Pak. J. Statist.
2025 Vol. 41(1), 91-102

LOCAL MODEL SELECTION UNDER RIGHT CENSORING

Izzessalam Telitel[1] and **Mohamed Boukeloua**[2§]
[1] Université Frères Mentouri Constantine 1, Algeria
Email: izzessalam.tlitel@doc.umc.edu.dz
[2] Laboratoire de Génie des procédés pour le Développement Durable et les Produits de Santé (LGPDDPS), Ecole Nationale Polytechnique de Constantine, Constantine, Algeria
and Laboratoire de Biostatistique, Bioinformatique et Méthodologie Mathématique Appliquées aux Sciences de la Santé´(BIOSTIM), Faculté de Médecine, Université Salah Boubnider Constantine 3, Alegria
[§] Corresponding author Email: boukeloua.mohamed@gmail.com
mohamed.boukeloua@enp-constantine.dz

ABSTRACT

In this work, we address the problem of local model selection with right censored data. First, we present the Kaplan-Meier empirical measure which we use to estimate the local power divergence between the unknown density of interest and a candidate model. Then, we present the asymptotic properties of the estimate that we propose. On the basis of these properties, we introduce a local divergence information criterion for model selection in the case of right censoring. Finally, we apply our criterion on a real dataset of the acute myelocytic leukemia patients who received bone marrow transplantation. Among many mixture models with normal components, we find that a mixture of three normal components is the best model to describe the distribution of this dataset in different parts of its support.

KEYWORDS

Local model selection, right censored data, power divergence, Kaplan-Meier empirical measure, acute myelocytic leukemia dataset.

1. INTRODUCTION

Since the pioneer work of Akaike (1973), the model selection problems continue to draw the interest of the statisticians. Many authors still deal with this problems such as Shang et al. (2024) who proposed a generalized expectation model selection algorithm for latent variable selection in multidimensional item response theory models, Mamun and Paul (2023) who studied the properties of the forward selection, backward elimination, and stepwise selection in generalized linear models, dealing with the normal linear regression as a special case, Nabika et al. (2024) who proposed an active learning with model selection method using multiple parametric models in order to improve the efficiency of spectral experiments and Wen et al. (2024) who elaborated an approach for building model confidence sets and gave a theoretical lower bound on the degree of confidence in the model confidence sets they built. Furthermore, the theory of ϕ divergences between measures, introduced by Csiszar (1963), has been widely applied in statistics. Broniatowski and Keziou (2009) used this theory to study some parametric models. Bouzebda and Keziou (2010) and Boukeloua (2021) used it to study semiparametric copula models for complete and censored data, respectively. Salicrú et al. (1994) proposed ϕ divergence tests in parametric models. Moreover, Basu et al. (1998) introduced a power divergence measure between an unknown density of interest and a parametric model, used in the model selection problem.

The approaches we have cited until now deal with the distribution of interest globally, i.e., on its whole support. However, it may happen that the results of the analysis change if we consider only a specific zone of this support. In such situations, Avlogiaris et al. (2016a) introduced local ϕ divergences that allow to quantify the dispersion between two distributions only on a part of their support. Using these local ϕ divergences, Avlogiaris et al. (2016b) proposed local tests in parametric models. These local tests have been extended to the case of right censored data by Boukeloua (2024). Moreover, Avlogiaris et al. (2019) introduced a local divergence information criterion for model selection based on the work of Basu et al. (1998). For our part, we extend the work of Avlogiaris et al. (2019) to the context of right censoring. In this context, instead of observing the variable of interest, we observe the minimum between this variable and another censoring variable, as well as an indicator of censorship, indicating which variable is observed. We start by presenting the Kaplan-Meier empirical measure which substitutes the empirical measure in the case of right censored data. Basing on this measure, we propose an estimate to the local power divergence between the unknown density of interest and a candidate model and we give the asymptotic properties of this estimate. Using these properties, we propose by analogy with the work of Avlogiaris et al. (2019), a local divergence information criterion for model selection under right censoring. Furthermore, we apply our proposed criterion on a real dataset of the acute myelocytic leukemia patients who received bone marrow transplantation.

The rest of the paper is organized as follows. In Section 2, we introduce the local divergence information criterion for model selection under right censoring. In Section 3, we present the results of our application on the acute myelocytic leukemia dataset. We give some conclusions and perspectives in Section 4. Finally, we conclude with three appendices in which we give some results we need in our theoretical study as well as the acute myelocytic leukemia dataset.

2. LOCAL MODEL SELECTION CRITERION

Let X be a non-negative real random variable (r.r.v.) with unknown cumulative distribution function G and with probability density function g. The function g is the Radon-Nikodym derivative of the probability distribution of X with respect to a $\sigma-$finite measure m on $\left(\mathbb{R}_+, B\left(\mathbb{R}_+\right)\right)$ $\left(B(\mathbb{R}_+\right)$ being the Borel $\sigma-$algebra of $\mathbb{R}_+$). Moreover, consider a set of candidate models $F=\left\{f_\theta, \theta\in\Theta\right\}$ $\left(\Theta\subseteq\mathbb{R}^d\right)$, where f_θ is the Radon-Nikodym derivative of a probability measure P_θ with respect to m. We assume that X is right censored by a non-negative r.r.v. R, independent of X. More precisely, we have at disposal a sample $\left(Z_i = \min\left(X_i, R_i\right), \Delta_i = 1_{\{X_i \le R_i\}}\right)_{1\le i\le n}$ (where 1_E is the indicator function of the set E) of independent and identically distributed (i.i.d.) random variables having the same distribution as $\left(Z = \min(X,R), \Delta = 1_{\{X_i \le R_i\}}\right)$. In what follows, for any random variable V, S_V and T_V denote, respectively, the survival function and the upper endpoint of the support of V. Moreover, for any right continuous function φ: $\mathbb{R} \to \mathbb{R}$, we set $\varphi(x^-) = \lim_{\varepsilon \xrightarrow{>} 0} \varphi(x-\varepsilon)$ and $\Delta_\varphi(x) = \varphi(x) - \varphi(x^-)$ whenever the limit exists. Furthermore, for any vector or matrix A, we denote by A^T the transpose of A.

For any function ψ: $\mathbb{R} \to \mathbb{R}$, the advanced time transformation of ψ with respect to G is defined by

$$\tilde{\psi}(x) = \frac{1}{S_X(x)} \int_x^{+\infty} \psi(u)\, dG(u),\ \forall x \in \mathbb{R}$$

(see Efron and Johnstone (1990)).

Since X is not completely observed, we estimate G by the Kaplan-Meier estimator, defined by

$$G_n(x) = 1 - \prod_{i / Z'_i \le x} \left(1 - \frac{D(Z'_i)}{U(Z'_i)}\right)$$

where $(Z'_i)_{1\le i\le k}$ $(k \le n)$ are the distinct values in increasing order of $\left\{h_\omega, \omega \in \Theta^*\right\}$, $D(Z'_i) = \sum_{j=1}^n \Delta_j 1_{\{Z_i = Z'_i\}}$ is the number of real deaths at the instant Z and $U(Z'_i) = \sum_{j=1}^n 1_{\{Z_j \ge Z'_i\}}$ is the number of individuals at risk just before the instant Z'_i.

We can show that

$$G_n(x) = 1 - \prod_{j / Z_{(j)} \le x} \left(1 - \frac{n-j}{n-j+1}\right)^{\Delta_j},$$

where $Z_{(1)}, ..., Z_{(n)}$ are the order statistics of the observations $Z_1, ..., Z_n$.

This relation of the Kaplan-Meier estimator is the most used in practice. The empirical measure of G_n can be written as

$$P_n^{KM} = \sum_{i=1}^k \Delta_{G_n}\left(Z'_i\right) \delta_{Z'_i} = \sum_{i=1}^n \frac{\Delta_{G_n}\left(Z_i\right)}{D\left(Z_i\right)} \delta_{Z_i},\ \left(\frac{0}{0} := 0\right)$$

where δ_x, is the Dirac measure at the point x.

Furthermore, let $\omega \in \Theta^* \subseteq \mathbb{R}^M$ be a known parameter and let $\left\{h_\omega, \omega \in \Theta^*\right\}$ be a set of Radon-Nikodym derivatives with respect to m, of some parametric probability measures on $\mathbb{R}_+$, absolutely continuous with respect to m. We will use the following local version of the Basu, Harris, Hjort and Jones (BHHJ) power divergence (see Basu et al. (1998)) between g and f_θ, introduced by Avlogiaris et al. (2019).

$$D_\alpha^\omega\left(g, f_\theta\right)=\int_0^{+\infty} h_\omega\left(x\right)$$
$$\left(f_\theta^{1+\alpha}\left(x\right)-\left(1+\frac{1}{\alpha}\right)g\left(x\right) f_\omega^\alpha\left(x\right)+\frac{1}{\alpha} g^{1+\alpha}\left(x\right)\right)dm\left(x\right)$$

where $a > 0$ is the index parameter.

This local divergence can be obtained from the local-divergence defined in Avlogiaris et al. (2016a) (relation (4)), by choosing the convex function equal to

$$\phi_a\left(u\right)=u^{1+a}-\left(1+\frac{1}{a}\right)u^a+\frac{1}{a}.$$

The limit of $D_a^\omega\left(g, f_\theta\right)$ when a goes to 0 is the local Kullback-Leibler divergence and for $a=1$ it reduces to the square of the standard local L_2 distance between g and f_θ.

Moreover, let us define

$$W_a^\omega\left(\theta\right)=\int_0^{+\infty} h_\omega\left(x\right) f_\theta^{1+a}\left(x\right)dm\left(x\right)-\left(1+\frac{1}{a}\right)\int_a^{+\infty} h_\omega\left(x\right)g\left(x\right)f_\theta^a\left(x\right)dm\left(x\right)$$
$$=D_a^\omega\left(g, f_\theta\right)-\frac{1}{a}\int_0^{+\infty} h_\omega\left(x\right)g_\theta^{1+a}\left(x\right)dm\left(x\right)$$

Remarking that

$$W_a^\omega\left(\theta\right)=E_{f_\theta}\left(h_\omega\left(X\right) f_\theta^a\left(X\right)\right)-\left(1+\frac{1}{a}\right)E_g\left(h_\omega\left(X\right) f_\theta^a\left(X\right)\right)$$

(where E_f denotes the mathematical expectation under the probability density f), we can approach $W_a^\omega\left(\theta\right)$, as in Avlogiaris et al. (2019), by

$$Q_a^\omega\left(\theta\right)=\int_0^{+\infty} h_\omega\left(x\right) f_\theta^{1+a}\left(x\right)dm\left(x\right)-\left(1+\frac{1}{a}\right)\int_a^{+\infty} h_\omega\left(x\right) f_\theta^a\left(x\right)dP_n^{KM}\left(x\right)$$
$$=\int_0^{+\infty} h_\omega\left(x\right) f_\theta^{1+a}\left(x\right)dm\left(x\right)-\left(1+\frac{1}{a}\right)\sum_{i=1}^n h_\omega\left(Z_i\right) f_\theta^a\left(Z_i\right)\frac{\Delta_{G_n}\left(Z_i\right)}{D\left(Z_i\right)}$$

Assume that $g \in F$ and let $\hat{\theta}_n$, be a consistent and asymptotically normal estimator of θ_T, the true value of θ, constructed from the sample $\left(Z_1,\Delta_1\right),\ldots,\left(Z_n,\Delta_n\right)$. The following theorem gives an explicit form of the local expected overall discrepancy between g and f_θ, $E_g\left(W_a^\omega\left(\hat{\theta}_n\right)\right)$ which is an estimator of $W_a^\omega\left(\theta_T\right)$. Note that the results of this theorem are given under the assumption $T_X \le T_R$. This assumption is very employed in the right censoring setting and it allows, in particular, to apply the strong law of large numbers under right censoring (see Theorem 2 below and the remark that follows it).

Theorem 1.

Under the assumption $T_X \le T_R$, *we have*

1. $nE_g\left(W_a^\omega\left(\hat{\theta}_n\right)\right) = nQ_a^\omega\left(\hat{\theta}_n\right) + n(a+1)\left(\hat{\theta}_n - \theta_T\right)^T J^\omega\left(\theta_T\right)\left(\hat{\theta}_n - \theta_T\right)$

2. $E_g\left[n\left(\hat{\theta}_n - \theta_T\right)^T J^\omega(\theta_T)\left(\hat{\theta}_n - \theta_T\right)\right] = \sum_{i=1}^{r}\beta_i$

   where $\hat{\theta}_n = \arg\min_{\theta\in\Theta}\left[\int_0^{+\infty}\left(f_\theta^{1+a}(x) - \left(1+\frac{1}{a}\right)f_\theta^a(x)\right)dP_n^{KM}(x)\right]$

   $\beta_1, \beta_2, ..., \beta_r$. *are the non-zero eigenvalues of the matrix* $J^\omega\left(\theta_T\right)AVar\left(\theta_T\right)$

   With $r = rank\left(AVar\left(\theta_T\right)J^\omega\left(\theta_T\right)AVar\left(\theta_T\right)\right)$,

$$J^\omega\left(\theta_T\right) = \left(\int h_\omega(x) f_\theta^{1+a}(x)\frac{\partial \log f_\theta(x)}{\partial\theta_i}\frac{\partial \log f_\theta(x)}{\partial\theta_j}dm(x)\right)$$

$$AVar\left(\theta_T\right) = J^{-1}\left(\theta_T\right)K\left(\theta_T\right)J^{-1}\left(\theta_T\right)$$

   with

$$J\left(\theta_T\right) = \int_0^{+\infty}u_{\theta_T}(x) - u_{\theta_T}^T(x)\, f_{\theta_T}^{1+a}(x)dm(x)$$

   and

$$K\left(\theta_T\right) = \int_0^{+\infty}\left(u_{\theta_T}(x) - \tilde{u}_{\theta_T}(x)\right)\left(u_{\theta_T}(x) - \tilde{u}_{\theta_T}(x)\right)^T \frac{f_{\theta_T}^{1+2a}(x)}{S_R\left(x^-\right)}dm(x)$$

$$-\int_0^{+\infty}u_{\theta_T}(x) f_{\theta_T}^{1+a}(x)dm(x)\int_0^{+\infty}u_{\theta_T}^T(x)\, f_{\theta_T}^{1+a}(x)dm(x)$$

$$\left(u_\theta(x) = \nabla_\theta \log f_\theta(x)\right)$$

Proof:

Using the strong law of large numbers in the case of right censored data (see Theorem 2 below) and the central limit theorem in the same case (see Theorem 3 below), this theorem can be proved in the same way as Proposition 5 of Avlogiaris et al. (2019).

From this theorem, we define the local divergence information criterion as follows.

$$L_{a,n}^{\omega}\left(\hat{\theta}_n,\beta_1,\ldots,\beta_r\right)=nQ_a^{\omega}\left(\hat{\theta}_n\right)+\left(a+1\right)\sum_{i=1}^{r}\beta_i$$

In a certain zone of the support of X, determined by the density h_{ω}, this criterion allows to choose the most relevant model to describe the distribution of X, from a collection of possible models. For example, if we have to choose locally between two candidate models f^1 and f^2, we compare between $L_{a,n}^{\omega}\left(\hat{\theta}_n^{(1)},\beta_1,\ldots,\beta_r\right)$ and $L_{a,n}^{\omega}\left(\hat{\theta}_n^{(2)},\beta_1,\ldots,\beta_r\right)$ and the smallest between these two values determines the most relevant model.

3. REAL DATA APPLICATION

In this section, we will illustrate our proposed criterion of local model selection, on a dataset of the survival time of acute myelocytic leukemia patients after the bone marrow transplantation. This dataset was used by Copelan et al. (1991) and it is given in Appendix C. It is composed of 80 complete observations and 57 right censored observations. The survival times being recorded in days, we divide by 365 to treat them in years. First, we represent in the left panel of Figure 1, the censored data histogram (see Huzurbazar (2005)) of the obtained dataset. We remark that this histogram highlights at least two main bulks of observations separated by some gaps. These two bulks are presented in red in the figure (bulk B1 and bulk B2). So, we can see that a mixture of at least two normal components is a suitable model to describe this dataset. Starting out from this observation, we consider the following four candidate models.

- Model 1: Mixture of two normal components.
- Model 2: Mixture of three normal components.
- Model 3: Mixture of four normal components.
- Model 4: Mixture of five normal components.

We use the estimator $\hat{\theta}_n$ given in (1) to estimate the parameters of these models, with an index parameter of $a = 0.01$. As it is noted above, the limit of the local BHHJ power divergence $D_a^{\omega}\left(g, f_{\theta}\right)$, when a goes to 0, is the local Kullback-Leibler divergence. So, our choice of a makes the local power divergence D^{ω} close to the local Kullback-Leibler which is a very employed divergence. The obtained values for the estimators of the parameters are given in Table 1 below, where $\hat{\mu}_i$ (resp. $\hat{\sigma}_i$) is the estimator of the mean (resp. the standard deviation) of the ith normal component. Then, we present in the right panel of Figure 1, the density of each candidate model with the estimated parameters.

Table 1
Estimation of the Parameters of the Candidate Models

The Model	Estimation of the Parameters
Model 1	$\hat{\mu}^1 = 0.11, \hat{\sigma}^1 = 0.2, \hat{\mu}^2 = 7.07, \hat{\sigma}^2 = 0.1$
Model 2	$\hat{\mu}^1 = 0.11, \hat{\sigma}^1 = 0.11, \hat{\mu}^2 = 1.30, \hat{\sigma}^2 = 0.6, \hat{\mu}^3 = 7.10, \hat{\sigma}^3 = 0.1$
Model 3	$\hat{\mu}^1 = 0.11, \hat{\sigma}^1 = 0.08, \hat{\mu}^2 = 1.02, \hat{\sigma}^2 = 0.5, \hat{\mu}^3 = 6, \hat{\sigma}^3 = 0.6,$ $\hat{\mu}^4 = 6.95, \hat{\sigma}^4 = 0.04$
Model 4	$\hat{\mu}^1 = 0.11, \hat{\sigma}^1 = 0.07, \hat{\mu}^2 = 0.66, \hat{\sigma}^2 = 0.37, \hat{\mu}^3 = 2.8, \hat{\sigma}^3 = 1,$ $\hat{\mu}^4 = 6, \hat{\sigma}^4 = 0.6, \hat{\mu}^5 = 6.95, \hat{\sigma}^5 = 0.01$

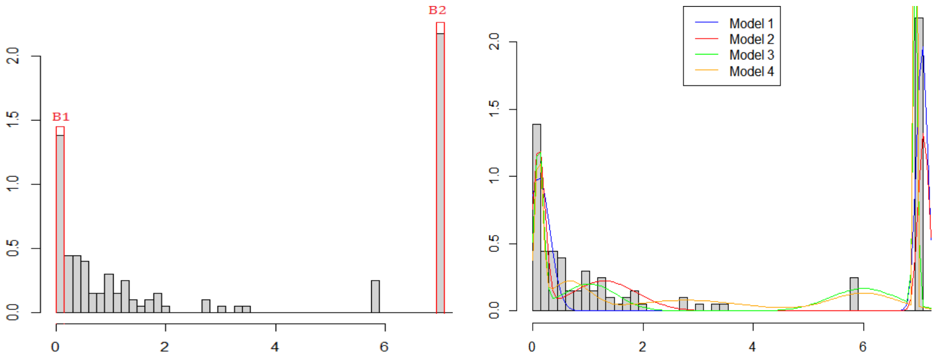

Figure 1: Censored Data Histogram of the Acute Myelocytic Leukemia Data

Moreover, we calculate the local divergence information criterion for each candidate model. For that, we use a normal kernel with mean μ and variance $(0.1)^2$, with different values of μ, allowing to consider different zones of the support of the variable of interest. The results we obtain are presented in Table 2 below. We also indicate, in each zone, the selected model which corresponds to the smallest value of the information criterion. We remark that for all considered zones, the best model is model 2 (mixture of three normal components), so it is the model that best describes the data in the different parts of the support of the variable of interest.

Table 2
The Local Divergence Information Criterion for Different Choices of the Kernel Function

μ	Model 1	Model 2	Model 3	Model 4	Selected Model
0.1	9.4×10^{10}	1.7497×10^4	3.2×10^{10}	3.5×10^8	Model 2
1	6.7×10^{11}	7.8159×10^3	1.2×10^8	1.8×10^8	Model 2
2	2.6×10^{11}	3.2157×10^4	3.4×10^9	6.7×10^{10}	Model 2
3	1.3×10^{11}	5.3312×10^4	2.5×10^9	5.4×10^{11}	Model 2
6	3×10^9	-8.3605×10^2	4.5×10^9	3.8×10^{10}	Model 2
7	3×10^{10}	3.5723×10^3	8.7×10^6	2×10^{10}	Model 2

In order to deeply investigate the distribution of the studied survival time, we present in Figure 2 the density function, the survival function and the failure rate of Model 2 (with the estimated parameters given in Table 1 above). Moreover, the calculation of the median, the mean and the standard deviation of this model gives the results presented in Table 3.

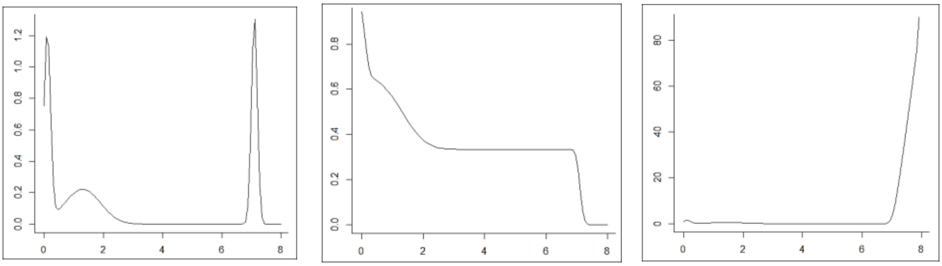

Figure 2: Density, Survival Function and Failure Rate of Model 2 (From left to right)

Table 3
Characteristics of Model 2

Median	1.3
Mean	2.8367
Standard Deviation	3.0223

4. CONCLUSION

In the context of right censored data, we have proposed a local divergence information criterion allowing model selection in specific parts of the support of the studied distribution. This criterion generalizes the one introduced by Avlogiaris et al. (2019). To obtain the theoretical results, we have based on the strong law of large numbers and the central limit theorem under right censoring. Moreover, we have applied our proposed criterion on a real dataset of the acute myelocytic leukemia patients. In particular, we have locally selected the best model that describes this dataset among some mixture models with normal components. The results we have obtained show that in many parts of the support of the studied distribution, a mixture model with three normal components is the best model that fits the dataset. Our approach can also be applied in many other fields involving the right censorship. For example, in the reliability for the study of the operating time of components, in insurance for the study of the survival time of insureds, in economy for the study of the effectiveness of financial policies and in environmental studies where certain variables may not be completely observed due to the detection ability of the measurement tools. In the future, it would be interesting to consider other types of censoring such as doubly or interval censored data. For that, we first have to look at the generalization of the strong law of large numbers and the central limit theorem for these types of data. Then, it would be easy to show similar results to Theorem 1 above and to construct the local model selection criterion in these cases. It would also be interesting to look at other local statistical inference problems, with censored data, such as local homogeneity tests and local independence tests.

ACKNOWLEDGMENTS

The authors wish to thank the editor and the reviewers for their valuable suggestions and comments.

REFERENCES

1. Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Proceeding of the Second International Symposium on Information Theory, Petrov B.N. and Cs´aki F., Eds., Akad´emiai Kiad´o, Budapest*, pages 276-281.
2. Avlogiaris, G., Micheas, A. and Zografos, K. (2016a). On local divergences between two probability measures. *Metrika*, 79(3), 303-333.
3. Avlogiaris, G., Micheas, A. and Zografos, K. (2016b). On testing local hypotheses via local divergence. *Statistical Methodology*, 31, 20-42.
4. Avlogiaris, G., Micheas, A.C. and Zografos, K. (2019). A criterion for local model selection. *Sankhya A: The Indian Journal of Statistics*, 81(2), 406-444.
5. Basu, A., Harris, I.R., Hjort, N.L. and Jones, M. (1998). Robust and efficient estimation by minimizing a density power divergence. *Biometrika*, 85(3), 549-559.
6. Boukeloua, M. (2021). Study of semiparametric copula models via divergences with bivariate censored data. *Communications in Statistics-Theory and Methods*, 50(23), 5429-5452.
7. Boukeloua, M. (2024). Testing local hypotheses with different types of in complete data. *Turkic World Mathematical Society, Journal of Applied and Engineering Mathematics*, 14(4), 1374-1389.
8. Bouzebda, S. and Keziou, A. (2010). New estimates and tests of independence in semiparametric copula models. *Kybernetika*, 46(1), 178-201.
9. Broniatowski, M. and Keziou, A. (2009). Parametric estimation and tests through divergences and the duality technique. *Journal of Multivariate Analysis*, 100(1), 16-36.
10. Copelan, E.A., Biggs, J.C., Thompson, J.M., Crilley, P., Szer, J., Klein, J.P., Kapoor, N., Avalos, B.R., Cunningham, I. and Atkinson, K. (1991). Treatment for acute myelocytic leukemia with allogeneic bone marrow transplantation following preparation with bucy2. *The American Society of Hematology*, 78(3), 838-843.
11. Csiszàr, I. (1963). Eine informationstheoretische ungleichung und ihre anwendung auf den beweis der ergodizität von markoffschen ketten. *Magyar Tud. Akad. Mat. Kutato Int. Közl.*, 8, 85-108.
12. Efron, B. and Johnstone, I. M. (1990). Fisher's information in terms of the hazard rate. *The Annals of Statistics*, 18(1), 38-62.
13. Huzurbazar, A.V. (2005). A censored data histogram. *Communications in Statistics—Simulation and Computation*, 34(1), 113-120.
14. Mamun, A. and Paul, S. (2023). Model selection in generalized linear models. *Symmetry*, 15(10), 1905.
15. Nabika, T., Nagata, K., Mizumaki, M., Katakami, S. and Okada, M. (2024). Bayesian active learning with model selection for spectral experiments. *Scientific Reports*, 14(1), 3680.

16. Salicrú, M., Morales, D., Menéndez, M.L. and Pardo, L. (1994). On the applications of divergence type measures in testing statistical hypotheses. *Journal of Multivariate Analysis*, 51(2), 372-391.
17. Shang, L., Zheng, Q.-Z., Xu, P.-F., Shan, N. and Tang, M.-L. (2024). A generalized expectation model selection algorithm for latent variable selection in multidimensional item response theory models. *Statistics and Computing*, 34(1), 49.
18. Stute, W. and Wang, J.-L. (1993). The strong law under random censorship. *Ann. Statist.*, 21(3), 1591-1607.
19. Wen, F., Jiang, J., and Luan, Y. (2024). Model selection path and construction of model confidence set under high-dimensional variables. *Mathematics*, 12(5), 664.
20. Zhou, M. (2015). *Empirical Likelihood Method in Survival Analysis*, volume 79. CRC Press.

APPENDIX-A

Strong Law of Large Numbers under Right Censoring

Theorem 2 (Stute and Wang (1993), Theorem 1.1)

Let be a Borel-measurable function on $\mathbb{R}$ *such that* $\int \left|\varphi(x)\right| dG(x) < \infty$

We have

$$\lim_{n\to\infty} \int \varphi(x) dG_n(x) = \int \varphi(x) dG(x)$$

with probability 1 and in the mean.

Note that the limit is equal to $\int \varphi(x) dG(x)$ since G is continuous and $T_X \le T_R$ (see Remark 3 of Stute and Wang (1993)).

APPENDIX-B

Central Limit Theorem under Right Censoring

Theorem 3 (Zhou (2015), Lemma 21).

Let $\varphi = (\varphi_1, \varphi_2, ..., \varphi_3) : \mathbb{R} \to \mathbb{R}^d$ *be a function satisfying* $\int \varphi_i(x) dG(x) = 0$ *, for all* $i \in \{1,...,d\}$ *. We have*

$$\int \varphi(x) dG_n(x) \xrightarrow{D} N(0, \Sigma \quad), \text{ as } n \to \infty$$

where $\xrightarrow{D}$ *denotes the convergence in distribution and* $\Sigma = (\sigma_{ij})_{1\le i, j\le d}$ *with*

$$\sigma_{ij} = \int \left[\varphi_i(x) - \tilde{\varphi}_i(x)\right]\left[\varphi_j(x) - \tilde{\varphi}_j(x)\right] \frac{dG(x)}{S_R(x^-)}, \forall i, j \in \{1,...,d\}$$

APPENDIX-C

The Acute Myelocytic Leukemia Dataset

Observed time	Censoring indicator	Observed time	Censoring indicator	Observed time	Censoring indicator	Observed time	Censoring indicator
2081	0	1	1	162	1	2133	0
1602	0	107	1	262	1	1238	0
1496	0	269	1	1384	1	1631	0
1462	0	350	0	414	1	2024	0
1433	0	2569	0	2204	1	1345	0
1377	0	2506	0	1063	1	1136	0
1330	0	2409	0	481	1	845	1
996	0	2218	0	105	1	491	1
226	0	1857	0	641	1	162	1
1199	0	1829	0	390	1	1298	1
1111	0	1562	0	288	1	121	1
530	0	1470	0	522	1	2	1
1182	0	1363	0	79	1	62	1
1167	0	1030	0	1156	1	265	1
418	1	860	0	583	1	547	1
417	1	1258	0	48	1	341	1
276	1	2246	0	431	1	318	1
156	1	1870	0	1074	1	195	1
781	1	1799	0	393	1	469	1
172	1	1709	0	10	1	93	1
487	1	1674	0	53	1	515	1
716	1	1568	0	80	1	183	1
194	1	1527	0	35	0	105	1
371	1	1324	0	1499	1	128	1
526	1	957	0	704	1	164	1
122	1	932	0	653	1	129	1
1279	1	847	0	222	0	122	1
110	1	848	0	1356	0	80	1
243	1	1850	0	2640	0	677	1
86	1	1843	0	2430	0	73	1
466	1	1535	0	2252	0	168	1
262	1	1447	0	2140	0	74	1

Observed Time	Censoring Indicator
16	1
248	1
732	1
105	1
392	1
63	1
97	1
153	1
363	1