Un nuevo proceso de difusión para modelizar patrones de crecimiento mixto exponencial-Gompertz

Gutiérrez, R., Rico, N., Román, P., Romero, D. y Torres, F. *Universidad de Granada*

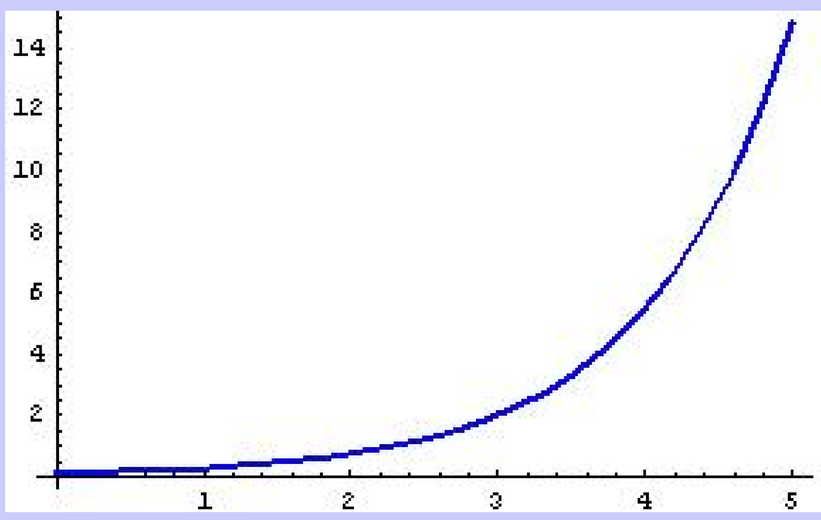
Resumen

En este artículo se propone un nuevo proceso de difusión no homogéneo, el cual permite modelizar patrones de crecimiento que combinan una forma exponencial con una sigmoidal de tipo Gompertz, expresado en términos de variables continuas en el tiempo. La principal innovación de este proceso es que engloba al lognormal y al Gompertz, permitiendo modelizar datos que presentan un comportamiento mixto entre exponencial y Gompertz. Para su obtención, llevada a cabo por diferentes métodos, se ha buscado un proceso asociado a una curva resultante de la combinación de las mencionadas anteriormente. Además se realiza un estudio amplio del mismo, incluyendo sus características principales. Finalmente, se lleva a cabo un estudio inferencial basado en muestreo discreto y se incluye una aplicación sobre datos simulados, mostrándose su utilidad para diferentes tipos de comportamientos.

Propuesta de una curva de crecimiento mixto exponencial-Gompertz

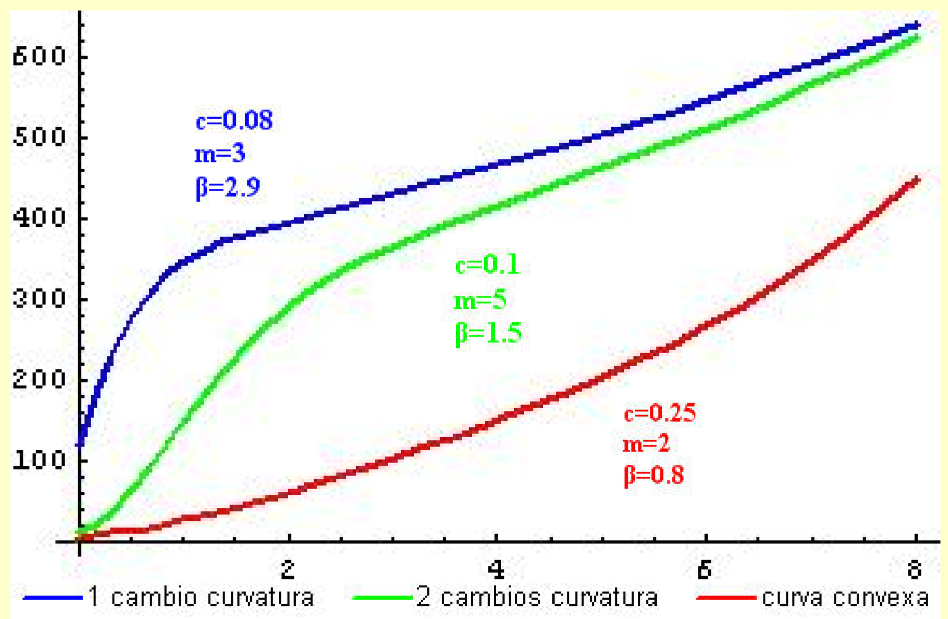
Curva de crecimiento exponencial

$$f_1(t) = e^{c(t-t_0)}, \quad t \geqslant t_0,\ c \in \mathbb{R}^+$$



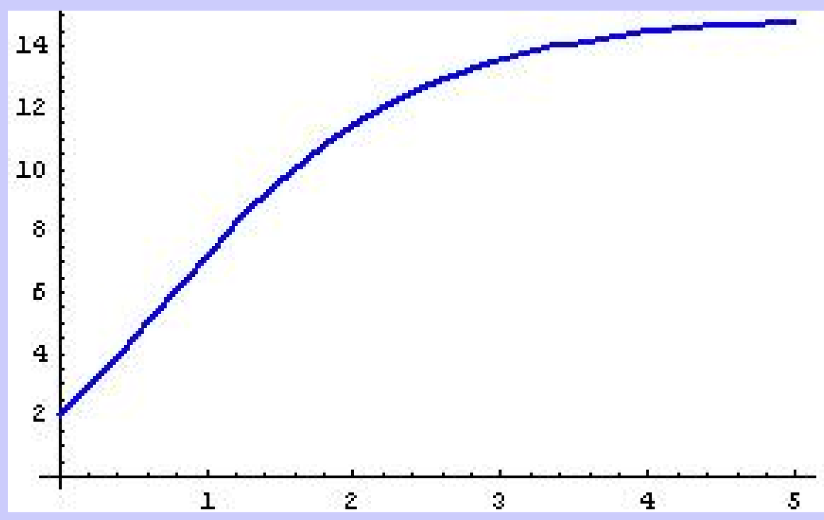
Curva de crecimiento mixto exponencial-Gompertz

$$f(t) = x_0 \exp\left(-\frac{m}{\beta}\left(e^{-\beta t} - e^{-\beta t_0}\right) + c(t-t_0)\right), \quad t \geqslant t_0,\ m > \beta > 0,\ c \in \mathbb{R}^+$$



Curva de crecimiento Gompertz

$$f_2(t) = x_0 \exp\left(-\frac{m}{\beta}\left(e^{-\beta t} - e^{-\beta t_0}\right)\right), \quad t \geqslant t_0,\ m > \beta > 0$$



Proceso mixto lognormal-Gompertz

Condiciones para la obtención

Modificando la ecuación de Fokker-Planck del proceso lognormal $(m \to me^{-\beta t} + c)$ se obtiene

$$\frac{\partial f}{\partial t} = -\frac{\partial}{\partial x}\left[\left(me^{-\beta t} + c\right) x f\right] + \frac{\sigma^2}{2}\frac{\partial^2}{\partial x^2}[x^2 f]$$

cuya solución cuando $\sigma^2 \to 0$, con la condición inicial $\lim_{t\to t_0} f(x, t/x_0, t_0) = \delta(x - x_0)$, es

$$f(x, t/x_0, t_0) = \delta\left(x - \underbrace{x_0 \exp\left(-\frac{m}{\beta}\left(e^{-\beta t} - e^{-\beta t_0}\right) + c(t-t_0)\right)}_{\downarrow}\right)$$

Curva exponencial-Gompertz.

Definición

$\{X(t), t \geqslant t_0\}$ proceso de difusión con valores en $\mathbb{R}^+$ y momentos infinitesimales

$$A_1(x,t) = \left(me^{-\beta t} + c\right) x$$
$$A_2(x,t) = \sigma^2 x^2$$

donde $m > \beta > 0$, $c \in \mathbb{R}^+$, $\sigma > 0$ y distribución inicial

$$X(t_0) \sim \Lambda(\mu_0, \sigma_0^2), \quad \mu_0 \in \mathbb{R},\ \sigma_0^2 \geqslant 0.$$

Características del proceso

- $X(t)|X(s) = y \sim \Lambda_1\left[\ln y - \frac{m}{\beta}\left(e^{-\beta t} - e^{-\beta s}\right) + \left(c - \frac{\sigma^2}{2}\right)(t-s); (t-s)\sigma^2\right]$
- $E[X(t)] = E[X(t_0)] \exp\left(\frac{m}{\beta}\left[e^{-\beta t_0} - e^{-\beta t}\right] + c(t-t_0)\right)$
- $Moda[X(t)] = Moda[X(t_0)] \exp\left(\frac{m}{\beta}\left[e^{-\beta t_0} - e^{-\beta t}\right] + c(t-t_0) - \frac{3\sigma^2}{2}(t-t_0)\right)$
- $Cuantil_\alpha[X(t)] = Cuantil_\alpha[X(t_0)] \exp\left(\frac{m}{\beta}\left[e^{-\beta t_0} - e^{-\beta t}\right] + c(t-t_0)\right)$
  $\times \exp\left(-\frac{\sigma^2}{2}(t-t_0) + z_{1-\alpha}\left[\sqrt{\sigma^2(t-t_0) + Var[\ln(X(t_0))]} - \sqrt{Var[\ln(X(t_0))]}\right]\right)$,
  donde $z_{1-\alpha}$ es el valor de una normal estándar que deja a la derecha una probabilidad de $1-\alpha$.

Estimación

Estimadores Máximo Verosímiles

$$\widehat{a} = \frac{1}{\widehat{b}^h - 1}\frac{A_{1,\widehat{b}}A^*_{3,\widehat{b}} - A^*_{1,\widehat{b}}A_{3,\widehat{b}}}{A^*_{1,\widehat{b}}A_{2,\widehat{b}} - A_{1,\widehat{b}}A^*_{2,\widehat{b}}} \quad \widehat{c} = \frac{\widehat{\sigma}^2}{2} - \frac{1}{h}\frac{A^*_{2,\widehat{b}}A_{3,\widehat{b}} - A_{2,\widehat{b}}A^*_{3,\widehat{b}}}{A^*_{1,\widehat{b}}A_{2,\widehat{b}} - A_{1,\widehat{b}}A^*_{2,\widehat{b}}}$$

$$\widehat{\sigma}^2 = \frac{1}{h(k-d)}\left(\frac{A_{4,\widehat{b}}A^*_{1,\widehat{b}}A_{2,\widehat{b}} - A_{4,\widehat{b}}A_{1,\widehat{b}}A^*_{2,\widehat{b}} + A_{1,\widehat{b}}A^*_{3,\widehat{b}}A_{3,\widehat{b}} - A^*_{1,\widehat{b}}A^2_{3,\widehat{b}}}{A^*_{1,\widehat{b}}A_{2,\widehat{b}} - A_{1,\widehat{b}}A^*_{2,\widehat{b}}}\right.$$
$$\left.+\frac{A_{3,\widehat{b}}A^*_{2,\widehat{b}}A_{5,\widehat{b}} - A^*_{3,\widehat{b}}A_{2,\widehat{b}}A_{5,\widehat{b}}}{A^*_{1,\widehat{b}}A_{2,\widehat{b}} - A_{1,\widehat{b}}A^*_{2,\widehat{b}}}\right) \quad \widehat{\mu}_1 = \frac{1}{d}\sum_{i=1}^{d}\ln x_{i1} \quad \widehat{\sigma}_1^2 = \frac{1}{d}\sum_{i=1}^{d}(\ln x_{i1} - \widehat{\mu}_1)^2$$

donde

$$A_{1,b} = \sum_{i=1}^{d}\sum_{j=2}^{n_i} b^{t_{ij-1}} \quad A_{2,b} = \sum_{i=1}^{d}\sum_{j=2}^{n_i} b^{2t_{ij-1}} \quad A_{3,b} = \sum_{i=1}^{d}\sum_{j=2}^{n_i} b^{t_{ij-1}}\ln\frac{x_{ij}}{x_{ij-1}}$$
$$A^*_{1,b} = \sum_{i=1}^{d}\sum_{j=2}^{n_i} t_{ij-1}b^{t_{ij-1}} \quad A^*_{2,b} = \sum_{i=1}^{d}\sum_{j=2}^{n_i} t_{ij-1}b^{2t_{ij-1}} \quad A^*_{3,b} = \sum_{i=1}^{d}\sum_{j=2}^{n_i} t_{ij-1}b^{t_{ij-1}}\ln\frac{x_{ij}}{x_{ij-1}}$$
$$A_{4,b} = \sum_{i=1}^{d}\sum_{j=2}^{n_i} \ln^2\frac{x_{ij}}{x_{ij-1}} \quad A_{5,b} = \sum_{i=1}^{d}\sum_{j=2}^{n_i} \ln\frac{x_{ij}}{x_{ij-1}}$$

y $\widehat{b}$ verifica la ecuación

$$A_{4,\widehat{b}} + \widehat{a}(\widehat{b}^h - 1)A_{3,\widehat{b}} + \left(\frac{\widehat{\sigma}^2}{2} - \widehat{c}\right)hA_{5,\widehat{b}} - \widehat{\sigma}^2 h(k-d) = 0,$$

la cual debe ser tratada por métodos numéricos.

Función de verosimilitud

La función de densidad de transición puede escribirse como

$$f(x_{ij}, t_{ij}|x_{ij-1}, t_{ij-1}) = \frac{1}{x_{ij}\sqrt{2\pi\sigma^2(t_{ij} - t_{ij-1})}} \times$$
$$\times \exp\left(-\frac{1}{2}\frac{\left[\ln\frac{x_{ij}}{x_{ij-1}} + a\left(b^{t_{ij}} - b^{t_{ij-1}}\right) - \left(c - \frac{\sigma^2}{2}\right)(t_{ij} - t_{ij-1})\right]^2}{(t_{ij} - t_{ij-1})\sigma^2}\right)$$

donde $a = \dfrac{m}{\beta}$, $b = e^{-\beta}$ y $k = \sum_{i=1}^{d} n_i$.

Si suponemos que $X(t_{i1}) \rightsquigarrow \Lambda_1[\mu_1; \sigma_1^2]$, $i = 1, \ldots, d$, el logaritmo de la verosimilitud, asociada a una muestra de d trayectorias observadas en instantes t_{ij} $(j = 1, \cdots, n_i; i = 1, \cdots, d)$, es

$$\ln L_{x_{ij}}(\mu_1, \sigma_1^2, a, b, c, \sigma^2) = -\frac{d}{2}\ln\sigma_1^2 - \frac{k}{2}\ln(2\pi) - \frac{k-d}{2}\ln\sigma^2 - \sum_{i=1}^{d}\ln x_{i1} -$$
$$-\frac{1}{2\sigma_1^2}\sum_{i=1}^{d}[\ln x_{i1} - \mu_1]^2 - \sum_{i=1}^{d}\sum_{j=2}^{n_i}\ln x_{ij} - \frac{1}{2}\sum_{i=1}^{d}\sum_{j=2}^{n_i}\ln(t_{ij} - t_{ij-1}) -$$
$$-\frac{1}{2}\sum_{i=1}^{d}\sum_{j=2}^{n_i}\frac{\left[\ln\frac{x_{ij}}{x_{ij-1}} + a\left(b^{t_{ij}} - b^{t_{ij-1}}\right) + \left(\frac{\sigma^2}{2} - c\right)(t_{ij} - t_{ij-1})\right]^2}{(t_{ij} - t_{ij-1})\sigma^2}.$$

Método alternativo de estimación

1. Se toma i=0 y se parte de un valor inicial de b, comprendido entre 0 y 1, $b^{(0)}$. Se calcula a partir de él $\beta^{(0)} = -\ln b^{(0)}$.
2. Se obtienen los valores $a_{b^{(i)}}$, $m^{(i)}$, $\sigma^2_{b^{(i)}}$ y $c_{b^{(i)}}$ mediante los estimadores máximo verosímiles y la ecuación
   $$m^{(i)} = a_{b^{(i)}}\beta^{(i)}.$$
3. Teniendo en cuenta que los instantes en los que se alcanzan los puntos de inflexión de las trayectorias son los mismos en todas ellas, se puede obtener un nuevo valor del parámetro $\beta^{(i+1)}$, mediante la resolución de la ecuación
   $$e^{-\beta^{(i+1)}t_1} = \frac{\beta^{(i+1)} - 2c_{b^{(i)}} - \sqrt{\beta^{(i+1)}(\beta^{(i+1)} - 4c_{b^{(i)}})}}{2m^{(i)}}$$
   donde t_1 es el instante de tiempo en el cual se produce el cambio de curvatura de cóncava a convexa y, a partir del valor obtenido para $\beta^{(i+1)}$, se puede obtener $b^{(i+1)}$.
4. Se hace i=i+1, se vuelve al paso 2 y se repite el proceso hasta conseguir la convergencia, entendiendo la misma en términos de una tolerancia, entre dos etapas sucesivas, en términos de b.

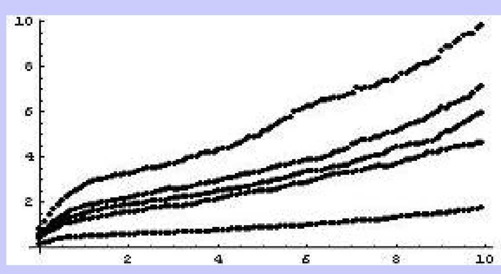
Aplicación a datos simulados

Simulación de trayectorias

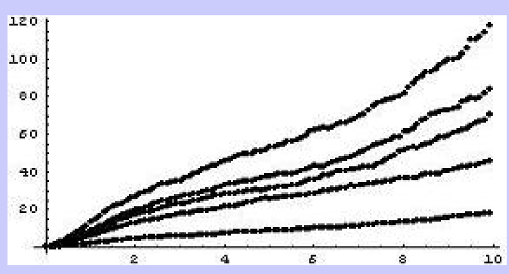
Trayectorias del proceso mixto lognormal-Gompertz simuladas mediante los algoritmos derivados de la resolución numérica de ecuaciones diferenciales estocásticas (Kloeden et al. (1994)), en 1000 instantes de tiempo, tomados desde $t_0 = 0$ con saltos de amplitud 0.01.

Caso 1
m=3
β=2.9
c=0.15
σ^2=0.001

Caso 2
m=5
β=1.5
c=0.15
σ^2=0.001



Caso de 1 sólo cambio de curvatura

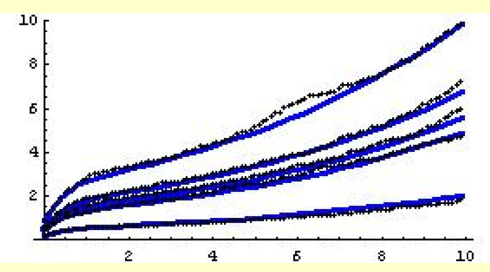


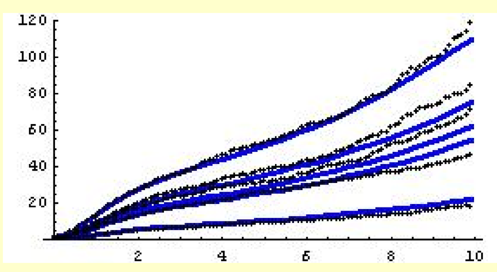
Caso de 2 cambios de curvatura

Estimaciones (Newton-Raphson)

Caso 1
$\widehat{m}$=3.19982
$\widehat{\beta}$=3.10781
$\widehat{c}$=0.141785
$\widehat{\sigma}^2$=0.000875252

Caso 2
$\widehat{m}$=5.07102
$\widehat{\beta}$=1.52482
$\widehat{c}$=0.153518
$\widehat{\sigma}^2$=0.00101747





Estimaciones (Método iterativo)

Caso 1
$\widehat{m}$=3.15551
$\widehat{\beta}$=3.0181
$\widehat{c}$=0.140202
$\widehat{\sigma}^2$=0.000926178

Caso 2
$\widehat{m}$= 5.03648
$\widehat{\beta}$= 1.5006
$\widehat{c}$=0.150364
$\widehat{\sigma}^2$= 0.000902611