Bayesian Likelihood Application on Multivariate Survival Functions

Judith H. Parkinson

Department of Mathematics, University of Salzburg, Austria

Abstract

Research stay over 3 months to conduct joint research on application of Bayesian statistics on multivariate survival analysis at *Georgia State University*, Atlanta, USA together with Prof. Dr. Yichuan Zhao, Department of Mathematics and Statistics.

Preprint submitted to Austrian Marshall Plan Foundation, Vienna *April 18, 2019*

Contents

1. Timeline

The local research was conducted over a time frame of a little bit over three months. Beginning on the 2nd of October 2018 we started with the on-site research. It was ended on January 20th 2019.

The research was divided into several sections. Those milestones are as followed:

- Literature Research
- Mathematical Derivation
- Implementation and Validation of the Mathematical Results
- Interpretation of the Findings

The first part could already be done before my actual arrival in the USA. After my arrival at the *Georgia State University* Professor Yichuan Zhao discussed our findings in the literature on multivariate survival analysis and Bayesian statistics. We then decided which approaches seemed most promising and would have a high impact on future results. After prioritizing the research tasks we then started with the actual work. Exchanging and discussing ideas during weekly meetings we then derived the mathematical results for the first approach. After running simulations and analyzing the results we then started the second approach. The task procedure was similar to the first approach. Due to the exceedingly high run times of the simulations of the second research task the validation of that approach could not be fully conducted until the end of the research stay at *Georgia State University*. Instead simulations are continued to run and expected to be completed at the beginning of the summer.

In total the complete research period is expected to last a full year where the main tasks could be completed during the on-site research and the follow-ups are done via Mail.

The results of the first research task where presented at the 14th "Workshop on Stochastic Models, Statistics and Their Applications" from March 6th to March 8th, 2019 in Dresden during a talk in a session. The results of the second research task shall be presented to an appropriate audience at another conference in 2019. Additionally, the results of the second approach are being prepared for publication with a quality journal.

2. Introduction

The task of statisticians is to answer research questions using collected data. It includes everything from setting up a study design, over calculating test statistics, to drawing inference. However, even without a specific research question new methods to test and analyze data sets can be developed to help improve future research. As the application range of statistics is broad and different applications consider different data structures we will first introduce the data structure and methodology on which we will focus during our research.

The statistical description and quantification of the length of time before an event is dubbed survival or time-to-event analysis. The times being analyzed are defined to be the time until the occurrence of an event of interest, such as death, recovery, or transition above or below a pre-specified threshold. Those times, often referred to as failure times, are described by the survival function $S(t), t \geq 0$, which is the probability to be event free up to time point t. Alternatively, they can be expressed through the hazard function, which is defined in the univariate case as

$$h(t) = \lim_{\Delta t \searrow 0} \frac{P(t \leq X < t + \Delta t | X \geq t)}{\Delta t}, \; t \geq 0$$

where X denotes the failure or survival time. Basically one can consider X to be a non-negative random variable.

Typically one is not capable of observing all event times during a study due to drop out or end of the study. So, additionally to the failure time, researchers consider a non-negative random variable C, the so called censoring time, which is typically assumed to be non-informative. One then observes neither X nor C but yet two other random variables, namely, $T := \min\{X, C\}$ and $\delta := \mathbb{1}\{X \leq T\}$. In the case of multivariate survival times the minimum is considered componentwise.

In a multivariate setting things are slightly more complex. Instead of an univariate survival time a vector of several survival times is considered. Most commonly the vector of survival times contains the event times within a single subject. In medicine this could be a patient where the time to blindness in each eye is measured or the time to failure in the right and the left lung. Both would be examples for the special multivariate case with two dimensions, namely the bivariate case. In engineering it could be time until breakdown of several mechanical components of a car.

Depending on the structure and what one is interested in multiple different aspects of survival analysis can be analyzed. If one has competing risks, i.e. an

event occurs due to one of multiple causes, only in one of the components will an event occurred while in the others no event can occur after that first event. Or one might not be interested in the exact event time in each component but instead only in the very first event per subject. We are interested in the multivariate survival function where each component or event time per subject is of interest. For a random, non-negative, d-dimensional vector of survival times $X = (X_1, \ldots, X_d)^T$ we will study the survival function defined as $S(x_1, \ldots, x_d) := P(X_1 \geq x_1, \ldots, X_d \geq x_d)$ with $(x_1, \ldots, x_d)^T \in \mathbb{R}_+^d$.

Methods to estimate the multivariate survival function of X based on the information contained in T and δ have been proposed, yet there currently exists no estimator as well established as the estimators in the univariate case, like the Kaplan-Meier estimator, see Kaplan and Meier (1958), or the Nelson-Aalen estimator, as first mentioned in Nelson (1969). The extensive dealing with the problematic in multivariate survival analysis has not such a long history, thus it is of interest to keep developing new test statistics that satisfy certain criteria. Those range from robustness, i.e. obtaining a good power no matter the true underlying hazard rate under the alternative, to good small sample size behavior, i.e. obtaining the nominal error rates fast. A pharmaceutical study using a method with bad small sample size properties is likely to get rejected by the ethical board, as it would require higher sample sizes as other methods. Medical or pharmaceutical areas are considered to be primary application fields of survival analysis, and there this problematic arises often and needs to be dealt with appropriately. The analysis of multivariate survival times is in the time of massive data ascertainment of high importance due to its manifold application areas. Next to the medical and pharmaceutical research area it can be applied to various others as well. Literature on those fields is plentiful, for example Perrigot et al. (2004), Czado (2002), and Scherm and Ojiambo (2004), to name only a few. A general overview over multivariate survival analysis can be found in Clark et al. (2003a), Clark et al. (2003b), Clark et al. (2003c), and Clark et al. (2003d).

Data sets with small sample sizes are often considered problematic, as test methods will often only start delivering reliable results with a high enough number of samples. Especially in survival analysis this can be problematic as often the data will be partly censored and thus decrease the amount of information that is available for the test statistics. One way to conquer this problematic is to assume a prior of the data which, if chosen correctly, can help improve the steadiness of the test method.

This research area is known as Bayesian statistic. The main idea behind it is to use prior beliefs to obtain more accurate results. Bayesian statistics are, in contrast

to other probability interpretations, subjective. As it is an easy method to include prior information into a new study and can be used for any data, irrespective of its structure, it has become quite popular in several applied research areas. Especially in the era of high-performance computers and new algorithms as for example the Markov Chain Monte Carlo, the Bayesian Statistic has found a home in Data Science, compare Gelman et al. (2013).

It's origin is based on the Bayes' theorem as it was first mentioned in some notes by M. Bayes that were published in Bayes and Price (1763) two years after the death of the author. Bayes' theorem considers that given two events A and B the conditional probability of A given that event B occurred can be calculated using the individual probabilities of A and B as well as the conditional probability of B given A. More explicitly, the theorem states that

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}.$$

Now $P(A)$ can be interpreted as the prior probability which basically represents ones prior beliefs and $P(B)$ is then the standardizing factor, as later explained.

Unlike the frequentist interpretation of probability the Bayesian statistic often requires a parametric modeling of the data as otherwise the calculations would become too complex and could not be correctly interpreted anymore. A parametric model makes it also easier to quantify the conditional probability $P(B|A)$. Based on a parametric model with a single parameter or a vector of parameters, the Bayesian statistic assigns each parameter a probability. This probability can, loosely spoken, be interpreted as the subjective beliefs on the correctness of the parameter. A parameter with a higher probability is more likely to be the true parameter than one with a low probability. As the probability $P(B)$ is equal for all prior beliefs, yet hard to calculate, it is sufficient to simply calculate $P(B|A) \cdot P(A)$ to obtain the best estimator. More precisely, it holds true that $P(A|B) \alpha P(B|A) \cdot P(A)$. Even though $P(B)$ is hard to calculate beforehand, one can obtain it by the knowledge that the sum or integral over all probabilities $P(A|B)$ must be equal to one. The probability $P(A|B)$ is often referred to as the posterior while $P(A)$ is referred to as the a priori.

A general overview on Bayesian Theory can be found in Bernardo and Smith (1994). While the prior can be applied in different forms to the data, we will apply the prior to some functional θ, as we will explain more detailed in the next section. The validity of the posterior results to use them for inference when using Bayesian prior on the functional was for example analyzed in Lazar (2003).

As mentioned above Bayesian statistic is only one way to interpret probabilities. The two other main interpretations are the frequentistic or empirical and the logical. A brief discussion over all three can be found in von Weizsäcker (1992).

The current lack of appropriate methods to analyze all data sets from the various areas, the constantly growing expectations that statisticians obtain better results and draw more accurate inference with high-dimensional data from undersized data set, require constant research on the field of survival analysis.

One won't be able to find a method that is appropriate for all realized data sets no matter the origin. Thus our focus was only concentrated on the small subsection of application fields in medical and pharmaceutical studies.

In this report we will explain in detail our ideas for new test methods, their mathematical validations, their simulation performances and an overall assessment of the approaches.

While introducing the first idea we will additionally explain the general procedure steps that are being followed while constructing new test methods. Without stating them again we will follow those steps during the second approach as well.

3. Bayesian Likelihood Application on Multivariate Survival Times under Constraints

The first approach of the research conducted at *Georgia State University* was based on Parkinson (2019). In that paper a method to conduct hypothesis testing on multivariate survival times was introduced. Using multiple constraints the introduced method showed good performances in multiple simulations settings. Those, however, came at the cost of a high required sample size making the introduced test statistic unfavorable for real data sets with only a small sample size. Yet the proposed idea should be made available for those cases which often occur in medicine. To this end we introduced a Bayesian statistic to the original method in the hopes of an improved sample size behavior.

3.1. Introduction

Many tests analyzing the survival experience of one or more populations, require that one chooses a weight function. The weight functions assign each time point a non-negative value that basically expresses how critical that time point is for the analysis. If for example one tests two medical procedures to see which has the better outcome and one suspects bigger differences in the early time points one assigns higher weights to them as to the later time points. It is possible to select an optimal weight function if the true shape of the hazard ratio under the alternative

hypothesis is known. However, in general precisely this is unknown and, thus, one is not capable of getting the best result with regard to a specific data set. In certain situations the wrong choice of weight function leads to an immense power reduction. Illustrations of this were shown with various data sets, see Kosorok and Lin (1999), Klein and Moeschberger (1997). Different methods in the univariate case were proposed to conquer the problem, see e.g., Brendel et al. (2014) and Bathke et al. (2009).

Most commonly used for multivariate survival data is the Cox model. Also, the accelerated failure model is often applied. Yet, both, just like many other of the multivariate models, have limitations. Some assume a parametric model, like the Cox model which assumes proportional hazards. Other methods lack the ability to correctly take the dependency structure of the underlying data into account. Many methods that are primary based on univariate or component wise test statistics conquer this problem by integrating the covariance matrix into the multivariate test statistic, as done, for example, by Wei and Lachin (1984).

The goal of the new method obviously is to conquer the problems that arise when setting up models and tests for multivariate survival times.

The approach of Bathke et al. (2009) was adapted to the multivariate setting by Parkinson (2019). A multivariate test statistic was developed based on conditional hazards. Applying constraints onto the conditional hazards the test was effective against several different kinds of alternatives. Now, modifying the test statistic to be based on a parametric model, instead of the non-parametric model, we used Bayesian priors to adapt that approach to obtain better small sample size behavior.

3.2. Mathematical Definitions and Theory

Due to the fact that the mathematical background of the one-sample and the two-sample case is quite similar, we will only focus on the one-sample case when introducing the notations and the mathematical theory. Further, similar to Parkinson (2019), we will only consider the bivariate case, even though the method could easily be extended to the true multivariate case.

Let $X_1, \ldots, X_n$ be independent, identically distributed d-dimensional observations with $X_i := (X_{1i}, X_{2i})^T$, distribution function F_0, and marginal distribution functions F_{01} and F_{02}. Further we denote the corresponding cumulative hazard function by Λ_0 and the marginal cumulative hazard functions by Λ_{01} and Λ_{02}. As we will allow for right censoring we only observe

$$T_{ji} = \min\{X_{ji}, C_{ji}\} \text{ and } \delta_{ji} = \mathbb{1}\{X_{ji} \leq C_{ji}\}\ ,\ i = 1, \ldots, n,\ j = 1, 2, \qquad (1)$$

with $C_i = (C_{1i}, C_{2i})^T$ the censoring times. For the censoring times we will assume them to be as well independent, identically distributed with distribution function G_0 and independent of the X_i's.

Following an approach by Dabrowska (1988) a joint bivariate survival function $S(x,y) = P(X \geq x, Y \geq y)$ and thus conditional hazards can be defined. We denote those hazards by

$$\begin{aligned}
\lambda_{11}^{10}(t)dt &= P(t \leq X_1 \leq t+dt | X_1 \geq t, X_2 > t) \,, \\
\lambda_{11}^{01}(t)dt &= P(t \leq X_2 \leq t+dt | X_1 > t, X_2 \geq t) \,, \\
\lambda_{10}^{10}(t)dt &= P(t \leq X_1 \leq t+dt | X_1 \geq t, X_2 < t) \,, \\
\lambda_{01}^{01}(t)dt &= P(t \leq X_2 \leq t+dt | X_1 < t, X_2 \geq t) \,.
\end{aligned}$$

The empirical likelihood for n observations is the product $V = \prod_{i=1}^{n} V_i$ with

$$\begin{aligned}
V_i = \prod_t \Big[& (1-\lambda_{11}^{10}(t)dt)^{\mathbb{1}\{T_{1i}>t\}\mathbb{1}\{T_{2i}>t\}} \cdot (\lambda_{11}^{10}(t)dt)^{\mathbb{1}\{T_{1i}=t\}\mathbb{1}\{T_{2i}>t\}\delta_{1i}} \\
& \cdot (1-\lambda_{11}^{01}(t)dt)^{\mathbb{1}\{T_{1i}>t\}\mathbb{1}\{T_{2i}>t\}} \cdot (\lambda_{11}^{01}(t)dt)^{\mathbb{1}\{T_{1i}>t\}\mathbb{1}\{T_{2i}=t\}\delta_{2i}} \\
& \cdot (1-\lambda_{10}^{10}(t)dt)^{\mathbb{1}\{T_{1i}>t\}\mathbb{1}\{T_{2i}<t\}\delta_{2i}} \cdot (\lambda_{10}^{10}(t)dt)^{\mathbb{1}\{T_{1i}=t\}\mathbb{1}\{T_{2i}<t\}\delta_{2i}\delta_{1i}} \\
& \cdot (1-\lambda_{01}^{01}(t)dt)^{\mathbb{1}\{T_{1i}<t\}\mathbb{1}\{T_{2i}>t\}\delta_{1i}} \cdot (\lambda_{01}^{01}(t)dt)^{\mathbb{1}\{T_{1i}<t\}\mathbb{1}\{T_{2i}=t\}\delta_{2i}\delta_{1i}} \Big] \,.
\end{aligned}$$

Let t_{jl}, $j = 1,2$, be the ordered distinct time points of jumps in the j-th component. For simplicity in notation we will denote the hazards at time point t_{jl} by $a_l := \lambda_{11}^{10}(t_{1l})dt$, $b_l := \lambda_{11}^{01}(t_{2l})dt$, $c_l := \lambda_{10}^{10}(t_{1l})dt$, and $d_l := \lambda_{01}^{01}(t_{2l})dt$. Then the empirical log-likelihood function, denoted by EL in the following, is given by

$$\begin{aligned}
& \sum_{t_{1l}} \Big[\log(1-a_l) \sum_i (\mathbb{1}\{T_{1i} > t_{1l}\}\mathbb{1}\{T_{2i} > t_{1l}\}) + \log(a_l) \sum_i (\mathbb{1}\{T_{1i} = t_{1l}\}\mathbb{1}\{T_{2i} > t_{1l}\}\delta_{1i}) \Big] \\
& + \sum_{t_{2l}} \Big[\log(1-b_l) \sum_i (\mathbb{1}\{T_{1i} > t_{2l}\}\mathbb{1}\{T_{2i} > t_{2l}\}) + \log(b_l) \sum_i (\mathbb{1}\{T_{1i} > t_{2l}\}\mathbb{1}\{T_{2i} = t_{2l}\}\delta_{2i}) \Big] \qquad (2) \\
& + \sum_{t_{1l}} \Big[\log(1-c_l) \sum_i (\mathbb{1}\{T_{1i} > t_{1l}\}\mathbb{1}\{T_{2i} < t_{1l}\}\delta_{2i}) + \log(c_l) \sum_i (\mathbb{1}\{T_{1i} = t_{1l}\}\mathbb{1}\{T_{2i} < t_{1l}\}\delta_{1i}\delta_{2i}) \Big] \\
& + \sum_{t_{2l}} \Big[\log(1-d_l) \sum_i (\mathbb{1}\{T_{1i} < t_{2l}\}\mathbb{1}\{T_{2i} > t_{2l}\}\delta_{1i}) + \log(d_l) \sum_i (\mathbb{1}\{T_{1i} < t_{2l}\}\mathbb{1}\{T_{2i} = t_{2l}\}\delta_{1i}\delta_{2i}) \Big]
\end{aligned}$$

where the sums over t_{jl} are only taken over those time points where a_l, respectively b_l, c_l, and d_l are truly greater than zero, and excluding the last time point, as there the hazards could be one.

Unlike Parkinson (2019) we now have to focus on the parametric case, which depicts a limitation to the original method as this was non-parametric. Next to the general limitations one has to deal with, we additionally encountered the problematic that no direct formula for the maximum likelihood estimator could be derived but instead intensive numerical optimization procedures are required. We later used pre-implemented functions in R to obtain the values of the parameter(s) that maximized the function EL.

Assuming that the underlying distribution function is defined through a single parameter or a vector of parameters, denoted by θ, the true conditional hazards are functions dependent on the parameter(s) and the time point at which the functions are evaluated, i.e. $a_l = \lambda_{11}^{10}(t_{1l}, \theta)dt$, ..., $a_l = \lambda_{01}^{01}(t_{2l}, \theta)dt$.

Now, maximizing equation (2) would result in the previous results as provided in Parkinson (2019). We however want to include some prior information into the test statistic. To do this we need to assign each vector of parameters a prior likelihood. The prior likelihood of a parameter vector θ will be denoted by $\pi(\theta)$. Instead of only considering the empirical log-likelihood function as given in equation (2) we will maximize the following function

$$EL_{orig}(\theta) \cdot \pi(\theta),$$

where $EL_{orig}(\theta)$ is the empirical likelihood function with the conditional hazards based on the parameter(s) θ plugged in. Alternatively, we can maximize the logarithm of the upper function, namely

$$EL(\theta) + \log(\pi(\theta)) \tag{3}$$

with $EL(\theta)$ the empirical log-likelihood function, again with the conditional hazards based on the parameter(s) θ plugged in and $\log(\pi(\theta))$ the logarithm of the prior likelihood of θ.

The conditional hazards evaluated at a specific time point t are determined through the optimal parameter(s) θ which are simply the parameter(s) that maximizes equation (3). We will denote those optimal parameter(s), i.e. those maximizing equation 3 by θ^* and the optimal conditional hazards a_l^*, b_l^*, c_l^*, and d_l^*.

This gives us the first part of the approach. For the second part we now need to introduce some more definitions before we can get to the actual results.

As we want to impose constraints onto the empirical log-likelihood function and for each added constraint the degrees of freedom of the limiting distribution increases, we want to combine the conditional hazards component wise to limit the number of constraints. Through the fact that for each extra constraint the number

of degrees of freedom increases by one, and the higher the number of degrees of freedom the higher the required sample size it conflicts with our interest of reducing the required sample size. It is thus critical to impose the constraints correctly.

To be able to do so we need to define the following two hazards

$$v_{1l} = P(T_2 > t_{1l} | T_1 \geq t_{1l}) \cdot a_l + P(T_2 < t_{1l} | T_1 \geq t_{1l}) \cdot c_l.$$

and

$$v_{2l} = P(T_1 > t_{2l} | T_2 \geq t_{2l}) \cdot b_l + P(T_1 < t_{2l} | T_2 \geq t_{2l}) \cdot d_l.$$

The constraints that are then imposed on the conditional hazards are given in the form of

$$\sum_l g_{jr}(t_{jl}) \log(1 - v_{jl}) = \mu_{jr} \tag{4}$$

for the components $j = 1,2$ and the r-th constraint function $g_{jr}, r = 1,\ldots,k$. The constraint functions must all be non-negative and predictable with respect to a filtration $\mathscr{F}_t = \sigma\{T_{1i}\mathbf{1}\{T_{1i} \leq t\}; \delta_{1i}\mathbf{1}\{T_{1i} \leq t\}; T_{2i}\mathbf{1}\{T_{2i} \leq t\}; \delta_{2i}\mathbf{1}\{T_{2i} \leq t\}; i = 1,\ldots,n\}$, which is identical to the one provided in Parkinson (2019).

The empirical log-likelihood function under constraint, denoted by G_C, is then given through

$$EL(\lambda) + \sum_{j=1}^{2}\sum_{r=1}^{k} n\lambda_{jr}\left[\mu_{jr} - \sum_i g_{jr}(t_{ji}) \log(1 - v_{ji})\right],$$

where $EL(\lambda)$ denotes the log-likelihood function, as given in (2) with the modified hazards $a_l(\lambda),\ldots,d_l(\lambda)$. In this setting λ is the Lagrange multiplier and $a_l(\lambda),\ldots,d_l(\lambda)$ are the parametric, conditional hazards that solve the constraint function as given in equation (4). Again the last sum ranges only over the distinct time points excluding the last.

Unlike in the first part we now do not include the prior into the function. As we want the value of the function parameters that maximizes G_C, i.e. solve the constraints, we do not want prior beliefs included as they could, if incorrect, redirect us from the optimal parameters, i.e. those parameters that solve the constraints as given in equation (4).

Unfortunately, as in the first part, we yet again can not provide a direct formula for the optimal parameter(s) θ or the conditional hazards. Instead we have to use numerical computations yet again.

Let us consider $\vartheta := (\vartheta_{11}, \ldots, \vartheta_{1k}, \vartheta_{21}, \ldots, \vartheta_{2k})^T$ a $(2k)$-dimensional parameter defined via the components' cumulative hazard functions Λ_{01} and Λ_{02},

$$\vartheta_{jr} = \int g_{jr}(t) \log(1 - d\Lambda_{0j}(t)) \ , \ r = 1, \ldots, k, \ j = 1, 2,$$

and a hypothesis testing problem

$$H_0 : \vartheta_{jr} = \mu_{jr} \ \forall j = 1, \ldots, d, r = 1, \ldots, k \ \text{ vs. } \ H_A : \vartheta_{jr} \neq \mu_{jr} \text{ for some } j \text{ and } r, \quad (5)$$

where $g_{jr}(t)$ are some non-negative functions, $\vartheta_j := (\vartheta_{j1}, \ldots, \vartheta_{jk})^T$ and $\mu_j = (\mu_{j1}, \ldots, \mu_{jk})^T$ is a vector of constants.

Let the test statistic in terms of hazards be given by

$$W = -2\{\max G_C - \max EL(\theta)\} \ , \quad (6)$$

where $EL(\theta)$ is empirical log-likelihood evaluated as given in (2). Both empirical log-likelihood functions will be evaluated at the maximum likelihood estimators as explained previously.

Theorem 3.1. *Suppose that the null hypothesis H_0, as defined in equation* (5)*, holds for non-negative, random functions $g_{jr}(t)$ that are predictable with respect to the filtration $\mathscr{F}_t = \sigma\{T_{1i}\mathbf{1}\{T_{1i} \leq t\}; \delta_{1i}\mathbf{1}\{T_{1i} \leq t\}; T_{2i}\mathbf{1}\{T_{2i} \leq t\}; \delta_{2i}\mathbf{1}\{T_{2i} \leq t\}; i = 1, \ldots, n\}$. Then, under some further regularity conditions as stated in Parkinson (2019), the test statistic W has asymptotically a chi-squared distribution with $2k$ degrees of freedom, where k is the number of constraints per dimension.*

For the interpretation of the constraints, and thus the null hypothesis, consult Section 4 'Applications' of Parkinson (2019). Examples of random but predictable constraint functions are provided in detail as well there.

3.3. Implementation and Simulations

Next to mathematical proofs simulation results are provided in statistical findings. This is due to the fact that most statistical results hold only asymptotically, i.e. for the sample size n converging to infinity. Now, obviously in real life the sample size is a pre-specified number or, if not, most certainly not converging to infinity. It is therefore standard to provide simulations, based on, preferred, realistic settings, for smaller sample size to show the accuracy of the method, even if one of the assumptions is obviously not fulfilled.

Carefully assessing the proofs provided in Parkinson (2019) we analyzed the likeliness that the proof of the new approach could be conducted in a similar manner. The assumptions needed in the original proof are still fulfilled in the new setting which is why it seems reasonable that the proof of the new method works as well.

Now, before writing down the proof in a detailed manner we started to conduct several simulations. All simulations were conducted using R (R version 3.5.2, R Core Team, 2018). To be able to compare the Bayesian approach with the original approach we followed the simulations settings as provided in Parkinson (2019).

For both components three constraints are being applied. The constraints are given by the deterministic functions

$$g_1(t) = \exp\{-t\}, \quad g_2(t) = 0.5t\mathbb{1}\{t \leq 1\}, \quad g_3(t) = \mathbb{1}\{0.5 < t < 1.5\},$$

as already used in the simulations of Parkinson (2019). We used the following distributions to generate the bivariate random variables:

$$X_1 \sim \exp\{0.9\}, \quad X_2 \sim \exp\{0.6\}, \quad C_1 \sim \exp\{0.5\}, \quad C_2 \sim \exp\{0.3\}.$$

The data that was then actually observed was then created as explained in Section 2. As we started off with the simplest setting model we assumed that the two components were independent. Censoring accumulated to roughly one third in each of the components and overall 42% of the observed data $T_1, \ldots, T_n$ could be observed completely, i.e. were uncensored in both components.

As for the priors we considered in total six different prior functions. For all six the likelihood of the prior was assumed to be normally distributed, i.e. $\pi(\theta) \sim N(\cdot, \cdot)$. Three of them had the true value of θ as expectation, while the other three had incorrect values as expectation. For each of the three we had different variances ranging from 0.01, 0.1 to 0.5.

The parametric model we considered for our test statistic is based on four conditional hazards, such that we need to estimate a four-dimensional vector of parameter(s). In the above setting, due to the independence between the components, the true parameter vector was given by $\theta_0 = (0.9, 0.9, 0.6, 0.6)^T$.

Setting the sample size to 200 we simulated each of the six choices of prior beliefs 500 times. A resulting Q-Q plot is provided in Figure 1. The plot is a scatter plot of the simulated quantiles and the theoretical quantiles. If the derived asymptotic distribution of the test statistic is correct the plotted points run along the red line. The lower, green, horizontal line corresponds to the 90% quantile of the theoretical quantiles and the upper, blue, horizontal line corresponds to the

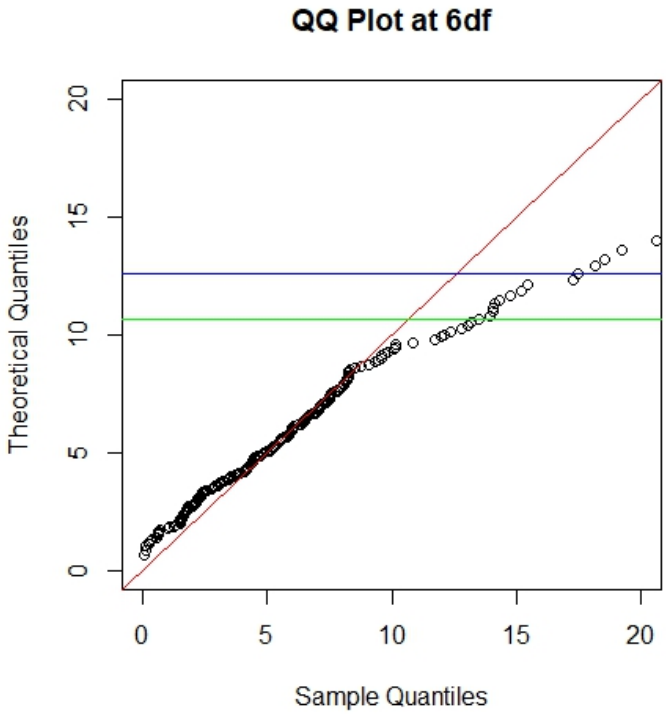

Figure 1: Q-Q plot of the empirical quantiles vs. $\mathcal{X}_6^2$ percentiles for sample size 200 in the one sample case, corresponding to the theoretical result in Theorem 3.1. The prior was normal distributed with expectation θ_0, variance 0.1, and covariances of zero between all components.

95% quantile. As it can be seen the simulated quantiles follow the theoretical quantiles roughly.

Looking at Figure 2 which contains the corresponding simulated results of the original method we can however tell that our simulations were not too great. The quantiles break off with the new method already before the 90% quantile while with the original method it did not break off until roughly the 97.5% quantile.

Similar results could unfortunately be observed with the other settings for the distribution of the prior likelihood. This left us puzzled on how this could be as the application of the a Bayesian prior is well known for and very effective in decreasing the required sample size or, as in our case, obtaining more accurate results with the same sample size.

To figure out where the problem arose from we applied a non-informative or,

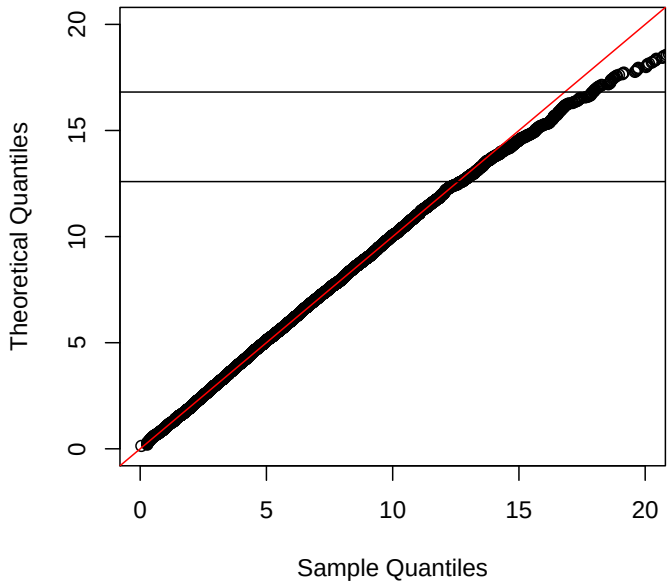

Figure 2: Q-Q plot of the empirical quantiles vs. $\mathscr{X}^2_6$ percentiles for sample size 200 in the one sample case based on the original method and 10000 simulation runs.

also called, flat prior to the empirical log-likelihood function. This meant that we ran simulations on the parametric model without the influence of prior beliefs.

We observed similar results as in the case with prior knowledge. However a slight drop in performance could be observed anyhow. This lead to the conclusion that the application of a Bayesian prior to the approach had a positive impact. The loss of performance was assumed to come from the change from the non-parametric to the parametric model. Simulating higher sample sizes the performance of the new test improved slowly but gradually. Additionally, we analyzed the simulation results of the smaller sample sizes step by step to figure out what exactly caused the parametric model to fail.

3.4. Discussion

Even though it seemed reasonable for the approach to work, the simulations proved us otherwise. Analyzing the results more detailed we noticed that the bad simulation behavior was not a result of the application of a prior to the data but

emerged from the fact that we switched from a non-parametric to a parametric model. It is well known that the parametric model is less flexible and in cases where the model structure is complex this can lead to challenges when estimating the model parameters. This sometimes leads to an increased required sample size as it seems to be the case in our setting.

Working with conditional hazards at different time points they can be estimated well for those time points where the situation we are conditioning on is likely. Looking at c_l, the hazard for the first component conditional on the fact that an event has already occurred in the second component. Now for small time points it is unlikely that an event will have already occurred in the second component. This led to an underestimation of the true value for small time points and a slight overestimation for larger time points. In the non-parametric model this wasn't much of a problem as there c_l could be estimated for every time point individually. In the parametric model though the estimators for $c_1, c_2, \ldots$ all depend on each other through the common parameter. This led, in the case of c_l, to a decreased estimator of the true parameter for that conditional hazard. Similar for d_l while a_l and b_l both had an inflated estimator. The sample size that would be needed to get fitting estimators would thus be significantly higher than the one needed in the original method. Small simulations showed that a sample size of at least $n = m = 1000$ would be necessary to obtain decent results. Applying the priors onto the parametric model helped to keep this problem constrained but couldn't get the performance to the level of the non-parametric model.

Although the method sounded promising it turned out to be a dead end. Further ideas of improving the original method in several aspects were thus rejected as they all would have required the usage of the parametric model, which clearly was the reason the approach failed to provide better simulation results than the original method as given in Parkinson (2019).

4. Bayesian Likelihood Application on Empirical Likelihood for the Bivariate Survival Times

After the non-satisfying results of our first idea, we proceeded with a different idea. To do so we consulted the paper Huang and Zhao (2017). In that paper a method to conduct hypothesis testing on bivariate survival time was introduced. After a first analysis of two already existing approaches Huang and Zhao (2017) adapted the two to the effect that empirical likelihood could be conducted to obtain confidence intervals. The approach via the empirical likelihood proved to be more reliable than the original approach in the reported simulations. Our goal

was yet again to obtain even better results by applying a Bayesian prior to the empirical likelihood ratio statistic. Not only were we hoping for an improvement in the sample size but also for an increased expected coverage probability of the confidence intervals while their average length should not increase.

4.1. Introduction

Bivariate survival data is commonly analyzed and new estimation and test methods for it are developed constantly. For univariate censoring, Lin and Ying (1993) and Wang and Wells (1997) both have developed path-dependent non-parametric estimators of the survival function. Dabrowska (1988) estimated it using a product integral representation for general bivariate censoring. Several other estimators for the bivariate survival function have been proposed. Overviews, as well as analysis of the performances of the estimators, have been provided as well, for instance, Gill et al. (1995), and Wang and Zafra (2009).

Often bivariate survival times are observed in one subject. In those cases univariate censoring, for example, drop out of the subject, is reasonable. Several of the older estimators for bivariate survival functions under univariate censoring have been adapted and improved, compare, for example, Huang and Zhao (2017). Huang and Zhao (2017) provided empirical likelihood confidence interval estimation for the Lin-Ying estimator, Lin and Ying (1993), and for the Wang-Wells estimator Wang and Wells (1997). As the empirical likelihood, as introduced by Owen (2001), is a reliable non-parametric approach for constructing confidence intervals the new approach outperformed the original approach and one of the adaptions in the provided simulations.

As the expected coverage probability in several of the small sample size simulations as given in Huang and Zhao (2017) stayed below its nominal level we aimed to improve the method by applying prior knowledge to the empirical likelihood approach for the Lin-Ying estimator.

In the following we will first provide previous results that are needed for the new method before stating the results of the improved approach. The notation in this work corresponds to the notation of Huang and Zhao (2017). While Huang and Zhao (2017) considered two estimators, namely those introduced by Lin and Ying (1993) and Wang and Wells (1997), we only focused on the first of the later two.

4.2. Mathematical Definitions and Theory

Let (X_i, Y_i), $i = 1, \ldots, n$ be independent, identically distributed bivariate survival times with survival function $S(x,y) = P(X \geq x,\, Y \geq y)$. Denote the marginal

survival functions as $S_1(x) = P(X \geq x)$ and $S_2(y) = P(Y \geq y)$. Further, let C_1, ..., C_n be univariate right-censoring times with survival function $G(t) = P(C \geq t)$ which are independent, identically distributed and independent of (X_i, Y_i). Due to the right censoring we only observe $(\tilde{X}_i, \tilde{Y}_i, \delta_i^X, \delta_i^Y)$ where $\tilde{X}_i = \min\{X_i, C_i\}$, $\tilde{Y}_i = \min\{Y_i, C_i\}$, $\delta_i^X = \mathbb{1}\{X_i \leq C_i\}$ and $\delta_i^Y = \mathbb{1}\{Y_i \leq C_i\}$. Denote the survival function of the actually observed data by $H(x,y) = P(\tilde{X} \geq x, \tilde{Y} \geq y)$. In the following we denote $x \wedge y = \min(x,y)$ and $x \vee y = \max(x,y)$.

Lin and Ying (1993) denoted the bivariate survival function as

$$S(x,y) = \frac{H(x,y)}{G(x \vee y)},$$

and suggested to estimate it by

$$\hat{S}(x,y) = \frac{\hat{H}(x,y)}{\hat{G}(x \vee y)},$$

where $\hat{H}(x,y) = \sum_{i=1}^{n} \mathbb{1}\left(\tilde{X}_i \geq x, \tilde{Y}_i \geq y\right)/n$ is the empirical estimator of $H(x,y)$ and $\hat{G}(\cdot)$ is the Kaplan-Meier estimator of $G(\cdot)$ based on the event time $\tilde{C}_i = \tilde{X}_i \vee \tilde{Y}_i$ and censoring indicator $\delta_i^C = 1 - \delta_i^X \delta_i^Y$.

Lemma 4.1. *[Lin and Ying(1993)] For any fixed $(x,y) \in [0,\tau]^2$, where τ satisfies $S(\tau,\tau)G(\tau) > 0$, $\sqrt{n}\{\hat{S}(x,y) - S(x,y)\}$ converges to a zero mean normal distribution with variance σ^2 as $n \to \infty$ where*

$$\sigma^2 = \frac{S(x,y)}{G(x \vee y)} - S^2(x,y) \left[1 - \int_0^{x \vee y} \frac{dG(u)}{G^2(u) P(X \vee Y \geq u)} \right].$$

For the proof of the 4.1 please consult Lin and Ying (1993).

To provide confidence intervals for the true value of the survival function evaluated at a fixed time point $(x,y) \in [0,\tau]^2$ Lin and Ying (1993) provided random variables W_i, $i = 1, \ldots, n$, based on which an empirical likelihood ratio could be constructed. Previous simulations already showed a low coverage accuracy for small sample sizes when using the empirical variance estimator. However the provided estimator proved to be not reliable either as the W_i's are not asymptotically independent, identically distributed as shown by Huang and Zhao (2017).

A better variable to calculate confidence intervals was provided by Huang and Zhao (2017). It is given by

$$U_i = \frac{\mathbb{1}\{\tilde{X}_i \geq x, \tilde{Y}_i \geq y\}}{\hat{G}(x \vee y)} - S(x,y) \left[1 - \int_0^{x \vee y} \frac{\delta_i^C \, d\mathbb{1}\{\tilde{C}_i \leq u\} - \mathbb{1}\{\tilde{C}_i \geq u\} \, d\hat{\Lambda}_C(u)}{\hat{G}(u)\hat{P}(X \vee Y \geq u)} \right], \tag{7}$$

where $\hat{\Lambda}_c(\cdot)$ is the Nelson-Aalen estimator of the cumulative hazard function $\Lambda_C(\cdot)$ for the censoring time variable C and

$$\hat{P}(X \vee Y \geq u) := n^{-1} \sum_{i=1}^{n} \frac{\mathbb{1}\{\tilde{C}_i \geq u\}}{\hat{G}(u)}.$$

Now those U_i, $i = 1, \ldots, n$, are asymptotically independent, identically distributed.

Now those U_i, $i = 1, \ldots, n$, can be used to obtain the EL ratio, see Owen (2001). For a given (x,y) the EL ratio evaluated at $\theta = S(x,y)$ is defined as

$$R(\theta) = \sup_{\{p_i\}} \left\{ \prod_{i=1}^{n} (n p_i) : \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i U_i(\theta) = 0, \; p_i \geq 0, i = 1, \ldots, n \right\}. \tag{8}$$

Applying the Lagrange multiplier method, we obtain the supremum for $p_i = n^{-1}(1 + \lambda U_i)^{-1}$. Further the empirical log likelihood ratio is given by $l(\theta) = -2 \log R(\theta) = 2 \sum_{i=1}^{n} \log\{1 + \lambda U_i(\theta)\}$, where $\lambda = \lambda(\theta)$ is the solution of

$$\frac{1}{n} \sum_{i=1}^{n} \frac{U_i(\theta)}{1 + \lambda U_i(\theta)} = 0 \, . \tag{9}$$

Denote by $\theta_0 = \theta_0(x,y)$ the true value of the survival function at a fixed (x,y) in the following. As given by Huang and Zhao (2017) and following Wilks' theorem, the following lemma states the convergence of the log likelihood ratio at the true value.

Lemma 4.2. *For any fixed $(x,y) \in [0,\tau]^2$, where τ satisfies $S(\tau,\tau)G(\tau) > 0$, $l(\theta_0)$ converges in distribution to $\mathscr{X}_1^2$ as $n \to \infty$, where $\mathscr{X}_1^2$ is a standard chi-square random variable with one degree of freedom.*

Even though the simulations provided by Huang and Zhao (2017) showed good results for the expected coverage probability of the confidence intervals the required sample size to obtain the nominal level could be improved.

To this end we propose the application of prior information onto the log likelihood ratio. For a fixed time point (x,y) let $\pi(\theta)$ be the prior likelihood of the value θ for $S(x,y)$. Then the log likelihood ratio under prior is given by

$$l^*(\theta) = 2\left(\sum_{i=1}^{n}\log\{1+\lambda U_i(\theta)\} - \pi(\theta)\right). \tag{10}$$

Following the reasoning of several papers, f.e. Lazar (2003) and Cheng and Zhao (2019), the following theorem holds true.

Theorem 4.3. *Let $\theta(F)$ be the functional of interest with no nuisance parameter. Under standard regularity conditions, as $n \to \infty$, the posterior distribution of $\theta(F)$ converges to a normal distribution with mean m_n and variance J_n where*

$$m_n = J_n^{-1}(J_0 m_0 + J(\hat{\theta}_n)\hat{\theta}_n),\ \ J_n = J_0 + J(\hat{\theta}_n); \tag{11}$$

with m_0 the prior mode, $\hat{\theta}_n$ the profile maximum empirical likelihood estimate of $\theta(F)$, J_0 minus the second derivative of the log prior distribution evaluated at m_0, and $J(\hat{\theta}_n)$ minus the second derivative of the log empirical likelihood evaluated at $\hat{\theta}_n$.

For the proof of the above theorem please consult Lazar (2003).

Even though this provides a nice framework for the construction of confidence intervals we encounter a problem when calculating the mean and the variance exactly. Due to the fact that the derivation of the log empirical likelihood involves the derivation of the Lagrange multiplier we cannot derive it analytically. Correspondence with authors of other publications and consulting various papers with similar results did not result in any satisfactory solution to this problem. Some used additional simulations to obtain the confidence intervals that can not be performed in that manner with a real data set while others simply ignored the fact that the Lagrange multiplier is depended on the parameter θ. Numerical approximation of the second derivative proved to be unreliable in several simulations, especially for moderate to high sample sizes.

To this end, we suggested a bootstrap approach to obtain the boundaries of the confidence intervals. Bootstrapping is a commonly used approach for small sample size data sets or data sets where a empirical variance estimator is suspected to be unreliable. There are several different ways to bootstrap data. We chose a plain bootstrap where "new" data sets are created using the observations of the original data set. Out of the original n observations one randomly draws n observations,

where the same observation can be drawn multiple times while others might not be drawn at all. This new set of observations is then a new bootstrapped sample set. Repeating this procedure several times one then can build estimators on not only one but several data sets.

Theorem 4.4. *Let (X_j^*, Y_j^*), $j = 1, 2, \ldots, k$, be bootstrapped sample sets with each n "new" observations of the original bivariate survival times $(X_1, Y_1), \ldots, (X_n, Y_n)$. Now under standard regularity conditions, as $m \to \infty$, it holds that $E(l_{n,\cdot}^*) \to m_n$ and $Var(l_{n,\cdot}^*) \to J_n$, where $l_{n,j}^*$ are estimators for $l^*(\theta_0)$ based on the bootstrap sample. More precisely $l_{n,j}^*$ is the maximum empirical log likelihood estimator under prior based on the j-th bootstrap sample. $l_{n,\cdot}^*$ is the random variable that is being estimated by $l_{n,1}^*, \ldots, l_{n,k}^*$.*

Based on Theorem 4.4 we can construct confidence intervals for the true value of the survival function at a pre-specified time point (x, y). For each data set we calculate the mean of the bootstrap estimators $l_{n,j}^*$, $j = 1, \ldots, k$, where k is the number of repetitions for the bootstrap. Next we can calculate a variance estimator based on those created samples, namely the empirical variance of the bootstrap estimator. Based on the mean and the variance we can then calculate a 95% confidence interval, denoted by CI_q, based on the quantiles for $l^*(\theta_0)$. The actual confidence interval for the parameter θ is then the interval $[a, b]$ which includes all θ's for which $l^*(\theta) \in CI_q$.

4.3. Implementation and Simulations

All simulations were conducted using R (R version 3.5.2, R Core Team, 2018). To be able to compare the Bayesian approach with the original approach we followed the simulations settings as provided in Huang and Zhao (2017).

Due to some misreported numbers all true values of the survival function of all settings as well as all simulations were calculated again to ensure that the simulation results are comparable.

In the first setting we considered independent components. The first component was distributed according an exponential distribution with parameter 1. The second component followed a log-normal distribution with expectation 0.1 and variance 0.5^2. In this case the true value of the survival function evaluated at a specified time point $S(x, y)$ can be calculated by

$$e^{-x}\left(1 - \Phi\left(\frac{\log y - 0.1}{0.5}\right)\right).$$

The univariate censoring was created according to an exponential distribution once with the parameter 0.25, which resulted in censoring of 20% in the first component and 25% in the second component, and parameter 0.65, which lead to 40% and 53% censoring in the two components. Again, we only observed the first-event vector, i.e. $T = (\min\{X_1, C\}, \min\{X_2, C\})^T$, and whether an event actually occurred at the observed time point.

For each setting we applied six different prior likelihoods onto the data sets. All six priors were normally distributed. The first three had expectation $\mu_0 = S(x,y)$, namely the correct value, the other three were assigned an incorrect expectation μ_1 which, dependent on the true value $S(x,y)$, was chosen in a manner to be far from the true value. For the variance we used three different values $\sigma_1 = 1$, $\sigma_2 = \sqrt{1/2}$, and $\sigma_3 = \sqrt{1/10}$.

For each setting we ran 1000 simulations. The number of repetitions for the bootstrap confidence interval depended on the sample size. For smaller sample sizes 1000 redraws of the original data set were analyzed. For the sample size of 75 2000 repetitions were run. The number of repetitions is limited due to the long simulation results, ranging from two days for the sample size of 30 and 1000 bootstraps to 15 days for the sample size of 75 and 2000 bootstraps for all six prior believes.

The value of the survival function was then estimated for several different time points. Next to the estimated survival probability, confidence intervals and the expected coverage probability, i.e. the percentage of confidence intervals that included the true value, were calculated. Additionally, we calculated the average length of the confidence intervals. In the following tables the current simulation results are provided.

What can be observed in Table 1 is that the modified approach performs better than the original method in these settings. For both censoring rates, as occurring in the first component, the expected coverage probability increased when applying prior information. This was irrespective of the fact whether the prior knowledge was correct. For incorrect priors the average length of the confidence intervals was increased. The increase of expected coverage probability came at the cost of longer confidence intervals which was especially noticeable for the higher censoring rate of 40% in the first and 53% in the second component. Almost no difference can be detected when comparing the influence of the variance for σ_1 and σ_2 onto the outcome. Even though the numbers are the same in Table 1, tiny differences could be observed in the average length of the confidence intervals, as the confidence intervals for σ_2 are marginal slimmer.

In the next table, Table 2, the simulation results for very small sample sizes

$S(x,y)$	Cens.	no	μ_0	μ_0	μ_0	μ_1	μ_1	μ_1
x,y	Rate	prior	σ_1	σ_2	σ_3	σ_1	σ_2	σ_3
0.7506	20%	0.909	0.977	0.977	0.982	0.979	0.982	0.991
0.25, 0.45		(0.26)	(0.29)	(0.29)	(0.28)	(0.29)	(0.29)	(0.29)
	40%	0.872	0.992	0.992	0.993	0.993	0.993	0.998
		(0.29)	(0.40)	(0.40)	(0.39)	(0.40)	(0.40)	(0.42)
0.4686	20%	0.944	0.951	0.951	0.955	0.951	0.949	0.944
0.7, 0.5		(0.29)	(0.31)	(0.31)	(0.30)	(0.31)	(0.31)	(0.30)
	40%	0.931	0.984	0.986	0.989	0.984	0.984	0.979
		(0.33)	(0.47)	(0.47)	(0.46)	(0.47)	(0.47)	(0.46)
0.3329	20%	0.960	0.967	0.967	0.973	0.970	0.970	0.975
0.8, 0.8		(0.28)	(0.29)	(0.29)	(0.29)	(0.29)	(0.29)	(0.30)
	40%	0.937	0.993	0.993	0.994	0.993	0.993	0.997
		(0.33)	(0.46)	(0.46)	(0.45)	(0.46)	(0.46)	(0.46)

Table 1: Expected Coverage Probability for several different settings and sample size 50. The average length of the confidence intervals are provided in the brackets.

are provided. The data sets contained 30 observations. Just as in the case of sample size $n = 50$ the usage of prior information increased the expected coverage probability. For censoring of 20% in the first component, the average length of the confidence was comparable for the setting without the prior, i.e. the original approach, as well as the settings with the priors. In all settings the results were reasonable and showed slightly worse results than when having larger data sets. For censoring of 40%, the simulation results were not too good. Even though the expected coverage probability increased when applying prior knowledge, the resulting average confidence interval length is unreasonable high. As the survival function ranges between zero and one, an average length of 0.5 for the confidence interval is not really informative. A reasonable number of sample size should be thus provided if the censoring rate is high.

Looking at the moderate sample sizes, as given in Table 3, we observe that the original method obtained the nominal level for the sample size of 75. The application of priors again had a positive impact on the expected coverage probability. The observations as in the other sample sizes concerning the behavior for higher censoring could be noticed again. Even though an expected coverage probability of over 99%, for censoring of 40% and 53% in the two components, is incredible the cost of the high average confidence interval length makes this approach unfavorable for high censoring rates.

$S(x,y)$	Cens.	no	μ_0	μ_0	μ_0	μ_1	μ_1	μ_1
x,y	Rate	prior	σ_1	σ_2	σ_3	σ_1	σ_2	σ_3
0.7506	20%	0.934	0.993	0.993	0.994	0.992	0.991	0.982
0.25, 0.45		(0.32)	(0.38)	(0.38)	(0.37)	(0.38)	(0.39)	(0.40)
	40%	0.887	0.994	0.994	0.995	0.994	0.994	0.993
		(0.50)	(0.48)	(0.48)	(0.47)	(0.47)	(0.48)	(0.50)
0.4686	20%	0.935	0.958	0.958	0.966	0.953	0.952	0.945
0.7, 0.5		(0.36)	(0.38)	(0.38)	(0.37)	(0.38)	(0.38)	(0.37)
	40%	0.936	0.974	0.977	0.981	0.974	0.974	0.968
		(0.41)	(0.53)	(0.52)	(0.51)	(0.53)	(0.52)	(0.51)

Table 2: Expected Coverage Probability for several different settings and sample size 30. The average length of the confidence intervals are provided in the brackets.

$S(x,y)$	Cens.	no	μ_0	μ_0	μ_0	μ_1	μ_1	μ_1
x,y	Rate	prior	σ_1	σ_2	σ_3	σ_1	σ_2	σ_3
0.7506	20%	0.958	0.976	0.976	0.977	0.976	0.976	0.973
0.25, 0.45		(0.20)	(0.22)	(0.22)	(0.22)	(0.22)	(0.22)	(0.23)
	40%	0.941	0.996	0.996	0.996	0.996	0.995	0.994
		(0.22)	(0.35)	(0.35)	(0.35)	(0.35)	(0.35)	(0.36)
0.4686	20%	0.947	0.969	0.970	0.972	0.969	0.970	0.965
0.7, 0.5		(0.24)	(0.27)	(0.27)	(0.26)	(0.27)	(0.27)	(0.26)
	40%	0.947	0.998	0.998	0.999	0.998	0.998	0.997
		(0.27)	(0.45)	(0.45)	(0.44)	(0.45)	(0.45)	(0.44)

Table 3: Expected Coverage Probability for several different settings and sample size 75. The average length of the confidence intervals are provided in the brackets.

Further simulations with higher censoring rates and sample sizes of 35, 50, and 75 are omitted here. These simulations indicated yet again what the other two censoring rates already implied. The application of prior knowledge to the data sets improved the expected coverage probability in comparison to the original approach. It however came again at the cost of an average longer confidence intervals. It could further be observed that the original method of Huang and Zhao (2017) dropped well below the nominal level as the information contained in those small data sets was not sufficient to draw accurate estimators. This influenced the expected coverage probability of the adapted method as well, which stayed slightly below the nominal level.

4.4. Discussion

The second approach worked well. As no changes to the underlying model had to be conducted there occurred no problems when applying the prior likelihood to the test statistics.

The application of prior knowledge to the original approach by Huang and Zhao (2017) led to the desired results. For the same sample size an increase of expected coverage probability could be observed in the previous simulations. Further using this adaption the approach already obtained the nominal level for smaller sample sizes than required for the approach without a Bayesian prior.

It nonetheless has to be noted that the here introduced approach is not suitable for data sets with high censoring. However, this can be traced back to the already insufficient behavior of the original Lin-Ying estimator in scenarios with high censoring. Also the high run time, as well as the on average very long confidence intervals, makes the new approach unfavorable for very large sample sizes.

Overall we can still conclude that this idea was successful and represents an improvement of the original method in the so far analyzed simulations. Additional to the already conducted simulations we aim to perform further practical analysis of the behavior of the Bayesian approach. To do so we will simulate additional data with other marginal distributions. We hope that by doing so we will be able to tell the influence of heavily tailed marginals. Next to playing with the marginals we aim to analyze the influence of correlation or dependency structures to the performance of the approach. Not always will a method be able to obtain the nominal level if it fails to detect the correlation or dependence between the components correctly. Again we will let ourselves be guided by the simulation settings of Huang and Zhao (2017).

In further research we also intend to analyze the influence of the prior more detailed. We want to know if it is possible to provide a guideline on which prior distribution is most fitting depending on the underlying data. This not only includes the analysis of different variances on the outcome but further the general shape of the prior distribution. For example it could be possible that a skewed prior distribution might lead to a better performance for certain data structures than a symmetric one.

Huang and Zhao (2017) not only designed empirical likelihood confidence intervals for the Lin-Ying estimator but also for the Wang-Wells estimator. They noticed in the simulations that both tests performed similar under low censoring rates, regarding the expected coverage probability as well as the average length. For higher censoring rates the Wang-Wells estimator outperformed the Lin-Ying

estimator, even if only regarding the average length of the confidence interval in some cases. We thus aim to check if similar behavior still shows when applying Bayesian knowledge. To do so we will conduct further simulations for the Wang-Wells estimator using the same simulation settings as already used with the Lin-Ying estimator.

5. Conclusion and Outlook

During my three months research stay at *Georgia State University* we intensively conducted research on application of Bayesian statistic onto multivariate survival analysis.

In a first approach we modified an approach by Parkinson (2019). Applying a Bayesian prior onto the already existing method we hoped to decrease the required sample size. Using conditional hazards and applying constraints onto those the method by Parkinson (2019) showed overall reliable and solid results, especially in moderate sample size simulations. As the test statistic was originally based on a non-parametric model we first modified that approach to a parametric model. We then validated the mathematical derivation of the results by calculations. Further, we conducted simulations to show the moderate sample size behavior of the proposed method. Those were however not satisfying.

In a second approach we modified an approach by Huang and Zhao (2017). Again we applied a Bayesian prior onto an already existing method in the hopes of obtaining a higher expected coverage probability. Unlike in the first approach the original model did not have to be modified as instead considering the full range of the survival function the estimation was only performed for a pre-determined time point. We derived and proved the mathematical correctness of the new approach. Afterwards we conducted several simulations to confirm the validity of the found results. The current simulation results indicated that the modification of the original method improved its performance and thus should be preferred over the approach as provided by Huang and Zhao (2017).

Overall the joint research with Professor Zhao turned out other than expected but very informative. I received a great introduction to Bayesian analysis and its application to survival analysis. The profound knowledge of Professor Zhao on the general field of Bayesian statistic offered an unique opportunity to expand my knowledge on this versatile research area.

Unfortunately, the first research task turned out to be unsuccessful with regard to being able to publish the findings. The second task on the other hand, even

though the work is still ongoing, will be sufficient for a good publication in an established journal once it is finished.

In the next few weeks, additional simulations will be conducted with several different settings and sample sizes to see how the second approach performs under varying conditions. Additional to the confidence interval based on the bootstrapping we will construct confidence intervals based on a Monte Carlo Markov Chain. We will assess which of the two methods provides better simulation results to be able to provide the best possible approach.

After the simulations are finished we aim to continue working together to adopt the estimator of Wang and Wells (1997) as modified by Huang and Zhao (2017) as well. We hope that applying Bayesian prior knowledge to the test statistic will improve performance of it.

The research as financed by the Austrian Marshal Plan Foundation, Vienna, was hopefully only the beginning of a long term cooperation between the Department of Mathematics and Statistics of the *Georgia State University* and the Department of Mathematics of the *University of Salzburg*.

Additionally, the research stay at *Georgia State University* offered the opportunity to reach out to other colleagues in the USA. To this end, we managed to start a new project with the *City University of New York*. Another project with the *Naval Postgraduate School* will hopefully be launched during this year. Both will encourage the future American-Austrian academic exchange.

Acknowledgments.
This work was financially supported by a grant from the National Science Foundation, FWF I 2697, and by the Austrian Marshall Plan Foundation, Vienna (Marshallplan Jubiläumsstiftung).
I would like to thank Yichuan Zhao for inviting me to work with him at the *Georgia State University* and his valuable guidance. Additionally, I would like to thank my supervisor Professor Arne Bathke, *University of Salzburg*, for his full support before, during, and after the research stay.

References

Bathke, A., Kim, M.O., Zhou, M., 2009. Combined multiple testing by censored empirical likelihood. Journal of Statistical Planning and Inference 139, 814–827. Doi: 10.1016/j.jspi.2008.05.041.

Bayes, M., Price, M., 1763. An essay towards solving a problem in the doctrine of chances. by the late rev. mr. bayes, f. r. s. communicated by mr. price, in a letter

to john canton, a. m. f. r. s. Philosophical Transactions of the Royal Society of London 53, 370–418.

Bernardo, J.M., Smith, A.F.M., 1994. Bayesian Theory. Chichester: John Wiley & Sons.

Brendel, M., Janssen, A., Mayer, C.D., Pauly, M., 2014. Weighted logrank permutation tests for randomly right censored life science data. Scandinavian Journal of Statistics 41, 742–761. Doi: 10.1111/sjos.12059.

Cheng, Y., Zhao, Y., 2019. Bayesian jackknife empirical likelihood. Biometrika .

Clark, T.G., Bradburn, M.J., Love, S.B., Altman, D.G., 2003a. Survival analysis part i: basic concepts and first analyses. British Journal of Cancer 89, 232–238. Doi: 10.1038/sj.bjc.6601118.

Clark, T.G., Bradburn, M.J., Love, S.B., Altman, D.G., 2003b. Survival analysis part ii: multivariate data analysis – an introduction to concepts and methods. British Journal of Cancer 89, 431–436. Doi: 10.1038/sj.bjc.6601119.

Clark, T.G., Bradburn, M.J., Love, S.B., Altman, D.G., 2003c. Survival analysis part iii: multivariate data analysis–choosing a model and assessing its adequacy and fit. British Journal of Cancer 89, 605–611. Doi: 10.1038/sj.bjc.6601120.

Clark, T.G., Bradburn, M.J., Love, S.B., Altman, D.G., 2003d. Survival analysis part iv: further concepts and methods in survival analysis. British Journal of Cancer 89, 781–786. Doi: 10.1038/sj.bjc.6601117.

Czado, R., 2002. Application of Survival Analysis Methods to Long Term Care Insurance. Technical Report 268. Ludwig Maximilian University Munich, Sonderforschungsbereich 386. http://epub.ub.uni-muenchen.de/.

Dabrowska, D.M., 1988. Kaplan-meier estimate on the plane. The Annals of Statistics 16, 1475–1489. Doi: 10.1214/aos/1176351049.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B., 2013. Bayesian Data Analysis. Chapman and Hall/CRC.

Gill, R., van der Laan, M., Wellner, J., 1995. Inefficient estimators of the bivariate survival function for three models. Annales de I'I.H.P. Probabilités et Statistique 31, 545–597.

Huang, H., Zhao, Y., 2017. Empirical likelihood for the bivariate survival function under univariate censoring. Journal of Statistical Planning and Inference , doi: 10.1016/j.jspi.2017.10.002.

Kaplan, E.L., Meier, P., 1958. Individual nonparametric estimation from incomplete observations. Journal of the American Statistical Association 53, 457–481.

Klein, J.P., Moeschberger, M., 1997. Survival Analysis - Techniques for Censored and Trunctated Data. Springer, New York.

Kosorok, M.R., Lin, C.Y., 1999. The versatility of function-indexd weighted log-rank statistics. Journal of American Statistical Association 94, 330–332. Doi: 10.1080/01621459.1999.10473847.

Lazar, N.A., 2003. Bayesian empirical likelihood. Biometrika 90, 319–326.

Lin, D., Ying, Z., 1993. A simple nonparametric estimator of the bivariate survival function under univariate censoring. Biometrika 80, 573–581.

Nelson, W., 1969. Hazard plotting for incomplete failure data. Journal of Quality Technology 1, 27–52.

Owen, A.B., 2001. Empirical Likelihood. Chapman and Hall/CRC Monographs on Statistics and Applied Probability, Chapman and Hall/CRC.

Parkinson, J.H., 2019. Combined multiple testing of multivariate survival times by censored empirical likelihood. Scandinavian Journal of Statistics .

Perrigot, R., Cliquet, G., Mesbah, M., 2004. Possible applications of survival-analysis in franchising research. Int. Rev. of Retail, Distribution and Consumer Research 14, 129–143.

Scherm, H., Ojiambo, P., 2004. Applications of survival analysis in botanical epidemiology. Phytopathology 94.

Wang, W., Wells, M.T., 1997. Nonparametric estimators of the bivariate survival function under simplified censoring conditions. Biometrika 84, 863–880. Doi: 10.1093/biomet/84.4.863.

Wang, W., Zafra, P., 2009. Estimating bivariate survival function by volterra estimator using dynamic programming techniques. Journal of Data Science 7, 365–380.

Wei, L., Lachin, J., 1984. Two-sample asymptotically distribution-free tests for incomplete multivariate observations. Journal of American Statistical Association 79, 653–661. Doi: 10.1080/01621459.1984.10478093.

von Weizsäcker, C.F., 1992. Zeit und Wissen. Carl Hanser.