Properties of Gompertz data revealed with non-Gompertz integrable difference equation

Daisuke Satoh*

Abstract: Behaviour of the upper limit estimated by an unsuitable model (the logistic curve model) was mathematically analysed for data on an exact solution of the Gompertz curve model with integrable difference equations. The analysis contributes to identifying a suitable model because the behaviour is independent of noise included in actual data. A suitable model is indispensable for correct forecasts. The following results were proved. The estimated upper limit monotonically increases as the data size increases and converges to the upper limit estimated with the suitable model (the Gompertz curve model) as the data size approaches infinity. Therefore, the upper limit estimated with the logistic curve model is smaller than that estimated with the Gompertz curve model.

Subjects: Science; Mathematics & Statistics; Applied Mathematics; Mathematical Modeling; Non-Linear Systems; Statistics & Probability; Statistics; Multivariate Statistics; Statistics for Business, Finance & Economics;

Keywords: Forecasting; Gompertz curve model; model selection; difference equations; exact solutions; applied discrete systems

ABOUT THE AUTHOR
Daisuke Satoh received his B.E. and M.E. degrees in electronics and communication engineering in 1992 and 1994 and his Ph.D. in information and computer science in 2002 from Waseda University, Tokyo, Japan. He joined NTT in 1994. He is currently working as a senior research engineer at NTT Network Technology Laboratories. His research interests include applied discrete systems, software reliability, and teletraffic issues. He received the Best Author Award from JSIAM in 2011 and the Best Presentation Award at the 79th annual Convention of the Japanese Psychological Association in 2016. Dr. Satoh is a member of JSIAM, IEICE and ORSJ.

PUBLIC INTEREST STATEMENT
The Gompertz curve model has been used for forecasting in various areas. Accurate forecasting requires not only accurate parameter estimation but also appropriate model selection. This paper contributes to selecting an appropriate model because it provides identification of Gompertz data. It analyses the behaviour of the saturation level estimated by the logistic curve model as an inappropriate model when the data size increases. It proves the following results. The estimated saturation level monotonically increases as the data size increases and converges to the saturation level estimated with the appropriate model as the data size approaches infinity. Therefore, the saturation level estimated with the logistic curve model is smaller than that estimated with the Gompertz curve model. This discrete analysis is not made possible by an ordinary forward or central difference equation but by difference equations that have exact solutions; the exact solutions are on the exact solutions of the differential equations.
1. Introduction
The Gompertz curve model has been used as a growth curve model for forecasting in various areas of applied research such as marketing (Bemmaor, 1992; Franses, 1994a, 1994b; Kaldasch, 2011; Meade, 1984), economics (Figueira, Moura, & Ribeiro, 2011), management (Rządkowski, Głażewska, & Sawińska, 2015), the construction industry (Zhang, Wang, Yang, Zhang, & Lv, 2014), and software reliability engineering (Ohishi, Okamura, & Dohi, 2009; Pavlov, Spasov, Rahné, & Kyurkchiev, 2018; Satoh, 2000; Yamada, 1992) since it was proposed for specifying a mortality law in actuarial science (Gompertz, 1825).

Forecasting with the Gompertz curve model is accomplished by estimating parameters in the model from an obtained dataset. The Gompertz curve model has to be suitable for the obtained dataset when it is used for forecasting because forecasting using an unsuitable model can result in seriously incorrect forecasts (Chu et al., 2009; Martino, 2003; Yamakawa, Rees, Salas, & Alva, 2013).

The Gompertz curve model is difficult to identify as suitable for a given dataset from goodness-of-fit measures because regression analysis with a forward or central difference equation cannot recover parameters completely even when data are picked from the exact solution of the suitable differential model. Solutions of the difference equations are generally different from those of differential equations. Parameter estimates by a non-linear least squares estimation may not converge or may be sensitive to starting values, and algorithms may not provide a global optimum. Although the identification from goodness-of-fit measures is relatively correct in the final phase, forecasting in the final phase has little meaning. Forecasting is more variable in the earlier phase, where it has much more meaning.

Forecasting using an integrable difference equation that has an exact solution enables us to forecast accurately in the early phase (Satoh, 2000, 2001; Satoh & Yamada, 2002; Yamada, Inoue, & Satoh, 2002). Integrable difference equations (Hirota, 1979, 2000; Hirota & Takahashi, 2003; Morisita, 1965) and their application for forecasting (Satoh, 2000, 2001; Satoh & Uchida, 2010; Satoh & Yamada, 2001, 2002; Yamada et al., 2002) have been studied. A solution of an integrable difference equation plots discrete values on the exact solution of the original differential equation. General difference equations such as forward and central difference equations do not have this property. The forecasting using an integrable difference equation is simple because it uses regression analysis. Regression analysis with an integrable difference equation recovers parameters completely when data are picked from an exact solution of a suitable differential model (Satoh, 2000, 2001; Satoh & Yamada, 2001, 2002).

Estimated parameters change as the data size increases due to noise included in actual data even when a suitable model with an integrable difference equation is applied to the data. These changes depend on the noise. Parameters estimated by an unsuitable model change as the data size increases, too. However, these changes depend not only on the noise but also on the unsuitable model itself. If behaviour of parameters estimated by an unsuitable model can be analysed for data on an exact solution of a suitable model (that is, data without noise), the analysis will enable us to identify the suitable model because the analysis will be independent of noise.

A mathematical analysis using an integrable difference equation (“applied discrete systems”) is possible thanks to an exact solution of a difference equation. An integrable difference equation enables us to conduct regression analysis for a mathematical formula and regression coefficients as a mathematical formula even though we usually conduct regression analysis for numerical data and obtain regression coefficients numerically. Applied discrete systems enable us to analyse the behaviour of parameters estimated by an unsuitable model, whereas a non-linear least squares estimation cannot analyse the behaviour of parameters estimated by an unsuitable model algebraically because it is a numerical analysis.

This paper reveals properties of the Gompertz data with the logistic curve model (Hirota, 1979; Morisita, 1965; Satoh & Yamada, 2002) as an unsuitable model. The logistic curve model is also used for forecasting in various areas. If a given dataset does not have the properties of the
Gompertz data, forecasting using the Gompertz curve model can result in seriously incorrect forecasts. If the dataset has the properties, the Gompertz curve model is one of the best models that describe the data.

The rest of this paper is structured as follows: Section 2 presents an exact solution of the Gompertz curve model. Section 3 presents discrete logistic curve models that have exact solutions as unsuitable models to reveal properties of data on the exact solution of the Gompertz curve model. Section 4 derives a regression equation from the discrete logistic curve models with data on the exact solution of the Gompertz curve model. Section 5 states three lemmas and two theorems. The first theorem is that an estimated upper limit increases monotonically as the data size increases, which occurs as time elapses. The second theorem is that the estimated upper limit converges to the real upper limit, i.e. that estimated by the suitable model when the data size approaches infinity. Section 6 describes the practical use of the theorems for data that include noise. Finally, Section 7 concludes the paper with a summary.

2. Gompertz data
The Gompertz curve model is described as the following differential equation:

\[
\frac{dx(t)}{dt} = (\log b) x(t) \log \frac{x(t)}{k},
\]

where \( x(t) \) is a cumulative number up to time \( t \). By integrating Eq. (1) and assuming that \( x(0) = ka \), \( x(t) \) is written as

\[
x(t) = ka^b / (k > 0, \ 0 < a < 1, \ 0 < b < 1),
\]

where \( a, b, c \) and \( k \) are parameters estimated through regression analysis. Parameter \( k \) is an upper limit, parameter \( a \) is \( x(0)/k \), and parameter \( b_c \) is \( \log a / \log x \). Eq. (1) is discretised as

\[
x_{i+1} = x_i \left( \frac{x_i}{k} \right)^{\delta \log b}.
\]

The exact solution of Eq. (3) is

\[
x_i = ka^{(1+\delta \log b)},
\]

where

\[-1 < \delta \log b < 0.\]

The exact solution of the difference equation (4) is on that of the differential equation (2) when \( b_c \) is regarded as \( 1 + \delta \log b \). We call the data “Gompertz data” if the data satisfy Eq. (4). A regression equation for estimating the parameter values is obtained from Eq. (1) as

\[
Y_i = A_n - B_n \log x_i, (i = 1, 2, \ldots n),
\]

where

\[
Y_i = \log x_{i+1} - \log x_i.
\]

Parameters \( k, a, \) and \( b \) are estimated from regression analysis as

\[
\hat{k} = \exp \left( \frac{A_n}{B_n} \right),
\]

\[
\hat{a} = \exp \left( \frac{\sum_{i=1}^{n} \log \frac{x_i}{k}}{\sum_{i=1}^{n} (1 + \delta \log b)^i} \right), \text{ and}
\]

\[
\hat{b} = \frac{\sum_{i=1}^{n} \log \frac{x_i}{k}}{\sum_{i=1}^{n} (1 + \delta \log b)^i}.
\]
\[ \hat{b} = \exp \left( -\frac{\hat{B}_n}{\hat{\delta}} \right), \]  

(10)

where \( \hat{a}, \hat{b}, \hat{k}, \hat{A}_n, \) and \( \hat{B}_n \) are estimates of \( a, b, k, A_n, \) and \( B_n \). Note that \( Y_i \) in Eq. (6) is independent of time-interval \( \delta \) in Eq. (3) because \( \delta \) is not used in Eq. (6). The same estimates \( \hat{k}, \hat{a}, \) and \( \delta \log \hat{b} \) are obtained for any \( \delta \) value.

### 3. Discrete logistic curve model

The logistic curve model is described as the following differential equation:

\[
\frac{dx(t)}{dt} = \frac{r}{k} x(t) (k - x(t)),
\]  

(11)

where \( x(t) \) is a cumulative number up to \( t \). By integrating Eq. (11), \( x(t) \) is written as

\[
x(t) = \frac{k}{1 + m \exp(-rt)},
\]  

(12)

(Morisita, 1965) obtained the difference equation of Eq. (11) as

\[ x_{i+1} - x_i = \delta \frac{r_{dm}}{k} x_i (k - x_i). \]  

(13)

The exact solution of Eq. (13) is

\[
x_i = \frac{k}{1 + m (1 - \delta r_{dm})^2},
\]  

(14)

where \( m \) is obtained from \( x_0 = k/(1 + m) \) and \( t_i = i \delta \). (Hirota, 1979) obtained another difference equation:

\[ x_{i+1} - x_i = \delta \frac{r_{dh}}{k} x_i (k - x_i+1). \]  

(15)

The exact solution of Eq. (15) is

\[
x_i = \frac{k}{1 + m \left( \frac{1}{1 + \delta_{dh}} \right)^2},
\]  

(16)

where \( m \) is obtained from \( x_0 = k/(1 + m) \) and \( t_i = i \delta \).

The same regression equation is obtained from both difference equations by (Satoh & Yamada, 2002) as

\[ Y_i = A_n - B_n x_{i+1}, \quad (i = 0, 1, 2, \ldots, n - 1), \]  

(17)

where

\[
Y_i = \frac{x_{i+1}}{x_i}.
\]  

(18)

Parameters \( k, r_{dm}, r_{dh}, \) and \( m \) are estimated from regression analysis as

\[
\hat{k} = \frac{\hat{A}_n - 1}{\hat{B}_n},
\]  

(19)

\[
\delta \hat{r}_{dm} = 1 - \frac{1}{\hat{A}_n},
\]  

(20)

\[
\delta \hat{r}_{dh} = \hat{A}_n - 1,
\]  

(21)
\[ m = \begin{cases} \frac{\sum_{i=1}^{n}(k-x_i)}{\sum_{i=1}^{n}(x_i-k)} & \text{(for Eq. (13))} \\ \frac{\sum_{i=1}^{n}(k-x_i)}{\sum_{i=1}^{n}(x_i-k)} & \text{(for Eq. (15))} \end{cases} \]  

(22)

where \( n \) represents the data size, and \( \hat{A}_n \) and \( \hat{B}_n \) are estimates of \( A_n \) and \( B_n \). The same estimates \( \hat{k}, \delta \hat{r}_{dm}, \) and \( \hat{m} \) are obtained for any \( \delta \) because \( \delta \) is not included in Eq. (17). The relationship among \( r_{dm}, r_{dh}, \) and \( r \):

\[ 1 - \delta r_{dm} = \frac{1}{1 + \delta r_{dh}} = \exp(-\delta r) = \frac{1}{\hat{A}} \]

(23)

holds (see (Satoh & Yamada, 2002)). The same estimate of \( \hat{m} \) is obtained because of Eq. (23).

4. Analysis of Gompertz data with logistic model

The Gompertz data: \( x_i, (i = 1, 2, \ldots, n) \) satisfy

\[ \frac{x_{i+1}}{x_i} = \left( \frac{x_i}{k} \right)^{\delta \log b} \]

(24)

Regression analysis with the logistic curve model obtains

\[ y_i = \frac{x_{i+1}}{x_i} = \left( \frac{x_i}{k} \right)^{\delta \log b} = A_n - B_n x_{i+1} + \epsilon_i, \]

(25)

(26)

where \( A_n \) and \( B_n \) satisfy

\[ \min \sum_{i=0}^{n-1} \left\{ \left( \frac{x_i}{k} \right)^{\delta \log b} - (A_n - B_n x_{i+1}) \right\}^2. \]

(27)

Regression analysis determines \( A_n \) and \( B_n \) through the regression equation:

\[ y_i = \frac{x_{i+1}}{x_i} = \left( \frac{x_i}{k} \right)^{\delta \log b} = A_n - B_n x_{i+1} + \epsilon_i, \quad x_i = k. \]

(28)

The regression coefficients \( A_n \) and \( B_n \) are obtained as follows:

\[ A_n = \bar{y}_n + B_n \bar{x}_n = \frac{A_n'}{V(X_n)}, \]

(29)

\[ B_n = -\frac{\text{Cov}(X_n, Y_n)}{V(X_n)} = \frac{B_n'}{V(X_n)}, \]

(30)

where

\[ \bar{x}_n = \frac{1}{n} \sum_{i=1}^{n} x_i; \]

(31)

\[ \bar{y}_n = \frac{1}{n} \sum_{i=1}^{n} y_i; \]

(32)

\[ V(X_n) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x}_n)^2 = \frac{1}{n} \sum_{i=1}^{n} x_i^2 - \bar{x}_n^2 > 0; \]

(33)

\[ \text{Cov}(X_n, Y_n) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x}_n)(y_i - \bar{y}_n) = \frac{1}{n} \sum_{i=1}^{n} x_i y_i - \bar{x}_n \bar{y}_n. \]

(34)
\[ A'_n = \left( \frac{1}{n} \sum_{i=1}^{n} x_i^2 \right) \bar{y}_n - \bar{x}_n \left( \frac{1}{n} \sum_{i=1}^{n} x_i y_i \right). \]  

(35)

\[ B'_n = -\text{Cov}(x_n, y_n). \]  

(36)

The estimated upper limit is obtained from Eq. (19), where \( A_n \) and \( B_n \) in Eq. (19) are obtained from Eqs. (29) and (30).

5. Behaviour of upper limit estimated with unsuitable model

The upper limit estimated with the logistic curve model is proved to increase monotonically as the data size increases through the following three lemmas.

5.1. Behaviour of slope of regression equation

Lemma 5.1. The sign reversal of the slope of the regression equation is positive and monotonically decreases, i.e.

\[ 0 < B_{n+1} < B_n \]  

(37)

holds for \( n \geq 2 \).

Proof. We define

\[ f_g(x) = \left( \frac{x}{k} \right)^{\delta \log b}, \]  

(38)

which is the data curve of Eq. (25). Thus,

\[ \frac{df_g(x)}{dx} = \frac{\delta \log b}{k} \left( \frac{x}{k} \right)^{\delta \log b - 1} < 0 \]  

(39)

and

\[ \frac{d^2f_g(x)}{dx^2} = \frac{\delta \log b (\delta \log b - 1)}{k^2} \left( \frac{x}{k} \right)^{\delta \log b - 2} > 0 \]  

(40)

are obtained because of Eq. (5). Hence, function \( f_g(x) \) decreases as \( x \) increases and is a downward convex function.

From the property of a simple linear regression, the following equation

\[- B_n V(x_n) = \text{Cov}(x_n, y_n) \]  

(41)

holds for \( n \geq 2 \) because \(- B_n \) is the slope of the regression equation (Eq. (26)). Since function \( f_g(x) \) decreases monotonically as \( x \) increases,

\[ \text{Cov}(x_n, y_n) < 0. \]  

(42)

Thus,

\[ 0 < B_n \]  

(43)

is obtained for \( n \geq 2 \) because of Eqs. (33), (41), and (42).

Thus,

\[- (n + 1)B_{n+1} V(x_{n+1}) = (n + 1)\text{Cov}(x_{n+1}, y_{n+1}) \]  

(44)
is obtained from Eq. (41). The left-hand side of Eq. (44) is rewritten as

\[-(n + 1)B_{n-1}V(x_{n+1})\]

\[= -B_{n+1} \sum_{i=1}^{n} (x_i - \bar{x}_n)^2\]

\[= -B_{n+1} \left( \sum_{i=1}^{n} (x_i - \bar{x}_n + \bar{x}_n - \bar{x}_{n+1})^2 + (\bar{x}_{n+1} - \bar{x}_n)^2 \right)\]

\[= -B_{n+1} \left( \sum_{i=1}^{n} (x_i - \bar{x}_n)^2 + n(\bar{x}_{n+1} - \bar{x}_n)^2 + (\bar{x}_{n+1} - \bar{x}_n)^2 \right)\]

\[= -B_{n+1} \left( \sum_{i=1}^{n} (x_i - \bar{x}_n)^2 + n(\bar{x}_{n+1} - \bar{x}_n)^2 + n^2(\bar{x}_{n+1} - \bar{x}_n)^2 \right)\]

\[= -B_{n+1} \left( \sum_{i=1}^{n} (x_i - \bar{x}_n)^2 + n(n + 1)(\bar{x}_{n+1} - \bar{x}_n)^2 \right)\]

\[= -nB_{n+1} \left\{ V(x_n) + (n + 1)(\bar{x}_{n+1} - \bar{x}_n)^2 \right\}. \quad (45)\]

where

\[(n + 1)\bar{x}_{n+1} = n\bar{x}_n + x_{n+1} \quad (46)\]

is used. The right-hand side of Eq. (44) is rewritten as

\[n + 1)\text{Cov}(x_{n+1}, y_{n+1})\]

\[= \sum_{i=1}^{n} (x_i - \bar{x}_n + \bar{x}_n - \bar{x}_{n+1})(y_i - \bar{y}_n + \bar{y}_n - \bar{y}_{n+1}) + (x_{n+1} - \bar{x}_n)(y_{n+1} - \bar{y}_{n+1})\]

\[= \sum_{i=1}^{n} (x_i - \bar{x}_n)(y_i - \bar{y}_n) + n(\bar{x}_{n+1} - \bar{x}_n)(\bar{y}_{n+1} - \bar{y}_n) + (x_{n+1} - \bar{x}_n)(\bar{y}_{n+1} - \bar{y}_{n+1})\]

\[= \sum_{i=1}^{n} (x_i - \bar{x}_n)(y_i - \bar{y}_n) + n(\bar{x}_{n+1} - \bar{x}_n)(\bar{y}_{n+1} - \bar{y}_n) + n^2(\bar{x}_{n+1} - \bar{x}_n)(\bar{y}_{n+1} - \bar{y}_n)\]

\[= n\{\text{Cov}(x_n, y_n) + (n + 1)(\bar{x}_{n+1} - \bar{x}_n)(\bar{y}_{n+1} - \bar{y}_n)\}\]

\[= -n\{B_nV(x_n) - (n + 1)(\bar{x}_{n+1} - \bar{x}_n)(\bar{y}_{n+1} - \bar{y}_n)\}. \quad (47)\]

where Eq. (46) and

\[(n + 1)\bar{y}_{n+1} = n\bar{y}_n + y_{n+1} \quad (48)\]

are also used. Hence,

\[B_{n+1} \left\{ V(x_n) + (n + 1)(\bar{x}_{n+1} - \bar{x}_n)^2 \right\} = B_nV(x_n) - (n + 1)(\bar{x}_{n+1} - \bar{x}_n)(\bar{y}_{n+1} - \bar{y}_n) \quad (49)\]

is obtained from Eqs. (45) and (47). Furthermore, the right-hand side of Eq. (49) is rewritten as

\[B_nV(x_n) - (n + 1)(\bar{x}_{n+1} - \bar{x}_n)(\bar{y}_{n+1} - \bar{y}_n)\]

\[= B_n \left\{ V(x_n) + (n + 1)(\bar{x}_{n+1} - \bar{x}_n)^2 \right\} \]

\[-(n + 1)B_n(\bar{x}_{n+1} - \bar{x}_n)^2 - (n + 1)(\bar{x}_{n+1} - \bar{x}_n)(\bar{y}_{n+1} - \bar{y}_n)\]
\[
\begin{align*}
= B_n \left\{ V(x_n) + (n + 1)(\bar{x}_{n+1} - \bar{x}_n)^2 \right\} \\
- (n + 1)(\bar{x}_{n+1} - \bar{x}_n) \left\{ B_n(\bar{x}_{n+1} - \bar{x}_n) + \bar{y}_{n+1} - \bar{y}_n \right\} \\
= B_n \left\{ V(x_n) + (n + 1)(\bar{x}_{n+1} - \bar{x}_n)^2 \right\} \\
- (\bar{x}_{n+1} - \bar{x}_n) \left\{ B_n(\bar{x}_{n+1} - \bar{x}_n) + y_{n+1} - y_n \right\} \\
= B_n \left\{ V(x_n) + (n + 1)(\bar{x}_{n+1} - \bar{x}_n)^2 \right\} \\
- (\bar{x}_{n+1} - \bar{x}_n) \left\{ B_n(\bar{x}_{n+1} - \bar{x}_n) + y_{n+1} - (A_n - B_n \bar{x}_n) \right\} \\
= B_n \left\{ V(x_n) + (n + 1)(\bar{x}_{n+1} - \bar{x}_n)^2 \right\} \\
- (\bar{x}_{n+1} - \bar{x}_n) \left\{ y_{n+1} - (A_n - B_n \bar{x}_{n+1}) \right\}.
\end{align*}
\]  

by using Eqs. (46) and (48) and the property that a regression equation satisfies the average, i.e.
\[
y_n = A_n - B_n \bar{x}_n.
\]  

Hence, Eq. (49) is rewritten as
\[
(B_{n+1} - B_n) \left\{ V(x_n) + (n + 1)(\bar{x}_{n+1} - \bar{x}_n)^2 \right\} = - (\bar{x}_{n+1} - \bar{x}_n) \left\{ y_{n+1} - (A_n - B_n \bar{x}_{n+1}) \right\}.
\]  

We have
\[
\bar{x}_n < \bar{x}_{n+1},
\]  

because \(x_n\) monotonically increases. Furthermore, the regression equation (Eq. (26)) intercepts the data curve of Eq. (25) at two points between \((x_1, y_1)\) and \((x_n, y_n)\) because Eq. (25) is a downward convex function. Thus,
\[
A_n - B_n x_{n+1} < y_{n+1}
\]  

is obtained. As a result,
\[
B_{n+1} < B_n
\]  

is obtained from Eqs. (52), (53), and (54).

\[\square\]

5.2. Magnitude relationship between y-coordinates of two consecutive regression lines at \(\bar{x}_{n+1}\)

**Lemma 5.2.** The following magnitude relationship between y-coordinates of two consecutive regression lines at \(\bar{x}_{n+1}\)
\[
A_n - B_n \bar{x}_{n+1} < A_{n+1} - B_{n+1} \bar{x}_{n+1}
\]  

holds for \(n \geq 2\).
Proof. We have
\[ nA_n = n\bar{y}_n + nB_n\bar{x}_n \]  
(57)
from Eq. (51). By using Eqs. (46) and (48), Eq. (57) is rewritten as
\[ nA_n = n\bar{y}_n + nB_n\bar{x}_n = (n + 1)\bar{y}_{n+1} - y_{n+1} + B_n\{ (n + 1)\bar{x}_{n+1} - x_{n+1} \} \]
\[ = (n + 1)\bar{y}_{n+1} + (n + 1)B_n\bar{x}_{n+1} - (n + 1)B_n\bar{x}_n - y_{n+1} - B_nx_{n+1} \]
\[ = (n + 1)(\bar{y}_{n+1} + B_n\bar{x}_{n+1} - (n + 1)(B_n - B_n)\bar{x}_{n+1} - y_{n+1} - B_nx_{n+1} \]
\[ = (n + 1)A_{n+1} - (n + 1)(B_n - B_n)\bar{x}_{n+1} - y_{n+1} - B_nx_{n+1}. \]  
(58)
Eq. (58) is rewritten as
\[ (n + 1)(A_{n+1} - A_n - (B_n - B_n)\bar{x}_{n+1}) = y_{n+1} - (A_n - B_nx_{n+1}). \]  
(59)
The right-hand side of Eq. (59) is positive because of Eq. (54). Hence,
\[ A_n - B_n\bar{x}_{n+1} < A_{n+1} - B_{n+1}\bar{x}_{n+1} \]
(60)
is obtained.

5.3. Value of \( y \)-coordinate of intersection of two consecutive regression lines

Lemma 5.3. The value of \( y \)-coordinate \( (y_p) \) of the intersection of two consecutive regression lines is larger than 1, i.e.
\[ y_p = \frac{B_nA_{n+1} - A_nB_{n+1}}{B_n - B_{n+1}} > 1 \]  
(61)
holds for \( n \geq 2 \).

Proof. By using Eqs. (46), (48), and (51), \( y_p \) is rewritten as follows,
\[ y_p = \frac{B_nA_{n+1} - A_nB_{n+1}}{B_n - B_{n+1}} = \frac{B_n(\bar{y}_{n+1} + B_n\bar{x}_{n+1} - (y_n + B_nx_n)B_{n+1}}{B_n - B_{n+1} \]
\[ = \frac{B_n\bar{y}_{n+1} - y_n + B_nB_n(\bar{x}_{n+1} - x_n)}{B_n - B_{n+1}} = \frac{B_{n+1}(y_{n+1} - y_n) + B_nB_n(\bar{x}_{n+1} - x_n)}{B_n - B_{n+1}} \]
\[ = \frac{1}{B_n \left( \frac{1}{B_{n+1}} - \frac{1}{B_n} \right)} \{ \bar{y}_{n+1} - \bar{y}_n + B_n(\bar{x}_{n+1} - \bar{x}_n) \}. \]  
(62)
We have
\[ \bar{y}_n > 1 \]  
(63)
for \( n \geq 1 \) because \( y_i > 1 \) for \( i \geq 1 \). From Eqs. (43) and (55),
\[ B_n \left( \frac{1}{B_{n+1}} - \frac{1}{B_n} \right) > 0 \]  
(64)
is obtained. Then, we also rewrite
\[ (n + 1) \{ \bar{y}_{n+1} - \bar{y}_n + B_n(\bar{x}_{n+1} - \bar{x}_n) \} = (n + 1) \{ \bar{y}_{n+1} - (A_n - B_n\bar{x}_n) + B_n(\bar{x}_{n+1} - \bar{x}_n) \} \]
\[ = (n + 1) \{ \bar{y}_{n+1} - (A_n - B_n\bar{x}_n) \} \]
\[ = n\bar{y}_n + y_{n+1} - (n + 1)A_n + B_n(nk_n + x_{n+1}) \]
\[ = n(y_n - (A_n - B_n\bar{x}_n)) + y_{n+1} - A_n + B_nx_{n+1} \]
\[ = y_{n+1} - (A_n - B_nx_{n+1}) > 0. \]  
(65)
Hence,
\[ y_p > 1 \]  
(66)
is obtained from Eqs. (63), (64), and (65).

### 5.4. Monotonic increase of estimated upper limit

**Theorem 5.4.** Let data $X_i$ $(i = 0, 1, 2, \ldots, n, n \geq 2)$ be described by the Gompertz curve model. The function of estimates of the upper limit defined as Eq. (19),
\[ \hat{k}_n \overset{def}{=} \frac{A_n - 1}{B_n} \]  
(67)
monotonically increases on $n \geq 2$, where parameters $A_n$ and $B_n$ are obtained through the regression equation (Eq. (28)), i.e.
\[ \hat{k}_{n+1} - \hat{k}_n = \frac{A_{n+1} - 1}{B_{n+1}} \frac{A_n - 1}{B_n} > 0 \]  
(68)
for $n \geq 2$.

**Proof.** The $x$-coordinate of the point on the intersection of two consecutive regression lines $x_p$ is smaller than $x_{n+1}$ because of Eqs. (37) and (56). Thus, we have
\[ A_n - B_n x < A_{n+1} - B_{n+1} x \]  
(69)
for $x > x_p$. Because of Eq. (61),
\[ x_p < \hat{k}_n \]  
(70)
is obtained for $n \geq 2$. From Eq. (67), the estimated upper limit $\hat{k}_n$ is the $x$-coordinate of a point on the intersection of
\[ y = A_n - B_n x, \]  
(71)
y = 1.  
(72)
By subtracting 1 from both sides of Eq. (69),
\[ A_n - 1 - B_n x < A_{n+1} - 1 - B_{n+1} x \]  
(73)
is obtained where $x > x_p$. By substituting $\hat{k}_n(x_p)$ for $x$,
\[ A_n - 1 - B_n \hat{k}_n = 0 < A_{n+1} - 1 - B_{n+1} \hat{k}_n \]  
(74)
is obtained. Hence,
\[ \hat{k}_n < \frac{A_{n+1} - 1}{B_{n+1}} = \hat{k}_{n+1} \]  
(75)
is obtained for $n \geq 2$. □

### 5.5. Limit of upper limit

**Theorem 5.5.** The limit of the upper limit converges to the upper limit estimated with the suitable model as
\[ \lim_{n \to \infty} \hat{k}_n = \lim_{n \to \infty} \frac{A_n - 1}{B_n} = k. \]  
(76)
Proof. We have
\[ A_n - B_nx_n \leq y_n, \]  
(77)
because the regression equation (Eq. (26)) intercepts the data curve of Eq. (25) at two points between \((x_1, y_1)\) and \((x_n, y_n)\) and Eq. (25) is a downward convex function. We also have
\[ y_n < \bar{y}_n \]  
(78)
because function \(f_g(x)\) decreases as \(x\) increases from Eq. (39). Thus, we have
\[ \frac{A_n - 1}{B_n} \leq x_n + \frac{y_n - 1}{B_n} \]  
(79)
from Eq. (77) and
\[ \bar{x}_n + \frac{y_n - 1}{B_n} < \frac{A_n - 1}{B_n} \]  
(80)
from Eq. (78). Hence,
\[ \bar{x}_n + \frac{y_n - 1}{B_n} < \frac{A_n - 1}{B_n} \leq x_n + \frac{y_n - 1}{B_n} \]  
(81)
is obtained from Eqs. (79) and (80). We have
\[ \lim_{n \to \infty} x_n = k. \]  
(82)
We also have
\[ \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} x_i = k, \]  
(83)
because of the Cesàro means theorem (Hardy, 1991). From Eq. (82), we have
\[ \lim_{n \to \infty} y_n = \lim_{n \to \infty} \frac{X_{n+1}}{X_n} = 1. \]  
(84)
The regression equation (Eq. (26)) intercepts the curve of Eq. (24) at two points: \((x_L, y_L), (x_R, y_R)\) between \((x_1, y_1)\) and \((x_n, y_n)\). Hence, there exists \(\xi\) such that
\[ B_n = -\frac{d}{dx} \left( \frac{x^{\delta \log b}}{b^{\delta \log b}} \right) \bigg|_{x=\xi} = -\frac{\delta \log b}{k} \left( \frac{\xi}{k} \right)^{\delta \log b - 1} = h(\xi) \]  
(85)
from the mean-value theorem, where
\[ x_1 < x_L < \xi < x_R < x_n, \]  
(86)
\[ A_n - B_n x_L = \left( \frac{x_L}{k} \right)^{\delta \log b}, \]  
(87)
\[ A_n - B_n x_R = \left( \frac{x_R}{k} \right)^{\delta \log b}, \]  
(88)
\[ x_L < x_R. \]  
(89)
From Eq. (53), \(\bar{x}_n\) monotonically increases as \(n\) increases. Also, \(x_n\) monotonically increases as \(n\) increases because \(y_n = x_{n+1}/x_n\) monotonically decreases as \(n\) increases. Furthermore, \(B_n\) monotonically decreases as \(n\) increases from Eq. (37). Hence, both \(x_L\) and \(x_R\) monotonically increase as \(n\) increases. As a result, \(\xi\) also monotonically increases as \(n\) increases. Since \(B_n\) monotonically decreases as \(n\) increases,
\[
\lim_{\xi \to k} \frac{b}{\delta \log b} < \lim_{n \to \infty} B_n < B_2 \tag{90}
\]

holds. Therefore, the limit of \(B_n\) is positive and finite because
\[
-\frac{\delta \log b}{k} > 0. \tag{91}
\]

From Eqs. (81), (82), (83), (84), and (90),
\[
\lim_{n \to \infty} \left( x_n + \frac{y_n - 1}{B_n} \right) = \lim_{n \to \infty} \left( x_n + \frac{y_n - 1}{B_n} \right) = k \tag{92}
\]
is obtained. Therefore, Eq. (76) holds from the squeeze theorem.

\section{6. Practical use}

The properties of the upper limit estimated by the logistic curve model and the slope of the regression equation are revealed in Sections 5.1 to 5.5. Furthermore, the upper limit is found to be smaller than that estimated by the Gompertz curve model from Theorems 5.4 and 5.5.

For practical use, we monitor whether the observed data conform to the three Lemmas, Theorem 5.4, and the magnitude relationship between the upper limits estimated by both models every time a new datapoint is obtained. If they do, the Gompertz curve model is one of the most suitable models. If they do not, the data are not regarded as Gompertz data because actual data with small noise retain these properties. If noise is so large that it drowns out the properties, the data are no longer described by the Gompertz curve model.

To confirm that data with small noise retain these properties, a sample dataset was generated with the non-homogeneous Poisson process (NHPP) model whose mean value function was a modified exact solution of the Gompertz equation (Yamada, 1992). The NHPP is a Poisson type model that takes the number of events per unit of time as independent Poisson random variables and is often used in software reliability engineering (Lyu, 1996; Yamada, 2014; Yamada & Tamura, 2016). The mean value function is
\[
x'_i = ka^{1+\delta \log b}' - ka \tag{93}
\]
to satisfy \(x'_0 = 0\) for the NHPP model (Yamada, 1992). An algorithm generating one realization for the NHPP in (Cinlar, 1975) was used with parameters:
\[
\begin{align*}
k & = 1000 \tag{94} \\
a & = 0.001 \tag{95} \\
\delta \log b & = \frac{\log (1 + a)}{\log a} - 1 \tag{96} \\
\delta & = 1 \tag{97}
\end{align*}
\]
in Eq. (4). Generated sample data with the NHPP model are shown in Figure 1.

Both upper limits estimated by the Gompertz and logistic curve models are shown with the first \(n\) available data points in Figure 1. The differences in upper limits estimated with the first \(n + 1\) and \(n\) available data points are shown in Figure 2.

The monotonic increase in the upper limits estimated by the logistic curve model was conserved at \(n \geq 8\) from Figure 2 even when data included noise, whereas the upper limits estimated by the Gompertz curve model fluctuated in the first 26 data points and slowly decreased after.
Both sign changed slopes of the regression equations estimated by the Gompertz and logistic curve models are shown with the first \( n \) available data points in Figures 2 and 2. Figure 2 is an enlarged view of the sign changed slopes estimated by the logistic curve model to clarify the estimated values. The differences in the sign changed slopes estimated with the first \( n + 1 \) and \( n \) available data points are shown in Figure 2.
The sign changed slopes estimated by the logistic curve model maintained positive values and monotonically decreased at \( n \geq 8 \) from Figure 3 even when data included noise, whereas those estimated by the Gompertz curve model fluctuated in the first 17 data points and slowly increased after.

7. Conclusion

Behaviour of the upper limit estimated with the logistic curve model as an unsuitable model was analysed for data on an exact solution of the Gompertz curve model. It was found and proved that the estimated upper limit monotonically increases as the data size increases, i.e. time elapses. It was also proved that the upper limit converges to the upper limit estimated with the suitable
model (the Gompertz curve model) as the data size approaches infinity. Therefore, the upper limit estimated with the logistic curve model is smaller than that estimated with the Gompertz curve model as the suitable model for data on the exact solution of the Gompertz curve model. These results were obtained by “applied discrete systems.” These results also contribute to identifying a suitable model even though actual data include noise because the estimated upper limit with low noise must behave similarly to that without noise. It was confirmed that the proved properties are conserved except in the very first phase with sample data with noise. If noise is so large that it drown out the properties, the data should no longer be analysed by the Gompertz curve model. If the above results are not observed from a given dataset, the data should not be analysed with the Gompertz curve model. Forecasting using the Gompertz curve model can result in seriously incorrect forecasts. However, if the above results are observed, the Gompertz curve model is one of the most suitable models. Further study will prove that a series of the estimated upper limits is upward or downward convex. The robustness of the behaviour of the upper limit against noise included in actual data is also to be further studied.

**Funding**
The author received no direct funding for this research.

**Author details**
Daisuke Satoh
E-mail: daisuke@mis.inpec.or.jp
ORCID ID: http://orcid.org/0000-0002-4086-7077
NTT Network Technology Laboratories, Musashino-shi, Tokyo, 180-8585, Japan.

**Description**
Behaviour of the upper limit estimated by an unsuitable model (the logistic curve model) was mathematically analysed for data on an exact solution of the Gompertz curve model with integrable difference equations. The analysis contributes to identifying a suitable model because the behaviour is independent of noise included in actual data. A suitable model is indispensable for correct forecasts. The following results were proved. The estimated upper limit monotonically increases as the data size increases and converges to the upper limit estimated with the suitable model (the Gompertz curve model) as the data size approaches infinity. Therefore, the upper limit estimated with the logistic curve model is smaller than that estimated with the Gompertz curve model.

**Citation information**
Cite this article as: Properties of Gompertz data revealed with non-Gompertz integrable difference equation, Daisuke Satoh, Cogent Mathematics & Statistics (2019), 6: 1596552.

**References**
Bennmor, A. C. (1992). Modeling the diffusion of new durable goods: Word-of-mouth effect versus consumer heterogeneity. In G. Laurent, G. L. Lilien, & B. Pras (Eds.), Research traditions in marketing (Vol. 5, pp. 201–223). Dordrecht: International Series in Quantitative Marketing Springer.

Chu, W. L., Wu, F. S., Kao, K. S., & Yen, D. C. (2009). Diffusion of mobile telephony: An empirical study in Taiwan. Telecommunications Policy, 33, 506–520. doi:10.1016/j.telpol.2009.07.003

Clelar, E. (1975). Introduction to stochastic process. Englewood Cliffs, NJ: Prentice-Hall.

Figueira, F. C., Moura, N. J., Jr., & Ribeiro, M. (2011). The Gompertz-Pareto income distribution. Physica A, 390, 689–698. doi:10.1016/j.physa.2010.10.014

Franses, P. H. (1994d). Gompertz curves with seasonality. Technological Forecasting and Social Change, 45, 267–297. doi:10.1016/0040-1625(94)90051-5

Franses, P. H. (1996b). Modeling new product sales; an application of cointegration analysis. International Journal of Research in Marketing, 11, 491–502. doi:10.1016/0167-8116(94)00013-1

Gompertz, B. (1825). On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. Philosophical Transactions of the Royal Society of London (Vol. 115, pp. 513–583).

Hardy, G. H. (1991). Divergent series (2nd ed.). New York, NY: American Mathematical Society.

Hirota, R. (1979). Nonlinear partial difference equations. V. nonlinear equations reducible to linear equations. Journal of the Physical Society of Japan, 46, 312–319. doi:10.1143/JPSJ.46.312

Hirota, R. (2000). Lecture on discrete equations. Saiensusha Library for Senior & Graduate Courses (SGC-8), pp. 64–70, In Japanese.

Hirota, R., & Takahashi, D. (2003). Discrete and ultradiscrete systems. Tokyo: Kyoritsushuppan (in Japanese).

Kaldasch, J. (2013). Evolutionary model of an anonymous consumer durable market. Physica A, 390, 2692–2715. doi:10.1016/j.physa.2011.02.039

Lyu, M. R. (ed.). (1996). Handbook of software reliability engineering. New York: McGraw-Hill.

Martino, J. P. (2003). A review of selected recent advances in technological forecasting. Technological Forecasting & Social Change, 70, 719–733. doi:10.1016/S0040-1625(02)00375-X

Meade, N. (1984). The use of growth curves in forecasting market development—A review and appraisal. Journal of Forecasting, 3, 429–451. doi:10.1002/FSF.1999–131X

Morisita, M. (1965). The fitting of the logistic equation to the rate of increase of population density. Researches on Population Ecology, 7, 52–55. doi:10.1007/BF02518815

Ohishi, K., Okamura, H., & Dohi, T. (2009). Gompertz software reliability model: Estimation algorithm and empirical validation. The Journal of Systems and Software, 82, 535–543. doi:10.1016/j.jss.2008.11.840

Pavlov, N., Spasov, G., Rahnev, A., & Kyurkchiev, N. (2018). A new class of Gompertz-type software reliability models. International Electronic Journal of Pure and Applied Mathematics, 12, 43–57.

Rządkowski, G., Głączewska, I., & SawIńska, K. (2015). The Gompertz function and its applications in management. Foundations of Management, 7, 185–190. doi:10.1515/fman-2015-0035

Satoh, D. (2000). A discrete Gompertz equation and a software reliability growth model. IEICE Transactions, E83-D(7), 1508–1513.
Satoh, D. (2001). A discrete Bass model and its parameter estimation. *Journal of the Operations Research Society of Japan*, 44, 1–18. doi:10.15807/jorsj.44.1

Satoh, D., & Uchida, M. (2010). Computer worm model describing infection via e-mail. *Bulletin of the Japan Society for Industrial and Applied Mathematics*, 20, 50–55. (in Japanese).

Satoh, D., & Yamada, S. (2001). Discrete equations and software reliability growth models, Proceedings of 12th international symposium on software reliability engineering (pp. 176–184). Hong Kong.

Satoh, D., & Yamada, S. (2002). Parameter estimation of discrete logistic curve models for software reliability assessment. *Japan Journal of Industrial and Applied Mathematics*, 19, 39–53. doi:10.1007/BF03167447

Yamada, S. (1992). A stochastic software reliability growth model with Gompertz curve. *Transactions of Information Processing Society of Japan*, 33, 964–969. (in Japanese).

Yamada, S. (2014). *Software reliability modeling—fundamentals and applications*. Tokyo: Springer.

Yamada, S., Inoue, S., & Satoh, D. (2002). Statistical data analysis modeling based on difference equations for software reliability assessment. *Transactions of the Japan Society for Industrial and Applied Mathematics*, 12, 155–168. (in Japanese).

Yamada, S., & Tamura, Y. (2016). OSS reliability measurement and assessment. Switzerland: Springer.

Yamakawa, P., Rees, G. H., Solas, J. M., & Alva, N. (2013). The diffusion of mobile telephones: An empirical analysis for Peru. *Telecommunications Policy*, 37, 594–606. doi:10.1016/j.telpol.2012.12.010

Zhang, Y., Wang, Y., Yang, X., Zhang, Y., & Lv, H. (2014). Empirical study on the growth model and the division of lifecycle phases about the Chinese construction industry. *The Open Construction and Building Technology Journal*, 8, 369–376. doi:10.2174/1874836801408010369

© 2019 The Author(s). This open access article is distributed under a Creative Commons Attribution (CC-BY) 4.0 license.

You are free to:
- Share — copy and redistribute the material in any medium or format.
- Adapt — remix, transform, and build upon the material for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:
- Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made.
- No additional restrictions — You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.