Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
http://siba-ese.unisalento.it/index.php/ejasa/index
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v12n2p380

Analysis of breast cancer data in framework of a GPD model with interval censoring
By Dakhmouche, Bahi

Published: 14 October 2019

Electronic Journal of Applied Statistical Analysis
Vol. 12, Issue 2, 2019, 380-391
DOI: 10.1285/i20705948v12n2p380

Analysis of breast cancer data in framework of a GPD model with interval censoring

Meghlaoui Dakhmouche and Oussama Bahi*

Laboratory of Mathematical Modelling and Simulation, University of Frères Mentouri Constantine 1, 25000 Constantine, Algeria

Published: 14 October 2019

In this work, we are interested in a hypothesis testing problem within the framework of a Generalized Pareto Distribution (GPD) model with interval censoring. For this purpose, we first develop the calculation of the likelihood function, using conditional probabilities, to achieve the same expression proposed by Klein and Moeschberger. Next, we show that the properties of the maximum pseudo-likelihood estimates of the model parameters and essentially the asymptotic normality are preserved. Finally, we built a hypothesis testing to compare two types of breast cancer treatment as part of the model mentioned above. As a result, we can distinguish which treatment lengthens the comfort time of the patients.

Keywords: Interval censoring, likelihood function, GPD model, asymptotic normality, hypothesis testing

1 Introduction

In practice, many problems are better formalized within censored models. To illustrate, in medicine and biology, censored data appears in clinical trial for survival studies (Lawless, 2003). Moreover, in finance, the reinsurance's companies face some censored data due to unreported or non-settled information claims (Albrecher et al., 2017). In an overview of available literature, a lot of contributions have dealt with censored models on the right (Lawless, 2003), left (Samson et al., 2006) or right and left (Turnbull,

*Corresponding author: bahi.oussama2@gmail.com.

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

1974). Furthermore, the concept of interval censoring has been introduced (Klein and Moeschberger, 1997) to generalize different types of censoring. Statistical inference in such models has been developed using more generally non-parametric or semi-parametric methods (Gentleman and Geyer, 1994, Clifford Anderson-Bergman, 2016, Kalbfleisch and Prentice, 2002, Lin et al., 2016).

In this work, we aim to develop a statistical inference within a parametric model by mean of the study of a corpus of data (Lindsey and Ryan, 1998) concerning two types of breast cancer treatment. We want to compare these two treatments using a hypothesis testing strategy built around the tail indexes of empirical data distributions for both types of cure.

In the second Section, we reformulate the likelihood function using a different calculation technique to achieve the same expression as in (Klein and Moeschberger, 1997). The estimates of the model parameters are obtained by maximizing a pseudo-likelihood function. Furthermore, we propose a proof of the asymptotic normality of model parameter estimates because this property is essential for the construction of a hypothesis testing.

In the last Section, we define the variable of interest X as the duration of the aesthetic comfort of the patients breasts, that is the time elapsed between the beginning of the treatment and the appearance of breast retraction. Without treatment, breast retraction appears fairly quickly. However, treatment by radiotherapy or chemotherapy extends the patient's aesthetic comfort time. Therefore, we can assume that X is an excess above a threshold. Thereby, we reposition our study in the context of a GPD model. Afterwards, we consider a test's statistic based on the difference between the tail indexes of the distributions and standardized using the results of the previous sections as in Wald (Wald, 1943). Finally, the rejection region is determined and the power of the test is calculated.

2 Parametric model with interval censoring

Let X be a random variable with probability distribution function F_X, defined on the probability space $\left(\mathbb{R}_+, \mathcal{B}_{\mathbb{R}_+}, P_{\boldsymbol{\theta}}\right)_{\boldsymbol{\theta}\in\Theta}$. Let us suppose that the set Θ is compact in $\mathbb{R}^d$ and contains the true value. Moreover, let us assume that the model is identifiable, that the application $\boldsymbol{\theta} \to P_{\boldsymbol{\theta}}$ is injective for all $\boldsymbol{\theta}$ in Θ. Data collection for some random phenomena is subject to interval censoring, so instead of observing the variable of interest X, we observe the triplet (Y, Z, Δ) which takes its values in $\mathcal{D} = \mathbb{R}_+^2 \times \{1, 2, 3\}$. The components Y and Z are two random variables absolutely continuous with respect to the Lebesgue measure. We assume that the joint density of Y and Z does not depend on $\boldsymbol{\theta}$, and that the variable of interest X is independent of the pair (Y, Z). The component Δ is a discrete random variable defined such as

$$\Delta = \begin{cases} 1 \text{ if } Y < X \leq Z \\ 2 \text{ if } X > Z \\ 3 \text{ if } X \leq Y \end{cases}.$$

The joint probability distribution of the triplet (Y, Z, Δ) is such that

$$\begin{aligned} P\left(y \leq Y < y+dy, z \leq Z < z+dz, \Delta = \delta\right) &= P\left(y \leq Y < y+dy, z \leq Z < z+dz\right) \\ & \quad P\left\{\Delta = \delta \,/ (y \leq Y < y+dy, z \leq Z < z+dz)\right\}. \end{aligned}$$

It obvious that,

$$\lim_{dy \to 0\ ,\ dz \to 0} \frac{P\left(y \leq Y < y+dy, z \leq Z < z+dz\right)}{dydz} = f_{(Y,Z)}(y,z),$$

where $f_{(Y,Z)}$ is the joint density function of the couple (Y, Z). Moreover, for $\delta = 1, 2, 3$ we have

$$\begin{aligned} \lim_{dy \to 0,\ dz \to 0} P\left\{(\Delta = 1) / (y \leq Y < y+dy, z \leq Z < z+dz)\right\} &= P\left(y < X \leq z\right), \\ \lim_{dy \to 0,\ dz \to 0} P\left\{(\Delta = 2) / (y \leq Y < y+dy, z \leq Z < z+dz)\right\} &= P\left(X > z\right), \\ \lim_{dy \to 0,\ dz \to 0} P\left\{(\Delta = 3) / (y \leq Y < y+dy, z \leq Z < z+dz)\right\} &= P\left(X \leq y\right). \end{aligned}$$

The distribution of the conditional random variable $(\Delta = \delta / Y = y, Z = z)$ can be summarized as:

$$\begin{aligned} P\left(\Delta = \delta | Y = y, Z = z\right) &= \lim_{dy \to 0, dz \to 0} P\left(\Delta = \delta \,/ y \leq Y < y+dy, z \leq Z < z+dz\right) \\ &= P\left(y \leq X < z\right)^{\mathbf{1}_{\{\delta=1\}}(\delta)} P\left(X < y\right)^{\mathbf{1}_{\{\delta=2\}}(\delta)} P\left(X \geq z\right)^{\mathbf{1}_{\{\delta=3\}}(\delta)}. \end{aligned}$$

Let $f_{(Y,Z,\Delta)}$ be the density function of the triplet (Y, Z, Δ) defined such that

$$\lim_{dy \to 0, dz \to 0} P\left(y \leq Y < y+dy, z \leq Z < z+dz, \Delta = \delta\right) = f_{(Y,Z,\Delta)}(y,z,\delta),$$

where

$$\begin{aligned} f_{(Y,Z,\Delta)}\left(y,z,\delta\right) &= f_{(Y,Z)}(y,z) P\left(y \leq X < z\right)^{\mathbf{1}_{\{\delta=1\}}(\delta)} P\left(X \leq y\right)^{\mathbf{1}_{\{\delta=2\}}(\delta)} P\left(X > z\right)^{\mathbf{1}_{\{\delta=3\}}(\delta)} \\ &= f_{(Y,Z)}(y,z) \left(F_X(z) - F_X(y)\right)^{\mathbf{1}_{\{\delta=1\}}(\delta)} F_X(y)^{\mathbf{1}_{\{\delta=2\}}(\delta)} \left(1 - F_X(z)\right)^{\mathbf{1}_{\{\delta=3\}}(\delta)}. \end{aligned}$$

Let $(Y_k, Z_k, \Delta_k)_{k=1,\ldots,n}$ be a random sample of the triplet (Y, Z, Δ) whose distribution function depends on the parameter $\boldsymbol{\theta}$. The likelihood function of this sample is

$$\begin{aligned} L\left(\boldsymbol{\theta}\right) &= \prod_{k=1}^{n} f_{(Y,Z,\Delta)}(y_k, z_k, \delta_k) \\ &= \prod_{k=1}^{n} f_{(Y,Z)}(y_k, z_k) \prod_{k=1}^{n} U\left(y_k, z_k, \delta_k; \boldsymbol{\theta}\right) \end{aligned}$$

where

$$U\left(y_k, z_k, \delta_k; \boldsymbol{\theta}\right) = \left(F_X\left(z_k; \boldsymbol{\theta}\right) - F_X\left(y_k; \boldsymbol{\theta}\right)\right)^{1_{\{\delta_k=1\}}(\delta_k)} \left(1 - F_X\left(z_k; \boldsymbol{\theta}\right)\right)^{1_{\{\delta_k=2\}}(\delta_k)} F_X\left(y_k; \boldsymbol{\theta}\right)^{1_{\{\delta_k=3\}}(\delta_k)}$$

The maximum likelihood estimator of $\boldsymbol{\theta}$ is obtained by maximizing the log-pseudo-likelihood $\ell(\boldsymbol{\theta}) = \log L(\boldsymbol{\theta}) = \sum_{k=1}^{n} \log U(y_k, z_k, \delta_k; \boldsymbol{\theta})$, so the maximum pseudo-likelihood estimator $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}\in\Theta} \ell(\boldsymbol{\theta}).$$

In the sequel, we are going to prove the asymptotic normality of the estimator $\widehat{\boldsymbol{\theta}}$.

2.1 Asymptotic normality of the estimator

Let us define the score function D to simplify the notation in the sequel, as

$$D(Y, Z, \Delta; \boldsymbol{\theta}) = \left(\frac{\partial}{\partial\theta_1} \log U(Y, Z, \Delta; \boldsymbol{\theta}), ..., \frac{\partial}{\partial\theta_d} \log U(Y, Z, \Delta; \boldsymbol{\theta}) \right).$$

Moreover, let us assume that the function $U(y, z, \delta; \boldsymbol{\theta})$ is smooth and twice continuously differentiable with respect to $\boldsymbol{\theta}$ in a neighborhood of the true value for all $(y, z, \delta) \in \mathcal{D}$ (condition C_1).

Lemma 1 *If the function $U(Y, Z, \Delta; \boldsymbol{\theta})$ satisfies the condition C_1, then*

$$E\left[D(Y, Z, \Delta; \boldsymbol{\theta})\right] = 0$$

Proof. In fact, for every i = 1,..., d, we have

$$\begin{aligned} E\left(\frac{\partial}{\partial\theta_i} \log U(Y, Z, \Delta; \theta_i) \right) &= \iint \sum_{\delta\in\{1,2,3\}} \frac{\frac{\partial}{\partial\theta_i} U(y, z, \delta; \theta_i)}{U(y, z, \delta; \theta_i)} U(y, z, \delta; \theta_i) f(y, z) dydz \\ &= \frac{\partial}{\partial\theta_i} \iint \sum_{\delta\in\{1,2,3\}} f_{Y,Z,\Delta}(y, z, \delta; \theta_i) dydz = 0. \end{aligned}$$

■

Moreover, let us suppose that the information matrix $\mathcal{I}(\boldsymbol{\theta})$ exist for all $(i, j) \in \{1, ..., d\}^{\otimes 2}$ (condition C_2).

Lemma 2 *If the function $U(Y, Z, \Delta; \boldsymbol{\theta})$ satisfies the condition C_2, then*

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} D(Y_k, Z_k, \Delta_k; \boldsymbol{\theta}) \xrightarrow{\mathbb{D}} \mathcal{N}\left(\mathbf{0}, \mathcal{I}(\boldsymbol{\theta})\right).$$

Proof. Considering the Lemma 1, it is obvious that the covariance matrix of the vector $D(Y, Z, \Delta; \boldsymbol{\theta})$ coincides with the Fisher information matrix $\mathcal{I}(\boldsymbol{\theta})$, so according to the central limit theorem we have

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} D(Y_k, Z_k, \Delta_k; \boldsymbol{\theta}) \xrightarrow{\mathbb{D}} \mathcal{N}(\mathbf{0}, \mathcal{I}(\boldsymbol{\theta})).$$

■

Finally, let us assume that the function $U(y, z, \delta; \boldsymbol{\theta})$ is continuous for all $(y, z, \delta) \in \mathcal{D}$ and for all $\boldsymbol{\theta} \in \Theta$ (condition C_3), and that the Hessian matrix $J(\boldsymbol{\theta})$ exists and is invertible for all $(i, j) \in \{1, ..., d\}^{\otimes 2}$ (condition C_4).

Proposition 3 *If the function $U(Y, Z, \Delta; \boldsymbol{\theta})$ satisfies the conditions C_1, C_2, C_3 and C_4, then*

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\right) \xrightarrow{\mathbb{D}} \mathcal{N}\left(\mathbf{0}, \Sigma_{\boldsymbol{\theta}}\right),$$

where

$$\Sigma_{\boldsymbol{\theta}} = J(\boldsymbol{\theta})^{-1}\mathcal{I}(\boldsymbol{\theta})J(\boldsymbol{\theta})^{-1}.$$

Proof. The score function $D(Y_k, Z_k, \Delta_k; \widehat{\boldsymbol{\theta}}_n)$ equal to zero at $\widehat{\boldsymbol{\theta}}$ since $\widehat{\boldsymbol{\theta}}$ maximizes $\log U(Y, Z, \Delta; \boldsymbol{\theta})$. Then, using the Taylor-Young formula in the neighborhood of the true value of θ, we can write

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n} D(Y_k, Z_k, \Delta_k; \widehat{\boldsymbol{\theta}}_n) = \frac{1}{\sqrt{n}}\sum_{k=1}^{n} D(Y_k, Z_k, \Delta_k; \boldsymbol{\theta}) + J_n(\boldsymbol{\theta})\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) + R_n = 0, \quad (1)$$

where $J_n(\boldsymbol{\theta})$ is such as

$$J_n(\boldsymbol{\theta}) = \left(J_n^{i,j}(\boldsymbol{\theta})\right)_{1\leq i,j\leq d} = \left(\frac{1}{n}\sum_{k=1}^{n}\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log U(Y_k, Z_k, \Delta_k; \boldsymbol{\theta})\right)_{1\leq i,j\leq d}.$$

$J_n^{i,j}(\boldsymbol{\theta})$ is the empirical mean of the function $H_{i,j} = \frac{\partial^2}{\partial\theta_i\partial\theta_j}\log U(Y_k, Z_k, \Delta_k; \boldsymbol{\theta}), 1 \leq i, j \leq d$. Then, using the weak law of large numbers we claim that

$$J_n(\theta) \xrightarrow{\mathbb{P}} J(\theta). \quad (2)$$

Therefore, we can say that $J_n(\boldsymbol{\theta})$ is almost surely invertible.
Finally, by Lemma 2 and using the equation 1, we can write that

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\right) \xrightarrow{\mathbb{D}} \mathcal{N}\left(\mathbf{0}, J(\boldsymbol{\theta})^{-1}\mathcal{I}(\boldsymbol{\theta})J(\boldsymbol{\theta})^{-1}\right).$$

■

3 Analysis of breast cancer data

3.1 Statement of the problem

A study (Lindsey and Ryan, 1998) was performed to compare the state of patients with early breast cancer treated by radiotherapy and chemotherapy together with the state of those treated by radiotherapy alone in terms of aesthetic effects. The subjects of the study are patients who were treated at the Joint Center for Radiation Therapy in Boston between 1976 and 1980. The objective of this study is to compare the results obtained

with the two treatments in terms of the deterioration of the aesthetic state of the breast of the patients. These patients are examined at clinical visits every 4 to 6 months. Monitoring intervals can be longer depending on the circumstances of each patient. Interval censoring data occur naturally in medical studies requiring periodic monitoring. Indeed, a person under periodic monitoring may miss visits and return in an altered state. The results of this study are presented in Table 1 where the intervals indicate the period of time during which the aesthetic state has deteriorated. For example, if an observation is coded $(6; 14)$, then at the 6^{th} month the patient shows no aesthetic deterioration of the breast, but at 14^{th} month the retraction of the breast is present. If the observation is coded $(0; 5)$, then the patient shows an aesthetic deterioration of the breast before the study began, but that was figured out at the 5^{th} month of the study. If an observation is coded $(11, 61)$, then at the 11^{th} month the patient shows no aesthetic deterioration of the breast and maintained that until the end of the study at the 61^{th} month.

Table 1: Breast cancer data with two different treatments

Therapy = RC		Status	Therapy = R		Status
Left	Right		Left	Right	
8	12	1	45	61	2
0	22	3	6	10	1
24	31	1	0	7	3
17	27	1	46	61	2
17	23	1	46	61	2
24	30	1	7	16	1
16	24	1	17	61	2
13	61	2	7	14	1
11	13	1	37	44	1
16	20	1	0	8	3
18	28	1	4	11	1
17	26	1	15	61	2
32	61	2	11	15	1
23	61	2	22	61	2
44	48	1	46	61	2
14	17	1	46	61	2
0	5	3	25	37	1
5	8	1	46	61	2
12	20	1	26	40	1
11	61	2	46	61	2
33	40	1	27	34	1
31	61	2	36	44	1
13	39	1	46	61	2
19	32	1	36	48	1
34	61	2	37	61	2
13	61	2	40	61	2
16	24	1	17	25	1
35	61	2	46	61	2
15	22	1	11	18	1
11	17	1	38	61	2
22	32	1	5	12	1
10	35	1	37	61	2
30	34	1	0	5	3
13	61	2	18	61	2
10	17	1	24	61	2
8	21	1	36	61	2
4	9	1	5	11	1
11	61	2	19	35	1
14	19	1	17	25	1
4	8	1	24	61	2
34	61	2	32	61	2
30	36	1	33	61	2
18	24	1	19	26	1
16	60	1	37	61	2
35	39	1	34	61	2
21	61	2	36	61	2
11	20	1			
48	61	2			

The esthetic deterioration of the breast is inevitable for a patient with breast cancer. So if the treatment is effective, the comfort time X for a treated patient tends to become longer. As a result, the tail of the X distribution for the treated patients tends to thicken in the case of effective treatment. Thus, to compare the effectiveness of the two types of treatment, we compare the distribution functions of the patients treated with

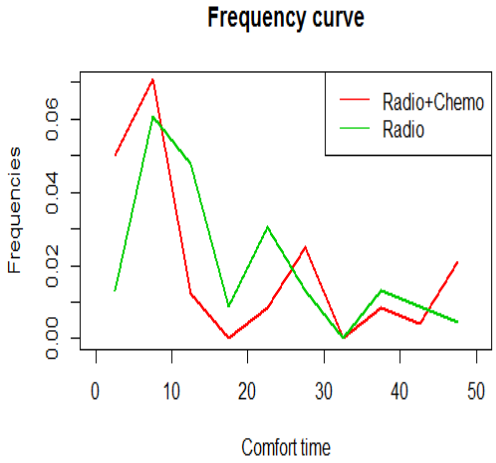

Figure 1: Frequency curves of the two empirical distributions of comfort time of the patients treated with chemotherapy and radiotherapy and the estimated comfort time of the patients treated with radiotherapy alone

chemotherapy and radiotherapy and radiotherapy alone using their tail indexes.

3.2 Estimation of model parameters

Let $\left\{X_i^{(0)}\right\}_{1\leq i\leq n_1}$ be the comfort time of patients who had been treated with chemotherapy after initial radiotherapy and $\left\{X_j^{(1)}\right\}_{1\leq j\leq n_2}$ be the comfort time of the patients who received only an equivalent dose of radiotherapy (Table1), and let $F(x^{(m)},\gamma_m,\sigma_m)$, $m=0$ or 1, be their distribution functions defined as follows:

$$F(x^{(m)};\gamma_m,\sigma_m)=1-\left(1+\gamma_m\frac{x^{(m)}}{\sigma_m}\right)^{-\frac{1}{\gamma_m}} \qquad \text{where } x^{(m)}>-\frac{\sigma_m}{\gamma_m}.$$

The model under consideration is subject to interval censoring. Therefore, the parameters estimation can be carried out using the maximum pseudo-likelihood method (Section 2). Let $\widetilde{\ell}(\gamma_m,\sigma_m)$ be the pseudo-likelihood function defined such as

$$\begin{aligned}\widetilde{\ell}(\gamma_m,\sigma_m) \;=\; & \sum_{k=1}^{n}\Big[1_{\{\delta_k=1\}}(\delta_k)\log\left((1+\frac{\gamma_m}{\sigma_m}y_k)^{-\frac{1}{\gamma_m}}-(1+\frac{\gamma_m}{\sigma_m}z_k)^{-\frac{1}{\gamma_m}}\right)\\ & -1_{\{\delta_k=2\}}(\delta_k)\frac{1}{\gamma_m}\log(1+\frac{\gamma_m}{\sigma_m}z_k)+1_{\{\delta_k=3\}}(\delta_k)\log\left\{1-(1+\frac{\gamma_m}{\sigma_m}y_k)^{-\frac{1}{\gamma_m}}\right\}\Big].\end{aligned}$$

To maximize $\widetilde{\ell}(\gamma_m,\sigma_m)$, we use the numerical method of Nelder-Mead (Nelder and Mead, 1965). The results of this optimization are summarized in the following table:

Table 2: GPD parameters estimation

Parameters	Estimation	Parameters	Estimation
γ_0	-0.4123641	γ_1	-0.6472023
σ_0	28.1896309	σ_1	73.9606626

The figure below shows the two survival function of GPD distribution with the estimated parameters:

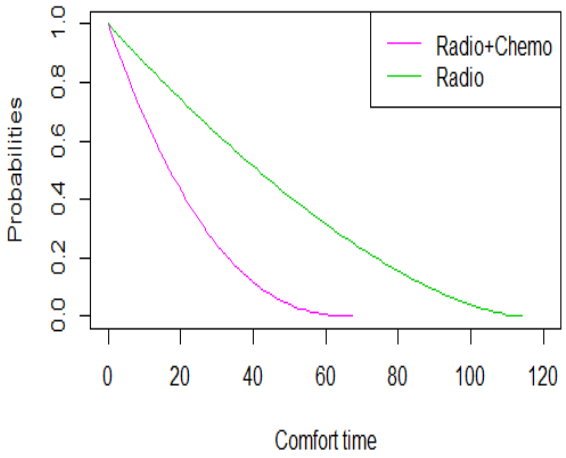

Figure 2: Two survival functions of the estimated comfort time of the patients treated with chemotherapy and radiotherapy and the estimated comfort time of the patients treated with radiotherapy alone.

3.3 Comparison between the two treatments

Figure 2 shows that the tail of the estimated comfort time distribution function of the patients treated with chemotherapy and radiotherapy is thinner than the tail of the distribution function of comfort time of the patients treated with radiotherapy alone. In other words, the aesthetic comfort time for the patients treated with chemotherapy and radiotherapy is shorter. That is, the treatment with chemotherapy and radiation does not lengthen the comfort time of the patients.
The comparison of the two treatments can be obtained simply by solving the testing problem of the null hypothesis H_0:'$\gamma = \gamma_0 = \gamma_1$' against the alternative hypothesis H_1 :'$\gamma = \gamma_1 < \gamma_0$'. There are several methods for determining a decision rule for this test. It is known that Neyman type tests and Wald type tests are equivalent for this kind of hypothesis (Toulemonde, 2008). However, for practical reasons we will proceed as in Wald (Wald, 1943) by standardizing the quantity $\widehat{\gamma}_0 - \widehat{\gamma}_1$ whose asymptotic variance $V = V_0 + V_1$ is given in the proposition 3 such that

$$V = \frac{J(\boldsymbol{\gamma}_0)^{-1}\mathcal{I}(\boldsymbol{\gamma}_0)J(\boldsymbol{\gamma}_0)^{-1}}{n_0} + \frac{J(\boldsymbol{\gamma}_1)^{-1}\mathcal{I}(\boldsymbol{\gamma}_1)J(\boldsymbol{\gamma}_1)^{-1}}{n_1}. \tag{3}$$

Let us define the statistic T such as

$$T = \frac{\widehat{\gamma}_0 - \widehat{\gamma}_1}{\sqrt{\widehat{\widehat{V}}}},$$

where $\widehat{V}$ is obtained by substituting γ_i by $\widehat{\gamma}_i$, i=0,1 in the relation 3.
Under the null hypothesis the statistic T has a standard normal distribution. Thus, the decision rule can be defined through the rejection region defined as:

$$R_c = \{T > q_{1-\alpha}\},$$

where the normal quantile $q_{1-\alpha} = \phi^{-1}(1-\alpha)$ and where ϕ is the distribution function of the standard normal distribution.
Moreover, under the hypothesis H_1, the statistic T is of normal distribution with mean $(\gamma_0 - \gamma_1)$ and variance $\widehat{V}$. Therefore, the power of this hypothesis testing can be expressed as

$$1 - \beta = P_{H_1}(R_c) = 1 - \phi\left(\frac{\phi^{-1}(1-\alpha) - (\gamma_0 - \gamma_1)}{\sqrt{\widehat{V}}}\right)$$

With the help of our script written under (Team et al., 2013), the value of the statistic T and the power $1 - \beta$ of the test are calculated such that

$$T = 2.475412 \quad , \quad q_{1-\alpha} = 1.644854 \quad \text{and} \quad 1 - \beta = 0.7968884,$$

as $T > q_{1-\alpha}$, we reject the hypothesis H_0.
As a conclusion, we can state that the treatment with radiotherapy and chemotherapy together does not improve the aesthetic comfort of patients with breast cancer, but it tends to damage it.

4 Conclusion

We have shown throughout our work that, in parametric models with interval censoring, the parameter estimates preserve their essential properties, namely efficiency and asymptotic normality. In addition, we have built a hypothesis testing and we calculated its power $1 - \beta$, which is not always possible. Therefore, we can claim that our test is as powerful as the Wald or Pearson type tests. Moreover, we can point out that it would be interesting to develop a specific method to fit probability distributions to interval-censored data. Hence, the distribution of these data can be confirmed instead of being supposed, thus many practical problems can be formalized and analyzed more accurately.

References

Albrecher, H., Teugels, J. L., and Beirlant, J. (2017). *Reinsurance: actuarial and statistical aspects*. John Wiley & Sons.

Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. (2006). *Statistics of extremes: theory and applications*. John Wiley & Sons.

Clifford Anderson-Bergman, Y. Y. (2016). Computing the log concave npmle for interval censored data. *Statistics and Computing*, 26(4):813–826.

Dorey, F. J., Little, R. J., and Schenker, N. (1993). Multiple imputation for threshold-crossing data with interval censoring. *Statistics in medicine*, 12(17):1589–1603.

Esteban, M. and Morales, D. (1998). On the asymptotic distribution of maximum likelihood estimators with doubly censored data. *Kybernetes*, 27(8):940–951.

Finkelstein, D. M. and Wolfe, R. A. (1985). A semiparametric model for regression analysis of interval-censored failure time data. *Biometrics*, pages 933–945.

Gentleman, R. and Geyer, C. J. (1994). Maximum likelihood for interval censored data: Consistency and computation. *Biometrika*, 81:618–623.

Gourieroux, C., Monfort, A., and Trognon, A. (1984). Pseudo maximum likelihood methods: Applications to poisson models. *Econometrica*, 52:701–720.

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The statistical analysis of failure time data*. John Wiley et Sons, New York; Chichester.

Klein, J. P. and Moeschberger, M. L. (1997). *Survival analysis: techniques for censored and truncated data*. Springer-Verlag Inc, Berlin; New York.

Klein, J. P. and Moeschberger, M. L. (2006). *Survival analysis: techniques for censored and truncated data*. Springer Science & Business Media.

Lawless, J. F. (2003). *Statistical models and methods for lifetime data*. John Wiley et Sons, New York; Chichester.

Lin, J., Sinha, D., Lipsitz, S., and Polpo, A. (2016). Semiparametric analysis of interval-censored survival data with median regression model. In *Statistical Applications from*

Clinical Trials and Personalized Medicine to Finance and Business Analytics, pages 149–163. Springer.

Lindsey, J. C. and Ryan, L. M. (1998). Methods for interval-censored data. *Statistics in medicine*, 17(2):219–238.

Nelder, J. A. and Mead, R. (1965). A simplex method for function minimization. *The computer journal*, 7(4):308–313.

Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. In *Handbook of econometrics, Vol. IV*, volume 2 of *Handbooks in Econom.*, pages 2111–2245. North-Holland, Amsterdam.

Odell, P. M., Anderson, K. M., and D'Agostino, R. B. (1992). Maximum likelihood estimation for interval-censored data using a weibull-based accelerated failure time model. *Biometrics*, pages 951–959.

Reiss, R.-D. and Thomas, M. (2007). *Statistical analysis of extreme values with applications to insurance, finance, hydrology and other fields*. Birkhäuser Verlag, Basel, third edition. With 1 CD-ROM (Windows).

Rücker, G. and Messerer, D. (1988). Remission duration: An example of interval-censored observations. *Statistics in Medicine*, 7(11):1139–1145.

Samson, A., Lavielle, M., and Mentré, F. (2006). Extension of the saem algorithm to left-censored data in nonlinear mixed-effects model: Application to hiv dynamics model. *Computational Statistics & Data Analysis*, 51(3):1562–1574.

Team, R. C. et al. (2013). R: A language and environment for statistical computing.

Thomas, M., Lemaitre, M., Wilson, M. L., Viboud, C., Yordanov, Y., Wackernagel, H., and Carrat, F. (2016). Applications of extreme value theory in public health. *PloS one*, 11(7):e0159312.

Toulemonde, G. (2008). *Estimation et tests en théorie des valeurs extrêmes*. PhD thesis, Université Pierre et Marie Curie-Paris VI.

Turnbull, B. W. (1974). Nonparametric estimation of a survivorship function with doubly censored data. *Journal of the American Statistical Association*, 69:169–173.

Van Der Vaart, A. W. (1998). *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge.

Varadhan, R., Gilbert, P., et al. (2009). Bb: An r package for solving a large system of nonlinear equations and for optimizing a high-dimensional nonlinear objective function. *Journal of Statistical Software*, 32(4):1–26.

Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical society*, 54(3):426–482.