Goodness-of-fit tests for the Gompertz distribution

Adam Lenart and Trifon I. Missov

Max Planck Institute for Demographic Research
Konrad-Zuse-Strasse 1
18057 Rostock, Germany

Abstract

The Gompertz distribution is often fitted to lifespan data, however testing whether the fit satisfies theoretical criteria was neglected. Here five goodness-of-fit measures, the Anderson-Darling statistic, the Kullback-Leibler discrimination information, the correlation coefficient test, testing for the mean of the sample hazard and a nested test against the generalized extreme value distributions are discussed. Along with an application to laboratory rat data, critical values calculated by the empirical distribution of the test statistics are also presented.

1 Introduction

Goodness-of-fit tests determine if the empirical distribution of the data satisfy the assumptions of theoretical distributions. While the Gompertz distribution is routinely used in demography, biology, actuarial and medical science, according to my best knowledge, there is no analysis published about the goodness-of-fit for it. However, the Gompertz distribution is a degenerate generalized extreme value distribution for the minima and an abundance of goodness-of-fit tests exist in the literature for other extreme value distributions (see e.g., Hosking 1984).

In a landmark paper Shannon (1948) defined the entropy of distributions and Kullback and Leibler (1951) were the first to measure the distance between probability distributions based on their entropy. Later, Song (2002) operationalized the Kullback-Leibler distance to test the goodness-of-fit of distributions. Recently, Pérez-Rodríguez et al. (2009) applied it to the Gumbel distribution.

In another important article, Anderson and Darling (1952) developed the Anderson-Darling test that later Stephens (1977) analyzed in the context of extreme value distributions. Sinclair et al. (1990) modified the Anderson-Darling test to allow different weighting schemes that emphasize either the lower or the upper tail of the distributions.

Filliben (1975) used the Pearson correlation coefficient to check the correlation between expected statistics of a theoretical distribution and sample statistics. The correlation coefficient test was the most popular in hydrology (Vogel 1986; Kinnison 1989) to assess the fit of extreme value distributions.

The likelihood ratio test naturally arises to account for the differences between the Gompertz and other extreme value distributions. The generalized extreme value distribution is characterized by μ, location, σ, scale and ξ shape parameters. For $\xi = 0$, the generalized extreme value distribution reduces to the Gumbel, and the Gompertz distribution is a reversed and truncated Gumbel distribution with additional correlation between its parameters a and b. The different parametrization of the Gompertz distribution removes it from location-scale family of distributions.

This paper will first briefly describe each of these tests and apply them to the Gompertz distribution. Additionally, based on the observation that the maximum likelihood estimators of a and b equal the mean of the sample hazard in the Gompertz distribution (Lenart 2012), a fifth test will be defined. The final sections of the paper compare the power of the tests against alternative distributions and derive critical values of them based on Monte Carlo simulation experiments. An application of the tests to laboratory rat data is also discussed.

2 Application: goodness-of-fit to laboratory rat data

The goodness-of-fit tests defined above can be readily used to check if empirical data is Gompertz distributed. As an example, individual life span data of rats will be used. The analyzed data was collected by Vladimir N. Anisimov at the N.N.Petrov Research Institute of Oncology, St.Petersburg, Russia to test carcinogenicity and it is now published in the Biodemographic Database (BDB) at the Max Planck Institute for Demographic Research. Here we will use only the rats in the control group, $n = 51$ females and $n = 46$ males. The data is fully observed and the number of survivors was recorded every day. (Please see Figure 1) for the estimated hazard and the Kaplan-Meier survival function and Table 1 for descriptive statistics of the dataset. The hazard estimation was carried out by the same varying kernel width estimation procedure as mentioned earlier. The Gompertz fit to the data show very wide confidence intervals which were estimated by the delta method.

The goodness-of-fit statistics in general do not reject the null hypothesis that both the distribution of death of both the male and the female rats is Gompertz. (Table 2) While the maximum likelihood estimate of a of the male rats is higher than $\hat{a}$ of the female rats, the estimated daily rate of aging parameter, $\hat{b}$ is lower, leading to a cross-over of mortality later in life. (Figure 1)

This result is corroborated by the non-parametric estimates. However, because of the low sample size, the confidence bands are very wide. In spite of that, by looking at the goodness-of-fit statistics and their respective critical values in the appendix, it can be seen that the

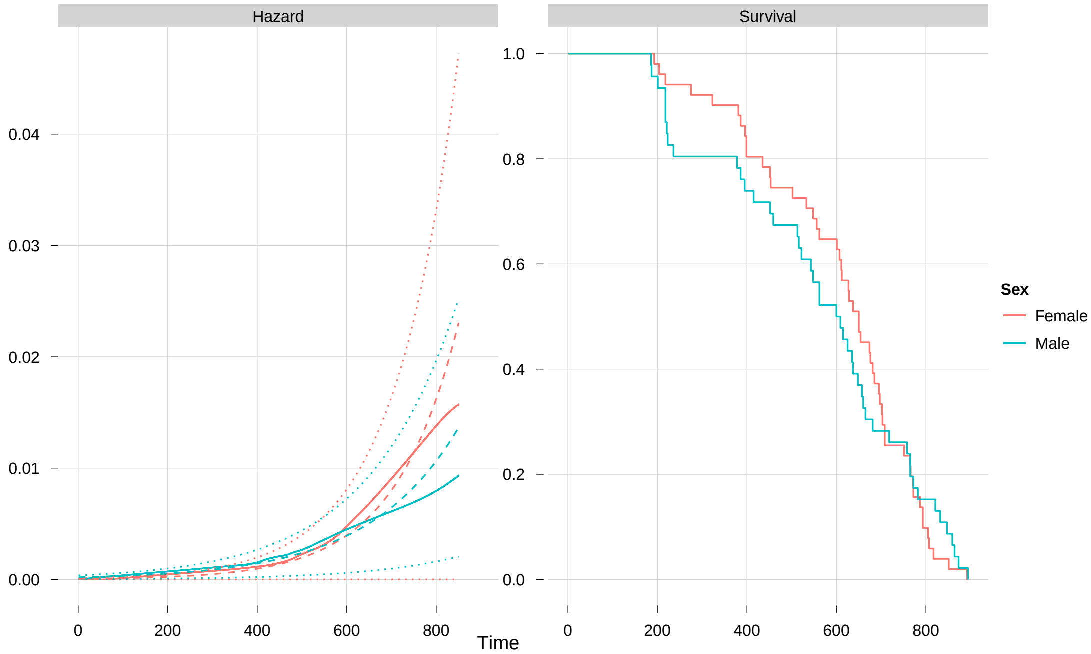

Figure 1: Hazard and survival of the rat data. On the left panel, the solid line corresponds to the non-parametric hazard estimate, the dashed line to the Gompertz fit and the dotted lines are the 95% confidence intervals of the fitted Gompertz hazard.

null is not rejected either by the Anderson-Darling ($0.384 < 0.63$ and $0.55 < 0.62$) and the correlation coefficient ($0.991 > 0.973$ and $0.983 > 0.976$) test statistics at $\alpha = 0.1$. The Kullback-Leibler statistic does not reject the null hypothesis at $\alpha = 0.1$ in the case of females ($5.6\times\ 10^{-6} < 0.25$) but would reject it in the case of males, however at higher *alpha*, the null hypothesis is not rejected ($0.23 < 0.24$). The likelihood ratio test also confirms that the Gompertz distribution fits the data as well as the generalized extreme value distribution (its shape parameter equals to 0) at $\alpha = 0.1$ for both females ($0.895 < 2.71$) and males (2.165 < 2.71). The quantiles of the normal distribution estimated by the test for the mean of the sample hazard envelops 0 only at $\alpha = 0.01$, otherwise it would reject the null. However, this test is not reliable at this low sample size because of the difficulties of estimating the sample hazard.

Sex	n	Min	q_1	$\widetilde{x}$	$\bar{x}$	q_3	**Max**	s	**IQR**
Female	51	192.5	477.0	649.5	603.2	729.0	891.5	177.9	252
Male	46	185.5	399.5	604.0	559.1	747.5	893.5	219.4	348

Table 1: Descriptive statistics of life spans of 51 female and 46 male rats (days)

Sex	$\hat{a}$	$\hat{b}$	$\bar{\mu}_{\alpha=0.01}$	r	KL	AD	LR
Female	5.7×10^{-5}	0.007	-0.0014 – 0.0029	0.991	$5.6\times\ 10^{-6}$	0.384	0.895
Male	$1.9\times\ 10^{-3}$	0.005	-0.00012 – 0.0034	0.983	0.23	0.55	2.165

Table 2: Calculated Gompertz goodness-of-fit test statistics for the dataset of 51 female and 46 male rats

3 Discussion

The comparison of the power of the tests show that the Anderson-Darling statistic is the most powerful in rejecting the null that the empirical distribution comes from the Gompertz distribution when it was simulated from an alternative distribution. The Anderson-Darling statistic implemented by its computing formula is also the simplest and the quickest to run, and an important advantage of it is that for low values of a, the distribution of the statistic is independent from the Gompertz a and b parameters.

The correlation coefficient test also efficiently refutes other alternative distributions, however, when the alternative distribution is closely related to the Gompertz, such as in the case of Weibull and Gamma-Gompertz distributions, the power of the correlation coefficient test drops. As Legates and McCabe Jr (1999) noted, the tests based on correlation are overly sensitive to outliers and insensitive to proportional differences between the expected and the observed values.

The main problem with testing the mean of the sample hazard lies in the estimation of the sample hazard. Here locally optimal varying kernels were used (Müller and Wang 1994) and as the variance of the kernel hazard estimate was not taken into account, the probability of committing Type I error at $\alpha = 0.05$ reduced to 5% only by $n = 400$. Other sample hazard estimators would necessarily yield different efficiency and critical values.

Juxtaposed with the results for the Gumbel distribution (Pérez-Rodríguez et al. 2009), the Kullback-Leibler test performs unexpectedly poorly relative to the other tests. The main disadvantage of the Kullback-Leibler test lies in the estimation of the sample entropy and choosing the optimal window width for that as it can vary from dataset to dataset with the same sample size and a different window width entails different critical values of the statistic. In the Appendix, the optimal window width is not reported as it was optimized at each draw from the Gompertz distribution, therefore the critical values correspond to an average of the optimal window widths that can be expected when sampling from the Gompertz distribution.

The likelihood ratio test is a powerful test when the alternative distribution is from the generalized extreme value family. A positive externality of the test is that the shape parameter of the generalized extreme value distribution, ξ has to be estimated during the testing procedure. If $\xi < 0$ and the likelihood ratio at the chosen significance level rejects the null hypothesis that $\xi = 0$, than the empirical distribution can be better fitted by a Weibull distribution than by a Gompertz. If $\xi > 0$, the empirical distribution is more likely to be Fréchet-type than Gompertz (Jenkinson 1955).

References

Anderson, T. and D. Darling. 1952. "Asymptotic theory of certain" goodness of fit" criteria based on stochastic processes." *The Annals of Mathematical Statistics* 23(2):193–212.

Filliben, J. 1975. “The probability plot correlation coefficient test for normality.” *Technometrics* 17(1):111–117.

Hosking, J. 1984. “Testing whether the shape parameter is zero in the generalized extreme-value distribution.” *Biometrika* 71(2):367–374.

Jenkinson, A. 1955. “The frequency distribution of the annual maximum (or minimum) values of meteorological elements.” *Quarterly Journal of the Royal Meteorological Society* 81(348):158–171.

Kinnison, R. 1989. “Correlation coefficient goodness-of-fit test for the extreme-value distribution.” *The American Statistician* 43(2):98–100.

Kullback, S. and R. Leibler. 1951. “On information and sufficiency.” *The Annals of Mathematical Statistics* 22(1):79–86.

Legates, D. and G. McCabe Jr. 1999. “Evaluating the use of “goodness-of-fit” measures in hydrologic and hydroclimatic model validation.” *Water Resources Research* 35(1):233–241.

Lenart, A. 2012. “The Gompertz distribution and maximum likelihood estimation of its parameters-a revision.” MPIDR Working Paper WP-2012-008.

Müller, H. and J. Wang. 1994. “Hazard rate estimation under random censoring with varying kernels and bandwidths.” *Biometrics* 50(1):61–76.

Pérez-Rodríguez, P., H. Vaquera-Huerta and J. Villaseñor-Alva. 2009. “A Goodness-of-Fit Test for the Gumbel Distribution Based on Kullback–Leibler Information.” *Communications in StatisticsTheory and Methods* 38(6):842–855.

Shannon, C. 1948. “A mathematical theory of communication.” *The Bell Systems Technical Journal* 27:379–423, 623–656.

Sinclair, C., B. Spurr and M. Ahmad. 1990. “Modified Anderson darling test.” *Communications in Statistics-Theory and Methods* 19(10):3677–3686.

Song, K. 2002. “Goodness-of-fit tests based on Kullback-Leibler discrimination information.” *IEEE Transactions on Information Theory* 48(5):1103–1117.

Stephens, M. 1977. “Goodness of fit for the extreme value distribution.” *Biometrika* 64(3):583–588.

Vogel, R. 1986. “The probability plot correlation coefficient test for the normal, lognormal, and Gumbel distributional hypotheses.” *Water Resources Research* 22(4):587–590.