

Estimación de la función de Gompertz considerando la correlación entre los errores. Dos aplicaciones: COVID-19 y Biogás

Estimation of the Gompertz function considering the correlation between the errors. Two applications: COVID-19 and Biogas

Batún José, Cool Rubén, Guerrero Ernesto, Pantí Henry

Recepción: 23/11/2022 **Aceptación:** 07/06/2023 **Publicación:** 31/07/2023

Abstract This article aims to illustrate the two-stage procedure to estimate the parameters of the Gompertz function when the errors are correlated. Throughout the article, the mathematical development of the properties of the Gompertz function is presented, as well as its interpretation. Similarly, the procedure is applied in two contexts: infections by COVID-19 in Yucatan and the production of Biogas during the anaerobic digestion of organic waste. In both cases, the Box-Jenkins methodology is used to determine the parameters of the time series that models the errors.

Keywords Correlated Errors, Generalized Least Square, Gompertz Function, Time Series.

José Luis Batún Cutz, Ph.D.

Docente, Universidad Autónoma de Yucatán, UADY, Facultad de Matemáticas, Mérida, Yucatán, e-mail: jbatun@correo.uady.mx,

 <https://orcid.org/0000-0001-7290-6993>

Rubén Alejandro Cool Padilla, M.Sc.

Técnico Académico, Universidad Autónoma de Yucatán, UADY, Facultad de Matemáticas, Mérida, Yucatán, e-mail: ruben.cool@correo.uady.mx,

 <https://orcid.org/0000-0001-9028-5940>

Ernesto Antonio Guerrero Lara, M.Sc.

Docente, Universidad Autónoma de Yucatán, UADY, Facultad de Matemáticas, Mérida, Yucatán, e-mail: ernesto.guerrero@correo.uady.mx,

 <https://orcid.org/0000-0002-9438-7474>

Henry Gaspar Pantí Trejo, Ph.D.

Docente, Universidad Autónoma de Yucatán, UADY, Facultad de Matemáticas, Mérida, Yucatán, e-mail: henry.panti@correo.uady.mx,

 <https://orcid.org/0000-0003-2416-5240>

Resumen El presente artículo tiene como objetivo ilustrar el procedimiento en dos etapas para estimar los parámetros de la función de Gompertz cuando los errores están correlacionados. A lo largo del artículo se presenta el desarrollo matemático de las propiedades que tiene la función de Gompertz así como su interpretación. De igual forma se aplica el procedimiento en dos contextos: contagios por COVID-19 en Yucatán y en la producción de Biogás durante la digestión anaeróbica de desperdicios orgánicos. En ambos casos se hace uso de la metodología de Box-Jenkins para determinar y estimar los parámetros de la serie de tiempo que modela a los errores.

Palabras Claves Errores Correlacionados, Función de Gompertz, Mínimos Cuadrados Generalizados, Series de Tiempo.

1 Introducción

Modelar matemáticamente algún fenómeno de interés nos permite comprender su comportamiento y en consecuencia tomar mejores decisiones.

Un modelo sugerido cuando la variable de estudio presenta un crecimiento a través del tiempo de tal forma que a largo plazo se aproxima a un valor fijo finito es el relacionado con la función de Gompertz, principalmente porque posee propiedades matemáticas que en el contexto de los problemas a estudiar, los parámetros que la definen y las funciones asociadas como las derivadas, suelen ser claras y de fácil interpretación. La gráfica de la función de Gompertz pertenece a la familia de curvas sigmoidales que tienen como característica principal la forma de S, por lo que en el principio son convexas y posterior al punto de inflexión, cóncavas, es decir, antes del punto de inflexión la velocidad de crecimiento aumenta a través del tiempo y posterior al punto de inflexión disminuye. Otra característica es que tienen una asíntota horizontal.

La popularidad de la función de Gompertz recae en la utilidad que ha mostrado a lo largo de estos años para modelar el crecimiento con respecto al tiempo de alguna característica de una población. Por ejemplo, en Przybylski y García (2004) se utiliza un modelo de Gompertz para modelar el peso de peces; en Cooper (2005) se plantea un modelo de Gompertz para medir el peso de la cabeza de avestruces; en Torrealba, Conde, y Hernández (2020) se presenta un modelo de Gompertz que describe el crecimiento en el tiempo del número acumulado de casos positivos por COVID-19 en México. Como éstos se pueden encontrar otros ejemplos adicionales de fenómenos modelados mediante la función de Gompertz.

En los ejemplos mencionados, el ajuste a la función de Gompertz se hace sin considerar la aleatoriedad de los datos, más aún, cuando se grafica la curva ajustada y los datos se observan variaciones que son aleatorias, estas variaciones representan los errores. En este artículo se presenta y se aplica, en dos conjuntos de datos, un método que considera la estructura de aleatoriedad

de las variaciones para hacer un nuevo ajuste a la función de Gompertz, en particular se consideran modelos cuando las variaciones se comportan de acuerdo a una serie de tiempo $Ar(p)$.

Adicionalmente se demuestran e interpretan las propiedades matemáticas de la función de Gompertz relacionadas con la monotonía de la primera y segunda derivada así como de sus puntos críticos; se proporciona una aproximación al crecimiento por unidad de tiempo de la función de Gompertz, se presenta un criterio con un gráfico de dispersión que sugiere que la función de Gompertz se puede ajustar a los datos y se da una sugerencia de valores iniciales para los parámetros cuando el ajuste se realiza por métodos numéricos.

La estructura del artículo es la siguiente: en la Sección 2 se presenta la función de Gompertz, sus propiedades, una aproximación al crecimiento por unidad de tiempo y un criterio gráfico que sugiere usar la función de Gompertz; en la Sección 3 se presenta un análisis e interpretaciones con las derivadas de orden superior (segunda y tercera derivada) de la función de Gompertz; en la Sección 4 se explica la metodología para ajustar la función de Gompertz considerando la estructura de los errores que se tienen en una primera etapa; en la Sección 5 se muestran dos aplicaciones de esta metodología en el contexto de COVID-19 y Biogás ajustando los parámetros del modelo a través de métodos numéricos.

2 La Función de Gompertz

Desde el punto de vista matemático, la función de Gompertz presenta propiedades que hacen frecuente su aplicación como modelo y que son explicadas en este artículo. Es importante señalar que algunas de las propiedades que aquí se detallan son mencionadas en diversos artículos sin proporcionar el procedimiento matemático que las justifica.

Definition 1. Sean α, β, γ reales positivos. La función de Gompertz es la función real valuada definida por

$$f(t) = \alpha e^{-\beta e^{-\gamma t}}, \quad t \geq 0. \quad (1)$$

Considerando que la función de Gompertz es utilizada para modelar crecimiento de poblaciones, a partir de la definición, se tienen las siguientes propiedades e interpretaciones

1. Siempre es positiva y el valor inicial está dado por $f(0) = \alpha e^{-\beta}$. Este valor es usualmente interpretado como el tamaño de la población inicial;
2. Es estrictamente creciente, por lo que es adecuada para modelar el crecimiento de poblaciones;
3. Tiene una asíntota horizontal dada por el valor de α , esto debido a que

$$\lim_{t \rightarrow \infty} f(t) = \alpha$$

y representa el tamaño al que se va aproximando la población cuando ha pasado mucho tiempo.

En la Figura 1 se ilustra la gráfica de la función de Gompertz y se pueden visualizar las propiedades 1, 2 y 3, y el punto de inflexión que será definido y calculado más adelante.

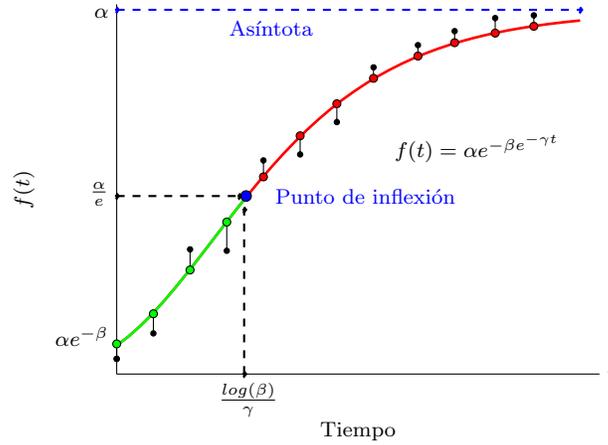


Figura 1: Ejemplo de la Gráfica de la Función de Gompertz

Observación 1

- Aunque estrictamente tenemos como condición que $\beta > 0$, para propósitos de este trabajo asumiremos que $\beta > \frac{3 + \sqrt{5}}{2}$ pues en caso contrario, algunos puntos de interés que mencionaremos a lo largo del artículo se estarían presentando en valores de $t < 0$, lo cual no sucede cuando los datos son recolectados desde que se empieza a presentar el crecimiento de la característica que se está modelando.

La derivada de la función de Gompertz también es empleada en la modelación, ésta permite conocer la tasa instantánea del crecimiento de la función y también aproxima al crecimiento de la función de Gompertz por unidad de tiempo, tal como será demostrado en la Proposición 1. La primera derivada de la función de Gompertz está dada por la expresión siguiente

$$f'(t) = \gamma \beta e^{-\gamma t} f(t), \quad t \geq 0. \quad (2)$$

Por otra parte, la segunda derivada de la función de Gompertz satisface

$$f''(t) = \gamma^2 \beta e^{-\gamma t} (\beta e^{-\gamma t} - 1) f(t).$$

Se puede observar que $f''(t)$ se anula cuando $\beta e^{-\gamma t} - 1 = 0$, es decir, cuando $\mathbf{t}_1 = \frac{\log(\beta)}{\gamma}$. Además, ya que $\beta e^{-\gamma t} - 1 > 0$ para $t \in [0, \mathbf{t}_1)$ y $\beta e^{-\gamma t} - 1 < 0$ para $t \in (\mathbf{t}_1, \infty)$, se asegura que en \mathbf{t}_1 , la función $f'(\mathbf{t}_1)$ alcanza un máximo global. El valor de t donde la segunda derivada se anula corresponde a la abscisa del punto de inflexión de la función original, de esta forma $(\mathbf{t}_1, \alpha e^{-1})$ es el punto de inflexión para la función de Gompertz. En la Figura 1 se indica con color verde el intervalo donde la gráfica es cóncava y de color rojo donde es convexa.

Si no se considera la Observación 1, en particular si $0 < \beta \leq 1$, entonces el punto de inflexión se presentaría antes del tiempo $t = 0$, es decir, desde antes de empezar a recolectar la información, los datos ya alcanzaron su velocidad de crecimiento máxima, lo cuál no sería adecuado al modelar el crecimiento de una característica de la población de estudio.

Adicionalmente, la ordenada del punto de inflexión es $\frac{\alpha}{e} \approx 0.368\alpha$, esto significa que en el punto de inflexión el tamaño de población es aproximadamente 36.8% del que se tendrá a largo plazo.

El siguiente resultado muestra que el crecimiento por unidad de tiempo de la función de Gompertz, se puede aproximar a través de su derivada.

Proposition 1. *Sea $f(t)$ la función de Gompertz. Cuando $\gamma \rightarrow 0$, la función $f(t)$ satisface la siguiente convergencia uniforme*

$$\sup_{t \geq 0} |f(t+1) - f(t) - f'(t)| \rightarrow 0.$$

Demostración. Observe que

$$\frac{f(t+1)}{f(t)} = e^{(1-e^{-\gamma})\beta e^{-\gamma t}}, \quad (3)$$

de donde

$$f(t+1) - f(t) = \left(e^{(1-e^{-\gamma})\beta e^{-\gamma t}} - 1 \right) f(t). \quad (4)$$

Sean $\lambda = \frac{1 - e^{-\gamma}}{\gamma}$ y $r(t) = \gamma \beta e^{-\gamma t}$, entonces

- $0 < \lambda \leq 1$, ya que $1 - e^{-\gamma} < \gamma$ para $\gamma > 0$,
- $\beta^{-1} \lambda r(t) \leq 1$, debido a que $\beta^{-1} \lambda r(t) = (1 - e^{-\gamma}) e^{-\gamma t}$,
- $\beta^{-1} r(t) = \gamma e^{-\gamma t} \leq \gamma$ para $\gamma > 0$.

Las desigualdades anteriores y el hecho que $|f(t)| \leq \alpha$ y $|e^{\tau t} - 1| \leq \tau(e^t - 1)$ para toda $t \geq 0$ y $0 < \tau \leq 1$, junto con la ecuación (4), implican lo siguiente

$$\begin{aligned}
|f(t+1) - f(t)| &\leq \alpha |e^{\beta^{-1}\lambda r(t)\beta} - 1| \\
&\leq \alpha \beta^{-1} \lambda r(t) (e^\beta - 1) \\
&\leq \alpha \beta^{-1} r(t) (e^\beta - 1) \\
&\leq \alpha \gamma (e^\beta - 1).
\end{aligned}$$

Por otro lado, por la ecuación (2) se obtiene $|f'(t)| = |r(t)f(t)| \leq \alpha\beta\gamma$. Por lo tanto,

$$\begin{aligned}
|f(t+1) - f(t) - f'(t)| &\leq |f(t+1) - f(t)| + |f'(t)| \\
&\leq \alpha\gamma(e^\beta - 1) + \alpha\beta\gamma \\
&\leq K\gamma,
\end{aligned}$$

donde $K = \alpha(e^\beta + \beta - 1)$. De aquí se sigue el resultado. ■

Observación 2

- A partir de la ecuación (3) se tiene

$$\left(\frac{f(t+1)}{f(t)} \right)^{1/(e^{-\gamma}-1)} = \frac{f(t)}{\alpha}.$$

De lo anterior se deduce

$$\log(f(t+1)) = e^{-\gamma} \log(f(t)) + (1 - e^{-\gamma}) \log \alpha. \quad (5)$$

Esta identidad asegura que los puntos $(\log f(t), \log f(t+1))$ están situados sobre una línea recta con pendiente $e^{-\gamma}$. Si $\gamma \rightarrow 0$, la recta mencionada tiende a ser la recta identidad. La utilidad de la ecuación (5) recae en que permite realizar un análisis exploratorio para determinar si la gráfica de una función de Gompertz se ajusta de acuerdo a un conjunto de datos, es decir, dada una muestra $\{(t_1, x_1), (t_2, x_2), \dots, (t_n, x_n)\}$ donde x_t representa el dato observado al tiempo t , entonces con base en la ecuación (5) se permite evaluar de forma empírica si el conjunto de datos tiene un comportamiento acorde a la gráfica de una función de Gompertz. Si el conjunto de datos observados $\mathbf{x} = [x_1, x_2, \dots, x_n]'$ satisface la ecuación (1), entonces de la ecuación (5), se esperaría que

$$\log(x_{t+1}) \approx e^{-\gamma} \log(x_t) + (1 - e^{-\gamma}) \log \alpha, \quad (6)$$

esto es, que la gráfica de dispersión de $\log(x_t)$ vs $\log(x_{t+1})$ presente una tendencia similar a la de una línea recta.

3 Velocidad y Aceleración de Crecimiento de la Función de Gompertz

La primera derivada de la función de Gompertz representa la velocidad de crecimiento y su comportamiento es tal que al inicio aumenta hasta alcanzar una velocidad máxima, posteriormente la velocidad de crecimiento disminuye y se estabiliza a una velocidad de crecimiento pequeña. La importancia de conocer los puntos de inflexión de la primera derivada es que permiten localizar los puntos donde se presentan cambios críticos en la velocidad de crecimiento. El comportamiento de la velocidad de crecimiento y sus puntos críticos se vinculan con las funciones $f''(t)$ (aceleración de crecimiento) y $f'''(t)$.

La tercera derivada de la función de Gompertz está dada por la siguiente expresión

$$f'''(t) = \gamma^3 \beta e^{-\gamma t} f(t) \left[(\beta e^{-\gamma t})^2 - 3\beta e^{-\gamma t} + 1 \right].$$

La función $f'''(t)$ se anula cuando $(\beta e^{-\gamma t})^2 - 3\beta e^{-\gamma t} + 1 = 0$, es decir, cuando $\mathbf{t}_2 = -\log(\frac{3+\sqrt{5}}{2\beta})/\gamma$ y $\mathbf{t}_3 = -\log(\frac{3-\sqrt{5}}{2\beta})/\gamma$. La Observación 1 garantiza que t_2 y t_3 son positivos. Además, $f'''(t) > 0$ para $t \in (0, \mathbf{t}_2)$, $f'''(t) < 0$ para $t \in (\mathbf{t}_2, \mathbf{t}_3)$ y $f'''(t) > 0$ para $t \in (\mathbf{t}_3, \infty)$. De lo anterior se deduce que en el punto \mathbf{t}_2 la función $f''(\mathbf{t}_2)$ alcanza un máximo y en el punto \mathbf{t}_3 , la función $f''(\mathbf{t}_3)$ alcanza un mínimo global. Los valores de t donde la tercera derivada se anula corresponden a las abscisas de los puntos de inflexión de la función $f'(t)$. Por lo tanto, los puntos de inflexión para la función $f'(t)$ son

$$\left(\mathbf{t}_2, \frac{\alpha\gamma(3+\sqrt{5})}{2} e^{-\frac{3+\sqrt{5}}{2}} \right) \text{ y } \left(\mathbf{t}_3, \frac{\alpha\gamma(3-\sqrt{5})}{2} e^{-\frac{3-\sqrt{5}}{2}} \right). \quad (7)$$

En la Figura 2 se presentan las gráficas de la función de Gompertz y sus primeras tres derivadas. El tiempo \mathbf{t}_1 corresponde a la abscisa del punto máximo de la primera derivada y del punto de inflexión de la función de Gompertz. Se puede demostrar algebraicamente que $\mathbf{t}_2 < \mathbf{t}_1 < \mathbf{t}_3$ bajo la condición de la Observación 1. Por otra parte, en el intervalo $[0, \mathbf{t}_2]$ la función $f'(t)$, que representa la velocidad de crecimiento, aumenta ya que su aceleración $f'''(t)$ es positiva, luego en el intervalo $(\mathbf{t}_2, \mathbf{t}_3]$ la aceleración de $f'(t)$ es negativa lo que quiere decir que $f'(t)$ crecerá en una parte pero luego comenzará a decrecer. El tiempo donde se presenta este cambio es en \mathbf{t}_1 , mientras que en el tiempo \mathbf{t}_2 y \mathbf{t}_3 es donde cambia de signo $f'''(t)$.

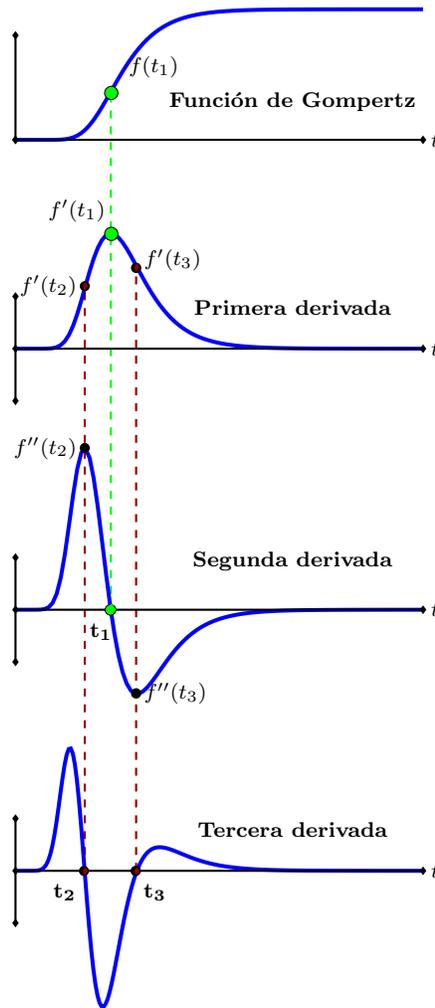


Figura 2: Gráfica de la Función de Gompertz y sus derivadas

4 Estimación en dos etapas para los parámetros del modelo de Gompertz

Para ajustar una función de Gompertz se asume que la tasa de crecimiento se incrementa de forma monótona hasta que alcanza un nivel máximo y después decrece en forma monótona asintóticamente. El modelo de Gompertz es de carácter determinista, esto significa que en caso de tener dos poblaciones con las mismas condiciones iniciales, el comportamiento de ambas debe ser

el mismo. Sin embargo, se sabe que en la realidad no sucede de esta forma, ya que se presentan variaciones que se consideran aleatorias y que no son capturadas por estos modelos deterministas.

En el gráfico de dispersión de una serie de datos se puede observar la aleatoriedad presente en los fenómenos tal y como se ilustró en la Figura 1. Las distancias entre cada dato y estimación dada por la curva ajustada son los errores que, cuando los datos son recolectados en forma secuencial, suelen estar correlacionados; por lo que estos errores pueden ser modelados mediante una serie de tiempo. En Seber y Wild (2003) se presenta una metodología que estima los parámetros para un modelo de Gompertz en dos etapas considerando la aleatoriedad de los errores de un primer ajuste, cuando éstos se comportan de acuerdo con una serie de tiempo $AR(p)$. A continuación se describe esta metodología para la estimación de los parámetros del modelo en el contexto de la función de Gompertz $f(t, \Theta)$ siendo $\Theta = (\alpha, \beta, \gamma)$ el vector de parámetros.

Supongamos que tenemos n valores de la variable de interés $\mathbf{x} = [x_1, \dots, x_n]'$ que se recolectaron en los tiempos $t = [t_1, \dots, t_n]$, es decir, se tiene una muestra $\{(t_1, x_1), (t_2, x_2), \dots, (t_n, x_n)\}$. El procedimiento es el siguiente

Etapas 1.

(i) Se ajusta la función de Gompertz de forma determinista. Con base en la ecuación (6) se pueden proporcionar valores iniciales de los parámetros γ y α , necesarios en el proceso de estimación de la función de Gompertz vía métodos numéricos. Se calculan los estimadores $\Theta = (\alpha, \beta, \gamma)$ de los parámetros de la función de Gompertz $f(t; \Theta)$ por el método de mínimos cuadrados, es decir, minimizando la suma de las diferencias al cuadrado entre el dato observado x_t y $f(t; \Theta)$. De manera matricial, la función de mínimos cuadrados tiene la siguiente forma

$$SS(\Theta) = [\mathbf{x} - \mathbf{f}(\Theta)]'[\mathbf{x} - \mathbf{f}(\Theta)],$$

donde $\mathbf{f}(\Theta) = [f(t_1; \Theta), f(t_2; \Theta), \dots, f(t_n; \Theta)]'$.

Se denota por $\tilde{\Theta} = (\hat{\alpha}, \hat{\beta}, \hat{\gamma})$ y $\tilde{f}(t) = f(t; \tilde{\Theta})$ a los estimadores y a la función de Gompertz estimada, respectivamente.

(ii) Se calculan los errores η_t definidos por $\eta_t = x_t - \tilde{f}(t)$ y se verifica que se modelen mediante una serie de tiempo autorregresiva $AR(p)$. Se calcula la matriz de correlación \mathbf{V} de la serie de tiempo. La componente aleatoria $AR(p)$ se determina y estima con base en la metodología Box-Jenkins que se presenta en Box, Jenkins, Reinsel, y Ljung (2016). Por lo tanto

$$\eta_t = \phi_1 \eta_{t-1} + \phi_2 \eta_{t-2} + \dots + \phi_p \eta_{t-p} + \varepsilon_t, \quad (8)$$

donde $\{\varepsilon_t\}$ es un proceso de ruido blanco Gaussiano con media cero y varianza constante σ^2 .

Etapa 2.

Se estima de nuevo el vector de parámetros Θ a través de mínimos cuadrados generalizados o ponderados, es decir, se minimiza la siguiente expresión

$$[\mathbf{x} - \mathbf{f}(\Theta)]' \mathbf{V}^{-1} [\mathbf{x} - \mathbf{f}(\Theta)], \quad (9)$$

donde $\mathbf{f}(\Theta) = [f(t_1; \Theta), f(t_2; \Theta), \dots, f(t_n; \Theta)]'$.

Para efectuar la Etapa 2 se puede considerar que mediante la descomposición de Cholesky se tiene que la matriz de correlación $\mathbf{V} = \mathbf{U}'\mathbf{U}$, donde \mathbf{U} es una matriz triangular superior. Por lo tanto, la ecuación (9) es equivalente a

$$\begin{aligned} [\mathbf{x} - \mathbf{f}(\Theta)]' \mathbf{V}^{-1} [\mathbf{x} - \mathbf{f}(\Theta)] &= [\mathbf{x} - \mathbf{f}(\Theta)]' \mathbf{R}'\mathbf{R} [\mathbf{x} - \mathbf{f}(\Theta)] \\ &= [\mathbf{z} - \mathbf{k}(\Theta)]' [\mathbf{z} - \mathbf{k}(\Theta)]. \end{aligned} \quad (10)$$

donde $\mathbf{R} = (\mathbf{U}')^{-1}$, $\mathbf{z} = \mathbf{R}\mathbf{x}$ y $\mathbf{k}(\Theta) = \mathbf{R}\mathbf{f}(\Theta)$.

Por lo tanto, la segunda etapa se reduce a minimizar mediante mínimos cuadrados la expresión $[\mathbf{z} - \mathbf{k}(\Theta)]' [\mathbf{z} - \mathbf{k}(\Theta)]$. La estimación de parámetros en este procedimiento en dos etapas requiere métodos computacionales; en este artículo se utilizará el método de Levenberg–Marquardt. Note que este segundo ajuste ya considera el factor aleatorio de las variaciones que tienen con respecto al primer ajuste.

5 Aplicaciones

5.1 COVID-19 en Yucatán

Al inicio de la pandemia por COVID-19, la función de Gompertz fue ajustada a los datos del número de casos positivos por COVID-19 que se acumularon a través del tiempo, como fue el caso descrito en Ahmadi, Fadaei, Shirani, y Rahmani (2020) y Torrealba et al. (2020) para Irán y México, respectivamente.

Para ilustrar el procedimiento en dos etapas descrito en este artículo, se consideran los datos correspondientes a los registros de la Secretaría de Salud en el estado de Yucatán durante 80 días a partir de que se presentó el primer caso positivo por COVID-19 el día 13 de marzo de 2020 y hasta que concluyó la Jornada Nacional de Sana Distancia el 31 de mayo de 2020. La información está dada por $\mathbf{x} = [x_1, \dots, x_{80}]'$, siendo x_t el número total de casos positivos acumulados hasta el día t .

Al realizar un análisis exploratorio de la información, se observa en la Figura 3 que la gráfica de $\log(x_t)$ vs $\log(x_{t+1})$ sugiere que el modelo dado por la ecuación (1) es adecuado para los datos, ya que se presenta una tendencia de línea recta, con excepción de una pequeña parte al inicio de la pandemia. La tendencia se hace más visible conforme se avanza en los días y se cuenta con mayor información respecto al número acumulado de casos positivos.

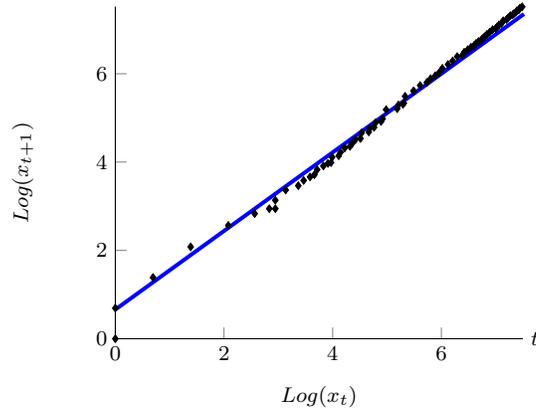


Figura 3: Gráfica de $\text{Log}(x_t)$ vs $\text{Log}(x_{t+1})$

De acuerdo con la metodología descrita en la Sección 4, la Etapa 1 consiste en la estimación de los parámetros $\Theta = (\alpha, \beta, \gamma)$ de la función de Gompertz. Puesto que la función de Gompertz no es lineal, se requieren de métodos de mínimos cuadrados no lineales para minimizar $SS(\Theta)$, optando por el método de Levenberg–Marquardt que se ilustra en Sauer (2012) y que hace uso de las derivadas parciales de la función de Gompertz para estimar los parámetros, siendo estas

$$\begin{aligned}\frac{\partial}{\partial \alpha} f(t; \Theta) &= \frac{1}{\alpha} f(t; \alpha, \beta, \gamma). \\ \frac{\partial}{\partial \beta} f(t; \Theta) &= -e^{-\gamma t} f(t; \alpha, \beta, \gamma). \\ \frac{\partial}{\partial \gamma} f(t; \Theta) &= \beta t e^{-\gamma t} f(t; \alpha, \beta, \gamma).\end{aligned}$$

Las estimaciones resultantes de los parámetros fueron $\tilde{\alpha} = 5234.89$, $\tilde{\beta} = 9.0921$ y $\tilde{\gamma} = 0.0268$.

Se procedió a calcular los errores $\eta_t = x_t - \tilde{f}(t)$ e identificar el orden del modelo $AR(p)$ al que se ajustan. En la Figura 4 se presenta la gráfica temporal de los errores, la cual sugiere que no hay evidencia en contra de la estacionariedad de la serie de tiempo.

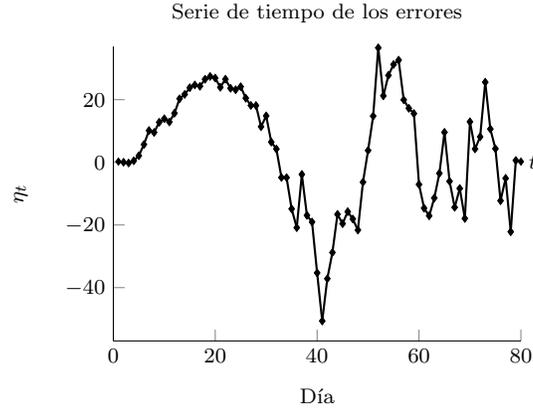


Figura 4: Gráfica de la serie de tiempo η_t

Por otra parte, el comportamiento de la Función de Autorrelación (ACF) estimada es tal que decrece lentamente y el comportamiento de la Función de Autorrelación Parcial (PACF) muestral es significativamente diferente de cero en el primer Lag y se desvanece a partir del segundo Lag, por lo que la metodología Box-Jenkins sugiere un modelo $AR(1)$ para este conjunto de datos. Este modelo estaría sugiriendo una dependencia entre el número acumulado de casos positivos por COVID-19 de un día con el dato del día anterior solamente.

Con una significancia de 0.05 se realizó la prueba de raíces unitarias de Dickey-Fuller y se rechazó la hipótesis de que se tengan raíces unitarias, concluyendo que la serie $\{\eta_t\}$ es estacionaria. Se ajustó el modelo $AR(1)$ y se verificó que los errores ε_t del modelo $AR(1)$ satisfacen el supuesto de un comportamiento de ruido blanco Gaussiano. La Prueba Box-Ljung ($p\text{-valor}=0.1047$) y la prueba Kolmogorov-Smirnov de Normalidad ($p\text{-valor}=0.1444$), no presentan evidencia en contra de estos supuestos. Con base en lo anterior, la estimación del modelo estadístico presentado por la ecuación (8) para la serie de tiempo η_t que modela a los errores, está dada por

$$\eta_t = \widetilde{\phi}_1 \eta_{t-1} + \varepsilon_t$$

donde $\widetilde{\phi}_1 = 0.8608$ y $\varepsilon_t \sim N(0, 85.57)$. En Seber y Wild (2003) se indica que la estimación de la inversa de la matriz de correlación de un modelo $AR(1)$.

$$\mathbf{V}^{-1} = \frac{1}{1 - \tilde{\phi}_1^2} \begin{pmatrix} 1 & -\tilde{\phi}_1 & 0 & \cdots & 0 & 0 \\ -\tilde{\phi}_1 & 1 + \tilde{\phi}_1^2 & -\tilde{\phi}_1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 + \tilde{\phi}_1^2 & -\tilde{\phi}_1 \\ 0 & 0 & 0 & \cdots & -\tilde{\phi}_1 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 3.86 & -3.32 & 0 & \cdots & 0 & 0 \\ -3.32 & 6.72 & -3.32 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 6.72 & -3.32 \\ 0 & 0 & 0 & \cdots & -3.32 & 3.86 \end{pmatrix}.$$

Con base en la ecuación (10) se minimizó $[\mathbf{z} - \mathbf{k}(\Theta)]' [\mathbf{z} - \mathbf{k}(\Theta)]$ mediante el método de Levenberg-Marquardt. La nueva estimación del vector de parámetros es $\hat{\Theta} = (5205.47, 9.2074, 0.0272)$. Por lo tanto, el modelo ajustado de la función de Gompertz es $\hat{f}(t) = 5205.47e^{-9.2074e^{-0.0272t}}$.

Con base en el ajuste de $\hat{f}(t)$ se puede estimar su derivada $\hat{f}'(t)$ y el tiempo donde se localiza el punto inflexión \mathbf{t}_1 de la función de Gompertz. La curva ajustada y los datos observados se presentan en la Figura 5, de igual manera se presentan los casos diarios de pacientes positivos por COVID-19 y la curva $\hat{f}'(t)$ ajustada.

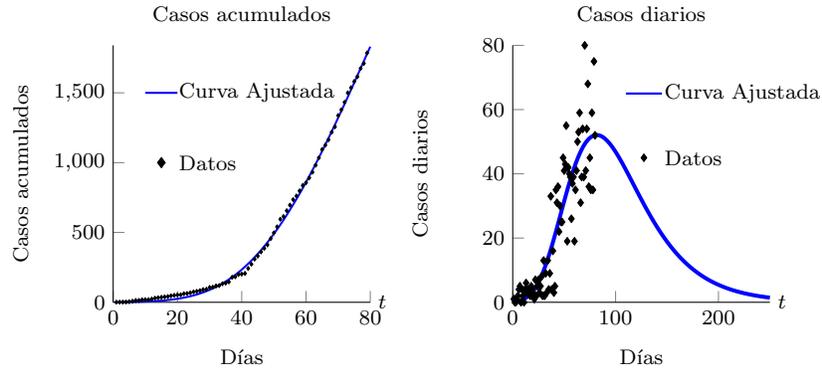


Figura 5: Casos Positivos por COVID-19

El punto de inflexión de la función de Gompertz se localiza en el tiempo $\mathbf{t}_1 = 81.61$, es decir, alrededor del día 81 se hubiera tenido el día pico si se hubiesen mantenido las restricciones impuestas durante la Jornada Nacional de Sana Distancia. Por otra parte, de acuerdo con la ecuación (7) se estima que en el día $\mathbf{t}_2=46$ se presenta un mayor cambio en el aumento de los casos diarios por COVID-19 con respecto al día anterior. Por otro lado, se estima

que en el día $t_3=117$ se presenta un mayor cambio en la disminución de los casos diarios por COVID-19 con respecto al día anterior. De igual manera, de acuerdo con las propiedades presentadas en la ecuación (1) se deduce que el número de contagios que se hubiera esperado tener bajo estas condiciones sería 5,205.

5.2 Biogás

El desperdicio orgánico contiene numerosas componentes fáciles de degradar con ayuda de la digestión anaeróbica (DA). El biogás producido mediante la DA es separado fácilmente y utilizado en distintas formas como una energía limpia. Por lo tanto, la DA tiene la ventaja de ser un proceso simple, con bajo impacto ambiental, altos beneficios económicos y un alto potencial para el reciclado del desperdicio orgánico. Existe literatura que incorpora la función de Gompertz para modelar la producción de biogás que se genera a través de la DA en los residuos orgánicos como por ejemplo Newton, Omodia, y Tamuno (2016) y Zhu, Yang, y Cheng (2019). Por otro lado, en la investigación presentada por Hosseini, Azizi, Baziyar, Hafez, y Elbeshbishy (2019) se compara la producción de biogás que se genera mediante diferentes fuentes de inóculo en muestras tomadas de una planta de aguas residuales localizada en la ciudad de Toronto. Por lo que proponemos aplicar la metodología en dos etapas usando la función de Gompertz para los datos de Hosseini et al. (2019). La información está dada por $\mathbf{x} = [x_1, \dots, x_{36}]'$ siendo x_t la producción acumulada de biohidrógeno en residuos que fueron sometidos a una temperatura de $38^\circ C$ desde el primer día hasta el día t .

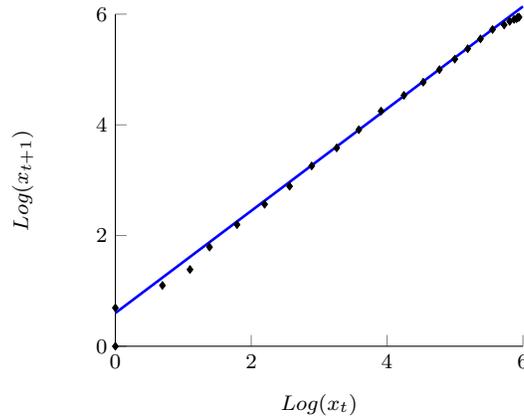


Figura 6: Gráfica de $\text{Log}(x_t)$ vs $\text{Log}(x_{t+1})$

Al realizar el diagrama de dispersión que se ilustra en Figura 6, se determinó que el modelo dado por la ecuación (1) es adecuado ya que se presenta una tendencia de línea recta.

De acuerdo con la metodología de la Sección 4 se procedió a realizar una primera estimación de los parámetros $\Theta = (\alpha, \beta, \gamma)$ por mínimos cuadrados y cuyo resultado fue $\tilde{\alpha} = 399.97$, $\tilde{\beta} = 125.35$, $\tilde{\gamma} = 0.2606$. Por lo tanto, se tiene que el modelo determinista estimado es $\tilde{f}(t) = 399.97e^{-125.35e^{-0.2606t}}$.

Posteriormente se procedió a obtener la serie de tiempo de los errores definida por $\eta_t = x_t - \tilde{f}(t)$ y cuya gráfica se presenta en la Figura 7, misma que sugiere que no hay evidencia en contra de la estacionariedad de la serie de tiempo.

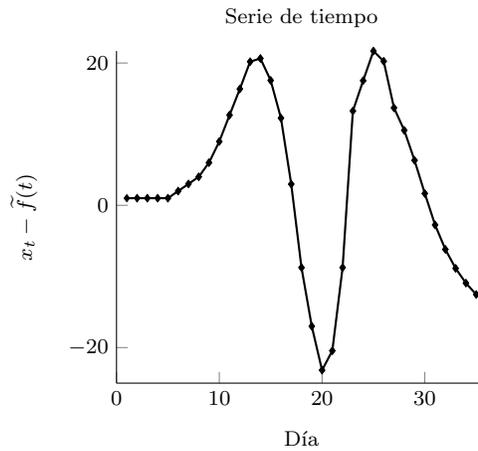


Figura 7: Gráfica temporal de la serie de tiempo η_t

Por otra parte, el comportamiento de la Función de Autorrelación (ACF) estimada es tal que decrece lentamente y el comportamiento de la Función de Autorrelacion Parcial (PACF) muestral es significativamente diferente de cero en el primer y segundo Lag y se desvanece a partir del tercer Lag, por lo que la metodología Box-Jenkins sugiere un modelo $AR(2)$ para este conjunto de datos. Este modelo estaría sugiriendo una dependencia entre la producción de biogás acumulada de un día con el biogás producido de los dos días anteriores.

Con una significancia de 0.05 se realizó la prueba de raíces unitarias de Dickey-Fuller y se rechazó la hipótesis de que se tengan raíces unitarias, es decir, la serie $\{\eta_t\}$ es estacionaria.

Se ajustó el modelo $AR(2)$ y se verificó que los errores ε_t del modelo $AR(2)$ satisfacen el supuesto de un comportamiento de ruido blanco Gaussiano. La prueba Box-Ljung ($p - valor = 0.4408$) y la prueba Kolmogorov-Smirnov de Normalidad ($p - valor = 0.5406$), no presentan evidencia en contra de

estos supuestos. De lo anterior, el modelo estadístico ajustado para la serie de tiempo de los errores está dado por

$$\eta_t = \widetilde{\phi}_1 \eta_{t-1} + \widetilde{\phi}_2 \eta_{t-2} + \varepsilon_t,$$

donde $\widetilde{\phi}_1 = 1.578$, $\widetilde{\phi}_2 = -0.780$ y $\varepsilon_t \sim N(0, 11.57)$.

Para la Etapa 2 del método se necesita la matriz de correlaciones de la serie de tiempo η_t . De acuerdo con Novales (1993) se tiene que la correlación entre η_t y η_{t+k} cuando la serie se comporta de acuerdo con un $AR(2)$ está dado por

$$\rho_1 = \frac{\phi_1}{1 - \phi_2}, \quad \rho_2 = \frac{\phi_1^2}{1 - \phi_2} + \phi_2$$

y en general

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}, \quad k \geq 2. \quad (11)$$

Mediante herramientas computacionales y con base en la ecuación (11) se puede construir la matriz de correlación V_{ϕ_1, ϕ_2} y su inversa V_{ϕ_1, ϕ_2}^{-1} . De acuerdo con la ecuación (10) se minimizó $[\mathbf{z} - \mathbf{k}(\Theta)]' [\mathbf{z} - \mathbf{k}(\Theta)]$ mediante el método de Levenberg-Marquardt. Los valores estimados fueron $\hat{\Theta} = (397.14, 180.27, 0.25)$. Por lo tanto, el modelo ajustado de la función de Gompertz es $\hat{f}(t) = 397.14e^{-180.27e^{-0.25t}}$.

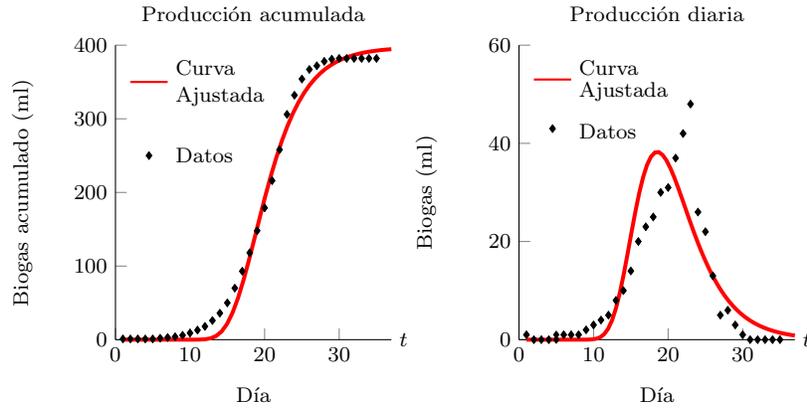


Figura 8: Producción de biogás

La curva ajustada y los datos observados se presentan en la Figura 8. El punto de inflexión de la función de Gompertz se localiza en el tiempo $\mathbf{t}_1 = 18.85$, es decir, alrededor del día 19 se estima que se tiene una mayor producción de biogás. Por otra parte, de acuerdo con la ecuación (7) se estima que en el día $\mathbf{t}_2 = 16.92$ se presenta un mayor cambio en el aumento de la pro-

ducción diaria de biogás con respecto al día anterior. Por otro lado, se estima que en el día $t_3=24.62$ se presenta un mayor cambio en la disminución de la producción diaria de biogás con respecto al día anterior. De igual manera, de acuerdo con las propiedades presentadas en la ecuación (1) se deduce que la cantidad de producción acumulada que se hubiera esperado tener a largo plazo fue 397.14 ml.

6 Conclusiones

La función de Gompertz ha sido útil para modelar fenómenos en los que se presenta un crecimiento a través del tiempo y en un periodo de tiempo largo este crecimiento se va aproximando a un cierto valor, el cual se interpreta como el tamaño al que se va aproximando la población cuando ha pasado mucho tiempo.

La metodología en dos etapas minimiza la suma ponderada de los errores considerando su correlación. Aunque en el desarrollo de las aplicaciones de este artículo, los errores se comportan de acuerdo a un modelo $AR(p)$, la metodología en dos etapas se puede aplicar cuando se comportan de acuerdo a una serie de tiempo $ARMA(p, q)$. Más aún, es posible realizar un proceso iterativo entre la Etapa 1 y Etapa 2 para obtener nuevas estimaciones de los parámetros.

Por otra lado, en este artículo se utilizó como curva de tendencia a la función de Gompertz, pero la metodología aplica para cualquier curva.

Finalmente aunque en muchas ocasiones los datos recolectados se ajustan en forma determinística a una función, cuando se desea considerar la aleatoriedad presente en la información, los modelos probabilísticos suelen ser de mayor utilidad porque permiten hacer inferencias y pronósticos. Estos temas serán estudiados en futuros trabajos.

7 Agradecimientos

Se agradece el apoyo parcial por parte del proyecto CONACYT CB-2015-01-252996: Modelos con estructuras de dependencia y sus aplicaciones II.

8 Bibliografía

Referencias

- Ahmadi, A., Fadaei, Y., Shirani, M., y Rahmani, F. (2020). Modeling and Forecasting Trend of COVID-19 Epidemic in Iran until May 13, 2020. *Medical Journal of The Islamic Republic of Iran*, 34(27). doi: 10.34171/mjiri.34.27
- Box, G., Jenkins, G., Reinsel, G., y Ljung, G. M. (2016). Time Series Analysis. Forecasting and Control (5th ed). *John Wiley and Sons Inc.*
- Cooper, R. (2005). Growth in the Ostrich (*Struthio Camelus* var. *Domesticus*). *Animal Science Journal*, 76, 1-4. doi: 10.1111/j.1740-0929.2005.00230
- Hosseini, K. E., Azizi, A., Bazyar, L. A., Hafez, H., y Elbeshbishy, E. (2019). Comparison of Liquid and Dewatered Digestate as Inoculum for Anaerobic Digestion of Organic Solidwastes. *Waste Management*, 87, 228-236. doi: 10.1016/j.wasman.2019.02.014
- Newton, C., Omodia, Y., y Tamuno, E. (2016). Development of Mathematical Models and Application of the Modified Gompertz Model for Designing Batch Biogas Reactors. *Waste and Biomass Valorization*, 7, 543-550. doi: 10.1007/s12649-016-9482-8
- Novalés, A. (1993). *Econometría*. *McGraw-Hill*.
- Przybylski, M., y García, E. (2004). Age and Growth of European Bitterling (*Rhodeus Sericeus*) in the Wieprz-Krzna Canal, Poland. *International Journal of Ecohydrology and Hydrobiology*, 4(2), 207-213.
- Sauer, T. (2012). *Numerical Analysis*. *Pearson Education, Inc.*
- Seber, G., y Wild, C. (2003). *Nonlinear Regression*. *John Wiley and Sons Inc.*
- Torrealba, O., Conde, R., y Hernández, A. (2020). Modeling and Prediction of COVID-19 in Mexico Applying Mathematical and Computational Models. *Chaos, Solitons Fractals*, 138, 1-8. doi: 10.1016/j.chaos.2020.109946
- Zhu, H., Yang, J., y Cheng, X. (2019). Application of Modified Gompertz Model to Study on Biogas Production from Middle Temperature Codi-

gestion of Pig Manure and Dead pigs. *E3S Web of Conferences* 118..
doi: 10.1051/e3sconf/201911803022