \gamma} - 4Z_1 = 0 & (L3) \end{cases}$$

donde

$$\begin{aligned} X_1^{\eta, \gamma} &= \sum_{i=1}^d \sum_{j=2}^{n_i} \frac{W_{ij}^{\eta, \gamma}}{S_{ij}^{\eta, \gamma}}, & X_2^{\eta, \gamma} &= \sum_{i=1}^d \sum_{j=2}^{n_i} \frac{W_{ij}^{\eta, \gamma}}{(t_{ij} - t_{i,j-1}) S_{ij}^{\eta, \gamma}} \log \left( \frac{x_{ij}}{x_{i,j-1}} \right), \\ X_3^{\eta, \gamma} &= \sum_{i=1}^d \sum_{j=2}^{n_i} \frac{T_{ij}^{\eta, \gamma} W_{ij}^{\eta, \gamma}}{(t_{ij} - t_{i,j-1}) S_{ij}^{\eta, \gamma}}, & X_4^{\eta, \gamma} &= \sum_{i=1}^d \sum_{j=2}^{n_i} \frac{V_{ij}^{\eta, \gamma}}{S_{ij}^{\eta, \gamma}}, \\ X_5^{\eta, \gamma} &= \sum_{i=1}^d \sum_{j=2}^{n_i} \frac{V_{ij}^{\eta, \gamma}}{(t_{ij} - t_{i,j-1}) S_{ij}^{\eta, \gamma}} \log \left( \frac{x_{ij}}{x_{i,j-1}} \right), & X_6^{\eta, \gamma} &= \sum_{i=1}^d \sum_{j=2}^{n_i} \frac{T_{ij}^{\eta, \gamma} V_{ij}^{\eta, \gamma}}{(t_{ij} - t_{i,j-1}) S_{ij}^{\eta, \gamma}}, \end{aligned}$$

$$Y_1^{\eta,\gamma} = \sum_{i=1}^d \sum_{j=2}^{n_i} \frac{T_{ij}^{\eta,\gamma}}{t_{ij} - t_{i,j-1}} \log \left( \frac{x_{ij}}{x_{i,j-1}} \right), \quad Y_2^{\eta,\gamma} = \sum_{i=1}^d \sum_{j=2}^{n_i} \frac{(T_{ij}^{\eta,\gamma})^2}{t_{ij} - t_{i,j-1}} \log \left( \frac{x_{ij}}{x_{i,j-1}} \right),$$

$$Z_1 = \sum_{i=1}^d \sum_{j=2}^{n_i} \frac{\log^2 \frac{x_{ij}}{x_{i,j-1}}}{t_{ij} - t_{i,j-1}}, \quad Z_2 = \sum_{i=1}^d (t_{i,n_i} - t_{i1})$$

y de la ecuación (L1) se obtiene que  $\sigma_{\eta,\gamma}^2 = 2S^{\eta,\gamma}$ , donde

$$S^{\eta,\gamma} = \frac{X_3^{\eta,\gamma} - X_2^{\eta,\gamma}}{X_1^{\eta,\gamma}}$$

que puede sustituirse en las ecuaciones (L2) y (L3), resultando finalmente el sistema

$$\begin{cases} X_6^{\eta,\gamma} - S^{\eta,\gamma} X_4^{\eta,\gamma} - X_5^{\eta,\gamma} = 0 \\ S^{\eta,\gamma} (S^{\eta,\gamma} Z_2 + 2(n-d)) + (2Y_1^{\eta,\gamma} - Y_2 X_6^{\eta,\gamma}) - Z_1 = 0. \end{cases}$$

La solución a este sistema proporciona la estimación máximo-verosímil de  $\eta$  y  $\gamma$ , de donde se obtiene la de  $\sigma^2$  dada por  $2S^{\hat{\eta},\hat{\gamma}}$ . La estimación de cualquier función paramétrica puede ser estimada aplicando el Teorema de Invarianza de Zenha.

Para este caso se vuelve a hacer una simplificación de las ecuaciones obtenidas. Se llegará a que el número final de ecuaciones a resolver será 2, obteniendo como en el caso de Gompertz, un sistema de ecuaciones muy complejo. Por esta razón es necesario el uso de procedimientos numéricos, de los cuales la mayoría necesitan una solución inicial para ser aplicados.

Con el objetivo de obtener una buena solución inicial, se han propuesto algunas alternativas basadas en la información proporcionada por las trayectorias observadas del proceso, como puede verse en [5]. Aquí se habla de dos procedimientos basados en las relaciones existentes entre los parámetros de la curva logística a través del punto de inflexión.

En este trabajo se harán estimaciones con los algoritmos S.A., y VNS, igual que para el proceso Gompertz. Las consideraciones que se hagan sobre este algoritmo se verán en el correspondiente capítulo (capítulo 5, de aplicación de algoritmos a datos simulados).





## Capítulo 5

# APLICACIÓN DE LOS ALGORITMOS

### 5.1. APLICACIÓN CON DATOS SIMULADOS

Para la simulación del proceso de Wiener, podemos tener en cuenta que viene representado por la siguiente expresión

$$X(t) = x_0 \exp \left( \sigma W(t - t_0) + \int_{t_0}^t h(u) du - \frac{\sigma^2}{2}(t - t_0) \right),$$

por lo que bastaría con simular  $W(t - t_0)$  proceso de Wiener, y evaluar la anterior expresión para obtener una simulación del proceso.

#### 5.1.1. Simulación de los procesos

Con la función construida para simular el proceso de Wiener, basta evaluar la expresión. Se construye otra función en R para la simulación del proceso logístico, teniendo en cuenta que en este caso

$$\int_{t_0}^t h(s) ds = \int_{t_0}^t m e^{-\beta s} ds = \frac{m}{\beta} (e^{-\beta t_0} - e^{-\beta t}) \quad \text{Gompertz}$$

$$\int_{t_0}^t h(s) ds = \int_{t_0}^t \frac{bc}{e^{cs} + b} ds = \log \left( \frac{1 + b e^{-ct_0}}{1 + b e^{-ct}} \right) \quad \text{Logístico}$$

Ahora se introduce una breve descripción sobre los parámetros de los algoritmos desarrollados:

1. **Simulated Annealing:** el entorno de soluciones (único) se introduce mediante matriz de  $p \times 2$ , con  $p$  el número de parámetros (3 en caso Gompertz y logístico), donde las componentes de cada fila son los extremos donde se mueve el parámetro. En el algoritmo se introduce manualmente ya que esta misma función se usará en el VNS descendente, aunque para aplicar solamente el S.A. no es necesario que dependa del entorno, si no que puede dejarse fijo siempre, y sea el espacio total de soluciones considerado.

En los algoritmos adaptados a los casos concretos, ni el entorno, ni la solución inicial forman parte de los parámetros correspondientes a la función que define cada algoritmo, sino que van integrados dentro de la misma. Los parámetros comunes en ambos casos son

- a) **temp0**: temperatura inicial (que debe ser alta).
- b) **alpha**: constante de enfriamiento. Si **temp** es la temperatura en la  $k$ -ésima iteración, en la siguiente la temperatura será  $\alpha \cdot \text{temp}$ , debiendo ser  $\alpha < 1$ .
- c) **x**: trayectorias observadas.
- d) **t0**: En el caso del S.A. para el proceso de Gompertz, se debe introducir el tiempo inicial  $t_0$ . Para el Logístico no es necesario.
- e) **h**: longitud de paso para el proceso de Wiener.
- f) **n**: número de observaciones de cada trayectoria.
- g) **d**: número total de trayectorias.
- h) **M**: la longitud de la cadena. Fijada una temperatura en cierta iteración, se completa dentro de ésta un ciclo de  $M$  iteraciones, donde generalmente  $M=50$ .

2. **VNS descendente**: aquí se hace un híbrido con el S.A.. Se trata de la versión descendente de este algoritmo introduciendo la función `simann.entorno.gompertz/logistico` en el paso de búsqueda local. En este caso el parámetro longitud de cadena queda determinado por el valor 50 por defecto, y la función del VNS no dependerá de él.

- a) **p**: número de parámetros de la función objetivo. Aunque en los casos estudiados siempre hay 3 parámetros, se deja así porque la función `vns.descendente` es válida para más tipos de funciones objetivo que puedan tener más o menos parámetros.
- b) **kmax**: número máximo de iteraciones en la búsqueda de solución al ir recorriendo los entornos. Para la versión reducida se requerirá un número alto, para la versión descendente, que es la que se aplica en este caso, se usa un número bajo (no más de 20) dando un buen resultado.
- c) **f**: la función objetivo.
- d) **temperatura**: temperatura inicial usada en el S.A..
- e) **coldconst**: constante de enfriamiento usada en el S.A..
- f) **x**: trayectorias observadas.
- g) **ini**: solución inicial.
- h) **h**: longitud de paso para el proceso de Wiener.
- i) **n**: número de observaciones de cada trayectoria.
- j) **d**: número total de trayectorias.
- k) **entorno**: entorno donde busco las soluciones, construido de la misma forma que en el S.A..

En el caso de la aplicación del VNS híbrido con S.A., se puede establecer el número de iteraciones máximo como  $k_{\text{máx}} = 10$ . ya que inicialmente, ante una posible solución generada por el S.A., esta cantidad de iteraciones proporciona un margen para poder encontrar mejora (menos de 10 podría ser poco). Es claro que cuanto mayor sea  $k_{\text{máx}}$  más alta será la probabilidad de obtener mejores soluciones, pero esto implica un aumento muy significativo del tiempo de ejecución.

## 5.2. ESTIMACIÓN EN EL PROCESO DE GOMPERTZ CON DATOS SIMULADOS:

Se dan algunos ejemplos variando los parámetros del algoritmo para ver cómo afecta a los resultados. El objetivo es ilustrar cómo han de elegirse los distintos parámetros. Se dará el error relativo obtenido, esto es:

$$\frac{|m - \hat{m}|}{m} + \frac{|\beta - \hat{\beta}|}{\beta} + \frac{|\sigma - \hat{\sigma}|}{\sigma} \quad \text{para el Gompertz}$$

$$\frac{|b - \hat{b}|}{b} + \frac{|c - \hat{c}|}{c} + \frac{|\sigma - \hat{\sigma}|}{\sigma} \quad \text{para el Logístico}$$

y también el valor que toma la función objetivo.

### 5.2.1. Estimaciones con Simulated Annealing

En los ejemplos que se dan, suponemos que las observaciones en las distintas trayectorias se realizarán simultáneamente, siendo  $t_{i_0j} = t_{i_1j} = t_j$ , para cada  $1 \leq i_0, i_1 \leq d$ , con  $t_0 \leq t_j \leq t_1$ , y los instantes de tiempo son equiespaciados, con longitud de paso  $h$ . Se van a hacer estimaciones de los parámetros mediante el algoritmo del S.A., para lo cual habrá que hacer algunas consideraciones sobre los mismos:

Se quiere estimar los parámetros  $m$ ,  $\beta$  y  $\sigma^2$  y para ello, en vista de la forma que tiene la función, puede considerarse la reparametrización dada por

$$a = \frac{m}{\beta}$$

$$b = e^{-\beta}$$

y así nuestra función objetivo (quitando la parte de la distribución inicial), viene dada por:

$$f(a, b, \sigma^2) = \frac{N-d}{2} \log(\sigma^2) + \frac{1}{2h\sigma^2} \sum_{i=1}^d \sum_{j=2}^{n_i} \left( \log \frac{x_{ij}}{x_{i,j-1}} + ab^{t_{i,j-1}}(b^h - 1) + \frac{\sigma^2 h}{2} \right)^2.$$

Sobre el espacio de soluciones se observa que:

1.  $b \in ]0, 1[$ , luego puede considerarse un número aleatorio cualquiera entre 0 y 1 en la solución inicial.
2. Para  $\sigma^2$  no es recomendable la modelización del problema mediante curvas de crecimiento de tipo Gompertz cuando la varianza es grande. Por ello se acota  $\sigma \in ]0, 0.1]$ .
3. Aunque en principio  $a > 0$  sin restricción, en vista de la muestra de la que se dispone se observa que para cada trayectoria, si  $k_i$  es la cota superior, han de tomarse los parámetros de forma que

$$x_{i0} e^{ab^{t_0}} < k_i$$

que equivale a que

$$a < \frac{\log \left( \max \frac{k_i}{x_{i0}} \right)}{b^{t_0}}$$

para  $b \in ]0, 1[$  ya elegido, y por otro lado es adecuado establecer como cota inferior

$$a > \frac{\log \left( \min \frac{k_i}{x_{i0}} \right)}{b^{t_0}}$$

Así ya tenemos el espacio de soluciones

$$\Omega = \left[ \frac{\log \left( \min \frac{k_i}{x_{i0}} \right)}{b^{t_0}}, \frac{\log \left( \max \frac{k_i}{x_{i0}} \right)}{b^{t_0}} \right] \times ]0, 1[ \times ]0, 0.01[$$

La *temperatura inicial* se elige observando los incrementos de la función objetivo cuando se eligen soluciones aleatorias que mejoran la solución. Se ha observado, experimentalmente en este caso, esta temperatura suele ser como mínimo de 3000. Algunas temperaturas obtenidas para los datos con los que se trabaja son: 2789.271, 9103.044, 27829.6, 77824.55, 38709.2, 13078.51, 22891.48. Se van a ver ejemplos con 2 temperaturas, y otro con temperatura elegida de la forma descrita anteriormente. El efecto de una mayor temperatura, en vista del algoritmo, es que el movimiento por el espacio de soluciones sea mayor, ya que la probabilidad exponencial que aparece se hace mayor cuanto mayor sea esta temperatura, saltando con más frecuencia de soluciones mejoradas respecto a la anterior.

Para la *longitud de la cadena* se sabe que es suficiente con 50.

La *constante de enfriamiento* suele elegirse entre 0.8 y 0.99, y esta proporciona una exploración más intensa (mayor número de iteraciones) del espacio de soluciones. Se observa que el crecimiento de ésta proporciona un incremento del número total de operaciones cuyo aumento es de mayor intensidad que lineal. El criterio elegido es que la temperatura vaya decreciendo en cada paso mediante la regla

$$t_{k+1} = \alpha t_k$$

que equivale a

$$t_{k+1} = \alpha^k t_0$$

con  $t_0$  la temperatura inicial elegida. Se hace esto hasta que la temperatura  $t_k$  sea inferior a 0.01, que equivale a que el número de iteraciones venga dado por la parte entera de

$$-\frac{\log(t_0 \times 10)}{\log(\alpha)} + 1$$

Se muestra una tabla con algunos ejemplos del número de iteraciones totales por etapa (el número total de operaciones resulta de multiplicar por 50, que es la longitud de la cadena):

Constante de enfriamiento $\alpha$	Temperatura inicial $T_0$	Número de iteraciones
0.8	600	39
	5000	49
	13000	53
0.85	600	54
	5000	67
	13000	73
0.9	600	83
	5000	103
	13000	112
0.95	600	170
	5000	211
	13000	230
0.99	600	866
	5000	1077
	13000	1172
0.999	600	8696
	5000	10815
	13000	11770

La siguiente tabla recoge los resultados sobre la simulación de 10 trayectorias con 50 observaciones cada una de un proceso de Gompertz de parámetros  $m = 7$ ,  $\beta = 2$  y  $\sigma = 0.03$ , donde el algoritmo se ejecuta bajo temperaturas iniciales de 5000, 10000 y una temperatura generada automáticamente estimando la media de los incrementos de 100 soluciones aleatorias generadas (esto se hace usando la función `temperatura.gompertz` que puede verse en el apéndice B). Para estos casos se toma constante de enfriamiento de  $\alpha = 0.95$  en los dos primeros casos, y  $\alpha = 0.99$  en el último:

	$T_0 = 5000, \alpha = 0.95$	$T_0 = 10000, \alpha = 0.95$	$T_0$ por estimación, $\alpha = 0.99$
$m = 7$	<b>7.048878</b>	<b>7.010974</b>	<b>7.03245</b>
$\beta = 2$	<b>2.017242</b>	<b>2.001668</b>	<b>2.00365</b>
$\sigma = 0.03$	<b>0.0304859</b>	<b>0.02931701</b>	<b>0.02868814</b>
Error	<b>0.03510043</b>	<b>0.02516788</b>	<b>0.05018945</b>
Valor de la función	<b>-1475.932</b>	<b>-1477.613</b>	<b>-1476.841</b>

La **evaluación de la función objetivo sobre los parámetros reales** da como resultado -1477.472, siendo

una diferencia muy pequeña respecto a los parámetros estimados. Así observamos que estos resultados se aproximan bastante bien en este caso. La temperatura inicial influye a favor de una mejor exploración por todo el entorno, mientras que la intensidad de búsqueda nos la da la constante de enfriamiento. Con las temperaturas escogidas, para este caso, se ha visto que era suficiente ya que apenas hay diferencia, y por ello mismo puede ocurrir que para una constante de enfriamiento más alta tampoco haya una mejora notable cuando los resultados ya son buenos con constantes más bajas.

Vemos a continuación otro ejemplo donde se simulan 10 trayectorias de un proceso de Gompertz de parámetros  $m = 1.7$ ,  $\beta = 0.5$  y  $\sigma = 0.01$ , con 500 observaciones cada una, cogiendo una muestra de 50 trayectorias (se eligen de 10 en 10, es decir la primera observación, la onceava,...). La temperatura es dada por estimación de la media de incrementos de valores de parámetros evaluados por la función objetivo que van mejorando la solución de ésta en 100 repeticiones, siendo esta temperatura de 6905.103 en este caso, y el valor de la función objetivo es de -2019.472. Los resultados obtenidos son:

	$\alpha = 0.8$	$\alpha = 0.9$	$\alpha = 0.95$	$\alpha = 0.99$
$m = 1.7$	<b>1.612942</b>	<b>1.713575</b>	<b>1.701889</b>	<b>1.721642</b>
$\beta = 0.5$	<b>0.4720981</b>	<b>0.4907378</b>	<b>0.5020462</b>	<b>0.5018289</b>
$\sigma = 0.01$	<b>0.0117095</b>	<b>0.0164135</b>	<b>0.0147045</b>	<b>0.01221116</b>
Error	<b>0.2779644</b>	<b>0.6680443</b>	<b>0.4756537</b>	<b>0.2375044</b>
Valor de la función	<b>-1882.406</b>	<b>-1906.483</b>	<b>-1957.265</b>	<b>-1988.411</b>

En el siguiente ejemplo se simulan 10 trayectorias con 500 observaciones, tomando una muestra igual que en el ejemplo anterior. Los parámetros son  $m = 1.5$ ,  $\beta = 0.3$  y  $\sigma = 0.02$ , y el valor de la función objetivo para los parámetros anteriores es de -1688.669. La temperatura de la que se parte se genera igual que el ejemplo anterior, y es de 7193.21. Los resultados obtenidos son:

	$\alpha = 0.8$	$\alpha = 0.9$	$\alpha = 0.95$	$\alpha = 0.99$
$m = 1.5$	<b>1.49677</b>	<b>1.490546</b>	<b>1.505176</b>	<b>1.507665</b>
$\beta = 0.3$	<b>0.3034968</b>	<b>0.3003529</b>	<b>0.3006636</b>	<b>0.302429</b>
$\sigma = 0.02$	<b>0.01752425</b>	<b>0.02012989</b>	<b>0.02098584</b>	<b>0.2053548</b>
Error	<b>0.137547</b>	<b>0.01397355</b>	<b>0.05495492</b>	<b>0.03998086</b>
Valor de la función	<b>-1681.378</b>	<b>-1687.798</b>	<b>-1685.376</b>	<b>-1686.78</b>

### 5.2.2. Estimaciones con VNS

Se aplica el VNS en su versión descendente a partir de una solución del S.A.. El paso de búsqueda local en este algoritmo se realiza mediante el S.A., y por la forma del algoritmo se tiene asegurada una mejora (en el peor de los casos nos quedaríamos con la solución inicial introducida). Para garantizar la mejora, la solución inicial ha de ser de una aproximación de igual ó inferior precisión a la que suele generar el algoritmo de búsqueda local (S.A. en este caso), ya que si es muy superior probablemente el algoritmo devuelva la solución inicial.

A partir de la solución dada para temperatura inicial 5000, se busca mejorar ésta aplicando VNS. Por lo comentado en el anterior párrafo consideramos los mismos parámetros del S.A. que han generado la solución inicial, que son 5000 de temperatura inicial con  $\alpha = 0.95$ , y se obtiene una mejora:

VNS híbrido con S.A. con $T_0 = 5000$ , $\alpha = 0.95$	
$m = 7$	<b>7.022239</b>
$\beta = 2$	<b>2.007685</b>
$\sigma = 0.03$	<b>0.03114149</b>
Error	<b>0.04506904</b>
Valor de la función	<b>-1476.372</b>

Tomando como solución inicial la del ejemplo de 10 trayectorias con 50 observaciones de parámetros  $m = 1.7$ ,  $\beta = 0.5$  y  $\sigma = 0.01$  se obtienen los siguientes resultados:

VNS híbrido con S.A. con $T_0 = 6905.103$ , $\alpha = 0.8$	
$m = 1.7$	<b>1.682834</b>
$\beta = 0.5$	<b>0.4930662</b>
$\sigma = 0.01$	<b>0.01071211</b>
Error	<b>0.0951763</b>
Valor de la función	<b>-2011.414</b>

Tomando como solución inicial la del ejemplo de 10 trayectorias con 50 observaciones de parámetros  $m = 1$ ,  $\beta = 0.3$  y  $\sigma = 0.02$  se obtienen los siguientes resultados

VNS híbrido con S.A. con $T_0 = 7193.21$ , $\alpha = 0.8$	
$m = 1.5$	<b>1.505545</b>
$\beta = 0.3$	<b>0.3020889</b>
$\sigma = 0.02$	<b>0.01910976</b>
Error	<b>0.05517139</b>
Valor de la función	<b>-1688.814</b>

En este caso la estimación del VNS mejora a todas las anteriores.

### 5.3. ESTIMACIÓN EN EL PROCESO LOGÍSTICO CON DATOS SIMULADOS:

En los ejemplos que se dan, se las observaciones en las distintas trayectorias se realizarán simultáneamente, siendo  $t_{i_0j} = t_{i_1j} = t_j$ , para cada  $1 \leq i_0, i_1 \leq d$ , con  $t_0 \leq t_j \leq t_1$ , y los instantes de tiempo son equiespaciados, con longitud de paso  $h$ .

Ahora se han de hacer algunas consideraciones sobre la función objetivo a la que se aplicará el algoritmo:

Se quiere estimar los parámetros  $b$ ,  $c$  y  $\sigma^2$  y para ello, en vista de la forma que tiene la función, puede considerarse la reparametrización dada por

$$\eta = \frac{1}{b}$$

$$\gamma = e^{-c}$$

y así nuestra función objetivo (quitando la parte de la distribución inicial), viene dada por:

$$f(\eta, \gamma, \sigma^2) = \frac{N-d}{2} \log(\sigma^2) + \frac{1}{2h\sigma^2} \sum_{i=1}^d \sum_{j=2}^{n_i} \left( \log \frac{x_{ij}}{x_{i,j-1}} + \log \left( \frac{\eta + \gamma^{i,j-1}}{\eta + \gamma^{i,j}} \right) + \frac{\sigma^2 h}{2} \right)^2.$$

Sobre el espacio de soluciones se observa que:

1. Teniendo en cuenta que es  $b > e^{ct_0} = 1$ , entonces  $\frac{1}{b} = \eta < 1$ , de donde el entorno para  $\eta$  es  $]0, 1[$ .
2. Para  $\sigma^2$  no es recomendable la modelización del problema mediante curvas de crecimiento de tipo Gompertz cuando la varianza es grande. Por ello se acota  $\sigma \in ]0, 0.1]$ .
3. Es claro que  $\gamma = e^{-c} < 1$ , ya que  $c > 0$ .

Así ya tenemos el espacio de soluciones

$$\Omega = ]0, 1[ \times ]0, 1[ \times ]0, 0.01[$$



La *temperatura inicial* se elige observando los incrementos de la función objetivo cuando se eligen soluciones aleatorias que mejoran la solución. Se van a ver ejemplos con 3 temperaturas, para apreciar la mejora. El efecto de una mayor temperatura, en vista del algoritmo, es que el movimiento por el espacio de soluciones es mayor, ya que la probabilidad exponencial que aparece se hace mayor cuanto mayor sea esta temperatura, saltando con más frecuencia de soluciones mejoradas respecto a la anterior. Se han obtenido las siguientes temperaturas tras realizar este procedimiento 10 veces con los datos simulados en este caso: 280.7666, 254.2167, 250.4168, 246.972, 231.3082, 209.0421, 223.0629, 211.6667, 272.1505, 221.6435. Así, con las temperaturas que se van a tomar se tiene asegurada una exploración completa del entorno de soluciones.

Para la *longitud de la cadena* se sabe que es suficiente con 50, y las consideraciones sobre la temperatura inicial y constante de enfriamiento son las mismas en que en el caso del proceso Gompertz.

Se simulan 10 trayectorias de un proceso logístico de parámetros  $b = 3.5$  y  $c = 2$  con 50 observaciones cada una. Los resultados son:

	$T_0 = 5000, \alpha = 0.95$	$T_0 = 10000, \alpha = 0.95$	$T_0$ por estimación, $\alpha = 0.99$
$b = 3.5$	<b>3.616794</b>	<b>3.632264</b>	<b>3.371769</b>
$c = 2$	<b>2.085345</b>	<b>1.901275</b>	<b>2.03042</b>
$\sigma = 0.05$	<b>0.05277421</b>	<b>0.04808643</b>	<b>0.04993079</b>
Error	<b>0.1315264</b>	<b>0.2152548</b>	<b>0.05323163</b>
Valor de la función	<b>-1214.11</b>	<b>-1218.003</b>	<b>-1220.652</b>

El valor de la función evaluada en los parámetros reales es de -1221.775. Las aproximaciones en este caso quizá no sean tan buenas como las conseguidas para el Gompertz, aunque en vista de los valores que alcanza la función objetivo de cada parámetro, se puede decir que son bastante buenas.

Vemos a continuación otro ejemplo donde se simulan 10 trayectorias de un proceso logístico de parámetros  $b = 1.7$ ,  $c = 0.5$  y  $\sigma = 0.01$ , con 500 observaciones cada una, cogiendo una muestra de 50 trayectorias (se eligen de 10 en 10, es decir la primera observación, la onceava,...). La temperatura es dada por estimación de la media de incrementos de valores de parámetros evaluados por la función objetivo que van mejorando la solución de ésta en 100 repeticiones y multiplicando el valor por 50 para asegurar que sea buena, siendo esta temperatura de 9756.51 en este caso, y el valor de la función objetivo en los parámetros de -2048.262. Los resultados obtenidos son:

	$\alpha = 0.8$	$\alpha = 0.9$	$\alpha = 0.95$	$\alpha = 0.99$
$b = 1.7$	<b>1.769577</b>	<b>1.858648</b>	<b>1.572231</b>	<b>1.688239</b>
$c = 0.5$	<b>0.4720701</b>	<b>0.5160203</b>	<b>0.4983281</b>	<b>0.4666135</b>
$\sigma = 0.01$	<b>0.01002311</b>	<b>0.01337043</b>	<b>0.01184026</b>	<b>0.0105843</b>
Error	<b>0.09909815</b>	<b>0.4624059</b>	<b>0.262581</b>	<b>0.1321157</b>
Valor de la función	<b>-2036.65</b>	<b>-1974.942</b>	<b>-2012.46</b>	<b>-2022.12</b>

En el siguiente ejemplo se simulan 10 trayectorias con 500 observaciones, tomando una muestra igual que en el ejemplo anterior. Los parámetros son  $b = 1$ ,  $c = 0.2$  y  $\sigma = 0.02$ . La temperatura de la que se parte se genera igual que el ejemplo anterior, y es de 7193.21, y el valor de la función objetivo para los parámetros a estimar es de -1701.719. Los resultados obtenidos son:

	$\alpha = 0.8$	$\alpha = 0.9$	$\alpha = 0.95$	$\alpha = 0.99$
$b = 1$	<b>1.138522</b>	<b>1.042946</b>	<b>1.022885</b>	<b>1.03001</b>
$c = 0.2$	<b>0.2278102</b>	<b>0.2000265</b>	<b>0.2501893</b>	<b>0.1746215</b>
$\sigma = 0.02$	<b>0.02015811</b>	<b>0.02194636</b>	<b>0.02043843</b>	<b>0.01976222</b>
Error	<b>0.2854781</b>	<b>0.141589</b>	<b>0.2956929</b>	<b>0.1687912</b>
Valor de la función	<b>-1679.873</b>	<b>-1691.121</b>	<b>-1676.953</b>	<b>-1700.731</b>

### 5.3.1. Estimaciones con VNS

Se aplica el VNS en su versión descendente a partir de una solución del S.A.. El paso de búsqueda local en este algoritmo se realiza mediante el S.A., y por la forma del algoritmo se tiene asegurada una mejora (en el peor de los casos nos quedaríamos con la solución inicial introducida). Para garantizar la mejora, la solución inicial ha de ser de una aproximación de igual ó inferior precisión a la que suele generar el algoritmo de búsqueda local (S.A. en este caso), ya que si es muy superior probablemente el algoritmo devuelva la solución inicial.

A partir de la solución dada para temperatura inicial 5000, se busca mejorar ésta aplicando VNS. Por lo comentado en el anterior párrafo consideramos los mismos parámetros del S.A. que han generado la solución inicial, que son 5000 de temperatura inicial con  $\alpha = 0.95$ , y se obtiene una mejora:

VNS híbrido con S.A. con $T_0 = 5000$ , $\alpha = 0.95$	
$b = 3.5$	<b>3.531381</b>
$c = 2$	<b>1.915925</b>
$\sigma = 0.05$	<b>0.04852347</b>
Error	<b>0.1547861</b>
Valor de la función	<b>-1219.619</b>

En vista del valor que toma la función objetivo, observamos que se obtiene una solución bastante mejor, de hecho mejora a la aproximación dada para una temperatura inicial de 10000.

Para el ejemplo de parámetros  $b = 1.7$ ,  $c = 0.5$  y  $\sigma = 0.01$ , partiendo de una solución aleatoria cualquiera del espacio de soluciones, se han obtenido los siguientes resultados:

VNS híbrido con S.A. con $T_0 = 9756.51$ , $\alpha = 0.9$	
$b = 1.7$	<b>1.783198</b>
$c = 0.5$	<b>0.4702115</b>
$\sigma = 0.01$	<b>0.008430294</b>
Error	<b>0.2654876</b>
Valor de la función	<b>-2028.112</b>

Tomando como solución inicial la del ejemplo de 10 trayectorias con 50 observaciones de parámetros  $b = 1$ ,  $c = 0.3$  y  $\sigma = 0.02$  se obtienen los siguientes resultados

VNS híbrido con S.A. con $T_0 = 10230.16$ , $\alpha = 0.9$	
$b = 1$	<b>1.009801</b>
$c = 0.2$	<b>0.2122234</b>
$\sigma = 0.02$	<b>0.01858356</b>
Error	<b>0.1417402</b>
Valor de la función	<b>-1699.453</b>

## 5.4. CONCLUSIONES

Respecto al S.A., se observa que a mayor temperatura no necesariamente se obtienen mejores resultados. La importancia fundamental de la elección de temperatura reside en asegurarnos un recorrido lo más uniforme posible por todo el espacio de soluciones. Sí que es fundamental una constante de enfriamiento lo más alta posible, ya que ésta asegura una explotación más intensa (por ejemplo se puede ver la tabla donde se muestra el crecimiento del número de iteraciones cuando aumenta la constante de enfriamiento, que crece indefinidamente cuando ésta se acerca a 1). Generando automáticamente la temperatura, y tomando como constante de enfriamiento  $\alpha$  entre 0.9 y 0.99 nos da bastante garantía de obtener una buena aproximación.

Respecto al VNS se pueden seguir algunas estrategias:

1. La más simple, que sería partir de una solución inicial aleatoria. El problema de esta estrategia es que, aunque simple, no garantiza la mejora sobre el S.A., y el tiempo de ejecución es mayor, ya que éste será menor cuanto mejor sea la solución inicial. Para tener alguna garantía de que mejore al S.A. ha de tomarse con los mismos parámetros que éste, y siempre hay posibilidad de que la solución devuelta sea la inicial que se ha introducido.
2. Como segunda opción se puede tomar como solución inicial una estimación hecha mediante S.A., y esto garantiza la mejora de la solución (ó en el peor de los casos se obtiene la misma), por cómo se construye el algoritmo VNS. Así se reduce el tiempo de ejecución del algoritmo, y además si partimos de una solución que de por si es buena obtendremos mejores resultados.

Se ha de tener en cuenta en esta estrategia que ha de tomarse la constante de enfriamiento igual en ambos pasos (S.A. y VNS con la solución del S.A.), ya que posiblemente si en el VNS se parte de una constante de enfriamiento inferior a la del S.A., la solución final apenas mejorará a la inicial, si es que la mejora.

## Apéndice A

# DEFINICIONES Y RESULTADOS QUE SE USARÁN

### A.1. SOBRE PROCESOS DE DIFUSIÓN

La clase de procesos estudiados en este trabajo, como se verá, se encuentran dentro de la amplia clase de procesos llamada **procesos de Markov**, que son aquellos procesos  $\{X(t) / t \in T\}$  que verifican **la propiedad de Markov**:

$$\forall t_0 < t_1 < \dots < t_n < t_{n+1} \in T \quad P(X(t_{n+1}) \leq x_{n+1} \mid X(t_n) \leq x_n, \dots, X(t_0) = x_0) = P(X(t_{n+1}) \leq x_{n+1} \mid X(t_n) \leq x_n).$$

Esta propiedad indica el comportamiento del proceso en el instante  $t = t_n$ , que sólo depende del estado del proceso en la etapa anterior, pudiendo prescindir de la información del resto de etapas.

Las probabilidades  $P(X(t) \in A \mid X(s) = x)$  cuando  $s < t$  instantes, y  $x$  un valor del espacio de estados se conoce como *probabilidades de transición*, y puede verse como una función de cuatro variables  $A, t, x, s$  que verifica:

1.  $P(A, t; \cdot, s)$  medible sobre el espacio de estados  $\forall s < t$ ,  $\forall A \in \mathcal{B}$   $\sigma$ -álgebra de Borel.
2.  $P(\cdot, t; x, s)$  es una medida de probabilidad sobre  $\mathcal{B}$ , para todo  $x$  en el espacio de estados y  $\forall s < t$ .
3. *Ecuación de Chapman-Kolmogorov*:  $\forall s \leq u \leq t$ ,  $\forall A \in \mathcal{B}$  y  $\forall x$  se verifica

$$P(A, t; x, s) = \int P(B, t; y, u) P(dy, u; x, s)$$

donde la integral se extiende sobre el espacio de estados.

Se considera el caso de procesos en tiempo continuo y espacio de estados continuo;  $f(x, t|y, s)$  las densidades de transición asociadas a las probabilidades de transición apropiadas.

Pueden reescribirse las anteriores ecuaciones en términos de las densidades

*Propiedad de Markov:*

$$f(x_n, t_n | x_1, t_1; \dots; x_{n-1}, t_{n-1}) = f(x_n, t_n | x_{n-1}, t_{n-1}) \quad \forall t_1 < \dots < t_n \in T$$

*Ecuación de Chapman-Kolmogorov:*

$$f(x, t | y, s) = \int f(x, t | z, \tau) f(z, \tau | y, s) dz \quad \forall x, y, \quad \forall s \leq \tau \leq t$$

Indican la transición de  $y$  en el instante  $s$  a  $x$  en el instante  $t$ , pasando por los estados  $z$  en  $\tau$  cuando  $s \leq \tau \leq t$ .

Con la propiedad de Markov se relaciona la distribución finito-dimensional, con la distribución inicial y las transiciones:

$$f(x_0, t_0; \dots; x_n, t_n) = f(x_0, t_0) \prod_{i=1}^n f(x_i, t_i; x_{i-1}, t_{i-1})$$

y un caso interesante es cuando  $P(A, t; x, s) = P(A, t + h; x, s + h) \quad \forall x, \forall A \in \mathcal{A}, \forall s \leq t, \forall h \in \mathbb{R}$ ; diciéndose que las probabilidades de transición son estacionarias y el proceso de Markov se dice **homogéneo en el tiempo**.

**Propiedades A.1.1** (de procesos de Markov). 1.  $\{X(t) / t \in T\}$  proceso de Markov,  $g$  función medible-Borel con inversa  $\Rightarrow \{g(X(t)) / t \in T\}$  es un proceso de Markov.

2. Si  $\{X(t) / t \in T\}$  proceso con incrementos independientes para el cual existen densidades finito-dimensionales  $\Rightarrow$  es un proceso de Markov.

La última propiedad será usada para probar que los procesos que se van a considerar son de Markov.

### A.1.1. Ecuaciones de Kolmogorov y de Fokker-Planck en procesos de difusión

Las ecuaciones de Kolmogorov y Fokker-Planck son:

$$\frac{\partial f(x, t | y, s)}{\partial t} = -\frac{\partial}{\partial x} [A_1(x, t) f(x, t | y, s)] + \frac{1}{2} \frac{\partial^2}{\partial x^2} [A_2(x, t) f(x, t | y, s)] \quad (\text{A.1})$$

$$\frac{\partial f(x, t | y, s)}{\partial s} + A_1(y, s) \frac{\partial f(x, t | y, s)}{\partial y} + \frac{A_2(y, s)}{2} \frac{\partial^2 f(x, t | y, s)}{\partial y^2} = 0 \quad (\text{A.2})$$

Los procesos de difusión verifican las ecuaciones de Kolmogorov. Se citan algunos resultados que muestran la relación entre los procesos de difusión y las ecuaciones de Kolmogorov y Fokker-Planck:

**Teorema A.1.1.** Sea  $\{X(t) / t_0 \leq t \leq T\}$  un proceso de difusión. Supongamos que para cada  $(x, t), F(x, t | y, s)$  es dos veces derivable respecto de  $y$ , siendo dichas derivadas continuas y acotadas. Entonces  $F(x, t | y, s)$  es derivable respecto a  $s$  y verifica la ecuación atrasada de Kolmogorov (A.2) con la condición  $\lim_{s \uparrow t} F(x, t | y, s) = \lim_{s \uparrow t} P[X(t) \leq x | X(s) = y] = \begin{cases} 1 & x \geq y \\ 0 & x < y \end{cases}$ .

En caso de existir densidades de transición, se verifica la ecuación atrasada para ellas con condición inicial

$$\lim_{s \uparrow t} f(x, t | y, s) = \delta(x - y)$$

ya que  $\delta$  es la derivada de la función de distribución  $F$ .

**Teorema A.1.2.** Sea  $\{X(t) / t_0 \leq t \leq T\}$  un proceso de difusión. Supongamos que existen  $\frac{\partial f(x, t | y, s)}{\partial t}$ ,  $\frac{\partial [A_1(x, t)f(x, t | y, s)]}{\partial x}$  y  $\frac{\partial^2 [A_2(x, t)f(x, t | y, s)]}{\partial x^2}$  y son continuas. Entonces  $f(x, t | y, s)$  verifica la ecuación adelantada de Kolmogorov o ecuación de Focker-Planck (A.1) con la condición  $\lim_{t \downarrow s} f(x, t | y, s) = \delta(x - y)$ .

En la obtención de las ecuaciones diferenciales se han hecho suposiciones acerca de las probabilidades de transición, pero cuando no se conocen éstas, pueden venir dadas en términos de los momentos infinitesimales:

**Teorema A.1.3.** *Suponamos que los momentos infinitesimales  $A_1$  y  $A_2$  verifican, para todo valor  $x$  del espacio de estados y  $\forall t \in [t_0, T]$ , las siguientes condiciones:*

1. *Existen unas constantes positivas  $\sigma_0$  y  $k$  tales que*

a)  $|A_1(x, t)| \leq k\sqrt{1+x^2}.$

b)  $0 < \sigma_0 \leq \sqrt{A_2(x, t)} \leq k\sqrt{1+x^2}.$

2. *(Condición de Hölder). Existen constantes positivas  $\gamma$  y  $k$  tales que*

a)  $|A_1(x, t) - A_1(y, t)| \leq k|x - y|^\gamma.$

b)  $|\sqrt{A_2(x, t)} - \sqrt{A_2(y, t)}| \leq k|x - y|^\gamma.$

*Entonces se verifica:*

1. *La ecuación atrasada tiene una única solución sujeta a la condición frontera del  $T^a$  anterior. Además, para  $t > s$ ,  $F(x, t|y, s)$  es derivable respecto de  $x$ , admitiendo densidad, luego también verificará la ecuación atrasada con condición frontera del tipo delta de Dirac.*

2. *Existe un proceso de Markov  $\{X(t) / t \in [t_0, T]\}$  con trayectorias continuas, que verifica las condiciones de proceso de difusión y que tiene por función de distribución de transición  $F(x, t|y, s)$ .*

3. *Si, además, las condiciones del enunciado son cumplidas por  $\frac{\partial A_1(x, t)}{\partial x}$ ,  $\frac{\partial A_2(x, t)}{\partial x}$  y  $\frac{\partial^2 A_2(x, t)}{\partial x^2}$  entonces la función  $f(x, t|y, s) = \frac{\partial F(x, t|y, s)}{\partial x}$  es la única solución fundamental de la ecuación adelantada.*

4. *Si  $\gamma = 1$ , entonces  $f(x, t|y, s)$  es la densidad de transición de la única solución de la ecuación integral estocástica*

$$X(t) = X(t_0) + \int_{t_0}^t A_1(X(s), s) ds + \int_{t_0}^t \sqrt{A_2(X(s), s)} dW(s).$$

Sobre ecuaciones diferenciales estocásticas será usado el siguiente resultado:



**Teorema A.1.4.** Consideremos la ecuación diferencial estocástica

$$dX(t) = a(X(t), t)dt + b(X(t), t)dW(t)$$

$$X(0) = x_0, \quad t_0 \leq t \leq T < \infty$$

donde  $W(t)$  representa el proceso de Wiener estándar y  $x_0$  es una variable aleatoria independiente de  $W(t) - W(t_0)$  para  $t \geq t_0$ . Supongamos que las funciones  $a$  y  $b$  están definidas y son medibles en  $[t_0, T] \times \mathbb{R}$  y verifican las siguientes condiciones: existe  $K > 0$  constante tal que

1. **Condición de Lipschitz:**  $|a(x, t) - a(y, t)| + |b(x, t) - b(y, t)| \leq K|x - y|$ ,  $\forall t \in [t_0, T]$ ,  $\forall x, y \in \mathbb{R}$ .

2. **Restricción sobre el crecimiento:**  $|a(x, t)|^2 + |b(x, t)|^2 \leq K(1 + |x|^2)$ ,  $\forall t \in [t_0, T]$ ,  $\forall x \in \mathbb{R}$ .

Entonces la ecuación inicial del enunciado tiene una única solución en  $[t_0, T]$  y con valores en  $\mathbb{R}$ , continua con probabilidad uno, que satisface la condición inicial; esto es, si  $X(t)$  e  $Y(t)$  son soluciones de dicha ecuación con igual valor inicial  $x_0$ , entonces

$$P\left(\sup_{t_0 \leq t \leq T} |X(t) - Y(t)| > 0\right) = 0.$$

**Teorema A.1.5** (Fórmula de Itô). Sea  $X(t)$  un proceso con diferencial

$$dX(t) = a(t)dt + b(t)dW(t)$$

con  $a, b \in \mathcal{H}_2[0, T]$ , y sea  $f(x, t)$  una función continua con derivadas parciales continuas. Entonces

$$\begin{aligned} df(X(t), t) &= \left( \frac{\partial f(X(t), t)}{\partial t} + a(t) \frac{\partial f(X(t), t)}{\partial x} + \frac{b(t)^2}{2} \frac{\partial^2 f(X(t), t)}{\partial x^2} \right) dt + b(t) \frac{\partial f(X(t), t)}{\partial x} dW(t) = \\ &= \left( \frac{\partial f(X(t), t)}{\partial t} + \frac{b(t)^2}{2} \frac{\partial^2 f(X(t), t)}{\partial x^2} \right) dt + \frac{\partial f(X(t), t)}{\partial t} dX(t) \end{aligned}$$

Esta regla puede generalizarse, en caso de que se tenga  $a(X(t), t)$ ,  $b(X(t), t)$  a:

$$df(X(t), t) = \left( \frac{\partial f(X(t), t)}{\partial t} + a(X(t), t) \frac{\partial f(X(t), t)}{\partial x} + \frac{b(X(t), t)^2}{2} \frac{\partial^2 f(X(t), t)}{\partial x^2} \right) dt + b(X(t), t) \frac{\partial f(X(t), t)}{\partial x} dW(t)$$

**Teorema A.1.6.** Consideremos la ecuación diferencial estocástica anterior, con  $a$  y  $b$  funciones determinísticas, y con solución

$$X(t) = x_0 + \int_{t_0}^t a(s)ds + \int_{t_0}^t b(s)dW(s),$$

siendo  $x_0$  una constante. Entonces

$$X(t) \sim \mathcal{N}_1 \left[ x_0 + \int_0^t a(s)ds ; \int_0^t b^2(s)ds \right].$$

## A.2. SOBRE LA DISTRIBUCIÓN NORMAL MULTIVARIANTE Y LOGARÍTMICO-NORMAL

Si  $\mathbf{Y} \sim \mathcal{N}_p[\boldsymbol{\mu}, \Sigma]$ , con  $\Sigma_{p \times p}$  simétrica definida positiva y  $\boldsymbol{\mu} \in \mathbb{R}^p$ , distribución *normal multivariante*, su densidad es de la forma

$$f(x) = \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} \exp \left( -\frac{1}{2} (x - \boldsymbol{\mu})' \Sigma^{-1} (x - \boldsymbol{\mu}) \right)$$

con  $x \in \mathbb{R}^p$ .  $E[\mathbf{Y}] = \boldsymbol{\mu}$ , y media y varianza

$$E[\mathbf{Y}] = \boldsymbol{\mu}$$

$$\text{Cov}[\mathbf{Y}] = \Sigma$$

y respecto a las distribuciones marginales y condicionadas, particionando en la forma  $\mathbf{Y} = (\mathbf{Y}'_{(1)} | \mathbf{Y}'_{(2)})$  con  $\mathbf{Y}'_{(1)}$   $q$ -dimensional, y  $\mathbf{Y}'_{(2)}$   $(p - q)$ -dimensional, suponiendo además las particiones correspondientes a  $\boldsymbol{\mu}$  y  $\Sigma$ :

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_{(1)} \\ \boldsymbol{\mu}_{(2)} \end{pmatrix} \quad ; \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

se tienen las siguientes

**Propiedades A.2.1.** En las condiciones anteriores:

1.  $\mathbf{Y}_{(1)} \sim \mathcal{N}_q[\boldsymbol{\mu}_{(1)}, \Sigma_{11}]$  y  $\mathbf{Y}_{(2)} \sim \mathcal{N}_{(p-q)}[\boldsymbol{\mu}_{(2)}, \Sigma_{22}]$ .
2. Si  $\Sigma_{12} = \Sigma_{21} = 0$  entonces  $\mathbf{Y}_{(1)}$  y  $\mathbf{Y}_{(2)}$  son independientes.
3.  $\mathbf{Y}_{(1)} - \Sigma_{12}\Sigma_{22}^{-1}\mathbf{Y}_{(2)} \sim \mathcal{N}_q[\boldsymbol{\mu}_{(1)} - \Sigma_{12}\Sigma_{22}^{-1}\boldsymbol{\mu}_{(2)}; \Sigma_{11.2}]$  donde  $\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ .
4.  $\mathbf{Y}_{(2)}$  y  $\mathbf{Y}_{(1)} - \Sigma_{12}\Sigma_{22}^{-1}\mathbf{Y}_{(2)}$  son independientes.
5.  $\mathbf{Y}_{(1)} | \mathbf{Y}_{(2)} = y_{(2)} \sim \mathcal{N}_q[\boldsymbol{\mu}_{(1)} + \Sigma_{12}\Sigma_{22}^{-1}(y_{(2)} - \boldsymbol{\mu}_{(2)}); \Sigma_{11.2}]$ .

Se usará la última propiedad del resultado anterior.

Dado  $\mathbf{Y} \sim \mathcal{N}_p[\boldsymbol{\mu}, \Sigma]$ , se puede considerar un vector  $p$ -dimensional  $\mathbf{X}$ , tal que  $\log(\mathbf{X}) = (\log(X_1), \dots, \log(X_p))$  componentes que verifiquen que  $\log(X_i) = \log(Y_i)$  para  $i = 1 \dots, p$ . En esta situación se dice que la distribución

de  $\mathbf{X}$  es *logarítmico normal multivariante*, denotada por  $\Lambda_p[\mu; \Sigma]$ . Tiene densidad con expresión

$$f(x) = \prod_{i=1}^p \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}} x_i} \exp \left( -\frac{1}{2} (\log(x) - \mu)' \Sigma^{-1} (\log(x) - \mu) \right)$$

con momentos dados por

$$E[X_1^{s_1} \dots X_p^{s_p}] = \exp \left\{ s' \mu + \frac{1}{2} s' \Sigma s \right\}$$

cuando  $s = (s_1, \dots, s_p)'$ , de donde en particular se tiene

$$\begin{aligned} E[X_i] &= e^{\mu_i + \frac{\sigma_i^2}{2}} \\ E[X_i^2] &= e^{2\mu_i + 2\sigma_i^2} & \implies & \text{Var}[X_i] = e^{2\mu_i + \sigma_i^2} (e^{\sigma_i^2} - 1) \\ E[X_i X_j] &= e^{\mu_i + \mu_j + \frac{\sigma_i^2}{2} + \frac{\sigma_j^2}{2} + \sigma_{ij}} & \implies & \text{Cov}(X_i, X_j) = e^{\mu_i + \mu_j + \frac{\sigma_i^2}{2} + \frac{\sigma_j^2}{2} + \sigma_{ij}} (e^{\sigma_{ij}} - 1) \end{aligned}$$

Otras propiedades a destacar son:

1.  $\mathbf{X} = (\mathbf{X}'_{(1)} | \mathbf{X}'_{(2)})$  con  $\mathbf{X}_{(1)}$   $q$ -dimensional, entonces  $\mathbf{X}_{(1)} \sim \Lambda_q(\mu_{(1)}, \Sigma_{11})$ .
2.  $\mathbf{X}_{(2)} | X_{(1)} = x_{(1)} \sim \Lambda_{(p-q)}[\mu_{(2)} + \Sigma_{21} \Sigma_{11}^{-1} (\log(x_{(1)}) - \mu_{(1)}); \Sigma_{22.1}]$ .
3. Si  $X \sim \Lambda_1[\mu, \sigma]$ , entonces

a) Su moda viene dada por:

$$\text{Moda}(t) = \exp(\mu - \sigma^2).$$

b) El cuantil de orden  $\alpha$  viene dado por:

$$P_\alpha(X) = \exp(\mu + z_\alpha \sigma)$$

donde  $z_\alpha$  es el punto crítico de la distribución normal estándar que deja una probabilidad acumulada a su izquierda igual a  $\alpha$ .



## Apéndice B

# CÓDIGOS DE R

### B.1. SIMULACIÓN DE PROCESOS GOMPERTZ Y LOGÍSTICO

Se comienza simulando el proceso de Wiener mediante:

```
wiener<-function(n,tra,h)
{
  wiener<-array(0,c(n,tra))
  for(i in 1:tra) wiener[,i]<-cumsum(c(0,rnorm(n-1,0,sqrt(h))))
  print(wiener)
}
```

#### B.1.1. Simulación del proceso de Gompertz

Se empieza definiendo la función

$$x(t) = \frac{m}{\beta} (e^{-\beta t_0} - e^{-\beta t})$$

dada por

```
gompertz<-function(t0,t,m,beta,sigma)
{
  m*(exp(-beta*t0)-exp(-beta*t))/beta
}
```

Ahora se define una función que genera tra trayectorias simuladas de un proceso de Gompertz:

```
sim.gompertz<-function(t0,t1,x0,m,beta,sigma,tra,n,Ini,m0,s0)
{
  h<-abs(t1-t0)/n

  t<-rep(0,n+1)
```

```

for(i in 0:n)
  t[i+1]<-t0+h*i

if(Ini==0)
  h<-abs(t1-t0)/n

t<-rep(0,n+1)

for(i in 0:n)
  t[i+1]<-t0+h*i

if(Ini==0)
  total<-x0*exp(sigma*wiener(n+1,tra,h)+gompertz(t0,t,m,beta,sigma)-sigma^2*(t-t0)/2)

if(Ini==1)
  total<-exp(m0+s0^2/2+sigma*wiener(n+1,tra,h)+gompertz(t0,t,m,beta,sigma)-sigma^2*(t-t0)/2)

total
}

```

### B.1.2. Simulación del proceso logístico

Análogamente se definen las siguientes funciones para simular el proceso logístico

```

logistico<-function(t0,t,b,c)
log((1+b*exp(-c*t0))/(1+b*exp(-c*t)))

```

La simulación del proceso logístico viene dada por:

```

sim.logistico<-function(t0,h,x0,b,c,sigma,tra,n,Ini,m0,s0)
{
  t<-rep(0,n+1)

  for(i in 0:n)
    t[i+1]<-t0+h*i

  if(Ini==0)
    total<-x0*exp(sigma*wiener(n+1,tra,h)+logistico(t0,t,b,c)-sigma^2*(t-t0)/2)

  if(Ini==1)
    total<-exp(m0+s0^2/2+sigma*wiener(n+1,tra,h)+logistico(t0,t,b,c)-sigma^2*(t-t0)/2)
}

```

```
total
}
```

Puede ser útil la siguiente función para extraer muestras de trayectorias simuladas:

```
submuestra<-function(x,k)
{
  i<-2
  total<-x[1,]
  while(k*i<=dim(x)[1])
  {
    total<-rbind(total,x[k*i,])
    i<-i+1
  }
  total
}
```

y otra para el cálculo del error relativo:

```
error<-function(x,estim)
{
  error<-0
  for(i in 1:length(x))
    error<-error+abs(x[i]-estim[i])/x[i]
  error
}
```

## B.2. CÓDIGO DEL SIMULATED ANNEALING EN R ADAPTADO AL PROBLEMA DE ESTIMACIÓN EN EL MODELO DE GOMPERTZ

Primero se define una función llamada `entorno.gompertz` dependiente de dos parámetros: matriz (en la práctica se corresponde con trayectorias de un proceso de Gompertz), y un real positivo  $t_0$  que denota el tiempo inicial. La función devuelve una matriz  $3 \times 3$  cuyas 2 primeras columnas determinan el entorno de la solución, y la tercera la condición inicial elegida. El código sería

```
entorno.gompertz<-function(x,t0)
{
  ini<-rep(0,p)

  ini[2]<-runif(1)
```

```

ini[3]<-runif(1)*0.01

cotassup<-rep(0,d)

for(k in 1:d)
  cotassup[k]<-max(x[,k])/x[k,1]

eizdo<-log(min(cotassup))
edcho<-log(max(cotassup))

ini[1]<-entorno[1,1]+(entorno[1,2]-entorno[1,1])*runif(1)

ent1<-c(eizdo,edcho)
ent2<-c(0,1)
ent3<-c(0,0.01)

cbind(rbind(rbind(ent1,ent2),ent3),ini)
}

```

La temperatura puede establecerse mediante la siguiente función:

```

temperatura.gompertz<-function(x,t0,h,n,d)
{
  ent<-entorno.gompertz(x,t0)[,1:2]
  solaleatoria1<-entorno.gompertz(x,t0)[,3]
  tempini<-0
  k<-1

  for(i in 1:100)
  {
    solaleatoria2<-entorno.gompertz(x,t0)[,3]

    if((f.gompertz(solaleatoria1,x,h,n,d)-f.gompertz(solaleatoria2,x,h,n,d))<0)
    {
      k<-k+1
      tempini<-tempini-(f.gompertz(solaleatoria1,x,h,n,d)-f.gompertz(solaleatoria2,x,h,n,d))
    }
    solaleatoria1<-solaleatoria2
  }
  tempini<-tempini/k
}

```



```
tempini
}
```

La función objetivo en este caso es:

```
f.gompertz<-function(param,x,h,n,d)
{
  N<-n*d
  tiempos<-rep(0,n+1)
  for(i in 0:n)
    tiempos[i+1]<-t0+i*h

  suma<-0
  for(i in 1:d)
  {
    for(j in 2:n)
      suma<-suma+(log(x[j,i]/x[j-1,i])+param[1]*param[2]^tiempos[j-1]*...
        ...*(param[2]^h-1)+param[3]*h/2)^2
  }

  ((N-d)/2)*log(param[3])+suma/(2*h*param[3])
}
```

El algoritmo del S.A. para el proceso de Gompertz escrito en R sería:

```
simann.gompertz<-function(temp,alpha,x,h,n,d)
{
  ini<-entorno.gompertz(x,t0,d)[,3]

  while(temp>0.1)
  {
    for(k in 1:50)
    {
      sol<-entorno.gompertz(x,t0,d)[,3]

      delta<-f.gompertz(sol,x,h,n,d)-f.gompertz(ini,x,h,n,d)
      if(delta<=0)
        ini<-sol
      else
      {
        if(exp(-delta/temp)>runif(1))
          ini<-sol
      }
    }
  }
}
```

```

    }
  }
  temp<-temp*alpha
}
ini
}

```

Se reescribe el mismo algoritmo, pero dependiente del entorno, ya que ésta versión del algoritmo se aplica en el VNS que requiere especificar el entorno en cada etapa.

```

simann.gompertz.entorno<-function(ini,temp,alpha,x,h,n,d,entorno)
{
  while(temp>0.1)
  {
    for(k in 1:50)
    {
      sol<-entorno[,1]
      for(i in 1:3)
      {
        while(sol[i]<=entorno[i,1]||sol[i]>=entorno[i,2])
          sol[i]<-entorno[i,1]+(entorno[i,2]-entorno[i,1])*(runif(1)-runif(1))/2
      }

      delta<-f.gompertz(sol,x,h,n,d)-f.gompertz(ini,x,h,n,d)
      if(delta<=0)
        ini<-sol
      else
      {
        if(exp(-delta/temp)>runif(1))
          ini<-sol
      }
    }
    temp<-temp*alpha
  }
  ini
}

```

A partir de la anterior función, puede definirse otra que no dependa de la temperatura inicial, que vendría dada por:

```

sa.gompertz<-function(x,t0,h,n,d)
simann.gompertz(temperatura.gompertz(x,t0,h,n,d),0.99,x,t0,h,n,d)

```

Lo que se ha hecho es integrar la temperatura dentro de la función, asegurando que ésta es adecuada. Como se ha visto en algunos ejemplos una elección de la temperatura baja da lugar a resultados no muy buenos. Tampoco depende de la constante de enfriamiento, que por defecto se ha dejado en 0.99.

### B.3. CÓDIGO DEL SIMULATED ANNEALING EN R ADAPTADO AL PROBLEMA DE ESTIMACIÓN EN EL MODELO LOGÍSTICO

En este caso no es necesaria una función para definir el entorno. Para elegir una temperatura adecuada se puede usar la siguiente función:

```
temperatura.logistico<-function(x,h,n,d)
{
  ent<-rbind(rbind(c(0,1),c(0,1)),c(0,1))
  solaleatoria1<-c(runif(1),runif(1),runif(1)*0.01)
  tempini<-0
  k<-1

  for(i in 1:100)
  {
    solaleatoria2<-c(runif(1)*(ent[1,2]-ent[1,1]),runif(1)*(ent[2,2]-ent[2,1]),...)

    if((f.logistico(solaleatoria1,x,h,n,d)-f.logistico(solaleatoria2,x,h,n,d))<0)
    {
      k<-k+1
      tempini<-tempini-(f.logistico(solaleatoria1,x,h,n,d)-...
                        ....f.logistico(solaleatoria2,x,h,n,d))
    }
    solaleatoria1<-solaleatoria2
  }
  tempini<-tempini/k
  tempini
}
```

La función objetivo viene dada por:

```
f.logistico<-function(param,x,h,n,d)
{
  N<-n*d
  tiempos<-rep(0,n+1)
  for(i in 1:n)
```

```

    tiempos[i+1]<-t0+i*h

    suma<-0
    for(i in 1:d)
    {
      for(j in 2:n)
      suma<-suma+(log(x[j,i]/x[j-1,i])-log((param[1]+param[2]^tiempos[j-1])/...
      .../(param[1]+param[2]^tiempos[j]))+param[3]*h/2)^2
    }

    ((N-d)/2)*log(param[3])+suma/(2*h*param[3])
  }

```

El algoritmo del S.A. para el proceso Logístico escrito en R sería:

```

simann.logistico<-function(temp,alpha,x,h,n,d)
{
  ini<-c(runif(1),runif(1),runif(1)*0.01)

  while(temp>0.1)
  {
    for(k in 1:50)
    {
      sol<-c(runif(1),runif(1),runif(1)*0.01)

      delta<-f.logistico(sol,x,h,n,d)-f.logistico(ini,x,h,n,d)
      if(delta<=0)
        ini<-sol
      else
      {
        if(exp(-delta/temp)>runif(1))
          ini<-sol
      }
    }
    temp<-temp*alpha
  }
  ini
}

```

Se reescribe el mismo algoritmo, pero dependiente del entorno, ya que ésta versión del algoritmo se aplica en el VNS que requiere especificar el entorno en cada etapa.

```

simann.logistico.entorno<-function(ini,temp,alpha,x,h,n,d,entorno)
{
  while(temp>0.1)
  {
    for(k in 1:50)
    {
      sol<-rep(0,3)
      for(i in 1:3)
        sol[i]<-entorno[i,1]+(entorno[i,2]-entorno[i,1])*runif(1)

      delta<-f.logistico(sol,x,h,n,d)-f.logistico(ini,x,h,n,d)
      if(delta<=0)
        ini<-sol
      else
      {
        if(exp(-delta/temp)>runif(1))
          ini<-sol
      }
    }
    temp<-temp*alpha
  }
  ini
}

```

A partir de la anterior función, puede definirse otra que no dependa de la temperatura inicial, que vendría dada por:

```

sa.logistico<-function(x,h,n,d)
simann.logistico(temperatura.gompertz(x,h,n,d),0.99,x,t0,h,n,d)

```

Lo que se ha hecho es integrar la temperatura dentro de la función, asegurando que ésta es adecuada. Como se ha visto en algunos ejemplos una elección de la temperatura baja da lugar a resultados no muy buenos. Tampoco depende de la constante de enfriamiento, que por defecto se ha dejado en 0.99.

## B.4. CÓDIGO DEL VNS PARA DISTINTAS VERSIONES:

Se implementan distintas versiones del VNS, de las que se habló en el segundo capítulo. Salvo la versión VNS reducida, el algoritmo de búsqueda local usado ser el S.A., obteniéndose lo que se llama un algoritmo híbrido:

### B.4.1. VNS reducida

Es la versión menos eficiente ya que la nueva solución es elegida aleatoriamente. No se ha usado para la estimación, aunque el código sería el siguiente:

```
vns.gompertz.reducida<-function(kmax,f)
{
  aini<-eizdo+(edcho-eizdo)/2
  bini<-0.5
  sigma2ini<-0.005
  centro<-c(aini,bini,sigma2ini)
  ini<-centro

  paso<-c((edcho-eizdo)/(2*kmax),1/(2*kmax),0.01/(2*kmax))

  k<-1
  while(k<=kmax)
  {
    if(k==1)
      ent<-cbind(centro-paso,centro+paso)

    sol<-c(runif(1)*(ent[1,2]-ent[1,1]),runif(1)*(ent[2,2]-ent[2,1]),runif(1)*
      *(ent[3,2]-ent[3,1]))

    if(f(sol)<f(ini))
    {
      ini<-sol
      k<-1
    }
    else
      k<-k+1

    ent<-ent+cbind(-paso,paso)
  }
  sol
}
```

### B.4.2. VNS descendente

Este es el algoritmo que se usa, ya que es un híbrido con el S.A., y en el paso de buscar una solución óptima es éste último algoritmo el que se usa mediante la función `simann.entorno` antes construida:

**1. Para el proceso tipo Gompertz:**

```

vns.descendente.gompertz<-function(p,kmax,temperatura,coldconst,x,t0,h,n,d)
{
  entorno<-entorno.gompertz(x,t0)[,1:2]

  centro<-rep(0,p)
  for(i in 1:p)
    centro[i]<-centro[i]+(entorno[i,1]+entorno[i,2])/2
  ini<-centro

  paso<-rep(0,p)
  for(i in 1:p)
    paso[i]<-(entorno[i,2]-entorno[i,1])/(2*kmax)

  k<-1
  while(k<=kmax)
  {
    if(k==1)
      ent<-cbind(centro-paso,centro+paso)

    sol<-simann.gompertz.entorno(ini,temperatura,coldconst,x,h,n,d,ent)
    if(f.gompertz(sol,x,h,n,d)<f.gompertz(ini,x,h,n,d))
    {
      ini<-sol
      k<-1
    }
    else
      k<-k+1

    ent<-ent+cbind(-paso,paso)
  }
  sol
}

```

**2. Para el proceso logístico:**

```

vns.descendente.logistico<-function(p,kmax,temperatura,coldconst,x,h,n,d)
{
  entorno<-rbind(rbind(c(0,1),c(0,1)),c(0,1))

```

```
centro<-rep(0,p)
for(i in 1:p)
  centro[i]<-centro[i]+(entorno[i,1]+entorno[i,2])/2
ini<-centro

paso<-rep(0,p)
for(i in 1:p)
  paso[i]<-(entorno[i,2]-entorno[i,1])/(2*kmax)

k<-1
while(k<=kmax)
{
  if(k==1)
    ent<-cbind(centro-paso,centro+paso)

  sol<-simann.logistico.entorno(ini,temperatura,coldconst,x,h,n,d,ent)
  if(f.logistico(sol,x,h,n,d)<f.logistico(ini,x,h,n,d))
  {
    ini<-sol
    k<-1
  }
  else
    k<-k+1

  ent<-ent+cbind(-paso,paso)
}
sol
}
```



# Bibliografía

- [1] R. Gutiérrez, P. Román-Román, D. Romero, and F. Torres-Ruiz, “Inference and first-passage-times for the lognormal diffusion process with exogenous factors: Application to modelling in economics,” *APPLIED STOCHASTIC MODELS IN BUSINESS and INDUSTRY*, 1999.
- [2] A. Laird, “Dynamics of tumor growth,” *Division of Biological and Medical Research, Argonne National Laboratory (Illinois)*, 1964.
- [3] G. Albano, V. Giorno, P. Román-Román, D. Romero, and F. Torres-Ruiz, “Inferring the effect of therapy on tumors hoving stochastic gompertzian growth,” *Elsevier*, 2011.
- [4] D. Mitra, F. Romeo, and A. Sangiovanni-Vincentelli, “Convergence and finite time behavior of simulated annealing,” *Adv. Appl. Probab.*, 1986.
- [5] P. Román-Román and F. Torres-Ruiz, “Modelling logistic growth by a new diffusion process: Application to biological systems,” *Elsevier*, 2012.
- [6] P. Román, D. Romero, and F. Torres-Ruiz, “Estimating the parameters of a gompertz diffusion process by means of simulated annealing,” *Elsevier*, 2011.
- [7] R. Gutiérrez, N. Rico, P. Román-Román, and F. Torres-Ruiz, “Approximate and generalized confidence bands for some parametric functions of the lognormal diffusion process with exogenous factors,” *Scientiae Mathematicae Japonicae Online*, 2006.
- [8] R. Gutiérrez, P. Román-Román, D. Romero, J. Serrano, and F. Torres-Ruiz, “A new gompertz diffusion process with application to tumor growth,” *Elsevier*, 2006.
- [9] T. Modis, “Us nobel laureates: Logistic growth versus volterra-lotka,” *Elsevier*, 2010.
- [10] B. Gallhager, “Peak oil analyzed with a logistic function and idealized hubbert curve,” *Energ. Policy*, 2011.
- [11] A. Giovanis and C. Skiadas, “A new modeling approach investigating the diffusion speed of mobile telecommunication services in eu-15,” *Springer Science+Business Media*, 2006.
- [12] M. Gendreau and J.-V. Potvin. Springer.
- [13] I. Boussaïd, J. Lepagnot, and P. Siarry, “A survey on optimization metaheuristics,” *Elsevier*, 2013.

- [14] L. Goldstein and M. Waterman, "Neighborhood size in the simulated annealing algorithm," *Mathl. Comput. Modelling*, 1988.
- [15] J. Franco and P. Tabares, "Aplicación del simulated annealing al problema de las n reinas," *Scientia et Technica*, 2005.
- [16] K. Dowsland and B. Díaz, "Heuristic design and fundamentals of the simulated annealing," *Mathl. Comput. Modelling*, 2003.
- [17] B. Abbasi, A. Hamid, E. Jahromi, J. Arkat, and M. Hosseinkouchack, "Estimating the parameters of weibull distribution using simulated annealing algorithm," *Springer Science+Business Media*, 2006.
- [18] S. Kirkpatrick, C. Gelatt, and M. Vecchi, "Optimization by simulated annealing," *American Association for the Advancement of Science*, 1983.
- [19] L. Ingber, "Simulated annealing: practice vs theory," *Mathl. Comput. Modelling*, 1993.
- [20] W. Ben-Ameur, "Computing the initial temperature of simulated annealing," *Kluwer Academic Publishers*, 2003.
- [21] P. Hansen, N. Mladenovic, and J. Moreno, "Búsqueda de entorno variable," *Revista Iberoamericana de Inteligencia Artificial*, 2003.
- [22] N. Mladenovic, "Variable neighborhood search," *Serbian Academic of Science*.
- [23] N. Mladenovic, M. Drazic, V. Kovacevic-Vujcic, and M. Cangalovic, "General variable neighborhood search for the continuous optimization," *Elsevier*, 2006.
- [24] N. Zoraghi, B. Abbasi, S. Taghi, and M. Abdi, "Estimating the four parameters of the burr iii distribution using a hybrid method of variable neighborhood search and iterated local search algorithms," *Elsevier*, 2012.
- [25] P. Hansen and N. Mladenovic, "Variable neighborhood search: Principles and applications," *Elsevier*, 2012.
- [26] B. Abbasi, A. Hamid, M. Abdi, and Y. Faize, "A hybrid variable neighborhood search and simulated annealing algorithm to estimate the three parameters of the weibull distribution," *Elsevier*, 2010.
- [27] P. Román-Román, D. Romero, and F. Torres, "Apuntes de la asignatura 'cálculo y modelización estocástica. procesos de difusión' y 'aplicaciones de los modelos de difusión en fenómenos de crecimiento en ciencias medioambientales y economía'," *Universidad de Granada*, 2013.