By: Óscar Gómez

Inferring Steam owners through differential calculus

Introduction

Here's a basic explanation of my idea, it came to me when I was studying the behavior of charged particles in Semiconductors (I'm doing a physics degree). When certain semiconductors are stimulated by light pairs of charged particles start generating (holes and electrons, with positive and negative charge), however this charged particles tend to recombine, and the more charged particles there are in the semiconductor the faster they do so. This behavior can be explained through this differential equation:

$$\frac{d\Delta \mathrm{n}}{dt} = -\frac{\Delta n}{\tau} + G \qquad (1)$$

Where Δn represents the number of negative charged particles (it would be the same with positive particles), G represent the number of pairs generated by light and τ represents the average time a charged particle exists before recombination.

The world of quantum physics is chaotic, so a charged particle might exist in the semiconductor much longer than τ (or for a time much shorter), something similar happens with G the generation coefficient at a given time might be higher or lower. However, going back to the macroscopic world this equation has proven to give accurate predictions.

This equation can be applied to represent the total number of players using a game at a given time in Steam given the different habits that each person might have.

Methodology and result

Let's start with the first day, when the game comes out:

$$\frac{dP}{dt} = -\frac{P}{\tau} + G \qquad (2)$$

P represents the number of players at a time t, τ is the average time which players spend playing the video game every day and G is the generation rate of **new players.** I'll consider τ and G constants, even though I'm aware that more new players might generate at certain times of the day, this can be easily solved by transforming G into a periodic function (Fourier series) and the differential equation would still have an analytic solution (numeric resolution of differential equations is computationally intense).

The analytical solution of the equation will vary depending on the initial conditions, given the fact that Steam releases the average number of players of its games everyday I'll start by solving the equation for the first day and using it to calculate the number of owners, and then I'll solve the equations for the next days, which is more complex.

By: Óscar Gómez

Our objective is finding G, because knowing at what rate the new owners generate means we can calculate the total number of owners of the day:

$$O = \int_0^{t_{day}} G(t)\,dt \cong G \cdot t_{day} \qquad (3)$$

O represents the number of new owners. Even if G is a periodic function this integral should be easy to solve.

Solving (2) using the initial conditions of the first day, P(0)=0, gives us the following equation:

$$P(t) = G\tau\left(1 - e^{-\frac{t}{\tau}}\right) \qquad (4)$$

If you plot this equation you will see it is very similar to the P(t) of the first days in many games. Using this equation finding the average players per day is very easy:

$$\bar{P} = \frac{1}{t_{day}}\int_0^{t_{day}} P(t)\,dt = G\tau\left[1 + \frac{\tau}{t_{day}}\left(e^{\frac{-t_{day}}{\tau}} - 1\right)\right] \qquad (5)$$

Steam gives us the average number of players of the day so calculating G and then O is easy:

$$G = \frac{\bar{P}}{\tau}\left[1 + \frac{\tau}{t_{day}}\left(e^{\frac{-t_{day}}{\tau}} - 1\right)\right]^{-1} \qquad (6)$$

$$O \cong t_{day}\frac{\bar{P}}{\tau}\left[1 + \frac{\tau}{t_{day}}\left(e^{\frac{-t_{day}}{\tau}} - 1\right)\right]^{-1} \qquad (7)$$

I know that the total number of players is not exactly the same as the number of new owners, I don't exactly get why someone would buy a game and not play it even once but I'll discus some of these issues in the next section. For the time being let us consider that the total new players of the day equals the number of new owners.

For the next day it is necessary to separate P in (2) into two different groups, P_1 and P_0, the first one is assigned to players who play for the first time (AKA new owners) and previous owners who are playing for a second time:

$$\frac{dP}{dt} = \frac{dP_1}{dt} + \frac{dP_0}{dt} = -\frac{P_1}{\tau_1} + G - \frac{P_0}{\tau_0} + R_0 \qquad (8)$$

We have introduced R_0 which is the generation rate of old players, it is a replacement of what was their generation rate G on the first day, τ_1 and τ_0 would be the average playtimes, and G is the generation rate of new players.

This gives us a more complicated solution:

$$\bar{P} = G\tau_1\left[1 + \frac{\tau_1}{t_{day}}\left(e^{-\frac{t_{day}}{\tau_1}} - 1\right)\right] + \frac{\tau_0}{t_{day}}(P_0^0 - R_0\tau_0)\left(1 - e^{-\frac{t_{day}}{\tau_0}}\right) + R_0\tau_0 \qquad (9)$$

Where P_0^0 is used as an initial condition to solve the differential equation, it represents the number of players who bought the game the previous day (P_0) that were playing at the end of it. From this point the calculation of G and O is trivial.

As it has been done for the second day, the players will be divided in groups depending on the days the have been playing the game. For example, P_3 will be assigned to groups of players that started playing 3 days ago. Using these criteria the final solution for n days is presented:

$$O_n = O_{n-1} + \frac{t_{day}}{\tau_n}\left\{\bar{P}_n - \sum_{i=0}^{n-1}\left[\frac{\tau_i}{t_{day}}\left(P_i^0 - R_i\tau_i\right) - R_i\tau_i\right]\right\}\left[1 + \frac{\tau_n}{t_{day}}\left(e^{-\frac{t_{day}}{\tau_n}} - 1\right)\right]^{-1} \quad (10)$$

This is the formula to calculate the total owners at n days since the game has been published. Note that players have been divided per days, but any unit of time is valid as long as the parameters are properly changed, I personally recommend choosing weeks as a time unit, however I will keep on using t_{day} because I do not want to create confusion with the notation. My intuition tells me that it does not matter if the players play more than once per unit of time, if they did so rising τ to match the total playtime would be enough, but I'd like to do a simulation on that just to be sure.

The determinations of the parameters τ, R and P^0 is the main issue of this method. In the next section I give some ideas on how to find the parameters.

Parameter determination

In order to determine the owners of the game it is necessary to have a source of information that provides the number of players of each game at any given time, I have only been able to extract the number of current players of the top 100 games, since Steamdb offers charts with continuous representations of the number of players I am going to assume that there is a way even though I could not find it.

The other necessary feature are the parameters τ and R for each unit of time, P^0 is a product of τ and R:

$$P_i^0 = \left(P_{i-1}^0 - R_{i-1}\tau_{i-1}\right)e^{-\frac{t_{day}}{\tau_{i-1}}} + R_{i-1}\tau_{i-1} \quad (11)$$

The difficulty of finding these parameters lies in the capacity for rapidly finding users who own the game to gather their data, in order the establish τ and R for each unit of time it is necessary to have $\tau(t_{played})$ and $R(t_{played})$ (t_{played} is the total hours spend playing a game, the user's Steam profile provides this information), obviously we'll have less data for shorter playtimes that's why it is important to find owners the firsts days when everybody has just started playing.

The method generates the necessity to search Steam users by their interests and their lists of desired games instead of a randomized search, I do not know if you already have a database for that, but It is necessary to start compilating information, the more data we have on users the easier it will to find owners. I am currently not sure on how many users would be appropriate to obtain the parameters with enough accuracy, but I'll run a simulation in the future.

The advantage of this parameters is that once you know them, they are transferable, since we can suppose that they evolve the same way regardless of when the players bought the game. For example, τ will be the same for players who have bought the game the current week than for the ones who will buy it the next one:

$$\tau_i\ (n\ weeks) = \tau_{i+1}\ (n\ weeks) \quad (12)$$

R is not the same than τ because it depends on the total number of players that incorporated the first day:

$$R_i\ (n\ weeks) \neq R_{i+1}(n\ weeks) \quad (13)$$

$$O_i x_i(n\ weeks) \neq O_{i+1} x_{i+1}(n\ weeks) \quad (14)$$

However, it is logical to assume:

$$x_i\ (n\ weeks) = x_{i+1}(n\ weeks) \quad (15)$$

Where x_i (n weeks) is the proportion of players who acquired the game after i weeks of its release who have played the game after n weeks of buying it, obviously x_i(0 weeks)=1.

Advantages of the method

- It calculates the owners who have incorporated every unit of time, this opens the possibility of inferring the game's total revenue in Steam.

- Knowing the habits of the Users who play a specific kind of game could allow to determinate the number of owners in games with insufficient data. With enough data about the users it would even be possible to predict the owners of a game before it has been released.

- Assuming that the Users do not change their habits in the different platforms, the information compilated in Steam could help to predict the owners in platforms such as GOG, which offers less information.