Univariate Parametric Survival Analysis using GS-distributions

A. Sorribas∗, J. M. Muiño, M. Rué and J. Fibla

Universitat de Lleida

Abstract

The GS-distribution is a family of distributions that provide an accurate representation of any unimodal univariate continuous distribution. In this contribution we explore the utility of this family as a general model in survival analysis. We show that the survival function based on the GS-distribution is able to provide a model for univariate survival data and that appropriate estimates can be obtained. We develop some hypotheses tests that can be used for checking the underlying survival model and for comparing the survival of different groups.

1 Introduction

Survival analysis refers to the analysis of time-to-event processes when a censoring mechanism prevents the observation of the precise time at what the event occurs in some of the individuals. A typical situation arises in medical studies where the goal is to estimate the time that a given patient will survive after a treatment and to evaluate the different factors that can influence this process. Censoring appears when a patient is lost to follow-up or when the event occurs between two observation times. In all those cases, estimation of the survival function $P(T > t) = S(t)$ can be obtained by constructing the appropriate likelihood. For example, in a right-censoring scheme some events are not

observed due to loss of follow-up. In that case the likelihood can be constructed as (see for instance Hosmer & Lemeshow, 1999; and Klein & Moeschberger, 1997)

$$L \propto \prod_{i \in U} f(x_i) \prod_{i \in R} S(t_i) \quad i = 1, \ldots, n \tag{1}$$

where U indicates the set of uncensored event times and R the set of right censored times. More complicated censoring schemes can be accommodated by considering the contribution of each observation to the likelihood. If a parametric model exists for the survival process, the analysis is straightforward and reduces to estimating the corresponding parameters. However, in most cases such a model is unknown making it necessary to consider non-parametric approaches. For example, the Kaplan-Meier estimator provides a good estimation of $S(t)$ for right-censored data (see for instance Klein & Moeschberger, 1997). Methods for estimating $S(t)$ for interval-censored data and other censoring schemes are still under investigation (Zhao & Sun, 2004).

If we could identify a sufficiently general family of distributions that can be used as a general model for survival analysis, we could use it as an alternative to the non-parametric approach for any censoring scheme. In this work, we introduce the GS-distribution as a family of distributions that can be used for this purpose in univariate unimodal continuous survival processes. We shall briefly recall the GS-distribution family of distributions and then we will introduce its use in survival analysis.

1.1 GS-distribution

The GS-distribution is defined as (Muiño *et al.* (2006)):

$$\frac{dF}{dx} = \alpha F^g (1 - F^k)^\gamma \quad F(x_0) = F_0 \tag{2}$$

Here, $F = P(X \leq x)$ and α, g, k, γ are real non-negative parameters. F_0 is the value of the cumulative distribution at x_0. Without lack of generality, we can take $F_0 = 0.5$. In such a case, x_0 is the median. We shall refer to this distribution as GSD$[F_0, x_0, \alpha, g, k, \gamma]$. This family of distributions provides a useful model for accurately represent univariant unimodal continuous distributions. It can be shown that some distributions are exactly represented by the GS-distribution. For instance, the logistic case corresponds to $g = \gamma$ and $k = 1$. The uniform, the exponential, and some Beta and **F** distributions are also exact cases. All symmetric distributions correspond to the special cases $g = \gamma$ and $k = 1$ (see Muiño *et al.* (2006)). Other distributions are accurately approximated by a GS-distribution, although they do not correspond to an exact case (see Muiño *et al.* (2006) for a complete discussion and examples). This family generalizes the S-distribution which corresponds to the case $\gamma = 1$ (Voit 1992, 2000; Voit & Schwacke, 1998; Voit & Sorribas, 2000; Voit *et al.* 1995). Other families based on quantiles are also included.

For instance, the case $k = 1$ corresponds to the Q-distribution family (see Turner & Pruitt, 1978; Parzen, 1979; Kamps, 1991; Jones 2002, 2004).

As the GS-distribution is defined as a differential equation, computation of its properties and its use for data modeling requires appropriate algorithms (Voit 1992; Voit & Schwake, 2000; Sorribas *et al.* 2000, 2002; Hernández-Bermejo & Sorribas, 2001; March *et al.* 2002). Due to the flexibility of this family, it can be used as a general model that can appropriately fit observed data. As the GS-distribution can approximate almost any univariate unimodal continuous distribution, fitting a GS-distribution is equivalent to selecting the best distribution that can describe the data. As the GS-distribution is quite general, if necessary we can introduce some constraints that limit the fit. For example, if we want to consider only symmetric distributions, we can fit a GS-distribution with $g = \gamma$ and $k = 1$.

In this work we consider using this family as a parametric family for survival analysis (see also Yu & Voit, 1995). In the next section, we shall introduce this model in the context of survival analysis and show its utility with uncensored data. Then we will move to comparing survival curves using this model and to more complex censoring schemes.

2 Survival model based on GS-distributions

The survival function is related to the distribution function as:

$$S(t) = 1 - F(t) \qquad (3)$$
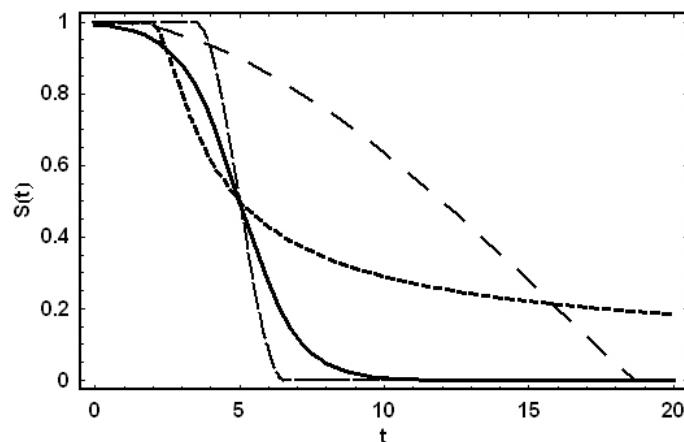
Figure 1: *Survival functions based on GS-distributions: GSD[0.5, 5.0, 1.0, 1.0, 1.0, 1.0] (continuous line), GSD[0.5, 5.0, 1.0, 0.5, 1.0, 3.0] (short dashed line), GSD[0.5, 5.0, 1.0, 0.5, 1.0, 0.5] (medium dashed line), GSD[0.5, 12.0, 0.1, 0.4, 1.0, 0.1] (long dashed line).*

According to this definition, the survival function $S(t)$ based in a GS-distribution is the solution of the differential equation:

$$\frac{dS}{dt} = -\frac{dF}{dt} = -\alpha F^g(1 - F^k)^\gamma \quad S(t_0) = 1 - F(t_0) \tag{4}$$

Some examples of GS-distribution based survival functions are shown in Figure 1.

Using this model, the hazard function can be written as:

$$h(t) = \frac{f(t)}{S(t)} = \alpha F^g(1 - F^k)^\gamma(1 - F)^{-1} \tag{5}$$

The survival function and the corresponding hazard function, for a set of parameters $(F_0, t_0, \alpha, g, k, \gamma)$, will correspond to the solution of the following set of equations:

$$\frac{dh}{dt} = \frac{F^{2g-1}(1-F^k)^{2\gamma-1}\alpha^2(F(1-F^k)+(1-F)((g+k\gamma)F^k+g))}{(1-F)^2} \quad h(t_0) = f(t_0)/S(t_0)$$

$$\frac{dS}{dt} = -\alpha F^g(1 - F^k)^\gamma \qquad S(t_0) = 1 - F(t_0)$$

$$\frac{dF}{dt} = \alpha F^g(1 - F^k)^\gamma \qquad F(t_0) = F_0 \tag{6}$$

The hazard functions corresponding to the survival functions in Figure 1 are shown in Figure 2.
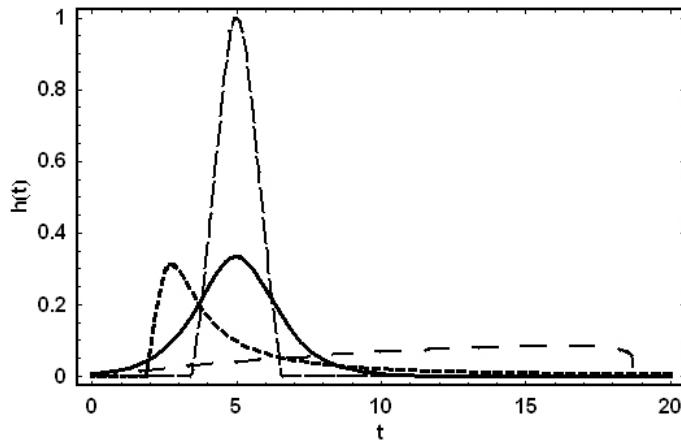
Figure 2: *Hazard functions corresponding to the survival functions in Figure 1. Lines follow the same scheme as in that figure.*
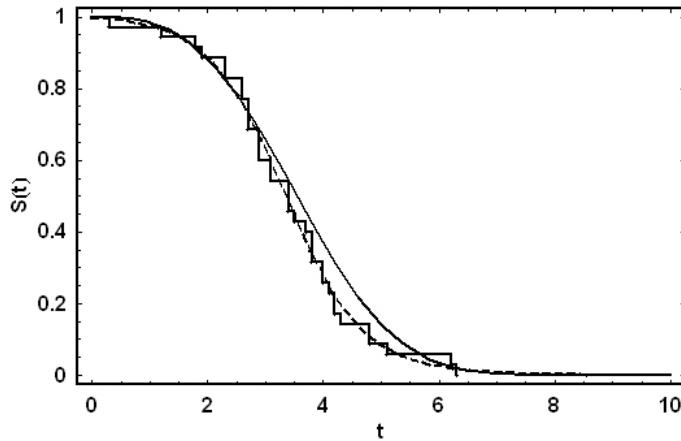
***Figure* 3**: *Fitting a GS-distribution to uncensored survival data. Data is generated from a Weibull model with shape parameter 3 and scale parameter 4. The original model corresponds to the continuous line. The fitted GS-distribution model is indicated with a dashed line. The step function corresponds to the Kaplan-Meier estimator.*

2.1 Fitting the GS-survival function

The structure of the observations is a fundamental issue for parameter estimation in survival analysis. As censoring and truncation impose special conditions, we shall begin with uncensored data to discuss the basics of maximum-likelihood estimation using the GS-distribution survival model. Let us suppose that we have a set of observations $t_1, t_2, \ldots, t_n$ where t_i indicates the survival time for a given individual, i.e. the exact time at which the event of interest is observed. For this situation, the likelihood function is simply:

$$L(0.5, t_0, \alpha, g, k, \gamma) \propto \prod_{i=1}^{n} f(t_i) = \prod_{i=1}^{n} \alpha F(t_i)^g (1 - F(t_i)^k)^\gamma \qquad (7)$$

To compute the likelihood for a given set of parameters, we shall follow a numerical procedure. First, we compute the $F(t_i)$ for a given survival time t_i. This requires integration of the GS-distribution equation from t_0 to t_i. Then, once all the $F(t_i)$ values are obtained, we calculate the likelihood using equation (7). Maximum-likelihood estimation will be performed by numerically finding the set of parameters that maximize equation (7). This requires an iterative computation of this equation until a maximum is reached. Although we have implemented this procedure in Mathematica, any alternative program could be used. At this point, we are developing **R** routines for using this model.

As an example, we generate survival times from a Weibull model and fit the data using the procedure discussed above (Figure 3). The corresponding Kaplan-Meier

estimation of the survival curve is also computed for comparison. It can be seen that the model provides an appropriate representation for the data. One advantage is that we obtain a parametric model that can be used for other purposes. For instance, using the quantile function for a GS-distribution (Muiño *et al.*, 2006) we can generate random survival times.

2.2 Using the GS-distribution as a model for survival data generation

The GS-distribution can be used as a model for generating survival data. First, we recall that the quantile x_q that is the solution of the equation $F(x_q) = q$ for a GS-distribution can be computed as (Muiño *et al.* 2006):

$$x_q = x_0 + \frac{B_{t_0^k, q^k}\left(\frac{1-g}{k}, 1-\gamma\right)}{\alpha\, k} \tag{8}$$

where B is the incomplete Beta function. According to this result, we can use this expression for obtaining a sample of values of the GS-distribution, which would correspond to a sample of survival times for the corresponding model. First, we generate a sample of a uniform distribution between 0 and 1. Then, we use those numbers as values of q in (refquantile) to obtain a sample of the required GS-distribution. This may be useful in testing some procedures in survival analysis. As an example, we generate a random sample of survival data from a GSD[0.5, 10, .3, .6, 1, 1] ($n = 50$) (Figure 4). Censored data can be obtained using a similar procedure and taking into account the corresponding censoring scheme.
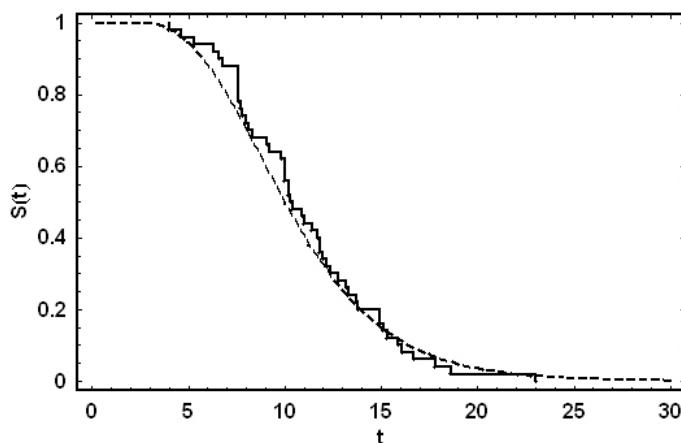
Figure 4: *Uncensored random sample from a survival model. The original model (dashed line) corresponds to a GSD[0.5, 10.0, 0.3, 0.6, 1.0, 1.0]. The step function corresponds to the Kaplan-Meier estimator from the generated data.*

### 2.3 Fitting S-survival functions on different censoring schemes

As the GS-distribution provides a survival model that is general for the unimodal continuous case, we can obtain the likelihood for any censoring scheme. According to the discussion above, computation of the likelihood requires a numerical procedure in all cases.

2.3.1 Right censoring

Suppose we have uncensored (U) observations and right-censored (R) observations. For the R cases, the contribution to the likelihood, in the general case of non-informative censoring, corresponds to terms $P(T \geq t_i) = S(t_i) = 1 - F(t_i)$. Thus, the likelihood is:

$$L(0.5, t_0, \alpha, g, k, \gamma) \propto \left(\prod_{i \in U} f(t_i) \right) \left(\prod_{i \in R} S(t_i) \right) = \left(\prod_{i \in U} f(t_i) \right) \left(\prod_{i \in R} (1 - F(t_i)) \right) \qquad (9)$$

This computation can be made with a slight modification of the procedure for the uncensored case. Here, for the censored observations, we just need to compute $F(t_i)$ as indicated in the previous section. As an example, we generate data from a Weibull model. We produce censored observations and fit a GS-distribution model and obtain the Kaplan-Meier estimator. As we can see in Figure 5 the GS-distribution produces an appropriate estimation also in this case.
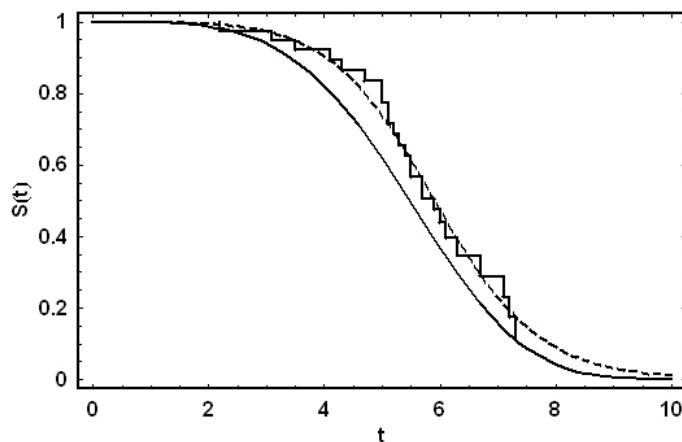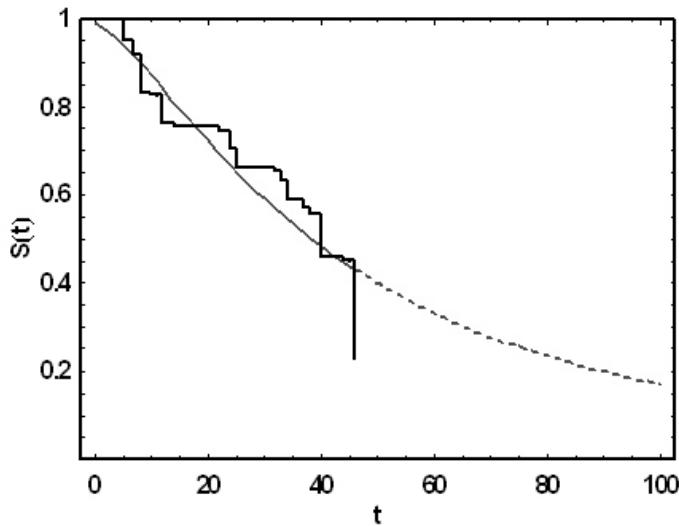

Figure 5: *Fitting a GS-distribution survival model to right-censored data. The original model is indicated by a continuous line. The fitted GS-distribution model is indicated by a dashed line. The step line corresponds to the Kaplan-Meier estimator.*

2.3.2 Interval censoring

A given observation is interval censored if it is known to have happened between two time points L_i and R_i. In that case, the contribution to the likelihood is $P(L_i \leq T \leq R_i) = F(R_i) - F(L_i) = S(L_i) - S(R_i)$. If we indicate by I the set of interval censored observations, the likelihood can be written as:

$$L(0.5, t_0, \alpha, g, k, \gamma) \propto \left(\prod_U f(t_i) \right) \left(\prod_{i \in R} S(t_i) \right) \left(\prod_{i \in I} (S(L_i) - S(R_i)) \right) \qquad (10)$$

Figure 6: *Fitting a GS-distribution survival model to interval-censored data corresponding to the example 5.2 in Klein and Moeschberger (1997). The fitted GS-distribution model is indicated by a continuous and a dashed line. The step line corresponds to the estimation provided by Klein and Moeschberger (1997). See text for discussion.*

As in previous cases, the required values of $S(t)$ can be computed numerically using the GS-distribution to obtain the likelihood. Maximization of L would provide the best GS-distribution survival model according to the data. Following this procedure, we can accommodate other censoring schemes. As an example of this procedure, we have used a set of data discussed by Klein & Moeschberger (1997, example 5.2). These data refers to breast cancer patients that follow a radiotherapy treatment. The time to cosmetic deterioration (in months) is collected. For many patients, deterioration appears between two visits. For instance, and given patient present cosmetic deterioration in its visit corresponding to 11 months after therapy. In the previous visit at 4 months no deterioration was observed. Thus, for this patient the event has occurred within (4, 11] months. This kind of data corresponds to an interval-censoring scheme. In Figure 6 we show the results of fitting a GS-distribution survival model. For comparison, the

estimation provided by Klein and Moeschberger is also provided. The GS-distribution procedure provides an estimation that allows extrapolating the results to times above the last observed time. Although this prediction must be considered with care, it seems reasonable according to the available data.

3 Hypothesis tests using GS-distribution survival models

3.1 Testing a given survival model

As the GS-distribution can accurately approximate unimodal continuous distributions, we can use it for defining a test that evaluates if the observed data follows a given model. This test can be defined as follows. First, a GS-distribution representation is obtained for the model to test (see Muiño *et al.* (2006) for a discussion on obtaining appropriate GS-distributions for a given statistical distribution). This model will be considered as the model under H_0. Second, we fit a unconstrained GS-distribution to the data. This is considered the estimated survival model under H_1. Then, a likelihood ratio test is applied to evaluate the hypothesis that the reference model is appropriate for the observed data. As an example, consider that we want to test if our data follow a Weibull model with shape parameter 4 and scale parameter 3. This distribution corresponds to a GS-distribution with parameters GSD[0.5, 2.74, 2.66, 0.92, 0.38, 0.7]. In that case, the corresponding hypotheses are:

$$H_0 \quad : \quad S(t) \text{ follows a model GSD[0.5, 2.74, 2.66, 0.92, 0.38, 0.7]}$$
$$H_1 \quad : \quad S(t) \text{ follows a model GSD[0.5,x}_0\text{,}\alpha\text{,g,k,}\gamma]$$

To test these hypotheses, we compute the resulting likelihood $L(\omega)$ under H_0 using a GSD[0.5, 2.74, 2.66, 0.92, 0.38, 0.7] as a model. Then we fit a GSD[0.5,x$_0$,α,g,k,γ] and obtain the likelihood under H_1, i.e. $L(\Omega)$. The likelihood ratio test is:

$$\Lambda = \frac{L(\omega)}{L(\Omega)} \tag{11}$$

The statistic $U = -2ln(\Lambda)$ follows a χ^2 distribution with 5 degrees of freedom. As an example, we generate a random sample of 35 survival times from a Weibull distribution with shape parameter 4 and scale parameter 3. The obtained sample is:

3.9	3.7	3.3	2.2	1.1	2.6	2.3
2.6	2.2	3.7	3.8	2.3	2.4	2.5
2.9	3.1	2.9	2.5	3.8	2.2	1.0
4.3	2.3	2.3	1.6	1.6	2.8	1.4
1.5	3.1	2.3	2.9	3.4	3.3	2.4

The log likelihood corresponding to a GSD([0.5, 2.74, 2.66, 0.92, 0.38, 0.7] is equal to -42.1479. Fitting an unconstrained GSD$[0.5,x_0,\alpha,g,k,\gamma]$ gives a log likelihood of -41.23. With that, $U = -2ln(\Lambda) = 1.84$. With 5 degrees of freedom the significance of this result is 0.643, which leads to not reject H_0. Other hypotheses can be tested following the same procedure.

For example, we can test some specific restrictions on the shape of the survival function. For instance, we can define a test for evaluating if the data follows a symmetric survival model. In that case, if we recall that a symmetric GS-distribution corresponds to the case $g = \gamma$ and $k = 1$, we have

$$
\begin{aligned}
H_0 &: \quad S(t) \text{ follows a model GSD}[0.5,x_0,\alpha,g,1,g] \\
H_1 &: \quad S(t) \text{ follows a model GSD}[0.5,x_0,\alpha,g,k,\gamma]
\end{aligned}
\tag{12}
$$

In that case, the resulting statistic $U = -2ln(\Lambda)$ follows a χ^2 distribution with 2 degrees of freedom.

3.2 Comparing GS-survival curves

Since we have a parametric model for the survival function, we can compare two groups using the likelihood-ratio test. In general, we want to test if the survival function in one group $S^a(t)$ is different than the survival function in another $S^b(t)$. This can be indicated as:

$$
\begin{aligned}
H_0 &: S^a(t) &=& S^b(t) \quad \forall t \\
H_1 &: S^a(t) &\neq& S^b(t) \quad \text{for some } t
\end{aligned}
\tag{13}
$$

Under H_0 the likelihood $L(\omega)$ can be obtained by fitting a GS-distribution model with common parameters for both groups:

$$
L(\omega) = L^a_{0.5,t_0,\alpha,g,k,\gamma} L^b_{0.5,t_0,\alpha,g,k,\gamma}
\tag{14}
$$

Under H_1 the likelihood $L(\Omega)$ can be obtained by fitting a GS-distribution to each of both groups:

$$
L(\Omega) = L^a_{0.5,t_{0_a},\alpha_a,g_a,k_a,\gamma_a} L^b_{t_{0_b},\alpha_b,g_b,k_b,\gamma_b}
\tag{15}
$$

In that case, we estimate 5 parameters under H_0 and 10 parameters under H_1. Accordingly, the statistic $U = -2ln(\Lambda)$ follows a χ^2 distribution with 5 degrees of freedom. This test is an alternative to the log-rank test and can be used with any censoring scheme for which an estimation of the likelihood can be obtained.

It is important to indicate that this test can be used independently of the censoring scheme. In the previous sections we have shown that the likelihood can be numerically obtained for any of such schemes, which provides the required information for hypothesis testing through the likelihood ratio test. Although more research is required to test the performance of this procedure, our preliminary results indicate that hypothesis testing based on GS-distributions may be a practical alternative to existing non-parametric methods.

3.3 An application to clinical data

Progression rate to the acquired immunodeficiency syndrome (AIDS) in HIV-1 infected patients is caused by a combination of environmental and genetic risk factors. The most significant association for host risk factor has been attributed to the C-C chemokine (CCR) receptor 5 (CCR5) gene variant Δ32. Homozygous individuals for a 32bp deletion in the CCR5 coding region (*CCR*5Δ32, allele) do not synthesize CCR5 protein. Those individuals are resistant to infection by HIV-1 macrophage tropic strains that use CCR5 as coreceptor. In addition, a slow progression rate to AIDS has been attributed to heterozygous patients (Huang, 1996). We have studied the role of *CCR*5Δ32 genotype status on disease progression rate to AIDS in a seroprevalent cohort of HIV-1 infected patients belonging to the injection drug use (IDU) risk group.

Given the natural history of HIV infection, HIV-1 seropositive patients could have been infected a few months to several years before their seropositivity was recorded. To reduce the lack of precision on seroconversion date, only patients with a first HIV positive test prior to 31 December 1989 were selected (137 patients). In addition, patients with an age lower than 20 years at the first HIV positive test were also included (42 patients). For patients infected prior to 31 December 1989, seroconversion date was estimated as the mean time between the date at the first HIV positive test and 1 January 1981, which correspond to the earliest year for the first AIDS case reported in Catalonia (Vilaseca, 1982). In addition, for patients with age lower than 20 years old in their first HIV positive test, seroconversion date was estimated as the mean time between the date of the first HIV positive test and the date at age 15 years. According to HIV epidemiological data from the Centro Nacional de Epidemiología in Spain, in the 1981-2004 period only 3 AIDS cases have been reported from IDU individuals with age less than 15 years old (Centro Nacional de Epidemiología, 2005). HIV-1 disease progression was analyzed according to Centers for Disease Control and Prevention (CDC) 1993 criteria (Centers for Disease Control and Prevention, 1993). The resulting data were analyzed using the Kaplan-Meier estimator and no significant differences were found among the genotypes. In Figure 7, we show the estimated survival function for both genotypes obtained by the GS-distribution method (the corresponding Kaplan-Meir estimation is not shown for clarity). The result of the corresponding test was similar to

the Kaplan-Meier method. As this result is obtained by guessing the infection time, one may consider other alternatives.

For instance, instead of taking the mean time as indicated before, we can consider that the time to developing AIDS is greater than t_i, i.e. the time between the first seropositive test and the moment they develop AIDS, and less that $t_i + T_i$, where T_i is the time from first reasonable date for infection, as computed above, to the day of the first seropositivity. Thus, independently of the final outcome of AIDS, all the observations will contribute to the likelihood by a factor of $P(t_i \leq T \leq t_i + T_i)$. This procedure is less restrictive than the method of computing the mean of the interval. In Figure 7 we present the results of estimating the time to AIDS following this strategy. We can see that the estimations obtained by this new strategy is different than the results corresponding for the first method. The procedure based on interval censored data estimates a shorter time to developing AIDS than the results computed from the mean method of estimating the time of infection. In both cases, the differences are not significant between both genotypes. In that case, independently of the method used, we can conclude that among IDU users, the CCR5 genotype is not likely to produce a different rate of progression to clinical AIDS.
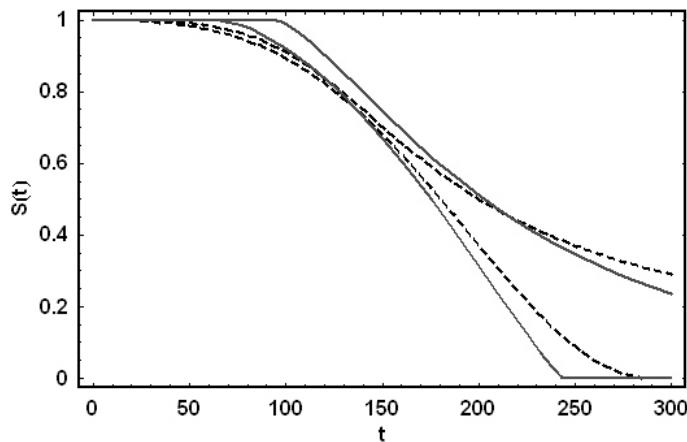
Figure 7*: GS-distributions fitted to data of HIV infected patients. The dashed line indicates the estimated survival functions for the wild type CCR5 gentoptype (n = 151). The continuous line indicates the estimation for the mutant (n = 27) genotype. The steepest curves correspond to the estimation using interval censoring. The other curves correspond to the estimation using the mean method for computing the time of infection. See text for details.*

These results show the interest of further analyzing the influence of the method used for obtaining an estimation of the date of infection. Clearly, considering an interval censored approach may be less subjective as we only establish a reasonable interval in which the event has occurred. The usual method of taken the mean of this interval as the time of infection is somehow more subjective. A careful study of the influence of this choice may shed new light on the problem of progression rate to AIDS. The GS-

distribution approach introduced in this paper would provide a valuable tool for this analysis.

4 Conclusions

We have introduced the GS-distribution as a parametric model for univariate survival analysis. We have shown its utility in simulation examples and we have developed the basis for hypothesis testing using this approach. The use of the GS-distribution as a general model allows for obtaining a parametric representation for survival data independently of the censoring scheme. Once obtained, we can use this model for further exploring the survival process. The tools introduced by Muiño *et al.* (2006) for the GS-distribution can be used for this purpose.

Our results show that the GS-distribution survival function can be used to easily obtain estimates of the survival function in different censoring schemes. Group comparison is straightforward through the likelihood ratio test. Although a systematic study is yet to be performed, our preliminary results indicate that the GS-distribution based method can be an interesting alternative to non-parametric methods. Using this approach, we could even compare two groups when the samples obey different censoring schemes. As a drawback, the extension of this model to cope with covariables is not straightforward. We shall explore practical solutions for this problem in the near future.

References

Centro Nacional de Epidemiología. (2005). *Vigilancia Epidemiológica del SIDA en España. Registro nacional de casos de SIDA*. Available at: http://193.146.50.130/htdocs/sida/sidavih.htm

Centers for Disease Control and Prevention. (1993). From the Centers for Disease Control and prevention. 1993 revised classification system for HIV infection and expanded surveillance case definition for AIDS among adolescents and adults. *Jama*, 269, 460.

Huang, Y., Paxton, W.A., Wolinsky, S.M., Neumann, A.U., Zhang, L., He, T., Kang, S., Ceradini, D., Jin, Z., Yazdanbakhsh, K., Kunstman, K., Erickson, D., Dragon, E., Landau, N.R., Phair, J., Ho, D.D. and Koup, R.A. (1996). The role of a mutant CCR5 allele in HIV-1 transmission and disease progression. *Nature Medicine*, 2, 1240-1243.

Hosmer, D.W.Jr., Lemeshow, S. (1999). *Applied Survival Analysis*. New York: John Wiley.

Hernández-Bermejo, B. and Sorribas, A. (2001). Analytical quantile solution for the S-distribution, random number generation and statistical data modeling. *Biometrical Journal*, 43, 1017-1025

Jones, M.C., (2002). The complementary beta distribution. *Journal of Statistical Planning and Inference*, 104, 329-337.

Jones, M.C., (2004). Families of distributions arising from distributions of order statistics. *Test*, 13, 1-43.

Kamps, U., (1991). A general recurrence relation for moments of order statistics in a class of probability distributions and characterizations. *Metrika*, 38, 215-225.

Klein, K.P., Moeschberger, M.L., (1997). *Survival Analysis*. New York: Springer.

March, J., Trujillano, J, Tort,M. and Sorribas, A. (2003). Estimating conditional distributions using a method based on S-distributions: Reference percentile curves for body mass index in Spanish Children *Growth, Development, and Aging*, **67**, 59-72.

Muiño, J.M., Voit, E.O. and Sorribas, A. (2006). GS-distributions: A new family of distributions for continuous unimodal variables. *Computational Statistics & Data Analysis*, 50, 2769-2798.

Parzen, E., (1979). Nonparametric statistical data modelling (with comments). *Journal of the American Statistical Association*, 74, 105-131.

Sorribas, A., March, J. and Trujillano, J. (2002). A new parametric mtehod based on S-distributions for computing Receiver Operating Characteristic curves for continuous diagnostic tests. *Statistics in Medicine*, 21, 1215-1235.

Sorribas, A., March, J. and Voit, E. O. (2000). Estimating age-related trends in cross-sectional studies using S-distributions *Statistics in Medicine*, 19, 697-713.

Turner, M.E. and Pruitt, K.M. (1978). A common basis for survival, growth, and autocatalysis *Mathematical Biosciences*, 39, 113-123.

Vilaseca, J., Arnau, J.M., Bacardi, R., Mieras, C., Serrano, A. and Navarro, C. (1982). Kaposi's sarcoma and toxoplasma gondii brain abscess in a Spanish homosexual. *Lancet*, 1, 572.

Voit, E.O. (1992). The S-distribution: A tool for approximation and classification of univariate, unimodal probability distributions. *Biometrical Journal*, 7, 855-878.

Voit, E.O. (2000). A maximum likelihood estimator for shape parameters of S-distributions. *Biometrical Journal*, 42, 471-479.

Voit, E.O. and Schwacke,L.. (1998). Scalability properties of the S-distribution, *Biometrical Journal*, 40, 665-684.

Voit, E.O. and Schawcke, L.H. (2000). Random number generation from right-skewed, symmetric and left-skewed distributions. *Risk Analysis*, 20, 59-71.

Voit, E.O. and Sorribas, A. (2000). Computer modeling of dynamically changing distributions of random variables *Mathematical and Computer Modelling*, 31, 217-225.

Voit, E.O., Balthis, W.L. and Holser, R.A. (1995). Hierarchical Monte-Carlo modeling with S-distributions: concepts and illustrative analysis of mercury contamination in king mackerel. *Environmental International*, 21, 627-635.

Yu, S.Y. and Voit E.O. (1995). A simple, flexible failure model. *Biometrical Journal*, 37, 595-609.

Zhao, Q. and Sun, J. (2004). Generalized log-rank test for mixed interval-censored failure time data. *Statistics in Medicine,* 23, 1621-1629.