

# Parameter Estimation for the Gompertz Distribution

## Stat 624 Project

*W. Zachary Horton*

Brigham Young University

**Abstract.** The Gompertz distribution is widely used for modeling adult lifespans. It can be described by a rate parameter and a shape parameter,  $a$  and  $b$  respectively, which influence the shape and skew of the density function. We discuss parameter estimation through log-hazard regression, maximum likelihood estimation, and bayesian methods. Through simulation studies, we find that log-hazard regression is sensitive to small datasets. We apply these estimators to an adult mortality data set to see how well they fit the data. We find that the all three fit the data with about the same degree of error, and that the maximum likelihood and bayes estimators converge towards similar values. We conclude that, due to extended lifespans in recent decades, the Gompertz distribution may not be appropriate to model current adult mortality.

## 1 Properties of the Gompertz Distribution

### 1.1 Introduction

The Gompertz distribution, named after the British mathematician Benjamin Gompertz, is most commonly used to model patterns of adult lifespans. It is based on the Gompertz Law of Mortality which states that the force of mortality, or hazard rate  $h$ , for someone of age  $x$  is given by

$$h(x|a, b) = ae^{bx}, \quad 0 \leq x < \infty, \quad a > 0, \quad b > 0 \quad (1.1)$$

where  $a$  is the baseline force of mortality and  $b$  represents an aging rate. The Gompertz distribution has historically been used to construct actuarial life tables and fit similar patterns in demography. Recently, it has been used to model failure times of computer processes and has been classified as an extreme value distribution modeling minimums.

## 1.2 Distribution Properties

The Gompertz distribution has support  $0 \leq x < \infty$ . The hazard function of the Gompertz distribution can be used to derive the cumulative density function (CDF). Actuarial mathematics texts state the relationship between the hazard function and the CDF for any vector of parameters  $\boldsymbol{\theta}$  as

$$F(x|\boldsymbol{\theta}) = 1 - e^{-\int_0^x h(t|\boldsymbol{\theta})dt} \quad (1.2)$$

Therefore we see the CDF of the Gompertz distribution is

$$F(x|a, b) = 1 - e^{-\frac{a}{b}(e^{bx}-1)} \quad (1.3)$$

and the probability density function (PDF) is

$$f(x|a, b) = ae^{bx}e^{-\frac{a}{b}(e^{bx}-1)} \quad (1.4)$$

Note that the PDF is the product of the hazard function and the survival function, which is expected based on properties of hazard functions. Section 1.5 shows graphical properties of the PDF and CDF.

## 1.3 Parameter Properties

The rate parameter  $a$  represents the baseline mortality and is the value of the density at  $x = 0$ . The shape parameter  $b$  represents an aging factor, but this is less interpretable. Both  $a$  and  $b$  must be greater than zero.

## 1.4 Other Properties

Other interesting properties about the Gompertz distribution include:

- The mean and variance don't have closed forms.
- As  $b$  approaches 0, the Gompertz distribution becomes an exponential distribution with rate  $a$ .
- The Gompertz Distribution is a negated truncated Gumbel distribution which means that if  $X$  follows a Gumbel distribution and samples are drawn until a negative value is obtained, then the random variable  $Y = -X$  follows a Gompertz distribution.

### 1.5 Graphical Properties

Figure 1 shows different shapes of the Gompertz distribution density. The direction of skew is mostly influenced by the  $a$  parameter, where the density is left-skewed when  $a$  is close to zero. The  $b$  parameter is less interpretable, but it serves as a sort of inverse variance parameter, meaning that increasing  $b$  will force the density to get taller and narrower. To demonstrate the possible shapes of a Gompertz distribution density it is useful to visualize the Gompertz density as a right-skewed Gumbel distribution that has been flipped around and truncated somewhere along the x-axis. Figure 3 shows an illustration of this exercise.

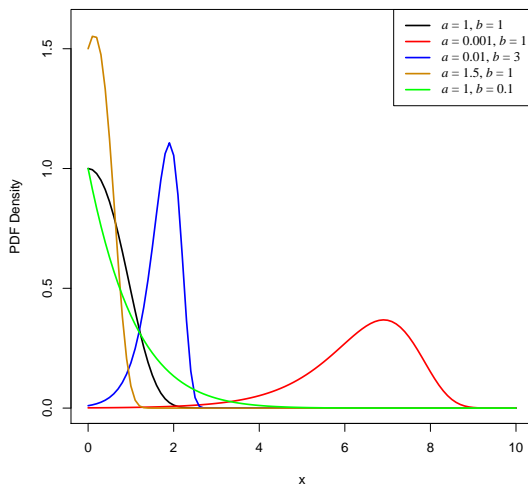


Fig. 1: Shapes of the Gompertz PDF

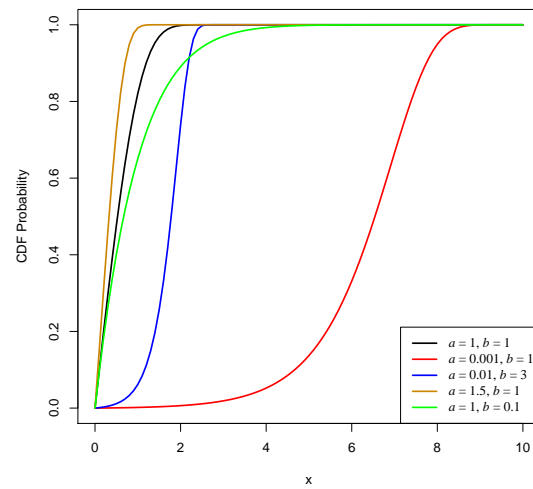


Fig. 2: Shapes of the Gompertz CDF

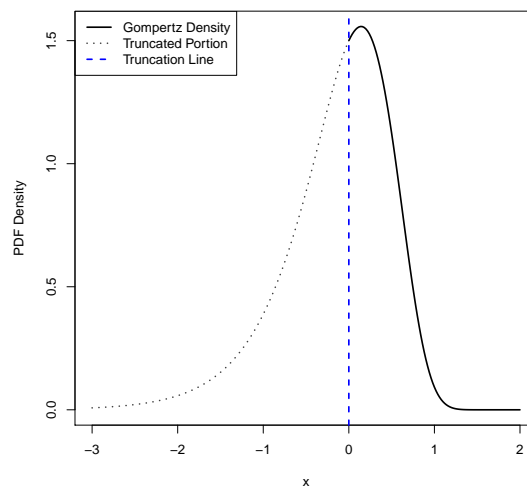


Fig. 3: Visualizing Gompertz as Truncated Gumbel

## 2 Methods of Parameter Estimation

In this section we discuss three methods of parameter estimation for the Gompertz distribution: log-hazard regression, maximum likelihood estimation, and bayesian estimation. For each we derive the method of estimation, discuss any relevant properties, and describe the computation algorithm.

### 2.1 Log-Hazard Regression

The log-hazard regression estimator (LHRE) is based on the observation that the log of the Gompertz hazard function given in (1.1) is a linear function:

$$\ln(h(x)) = \ln(a) + bx \quad (2.1)$$

This result suggests that simple linear regression of the log-hazard rate can be used to estimate parameter values. Estimating the hazard rate at each data point  $x$  can be done empirically. Actuarial mathematics texts define the hazard function for any parameter vector  $\theta$  as the ratio of the density function and the survival function:

$$h(x|\theta) = \frac{f(x|\theta)}{S(x|\theta)} \quad (2.2)$$

Empirical estimation of this ratio is done by estimating its components. The survival function at  $x$  can be estimated by the percent of data points greater than or equal to  $x$ . A simple method of estimating the density at  $x$  is to find the percent of data points equal to  $x$ . This estimation may not be accurate if the data are continuous. However, the Gompertz distribution is typically used to model discrete integer ages at death, in which case this method is sufficient. We denote the estimated hazard rate at  $x$  as  $\hat{h}(x_i)$ . The formulas for the least squares estimates can be derived through basic linear models principles, so the full derivation won't be included here. We now describe the method of computation. Begin by computing  $\bar{x}$ , the data mean of the  $n$  data points and  $\bar{h}$ , the mean of the estimated hazard values. The estimated slope  $\hat{b}$  is computed as

$$\hat{b} = \frac{\sum_{i=1}^n (x_i - \bar{x})(\hat{h}(x_i) - \bar{h})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (2.3)$$

The estimated intercept  $\ln(\hat{a})$  is computed and exponentiated to obtain  $\hat{a}$ :

$$\ln(\hat{a}) = \bar{h} - \hat{b}\bar{x} \implies \hat{a} = e^{\bar{h} - \hat{b}\bar{x}} \quad (2.4)$$

Note that the LHRE is the only estimator we will consider that can be found analytically. We will use the LHRE as the initial values for the numerical methods required to find the other estimators. For large

data sets, computational techniques such as lazy caching can be used to speed up calculations, making this method considerably faster than the other two.

## 2.2 Maximum Likelihood Estimation

The maximum likelihood estimator (MLE) of a parameter is a well explored and understood estimation technique. Here we detail the log-likelihood and other functions needed to obtain the MLE's for each parameter. Given a data vector  $\mathbf{x}$  with length  $n$ , the likelihood function of the Gompertz distribution is

$$L(a, b|\mathbf{x}) = \prod_{i=1}^n f(x_i|a, b) \quad (2.5)$$

The log-likelihood is then

$$\begin{aligned} \ell(a, b|\mathbf{x}) &= \ln(L(a, b|\mathbf{x})) = \sum_{i=1}^n \ln(f(x_i|a, b)) \\ &= \sum_{i=1}^n \left( \ln(a) + bx_i - \frac{a}{b}(e^{bx_i} - 1) \right) \end{aligned} \quad (2.6)$$

The partial derivatives of the log-likelihood with respect to  $a$  and  $b$  are then obtained:

$$\frac{\partial \ell(a, b|\mathbf{x})}{\partial a} = \sum_{i=1}^n \left( \frac{1}{a} - \frac{1}{b}(e^{bx_i} - 1) \right) \quad (2.7)$$

$$\frac{\partial \ell(a, b|\mathbf{x})}{\partial b} = \sum_{i=1}^n \left( x_i - \frac{a}{b}(x_i e^{bx_i}) + \frac{a}{b^2}(e^{bx_i} - 1) \right) \quad (2.8)$$

Then the partial derivatives are set equal to zero. We now solve for the MLE's,  $\hat{a}$  and  $\hat{b}$ . Using a little algebra on the partial derivative with respect to  $a$ , it can be shown that

$$\hat{a} = \frac{n\hat{b}}{\sum_{i=1}^n (e^{\hat{b}x_i} - 1)} \quad (2.9)$$

By plugging this value back into the partial derivative with respect to  $b$  and simplifying, we see that  $\hat{b}$ , the MLE of  $b$ , is the solution to the equation

$$0 = \sum_{i=1}^n (x_i) + \frac{n}{\hat{b}} - n \frac{\sum_{i=1}^n (x_i e^{\hat{b}x_i})}{\sum_{i=1}^n (e^{\hat{b}x_i} - 1)} \quad (2.10)$$

To solve this equation for  $\hat{b}$ , we use the secant method of root finding. The secant method is extremely similar to the Newton-Raphson algorithm. The Newton-Raphson algorithm can be summarized by the following

iterative formula:

$$\theta_{i+1} = \theta_i - \frac{f(\theta_i)}{f'(\theta_i)} \quad (2.11)$$

where  $f$  is the function of interest,  $f'$  is its derivative, and  $\theta_0$  is a provided initial guess for the root. The idea is that the values of  $\theta$  will get closer and closer to the root with every iteration. The secant method approximates the Newton-Raphson formula by approximating the value of the derivative using the slope between  $\theta_i$  and  $\theta_{i-1}$ . Then the iterative formula for the secant method is

$$\theta_{i+1} = \theta_i - f(\theta_i) \frac{\theta_i - \theta_{i-1}}{f(\theta_i) - f(\theta_{i-1})} \quad (2.12)$$

where  $\theta_0$  and  $\theta_1$  are two initial guesses for the root. The advantage to the secant method is that the derivative does not need to be found. Comparatively the secant method is just as efficient and accurate. A potential issue for these root-finding methods is the possibility of divergence. Divergence is rare in the univariate case, usually occurring in functions that are not continuous, have asymptotic features, or are not differentiable over all the domain. When divergence occurs, the algorithm must start over with slightly different starting values. We choose the LHRE for  $b$  ( $\hat{b}_L$ ) to be the first starting value and  $\frac{\hat{b}_L}{2}$  as the second. We chose this second value because it helped avoid divergence in our implementation of the secant method. After our secant method algorithm has converged to the MLE of  $b$  within some small margin of error, we plug it in to (2.9) and obtain the MLE of  $a$ . It is interesting to note that when the data set is large and discrete, the LHRE is very close to the MLE and the secant method rarely diverges.

### 2.3 Bayesian Estimation

The last method we discuss is Bayes estimators. The Bayes estimator is based on Bayes law, which says that the posterior distribution of a parameter  $\theta_i$  from a parameter vector  $\boldsymbol{\theta}$ , given a data vector  $\mathbf{x}$ , is given by

$$\pi(\theta_i|\mathbf{x}) \propto f(\mathbf{x}|\boldsymbol{\theta})\pi(\theta_i) \quad (2.13)$$

where  $f$  is the likelihood function and  $\pi(\theta_i)$  is the density of the prior distribution on  $\theta_i$ . The Bayes estimator, then, is the mean of the posterior distribution. Often it is quite difficult to analytically compute the mean, so Monte Carlo Markov Chain (MCMC) methods are used to sample from the posterior and approximate the mean, which we do here. We will use two MCMC methods to estimate our parameters: Conjugate Prior Draws for estimating  $a$  and Metropolis-Hastings sampling for estimating  $b$ . Both of these will fit into a Gibbs sampling framework where we iteratively sample from each parameter's posterior. Note that the prior

distribution we chose for  $a$  and  $b$  are gamma distributions with shape parameter  $\alpha = 1$  and scale parameter  $\beta$  as the respective MLE estimate.

### 2.3.1 Conjugate Prior on $a$

When the  $b$  parameter in the Gompertz distribution is known, the gamma distribution serves as a conjugate prior for  $a$ . This means the posterior distribution  $a|\mathbf{x}$  follows another gamma distribution. Once the formulas for the parameters of the posterior are determined, we will use Gibbs sampling to generate values and approximate the mean. We now derive the posterior distribution parameter formulas. Let  $X$  follow a Gompertz( $a, b$ ) distribution where  $a$  follows a Gamma distribution with shape parameter  $\alpha = 1$  and scale parameter  $\beta = \hat{a}_M$  where  $\hat{a}_M$  is the MLE estimate of  $b$ . Then the prior density of  $a$  is

$$\pi(a) = \frac{1}{\hat{a}_M} e^{-a/\hat{a}_M} \quad (2.14)$$

and the likelihood function is given by

$$f(\mathbf{x}|a, b) = \prod_{i=1}^n a e^{bx_i} e^{-\frac{a}{b}(e^{bx_i} - 1)} = a^n \exp \left\{ b \sum_{i=1}^n x_i \right\} \exp \left\{ \frac{-a}{b} \sum_{i=1}^n (e^{bx_i} - 1) \right\} \quad (2.15)$$

Then, plugging into the formula for Bayes law results in

$$\begin{aligned} \pi(a|\mathbf{x}) &\propto f(\mathbf{x}|a, b)\pi(a) = a^n \exp \left\{ b \sum_{i=1}^n x_i \right\} \exp \left\{ \frac{-a}{b} \sum_{i=1}^n (e^{bx_i} - 1) \right\} \frac{1}{\hat{a}_M} e^{-a/\hat{a}_M} \\ &\propto a^n \exp \left\{ -a \left( \frac{1}{\hat{a}_M} + \frac{1}{b} \sum_{i=1}^n (e^{bx_i} - 1) \right) \right\} \end{aligned} \quad (2.16)$$

Therefore,  $a|\mathbf{x}$  is a gamma distribution with shape parameter  $\alpha^*$  and scale parameter  $\beta^*$  where

$$\alpha^* = n + 1 \quad (2.17)$$

$$\beta^* = \left( \frac{1}{\hat{a}_M} + \frac{1}{b} \sum_{i=1}^n (e^{bx_i} - 1) \right)^{(-1)} \quad (2.18)$$

Recall that this result is conditioned on  $b$  being known. Therefore, this conclusion will only be useful in the context of a Gibbs sampling process.

### 2.3.2 Metropolis-Hastings Sampling for $b$

The Metropolis-Hastings algorithm is an MCMC technique used to sample from a distribution where traditional sampling techniques are difficult or impossible. In our case, the posterior distribution  $b|\mathbf{x}$  is not a

recognizable distribution and will be quite difficult to integrate, so Metropolis-Hastings sampling provides a simple alternative. The Bayes estimate for  $b$  is the mean of the posterior distribution, which will be approximated by the mean of the Metropolis-Hastings draws. The algorithm works by accepting and rejecting values from a candidate distribution based on a computed ratio. Similar to the Newton-Raphson method, an initial value must be chosen. If this value is not a reasonable value in the posterior, then some of the MCMC draws will be spent moving into an acceptable region. In other words, the draws take time to converge on the true posterior distribution. To account for this, we burn off some of the initial draws and observe some diagnostics to check for convergence. The diagnostics we will use are:

- Acceptance Rate, should be between 20% and 60%.
- Trace Plots, should appear to have converged on a band of plausible values.
- Auto Correlation, lag correlation should drop off quickly, not extending beyond 50-60.

For our estimation, we will use a normal distribution centered at the previously accepted point for drawing candidate values. The variance of our candidate distribution will influence convergence and the diagnostics so it will be optimized for each MCMC scenario. Our initial value for  $b$  will be  $\hat{b}_M$  the MLE for  $b$ . Our prior distribution on  $b$  is a gamma distribution with shape parameter  $\alpha = 1$  and scale parameter  $\beta$  as  $\hat{b}_M$ , the MLE for  $b$ . We will now derive the ratio to be used and describe the algorithm in detail. The Metropolis-Hastings iteration specific to our situation is described by the following conditional statement:

$$\text{If } U_j \leq \frac{L(a, b_c | \mathbf{x}) \pi(b_c)}{L(a, b_{j-1} | \mathbf{x}) \pi(b_{j-1})}, \text{ Then } b_j = b_c, \text{ Otherwise } b_j = b_{j-1} \quad (2.19)$$

where  $L$  is the likelihood function given in (2.5) and (2.15),  $\pi$  is the prior density function,  $U_j$  is a random variable distributed uniformly from zero to one,  $b_{j-1}$  is the most recent draw,  $b_c$  is the candidate draw from the candidate normal distribution centered at  $b_{j-1}$ , and  $j$  is the draw index. Note that the prior density function is given as

$$\pi(b) = \frac{1}{\hat{b}_M} e^{-b/\hat{b}_M} \quad (2.20)$$

An interesting property of the Metropolis-Hastings algorithm is that no iteration is repeated when there is a rejection. Instead of repeating upon rejecting a candidate value, the previous value is considered to be the new value and is entered again into the chain of draws. Also, notice that the ratio is simply the posterior density evaluated at the candidate point divided by the posterior density at the last draw. Computationally this can cause issues since likelihoods can become very small. Taking the log yields the logically equivalent condition

$$\ln(U_j) \leq \ell(a, b_c | \mathbf{x}) + \ln(\pi(b_c)) - \ell(a, b_{j-1} | \mathbf{x}) - \ln(\pi(b_{j-1})) \quad (2.21)$$



where  $\ell$  is the log-likelihood function given in (2.6). Note that the MLE is a good approximation of the Bayes estimator, which is why we use it as a starting value. This result is conditioned on  $a$  being a known value, which means this procedure will only be useful in our situation within a Gibbs sampling context.

### 3 Simulation Studies

In this simulation, we study properties of our parameter estimates. We simulate data from several Gompertz distributions with different parameter and sample size combinations to examine their performance. Well performing estimates should be reasonably and consistently close to the true values. For each combination of parameters and sample size we compute bias and mean squared error (MSE).

$a$	$b$
1.0	1.0
0.001	1.0
0.01	3.0
1.5	2.0
1.0	0.1

Table 1: Parameter Combinations for Simluation

#### 3.1 Simulation Procedure

Table 1 contains the five different parameter combinations we use for simulation. These were chosen for their demonstration of different shapes that the Gompertz density can make. Figure 1 shows the density graphs for these parameters. We will be crossing these with three sample sizes: 20, 50, and 100, which were chosen as representatively small, medium, and large sample sizes. We use the inverse CDF method to sample from a Gompertz distribution. By letting the CDF given in (1.2),  $F(x|a, b) = u$  we solve for  $u$  and find the inverse CDF to be

$$F^{-1}(u|a, b) = \frac{1}{b} \ln(1 - \frac{b}{a} \ln(1 - u)) \quad (3.1)$$

which we plug uniformly distributed values into to generate random samples. In order to observe the behavior of the estimators, we run each scenario a thousand times and compute the bias and MSE. Table 2 shows all the scenarios we will test along with a numbered ID.

#### 3.2 Results

Tables 3, 4, and 5 contain the results of the simulation studies for LHRE, MLE, and Bayes estimators respectively. The tables are organized such that each row corresponds to a scenario, where the row number matches

	n	a	b
1	20	1.000	1.0
2	20	0.001	1.0
3	20	0.010	3.0
4	20	1.500	2.0
5	20	1.000	0.1
6	50	1.000	1.0
7	50	0.001	1.0
8	50	0.010	3.0
9	50	1.500	2.0
10	50	1.000	0.1
11	100	1.000	1.0
12	100	0.001	1.0
13	100	0.010	3.0
14	100	1.500	2.0
15	100	1.000	0.1

Table 2: Simulation Scenarios and ID

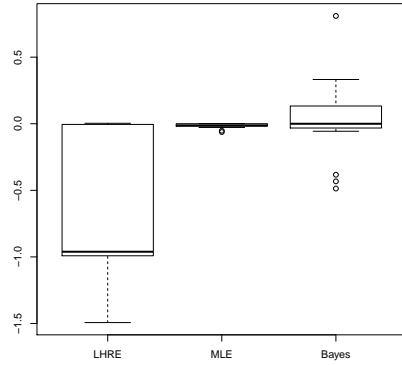
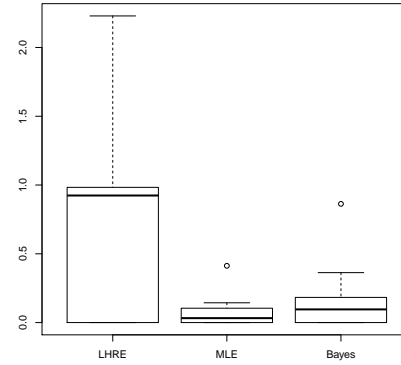
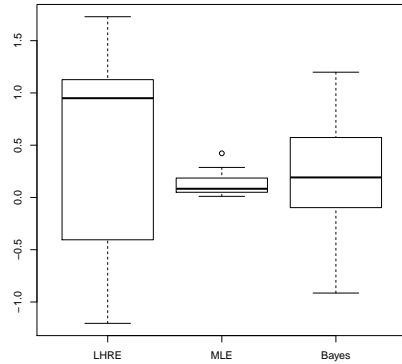
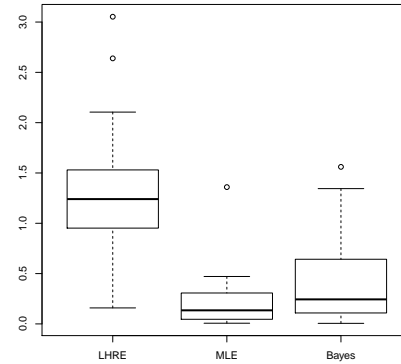
	Bias <i>a</i>	MSE <i>a</i>	Bias <i>b</i>	MSE <i>b</i>		Bias <i>a</i>	MSE <i>a</i>	Bias <i>b</i>	MSE <i>b</i>		Bias <i>a</i>	MSE <i>a</i>	Bias <i>b</i>	MSE <i>b</i>
1	-0.961	0.924	0.950	1.010	1	-0.062	0.142	0.287	0.471	1	-0.056	0.099	0.192	0.251
2	0.003	0.000	-0.415	0.196	2	0.000	0.000	0.070	0.046	2	-0.001	0.000	0.239	0.079
3	-0.003	0.000	-1.204	1.665	3	0.001	0.000	0.212	0.465	3	-0.008	0.000	1.198	1.561
4	-1.463	2.140	1.353	2.105	4	-0.020	0.413	0.422	1.360	4	-0.486	0.363	0.883	1.345
5	-0.950	0.903	0.921	0.922	5	-0.052	0.136	0.204	0.192	5	0.810	0.863	0.284	0.150
6	-0.985	0.970	1.093	1.240	6	-0.013	0.063	0.107	0.148	6	0.291	0.167	-0.381	0.244
7	0.000	0.000	-0.396	0.167	7	0.000	0.000	0.033	0.016	7	0.001	0.000	-0.105	0.020
8	-0.008	0.000	-1.145	1.396	8	0.000	0.000	0.086	0.135	8	-0.007	0.000	0.839	0.814
9	-1.486	2.207	1.589	2.639	9	-0.015	0.144	0.167	0.423	9	0.332	0.249	-0.915	1.017
10	-0.980	0.961	0.976	0.982	10	-0.028	0.045	0.083	0.046	10	-0.383	0.167	0.503	0.290
11	-0.993	0.985	1.161	1.371	11	-0.012	0.032	0.062	0.070	11	-0.433	0.198	0.644	0.471
12	-0.000	0.000	-0.392	0.159	12	0.000	0.000	0.011	0.007	12	-0.000	0.000	0.038	0.006
13	-0.009	0.000	-1.131	1.323	13	0.001	0.000	0.035	0.071	13	0.010	0.000	-0.316	0.138
14	-1.493	2.229	1.729	3.053	14	-0.004	0.073	0.077	0.188	14	0.149	0.096	-0.001	0.156
15	-0.990	0.981	1.014	1.044	15	-0.017	0.023	0.037	0.018	15	0.117	0.027	-0.089	0.008

Table 3: Log-Hazard Simulations

Table 4: MLE Simulations

Table 5: Bayes Simulations

a scenario specified in table 2 above. A glance at the table reveals how poorly the LHRE estimator does. It has the highest bias and MSE, however, this result is expected because the simulated data are continuous. Estimating hazard value with continuous data is often inaccurate since the probability of repeating a value is zero. Further insight about estimator performance can be gleaned from the boxplots in figures 4, 5, 6, and 7.

Fig. 4: Boxplot of Biases:  $a$ Fig. 5: Boxplot of MSE:  $a$ Fig. 6: Boxplot of Biases:  $b$ Fig. 7: Boxplot of MSE:  $b$ 

We observe the LHRE suffers from both significant bias and large MSE, while the MLE and Bayes estimators are relatively unbiased. Note that the MLE has the smallest bias spread and has a consistently smaller MSE than the Bayes estimator. In summary, the MLE estimator is both the most accurate and most consistent, but the Bayes estimator is sufficient as well. In choosing which estimator to compute in the future it would

seem logical to pick the one that is easiest to obtain. Both are relatively unbiased and have sufficiently low MSE to provide sound conclusions.

### 3.3 Model Misspecification

We briefly examine properties of our parameter estimates under a model misspecification. We choose the gamma distribution as the alternative model because of its shape. Both the gamma and the Gompertz distributions have positive data and parameter supports, however the gamma distribution is a right-skewed distribution. The Gompertz distribution may not fit well because the skews go in opposite directions. To explore this idea, we fit our estimates where 100 data points come from a  $\text{gamma}(\alpha = 3, \beta = 1)$ . Figure 8 shows how well it fits.

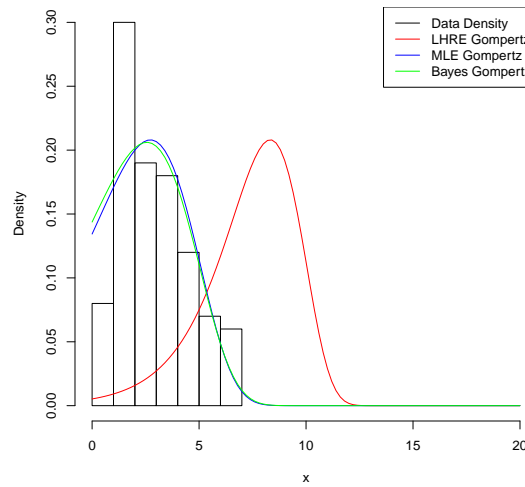


Fig. 8: Fitting Gompertz to  $\text{Gamma}(\alpha = 3, \beta = 1)$  data

Clearly, the LHRE estimated density performs very poorly. This is because the log-hazard function is not linear. The MLE and Bayes estimated densities fit decently, however this is only possible because of the large value for  $a$ . If samples from a gamma distribution were taken where there was very little density near zero, the fits would have been worse. In the case of modeling data near zero, the MLE and Bayes estimated Gompertz distributions are very flexible and can handle fitting even if the data does not come from a Gompertz random variable.

## 4 Applications: Mortality Data

The Centers for Disease Control and Prevention (CDC) published a mortality data set containing information about every death in the United States that occurred in the year 2015, over two-million data points. Note that we obtained the data set from kaggle.com where it was pulled from the CDC database and recompiled in a more flexible format (Centers for Disease Control and Prevention (2016)). The variable of interest is age at time of death, measured in integer years. See figure 9 for a histogram of the data. We remove data points where the age is less than 20 to restrict the observations to adults. Our question of interest is whether the Gompertz distribution should be used to analyze adult mortality in the United States. To answer this question, we will use our three methods of parameter estimation to fit the Gompertz distribution to the data. We will assess parameter estimation accuracy with confidence intervals and include a discussion about how they are computed. Finally we will analyze the results and discuss the appropriateness of using the Gompertz distribution to model mortality.

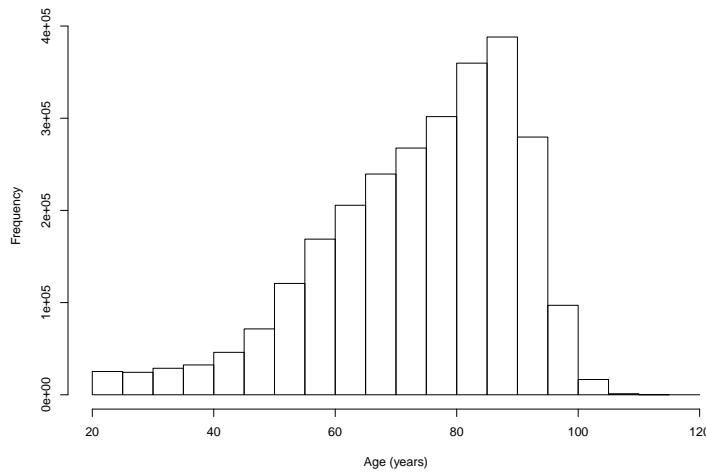


Fig. 9: Histogram of 2015 Mortality Data

### 4.1 Interval Estimation Techniques

We will use confidence intervals to assess the uncertainty in the LHRE and MLE estimators, and we will use a credible interval to assess the uncertainty in the Bayes estimator.

**4.1.1 Credibility Intervals** Credibility intervals for Bayesian estimators are relatively straight forward. A 95% credibility interval is computed by finding the 2.5<sup>th</sup> and 97.5<sup>th</sup> percentiles of the posterior distribution.

Since we have random draws from the posterior distributions from our MCMC process, we can use the corresponding quantiles to estimate the percentiles and obtain the credible interval. We will use an MCMC chain with ten-thousand draws in order to make sure all regions of the posterior sample space are represented correctly.

**4.1.2 Confidence Intervals** In order to obtain confidence intervals on our LHRE and MLE estimates we will use bootstrapping. To bootstrap, we will repeatedly resample from our data with replacement, computing the estimators each time. Computing the 2.5<sup>th</sup> and 97.5<sup>th</sup> quantiles of each estimator will yeild 95% confidence intervals. For each estimate we will compute a thousand bootstrap values. This is less than the number specified in section 4.1.1 because the bootstrapping algorithm generates independent values and thus explores the sample space more efficiently.

## 4.2 Results

Table 6 shows the estimated parameter values for each estimation method and tables 7 and 8 show the lower and upper interval bounds on the stimates. Notice that they are very similar. We will now examine the results from each method in detail.

	<i>a</i>	<i>b</i>
LHRE	0.0001728	0.0743
MLE	0.0001323	0.0781
Bayes	0.0001322	0.0781

Table 6: Parameter Estimation Results

	<i>a</i>	<i>b</i>
LHRE	0.0001719	0.0742
MLE	0.0001316	0.0780
Bayes	0.0001315	0.0780

Table 7: Interval Lower Bound

	<i>a</i>	<i>b</i>
LHRE	0.0001741	0.0743
MLE	0.0001332	0.0782
Bayes	0.0001329	0.0782

Table 8: Interval Upper Bound

### 4.2.1 LHRE

In order for the LHRE to be a viable method of estimation, the empirical log-hazard must be close to linear. Figure 10 confirms that this is roughly the case. However, the pattern is not perfectly linear, which is why the LHRE differs most from the other two estimators. Figure 11 shows that the LHRE fit is decent but it doesn't capture the mode of the data.

### 4.2.2 MLE

Figure 12 shows the fitted Gompertz density under MLE estimation. It also does not capture the mode but still fits decently. Note that even though the estimated parameter values for the MLE and the Bayes estimator are very close, the interval for the MLE is much wider.

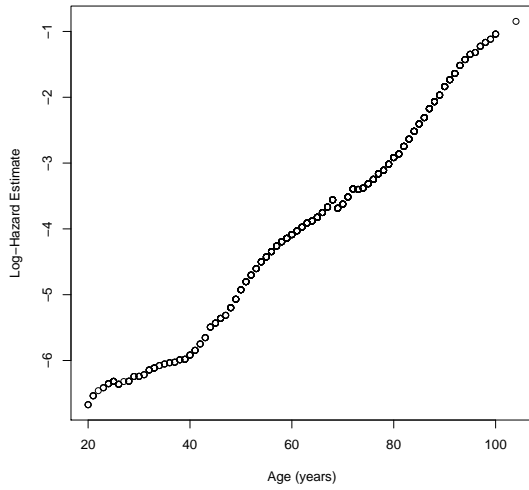


Fig. 10: Linearity of Estimated Log-Hazard Values

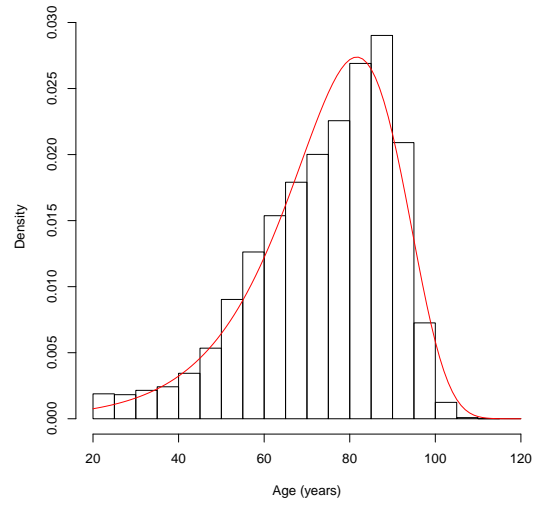


Fig. 11: Histogram of Mortality Data and LHRE fitted Gompertz Density

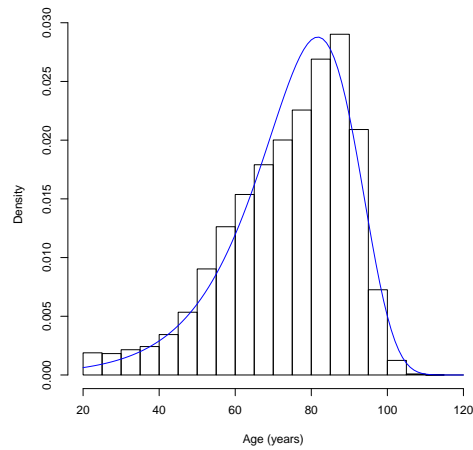
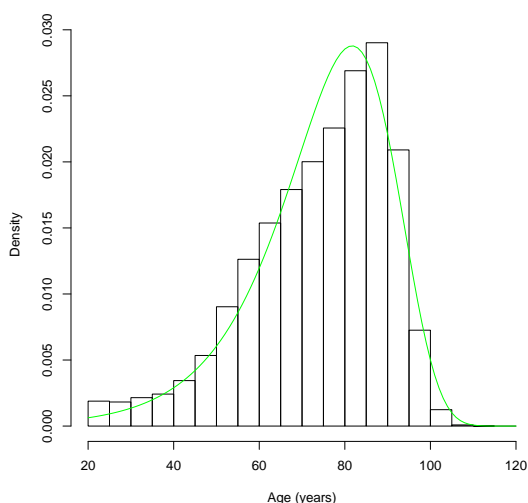


Fig. 12: Histogram of Mortality Data and MLE fitted Gompertz Density

### 4.2.3 Bayes Estimator

Figure 13 shows the Gompertz density under Bayesian estimation. Likewise with the other estimators, the fitted density is decent but does not capture the mode. The interval estimate for the Bayes estimator is the tightest interval of the three, indicating high levels of certainty about the best fitting parameters. Along with the density graph we have included diagnostic plots described in section 2.3 in the appendix. These plots are not ideal as there is a lot of correlation between the parameters, which is why we used ten times as many draws compared to bootstrapping. Increased sample size leads to a better explored sample space, which can be seen in the trace plots.



Acceptance Rate
0.508

Table 9

Fig. 13: Histogram of Mortality Data and Bayesian fitted Gompertz Density

## 4.3 Application Summary

We chose this data set because the Gompertz distribution was designed to model adult human mortality. The results from our estimation procedures reveal that the Gompertz may not be appropriate to analyze 2015 mortality data. Figure 14 shows all three fitted distributions over the data and how they are quite similar. Most notably, they all miss the mode of the data. The consistently erroneous shape in combination with consistently narrow confidence intervals indicate that the issue may not be with the estimation technique, but rather with the distributional assumption. In context, this means that people in the dataset are living longer than the Gompertz distribution would predict. In fact, this is a recently developed phenomenon called late-life mortality deceleration and has directly impacted the viability of using the Gompertz distribution for



modeling lifespan patterns in this century. For current or recent mortality data, it does not seem reasonable to use the Gompertz distribution as a model for that data.

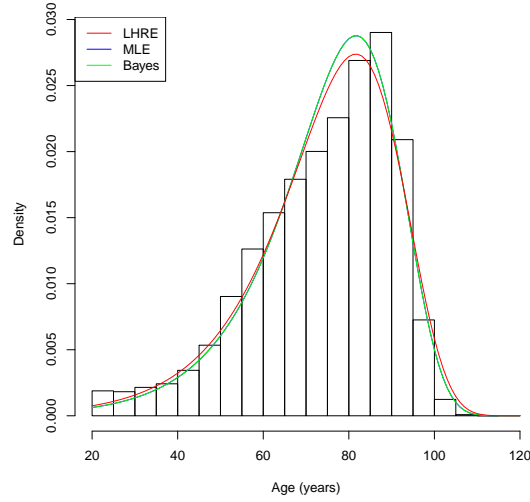


Fig. 14: Density Comparison of Estimation Methods Against Data

## 5 Summary

The Gompertz distribution is based on an exponential hazard rate. The log of this hazard rate can be used to crudely estimate the rate and shape parameters  $a$  and  $b$ . This method is prone to error when analyzing continuous data. Maximum likelihood estimation and Bayesian estimation are much more accurate and consistent. Both methods require iterative numerical methods and so the log-hazard estimates serve as initial values. Simulation studies find that maximum likelihood estimation and bayesian estimation are unbiased and comprable with maximum likelihood being slightly more consistent. Using a mortality data set, the Gompertz distribution is found to be a mediocre modeling tool for adult lifespans.

## 6 Appendix

These are the diagnostic plots referenced in section 4.2.3

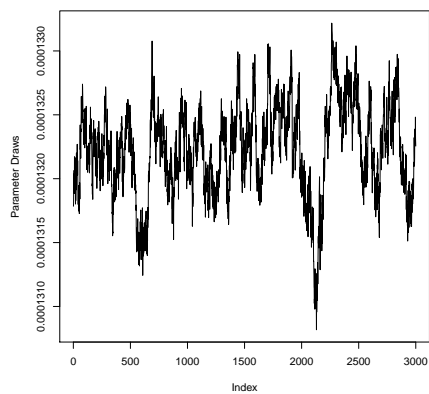


Fig. 15: Trace Plot of  $a$

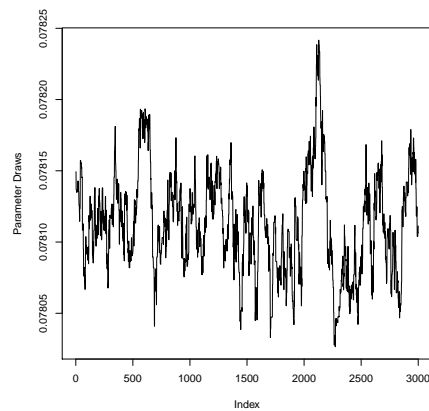


Fig. 16: Trace Plot of  $b$

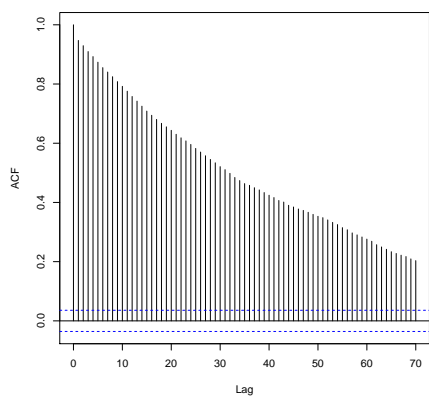


Fig. 17: Auto Correlation Function of  $a$

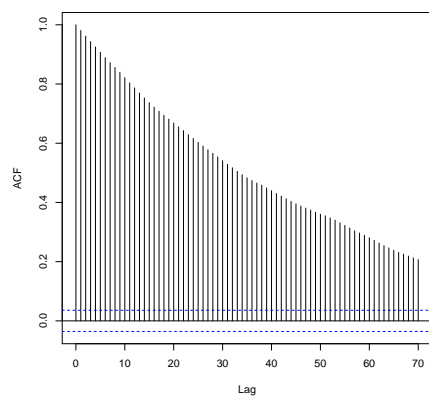


Fig. 18: Auto Correlation Function of  $b$

## Bibliography

Centers for Disease Control and Prevention (2016) Death in the United States, 2015. Accessed through Kaggle database, <https://www.kaggle.com/cdc/mortality>