STUDENT VERSION

Mathematics of Marriage

Erdi KARA
Mathematics and Statistics
Texas Tech University
Lubbock TX USA

STATEMENT

Have you ever wondered the age at which you will most likely get married? Don't tell me no, of course you did. Assuming that you are taking differential equation in your early 20s, if you are already married, you are so fast, congratulation! If not, this tutorial may partially tell you the time you will most likely say goodbye to the single life. Well, it may not be that much helpful to find the best match for you but it may tell you "Go and enjoy your life, you have time!" or "I think this is a good time to look around..." or "You'd better hustle your bustle!". Let's walk together and explore a model [1] which can help us out on our way to marriage.

MODEL

Construction

Let's think about the forces that cause people to marry. We will try to build a model based on the following assumptions:

1. *Social pressure*
   As the fraction of people married increase in our age group, we feel more pressure to get married.
2. *Age*
   Chances of marriage declines as we get older.

Activity 1

Do these assumptions seem reasonable? Write down several other factors which may effect the marriage of an individual in a cohort.

We will build a differential equation where our dependent variable is the fraction of individuals in a cohort already married. In this work, this cohort refers to the group of man or women who were born in a specific time period. For example, we can consider the women in US born between 1970 and 1974.

Let p_i be the i^{th} person's chances of marriage in a small time frame dt and $P = P(t)$ be the fraction of the cohort already married at time t. Notice that this implies $0 \leq P \leq 1$ Then based on the first assumption, we can write

$$\frac{dp_i}{dt} = qP \tag{1}$$

where q is a parameter which will take care of the second assumption. We will figure it out soon. The model looks nice but how do we determine each person's chances of marriage? This is not feasible. Let's find an expression for the probability at the cohort level. Assume that

$$\frac{dp_i}{dt} = \frac{dp_j}{dt.} \tag{2}$$

Activity 2

Discuss what this assumption actually means.

Let n be the number of people in the cohort and $m = m(t)$ be the number of people in that cohort who are already married. We can then write an expression for the rate of change in the number already married by adding the individual rate of change for those not married at time t.

$$\frac{dm}{dt} = \sum_{i=m+1}^{n} \frac{dp_i}{dt} = (n-m)\frac{dp_i}{dt} = (n-m)qP. \tag{3}$$

Let's divide both sides by n, noting $P = \dfrac{m}{n}$. We obtain

$$\frac{1}{n}\frac{dm}{dt} = \frac{d(\frac{m}{n})}{dt} = \frac{dP}{dt} = \Big(\frac{n-m}{n}\Big)qP = \Big(1 - \frac{m}{n}\Big)qP = q(1-P)P. \tag{4}$$

So we have

$$\frac{dP}{dt} = q(1-P)P. \tag{5}$$

In this form, (5) looks like a logistic differential equation. Now let's go back to (1) and focus on the term q. Using our second assumption, we claim that a person's chances of marriage decreases as he/she gets older. So it is reasonable to say

$$q = f(t)$$

where $f(t)$ is a decreasing function of time. We can actually do better. Let's further assume that individual i starts life with a marriage potential A_i which declines with a constant proportion b_i over the time. If we consider q at the cohort level, we can write

$$f(t) = Ab^t \tag{6}$$

where A is the average initial marriage potential of the cohort and b is the average deterioration term taken as $0 < b < 1$. [1] Later, we will see that these parameters have very interesting interpretations pertaining to the nature of the chosen group. For example, it may turn out to be that your parents' cohort was well ahead of yours' in terms of marriageability.

Activity 3

Interpret A and b and criticize the construction of $f(t)$. You can, for example, try to come up with a new function which is promising for single members of the cohort or simply say "well, that makes sense at this moment, but I should see the results...". Note down your ideas, we may use them later.

We are ready to complete our differential equation. If we substitute (6) into (5), we obtain

$$\frac{dP}{dt} = Ab^t(1-P)P. \tag{7}$$

Activity

(7) is a separable differential equation. Find its general solution.

After some simplifications, the general solution of (7) can be found as

$$P(t) = \frac{1}{1 + \frac{(1-C)e^{\frac{A}{ln(b)}}}{Ce^{\frac{Ab^t}{ln(b)}}}} \tag{8}$$

for a constant C. If we let $\ln(a) = \dfrac{A}{\ln(b)}$ for some positive number a and substitute into (8), we have

$$P(t) = \frac{1}{1 + \frac{(1-C)a}{Ca^{b^t}}}. \tag{9}$$

If we further let $k = \dfrac{C}{(1-C)a}$, we obtain the most simplified version of our solution

$$P(t) = \frac{1}{1 + \frac{1}{ka^{b^t}}} \tag{10}$$

Parameter Estimation

We can now proceed to the next step, parameter estimation. We must test the quality of our models against real data. As you already noticed, we have some parameters in the solution (10). How do we estimate these parameters?

Let's remember again that $P(t)$ is the fraction of the cohort already married at time t. For example, if we assume that the earliest marriage in the cohort is observed at the age of 18, then $P(0) = 1.7$ means, 1.7% of the cohort was married at 18 years old, $P(1) = 5.1$ means 5.1% of the group was married at 19 years old and notice that this 5% also includes those who married at the age of 18. Similarly, we interpret

[1] At least it is not 0, never give up!

$P(4) = 31.3$ as 31.3% of the cohort were already married at 22 years old and this 31% contains ages from 18 to 22 inclusive.[2] This way of data representation is called *cumulative distribution* and plays a central role in statistics. Once we have such a dataset, we can compute the parameters, k, a and b in (10)(thus A and P_0 as well) using a parameter estimation technique and then test our model for increasing values of t. To sum up, we have two tasks: find a relevant data set and evaluate the parameters.

Activity 4

"Data!data!data!" he cried impatiently. "I can't make bricks without clay." (Arthur Conan Doyle, The Adventure of the Copper Beeches). We can't complete our model without data. Now, your task is to find relevant data sets. You can look at [3]. Arrange your data in two columns, for example:

Age	Marriage(%)
18	2.37
19	4.51
20	6.93
...	...

Model Validation

Data Description

With the data in hand, we can now focus on parameter estimation. Our solution (10) includes a term ka^{b^t} which has the general term of so-called Gompertz function and there is a standard procedure to compute the parameters of a Gompertz function[2] for a given data set. To find these three parameters, three equations are created by dividing up observations into three equal parts. Then k, a and b can be computed explicitly from these equations. Since this task is not our main interest in this project, we will first introduce the data set and then estimate the corresponding parameters using Python programming language without going through the details. However, we will attach the required codes in Jupyter Notebook (DataAnalysis.ipynb) format along with the data files so that you can easily modify the code and estimate the parameters based on your data. The first dataset we will use is the cumulative first marriages for white women born in 1920-1924 in US. This data can be found in [1]. Real values and the predicted values by the model (10) can be seen in table below.

We find the best parameters in (10) to be;

$$k = 13.259 \quad a = 0.001 \quad b = 0.854 \quad A = 1.046 \tag{11}$$

We can also plot the results. As we can see from the figure below, the model closely agrees with the actual data. Thus, we can claim that as long as we have reliable data, we can use our model to make a prediction about the marriage patters of selected birth cohorts.

We want to investigate how marriage patterns change for particular birth cohorts over the years. To achieve this, we need a data set consisting of cumulative first marriage of people born between particular

[2] Notice that instead of using the actual ages, we prefer to translate the ages to numbers starting from 0. This is very common practice in time dependent processes.

ages	Actual Value	Predicted
15	1.8	1.68
16	4.6	4.32
17	9.6	9.35
18	17.6	17.30
19	27.5	27.67
20	38.1	39.06
21	48.3	49.90
22	57.5	59.22
23	65.3	66.70
24	71.6	72.51
25	76.6	76.94
26	80.6	80.31
27	83.3	82.88
28	85.2	84.86
29	86.7	86.40
30	87.9	87.61
31	88.7	88.57
32	89.4	89.34
33	90	89.96
34	90.5	90.46
35	90.9	90.88
36	91.2	91.21
37	91.4	91.50
38	91.6	91.73

Table 1: Cumulative First Marriages for White Women Born in 1920-24 in the United States: Observed and Calculated Percentage

years. One obvious source might be the publications of the US Census Bureau. We could not directly find such data among the US Census Data. One can however create the cumulative marriage fractions by comparing the various relevant sources in the census records. This task can be quite challenging since it may potentially require accessing unpublished results as well. Thus, we will use a survey data given in [4]. The marital history by sex for selected birth cohorts from this publication can be seen in the Tables 2 and 3 below. Since the survey is carried out in 2009, the cohort which had not reached at the indicated age are represented by 0.

When we consider the general theme of this paper, we can see that this data set does not fully meet our needs. The first problem is that we are missing the intermediate ages since the data is given in 5-year intervals. Notice that we have only 7 data points, namely ages from 20 to 50. Recall from equation (10) that we have 3 parameters to be determine but 7 data points result in poor parameter estimation so we can not rely on parameters computed with few data points. To work around this problem, we will insert 4 additional

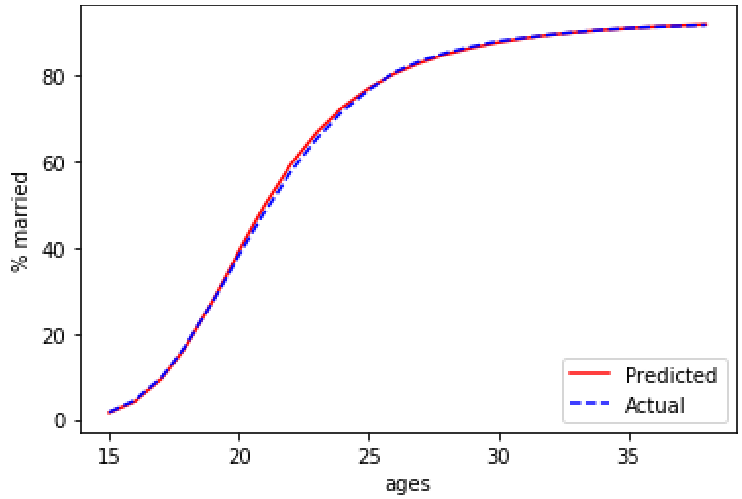


Figure 1: Model Predictions for the White Women Born in 1920-1924

Age	1940-1944	1945-1949	1950-1954	1955-1959	1960-1964	1965-1969	1970-1974	1975-1979
20	21.1	22.3	21.9	17.8	14.2	12.2	10.1	8.7
25	66.1	65.5	57.6	49.7	43.9	39.6	35.9	24.3
30	83.1	80.1	73.3	66.9	63.8	61.6	59.7	0
35	88.8	86.1	80.5	75.4	74.6	75.3	0	0
40	91.2	89.3	84.2	81.2	80.6	0	0	0
45	92.7	91.3	86.8	84.5	0	0	0	0
50	94	92.5	88.6	0	0	0	0	0

Table 2: Cumulative Marriage Rates for Men Ever Married

data points into each 5-year age intervals by equally dividing the cumulative marriage rates. By this way, we will obtain 30 data points in total representing all ages from 20 to 50 and we will treat these values as the observed values. If we apply this approach for Table (2), we can obtain the values given in Table (4). We display the ages from 20 to 35 for convenience, while the rest can be found accordingly.

Using the Table 4, we compute and compare the parameters for different birth cohorts. Although we *artificially* add new data points, we can claim that this should not severely effect the reliability of the model since the values we add are consistent with the overall behavior of the marriage pattern. For example, we observe from Table (2) that 22.3% of the people born between 1945-1949 were already married at the age of 20 and this percentage is 65.5% when the cohort turned into 25. Based on this range, one would not expect that 60% of the cohort would get marry at the age of 21. We can predict that this percentage should be somewhere around 30%.

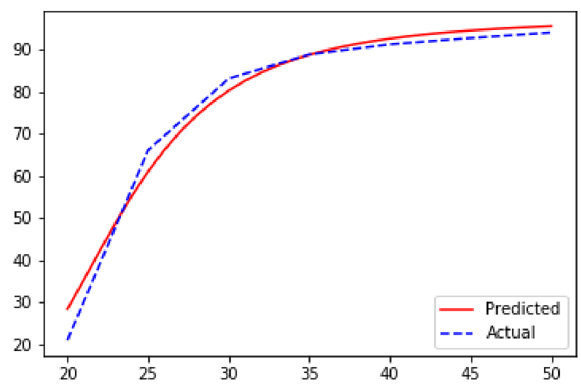
Results

Let's first take the birth cohort 1940-1944 and 1945-1949 from Table (2) and compute the values predicted by the model. The figure below shows that the model is in close agreement with the interpolated data. Note

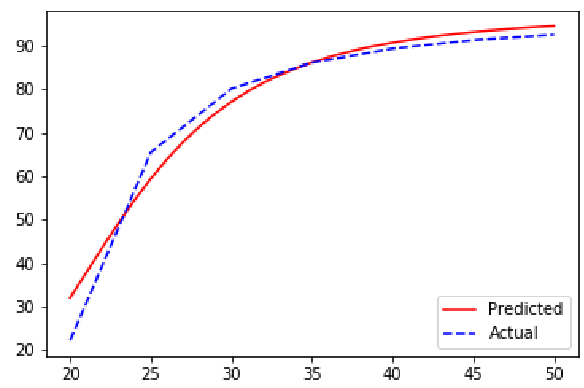
Age	1940-1944	1945-1949	1950-1954	1955-1959	1960-1964	1965-1969	1970-1974	1975-1979
20	48.1	43.1	38.9	33.4	28.5	22.9	20.2	18.5
25	78.2	76.9	69.2	62.5	56.2	52.1	50.2	47.3
30	86.6	85	79.7	74.8	71.9	71.2	70.8	0
35	89.7	88.4	84.8	81.6	79.7	79.6	0	0
40	91.4	90.2	87.5	85.3	83.7	0	0	0
45	92.5	91.5	89.2	87.6	0	0	0	0
50	93.2	92.2	90.3	0	0	0	0	0

Table 3: Cumulative Marriage Rates for Women Ever Married

that we had also confirmed these results in Figure (1) on a real and complete data set.



(a) Cohort 1940-1944



(b) Cohort 1945-1949

Figure 2: Model Predictions for Two Different Birth Cohorts

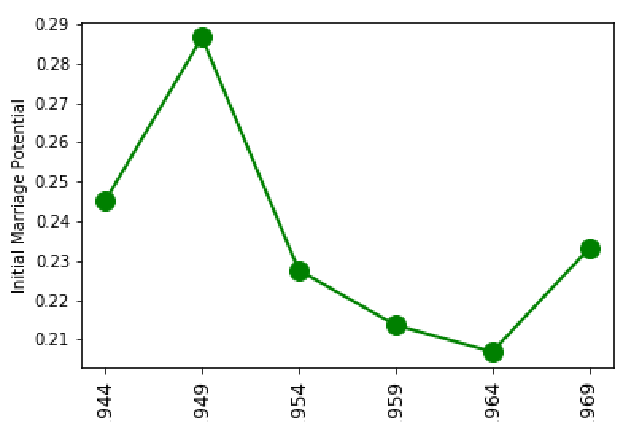
We introduced two interesting terms in (6), the initial marriage potential A and and the deterioration rate b. Let's compute these parameters by considering the ages from 20 to 35 for the interpolated marriage rates of man and female. Results can be seen in the tables below.

Age	1940-1944	1945-1949	1950-1954	1955-1959	1960-1964	1965-1969	1970-1974	1975-1979
20	21.1	22.3	21.9	17.8	14.2	12.2	10.1	8.7
21	30.1	30.94	29.04	24.18	20.14	17.68	15.26	11.82
22	39.1	39.58	36.18	30.56	26.08	23.16	20.42	14.94
23	48.1	48.22	43.32	36.94	32.02	28.64	25.58	18.06
24	57.1	56.86	50.46	43.32	37.96	34.12	30.74	21.18
25	66.1	65.5	57.6	49.7	43.9	39.6	35.9	24.3
26	69.5	68.42	60.74	53.14	47.88	44	40.66	0
27	72.9	71.34	63.88	56.58	51.86	48.4	45.42	0
28	76.3	74.26	67.02	60.02	55.84	52.8	50.18	0
29	79.7	77.18	70.16	63.46	59.82	57.2	54.94	0
30	83.1	80.1	73.3	66.9	63.8	61.6	59.7	0
31	84.24	81.3	74.74	68.6	65.96	64.34	0	0
32	85.38	82.5	76.18	70.3	68.12	67.08	0	0
33	86.52	83.7	77.62	72	70.28	69.82	0	0
34	87.66	84.9	79.06	73.7	72.44	72.56	0	0
35	88.8	86.1	80.5	75.4	74.6	75.3	0	0

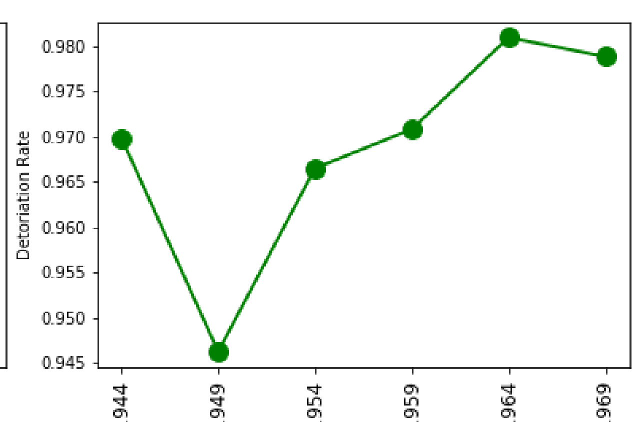
Table 4: Interpolated Marriage Rates for Table (2)

Params	1940-1944	1945-1949	1950-1954	1955-1959	1960-1964	1965-1969
A	0.36223	0.36341	0.30306	0.29722	0.29282	0.27701
b	0.95558	0.93815	0.94303	0.93864	0.94986	0.96657
k	912.74167	97.15599	55.37289	26.90764	57.79610	581.01011
a	0.00034	0.00337	0.00570	0.00915	0.00337	0.00029

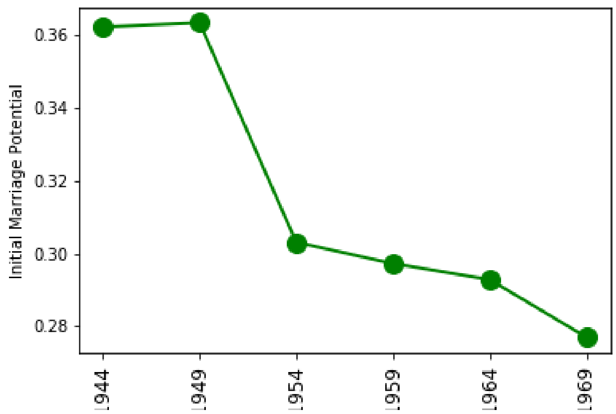
Table 5: Initial Marriage Potentials and Deterioration Rates for Men



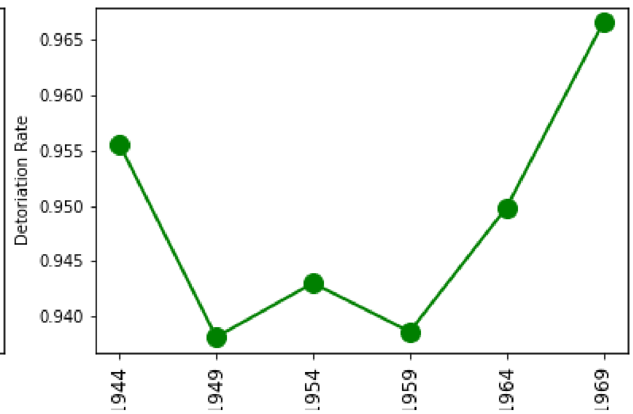
(a) Initial marriage potentials for women



(b) Deterioration rates for women



(a) Initial arriage potentials for men



(b) Deterioration rates for men

Params	1940-1944	1945-1949	1950-1954	1955-1959	1960-1964	1965-1969
A	0.24516	0.28658	0.22758	0.21363	0.20687	0.23300
b	0.96983	0.94621	0.96649	0.97080	0.98094	0.97885
k	2857.33467	138.64908	537.31038	731.75656	20170.96497	17839.41488
a	0.00034	0.00561	0.00126	0.00074	0.00002	0.00002

Table 6: Initial Marriage Potentials and Deterioration Rates for Women

The results above can give us valuable insight about how the marriage patterns have changed over the years. We can draw the following conclusions from these results;

1. Initial marriage potentials of men are always greater than women's in all birth cohorts. In other words, compared to woman, men cohorts historically had a greater chance of marriageability.
2. Except the birth cohort 1965-1969 of women, there is a steady decrease in marriage potentials for both groups after 1945-1949 cohort. Although we don't have a reliable data after the last cohort, the results do not seem very promising for both cohorts.
3. For both men and women, there was a sharp fall in marriageability between the second and third birth cohorts. Can we come up with possible explanations about the causes of this change?
4. Recall that the smaller the deterioration rate, the faster the initial marriage potential decreases. From these figures, we observe that whenever there is a decrease in the marriage potentials, there is an increase in the deterioration rates. This means that although the later generations had a smaller marriageability when compared to the previous ones, their chance of marriage did not fade away quickly due to the relatively large deterioration rates. If we were given such information without any mathematics behind it, we might say, "Well if my chance of marriage will not dramatically decay for my generation over the years, then why not enjoy the single life for sometime or work for my career without thinking about the responsibilities of marriage?". This is exactly what we observe from the Table 2 and 3. If we examine the marriage rates, at 20s for example, they are steadily decreasing, i.e. people tend to get marry in later ages.

Activity 5

You have probably heard about the popular adage saying that "There ain't no such thing as a free lunch". Discuss the inverse proportionality between the initial marriage potential and the deterioration rates from the perspective of the this adage.

REFERENCES

[1] Hernes, Gudmund. 1972. The process of entry into first marriage. *American Sociological Review*. 173-182.

[2] Prescott, Raymond B. 1920. Law of Growth in Forecasting Demand. *Journal of the American Statistical Association*. 18: 471-479.

[3] United States Census. 2020. https://www.census.gov/library/publications.html. Accessed 5 April 2020.

[4] Kreider, Rose M. and Renee Ellis. 2009. *Number, Timing, and Duration of Marriages and Divorces: 2009*. Washington DC: Current Population Reports, P70-125, U.S. Census Bureau.