Article

Not peer-reviewed version

# Flexible Hazard Modeling with Segmented Distributions

Guillermo Martínez Flórez , Rafael Bráz Azevedo Farias , Carlos Javier Barrera Causil *

Posted Date: 14 April 2025

doi: 10.20944/preprints202504.1011.v1

Keywords: segmented distributions; proportional hazard models; survival analysis; structural change; maximum likelihood; Laplace distribution; right censoring

*Article*

# Flexible Hazard Modeling with Segmented Distributions

**Guillermo Martínez-Flórez [1,†], Rafael Bráz Azevedo Farias [2,†] and Carlos Barrera-Causil [3,*,†]**

1 Departamento de Matemáticas y Estadística, Facultad de Ciencias Básicas, Universidad de Córdoba, Montería 230002, Colombia
2 Departamento de Estatística e Matemática Aplicada, Universidade Federal do Ceará, Ceará 60440-900, Brazil
3 Grupo de Investigación Davinci, Facultad de Ciencias Exactas y Aplicadas, Institución Universitaria ITM, Medellín 050034, Colombia (C.B.)
* Correspondence: carlosbarrera@itm.edu.co
† These authors contributed equally to this work.

**Abstract:** This paper introduces a flexible family of segmented proportional hazard distributions designed to model abrupt changes in hazard rates, which are often observed in medical and engineering applications. The proposed framework generalizes the proportional hazard transformation to segmented distributions, including new forms of the Rayleigh, log-logistic, Lindley, and Laplace PH models. We develop a maximum likelihood estimation procedure incorporating right censoring, a key feature of real-world survival data. The segmented hazard models effectively capture structural breaks in the hazard function, providing a robust alternative to traditional survival models that assume constant hazard dynamics. A case study based on IQ score data illustrates the improved flexibility and interpretability of the segmented Laplace PH model in detecting latent change points. The proposed models enhance the capacity to model complex survival patterns with abrupt changes in risk, contributing to a deeper understanding of dynamic hazard processes.

**Keywords:** segmented distributions; proportional hazard models; survival analysis; structural change; maximum likelihood; Laplace distribution; right censoring

---

## 1. Introduction

Proportional hazard models have become essential tools in survival analysis, particularly in fields like medicine and engineering, where time-to-event data often play a critical role in understanding treatment effects, risk factors, and patient outcomes [21]. These models are widely used to analyze survival data by adjusting the hazard function and incorporating time-dependent covariates, which are highly prevalent in real-world medical studies. However, traditional models often assume constant hazard rates over time, failing to account for abrupt changes in failure or risk rates that are frequently observed in clinical and environmental studies.

In medical research, for instance, sudden changes in failure rates may occur as a result of treatment interventions, disease progression, or other factors. A patient undergoing a new medication regimen may experience different risk levels before, during, and after treatment, with the hazard rate shifting accordingly. Similarly, environmental studies often exhibit abrupt changes in variables like pollution levels or temperature due to external factors, leading to significant variations in failure rates over time [5,10].

Traditional models struggle to capture these dynamic changes. For this reason, there is a growing need for more flexible statistical models that can accommodate multiple change points in the hazard function. Recent approaches, such as the exponential piecewise model [9], power piecewise exponential model [11], and exponentiated Weibull hazard function [16], have made progress in modeling these complexities. However, these methods are often limited to specific distributions and lack the versatility needed for more complex data structures.

 doi:10.20944/preprints202504.1011.v1



In this paper, we introduce a family of segmented proportional hazard models that extend beyond current methodologies by offering greater flexibility in modeling data with multiple change points in hazard rates. Our proposed models, including segmented versions of the Rayleigh, log-logistic, Lindley, and Laplace proportional hazard distributions, provide a powerful tool for analyzing survival data across various fields, particularly in medicine and engineering, where abrupt changes are common. By incorporating a censoring mechanism and employing maximum likelihood estimation, our models offer a robust alternative for analyzing complex survival data with high degrees of variability over time or space.

The remainder of this paper is organized as follows. Section 2 introduces the Proportional Hazard Distribution (PHD), which serves as the cornerstone for constructing the subsequent segmented models. Section 3 presents the proposed family of segmented proportional hazard distributions, including the segmented Rayleigh, log-logistic, Lindley, and Laplace variants (SLPH). In Section 4, we describe the maximum likelihood estimation procedure for these models, including the treatment of right-censored data, following the approach of [6]. Section 5 illustrates the practical application of the proposed methodology using real data, comparing the performance of the SLPH model against various alternative models such as the segmented Laplace (LS), segmented half-normal (HNS), segmented half-normal proportional hazard (HNSHP), standard Laplace, skew-normal (SN), Gamma, normal, and mixture of normals. Finally, Section 6 concludes the paper with a summary of findings and directions for future research.

## 2. Proportional Hazard Distribution

The Proportional Hazard Distribution (PHD) is introduced as an extension of the minimum distribution in the sample, wherein $n \in \mathbb{Z}$ is replaced by $\alpha \in \mathbb{R}^+$, for a random variable $Z$ with a Probability Density Function (PDF) denoted as $f$ and a Cumulative Distribution Function (CDF) denoted as $F$. As a result, this distribution can be regarded as that of a fractional order statistic. We proceed to define the PHD distribution (as seen in [15]).

Let $F$ be a continuous Cumulative Distribution Function (CDF) with a Probability Density Function (PDF) denoted as $f$, and a hazard function $h = f/(1-F)$. We state that a random variable $Z$ follows a proportional hazard distribution associated with $F$ $(f)$ and the parameter $\alpha > 0$ if its probability density function is given by

$$\varphi_F(z;\alpha) = \alpha f(z)[1-F(z)]^{\alpha-1}, \quad z \in \mathbb{R}, \tag{1}$$

where $\alpha$ is a positive real number. We use the notation $Z \sim PHD(\alpha)$. The distribution function of the $PHD$ model is defined as

$$G_F(z) = 1-[1-F(z)]^{\alpha}, \quad z \in \mathbb{R}. \tag{2}$$

The risk function of this model is expressed as $h_G(z;\alpha) = \alpha h(z)$, where $h(\cdot)$ represents the risk function of the Probability Density Function (PDF) $f(\cdot)$. In other words, the risk function is proportional to that of the PDF $f(\cdot)$, which explains the nomenclature "pdf with proportional hazard."

Within the proportional hazard family of distributions, in cases where $f(\cdot)$ is parameterized by a distribution family indexed by a parameter vector $\theta$, it suffices for $h(x,\theta)$ to be fully determined. This ensures the identifiability of the parameters within the PHD model.

In effect, much like [2], we consider two Probability Density Functions (PDFs) belonging to the proportional hazard family, characterized by the vectors $(\theta_1,\alpha_1)$ and $(\theta_2,\alpha_2)$. Since both distributions adhere to the proportional hazard principle, we have $h_i(x;\theta_i,\alpha_i) = \alpha_i h(x;\theta_i)$. Now, if for $(\theta_1,\alpha_1) \neq (\theta_2,\alpha_2)$ we find $\alpha_1 h(x;\theta_1) = \alpha_2 h(x;\theta_2)$, then $\alpha_1/\alpha_2 = h(x;\theta_2)/h(x;\theta_1)$. Notably, if $\theta_1 = \theta_2$, it yields $\frac{h(x;\theta_2)}{h(x;\theta_1)} = 1$, implying $\alpha_1 = \alpha_2$, which in turn contradicts the assumption of distinct parameters for the two distributions.

Now, when $\theta_1 \neq \theta_2$, at least one pair $x_i \neq x_{i'}$ exists such that $\frac{h(x;\theta_2)}{h(x;\theta_1)} \neq \frac{h(x_{i'};\theta_2)}{h(x_{i'};\theta_1)}$. Since $\alpha_1 = \alpha_2$ is constant, it follows that $\alpha_1 \neq \alpha_2$. These findings underscore the impossibility of distinct parameter sets

 doi:10.20944/preprints202504.1011.v1



leading to the same risk function. Consequently, it suffices for the risk function to be determined to ensure the identifiability of model parameters.

Now, we consider the scenario in which the density function $f(\cdot;\theta)$ takes on a segmented form, resulting in a more adaptable model for explaining abrupt alterations in the intermittent failure rate caused by diverse experimental factors within specific time frames or spaces.

## 3. Segmented Proportional Hazard Distribution

If the risk function, $h(x;\theta)$, of the base PDF $f(\cdot;\theta)$ exhibits distinct behaviors within different sections of its range, it becomes feasible to partition the domain of $h(\cdot;\theta)$ into $L$ intervals delimited by points $l_1 < l_2 < \cdots < l_{L-1} < l_L < \infty$. This allows for the expression of the failure rate ratio in the interval $[l_i,\ l_{i+1})$ as $h_i(x)$. Consequently, the segmented proportional hazard function $h(x)$, with $h_i(x) > 0$, can be defined as follows:

$$h(x) = \begin{cases} 0, & \text{if } x < l_1, \\ h_1(x), & \text{if } l_1 \leq x < l_2, \\ h_2(x), & \text{if } l_2 \leq x < l_3, \\ \vdots & \\ h_i(x), & \text{if } l_i \leq x < l_{i+1}, \\ \vdots & \\ h_L(x), & \text{if } l_L \leq x. \end{cases} \tag{3}$$

In a similar way, we have that

$$S(x) = 1 - F(x) = \begin{cases} S_1(x), & \text{if } l_1 \leq x < l_2, \\ S_2(x), & \text{if } l_2 \leq x < l_3, \\ \vdots & \\ S_i(x), & \text{if } l_i \leq x < l_{i+1}, \\ \vdots & \\ S_L(x), & \text{if } l_L \leq x, \end{cases} \tag{4}$$

Here, $F_i(x) = 1 - S_i(x)$ must be a monotonically non-decreasing and continuous function, while $F(x)$ must be continuous at the common points specifically, $F_i(l_{i+1}) = F_{i+1}(l_{i+1})$ for $i = 1, 2, 3, \cdots, L$. Similarly, $f_i(\cdot)$, the Probability Density Function (PDF) of $F_i(x)$, should also be continuous at these common points, i.e., $f_i(l_{i+1}) = f_{i+1}(l_{i+1})$. These conditions ensure that $F(\cdot)$ is a continuous and differentiable (smooth) function, and that $f(\cdot)$ is continuous at the shared points. As a result, it can be deduced that $S_i(l_{i+1}) = S_{i+1}(l_{i+1})$, $f_i(l_{i+1}) = f_{i+1}(l_{i+1})$, and $h_i(l_{i+1}) = h_{i+1}(l_{i+1})$ signifying that they are all monotonically continuous functions.

By defining the risk function, we have that

$$f_i(x) = h_i(x) S_i(x) \quad x \in \mathbb{R}, \;\; i = 1, 2, 3, \cdots, L. \tag{5}$$

Now, according to the definition of the cumulative risk function, it follows that

$$H_i(x) = -\log[S_i(x)] = \int_{x_{min}}^{x} h_i(y) dy,$$

that is,

$$\begin{aligned} S_i(x) &= \exp[-H_i(x)] \\ &= \exp\left[-\int_{l_1}^{x} h_i(y)dy\right] \\ &= \exp\left[-\int_{l_1}^{l_2} h_1(y)dy - \int_{l_2}^{l_3} h_2(y)dy - \cdots - \int_{l_{i-1}}^{l_i} h_{i-1}(y)dy - \int_{l_i}^{x} h_i(y)dy\right] \\ &= \exp\left[-\sum_{j=1}^{i-1}\int_{l_j}^{l_{j+1}} h_j(y)dy - \int_{l_i}^{x} h_i(y)dy\right] \end{aligned} \tag{6}$$

So, calling

$$\psi_i = \int_{l_i}^{l_{i+1}} h_i(y)dy \quad \text{and} \quad \Lambda_i(x) = \int_{l_i}^{x} h_i(y)dy$$

for $l_i \leq x < l_{i+1}$, the $i-$th component of the survival function is represented by:

$$\begin{aligned} S_1(x) = h_1(x)\exp(-\Lambda_1(x)), \quad \text{for } i = 1, \\ S_i(x) = \exp\left[-\sum_{j=1}^{i-1}\psi_j - \Lambda_i(x)\right], \quad \text{for } i = 2, 3, \ldots, L. \end{aligned} \tag{7}$$

By substituting (7) into (5), we obtain the component of the $i$-th group of the segmented Probability Density Function (PDF) of the risk function, denoted as (HSM) $f(\cdot)$. The segmented PDF of $f$ will be indicated as $X \sim HSM_f$.

The cumulative hazard distribution function of the proportional hazard family can be expressed in terms of the survival function $S_F(\cdot)$ as

$$G_F(x) = 1 - [1 - F(x)]^{\alpha} = 1 - [S(x)]^{\alpha}. \tag{8}$$

Consequently, the proportional hazard extension of the segmented function (HPSM) $SM_f$ is derived by substituting (4) into (8). The PDF of this extension is given by

$$g_F(x) = \alpha f(x)[S(x)]^{\alpha-1}. \tag{9}$$

Based on equation (5), the density (9) can be expressed as

$$g_F(x) = \alpha h(x)[S(x)]^{\alpha}. \tag{10}$$

The $i$-th component of the $i$-th group of the random variable $X$ is given by:

$$\begin{aligned} g_1(x) &= \alpha h_1(x)\exp(-\alpha\Lambda_1(x)), \quad \text{for } i = 1, \\ g_i(x) &= \alpha h_i(x)\exp\left[-\alpha\sum_{j=1}^{i-1}\psi_j - \alpha\Lambda_i(x)\right], \quad l_i \leq x < l_{i+1}, \quad \text{for } i = 2, 3, \ldots, L. \end{aligned} \tag{11}$$

This Probability Density Function (PDF) will be referred to as $X \sim HPSM_f(\alpha)$. In cases where $f(\cdot)$ is parameterized by a vector of parameters $\theta$, we will denote it as $X \sim HPSM_f(\theta, \alpha)$.

Note that when $\alpha = 1$, the segmented function of the density $f(x)$ is obtained. Similarly, if $L = 1$, the Probability Density Function (PDF) of the proportional hazard is derived. When $\alpha = L = 1$, we retrieve the PDF $f(x)$.

Based on the preceding equation, the survival function of the $i$-th group is given by:

$$S_{1-G}(x) = \exp[-\alpha\Lambda_1(x)], \quad \text{for } i = 1$$

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 April 2025 doi:10.20944/preprints202504.1011.v1



$$S_{i-G}(x) = \exp\left[-\alpha \sum_{j=1}^{i-1} \psi_j - \alpha\Lambda_i(x)\right], \quad \text{for } i = 2,3,\ldots,L$$

Similarly, the Cumulative Distribution Function (CDF) can be expressed as:

$$G_1(x) = 1 - \exp[-\alpha\Lambda_1(x)], \quad \text{for } i = 1$$

$$G_i(x) = 1 - S_{i-G}(x) = 1 - \exp\left[-\alpha \sum_{j=1}^{i-1} \psi_j - \alpha\Lambda_i(x)\right], \quad \text{for } i = 2,3,\ldots,L$$

Additionally, the hazard function of the $i$-th group is given by:

$$h_{i-G}(x) = \alpha h_i(x).$$

Now, as $S(x) \sim U(0,1)$, we can express $(0,1)$ as the union $\cup_{i=1}^{L}(u_i, u_{i+1})$, where $u_i = 1 - S_i(x_i)$. Thus, for $u \in (u_i, u_{i+1})$, the $p$-th percentile of $X$ is given by

$$t_p = \begin{cases} \Lambda_1^{-1}\left(-\frac{1}{\alpha}\log(u)\right), & \text{if } l_1 \le x < l_2, \\ \Lambda_i^{-1}\left(-\frac{1}{\alpha}\log(u) - \sum_{j=1}^{i-1}\psi_j\right), & \text{if } l_i \le x < l_{i+1}, \quad i = 2,3,\ldots,L. \end{cases} \tag{12}$$

Here, $\Lambda_i^{-1}(\cdot)$ represents the inverse function of $\Lambda_i(\cdot)$.

The moments of the segmented random variable $X$ will depend on each distribution family under discussion. In general, the $r$-th moment of $X$ can be expressed as

$$\begin{aligned} \mathbb{E}(X^r) &= \int_{-\infty}^{\infty} x^r g(x)dx \\ &= \sum_{i=1}^{L} \int_{l_i}^{l_{i+1}} x^r g_i(x)dx \\ &= \int_{l_1}^{l_2} x^r h_1(x)\exp[-\alpha\Lambda_1(x)]dx \\ &\quad +\alpha \sum_{i=1}^{L} \exp\left[-\alpha \sum_{j=1}^{i-1}\psi_j\right] \int_{l_i}^{l_{i+1}} x^r h_i(x)\exp[-\alpha\Lambda_i(x)]dx \\ &= q_1 + \sum_{i=1}^{L} \exp\left[-\alpha \sum_{j=1}^{i-1}\psi_j\right] q_i, \end{aligned} \tag{13}$$

where $q_i = \alpha \int_{l_i}^{l_{i+1}} x^r h_i(x)\exp[-\alpha\Lambda_i(x)]dx$ for $i = 1,2,3,\ldots,L$.

Similarly, the moment-generating function of $X$ is given by

$$\begin{aligned} M_X(t) &= \int_{-\infty}^{\infty} \exp(tx)g(x)dx \\ &= \sum_{i=1}^{L} \int_{l_i}^{l_{i+1}} \exp(tx)g_i(x)dx \\ &= \int_{l_1}^{l_2} h_1(x)\exp[-\alpha\Lambda_1(x) + tx]dx \\ &\quad +\alpha \sum_{i=1}^{L-1} \exp\left[-\alpha \sum_{j=1}^{i-1}\psi_j\right] \int_{l_i}^{l_{i+1}} h_i(x)\exp[-\alpha\Lambda_i(x) + tx]dx \\ &= p_1 + \sum_{i=1}^{L-1} \exp\left[-\alpha \sum_{j=1}^{i-1}\psi_j\right] p_i, \end{aligned} \tag{14}$$

where $p_i = \alpha \int_{l_i}^{l_{i+1}} h_i(x)\exp[-\alpha\Lambda_i(x) + tx]dx$ for $i = 1,2,3,\ldots,L$.

 doi:10.20944/preprints202504.1011.v1



Next, we present the PDF, CDF, survival, and hazard functions of the segmented Rayleigh, Log-logistic, Lindley, and Laplace distributions.

### 3.1. Segmented Rayleigh Proportional Hazard Distribution

For the density function of the Rayleigh distribution, we have the hazard function $h_i(t;\sigma_i) = \frac{t}{\sigma_i^2}$, which leads to the PDF of the segmented proportional hazard represented by:

$$g_i(t) = \begin{cases} \frac{\alpha t}{\sigma_1^2} \exp\left(-\frac{\alpha t^2}{2\sigma_1^2}\right), & \text{if } i = 1, \\ \frac{\alpha t}{\sigma_i^2} k_{iR}^{\alpha} \exp\left(-\frac{\alpha}{2\sigma_i^2}(t^2 - t_i^2)\right), & \text{if } t_i \le t < t_{i+1}, \ i = 2, 3, \cdots, L, \end{cases} \tag{15}$$

where $k_{iR} = \exp\left(-\frac{1}{2}\sum_{j=1}^{i} \frac{t_{j+1}-t_j}{\sigma_j^2}\right)$.

Thus, the distribution, survival, and hazard functions of this model are given by:

$$G_i(t) = \begin{cases} 1 - \exp\left(-\frac{\alpha t^2}{2\sigma_1^2}\right), & \text{if } i = 1, \\ 1 - k_{iR}^{\alpha} \exp\left(-\frac{\alpha}{2\sigma_i^2}(t^2 - t_i^2)\right), & \text{if } t_i \le t < t_{i+1}, \ i = 2, 3, \cdots, L. \end{cases} \tag{16}$$

$$S_i(t) = \begin{cases} \exp\left(-\frac{\alpha t^2}{2\sigma_1^2}\right), & \text{if } i = 1, \\ k_{iR}^{\alpha} \exp\left(-\frac{\alpha}{2\sigma_i^2}(t^2 - t_i^2)\right), & \text{if } t_i \le t < t_{i+1}, \ i = 2, 3, \cdots, L-1. \end{cases} \tag{17}$$

$$h_{iG}(t;\sigma_i) = \frac{\alpha t}{\sigma_i^2}, \quad i = 1, 2, 3, \cdots, L.$$

### 3.2. Segmented Log-logistic Proportional Hazard Distribution

The hazard function of the log-logistic distribution for the $i$-th group is given by $h_i(t) = \frac{\beta_i \lambda_i t^{\beta_i - 1}}{1+\lambda_i t^{\beta_i}}$, which leads to the expression of the PDF for the segmented proportional hazard as:

$$g_i(t) = \begin{cases} \frac{\alpha\beta_1\lambda_1 t^{\beta_1-1}}{(1+\lambda_1 t^{\beta_1})^{\alpha+1}}, & \text{if } i = 1, \\ \frac{\alpha\beta_i\lambda_i t^{\beta_i-1}}{(1+\lambda_i t^{\beta_i})^{\alpha+1}} \frac{\prod_{j=1}^{i}(1+\lambda_j t_j^{\beta_j})^{\alpha}}{\prod_{j=1}^{i-1}(1+\lambda_j t_{j+1}^{\beta_j})^{\alpha}}, & \text{if } t_i \le t < t_{i+1}, \ i = 2, 3, \cdots, L. \end{cases} \tag{18}$$

The survival function is expressed by

$$S_i(t) = \begin{cases} \frac{1}{(1+\lambda_1 t^{\beta_1})^{\alpha}}, & \text{if } i = 1, \\ \frac{1}{(1+\lambda_i t^{\beta_i})^{\alpha}} \frac{\prod_{j=1}^{i}(1+\lambda_j t_j^{\beta_j})^{\alpha}}{\prod_{j=1}^{i-1}(1+\lambda_j t_{j+1}^{\beta_j})^{\alpha}}, & \text{if } x_i \le x < x_{i+1}, \ i = 2, 3, \cdots, L, \end{cases} \tag{19}$$

as long as the hazard function is given by

$$h_{iLL}(t) = \frac{\alpha\beta_i\lambda_i t^{\beta_i-1}}{1+\lambda_i t^{\beta_i}}.$$

### 3.3. Segmented Lindley Proportional Hazard Distribution

The hazard function of the Lindley distribution's PDF is given by

$$h_i(t) = \frac{\lambda_i^2(1+t)}{1+\lambda_i(t+1)},$$

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 April 2025 doi:10.20944/preprints202504.1011.v1



then, after intensive calculations, the segmented proportional hazard PDF is given by

$$g_i(t) = \begin{cases} \frac{\alpha\lambda_1^2(1+t)(1+\lambda_1(1+t))^{\alpha-1}}{(1+\lambda_1)^\alpha} \exp(-\alpha\lambda_1 t), & \text{if } i = 1, \\ \frac{\alpha\lambda_i^2(1+t)(1+\lambda_i(1+t))^{\alpha-1}}{(1+\lambda_i(t_i+1))^\alpha} \frac{\prod_{j=1}^{i-1}(1+\lambda_j(t_{j+1}+1))^\alpha}{\prod_{j=1}^{i-1}(1+\lambda_j(t_j+1))^\alpha} \exp(-\alpha\lambda_i(t-t_i)- & \text{if } t_i \le t < t_{i+1}, \\ \alpha\sum_{j=1}^{i-1}\lambda_j(t_{j+1}-t_j)), & i = 2,3,\cdots,L. \end{cases} \tag{20}$$

Therefore, the survival function is given by

$$S_i(t) = \begin{cases} \frac{(1+\lambda_1(1+t))^\alpha}{(1+\lambda_1)^\alpha} \exp(-\alpha\lambda_1 t), & \text{if } i = 1, \\ \frac{(1+\lambda_i(1+t))^\alpha}{(1+\lambda_i(t_i+1))^\alpha} \frac{\prod_{j=1}^{i-1}(1+\lambda_j(t_{j+1}+1))^\alpha}{\prod_{j=1}^{i-1}(1+\lambda_j(t_j+1))^\alpha} \exp(-\alpha\lambda_i(t-t_i)- & \text{if } t_i \le t < t_{i+1}, \\ \alpha\sum_{j=1}^{i-1}\lambda_j(t_{j+1}-t_j)), & i = 2,3,\cdots,L, \end{cases} \tag{21}$$

where the segmented proportional hazard function is obtained

$$h_{iH}(t) = \frac{\alpha\lambda_i^2(1+t)}{1+\lambda_i(t+1)}.$$

*3.4. Segmented Laplace Proportional Hazard Distribution*

The PDF of a Laplace random variable with parameters $\mu$ and $\lambda$ is given by:

$$f_L(x) = \begin{cases} \frac{\lambda}{2}\exp(\lambda(x-\mu)), & \text{if } x < \mu, \\ \frac{\lambda}{2}\exp(-\lambda(x-\mu)), & \text{if } x \ge \mu, \end{cases} \tag{22}$$

where the survival and hazard functions are expressed, respectively, by

$$S_L(x) = \begin{cases} 1-\frac{1}{2}\exp(\lambda(x-\mu)), & \text{if } x < \mu, \\ 1-\frac{1}{2}\exp(-\lambda(x-\mu)), & \text{if } x \ge \mu, \end{cases} \tag{23}$$

and

$$h_L(x) = \begin{cases} \frac{\frac{\lambda}{2}\exp(\lambda(x-\mu))}{1-\frac{1}{2}\exp(\lambda(x-\mu))}, & \text{if } x < \mu, \\ \lambda, & \text{if } x \ge \mu. \end{cases} \tag{24}$$

Then, for $i = 1$, the proportional hazard density is obtained from Laplace

$$g_1(x) = \begin{cases} \frac{\alpha\lambda_1}{2} \frac{\exp(\lambda_1(x-\mu_1))(1-\frac{1}{2}\exp(\lambda_1(x-\mu_1)))^{\alpha-1}}{(1-\frac{1}{2}\exp(\lambda_1(x_1-\mu_1)))^\alpha}, & \text{if } x < \mu_1, \\ \frac{\alpha\lambda_1}{2}\exp\left(-\frac{\alpha\lambda_1}{2}(x-x_1)\right), & \text{if } x \ge \mu_1, \end{cases} \tag{25}$$

while if $x_i \le x < x_{i+1},\ i = 2,3,\cdots,L$

$$g_i(x) = \begin{cases} \frac{\alpha\lambda_i}{2} \frac{\exp(\lambda_i(x-\mu_i))(1-\frac{1}{2}\exp(\lambda_i(x-\mu_i)))^{\alpha-1}}{(1-\frac{1}{2}\exp(\lambda_i(x_i-\mu_i)))^\alpha} \prod_{j=1}^{i-1}\left(\frac{1-\frac{1}{2}\exp(\lambda_j(x_{j+1}-\mu_j))}{1-\frac{1}{2}\exp(\lambda_j(x_j-\mu_j))}\right)^\alpha, & \text{if } x < \mu_i, \\ \frac{\alpha\lambda_i}{2}\exp\left(-\frac{\alpha\lambda_i}{2}(x-x_i) - \frac{\alpha}{2}\sum_{j=1}^{i-1}\lambda_j(x_{j+1}-x_j)\right), & \text{if } x \ge \mu_i. \end{cases} \tag{26}$$

The survival function of the segmented proportional hazard for $i = 1$ is given by

$$S_1(x) = \begin{cases} \frac{(1-\frac{1}{2}\exp(\lambda_1(x-\mu_1)))^\alpha}{(1-\frac{1}{2}\exp(\lambda_1(x_1-\mu_1)))^\alpha}, & \text{if } x < \mu_1, \\ \exp\left(-\frac{\alpha\lambda_1}{2}(x-x_1)\right), & \text{if } x \ge \mu_1, \end{cases} \tag{27}$$

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 April 2025 doi:10.20944/preprints202504.1011.v1



while if $x_i \leq x < x_{i+1}, \ \ i = 2, 3, \cdots, L$

$$S_i(x) = \begin{cases} \frac{(1-\frac{1}{2}\exp(\lambda_i(x-\mu_i)))^\alpha}{(1-\frac{1}{2}\exp(\lambda_i(x_i-\mu_i)))^\alpha} \prod_{j=1}^{i-1}\left(\frac{1-\frac{1}{2}\exp(\lambda_j(x_{j+1}-\mu_j))}{1-\frac{1}{2}\exp(\lambda_j(x_j-\mu_j))}\right)^\alpha, & \text{if } x < \mu_i, \\ \exp\left(-\frac{\alpha\lambda_i}{2}(x-x_i) - \frac{\alpha}{2}\sum_{j=1}^{i-1}\lambda_j(x_{j+1}-x_j)\right), & \text{if } x \geq \mu_i, \end{cases} \tag{28}$$

therefore, the hazard function is given by

$$h_i(x) = \begin{cases} \frac{\alpha\lambda_1}{2}\frac{\exp(\lambda_1(x-\mu_1))}{1-\frac{1}{2}\exp(\lambda(x-\mu_1))}, & \text{if } x < \mu_i, \\ \alpha\lambda_i, & \text{if } x \geq \mu_i. \end{cases} \tag{29}$$

## 4. Estimation

For the estimation of the parameters of the segmented proportional hazard distribution, we will employ the maximum likelihood estimate. Suppose that $g(x;\theta,\alpha)$ represents the PDF of the segmented proportional hazard distribution indexed by the $\alpha$ parameter and the parameter vector $\theta = (\theta_1, \theta_2, \cdots, \theta_i, \cdots, \theta_L)$, where for $i = 1, 2, \cdots, L$, we have $\theta_i = (\theta_{i1}, \theta_{i2}, \cdots, \theta_{ia}, \cdots, \theta_{ib})$. This sets the framework for defining,

$$I_i(x) = \begin{cases} 1, & \text{if } l_i \leq x < l_{i+1}, \\ 0, & \text{otherwise.} \end{cases} \tag{30}$$

The PDF $g(x;\theta,\alpha)$ can be expressed in the following form:

$$g(x;\theta,\alpha) = \alpha\prod_{i=1}^{L} h_i^{I_i(x)}(x,\theta_i)\exp\left(-\alpha\sum_{i=1}^{L} I_i(x)\Lambda_i(x,\theta_i) - \alpha\sum_{i=2}^{L}\sum_{j=1}^{i-1} I_i(x)\psi_j(\theta_j)\right), \tag{31}$$

Assuming that the elements of the partition $l_1, l_2, \ldots, l_L$ are known and given a sample of size $n = n_1 + n_2 + \cdots + n_i + \cdots + n_L$, obtained from observations $x_1, x_2, \ldots, x_n$, the likelihood function of $(\theta, \alpha)$ is expressed as follows:

$$\begin{aligned} L(\theta,\alpha) &= \alpha^n \prod_{i=1}^{L}\prod_{l_i \leq x_k < l_{i+1}} h_i(x_k,\theta_i)\exp\Bigg(-\alpha\sum_{i=1}^{L}\sum_{l_i \leq x_k < l_{i+1}}\Lambda_i(x_k,\theta_i) - \\ &\quad \alpha\sum_{j=1}^{L-1}\left(\psi_j(\theta_j)\sum_{i=j+1}^{L} n_i\right)\Bigg). \end{aligned} \tag{32}$$

Subsequently, the log-likelihood function is given by

$$\begin{aligned} \ell(\theta,\alpha) &= n\log(\alpha) + \sum_{i=1}^{L}\sum_{l_i \leq x_k < l_{i+1}}\log(h_i(x_k,\theta_i)) - \alpha\sum_{i=1}^{L}\sum_{l_i \leq x_k < l_{i+1}}\Lambda_i(x_k,\theta_i) - \\ &\quad \alpha\sum_{j=1}^{L-1}\left(\psi_j(\theta_j)\sum_{i=j+1}^{L} n_i\right). \end{aligned} \tag{33}$$

Consequently, the components of the score function are given by:

$$U(\theta_{ia}) = \sum_{l_i \leq x_k < l_{i+1}} \frac{h'_{\theta_{ia}}}{h_i(x_k,\theta_i)} - \alpha\sum_{l_i \leq x_k < l_{i+1}}\Lambda'_{\theta_{ia}} - \alpha\psi'_{\theta_{ia}}\sum_{i=j+1}^{L} n_i, \tag{34}$$

where $i = 1, 2, \ldots, L$ and $a = 1, 2, \ldots, b$. Here, $h'_{\theta_{ia}}$ denotes the first derivative of $h_i(x_k, \theta_i)$ with respect to $\theta_{ia}$, and the same notation is used for $\Lambda_i$ and $\psi_i$.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 April 2025 doi:10.20944/preprints202504.1011.v1



The derivative with respect to $\alpha$ is given by:

$$U(\alpha) = \frac{n}{\alpha} - \sum_{i=1}^{L} \sum_{l_i \leq x_k < l_{i+1}} \Lambda_i(x_k, \theta_i) - \sum_{j=1}^{L-1} \left( \psi_j(\theta_j) \sum_{i=j+1}^{L} n_i \right). \tag{35}$$

Equating (34) and (35) to zero gives the score equations, the solution of which, using iterative numerical methods, leads to the maximum likelihood estimates of the parameters of the segmented proportional hazard model. It should be noted that there are existing statistical software and libraries, such as the R project, that provide functions like `optim` and `nlm`, among others, which can be used to obtain maximum likelihood estimates for any proposed model. From the equation $U(\alpha) = 0$, we obtain

$$\alpha(\theta_i) = \frac{n}{\sum_{i=1}^{L} \sum_{l_i \leq x_k < l_{i+1}} \Lambda_i(x_k, \theta_i) + \sum_{j=1}^{L-1} \left( \psi_j(\theta_j) \sum_{i=j+1}^{L} n_i \right)}. \tag{36}$$

Subsequently, substituting (36) into (33) yields the profiled log-likelihood of $\theta_i$, removing the constant term:

$$\begin{aligned} \ell_p(\theta) \quad = \quad & -n \log \left( \sum_{i=1}^{L} \sum_{l_i \leq x_k < l_{i+1}} \Lambda_i(x_k, \theta_i) + \sum_{j=1}^{L-1} \left( \psi_j(\theta_j) \sum_{i=j+1}^{L} n_i \right) \right) + \\ & \sum_{i=1}^{L} \sum_{l_i \leq x_k < l_{i+1}} \log(h_i(x_k, \theta_i)) + n(\log(n) - 1). \end{aligned} \tag{37}$$

We now proceed to determine the observed and expected information matrices in general terms. These matrices are of significant interest because, for large samples, their inverses evaluated at the maximum likelihood estimates (MLE) of the parameters yield the asymptotic covariance matrix. This matrix, in turn, is used to determine the standard errors of the parameter estimates. Under the assumption of asymptotic normality and certain regularity conditions, this allows for the calculation of confidence intervals for the parameters and the execution of hypothesis tests. Furthermore, certain optimization methods use the information matrix to compute the MLEs of the parameters. The existence of these matrices depends on the specific configuration of $f$ and $F$ in each case.

Given that the elements of the information matrix are defined as the negative second derivatives of the log-likelihood function's elements, these elements will depend on the properties of $f$ and $F$ as formulated in the particular case's PDF. In this sense, the conditions and properties of the observed and expected information matrices are contingent upon these configurations. Assuming that the second derivatives of $h_i(x_k, \theta_i)$, $\Lambda_i(x_k, \theta_i)$, and $\psi_j(\theta_j)$ exist (which is guaranteed since both $f$ and $F$ are continuous, as well as $h$), the elements of the observed information matrix ($K(\theta, \alpha)$) are given by

$$\begin{aligned} \kappa_{\theta_{ia}\theta_{ia'}} \quad = \quad & -\sum_{l_i \leq x_k < l_{i+1}} \frac{h''_{\theta_{ia}\theta_{ia'}} - h'_{\theta_{ia}} h'_{\theta_{ia'}}}{h_i^2(x_k,\theta_i)} + \alpha \sum_{l_i \leq x_k < l_{i+1}} \frac{\Lambda''_{\theta_{ia}\theta_{ia'}} - \Lambda'_{\theta_{ia}} \Lambda'_{\theta_{ia'}}}{\Lambda_i^2(x_k,\theta_i)} + \\ & \alpha \frac{\psi''_{\theta_{ia}\theta_{ia'}} - \psi'_{\theta_{ia}} \psi'_{\theta_{ia'}}}{\psi_i^2(x_k,\theta_i)} \sum_{i=j+1}^{L} n_i, \end{aligned} \tag{38}$$

where $h''_{\theta_{ia}\theta_{ia'}}$ represents the second derivative of $h_i(x_k, \theta_i)$, with respect to $\theta_{ia}$ and $\theta_{ia'}$ respectively, the same notation is used for $\Lambda_i$ and $\psi_i$.

$$\kappa_{\theta_{ia}\alpha} = \sum_{l_i \leq x_k < l_{i+1}} \Lambda'_{\theta_{ia}} + \psi'_{\theta_{ia}} \sum_{i=j+1}^{L} n_i, \tag{39}$$

and

$$\kappa_{\alpha\alpha} = \frac{n}{\alpha^2}. \tag{40}$$

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 April 2025 doi:10.20944/preprints202504.1011.v1



The elements of the information matrix ($I(\theta,\alpha)$) are obtained by calculating the expectation of the elements of the observed information matrix, denoted as $I_{\gamma_l\gamma_{l'}}$, where $\gamma_l$ spans the parameters $\theta_{ia}$ for $i = 1, 2, \cdots, L$ and $a = 1, 2, \cdots, b$, as well as the parameter $\alpha$.

Under conditions of regularity, employing asymptotic theory for the maximum likelihood estimator vector of the parameter vector $(\theta, \alpha)$, we observe that as $n \to \infty$, $K(\theta,\alpha) \to I(\theta,\alpha)$. Using the same argument of stochastic convergence, it follows that $(\hat{\theta}, \hat{\alpha})$ converges in distribution to a normal distribution, that is,

$$\sqrt{n}((\hat{\theta},\hat{\alpha}) - (\theta,\alpha)) \xrightarrow{D} N_{Lp+1}(\mathbf{0}, I^{-1}(\theta,\alpha)).$$

### *4.1. Censoring*

We now introduce a point of censoring, denoted as $C$, in the distribution of the segmented proportional hazard. In studies of survival data, it is common to encounter various types of censoring [18]. Therefore, for the $k$-th observation, the censored variable in the segmented proportional hazard is defined as $Y_k = \min(X_k, C_k)$, where $k = 1, 2, \cdots, n$. It is assumed that the value of censoring for the $k$-th individual does not depend on $C_k$. To this end, we define the indicator variable for censoring as

$$\delta_k = \begin{cases} 0, & \text{if the individual is censored,} \\ 1, & \text{otherwise} \end{cases} \tag{41}$$

The density function for the case of right-censoring ($\delta_k = 0$ for $x_k \geq C$) for the $i$-th group is given by:

$$\begin{aligned} g_{Ci}(x_k) &= g_{Ci}^{\delta_k}(x_k) S_{i-G}^{1-\delta_k}(C) \\ &= \alpha^{\delta_k} h_i^{\delta_k}(x_k) \exp\left[-\alpha\delta_k \sum_{j=1}^{i-1} \psi_j - \alpha\delta_k \Lambda_i(x_k) + (1-\delta_k)\left(-\alpha \sum_{j=1}^{i-1}\psi_j - \alpha\Lambda_i(C)\right)\right] \\ &= \alpha^{\delta_k} h_i^{\delta_k}(x_k) \exp\left[-\alpha \sum_{j=1}^{i-1} \psi_j - \alpha\delta_k\Lambda_i(x_k) - \alpha(1-\delta_k)\Lambda_i(C)\right]. \end{aligned} \tag{42}$$

Note that if $\alpha = 1$, the segmented censored density of $f(x)$ is obtained. Similarly, if $L = 1$, the censored PDF of the proportional hazard is achieved, and if $\alpha = L = 1$, the censored PDF of $f(x)$ is obtained.

For a random sample, $x_1, x_2, \cdots, x_n$, the log-likelihood function is given by

$$\begin{aligned} \ell_C(\theta,\alpha) = \;& \log(\alpha) \sum_{k=1}^{n} \delta_k + \sum_{i=1}^{L} \sum_{l_i \leq x_k < l_{i+1}} \delta_k \log(h_i(x_k,\theta_i)) \\ & -\alpha \sum_{i=1}^{L} \sum_{l_i \leq x_k < l_{i+1}} \delta_k \Lambda_i(x_k,\theta_i) \\ & -\alpha \sum_{i=1}^{L} \sum_{l_i \leq x_k < l_{i+1}} (1-\delta_k)\Lambda_i(C,\theta_i) \\ & -\alpha \sum_{j=1}^{L-1} \left(\psi_j(\theta_j) \sum_{i=j+1}^{L} n_i\right), \end{aligned} \tag{43}$$

where $\delta_k = 0$ for $x_k \geq C$. The estimation method remains an iterative process, using the score equations as in the case of the uncensored scenario. Profiled likelihood can also be employed since, for the $\alpha$ parameter, it follows that:

$$\alpha(\theta) = \frac{\sum_{k=1}^{n} \delta_k}{\sum_{i=1}^{L} \sum_{l_i \leq x_k < l_{i+1}} (\delta_k \Lambda_i(x_k,\theta_i) + (1-\delta_k)\Lambda_i(C,\theta_i)) + \sum_{j=1}^{L-1} \left(\psi_j(\theta_j) \sum_{i=j+1}^{L} n_i\right)}. \tag{44}$$

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 April 2025 doi:10.20944/preprints202504.1011.v1



Until now, it has been assumed that the number of exchange points, denoted as $L$, and the values $l_1,\ l_2,\cdots,l_L$ are known. However, this assumption does not hold in practice, necessitating the exploration of alternatives that can help identify these points or provide their estimates. The first approach to address this challenge involves analyzing potential change points in the empirical cumulative hazard distribution derived from the data and considering these points as the actual change points in the distribution.

Another possible alternative involves estimating the change points through an iterative process. This process entails creating partitions for $i = 1, 2, \cdots, L$ and estimating potential change points $l_1, l_2, \cdots, l_L$ within each partition. Subsequently, the log-likelihood function is maximized for each partition to determine the most suitable one. Comparison criteria for models, such as AIC, BIC, and CAIC, can also be employed to assist in partition selection.

In reality, it is generally not expected to have a partition with numerous change points. Instead, real-world data tends to exhibit multiple instances of subtle changes in the cumulative hazard distribution. Therefore, the partition with a limited number of change points is often more representative of the underlying distribution.

## 5. Illustration

Here, we illustrate the segmented proportional hazard model using a real dataset. We consider the IQ scores data of 52 white males hired by a large insurance company in 1971, known as the Otis IQ Scores dataset. These data were previously studied by [19] [see also [12]]. The cumulative hazard function based on this dataset is provided in Figure 1 - (c).

We fit various families of segmented distributions to the data, including segmented Laplace, segmented Laplace proportional hazard, segmented half-normal, and segmented half-normal proportional hazard. Additionally, we adjust other distributions, such as the Laplace distribution, skew-normal, Gamma, normal, and the mixture of normals (MN).

To discriminate and choose the best model among the aforementioned options, we employed the Akaike Information Criterion (AIC) [1], the corrected Akaike Information Criterion (CAIC) proposed by [4], and the Hannan-Quinn criterion (HQC) introduced by [13].

These criteria are defined as follows:

$$AIC = -2\ell(\hat{\theta}) + 2p, \quad CAIC = -2\ell(\hat{\theta}) + \frac{2n(p+1)}{n-p-2} \quad \text{and} \quad HQIC = -2\ell(\hat{\theta}) + 2p\log(\log(n)).$$

Table 1 displays the AIC, CAIC, and HQIC values for the various fitted models, along with the estimated partition values for the case of $L = 2$.

**Table 1.** AIC, CAIC and HQIC criteria for the fitted models.

| Distribution | $\hat{l}_1$ | AIC | BIC | HQIC |
|---|---|---|---|---|
| Laplace | | 641.708 | 643.9977 | 643.6944 |
| LS | 108.0222 | 641.6922 | 644.7422 | 646.6569 |
| SLPH | 111.9693 | 637.8516 | 641.2693 | 643.8092 |
| HNS | 99.9687 | 682.7748 | 685.2626 | 685.7536 |
| HNSHP | 99.9687 | 683.4016 | 686.1423 | 687.3734 |
| SN | | 644.612 | 647.0998 | 647.5908 |
| Gamma | | 642.9468 | 645.2360 | 644.9327 |
| Normal | | 646.3490 | 648.6382 | 648.3349 |
| MN | | 646.4074 | 649.4574 | 651.3721 |

The Tables 2 and 3 contain the maximum likelihood estimates of the fitted models.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 April 2025 doi:10.20944/preprints202504.1011.v1



**Table 2.** Estimation of the parameters, along with their standard errors, for the Laplace, LS, SLPH, HNS, and HNSHP models.

| Estimate | Laplace | LS | SLPH | HNS | HNSHP |
|---|---|---|---|---|---|
| $\hat{\mu}_1$ | 112.0004 | 110.6122 | 105.9556 | | |
| | (0.0308) | (1.1075) | (0.4240) | | |
| $\hat{\mu}_2$ | | 113.3408 | 111.9993 | | |
| | | (0.9016) | (0.1693) | | |
| $\hat{\sigma}_1$ | 7.1859 | 5.6885 | 5.9306 | 1382.2189 | 272.5887 |
| | (0.7706) | (1.3427) | (1.1603) | (598.4787) | (108.9069) |
| $\hat{\sigma}_2$ | | 7.3183 | 2.1697 | 6.2681 | 17.7197 |
| | | (1.0452) | (0.5466) | (2.5036) | (0.1300) |
| $\hat{\alpha}$ | | | 0.3093 | | 0.1867 |
| | | | (0.0638) | | (0.0214) |

**Table 3.** Estimation of the parameters, along with their standard errors, for the SN, Gamma, Normal, and MN models.

| Estimate | SN | Gamma | Normal | MN |
|---|---|---|---|---|
| $\hat{\mu}_1$ | 105.724 | | 112.862 | 111.6556 |
| | (3.119) | | (1.028) | (2.1347) |
| $\hat{\mu}_2$ | | | | 113.7674 |
| | | | | (2.5908) |
| $\hat{\sigma}_1$ | 11.910 | 1.2396 | 9.589 | 35.9973 |
| | (2.076) | (0.1880) | (0.2349) | (42.1964) |
| $\hat{\sigma}_2$ | | | | 130.2105 |
| | | | | (81.5536) |
| $\hat{\alpha}$ | 1.143 | 139.9127 | | 0.4278 |
| | (0.684) | (21.1856) | | (0.6189) |

Initially, the Laplace model is compared to the SLPH model using hypothesis tests.

$$H_0 : (L, \alpha) = (1, 1) \text{ versus } H_1 : (L, \alpha) \neq (1, 1).$$

Using the likelihood ratio statistic,

$$\Lambda = \frac{\ell_L(\hat{\boldsymbol{\theta}})}{\ell_{SLPH}(\hat{\boldsymbol{\theta}})}$$

we obtain

$$-2\log(\Lambda) = 11.857,$$

The result is greater than the value of $\chi^2_{2.95\%} = 5.99$. Thus, the SLPH model stands as a suitable alternative for fitting the dataset. Additionally, the SLPH model is compared to the LS model, and the Laplace model is compared to the LS model using hypothesis tests.

$$H_{01} : \alpha = 1 \text{ versus } H_{11} : \alpha \neq 1, \quad \text{and} \quad H_{02} : L = 1 \text{ versus } H_{12} : L \neq 1,$$

respectively, using the likelihood ratio statistics

$$\Lambda_1 = \frac{\ell_{LS}(\theta)}{\ell_{SLPH}(\theta)} \quad \text{and} \quad \Lambda_2 = \frac{\ell_L(\theta)}{\ell_{LS}(\theta)}.$$

After numerical evaluations, we obtain

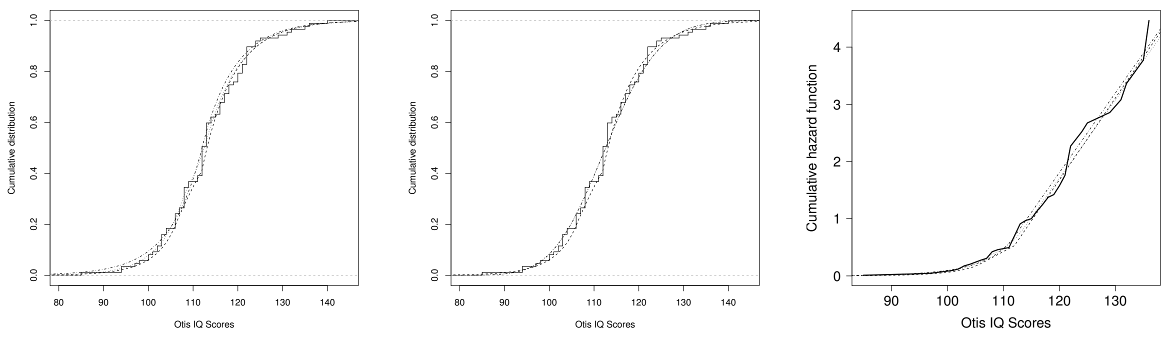
$$-2\log(\Lambda_1) = 5.8406 \quad \text{and} \quad -2\log(\Lambda_2) = 6.0164,$$

 doi:10.20944/preprints202504.1011.v1



This result is greater than the value of $\chi^2_{1,95\%} = 3.84$. The SLPH model exhibits the best fit among the other models.

Figure 1 - (a), (b), and (c) depict the CDF of the adjusted distributions and the cumulative hazard function derived from parameter estimates. These figures demonstrate the good fit of the segmented Laplace proportional hazard model (SLPH).




(a) (b) (c)

**Figure 1.** (a) Empirical CDF (solid line), SLPH (dashed line), LS (dotted line), and Laplace (dotted and dashed line). (b) Empirical CDF (solid line), SLPH (dashed line), Gamma (dotted line), and SN (dotted and dashed line). (c) Cumulative hazard function: empirical (solid line), SLPH (dashed line), LS (dotted line), and Laplace (dotted line and dashed strokes).

## 6. Concluding Remarks

This study addresses the growing need for advanced statistical models in survival analysis, particularly when dealing with data that may exhibit abrupt changes in the hazard rates during the observation period. Traditional models often fail to capture these complexities, especially when handling skewed distributions and heavy-tailed data. The segmented proportional hazard models we propose in this paper contribute to filling this gap by offering enhanced flexibility for accurately modeling such behaviors.

Our introduction of segmented proportional hazard models, including the Rayleigh, log-logistic, Lindley, and Laplace distributions, provides a new framework for capturing the nuances in survival data, particularly in medical research where sudden shifts in failure rates are common. By incorporating a censoring mechanism, our approach also addresses a critical aspect of real-world survival data that has previously been overlooked in segmented models. This makes our model well-suited for medical applications, such as the analysis of time-to-event data or survival times in clinical trials.

Furthermore, our development of the maximum likelihood estimation procedure for these models, coupled with its proven asymptotic properties, strengthens the theoretical foundation of segmented hazard models in survival analysis. The ability of our method to adapt to varying data patterns is supported by real data applications, such as the analysis of IQ scores, where the segmented Laplace hazard proportional model demonstrated superior performance. This finding underscores the model's practical utility in medical research, particularly for analyzing datasets that exhibit non-normality, censoring, and abrupt changes in hazard rates.

**Author Contributions:** Conceptualization, G.M., R.B.A.F., and C.B.; data curation, G.M., R.B.A.F., and C.B.; formal analysis, G.M. and R.B.A.F.; funding acquisition, C.B.; investigation, G.M., R.B.A.F., and C.B.; methodology, G.M.; project administration, G.M. and C.B.; resources, G.M., R.B.A.F., and C.B.; supervision, G.M., R.B.A.F., and C.B.; visualization, C.B.; writing—original draft, G.M., R.B.A.F., and C.B.; writing—review and editing, C.B. All authors have read and agreed to the published version of the manuscript.

**Funding:** This project was supported in a collaborative and non-monetary way by the Universidad de Córdoba through the SFCB-07-23 project (Modelos de regresión segmentados con respuestas distribuidas Hazard proporcional como herramienta para el análisis estadístico de datos), and by the Institución Universitaria ITM, which contributed time and resources for the researchers.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** G.M. acknowledges the support provided by Universidad de Córdoba, Montería, Colombia. R.B.A.F. thanks the Universidade Federal do Ceará-Brazil. C.B. expresses his sincere gratitude to the Institución Universitaria ITM.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

| | |
|---|---|
| PHD | Proportional hazard distribution |
| HPSM | Proportional hazard segmented model |
| MN | Mixture of normals |
| HNS | Segmented half-normal |
| HNSHP | Segmented half-normal proportional hazard |
| LS | Segmented Laplace |
| SLPH | Segmented Laplace proportional hazard |
| SM | Segmented model of hazard function |
| SN | Skew normal |

## References

1. Akaike H. A new look at the statistical model identification. *IEEE Trans. Autom. Control* **1974**, *19*, 716–723.
2. Agamez-Montalvo; G.S. Modelos de mistura finita usando a classe de distribuições alpha potência. Doctoral thesis, Universidade de Sao paulo, Brasil.
3. Bathiany, S.; Dijkstra, H.; Crucifix, M.; Dakos, V.; Brovkin, V.; Williamson, M.S.; Lenton, T.M.; Scheffer, M. Beyond bifurcation: using complex models to understand and predict abrupt climate change *Dynamics and Statistics of the Climate System* **2016**, *1(1)*, 2059-6987.
4. Cavanaugh, J. E. Unifying the derivations for the Akaike and corrected Akaike information criteria. *Statistics & Probability Letters* **1997**, *33(2)*, 201-208.
5. Chen, X.; Baron, M. Change-point analysis of survival data with application in clinical trials. *Open Journal of Statistics* **2014**, *4(09)*, 663.
6. Chen, Wx.; Long, Cx.; Yang, R.; Dong-sen Y. Maximum Likelihood Estimator of the Location Parameter under Moving Extremes Ranked Set Sampling Design. *Acta Math. Appl. Sin. Engl.* **2021**, *37*, 101–108.
7. Coelho-Barros, E.A.; Achcar J.A.; Martinez, E.Z.; Davarzani, N.; Grabsch, H.I. Bayesian Inference for the Segmented Weibull Distribution. *Colombian Journal of Statistics* **2019**, *42(2)*, 225–243.
8. Conners, T.E. Segmented models for stress-strain diagrams. *Wood Sci.Technol.* **1989**, *23*, 65–73.
9. Feigl, P.; Zelen, M. Estimation of exponential survival probabilities with concomitant information. *Biometrics* **1965**, *21*, 826-838.
10. Friedman, M. Piecewise exponential models for survival data with covariates. *Ann Statist* **1982**, *10*, 101-113.
11. Gómez, Y. M.; Gallardo, D. I.; Arnold, B. C. The power piecewise exponential model. *Journal of Statistical Computation and Simulation* **2018**, *88(5)*, 825-840.
12. Gupta, R.C.; Brown, N. Reliability studies of skew normal distribution and its application to a strengthstress model. *Commun Stat Theory Method* **2001**, *30(11)*, 2427-2445.
13. Hannan, E.; Quinn, B. The determinationof the order of an autoregression. *Journal of the Royal Statistical Society. Series B* **1979**, *41(2)*, 190-195.
14. Li, H.; Zuo, H.; Su, Y.; Xu, J; Yin, Y. Study on segmented distribution for reliability evaluation. *Chin J Aeronaut* **2016**, *30(1)*, 310-329.
15. Martínez-Flórez, G.; Moreno-Arenas, G.; Vergara-Cardoso, S. Properties and Inference for Proportional Hazard Models. *Colombian Journal of Statistics* **2013**, *36*, 95–114.

 doi:10.20944/preprints202504.1011.v1




16. Mazucheli, J.; Coelho-Barros, E.A.; Achcar, J.A. Inferences for the change-point of the exponentiated Weibull hazard function. *REVSTAT* **2012**, *10(3)*, 309-322.
17. Perevaryukha, A.Y. Modeling Abrupt Changes in Population Dynamics with Two Threshold States. *Cybern Syst Anal* **2016**, *52*, 623-630.
18. Reich, B.J.; Smith, L.B. Bayesian quantile regression for censored data. *Biometrics* **2013**, *69(3)*, 651-660.
19. Roberts H.V. Data analysis for managers with minitab. Scientific Press, Redwood City. **1988**.
20. R Core Team R: A language and environment for statistical computing. R Foundation for Statistical Computing. SVienna, Austria. **2022**.
21. Samawi, H., Yu, L., & Yin, J. On Cox proportional hazards model performance under different sampling schemes. *PLOS ONE*, **2023**, *18*(4), e0278700.
22. Zhang, S.; Zhang, H.; Li, J.; Li, J. AGCT: a hybrid model for identifying abrupt and gradual change in hydrological time series *Environ Earth Sci* **2019**, *78(43)*, https://doi.org/10.1007/s12665-019-8453-3.