Maximum Likelihood Estimation
The Exponential Distribution

Gary Schurman MBE, CFA

March, 2023

In this white paper we will use Maximum Likelihood Estimation (MLE) to estimate the hazard rate parameter of an Exponentiallly-distributed random variable. To that end we will work through the following hypothetical problem...

Our Hypothetical Problem

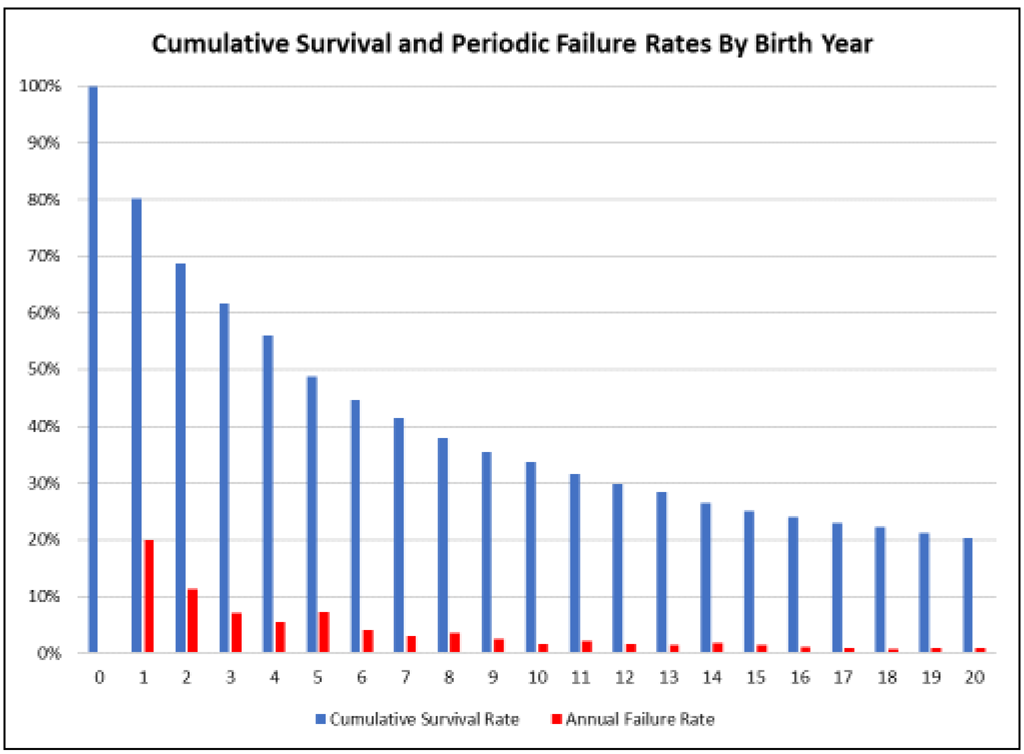
We are given the following startup survival and failure rates by birth year...

Source: https://www.jpmorganchase.com/institute/research/small-business/small-business-dashboard/longevity

Table 1: JP Morgan Data

Birth Year	Survival Rate	Failure Rate
0	1.0000	–
1	0.8010	0.1990
2	0.6870	0.1140
3	0.6160	0.0710
4	0.5600	0.0560
5	0.4880	0.0720
6	0.4470	0.0410
7	0.4160	0.0310
8	0.3800	0.0360
9	0.3550	0.0250
10	0.3380	0.0170
11	0.3160	0.0220
12	0.2990	0.0170
13	0.2840	0.0150
14	0.2660	0.0180
15	0.2510	0.0150
16	0.2400	0.0110
17	0.2300	0.0100
18	0.2220	0.0080
19	0.2130	0.0090
20	0.2040	0.0090
Total	–	0.7960

Table 2: Survival And Failure Rate Graph By Birth Year



Question 1: Using the data above, what is the maximum likelihood estimated hazard rate?

Question 2: Graph the actual distribution against the estimated exponential distribution.

The Probability Density Function

We will define the parameter λ to be hazard rate of an Exponential distribution. We will define the function $g(t)$ to be the observed survival rate at time t. The equation for the expected survival rate at time t is...

$$\mathbb{E}\bigg[g(t)\bigg] = \text{Exp}\bigg\{-\lambda\,t\bigg\} \tag{1}$$

We will define the function $f(t)$ to be the probability density function of an Exponentially-distributed random number t. The equation for the probability density function of the Exponential distribution is... [1]

$$f(t_i) = \lambda\,\text{Exp}\bigg\{-\lambda\,t\bigg\} \tag{2}$$

The equation for the natural log of Equation (2) above is...

$$\ln\bigg(f(t_i)\bigg) = \ln(\lambda) - \lambda\,t \tag{3}$$

The Maximum Log Likelihood Function

Maximum likelihood estimation is a method of estimating the parameters of an assumed probability distribution, given some observed data. This is achieved by maximizing a likelihood function so that, under the assumed statistical model, the observed data is most probable.

We will define the function ML to be the maximum likelihood function. Given that the random observations in our sample are independent, using Equation (2) above the joint probability of observing all of the data points in our sample is...

$$ML = \prod_{i=1}^{N} f(t_i) \tag{4}$$

We will define the variable MLL to be the maximum log likelihood function and the variable N to be the number of observations in our sample above. Using Equations (3) and (4) above the equation for the maximum log likelihood function is...

$$MLL = \ln\bigg(\prod_{i=1}^{N} f(t_i)\bigg) = \sum_{i=1}^{N} \ln(\lambda) - \lambda \sum_{i=1}^{N} t_i \tag{5}$$

Using Appendix Equations (10) and (11) below the equation for the derivative of Equation (5) above with respect to the hazard rate λ is...

$$\frac{\delta\,MLL}{\delta\,\lambda} = N\,\lambda^{-1} - \sum_{i=1}^{N} t_i \tag{6}$$

The estimate the hazard rate of the Exponential distribution that is consistent with our hypothetical problem above we want to maximize the joint probability of observing all of the data points in our sample. We do this by setting the derivative in Equation (6) above to zero and then solving for λ.

Using Equation (6) above the equation for the value of the hazard rate λ is...

$$\text{if...}\;\; \frac{\delta\,MLL}{\delta\,\lambda} = 0 \;\;\text{...then...}\;\; \lambda = \left[\frac{1}{N}\sum_{i=1}^{N} t_i\right]^{-1} \tag{7}$$

The Solution To Our Hypothetical Problem

Using the data in Table 1 above our sample statistics are...

$$N = 21 \;\;\text{...and...}\;\; \sum_{i=1}^{N} t_i = 210 \tag{8}$$

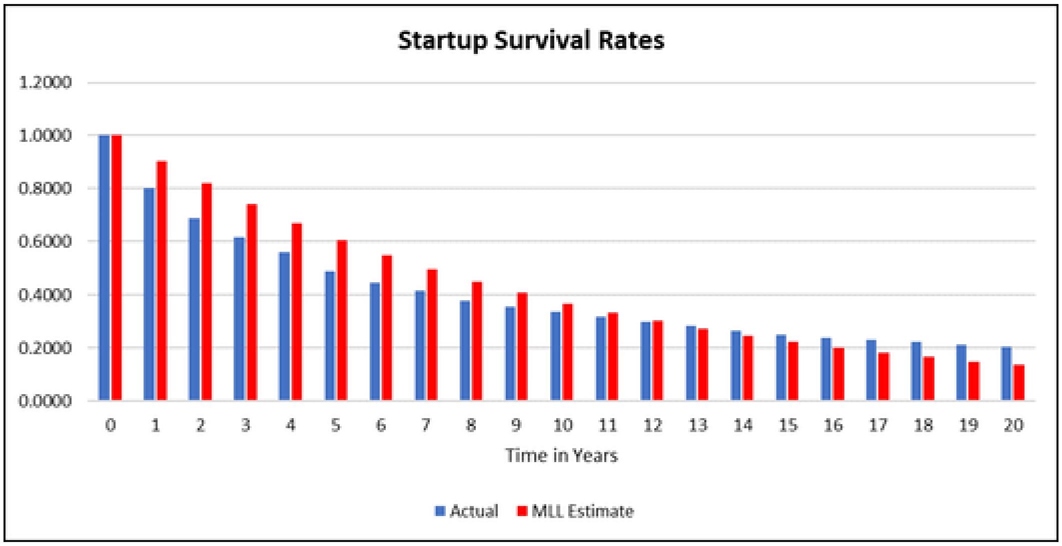
Question 1: Using the data above, what is the maximum likelihood estimated hazard rate?

Using Equations (7) and (8) above the equation for the hazard rate λ is...

$$\lambda = \left[\frac{1}{21} \times 210\right]^{-1} = 0.1000 \tag{9}$$

Question 2: Graph the actual distribution against the estimated exponential distribution.

Using Equations (1) and (9) above and the data in Table 1 above the graph of the actual survival rate versus the expected survival rate using our estimate of λ above is...



References

[1] Gary Schurman, *The Exponential Distribution - The Mathematics,* March, 2012.

Appendix

A. The solution to the following derivative equation is...

$$\frac{\delta}{\delta\lambda}\left[\sum_{i=1}^{N} \ln(\lambda)\right] = \sum_{i=1}^{N} \frac{1}{\lambda} = N\,\lambda^{-1} \tag{10}$$

B. The solution to the following derivative equation is...

$$\frac{\delta}{\delta\lambda}\left[\lambda \sum_{i=1}^{N} t\right] = \sum_{i=1}^{N} t \tag{11}$$