MATHEMATICALLY MODELLING HCG IN WOMEN WITH GESTATIONAL TROPHOBLASTIC DISEASE USING EXPONENTIAL INTERPOLATION

Catherine Costigan
Sabin Tabirca
School of Mathematics
University College Cork

John Coulter
Cork University Hospital
University College Cork

Ernest Scheiber
Department of Informatics
Transylvania University of Brasov
Romania

KEYWORDS

Mathematical Modelling; Gestational Trophoblastic Disease; Logarithmic Transformation; hCG; Exponential Interpolation

ABSTRACT

Exponential interpolation is a problem in mathematics which can have very useful results in many areas. One area which is particularly of interest here is using exponential interpolation to model human chorionic gonadotropin (hCG) levels in women with gestational trophoblastic disease as it has been previously reported that the hCG measurements in these women follow an exponential curve.

INTRODUCTION

Interpolation is the process of estimating the value of a function, $f(x)$, whose value is known at specific points, say $x_0 < x_1 < \ldots < x_n$, by a second function $g(x)$ so that $f(x_i) = g(x_i)$ for each $i \in \{0,1,\ldots,n\}$. The values of the function $f(x)$ at other points in the interval (x_0,x_n) can then be estimated using the function $g(x)$. In most interpolation schemes, the function $g(x)$ is found to be unique and either an equation or an algorithm can be used to calculate it. For example, the linear, polynomial and spline interpolation methods provide nice equations for the unique function $g(x)$. Furthermore, theoretical results are also available for these methods to evaluate the estimation error of $\|f(x) - g(x)\|$ (see (Phillips, 2003) for a good review of interpolation methods).

Interpolation is used in many different areas including economics, engineering and medical simulation for a variety of reasons. One reason that interpolation is used is if the actual function $f(x)$ is known explicitly but is too complicated to use in a computer programme. In this case an interpolating function $g(x)$ that agrees closely with the actual function is used. Another situation where interpolation is used is when the value of the function $f(x)$ is only known at certain points. In this case an interpolating function $g(x)$ is used so that the function can be estimated at other points (Szidarovszky & Yakowitz, 2013).

Exponential interpolation considers the function $g(x)$ as a linear combination of exponentials

$$g(x) = B + \sum_{i=1}^{n} A_i \cdot e^{\alpha_i \cdot x}. \quad (1)$$

In this case, the *Exponential Interpolation* problem can be formulated as follows:

Definition 1. *Given a function $f : [a,b] \to \mathbb{R}$ which is twice differentiable and $2n+1$ points*

$$a \leq x_0 < x_1 < \ldots < x_{2n+1} \leq b$$

then find the function $g(x)$ given by equation 1 so that $f(x_i) = g(x_i)$, $i = 1,2,...,2n+1$.

Usually, in exponential interpolation, it is requaired to find a method to calculate the $2n+1$ coefficients, B and $A_i, \alpha_i, i = 1,2,\ldots,2n$, and then to prove that these coefficients are unique. It is also useful to have an evaluation for the interpolation error $|f(x) - g(x)|$.

Exponential Interpolation, often called interpolation with Dirichlet polynomials, has been investigated quite intensively for the last decades in order to solve the above problems (Ammar, Dayawansa, & Martin, 1991), however the interpolation methods proposed are either difficult to implement or computational expensive. More effective approaches can be developed by some direct approaches especially in some particular cases of equation 1.

One particular form is given by $g(x) = A \cdot e^{-\alpha \cdot x} + B$, which occurs in various technical and medical decay problems. A problem of interest for us is the variation of the hCG marker for Gestational Trobhoblastic Disease (GTD). GTD is a term used to describe a range of illnesses in women, such as complete and partial hydatidiform mole, choriocarcinoma, placental cite trophoblastic tumour and epithelioidtrophoblast tumour. All of these conditions arise from human trophoblastic tissue (Young et al., 2011). Upon diagnosis, the hydatidiform mole

Proceedings 30th European Conference on Modelling and Simulation ©ECMS Thorsten Claus, Frank Herrmann, Michael Manitz, Oliver Rose (Editors)
ISBN: 978-0-9932440-2-5 / ISBN: 978-0-9932440-3-2 (CD)

should be surgically evacuated. Usually, after this procedure the hCG levels will drop exponentially (Schoeberl, 2007). Interpreting the drop in hCG levels is a useful way to decide on a treatment plan for the patients, and to decide if the use of chemotherapy is required (Schoeberl, 2007).

The aim of this study is to mathematically model hCG level in women with gestational trophoblastic disease using Exponential Interpolation. It has been suggested previously in (Almufti et al., 2014) that in the future, mathematical models may be used to describe tumour biomarker production. This could be in relation to tumour size, the effectiveness of treatment and biomarker elimination. It was even thought that based on a few time points for each patient, mathematical models may enable early prediction of changes in disease burden and treatment efficacy, , however little concrete research work has been done in this direction. It was reported in (You et al., 2010) and (You et al., 2013) that the model that hCG follows after evacuation is of the form

$$g(t) = Ae^{-\alpha t} + B. \tag{2}$$

Firstly, this article investigates the exponential interpolation based on equation 2 in order to calculate the coefficients A, B and α. Then we will prove that these coefficients are unique and will evaluate the interpolation error. Finally, we present how this interpolation can be applied to predict the evolution of the hCG marker in GTD.

EXPONENTIAL INTERPOLATION FOR $n = 1$

The interpolation process needs to find the exponential function

$$g(t) = Ae^{-\alpha t} + B \tag{3}$$

that goes through three points (x_i, y_i) e.g. $g(x_i) = y_i,\ i = 0, 1, 2$. We assume that

$$x_0 < x_1 < x_2 \Rightarrow y_0 > y_1 > y_2. \tag{4}$$

Proposition 1. *Given the points (x_0, y_0),(x_1, y_1) and (x_2, y_2), the parameters A and B can be calculated using the formulas so that $y_i = g(x_i)$ for each $i \in \{0, 1, 2\}$.*

$$A = \frac{y_1 - y_0}{e^{-\alpha x_1} - e^{-\alpha x_0}}, \quad B = y_0 - Ae^{-\alpha x_0} \tag{5}$$

and α can be found by finding a root of the equation

$$\frac{e^{-\alpha(x_1 - x_0)} - 1}{e^{-\alpha(x_2 - x_0)} - 1} = \frac{y_1 - y_0}{y_2 - y_1} \tag{6}$$

Proof. Insisting the function passes through the points $\{x_0, y_0\}$,$\{x_1, y_1\}$ and $\{x_2, y_2\}$ means the the equations

$$\begin{aligned} Ae^{-\alpha x_0} + B &= y_0 \\ Ae^{-\alpha x_1} + B &= y_1 \\ Ae^{-\alpha x_2} + B &= y_2 \end{aligned} \quad . \tag{7}$$

must be satisfied. Solving for B gives $B = y_0 - Ae^{-\alpha x_0}$ and substituting it into the second equation to obtain $A = \frac{y_1 - y_0}{e^{-\alpha x_1} - e^{-\alpha x_0}}$. To get the equation for α both of these can be substituted into the third equation to give

$$\frac{y_1 - y_0}{e^{-\alpha x_1} - e^{-\alpha x_0}} e^{-\alpha x_2} + y_0 - \frac{y_1 - y_0}{e^{-\alpha x_1} - e^{-\alpha x_0}} e^{-\alpha x_0} = y_2.$$

Tidying this up gives

$$\frac{e^{-\alpha(x_1 - x_0)} - 1}{e^{-\alpha(x_2 - x_0)} - 1} = \frac{y_1 - y_0}{y_2 - y_0} \tag{8}$$

as desired. □

Theorem 1. *Equation 6 has a unique solution for α.*

Proof. From equation 2 $g(x) = Ae^{-\alpha x} + B$. Then

$$g'(x) = -A\alpha e^{-\alpha x} \tag{9}$$

From this it is clear that as long as $A \cdot \alpha \neq 0$ then $g(x)$ is strictly monotone. Since for $x_0 < x_1 < x_2 \Rightarrow y_0 > y_1 > y_2$ we have that $g(x)$ is decreasing i.e. $A \cdot \alpha > 0$. Concentrating on this case, let

$$p = x_1 - x_0 > 0 \tag{10}$$

$$q = x_2 - x_0 > 0, q > p \tag{11}$$

$$k = \frac{y_1 - y_0}{y_2 - y_0} \in (0, 1) \tag{12}$$

Now, equation 6 can be rewritten in terms of p, q and k to give

$$m(\alpha) = \frac{e^{-p\alpha} - 1}{e^{-q\alpha} - 1} - k = 0. \tag{13}$$

The function $m(\alpha)$ can be extended at 0 by $m(0) = \frac{p}{q} - k$ so that it is continuous and differentiable over $\mathbb{R}$. By looking at the behaviour of $m(\alpha)$ it can be shown that a unique solution for $m(\alpha) = 0$ exists.

$$\lim_{\alpha \to \infty} m(\alpha) = 1 - k > 0 \tag{14}$$

and

$$\lim_{\alpha \to -\infty} m(\alpha) = -k < 0. \tag{15}$$

As a consequence of this, a solution of equation 8 can be found. In order to show that the solution for α is unique it will be shown that the equation $m(\alpha)$ is strictly monotone. Finding the derivative of $m(\alpha)$ gives

$$m'(\alpha) = e^{(q-p)\alpha} \cdot \frac{e^{p\alpha} - 1}{e^{q\alpha} - 1} \cdot \left(\frac{p}{e^{p\alpha} - 1} - \frac{q}{e^{q\alpha} - 1} \right). \tag{16}$$

Multiplying the top and bottom by α gives

$$m'(\alpha) = e^{(q-p)\alpha} \cdot \frac{e^{p\alpha} - 1}{e^{q\alpha} - 1} \cdot \left(\frac{p\alpha}{e^{p\alpha} - 1} - \frac{q\alpha}{e^{q\alpha} - 1} \right) \cdot \frac{1}{\alpha}. \tag{17}$$

Examining each part of $m'(\alpha)$ to check to see if $m'(\alpha)$ is positive or negative, since $q > p$ it is clear that $e^{(q-p)\alpha}$ and $\frac{e^{p\alpha}-1}{e^{q\alpha}-1}$ are always positive. $\frac{1}{\alpha}$ is positive for positive

α and negative for negative α so it just remains to find the sign of $\frac{p\alpha}{e^{p\alpha}-1} - \frac{q\alpha}{e^{q\alpha}-1}$.

Looking at the function $\psi(t) = \frac{t}{e^t-1}$, $\psi'(t) = \frac{1-e^t(t-1)}{(e^t-1)^2}$, we can find that $\psi'(t)$ is always negative as the numerator is negative and the denominator is positive. This implies that the function $\psi'(t)$ is decreasing. This means that $\frac{p\alpha}{e^{p\alpha}-1} - \frac{q\alpha}{e^{q\alpha}-1}$ is positive for positive α and negative for negative α.

So overall $m'(\alpha) > 0$ for all values of α, which means that $m(x)$ is strictly increasing. So $m(\alpha)$ is injective so there is a unique solution to equation 6. □

Theorem 2. *Given a function $f : [a,b] \to \mathbb{R}$ n times differentiable and n points $\{x_i\}_{i=0}^{n-1}$ so that $f(x_i) = 0$ for all $i \in \{0,\ldots,n-1\}$, then for all $x \in (a,b)$ there is a number $c \in (a,b)$ so that*

$$f(x) = \frac{f^{(n)}(c)}{n!}\prod_{i=0}^{n-1}(x-x_i) \tag{18}$$

This is just a consequence of the interpolation error theorem (Phillips, 2003) in which the interpolation polynomial is 0 since all the values $f(x_i) = 0$ are 0.

Proposition 2. *Given a function $f : [a,b] \to \mathbb{R}$ 3 times differentiable and $g(x)$ the exponential function defined in equation 2 so that $f(x_i) = g(x_i) = y_i,\ i = 0,1,2$ (with A, B and α satisfying equations 7). Then the interpolation error is provided by:*

$$|f(x)-g(x)| < \frac{\max_{x\in(a,b)}|f'''(x)+A\alpha^3e^{-\alpha t}|}{6}|(x-x_1)(x-x_2)(x-x_3)| \tag{19}$$

Proof. Define the function

$$m(x) = f(x) - g(x) \tag{20}$$

with the properties $m(x)$ is 3 times differentiable and $m(x_0) = m(x_1) = m(x_2) = 0$. By theorem 2 there is some $c \in (x_0,x_2)$ so that

$$m(x) = \frac{m'''(c)}{6}(x-x_0)(x-x_1)(x-x_2) \tag{21}$$

This means that

$$|f(x)-g(x)| < \frac{\max_{x\in(x_0,x_2)}|m'''(x)|}{6}|(x-x_0)(x-x_1)(x-x_2)| \tag{22}$$

□

Remark 1. *As a consequence of proposition 2*

$$|f(x)-m(x)| < \frac{\max_{x\in(x_0,x_2)}|f'''(x)|+A\alpha^3e^{-\alpha x_0}}{6}|(x-x_0)(x-x_1)(x-x_2)| \leq \frac{\max_{x\in(x_0,x_2)}|f'''(x)|+A\alpha^3}{6}|(x-x_0)(x-x_1)(x-x_2)|. \tag{23}$$

From experiments it was found that the value of A is very large, at least in the order of 10^3. So this error term is not small enough to ensure that the errors would be sufficiently small when using this method.

IMPLEMENTING EXPONENTIAL INTERPOLATION FOR $n = 1$

The method described in proposition 1 was implemented in C# in order to fit the model 2 to the data. The Newton-Raphson method was used on equation 6 in order to find the root. This method is an iterative root finding technique that uses an initial guess together with the value of the function at that point and its derivative at the point in order to make a better approximation to the root. The $n+1^{th}$ guess is found using the formula

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \tag{24}$$

The process continues until the difference $x_{n+1} - x_n < \varepsilon$, where ε is the accuracy of the root desired. On implementing the Newton-Raphson method, the values of the derivatives of 6 were estimated numerically at each iteration (Szidarovszky & Yakowitz, 2013) (Press, Teukolsky, Vetterling, & Flannery, 1992).

Algorithm 1 Finding A, B and α using Interpolation

input: $x_0, y_0, x_1, y_1, x_2, y_2$
α = newtonraphson(equation 8)
A = findA(α) {using equation 5}
B = findB(α, A) {using equation 5}
output: A,B, α

In order to test if the method of interpolation works well for fitting a curve to data points some points were chosen from a curve in the form of 2 with known values for A, B and α. In this case, the values for A, B and α chosen were 1000, 7 and 0.5 respectively. Using the method described in algorithm 1 the parameters

$$\begin{aligned} A &= 999.999 \\ B &= 7.0009 \\ \alpha &= 0.5000 \end{aligned}$$

were found. This shows that this method for interpolation works well. The error in the estimated parameters is negligible.

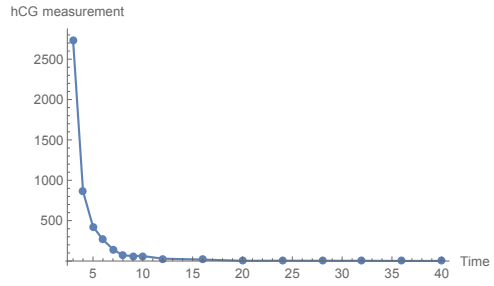

Figure 1: hCG measurements for a Typical Patient

MODELLING hGC MARKER FOR GTD

Upon diagnosis, the hydatidiform mole should be surgically evacuated. Usually, after this procedure the hCG levels will drop exponentially (Schoeberl, 2007). Sometimes, however, after a possible drop at the start the hCG levels will plateau or rise again. Sustained high levels of hCG are an identifier of gestational trophoblastic neoplasia (GTN), these high levels of hCG have also been an indicator for the need for chemotherapy (Schoeberl, 2007). Interpreting the drop in hCG levels is a useful way to decide on a treatment plan for the patients, and to decide if the use of chemotherapy is required (Schoeberl, 2007).

Each woman in the UK, and now in Ireland, diagnosed with GTD should be registered with a specialist centre for hCG surveillance as it is necessary to monitor hCG levels in these women since it is known that about 15% of women with a complete hydatidiform mole will require chemotherapy (Seckl, Sebire, & Berkowitz, 2010). Currently in these centres in the UK hCG is measured every 2 weeks following evacuation until the values are within normal range. After this, hCG concentrations are measured monthly for 6 months from the evacuation date (Seckl et al., 2010). Therefore, a number of blood tests (usually about 10 - 14) are conducted fortnightly in order to asses the evolution of the hCG marker. In most patients' cases, these hCG levels are decreasing suggesting that the chemotherapy treatment is not needed. In practice it often happens that a patient would miss some blood tests so that there may be a discontinuity of the data. In this case, we can apply exponential interpolation to best fit the curve $g(x) = Ae^{-\alpha x} + B$ through the available data, in order to estimate the missing data and to predict the future data.

A TYPICAL PATIENT

The hCG measurements for a typical patient are shown in figure 1.

Assume the patient has missed her blood test at week 5, but the results are known for week 3, 5 and 6. Using interpolation to find the coefficients A, B and α in the model 2, the values found are

$$\begin{aligned} A &= 152,149.769 \\ B &= 222.544 \\ \alpha &= 1.368. \end{aligned}$$

Using these values in the model 2 it can be shown that this corresponds to a value of 385.36 at week 5, which means the relative error is 0.09. This is a good approximation to the actual value.

Another situation in which interpolation might be used is in predicting future measurements. Say the hCG measurements for weeks 3, 4 and 5 are known and an approximation for the measurement for week 6 is required. In this case the coefficients found are

$$\begin{aligned} A &= 193,159.163 \\ B &= 293.621 \\ \alpha &= 1.457. \end{aligned}$$

The actual hCG level for week 6 is 264, using this technique the approximation is 324.45, which corresponds to a relative error of 0.22. This is larger than the error for the previous test but this is to be expected as interpolation works best for approximating values within the interval used to generate the curve.

TESTING ON MULTIPLE WOMEN

Firstly, the second, third and fifth points from each data set were used for interpolation and then the fourth point was used to check the size of the error, the same way as in the previous section. The results of this can be seen in table 1. Each row in the table represents a different womans data. The actual value is the fourth measured value of hCG in the blood of each patient. The error represents the difference between the actual fourth measurement and the value predicted using the curve that uses the values of A, B and α found using interpolation for each data set. The percentage error is the absolute value of the error, divided by the actual measurement. It can be seen that 7 out of the 21 points have a percentage error of less than 0.15. This seems to be reasonable, however, some of the points have very large errors, with 5 of the data points having percentage errors of more than 1. These errors are most likely due to the fact that the hCG levels drop exponentially fast so measurements taken in later weeks are small. This means that even a small deviation from the curve will result in a large relative error.

A test was then carried out using the second, third and forth data points to find the values of the coefficients for the model 2, and then estimating the value for the fifth measurement. The results are shown in table 2.

data set	actual value	error	relative error
1	426	-40.64	0.09
2	1,250	-97.73	0.07
3	236	68.08	0.28
4	56	-11.91	0.21
5	5	-0.62	0.12
6	163	-21.32	0.13
7	295	-36.79	0.12
8	1,979	-770.59	0.38
9	388	1,003.83	2.58
10	34	-4.96	0.14

Table 1: Errors for predicting the fourth hCG measurement using interpolation

data set	actual value	error	relative error
1	264	60.45	0.22
2	357	158.42	0.44
3	290	-82.11	0.28
4	17	14.98	0.88
5	3	1.2	0.4
6	63	28.81	0.45
7	213	45.65	0.21
8	504	840.41	1.66
9	113	-113	1
10	26	5.312	0.20

Table 2: Errors for predicting the fifth hCG measurement using interpolation

Some women had measurement of 0 for their fifth measurement. These women were left out of this test. The results of this experiment are not quite as good as the previous experiment. This is to be expected however. Even though a lot of the cases have large errors, some of the cases have small enough errors, for instance case 14 has a relative error of just 0.2.

CONCLUSION

In this study interpolation was implemented. Some nice results were found about using interpolation to calculate values of A, B and α for the model 2 including a proof that a unique solution for the values of A, B and α exists.

The method described in this study was tested on simulated data and it works very well for finding the values of A, B and α. Estimating missing hCG measurements in the data was quite successful, with relative errors as small as 0.07 in the data sets tested. This method was not as accurate for predicting measurements outside of the values used to calculate the coefficients, In saying that, relative errors as small as 0.2 were observed.

There are two reasons the relative errors for predicting future measurements are larger than the relative errors for estimating missing measurements. The first is that interpolation is known to be more accurate when estimating values within the interval used to calculate the coefficients. The other is that since the hCG levels are smaller for later weeks, the relative errors are larger for these values even if the absolute error stays the same.

References

Almufti, R., Wilbaux, M., Oza, A., Henin, E., Freyer, G., Tod, M., ... You, B. (2014). A critical review of the analytical approaches for circulating tumor biomarker kinetics during treatment. *Annals of oncology*, *25*(1), 41–56.

Ammar, G., Dayawansa, W., & Martin, C. (1991). Exponential interpolation: theory and numerical algorithms. *Applied Mathematics and Computation*, *41*(3), 189–232.

Phillips, G. M. (2003). *Interpolation and approximation by polynomials* (Vol. 14). Springer Science & Business Media.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1992). *Numerical recipes: The art of scientific computing (cambridge.* Cambridge Univ. Press.

Schoeberl, M. R. (2007). A model for the behavior of β-hcg after evacuation of hydatidiform moles. *Gynecologic oncology*, *105*(3), 776–779.

Seckl, M. J., Sebire, N. J., & Berkowitz, R. S. (2010). Gestational trophoblastic disease. *The Lancet*, *376*(9742), 717–729.

Szidarovszky, F., & Yakowitz, S. J. (2013). *Principles and procedures of numerical analysis* (Vol. 14). Springer.

You, B., Harvey, R., Henin, E., Mitchell, H., Golfier, F., Savage, P., ... Seckl, M. (2013). Early prediction of treatment resistance in low-risk gestational trophoblastic neoplasia using population kinetic modelling of hcg measurements. *British journal of cancer*, *108*(9), 1810–1816.

You, B., Pollet-Villard, M., Fronton, L., Labrousse, C., Schott, A.-M., Hajri, T., ... others (2010). Predictive values of hcg clearance for risk of methotrexate resistance in low-risk gestational trophoblastic neoplasias. *Annals of oncology*, *21*(8), 1643–1650.

Young, T., Coleman, R., Hancock, B., Drew, D., Wilson, P., Tidy, J., et al. (2011). Predicting gestational trophoblastic neoplasia (gtn): is urine hcg the answer? *Gynecologic oncology*, *122*(3), 595–599.