# The Problem of Separate Hypotheses via Mixture Models

Marcelo de Souza Lauretto*,†, Silvio R. de Faria Jr.*, Basilio B. Pereira**, Carlos A. B. Pereira* and Julio M. Stern*,†

*University of Sao Paulo, Institute of Mathematics and Statistics
†lauretto@ime.usp.br jstern@ime.usp.br
**University of Rio de Janeiro, Medical School and COPPE

**Abstract.** This article describes the Full Bayesian Significance Test for the problem of separate hypotheses. Numerical experiments are performed for the Gompertz vs. Weibull life span test.

**Keywords:** FBST, Life span, Model selection, Separate hypotheses, Significance tests, Reliability.

## INTRODUCTION

An important problem in statistics inference consists of deciding which of $m$ alternative models, $f_k(x,\psi_k)$, more adequately fits a given dataset. When the candidate models $f_k$ have distinct (unrelated) functional forms, it is usual to call them "separate" models (or hypotheses). Many discriminate models have been developed, which counterpoise a (null) model $f_1(x,\psi_1)$ against one alternative model $f_2(x,\psi_2)$, providing a measure of evidence in data favoring model 1 over model 2 [1,18]. However, these methods are not capable of give a straight answer when neither candidate model individually describes well the data. Non-parametric tests (e.g. Goodness-of-fit and Kolmogorov-Smirnov), on the other hand, have a comparatively slow convergence rate.

In this article we analyze this problem in the context of mixture models, see [14]. The basic distribution of this statistical model is a weighted sum of two or more candidate pdf's. Deciding if the data comes from a specific distribution is to test if the other distributions weights equal 0. Under this formulation, if neither model describes adequately the data, the test is capable of give a direct answer – a high evidence against all candidate models. As a numerical example we use a classical problem in reliability analysis, the Gompertz vs. Weibull life span, see [11,12].

The Fully Bayesian Significance Test (FBST) is presented by Pereira and Stern [19] as a coherent Bayesian significance test. The FBST is intuitive and has a geometric characterization. In this article the parameter space, $\Theta$, is a subset of $R^n$, and the hypothesis is defined as a further restricted subset defined by vector valued inequality and equality constraints: $H:\theta\in\Theta_H$ where $\Theta_H=\{\theta\in\Theta\,|\,g(\theta)\leq 0\wedge h(\theta)=0\}$. For simplicity, we often use $H$ for $\Theta_H$. We are interested in precise hypotheses, with $\dim(H)<\dim(\Theta)$ . $f(\theta)$ is the posterior probability density function.

The computation of the evidence measure used on the FBST is performed in two steps: The optimization step consists of finding $f^*$ and $\widehat{f}$, the constrained (over H) and unconstrained maxima of the posterior. The integration step consists of integrating the

CP954, 27th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering,
edited by K. H. Knuth, A. Caticha, J. L. Center, Jr., A. Giffin, and C. C. Rodríguez

posterior density over the Tangential Set, $\overline{T}$ where the posterior is higher than anywhere in $H$, i.e., $\overline{T} = \{\theta \in \Theta : f(\theta) > f^*\}$, $f^* = \max_H f(\theta) = f(\theta^*)$, $\widehat{f} = \max_\Theta f(\theta) = f(\widehat{\theta})$, $\overline{\mathrm{Ev}}(H) = \Pr(\theta \in \overline{T} \,|\, x) = \int_{\overline{T}} f(\theta) d\theta$.

$\overline{\mathrm{Ev}}(H)$ is the evidence against $H$, and $\mathrm{Ev}(H) = 1 - \overline{\mathrm{Ev}}(H)$ is the evidence supporting (or in favor of) $H$. Intuitively, if $\overline{\mathrm{Ev}}(H)$ is "large", $\overline{T}$ is "heavy", and the hypothesis set is in a region of "low" posterior density, meaning a "strong" evidence against $H$.

Let us consider the cumulative distribution of the evidence value against the hypothesis, $\overline{V}(\tau) = \Pr(\overline{\mathrm{Ev}} \leq \tau)$, given $\theta^0$, the true value of the parameter. Under appropriate regularity conditions, for increasing sample size, $n \to \infty$, we can state the following:
- If $H$ is false, $\theta^0 \notin H$, then $\overline{\mathrm{Ev}}$ converges (in probability) to one, that is, $\overline{V}(\tau) \to \delta(1)$.
- If $H$ is true, $\theta^0 \in H$, then $\overline{V}(\tau)$, the confidence level, is approximated by the function $\overline{W}(t,h,\tau) = \mathrm{Chi2}\left(t-h, \mathrm{Chi2}^{-1}(t,c)\right)$, where $t = \dim(\Theta)$, $h = \dim(H)$ and $\mathrm{Chi2}(k,x)$ is the cumulative chi-square distribution with $k$ degrees of freedom.

Hence, for large $n$, to reject $H$ with a level of significance $\delta$, we set $\tau = \overline{W}^{-1}(t,h,1-\delta)$, i.e. set $\tau$ such that $\overline{W}(t,h,\tau) = 1-\delta$.

Several FBST applications and examples, efficient computational implementation, interpretations, and comparisons with other techniques for testing sharp hypotheses, can be found in the authors' papers in the reference list. For a FBST review see the on line document [21].

## WEIBULL AND GOMPERTZ DISTRIBUTIONS

In this paper we analyze the Gompertz vs. Weibull life span model selection problem. For the importance and interpretation of this problem see [11].

The Weibull hazard and probability density functions, for a failure time $x \geq 0$, given the shape and characteristic life (or scale) parameters, $\beta > 0, \gamma > 0$, are:

$$h_W(x \,|\, \beta,\gamma) = \beta x^{\beta-1}/\gamma^\beta \ , \ \ f_W(x \,|\, \beta,\gamma) = (\beta x^{\beta-1}/\gamma^\beta)\exp(-(x/\gamma)^\beta) \ .$$

The Gompertz hazard and probability density functions, for a failure time $x \geq 0$, given the parameters, $\alpha > 1, \lambda > 0$, are:

$$h_G(x \,|\, \alpha,\lambda) = \lambda\alpha^x \ , \ \ f_G(x \,|\, \alpha,\lambda) = \lambda\alpha^x \exp(-(\alpha^x - 1)\lambda/\log\alpha) \ .$$

The Gompertz distribution exhibits a strong nonlinear correlation between the parameters $\alpha$ and $\lambda$, see Figure 1A. This correlation explains the *compensation law of mortality*, which states that higher values for the parameter $\alpha$ are compensated by lower values of parameter $\lambda$ in different populations of a given species: $ln(\lambda) = ln(M) - B\alpha$, where $B$ and $M$ are universal species-specific invariants, see [11]. As a result, the Gompertz density in its original form is not log-concave. As we shall discuss later, we use adaptive samplers for the parameters, which depend on the shape of the density function – preferably log-concave distributions. In order to separate the parameters $\alpha$ and $\lambda$, diminishing this nonlinear dependence and enhancing the shape of density function for sampling, we adopt the reparameterization $u = 1/\log\alpha$ and $v = \log(\log\alpha)/\lambda)$, suggested by Meeker and Escobar [17], see Figure 1B.

The log-likelihoods of Weibull and (reparameterized) Gompertz models and their respective gradients (used for maximum likelihood estimation) are:

$$\begin{array}{rcl} L_W(\beta,\gamma|X) &=& n\log\beta - n\beta\log\gamma + (\beta+1)\sum_j \log x - \sum_j (x_j/\gamma)^\beta \ , \\ dL_W/d\beta &=& n/\beta - n\log\gamma + \sum_j \log x_j - \sum_j (x_j/\gamma)^\beta \log(x_j/\gamma) \ , \\ dL_W/d\gamma &=& -n\beta/\gamma + \beta/\gamma \sum_j (x_j/\gamma)^\beta \ , \\ L_G(u,v|X) &=& -n\log u - nv + \sum_j x_j/u + n/\exp(v) - \sum_j \exp(x_j/u - v) \ , \\ dL_G/du &=& -n/u - \sum_j x_j/u^2 + \sum_j x_j/u^2 \exp(x_j/u - v) \ , \\ dL_G/dv &=& -n - n/\exp(v) + \sum_j \exp(x_j/u - v) \ . \end{array}$$

## MIXTURES OF SEPARATE MODELS

Given a dataset $X = \{x_1, x_2 \dots x_n\}$ and distinct alternative probability densities, $f_1(X \mid \psi_1), f_2(X \mid \psi_1), \dots, f_m(X \mid \psi_m)$, where $\psi_k$ are (vector) parameters, the problem of interest is to measure the evidence in favour of each model for fitting the dataset. In this paper, the consider a general model including all candidate distributions, where the choice of a specific distribution is a special case. The origin of this model comes in the work of Cox [7], who suggested that, in the presence of two alternative models, the p.d.f of data could be taken proportional to

$$f(x \mid w, \psi_1, \psi_2) \propto f_1(x \mid \psi_1)^{w_1} f_2(x \mid \psi_2)^{w_2} \ , \ \ w > 0 \mid w\mathbf{1} = w_1 + w_2 = 1 \ ,$$

Then, deciding if the model 1 is adequate to describe the data is to test the hypothesis $H_1 : w_1 = 1$ against the hypothesis $w \neq 1$. Atkinson [2] developed this idea for some distributions of the exponential class, writing the density as

$$f(x \mid w, \psi_1, \psi_2) = \frac{f_1(x \mid \psi_1)^{w_1} f_2(x \mid \psi_2)^{w_2}}{\int f_1(y \mid \psi_1)^{w_1} f_2(y \mid \psi_2)^{w_2} dy} \ .$$

In this paper, we consider that the p.d.f. of data is a convex linear combination of the fixed candidate densities: denoting $\theta = [w, \psi_1, \dots \psi_m]$,

$$f(x \mid \theta) = w_1 f_1(x \mid \psi_1) + \dots + w_m f_m(x \mid \psi_m) \ , \ \ w \geq 0 \mid w\mathbf{1} = 1 \ .$$

The likelihood then is

$$f(X \mid \theta) = \prod_{j=1}^{n} \sum_{k=1}^{m} w_k f_k(x_j \mid \psi_k) \ .$$

Here it is important to remember some key concepts of mixture models. In mixture analysis for unsupervised classification, we assume that the data come from one or more subpopulations (classes), distributed under distinct densities. The evidence in favor of the existence of more than one subpopulation will be higher if some subsets of data are more adequately fitted by a particular component of the mixture, where other subsets are better fitted by another components. In order to detect this situation, the mixture model must be able to infer the (probability of) data classifications. The real classifications are considered non observable and, for this reason, called *hidden* or *latent* variables. The problem of deciding if one single candidate distribution fits adequately the data is analogous to decide the number of components in a traditional mixture model, and the behavior of the system will be also similar: if the candidate model does not fit well the data, some observed points may be better described by a particular component of mixture, where the remaining will be better fitted by other components.

A sample $j$ of class $k = c(j)$ is distributed with density $f_k(x_j \mid \psi_k)$. The boolean classification matrix $Z$ indicates whether or not $x_j$ is of class $k$, i.e. $z_j^k = 1$ iff $c(j) = k$. Conditioning on the latent variables we can rewrite:

$$\begin{aligned} f(x_j \mid \theta) &= \sum_{k=1}^m f_k(x_j \mid \theta, z_k^j) f(z_j^k \mid \theta) = \sum_{k=1}^m w_k f_k(x_j \mid \psi_k) \ , \\ f(X \mid \theta) &= \prod_{j=1}^n f(x_j \mid \theta) = \prod_{j=1}^n \sum_{k=1}^m w_k f_k(x_j \mid \psi_k) \ . \end{aligned}$$

Given the mixture parameters, $\theta$, and the observed data, $X$, the conditional classification probability matrix, $P = f(Z \mid X, \theta)$, is given by:

$$p_k^j = f(z_j^k \mid x_j, \theta) = \frac{f_k(z_j^k, x_j \mid \theta)}{f(x_j \mid \theta)} = \frac{w_k f_k(x_j \mid \psi_k)}{\sum_{k=1}^m w_k f(x_j \mid \psi_k)} \ .$$

We use $y_k$ for the number of samples of class $k$, i.e. $y_k = \sum_j z_j^k$, or $y = Z\mathbf{1}$.

The density for the "completed" data, $X, Z$, is:

$$f(X, Z \mid \theta) = \prod_{j=1}^n f_{\psi_{c(j)}}(x_j \mid \psi_{c(j)}) f(z_j^k \mid \theta) = \prod_{k=1}^m \left( {w_k}^{y_k} \prod_{j \mid c(j)=k} f_k(x_j \mid \psi_k) \right) \ .$$

In the remaining of this section we discuss the FBST formulation for the Weibull vs. Gompertz mixture model. The conjugate prior for a multinomial distribution is a Dirichlet distribution:

$$\begin{aligned} M(y \mid n, w) &= n! / (y_1! \ldots y_m!) \ {w_1}^{y_1} \ldots {w_m}^{y_m} \ , \\ D(w \mid y) &= \Gamma(y_1 + \ldots + y_k) / (\Gamma(y_1) \ldots \Gamma(y_k)) \prod_{k=1}^m {w_k}^{y_k - 1} \ , \end{aligned}$$

with $w > \mathbf{0}$ and $w\mathbf{1} = 1$. Prior information given by $\dot{y}$, and observation $y$, result in the posterior parameter $\ddot{y} = \dot{y} + y$. Here we take the non-informative prior given by $\dot{y} = \mathbf{1}$. We also consider a improper uniform prior for $(\beta, \gamma, u, v)$. Therefore, the posteriori is

$$\begin{aligned} f(\theta \mid X) &\propto f(X \mid \theta) = \prod_{j=1}^n \left( p_j^1 w_1 f_W(x_j \mid \beta, \gamma) + p_j^2 w_2 f_G(x_j \mid \alpha, \lambda) \right) \ , \\ p_j^1 &= \frac{w_1 f_W(x_j \mid \beta, \gamma)}{w_1 f_W(x_j \mid \beta, \gamma) + w_2 f_G(x_j \mid \alpha, \lambda)} \ , \ p_j^2 = 1 - p_j^1 \ . \end{aligned}$$

The hypotheses of interest are $H_1 : w_1 = 1 \wedge w_2 = 0$ and $H_2 : w_1 = 0 \wedge w_2 = 1$. The FBST procedure for testing $H_k, k = 1, 2$ consists of two steps:

- Estimate the maximum of the log-likelihood $L_k^*$ under $H_k$, which corresponds to the maximum log-likelihood under the corresponding single component distribution.

- Estimate the e-value supporting the hypothesis $H_k$, that is, the ratio

$$\mathrm{Ev}(H_k) = \frac{\int_{T_k} f(\theta \,|\, X) d\theta}{\int_{\Theta} f(\theta \,|\, X) d\theta} \; , \; T_k = \{\theta \in \Theta \,|\, L(\theta) \le L_k^*\} \; .$$

Notice that since the likelihood normalization constant is the same for both numerator and denominator, so it is cancelled and can therefore be ignored in the computational procedure. For the optimization step, we used the `Algencan-Tango` solver, which source code and detailed description are freely distributed (see internet link at the reference), see [4,5].

In order to perform the integration over the posterior measure, we used a Gibbs sampling Markov Chain Monte Carlo algorithm, MCMC. Given the current vector parameter $\theta^i$, we compute $P$. Given $P$, we draw $Z$ from $f(z_j \,|\, p_j)$, a simple multinomial distribution. Given the latent variables, $Z$, we separate the samples of classes 1 and 2. In the Weibull component, we draw a parameter value $[\beta^{i+1}, \gamma^{i+1}]$ with density proportional to the partial likelihood $\prod_{j|c(j)=1} fw(x_j \,|\, \beta, \gamma)$. The same idea is applied to draw the Gompertz parameters $[\alpha^{i+1}, \lambda^{i+1}]$. Given $\ddot{y} = Z\mathbf{1} + \dot{y}$, we can draw a new weight vector $[w_1^{i+1}, w_2^{i+1}]$ using a Dirichlet distribution $D(w \,|\, \ddot{y}_1, \ddot{y}_2)$. At the end of iteration $(i)$, we have a new vector parameter $\theta^{i+1} = [w_1^{i+1}, w_2^{i+1}, \beta^{i+1}, \gamma^{i+1}, \alpha^{i+1}, \lambda^{i+1}]$, and can begin iteration $(i+1)$.

We do not know a direct method to draw the parameters from the Weibull or Gompertz likelihood. For this purpose we used the adaptive sampler `HITRO`, see [13,20,22]. `HITRO` combines the multivariate Ratio-of-Uniforms method with the Hit-and-Run sampler. The Ratio-of-Uniforms transformation maps the region below the p.d.f $f$, i.e. $G(f) = \{(x, y) : 0 < y < f(x)\}$ into the region

$$A(f) = A_{r,m}(f) = \left\{ (u, v) : 0 < v < f\left(\frac{u}{v^r} + m\right)^{1/(rn+1)} \right\}$$

by means of the transformation

$$(u, v) \mapsto (x, \, y) = \left(\frac{u}{v^r} + m, \; v^{rn+1}\right) \; .$$

The vector $m$ must be a point near the mode (in our implementation, we set $m$ as the mode). The method relies on the theorem that, if $(u, v)$ is uniformly distributed over $A(f)$, then $x = u/v^r + m$ has probability density function $f(x)/\int f(z)dz$. The Hit-and-run sampler is used for generating points $(u, v)$ uniformly over $A(f)$.

## NUMERICAL EXPERIMENTS AND FINAL REMARKS

We run some numerical experiments in order to evaluate the FBST performance on our problem of separate models. The experiments were based on the IBGE data bank for the mortality of Brazilian male population in the year of 2005, available on line at *http://www.ibge.gov.br /home/estatistica/populacao/tabuadevida/2005/default.shtm.* We used the mortality rate table from ages 5 to 80, hence avoiding the early infancy or burn-in period, see [3,17].

The experiments were based on simulated data, drawn from four distributions, the parameters of which have always been chosen to provide the best fit to the IBGE data bank. The distributions fitted were: (1)-Weibull, (2)-Gompertz, (3)-Gamma, and (4)-Beta (rescaled), see Figure 2. Our main interest was to measure the convergence rate of correct decisions, concerning the acceptance / rejection of the Weibull vs. Gompertz hypotheses, when using the FBST on the mixture model. Of course, in cases (1) and (2) we want to accept the correct hypothesis and reject the false one, whereas in cases (3) and (4) we want to reject them both.

As acceptance / rejection threshold, we adopted the critical level $\tau$ according to criterion presented in section 1, with a significance level of 5%. Since the mixture model and the restricted model have 5 and 2 degrees of freedom, respectively, we have $\tau = \overline{W}^{-1}(5,2,0.95) = 0.83$. Therefore, we reject $H$ if $\overline{\mathrm{Ev}}(H) > 0.83$, or equivalently if $\mathrm{Ev}(H) < 0.17$. Using each of the four fitted distributions we generated 500 samples of size $n =$ 30, 50, 75, 100, 150, 200, 300, 400 and 500.

We have compared the performance of the FBST with the Kolmogorov–Smirnov (KS) test, [9]. In this test, the goodness of fit measure is taken to be the Kolmogorov distance $D_n^* = D(F_n, F^*) = sup_x |F_n(x) - F^*(x \mid \theta)|$, where $F_n$ denotes the sample (empirical) distribution and $F^*$ denotes the theoretical distribution to be tested. Due to difficulty in estimate $\theta$ which minimizes $D(F_n, F^*)$, it is usually adopted the maximum likelihood estimator for $\theta$. Kolmogorov and Smirnov demonstrated in 1930's that, if the null hypothesis $F(X) = F^*(X \mid \theta)$, then $lim_{n\to\infty} Pr(\sqrt{n} D_n^* \le t) = 1 - 2\sum_{i=1}^{\infty}(-1)^{i-1}\exp(-2i^2t^2)$. The distribution at the right side of this equation allows one to compute the significance (p-value) of $D_n^*$. For a meaningful comparison, we also used a 5% significance level.

The hole batch of 500 simulations for each of the 4 cases and 9 sample sizes, took about 2 day of computation on a Intel Pentium server, or about 10 seconds per test. Computing time was dominated by *Hitro*, a flexible and robust but generic subroutine. Hence, its substitution by a tailor made and more efficient sampler could enhance the program computational performance.

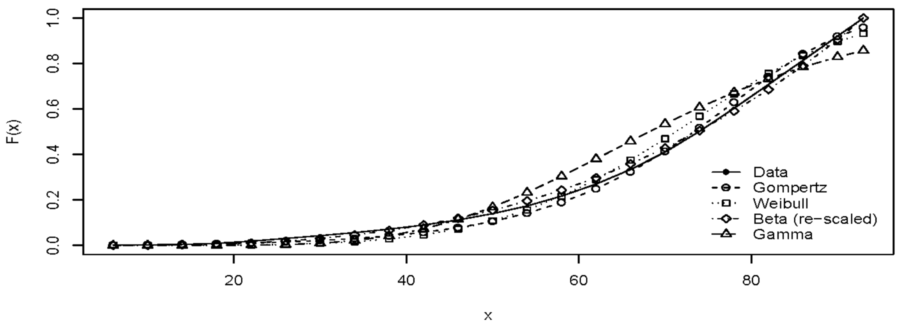
Figure 3 summarized the correct decision rates in the numerical simulations. The Weibull distribution can approximate very well a Gamma distribution. This explains the relatively slow convergence in the decision to reject the Weibull hypothesis in the simulations from the Gamma.

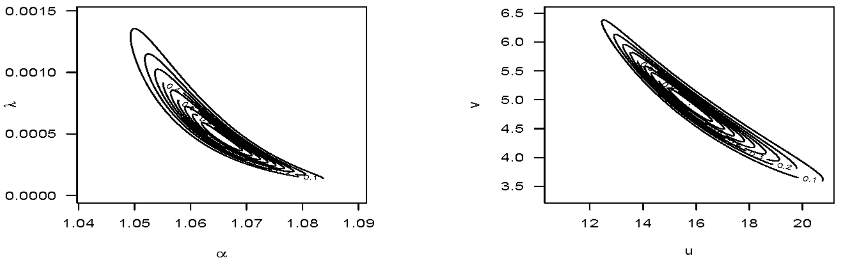
As expected, the FBST had a good performance. Moreover its implementation is straightforward, following the guidelines presented in [19,21]. It would be interesting to replace the Kolmogorov-Smirnov benchmark with a parametric alternative, like some form of jump MCMC. However, as far as the authors know, none is available at this time. The authors intend to collaborate with other research groups in order to develop and implement such algorithms.

Acknowledgments: The authors are grateful for the support of CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico, and FAPESP - Fundação de Amparo à Pesquisa do Estado de São Paulo.

# REFERENCES

1. M.A.Araujo, B.B.Pereira (2007). A Comparison of Bayes Factors for Separated Models: Some Simulation Results. *Communications in Statistics-Simulation and Computation* 36(2), 297–309.
2. A.C.Atkinson (1970). A Method for Discriminating Between Models. *J. R. Statist Soc. B* 32, 323–354.
3. R.E.Barlow, F.Prochan (1981). *Statistical Theory of Reliability and Life Testing Probability Models.* Silver Spring: To Begin With.
4. E.G.Birgin, J.M.Martnez and M.Raydan (2000). Nonmonotone Spectral Projected Gradient Methods on Convex Sets. *SIAM Journal on Optimization*, 10, 1196-1211. Software and documentation available at *www.ime.usp.br/∼egbirgin/tango/*.
5. E.G.Birgin , J.M.Martnez (2002). Large-scale Active-Set Box-Constrained Optimization Method with Spectral Projected Gradients. *Computational Optimization and Applications*, 23, 101-125.
6. W.Borges, J.M.Stern (2006). Evidence and Compositionality. p.307-315 in J.Lawry et al., *Soft Methods for Integrated Uncertainty Modelling.* NY: Springer.
7. D.R.Cox (1962). Further Results on Tests of Separated Families of Hypotheses. *J. R. Statist Soc. B* 24, 406-423.
8. D.R.Cox and D.Oakes (1984). *Analysis of Survival Data.* Monographs on Statistics and Applied Probability, Chapman & Hall.
9. M.H.DeGroot (1986). *Probability and Statistics.* Addison–Wesley.
10. B.Dodson (1994). *Weibull Analysis.* Milwaukee: ASQC Quality Press.
11. L.A.Gavrilov and N.S.Gavrilova (1991). *The Biology of Life Span: A Quantitative Approach.* New York: Harwood Academic Publisher.
12. L.A.Gavrilov and N.S.Gavrilova (2001). The Reliability Theory of Aging and Longevity. *J. Theor. Biol.* 213, 527–545.
13. R.Karawatzki, J.Leydold, K.Pötzelberger (2005). Automatic Markov Chain Monte Carlo Procedures for Sampling from Multivariate Distributions. Department of Statistics and Mathematics Wirtschaftsuniversität Wien Research Report Series. Report 27, December 2005. Software available at *http://statistik.wu-wien.ac.at/arvag/software.html.*
14. M.S.Lauretto, J.M.Stern (2005). FBST for Mixture Model Selection. Maxent'2005, *AIP Conf. Proc.* 803, 121–128.
15. M.R.Madruga, L.G.Esteves, S.Wechsler (2001). On the Bayesianity of Pereira-Stern Tests. *Test*, 10, 291–299.
16. M.R.Madruga, C.A.B.Pereira, J.M.Stern (2003). Bayesian Evidence Test for Precise Hypotheses. *Journal of Statistical Planning and Inference,* 117, 185–198.
17. W.Q.Meeker and L.A.Escobar (1998). Statistical Methods for Reliability Data. Wiley Series in Probability and Statistics.
18. B.B.Pereira (2005). Separate Families of Hypotheses. in P.Armitage, T.Colton (eds.) *Encyclopedia of Biostatistics* (2nd.ed.). 7, 4881-4886. NY: Wiley.
19. C.A.B.Pereira, J.M.Stern, (1999). Evidence and Credibility: Full Bayesian Significance Test for Precise Hypotheses. *Entropy Journal*, 1, 69–80.
20. R.L.Smith (1984). Efficient Monte Carlo Procedures for Generating Points Uniformly Distributed over Bounded Regions. *Operations Research 32*, 1296–1308.
21. J.M.Stern (2007). Cognitive Constructivism, Eigen-Solutions, and Sharp Statistical Hypotheses. *Cybernetics and Human Knowing,* 14,1, 9-36.
24. I.Vaduva (1984). Computer Generation of Random Vectors Based on Transformation of Uniformly Distributed Vectors. In *Probability Theory, Proc. 7th Conf.*, Brasov/Rom, 589–598.
25. J.C.Wakefield, A.E.Gelfand, A.F.Smith (1991). Efficient Generation of Random Variates via Ratio-of-Uniforms Method. *Statist. Comput.* 1,2, 129-133.





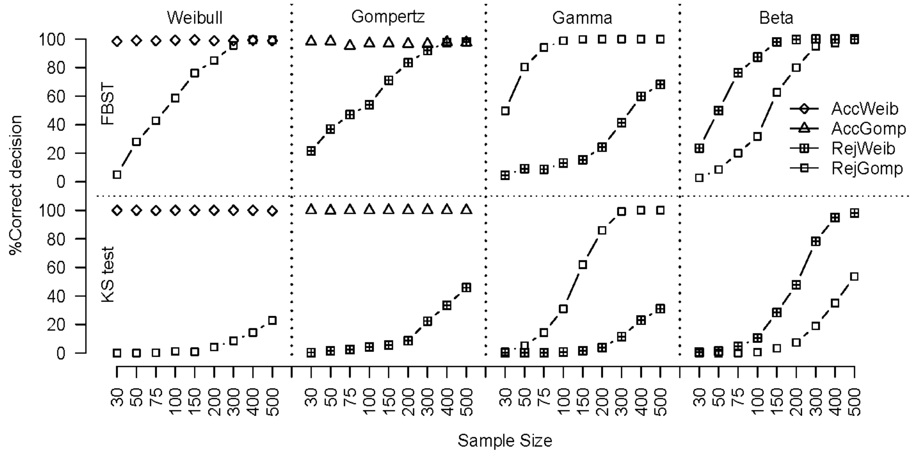


**Figure 1:** IBGE Brazilian mortality rates and fitted distributions.
**Figure 2A,B:** Contour plots for Gompertz density and reparameterization.
**Figure 3:** Correct decision rates on numerical simulations.