Revista ARACÊ
ISSN: 2358-2472

APPLICATION OF BIVARIATE CONDITIONAL INVERSE-GAUSSIAN DISTRIBUTION TO CANCER SURVIVAL ANALYSIS

https://doi.org/10.56238/arev7n5-141

Date of submission: 04/08/2025 **Date of publication:** 05/08/2025

Olubunmi Temitope Olorunpomi[1], Toluwani John Dare[2], Oluwatosin Falebita[3] and Ayodeji Benjamin Olorungbon[4]

ABSTRACT

This investigation explores the application of the Bivariate Inverse-Gaussian Distribution (BCIGD) to model the time to relapse (TR) and time to death (TD) with sequential medical events. A moderate negative linear correlation between TR and TD was pragmatic in the scatter plot. Thus, the inverse relationship suggests that patients with a longer TR tend to have shorter TD and vice versa. The histogram plot showed that TR has a bimodal distribution with two peaks; while TD has a multimodal distribution. The Kolmogorov-Smirnov tests suggested that both TR and TD data follow a normal dis-tribution for p-values greater than a significance level of 0.05. The prominent peak of the contour plot indicates the region of highest probability density for TR and TD is centred on 0.5. The rapid decay of density values beyond 1.5 for TR and TD aligns with the inverse-Gaussian distribution's tendency to concentrate values near the mean with a strident drop-off as values diverge from this central region. These findings support the use of the inverse-Gaussian distribution in this area and highlight the potential of numerical integration to understand probability density behaviour in survival analysis and medi-cal event modelling.

Keywords: Bivariate Inverse-Gaussian Distribution. Kolmogorov-Smirnov test. Medical Event Modelling. Sequential Medical Event. Time to Death. Time to Relapse.

[1] Department of Statistics, Federal University Lokoja, Lokoja-Kogi State, P.M.B 1154, NIGERIA.
E-mail: olubunmi.olorunpomi@fulokoja.edu.ng
[2] Department of Statistics, Federal University Lokoja, Lokoja-Kogi State, P.M.B 1154, NIGERIA.
[3] Department of Computer Science, Colorado State University, 80523, UNITED STATES.
[4] Department of Statistics, Federal University Lokoja, Lokoja-Kogi State, P.M.B 1154, NIGERIA.

Revista ARACÊ
ISSN: 2358-2472

INTRODUCTION

Statistical distributions are essential tools in data analysis, offering a mathematical basis for describing how data points are spread across different values. The inverse-Gaussian distribution is particularly notable for modelling time-to-event data, especially within stochastic processes. Characterized by its mean and shape parameters, the Weibull distribution effectively represents the distribution of times between events in a Poisson process with drift, making it highly relevant in various fields such as survival analysis, reliability engineering, and econometrics ([3]).

Extending the concept of the Wald distribution to handle dependencies between two variables leads to the Bivariate Conditional Inverse-Gaussian Distribution (BCIGD). This extension is crucial in real-world applications where variables are often interdependent. For example, in reliability engineering, the time until failure of a system may depend on various correlated factors such as load and temperature. Similarly, in financial modelling, the time to reach a certain financial threshold can be influenced by multiple interacting economic indicators. The BCIGD provides a robust framework for modelling such interdependencies, which are increasingly recognized as essential in fields ranging from finance to environmental science ([1], [9], [5]).

The importance of understanding relationships between random variables cannot be overstated. This is particularly true in fields like economics, engineering, and biomedical sciences, where variables are often interdependent, and understanding these dependencies is key to accurate modelling and prediction. The BCIGD extends the univariate Wald distribution to a bivariate context, allowing for the modelling of one variable conditional on the value of another. This approach provides a more nuanced and accurate representation of the underlying data, which is critical in improving predictive accuracy and understanding complex systems. Recent studies underscore the value of such models in a variety of applications, including risk assessment, resource management, and predictive analytics ([5]).

Moreover, the BCIGD’s ability to handle dependencies between variables makes it a valuable tool for researchers and practitioners. By providing a joint probability density function that incorporates the conditional relationships between variables, the BCIGD enables more precise estimations and inferences. This is particularly useful in areas like survival analysis, where the time until an event occurs can be influenced by multiple, interdependent factors. Leveraging the BCIGD can lead to significant advancements in

Revista ARACÊ
ISSN: 2358-2472

both theoretical and applied statistics, as demonstrated by recent advancements in the field ([2], [7]).

The rationale for this assessment description is to build upon the introductory research of [2] and [5] in order to disentangle the complexities and implications of bivariate conditional extension. This work not only seeks to enhance our theoretical understanding but also to demonstrate practical applications across various disciplines and further solidifying the BCIGD's role in modern statistical analysis. ([6])

MATERIALS AND METHODS

This study focuses on developing and validating the Bivariate Conditional Inverse-Gaussian Distribution (BCIGD) as a robust statistical tool for modeling the conditional dependencies between two interdependent variables. Extending families of distributions to enhance flexibility and applicability is a well-established approach in statistical modeling, particularly for capturing the complex dynamics inherent in real-world data. BCIGD represents such an extension, providing a powerful tool for analyzing interdependent lifetime data across various fields, including medicine, engineering, finance, and public health. The BCIGD's ability to model the conditional relationships between two variables offers a significant advantage in accurately reflecting the underlying processes that drive these dependencies.

BIVARIATE CONDITIONAL INVERSE-GAUSSIAN DISTRIBUTION

The Bivariate Inverse-Gaussian Distribution (BIGD) extends the traditional Wald distribution to model the joint behaviour of two interdependent variables. This extension is crucial for accurately capturing the dependencies and interactions between variables in real-world scenarios. In the BCIGD, if a random variable X follows inverse-Gaussian distribution with parameters μ_X and λ_X, and another variable Y given X = x follows inverse-Gaussian distribution with parameters dependent on x, the joint probability density function (PDF) of X and Y can be expressed as:

$$f_{X,Y}(x,y) = f_X(x) \cdot f_{Y|X}(x,y) \qquad (1)$$

Revista ARACÊ
ISSN: 2358-2472

This formulation allows for the modelling of one variable conditional on the other, providing a flexible and nuanced approach to understanding their relationship. ([4], [7], [8]). The joint PDF under the BIGD is given by:

$$f_{X,Y}(x,y) = \frac{\lambda_X \lambda}{2\pi (xy)^{3/2}} \exp\left(-\frac{\lambda_X (x-\mu_X)^2}{2\mu_X^2 x} - \frac{\lambda (y-x)^2}{2x^2 y}\right) \quad (2)$$

The joint PDF represents the likelihood of different pairs of values occurring together. It ensures that the sum of probabilities over all possible values equals 1, while also capturing the correlation and potential causal relationships between the variables. This understanding is essential for accurately modelling dependencies and interactions in various applications across fields such as finance, engineering, and biology.

Theorem 1: The normalization condition ensures that the joint PDF $f_{X,Y}(x,y)$ is non-negative for all x > 0 and y > 0 and the total probability over all possible values of X and Y sums to 1; indicating that, the joint distribution represents a valid probability density function given as:

$$\int_0^\infty \int_0^\infty f_{X,Y}(x,y)\, dxdy = 1 \qquad (3)$$

Proof:

To prove the normalization condition for BCIGD, we need to show that the double integral (4) of the joint PDF of equation (2) over all possible values of X and Y equals 1.

$$\int_0^\infty \int_0^\infty \frac{\lambda_X \lambda}{2\pi (xy)^{3/2}} \exp\left(-\frac{\lambda_X (x-\mu_X)^2}{2\mu_X^2 x} - \frac{\lambda_X (y-x)^2}{2x^2 y}\right) dxdy \quad (4)$$

$$\Rightarrow \int_0^\infty \left(\int_0^\infty \frac{\lambda_X \lambda}{2\pi (xy)^{3/2}} \exp\left(-\frac{\lambda_X (x-\mu_X)^2}{2\mu_X^2 x} - \frac{\lambda_X (y-x)^2}{2x^2 y}\right) dy \right) dx \quad (5)$$

The inner integral of (5) is given by:

$$\int_0^\infty \lambda^2 \frac{1}{2\pi} y^{\frac{3}{2}} \exp\left(-\frac{\lambda (y-x)^2}{2x^2 y}\right) dy$$

Revista ARACÊ
ISSN: 2358-2472

The expression inside the integral is the conditional PDF of the BCIGD with parameters x and λ:

$$\Rightarrow \quad f_{Y|X}\left(y|x\right)=\lambda^2\frac{1}{2\pi}y^{\frac{3}{2}}\exp\left(-\frac{\lambda\left(y-x\right)^2}{2x^2y}\right)(6)$$

Substituting equation (6) into (1), we obtain:

$$\int_0^\infty \lambda^2\frac{1}{2\pi}y^{\frac{3}{2}}\exp\left(-\frac{\lambda\left(y-x\right)^2}{2x^2y}\right)=1 \qquad (7)$$

Let $z=\frac{y}{x}$; then, y = xz and dy = xdz. The integral of (7) becomes:

$$\int_0^\infty \lambda^2\frac{1}{2\pi}xz^{\frac{3}{2}}\exp\left(-\frac{\lambda\left(xz-x\right)^2}{2x^3z}\right)=1 \qquad (8)$$

Simplifying the exponent of equation (8); we have that:

$$-\frac{\lambda\left(xz-x\right)^2}{2x^3z}=-\frac{\lambda\left(z-1\right)^2}{2xz}$$

Thus, the integral of equation (8) becomes:

$$\int_0^\infty \lambda^2\frac{1}{2\pi}x^{\frac{1}{2}}z^{\frac{3}{2}}\exp\left(-\frac{\lambda\left(z-1\right)^2}{2xz}\right)dz \qquad (9)$$

By definition, the inverse-Gaussian distribution integrates to 1 over its entire domain. The normalizing factor ensures that equation (3) is satisfied. Since the joint PDF of (2) is valid for BCIGD; then, equation (9) holds.

The outer integral of (5) is given by:

Revista ARACÊ
ISSN: 2358-2472

$$\int_0^{\infty} \frac{\lambda_X}{2\pi x^3} \exp\left(-\frac{\lambda_X (x-\mu_X)^2}{2\mu_X^2 x}\right) dx$$

This remaining integral is the marginal PDF $f_X(x)$ which follows the inverse-Gaussian distribution with parameters μ_X and λ_X:

$$f_X(x) = \frac{\lambda_X}{2\pi x^3} \exp\left(-\frac{\lambda_X (x-\mu_X)^2}{2\mu_X^2 x}\right)$$

Therefore;

$$\int_0^{\infty} \frac{\lambda_X}{2\pi x^3} \exp\left(-\frac{\lambda_X (x-\mu_X)^2}{2\mu_X^2 x}\right) dx = 1 \qquad (10)$$

The Inverse-Gaussian Distribution has a PDF that integrates to 1 over its entire domain by definition. The direct evaluation of the double integral of the joint PDF $f_{X,Y}(x,y)$ over all x and y confirms that it integrates to 1. This verifies that BCIGD is a valid probability density function that satisfies the normalization condition, ensuring that the total probability over all possible values of X and Y is equal to 1. This property makes the BCIGD a proper joint distribution for modelling the relationship between the two interdependent random variables X and Y.

PROPERTIES OF BCIGD

Marginal Distribution of Y

Theorem 2: The marginal distribution of Y under BCIGD can be found by integrating the joint PDF $f_{X,Y}(x,y)$ over all possible values of X.

Given the joint PDF:

$$f_{X,Y}(x,y) = \lambda_X \lambda^2 \sqrt{\frac{2}{\pi}} (xy)^{\frac{3}{2}} \exp\left(-\frac{\lambda_X (x-\mu_X)^2}{2\mu_X^2 x} - \frac{\lambda (y-x)^2}{2x^2 y}\right) (11)$$

We substitute the joint PDF into the integral to find $f_Y(y)$

Revista ARACÊ
ISSN: 2358-2472

$$f_Y(y) = \int_0^\infty \sqrt{\frac{2}{\pi}}(xy)^{\frac{3}{2}} \exp\left(-\frac{\lambda_X(x-\mu_X)^2}{2\mu_X^2 x} - \frac{\lambda(y-x)^2}{2x^2 y}\right) dx \quad (12)$$

Similarly, we can find the conditional distribution of X given Y = y using the joint PDF we need to integrate it with respect to x from 0 to ∞ to find $f_{X\backslash Y}(y)$.

However, this integral is quite complex and does not have a simple closed-form solution. Therefore, we might need to look for alternative methods or approximations depending on specific values of μ_X, λ_X, and λ.

If $\lambda_X = \mu_X = 1$, we can make use of the fact that $f_{Y|X}(y|x)$ is a Wald distribution, so $f_Y(y)$ can be found by integrating a Wald distribution over all possible values of X.

Let $f_{Y|X}(y|x)$ denote the conditional PDF of Y given X = x; then:

$$f_{Y|X}(y|x) = \frac{\lambda^{\frac{3}{2}}}{2\pi} \exp\left(-\frac{\lambda(y-x)^2}{2x^2 y}\right)$$

Let the marginal PDF of X be given by:

$$f_X(x) = \frac{\lambda_X^2}{2\pi x^3} \exp\left(-\frac{\lambda_X(x-\mu_X)^2}{2\mu_X^2 x}\right)$$

We now integrate $f_{Y|X}(y|x)$ over all possible values of X and then insert it into the expression for $f_Y(y)$:

$$\int_0^\infty f_{Y|X}(y|x)\, f_X(x)\, dx$$

This integral represents the convolution of $f_{Y|X}(y|x)$ and $f_X(x)$, which can be complex and may not have a simple closed-form solution.

However, if $\lambda_X = \mu_X = 1$, then $f_X(x)$ simplifies to the standard Inverse-Gaussian distribution, and we can use known properties of the Wald distribution to find $f_Y(y)$ by integrating over all possible values of X.

Revista ARACÊ
ISSN: 2358-2472

Joint Moment of X and Y

The joint moments of X and Y under BCIGD can be calculated using the joint PDF $f_{X,Y}(x,y)$. The n-th joint moment of X and Y is given by:

$$E[X^nY^m] = \int_0^\infty \int_0^\infty x^n y^m f_{X,Y}(x,y)\,dx\,dy \quad (13)$$

Where

$$f_{X,Y}(x,y) = \lambda_X\lambda^2 \frac{1}{2\pi}(xy)^{\frac{3}{2}} \exp\left(-\frac{\lambda_X(x-\mu_X)^2}{2\mu_X^2 x} - \frac{2x^2y\lambda(y-x)^2}{2}\right)$$

Simplifying the Integrand, we have

$$E[X^nY^m] = \frac{\lambda_X\lambda}{2\pi}\int_0^\infty \int_0^\infty x^{n-3/2}y^{m-3/2}(xy)^{3/2}\exp\left(-\frac{\lambda_X(x-\mu_X)^2}{2\mu_X^2 x} - \frac{\lambda(y-x)^2}{2x^2y}\right)dx\,dy$$

This can be rewritten as:

$$E\left[X^nY^m\right] = \frac{\lambda_X\lambda}{2\pi}\int_0^\infty\int_0^\infty (xy)^{3/2}\,x^{n-3/2}y^{m-3/2}\exp\left(-\frac{\lambda_X(x-\mu_X)^2}{2\mu_X^2 x} - \frac{\lambda_X(y-x)^2}{2x^2y}\right)dxdy \quad (14)$$

Applying numerical methods for specific cases of n and m and using arbitrary values for the parameters, we choose:

$\lambda X = 2, \qquad \lambda = 3, \ \mu X = 1.5$

The joint moment E[XY] is given by:

$$E[XY] = \int_0^\infty \int_0^\infty x\cdot y\cdot \frac{\lambda_X\lambda}{2\pi(xy)^{3/2}}\exp\left(-\frac{\lambda_X(x-\mu_X)^2}{2\mu_X^2 x} - \frac{\lambda(y-x)^2}{2x^2y}\right)dx\,dy$$

$$E[XY] = \int_0^\infty \int_0^\infty x\cdot y\cdot \frac{\lambda_X\lambda}{2\pi(xy)^{3/2}}\exp\left(-\frac{\lambda_X(x-\mu_X)^2}{2\mu_X^2 x} - \frac{\lambda(y-x)^2}{2x^2y}\right)dx\,dy$$

Substituting the chosen values into the equation, we obtain:

Revista ARACÊ
ISSN: 2358-2472

$$E[XY] = \int_0^\infty \int_0^\infty \frac{2 \cdot 3}{2\pi \cdot 1} \cdot (xy) \exp\left(-\frac{2(x-1.5)^2}{2 \cdot 1.5^2 x} - \frac{3(y-x)^2}{2x^2 y}\right) dx\, dy$$

$$E[XY] = \int_0^\infty \int_0^\infty \frac{6}{2\pi} \cdot (xy) \exp\left(-\frac{2(x-1.5)^2}{2 \cdot 1.5^2 x} - \frac{3(y-x)^2}{2x^2 y}\right) dx\, dy$$

This simplifies to:

$$E[XY] = \frac{3}{\pi} \int_0^\infty \int_0^\infty \frac{1}{xy} \exp\left(-\frac{(x-1.5)^2}{2 \cdot 1.125x} - \frac{3(y-x)^2}{2x^2 y}\right) dxdy$$

The numerical integration for the joint moment E[XY] with the given parameters to be:

$$E[XY] \approx 3.41$$

The calculation has a very small error margin of approximately 2.34×10^{-8}, indicating high accuracy.

Marginal Moment and Covariance

The marginal moment $E\left[X^n\right]$ is given by:

$$E[X^n] = \int_0^\infty x^n \cdot \frac{\lambda_X^2}{2\pi x^3} \exp\left(-\frac{\lambda_X (x-\mu_X)^2}{2\mu_X^2 x}\right) dx$$

The marginal moment $E[X^n]$ is obtained as follows from the Conditional Expectation $E[Y^m \mid X = x]$:

$$E[Y^m \mid X = x] = \int_0^\infty y^m \frac{\lambda^2}{2\pi y^{3/2}} \exp\left(-\frac{\lambda (y-x)^2}{2x^2 y}\right) dy.$$

To obtain the marginal moment $E[Y^m]$, we marginalize the conditional expectation over the marginal distribution of X, which has the PDF:

$$f_X(x) = \frac{\lambda_X^2}{2\pi x^3} \exp\left(-\frac{\lambda_X (x-\mu_X)^2}{2\mu_X^2 x}\right)$$

Revista ARACÊ
ISSN: 2358-2472

Thus, the marginal moment is:

$$E[Y^m] = \int_0^\infty E[Y^m \mid X = x] f_X(x)\, dx.$$
$$E[Y^m] = \int_0^\infty \left(\int_0^\infty y^m f_{Y|X}(y \mid x)\, dy \right) f_X(x)\, dx.$$

Combining the expressions and substitute the expression for $E[Y^m \mid X = x]$ into the above integral; we obtain:

$$E[Y^m] = \int_0^\infty \left(\int_0^\infty y^m \frac{\lambda^2}{2\pi y^{3/2}} \exp\left(-\frac{\lambda(y-x)^2}{2x^2 y} \right) dy \right) \cdot \frac{\lambda_X^2}{2\pi x^3} \exp\left(-\frac{\lambda_X (x-\mu_X)^2}{2\mu_X^2 x} \right) dx.$$
$$E[Y^m] = \int_0^\infty \left(\int_0^\infty y^{m-\frac{3}{2}} \frac{\lambda^2}{2\pi} \exp\left(-\frac{\lambda(y-x)^2}{2x^2 y} \right) dy \right) \cdot \frac{\lambda_X^2}{2\pi x^3} \exp\left(-\frac{\lambda_X (x-\mu_X)^2}{2\mu_X^2 x} \right) dx.$$

Substitute the expression for $E[Y^m \mid X = x]$ into the above integral; we obtain:

$$E[Y^m] = \int_0^\infty \left(\int_0^\infty y^m \frac{\lambda^2}{2\pi y^{3/2}} \exp\left(-\frac{\lambda(y-x)^2}{2x^2 y} \right) dy \right) \cdot \frac{\lambda_X^2}{2\pi x^3} \exp\left(-\frac{\lambda_X (x-\mu_X)^2}{2\mu_X^2 x} \right) dx.$$

Where the covariance of X and Y is given by:

$$\mathrm{Cov}(X,Y) = E[XY] - E[X]E[Y]$$

Given $E[X] = \mu_X$ and $E[Y] = \mu_X$, we have that:

$$E[XY] = E[XE[Y \mid X]] = E[X^2]$$

For a Wald distribution; $X \sim \mathrm{Wald}(\mu_X, \lambda_X)$:

$$\Rightarrow \quad E\left[X^2\right] = \mu_X^2 \left(1 + \frac{1}{\lambda_X}\right) \therefore \quad \mathrm{cov}(X,Y) = \mu_X^2 \left(1 + \frac{1}{\lambda_X}\right) - \mu_X \mu_X$$
$$\mathrm{cov}(X,Y) = \mu_X^2 \left(\frac{1}{\lambda_X}\right)$$

METHOD OF PARAMETER ESTIMATION

The parameter estimation of BCIGD can be derived using Maximum Likelihood Estimation.

Revista ARACÊ
ISSN: 2358-2472

Given the joint PDF of X and Y:

$$f_{X,Y}(x,y) = \frac{\lambda_X \lambda}{2\pi (xy)^{3/2}} \exp\left(-\frac{\lambda_X (x-\mu_X)^2}{2\mu_X^2 x} - \frac{\lambda (y-x)^2}{2x^2 y}\right)$$

Log-Likelihood Function

Let$\{(x_i, y_i)\}_{i=1}^{n}$ be a sample of *n* pairs of observations. The likelihood function L(μX,λX,λ) is the product of the joint PDFs for each observation:

$$L(\mu_X, \lambda_X, \lambda) = \prod_{i=1}^{n} f_{X,Y}(x_i, y_i)$$

The log-likelihood function $\ell(\mu_X,\lambda_X,\lambda)$ is:

$$\ell(\mu_X, \lambda_X, \lambda) = \sum_{i=1}^{n} \ln f_{X,Y}(x_i, y_i)$$

Substituting the joint PDF into the log-likelihood function:

$$\ell(\mu_X, \lambda_X, \lambda) = \sum_{i=1}^{n} \ln\left[\frac{\lambda_X \lambda}{2\pi (x_i y_i)^{3/2}} \exp\left(-\frac{\lambda_X (x_i-\mu_X)^2}{2\mu_X^2 x_i} - \frac{\lambda (y_i-x_i)^2}{2x_i^2 y_i}\right)\right]$$

Simplifying inside the logarithm:

$$\ell(\mu_X, \lambda_X, \lambda) = \sum_{i=1}^{n}\left[\ln(\lambda_X) + \ln(\lambda) - \ln(2\pi) - \frac{3}{2}\ln(x_i) - \frac{3}{2}\ln(y_i) - \frac{\lambda_X (x_i-\mu_X)^2}{2\mu_X^2 x_i} - \frac{\lambda (y_i-x_i)^2}{2x_i^2 y_i}\right]$$

Separating the sums:

$$\ell(\mu_X, \lambda_X, \lambda) = n\ln(\lambda_X) + n\ln(\lambda) - n\ln(2\pi) - \frac{3}{2}\sum_{i=1}^{n}\ln(x_i) - \frac{3}{2}\sum_{i=1}^{n}\ln(y_i) - \frac{\lambda_X}{2\mu_X^2}\sum_{i=1}^{n}\frac{(x_i-\mu_X)^2}{x_i}$$
$$-\frac{\lambda}{2}\sum_{i=1}^{n}\frac{(y_i-x_i)^2}{x_i^2 y_i}$$

Revista ARACÊ
ISSN: 2358-2472

Derivatives of the Log-Likelihood Function

To find the maximum likelihood estimates (MLEs) of μ_X, λ_X, and λ, we take the partial derivatives of $\ell(\mu_X, \lambda_X, \lambda)$ with respect to each parameter, set them to zero, and solve for the parameters.

Partially differentiating with respect to μ_X and setting it to zero; we obtain:

$$\frac{\partial \ell}{\partial \mu_X} = -\frac{\lambda_X}{2\mu_X^2}\sum_{i=1}^{n}\frac{(x_i-\mu_X)x_i}{x_i} = -\frac{\lambda_X}{2\mu_X^2}\sum_{i=1}^{n}(x_i-\mu_X)$$

$$-\frac{\lambda_X}{2\mu_X^2}\sum_{i=1}^{n}(x_i-\mu_X) = 0$$

$$\sum_{i=1}^{n}(x_i-\mu_X) = 0$$

$$\sum_{i=1}^{n}x_i - n\mu_X = 0$$

$$\mu_X = \frac{1}{n}\sum_{i=1}^{n}x_i$$

Partially differentiating with respect to λ and setting it to zero; we obtain:

$$\frac{\partial \ell}{\partial \lambda} = \frac{n}{\lambda} - \frac{1}{2}\sum_{i=1}^{n}\frac{(y_i-x_i)^2}{x_i^2 y_i}$$

$$\frac{n}{\lambda} - \frac{1}{2}\sum_{i=1}^{n}\frac{(y_i-x_i)^2}{x_i^2 y_i} = 0$$

$$\lambda = \frac{2}{n}\sum_{i=1}^{n}\frac{(y_i-x_i)^2}{x_i^2 y_i}$$

Therefore, the maximum likelihood estimates for the parameters μ_X, λ_X, and λ are:

$$\hat{\mu}_X = \frac{1}{n}\sum_{i=1}^{n}x_i$$

$$\hat{\lambda}_X = \frac{2}{n}\sum_{i=1}^{n}\frac{(x_i-\hat{\mu}_X)^2}{x_i}$$

$$\hat{\lambda} = \frac{2}{n}\sum_{i=1}^{n}\frac{(y_i-x_i)^2}{x_i^2 y_i}$$

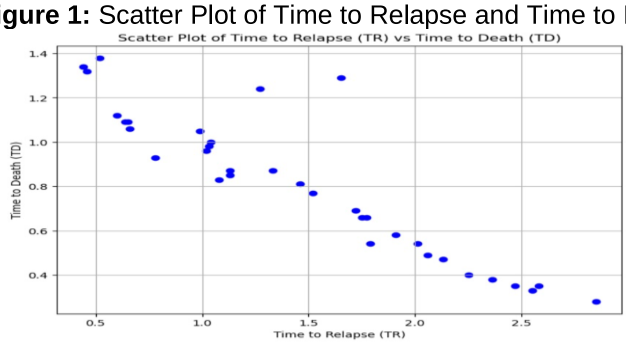
Revista ARACÊ
ISSN: 2358-2472

RESULTS AND DISCUSSION

The graphics, probabilities and consistent measures of the regression were acquired by fitting the imprinting of Python and R programming languages.

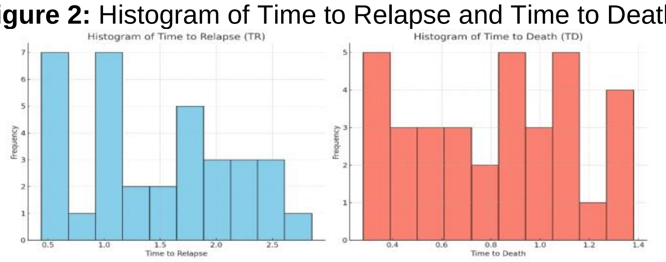
The scatter plot of figure 1 identifies and visualizes relationships between TR and TD. We also observed outlier points for some TD that has a much larger TR than the others. Again, frequently moderate negative linear correlation between TR and TD was pragmatic. Thus, as TR increases, TD tends to decrease. The inverse relationship suggests that patients with a longer TR tend to have shorter TD and vice versa. This pattern aligns with the survival analysis expectations in medical context. It implies that the occurrence of a relapse earlier in the treatment might lead to a prolonged survival time, while a delayed relapse associated with a shorter survival time following that relapse. The relationship is also consistent with the bivariate conditional inverse-Gaussian distribution's nature, capturing interdependences between two time-to-event variables in a way that reflects their joint probability behaviour.

Figure 1: Scatter Plot of Time to Relapse and Time to Death



The histograms of Time to Relapse and Time to Death in figure 2, demonstrated the shape of the distribution of each variable separately. It can be seen that TR has a bimodal distribution with two peaks; while, TD has a multimodal distribution with three peaks.

Figure 2: Histogram of Time to Relapse and Time to Death



Revista ARACÊ
ISSN: 2358-2472

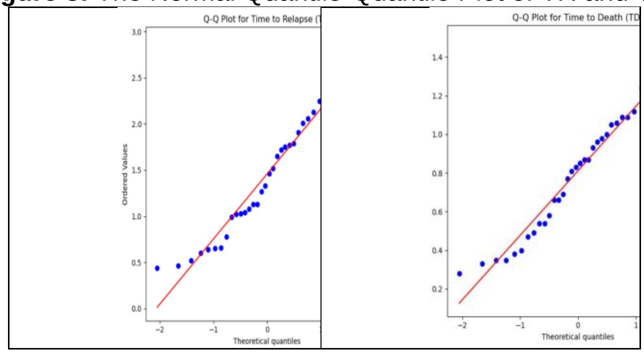
The normal quantile-quantile (QQ) plot of figure 3 compares the quantiles of numeric variables of TR and TD to the quantiles of a normal distribution and their linear relationship. The regression line showed a positive linear trend in TR and TD datasets. In a Q-Q plot, if the lower end deviates from the straight line, it indicates that the distribution has heavier tails (more extreme values) or lighter tails (fewer extreme values) in the lower end compared to a normal distribution.

It can be seen that the bottom end of the Q-Q plot for TR deviates from the straight line but the upper end does not; thus, the distribution has a longer tail to its left. In order words, it is left-skewed **or** negatively skewed.

Again, the distribution of TD has a fat tail since both ends of the Q-Q plot deviate from the straight line while its center follows a straight line. The points at the upper end are curving upward which suggests a heavy-tailed distribution, meaning there are more extreme values in the data than would be expected under a normal distribution. This might indicate positive skewness or an excess of high outliers.

Also, the points at the lower end are curving downward which suggests a light-tailed distribution, where there are fewer extreme values than expected. This could indicate negative skewness or a lack of high outliers.

Figure 3: The Normal Quantile-Quantile Plot of TR and TD



The Kolmogorov-Smirnov (KS) tests were carried out in table 1 in order to confirm the distributional properties of TR and TD. The Kolmogorov-Smirnov tests suggested that we cannot reject the null hypothesis for either dataset, suggesting that both TR and TD data follow a normal distribution. These findings indicate that assuming normality for these datasets is statistically justified.

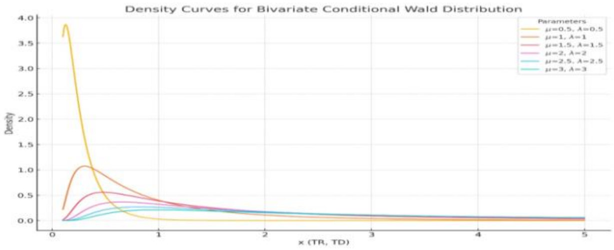
Revista ARACÊ
ISSN: 2358-2472

Table 1. Kolmogorov-Smirnov tests

KS Statistic for TR	P-Value
0.12626939063412734	0.6056346549803906
KS Statistic for TD	P-Value
0.09383937665287839	0.8984092162458037

The peak in the 3D plot where the contour is highest indicates the region with the greatest probability density. This suggests a specific combination of TR and TD values, approximately around $TR \approx 0.5$ and $TD \approx 0.5$ where the likelihood is the highest. This peak area represents the most probable values for relapse and time to death, given the parameters of the bivariate conditional inverse-Gaussian distribution.

Figure 4: Density Curve of the Bivariate Conditional Inverse-Gaussian Distribution for various Parameter Values



The 3D contour plot for the dataset in figure 5 demonstrates the behavior of the bivariate conditional inverse-Gaussian distribution, where high density is concentrated around shorter times for both TR and TD, and density decline rapidly as these times increase. This is typical in survival models where earlier events are more probable than elongated consequences.

Figure 5: 3D Contour Plot for the Dataset

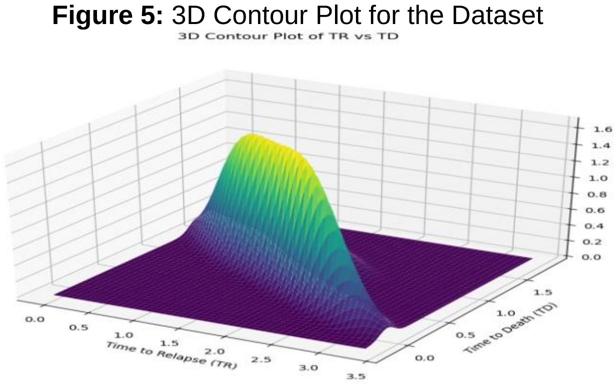


Table 2 below shows the cumulative probability values (integral values) for the distribution under different parameter settings.

Revista ARACÊ
ISSN: 2358-2472

Table 2. The results of numerical integration of the area under curve of the marginal distribution y for Fig. 1

μ_TR	μ_TD	λ_TR	λ_TD	Integral Value
0.5	0.5	1	1	0.9796
0.5	0.5	1	2	0.9895
0.5	0.5	2	1	0.9895
0.5	0.5	2	2	0.9994
0.5	1	1	1	0.9394
0.5	1	1	2	0.9685
0.5	1	2	1	0.9488
0.5	1	2	2	0.9782
0.5	1.5	1	1	0.8642
0.5	1.5	1	2	0.8866
0.5	1.5	2	1	0.8728
0.5	1.5	2	2	0.8955
1	0.5	1	1	0.9394
1	0.5	1	2	0.9488
1	0.5	2	1	0.9685
1	0.5	2	2	0.9782
1	1	1	1	0.9008
1	1	1	2	0.9287
1	1	2	1	0.9287
1	1	2	2	0.9574
1	1.5	1	1	0.8287
1	1.5	1	2	0.8502
1	1.5	2	1	0.8543
1	1.5	2	2	0.8765
1.5	0.5	1	1	0.8642
1.5	0.5	1	2	0.8728
1.5	0.5	2	1	0.8866
1.5	0.5	2	2	0.8955
1.5	1	1	1	0.8287
1.5	1	1	2	0.8543
1.5	1	2	1	0.8502
1.5	1	2	2	0.8765
1.5	1.5	1	1	0.7623
1.5	1.5	1	2	0.7821
1.5	1.5	2	1	0.7821
1.5	1.5	2	2	0.8025

CONCLUSIONS

The analysis of time to relapse and time to death data using the bivariate conditional inverse-Gaussian distribution provides valuable insights into the probabilistic structure of these medical events. The contour plot reveals that the highest probability density is

Revista ARACÊ
ISSN: 2358-2472

centered on specific TR and TD values with density rapidly decreasing as either variable increase, reflecting the Invers-Gaussian distribution's nature. The Kolmogorov-Smirnov tests confirmed that both TR and TD datasets are normally distributed which validates the statistical assumption used in this analysis. The numerical integration values obtained for various parameters accentuate the tractability and applicability of the bivariate inverse-Gaussian distribution in capturing complex interdependent survival times in a medical perspective. This finding supports the use of inverse-Gaussian distribution in this area and highlights the potential numerical integration as a tool for better understanding probability density behavior in survival analysis. The numerical integration that was performed over a range of parameters for the bivariate conditional inverse-Gaussian distribution provides integral values across selected TR and TD ranges. These values offer a quantitative understanding of the probability density behaviour and serve as a foundation for further research into survival analysis and medical event modelling.

The analysis of time to relapse and time to death data using the bivariate conditional inverse-Gaussian distribution provides valuable insights into the probabilistic structure of these medical events. The contour plot reveals that the highest probability density is centered on specific TR and TD values with density rapidly decreasing as either variable increase, reflecting the Invers-Gaussian distribution's nature. The Kolmogorov-Smirnov tests confirmed that both TR and TD datasets are normally distributed which validates the statistical assumption used in this analysis. The numerical integration values obtained for various parameters accentuate the tractability and applicability of the bivariate inverse-Gaussian distribution in capturing complex interdependent survival times in a medical perspective. This finding supports the use of inverse-Gaussian distribution in this area and highlights the potential numerical integration as a tool for better understanding probability density behavior in survival analysis. The numerical integration that was performed over a range of parameters for the bivariate conditional inverse-Gaussian distribution provides integral values across selected TR and TD ranges. These values offer a quantitative understanding of the probability density behaviour and serve as a foundation for further research into survival analysis and medical event modelling.

Revista ARACÊ
ISSN: 2358-2472

REFERENCES

1. Athanassoulis, G. A., Skarsoulis, E. K., & Belibassakis, K. A. (1994). Bivariate distributions with given marginals with an application to wave climate description. Applied Ocean Research, 16(1),1–17. https://doi.org/10.1016/0141-1187(94)90010-8

2. Edwards, J. (2023, September 18). Leveraging predictive analytics for risk management. LinkedIn. https://www.linkedin.com/pulse/leveraging-predictive-analytics-risk-management-dr-jeffrey/

3. Gongsin, I. E., & Saporu, F. W. O. (2020). A bivariate conditional Weibull distribution with application. *African Matematics, 31*(3), 565–583. https://doi.org/10.1007/s13370-019-00742-8.

4. Haberl, Helmut & Steinberger, Julia & Plutzar, Christoph & Erb, Karl-Heinz & Gaube, Veronika & Gingrich, Simone & Krausmann, Fridolin. (2012). Natural and socioeconomic determinants of the embodied human appropriation of net primary production and its relation to other resource use indicators. Ecological indicators. 23. 222-231. 10.1016/j.ecolind.2012.03.027.

5. Johnson, N. L., Kotz, S., & Balakrishnan, N. (1995). *Continuous univariate distributions, volume 2* (Vol. 289). John wiley & sons. Kozubowski, T. J., & Podg´orski, K. (2018). A comprehensive study of multivariate distributions with applications. *Statistical Science*, 33(4), 548-563.

6. Jondeau, E., & Rockinger, M. (2006). The copula-GARCH model of conditional dependencies: An international stock market application. Journal of International Money and Finance, 25(5), 827–853. doi.org/10.1016/j.jimonfin.2006.04.007 https://doi.org/10.1016/j.jimonfin.2006.04.007

7. Kynigakis, I., & Panopoulou, E. (2023). Modeling the distribution of key economic indicators in a data-rich environment: New empirical evidence.*SSRN*. https://doi.org/10.2139/ssrn.4606867

8. McCarthy, R. V., McCarthy, M. M., Ceccucci, W., Halawi, L., McCarthy, R. V., McCarthy, M. M., & Halawi, L. (2022). Applying predictive analytics (pp. 89–121). Springer International Publishing. https://doi.org/10.1007/978-3-030-83070-0

9. Pham, H. (2022). *Statistical reliability engineering*. Springer International Publishing. https://doi.org/10.1007/978-3-030-76904-8