Two-sample location-scale estimation from censored data

Rianka Bhattacharya and Sundarraman Subramanian
Center for Applied Mathematics and Statistics
Department of Mathematical Sciences
New Jersey Institute of Technology
USA

Abstract
When two survival functions belong to a location-scale family of distributions, and the available two-sample data are each right censored, the location and scale parameters can be estimated using a minimum distance criterion combined with Kaplan–Meier quantiles. In this paper, it is shown that using the estimated quantiles from a semiparametric random censorship framework produces improved parameter estimates. The semiparametric framework was originally proposed for the one-sample case (Dikta, 1998), and uses a model for the conditional probability that an observation is uncensored given the observed minimum. The extension to the two-sample setting assumes the availability of good fitting models for the group-specific conditional probabilities. When the models are correctly specified for each group, the new location and scale estimators are shown to be asymptotically as or more efficient than the estimators obtained using the Kaplan–Meier based quantiles. Individual and joint confidence intervals for the parameters are developed. Simulation studies show that the proposed method produces confidence intervals that have correct empirical coverage and that are more informative. The proposed method is illustrated using two real data sets.

KEY WORDS: Censoring rate; Cauchy link; Empirical coverage probability; Functional delta method; Gaussian process; Power function.

1 Introduction

In statistical analysis of survival data, it is often of interest to determine the difference, if any, between two treatment effects. When the treatment-specific populations are known to be each normally distributed and when the two-sample data are each completely uncensored, the standard \( t \)-test can be used to discriminate between the treatments. But the normal family is only one member of the location-scale family of distributions, among several others, and the \( t \)-test would be inadequate for non-normal families. There may be the additional complication due to censoring which the \( t \)-test was not designed to handle. On the other hand, when the underlying distributions are unknown, the Wilcoxon test is used. However, it must be noted that, when the samples come from a location-scale family of distributions, an inferential method that also incorporates the available model information into the analysis should perform better. Indeed, when distributional difference in location and scale is suspected, the Generalized Wilcoxon test for detecting differences is inadequate, see page 211 of Kalbfleisch and Prentice (1980).
INTRODUCTION

For the general two-sample problem, the group-specific empirical distribution functions, or their Kaplan–Meier (KM) counterparts in the case of the random censorship model (RCM), provide the basic resource for inference. For two-sample location-scale, Zhang and Li (1996) suggested a heuristic method using simple quantiles for inference. Note that in this setting the two continuous distribution functions $F_1$ and $F_2$ are related through the equation $F_1(t) = F_2(a + bt)$, where $a \in \mathbb{R}$ and $b > 0$.

For RCM two-sample location-scale inference, the standard method is to employ a minimum distance criterion, more specifically, the Cramér-von Mises type discrepancy involving either the KM estimators of the survival functions, $S_i(t) = 1 - F_i(t)$, $i = 1, 2$, or their quantiles. Hsieh (1996), however, constructed a regression setup that was based on the KM quantile process and showed that his generalized least squares estimator is semiparametric efficient; see also Hsieh (1995) for the uncensored case. Some limitations, including practical utility of Hsieh’s estimator, are pointed out by Potgieter and Lombard (2012), however. Koul and Yang (1989) applied the Cramér-von Mises type discrepancy to the two KM estimators but focused only on the two-sample scale model; the extension to the location-scale model can be complicated. The Cramér-von Mises type discrepancy combined with quantiles is seen to be a very convenient method for estimating the model parameters $a$ and $b$, as evidenced by the fact that, under the location-scale model assumption, the quantile functions for the two groups at each point $t \in (0, 1)$ are linearly related (Parzen, 1979; Zhang and Yu, 2002; Potgieter and Lombard, 2012); that is, $Q_2(t) = a + bQ_1(t)$, so that minimizing $\hat{S}(a, b)$, an estimate of $S(a, b) = \int \{Q_2(s) - a - bQ_1(s)\}^2 dG(s) := E \left( \{Q_2 - a - bQ_1\}^2, G \right)$(1.1), where $G(s)$ is a positive measure on $(0, 1)$, presents a viable option, unlike the approach founded on the KM estimators of $S_i(t)$, $i = 1, 2$. In Eq. (1.1), $E(\cdot, \cdot)$ denotes the integral of the first argument with respect to the second argument. Zhang and Yu (2002) developed estimation of $\theta = (a, b)'$ using Eq. (1.1), where they plugged in the KM quantile function estimators to obtain $\hat{S}(a, b)$, which they minimized to yield $\hat{\theta}_n$, their estimator of $\theta$. Under some regularity conditions, Zhang and Yu (2002) derived the large sample distribution of $\hat{\theta}_n$ via the delta method combined with standard large sample theory for weighted KM statistics.

In this paper, we showcase the efficacy of utilizing alternate quantiles, obtained from semiparametric survival function estimators, for two-sample location-scale inference. These survival function estimators arise from the framework of semiparametric random censorship models, SRCMs henceforth, introduced by Dikta (1998). For the $i$th group, $i = 1, 2$, start with the parametric model $m_i(t, \gamma_i)$ for the conditional probability that an observation is uncensored given the observed minimum, and use the $i$th group sample data to obtain $\hat{\gamma}_i$, the maximum likelihood estimator (MLE) of $\gamma_i \in \Gamma_i \subset \mathbb{R}^k$. The estimated $m_i$ is used as a “surrogate” for the censoring indicator in the $i$th group, leading to a semiparametric estimator of the group-specific subdistribution function corresponding to uncensored failures. Plugging in this last estimator and the usual “at-risk” function into a standard sequence of mappings (Gill and Johansen, 1990) yields the group-specific SRCM-based survival function estimator, see section 2.1 for details.
INTRODUCTION

There are compelling reasons why it would be desirable to incorporate SRCMs into the two-sample location-scale analysis. The KM estimator, although the primary choice under the RCM, ceases to provide optimal performance under the framework of SRCMs. Specifically, Dikta (1998) proved that, under correct model specification, the asymptotic variance of the SRCM-based survival function estimator is no greater than that of the KM estimator, with equality attained only in rare and unrealistic cases. In fact, from a more general result derived by Dikta (2014) recently, it is now evident that the SRCM-based survival function estimator is asymptotically efficient with respect to the class of all regular estimators under the SRCM. It stands to reason that, if proper parametric models can be identified for each group-specific conditional probability function, inference for the censored two-sample location-scale problem can be improved. Note that the function to be estimated from binary response data is a “success probability” function, for which models are readily available. In reality, fitting the standard Cauchy model to estimate this function appears to produce better estimates most of the times, see Subramanian (2012) and Mondal and Subramanian (2014); see section 5 for discussion about various possible models for analyzing the binary response data. Furthermore, when the censoring is rather heavy, the KM estimator has fewer jumps leading to a patchy result, which is not a problem with an SRCM-incorporated analysis. Finally, when the censoring indicators are missing at random for a subset of the study subjects (Subramanian, 2004), RCM-based inference for the location-scale problem becomes difficult compared with our proposed SRCM-based inference. Indeed, with minor modifications, the SRCMs approach readily applies to the case of missing censoring indicators, see section 5 for some discussion.

We propose to plug in the SRCM-based quantiles into Eq. (1.1) and obtain $\hat{\theta}$ as the minimizer of the resulting criterion function. Under mostly the same regularity conditions as in Zhang and Yu (2002) we derive the limiting distribution of $\hat{\theta}$, from which we are able to obtain confidence intervals for $a$, $b$ or any function of the two parameters thereof. Numerical results reported in section 3 indicate that fitting the standard probit or Cauchy link for the binary response data produces estimators with approximately correct coverage and relative reduction over the estimators based on KM quantiles amounting to between 5% and 15%. The Cauchy fit provides the overall coverage closest to the nominal value. A power study confirms the superiority of SRCMs over RCMs. A theoretical analysis of the asymptotic variances reinforces the numerical evidence. Specifically, when the models are correctly specified, we show that the proposed estimators are asymptotically as or more efficient than Zhang and Yu’s (2002) estimators. Thus, there appears to be strong theoretical and numerical support for SRCMs to be incorporated into censored two-sample location-scale analysis.

The paper is organized as follows. In section 2, we review the SRCMs and then present our proposed approach. In section 3, we present our simulation results as well as our analysis of the mouse leukemia data. In section 4, we illustrate our method using a mouse leukemia data set (Kalbfleisch and Prentice, 1980) and an acute myelogenous data set (Klein and Moeschberger, 2005). In section 5, we give some concluding remarks.
2 Semiparametric location-scale estimators

Henceforth, the subscript $i$ will denote the group indicator, where $i = 1, 2$. Let $Y_i$ be independent random variables with distribution functions $F_i$ and density functions $f_i$. As discussed in the introduction section, $F_2(t) = F_1((t - a)/b)$ for all $t$, and some fixed constants $a \in \mathbb{R}^1$ and $b > 0$. Equivalently, $Y_2$ and $a + bY_1$ have the same distribution. We denote the quantile functions by $Q_i(s)$, where $Q_i(s) = \sup\{t : F_i(t) \leq s\}$. Let $Z_i$ be independent censoring variables. We observe $n_i$ copies of $(X_i, \delta_i) = (\min\{Y_i, Z_i\}, I(Y_i \leq Z_i))$, where $I(\cdot)$ is an indicator function, and $n_i$ are the group sample sizes. Let $H_i$ denote the distribution functions of $X_i$ and let $\hat{H}_i(t)$ denote their empirical estimators. Also, let $m_i(x, \gamma_{i0}) = E(\delta_i|X_i = x)$.

2.1 Review of SRCMs

In the framework of Dikta’s SRCMs, a parametric model is fitted to estimate $m_i(x, \gamma_{i0}) = E(\delta_i|X_i = x)$, the conditional expectation of the censoring indicator given the observed minimum. Here, $m_i(x, \gamma_i)$ is a known function with unknown $\gamma_i \in \Gamma_i \subset \mathbb{R}^k$. As in Subramanian (2009) or Subramanian and Zhang (2013), and as will be seen in section (2.2) as well, it will be most convenient to describe $\hat{S}_i(t) = 1 - \hat{F}_i(t)$, Dikta’s (1998) group-specific semiparametric survival function estimators, in Gill and Johansen’s (1990) formulation.

Write $H_i^{(1)}(t) = E(m_i I_{[0,t]}, H_i) \equiv P(X_i \leq t, \delta_i = 1)$, with $\hat{H}_i^{(1)}(t)$, its estimator, being obtained by replacing $m_i(s, \gamma_{i0})$ and $H_i(s)$ with the estimates $m_i(s, \hat{\gamma}_i)$ and $\hat{H}_i(s)$ respectively. Writing $\hat{R}_i = 1 - \hat{H}_i(t)$, an estimator of $R_i = 1 - H_i(t)$, the SRCM-based estimator of $\Lambda_i(t)$, the group-specific cumulative hazard, denoted by $\hat{\Lambda}_i(t)$, is obtained via the following sequence of mappings, see Gill and Johansen (1990) or Subramanian (2009):

\[
(\hat{H}_i^{(1)}, \hat{R}_i) \rightarrow \left(\frac{\hat{H}_i^{(1)}}{\hat{R}_i}, \frac{1}{\hat{R}_i}\right) \rightarrow \int_{[0,\cdot]} \frac{1}{\hat{R}_i} d\hat{H}_i^{(1)} \equiv \hat{\Lambda}_i, i = 1, 2. \tag{2.1}
\]

The group-specific survival function estimator $\hat{S}_i(t)$ is computed via the product integral mapping, given by

\[
\prod_{[0,\cdot]} \left(1 - \frac{1}{\hat{R}_i} d\hat{H}_i^{(1)}\right) \equiv \prod_{[0,\cdot]} \left(1 - d\hat{\Lambda}_i\right) \equiv \hat{S}_i, \quad i = 1, 2. \tag{2.2}
\]

By Theorem II.6.2 of Andersen et al. (1993), also known as the Duhamel equation, we have

\[\hat{F}_i(u) - F_i(u) = \hat{S}_i(u) \int_0^u \frac{\hat{S}_i(v-)}{S_i(v)} d\left(\hat{\Lambda}_i(v) - \Lambda_i(v)\right).\]

By Theorem 2.4 of Dikta (1998), $\hat{\Lambda}_i(t)$ and $\hat{S}_i(t)$ are each uniformly strongly consistent on the interval $[0, \tilde{\tau}_i]$, where $\tilde{\tau}_i < \tau_{H_i}$ and $\tau_{H_i}$ denotes the right end-point of the support of $H_i$. This means that when centered processes have estimates that are multipliers, such as, for example, $\hat{S}_i(s)$, those may be replaced by the respective true quantities, with remainder term equal to $o_p(n^{-1/2})$, uniformly over $[0, \tilde{\tau}_i]$. Therefore,

\[\hat{F}_i(u) - F_i(u) = S(u) \left(\hat{\Lambda}_i(u) - \Lambda(u)\right) + o_p(n^{-1/2}), \tag{2.3}\]
uniformly over $[0, \tau_i]$, and it suffices to focus on $n^{1/2} \left( \hat{L}_i - L_i \right)$, the normalized cumulative hazard process.

When the model for $m_i(t)$ is correctly specified, the weak convergence of the basic centered bivariate process $n^{1/2} \left( \hat{H}^{(1)}_i(t) - H^{(1)}_i(t), \hat{R}_i(t) - R_i(t) \right)$ to the zero-mean bivariate Gaussian process $(Z^{(1)}_{H}, Z_{R_i})$, combined with compact differentiability of the sequence of mappings in (2.1) and Gill and Johansen’s (1990) functional delta method, yields the weak convergence in $D[0, \tau_i]$ of $n^{1/2} \left( \hat{L}_i(t) - L_i(t) \right)$ to a zero-mean Gaussian process. Specifically, analogous to page 1537 of Gill and Johansen (1990), as $n \to \infty$, and, in $D[0, \tau_i]$,

$$n^{1/2} \left( \hat{L}_i - L_i \right) \xrightarrow{D} \int_{[0,1]} \frac{1}{R_i(u)} d\mathbb{W}_i, \quad i = 1, 2,$$

where $\mathbb{W}_i = Z^{(1)}_{H} - \int_{[0,1]} Z_{R_i} \, d\Lambda_i$. To specify the limiting covariance function, for each $r = 1, \ldots, k$, let $D_r(m_i(t, \gamma_i))$ denote the partial derivative of $m_i(t, \gamma_i)$ with respect to $\gamma_i$. Write $\text{Grad}(m_i(t, \gamma_i)) = [D_1(t, \gamma_i), \ldots, D_k(t, \gamma_i)]^T$, and let $I(\gamma_{io})$ denote the Fisher information with $(r, s)$ element given by

$$I_{r,s}(\gamma_{io}) = E \left( \frac{D_r(m_i(X_i, \gamma_{io})) D_s(m_i(X_i, \gamma_{io}))}{m_i(X_i, \gamma_{io})(1 - m_i(X_i, \gamma_{io}))} \right), \quad i = 1, 2. \tag{2.5}$$

Note that, for $i = 1, 2$, the $n_i = 1$ counterpart of $\mathbb{W}_i(t)$ is given by [see Dikta (1998) or Subramanian (2004)]

$$\frac{\delta_i - m_i(X_i, \gamma_{io})}{m_i(X_i, \gamma_{io})(1 - m_i(X_i, \gamma_{io}))} \int_0^t \alpha_i(u, \gamma_i) \, dH_i(u) + \int_0^t m_i(u, \gamma_{io}) \, dI(X_i \leq u) - \int_0^t I(X_i \geq u) \, d\Lambda_i(u), \tag{2.6}$$

where $\alpha_i(u, v) = (\text{Grad}(m_i(u, \gamma_{io})))^T I^{-1}(\gamma_{io}) \text{Grad}(m_i(v, \gamma_{io}))$. For $0 \leq t_1 \leq t_2 \leq \tau_i$, direct computation of the covariance function of the right hand side (RHS) of Eq. (2.4), using Eq. (2.6) for $\mathbb{W}_i(t)$, leads to

$$V_i(t_1, t_2) = \int_0^{t_1} \frac{m_i^2(s, \gamma_{io})}{R_i^2(s)} \, dH_i(s) + \int_0^{t_1} \int_0^{t_2} \alpha_i(u, v) \frac{R_i(u) R_i(v)}{R_i^2(s)} \, dH_i(u) \, dH_i(v), \quad i = 1, 2. \tag{2.7}$$

When $m_i(t)$ is correctly specified, $V_i(t, t)$ is no greater than the asymptotic variance of the Nelson–Aalen estimator of $L_i(t)$ (Dikta, 1998). In the next section, it will be demonstrated that, under correct specification of $m_i(t)$, this efficiency of the SRCMs leads to improved location and scale parameter estimates.

### 2.2 Estimator and large-sample distribution

For $0 < p < 1$, the group-specific quantile function estimator is defined by $\hat{Q}_i(p) = \sup \{ t : \hat{F}_i(t) \leq p \}$. By Proposition II.8.4 of Andersen et al. (1993) and the functional delta method, we have

$$\hat{Q}_i(p) - Q_i(p) = -\frac{\hat{F}_i(F_i^{-1}(p)) - F_i(F_i^{-1}(p))}{f_i(F_i^{-1}(p))} + o_p(n^{-1/2}), \quad i = 1, 2. \tag{2.8}$$

The weak convergence of $\hat{Q}_i - Q_i$ in $D[p_0, p_1]$, where $0 < p_0 < p_1 < 1$, can be deduced from Eq. (2.8) and the weak convergence of $\hat{F}_i - F_i$, see section 2.1. The estimators $\hat{Q}_1$ and $\hat{Q}_2$ are plugged into Eq. (1.1) to obtain $\hat{S}(a, b)$, minimizing which yields the set of linear equations expressed in matrix notation as $A\theta = d$, where

$$\theta = (a, b)^T, \quad d = (d_1, d_2)^T$$

with $d_j = E(\hat{Q}_i^{-1}(\hat{Q}_2^2, G), j = 1, 2$, and $A = (A_{ij})_{2 \times 2}$ with

$$A_{ij} = E \left( \hat{Q}_i^{-1} \hat{Q}_2^{-1} \right) := \int \hat{Q}_i^{-1} \hat{Q}_2^{-1} \, dG, \quad i = 1, 2, \quad j = 1, 2.$$
Let $\hat{\theta}$ denote the minimum distance estimator (MDE) of $\theta$. Assuming that $\tilde{Q}_1$ is not a constant, $A$ is non-singular by the Cauchy-Schwarz inequality. Then $\hat{\theta} = A^{-1}d$.

Let $\tau_{i,l}$ and $\tau_{i,u}$ denote the infimum and supremum of the support set of the measure induced by $G \circ F_i$, where $\circ$ denotes composition. Some regularity conditions given by Zhang and Yu (2002) will be needed to prove the main theorem. The conditions ensure that (i) the entries of $A$ and $d$ are finite numbers; (ii) the densities $f_i$ do not vanish over the specified range of integration; (iii) remainder terms are $o_p(1)$ uniformly over the support set of the measure $dG$; (iv) the elements of $\Sigma_{SRM}$, defined below, are finite.

A1 For each $i = 1, 2$, the quantities $\tau_{i,u}$ satisfy $\tau_{i,u} < \tau_H$.

A2 The cumulative distribution function (CDF) $F_i$ is strictly increasing on $[0, \tau_{i,u}]$.

A3 The CDF $F_i$ is absolutely continuous with density $f_i$, $i = 1, 2$.

A4 For some $\epsilon > 0$, $E \{1(x : f_1(x) < \epsilon \text{ or } f_2(x) < \epsilon), G(F_1(x)) + G(F_2(x))\} = 0$.

Note that A1 implies assumption A4 of Zhang and Yu (2002).

To precisely describe the large sample distribution of $\hat{\theta}$, we define

$$e_i(s) = E(S_i I(s, \tau_{i,u})/f_i G \circ F_i), \quad i = 1, 2,$$

$$\tilde{e}_{ij}(s) = E((Q_{ij} \circ F_i) S_i I(s, \tau_{i,u})/f_i G \circ F_i), \quad i = 1, 2, \quad j = 1, 2,$$

and further define $\Sigma_{SRM} = (\sigma_{rs})_{4 \times 4}$, where $\sigma_{13} = \sigma_{23} = 0$ and

$$\sigma_{11} = \frac{1}{n_1} \left[ \int_{\tau_{1,l}}^{\tau_{1,u}} \frac{m_1^2(s, \gamma_{10})}{R_1(s)} e_1^2(s) \, dH_1(s) + \int_{\tau_{1,l}}^{\tau_{1,u}} \frac{\alpha_1(s, t)}{R_1(s) R_1(t)} e_1(s) e_1(t) \, dH_1(s) \, dH_1(t) \right],$$

$$\sigma_{21} = \frac{2}{n_1} \left[ \int_{\tau_{1,l}}^{\tau_{1,u}} \frac{m_1^2(s, \gamma_{10})}{R_1(s)} e_1(s) \tilde{e}_{11}(s) \, dH_1(s) + \int_{\tau_{1,l}}^{\tau_{1,u}} \frac{\alpha_1(s, t)}{R_1(s) R_1(t)} e_1(s) \tilde{e}_{11}(t) \, dH_1(s) \, dH_1(t) \right],$$

$$\sigma_{22} = \frac{4}{n_1} \left[ \int_{\tau_{1,l}}^{\tau_{1,u}} \frac{m_1^2(s, \gamma_{10})}{R_1(s)} \tilde{e}_{11}^2(s) \, dH_1(s) + \int_{\tau_{1,l}}^{\tau_{1,u}} \frac{\alpha_1(s, t)}{R_1(s) R_1(t)} \tilde{e}_{11}(s) \tilde{e}_{11}(t) \, dH_1(s) \, dH_1(t) \right],$$

$$\sigma_{33} = \frac{1}{n_2} \left[ \int_{\tau_{2,l}}^{\tau_{2,u}} \frac{m_2^2(s, \gamma_{20})}{R_2(s)} e_2^2(s) \, dH_2(s) + \int_{\tau_{2,l}}^{\tau_{2,u}} \frac{\alpha_2(s, t)}{R_2(s) R_2(t)} e_2(s) e_2(t) \, dH_2(s) \, dH_2(t) \right],$$

$$\sigma_{41} = \frac{1}{n_1} \left[ \int_{\tau_{1,l}}^{\tau_{1,u}} \frac{m_1^2(s, \gamma_{10})}{R_1(s)} e_1(s) \tilde{e}_{12}(s) \, dH_1(s) + \int_{\tau_{1,l}}^{\tau_{1,u}} \frac{\alpha_1(s, t)}{R_1(s) R_1(t)} e_1(s) \tilde{e}_{12}(t) \, dH_1(s) \, dH_1(t) \right],$$

$$\sigma_{42} = \frac{2}{n_1} \left[ \int_{\tau_{1,l}}^{\tau_{1,u}} \frac{m_1^2(s, \gamma_{10})}{R_1(s)} \tilde{e}_{12}(s) \tilde{e}_{12}(t) \, dH_1(s) + \int_{\tau_{1,l}}^{\tau_{1,u}} \frac{\alpha_1(s, t)}{R_1(s) R_1(t)} \tilde{e}_{12}(s) \tilde{e}_{12}(t) \, dH_1(s) \, dH_1(t) \right],$$

$$\sigma_{43} = \frac{1}{n_2} \left[ \int_{\tau_{2,l}}^{\tau_{2,u}} \frac{m_2^2(s, \gamma_{20})}{R_2(s)} e_2(s) \tilde{e}_{21}(s) \, dH_2(s) + \int_{\tau_{2,l}}^{\tau_{2,u}} \frac{\alpha_2(s, t)}{R_2(s) R_2(t)} e_2(s) \tilde{e}_{21}(t) \, dH_2(s) \, dH_2(t) \right],$$

$$\sigma_{44} = \frac{1}{n_1} \left[ \int_{\tau_{1,l}}^{\tau_{1,u}} \frac{m_1^2(s, \gamma_{10})}{R_1(s)} \tilde{e}_{12}^2(s) \, dH_1(s) + \int_{\tau_{1,l}}^{\tau_{1,u}} \frac{\alpha_1(s, t)}{R_1(s) R_1(t)} \tilde{e}_{12}(s) \tilde{e}_{12}(t) \, dH_1(s) \, dH_1(t) \right],$$

$$\sigma_{44} = \frac{1}{n_2} \left[ \int_{\tau_{2,l}}^{\tau_{2,u}} \frac{m_2^2(s, \gamma_{20})}{R_2(s)} \tilde{e}_{21}^2(s) \, dH_2(s) + \int_{\tau_{2,l}}^{\tau_{2,u}} \frac{\alpha_2(s, t)}{R_2(s) R_2(t)} \tilde{e}_{21}(s) \tilde{e}_{21}(t) \, dH_2(s) \, dH_2(t) \right].$$

See Zhang and Yu (2002) for the corresponding RCM version, denoted here by $\Sigma_{RCM}$. Writing $n = n_1 + n_2$. 6
let \( \lim_{n \to \infty} n_1/n = \nu \in (0,1) \). Theorem 1 gives the large sample distribution of \( \hat{\theta} \).

**Theorem 1** Suppose that conditions A1–A4 hold. Then, \( n^{1/2}(\hat{\theta} - \theta_0) \) converges in distribution to a bivariate normal distribution with mean vector \( \mathbf{0} \) and covariance matrix \( \Sigma = nD^T \Sigma_{\text{SRCM}} D \). Furthermore, the SRCMs two-sample location-scale estimation is as or more efficient than its RCM counterpart, in the sense that

\[
\eta^T (\Sigma_{\text{RCM}} - \Sigma_{\text{SRCM}}) \eta \geq 0,
\]

where \( \eta = (\eta_1, \eta_2, \eta_3, \eta_4)^T \).

**Proof** As in Zhang and Yu (2002), \( \hat{\theta} \) is a function of the vector

\[
U = \left( E \left( Q_1, G \right), E \left( Q_1^2, G \right), E \left( Q_2, G \right), E \left( Q_1Q_2, G \right) \right)^T \equiv (U_1, U_2, U_3, U_4)^T.
\]

Writing \( c = E(1, G) \), it follows that \( \hat{\theta} = g(U) \equiv (g_1, g_2)^T \), where \( u = (u_1, u_2, u_3, u_4)^T \), and \( g(u) = (g_1, g_2)^T \) with

\[
g_1 = \frac{u_2u_3 - u_1u_4}{cu_2 - u_1^2}, \quad g_2 = \frac{cu_4 - u_1u_3}{cu_2 - u_1^2}.
\]

Write \( U = (U_1, U_2, U_3, U_4)^T \), where \( U_j \) is obtained from \( U \) by replacing the quantile function estimate with its true value. Apply Proposition II.8.4 of Andersen et al. (1993), the chain rule of differentiation, and the functional delta method to conclude that \( n^{1/2} \left( \hat{U} - U \right) \) has the same asymptotic distribution as

\[
n^{1/2} \int \begin{pmatrix} (G \circ F_1)' \cdot 1/f_1 & 0 & 0 \\
(G \circ F_1)' \cdot 2(Q_1 \circ F_1)/f_1 & 0 & (G \circ F_2)' \cdot 1/f_2 \\
(G \circ F_1)' \cdot (Q_2 \circ F_1)/f_1 & (G \circ F_2)' \cdot (Q_2 \circ F_2)/f_2 & \end{pmatrix} \begin{pmatrix} \hat{F}_1(u) - F_1(u) \\
\hat{F}_2(u) - F_2(u) \end{pmatrix} \ du. \tag{2.9}
\]

Now let

\[
W(u) = \begin{pmatrix}
E \left( S_1 I_{(u, \tau_{1,u})} / f_1, G \circ F_1 \right) & 0 & 2E \left( (Q_1 \circ F_1) S_1 I_{(u, \tau_{1,u})} / f_1, G \circ F_1 \right) \\
2E \left( (Q_1 \circ F_1) S_1 I_{(u, \tau_{1,u})} / f_1, G \circ F_1 \right) & 0 & E \left( S_2 I_{(u, \tau_{2,u})} / f_2, G \circ F_2 \right) \\
E \left( (Q_2 \circ F_1) S_1 I_{(u, \tau_{2,u})} / f_1, G \circ F_1 \right) & E \left( (Q_1 \circ F_2) S_2 I_{(u, \tau_{2,u})} / f_2, G \circ F_2 \right) & e_2(u) \\
e_1(u) & 0 & \tilde{e}_{11}(u) \\
2\tilde{e}_{11}(u) & 0 & e_2(u) \\
\tilde{e}_{12}(u) & \tilde{e}_{21}(u) & \end{pmatrix},
\]

Plugging (2.3) into (2.9) and interchanging the order of integration, it follows that

\[
\hat{U} - U = \int W(u) d \begin{pmatrix} \hat{\Lambda}_1(u) - \Lambda_1(u) \\
\hat{\Lambda}_2(u) - \Lambda_2(u) \end{pmatrix} + o_p(n^{-1/2}). \tag{2.10}
\]

Eq. (2.10) shows that \( n^{1/2}(\hat{U} - U) \) is asymptotically equivalent to a weighted-semiparametric-cumulative-hazards statistic. The asymptotic distribution of the vector \( n^{1/2}(\hat{U} - U) \) can be deduced from the weak convergence of \( n^{1/2}(\hat{\Lambda}_i - \Lambda_i), i = 1,2 \), which is, of course, well established, see section 2.1. That is, \( n^{1/2}(\hat{U} - U) \) converges in distribution to a multivariate normal with mean vector \( \mathbf{0} \) and covariance matrix approximated by \( \Sigma_{\text{SRCM}} \). We show an illustrative derivation by applying the functional delta method to the first component

\[
\hat{U}_1 - U_1 = \int_{\tau_{1,u}}^{\tau_{1,u}} e_1(u) d \left( \hat{\Lambda}_1(u) - \Lambda_1(u) \right) + o_p(n^{-1/2}).
\]
Indeed, assuming continuity of $S_i$, the method of proof given on page 1537 of Gill and Johansen (1990) can be followed exactly, permitting one to conclude that, as $n \to \infty$,

$$n^{1/2} \left( \hat{U}_i - U_i \right) \overset{D}{\to} \int_{\tau_1}^{\tau_{1:n}} e_1(u) \frac{dZ_{H_1}(1) - Z_{R_1} d\Lambda_1}{R_1(u)} \equiv \int_{\tau_1}^{\tau_{1:n}} e_1(u) \frac{d\mathbb{W}_1}{R_1(u)}.$$

Direct computation of the covariance function of the RHS of the above equation, using Eq. (2.6) for $\mathbb{W}_i(t)$, gives $\sigma_{11}$. Other elements of $\Sigma_{RCM}$ can be obtained in a similar way.

The asymptotic distribution of $n^{1/2}(\hat{\theta} - \theta_0)$ follows exactly as in Zhang and Yu (2002). Specifically,

$$\hat{\theta} - \theta_0 = \left( \frac{dg}{du} \right)^T \Bigr|_{u=U} (\hat{U} - U) + o_p(n^{-1/2}),$$

where

$$\left( \frac{dg}{du} \right)^t = \begin{pmatrix} \frac{dg_1}{du_1} & \frac{dg_1}{du_2} & \frac{dg_1}{du_3} & \frac{dg_1}{du_4} \\ \frac{dg_2}{du_1} & \frac{dg_2}{du_2} & \frac{dg_2}{du_3} & \frac{dg_2}{du_4} \end{pmatrix} 1 \leq i, j \leq 4.$$

Writing $D = \frac{dg}{du} |_{u=U}$, it follows that $n^{1/2}(\hat{\theta} - \theta_0)$ converges in distribution to a zero-mean bivariate normal distribution with covariance matrix $\Sigma = nD^T \Sigma_{RCM} D$.

To prove efficiency, let $\rho_1(s) = \eta_1 + 2\eta_2 \bar{e}_{11}(s) + \eta_3 \bar{e}_{12}(s)$ and $\rho_2(s) = \eta_1 + \eta_4 \bar{e}_{21}(s)$. Standard calculations yield

$$\eta^T \Sigma_{RCM} \eta = \int_{\tau_1}^{\tau_{1:n}} \frac{m_1^2(s, \gamma_{10})}{R_1^2(s)} \rho_1^2(s) dH_1(s) + \int_{\tau_1}^{\tau_{1:n}} \alpha_1(s, t) \rho_1(s) \rho_1(t) dH_1(s) dH_1(t) + \int_{\tau_1}^{\tau_{1:n}} \frac{m_2^2(s, \gamma_{20})}{R_2^2(s)} \rho_2^2(s) dH_2(s) + \int_{\tau_1}^{\tau_{1:n}} \alpha_2(s, t) \rho_2(s) \rho_2(t) dH_2(s) dH_2(t).$$

For RCM-based two-sample location-scale estimation (Zhang and Yu, 2002), it is seen that

$$\eta^T \Sigma_{RCM} \eta = \int_{\tau_1}^{\tau_{1:n}} \frac{1}{R_1^2(s)} \rho_1^2(s) dH_1^{(1)}(s) + \int_{\tau_1}^{\tau_{1:n}} \frac{1}{R_2^2(s)} \rho_2^2(s) dH_2^{(1)}(s).$$

Since $m_i^2(s, \gamma_{i0}) dH_i(s) = m_i(s, \gamma_{i0}) dH_i^{(1)}(s)$, it is seen that

$$\eta^T \left( \Sigma_{RCM} - \Sigma_{RCM} \right) \eta = \int_{\tau_1}^{\tau_{1:n}} \frac{1 - m_1(s, \gamma_{10})}{R_1^2(s)} \rho_1^2(s) dH_1^{(1)}(s) + \int_{\tau_1}^{\tau_{1:n}} \frac{1 - m_2(s, \gamma_{20})}{R_2^2(s)} \rho_2^2(s) dH_2^{(1)}(s).$$

Since $m_i^2(s, \gamma_{i0}) dH_i(s) = m_i(s, \gamma_{i0}) dH_i^{(1)}(s)$, it is seen that

$$\eta^T (\Sigma_{RCM} - \Sigma_{RCM}) \eta = B_1 - B_2 + B_3 - B_4.$$
To show that \( \eta^T (\Sigma_{RCM} - \Sigma_{SRCM}) \eta \geq 0 \), it suffices to show that \( B_{2i-1} - B_{2i} \geq 0, \ i = 1, 2 \). Note that

\[
B_{2i} = \langle \beta_i, \varphi^{-1}(\gamma_{\omega}) \beta_i \rangle, \ i = 1, 2,
\]

where \( \beta_i = (\beta_{i1}, \ldots, \beta_{ik})^T \) and

\[
\beta_{ir} = \mathbb{E} \left( \frac{D_r(m_i(X_i, \gamma_{\omega})))}{R_i(X_i)} \rho_i(X_i) 1_{[\tau_{i,1} \leq X_i, X_i, \leq \tau_{i,2}]} \right), \ r = 1, \ldots, k.
\]

Let \( h = (h_1, \ldots, h_k)^T \in \mathbb{R}^k \setminus \{0\} \). Since \( I^{-1}(\gamma_{\omega}), i = 1, 2 \), are each positive definite, by a well-known argument (cf. 1f1.1, Rao, 1973) it follows that

\[
\sup_{h \in \mathbb{R}^k \setminus \{0\}} \frac{\langle h, \beta_i \rangle^2}{\langle h, I(\gamma_{\omega}) h \rangle} = \langle \beta_i, \varphi^{-1}(\gamma_{\omega}) \beta_i \rangle \equiv B_{2i}, \ i = 1, 2.
\]  

(2.13)

From Eq. (2.13) it suffices to show for an arbitrary \( h \in \mathbb{R}^k \setminus \{0\} \) that

\[
\frac{\langle h, \beta_i \rangle^2}{\langle h, I(\gamma_{\omega}) h \rangle} \leq B_{2i-1}, \ i = 1, 2,
\]

which follows exactly as in the concluding part of the proof of Corollary 2.7 of Dikta (1998). □

**Remark 1** The methods of Dikta, Ghorai, and Schmidt (2005) can also be employed to derive the asymptotic distribution of \( \hat{\Theta} \). Here, we have used Gill and Johansen’s (1990) functional delta method.

**Remark 2** As pointed out by Zhang and Yu (2002), for a particular choice of \( G \), the estimator of the density \( f_i \) would not be needed, which makes computing the elements of the matrix \( \Sigma_{SRCM} \) easier. The choice \( \frac{dG(F)}{dF} \propto 1(F \in [p_0, p_1]) \), where \( p_0, 1 - p_1 \geq \epsilon > 0 \) ensures this.

**Remark 3** Writing \( \theta_n \) for the RCM-based estimator of \( \theta \), it follows from theorem 1 that the asymptotic variance of \( n^{1/2} \varphi^T (\hat{\Theta} - \theta_0) \) is no greater than that of \( n^{1/2} \varphi^T (\theta_n - \theta_0) \), where \( \varphi \in \mathbb{R}^2 \).

**Remark 4** A 100(1-\( \alpha \))% Wald-type confidence interval for \( \varphi^T \theta_0 \) is given by \( \varphi^T \hat{\Theta} \pm z_{\alpha/2} \hat{\psi} \), where \( \hat{\psi} \) is the estimated standard error of \( \varphi^T \hat{\Theta} \) and \( z_\alpha \) denotes the upper-\( \alpha \) quantile of the standard normal distribution. In simulations, \( \varphi = (1, 0)^T, (0, 1)^T, (1, 1)^T \) are used to obtain confidence intervals for \( a, b \) and \( a+b \) respectively.

**Remark 5** Let \( \hat{\Theta} \) denote the estimated covariance matrix of \( n^{1/2}(\hat{\Theta} - \theta_0) \). To test the null hypothesis \( H_0 : \theta = \theta_0 \) against the alternative \( H_1 : \theta \neq \theta_0 \) the test statistic \( n(\hat{\Theta} - \theta_0)^T \hat{\Sigma}^{-1}(\hat{\Theta} - \theta_0) \) has a chi-squared distribution with 2 degrees of freedom. Furthermore, to test a marginal hypothesis, say, \( H_0 : b = b_0 \) against the alternative \( H_1 : b \neq b_1 \) the test statistic \( n((\hat{b} - b_0)/\hat{\sigma}_b)^2 \), where \( \hat{\sigma}_b \) denotes the estimated standard error of \( \hat{b} \), has a chi-squared distribution with 1 degree of freedom. The power of the latter test for the proposed SRCMs vs standard RCM is analyzed through simulations, see section 3.

### 3 Simulation results

In this section, we present simulation results comparing the performance of the proposed (SRCM) and existing (RCM) estimators of the location and scale parameters. First we compare the confidence intervals of \( \theta \) obtained by the competing methods. Then we report the results of a power study.
3.1 Wald-type confidence intervals

The data were generated from a two-sample location-scale model and the empirical coverage probabilities (ECPs) and estimated average widths (EAWs) of the 95% confidence intervals for \(a\), \(b\), and \(a+b\) were computed. The joint confidence regions for \(a\) and \(b\) obtained by using the competing approaches were also compared. For joint confidence regions, however, an “estimated average area” (EAA), obtained by averaging over replications the products of the lengths of the simultaneous confidence intervals for \(a\) and \(b\), was computed.

The failure-time distributions were each taken to be independent normal; that is, \(Y_i \sim N(\mu_i, \sigma_i^2), i = 1, 2\). The censoring distributions were also each taken to be independent normal, independent of the \(Y_i\); that is, \(Z_i \sim N(\mu_c, \sigma_c^2)\). Let \(\Phi(\cdot)\) and \(\phi(\cdot)\) denote the standard normal cumulative distribution and probability density functions respectively. The censoring rate (CR) for each group, expressed as a function \(\mu, \mu_c, \sigma, \sigma_c\), is given by:

\[
C(\theta) = 1 - \int_c \Phi \left( \frac{y - \mu}{\sigma} \right) \phi \left( \frac{y - \mu_c}{\sigma_c} \right) dy.
\]

The two group sample sizes were each 100, and the ECPs and EAWs were based on 2000 replications. The parameters of the two failure-time distributions were taken as \(\mu_1 = 0, \mu_2 = 1, \sigma_1 = 1, \sigma_2 = 2\). Since, for the setting investigated, \(a = \mu_2 - \frac{\sigma_2}{\sigma_1} \mu_1\) and \(b = \frac{\sigma_2}{\sigma_1}\), the true values of \(a\) and \(b\) are 1 and 2 respectively. The first censoring distribution parameters \(\mu_{c1}\) and \(\sigma_{c1}\) were taken to give a fixed 40% CR for the first sample. For the second sample, different values of the censoring parameters \(\mu_{c2}\) and \(\sigma_{c2}\) were taken to give CRs between 10% and 40%. The measure \(G\) was taken as the uniform distribution over \([p_0, p_1]\), where \(p_0 = 0.01\) and \(p_1 = 0.96\). For computing the proposed estimators, both the probit and the Cauchy models were fitted from each sample data:

\[
\begin{align*}
    m_i^{\text{(Probit)}}(x, \gamma_i) &= \Phi(\gamma_{i1} + \gamma_{i2}x), & i = 1, 2, \\
    m_i^{\text{(Cauchy)}}(x, \gamma_i) &= 0.5 + \frac{1}{\pi} \arctan(\gamma_{i1} + \gamma_{i2}x), & i = 1, 2.
\end{align*}
\]

Note that both models are misspecified for \(m_i, i = 1, 2\), the closed forms of each being intractable. In figure 1, the ECPs of the 95% confidence intervals/regions of \(a\), \(b\), \(a+b\), and \((a,b)\), are presented as a function of the CR. The Bonferroni method was used to obtain the 95% joint confidence regions for \((a,b)\). Since the family confidence coefficient or the overall level of significance was 0.05, each individual confidence interval was computed using the confidence coefficient 0.025. Thus the confidence interval for \(a\) and \(b\) was given by

\[
\hat{a} \mp z_{0.0125} \hat{\sigma}_a, \quad \hat{b} \mp z_{0.0125} \hat{\sigma}_b.
\]

Even though the parametric models were misspecified, the proposed confidence intervals/regions produced coverage approximately close to the nominal value of 95%.

In figure 2, we present the percent relative reductions (PRRs) in EAW/EAA of the proposed confidence intervals/regions over the RCM-based ones. For the location parameter \(a\), the PRR in EAW of the proposed
Figure 1: The empirical coverage probabilities (ECPs) of the 95% confidence intervals/regions as a function of the censoring rate (CR) for the second sample, with the CR of the first sample fixed at 40%.
3.1 Wald-type confidence intervals

95% confidence interval over the RCM based counterpart ranges between 5% and 7%, while for the scale parameter $b$ and for the sum $a + b$ the PRR varies between 10% and 14%. The PRR in EAA of the proposed 95% joint confidence region of $(a, b)$ varies between 16% to 20% over the RCM-based method.

![Graph showing percentage reduction in estimated average width/area](image)

Figure 2: The percentage relative reduction in estimated average width/area of the proposed confidence intervals/regions over RCM-based ones as a function of the censoring rate (CR) for the second sample, with the CR of the first sample fixed at 40%.

Two more sets of simulation studies were also carried out. For both studies $a = 0$ and $b = 1$. However, the CR for the first sample was different for each study, being held at 40% for one and 30% for the other. The CR for the second sample was taken as before and consisted of several values between 10% and 40%. The results are plotted in figures 3–6 as a function of the CR in the second sample. The results were as before, with the proposed confidence intervals providing approximately correct coverage and with the Cauchy fitted estimators performing best overall. Furthermore, the proposed confidence intervals provided between 5% and 20% PRR in EAW/EAA over the RCM based ones as before.
3.1 Wald-type confidence intervals

SIMULATION RESULTS

Figure 3: The empirical coverage probabilities (ECPs) of the 95% confidence intervals/regions as a function of the censoring rate (CR) for the second sample, with the CR of the first sample fixed at 40% for $a = 0, b = 1$. 
3.1 Wald-type confidence intervals

SIMULATION RESULTS

Figure 4: The percentage relative reduction in estimated average width/area of the proposed confidence intervals/regions over RCM-based ones as a function of the censoring rate (CR) for the second sample, with the CR of the first sample fixed at 40% for $a = 0, b = 1$. 
Figure 5: The empirical coverage probabilities (ECPs) of the 95% confidence intervals/regions as a function of the censoring rate (CR) for the second sample, with the CR of the first sample fixed at 30% for $a = 0, b = 1$. 
3.2 A power study

Figure 6: The percentage relative reduction in estimated average width/area of the proposed confidence intervals/regions over RCM-based ones as a function of the censoring rate (CR) for the second sample, with the CR of the first sample fixed at 30% for \( a = 0, b = 1 \).

3.2 A power study

Here we compare the power of the competing methods to reject a false null hypothesis. We considered the marginal test \( H_0 : b = 2 \) versus \( H_1 : b \neq 2 \). We generated the data from the normal location-scale family, always keeping \( a = 1 \) but varying \( b \) over the interval \((1, 4)\). For each chosen \( b \), both samples were simulated with 40% CR, and we computed the proportion of times that the test statistic 
\[
Z = n^{1/2} \left( \hat{b} - b_0 \right) / \hat{\sigma}_b
\]
gave a value that was more extreme than the critical value computed under \( H_1 \). Specifically, we rejected \( H_0 \) when
\[
\left| Z - n^{1/2} \frac{b_0 - b}{\hat{\sigma}_b} \right| \geq z_{\alpha/2},
\]
where \( b_0 = 2 \) and \( z_{\alpha} \) is the upper-\( \alpha \) quantile of \( \Phi \). For each chosen \( b \), the empirical power, which is the proportion of rejected \( H_0 \), was based on 2,000 tests. The empirical power is plotted as a function of \( b \) for all the three procedures, namely, the RCM and SRCM with probit and Cauchy fits. The value of \( \alpha \) was 0.05. Note that when \( H_0 \) is true, our test statistic can be approximated by the standard normal distribution. The
power function in figure 7 shows that as \( b \) deviates away from the null value \( b_0 = 2 \), the power of the each test increases, and, at \( b_0 = 2 \), the power of each test is very close to 0.05, which is the level of significance. The proposed SRCM-based tests have greater power than the RCM-based test, especially for higher values of \( b \).

Figure 7: Empirical power of competing tests, when testing \( H_0 : b = 2 \) against \( H_1 : b \neq 2 \), as a function of \( b \), for 40% censoring rate for each sample.

4 Real data illustrations

Two real example data sets are analyzed to illustrate the proposed methodology. These are the mouse leukemia data given in Kalbfleisch and Prentice (1980) and the highly censored acute myelogenous data given in Klein and Moeschberger (2005).

4.1 Analysis of the mouse leukemia data

First we illustrate the proposed procedure with the mouse-leukemia data set given in Kalbfleisch and Prentice (1980). Of the 75 mice with thymic leukemia, \( n_1 = 36 \) are males and \( n_2 = 39 \) are females. One male lifetime observation of 26 was discarded as an extreme outlier. There are 2 censored observations in the male group and 5 in the female group. From figure 8, which shows box plots of the logarithm of the lifetime values for both males and females, there appears to be a distributional difference among male and female mice not
only in location but also in scale. However, the Generalized Wilcoxon test fails to detect any distributional difference, see page 211 of Kalbfleisch and Prentice (1980).

To incorporate SRCMs, we examined the adequacy of several models using the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). Table 1 below shows the AIC and BIC values for the different models for both the samples of the mouse leukemia data. The Cauchy model provided the lowest AIC and BIC compared to all other models.

<table>
<thead>
<tr>
<th>Model</th>
<th>First sample</th>
<th>Second sample</th>
<th>First sample</th>
<th>Second sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probit</td>
<td>23.80424</td>
<td>33.95392</td>
<td>26.97128</td>
<td>37.33168</td>
</tr>
<tr>
<td>Logit</td>
<td>23.52002</td>
<td>32.78232</td>
<td>26.68705</td>
<td>36.16008</td>
</tr>
<tr>
<td>Cauchy</td>
<td><strong>16.2652</strong></td>
<td><strong>26.43407</strong></td>
<td><strong>19.43224</strong></td>
<td><strong>29.81183</strong></td>
</tr>
<tr>
<td>Complimentary log-log</td>
<td>23.97008</td>
<td>35.0694</td>
<td>27.13711</td>
<td>38.44716</td>
</tr>
</tbody>
</table>

Table 1: AIC and BIC of different models used for the mouse leukemia data.

Assuming that \( c = p_1 - p_0 = 0.96 - 0.01 = 0.95 \), the parameter estimates, width/area of the confidence intervals/regions, and percent reduction in width/area of proposed over the RCM based approach are given in
4.2 Analysis of the acute myelogenous data

These data with high CRs pertain to acute myelogenous leukemia (AML) patients, who were placed in two groups (Klein and Moeschberger, 2005). One group is the AML low risk while the other is the high risk group. The time to death or time on study for each patient in days was observed. There are 54 patients in the

tables 2 and 3. The Cauchy fitted estimates produced confidence intervals with the smallest width. In figure 9 we present the error bar plots, showing the point estimates accompanied by the confidence intervals. The efficacy of the Cauchy model is again confirmed.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>(a) Estimate</th>
<th>Width</th>
<th>PRR</th>
<th>(b) Estimate</th>
<th>Width</th>
<th>PRR</th>
</tr>
</thead>
<tbody>
<tr>
<td>RCM</td>
<td>0.12369</td>
<td>0.0822</td>
<td>—</td>
<td>0.95959</td>
<td>0.0504</td>
<td>—</td>
</tr>
<tr>
<td>SRCM (Cauchy)</td>
<td>0.13673</td>
<td>0.0756</td>
<td><strong>8.0292</strong></td>
<td>0.96019</td>
<td>0.0494</td>
<td><strong>1.9841</strong></td>
</tr>
<tr>
<td>SRCM (Probit)</td>
<td>0.14278</td>
<td>0.0804</td>
<td>2.1898</td>
<td>0.96043</td>
<td>0.0528</td>
<td>-4.7619</td>
</tr>
</tbody>
</table>

Table 2: Point estimates, widths of 95% confidence intervals, and percentage relative reduction (PRR) in width over RCM for the mouse leukemia data.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>(a + b) Estimate</th>
<th>Width</th>
<th>% Reduction</th>
<th>((a, b)) Area</th>
<th>% Reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>RCM (Kaplan–Meier)</td>
<td>1.08328</td>
<td>0.1011</td>
<td>—</td>
<td>0.00542</td>
<td>—</td>
</tr>
<tr>
<td>SRCM (Cauchy link)</td>
<td>1.09693</td>
<td>0.09342</td>
<td><strong>7.5964</strong></td>
<td>0.00488</td>
<td><strong>9.9631</strong></td>
</tr>
<tr>
<td>SRCM (Probit link)</td>
<td>1.10322</td>
<td>0.08403</td>
<td>16.8843</td>
<td>0.00485</td>
<td>10.5166</td>
</tr>
</tbody>
</table>

Table 3: Point estimates, width/area of 95% confidence interval/region, and percentage reduction over RCM.

Figure 9: Error bar plots for \(a\), \(b\), and \(a + b\), showing point estimates accompanied by confidence intervals.

4.2 Analysis of the acute myelogenous data

These data with high CRs pertain to acute myelogenous leukemia (AML) patients, who were placed in two groups (Klein and Moeschberger, 2005). One group is the AML low risk while the other is the high risk group. The time to death or time on study for each patient in days was observed. There are 54 patients in the
low risk of which 31 are censored, and there are 45 patients in the high risk group of which 11 are censored. From figure 10, it is apparent that the distribution of the lifetimes of patients in the two groups may differ not only in location but also in scale.

Figure 10: Boxplot of the logarithm of the lifetime data for patients affected with Acute myeloid leukemia (AML).

An AIC/BIC analysis indicated that the Cauchy model would fit both samples the best, see table 4. The next best fit was the logit, so was also included in the analysis.

<table>
<thead>
<tr>
<th>Model</th>
<th>AIC Sample 1</th>
<th>AIC Sample 2</th>
<th>BIC Sample 1</th>
<th>BIC Sample 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probit</td>
<td>59.342</td>
<td>46.057</td>
<td>63.357</td>
<td>49.715</td>
</tr>
<tr>
<td>Logit</td>
<td>56.021</td>
<td>40.933</td>
<td>60.035</td>
<td>44.59</td>
</tr>
<tr>
<td>Cauchy</td>
<td>41.781</td>
<td>24.325</td>
<td>45.796</td>
<td>27.982</td>
</tr>
<tr>
<td>Complimentary log-log</td>
<td>64.278</td>
<td>49.621</td>
<td>68.292</td>
<td>53.278</td>
</tr>
</tbody>
</table>

Table 4: AIC and BIC of different models for fitting the acute myelogenous data

Taking $c = p_1 - p_0 = 0.96 - 0.01 = 0.95$ as before, the parameter estimates, width/area of the confidence intervals/regions, and PRR in width/area of proposed over the RCM based approach are given in tables 5
and 6. Again, the Cauchy fitted estimates produced confidence intervals with the smallest width. From the boxplot, figure 10, it is apparent that the RCM-based approach underestimates the location parameter, due perhaps to the high CR in the first sample. Error bar plots, showing the point estimates accompanied by the confidence intervals, are presented in figure 11. The efficacy of the Cauchy model is yet again confirmed.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$a$ Estimate</th>
<th>Width</th>
<th>PRR</th>
<th>$b$ Estimate</th>
<th>Width</th>
<th>PRR</th>
</tr>
</thead>
<tbody>
<tr>
<td>RCM</td>
<td>-0.1921</td>
<td>3.3775</td>
<td></td>
<td>0.8471</td>
<td>0.5349</td>
<td></td>
</tr>
<tr>
<td>SRCM (Cauchy)</td>
<td>-0.5432</td>
<td>3.2431</td>
<td>3.979</td>
<td>0.9115</td>
<td>0.4899</td>
<td>8.413</td>
</tr>
<tr>
<td>SRCM (Logit)</td>
<td>-0.7816</td>
<td>3.4453</td>
<td>-2.007</td>
<td>0.9499</td>
<td>0.5028</td>
<td>6.001</td>
</tr>
</tbody>
</table>

Table 5: Point estimates, widths of 95% confidence intervals, and percent relative reduction (PRR) in width over RCM for the acute myelogenous data.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$a + b$ Estimate</th>
<th>Width</th>
<th>PRR</th>
<th>$(a, b)$ Estimate</th>
<th>Area</th>
<th>PRR</th>
</tr>
</thead>
<tbody>
<tr>
<td>RCM (Kaplan–Meier)</td>
<td>0.6549</td>
<td>2.8692</td>
<td>2.3597</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRCM (Cauchy link)</td>
<td>0.3683</td>
<td>2.7812</td>
<td>3.067</td>
<td>2.2626</td>
<td>4.115</td>
<td></td>
</tr>
<tr>
<td>SRCM (Logit link)</td>
<td>0.1683</td>
<td>2.9704</td>
<td>-3.527</td>
<td>2.0753</td>
<td>12.052</td>
<td></td>
</tr>
</tbody>
</table>

Table 6: Point estimates, width/area of 95% confidence intervals/regions, and percent relative reduction (PRR) over RCM for the acute myelogenous data.

![Figure 11: Error bar plots for $a$, $b$, and $a + b$, showing point estimates accompanied by confidence intervals for the acute myelogenous data.](image-url)
5 Concluding discussion

In this paper a semiparametric approach for estimating the censored two-sample location-scale model parameters has been proposed. This development was carried out by exploiting SRCMs, more specifically semiparametric quantile function estimators. The success of SRCMs is predicated on specifying a parametric model for each group-specific conditional non-censoring probability given the observed minimum as the single covariate. A good-fitting model from among logistic, probit, Cauchy, complementary log-log, generalized proportional hazards (Dikta, 1998), among others, can be supplied for improved estimation and inference. In general, the search for a good-fitting binary regression model can be narrowed down by first discarding ill-fitting models using model-checking methods, such as the one introduced by Dikta, Kvesic, and Schmidt (2006). As was followed in both the illustrations given in this paper, an apt model can then be zeroed-in as the one that gives the smallest AIC and BIC. Theoretical and numerical studies offer good support for incorporating SRCMs into censored two-sample location-scale analysis. In simulation studies, SRCMs using the Cauchy link provided the best empirical coverage probabilities, which were closest to the pre-specified nominal level. Although the parametric models that were fitted were misspecified, they still offered considerable improvements over the approach that used the nonparametric (RCM) quantile function estimators. Numerical studies indicate that, in practice, in the absence of knowledge concerning the true parametric model, the Cauchy model may be fitted to obtain improved location-scale parameter estimates.

When there are missing censoring indicators (MCIs), meaning that the censoring indicator may be missing for a subset of study subjects, the proposed methodology can still be employed, indicating its flexibility over the RCM-based approach which cannot handle MCIs easily. The RCM approach would require estimation of conditional functions, which would also require the user to supply optimal window widths. With SRCMs, however, parameters of the model for the binary response data can still be estimated from the complete cases (Subramanian, 2004), and the location-scale parameter estimates can be obtained as described herein.

An important issue concerns the question whether a location-scale model would be a tenable one for a given data set. For uncensored data, Doksum and Sievers (1976) computed simultaneous confidence bands (SCBs) for a vertical shift process, using which one may test the adequacy of a location-scale hypothesis by determining whether or not a straight line fits within the SCB. Henry, Wells and Tiwari (1994) implemented the RCM extension of Doksum and Siever’s (1976) method. Hall, Lombard, and Potgieter (2013) investigated this problem for uncensored data by using the Cramér-von-Mises type of discrepancy combined with empirical characteristic functions. Research on this front utilizing SRCMs is in progress currently.

In conclusion, as shown in this paper, the SRCMs are highly useful tools for obtaining improved estimation and inference in censored data models. They offer great promise that SRCMs will provide substantial improvements over existing RCM-based methods in other applications, including testing for model adequacy.
6 References


