Fitting the Portuguese population from 1850 to 2010 to a logistic growth model

Carlos Francisco Barbosa[1], Michael Rothwell[1]

[1]Oporto British School, Porto, Portugal

ABSTRACT

This work explores how the Portuguese population fits a logistic growth model. The present study is divided into two main sections. The first one consists of the qualitative and quantitative study of the logistic equation. Qualitatively, I will look at various aspects of the differential equation, such as the equilibria and their stability and possible inflections of solutions. Quantitatively, I will use the separation of variables to find explicit solutions. Given the lack of accuracy in the linear fitting to the proportional growth rate against the population, the rate of growth being proportional to the current population. In the second chapter, I attempted a polynomial trendline fitting to the growth rate against the population. This led the focus to creating an adapted form of the logistic curve that fits the Portuguese population from 1850 to 2010. With a certain degree of accuracy, the adapted form of the logistic growth model fits the Portuguese population in the period mentioned.

Introduction

My interest for the logistic equation stemmed from my research in Economics. The logistic function finds applications in a large range of fields including Demography (Petrescu, 2009; Barrocas, 2001). Population changes, in the field of Economics, dictate the amount of investments, consumption and government expenditure, essentially internal demand, which is a key component in an economy because of its strong correlation with the GDP (Blink & Dorton, 2012). It is therefore crucial for an economist to be aware of such a huge internal influence. During my research on this topic, I came across another investigation from 1945, by a Portuguese mathematician named Joaquim José Paes Moraes (1914-2008) (INE 2008), entitled "Sobre o acerto duma logística à população portuguesa" (Moraes, 1945), that modelled the Portuguese population using the logistic model. This study motivated me further into exploring this topic. Also, another factor that contributed to this was the global problem of population growth due to the increasing global population that results in scarcer resources and potential living space problems, all of these made me want to look further into the population of my home country, Portugal.

The English demographer and political economist Thomas Malthus was one of the first to express the idea on a model for populations, commonly referred to as the exponential growth model (Seidl & Tisdell, 1999). He alleged that populations followed exponential increases if there was an abundance of resources this would lead ultimately to a lack of resources. At the time, this was a counter argument to the idea that the industrial revolution would bring an end to poverty and therefore quite controversial. This triggered the appearance of the logistic growth model by the Belgian mathematician Pierre François Verhulst in 1838 (Verhulst, 1838, 1845, 1847). Verhulst derived his logistic equation after he had read "An Essay on the Principle of Population" by Malthus (Kucharavy & De Guio, 2011).

The logistic differential equation

The first population models were represented by an exponential growth with no limiting factor, the rate of growth being proportional to the current population, where $\frac{dP}{dt}$ is the proportional growth rate and r is the proportionality constant, as it is depicted in Figure 1.
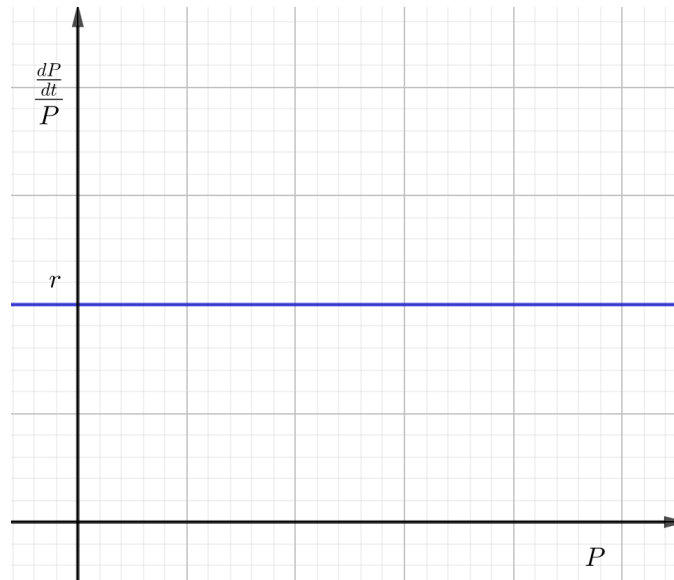

Figure 1. Graph showing constant proportional growth rate.

The rate of growth being proportional to the current population is given by,

$$\frac{\frac{dP}{dt}}{P} = r \qquad\qquad \text{[Eq. 1]}$$

Or, equivalently,

$$\frac{dP}{dt} = rP \qquad (if\ P \neq 0) \qquad\qquad \text{[Eq. 2]}$$

Note that, $P(t) = 0$ is a solution of Eq. 2 but not from Eq. 1. However, $P(t) = 0$ is an important solution since in general wherever $\frac{dy}{dx} = f(y)$ then the solutions of the equation $f(y) = 0$, say, y^*, will generate constant solutions namely $y(x) = y^*$ as $\frac{dy}{dx} = 0$. These solutions are denoted as *equilibria*. To solve the equation, we will use separation of variables,

$$\int \frac{1}{P}\frac{dP}{dt}\,dt = \int r\,dt$$
$$\int \frac{1}{P}dP = \int r\,dt$$

Since P is a population, it is always positive therefore,

$$\ln P = rt + c$$
$$P = e^{rt+c}$$
$$\text{Letting } A = e^c, \text{we have } P = Ae^{rt}$$
$$(P' = Are^{rt} \therefore P' = rP)$$

However, most populations are restrained to certain resources that limit their growth, therefore, instead of the per capita growth rate being constant we will assume that it will decrease linearly as P grows, this is represented by the equation below (Figure 2), *i.e.,*

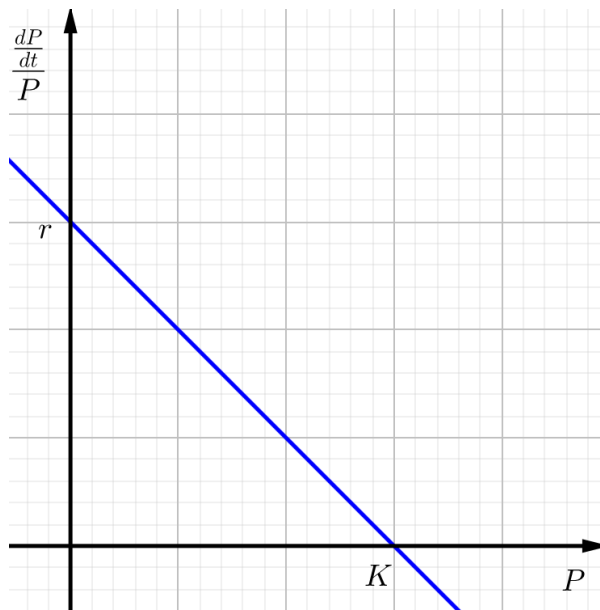$$\frac{\frac{dP}{dt}}{P} = -\frac{r}{K} \cdot P + r \qquad \text{[Eq. 3]}$$

Figure 2. Graph showing a linearly decreasing proportional growth rate.

Using the equation of graph above (Eq. 3) we have,

$$\frac{\frac{dP}{dt}}{P} = -\frac{r}{K} \cdot P + r$$

$$\therefore \frac{dP}{dt} = rP\left(1 - \frac{P}{K}\right), \quad \text{(if } P \neq 0\text{)} \qquad \text{[Eq. 4]}$$

Again, $P(t) = 0$ is a solution of Eq. 4 but not from Eq. 3. Also, when $P = K$, $\frac{dP}{dt} = 0$, therefore will be a point of equilibrium. This can be summarised in the diagram displayed in Figure 3, which is called phase line (MIT, 2011).

Figure 3. Phase line (MIT, 2011) to illustrate equilibria of Eq. 3.

The phase line (MIT, 2011) depicts the asymptotic behaviour, as $t \to +\infty$, for all solutions. The behaviour of the solutions is the following: $P(t) = 0$ is an unstable equilibrium, also referred to as a *repeller* (University of Arizona, 2008), since all solutions near this point move away from it. $P = K$, the point of equilibrium, is usually called *carrying*

capacity (Britannica, 2019) because $f(P) = rP\left(1 - \frac{K}{P}\right)$ is 0 when $P = K$, for the following reason, if $P > K$ (in Eq. 4), then $\frac{dP}{dt} < 0$ and therefore the solutions are decreasing. On the other hand, if $0 < P < K$, then $\frac{dP}{dt} > 0$ and therefore the solutions are increasing. Since, $\frac{dP}{dt} = 0$ at $P(t) = K$, then all other solutions must converge to K. For $P(t) = K$, we can conclude all solutions will converge towards this point, that is $\lim_{t \to \infty} P(t) \to K$ [except $P(t) = 0$] and for this reason, it can be called a *global attractor* or *asymptotically stable* (University of Arizona, 2008).

Point of inflection

In order to better understand the behaviour of the solutions of the differential equation, we will look for any inflection points that are the zeros of the second derivative of P with respect to *t*.

From Eq. 4, shown in Figure 4,

$$\frac{dP}{dt} = rP - \frac{rP^2}{K}$$

$$\frac{d^2P}{dt^2} = \frac{d}{dt}\left(\frac{dP}{dt}\right) = \frac{d}{dt}(f(P)) = \frac{df}{dP} \cdot \frac{dP}{dt}$$

Hence, using the chain rule, the zeros of the second derivate of P are those of $\frac{dP}{dt}$ and of $\frac{df}{dP}$. Thus, the inflection points must be the critical points of f. In this case f is a parabola with zeros 0 and K, hence the vertex is $\frac{K}{2}$, as shown in Figure 4.
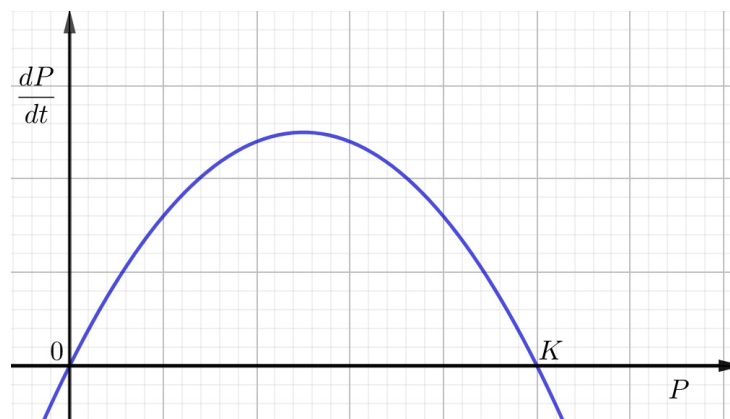

Figure 4. Parabola of the Logistic differential equation.

Nevertheless, the general procedure for, calculating the 0 of the second derivative of P with respect to *t*, for any given function f is given by,

$$\frac{d^2P}{dt^2} = \frac{d}{dt}\left(\frac{dP}{dt}\right) = \frac{d}{dt}\left(rP\left(1 - \frac{P}{K}\right)\right)$$
$$= \frac{d}{dt}\left(rP\left(1 - \frac{P}{K}\right)\right)$$
$$= r\frac{dP}{dt}\left(1 - \frac{P}{K}\right) + rP\left(-\frac{1}{k}\frac{dP}{dt}\right)$$
$$= \frac{dP}{dt}\left(r\left(1 - \frac{P}{K}\right) - r\frac{P}{K}\right)$$
$$= \frac{dP}{dt}\left(r\left(1 - \frac{2P}{K}\right)\right)$$

If $\frac{d^2P}{dt^2} = 0$, then, $\frac{dP}{dt} = 0$ or $r\left(1 - \frac{2P}{K}\right) = 0$. Since $\frac{dP}{dt} = 0$ corresponds to the equilibrium points, we will only consider $r\left(1 - \frac{2P}{K}\right) = 0$. Therefore, $r - 2r\frac{P}{K} = 0$, hence, $P = \frac{K}{2}$ as expected. The inflection point allows us to understand the point where the population changes from increasing at an increasing rate to increasing at a decreasing rate. If $P < \frac{K}{2}$ then $\frac{d^2P}{dt^2}$ is positive thus the graph of $P(t)$ is concave up, if $P > \frac{K}{2}$ then $\frac{d^2P}{dt^2}$ is negative so the graph of $P(t)$ concave down. The inflection illustrates the point where resources become scarcer and, thus, forces the population to increase at a slower rate. We can also conclude that all solutions where $P(0) > K$, P will always remain greater than K, therefore will not have any points of inflection.

Explicit solutions

In the present section, we will solve explicitly the logistic differential equation, using Eq. 4, the linearly decreasing proportional growth rate:

$$\frac{dP}{dt} = rP\left(1 - \frac{P}{K}\right)$$

Using separation of variables,

$$\frac{1}{P\left(1 - \frac{P}{K}\right)}\frac{dP}{dt} = r$$

$$\frac{K}{P(K-P)}\frac{dP}{dt} = r$$

$$\int \frac{K}{P(K-P)}\,dP = \int r\,dt$$

By using partial fractions,

$$\frac{K}{P(K-P)} = \frac{A}{P} + \frac{B}{K-P}$$

Where A and B are constants to be determined, thus,

$$K = A(K-P) + PB$$

When $P = 0$: $K = AK \therefore A = 1$ and when $P = K$: $K = KB \therefore B = 1$, thus,

$$\frac{K}{P(K-P)} = \frac{1}{P} + \frac{1}{K-P}$$

$$\int \frac{A}{P} + \frac{B}{K-P}\,dP = \int r\,dt$$

$$\int \frac{1}{P}\,dP + \int \frac{1}{K-P}\,dP = rt + c$$

As before, P is a population and so it is always positive, yet as P can be larger than K, $(K-P)$ can be negative.

$$\ln|P| - \ln|K-P| = rt + c$$

$$\ln\left|\frac{P}{K-P}\right| = rt + c$$

$$\left|\frac{P}{K-P}\right| = e^{rt}e^c \qquad \text{[Eq. 5]}$$

$$\frac{P}{K-P} = Se^{rt} \qquad \text{[Eq. 6]}$$

where $S = \pm e^c$ (hence, $S \in \mathbb{R}\backslash\{0\}$)

$$P = (K-P)Se^{rt}$$

$$P(1 + Se^{rt}) = KSe^{rt}$$

$$P(t) = \frac{KSe^{rt}}{1 + Se^{rt}}$$

In fact, we can allow S to be 0 since this would imply $P(t) = 0$, which is a constant solution of the equation as found before. However, the other constant solution $P(t) = K$ cannot be included in the same formula.

If we substitute $t = 0$ in Eq. 6, we obtain,

$$S = \frac{P(0)}{K - P(0)}$$

Let us denote the initial population $P(0) = P_0$, thus,

$$P(t) = \frac{K \left(\frac{P_0}{K - P_0}\right) e^{rt}}{1 + \left(\frac{P_0}{K - P_0}\right) e^{rt}}$$

$$= \frac{P_0 K e^{rt}}{K - P_0 + P_0 e^{rt}}$$

$$= \frac{P_0 K}{(K - P_0)e^{-rt} + P_0} \quad [\text{Eq. 7}]$$

Eq. 7 represents the explicit solution of the logistic differential equation. When $t \to \infty$ and $P_0 \neq 0$, $e^{-rt} \to 0$ and therefore,

$$\lim_{t \to \infty} P(t) = \frac{P_0 K}{P_0} = K$$

This can also be illustrated by the direction field (Figure 5). For this representation, I have used $K = 10$, $r = 1$ and, in order to emphasize the points of inflection, t was restricted to an interval from 0 to 10. Also highlighting the fact that K is the *carrying capacity* (Britannica, 2019) of the population.
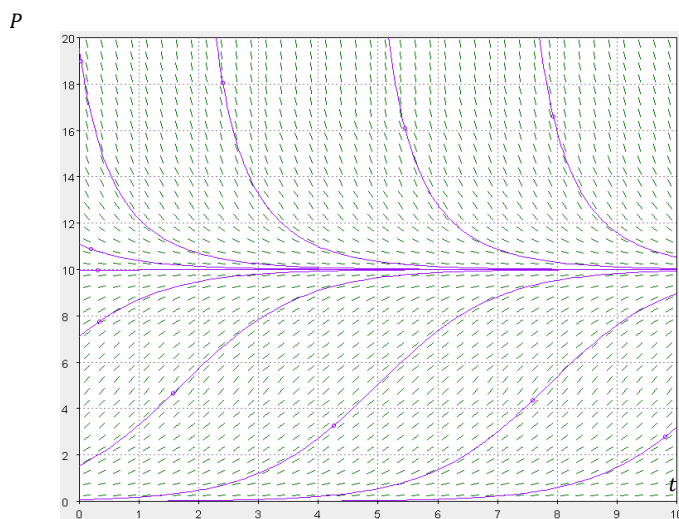

Figure 5. Direction field and graph of solutions for the Logistic differential equation.

Fitting a logistic curve to the Portuguese population from 1850 to 2010

To find out if the population could be modelled using the Logistic differential equation (Eq. 4), I will use a method that is based on one of the techniques used by Lipkin and Smith (2004). This will be an attempt to fit a linear function to the values of $\frac{\frac{\Delta P}{\Delta t}}{P}$ (the proportional growth). Since,

$$\frac{\frac{dP}{dt}}{P} = -\frac{r}{K}P + r$$

From the obtained line $y = mx + c$, we get $m = \frac{-r}{K}$, thus $K = \frac{-r}{m}$ and $c = r$. In order to approximate the values of $\frac{dP}{dt}$ we will use the symmetric differences, as sketched by Figure 6.
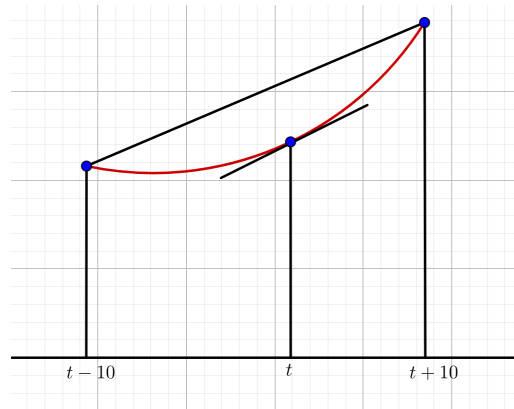
Figure 6. Estimation of the slope $\frac{dP}{dt}$ at a given point in time t.

For each particular value of *t* we use the following equation:

$$\frac{\Delta P}{\Delta t} = \frac{P(t+10) - P(t-10)}{20}$$

We are able to approximate the slope at time *t*, by using the value before and after, obviously, given data for the period between 1840 and 2010, we will be able to estimate the value of $\frac{dP}{dt}$ from 1850 to 2000 (Lipkin & Smith, 2004).

Using Microsoft® Excel (Microsoft Corp., USA), I will plot the ratios of the slope against the values of $P(t)$, as it is shown in the Figure 7.
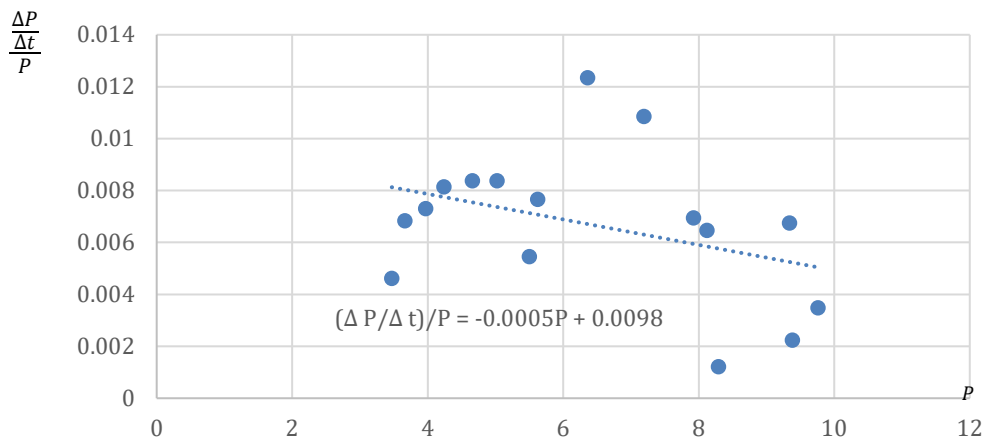

Figure 7. Linear fittings to the proportional growth rate against the population. The dots are the values of $(\Delta P/\Delta t)/P$ and the dotted line is the linear fit of $(\Delta P/\Delta t)/P$.

From the linear fit of Figure 6 we get $r = c = 0.0098$ and $K = \frac{-r}{m} = 19.6$. Visually we can see that this is not linear, this can also be seen by comparing the K value with the actual population values, thus we can conclude it is a poor approximation. In order to look for a better trend, once again using Microsoft® Excel (Microsoft Corp., USA), I plotted $\frac{\Delta P}{\Delta t}$ against P and looked for a quadratic fitting, given by excel, instead of plotting $\frac{\Delta P}{\Delta t}/P$ against P and use a linear trendline, as shown in Figure 8.
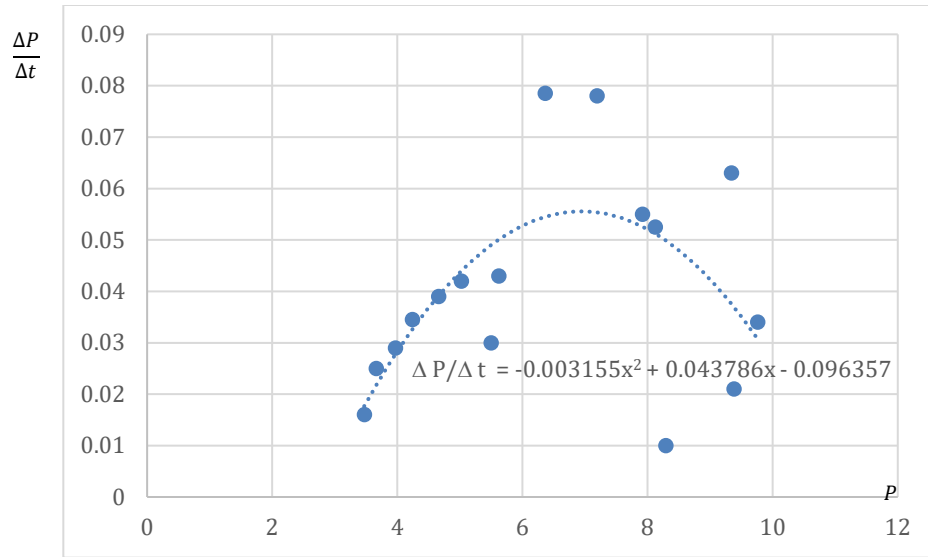
Figure 8. Polynomial trendline fitting to the growth rate against the population.

Since it was no longer possible to use the logistic form of the differential equation, we will use a adapted version. The limitation of using this method is that the equation of the polynomial trendline yields a new model of the form,

$$\frac{dP}{dt} = aP^2 + bP + c$$

with $c \neq 0$ and therefore, would move away from the logistic model as there is no constant in the logistic differential equation. Nevertheless, I rewrote this equation in the form,

$$\frac{dP}{dt} = r(P - A)\left(1 - \frac{P}{K}\right)$$

[Eq. 8]

to preserve similarities with Eq. 4 and so that we enhance the existence of two equilibria. the equilibria are the zeros, $A = 2.7426$ and $K = 11.136$, of the polynomial fit given by excel. Where A is the *lower carrying capacity* and K is the *carrying capacity*. Since the values of A and K satisfy the latter requirement and the quadratic fit in Figure 8 is better than the linear fit in Figure 7, I decided to change the aim and fit the Portuguese population to the adapted form of the logistic equation.

Note that if P is a solution of Eq. 8 then $Q = P - A$ is a solution of a logistic differential equation. Namely,

$$\frac{dQ}{dt} = \frac{dP}{dt} = r(P - A)\left(1 - \frac{P}{K}\right)$$

$$\frac{dQ}{dt} = rQ\left(1 - \frac{Q + A}{K}\right)$$

Since,

$$1 - \frac{Q + A}{K} = 1 - \frac{Q}{K} - \frac{A}{K} = \frac{K - A}{K} - \frac{Q}{K}$$

Thus,

$$\frac{dQ}{dt} = r\left(\frac{K - A}{K}\right)Q\left(1 - \frac{Q}{K - A}\right) = r^*Q\left(1 - \frac{Q}{K^*}\right)$$

Which is in the previously seen form of the logistic differential equation, Eq. 4. However, $r^* = r \cdot \frac{K-A}{K}$ and $K^* = K - A$. By using Eq. 7 and substituting $r*$ and $K*$ we have,

$$Q(t) = \frac{Q_0 K^*}{(K^* - Q_0)e^{-r^*t} + Q_0}$$

Thus,

$$P(t) = \frac{Q_0 K^*}{(K^* - Q_0)e^{-r^*t} + Q_0} + A$$

[Eq. 9]

$$= \frac{(P_0 - A) \cdot (K - A)}{(K - P_0)e^{-r^*t} + P_0 - A} + A$$

In order to find r^* we need to find r, using Eq. 8 in relation to the quadratic of Figure 8, $-rA = -0.096357$, thus $r = 0.035133$. Therefore, $r^* = r \cdot \frac{K-A}{K} = 0.026480$.

In Figure 9 we show the corresponding phase line (MIT, 2011) displaying the asymptotic behaviour of the solutions. I would expect that our solution curve to have a lower limit (as $t \to -\infty$) below the minimum data value used, 3.34, and an upper limit (as $t \to +\infty$) above the maximum data value used, 10.25.

Figure 9. Phase line (MIT, 2011) to illustrate equilibria of Eq. 8.

From the phase line we can see that A is a *repellor* (University of Arizona, 2008), since when $0 \le P < A$, $(P - A)$ will be negative thus, $\frac{dP}{dt} < 0$ and therefore the solutions are decreasing. Illustrating that the population will eventually face extinction, which is a situation not seen before in the logistic case. When $A < P < K$ then $\frac{dP}{dt} > 0$ and therefore the solutions are increasing. Equilibrium solution K will behave the same as in the logistic case, the *carrying capacity* and *global attractor* (University of Arizona, 2008).
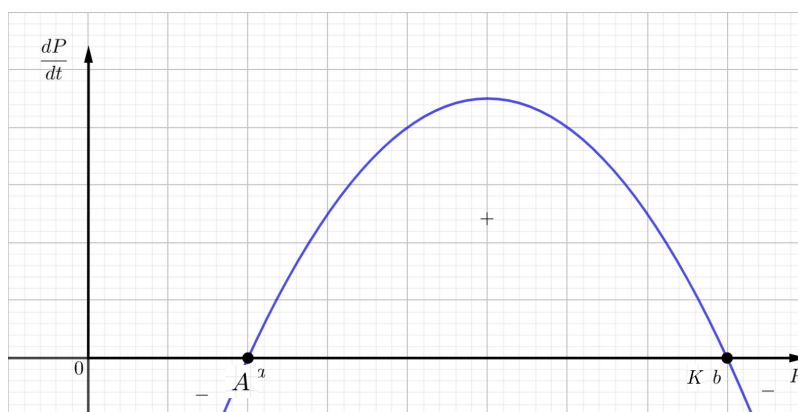

Figure 10. Parabola of the new model.

Using the same method as in Figure 4, in Figure 10 we can see that the new inflection point is attained at the maximum of the parabola, which is $P = \frac{A+K}{2}$. Therefore the inflection point is at $P = 6.9393$.

In Figure 11 we are now able to plot the graph of the solution, with the values of A, K, r^* found earlier in Eq. 9, $P_0 = 3.34$ and the initial condition t, where $t = 0$ at 1840. Thus,

$$P(t) = \frac{5.0142}{(7.796)e^{-0.026480t} + 0.5974} + 2.7426$$

[Eq. 10]

To facilitate the understating of Eq. 10 and the graph of $P(t)$ I will introduce a new variable T, where T is the year, therefore $t = T - 1840$ and $T = t + 1840$. Thus,

$$P(T) = \frac{5.0142}{(7.796)e^{-0.026480(T-1840)} + 0.5974} + 2.7426$$
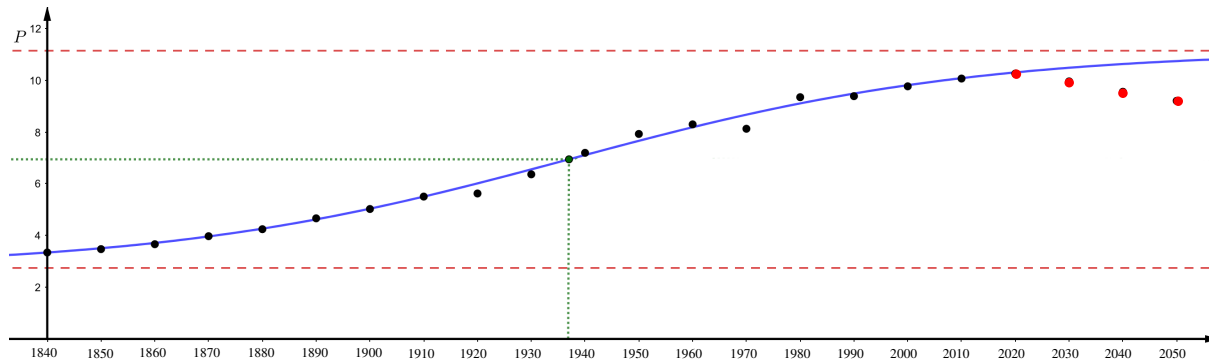
[Eq. 11]

Figure 11. Graph of my adapted form of the logistic model (Eq. 11). The blue solid curves correspond to the adapted form of the logistic equation, both red dashed lines correspond to upper and lower asymptotes, the green dot represents the point of inflection, the black dots correspond to the actual data from official sources and the red dots correspond to projected data also from official sources (see Appendix)

Note that I could have looked for other curves by using different initial conditions e.g. $P(0) = (1850)$ in order to look for a better fit. However, this was not necessary since this curve has a very good fit. Using the value of the inflection point found earlier ($P = 6.9393$), we can see that it lies between the year 1930 and 1940. As mentioned in the introduction, Moraes also attempted to fit a curve to the Portuguese population (but using data from 1840 to 1940) and, in my opinion, compared with his curve this model has a better fit.

Conclusion

In conclusion, I met the aim of the exploration since I was able to graph the adapted form of the logistic model with a very good degree of accuracy compared to the real population values. This adapted model of the logistic has only a good fitting for data ranging from1840 to 2010 and the projection of 2020, as from this point onwards the Portuguese population is expected to decrease and, therefore, this model does not illustrate population decreases. Also, other models could be fitted to the population, for example, the Gompertz model (Tjørve & Tjørve, 2017), which has a different function defining $\frac{dP}{dt}$. A potential improvement to my model would be to use the second method of the Lipkin and Smith (2004), which improves the initial model (based on approximating the derivative using symmetric differences) by fitting the logarithmic form of the solution (Eq. 5) to a linear function of t. I would also look at the interpretation of the new parameter A in Eq. 8, which is not part of the standard logistic equation.

Finally, this work has shown me that this fitting process is certainly not accurate, since it is too complex to capture all external influences, such as the 1929 world economic crisis, in one mathematical equation, as each of these external factors may have logistic equations to model their progress, but with different parameters. However, it is

remarkable that the adapted form of the logistic differential equation gives a curve that models, with a great degree of accuracy, the Portuguese population for 170 years.

Acknowledgments

References

Barrocas, J.M. (2001). Logística (curva). In *Enciclopédia Verbo Luso-Brasileira de Cultura* (Século XXI ed., pp. 213-214). Editorial Verbo.

Blink, J. and Dorton, I. (2012). *Economics* (2nd ed.), Oxford University Press.

Britannica, T. Editors of Encyclopaedia (2019). Carrying capacity. Encyclopedia Britannica. https://www.britannica.com/science/carrying-capacity

INE – Instituto Nacional de Estatística. (2008). *Revista de Estudos Demográficos*, *44*, p. 3. https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_publicacoes&PUBLICACOESpub_boui=45180054&PUBLICACOESmodo=2

INE – Instituto Nacional de Estatística. (2014). População residente em Portugal com tendência para diminiuição e envelhecimento. *Destaque – Informação à Comunicação Social*, p. 7. https://www.ine.pt/ngt_server/attachfileu.jsp?look_parentBoui=218948085&att_display=n&att_download=y

INE – Instituto Nacional de Estatística. (2016). População residente. *Estatísticas Demográficas 2015*, *75*, p. 22. https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_publicacoes&PUBLICACOESpub_boui=275533085&PUBLICACOEStema=55466&PUBLICACOESmodo=2

Kucharavy, D. & De Guio, R. (2011). Application of S-shaped curves. *Procedia Engineering*, *9*, 559–572.

Lipkin, L. & Smith, D. (2004). Logistic Growth Model. Mathematical Association of America. Retrieved January 15, 2019, from https://www.maa.org/press/periodicals/loci /joma/logistic-growth-model-introduction

MIT (2011). Phase Lines. Retrieved April 2021, from https://ocw.mit.edu/courses/mathematics/18-03sc-differential-equations-fall-2011/unit-i-first-order-differential-equations/first-order-autonomous-differential-equations/MIT18_03SCF11_s10_2text.pdf

University of Arozina (2008) . Retrieved April 2021, https://www.math.arizona.edu/~hhmi/Jim250AB/250b/250b_11_21Feb08.pdf

Moraes, J.J.P. (1945). Sobre o acerto da logística à população portuguesa. *Revista do Centro de Estudos Demográficos, 1*, 105-118.

Petrescu, E. (2009). A statistical distribution useful in product lifecycle modelling. *Management and Marketing*, *4*(2), 165-170. http://www.managementmarketing.ro/pdf/articole/145.pdf

Seidl, I. & Tisdell, C. A. (1999). Carrying capacity reconsidered: from Malthus' population theory to cultural carrying capacity. Ecological Economics, 31(3), 395–408. https://doi.org/10.1016/S0921-8009(99)00063-4

Tjørve, K.M.C. & Tjørve, E. (2017). The use of Gompertz models in growth analyses, and new Gompertz-model approach: An addition to the Unified-Richards family. PLoS ONE, 12(6), 1-17. https://doi.org/10.1371/journal.pone.0178691

Verhulst, P.-F. (1838). Notice sur la loi que la population suit dans son accroissement. In A. Quetelet (Ed.), *Correspondance Mathématique et Physique de l'Observatoire de Bruxelles*, (*4*th ed., pp. 113-121).[1]

Verhulst, P.-F. (1845). Recherches mathématiques sur la loi d'accroissement de la population. *Nouveaux Mémoires de l'Académie Royal des Sciences et Belles-Lettres de Bruxelles*, *18*, 1-42.[5]

University of Arozina (2008) MATH 250b – Section 2. Retrieved April, 2021, https://www.math.arizona.edu/~hhmi/Jim250AB/250b/250b_11_21Feb08.pdf
Verhulst, P.-F. (1847). Deuxième mémoire sur la loi d'accroissement de la population. *Mémoires de l'Académie Royal des Sciences, des Lettres des Beaux-Arts de Belgique*, *20*, 1-32.

The Free Encyclopedia (2019). Generalised logistic function. Retrieved January 16, 2019, from https://en.wikipedia.org/wiki/Generalised_logistic_function