NORTH-HOLLAND

Adaptive Bayes Estimators for Parameters of the Gompertz Survival Model

Malwane M. Ananda, Rohan J. Dalpatadu, and Ashok K. Singh

Department of Mathematical Sciences
University of Nevada
Las Vegas, Nevada 89154

Transmitted by John Casti

ABSTRACT

The two-parameter Gompertz model is a commonly used survival time distribution in actuarial science and reliability and life testing. The estimation of the parameters of this model is numerically involved. We consider the estimation problem in a Bayesian framework and give the Bayesian estimators of parameters in terms of single numerical integrations. We propose an adaptive Bayesian estimation procedure by putting a prior only on one parameter and finding the other parameter by minimizing the distance between empirical and parametric cumulative distribution functions. This easily computable (even for large samples) adaptive Bayesian procedure is compatible with the exact Bayesian procedure. In particular, numerical integration for computing the exact Bayesian procedure is difficult for large samples. Furthermore, for the no prior information situation, a noninformative adaptive Bayes procedure is given. Some examples of the proposed adaptive method along with a comparison with other existing methods are given. Monte Carlo simulation has been used to compare the existing procedures with the proposed procedures.

1. INTRODUCTION

Gompertz distribution is a widely used distribution in life testing and reliability studies. We consider the problem of estimating parameters of the

APPLIED MATHEMATICS AND COMPUTATION 75:167–177 (1996)

655 Avenue of the Americas, New York, NY 10010

0096-3003/96/$15.00
SSDI 0096-3003(95)00124-Z

two-parameter Gompertz distribution. Suppose the random variable X has the two parameter Gompertz distribution $f_X(x)$,

$$f_X(x) = bc^x \exp\left\{\frac{b}{\ln c}(1 - c^x)\right\}, \qquad x > 0 \tag{1}$$

with hazard rate $\lambda(x)$ is

$$\lambda(x) = bc^x, \qquad x > 0 \tag{2}$$

where both b and c are unknown parameters and $b > 0$ and $c > 1$.

Based on a random complete sample $x_1, x_2, \ldots, x_n$ we would like to estimate these parameters. There are many estimation methods available in the literature for these parameters: maximum likelihood estimators, percentile estimators (estimators based on two selected percentiles; for details see London [1]), minimum chi-squared estimators, minimum modified chi-squared estimators, least square estimators, (for details see Elandt-Johnson and Johnson [2] or London [1]), etc. But none of these estimators has closed forms, and estimation of these parameters involve numerical computations. For example, maximum likelihood estimators of these parameters can be found by maximizing the likelihood function

$$L = b^n c^{\Sigma x_i} \exp\left\{b\left(n - \sum c^{x_i}\right)/\ln c\right\}$$

or, alternatively, by finding the solution for the following two nonlinear equations

$$n/b + \left(n - \sum c^{x_i}\right)/\ln c = 0$$

$$\Sigma x_i/c - b\left(\sum x_i c^{x_i - 1}\right)/\ln c - b(n - \Sigma c^{x_i})/\left\{c(\ln c)^2\right\} = 0.$$

We consider the problem from the Bayesian point of view using two independent priors; a gamma prior for the parameter b and any other prior on the parameter c. The Bayesian estimators for these parameters are given as a ratio of one-dimensional integrals. For small samples, one can evaluate these integrals numerically. But for large samples, evaluating these integrals or getting an approximation to these integrals is difficult.

We propose adaptive Bayesian estimators for the above problem by putting a prior only on one parameter and finding the other unknown parameter by minimizing the distance between empirical and parametric cumulative distribution functions. The proposed procedure works well for large n as well as for small n. Furthermore, we derive the estimators corresponding to noninformative priors. Examples for these procedures are given using actual data sets. Monte Carlo simulation is used to compare these methods with the maximum likelihood procedure.

2. BAYESIAN ESTIMATORS

Consider the gamma prior distribution (with parameters $\alpha > 0$ and $\beta > 0$) on the parameter b and any other prior $g(c)$ on the parameter c. Assuming that the parameters b and c are independent, the joint prior distribution of b and c is

$$\pi(b, c) = \frac{e^{-b/\beta} b^{\alpha-1}}{\Gamma(\alpha)\beta^{\alpha}} g(c) \tag{3}$$

and one can show that the posterior distribution of b and c given the data $\pi(b, c/\text{data})$ is proportional to

$$\pi(b, c/\text{data}) \propto b^{n+\alpha-1} c^{\Sigma x_i} g(c) \exp\{b(n - \Sigma c^{x_i})/\ln c - b/\beta\}.$$

Easy calculation shows that the Bayesian estimators $\hat{b}$ and $\hat{c}$ (assuming the quadratic loss) of the parameters b and c are

$$\hat{b} = \frac{(n+\alpha)\int_1^{\infty} c^{\Sigma x_i} g(c)\{1/\beta + (\Sigma c^{x_i} - n)/(\ln c)\}^{-(n+\alpha+1)}\, dc}{\int_1^{\infty} c^{\Sigma x_i} g(c)\{1/\beta + (\Sigma c^{x_i} - n)(\ln c)\}^{-(n+\alpha)}\, dc}$$

and

$$\hat{c} = \frac{\int_1^{\infty} c^{\Sigma x_i + 1} g(c)\{1/\beta + (\Sigma c^{x_i} - n)/(\ln c)\}^{-(n+\alpha)}\, dc}{\int_1^{\infty} c^{\Sigma x_i} g(c)\{1/\beta + (\Sigma c^{x_i} - n)(\ln c)\}^{-(n+\alpha+1)}\, dc}.$$

If one uses the noninformative prior

$$\pi(b, c) = b^{\alpha-1} db\, dc; \qquad b > 0, \quad c > 1, \tag{4}$$

the generalized Bayes estimator for these parameters are

$$\hat{b} = \frac{(n+\alpha)\int_1^\infty c^{\Sigma x_i} g(c)\{(\Sigma c^{x_i} - n)/(\ln c)\}^{-(n+\alpha+1)}\, dc}{\int_1^\infty c^{\Sigma x_i} g(c)\{(\Sigma c^{x_i} - n)/(\ln c)\}^{-(n+\alpha)}\, dc}$$

and

$$\hat{c} = \frac{\int_1^\infty c^{\Sigma x_i + 1} g(c)\{(\Sigma c^{x_i} - n)/(\ln c)\}^{-(n+\alpha)}\, dc}{\int_1^\infty c^{\Sigma x_i} g(c)\{(\Sigma c^{x_i} - n)/(\ln c)\}^{-(n+\alpha+1)}\, dc}.$$

These integrals must be evaluated numerically. For small n, this can be done quite easily using any integration program such as QDAGS in IMSL. For large n, calculating these integrals is difficult. Usually this is done by using the Tierney and Kadane [3] type or some other type of approximation. These approximations require the value of the second derivative of the integrand at the maximum. Because of this, these methods do not provide accurate answers for the problem at hand.

3. ADAPTIVE BAYESIAN ESTIMATORS

First let us assume that the value of the parameter c is known. Also we assume that the prior information about the parameter b can be expressed using the gamma prior with parameters α and β,

$$\pi(b) = e^{-b/\beta} b^{\alpha-1}/\{\Gamma(\alpha)\beta^\alpha\}. \tag{5}$$

Easy calculation shows that the posterior distribution of b given the data, $\pi(b/\text{data})$ is proportional to

$$\pi(b/\text{data}) \propto b^{n+\alpha-1} \exp\{-b/\beta + b(n - \Sigma c^{x_i})/\ln c\}$$

and the Bayesian estimator of b, $\hat{b}$ is

$$\hat{b} = \frac{(n+\alpha)}{\{1/\beta + (\Sigma c^{x_i} - n)/\ln c\}}.$$

Now we find the optimum c value by minimizing the "distance" between empirical cdf ($F_{\text{emp}}(x)$) and the parametric cdf $F(x)$. Here $F_{\text{emp}}(x) = $ (#of obs. $\leqslant x)/n$ and $F(x) = 1 - \exp\{b(1 - c^x)/\ln c\}$.

We use following distance criteria in our analysis.

(a) Area (A_1) between $F_{\text{emp}}(x)$ and $\hat{F}(x)$:

$$A_1 = \int_0^\infty \left| F_{\text{emp}}(x) - \hat{F}(x) \right| dx \tag{6}$$

(b) Anderson and Darling [4] A^2 statistic:

$$A^2 = \int_0^\infty \frac{n}{(1 - F(x))\,F(x)} \left[F_{\text{emp}}(x) - F(x) \right]^2 f(x)\, dx.$$

This is the expected squared deviation of $F_{\text{emp}}(x)$ and $F(x)$, with the deviations weighted by the inverse variance of $F_{\text{emp}}(x)$. By replacing $F(x)$ and $f(x)$ by their estimated values $\hat{F}(x)$ and $\hat{f}(x)$, one gets the A^2 statistic as

$$A^2 = \int_0^\infty \frac{n}{\left(1 - \hat{F}(x)\right)\hat{F}(x)} \left[F_{\text{emp}}(x) - \hat{F}(x) \right]^2 \hat{f}(x)\, dx$$

which can be written as

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^{n} (2i - 1)\left\{ \ln\left[\hat{F}(x_i) \cdot \left(1 - \hat{F}(x_{n-i+1})\right)\right]\right\}. \tag{7}$$

Studies by Stephens [5] show that, in many important situations, the A^2 statistic provides a better measure for the departure between $\hat{F}(x)$ and $F_{\text{emp}}(x)$ than the Kolmogorove-Smirnov D statistic, which is defined as

$$D = \sup_x \left| F_{\text{emp}}(x) - \hat{F}(x) \right|.$$

If no prior information is available about the parameter b, one can use the noninformative generalized prior, which has the density

$$\pi(b) = b^{\alpha - 1} \qquad 0 < b < \infty. \tag{8}$$

When c is known, the Jeffreys noninformative prior [6] is proportional to $\pi(b) \propto \sqrt{|I(b)|}$, where $I(b)$ is the Fisher information. So the Jeffreys noninformative prior is $\pi(b) = 1/b$, which corresponds to $\alpha = 0$, $\beta = \infty$ with the notation given in (5). The posterior density of b with respect to the prior given in (8) is

$$\pi(b/\text{data}) \propto b^{n+\alpha-1} \exp\{-b(\Sigma c^{x_i} - n)/\ln c\}$$

and the generalized Bayesian estimator of b is

$$\hat{b} = \frac{(n+\alpha)\ln c}{(\Sigma c^{x_i} - n)}.$$

Again one can find c by minimizing the area A_1 given in (6) or the Anderson and Darling [4] A^2 statistic given in (7). Notice that, when c is known, the generalized Bayesian estimator of b with respect to the Jeffreys noninformative prior is the ML estimator of b.

In the next section, a couple of examples are given for the proposed methods along with a comparison of existing methods. Monte Carlo simulations were carried out to compare the performances of these methods and their results are given in Section 5.

4. CASE EXAMPLES

EXAMPLE 1. The following data from Hoel [7] represents time (in days) at death of 39 irradiated mice. These life times in days are 40, 42, 51, 62, 163, 179, 206, 222, 228, 249, 252, 282, 324, 333, 341, 366, 385, 407, 420, 431, 441, 461, 462, 482, 517, 517, 524, 564, 567, 586, 619, 620, 621, 622, 647, 651, 686, 761, 763.

Elandt-Johnson and Johnson [2] showed that these data follow a two-parameter Gompertz distribution. Since we do not have any prior information, we cannot calculate the exact Bayesian procedure. We calculate the adaptive generalized Bayesian estimators (with respect to the noninformative prior in (5) with $\alpha = 0$ and $\beta = \infty$) by minimizing 1) area, and 2) the Anderson and Darling statistic. The chi-squared goodness-of-fit test was used to compare these methods with the other methods. Results of the analysis are given in the Table 1. The p value with the maximum likelihood procedure is 0.307. The p values with the adaptive generalized Bayes procedures are 0.791 (by minimizing the area) and 0.729 (by minimizing the

TABLE 1
RESULTS FROM EXAMPLE 1

Time interval	Observed number	Expected number of deaths			
		M_1	M_2	M_3	M_4
0–100	4	6.901	4.444	2.592	2.723
100–200	2	6.763	5.306	3.729	3.845
200–300	6	6.330	6.002	5.104	5.165
300–400	5	5.603	6.306	6.458	6.422
400–500	7	4.635	5.986	7.232	7.090
500–600	6	3.532	4.946	6.700	6.527
600–700	7	2.436	3.381	4.632	4.557
700–800	2	1.490	1.785	2.048	2.096
$df =$		3	3	3	3
$\chi^2 =$		10.29	3.601	1.044	1.299
p value =		0.016	0.307	0.791	0.729

M_1: Percentile estimators (based on 25th and 75th percentiles); $\hat{c} = 1.00195$; $\hat{b} = 0.00176393$.

M_2: Maximum likelihood estimators; $\hat{c} = 1.00321$; $\hat{b} = 0.00102648$.

M_3: Adaptive generalized Bayes (by minimizing the area, $\alpha = 0$ and $\beta = \infty$); $\hat{c} = 1.00453$; $\hat{b} = 0.00054404$.

M_4: Adaptive generalized Bayes (by minimizing Anderson and Darling statistic, $\alpha = 0$ and $\beta = \infty$); $\hat{c} = 1.00438$; $\hat{b} = 0.00057717$.

Anderson and Darling statistic). Here, performances of the adaptive noninformative procedures are much better than all the other procedures.

EXAMPLE 2. For the second example, we use data from Elandt-Johnson and Johnson [2, p. 136] (original data given by Kimball [8]). These are mortality data for 208 mice, which were exposed to gamma radiation. Maximum likelihood, minimum chi-square, minimum modified chi-square estimators of parameters b and c are given in Elandt-Johnson and Johnson [2]. Since we do not have any prior information, we analyze this data set using the noninformative prior. Results of these analyses (estimated parameter values and chi-square comparison) are given in the Table 2. Because minimum chi-square estimators and minimum modified chi-square estimators are readily available from Elandt-Johnson and Johnson, those methods are also used in the chi-square comparison. While the p value for the ML procedure is 0.178, p values for the adaptive procedures are 0.188 (by minimizing area) and 0.200 (by minimizing Anderson and Darling statistic).

TABLE 2
RESULTS FROM EXAMPLE 2

Time interval	Observed number	Expected number of deaths M_1	M_2	M_3	M_4	M_5	M_6
0–50	3	27.341	6.446	6.664	7.031	6.595	5.717
50–60	3	5.024	3.497	3.580	3.721	3.542	3.188
60–70	6	4.888	5.054	5.152	5.316	5.095	4.667
70–80	6	4.752	7.229	7.335	7.515	7.256	6.766
80–90	16	4.621	10.183	10.286	10.461	10.176	9.667
90–100	14	4.494	14.02	14.10	14.23	13.95	13.51
100–110	25	4.371	18.68	18.70	18.74	18.53	18.29
110–120	20	4.249	23.71	23.63	23.51	23.45	23.60
120–130	32	4.132	28.01	27.81	27.50	27.68	28.33
130–140	25	4.019	29.79	29.50	29.05	29.48	30.59
140–150	27	3.908	27.15	26.89	26.46	27.04	28.22
150–160	13	3.799	19.76	19.66	19.47	19.96	20.65
160–170	11	3.694	10.390	10.459	10.536	10.77	10.76
170–180	7	3.593	3.417	3.522	3.676	3.705	3.427
$df =$		9	9	9	9	9	9
$\chi^2 =$		668.7	12.66	12.46	12.25	12.60	13.82
p value =		0	0.178	0.188	0.200	0.181	0.129

M_1: Percentile estimators (based on 10th and 90th percentiles); $\hat{c} = 1.00003$; $\hat{b} = 0.00281676$.

M_2: Maximum likelihood estimators; $\hat{c} = 1.03975$; $\hat{b} = 0.00020389$.

M_3: Adaptive generalized Bayes estimator (by minimizing the area, $\alpha = 0$ and $\beta = \infty$); $\hat{c} = 1.03935$; $\hat{b} = 0.00021348$.

M_4: Adaptive generalized Bayes estimator (by minimizing Anderson and Darling statistic, $\alpha = 0$ and $\beta = \infty$); $\hat{c} = 1.03871$; $\hat{b} = 0.000230039$.

M_5: Minimum chi-square estimators; $\hat{c} = 1.03931$; $\hat{b} = 0.0002115$ (from Johnson and Johnson [2]).

M_6: Minimum modified chi-square estimators; $\hat{c} = 1.04090$; $\hat{b} = 0.000174$ (from Johnson and Johnson [2]).

Here too, adaptive noninformative procedures are doing slightly better than all the other procedures.

5. RESULTS OF SIMULATION STUDY

In this section, results of two simulation studies are given (one simulation study for a small sample–sample size = 10; and the other one for a large sample–sample size = 30). Simulations were carried out as follows. Parame-

TABLE 3
FIRST SIMULATION EXAMPLE WITH SAMPLE SIZE $n = 10$

Method		Percentiles of estimated parameters (out of 1000 iterations)				
		10th	25th	50th	75th	90th
M_1	$\hat{c}$:	1.0000	1.0000	1.0004	1.0618	1.2897
	$\hat{a}$:	0.0562	0.0842	0.1233	0.1965	0.2989
M_2	$\hat{c}$:	1.0001	1.0385	1.1142	1.1848	1.2735
	$\hat{a}$:	0.0357	0.0567	0.0849	0.1276	0.1734
M_3	$\hat{c}$:	1.0023	1.0523	1.1063	1.1828	1.2771
	$\hat{a}$:	0.0338	0.0547	0.0875	0.1269	0.1658
M_4	$\hat{c}$:	1.0001	1.0211	1.0805	1.1565	1.2451
	$\hat{a}$:	0.0389	0.0637	0.0994	0.1411	0.1825
M_5	$\hat{c}$:	1.0694	1.0879	1.1161	1.1373	1.1503
	$\hat{a}$:	0.0558	0.0689	0.0875	0.1093	0.1361
M_6	$\hat{c}$:	1.0055	1.0523	1.1051	1.1777	1.2737
	$\hat{a}$:	0.0346	0.0551	0.0873	0.1270	0.1671
M_7	$\hat{c}$:	1.0000	1.0254	1.0823	1.1554	1.2377
	$\hat{a}$:	0.0390	0.0633	0.0988	0.1401	0.1861

Actual values: $c = 1.06783$, $b = 0.10129$, $\alpha = 0.5$, $\beta = 0.25$, $n = 10$.

M_1: Percentile method.

M_2: Maximum likelihood method.

M_3: Adaptive Bayes (using partial prior information and by minimizing area); $\alpha = .5$; $\beta = .25$.

M_4: Adaptive Bayes (using partial prior information and by minimizing Anderson and Darling statistic); $\alpha = .5$; $\beta = .25$.

M_5: Exact Bayesian method (using full prior information); $\alpha = .5$; $\beta = .25$.

M_6: Adaptive generalized Bayes (using no prior information and by minimizing area); $\alpha = 0$; $\beta = \infty$.

M_7: Adaptive generalized Bayes (using no prior information and by minimizing area); $\alpha = 0$; $\beta = \infty$.

ter value b was generated from the Gamma distribution with $\alpha = 0.5$ and $\beta = 0.25$. Parameter value c was generated from the uniform distribution $U(1.02, 1.22)$. These generated parameter values are $b = 0.10129$ and $c = 1.06783$. Using these values as actual population values, 1000 random samples were generated for each simulation example (for the first simulation example, sample size is 10 and for the second simulation example, sample size is 30). For each simulated sample, parameters b and c were estimated using the percentile method (based on the 25th and 75th percentile), the maximum likelihood method, the exact Bayesian method (using full prior

TABLE 4
SECOND SIMULATION EXAMPLE WITH SAMPLE SIZE $n = 30$

Method		Percentiles of estimated parameters (out of 1000 iterations) 10th	25th	50th	75th	90th
M_1	$\hat{c}$	1.0000	1.0002	1.0007	1.0458	1.0818
	$\hat{a}$	0.0778	0.0943	0.1126	0.1393	0.1874
M_2	$\hat{c}$	1.0001	1.0364	1.0810	1.1130	1.1459
	$\hat{a}$	0.0601	0.0767	0.0977	0.1291	0.1604
M_3	$\hat{c}$	1.0288	1.0523	1.0788	1.1097	1.1474
	$\hat{a}$	0.0610	0.0775	0.0968	0.1179	0.1419
M_4	$\hat{c}$	1.0188	1.0440	1.0700	1.1034	1.1365
	$\hat{a}$	0.0652	0.0808	0.1018	0.1249	0.1477
M_5	$\hat{c}$	1.0294	1.0523	1.0788	1.1083	1.1457
	$\hat{a}$	0.0611	0.0773	0.0971	0.1180	0.1417
M_6	$\hat{c}$	1.0194	1.0445	1.0700	1.1023	1.1348
	$\hat{a}$	0.0654	0.0809	0.1017	0.1246	0.1476

Actual Values: $c = 1.06783$, $b = 0.10129$, $\alpha = 0.5$, $\beta = 0.25$, $n = 30$.
M_1: Percentile method.
M_2: Maximum likelihood method.
M_3: Adaptive Bayes (using partial prior information and by minimizing area); $\alpha = .5$; $\beta = .25$.
M_4: Adaptive Bayes (using partial prior information and by minimizing Anderson and Darling statistic); $\alpha = .5$; $\beta = .25$.
M_5: Adaptive generalized Bayes (using no prior information and by minimizing area); $\alpha = 0$; $\beta = \infty$.
M_6: Adaptive generalized Bayes (using no prior information and by minimizing Anderson and Darling statistic); $\alpha = 0$; $\beta = \infty$.

knowledge), the generalized Bayes method, and the proposed adaptive Bayes methods. Because of numerical difficulties, (since the sample size is large), the exact Bayesian procedure was not used in the second simulation example. For these 1000 iterations, 10th, 25th, 50th, 75th, and 90th percentiles of these estimated parameters are given in Table 3 and Table 4.

These simulation results show that overall performances of these adaptive procedures are better than those of the maximum likelihood procedure and the percentile estimation procedure. Furthermore, these results show that (this comparison is done only for small n) the performances of the adaptive Bayes procedures are quite close to the performances of the exact Bayesian procedure. It is also important to notice that, even for large n, not like the exact Bayesian procedure, calculation of these adaptive Bayes procedures (in particular, by minimizing the Anderson and Darling statistic) is much easier. So using these adaptive Bayes procedures, not only can one incorporate some prior information, but also one may gain some numerical advantages.

REFERENCES

1 D. London, *Survival Models and Their Estimation.* ACTEX publications, 1988.
2 R. C. Elandt-Johnson, and N. L. Johnson, *Survival Models and Data Analysis*, John Wiley and Sons, New York, 1980.
3 L. Tierney, and J. B. Kadane, (1986): Accurate approximations for posterior moments and marginal densities, *J. Am. Statist. Assoc.* 81:82–86 (1986).
4 T. W. Anderson, and D. A. Darling, A test of goodness of fit, *J. Am. Statist. Assoc.* 49:765–769 (1954).
5 M. A. Stephens, EDF statistics for goodness of fit and some comparisons, *J. Am. Statist. Assoc.* 69:730–737 (1974).
6 H. Jeffreys, *The Theory of Probability*, Oxford University Press, Oxford, UK, 1961.
7 D. G. Hoel, A representation of mortality data by competing risks, *Biometrics*, 28:475–488 (1972).
8 A. W. Kimball, Estimation of mortality intensities in animal experiments, *Biometrics*, 16:505–521 (1960).